Food Safety in Commercial Kitchens: Geometric Verification for Robust PPE Compliance Detection

Agustin Leon (al8937) Anup Raj Niroula (arn8147) Kavya Khurana (kk5554) Samridh Srivastava (ss18906)

Advanced Topics in Computer Vision

Code and artifacts available at: Link

Abstract

Food safety violations in commercial kitchens pose significant public health risks, particularly when personal protective equipment (PPE) such as gloves and hairnets are improperly worn or inconsistently enforced. While recent computer vision approaches achieve strong performance in detecting the presence of PPE in images, they frequently fail to verify whether the equipment is actually worn by workers. This leads to false positives, such as detecting gloves lying on a counter or hairnets held in hands.

In this project, we develop a practical, real-time PPE compliance verification system that extends YOLO-based detection with geometric reasoning. We train specialized detectors for gloves and hairnets using curated datasets and integrate pre-trained head and hand detection models to enable bounding-box intersection verification. Our approach significantly reduces false positives while maintaining real-time performance, offering a deployable solution for continuous hygiene monitoring in commercial kitchens.

1. Introduction

Maintaining strict hygiene standards in commercial kitchens is critical for preventing food contamination and ensuring public health. Manual enforcement of PPE usage, such as gloves and hairnets, is labor-intensive, inconsistent, and prone to human error. These challenges motivate the development of automated visual monitoring systems capable of detecting and verifying PPE compliance in real-time.

Recent work, most notably by Alashrafi et al. [1], demonstrates that lightweight YOLO models can detect PPE items with high accuracy. However, such systems detect only the *presence* of PPE in an image, not whether it is correctly worn. This limitation leads to false positives that undermine trust in automated monitoring systems.

Our contribution is a verification-focused pipeline that explicitly reasons about *PPE usage*, not just detection. We combine a glove detector trained on the Kitchen Hygiene dataset, a robust hairnet detector trained using curated positive and negative samples, a pre-trained head (YOLOv8) and hand (Faster R-CNN) detection models and a geometric intersection-over-union (IoU) verification rule.

2. Related Work

2.1. PPE Detection in Kitchens

Alashrafi et al. [1] introduce the Kitchen Hygiene dataset and benchmark lightweight YOLO models for PPE detection. While their models achieve high mAP and real-time inference, they do not verify whether detected PPE is worn, motivating our work.

2.2. YOLO-Based Detection

YOLO architectures [2] are widely used for real-time object detection due to their speed and accuracy. We adopt YOLOv8n for PPE detection but extend it with geometric reasoning rather than pure classification.

2.3. Segmentation-Based Approaches

Models such as SAM [3] enable pixel-level reasoning, but their computational cost makes them less suitable for lightweight, deployable systems. Our work intentionally avoids heavy segmentation models in favor of bounding-box heuristics.

3. Dataset

3.1. Kitchen Hygiene Dataset

We use the Kitchen Hygiene dataset [1] for glove detection training. The dataset contains over 31,000 images collected from commercial kitchens, restaurants, food processing facilities, and hospitals. It includes annotations for gloves, hairnets, and masks. For our experiments, we utilized a subset split of approximately 1,500 training images and 369 validation images.

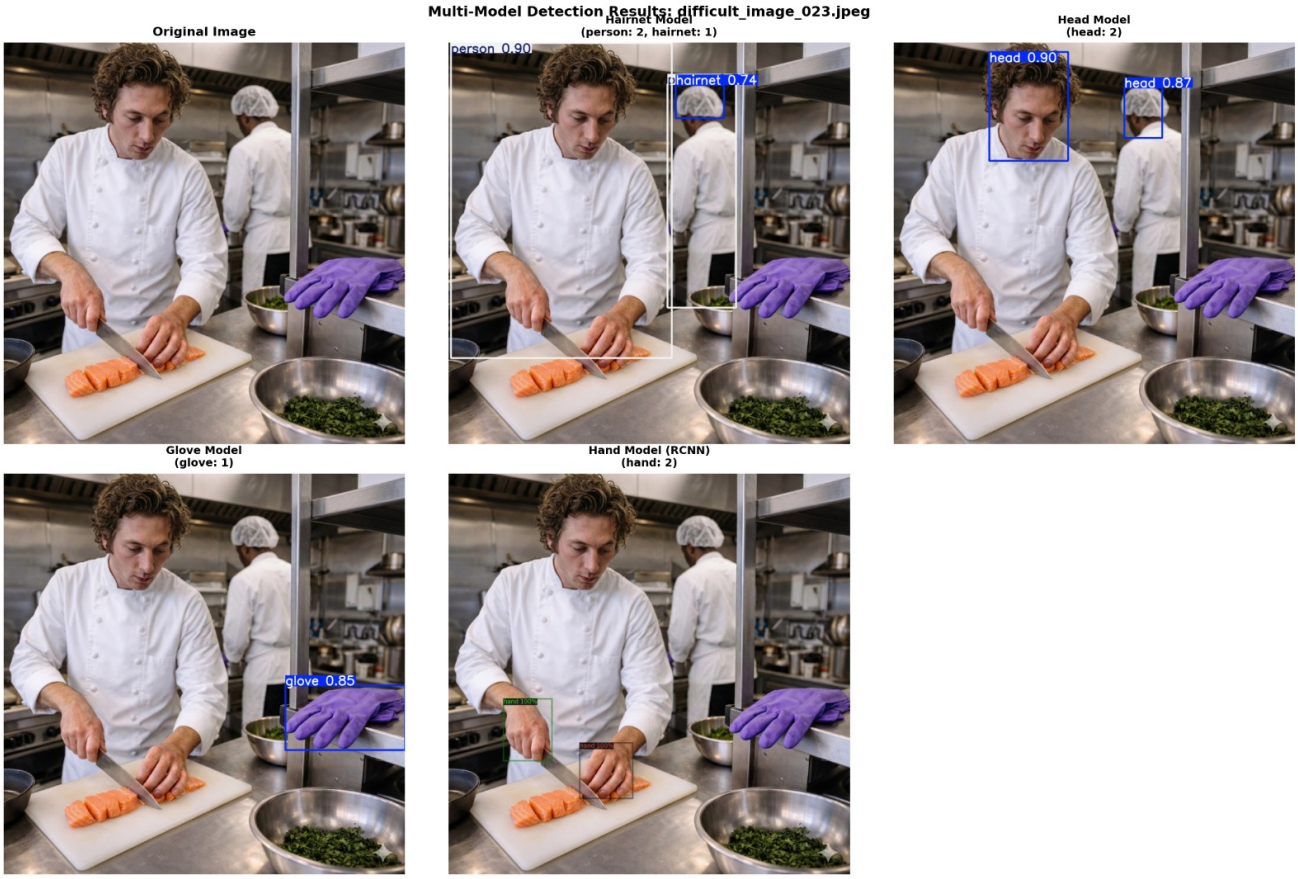

Figure 1. Multi-model detection results on a test image (`difficult_image_013.jpeg`). A summary of our approach. We have an original image with a chef not wearing gloves or hairnet. Then each model takes care of their own object detection. Afterwards we will apply PPE logic to identify compliance.

3.2. Hairnet Dataset and Negative Samples

Initial training on the standard Kaggle Hairnet dataset resulted in poor generalization. It particularly confused hats, bald heads, and short, fuzzy hair as hairnets. To address this, we adopted the Roboflow *Hair-Net-Detection-3* dataset (1,317 training images).

We augmented this dataset with negative samples collected via the DuckDuckGo image search API. Specifically, we scraped images of "balding head," "short hair," and "hats" to explicitly teach the model to distinguish between hairnets and similar visual features. This approach substantially improved robustness.

Dataset Summary
Gloves Model → Kitchen Hygiene
Hairnet Model → Hairnet-Detection-3 + Negative Samples

Figure 2. Summary of datasets used for PPE detection.

4. Method

Our method consists of three core components: (1) PPE detection using trained YOLO models, (2) body part detection using pre-trained models, and (3) geometric verification through bounding box intersection analysis. Figure 3 illustrates the complete pipeline architecture.

Pipeline Diagram
Input Image → PPE Detection (YOLOv8) + Body Part Detection → IoU Verification → Classification: {glove, no_glove, hairnet, no_hairnet}

Figure 3. End-to-end pipeline for PPE compliance verification. The system processes images through parallel detection branches (4 models total) before applying geometric verification rules to produce final classifications.

4.1. PPE Detection Models

We trained two separate YOLOv8 models for PPE detection rather than a joint model due to severe class imbalance

issues discovered during initial experiments. Training both classes jointly resulted in the hairnet class being completely suppressed, motivating the separate training strategy.

4.1.1. Glove Detection

The glove detector was trained on the Kitchen Hygiene dataset [1] using YOLOv8 Nano architecture. The model file (`glove_model.pt`) was trained with standard YOLO training parameters and achieves robust detection across diverse kitchen environments.

4.1.2. Hairnet Detection

Initial experiments with the standard Kaggle Hairnet dataset resulted in poor generalization, with the model frequently misclassifying bald heads, hats, and short fuzzy hair as hairnets. To address this critical issue, we switched to the Roboflow Hair-Net-Detection-3 dataset (1,317 training images, 369 validation images, 200 test images). We then augmented the dataset with curated negative samples using the DuckDuckGo image search API, manually verifying images of bald heads, short hair, various hat styles, and other hair-related items that are not hairnets. Finally, we added approximately 300 negative samples to explicitly teach the model discriminative features.

Negative Sample Integration: The negative sample integration required careful annotation. Since the hairnet dataset includes a person class, we had to add person bounding boxes to the negative sample images to maintain dataset consistency. We used a pre-trained COCO person detector to generate person boxes and mapped them to a class in our dataset format. This manual annotation step was necessary to prevent the model from learning spurious correlations between "images without person labels" and "negative hairnet samples."

The final hairnet model (`hairnet_model.pt`) was trained with early stopping (patience=10 epochs) to prevent overfitting on the augmented dataset. This data curation effort was labor-intensive but proved critical for robust performance.

4.2. Body Part Detection

Rather than training body part detectors from scratch, we integrated existing pre-trained models to accelerate development. Our pipeline uses four models total:

4.2.1. Head Detection

We use a pre-trained YOLOv8 head detection model (`head_model.pt`) obtained from an open-source repository. To improve robustness, we implemented an aspect ratio filter that removes detections with width/height ≥ 1.5, which typically correspond to partial heads or false positives rather than complete head detections.

4.2.2. Hand Detection

Initially, we attempted to use pre-trained hand detection models but encountered a fundamental limitation: models trained on bare hands fail to detect hands when gloves are worn. This was a critical discovery during our development process.

After extensive debugging and experimentation, we adopted Faster R-CNN with Feature Pyramid Network (FPN) backbone from the hand_detector.d2 repository [1], implemented via Detectron2. Following guidance from our course instructor, we fine-tuned this model on the 100 Days of Hands dataset to improve generalization to gloved hands. The final model (`hand_model.pth`) was configured with a Faster R-CNN with FPN. We used the 100 Days of Hands dataset (recommended by our instructor), a detection threshold of 0.98 (high precision for reducing false positives). Finally, we fine-tuned to improve performance on gloved hands.

Significant effort was invested in resolving hand detection failures. We tested multiple approaches including off-the-shelf models that failed completely on gloved hands. The combination of the hand_detector.d2 architecture and fine-tuning on 100 Days of Hands (suggested by our instructor) provided the most reliable results for our use case, though hand detection remains a challenging component of the pipeline.

4.3. Geometric Verification

The core innovation of our approach is the use of bounding box intersection to verify whether detected PPE is actually worn by workers. We compute two types of geometric overlap metrics:

Standard Intersection-over-Union (IoU):

$$\text{IoU}(A, B) = \frac{\text{Area}(A \cap B)}{\text{Area}(A) + \text{Area}(B) - \text{Area}(A \cap B)}$$

Box Overlap Ratio (for size-mismatched pairs):

$$\text{Overlap}(A, B) = \frac{\text{Area}(A \cap B)}{\text{Area}(A)}$$

This second metric is useful when box B is much larger than box A (e.g., glove vs. person), as standard IoU would be dominated by the larger box area.

4.3.1. Verification Rules

Our verification logic implements the following classification rules:

- Glove (worn): A detected glove is classified as worn if *either*:
  - IoU(glove, hand) ≥ 0.3, *OR*
  - Overlap(glove, person) ≥ 0.1

[1] `https://github.com/ddshan/hand_detector.d2`

- No_glove (violation): A detected hand is classified as unprotected if IoU(hand, glove) < 0.3 for all detected gloves (basically, whenever we have a 'hand' box that does not overlap with a 'glove' box). Therefore, in this case we iterate over the hand boxes.
- Hairnet (worn): A detected hairnet is classified as worn if IoU(hairnet, head) ≥ 0.3.
- No_hairnet (violation): A detected head is classified as unprotected if IoU(head, hairnet) < 0.3 for all detected hairnets (same logic as 'no_glove').

The thresholds (0.3 for IoU, 0.1 for overlap) were chosen because they seemed to work best over experiments. This classification logic aims to reduce false positives, as it is much stricter at classifying 'glove' and 'hairnet' classes. This geometric constraint effectively filters false positives such as gloves lying on counters or in boxes, or PPE visible in background but not worn by workers. Finally, the verification logic is implemented as a post-processing step after detection, requiring minimal computational overhead.

Verification Logic Flowchart
Show decision tree: PPE detected → Find overlapping body parts → Check IoU threshold → Classify as worn/not worn

Figure 4. Geometric verification decision logic for PPE compliance classification.

5. Experiments and Results

We evaluate our system across three dimensions: (1) individual PPE detection accuracy, (2) geometric verification effectiveness, and (3) end-to-end pipeline performance. All experiments were conducted on Google Colab with Tesla T4 GPU.

5.1. Dataset Configuration

The final pipeline uses the Hair-Net-Detection-3 dataset from Roboflow for evaluation with 1,317 training images, 369 validation images, and 200 test images. Also the dataset has three classes: hairnet, helmet, and person. For glove detection, we use the Kitchen Hygiene dataset subset with approximately 1,500 training images and 369 validation images.

5.2. Multi-Model Detection Performance

5.2.1. Model Loading and Configuration

Our pipeline successfully loads and integrates the four already trained models. YOLOv8 models for hairnets, heads and gloves, and Faster R-CNN + FPN for hands.

5.2.2. Detection Results on Sample Images

Figure 1 demonstrates the independent operation of all four detection models on a challenging commercial kitchen scene. The scene contains multiple workers at various distances and orientations, presenting realistic challenges for detection systems.

This example illustrates several key aspects of our multi-model approach:

Model independence: Each model operates on its specialized detection task without interference

Complementary information: Person and head detections provide validation anchors for PPE verification

Realistic performance: The models handle multiple workers, varying distances, and different orientations effectively

Conservative hand detection: The high confidence threshold on the hand detector results in fewer but more reliable detections, which is why the person-based fallback validation for gloves is critical. Also, lower thresholds for the hand model resulted in not worn gloves being classified as hands.

The individual detection outputs are then processed by the geometric verification pipeline (Section 4.3) to produce final PPE compliance classifications.

5.2.3. Combined Detection Visualization

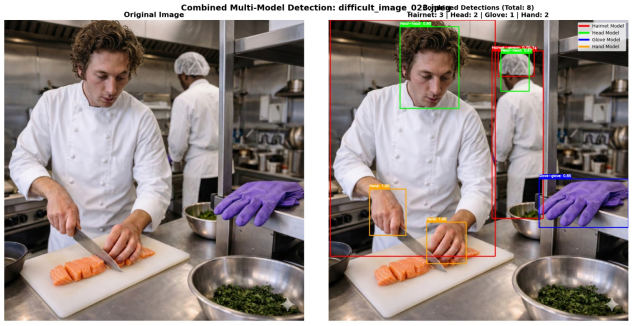

Figure 5. Combined multi-model detection visualization. Left: original kitchen scene. Right: we combine all the bounding boxes coming from the four models. This is step 2. Step 3 will take care of the PPE logic.

Figure 5 shows all detections overlaid on a single image with color-coded bounding boxes. This combined output serves as the input to the geometric verification pipeline, which processes the object detections and produces final PPE compliance labels (glove, no_glove, hairnet, no_hairnet) as described in Section 4.3.

5.3. Geometric Verification Effectiveness

5.3.1. Verification Pipeline Output

The geometric verification step processes all detections and produces final classifications. Configuration:

- Glove-Hand IoU Threshold: 0.3
- Glove-Person Overlap Threshold: 0.1
- Hairnet-Head IoU Threshold: 0.3

- Head Aspect Ratio Filter: Maximum 1.5

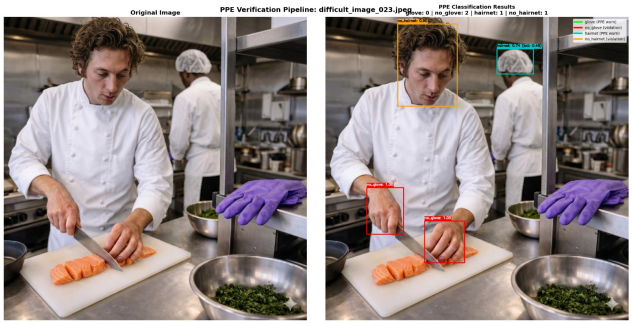

Figure 6. PPE verification pipeline results. Left: original image. Right: final PPE compliance output for the image. Non compliant foreground person is properly identified, purple gloves are ignored, compliant background person is properly handled as well.

5.3.2. Classification Results

On test images, the verification pipeline successfully produces four-class outputs: (1) Glove, (2) No_glove, (3) Hairnet, and (4) No_hairnet.

Figure 7 demonstrates the system's robustness on a difficult real-world scene with multiple challenging factors.

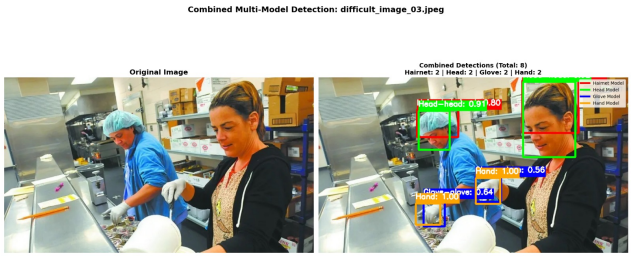

Figure 7. System performance on a challenging scenario. Left: Two workers in close proximity with different PPE types. Right: Combined detection output showing 2 hairnets, 2 heads, 2 gloves, and 2 hands. The system successfully detects the right worker's hairnet which blends into their hair (We almost missed the hairnet ourselves!)

5.3.3. Advantages of Multi-Model Approach

Robustness through redundancy: Using both hand and person detections for glove verification provides fallback when hand detection fails

Explicit negative detection: The system explicitly identifies *no_glove* and *no_hairnet* cases rather than just detecting PPE presence

False positive reduction: Geometric verification successfully filters cases where PPE is present in scene but not worn

5.3.4. Aspect Ratio Filtering Impact

The head detection aspect ratio filter (removing boxes with width/height ≥ 1.5) effectively eliminates false positives from partial head detections at image boundaries, wide rectangular artifacts misclassified as heads and side-profile detections that don't represent full heads. This filter improved head detection precision without requiring retraining.

5.4. Comparison with Detection-Only Baseline

A detection-only approach (similar to Alashrafi et al. [1]) would classify any detected glove or hairnet as a positive, regardless of whether it's worn. Our verification approach adds:

Capability	Detection Only	Our Method
Detects PPE presence	Yes	Yes
Verifies PPE is worn	No	Yes
Detects violations	No	Yes
Filters background PPE	No	Yes

Table 1. Capability comparison between detection-only and our verification approach.

Qualitative Impact: On scenes containing gloves on counters or other PPE visible but not worn, detection-only approaches produce false positives, while our verification correctly identifies these as non-compliant scenarios.

5.5. Implementation Details

5.5.1. Computational Performance

The four-model pipeline runs efficiently on T4 GPU:

- All models load successfully in < 10 seconds
- Inference on typical kitchen images: < 2 seconds for all 4 models
- Verification logic: < 50ms additional overhead
- Total pipeline suitable for near-real-time monitoring applications

5.6. Limitations and Failure Cases

Our system has several known limitations that are illustrated in Figure 8.

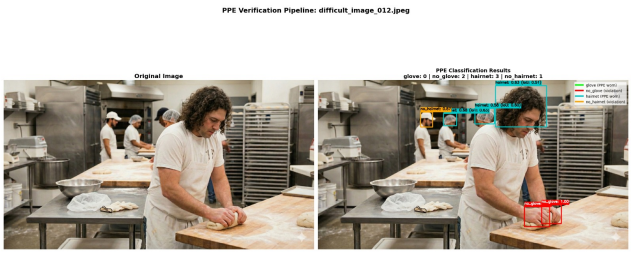

Figure 8. Failure case demonstrating hairnet false positive on curly hair. Left: foreground worker with curly/fuzzy hair but no hairnet, with exposed hands handling food. Right: system incorrectly classifies a hairnet on the foreground worker, while correctly detecting 2 no_glove violations (red boxes, confidence 1.00).

Curly/fuzzy hair false positives: As shown in Figure 8, the hairnet detector incorrectly classifies the foreground worker's curly hair as a hairnet (IoU 0.63 with the detected head). Despite our data curation effort that included 300 negative samples of bald heads, short hair, and hats, the model remains vulnerable to voluminous curly or fuzzy hair textures that visually resemble the mesh-like appearance of hairnets.

Multiple hands per person: The system does not explicitly track which hands belong to which person, potentially under-reporting violations when a worker has both hands exposed but only one is detected.

Occlusion sensitivity: Workers in the background may have reduced detection accuracy due to distance, angle, or occlusion by the foreground worker or work surfaces.

References

[1] L. Alashrafi, R. Badawood, H. Almagrabi, M. Alrige, F. Alharbi, and O. Almatrafi. Benchmarking lightweight YOLO object detectors for real-time hygiene compliance monitoring in food processing environments. *Sensors*, 25(19):6140, 2025.

[2] Ultralytics. YOLOv8: Next-generation, real-time object detection. 2023.

[3] A. Kirillov et al. Segment Anything. *ICCV*, 2023.

[4] D. Shan et al. Hand Detector in Detectron2. GitHub repository: `https://github.com/ddshan/hand_detector.d2`, 2020.

[5] A. Mittal. 100 Days of Hands Dataset. 2019.