Automated Skin Lesion Detection in Total Body Photography: A Multi-Architecture Benchmark

Muhammad Yamin[1], Praveen Kumar Murali[1], Aritra Das[1], Hayat Rajani[1], Anup Saha[1*], Sheikh Abdul Wahid[2], Solomon Chibuzo Nwafor[1], Sana Nazari[1], Ferrera Nuria[3], Laura Serra-García[3], Clément Lenoir[3], Nuno Gracias[1], Josep Malvehy[3,4], and Rafael Garcia[1]

[1] Computer Vision and Robotics Research Institute, University of Girona, Girona, Spain
[2] University of the Algarve, Faro, Portugal
[3] Dermatology Department, Hospital Clínic Barcelona, Universitat de Barcelona, Barcelona, Spain
[4] CIBER de Enfermedades Raras, Instituto de Salud Carlos III, Barcelona, Spain
[*] Corresponding author: anup.saha@udg.edu

Abstract. Automated skin lesion detection is an essential first step for dermatological screening, yet most work targets lesion classification rather than localisation. Total body photography (TBP) sharpens this gap, producing hundreds of images per patient that each require reliable localisation at scale. We present, to our knowledge, the first systematic multi-architecture benchmark for lesion detection in TBP imagery. Rather than proposing a new detector, we examine design choices that shape detector performance, providing insights to guide their development. Using the iToBoS dataset, we evaluate seven architectures across fourteen configurations: four YOLO variants, two transformer set-prediction detectors (MS-DETR, RT-DETRv2-S), and a distribution-refinement detector (D-FINE-S). Rather than an mAP leaderboard, we characterise reliability through error decomposition (TIDE), AP across IoU thresholds, a confidence × NMS-IoU sweep, and EigenCAM saliency. YOLOv12s with full augmentation gives the best F1 (0.640), while D-FINE-S leads on mAP50 (0.674) and precision (0.707) through probabilistic coordinate refinement, with the smallest dependence on augmentation (+0.17 pp F1 versus +0.8–1.9 pp for the others), and EigenCAM shows its activations are lesion-focused where YOLOv12s exhibits signs of the Clever Hans effect; we therefore recommend D-FINE-S as the primary model. The sweep exposes four distinct confidence regimes (a YOLO plateau, MS-DETR threshold-invariance, and calibrated conf = 0.50 operation for RT-DETRv2-S and D-FINE-S), and TIDE reveals complementary failure modes with distinct clinical implications.

Keywords: Skin lesion detection · Total body photography · YOLO · RT-DETR · MS-DETR · D-FINE · NMS-free detection

1 Introduction

Skin cancer is among the most prevalent cancers worldwide, and melanoma accounts for most skin-cancer deaths despite its low incidence. Early detection is critical: five-year survival exceeds 98% for localised melanoma but falls sharply once the disease spreads [18]. In routine practice lesions are assessed by visual examination and dermoscopy, which is time consuming, depends on clinician expertise and is subject to inter-observer variability. Total body photography (TBP) addresses this setting by documenting the entire skin surface in a single automated session, capturing hundreds of standardised high-resolution images while preserving the anatomical context in which lesions occur. Unlike wide-field photography, which records a single region in one 2D image, TBP captures the whole surface under calibrated conditions, supporting longitudinal surveillance of new, evolving, or "ugly duckling" lesions that signal possible malignancy. It has become an important tool for melanoma surveillance in high-risk individuals.

However, TBP introduces a computational challenge that has received little attention. Before a lesion can be classified it must first be localised within large patches of skin (tens of cm^2), yet most dermatological AI target classification on pre-cropped, lesion-centred images [5,4], assuming the lesion is already identified. This fails in TBP, where lesions sit within complex backgrounds of normal anatomy, variable skin texture, hair, shadows, and imaging artefacts. Detection is therefore a prerequisite for both automated screening and longitudinal monitoring. Despite this, lesion detection in TBP remains largely unexplored, primarily due to the lack of annotated data. The recent release of the iToBoS dataset [16] addresses this gap by providing a large-scale collection of bounding-box-annotated skin-region images acquired with a clinical TBP system.

We present the first systematic multi-architecture benchmark and mechanistic analysis of modern detectors for lesion detection in TBP. Rather than proposing a new detector, we ask which architectural choices most aid localisation, how localisation errors arise, and how detector behaviour should be characterised beyond aggregate scores, reporting all results under a reproducible protocol. Aggregate mAP conflates independent design choices and explains little about *why* one detector localises better than another, so we complement it with error decomposition, behaviour across IoU thresholds, threshold-regime analysis, and saliency. Saliency addresses a significant clinical concern regarding reliability and generalization: a detector can localize a lesion correctly while relying on surrounding skin context rather than the lesion itself, a "Clever Hans" effect [9] that box-overlap metrics cannot reveal. Our contributions are as follows:

- An open-source, multi-architecture benchmark for skin lesion detection, covering fourteen experimental configurations against the iToBoS challenge-winning baseline.
- A mechanistic evaluation protocol beyond mAP, combining error decomposition [3], the AP-versus-IoU-threshold curve, identification of four distinct confidence operating regimes, and EigenCAM [13] to test whether predictions rest on lesion evidence rather than background context.

– A systematic ablation separating augmentation-driven from architecture-driven gains, quantifying each detector's reliance on external regularisation.
– Inference-speed benchmarking characterising the accuracy–efficiency trade-off relevant to clinical deployment at TBP scale.

2 Related Work

A small but growing body of work studies lesion *localisation* from wide-field rather than lesion-centred images, the regime most relevant to TBP. Early efforts operated on 2D wide-area photographs: Lee et al. [10] counted moles from back images, and Mirzaalian et al. [12] matched lesions across back images using anatomical landmarks. Korotkov et al. [8] built a photogrammetry-based total-body scanner tracking pigmented lesions across acquisitions. Closer to clinical screening, Soenksen et al. [19] applied deep networks to wide-field photographs to flag suspicious lesions via the "ugly duckling" criterion, and Birkenfeld et al. [2] used wide-field images for computer-aided triage.

Closest to our work is Skin3D [23], which detects and tracks pigmented lesions on 3D total-body textured meshes by unwrapping the surface to a 2D texture image, localising with a Faster R-CNN [14], and mapping detections back to 3D for graph-based matching on 3DBodyTex [17]. Notably, it reports that strict IoU matching suits small pigmented lesions poorly, motivating a centroid-overlap criterion. Our study is distinct in three respects: we operate on the 2D tiles of the VECTRA WB360 system via iToBoS [16] rather than unwrapped texture maps; we benchmark seven contemporary real-time architectures under one protocol where Skin3D evaluates a single detector; and we add a mechanistic analysis of *why* detectors succeed or fail on TBP imagery.

3 Methodology

For TBP, reliable lesion localisation is the bottleneck, as the detected box determines which skin region is cropped and classified. While classification remains non-trivial, overall performance is predominantly driven by localisation quality. The models perform single-class detection (lesion vs. background). We therefore benchmark seven real-time detectors whose loss functions optimise the regression target directly. These are four CNN-based YOLO variants, **YOLOv8s** [7] (the Kaggle challenge-winning architecture [20]), **YOLOv11s** [6], **YOLOv12s** [22] (attention blocks relating spatially distant regions, aiding context around lesions), and the NMS-free **YOLO-26** [21], all combining Distribution Focal Loss regression with Binary Cross-Entropy or VariFocal classification; two transformer set-prediction detectors, **MS-DETR** [15] and **RT-DETRv2-S** [24], NMS-free by construction via Hungarian matching; and **D-FINE-S** [11], which reframes regression as iterative refinement of per-coordinate distributions (Fine-Grained Distribution Loss over K=17K bins) and retains NMS.

We use the iToBoS dataset [16] for all experiments. For reproducibility, we adopt the publicly released challenge partition, comprising 8,473 images and

29,403 bounding-box annotations from 49 distinct patients for model training. As preprocessing, a Gaussian blur was applied to detected background regions (scanner frame, room surfaces) to suppress environmental noise, while the patient skin surface was retained at full quality; all images are 1024^2 px with no spatial resizing applied.

All associated code and trained models will be made available at `https://github.com/CIRS-Girona/tbp_lesion_detection`

3.1 Training Protocol

To isolate the effect of data augmentation, each architecture was trained under two conditions. The *Full-Aug* setting combined mosaic composition (four-image concatenation, $p = 0.85$–0.99), horizontal and vertical flips, HSV saturation/value jitter, random scale, translate, and minor rotation, with parameters set by the hyperparameter sweep. Strong hue jitter and heavy shear were deliberately excluded: in dermatology the colour signature and border morphology of a lesion carry diagnostic information, and corrupting them would degrade model quality. The *No-Aug* setting retained only horizontal flip, testing the performance ceiling on unmodified clinical images.

For each YOLO architecture we ran a 25-trial Bayesian sweep with Weights & Biases [1], maximising validation F1 over 40 epochs per trial. For MS-DETR and D-FINE-S we used 15 trials of 30 epochs each, reflecting their higher per-trial cost. The sweep spanned 15 parameters: learning rate (`lr0`, `lrf`), momentum, weight decay, warmup epochs, box/cls/dfl loss weights, IoU threshold, dropout, and augmentation parameters. The best configuration from each sweep was used for the full 100-epoch run.

All models used the AdamW optimiser, a cosine schedule with linear warm-up, and patience-based early stopping (20 epochs on validation F1). Training ran on a single NVIDIA A100 80 GB GPU with batch size of 8.

3.2 Evaluation Protocol

All models were evaluated on the held-out test set at fixed IoU threshold of 0.7. Beyond aggregate scores, three diagnostics characterised localisation behaviour. First, a two-dimensional sweep over an 11×9 grid of confidence (0.05–0.70) and NMS-IoU (0.30–0.75) thresholds maps each model's operating regime and quantified its sensitivity to both parameters. Second, TIDE [3] decomposed the mAP gap into localisation error, duplicates, false positives, and missed-detections. Third, AP was recomputed across the nine COCO IoU thresholds (0.50–0.95) to expose sensitivity to clinician annotation variability that single-threshold mAP conceals. Finally, for the two leading detectors (D-FINE-S and YOLOv12s) we used EigenCAM [13] on the backbone features to analyze whether predictions rely on lesion morphology or on spurious cues.

4 Results and Discussion

4.1 Cross-Architecture Performance

Table 1 reports all fourteen configurations on the held-out test set. Two models define the practical frontier: YOLOv12s with augmentations attains the highest F1 (0.640) at conf = 0.20, while D-FINE-S with augmentations attains the highest mAP50 (0.674) and the highest precision (0.707) at conf = 0.50. The two are effectively tied on F1 (0.1 pp apart) but represent different optima, a peak operating point versus stronger box-regression area.

The F1 spread across full-augmentation models is only 1.3 pp (rank 1 to rank 7), indicating that the current F1 ceiling is set by dataset characteristics rather than architecture. At the deployment threshold, recall ranges from 0.569 to 0.625, so roughly 37–43% of annotated lesions are still missed. Clinically this is the main limitation: on a moderate-scale TBP corpus the binding constraint is data, not detector choice, and no architecture in this study yet reaches recall suitable for unsupervised screening. All evaluated detectors improve on the prior Kaggle challenge-winning baseline (F1 = 0.598 at its default configuration).

Inference cost (Table 2) separates the candidates further. YOLOv26s is the throughput leader (46.5 FPS GPU) by removing NMS, and D-FINE-S combines competitive accuracy with low GPU latency (14.7 ms), whereas MS-DETR is impractical on CPU (0.5 FPS). For deployment the relevant trade-off is between D-FINE-S (precision, calibrated gate) and the YOLO variants (recall, speed).

TIDE decomposition [3] exposes a clinically important split between the two leading detectors (Table 3). YOLOv12s has very few background false positives but a high miss count, whereas D-FINE-S inverts this, with few misses but many background candidates that are then removed by its conf = 0.50 gate, which is the reason for its high precision. Localisation errors are comparable across both, so neither holds a decisive box-regression advantage. Although YOLOv12s has the least false positives, it fails to detect a much larger proportion of lesions—a significant failure mode in the clinical setting.

4.2 Four Confidence Operating Paradigms

Sweeping confidence (0.05–0.70) and NMS-IoU (0.30–0.75) reveals four qualitatively distinct confidence behaviours (Fig. 1), which we believe are new to the skin-lesion detection literature and directly govern how a clinician must set each model's operating point. YOLO models share a low-confidence plateau: F1 is flat between conf = 0.05 and 0.25, making conf = 0.20 a robust default. RT-DETRv2-S and D-FINE-S are calibrated to a high, clinically intuitive threshold (conf = 0.50), behaving as a reliable detection gate that fires only when confident. MS-DETR is completely threshold-invariant: its Hungarian matching produces a bimodal score distribution (matched >0.70, unmatched <0.05), so every confidence and NMS-IoU combination yields identical metrics and no tuning is required. The practical message for deployment is that the operating point is an architectural property, not a free hyperparameter to be guessed per site.

Table 1: Cross-architecture comparison on held-out test set. mAP50 is reported at conf = 0.001 (YOLO/MS-DETR) or at best-F1 threshold (D-FINE-S†, RT-DETRv2-S†). mAP50-95 is reported at conf = 0.001. P: Precision; R: Recall

Rank	Model	Aug	Conf	F1	mAP50	mAP50-95	P	R	NMS-free
1	**YOLOv12s**	Full	0.20	**0.640**	0.652	0.351	0.680	0.604	No
2	D-FINE-S	Full	0.50	0.639	**0.674**†	0.361	**0.707**	0.583	No
3	YOLOv26s	Full	0.20	0.634	0.644	0.343	0.680	0.593	Yes
4	RT-DETRv2-S	Full	0.50	0.633	0.668†	0.361	0.703	0.576	Yes
5	YOLOv8s (Kaggle)	Full	0.30	0.632	0.644	0.342	0.656	0.609	No
6	YOLOv11s	Full	0.20	0.631	0.643	0.343	0.657	0.606	No
7	YOLOv8s (ours)	Full	0.20	0.627	0.622	0.328	0.629	0.625	No
8	D-FINE-S	None	0.50	0.637	0.659†	0.344	0.658	0.617	No
9	YOLOv12s	None	0.20	0.630	0.619	0.320	0.659	**0.620**	No
10	RT-DETRv2-S	None	0.50	0.621	0.649†	0.341	0.612	0.614	Yes
11	YOLOv8s (ours)	None	0.20	0.619	0.602	0.316	0.661	0.582	No
12	YOLOv11s	None	0.20	0.616	0.609	0.314	0.660	0.578	No
13	YOLOv26s	None	0.20	0.615	0.596	0.312	0.670	0.569	Yes
14	MS-DETR	Full	–	0.614	0.642	0.341	0.643	0.588	Yes
15	MS-DETR	None	–	0.599	0.616	0.324	0.642	0.561	Yes
16	YOLOv8s (Kaggle)	None	0.325	0.598	0.565	0.289	0.654	0.551	No

Table 2: Inference speed. GPU: NVIDIA A100 80 GB, 1024 px. CPU: server Intel Xeon Silver, 1024 px. GFLOPs at 640 px for YOLO variants and MS-DETR.

Model	Params	GFLOPs	Size	GPU Lat.	GPU FPS	CPU Lat.	CPU FPS	NMS-free
YOLOv26s	10.01M	22.8	19.4 MB	**21.5 ms**	**46.5**	206.3 ms	4.8	Yes
YOLOv12s	9.29M	21.7	18.1 MB	28.4 ms	35.2	333.1 ms	3.0	No
YOLOv11s	9.46M	21.7	18.3 MB	16.3 ms	61.3	161.5 ms	6.2	No
YOLOv8s	11.17M	28.8	21.5 MB	11.9 ms	84.2	158.3 ms	6.3	No
RT-DETRv2-S	20.08M	76.0	80.8 MB	23.3 ms	42.9	214.9 ms	4.7	Yes
MS-DETR	13.14M	58.3	51.6 MB	29.6 ms	33.8	1995.6 ms	0.5	Yes
D-FINE-S	10.18M	29.6	41.4 MB	14.7 ms	67.8	279.8 ms	3.6	No

Table 3: TIDE error decomposition on the held-out test set (lower is better). Loc = localisation error, FP = False Positives, FN = False Negatives.

Model	Loc.	Duplicates	FP	FN
D-FINE-S (Full-Aug)	4.35	0.21	**20.55**	**1.02**
D-FINE-S (No-Aug)	3.53	1.49	19.13	1.36
YOLOv12s (Full-Aug)	**3.33**	**0.01**	8.44	23.81

NMS-IoU has negligible effect throughout (F1 spread $\leq$0.83 pp for YOLO; 0.00 pp for MS-DETR). In TBP, lesions occupy distinct anatomical sites and rarely overlap, so two boxes for one image typically have IoU below 0.10 and suppression is never triggered. The NMS-free advantage of YOLOv26s and the transformers is thus neutralised on accuracy and surfaces only as a speed gain; on dense-object domains the conclusion would differ.

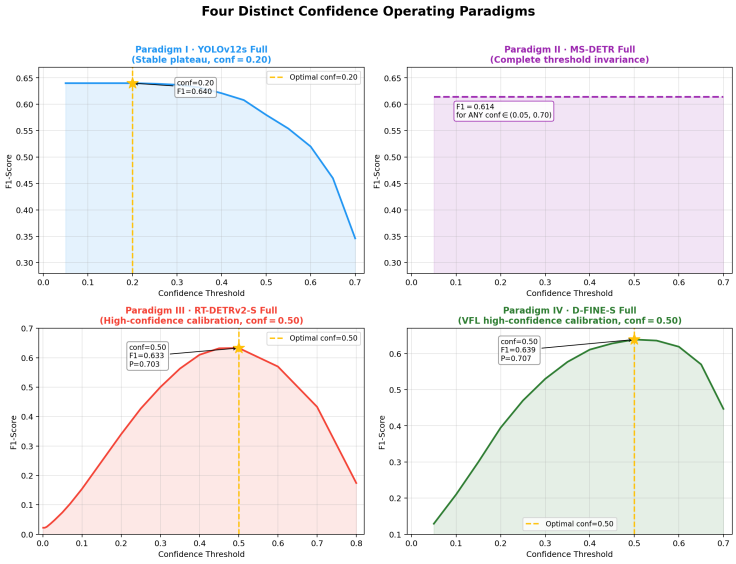

Fig. 1: Four confidence operating paradigms. Gold stars mark each model's recommended clinical operating point.

4.3 Saliency and the Clever Hans Check

A correct box does not guarantee correct evidence. Applying EigenCAM to the two leading detectors (Fig. 2) shows a meaningful reliability difference. D-FINE-S activations are spatially distributed with elevated response at lesion sites, consistent with global context modelling. YOLOv12s activations concentrate on background regions or high-contrast hair and follicular texture even where no confident detection exists, a candidate Clever Hans signal [9] suggesting the head may exploit skin-texture shortcuts rather than lesion features. This matters clinically because a detector keyed to hair contrast may degrade on hairless sites or in patients with sparse body hair. The evidence is qualitative (EigenCAM probes the backbone, not the task head), so we frame it as motivating two follow-ups before clinical deployment: evaluating both models on hairless sites, and applying a quantitative faithfulness test (insertion/deletion AUC) to measure the share of output attributable to lesion versus background.

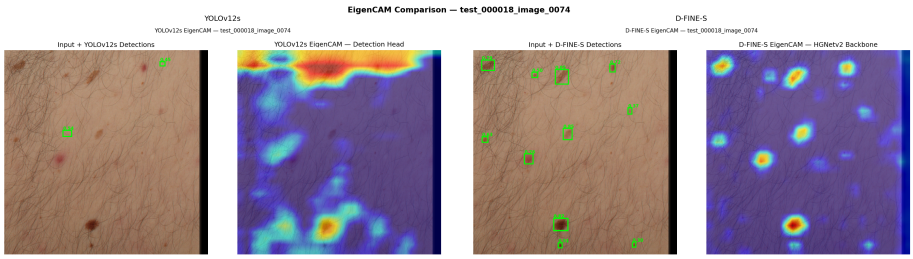

Fig. 2: EigenCAM saliency (red = high activation). D-FINE-S responds at lesion sites with distributed context; YOLOv12s concentrates on hair and follicular texture, a potential Clever Hans shortcut.

4.4 Localisation Under Tight IoU

Across COCO IoU thresholds, D-FINE-S Full falls from AP@50 = 67.4% to AP@75 = 34.5% and AP@95 = 0.3%. The steep AP@50-to-AP@75 drop reflects clinician annotation variability rather than detector failure: dermoscopic lesion borders carry inherent inter-annotator disagreement, so tight IoU rarely aligns with a single annotated boundary. This agrees with Skin3D [23], where strict IoU matching of small pigmented lesions proved unreliable and motivated a centroid-overlap criterion. For clinicians the implication is that high-IoU AP penalises annotation noise, not localisation quality, and centroid- or region-level agreement is the more meaningful target for TBP triage.

4.5 Augmentation is Architecture-Dependent

Full augmentation helps six of seven architectures (+0.8 to +1.9 pp F1), consistent with augmentations extending productive learning beyond the point at which a relatively small training set saturates (diversity exhaustion). D-FINE-S is the exception, with a near-zero delta (+0.17 pp): its Fine-Grained Distribution Loss (FGL), which supervises a full coordinate distribution rather than a point estimate, supplies implicit regularisation roughly equivalent to external augmentation. Notably D-FINE-S without augmentations (F1 = 0.637) still beats every other *No-Aug* model and most *Full-Aug* models, which is useful where augmentation pipelines are unavailable. The benefit comes from the loss, providing strong implicit regularisation: predicting a calibrated 17-bin coordinate distribution prevents overfitting, reducing the dependence on external regularization.

5 Conclusion

We present a systematic benchmark for skin lesion detection on the iToBoS TBP dataset, spanning fourteen configurations across seven real-time architectures, complementing aggregate metrics with a mechanistic protocol aimed at detector

reliability rather than leaderboard position. D-FINE-S emerges as the primary model, attaining the highest mAP50 (0.674) and lowest false negatives (1.02) with the smallest dependence on augmentation, and EigenCAM confirms its activations are lesion-focused. However, three key findings carry clinical weight. First, confidence operating behaviour is an architectural property, not a free parameter: YOLO models share a low-confidence plateau (conf = 0.20), MS-DETR is completely threshold-invariant, and RT-DETRv2-S and D-FINE-S operate at a calibrated conf = 0.50 gate, so the operating point need not be guessed per site. Second, TIDE reveals complementary failure modes, background false positives for D-FINE-S and missed detections for YOLOv12s, that should govern how each is operated. Third, the steep fall of AP at tight IoU reflects clinician annotation variability, not localisation failure, indicating centroid- or region-level agreement as the better target for TBP triage. Recall remains the limiting factor at this scale, making these detectors decision-support tools, not autonomous screens. With this, we establish a new benchmark for future methods, setting D-FINE-S as the performance threshold to be surpassed without compromising the meaningfulness of the resulting attribution maps.

Acknowledgments. This study was supported by the European Academy of Dermatology and Venereology (EADV) through the project HARMONY: A Holistic AI-Driven Approach to Melanoma Risk Assessment Integrating Clinical, Phenotypic, Genotypic, and Imaging Data (reference no. PPRC-20250011).

Disclosure of Interests. J. Malvehy is co-founder of Diagnosis Dermatologica, holding stock and receiving fees from Dermavision Solutions. J. Malvehy is the chairman of the Task Force on Artificial Intelligence of the EADV. J. Malvehy is a Medical Consultant for Canfield Scientific, Inc. and receives fees for lecturing from Fotofinder.

References

1. Biewald, L.: Experiment tracking with Weights and Biases (2020), `https://www.wandb.com`
2. Birkenfeld, J.S., Tucker-Schwartz, J.M., Soenksen, L.R., Avilés-Izquierdo, J.A., Marti-Fuster, B.: Computer-aided classification of suspicious pigmented lesions using wide-field images. Computer Methods and Programs in Biomedicine **195**, 105631 (2020)
3. Bolya, D., Foley, S., Hays, J., Hoffman, J.: TIDE: A general toolbox for identifying object detection errors. In: Computer Vision – ECCV 2020. Lecture Notes in Computer Science, vol. 12348, pp. 558–573. Springer (2020). `https://doi.org/10.1007/978-3-030-58580-8_33`
4. Codella, N.C.F., et al.: Skin lesion analysis toward melanoma detection: A challenge at the 2017 ISBI. In: Proc. IEEE International Symposium on Biomedical Imaging (ISBI). pp. 168–172 (2018)
5. Esteva, A., et al.: Dermatologist-level classification of skin cancer with deep neural networks. Nature **542**(7639), 115–118 (2017)
6. Jocher, G., Chaurasia, A., Qiu, J.: Ultralytics YOLO11. Ultralytics (2024), `https://github.com/ultralytics/ultralytics`

7. Jocher, G., et al.: YOLOv8 by Ultralytics. Ultralytics (2023), https://github.com/ultralytics/ultralytics
8. Korotkov, K., Quintana, J., Puig, S., Malvehy, J., Garcia, R.: A new total body scanning system for automatic change detection in multiple pigmented skin lesions. IEEE Transactions on Medical Imaging **34**(1), 317–338 (2015)
9. Lapuschkin, S., Wäldchen, S., Binder, A., Montavon, G., Samek, W., Müller, K.R.: Unmasking clever hans predictors and assessing what machines really learn. Nature Communications **10**(1), 1096 (2019). https://doi.org/10.1038/s41467-019-08987-4
10. Lee, T.K., Atkins, M.S., King, M.A., Lau, S., McLean, D.I.: Counting moles automatically from back images. IEEE Transactions on Biomedical Engineering **52**(11), 1966–1969 (2005)
11. Ma, Y., et al.: D-FINE: Redefine regression task in DETRs as fine-grained distribution refinement. In: Proc. International Conference on Learning Representations (ICLR). Singapore (2025)
12. Mirzaalian, H., Lee, T.K., Hamarneh, G.: Skin lesion tracking using structured graphical models. Medical Image Analysis **27**, 84–92 (2016)
13. Muhammad, M.B., Yeasin, M.: Eigen-CAM: Class activation map using principal components. In: Proc. International Joint Conference on Neural Networks (IJCNN). pp. 1–7 (2020)
14. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems (NeurIPS) (2015)
15. Saha, A., et al.: MS-DETR: Multi-scale detection transformer (2024), arXiv preprint
16. Saha, A., et al.: Skin region images extracted from 3D total body photographs for lesion detection. Scientific Data **12**(1), 1442 (2025). https://doi.org/10.1038/s41597-025-05483-x
17. Saint, A., Ahmed, E., Shabayek, A.E.R., Cherenkova, K., Gusev, G., Aouada, D., Ottersten, B.: 3DBodyTex: Textured 3D body dataset. In: Proc. International Conference on 3D Vision (3DV). pp. 495–504 (2018)
18. Siegel, R.L., Miller, K.D., Wagle, N.S., Jemal, A.: Cancer statistics, 2023. CA: A Cancer Journal for Clinicians **73**(1), 17–48 (2023)
19. Soenksen, L.R., Kassis, T., Conover, S.T., et al.: Using deep learning for dermatologist-level detection of suspicious pigmented skin lesions from wide-field images. Science Translational Medicine **13**(581), eabb3652 (2021)
20. Solomon, T.: Skin lesion detection on iToBoS dataset. Kaggle Competition Solution (2024), https://www.kaggle.com/competitions/itobos-2024-detection
21. Wang, X., et al.: YOLO26: Real-time NMS-free object detection (2025), arXiv preprint
22. Wang, Y., et al.: YOLOv12: Attention-centric real-time object detectors. arXiv preprint arXiv:2502.12524 (2025)
23. Zhao, M., Kawahara, J., Abhishek, K., Shamanian, S., Hamarneh, G.: Skin3D: Detection and longitudinal tracking of pigmented skin lesions in 3D total-body textured meshes. Medical Image Analysis **77**, 102329 (2022). https://doi.org/10.1016/j.media.2021.102329
24. Zhao, Y., et al.: DETRs beat YOLOs on real-time object detection. In: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 16965–16974 (2024)