Propuesta de Proyecto de Investigación
Maestría en Ciencias y Tecnologías de la Información

21 de octubre del 2021

1. **Nombre del proyecto:** Análisis comparativo de algoritmos para la generación de deepfakes.

2. **Responsable(s)**
Dr. Pedro Lara Velázquez
Universidad Autónoma Metropolitana, cubículo T-145, plara@xanum.uam.mx

Dr. Sergio Gerardo de los Cobos Silva
Universidad Autónoma Metropolitana, cubículo T-103, cobos@xanum.uam.mx

3. **Área(s) de conocimiento relacionada(s) con el proyecto**
Optimización e inteligencia artificial

Observación: Es recomendable que el estudiante tenga conocimientos de redes neuronales.

4. **Descripción del proyecto**
   - Contexto
     Deepfake es un acrónimo de las palabras en inglés *fake*, que significa falso, y *deep*, que hace alusión al *deep learning* o aprendizaje profundo, y se define como una técnica de aprendizaje no supervisado, que combina diferentes tipos de redes neuronales para alterar una imagen o video, produciendo un resultado falso pero realista. Los primeros videos conocidos, generados mediante el uso de esta técnica, fueron subidos en 2017 a Reddit, por un usuario con el alias *Deepfake*. Desde entonces, el uso y desarrollo de videos deepfake ha crecido ampliamente, de tal manera que cada vez es más fácil crear resultados más convincentes. Lamentablemente, desde sus orígenes, el uso de esta técnica ha estado estrechamente relacionado con actividades poco éticas, lo cual ha abierto un debate sobre las acciones que deben tomarse para controlarlo, restringirlo o incluso penalizarlo. También es importante mencionar, que parte de su desarrollo se ha visto favorecido por usos completamente válidos, por ejemplo, para el cine se han desarrollado sofisticados programas, que ayudan a sustituir el rostro de actores, con imágenes de ellos mismos pero más jóvenes, extraídos de películas o videos anteriores.

   - Motivación
     Entender el funcionamiento de las herramientas diseñadas para producir videos deepfake, requiere de sólidos conocimientos en programación, y en técnicas de aprendizaje profundo como redes generativas adversarias, redes neuronales convolucionales y autoencoders. Por lo tanto, se convierte en una tarea que pone a prueba los conocimientos de cualquier persona interesada en aprendizaje automatizado. De manera adicional, este proyecto forma parte de una línea de investigación más amplia, destinada a la detección del deepfake.

   - Aporte esperado al área de conocimiento
     Se analizarán y compararán al menos dos algoritmos basados en redes neuronales para la generación de videos deepkafe.

5. **Objetivos**
   Objetivo general

Determinar cuál es la mejor herramienta de acceso gratuito y código abierto para la realización de deepfakes.

Objetivos particulares

1. Estudiar los conceptos más importantes sobre redes neuronales básicas, especialmente redes neuronales generativas adversarias, redes neuronales convolucionales y autoencoders.
2. Seleccionar el conjunto de herramientas, de código abierto, para la generación de deepfakes que serán empleadas en este proyecto.
3. Modificar los códigos de las herramientas seleccionadas para estandarizar algunas de sus características, como el cálculo de la función de pérdida.
4. Establecer un diseño de experimentos adecuado al proyecto, detallando las instancias a usar y los parámetros a medir.
5. Análisis y comparación de los resultados obtenidos.
6. Reportar los resultados obtenidos en la Idónea Comunicación de Resultados (ICR).

6. Metodología

- Realizar una investigación en libros especializados en redes neuronales para conocer sus fundamentos, dando especial énfasis a las redes generativas adversarias, convolucionales y autoencoders ya que se considera que estas arquitecturas son las más prometedoras para el problema de generación de deepfakes.
- Realizar una investigación más profunda sobre el tema en revistas especializadas para determinar el tipo de redes neuronales más empleadas en este problema.
- Con el conocimiento adquirido al desarrollar el estado de arte se realizará la selección de los algoritmos generadores de deepfakes que se emplearán en el análisis comparativo.
- Modificar los hiper-parámetros de los algoritmos, cuando sea necesario o posible, con el objetivo de que los experimentos se realicen en condiciones equivalentes, que permitan comparaciones válidas de los experimentos.
- Seleccionar un conjunto de videos cortos con características adecuadas para la generación de deepfakes, por ejemplo mismo tipo de iluminación, personas con medidas antropométricas del rostro parecidas, y mismo tono de piel.
- Comparar los resultados obtenidos de forma cuantitativa por medio del costo de la función de pérdida y de forma cualitativa mediante la realización de encuestas.
- Reportar los resultados obtenidos en la Idónea Comunicación de Resultados (ICR).

7. Calendarización de actividades

Trimestre 1: Estudio de redes neuronales básicas, redes neuronales generativas adversarias, redes neuronales convolucionales y autoencoders.
Trimestre 2: Selección y análisis de herramientas para la generación de deepfakes.
Trimestre 3: Análisis del desempeño de las herramientas seleccionadas en diferentes instancias. Entrega de la versión final de la Idónea Comunicación de Resultados.
Trimestre 4: Revisión de los sinodales de la Idónea Comunicación de Resultados. Presentación del examen de grado.

8. Infraestructura necesaria y disponible

Una computadora con Windows o Linux, tarjeta de video Nvidia y Python para programar.

9. Lugar de realización

El proyecto puede realizarse en el cubículo T-103, o de forma remota (no presencial) dependiendo de las circunstancias.

10. Entregables

- Al concluir el Proyecto de investigación I se entregará un reporte en formato de artículo con los avances obtenidos en los objetivos particulares 1 y 2.
- Al terminar el Proyecto de investigación II se entregará una primera versión de la ICR con los avances obtenidos hasta el objetivo particular 5.
- La ICR deberá entregarse al finalizar el Proyecto de Investigación III.

11. Referencias bibliográficas básicas

1. A. Siarohin, S. Lathuilière, S. Tulyakov, E. Ricci, N. Sebe, (2020). First Order Motion Model for Image Animation. Disponible en: https://arxiv.org/abs/2003.00196

2. Deepfacelab. Disponible en: https://github.com/iperov/DeepFaceLab

3. Faceswap. Disponible en: https://faceswap.dev/

4. FaceSwapGan. Disponible en: https://github.com/shaoanlu/faceswap-GAN

5. R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, J. Ortega-Garcia, (2020). Deepfakes and beyond: A Survey of face manipulation and fake detection. Information Fusion, 64, pp. 131-148. https://doi.org/10.1016/j.inffus.2020.06.014

6. SimSwap. Disponible en: https://github.com/neuralchen/SimSwap

7. Y. Nirkin. Y. Keller, T. Hassner, (2020). FSGAN: Subject Agnostic Face Swapping and Reenactment. Disponible en: https://arxiv.org/abs/1908.05932