Robustness via Cross-Domain Ensembles

Teresa Yeo*, Oğuzhan Fatih Kar*, Amir Zamir

EPFL

crossdomain-ensembles.epfl.ch



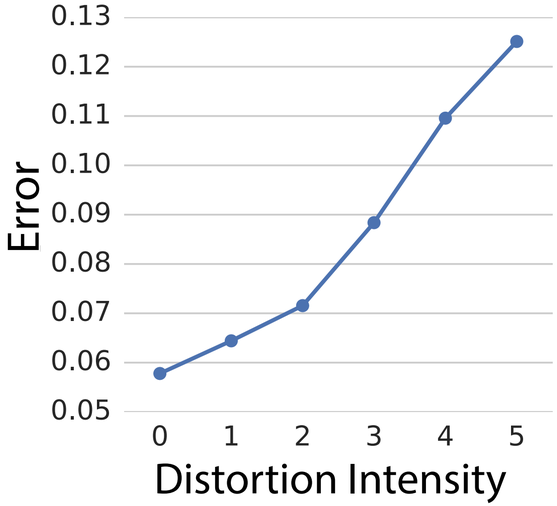
Neural networks are not robust under distribution shifts

Clean input (Intensity 0)

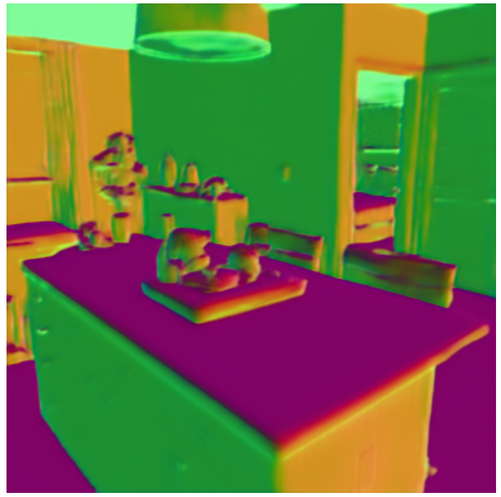
Distorted
(Defocused Blur, Intensity 2)

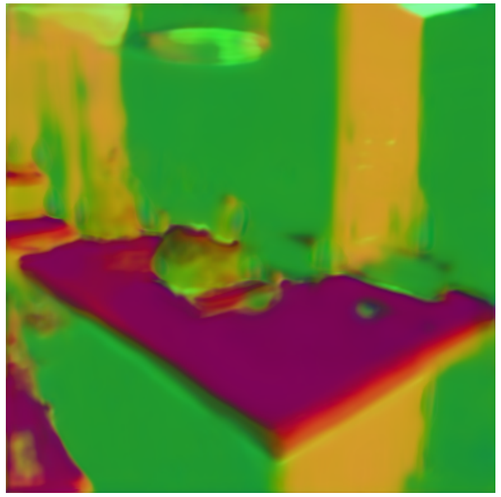
Distorted
(JPEG Compressed, Intensity 2)

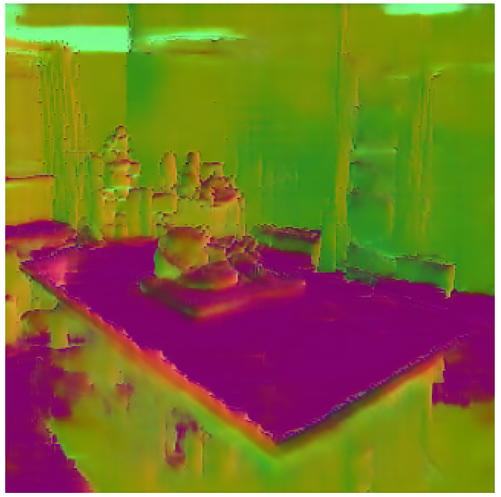
Query Image



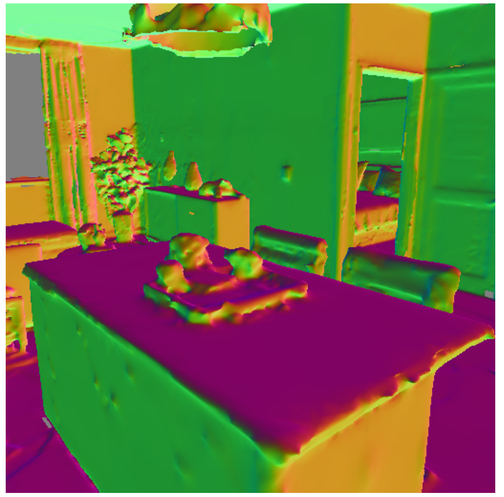
Prediction





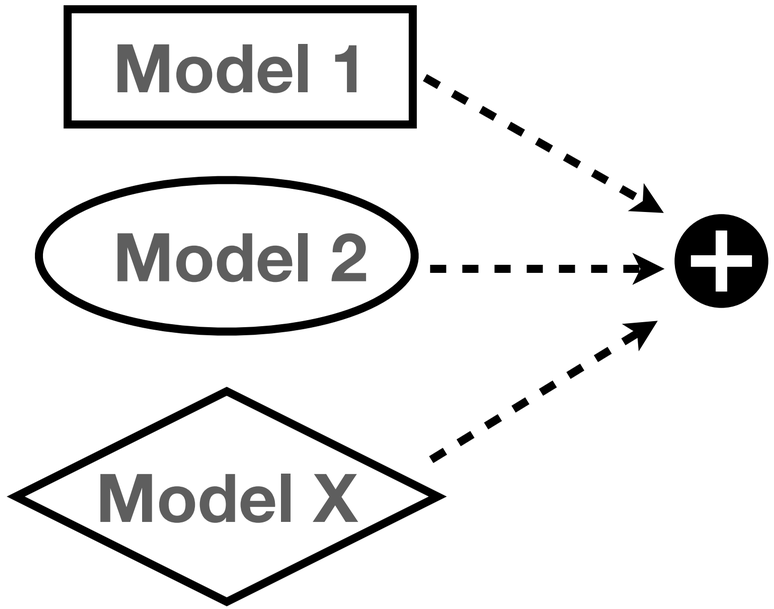


Ground Truth



Ensembles

- **Core idea**: using diversity to de-correlate errors
  - Sources of diversity
    - Different initialisations[1]
    - Hyperparameters[2]
    - Network architectures[3]
    - Loss constraints[4]



1. *Lakshminarayanan, et. al.* 2017.
2. *Wenzel et. al.* 2020.
3. *Zaidi et. al.* 2020.
4. *Yang et. al.* 2020.

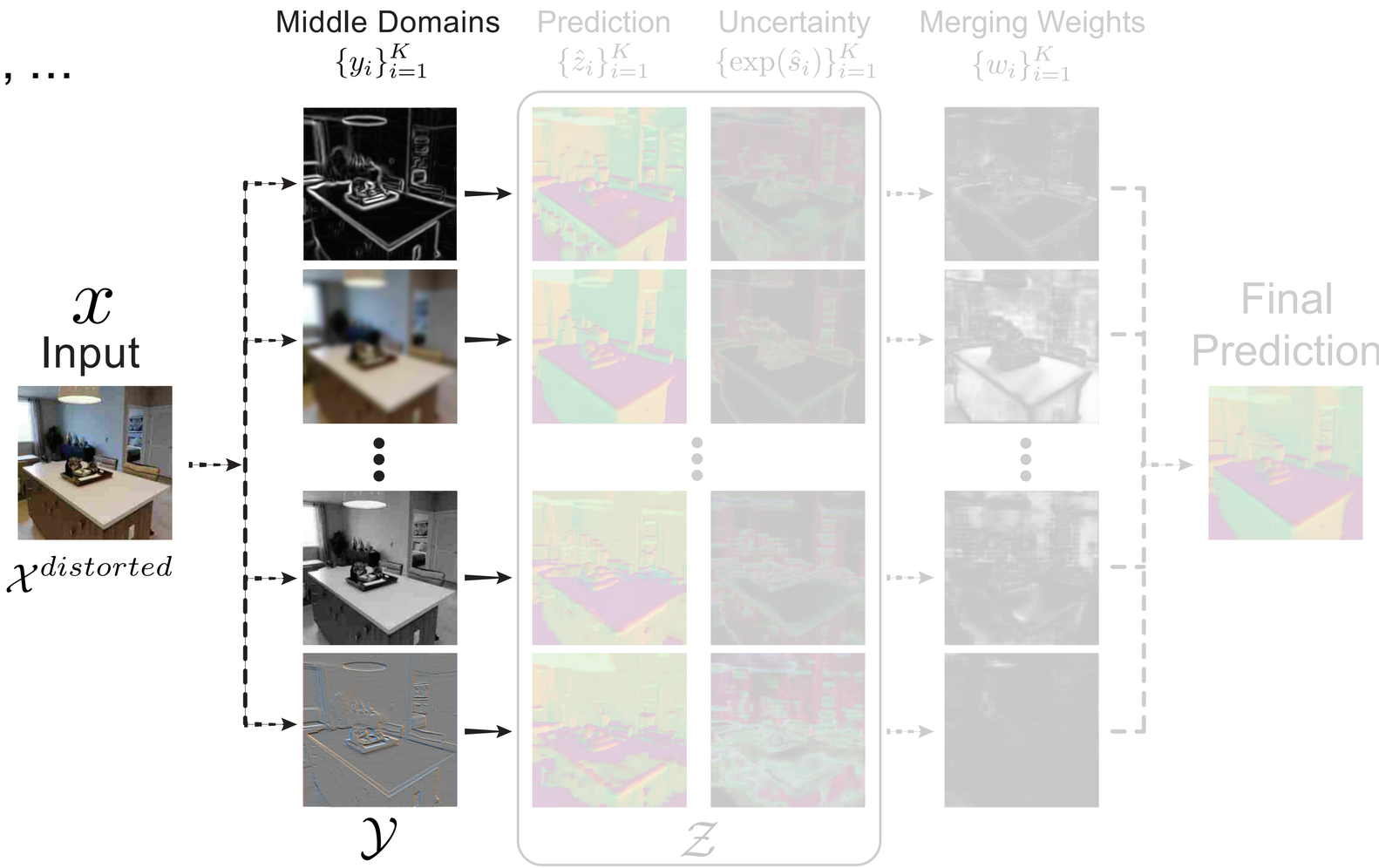
Our Proposed Method

- Diverse set of cues (middle domains)
- Merging based on uncertainty

Our Proposed Method

Extracting middle domains

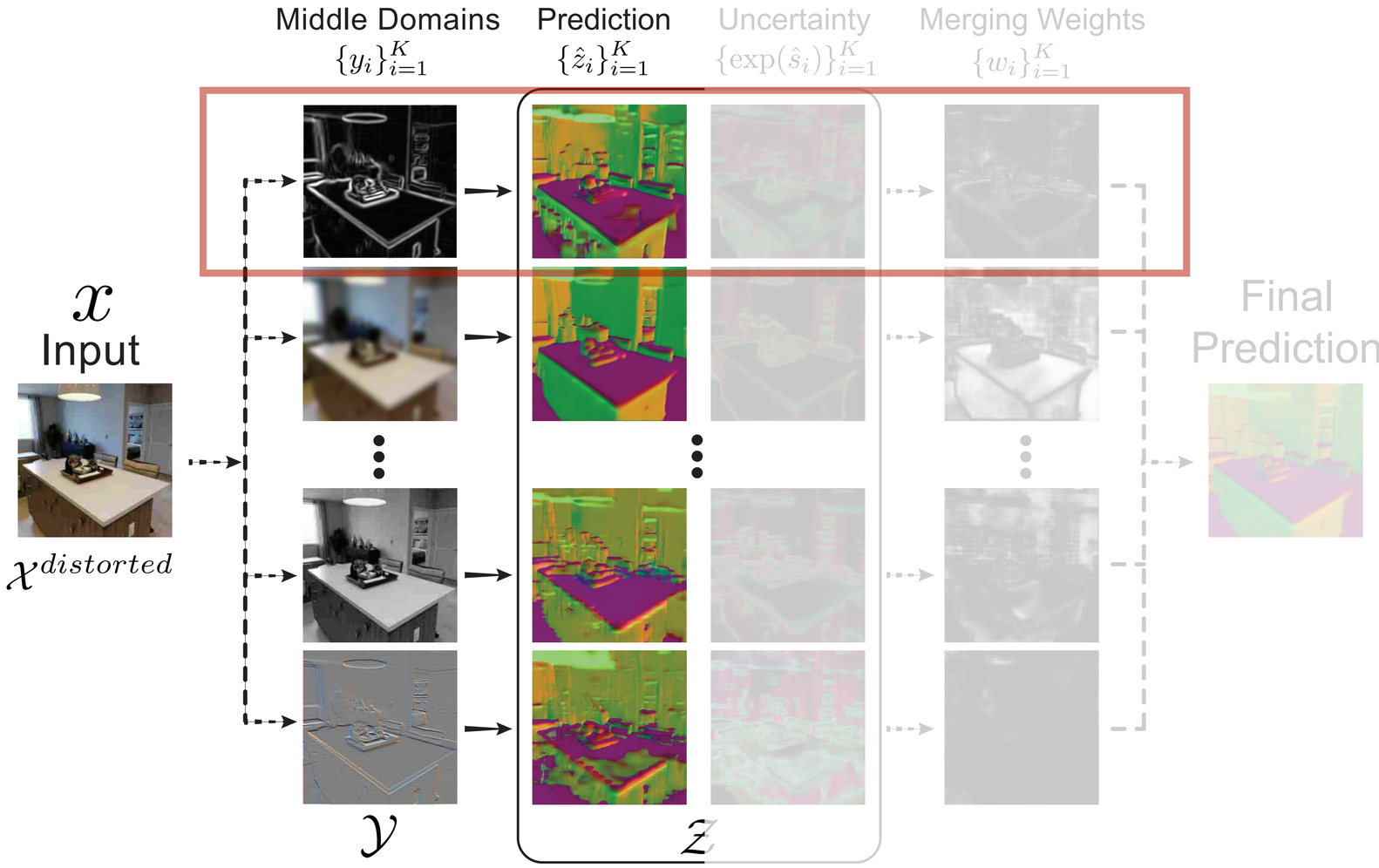
- 2D edges, low-pass filtering, …
- No labels needed



Our Proposed Method

Learning from middle domains

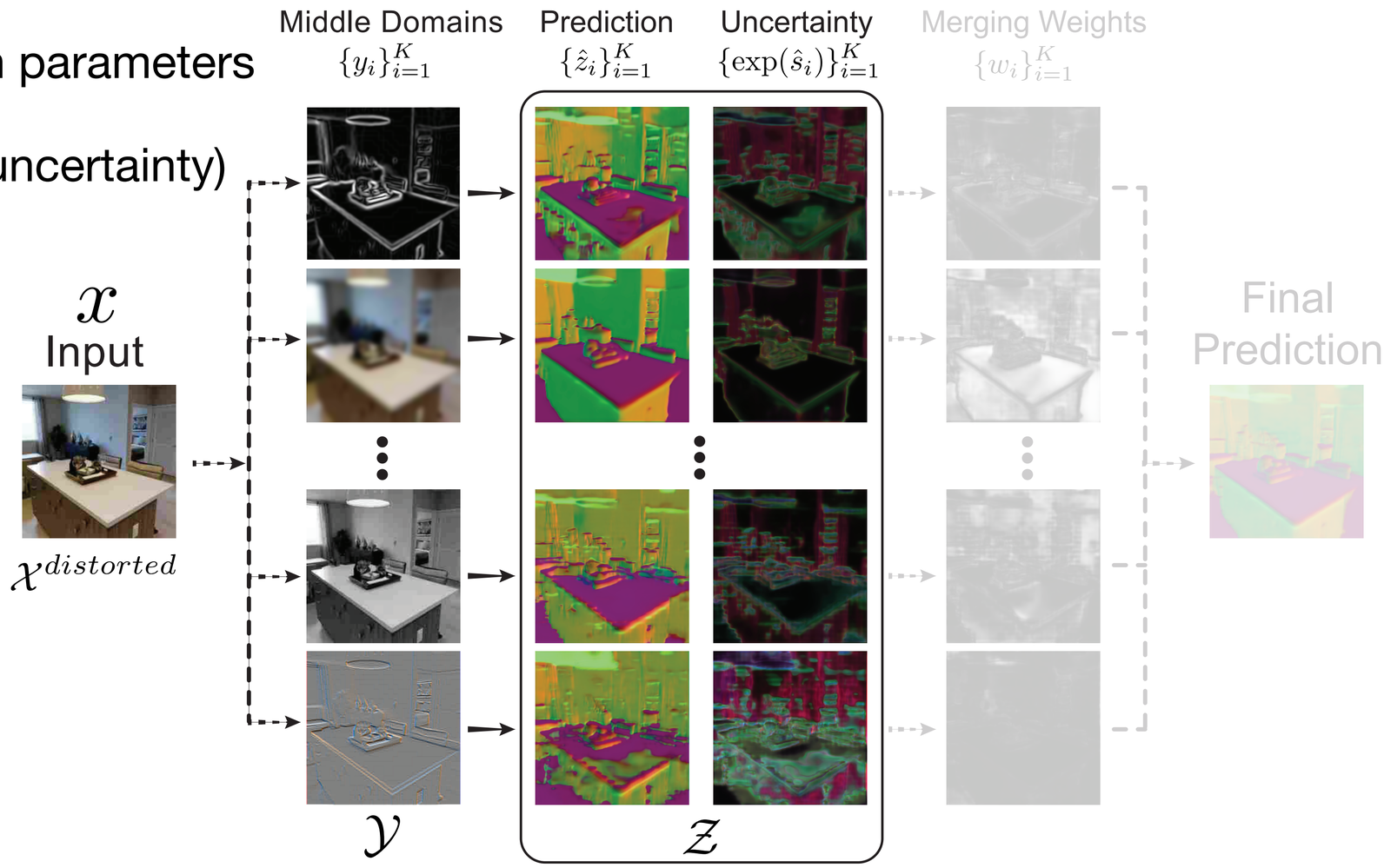
- Diverse predictions



Our Proposed Method

Predicting uncertainties

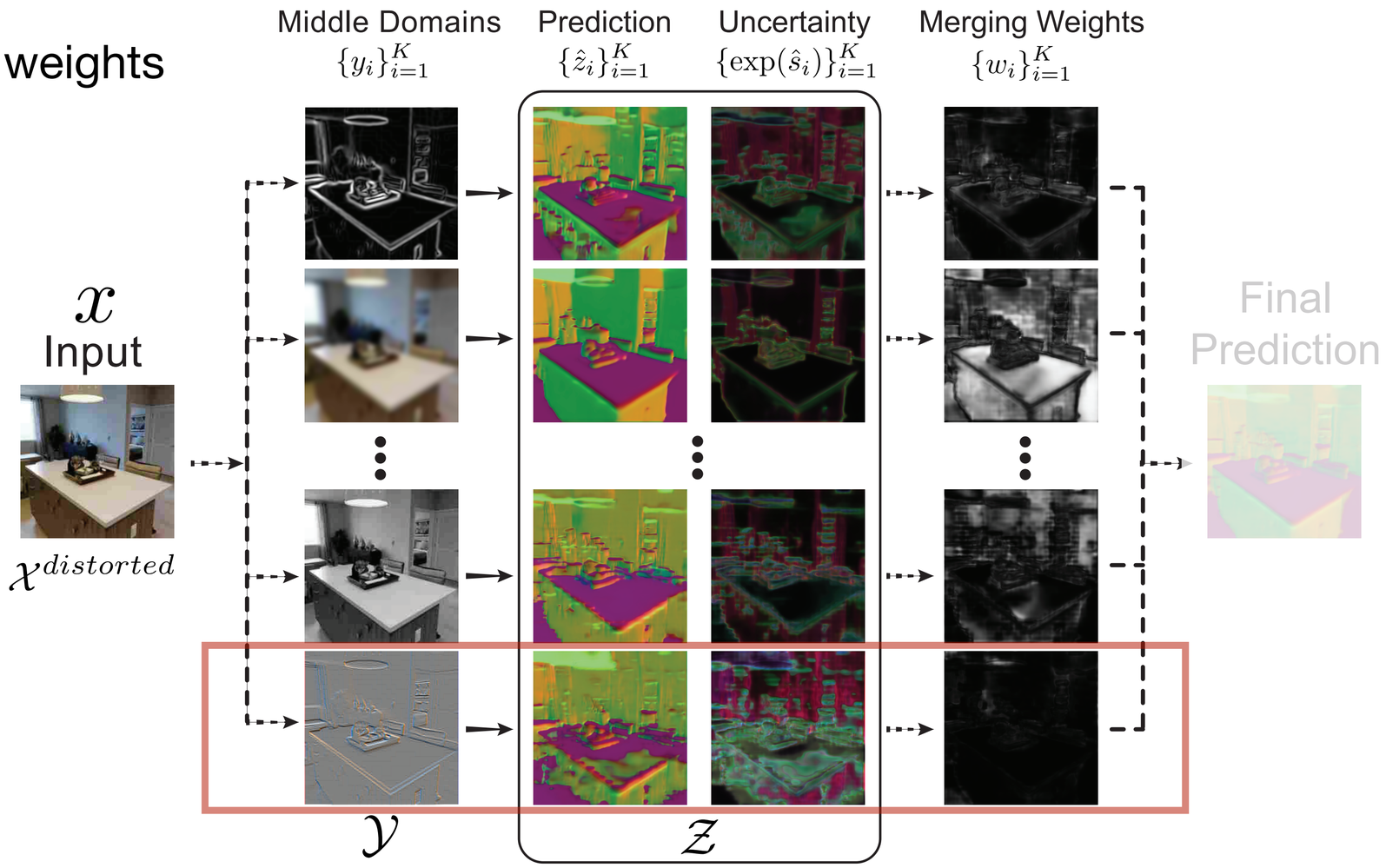
- Output: Laplace distribution parameters
- (mean,sigma)→(prediction,uncertainty)
- Use likelihood loss



Our Proposed Method

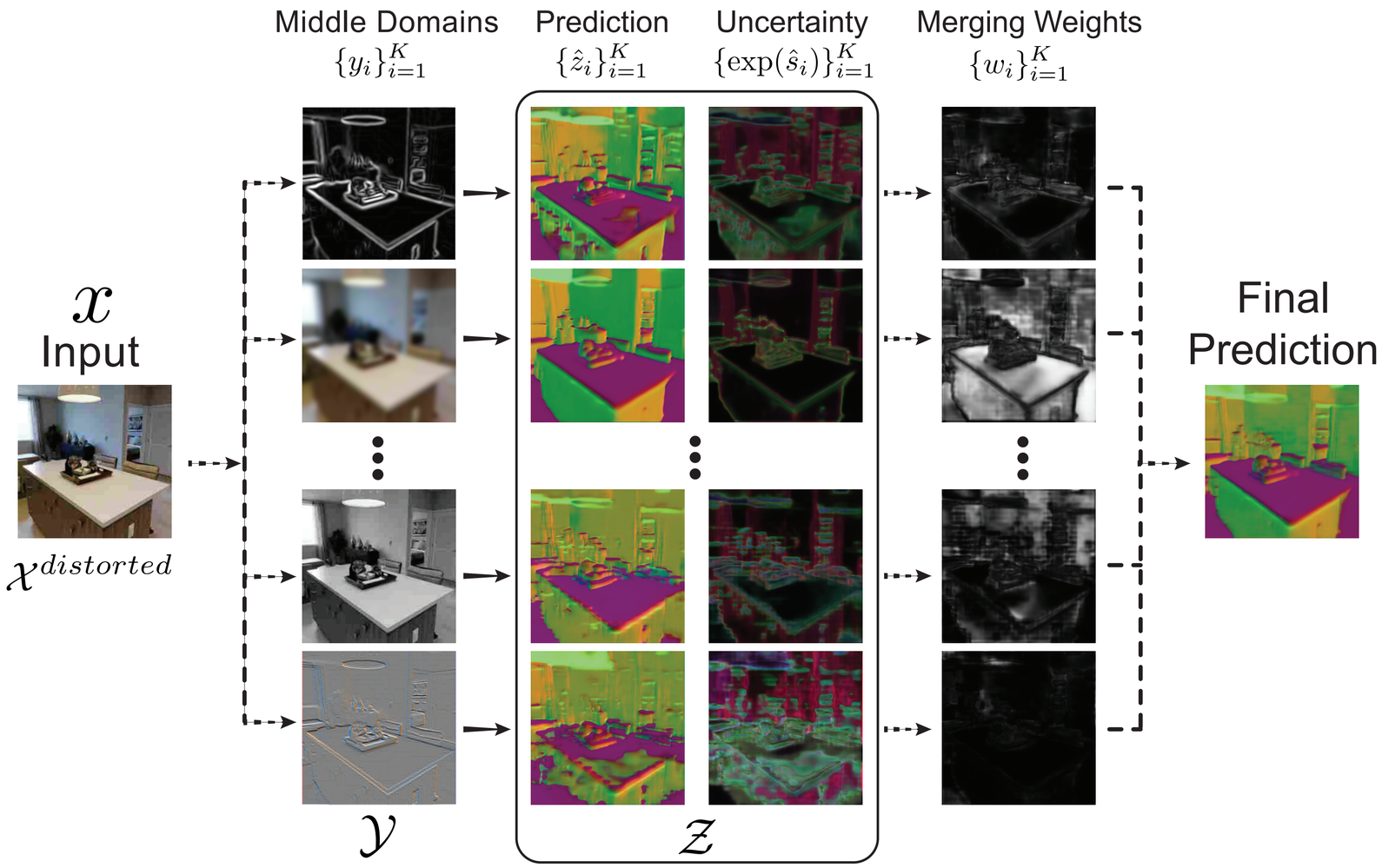
Using uncertainties as weights

- Higher uncertainty → lower weights

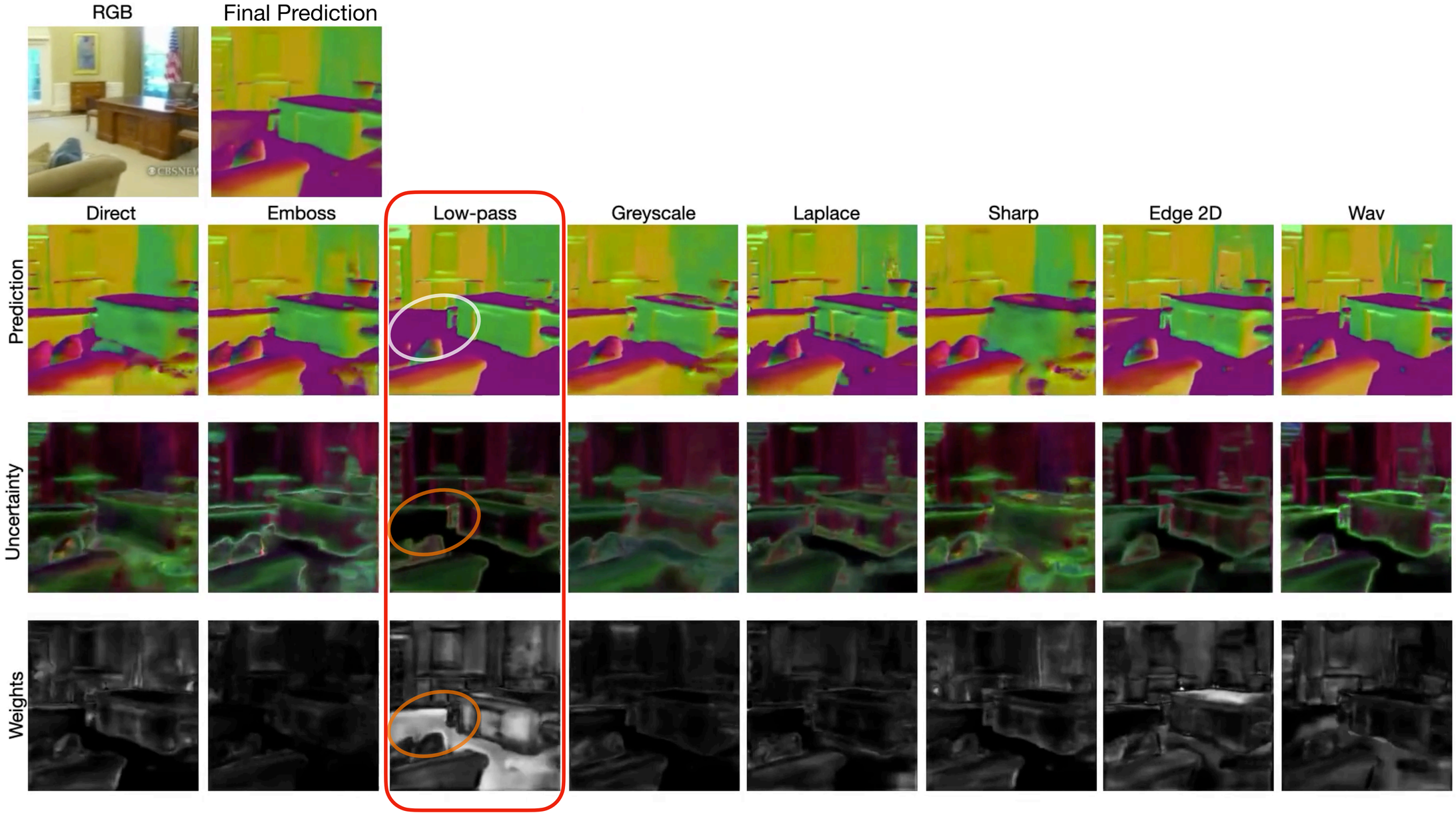


Our Proposed Method

Getting the final prediction



Our Proposed Method



Uncertainty Estimates are Overconfident

Calibrating uncertainties via sigma training

- Output: Laplace distribution parameters
  - (mean,sigma)→(prediction,uncertainty)
- Training with out-of-distribution data
- $\mathscr{L}_{\text{sigma training}} = \mathscr{L}_{\text{sigma calibration}} + \mathscr{L}_{\text{mean grounding}}$

  Help uncertainties increase under distorted inputs

  Keeps predictions fixed

Addressing Overconfident Incorrect Predictions

- Before sigma training → overconfident predictions
- Similar for deep ensembles

Addressing Overconfident Incorrect Predictions

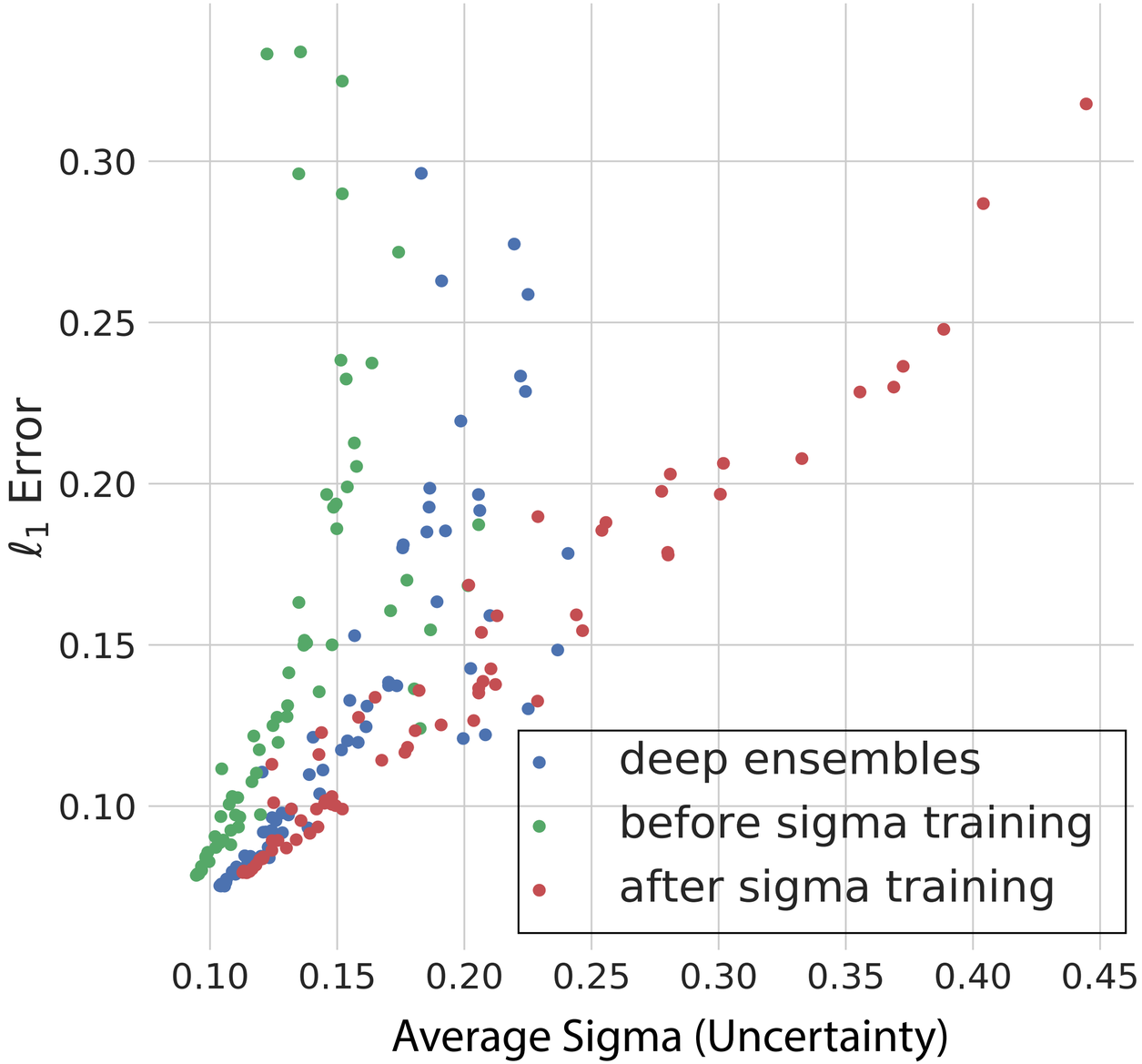
- After sigma training → calibrated uncertainties

Addressing Overconfident Incorrect Predictions

- Predictions are not updated

Addressing Overconfident Incorrect Predictions

- Before sigma training, deep ensemble: Poor correlation with error
- After sigma training: Strong correlation with error



Sample Results

Sample Results

Key Takeaways

- We obtain *notable improvements* in robustness with our method compared to several baselines
  - Against *non-adversarial* shifts (Common Corruptions[2]) and *adversarial* (I-FGSM[1])
  - For several *tasks* and *datasets*
    - Dense pixel-wise regression (surface normals, reshading, depth) on Taskonomy[3], Replica[4], Habitat[5]
    - Object classification on ImageNet[6,7] and CIFAR[8]

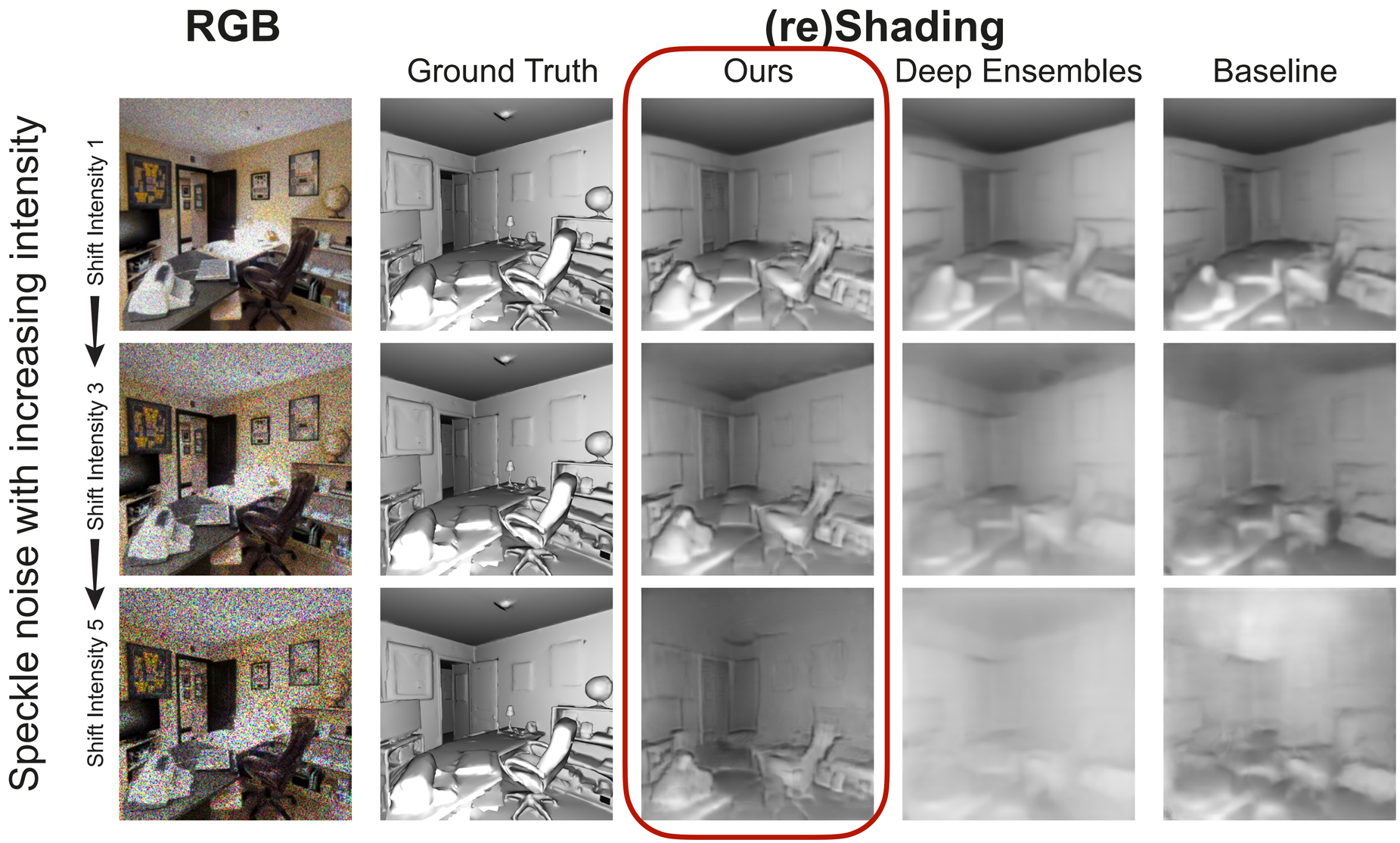
[1] *Kurakin et. al. 2016.*
[2] *Hendrycks et. al. 2019.*
[3] *Zamir et. al. 2018.*
[4] *Straub et. al. 2019.*
[5] *Savva et. al. 2019.*
[6] *Russakovsky et. al. 2015*
[7] *Shankar et. al. 2019*
[8] *Krizhevsky et. al. 2009*

Qualitative results

Under increasing speckle noise

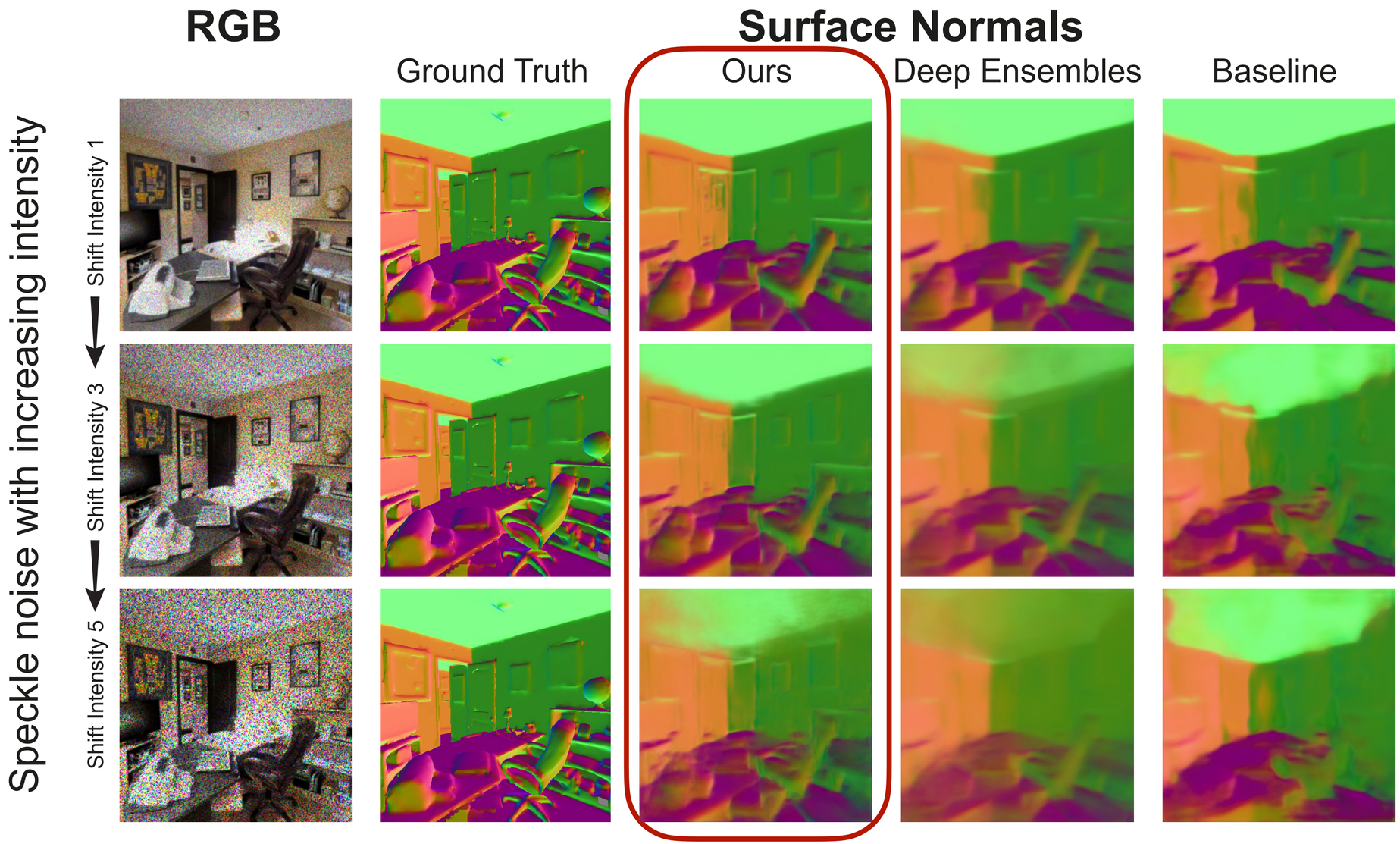
Qualitative results

Under increasing speckle noise



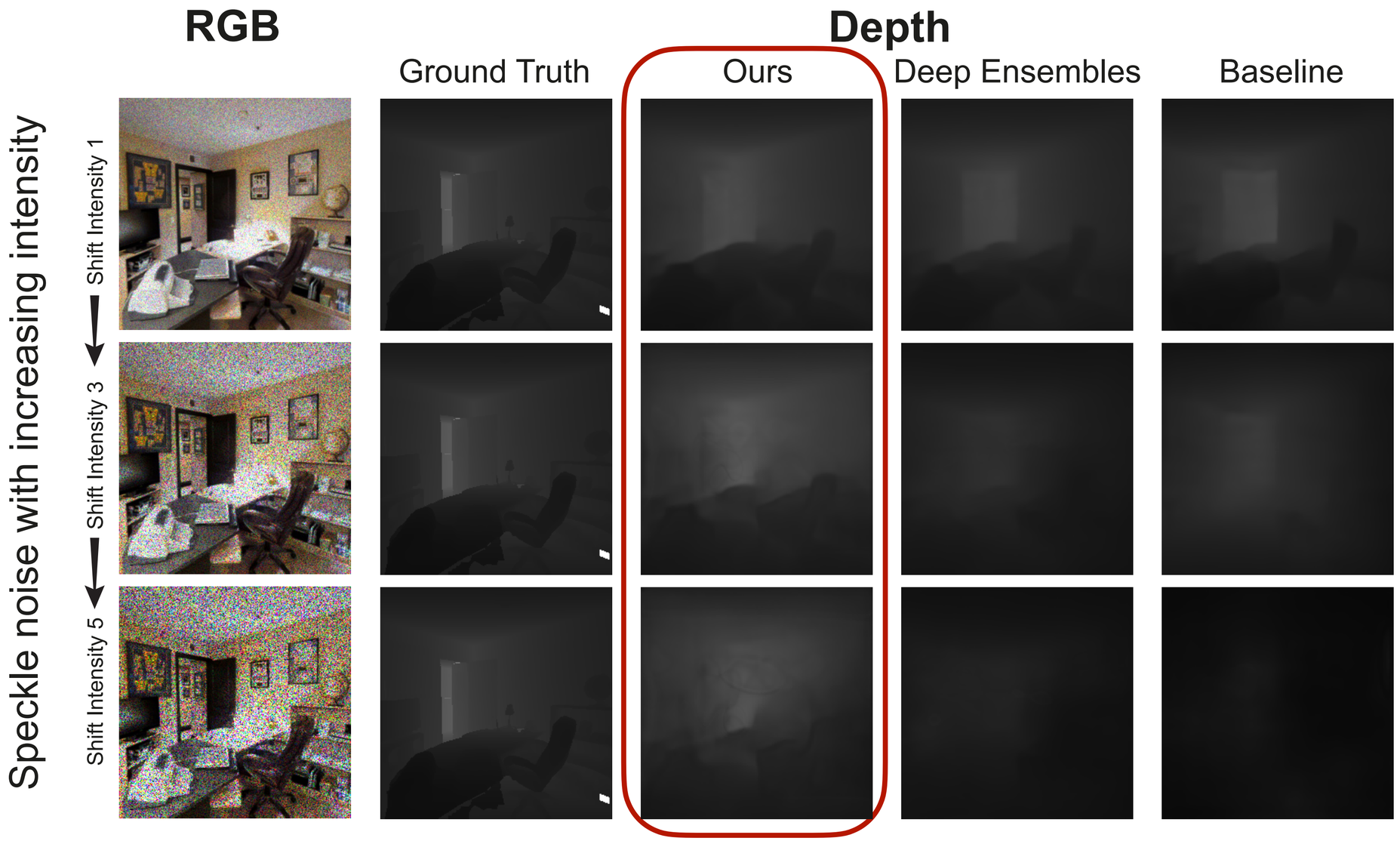
Qualitative results

Under increasing speckle noise



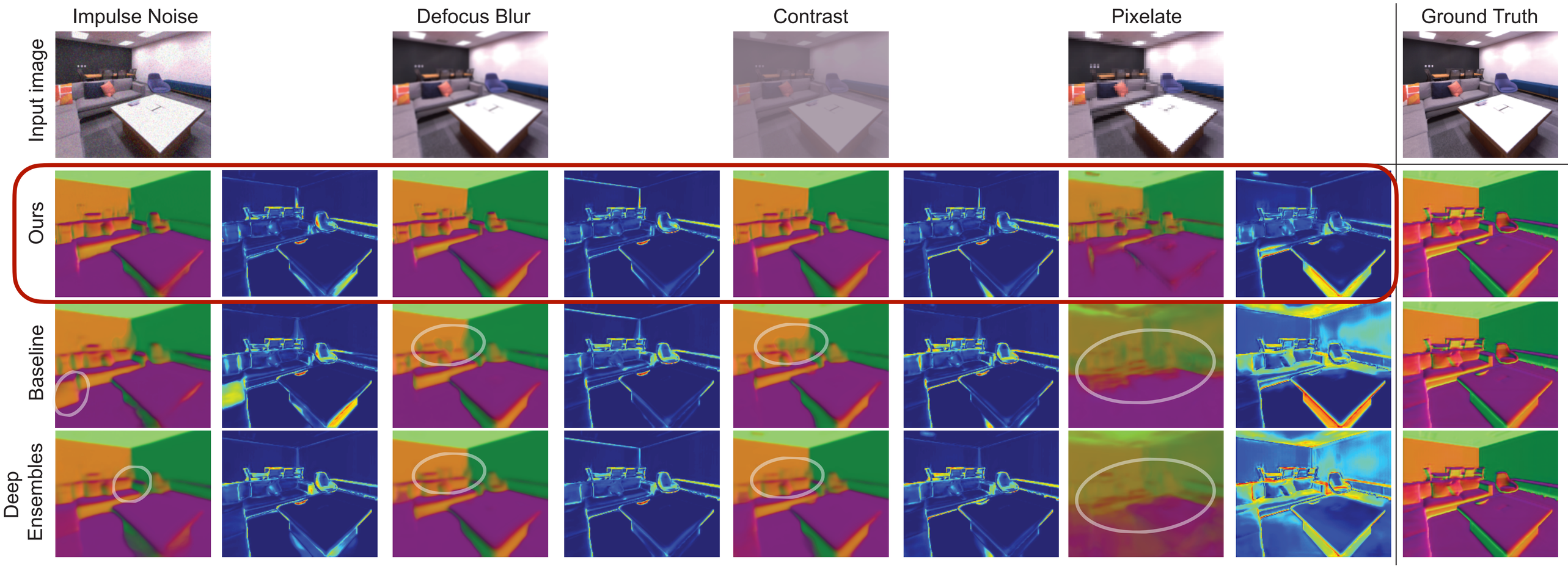
Qualitative results

Under increasing speckle noise



Qualitative results

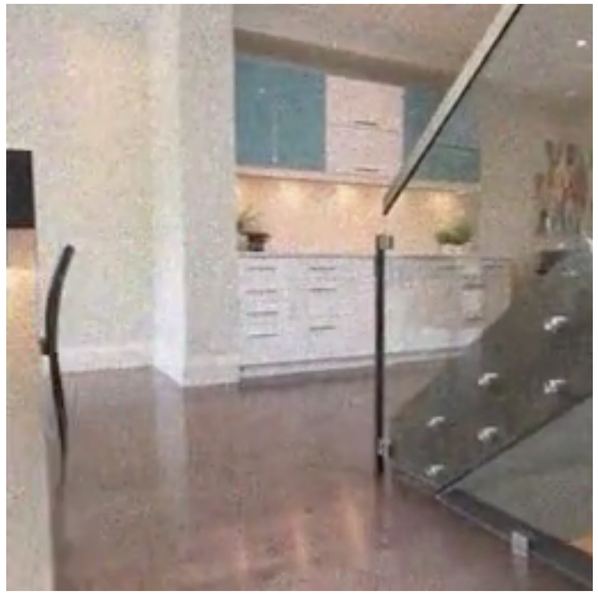
Under 4 unseen distortions



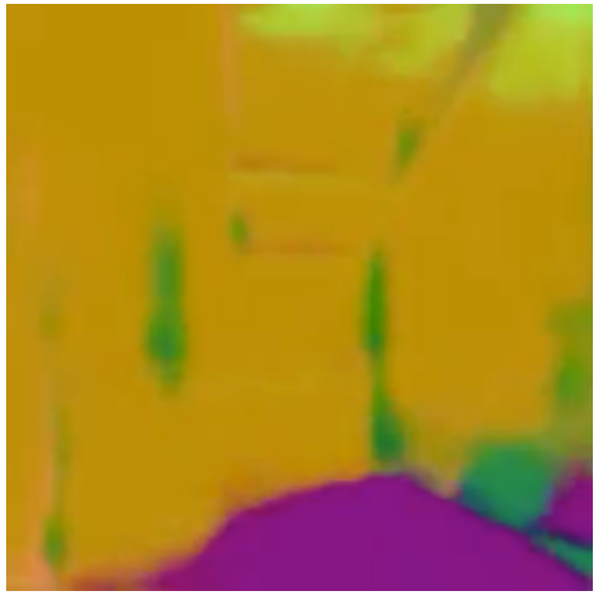
Qualitative results

Video corrupted with increasing shot noise

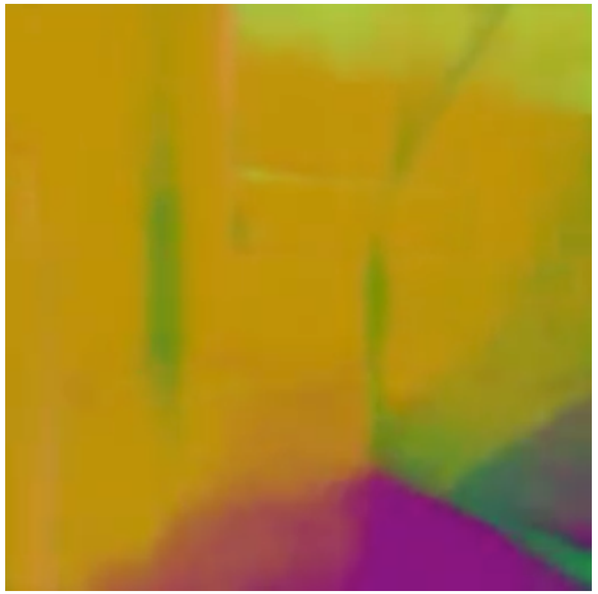
RGB



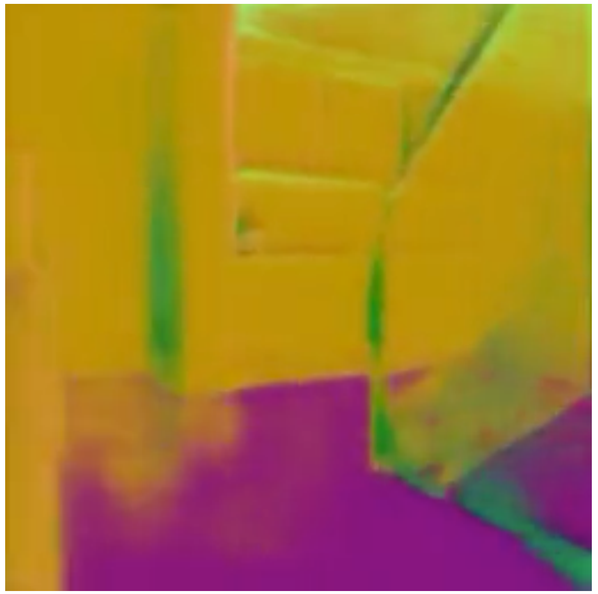
Baseline



Deep Ensembles



Ours

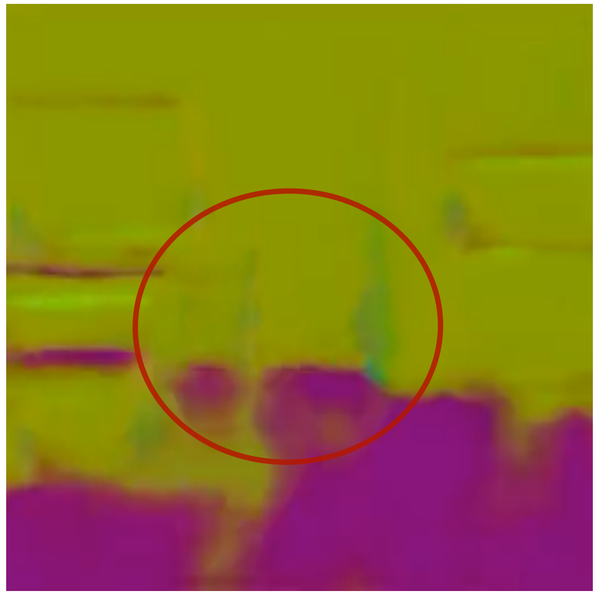


Qualitative results

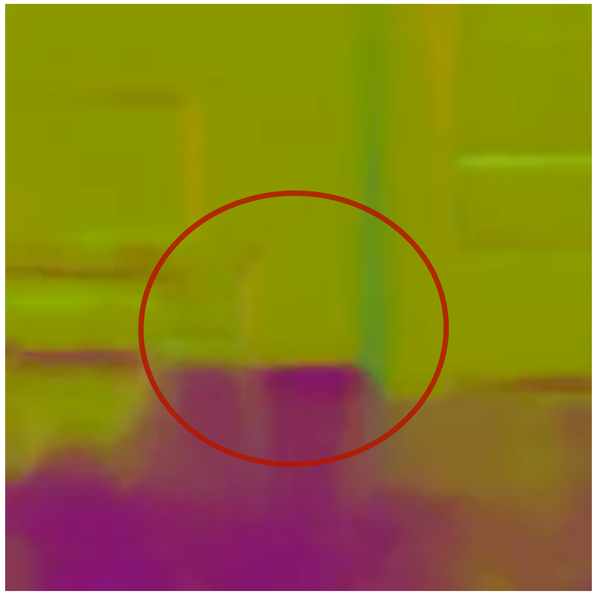
Video corrupted with increasing shot noise

RGB

Baseline



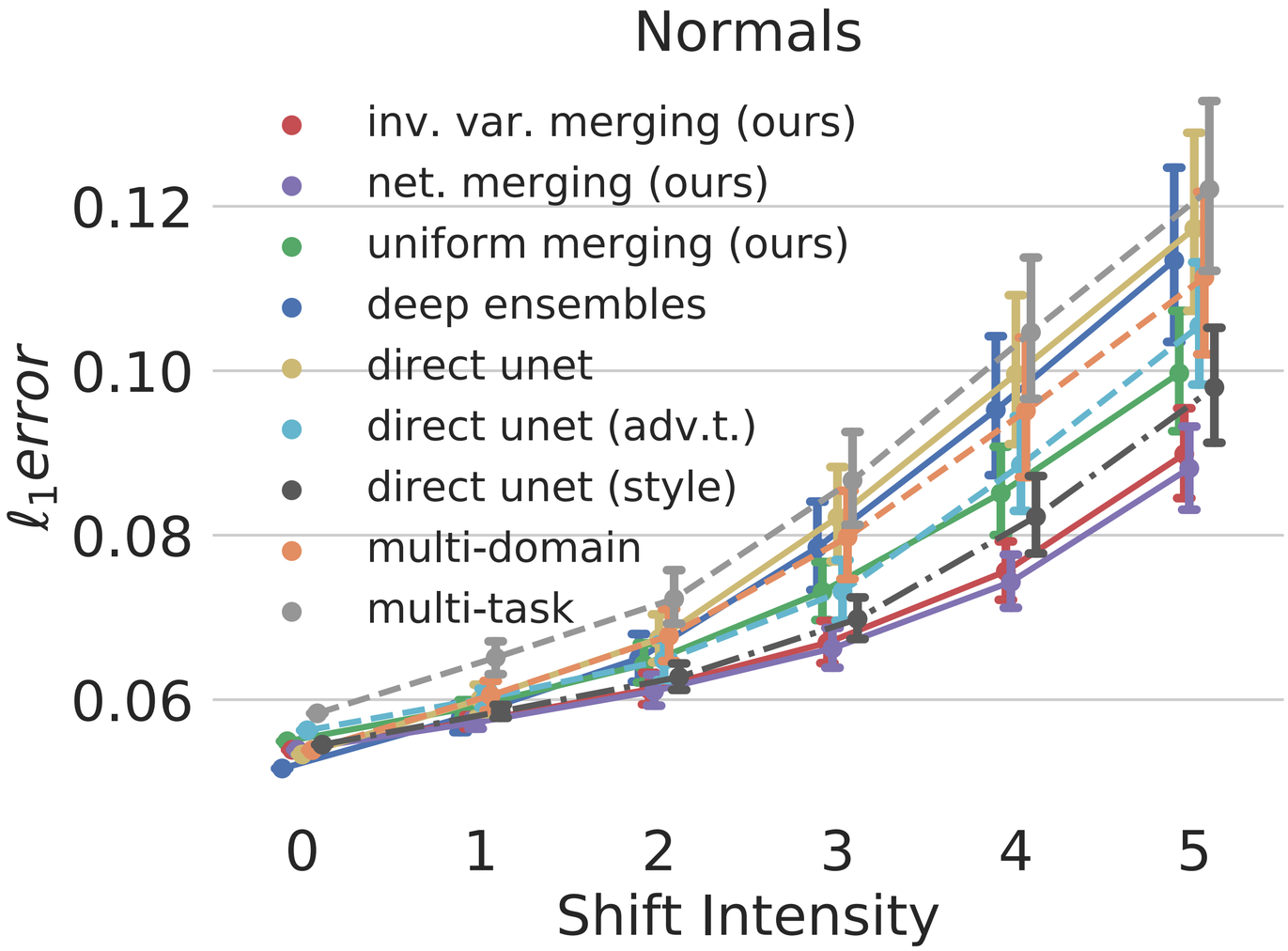
Deep Ensembles



Ours

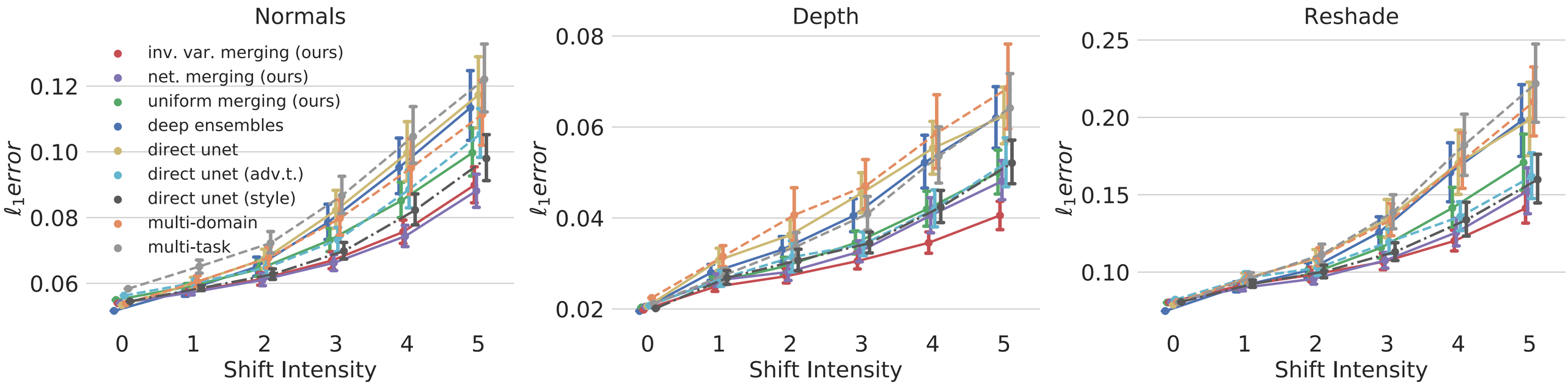
Taskonomy + Common Corruptions

- 11 unseen distortions
- Lower is better



Taskonomy + Common Corruptions

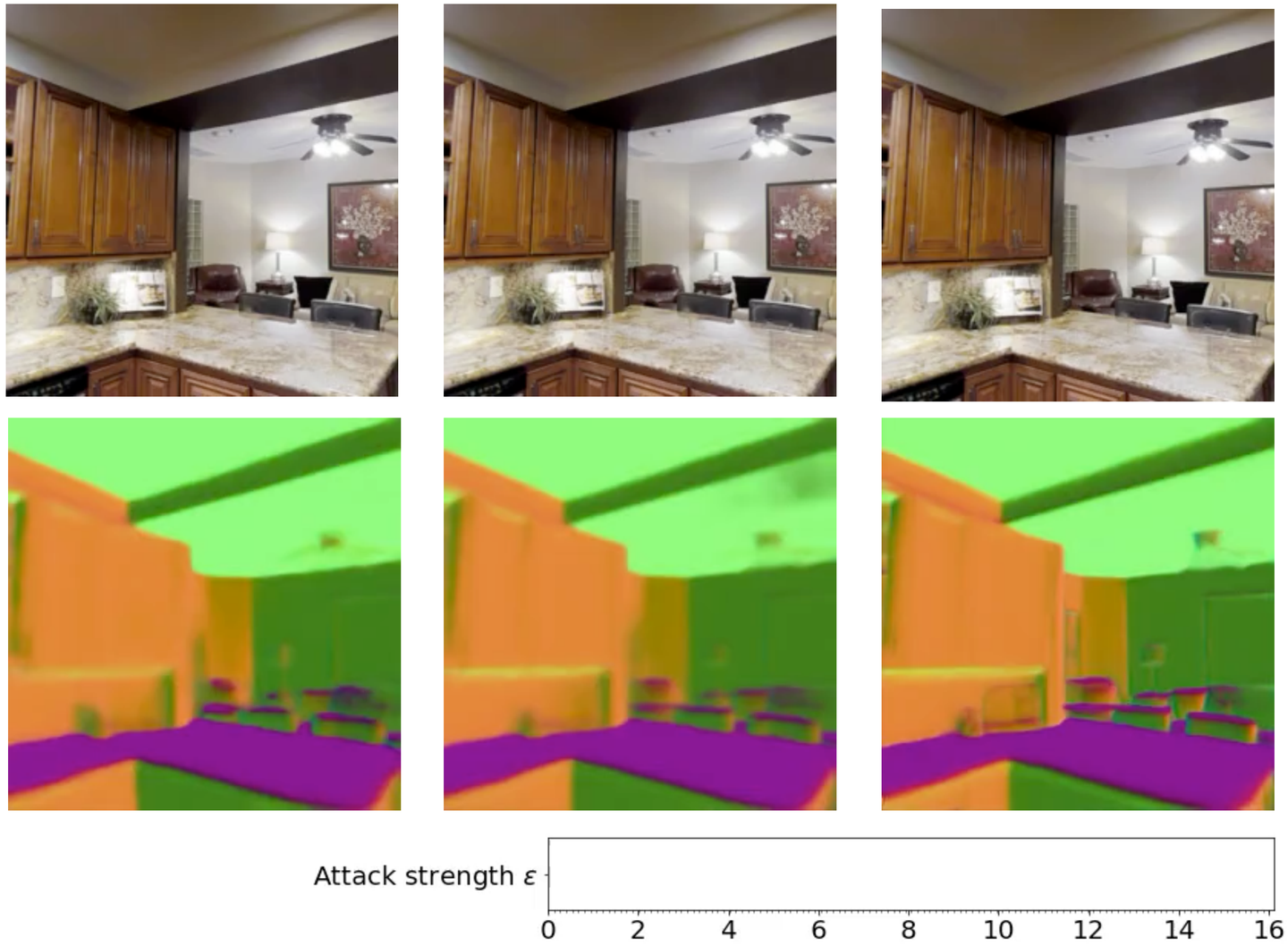
- Our method consistently outperforms the baselines



Taskonomy + adversarial attacks

Image corrupted with increasing I-FGSM attack

Baseline UNet | Deep Ensembles | **Ours**



Taskonomy + adversarial attacks

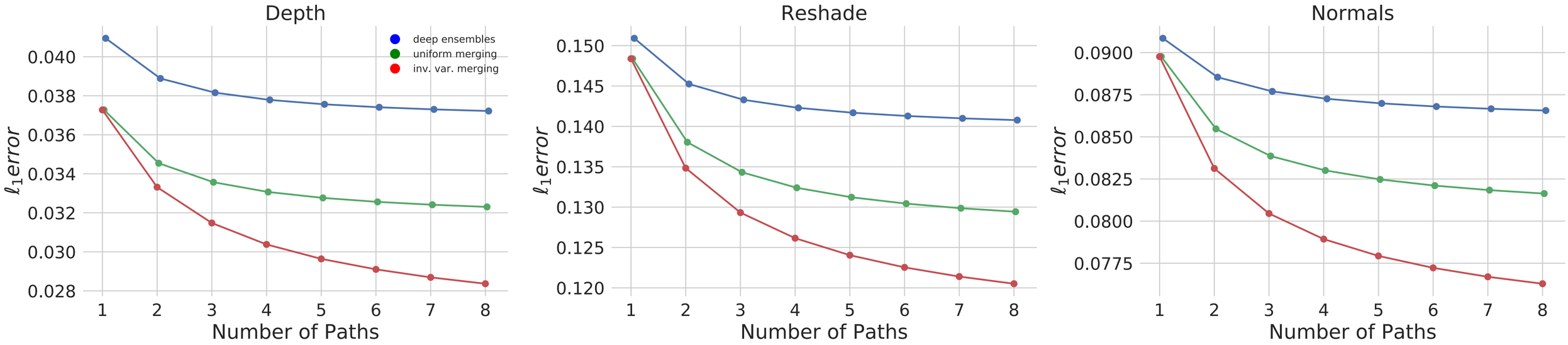
- Lower is better
- Improved robustness against I-FGSM attacks *without adversarial training*
- More challenging to fools all paths simultaneously

	Normal				Reshade				Depth			
Method \ ϵ	2	4	8	16	2	4	8	16	2	4	8	16
Baseline UNet	8.23	11.53	13.03	14.37	17.92	22.78	27.26	34.40	5.50	6.76	8.36	9.80
Deep ensembles	**7.49**	11.13	13.36	15.65	15.66	21.95	27.75	34.98	5.45	6.68	8.27	10.52
Inv. var. merging	7.60	**8.89**	**10.40**	**12.77**	**15.56**	**16.55**	**18.93**	**22.01**	**4.94**	**4.99**	**5.93**	**6.75**

Ablation studies

Effect of increasing number of paths & Role of uncertainty

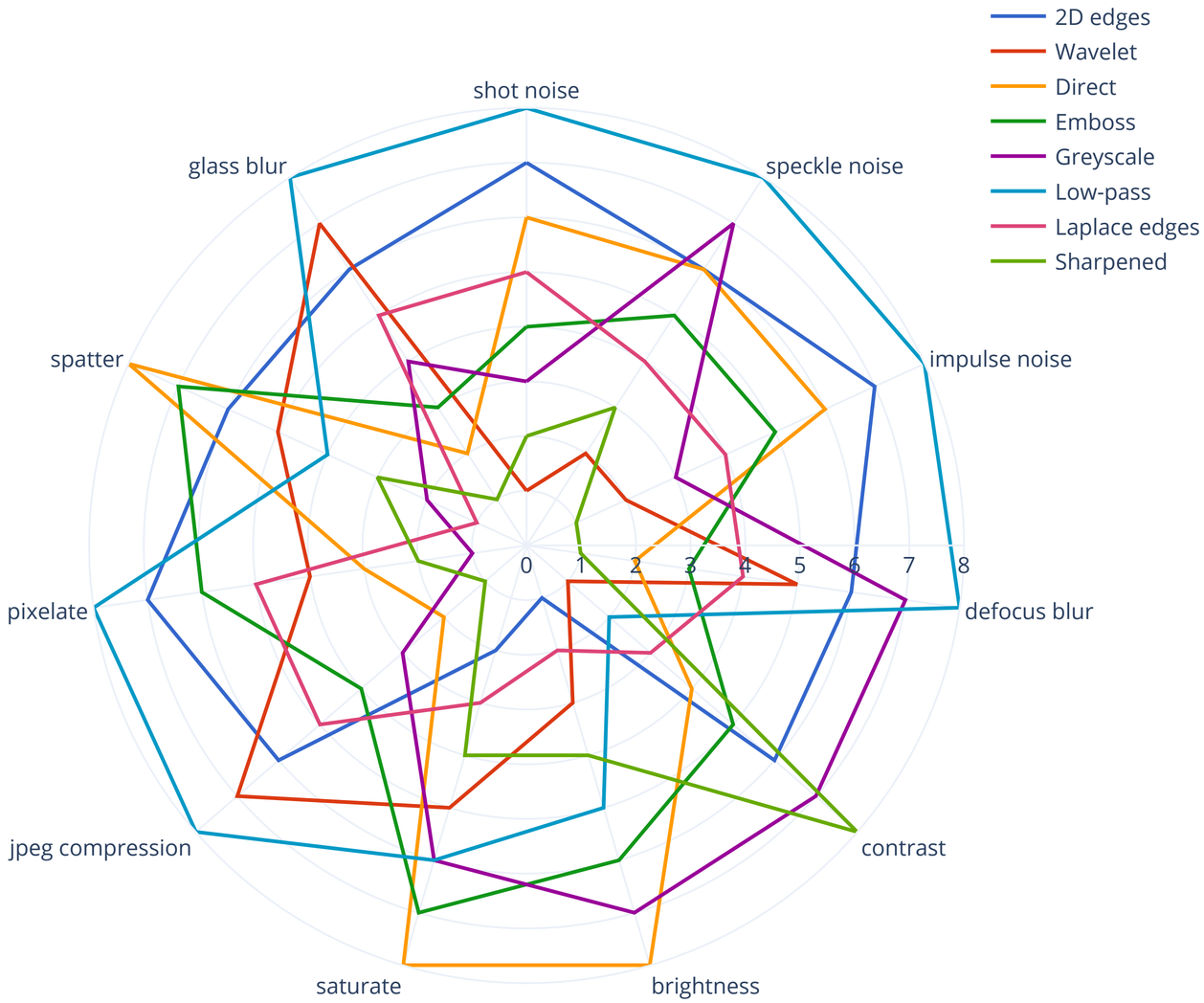
- More paths → larger performance gap
- Using uncertainties as merging weights boosts performance



Ablation studies

Importance of each middle domain

- Order of best performing paths under different distortions
- Most important = 8, least important =1
- *Low-pass:* helpful for *Noise* distortions
- *Sharpened:* helpful for *Contrast* distortion



Imagenet-C

- Our method outperforms deep ensembles by only using middle domains

			Noise			Blur				Weather				Digital			
Method	Clean error	Avg.	Gauss.	Shot	Impulse	Defocus	Glass	Motion	Zoom	Snow	Frost	Fog	Bright	Contrast	Elastic	Pixel	JPEG
Baseline ResNet-50	24.4	76.2	73.0	74.7	78.8	79.9	92.1	81.5	82.5	75.2	75.6	64.0	59.2	65.3	90.6	74.8	75.9
Deep ensembles	**21.5**	70.4	67.4	69.7	72.5	73.4	87.4	76.1	**76.9**	70.3	70.3	60.1	52.4	61.7	83.8	66.4	68.3
Ours	21.6	**67.9**	**66.6**	**68.6**	**71.2**	**71.7**	**82.1**	**75.6**	77.3	**69.1**	**67.2**	**59.1**	**51.3**	**55.8**	**82.1**	**54.4**	**65.9**

Summary

Key Takeaways

- Using middle domains promotes ensemble diversity with a *negligible increase in computational cost*
- The uncertainty based merging select regions from the best performing path
- We improved robustness compared to several baselines under distribution shifts (*common corruptions*, *adversarial attacks)* for *classification* and *regression* tasks
- Furthermore, improvements in robustness does not sacrifice performance on in-distribution data

Robustness via Cross-Domain Ensembles

Teresa Yeo*, Oğuzhan Fatih Kar*, Amir Zamir

EPFL

crossdomain-ensembles.epfl.ch