

Monocular Depth Estimation for Spacecraft: Combining Relative and Scale Information

Siddharth Singh
Faculty of Engineering and Applied Sciences (FEAS)
Cranfield University
United Kingdom, MK43 0AL
siddharth.singh.026@cranfield.ac.uk

Dr. Leonard Felicetti
Reader in Space Robotics and GNC, FEAS
Cranfield University
United Kingdom, MK43 0AL
leonard.felicetti@cranfield.ac.uk

Prof. Hyo Sang Shin
Autonomous and Intelligent Systems Laboratory
Cho Chun Shik Graduate School of Mobility
South Korea, 34051
hyosangshin@kaist.ac.kr

Prof. Antonios Tsourdos
Director of AI, Robotics & Space, FEAS
Cranfield University
United Kingdom, MK43 0AL
a.tsourdos@cranfield.ac.uk

Abstract—This paper proposes a novel object-specific depth estimation method for space applications, combining multi-scale feature fusion, transformer improvements, and frequency-based refinement in the Discrete Cosine domain (DCT). It addresses challenges like high contrast, noise, and limited compute by leveraging segmentation maps to focus computation. The method avoids iterative full-image processing and aims for real-time, robust performance. Benchmarks are performed on spacecraft such as LRO and Dream Chaser, showcasing state-of-the-art depth accuracy, geometric learning and structural reconstruction showcasing sub 0.5 meter accuracy.

process over the whole image. These advancements allow for an object-specific depth estimation algorithm for space applications with a lesser computational requirement. We benchmark our algorithm and reference algorithms on our own synthetic datasets featuring LRO and Dream Chaser to demonstrate the generalizability and robustness of our algorithm across spacecraft of different scales and geometric structures.

TABLE OF CONTENTS

1. INTRODUCTION.....	1
2. RELATED WORKS	1
3. PROPOSED METHODOLOGY	2
4. EXPERIMENTS AND RESULTS	2
5. RESULTS	4
6. CONCLUSION	6
ACKNOWLEDGEMENTS.....	6
REFERENCES	6
BIOGRAPHY	7

1. INTRODUCTION

Future mission concepts for active debris removal or servicing require autonomous spacecraft to perform rendezvous or close proximity operations to targets. Most algorithms rely on cameras or sensors (e.g., Lidars) for relative navigation. Depth estimation generally provides the most useful information (e.g. relative distance from the target) but is a complex challenge due to the noisy, textureless and high contrast images. To overcome these challenges of visual noise, compute restraint and object-specific tasks, we propose a novel method for object-specific depth estimation. We take the learnings of multi-scale feature fusion ([1]), the improvements from transformer architectures ([2],[3], etc.), and the introduction of the use of CRFS in the Discrete Cosine Domain (DCT) by [4]. Our algorithm uses DCT-CRFS at a multi-scale level (Multi-Scale Transformers) using an upscaling and refining methodology with additional information from segmentation maps to reduce unnecessary computing and avoid an iterative

2. RELATED WORKS

Most current space applications directly regress 6D pose using machine learning or deep learning models due to computational resource constraints. These approaches typically lead to black-box regressions for depth and location estimates. Although more sophisticated machine learning or deep learning models offer higher accuracy in depth prediction, they often remain computationally expensive and unsuitable for real-time deployment on space-grade flight hardware, which is typically limited in processing capability and memory.

Computationally lightweight models such as **DepthPro** [5] use multi-scale prediction architectures to enable fast inference but still suffer from the inherent opacity of deep learning methods. Transformer-based models including **Marigold** [6], **ZoeDepth** [3], **MiDaS** [2], **DepthART** [7], and **MonoDepth** [1] deliver excellent accuracy across varied domains. However, their high computational cost makes them impractical for embedded or real-time space operations.

DCDepth [4] introduces a novel depth estimation technique leveraging the Direct Cosine Manifold to improve upon traditional Conditional Random Fields (CRFs) and Markov Random Fields (MRFs) [8]. Despite its improved accuracy, its iterative process can become computationally expensive depending on input resolution.

Deployment Challenges in Space

Deploying depth estimation models for space missions entails unique challenges beyond those in terrestrial scenarios:

- **Hardware Constraints:** space-rated hardware such as RAD750 or Leon3 CPUs operate under strict limits in memory and processing speed, rendering large transformer models infeasible, but the introduction of space-rated Jetson improves this situation [9].
- **Domain Shift:** space imagery often lacks background objects and exhibits extreme lighting, contrast, and reflectivity

variations. This shift from Earth-based datasets necessitates domain adaptation or synthetic data training.

- **Visual Conditions:** harsh lighting, low contrast, and specular reflections complicate depth estimation. Robust models must account for these factors or incorporate multimodal inputs (e.g., IR or stereo).
- **Model Interpretability:** due to the high stakes of space operations, black-box predictions are risky. Models must be explainable and provide confidence metrics.

Recent Approaches

Several new methods have emerged to address these trade-offs:

- **Marigold** [6] reuses pre-trained diffusion models (e.g., Stable Diffusion) for zero-shot monocular depth estimation.
- **DepthART** [7] reframes the problem as an autoregressive refinement task, allowing for model self-improvement during inference.
- **MiDaS v3.1** [2] offers multiple ViT backbones to explore performance-runtime trade-offs.
- **Lite-Mono** [10] combines CNNs and transformers with modules like Consecutive Dilated Convolutions (CDC) for high accuracy at low computational cost.
- **Token-Sharing Transformer (TST)** [11] uses global token sharing to enable real-time inference on platforms like NVIDIA Jetson.
- **NeW CRFs** [8] enhances CRF optimization through multi-head attention in localized windows for improved scalability.
- **DepthPro** [5] utilizes multi-scale prediction and upsampling strategy using multi-scale transformers (DINO, ViT, BEiT) and feature fusion.
- **FastDepth** [12] implements a simpler feature upscaling-based dense prediction strategy by using an image classification network (CNN Encoder) as a backbone.

Limitations of Monocular Metric Depth Estimation in Space Applications

State-of-the-art monocular metric depth estimation (MMDE) models, MonoDepth, Midas, and FastDepth [12] achieve high performance on traditional datasets but face critical limitations when applied to space imagery. The foremost issue is **scale ambiguity**, inherent to monocular setups, which impairs recovery of absolute distances on planetary terrains unless supplementary cues (e.g., stereo, IMU, or orbital priors) are provided. Models like Marigold [6] and Apple DepthPro [5] are highly **domain-specific** and poorly generalize to space environments due to the absence of Earth-like textures, structured lighting, or familiar object geometries. Transformer-based methods such as NewCRFs and DC-Depth offer improved spatial understanding, but their high computational requirements and poor real-time performance make them unsuitable for onboard spacecraft deployment.

We draw on insights from these methodologies and propose a hybrid approach that reduces computational load while implementing state-of-the-art methods for fast and accurate Object-Specific Depth Estimation, as presented in the following section.

3. PROPOSED METHODOLOGY

Taking the improvements from DC-Depth[4], Ada-Bins[13] and light-weight methods like Apple’s Depth-Pro, we propose the hybrid architecture in Fig.1: a segmentation map aided depth estimation model, which subdivides the problem

into range prediction and normalized surface depth estimation. Incoming images are processed through a lightweight network (ResNet50/EfficientNetV2) to obtain a segmentation map and global image features. Using this segmentation map, we crop out our target and resize it to 224×224 while maintaining the aspect ratio to preserve spatial information and relationships for input to the modified DC-Depth algorithm. It uses a Swin-Transformer for local features and a DCT-downsample-based feature fusion to predict depth progressively in the frequency domain, enabling smooth transitions between surfaces while preserving edge information. This provides a normalized depth map, which is upsampled to the original Segmentation crop size. Using features from the EfficientNetV2 (Global Features) and Swin-T (Local Features) (See Fig.1), we use the Ada-bins strategy to predict minimum and maximum depth as shown in Eq.1.

$$N_D = \frac{D_m - Min_D}{Max_D - Min_D} \quad (1)$$

where :

$$Max_D = Min_D + Disp.$$

$$Disp \leq Max\ Target\ Dimension$$

where N_D is the normalized depth map ($0 - 1m$), D_m being the metric Depth map, Min_D & Max_D are the minimum and maximum depth in the given depth map (D_m), respectively. The relationship between Max_D and Min_D is based upon the maximum geometrically possible distance between two points on the spacecraft, i.e, the Max_d cannot be greater than the maximum possible dimension of the target since both points, Min_d , Max_d are on the same target surface. This we call Disparity, defined as $Disp.$ in Eq.1. The maximum possible value for Disparity is the maximum possible spacecraft dimension. This allows the DC-Depth algorithm to have a uniform manifold to learn the 3D structure of the given spacecraft on a normalized scale. We train our model for the Dream Chaser spacecraft and LRO satellite separately as our method is a target-specific solution.

Dataset

To benchmark our dataset for spacecraft and objects of different scales, we also set up a Blender-based data-generation pipeline [14], which has been used in previous works for similar data-generation tasks. We generate images, segmentation maps and depth maps for training and validation, with our training set being 16,800 samples and test set at 2200 samples. Our camera intrinsic properties for all generated data were constant for the close-range and long-range datasets accordingly. To increase robustness of the algorithm, given different camera parameters for long range estimation ($30 - 150m$) and close range ($5 - 30m$), we train the same algorithm on 2 different datasets with different camera setups, with the first long range dataset based on a small FOV (long range camera) and second dataset on a wide FOV (close range camera). Both generated datasets follow the distribution of training and testing described above. Fig.2 shows examples of images, segmentation maps and depth maps generated.

4. EXPERIMENTS AND RESULTS

For a robust solution, we train the same network on 2 datasets: long-range data ($30 - 150m$) and close-range data ($5 - 30m$), with the key idea being to increase estimation performance over the two different manifolds of operation, where we can get away with lower accuracy at far ranges but require high accuracy for RPO activities. We train the same network,

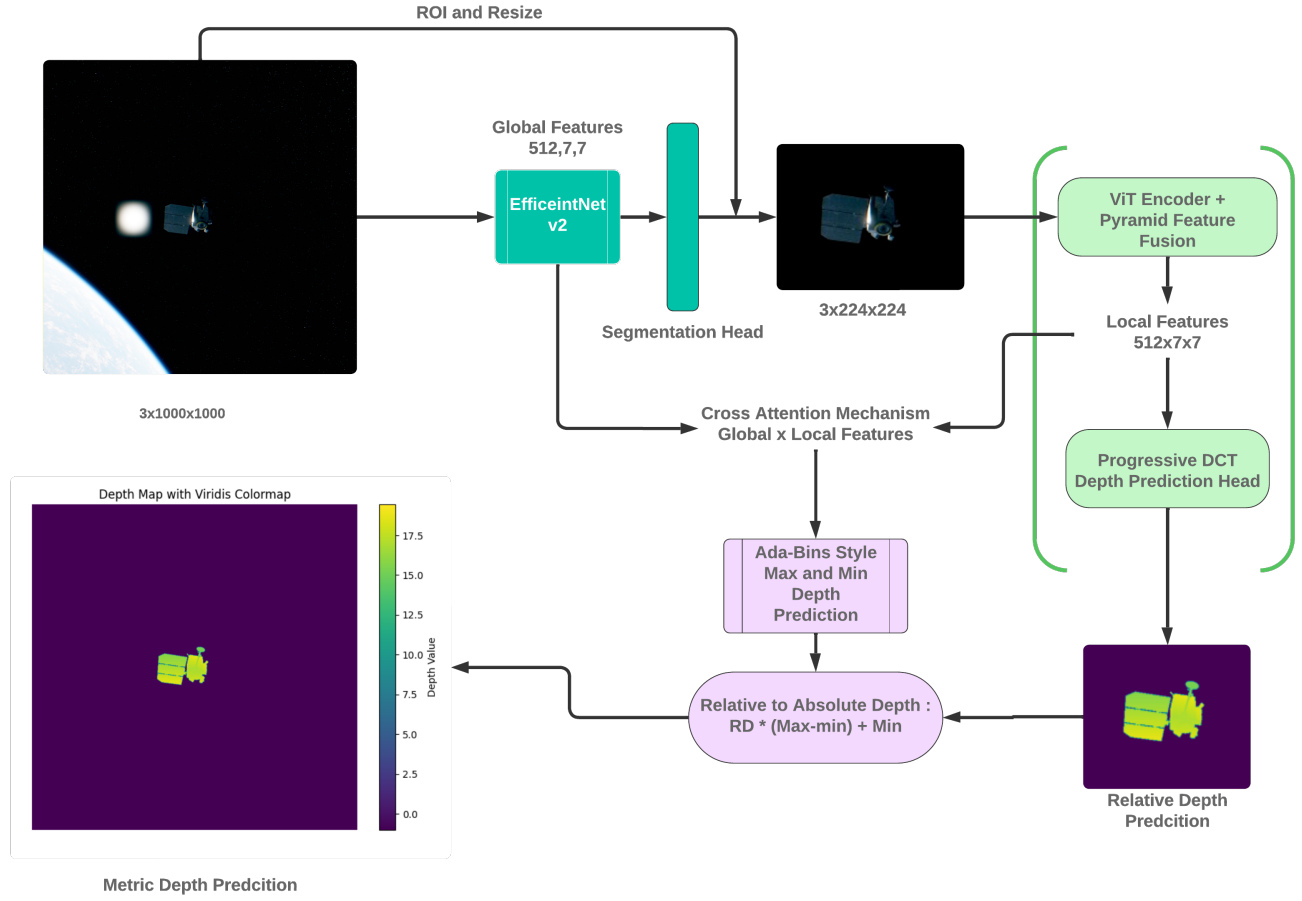


Figure 1. Proposed Method: Combining Local and Global Information for normalized depth and object range prediction (meters)

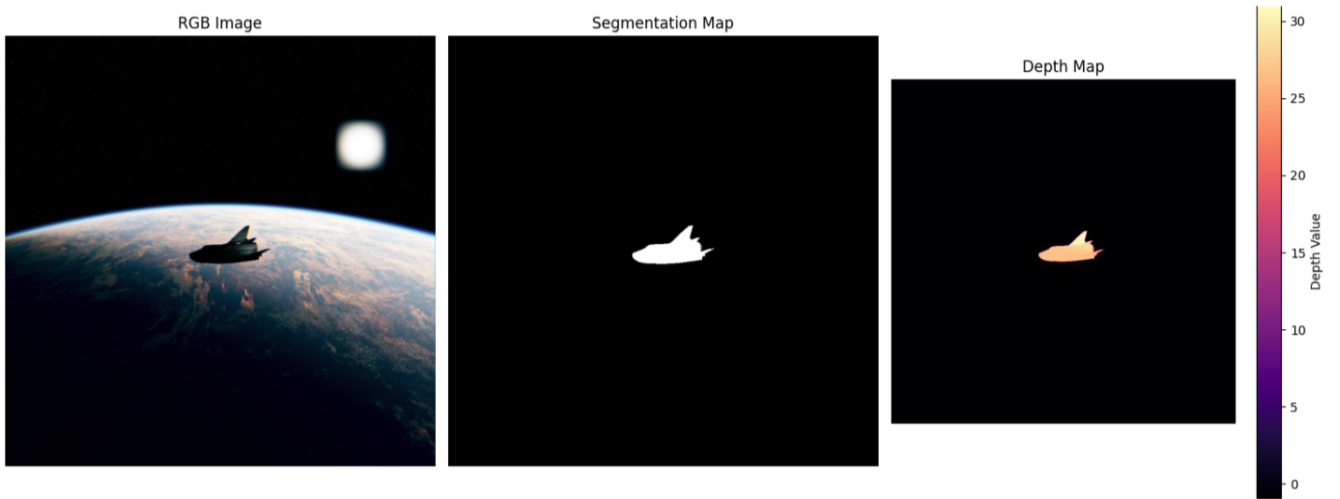


Figure 2. Data-generation : Realistic lighting conditions with added lens distortion, noise and glare for the Dream Chaser Spacecraft, extracting segmentation map, metric depth and pose

i.e., same architecture, same number of parameters, so the only difference is the weight files in use. To have a more robust and generalized network, we would have to increase both the datasets and model size to achieve this. Instead, we make a trade-off and switch the weight files only after the algorithm predicts depths in the range of $30 - 40m$. This helps us reduce compute requirements but increases storage requirements. Currently, we are operating (train, test and evaluation) at 'float32' precision, with weight files of size 440 MB. During hardware testing, we plan to make use of quantizations and compilation techniques to reduce model size further.

We discuss absolute errors in depth predictions results for:

1. Generalized model: model trained on complete range ($5 - 150m$) with the same model size, higher number of epochs and lower batch size, to help training. The model is evaluated on a random test set, generated with different lighting and background illumination;
2. Long-range Model: model trained on a subset of the complete range ($30 - 150m$);
3. Close-range Model: trained on closer ranges with different camera parameters ($5 - 30m$).

Training, testing and inference time benchmarks for all networks were done on NVIDIA RTX A5000 GPU. All 3D/point cloud Projections (Fig.3, Fig.4) are post denoising/outlier removal.

Losses

Previous methods use loss functions such as Scale-Invariant Log Loss (SI-log loss) [15] for normalized/relative depth estimation, as it creates higher fidelity losses at normalized scales due to log's sensitivity to small changes between $0 - 1$ input values. For metric depth estimation applications, losses like Mean Absolute Error (MAE), Mean Squared Error (MSE) or Root MSE (RMSE) are used for their separate advantages, where functions like MAE and RMSE are good for being sensitive to large outliers, but do not impose a threshold for good losses. The Huber Loss (Combination of MAE and MSE) allows for switching between said functions based on the threshold value.

Given that our system is split into two parts, namely the normalized depth map prediction and min-max depth prediction, we use two loss functions separately. We use the SI-log loss for our normalized prediction (as shown in Eq.2 where $\hat{y}_i$ is ground truth depth map and y_i is the predicted depth map) with an algebraic sum of Huber Loss for min and max depth separately (as in Eq.3, where $a = \max_{gt}/\min_{gt} - \max_p/\min_p$, $\max_{gt}/\min_{gt}$ are ground truth values of max or min depth and $\max_p/\min_p$ are predicted max or min values). We also add an Earth Movers Distance (L_{EMD} , in Eq.4) metric to help the predicted depth maps obey shape/structural consistency after 3D projection into point clouds. For this, we 3D project our predicted depth maps and ground truth using the available intrinsic camera matrix. The final Loss function (L_T) is shown in Eq.4 as an algebraic sum of SI loss for normalized depth prediction, Huber Loss for the max and min predicted depths and L_{EMD} for structural consistency in depth prediction.

$$L_{SI} = \frac{1}{n} \sum_{i=1}^n (\log \hat{y}_i - \log y_i)^2 - \frac{1}{n^2} \left(\sum_{i=1}^n (\log \hat{y}_i - \log y_i) \right)^2 \quad (2)$$

$$L_\delta(a) = \begin{cases} \frac{1}{2}a^2 & \text{if } |a| \leq \delta \\ \delta (|a| - \frac{1}{2}\delta) & \text{otherwise} \end{cases} \quad (3)$$

$$L_T = L_{SI} + L_\delta(a) + 0.5 * L_{EMD} \quad (4)$$

5. RESULTS

Table 1 compares performance over range brackets to see that the original approach, the generalized model, provides good depth prediction over long ranges but collapses completely due to the lack of enough data and the large dataset range for small sizes of the model. We help improve this by training the same network on a dataset of only close approaches. We see significant improvement in depth prediction performance with an average accuracy of $0.23m$ (EfficientNetv2) from $5 - 30m$ compared to the global model at $2.17m$ that trains on the whole range. We also do a close-range training and evaluation campaign for the LRO satellite, achieving an accuracy of $0.33m$ for the range of $0 - 35m$. To display the improvements of our methodology, we benchmark it against state-of-the-art real-time depth estimation algorithms, FastDepth[12], RT-MonoDepth [18], VA-DepthNet [19] algorithms.

As we can see in Table 2, *Our* algorithm outperforms lightweight networks [18], [12] over the entire range of $0 - 30$ meters. These algorithms do perform slightly better than our algorithm in the extremely close range i.e., $< 5m$, but drastically lose performance after that. *Our* approach, on the other hand, is able to maintain a stable performance of $0.3m$ error throughout the entire range of operation. Our model's increase in compute time and parameter count is also due to the larger input and output sizes we operate on, compared with the other benchmarked algorithms. We expect to see a decrease in computational requirements and an increase in inference rates when we downsize the input. We choose not to downsize the input drastically to not lose image information, as our Minimum and Maximum depth predictions are dependent on the global and local image features combined (see Fig.1, Eq.1). Results provided for VA-DepthNet[19] in Tables 3 and 2 (underlined), are numerically sound and seem like the best performing networks, but if we compare them in Fig.4 and Fig.3, we see that there is not structural understanding in the predicted depth or projected point cloud. The depth accuracy metrics do produce a state-of-the-art average for error, but when comparing ground truth to the predicted, we see the error, which is hidden by the good average metrics. This is also present because of our use of Earth Movers Distance (See Eq.4), which is an optimal transport metric, calculating the shape matching error post 3D projection. We can see the added advantages of using more structural/geometry-based loss functions, instead of simpler Root Mean Squared Error (RMSE) or Mean Absolute Error (MAE), in Fig.4 (d) and 3 (d), comparing our 3D projections versus other algorithms.

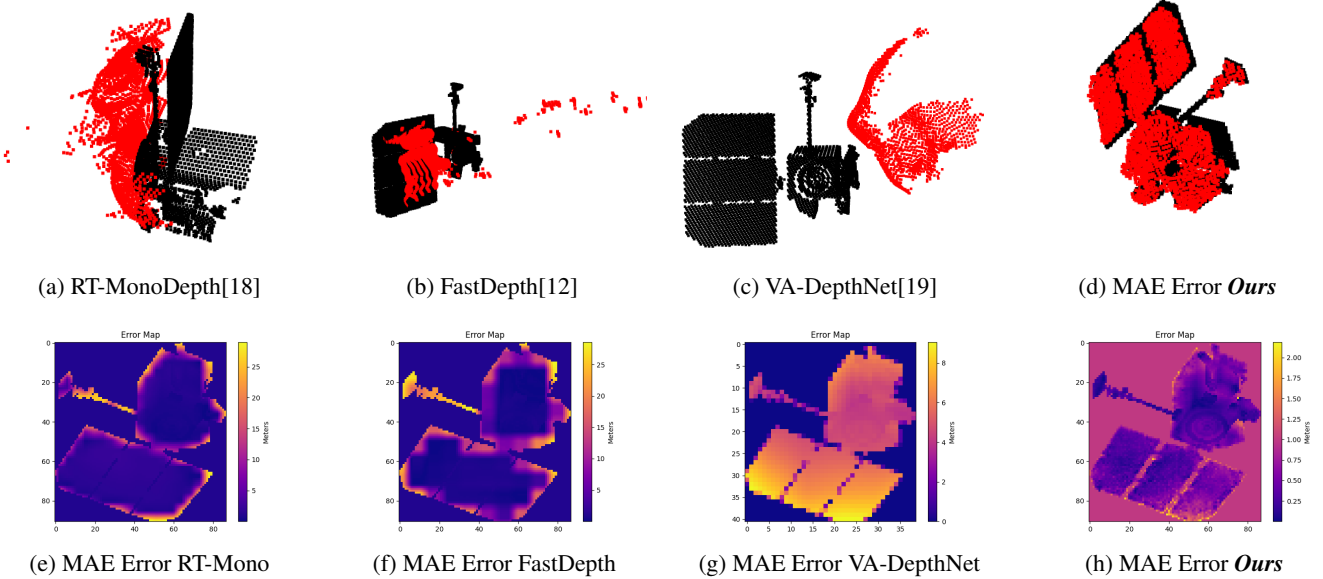
Through our 3D projected samples and depth prediction comparison in Fig.3, we can observe how our network captures the structural information of curves and flat surfaces in the

Table 1. Mean Absolute Error (meters) over different Range brackets for the Dream Chaser Spacecraft

Model	Range Brackets (meters)			
	<20	20-50	50-100	>100
Gen. Model	2.172	0.593	1.283	-
Long Rng. Model	1.437	0.650	1.148	1.558
Close Approach	<5	5-10	10-20	20-30
Close Rng. Model (ResNet[16])	0.812	0.616	0.585	0.661
Close Rng. Model (EfficientNetV2[17])	0.113	0.118	0.236	0.293

Table 2. Quantitative comparison of monocular depth estimation models for LRO Satellite

Model	Params. (M)	Input/Output size ($h \times w$)	Range Brackets MAE (m)				Average Metrics			Time (ms)
			<5	5-10	10-20	20-30	MAE	RMSE	$\delta 1$	
RT-MonoDepth[18]	2.8	224×224	0.143	0.270	0.972	2.557	1.655	3.279	0.860	10.61
FastDepth[12]	4.0	224×224	0.163	0.373	1.486	4.036	2.587	4.376	0.776	7.02
VA-DepthNet[19]	69.6	448×448	0.861	0.296	0.038	0.014	0.129	0.421	0.020	54.08
Ours	109	1000×1000	0.267	0.337	0.331	0.348	0.335	0.488	0.965	78.92

**Figure 3.** 3D projection and MAE results of the proposed method: (a)–(d) represent 3D Projections (Ground Truth = Black, Predicted = Red) and (e)–(h) are metric error maps for the same

images while predicting depth. Methods like RT-Mono, FastDepth, VA-Depth, provide good average results, but sub-figures (e) to (g) in Fig.3 show how the error is distributed, increasing closer to the edges and changes in surface, whereas in sub-figure (h) our model has an even distribution of error, inhibiting the improvements made by breaking down the process of Metric Depth prediction into Structural Depth Prediction + Min-Max Metric range prediction (See Fig.1).

In Table 3 and Fig.4, we observe that the compared networks perform better than on the LRO test set, which is partially due to the simpler structure of the Dream Chaser Spacecraft, i.e., the lack of sharp edges and multiple extruding surfaces. They still follow the general trend of performing better in extreme close ranges, i.e from $0-5m$, which can be explained

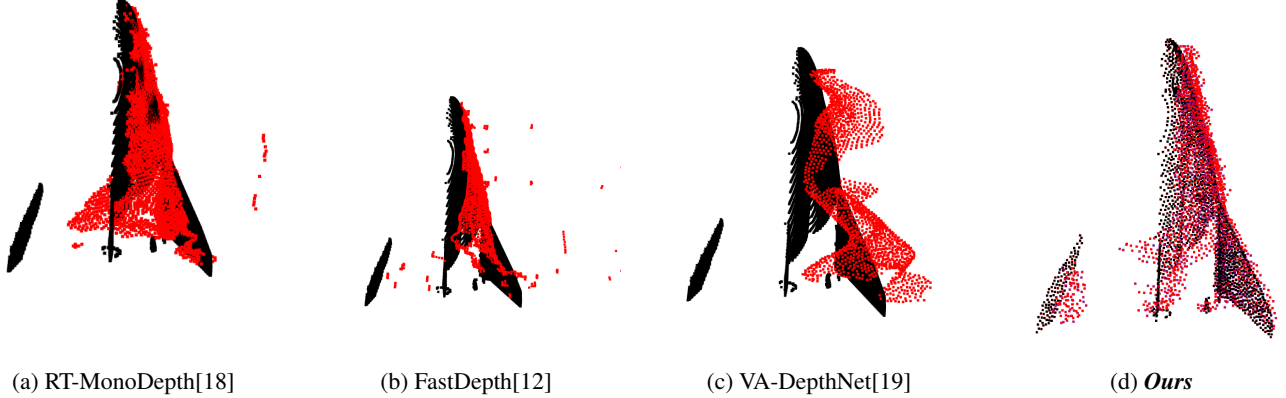
by the specific use of the Scale-Invariant Log Loss Function (see Eq.2), which prioritises and enhances losses for smaller values ($0-5m$) over larger values ($5+m$) while calculating error. Another reason for their improved performance is the greater availability of image information as the object comes closer; these algorithms have more data to regress depth. Compared to our algorithm, we see a small drop in prediction accuracy over the entire range, which can be attributed to the better use of the SI-Log Loss function due to our split methodology of normalized predictions and bin-range prediction.

Limitations

Unlike the end-to-end single-stage networks (FastDepth[12], RT-Mono[18], VA-Depth[19]), our accuracy doesn't increase

Table 3. Quantitative comparison of monocular depth estimation models for Dream Chaser Spacecraft

Model	Params. (M)	Input/Output size ($h \times w$)	Range Brackets MAE (m)				Average Metrics			Time (ms)
			<5	5-10	10-20	20-30	MAE	RMSE	$\delta 1$	
RT-MonoDepth[18]	2.8	224×224	0.063	0.118	0.452	0.9920	0.556	1.495	0.945	8.55
FastDepth[12]	4.0	224×224	0.163	0.373	1.486	4.036	0.932	2.175	0.898	7.25
VA-DepthNet[19]	69.6	448×448	0.472	0.336	<u>0.069</u>	<u>0.054</u>	<u>0.107</u>	0.421	<u>0.046</u>	54.08
Ours	109	1000×1000	0.113	0.118	0.236	0.293	0.213	0.790	0.968	78.92

**Figure 4.** 3D projection for Dream Chaser Spacecraft: (a)–(d) Ground Truth = Black, Predicted = Red

as the target gradually approaches due to our min-max prediction strategy. Our normalized structural depth predictions are still pristine, but we lose metric accuracy in the min-max predictions. This is one of the drawbacks of the adaptive binning strategy implemented from Ada-Bins[13]. Our implementation relies on the cross attention between local features (ViT features) and Global Features (EfficientNetv2 Features). However, once the target object fills the field of view (FOV), these features begin to represent the same image information across different feature dimensions. Hence, we are unable to extract any more useful information for our min-max depth prediction. This limitation, on the other hand, helps us maintain stability and increased prediction accuracy across the entire range of operations.

Currently, compared with FastDepth and RT-Mono, we are computationally more demanding, as all inference on their benchmark is performed in *float16*. In contrast, our algorithm is unoptimized and currently runs in *float32* precision. With further optimization and pruning following the DC-Depth Paper[4], the inference time can be reduced while maintaining accuracy.

6. CONCLUSION

This article presented a new logically devised architecture for spacecraft-specific metric depth estimation. Our proposed algorithm addresses the challenges faced by previous methods [4], [18], [12], [19], which exhibit a drop-off in prediction accuracy as the depth range increases. Previous methods have used an end-to-end approach to achieve computational lightness, but they suffer from accuracy losses due to this trade-off. These algorithms achieve good average results in MAE and RMSE, but when we analyse the $\delta 1$ metric and 3D projections, we observe a lack of structural understanding

and dependence on current visual information for depth prediction. On the contrary, our algorithm predicts a region of interest using global features and a segmentation map. This selected region of interest is used for actual structural depth computations, thereby reducing computational load, unlike other methodologies that pass the entire image through their neural network pipeline. With these arguments, we presented a new Metric Depth Estimation algorithm that takes into account the progress in computer vision and deep learning applications and provides an explainable, lightweight, real-time inference depth-prediction architecture for space rendezvous and proximity operations.

Future Work

Future work revolves around designing filters for our proposed network switching algorithm, optimization and quantization of our proposed methodology for embedded applications. We plan to test this algorithm with Point Cloud Pose Estimation pipelines, to prove a non-black box, real-time, multi-stage spacecraft pose estimation algorithm, which can further be integrated into a guidance and control architecture.

ACKNOWLEDGEMENTS

This research was supported by the Inha University funded project Precision Guidance and Navigation for Deep Space Exploration (project number P20962).

REFERENCES

- [1] C. Godard, O. M. Aodha, and G. J. Brostow, “Unsupervised monocular depth estimation with left-

- right consistency,” *CoRR*, vol. abs/1609.03677, 2016. [Online]. Available: <http://arxiv.org/abs/1609.03677>
- [2] R. Ranftl, K. Lasinger, D. Hafner, K. Schindler, and V. Koltun, “Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer,” 2020. [Online]. Available: <https://arxiv.org/abs/1907.01341>
- [3] S. F. Bhat, R. Birkel, D. Wofk, P. Wonka, and M. Müller, “Zoedepth: Zero-shot transfer by combining relative and metric depth,” 2023. [Online]. Available: <https://arxiv.org/abs/2302.12288>
- [4] K. Wang, Z. Yan, J. Fan, W. Zhu, X. Li, J. Li, and J. Yang, “Dcdepth: Progressive monocular depth estimation in discrete cosine domain,” 2024. [Online]. Available: <https://arxiv.org/abs/2410.14980>
- [5] A. Bochkovskii, A. Delaunoy, H. Germain, M. Santos, Y. Zhou, S. R. Richter, and V. Koltun, “Depth pro: Sharp monocular metric depth in less than a second,” 2024. [Online]. Available: <https://arxiv.org/abs/2410.02073>
- [6] B. Ke, K. Qu, T. Wang, N. Metzger, S. Huang, B. Li, A. Obukhov, and K. Schindler, “Marigold: Affordable adaptation of diffusion-based image generators for image analysis,” 2025. [Online]. Available: <https://arxiv.org/abs/2505.09358>
- [7] B. Gabdullin, N. Konovalova, N. Patakin, D. Senushkin, and A. Konushin, “Depthart: Monocular depth estimation as autoregressive refinement task,” 2024. [Online]. Available: <https://arxiv.org/abs/2409.15010>
- [8] W. Yuan, X. Gu, Z. Dai, S. Zhu, and P. Tan, “New crfs: Neural window fully-connected crfs for monocular depth estimation,” 2022. [Online]. Available: <https://arxiv.org/abs/2203.01502>
- [9] AITECH. Aitech systems. [Online]. Available: <https://aitechsystems.com/>
- [10] N. Zhang, F. Nex, G. Vosselman, and N. Kerle, “Lite-mono: A lightweight cnn and transformer architecture for self-supervised monocular depth estimation,” 2023. [Online]. Available: <https://arxiv.org/abs/2211.13202>
- [11] D.-J. Lee, J. Y. Lee, H. Shon, E. Yi, Y.-H. Park, S.-S. Cho, and J. Kim, “Lightweight monocular depth estimation via token-sharing transformer,” 2023. [Online]. Available: <https://arxiv.org/abs/2306.05682>
- [12] Wofk, Diana and Ma, Fangchang and Yang, Tien-Ju and Karaman, Sertac and Sze, Vivienne, “FastDepth: Fast Monocular Depth Estimation on Embedded Systems,” in *IEEE International Conference on Robotics and Automation (ICRA)*, 2019.
- [13] S. Farooq Bhat, I. Alhashim, and P. Wonka, “Adabins: Depth estimation using adaptive bins,” in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Jun. 2021, p. 4008–4017. [Online]. Available: <http://dx.doi.org/10.1109/CVPR46437.2021.00400>
- [14] L. F. A. T. Siddharth Singh, Hyo Sang Shin, “Point cloud completion based pose estimation for spacecraft,” in *11th European Conference for AeroSpace Sciences (EUCASS 2025)*, 2025.
- [15] D. Eigen, C. Puhrsch, and R. Fergus, “Depth map prediction from a single image using a multi-scale deep network,” 2014. [Online]. Available: <https://arxiv.org/abs/1406.2283>
- [16] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” 2015. [Online]. Available: <https://arxiv.org/abs/1512.03385>
- [17] M. Tan and Q. V. Le, “Efficientnetv2: Smaller models and faster training,” 2021. [Online]. Available: <https://arxiv.org/abs/2104.00298>
- [18] C. Feng, C. Zhang, Z. Chen, W. Hu, and L. Ge, “Real-time monocular depth estimation on embedded systems,” 2024. [Online]. Available: <https://arxiv.org/abs/2308.10569>
- [19] C. Liu, S. Kumar, S. Gu, R. Timofte, and L. V. Gool, “Va-depthnet: A variational approach to single image depth prediction,” 2023. [Online]. Available: <https://arxiv.org/abs/2302.06556>

BIOGRAPHY



Siddharth Singh received his MSc in Applied Artificial Intelligence at Cranfield University, United Kingdom with a focus on vision based estimation and prediction systems. Currently, he is a PhD student at Cranfield University with research interests in autonomous space rendezvous and proximity operations using vision systems.



Prof. Hyo Sang Shin received the BSc in aerospace engineering from Pusan National University, Busan, South Korea, in 2004, the MSc degree in aerospace engineering with focus on flight dynamics, guidance, and control from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2006, and the PhD degree in cooperative missile guidance from Cranfield University, Cranfield, U.K., in 2011. Currently the Head of Autonomous and Intelligent Systems Group and a Professor of Guidance, Navigation, and Control with Cranfield University and a Professor in Cho Chun Shik Graduate School of Mobility, KAIST, Daejeon 34141, Republic of Korea with interests/expertise in multiple target tracking, probabilistic target detection, and distributed control of multiple agent systems



Dr. Leonard Felicetti received his M.Sc. *summa cum laude* from Sapienza - University of Rome (Italy) in 2008 with a dissertation on the Attitude Dynamics and Control of a Cubesat implementing a Drag-Balance Instrument for measuring the Atmospheric Drag in Low Earth Orbit. He received his Ph.D. in the 'Dynamics, Guidance and Control of Multiple Platform Space Systems' in 2013. Currently he is a Reader in Space Robotics and GNC and Course Director of the MSc in Astronautics and Space Engineering at Cranfield University. His research activity focuses on Space Robotics, Attitude and Orbit Control Systems (AOCS) and Guidance, Navigation and Control (GNC) of Space Systems.



Dr. Antonios Tsourdos obtained an MEng on Electronic, Control and Systems Engineering from the University of Sheffield (1995), an MSc on Systems Engineering from Cardiff University (1996) and a PhD on Nonlinear Robust Flight Control Design and Analysis from Cranfield University (1999). He is a Professor of Autonomous Systems and Control with Cranfield University and Head of the Centre of Autonomous and Cyber-Physical Centre. He is a fellow of the Royal Aeronautical Society. Professor Tsourdos' research is focused on guidance, control and navigation for single and multiple autonomous vehicles as well as mission planning and multi-agent systems.



DSPACE

<https://dspace.org/>

Monocular depth estimation for spacecraft: combining relative and scale information

Singh, Siddharth; Shin, Hyo-Sang; Felicetti, Leonard; Tsourdos, Antonios

2026-03-07

Institute of Electrical and Electronics Engineers (IEEE)

<https://doi.org/10.1109/aero66936.2026.11519926>

Singh S, Shin HS, Felicetti L, Tsourdos A. (2026) Monocular depth estimation for spacecraft: combining relative and scale information. In: Proceedings of the 2026 IEEE Aerospace Conference, 7-14 Mar 2026, Big Sky, USA