Supplementary material for Invisibility Cloak: Hiding Anomalies in Videos via Adversarial Machine Learning Attacks

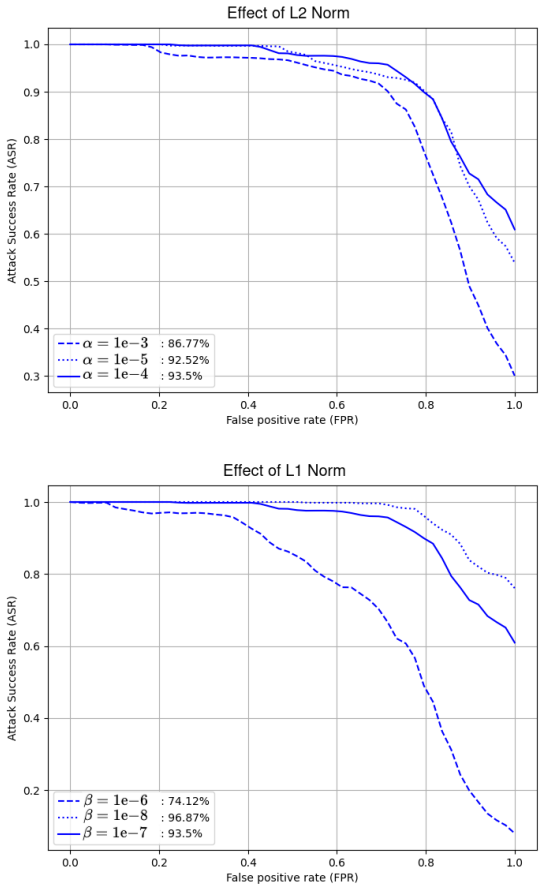

Figure 1. Attack success rate plots for effect of L2-norm and L1-norm regularization under the condition S3R→S3R|UCF→UCF presented in the main paper under section 6.2

A. Effect of α and β Plots

We provide the plots for the results showcasing the effect of α and β corresponding to Table 5 in the main paper. Figure 1 shows the plots for the effect of L2-norm and L1-norm regularization on the performance of the attack under the condition S3R→S3R|UCF→UCF. We see that L1 norm has a significant effect on the performance.

B. Transferability Plots

In Section 5.3, we discuss the transferability of the attack across model and data domain. Due to limited space in the main paper, we provide here all success rate vs. false positive rate plots for cross-model and cross-data domain analysis. Figure 2 shows plots for the transferability between all models while Figure 3 shows the plots for the transferability across data domains for each attack.

C. Demo

A demo code for our method can be accessed at the following link: https://drive.google.com/file/d/1RbTK4OFKj9vueUSym_Tdzc8CiyJaqH_r/view?usp=sharing. The zip file contains 3 folders: "MGFN_UCF", "S3R_UCF" and "REWARD_UCF". These folders contain all the pre-trained attack models as discussed in Section 5.2 in the main paper trained on UCF Crime. We provide pre-trained S3R, MGFN and REWARD models that can be used as target models. We also provide a video sample to attack, it can be found in the CLOAK folder as "sample_video.avi". Edit the script "test_video_sample.py" to try the demo. Edit the code as follows:

Line 253: M1 = "Attack model name" (Edit here to select the attack model M1)

Line 254: M2 = "S3R" (Edit here to select the target model M2)

"Attack model name" can be found in either of the 3 folders mentioned above. The output of the demo will produce the perturbed video, the perturbation mask and a plot illustrating the anomaly scores before and after the attack. You can choose between "S3R" and "MGFN" and "REWARD" for M2.

Libraries Required: numpy, torch, matplotlib, einops, opencv.

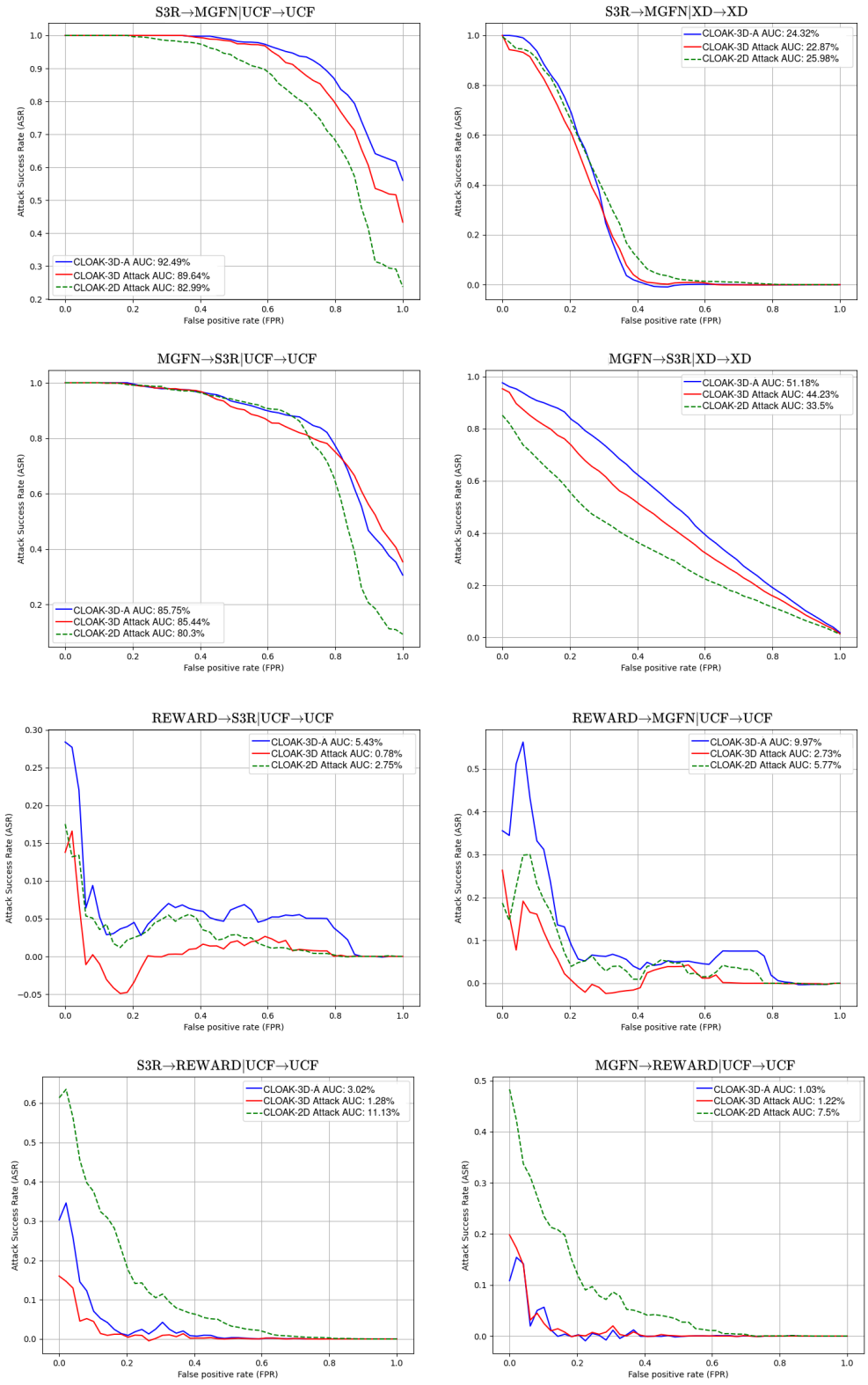

Figure 2. Attack success rate plots for all the cross-model transferability results presented in the main paper in Section 5.3.

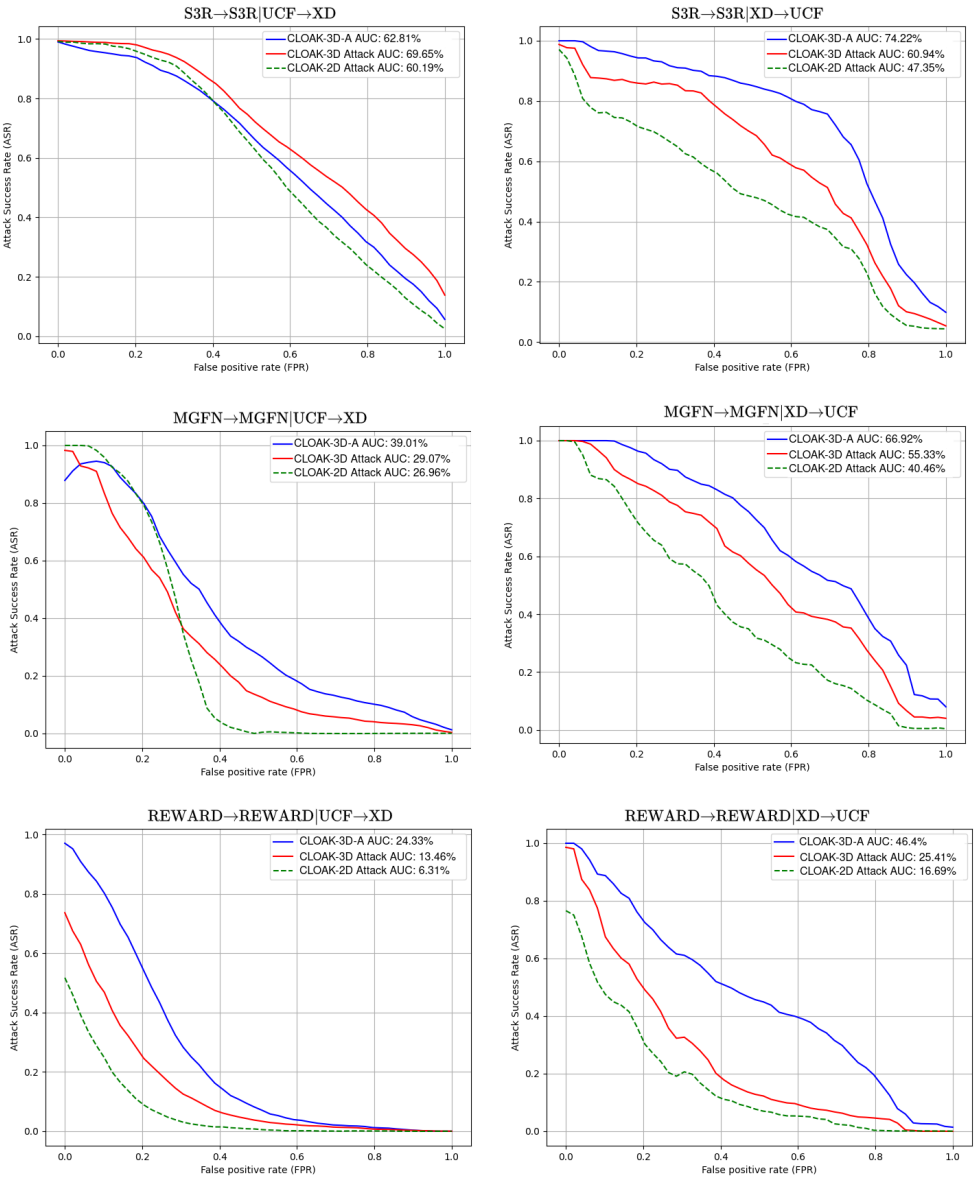

Figure 3. Attack success rate plots for all the cross-data domain transferability results presented in the main paper in Section 5.3.