

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

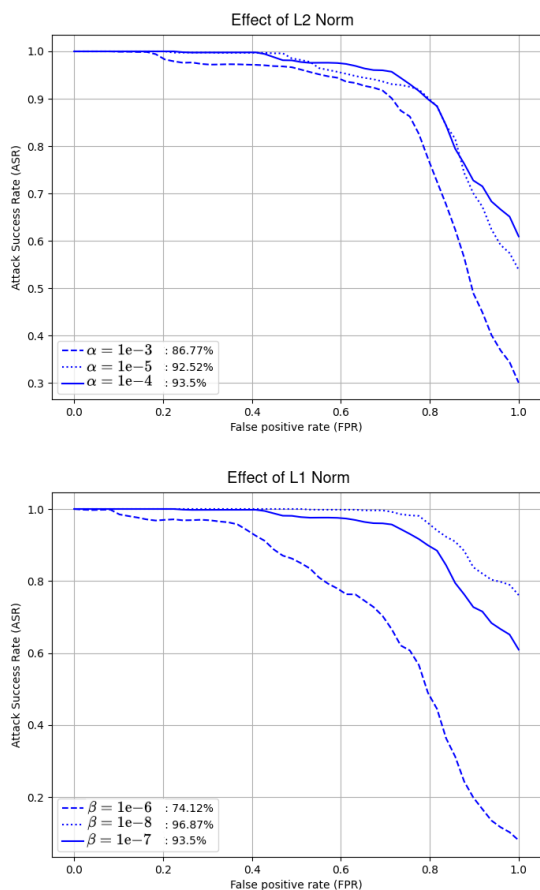


Figure 1. Attack success rate plots for effect of L2-norm and L1-norm regularization under the condition $S3R \rightarrow S3R | UCF \rightarrow UCF$ presented in the main paper under section 6.2

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 $S3R \rightarrow S3R | UCF \rightarrow UCF$. 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/1RbTK40FKj9vueUSym_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.

054
055
056
057
058
059
060
061
062
063
064
065
066
067
068
069
070
071
072
073
074
075
076
077
078
079
080
081
082
083
084
085
086
087
088
089
090
091
092
093
094
095
096
097
098
099
100
101
102
103
104
105
106
107

108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161

162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215

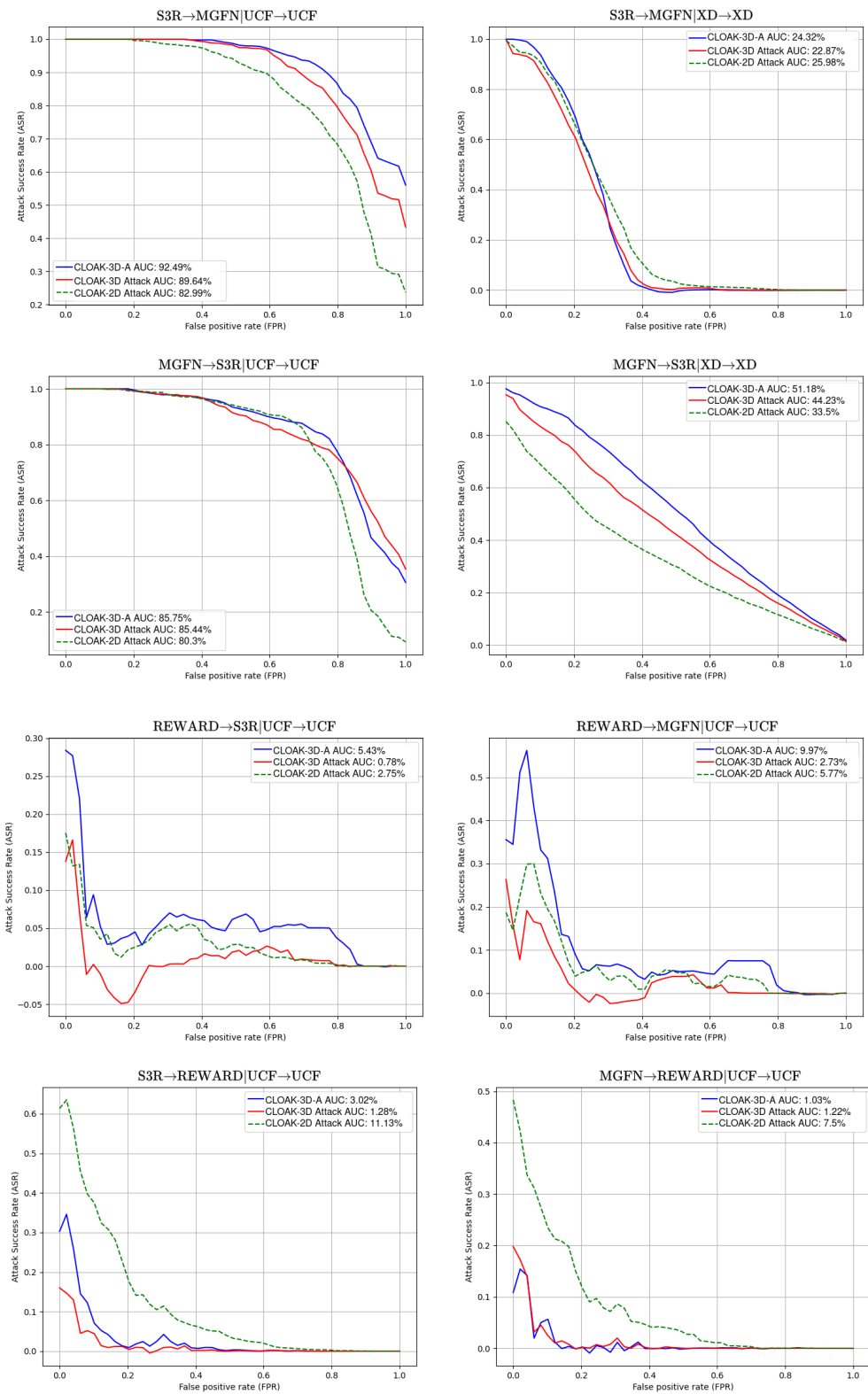


Figure 2. Attack success rate plots for all the cross-model transferability results presented in the main paper in Section 5.3.

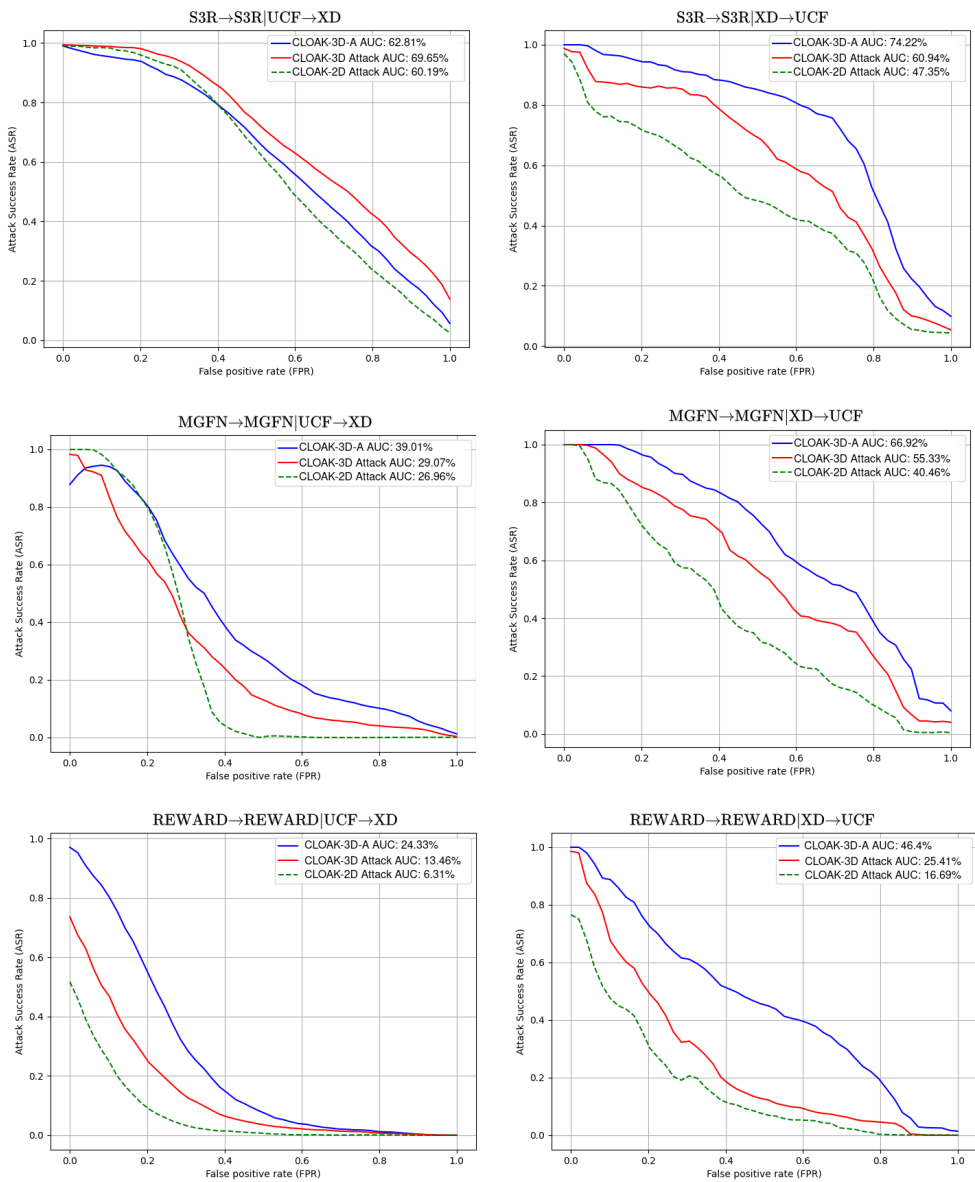


Figure 3. Attack success rate plots for all the cross-data domain transferability results presented in the main paper in Section 5.3.

216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269

270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323