QESAR: Query Effective Decision-based Attack on Skeletal Action Recognition Supplemental Document

Zi Kang[0000−0002−1863−2802], Yumei Zhang[0009−0005−0003−6948], Rui Zhang[0000−0002−4117−2656], Yanan Jiang[0000−0002−2377−0492], and Hui Xia(✉)[0000−0001−7326−5796]

Computer Science and Technology, Ocean University of China, Qingdao 266100, China
kangzi@stu.ouc.edu.cn, zym8004@stu.ouc.edu.cn, zhangrui0504@stu.ouc.edu.cn, jyniw0923@163.com, xiahui@ouc.edu.cn

1 Additional Content

The supplementary materials include content that needs to be included in the main body of the paper due to space constraints.

1.1 Hierarchical Joint Perturbation

We select appropriate perturbation directions based on the skeletal partition diagram shown in Fig.1 to ensure more accurate gradient direction estimation.

1.2 Impact of Hierarchical Joint Perturbation

Fig.2, Fig.3, and Fig.4 illustrate the impact of the hierarchical joint perturbation method on imperceptibility and query volume under different settings. In Fig.2, the joint position deviation, joint position acceleration deviation, joint angle acceleration deviation, and query of adversarial examples generated by QESAR with hierarchical joint perturbation and without hierarchical joint perturbation for ST-GCN and SGN action recognition models in the targeted attack setting on the HDM05 dataset are compared. Fig.3 shows the joint position deviation, joint position acceleration deviation, joint angle acceleration deviation, and query of adversarial examples generated by QESAR with hierarchical joint perturbation and without hierarchical joint perturbation for ST-GCN and SGN models in the untargeted attack setting on the NTU dataset. Fig.4 presents the joint position deviation, joint position acceleration deviation, joint angle acceleration deviation, and query of adversarial samples generated by QESAR with hierarchical joint perturbation and without hierarchical joint perturbation for ST-GCN and SGN models in the targeted attack setting on the NTU dataset. From Fig.2, Fig.3, and Fig.4, it can be observed that the curves of QESAR with hierarchical joint perturbation are lower than those of QESAR without hierarchical joint perturbation, indicating that QESAR with hierarchical joint perturbation improves the imperceptibility of adversarial examples at lower query volumes.

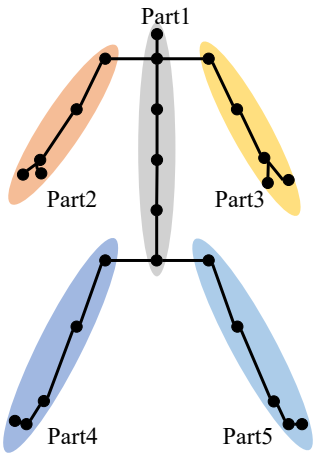
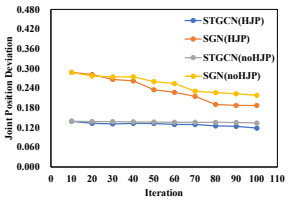
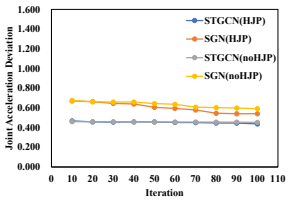


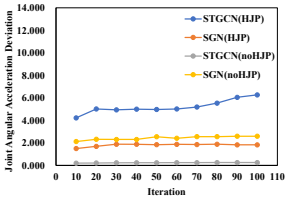
Fig. 1: Skeletal Partition Diagram. The body is divided into limbs and torso based on the different functions of body regions.



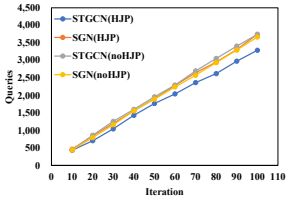
(a) Joint position deviation



(b) Joint position acceleration deviation

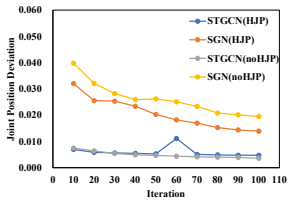


(c) Joint angular acceleration deviation

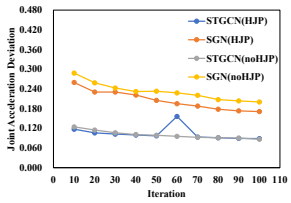


(d) Queries

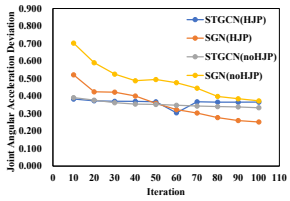
Fig. 2: Under the targeted attack settings, we compare the joint position deviation, joint position acceleration deviation, joint angle acceleration deviation, and query of adversarial examples generated by QESAR with and without hierarchical joint perturbation for the ST-GCN and SGN action recognition models on the HDM05 dataset. In the comparison, 'HJP' refers to QESAR with hierarchical joint perturbation, and 'noHJP' refers to QESAR without hierarchical joint perturbation.



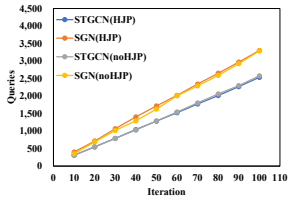
(a) Joint position deviation



(b) Joint position acceleration deviation

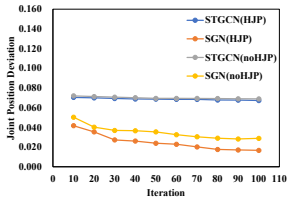


(c) Joint angular acceleration deviation

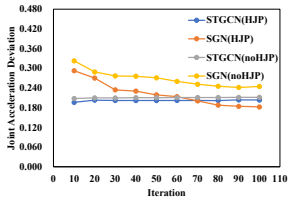


(d) Queries

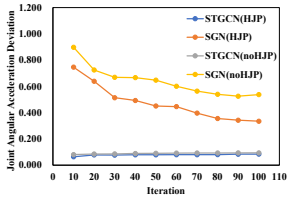
Fig. 3: Under the untargeted attack settings, we compare the joint position deviation, joint position acceleration deviation, joint angle acceleration deviation, and query of adversarial examples generated by QESAR with and without hierarchical joint perturbation for the ST-GCN and SGN action recognition models on the NTU dataset.



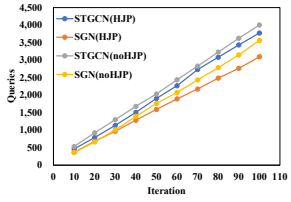
(a) Joint position deviation



(b) Joint position acceleration deviation



(c) Joint angular acceleration deviation



(d) Queries

Fig. 4: Under the targeted attack settings, we compare the joint position deviation, joint position acceleration deviation, joint angle acceleration deviation, and query of adversarial examples generated by QESAR with and without hierarchical joint perturbation for the ST-GCN and SGN action recognition models on the NTU dataset.