RSNA® Radiological Society of North America

RADIOLOGICAL SOCIETY OF NORTH AMERICA
820 JORIE BLVD, OAK BROOK, IL 60523
TEL 1-630-571-2670 FAX 1-630-571-7837
RSNA.ORG

NEWS

Generative Neural Network Inserting or Removing Cancer into Mammograms Fools Radiologists and Deep Learning Alike: Example of an Adversarial Attack

Wednesday 11:10-11:20 AM | SSK02-05 | E451B

PURPOSE

To investigate whether a cycle-consistent generative adversarial network (CycleGAN) can insert or remove cancer-specific features into mammographic images in a realistic fashion.

METHOD AND MATERIALS

From two publicly available datasets (BCDR and INbreast) 680 mammographic images from 334 patients were selected, 318 of which exhibited potentially cancerous masses, and 362 were healthy controls. We trained a CycleGAN, using two pairs of generator and discriminator networks to convert cancerous breast images to healthy and back, and vice versa for the controls, without the need for paired images. The network, implemented in TensorFlow, was trained for 40 epochs on an augmented dataset enlarged ten-fold by random rotation, scaling, and contrast perturbations. To investigate how realistic the images appear, we randomly selected 20 image pairs of original and generated images, and 10 single images of each category (60 images in total). The images were presented to three radiologists (5 and 3 years of experience, and PGY-5 resident) who rated them on a 5-point Likert-like scale and had to indicate whether the image was real or generated/modified. The readout was analysed with a receiver-operating-characteristics (ROC) analysis, performance was expressed as area under the ROC curve (AUC).

RESULTS

For the most experienced radiologist, the modifications introduced by CycleGAN reduced diagnostic performance, with the AUC dropping from 0.85 to 0.63 (p=0.06), respectively, while the two less experienced ones seemed unaffected at a lower baseline performance (AUC 0.75 vs. 0.77 and 0.67 vs. 0.69). None of the radiologists could reliably detect which images were real and which were modified by CycleGAN (AUC 0.50-0.66).

CONCLUSION

CycleGAN can inject or remove malignant features into mammographic images while retaining their realistic appearance. These artificial modifications may lead to false diagnoses.

CLINICAL RELEVANCE/APPLICATION

Modern adversarial attacks may go undetected by humans as well as deep learning algorithms, and could be used in cyber warfare. It is vital to secure healthcare devices and information systems against such attacks mediated by neural networks.

CTrl-Alt-Radiate?

Tuesday 3:40-3:50 PM | SSJ13-05 | Room: N230B

PURPOSE

Computed Tomography (CT) is an essential and commonly used X-ray generator modality that uses ionizing X-ray radiation to produce images. The CT modality consists of an ecosystem of components, which communicate with each other within the CT's ecosystem. As technology advances, the CT's ecosystem is becoming more connected to the hospital's network and the Internet, exposing it to a variety of security vulnerabilities and threats to potential cyber-attacks. The combination of ionizing radiation, potentially harmful to patients, and security vulnerabilities to cyber-attacks results in possible dangerous scenarios that compromise patients' safety. To illustrate the importance of the topic, we demonstrate how we hacked a CT.

METHOD AND MATERIALS

We present a step-by-step implementation of how we bypass current security protection mechanisms of a CT in order to manipulate its behavior, making it potentially dangerous to patients. This attack demonstrates how additional cyber-attacks on medical imagining devices (MIDs) can be similarly implemented. To accurately measure the potential damage to patients' health, we use a phantom device (i.e., a CT radiation measuring device), and analyze the risks that such attack can cause. Furthermore, we demonstrate how to exploit our cyber-attack covertly, so that it is difficult to detect it using current security solutions; thus, such attack may have long-term effects on a large-scale of the population.

RESULTS

A live demonstration of how we hacked a CT device and how we manipulated its behavior to create various dangerous scenarios for patients' health. Moreover, we analyze this attack in depth, to better understand the potential impacts of such attacks.

CONCLUSION

CT and MIDs are vulnerable to cyber-attacks; we demonstrate forcefully that hacking CT and MIDs is no longer theoretical. By analyzing the potential impacts to patients, we can conclude that such impacts are critical and must be dealt with urgently.

CLINICAL RELEVANCE/APPLICATION

This calls for an immediate improvement of CTs and MIDs security and further mitigation of risks to patients.