

---

# Supplementary Material of Neural Inverse Rendering for General Reflectance Photometric Stereo

---

Tatsunori Taniai<sup>1</sup> Takanori Maehara<sup>1</sup>

<sup>1</sup> RIKEN AIP, Nihonbashi, Tokyo, Japan

In this supplementary material, we provide results for all the ten benchmark scenes that are omitted in the paper due to the page limitation. Specifically, we provide the following additional results.

- A. Visual comparisons of our method and existing methods.
- B. Our image reconstruction results.
- C. Convergence analyses of our method by different types of weak supervision.

## A. Additional visual comparisons

In Figures A1–A10, we show surface normal estimates and their angular error maps for all ten scenes from the DiLiGenT dataset, comparing our method with three methods by Santo et al. (2017), Shi et al. (2014), and Ikehata & Aizawa (2014), and also the baseline least squares method. Here, the best result for each scene is obtained by either of the top four methods (where 8 of 10 best results are by our method), whose mean angular error is shown by a bold font number.

## B. Additional image reconstruction results

In Figures A11–A20, we show our image reconstruction results for all ten scenes from the DiLiGenT dataset. For each scene we show 6 images from 96 observed images, and for each observed image we show our synthesized image, intermediate reflectance image, and reconstruction error map. Here, the reflectance images represent multiplication of spatially-varying BRDFs and cast shadows under a particular illumination condition. We can clearly see cast shadows in reflectance images appearing in the results of BEAR and BUDDHA. Note that for better visualization, the image intensities are scaled by a factor of  $255/2$  after the proposed global scaling normalization.

## C. Additional convergence analyses

In Figures A21–A30, we show convergence analyses for all ten scenes from the DiLiGenT dataset, where the proposed early-stage supervision is compared with no and all-stage supervision. Training without supervision is very unstable, while training with all-stage supervision is strongly biased by inaccurate least square priors. Note that for better comparison, the median profiles of the proposed early-stage weak supervision (red solid lines) are overlaid in the plots of no and all-stage supervision.

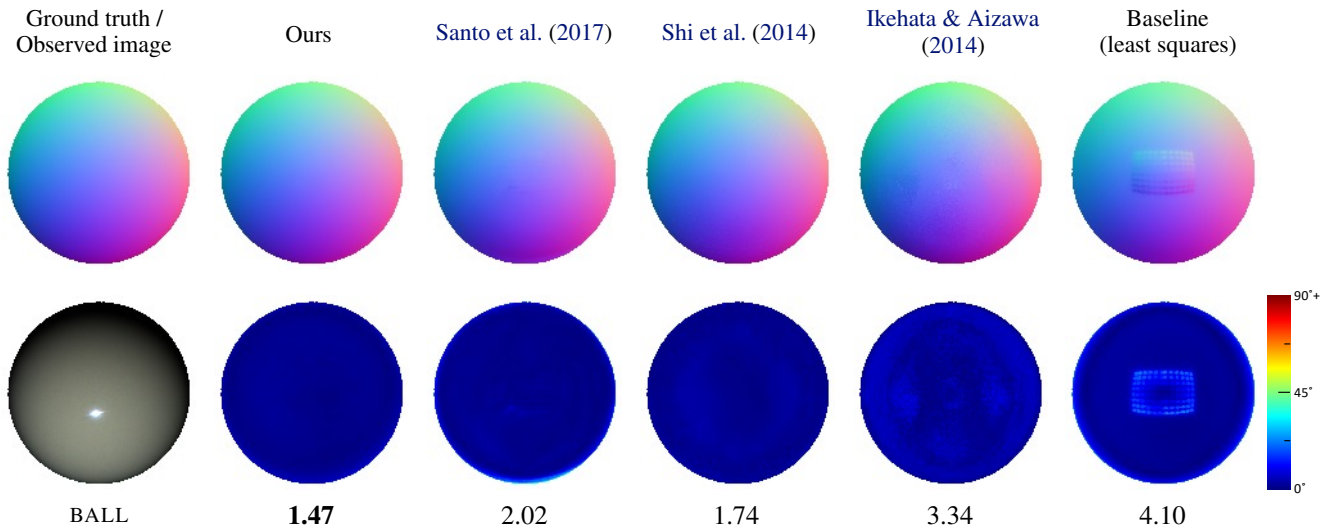


Figure A1. **Visual comparisons for BALL scene.** From left to right columns, we show 1) the ground truth normal map and an observed image, 2) our normal estimate and its error map, 3–6) four pairs of a normal estimate and its error map by three state-of-the-art methods (Santo et al., 2017; Shi et al., 2014; Ikehata & Aizawa, 2014) and the baseline least squares method.

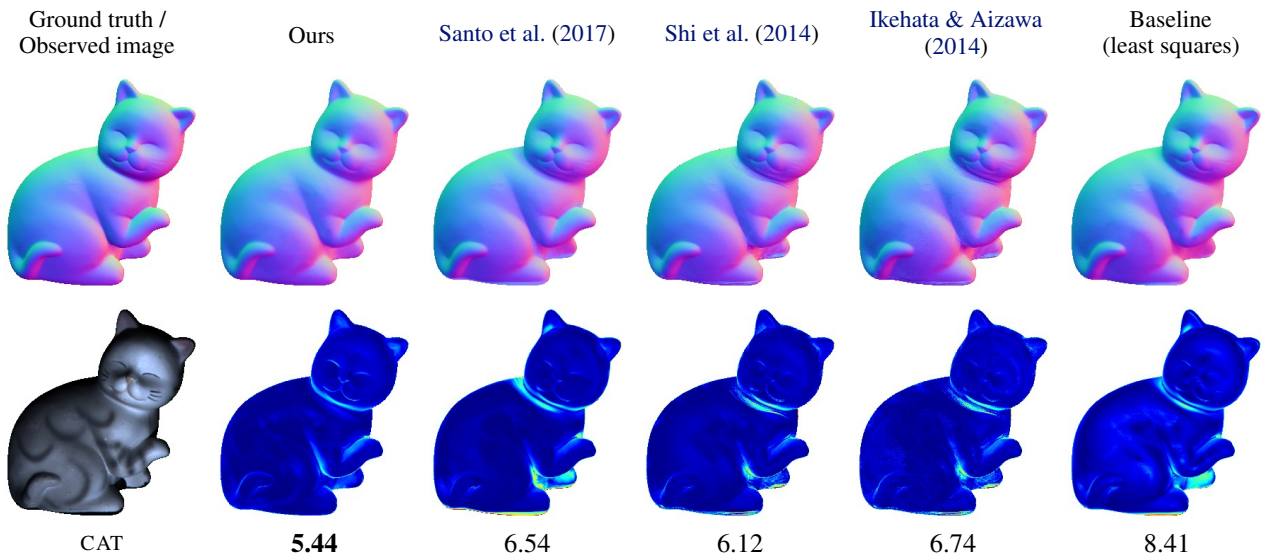


Figure A2. **Visual comparisons for CAT scene.** See also explanations in Fig. A1.

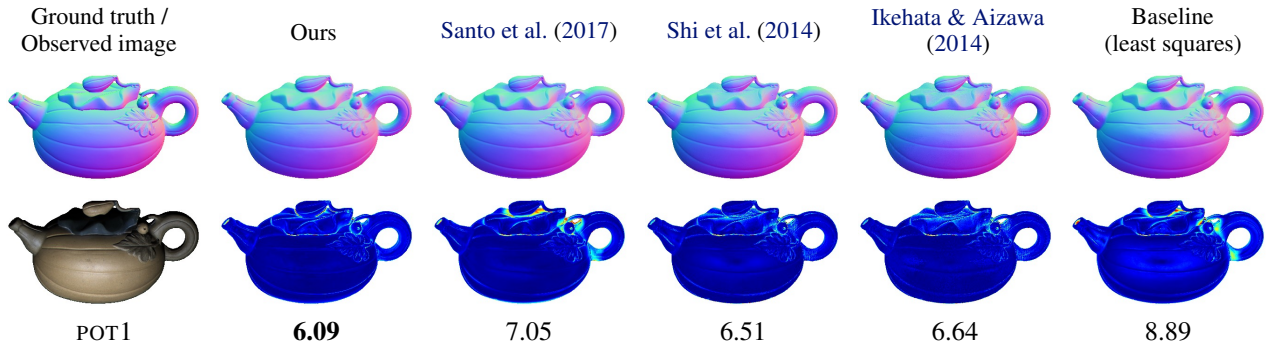


Figure A3. Visual comparisons for POT1 scene. See also explanations in Fig. A1.

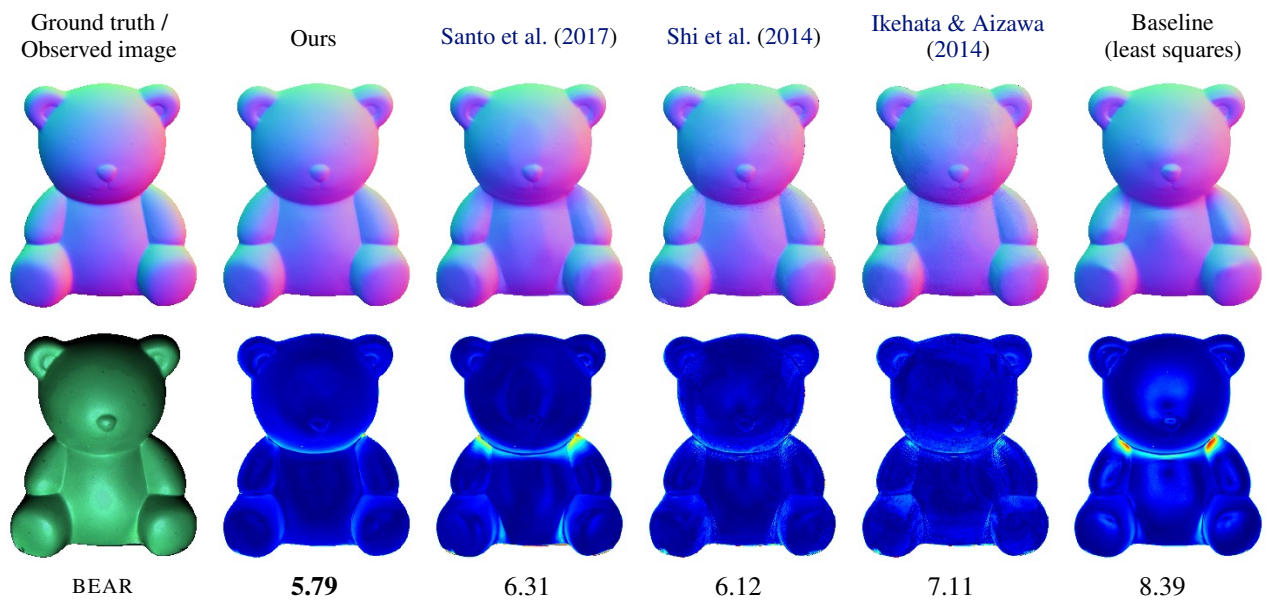


Figure A4. Visual comparisons for BEAR scene. See also explanations in Fig. A1.

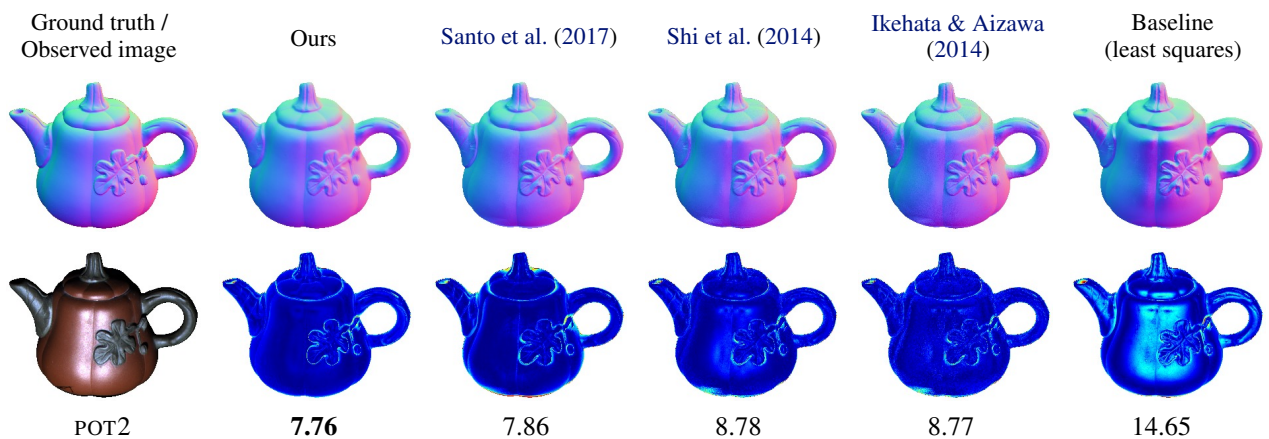


Figure A5. Visual comparisons for POT2 scene. See also explanations in Fig. A1.



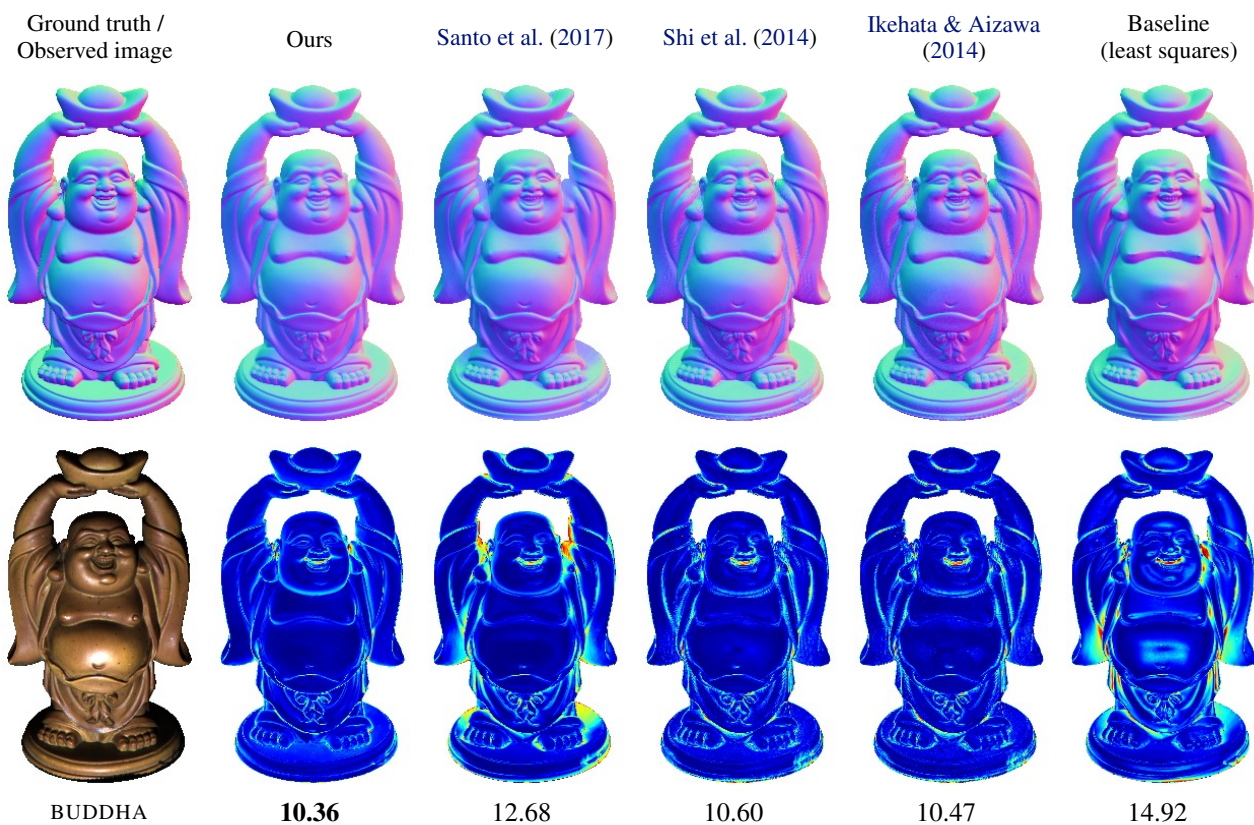


Figure A6. Visual comparisons for BUDDHA scene. See also explanations in Fig. A1.

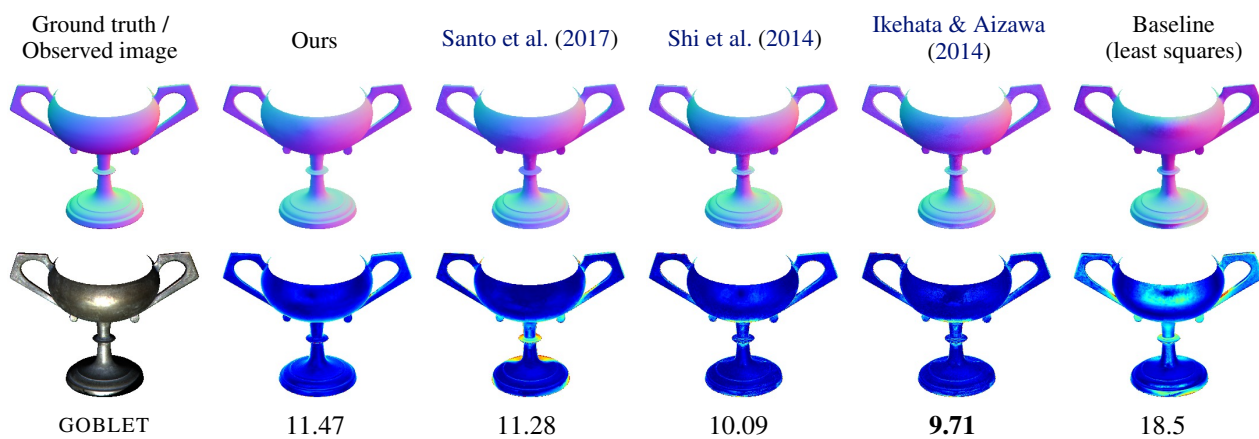


Figure A7. Visual comparisons for GOBLET scene. See also explanations in Fig. A1.

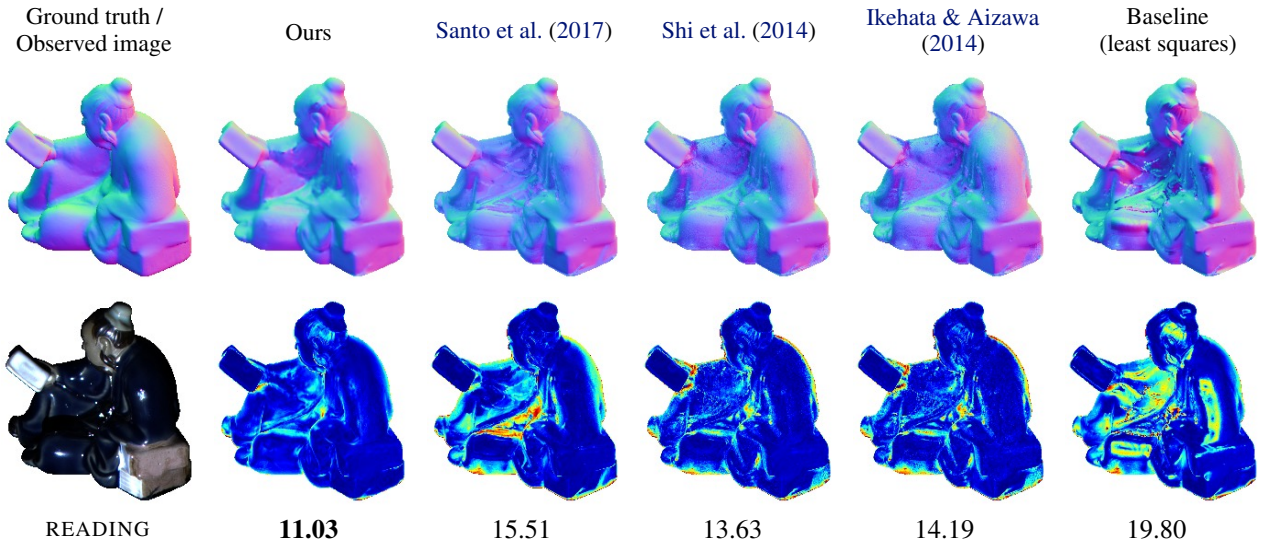


Figure A8. Visual comparisons for READING scene. See also explanations in Fig. A1.

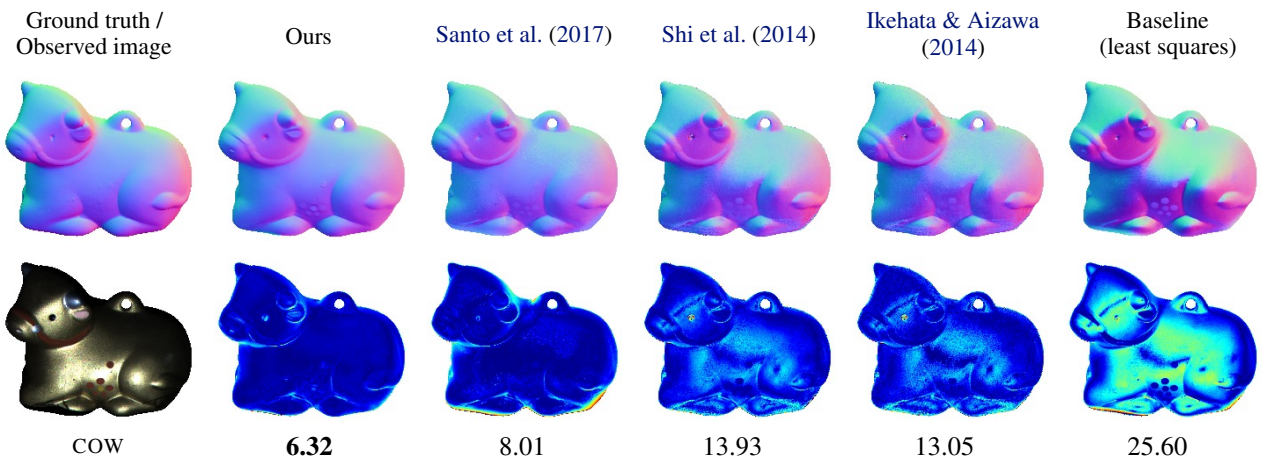


Figure A9. Visual comparisons for COW scene. See also explanations in Fig. A1.

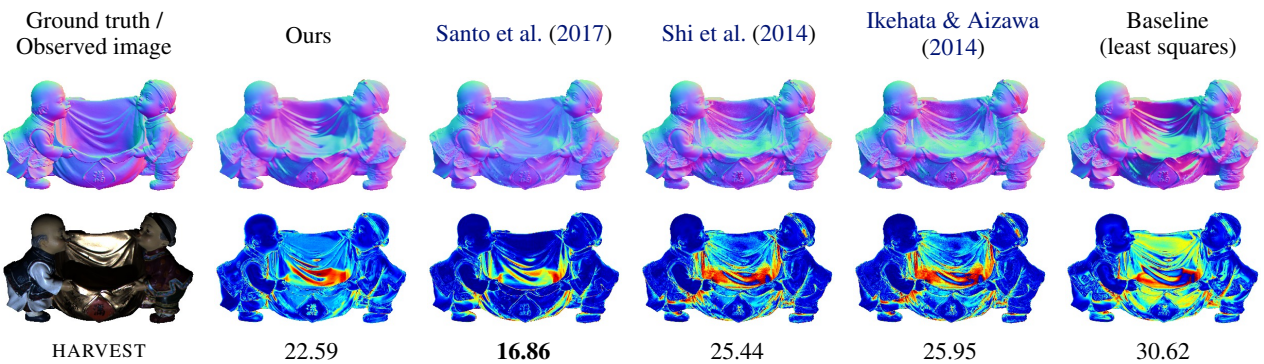


Figure A10. Visual comparisons for HARVEST scene. See also explanations in Fig. A1.

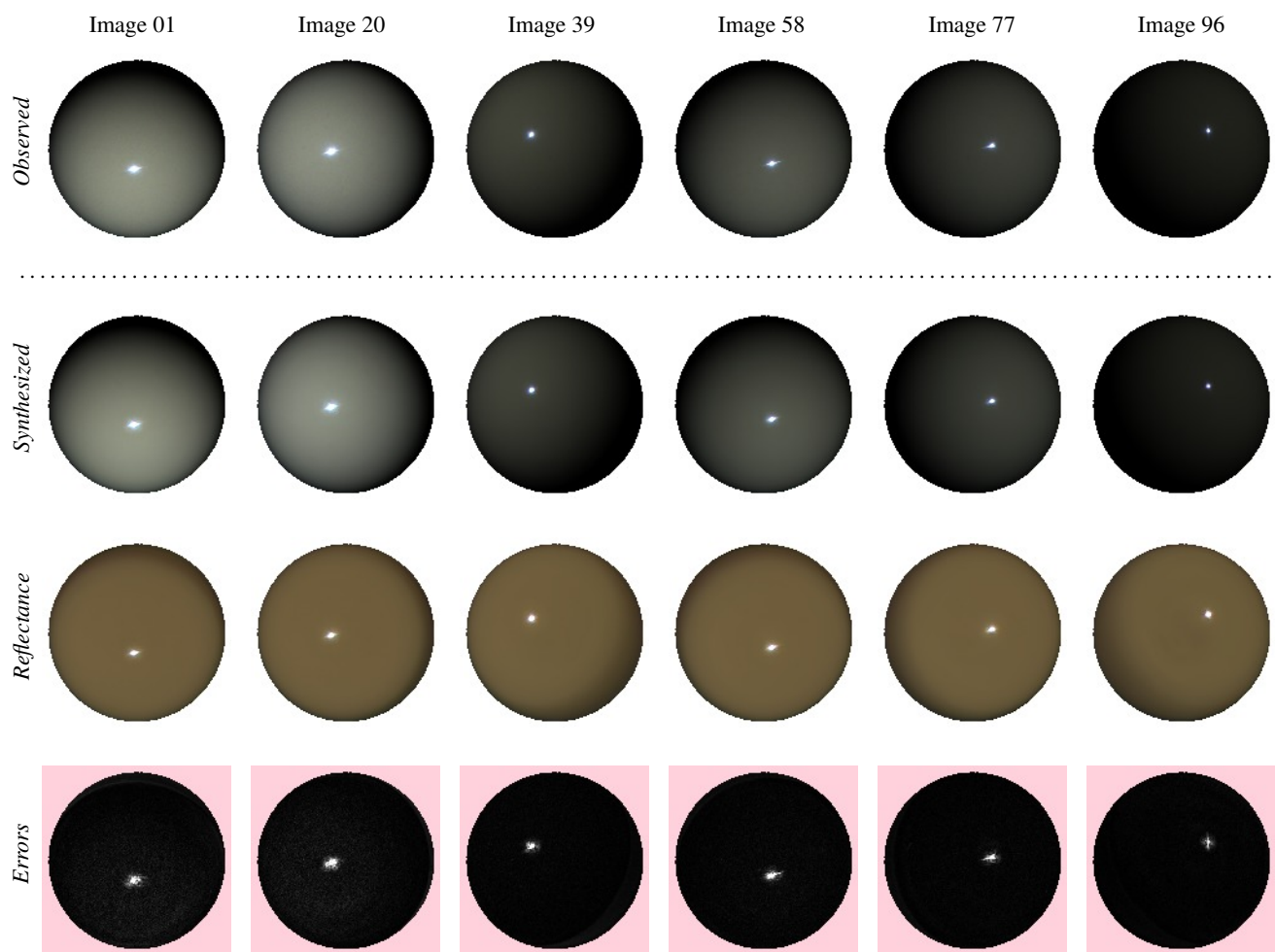


Figure A11. **Our image reconstruction results for BALL scene.** We show our image reconstruction results for six selected images. From top to bottom, we show observed images, our synthesized images, our reflectance images, and reconstruction errors between observed and synthesized images. The error maps show absolute intensity differences scaled by a factor of 10.



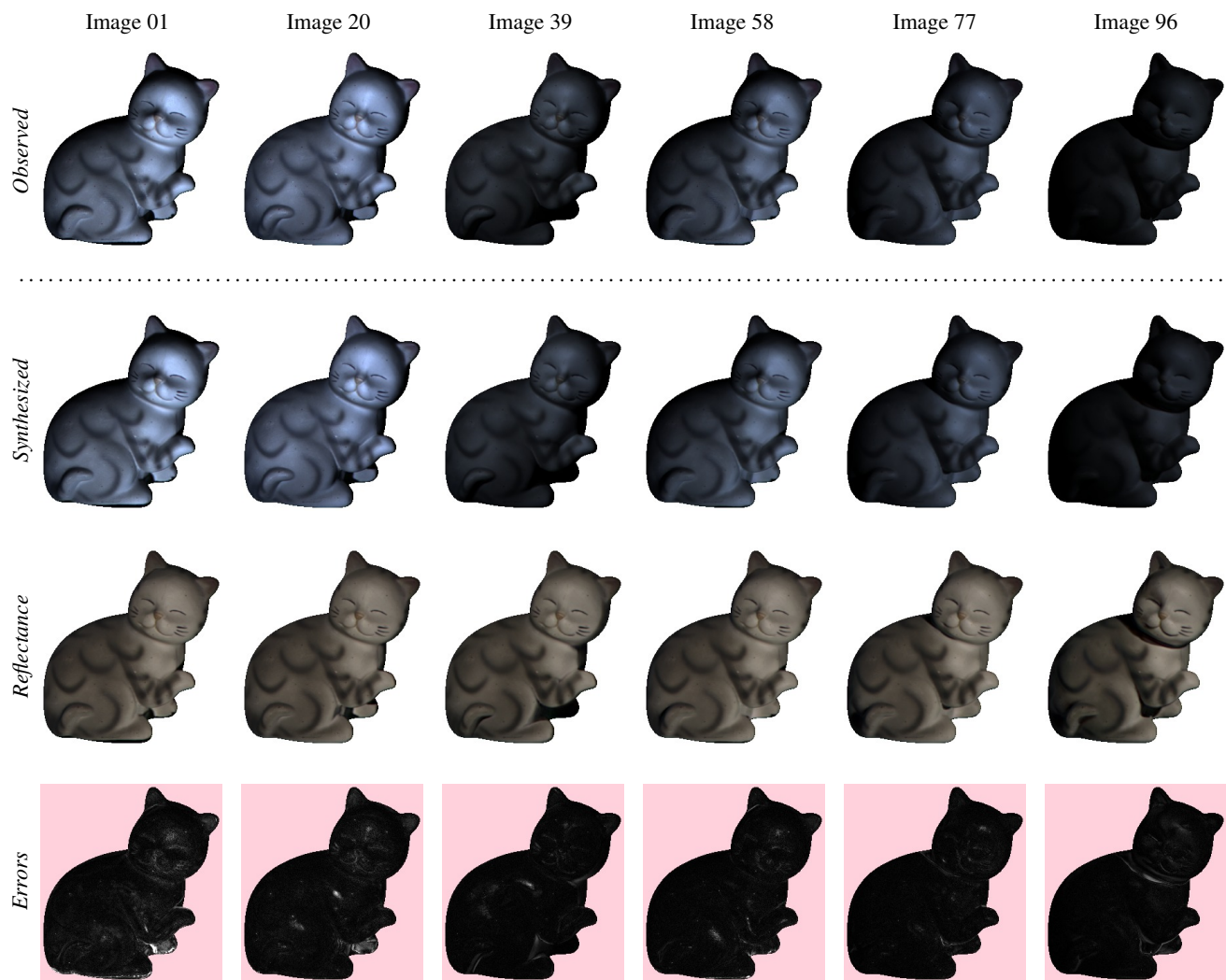


Figure A12. Our image reconstruction results for CAT scene. See also explanations in Fig. A11.



Figure A13. Our image reconstruction results for POT1 scene. See also explanations in Fig. A11.



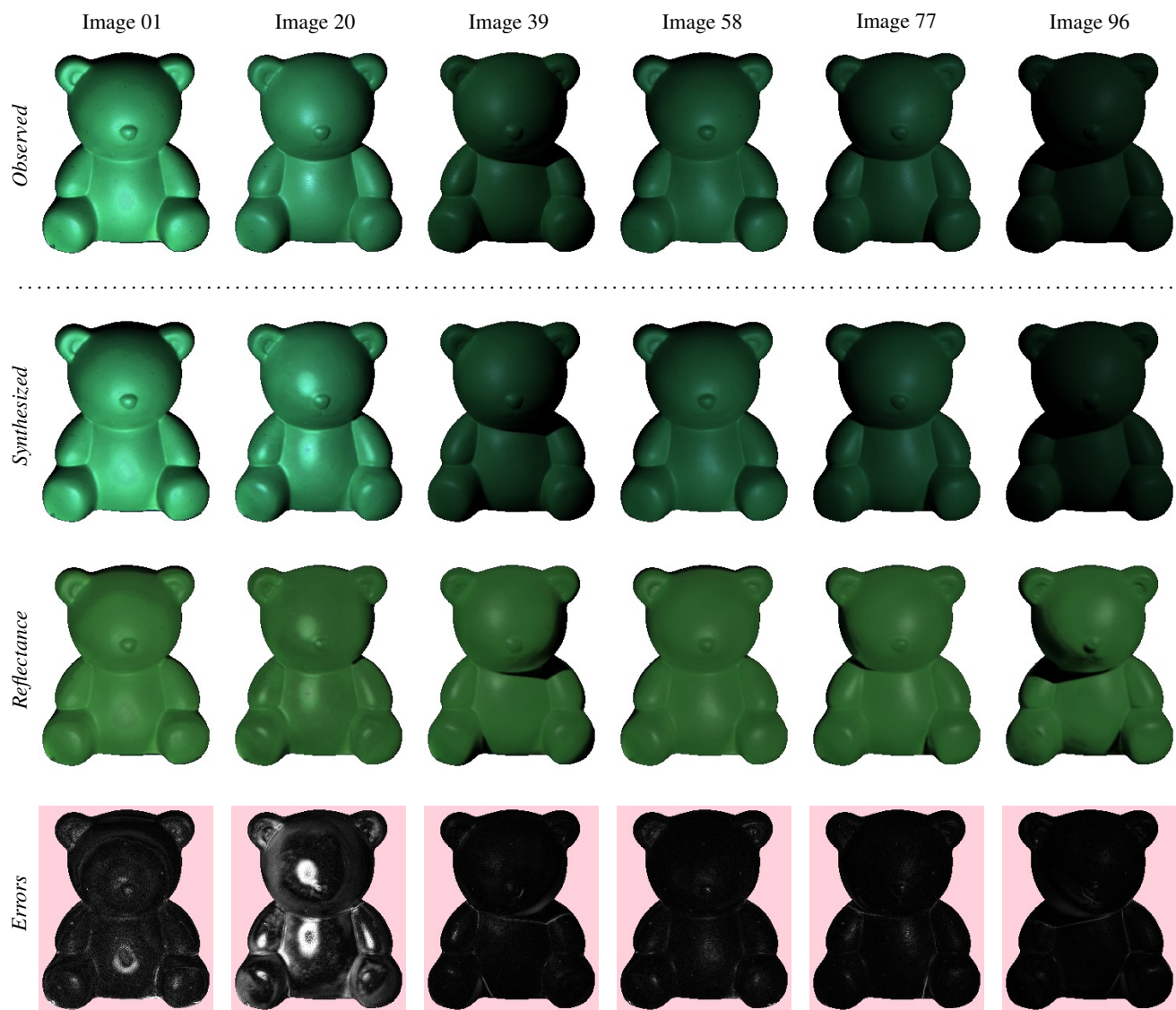


Figure A14. Our image reconstruction results for BEAR scene. See also explanations in Fig. A11.

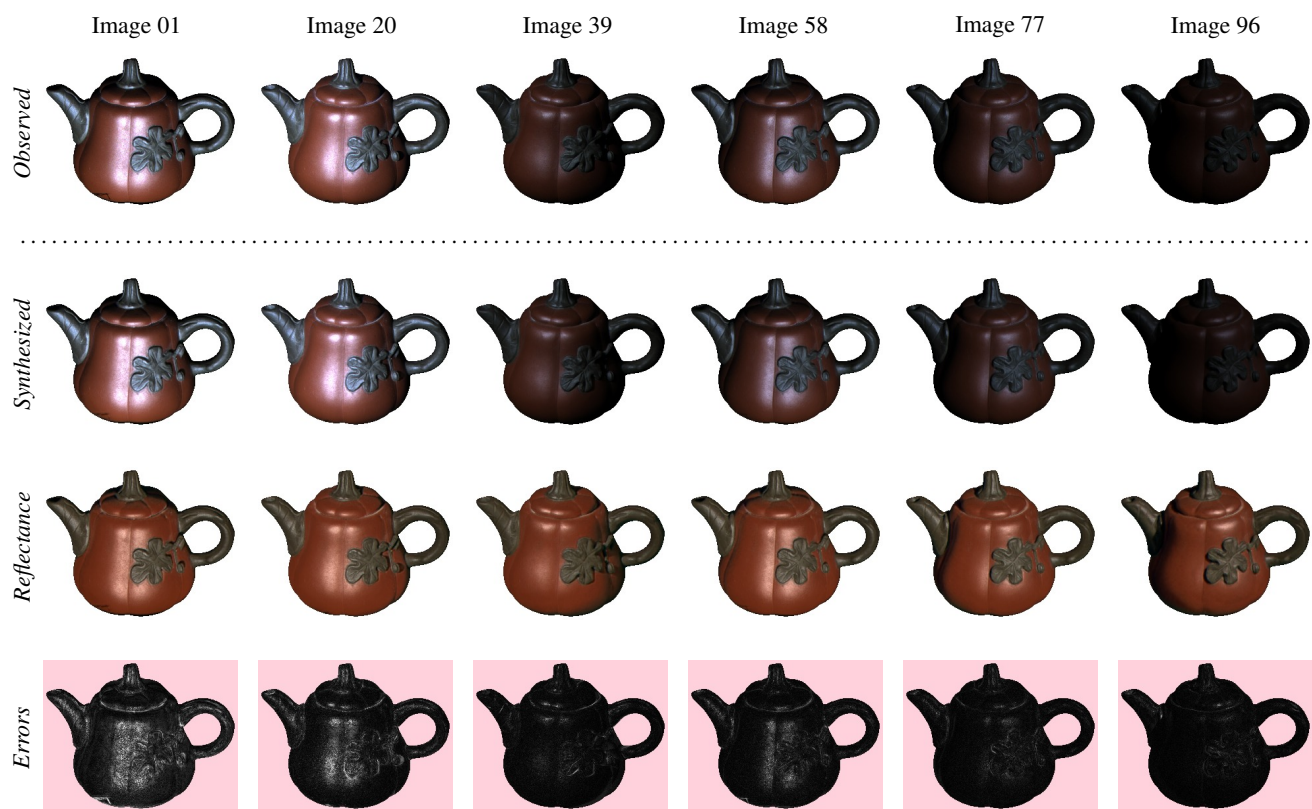


Figure A15. Our image reconstruction results for POT2 scene. See also explanations in Fig. A11.



Figure A16. Our image reconstruction results for BUDDHA scene. See also explanations in Fig. A11.





Figure A17. Our image reconstruction results for GOBLET scene. See also explanations in Fig. A11.

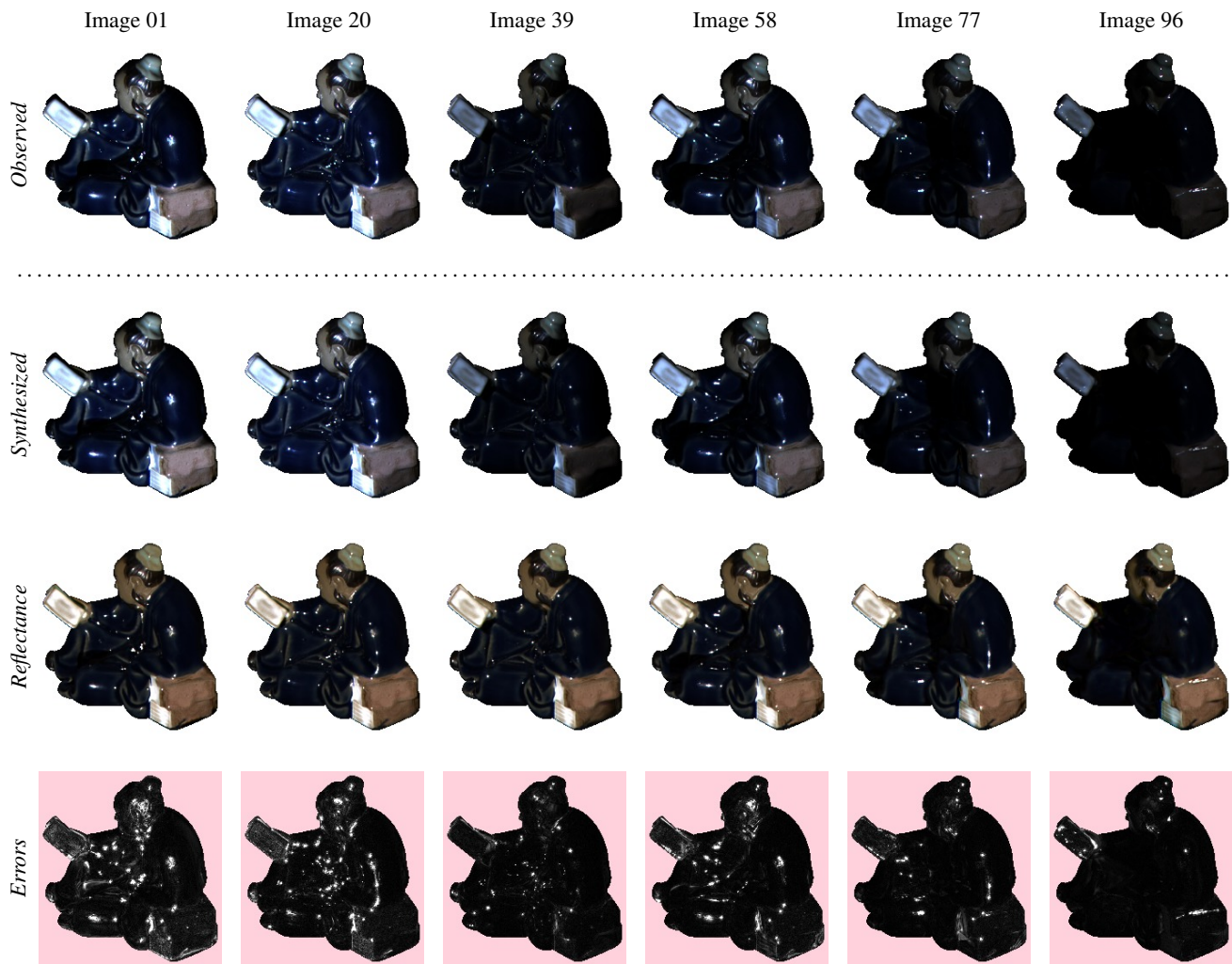


Figure A18. Our image reconstruction results for READING scene. See also explanations in Fig. A11.

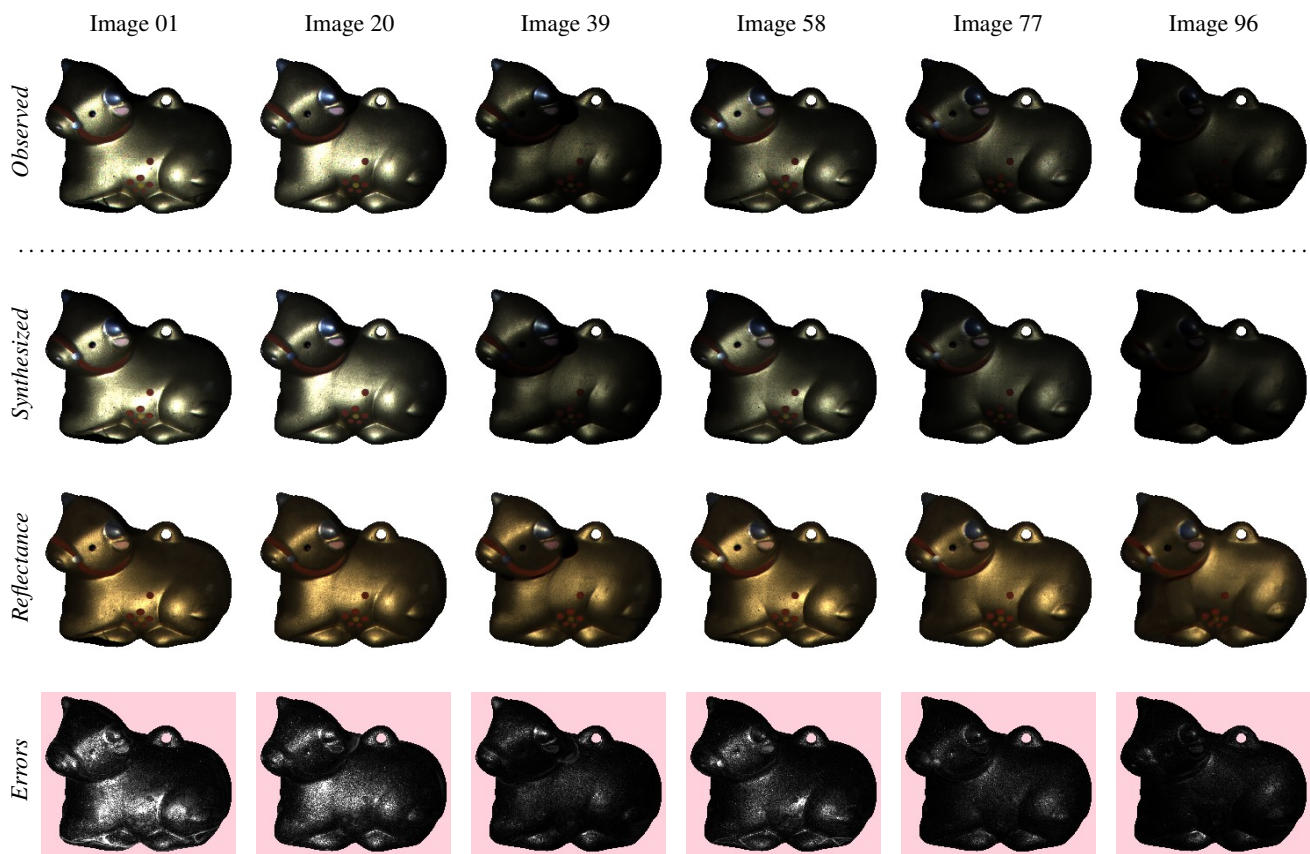


Figure A19. Our image reconstruction results for COW scene. See also explanations in Fig. A11.



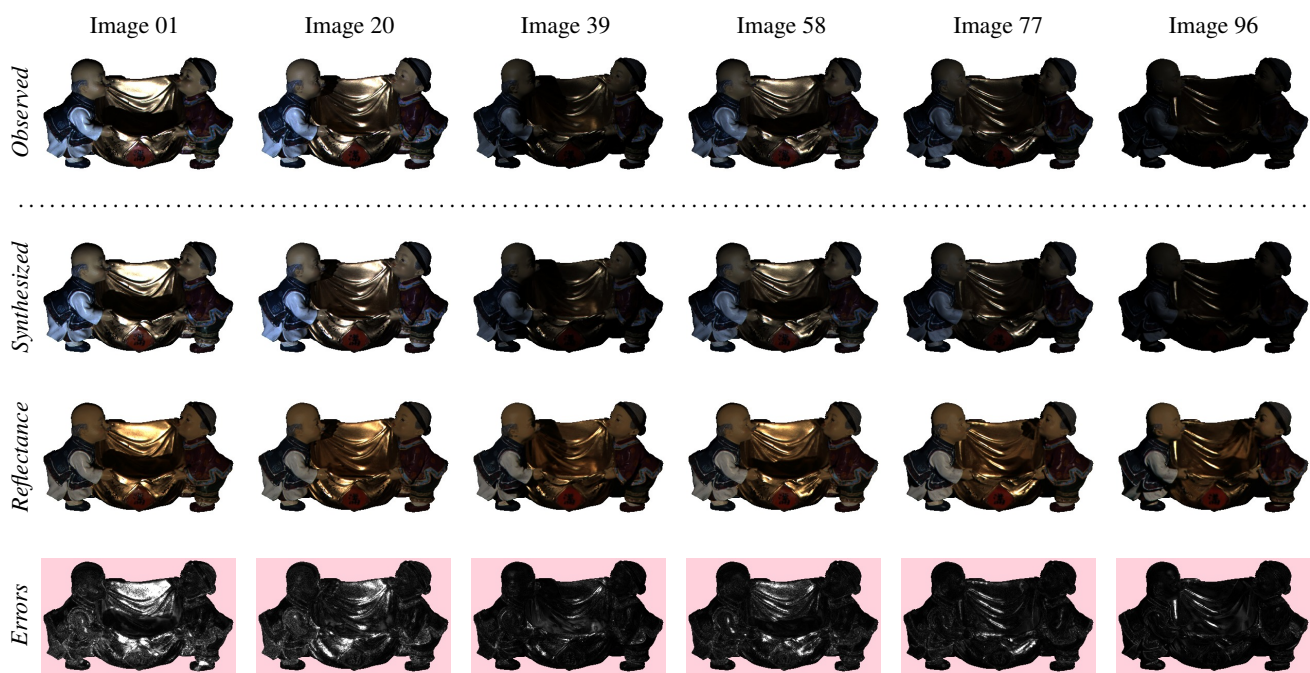


Figure A20. Our image reconstruction results for HARVEST scene. See also explanations in Fig. A11.

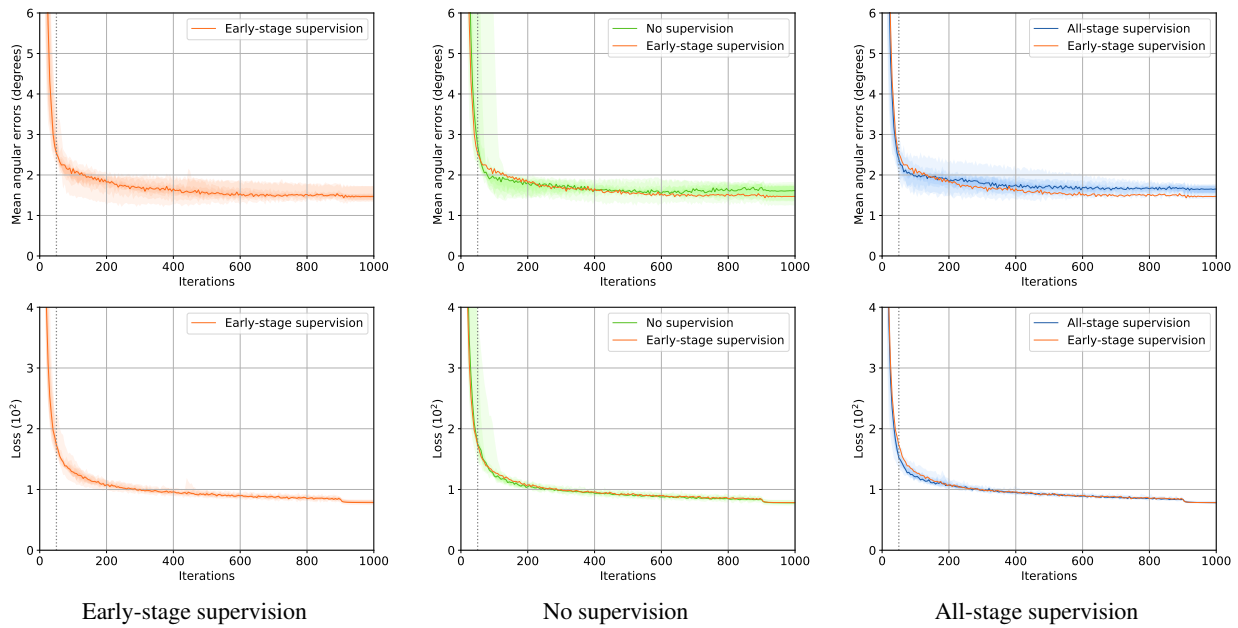


Figure A21. Convergence analysis with different types of weak supervision for BALL scene. We compare training curves by using the proposed early-stage supervision (left) with no supervision (middle), and all-stage supervision (left). For each case, profiles of mean angular errors and loss values are shown in top and bottom, respectively, which are visualized by distributions of 11 rounds run (colored region) and medians (solid line). Vertical lines at  $t = 50$  show termination of early-stage supervision.

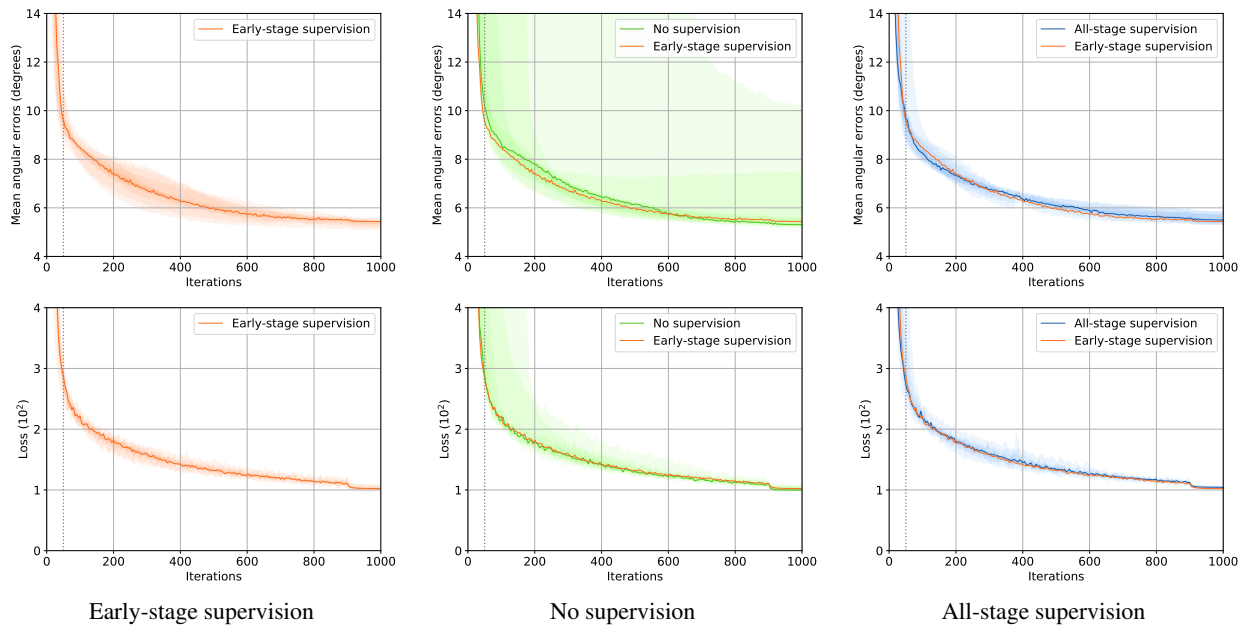


Figure A22. Convergence analysis with different types of weak supervision for CAT scene. See also explanations in Fig. A21.

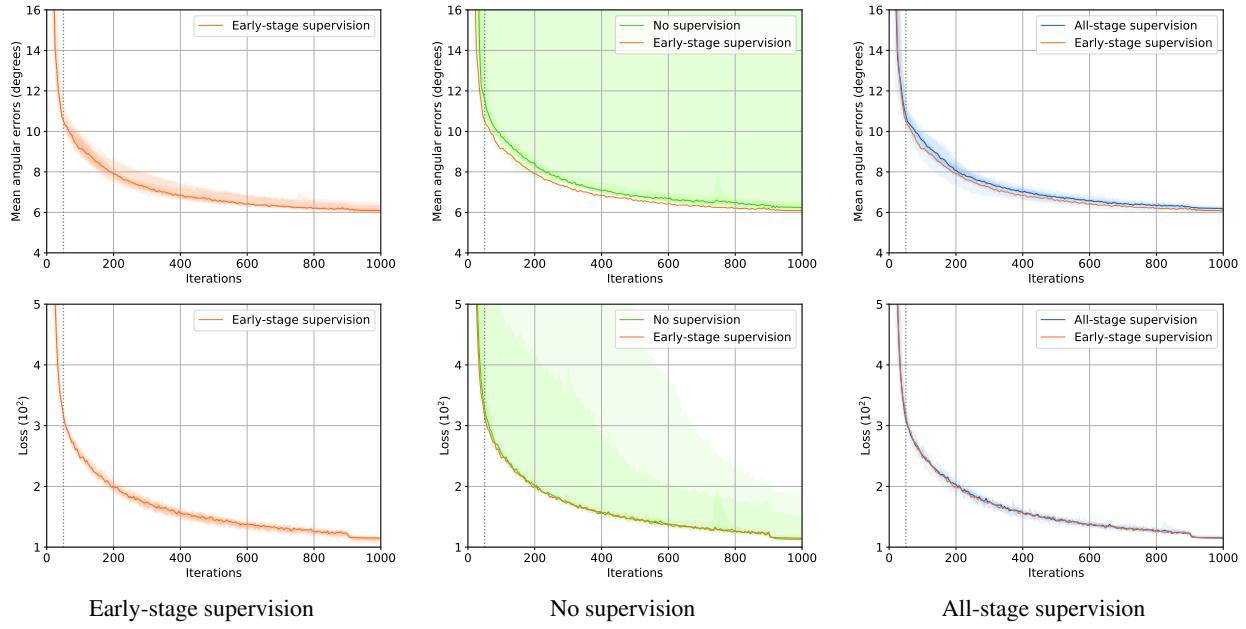


Figure A23. Convergence analysis with different types of weak supervision for POT1 scene. See also explanations in Fig. A21.

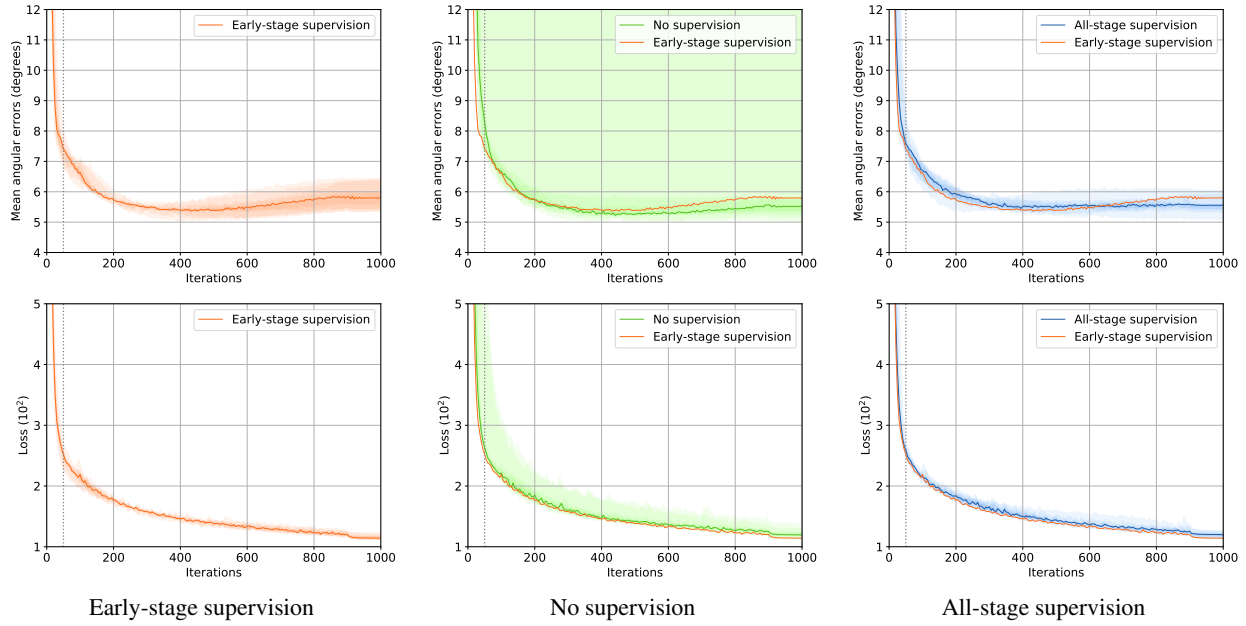


Figure A24. Convergence analysis with different types of weak supervision for BEAR scene. See also explanations in Fig. A21.



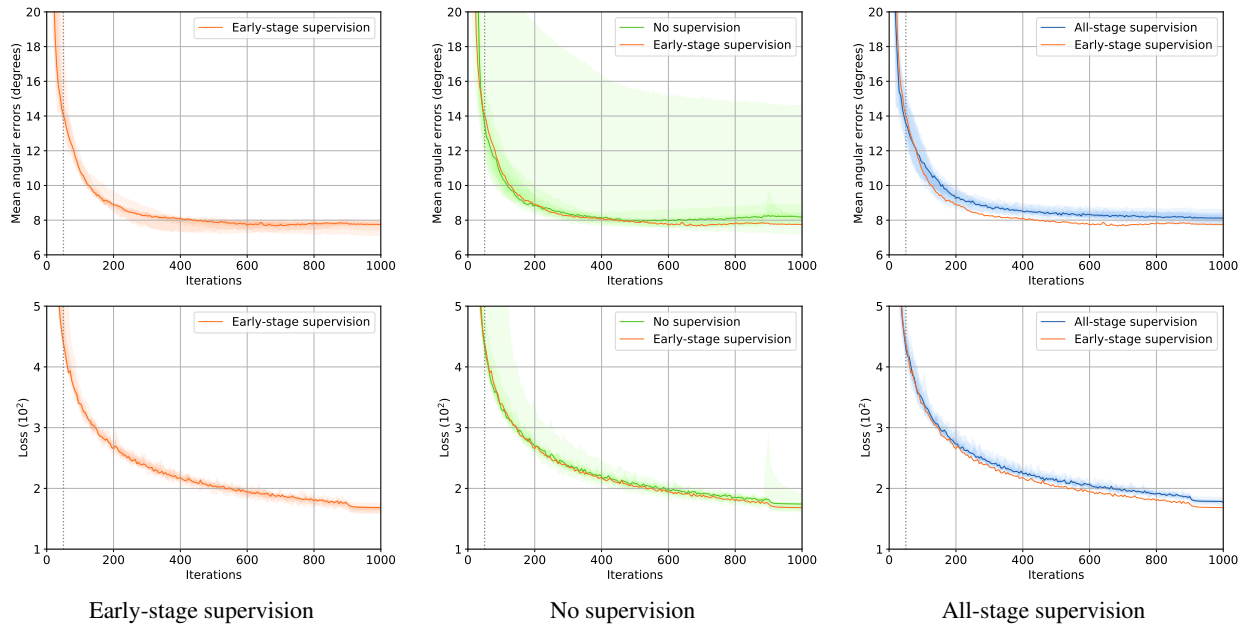


Figure A25. Convergence analysis with different types of weak supervision for POT2 scene. See also explanations in Fig. A21.

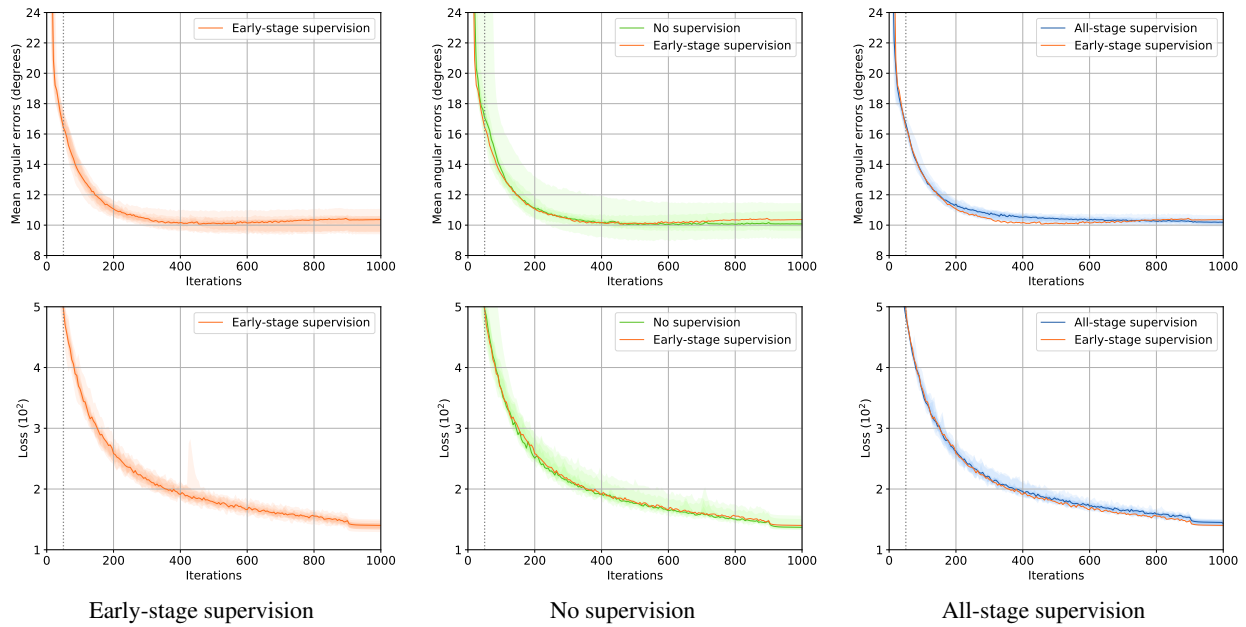


Figure A26. Convergence analysis with different types of weak supervision for BUDDHA scene. See also explanations in Fig. A21.

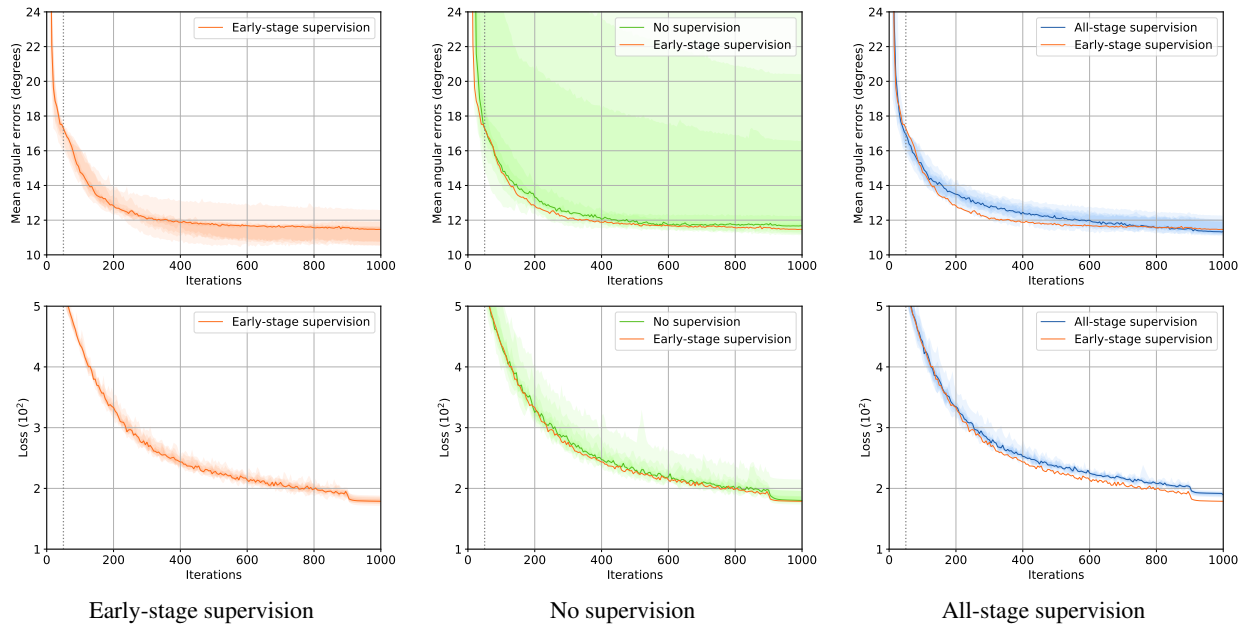


Figure A27. Convergence analysis with different types of weak supervision for GOBLET scene. See also explanations in Fig. A21.

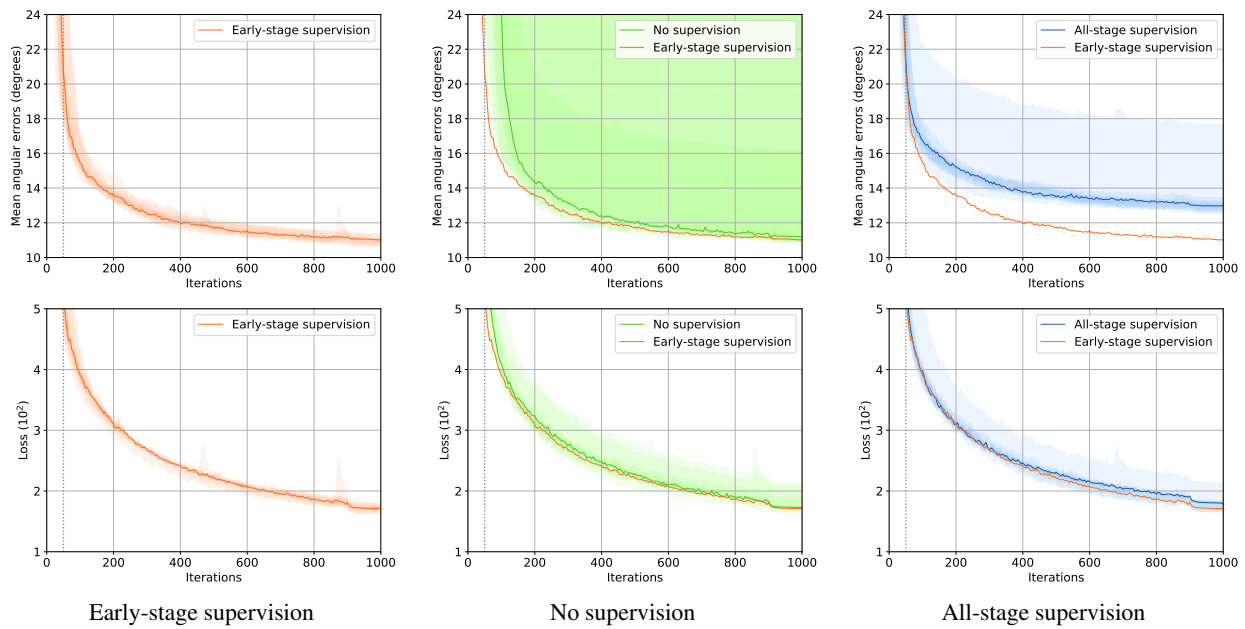


Figure A28. Convergence analysis with different types of weak supervision for READING scene. See also explanations in Fig. A21.

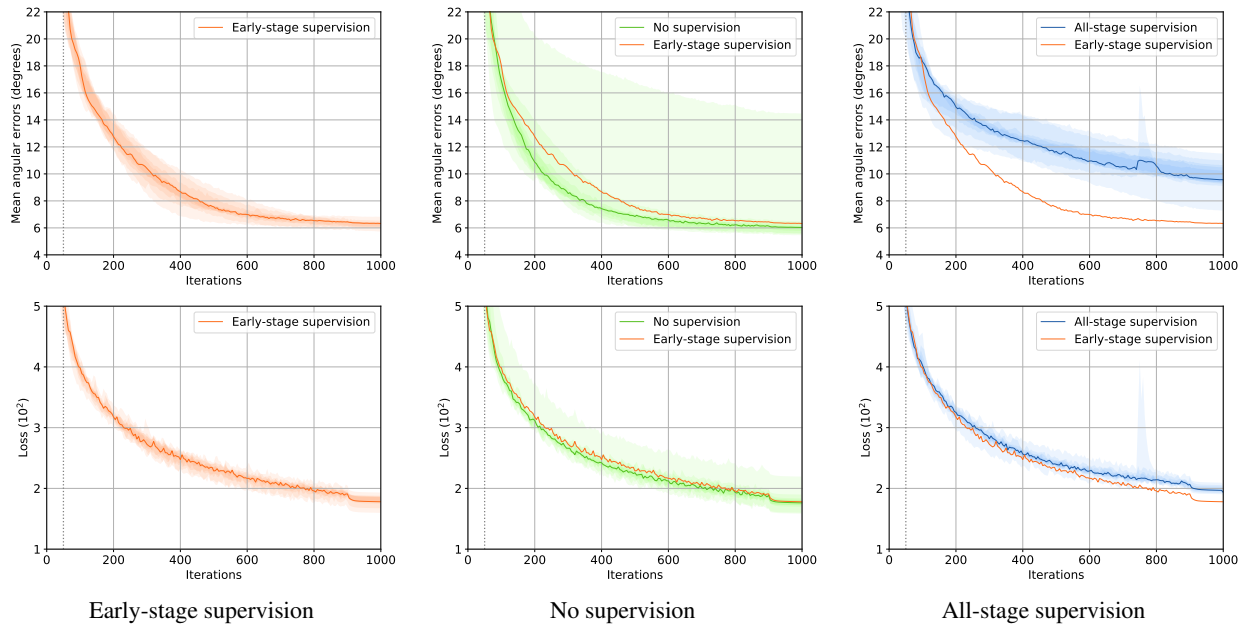


Figure A29. Convergence analysis with different types of weak supervision for COW scene. See also explanations in Fig. A21.

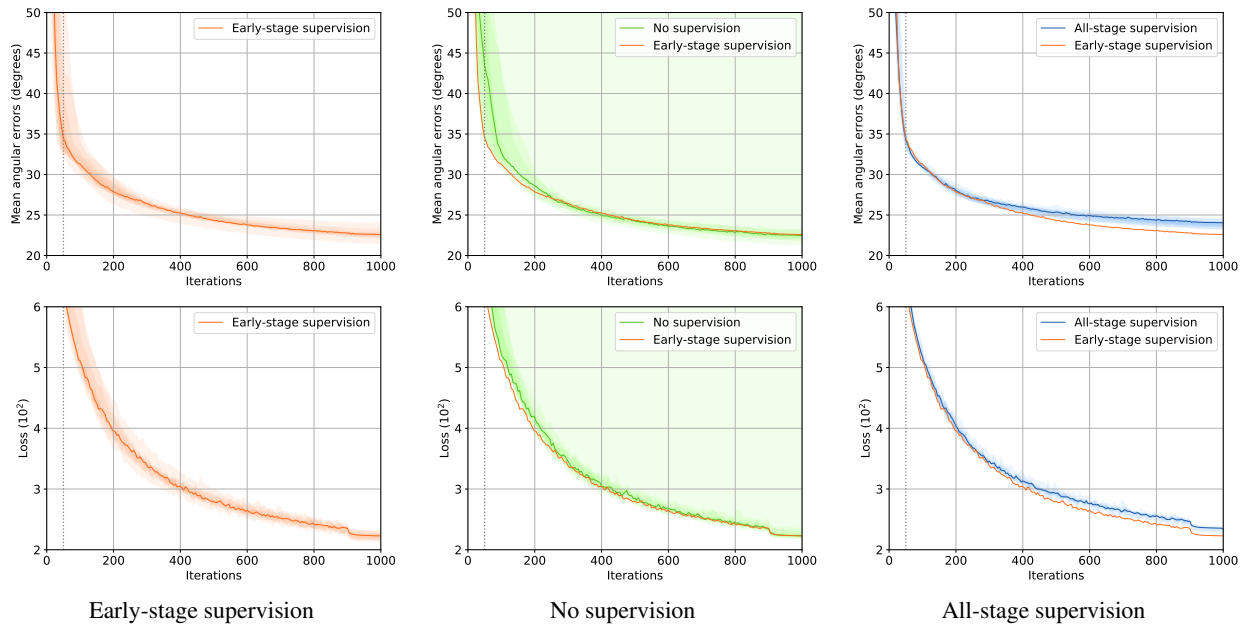


Figure A30. Convergence analysis with different types of weak supervision for HARVEST scene. See also explanations in Fig. A21.