Chiranjeev
P21CS007
chiranjeev.1@iitj.ac.in

Assignment - 2
Research Problems

25-09-2021

1. **Explainability in Cloned Voice:** In voice cloning, a speaker's voice is given as an input to the model and we attempt to generate the speaker's voice(output voice) but with a different context. So as to know which features of the original input voice the network is considering to generate the output voice. The explainability is lacking here. As we know output voice resembles the input voice in accent, tone, pitch and other properties. There is no mechanism for explainability like Grad-CAM for images, so we need to find which input features are used and given weightage respectively in order to generate the cloned output voice.

2. **Camouflage attack for Army Drones:** Enemies use Camouflage to protect personnel and equipment from observation by armed forces. In practice this means applying colour and materials to equipment of all kinds, including vehicles, camps, guns to conceal it from observation or to make it appear as something else(mimicry of aerial location, trees etc.). This camouflage can fool the drones used by the army and act as an attack to the drone's captured videos. Thus the enemy equipments can't be detected or can be detected incorrectly.

3. **Holographic AR attacks for Object Recognition model:** In this attack the person's or any object's 3D hologram can be projected in space which can be detected as the original/real person or object by the Object Recognition model. This Holographic AR attack can be used to generate false alarms by recognizing it as Original object or person.

4. **Counterfactuals in detecting Parkinson Disease by Handwriting:** For diagnosing Parkinson disease in humans, the person has to write on Tablet with the electronic pencil which is used to detect at which angle the person is holding the pencil and with how much pressure the person is writing. The person has to draw various patterns, one of them is the Spiral pattern and various other patterns. By analysing handwriting through these features the model tells whether a person is having Parkinson disease or not. But what about the person's age and person's mental state when a person is writing with pencil because person in different mind states can write differently and people with less age can also face Parkinson's disease. So, how will the model deal with this situation?

5. **Distribution Shift in images:** There can be distribution shift in images captured from various cameras, which can make our model learn sensor information rather than properly focussing on the part of image it has to focus on in order to predict the correct label. We can take an example of COVID-19 situation, people who are suffering from covid was going to Covid hospitals to get there X-ray or CT-Scan done of their lungs to detect covid in lungs but people who are in doubt

of very mild covid generally prefer to go to normal Hospital in order to stay safe by not coming in contact with the people who are having covid. So here Covid and Non-Covid images of lungs are captured from different cameras/machine sensors. It can lead to distribution shift in images of Covid and Non-Covid images and when we pass these images to the deep learning model, then it may lead to learn sensor information rather than focussing on lungs part to detect covid patches in lungs or whether the person is not having any covid patches.

6. **We can extend problem 1 to another level:** By leading its way to **Voice Conversion**. Here the Speaker's input voice is converted to another desirable person's voice (output voice). We want explainability in the voice of the speaker (input voice) and output voice both i.e., which features of input voice and output voice are learned for conversion by deep learning's Neural Network model. If this is achieved we can also use this to prevent attack by voice because one person can mimic another person's voice.