Enhancing the Fairness and Performance of Edge Cameras with Explainable AI

Truong Thanh Hung Nguyen[1,2], Vo Thanh Khang Nguyen[1], Quoc Hung Cao[1], Van Binh Truong[1], Quoc Khanh Nguyen[1], Hung Cao[2]

[1] Quy Nhon AI, FPT Software, Vietnam
[2] Analytics Everywhere Lab, University of New Brunswick, Canada

January 2024

Table of Content

01 Research Question

Current problem of Human Detection models

While we are a security camera, namely akaCam, we find out that Multi-object tracking models (i.e. YOLOX (Ge et al., 2021), Bytetrack (Zhang et al., 2021)) meet problems with:

- Obscured bodies (Fig (a).)
- Physically disabled individuals (Fig (b).)

leading to vulnerability and unfairness of security cameras.

Due to their “black-box” nature, there is a call for debugging and improvement techniques to enhance the safety and fairness of AI human detection models on security cameras.

(a)

(b)

Current call for Explainable AI

BEFORE

Training data → ML process → Learned Function → Decision Cat: 0.76 → End-user

- Why did you do that?
- Why not something else?
- When do you succeed?
- When do you fail?
- When can I trust you?
- How do I correct an error?

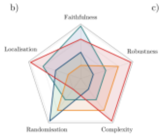
TODAY

Training data → New ML process → (Explainable) Learned Func. | Explainable decision ⇄ End-user

- I understand why
- I understand why not
- I know when you succeed
- I know when you fail
- I know when to trust you
- I know why you erred

TOMORROW

Evaluation



Training data → New ML process → (Explainable) Learned Func. | Explainable decision ⇄ End-user

Feedback

- Which explanation is better?
- How does this explanation mean?
- Can I trust the explanation?
- How can I further utilize the explanation?

XAI in Model Improvement on videos

1. “How can we enhance model performance based on the explanation?”
2. No study has proposed a framework for debugging human detection models.

We present a **debugging framework** for human detection models,
utilizing **XAI as a diagnostic tool** to identify problems
and enhance model fairness and performance.

02 Methodology & Experiment

Debugging Framework

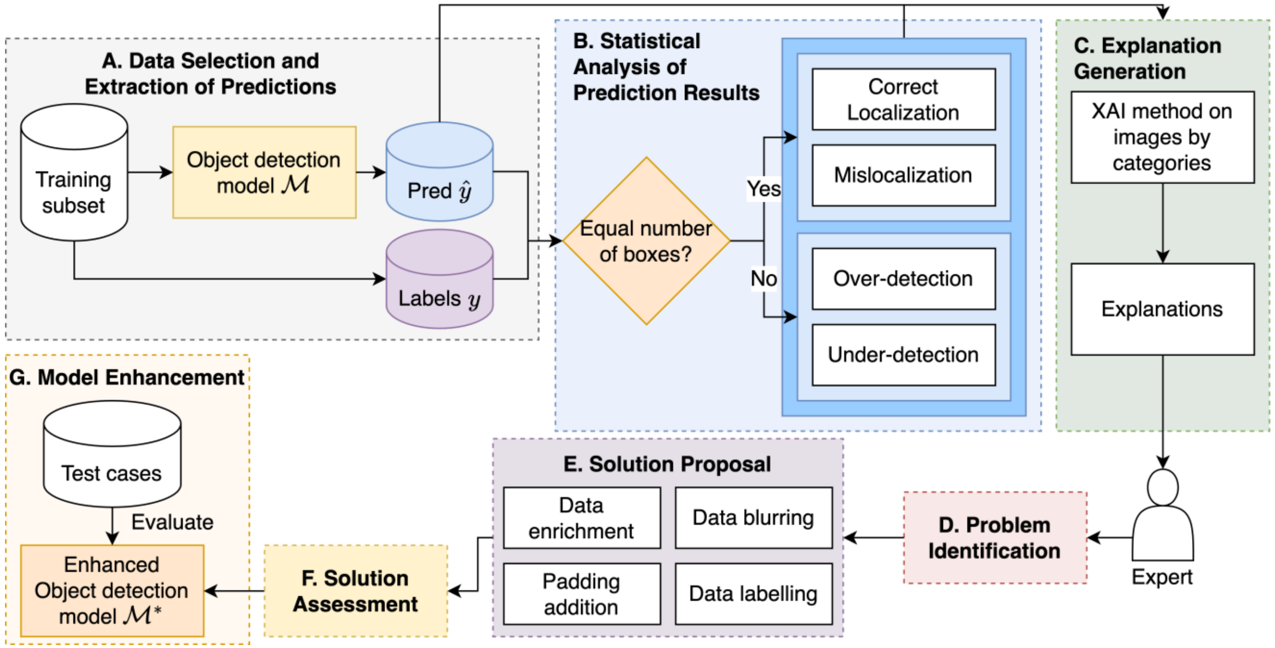
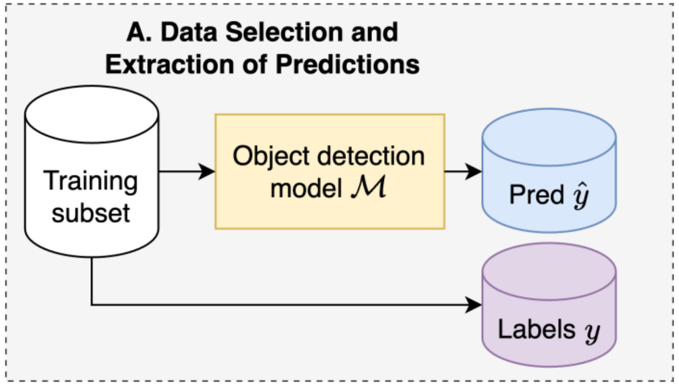


Fig 1. Our debugging framework for Human Detection models.

- In our experiment, we use a pre-trained Bytetrack model (Zhang et al., 2021) as the human detection model.
- The Bytetrack model consists of two components: YOLOX model for individual detection and Byte stage for processing detected boxes

Step 1 – Data Selection & Predictions Extraction



- This step involves selecting a subset of the training dataset to enhance the model.
- We divide CrownHuman dataset into 15000/4370/5000 for training/validation/test.
- Training and validation sets collectively contain 470K human instances.
- The data subset (random 1000 training samples) is input into the model to generate predictions, which are then analyzed and compared with ground truth.

Step 2 – Statistical Analysis of Prediction Results

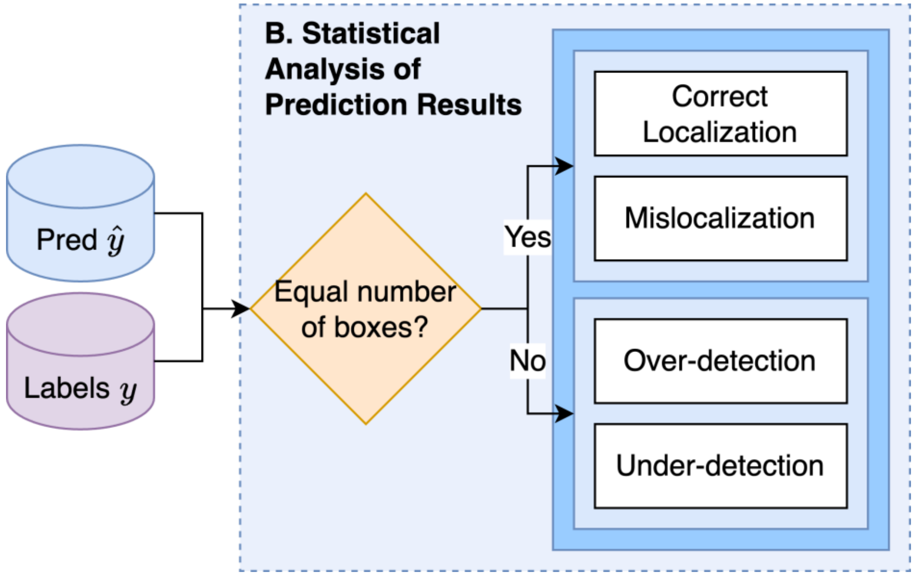


TABLE I
THE CATEGORIES OF 1000 IMAGES IN THE SUBSET.

Case	Number of images
Under-detection	855
Over-detection	17
Correct Localization	108
Mislocalization	20

After obtaining predictions from the model, they are systematically categorized based on being compared with ground-truth (GT) labels.

This classification is adjudicated by field experts, focusing on 4 categories:

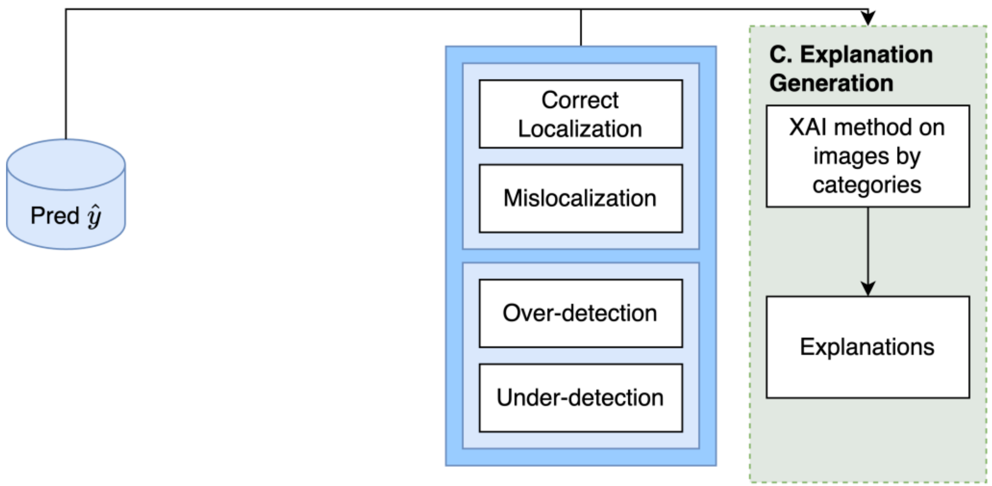
- Different number of detected people versus GT:

  Under-detection (855 images)**:** If the model detects fewer people than GT, otherwise, **Over-detection** (17 images).

- Same number of detected people versus GT:

  Correct Localization (108 images): all box pairs have IoU (Intersection over Union) ≥ 0.5, otherwise, **Mislocalization** (20 images).

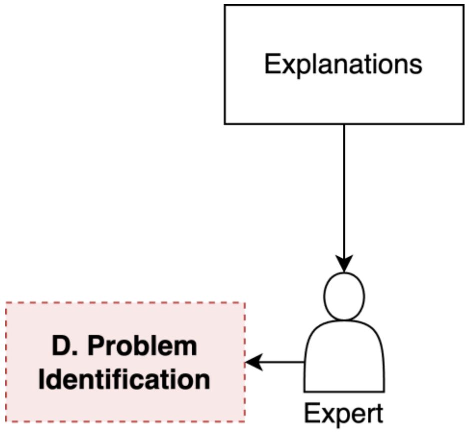
Step 3 – Explanation Generation



- In this step, an XAI method is applied to generate explanations for each category of images.
- We employ D-RISE (Petsiuk et al., 2020) to extract explanations for YOLOX, using the final box coordinates predicted by the Bytetrack model.
- These explanations help experts diagnose the cause of incorrect predictions in the next step.

Step 4 – Problem Identification

- Based on the XAI results, experts analyze specific categories identified in the statistical analysis to detect potential biases and errors.
- Fig. 3 indicate Bytetrack’s focus on entire human bodies, exposing its struggle to detect individuals showing only their heads.
- Hence, Bytetrack’s challenge in spotting partially visible humans (obscured bodies, physically disabled individuals) emerges as a key concern needing attention and resolution



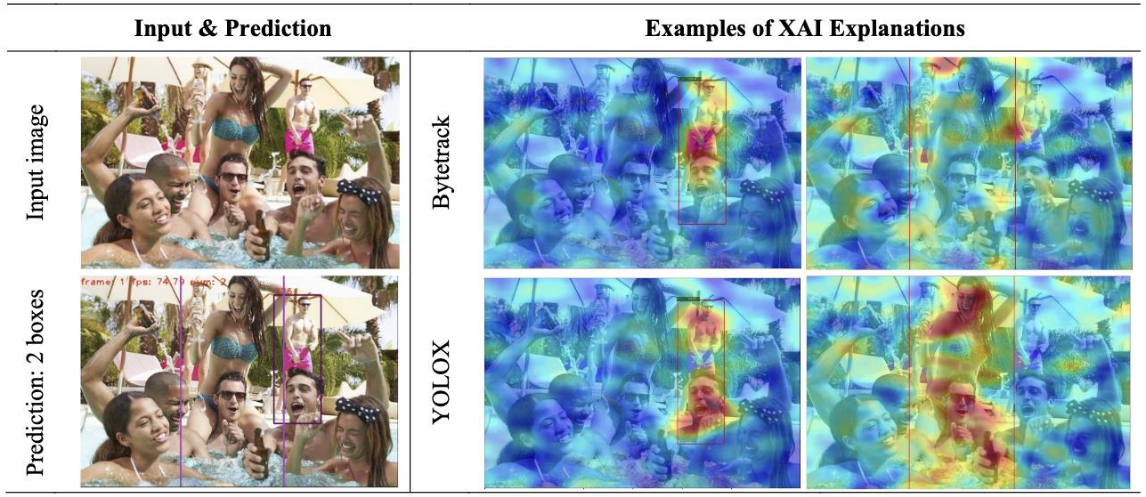
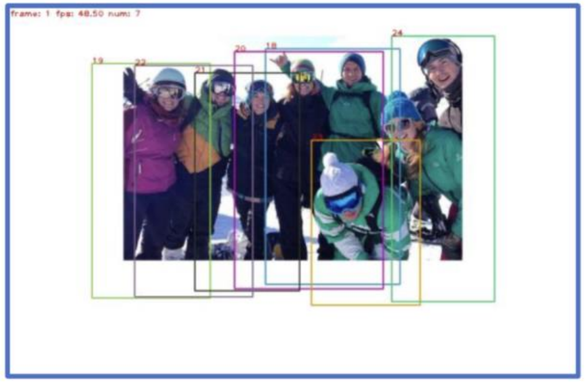


Fig. 3. Examples of XAI Explanations with Bytetrack and YOLOX model. In which, each image in the second column is the XAI Explanations for a corresponding box.

Step 5 – Solution Proposal



Labels box coordinates are outside the image

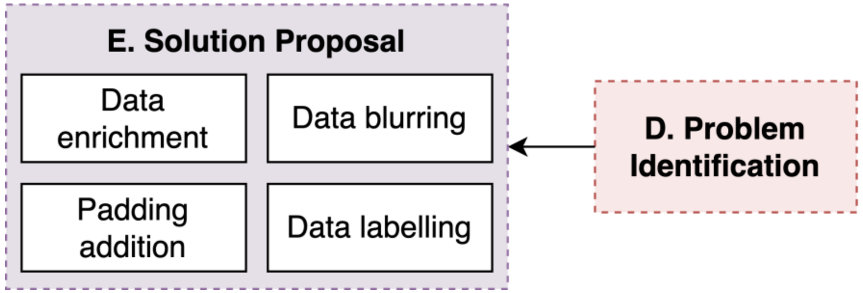
In this phase, experts analyze the dataset and model to identify causes of errors.

Dataset issue:

- Labels have body bias
- Labels box coordinates are outside the image

Bytetrack issue:

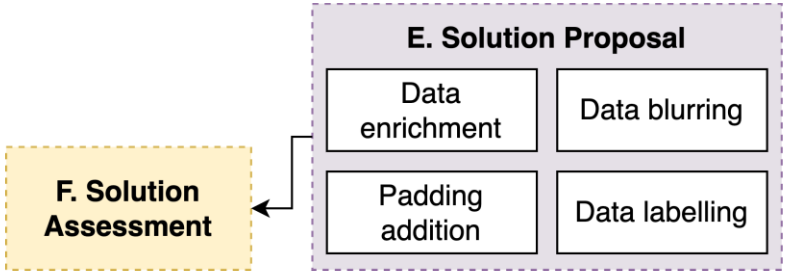
- Bytetrack tries to resolve occluded objects. For head-only images, Bytetrack expects an associated body.



Proposed solutions:

- **Data enrichment:** Add images with mostly obscured body sections.
- **Data blurring:** Based on XAI explanations, blur bodies to make the model focus on heads.
- **Padding:** Ensure bounding boxes are fully within images.
- **Relabeling:** Adjust bounding boxes to remain inside the image.

Step 6 – Solution Assessment



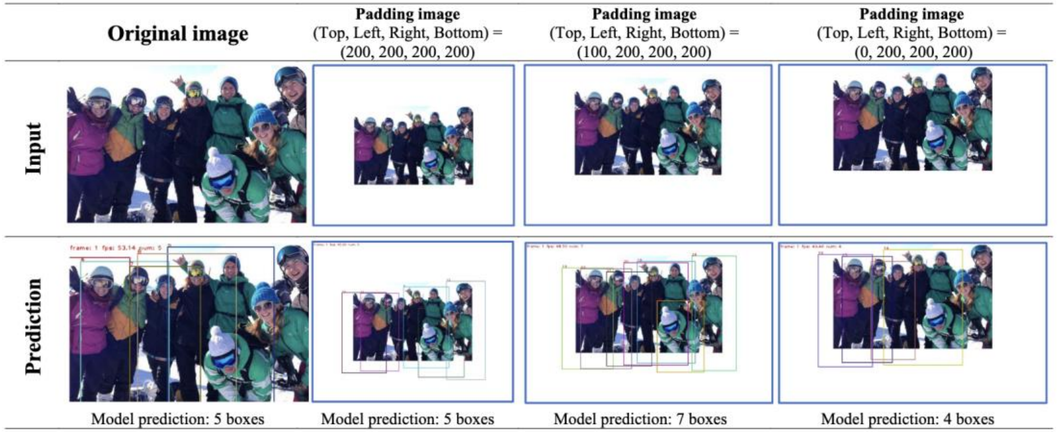


Fig. 5. Example of padding result. (Top, Left, Right, Bottom) = (100, 200, 200, 200) signifies padding of 100, 200, 200, and 200 pixels respectively on the top, left, right, and bottom.

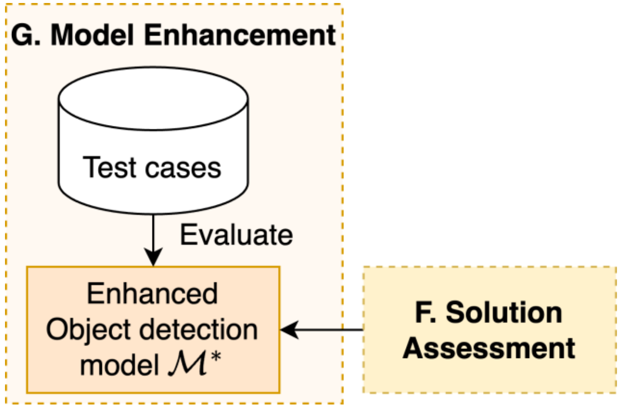
This step involves assessing the feasibility of proposed solutions.

The infeasible solutions are identified and eliminated, allowing for the selection of the most suitable solution.

- **Data enrichment:** The current dataset already has partly hidden figures, so more data might not help much.
- **Data blurring:** Effective for image classification but might not suit human detection where only humans are predicted.
- **Padding:** While sometimes effective, as in Fig. 5, it often fails, especially when objects obstruct people.
- **Relabeling:** Given dataset inconsistencies and variant model features, relabeling seems promising.

Following this analysis, **relabeling** emerges as the most impactful solution.

Step 7 – Model and Dataset Enhancement



- The chosen solution (**Relabeling**) is implemented to fine-tune the model by constraining bounding box coordinates within the image dimensions on CrownHuman dataset.
- Then, subsequent model refinement occurs over 10 epochs.
- The improvement is evaluated through a comparative analysis of the model's performance before and after fine-tuning.

03 Results

Training set Testing

- We test a 1000-image subset after refining the model.
- Both quantitative and qualitative evaluations are made against the original model, as seen in Table II and Fig. 4.
- The updated model better localizes in 855 “Under-detection” images, improving by 21 cases.

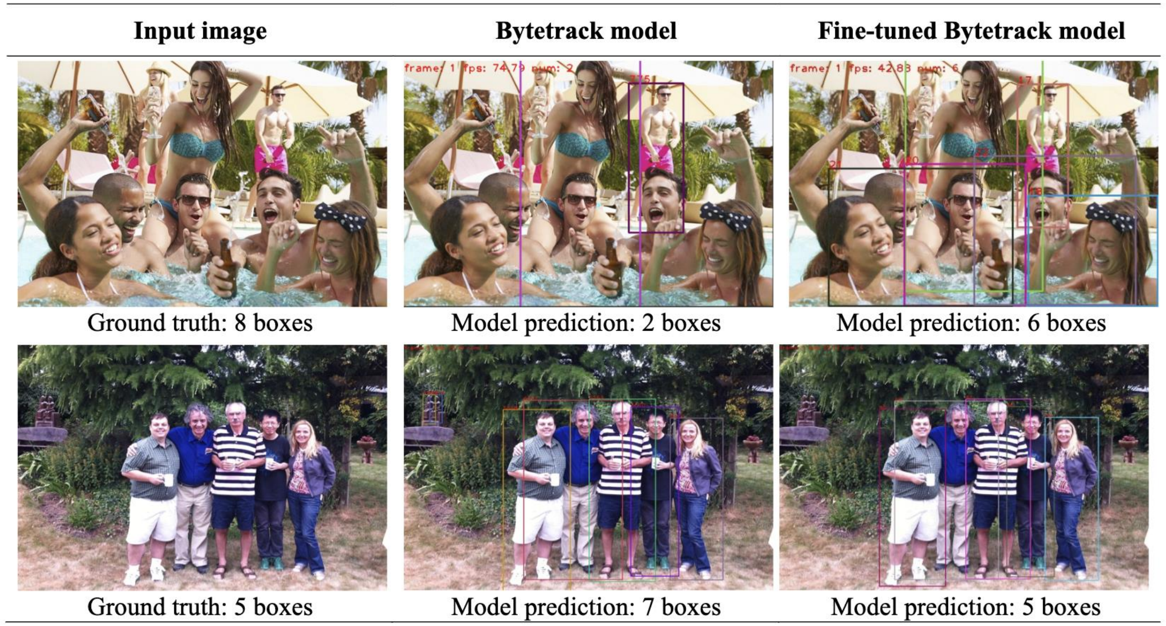


Fig. 4. Predictions of the Bytetrack model before and after fine-tuning.

TABLE II

STATISTICAL RESULT PRE-TRAINED MODEL VERSUS FINE-TUNED MODEL. THE ARROW ↑/↓ INDICATES THE HIGHER/LOWER VALUE, THE BETTER. THE BOLD INDICATES THE BETTER RESULT.

Case	Pre-trained model	Fine-tuned model
Under-detection (↓)	855	**834**
Over-detection (↓)	17	**13**
Correct Localization (↑)	108	**133**
Mislocalization (↓)	20	20

Images of Disabled Individuals & Partially obscured individuals in security footage

The refined model exhibits enhanced detection capabilities on images of physically disabled individuals.

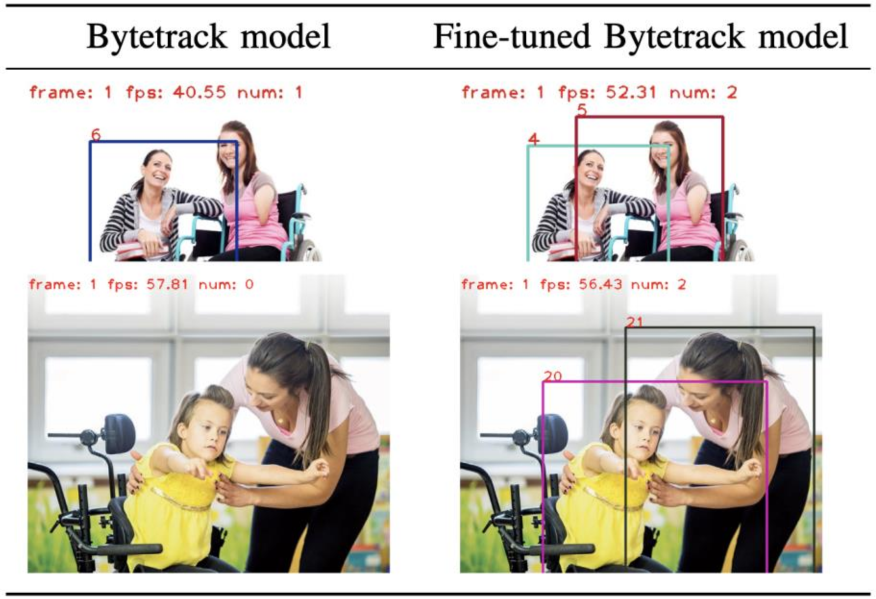


Fig. 6. Model's prediction on physically disabled person images. After fine-tuning, the model performs better than the original pre-trained model.

We assess the model in real-life contexts, like office security footage where people might be partly hidden. Post-refinement performance shows improvement.

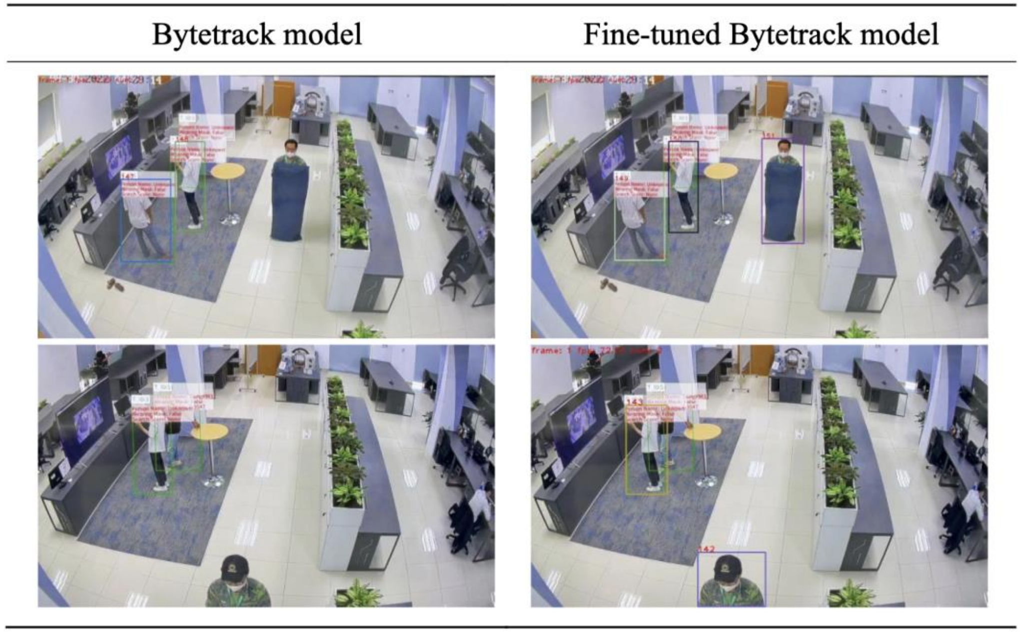
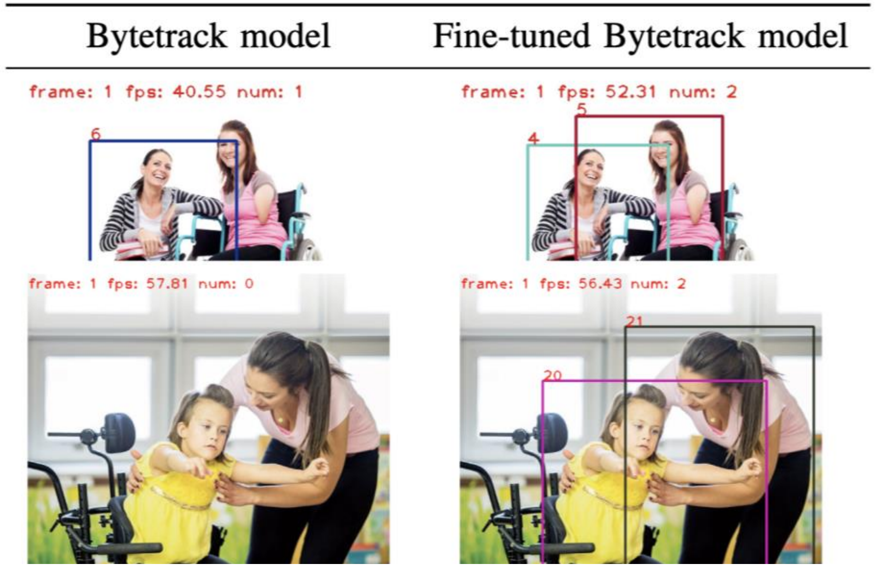


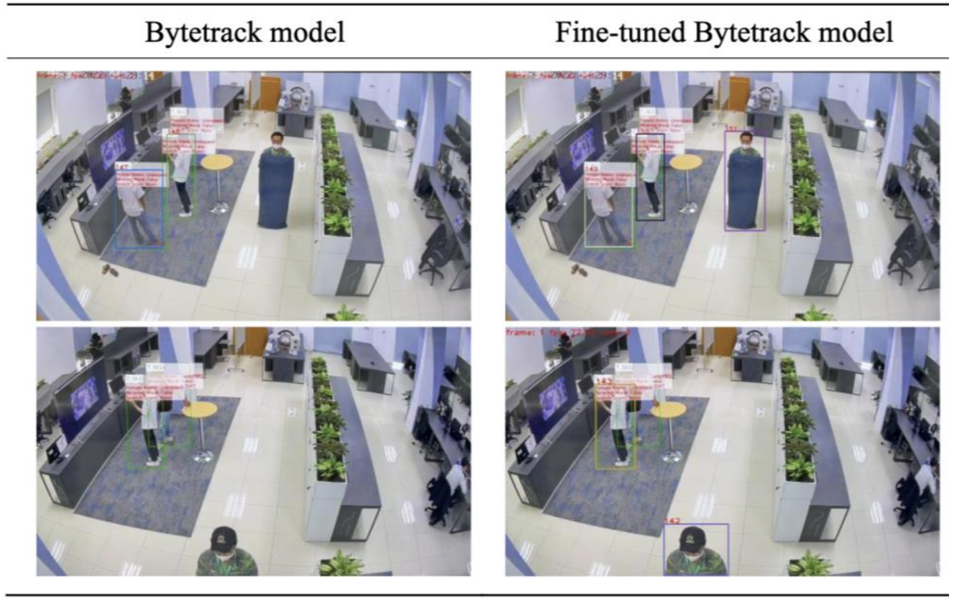
Fig. 7. Model's prediction on a security camera. The fine-tuned model performs better than the original pre-trained model detecting covered people.

04 Conclusion & Future Work

Our contribution & Future Work

- We propose a debugging model framework with XAI for the human detection problem.
- Based on the XAI explanations, the experts can identify problems and propose solutions to improve the model and the dataset.
- The problem of unfairness and under performance of the Bytetrack model is in data labeling, showing that the label can make the model biased.
- Our framework can be extended to other object detection problems that require focused consideration of specific classes.





References

Petsiuk, V., Jain, R., Manjunatha, V., Morariu, Vlad I, Mehra, A., Ordonez, V., & Saenko, K. (2020). Black-box Explanation of Object Detectors via Saliency Maps. ArXiv.org. https://arxiv.org/abs/2006.03204

Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2021). ByteTrack: Multi-Object Tracking by Associating Every Detection Box. ArXiv.org. https://arxiv.org/abs/2110.06864

Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). YOLOX: Exceeding YOLO Series in 2021. ArXiv.org. https://arxiv.org/abs/2107.08430