FALL 2023 FINALS.

QUESTION 1:

Adding noise to images of sneakers could cause the neural network to misclassify them as t-shirts for a variety of reasons, including feature perturbation, adversarial attacks, and training data bias.

Feature Perturbation; neural networks rely on learning distinguishing features to make accurate predictions. If the introduced noise obscures or distorts the essential features that distinguish t-shirts from trainers, the network may struggle to identify the correct class.

Another thing to think about is the possibility of Adversarial Attacks. These involve deliberately modifying input data in order to deceive a neural network. The additional noise could be designed to exploit network vulnerabilities, causing the network to misinterpret the altered sneaker image as a t-shirt. Adversarial attacks frequently take advantage of neural networks' sensitivity to small, carefully designed perturbations.

Additionally, **Training Data Bias** could contribute to misclassifications. If the data used to train the neural network is biased or lacks diversity, the network may develop a skewed understanding of the defining features of sneakers or t-shirts. In such instances, the introduced noise might capitalize on these biases, resulting in inaccurate classifications.

QUESTION 2:

Adding the same noise (delta) to different images of sneakers can influence the neural network's performance and predictions in several ways and possible outcomes can be:

Consistent Misclassifications: If the noise consistently disrupts the critical features on which the neural network relies to distinguish trainers from other categories, it may result in consistent misclassifications. Simply put, the neural network may consistently identify the modified sneaker images as belonging to a different category, such as t-shirts.

Impact on Pattern Recognition: The added noise may impair the neural network's ability to identify recurring patterns across different sneaker images. If the noise consistently alters or obscures essential features, the network might generalize this pattern across different instances of sneakers, causing a consistent trend of misclassifications.

Adversarial Influence: In the case of an adversarial attack, where the noise is designed to exploit flaws in the neural network's decision-making boundaries, applying this noise to different sneaker images repeatedly may deceive the network into misclassifying them. Adversarial attacks are typically designed to cause disruptions that consistently result in the desired misclassification outcome. The impact of the noise on the network's predictions can vary based on factors such as the nature of the noise, the architecture of the neural network, and the robustness of the model.

QUESTION 3:

Enhancing the robustness of a neural network against adversarial attacks or noise requires implementing specific strategies. Here are some common approaches:

1. Adversarial Training: Train the neural network by exposing it to adversarial examples. Introduce noisy or perturbed images during the training process and incorporate them into the training dataset. This enables the model to learn how to recognize and adapt to such perturbations, improving its overall robustness.

2. Regularization Techniques: Implement regularization techniques during training, such as dropout or weight regularization. These methods help prevent the neural network from overly relying on specific features, promoting better generalization to inputs that may be perturbed or noisy.

3. Use Robust Activation Functions: Opt for activation functions that exhibit reduced sensitivity to small changes in input, such as the rectified linear unit (ReLU). Certain activation functions are more robust and can assist the network in effectively handling input with added noise or perturbations.