Deepfakes and Other Controversial Computer Vision Issues

Jacob Vesco

CSCI 494

Outline

- What are deepfakes
- History of deepfakes
- The problems and issues deepfakes pose
- Computer vision – high level overview of some algorithms
- A few other issues with some computer vision applications
- Conclusion

What are Deepfakes

- Artificially created fake media
- Look convincing, but entirely fictional
- Like a more modern, more algorithm heavy photoshop

What are Deepfakes

- Used to make images of fake events
- Mostly, they are used for innocent fun
  - Some for comedic effect or shock value
  - But, as of September 2019, 96% were used for adult entertainment
- However, they can be used maliciously

History of Deepfakes

- The first actual deepfake occurred in 2017 on Reddit
- But deepfakes go back to 1997 (theorized at this time)
- The first iterations were easy to spot
- Current deepfakes are harder to spot
- Some forms are now readily available for the public

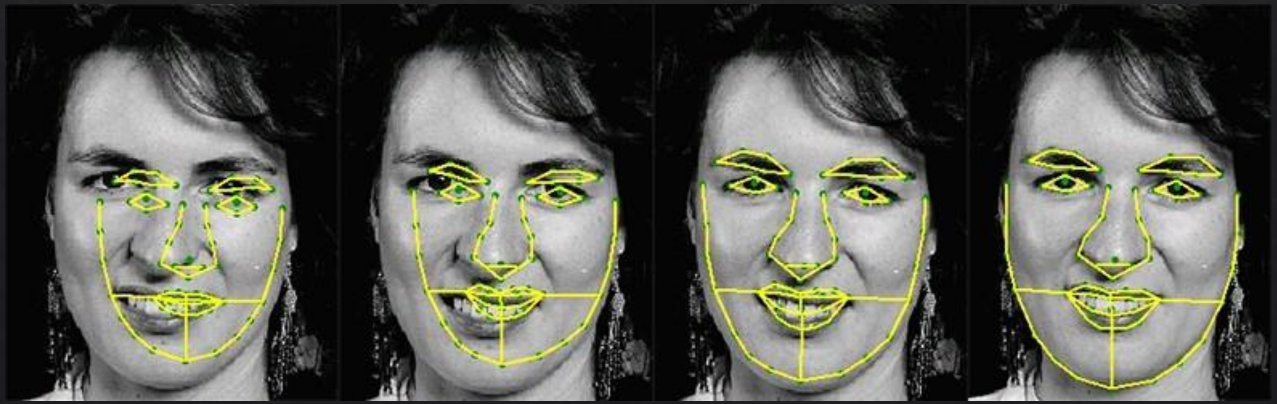
History of Deepfakes

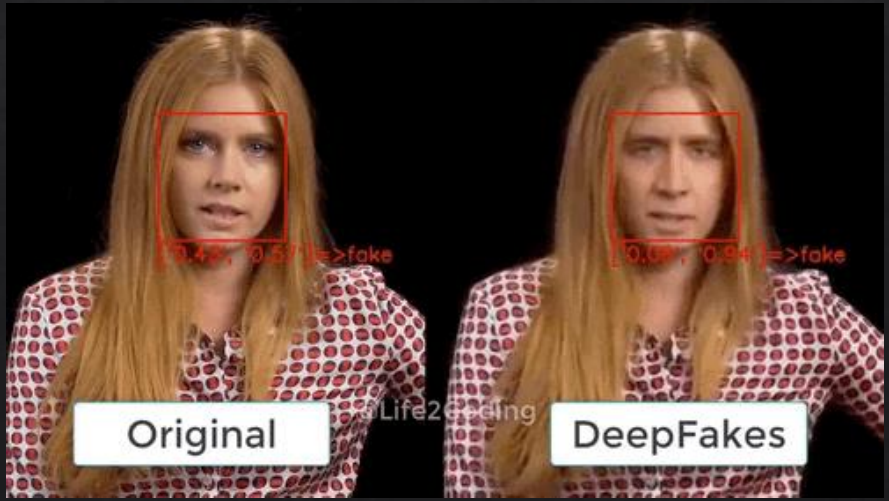
- The 1997 paper (Video Rewrite Program)
  - Written by Christoph Bregler, Michele Covell, and Malcolm Slaney
  - Developed a program to automate movie studio work
  - Not necessarily a deepfake, but first program similar
  - Could create faces from audio
  - First to animate convincingly

History of Deepfakes

History of Deepfakes

- The 2001 paper (Active Appearance Models)
  - Written by Tomothy F. Cootes, Gareth J. Edwards, and Christopher J. Taylor
  - Algorithm would match a shape to an image using a statistical model
  - Big step forward for face matching and tracking





History of Deepfakes

- “Face2Face” and “Synthesizing Obama” projects
  - Proved deepfakes to be obtainable with consumer level hardware
  - Improved graphical quality to look photorealistic
- Face2Face – Modifying the face of a target video with that of an actor in real time
- Synthesizing Obama
  - More akin to the deepfakes we have now
  - Like the 1997 program, but with major improvements

History of Deepfakes

Face2Face

History of Deepfakes

Synthesizing Obama

History of Deepfakes

- The big surge of deepfake popularity happened in 2017 on Reddit
- Deepfakes became increasingly popular for pornographic use
- The subreddit, titled r/deepfakes, had around 90,000 members
  - This subreddit, now banned, is responsible for Reddit updating their policy on pornographic content
- Safer deepfake content has emerged since

The Problem

- Deepfakes can be convincing, near photorealistic
- With that comes more problems than most realize
- Consider weaponized deepfakes
  - Politics
  - Society
- Consider an entire population witnesses a deepfake

The Problem

- Deepfake pornography – no longer innocent
- Nonconsensual, and convincing videos
- Celebrities targeted, and must deal with the consequences
  - Gal Gadot, Wonder Woman actress, was one of the first to be targeted
- Mostly, these are videos targeting and harassing women

The Problem

- Socially, deepfakes can be used to harass
- Careers could be ruined from fabricated events
- Lives could be irreversibly altered
- The technology is readily available, meaning anybody can fabricate events

The Problem

- Politically, the ramifications can be catastrophic
- According to The Brookings Institution, deepfakes can:
  - Manipulate elections, cause institutional distrust, undermine public safety, destroy reputations, and more
- AI generated propaganda
- Governments have already experienced problems with deepfakes…

The Problem

- The incident with Ali Bongo, the president of Gabon
  - Missing from the public's eyes for too long, rumors spread of his health
  - A video of Ali emerged, but it was suspicious, and accused of being a deepfake
  - Rapid destabilization occurred, and the military launched a coup
  - Ali has appeared since, and remains in office today
  - https://fb.watch/3_soS3fDBm/ - The video in question
- 2 incidents in Malaysia and Brazil
  - Both claimed incriminating footage were deepfakes
  - No one can prove they were or were not deepfakes

Computer Vision

- Deepfakes are a part of computer vision
- Computer vision is essentially how computers see and respond to real world imagery
- Think of programs like item detection, face recognition, and object tracking
- Deepfakes is a conjunction of “deep learning” and “fake”
  - Meaning they utilize deep learning AI algorithms to create fake events

Computer Vision

- Algorithm for “face-swap” deepfakes
  - Run thousands of face-pics of the two people through an encoder
  - A decoder then is meant to recover the faces from compressed images
  - One decoder should recover one person’s face, another decoder for the second
  - Simply feed compressed images of one person into the wrong decoder, and visa versa
  - Must be done for every video frame for a convincing deepfake

Computer Vision

- Algorithm for creating fake people
  - Generative adversarial network (Gan)
  - Two AI algorithms work against each other. The generator and the discriminator
  - The generator turns random noise into an image
  - The discriminator then gets fed a stream of real images + the generated images
  - Run this cycle multiple times, and images of fake but realistic people will emerge

Other Controversies

- Deepfakes are a big issue in computer vision, but there are other notable problems
  - Mass surveillance and facial recognition
  - Right to privacy and security
  - Self driving vehicles

Conclusion

Resources

- https://medium.com/@songda/a-short-history-of-deepfakes-604ac7be6016
- https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them
- https://www.forbes.com/sites/robtoews/2020/05/25/deepfakes-are-going-to-wreak-havoc-on-society-we-are-not-prepared/?sh=2d7727597494