## Black-Box-Base Generic Perturbation Generation Method under Salient Graphs

Authors: Dingyang Hu, Dan Liu

**Abstract :** DNN (Deep Neural Network) deep learning models are widely used in classification, prediction, and other task scenarios. To address the difficulties of generic adversarial perturbation generation for deep learning models under black-box conditions, a generic adversarial ingestion generation method based on a saliency map (CJsp) is proposed to obtain salient image regions by counting the factors that influence the input features of an image on the output results. This method can be understood as a saliency map attack algorithm to obtain false classification results by reducing the weights of salient feature points. Experiments also demonstrate that this method can obtain a high success rate of migration attacks and is a batch adversarial sample generation method.

Keywords: adversarial sample, gradient, probability, black box

Conference Title: ICSLP 2022: International Conference on Speech and Language Processing

**Conference Location :** San Francisco, United States

Conference Dates: November 03-04, 2022