



OPEN Enhanced real-time 6d pose estimation for automatic recovery of in-flight UAVs using distance-aware keypoint heatmaps

Mingyu Jeong & Andrew Jaeyong Choi✉

Unmanned Aerial Vehicles (UAVs) play crucial roles across various fields but face significant challenges when control is lost or securing a safe runway is not feasible. Under these circumstances, reliable UAV recovery systems are essential for safe retrieval, requiring precise and real-time localization. This paper proposes an automatic in-flight UAV recovery system using an unmanned ground vehicle when securing a safe runway is challenging such as unstructured road, and presents a practical approach for real-time six-degree-of-freedom (6-DOF) pose estimation of an in-flight UAV using monocular RGB images and deep learning-based heatmap keypoint detection. To enhance the accuracy of the pose estimation, we propose an Adaptive Sigma technique for keypoint detection, which adjusts the sigma values of the keypoint heatmap based on the distance between the camera and the UAV. Thus, the proposed method can robustly adapt to changes in the distance of UAV to improve the keypoint localization performance. By utilizing the predicted keypoints in a Perspective-n-Point algorithm, the 6-DOF pose information of the UAV can be obtained in real time. The proposed Adaptive Sigma to heatmap-based keypoint detection improves the Percentage of Correct Keypoints by up to 2.5%, with consistently high performance across various state-of-the-art backbone architectures. The proposed method qualitatively evaluated in challenging scenarios such as various altitudes, significant tilt, and motion blur.

Unmanned Aerial Vehicles (UAVs) are aircraft operated remotely or autonomously without an onboard human pilot. With advancements in UAV technology, their applications have significantly expanded, particularly in military contexts for reconnaissance, surveillance, and tactical strikes.

However, UAVs such as fixed-wing models inherently require a runway for safe landing. To address this limitation, recent research has investigated alternative recovery methods that operate without the need for a runway^{1–3}. But, they often risk causing structural damage or the loss of sensitive data. To address these issues, a potential solution to these drawbacks is an autonomous in-flight recovery system using a ground vehicle to capture the UAV, as conceptually illustrated in Fig. 1. For such a system to be feasible, the ground vehicle must be capable of robust and accurate real-time 6-DOF pose estimation of the UAV. This paper specifically focuses on the final docking stage (Phase 3) of this concept, where precise localization is crucial for a successful capture.

From the perspective of the recovery UGV, obtaining the UAV's accurate and real-time 6-DOF (six-degree-of-freedom) pose is a key requirement. To address this in a practical and cost-effective manner, we propose a method for real-time 6-DOF pose estimation using only a monocular RGB camera, avoiding the need for expensive external localization equipment. Our approach is a two-stage process: we first detect the 2D keypoints of the in-flight UAV using a heatmap-based deep learning approach, and then solve the Perspective-n-Point (PnP) problem to compute its 6D pose.

Unlike conventional heatmap-based keypoint detection methods, this paper introduces the concept of Adaptive Sigma to achieve more effective precision in UAV localization. The Adaptive Sigma technique adjusts the Gaussian distribution's sigma value according to the distance between the UAV and the camera. This adjustment is crucial as it allows the model to account for variations in the apparent size of the UAV in the captured image, enhancing the accuracy of keypoint localization.

The performance of this approach was evaluated through extensive experiments using a custom dataset labeled with UAV keypoints. The experiments included comparisons with various state-of-the-art backbone architectures. The effectiveness of the Adaptive Sigma method was demonstrated through improvements in the Percentage of Correct Keypoints (PCK) metric, frame rates (FPS), and model parameters.

School of Computing, Gachon University, Seongnam 1342, South Korea. ✉email: andrewjchoi@gachon.ac.kr

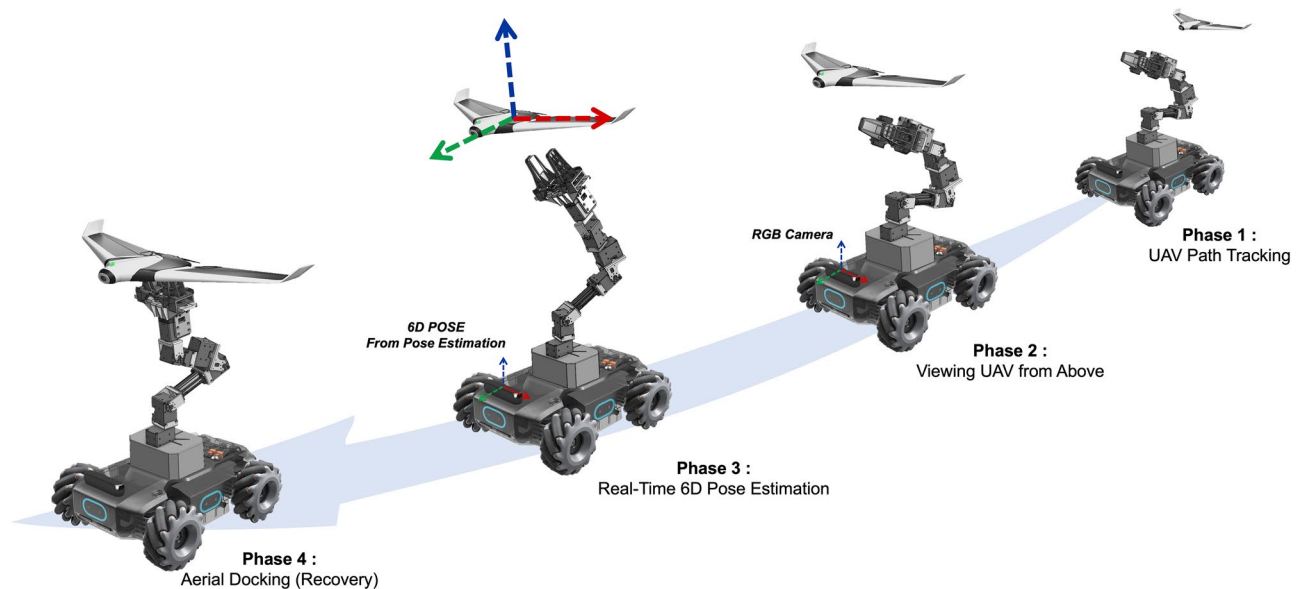


Fig. 1. Conceptual overview of the in-flight UAVs automatic recovery system using a mobile manipulator robot platform with 6D pose estimation. While the figure illustrates the overall system-level goal, this paper focuses on evaluating the 6D pose estimation module, with experimental validation conducted using a quadcopter for practical lab-scale implementation.

In summary, the main contributions of this paper are:

- We propose a distance-aware Adaptive Sigma technique for keypoint heatmap generation, which improves 2D localization accuracy by explicitly modeling the UAV's scale variations in monocular images.
- We conduct a comprehensive empirical study showing that Adaptive Sigma consistently boosts keypoint detection performance (up to 2.5% in PCK) across diverse modern backbone architectures.
- We present a practical and real-time UAV 6D pose estimation framework on an embedded platform, demonstrating robustness under challenging conditions such as severe tilt and motion blur in in-flight recovery scenarios.

Related works

UAVs automatic recovery systems

Unmanned Aerial Vehicles (UAVs) are increasingly utilized in various applications, especially in military operations for reconnaissance, surveillance, and tactical strikes. Given the high costs associated with UAVs and the critical information they may carry, ensuring their safe recovery after missions is imperative. Efficient recovery and redeployment of UAVs not only enhance their operational range but also improve cost-effectiveness.

However, reliable recovery of UAVs poses significant challenges, particularly when control is lost or suitable landing zones are unavailable. Such scenarios introduce substantial risks, including the potential loss of valuable equipment and sensitive data. Recent studies have explored UAV recovery in different environments—marine, road, and aerial spaces—to address these challenges⁴.

One notable approach involves using two multirotor UAVs equipped with nets to intercept an incoming UAV mid-air⁵. While this aerial recovery technique showcases innovative solutions, our research adopts a different strategy. We focus on tracking and capturing descending UAVs from the ground using Unmanned Ground Vehicles (UGVs) equipped with robotic manipulators. This ground-based approach aims to provide a practical and efficient solution for UAV recovery operations. By obtaining precise 6-DOF pose information of the UAV from a ground perspective, our method enhances the accuracy and reliability of the recovery process. This approach not only mitigates the risks associated with uncontrolled UAV descents but also contributes to the overall safety and effectiveness of UAV deployments.

In summary, effective UAV recovery methods are crucial due to the high value of UAVs and the potential risks involved in their operations. Our research contributes to this field by proposing a ground-based tracking and capture system that ensures the safe and efficient recovery of UAVs, addressing the challenges posed by loss of control and lack of suitable landing zones.

6D rigid body pose estimation

6D pose estimation for rigid objects aims to determine the 3D rotation and 3D translation of an object relative to a camera. Modern approaches can be broadly categorized into correspondence-based methods and direct regression methods. We review these two mainstreams and then focus on heatmap-based techniques, which our work is built upon.

1. Correspondence-based methods: These methods follow a two-stage pipeline: first, they establish correspondences between 2D image pixels and the object's 3D coordinates, and second, they solve for the 6D pose using a Perspective-n-Point (PnP) algorithm. While early works relied on hand-crafted features like SIFT, recent deep learning-based methods have shown significant improvements. For instance, PVNet⁶ pioneered a pixel-wise voting approach, where each pixel votes for the direction of 3D keypoints, making it robust to heavy occlusion. Another notable method, GDRNet⁷, learns an intermediate geometry-aware representation to establish dense correspondences, proving effective even for texture-less objects. These methods demonstrate that high-quality correspondences are key to accurate pose estimation.
2. Direct Regression Methods: In contrast to correspondence-based approaches, direct regression methods aim to predict the 6D pose from the input image in an end-to-end fashion. PoseCNN⁸ was an early attempt that regressed pose parameters directly, but its accuracy was limited, partly due to the challenges in regressing discontinuous rotation representations. More recent works have tried to overcome this by designing better loss functions or employing iterative refinement schemes^{9,10}. However, these methods often exhibit lower accuracy compared to correspondence-based approaches, as they do not explicitly leverage geometric constraints during the pose calculation.

Recent foundation models, such as FoundationPose¹¹, achieve strong generalization to novel objects using a complex render-and-compare framework. However, this approach's computational overhead from rendering numerous pose hypotheses makes it challenging for real-time, embedded applications. Additionally, their optimal performance often relies on RGB-D input, which is impractical for outdoor scenarios due to sunlight interference with depth sensors. In contrast, our work employs a direct, keypoint-based method using only RGB images, targeting a lightweight and robust solution for real-time outdoor operations.

Our research builds on the correspondence-based paradigm, leveraging heatmap-based keypoint detection, a technique proven effective in human pose estimation (HPE)¹². This method involves two main steps: first, predicting a probabilistic heatmap for each predefined keypoint on the object; second, using these detected 2D keypoints and their known 3D model coordinates to solve for the object's 6D pose via a Perspective-n-Point (PnP) algorithm¹³. As illustrated in Fig. 1, this corresponds to Phase 3 of the UAV recovery process, where reliable 6D pose estimation is essential for successful docking. During this phase, the UAV captured by the ground camera may appear at varying distances and scales—sometimes fully visible, and other times only partially within the field of view. To ensure consistent pose estimation under such conditions, we formulate the task as a Perspective-n-Point (PnP) problem that remains solvable even when only a subset of keypoints is visible.

While this heatmap-PnP pipeline is robust, a common limitation in existing works is the use of a fixed-sigma Gaussian kernel to generate ground-truth heatmaps for training. This approach overlooks the significant scale variations of an object as its distance to the camera changes, potentially degrading localization accuracy. To address this gap, we introduce a Distance-Aware Adaptive Sigma technique. Unlike these previous works, our key idea is to dynamically adjust the sigma value of the Gaussian kernel based on the object's distance, allowing the model to learn scale-adaptive heatmaps and leading to more accurate keypoint localization across varying scales. This simple modification enhances keypoint precision across various distances, thereby improving the accuracy of the final 6D pose estimated by the PnP solver, making our framework more suitable for dynamic, real-world operations.

Problem formulation

As shown in Fig. 1, in the context of an Unmanned Aerial Vehicle (UAV) recovery system, we consider a scenario where a UAV needs to land in an unstructured environment without access to a conventional runway. The recovery system comprises an Unmanned Ground Vehicle (UGV) equipped with a robotic manipulator designed to capture the descending UAV using its end effector. The successful execution of this task necessitates precise localization of the UAV relative to the UGV.

The primary objective is to determine the 6D pose of the UAV as it descends, using an RGB camera mounted on the UGV. This camera is oriented upwards, ensuring continuous monitoring of the UAV's underside during its descent. The task is to predict the UAV's pose (R orientation and t position) in real-time using RGB Image captured by the camera.

This estimated pose is then utilized within a predefined transformation (TF) system to guide the UGV's manipulation system for capturing the UAV.

We replaced the fixed-wing UAV with a quadcopter (multirotor UAV) due to limitations related to the flying speed and to conduct the lab-scale experiments with validation

6-DoF pose estimation

To address the challenge of 6-DOF pose estimation for UAVs, we adopt a correspondence-based method. Directly acquiring ground truth pose data for training a model to regress 6-DOF pose is prohibitively expensive. Therefore, we employ a heatmap-based approach to predict keypoints on the UAV, leveraging these predictions as input to a PnP solver to determine the final 6-DOF pose.

6-DoF pose estimation pipeline

Our proposed 6D pose estimation pipeline is illustrated in Fig. 2. Given a monocular RGB image of the UAV captured by the camera, we first extract image features using a backbone network. These features are then passed to a decoder, which generates heatmaps representing the likelihood of each keypoint's presence for each predefined class. Keypoint locations are extracted from these heatmaps through post-processing. Finally, by establishing correspondences between the extracted 2D keypoints and a priori defined 3D points on the UAV, we utilize a PnP solver, along with camera intrinsics, to compute the 6D pose of the UAV.

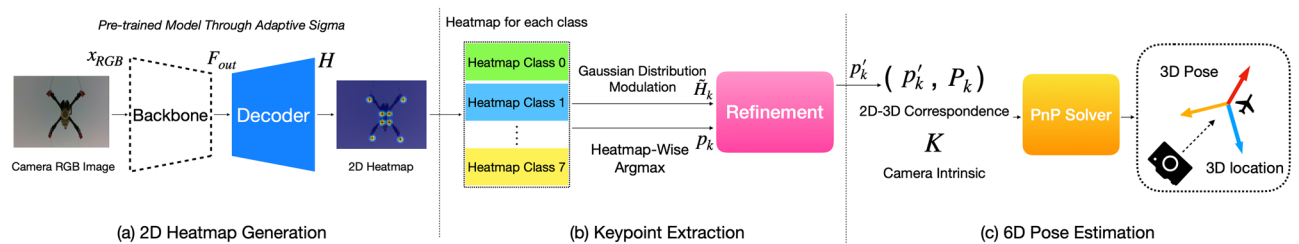


Fig. 2. Our 6D pose estimation flow for UAV localization.

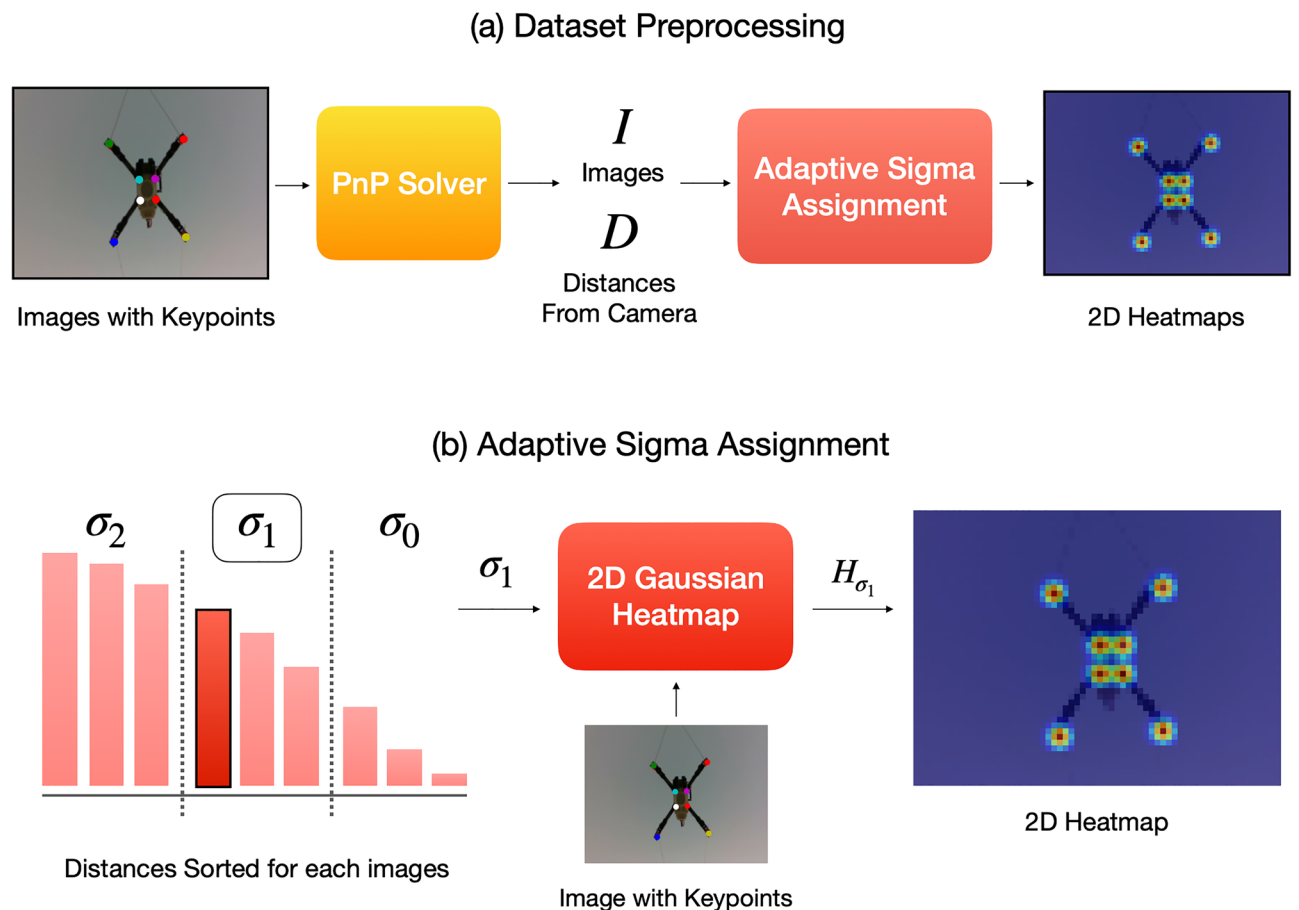


Fig. 3. Overview of adaptive sigma assignment for keypoint heatmap generation during model training: (a) illustrates heatmap creation process based on the object in all images, and (b) illustrates 2d heatmap generation by determining sigma based on the distance between objects and the camera.

Distance-aware sigma adaptation for heatmap generation

In heatmap-based keypoint detection tasks, keypoints within an image are typically represented using heatmaps generated from 2D Gaussian distributions. These distributions are centered at the keypoints' 2D coordinates and characterized by a predefined sigma (σ). However, applying a uniform sigma fails to account for variations in the object's apparent size due to changing distance from the camera, which can limit the model's localization precision.

As shown in Fig. 3, We propose a method to adaptively adjust the sigma of the Gaussian distribution based on the actual distance between the object and the camera when generating heatmaps. The apparent size of the object in the image diminishes as it moves further from the camera, altering the image's informational characteristics. By leveraging these variations, the model can more easily make accurate predictions. For instance, when an object is distant and appears smaller in the image, using a smaller sigma provides more precise positional information. The heatmaps generated using this method are utilized for training the Backbone and Decoder, as illustrated in Fig. 2a.

Variations in the object's apparent size in the image also impact existing models. In transformer-based models like Vision Transformers (ViT)¹⁴, the number of image patches that contain the object varies depending on the object's apparent size in the image. In convolutional neural network (CNN) models, the amount of information captured by the receptive field varies depending on the object's dimensions within the image. By exploiting these characteristics, we can enhance the linearity of the model's localization capabilities. By adjusting sigma to align with the object's actual distance and its apparent size in the image, we reduce the complexity of heatmap generation for the model and achieve more accurate keypoint localization. To calculate the distance between the object and the camera, various methods are available. In this study, we employ the Perspective-n-Point (PnP) algorithm using pre-labeled keypoints to estimate this distance.

It is crucial to clarify that this distance calculation is performed as a one-time, offline preprocessing step on the training dataset. We use the manually annotated ground-truth (GT) 2D keypoints to compute a ground-truth distance for each training image. This pre-computed distance is then saved and used to assign the corresponding adaptive sigma value during model training. This approach ensures that the training process relies only on pre-established ground-truth information. While manual annotation of 2D keypoints is labor-intensive, this method provides the significant advantage of generating the necessary pose ground-truth data, including distance, without requiring expensive external sensors, thus making the dataset construction process more practical.

We begin by computing the distances between the object and the camera for all images in the dataset. Based on these distances, we define a range for sigma values and decide on the number of stages to divide this range. The calculated object distances are then partitioned into the predetermined number of stages (bins), and a specific sigma value is assigned to each distance bin. The sigma values are calculated using the following equations:

$$\sigma_{\text{range}} = \left[\sigma_{\min} + i \cdot \frac{\sigma_{\max} - \sigma_{\min}}{\text{stages} - 1} \right], \quad \text{for } i \in \{0, 1, 2, 3\} \quad (1)$$

Where:

- $\sigma_{\min}$ and $\sigma_{\max}$ are the minimum and maximum sigma values, respectively.
- stages is the number of bins the distances are divided into.
- i represents the stage index.

An example of Heatmap with Adaptive Sigma is introduced in Fig. 5. The generated heatmaps using this adaptive sigma can be seamlessly integrated into existing training methods without any additional modifications to the model. By implementing this adaptive sigma method, we enhance the model's ability to localize keypoints accurately by aligning the heatmap generation process with the varying apparent sizes of objects within the images.

Heatmap prediction

We adopted ViTPose¹⁵ as our baseline for heatmap-based pose estimation. As shown in Fig 2a, the ViTPose-based architecture. ViTPose has shown excellent performance in human pose estimation by leveraging a basic ViT structure and simple deconvolution. We applied various state-of-the-art backbones to this structure to identify the optimal one that could meet the specific requirements of our experiments, including parameters, FPS, and other metrics.

The decoder consists of two consecutive deconvolution layers, each followed by Batch Normalization and ReLU activation functions. All deconvolution layers have an output channel size of 256, and their kernel sizes vary depending on the backbone used, as shown in Table 1. Specifically, when the kernel size is 4, we set the stride and padding to 2 and 1, respectively; when the kernel size is 8, we set them to 4 and 2. With these configurations, we aim to increase the resolution of the input feature maps by factors of two and four, respectively.

$$D = \text{ReLU}(\text{BatchN}(\text{Deconv}_{4 \times 4}(F_{\text{out}}))) \quad (2)$$

Here, F_{out} is the reshaped output of the Backbone's embedding vector (N_k, N_T) into an image form. The reshaping form is as follows:

$$F_{\text{out}} \in \mathbb{R}^{(N_k, \frac{H}{P}, \frac{W}{P})} \quad (3)$$

where P is the patch size, H and W are the height and width of the input image respectively, N_k is the size of the embedding vector, and N_T is the total number of tokens.

After the first deconvolutional layer, the resolution of the image is doubled:

$$D \in \mathbb{R}^{(N_k, \frac{H}{P} \times 2, \frac{W}{P} \times 2)} \quad (4)$$

This upsampling process continues through the second deconvolution layer, ultimately producing heatmaps with the desired output resolution for keypoint prediction. Finally, we generate the required class heatmaps through a 1×1 convolution layer. If the input passes through two deconvolution layers with a kernel size of 4, the final heatmap size is as follows:

$$H = \text{Conv}_{1 \times 1}(D) \in \mathbb{R}^{(N_{\text{class}}, \frac{H}{P} \times 4, \frac{W}{P} \times 4)} \quad (5)$$

Backbone	Decoder (Kernels)	Input Size	Heatmap Size	Params (M)	FLOPs (G)	FPS [Jetson]	PCK _{avg}	PCK _{.01}	PCK _{.0075}	PCK _{.005}	PCK _{.0025}
EfficientViT-M ¹⁸	8, 8	224×224	64×64	9.4	1.21	26	84.2	93.7	91.8	87.4	64.0
+Adaptive	8, 8	224×224	64×64	9.4	1.21	26	86.1 (+1.9)	95.3	93.7	88.7	66.7
PoolFormer-S1 ¹⁹	8, 4	224×224	56×56	15.6	2.95	75	90.7	96.4	95.8	94.0	76.5
+Adaptive	8, 4	224×224	56×56	15.6	2.95	75	92.3 (+1.6)	97.4	96.8	95.1	79.7
EfficientNet-B2 ²⁰	8, 4	256×192	64×48	17.0	5.17	26	94.9	98.1	98.0	96.5	87.0
+Adaptive	8, 4	256×192	64×48	17.0	5.17	26	95.2 (+0.3)	98.2	97.8	96.8	88.0
FastViT-S12 ²¹	8, 4	224×224	64×64	21.5	2.52	85	91.4	97.1	96.6	95.0	77.0
+Adaptive	8, 4	224×224	64×64	21.5	2.52	85	92.6 (+1.2)	97.0	96.5	95.3	81.5
ViT-S/16 ^{*14}	4, 4	256×192	64×48	28.8	6.42	97	94.1	98.0	97.8	96.9	83.5
+Adaptive	4, 4	256×192	64×48	28.8	6.42	97	94.0 (-0.1)	98.4	98.1	96.7	82.9
CMT-T ²²	8, 4	224×224	56×56	30.4	3.16	22	93.1	98.0	97.6	96.1	80.6
+Adaptive	8, 4	224×224	56×56	30.4	3.16	22	94.0 (+0.9)	98.6	98.2	96.0	83.0
CoAtNet-0 ²³	8, 4	224×224	56×56	30.6	4.84	48	92.0	97.8	97.3	95.7	77.0
+Adaptive	8, 4	224×224	56×56	30.6	4.84	48	94.2 (+2.2)	98.1	97.7	96.4	84.6
Swin-T ²⁴	8, 4	224×224	56×56	33.5	5.95	44	89.8	97.4	96.7	93.8	71.3
+Adaptive	8, 4	224×224	56×56	33.5	5.95	44	90.9 (+1.1)	97.2	96.5	94.4	75.6
MobileOne-S3 ²⁵	8, 4	256×192	64×48	34.6	4.29	60	92.4	96.4	95.9	94.5	82.9
+Adaptive	8, 4	256×192	64×48	34.6	4.29	60	91.5 (-0.9)	97.1	96.5	94.7	77.7
LiTV2-S pan2022fast	8, 4	224×224	56×56	40.7	5.01	-	92.5	97.7	97.0	95.2	80.1
+Adaptive	8, 4	224×224	56×56	40.7	5.01	-	95 (+2.5)	98.7	98.2	96.6	86.3
ConvNeXt-V2-T ²⁷	8, 4	224×224	56×56	41.5	5.91	54	89.8	94.6	94.0	92.6	78.1
+Adaptive	8, 4	224×224	56×56	41.5	5.91	54	92.1 (+2.3)	95.6	95.2	93.9	83.6
MambaVision-T ²⁸	8, 4	224×224	56×56	42.6	5.81	42	93.3	98.6	98.3	96.9	79.4
+Adaptive	8, 4	224×224	56×56	42.6	5.81	42	95 (+1.7)	98.4	98.2	97.3	85.9
RegNetY-3.2G ²⁹	8, 4	224×224	56×56	45.2	5.22	35	94.1	98.0	97.6	96.6	84.1
+Adaptive	8, 4	224×224	56×56	45.2	5.22	35	95.7 (+1.6)	98.6	98.3	97.3	88.5

Table 1. Comparison of different state-of-the-art Backbones on Drone Keypoints Estimation. * denotes the model hyperparameter where $Embed_d$ is set to 512, Depth to 8, $Head_{num}$ to 6, and MLP size to 2048. FPS was measured on Jetson AGX Orin 64GB Development Kit using FP32. Decoder (Kernels) means the size of each kernel in Deconv layer. For instance, an entry of '8,8' indicates two Deconv layers, each using an 8x8 kernel filter.

Keypoint extraction & PnP for pose estimation

As shown in Fig. 2b, to extract 2D keypoints from the generated heatmaps for use in the PnP algorithm, we utilize a post-processing step involving an argmax operation followed by an unbiased decoding method. First, the initial keypoint predictions are obtained by finding the pixel location with the highest confidence score in each heatmap channel. This can be expressed as:

$$p_k = \arg \max_{x,y} H_k(x,y) \quad (6)$$

where p_k represents the predicted 2D coordinates of the k -th keypoint, and $H_k(x,y)$ denotes the heatmap value for the k -th keypoint at pixel location (x,y) .

To mitigate potential bias introduced by the argmax operation, we apply an unbiased decoding method proposed by Zhang et al.¹⁶. This method applies a Gaussian distribution modulation to the heatmaps before extracting the keypoints, resulting in more accurate predictions.

After obtaining the initial keypoint predictions using the argmax operation, the heatmaps undergo Gaussian blurring. This step smooths the heatmap values, which helps in better capturing the true peak of the distribution around the keypoint. The Gaussian blur adjusts the heatmap such that the influence of surrounding pixels is considered, reducing the bias caused by simply selecting the maximum value. This can be expressed as:

$$\tilde{H}_k(x,y) = (H_k * G)(x,y) \quad (7)$$

where $\tilde{H}_k(x,y)$ represents the smoothed heatmap for the k -th keypoint, $H_k(x,y)$ is the original heatmap value at pixel location (x,y) , and $G(x,y)$ denotes the Gaussian kernel.

Subsequently, a Taylor expansion is applied around the peak location identified by the argmax. This mathematical technique refines the keypoint position by considering the gradient and curvature of the heatmap at the peak. This can be expressed as:

$$p'_k = p_k + \nabla \tilde{H}_k(x_0, y_0) \cdot \Delta \quad (8)$$

where p'_k is the refined 2D coordinates of the k -th keypoint, $p_k = (x_0, y_0)$ is the initial keypoint position obtained from the argmax, $\nabla \tilde{H}_k(x_0, y_0)$ is the gradient of the smoothed heatmap at the initial peak location, and Δ is a small adjustment vector based on the gradient and curvature.

As shown in Fig. 2c, by incorporating this keypoint extraction process, we obtain accurate 2D keypoint locations from the heatmaps, which are then used in conjunction with the PnP algorithm to estimate the 6D pose of the UAV.

Given a set of n 2D-3D correspondences, $\{p_i, P_i\}_{i=1}^n$, where p_i represents the predicted 2D projection of the i -th 3D point P_i on the UAV, the PnP algorithm aims to find the rotation matrix R and translation vector t that best project the 3D points onto the corresponding 2D points. This can be formulated as a minimization problem:

$$R, t = \arg \min_{R, t} \sum_{i=1}^n \|p_i - \pi(RP_i + t)\|^2 \quad (9)$$

where $\pi(\cdot)$ represents the camera projection function. By solving this optimization problem, we obtain the 6D pose of the UAV, represented by the rotation R and translation t .

Experiment

In this section, we evaluate the effectiveness of our proposed heatmap-based keypoint detection method for UAV 6-DOF pose estimation. Our primary objective is to determine whether accurate keypoint positions can be obtained to establish reliable 2D-3D correspondences, which are essential for solving the PnP problem and, ultimately, determining the UAV's pose. Another important aspect of our evaluation is to assess the real-time capability of the proposed method, which is critical for its practical application in UAV recovery operations. To evaluate keypoint detection accuracy, we use the Percentage of Correct Keypoints (PCK)¹⁷, a metric widely used in human pose estimation.

We also aim to verify the impact of incorporating the Adaptive Sigma technique on keypoint detection accuracy and to conduct a qualitative evaluation of the 6D pose obtained through the PnP solution. This evaluation includes assessing the accuracy of pose estimation in challenging scenarios, such as images with drones in highly tilted poses or with motion-induced blur, by plotting the calculated pose.

Dataset

To train the Heatmap Estimation Model, we created a dataset with Drone Keypoint annotations. We captured drone footage on a building rooftop and labeled keypoints for a total of 8 classes. The dataset comprises four different videos, each shot at different times. The training dataset contains a total of 8,598 images, and the test dataset consists of 1,801 images. The sequences of the video recordings are as follows:

1. The drone approaches the camera in a straight line.
2. The drone ascends away from the camera in a straight line.
3. The drone descends towards the camera while shaking.
4. The drone ascends away from the camera while shaking.

The recorded videos were annotated using the Computer Vision Annotation Tool (CVAT). As shown in Fig. 4, the 8 classes are labeled starting from the right front leg of the drone in a counterclockwise direction as classes 1 through 4, and from the right front corner of the inner docking structure in a counterclockwise direction as classes 5 through 8 (Fig. 5).

Implementation details

We conducted training and testing using our custom dataset. The model was trained on a single RTX 4090 GPU employing the Automatic Mixed Precision (AMP) method. Input images were resized based on the Backbone architecture. We used AdamW as the optimizer with a learning rate of $1e-3$ for generalization, and experimentally found that AdamW demonstrated better generalization performance compared to Adam. The model are trained for 1500 epochs, and the best results among the measurements were reported for each epoch.

To enhance the model's generalization performance, Data augmentation techniques were applied. Specifically, ColorJitter and either a 180-degree rotation or a random rotation within ± 30 degrees were utilized to increase the training data.

In Table 1, FPS measurements were conducted on a Jetson AGX Orin 64GB Development Kit using FP32 precision to verify whether real-time execution is feasible.

To assess the effectiveness of the proposed method, we compare two settings: (i) a baseline with the heatmap sigma fixed at 1, and (ii) the Adaptive Sigma approach with $\sigma_{\min} = 0.5$ and $\sigma_{\max} = 1.5$.

We utilized the Efficient Perspective-n-Point (EPnP) algorithm provided by the OpenCV library, and the camera intrinsics were set to the values of the camera used to capture the dataset. The 3D coordinates of the UAV's keypoints were manually measured and defined in a coordinate system centered at the UAV's center of gravity, which also serves as the center of the docking area.

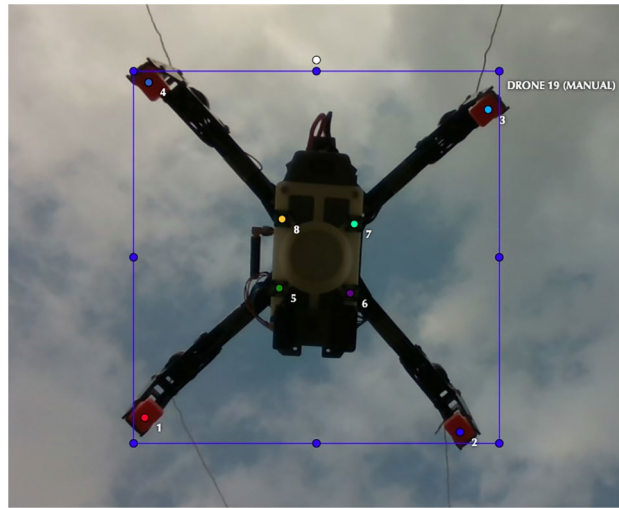


Fig. 4. Labeled keypoints of a drone in our dataset: classes 1 to 4 indicate the drone's legs in a counterclockwise direction, and classes 5 to 8 represent the inner docking structure corners in a counterclockwise sequence.

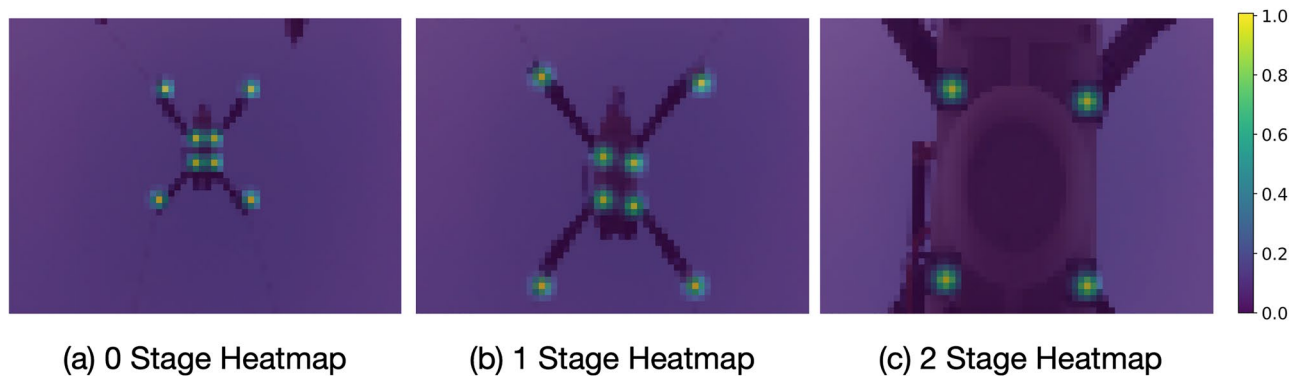


Fig. 5. Examples of heatmaps generated using adaptive sigma based on object distance: (a) 0.78m distance with 0.85 sigma, (b) 0.47m distance with 1.06 sigma, and (c) 0.11m distance with 1.28 sigma.

Evaluation metrics

This paper presents a practical approach for 6D pose estimation of automatic recovery UAVs in the absence of target pose information. Due to the difficulty in directly evaluating pose accuracy quantitatively, we present our results through visualizations, showcasing the effectiveness of our approach.

Our method utilizes a correspondence-based approach, where the performance of the PnP algorithm heavily relies on the accuracy of the 2D-3D correspondences. Therefore, we quantitatively evaluate our approach by comparing the performance of various model architecture for heatmap-based keypoint extraction on the target drone.

To construct the optimal architecture that meets the requirements of the task, we present three key metrics: Percentage of Correct Keypoints (PCK), frames per second (FPS) of the entire pipeline, and model parameter count.

PCK, a commonly used metric in human pose estimation as introduced by¹⁷, measures the accuracy of keypoint localization. It calculates the proportion of correctly predicted keypoints, where a keypoint is considered correct if the Euclidean distance between the predicted and ground truth keypoint locations is within a predefined threshold. This can be expressed as:

$$PCK@t = \frac{1}{N} \sum_{i=1}^N I(d(p_i, g_i) < t) \quad (10)$$

Where:

- N is the total number of keypoints.
- p_i represents the predicted 2D coordinates of the i -th keypoint within an image.

- g_i represents the ground truth 2D coordinates of the i -th keypoint within an image.
- $d(p_i, g_i)$ calculates the Euclidean distance between the predicted and ground truth keypoint locations.
- t is the predefined distance threshold.
- $I(\text{condition})$ is an indicator function that returns 1 if the condition is true, and 0 otherwise.

By analyzing these metrics, we can comprehensively assess the accuracy, efficiency, and complexity of our proposed method for 6D pose estimation of automatic recovery UAVs.

Experiment results and discussion

Quantitative results

Table 1 presents the quantitative results of our experiments, comparing various state-of-the-art backbone architectures in terms of keypoint detection accuracy, computational efficiency, and model complexity. The PCK metric at different thresholds provides a comprehensive assessment of the models' ability to localize keypoints accurately.

Our results demonstrate that the proposed heatmap-based keypoint detection method effectively identifies the UAV's feature points across a wide range of backbone architectures. Across all backbones tested, the average PCK scores are consistently high, indicating that the majority of keypoints are localized within acceptable error margins. This high level of accuracy is essential for establishing precise 2D-3D correspondences required for reliable 6D pose estimation through the PnP algorithm.

Lightweight models such as EfficientViT-M0 and PoolFormer-S12 achieve average PCK scores of 84.2% and 90.7%, respectively, demonstrating that even models with fewer parameters can effectively detect keypoints. Mid-sized models like FastViT-S12 and ViT-S/16 show higher average PCK scores of 91.4% and 94.1%, respectively. Larger models, including LiTV2-S and RegNetY-3.2G, achieve even higher average PCK scores of 92.5% and 94.1%, respectively. These results indicate that our heatmap-based keypoint detection method is effective across various model sizes and architectures, confirming its robustness and general applicability for localizing drone-specific features.

To further enhance keypoint localization accuracy, we introduced the Adaptive Sigma technique in the heatmap generation process. By adjusting the sigma value based on the UAV's distance from the camera, the model better accounts for variations in the object's apparent size, leading to improved keypoint detection performance. As shown in Table 1, incorporating Adaptive Sigma results in consistent improvements in PCK across most models. When applying Adaptive Sigma, we observe improvements in average PCK scores across different groups of backbones. Lightweight models, such as EfficientViT-M0 and PoolFormer-S12, see improvements of 1.9% and 1.6%, respectively. Mid-sized models like FastViT-S12 and CoAtNet-0 exhibit improvements of 1.2% and 2.2%, respectively. Larger models, including LiTV2-S and RegNetY-3.2G, show significant improvements of 2.5% and 1.6%, respectively. Notably, even for metrics demanding high precision, such as PCK@0.0025, using the Adaptive Sigma approach yields substantial performance gains. These enhancements indicate that the Adaptive Sigma method effectively enhances the model's ability to localize keypoints accurately across different architectures.

Furthermore, we assessed the real-time capability of our method to ensure practical usability in UAV recovery operations. The FPS measurements in Table 1 demonstrate that our models maintain high computational efficiency, with several models achieving over 50 FPS on the Jetson AGX Orin platform. This confirms that our approach is suitable for real-time applications, balancing both accuracy and efficiency.

We excluded the FPS measurement for the LiTV2-S backbone because its PyTorch implementation did not function properly in the Jetson environment. Additionally, the ViT backbone exhibits higher FPS compared to its FLOPs, which we attribute to its design being optimized for efficient computation within PyTorch.

To assess the statistical significance of the observed improvements, we performed a one-tailed paired t-test on the per-image PCK scores. Figure 6 visualizes the comparison at the strictest threshold (PCK@0.0025) and summarizes the statistical significance of the improvements.

As shown in the Fig. 6, our Adaptive Sigma method leads to a statistically significant performance increase for the vast majority of the tested architectures. Models spanning various families, from the lightweight EfficientViT-M0 to the larger LiTV2-S, demonstrate highly significant gains (marked with **, $p < 0.005$). This provides strong evidence that our proposed method effectively enhances the precision of keypoint localization. Interestingly, the models that did not show significant improvement, such as the pure ViT, are those with architectural properties that may not fully align with our scale-aware learning targets, as we will discuss further in the ablation study.

Conversely, models that did not exhibit significant improvements, such as ViT-S/16 and MobileOne-S3, possess distinct architectural properties that may limit the benefits of the Adaptive Sigma mechanism. Unlike hybrid transformer backbones, ViT-S/16 is a pure transformer architecture that relies solely on global self-attention from its earliest layers, thereby lacking strong local inductive biases. We hypothesize that, in such architectures, the dynamically varying precision of the heatmap targets may interfere with the model's ability to establish spatial correspondences from scratch. In the case of MobileOne-S3, which employs a highly sensitive re-parameterization scheme, the non-stationary nature of the learning targets could hinder the stable convergence required for effective branch fusion, ultimately resulting in performance degradation.

To better evaluate the effectiveness of Adaptive Sigma across different distances, we conducted an ablation study by dividing the validation set into four distance bins. We selected three representative backbones—PoolFormer-S12 (15.6M), FastViT-S12 (21.5M), and LiTV2-S (40.7M)—which represent small, medium, and large parameter scales, respectively. This setup allows us to examine how adaptive target scaling behaves across architectures of different capacities.

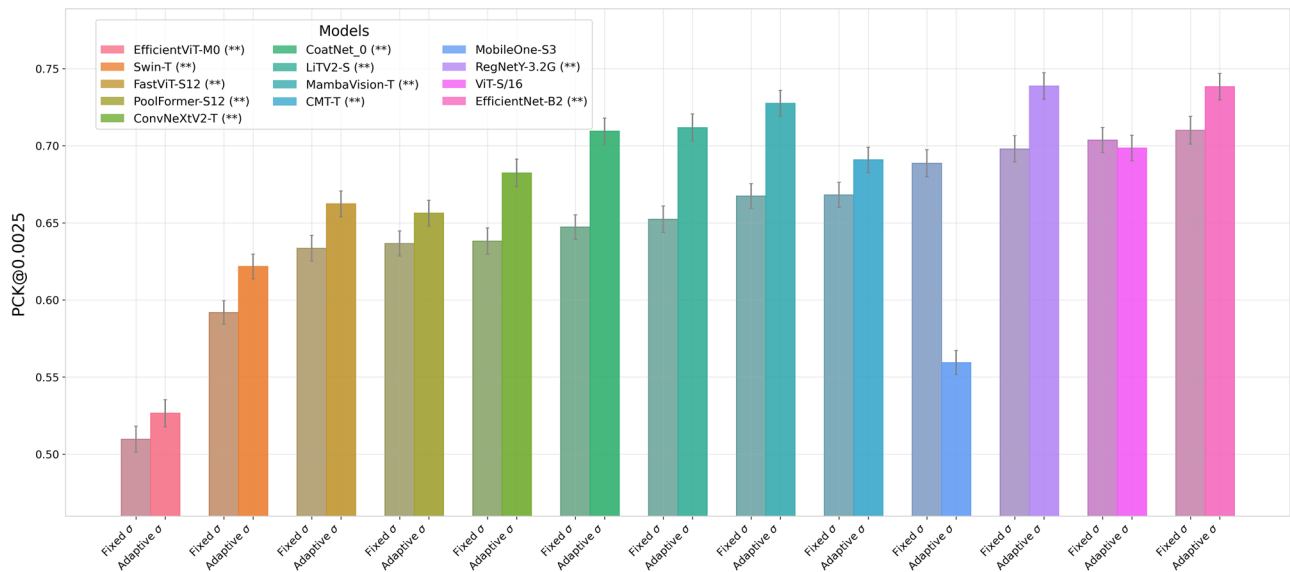


Fig. 6. Statistical significance of PCK improvements with adaptive sigma. The bar chart compares the macro-average PCK@0.0025 between the Fixed σ baseline and our Adaptive σ method for each backbone model. Error bars represent the standard error of the mean across all images in the validation set, with a typical magnitude of approximately 0.008. Model names marked with (**) indicate that the improvement from Adaptive σ is highly statistically significant ($p < 0.005$), as determined by a one-tailed paired t-test.

Backbone	Method	Far Bin	Mid-Far Bin	Mid-Near Bin	Near Bin
		($\sigma = 0.85$)	($\sigma \approx 1.07$)	($\sigma \approx 1.28$)	($\sigma = 1.50$)
PoolFormer-S12	Fixed Sigma	71.15	85.83	58.52	38.25
	+Adaptive	85.99 (+20.86%)	84.49 (-1.56%)	58.13 (-0.67%)	38.23 (-0.05%)
FastViT-S12	Fixed Sigma	83.41	81.49	55.97	33.32
	+Adaptive	88.21 (+5.75%)	83.39 (+2.33%)	56.74 (+1.38%)	40.29 (+20.92%)
LiTV2-S	Fixed Sigma	83.28	87.55	57.57	29.17
	+Adaptive	91.66 (+10.06%)	91.85 (+4.91%)	66.43 (+15.39%)	38.64 (+32.46%)
Bin	Samples (%)				
Far	406 (22.5%)				
Mid-Far	698 (38.8%)				
Mid-Near	351 (19.5%)				
Near	346 (19.2%)				
Total	1801				

Table 2. Ablation study comparing Fixed σ and Adaptive σ performance (PCK@0.0025) across different distance bins for representative backbones. The value in parentheses indicates the percentage point change from the baseline. The distribution of the validation dataset across these bins is detailed on the right.

As shown in Table 2, the results show a clear and consistent trend. Adaptive Sigma brings the largest improvements at the extreme distance ranges, where using a fixed sigma becomes less effective. For example, LiTV2-S achieves gains of +10.06% in the farthest bin and an even larger +32.46% in the nearest bin. Likewise, PoolFormer-S12 improves by +20.86% in far-distance cases. Although the mid-distance bins show smaller changes, the strong improvements at both ends demonstrate the robustness and adaptability of our method. These results indicate that adjusting sigma dynamically helps maintain accurate localization across different scales-especially under the challenging conditions of UAV recovery scenarios.

Qualitative results

While quantitative metrics such as PCK provide valuable insights into the accuracy of keypoint localization, it is essential to qualitatively evaluate the overall effectiveness of our proposed 6D pose estimation method in practical scenarios. To this end, we present qualitative results that demonstrate the capability of our method to accurately estimate the 6D pose of the UAV under various challenging conditions.

We selected three backbone networks-MambaVision, FastViT-S12, and LiTV2-S-based on a balanced consideration of performance, model size, computational efficiency, and FPS in our quantitative evaluations

We evaluated our method in several challenging scenarios:

1. UAV at varying proximities to the camera: The UAV appears at different scales within the image due to changes in proximity, which affects the resolution and detail of the UAV's features, making keypoint detection more challenging. To simulate these varying scales, the UAV was positioned at diverse altitudes.
2. UAV with significant tilt: The UAV may be significantly tilted relative to the camera, resulting in foreshortening and occlusion of certain keypoints.
3. Motion blur due to UAV movement: Rapid movement of the UAV can cause motion blur in the images, potentially degrading the quality of keypoint detection.

Figures 7 and 8 illustrate the qualitative results for these scenarios using the MambaVision, FastViT-S12, and LiTV2-S backbones, respectively. In each figure, we overlay the projected 3D model of the UAV onto the input image using the estimated 6D pose parameters. The alignment between the projected model and the actual UAV in the image serves as an indicator of the pose estimation accuracy.

In the varying proximities to the camera scenario, our method successfully estimates the 6D pose of the UAV across different scales within the image due to changes in distance. The Adaptive Sigma technique adjusts the sigma value in the heatmap generation process to account for the UAV's varying apparent size resulting from its proximity to the camera. This adjustment allows for accurate keypoint localization regardless of distance, leading to precise alignment of the projected 3D model with the UAV in the image.

In the case of significant tilt, the UAV presents unusual angles and potential occlusions. Our method demonstrates robustness in handling such challenging orientations. The keypoints are accurately detected even when some parts of the UAV are less visible, and the estimated 6D pose accurately reflects the UAV's orientation.

For the motion blur scenario, where the UAV's rapid movement causes blur, our method maintains reliable keypoint detection and pose estimation. Despite the degraded image quality, the model can infer the keypoint positions effectively. The use of heatmap-based keypoint detection proves advantageous in such situations, as it allows the model to consider probabilistic distributions over keypoint locations, rather than relying on precise pixel values.

In the representative scenarios and with the selected backbone networks, the qualitative results confirm that our proposed method provides accurate 6D pose estimations. By focusing on specific cases-such as the UAV at various altitude, with significant tilt, or experiencing motion blur-we demonstrate that accurate 2D-3D correspondences have been established, enabling the PnP algorithm to compute reliable pose parameters.

Both methods produce reasonable pose estimates without large deviations, which can be attributed to the robustness of the EPnP solver that yields a valid pose as long as the keypoints are reasonably accurate. However, by applying the Adaptive Sigma technique, the accuracy of keypoint localization is further improved, resulting in a more precise and stable 6D pose. This improvement is also quantitatively verified by the reduction in the mean difference **Diff(mean)** values within the images, indicating more accurate 2D-3D correspondences and, consequently, higher overall pose estimation precision.

Future works

Our proposed method has demonstrated effectiveness in real-time 6D pose estimation for UAV recovery operations, successfully utilizing heatmap-based keypoint detection and the PnP algorithm to obtain accurate UAV localization. This approach provides a practical solution that balances accuracy and computational efficiency, making it suitable for real-world applications where timely and precise pose estimation is essential.

However, despite these strengths, there are inherent limitations associated with the reliance on keypoint detection and the PnP solver. The PnP algorithm requires precise 2D-3D correspondences to compute an accurate 6D pose, necessitating the extraction of highly accurate keypoints. The creation of training datasets and the manual labeling of keypoints is a time-consuming and labor-intensive process, which can limit the scalability and applicability of the method to different UAV models or scenarios. Furthermore, in situations involving occlusion or extreme viewing angles, some keypoints may not be detectable, leading to less reliable 2D-3D correspondences and potentially inaccurate pose estimations. These challenges suggest that moving towards methods that can directly predict the 6D pose without relying solely on keypoint detection could enhance robustness and accuracy.

Future work will explore end-to-end approaches for 6D pose estimation that can directly regress pose parameters from image data. Leveraging deep learning techniques that integrate feature extraction and pose estimation into a unified framework may alleviate the dependency on manual keypoint labeling and improve performance under challenging conditions such as occlusions. Additionally, to address the dataset creation burden, we plan to this challenge by leveraging synthetic data augmentation and domain adaptation techniques. Recent advances in Human Pose Estimation (HPE) have shown that training with synthetic data and employing domain adaptation can improve model generalization to real-world data³⁰. Utilizing such synthetic datasets would also overcome a key limitation of the current study: the absence of direct pose ground truth. Simulation environments can provide precise 6D pose labels, enabling a more direct and rigorous evaluation of pose accuracy using standard metrics like ADD(-S). In future work, we plan to construct a large-scale synthetic dataset within a high-fidelity simulation environment to enable cross-model validation across different UAV types. By leveraging these simulation-based trainings, we aim to achieve robust sim-to-real transfer and ensure the proposed framework's applicability to diverse real-world scenarios.

Conclusion

We proposed a practical approach for real-time 6-DOF pose estimation of UAVs using monocular RGB imagery, specifically designed for the near-range docking phase in autonomous UAV recovery systems. By introducing

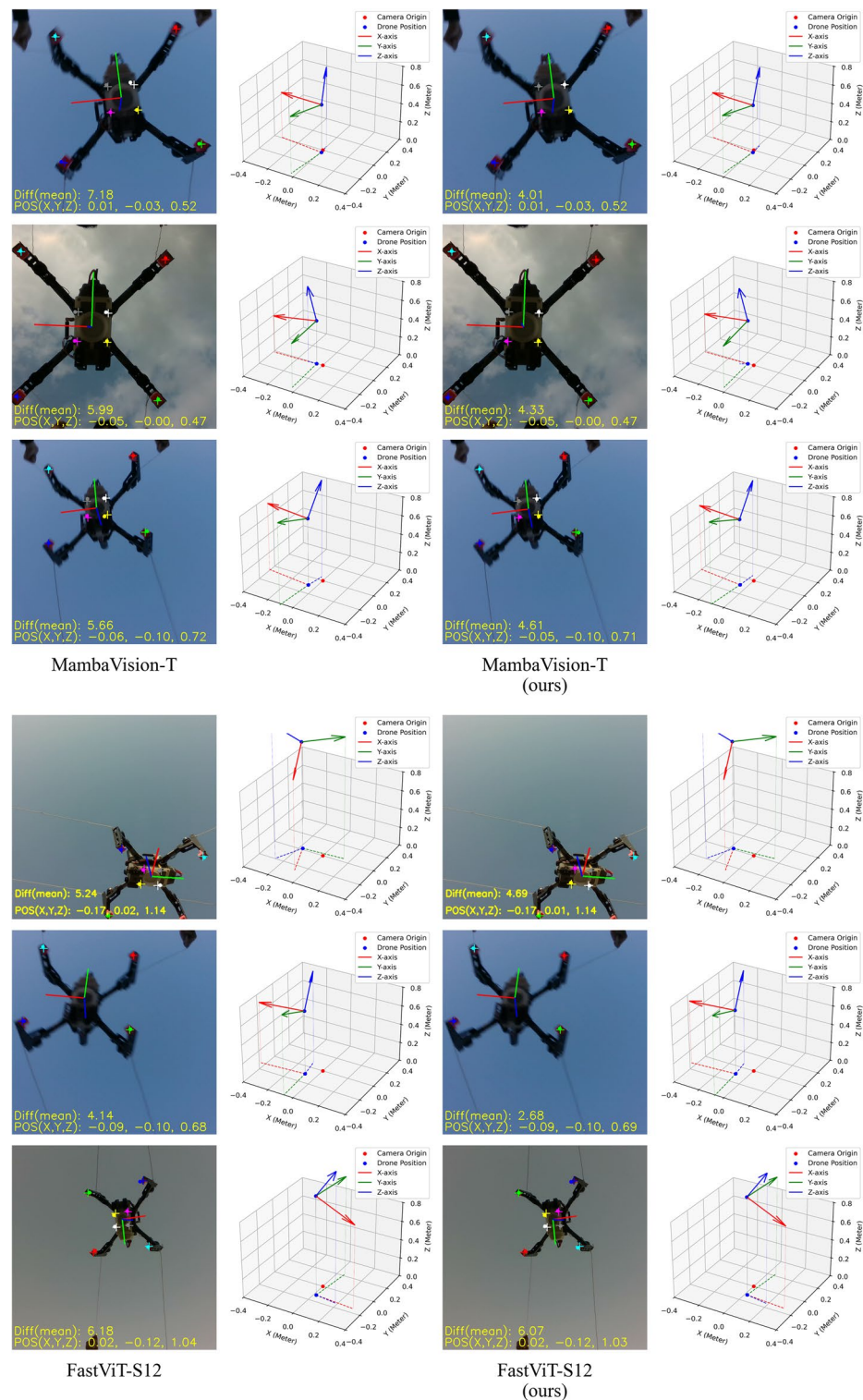


Fig. 7. Qualitative comparison of 6D pose estimation results using the **MambaVision** and **FastViT-S12** backbone, highlighting the improvements achieved by applying the Adaptive Sigma technique.

a novel Adaptive Sigma technique for heatmap-based keypoint detection, we enhanced localization accuracy according to the UAV's distance from the camera. By establishing accurate 2D–3D correspondences and solving for the pose through the PnP algorithm, the proposed method achieved high Percentage of Correct Keypoints (PCK) scores and demonstrated robustness under challenging conditions such as varying distances, significant UAV tilts, and motion blur.

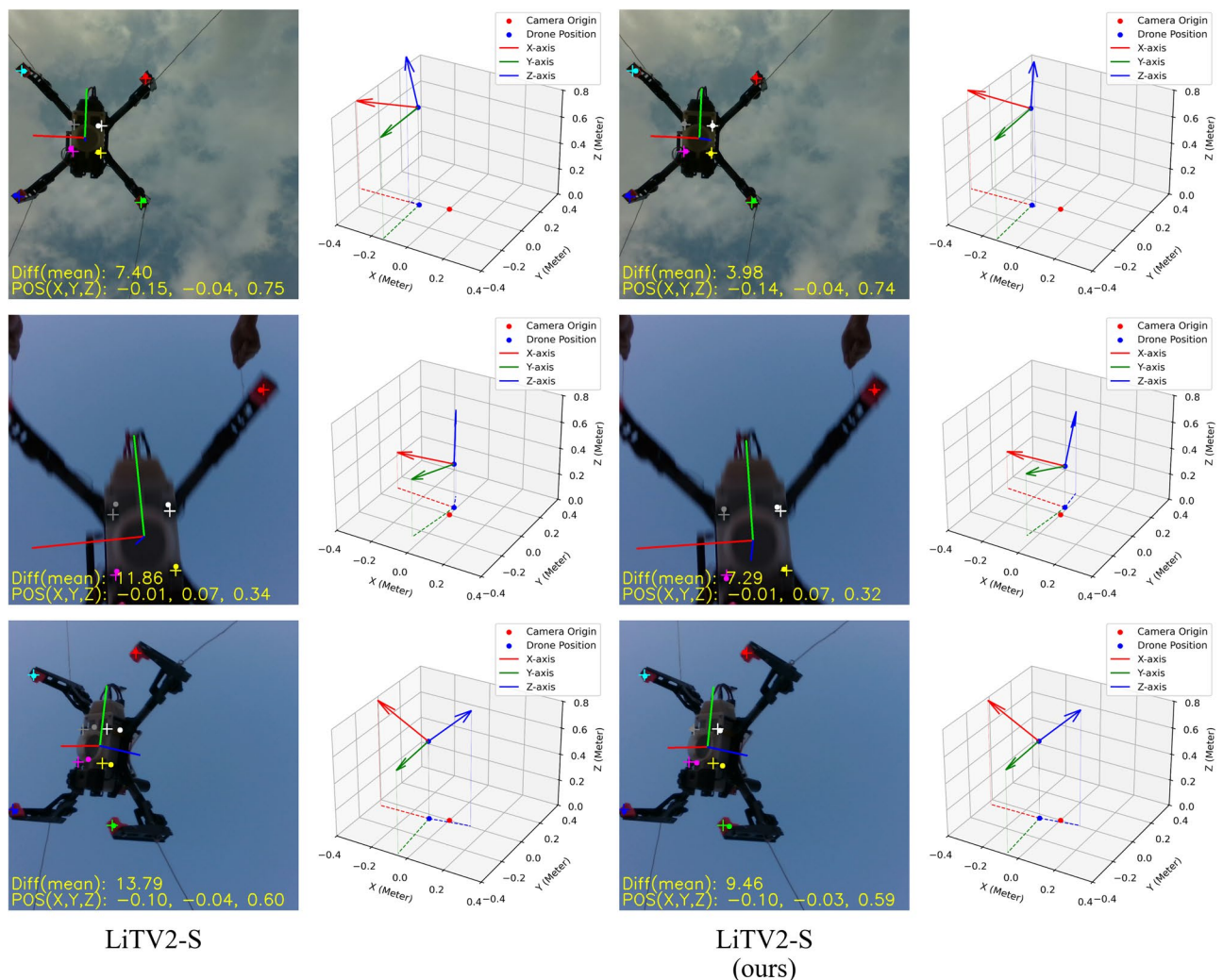


Fig. 8. Qualitative comparison of 6D Pose Estimation results using the LiTV2-S backbone, highlighting the improvements achieved by applying the Adaptive Sigma technique.

This study serves as a foundational module for high-precision, near-range localization in UAV recovery scenarios. Extensive experiments with multiple state-of-the-art backbone architectures verified both the accuracy and real-time capability of the proposed approach. Nevertheless, realizing a fully autonomous recovery pipeline will require integrating this fine-grained localization with long-range detection and mid-range tracking modules.

While acknowledging limitations related to the reliance on keypoint detection and manual dataset creation, our work provides useful insights for future research. Future efforts should focus on developing a unified framework that seamlessly integrates precise pose estimation with hierarchical tracking and control components. In addition, exploring direct 6-DOF pose regression and leveraging synthetic data for training could further enhance robustness and generalization across diverse UAV platforms and real-world environments.

Data availability

The datasets generated and/or analysed during the current study are available in the ph-10 huggingface repository at <https://huggingface.co/datasets/jaezic/ph-10>

Received: 12 August 2025; Accepted: 3 December 2025

Published online: 10 December 2025

References

1. Gryte, K., Sollie, M. L. & Johansen, T. A. Control system architecture for automatic recovery of fixed-wing unmanned aerial vehicles in a moving arrest system. *J. Intell. Robot. Syst.* **103**, 73 (2021).
2. Chu, L. et al. Aerodynamic configuration and control optimization for a novel horizontal-rope shipborne recovery fixed-wing uav system. *Aerosp. Sci. Technol.* **137**, 108253 (2023).

3. Zeng, W., Liu, J., Shan, J. & Wei, L. Optimization of deep stall landing trajectory for morphing aided uav. In *2021 33rd Chinese Control and Decision Conference (CCDC)*, 552–557 (IEEE, 2021).
4. Wu, J. et al. Design and optimization of uav aerial recovery system based on cable-driven parallel robot. *Biomimetics* **9**, 111 (2024).
5. Kim, H. J. et al. Fully autonomous vision-based net-recovery landing system for a fixed-wing uav. *IEEE/ASME Transac. Mech.* **18**, 1320–1333 (2013).
6. Peng, S., Liu, Y., Huang, Q., Zhou, X. & Bao, H. Pvnnet: Pixel-wise voting network for 6dof pose estimation. In *Proc. of the IEEE/CVF conference on computer vision and pattern recognition*, 4561–4570 (2019).
7. Wang, G., Manhardt, F., Tombari, F. & Ji, X. Gdr-net: Geometry-guided direct regression network for monocular 6d object pose estimation. In *Proc. of the IEEE/CVF conference on computer vision and pattern recognition*, 16611–16621 (2021).
8. Xiang, Y., Schmidt, T., Narayanan, V. & Fox, D. Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes. arXiv 2017. *arXiv preprint arXiv:1711.00199*.
9. Iwase, S., Liu, X., Khirrodar, R., Yokota, R. & Kitani, K. M. Repose: Fast 6d object pose refinement via deep texture rendering. In *Proc. of the IEEE/CVF International Conference on Computer Vision*, 3303–3312 (2021).
10. Li, H., Lin, J. & Jia, K. Dcl-net: Deep correspondence learning network for 6d pose estimation. In *European Conference on Computer Vision*, 369–385 (Springer, 2022).
11. Wen, B., Yang, W., Kautz, J. & Birchfield, S. Foundationpose: Unified 6d pose estimation and tracking of novel objects. In *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17868–17879 (2024).
12. Zheng, C. et al. Deep learning-based human pose estimation: A survey. *ACM Comput. Surv.* **56**, 1–37 (2023).
13. Lu, X. X. A review of solutions for perspective-n-point problem in camera pose estimation. In *Journal of Physics: Conference Series*, **1087**, 052009 (IOP Publishing, 2018).
14. Dosovitskiy, A. et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).
15. Xu, Y., Zhang, J., Zhang, Q. & Tao, D. Vitpose: Simple vision transformer baselines for human pose estimation. *Adv. Neural Inf. Process. Syst.* **35**, 38571–38584 (2022).
16. Zhang, F., Zhu, X., Dai, H., Ye, M. & Zhu, C. Distribution-aware coordinate representation for human pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7093–7102 (2020).
17. Yang, Y. & Ramanan, D. Articulated pose estimation with flexible mixtures-of-parts. In *CVPR 2011*, 1385–1392 (IEEE, 2011).
18. Liu, X. et al. Efficientvit: Memory efficient vision transformer with cascaded group attention. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 14420–14430 (2023).
19. Yu, W. et al. Metaformer is actually what you need for vision. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10819–10829 (2022).
20. Tan, M. & Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, 6105–6114 (PMLR, 2019).
21. Vasu, P. K. A., Gabriel, J., Zhu, J., Tuzel, O. & Ranjan, A. Fastvit: A fast hybrid vision transformer using structural reparameterization. In *Proceedings of the IEEE/CVF international conference on computer vision*, 5785–5795 (2023).
22. Guo, J. et al. Cmt: Convolutional neural networks meet vision transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 12175–12185 (2022).
23. Dai, Z., Liu, H., Le, Q. V. & Tan, M. Coatnet: Marrying convolution and attention for all data sizes. *Adv. Neural Inf. Process. Syst.* **34**, 3965–3977 (2021).
24. Liu, Z. et al. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, 10012–10022 (2021).
25. Vasu, P. K. A., Gabriel, J., Zhu, J., Tuzel, O. & Ranjan, A. Mobileone: An improved one millisecond mobile backbone. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 7907–7917 (2023).
26. Pan, Z., Cai, J. & Zhuang, B. Fast vision transformers with hilo attention. *Adv. Neural Inf. Process. Syst.* **35**, 14541–14554 (2022).
27. Woo, S. et al. Convnext v2: Co-designing and scaling convnets with masked autoencoders. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 16133–16142 (2023).
28. Hatamizadeh, A. & Kautz, J. Mambavision: A hybrid mamba-transformer vision backbone. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 25261–25270 (2025).
29. Radosavovic, I., Kosaraju, R. P., Girshick, R., He, K. & Dollár, P. Designing network design spaces. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10428–10436 (2020).
30. Peng, Q., Zheng, C. & Chen, C. Source-free domain adaptive human pose estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4826–4836 (2023).

Author contributions

A.J.C. conceived the original idea and supervised the project. M.J. designed and conducted the experiments, developed the methodology, and analyzed the results. M.J. wrote the initial draft of the manuscript. Both authors discussed the results and contributed to the final manuscript.

Funding

This work was supported by the Gachon University research fund of 2024(GCU-202500670001)

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.J.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025