1. Adversial attack should be using few images from ‘cifar 10’ or prepared 10 images and

Attack them and check if it works…

Or using the whole cifar 10\_test dataset to do attack and average how much percentage got ‘fooled’.

2. Checking attention weights to compare???

Before attack (parameters)

After attack (changed parameters)