Online Resource 1

How Much Can Per-Image Classical Filter Selection Help Frozen Object Detectors in Low Light and Haze?

Signal, Image and Video Processing

Melikşah Beşir and E. İlhan Konukseven

Corresponding author: Melikşah Beşir, Robotics Graduate Program, Middle East Technical University, Ankara, Türkiye; besir.meliksah@metu.edu.tr

This file contains the formal statement behind the re-measurement test (Section S1), the full result tables (Tables S1–S7) and qualitative examples (Figure S1). The per-image true-positive counts from which all tables except the mAP values can be recomputed are provided as Online Resource 2.

Terms used in the tables. *Baseline* is no filtering (the identity filter). *Static filter* is the best single filter, applied to every image; in Tables S1–S3 it is chosen on all evaluation images, and in Tables S4–S7 it is chosen on the training folds and scored on held-out images. *In-sample oracle* is the per-image oracle scored on the copy on which it chose; *re-scored oracle* is the same choice scored on the other noisy copy ("oracle checked on the other copy" in the main text). *Router* is the ridge-regression router on eight global image statistics. *Regret* is the shortfall in true positives (TP) relative to the oracle on the same image; *normalized regret* divides it by the number of annotated objects. *Transfer* is the share of the in-sample oracle's gain over the baseline that survives re-scoring.

S1 Why the two oracle values bracket the reproducible gain

For image x and filter $f \in \mathcal{F}$, let $s_f^A(x)$ and $s_f^B(x)$ be the TP counts on the two noisy copies, and let $\mathcal{O}_A(x)$ and $\mathcal{O}_B(x)$ be the sets of filters that reach the maximum TP count on each copy. The re-scored oracle averages the other copy's count over the tied set,

$$S_{\mathrm{re}} = \tfrac{1}{2}\,\mathbb{E}_x\Big[\tfrac{1}{|\mathcal{O}_A(x)|}\textstyle\sum_{f\in\mathcal{O}_A(x)} s_f^B(x) + \tfrac{1}{|\mathcal{O}_B(x)|}\sum_{f\in\mathcal{O}_B(x)} s_f^A(x)\Big], \tag{1}$$

and the in-sample oracle is $S_{\mathrm{orc}} = \frac{1}{2}\mathbb{E}_x[\max_f s_f^A(x) + \max_f s_f^B(x)]$. Let

$$\mu_f(x) = \mathbb{E}\big[s_f^A(x) \mid x\big] = \mathbb{E}\big[s_f^B(x) \mid x\big], \qquad V_{\mathrm{rep}} = \mathbb{E}_x\left[\max_{f\in\mathcal{F}} \mu_f(x)\right], \tag{2}$$

where the conditional expectations are over the noise. V_{rep} is the value of a per-image choice that knows the expected TP count of every filter on every image, so $V_{\mathrm{rep}} - S_{\mathrm{base}}$ is the per-image gain that survives re-measurement.

Proposition S1 (Bracketing; after van Hasselt, 2010)**.** *Suppose the two copies of each image receive independent, identically distributed perturbations, and the mapping from a perturbed image to its TP counts is deterministic, as it is for a frozen detector. Then* $\mathbb{E}[S_{\mathrm{re}}] \le V_{\mathrm{rep}} \le \mathbb{E}[S_{\mathrm{orc}}]$*, where the expectations are over the perturbations. The left inequality holds for any choice made on one copy and scored on the other, including the first tied filter in bank order.*

Proof. The argument follows Lemma 1 of van Hasselt (2010, *Double Q-learning*, NeurIPS), extended to the average over tied filters. Fix x. The optimal set $\mathcal{O}_A(x)$ depends only on the perturbation of copy A, and the counts $s_f^B(x)$ depend only on the independent perturbation of copy B. Hence

$$\mathbb{E}\Big[\tfrac{1}{|\mathcal{O}_A(x)|}\textstyle\sum_{f\in\mathcal{O}_A(x)} s_f^B(x) \;\Big|\; x, \mathcal{O}_A(x)\Big] = \tfrac{1}{|\mathcal{O}_A(x)|}\sum_{f\in\mathcal{O}_A(x)} \mu_f(x) \le \max_{f\in\mathcal{F}} \mu_f(x),$$

and the same holds with A and B exchanged. Averaging over the two directions and over images gives $\mathbb{E}[S_{\mathrm{re}}] \le V_{\mathrm{rep}}$. For the right inequality, $\max_f \mu_f(x) = \max_f \mathbb{E}[s_f^A(x) \mid x] \le \mathbb{E}[\max_f s_f^A(x) \mid x]$, and likewise for copy B. □

The in-sample oracle is therefore an upper estimate (the optimizer's curse; Smith and Winkler, 2006) and the re-scored oracle, the double estimator of van Hasselt (2010), is a lower estimate. The best and worst tied filters on the target copy (Table S4) use the target outcome, so the bound does not apply to them. A router $\pi(x) = g(X(x))$ that uses image features X computed from the unperturbed image cannot exceed V_{rep} either: its best achievable value $V^*(X) = \mathbb{E}_X[\max_f \mathbb{E}[s_f(x) \mid X]]$ satisfies $S_{\mathrm{stat}} \le V^*(X) \le V_{\mathrm{rep}}$ by conditioning. Our experiments estimate the value of one router family, not $V^*(X)$.

Table S1: RQ1 on the original images: corpus mAP@50 of the baseline (no preprocessing), the best static filter, and oracle-selected detections. The oracle chooses, for each image, a filter that maximizes the TP count; the resulting mAP is a consequence of that choice, not an upper bound on mAP. Gains are mAP@50 points over the baseline.

Data	Detector	Baseline	Best static filter (mAP)	Oracle	Gain, static	Gain, oracle
ExDark	YOLOv3u	66.4	$\gamma = 0.8$ (66.8)	69.5	+0.37	+3.14
	YOLOv5lu	65.9	$\gamma = 0.8$ (66.5)	69.6	+0.55	+3.67
	YOLOv8l	66.8	$\gamma = 0.8$ (66.8)	70.0	+0.01	+3.22
	RF-DETR	71.9	$\gamma = 0.8$ (72.9)	74.0	+0.98	+2.07
RTTS	YOLOv3u	62.3	CLAHE, clip 4 (63.6)	66.0	+1.26	+3.73
	YOLOv5lu	63.4	CLAHE, clip 2 (64.1)	66.3	+0.73	+2.91
	YOLOv8l	63.2	CLAHE, clip 2 (64.3)	65.8	+1.09	+2.60
	RF-DETR	70.9	CLAHE, clip 2 (71.9)	73.4	+1.01	+2.54

S2 Full results

Table S2: Absolute and normalized TP regret on the original images for the baseline (B) and the best static filter (S), chosen on all evaluation images. Absolute regret is the mean TP shortfall per image relative to the oracle; normalized regret divides it by the number of annotated objects. "Ties oracle" and "within 1 TP" are the shares of images with zero regret and with a regret of at most one TP.

Data	Detector	B, absolute	S, absolute	B, normalized	S, normalized	B: ties oracle / within 1 TP (%)	S: ties oracle / within 1 TP (%)
ExDark	YOLOv3u	0.214	0.181	0.076	0.064	80.6 / 98.2	83.4 / 98.6
	YOLOv5lu	0.200	0.166	0.071	0.061	82.5 / 97.8	84.9 / 98.8
	YOLOv8l	0.194	0.172	0.066	0.061	82.1 / 98.8	84.4 / 98.6
	RF-DETR	0.192	0.154	0.061	0.050	84.5 / 97.2	86.6 / 98.6
RTTS	YOLOv3u	0.604	0.448	0.058	0.049	62.7 / 86.2	67.9 / 92.2
	YOLOv5lu	0.573	0.391	0.051	0.040	64.8 / 87.8	72.0 / 92.9
	YOLOv8l	0.551	0.425	0.058	0.047	64.7 / 88.3	69.7 / 92.5
	RF-DETR	0.435	0.351	0.041	0.033	69.7 / 91.1	74.8 / 92.9

Table S3: TP regret on the original images by the number of annotated objects per image, for the baseline (B) and the best static filter (S) chosen on all evaluation images.

Data	Detector	Objects	Images	B, normalized	S, normalized	B, absolute	S, absolute
ExDark	YOLOv3u	1	567	0.079	0.069	0.079	0.069
		2–5	684	0.082	0.067	0.241	0.196
		6–10	160	0.053	0.043	0.400	0.325
		> 10	56	0.051	0.050	0.714	0.732
	YOLOv5lu	1	567	0.083	0.076	0.083	0.076
		2–5	684	0.066	0.054	0.199	0.158
		6–10	160	0.057	0.047	0.412	0.344
		> 10	56	0.051	0.041	0.786	0.661
	YOLOv8l	1	567	0.067	0.069	0.067	0.069
		2–5	684	0.070	0.058	0.202	0.173
		6–10	160	0.051	0.047	0.381	0.350
		> 10	56	0.058	0.046	0.857	0.696
	RF-DETR	1	567	0.055	0.044	0.055	0.044
		2–5	684	0.067	0.057	0.199	0.165
		6–10	160	0.056	0.044	0.419	0.325
		> 10	56	0.058	0.044	0.857	0.643
RTTS	YOLOv3u	1	122	0.025	0.033	0.025	0.033
		2–5	301	0.063	0.054	0.229	0.199
		6–10	200	0.069	0.055	0.535	0.430
		> 10	241	0.061	0.045	1.423	0.983
	YOLOv5lu	1	122	0.016	0.025	0.016	0.025
		2–5	301	0.054	0.043	0.186	0.140
		6–10	200	0.059	0.042	0.455	0.330
		> 10	241	0.059	0.041	1.436	0.942
	YOLOv8l	1	122	0.025	0.049	0.025	0.049
		2–5	301	0.066	0.044	0.216	0.146
		6–10	200	0.069	0.052	0.530	0.395
		> 10	241	0.055	0.044	1.253	0.988
	RF-DETR	1	122	0.008	0.008	0.008	0.008
		2–5	301	0.046	0.041	0.169	0.150
		6–10	200	0.048	0.033	0.355	0.260
		> 10	241	0.046	0.035	1.050	0.851

Table S4: Tie handling in the reproducibility test (noisy copies). The first two columns give the mean optimal-set size and the share of images with a tied optimum on copies A and B. The remaining columns give the re-scored oracle headroom over the out-of-fold static filter, in TP per image, when the source choice is the uniform average over all optimal filters (primary estimate, as in Table 2 of the main text), the first optimal filter in bank order, or the best or worst tied filter on the target copy. The last two inspect the target outcome and only bound the effect of ties. Brackets: 95% bootstrap CIs.

Data	Detector	Optimal-set size (A/B)	Tied optimum (%)	Uniform average	First in bank order	Best tied	Worst tied
ExDark	YOLOv3u	11.40/11.31	93.0/93.3	0.092 [0.075, 0.110]	0.091 [0.071, 0.111]	0.190 [0.170, 0.211]	−0.126 [−0.149, −0.103]
	YOLOv5lu	11.67/11.53	93.9/92.9	0.082 [0.065, 0.099]	0.080 [0.061, 0.099]	0.176 [0.156, 0.197]	−0.123 [−0.145, −0.100]
	YOLOv8l	11.61/11.55	95.0/94.4	0.064 [0.048, 0.080]	0.064 [0.047, 0.082]	0.156 [0.137, 0.176]	−0.159 [−0.180, −0.137]
	RF-DETR	12.37/12.40	95.8/96.0	0.090 [0.074, 0.108]	0.089 [0.071, 0.106]	0.148 [0.129, 0.167]	−0.068 [−0.088, −0.047]
RTTS	YOLOv3u	9.61/9.53	87.0/85.2	0.185 [0.145, 0.228]	0.148 [0.103, 0.194]	0.381 [0.336, 0.432]	−0.120 [−0.166, −0.072]
	YOLOv5lu	9.52/9.48	84.5/84.6	0.166 [0.124, 0.211]	0.126 [0.081, 0.172]	0.346 [0.302, 0.394]	−0.115 [−0.165, −0.063]
	YOLOv8l	9.58/9.53	85.0/86.6	0.181 [0.137, 0.228]	0.134 [0.086, 0.186]	0.373 [0.328, 0.422]	−0.127 [−0.177, −0.072]
	RF-DETR	10.62/10.55	89.0/88.2	0.150 [0.114, 0.189]	0.112 [0.071, 0.153]	0.286 [0.245, 0.329]	−0.108 [−0.152, −0.062]

Table S5: Reproducibility test with 95% bootstrap CIs for every quantity (noisy copies). Columns 3 to 5 are gains over the baseline in TP per image; column 6 is the reproducible headroom over the out-of-fold static filter; transfer is the share of the oracle's gain over the baseline that survives re-scoring.

Data	Detector	Oracle, in-sample	Oracle, re-scored	Static	Re-scored oracle − static	Transfer (%)
ExDark	YOLOv3u	0.256 [0.231, 0.282]	0.123 [0.103, 0.144]	0.031 [0.015, 0.047]	0.092 [0.075, 0.110]	48.1 [43.1, 52.6]
	YOLOv5lu	0.254 [0.229, 0.280]	0.133 [0.112, 0.154]	0.051 [0.035, 0.069]	0.082 [0.065, 0.099]	52.4 [47.7, 56.8]
	YOLOv8l	0.242 [0.218, 0.268]	0.113 [0.093, 0.133]	0.049 [0.033, 0.065]	0.064 [0.048, 0.080]	46.6 [41.5, 51.3]
	RF-DETR	0.210 [0.185, 0.236]	0.137 [0.116, 0.159]	0.046 [0.030, 0.063]	0.090 [0.074, 0.108]	65.1 [60.3, 69.4]
RTTS	YOLOv3u	0.637 [0.573, 0.704]	0.333 [0.279, 0.391]	0.148 [0.092, 0.206]	0.185 [0.145, 0.228]	52.3 [47.3, 56.8]
	YOLOv5lu	0.696 [0.626, 0.771]	0.385 [0.320, 0.451]	0.219 [0.161, 0.280]	0.166 [0.124, 0.211]	55.3 [50.0, 59.9]
	YOLOv8l	0.640 [0.577, 0.706]	0.313 [0.256, 0.372]	0.133 [0.078, 0.188]	0.181 [0.137, 0.228]	48.9 [43.2, 54.1]
	RF-DETR	0.524 [0.465, 0.586]	0.297 [0.246, 0.351]	0.148 [0.101, 0.196]	0.150 [0.114, 0.189]	56.8 [51.3, 61.7]

Table S6: Ridge router with the log-compressed Laplacian feature (primary analysis; noisy copies), in TP per image with 95% bootstrap CIs. Router models and static filters are selected within outer training folds and scored on the other copy of the held-out images.

Data	Detector	Router − baseline	Static − baseline	Router − static
ExDark	YOLOv3u	+0.0297 [0.0125, 0.0464]	+0.0309 [0.0151, 0.0471]	−0.0013 [−0.0108, 0.0082]
	YOLOv5lu	+0.0427 [0.0253, 0.0606]	+0.0515 [0.0346, 0.0691]	−0.0088 [−0.0194, 0.0018]
	YOLOv8l	+0.0404 [0.0240, 0.0570]	+0.0487 [0.0332, 0.0651]	−0.0083 [−0.0170, 0.0003]
	RF-DETR	+0.0435 [0.0255, 0.0620]	+0.0464 [0.0299, 0.0635]	−0.0029 [−0.0124, 0.0065]
RTTS	YOLOv3u	+0.1287 [0.0760, 0.1811]	+0.1480 [0.0922, 0.2061]	−0.0193 [−0.0375, −0.0015]
	YOLOv5lu	+0.2230 [0.1660, 0.2814]	+0.2188 [0.1609, 0.2795]	+0.0043 [−0.0164, 0.0248]
	YOLOv8l	+0.1280 [0.0779, 0.1795]	+0.1325 [0.0781, 0.1881]	−0.0045 [−0.0225, 0.0132]
	RF-DETR	+0.1313 [0.0875, 0.1780]	+0.1476 [0.1007, 0.1962]	−0.0163 [−0.0319, 0.0000]

Table S7: Sensitivity analysis: the same router with the raw instead of the log-compressed Laplacian variance (noisy copies), in TP per image with 95% bootstrap CIs. All other settings are identical to Table S6.

Data	Detector	Router − baseline	Static − baseline	Router − static
ExDark	YOLOv3u	+0.0305 [0.0142, 0.0469]	+0.0309 [0.0151, 0.0471]	−0.0004 [−0.0098, 0.0087]
	YOLOv5lu	+0.0434 [0.0261, 0.0607]	+0.0515 [0.0346, 0.0691]	−0.0081 [−0.0183, 0.0019]
	YOLOv8l	+0.0369 [0.0206, 0.0539]	+0.0487 [0.0332, 0.0651]	−0.0118 [−0.0204, −0.0033]
	RF-DETR	+0.0391 [0.0224, 0.0562]	+0.0464 [0.0299, 0.0635]	−0.0074 [−0.0152, 0.0005]
RTTS	YOLOv3u	+0.1238 [0.0709, 0.1769]	+0.1480 [0.0922, 0.2061]	−0.0242 [−0.0449, −0.0028]
	YOLOv5lu	+0.2284 [0.1698, 0.2889]	+0.2188 [0.1609, 0.2795]	+0.0096 [−0.0140, 0.0331]
	YOLOv8l	+0.1277 [0.0758, 0.1793]	+0.1325 [0.0781, 0.1881]	−0.0049 [−0.0269, 0.0168]
	RF-DETR	+0.1248 [0.0807, 0.1716]	+0.1476 [0.1007, 0.1962]	−0.0228 [−0.0400, −0.0054]

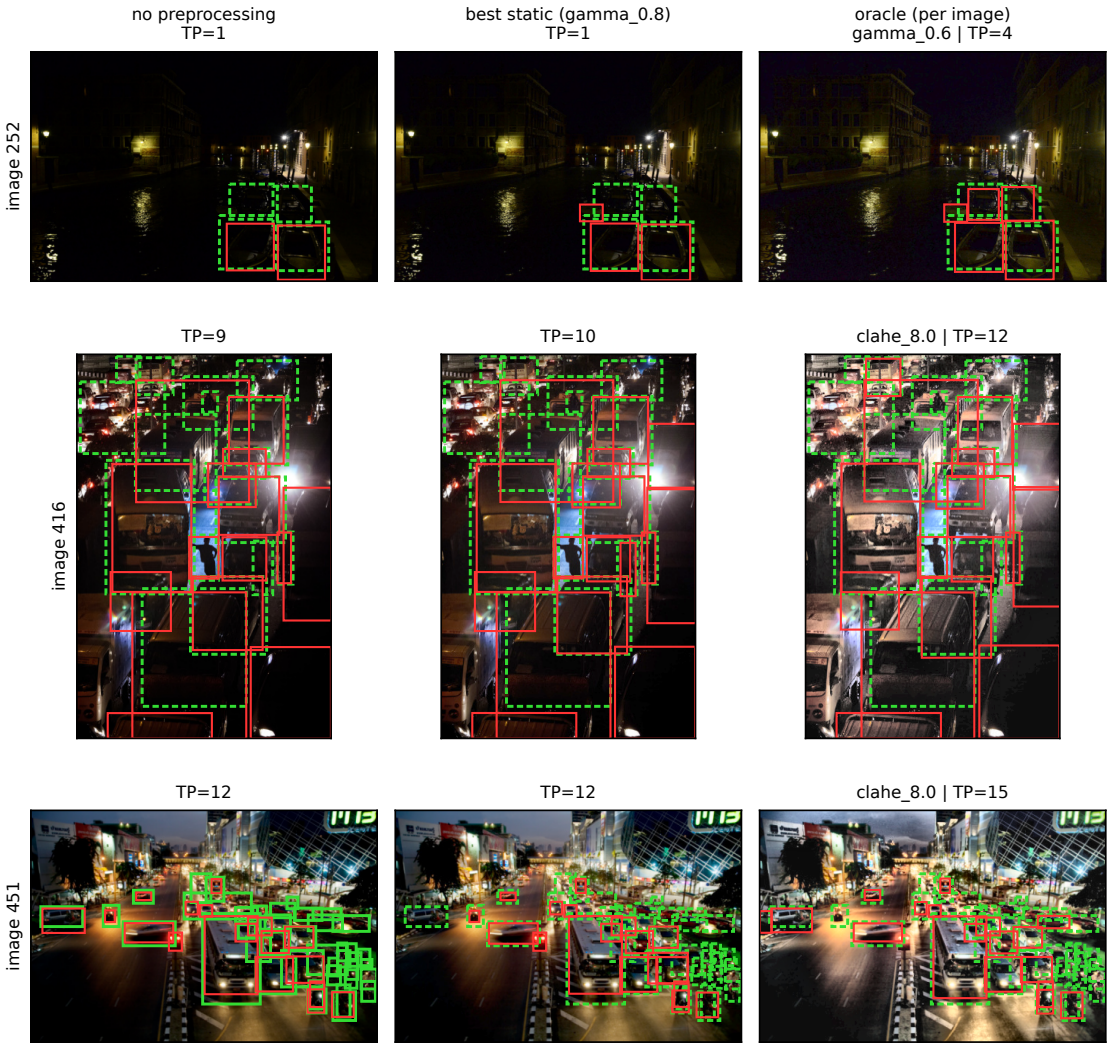


Figure S1: The three ExDark images with the largest oracle gain for YOLOv8l. Columns show the baseline (no preprocessing), the best static filter, and one TP-maximizing filter. Dashed green boxes are ground truth, solid red boxes are detections with confidence above 0.25, and true positives are counted at an IoU of 0.5. These images were chosen for their large gain and are not representative of a typical image. Images from the ExDark dataset (Loh and Chan, 2019, Computer Vision and Image Understanding 178:30–42).