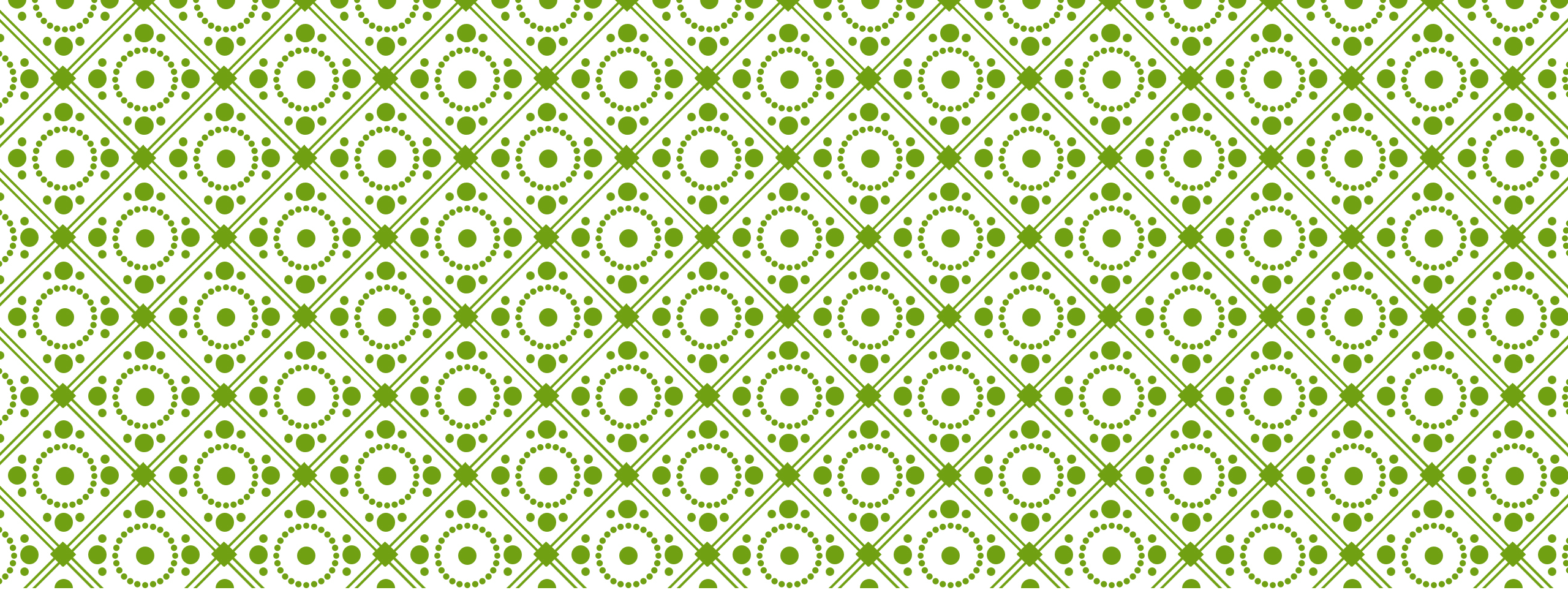


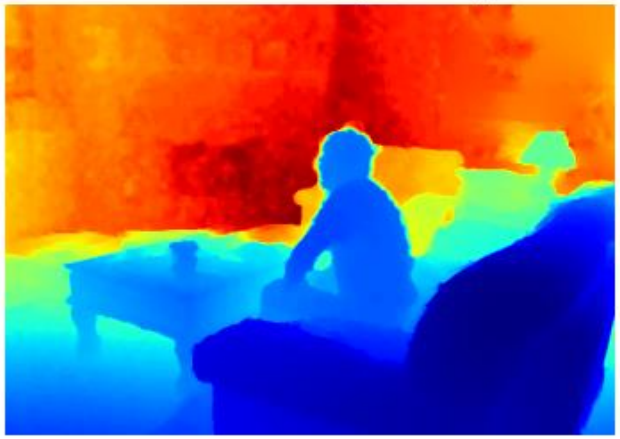
DENOISING DEPTH IMAGES USING RGB IMAGES

Vasu Eranki

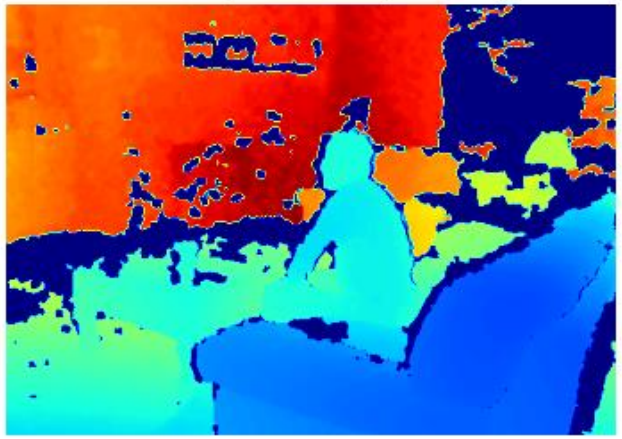
MOTIVATION

Color Image

Ground Truth Depth Image



Actual Captured Depth Image



- Commercial Depth Cameras suffer from multiple sources of noise which can severely degrade the image quality.
- Having cleaner depth maps can help with downstream tasks such as Segmentation and Object Detection
- The goal of this project is to leverage the information present in the RGB image to further denoise the depth map.

LITERATURE SURVEY & NOVELTY

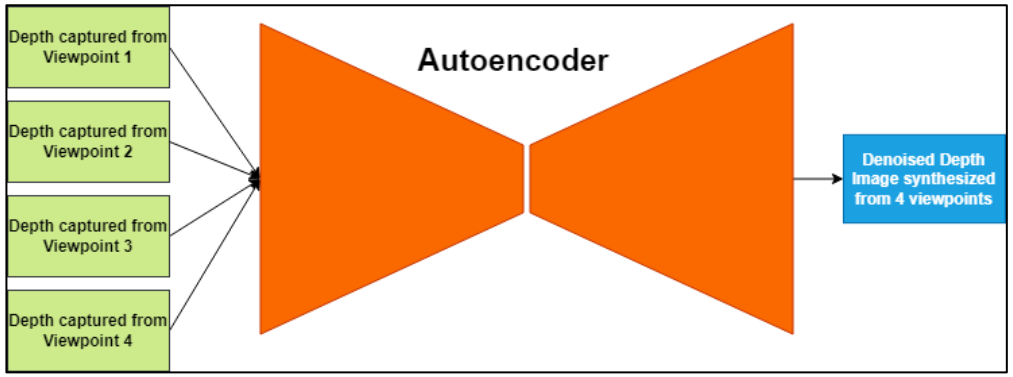
Computational Imaging Method [1]

- Same scene is taken at slightly different viewpoints, then using
- Uses the fact that noise isn't static, to remove noise from the depth image

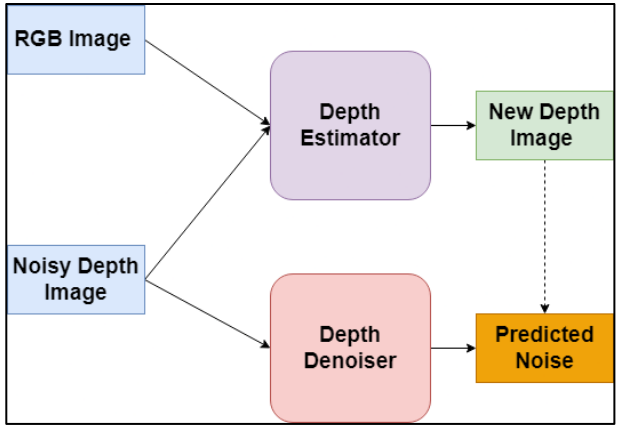
Leveraging Task Similarities [2]

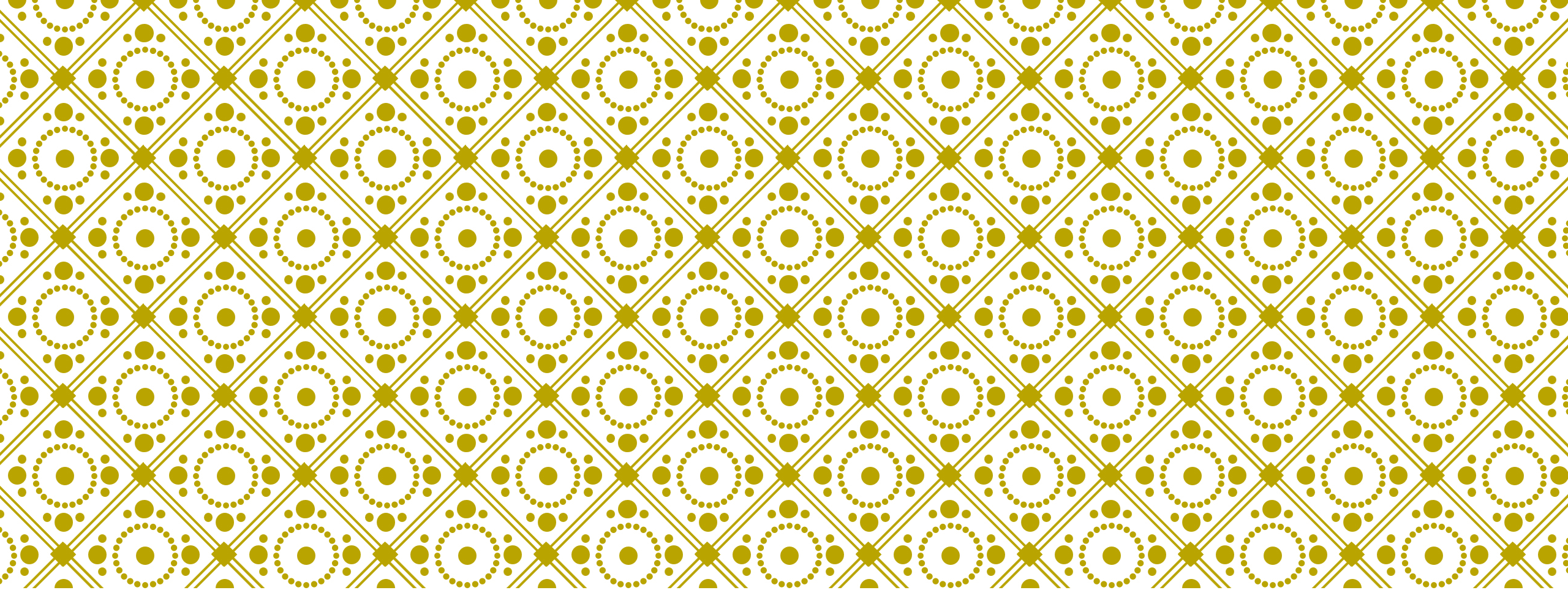
- Uses a depth estimator to create noisy-clean pairs which are then passed through a depth denoiser.
- Requires training both networks in parallel.

Leveraging Computational Imaging [1]



Leveraging Task Similarities [2]



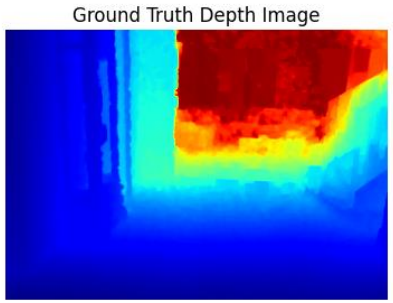


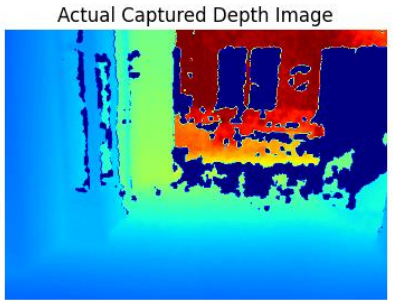
TECHNICAL METHODS

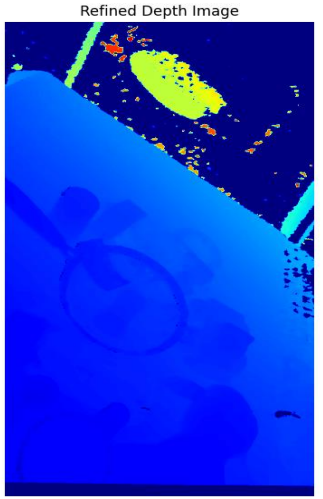
DATASET

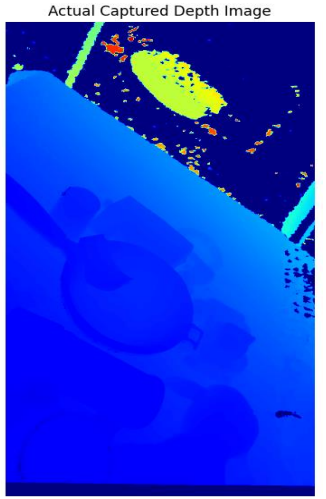
Dataset	Types of Images
NYU Depth Dataset [3]	Microsoft Kinect – Noisy and Clean
TransCG [4]	Intel L515 – Noisy and Refined Depth

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





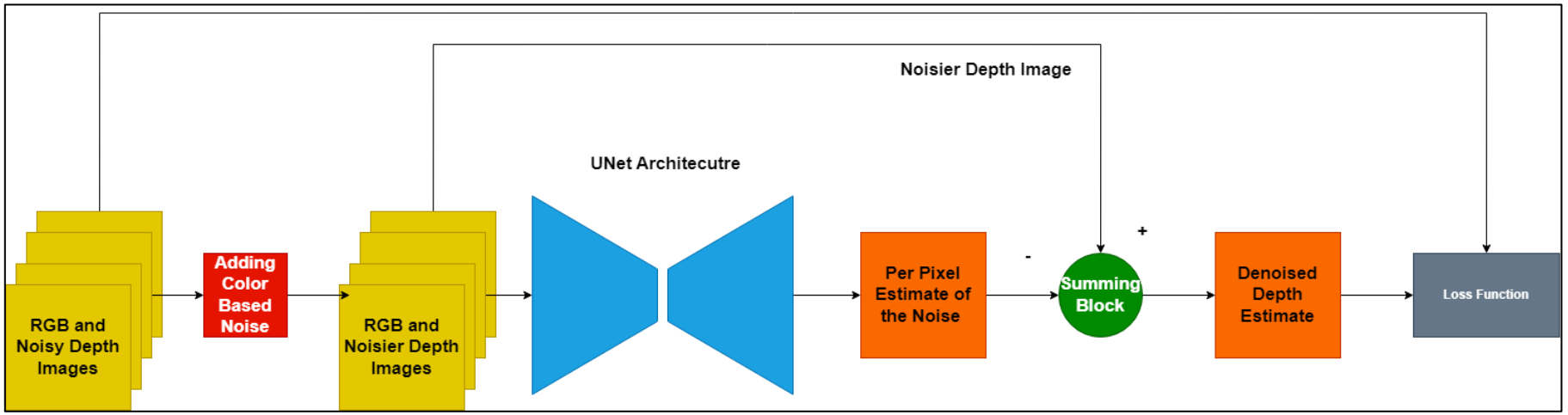




CREATING A COLOUR – NOISE FUNCTION

- ToF Sensors suffer from both distance dependent and intensity dependent noise.
- To mitigate the impact of this
  - Averaging the sensor readings over multiple seconds
  - Fixed distance of 0.5m
  - Sensor readings captured with minimal external light
- This experiment was repeatedly done over the course of 2 weeks, was subsequently then used to create a parametric noise model.

MODEL ARCHITECTURE & TRAINING



Three loss functions were tried and their results will be discussed in the following slides:

- Mean Squared Error Loss, additional constraints were levied such as latent space sparsity constraints or an additional signal via a downstream task (Semantic Segmentation).
- Following Hyperparameters were used to prevent overfitting and to stabilise the training process
  - Learning Rate Scheduler. At each epoch the learning rate is reduced by 10%
  - Dropout of 50% on the Depth Channel
  - L2 Regularization of 10^{-4}
- Total Number of trainable parameters = 2.6 Million, trained on the NYU Dataset for 10 Epochs and then in a zero-shot manner evaluated on the TransCG dataset.

RESULTS

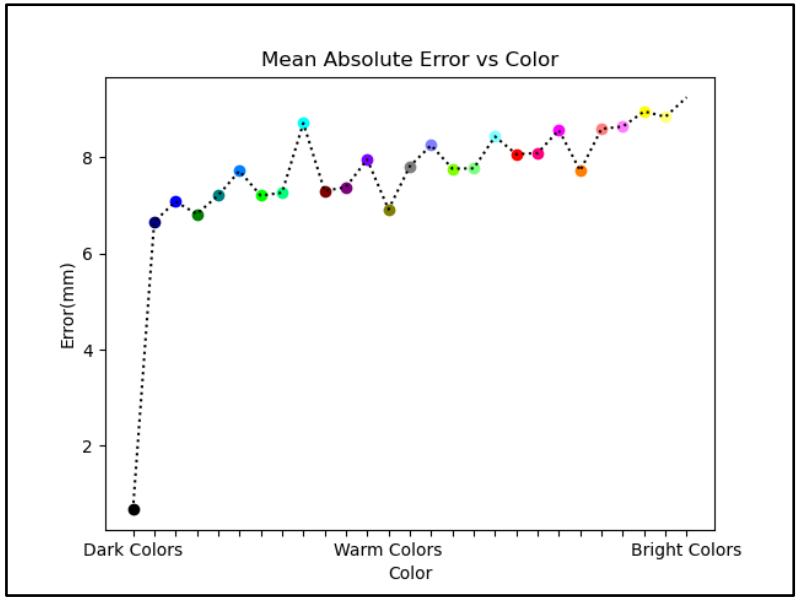
USING OUR RGB INFORMATION TO ESTIMATE THE NOISE

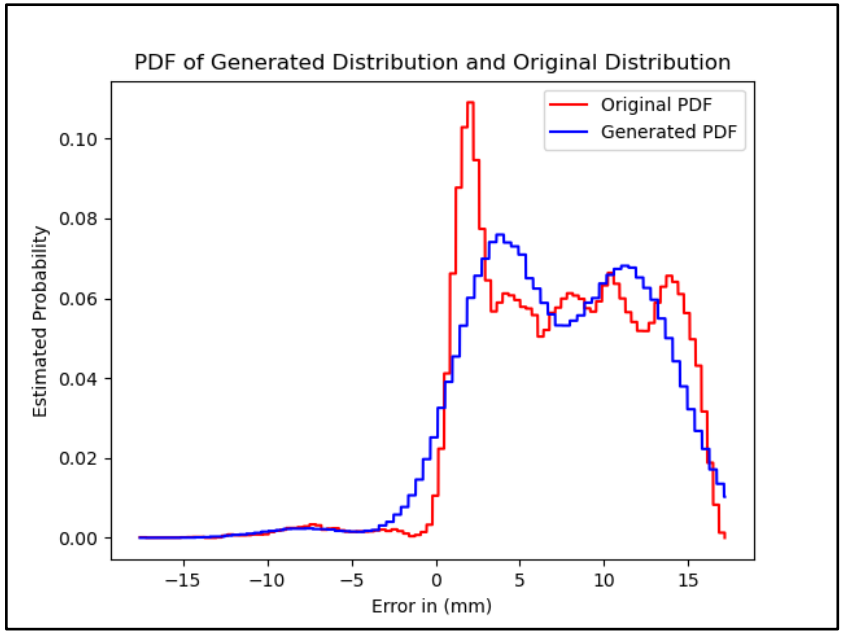
Results of Noise

- Brighter colours like yellow, pink and light green tend to have more noise in their associated depth readings
- Noise isn't channel independent. There's a relationship between R,G and B.
- The source of this error is in the sensor design, since all physics based reflection models contradict our results.

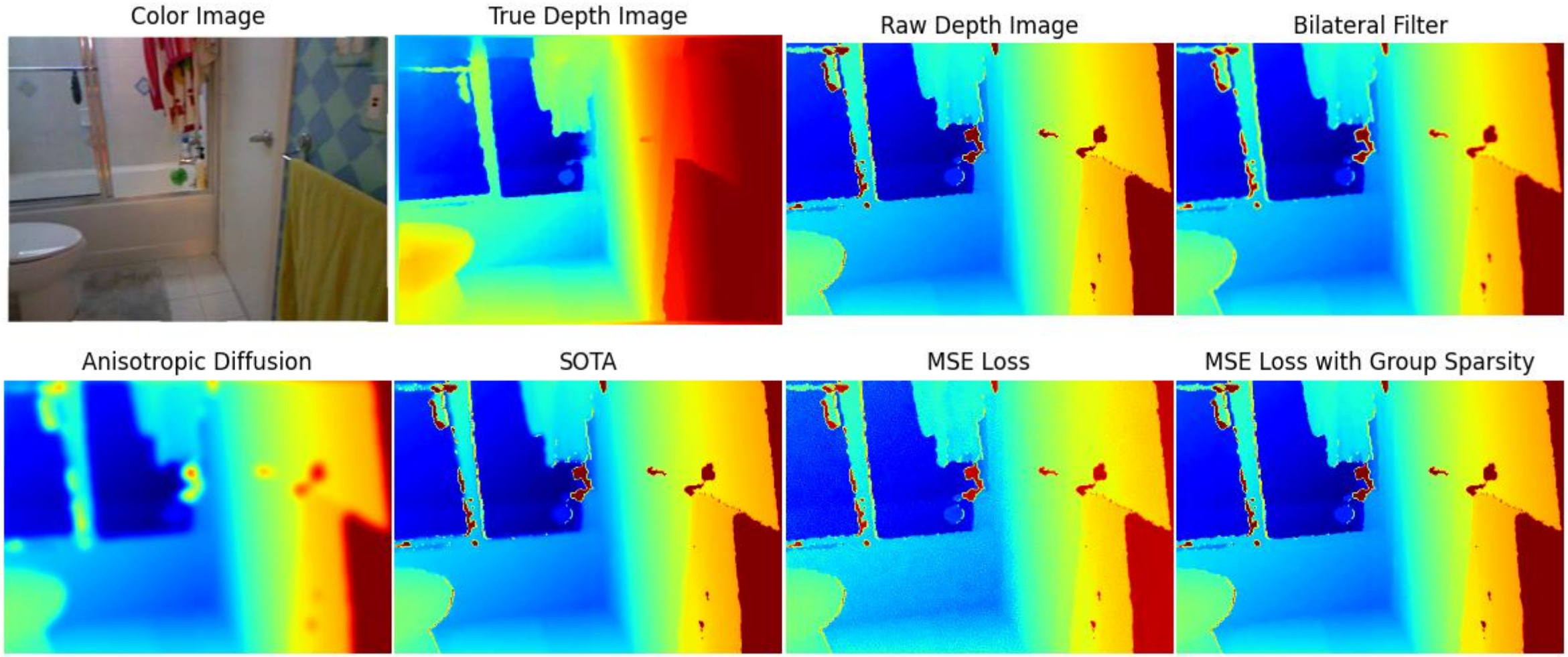
Using this for training a NN

- A Gaussian Mixture Model with 3 Components (R,G,B) was used to approximate the distribution of errors across colours.

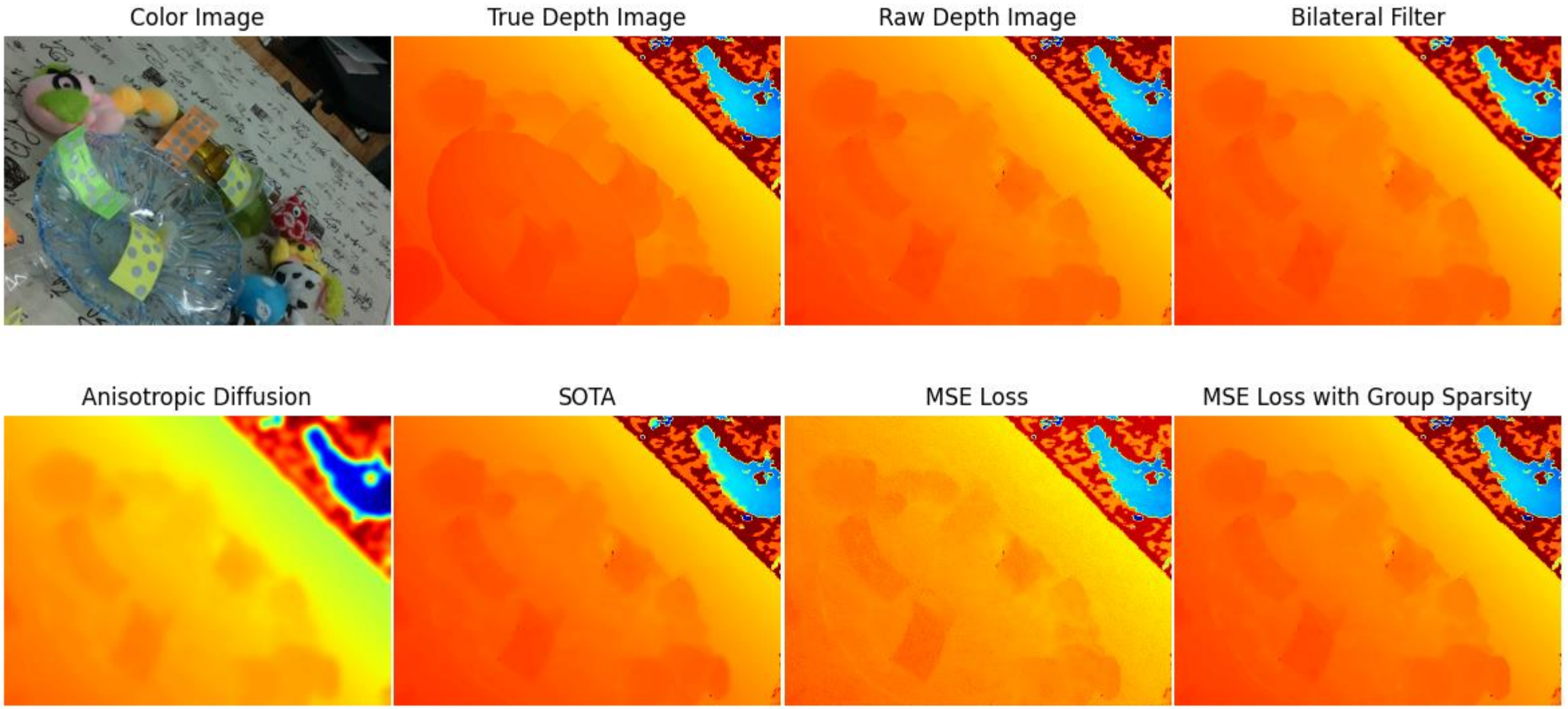




GENERATED OUTPUTS (NYU DEPTH DATASET [3])



GENERATED OUTPUTS (TRANSCG DATASET [4])



EXPERIMENTAL EVALUATION

MAE and RMSE were calculated on the masked image (only valid pixels were used)		**NYU Depth Dataset[2]**		**TransCG Dataset [3]**	
		MAE	RMSE	MAE	RMSE
Classical Computer Vision (Out of the Box Implementations)	Bilateral Filter	16.41mm	37.62mm	41.03mm	84.90mm
	Anisotropic Diffusion based Filter	44.34mm	196.89mm	49.24mm	169.32mm
CNN based Method	Current SOTA [1]	**8.58mm**	30.15mm	**11.02mm**	37.78mm
Control Group	MSE (w AWGN Noise)	16.74mm	36.30mm	31.01mm	42.12mm
Proposed Architectures **Base Model was a UNet**	MSE	11.75mm	**30.05mm**	35.99mm	46.30mm
	MSE w Representation Loss	10.01mm	**24.73mm**	16.35mm	**32.45mm**
	MSE w training on Downstream Tasks	15.31mm	34.21mm	37.81mm	49.05mm

EXPERIMENTAL EVALUATION

Models		Inference Time
Classical Computer Vision	Bilateral Filter	22ms
	Anisotropic Diffusion based Filter	0.64s
CNN Based Architecture	Current SOTA [1]	16ms – On a T4 GPU (8GB RAM)
Proposed Architecture Base Model was a UNet	UNet (MSE/ MSE w Representation Loss/ MSE w training on downstream tasks)	12.8ms – On a T4 GPU (8GB RAM)

CONCLUSION AND FUTURE SCOPE

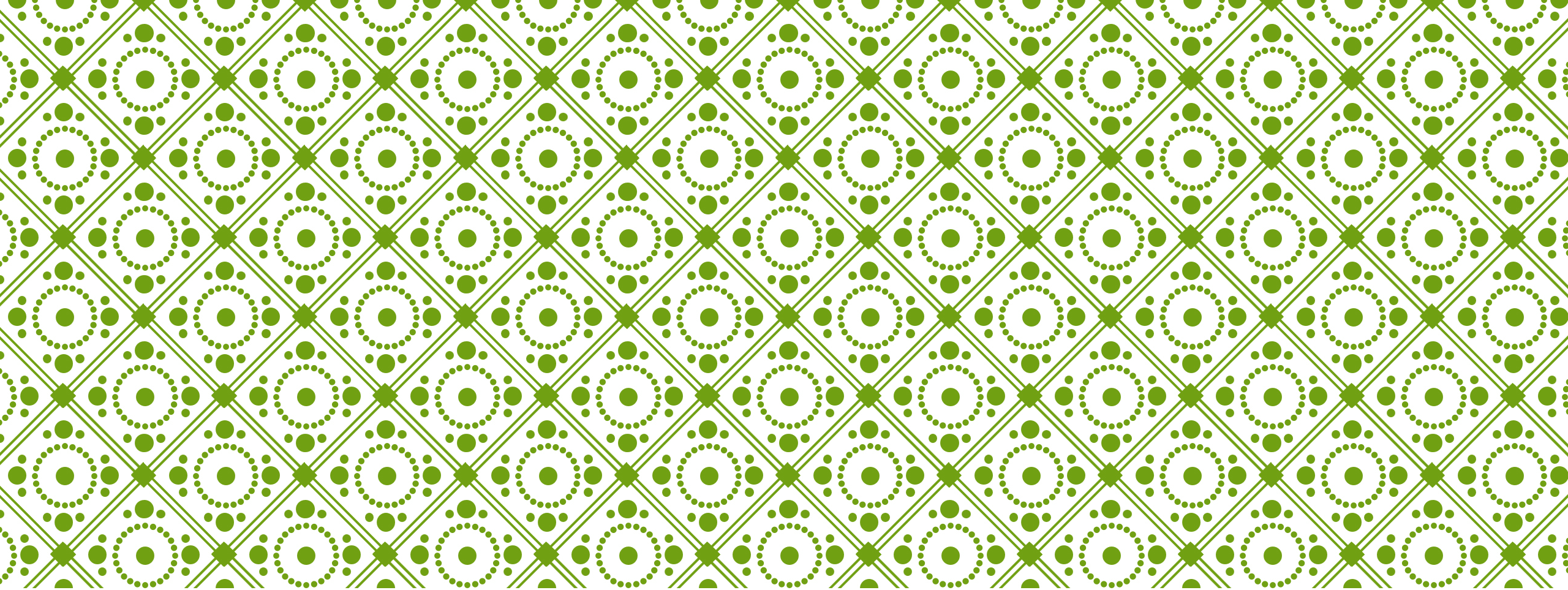
CONCLUSION

- The colour offers some useful information on the noise in the depth map, and it can be approximated with a mixture of gaussians.
- The U-Net model trained on the MSE loss with an additional sparsity constraint is able to generalise well and can work in a zero-shot setting as well, making it device agnostic.
- Injecting colour based noise into a model helped it to learn, since each of the proposed model outperformed the model trained AWGN noise.

FUTURE SCOPE

Ideas for future directions:

- Using diffusion models to denoise images since the training process is quite similar and recent literature [7]
- Convex optimization based methods which focus on reducing the # of eigenvalues in the image. Such methods are interpretable, work in a zero-shot manner [6]
- Leverage embeddings to find a sparse representation of images, effectively denoising them.



THANK YOU FOR YOUR TIME

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] Fan, L., Li, Y., Jiang, C. and Wu, Y., 2022, May. Unsupervised Depth Completion and Denoising for RGB-D Sensors. In *2022 International Conference on Robotics and Automation (ICRA)* (pp. 8734-8740). IEEE.

[3] 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.

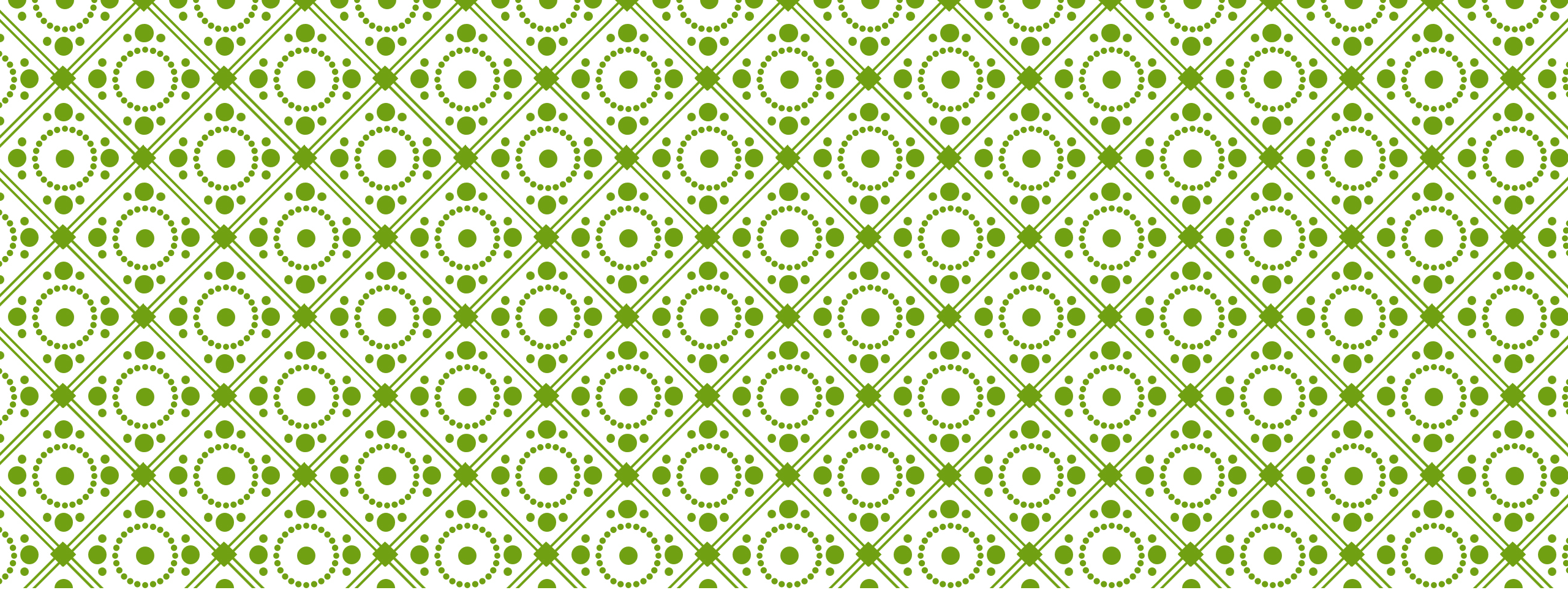
[4] Fang, H., Fang, H.S., Xu, S. and Lu, C., 2022. Transcg: A large-scale real-world dataset for transparent object depth completion and a grasping baseline. *IEEE Robotics and Automation Letters*, *7*(3), pp.7383-7390.

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

REFERENCES

[6]Gu, S., Zhang, L., Zuo, W. and Feng, X., 2014. Weighted nuclear norm minimization with application to image denoising. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2862-2869).

[7] Yang, C., Liang, L. and Su, Z., 2023. Real-World Denoising via Diffusion Model. *arXiv preprint arXiv:2305.04457*.



ADDITIONAL SLIDES

CDF OF GENERATED GAUSSIAN MIXTURE MODEL