Overthinking Causes Hallucination: Tracing Confounder Propagation in Vision Language Models

Supplementary Material

LR	Maximum depth	No. of estimators	F1 (%)
0.1	3	100	71.39
0.1	3	200	72.43
0.1	3	300	72.72
0.1	5	100	73.59
0.1	5	200	74.25
0.1	5	300	73.06
0.1	10	100	73.42
0.1	**10**	**200**	**74.38**
0.1	10	300	74.19
0.05	3	100	71.02
0.05	3	200	71.59
0.05	3	300	72.16
0.05	5	100	72.40
0.05	5	200	72.83
0.05	5	300	72.81
0.05	10	100	73.43
0.05	10	200	72.87
0.05	10	300	74.19
0.01	3	100	54.53
0.01	3	200	66.30
0.01	3	300	68.97
0.01	5	100	58.18
0.01	5	200	68.28
0.01	5	300	71.48
0.01	10	100	61.18
0.01	10	200	68.06
0.01	10	300	70.18

Table 1. Results of hyperparameter tuning with GB model.

A. Hyperparameters Optimization

The hyperparameters of GB and MLP variants of the detector are finetuned using Grid Search on a selected set of parameters. The training dataset is split into a train and validation set with a 10% allocation for the validation set. The models are finetuned to improve the F1-score on the validation set.

GB Optimization: The optimal value of the number of estimators, learning rate (LR) and maximum depth parameters are chosen from the set {100, 200, 300}, {0.1, 0.05, 0.01}, and {3, 5, 10} respectively. The performance of the model under different hyperparameter combinations is shown in Tab. 1.

MLP Optimization: The optimal value of hidden units, learning rate (LR) and optimizers are chosen from the set {32, 64, 128}, {0.1, 0.01, 0.001}, and {Adam, SGD} respectively. The performance of the model on validation set under different hyperparameter combinations is shown in Tab. 2.

Hidden units	LR	Optimizer	F1 (%)
32	0.1	Adam	69.50
32	0.1	SGD	77.81
32	0.01	Adam	73.97
32	0.01	SGD	73.48
32	0.001	Adam	76.04
32	0.001	SGD	72.99
64	0.1	Adam	75.02
64	0.1	SGD	77.02
64	0.01	Adam	73.94
64	0.01	SGD	78.09
64	0.001	Adam	75.37
64	0.001	SGD	72.82
128	0.1	Adam	72.77
128	0.1	SGD	74.66
128	0.01	Adam	75.66
128	**0.01**	**SGD**	**78.31**
128	0.001	Adam	74.91
128	0.001	SGD	73.65

Table 2. Results of hyperparameter tuning with MLP model.

B. Semantic Alignment Calculation

Maximum Semantic Alignment We employ the sentence-transformer all-MiniLM-L6-v2 [5] to embed tokens decoded from different VLM layers. For each generated token, we compute the cosine similarity between its final-layer embedding and the embeddings of tokens from all intermediate layers. The semantic alignment is defined as the maximum similarity across layers, as shown in Algorithm 1.

Layer-wise Semantic Alignment We analyze how intermediate layers influence the final prediction by computing the cosine similarity between the final-layer token (both hallucinated and real) and the tokens decoded at each layer. The layer-wise alignment curves for LLaVA-1.5, Gemma-3, and Qwen3-VL in Fig. 1 show that semantic similarity becomes high from the middle layers onward, indicating that

Algorithm 1: Maximum Semantic Alignment Computation

Input: Top-1 tokens at each layer $x_1, \ldots, x_L$ and sentence transformer T
Output: Semantic alignment value S_{align}
Initialize: $S_{\text{align}} \leftarrow 0$;
Embed the final-layer token: $x_L^{\text{emb}} \leftarrow T(x_L)$;
foreach $x_i \in \{x_1, \ldots, x_{L-1}\}$ **do**
  if $x_i \neq x_L$ **then**
    Embed token: $x_i^{\text{emb}} \leftarrow T(x_i)$;
    Compute cosine similarity:
    $S \leftarrow \text{cosine\_similarity}(x_L^{\text{emb}}, x_i^{\text{emb}})$;
    if $S > S_{align}$ **then**
      $S_{\text{align}} \leftarrow S$;
return S_{align};

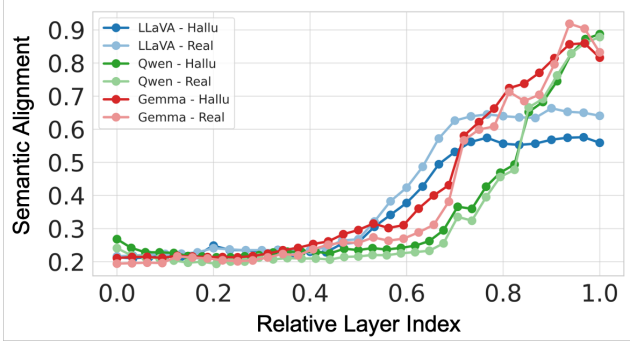


Figure 1. Semantic alignment of hallucinated (hallu) and real (real) token instances across the layers (relative index) for LLaVA-1.5, Gemma-3 and Qwen3-VL models.

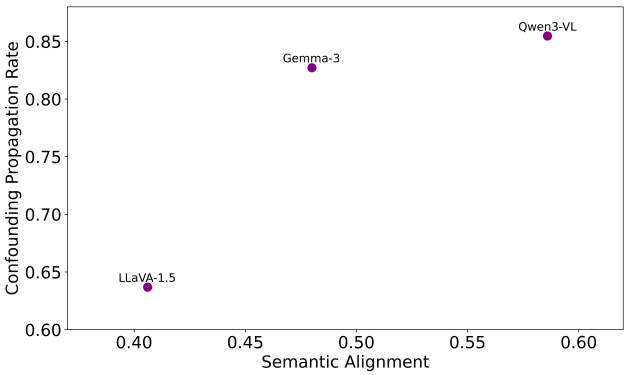


Figure 2. Semantic Alignment vs Confounding Proppagation Rate in LLaVA-1.5, Qwen3-VL and Gemma-3 models.

these intermediate representations strongly shape final outputs.

C. Strong Scene-prior Examples Selection

We use the `facebook/bart-large-mnli` model [3] solely as an embedding encoder. The generated image description is embedded and compared (via cosine similarity) against the embeddings of all predefined scene labels (SCENE_LABELS = [“beach”, “playground”, “forest”, “park”, “street”, “mountains”, “desert”, “kitchen”, “living room”, “bedroom”, “office”, “classroom”, “outdoors”, “indoors”, “city”, “festival”, “space”, “dining room”, “ocean”, “lake”, “river”].). The label with the highest similarity is treated as the inferred scene context, which allows us to estimate scene priors without requiring ground-truth scene annotations. A sample is marked as having a strong scene prior when the generated object token shows similarity above 0.6 with this inferred (maximum-similarity) scene label. The process is detailed in Algorithm 2.

Algorithm 2: Extract Strong Scene Prior Cases

Input: Dataset D, scene extraction model S, sentence transformer T
Output: Filtered dataset D'
Initialize: $D' \leftarrow \{\}$;
Set similarity threshold $t \leftarrow 0.6$;
foreach $d \in D$ **do**
  Extract scene label:
  $d_{\text{scene}} \leftarrow S(d_{\text{description}}, \text{SCENE\_LABELS})$;
  Let the next token of instance d be d_{x_t};
  Embed scene and token:
  $d_{\text{scene}}^{\text{emb}} \leftarrow T(d_{\text{scene}})$;
  $d_{x_t}^{\text{emb}} \leftarrow T(d_{x_t})$;
  Compute cosine similarity:
  $c \leftarrow \text{cosine\_similarity}(d_{\text{scene}}^{\text{emb}}, d_{x_t}^{\text{emb}})$;
  if $c > t$ **then**
    Append d to D';
return D';

D. Attention-based Methods Failures

We observe that attention-based methods such as SVAR fail to detect hallucinated objects when strong scene priors are present. As illustrated in Fig. 3, several hallucinated tokens receive high intermediate-layer attention (layers 5–18), yet SVAR incorrectly classifies them as real because attention magnitude remains high. To quantify this effect, we isolate strong scene-prior cases using Algorithm 2 and evaluate both SVAR and our method. As shown in Tab. 4, our approach achieves substantially higher AUC (86.36%) and F1 (82.59%), demonstrating its robustness in context-biased scenarios where SVAR fails.

Table 3. Impact of overthinking score integration in existing hallucination detection methods.

	w/o. Overthinking Score		w. Overthinking Score	
Method	AUC	F1	AUC	F1
SVAR [2]	85.12	69.35	**86.67**	**75.06**
HalLoc [4]	80.38	73.68	**88.53**	72.97
MetaToken (LR) [1]	85.41	72.88	**87.67**	70.09
MetaToken (GB) [1]	88.95	75.95	**89.15**	**76.14**
MetaToken (MLP) [1]	86.81	73.89	**89.23**	72.35

Method	**AUC**	**F1**
SVAR	76.92	48.86
Ours (GB)	**86.36**	**82.59**

Table 4. Performance comparison (%) of SVAR and overthinking-based approach on strong scene prior cases.

	AUC
Full features	83.33
w/o. entropy	82.91
w/o. image attention	82.42
w/o. text attention	80.49
w/o. Overthinking Score	**79.93**

Table 5. Comparison of performance (%) drop by ablating individual features.

Method	Inference time (s)
Greedy Search	4.21
SVAR [2]	8.07
HalLoc [4]	5.03
MetaToken (LR) [1]	5.35
MetaToken (GB) [1]	5.42
MetaToken (MLP) [1]	5.53
Ours (LR)	5.61
Ours (GB)	5.77
Ours (MLP)	5.74

Table 6. Inference time of various hallucination detection methods.

E. S-OT Improves Other Baselines

To highlight the benefit of the Overthinking Score, we add S-OT to the feature sets of existing detectors and compare performance. As shown in Tab. 3, every method improves substantially once S-OT is included, demonstrating that S-OT provides a meaningful performance boost. Remarkably, S-OT is only a single scalar, yet it delivers greater benefit than many multidimensional attention and entropy features, showing that it captures a uniquely informative signal missing in prior methods.

F. Features Importance

Feature Ablation To understand the importance of each feature in our detection model, we ablate one feature at a time and observe the feature causing the largest performance drop with the GB-variant for hallucination detection in the LLaVA-1.5 baseline. Since the overthinking score is a singleton feature compared to the other features, for a fair comparison, we consider the average of image attention, text attention and entropy features across the layers. As shown in Tab. 5, the largest performance drop (83.33% $\rightarrow$ 79.93%) is seen when the overthinking score is removed.

SHAP Analysis We estimate the significance overthinking score over other features such as image attention, text attention and entropy using the average SHAP value of the compound features on MS-COCO. The results of SHAP values for the Entropy/Text Attention/Image Attention and **Overthinking Scores** are: 0.002/0.004/0.004/**0.007** respectively. Overthinking score has the highest importance (0.007 SHAP value) compared to other features.

G. Computational Cost

Tab. 6 shows that our method incurs only 36% additional computational cost in terms of inference time compared to the default greedy search method.

H. Confounder Propagation Examples

The Fig. 4 shows some examples of confounding propagation in LLaVA-1.5.

I. Qualitative Results

We provide qualitative examples in Fig. 5, 6, 7, 8.

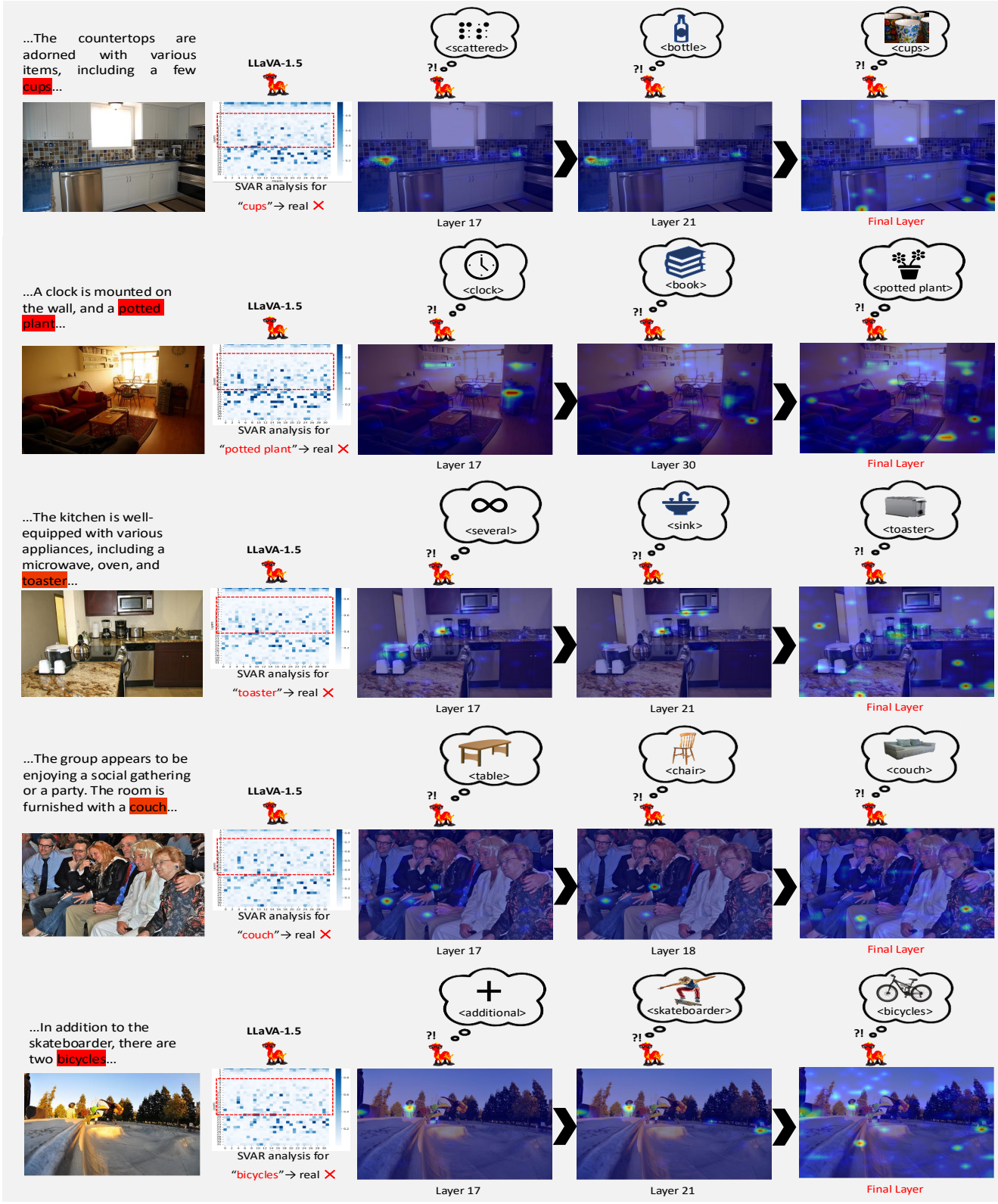
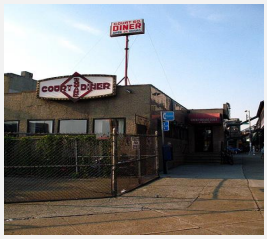


Figure 3. The above examples illustrates that SVAR fails to detect hallucinated objects when the attention values are active in the intermediate layers.

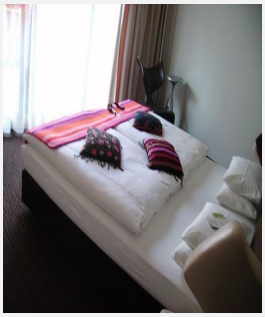


The image features a restaurant called "Cooter's Diner" with a large neon sign on top of the building. The sign is prominently displayed, drawing attention to the restaurant. The diner is located on a street corner, and there are several cars parked nearby. In addition to the cars, there are two___

Confounders: teen → pairs → sets → people

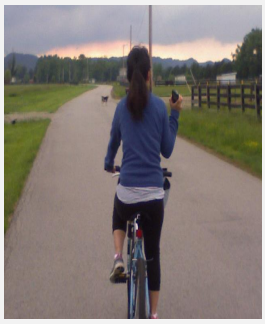
The image depicts a group of young men gathered in a kitchen, preparing food and enjoying each other's company. There are four people in the scene, with one person standing near the refrigerator, another near the__

Confounders: table → counter → sink



The image features a hotel room with a large bed, which is neatly made and ready for guests. The bed has two white pillows on it, one on the left side and the other on the right side. There are also two colorful blankets on the bed, one on the left side and the other on the right side. In addition to the bed, there is a chair located near the right side of the room, and a__

Confounders: clock → table → chair → couch



The image features a woman riding a bicycle down a country road, enjoying her ride. She is wearing a blue shirt and black pants, and she has a cell phone in her hand. The woman is also accompanied by a dog, which is walking beside her on the road. In the background, there are several__

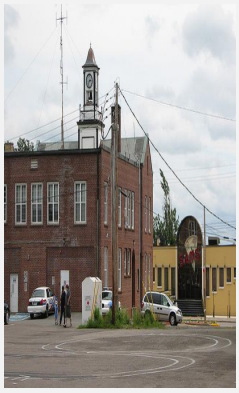
Confounders: buildings → vehicles → houses

The image features a cake with a black bird, possibly a crow, as the main decoration. The bird is perched on top of the cake, and the cake itself is placed on a___

Confounders: top → board → table

The image features a kitchen counter filled with a variety of fresh fruits and vegetables. There are several apples scattered across the counter, with some placed closer to the center and others towards the edges. A bunch of__

Confounders: fresh → green → bananas



The image features a large brick building with a clock tower on top, located in a town square. The building appears to be a courthouse, and there are several people walking around the area. In total, there are five people visible in the scene, with some walking closer to the building and others further away. There are also several cars parked around the courthouse, with three cars on the left side and two cars on the right side of the building. Additionally, there is a__

Confounders: bus → car → clock → truck

The image features a kitchen scene with a person standing next to a stove, preparing a meal. The stove is filled with a pot of pasta, and there are several bowls and a spoon placed on the countertop. The person is standing near the stove, possibly stirring the pasta or attending to other tasks in the kitchen. In addition to the main dish, there are multiple___

Confounders: tomato → carrots

The image features a woman standing in a living room, attentively watching a dog playing with a toy. The dog is located near the center of the room, with the toy in its mouth. The woman appears to be observing the dog's actions and enjoying the playful moment. The living room is furnished with a___

Confounders: table → chair → television → couch

The image features a silver computer keyboard sitting on a desk. The keyboard is positioned towards the left side of the desk, with a mouse placed on the right side. The mouse is located closer to the top right corner of the desk. In addition to the keyboard and mouse, there is a___

Confounders: keyboard → mouse → remote

Figure 4. Some examples of confounding propagation in LLaVA-1.5.

Qualitative Examples

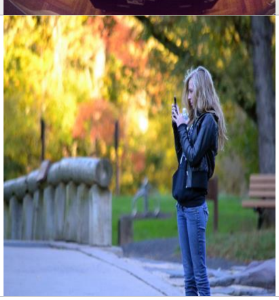
Image	Ours	MetaToken	HalLoc	SVAR
	...The kitchen is well equipped with various appliances, including a sink, oven, and refrigerator. There is a stove top oven situated in the middle of the kitchen, and a microwave is placed above it. The kitchen is also adorned with several decorative items, such as a vase , a bowl , and a couple of cups . A clock is mounted on the wall,...	...The kitchen is well equipped with various appliances, including a sink, oven, and refrigerator. There is a stove top oven situated in the middle of the kitchen, and a microwave is placed above it. The kitchen is also adorned with several decorative items, such as a vase , a bowl , and a couple of cups . A clock is mounted on the wall,...	...The kitchen is well equipped with various appliances, including a sink, oven, and refrigerator. There is a stove top oven situated in the middle of the kitchen, and a microwave is placed above it. The kitchen is also adorned with several decorative items, such as a vase , a bowl , and a couple of cups . A clock is mounted on the wall,...	...The kitchen is well equipped with various appliances, including a sink, oven, and refrigerator. There is a stove top oven situated in the middle of the kitchen, and a microwave is placed above it. The kitchen is also adorned with several decorative items, such as a vase , a bowl , and a couple of cups . A clock is mounted on the wall,...
	...The room has a cozy atmosphere, with a couch in the background and a dining table nearby. There are several items scattered around the room, including a cell phone, a laptop, and a couple of books. A handbag can also be seen placed on the floor...	...The room has a cozy atmosphere, with a couch in the background and a dining table nearby. There are several items scattered around the room, including a cell phone, a laptop, and a couple of books. A handbag can also be seen placed on the floor...	...The room has a cozy atmosphere, with a couch in the background and a dining table nearby. There are several items scattered around the room, including a cell phone, a laptop, and a couple of books. A handbag can also be seen placed on the floor.	...The room has a cozy atmosphere, with a couch in the background and a dining table nearby. There are several items scattered around the room, including a cell phone, a laptop, and a couple of books. A handbag can also be seen placed on the floor...
	...A backpack is also visible in the room, resting on the floor near the couch . A cat is present in the room,..	...A backpack is also visible in the room, resting on the floor near the couch . A cat is present in the room,..	...A backpack is also visible in the room, resting on the floor near the couch . A cat is present in the room,..	...A backpack is also visible in the room, resting on the floor near the couch . A cat is present in the room,..
	...The scene takes place in a park, with a bench visible in the background. There are several birds in the scene, with one bird perched on a nearby tree branch and two others flying in the sky. The presence of birds adds a sense of liveliness to the park setting...	...The scene takes place in a park, with a bench visible in the background. There are several birds in the scene, with one bird perched on a nearby tree branch and two others flying in the sky. The presence of birds adds a sense of liveliness to the park setting...	...The scene takes place in a park, with a bench visible in the background. There are several birds in the scene, with one bird perched on a nearby tree branch and two others flying in the sky. The presence of birds adds a sense of liveliness to the park setting...	...The scene takes place in a park, with a bench visible in the background. There are several birds in the scene, with one bird perched on a nearby tree branch and two others flying in the sky. The presence of birds adds a sense of liveliness to the park setting...

Figure 5. Qualitative Results.

	...The street is lined with a variety of street signs, including a "One Way" sign and a "No Standing" sign. The "One Way" sign is positioned on the left side of the street, while the "No Standing" sign is located further down the street. There are also a few traffic lights visible in the scene, with one on the left side of the street and another on the right side. The traffic lights are positioned at different heights,...	...The street is lined with a variety of street signs, including a "One Way" sign and a "No Standing" sign. The "One Way" sign is positioned on the left side of the street, while the "No Standing" sign is located further down the street. There are also a few traffic lights visible in the scene, with one on the left side of the street and another on the right side. The traffic lights are positioned at different heights,...	...The street is lined with a variety of street signs, including a "One Way" sign and a "No Standing" sign. The "One Way" sign is positioned on the left side of the street, while the "No Standing" sign is located further down the street. There are also a few traffic lights visible in the scene, with one on the left side of the street and another on the right side. The traffic lights are positioned at different heights,...	...The street is lined with a variety of street signs, including a "One Way" sign and a "No Standing" sign. The "One Way" sign is positioned on the left side of the street, while the "No Standing" sign is located further down the street. There are also a few traffic lights visible in the scene, with one on the left side of the street and another on the right side. The traffic lights are positioned at different heights,...
	...They are all dressed in white, which suggests a formal event. A man is holding a pair of scissors, ready to cut the ribbon,...	...They are all dressed in white, which suggests a formal event. A man is holding a pair of scissors, ready to cut the ribbon,...	...They are all dressed in white, which suggests a formal event. A man is holding a pair of scissors, ready to cut the ribbon,...	...They are all dressed in white, which suggests a formal event. A man is holding a pair of scissors, ready to cut the ribbon,...
	...The image features a beautiful butterfly perched on an orange, which is placed on a table. The butterfly is positioned on the orange,...	...The image features a beautiful butterfly perched on an orange, which is placed on a table. The butterfly is positioned on the orange,...	...The image features a beautiful butterfly perched on an orange, which is placed on a table. The butterfly is positioned on the orange,...	...The image features a beautiful butterfly perched on an orange, which is placed on a table. The butterfly is positioned on the orange,...
	...In the background, there are a few boats scattered across the water, adding to the lively atmosphere of the scene...	...In the background, there are a few boats scattered across the water, adding to the lively atmosphere of the scene...	...In the background, there are a few boats scattered across the water, adding to the lively atmosphere of the scene...	...In the background, there are a few boats scattered across the water, adding to the lively atmosphere of the scene...
	...The image features a man sitting on a bench in a hallway, possibly in a museum or an art gallery. He is wearing a suit and tie, and appears to be looking at a sculpture or a piece of artwork nearby....	...The image features a man sitting on a bench in a hallway, possibly in a museum or an art gallery. He is wearing a suit and tie, and appears to be looking at a sculpture or a piece of artwork nearby....	...The image features a man sitting on a bench in a hallway, possibly in a museum or an art gallery. He is wearing a suit and tie, and appears to be looking at a sculpture or a piece of artwork nearby....	...The image features a man sitting on a bench in a hallway, possibly in a museum or an art gallery. He is wearing a suit and tie, and appears to be looking at a sculpture or a piece of artwork nearby....
	...The image features a dining table with a plate of food, including a sandwich and a bowl of soup. The sandwich is cut...	...The image features a dining table with a plate of food, including a sandwich and a bowl of soup. The sandwich is cut...	...The image features a dining table with a plate of food, including a sandwich and a bowl of soup. The sandwich is cut...	...The image features a dining table with a plate of food, including a sandwich and a bowl of soup. The sandwich is cut...

Figure 6. More Qualitative Results.

	...The people are standing close to each other, creating a sense of unity and togetherness. In the background, there are several cars parked, indicating that the gathering might be taking place...	...The people are standing close to each other, creating a sense of unity and togetherness. In the background, there are several cars parked, indicating that the gathering might be taking place...	...The people are standing close to each other, creating a sense of unity and togetherness. In the background, there are several cars parked, indicating that the gathering might be taking place...	...The people are standing close to each other, creating a sense of unity and togetherness. In the background, there are several cars parked, indicating that the gathering might be taking place...
	...There are several other people in the scene, some of them standing near the ramp, while others are scattered around the area. A few of them are wearing backpacks, with one located near the center of the scene and another towards the right side...	...There are several other people in the scene, some of them standing near the ramp, while others are scattered around the area. A few of them are wearing backpacks, with one located near the center of the scene and another towards the right side...	...There are several other people in the scene, some of them standing near the ramp, while others are scattered around the area. A few of them are wearing backpacks, with one located near the center of the scene and another towards the right side...	...There are several other people in the scene, some of them standing near the ramp, while others are scattered around the area. A few of them are wearing backpacks, with one located near the center of the scene and another towards the right side...
	...Another person is visible in the background, but they are not the main focus of the scene. The room has a comfortable atmosphere, with a chair placed near the couch and a potted plant located in the corner. The couch occupies a significant portion of the room...	...Another person is visible in the background, but they are not the main focus of the scene. The room has a comfortable atmosphere, with a chair placed near the couch and a potted plant located in the corner. The couch occupies a significant portion of the room...	...Another person is visible in the background, but they are not the main focus of the scene. The room has a comfortable atmosphere, with a chair placed near the couch and a potted plant located in the corner. The couch occupies a significant portion of the room...	...Another person is visible in the background, but they are not the main focus of the scene. The room has a comfortable atmosphere, with a chair placed near the couch and a potted plant located in the corner. The couch occupies a significant portion of the room...
	...In addition to the hot dogs, there are two cups on the table, one near the left hot dog and the other near the right hot dog...	...In addition to the hot dogs, there are two cups on the table, one near the left hot dog and the other near the right hot dog...	...In addition to the hot dogs, there are two cups on the table, one near the left hot dog and the other near the right hot dog...	...In addition to the hot dogs, there are two cups on the table, one near the left hot dog and the other near the right hot dog...
	...A few individuals are holding cell phones, likely capturing the moment or taking pictures of the unique encounter...	...A few individuals are holding cell phones, likely capturing the moment or taking pictures of the unique encounter...	...A few individuals are holding cell phones, likely capturing the moment or taking pictures of the unique encounter...	...A few individuals are holding cell phones, likely capturing the moment or taking pictures of the unique encounter...

Figure 7. More Qualitative Results.

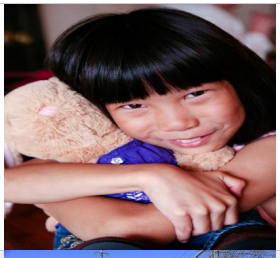
	The image features a young girl sitting on a chair, holding a brown teddy bear in her arms.	The image features a young girl sitting on a chair, holding a brown teddy bear in her arms.	The image features a young girl sitting on a chair, holding a brown teddy bear in her arms.	The image features a young girl sitting on a chair, holding a brown teddy bear in her arms.
	...The street is lined with various vehicles, including cars and a truck. There are two cars on the left side of the street, one behind the other...	...The street is lined with various vehicles, including cars and a truck. There are two cars on the left side of the street, one behind the other...	...The street is lined with various vehicles, including cars and a truck. There are two cars on the left side of the street, one behind the other...	...The street is lined with various vehicles, including cars and a truck. There are two cars on the left side of the street, one behind the other...
	...In addition to the sink and mirror, there is a toilet located in the background, slightly to the left of the sink...	...In addition to the sink and mirror, there is a toilet located in the background, slightly to the left of the sink...	...In addition to the sink and mirror, there is a toilet located in the background, slightly to the left of the sink...	...In addition to the sink and mirror, there is a toilet located in the background, slightly to the left of the sink...
	...In addition to the clock tower, there are two traffic lights visible in the scene, one on the left side and another on the right side of the image...	...In addition to the clock tower, there are two traffic lights visible in the scene, one on the left side and another on the right side of the image...	...In addition to the clock tower, there are two traffic lights visible in the scene, one on the left side and another on the right side of the image...	...In addition to the clock tower, there are two traffic lights visible in the scene, one on the left side and another on the right side of the image...
	...In addition to the people and the dog, there are two birds in the scene, one perched on a tree branch and the other flying in the sky. A handbag can be seen placed on the ground near the bench,...	...In addition to the people and the dog, there are two birds in the scene, one perched on a tree branch and the other flying in the sky. A handbag can be seen placed on the ground near the bench,...	...In addition to the people and the dog, there are two birds in the scene, one perched on a tree branch and the other flying in the sky. A handbag can be seen placed on the ground near the bench,...	...In addition to the people and the dog, there are two birds in the scene, one perched on a tree branch and the other flying in the sky. A handbag can be seen placed on the ground near the bench,...

Figure 8. More Qualitative Results.

References

[1] Laura Fieback, Jakob Spiegelberg, and Hanno Gottschalk. Metatoken: Detecting hallucination in image descriptions by meta classification. In *VISIGRAPP : VISAPP*, 2024. 3

[2] Zhangqi Jiang, Junkai Chen, Beier Zhu, Tingjin Luo, Yankun Shen, and Xu Yang. Devils in middle layers of large vision-language models: Interpreting, detecting and mitigating object hallucinations via attention lens. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 25004–25014, 2025. 3

[3] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online, 2020. Association for Computational Linguistics. 2

[4] Eunkyu Park, Minyeong Kim, and Gunhee Kim. Halloc: Token-level localization of hallucinations for vision language models. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 29893–29903, 2025. 3

[5] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2019. 1