Research Project : SEM 7

Weekly Reports

TEAM : 13

Topic :

Detection of Deep Fakes using a hybrid neural network architecture

Mentor :

Dr.Mohendra Roy

Team :

Ishan Mistry (18BIT033)
Alister Rodrigues (18BIT041)
Khush Joshi (18BIT056)
Manali Shah (18BIT060)
Mansi Raveshia (18BIT063)

WEEK - 1

(a) Problem statement :

Detection of Deep Fakes using a hybrid neural network architecture :
The Deep Fakes out there need to be kept in check and for that very reason we need to elevate our detection techniques too, in order to tackle them.This project requires studying the various implemented detection schemes and use the insights to devise a new and possibly better hybrid architecture for the detection of Deep Fakes.

(b) Motivation :

People have a tendency to readily believe what they see, quite often overlooking the credibility of the source of the video/image/text.Although people are getting more and more vigilant these days, becoming skeptical towards blindly trusting the media, but there is a caveat to this self established trust.This is a subtle case of misinformation.

Humans are believed to have six degrees of separation between you and your farthest friend. “Six degrees of separation” is the idea that all people on average are six, or fewer, social connections away from each other. As a result, a chain of "friend of a friend" statements can be made to connect any two people in a maximum of six steps.Moreover the situation has been exacerbated by the various channels of media/communication that are consumed these days - News groups/conspiracy groups/Social Media. There is no one to substantiate the validity of these seemingly true sources, as it is a result of the mere misinformation that has been made really easy by the current forms of media.

At first glance, these matters might seem trivial and even not alarming at all, because in most cases the impact of these is not directly visible.But there are times when things could go down-hill and have serious repercussions.For example, at times such information could be used to belittle someone or could be a philippic , aimed to damage the reputation of someone in power.This when combined with the current power of the spreading information can be treacherous.

These subtle but misinformed texts or images are not easy to track and the process could be cumbersome as this spurious information is gleaned into everyday lives to a point that they come out as obvious and evidently become the truth. The facts and figures presented/spread could still be validated with the human expertise in the specific domains - Politics/Sports/Daily Sciences etc. But with advances in Deep Learning, deep fakes have become so realistic that it is nearly impossible for us to visibly differentiate a true video from a deep fake video.This gives rise to the need of having accurate detection

techniques for their detection and hence their removal from the web.The bad people keep getting better and better and so shall we, in order to counter them.

(c) Objectives :

- **Analyze the various prevailing implementations:**
  Study of the various currently popular architectures and learn what exactly is it that makes these architectures good at their job.

- **Studying the characteristics of the Deep fakes images and regular images**
  This is essentially crucial and could give us a boost in critically selecting models/layers

- **Devising a novel technique if possible for the detection**
  Combining both the stated objectives enable us to think critically and approach solving the problem from a novel perspective.This implementation could then be hosted online on a small web app, made on StreamLit, so that it is available to others for use. This way the project becomes our little contribution back to the Society as well to the Deep Learning community.

(d) Roadmap :

Major Checkpoints :

1. Observing the trends/architectures : Literature Survey
2. Start by implementing a few of the current models - Experimenting with them/Looking for vulnerabilities
3. Consolidate our work along with the real world testing.
4. Make our findings available to people, in the form of a simple web app and also to the Deep Learning Community in form of a paper

Detailed schedule(Tentative) :

Week 1 :
Settling on the Logistics for Literature Survey and Too
Week 2 :
Literature Survey + Insights from the papers
Week 3 :
Getting Familiar with the tools + Recapitulation of the insights
Week 4 :
Coming up with hypothesis/experiments + Setting up environments

Week 5 :

Github Project setup + Modelling the Neural Network + Implementation

Week 6 :

Testing the network + Validation + Conducting Experiments

Week 7 :

Concluding the experiments + Working on the paper + Conference submission

Week 8 :

Submission wrap up + Concluding the project

(e) Work Done so far:

Logistics :

- Setting up the System for managing the details from the literature survey
- Zotero Management - for ease of future citations/ reference management

Literature Survey:

For literature survey we have started looking for papers and are currently briefly going through the abstracts and will redistribute the papers for in- depth readings as we progress further.

1. Deep Learning for Deepfakes Creation and Detection: A Survey - Link -
2. Deep-fakes Generation and Detection: State-of-the-art, open challenges, countermeasures, and way forward - Link
3. Recurrent Convolutional Strategies for face manipulation detection in videos - Link
4. An exploratory analysis on visual counterfeits using Conv-LSTM hybrid architecture - Link
5. Deepfakes detection technique using deep learning : A Survey - Link
6. Deep Fake Detection : Survey of Facial Manipulation Detection Solutions - Link
7. Media Forensics and Deep Fakes: An Overview - Link

-