

100%

Despite neural network-based speaker recognition systems (SRS) have enjoyed significant success, they are proved to be quite vulnerable to adversarial examples. In practice, the SRS model parameters are not always available. Attackers have to probe the model only via querying, and such decision-based attacking merely relies on the output label is quite challenging. This letter proposes a two-step query-efficient decision-based attack based on local low-frequency perturbation. Specifically, instead of imposing perturbation on the entire audio sample, a local attacking region is firstly sought, confining the perturbed distortion to a local region. Second, considering that the majority of energy concentrates on the low-frequency bands, the proposed method suggests performing perturbation generation in the low-frequency domain. Experimental results demonstrate that, compared with the recent methods, our method could implement target attacking to SRS with a higher attacking success rate, at the cost of much lower queries and adversarial perturbation.

60%

Despite neural network-based speaker recognition systems (SRS) have enjoyed significant success, they are proved to be quite vulnerable to adversarial examples. In practice, the SRS model parameters are not always available. Attackers have to probe the model only via querying, and such decision-based attacking merely relies on the output label is quite challenging. This letter proposes a two-step query-efficient decision-based attack based on local low-frequency perturbation. Specifically, instead of imposing perturbation on the entire audio sample, a local attacking region is firstly sought, confining the perturbed distortion to a local region. Second, considering that the majority of energy concentrates on the low-frequency bands, the proposed method suggests performing perturbation generation in the low-frequency domain. Experimental results demonstrate that, compared with the recent methods, our method could implement target attacking to SRS with a higher attacking success rate, at the cost of much lower queries and adversarial perturbation.

50%

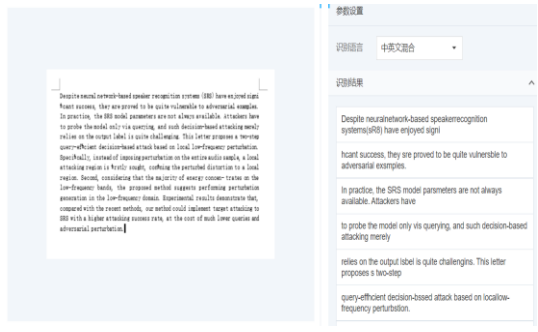
Despite neural network-based speaker recognition systems (SRS) have enjoyed significant success, they are proved to be quite vulnerable to adversarial examples. In practice, the SRS model parameters are not always available. Attackers have to probe the model only via querying, and such decision-based attacking merely relies on the output label is quite challenging. This letter proposes a two-step query-efficient decision-based attack based on local low-frequency perturbation. Specifically, instead of imposing perturbation on the entire audio sample, a local attacking region is firstly sought, confining the perturbed distortion to a local region. Second, considering that the majority of energy concentrates on the low-frequency bands, the proposed method suggests performing perturbation generation in the low-frequency domain. Experimental results demonstrate that, compared with the recent methods, our method could implement target attacking to SRS with a higher attacking success rate, at the cost of much lower queries and adversarial perturbation.

40%

Despite neural network-based speaker recognition systems (SRS) have enjoyed significant success, they are proved to be quite vulnerable to adversarial examples. In practice, the SRS model parameters are not always available. Attackers have to probe the model only via querying, and such decision-based attacking merely relies on the output label is quite challenging. This letter proposes a two-step query-efficient decision-based attack based on local low-frequency perturbation. Specifically, instead of imposing perturbation on the entire audio sample, a local attacking region is firstly sought, confining the perturbed distortion to a local region. Second, considering that the majority of energy concentrates on the low-frequency bands, the proposed method suggests performing perturbation generation in the low-frequency domain. Experimental results demonstrate that, compared with the recent methods, our method could implement target attacking to SRS with a higher attacking success rate, at the cost of much lower queries and adversarial perturbation.

Images

Results of  
Tencent OCR



Results of  
Alibaba OCR

