

UAVD4L: A Large-Scale Dataset for UAV 6-DoF Localization

Anonymous 3DV submission

Paper ID 298

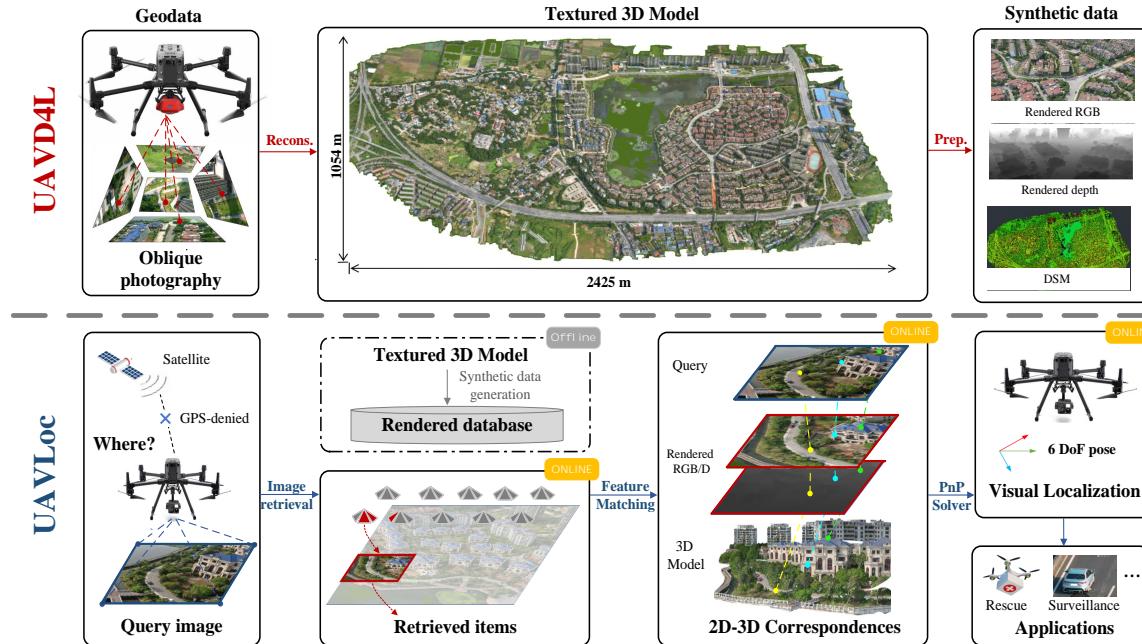


Figure 1. Top. We introduce a large-scale dataset for the 6-DoF localization of UAVs. The dataset includes a 3D reference textured model, which enables the generation of synthetic data such as rendered RGB and depth images, as well as a Digital Surface Model (DSM). **Bottom.** We also develop an offline-and-online pipeline for performing 6-DoF localization of UAVs in GPS-denied environments.

Abstract

Despite significant progress in global localization of Unmanned Aerial Vehicles (UAVs) in GPS-denied environments, existing methods remain constrained by the availability of datasets. Current datasets often focus on small-scale scenes and lack viewpoint variability, accurate ground truth (GT) pose, and UAV build-in sensor data. To address these limitations, we introduce a large-scale 6-DoF UAV dataset for localization (UAVD4L) and develop a two-stage 6-DoF localization pipeline (UAVLoc), which consists of offline synthetic data generation and online visual localization. Additionally, based on the 6-DoF estimator, we design a hierarchical system for tracking ground target in 3D space. Experimental results on the new dataset demonstrate

the effectiveness of the proposed approach. The code and dataset will be publicly released upon the publish of the paper.

014
015
016

1. Introduction

017

Global localization of Unmanned Aerial Vehicles (UAVs) plays an important role in various applications such as cargo transport [25], surveillance [18], and search-rescue tasks [5, 41]. While most established works rely on global satellite navigation systems (GNSS) [3, 11, 28] for global position, vulnerabilities in GNSS signal reception [1, 22, 45] have led to the development of visual solutions [32] for GPS-denied environments.

018
019
020
021
022
023
024
025
026

However, compared with ground-level localization of

027 cellphones or VR devices [34, 36, 38], the development of
028 visual localization of UAVs lags behind. We believe this is
029 partly due to the scarcity of open-source datasets for localizing
030 airborne platforms in the academic community. Among
031 the most evaluated open-source datasets, limitations include
032 restricted to small scenes [10], evaluation of only 3-DoF pos-
033 itions [53], limited viewpoint variability [16], lack of accu-
034 rate ground truth (GT) [19], concentration on learning-
035 based regression methods [47], and disregard for additional
036 inertial sensors [29].

037 To facilitate research in this area, our first contribution is
038 to introduce a **novel large-scale dataset for 6-DoF UAV**
039 **localization**, as shown in the top part of Figure 1. The
040 dataset comprises a textured 3D reference model recon-
041 structed from aerial oblique photography, covering approx-
042 imately 2.5 million square meters. It enables the genera-
043 tion of various synthetic data, including rendered RGB and
044 depth images, and a Digital Surface Map (DSM). Query im-
045 ages are captured and sampled from five distinct flight tra-
046 jectories, with varying heights (50-200 meters), viewpoints
047 (pitch angles of 15-70 degrees), and acquisition positions.
048 Rather than relying on built-in sensor information from the
049 UAV, such as Global Positioning System(GPS) or Real-time
050 kinematic(RTK), to provide 3-DoF ground truth positions,
051 we use manual tie points to register query sequences to the
052 reference map for more accurate 6-DoF ground truth poses.
053 In addition, sensor data (i.e., the rotation priors from the
054 Inertial Measurement Unit (IMU)) is recorded to assist further
055 localization algorithms.

056 Second, we develop a new **two-stages 6-DoF UAV lo-**
057 **calization pipeline for GPS-denied environments**, which
058 consists of offline synthetic data generation and online vi-
059 sual localization, as illustrated in the bottom portion of Fig-
060 ure 1. In the offline stage, our approach employs render-
061 ing techniques to obtain synthetic views and depth maps at
062 different virtual viewpoints (with varying heights and direc-
063 tions) to fully represent the dense 3D model of the scene. In
064 the online stage, inspired by the recent ground-level visual
065 localization method, SensLoc [48], we combine camera in-
066 formation with prior data from UAV equipment. Specif-
067 ically, we first use rotation information from the build-in
068 IMU to constrain search space for identifying relevant ref-
069 erence views of the query image. Then, feature matches
070 between the query image and the top- k retrieved database
071 images are established and lifted to 2D-3D correspondences
072 with the help of depthmaps. Finally, a gravity-guided PnP
073 RANSAC [7, 21] is employed to estimate the camera pose.

074 Finally, based on the 6-DoF localization results of UAVs,
075 we design a **hierarchical system to track designated ob-**
076 **jects located on the ground**, such as pedestrians or vehi-
077 cles. This system employs a wide-angle lens (offering a
078 larger field of view) to determine the 6-DoF pose of the
079 UAV, and a zoom lens to identify targets, facilitating more

080 accurate target extraction on the 2D image. A ray-tracing
081 technique [21] is utilized to project 2D targets on image
082 plane onto their absolute positions in the 3D map. h

2. Related Work

2.1. Ground-level Visual Localization

083 Conventionally, most established methods [34, 36, 37, 48]
084 perform ground-level visual localization by establishing
085 2D-3D correspondences between the query image and the
086 reconstructed sparse point clouds. The camera pose is then
087 recovered using a PnP solver [20, 26] with RANSAC [14].
088 To scale to large scenes, an image retrieval phase [2, 15, 17]
089 is often used to identify a small set of potentially relevant
090 database images. Recently, HLoc [36, 37] has provided
091 a comprehensive toolbox that integrates existing methods
092 and achieves a promising result. To address the robust-
093 ness, efficiency, and flexibility issues of HLoc, subsequent
094 research includes replacing the sparse point clouds with
095 dense model [34], and introducing sensor prior informa-
096 tion [39, 48].

097 Inspired by these works, we propose a two-stage pipeline
098 that effectively explores synthetic view representation of the
099 dense model and successfully combines sensor prior with
100 visual signal for localizing the 6-DoF pose of UAVs.
101

2.2. GPS-denied UAV Localization

102 Compared with ground-level visual localization, pose esti-
103 mation of UAVs in GPS-denied environments has received
104 less attention. Some previous works [44, 46, 49, 50] treat
105 UAV localization as a scene recognition task, where the
106 goal is to identify the most similar images to a query im-
107 age within a geo-tagged database. The performance of this
108 approach is influenced by the density and distribution of im-
109 ages within the database and can result in significant errors.
110 Other works [6, 9, 27, 33, 47] aim to calculate the 6-DoF
111 camera pose and can be broadly classified into two cate-
112 gories. The first category [40, 47] involves using neural net-
113 works to implicitly encode the scene and regress an absolute
114 pose for an input image. These methods have the advantage
115 of occupying small memory and having fast inference time,
116 but their localization accuracy is limited. The second cat-
117 egory [8] follows HLoc [36, 37] to explicitly use the map
118 to render synthetic views upon Google Earth and recovers
119 6-DoF pose via image retrieval, feature matching and PnP
120 RANSAC. Our work is similar to the second one, but we
121 explore richer rendering viewpoints to represent the scene,
122 and make full use of sensor priors to achieve more accurate
123 6-DoF localization of UAVs.

2.3. UAV Localization Datasets

124 The development of UAV localization research is largely
125 driven by the availability of datasets, as summarized in Ta-
126

Name	Reference source	UAV viewpoint	Map size	Estimated pose DoF	Ground truth	Open source
Chen et al. [8]	Google Earth	arbitrary	$400 \times 400 m^2$	6-DoF	GPS and IMU	no
Goforth et al. [16]	United States Earth	top view	$0.85 km$	6-DoF	RTK-GPS and IMU	yes
Kinnari et al. [23]	georeferenced orthophotos	arbitrary	$200 \times 200 m^2$	6-DoF	RTK-GPS and IMU	no
Kinnari et al. [24]	Google Earth	arbitrary	-	3-DoF (x,y,heading)	RTK-GPS and IMU	no
Patel et al. [35]	Google Earth	top view	$300 \times 350 m^2$	4-DoF (x,y,z,heading)	RTK-GPS and IMU	no
Gurgu et al. [19]	Google Earth	top view	-	2-DoF (x,y)	RTK-GPS	yes
Zheng et al. [53]	Google Earth and Google Map	arbitrary	-	2-DoF (x,y)	sim from Google Earth	yes
Yan et al. [47]	render and real geographic image	arbitrary	$2.7 \text{ million } m^2$	6-DoF	RTK-GPS and ground control point	yes
Cisneros et al. [10]	LiDAR, real image	arbitrary	$50 \times 50 m^2$	6-DoF	RTK-GPS and ground control point	yes
Ours	3D model, geographic image, DSM	arbitrary	$2.5 \text{ million } m^2$	6-DoF	RTK-GPS and ground control point	yes

Table 1. **Overview of the existing UAV localization datasets.** Several key attributes are taken into account: the source of reference data, the viewpoint, degrees of freedom used to measure the estimated pose, how to acquire ground truth data, and whether or not the dataset is open source.

129 However, existing UAV datasets face several challenges that limit their effectiveness in advancing the field.
130 These challenges include: 1) **Limited Degrees of Freedom.** Many datasets, such as University-1652 [53] and
131 ALTO [10], provide only 3 degrees of freedom (DoF) for
132 localization, which may not be sufficient for certain applications.
133 2) **Top-View Bias.** Some datasets, such as those
134 presented in [16] and [19], focus exclusively on top-view
135 images generated from Google Earth or real drones. How-
136 ever, these datasets do not represent the diverse scenarios of
137 real-world UAV usage.
138 3) **Regression-Based Approaches.** Datasets such as CrossLoc [47] provide rich data sources in-
139 cluding real images, synthetic images, and semantic information
140 for regression-based localization approaches. However,
141 these approaches may not be suitable for structure-
142 based localization algorithms.
143

144 These challenges highlight the need for more comprehensive and diverse datasets to advance the field of UAV
145 localization.
146

3. Dataset

149 We present a large-scale dataset, UAVD4L, that covers an
150 area of approximately 2.5 million square meters and in-
151 cludes a diverse range of urban and rural scenes, including
152 buildings, streets, vegetation, and a lake. A visualization of
153 the dataset is provided in Figure 2.

3.1. Reference Map Collection

154 To create the reference map, we captured high-resolution
155 aerial images $\{I_o\}$ using a DJI M300 RTK drone¹ equipped
156 with a five-eye oblique SHARE PSDK 102s camera². To
157 ensure sufficient and uniform coverage of the area, we pre-
158 planned a grid fashion flight path, which was automatically
159 executed by the drone’s flight control system. We then ap-
160 plied modern 3D reconstruction techniques to produce tex-
161 tured mesh models and align them with the real geographic
162 coordinate system using built-in RTK measurements and
163



Figure 2. **Distribution of query images.** The query sequence consists of five trajectories, with the first four covering urban areas characterized by a high density of buildings, while the fifth trajectory covers a rural area with predominantly vegetation. The red locators, numbered from 1 to 5, represent the shooting positions.

164 ground control points. More details about reference map
165 collection is provided in the supplementary materials.

3.2. Query Image Collection

166 We collect all query images $\{I_q\}$ using a DJI M300 RTK
167 mounted with a DJI H20T camera³. To enhance the diver-
168 sity of query images, we manually captured images cross
169 different regions, flight altitudes and a diverse range of cap-
170 turing angles. In addition, to simulate the different shoot-
171 ing habits of drone pilots, we captured query images using
172 two modes: continuous shooting at 2-second intervals and
173 random manual shooting. The supplementary material pro-
174 vides statistics on the number of query images for each tra-
175 jectory.
176

3.3. Query GT Generation

177 In this work, we adopt a scalable semi-automatic annotation
178 method to generate pseudo GT poses for query images. Our
179 method is capable of producing pose annotations for hun-
180 dreds of query images with minimal human intervention.
181 The main steps of the method are as follows: First, we use
182 the Structure-from-Motion (SfM) method to separately re-

¹<https://enterprise.dji.com/cn/matrice-300>

²<https://shareuav.cn/V3S>

³<https://enterprise.dji.com/cn/zenmuse-h20-series>

184 construct sparse point clouds from the oblique photographic
 185 image $\{I_o\}$ and multiple query sequences $\{I_q\}$. Next, we
 186 manually select some tie-points between query sequence
 187 and the oblique photographic images and perform registra-
 188 tion based on these tie-points. Finally, we refine SfM block
 189 using bundle adjustment to achieve a whole 3D model with
 190 oblique photographic and query images.



Figure 3. **GT poses quality on UAVD4L.** Pixel-aligned renderings of the estimated camera pose confirm that the poses are sufficiently accurate for evaluation.

191 To evaluate the accuracy of the reference 3D model \mathcal{M} ,
 192 we report the median reprojection error and the median re-
 193 projection error per tie-point, which are 0.82 pixels and 0.62
 194 pixels, respectively. Besides, to ensure the accuracy of the
 195 pseudo GT pose, we conduct a visual inspection, and se-
 196 lect the images whose pseudo GT pose rendered image is
 197 aligned with the original query image. Figure 3 illuminates
 198 an example where the pixel-alignment between rendered
 199 and original query image indicates that the GT poses are
 200 accurate.

201 4. Method

202 Given a textured reference 3D model \mathcal{M} with real-world
 203 geographical coordinates and the sensor information S_q of
 204 query image I_q , our goal is to estimate the 6-Dof pose ζ_q of
 205 the query image I_q . To achieve this, we propose a two-stage
 206 localization pipeline. In the first stage, the database images
 207 are rendered offline according to carefully selected virtual
 208 viewpoints (Section 4.1). In the second stage, we conduct
 209 a 6-DoF localization pipeline online (Section 4.2). Figure 4
 210 provides an overview of the proposed method. Furthermore,
 211 based on the results of UAV 6-DoF localization, we design
 212 a hierarchical system capable of tracking ground targets in
 213 3D space (Section 4.3).

214 4.1. Synthetic Data Generation.

215 Unlike ground-level localization situations, query image I_q
 216 and oblique images $\{I_o\}$ are often captured at drastically
 217 varying positions and directions, posing a significant chal-
 218 lenge for image retrieval [2, 15, 17] and feature matching al-
 219 gorithms [4, 31, 42]. To address this challenge, inspired by
 220 view synthesis work [30, 43, 47, 51, 52], we generate com-
 221 prehensive synthetic reference data, including RGB images
 222 $\{I_r\}$ and depth maps $\{D_r\}$, to represent the dense map of
 223 the scene.

224 Specifically, we first leverage the geographical boundary
 225 of 3D map \mathcal{M} to identify the area where view synthesis is
 226 required. Then the synthetic data generation can be divided
 227 into two parts: 1) For translation, we arrange the virtual
 228 viewpoints ζ above the 3D map and establish two hyperpa-
 229 rameters for horizontal interval distance of α_t and vertical
 230 height of H , which are dynamically adjusted to ensure a
 231 uniform coverage of the entire geographic region. 2) For
 232 rotation, we set multiple θ_{pitch} and an appropriate interval
 233 α_{theta} of yaw to cover 360° at the same position. Based
 234 multi view and multi height render method, it can enhance
 235 adaptation to the scale invariance and rotational invariance
 236 of query images.

237 4.2. Visual Localization

238 **Rotation-guided image retrieval.** Given a query image
 239 I_q , we aim to identify a set of commonly visible images
 240 ${}^s\mathcal{I} = \{I_{r1}, I_{r2}, \dots, I_{rk}\}$ within the reference images $\{I_r\}$,
 241 where k represents the number of retrieved items. A gen-
 242 eral solution involves mapping the images $I_q, \{I_r\}$ into a
 243 compact feature space using an embedding function $f(\cdot)$,
 244 followed by searching for the nearest neighbors of I_q us-
 245 ing the distance metric $d(f(I_q), f(I_r))$. However, simply
 246 using a global feature to find retrieved items may lead to
 247 some mistakes, as different UAV images share significant
 248 viewpoint and scale changes.

249 To improve the retrieval accuracy, inspired by
 250 SensLoc [48], we use the metadata from UAV sensors
 251 S , especially the rotation information, as a prior to pre-
 252 filter incorrect reference candidates. Considering a richer
 253 variety of viewpoints in drones, we impose constraints on
 254 each degree of rotation angle, consisting of θ_{roll} , θ_{yaw} ,
 255 θ_{pitch} , as shown in Equation 1.

$$\begin{aligned} & \{ || -\arcsin(R_{31}^S) - \theta_{roll}^r || \leq \gamma_o, \\ & \quad || \arctan(\frac{R_{21}^S}{R_{11}^S}) - \theta_{yaw}^r || \leq \gamma_o, \\ & \quad || \arctan(\frac{R_{32}^S}{R_{33}^S}) - \theta_{pitch}^r || \leq \gamma_o, \\ & \quad R_{31}^S \neq \pm 1 \}, \end{aligned} \quad (1)$$

256 where γ_o is a orientation threshold and R^S represents the

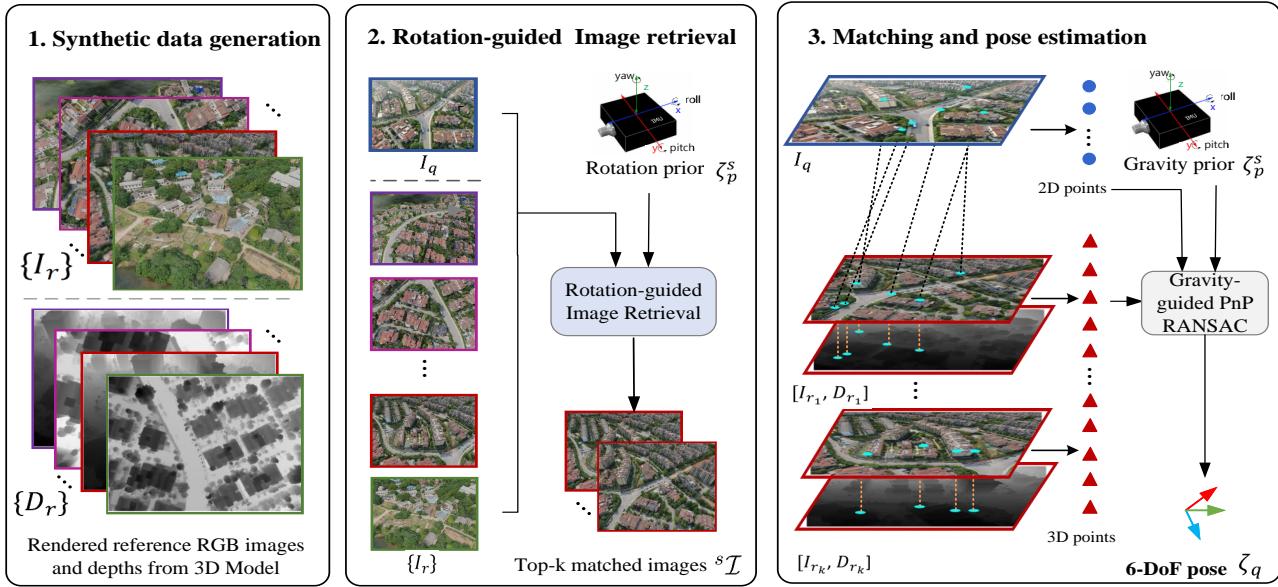


Figure 4. **Overview of the proposed method.** 1. We generate comprehensive synthetic data from textured 3D model, including RGB images I_r and depthmaps D_r (Section 4.1). 2. For each query image I_q , we use an image retrieval algorithm combined with rotation sensor prior S to find the top- k relevant images. 3. Then, we apply feature detection and matching algorithm to establish the 2D-3D correspondence between the query image I_q and the relevant images I_r . A gravity-guided PnP RANSAC is used to obtain the pose ζ_q of the UAV (Section 4.2).

rotation matrix from the sensor prior S . θ^r represents the rotation angle of the reference image. After pre-filtering incorrect candidate data using rotation information as a prior, we are able to efficiently and accurately determine k nearest neighbor ${}^s\mathcal{I}$ according to $d(f(I_q), f({}^s\mathcal{I}))$.

Matching and pose estimation. We adopt learnable feature matching techniques to establish 2D-2D correspondences between the query image I_q and the retrieved synthetic images ${}^s\mathcal{I}$. Since depthmaps ${}^s\mathcal{D}$ are also rendered during the offline stage of synthetic data generation, these 2D-2D correspondences can be lifted to 2D-3D correspondences via back-projection.

For pose estimation, we follow SensLoc [48] and incorporate a gravity verification module into the PnP RANSAC process. During each RANSAC iteration, we compute the deviation d_ϵ between the gravity direction of the sensor pose ζ_s^g and the hypothetical pose ζ_{hyp}^g , as shown in Equation 2. If this deviation d_ϵ is less than a predefined stopping threshold γ_ϵ , we halt the RANSAC iterations prematurely.

$$d_\epsilon = \arccos(\zeta_s^g \cdot \zeta_{hyp}^g) \quad (2)$$

4.3. Hierarchical Target Tracking

Based on the proposed 6-DoF UAV pose estimation approach mentioned above, we design a hierarchical system

for target tracking on the ground, as illustrated in Figure 5.

The system comprises two cameras: a fixed wide-angle lens camera and a zoom lens camera, which are carefully calibrated. The wide-angle lens, with a broader vision for global and local feature extraction, captures images for online 6-DoF localization. The zoom lens captures images for specific target detection, where the target occupies a sufficient area. Afterwards, the 3D position of the target is estimated using ray-tracing [7, 21] based on a digital elevation model (DEM), as detailed in the supplementary material.

4.4. Implementation Details

In the offline synthetic reference data generation process, we generate virtual viewpoints with absolute altitude $H = 100m$ and $150m$, the horizontal interval $\alpha_t = 50m$ and $\alpha_t = 75m$, respectively, and generate 16 images with different angles ($\theta_{pitch}^r = 0^\circ$ or 45° , $\alpha_{theta} = 45^\circ$) for each same position. In the image retrieval step, the angle threshold γ_o is set to 30° . In the gravity-guided PnP solver step, we set the gravity angle threshold $\gamma_\epsilon = 2^\circ$. We perform all experiments on a PC equipped with Intel Core i9-11900K processor, GTX 3090 graphics card (24 GB RAM) and Ubuntu 18.04 operating system. We implement the proposed pipeline in Python alone with Pytorch.

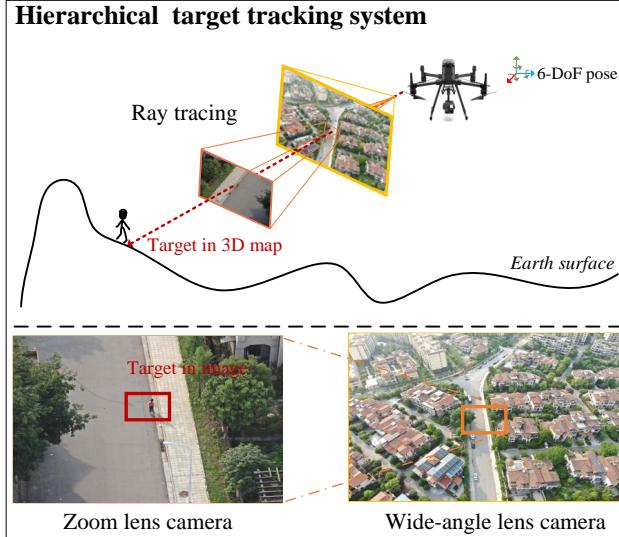


Figure 5. **The hierarchical target tracking system** consists of two lens: a wide-angle lens camera and a zoom lens camera. The wide-angle lens camera is used to recover the 6-DoF pose of the UAV, while the zoom lens camera is used to accurately detect the target. A ray-tracking technique is employed to track designated targets in 3D space.

5. Experiment

5.1. Image Retrieval

Baselines. In this study, we evaluate the performance of several global descriptors at image retrieval stage, including NetVLAD [2], AP-GeM [17], and OpenIBL [15]. Additionally, we conduct controlled experiments to verify the effectiveness of using a rotation sensor prior.

Evaluation protocol. Retrieval results are considered correct if they share a sufficient overlapping area with the query image, where the overlap percentage $P_{overlap}$ is greater than 50%. The overlap area between two images is calculated by transforming pixels from one image to the other based on their depth maps. Please refer to the supplementary materials for more details.

Results. The retrieval results are shown in Table 2. We maintain a rank list of size k for each query and report recall and precision metrics. The experimental results show that all three representative retrieval methods are capable of finding correct retrieved items, even when the rotation sensor is not adapted. This can be attributed to the generation of comprehensive virtual views across the scene, which contributes to the availability of more visual similar images. Among all options, OpenIBL [15] with rank list $k = 3$ and employing sensor priors leads the benchmark, and is chosen

to find a visible reference set in the localization pipeline.

Global feature	Prior	R@1	R@3	P@3
NetVLAD [2]	-	73.6	98.08	74.42
	Rot	79.9	98.08	77.8
AP-GeM [17]	-	84.62	96.5	72.55
	Rot	85.31	97.55	73.14
OpenIBL [15]	-	83.74	98.43	76.69
	Rot	84.44	98.78	79.08

Table 2. **Image retrieval results.** We report the top- k recall and precision for $k = 1, 3$, using global image features and considering the use of rotation priors.

328
329
330
331
332
333
334
335
336

5.2. Visual Localization

Baselines. We evaluate the performance of various key-point descriptors, including SIFT [12] and Superpoint [13], as well as matchers such as Nearest Neighbor and the learnable SuperGlue [37]. We also compare the results with those obtained using the detector-free matcher LoftR [42]. For pose estimation, we employ the PnP RANSAC solver, with and without gravity information.

337
338
339
339

Evaluation protocol. Our evaluation follows the standard localization benchmark protocol [38, 47], reporting results under the thresholds of $(1m, 1^\circ)$, $(3m, 3^\circ)$ and $(5m, 5^\circ)$.

340
341
342
343
344
345
346
347
348
349
350
351
352

Results. The quantitative results are presented in Table 3. The Superpoint+Superglue outperforms other counterparts in all metric except $(1m, 1^\circ)$. This may be attributed to the fact that query images of UAVD4L were taken under favorable lighting conditions and feature rich textures that facilitate keypoint detection by SuperPoint. Additionally, we found that the gravity-guided PnP RANSAC method [48] enhances localization accuracy regardless of the matching algorithm used. Qualitative results can be found in Figure 6 and Figure 7. Figure 6 provide the feature matching results, while Figure 7 provides a comparison between rendered estimated pose images and their corresponding original images, further demonstrating the precision of our results.

340
341
342
343
344
345
346
347
348
349
350
351
352

5.3. Target Tracking

To evaluate the target tracking performance of the proposed method, we capture people and car from two trajectories, and using the data recorded by RTK as their GT position.

353
354
355
356

Evaluation protocol. We adopt the translation difference between the ground truth 3D coordinates and the computed 3D coordinates to represent the target tracking estimation error.

357
358
359
360

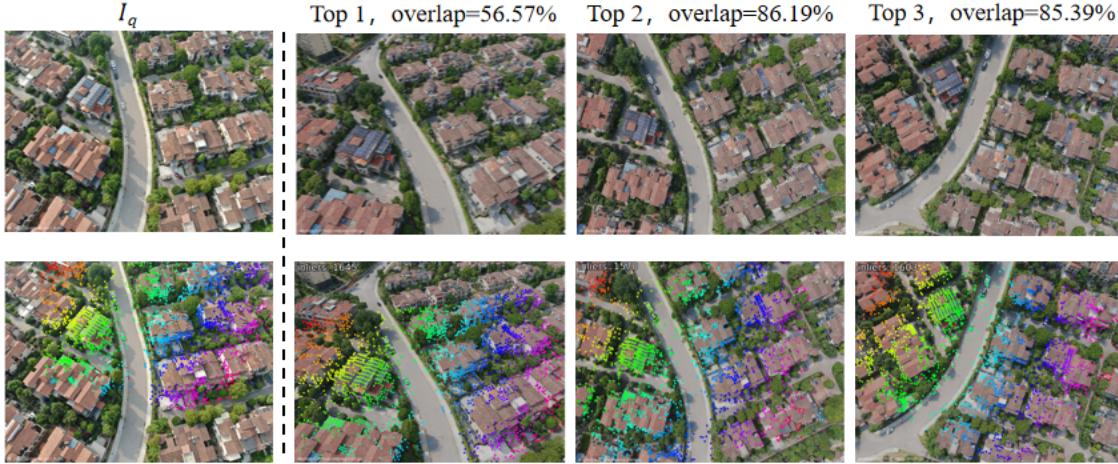


Figure 6. **Qualitative results.** The top part of the figure illustrates an example retrieval result using OpenIBL. The bottom part displays the local feature matching results by employing SuperPoint and SuperGlue. The matched correspondences are visualized using the same color scheme.

Method	-			Gravity-guided		
	(1m, 1°)	(3m, 3°)	(5m, 5°)	(1m, 1°)	(3m, 3°)	(5m, 5°)
LoFTR(Top @1)	39.34	89.34	89.86	43.18	91.78	93.01
LoFTR(Top @3)	41.26	93.53	93.88	43.18	93.88	94.58
SIFT + NN(Top @3)	34.79	91.26	93.01	34.09	90.73	91.78
SPP + NN(Top @3)	41.96	93.71	94.06	43.71	94.93	95.45
SPP + SPG(Top @3)	40.73	94.76	94.76	44.58	96.15	96.68

Table 3. **Visual localization results on UAVD4L.** We evaluate the localization performance of different feature detectors and descriptors using the two-stages UAV localization pipeline with /without sensor priors. Note that all methods use OpenIBL for image retrieval.

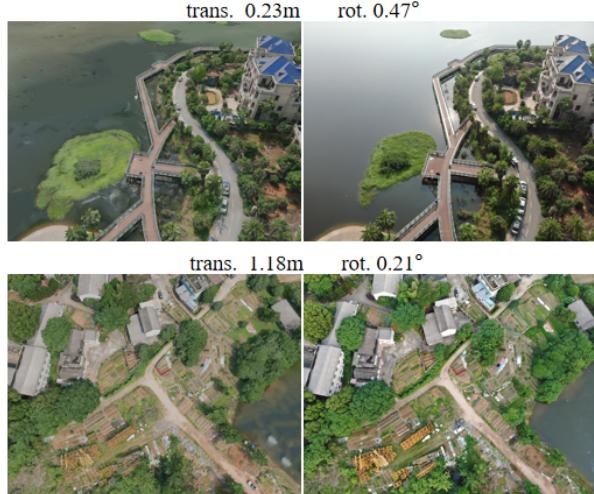


Figure 7. **Visual results of images rendered from the estimated poses.** The localization error for each pose is shown at the top of the image.

361 **Results.** Figure 8 shows two sequential UAV localization
362 and target tracking results, with and without the use of sen-

sor information for localization.

363 Due to the accurate UAV localization, the target position
364 achieves meter-level accuracy. Figure 9 visualizes the
365 difference between the estimated target trajectory and the true
366 trajectory under different UAV localization results. It can be
367 seen that target estimation using sensor-guided localization
368 results in a smaller position error.

Hierarchical render (m, \circ)	(1m, 1°) / (3m, 3°) / (5m, 5°)
$H = 150, \theta_{pitch}^r = 45$	32.69 / 79.2 / 79.37
$H = 150 \& 100, \theta_{pitch}^r = 45$	42.63 / 89.86 / 90.03
$H = 150 \& 100, \theta_{pitch}^r = 45 \& 0$	44.58 / 96.15 / 96.68

Table 4. **Ablation study.** The visual localization results of different hierarchical render setting are reported on the UAVD4L.

5.4. Ablation Studies

To fully understand the impact of different hierarchical render settings on visual localization, we evaluate three render setting with results show in Table 4. Only render single level of altitude and angle obtains the lowest accuracy, while render multiple levels of attitude and angle attains the favor-

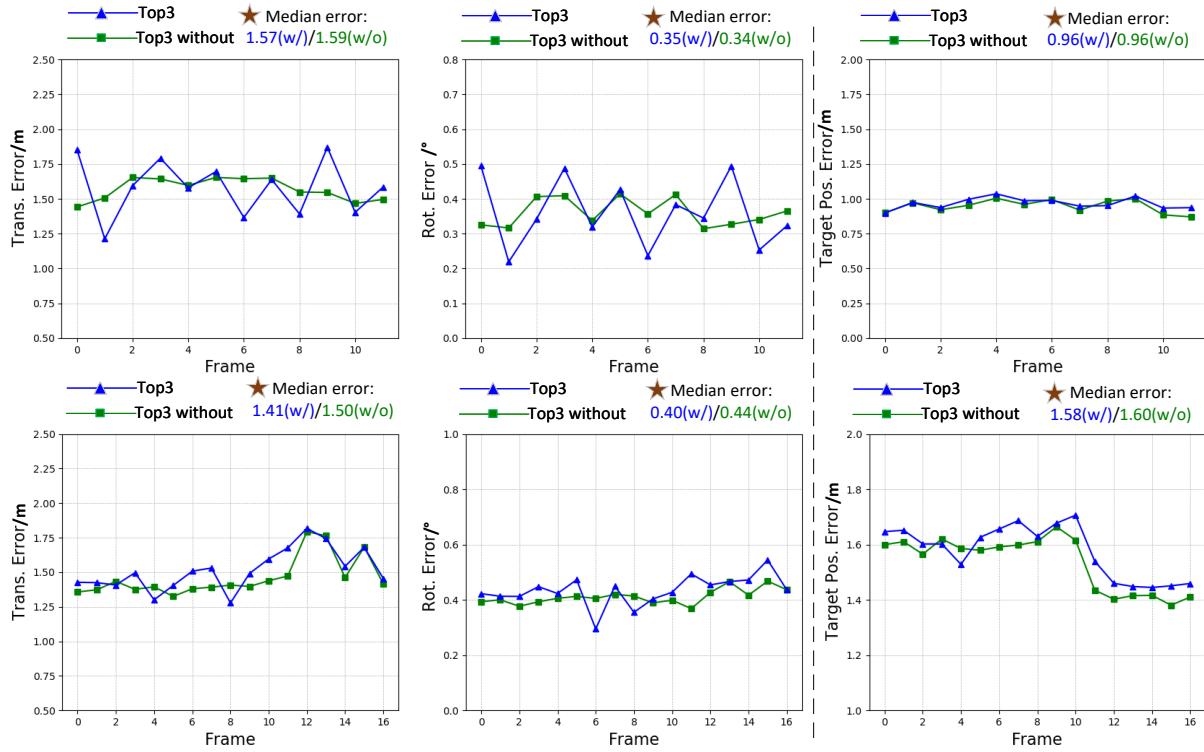


Figure 8. **Localization error results for different trajectory.** For two target tracking trajectories, we report the UAV localization and target tracking error with or without sensor prior.

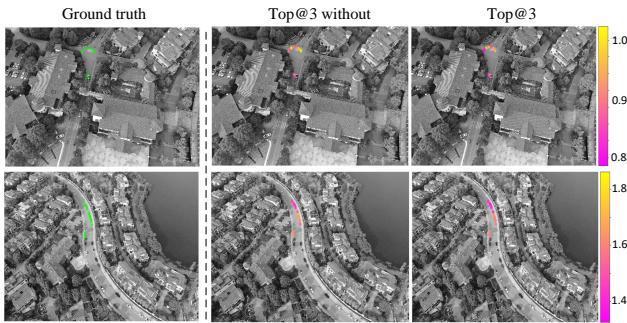


Figure 9. **Qualitative results of target tracking.** The left side of the figure displays the ground truth positions of the tracking target, represented by green points. The right side of the figure shows the estimated target position error, measured in meters as a translation error.

able accuracy. This indicates increasing render layer by altitude and angle improves the results. We conduct these experiment using the same image retrieval (OpenIBL), feature point detection and matching (SPP+SPG), and the gravity-guided PnP RANSAC on UAVD4L.

6. Conclusion

We introduce a large-scale dataset, UAVD4L, for UAV 6-DoF localization in GPS-denied environments. The dataset contains a world-aligned textured 3D reference model and query images with accurate GT poses. Additionally, we implement a novel two-stage 6-DoF localization pipeline that fully utilizes comprehensive data rendering and rotation sensor information. Furthermore, based on the 6-DoF pose estimator, we design a hierarchical system to track designated objects on the ground and output their 3D positions. We anticipate that the UAVD4L dataset will serve as a valuable resource and benchmark for future research in this field.

References

- [1] Abdulla Al-Kaff, David Martin, Fernando Garcia, Arturo de la Escalera, and José María Armingol. Survey of computer vision algorithms and applications for unmanned aerial vehicles. *Expert Systems with Applications*, 92:447–463, 2018. 1
- [2] Relja Arandjelovic, Petr Gronat, Akihiko Torii, Tomas Pajdla, and Josef Sivic. Netvlad: Cnn architecture for weakly supervised place recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 400–408, 2016. 1

381
382
383
384
385
386
387
388
389
390
391
392
393

394
395
396
397
398
399
400
401
402

- 403 conference on computer vision and pattern recognition,
404 pages 5297–5307, 2016. 2, 4, 6
- [3] Ganesan Balamurugan, J Valarmathi, and VPS Naidu. Survey on uav navigation in gps denied environments. In *2016 International conference on signal processing, communication, power and embedded system (SCOPES)*, pages 198–204. IEEE, 2016. 1
- [4] Giovanni Barbarani, Mohamad Mostafa, Hajali Bayramov, Gabriele Trivigno, Gabriele Berton, Carlo Masone, and Barbara Caputo. Are local features all you need for cross-domain visual place recognition? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6154–6164, 2023. 4
- [5] Mesay Belete Bejiga, Abdallah Zeggada, Abdelhamid Nouffidj, and Farid Melgani. A convolutional neural network approach for assisting avalanche search and rescue operations with uav imagery. *Remote Sensing*, 9(2):100, 2017. 1
- [6] Mollie Bianchi and Timothy D Barfoot. Uav localization using autoencoded satellite images. *IEEE Robotics and Automation Letters*, 6(2):1761–1768, 2021. 2
- [7] Yiming Cai, Yao Zhou, Hongwen Zhang, Yuli Xia, Peng Qiao, and Junsuo Zhao. Review of target geo-location algorithms for aerial remote sensing cameras without control points. *Applied Sciences*, 12(24):12689, 2022. 2, 5
- [8] Shuxiao Chen, Xiangyu Wu, Mark W Mueller, and Koushil Sreenath. Real-time geo-localization using satellite imagery and topography for unmanned aerial vehicles. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2275–2281. IEEE, 2021. 2, 3
- [9] Junho Choi and Hyun Myung. Brm localization: Uav localization in gnss-denied environments based on matching of numerical map and uav images. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4537–4544. IEEE, 2020. 2
- [10] Ivan Cisneros, Peng Yin, Ji Zhang, Howie Choset, and Sebastian Scherer. Alto: A large-scale dataset for uav visual place recognition and localization. *arXiv preprint arXiv:2207.12317*, 2022. 2, 3
- [11] Andy Couturier and Moulay A Akhloufi. A review on absolute visual localization for uav. *Robotics and Autonomous Systems*, 135:103666, 2021. 1
- [12] Lowe David. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60:91–110, 2004. 6
- [13] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Superpoint: Self-supervised interest point detection and description. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 224–236, 2018. 6
- [14] Martin A Fischler and Robert C Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. 2
- [15] Yixiao Ge, Haibo Wang, Feng Zhu, Rui Zhao, and Hongsheng Li. Self-supervising fine-grained region similarities for large-scale image localization. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, Au-*
- gust 23–28, 2020, *Proceedings, Part IV 16*, pages 369–386. Springer, 2020. 2, 4, 6
- [16] Hunter Goforth and Simon Lucey. Gps-denied uav localization using pre-existing satellite imagery. In *2019 International conference on robotics and automation (ICRA)*, pages 2974–2980. IEEE, 2019. 2, 3
- [17] Albert Gordo, Jon Almazan, Jerome Revaud, and Diane Larlus. End-to-end learning of deep visual representations for image retrieval. *International Journal of Computer Vision*, 124(2):237–254, 2017. 2, 4, 6
- [18] Emilia Guisado-Pintado, Derek WT Jackson, and David Rogers. 3d mapping efficacy of a drone and terrestrial laser scanner over a temperate beach-dune zone. *Geomorphology*, 328:157–172, 2019. 1
- [19] Marius-Mihail Gurgu, Jorge Peña Queralta, and Tomi Westerlund. Vision-based gnss-free localization for uavs in the wild. In *2022 7th International Conference on Mechanical Engineering and Robotics Research (ICMERR)*, pages 7–12. IEEE, 2022. 2, 3
- [20] Bert M Haralick, Chung-Nan Lee, Karsten Ottenberg, and Michael Nölle. Review and analysis of solutions of the three point perspective pose estimation problem. *International journal of computer vision*, 13:331–356, 1994. 2
- [21] Chao Huang, Hongmei Zhang, and Jianhu Zhao. High-efficiency determination of coastline by combination of tidal level and coastal zone dem from uav tilt photogrammetry. *Remote Sensing*, 12(14):2189, 2020. 2, 5
- [22] Christoforos Kanellakis and George Nikolakopoulos. Survey on computer vision for uavs: Current developments and trends. *Journal of Intelligent & Robotic Systems*, 87:141–168, 2017. 1
- [23] Jouko Kinnari, Francesco Verdoja, and Ville Kyrki. Gnss-denied geolocation of uavs by visual matching of on-board camera images with orthophotos. In *2021 20th International Conference on Advanced Robotics (ICAR)*, pages 555–562. IEEE, 2021. 3
- [24] Jouko Kinnari, Riccardo Renzulli, Francesco Verdoja, and Ville Kyrki. Lsvl: Large-scale season-invariant visual localization for uavs. *arXiv preprint arXiv:2212.03581*, 2022. 3
- [25] Edward Larson, Arthur Ianuzzi, Renier Oliva, Dennis Lorence, Graham Sheffer, et al. Autonomous underwater survey apparatus and system, 2021. US Patent 11,072,405. 1
- [26] Vincent Lepetit, Francesc Moreno-Noguer, and Pascal Fua. Ep n p: An accurate o (n) solution to the p n p problem. *International journal of computer vision*, 81:155–166, 2009. 2
- [27] Yu Liu, Jing Bai, Gang Wang, Xiaobo Wu, Fangde Sun, Zhengqiang Guo, and Hujun Geng. Uav localization in low-altitude gnss-denied environments based on poi and store signage text matching in uav images. *Drones*, 7(7):451, 2023. 2
- [28] Yuncheng Lu, Zhucun Xue, Gui-Song Xia, and Liangpei Zhang. A survey on vision-based uav navigation. *Geospatial information science*, 21(1):21–32, 2018. 1
- [29] Alina Marcu, Dragos Costea, Emil Slusanschi, and Marius Leordeanu. A multi-stage multi-task neural network

- 517 for aerial scene interpretation and geolocalization. *arXiv preprint arXiv:1804.01322*, 2018. 2
- 518
- 519 [30] Arthur Moreau, Nathan Piasco, Dzmitry Tsishkou, Bogdan
- 520 Stanciulescu, and Arnaud de La Fortelle. Lens: Localization
- 521 enhanced by nerf synthesis. In *Conference on Robot Learning*,
- 522 pages 1347–1356. PMLR, 2022. 4
- 523 [31] Muhammad Hamza Mughal, Muhammad Jawad Khokhar,
- 524 and Muhammad Shahzad. Assisting uav localization via
- 525 deep contextual image matching. *IEEE Journal of Selected*
- 526 *Topics in Applied Earth Observations and Remote Sensing*,
- 527 14:2445–2457, 2021. 4
- 528 [32] Ahmed Nassar and Mohamed ElHelw. Aerial imagery reg-
- 529 istration using deep learning for uav geolocalization. *Deep*
- 530 *Learning in Computer Vision: Principles and Applications*,
- 531 pages 183–210, 2020. 1
- 532 [33] Nipun D Nath, Chih-Shen Cheng, and Amir H Behzadan.
- 533 Drone mapping of damage information in gps-denied dis-
- 534 aster sites. *Advanced Engineering Informatics*, 51:101450,
- 535 2022. 2
- 536 [34] Vojtech Panek, Zuzana Kukelova, and Torsten Sattler.
- 537 Meshloc: Mesh-based visual localization. In *European Con-*
- 538 *ference on Computer Vision*, pages 589–609. Springer, 2022.
- 539 2
- 540 [35] Bhavit Patel, Timothy D Barfoot, and Angela P Schoell-
- 541 lig. Visual localization with google earth images for ro-
- 542 bust global pose estimation of uavs. In *2020 IEEE Inter-*
- 543 *national Conference on Robotics and Automation (ICRA)*,
- 544 pages 6491–6497. IEEE, 2020. 3
- 545 [36] Paul-Edouard Sarlin, Cesar Cadena, Roland Siegwart, and
- 546 Marcin Dymczyk. From coarse to fine: Robust hierarchical
- 547 localization at large scale. In *Proceedings of the IEEE/CVF*
- 548 *Conference on Computer Vision and Pattern Recognition*,
- 549 pages 12716–12725, 2019. 2
- 550 [37] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz,
- 551 and Andrew Rabinovich. Superglue: Learning feature
- 552 matching with graph neural networks. In *Proceedings of*
- 553 *the IEEE/CVF conference on computer vision and pattern*
- 554 *recognition*, pages 4938–4947, 2020. 2, 6
- 555 [38] Paul-Edouard Sarlin, Ajaykumar Unagar, Mans Larsson,
- 556 Hugo Germain, Carl Toft, Viktor Larsson, Marc Pollefeys,
- 557 Vincent Lepetit, Lars Hammarstrand, Fredrik Kahl, et al.
- 558 Back to the feature: Learning robust camera localization
- 559 from pixels to pose. In *Proceedings of the IEEE/CVF con-*
- 560 *ference on computer vision and pattern recognition*, pages
- 561 3247–3257, 2021. 2, 6
- 562 [39] Paul-Edouard Sarlin, Mihai Dusmanu, Johannes L
- 563 Schönberger, Pablo Speciale, Lukas Gruber, Viktor
- 564 Larsson, Ondrej Miksik, and Marc Pollefeys. Lamar:
- 565 Benchmarking localization and mapping for augmented
- 566 reality. In *European Conference on Computer Vision*, pages
- 567 686–704. Springer, 2022. 2
- 568 [40] Akshay Shetty and Grace Xingxin Gao. Uav pose estimation
- 569 using cross-view geolocalization with satellite imagery. In
- 570 *2019 International Conference on Robotics and Automation*
- 571 (*ICRA*), pages 1827–1833. IEEE, 2019. 2
- 572 [41] Mario Silvagni, Andrea Tonoli, Enrico Zenerino, and Mar-
- 573 cello Chiaberge. Multipurpose uav for search and rescue op-
- 574 erations in mountain avalanche events. *Geomatics, Natural*
- 575 *Hazards and Risk*, 8(1):18–33, 2017. 1
- 576 [42] Jiaming Sun, Zehong Shen, Yuang Wang, Hujun Bao, and
- 577 Xiaowei Zhou. Loftr: Detector-free local feature matching
- 578 with transformers. In *Proceedings of the IEEE/CVF con-*
- 579 *ference on computer vision and pattern recognition*, pages
- 580 8922–8931, 2021. 4, 6
- 581 [43] Hajime Taira, Masatoshi Okutomi, Torsten Sattler, Mircea
- 582 Cimpoi, Marc Pollefeys, Josef Sivic, Tomas Pajdla, and Ak-
- 583 ihiko Torii. Inloc: Indoor visual localization with dense
- 584 matching and view synthesis. In *Proceedings of the IEEE*
- 585 *Conference on Computer Vision and Pattern Recognition*,
- 586 pages 7199–7209, 2018. 4
- 587 [44] Songbing Wu, Chun Du, Hao Chen, and Ning Jing. Coarse-
- 588 to-fine uav image geo-localization using multi-stage lucas-
- 589 kanade networks. In *2021 2nd Information Communication*
- 590 *Technologies Conference (ICTC)*, pages 220–224. IEEE,
- 591 2021. 2
- 592 [45] Yingxiao Xu, Long Pan, Chun Du, Jun Li, Ning Jing, and
- 593 Jiangjiang Wu. Vision-based uavs aerial image localization:
- 594 A survey. In *Proceedings of the 2nd ACM SIGSPATIAL*
- 595 *International Workshop on AI for Geographic Knowledge Dis-*
- 596 *covery*, pages 9–18, 2018. 1
- 597 [46] Yingxiao Xu, Songbing Wu, Chun Du, Jun Li, and Ning
- 598 Jing. Uav image geo-localization by point-line-patch feature
- 599 matching and iclk optimization. In *2022 29th International*
- 600 *Conference on Geoinformatics*, pages 1–7. IEEE, 2022. 2
- 601 [47] Qi Yan, Jianhao Zheng, Simon Reding, Shanci Li, and Iordan
- 602 Doytchinov. Crossloc: Scalable aerial localization as-
- 603 sisted by multimodal synthetic data. In *Proceedings of*
- 604 *the IEEE/CVF Conference on Computer Vision and Pattern*
- 605 *Recognition*, pages 17358–17368, 2022. 2, 3, 4, 6
- 606 [48] Shen Yan, Yu Liu, Long Wang, Zehong Shen, Zhen Peng,
- 607 Haomin Liu, Maojun Zhang, Guofeng Zhang, and Xiaowei
- 608 Zhou. Long-term visual localization with mobile sensors.
- 609 In *Proceedings of the IEEE/CVF Conference on Computer*
- 610 *Vision and Pattern Recognition*, pages 17245–17255, 2023.
- 611 2, 4, 5, 6
- 612 [49] Peng Yin, Lingyun Xu, Ji Zhang, Howie Choset, and Sebas-
- 613 tian Scherer. i3dloc: Image-to-range cross-domain localiza-
- 614 tion robust to inconsistent environmental conditions. *arXiv*
- 615 *preprint arXiv:2105.12883*, 2021. 2
- 616 [50] Peng Yin, Ivan Cisneros, Shiqi Zhao, Ji Zhang, Howie
- 617 Choset, and Sebastian Scherer. isimloc: Visual global lo-
- 618 calization for previously unseen environments with simulated
- 619 images. *IEEE Transactions on Robotics*, 2023. 2
- 620 [51] Jiahui Zhang, Shitao Tang, Kejie Qiu, Rui Huang,
- 621 Chuan Fang, Le Cui, Zilong Dong, Siyu Zhu, and Ping
- 622 Tan. Rendernet: Visual relocalization using virtual view-
- 623 points in large-scale indoor environments. *arXiv preprint*
- 624 *arXiv:2207.12579*, 2022. 4
- 625 [52] Zichao Zhang, Torsten Sattler, and Davide Scaramuzza. Ref-
- 626 erence pose generation for long-term visual localization via
- 627 learned features and view synthesis. *International Journal of*
- 628 *Computer Vision*, 129:821–844, 2021. 4
- 629 [53] Zhedong Zheng, Yunchao Wei, and Yi Yang. University-
- 630 1652: A multi-view multi-source benchmark for drone-
- 631 based geo-localization. In *Proceedings of the 28th ACM*

3DV 2024 Submission #298. CONFIDENTIAL REVIEW COPY. DO NOT DISTRIBUTE.

632
633

international conference on Multimedia, pages 1395–1403,
2020. 2, 3

UAVD4L: A Large-Scale Dataset for UAV 6-DoF Localization

Supplementary Material

Lens number	5
Tilt angle	45°
Image resolution	6144 × 4096
Focal lens	Downward:25mm, sideward:35mm
Sensor size	APS-C Format (23.5 × 15.6mm)

Table 1. SHARE PSDK 102s camera parameters.

634

A. Dataset Details

635

A.1. Reference Map Collection

636

We collect the reference images via aerial oblique photography, with the capture devices visualized in Figure 1. Specifically, we equip a quadcopter UAV, DJI M300 RTK¹ with a five-eye oblique camera, SHARE PSDK 102s². The detailed parameters of the five-eye camera are provided in Table 1.

642

A group of five-eye images captured in one shot of SHARE PSDK 102s is shown in Figure 2, with one lens capturing downward and four lenses capturing sideward. And the reconstructed 3D textured model is illuminated in Figure 3.



Figure 1. Reference map capturing devices. (a). The UAV platform DJI M300 RTK. (b). The five-eye camera SHARE 102s.

647

A.2. Query Image Collection

648

We collect all query images $\{I_q\}$ using a DJI M300 RTK UAV mounted with a DJI H20T camera³, which has two RGB cameras, including a zoom lens camera and a wide-angle lens camera. More details are provided in Table 2.

649

The query images are captured at different time and positions, with each image in the sequence taken 2 seconds apart. The capturing details of each trajectory are provided in Table 3. We visualize the capturing device and the flight trajectory of one of the sequences, as shown in Figure 4.

¹<https://enterprise.dji.com/cn/matrice-300>

²<https://shareuav.cn/V3S>

³<https://enterprise.dji.com/cn/zenmuse-h20-series>



Figure 2. Images captured by the SHARE 102s at one shot. The five-eye camera owns forward, backward, left, right, and downward directions.



Figure 3. Textured 3D model reconstruction. Given a large number of aerial oblique photographs, we produce a high-quality 3D textured model. The yellow boxes show the details of the partial 3D model. Please zoom in to see the details.

Camera type	Camera parameters
Zoom lens	Sensor size
	Focal lens
	Image resolution
Wide-angle lens	1.17" CMOS
	6.83 – 119.94mm
	5184 × 3888
Wide-angle lens	1.23" CMOS
	4.5mm
	4056 × 3040

Table 2. The DJI H20T camera parameters.

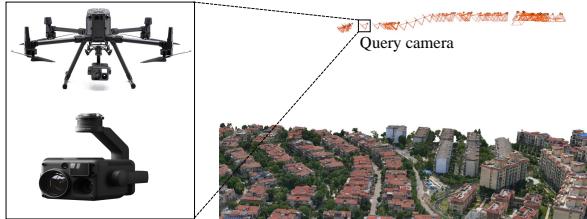


Figure 4. **Query image collection.** The capture device DJI M300 RTK mounted H20T is shown on the left, and the one query trajectory is shown on the right.

Seq.	Num.	Height(m)	$\theta_{pitch}(\circ)$
1	45	50-70	22-45
2	216	120-180	37-70
3	151	120-150	55-70
4	103	110-180	45-70
5	58	160-180	50-60

Table 3. The details of the different query trajectories

A.3. Query GT Generation

We provide additional GT rendering results in Figure 5 to demonstrate the accuracy of our generated GT poses for evaluation.

B. Method Details

B.1. Synthetic Data Generation

In UAVD4L, we generate virtual viewpoints at different spatial positions and produce 16 images for each position with varying headings and pitches, as illustrated in Figure 6.

B.2. Hierarchical Target Tracking

In this section, we introduce the details of our target localization algorithm, which is based on DEM-guided sampling and comparison [7]. The main scheme is illustrated in Figure 7, and the entire process can be divided into two steps:

- 1) **Target ray calculation.** After using the wide-angle lens camera to obtain the 6-DoF of the UAV camera system, we manually detect the 2D pixel position (x, y) of the target on the zoom lens image. We then unproject a ray l from the camera optical center c to the target position (x, y) in 2D plane, where $l = \{x_t = a_1 z_t + b_1, y_t = a_2 z_t + b_2\}$. Please note that the zoom lens and wide-angle cameras have been carefully calibrated.
- 2) **Calculate the intersection point.** Through DEM, we obtain the lowest height $z(min)$ of the area and calculate the potentially lowest ray-and-map intersection point $p(min) = (x_t(min), y_t(min), z_t(min))$ using $\{x_t(min) = a_1 z(min) + b_1, y_t(min) = a_2 z(min) + b_2, z_t(min) = z(min)\}$. We sample n points between

Rot	R@1	R@3	P@3
None	83.74	98.43	76.69
$\gamma_o = 60^\circ$	83.74	98.51	77.45
$\gamma_o = 50^\circ$	84.13	98.51	77.51
$\gamma_o = 40^\circ$	84.27	98.51	77.74
$\gamma_o = 30^\circ$	84.44	98.78	79.08

Table 4. **Image retrieval results.** We report the top- k recall and precision for $k = 1, 3$, based OpenIBL [15] and using different rotation threshold γ_o .

camera optical center c and potential lowest point p with varying height z_t and compare it with DEM height $h_t = DEM(x_t, y_t)$ at 2D coordinate (x_t, y_t) . We find the one with the lowest absolute height difference $z_t(opt) = argmin||z_t - h_t||$, and recover the 3D target position using $\{x_t(opt) = a_1 z_t(opt) + b_1, y_t(opt) = a_2 z_t(opt) + b_2, z_t(opt) = z_t(opt)\}$. The number of sampling number n is set to 500.

C. Experiment Details

C.1. Image Retrieval

More ablation studies. In this study, we evaluate the performance of rotation threshold γ_o for image retrieval under the OpenIBL [15] global descriptor. The retrieval results in terms of recall and precision are shown in Table 4.

Overlap computation. We assess the correctness of retrieval results based on the overlap rate. Retrieval results are considered correct if they share a sufficient overlapping area with the query image, where the overlap percentage $P_{overlap}$ is greater than 50%. We calculate the overlap as follows:

$$\begin{aligned} p_{x,y} &\in \mathcal{IR}_r, \quad p_{x,y}^t = K_q \varepsilon_q K_r^{-1} \varepsilon_r^{-1} p_{x,y}, \\ \mathcal{IR}_{overlap} &= \sum \{p_{x,y}^t \in \mathcal{IR}_q\}, \\ P_{overlap} &= \mathcal{IR}_{overlap} / \sum \mathcal{IR}_r \times 100\% \end{aligned} \quad (3)$$

where \mathcal{IR}_r denotes the pixel region of the reference image, $p_{x,y}$ denotes a pixel point within the \mathcal{IR}_r region, (K_q, ε_q) and (K_r, ε_r) denote the intrinsic and extrinsic of query and reference image, respectively. As we have RGB $\{I_r\}$ and depth $\{D_r\}$ images for synthetic reference views, we can compute the pixel transformation from the reference image to the query image by projection formula. We use $\mathcal{IR}_{overlap}$ to count the number of the transformed pixel point $p_{x,y}^t$ that fall within the query image region \mathcal{IR}_q , and $P_{overlap}$ represents the overlap percentage.



Figure 5. **GT poses quality on UAVD4L.** Using the GT poses, we render images from the textured 3D model, and compare them to the original query image.

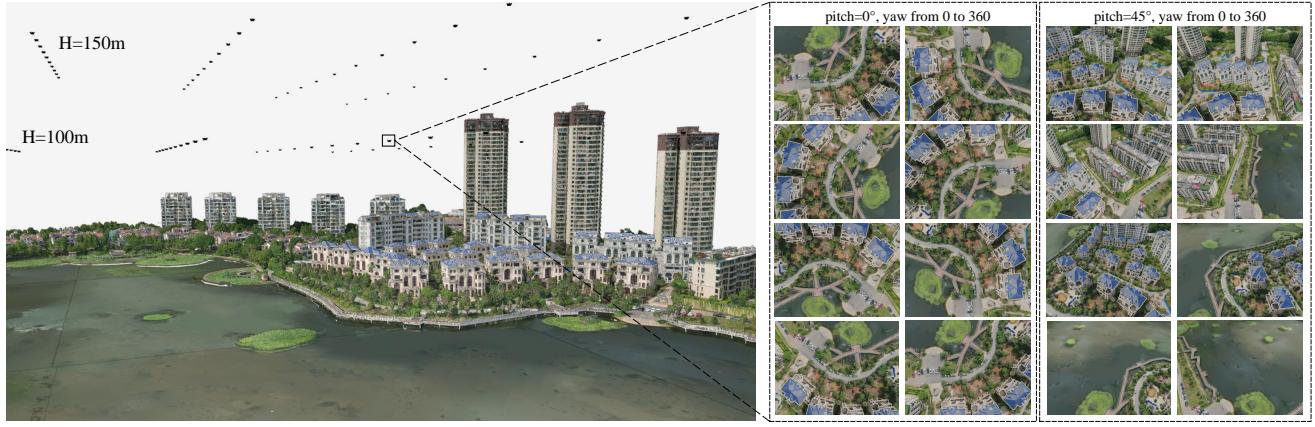


Figure 6. **Synthetic data generation.** We show the distribution of virtual positions in a specific region (left), with different capturing altitudes (100 and 150 meters) and intervals (50 and 75 meters). The 16 synthesized images for each position are provided for visualization, with varying pitch and yaw angles.

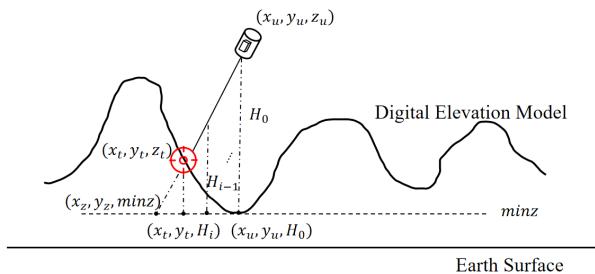


Figure 7. **Schematic of target localization based on DEM.**