Domain Generalization for In-Orbit 6D Pose Estimation

Antoine Legrand *
UCLouvain, Louvain-la-Neuve, 1348, Belgium
KU Leuven, Leuven, 3001, Belgium
Aerospacelab, Mont-Saint-Guibert, 1435, Belgium

Renaud Detry †
KU Leuven, Leuven, 3001, Belgium

Christophe De Vleeschouwer ‡
UCLouvain, Louvain-la-Neuve, 1348, Belgium

We address the problem of estimating the relative 6D pose, *i.e.*, position and orientation, of a target spacecraft, from a monocular image, a key capability for future autonomous Rendezvous and Proximity Operations. Due to the difficulty of acquiring large sets of real images, spacecraft pose estimation networks are exclusively trained on synthetic ones. However, because those images do not capture the illumination conditions encountered in orbit, pose estimation networks face a domain gap problem, *i.e.*, they do not generalize to real images. Our work introduces a method that bridges this domain gap. It relies on a novel, end-to-end, neural-based architecture as well as a novel learning strategy. This strategy improves the domain generalization abilities of the network through multi-task learning and aggressive data augmentation policies, thereby enforcing the network to learn domain-invariant features. We demonstrate that our method effectively closes the domain gap, achieving state-of-the-art accuracy on the widespread SPEED+ dataset. Finally, ablation studies assess the impact of key components of our method on its generalization abilities.

I. Introduction

Future space missions such as space debris mitigation [1–3] or on-orbit servicing [4, 5] (*e.g.*, refueling, repairing or inspection of end-of-life satellites) are gaining in interest among both space agencies [1, 2, 4] and private companies [3, 5]. These missions involve Rendezvous and Proximity Operations (RPOs) between two spacecraft. While some previous missions conducted RPOs via tele-operation, autonomous operations are preferred nowadays because they are considered safer and cheaper. This motivates the development of Guidance Navigation & Control (GNC) systems that allow a spacecraft to navigate by itself at close range of its target [6]. A key component of this GNC system is the Navigation

*PhD Researcher, UCLouvain, ICTEAM, Dept. Electrical Engineering; KU Leuven, Dept. Electrical Engineering; and Aerospace Engineer, Aerospacelab, (antoine.legrand@uclouvain.be).
†Associate Professor, KU Leuven, Dept. Electrical Engineering; KU Leuven, Dept. Mechanical Engineering
‡Research Director, UCLouvain, ICTEAM, Dept. Electrical Engineering.

subsystem that has to estimate the 6D pose, *i.e.*, position and orientation, of the target spacecraft relative to the servicer.

The complexity of this estimation depends on whether the RPO is cooperative or uncooperative. In the former case, the target spacecraft is equipped with known fiduciary markers or inter-spacecraft communications capabilities that help the servicer in estimating their relative pose. In the latter case, the servicer has to estimate the pose from the sole information provided by its sensors. While cooperative RPOs have been the norm over the past, they do not suit with the requirements of future on-orbit servicing missions where the targets were not designed for cooperative operations. Since most of those RPOs involve known spacecraft, we study the case of an operation targeting an uncooperative spacecraft whose CAD model is available.

Different sensors can be used to observe the target spacecraft [7], such as LIDAR systems [8, 9], infrared [10, 11], time-of-flight [12] or stereo cameras [13]. However, their cost, mass, bulkiness, and power consumption make them unsuitable for low-cost missions in Low Earth Orbit. Over the past few years, the community started considering single cameras as the cheapest solution to the spacecraft pose estimation problem. Indeed, due to their low mass, low cost, and low power consumption, they can easily be integrated in less expensive satellites. Estimating the 6D pose of a target spacecraft from a single image is a complex task. The orbital lighting conditions differ significantly from those encountered on Earth. Since there is no atmospheric diffusion, the captured images suffer from large contrasts between exposed and shadowed spacecraft parts. This is further aggravated by the specular materials of which the spacecraft is made up. As there is no dust contamination, the spacecraft's surfaces remain specular forever. In addition, a spacecraft is often nearly symmetrical so that the proposed solution should be able to resolve ambiguities from subtle features. Finally, the pose estimation pipeline must run on limited resources provided by space-grade hardware [14, 15].

Deep-learning based solutions [16–18] have been proposed to solve the spacecraft pose estimation problem but they are exclusively trained on synthetic images that hardly capture the illumination conditions encountered in Low-Earth Orbit and rely on a simplified spacecraft model. Due to these mismatches between the synthetic and real domains, those solutions encounter significant generalization issues. This domain gap problem, *i.e.*, the mismatch between the synthetic training domain and the real test domain, has recently attracted a strong interest with the Spacecraft Pose Estimation Challenge (SPEC) 2021 [19] hosted by the European Space Agency and Stanford University. Several works [19–21] address the domain gap by exploiting techniques such as Generative Adversarial Networks (GANs) [21, 22], adversarial training [19] or pseudo-labeling [20]. Although those works successfully bridge the gap towards a specific target domain, they rely on domain adaptation strategies, *i.e.*, techniques that exploit priors on the target domain to bridge the domain gap. As a result, those techniques are unsuitable for real use cases where little information is known about the target domain. Unlike them, our work aims at bridging the domain gap through domain generalization, *i.e.*, techniques that require no prior knowledge of the target and aim at enlarging the range of domains covered by the trained model.

Our work addresses the domain generalization problem by introducing a learning strategy that relies on aggressive data augmentation policies, *i.e.*, a form of Domain Randomization, and multi-task learning to prevent the network from

overfitting on the training domain and enforce it to learn domain-invariant features. This strategy is applied on a novel, fully neural based, pose estimation network which embeds most of the salient components integrated in recent solutions proposed in the related literature. It relies on an intermediate detection of keypoints to recover the spacecraft relative 6D pose, decoupling the image processing step from the pose estimation process. We show that our method achieves State-of-The-Art accuracy without requiring on-orbit images at training.

II. Related Works

The first methods proposed for spacecraft pose estimation, such as the Sharma-Ventura-D'Amico method [23], only achieved accurate estimates with a high failure rate. Sharma and D'Amico [17] further introduced SPN, the first deep learning-based solution. The problem gained in interest with SPEC2019 [24]. Several methods emerged from SPEC2019. Chen *et al*. [16] won the challenge with an approach based on the prediction of pre-defined keypoints through a HRNet [25] backbone. The pose was then computed by solving a non-linear least squares problem using the Levenberg–Marquardt algorithm [26]. Park *et al*. [18] used a MobileNet-v2 [27] to regress keypoints, which were then used by a Perspective-n-Point (PnP) solver to predict the spacecraft pose. Similarly, Gerard *et al*. [28] predicted pre-defined keypoints through a DarkNet-53 [29] and estimated the pose using EPnP [30] combined with RANSAC [31]. WDR [32] improved this solution by considering a hierarchical architecture that predicts keypoints at different scales. Unlike all those works relying on a geometrical optimization problem, Proença and Gao [33] proposed an end-to-end solution, URSONet. They exploited a ResNet architecture [34] to extract features which were then used to regress the spacecraft position and predict its relative orientation through probabilistic quaternion fitting. Following SPEC2019, several works addressed the problem through end-to-end approaches [35, 36] or keypoint-based ones [37, 38]. Some of them [39, 40] further focused on lightweight approaches to meet the complexity requirements of space-grade hardware. In this paper, we introduce a novel method that is both keypoint-based and end-to-end. We further consider lightweight networks to decrease its computational complexity.

SPEC2019 also highlighted that all deep learning-based solutions overfit the synthetic training domain and therefore achieve a limited accuracy on real images. This problem, known as the domain gap, motivated a second edition, SPEC2021 [19], that brought several approaches to deal with the domain gap. They correspond either to domain adaptation or to domain generalization strategies. While the former assume that the target domain is known, the latter do not rely on any prior on this domain. As a result, *domain adaptation techniques use images from the target domain* during training while *domain generalization techniques only exploit synthetic images.* All top-scoring methods were unsurprisingly relying on domain adaptation techniques. The *TangoUnchained* team [19] used data augmentation policies and adversarial landmark regression. Wang *et al*. [21] further explored the use of adversarial techniques by using Cycle-GAN [41] to train the network on semi-real images, and self-training. Pérez-Villar *et al*. [20] relied on pseudo-labeling of the HIL domains. Finally, Legrand *et al*. [42] exploited Neural Radiance Fields [43] to augment the

size of a training set made of few real images to train a pose estimator. Regarding the domain generalization strategies, Park and D'Amico [44] introduced SPNv2 which relies on multi-task learning and image augmentation policies to bridge the domain gap. In addition, they proposed an Online Domain Refinement (ODR) technique to fine-tune a pre-trained model on-board during the mission to achieve on-orbit domain adaptation. Park and D'Amico further improved their work with SPNv3 [45] which considered additional data augmentations such as DeepAugment [46] and RandConv [47], as well as transformer-based architectures [48]. Ulmer *et al*. [49] relied on dense 2D-3D correspondences prediction combined with pose hypothesis and refinement as well as data augmentation techniques. Finally, Cassinis *et al*. [50] exploited augmentation techniques to train a HRNet [25] to perform keypoint regression. This paper bridges the domain gap through a learning strategy that aims at domain generalization, *i.e*., assuming no prior knowledge on the target domain. Our method builds on the domain generalization paradigm followed by SPNv2 [44] and Cassinis *et al*. [50] but offers a different perspective from those works. While the main purpose of the multi-task learning framework used to train the backbone of SPNv2 is to promote task-agnostic features, our framework further encourages the learning of a geometry-aware representation through the auxiliary face segmentation task. Similarly, while the Data Augmentation policies of SPNv2 and Cassinis *et al*. aimed at enlarging the training set distribution by synthesizing plausible samples, our Domain Randomization strategy goes one step further. Our augmentation policies are designed to aggressively transform the original images through, e.g., deletion of parts of the images, simulation of severe over/under-exposure, or texture alterations. This forces the network to exploit more subtle features.

III. Method

This section describes our method for enlarging the range of image domains properly managed when estimating the target 6D pose, *i.e*., position and orientation. It relies on a novel architecture as well as a domain generalization learning strategy. Section III.A describes the network architecture while the domain generalization strategy is developed in Section III.B, which introduces our Multi-Task Learning (MTL) framework, and Section III.C, which presents our domain randomization strategy.

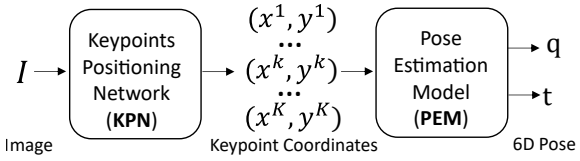


Fig. 1 Architecture overview. The target 6D pose is predicted by the Pose Estimation Model (PEM) trough pre-defined keypoints coordinates regressed from the input image by the Keypoint Positioning Network (KPN).

A. Network Architecture

The input image depicts a spacecraft at a variable distance from the camera. As a result, the spacecraft may occupy a substantial portion of the image or only a small part. To achieve scale invariance and limit the computational complexity of the method, only a crop of the original image centered on the spacecraft is provided to the subsequent network. We assume that a bounding box centered on the target is available, *e.g.*, through an object detection network [16, 18] or using a pose estimate obtained at a previous time step. This bounding box is used to crop the image and resize it to a resolution $W \times H$. As depicted in Figure 1, this image is processed through a Keypoint Positioning Network (KPN) to predict the coordinates of K pre-defined keypoints. Those coordinates are then processed by a Pose Estimation Model (PEM) which regresses the 6D pose through an attention-based encoder. This method decouples the image processing, *i.e.*, keypoint regression, from the pose estimation, thereby primarily restricting the generalization issue to the keypoint estimation module and leveraging the pose estimation robustness to deal with erroneous keypoints. Section III.A.1 describes the KPN architecture while the Pose Estimation Model is presented in Section III.A.2.

1. Keypoint Positioning Network

The Keypoint Positioning Network (KPN) aims at estimating the 2D coordinates of K pre-defined keypoints from the image. For this purpose, we rely on a first step of histogram equalization to bring the histograms of all the images, *i.e.*, training and testing, closer to a uniform distribution which results in a reduced gap between the domains. Section IV.C.1 provides an ablation study that demonstrates the interest of this equalization on the network generalization abilities.

The KPN relies on a high-resolution network backbone, *i.e.*, HRNet [25] that predicts feature maps at multiple resolutions. Those multi-resolution maps are then aggregated into K heatmaps of size ($W/4$, $H/4$) using a 4-layers Bi-directional Feature Pyramid Network (BiFPN) [51]. The normalized keypoint coordinates (x_N^k,y_N^k) are then obtained by applying a DSNT [52], *i.e.*, Differentiable Spatial to Numerical Transform, to each heatmap h^k. The DSNT is used because it achieves a better prediction accuracy and is differentiable. The normalized x and y coordinates of the k^{th} keypoint are then computed as follows,

$$x_N^k = \sum_{i=1}^{W/4} \sum_{j=1}^{H/4} h_N^k(i,j)x(i,j) \quad \text{and} \quad y_N^k = \sum_{i=1}^{W/4} \sum_{j=1}^{H/4} h_N^k(i,j)y(i,j), \tag{1}$$

where $x(i,j)$ and $y(i,j)$ are the x and y normalized coordinates at location (i,j), while h_N^k denotes the heatmap h^k normalized so that each heatmap sums up to 1.

Finally, the keypoint coordinates in the full-resolution input image (x^k,y^k) are computed as

$$(x^k, y^k) = (x_0 + x_N^k w, y_0 + y_N^k h), \tag{2}$$

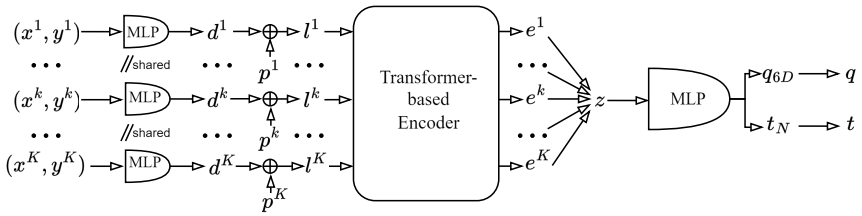


Fig. 2 Overview of our Pose Estimation Model which relies on a transformer-based encoder to predict the target 6D pose from the coordinates of K pre-defined keypoints.

where $[x_0, y_0, w, h]$ comes from the bounding box used to crop the spacecraft from the full resolution image.

The KPN is trained by minimizing the average Euclidean distance between the predicted and ground-truth normalized keypoint coordinates, (x_N^k, y_N^k) and $(\hat{x}_N^k, \hat{y}_N^k)$, respectively, *i.e.*,

$$L_{Kpts} = \frac{1}{K} \sum_{k=1}^{K} \sqrt{(x_N^k - \hat{x}_N^k)^2 + (y_N^k - \hat{y}_N^k)^2}. \tag{3}$$

2. Pose Estimation Model

As illustrated in Figure 2, the Pose Estimation Model (PEM) aims at predicting the spacecraft 6D pose, *i.e.*, position t and orientation q, from the 2D coordinates of the K pre-defined keypoints estimated by the KPN. First, to express the input coordinates in a representation that is more suited to the pose estimation task, each of those K 2D coordinates, (x^k,y^k), is normalized and mapped to a higher dimension through a Multi-Layer Perceptron (MLP) resulting in an embedding d^k of dimension D. Then, for each keypoint, a learned positional embedding p^k is added to d^k so that the resulting embedding, l^k, implicitly contains information on which physical keypoint is represented. The resulting K embeddings are fed into the encoder of a transformer which outputs K embeddings of the same dimension, which are concatenated into a single vector that is fed in a MLP made of three hidden layers with 256 neurons per layer. It outputs the normalized 3D position of the target in the camera frame as well as a 6D vector representing the rotation that aligns the spacecraft inertial frame to the camera frame. Predicting the rotation as a 6D vector has been inspired by Zhou *et al*. [53], who demonstrated that regressing rotations represented by Euler angles, axis-angle or quaternions was inefficient due to their discontinuous representation of the SO(3) space. At inference, the translation t in meters is computed from the normalized regressed translation vector while the 6D-vector representing the rotation is mapped to the corresponding quaternion q. A salient specificity of our architecture lies in the transformer-based encoder. Its purpose is to identify, through self-attention layers, consistent patterns in the input embeddings, thereby providing outliers-resilience capabilities to the model. Section IV.C.4 provides an ablation study on the PEM architecture.

The Pose Estimation Model is trained on top of the KPN. In this phase, only the PEM weights are optimized by

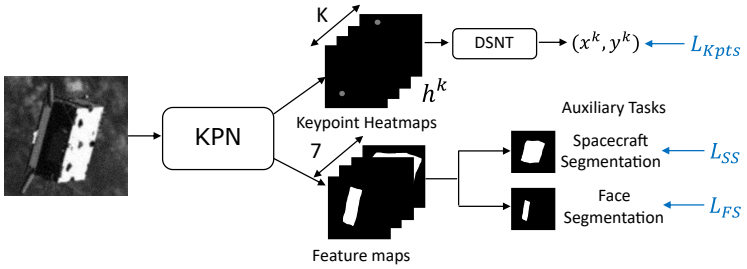


Fig. 3 Using our MTL framework, the network is supervised on the regression of keypoint coordinates and on two auxiliary tasks that aim at segmenting the spacecraft and each of its faces.

minimizing the PEM loss, *i.e.*, L_{PEM}, which combines the normalized translation error with the error between the ground-truth and predicted 6D representations of the rotations, *i.e.*,

$$L_{PEM} = \frac{\left\|t - \hat{t}\right\|_2}{||t||_2} + ||q_{6D} - \hat{q}_{6D}||_1 , \tag{4}$$

where t and $\hat{t}$ are the predicted and ground-truth positions, q_{6D} and $\hat{q}_{6D}$ are the 6D representations of the predicted and ground-truth rotations.

B. Multi-Task Learning

Inspired by previous works [19, 44] in domain generalization, our training procedure relies on multi-task learning to improve the KPN generalization abilities. In a multi-task framework, the learning of a main task is leveraged by learning auxiliary tasks in parallel with the main one. In our method, in addition to the main task of regressing keypoint coordinates, two segmentation tasks are considered. The first one aims at segmenting the spacecraft while the second one aims at segmenting each of the six faces of the spacecraft body. This multi-task learning strategy is illustrated in Figure 3, where the KPN outputs K keypoint heatmaps and seven feature maps used by the auxiliary segmentation tasks.

In addition to the loss on the keypoints (see Equation (3)), the KPN is therefore trained on both auxiliary tasks through Binary Cross Entropy losses [54], *i.e.*, L_{SS} and L_{FS}. The total loss back-propagated through the KPN, L_{KPN}, therefore sums up the three losses weighted by two hyper-parameters, β_{Kpts} and β_{Multi}, that determine the impact that the auxiliary tasks should have on the KPN training, *i.e.*,

$$L_{KPN} = \beta_{Kpts} L_{Kpts} + \beta_{Multi}(L_{SS} + L_{FS}). \tag{5}$$

C. Domain Randomization

The origin of the domain gap comes from the two main limitations of the image rendering tool used to generate the training set. Firstly, the illumination conditions encountered on-orbit are not captured by the rendering tool. Indeed, the

specular spacecraft parts reflect the oncoming light which results in over-exposed images or image artifacts such as lens flares. Furthermore, due to the lack of atmospheric diffusion, the spacecraft parts that are not directly illuminated cannot be seen. However, those over and under exposure impairments are poorly captured by the rendering tool. Secondly, the CAD model used for generating the training set suffers from mismatches with the actual spacecraft. For example, in the SPEED+ dataset [55], the Multi-Layer Insulation (MLI) represented in the CAD model is over-simplified compared to the actual one, as illustrated in Figure 4. These illumination and CAD mismatches are responsible for the domain gap problem. Hence, we introduce a domain randomization [56] technique that aims at randomizing the training domain through an aggressive data augmentation strategy.

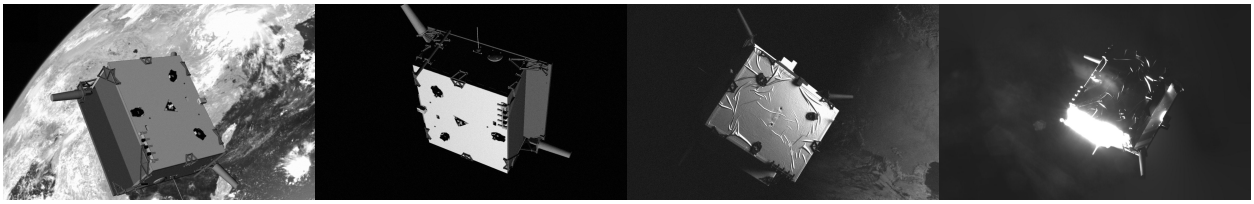


Fig. 4 SPEED+ [55] synthetic (left) and HIL images (right). HIL images exhibit highly textured surfaces, due to the Multi-Layer Insulation (MLI), that are not properly rendered in the synthetic images.

Several techniques are considered to achieve invariance against the illumination conditions in addition to the classical Gaussian noise and Brightness & Contrast augmentations. Hide&Seek [57] is used to deal with the under-exposed spacecraft parts by randomly erasing parts of the image while an exposure augmentation technique, inspired from PostLamp [58], is applied on multiple randomly selected points in the image in order to improve the network ability to cope with over-exposed spacecraft parts.

In addition, since Convolutional Neural Networks tend to overfit on the image texture [59] and since this texture is not properly rendered in synthetic images due to CAD mismatches, we introduce a texture augmentation technique that consists in applying either Neural Style Transfer [60] or a Fourier-based augmentation. This Fourier-based augmentation is inspired by previous arts [61, 62] and consists in computing the Fast Fourier Transform (FFT) [63] of the image and adding noise to its magnitude while preserving its phase. The augmented image is then reconstructed from the noisy magnitude and the preserved phase. Since the FFT phase is intact, the image semantic, *i.e.*, the spacecraft pose and shape, is conserved while its aspect is corrupted, resulting in a texture-augmented image. Figure 5 depicts some images from SPEED+ along their counterpart augmented with each of the considered data augmentations.

Finally, the domain randomization is achieved by applying the augmentation techniques in an approach similar to RandAugment [64]. We define a set of transforms composed of 6 data augmentation policies, *i.e.*, (i) Gaussian noise, (ii) Brightness & Contrast, (iii) Hide&Seek, (iv) Exposure, (v) Texture and (vi) no augmentation. For each image of the train set, at each epoch, we randomly pick three different augmentation policies from the set and apply them consecutively. As illustrated in Figure 6, this results in significantly randomized training images that preserve the semantic of the

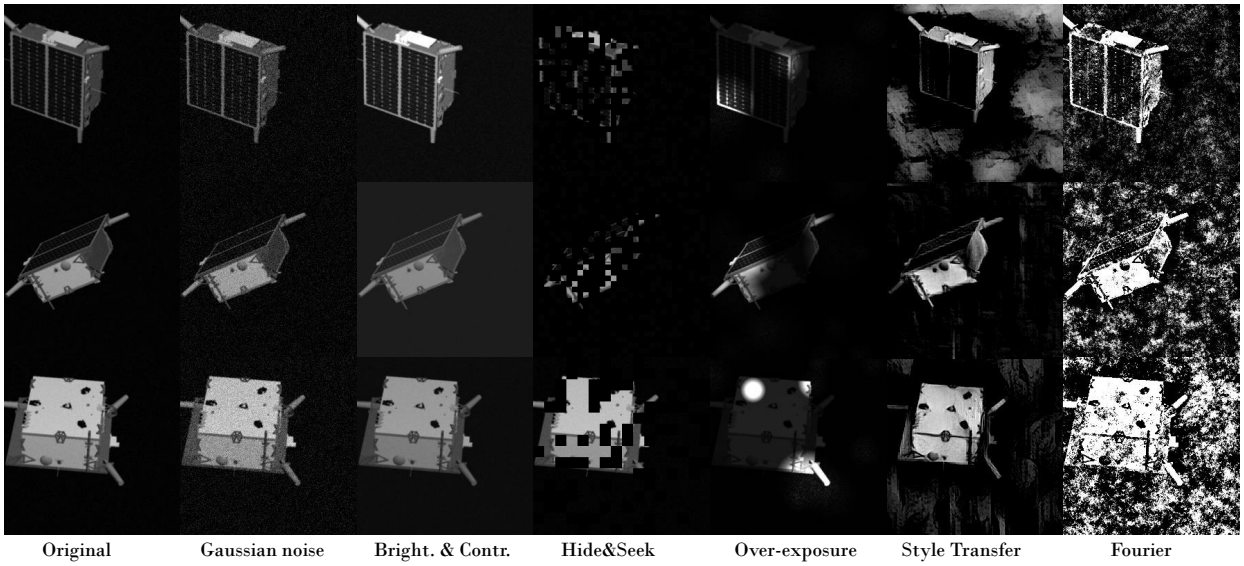


Fig. 5 SPEED+ [55] synthetic images and their augmented versions, considering the different data augmentation techniques used in this paper.

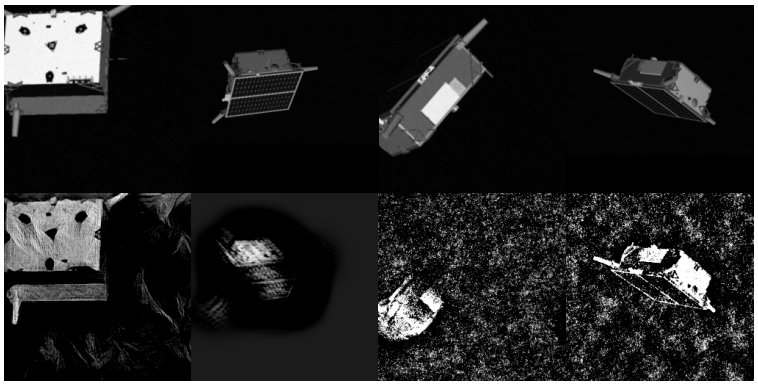


Fig. 6 Examples of our domain randomization technique. Original (top) and augmented (bottom) images from SPEED+ [55]. The image semantic is preserved but the texture and illumination significantly differ.

training set while significantly enlarging its distribution, consequently increasing the network generalization abilities.

IV. Experiments

This section assesses the effectiveness of our domain generalization strategy. Section IV.A describes the dataset, metrics and network settings used in our experiments. Section IV.B quantitatively evaluates the accuracy of our method and compares it with previous works. Finally, Section IV.C provides ablation studies that demonstrate the impact of the components of our method on its ability to bridge the domain gap.

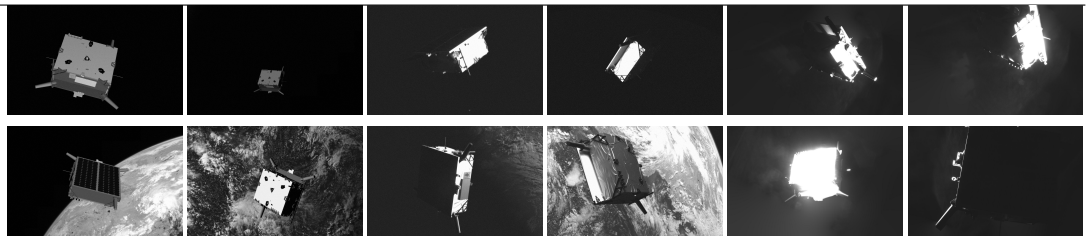
A. Experiment setup and evaluation metrics

The following experiments were conducted on the SPEED [17] and SPEED+ [55] datasets, used in the Spacecraft Pose Estimation Challenges (SPECs) 2019 [24] and 2021 [19], respectively, co-organized by the European Agency and Stanford University. SPEED+ [55] includes both synthetic and Hardware-In-the-Loop (HIL) images depicting TANGO, the target spacecraft of the PRISMA mission [65], as well as the associated pose labels, while SPEED [17] only contains synthetic images depicting the same target and the associated pose labels. Table 1 illustrates the three different domains encountered in SPEED+ [55].

The HIL images of SPEED+ [55], dubbed "*real*" in the rest of this paper, were captured in the TRON facility [66] which mimics the lighting conditions observed in Low Earth Orbit. Different illumination conditions are replicated by *lightbox* and *sunlamp*, the two HIL test set of SPEED+. *Lightbox* is made of 6740 images where the Earth albedo is replicated using light boxes while *sunlamp* is made of 2791 images where the direct illumination of the Sun is replicated using a metal halide lamp. In both domains, the distance from the target to the servicer ranges from 2.5 to 9.5 meters.

59,960 grayscale images of resolution 1920x1200 were generated through an OpenGL pipeline to form the synthetic set of SPEED+ [55]. The images were then processed with Gaussian blur and white Gaussian noise. In half of them, real satellite images of Earth were added in background. The distance from the target to the servicer ranges from 2.2 to 10 meters. SPEED [17] contains 15,000 synthetic images generated under the same conditions as in SPEED+, apart from the distance of the target which ranges from 3m to 50 meters. In the following experiments, to further increase the number of images seen by the network, both synthetic sets are combined and split into training and validation sets (80%-20%). Face segmentation masks are not available in SPEED+. Nevertheless, they are approximated using the keypoint coordinates of the spacecraft body. For each face of each training image, the coordinates of its 4 corners determines a quadrilateral. For each point within that quadrilateral, the point is considered as part of the face mask only if it is not occluded by another face. The purpose of the face segmentation mask being limited to encouraging the

Table 1 Overview of the SPEED+ dataset [55]. *Sunlamp* and *Lightbox* are two Hardware-In-the-Loop (HIL) sets which are used only for validation. Because of the direct illumination of the Sun, which causes over-exposed images, *Sunlamp* is the most challenging test set. The main differences between the HIL images and the synthetic ones are the illumination conditions and the spacecraft texture.

SPEED +	Synthetic	*Lightbox*	*Sunlamp*
Domain (Use)	Synthetic (Train)	HIL (Test)	
Illumination	Synthetic	Diffuse	Direct
Examples			

network to learn a geometry-aware representation, the supervision of the auxiliary face segmentation task does not require perfect labels.

In all experiments, the ROI is cropped based on the ground-truth keypoint locations. The KPN and PEM are trained according to the strategy described in Section III using the Adam [67] optimizer and cosine annealing [68] with an initial learning rate of $1e^{-3}$, a momentum of 0.9 and a batch size of 64 images. The KPN is trained with a weight decay of $1e^{-3}$ for 400 epochs while the PEM is trained with a weight decay of $1e^{-4}$ for 50 epochs. Following common practices on SPEED [16, 18], the $K = 11$ keypoints considered in the pipeline are the 8 corners of the spacecraft body and the top of its three antennas. The KPN input resolution is fixed to 256x256. The PEM embeddings contain 64 features. Its coordinate MLP contains a single hidden layer of 256 neurons. The attention-based encoder is made of six encoder layers, each of them consisting of a self-attention layer and a feed-forward layer of dimension 256.

Our pipeline is evaluated on the SPEED+ score, S_P^*, introduced in SPEC2021 [19], that averages the pose scores on the N samples of a test set. For each image in the set, the translation error, e_t, is computed as the norm of the difference between the predicted and ground-truth positions, t and $\hat{t}$, respectively, *i.e.*,

$$e_t = \left\| t - \hat{t} \right\|, \tag{6}$$

while the normalized translation error, $\bar{e}_t$, is defined as the ratio between the translation error and the norm of the ground-truth position, *i.e.*,

$$\bar{e}_t = \frac{e_t}{\left\| \hat{t} \right\|}. \tag{7}$$

The rotation error, e_q is the angular error between the predicted and ground-truth quaternions, q and $\hat{q}$, respectively, *i.e.*,

$$e_q = 2 \arccos\left(\left| \hat{q} q^T \right| \right). \tag{8}$$

Since the ground-truth positions and rotations on the HIL domains are not perfectly accurate because of a limited calibration accuracy, the estimates are considered as perfect if they are below a certain threshold, *i.e.*,

$$e_q^* = \begin{cases} 0, & \text{if } e_q < 0.00295\text{rad} \\ e_q, & \text{otherwise,} \end{cases} \quad \text{and} \quad \bar{e}_t^* = \begin{cases} 0, & \text{if } \bar{e}_t < 2.173\text{mm/m} \\ \bar{e}_t, & \text{otherwise} \end{cases} \tag{9}$$

Three metrics are used to evaluate the accuracy of the method. The average translation and rotation errors, E_T and E_R, respectively, are computed as the average of the images translation and rotation errors over the test set, *i.e.*,

$$E_T = \frac{1}{N} \sum_{i=1}^{N} e_t^{(i)}, \quad \text{and} \quad E_R = \frac{1}{N} \sum_{i=1}^{N} e_q^{(i)}. \tag{10}$$

Finally, the SPEED+ score, S_P^*, is defined as the sum of the normalized translation and rotation errors, averaged over the whole test set, *i.e*.

$$S_P^* = \frac{1}{N}\sum_{i=1}^{N}\left(\bar{e}_t^{*(i)} + e_q^{*(i)}\right). \tag{11}$$

B. Evaluation

As pointed out in Table 2, our method successfully bridges the domain gap. It achieves average errors of 9.3cm and 4.3° on *Lightbox* and 14cm and 6.9° on *Sunlamp*. On *Lightbox*, the median errors are of 6cm and 1.6° while those median errors are of 9cm and 2.5° on *Sunlamp*. On the SPEED+ synthetic validation set, the median errors are of 3 cm and 1°. Since those median errors are not significantly better than the ones obtained on *Lightbox* and *Sunlamp*, we conclude that the proposed approach successfully bridges the domain gap.

Table 2 also summarizes the accuracy obtained by State-of-The-Art methods. Our method outperforms all those which follow a domain generalization strategy, *i.e*., which do not rely on the knowledge of the test set at training, except for the EagerNet method [49], which is computationally more intensive. Furthermore, our method is on par with most domain adaptation approaches, *i.e*., which exploits the test domains during the network training. This highlights that a proper learning strategy can bridge the domain gap without using images from the target domain.

Table 2 also highlights the accuracy of our method used with a Lite-HRNet [69] backbone. Even, if this network is not as accurate as its HRNet-based [25] equivalent, it is less computationally intensive, *i.e*., the number of operations is divided by 5, while only doubling the errors. Given the low computing capabilities offered by space-grade hardware, it therefore presents an interest for future missions. This further demonstrates that large neural networks are not required

Table 2 Overview of the performance metrics averaged on the test datasets for different methods. Our method, which relies on a domain generalization strategy, outperforms the other domain generalization strategies, except EagerNet [49] which is more complex than the proposed approach. Furthermore, our method is on par with most domain adaptation strategies. Finally, our Lite-HRNet based variant achieves decent performance given its reduced computational load.

Method	Domain	Complexity	Lightbox			Sunlamp		
		[GFlops]	$E_T[m]$	$E_R[°]$	$S_P^*[/]$	$E_T[m]$	$E_R[°]$	$S_P^*[/]$
TangoUnchained [19]	Adaptation	$\gg$ Ours	0.11	3.19	0.07	0.09	4.30	0.09
Pérez-Villar *et al*. [20]	Adaptation	$\gg$ Ours	-	4.59	0.10	-	2.81	0.06
Wang *et al*. [21]	Adaptation	$\gg$ Ours	-	6.66	0.16	-	2.73	0.06
SPNv2 [44] w. ODR	Adaptation	20.8	0.15	5.58	0.12	0.16	9.79	0.20
EagerNet [49]	Generalization	>19	0.09	1.75	0.04	0.013	2.66	0.06
SPNv2 [44] w/o. ODR	Generalization	20.8	0.22	7.99	0.17	0.23	10.37	0.22
Ours (HRNet)	Generalization	6.3	**0.09**	**4.32**	**0.09**	**0.14**	**6.94**	**0.14**
Ours (Lite-HRNet)	Generalization	1.2	**0.14**	**7.42**	**0.15**	**0.20**	**15.12**	**0.30**

to successfully bridge the domain gap, contrary to what is commonly accepted [70].

C. Ablation Studies

This section aims at investigating the role of key components of our approach in its final accuracy. Sections IV.C.1, IV.C.2 and IV.C.3 analyze the impact of, respectively, the histogram equalization, the multi-task learning and the domain randomization, on the generalization abilities of our method. Finally, Section IV.C.4 focuses on the PEM architecture.

1. Histogram Equalization

Performing histogram equalization on the KPN input aims at reducing the distance between the test and training domains by enforcing the histograms of all images, *i.e.*, from both train and test domains, to follow an uniform distribution. As a result, the parts of the images that are over-exposed are shadowed while the ones that are under-exposed appear brighter. This decreases the gap between the domains and therefore improves the generalization abilities of the network. As pointed out in Table 3, this equalization decreases the errors by 25% on both sets.

2. Multi-Task Learning

As stated in Section III.B, our domain generalization strategy relies on multi-task learning to force our network to learn more generic features through auxiliary tasks. At inference, the auxiliary tasks are not inferred, so that the multi-task learning strategy does not increase the computational complexity of the network at inference. Our method relies on both spacecraft segmentation and faces segmentation as auxiliary tasks. As pointed out in Table 4, adding an auxiliary task decreases the errors by 30% on *Lightbox* and 25% on *Sunlamp*, compared to a network trained on the heatmap-based keypoint regression task only. Furthermore, if the network is trained on both segmentation tasks, the errors are decreased by 40% on *Lightbox* and 45% on *Sunlamp*.

3. Domain Randomization

As explained in Section III.C, our domain generalization strategy relies on domain randomization to enlarge the training set distribution in order to force the network to learn domain-invariant features through aggressive data

Table 3 Average performance metrics achieved by our method on *Lightbox* and *Sunlamp* with or without performing histogram equalization on the KPN input. Histogram Equalization improves the generalization abilities of our method.

	Lightbox			Sunlamp		
Equal.	$E_T[m]$	$E_R[^\circ]$	$S_P^*[/]$	$E_T[m]$	$E_R[^\circ]$	$S_P^*[/]$
×	0.11	6.20	0.13	0.14	8.55	0.17
✓	**0.09**	**4.32**	**0.09**	**0.14**	**6.94**	**0.14**

Table 4 Average performance metrics achieved on *Lightbox* and *Sunlamp* for different multi-task learning strategies, *i.e.*, considering (i) no auxiliary task, (ii) only spacecraft segmentation, or (iii) only faces segmentation, or (iv) both segmentation tasks as auxiliary tasks. The network generalizes better when trained on both auxiliary tasks.

Aux. Segm.	Lightbox			Sunlamp		
Tasks	$E_T[m]$	$E_R[^\circ]$	$S_P^*[/]$	$E_T[m]$	$E_R[^\circ]$	$S_P^*[/]$
No Aux. Task	0.13	7.40	0.15	0.18	12.87	0.26
Spacecraft	0.10	5.22	0.11	0.14	9.47	0.19
Faces	0.11	4.77	0.10	0.15	9.67	0.19
Both	**0.09**	**4.32**	**0.09**	**0.14**	**6.94**	**0.14**

augmentation techniques. Those data augmentations consist of Gaussian Noise, Brightness & Contrast, Hide&Seek [57], texture and exposure [58] augmentations, resulting in distorted training images. Figure 6 provides some examples of training images augmented through this domain randomization strategy.

Table 5 presents the average errors achieved by our method trained using different data augmentation strategies. With no data augmentation, the network completely overfits on the synthetic train set and provides average errors of 0.51m and 40.7° on *Lightbox* and 1.6m and 106° on *Sunlamp*. The testing images are so different from the images seen during training that the network has not learnt any robust features and does not generalize at all. Table 5 also provides the average errors achieved by the network when it is trained with a single data augmentation randomly applied to 50% of the images. While the usual augmentations, *i.e.*, brightness & contrast, hide&seek and Gaussian noise, reduce the average errors by 55% on *Lightbox* and 25% on *Sunlamp*, our texture-based augmentation achieves a reduction of 75% of the average pose score on both test sets. Interestingly, the exposure augmentation brings little gain, *i.e.*, 13% on *Lightbox* while it decreases by 75% the pose score on *Sunlamp*. Combining these augmentations through a RandAugment [64] policy significantly reduces the pose score, *i.e.*, by 89% on *Lightbox* and 93% on *Sunlamp*.

Finally, Table 5 also shows the average errors achieved when the network is trained using RandAugment [64] with all data augmentation policies but one. The augmentations that contribute the most to the generalization abilities of the network are the texture and exposure ones. Without either of these techniques, the average errors on *Sunlamp* increase by 65%. This can be explained from the fact that the exposure augmentation produces artifacts that are close to the ones observed in *Sunlamp* while the texture augmentation is well suited for this domain randomization task because it preserves the semantics of the image while randomizing its style.

4. PEM Architecture

Both HRNet [25] and Lite-HRNet [69] have been used in spacecraft pose estimation pipelines [16, 40]. The novelty of our approach comes from the attention-based Pose Estimation Model used to predict the 6D pose from a set of predicted keypoint coordinates using a continuous 6D rotation representation [53]. This section evaluates the impact of

Table 5 Average performance metrics achieved on *Lightbox* and *Sunlamp* for data different data augmentation strategies. While our network trained without Data Augmentation does not generalize at all, the same network trained with our Domain Randomization technique achieves a state-of-the-art accuracy. This demonstrates the interest of aggressive data augmentation policies for Domain Generalization. The two augmentations that contribute the most to the generalization capabilities are the texture and exposure ones.

Data Augmentation Techniques					Lightbox			Sunlamp		
Brightness	Hide&Seek	Exposure	Texture	Noise	$E_T[m]$	$E_R[^\circ]$	$S_P^*[/]$	$E_T[m]$	$E_R[^\circ]$	$S_P^*[/]$
					0.51	40.71	0.79	1.64	106.33	2.14
✓					0.24	18.32	0.36	0.90	79.91	1.56
	✓				0.18	14.24	0.28	0.632	74.22	1.40
		✓			0.53	34.47	0.69	2.318	35.5	1.82
			✓		0.18	9.29	0.19	0.373	26.89	0.53
				✓	0.29	18.79	0.38	1.544	87.40	1.78
	✓	✓	✓	✓	0.11	5.55	0.11	0.15	9.18	0.19
✓		✓	✓	✓	0.10	**4.31**	0.09	0.14	7.83	0.16
✓	✓		✓	✓	0.10	4.75	0.10	0.17	12.03	0.24
✓	✓	✓		✓	0.10	4.72	0.10	0.17	12.37	0.24
✓	✓	✓	✓		0.12	6.25	0.13	**0.14**	7.39	0.15
✓	✓	✓	✓	✓	**0.09**	4.32	**0.09**	**0.14**	**6.94**	**0.14**

some of its design choices on its accuracy.

Table 6 presents the average errors achieved by our method considering several variations from the proposed PEM architecture on two test sets, *i.e.*, *Lightbox* and *Sunlamp*. It shows that even if directly predicting the pose from the keypoints through a MLP works, expanding the dimension of the keypoints before processing them through this MLP improves the performances on all sets. In addition, pre-processing the keypoints through an attention-based encoder further improves the accuracy. Finally, adding a positional embedding to the keypoint embeddings provides additional gains. Table 6 also shows that the use of the continuous 6D representation for regressing rotations significantly increases the accuracy of the network.

V. Conclusion

This paper addressed the domain gap problem of the spacecraft pose estimation task. Our solution follows a usual architecture in spacecraft pose estimation, *i.e.*, heatmap-based keypoint regression followed by pose estimation. This two-step approach concentrates the image processing generalization issue on the keypoint regression part, and leverages on the robustness of the pose estimation module to deal with erroneous keypoint predictions. In our work, the pose estimation exploits an attention-based encoder to achieve robust pose prediction, in the form of a continuous 6D rotation and normalized 3D position, from the keypoint coordinates.

We introduced an efficient learning strategy that relies on histogram equalization, multi-task learning and domain

Table 6 Average performance metrics achieved by our method on *Lightbox* and *Sunlamp*, considering simplified Pose Estimation Models. (a) consists in a simple MLP applied on all keypoint coordinates. (b) adds to (a) a mapping of the keypoint coordinates to an embedding in a higher dimensional space. (c) processes those embeddings through an attention-based encoder before passing them to the MLP. (d) adds positional embeddings at the input of the encoder. Finally, the benefits of the continuous 6D rotation representation over the quaternion representation are evaluated.

PEM Arch.	Rotation	Lightbox			Sunlamp		
	format	$E_T[m]$	$E_R[^\circ]$	$S_P^*[/]$	$E_T[m]$	$E_R[^\circ]$	$S_P^*[/]$
(a): MLP	6D	0.12	4.60	0.10	0.16	7.38	0.16
(b): (a)+EMB	6D	0.11	4.55	0.10	0.16	7.32	0.16
(c): (b)+ENC	6D	0.10	4.37	0.09	0.14	7.08	0.15
(d): (c)+PE	6D	**0.09**	**4.32**	**0.09**	**0.14**	**6.94**	**0.14**
(d)	4D	0.11	7.65	0.15	0.14	10.10	0.20

randomization *i.e.*, aggressive data augmentations. Unlike previous works that followed a domain adaptation strategy, *i.e.*, aimed at reducing the gap by exploiting the target domain during the network training, our domain generalization strategy enables the training of our network using only synthetic images, and without increasing the inference computational cost. We successfully validated our method on the widespread SPEED+ dataset. The roles of the different components of our approach were evaluated through extensive ablation studies.

Funding Sources

The research was funded by Aerospacelab and the Walloon Region through the Win4Doc program. C. De Vleeschouwer is a Research Director of the Fonds de la Recherche Scientifique – FNRS.

Acknowledgments

Computational resources have been provided by the Consortium des Équipements de Calcul Intensif (CÉCI), funded by the Fonds de la Recherche Scientifique de Belgique (F.R.S.-FNRS) under Grant No. 2.5020.11 and by the Walloon Region.

References

[1] Forshaw, J. L., Aglietti, G. S., Navarathinam, N., Kadhem, H., Salmon, T., Pisseloup, A., Joffre, E., Chabot, T., Retat, I., Axthelm, R., Barraclough, S., Ratcliffe, A., Bernal, C., Chaumette, F., Pollini, A., and Steyn, W. H., "RemoveDEBRIS: An in-orbit active debris removal demonstration mission," *Acta Astronautica*, Vol. 127, 2016, pp. 448–463. https://doi.org/10.1016/j.actaastro.2016.06.018.

[2] Reed, B. B., Smith, R. C., Naasz, B. J., Pellegrino, J. F., and Bacon, C. E., "The restore-L servicing mission," *AIAA space 2016*, 2016, p. 5478. https://doi.org/10.2514/6.2016-5478.

[3] Biesbroek, R., Aziz, S., Wolahan, A., Cipolla, S.-f., Richard-Noca, M., and Piguet, L., "The clearspace-1 mission: ESA and clearspace team up to remove debris," *Proceedings of the 8th Conference on Space Debris*, 2021, pp. 1–3.

[4] Henshaw, C. G., "The darpa phoenix spacecraft servicing program: Overview and plans for risk reduction," *International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS)*, European Space Agency, 2014.

[5] Pyrak, M., and Anderson, J., "Performance of Northrop Grumman's Mission Extension Vehicle (MEV) RPO imagers at GEO," *Autonomous Systems: Sensors, Processing and Security for Ground, Air, Sea and Space Vehicles and Infrastructure 2022*, Vol. 12115, SPIE, 2022, pp. 64–82. https://doi.org/10.1117/12.2631524.

[6] Ventura, J., "Autonomous proximity operations for noncooperative space targets," Ph.D. thesis, Technische Universität München, 2016.

[7] Opromolla, R., Fasano, G., Rufino, G., and Grassi, M., "A review of cooperative and uncooperative spacecraft pose determination techniques for close-proximity operations," *Progress in Aerospace Sciences*, Vol. 93, 2017, pp. 53–72. https://doi.org/10.1016/j.paerosci.2017.07.001.

[8] Opromolla, R., Fasano, G., Rufino, G., and Grassi, M., "Pose estimation for spacecraft relative navigation using model-based algorithms," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 53, No. 1, 2017, pp. 431–447. https://doi.org/10.1109/TAES.2017.2650785.

[9] Christian, J. A., and Cryan, S., "A survey of LIDAR technology and its use in spacecraft relative navigation," *AIAA Guidance, Navigation, and Control (GNC) Conference*, 2013, p. 4641. https://doi.org/10.2514/6.2013-4641.

[10] Shi, J.-F., Ulrich, S., Ruel, S., and Anctil, M., "Uncooperative spacecraft pose estimation using an infrared camera during proximity operations," *AIAA SPACE 2015 conference and exposition*, 2015, p. 4429. https://doi.org/10.2514/6.2015-4429.

[11] Rondao, D., Aouf, N., and Richardson, M. A., "ChiNet: Deep recurrent convolutional learning for multimodal spacecraft pose estimation," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 59, No. 2, 2022, pp. 937–949. https://doi.org/10.1109/TAES.2022.3193085.

[12] Martínez, H. G., Giorgi, G., and Eissfeller, B., "Pose estimation and tracking of non-cooperative rocket bodies using time-of-flight cameras," *Acta Astronautica*, Vol. 139, 2017, pp. 165–175. https://doi.org/10.1016/j.actaastro.2017.07.002.

[13] Pesce, V., Lavagna, M., and Bevilacqua, R., "Stereovision-based pose and inertia estimation of unknown and uncooperative space objects," *Advances in Space Research*, Vol. 59, No. 1, 2017, pp. 236–251. https://doi.org/10.1016/j.asr.2016.10.002.

[14] Cosmas, K., and Kenichi, A., "Utilization of FPGA for onboard inference of landmark localization in CNN-Based spacecraft pose estimation," *Aerospace*, Vol. 7, No. 11, 2020, p. 159. https://doi.org/10.3390/aerospace7110159.

[15] Leon, V., Lentaris, G., Soudris, D., Vellas, S., and Bernou, M., "Towards Employing FPGA and ASIP Acceleration to Enable Onboard AI/ML in Space Applications," *2022 IFIP/IEEE 30th International Conference on Very Large Scale Integration (VLSI-SoC)*, IEEE, 2022, pp. 1–4. https://doi.org/10.1109/VLSI-SoC54400.2022.9939566.

[16] Chen, B., Cao, J., Parra, A., and Chin, T.-J., "Satellite pose estimation with deep landmark regression and nonlinear pose refinement," *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019, pp. 0–0. https://doi.org/10.1109/ICCVW.2019.00343.

[17] Sharma, S., and D'Amico, S., "Pose estimation for non-cooperative rendezvous using neural networks," *arXiv preprint arXiv:1906.09868*, 2019. https://doi.org/10.48550/arXiv.1906.09868.

[18] Park, T. H., Sharma, S., and D'Amico, S., "Towards robust learning-based pose estimation of noncooperative spacecraft," *arXiv preprint arXiv:1909.00392*, 2019. https://doi.org/10.48550/arXiv.1909.00392.

[19] Park, T. H., Märtens, M., Jawaid, M., Wang, Z., Chen, B., Chin, T.-J., Izzo, D., and D'Amico, S., "Satellite Pose Estimation Competition 2021: Results and analyses," *Acta Astronautica*, 2023. https://doi.org/10.1016/j.actaastro.2023.01.002.

[20] Pérez-Villar, J. I. B., García-Martín, Á., and Bescós, J., "Spacecraft Pose Estimation Based on Unsupervised Domain Adaptation and on a 3D-Guided Loss Combination," *European Conference on Computer Vision*, Springer, 2022, pp. 37–52. https://doi.org/10.1007/978-3-031-25056-9_3.

[21] Wang, Z., Chen, M., Guo, Y., Li, Z., and Yu, Q., "Bridging the Domain Gap in Satellite Pose Estimation: a Self-Training Approach based on Geometrical Constraints," *IEEE Transactions on Aerospace and Electronic Systems*, 2023. https://doi.org/10.1109/TAES.2023.3250385.

[22] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., "Generative adversarial nets," *Advances in neural information processing systems*, Vol. 27, 2014.

[23] Sharma, S., Ventura, J., and D'Amico, S., "Robust model-based monocular pose initialization for noncooperative spacecraft rendezvous," *Journal of Spacecraft and Rockets*, Vol. 55, No. 6, 2018, pp. 1414–1429. https://doi.org/10.2514/1.A34124.

[24] Kisantal, M., Sharma, S., Park, T. H., Izzo, D., Märtens, M., and D'Amico, S., "Satellite pose estimation challenge: Dataset, competition design, and results," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 56, No. 5, 2020, pp. 4083–4098. https://doi.org/10.1109/TAES.2020.2989063.

[25] Sun, K., Xiao, B., Liu, D., and Wang, J., "Deep high-resolution representation learning for human pose estimation," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 5693–5703. https://doi.org/10.1109/CVPR.2019.00584.

[26] Marquardt, D. W., "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the society for Industrial and Applied Mathematics*, Vol. 11, No. 2, 1963, pp. 431–441. https://doi.org/10.1137/0111030.

[27] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C., "Mobilenetv2: Inverted residuals and linear bottlenecks," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520. https://doi.org/10.1109/CVPR.2018.00474.

[28] Gerard, K., "Segmentation-driven satellite pose estimation," *Kelvins Day Presentation, URL: https://indico. esa.int/event/319/attachments/3561/4754/pose_gerard_ segmentation.pdf*, Vol. 9, 2019.

[29] Redmon, J., and Farhadi, A., "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018. https://doi.org/10.48550/arXiv.1804.02767.

[30] Lepetit, V., Moreno-Noguer, F., and Fua, P., "EPnP: An accurate O(n) solution to the PnP problem," *International journal of computer vision*, Vol. 81, 2009, pp. 155–166. https://doi.org/10.1007/s11263-008-0152-6.

[31] Fischler, M. A., and Bolles, R. C., "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, Vol. 24, No. 6, 1981, pp. 381–395. https://doi.org/10.1145/358669.35869.

[32] Hu, Y., Speierer, S., Jakob, W., Fua, P., and Salzmann, M., "Wide-depth-range 6d object pose estimation in space," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 15870–15879. https://doi.org/10.1109/CVPR46437.2021.01561.

[33] Proença, P. F., and Gao, Y., "Deep learning for spacecraft pose estimation from photorealistic rendering," *2020 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2020, pp. 6007–6013. https://doi.org/10.1109/ICRA40945.2020.9197244.

[34] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.

[35] Sonawani, S., Alimo, R., Detry, R., Jeong, D., Hess, A., and Amor, H. B., "Assistive relative pose estimation for on-orbit assembly using convolutional neural networks," *arXiv preprint arXiv:2001.10673*, 2020. https://doi.org/10.48550/arXiv.2001.10673.

[36] Legrand, A., Detry, R., and De Vleeschouwer, C., "End-to-end Neural Estimation of Spacecraft Pose with Intermediate Detection of Keypoints," *Computer Vision – ECCV 2022 Workshops*, 2023, pp. 154–169. https://doi.org/10.1007/978-3-031-25056-9_11.

[37] Black, K., Shankar, S., Fonseka, D., Deutsch, J., Dhir, A., and Akella, M. R., "Real-time, flight-ready, non-cooperative spacecraft pose estimation using monocular imagery," *arXiv preprint arXiv:2101.09553*, 2021. https://doi.org/10.48550/arXiv.2101.09553.

[38] Cassinis, L. P., Menicucci, A., Gill, E., Ahrns, I., and Sanchez-Gestido, M., "On-ground validation of a CNN-based monocular pose estimation system for uncooperative spacecraft: Bridging domain shift in rendezvous scenarios," *Acta Astronautica*, Vol. 196, 2022, pp. 123–138. https://doi.org/10.1016/j.actaastro.2022.04.002.

[39] Posso, J., Bois, G., and Savaria, Y., "Mobile-ursonet: an embeddable neural network for onboard spacecraft pose estimation," *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*, IEEE, 2022, pp. 794–798. https://doi.org/10.1109/ISCAS48785.2022.9937721.

[40] Carcagnì, P., Leo, M., Spagnolo, P., Mazzeo, P. L., and Distante, C., "A Lightweight Model for Satellite Pose Estimation," *International Conference on Image Analysis and Processing*, Springer, 2022, pp. 3–14. https://doi.org/10.1007/978-3-031-06427-2_1.

[41] Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A., "Unpaired image-to-image translation using cycle-consistent adversarial networks," *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[42] Legrand, A., Detry, R., and De Vleeschouwer, C., "Leveraging Neural Radiance Fields for Pose Estimation of an Unknown Space Object during Proximity Operations," *arXiv preprint arXiv:2405.12728*, 2024. https://doi.org/10.48550/arXiv.2405.12728.

[43] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., and Ng, R., "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, Vol. 65, No. 1, 2021, pp. 99–106. https://doi.org/10.1145/3503250.

[44] Park, T. H., and D'Amico, S., "Robust multi-task learning and online refinement for spacecraft pose estimation across domain gap," *Advances in Space Research*, 2023. https://doi.org/10.1016/j.asr.2023.03.036.

[45] Park, T. H., and D'Amico, S., "Bridging Domain Gap for Flight-Ready Spaceborne Vision," *arXiv preprint arXiv:2409.11661*, 2024. https://doi.org/10.48550/arXiv.2409.11661.

[46] Hendrycks, D., Basart, S., Mu, N., Kadavath, S., Wang, F., Dorundo, E., Desai, R., Zhu, T., Parajuli, S., Guo, M., Song, D., Steinhardt, J., and Gilmer, J., "The many faces of robustness: A critical analysis of out-of-distribution generalization," *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 8340–8349. https://doi.org/10.1109/ICCV48922.2021.00823.

[47] Xu, Z., Liu, D., Yang, J., Raffel, C., and Niethammer, M., "Robust and generalizable visual representation learning via random convolutions," *Proceedings of the International Conference on Learning Representations*, 2021. https://doi.org/10.48550/arXiv.2007.13003.

[48] Xu, Y., Zhang, J., Zhang, Q., and Tao, D., "Vitpose: Simple vision transformer baselines for human pose estimation," *Advances in neural information processing systems*, Vol. 35, 2022, pp. 38571–38584.

[49] Ulmer, M., Durner, M., Sundermeyer, M., Stoiber, M., and Triebel, R., "6d object pose estimation from approximate 3d models for orbital robotics," *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2023, pp. 10749–10756. https://doi.org/10.1109/IROS55552.2023.10341511.

[50] Cassinis, L. P., Park, T. H., Stacey, N., D'Amico, S., Menicucci, A., Gill, E., Ahrns, I., and Sanchez-Gestido, M., "Leveraging neural network uncertainty in adaptive unscented Kalman Filter for spacecraft pose estimation," *Advances in Space Research*, Vol. 71, No. 12, 2023, pp. 5061–5082. https://doi.org/10.1016/j.asr.2023.02.021.

[51] Tan, M., Pang, R., and Le, Q. V., "Efficientdet: Scalable and efficient object detection," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10781–10790. https://doi.org/10.1109/CVPR42600.2020.01079.

[52] Nibali, A., He, Z., Morgan, S., and Prendergast, L., "Numerical coordinate regression with convolutional neural networks," *arXiv preprint arXiv:1801.07372*, 2018. https://doi.org/10.48550/arXiv.1801.07372.

[53] Zhou, Y., Barnes, C., Lu, J., Yang, J., and Li, H., "On the continuity of rotation representations in neural networks," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5745–5753. https://doi.org/10.1109/CVPR.2019.00589.

[54] Good, I. J., "Rational decisions," *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 14, No. 1, 1952, pp. 107–114. https://doi.org/10.1111/j.2517-6161.1952.tb00104.x.

[55] Park, T. H., Märtens, M., Lecuyer, G., Izzo, D., and D'Amico, S., "SPEED+: Next-generation dataset for spacecraft pose estimation across domain gap," *2022 IEEE Aerospace Conference (AERO)*, IEEE, 2022, pp. 1–15. https://doi.org/10.1109/AERO53065.2022.9843439.

[56] Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., and Abbeel, P., "Domain randomization for transferring deep neural networks from simulation to the real world," *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, IEEE, 2017, pp. 23–30. https://doi.org/10.1109/IROS.2017.8202133.

[57] Singh, K. K., Yu, H., Sarmasi, A., Pradeep, G., and Lee, Y. J., "Hide-and-seek: A data augmentation technique for weakly-supervised localization and beyond," *arXiv preprint arXiv:1811.02545*, 2018. https://doi.org/10.48550/arXiv.1811.02545.

[58] Sakkos, D., Shum, H. P., and Ho, E. S., "Illumination-based data augmentation for robust background subtraction," *2019 13th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, IEEE, 2019, pp. 1–8. https://doi.org/10.1109/SKIMA47702.2019.8982527.

[59] Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., and Brendel, W., "ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness," *arXiv preprint arXiv:1811.12231*, 2018. https://doi.org/10.48550/arXiv.1811.12231.

[60] Jackson, P. T., Atapour-Abarghouei, A., Bonner, S., Breckon, T. P., and Obara, B., "Style Augmentation: Data Augmentation via Style Randomization," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 83–92. https://doi.org/10.48550/arXiv.1809.05375.

[61] Xu, Q., Zhang, R., Zhang, Y., Wang, Y., and Tian, Q., "A fourier-based framework for domain generalization," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 14383–14392. https://doi.org/10.1109/CVPR46437.2021.01415.

[62] Xu, Q., Zhang, R., Fan, Z., Wang, Y., Wu, Y.-Y., and Zhang, Y., "Fourier-based augmentation with applications to domain generalization," *Pattern Recognition*, Vol. 139, 2023, p. 109474. https://doi.org/10.1016/j.patcog.2023.109474.

[63] Cooley, J. W., and Tukey, J. W., "An algorithm for the machine calculation of complex Fourier series," *Mathematics of computation*, Vol. 19, No. 90, 1965, pp. 297–301. https://doi.org/10.2307/2003354.

[64] Cubuk, E. D., Zoph, B., Shlens, J., and Le, Q. V., "Randaugment: Practical automated data augmentation with a reduced search space," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 2020, pp. 702–703. https://doi.org/10.1109/CVPRW50498.2020.00359.

[65] Gill, E., D'Amico, S., and Montenbruck, O., "Autonomous formation flying for the PRISMA mission," *Journal of Spacecraft and Rockets*, Vol. 44, No. 3, 2007, pp. 671–681. https://doi.org/10.2514/1.23015.

[66] Park, T. H., Bosse, J., and D'Amico, S., "Robotic testbed for rendezvous and optical navigation: Multi-source calibration and machine learning use cases," *arXiv preprint arXiv:2108.05529*, 2021. https://doi.org/10.48550/arXiv.2108.05529.

[67] Kingma, D. P., and Ba, J., "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014. https://doi.org/10.48550/arXiv.1412.6980.

[68] Loshchilov, I., and Hutter, F., "Sgdr: Stochastic gradient descent with warm restarts," *arXiv preprint arXiv:1608.03983*, 2016. https://doi.org/10.48550/arXiv.1608.03983.

[69] Yu, C., Xiao, B., Gao, C., Yuan, L., Zhang, L., Sang, N., and Wang, J., "Lite-hrnet: A lightweight high-resolution network," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 10440–10450. https://doi.org/10.1109/CVPR46437.2021.01030.

[70] Cassinis, L. P., "Monocular Vision-Based Pose Estimation of Uncooperative Spacecraft," Ph.D. thesis, TU Delft, 2022.