ributions of BSD at 1.0 Detection Threshold L2CarliniWagnerAttack on C 100% 80% Classification accuracy 60% JSD 10 Accuracy JSD 40 Accuracy 40% MagNet Accuracy BSD Accuracy Class Divergence Detector All JSD detectors 20% Distance Based Detections No Defense Accuracy 0% 20 40 60 80 100

Confidence in L2CarliniWagnerAttack