Adversarial Patches

11/3/2025

60 Points Possible

Attempt 1

In Progress

NEXT UP: Submit Assignment

Add Comment

Unlimited Attempts Allowed

Details

Adversarial Attacks (Adversarial Patches)

Instructions

Create your own adversarial patch. We will use the Torchvision ResNet34 model trained on a small version of ImageNet to test your patch. You will need to use a class from the imagenet_classes.txt file (see GitHub) for your patch.

```
torchvision.models.resnet34(weights='IMAGENET1K_V1')
```

In addition to creating your adversarial patch, you must apply some creative component to your patch. Here are some ideas, but this is open-ended:

- “Disguise” the patch in a sticker, like they did in Brown, et.al.
- “Disguise” your patch in something else (ie clothing, jewelry, accessories, household items)
- Combine two patches into one and test the results (what happens?)
- Send a secret message using a series of patches

Submission

Your adversarial patch will be tested live in class, so bring a physical version of your patch (remember if color is something utilized in your patch, ensure you print it in color!)

If you would like to test your patch using the same setup we will in class, visit: [https://resnet34-classifier.streamlit.app/ (https://resnet34-classifier.streamlit.app/)](https://resnet34-classifier.streamlit.app/)

Submit a GitHub repository containing a Google Colab notebook to run your assignment. Any scripts or notebook should be well documented and easy to follow. Ensure your Google Colab notebook has the necessary documentation to run it. (See example in class repository).

Rubric

Presentation (20 points)

- A physical patch is brought on testing day
- The physical patch affects the results of the model in the intended manner

Code (25 points)

- Notebook/scripts are well documented and includes details and references to any material used
- Code implementing the adversarial techniques is clear and well documented

Creativity (15 points)

- The extension of the adversarial patch is creative

Edit View Insert Format Tools Table

12pt Paragraph

p

0 words

[(https://canvas.duke.edu/courses/62464/modules/items/572884)](https://canvas.duke.edu/courses/62464/modules/items/572884)

Assignment

[(https://canvas.duke.edu/courses/62464/modules/items/572883)](https://canvas.duke.edu/courses/62464/modules/items/572883)