Alexnet accuracy on cifar 10: 0.7

Image testing 2021 (correct: ship): AlexNet predicts with Ship, adversarial input also correctly predicted with ship



Image testing 2024 (correct: frog): AlexNet predicts with frog, adversarial input 21: frog, adversarial input 46: frog



Testing with a high confidence interval attack:



Image testing 2026 (correct: automobile): AlexNet predicts with automobile, adversarial input 37: automobile, adversarial 54: automobile



Image testing, resnet mispredicts as cat, alexnet correctly predicts it as a dog.



Image testing (frog, no number): Alexnet predicts frog, adversarial input 40: frog, 47: frog

From the vidnerova paper: Observations above indicate that adversarial examples are more likely to be shared between similar models, for example between individual variants of SVM. Models that are of different nature, like convolutional network or DT, share adversarial examples with the rest of models very rarely