aerospace

Article

Non-Cooperative Spacecraft Pose Estimation Based on Feature Point Distribution Selection Learning

Haoran Yuan [1], Hanyu Chen [2], Junfeng Wu [1,*] and Guohua Kang [1,*]

[1] School of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China; yuanhaoran@nuaa.edu.cn

[2] Jiangsu Shuguang Opto-Electronics Co., Ltd., Taizhou 225532, China; hanyuchen@nuaa.edu.cn

* Correspondence: awublack@126.com (J.W.); kanggh@nuaa.edu.cn (G.K.)

Abstract: To address the limitations of inadequate real-time performance and robustness encountered in estimating the pose of non-cooperative spacecraft during on-orbit missions, a novel method of feature point distribution selection learning is proposed. This approach utilizes a non-coplanar key point selection network with uncertainty prediction, pioneering in its capability to accurately estimate the pose of non-cooperative spacecraft, thereby representing a significant advancement in the field. Initially, the feasibility of designing a non-coplanar key point selection network was analyzed based on the influence of sensor layout on the pose measurement. Subsequently, the key point selection network was designed and trained, leveraging images extracted from the spacecraft detection network. The network detected 11 pre-selected key points with distinctive features and was able to accurately predict their uncertainties and relative positional relationships. Upon selection of the key points exhibiting low uncertainty and non-coplanar relative positions, we utilized the EPnP algorithm to achieve accurate pose estimation of the target spacecraft. Our experimental evaluation on the SPEED dataset, which comes from the International Satellite Attitude Estimation Competition, validates the effectiveness of our key point selection network, significantly enhancing estimation accuracy and timeliness compared to other monocular spacecraft attitude estimation methods. This advancement provides robust technological support for spacecraft guidance, control, and proximity operations in orbital service missions.

Keywords: pose estimation; uncertainty prediction; key point detection; non-cooperative spacecrafts; deep learning

check for updates

Citation: Yuan, H.; Chen, H.; Wu, J.; Kang, G. Non-Cooperative Spacecraft Pose Estimation Based on Feature Point Distribution Selection Learning. *Aerospace* **2024**, *11*, 526. https://doi.org/10.3390/aerospace11070526

Academic Editor: M. Reza Emami

Received: 12 May 2024
Revised: 22 June 2024
Accepted: 26 June 2024
Published: 27 June 2024

1. Introduction

To address the pressing sustainability crisis in near-Earth space, alleviate congestion in Earth's orbit, and extend the operational lifespan of geostationary satellites, the expedited development of safe and autonomous rendezvous and docking capabilities holds paramount importance [1]. Central to this endeavor is the acquisition of the target spacecraft's accurate, real-time pose. This enables autonomous servicing spacecraft to generate both safe and energy-efficient rendezvous and docking trajectories, grounded in real-time attitude estimations of the target [2]. However, the targets we encounter often consist of uncooperative entities, such as defunct satellites and debris, which are unable to provide identifiable markers or active communication links. Consequently, attitude estimation must solely rely on onboard sensors and computing capabilities. In comparison to active sensors, like those used for light detection and ranging, vision sensors boast simpler structures, lower mass, and reduced power requirements, facilitating their widespread utilization [3]. Nevertheless, to achieve autonomous attitude estimation, it is crucial to precisely extract attitude information from a sequence of images captured by a monocular camera, leveraging efficient and powerful computer vision algorithms. As camera technology continues to advance, the precision of attitude estimation steadily improves, providing robust technical support for addressing the sustainability crisis in near-Earth space [4].

In this paper, we aim to efficiently and swiftly determine the attitude of a non-cooperative spacecraft from a sequence of 2D images. The principal challenges in this task arise from the intricate nature of the target's distance and the surrounding environment, encompassing scenarios where the target is exceedingly remote, the backgrounds display a diminished signal-to-noise ratio, or the spacecraft is overshadowed. Initially, Drom [5] proposed an inverse-projection approach capable of estimating the positional attitude of a three-dimensional object in space from a single projected image. Subsequently, a comprehensive attitude analysis solution method was introduced. Following Drom's work, Horaud [6] presented a perspective-based four-point solution, while Lepetit [7] developed a non-iterative solution to the Perspective-n-Point (PnP) problem, leveraging an enhanced initialized Gaussian–Newton scheme to enhance computational efficiency and measurement accuracy. However, these model-based methods necessitate a substantial number of feature matches prior to pose estimation, limiting their real-time application potential. Alternatively, non-model-based algorithms often rely on target-specific characteristics, utilizing scale-invariant features like corner points to initialize the 3D geometric configuration of the spacecraft. In addition, many researchers have applied relative attitude methods to satellite missions. Chiodini, S [8] discussed two approaches to proximity based on monocular camera systems: the Sharma–Ventura–D'Amico (SVD) method and the silhouette matching method, which investigated the problem of initializing the relative attitude between a chaser and a non-cooperative target satellite; B. E. Tweddle [9] solved the external orientation problem for four-point features using a globally convergent nonlinear iterative algorithm; and S. Sharma [10] synthesized detected features using simple geometric constraints to greatly reduce the search space for the feature correspondence problem. These methods seek to establish the optimal correspondence between known 2D images and 3D feature perspectives to obtain the positional attitude. Nevertheless, if these methods lose sight of target features in certain scenarios, their accuracy suffers a significant decline.

In recent years, with the rapid development of neural networks, researchers have proposed numerous innovative learning-based algorithms for the problem of spacecraft pose estimation from monocular images, aiming to improve the detection accuracy. Among them, the features detected using CNN networks (the feedforward neural networks containing convolutional computation with deep structure) exhibit higher accuracy and stability compared to traditional methods. There are some methods of estimating the relative pose between two satellites based on non-AI methods. Furthermore, in order to bridge the gap between on-board computer vision applications and ground-based validation, SLAB (the Space Rendezvous Laboratory of Stanford University) and ACT (the Advanced Concepts Team of the European Space Agency) co-organized the International Satellite Attitude Estimation Competition. Based on the dataset from this competition, Chen [1] and Park [11] et al. took a similar approach to attitude estimation: they both used 2D pixel coordinates of key points and a pre-acquired wireframe model of the spacecraft to estimate attitude. One approach used heatmaps to process the images, while the other used features from CNN networks for deepening. We also conduct a study on this dataset to improve the pose estimation accuracy based on the effect of sensor distribution on pose measurement, and then select non-coplanar feature points for pose estimation. Our main contributions can be summarized as follows:

1. Utilizing the PnP angular error propagation model alongside the mathematical theory of sensor mounting configuration optimization, we can effectively analyze and guide the impact of coplanar and non-coplanar surfaces at crucial points on the accuracy of pose estimation.
2. In the process of detecting key points, the region encompassing the spacecraft's location is first identified. Subsequently, an uncertainty prediction is conducted for these key points. Based on theoretical knowledge, key points that exhibit high uncertainty are eliminated. To enhance accuracy, a non-coplanar feature point selection network

incorporating uncertainty is proposed. Finally, the bit pose is estimated utilizing the Efficient Perspective-n-Point (EPnP) algorithm.

3. We fully experimented with the SPEED dataset and compared it with the key point detection methods in various cases and found that our method can reduce the average error of the pose estimation by 61.3%.

2. Related Work

Methods for automatically extracting features of the target spacecraft for pose estimation through deep learning can be categorized as follows: two-stage methods, end-to-end methods, domain-adapted methods, and lightweight methods [12]. Among these methods, the two-phase approach is more commonly used, which usually utilizes a target detection network to intercept the region where the spacecraft is located, uses a key point prediction network to extract the key points in the spacecraft image, and finally solves the positional relationship using the PnP approach, as shown in Figure 1.

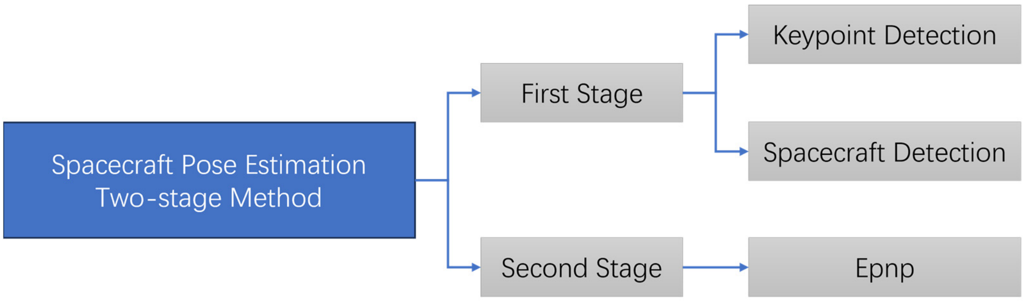

Figure 1. Two-stage approach to spacecraft attitude estimation.

Algorithms for detecting aircrafts are essential for localizing the critical point on spacecraft. When selecting algorithms, we usually go for single-stage target detection methods like YOLO (You Only Look Once) rather than two-stage methods like RCNN (Region-based Convolutional Neural Network), Fast RCNN, FPN (Feature Pyramid Networks), etc. This is due to the fact that, despite having good detection accuracy, two-stage target detection algorithms have a slow detection speed that is unable to satisfy real-time demands. The single-stage target detection algorithm, on the other hand, achieves a good balance between speed and accuracy by greatly enhancing detection speed while retaining high detection accuracy. For estimating the attitude of the spacecraft, the detection speed is particularly crucial. The single-stage target detection algorithm becomes our first option because we do not require a very high target detection accuracy.

In order to improve the accuracy of the pose estimation, the detection of key points is crucial. In terms of key point detection methods, pose regression, segmentation driving, heatmap prediction, and bounding box detection are four commonly used methods. Because of its ease of use and efficiency, the key point pose regression method is currently in high demand. For the first time, Chen [1] employed HRNet (High-Resolution Net) to regress the 2D key points straight to 1*1*2N vectors, giving each key point a region of the same size. They then solved a nonlinear least squares optimization problem to determine the relative position. Park [11] proposed a YOLO-v2 network-based MobileNet-v2 network for positional modeling using the YOLO-v2 architecture; in order to eliminate network layers from the original Efficient-Nets that do not support hardware deployment, Lotti [13] et al. proposed an Efficient-Net-Lite-based key point regression model. Furthermore, the heatmap prediction method demonstrates its own benefits. In order to more intuitively represent the distribution of key points, it regresses the heatmap, encoding probabilities based on key point locations. While Huo [14] et al. proposed a lightweight hybrid architecture that combines YOLO with heatmap regression sub-networks for key point prediction, which further improves the prediction accuracy, Gao's [15] work skillfully converts the prediction of direction vectors into the regression of heatmaps, improving the prediction accuracy. Another interesting method is the bounding box prediction model. Li [2] et al.

innovatively formulated key point prediction as a problem of key point bounding box detection, where the location of key points is determined by predicting closed bounding boxes and their confidence scores. This approach not only improves the detection accuracy but also provides more reliable data support for the subsequent positional solving.

In addition to two-stage methods, end-to-end methods, domain-adapted methods, and lightweight methods have their own advantages and disadvantages. The end-to-end method directly realizes the conversion from the original image to the bit-pose estimation by constructing a complex network model, which simplifies the processing flow but may sacrifice a certain degree of accuracy, and generally uses the bit-pose error to compute the loss function to train the model. Sharma uses CNN to extract the features in the image and uses the fully connected layer to output the six-dimensional vectors as the predicted 6D postures without intermediate phases. The domain adaptation approach, on the other hand, addresses the special nature of spacecraft images by introducing domain adaptation techniques to improve the generalization ability of the model in different scenarios. The lightweight approach focuses on reducing the computational and storage requirements of the model, making it more suitable for resource-constrained on-board computer vision applications.

3. Method

The overall flow of our method is depicted in Figure 2. Initially, we preprocess the existing dataset, analyze the sample labels, reconstruct the target in 3D using the sample data, and subsequently obtain the corresponding training data through projection processing. Subsequently, for each input image, we devise a target detection network to train it and derive the spacecraft's location and region. Then, we input the cropped spacecraft image into the key point detection network to detect the location of key points. After screening the uncertainty and key point structure, we determine the pixel coordinates of the key points to be settled in position. Finally, we perform pose solving using EPnP to obtain the relative pose parameters and arrive at the pose estimation results.

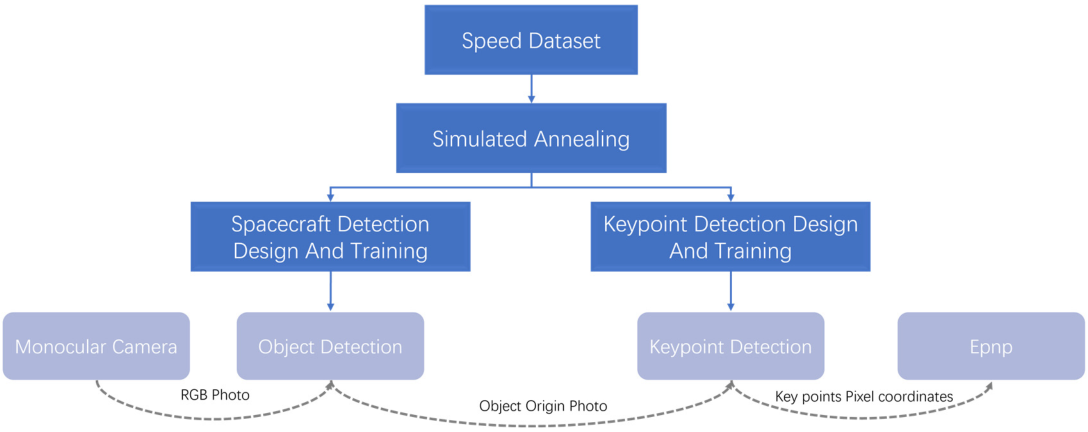

Figure 2. Flow of our proposed method for estimating the pose of a spacecraft from a monocular image.

3.1. Sensor Layout

Pose estimation of non-cooperative targets lacks explicit labeling information, and thus, we usually rely on on-board sensors working in concert with computers to achieve our goals. With the help of neural networks, we are able to detect a series of key points and then apply the PnP algorithm to accurately compute the positional relationship. Therefore, the importance of selecting which candidate key points to use as the basis for estimation cannot be overstated. By filtering out the accurate key points, we can improve the accuracy and reliability of the pose estimation and lay a solid foundation for subsequent use.

Wu's team [16] rigorously constructed an error propagation model for the laser lighthouse system and innovatively proposed the AP (Angle-Precision) mapping method to visualize the distribution of the position and attitude measurement accuracy of these sensor-mounted configurations under different observation viewpoints, aiming at solving the problem of the distribution of the measurement accuracy under the observation of different site locations. The team chose square and tetrahedral configurations for simulation analysis and set the observation distance r to 10 m, and the root mean square of the angular noise is 4.01×10^{-5} rad (0.0023°). The simulated "observation view angle-measurement accuracy mapping" of the two configurations is shown in Figure 3, which visualizes the relationship between the attitude, the measurement accuracy of the position, and the observation view angle.

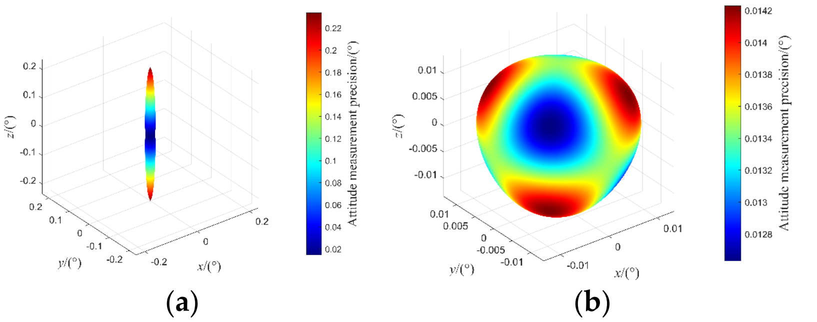

Figure 3. AP spherical mapping for coplanar and non-coplanar sensor configurations. (**a**) AP spherical mapping for square sensor mounting configuration. (**b**) AP spherical mapping for square tetrahedral sensor configuration.

The "x, y, z" axes in degrees represent the attitude measurement accuracy variation intervals of the three axes, which enables the visualization of the attitude measurement accuracy distribution with the observation viewpoint.

The experimental comparison results indicates that the measurement accuracy of the laser beacon in the non-coplanar configuration is more uniformly distributed with the observation viewpoint and is less likely to be drastically fluctuated due to changes in the observation viewpoint. Therefore, the non-coplanar configuration can effectively inhibit the dispersion of the measurement error and improve the stability and reliability of the measurement.

Based on the above theory, we reasonably hypothesize that the configuration of the key points is similar to that of the sensors, i.e., by adopting a non-coplanar key point selection method, the effect of the pose estimation will be better and the accuracy will be higher. In order to verify this inference, we next designed and implemented several comparative experiments, with a view to further confirming and optimizing our theoretical model through empirical data.

3.2. Three-Dimensional Reconstruction

In this paper, the world coordinate system refers to the target spacecraft coordinate system.

In the available dataset, each image is labeled in detail with the specific position and attitude of the spacecraft, while the information of the camera used in the dataset is known, and using the imaging model of the pinhole camera, as Equation (1) shows, we are able to project points onto the image plane:

$$Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{pmatrix} 1/d_x & 0 & u_0 \\ 0 & 1/d_y & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} f_x & 0 & 0 & 0 \\ 0 & f_y & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{1}$$

where $[X_w, Y_w, Z_w]^T$ is the position of the point in the world coordinate system, u and v represent the pixel coordinates in the image coordinate system, Z_C refers to the depth (distance from the camera) corresponding to the point, R is the pose matrix from the world coordinate system to the camera system, t is the translation from the world system to the camera system, f_x and f_y are the focal lengths of the two axes of the camera, d_x and d_y are the widths of the pixels of the camera in the direction of the two axes, and u_0 and v_0 are the offsets of the centers of the pixel planes with respect to the pixel coordinate origin [17].

In order to provide data labels for the subsequent target detection network, as well as the key point detection network, we selected 11 points with strong visual features as key points. Using labelImg1.8.0 software, we manually labeled the pixel values of these visible key points and subsequently solved Equation (2), minimizing the reprojection error of the key points to find their 3D coordinates [14].

$$\min_{X_{i,j}^{3D}} \sum_{i,j} \| X_{i,j}^{2D} - K_{T_j^*} X_{i,j}^{3D} \|_2^2 \tag{2}$$

where $X_{i,j}^{2D}$ denotes the $2D$ coordinates of the j-th key point in the i-th image, $X_{i,j}^{3D}$ denotes the corresponding $3D$ coordinates, T_{i}^* is the ground truth pose of image i that can be obtained from the dataset, and $K_{T_{\mathrm{j}}^*}$ denotes the point-to-image plane projection transformation.

The optimal solution cannot make Equation (2) zero due to the presence of noise. Therefore, the least squares method is used to obtain the optimal solution. We are able to construct a system of N linear equations through N graphs to form a super-determined linear equation, as shown in Equation (3):

$$As = b \tag{3}$$

where A is a 2N × 3 matrix and b is a 2N × 1 matrix. s can be obtained as the optimal solution:

$$s = (X_k, Y_k, Z_k)^T = \left(A^T A\right)^{-1} A^T b \tag{4}$$

The 3D coordinates of the key points are finally obtained, and from this, a spacecraft wireframe model can be obtained.

We were able to precisely calculate the 2D coordinates of the 11 important spots on the picture plane using the previously determined 3D coordinates, the true values of the quaternion and position supplied by each image in the Speed dataset, and the camera imaging model. In addition to being effective, this method saves the time and effort needed to manually identify a large number of photos. By using this method, we can obtain precise 2D coordinate data without having to rely on laborious human processes. In addition, since we need to determine whether or not it is coplanar, we also need to make coplanar and non-coplanar determinations for each point, so using the known coplanar and non-coplanar set of points and the 2D coordinate information mentioned above, the dataset for the subsequent task is jointly constructed.

3.3. Spacecraft Detection Network

In order to meet the stringent requirements of spacecraft pose estimation, we design the network for the accuracy and speed of the detection network. Considering the excellent performance of the yolov8 algorithm in the field of target detection, we decided to design and train the spacecraft detection network based on this algorithm.

First, we transformed the dataset obtained from the previous session into yolo format to fit our network training needs. The generation of 2D bounding boxes is particularly critical in this process. We utilize ground truth attitude information to accurately project 3D key points onto the image to obtain accurate 2D bounding boxes. To ensure that the spacecraft can be completely enveloped within the bounding box while minimizing the loss of key points, we enlarged the center frame of the bounding box [18].

Next, we designed the network structure and related parameters. The whole network structure is divided into three parts: backbone network, neck part, and head part [19]. The backbone network is responsible for extracting rich feature information from the input image, which lays a solid foundation for the subsequent detection tasks, while the neck part performs multi-scale fusion to make full use of the feature information at different scales and to improve the adaptability of the detection network to different sizes of spacecraft. Finally, the head part is shown in Figure 4, which is capable of outputting predicted values at three different scales, including information such as the location, size, and category of the bounding box.

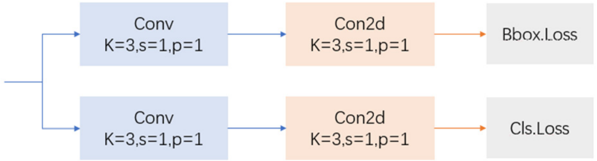

Figure 4. Head section of the spacecraft detection network.

Conv denotes a one-dimensional convolution, where only the width is convolved and not the height, and Conv2d denotes a two-dimensional convolution, where both the width and height are convolved. Furthermore, K represents the convolutional kernel size, s denotes the step size, and p denotes the additional boundary values added around the input data.

With this design, our spacecraft detection network is not only able to guarantee the detection accuracy but also meet the real-time requirement of spacecraft position estimation.

3.4. Key Point Detection Network

Considering the different relative distances between the spacecraft and the camera, we designed three rectangular frames of different sizes for each pixel to accommodate various detection scenarios. The framework of our final design of the KDN is shown in Figure 5, including the CSPDarknet backbone network, which combines the CSPNet (Cross-Stage Partial Network) and Darknet, multi-scale feature fusion, the design of each branch, and the final detection head [2].

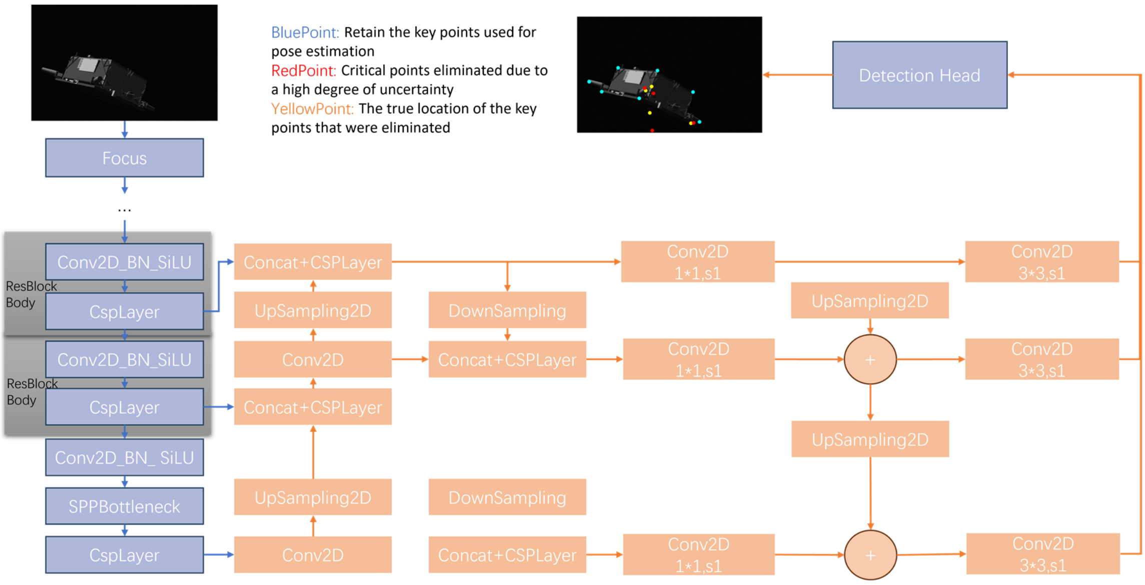

Figure 5. KDN framework.

In the KDN framework, the specific operation of the focus structure is to perform inter-pixel extraction in the direction of rows and columns in an image to form a new feature layer. Furthermore, the structure of CSPNet is somewhat similar to Residual, which is also divided into two parts, left and right, with the main part performing the stacking of residual blocks, and the other part acting like residual edges, which are connected to the end of the main part after a small amount of processing. The SiLU activation function is an improved version of Signoid and ReLU with lower and no upper bounds, smoothness, and non-monotonicity, which is better than ReLU on deep models.

We choose CSPDarknet as the main network in the network architecture to extract multi-scale features from the input photos. These features gradually decrease in size and increase in the number of channels as the network goes deeper. Convolutional layers are used to grab important information from the image in order to enhance detection performance even more. To improve the nonlinear representation of the network [20], activation functions like SiLU are used in conjunction with convolutional layers that vary in size and number of convolutional kernels.

Multi-scale feature fusion is a core aspect of our network design. We employ a feature pyramid network (FPN) to integrate feature information at different scales. Through down-sampling and up-sampling operations, we create feature maps at different scales and fuse them together through join operations. This fusion helps the network to fully capture the contextual information, which improves the accuracy and stability of key point detection.

Based on the feature fusion, we designed several network branches to accomplish different tasks. One branch is responsible for generating anchor points for determining possible object or key point locations; another branch is used to determine the uncertainty of key points to distinguish between high and low uncertainty frames [21]; and yet another branch is used to determine the relative positional relationship of key points, i.e., whether they are coplanar or not. These branches work together to achieve accurate detection of the spacecraft key points. Finally, we input all the features to the detection head for the final detection of the key points. The detection head outputs the precise position information of the critical point based on the input features and the outputs of the previous branches. We designed this network structure to fully utilize multi-scale features and contextual information to achieve accurate detection of spacecraft key points.

For detection and classification, we minimize the loss function L_{det}, which is commonly used in target detection.

$$\begin{cases} L_{\mathrm{det}} = \frac{1}{N}\sum\limits_{i=1}^{N}\left(L_{reg}(b_i,\tilde{b}_i) + L_{cls}(c_i,\tilde{c}_i) + L_{conf}(C_i,\tilde{C}_i)\right) \\ L_{reg}(b_i,\tilde{b}_i) = (x_i-\tilde{x}_i)^2 + (y_i-\tilde{y}_i)^2 + (\sqrt{w_i}-\sqrt{\tilde{w}_i})^2 + \left(\sqrt{h_i}-\sqrt{\tilde{h}_i}\right)^2 \\ L_{cls}(c_i,\tilde{c}_i) = -\sum\limits_{k=1}^{K}\left[\tilde{c}_{i,k}\log_2(c_{i,k}) + (1-c_{i,k})\log_2(1-c_{i,k})\right] \\ L_{conf}(C_i,\tilde{C}_i) = -\sum\limits_{k=1}^{K}\tilde{C}_{i,k}\log_2(C_{i,k}) - \sum\limits_{k=1}^{K}(1-C_{i,k})\log_2(1-C_{i,k})] \end{cases} \tag{5}$$

b_i, c_i, and C_i represent the box, key point category, and confidence level predicted by KDN for the *i*-th image. Moreover, x and y denote the pixel coordinates of the center point of the prediction frame on the image, and w and h denote the width and height of the prediction frame. Furthermore, $\tilde{b}_i$, $\tilde{c}_i$, and $\tilde{C}_i$ represent the corresponding labels. $L_{reg}(\bullet)$ denotes the MSE loss function, $L_{cls}(\bullet)$ and $L_{conf}(\bullet)$ denote the cross-entropy loss function, and N denotes the number of images in each batch.

Furthermore, for uncertainty prediction, we minimize the following loss function:

$$L_{uncertain} = \frac{1}{N}\sum_{i=1}^{N} L_{uncertain}(U_i,\tilde{U}_i) \tag{6}$$

where U_{i} represents predicted uncertainty, i.e., the probability of whether there is a target for each key point, and $\tilde{U}_{\mathrm{i}}$ represents the corresponding label. $L_{\mathrm{uncertain}}$ can be written as follows:

$$L_{uncertain}(U_i,\tilde{U}_i) = \frac{1}{K}\sum_{k=1}^{K}\left[-\tilde{U}_{i,k}\log_2 U_{i,k} - (1-\tilde{U}_{i,k})\log_2(1-U_{i,k})\right] \tag{7}$$

where $U_{i,k}$ represents predicted uncertainty for the *k*-th key point, and $\tilde{U}_{i,k}$ represents the corresponding label. The uncertainty label for the *k*-th key point can be calculated as follows:

$$\widetilde{U}_{i,k} = \begin{cases} 1 & \frac{L_{cls}(c_{i,k}, \widetilde{c}_{i,k})}{\log K} > 1 \\ \frac{1 - IOU(b_{i,k}, \widetilde{b}_{i,k}) + \frac{L_{cls}(c_{i,k}, \widetilde{c}_{i,k})}{\log K}}{2} & else, \end{cases} \tag{8}$$

where $IOU(\bullet)$ is the intersection ratio of the predicted box $\mathrm{b_i}$ and the ground truth box $\widetilde{\mathrm{b}}_\mathrm{i}$. K is the number of key point classes. The subscript k indicates that the variable is related to the k-th key point. To guide the KDN to realize the joint prediction of categorical uncertainty and regression uncertainty, the loss function of our KDN is defined as follows:

$$L = L_{\mathrm{det}} + L_{uncertain} \tag{9}$$

3.5. Pose Estimation

After obtaining the 3D coordinates and corresponding 2D coordinates of the key points, we utilize the EPnP algorithm to accurately compute the 6D attitude of the spacecraft [20]. To improve the accuracy of the attitude estimation, we first evaluated the uncertainty of the key points. An uncertainty threshold was set, and for the key points exceeding this threshold, we eliminated them due to their high uncertainty to ensure the reliability of the attitude estimation. Subsequently, for the key points detected by thresholding, we further classified the key point structure. This step aimed to identify the key points that were most critical for bit-pose settlement and determined their pixel coordinates [21]. In particular, we conducted comparative experiments between key points of coplanar and non-coplanar configurations to explore the effect of different configurations on the accuracy of bit-pose estimation. Eventually, we used all the filtered and classified candidate key points to identify the 6D position of the spacecraft via the EPnP algorithm.

This method not only improves the accuracy of the pose estimation but also provides a reliable database for our subsequent analysis and optimization.

4. Experiment

4.1. Datasets and Implementation Details

Vision mission training requires large image datasets, and obtaining accurate attitude labels is particularly costly for any target spacecraft in different orbits, so synthetic images are usually rendered using 3D models.

The SPEED dataset, which comes from the International Satellite Attitude Estimation Competition, comprises 300 Tango-modeled real satellite images that were taken by actual cameras at the Rendezvous and Optical Navigation Testbed (TRON) facility at the Space Rendezvous Laboratory (SLAB) at Stanford University [22] and 15,000 satellite images from the PRISMA mission that were synthesized using OPENGL rendering. Satellite photographs are shown in Figure 6. This dataset's model images have high-fidelity illumination conditions, a wide range of attitude distributions, and annotations with camera-corresponding parameters R and T that make it possible to obtain precise attitude labels.

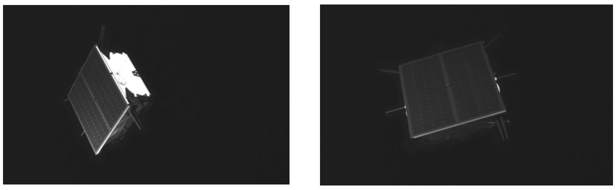

Figure 6. Tango model real satellite images.

We compared our method to that of Park and Chen, who both obtained good results using a two-stage method to estimate spacecraft position on the SPEED dataset. The term "two-stage approach" refers to the process of first employing deep convolutional networks to extract important points from the spacecraft image and then using the PnP method to solve the position. In general, the use of two-stage approaches for position estimation is innovative because it allows for the discovery of important points [3]. Park and Chen's methods differ in that they employ distinct techniques for detecting key points and determining the 2D coordinates of those locations. Park used the YOLO-v2 architecture and proposed a regression-based method for key point detection based on the position model of the MobileNet-v2 network [12]; Chen used HRNet to predict heatmaps for each key point and selected the location with the highest probability to determine the coordinates [11]. In order to compare with their methods, only the bit-position estimation method was compared, and we adopted the data enhancement method used by Park to maintain the consistency of the image processing.

During the experiments, we used the Adaptive Momentum Estimation (Adam) optimizer for 500 epochs of training with a learning rate of 0.001, a batch size of 64, a momentum of 0.8, and a weight decay of 5×10^{-4}.

4.2. Evaluation Metrics

In order to quantitatively evaluate our final pose estimation results, we used the evaluation metrics provided by ESA to define the estimation errors for translation, orientation, and 6D pose.

For the *i*-th image, the translation error, orientation error, and pose estimation error are calculated as follows:

$$E_{T,\mathrm{i}} = \frac{||\mathrm{x_i} - \widetilde{\mathrm{x}}_i||_2}{||\mathrm{x}_i||_2} \tag{10}$$

$$E_{R,i} = 2\arccos(< \widetilde{q}_i, q_i >) \tag{11}$$

$$E_\mathrm{i} = E_{T,i} + E_{R,i} \tag{12}$$

where x_i denotes the predicted translation vector, $\widetilde{x}_i$ denotes the true translation vector, q_i denotes the predicted direction vector, $\widetilde{q}_i$ denotes the true direction vector, $||\cdot||$ denotes the second-paradigm number of the two vectors, $\langle \cdot, \cdot \rangle$ calculates the angle between the two vectors, and the error of the bit-pose estimation can be expressed as the sum of the translation error and the direction error. For the whole dataset, we can evaluate the pose estimation results by the six metrics of median and mean of the above errors. For example, the formula for *meanE* is

$$meanE = \frac{1}{N}\sum_{i=1}^{N} E_i \tag{13}$$

4.3. Experimental Results and Comparison with Benchmark Experiments

4.3.1. Performance with Synthetic Images

In this section, we utilized 1000 synthetic images for testing. Initially, an uncertainty test was conducted for the detected key points, followed by the execution of comparative experiments. From the Table 1, it can be seen that since the loss function and model construction are mainly driven by the pose data, the accuracy becomes lower for the average position error, but after adding the uncertainty judgment, the accuracy of most of the metrics becomes higher, and the effect of the average accuracy of the total error is better than the algorithm with added uncertainty. Consequently, we conducted experiments to select coplanar and non-coplanar key points, basing our selection on the key points filtered through uncertainty screening.

Table 1. Comparison of experiments with and without uncertainty options.

Method	medianET	medianER	medianE	meanET	meanER	meanE
Ours_Uncertainty	0.0042	0.0079	0.0121	0.0159	0.0207	0.0366
Ours_Non	0.0058	0.0118	0.0176	0.0152	0.0254	0.0406

Different numbers of key points were selected for the experiment and the results are shown in Table 2. From the table, we can see that by choosing the same number of key points, the error of the pose estimation is higher for the key points with the coplanar algorithm than the non-coplanar one; when choosing the same coplanar or non-coplanar algorithm, the more the number of key points is, the higher the accuracy of the pose estimation is.

Table 2. Performance of coplanar and non-coplanar algorithms on 1000 synthetic images.

Method	medianET	medianER	medianE	meanET	meanER	meanE
4 points (coplanarity)	0.0215	0.0631	0.0846	0.0382	0.0681	0.1063
5 points (coplanarity)	0.0174	0.0458	0.0632	0.0294	0.0635	0.0929
5 points (non-coplanarity)	0.0056	0.0163	0.0219	0.0203	0.0498	0.0701
7 points (non-coplanarity)	0.0036	0.0075	0.0111	0.0046	0.0102	0.0148

After careful selection of the key points and rigorous uncertainty screening, we have arrived at the definitive estimation performance, which is depicted in Figure 7. A comparison of our algorithm with other algorithms is shown in Table 3. It can be seen that our method performs better than Park's and Chen's in six indicators, and our method can reduce the average error of the bit-position estimation by up to 61.3% with fewer parameters and faster computation.

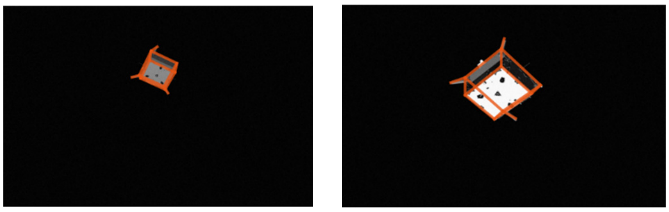

Figure 7. Performance of our algorithm for bit-position estimation on synthetic images.

Table 3. Comparison of our algorithm with Chen's and Park's algorithm on 1000 synthetic images.

Method	medianET	medianER	medianE	meanET	meanER	meanE
Ours	0.0036	0.0075	0.0111	0.0046	0.0102	0.0148
Chen	0.0047	0.0118	0.0172	0.0083	0.0299	0.0383
Park	0.0198	0.0539	0.0783	0.0287	0.0929	0.1216

4.3.2. Performance with Real Images

In this section, we utilized five real images for testing.

The accuracy of our coplanar or non-coplanar algorithms were degraded due to the large gap between the training and test set domains. Compared with other algorithms, all three methods tested on synthetic images were not as good as on real images, some of Chen's estimation results were worse, and the accuracy of our method was a little bit higher compared to the other two methods, which can prove that the generalization ability of our method is better (Tables 4 and 5).

Table 4. Performance of coplanar and non-coplanar algorithms on 5 real images.

Method	medianET	medianER	medianE	meanET	meanER	meanE
4 points (coplanarity)	0.0845	0.1259	0.2104	0.2196	0.2593	0.4789
7 points (coplanarity)	0.0712	0.0572	0.1284	0.0511	0.1645	0.2156
7 points (non-coplanarity)	0.0563	0.0247	0.0810	0.0345	0.1024	0.1369
11 points (non-coplanarity)	0.0158	0.0196	0.0354	0.0412	0.0901	0.1313

Table 5. Comparison of our algorithm with Chen’s and Park’s algorithm on 5 synthetic images.

Method	medianET	medianER	medianE	meanET	meanER	meanE
Ours	0.0158	0.0196	0.0354	0.0412	0.0901	0.1313
Chen	0.1253	0.2342	0.3595	0.1793	0.5457	0.7250
Park	0.0842	0.0965	0.1807	0.1135	0.1350	0.2485

5. Conclusions

A non-coplanar feature point selection network with uncertainty, exploiting the effect of sensor layout on pose measurement, is proposed in this paper, and the idea of integrating region detection into the spacecraft critical point detection task for estimating the pose of non-cooperative spacecraft is verified. With the reduction in the network parameters, the method proposed in this paper achieves a 61.3% reduction in the pose estimation error. In future work, we will investigate how to adaptively correct or remove the deviating bit-position quantities with large deviations and adaptively screen the critical points with high uncertainty instead of setting a given threshold to accommodate the trade-off between accuracy and efficiency.

Author Contributions: Conceptualization, H.Y. and J.W.; methodology, H.Y.; software, H.Y.; validation, H.C. and H.Y.; formal analysis, H.C.; investigation, H.Y.; resources, H.Y. and H.C.; data curation, H.Y.; writing—original draft preparation, H.Y.; writing—review and editing, J.W. and G.K.; visualization, H.Y.; supervision, J.W. and G.K.; funding acquisition, J.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by Jiangsu Province Entrepreneurship and Entrepreneurship Talent Project.

Data Availability Statement: The data presented in this study are openly available at https://kelvins.esa.int/satellite-pose-estimation-challenge/, accessed on 17 June 2024.

Conflicts of Interest: Author Hanyu Chen was employed by Jiangsu Shuguang Opto-Electronics Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Chen, B.; Cao, J.; Parra, A.; Chin, T.-J. Satellite Pose Estimation with Deep Landmark Regression and Nonlinear Pose Refinement. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Republic of Korea, 27–28 October 2019; pp. 2816–2824.
2. Li, K.; Zhang, H.; Hu, C. Learning-Based Pose Estimation of Non-Cooperative Spacecrafts with Uncertainty Prediction. *Aerospace* **2022**, *9*, 592. [CrossRef]
3. Kisantal, M.; Sharma, S.; Park, T.H.; Izzo, D.; Märtens, M.; D’Amico, S. Satellite Pose Estimation Challenge: Dataset, Competition Design, and Results. *IEEE Trans. Aerosp. Electron. Syst.* **2020**, *56*, 4083–4098. [CrossRef]
4. Park, T.H.; Märtens, M.; Jawaid, M.; Wang, Z.; Chen, B.; Chin, T.-J.; Izzo, D.; D’Amico, S. Satellite Pose Estimation Competition 2021: Results and Analyses. *Acta Astronaut.* **2023**, *204*, 640–665. [CrossRef]

5. Dhome, M.; Richetin, M.; Lapreste, J.-T.; Rives, G. Determination of the Attitude of 3D Objects from a Single Perspective View. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 1265–1278. [CrossRef]
6. Dhome, M.; Richetin, M.; Lapreste, J.T.; Rives, G. The Inverse Perspective Problem from a Single View for Polyhedra Location. In Proceedings of the Proceedings CVPR '88: The Computer Society Conference on Computer Vision and Pattern Recognition, Ann Arbor, MI, USA, 5–9 June 1988; pp. 61–66.
7. Lepetit, V.; Moreno-Noguer, F.; Fua, P. EPnP: An Accurate O(n) Solution to the PnP Problem. *Int. J. Comput. Vis.* **2009**, *81*, 155–166. [CrossRef]
8. Experimental Evaluation of Pose Initialization Methods for Relative Navigation Between Non-Cooperative Satellites | IEEE Conference Publication | IEEE Xplore. Available online: https://ieeexplore.ieee.org/document/9856146 (accessed on 17 June 2024).
9. Relative Computer Vision-Based Navigation for Small Inspection Spacecraft | Journal of Guidance, Control, and Dynamics. Available online: https://arc.aiaa.org/doi/10.2514/1.G000687 (accessed on 17 June 2024).
10. Sharma, S.; Beierle, C.; D'Amico, S. Pose Estimation for Non-Cooperative Spacecraft Rendezvous Using Convolutional Neural Networks. In Proceedings of the 2018 IEEE Aerospace Conference, Big Sky, MT, USA, 3–10 March 2018; pp. 1–12.
11. Park, T.H.; Sharma, S.; D'Amico, S. Towards Robust Learning-Based Pose Estimation of Noncooperative Spacecraft. *arXiv* **2019**, arXiv:1909.00392.
12. Zhu, W.S.; Mou, J.Z.; Li, S.; Han, F. Advances in spacecraft attitude estimation based on deep learning. *J. Astronaut.* **2023**, *44*, 1633–1644.
13. Lotti, A.; Modenini, D.; Tortora, P.; Saponara, M.; Perino, M.A. Deep Learning for Real Time Satellite Pose Estimation on Low Power Edge TPU. *arXiv* **2022**, arXiv:2204.03296.
14. Huo Fast and Accurate Spacecraft Pose Estimation From Single Shot Space Imagery Using Box Reliability and Keypoints Existence Judgments—DOAJ. Available online: https://doaj.org/article/ff655c5ff2ee4ae38c601f93e0a5b832 (accessed on 4 March 2024).
15. Proença, P.F.; Gao, Y. Deep Learning for Spacecraft Pose Estimation from Photorealistic Rendering. In Proceedings of the 2020 IEEE International Conference on Robotics and Automation (ICRA), Paris, France, 31 May–31 August 2020; pp. 6007–6013.
16. Chen, H.Y.; Wu, J.F.; Kang, G.H.; Wu, J.Q.; Li, X. A cooperative spacecraft position measurement method based on laser beacon and its accuracy analysis. *J. Instrum.* **2023**, *44*, 101–110. [CrossRef]
17. Wang, G. Neural Network Based Position Measurement Method for Non-Cooperative Spacecraft. Master's Thesis, Harbin Institute of Technology, Harbin, China, 2022.
18. Pan, J.; Ren, D.; Shi, Y.; Wang, L. Research on monocular position estimation algorithm based on improved YOLO6D. *Sens. Microsyst.* **2024**, *43*, 44–47+51. [CrossRef]
19. Pasqualetto Cassinis, L.; Fonod, R.; Gill, E. Review of the Robustness and Applicability of Monocular Pose Estimation Systems for Relative Navigation with an Uncooperative Spacecraft. *Prog. Aerosp. Sci.* **2019**, *110*, 100548. [CrossRef]
20. Axelrad, P.; Ward, L.M. Spacecraft Attitude Estimation Using the Global Positioning System—Methodology and Results for RADCAL. *J. Guid. Control. Dyn.* **1996**, *19*, 1201–1209. [CrossRef]
21. Fan, R.; Xu, T.-B.; Wei, Z. Estimating 6D Aircraft Pose from Keypoints and Structures. *Remote Sens.* **2021**, *13*, 663. [CrossRef]
22. Kelvins—Pose Estimation Challenge—Home. Available online: https://kelvins.esa.int/satellite-pose-estimation-challenge/ (accessed on 23 April 2024).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.