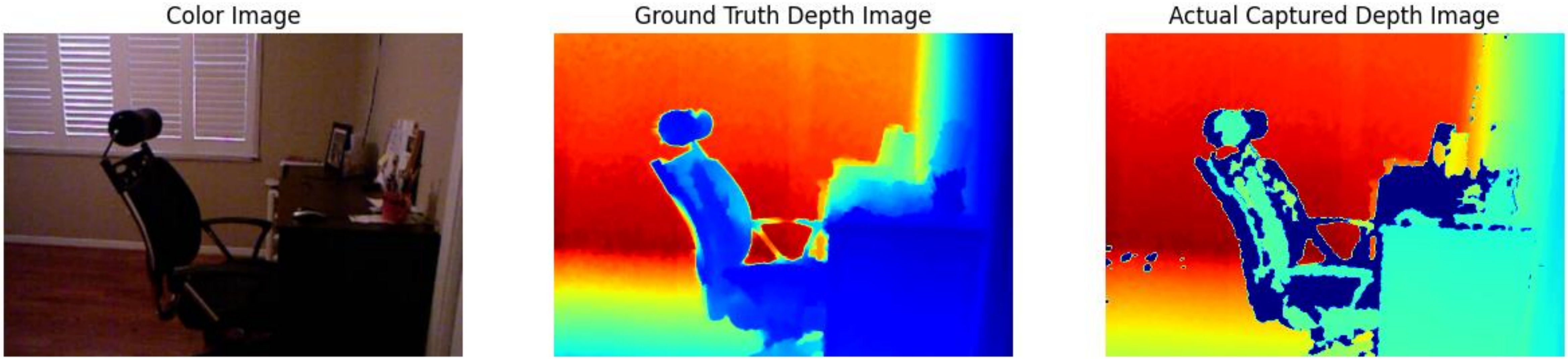
DENOISING DEPTH IMAGES USING RGB IMAGES

Vasu Eranki

MOTIVATION & OBJECTIVES

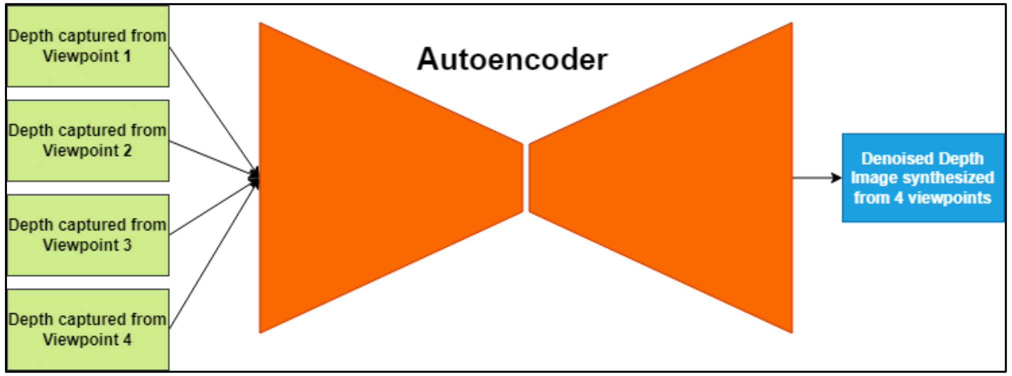


- Practical Depth Images captured can be quite noisy and are usually not denoised which affects downstream tasks and datasets
- Color and Scene information can be used since its less noisy than the Depth Image
- Having denoised depth images closer to the ground truth can help improve the confidence levels of downstream tasks
- The goal of this project is to leverage the information present inside the color image to help further denoise a depth image. (Without having access to pairs of clean-noisy depth images)

TECHNICAL APPROACH AND NOVELTY

Current State of the Art (SOTA) [1]

- Images captured at slightly different angles have:
  - Similar information about the scene
  - Different amounts of noise present in it
  - Trained in a self-supervised manner



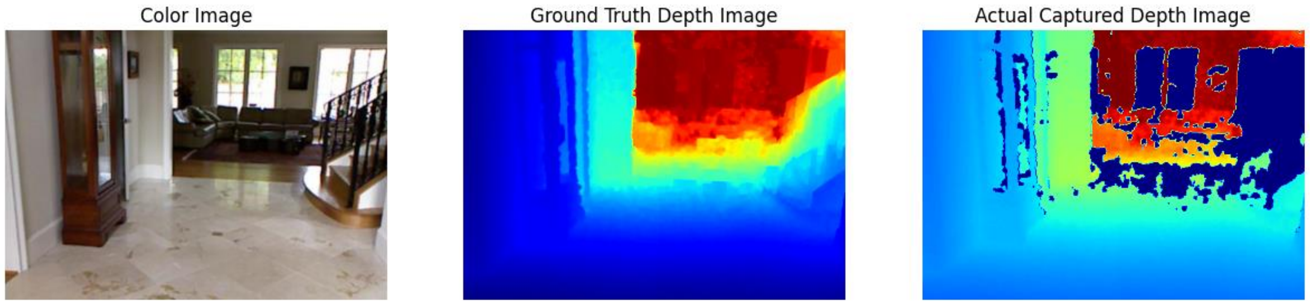
- Limitations:
  - Requires 4 different depth images for training and there is no such dataset that is present

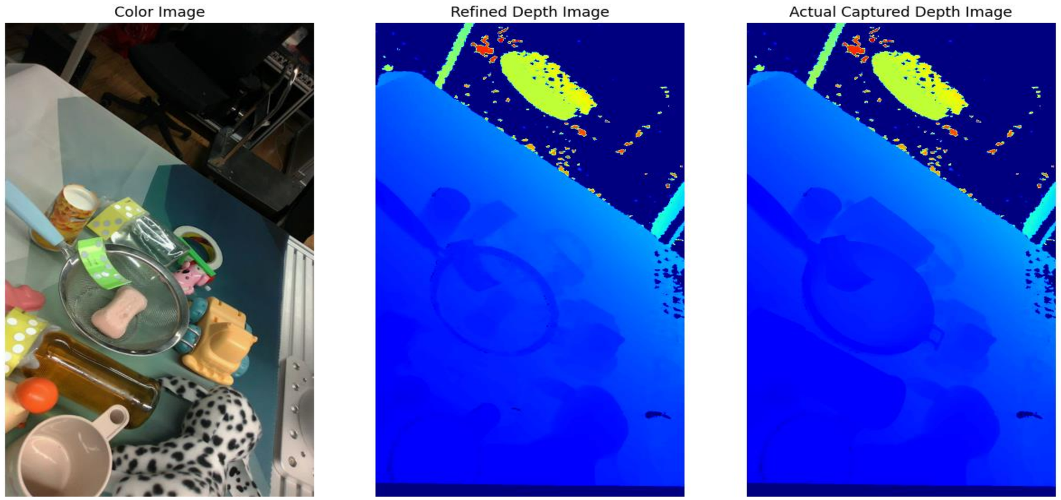
- Proposed Approach
  - Train a denoiser which is given both RGB and Depth
  - <u>Should use the scene information from RGB to denoise the Depth in a self-supervised manner</u>

METHODS

Dataset	Types of Images
NYU Depth Dataset [2]	Microsoft Kinect – Noisy and Clean
TransCG [3]	Intel L515 – Noisy and Refined Depth
CLUBS [4]	Intel D415 – Noisy

Hardware	Type
Intel RealSense L515 [5]	LiDAR based Depth Sensor





EVALUATION AND METRICS

Typical Values Seen in Literature

Metrics	NYU Depth Dataset [2]	TransCG Dataset [3]
Mean Absolute Error (MAE)	~5mm	~12mm
RMSE (Root Mean Squared Error)	~21mm	~35mm

RMSE and MAE are the most common metrics used for evaluating denoising based models

Algorithm	Time to denoise a single frame
Bilateral Filter	22ms
Anisotropic Diffusion	0.64s
SOTA [1]	16ms – On a T4 GPU

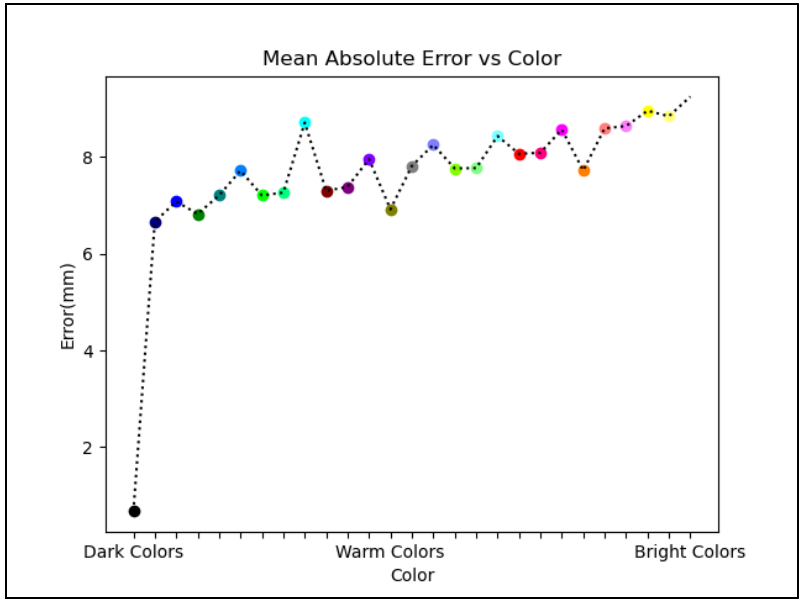
A secondary goal of this project is to not make an overly complex system. A metric that helps this is the time taken to denoise a single depth frame.

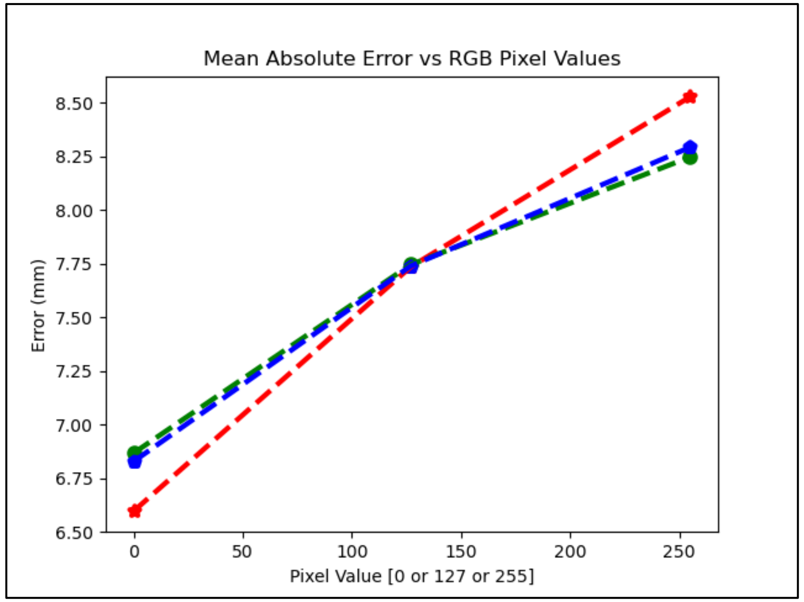
CURRENT STATUS

Simulated various colours through a monitor and captured the depth readings in a dark room:

Conclusions:

- Colors with stronger intensities generate more noise (like white, yellow, orange)
- The noise introduced by one channel is not independent of the others.





CURRENT STATUS AND NEXT STEPS

Setting Baseline Performance (Errors are in millimetre (mm))

		NYU Depth Dataset[2]		TransCG Dataset [3]	
		MAE	RMSE	MAE	RMSE
Classical Computer Vision	Bilateral Filter	16.41mm	37.62mm	41.03mm	84.90mm
	Anisotropic Diffusion based Filter	44.34mm	196.89mm	49.24mm	169.32mm
Data Driven Method	Current SOTA [1]	8.58mm	30.15mm	11.02mm	37.78mm

Next Steps

- Develop a Noise model from the data collected
- Apply it during training in one of two ways:
  - By augmenting our training data with a colour-informed noise function
  - Create a loss function that forces the model to leverage the colour information.

REFERENCES

[1] Sterzentsenko, V., Saroglou, L., Chatzitofis, A., Thermos, S., Zioulis, N., Doumanoglou, A., Zarpalas, D. and Daras, P., 2019. Self-supervised deep depth denoising. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 1242-1251).

[2] Silberman, N., Hoiem, D., Kohli, P. and Fergus, R., 2012. Indoor segmentation and support inference from rgbd images. In Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part V 12 (pp. 746-760). Springer Berlin Heidelberg.

[3] Fang, H., Fang, H.S., Xu, S. and Lu, C., 2022. Transcg: A large-scale real-world dataset for transparent object depth completion and a grasping baseline. *IEEE Robotics and Automation Letters*, *7*(3), pp.7383-7390.

[4] Novkovic, T., Furrer, F., Panjek, M., Grinvald, M., Siegwart, R. and Nieto, J., 2019. CLUBS: An RGB-D dataset with cluttered box scenes containing household objects. The International Journal of Robotics Research, 38(14), pp.1538-1548

[5] https://www.intelrealsense.com/lidar-camera-l515/