Politecnico di Torino

SmartData@PoliTO

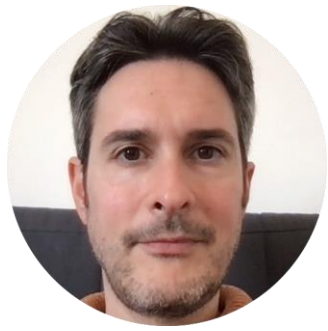
SmartData

June 22nd, 2022, 5:00 PM CEST

SmartTalk: Covivio

Tiziano Bianchi

Associate Professor - Politecnico di Torino



Robust Deep Networks via Gaussian Class-Conditional Loss

ABSTRACT

Deep networks have shown remarkable performance in complex visual and classification tasks. However, their use in security sensitive applications raises serious concerns, as they can be targeted by adversaries. One of the most severe threats is represented by adversarial perturbations, a collection of methods that interfere with neural networks input data in order to produce undesired outputs.

In this talk, I will present the Gaussian Class-Conditional Simplex (GCCS) loss, a novel approach for training deep networks with improved classification accuracy and adversarial robustness. The proposed method learns a mapping of the input classes onto Gaussian target distributions in the latent space, such that a hyperplane can be used as the optimal decision surface. Results show that GCCS provides improved robustness against adversarial perturbations, outperforming models trained with conventional adversarial training.

SmartTalks Season 2