Final Project Progress Report

Important: In your report, please 1) Make it very clear who on the team contributed what, and 2) Include the dates of at least *two* meetings you had with your mentor.
Team name: ***Pixel Pioneers***
TA name: ***Hannah***

Note: when submitting this document to Gradescope, make sure to add all other team members to the submission. This can be done on the submission page after uploading.

Progress Report Instructions

Before writing your progress report, you should have met with your TA and talked through your progress.

Team contributions

Please describe in one paragraph (3–4 sentences) per team member what each of you contributed to the project so far.

Person 1 Completed a comprehensive review of over 20 research papers and articles focusing on adversarial attack methods and their transferability across models; Identified key techniques that enhance transferability, including gradient-based methods and decision boundary analysis; Presented a summary of defense mechanisms and their effectiveness against state-of-the-art adversarial attacks.

Person 2 Through research, confirm the types of popular neural network models. Then, write code to build and run part of the neural network models for image classification tasks. Collect a sufficient amount of image samples from different categories, and preprocess these images. Use the neural network to conduct a preliminary validation of the classification effectiveness on these unattacked images.

Person 3 I went through a significant amount of literature to learn about different transfer attack methods. This included digging into gradient-based attacks, which are pretty common, and also checking out input transformation-based attacks. Along the way, I also summarized some of the newer algorithms designed to create adversarial examples. This deep dive has given me a thorough understanding of how our model is structured and has helped me identify which attack methods might be most effective for us to incorporate.