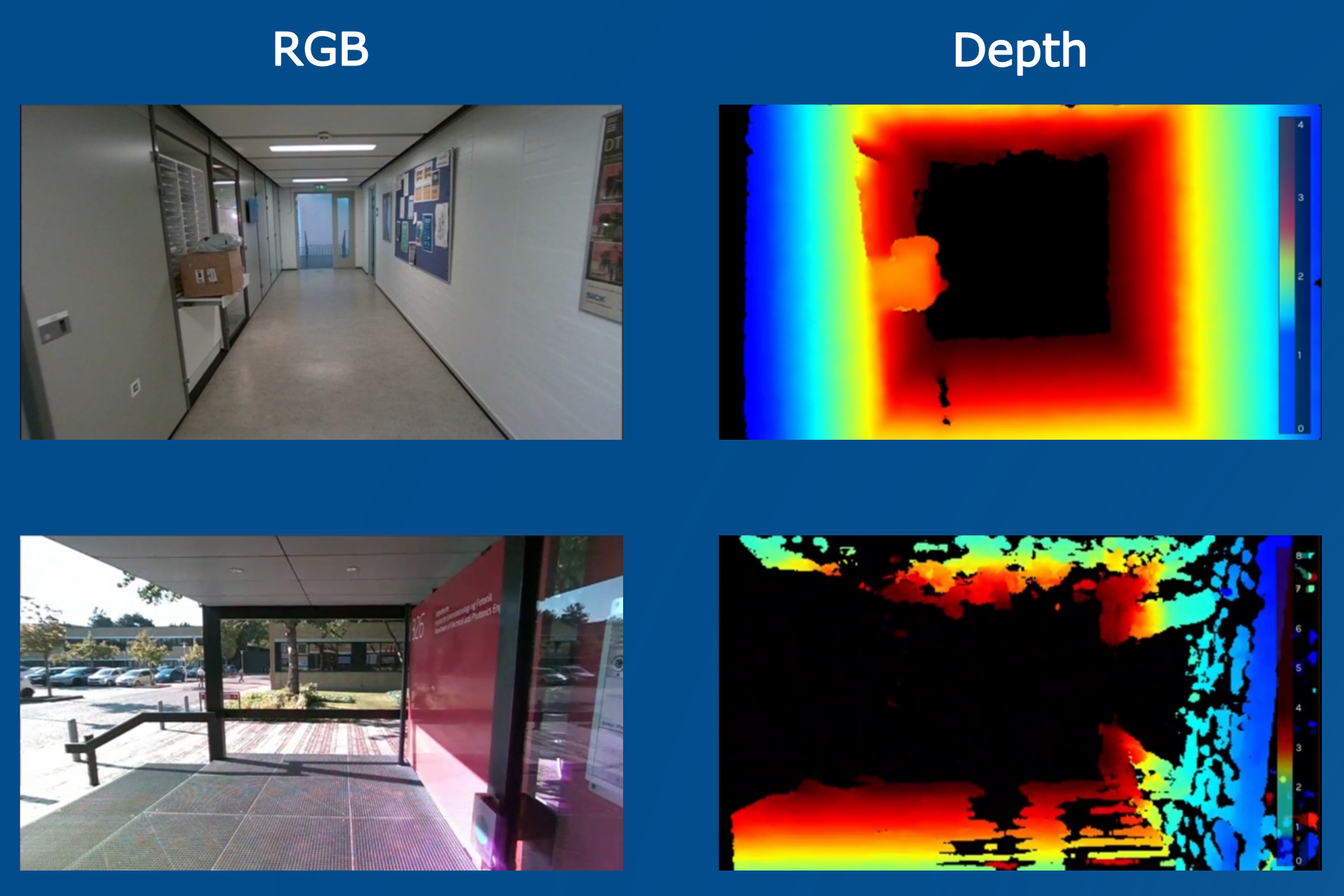
SteeredMarigold: Steering Diffusion Towards Depth Completion of Largely Incomplete Depth Maps

Jakub Gregorek, Lazaros Nalpantidis*

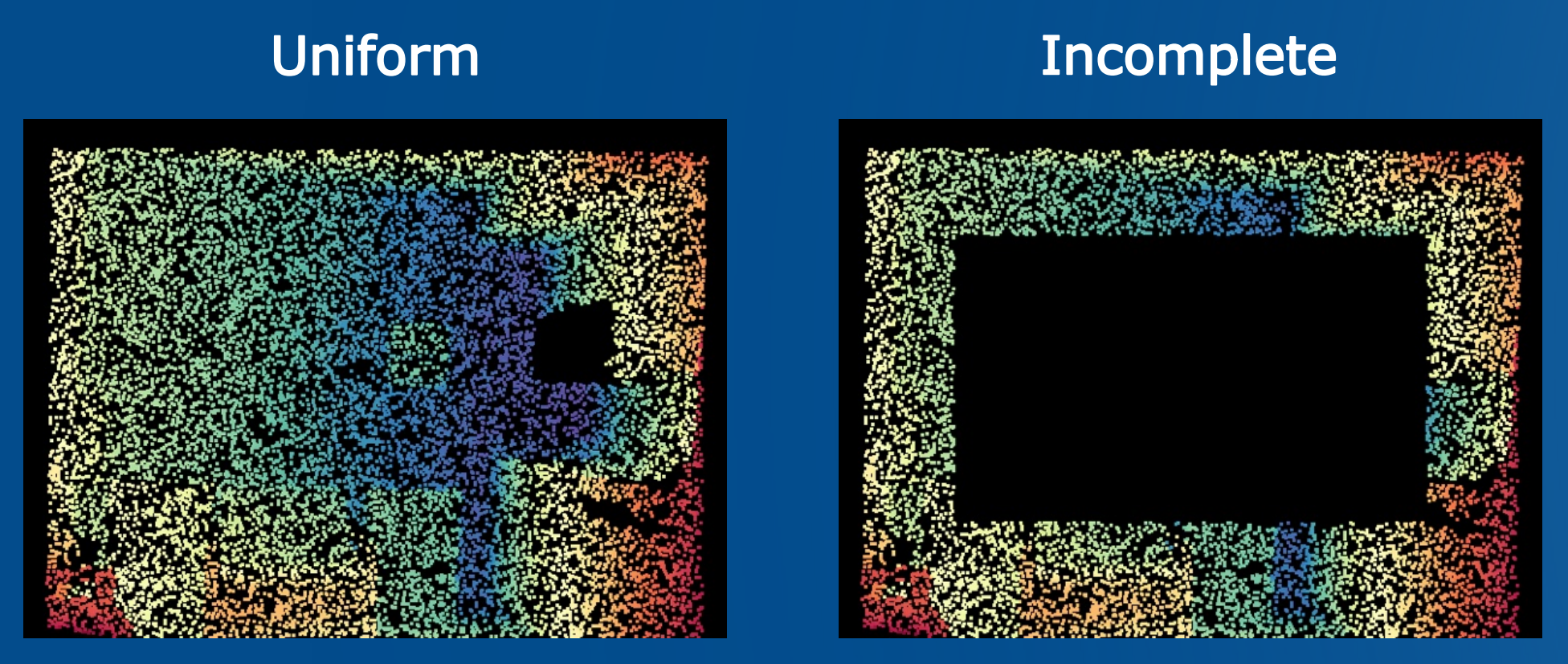
Page & Code

1. Motivation

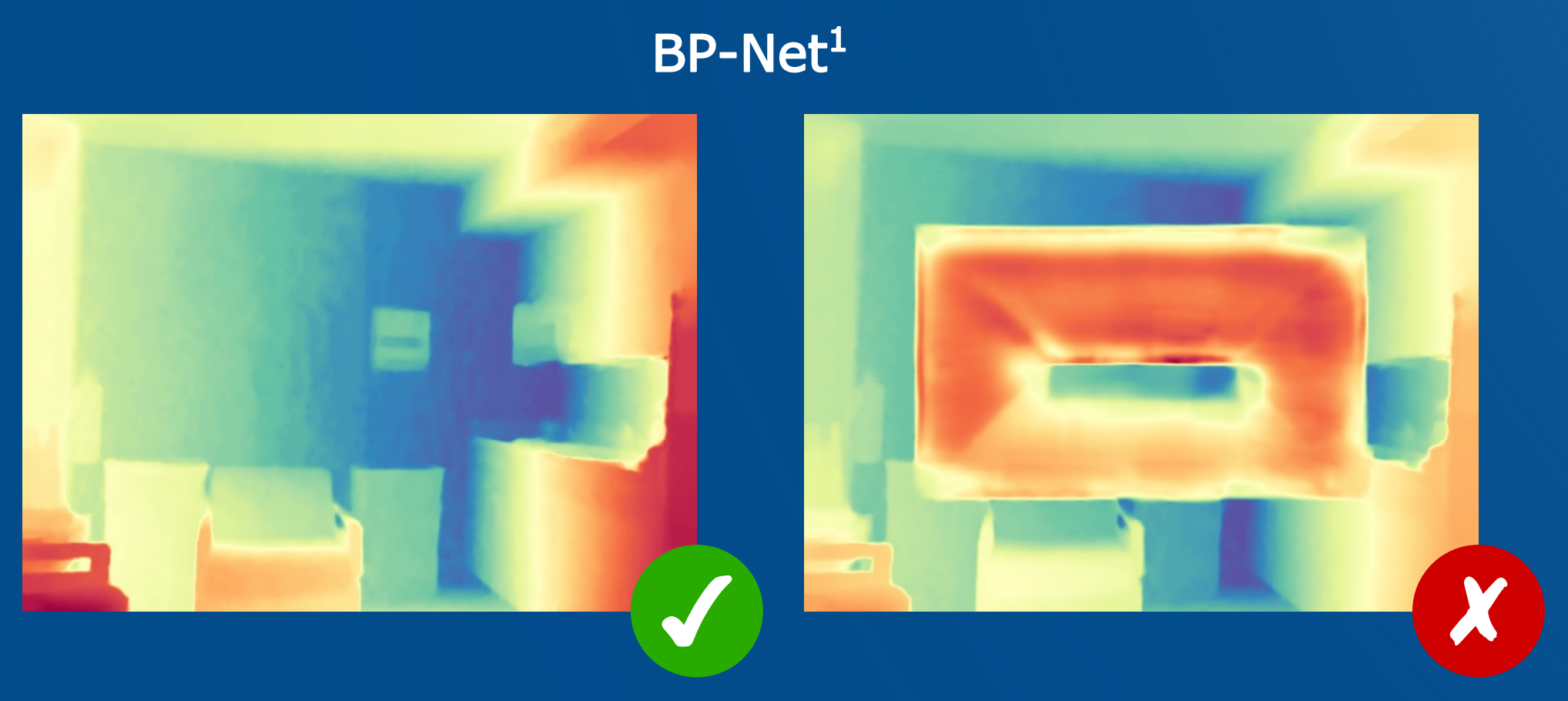
Real world depth maps captured by RGB-D sensors exhibit large areas with missing depth measurements. Among common causes are uniform or repetitive textures, specular surfaces, dust, rain or fog, or limited sensor range.

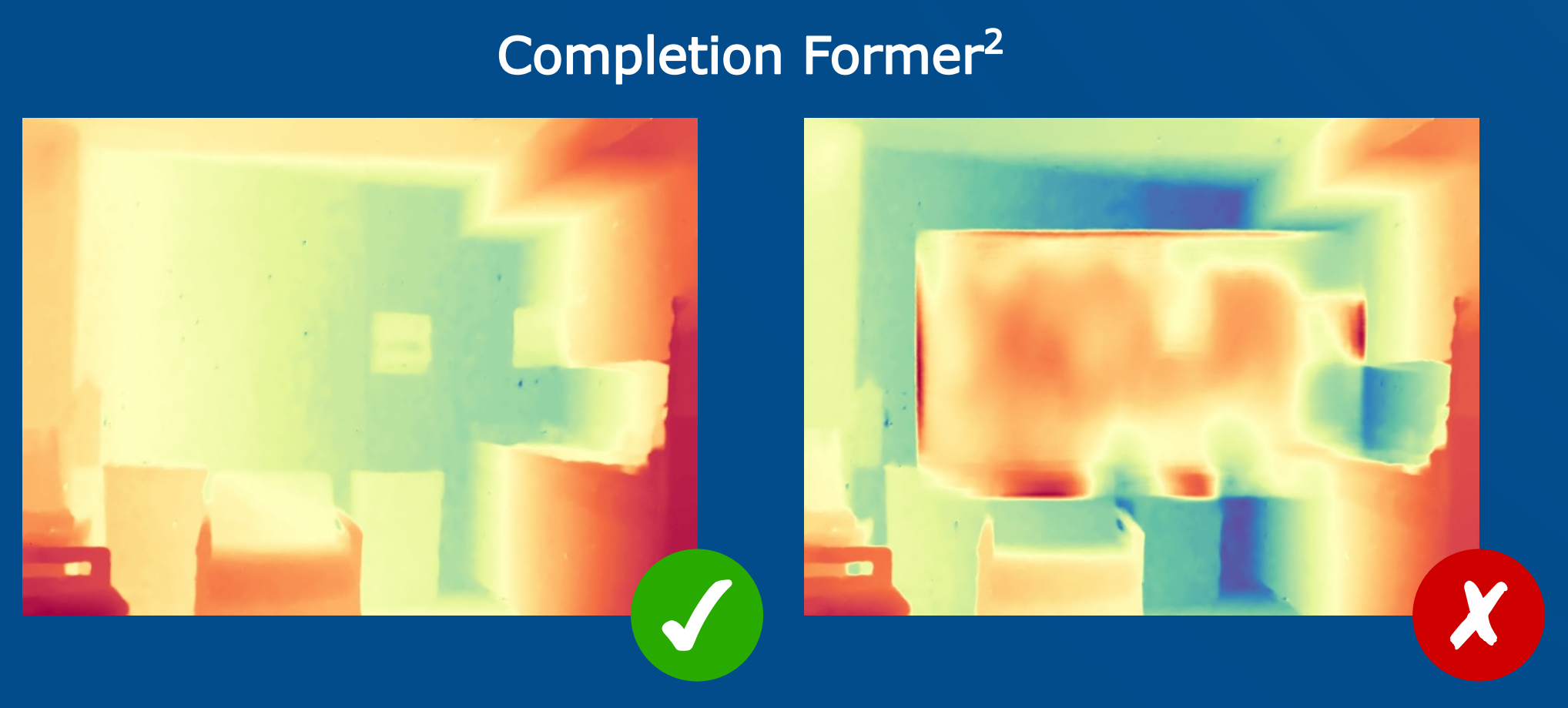


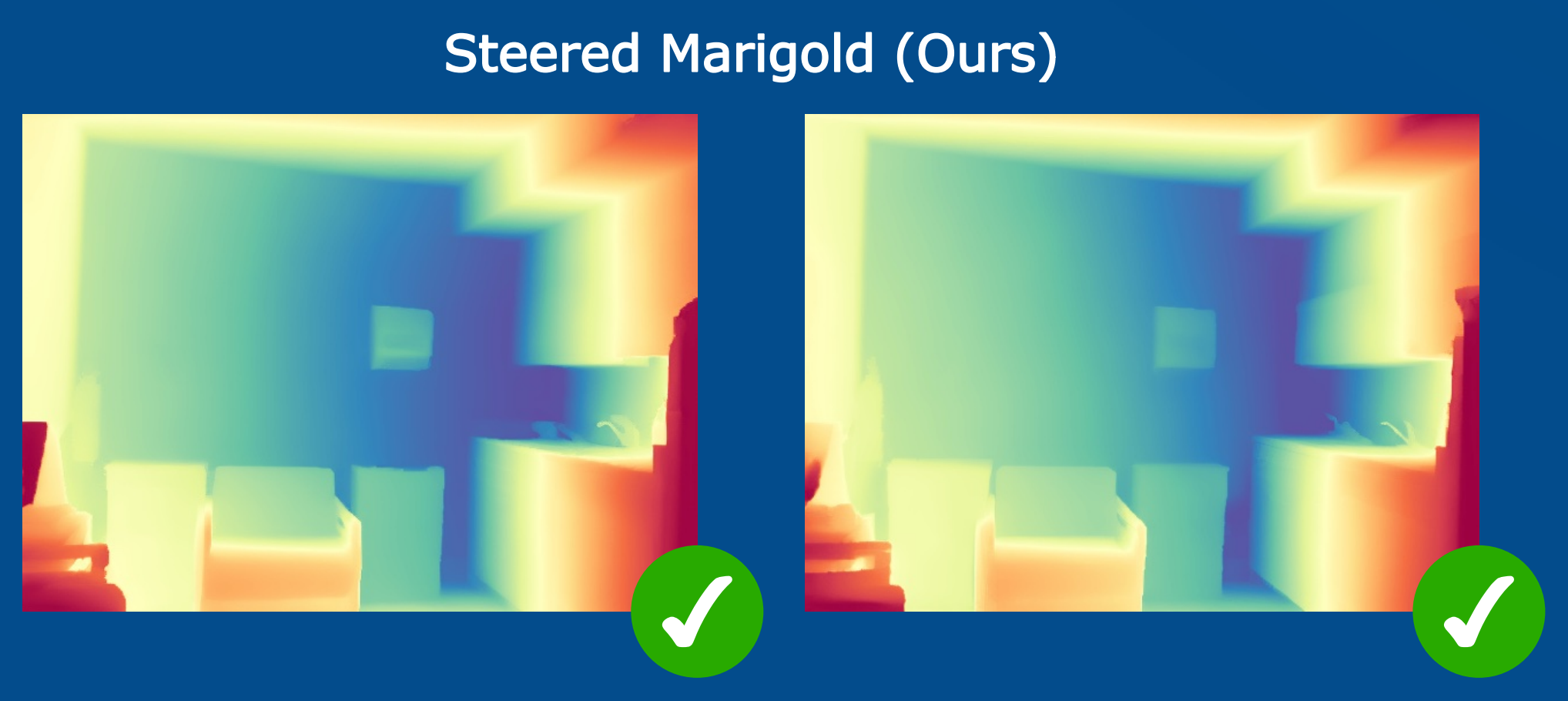
The majority of depth completion methods assume depth measurements uniformly covering all areas of the scene.



When the methods are presented with largely incomplete depth maps, their performance significantly degrades.







2. Solution

Our training-free zero-shot depth completion method Steered Marigold utilizes the pretrained monocular depth estimator Marigold[3]. We follow Denoising Diffusion Probabilistic Models[4], with with timesteps 0< t < T, variance schedule $\beta_1 \dots \beta_T$ and noising process:

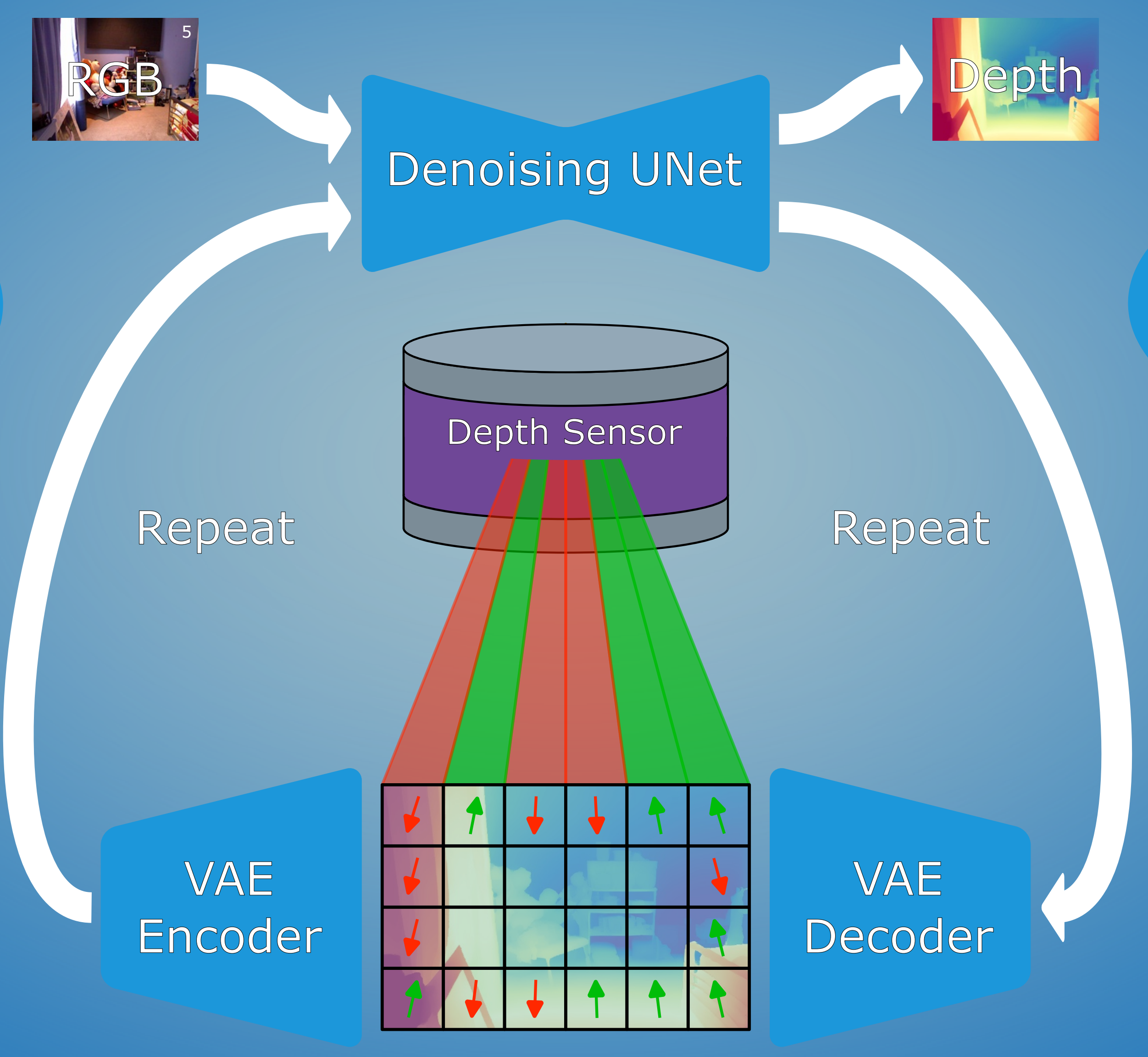
- $q(\mathbf{x}_t|\mathbf{x}_{t-1}) \coloneqq \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I})$ noisy sample is expressed as: $\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$

where: $\bar{\alpha}_t = \prod_{s=1}^{t}\alpha_s$ $\alpha_t = 1-\beta_t$ $\epsilon \sim \mathcal{N}(0, \mathbf{I})$

Denoising process defined as:

- $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) \coloneqq \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$

$$\begin{cases} \boldsymbol{\mu}_\theta = \dfrac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\tilde{\mathbf{x}}_0 + \dfrac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t \\ \sigma_t^2 = \dfrac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \end{cases}$$

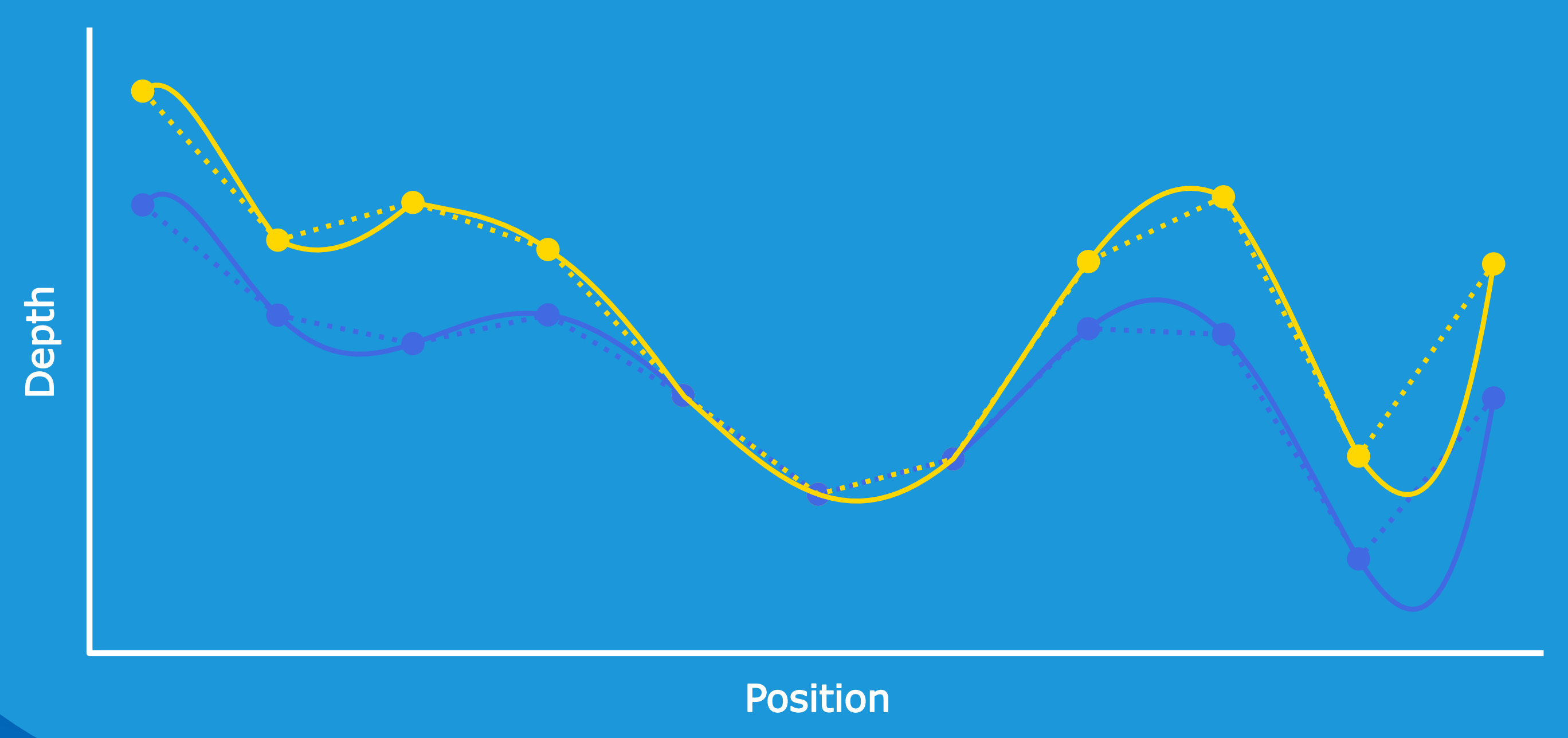


The model works in the latent space of VAE. Which encodes RGB condition, decodes the result, and plays a role in the steering process.

Estimate a clean data sample:

$$\tilde{\mathbf{x}}_0 = \sqrt{\bar{\alpha}_t}\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\mathbf{v}_\theta(\mathbf{x}_t)$$

$$\mathbf{x}_{t-1} = \mathbf{x}_{t-1} + \lambda \cdot (\mathcal{E}(\tilde{\mathbf{x}}_0^{\mathcal{D}} - \phi_1(\tilde{\mathbf{x}}_0^{\mathcal{D}}, P) + \phi_2(\tilde{\mathbf{x}}_0^{\mathcal{D}}, \mathbf{c}, P)) - \tilde{\mathbf{x}}_0)$$

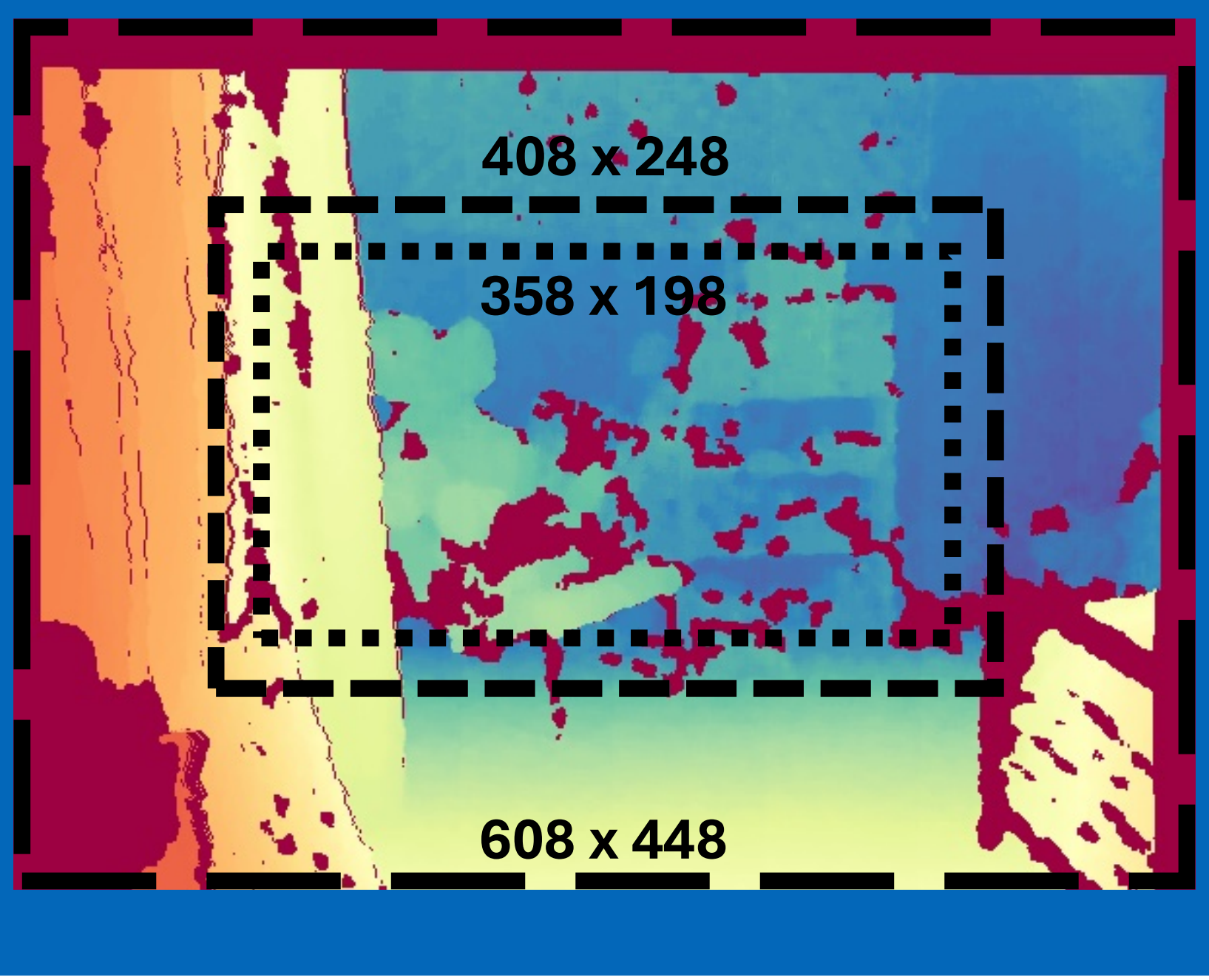


- steering as an update of x_{t-1}
- defined in the latent space
- utilizes encoder ε and decoder D
- ϕ_1 and ϕ_2 are overcoming sparsity
- depth estimate and c give the direction
- steering factor λ scales the step

3. Results

- evaluated on NYUv2[5] test dataset
- metrics for depth estimation and completion
- three evaluation areas (see figure on the right)
- depth **c** sampled only outside 408 x 248
- **c** is used to compute shift and scale values
- for more results scan the QR code on top

Evaluation Zones

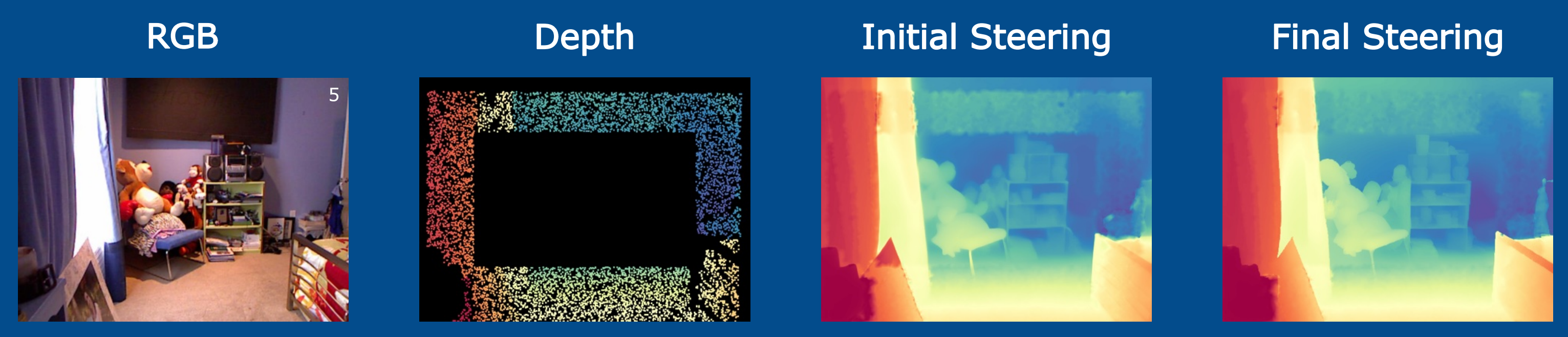


Error decreases even in areas without depth samples available!

Evaluation Area	Method	REL ↓		δ_1 ↑		RMSE ↓		MAE ↓	
608 × 448	Marigold	0.0618		0.9559		0.2555		0.1599	
	SteeredMarigold	0.0352	↓ 43.04%	0.9834	↑ 2.88%	0.1854	↓ 27.43%	0.0960	↓ 39.96%
408 × 248	Marigold	0.0610		0.9570		0.2794		0.1799	
	SteeredMarigold	0.0510	↓ 16.39%	0.9718	↑ 1.55%	0.2586	↓ 7.44%	0.1523	↓ 15.34%
358 × 198	Marigold	0.0630		0.9536		0.2906		0.1912	
	SteeredMarigold	0.0573	↓ 9.04%	0.9646	↑ 1.15%	0.2850	↓ 1.92%	0.1743	↓ 8.84%

4. Conclusion

- we proposed a training-free zero-shot method depth completion method: SteeredMarigold
- the method is capable of completing the largely incomplete depth maps where other methods fail
- the proposed steering process improves performance in the areas with no depth samples available
- the model harmonizes the areas not covered by depth measurements with the steering direction
- the method inherits its qualities from the successful depth estimator Marigold
- while the method does not require training, it comes with a large computational cost



References

[1] Jie Tang et al., Bilateral Propagation Network for Depth Completion, CVPR 2024
[2] Youmin Zhang et al., CompletionFormer: Depth Completion with Convolutions and Vision Transformers, CVPR 2023
[3] Bingxin Ke et al., Repurposing Diffusion-Based Image Generators for Monocular Depth Estimation, CVPR 2024
[4] Jonathan Ho et al., Denoising Diffusion Probabilistic Models, NeurIPS 2020
[5] Nathan Silberman et al., Indoor Segmentation and Support Inference from RGBD Images, ECCV 2012

This work has been funded and supported by the EU Horizon Europe project “RoBétArmé” under the Grant Agreement 101058731.

IEEE INTERNATIONAL CONFERENCE ON ROBOTICS AND AUTOMATION

ATLANTA 2025

*Department of Electrical and Photonics Engineering
DTU - Technical University of Denmark
{jagre, lanalpa}@dtu.dk