3D Common Corruptions and Data Augmentation
Supplementary Material

1. Overview

We provide further details and evaluations in the supplementary material, as outlined below.

1. **Overview video** providing a short summary of the paper and qualitative results (on the project page)

2. Quantitative evaluations:

   - Performance of robustness mechanisms on 3D Common Corruptions (3DCC) and Common Corruptions [6] (2DCC) (Sec. 2.1)
   - Analyzing the robustness of panoptic segmentation models against occlusion changes in 3DCC (Sec. 2.2)
   - Full affinity matrices for 3DCC and 2DCC (Sec. 2.3)
   - Additional analysis on the effectiveness of using predicted depth to generate 3DCC (Sec. 2.4)
   - Comparing robust ImageNet models on 2DCC and 3DCC benchmarks (Sec. 2.5)
   - Performance of 3D data augmentation (Sec. 2.6)

3. Qualitative results of 3D data augmentation:

   - **Video evaluations** performed by applying the proposed method and the baselines frame-by-frame to several videos (Sec. 3.1)
   - Additional qualitative evaluations (Sec. 3.2)

4. Further method details for data augmentation mechanisms (Sec. 4.1)

5. Further implementation details for corruptions (Sec. 4.2)

6. Visualizations of 3DCC and 2DCC for different shift intensities (Sec. 5)

7. Our development **code** with documentation (on the project page)

2. Quantitative Results

2.1. Robustness mechanisms against 3DCC and 2DCC

Figure 6 shows ℓ_1 errors of robustness mechanisms against individual corruptions in 3DCC for surface normals estimation. Figure 7 shows the same result for depth estimation. For both tasks, 3DCC leads to significantly degraded predictions for models trained with robustness mechanisms. For completeness, we also provide performances of these models against corruptions in 2DCC, in Figures 8 and 9.

2.2. Robustness of panoptic segmentation models against occlusion changes

We evaluate the robustness of two panoptic segmentation models from [3] against occlusion changes in 3DCC. The models are trained on Omnidata [3] and Taskonomy [15] datasets with a Detectron [12] backbone. The Detectron was initialised with an ImageNet pre-trained ResNet50 backbone with the first two layers frozen during training.

Figure 1 shows the effect of occlusion on the predictions of models, i.e. how the models' intersection over union (IoU) scores change with increasing occlusion, for selected objects. This is computed on the test scenes from Replica [11]. We defined the occlusion ratio as the number of occluded pixels of the occluded object divided by the summation of occluded and visible pixels of the object. As expected, we see a decrease in IoU as occlusion increases.

2.3. Redundancy between 3DCC and 2DCC

We provide in Figures 10 and 11 the full affinity matrices between 2DCC and 3DCC by computing the correlations of ℓ_1 errors made in the surface normals and depth estimation tasks, respectively. Figure 12 shows the same result by computing ℓ_1 errors in the RGB domain. As can be seen, 3DCC yields lower correlations both intra-benchmark and against 2DCC.

2.4. Effectiveness of predicted depth to generate 3DCC

We compare the effectiveness of using MiDaS [10] depth estimation model to generate 3DCC against two control baselines in Fig 5. The first one is *incorrect instance depth*

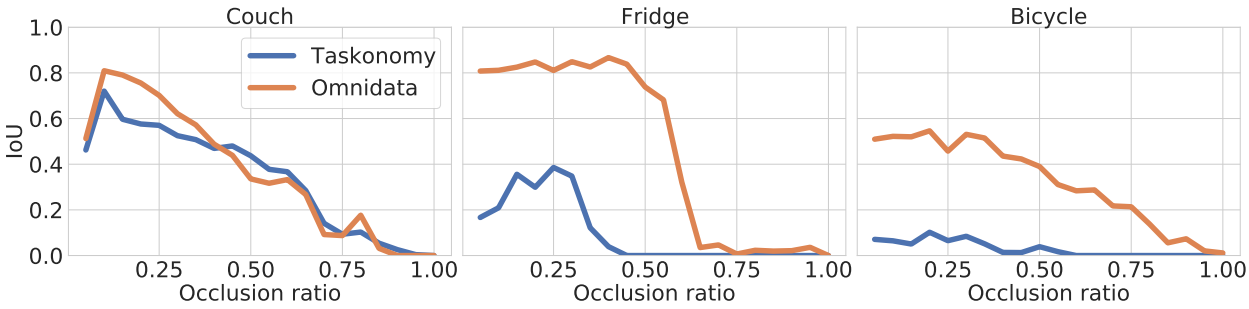

Figure 1. **Robustness against occlusion changes in 3DCC.** The plot shows the intersection over union (IoU) scores of Detectron models [12] trained on Taskonomy [15] or Omnidata [3], for different objects over a range of occlusion ratios. The occlusion ratio is defined as the number of occluded pixels of the occluded object divided by the summation of occluded and visible pixels of the object. This is computed over the test scenes of Replica [11]. The plots show that the models predictions degrade with higher occlusion levels, and that the Omnidata trained model is generally more robust than the Taskonomy ones.

where we randomly swap depth predictions for a given RGB image with another depth prediction. The second one is *blind guess depth* which minimizes the expected likelihood loss in the training dataset, hence it is a statistically informed guess reflecting the dataset regularities [13, 14] (See Fig. 4 for its visualization). As shown in Fig. 5, the predicted depth yields higher correlation with the ground truth compared to control baselines, showing that it can be used to generate 3DCC for datasets without 3D information.

2.5. Comparing robust ImageNet models on 2DCC and 3DCC benchmarks

We compare performances of the robust ImageNet models from RobustBench and ImageNet-C leaderboards below (See the links for full model names) in Figure 2. A quick look at the scatter plot (top) shows that the general trends between 2DCC and 3DCC are similar, an observation also made in [9] even when the corruptions are *designed to be dissimilar to 2DCC*. Hence, this is also expected for our case as 2D and 3D corruptions are not completely disjoint (expected). But, we observe notable differences between the two benchmarks in local trends, e.g. in the ellipsoid regions certain robustness mechanisms improved performance on 2DCC while being ineffective against 3DCC.

The bar plot (bottom) comparing a subset of corruptions that exists in both benchmarks (e.g. 2D defocus blur vs its 3D version), further reflects the differences where **1.** all models have *consistently higher* errors on 3D corruptions compared to their 2DCC counterparts and **2.** certain models, e.g. AugMix and AugMix+DA, face a larger drop in performance on 3DCC compared to the other models, indicating that AugMix may be biased towards 2DCC. Thus, 3DCC evaluations can be informative during model development as they **expose nonlinear trends and vulnerabilities that are not captured by 2DCC** (also discussed in Sec. 5.2.2).

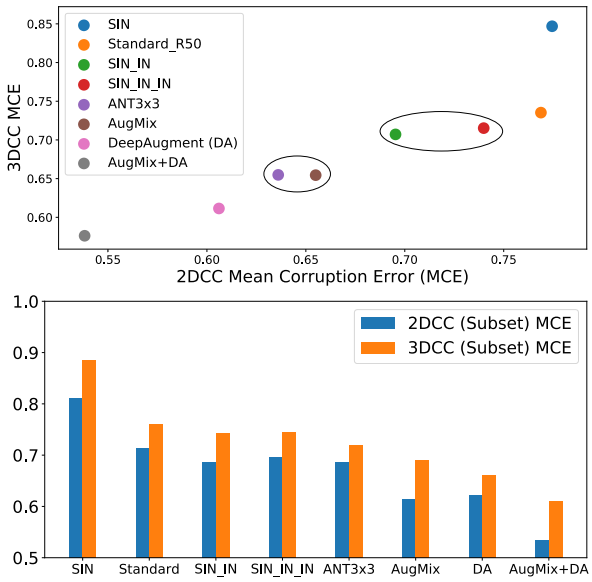

Figure 2. Comparing ImageNet models on 2DCC and 3DCC benchmarks (i.e. ImageNet-C vs ImageNet-3DCC). **Top:** Comparison of mean corruption errors (MCEs) on 2DCC and 3DCC. **Bottom:** Comparison of MCEs for a subset of corruptions that exists in both benchmarks (e.g. 2D defocus blur vs its 3D version). See Sec. 2.5 for details.

2.6. Performance of 3D data augmentation

We show in Fig. 13 the performance of 3D data augmentation and baselines on individual corruptions from 3DCC for surface normals estimation task. Similarly, Fig. 14 shows the performance on 2DCC and Table 1 provides full performance metrics on OASIS [2] benchmark. As can be seen from the results, 3D data augmentation notably boosts robustness.

3. Qualitative Results

3.1. Video evaluations

We perform evaluations on clips from manually collected DSLR data, YouTube videos, Adobe After Affects (AE) generated corrupted data (Sec.5.2.3 in the main paper), and sample queries from OASIS [2]. They suggest the proposed 3D data augmentation yields notably sharper and more accurate predictions with less flickering, compared to baselines. We recommend watching the clips on the project page.

3.2. Additional queries

In addition to video evaluations in Sec. 3.1 we provided additional results on OASIS and AE datasets in Figures 15 and 16, again suggesting 3D data augmentation is beneficial for improving robustness.

4. Further method details

4.1. Data augmentation mechanisms

Below we provide additional details about the training procedures of data augmentation models. All models were finetuned from the Baseline UNet (T+UNet) with an equal number of clean and augmented images.

Adversarial training: The adversarial examples are generated from an I-FSGM attack with $\epsilon = (0 - 16]$. To generate the I-FGSM attack [8], we apply the following:

$$X_0 = X, \tag{1}$$

$$X_{n+1} = Clip_{X,\epsilon}\{X_n + \alpha \text{sign}(\nabla J(X_n, y))\} \tag{2}$$

where J is the loss function. Similar to [8], we set $\alpha = 1$ in our experiments and the number of iterations given by $N = \min(4 + \epsilon, 1.25\epsilon)$.

Style [4]: We applied the AdaIN style transfer [7]. The stylization coefficient is randomly selected from the range $[0.1, 0.5]$.

DeepAugment [5]: We use the same perturbations as [5], with the exception of the ones that change the scene geometry, e.g. rotation, flipping.

4.2. Implementing corruptions

We release the full open source code of our pipeline, which enables using the implemented corruptions on any dataset. Below, we provide further details about implementing corruptions.

Depth of field: We divide the scene into two regions using hyperfocal distance [1] which is the focus distance yielding the largest depth of field. We then define *near focus* region as parts of the scene closer to camera than hyperfocal distance (and vice versa for far focus). After picking the focus region, we perform the blurring (See Fig. 3 right of main paper for an illustration.)

	Angular error°		% within t°			Relative Normal	
	Mean	Median	11.25°	22.5°	30°	AUCo	AUCp
T+UNet	30.49	22.93	23.18	49.24	61.12	0.6095	0.5953
T+DPT	32.13	25.68	18.82	44.06	57.13	0.6078	0.5484
OASIS	24.63	19.06	30.10	57.34	69.91	0.5693	0.5490
O+DPT	24.42	18.46	28.82	59.53	72.39	0.6320	0.5484
O+DPT+2DCC	23.67	17.75	30.24	61.22	73.83	0.6287	0.6806
O+DPT+2DCC+3D **(Ours)**	24.65	18.53	28.89	58.97	71.89	0.6251	0.6796
Ours (+X-TC [14])	23.89	18.34	28.66	60.00	73.29	0.6264	0.6928

Table 1. **Evaluations on OASIS.** Similar to Table 1 in the main paper, but results for more metrics are shown.

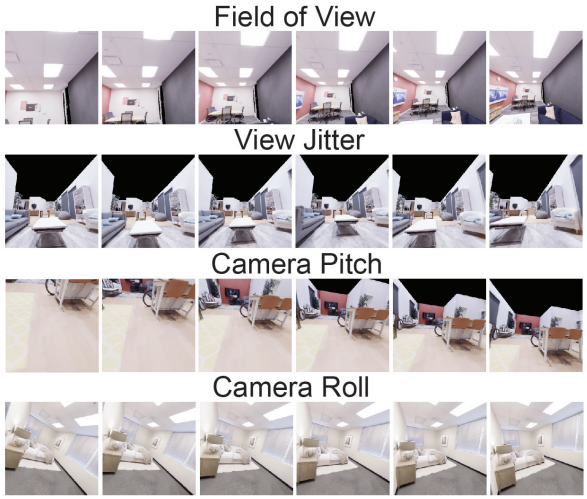

Figure 3. Visualizations of view change corruptions from 3DCC for different sampled angles.

Video: We use *ffmpeg* scripts to generate video-based corruptions.

Semantics: We use Replica [11] dataset which comes with 3D mesh annotations. To obtain occlusion masks, while it is possible to perform ray-tracing, it could be expensive and time-consuming. Thus, we also investigated an alternative approach and modified the mesh by removing all the objects except the target one and rendering semantic masks. Performing a second rendering when all objects are in the mesh, i.e. original state, yields the semantic masks with occlusions (e.g. blue masks in Fig 1 and Fig. 4 in the main paper). The difference between the two masks correspond to occlusion mask (e.g. red masks in Fig. 1 and Fig. 4 in the main paper).

5. Visualizing Corruptions

We show visualizations of corruptions from 3DCC and 2DCC for 5 shift intensities in Figures 17 and 18, respectively. Furthermore, we also show samples from *view changes* corruptions from 3DCC in Fig. 3.

References

[1] Fergus W Campbell. The depth of field of the human eye. *Optica Acta: International Journal of Optics*, 4(4):157–164,

Figure 4. Statistically informed blind guess depth prediction of Taskonomy dataset [13, 14].

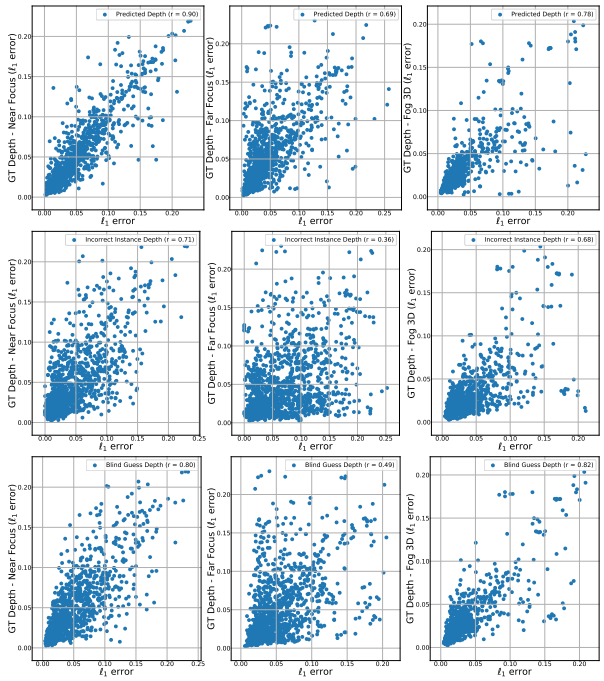

Figure 5. **Effectiveness of applying 3DCC without ground truth depth.** Similar to Fig. 12 in the main paper, but control baselines are also provided, namely *incorrect instance depth* and *blind guess depth*. The predicted depth from MiDaS [10] model yields the strongest correlations with the ground truth depth.

1957. 3

[2] Weifeng Chen, Shengyi Qian, David Fan, Noriyuki Kojima, Max Hamilton, and Jia Deng. Oasis: A large-scale dataset for single image 3d in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 679–688, 2020. 2, 3, 12

[3] Ainaz Eftekhar, Alexander Sax, Jitendra Malik, and Amir Zamir. Omnidata: A scalable pipeline for making multi-task mid-level vision datasets from 3d scans. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10786–10796, 2021. 1, 2

[4] Robert Geirhos, Patricia Rubisch, Claudio Michaelis, Matthias Bethge, Felix A Wichmann, and Wieland Brendel. Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv preprint arXiv:1811.12231*, 2018. 3

[5] Dan Hendrycks, Steven Basart, Norman Mu, Saurav Kadavath, Frank Wang, Evan Dorundo, Rahul Desai, Tyler Zhu, Samyak Parajuli, Mike Guo, et al. The many faces of robustness: A critical analysis of out-of-distribution generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8340–8349, 2021. 3

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

[7] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1501–1510, 2017. 3

[8] Alexey Kurakin, Ian Goodfellow, and Samy Bengio. Adversarial machine learning at scale. *arXiv preprint arXiv:1611.01236*, 2016. 3

[9] Eric Mintun, Alexander Kirillov, and Saining Xie. On interaction between augmentations and corruptions in natural corruption robustness. *arXiv preprint arXiv:2102.11273*, 2021. 2

[10] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *arXiv preprint arXiv:1907.01341*, 2019. 1, 4

[11] Julian Straub, Thomas Whelan, Lingni Ma, Yufan Chen, Erik Wijmans, Simon Green, Jakob J. Engel, Raul Mur-Artal, Carl Ren, Shobhit Verma, Anton Clarkson, Mingfei Yan, Brian Budge, Yajie Yan, Xiaqing Pan, June Yon, Yuyang Zou, Kimberly Leon, Nigel Carter, Jesus Briales, Tyler Gillingham, Elias Mueggler, Luis Pesqueira, Manolis Savva, Dhruv Batra, Hauke M. Strasdat, Renzo De Nardi, Michael Goesele, Steven Lovegrove, and Richard Newcombe. The Replica dataset: A digital replica of indoor spaces. *arXiv preprint arXiv:1906.05797*, 2019. 1, 2, 3

[12] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. https://github.com/facebookresearch/detectron2, 2019. 1, 2

[13] Teresa Yeo, Oğuzhan Fatih Kar, and Amir Zamir. Robustness via cross-domain ensembles. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12189–12199, October 2021. 2, 4

[14] Amir Zamir, Alexander Sax, Teresa Yeo, Oğuzhan Kar, Nikhil Cheerla, Rohan Suri, Zhangjie Cao, Jitendra Malik, and Leonidas Guibas. Robust learning through cross-task consistency. *arXiv preprint arXiv:2006.04096*, 2020. 2, 3, 4

[15] Amir R Zamir, Alexander Sax, William Shen, Leonidas J Guibas, Jitendra Malik, and Silvio Savarese. Taskonomy: Disentangling task transfer learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3712–3722, 2018. 1, 2

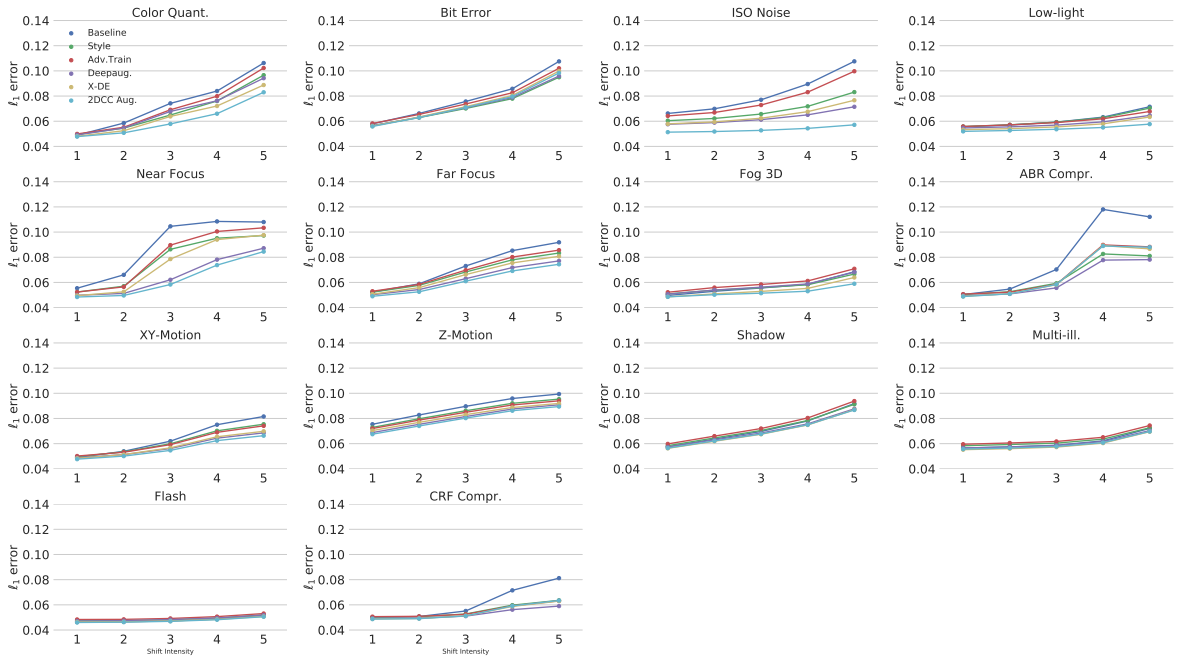

Figure 6. **Average ℓ_1 losses of robustness mechanisms against corruptions in 3DCC for surface normals estimation.** Similar to Fig. 6 in the main paper, but demonstrates the performance for individual corruptions in 3DCC.

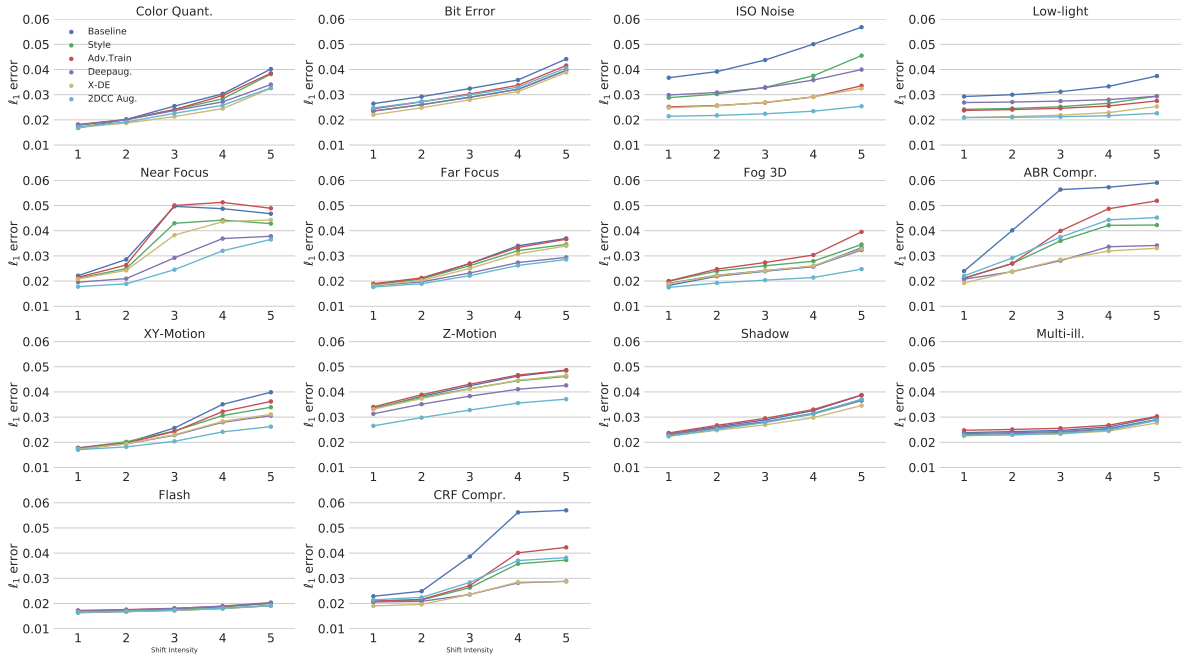

Figure 7. **Average ℓ_1 losses of robustness mechanisms against corruptions in 3DCC for depth estimation.** Similar to Fig. 6 in the main paper, but demonstrates the performance for individual corruptions in 3DCC.

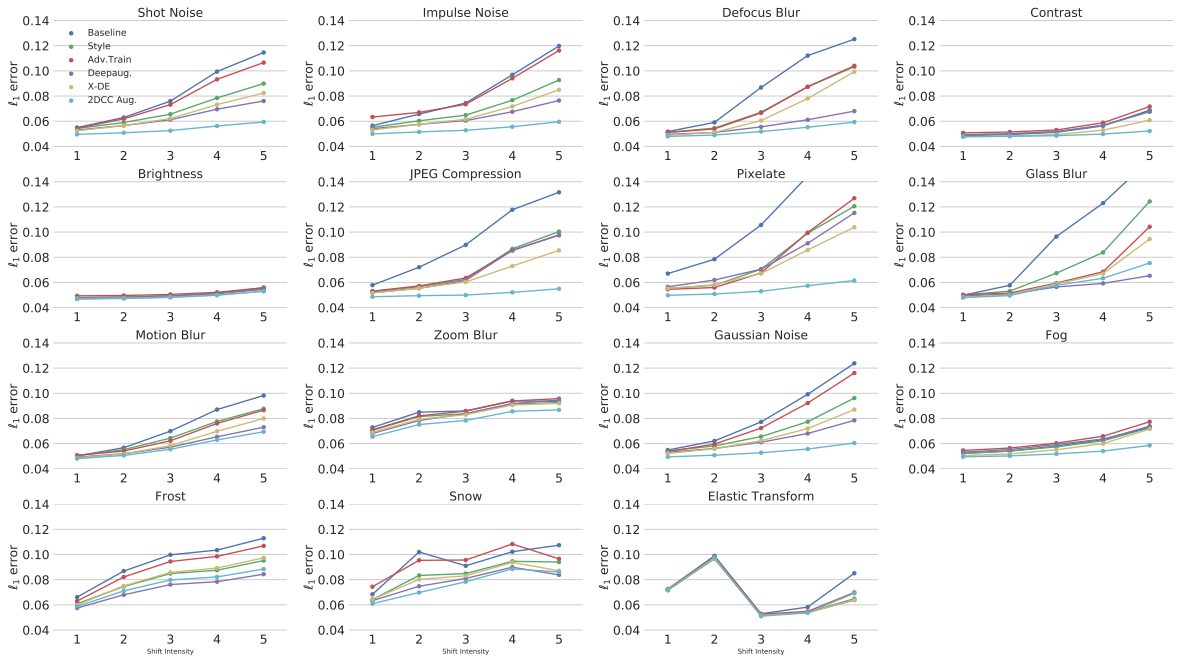

Figure 8. **Average ℓ_1 losses of robustness mechanisms against corruptions in 2DCC for surface normals estimation.** Similar to Fig. 6 in the main paper, but demonstrates the performance for individual corruptions in 2DCC.

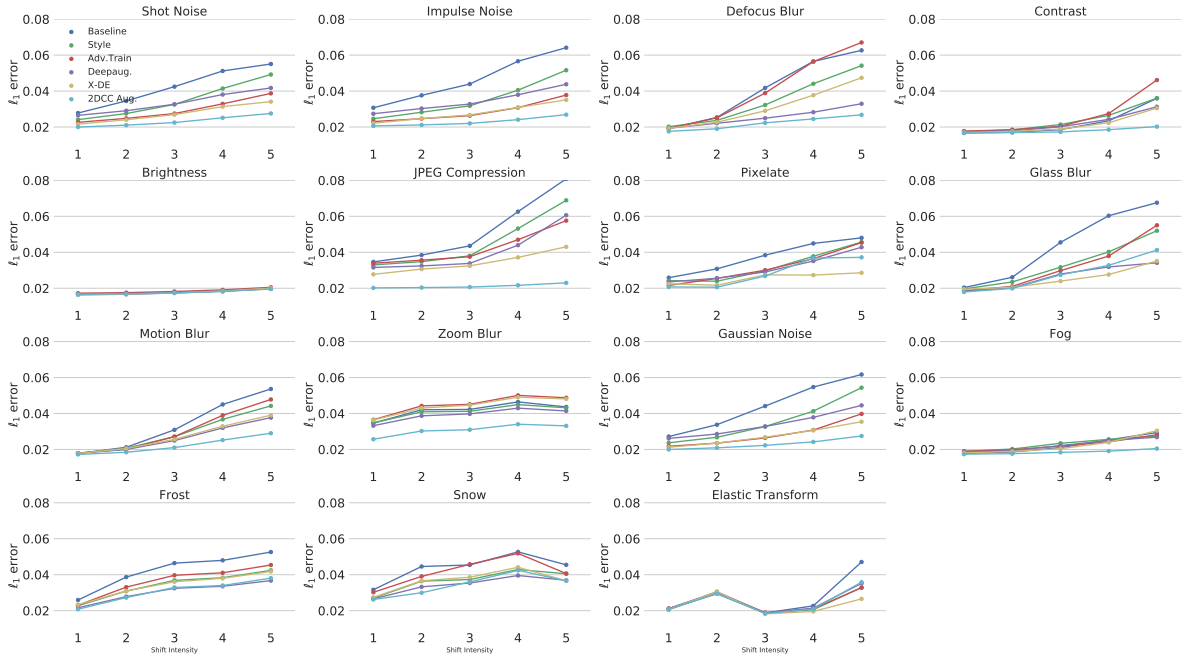

Figure 9. **Average ℓ_1 losses of robustness mechanisms against corruptions in 2DCC for depth estimation.** Similar to Fig. 6 in the main paper, but demonstrates the performance for individual corruptions in 2DCC.

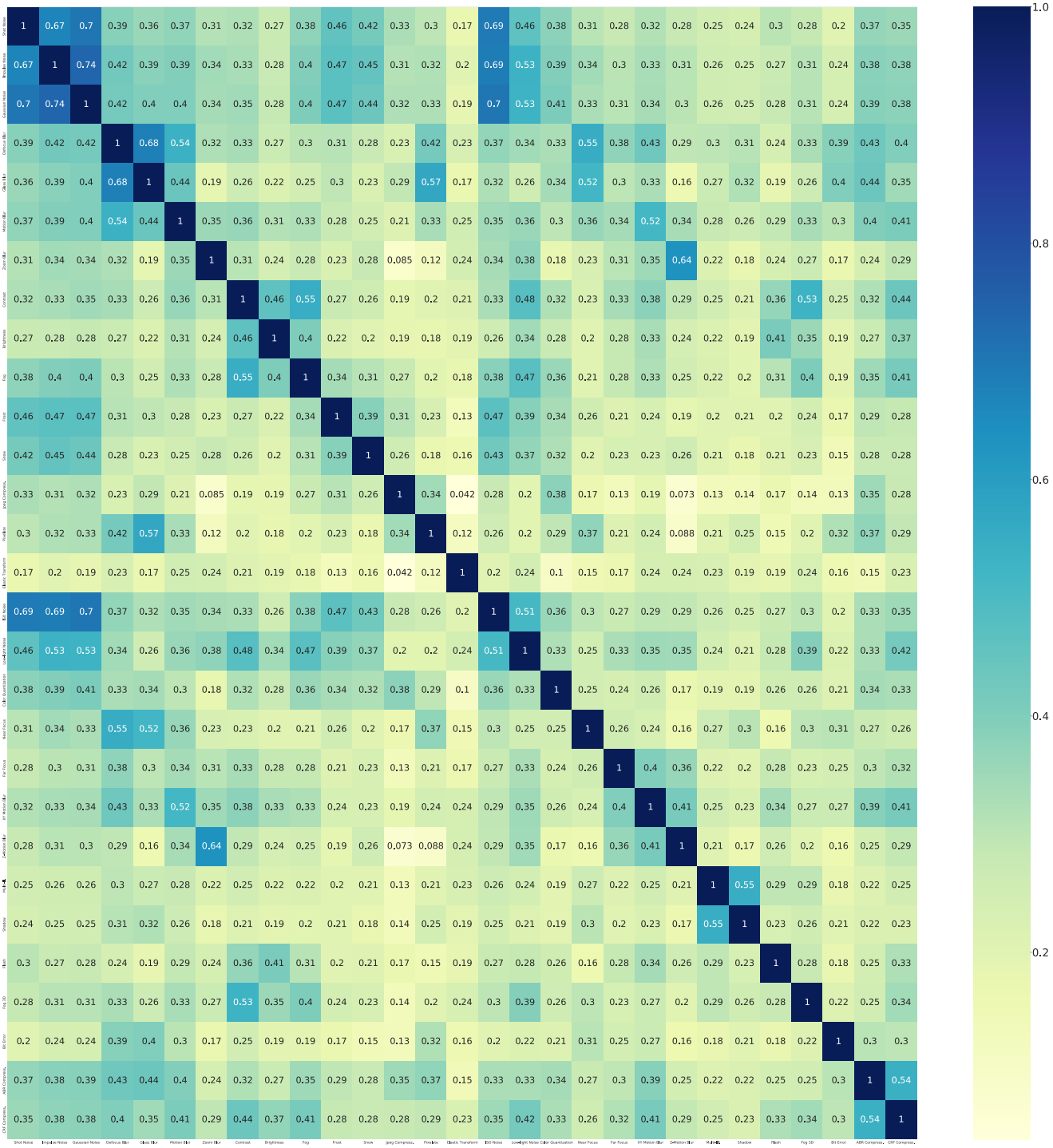

Figure 10. **Redundancy among corruptions in 2DCC and 3DCC in the ℓ_1 errors of surface normals prediction.** Similar to Fig. 9 in the main paper, but are shown for all corruptions.

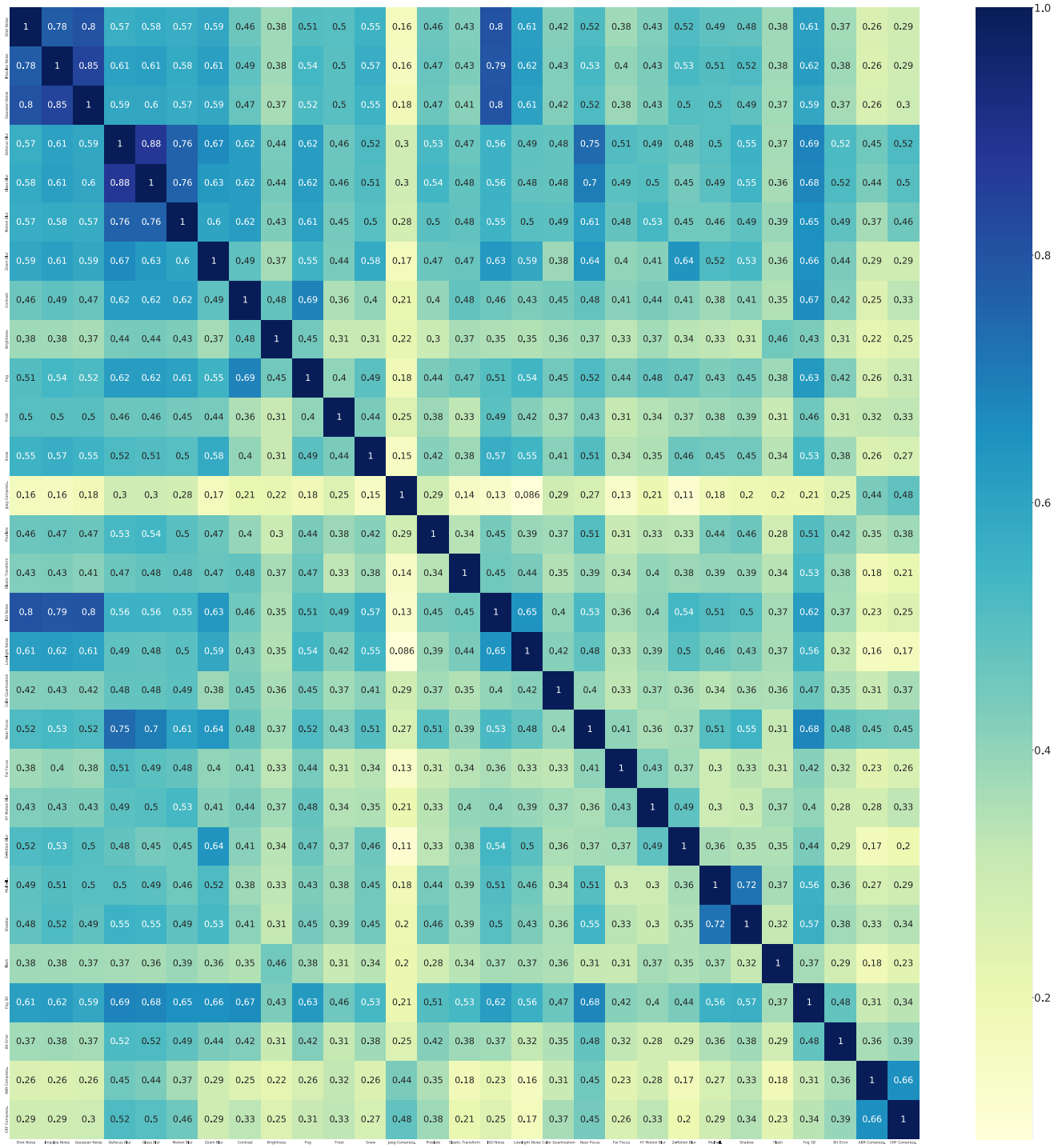

Figure 11. **Redundancy among corruptions in 2DCC and 3DCC in the ℓ_1 errors of depth prediction.** Similar to Fig. 9 in the main paper, but are shown for all corruptions.

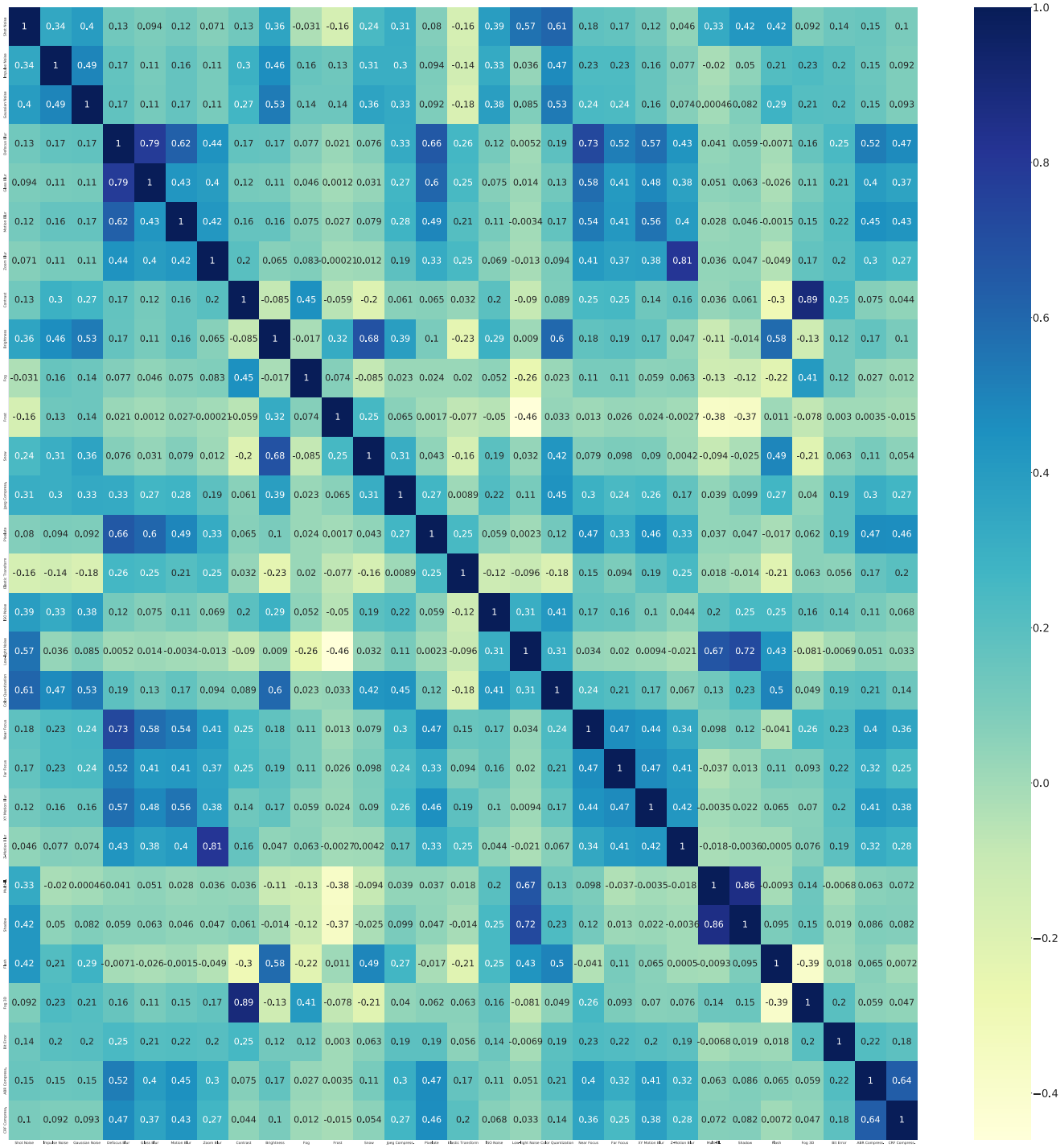

Figure 12. **Redundancy among corruptions in 2DCC and 3DCC in the ℓ_1 errors of RGB images.** Similar to Fig. 9 in the main paper, but are shown for all corruptions.

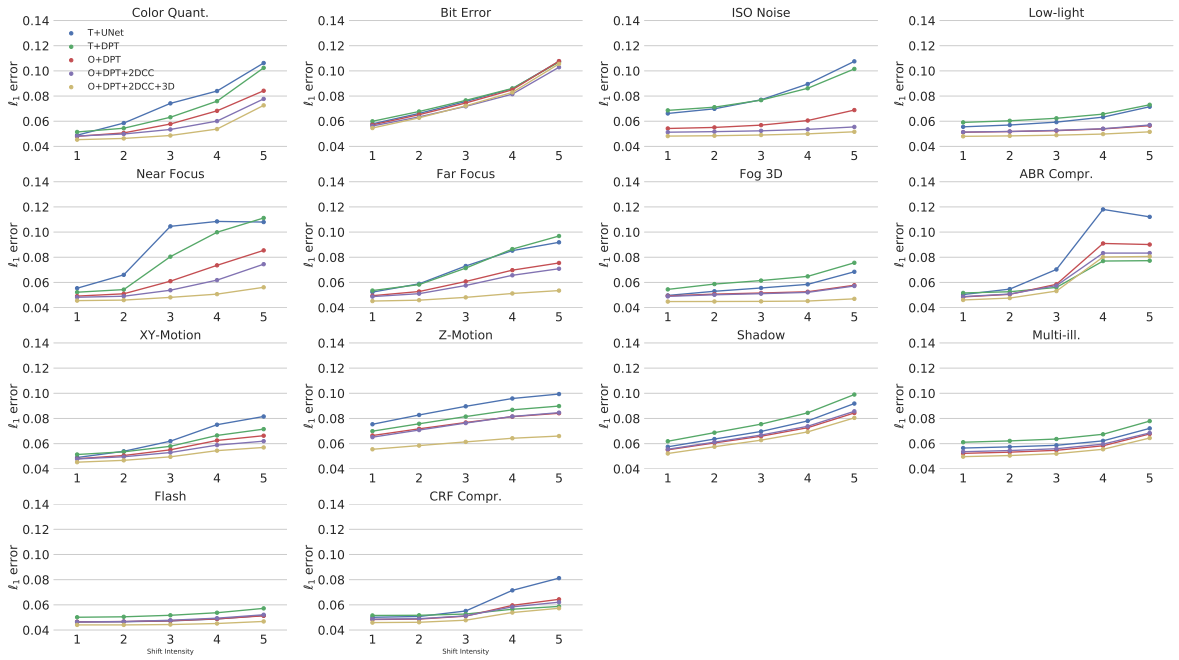

Figure 13. **Average ℓ_1 losses of 3D data augmentation and baselines against corruptions in 3DCC for surface normals estimation.** Similar to Tab. 1 in the main paper, but demonstrates the performance for individual corruptions in 3DCC.

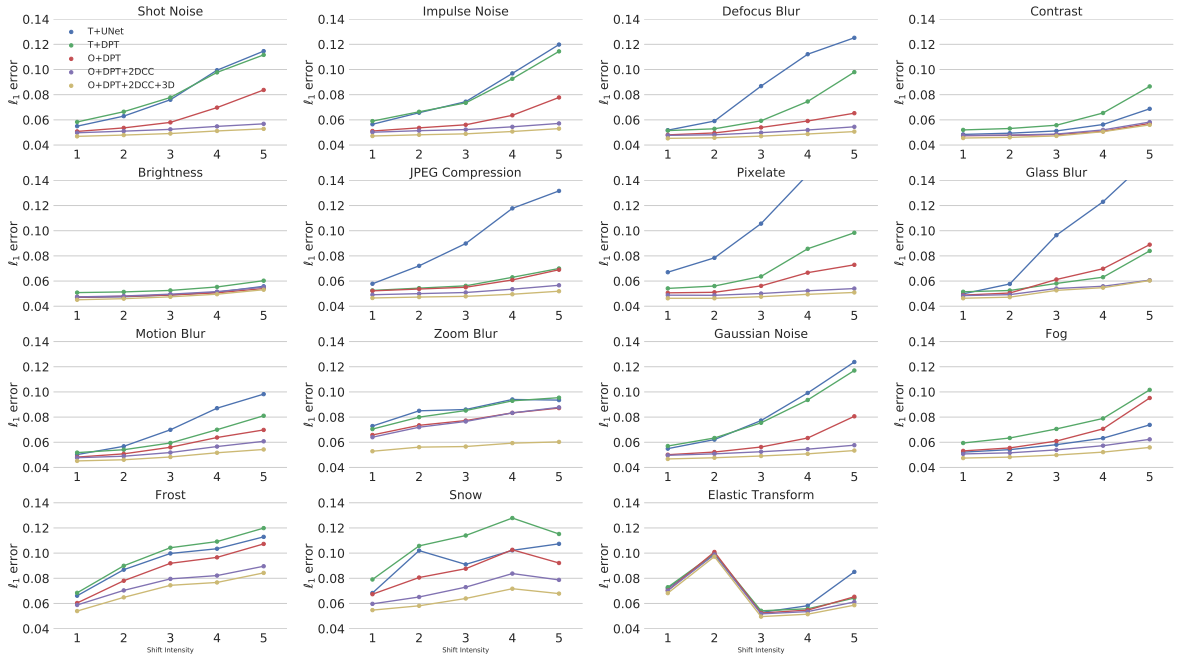

Figure 14. **Average ℓ_1 losses of 3D data augmentation and baselines against corruptions in 2DCC for surface normals estimation.** Similar to Tab. 1 in the main paper, but demonstrates the performance for individual corruptions in 2DCC.

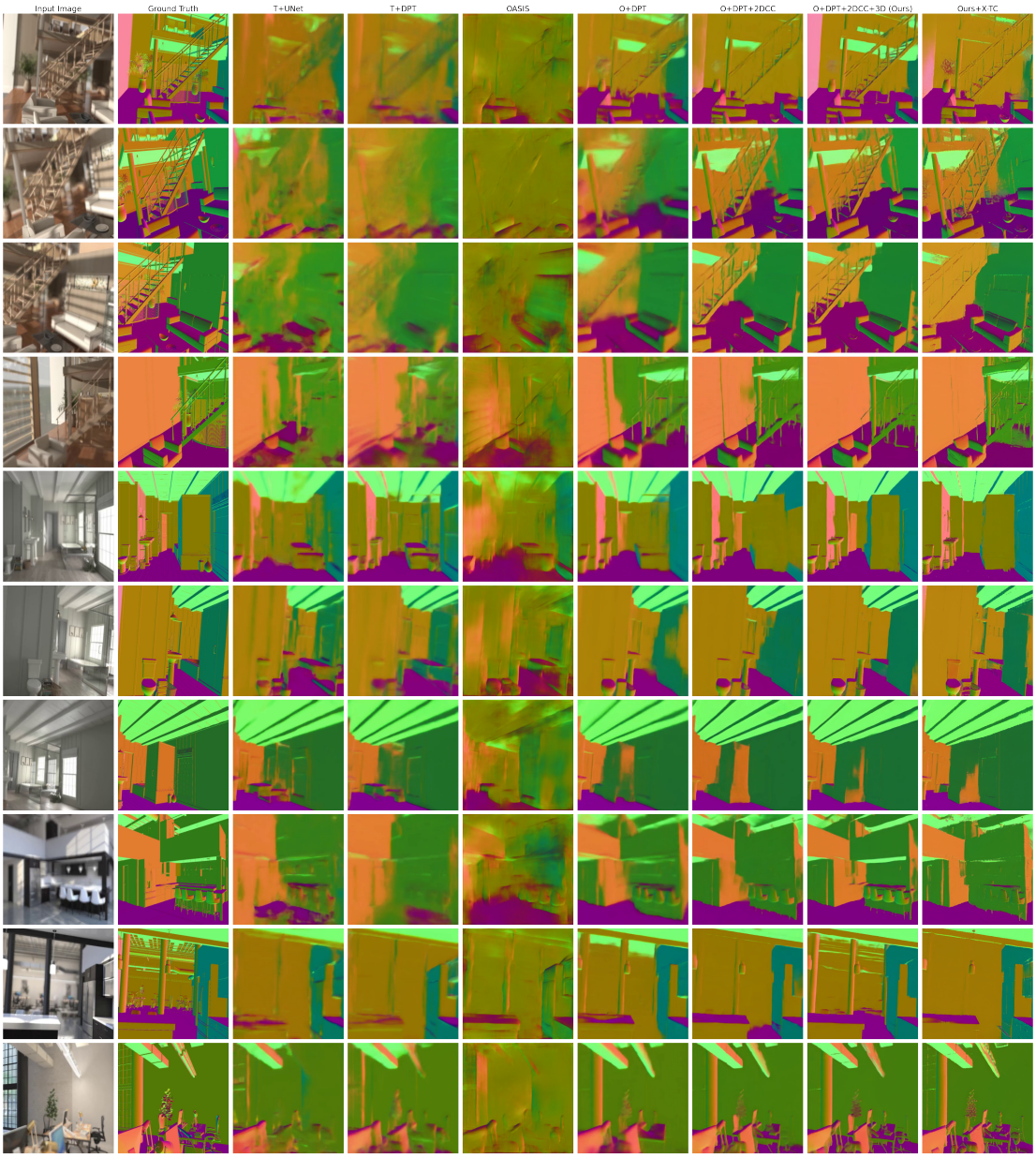

Figure 15. **Qualitative results on corruptions generated with After Effects.** An extension of the qualitative results in Fig. 8 of the main paper showing the benefits of 3D augmentation. See Sec. 5.2.3 of the main paper for details on how the corruptions were generated.

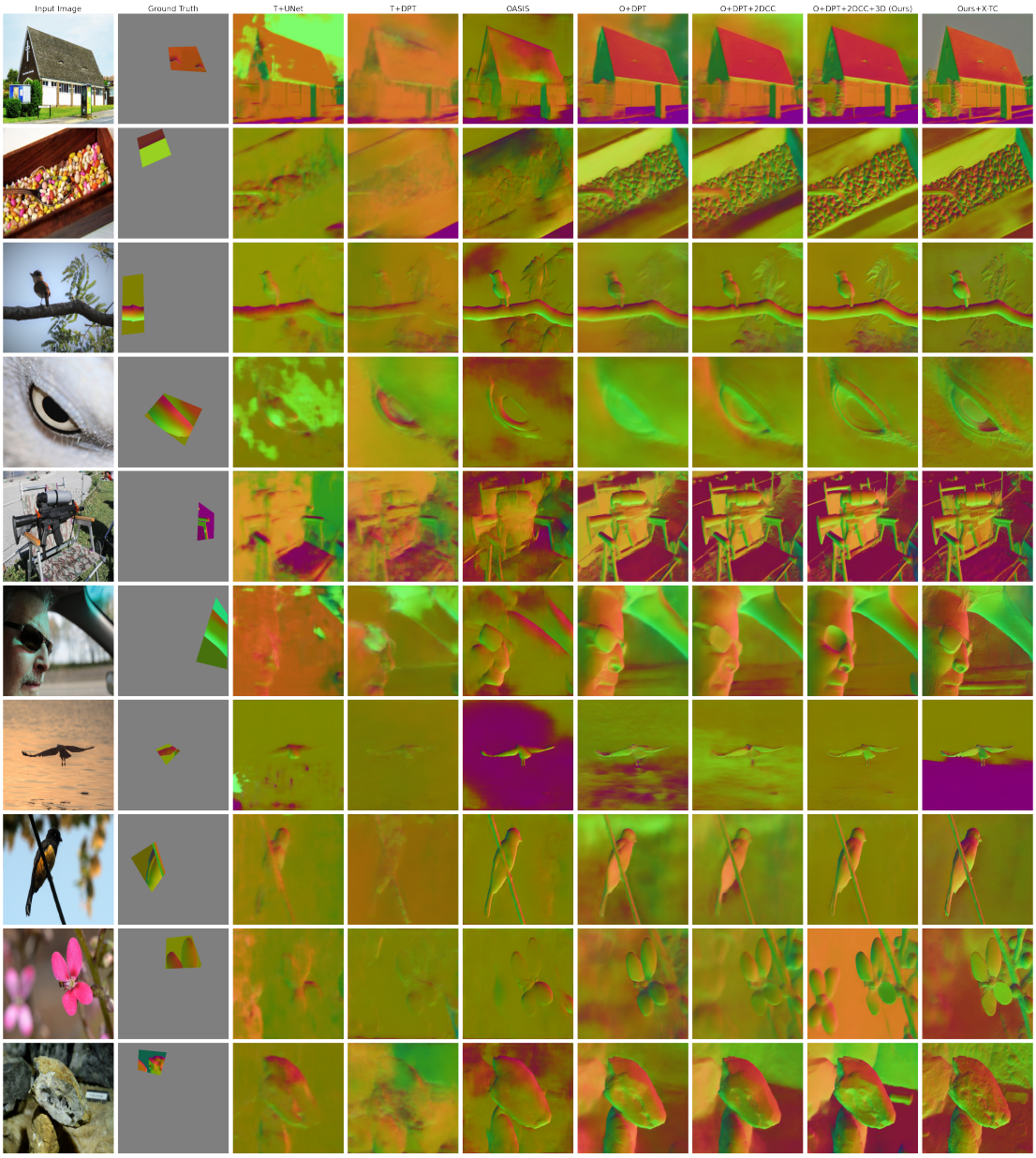

Figure 16. **Qualitative results on OASIS [2].** An extension of the qualitative results in Fig. 8 of the main paper showing the benefits of 3D augmentation.

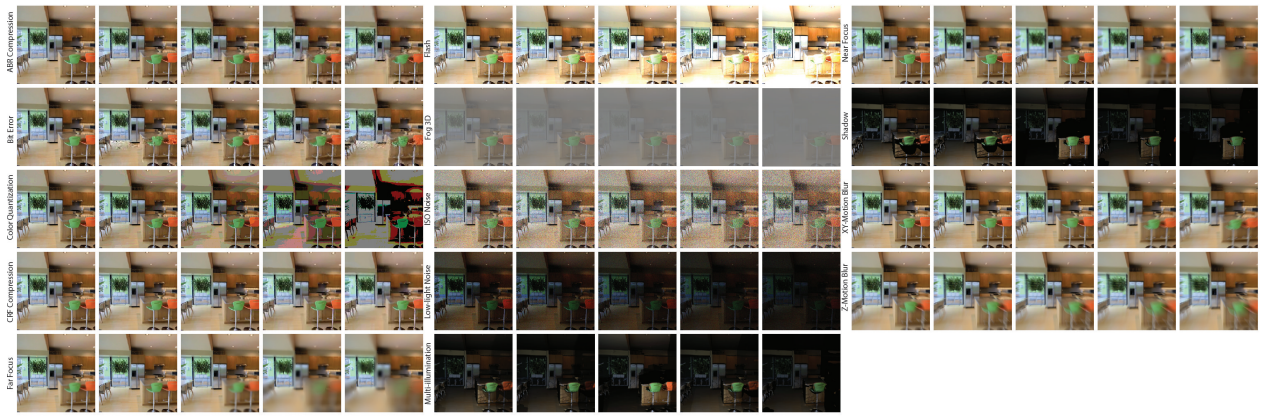

Figure 17. Visualizations of corruptions from 3DCC for 5 shift intensities.

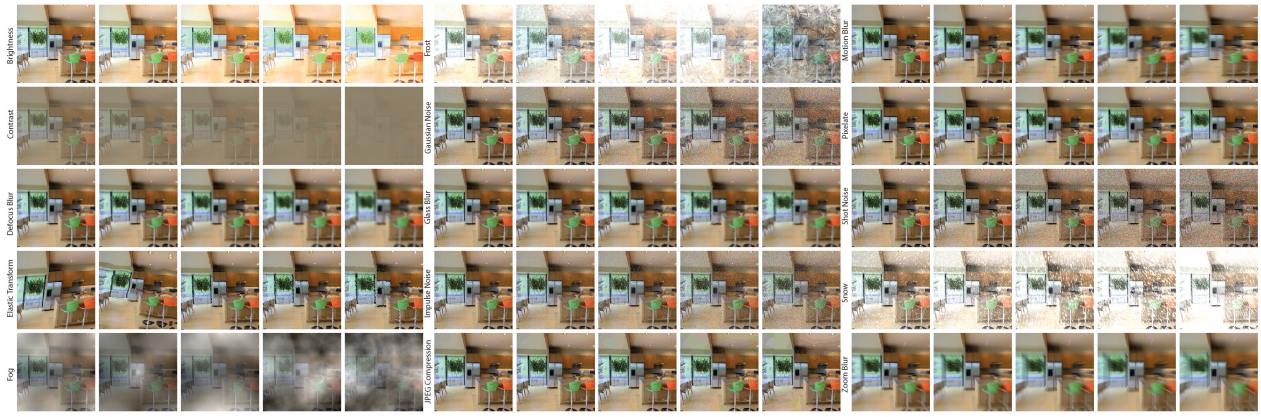

Figure 18. Visualizations of corruptions from 2DCC for 5 shift intensities.