Investigating the Impact of Input Resolution on Object Detection Performance Using EfficientDet for Signature Detection in Document Images

1st Ensophea Toch
School of Graduate Studies
Paragon International University
Phnom Penh, Cambodia
etoch@paragoniu.edu.kh

2nd Pakrigna Long
Department of Applied Mathematics and Statistics,
Institute of Technology of Cambodia
Phnom Penh, Cambodia
long.pakrigna@itc.edu.kh

***Abstract*—Automated signature detection is a critical component of intelligent document processing systems, supporting document authentication and fraud detection. This paper presents a controlled empirical investigation of the effect of input resolution on handwritten signature detection using EfficientDet-D0. Three resolutions—512×512, 640×640, and 768×768—were evaluated under identical training conditions on a self-collected corpus of 1,000 annotated document images. Detection performance, measured by mAP@0.5, follows a non-monotonic trend: it improves from 0.9000 at 512×512 to a peak of 0.9261 at 640×640, then declines slightly to 0.9233 at 768×768, even though the 768×768 configuration achieves the lowest validation loss. We attribute this pattern to interacting effects of BiFPN feature granularity, anchor-scale alignment, and optimization complexity, and we discuss the scope and limitations of a single-architecture, single-run study. The results indicate that input resolution is a domain-specific hyperparameter that should be empirically tuned rather than inherited from benchmark conventions.**

***Index Terms*—signature detection, object detection, EfficientDet, BiFPN, input resolution, document image analysis**

I. Introduction

Automated document workflows increasingly require robust signature localization across diverse layouts, scan qualities, and handwriting styles. Signature detection differs from conventional object detection in three respects: (1) signatures are extreme small objects, typically spanning less than 1% of a scanned A4 page; (2) they exhibit high intra-class variation, reflecting idiosyncratic motor habits rather than rigid shape templates; and (3) they lack a canonical spatial position within general (non-templated) documents.

EfficientDet [3], built on a Bidirectional Feature Pyramid Network (BiFPN), is a widely used efficient anchor-based detector. Its compound-scaling resolution schedule is calibrated against COCO's object-size distribution [4], which differs substantially from the signature-dominated size distribution found in document images. Whether this mismatch helps or hurts performance in the signature domain is not obvious a priori, and–to our knowledge–has not been systematically studied.

A. Contributions

(1) A controlled single-factor study isolating input resolution (512, 640, 768) for EfficientDet-D0 signature detection; (2) empirical evidence of non-monotonic accuracy vs. resolution, challenging the assumption that higher resolution is always better; (3) a mechanistic account of the observed pattern in terms of BiFPN granularity, anchor-scale alignment, and optimization complexity; and (4) practical guidance for resolution selection in document AI pipelines, together with an explicit discussion of the scope and limitations of the study.

II. Related Work

A. Signature Detection

Early signature localization relied on document-specific structural priors, such as template matching against form coordinates, and on handcrafted descriptors (e.g., HOG plus SVM) that generalized poorly to unstructured layouts [1]. Deep learning shifted the landscape: two-stage pipelines combining Faster R-CNN proposals with signature-specific classifiers improved recall on multi-page corpora [6]. Single-stage detectors such as SSD [8] and, more recently, YOLO-family models have been applied to signature and stamp detection with favorable speed–accuracy trade-offs; Yan et al. introduced ChiSig, a benchmark spanning detection, restoration, and verification for Chinese document signatures, and compared several deep detectors on the detection sub-task [10]. Zhang et al. further explored Copy–Paste augmentation for signature detection on technical documents with limited labeled data [11]. Transformer-based detectors (DETR and successors) remove the need for anchor design [7], [9] but remain harder to train stably on small datasets. Beyond static, image-based detection, recent work has also applied transformer architectures to the related task of online (dynamic) signature verification, modeling pen-stroke trajectories with a temporal-spatial graph attention transformer [5]; this line of work targets stroke-level dynamic verification rather than document-level localization, the focus of the present study. Two recent surveys further situate the present study: Wang et al. review advances in offline

handwritten signature recognition, highlighting resolution and small-object handling as open challenges across deep detection pipelines [16], and Pandey et al. provide an extensive survey of offline signature verification pipelines, underscoring the continued reliance on document-specific preprocessing choices such as input scale [17]. Recent studies emphasize that robust signature detection requires balancing localization accuracy, computational efficiency, and adaptability across diverse document formats, particularly when real-world samples contain noise, distortion, and complex visual backgrounds.

B. Small-Object Detection and Input Resolution

The COCO benchmark defines small objects by bounding-box area [4], and small objects are consistently the dominant source of mAP degradation [2]. Feature Pyramid Networks [2] and BiFPN [3] mitigate this via multi-scale fusion, but do not eliminate the underlying resolution sensitivity. He et al. and the EfficientNet compound-scaling law established that resolution jointly determines feature quality and compute cost [3], [13]. For detection specifically, resolution simultaneously affects feature granularity at each pyramid level and the size of the anchor proposal set; studies of aerial and pedestrian detection have reported diminishing or even reversed accuracy returns at very high resolutions [12], [14], consistent with the pattern we observe here in the document domain. Anchor-free and query-based detectors (e.g., Deformable DETR [9]) sidestep anchor-scale misalignment entirely and represent a natural point of comparison for future work.

III. Methodology

A. Dataset

The dataset comprises 1,000 document images containing handwritten signatures, annotated with Pascal VOC bounding boxes. **The corpus was collected and annotated internally by the authors** (not derived from an existing public benchmark), covering administrative forms, record sheets, official correspondence, and business documents, with scan resolutions from 100–300 DPI. Each image contains one annotated signature; annotation was performed by trained annotators with consensus verification for ambiguous cases. A single-class formulation was adopted to isolate localization performance from multi-class confounds. The dataset is partitioned 80/20 (800 train / 200 validation) via stratified sampling by document type; no augmentation is applied to validation. Training augmentation includes horizontal flipping, brightness/contrast jitter, and constrained random cropping. In the training split, signature widths range 87–412 px (mean 216, std 74) and heights 31–148 px (mean 79, std 29); roughly 68% qualify as "small objects" under the COCO convention at 512 resolution.

B. Model and Choice of Architecture

EfficientDet-D0 (EfficientNet-B0 backbone, 3 BiFPN layers at 64 channels, 3-layer shared prediction head, ≈3.9M parameters) was selected as the sole architecture because it represents lightweight, anchor-based single-stage detectors used in on-premises document-processing pipelines, where inference cost and memory footprint are critical. This choice limits conclusions to anchor-based, compound-scaled detectors; we do not claim the observed resolution sensitivity generalizes to other families (e.g., YOLO, DETR), and discuss this in Section VI. Anchor scale is $s = 4.0$ with ratios $\{1.0, 2.0, 0.5\}$ across pyramid levels P3–P7; at 640×640 the P3 anchor footprint (≈64×32 px) matches the median signature extent, motivating this configuration. The composite loss combines focal loss (classification, $\alpha = 0.25, \gamma = 1.5$) and Smooth L1 (box regression), $\mathcal{L} = \lambda_{cls}\mathcal{L}_{focal} + \lambda_{reg}\mathcal{L}_{SmoothL1}$, $\lambda_{cls} = \lambda_{reg} = 1.0$.

C. Training Protocol

All conditions share an identical protocol (Table I) except for input resolution, isolating resolution as the sole independent variable. AdamW [15] was used for its decoupled weight decay; cosine annealing avoided confounding abrupt LR drops with optimization artifacts; mixed-precision (FP16/FP32) training reduced memory use. The models were trained from identical initialization settings with the same augmentation strategy, batch size, optimizer parameters, and evaluation procedure to ensure reproducible comparisons. Checkpoints were selected based on validation performance, and the final model was evaluated on the held-out validation set using consistent detection metrics. This controlled design enables a direct assessment of how image resolution influences signature localization accuracy and computational efficiency. Fig. **??** and Fig. 1 show the shared cosine-annealed learning-rate schedule and the per-step training batch loss for all three resolutions.

TABLE I
TRAINING HYPERPARAMETERS (ALL EXPERIMENTS)

Hyperparameter	Value
Optimizer	AdamW [15]
Initial learning rate	1×10^{-3}
Weight decay	4×10^{-5}
LR schedule	Cosine annealing (to 0)
Batch size	16
Training epochs	30
Mixed precision	FP16/FP32
Warmup steps	500
Gradient clipping	max norm = 1.0

IV. Experimental Setup

Three conditions were evaluated: **E1 (512×512)**, EfficientDet-D0's canonical design-point resolution [3]; **E2 (640×640)**, the common YOLO-family deployment resolution; and **E3 (768×768)**, EfficientDet-D2's design-point resolution applied under D0's fixed capacity. All preprocessing, augmentation, optimizer state, dataset split, and evaluation protocol were held constant across conditions. Following the PASCAL VOC convention, primary evaluation uses mAP@0.5 over the full 200-image validation set. We selected this single-IoU-threshold metric for direct comparability with prior signature/document detection

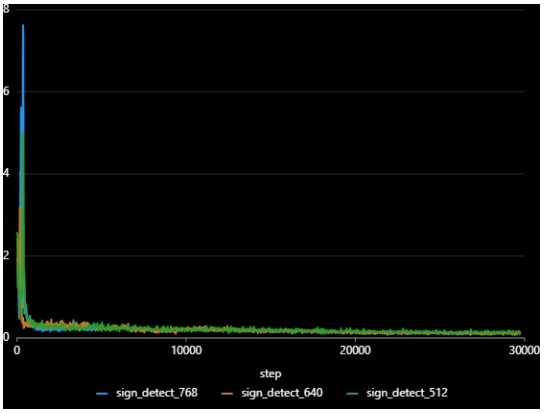


Fig. 1. Training batch loss for the three input resolution conditions (512×512, 640×640, and 768×768). The 768×768 model exhibits a higher initial loss and slower early convergence than the 512×512 and 640×640 models.

literature and because, in a single-class formulation, it isolates the localization-accuracy trend of interest without the additional variance introduced by averaging over ten COCO-style IoU thresholds on a small validation set; we return to this choice as a limitation in Section VI. In addition to accuracy evaluation, we recorded inference latency and computational behavior under each resolution setting to quantify the practical trade-off between localization improvement and deployment cost. Since higher resolutions increase pixel-level information but also require greater memory consumption and computation, this experimental design enables analysis of whether accuracy gains justify additional resource requirements. The comparison therefore provides insight into selecting an appropriate operating resolution for real-world document authentication systems with constrained hardware environments.

V. Results and Discussion

A. Quantitative Results

Table II summarizes mAP@0.5 across conditions. Accuracy improves substantially from 512 to 640 (ΔmAP $= +0.0261$, $+2.90\%$ relative), then regresses slightly from 640 to 768 (ΔmAP $= -0.0028$). The direction of this small but reproducible regression, under a fully controlled design, is the central empirical finding of this study.

TABLE II
Detection performance under varying input resolution

Resolution	mAP@0.5	Δ vs. 512	Rank
512×512	0.9000	–	3
640×640	**0.9261**	+2.90%	**1**
768×768	0.9233	+2.59%	2

B. Training Dynamics

At 512, validation loss begins to drift upward near step 10,000 with mild mAP oscillation, suggesting incipient overfitting at the model's design-point resolution. At 640, training and validation loss converge more stably, with no upward drift, and validation mAP reaches its peak (0.9261). At 768, initial training loss peaks near 6.0—roughly three times the 640 peak—reflecting a substantially more complex optimization landscape; despite reaching the lowest terminal validation loss (≈0.42) of the three conditions, final mAP (0.9233) does not surpass the 640 peak, indicating a loss–mAP discrepancy discussed below. Fig. 2 shows the validation mAP and validation loss curves underlying this comparison.

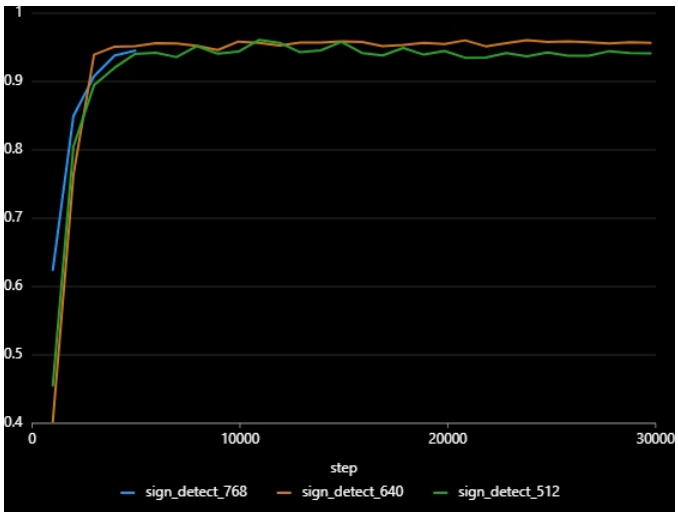


Fig. 2. Validation mAP@0.5 and validation loss curves across training steps for the three resolution conditions (512×512, 640×640, 768×768). The 640×640 curve reaches the highest and most stable validation mAP, while 768×768 attains the lowest validation loss without a corresponding mAP gain.

C. Feature Granularity and Non-Monotonic Behavior

At pyramid level P_l, spatial dimensions scale as $R/2^l \times R/2^l$. For a median-sized signature (≈200×60 px native), P_3 coverage increases from ≈7.7×2.4 feature vectors at 512 to ≈9.6×2.9 at 640 (a 56% increase), tracking the mAP gain from 0.9000 to 0.9261. At 768, P_3 coverage increases further (≈11.6×3.5) yet mAP declines slightly. We attribute this to four interacting, partially offsetting mechanisms:

(1) Feature dilution. EfficientDet-D0's fixed 64-channel, 3-layer BiFPN must encode a $(768/512)^2 = 2.25\times$ larger spatial volume with unchanged channel capacity, diluting per-location discriminative signal. *(2) Optimization complexity.* Total anchor proposals scale from ≈49,000 (640) to ≈70,500 (768), a 44% increase, straining convergence given only 800 training images and 30 epochs. *(3) Interpolation artifacts.* Native scan resolutions are typically below 768; upscaling amplifies high-frequency interpolation noise at ink-stroke boundaries, whereas downscaling to 512/640 acts as a low-pass filter. *(4) Anchor-scale misalignment.* The anchor scale $s = 4.0$ was tuned against the signature-size distribution at 640; at 768, signatures grow in rescaled pixel terms while anchor footprints remain fixed, lowering IoU with ground truth and degrading positive-sample quality.

D. Loss–mAP Discrepancy

The 768 condition achieves the lowest terminal validation loss but not the highest mAP. Since focal and Smooth L1 loss are correlated with, but not equivalent to, mAP at a fixed IoU threshold, a model can produce confident yet slightly mislocalized boxes that reduce combined loss without improving IoU-thresholded precision–recall. This is consistent with the anchor-misalignment hypothesis: noisier regression targets increase localization variance, which mAP@0.5 penalizes more directly than Smooth L1 loss does. This observation highlights that optimization objectives should be interpreted alongside task-specific evaluation metrics, especially for small object detection where minor localization errors can substantially affect final performance.

E. Computational Trade-off

Table III reports relative FLOPs (BiFPN cost scales $\approx O(R^2)$). The 768 configuration costs 2.25× the 512 baseline for a smaller accuracy gain (+0.0233) than 640 achieves at only 1.56× cost (+0.0261); 640×640 Pareto-dominates 768×768 on this task. This result suggests that moderate resolution scaling provides a more efficient balance between detection accuracy and computational requirements, particularly for resource-constrained document processing deployments.

TABLE III
RELATIVE COMPUTATIONAL COST VS. ACCURACY

Resolution	mAP@0.5	Rel. FLOPs	Anchors/img
512×512	0.9000	1.00×	≈32,000
640×640	0.9261	1.56×	≈49,000
768×768	0.9233	2.25×	≈70,500

This non-monotonic pattern is broadly consistent with prior findings in aerial and pedestrian detection, where very high input resolutions introduced diminishing or reversed returns due to background complexity and anchor over-generation [12], [14], suggesting the optimal resolution for anchor-based detectors is jointly determined by target object scale, backbone/fusion capacity, and source-image characteristics.

VI. Limitations and Threats to Validity

We report several limitations directly, rather than as an afterthought, since they bound how the results should be interpreted.

Single architecture. All experiments use EfficientDet-D0; we have not verified whether the non-monotonic pattern holds for other detector families (YOLO, DETR) or larger EfficientDet variants (D1–D7). The mechanistic account above (Section VD) generates architecture-specific predictions—e.g., that a higher-capacity BiFPN (as in D2) should mitigate the feature-dilution effect at 768—that future comparative studies could test directly.

Single run, no variance estimate. Each resolution condition was trained once. The 640 vs. 768 gap (0.9261 vs. 0.9233) is small enough that we cannot rule out run-to-run noise as a partial explanation; multi-seed replication with reported mean/standard deviation is needed before treating the ranking as statistically established, and we present it here as a reproducible single-run observation rather than a statistically validated effect.

Metric choice. We use mAP@0.5 rather than COCO-style mAP averaged over IoU 0.50:0.95. This choice was made for comparability with prior VOC-style signature/document detection work and because a single-class, single-threshold metric is less sensitive to the added variance of averaging across ten thresholds on a 200-image validation set; however, it is a coarser localization criterion, and we cannot confirm whether the observed ranking is preserved under stricter IoU thresholds without re-evaluation.

No cross-architecture or cross-baseline comparison. We do not compare against SSD, YOLO, or DETR-family detectors at matched resolutions, so we cannot say whether the resolution sensitivity we observe is specific to EfficientDet's anchor and BiFPN design or a more general property of signature detection.

No anchor-scale ablation. The anchor-misalignment hypothesis (Section VC) is proposed but not directly tested; an ablation varying anchor scale/aspect ratio at 768 would be needed to confirm it mechanistically rather than by inference from indirect evidence (IoU/loss patterns).

Single dataset. The corpus, while diverse in document type and scan quality, was collected internally at a single institution; generalization to other document ecosystems (e.g., different national forms, languages, or scanning hardware) is untested.

Dataset scale and annotation uncertainty. Although the dataset covers diverse document types, its size remains limited. Larger multi-source datasets and improved annotation analysis are needed to better evaluate robustness and generalization.

No per-image qualitative audit. The present figures report aggregate training and validation curves (Figs. **??**–2) rather than per-image bounding-box comparisons; a qualitative audit of detections on individual documents across resolutions is left for future work.

VII. Conclusion and Future Work

Across a controlled, single-factor study, EfficientDet-D0 signature detection accuracy improved from 512 to 640 resolution but declined slightly at 768, despite 768 achieving the lowest validation loss. The pattern is consistent with four interacting mechanisms—feature dilution, optimization complexity, interpolation artifacts, and anchor-scale misalignment—suggesting that input resolution is a domain-specific hyperparameter that should be tuned against the target object-size distribution rather than inherited from COCO-calibrated defaults or applied at the highest resolution available. For this task, 640×640 offers the best accuracy–efficiency trade-off among the settings tested.

Because the present study is scoped to a single architecture and a single run per condition, the most direct next steps are: (1) multi-seed replication with reported variance for the 640 and 768 conditions; (2) extending the resolution sweep to larger EfficientDet variants (D1–D7) and to

anchor-free/transformer-based detectors (YOLO, Deformable DETR [9]) to test whether the non-monotonic pattern is architecture-specific; (3) an anchor-scale/aspect-ratio ablation at 768 to directly test the anchor-misalignment hypothesis; (4) re-evaluation under COCO-style mAP (IoU 0.50:0.95) to test whether the ranking holds under stricter localization criteria; and (5) qualitative, per-image comparison of detections across resolutions to complement the aggregate mAP trend. More broadly, we hope this study encourages treating input resolution as a first-class, empirically validated design variable in document-domain object detection rather than a convention inherited from natural-image benchmarks.

Future research should further investigate adaptive resolution strategies, where the input size is dynamically selected based on document complexity, signature scale, and available computational resources. Such approaches could reduce unnecessary inference cost while maintaining detection reliability. In addition, integrating resolution-aware training strategies, multi-scale feature learning, and domain adaptation techniques may improve robustness when deploying signature detection systems across heterogeneous real-world document environments.

REFERENCES

[1] M. Blumenstein, X. Y. Liu, and B. Verma, "A modified direction feature for cursive character recognition," *Pattern Recognition*, vol. 40, no. 2, pp. 376–388, 2007. [Online]. Available: https://doi.org/10.1109/IJCNN.2004.1381140

[2] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," *CVPR*, 2017. [Online]. Available: https://arxiv.org/abs/1612.03144

[3] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," *CVPR*, 2020. [Online]. Available: https://arxiv.org/abs/1911.09070

[4] 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," *ECCV*, 2014. [Online]. Available: https://doi.org/10.1007/978-3-319-10602-1_48

[5] H. Yuan *et al.*, "Online handwritten signature verification based on temporal-spatial graph attention transformer," *arXiv preprint*, 2025. [Online]. Available: https://arxiv.org/abs/2510.19321

[6] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE TPAMI*, vol. 39, no. 6, pp. 1137–1149, 2017. [Online]. Available: https://arxiv.org/abs/1506.01497

[7] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," *ECCV*, 2020. [Online]. Available: https://doi.org/10.1007/978-3-030-58452-8_13

[8] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," *ECCV*, 2016. [Online]. Available: https://doi.org/10.1007/978-3-319-46448-0_2

[9] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable DETR: Deformable transformers for end-to-end object detection," *ICLR*, 2021. [Online]. Available: https://openreview.net/forum?id=gZ9hCDWe6ke

[10] K. Yan, Y. Zhang, H. Tang, C. Ren, J. Zhang, G. Wang, and H. Wang, "Signature detection, restoration, and verification: A novel Chinese document signature forgery detection benchmark," *CVPRW*, 2022, pp. 5163–5172. [Online]. Available: https://doi.org/10.1109/CVPRW56347.2022.00564

[11] Y. Zhang, J. Zhang, K. Yan, et al., "Handwritten Chinese signature detection with simple Copy–Paste augmentation on power plants technical documents," *Service Oriented Computing and Applications*, vol. 17, pp. 293–302, 2023. [Online]. Available: https://doi.org/10.1007/s11761-023-00374-8

[12] G.-S. Xia, X. Bai, J. Ding, Z. Zhu, S. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang, "DOTA: A large-scale dataset for object detection in aerial images," *CVPR*, 2018, pp. 3974–3983. [Online]. Available: https://doi.org/10.1109/CVPR.2018.00418

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CVPR*, 2016, pp. 770–778. [Online]. Available: https://doi.org/10.1109/CVPR.2016.90

[14] P. Zhu, L. Wen, D. Du, X. Bian, H. Ling, Q. Hu, and Q. Nie, "Detection and tracking meet drones challenge," *IEEE TPAMI*, vol. 44, no. 11, pp. 7380–7399, 2021. [Online]. Available: https://doi.org/10.1109/TPAMI.2021.3119563

[15] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *ICLR*, 2019. [Online]. Available: https://openreview.net/forum?id=Bkg6RiCqY7

[16] Z. Wang *et al.*, "Advances in offline handwritten signature recognition research: A review," *IEEE Access*, 2023. [Online]. Available: https://doi.org/10.1109/ACCESS.2023.3326471

[17] G. K. Pandey *et al.*, "Offline signature verification: An extensive survey," *ICSADL*, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/10933339