Supplementary material for “Attacking VQA Systems via Adversarial Background Noise”

This supplementary material contains the difference images for all the relevant images in the main paper. The difference images are scaled (scaling factor is in brackets) to make the difference apparent. This material also presents additional *successful* examples of the proposed attack.

1 Difference Images

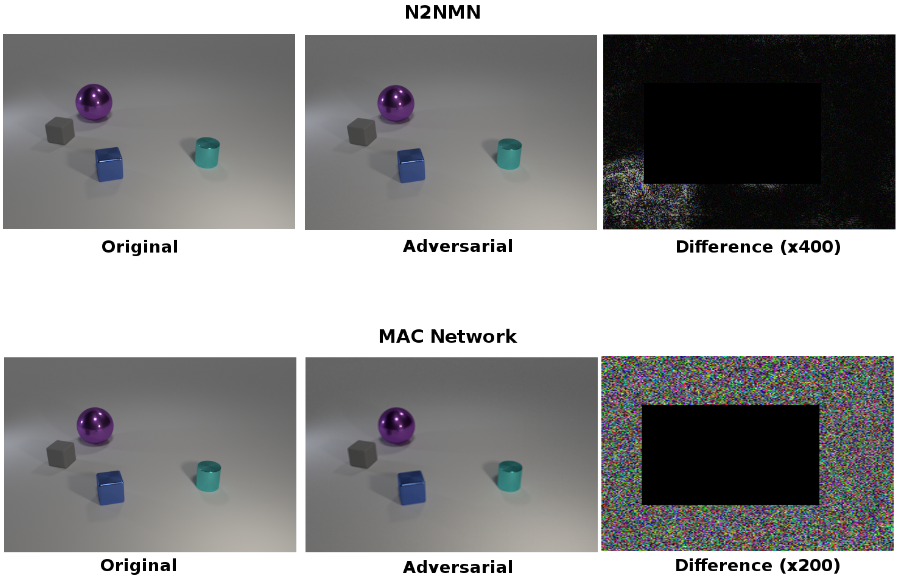


Figure 1: For the question “Are any big red matte things visible?”, both N2NMN and MAC network give the correct answer (“no”) when original image is given as input but incorrect answer (“yes”) when respective adversarial image is given as input.

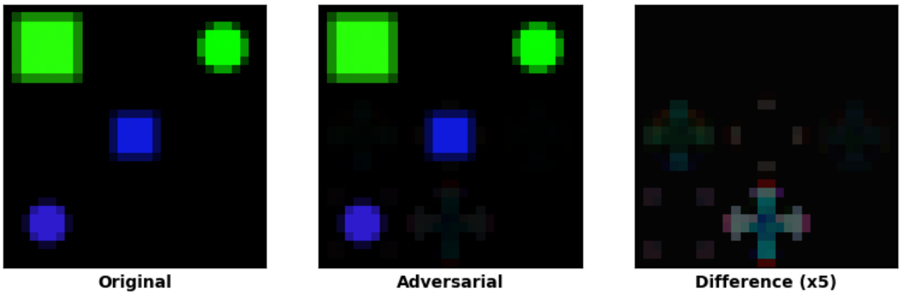


Figure 2: For the question “is a circle right of a blue shape?”, N2NMN gives the correct answer (“no”) when original image is given as input but incorrect answer (“yes”) when adversarial image is given as input.

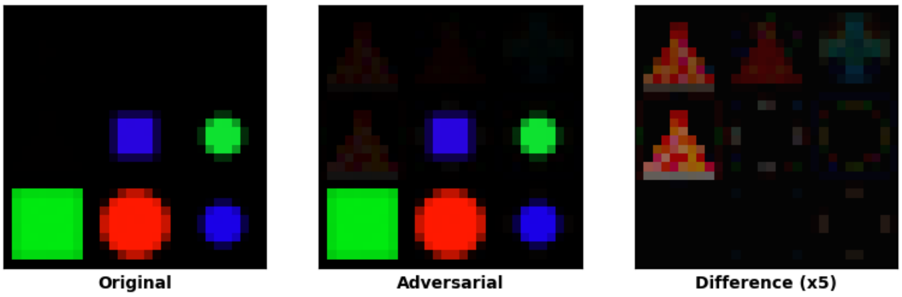


Figure 3: The answer for the question “is a blue shape right of a triangle?” changes from *no* to *yes* for the adversarial image.

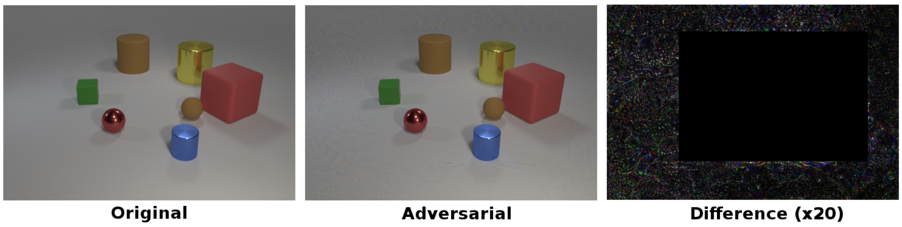


Figure 4: An example where the proposed attack gets target category correct. In this example, for the adversarial image, N2NMN predicts *cylinder* which belongs to the category of the target answer (cube) for the question “The big matte thing in front of the green rubber block is what color?”.

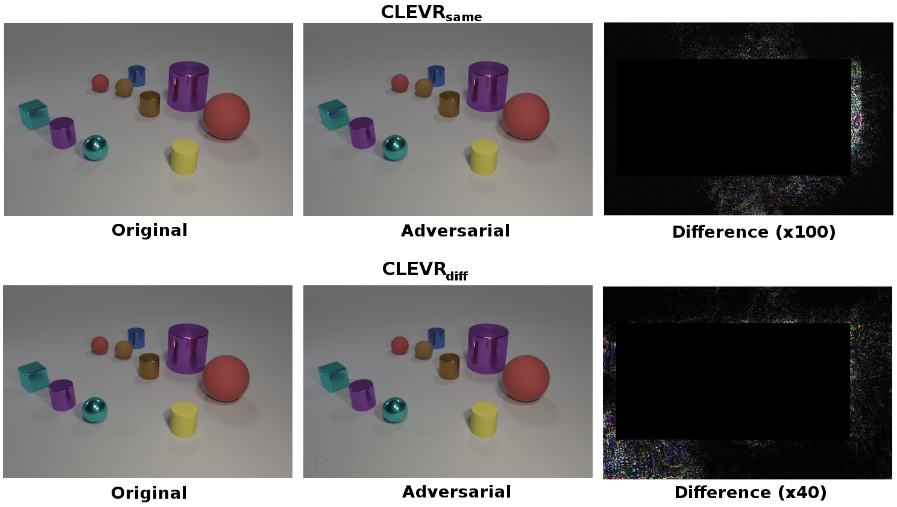


Figure 5: For the question “There is a big matte thing; what shape is it?”, N2NMN predicts *sphere* for the original image, *cylinder* for the adversarial image of CLEVR$_{\text{same}}$ and *metal* for the adversarial image of CLEVR$_{\text{diff}}$.

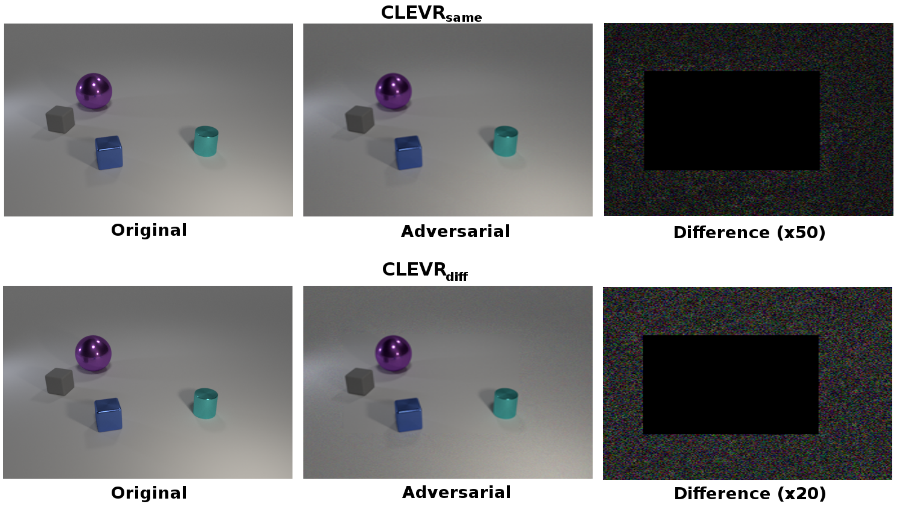


Figure 6: For the question "How many things are metal cylinders or metal blocks?", MAC network predicts 2 for the original image, 7 for the adversarial image of $\text{CLEVR}_{\text{same}}$ and *red* for the adversarial image of $\text{CLEVR}_{\text{diff}}$.

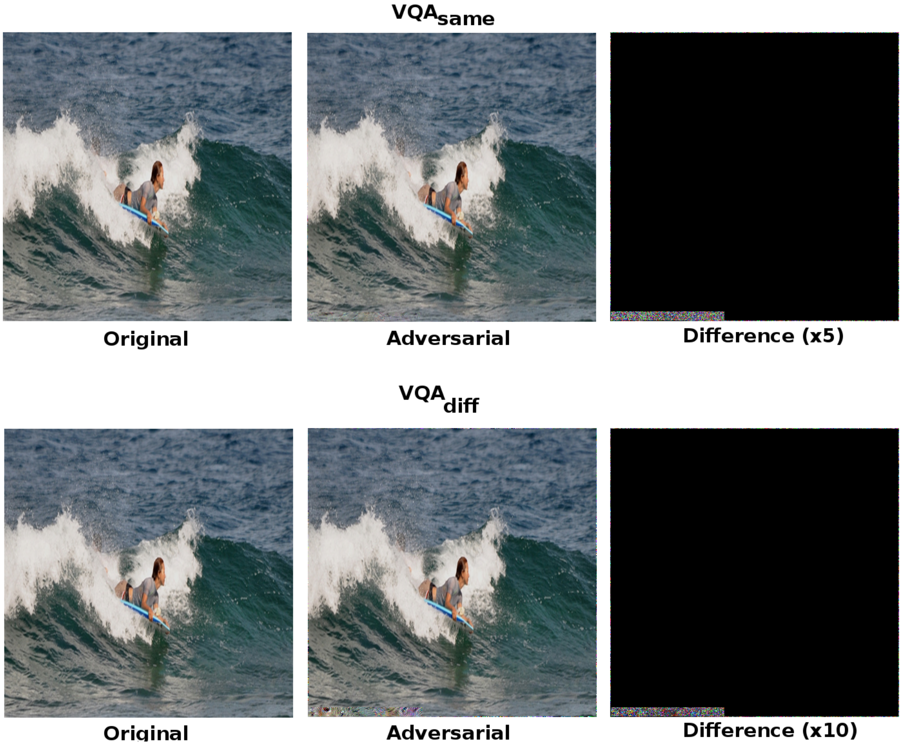


Figure 7: For the question "How many people are in the photo?", N2NMN predicts 1 for the original image, 6 for the adversarial image of VQA_{same} and *brown* for the adversarial image of VQA_{diff}.

2 More Examples of the proposed attack

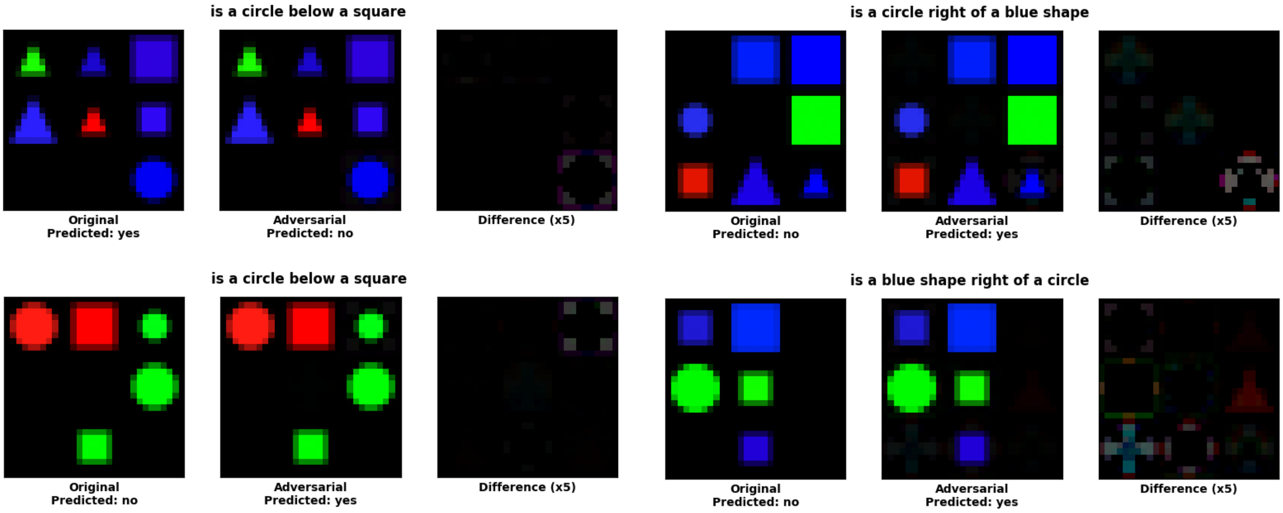


Figure 8: Examples from SHAPES

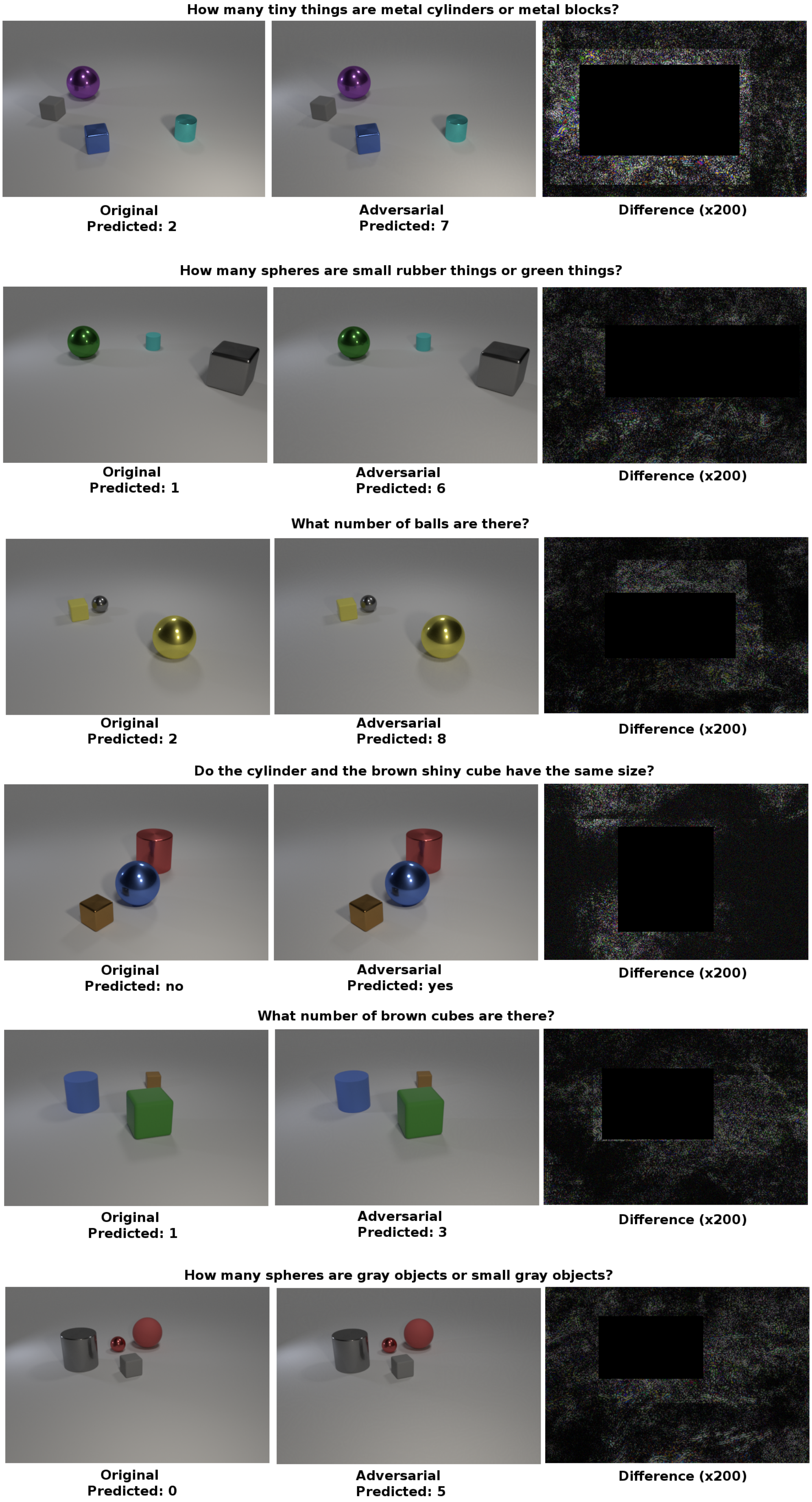
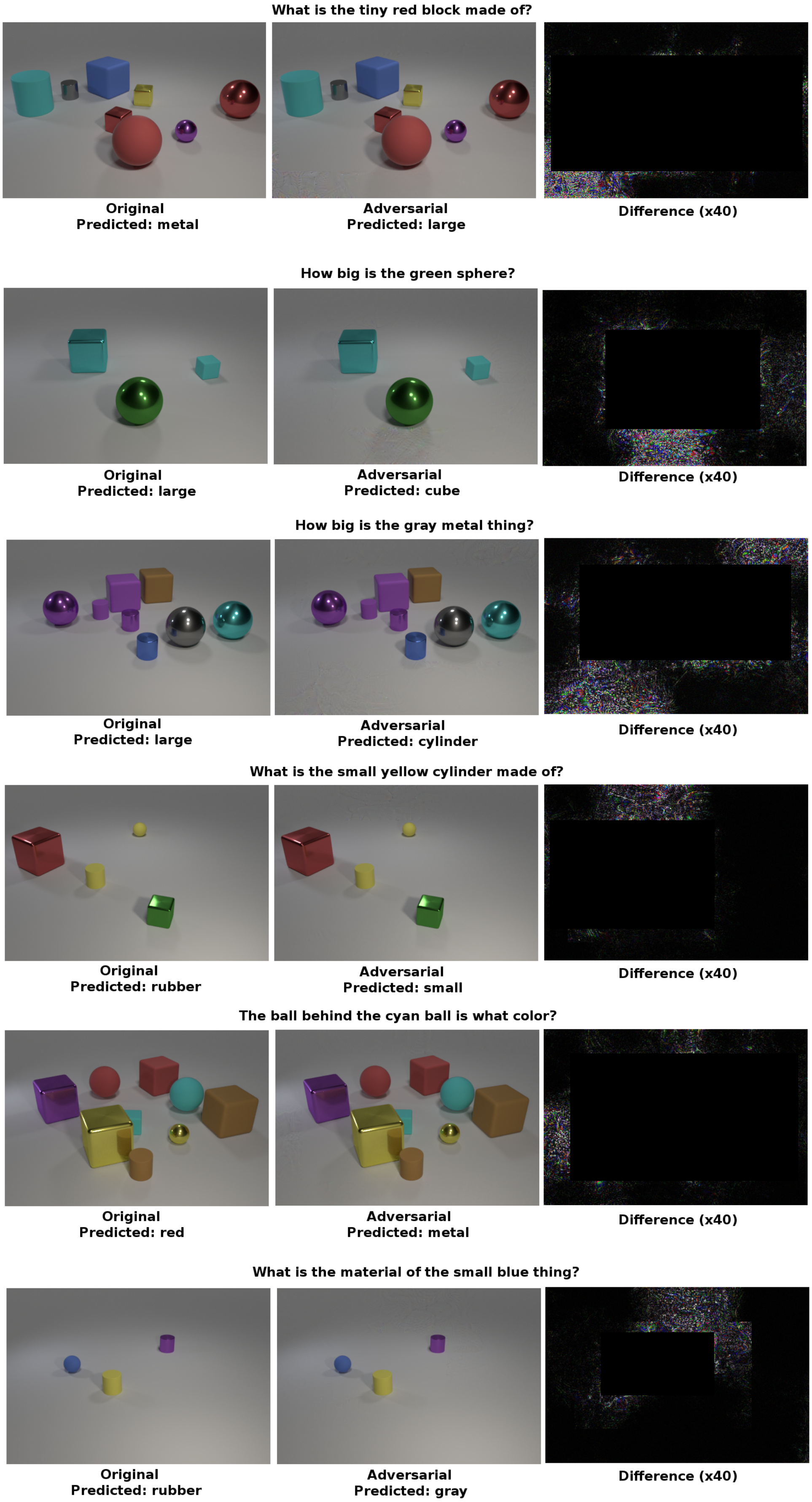


Figure 9: Examples for N2NMN on CLEVR$_{\text{same}}$



What is the tiny red block made of?

Original
Predicted: metal

Adversarial
Predicted: large

Difference (x40)

How big is the green sphere?

Original
Predicted: large

Adversarial
Predicted: cube

Difference (x40)

How big is the gray metal thing?

Original
Predicted: large

Adversarial
Predicted: cylinder

Difference (x40)

What is the small yellow cylinder made of?

Original
Predicted: rubber

Adversarial
Predicted: small

Difference (x40)

The ball behind the cyan ball is what color?

Original
Predicted: red

Adversarial
Predicted: metal

Difference (x40)

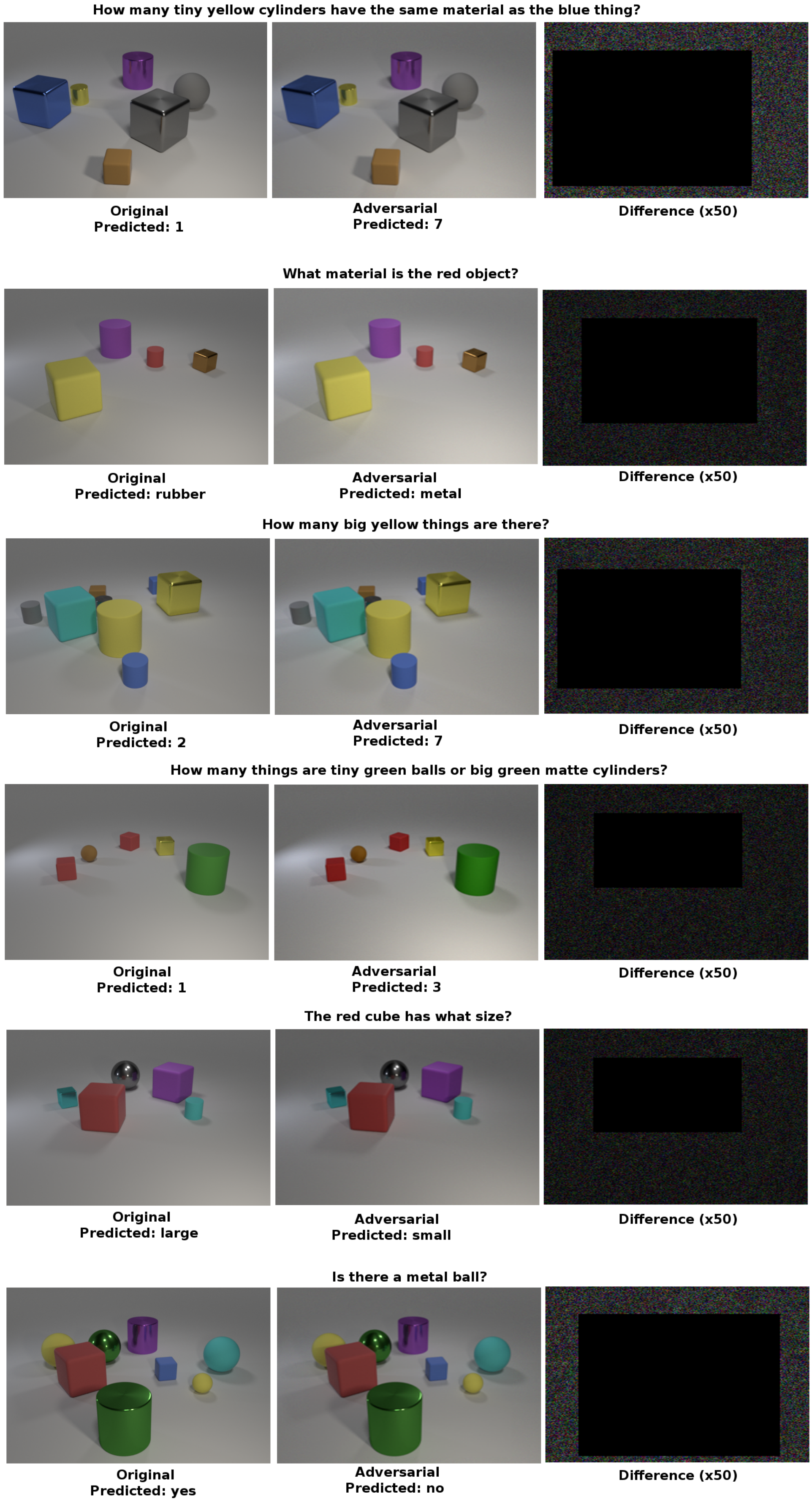
What is the material of the small blue thing?

Original
Predicted: rubber

Adversarial
Predicted: gray

Difference (x40)

Figure 10: Examples for N2NMN on $\text{CLEVR}_{\text{diff}}$



How many tiny yellow cylinders have the same material as the blue thing?

Original
Predicted: 1

Adversarial
Predicted: 7

Difference (x50)

What material is the red object?

Original
Predicted: rubber

Adversarial
Predicted: metal

Difference (x50)

How many big yellow things are there?

Original
Predicted: 2

Adversarial
Predicted: 7

Difference (x50)

How many things are tiny green balls or big green matte cylinders?

Original
Predicted: 1

Adversarial
Predicted: 3

Difference (x50)

The red cube has what size?

Original
Predicted: large

Adversarial
Predicted: small

Difference (x50)

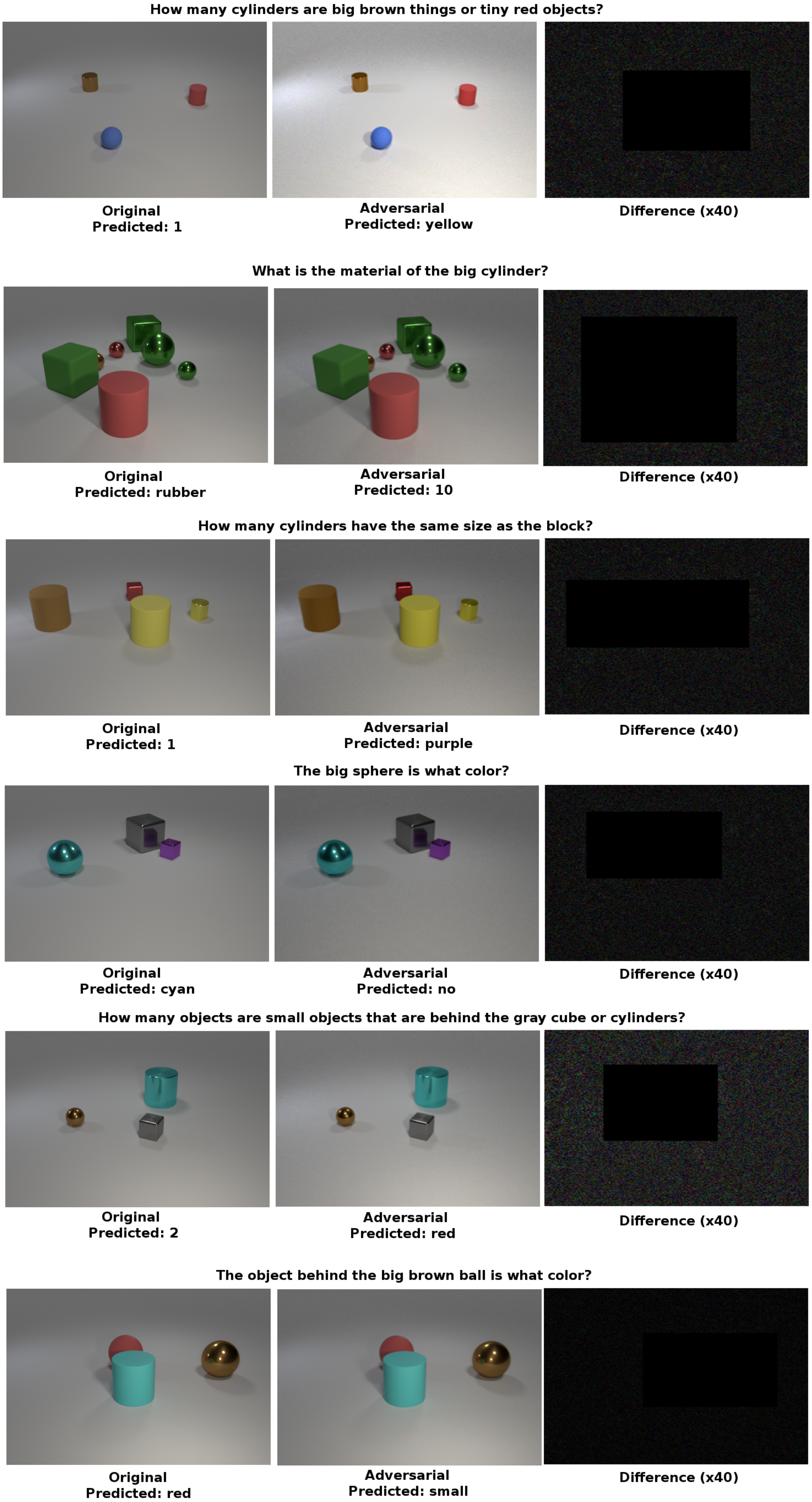
Is there a metal ball?

Original
Predicted: yes

Adversarial
Predicted: no

Difference (x50)

Figure 11: Examples for MAC network on CLEVR$_{\text{same}}$



How many cylinders are big brown things or tiny red objects?

Original
Predicted: 1

Adversarial
Predicted: yellow

Difference (x40)

What is the material of the big cylinder?

Original
Predicted: rubber

Adversarial
Predicted: 10

Difference (x40)

How many cylinders have the same size as the block?

Original
Predicted: 1

Adversarial
Predicted: purple

Difference (x40)

The big sphere is what color?

Original
Predicted: cyan

Adversarial
Predicted: no

Difference (x40)

How many objects are small objects that are behind the gray cube or cylinders?

Original
Predicted: 2

Adversarial
Predicted: red

Difference (x40)

The object behind the big brown ball is what color?

Original
Predicted: red

Adversarial
Predicted: small

Difference (x40)

Figure 12: Examples for MAC network on $\mathrm{CLEVR}_{\mathrm{diff}}$

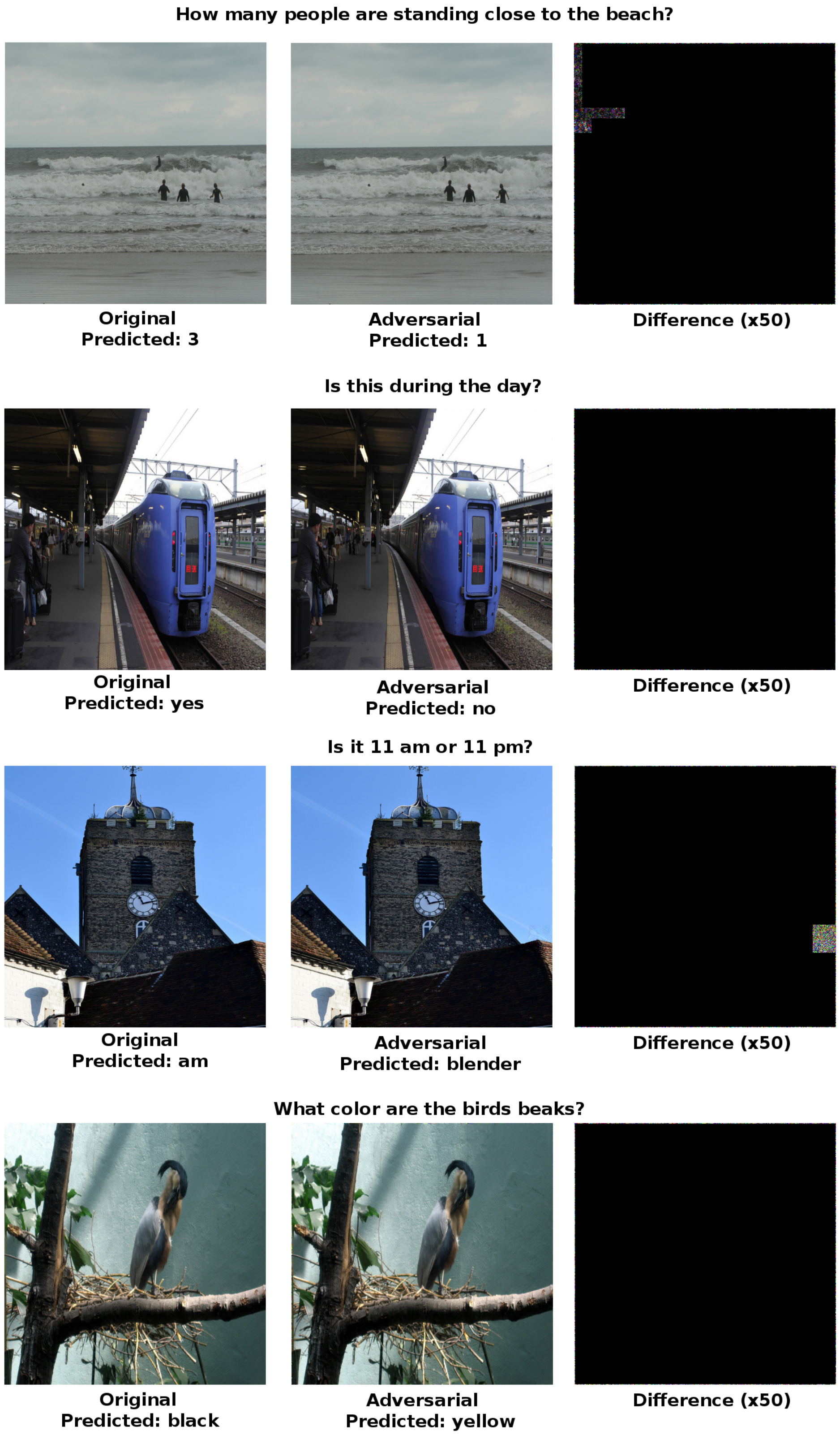


Figure 13: Examples for N2NMN on VQA_{same}

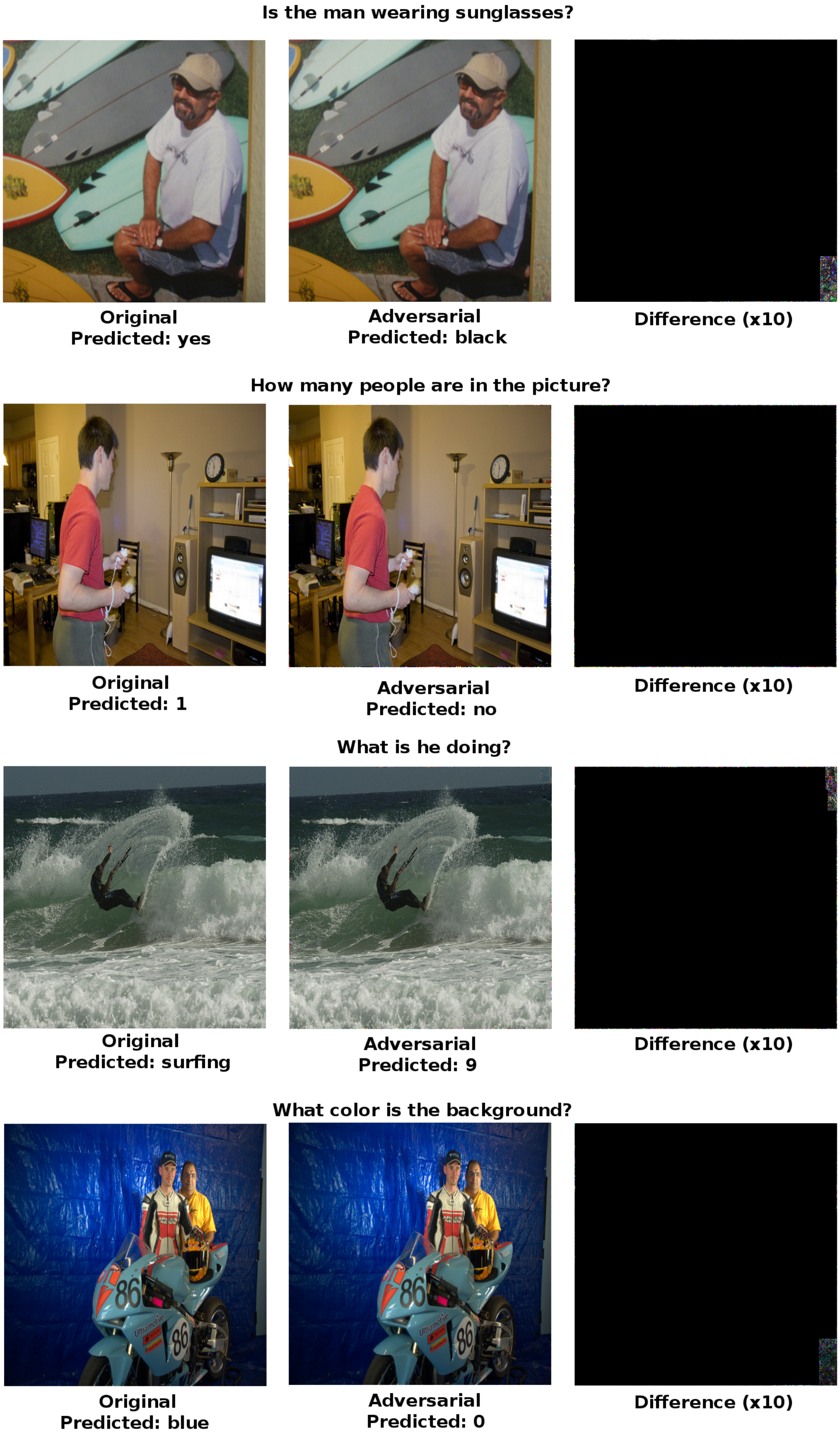


Figure 14: Examples for N2NMN on VQA$_{\text{diff}}$