Improved Handling of Motion Blur in Online Object Detection
Supplementary Material

Mohamed Sayed Gabriel Brostow
University College London
visual.cs.ucl.ac.uk/pubs/handlingMotionBlur/

1. Blur Discretization and Blur Space Segmentation

1.1. Discretization

In the main paper, P and E are held to discrete values. P_{1-3} are $[0.005, 0.001, 0.00005]$, where a lower value for P, P_3 for example, gives a more rectilinear blur. Since there are underlying random factors initialized for every blur kernel that are only influenced by P, some overlap exists between the type of blur kernels produced across different Ps. Exposures E_{1-5} are $[1/25, 1/10, 1/5, 1/2, 1]$. Note that for blur trained networks, we don't resize images to a canonical size before blurring; this acts as a mild regularizer and helps creates specialists that are flexible across a range of blur levels.

All mAP@50 scores are reported on the COCO minival set (5000 images). We use a fixed seed for every evaluation when generating blur kernels.

While our proposed model was trained with those discrete blur settings, the space of camera-induced blur is not so neatly quantized. To explore a larger cross-section of the continuous blur space, we evaluated a sweep across a random selection of exposures (horizontal axis) and blur types (vertical axis), comparing the original network against our **Specialized by Exposure Expanded Labels**. Each marker plotted in Fig. 1 is an evaluation on 2000 images from the COCO minival set. It visually summarizes that for sharp and barely-blurred images, our approach is negligibly better than the original model. But for essentially all other settings of induced motion blur, our model does measurably better.

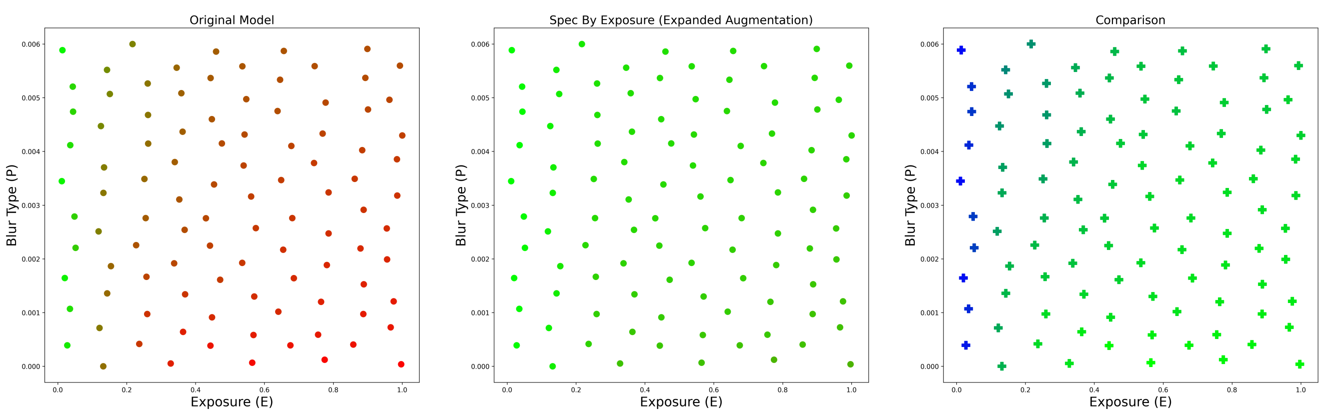

Figure 1. Comparison of the original model (ResNet50FPN trained on COCO) and our best model evaluated on expanded labels across a random selection of P and E values. Each marker is a representation of the accuracy (mAP@50) on an evaluation of 2000 images from the COCO minival. For the first two graphs (left to right), the greener the marker the closer it is to an mAP@50 of 61%. The redder it is, the closer it is to an mAP@50 of 0%. For the third graph, we visualize the difference between both networks; the greener the marker the larger the diffrence in mAP@50 between our best solution and the original network. The bluer the marker the less the difference is in accuracy. Naturally, at lower exposures, the original network holds up well, but as the exposure is ramped up, and particularly with more rectilinear blur (low P value), the difference is much larger.

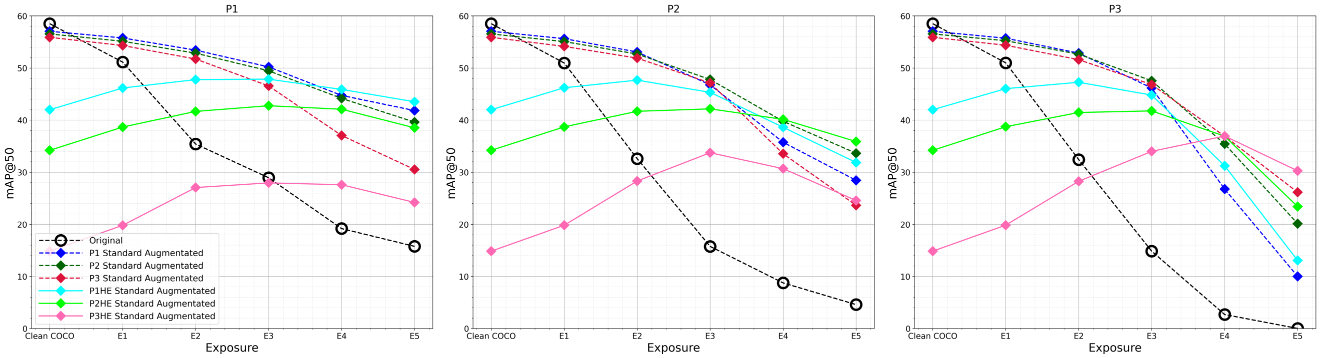

Figure 2. Standard augmented specialized networks performance across different blur types and exposures, evaluated on standard labels.

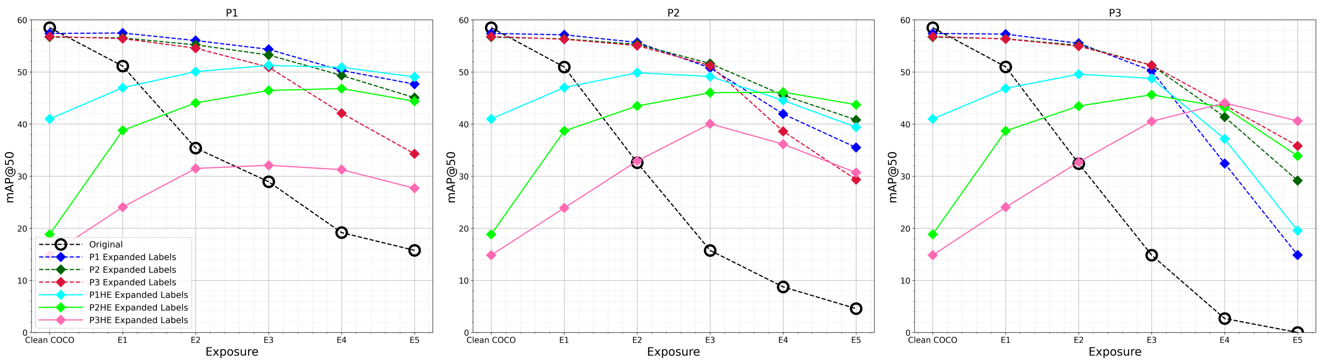

Figure 3. Expanded augmented specialized networks performance across different blur types and exposures, evaluated on expanded labels. Note how here and in Fig. 2, the each blur type specialized network tends to be better than its peers, especially at higher exposures. The high exposure HE networks outperform the rest at their respective blur type specialty at high exposures.

1.2. Segmenting Blur Space: By Type *vs.* By Exposure

All general augmented networks (non-specialized) are trained with a mixture of sharp COCO images (10%) and a random selection of blurry images across P_{1-3} and E_{1-5} (90%). **Spec by Type** networks are also trained on the same ratio, but are fixed to a specific P. The low exposure network in **Spec by Exposure** is trained on 25% sharp images and 75% blurry images from P_{1-3} and E_{1-3}; the three others are trained on 100% blurry images exclusively from a a specific P and $E_{4,5}$. The performance of these networks separately across blur levels is displayed in Fig. 2 and Fig. 3.

1.3. Blur Estimators

We use two flavors of a ResNet18 classification network, one for each type of bag of specialists. For Spec By Type, the blur estimator is trained to classify an image as belonging to one of 16 classes, clean or one of $\{P_{1-3} \times E_{1-5}\}$; it achieves 84% accuracy. For Spec by Exp, images are classified to one of four classes - clean and exposures in $\{E_{1-3}\}$, $\{P_1 \times E_{4,5}\}$, $\{P_2 \times E_{4,5}\}$, or $\{P_3 \times E_{4,5}\}$; this flavor achieves 93% accuracy.

Blur estimators are trained for 12 epochs with an initial learning rate of 0.02 (20 images) attenuated by a factor of 10 for each epoch in [3, 7, 10]. Unlike blur augmentation for detection, we resize images to a canonical size of 1333×800 before blurring.

2. Zero Centering Ablation

We show how kernel/label centering improves training and test-time accuracy. The main paper features results of evaluating on centered labels that match the barycenter of the kernel. In Fig. 4 we evaluate on non-aligned kernels and labels as well. Training and evaluating on centered kernels aligned to detection labels produces better scores, likely because the typically non-centered kernels are offset relative to the training bounding boxes. The Non-Centered model achieves the same scores when evaluated with and without centered kernels, indicating that the network has likely learned to find a vague localization and misses boxes altogether that it ought to have detected.

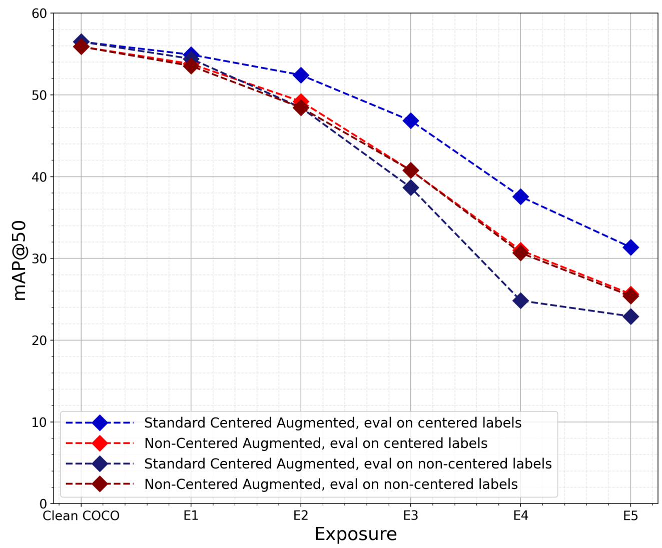

Figure 4. Comparison of different training and evaluation strategies. Results are averaged across the blur types P_{1-3}. Evaluating and training on kernels aligned to detection labels (Standard Augmentation and centered labels) scores best.

3. Expanded Labels and What the Network Outputs

Fig. 5 shows an example image with motion blur, and the output from both the **Standard Augmented** and the **Expanded Augmented** networks. The Expanded augmented network learns to predict bounding boxes that capture the superset of all spatial locations an object occupied during an exposure. This seems to be an easier objective for the network to learn. While one could argue that downstream tasks may prefer original-sized bounding boxes as shown computed by the Standard Augmented model, there is no good compromise there: the middle of the blurred object could be a "stale" image-space location compared to where the object is at the end of the exposure, in, *e.g.* a tracking-by-detection task. In qualitative examples, the expanded networks manage to detect bounding boxes that are otherwise missed by their standard counterparts, see Fig.5.

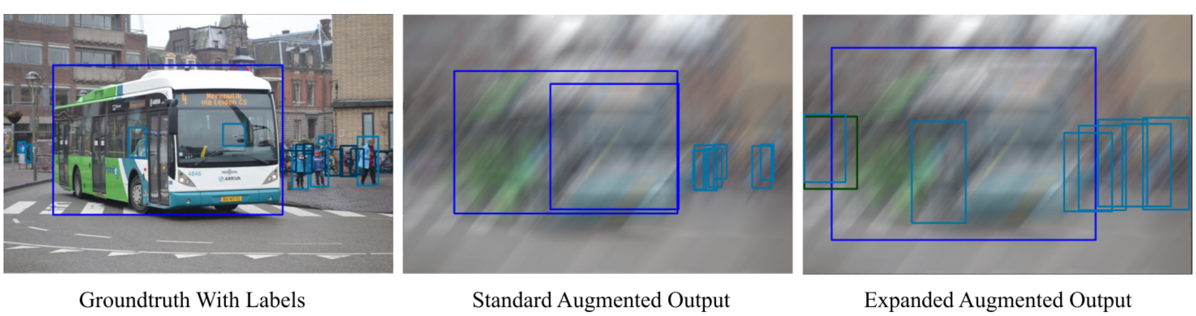

Figure 5. Left: Groundtruth image with COCO labels. Middle: Network output from the Standard Augment network. Right: Network output from the Expanded Augment network. Expanded augment networks learn to output boxes that represent the superset of all locations an object has been at during an exposure.

4. Minibatch Normalization as Schneider *et al*. [6]

In this late-breaking NeurIPS 2020 paper, results are reported for minibatch normalization on networks already trained with augmentation for blurry images. As per their algorithm, we perform minibatch normalization by finding the statistics of the activations of an input example, μ and σ, and computing a weighted sum with the training statistics using $N = 16$ and $n = 1$. This is done progressively in one forward pass. In Fig. 6, results are reported for the performance of the original model with this modification. We also experimented with finding an accurate estimate of the target distribution for blurry images by running a large portion of the train set under a specific type of blur and exposure as many times as there are batchnorm layers in the network with $n = 2048$, and using that as normalization statistics, but this did not constitute an improvement. Despite the appeal of this test-time approach, object detection was not substantially better off with it, so we excluded it from our final model.

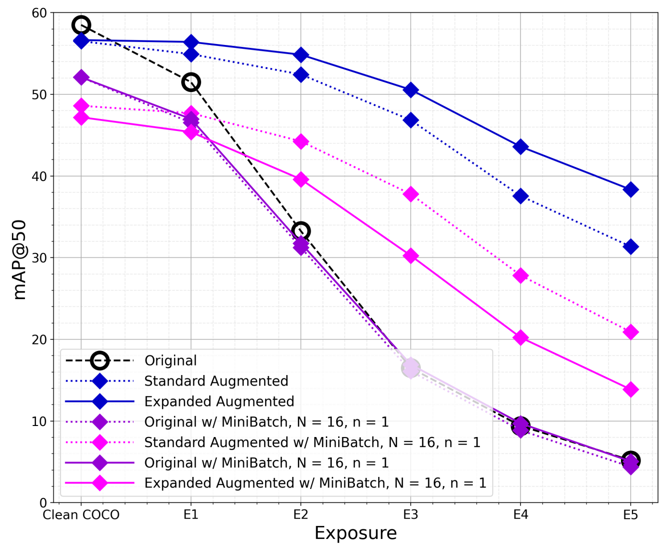

Figure 6. Comparison of using minibatch normalization on both the original network and blur augmented networks. For only this graph: solid lines are evaluation runs on expanded labels and dashed lines are evaluated on standard labels; the exception here is the original model which is evaluated on expanded labels. Results are averaged across P_{1-3}.

5. Defocus And Motion Blur

Our models are more resilient to camera defocus blur than the original. P_1 is close to simulating defocus blur since the camera trajectory loops in place. We Gaussian blur each motion-blur kernel with a random σ across all blur types and exposures. We report results in Fig. 7.

6. Real-World Blur Datasets

We evaluate our models on on two pseudo-real blur datasets, GOPRO [4] and REDS [3], and a real-world blur dataset, RealBlur [**?**], obtained using shutter tied cameras. These datasets don't have box annotations, so we utilize a state-of-the-art high accuracy detector, DetectoRS [5], to obtain pseudo-groundtruth bounding-boxes for evaluation. For evaluating expanded bounding boxes, we generate our own GOPRO testset using grountruth sharp frames and use flow computed using [7] for bounding-box expansion.

We stick to the canonical train and test sets when available. However when either the train and sets combined don't contain enough images for reliable evaluation or when we need to estimate flow on source high-frame rate images, we synthesize our own set of blurry images using sharp images from the datasets.

We perform standard evaluation on RealBlur, GOPRO, and REDS. RealBlur sharp and blurry frames are naturally aligned during capture and the alignment is further refined in the post process described in the paper [**?**]. GOPRO and REDS report a sharp frame as one in the middle of the window to synthesize a blurry frame. Although this doesn't necessarily equate to centering the blur kernel since movement can be asymmetric on either side of the sharp frame, we use this as our "standard" evaluation as it's the closest approximation given the data.

For RealBlur, we use both train and test sets (4,738 pairs) and set to the confidence threshold on pseudo-groundtruth boxes to 0.6. For REDS, we sample 5,000 frames from the train and validation, set the confidence threshold to 0.4, and allow only a

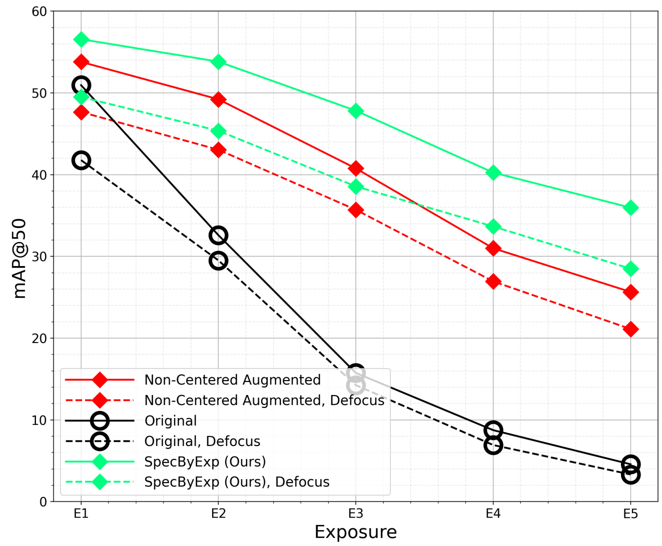

Figure 7. Models evaluated on motion blur with and without defocusing. E is motion blur extent; each point averages blur types P_{1-3}. Defocus is simulated by Gaussian blurring each motion-blur kernel with a random σ. Defocus hurts, but our model still performs well, especially compared to no-centering and the original network.

maximum of 20 images without bounding boxes. For GOPRO, we use the combination of the train and test sets (3214 pairs) and set the confidence threshold to 0.4.

For expanded evaluation, we synthesize 5,442 blurry GOPRO frames using the method and window size limits outlined in [4], and we set the confidence threshold on pseudo-groundtruth bounding boxes to 0.6. We omit empty scenes with no COCO object classes, namely `GOPR0374_11_00`, `GOPR0374_11_01`, `GOPR0374_11_02`, `GOPR0374_11_03`, `GOPR0857_11_00`, `GOPR0868_11_00`, `GOPR0868_11_02`, `GOPR0871_11_00`, and `GOPR0396_11_00`. During this process, we note the sharp frames used to synthesize blurry frames and obtain low-resolution flow fields using RAFT [7] to estimate where objects have moved during the exposure. We use low-resolution as apposed to the refined maps to avoid artefacts at object boundaries. We advect each bounding box corner using the estimated flow fields both forwards and backwards on either side of the sharp frame, stopping at the assigned blur window size. We then assign the bounding box corners to the super-set of both the original points and the advected points. These new boxes are estimates of the super-set location of where an object has been in an exposure, and are used during expanded evaluation.

7. Qualitative Results

You can find a video with qualitative results and a visual explanation of our method `visual.cs.ucl.ac.uk/pubs/handlingMotionBlur/`.

There, we show real world examples where our model, based on the two proposed remedies, manages to detect objects in many places where the original model fails, especially when the ratio of camera motion to object size is high. Following on from the quantitative experiments in the paper and here, we synthesize blurry COCO images (in the same spirit as [2, 1]) and show sample results in the video.

8. Results Tables

Table 1 and Table 2 contain the raw results used to generate Fig. 5 and Fig. 6 in the paper. Table 3 and Table 4 show raw numbers for generating Fig. 2 and Fig. 3.

Variant	Clean	E_1	E_2	E_3	E_4	E_5
Original	58.50	50.95	32.59	15.75	8.75	4.58
Deblur then Original	55.50	49.18	42.13	30.31	12.72	6.26
Deblur then Standard Augmented	53.90	51.47	48.44	40.35	23.85	15.86
Squint	55.65	54.30	51.76	46.21	37.24	31.39
AugMix (Non Expanded)	**59.34**	53.13	38.07	20.70	13.63	8.21
AugMix PixelLevel	58.93	51.68	32.10	14.84	9.12	4.48
Original w/ MiniBatch, N = 16, n = 1	52.10	46.53	31.25	16.10	8.86	4.40
Standard Augmented w/ MiniBatch, N = 16, n = 1	48.60	47.70	44.25	37.79	27.84	20.92
Non-Centered Augmented	55.91	53.80	49.22	40.77	31.00	25.66
Standard Augmented w/ NonSpatial Augmix	55.77	54.15	51.95	46.53	38.41	31.67
Standard Augmented	56.51	54.93	52.44	46.85	37.56	31.37
Spec By Type	56.50	55.39	52.33	**47.78**	**39.81**	33.84
Spec By Exposure (Ours)	58.55	**56.57**	**53.83**	47.74	40.21	**35.93**

Variant	Clean	E_1	E_2	E_3	E_4	E_5
Original	58.50	50.95	32.59	15.75	8.75	4.58
Deblur then Original	55.50	49.18	42.13	30.31	12.72	6.26
Deblur then Standard Augmented	53.90	51.47	48.44	40.35	23.85	15.86
Squint	55.65	54.30	51.76	46.21	37.24	31.39
AugMix (Non Expanded)	**59.34**	53.13	38.07	20.70	13.63	8.21
AugMix PixelLevel	58.93	51.68	32.10	14.84	9.12	4.48
Original w/ MiniBatch, N = 16, n = 1	52.10	46.53	31.25	16.10	8.86	4.40
Standard Augmented w/ MiniBatch, N = 16, n = 1	48.60	47.70	44.25	37.79	27.84	20.92
Non-Centered Augmented	55.91	53.80	49.22	40.77	31.00	25.66
Standard Augmented w/ NonSpatial Augmix	55.77	54.15	51.95	46.53	38.41	31.67
Standard Augmented	56.51	54.93	52.44	46.85	37.56	31.37
Spec By Type	56.50	55.39	52.33	**47.78**	39.81	33.84
Spec By Exposure (Ours)	58.55	**56.57**	**53.83**	47.74	**40.21**	**35.93**

Variant	Clean	E_1	E_2	E_3	E_4	E_5
Original	58.50	50.95	32.59	15.75	8.75	4.58
Deblur then Original	55.50	49.18	42.13	30.31	12.72	6.26
Deblur then Standard Augmented	53.90	51.47	48.44	40.35	23.85	15.86
Squint	55.65	54.30	51.76	46.21	37.24	31.39
AugMix (Non Expanded)	**59.34**	53.13	38.07	20.70	13.63	8.21
AugMix PixelLevel	58.93	51.68	32.10	14.84	9.12	4.48
Original w/ MiniBatch, N = 16, n = 1	52.10	46.53	31.25	16.10	8.86	4.40
Standard Augmented w/ MiniBatch, N = 16, n = 1	48.60	47.70	44.25	37.79	27.84	20.92
Non-Centered Augmented	55.91	53.80	49.22	40.77	31.00	25.66
Standard Augmented w/ NonSpatial Augmix	55.77	54.15	51.95	46.53	38.41	31.67
Standard Augmented	56.51	54.93	52.44	46.85	37.56	31.37
Spec By Type	56.50	55.39	52.33	47.78	39.81	33.84
Spec By Exposure (Ours)	58.55	**56.57**	**53.83**	**47.74**	**40.21**	**35.93**

Table 1. Raw numbers from Fig. 5 in the paper. Non-expanded labels used during evaluation. Results are on the COCO minival set under different blur parameters and exposure. From top to bottom, the blur type changes from P_1 to P_2 to P_3. Networks trained with blur augmentation would be trained on non-expanded labels.

Variant	Clean	E_1	E_2	E_3	E_4	E_5
Original	58.50	51.50	33.26	16.49	9.41	5.20
Deblur then Original	55.50	49.50	41.07	28.32	12.19	6.24
Squint Expanded Labels	56.15	56.25	54.53	50.09	42.92	37.66
AugMix Expanded Labels	51.80	46.62	34.21	18.99	11.32	6.15
AugMix PixelLevel	**58.93**	51.50	33.26	16.49	9.41	0.05
Original w/ MiniBatch, N = 16, n = 1	52.10	46.99	31.77	16.92	9.71	5.07
Expanded Labels w/ MiniBatch, N = 16, n = 1	47.20	45.39	39.61	30.26	20.23	13.88
Expanded Labels	56.65	56.42	54.86	50.57	43.60	38.35
Expanded Labels w/ NonSpatial Augmix	56.33	55.99	54.54	50.32	43.10	37.85
Spec By Type Expanded Labels	56.70	<u>56.75</u>	<u>55.23</u>	**51.59**	<u>45.55</u>	<u>40.81</u>
Spec By Exposure Expanded Labels (Our Best)	<u>58.62</u>	**58.01**	**56.40**	<u>50.97</u>	**46.37**	**43.78**

Variant	Clean	E_1	E_2	E_3	E_4	E_5
Original	58.50	51.50	33.26	16.49	9.41	5.20
Deblur then Original	55.50	49.50	41.07	28.32	12.19	6.24
Squint Expanded Labels	56.15	56.25	54.53	50.09	42.92	37.66
AugMix Expanded Labels	51.80	46.62	34.21	18.99	11.32	6.15
AugMix PixelLevel	**58.93**	51.50	33.26	16.49	9.41	0.05
Original w/ MiniBatch, N = 16, n = 1	52.10	46.99	31.77	16.92	9.71	5.07
Expanded Labels w/ MiniBatch, N = 16, n = 1	47.20	45.39	39.61	30.26	20.23	13.88
Expanded Labels	56.65	56.42	54.86	50.57	43.60	38.35
Expanded Labels w/ NonSpatial Augmix	56.33	55.99	54.54	50.32	43.10	37.85
Spec By Type Expanded Labels	56.70	<u>56.75</u>	<u>55.23</u>	**51.59**	<u>45.55</u>	<u>40.81</u>
Spec By Exposure Expanded Labels (Our Best)	<u>58.62</u>	**58.01**	**56.40**	<u>50.97</u>	**46.37**	**43.78**

Variant	Clean	E_1	E_2	E_3	E_4	E_5
Original	58.50	51.50	33.26	16.49	9.41	5.20
Deblur then Original	55.50	49.50	41.07	28.32	12.19	6.24
Squint Expanded Labels	56.15	56.25	54.53	50.09	42.92	37.66
AugMix Expanded Labels	51.80	46.62	34.21	18.99	11.32	6.15
AugMix PixelLevel	**58.93**	51.50	33.26	16.49	9.41	0.05
Original w/ MiniBatch, N = 16, n = 1	52.10	46.99	31.77	16.92	9.71	5.07
Expanded Labels w/ MiniBatch, N = 16, n = 1	47.20	45.39	39.61	30.26	20.23	13.88
Expanded Labels	56.65	56.42	54.86	50.57	43.60	38.35
Expanded Labels w/ NonSpatial Augmix	56.33	55.99	54.54	50.32	43.10	37.85
Spec By Type Expanded Labels	56.70	<u>56.75</u>	<u>55.23</u>	**51.59**	<u>45.55</u>	<u>40.81</u>
Spec By Exposure Expanded Labels (Our Best)	<u>58.62</u>	**58.01**	**56.40**	<u>50.97</u>	**46.37**	**43.78**

Table 2. Raw numbers from Fig. 6 in the paper. Expanded labels used during evaluation. Results are on the COCO minival set under different blur parameters and exposure. From top to bottom, the blur type changes from P_1 to P_2 to P_3.

Variant	Clean	E_1	E_2	E_3	E_4	E_5
Original	58.50	50.95	32.59	15.75	8.75	4.58
Low-Exposure Augmented	**58.55**	**56.57**	**53.83**	**47.83**	29.32	18.64
P1 Standard Augmentated	57.04	55.62	53.06	46.76	35.74	28.43
P2 Standard Augmentated	56.53	55.06	52.67	47.78	39.76	33.63
P3 Standard Augmentated	55.88	54.14	51.89	47.12	33.53	23.64
P1HE Standard Augmentated	41.98	46.20	47.66	45.32	**38.62**	**31.85**
P2HE Standard Augmentated	34.19	38.69	41.68	42.15	40.13	35.88
P3HE Standard Augmentated	14.84	19.81	28.30	33.73	30.67	24.56

Variant	Clean	E_1	E_2	E_3	E_4	E_5
Original	58.50	50.95	32.59	15.75	8.75	4.58
Low-Exposure Augmented	**58.55**	**56.57**	**53.83**	**47.83**	29.32	18.64
P1 Standard Augmentated	57.04	55.62	53.06	46.76	35.74	28.43
P2 Standard Augmentated	56.53	55.06	52.67	47.78	39.76	33.63
P3 Standard Augmentated	55.88	54.14	51.89	47.12	33.53	23.64
P1HE Standard Augmentated	41.98	46.20	47.66	45.32	38.62	31.85
P2HE Standard Augmentated	34.19	38.69	41.68	42.15	**40.13**	**35.88**
P3HE Standard Augmentated	14.84	19.81	28.30	33.73	30.67	24.56

Variant	Clean	E_1	E_2	E_3	E_4	E_5
Original	58.50	50.95	32.59	15.75	8.75	4.58
Low-Exposure Augmented	**58.55**	**56.57**	**53.83**	**47.83**	29.32	18.64
P1 Standard Augmentated	57.04	55.62	53.06	46.76	35.74	28.43
P2 Standard Augmentated	56.53	55.06	52.67	47.78	39.76	33.63
P3 Standard Augmentated	55.88	54.14	51.89	47.12	33.53	23.64
P1HE Standard Augmentated	41.98	46.20	47.66	45.32	38.62	31.85
P2HE Standard Augmentated	34.19	38.69	41.68	42.15	40.13	35.88
P3HE Standard Augmentated	14.84	19.81	28.30	33.73	**30.67**	**24.56**

Table 3. Raw numbers for standard augmented specialists performance (Fig. 2). Results are on the COCO minival set under different blur parameters and exposure. From top to bottom, the blur type changes from P_1 to P_2 to P_3. Networks are trained and evaluated on non-expanded "standard" labels under blur augmentation, with the exception of Original.

Variant	Clean	E_1	E_2	E_3	E_4	E_5
Original	$\underline{58.50}$	51.50	33.26	16.49	9.41	5.20
Low-Exposure Expanded Labels	**58.62**	**58.06**	**56.38**	50.97	33.95	22.40
P1 Expanded Labels	57.39	$\underline{57.13}$	$\underline{55.67}$	**50.78**	$\underline{41.93}$	$\underline{35.50}$
P2 Expanded Labels	56.68	56.34	55.30	$\underline{51.62}$	45.54	40.80
P3 Expanded Labels	56.80	56.28	55.06	51.22	38.61	29.34
P1HE Expanded Labels	40.99	47.00	49.85	49.13	**44.55**	**39.40**
P2HE Expanded Labels	18.84	38.67	43.47	46.02	46.12	43.72
P3HE Expanded Labels	14.84	23.90	32.92	40.05	36.14	30.67

Variant	Clean	E_1	E_2	E_3	E_4	E_5
Original	$\underline{58.50}$	51.50	33.26	16.49	9.41	5.20
Low-Exposure Expanded Labels	**58.62**	**58.06**	**56.38**	50.97	33.95	22.40
P1 Expanded Labels	57.39	$\underline{57.13}$	$\underline{55.67}$	50.78	41.93	35.50
P2 Expanded Labels	56.68	56.34	55.30	**51.62**	$\underline{45.54}$	$\underline{40.80}$
P3 Expanded Labels	56.80	56.28	55.06	$\underline{51.22}$	38.61	29.34
P1HE Expanded Labels	40.99	47.00	49.85	49.13	44.55	39.40
P2HE Expanded Labels	18.84	38.67	43.47	46.02	**46.12**	**43.72**
P3HE Expanded Labels	14.84	23.90	32.92	40.05	36.14	30.67

Variant	Clean	E_1	E_2	E_3	E_4	E_5
Original	$\underline{58.50}$	51.50	33.26	16.49	9.41	5.20
Low-Exposure Expanded Labels	**58.62**	**58.06**	**56.38**	50.97	33.95	22.40
P1 Expanded Labels	57.39	$\underline{57.13}$	$\underline{55.67}$	50.78	41.93	35.50
P2 Expanded Labels	56.68	56.34	55.30	**51.62**	45.54	40.80
P3 Expanded Labels	56.80	56.28	55.06	$\underline{51.22}$	$\underline{38.61}$	$\underline{29.34}$
P1HE Expanded Labels	40.99	47.00	49.85	49.13	44.55	39.40
P2HE Expanded Labels	18.84	38.67	43.47	46.02	46.12	43.72
P3HE Expanded Labels	14.84	23.90	32.92	40.05	**36.14**	**30.67**

Table 4. Raw numbers for expanded augmented specialists performance (Fig. 3). Results are on the COCO minival set under different blur parameters and exposure. From top to bottom, the blur type changes from P_1 to P_2 to P_3. Networks are trained and evaluated on expanded labels under blur augmentation, with the exception of Original.

References

[1] Dan Hendrycks and Thomas Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. *arXiv preprint arXiv:1903.12261*, 2019. 5

[2] Claudio Michaelis, Benjamin Mitzkus, Robert Geirhos, Evgenia Rusak, Oliver Bringmann, Alexander S Ecker, Matthias Bethge, and Wieland Brendel. Benchmarking robustness in object detection: Autonomous driving when winter is coming. *arXiv preprint arXiv:1907.07484*, 2019. 5

[3] Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019. 4

[4] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3883–3891, 2017. 4, 5

[5] Siyuan Qiao, Liang-Chieh Chen, and Alan Yuille. Detectors: Detecting objects with recursive feature pyramid and switchable atrous convolution. *arXiv preprint arXiv:2006.02334*, 2020. 4

[6] Steffen Schneider, Evgenia Rusak, Luisa Eck, Oliver Bringmann, Wieland Brendel, and Matthias Bethge. Improving robustness against common corruptions by covariate shift adaptation. *Advances in Neural Information Processing Systems*, 33, 2020. 3

[7] Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow. In *European Conference on Computer Vision*, pages 402–419. Springer, 2020. 4, 5