

Supplementary Material

Logical Implications for Visual Question Answering Consistency

Sergio Tascon-Morales Pablo Márquez-Neila Raphael Sznitman
University of Bern

{sergio.tasconmorales, pablo.marquez, raphael.sznitman}@unibe.ch



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Is it a party?	No	←	Do the people appear to be at work?	Yes

Ans. None: Yes 🚨
Ans. SQuINT: Yes 🚨
Ans. CP-VQA: Yes 🚨
Ans. Ours: No

Ans. None: Yes
Ans. SQuINT: Yes
Ans. CP-VQA: Yes
Ans. Ours: Yes



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Is the sky clear?	Yes	↔	Are there clouds in the sky?	No

Ans. None: Yes
Ans. SQuINT: Yes
Ans. CP-VQA: Yes
Ans. Ours: Yes

Ans. None: Yes 🚨
Ans. SQuINT: Yes 🚨
Ans. CP-VQA: Yes 🚨
Ans. Ours: No

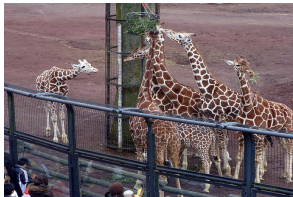


Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Is this animal in captivity?	Yes	←	Is there a fence behind the zebra?	Yes

Ans. None: Yes
Ans. SQuINT: No 🚨
Ans. CP-VQA: Yes
Ans. Ours: No 🚨

Ans. None: Yes
Ans. SQuINT: Yes
Ans. CP-VQA: Yes
Ans. Ours: Yes



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Is one of the giraffes a baby?	Yes	→	Are all of the giraffes adults?	No

Ans. None: Yes
Ans. SQuINT: Yes
Ans. CP-VQA: Yes
Ans. Ours: Yes

Ans. None: Yes 🚨
Ans. SQuINT: No
Ans. CP-VQA: Yes 🚨
Ans. Ours: No



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Are these elephants wild?	No	←	Are the elephants fenced in?	Yes

Ans. None: No
Ans. SQuINT: Yes 🚨
Ans. CP-VQA: No
Ans. Ours: No

Ans. None: Yes
Ans. SQuINT: Yes
Ans. CP-VQA: Yes
Ans. Ours: Yes



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Is the trolley overloaded?	No	←	Is the trolley normally loaded?	Yes

Ans. None: Yes 🚨
Ans. SQuINT: Yes 🚨
Ans. CP-VQA: No
Ans. Ours: Yes 🚨

Ans. None: Yes
Ans. SQuINT: Yes
Ans. CP-VQA: Yes
Ans. Ours: Yes

Figure 1. Additional qualitative examples from the Introspect dataset using BAN as backbone. Red siren symbols indicate inconsistent cases.



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Is it evening?	No	←	Is it sunny?	Yes

Ans. None: Yes 🚨
 Ans. SQuINT: Yes 🚨
 Ans. CP-VQA: Yes 🚨
 Ans. Ours: No

Ans. None: Yes
 Ans. SQuINT: Yes
 Ans. CP-VQA: Yes
 Ans. Ours: Yes



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Is the dog being friendly to the bird?	Yes	→	Is the dog biting the bird?	No

Ans. None: Yes
 Ans. SQuINT: Yes
 Ans. CP-VQA: Yes
 Ans. Ours: Yes

Ans. None: Yes 🚨
 Ans. SQuINT: No
 Ans. CP-VQA: Yes 🚨
 Ans. Ours: No

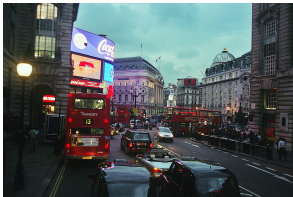


Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Is the pizza vegetarian?	Yes	↔	Is there any meat on the pizza?	No

Ans. None: No 🚨
 Ans. SQuINT: No 🚨
 Ans. CP-VQA: No 🚨
 Ans. Ours: Yes

Ans. None: No
 Ans. SQuINT: No
 Ans. CP-VQA: No
 Ans. Ours: No



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Is this a busy street?	Yes	↔	Is there a lot of traffic on the street?	Yes

Ans. None: Yes
 Ans. SQuINT: Yes
 Ans. CP-VQA: Yes
 Ans. Ours: Yes

Ans. None: Yes
 Ans. SQuINT: No 🚨
 Ans. CP-VQA: Yes
 Ans. Ours: Yes



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Is this meal vegan?	No	←	Is there meat on the plate?	Yes

Ans. None: No
 Ans. SQuINT: No
 Ans. CP-VQA: No
 Ans. Ours: Yes 🚨

Ans. None: Yes
 Ans. SQuINT: Yes
 Ans. CP-VQA: Yes
 Ans. Ours: Yes



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Is this a vegan dish?	Yes	→	Is there meat?	No

Ans. None: Yes
 Ans. SQuINT: Yes
 Ans. CP-VQA: Yes
 Ans. Ours: Yes

Ans. None: Yes 🚨
 Ans. SQuINT: No
 Ans. CP-VQA: No
 Ans. Ours: No



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Is the ground damp?	Yes	↔	Is the ground wet?	Yes

Ans. None: Yes
 Ans. SQuINT: No 🚨
 Ans. CP-VQA: No 🚨
 Ans. Ours: No 🚨

Ans. None: Yes
 Ans. SQuINT: Yes
 Ans. CP-VQA: Yes
 Ans. Ours: Yes

Figure 2. Additional qualitative examples from the Introspect dataset using BAN as backbone. Red siren symbols indicate inconsistent cases.



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Is this meal nutritious?	Yes	→	Is this food high in calories?	No

Ans. None: Yes
 Ans. SQuINT: Yes
 Ans. CP-VQA: Yes
 Ans. Ours: Yes

Ans. None: Yes 🚨
 Ans. SQuINT: Yes 🚨
 Ans. CP-VQA: No
 Ans. Ours: No



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Are the bananas ripe?	Yes	↔	Are the bananas ready to be eaten?	Yes

Ans. None: Yes
 Ans. SQuINT: No
 Ans. CP-VQA: No
 Ans. Ours: Yes

Ans. None: No 🚨
 Ans. SQuINT: No
 Ans. CP-VQA: No
 Ans. Ours: Yes



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Is this a sweet desert?	Yes	←	Does the desert have frosting?	Yes

Ans. None: No 🚨
 Ans. SQuINT: No 🚨
 Ans. CP-VQA: No 🚨
 Ans. Ours: Yes

Ans. None: Yes
 Ans. SQuINT: Yes
 Ans. CP-VQA: Yes
 Ans. Ours: Yes



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Is it wintertime?	No	→	Is there snow?	No

Ans. None: No
 Ans. SQuINT: No
 Ans. CP-VQA: No
 Ans. Ours: No

Ans. None: Yes 🚨
 Ans. SQuINT: No
 Ans. CP-VQA: No
 Ans. Ours: No



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Are these birds alive?	Yes	←	Are the birds eating?	Yes

Ans. None: Yes
 Ans. SQuINT: Yes
 Ans. CP-VQA: No 🚨
 Ans. Ours: Yes

Ans. None: Yes
 Ans. SQuINT: Yes
 Ans. CP-VQA: Yes
 Ans. Ours: Yes



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Are these animals real?	No	←	Are the animals made out of plastic?	Yes

Ans. None: Yes 🚨
 Ans. SQuINT: Yes 🚨
 Ans. CP-VQA: Yes 🚨
 Ans. Ours: No

Ans. None: Yes
 Ans. SQuINT: Yes
 Ans. CP-VQA: Yes
 Ans. Ours: Yes



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Do the girls appear to be in a home?	Yes	→	Is the child at school?	No

Ans. None: Yes
 Ans. SQuINT: Yes
 Ans. CP-VQA: Yes
 Ans. Ours: Yes

Ans. None: Yes 🚨
 Ans. SQuINT: Yes 🚨
 Ans. CP-VQA: Yes 🚨
 Ans. Ours: No

Figure 3. Additional qualitative examples from the Introspect dataset using BAN as backbone. Red siren symbols indicate inconsistent cases.



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Can people walk across the streets depicted in the image?	Yes	→	Is there a street?	Yes

Ans. None: Yes
 Ans. SQuINT: Yes
 Ans. CP-VQA: Yes
 Ans. Ours: Yes

Ans. None: No 🚨
 Ans. SQuINT: No 🚨
 Ans. CP-VQA: No 🚨
 Ans. Ours: Yes



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Does the weather appear to be pleasant?	No	←	Are the skies dark?	Yes

Ans. None: Yes 🚨
 Ans. SQuINT: Yes 🚨
 Ans. CP-VQA: Yes 🚨
 Ans. Ours: Yes 🚨

Ans. None: Yes
 Ans. SQuINT: Yes
 Ans. CP-VQA: Yes
 Ans. Ours: Yes



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Is this room decorated for the 1970s?	Yes	↔	Are the decorations consistent with the 1970's?	Yes

Ans. None: No 🚨
 Ans. SQuINT: No 🚨
 Ans. CP-VQA: Yes
 Ans. Ours: Yes

Ans. None: Yes
 Ans. SQuINT: Yes
 Ans. CP-VQA: Yes
 Ans. Ours: Yes



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Is the pizza hot?	Yes	↔	Is the cheese melted?	Yes

Ans. None: No 🚨
 Ans. SQuINT: No 🚨
 Ans. CP-VQA: No 🚨
 Ans. Ours: Yes

Ans. None: Yes
 Ans. SQuINT: Yes
 Ans. CP-VQA: Yes
 Ans. Ours: Yes



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Is the woman taking a selfie?	Yes	→	Is the woman holding a camera?	Yes

Ans. None: Yes
 Ans. SQuINT: Yes
 Ans. CP-VQA: Yes
 Ans. Ours: Yes

Ans. None: No 🚨
 Ans. SQuINT: Yes
 Ans. CP-VQA: No 🚨
 Ans. Ours: Yes



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Is the picture sepia?	Yes	→	Is the photo colorful?	No

Ans. None: Yes
 Ans. SQuINT: Yes
 Ans. CP-VQA: Yes
 Ans. Ours: Yes

Ans. None: Yes 🚨
 Ans. SQuINT: Yes 🚨
 Ans. CP-VQA: Yes 🚨
 Ans. Ours: Yes 🚨



Ground Truth

Question 1	Ans. 1	Relation	Question 2	Ans. 2
Can you ride this airplane?	No	←	Is the airplane smaller than the man?	Yes

Ans. None: Yes 🚨
 Ans. SQuINT: Yes 🚨
 Ans. CP-VQA: Yes 🚨
 Ans. Ours: No

Ans. None: Yes
 Ans. SQuINT: Yes
 Ans. CP-VQA: Yes
 Ans. Ours: Yes

Figure 4. Additional qualitative examples from the Introspect dataset using BAN as backbone. Red siren symbols indicate inconsistent cases.

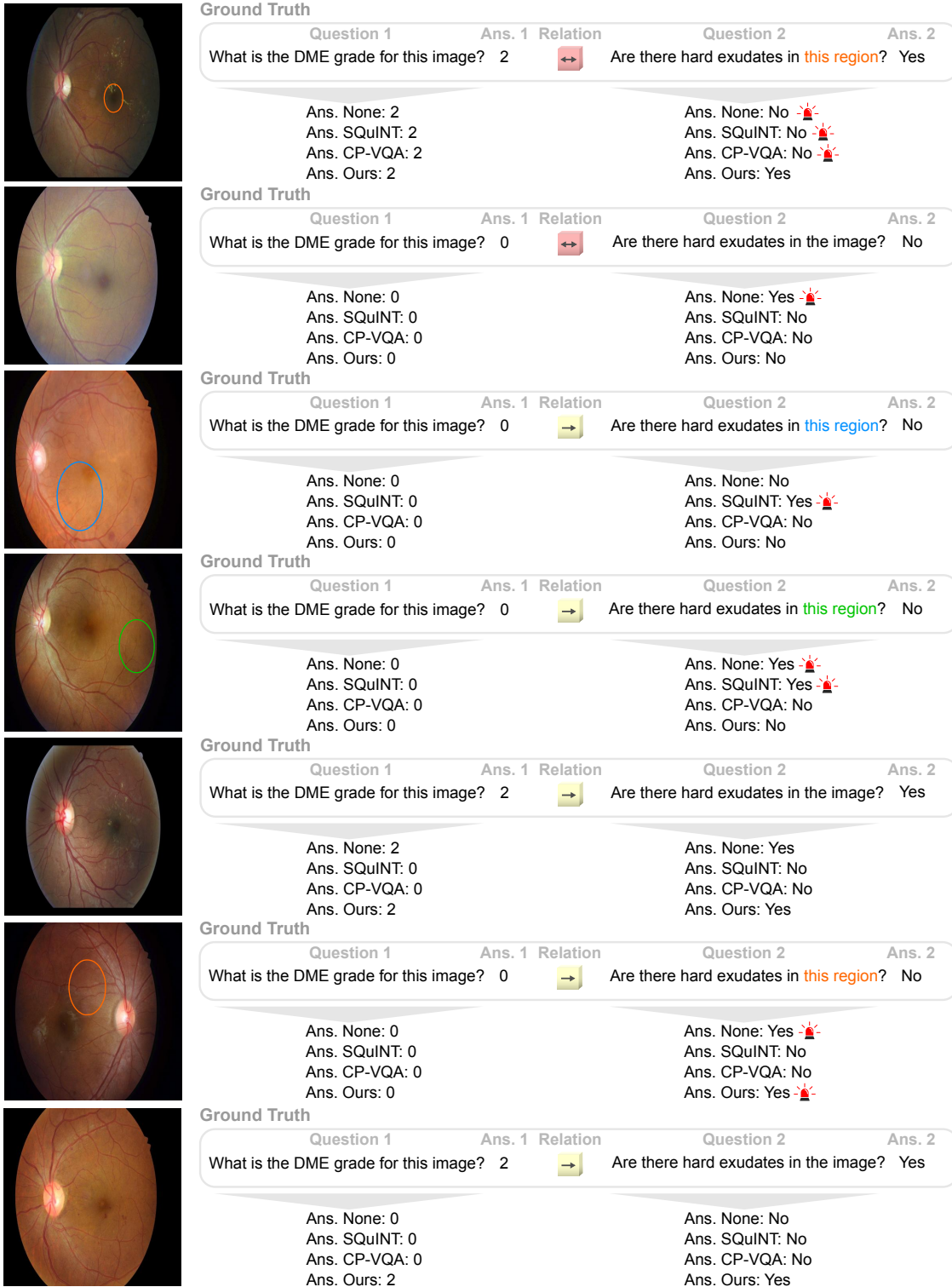


Figure 5. Additional qualitative examples from the DME dataset using MVQA as backbone. Red siren symbols indicate inconsistent cases. DME is a disease that is staged into grades (0, 1 or 2), which depend on the number of visual pathological features of the retina.

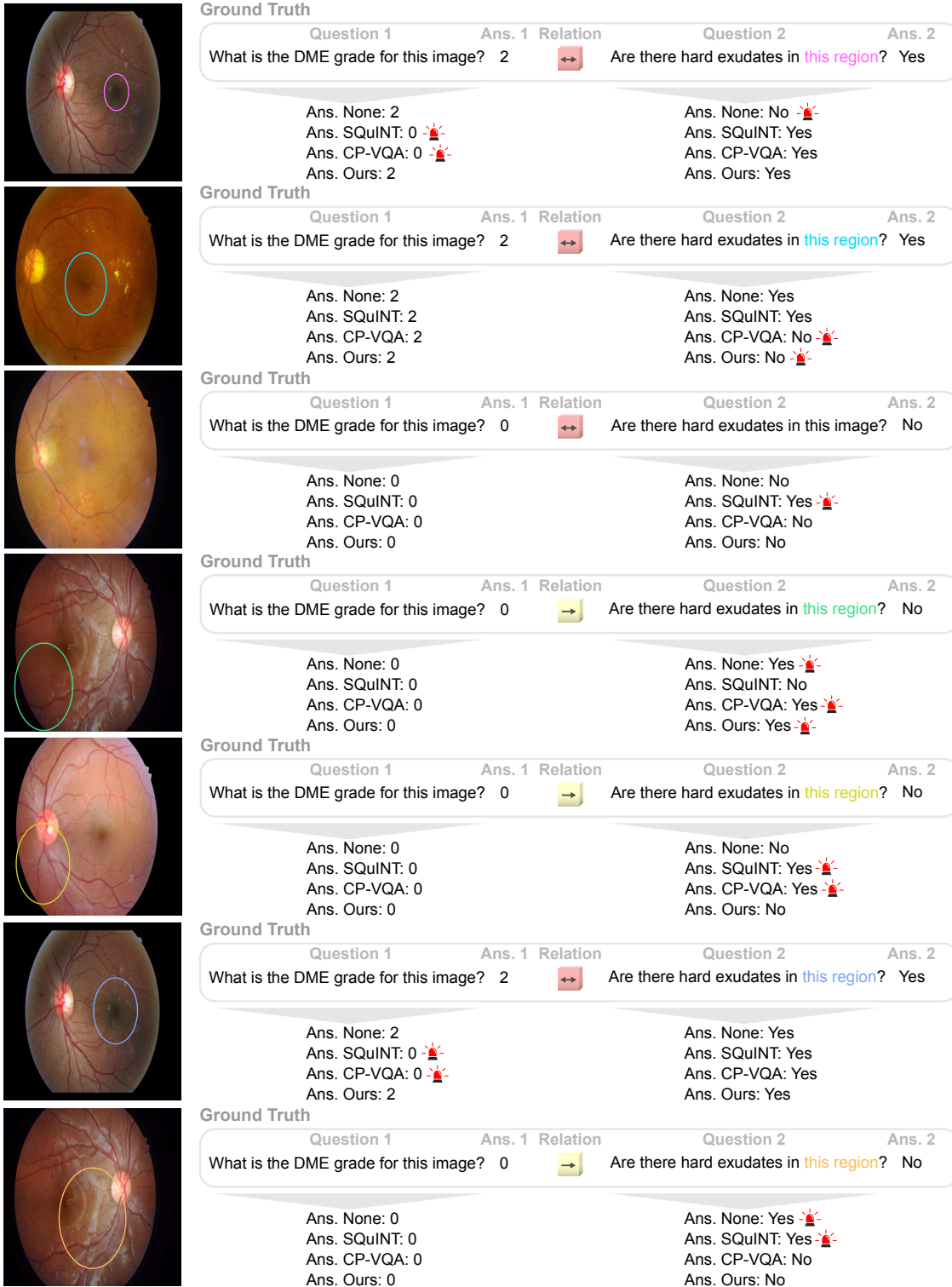


Figure 6. Additional qualitative examples from the DME dataset using MVQA as backbone. Red siren symbols indicate inconsistent cases. DME is a disease that is staged into grades (0, 1 or 2), which depend on the number of visual pathological features of the retina.