MACHINES CAN SEE

Computer Vision and Deep Learning summit

june 08. 2018. Moscow. Arbat Hall.

Adversarial Attacks on Black-box Face Recognition

Faces provide a natural means to recognize our friends, colleagues and relatives as well as to make ourselves recognized by others. Recent computer vision technology enables to scale face recognition to millions of faces. Modern methods of face recognition largely bypass human performance while relying on machine learning and neural networks. Despite this power, such methods can be vulnerable to attacks aiming to modify network outputs. Indeed, arbitrary changes to the network output can be produced by small and well-designed modifications of the network input, as known under Adversarial Examples. Applied to face recognition, adversarial examples imply that an attack can be designed to force a network to identify a person in the original image (Im1) as any other person on the planet (e.g. Im2) by making small modifications to the original image G(Im1) as shown below.

Im1

G(Im1)

Im2

While attacks on open networks are relatively straightforward, the design of attacks on “black-box” systems, e.g. networks with unknown structure and parameters, is more difficult. Our challenge ”Adversarial Attacks on Black-box Face Recognition” aims to test the vulnerability of black-box face recognition systems. Participants will be given an opportunity to compete and design the best attack forcing the system to recognize an image of a person A as person B, by applying small modifications to images of A.

Prizes

The winners will share a cash pool of 300,000 RUB and 3 NVIDIA GPU’s. The prizes will be announced at the annual conference MachinesCanSee organized on June 8, 2018 in Moscow

Timeline

14th May, 2018: Competition begins

6th June, 2018: Final submission date

VisionLabs MACHINES CAN SEE | machinescansee.com | join in | supported by NVIDIA