

# A Synthetic Dataset for Manometry Recognition in Robotic Applications

Pedro Antonio Rabelo Saraiva<sup>1</sup>, Enzo Ferreira de Souza<sup>1</sup>, João Manoel Herrera Pinheiro<sup>1</sup>,  
Thiago H. Segreto<sup>1</sup>, Ricardo V. Godoy<sup>1</sup>, and Marcelo Becker<sup>1</sup>

**Abstract**—This paper addresses the challenges of data scarcity and high acquisition costs in training robust object detection models for complex industrial environments, such as offshore oil platforms. Data collection in these hazardous settings often limits the development of autonomous inspection systems. To mitigate this issue, we propose a hybrid data synthesis pipeline that integrates procedural rendering and AI-driven video generation. The approach uses BlenderProc to produce photorealistic images with domain randomization and NVIDIA’s Cosmos-Predict2 to generate physically consistent video sequences with temporal variation. A YOLO-based detector trained on a composite dataset, combining real and synthetic data, outperformed models trained solely on real images. A 1:1 ratio between real and synthetic samples achieved the highest accuracy. The results demonstrate that synthetic data generation is a viable, cost-effective, and safe strategy for developing reliable perception systems in safety-critical and resource-constrained industrial applications.

## I. INTRODUCTION

The uninterrupted functioning and security of vital infrastructure, especially within the petroleum and natural gas sectors, are of paramount importance due to the inherent dangers associated with the handled materials and the severe repercussions of system malfunctions [1]–[3]. Conventional maintenance and inspection procedures in these intricate industrial settings are largely dependent on human personnel, who are frequently exposed to substantial hazards such as contact with toxic agents, extreme temperatures, and confined spaces [4], [5]. Accordingly, there is a strong demand for technologies that increase inspection effectiveness [6], [7] while simultaneously reducing human exposure in dangerous scenarios [8]–[10].

The introduction of autonomous legged robots, e.g., Unitree B2 [11] and ANYmal C [12], offers a promising alternative [13], [14]. Unlike wheeled or tracked systems that struggle with uneven terrain, legged platforms can operate in a reliable manner even in the presence of stairs and irregular surfaces [15], enabling routine, automated data capture that minimizes downtime and reduces human risk [16].

However, reliable perception remains a key bottleneck: robots must detect and localize facility-critical assets—valves, pipes, gauges, and safety equipment—to assess conditions and comply with procedures in visually cluttered plants [3], [17].

This work was carried out with the support of Petrobras, using resources from the R&D clause of the ANP, in partnership with the University of São Paulo (USP) and the intervening foundation Fundação de Apoio à Física e à Química (FAFQ), under Cooperation Agreement No. 2023/00013-7 and 2023/00016-6.

<sup>1</sup>Pedro Saraiva, Enzo Ferreira de Souza, Thiago H. Segreto, Ricardo V. Godoy, and Marcelo Becker are with the Department of Mechanical Engineering, University of São Paulo, São Carlos, Brazil. [becker@sc.usp.br](mailto:becker@sc.usp.br)

Acquiring and annotating large, diverse datasets for training modern detectors (e.g., YOLO) is time-consuming, costly, and often disruptive to operations; access constraints and safety risks further limit in-plant data collection [18]–[21].

To overcome these limitations, in this work, we leverage high-quality synthetic data at two complementary levels. First, we use BlenderProc—an open-source, procedural rendering pipeline—to generate photorealistic images with controlled domain randomization over geometry, illumination, materials, and sensors [22]–[24]. Second, we integrate NVIDIA Cosmos-Predict2, NVIDIA’s latest World Foundation Model (WFM) family for “physical AI” [25]–[27], via ComfyUI [28] workflows to synthesize physically plausible imagery and video priors from rare viewpoints, adverse lighting, and hard-to-stage events, thus broadening coverage beyond what is practical to render or capture in the field. Cosmos provides model variants and deployment tooling intended for robotics and autonomous vehicles (AV) pipelines, enabling scalable generation and curation that complements procedural graphics and reduces the sim-to-real gap in downstream perception [29].

This work introduces a practical hybrid synthetic plus real pipeline. Blending BlenderProc rendering with Cosmos-Predict2 generation for industrial asset perception, we report a case study on analog pressure gauges (manometers) and show that the same pipeline readily transfers to valves, digital or analog gauges and displays, flowmeters, placards, tags, and other safety-critical industrial devices with minimal changes, primarily swapping the 3D models. We build a large-scale, richly annotated synthetic dataset and train a YOLO-based detector, showing that synthetic data can substantially reduce reliance on costly and hazardous in-plant collection while improving recognition of safety-critical assets in complex environments.

The rest of this paper is organized as follows. Section II presents the related works. Section III details the unified pipeline procedural dataset creation with BlenderProc and AI-extended synthesis with Cosmos-Predict2 (ComfyUI) and training/validation of the object detector. Section IV presents quantitative and qualitative results, while Section V concludes with limitations and future work.

## II. RELATED WORK

Synthetic data has become a practical solution for the scarcity and safety constraints of in-plant data collection, with domain randomization and photorealistic rendering shown to reduce the sim-to-real gap for perception and control [30]. Early sim-to-real studies demonstrated that heavy

appearance randomization can transfer object localization and even collision-avoidance policies trained purely in simulation to the real world [31], [32]. Recent analyses formalize when and how randomization narrows the gap and how to tune distributions over simulator parameters [33], [34].

On the procedural/graphics side, toolchains such as BlenderProc provide modular, scriptable generation of photorealistic images with rich annotations and sensor simulation—now a common backbone for dataset bootstrapping in robotics and industrial vision [22]. Parallel efforts in commercial-grade simulators (NVIDIA Omniverse/Isaac Sim Replicator) expose APIs for large-scale synthetic data generation with domain randomization, sensor models, and annotators, frequently used in AV/robotics pipelines [35], [36]. Unity’s Perception stack similarly targets turnkey dataset creation with labelers and randomized scenes for object detection and pose estimation [37].

Complementary to procedural rendering, world foundation models (WFMs) have emerged as AI-native generators that simulate and predict physically plausible scenes directly from text/visual conditions. NVIDIA’s Cosmos platform targets “physical AI” use cases (robotics, AV, industrial vision), exposing Cosmos-Predict2 models for Video-to-World generation (2B and 14B variants) and post-training/fine-tuning workflows aimed at scalable coverage of rare viewpoints, adverse illumination, and hard-to-stage events [38]–[40]. Public releases and reports emphasize open accessibility and training on large corpora of robotics/drive videos, positioning Cosmos as a data engine to complement graphics-based pipelines rather than replace them [41], [42]. Community workflows via ComfyUI further reduce integration friction, enabling mixed procedural and generative curation where Cosmos assets are inserted into Blender/YOLO pipelines [43]. Taken together, recent work suggests that hybrid pipelines, which incorporate procedural rendering for precise labels and controllable physics, along with WFM-based synthesis for broader coverage and temporal variation, can significantly reduce reliance on hazardous and costly field acquisition in complex industrial environments.

### III. METHODOLOGY

Our methodology combines real data acquisition, synthetic data generation through BlenderProc, and AI-extended video synthesis using NVIDIA Cosmos-Predict2 integrated via ComfyUI. The unified pipeline is illustrated in Fig. 1.

#### A. Real Data Acquisition and Annotation

We collected approximately 2,500 real images and video frames of industrial manometers. These samples were annotated using a hybrid approach: manual bounding-box labeling with CVAT and semi-automatic segmentation with SAM2, as shown in Fig. 2. A small portion of these annotations was further reviewed by experts to ensure quality.



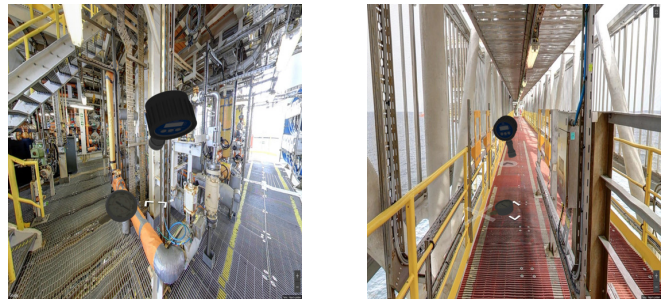
(a) Manual bounding boxes in CVAT.

(b) SAM2-assisted segmentation overlays.

Fig. 2: Examples of real data annotations: (a) manual boxes curated in CVAT; (b) semi-automatic pixel masks using SAM2, later spot-checked by experts.

#### B. Synthetic Data with BlenderProc

3D models of manometers were rendered with BlenderProc. To improve generalization and reduce overfitting, domain randomization was applied: random backgrounds (industrial textures and photos), lighting variations, camera pose sampling, and post-processing (noise, blur, chromatic aberration), with examples provided in Fig. 3. BlenderProc provides pixel-perfect labels, ensuring accurate segmentation and bounding-box masks.



(a) Manometer in a random background.

(b) An example of multiple blend files.

Fig. 3: BlenderProc synthetic samples.

#### C. AI-Extended Video Generation (ComfyUI + Cosmos-Predict2)

To complement static synthetic renders, we integrated AI-driven video extension, illustrated in Fig. 4. Real short clips of manometers were expanded using ComfyUI workflows with NVIDIA Cosmos-Predict2, enabling frame synthesis with temporal consistency, relighting, viewpoint changes, and motion blur. These pseudo-extended frames introduce temporal diversity (occlusion, reflections, jitter) that is difficult to model with static rendering. Pseudo-labels were propagated across frames using tracking and filtered with confidence thresholds, with human-in-the-loop auditing on a subset to mitigate noise.



(a) Relighting and viewpoint changes with motion blur. (b) A frame from a short clip made with Cosmos.

Fig. 4: AI-extended video frames produced via ComfyUI + NVIDIA Cosmos-Predict2.

#### D. Data Composition Strategy

All data sources were merged into a unified dataset. To investigate the impact of proportions, we defined scenarios with:

- Real only (2,500 images)
- Mixed (real + synthetic in ratios 1:1, 1:3)

Within the synthetic portion, we tested different splits between BlenderProc renders and AI-extended video frames, adopting 70% BlenderProc and 30% video-based synthesis as baseline.

#### E. Training and Evaluation

All scenarios were trained with the same YOLO-based architecture using identical hyperparameters. Evaluation was performed exclusively on a held-out set of real images. Metrics include mean Average Precision (mAP) at intersection-over-union (IoU) thresholds .5:.95, recall, Average Recall

(AR), per-class Average Precision (AP), and precision–recall (PR) curves. We also report normalized metrics (mAP divided by average instances per image), shown in Fig. 6, and perform ablations on domain randomization factors.

#### F. Computational Setup

We ran the experiments on a local workstation with  $2 \times$  NVIDIA RTX A2000 12 GB. The video synthesis process using ComfyUI required approximately 8 minutes to generate a 15-second video, which was subsequently decomposed at 20 frames per second, yielding an average of 300 frames per minute. In comparison, rendering 1000 images in Blender took approximately 13 minutes, corresponding to a throughput of 77 images per minute. Regarding the YOLO training times, Table I summarizes the total training duration for each dataset configuration.

TABLE I: Total training duration

Scenario	Training Time
Real Only	3h 18m 8s
Mix 1:1	1h 54m 45s
Mix 1:3	3h 45m 39s
Mix 0.5:0.5	1h 16m 47s

## IV. RESULTS

In this section we present the experimental protocol, justification for synthetic data generation, proportionality choices between real and synthetic samples, and the results obtained in different scenarios.

#### A. Experimental Protocol

Four training scenarios were defined: (i) only real images (2,500) and (ii) a mix of real and synthetic images.

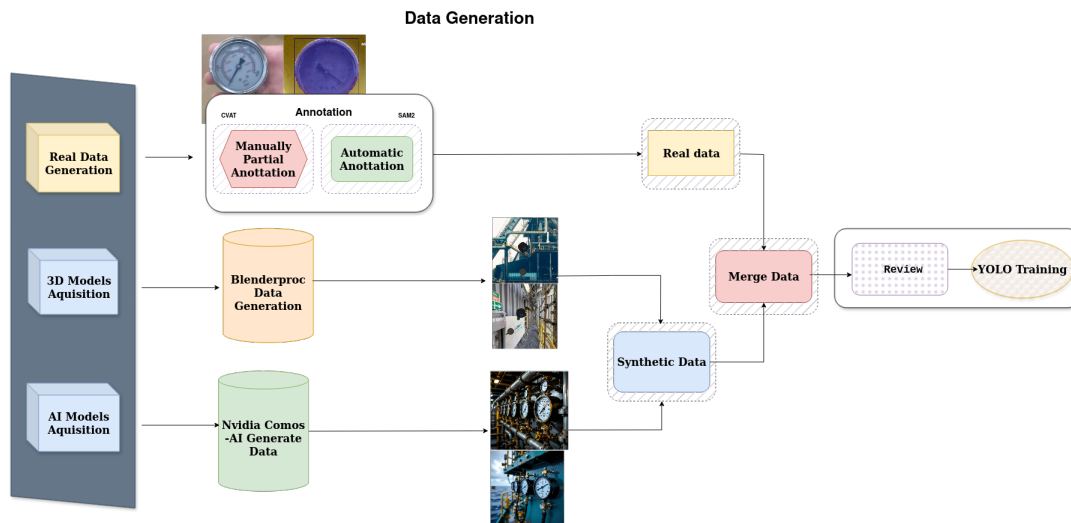


Fig. 1: Proposed methodology pipeline for dataset creation and training. The framework integrates three sources: (i) real data acquisition and annotation (manual via CVAT (*Computer Vision Annotation Tool*) and semi-automatic via SAM2 (*Segment Anything Model 2*)), (ii) synthetic renders generated in BlenderProc with domain randomization (backgrounds, illumination, pose, post-processing), and (iii) AI-extended video frames generated through ComfyUI with NVIDIA Cosmos-Predict2 (relighting, motion, occlusion). All data sources are reviewed, merged, and used for training the YOLO-based object detector.

For mixed datasets, we tested ratios of 0.5:0.5, 1:1, and 1:3 (real:synthetic). All scenarios were trained with the same YOLO-based detection architecture and evaluated on the same test split of real images.

### B. Justification

The adoption of synthetic data is motivated by three factors: 1) the cost and time of annotating real images are significantly higher than generating synthetic ones; 2) domain randomization techniques such as random backgrounds, lighting and camera poses reduce overfitting and bridge the sim-to-real gap; 3) prior work shows that mixing 5–20% real data with synthetic data is often enough to reach competitive performance.

### C. Proportionality of Synthetic Data

Considering the availability of 2,500 real photos, we defined proportionality as follows: - 1:1  $\rightarrow$  2,500 real + 2,500 synthetic - 1:3  $\rightarrow$  2,500 real + 7,500 synthetic

Inside the synthetic portion, we explored the impact of splitting between BlenderProc-rendered images and AI-extended video frames. BlenderProc provides proper labels and strong control over background, illumination, and pose, while video-based synthesis injects temporal variation such as blur, occlusion, and camera jitter. The baseline adopted was 70% BlenderProc and 30% video-based frames, which balances labeling accuracy and temporal diversity.

### D. Results

Table II shows dataset composition and Table III reports object detection performance.

TABLE II: Dataset composition (examples).

Scenario	Real Images	Synthetic Images	Total
Real only	2500	0	2500
Mix 1:1	2500	2500	5000
Mix 1:3	2500	7500	10000
Mix 0.5:0.5	1250	1250	2500

TABLE III: Performance comparison (updated with 0.5:0.5).

Scenario	mAP@[.5:.95]	Recall	AR
Real only	0.936	0.969	0.969
Mix 1:1	0.962	0.972	0.972
Mix 1:3	0.928	0.948	0.948
Mix 0.5:0.5	0.868	0.933	0.933

These results indicate that mixed 1:1 dataset outperforms the real-only baseline. In our experiments, the 1:1 mix achieved the highest mAP50-95, while the 1:3 mix still surpassed the real-only model but trailed the 1:1 configuration, as depicted in the training evolution in Fig. 5. These findings corroborate the broader consensus that blending a modest amount of real data with a larger volume of synthetic data can guide the detector toward realistic features while benefiting from the diversity and volume of synthetic samples.

### E. Additional Analyses

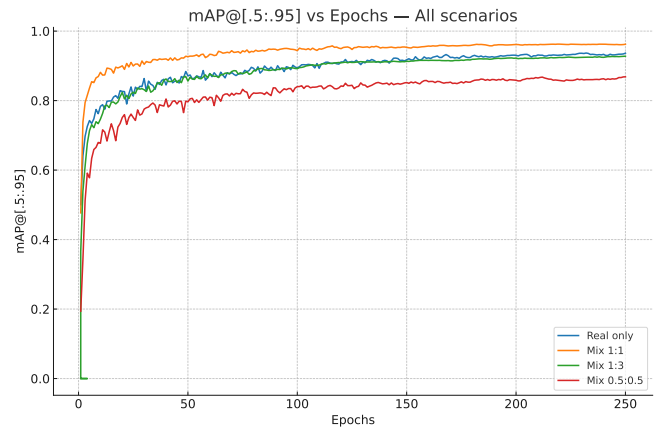


Fig. 5: These results indicate that adding synthetic data is beneficial, but excessive synthetic proportion (e.g., 1:3) does not yield further gains over Real-only and incurs additional compute cost.

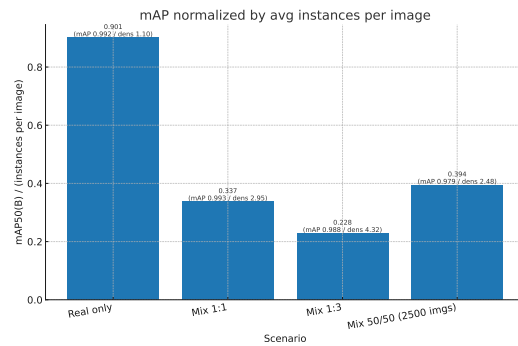


Fig. 6: Normalized mAP (0.5–0.95) per dataset scenario, corrected by the average number of instances per image. By normalizing precision scores with respect to the number of object instances, this metric mitigates the effect of dataset imbalance and varying annotation densities

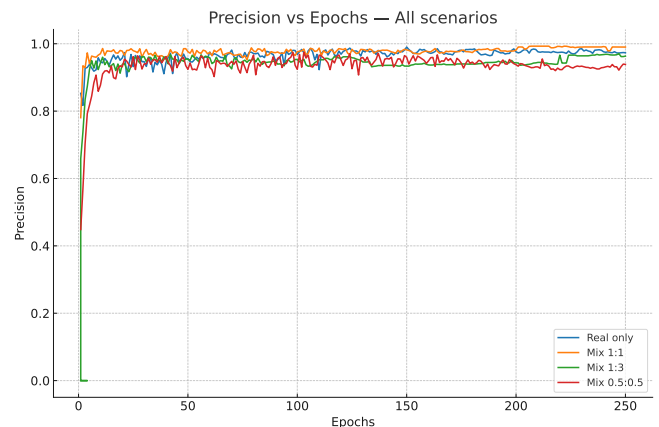


Fig. 7: Precision over training epochs for four training regimes. The synthetic-augmented configuration reaches high precision within the first epochs and remains near saturation. Synthetic data can accelerate the reduction of false positives.

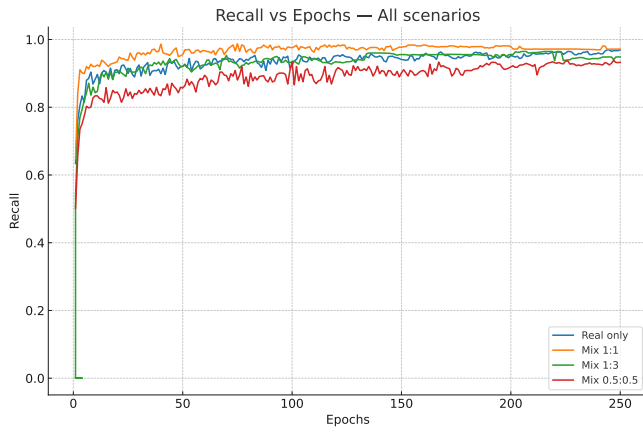


Fig. 8: Recall across training epochs for four training regimes. Mix 1:1 provides the most consistent recall gains. Real-only and Mix 1:3 display similar recall trajectories, whereas Mix 0.5:0.5 is lower yet notably better than a naive halving of the real dataset would suggest, evidencing the compensatory effect of synthetic data.

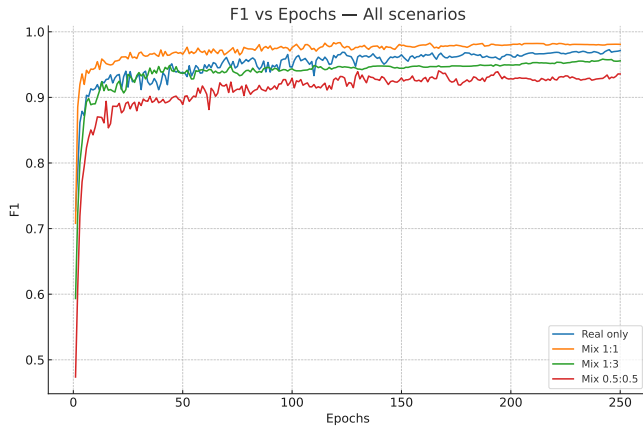


Fig. 9: F1-score over epochs for four training regimes. The curves mirror the recall advantage of Mix 1:1 while preserving high precision. Real-only and Mix 1:3 are nearly tied, and Mix 0.5:0.5 trails but benefits from synthetic augmentation given the reduced real data.

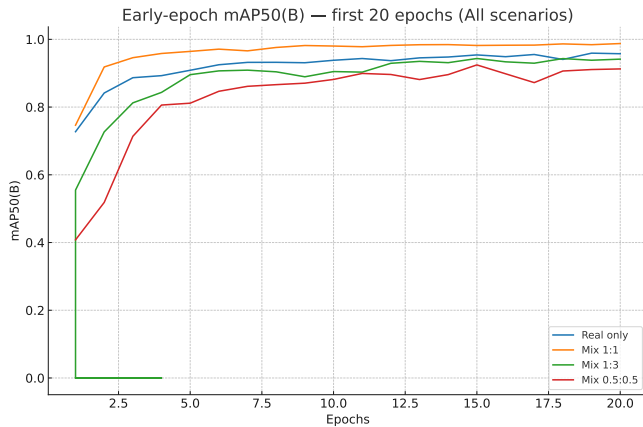


Fig. 10: Early-epoch mAP50(B) (first 20 epochs) comparing four training regimes.

TABLE IV: Number of instances in each scenario

Scenario	Instances
Real only	2753
Mix 1:1	13560
Mix 1:3	43242
Mix 0.5:0.5	6207

Figures 7–9 showcase the training dynamics of *Precision*, *Recall*, and *F1-score* across the four regimes (Real only, Mix 1:1, Mix 1:3, and Mix 0.5:0.5), enabling a side-by-side inspection of convergence behavior. A consistent pattern is that the Mix 1:1 configuration accelerates early learning relative to the Real-only baseline, reflected in higher early-epoch averages. In contrast, Mix 1:3 tends to underperform in the same interval and later converges to results that are practically tied with Real-only. The Mix 0.5:0.5 setting exhibits the slowest convergence and lowest asymptotic accuracy; nevertheless, despite using only half the real images, synthetic augmentation prevents a sharp degradation, indicating that synthetic data can partially compensate for small or imperfect real dataset. Taken together, these analyses suggest that synthetic data can significantly improve early optimization and stability, especially in Recall and the resulting F1 score, provided that the synthesis pipeline is properly configured; otherwise, suboptimal choices in synthetic generation may negate the expected benefits.

## V. CONCLUSIONS

This paper showed that combining synthetic and real images enhances detection performance and training stability in industrial settings. A 1:1 mixture achieves the highest overall accuracy, while real-only and 1:3 mixes yield similar results, indicating diminishing returns when synthetic data dominates. Even with only half the real images, synthetic augmentation mitigates performance drops, highlighting its value for small or imperfect datasets.

Limitations include evaluation on a single asset class and camera setup, focus on 2D detection only, potential temporal artifacts from video synthesized frames, and no benchmarking on embedded platforms. Future work will expand to multiple asset types, richer tasks (instance segmentation, 6D pose, needle angle/OCR), multimodal sensing, cross-domain adaptation, and on-robot validation with embedded hardware, incorporating active learning and quality/uncertainty monitoring. This approach can support more robust, cost-effective industrial perception systems.

## VI. DATA AVAILABILITY

The data that support the findings of this study are available in Figshare at [doi.org/10.6084/m9.figshare.29936768.v1](https://doi.org/10.6084/m9.figshare.29936768.v1), reference number 29936768 [44].

## REFERENCES

- [1] F. I. Khan and M. Haddara, “Risk-based maintenance (rbm): A new approach for process plant inspection and maintenance,” *Process Safety Progress*, vol. 23, no. 4, pp. 252–265, 2004.
- [2] J. Bhandari, E. Arzaghi, R. Abbassi, V. Garaniya, and F. Khan, “Dynamic risk-based maintenance for offshore processing facility,” *Process Safety Progress*, vol. 35, no. 4, pp. 399–406, 2016.

- [3] M. Hutter, R. Diethelm, S. Bachmann, P. Fankhauser, C. Gehring, V. Tsounis, A. Lauber, F. Guenther, M. Bjelonic, L. Isler, H. Kolvenbach, K. Meyer, and M. Hoepflinger, "Towards a generic solution for inspection of industrial sites," in *Field and Service Robotics*, M. Hutter and R. Siegwart, Eds. Cham: Springer International Publishing, 2018.
- [4] F. I. Khan and S. Abbasi, "Major accidents in process industries and an analysis of causes and consequences," *Journal of Loss Prevention in the Process Industries*, vol. 12, no. 5, pp. 361–378, 1999.
- [5] J.-F. Yang, P.-C. Wang, X.-Y. Liu, M.-C. Bian, L.-C. Chen, S.-Y. Lv, J.-F. Tao, G.-Y. Suo, S.-Q. Xuan, R. Li, J.-W. Zhang, C.-M. Shu, and Z. Dou, "Analysis on causes of chemical industry accident from 2015 to 2020 in chinese mainland: A complex network theory approach," *Journal of Loss Prevention in the Process Industries*, vol. 83, p. 105061, 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0950423023000918>
- [6] A. Khan, S. Gupta, and S. K. Gupta, "Emerging uav technology for disaster detection, mitigation, response, and preparedness," *Journal of Field Robotics*, vol. 39, no. 6, pp. 905–955, 2022.
- [7] Q. Ma, W. Liang, and P. Zhou, "A review on pipeline in-line inspection technologies," *Sensors*, vol. 25, no. 15, 2025. [Online]. Available: <https://www.mdpi.com/1424-8220/25/15/4873>
- [8] M. Fisher, R. C. Cardoso, E. C. Collins, C. Dadswell, L. A. Dennis, C. Dixon, M. Farrell, A. Ferrando, X. Huang, M. Jump, G. Kourtis, A. Lisitsa, M. Luckcuck, S. Luo, V. Page, F. Papacchini, and M. Webster, "An overview of verification and validation challenges for inspection robots," *Robotics*, vol. 10, no. 2, 2021.
- [9] M. Karkoub, O. Bouhali, and A. Sheharyar, "Gas pipeline inspection using autonomous robots with omni-directional cameras," *IEEE Sensors Journal*, vol. 21, no. 14, pp. 15 544–15 553, 2021.
- [10] W. E. Heap, S. Man, V. Bassari, S. Nguyen, E. B. Yao, N. A. Tripathi, N. D. Naclerio, and E. W. Hawkes, "Large-scale vine robots for industrial inspection: Developing a new framework to overcome limitations with existing inspection methods," *IEEE Robotics & Automation Magazine*, pp. 2–13, 2024.
- [11] Q. Li, F. Cicirelli, A. Vinci, A. Guerrieri, W. Qi, and G. Fortino, "Quadruped robots: Bridging mechanical design, control, and applications," *Robotics*, vol. 14, no. 5, 2025. [Online]. Available: <https://www.mdpi.com/2218-6581/14/5/57>
- [12] M. Hutter, C. Gehring, A. Lauber, F. Gunther, C. D. Bellicoso, V. Tsounis, P. Fankhauser, R. Diethelm, S. Bachmann, M. Bloesch, H. Kolvenbach, M. Bjelonic, L. Isler, and K. M. and, "Anymal - toward legged robots for harsh environments," *Advanced Robotics*, 2017.
- [13] J. Trevelyan, W. R. Hamel, and S.-C. Kang, *Robotics in Hazardous Applications*. Springer International Publishing, 2016.
- [14] J. Yang, C. Wang, B. Jiang, H. Song, and Q. Meng, "Visual perception enabled industry intelligence: State of the art, challenges and prospects," *IEEE Transactions on Industrial Informatics*, 2021.
- [15] M. Ramezani, M. Brandao, B. Casseau, I. Havoutis, and M. Fallon, "Legged robots for autonomous inspection and monitoring of offshore assets," in *Offshore Technology Conference*. OTC, 2020, p. D011S006R005.
- [16] C. Gehring, P. Fankhauser, L. Isler, R. Diethelm, S. Bachmann, M. Potz, L. Gerstenberg, and M. Hutter, "Anymal in the field: Solving industrial inspection of an offshore hvdc platform with a quadrupedal robot," in *Field and Service Robotics*, G. Ishigami and K. Yoshida, Eds. Singapore: Springer Singapore, 2021, pp. 247–260.
- [17] A. Bonci, P. D. Cen Cheng, M. Indri, G. Nabissi, and F. Sibona, "Human-robot perception in industrial environments: A survey," *Sensors*, vol. 21, no. 5, 2021.
- [18] N. Manakitsa, G. S. Maraslidis, L. Moysis, and G. F. Fragulis, "A review of machine learning and deep learning for object detection, semantic segmentation, and human action recognition in machine and robotic vision," *Technologies*, vol. 12, no. 2, 2024.
- [19] C. Shorten and T. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, 07 2019.
- [20] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 2016.
- [21] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," 2020. [Online]. Available: <https://arxiv.org/abs/2004.10934>
- [22] M. Denninger, M. Sundermeyer, D. Winkelbauer, Y. Zidan, D. Olefir, M. Elbadrawy, A. Lodhi, and H. Katam, "Blenderproc," 2019.
- [23] L. Eversberg and J. Lambrecht, "Generating images with physics-based rendering for an industrial object detection task: Realism versus domain randomization," *Sensors*, vol. 21, no. 23, 2021.
- [24] J. Dirr, D. Gebauer, J. Yao, and R. Daub, "Automatic image generation pipeline for instance segmentation of deformable linear objects," *Sensors*, vol. 23, no. 6, 2023.
- [25] NVIDIA Corporation. (2025) Nvidia cosmos: World foundation models for physical ai. Accessed: 2025-08-18. [Online]. Available: <https://www.nvidia.com/en-us/ai/cosmos/>
- [26] NVIDIA Research. (2025) Cosmos-predict2. Accessed: 2025-08-18. [Online]. Available: <https://research.nvidia.com/labs/dir/cosmos-predict2/>
- [27] NVIDIA Cosmos Team, "Cosmos-predict2," <https://github.com/nvidia-cosmos/cosmos-predict2>, 2025, apache-2.0 license, Accessed: 2025-08-18.
- [28] ComfyUI Authors. (2025) Comfyui examples: Nvidia cosmos predict2. Accessed: 2025-08-18. [Online]. Available: [https://comfyanonymous.github.io/ComfyUI\\_examples/cosmos\\_predict2/](https://comfyanonymous.github.io/ComfyUI_examples/cosmos_predict2/)
- [29] NVIDIA Corporation. (2025) Nvidia cosmos documentation. Accessed: 2025-08-18. [Online]. Available: <https://docs.nvidia.com/cosmos/index.html>
- [30] G. Paulin and M. Ivasic-Kos, "Review and analysis of synthetic dataset generation methods and techniques for application in computer vision," *Artificial Intelligence Review*, vol. 56, no. 9, pp. 9221–9265, 2023. [Online]. Available: <https://doi.org/10.1007/s10462-022-10358-3>
- [31] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," *arXiv preprint arXiv:1703.06907*, 2017.
- [32] F. Sadeghi and S. Levine, "Cad2rl: Real single-image flight without a single real image," *arXiv preprint arXiv:1611.04201*, 2016.
- [33] Q. Vuong, S. Vikram, H. Su, S. Gao, and H. I. Christensen, "How to pick the domain randomization parameters for sim-to-real transfer of reinforcement learning policies?" *arXiv preprint arXiv:1903.11774*, 2019. [Online]. Available: <https://arxiv.org/abs/1903.11774>
- [34] X. Chen *et al.*, "Understanding domain randomization for sim-to-real transfer," in *International Conference on Learning Representations (ICLR)*, 2022. [Online]. Available: <https://openreview.net/forum?id=T8vZHIRTyY>
- [35] NVIDIA Corporation. (2025) Omniverse replicator. Accessed: 2025-08-18. [Online]. Available: [https://docs.omniverse.nvidia.com/extensions/latest/ext\\_replicator.html](https://docs.omniverse.nvidia.com/extensions/latest/ext_replicator.html)
- [36] —. (2025) Perception data generation (replicator) — isaac sim. Accessed: 2025-08-18. [Online]. Available: [https://docs.isaacsim.omniverse.nvidia.com/latest/replicator\\_tutorials/index.html](https://docs.isaacsim.omniverse.nvidia.com/latest/replicator_tutorials/index.html)
- [37] S. Borkman *et al.*, "Unity perception: Generate synthetic data for computer vision," *arXiv preprint arXiv:2107.04259*, 2021. [Online]. Available: <https://arxiv.org/abs/2107.04259>
- [38] NVIDIA Corporation. (2025) Nvidia cosmos: Physical ai with world foundation models. Accessed: 2025-08-18. [Online]. Available: <https://www.nvidia.com/en-us/ai/cosmos/>
- [39] —. (2025) Predict2 model reference — cosmos. Accessed: 2025-08-18. [Online]. Available: <https://docs.nvidia.com/cosmos/latest/predict2/reference.html>
- [40] —. (2025) Predict2 quickstart guide — cosmos. Accessed: 2025-08-18. [Online]. Available: [https://docs.nvidia.com/cosmos/latest/predict2/quickstart\\_guide.html](https://docs.nvidia.com/cosmos/latest/predict2/quickstart_guide.html)
- [41] M. Liu. (2025) Nvidia makes cosmos world foundation models openly available to physical ai developer community. Accessed: 2025-08-18. [Online]. Available: <https://blogs.nvidia.com/blog/cosmos-world-foundation-models/>
- [42] NVIDIA Corporation. (2025) Why world foundation models will be key to advancing physical ai. Accessed: 2025-08-18. [Online]. Available: <https://blogs.nvidia.com/blog/world-foundation-models-advance-physical-ai/>
- [43] ComfyUI Authors. (2025) Nvidia cosmos predict2 — comfyui examples. Accessed: 2025-08-18. [Online]. Available: [https://comfyanonymous.github.io/ComfyUI\\_examples/cosmos\\_predict2/](https://comfyanonymous.github.io/ComfyUI_examples/cosmos_predict2/)
- [44] P. Saraiva, "Synthetic data manometry interpretation," Figshare, 2025, version 1.0, CC BY 4.0. [Online]. Available: [https://figshare.com/articles/figure/SyntheticData\\_ManometryInterpretation/29936768](https://figshare.com/articles/figure/SyntheticData_ManometryInterpretation/29936768)