Domain Decluttering: Simplifying Images to Mitigate Synthetic-Real Domain Shift and Improve Depth Estimation

Yunhan Zhao Shu Kong Daeyun Shin Charless Fowlkes

UCIrvine

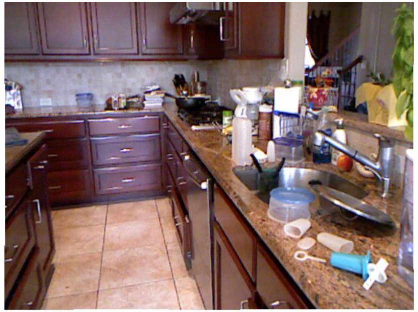
Carnegie Mellon University
The Robotics Institute

Leveraging Synthetic Data in Depth Prediction

Motivation & Goal

- Existing methods focus on translating images from synthetic-to-real, hoping to close low-level domain gap (*e.g.*, color & texture).
- We address the high-level domain gap, such as real-world clutter and novel objects absent in synthetic training data

real images

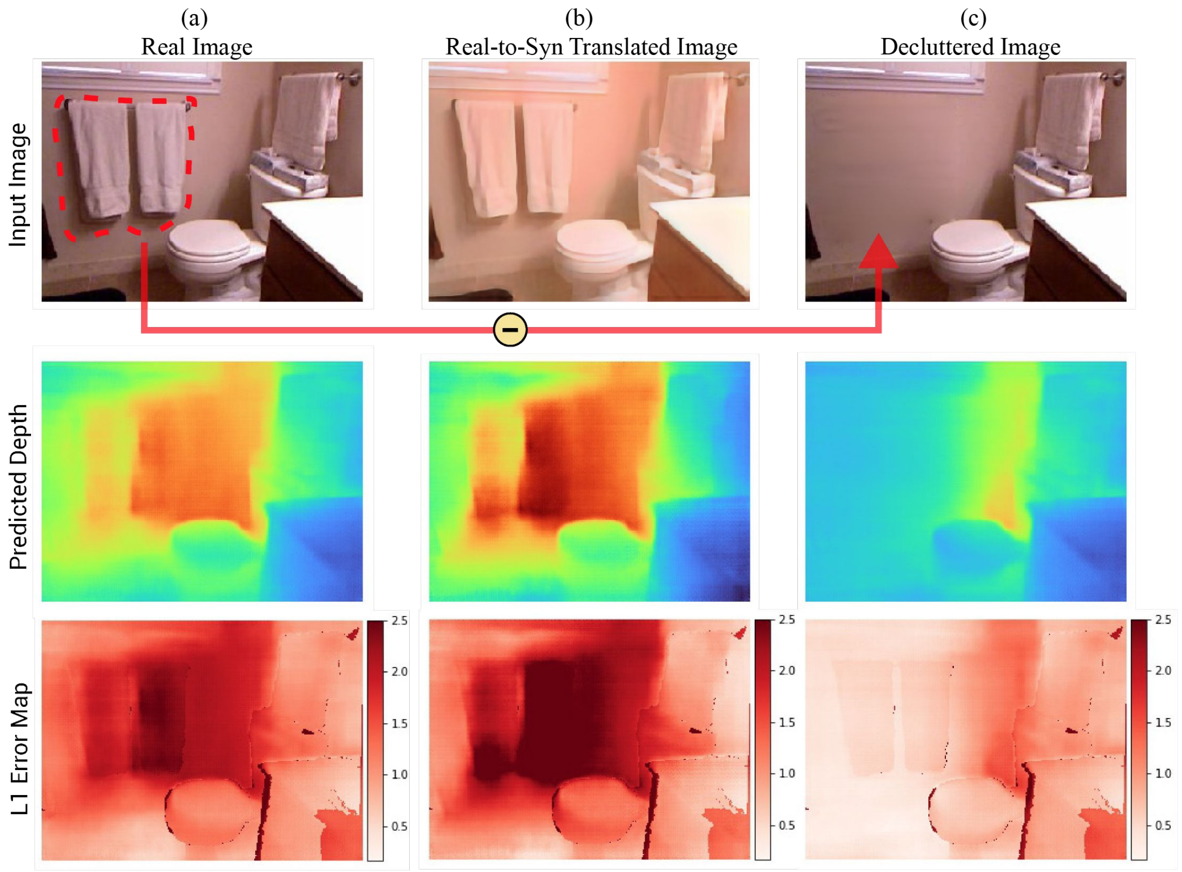


synthetic images

Philosophy - "Admit what you don't understand"

- *Decluttering*: learn to remove and inpaint "clutter" in real images.
- Real-to-synthetic translation of *decluttered* images to leverage model trained on synthetic data.

Robustness to Clutters and Novel Objects



Depth predictor...

(a) struggles on an image with "*clutter*", *e.g.*, towel as a novel object shown here.

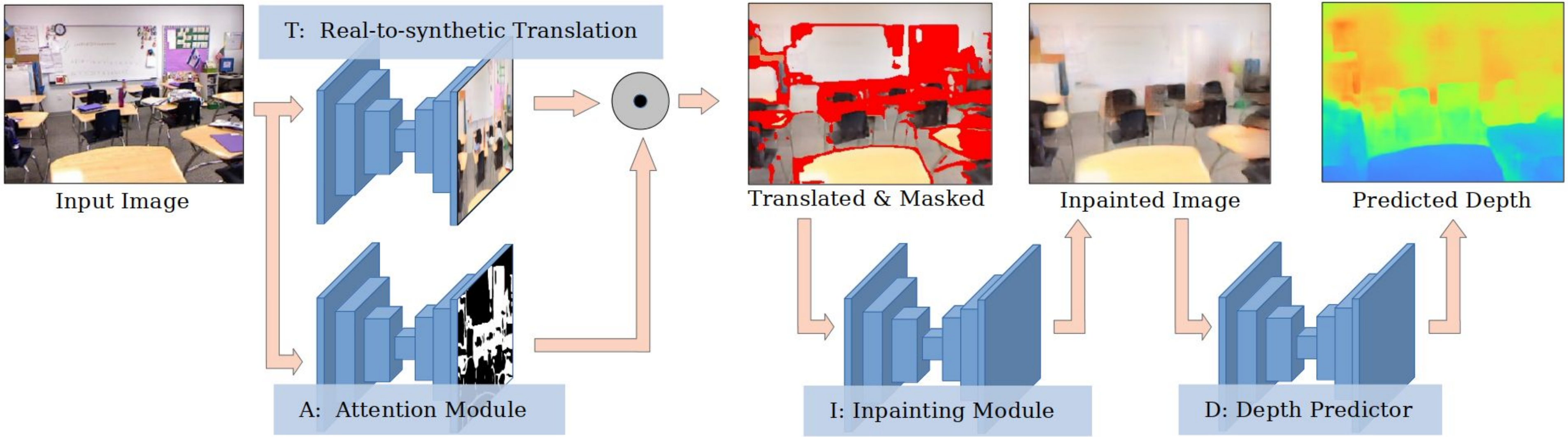
(b) may perform worse on a real-to-syn translated version, although translator and depth predictor are trained over large-scale synthetic data.

(c) produces much better depth estimate on the *decluttered* image, even though original regions are modified!

The Proposed Method: Attend-Remove-Complete (*ARC*)

We train the *ARC* model that can automatically ...

- ***Attend*** to the "cluttered regions" with module-A and remove them
- ***Complete*** these regions with module-I
- ***Translate*** images from real to synthetic with module-T
- ***Predict*** depth with module-D

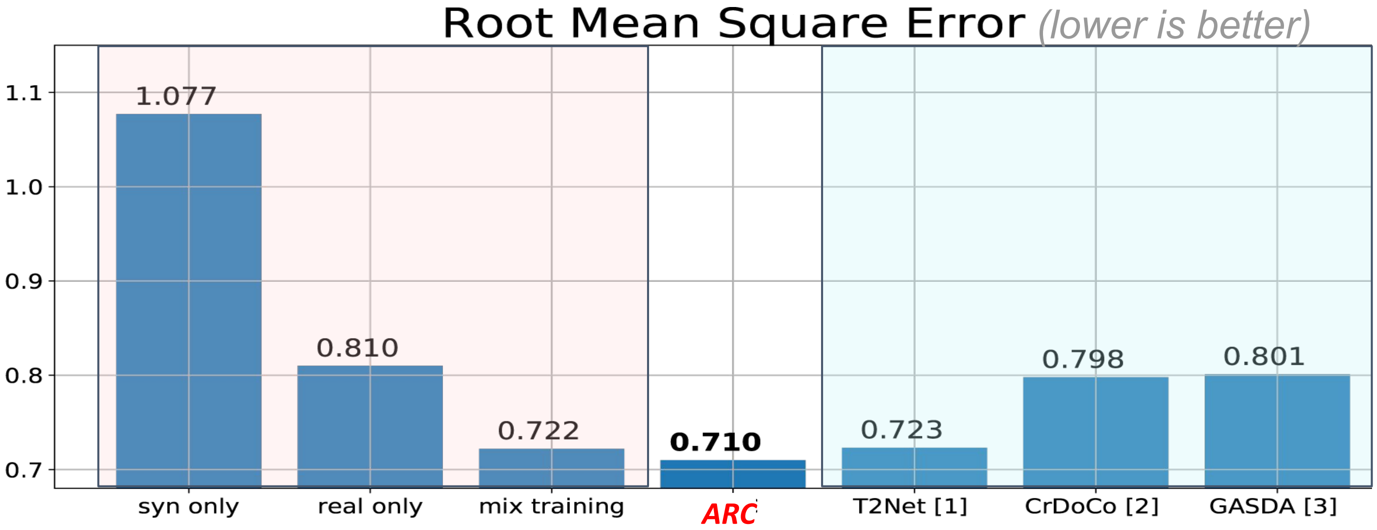


Experiment Snippet: *ARC* performs the best.

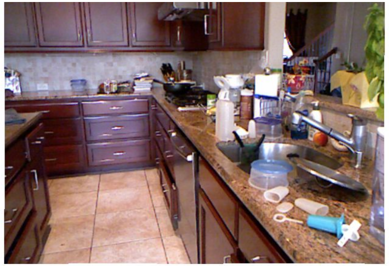
training set:
- ➢ 500 real images
- ➢ 5,000 synthetic images

testing set: 1,449 real images

Baselines:
- ➢ *syn only*: train with 5,000 synthetic images
- ➢ *real only*: train with 500 real images
- ➢ *mix training*: train with all above real&syn data



real images

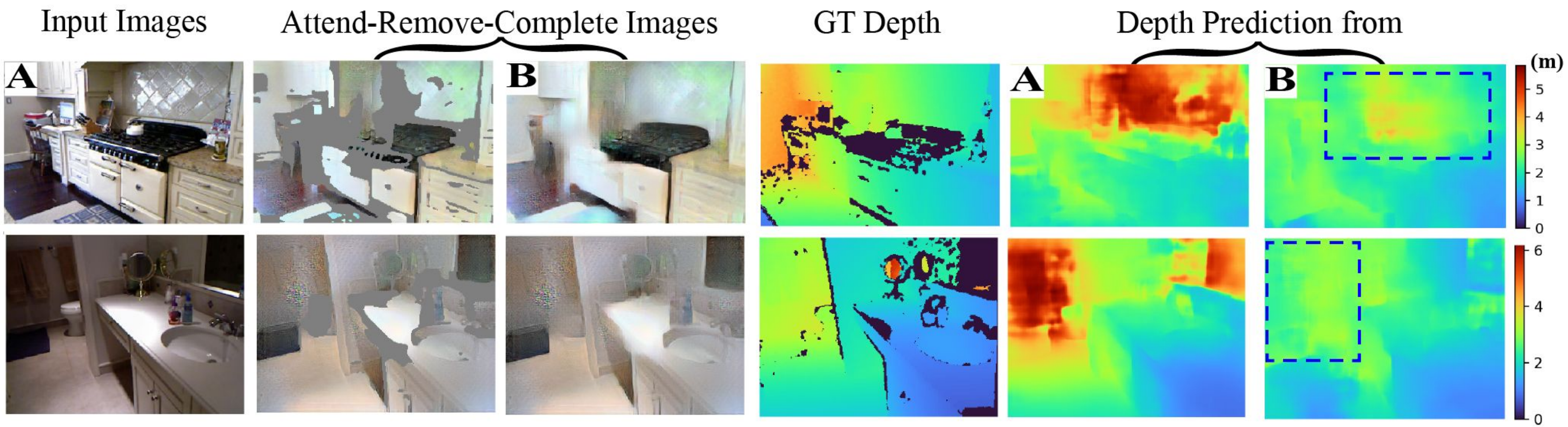


synthetic images

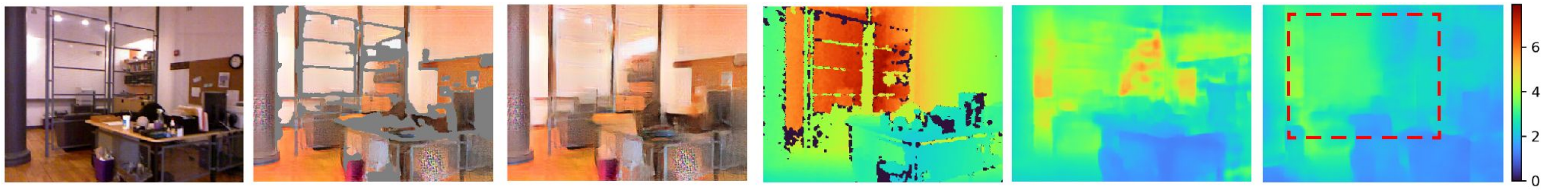
[1] Zheng et al. T2net: Synthetic-to-realistic translation for solving single-image depth estimation tasks. ECCV 2018
[2] Chen et al. Crdoco: Pixel-level domain transfer with crossdomain consistency. CVPR 2019
[3] Zhao et al. Geometry-aware symmetric domain adaptation for monocular depth estimation. CVPR 2019

Experiment Snippet: Qualitative Evaluation

- Visual improvements are visible in blue regions.



- Failure case happens with noticeable ambiguity, *e.g.*, glass in the red region.



Conclusions

- Depth-prediction models are not robust to novel objects and clutters.
- *ARC* avoids some of the failures by actively ignoring scene content it wasn't trained on.
- Previous domain-adaptation-by-translation methods are beneficial when no ground-truth is available for real images. But low-level adaptation is not helpful when some small amount of real-image supervision is available.

project website

Paper: https://arxiv.org/abs/2002.12114

Acknowledgements: This research was supported by NSF grants IIS-1813785, IIS-1618806, a research gift from Qualcomm, and a hardware donation from NVIDIA.