
ForestPersons: A Large-Scale Dataset for Under-Canopy Missing Person Detection

Deokyun Kim^{1*} Jeongjun Lee^{2*} Jungwon Choi^{2*} Jonggeon Park^{2*}
Giyoung Lee¹ Yookyung Kim¹ Myungseok Ki¹ Juho Lee² Jihun Cha¹

¹ Autonomous UAV Research Section, ETRI

² Kim Jaechul Graduate School of AI, KAIST

{deokyunkim, giyoung, yk.kim, serdong, jihun}@etri.re.kr
{lee.jeongjun, jungwon.choi, parkjonggeon, juholee}@kaist.ac.kr

Abstract

Detecting missing persons in forest environments remains a challenge, as dense canopy cover often conceals individuals from detection in top-down or oblique aerial imagery typically captured by Unmanned Aerial Vehicles (UAVs). While UAVs are effective for covering large, inaccessible areas, their aerial perspectives often miss critical visual cues beneath the forest canopy. This limitation underscores the need for under-canopy perspectives better suited for detecting missing persons in such environments. To address this gap, we introduce ForestPersons, a novel large-scale dataset specifically designed for under-canopy person detection. ForestPersons contains 96,482 images and 204,078 annotations collected under diverse environmental and temporal conditions. Each annotation includes a bounding box, pose, and visibility label for occlusion-aware analysis. ForestPersons provides ground-level and low-altitude perspectives that closely reflect the visual conditions encountered by Micro Aerial Vehicles (MAVs) during forest Search and Rescue (SAR) missions. Our baseline evaluations reveal that standard object detection models, trained on prior large-scale object detection datasets or SAR-oriented datasets, show limited performance on ForestPersons. This indicates that prior benchmarks are not well aligned with the challenges of missing person detection under the forest canopy. We offer this benchmark to support advanced person detection capabilities in real-world SAR scenarios. The dataset is publicly available at <https://huggingface.co/datasets/etri/ForestPersons>.

1 Introduction

Unmanned Aerial Vehicles (UAVs) have been widely used in Search and Rescue (SAR) missions because they can quickly cover large open areas. While early UAVs relied on manual operation, advances in navigation, path planning, and flight control technologies have enabled fully autonomous missions. Furthermore, hardware miniaturization has led to the development of Micro Aerial Vehicles (MAVs), and improvements in Simultaneous Localization and Mapping (SLAM) technologies have made GPS-denied navigation possible [1, 2]. These developments have extended UAV operations to challenging forest environments with dense and scattered obstacles. Recent studies have demonstrated that UAVs can perform safe navigation [3, 4], rapid path planning for exploration [5, 6, 7], and mapping tasks [8, 9]. Despite the growing feasibility of deploying MAVs in forested environments, detecting missing persons under dense canopies remains a fundamental challenge. Forests are environments where people are not typically present, and the abundance of vegetation causes significant

*Equal contribution

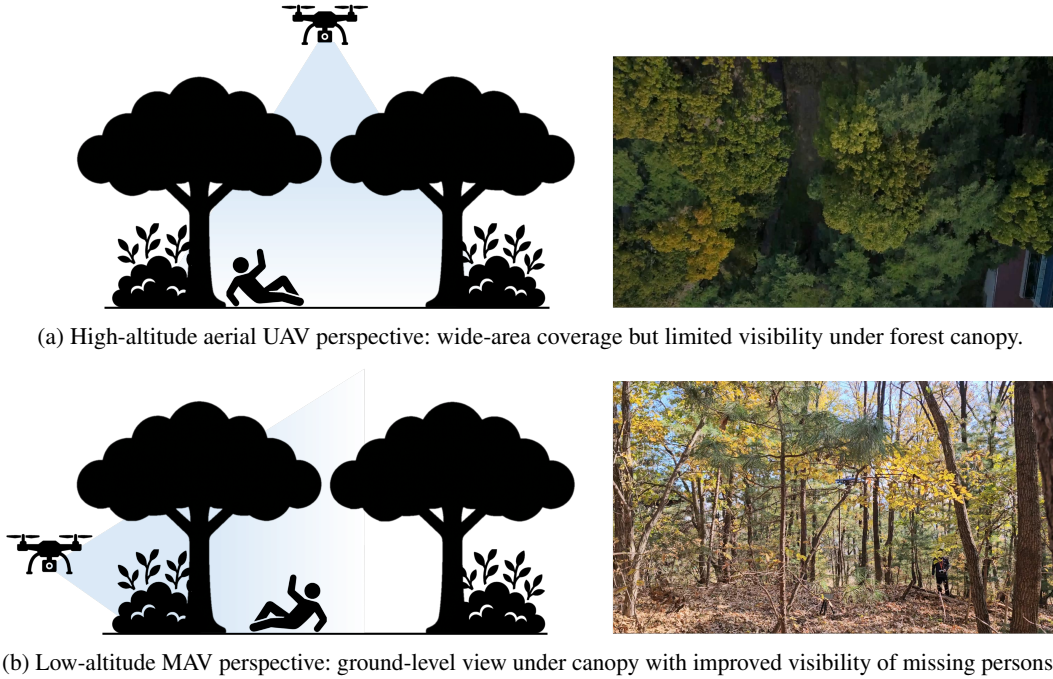


Figure 1: **Comparison of two UAV-based person search scenarios.** (a) High-altitude views offer wide-area coverage but often fail to detect targets due to canopy occlusion. (b) Low-altitude MAVs provide closer, ground-level views beneath the canopy, improving the chances of spotting missing persons despite vegetation occlusion.

and often unpredictable occlusions. Moreover, there is a lack of dedicated datasets targeting such under-canopy scenarios, limiting the ability of detection models to learn and generalize to these challenging conditions.

While several UAV-based datasets [10, 11, 12, 13] have been introduced to support SAR applications, most prior benchmarks are collected from high altitudes, typically using top-down or oblique perspectives. Although such aerial viewpoints provide broad coverage and are effective for detecting objects in open areas, they are less suitable for locating missing persons concealed beneath dense forest canopy. At high altitudes, individuals often appear as only a few pixels in the image. Dense foliage and uneven terrain further obstruct visibility, making reliable detection extremely challenging. Moreover, occlusions caused by vegetation are pervasive and vary unpredictably across different forest structures, exacerbating the difficulty of identifying partially visible or collapsed individuals.

To address this challenge, we introduce **ForestPersons**, a large-scale dataset specifically designed to support the training of models for detecting missing persons under forest canopies, where dense vegetation often causes severe occlusion and obstructs the visibility of human bodies. The dataset consists of 96,482 images and 204,078 annotated instances, collected across varying seasonal, weather, and lighting conditions, reflecting real-world under-canopy scenarios. Each person instance is annotated with bounding boxes and additional attributes including pose and visibility, which are particularly relevant to SAR applications. To the best of our knowledge, ForestPersons is the first benchmark explicitly designed for detecting persons under forest canopies, providing a foundation for developing and evaluating models in realistic SAR scenarios, and is expected to improve the likelihood of successful rescue of missing persons in real-world SAR missions.

2 Related work

2.1 UAV-Based Person Detection Datasets

Most prior UAV-based datasets capture people from top-down or oblique perspective at high altitudes as illustrated in Figure 1a. Over the past several years, large-scale datasets [14, 15, 16, 17, 18, 19]

Table 1: **ForestPersons vs. Others.** Comparison of ForestPersons with existing UAV-based datasets containing person class annotations.

Dataset	Scenario	Configuration	View Point	Data Volume		Attributes	
		Environments		#Images	#Annotations	Occlusion	Pose
HERIDAL [10]	SAR	Forest	Top-down	1,600	3,194	×	×
WiSARD [11]	SAR	Forest, Maritime	Oblique	44,588	74,204	×	×
SARD [12]	SAR	Forest	Oblique	1,981	6,532	×	✓
VTsAR [13]	SAR	Urban, Maritime, Forest	Top-down	12,465	19,956	×	×
VisDrone [14]	Surveillance	Urban	Oblique	10,209	147,747	✓	×
NII-CU [15]	Detection	Urban	Oblique	5,880	18,736	✓	×
Okutama-Action [16]	Detection	Urban	Oblique	77,365	524,649	×	✓
ForestPersons	SAR	Forest	Ground-level	96,482	204,078	✓	✓

containing high-resolution aerial imagery have been developed to support computer vision tasks such as object detection, tracking, and person recognition from aerial perspectives. Among these, VisDrone dataset [14] stands out as a comprehensive resource for drone-based computer vision applications, offering data captured using various drone-mounted cameras across diverse urban and country environments, locations, object types, and scene densities. Other notable general-purpose aerial datasets include NII-CU [15], which contains well-aligned RGB and thermal images with occlusion labels, and Okutama-Action [16], which provides aerial video for human action detection with bounding boxes and 12 action classes such as standing, sitting, and lying.

Several datasets have been proposed for various SAR applications. The HERIDAL [10] provides high-resolution imagery from mountainous regions, while the WiSARD [11] offers synchronized RGB and thermal data across diverse terrains and weather conditions. The SARD [12] and the recently proposed VTsAR [13] extend multimodal capabilities by incorporating real and synthetic RGB-thermal image pairs. Most UAV-based SAR datasets, however, are collected from high altitudes and primarily offer top-down or oblique viewpoints. While such perspectives are advantageous for efficiently covering wide areas, they are less effective in real SAR scenarios where missing persons are often located beneath dense foliage. In these environments, visibility is severely limited and occlusions caused by vegetation are frequent. As a result, This significantly reduces the chances of successfully detecting missing persons in aerial imagery. Table 1 summarizes the key attributes of representative UAV-based detection datasets.

2.2 Ground-Level Person Detection Datasets

As illustrated in Figure 1b, MAVs typically operate at low altitudes close to ground-level view. Given the similarity in viewpoints, ground-level person detection datasets are suitable training resources for under-canopy missing person detection models. Representative prior works include COCO [20], CrowdHuman [21], CityPersons [22], KITTI [23], and JRDB [24], which are widely used as benchmarks for developing and evaluating person detection models. These datasets provide high-resolution images captured in everyday environments, including annotations for bounding boxes, body joints, and occlusion states. They have supported the development of person detection models that are robust to partial occlusion and variations in human pose.

However, existing datasets primarily depict standing or walking individuals in typical indoor and outdoor environments where people are commonly found. These conditions differ substantially from those encountered in SAR missions conducted in forested environments. In real SAR scenarios, missing persons are often partially occluded by vegetation, sitting or lying beneath canopy cover, and subject to highly variable lighting and visibility conditions. Such characteristics are rarely captured in prior benchmarks, making existing datasets less suitable for training missing person detection models intended for under-canopy search operations.

3 ForestPersons

ForestPersons is a large-scale image dataset specifically developed for missing person detection in under-canopy forest environments, a key task in autonomous SAR missions. The dataset captures conditions that are common in under-canopy forest searches, where people may be partially or fully

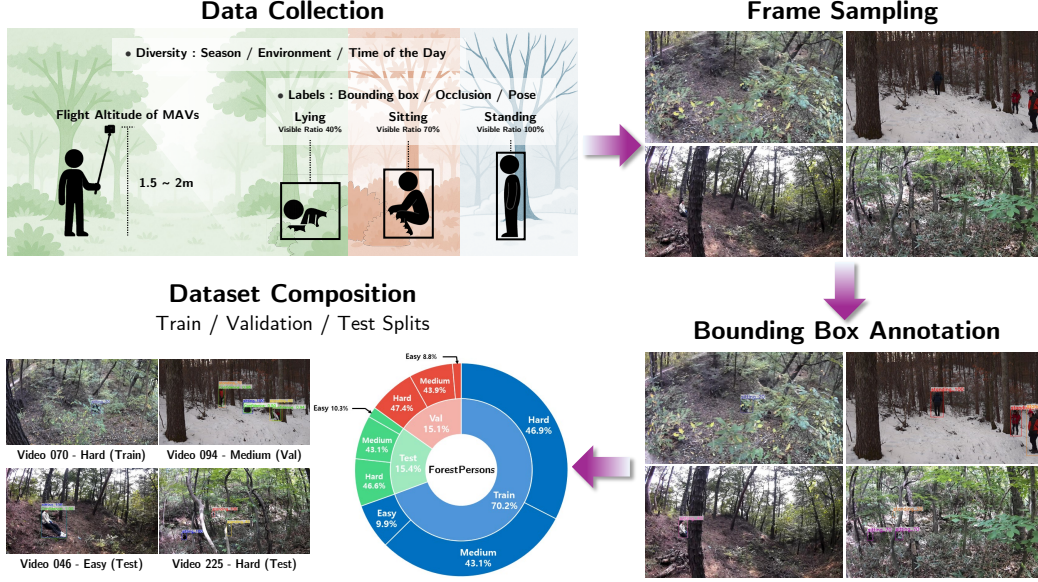


Figure 2: **Overview of ForestPersons composition pipeline.** The full process from data collection in forest environments to frame sampling from video sequences, bounding box annotation of missing persons, and difficulty-aware dataset splitting.

hidden by vegetation and can appear in various poses such as lying down, sitting, or standing. Unlike conventional person detection datasets that focus on images collected in places where people are typically found, ForestPersons targets under-canopy forest scenes, where dense foliage, seasonal shifts, and weather variability significantly impact visibility and scene appearance.

3.1 Data Collection and Frame Sampling

The ForestPersons dataset was constructed to simulate realistic SAR scenarios occurring under forest canopy conditions. As shown in Fig 2, videos were collected across diverse forest environments by simulating missing person situations that reflect plausible outcomes of fatigue or disorientation. Individuals were positioned in different postures such as lying on the ground, sitting, or standing. In these settings, they were naturally partially occluded by vegetation, branches, or uneven terrain. To emulate the viewpoints typically encountered by MAVs during under-canopy missions, handheld or tripod-mounted cameras were positioned at altitudes between 1.5 and 2.0 meters, approximating the expected flight height of MAVs.

The videos include scenes from different seasons, such as dense summer foliage that increases occlusion and winter settings with leafless trees and snow-covered terrain. Variations in weather, including clear skies, overcast conditions, and light rain, were incorporated to introduce changes in visibility and lighting. Temporal diversity was also considered by capturing footage at different times of day, primarily in the afternoon and at dusk. We deliberately included seasonal and temporal conditions in the videos to support the development of detection models that are robust to real-world SAR scenarios. Frames were extracted from the 377 video sequences collected as described above.

3.2 Annotation

Bounding boxes were annotated using the open-source COCO Annotator tool [25], following shared guidelines that required labeling only the visible portions of each individual. Given the dense vegetation and complex terrain characteristic of under-canopy environments, annotators were instructed to carefully delineate the visible contours of partially occluded individuals to ensure precise and consistent annotations.

In addition to bounding boxes, each person instance was annotated with two semantic attributes, pose and visibility level, to capture information relevant to practical SAR operations. The pose attribute provides cues about the physical state of an individual, while visibility level quantifies the degree of



Figure 3: **Visual samples from ForestPersons.** Images depicting individuals in diverse poses, occlusion levels, seasons, and forest environments.

visual difficulty caused by environmental occlusions. These interpretable categories are designed to reflect the visual conditions commonly encountered in real-world forest search scenarios.

Poses were categorized into three classes: standing, sitting, and lying. In cases where the posture of a person was ambiguous due to occlusion or background clutter, annotators referred to adjacent video frames to make informed decisions, based on shared annotation guidelines. Visibility levels were categorized into four levels based on the degree of occlusion caused by vegetation or terrain: a value of 20 indicates heavy occlusion where the individual is almost unrecognizable, 40 corresponds to partial occlusion with the person still identifiable, 70 denotes minor occlusion with most of the body clearly visible, and 100 represents full visibility without any occlusion. Representative examples of each visibility level and pose category under realistic forest conditions are presented in Figure 3.

Following the annotation of bounding box and semantic attributes, an automated and manual anonymization protocol was applied to remove personally identifiable facial information. Specifically, a face detector [26] was used to identify facial regions in all images, which were then blurred accordingly. Subsequently, a manual review was conducted to identify any remaining visible faces, and additional blurring was applied as needed to ensure complete anonymization.

3.3 Dataset Split and Statistics

With the data collection and annotation processes described above, ForestPersons comprises 96,482 images and 204,078 annotated person instances, each instance labeled with a bounding box, pose,

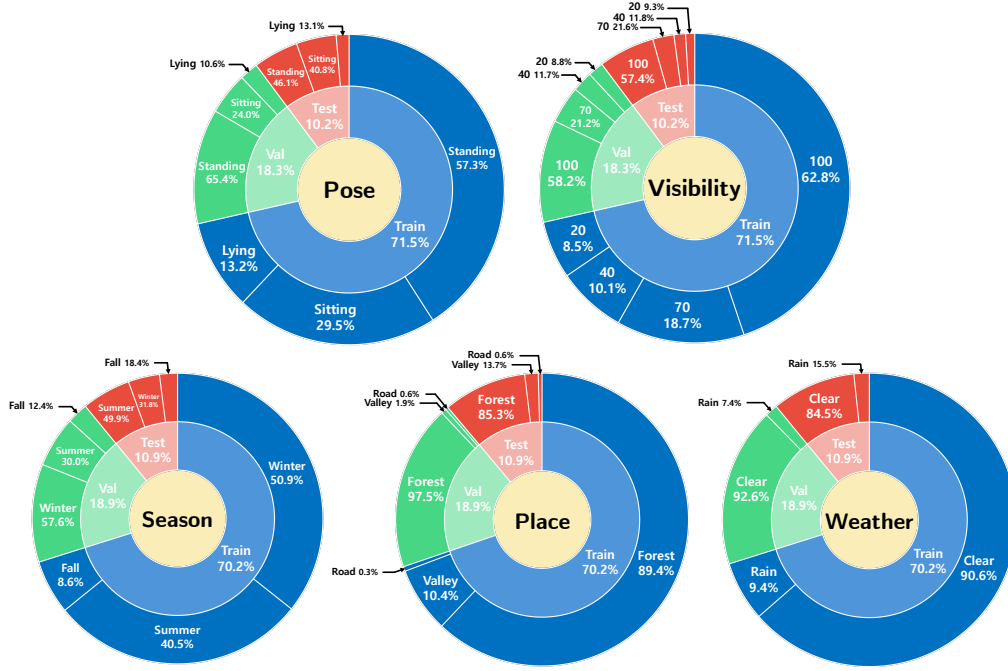


Figure 4: **Annotation statistics of ForestPersons.** Instance-level distribution for pose and visibility (Top) and image-level distribution for season, place, and weather (Bottom).



Figure 5: **ForestPersons samples by difficulty level.** Shown are representative video sequences from the easy, medium, and hard groups. Predicted boxes are shown with confidence scores, and ground-truth boxes are labeled as $\{pose\}_{[visibility\ level]}$.

and visibility level. To reduce annotator bias and mitigate the effects of human error, we designed a model-driven difficulty-aware dataset splitting strategy. In particular, to prevent overlap between temporally adjacent frames and to account for task difficulty, we split the dataset at the video sequence level. Each sequence was assigned a difficulty score based on the detection performance of a COCO-pretrained Faster R-CNN [27] implemented in Detectron2 [28], computed as $1 - AP_{50}$. Sequences were then grouped such that easy, medium, and hard samples were proportionally distributed across the training, validation, and test sets.

As shown in Figure 4, the training, validation, and test splits exhibit comparable distributions across seasons, location types, and weathers for images, as well as visibility levels and poses for the missing person instances. These distributions reflect biases introduced during the image collection process, despite efforts to cover a broad range of scenarios. To better simulate realistic SAR situations near forest entrances, a small number of videos recorded at forest edges (labeled as "Road") were also included in the dataset.

Representative examples from each difficulty group are shown in Figure 5, with one sample per row corresponding to easy ($\text{difficulty score} < 0.45$), medium ($0.45 \leq \text{score} < 0.75$), and hard ($\text{score} \geq 0.75$) levels, respectively. The final split consists of 67,686 images and 145,816 annotations for training, 18,243 images and 37,395 annotations for validation, and 10,553 images and 20,867 annotations for testing.

4 Experiments

4.1 Experiment Setting

Training object detection models. We evaluate a diverse set of widely adopted and representative object detection models. Specifically, we train models with YOLO-based [29] backbones (YOLOv3 [30] and YOLOX [31]), ResNet-50-based [32] backbones (RetinaNet [33] and Faster R-CNN [27]), a MobileNetV2-based [34] backbone (SSD [35]), and transformer-based [36] backbones (DETR [37] and DINO [38]). All models, except for DINO, are implemented using MMDetection framework [39], while DINO is implemented using detrex framework [40]. The training hyperparameters for each model are detailed in the Appendix. We conduct all experiments on NVIDIA RTX 3090 GPUs, except for DETR models, which were trained on NVIDIA A100 and A6000 GPUs.

Evaluation. We use Average Precision (AP) and Average Recall (AR) as the primary evaluation metrics. Specifically, both are computed over Intersection over Union (IoU) thresholds ranging from 0.5 to 0.95 at intervals of 0.05. We report $AP_{50:95}$ as the main metric, along with AP_{50} and AP_{75} , which correspond to IoU thresholds of 0.5 and 0.75, respectively. In SAR missions, where false negatives (i.e., missed detections of actual persons) can critically impact mission success, recall is especially important. We therefore report $AR_{50:95}$ to provide a complementary view of detection performance. Unless otherwise specified, we refer to $AP_{50:95}$ and $AR_{50:95}$ simply as AP and AR throughout the paper.

4.2 Limitations of Prior Datasets in Under-Canopy Environments

Prior SAR datasets, which are composed of aerial imagery, present challenges for detecting persons under-canopy due to the difference in viewpoint and limited visibility caused by vegetation. Meanwhile, publicly available ground-level person datasets do not adequately account for occlusions caused by dense vegetation, making them less suitable for these tasks. To demonstrate this limitation, we conduct experiments to assess the generalization capability of models trained on these prior datasets when applied to our proposed dataset. Specifically, we train object detection models using existing SAR datasets and conventional ground-level person datasets, and their performance was evaluated on the test set of our dataset.

Table 2: **Adaptation of prior datasets to under-canopy SAR tasks.** Performance comparison of Faster R-CNN [27] trained and tested on combinations of datasets: (Left) prior UAV-based SAR datasets and ForestPersons; (Right) prior ground-level person datasets and ForestPersons.

UAV-based SAR dataset					Ground-level person dataset				
Train	Test	AP	AP_{50}	AP_{75}	Train	Test	AP	AP_{50}	AP_{75}
SARD [12]	SARD	58.6	90.8	68.4	COCOPerson [20]	COCOPerson	54.0	82.5	58.2
	Ours	3.0	7.8	1.6		Ours	40.8	66.9	45.2
HERIDAL [10]	HERIDAL	35.0	70.8	29.3	CrowdHuman [21]	CrowdHuman	39.4	74.8	37.3
	Ours	0.2	0.3	0.2		Ours	31.9	58.8	31.0
WiSARD [11]	WiSARD	18.5	51.7	7.9	CityPersons [22]	CityPersons	38.7	62.5	42.1
	Ours	11.3	29.0	6.4		Ours	5.9	15.1	3.7

The results, summarized in Table 2, indicate that models trained on SAR data performed poorly on ForestPersons, and those trained on ground-level data also showed significant performance degradation due to occlusions from natural elements. Specifically, models trained on prior SAR datasets often fail to detect clearly exposed individuals, primarily due to viewpoint differences, especially the aerial perspective common in SAR data. Meanwhile, models trained on ground-level person datasets struggle with individuals who are partially occluded by vegetation or in non-standing poses such as sitting or lying. These findings highlight the limitations of relying solely on existing SAR and ground-level datasets for under-canopy SAR applications, thereby underscoring the necessity and relevance of our proposed dataset. The examples of failure cases of the object detection models trained with existing datasets are depicted in Figure 9 in the Appendix.

Table 3: **ForestPersons benchmark results.** Object detection model performance on validation and test splits of ForestPersons.

Detection Model	Validation Split				Test Split			
	AP	AP ₅₀	AP ₇₅	AR	AP	AP ₅₀	AP ₇₅	AR
YOLOv3 [30]	55.6	91.7	63.2	63.1	50.2	86.5	53.9	58.6
YOLOX [31]	56.8	92.9	65.2	62.5	51.0	89.0	54.4	58.2
RetinaNet [33]	64.1	96.0	75.8	70.4	64.2	93.9	74.4	70.9
Faster R-CNN [27]	64.2	95.6	76.5	69.6	64.4	92.7	75.4	70.0
SSD [35]	48.9	88.5	49.4	57.8	45.0	83.6	43.1	53.7
DETR [37]	55.3	93.0	59.9	68.0	53.9	88.7	59.4	67.9
DINO [38]	59.9	91.7	69.1	70.1	65.3	94.0	76.2	77.7

4.3 Dataset Benchmark Performance

We evaluated the baseline object detection models on ForestPersons and reported their benchmark performance on both the validation and test sets, as summarized in Table 3. Our results show that YOLO-based models, including YOLOv3 [30] and YOLOX [31], achieve AP scores of 50.2 and 51.0, respectively. ResNet-50-based models, including RetinaNet and Faster R-CNN [27], obtain AP scores of 64.2 and 64.4, respectively. The MobileNetV2-based model, SSD [35], achieves an AP of 45.0. Transformer-based models, including DETR [37] and DINO [38], achieve AP scores of 53.9 and 65.3, respectively.

4.4 Impact of Different Attributes on Detection Performance

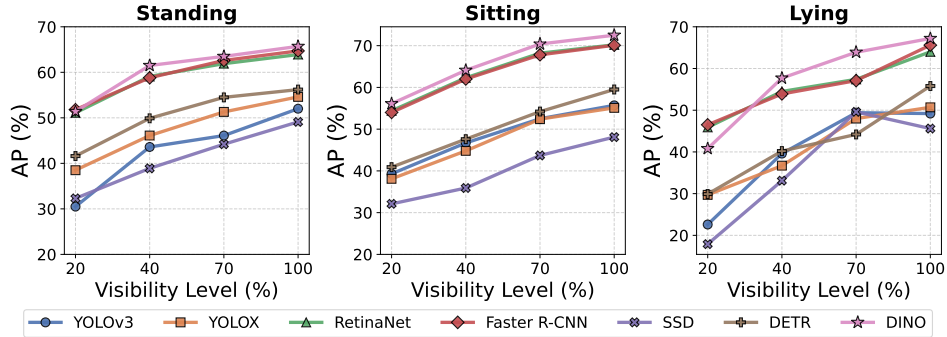


Figure 6: **Effect of visibility level on detection performance.** Detection precision improves as the visibility level increases across pose attributes.

Visibility diversity reflecting real-world SAR conditions. In under-canopy SAR tasks, it is natural that the difficulty of person detection increases as the degree of occlusion caused by surroundings becomes more severe. To simulate this challenge, ForestPersons includes human instances with varying levels of occlusion, which are carefully annotated with corresponding visibility level. Figure 6 shows that the performance of models trained on ForestPersons increases with the visibility level. The correlation between AP and visibility level empirically demonstrates the inherent difficulty of detecting heavily occluded individuals in under-canopy SAR tasks. The explicit annotation of pose and visibility level in ForestPersons enables systematic evaluation and facilitates the development of robust object detection models better suited for real-world SAR scenarios.

Effect of pose diversity on generalizability. In SAR tasks, it is important to collect data of individuals in a variety of poses since missing persons in forest environments may be found in diverse postures. However, most existing public person datasets predominantly consist of upright individuals, with standing poses comprising the vast majority. We hypothesize that this imbalance limits the generalizability of person detection models for SAR applications. To validate this hypothesis, we conduct an experiment using pose annotations in ForestPersons. Specifically, we trained object detection models using only samples labeled with standing poses and evaluated their performance on test samples categorized into standing, sitting, and lying poses, respectively.

Table 4: **Impact of various attributes on detection performance in ForestPersons.** Each object detection model was trained and evaluated using subsets of train and test data with unique attributes.

Train Attributes	(a) Pose						(b) Season								
	Standing			All Poses			Summer			Winter			All Seasons		
	Standing	Sitting	Lying	Standing	Sitting	Lying	Summer	Fall	Winter	Summer	Fall	Winter	Summer	Fall	Winter
YOLOv3 [30]	45.3	30.0	32.1	49.3	51.5	47.5	49.7	53.7	25.7	4.5	1.4	54.0	51.1	58.2	50.7
YOLOX [31]	47.3	30.3	31.7	52.2	50.6	47.9	56.8	57.1	17.2	5.5	1.5	60.0	50.0	53.6	56.5
RetinaNet [33]	57.5	47.2	43.8	62.3	66.3	60.3	63.4	66.3	43.8	14.6	4.7	63.4	66.0	73.2	63.1
Faster R-CNN [27]	58.0	47.0	42.2	63.1	66.1	61.0	65.7	66.9	34.6	18.7	11.7	61.5	65.9	71.6	64.0
SSD [35]	39.3	22.3	22.8	46.1	43.7	45.1	44.2	49.0	21.9	5.2	1.9	50.1	42.5	55.2	50.6
DETR [37]	43.2	29.4	26.2	54.1	54.3	48.4	31.9	41.9	22.0	8.4	3.3	54.8	53.2	63.8	57.1
DINO [38]	59.9	50.3	46.3	64.2	67.6	64.1	51.3	48.9	32.0	17.6	7.1	57.0	68.0	74.9	64.6

The results are presented in the Table 4 (a). Specifically, models trained solely on standing attribute exhibited significantly lower performance in detecting sitting and lying poses across all evaluated models. In contrast, models trained on the dataset with comprehensive pose annotations, achieved improved detection performance across all pose categories. These findings highlight the importance of collecting diverse human poses for SAR tasks. ForestPersons addresses this need by including underrepresented poses such as sitting and lying, which are often absent from conventional public datasets, making it more suitable for under-canopy person detection in SAR scenarios.

Effect of season diversity on generalizability. The visual appearance of forest environments can vary drastically across seasons due to changes in under-canopy vegetation density, foliage, and lighting conditions. These seasonal differences directly affect the visibility and occlusion patterns of individuals, which in turn influence detection difficulty. We presume that insufficient seasonal diversity in training data is likely to constrain the generalization capability of detection models under diverse environmental conditions. To demonstrate this, we conduct a controlled experiment using ForestPersons with explicit season labels, comparing models trained on a specific season (summer or winter) and tested on different seasons.

The results on the Table 4 (b) show a clear asymmetry in cross-season performance. Models trained on only summer images exhibited performance degradation when tested on winter images but maintained a relatively stable level of AP. In contrast, models trained solely on winter images showed a significant drop in performance when evaluated on summer and fall images. Notably, when models were trained on images from all seasons, they achieved consistent performance across all seasonal conditions. These findings highlight the importance of seasonally diverse training data for robust SAR performance, which our dataset fulfills by including images captured across different seasons.

5 Discussion and Conclusion

ForestPersons is the first large-scale dataset designed to detect missing persons in under-canopy forest environments. Unlike previous SAR benchmarks that focus on UAV-based aerial imagery, ForestPersons provides ground-level views from the perspective of MAVs, which are more suitable for detecting partially occluded individuals beneath forest canopies. The dataset includes annotations for various attributes, such as season, location type, weather, human pose, and visibility level, providing a basis for training and evaluating models under diverse and realistic SAR scenarios. We also anticipate that ForestPersons can contribute to autonomous SAR efforts using ground-based robotic platforms such as unmanned ground vehicles.

Limitations. Annotations were manually created and reviewed based on consistent guidelines, but may contain occasional noise or ambiguity due to the inherent subjectivity of visual interpretation. Frames from thermal or infrared cameras, which can provide important cues for locating missing persons in real SAR missions, are not included, as the dataset is limited to RGB imagery.

Societal and ethical considerations. To prevent misuse, only staged scenes with voluntary participants are included, and no personal information is present. The dataset will be released under a research-only license, with responsible and transparent use strongly encouraged.

References

- [1] Xu Liu, Guilherme V Nardari, Fernando Cladera, Yuezhan Tao, Alex Zhou, Thomas Donnelly, Chao Qu, Steven W Chen, Roseli AF Romero, Camillo J Taylor, et al. Large-scale autonomous flight with real-time semantic slam under dense forest canopy. *IEEE Robotics and Automation Letters*, 7(2):5512–5519, 2022.
- [2] Abraham Bachrach, Anton de Winter, Ruijie He, Garrett Hemann, Samuel Prentice, and Nicholas Roy. Range-robust autonomous navigation in gps-denied environments. In *2010 IEEE International Conference on Robotics and Automation*, Anchorage, Alaska, US, 2010. IEEE.
- [3] Sebastián Barbas Laina, Simon Boche, Sotiris Papatheodorou, Dimos Tzoumanikas, Simon Schaefer, Hanzhi Chen, and Stefan Leutenegger. Scalable autonomous drone flight in the forest with visual-inertial slam and dense submaps built without lidar. *arXiv preprint arXiv:2403.09596*, 2024.
- [4] Youkyung Hong, Suseong Kim, Youngsun Kwon, Sanghyouk Choi, and Jihun Cha. Safe and efficient exploration path planning for unmanned aerial vehicle in forest environments. *Aerospace*, 11(7):598, 2024.
- [5] Yunfan Ren, Fangcheng Zhu, Guozheng Lu, Yixi Cai, Longji Yin, Fanze Kong, Jiarong Lin, Nan Chen, and Fu Zhang. Safety-assured high-speed navigation for mavs. *Science Robotics*, 10(98):eado6187, 2025.
- [6] Laura Jarín-Lipschitz, Xu Liu, Yuezhan Tao, and Vijay Kumar. Experiments in adaptive replanning for fast autonomous flight in forests. In *2022 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2022.
- [7] Boyu Zhou, Yichen Zhang, Xinyi Chen, and Shaojie Shen. Fuel: Fast uav exploration using incremental frontier structure and hierarchical planning. *IEEE Robotics and Automation Letters*, 6(2):779–786, 2021.
- [8] Tzu-Jui Lin and Karl A. Sto. Autonomous surveying of plantation forests using multi-rotor uavs. *Drones*, 6(9):256, 2022.
- [9] Youngsun Kwon, Suseong Kim, Youkyung Hong, Sanghyouk Choi, and Jihun Cha. Online terrain mapping for exploring dense forests on unmanned aerial vehicles. In *2024 15th International Conference on Information and Communication Technology Convergence (ICTC)*, pages 1676–1680. IEEE, 2024.
- [10] Mirela Kundid Vasić and Vladan Papić. Improving the model for person detection in aerial image sequences using the displacement vector: A search and rescue scenario. *Drones*, 6(1):19, 2022.
- [11] Daniel Broyles, Christopher R Hayner, and Karen Leung. Wisard: A labeled visual and thermal image dataset for wilderness search and rescue. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 9467–9474. IEEE, 2022.
- [12] Sasa Sambolek and Marina Ivacic-Kos. Search and rescue image dataset for person detection (sard). *IEEE Dataport*, 2021.
- [13] Xiangqing Zhang, Yan Feng, Nan Wang, Guohua Lu, and Shaohui Mei. Aerial person detection for search and rescue: Survey and benchmarks. *Journal of Remote Sensing*, 5, 2025.
- [14] Pengfei Zhu, Longyin Wen, Dawei Du, Xiao Bian, Heng Fan, Qinghua Hu, and Haibin Ling. Detection and tracking meet drones challenge. *IEEE transactions on pattern analysis and machine intelligence*, 44(11):7380–7399, 2021.
- [15] Simon Speth, Artur Gonçalves, Bastien Rigault, Satoshi Suzuki, Mondher Bouazizi, Yutaka Matsuo, and Helmut Prendinger. Deep learning with rgb and thermal images onboard a drone for monitoring operations. *Journal of Field Robotics*, 39(6):840–868, 2022.

- [16] Mohammadamin Barekatin, Miquel Martí, Hsueh-Fu Shih, Samuel Murray, Kotaro Nakayama, Yutaka Matsuo, and Helmut Prendinger. Okutama-action: An aerial view video dataset for concurrent human action detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops (CVPRW)*, pages 28–35, 2017.
- [17] Dawei Du, Yuankai Qi, Hongyang Yu, Yifan Yang, Kaiwen Duan, Guorong Li, Weigang Zhang, Qingming Huang, and Qi Tian. The unmanned aerial vehicle benchmark: Object detection and tracking. In *Proceedings of the European conference on computer vision (ECCV)*, pages 370–386, 2018.
- [18] Jingru Zhu, Xiandong Wang, Yi Liu, Qianwei Ji, Zhao Zhao, and Shengke Wang. Uavtinydet: Tiny object detection in uav scenes. In *2022 7th International conference on image, vision and computing (ICIVC)*, pages 195–200. IEEE, 2022.
- [19] Yingying Liu, Fengqin Yao, Laihui Ding, Zhiwei Xu, Xiaogang Yang, and Shengke Wang. An image segmentation method based on transformer and multi-scale feature fusion for uav marine environment monitoring. In *2023 8th International Conference on Image, Vision and Computing (ICIVC)*, pages 328–336. IEEE, 2023.
- [20] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. In *European Conference on Computer Vision*, pages 740–755. Springer, 2014.
- [21] Shuai Shao, Zijian Zhao, Boxun Li, Tete Xiao, Gang Yu, Xiangyu Zhang, and Jian Sun. Crowdhuman: A benchmark for detecting human in a crowd. *arXiv preprint arXiv:1805.00123*, 2018.
- [22] Shanshan Zhang, Rodrigo Benenson, and Bernt Schiele. Citypersons: A diverse dataset for pedestrian detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, pages 3213–3221, 2017.
- [23] Andreas Geiger, Philip Lenz, and Raquel Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *2012 IEEE conference on computer vision and pattern recognition (CVPR)*, pages 3354–3361. IEEE, 2012.
- [24] Roberto Martin-Martin, Mihir Patel, Hamid Reza Tofighi, Abhijeet Sheno, JunYoung Gwak, Eric Frankel, Amir Sadeghian, and Silvio Savarese. Jrdb: A dataset and benchmark of egocentric robot visual perception of humans in built environments. *IEEE transactions on pattern analysis and machine intelligence*, 45(6):6748–6765, 2021.
- [25] Justin Brooks. Coco annotator. <https://github.com/jsbroks/coco-annotator/>, 2019.
- [26] Josep López. YOLOv8-Face: A yolov8-based face detection implementation. <https://github.com/Yusepp/YOLOv8-Face>, 2024.
- [27] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.
- [28] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. Detectron2. <https://github.com/facebookresearch/detectron2>, 2019.
- [29] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788, 2016.
- [30] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.
- [31] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430*, 2021.
- [32] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

- [33] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 2980–2988, 2017.
- [34] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
- [35] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. Ssd: Single shot multibox detector. In *Proceedings of the European conference on computer vision (ECCV)*, page 21–37, 2016.
- [36] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.
- [37] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-end object detection with transformers. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.
- [38] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2021.
- [39] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, Zheng Zhang, Dazhi Cheng, Chenchen Zhu, Tianheng Cheng, Qijie Zhao, Buyu Li, Xin Lu, Rui Zhu, Yue Wu, Jifeng Dai, Jingdong Wang, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. MMDetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*, 2019.
- [40] Tianhe Ren, Shilong Liu, Feng Li, Hao Zhang, Ailing Zeng, Jie Yang, Xingyu Liao, Ding Jia, Hongyang Li, He Cao, Jianan Wang, Zhaoyang Zeng, Xianbiao Qi, Yuhui Yuan, Jianwei Yang, and Lei Zhang. detrex: Benchmarking detection transformers, 2023.

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [\[Yes\]](#)

Justification: The paper clearly describes introduction of a large-scale dataset for under-canopy missing person detection in forest SAR scenarios. The claims are consistent with the core content, scope, and structure of the paper, particularly as described in Sec. 1 and Sec. 3.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: The dataset includes only RGB imagery and simulated missing persons, which may not fully capture the conditions and sensor modalities used in real-world SAR scenarios. See Sec. 5 for further details.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [\[Yes\]](#)

Justification: We propose ForestPersons under the assumption that prior UAV-based datasets or ground-level datasets are not suitable for under-canopy detection scenarios. The experiments presented in Sec. 4 are designed to empirically validate this assumption, demonstrating the effectiveness of the proposed dataset in addressing the identified limitations.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [\[Yes\]](#)

Justification: The paper fully discloses the information needed to reproduce the main experimental results, including detailed descriptions of the data collection protocols, annotation procedures, dataset composition, evaluation setup, dataset access, and baseline experiments. All essential details required to reproduce the main claims and conclusions are included in the paper.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

- (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The ForestPersons is publicly available under the CC BY-NC-SA 4.0 license and can be accessed via Hugging Face at <https://huggingface.co/datasets/etri/ForestPersons>.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: The paper provides complete details on both the data split strategy and the experimental setup, as described in Sec 3.3 and Sec. 4.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: we have included graphs with complete distributions of the dataset across key attributes such as season, environment type, visibility ratio, and pose.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: The paper reports baseline detection experiments using multiple detectors and provides detailed information on the compute environment, including GPU type as described in Sec. 4.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: In accordance with domestic regulations, all participants acting as missing persons provided prior informed consent and received appropriate compensation. Faces, which may constitute personally identifiable information, were blurred to protect privacy, as detailed in Sec. 3.2.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [Yes]

Justification: The ForestPersons dataset is intended to support missing person detection in SAR scenarios. To prevent misuse, it contains only staged scenes with voluntary participants and no personal information. Societal impacts are discussed in the main text and Section 5.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [Yes]

Justification: To prevent potential misuse, we applied face blurring and released the dataset under terms that restrict its use to research-only purposes.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All external resources used in this paper, such as baseline detector implementations and evaluation tools, are properly cited with credit to the original authors. Each resource was used in compliance with its respective license, and any reused components are explicitly stated in the paper or supplemental material.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. New assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: The ForestPersons dataset is available on Hugging Face under the CC BY-NC-SA 4.0 license at <https://huggingface.co/datasets/etri/ForestPersons>, with documentation covering its structure, annotation format, collection protocol, and intended use.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. Crowdsourcing and research with human subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [Yes]

Justification: Data collection and annotation were conducted by an agency contracted by our institute, with verification performed by in-house annotators. All individuals appearing in the dataset voluntarily participated in simulated missing person scenarios with prior written consent after being fully informed. Compensation was provided in accordance with the contract and local labor standards. No crowdsourcing platforms were used in this process.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.

- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. Institutional review board (IRB) approvals or equivalent for research with human subjects

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The dataset does not contain real missing persons or any sensitive personal or biometric information. All experiments were conducted on images and bounding box annotations that were processed to blur facial regions.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. Declaration of LLM usage

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The research does not involve the use of large language models (LLMs).

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.