Secure Systems Design and Technology Development. Veliko Tarnovo, Bulgaria
Proceedings of the International Scientific Conference, 2026. Volume I, 303-310

Analysis of the Human Body Positioning in Disaster Remediation

Svilen Neykov
Bulgarian Defence Institute
“prof. Tsvetan Lazarov”
Sofia, Bulgaria
svilen@nkv.me
ORCID: 0009-0005-3975-2781

Rosen Iliev
Bulgarian Defence Institute
“prof. Tsvetan Lazarov”
Sofia, Bulgaria
r.iliev@di.mod.bg

***Abstract*—The paper presents an approach to detection, classification and decision making for disaster relief scenarios, based on imaging data, pose estimations, depth perception and scene understanding, utilizing AI while performing on low power hardware.**

***Keywords*— AI, detection, classification, disaster relief, pose estimations, low power hardware.**

I. Introduction

Disaster remediation is one of those spheres that sit on the line between the military and the civilian space, as it is not a prime directive of the army to assist in these situations, but usually they are the ones with the most suitable equipment to lend a hand. As such, the armed forces of each nation should be interested in the most up-to-date ways to verify where their efforts will be the most valuable – which is where the chance to save a human life is.

Here we come to the most pressing issue – in a disaster remediation mission we can assume that there are victims in the whole region, but where would we find the ones that are most in need of assistance.

In this paper, we will address this question and propose a way to observe and evaluate where the disaster relief forces must concentrate their efforts in order to mitigate potential loss of life.

II. Materials and Methods

Both through Geo-political tensions rising and changes to the climate of the planet, disaster relief becomes more important with each passing day. One of the pillars in this is the localisation of impacted people in the area of the disaster.

With the advancement of technology this aspect can be off-loaded from ground-based search parties and aerial observers on planes/helicopters to a lot more mobile platforms, such as UAVs, UGVs and USVs.

The most perspective platform we can use for fast deployment and detection is the UAV, as aerial photography provides information for a very wide area. However, this now thrusts the responders onto the next hurdle – these platforms have a very limited carrying capacity for sensors, fuel/batteries and in general computational capabilities. This necessitates that the industry and researchers to develop more efficient algorithms that will be loaded unto these platforms.

A. Object Detection and Classification

As a visual animal, we perceive most of the world around us through sight. Our sight is sharp, but it relies on cues, movement, patterns and imperfections. However, as we are prone to boredom, tiredness and in general loss of sight, we are increasingly more reliant on the “digital eye”.

Sensors are our solution to the imperfections of the human eye and psyche. They can process information faster and can be adapted to a far greater extent to varying conditions. Thus, we have Computer Vision.

Computer Vision has a wide range of applications, from industrial applications for defect detection, sorting, maintenance, etc. through healthcare, by tumour detection, blood cell counting, etc., everyday applications such as licence plate reading for parking lot entry or highway control to the vast military domain.

One of the major computer vision aspects is Object Detection and Classification. Traditionally object detection has utilized sliding windows which select regions of the image, but it is very hard to find the optimal number of these windows. The traditional methods rely heavily on manually created features, which struggle with accuracy and robustness when viewing diverse objects and possess complex backgrounds. These features also have one additional flaw -they are not universal; this translates to

Online ISSN 3044-7771
https://doi.org/10.68302/std2026.vol1.149

poor multi-domain performance and high complexity and time-consumption in region selection. With Deep Learning (DL)booming in the industry, object detection utilizing DL has vastly improved and surpassed traditional object detection methods. DL-based algorithms use very intricate and strictly hierarchical architectures; these enable the training of more complex features. Also, owing to the significant learning capabilities of convolutional neural networks (CNNs) feature can be better represented than traditional methods. This in turn raises significantly the robustness and eliminates the manually created features. [1]

B. Pose Estimation

Pose estimation is a subset of the computer vision task that is to detect human key joints from general data sets. It is an important task in computer vision, skeleton-based action recognition can be based on its results to determine the type of motions. [2]

In this context the poses that are most likely to be of interest to the researcher or the people in the field is if you can detect an anomalous position of the body or extremities, which will signal the need for intervention.

C. Scene Understanding

Scene understanding in the context of Computer Vision is the Analysis of the frame as a whole, where apart from the detection, classification and pose estimation the disaster remediation teams or researchers need to know, not only if there is a person that is in need of assistance, but what are the conditions around him, is it viable to send help, what kind of risk is involved, or in short, what kind of help is needed.

D. UxVs

The use of multiple unmanned vehicles in recent years has skyrocketed. From the war in Ukraine and the incidents in the Baltic Sea we can see that the face of any conflict has changed significantly in the use of such platforms. Disaster relief is also not a stranger to these changes; multiple platforms are being developed and introduced:

- DJI Matrice 4TD / 4T (2026 Model): Recognized as a premier choice for 2026, offering 640×512 thermal imaging, 56× optical zoom, and IP55 weather resistance, allowing it to operate in rain, snow, and strong winds.
- Autel Robotics EVO Max 4T: Specialized for SAR with dual thermal sensors, 42-minute flight endurance, and 15km transmission range, making it ideal for large, complex search areas.
- Skydio X10: Highlights advanced AI for obstacle avoidance and autonomous flight paths, minimizing operator workload in dense urban or mountainous areas.
- WingtraRay (VTOL): Combines vertical takeoff and landing with fixed-wing endurance, capable of covering up to 350 square kilometers per mission, ideal for large-area wilderness searches.
- GDU UAV-P300: Introduced at CES 2026, this FPV drone features AI-powered "fog-penetrating" optical, electronic, and AI de-fogging tech for operation in poor weather.
- DJI Dock 3 (Drone-in-a-Box): Enables automated 24/7 deployment, where a drone automatically takes off, scans a pre-programmed area, and returns to charge without human intervention, ideal for rapid initial response.[3][4]

For the trends in UxVs we can see that the most vital areas are:

- AI-Enhanced Detection: AI algorithms distinguish human heat signatures from environmental heat sources (e.g., fires) in thermal data, increasing success rates.
- Thermal/Multispectral Integration: Drones are frequently carrying combined thermal, zoom, and RGB cameras to identify victims in low-light, fog, or through light foliage.
- Automatic 3D Mapping: Drones can generate real-time 3D models of destroyed structures, allowing responders to identify safe entry points.
- Fiber-Optic Data Links: To fight electronic warfare and signal jamming, new drones are using fiber-optic cables for secure, high-definition data transmission.[5]

With all of the aspects discussed above, the combination of which can lead to a greater development in the Disaster Relief systems and strategies that all countries need to address.

III. Results and Discussion

A. Setup

Utilizing state-of-the-art algorithms, we can assess all of the points in this paper, as one of the leading directions for development of detection models is bringing them to edge applications, therefore minimizing their resource footprint. [6]

1) The Model: One such model is the evolution of the YOLO (You Only Look Once) family of models, this is a popular series of state-of-the-art, real-time object detection systems designed for high speed and accuracy. Unlike traditional two-stage detectors (like R-CNN) that first find potential object areas and then classify them, YOLO treats object detection as a single regression problem, directly predicting bounding boxes and class probabilities from full images in a single evaluation.
The used model here is the new YOLO26. The performance improvements can be see in Fig.1 and Fig.2. There we can see the improvements that the new architecture, Fig.3 brings despite the lowering of the hardware

requirements.[6][7] One of the key features is that we can also use the same architecture for all 3 major tasks ahead: Detection, Classification and Pose Estimation.

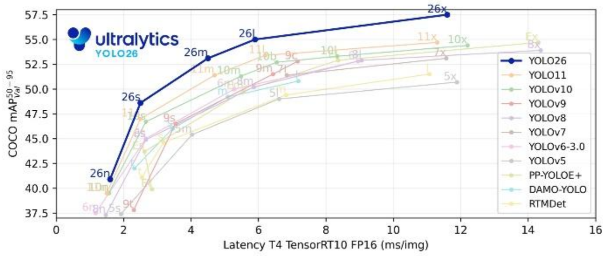

Fig. 1. Performance improvement of YOLO26. [6][7]

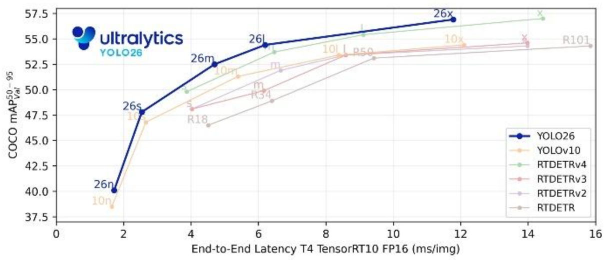

Fig. 2. E2E Performance improvement of YOLO26. [6][7]

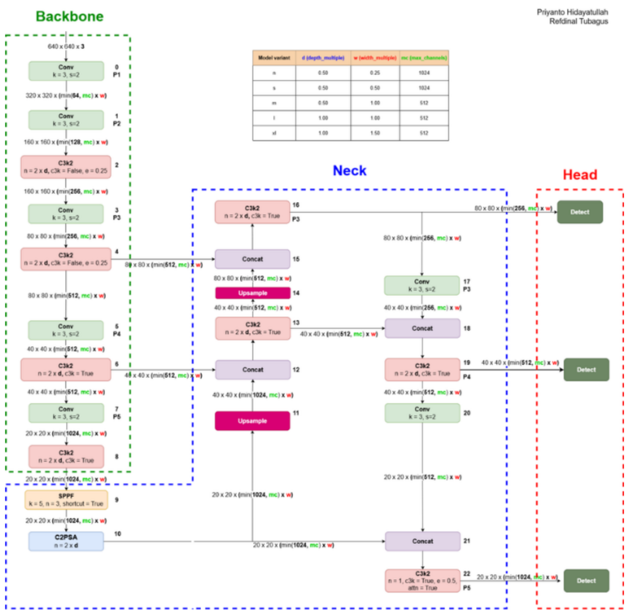

Fig. 3. Architecture of YOLO26. [7]

In the following tables we present the key performance indicators of the model used:

TABLE 1 YOLO26 DETECTION KPIS [6][7]

Model	size	mAP^{val}	mAP^{val}	Speed	Speed	params	FLOPs
	(pixels)	50-95	50-95(e2e)	CPU ONNX	T4 TensorRT10	(M)	(B)
				(ms)	(ms)		
YOLO26n	640	40.9	40.1	38.9 ± 0.7	1.7 ± 0.0	2.4	5.4
YOLO26s	640	48.6	47.8	87.2 ± 0.9	2.5 ± 0.0	9.5	20.7
YOLO26m	640	53.1	52.5	220.0 ± 1.4	4.7 ± 0.1	20.4	68.2
YOLO26l	640	55	54.4	286.2 ± 2.0	6.2 ± 0.2	24.8	86.4
YOLO26x	640	57.5	56.9	525.8 ± 4.0	11.8 ± 0.2	55.7	193.9

TABLE 2 YOLO26 CLASSIFICATION KPIS [6][7]

Model	size	acc	acc	Speed	Speed	params	FLOPs
	(pixels)	top1	top5	CPU ONNX	T4 TensorRT10	(M)	(B) at 224
				(ms)	(ms)		
YOLO26n-cls	224	71.4	90.1	5.0 ± 0.3	1.1 ± 0.0	2.8	0.5
YOLO26s-cls	224	76	92.9	7.9 ± 0.2	1.3 ± 0.0	6.7	1.6
YOLO26m-cls	224	78.1	94.2	17.2 ± 0.4	2.0 ± 0.0	11.6	4.9
YOLO26l-cls	224	79	94.6	23.2 ± 0.3	2.8 ± 0.0	14.1	6.2
YOLO26x-cls	224	79.9	95	41.4 ± 0.9	3.8 ± 0.0	29.6	13.6

TABLE 3 YOLO26 POSE KPIS [6][7]

Model	size	mAP^{pose}	mAP^{pose}	Speed	Speed	params	FLOPs
	(pixels)	50-95(e2e)	50(e2e)	CPU ON NX	T4 TensorRT10	(M)	(B)
				(ms)	(ms)		
YOLO26n-pose	640	57.2	83.3	40.3 ± 0.5	1.8 ± 0.0	2.9	7.5
YOLO26s-pose	640	63	86.6	85.3 ± 0.9	2.7 ± 0.0	10.4	23.9
YOLO26m-pose	640	68.8	89.6	218.0 ± 1.5	5.0 ± 0.1	21.5	73.1
YOLO26l-pose	640	70.4	90.5	275.4 ± 2.4	6.5 ± 0.1	25.9	91.3
YOLO26x-pose	640	71.6	91.6	565.4 ± 3.0	12.2 ± 0.2	57.6	201.7

As our focus is on edge applications and low-powered hardware, we will be using the n and s variants.

2) The Dataset: One of the biggest challenges in both the military and the disaster relief fields of study as always is the available datasets for training and development of appropriate models and pipelines.Especially as such, when we are trying to view deifferent poses. Fortunately there are examples of such datasets that can be used, one such dataset is the C2A Dataset: Human Detection in Disaster Scenarios [8]. The C2A (Combination to Application) Dataset is a resource designed to advance human detection in disaster scenarios using UAV imagery. This dataset addresses a critical gap in the field of computer vision and disaster response by providing a large-scale, diverse collection of synthetic images that combine real disaster scenes with human poses.This dataset combines sources such as the Aerial Image Dataset for Emergency Response Applications (AIDER) and the human poses derived from the Multiple Poses Human Body (LSP/MPII-MPHB) dataset. Key features include:

a) 10,215 high-resolution images

b) Over 360,000 annotated human instances

c) 5 human pose categories: Bent, Kneeling, Lying, Sitting, and Upright

d) 4 disaster scenario types: Fire/Smoke, Flood, Collapsed Building/Rubble, and Traffic Accidents

e) Image resolutions ranging from 123x152 to 5184x3456 pixels

f) Bounding box annotations for each human instance

Examples of these can be seen in Fig.4-6.

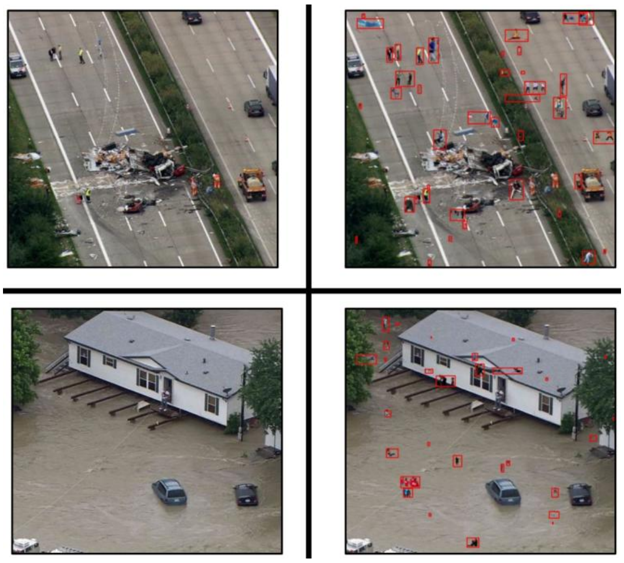

Fig. 4. Example of detection in C2A dataset. [8]

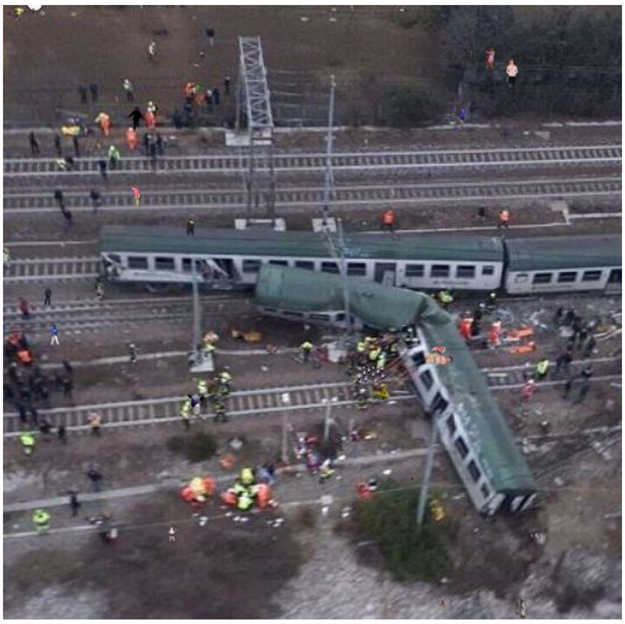

Fig. 5. Example of augmented image in C2A dataset. [8]

Fig. 6. Example of augmented image in C2A dataset. [8]

B. Results

We created a three-branch pipeline for posture classification in order to compare different ways of extracting discriminative information from the same input images. This was motivated by the fact that posture recognition can depend on global appearance and on local body configurations, therefore it is not optimal to have a single representation in all cases.

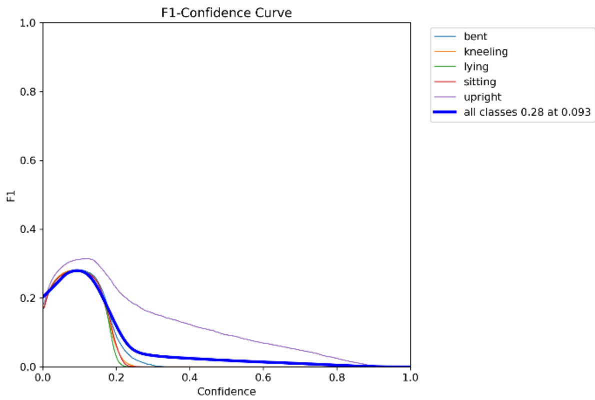

Fig. 7. F1 curve for Branch 1.

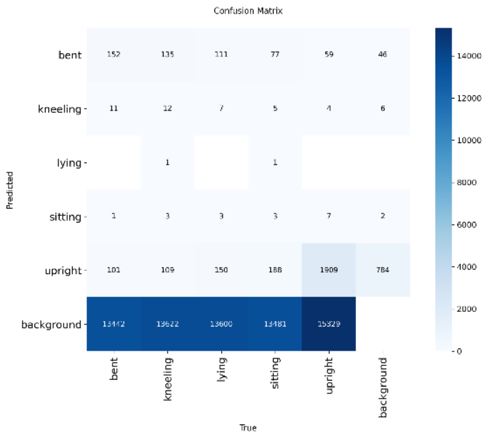

Fig. 8. Branch 1 Confusion Matrix.

1) Branch 1: Direct image-based detector, which had a role of providing a baseline, using the full visual context of the image. This included the silhouette, orientation and scene-level cues. This was expected to be the most robust one, as it does not depend on intermediate pose estimations and exploits all of the available visible discriminative information.

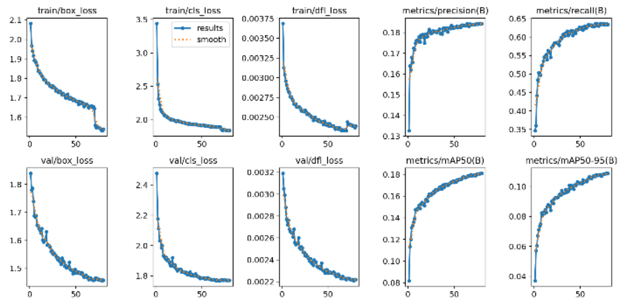

Fig. 9. Branch 1 results

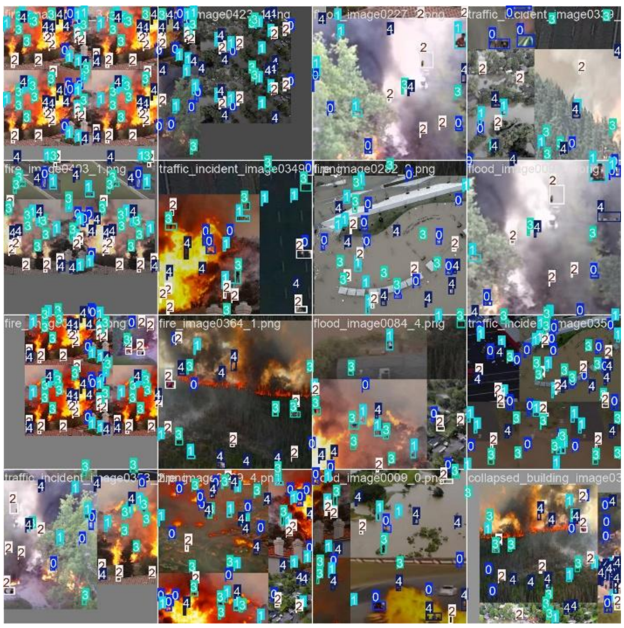

Fig. 10. Branch 1: Training Batch

Fig. 11. Branch 1: Validation batch, labels

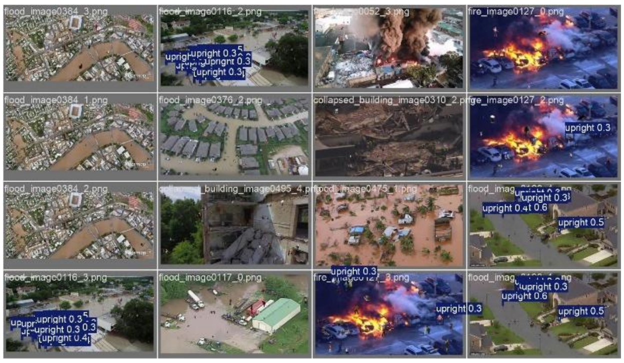

Fig. 12. Branch 1: Validation batch, predictions

2) Branch 2: This was designed around keypoints extracted from the entire image. The main idea was that an explicit pose representation may improve classification by applying a focus on the structure of the body, not on appearance. In practice, the full-image formulation introduced too much irrelevant context and did not isolate the person tightly enough. This significantly reduced the usefulness of the keypoint signal.

3) Branch 3: Here we used keypoints that were extracted from a cropped, more localized region of the image, centred on the identified person. Here we wanted to preserve the benefits of pose-based reasoning while reducing the background noise and to improve the concentration of the pose signal. This resulted in the most promising pose-based formulation, because it retained more useful information, compare to the full-image branch.

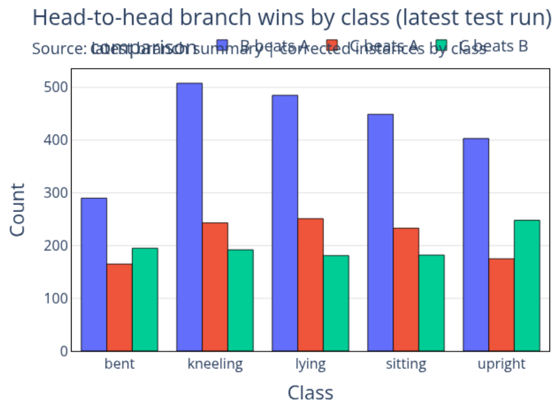

Fig. 13. Branch wins by class

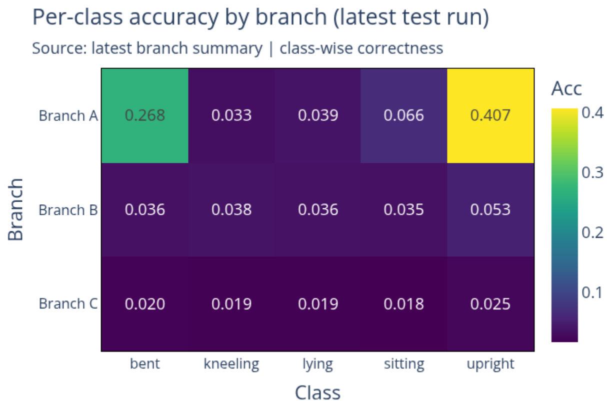

Fig. 14. Heatmap between the branches

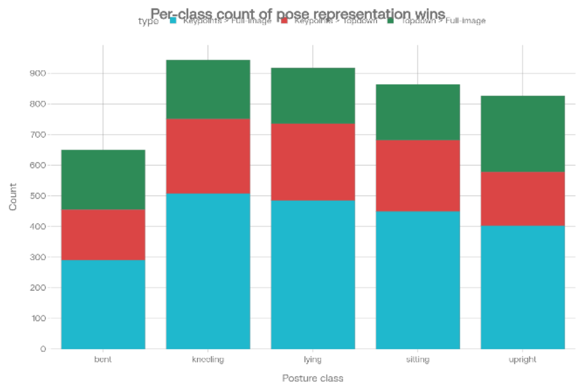

Fig. 15. Per-class wins

Fig. 16. Per-class rates

During the experimentation, we came to the conclusion, that the original full image keypoint branch was not providing a sufficient signal, the pipeline had to be changed in order to perform the extraction on a more localized crop rather than on the entire image. This improved the pose input and the whole branch was more interpretable.

An adjustment was made also to the evaluation strategy, so that it better reflects the complementary behaviour of the branches. Instead of presenting them as equal model candidates, we treat the direct detector as the primary reference method, the crop-based keypoint branch as the main pose-based method, and the full-image branch as an ablation showing the limitations of the original pose formulation.

Blood Detection is also very important in disaster relief scenarios. As such, a optional detector focusing on this was trained as well. It performs really well as seen in the figure 17.

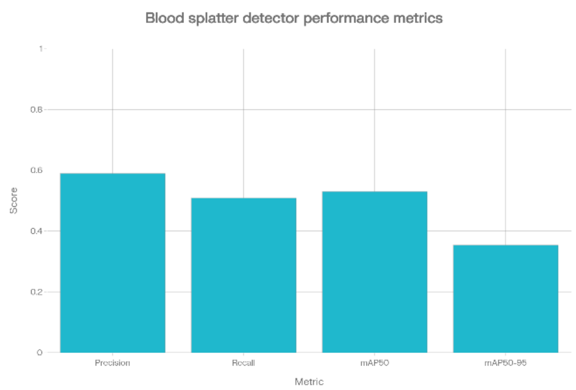

Fig. 17. Blood splatter metrics

It can be a very useful addition to the pipeline, as long as the other branches are performing as expected, as this can be a major decision making indicator.

C. Discussion

The main finding is that appearance-based detection remains as the most reliable approach, while pose-based classification becomes useful only after localization of the input. This suggests that the critical factor was not the pose information, but the quality of the pose representation. When the region for the body is tightly cropped, the model preserved a more discriminative structure and produced better results compared to the full-image branch.

This shows us that the design of the pipelines and the branches within matters as much, or even more that the choice of the detection model. A pose branch can fail not because the pose does not provide enough information, but because the input is too noisy or too diluted by the background context. This is evidenced by the revised crop-based branch, and the improved performance after localization.

Future work should be focused on converting the current branch-level combinations to a true learned fusion framework. The natural next step is confidence-aware ensembling, where branch outputs are weighed according to prediction certainty or class-specific reliability. Other steps could include probability calibration for more comparable detector and pose branches before fusion; adding a meta-classifier, which can learn when to trust each branch, this will probably outperform the fixed logic.

ACKNOWLEDGMENT

This work was supported by the NSP SD program, which has received funding from the Ministry of Education and Science of the Republic of Bulgaria under the grant agreement no. D01-74/19.05.2022.

REFERENCES

[1] Yibo Sun and Zhe Sun and Weitong Chen, The evolution of object detection methods https://www.sciencedirect.com/science/article/pii/S095219762400616X, 2024. https://doi.org/10.1016/j.engappai.2024.108458

[2] Rujing Yue, Zhiqiang Tian, Shaoyi Du, Action recognition based on RGB and skeleton data sets: A survey https://www.sciencedirect.com/science/article/abs/pii/S0925231222011596, 2022, https://doi.org/10.1016/j.neucom.2022.09.071

[3] Best Drone for Search and Rescue: Emergency Response Drones for Teams, [Online]. Available: https://www.flybyguys.com/blog/drones-for-search-and-rescue

[4] New AI-powered fog-penetrating first responder drone debuts at CES 2026, [Online]. Available: https://aerospaceglobalnews.com/news/ai-powered-search-rescue-drone-ces-2026/#:~:text=New%20AI%2Dpowered%20fog%2Dpenetrating,support%20search%20and%20rescue%20operations.&text=GDU%20says%20its%20new%20UAV,offers%20significantly%20improved%20imaging%20capabilities.

[5] Emergency and Rescue Robot Market, [Online]. Available: https://www.intelmarketresearch.com/emergencyrescue-robot-market-26181#:~:text=Startups%20like%20Unitree%20Robotics%20are,and%20Rescue%20Robot%20Market%20Trends

[6] Glenn Jocher and Jing Qiu, Ultralytics YOLO26, [Online]. Available: https://github.com/ultralytics/ultralytics

[7] Priyanto Hidayatullah, Refdinal Tubagus, YOLO26: A Comprehensive Architecture Overview and Key Improvements https://www.researchgate.net/publication/400855947_YOLO26_A_Comprehensive_Architecture_Overview_and_Key_Improvements, 2026.

[8] Ragib Amin Nihal, Benjamin Yen, Katsutoshi Itoyama, and Kazuhiro Nakadai, UAV-Enhanced Combination to Application: Comprehensive Analysis and Benchmarking of a HumanDetection Dataset for Disaster Scenarios, https://arxiv.org/pdf/2408.04922, 2024