IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 18 September 2026, accepted 3 October 2026, date of publication 8 October 2026, date of current version 19 October 2026.

Digital Object Identifier 10.1109/ACCESS.2026.3741782

RESEARCH ARTICLE

Separability by Design: Decoupling Classification and Localization Quality in Anchor-Free Object Detection

HADI MAT ROSLY[1], ZULKIFLI ZAINAL ABIDIN[2], AND HASAN FIRDAUS MOHD ZAKI[1,3], (Member, IEEE)

[1]Department of Mechatronics Engineering, International Islamic University Malaysia, Gombak, Kuala Lumpur 53100, Malaysia
[2]Integrated Engineering Department, Faculty of Engineering, Universiti Teknologi PETRONAS, Seri Iskandar, Perak 32610, Malaysia
[3]Centre for Unmanned Technologies (CUTe), International Islamic University Malaysia, Gombak, Kuala Lumpur 53100, Malaysia

Corresponding authors: Hasan Firdaus Mohd Zaki (hasanzaki@iium.edu.my) and Zulkifli Zainal Abidin (zulkifli.zainal@utp.edu.my)

The authors gratefully acknowledge the financial support of Universiti Teknologi PETRONAS (UTP), Malaysia, through the AMGTech Research Centre OPEX Incentive Grant (PGR Cost Centre No. 015J00-001) and the Centre for Graduate Studies (Cost Centre No. 015BD1-001). The authors also thank the Research Management Centre of International Islamic University Malaysia (IIUM) for funding through the Publication Facilitation Fee Scheme, and the Ministry of Higher Education Malaysia for support through the Fundamental Research Grant Scheme (FRGS) (Grant ID: FRGS/1/2025/ICT02/UIAM/02/1).

ABSTRACT Object detection research validates architectural decisions almost exclusively through mean average precision (*mAP*), a metric that certifies performance but conceals mechanism. We argue that principled improvement requires a detector whose classification and localization pathways are disentangled and therefore individually diagnosable. Starting from the convolutional process itself, we show that depth trades positional capacity for featural dimensionality at fixed rates set by the stage schedule, so a detection head reading both targets from one point accepts whatever proportion its depth supplies. From this decomposition we derive separability by design: classification and localization quality placed in architecturally distinct, separately supervised pathways, with an objectness route admitting localization evidence into the semantic score. Three falsifiable predictions follow, tested across eight configurations crossing head structure, box geometry and gradient coupling. All are supported. A quality field read where positional capacity survives raises its rank correlation with realized *IoU* from 0.051 to 0.484; two levels recover the localizability of a five-level pyramid, at 0.943 against 0.950 oracle recall; and the convolutional head commits its class field in the forward pass, reducing candidate entropy from 3.66 to 2.22 bits and uncorroborated rare survivors from 3.2% to 0.9% across 5000 COCO val2017 images. The prediction concerning the objectness route holds conditionally: the gate lifts the class score only where the class head can read the spatial neighborhood the objectness cue occupies, which requires both a convolutional head and a cue-preserving geometry. The contribution is a framework and a diagnostic method for mechanistic detector analysis that enables systematic architectural design, not a leaderboard result.

INDEX TERMS Object detection, anchor-free detection, feature entanglement, localization quality, model interpretability, error diagnosis.

I. INTRODUCTION

Object detection research has organized itself around a single metric benchmark. Architectural decisions, such as head design, assignment rules, loss shaping, and score fusion, are proposed, ablated, and accepted or discarded according to their effect on mean average precision (*mAP*). The metric is a fine certificate of *performance*, but it is silent on *mechanism*: it cannot say why a model fails, which component carries a weakness, or whether an improvement repaired a defect or papered over it. A research framework that aims to improve a detector by unfolding its mechanism, rather than by undirected metric search, needs instruments that *mAP* does not provide.

Surveying the ablation literature of single-stage detection, one structural trend recurs across otherwise diverse designs:

The associate editor coordinating the review of this manuscript and approving it for publication was Sawyer Duane Campbell.

© 2026 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

classification and box regression are computed from shared features. From SSD [1] and YOLO [2] through RetinaNet [3] and FCOS [4], the two task heads attach in parallel to the same feature map at each pyramid level. Later refinements, such as decoupled towers [5], task alignment [6], and quality fusion [3], [4], change how that shared representation is read, but none question whether it should be shared at all. Sharing remains a default rather than a derived property: no founding principle establishes that one representation is at once efficient and effective for two objectives as different as classification and box regression. This gap remains obscured by *mAP*, whose single aggregate score cannot reveal whether the shared features serve each head or only their average.

Prior interpretability work suggests it does not. Deep feature factorization (DFF) [7] decomposes learned detector features into a small number of concept directions, and applying it layer by layer shows how a network progressively segregates information as depth and stride grow. Read through the lens of information theory, the picture is one of competition: the class-relevant and geometry-relevant components of a shared feature contend for the same representational capacity, and the observed factorizations are consistent with an allocation that is neither controlled nor optimal for either task. Section III formalizes this with a chain-rule decomposition of the class information: a shared feature may segregate the two terms cleanly, but nothing in the training objective makes it do so, and nothing in *mAP* lets us check.

Entanglement is not merely an analysis inconvenience; it is a plausible fragility. A classifier whose class evidence is interwoven with localization geometry in one feature has no clean categorical posterior to fall back on: small input perturbations arising from distributional shift, corruption, or adversarial nudges can move the shared feature in directions that skew class inference even when the object's geometry is intact. Misclassification under out-of-distribution conditions would then be a structural consequence of where the two vector spaces live and compete rather than a probabilistic failure. This paper does not test that consequence. What it establishes directly is the entanglement signature itself, and the corruption results of § VI are reported as a null on the question rather than as support for it.

The signature is visible in the score field itself. Figure 1 plots FCOS's joint score against its maximum class logit and class probability for a single image: the score rises smoothly from a dense low-score mass with no separation anywhere along its range. Behind that mass is a diffuse distribution over categories, dozens of classes in play and many of them closely related, which no scoring threshold separates cleanly. We call this the spray strategy, or hedging. It is effective for *mAP*, because the threshold hides the dispersion, but the class decision is never committed in the head: misclassification lives in a threshold cutoff rather than in a representation one can supervise, and the closely related candidates that spray produces are the seeds from which confusion grows. None of this is visible in *mAP*.

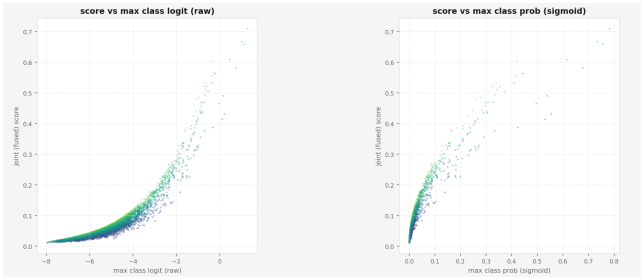


FIGURE 1. Joint score against maximum class logit and class probability for FCOS, one image. No separation along the score range: the class decision is deferred to the threshold rather than made in the head.

This builds our core motivation, in which we propose **separability by design**: classification and localization quality placed in architecturally distinct, separately supervised pathways, combined only at inference, so that the model's feature diagnostics disentangle: each pathway individually measurable, each weakness attributable, each repair targeted. Because a single architecture derived from a principle might succeed by coincidence, we verify it with our model, the Acyclic Integrated Concept (AIC), in the convolutional realization AIC-Conv under four configurations and, as a controlled verification, an FC classification head matched to FCOS (AIC-FC) under the same four, eight configurations in all. The theory issues three falsifiable predictions (§ III-A), each with a stated failure condition, all of which the experiments support.

Our contributions are:

1) A separability-by-design framework grounded in an information-theoretic decomposition of the convolutional process, which shows depth trading positional capacity for featural dimensionality at rates set by the stage schedule, and from which three falsifiable predictions follow. The predictions are tested across eight configurations crossing head structure, box geometry and gradient coupling.
2) A three-instrument diagnostic methodology, comprising score–quality rank correlation, candidate-class field analysis, and a rare-survivor audit, that separates localization from ranking and locates where semantic commitment occurs. Each instrument is tied to a stated prediction.
3) A per-level oracle-recall decomposition of the baseline pyramid, showing that contribution falls monotonically with depth and that the deepest levels are the weakest single-level performers, while the separated design recovers the localizability of the full pyramid from a reduced set of levels placed where positional capacity survives. The reduction is therefore an alignment with where localization is effective, not a concession against a larger pyramid.
4) Cross-model results over the val2017 set establishing that the convolutional design achieves the separation: the box head rather than the class score orders localization, the candidate class field commits in the forward

pass rather than at suppression, and uncorroborated survivors are substantially reduced. Adverse findings are reported alongside: localization error rising under separation, missed-object error rising for AIC-Conv, and corruption retention remaining unchanged.

II. RELATED WORK

A. SINGLE-STAGE DETECTION, THE SHARED HEAD, AND HEAD GEOMETRY

Single-stage detectors predict class and box jointly across a feature pyramid: SSD [1] and YOLO [2] attach classification and regression to a shared convolutional feature; RetinaNet [3] adds focal loss and parallel sub-networks over FPN [8] levels; FCOS [4] removes anchors, regresses four point-to-edge distances under an IoU-based regression loss [9], and multiplies a centerness term into the class score. YOLOX [5] decouples the classification and regression towers, OTA [10] formulates label assignment as optimal transport, and TOOD [6] aligns the two tasks explicitly, but both retain the shared input feature and a fused ranking score. Within this lineage, the spatial parameterization of the classification head is an under-examined variable. A fully-connected classification relation evaluates each candidate in isolation: it severs the class decision from the spatial neighborhood that produced the box, altering the classification–localization relation at the root. A convolutional 3×3 relation, by contrast, retains part of the localization context in the class pathway depending on the layer depth and stride at which it operates, since receptive-field growth trades spatial specificity for semantic abstraction. This distinction, largely invisible to mAP, is one we examine directly: in the oracle-recall analysis of § V-B (where localizability turns out to be nearly matched) and in the class-field structure of § V-D (where it is not).

B. MEASUREMENT INITIATIVES: ENTROPY AND RANK-CORRELATION DIAGNOSTICS

A smaller literature measures detectors rather than only scoring them. IoU-Net [11] introduced predicted-vs-actual IoU as a measurable quantity, later extended to single-stage detection [12], establishing rank agreement between a score and true localization quality as a diagnostic; TOOD [6] quantifies classification–localization misalignment per instance; and GFLv2 [13] reads localization quality off the statistics (effectively the entropy) of the learned box distribution, which is evidence that distributional shape carries diagnostic signal. We extend this measurement tradition in two directions: Spearman rank correlation between each head's score and realized IoU, per size bin, as a pathway-attribution instrument (§ IV-A); and Shannon entropy of the predicted-category distribution, at both the candidate and survivor stages, as a direct measure of where semantic commitment occurs (§ IV-B). Unlike fusion-oriented uses of these quantities, our aim is not a better ranking score but a measurable model.

C. ROBUSTNESS, INTERPRETABILITY, AND THE SEEDS OF MISCLASSIFICATION

Benchmarking under corruption [14] established that clean mAP overstates capability under distributional shift and established mean/relative performance under corruption (mPC/rPC) as the field-standard measures. We adopt these (§ VI) and connect them to a mechanism-level observation: a model that sprays closely-related classes to secure mAP carries within its candidate field the confusions it will later commit. This seeding is invisible in mAP, which rewards the hedging; it surfaces only in error decomposition and in our categorical diagnostics: as false-positive burden and quality-carried rare survivors (§ V-D) rather than in any aggregate score. Deep feature factorization [7] provides the representation-level counterpart: shared features factor into competing concept directions, motivating the expectation that entangled class evidence is susceptible to perturbation.

D. ERROR DECOMPOSITION WITH TIDE

TIDE [15] decomposes detection error into typed contributions (classification, localization, both, duplicate, background, missed) plus special false-positive and false-negative terms, each expressed as the AP gained if that error type were repaired (dAP). It is the natural complement to our diagnostics: where our instruments characterize how a model earns its detections (committed vs. sprayed), TIDE prices what the strategy costs. One caution attends cross-model TIDE reading: dAP terms are relative to each model's own AP, so the classification column is confounded by baseline accuracy and cross-model claims are safest on the false-positive/false-negative structure, where the spray signature is direct. This is important because false-positive burden is where system architecture integration matters most.

E. COLLABORATIVE AND MUTUAL-GUIDANCE LEARNING

A parallel line of work improves learners by coupling them. Co-training [16] trains separate learners on distinct views and lets each expand the other's training set; tri-training [17] generalizes this to three classifiers, labeling an example for one when the remaining two agree; and deep mutual learning [18] replaces one-way teacher-to-student transfer with an ensemble of peers that teach each other throughout training. Confucius tri-learning [19] is a recent instance in this lineage, training two classifiers and a generator jointly so that each classifier learns from the good examples its peer supplies and from the bad examples the generator supplies.

The three frameworks differ less in whether they add complementary supervision than in where it originates. Mutual guidance derives it from a peer learner's predictions and adversarial schemes from a synthesized distribution, whereas separability derives it from the ground-truth box geometry: a class-agnostic quality scalar supervised toward the IoU of the decoded box, made available to the classification pathway through the soft gate. The complementary signal is therefore already present in the annotation and merely unused on the class side by the conventional design, which is why it can

be introduced without an additional learner, an additional objective over labels, or generated data.

The coupling is also differently shaped. Mutual guidance is symmetric, between learners solving the same task and exchanging distributions over labels; after co-adaptation their contributions are not separately attributable, nor need they be, since accuracy is the objective. The soft gate is a deliberately impoverished channel by comparison, carrying a single scalar between pathways with different supervision targets. We do not forgo co-adaptation: both realizations deploy the gate without a stop-gradient, so classification reshapes the quality field it reads, and § V-C prices that choice directly. What the asymmetry preserves is not the absence of exchange but its attributability, since quality remains supervised toward IoU, classification toward class labels, and each readout separately measurable, so the cost of co-adaptation is a quantity we report rather than a confound we inherit. Mutual guidance and separability are complementary frameworks, the former adding coupling to buy accuracy and the latter shaping coupling to keep it attributable, and a combination would have to supervise the pathways jointly while keeping their readouts distinct.

III. THE SEPARATION FRAMEWORK

A. AN INFORMATION-THEORETIC FOUNDATION THAT MAKES PREDICTIONS

The framework originates in an observation about the convolutional process in the backbone. As depth increases, successive convolution and pooling operations enrich the channel dimension while contracting the spatial dimensions, increasing the residual uncertainty from which box coordinates must be recovered. The representation delivered to a coupled head is therefore an entanglement of classification and localization information, and disentangling the two is the founding motivation of the separation concept.

For a prediction Y drawn from a single undivided read-out over the tensor $X_i(F, L)$, the chain rule gives

$$I(Y; F, L) = I(Y; F) + I(Y; L|F) \tag{1}$$

The two terms are the contributions of the featural and spatial axes respectively, the second conditioned on the first. A coupled head reads both terms at whatever proportion obtains at its chosen depth, and accepts that proportion for both of its tasks. Separating the classification and box heads removes this constraint. The two predictions may then be drawn from different depths, allowing the box head to read from early layers where positional capacity remains high, and the class head from deeper layers where featural dimensionality is sufficient for category distributions to separate. A further consequence is that the class head, no longer required to support box regression, may be supplemented with objectness or quality information to sharpen the scope of the features it depends on. Three hypotheses were established from this concept.

H-1) Sharper localization quality from spatially enriched features: If information supporting box quality is concentrated where positional capacity is largest, a read-out drawn from that regime should yield a quality estimate more tightly correlated with realized IoU than one drawn from a representation shared with classification. A tighter correlation admits a lower IoU threshold at matched selection behavior.

H-2) Improved candidate ranking from quality coupling in the classification head: If the quality artifact produced by the box head carries localization information the class head cannot derive from its own axis, then a class head trained with that input coupled should show closer rank correspondence between its score and realized IoU than one trained with the input detached. The comparison is paired within variant across head type (FC, Conv) and box parameterization (ltrb, center+size), each detached configuration serving as the control for its non-detached counterpart. Because random initialization admits some spread between runs, a signed difference of at least 0.05 is required before a pair is read as informative: positive indicates coupling improves rank correspondence, negative that it degrades it.

H-3) Reduction of readout uncertainty: If the classification read-out is relieved of carrying localization information, supplying it instead with a localization factor from a separated head should reduce the uncertainty of the candidate class field, measured as Shannon entropy in bits, and reduce sporadic non-correlative leak-out, measured as the rare-survivor fraction (§ V-D).

These are directional predictions about variation relative to the principled design, not about absolute performance. Section V reports the outcomes.

B. OBJECTNESS AS A SOFT BRIDGE BETWEEN THE TWO PATHWAYS

Decoupling the heads separates where each target's information is read, but does not require the two pathways to stay isolated. The box-quality field provides a route between them that is narrower than a shared representation: the quality map is concatenated into the classification stream, so the classification logit becomes a downstream function of localization evidence. The route is typed here not as an architectural leak but as the model-space realization of objectness, a channel through which localization evidence is admitted into the semantic score.

The route carries gradient as well as activation. With the concatenation left undetached, the classification loss propagates into the box head, so the quality field is shaped toward the localization evidence that bears on the semantic score rather than toward geometric fit alone. The class head then conditions on a selected signal instead of drawing on the featural axis in full. A stop-gradient severs the adaptation while leaving the quality map readable, isolating the shaping of the field from its availability.

C. REALIZATIONS ACROSS A FACTORIAL DESIGN

All configurations share a ResNet-50 [20] + FPN backbone pretrained on COCO [21], operate on two pyramid levels

(strides 8 and 16), and use a decoupled box head supervised toward the *IoU* of the decoded box. Classification reads the box head's quality map through a soft gate that concatenates it onto the classification features; a single detach makes the gate read-only, so classification conditions on quality without reshaping it. Two factors are varied independently, the classification head and the box geometry, so that their effects remain attributable rather than confounded (Table 1).

TABLE 1. Factorial design configurations.

Component	AIC-FC	AIC-Conv
Classification head	fully-connected (per-box)	convolutional (3×3)
Box geometry	Original: Edge-distances (l,t,r,b)	Original: center+size (dx,dy,dw,dh)
	α: center+size (dx,dy,dw,dh)	**α**: edge-distances (l,t,r,b)
Backbone / FPN / levels	ResNet-50 + FPN, strides 8/16	ResNet-50 + FPN, strides 8/16

AIC-Conv, the principled design, uses a 3 × 3 convolutional classification head and is run in both geometries, center+size and the edge-distance variant (α), each with the gate coupled and detached. The center+size geometry is expected to keep objectness legible at each anchor point.

AIC-FC, the controlled variation, uses a fully connected per-box classification head and is run in both geometries: the canonical edge-distance parameterization, which by construction discards the anchor-point centroid objectness cue, and a center+size variant (α) that restores that cue at the anchor point. The two geometries separate the head factor from the box factor, since edge-distance withholds the objectness cue at the geometry level, while center+size supplies it but leaves the per-box head with no spatial neighborhood over which to read it. AIC-FC is therefore the configuration in which the principle is unsatisfied on the head side even where the geometry permits the cue, and it is subject to H-1 through H-3.

The design is examined in three stages. The first establishes whether separation holds at all: FCOS, AIC-Conv and AIC-FC are compared on the alignment between quality and realized *IoU*, testing whether a read-out drawn from the spatially dominated regime yields a geometric estimate better matched to attained overlap than one drawn from a shared representation. The second turns to the classification side, comparing AIC-FC against AIC-Conv on the uncertainty of the candidate class field in bits, so that the effect of head structure on readout confidence is assessed independently of the box pathway. The third varies the objectness route within each variant, comparing detached against non-detached configurations to establish whether the quality field must be shaped by the classification loss or need only be available to it.

IV. DIAGNOSTIC METHODOLOGY

Aggregate *mAP* cannot reveal whether an error is one of localization or of ranking, where a model commits semantically, or how it earns its detections. We therefore use three instruments, each tied to a stated prediction and computed from raw predictions before NMS except where noted.

A. SCORE–QUALITY RANK CORRELATION

Over matched predictions we compute Spearman correlation between realized *IoU* and the classification score, and between realized *IoU* and the quality score, attributing localization ranking to a pathway. The two are reported separately rather than combined, since a fused score depends on the combination rule and would confound the information content with that choice. This instrument tests H-1 and H-2.

B. CANDIDATE-CLASS FIELD ANALYSIS

For each image we characterize the distribution over categories of the pre-NMS candidates, reporting the effective support K as the number of distinct classes present and the Shannon entropy H in bits. These separate a sprayed field that keeps the full label set in play from a committed field concentrated on a few categories, and supply the readout-uncertainty measure of H-3.

C. RARE-SURVIVOR AUDIT

Comparing the candidate set against the surviving set, we report the fraction of surviving boxes in classes carrying six or fewer boxes on the image, together with their mean class confidence and quality. These are detections that survive without corroboration from the candidate field, and they supply the leak-out measure of H-3.

V. EXPERIMENTS AND DISCUSSION

A. SETUP AND PROTOCOL

1) TRAINING PROTOCOL

All variants were trained on MS-COCO train2017 and evaluated on val2017, covering the factorial configuration of § III-C: the convolutional realization (AIC-Conv) in both box geometries, center+size and edge-distance, each with and without the detached gate, plus a fully-connected variant (AIC-FC) matched to FCOS. A preliminary run with backbone, FPN, and class head frozen isolated the localization readout ahead of any joint training. The two heads differ in initialization as well as structure: AIC-FC inherits the pretrained FCOS classification head, already entangled with localization through centerness, whereas AIC-Conv trains its classification head from scratch.

AIC-FC was trained on a frozen ResNet-50 + FPN, first as detached and non-detached sub-variants (four epochs each). The non-detached configuration was additionally carried into a second phase of four epochs (two at 1×10^{-3}, two at 1×10^{-4}), eight in total, to probe extended optimization. Over the reduced-rate continuation the validation classification loss fell monotonically while *mAP* fell from its peak of 0.245 to 0.224; selection was therefore governed by the task metric rather than the loss, and the peak checkpoint was carried forward. Even at that checkpoint the classification logits showed no clean separability. Non-detached AIC-FC therefore serves as the controlled variant against

which AIC-Conv is compared, the latter developed from the separability concept rather than from further ablation of the FC head.

AIC-Conv was trained in three stages: heads only on a frozen backbone and FPN at 3×10^{-4} (epochs 1–4); backbone released at 1×10^{-5} with heads at 1×10^{-4} (epochs 5–10); both rates reduced by an order of magnitude (epochs 11–13). The epoch-12 checkpoint is evaluated throughout; epoch 13 differs by less than 0.001 *mAP*.

2) BASELINE AND INFERENCE CONFIGURATION

The FCOS baseline is the COCO_V1 checkpoint at score threshold 0.2, top-1000 candidates per level, NMS *IoU* 0.6 and 100 detections per image. The separated models use score threshold 0.2 and NMS *IoU* 0.3, each family evaluated at the suppression strength its geometry requires: dense levels at strides 8 and 16 emit spatially adjacent candidates.

3) LIMITATIONS AND STATISTICAL REPORTING

Cross-model comparisons place a fully converged fused model against short-schedule separated models. The evaluated population is controlled by floor-matching; the training budget is not. The AIC-FC record bounds how this should be read: continued optimization lowered the loss while *mAP* declined, so the short schedules are not truncated approaches to a better operating point. No claim rests on *mAP* magnitude, as the rank correlations and entropies are scale-free and the commitment result is per-image consistent across 5000 images. Dispersion statistics carry 95% bootstrap confidence intervals over the same 5000 val2017 images. Spearman correlations in Table 2 carry 95% confidence intervals from an image-level bootstrap ($B = 1000$) over matched predictions at $IoU \geq 0.5$; images rather than predictions are resampled, since predictions within an image are correlated. Single-GPU training at batch size 1 precludes multi-seed replication.

4) ATTRIBUTION OF THE COMPARISONS

The comparisons are not equally controlled, and the claims are distributed accordingly. The detached versus non-detached contrast of H-2 is drawn within each configuration, where pyramid depth, head structure, initialization, box geometry, schedule and suppression are held fixed and only the gradient path differs; each arm is nevertheless a single run, which is why H-2 carries a declared 0.05 decision rule. The within-AIC contrast of H-3, AIC-FC against AIC-Conv, shares pyramid levels, backbone architecture, score threshold and suppression but differs in head structure, classification-head initialization, box geometry, backbone fine-tuning and schedule, so the attribution to head structure is supported but not isolated. Every comparison against FCOS, in H-1, H-3 and § V-B, varies pyramid depth, head architecture, classification-head initialization, box parameterization, level assignment, training budget and suppression threshold simultaneously; these comparisons establish that the separated design reaches the reported operating points, not which of these factors produces the difference. Isolating the separation effect requires an FCOS-style baseline restricted to P2–P3 under matched initialization, schedule, assignment, score threshold and suppression, replicated over independent seeds, which we leave to future work. Until then, run-to-run variance remains unmeasured: the bootstrap intervals quantify sampling variance over images, not over training runs.

TABLE 2. Localization diagnostics by variant.

Metric	FCOS (P3–P7)	AIC-FC (P2–P3)	AIC-Conv ND (P2–P3)
all @0.5	0.9497	0.9418	0.9426
all @0.75	0.8141	0.8002	0.7808
all mean best *IoU*	0.8385	0.8194	0.8203
S @0.5 / @0.75	0.7143/ 0.3804	0.6734 / 0.3424	0.716 / 0.3681
M @0.5 / @0.75	0.9800 / 0.8094	0.9708 / 0.7915	0.9595 / 0.749
L @0.5 / @0.75	0.9962 / 0.9382	0.9983 / 0.9336	0.9952 / 0.9162
Spearman joint–*IoU* [ci low, ci high]	0.4869 [0.4823, 0.4911]	0.2530 [0.2475, 0.2586]	0.4064 [0.4015, 0.4115]
Spearman cls–*IoU* [ci low, ci high]	0.5022 [0.4976, 0.5064]	0.2124 [0.2071, 0.2180]	0.3162 [0.3111, 0.3213]
Spearman quality–*IoU* [ci low, ci high]	0.0509 [0.0476, 0.0546]	0.3693 [0.3602, 0.3783]	0.4841 [0.4785, 0.4900]
P2 (Share / Recall@50)	-	0.6413 / 0.9357	0.6593 / 0.8848
P3 (Share / Recall@50)	0.4223 / 0.5303	0.3587 / 0.7181	0.3407 / 0.7958
P4 (Share / Recall@50)	0.2296 / 0.5281	-	-
P5 (Share / Recall@50)	0.1842 / 0.4520	-	-
P6 (Share / Recall@50)	0.1224 / 0.2973	-	-
P7 (Share / Recall@50)	0.0415 / 0.1546	-	-

**At Optimal training cycle*

B. WHERE LOCALIZATION LIVES: A PER-LEVEL ANALYSIS

Aggregate *mAP* cannot show where in the pyramid localization is achieved, nor how the load is distributed across levels. Oracle recall decomposed by level can. Table 2 reports each level's share of best matches alongside its standalone oracle recall *IoU*@50.

The clearest signal concerns the quality field. FCOS's centerness sits at the floor as an *IoU* predictor, with a rank correlation of 0.051. Both separated variants raise it by an order of magnitude, to 0.484 for AIC-Conv and 0.369 for AIC-FC. The gap holds across two different head structures, so it tracks where the quality estimate is read rather than how it is read, thus supporting H-1. Part of this gap is expected by construction: centerness is supervised toward normalized center distance rather than overlap, whereas both separated quality heads are trained toward the IoU of the decoded box, so the magnitude of the rise is not attributable to read-out location alone. A matched baseline with an IoU-supervised quality branch would separate the two.

The per-level distribution shows why. In FCOS, the load is already concentrated at the shallow end: P3 and P4 carry 0.652 of best matches between them, and contribution falls monotonically thereafter, to 0.184, 0.122 and 0.042 at P5, P6 and P7. Standalone recall declines alongside it, from 0.530 at P3 to 0.155 at P7. The three deepest levels are retained by the architecture but contribute little of the localization that the pyramid achieves.

The separated design acts on that concentration rather than against it. AIC reads at P2 and P3 and takes two levels where FCOS takes five, with standalone recall of 0.885 and 0.796 for AIC-Conv against a maximum of 0.530 at any FCOS level. The two-level aggregate reaches 0.943 at $IoU@50$ against 0.950 for the five-level pyramid and nearly matches it on large objects, 0.9952 against 0.9962. The load carried by five levels is recovered from two.

One structural caveat bounds the per-level comparison. FCOS restricts regression targets to a size range at each level, so any single level's standalone recall is depressed for objects outside its assigned range, while the AIC levels each cover the full-size distribution. The per-level figures therefore favor the two-level design by construction, and the claim rests on the aggregate, which is unaffected.

Taken together, the two results describe the same effect from opposite sides. Reading the quality field where positional capacity survives raises its correspondence with attained overlap; concentrating the box pathway at those levels recovers the localizability of a deeper pyramid from a fraction of it. What the separation supplies is not additional information but a better organization of the information already present, with each target drawn from the regime that supports it.

C. QUALITY IN THE CLASS HEAD: THE DETACH PROBE

With localizability established, we ask where the objectness factor belongs. Table 3 compares the detached and non-detached arms within each architecture. The objective is to raise cls→IoU, how well the class score itself ranks localization, without sacrificing quality→IoU. H-2 predicts the gate allows the classifier to convert the quality signal into stronger cls→IoU, but only where the class head carries a usable objectness representation, which the center+size geometry preserves and the edge-distance geometry discards. The comparison tests how compatible each geometry, center+size versus edge-distance, is with the gate by tracking the detached and non-detached effect on cls→IoU and quality→IoU. The quantity of interest is the monotonicity of each correlation across arms rather than its absolute level. Differences below 0.05 are not read as informative; the threshold is the decision rule declared in H-2, not a measured seed spread, since run-to-run variance could not be estimated (§ V-A).

AIC-Conv behaves as predicted. Non-detachment raises cls→IoU in every size bin, by +0.083 on large objects and +0.092 in aggregate, with the small and medium gains of +0.044 and +0.031 falling below the threshold but retaining the same monotonicity. Quality→IoU does not pay for the gain: the non-detached arm is higher there as well, 0.426 against 0.347 in aggregate. With the centroid available to the class head, opening the gate lets the classifier use the quality signal, and the two correlations rise together rather than trading against one another.

The other three configurations do not reproduce this. In AIC-FC, detachment improves quality→IoU substantially, from 0.349 to 0.482 in aggregate, while cls→IoU remains flat under either arm at 0.364 and 0.361. AIC-FC α shows neither correlation responding to the gate: cls→IoU holds at 0.204 and 0.169 and quality→IoU at 0.464 and 0.462 across arms, suggesting this combination lacks compatibility, since neither correlation moves monotonically with the gate. In AIC-Conv α the same pattern as AIC-FC appears in weaker form, with quality→IoU moving from 0.362 to 0.419 and cls→IoU flat at 0.091 and 0.099. Across these configurations (AIC-FC, AIC-FC α and AIC-Conv α) the classification score does not respond to the gate; where quality→IoU does improve under detachment, the gain stays confined to the quality pathway. In AIC-FC the per-box read-out cannot convert the cue under either geometry, so under coupling the classification loss reshapes the quality field rather than reading from it, which is why detaching recovers quality→IoU most strongly there. An open gate and an improving quality field are together insufficient: without an objectness representation the class head can read, the signal does not reach the classification score.

TABLE 3. Spearman cls→*IoU* and quality→*IoU* by model, Size, and arm: AIC-FC (original→ edge-distance, α →center+size) vs. AIC-Conv (original→center+size, α →edge-distance), Non-Detach vs. Detach.

Metric	Size	AIC-FC		AIC-FC α		AIC-Conv		AIC-Conv α	
		ND	DT	ND	DT	ND	DT	ND	DT
cls→ *IoU*	S	0.478	0.433	0.312	0.275	0.16	0.116	0.130	0.129
	M	0.483	0.476	0.309	0.283	0.243	0.212	0.199	0.202
	L	0.359	0.358	0.169	0.132	**0.307**	0.224	0.079	0.091
	All	0.364	0.361	0.204	0.169	**0.334**	0.242	0.091	0.099
quality → *IoU*	S	0.530	0.530	0.456	0.411	0.365	0.357	0.433	0.466
	M	0.411	**0.524**	0.599	0.591	**0.592**	0.524	0.580	0.585
	L	0.358	**0.485**	0.436	0.446	**0.412**	0.332	0.357	**0.417**
	All	0.349	**0.482**	0.464	0.462	**0.426**	0.347	0.362	**0.419**

All variants trained for 4 epochs

The FC comparison carries an additional confound. Its classification head is the pretrained FCOS head, already entangled with localization through joint centerness training, whereas AIC-Conv trains its classification head from scratch. The absolute cls→IoU gap between the families therefore reflects initialization as well as head structure. The within-family detached-versus-non-detached contrast is unaffected, since initialization is held constant across the two arms, and it is that contrast the prediction concerns.

H-2 is supported in the conditional form in which it was stated: quality can lift cls→IoU, but only where the class head can read the spatial neighborhood the objectness cue occupies. Geometry alone is not sufficient: the center+size probe preserves the cue, yet the fully-connected per-box

head, scoring each candidate by a 1×1 read-out without spatial context, leaves cls→IoU flat under both geometries. The lift appears only where the convolutional head and the center+size geometry coincide. We therefore carry forward AIC-Conv non-detached, the only configuration that raises cls→IoU while leaving quality→IoU intact.

D. CANDIDATE-CLASS FIELD: SPRAY VERSUS COMMITMENT

We apply the candidate-class instrument to all three models. Fig. 2 and Table 4 characterize the surviving candidate field of a single evaluation pass; Table 5 reports both stages, candidate and survivor, over 5000 val2017 images with bootstrap intervals.

The two spatially-blind heads behave alike at the candidate stage. FCOS and AIC-FC hold 50.9 and 41.1 classes in play at entropies of 3.66 and 3.17 bits, while AIC-Conv holds 15.1 at 2.22 bits. The rank-frequency curves separate across their whole range rather than in the tail alone (Fig. 2a), and the concentration measure gives the same ordering: 5 classes carry 90% of AIC-Conv's candidate mass against 12 and 14 for the others.

That AIC-FC sprays as widely as FCOS while reading from separated levels locates the effect in the read-out rather than in the separation. A per-box 1×1 read-out scores each candidate without reference to its spatial neighborhood, so nothing in the scoring path suppresses categories the surrounding field does not support. This is why the FC head was not carried forward, and why its center+size variant appears only in the Table 3 diagnostic rather than as a full realization: the deficit lies in how the field is read in the class head, and changing the box geometry does not address it. The interaction between head structure and box geometry is therefore estimated only from the Table 3 diagnostic and read as indicative rather than conclusive; we state this as a limitation of the factorial design.

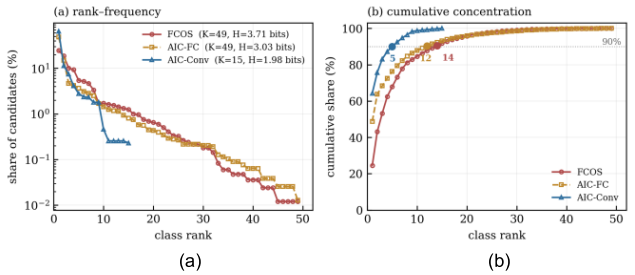


FIGURE 2. Candidate-class field of the three models. (a) Rank-frequency of surviving candidates over categories; (b) cumulative concentration, with the class count carrying 90% of candidate mass marked. The spatially-blind heads (FCOS, AIC-FC) spray across all K=49 categories; the convolutional head (AIC-Conv) commits to K=15.

The survivor stage sharpens the argument rather than qualifying it. AIC-Conv retains the lowest survivor entropy at 1.31 bits, but FCOS at 1.75 falls below AIC-FC at 2.13, reversing the candidate-stage order. The entropy-retained figures account for it: FCOS carries 48% of its candidate entropy through to the survivor set, against 59% for AIC-Conv and 67% for AIC-FC. FCOS arrives at a committed output by suppressing a hedged field, while AIC-Conv arrives there with a field that was already concentrated. Commitment achieved at NMS is commitment delegated to a threshold rather than earned in the scoring path, and it is the head, not the suppression stage, that the separation is designed to act on.

The leak-out measure is unambiguous, and its composition confirms the mechanism. Rare survivors account for 0.9% of AIC-Conv's survivors against 5.9% for AIC-FC and 3.2% for FCOS, and appear in roughly 29% of bootstrap resamples against 91% and 78%. Those that remain carry the highest mean quality of the three, at 0.827 against 0.774 and 0.639, with class confidence level across all models at approximately 0.09. A leaked survivor would be poorly localized, having entered on a spurious score rather than on evidence; a well-localized one is a genuine detection of a class the image supports sparsely. The reduction is therefore in leakage specifically, not in rare detections as a category.

TABLE 4. Candidate-Class field statistics by model (Single Evaluation Pass; Statistics Over 5000 Images in Table 5).

	FCOS	AIC-FC	AIC-Conv
Candidates	8,325	7,800	7,800
Classes	49	49	15
H (bits)	3.71	3.03	1.98
Top-1	24.5%	48.8%	64.3%
Top-5	67.9%	76.2%	90.0%
Rank at 90%	14	12	5

TABLE 5. Categorical dispersion over 5000 val2017 images (Bracketed Ranges are 95% Bootstrap CIs).

Metric	FCOS	AIC-FC	AIC-Conv
Candidate entropy, H_{cand}	3.66	3.17	2.22
Survivor entropy, H_{surv}	1.75	2.13	1.31
Candidate class count	50.9	41.1	15.1
Survivor class count	8.0	12.0	4.3
Entropy retained	48%	67%	59%
Rare survivor fraction, f	3.2%	5.9%	0.9%
Rare survivor, c_f	0.094	0.091	0.097
Rare survivor quality	0.639	0.774	0.827
Bootstraps w/ rare survivors	~78%	~91%	~29%

The decomposition corroborates this at the detection level (Table 6). Within the AIC pair, FalsePos falls from 22.23 for AIC-FC to 14.17 for AIC-Conv, tracking the rare-survivor result: the head that commits also admits fewer spurious detections. The comparison is drawn within the pair, since ΔAP terms are relative to each model's own baseline AP. Each model is evaluated at its native suppression setting, 0.6 for FCOS and 0.3 for the AIC variants, because the claim concerns what the scoring path delivers to suppression rather than what suppression recovers. Equalizing the threshold would test how each field responds to a common operating point; holding each at its own tests whether commitment was achieved before that point, which is what the separation is designed to do. One cost is visible in the same table: missed-object error for AIC-Conv, at 12.19, is roughly double that of

FCOS (6.04) and AIC-FC (5.91). AIC-FC shares the 0.2 score threshold and 0.3 suppression without showing the increase, so suppression strength alone does not account for it. One reading consistent with the candidate-field results is that a committed field holds fewer alternative candidates per object, leaving no fallback when the committed score falls below threshold. We have not isolated this mechanism and report the increase as an adverse finding.

TABLE 6. TIDE error decomposition at AP50 (dAP per Category; Score threshold 0.2 for all; Native NMS: FCOS 0.6, AIC 0.3).

Model	AP_{50}	Cls	Loc	Both	Dupe	Bkg	Miss	False Pos
FCOS (thr 0.2, NMS 0.6)	58.25	3.13	6.97	1.22	0.44	4.13	6.04	20.92
AIC-Conv (thr 0.2, NMS 0.3)	43.87	5.99	10.54	0.85	0.00	4.05	12.19	14.17
AIC-FC (thr 0.2, NMS 0.3)	46.65	4.74	11.02	1.22	0.00	4.66	5.91	22.23

H-3 is supported on both measures: candidate entropy falls from 3.66 bits to 2.22, and the rare-survivor fraction falls from 3.2% to 0.9%. Within the AIC pair, where more factors are matched, the same ordering holds: candidate entropy falls from 3.17 to 2.22 bits and the rare-survivor fraction from 5.9% to 0.9%. The reduction in readout uncertainty appears where the design predicted, in the head that retains spatial support for its scoring, and is achieved in the field itself rather than recovered at suppression.

VI. ROBUSTNESS UNDER CORRUPTION

Table 7 reports clean *mAP*, mean performance under corruption, and retention across a four-family subset of [14] on shared corrupted inputs. Retention is 0.593 for FCOS against 0.554 and 0.559 for the two-level variants, a relative robustness of 0.934 and 0.942. The separation neither improves nor materially degrades corruption retention.

This is reported as a null rather than a result. None of H-1 to H-3 concerns robustness, and commitment in the class field does not by itself imply resilience to input perturbation: a concentrated field commits more confidently, which under corruption can mean committing confidently to the wrong category. Retention holding within 7 per cent of the baseline while the head, pyramid depth and box geometry all change indicates that corruption behavior is governed by the backbone the three models share, and the question of whether commitment affects robustness requires an evaluation designed for it.

TABLE 7. Corruption robustness (Four-Family Subset of [14]; shared corrupted inputs).

Model	clean mAP	mPC	retention (rPC)	rel. robust.
FCOS (5-level)	0.391	0.232	0.593	1.000
AIC-Conv (2-level)	0.259	0.143	0.554	0.934
AIC-FC (2-level)	0.250	0.139	0.559	0.942

VII. CONCLUSION

We began from an observation about the convolutional process rather than about the detection head. As depth increases, successive operations enrich the channel dimension while contracting the spatial dimensions, so a prediction drawn from a single undivided read-out takes both contributions in whatever proportion its depth supplies. Separating the classification and box heads removes that constraint, and the three hypotheses of § III ask what follows when each target is read from the regime that supports it.

All three are supported. The quality field carries a usable localization signal where the coupled baseline's does not; two levels recover the localizability of five; and the class field commits in the head rather than at suppression, with uncorroborated survivors reduced accordingly. H-2 holds in the conditional form in which it was stated, since the gate transmits nothing the class head has no representation to receive.

Each result was predicted before measurement and each carried a stated failure condition, so the diagnosis is testable independently of aggregate performance. Two findings run against expectation and are reported as such: localization error rises rather than falls under separation, and corruption retention is unchanged. Both are consistent with the position that the separation acts on the scoring pathway, not on what the box regressor attains or on what the shared backbone represents. A third, the near doubling of missed-object error for AIC-Conv, lies on the scoring pathway itself and is not explained by the present instruments.

The concept remains at an early stage. Its clearest limitation is the fusion of quality into the classification score, realized here through a single gated concatenation. The results establish that the quality signal carries usable information; they do not establish that this is the best way to admit it, and the present arrangement is unlikely to be optimal. Whether a different fusion yields a more stable score ranking, and whether that stability relates to the commitment properties measured here, is the direction we pursue next, on the instruments introduced in § IV rather than on aggregate score.

REFERENCES

[1] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 21–37.

[2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[3] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2999–3007.

[4] Z. Tian, C. Shen, H. Chen, and T. He, "FCOS: Fully convolutional one-stage object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9626–9635.

[5] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.

[6] C. Feng, Y. Zhong, Y. Gao, M. R. Scott, and W. Huang, "TOOD: Task-aligned one-stage object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 3490–3499.

[7] E. Collins, R. Achanta, and S. Süsstrunk, "Deep feature factorization for concept discovery," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 336–352.

[8] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.
[9] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 658–666.
[10] Z. Ge, S. Liu, Z. Li, O. Yoshie, and J. Sun, "OTA: Optimal transport assignment for object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 303–312.
[11] B. Jiang, R. Luo, J. Mao, T. Xiao, and Y. Jiang, "Acquisition of localization confidence for accurate object detection," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 784–799.
[12] S. Wu, X. Li, and X. Wang, "IoU-aware single-stage object detector for accurate localization," *Image Vis. Comput.*, vol. 97, May 2020, Art. no. 103911.
[13] X. Li, W. Wang, X. Hu, J. Li, J. Tang, and J. Yang, "Generalized focal loss v2: Learning reliable localization quality estimation for dense object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 11627–11636.
[14] C. Michaelis, B. Mitzkus, R. Geirhos, E. Rusak, O. Bringmann, A. S. Ecker, M. Bethge, and W. Brendel, "Benchmarking robustness in object detection: Autonomous driving when winter is coming," 2019, *arXiv:1907.07484*.
[15] D. Bolya, S. Foley, J. Hays, and J. Hoffman, "TIDE: A general toolbox for identifying object detection errors," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 558–573.
[16] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proc. 11th Annu. Conf. Comput. Learn. Theory*, Jul. 1998, pp. 92–100.
[17] Z.-H. Zhou and M. Li, "Tri-training: Exploiting unlabeled data using three classifiers," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 11, pp. 1529–1541, Nov. 2005.
[18] Y. Zhang, T. Xiang, T. M. Hospedales, and H. Lu, "Deep mutual learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4320–4328.
[19] P. Ren, Z. Han, Z. Yu, and B. Zhang, "Confucius tri-learning: A paradigm of learning from both good examples and bad examples," *Pattern Recognit.*, vol. 163, Apr. 2025, Art. no. 111481.
[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[21] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 740–755.



HADI MAT ROSLY received the bachelor's and master's degrees in mechatronics engineering from International Islamic University Malaysia (IIUM), in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree in machine learning on a part-time basis. He is a Senior Software Developer with nine years of experience delivering production systems across the automotive, industrial, and enterprise domains, from PLC-integrated edge devices on factory floors to a Kubernetes-native platform now running across nine plants on four continents. He has led cross-functional teams of up to nine engineers and architects systems under Clean Architecture and SOLID principles. His research interests include anchor-free object detection, mechanistic diagnosis of detectors, and robustness to out-of-distribution inputs.



ZULKIFLI ZAINAL ABIDIN received the B.Eng. degree in computer and information engineering from the International Islamic University Malaysia (IIUM) in 2003, and the M.Sc. and Ph.D. degrees in electrical and electronics engineering (robotics) from Universiti Sains Malaysia (USM) in 2007 and 2013, respectively. He is currently an Associate Professor with the Department of Integrated Engineering, Universiti Teknologi PETRONAS (UTP). Prior to joining UTP, he was a faculty member with the Department of Mechatronics Engineering, IIUM, where he Co-Founded and served as the inaugural Director of the Centre for Unmanned Technologies (CUTe). He has served as a technical consultant for numerous public and multinational entities, securing over RM5 million in industry-driven research funding. He is a Chartered Engineer (C.Eng) with the Engineering Council UK, a Professional Engineer (Ir.) registered with the Board of Engineers Malaysia (BEM), a Professional Technologist (Ts.) with the Malaysia Board of Technologists (MBOT), and a Certified Geospatialist (Gs.). He also serves on the executive committees of SAE International Malaysia and the IEEE Oceanic Engineering Society. His research interests include multi-domain unmanned systems (marine, ground, and aerial), advanced embedded sensing, telematic control units, and intelligent control algorithms.



HASAN FIRDAUS MOHD ZAKI (Member, IEEE) received the Ph.D. degree in computer science from The University of Western Australia. He is currently an Associate Professor with the Department of Mechatronics Engineering, International Islamic University Malaysia (IIUM), and the Director of the Centre for Unmanned Technologies (CUTe). He is also a Certified Professional Technologist. His research interests include robotic vision, RGB-D object and scene recognition, machine learning, 3-D face analysis, action recognition, and embedded artificial intelligence, with applications in autonomous robotics, intelligent surveillance, and UAV-based sensing. His work focuses on perception, human–robot interaction, and real-time decision-making for intelligent automation and autonomous systems. He also collaborates with industry and government agencies as a consultant and subject matter expert, translating research advances into practical engineering solutions.