In this supplement, we first discuss visualization design alternative (Appendix A). We then provide the complete prompt templates used for GPT and LLaVA in the process of explanation generation (Appendix B). We also provide all slices and their explanations for the four detectors (car, dog, chair, and person) evaluated in our work, as well as a sample of individual detection explanations from which the slice explanations are derived (Appendix C). We include details on clustering in slice discovery (Appendix D), filtering techniques in slice explanations (Appendix E), experimental and implementation details (Appendix F), and Use Case 2 (Appendix G). We also include expert ratings on *VISLIX* (Appendix H) and a detailed comparison with *AttributionScanner* (Appendix I).

#### Appendix A: Visualization Design Alternative



Figure 1: A 2D density plot based on kernel density estimate.

Fig. 1 presents a design alternative to the 2D histogram-based density plot (View B1) that we have explored during VISLIX's development. This version, based on a Gaussian 2D kernel density estimate, provides a continuous and smoothed representation of the data distribution. We initially chose this approach for its visual smoothness, compared to the discrete blocks of a 2D histogram. However, we found that this design could be misleading, as it may still render regions without actual instances, which is particularly problematic when using brushing operations. The iteration indicates that, when addressing domain-specific questions, rule-of-thumb design principles might be misaligned with expert needs, underscoring the importance of continuous feedback throughout the design process to ensure accurate and effective visual communication.

#### Appendix B: Prompt Templates

**Individual explanations.** Fig. 2 shows the prompt template for asking a question to LLaVA about an image, which is used in answering both predefined questions and questions generated by GPT. Fig. 6 shows the template used to prompt GPT to ask a question about the context region of a False Positive, bridging the GPT-LLaVA conversation. We start by describing the detection background using answers to predefined questions, then task GPT with

© 2025 The Author(s). Computer Graphics Forum published by Eurographics and John Wiley & Sons Ltd. generating a new question. Specifically, we organize several predefined questions, such as those about weather and lighting conditions, and their answers as a dialog before the task description, serving as in-context learning examples. Once GPT generates a new question, we send it along with the context region to LLaVA to collect the answer. We then append the question and answer to the end of the dialog and prompt GPT to generate another question. This iterative process continues for 10 rounds or until GPT outputs "STOP". This template applies to False Negatives after slight modifications, as shown in Fig. 7.

After collecting answers to predefined and GPT-generated questions, we use the content from Fig. 6 to prompt GPT to generate the individual explanation (Fig. 3).

**Slice explanations.** Fig. 4 displays the prompt template for GPT to generate a slice explanation, based on individual explanations of FPs and FNs within this slice. Similarly, Fig. 5 exhibits the prompt of generating slice keywords using the same input but different tasks.

Appendix C: Slice and Individual Detection Explanations

**Slice explanations.** We provide all slices and their explanations generated by our framework corresponding to four object detectors analyzed in the expert study. These include the car detector, dog detector, chair detector, and person detector. Specifically, for each detector, each slice showcases 10 random samples of False Positives (FP) and False Negatives (FN), accompanied by its textual explanations, as with the format used in the expert study.

**Individual detection explanations.** We also present sample explanations for individual FPs and FNs from the four object detectors. Explanations that are clearly incorrect are highlighted in red.



Figure 2: Prompt template used for LLaVA to answer a question about an image.

**TASK:** Based on the provided information above, summarize the scenario in one sentence, then explain the most possible reason for the False Positive [*object\_name*] detection step by step. Limit to 70 words.

OUTPUT:

**Figure 3:** Task description used for GPT (Interpreter) to explain a False Positive detection.



Figure 4: Prompt templates of GPT (Summarizer) for creating slice explanations.

**INSTRUCTION:** In a [*object\_name*] detection model, a set of False Positives and False Negatives occurs. For each False Positive and False Negative, you are given sentences describing the scene and the error in a single row. As a Computer Vision expert, you concisely summarize these errors for those who haven't observed these pictures. In your output, you first describe the common feature of these environments, then identify the most primary reason of False Positives and False Negatives.

- You think step by step.

- Be concrete and explicit.

- You ONLY output the most likely reason and ignore other reasons. - Be succinct and limit your output within 70 words.

INPUT:

[Individual explanations for False Positives] [Individual explanations for False Negatives]

OUTPUT:

Figure 5: Prompt templates of GPT for creating slice keywords.

#### Appendix D: Details on Clustering in Slice Discovery

Our slice discovery method identifies data slices by first applying dimensionality reduction using UMAP, followed by clustering with HDBSCAN. The UMAP parameter setup largely follows prior work [MSRPC21, SDA\*20], including setting the minimum distance between points to 0 for a more precise representation of the data. However, McConville et al. [MSRPC21] sets the number of dimensions, k, to the number of distinct labels in the dataset, which is not applicable to our data. Instead, we have experimented with kvalues ranging from 2 to 10 and have observed no significant differences in the resulting slices when applying clustering methods. To preserve as much information as possible, we empirically set k to 10. Additionally, we fine-tune two parameters via grid search: the number of neighbors considered in UMAP and the minimum number of points required to form a core point in HDBSCAN. The optimal parameter combination is selected using the Silhouette metric [Rou87] that is widely used to measure the clustering quality.

Appendix E: Centrality-Based Filtering for Slice Explanations

To improve the quality of slice explanations, we adopted a centrality-based strategy to filter out noisy individual explanations within a slice. These explanations often arise from VLM/LLM hallucinations or suboptimal parameter configurations during slice discovery. Our approach involves converting each individual explanation into a sentence embedding and selecting explanations closest to the embedding center until 80% are retained or the token limit of 2000 is reached. Our preliminary experiments show that this strategy produces higher-quality and more consistent slice explanations compared to random selection.

For instance, in Fig. 8, we present a slice of false positives (FPs) from a car detector along with explanations generated using centrality- and random-based strategies. This slice primarily involves buses, with some focusing on trucks. Images are sorted by descending centrality based on the sentence embeddings of their explanations. Additionally, we highlight the individual explanations of the 5 most central FPs (1–5) and the 5 least central FPs (19–23), which are filtered out by the centrality-based method. The most central explanations typically include bus misclassifications and are accurately described, while the least central ones often contain noise or errors—such as truck misclassification, excessively long explanations, or hallucinations, highlighted in grey text. The centrality-based method thus provides a more focused and concise slice explanation by describing primarily buses.

#### Appendix F: Experimental and Implementation Details

We conducted experiments related to the VLMs, including BLIP-2 and LLaVA, on a Linux system equipped with an Intel Core i7-5930K CPU (12 cores, 3.50 GHz), 128 GB of RAM, and one Nvidia Titan X GPUs with 12 GB of memory. The GPT3.5 API was accessed through the LangChain library (version 0.2). For the visual analytics system, we utilized JavaScript, D3 (version 7.8.5), Material UI (version 6.1.9), and the React framework(version 18.3.1) for the frontend, while the backend was developed using Python (version 3.9) and Flask (version 3.0.3).

As shown in Tab. 1, for each evaluated object, we report the total time spent on generating context-aware embeddings, performing slice discovery, and generating all individual explanations. All of these data are precomputed before being integrated into the visual analytics system. Only the slice explanation is generated interactively when users test hypothetical slices. On average, the time to generate a slice explanation is 1.70 seconds (SD = 0.25).

Object	Embeddings	Slice discovery	Individual explanations		
Dog	1 min, 39.12 s	4 min, 15.12 s	1 h, 4 min, 28.60 s		
Car	2 min, 59.49 s	4 min, 23.24 s	1 h, 58 min, 45.59 s		
Chair	6 min, 49.77 s	5 min, 52.80 s	4 h, 55 min, 36.25 s		
Person	17 min, 13.64 s	12 min, 45.40 s	14 h, 35 min, 52.04 s		

**Table 1:** The total runtime for context-aware embedding generation, slice discovery, and individual explanation generation for each dataset.

#### INSTRUCTION:

You are an expert in Computer Vision, specializing in analyzing the reasons behind False Positive (FP) detections in object detection models. You are familiar with key concepts in this domain, including bounding box, ground truth, among others.

Your expertise covers various factors contributing to False Positives, including but not limited to, occlusion, size, visibility, viewpoint, localization error, background, spurious correlation, missing or inaccurate ground truth, and so on.

In the context of a [object name] detection model, a False Positive detection of a [object name] has occurred, which is depicted by a bounding box. The size of the detection region is [width] x [height] pixels.

The description of the False Positive detection region is: < [0]-detection caption] >.

Only parts of the ground truth [object\_name] are detected, the detected parts are: < [Q6-intersection\_caption] >. (Apply to partial detection) The detection region fully encloses the ground truth [object name], a description of the ground truth is:< [groundtruth caption] >. (Apply to over detection)

Furthermore, there's a cropped image that encompasses the False Positive detection and its surroundings, which is described as: < [02context caption]>.

Given the information provided above, ask me questions about the cropped image to maximize your information of reasoning about the False Positive. Don't mention "False Positive" in your question. Carefully asking me specific and informative questions. Each time ask only one question without giving an answer. I'll put my answer beginning with "Answer:".

#### Q&A about the image:

Question: What is the weather in this image? Answer: [Q3-weather] Question: How is the lighting condition in this image? Answer: [04-lighting] Question: Is there any obstruction blocking the [object\_name]? If so, what is obstructing the [object\_name]? (Apply to IoU>0.2) Answer: [*Q5-obstruction*]

#### TASK:

Next Question(You can still ask [count] questions. Please output "STOP" when no questions is needed). Question:

	Baseline		Original-Train set		Embedding-Train set		CLIP Score-Train set	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Slice 1	0.5471	0.7881	0.6483	0.7966	0.6452	0.8475	0.6821	0.8729
Slice 2	0.6071	0.7083	0.6282	0.6806	0.6105	0.8056	0.6129	0.7917
Slice 3	0.5196	0.6883	0.5543	0.6623	0.5729	0.7143	0.5895	0.7273
Slice 4	0.3879	0.6	0.4565	0.56	0.3712	0.6533	0.3952	0.6533
Slice 5	0.3907	0.7763	0.381	0.7368	0.5214	0.8026	0.537	0.7632
Slice 6	0.5775	0.7069	0.6338	0.7759	0.6184	0.8103	0.6912	0.8103
Slice 7	0.4146	0.7556	0.4684	0.8222	0.5362	0.8222	0.5968	0.8222
Slice 8	0.4188	0.7273	0.3507	0.6727	0.4343	0.7818	0.4286	0.7636
All	0.6089	0.799	0.6071	0.7777	0.6304	0.8325	0.6551	0.8271
mAP	0.7542		0.7265		0.7878		0.7845	

Figure 6: Prompt template used for promoting GPT (Questioner) to raise a question for a False Positive detection.

Table 2: Metrics on slices (Precision and Recall) and model (Precision, Recall, and mAP) across baseline and three training sets (Original, Embedding, CLIP Score). Values below baseline are gray. The highest values for each row are blue.

#### Appendix G: Details on Use Case 2

In the second use case, we fine-tune the person detector for an additional epoch using the original training data (Original), training data retrieved from image embeddings of the selected eight slices (Embedding), and training data retrieved from the textual descriptions of those slices (CLIP Score). The results for the three conditions are shown in Tab. 2, where the baseline represents the detector before fine-tuning; metric values lower than the baseline are colored gray, and the highest values for each row are colored blue. For the eight slices' metric values, the Original condition shows the most declines, with 8 of 16 values decreasing. In contrast, the Embedding and CLIP Score conditions each have only 1 decrease. Besides, each condition possesses 3 (Original), 6 (Embedding),

to ensure the model's overall performance, reflected by the mAP value. The Original condition, despite having the 3 highest metric values, has an mAP of 0.7265, lower than the baseline of 0.7542. Yet, the Embedding and CLIP Score conditions achieve similar mAP values of 0.7878 and 0.7845, approximately 3% higher than baseline.

### Appendix H: Expert Ratings of VISLIX

The experts have also evaluated VISLIX using a 5-point Likert scale, shown in Fig. 9. The results reflect a positive overall impression of VISLIX.

and 10 (CLIP Score) highest metric values. Although higher precision or recall values are generally desirable, they must be balanced

© 2025 The Author(s). Computer Graphics Forum published by Eurographics and John Wiley & Sons Ltd.

#### **INSTRUCTION:**

You are an expert in Computer Vision, specializing in analyzing the reasons behind False Negative (FN) detections in object detection models. You are familiar with key concepts in this domain, including bounding box, ground truth, among others.

Your expertise covers various factors contributing to False Positives, including but not limited to, occlusion, size, visibility of parts (truncation), viewpoint, localization error, missing or inaccurate ground truth, and so on.

In the context of a [*object\_name*] detection model, a False Negative has occurred in an image, meaning that the ground truth [*object\_name*] isn't correctly detected by the model. The ground truth is depicted by a bounding box. The size of the ground truth is [*width*] x [*height*] pixels. The description of the ground truth [*object\_name*] is: < [*Q1-groundtruth\_caption*] >. Only parts of the ground truth [*object\_name*] are detected, the detected parts are: < [*Q6-intersection\_caption*] >. (Apply to partial detection) There is a False Positive bounding box fully enclosing the ground truth [*object\_name*], a description of the False Positive region is: :< [FP\_*caption*] >. (Apply to over detection)

Furthermore, there's a cropped image that encompasses the ground truth [*object\_name*] and its surroundings, which is described as: < [*Q2-context\_caption*] >.

Given the information provided above, ask me questions about the cropped image to maximize your information of reasoning about the False Negative. Don't mention "False Negative" in your question. Carefully asking me specific and informative questions. Each time ask only one question without giving an answer. I'll put my answer beginning with "Answer:".

 Q&A about the image:

 Question: What is the weather in this image?

 Answer: [Q3-weather]

 Question: How is the lighting condition in this image?

 Answer: [Q4-lighting]

 Question: Is there any obstruction blocking the [object\_name]? If so, what is obstructing the [object\_name]? (Apply to IoU>0.2)

 Answer: [Q5-obstruction]

#### TASK:

Next Question(You can still ask [count] questions. Please output "STOP" when no questions is needed). Question:

Figure 7: Prompt template used for promoting GPT (Questioner) to raise a question for a False Negative detection.

Experts evaluated the quality of slice explanations using a Likert scale based on their agreement with the statement: "The edge case explanation is correct for this group." Overall, 65% of explanations received agreement ("Agree" or "Str. Agree"), 21% were rated as "Neutral", and 14% were disagreed with ("Disagree" or "Str. Disagree").

#### Appendix I: Detailed Comparison with AttributionScanner

AttributionScanner [XOG\*25] similarly clusters image embeddings to identify slices but is specifically designed to uncover spurious correlations [XDLM24] in classification models. To ensure a fair comparison, we adapt VISLIX to AttributionScanner's second use case that analyzes a bird classification model (landbird/waterbird) using the Waterbirds dataset. Specifically, we extract context-aware embeddings based on bird bounding boxes and make minor adjustments to explanation prompts to better align with the classification task. We apply both approaches to identify slices in the official validation split of 1, 199 images, with an overall accuracy of 85.74%. VISLIX provides 28 data slices with 84.93% average model accuracy per slice, while AttributionScanner provides 46 slices with an average model accuracy per slice being 86.90%. We further examine slices with an accuracy below 70%, significantly below the overall accuracy, resulting in 7 slices for AttributionScanner and 5 for VISLIX.

As shown in Fig. 10, for AttributionScanner, six slices are spu-

riously associated with water backgrounds (e.g., beach, ocean), leading to the misclassification of landbirds as waterbirds, while one slice is linked to a land background. However, since AttributionScanner uses attribution-weighted embeddings that emphasize model-attended regions (in red), they likely capture background elements and overlook the foreground for problematic instances, leading to dispersed bird species within each slice. In contrast, VIS-LIX's context-aware embeddings focus on birds and their surroundings, combined with HDBSCAN, producing more consistent slices that misclassify specific bird species or birds in particular environments. For example, in Fig. 10, we identify slices that misclassify gadwalls and mallards as landbirds (slice 1) and misclassify kingfishers near water bodies as waterbirds (slice 5). The slice explanations also reveal possible spurious correlations, such as "association of birds with water due to contextual cues like water bodies, reflections, or behaviors near water" for slice 5. In particular, we observe that the slice of "towhees in forests" (slice 16) is mislabeled as waterbirds in the data set, which in fact are landbirds.

These findings suggest that the model may be vulnerable to certain bird species, affected by spurious correlations, and reveal potential labeling issues. Moreover, we do not observe significant overlap between the two sets of slices, suggesting that the two approaches offer distinct insights that complement each other, leading to a more comprehensive validation.



#### Slice Explanation using centroid individual explanations

The common feature in the scenes is the presence of buses and cars in urban environments. The most primary reason for False Positives is the misclassification of buses as cars due to similarities in shape, size, and color, as well as occlusion and partial visibility of the buses.

#### Slice Explanation using random individual explanations

Common feature of these environments: The scenes generally involve urban or city street settings with various objects such as trucks, buses, cars, buildings, trees, and people present. The most common reason for False Positive car detection is the presence of large vehicles, such as trucks or buses, which have similar visual features to cars, leading to confusion in the object detection model.

#### Individual Explanations

1. The scenario is that a false positive car detection occurred in an image of a red double-decker bus driving down a city street. The most possible reason for the false positive detection is that the object detection model mistakenly identified the bus as a car due to similarities in shape, size, and appearance, and potentially due to occlusion of certain parts of the bus or errors in localization.

2. The scenario depicts a yellow school bus parked on a dirt road with people, trucks, and a small building in the background. The most possible reason for the False Positive car detection is that the model may have mistakenly identified the bus as a car due to similarities in shape, size, and color, as well as the presence of headlights on the bus. Additionally, occlusion and the partial visibility of the bus may have contributed to the incorrect detection.

3. The scenario is an urban street scene with a red bus parked next to a tree, a person standing nearby, two small red cars, and a traffic light in the background. The most possible reason for the False Positive CAR detection is that the object detection model mistakenly identified the red bus as a car due to its similar shape and size, as well as the presence of a partially visible car-like structure on the back end of the bus, leading to a misclassification.

4. The scenario is a sunny day with a blue and yellow bus parked on the side of a street, a man standing next to it, and cars parked nearby. The most possible reason for the False Positive car detection could be the similarity in color and shape between the bus and a car, leading to the model mistakenly detecting the bus as a car.

5. The scenario depicts a busy city street with a bus driving down the road. The most possible reason for the False Positive car detection is that the bus in the image shares some visual similarities with a car, such as its shape and color, leading the object detection model to mistakenly identify the bus as a car. Additionally, occlusion, size, and visibility of parts may have contributed to the false positive.

#### •••

19. The scenario is a rainy urban environment with a bus parked next to a building, a car parked behind the bus, and two people nearby. The most possible reason for the False Positive CAR detection could be the presence of similar features between the bus and a car, such as their shapes and colors, leading to a confusion in the object detection model. Additionally, the occlusion and size of the bus may have contributed to the false positive detection.

20. The scenario depicts a sunny day with a dark nighttime lighting condition, featuring a red car driving on top of a white truck carrying a load of fruits and vegetables. The most probable reason for the False Positive car detection is the spurious correlation between the red car on the truck and the actual ground truth car, leading to a misclassification by the object detection model.

21. The scenario is of a false positive car detection where a black bus is parked on the side of the road, partially visible, with a wet street and a building nearby. The most possible reason for the false positive car detection could be the misclassification of the bus as a car due to the following factors: 1. Occlusion: The bus is partially visible, making it difficult for the model to accurately detect its shape and distinguish it from a car. 2. Spurious ... (**Omitted due to the long length**)

22. The image depicts a city street with parked trucks, a partially visible car, a person walking, and a tree, and the most probable reason for the False Positive car detection is the occlusion caused by the large trucks, which blocked the view of the car's front end, resulting in only the back end of the car being detected.

23. The scenario involves a False Positive detection of a car in an image, which is actually a computer screen displaying a blue background with a white keyboard. The most possible reason for the False Positive detection could be the spurious correlation between the blue background on the computer screen and the color of the van in the image, leading the model to mistakenly identify the keyboard as a car.

**Figure 8:** A slice of false positives (FPs) in the car detector and their corresponding explanations. The images display all 23 FPs in red bounding boxes, sorted by the descending centrality of sentence embeddings derived from their individual explanations. Below, two slice-level explanations are shown: one generated using the top 80% most central individual explanations and another using a random 80% selection. We also show individual explanations of the 5 most central explanations (1–5) and the 5 least central ones (19–23) that are excluded by the centroid method. We highlighted problematic or irrelevant explanations in gray.



Figure 9: Expert ratings of the system on a five-point Likert scale.

#### References

- [MSRPC21] MCCONVILLE R., SANTOS-RODRIGUEZ R., PIECHOCKI R. J., CRADDOCK I.: N2D:(not too) deep clustering via clustering the local manifold of an autoencoded embedding. In 2020 25th international conference on pattern recognition (ICPR) (2021), IEEE, pp. 5145–5152. 2
- [Rou87] ROUSSEEUW P. J.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics 20 (1987), 53-65. 2
- [SDA\*20] SOHONI N., DUNNMON J., ANGUS G., GU A., RÉ C.: No subclass left behind: Fine-grained robustness in coarse-grained classification problems. Advances in Neural Information Processing Systems 33 (2020), 19339–19352. 2
- [XDLM24] XUAN X., DENG Z., LIN H.-T., MA K.-L.: SLIM: Spuriousness mitigation with minimal human annotations. In European Conference on Computer Vision (2024), Springer, pp. 215-231. 4
- [XOG\*25] XUAN X., ONO J. P., GOU L., MA K.-L., REN L.: AttributionScanner: A visual analytics system for model validation with metadata-free slice finding. IEEE Transactions on Visualization and Computer Graphics (2025). 4



6 of 7



**Figure 10:** Slices with accuracies below 70% identified by AttributionScanner (left, 7 slices) and VISLIX (right, 5 slices), for the waterbirds/landbirds classification model. Each slice displays 5 random examples and their predictions (Pred) and ground-truth labels (GT). For AttributionScanner, we show spurious features and attribution maps. For VISLIX, we provide generated slice explanations, and indicate species based on dataset metadata when they are consistent within the slice.

# **Edge Cases From a Car Detection Model**

We present 13 edge cases derived from a **car** detection model. Each case showcases 10 random samples of False Positives (FPs) and False Negatives (FNs), accompanied by a textual explanation summarizing the errors of this edge case.

FPs (indicated by red bounding boxes) denote instances where the model incorrectly identifies objects.

FNs (indicated by blue bounding boxes) denote instances where the model fails to detect the ground truth.

Edge Case 1 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** The common feature of these environments is that they are all in airport settings, with various vehicles and people present. The most primary reason for the False Positives is the misclassification of parts of trucks or airplanes as cars, due to similarities in appearance, occlusion, and background similarity. The most primary reason for the False Negatives is the occlusion caused by other vehicles, such as trucks or luggage carts, obstructing the view of the cars, leading to limited visibility and difficulties in detection.

Edge Case 2 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** The common feature in the scenes is the presence of buses and cars in urban environments. The most primary reason for False Positives is the misclassification of buses as cars due to similarities in shape, size, and color, as well as occlusion and partial visibility of the buses. The most primary reason for False Negatives is the occlusion and partial visibility of cars, leading to difficulties in accurately detecting and localizing them.

Edge Case 3 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common feature: All scenarios involve a car detection model in a parking lot scene with various obstructions and occlusions. Primary reason for False Positives: Similarity in color, shape, and size between objects like electrical boxes and a car, along with occlusion caused by fences or other vehicles. Primary reason for False Negatives: Occlusion caused by other vehicles, fences, or objects obstructing the view of the car, limited visibility of car parts, and challenging lighting conditions.

Edge Case 4 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation**: The common feature of these environments is the presence of occlusion and partial visibility of the cars. The most primary reason for False Positives is the model mistaking other objects, such as car doors, hoods, tires, or curtains, as cars due to their similarity in color, shape, or prominent appearance. The most primary reason for False Negatives is the occlusion caused by other cars, buildings, fences, mirrors, bushes, or people, obstructing the view of the ground truth cars and making it difficult for the model to accurately detect and localize them.

Edge Case 5 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** The common feature in these environments is that they contain elements that can be visually similar to cars or cause occlusion of the car itself. The primary reason for the False Positives is the misinterpretation of objects such as green objects, reflections in mirrors, and dashboard screens as cars. The primary reason for the False Negatives is the occlusion of the car by objects like people, dogs, and mirrors, as well as the limited viewpoint and inaccurate ground truth.

Edge Case 6 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation**: Common Feature: The scenarios involve partially visible cars in various environments, including garages, showrooms, and parking lots, with occlusion caused by people and other objects. Most likely reason for False Positives: Partial visibility of cars, particularly the open hood and windshield, combined with the presence of people or other objects that may confuse the model. Most likely reason for False Negatives: Occlusion caused by people and other cars, dim lighting conditions, and the presence of other objects/people obstructing the view of the cars. Edge Case 7 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common feature of the environments: The scenarios involve parking lots or commercial venues with multiple cars, people, and occlusion present. Most primary reason for False Positives: Similar shapes, colors, or characteristics in the surrounding objects confuse the object detection algorithm. Most primary reason for False Negatives: Occlusion caused by people, other cars, or trailers obstructing the view of the cars, combined with potential localization errors.

Edge Case 8 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation**: The common feature in these scenes is the presence of motorcycles and cars. The most primary reason for the False Positives is the confusion between motorcycles and cars due to similarities in appearance, occlusion, and proximity in the image. The most primary reason for the False Negatives is the occlusion and partial visibility of cars caused by motorcycles, people, trees, and other objects in the scene.

Edge Case 9 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common Feature: All scenarios involve partial or complete occlusion of the cars by different objects such as trucks, bushes, branches, trees, and obstructions created by the environment. Primary Reason for False Positives: The misinterpretation of partial visibility and occlusion by other objects as complete car detection. Primary Reason for False Negatives: The difficulty in accurately detecting cars due to partial or complete occlusion by various objects, such as trucks, bushes, branches, trees, and obstructions created by the environment.

Edge Case 10 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation**: Common Feature: The scenarios involve parking lots and busy streets with various cars and occlusion. False Positives: The most primary reason for false positives is occlusion and limited visibility of the car due to other parked vehicles, taxi cabs, or trees. The detection model mistakenly classifies the visible parts as cars. False Negatives: The most primary reason for false negatives is occlusion caused by other cars, trees, or people, leading to limited visibility of the car. The model struggles to detect the complete vehicle accurately in these scenarios.

Edge Case 11 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** The common feature of these environments is that they involve various factors that make it challenging for the car detection model to accurately detect and localize cars. The primary reason for False Positives is the presence of similar colors and shapes in the background or other objects. The primary reason for False Negatives is a combination of factors including occlusion, low visibility, dark lighting conditions, small car size, and blurry backgrounds.

Edge Case 12 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation**: The common feature of these environments is that they all involve parking lots or street scenes with partially visible cars and various occlusions. The primary reason for False Positives is the partial visibility of the car, occlusion by other objects, and misinterpretation of the visible parts as complete cars. The primary reason for False Negatives is occlusion caused by other cars, fences, trees, or buildings, along with the low angle, dim lighting conditions, and similar colors blending with the surroundings.





**Explanation:** Common feature: The scenarios involve various obstacles obstructing the view of the cars, such as tree branches, rocks, snow banks, and other cars. The lighting conditions are often dark, and the images are sometimes blurry or black and white. Primary reason for False Positives: The presence of other objects or structures, such as a clock tower, a black object, or a person, that resemble or confuse the model into incorrectly detecting them as cars. Primary reason for False Negatives: The occlusion caused by tree branches and other objects, combined with low visibility due to darkness, blurriness, or small size of the cars. The model also struggles to accurately detect the shape and features of the cars in these challenging environments.

# **Edge Cases From a Dog Detection Model**

We present 7 edge cases derived from a **dog** detection model. Each case showcases 10 random samples of False Positives (FPs) and False Negatives (FNs), accompanied by a textual explanation summarizing the errors of this edge case.

FPs (indicated by red bounding boxes) denote instances where the model incorrectly identifies objects.

FNs (indicated by blue bounding boxes) denote instances where the model fails to detect the ground truth.





**Explanation:** Common Feature: Grassy fields with sheep and other animals present. Most Primary Reason for False Positives: Misclassification due to similarities in appearance, size, shape, color, and occlusion between sheep and dogs. The dog detection model may also confuse other animals like deer, horses, and goats as dogs. Most Primary Reason for False Negatives: Partial occlusion and visibility of the dog's body, particularly the head, by surrounding objects such as trees and branches, leading to difficulty in accurate detection by the model.

Edge Case 2 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** The common feature of these environments is that they involve people interacting with dogs in various settings, including outdoor parks, benches, and indoor rooms. The most primary reason for False Positives is occlusion caused by people's arms, bodies, and objects, which partially cover the dog's body, leading to misinterpretation as a dog or misclassification based on similar features. The most primary reason for False Negatives is occlusion caused by people, leashes, and other dogs, which obstruct the view of the dog and make it difficult for the model to accurately detect and localize the dog.

Edge Case 3 | False Positives: Red bounding box ; False Negatives: Blue bounding box



**Explanation:** Common Feature: The common feature of these environments is the presence of objects or elements that visually resemble dogs, such as jackets, clothes, people, and occlusions caused by fences, grass, and other objects. Primary Reason for False Positives: The primary reason for false positives is the model's confusion between visually similar objects or elements, such as mistaking jackets, black Labrador retriever, and mountain range features as dogs. Primary Reason for False Negatives: The primary reason for false negatives is the occlusion or obstruction of the dog's body or features, such as fences, people, grass, and shadows, making it difficult for the model to accurately detect the complete dog.

### Edge Case 4 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** The common feature in these environments is the presence of multiple dogs and occlusion caused by other objects or dogs. The most primary reason for False Positives is the detection model only detecting parts of the dog, such as the rear end or face, without detecting the full body. The most primary reason for False Negatives is occlusion, where the ground truth dog is partially or fully hidden by other dogs or objects, making it challenging for the model to accurately detect the complete dog.

### Edge Case 5 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common Feature: Several false positives and false negatives occur due to occlusion or obstruction in the scene, where parts of the dog's body are covered or hidden from view. Primary Reason for False Positives: The primary reason for false positives is the model mistaking objects or elements in the scene, such as towels, pillows, patterns, or furniture, as part of the dog due to similarities in color, texture, or shape. Primary Reason for False Negatives: The primary reason for false negatives is the occlusion or obstruction caused by window frames, furniture, or other objects, which prevent the model from accurately detecting the complete dog. Dark lighting conditions can also contribute to false negatives.

Edge Case 6 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common Feature: The scenarios involve dogs in various locations such as beds, couches, and floors, often with occlusion from objects or people. Most Primary Reason for False Positives: Occlusion caused by objects or people obstructing the view of the dog, similarity in color and texture between the dog and the surroundings, and misinterpretation of background objects as separate dogs. Most Primary Reason for False Negatives: Obstruction of the dog's head or body by people or objects, resulting in limited visibility and inability to accurately detect the dog. Edge Case 7 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common Feature: The common feature in these scenarios is that they all involve scenes with cats and dogs resting together on a couch or bed. Most Primary Reason for False Positives: The most primary reason for the false positives is the misclassification caused by the partial visibility of the dogs and their similarity to the background or the cats in terms of color, shape, and context. Occlusion by other objects, such as blankets or furniture, also contributes to the false positives.

# **Edge Cases From a Chair Detection Model**

We present 15 edge cases derived from a **chair** detection model. Each case showcases 10 random samples of False Positives (FPs) and False Negatives (FNs), accompanied by a textual explanation summarizing the errors of this edge case.

FPs (indicated by red bounding boxes) denote instances where the model incorrectly identifies objects.

FNs (indicated by blue bounding boxes) denote instances where the model fails to detect the ground truth.

Edge Case 1 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common Feature: The common feature in these environments is the presence of animals (cats and dogs) and occlusion caused by them, as well as dim lighting conditions. Primary Reason for False Positives: The primary reason for false positives is the misclassification of animals (cats and dogs) as chairs due to similarity in color, shape, and texture, as well as occlusion caused by the animals obstructing the view of the chair. Primary Reason for False Negatives: The primary reason for false negatives is the occlusion caused by animals (cats and dogs) partially or fully blocking the view of the chair, making it difficult for the model to accurately detect the chair. Dim lighting conditions and the small size of the detected chair parts also contribute to false negatives.

Edge Case 2 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common feature of environments: The scenarios involve various indoor settings such as dining rooms, restaurants, and rooms with furniture. Most primary reason for False Positives: Similarity in color, shape, and features between the chair and other objects in the environment, as well as partial visibility and occlusion of the chair, leading to confusion in the detection model. Most primary reason for False Negatives: Occlusion of the chair by other objects in the environment, such as tables or other chairs, combined with partial visibility, dim lighting conditions, and similarity in color between the chair and the obstructing objects, making it challenging for the model to accurately detect the chair. Edge Case 3 | False Positives: Red bounding box ; False Negatives: Blue bounding box



Explanation: The scenarios involve various instances where the chair is partially visible and obstructed by other objects such as tables, desks, monitors, keyboards, and other items commonly found in workspaces or rooms. The primary reason for the false positive chair detections is the misinterpretation of partially visible chairs and tables as complete chairs. This could be due to the object detection model mistaking the partial visibility and occlusion as the complete chair, leading to incorrect detections. The primary reason for the false negative chair detections is the occlusion caused by other objects, such as tables, desks, monitors, keyboards, and furniture, obstructing the view of the chair. The occlusion makes it difficult for the model to accurately detect the entire chair, resulting in false negatives.

Edge Case 4 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common feature of these environments: - Outdoor scenes with various objects and occlusions, such as people, vehicles, boats, and structures like windows and railings. Primary reason for False Positives: - Confusion caused by occlusions, partial visibility, and visual similarities with other objects, such as people, vehicles, boat structures, and ladders. Primary reason for False Negatives: - Occlusion by obstructions like people, glass doors, windows, railings, benches, and boats, as well as partial visibility, complex backgrounds, and smaller size compared to other objects.

Edge Case 5 | False Positives: Red bounding box ; False Negatives: Blue bounding box





Explanation: The scenarios involve cluttered rooms or images with various objects and occlusion, making it challenging for the chair detection model to accurately identify the chair. The primary reason for false positives is the presence of objects with similar shapes, sizes, or colors to a chair, such as boxes, bags, or furniture, leading to confusion for the chair detection model. Additionally, partial visibility and occlusion of the chair by other objects contribute to inaccurate detection. The primary reason for false negatives is occlusion, where the chair is partially hidden or obstructed by other objects such as stacks of papers, bags, or furniture, making it difficult for the model to detect the chair accurately. Dim lighting conditions and the small size of the chair also contribute to the false negative detection.

### Edge Case 6 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common Feature: The common feature in these scenes is the presence of occlusion or obstruction, either by other objects or by the environment (e.g., lighting conditions). The primary reason for false positives is the confusion caused by partial occlusion or obstruction of the chair by other objects, such as beds, potted plants, wooden beams, or similar structures. This confusion leads the detection model to misclassify these objects as chairs. The primary reason for false negatives is the occlusion or obstruction of the chair by other objects, such as tables, towels, potted plants, lamps, couches, or people. The occlusion makes it difficult for the model to accurately detect and localize the chair, resulting in false negatives. Additionally, dim lighting conditions can further affect the model's ability to detect the chair accurately.

Edge Case 7 | False Positives: Red bounding box ; False Negatives: Blue bounding box





Explanation: Common Feature: The common feature in these scenarios is that they all involve living rooms with various furniture and objects, such as couches, tables, pillows, and chairs. The primary reason for the False Positives is the presence of objects that resemble chairs, such as couches, pillows, and tables, which can confuse the object detection model. Additionally, occlusion and partial visibility of the chairs contribute to the incorrect detection. The primary reason for the False Negatives is occlusion caused by other objects, such as pillows, vases, cabinets, tables, and people, which obstruct the view of the chairs and make it difficult for the model to accurately detect them. Additionally, the focus on other objects and the dim lighting conditions may divert attention away from the chairs

Edge Case 8 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common Feature: The common feature in these environments is the presence of occlusion, either due to objects partially covering the chair or the chair being partially visible. Primary Reason for False Positives: The primary reason for false positives is the occlusion caused by objects such as tablecloths, doors, dolls, fabric coverings, and other furniture pieces, leading to a partial detection of the chair's visible parts and confusing the model. Primary Reason for False Negatives: The primary reason for false negatives is the occlusion caused by tables, cabinets, shelves, plants, vases, curtains, and other objects obstructing the view of the chair, making it challenging for the model to accurately detect and localize the chair.

Edge Case 9 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common Feature: The common feature in these scenes is that they all involve dining tables or dining areas with chairs, often with occlusions or obstructions present. Primary Reason for False Positives: The primary reason for False Positives is the confusion caused by occlusions or obstructions, such as tables, objects on the table, or other chairs, leading to misclassification by the chair detection model. Primary Reason for False Negatives: The primary reason for False Negatives is the occlusion or partial visibility of the chairs, often caused by obstructions like potted plants, vases, tablecloths, or other objects, making it challenging for the model to accurately detect the chairs.

Edge Case 10 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common Feature: The common feature in these scenes is the presence of occlusion and obstruction, either by other objects or by the environment (e.g., tables, curtains, bags, plants, etc.). The scenes often have dim lighting conditions and a variety of other objects present. The most primary reason for False Positives is the occlusion caused by tables, obstructing the view of the chair. Additionally, low-angle perspectives, small chair sizes, and confusion with similar objects contribute to false positives. The most primary reason for False Negatives is occlusion caused by other objects such as tables, vases, suitcases, cushions, and rugs, obstructing the view of the chair. Dim lighting conditions, truncation, and partial visibility of the chair further contribute to false negatives.

### Edge Case 11 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** - The common feature in these environments is that they contain various occlusions and obstructions, such as people, objects, and fabrics, which partially or completely cover the chairs. - The most primary reason for False Positives is the model mistaking parts of the chair, such as armrests, backrests, or fabrics covering the chair, as complete chairs. This is due to similar colors, textures, and shapes between the chair and other objects in the scene. - The most primary reason for False Negatives is the occlusion and obstruction caused by people, objects, or fabrics partially covering the chairs. The limited visibility, combined with dim lighting conditions, make it difficult for the model to accurately detect the complete chairs.

Edge Case 12 | False Positives: Red bounding box ; False Negatives: Blue bounding box



**Explanation:** The common feature in these scenes is the presence of various objects and people, including chairs, in different environments such as restaurants, dining areas, dimly lit rooms, and cozy settings. False Positives: The confusion caused by similar visual features, such as shape, color, and texture, between the chairs and other objects in the scene, including stairs, railing, wooden steps, person's clothing, backrest, and chair-like patterns. Occlusion from objects and people in the scene also contributes to the false positives. False Negatives: The occlusion caused by people sitting on the chairs, obstructing the view of the chair's design and structure. Additionally, partial visibility of the chair, small size, dim lighting conditions, and the focus of the image on people and surroundings rather than the chair itself contribute to the false negatives.

Edge Case 13 | False Positives: Red bounding box ; False Negatives: Blue bounding box



**Explanation:** Common feature of these environments: - The images often contain people sitting at dining tables or on couches in dimly lit rooms. Most primary reason for False Positives: - Chairs being partially hidden or obstructed by people or objects, leading to confusion in the object detection model and resulting in false positive detections.

Most primary reason for False Negatives: - Occlusion caused by people sitting on or partially obstructing the chairs, making it difficult for the model to accurately detect and localize the complete chair.

## Edge Case 14 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common Feature: The common feature among these scenes is the presence of occlusion or partial visibility of the chair. False Positives: The primary reason for false positives is the presence of spurious correlation or confusion due to similar features, shapes, or colors between the chair and other objects in the scene, leading to misclassification. False Negatives: The primary reason for false negatives is the occlusion or obstruction of the chair, either by other objects or by the scene itself, making it difficult for the model to accurately detect the chair.

Edge Case 15 | False Positives: Red bounding box ; False Negatives: Blue bounding box



**Explanation:** Common Feature: The common feature in these environments is that they involve various forms of occlusion, such as objects, people, or furniture obstructing the view of the chairs. Primary Reason for False Positives: The primary reason for false positives is the model mistaking objects or textures in the scene, such as fabric, spheres, reflective surfaces, or wigs, as chairs due to similarities in color, shape, or texture. Primary Reason for False Negatives: The primary reason for false negatives is the combination of occlusion, small chair size, limited visibility, dim lighting conditions, and the model's difficulty in accurately detecting and localizing chairs that are partially visible or obstructed by other objects or people.

# **Edge Cases From a Person Detection Model**

We present 24 edge cases derived from a **person** detection model. Each case showcases 10 random samples of False Positives (FPs) and False Negatives (FNs), accompanied by a textual explanation summarizing the errors of this edge case.

FPs (indicated by red bounding boxes) denote instances where the model incorrectly identifies objects.

FNs (indicated by blue bounding boxes) denote instances where the model fails to detect the ground truth.

Edge Case 1 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common Feature: The common feature in these environments is the presence of bicycles and people wearing various outfits in outdoor settings. The primary reason for False Positives is the occlusion caused by the presence of bicycles, backpacks, chairs, and other objects, as well as the similarity in color, shape, and texture between these objects and a person. Additionally, the limited visibility and partial detection of body parts contribute to the misclassification. The primary reason for False Negatives is the occlusion caused by other objects, such as bicycles and people, obstructing the view of the person. The model struggles to detect the complete person when their body is partially hidden, leading to the false negative.

Edge Case 2 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common Feature of Environments: - Sunny outdoor scenes with various objects like boats, beaches, wooden piers, and rocky shorelines. Primary Reason for False Positives: - Occlusion and limited visibility due to obstructions, such as ropes, poles, and other people, leading to the model incorrectly identifying parts of a person's body as a complete person. Primary Reason for False Negatives: - Occlusion caused by objects like paddles, wooden planks, trees, and other people, obstructing the view of the person's body and making it difficult for the model to accurately detect the complete person.

Edge Case 3 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common Feature of Environments: - The scenarios all involve images taken on a boat or near water. Primary Reason for False Positives: - Occlusion and limited visibility of the person's body parts, leading to misinterpretation by the object detection model. Primary Reason for False Negatives: - Occlusion caused by objects, people, or body parts obstructing the view of the person, making it difficult for the model to accurately detect them.

## Edge Case 4 | False Positives: Red bounding box ; False Negatives: Blue bounding box



**Explanation:** Common Feature: The scenarios involve various environments such as buses, trains, rooms, and outdoor settings with obstructions, limited visibility, and dim lighting conditions. False Positives: The primary reason for False Positives is occlusion caused by objects like jackets, bags, seats, and other people, leading to the model mistakenly identifying these objects as body parts of a person. False Negatives: The primary reason for False Negatives is occlusion caused by obstructions like seats, other people, and objects, as well as limited visibility of the person's face and body due to hair, clothing, and lighting conditions.

Edge Case 5 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common Feature of Environments: The scenarios involve various environments such as parking lots, airports, grassy fields, dimly lit rooms, and outdoor areas with different objects and occlusions. Primary Reason for False Positives: The primary reason for false positives is the misclassification of objects or structures, such as cars, propellers, buildings, tires, and airplanes, as persons due to their shape, size, color, or similarity to human body parts. Primary Reason for False Negatives: The primary reason for false negatives is the occlusion caused by other people, objects, or structures in the scene that obstruct the view of the person, making it difficult for the model to accurately detect the person's body or features.

Edge Case 6 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** The scenarios involve various outdoor settings, such as sunny days, parking lots, grassy fields, and cityscapes, with cars and people present. The most likely reason for false positives is the presence of objects or occlusions that resemble human characteristics, such as car seats, red objects, reflections, or shadows. Occlusion and partial visibility also contribute to the false positive detection. The most common reason for false negatives is occlusion caused by objects like car windshields, steering wheels, side mirrors, and other cars, obstructing the view of the person. Additionally, focus on specific body parts, such as the face or head, and dim lighting conditions can contribute to false negatives.

## Edge Case 7 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** The common feature in these environments is the presence of dining tables or food- related scenes, where people are sitting and enjoying meals together. The primary reason for false positives is occlusion caused by objects, other people, or obstructions in the image. The detection model mistakenly identifies parts of objects, such as tables, utensils, and bags, or visible body parts, like arms, legs, or heads, as people due to their similarity in appearance and prominence in the image. The primary reason for false negatives is occlusion caused by obstructions in the image, such as hands, glasses, food, or other objects, which block the view of the person's body or face. The focus on specific body parts, like upper bodies or faces, and the presence of blurry or visually dominating elements further contribute to the missed detections.

### Edge Case 8 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common Feature: The common feature in all the scenarios is the presence of a bus, car, or train, with various objects and occlusions in the environment. Primary Reason for False Positives: The primary reason for false positives is the spurious correlation between objects in the scene (e.g., reflective surfaces, poles) and the appearance of a person, combined with occlusion and partial visibility. Primary Reason for False Negatives: The primary reason for false negatives is the occlusion caused by windows, window frames, and other objects, obstructing the view of the person and making it challenging for the model to accurately detect them.

Edge Case 9 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common Feature of Environments: The common feature in these environments is the presence of occlusion, low lighting conditions, and partial visibility of the person's body. Most Primary Reason for False Positives: The most primary reason for the false positives is the misinterpretation of objects or occlusions resembling a person, such as a dog's collar, curtains, bottles, or other objects with similar shape or color to a person. Most Primary Reason for False Negatives: The most primary reason for the false negatives is the occlusion caused by objects, such as fences, handbags, cell phones, or other people, obstructing the view of the person's body and making it difficult for the model to accurately detect the complete person.

Edge Case 10 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common feature: The common feature in these environments is the presence of dogs and their interaction with people. Most primary reason for False Positives: The most primary reason for False Positives is the confusion between the appearance, size, and shape of dogs and humans. The object detection model may mistake dogs for people due to similarities in body shape, size, and occlusion caused by the dog's presence. Most primary reason for False Negatives: The most primary reason for False Negatives is occlusion and partial visibility of the person's body caused by dogs. The presence of dogs obstructs the view of the person, making it difficult for the model to accurately detect them.

## Edge Case 11 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common feature of these environments: The environments in these scenarios involve people interacting with each other, often in close proximity, and with various occlusions and lighting conditions present. Primary reason for False Positives: The primary reason for False Positives in person detection is occlusion caused by objects or body parts obstructing the view of the person, leading to only partial detection and mistaken identification of non-person objects or body parts as a person. Primary reason for False Negatives: The primary reason for False Negatives in person detection is occlusion caused by objects, body parts, or other people obstructing the view of the person, making it difficult for the model to accurately detect and localize the person in the image.

Edge Case 12 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** The common feature in these environments is the presence of babies or young children, often in close proximity to adults, on couches or blankets, with various objects and toys in the surrounding area. The primary reason for the False Positives is the occlusion caused by the presence of blankets, toys, and other objects partially covering the babies or young children. This occlusion, along with the small size of the detected parts and similarity in color or shape between the objects and a person, leads to the misclassification as a person. The primary reason for the False Negatives is the occlusion caused by the presence of blankets, bottles, toys, and hands partially obstructing the view of the person. This occlusion makes it difficult for the model to accurately detect and localize the person's body and face.

Edge Case 13 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** The scenarios mostly involve outdoor settings with boats, water, and sunny or bright lighting conditions. - The primary reason for false positives in person detection is the presence of objects or patterns that resemble the appearance of a person, such as red shirts, helmets, life jackets, or boat structures. These similarities in color, shape, and size can trigger misclassifications. Primary Reason for False Negatives: - The primary reason for false negatives in person detection is occlusion and limited visibility of the person's body, often caused by objects, other people, or the angle of the photo. The lack of clear facial features or truncated body parts make it challenging for the model to accurately detect the entire person.

Edge Case 14 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common Feature: - Many scenarios involve dark lighting conditions or partially obscured views of the person. Primary Reason for False Positives: - Occlusion by objects such as bottles, tables, or other people, leading to confusion in the detection model. Primary Reason for False Negatives: - Partial visibility of the person's body, with only specific body parts being detected, making it difficult for the model to accurately detect the complete person.

## Edge Case 15 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common Feature: Dimly lit rooms, presence of occlusions, and objects resembling body parts. Primary Reason for False Positives: Occlusion caused by objects such as pillows, couches, desks, and plants, leading to incomplete or ambiguous features for person detection models. Primary Reason for False Negatives: Occlusion caused by furniture, magazines, curtains, and fans, obstructing the view of the person's body, combined with low angle viewpoints and dim lighting conditions.

## Edge Case 16 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** All scenarios involve dark rooms or images with limited lighting, occlusion, and partial visibility of the person's head or body. The most primary reason for false positives is the model mistaking the partially visible head or body parts, along with occlusions and similarities in color or shape, as complete persons due to the lack of contextual information and accurate ground truth. The most primary reason for false negatives is the occlusion caused by objects or body parts, such as chairs, curtains, or hair, combined with limited visibility, blurry images, and dark lighting conditions, making it challenging for the model to detect the complete person.

## Edge Case 17 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** The common feature in these environments is that they are all dimly lit rooms with various occlusions and obstructions present. The primary reason for false positives is the model mistakenly detecting objects such as clothing, scarves, hands, and body parts as complete persons due to their prominence in the image, lack of visibility of the rest of the body, and spurious correlations. The primary reason for false negatives is the occlusion caused by obstructions such as walls, objects, handbags, and other body parts, which prevent the model from accurately detecting the person's complete body.

### Edge Case 18 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common Feature of Environments: - Many scenarios involve low lighting conditions, blurry backgrounds, and occlusion. Most Primary Reason for False Positives: - The main reason for false positives is the presence of occlusion, partial visibility of body parts, and similarity in color or shape between the main object and a person class. Most Primary Reason for False Negatives: - The primary reason for false negatives is occlusion, limited visibility of body parts, and focus on specific body parts rather than the entire body.

## Edge Case 19 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** - The scenarios involve various indoor settings, often with dim lighting conditions and occlusions caused by objects, other people, or facial features like glasses or facial hair. Primary reason for False Positives: - Occlusion and limited visibility of the person's body or face, leading to the model mistakenly detecting partial features or objects as people. Primary reason for False Negatives: - Occlusion and obstruction caused by objects, other people, or the person themselves, preventing the model from accurately detecting the person's full body or face.

### Edge Case 20 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** The common feature among these environments is that they involve group settings with multiple people present, often posing for pictures or sitting together. The most primary reason for false positives is occlusion, where the presence of other people or objects obstructs the view of the ground truth person, leading to the model mistakenly detecting parts of the person as a separate entity. The most primary reason for false negatives is also occlusion, where other people or objects block the view of the ground truth person, making it difficult for the model to accurately detect the complete person. Additionally, factors such as dim lighting, viewpoint, and partial visibility of the person contribute to false negatives.

Edge Case 21 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common Environment Feature: - The scenarios generally involve outdoor settings, such as parks or grassy areas, with a mix of people and objects in the background. Most Primary Reason for False Positives: - Occlusion and limited visibility of the person's body, leading to misinterpretation by the object detection model. Most Primary Reason for False Negatives: - Partial visibility of the person's body, particularly the upper body or head, leading to a misclassification by the object detection model.

### Edge Case 22 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Common Feature: - The scenarios involve various objects or backgrounds that share visual similarities or correlations with a person, such as shape, color, texture, or occlusion. Primary Reason for False Positives: - The primary reason for the false positives is the spurious correlation between the objects or backgrounds and human features, leading to misclassification by the object detection model. Primary Reason for False Negatives: - The primary reason for the false negatives is the occlusion or partial visibility of the person, which makes it difficult for the object detection model to accurately detect the entire person.

Edge Case 23 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Many of the false positives and false negatives occur in scenarios with partial visibility of the person, occlusion, blurriness, and similarity in color or texture between the person and the background or other objects. The primary reason for false positives is the combination of partial visibility, occlusion, and similarity in color or texture between the person and the background or other objects, leading to misinterpretation by the object detection model. The primary reason for false negatives is the occlusion caused by objects or body parts obstructing the view of the person, making it difficult for the model to accurately detect and recognize the person as a valid person object.

Edge Case 24 | False Positives: Red bounding box ; False Negatives: Blue bounding box





**Explanation:** Many of the scenarios involve partially visible persons, occlusion, blurry images, dark lighting conditions, and focus on specific body parts or objects. The primary reason for false positives is the spurious correlation between the characteristics of objects or backgrounds (such as color, shape, or patterns) and the appearance of a person, leading to misclassification by the object detection model. The primary reason for false negatives is the obstruction of the person's body or face, blurry or dark images, limited visibility of body parts, and the focus on specific objects or body parts, making it difficult for the model to accurately detect and classify the person.

# **Individual Explanation Examples**

We present individual explanation examples for the four datasets used in the paper: Dog detector, Car detector, Chair detector, and Person Detector.

Each dataset includes False Positives and False Negatives, with obviously incorrect explanations marked as "Incorrect".

# **Dog Detector**

- Individual Explanations: 8 False Positives and 4 False Negatives.



The scenario is an image featuring two sheep, one in the foreground with long curly wool and the other in the background, standing on a dirt road in a sunny and bright environment. The most possible reason for the False Positive dog detection is that the sheep in the foreground with its long curly wool may have been mistakenly identified as a dog due to its appearance and size, leading to an incorrect classification.



The scenario is a beach scene with a Dalmatian dog walking and its reflection in the water. The most possible reason for the False Positive dog detection is that the detection model mistakenly identified the dog's reflection as a separate dog due to its similarity in appearance, size, and position, leading to a false positive detection.



The scenario is a false positive dog detection in an image of a small white dog wearing a sweater on a wooden floor, where only parts of the dog's face are detected. The most possible reason for the false positive detection is the occlusion caused by the dog wearing a sweater, which partially covers its body and obstructs the view of the rest of the dog.



The scenario is a cozy living room with a dog sitting on a couch, surrounded by books and a bookshelf in the background. The most possible reason for the False Positive dog detection is that the dog's presence and position on the couch, combined with the background of books and the cozy atmosphere, might have led the object detection model to erroneously classify other objects, such as the couch or books, as dogs.



The scenario involves a false positive dog detection in an image of a black dog and a brown teddy bear on a bed. The most possible reason for the false positive detection is that the dog and teddy bear exhibit similar color and texture characteristics, causing the dog detection model to mistakenly identify the teddy bear as a dog. Additionally, the occlusion and partial visibility of the teddy bear may have contributed to the confusion.



The scenario depicts a woman lying on a red surface, wearing a red hoodie, in a dimly lit room with a laptop nearby, giving off a relaxed atmosphere. The most likely reason for the False Positive dog detection is the presence of the woman's head, which may have features that resemble a dog's face, such as the hair and the top of the head, leading to a misclassification by the object detection model.



The scenario depicts a false positive detection of a dog in an image of a horse with a harness, two dogs, and a person in a grassy area, possibly due to the presence of a small white horse-like dog and the similarity in appearance between the horse and a dog, leading to misclassification by the dog detection model.



(Incorrect) The scenario depicts an image of a brown teddy bear on a wooden table, where a False Positive dog detection occurred due to the presence of a brown piece of food partially covered by a brown piece of paper, leading to a misclassification by the object detection model. The reasons for the False Positive may include the similarity in color and texture between the food and a dog, as well as the partial occlusion caused by the paper.



The scenario depicts a dog lying down in a grassy area, partially hidden by tall grass and bushes. The most possible reason for the False Negative in the dog detection is that the dog's body and face are partially obscured by the surrounding vegetation, making it difficult for the detection model to accurately identify and detect the dog. The dim lighting conditions and the similarity of the dog's fur color to the surrounding grass further contribute to the difficulty in detecting the dog.



The scenario involves a False Negative in a dog detection model, where a small brown dog's face is obstructed by a wooden object, resulting in the model failing to detect it. The possible reasons for the False Negative are: occlusion caused by the wooden object, partial visibility of the dog's body, dark lighting conditions, and the focus on the close-up of the dog's face instead of the entire dog.



The scenario involves a cropped image of two dogs lying down on a couch, with one dog partially obstructing the view of the ground truth dog, a Cocker Spaniel, which is lying down next to a stuffed animal. The most possible reason for the False Negative dog detection is the obstruction caused by the other dog and the stuffed animal, which partially covers the ground truth dog's face and body, making it difficult for the model to detect it accurately.



The scenario depicts a cropped image of a black dog walking on a sidewalk, with a red rope obstructing the view of the dog from the front, and the dog's shadow visible on the wall. The most possible reason for the False Negative in dog detection is the focus on the dog's rear end, making it difficult to see the dog's face and other body parts, combined with the occlusion caused by the red rope.

## **Car Detector**

- Individual Explanations: 12 False Positives and 4 False Negatives.



The scenario is a sunny day with a blue and yellow bus parked on the side of a street, a man standing next to it, and cars parked nearby. The most possible reason for the False Positive car detection could be the similarity in color and shape between the bus and a car, leading to the model mistakenly detecting the bus as a car.



The scenario involves a False Positive detection of a car in an image of a person riding a motorcycle, and the most possible reason for the False Positive is the presence of a partially visible car in the background, coupled with the spurious correlation between the car logo on the person's jacket and the ground truth car. The size, viewpoint, and localization error might have also contributed to the False Positive.



The scenario depicts a yellow school bus parked on a dirt road with people, trucks, and a small building in the background. The most possible reason for the False Positive car detection is that the model may have mistakenly identified the bus as a car due to similarities in shape, size, and color, as well as the presence of headlights on the bus. Additionally, occlusion and the partial visibility of the bus may have contributed to the incorrect detection.



The scenario is a car detection model detecting a false positive of a car in an image of a parking lot, where the car is partially obstructed by a tree and only parts of the car, specifically the car hood, are visible. The most possible reason for the false positive detection is that the model is mistaking the visible car hood for a full car due to its shape and similarity to a car's front end.



The scenario is that a false positive car detection occurred in an image of a red doubledecker bus driving down a city street. The most possible reason for the false positive detection is that the object detection model mistakenly identified the bus as a car due to similarities in shape, size, and appearance, and potentially due to occlusion of certain parts of the bus or errors in localization.



The scenario involves a False Positive detection of a car in a parking lot, where only parts of the car are detected, and the car is obstructed by the wing of a plane. The most possible reason for the False Positive detection could be the occlusion caused by the airplane wing, which partially covers the car and may confuse the object detection model. Additionally, the limited visibility of the car and the presence of other objects in the background, such as the truck, could also contribute to the False Positive.



The scenario depicts a sunny day where a man and a woman are enjoying a boat ride on the water. The most possible reason for the False Positive car detection is that the boat's shape and the person's sitting position in the boat might resemble the shape and size of a car, leading to a misclassification by the object detection model. Additionally, the presence of a steering device held by the person might further contribute to the misclassification.



The scenario is a sunny outdoor image featuring a white car parked next to a pole, with the side view mirror partially detected as a false positive due to the bright reflection caused by the sun shining on the mirror and the car's window, creating a visually striking scene that can affect the visibility of the detected parts of the car.



The scenario involves a false positive car detection in an image of a red motorcycle parked in a parking lot. The most possible reason for the false positive car detection could be the presence of spurious correlation, where the detection model mistakenly associates certain characteristics of the motorcycle, such as its red color and visible tail light, with those of a car.



The scenario involves a False Positive detection of a car in an image featuring a bottle of champagne with a car image on the label, where the car appears to be a small, silver convertible, and is partially visible. The most possible reason for the False Positive detection is that the car image on the champagne bottle label may have similarities with the car detection model's training data, leading to a spurious correlation and a false positive detection.



(Incorrect) The scenario involves a False Positive detection of a car in an image containing a main object of a train, parked on a snowy surface, with only a portion of it visible. The most possible reason for the False Positive detection is the similarity in color and partial overlap between the train and the main car, as well as the presence of other cars and a person in the scene, potentially causing confusion in the object detection model.



(Incorrect) The scenario is a false positive car detection in an image of a house with a clock tower, where the car parked in front of the house is mistakenly detected as a car. This false positive may have occurred due to the small size of the car in the image, the presence of other objects (such as the clock tower), and possibly unclear or inaccurate ground truth data.



The scenario involves a car detection model that has produced a False Negative, where the ground truth car is not detected correctly. The most possible reason for this False Negative is the occlusion caused by the chainlink fence obstructing the car, which prevents the model from accurately detecting the car's presence and features. The focus on the mesh grill, which is the main subject of the photo, further adds to the difficulty in detecting the entire car.



The scenario depicts a partially visible red sports car parked in the background of a crowded protest scene, with the most likely reason for the False Negative car detection being occlusion caused by the presence of a bicycle and a person holding a sign, obscuring the car from the model's view.



The scenario is that a white van parked on the side of the road is partially obscured by a man and a sign, resulting in a False Negative in the car detection model. The most possible reason for the False Negative is that the car is partially hidden and the detected parts are small and out of focus, making it difficult for the model to accurately detect the car.



The scenario depicts a sunny day with a congested highway, where a silver compact car is obstructed by a white car parked in front of it, resulting in a False Negative in the car detection model. The most possible reason for the False Negative is the occlusion caused by the white car, which hides most of the ground truth car, making it difficult for the model to detect the complete car accurately.

## **Chair Detector**

- Individual Explanations: 16 False Positives and 8 False Negatives.



The scenario is a black and white image of a kitchen with a partially visible white refrigerator next to a wall, and various objects on the countertop including a bowl, a bottle, a cup, and a spoon. The most possible reason for the False Positive chair detection could be the presence of objects with similar shape or texture to a chair, such as the bowl or the refrigerator, leading to confusion in the object detection model.



The scenario is an image featuring a black metal stand with a circular base surrounded by plants, where a False Positive detection of a chair has occurred. The most possible reason for the False Positive detection could be the presence of a black metal stand with a circular base, which shares visual similarities with a chair, such as shape and material, leading to confusion in the object detection model.



The scenario depicts a cozy dining setting with a partially visible chair obstructed by a table and a bottle of wine, leading to a False Positive detection in the chair detection model. The most likely reason for the False Positive is the occlusion caused by the table and wine bottle, which prevents the model from accurately detecting the complete chair and instead detects only the partially visible backrest.



The scenario depicts an image of a blue vase filled with plants on a dining table, where a False Positive detection of a chair has occurred. The most possible reason for the False Positive detection could be the presence of a blue flower pot, which may have a similar color and shape to a chair, leading to confusion in the object detection model.



The scenario depicts a false positive detection of a chair in an image of a barbecue grill on a patio surrounded by potted plants. The most possible reason for the false positive detection could be the presence of visual features, such as the shape and color of the grill, that resemble a chair and lead to a misclassification by the chair detection model.



The scenario is a close-up image of a person wearing blue jeans and a green shirt, holding a brown purse, with a kitchen environment in the background featuring a dining table and a chair near it, a bottle on the counter, and a cup close to the sink. The most possible reason for the False Positive CHAIR detection is that the model misinterpreted the shape and color of the person's legs and the surrounding objects as a chair.



The scenario involves a false positive chair detection in an image of a wooden pole, where the main object is a pole leaning against a wall. The most possible reason for the false positive detection is that the chair detector model mistakenly identified the top part of the pole as a chair due to similar shape and size characteristics, as well as the presence of dark wood texture.



The scenario involves a False Positive detection of a chair in an image featuring a striped pattern and a brown wall with a red candle on a shelf; the most possible reason for the False Positive detection is the presence of the striped pattern in the image, which visually resembles the shape and structure of a chair, leading to a misclassification by the chair detection model.

![](_page_52_Picture_0.jpeg)

The scenario is a cozy living room with a couch, chair, and other objects. The most possible reason for the False Positive CHAIR detection is that the chair's size and positioning, coupled with the presence of a pillow on the couch and a blanket on the back, may have created visual cues that resemble a chair, leading to a misclassification by the chair detection model. The background and lighting conditions may have also contributed to the false positive.

![](_page_52_Picture_2.jpeg)

The scenario depicts a group of people gathered around a dining table in a restaurant or boat, with a staircase and pole nearby. The most possible reason for the False Positive chair detection is that the chair model mistakenly identified elements of the staircase, such as the wooden steps or the metal railing, as chairs due to similarities in shape, color, or texture. Additionally, occlusion from objects and people in the scene may have contributed to the confusion.

![](_page_52_Picture_4.jpeg)

The scenario is a baby sitting on a wooden chair, with the chair partially obstructed by the baby's body and the main focus of the image being the baby and the chair. The most possible reason for the False Positive detection of a chair is that the detection model detected parts of the chair (backrest and armrest) but failed to detect the complete chair due to occlusion and the presence of the baby.

![](_page_52_Picture_6.jpeg)

The scenario is an image with a wooden fence, chairs, potted plants, and a white building, where a False Positive detection of a chair has occurred. The most possible reason for the False Positive chair detection is that the chair is partially visible and positioned in front of the wooden fence, causing the detection model to mistakenly classify parts of the fence as the chair due to similarities in color, texture, and shape between the chair and the fence.

![](_page_53_Picture_0.jpeg)

The scenario involves a false positive detection of a chair in an image of a man sitting on a red motorcycle parked on a sidewalk, with a grass background and a car and truck in the distance. The most possible reason for the false positive chair detection could be due to spurious correlation, where certain visual features in the motorcycle, such as its shape or color, might be correlating with the chair class in the model's training data.

![](_page_53_Picture_2.jpeg)

The scenario depicts a woman sitting on a white chair outdoors, with a grassy background and a black handbag nearby. The most possible reason for the False Positive chair detection could be the partial visibility of the chair, as it is positioned next to the person and only the back of the chair is visible, making it challenging for the object detection model to accurately identify the chair.

![](_page_53_Picture_4.jpeg)

(Incorrect) The scenario involves a false positive detection of a chair in an image of a cat sitting on a chair, partially blocking the view of the chair, in a dimly lit room with a potted plant next to it. The most possible reason for the false positive detection is the presence of the cat, which obstructs the view of the chair and causes the model to mistake the cat as a chair. Additionally, the dim lighting and partial visibility of the chair further contribute to the false positive.

![](_page_53_Picture_6.jpeg)

(Incorrect) The scenario is a cozy living room with a red leather chair placed next to a red couch, partially occluded, with a black pillow on it, surrounded by smaller black chairs, a dining table, and a TV, with a green carpet, dim lighting, and a sense of relaxation. The most possible reason for the False Positive CHAIR detection is the presence of the red leather chair next to the red couch, which may have led the model to mistakenly identify the couch as a chair. Additionally, the partial occlusion of the red leather chair by the couch might have contributed to the misclassification.

![](_page_54_Picture_0.jpeg)

The scenario involves a red chair that is partially visible in a room with other chairs and a dining table, and the most possible reason for the False Negative chair detection is the obstruction caused by another red chair that is blocking the view of the ground truth chair, preventing it from being detected accurately by the model.

![](_page_54_Picture_2.jpeg)

The scenario involves a cropped image of a woman sitting at a dining table, with a partially obstructed chair in the background. The most possible reason for the False Negative chair detection is that the focus of the image is on the person and the table setting, with the chair's back and sides not clearly visible. The detection model may have struggled to identify the chair due to its limited visibility and occlusion by the table.

![](_page_54_Picture_4.jpeg)

The scenario involves a small, partially obscured black chair in a dark room, obstructed by a red and yellow tablecloth and table runner, leading to a False Negative in chair detection. The most possible reason for the False Negative is that the chair's small size, partial occlusion, dark lighting conditions, and similarity in color to other objects in the image make it difficult for the model to accurately detect the chair.

![](_page_54_Picture_6.jpeg)

The scenario involves a cropped image of a man sitting on a partially visible chair with a woven, basketlike design, and the most possible reason for the False Negative chair detection is the obstruction caused by the man's arm covering the back of the chair, making it difficult for the model to detect the chair accurately.

![](_page_55_Picture_0.jpeg)

The scenario involves a partially obstructed white chair with a wooden frame and a design on the back, which is partially covered by a white cloth on a table set with various items, including a wine glass and a cup. The most possible reason for the False Negative chair detection is the occlusion caused by the tablecloth and glass, which obstruct the view of the chair's design and structure.

![](_page_55_Picture_2.jpeg)

The scenario is a cropped image of a kitchen with a white chair partially obstructed by a towel, a squareshaped wooden table near the chair, and a counter nearby. The most possible reason for the False Negative chair detection is the occlusion caused by the towel, which covers a significant portion of the chair, making it difficult for the model to accurately detect the chair.

![](_page_55_Picture_4.jpeg)

The scenario depicts a cropped image of a dining table with a partially visible black leather chair, obstructed by a potted plant, where the chair's appearance, color, and position can be discerned, leading to a False Negative in chair detection due to occlusion caused by the plant, partial truncation of the chair, and potentially inaccurate ground truth.

![](_page_55_Picture_6.jpeg)

(Incorrect) The scenario depicts a False Negative in chair detection, where a partially visible chair in a room with a wall is not correctly detected by the model. The most possible reason for this False Negative could be the occlusion caused by the wall, which obstructs the view of the chair and prevents the model from accurately detecting it. The chair's size, orientation, and partial visibility also contribute to the False Negative.

## **Person Detector**

- Individual Explanations: 24 False Positives and 8 False Negatives.

![](_page_56_Picture_2.jpeg)

The scenario involves a false positive person detection where a baby girl sleeping on a bed is mistakenly detected as a person, likely due to occlusion by a blanket and pillow, dim lighting conditions, the baby's face being covered, the baby's body not being fully visible, and the size of the baby being small compared to the bed.

![](_page_56_Picture_4.jpeg)

The scenario depicts a couple celebrating a special occasion in a sunny park, where a False Positive person detection occurred due to the occlusion caused by the man's tie and the woman's dress, the close-up view of the woman's face and upper body, and the absence of the rest of her body in the detection region, leading to incomplete detection and misclassification as a person.

![](_page_56_Picture_6.jpeg)

The scenario depicts a group of people, with one person holding a wooden stick, in an image where a False Positive person detection occurred. The most possible reason for the False Positive person detection is that the wooden stick being held by the person, along with the partial obstruction of the person's body by the wooden wall, may have caused confusion for the object detection model, leading to the incorrect classification of the wooden stick as a person.

![](_page_56_Picture_8.jpeg)

The scenario depicts a young man sitting on a couch, holding a large calico cat, with a partially detected person in the image. The most possible reason for the False Positive PERSON detection is that the person's face and arm were partially visible, leading to a misclassification by the object detection model. This could be due to occlusion by the cat, size of the detected parts, or other factors affecting the visibility of the person.

![](_page_57_Picture_0.jpeg)

The scenario depicts a person sitting at a table, with their hands resting on it, but only the person's hand and arm are visible in the image. The most possible reason for the False Positive PERSON detection could be due to occlusion, as the rest of the person's body is not entirely visible. and only partial body parts are detected. The model might have mistakenly associated the detected hand and arm with a full person.

![](_page_57_Picture_2.jpeg)

The scenario depicts two women sitting on a bus, with one woman hiding her face behind a blue towel, and various objects such as handbags and a backpack in the image. The most possible reason for the False Positive PERSON detection is that the person's face and arm, along with the hand holding the towel, are only partially visible and may not meet the criteria for a complete person detection. Additionally, the occlusion caused by the towel further complicates the detection.

![](_page_57_Picture_4.jpeg)

The scenario is a black and white image of a young girl standing next to a brown horse in front of a house, with the girl smiling and the horse wearing a bridle. The most possible reason for the False Positive PERSON detection is that the model may have mistakenly identified the horse's body or parts as a person due to similarities in shape, size, and context, potentially influenced by spurious correlations or missing ground truth.

![](_page_57_Picture_6.jpeg)

The scenario is an underwater image showing a person wearing a wetsuit and mask, partially submerged in water, surrounded by a blue ocean and a boat in the background. The most possible reason for the False Positive PERSON detection could be the presence of a black and green life preserver, which is partially submerged and partially visible, leading to confusion with a person's body parts.

![](_page_58_Picture_0.jpeg)

The scenario depicts a group of people sleeping on a plane, with a False Positive detection of a person wearing an orange shirt and glasses. The most possible reason for the False Positive detection could be the occlusion caused by the person's position, the partial visibility of the person's face due to resting on a pillow, and the presence of other objects such as a headrest and another person's head obstructing the view

![](_page_58_Picture_2.jpeg)

The scenario involves a closeup image of two men's faces, where a person detection model falsely detects a person due to occlusion caused by the heads of other people in the room, the focus on facial features, the close proximity of the detected parts to a bottle in the foreground, and potentially missing or inaccurate ground truth.

![](_page_58_Picture_4.jpeg)

The scenario is a man reading a newspaper in a dimly lit room, with parts of his body and the newspaper obstructed, and another person also reading a newspaper in the room. The most possible reason for the False Positive PERSON detection is that the occlusion caused by the man holding the newspaper in front of his face and the strap covering his body may have led to the detection model incorrectly identifying the visible parts of the man as a separate person. Additionally, the dim lighting condition and the presence of another person reading a newspaper in the room may have added confusion to the detection model.

![](_page_58_Picture_6.jpeg)

The scenario depicts a sunny day at an airport where a hot dog with toppings is being served, surrounded by parked trucks and cars, airplanes, and people. The most possible reason for the False Positive PERSON detection could be due to the presence of people in the vicinity of the hot dog, which might have led the object detection model to mistakenly identify a person in the image.

![](_page_59_Picture_0.jpeg)

The scenario is a man sitting in a dimly lit room, looking at a laptop, with two other people in the background. The most possible reason for the False Positive PERSON detection is that the model mistakenly classified the man's curly hair as a person due to its prominence in the image, the close-up view of the man's face, and the occlusion of his face by the laptop, leading to a misinterpretation of the detection region.

![](_page_59_Picture_2.jpeg)

The scenario depicts a crowded stadium with people sitting together, some holding cell phones, and a False Positive person detection occurred due to occlusion and limited visibility of the ground truth person caused by another person sitting in front of them and a person sitting next to them, along with the black and white color scheme adding difficulty in discerning details.

![](_page_59_Picture_4.jpeg)

The scenario depicts a group of men standing together at a baseball game, with one man wearing a white hat and another man with a visible tattooed arm. The most possible reason for the False Positive PERSON detection is that the occlusion caused by the other people standing around the person with the tattooed arm, as well as the limited visibility of the person's body, led to a partial detection that matched the model's criteria for a person.

![](_page_59_Picture_6.jpeg)

The scenario depicts an image of a woman wearing a blue dress, holding a red flower, with a person detection model producing a False Positive detection of a person. The most possible reason for the False Positive detection could be the presence of the red flower and the handbag, which obstruct the view of the person's body, causing the model to incorrectly classify the flower and handbag as parts of a person. Additionally, the model may be sensitive to the color and shape of the flower and bracelet, leading to a misclassification.

![](_page_60_Picture_0.jpeg)

The scenario is a woman posing in front of motorcycles at a gathering or show, with a False Positive person detection caused by the presence of a motorcycle seat decorated with a picture of a man and a sign on the handlebar, along with the occlusion and similarity in appearance between the motorcycle seat and a person.

![](_page_60_Picture_2.jpeg)

The scenario is a false positive person detection in an image of a brown horse wearing a bridle and harness, standing near a wall in a close-up perspective; the likely reason for the false positive is the presence of the horse's head, which is similar in appearance to a human head, combined with the occlusion caused by the bridle and harness, leading to a misclassification by the person detection model.

![](_page_60_Picture_4.jpeg)

The scenario is a man drinking from a glass in a room, where he is wearing a white tank top and the room has a TV and a keyboard. The most possible reason for the False Positive PERSON detection is that the model mistakenly identified the glass as a person due to its shape and position, and the presence of the man's hand holding the glass further contributed to the confusion.

![](_page_60_Picture_6.jpeg)

The scenario involves a person riding a bicycle in a sunny environment, with other people present. The most likely reason for the False Positive person detection is that the person's body parts, such as their leg and arm, were mistakenly identified as a person due to their proximity to the bicycle, their red glove, and their outfit, which were the secondary elements in the image

![](_page_61_Picture_0.jpeg)

The scenario is a group of people sitting around a dining table, enjoying a meal together, with a False Positive person detection in the image. The most possible reason for the False Positive detection could be the presence of multiple people and objects in the scene, which could cause confusion for the person detection model and lead to incorrect detections. Additionally, factors such as occlusion, viewpoint, and the positioning of the subjects could further contribute to the False Positive detection.

![](_page_61_Picture_2.jpeg)

(Incorrect) The scenario is that a false positive person detection occurred in an image of a cabinet filled with books, located in a room, with a wall in the background, and partially obscured by a person's head, possibly due to spurious correlation between the shape of the cabinet and a person's body, or due to the presence of books in the cabinet resembling a person's body shape.

![](_page_61_Picture_4.jpeg)

(Incorrect) The scenario is a restaurant scene with a man and a woman sitting at a table, surrounded by other people. The most possible reason for the False Positive person detection is the occlusion caused by the reflections of a woman and a man in the mirrors, which obstruct the view of the person. Additionally, the person is sitting down, facing away from the camera, and only parts of their face are visible, leading to a partial detection.

![](_page_61_Picture_6.jpeg)

(Incorrect) The scenario is set in an urban environment with a motorcycle parked next to a building, and a False Positive person detection occurred due to the reflection of a pair of black boots and a partially visible black mask in the mirror of the motorcycle, where the boots are positioned in front of the mask in the False Positive detection region.

![](_page_62_Picture_0.jpeg)

The scenario is a black and white image of a young girl sitting in water at night, wearing a white dress, with a focus on her silhouette and the dark background. The most possible reason for the False Negative person detection is the obstruction caused by the water and darkness, making it difficult for the model to accurately detect and classify the person in the image. Additionally, the lack of discernible facial features and the focus on the upper body further contribute to the False Negative.

![](_page_62_Picture_2.jpeg)

The scenario depicts a False Negative in person detection, where only parts of a person's face are detected, and the most possible reason for this is the obstruction caused by the person's own hair and other objects, such as a potted plant and a piece of paper, partially covering the person's face, as well as the dim lighting conditions and the presence of other people in the scene.

![](_page_62_Picture_4.jpeg)

The scenario depicts a group of men sitting at a dining table, enjoying beer and food, with one man holding a glass of beer. The most possible reason for the False Negative in person detection is that the person's head and shoulders are partially obscured by the table and other people, making it difficult for the model to detect the person accurately. Additionally, the dim lighting and crowded environment may further contribute to the False Negative.

![](_page_62_Picture_6.jpeg)

The scenario depicts a man sitting on a horse, with a woman standing next to him, in a dimly lit indoor setting, where the person detection model fails to detect the person due to occlusion by the horse and the woman, low visibility of the person's face and body parts, and the focus of the image being on the horse and its rider rather than the person.

![](_page_63_Picture_0.jpeg)

The scenario is a cropped image showing a group of people sitting in a boat on a river, with a person wearing a black jacket and a white shirt partially obstructed from view. The most possible reason for the False Negative in the person detection is the occlusion caused by the backs of the people sitting in front of the person, as the person's upper body and face are not clearly visible in the image.

![](_page_63_Picture_2.jpeg)

The scenario depicts a person standing behind a wooden fence in a dimly lit environment, wearing a yellow shirt and holding a handbag, with obstructed view, a small size, and only detected parts of the person visible, such as the back, backpack, and hair. The most possible reason for the False Negative detection could be the occlusion caused by the wooden fence, the person's silhouette making it difficult to detect the person's features, and the dark lighting conditions affecting the model's ability to accurately detect the person.

![](_page_63_Picture_4.jpeg)

The scenario depicts an image of a baby lying in a crib, partially covered by a blanket, with only the baby's face visible. The most possible reason for the False Negative in person detection is that the model fails to detect the person due to occlusion caused by the blanket and the baby's hand, as well as the limited visibility of the person's body and face in the image.

![](_page_63_Picture_6.jpeg)

(Incorrect) The scenario involves a person wearing a white shirt and hat, standing in a brightly lit area, with their arms obstructing their face, and the lower part of their body not clearly visible. The most possible reason for the False Negative in the person detection is the occlusion caused by the person's own arms, which prevents the model from accurately detecting the full extent of the person's presence.