An anomaly detection approach for backdoored neural networks: face recognition as a case study

Alex Unnervik

Original dataset

Training

Poisoned samples

An anomaly detection approach for backdoored neural networks: face recognition as a case study, Alexander Unnervik and Sébastien Marcel, BIOSIG 2022
Arxiv: 2208.10231
Backdoor attack results

An anomaly detection approach for backdoored neural networks: face recognition as a case study, Alexander Unnervik and Sébastien Marcel, BIOSIG 2022
Arxiv: 2208.10231
Backdoor attack detection method

Detector architecture

Results on first set of backdoored networks (loc)

Results on first set of backdoored networks (trig)