PROYECTO IVERES

VERIFICA

Plan de Recuperación, Transformación y Resiliencia

Financiado por la Unión Europea
NextGenerationEU

Cátedra RTVE UPC

Proyecto IVERES
La verificación de imágenes

11/07/2024

julio de 2024

Indice

- Un poco de historia sobre imágenes/videos falsos
- Tipologías de fakes en imagen/video
- Manipulación de imágenes (Forensics)
- Generación de caras
- Tecnicas de intercambio de caras: Reenactment, Swapping, Talking
- Tecnicas de detección de fakes generados por IA
- Generación de imágenes a partir de texto y su detección.
- Generación de vídeo a partir de texto.
- Estrategias de certificación de autenticidad.

PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Breve historia del fake en imágenes.

Ramses II. Batalla de Kadesh (1274 a.c)

Múltiples representaciones de la batalla en la que Ramses II explica como consigue vencer a los hititas.

Las representaciones se acompañan de poemas que son recitados por mensajeros reales en todas las ciudades. La batalla parece bien documentada por los egipcios.

Recientemente se descubrieron más de un centenar de papiros privados entre Ramses II I Hattusili III en los que se revela que Ramses II no logró conquistar Kadesh.

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Breve historia del fake en imágenes.

Ramses II. Batalla de Kadesh (1274 a.c)

Múltiples representaciones de la batalla en la que Ramses II explica como consigue vencer a los hititas.

Las representaciones se acompañan de poemas que son recitados por mensajeros reales en todas las ciudades. La batalla parece bien documentada por los egipcios.

Recientemente se descubrieron más de un centenar de papiros privados entre Ramses II I Hattusili III en los que se revela que Ramses II no logró conquistar Kadesh.

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Breve historia del fake en imágenes.

Ramses II. Batalla de Kadesh (1274 a.c)

Múltiples representaciones de la batalla en la que Ramses II explica como consigue vencer a los hititas.

Las representaciones se acompañan de poemas que son recitados por mensajeros reales en todas las ciudades. La batalla parece bien documentada por los egipcios.

Recientemente se descubrieron más de un centenar de papiros privados entre Ramses II I Hattusili III en los que se revela que Ramses II no logró conquistar Kadesh.

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

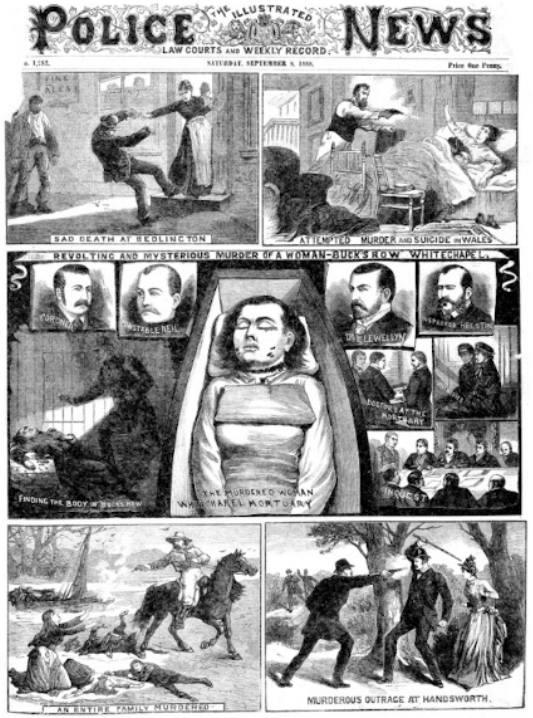
Breve historia del fake en imágenes.

Jack the Ripper 1888 - 1891

Aparecieron muchas noticias falsas y sensacionalistas: prensa amarilla (finales siglo XIX)

Para vender más ejemplares y captar la atención se publican historias exageradas y completamente falsas sobre los asesinatos.

Las noticias se apoyan sobre dibujos realizados manualmente, muchos de ellos xenófobos (especialmente contra los judíos)



PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Breve historia del fake en imágenes.

Primer libro impreso, 868 CE, China, “The Diamond Sutra”.

Gutenberg, 1450, impremta con tipos móviles.
1452, 180 ejemplares de la Biblia con 1300 páginas
El número de personas que aprenden a leer se dobla en pocos años.
Las monarquías empiezan a tener problemas para controlar los contenidos que se publican.
Se publican panfletos y ‘Libelles’ fuera de las fronteras para divulgar información falsa.

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Breve historia del fake en imágenes.

En 1750 -1770 aparecen multitud de ‘Libelles’, panfletos publicados en imprentas y distribuidos entre la población francesa criticando a la monarquía y especialmente a Maria Antonieta. Son bulos sin ningún tipo de control que difunden ideas sobre la vida de la monarquía.

El juicio a Maria Antonieta se acabo basando en la información publicada en los ‘Libelles’

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Breve historia del fake en imágenes.

- Benjamin Franklin 1780s. Publicación de periódicos en Francia, enviados a America, con noticias falsas sobre como los britanicos se asocian con los nativos.

- Elecciones Presidenciales. Washington vs Jefferson

- Naufragio USS Maine
-
- Alien Invasión (La guerra de los mundos)

- In the Event of a Moon Disaster. Emy Award 2021, MIT Center for Advanced Virtuality Reproduction, Francesca Pa netta, Halsey Burgund

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

In the Event of a Moon Disaster

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Positive Deepfakes

Star Wars Deep Fakes

Welcome to Chechnya, David France, Deepfakes to conceal true faces of LGBTQ activist facing persecution in Russia
See people emotions while protecting their identities

Disabled people that have lost voice.

David Beckam. Malaria message in 19 languages

Dali lives. Dali Museum St Petersburg, Florida

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Deepfake (Britannica)

- Synthetic Media, including image, video and audio, generated by AI technology that portray something that does not exist in reality or events that have never ocurred

PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Manipulación de imágenes (Photoshop)

Técnicas forénsicas para detectar si una imagen ha sido manipulada

1. Análisis de metadatos. EXIF Data. Buscar inconsistencias en los metadatos: fechas, cámara, parámetros, lugar, etc.
2. Análisis de artefactos de compresión. La manipulación puede alterar la estadística de los artefactos de compresión de un codificador.
3. Análisis del nivel de ruido
4. Análisis de bordes y dobles duplicados
5. Análisis de iluminación
6. Detección de clonación
7. Detección de huellas de edición

PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Manipulación de imágenes (Photoshop)

Herramientas utilizadas

1. **JPEGsnoop**. Herramienta gratuita que analiza los metadatos de la imagen. Descargable: https://jpegsnoop.en.softonic.com

2. **Forensically**. Detección de clonaciones, análisis de bordes y ruidos.
   Aplicativo web: https://29a.ch/photo-forensics/#forensic-magnifier

3. **Ghiro**. Análisis de imágenes forenses que proporciona informes detallados.

PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Manipulación de imágenes (Photoshop)

ExifTool

```
ExifTool Version Number         : 12.06
File Name                       : gambar3.jpg
Directory                       : D:/ITBTERCINTA/SEMESTER7/ForDig/Tugas 3
File Size                       : 1656 kB
File Modification Date/Time     : 2020:09:14 00:07:18+07:00
File Access Date/Time           : 2021:10:04 02:42:45+07:00
File Creation Date/Time         : 2021:10:04 01:10:35+07:00
File Permissions                : rw-rw-rw-
File Type                       : JPEG
File Type Extension             : jpg
MIME Type                       : image/jpeg
JFIF Version                    : 1.01
Exif Byte Order                 : Little-endian (Intel, II)
X Resolution                    : 300
Y Resolution                    : 300
Resolution Unit                 : inches
Software                        : GIMP 2.10.20
Modify Date                     : 2020:09:14 00:07:14
Color Space                     : sRGB
Compression                     : JPEG (old-style)
Photometric Interpretation      : YCbCr
Samples Per Pixel               : 3
Thumbnail Offset                : 292
Thumbnail Length                : 10657
Profile CMM Type                : Little CMS
Profile Version                 : 4.3.0
Profile Class                   : Display Device Profile
Color Space Data                : RGB
Profile Connection Space        : XYZ
Profile Date Time               : 2020:09:13 16:56:06
Profile File Signature          : acsp
Primary Platform                : Microsoft Corporation
CMM Flags                       : Not Embedded, Independent
Device Manufacturer             :
Device Model                    :
Device Attributes               : Reflective, Glossy, Positive, Color
Rendering Intent                : Perceptual
Connection Space Illuminant     : 0.9642 1 0.82491
Profile Creator                 : Little CMS
Profile ID                      : 0
Profile Description             : GIMP built-in sRGB
Profile Copyright               : Public Domain
Media White Point               : 0.9642 1 0.82491
Chromatic Adaptation            : 1.04788 0.02292 -0.05022 0.02959 0.99048 -0.01707 -0.00925 0.01508 0.75168
Red Matrix Column               : 0.43604 0.22249 0.01392
Blue Matrix Column              : 0.14305 0.06061 0.71393
Green Matrix Column             : 0.38512 0.7169 0.09706
Red Tone Reproduction Curve     : (Binary data 32 bytes, use -b option to extract)
Green Tone Reproduction Curve   : (Binary data 32 bytes, use -b option to extract)
Blue Tone Reproduction Curve    : (Binary data 32 bytes, use -b option to extract)
Chromaticity Channels           : 3
Chromaticity Colorant           : Unknown (0)
Chromaticity Channel 1          : 0.64 0.33002
Chromaticity Channel 2          : 0.3 0.60001
Chromaticity Channel 3          : 0.15001 0.06
Device Mfg Desc                 : GIMP
Device Model Desc               : sRGB
Image Width                     : 2475
Image Height                    : 2475
Encoding Process                : Progressive DCT, Huffman coding
Bits Per Sample                 : 8
Color Components                : 3
Y Cb Cr Sub Sampling            : YCbCr4:4:4 (1 1)
Image Size                      : 2475x2475
Megapixels                      : 6.1
Thumbnail Image                 : (Binary data 10657 bytes, use -b option to extract)
```

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Manipulación de imágenes (Photoshop)

ExifTool

```
ExifTool Version Number         : 12.06
File Name                       : gambar3.jpg
Directory                       : D:/ITBTERCINTA/SEMESTER7/ForDig/Tugas 3
File Size                       : 1656 kB
File Modification Date/Time     : 2020:09:14 00:07:18+07:00
File Access Date/Time           : 2021:10:04 02:42:45+07:00
File Creation Date/Time         : 2021:10:04 01:10:35+07:00
File Permissions                : rw-rw-rw-
File Type                       : JPEG
File Type Extension             : jpg
MIME Type                       : image/jpeg
JFIF Version                    : 1.01
Exif Byte Order                 : Little-endian (Intel, II)
X Resolution                    : 300
Y Resolution                    : 300
Resolution Unit                 : inches
Software                        : GIMP 2.10.20
Modify Date                     : 2020:09:14 00:07:14
Color Space                     : sRGB
Compression                     : JPEG (old-style)
Photometric Interpretation      : YCbCr
Samples Per Pixel               : 3
Thumbnail Offset                : 292
Thumbnail Length                : 10657
Profile CMM Type                : Little CMS
Profile Version                 : 4.3.0
Profile Class                   : Display Device Profile
Color Space Data                : RGB
Profile Connection Space        : XYZ
Profile Date Time               : 2020:09:13 16:56:06
Profile File Signature          : acsp
Primary Platform                : Microsoft Corporation
```

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Manipulación de imágenes (Photoshop)

ExifTool

```
Exif Byte Order                 : Little-endian (Intel, II)
X Resolution                    : 300
Y Resolution                    : 300
Resolution Unit                 : inches
Software                        : GIMP 2.10.20
Modify Date                     : 2020:09:14 00:07:14
Color Space                     : sRGB
Compression                     : JPEG (old-style)
Photometric Interpretation      : YCbCr
```

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Manipulación de imágenes (Photoshop)

JPEGSnoop

También muestra los metadatos EXIF.

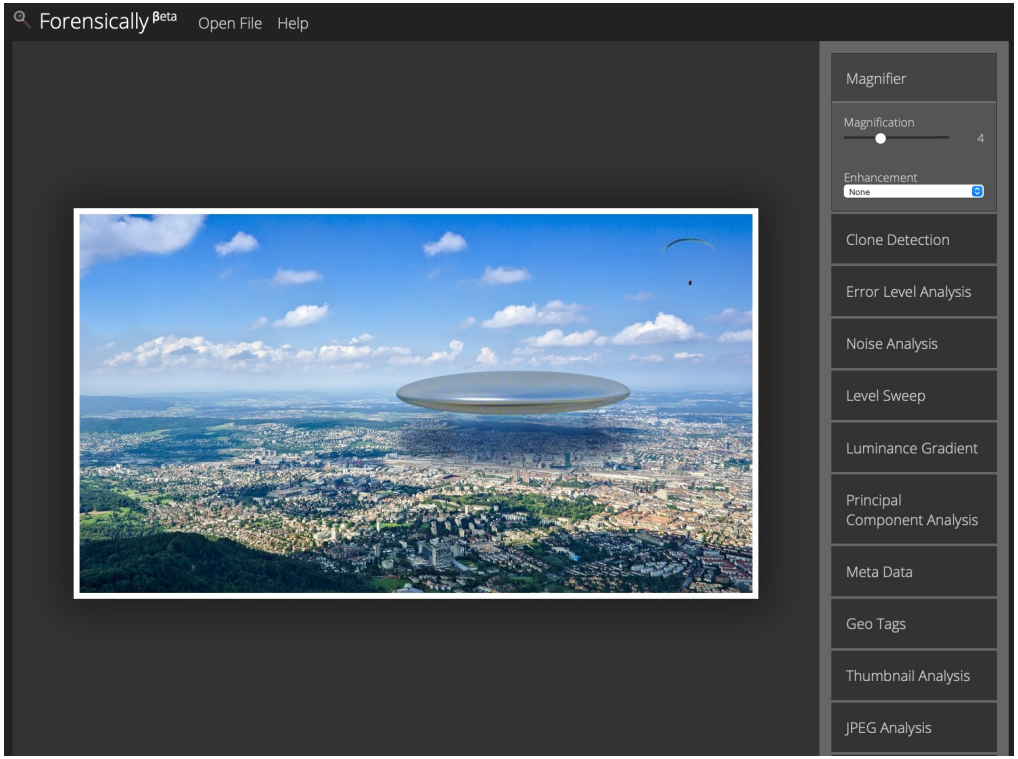
Además categoriza la imagen en clases en función de la probabilidad que sea original o haya sido editada

```
Based on the analysis of compression characteristics and EXIF metadata:

ASSESSMENT: Class 1 - Image is processed/edited

This may be a new software editor for the database.
If this file is processed, and editor doesn't appear in list above,
PLEASE ADD TO DATABASE with [Tools->Add Camera to DB]
```

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Manipulación de imágenes (Photoshop)

Forensically



https://29a.ch/photo-forensics/#forensic-magnifier

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Manipulación de imágenes (Photoshop)

Forensically. Funcionalidades

Magnifier: Una lupa para ampliar los píxeles en diferentes factores de amplitud. Puede amplificar las diferencias entre pixeles mediante mejoras de contraste y ecualizaciones automáticas.

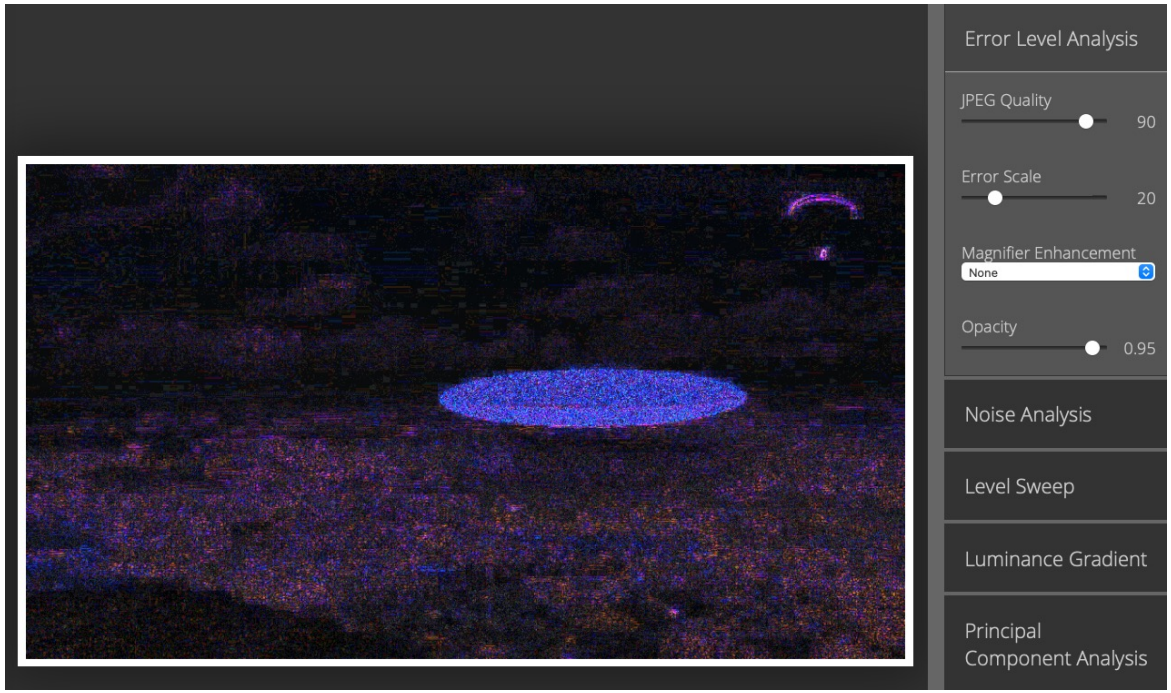
Clone Detection: Busca regiones de la imagen que sean una copia de otras regiones, que hayan sido editadas mediante cloning.

https://29a.ch/photo-forensics/#forensic-magnifier

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Manipulación de imágenes (Photoshop)

Forensically. ELA (Error Level Analysis)

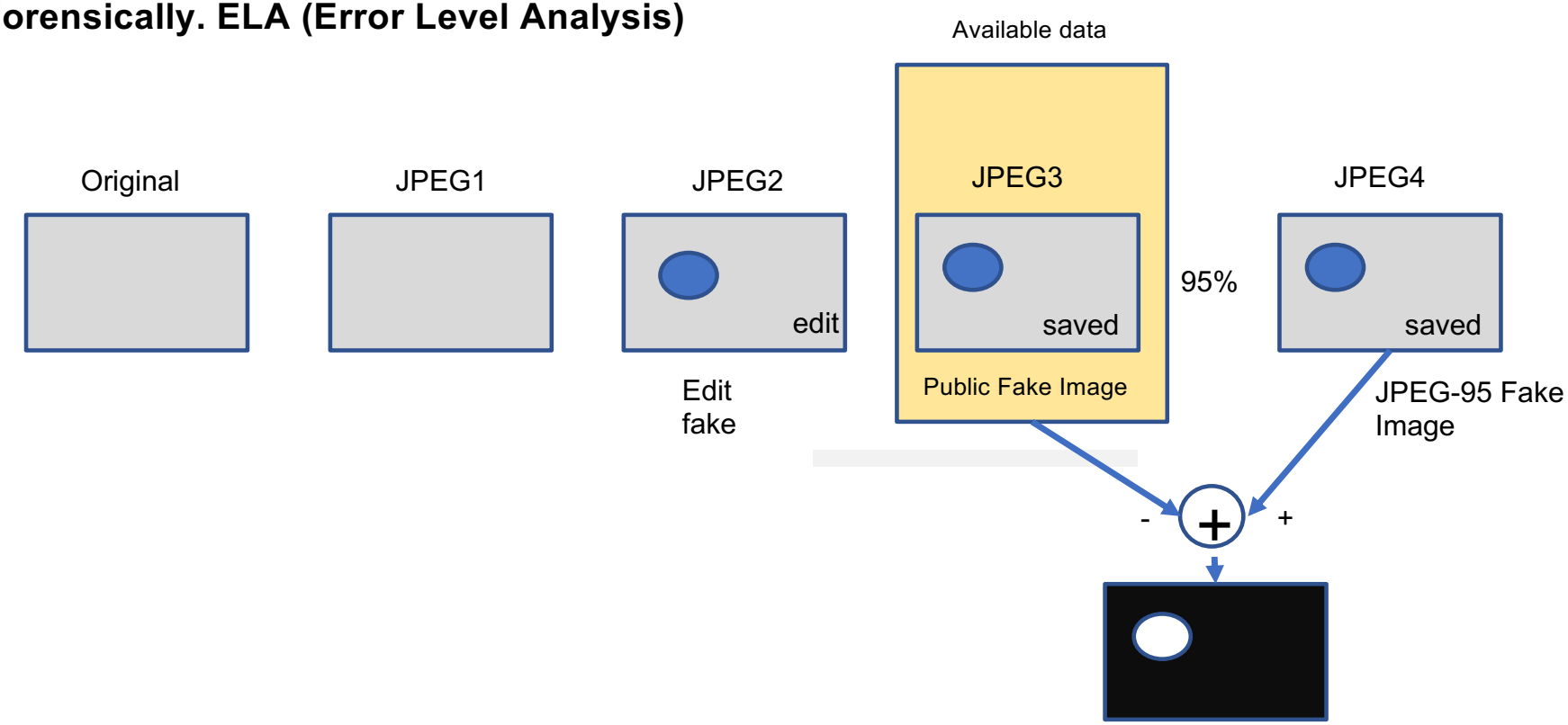


https://29a.ch/photo-forensics/#forensic-magnifier

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Manipulación de imágenes (Photoshop)

Forensically. ELA (Error Level Analysis)



PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Manipulación de imágenes (Photoshop)

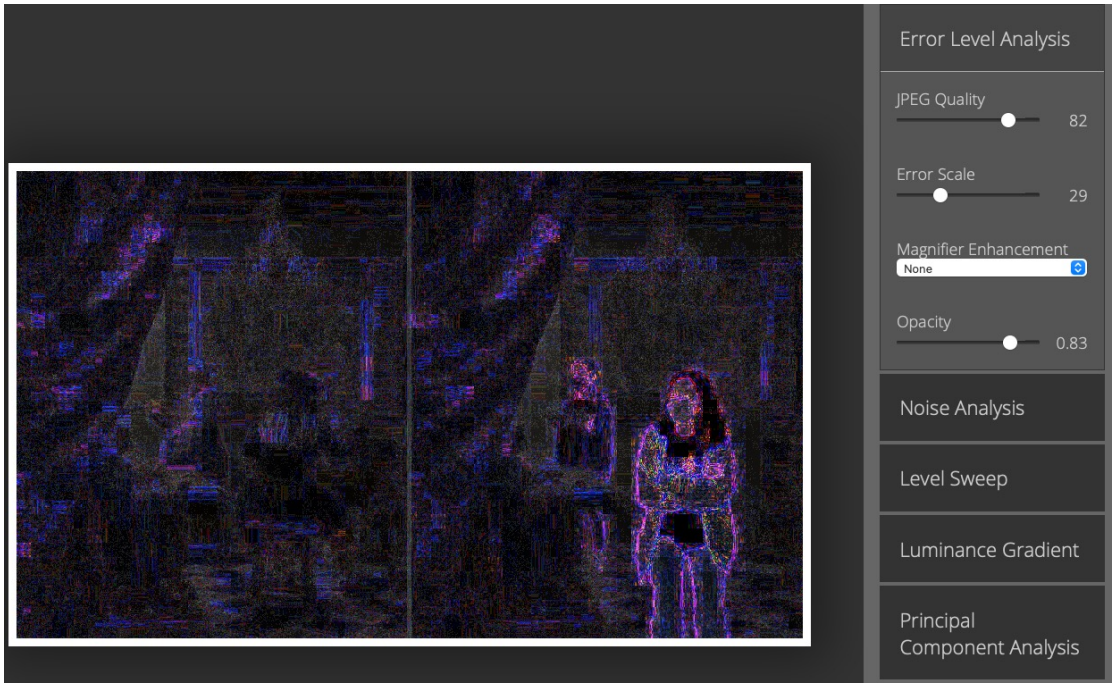
Forensically. ELA (Error Level Analysis)

https://29a.ch/photo-forensics/#forensic-magnifier

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Manipulación de imágenes (Photoshop)

Forensically. ELA (Error Level Analysis)



https://29a.ch/photo-forensics/#forensic-magnifier

PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Manipulación de imágenes (Photoshop)

Forensically. ELA (Error Level Analysis)

PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Manipulación de imágenes (Photoshop)

Forensically. ELA (Error Level Analysis)

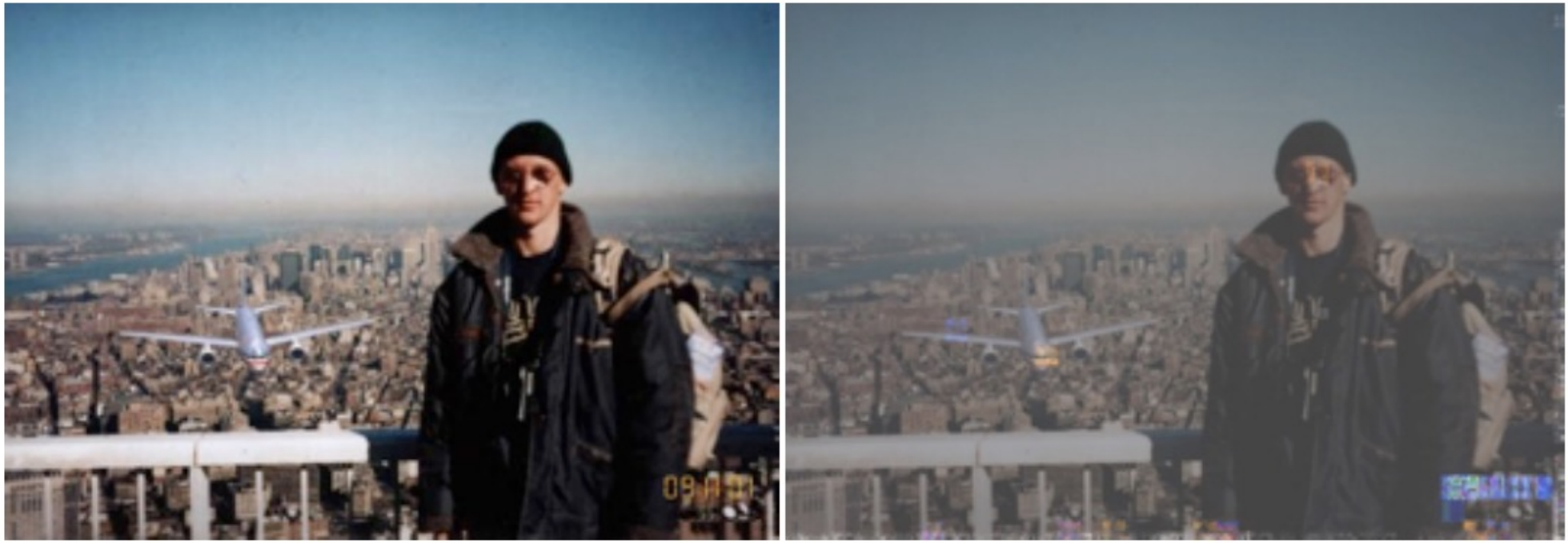


Figure 19. The crash picture and 95% ELA overlay.

PROYEC iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Manipulación de imágenes (Photoshop)

Forensically. Noise Analysis

El patrón de ruido cuando se genera una imagen manipulada (borrando objetos o moviendo objetos) es distinto al patrón de ruido original por lo que puede detectarse si el ruido se caracteriza estadísticamente mediante filtros y posteriormente se representa gráficamente

https://29a.ch/photo-forensics/#forensic-magnifier

PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Manipulación de imágenes (Photoshop)

Forensically. Noise Analysis

https://29a.ch/photo-forensics/#forensic-magnifier

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Manipulación de imágenes (Photoshop)

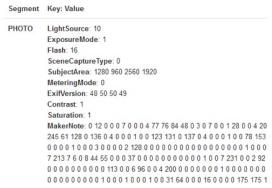
Forensically. ELA (Error Level Analysis)

https://29a.ch/photo-forensics/#forensic-magnifier

PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Manipulación de imágenes (Photoshop)

Ghiro. Automated tool designed to run forensics analysis (massive amount of images)

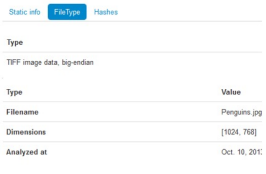


METADATA EXTRACTION

Metadata are divided in several categories depending on the standard they come from. Image metadata are extracted and categorized. For example: EXIF, IPTC, XMP.

GPS LOCALIZATION

Embedded in the image metadata sometimes there is a geotag, a bit of GPS data providing the longitude and latitude of where the photo was taken, it is read and the position is displayed on a map.



MIME INFORMATION

The image MIME type is detected to know the image type your are dealing with, in both contacted (example: image/jpeg) and extended form.

ERROR LEVEL ANALYSIS

Error Level Analysis (ELA) identifies areas within an image that are at different compression levels. The entire picture should be at roughly the same level, if a difference is detected, then it likely indicates a digital modification.

https://getghiro.org

PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Manipulación de imágenes (Photoshop)

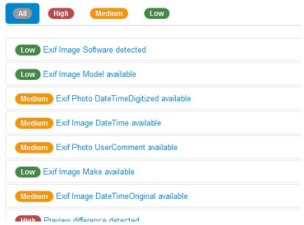
Ghiro. Automated tool designed to run forensics analysis (massive amount of images)

THUMBNAIL EXTRACTION

The thumbnails and data related to them are extracted from image metadata and stored for review.

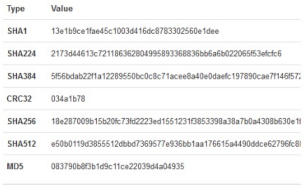
THUMBNAIL CONSISTENCY

Sometimes when a photo is edited, the original image is edited but the thumbnail not. Difference between the thumbnails and the images are detected.



SIGNATURE ENGINE

Over 120 signatures provide evidence about most critical data to highlight focal points and common exposures.



HASH MATCHING

Suppose you are searching for an image and you have only the hash. You can provide a list of hashes and all images matching are reported.

https://getghiro.org

PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Manipulación de imágenes (Photoshop)

Ghiro. Automated tool designed to run forensics analysis (massive amount of images)

Herramienta gratuita Open Source instalable en local. https://github.com/ghirensics/ghiro.git

Puede realizar tratamiento masivo de ficheros en directorios

Dispone de las siguientes funcionalidades:

- Extracción de metadatos (Exif, IPTC, XMP)
- Localización GPS (lectura de geotags, posicionamiento en mapa)
- MIME Información. Determina el tipo y codificación de la imagen (JPEG, TIFF, JP2, etc.)
- Error Level Analysis (ELA). El mismo algoritmo que en Forensically.
- Thumbnail extraction. Extracción de imágenes en miniatura para facilitar la revision
- Thumbnail consistency. Detecta inconsistencias entre una foto editada en la que no se ha editado el thumbnail
- Signature Engine. Utiliza la identificación y comparación de firmas digitales para detector manipulaciones y verificar el contenido de las imágenes
- Hash Matching

https://getghiro.org

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Manipulación de imágenes (Photoshop)

Ghiro. Automated tool designed to run forensics analysis (massive amount of images)

Signature Engine. A partir de la imagen intenta identificar el tipo de cámara con el que se ha realizado la fotografía y los artefactos de la compresión, teniendo en cuenta los puntos de enfoque, datos de exposición,etc. Tiene una base de datos con más de 200 cámaras y si identifica inconsistencias entre la cámara de los metadatos y la cámara identificada en el análisis de firmas advierte de la posible manipulación de la imagen.

Hash Analisys.

- Para cada imagen que entra en Ghiro se calcula un Hash y se compara con una base de datos de Hashes calculados previamente.
- Si los Hashes coinciden representa que la imagen no ha sido manipulada
- Si no coinciden se verifica que la imagen ha sido manipulada o es una nueva imagen.

https://getghiro.org

Manipulación de imágenes (Deep Fake)

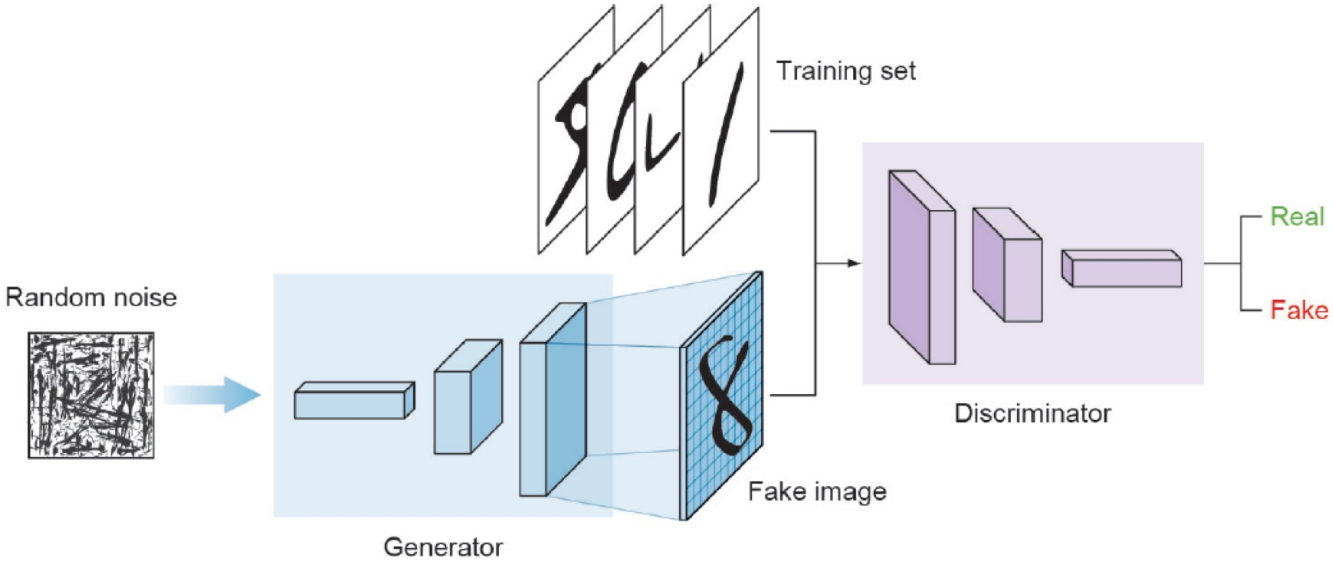
Técnicas más habituales

1. **Generación de caras ’nuevas’. Avatares**
2. **Generación con control de características.**
3. **Face Reenactment**
4. **Face Swapping**
5. **Talking Face**
6. **Text-driven generative images**
7. **In-Painting & Out-Painting Text-driven generative images**

PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Generación de Caras Artificiales

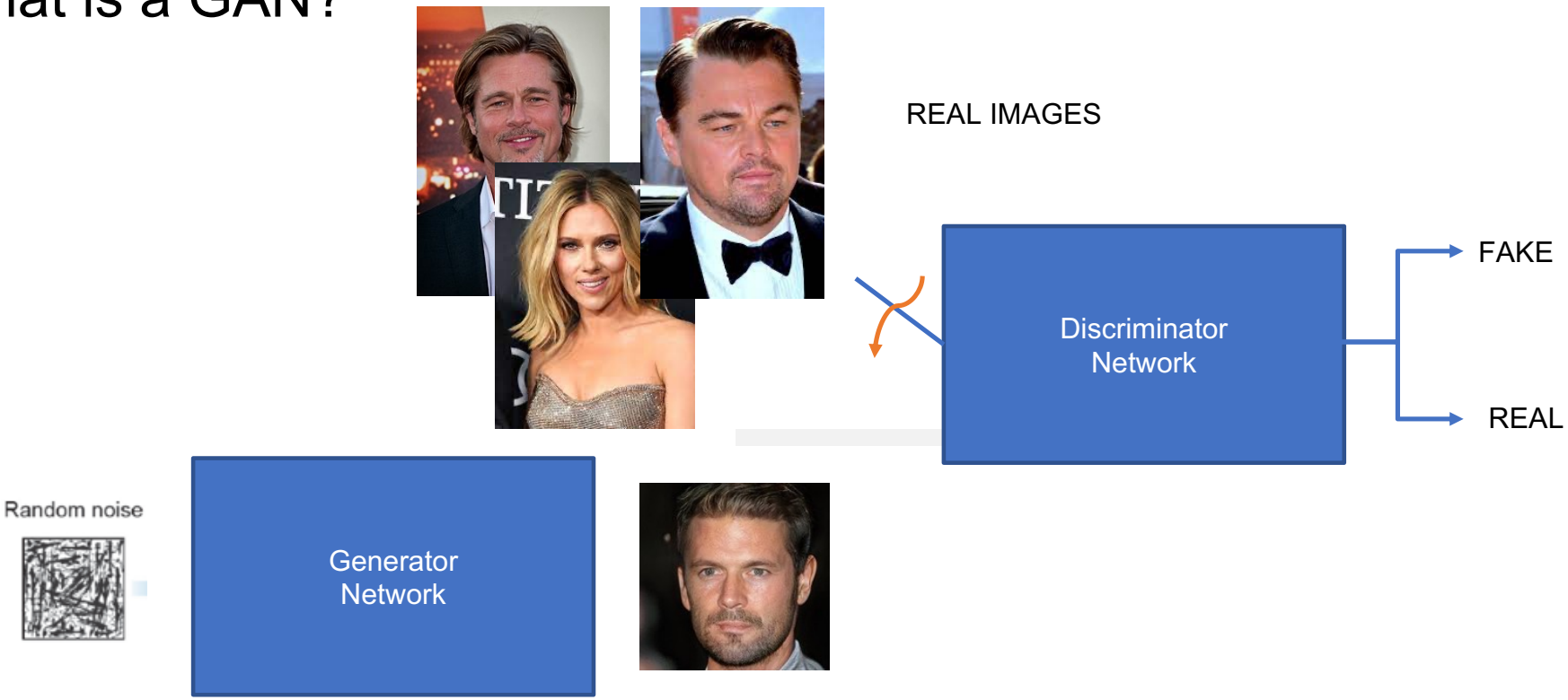
What is a GAN?



PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Generación de Caras Artificiales

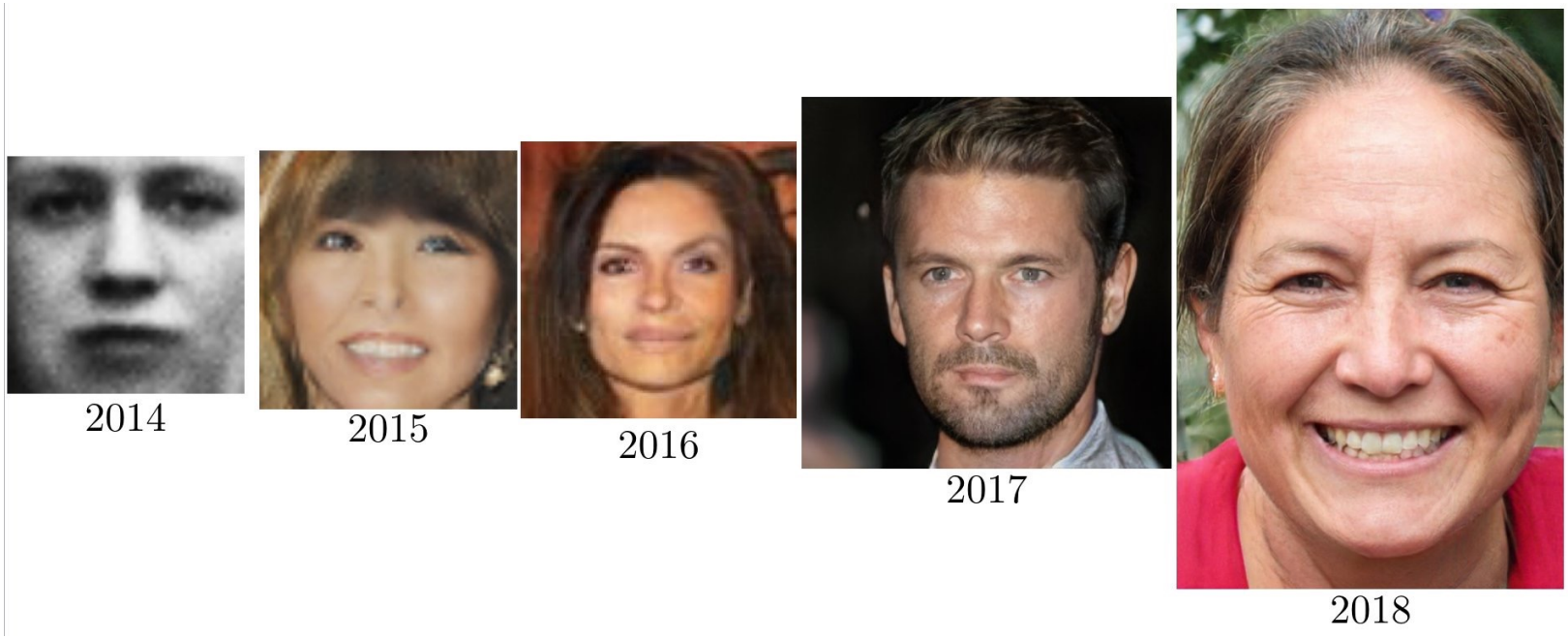
What is a GAN?



PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Ejemplos de caras generadas artificialmente

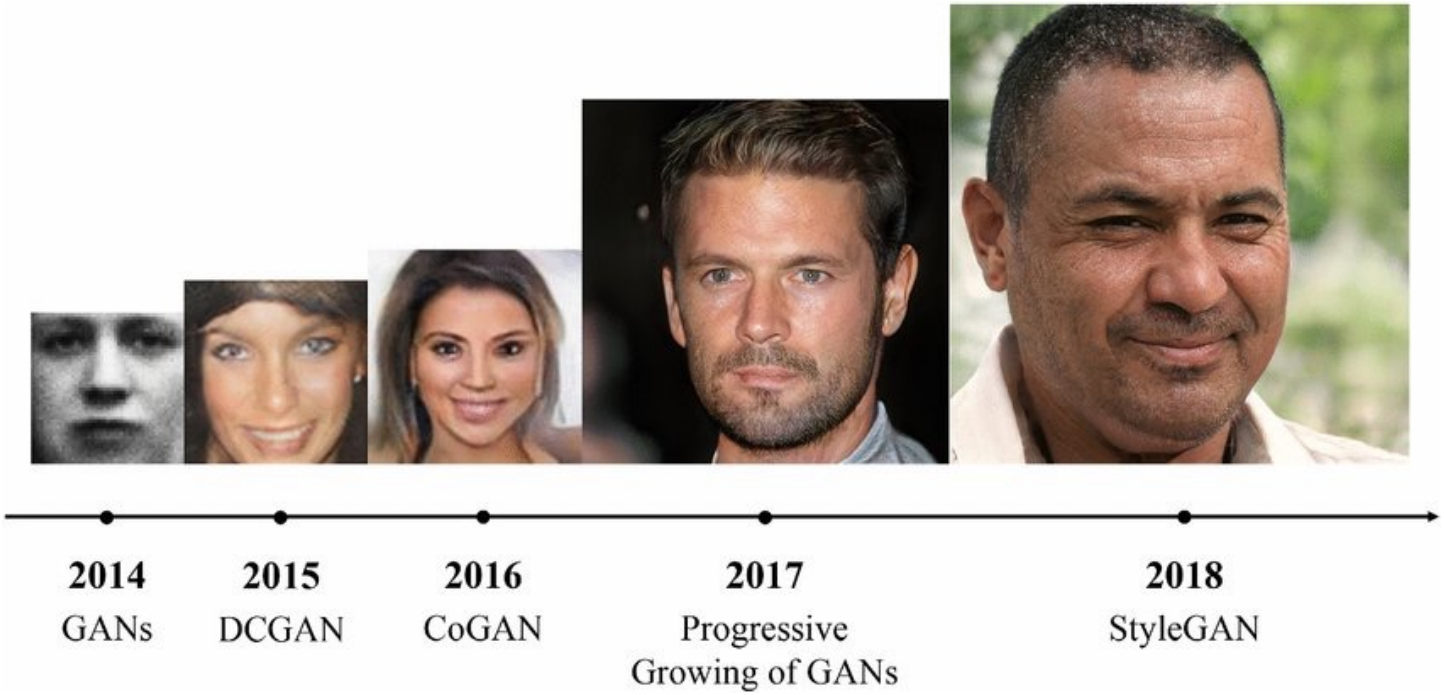
Examples of Images Generated by GANs



PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Ejemplos de caras generadas artificialmente

Examples of Images Generated by GANs



PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Otras aplicaciones de las GANs

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Otras aplicaciones de las GANs

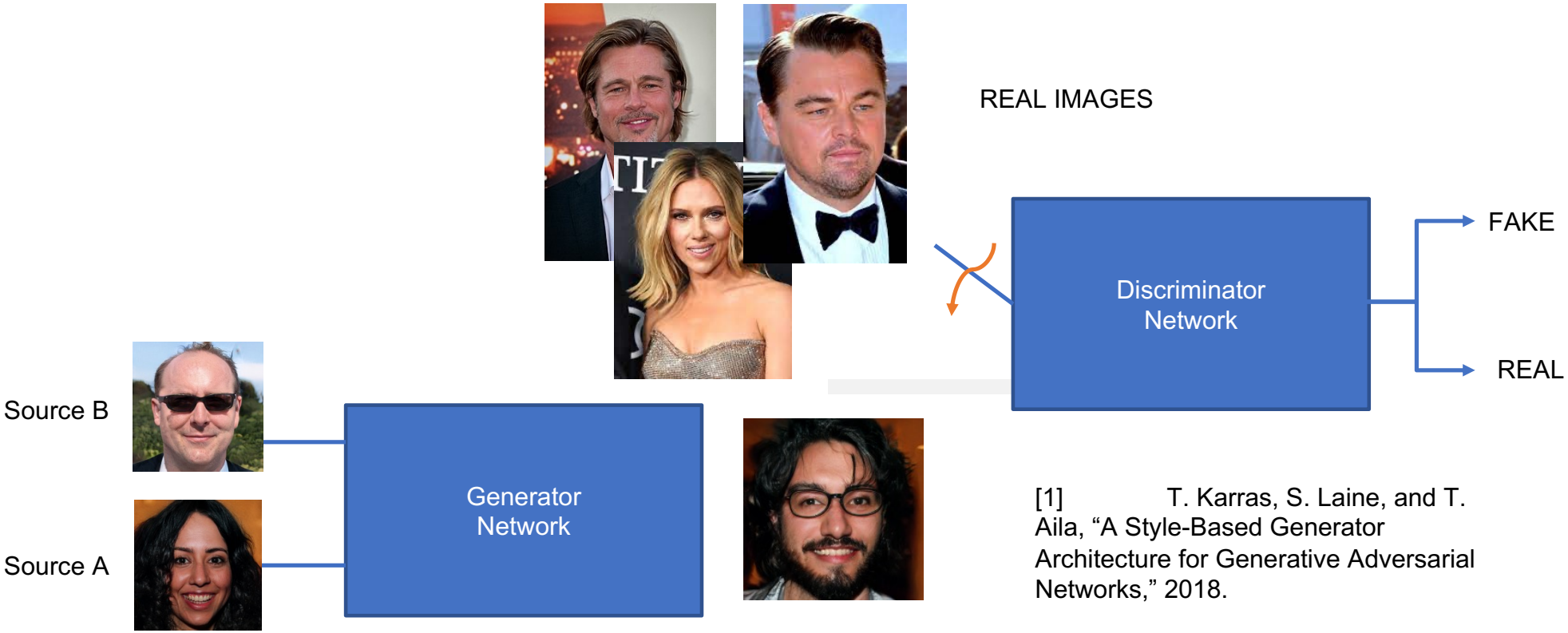
PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Otras aplicaciones de las GANs

Map to Aerial and Aerial to Photo

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Control de Características StyleGan1, 2 & 3



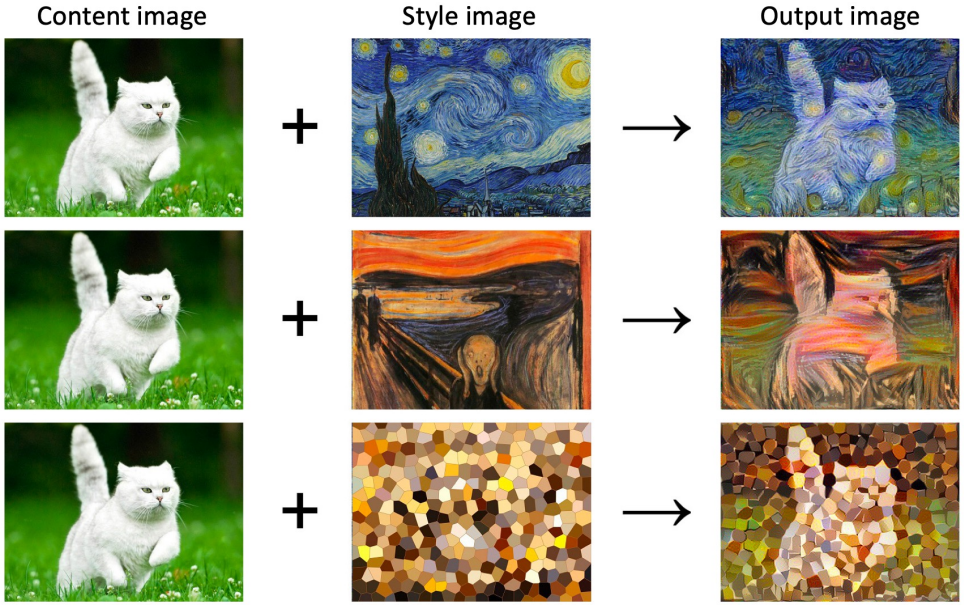
[1] T. Karras, S. Laine, and T. Aila, “A Style-Based Generator Architecture for Generative Adversarial Networks,” 2018.

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Conceptos previos: Transferencia de Estilo

StyleGAN is a method for fake face synthesis that gives some control to the person generating the images.

The idea is based on previous methods of Style Image Generation: STYLE TRANSFER



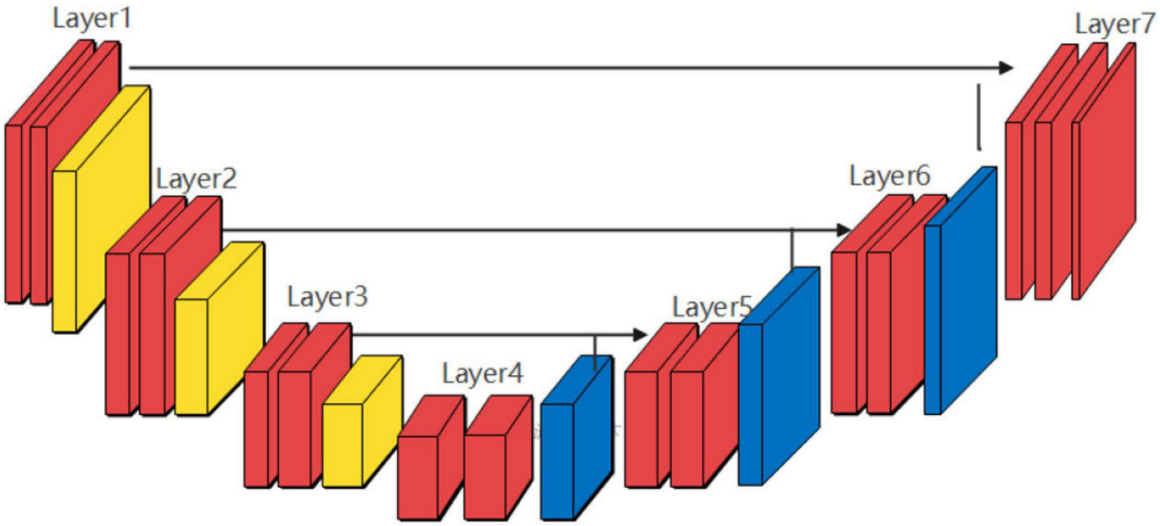
X. Huang and S. Belongie, “Arbitrary style transfer in real-time with adaptive instance normalization,” *5th Int. Conf. Learn. Represent. ICLR 2017 - Work. Track Proc.*, 2019.

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Conceptos previos: U-Nets, Autoencoders, VAE

U-Net Idea

Autoencoders



Conv+BatchNormalization+ReLU

Pooling operation

Upsampling Layer

Skip-Connection

PROYECTO iveres

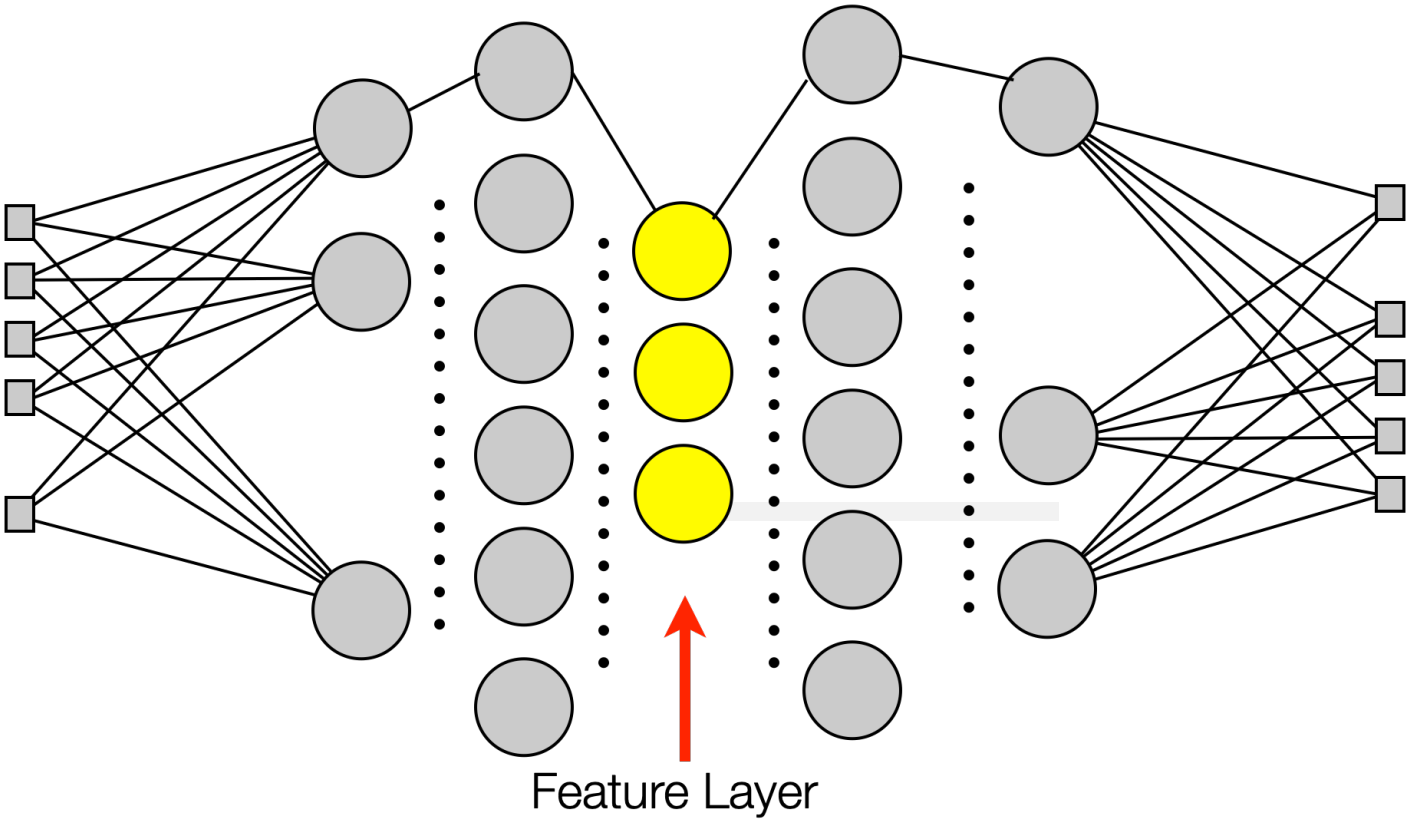
UPC

rtve

Plan de Recuperación, Transformación y Resiliencia

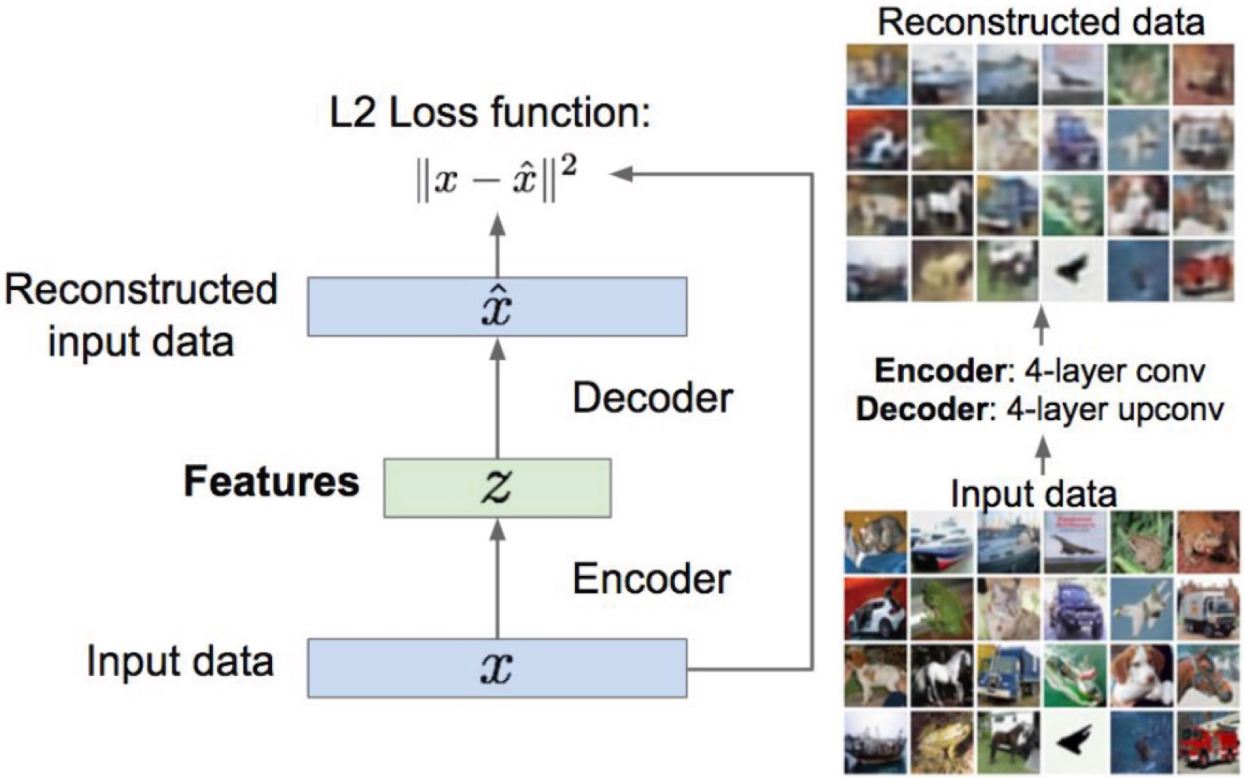
Financiado por la Unión Europea NextGenerationEU

Conceptos previos: U-Nets, Autoencoders, VAE



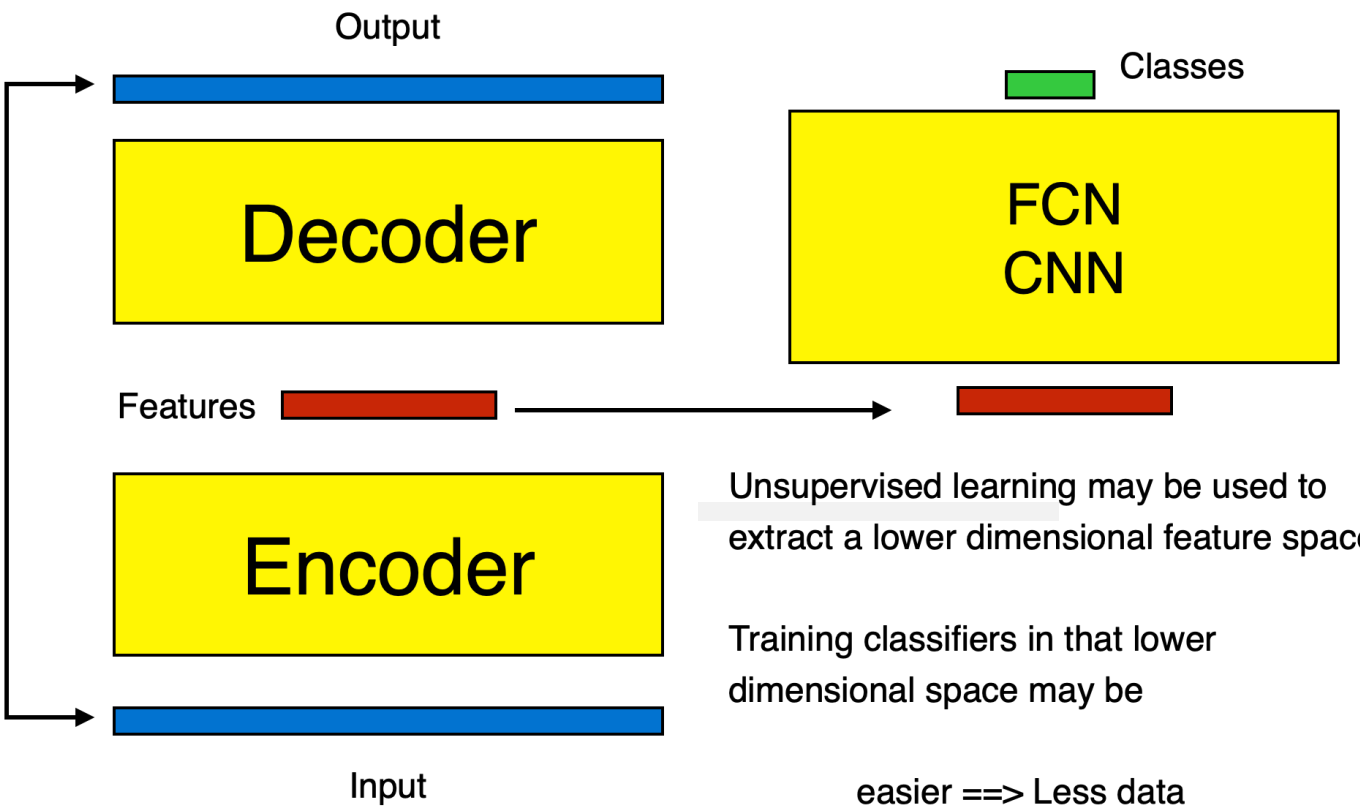
PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Conceptos previos: U-Nets, Autoencoders, VAE



PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Conceptos previos: U-Nets, Autoencoders, VAE



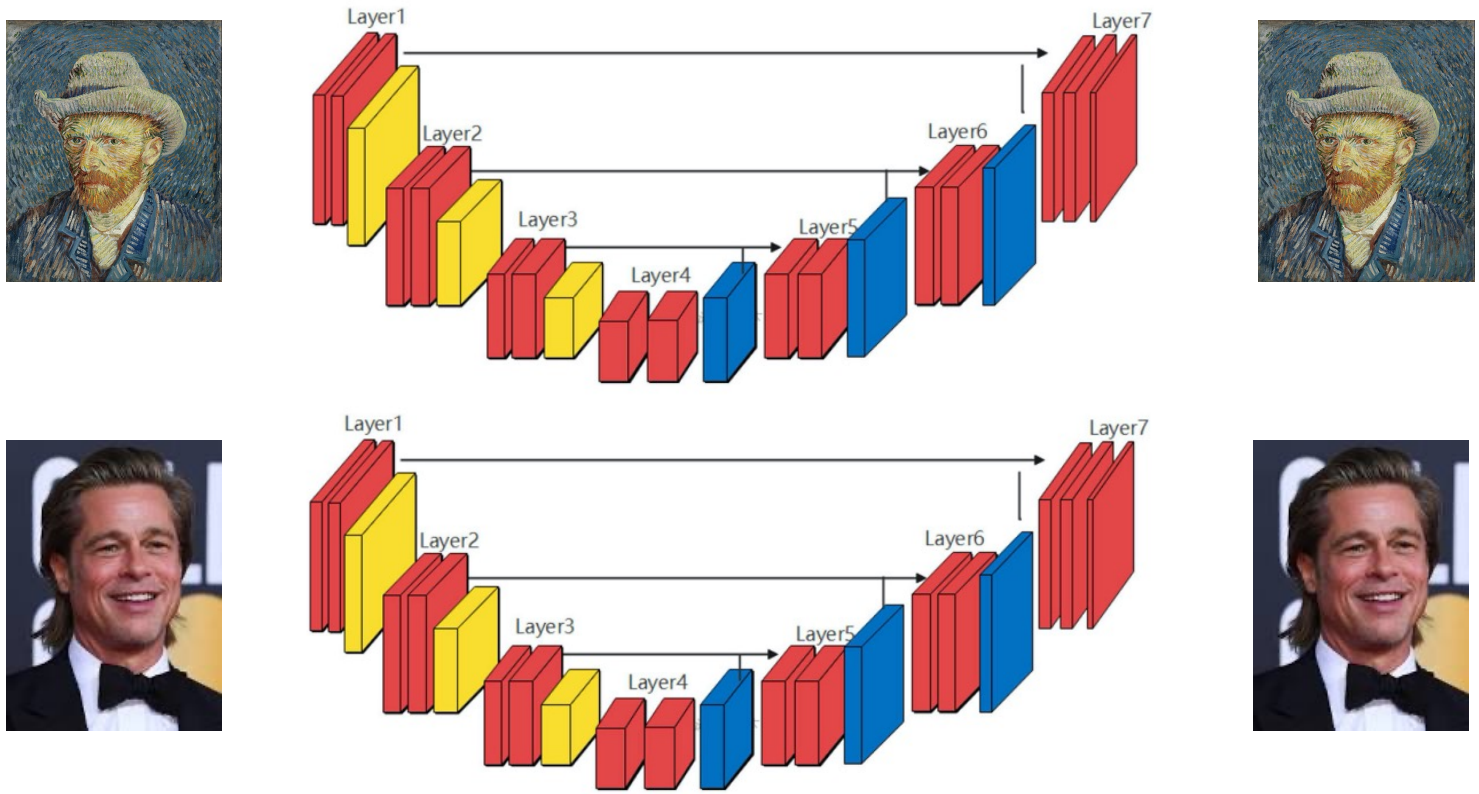
Unsupervised learning may be used to extract a lower dimensional feature space.

Training classifiers in that lower dimensional space may be

easier ==> Less data

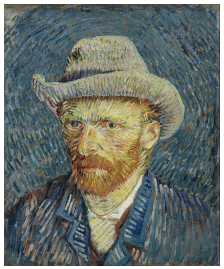
PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

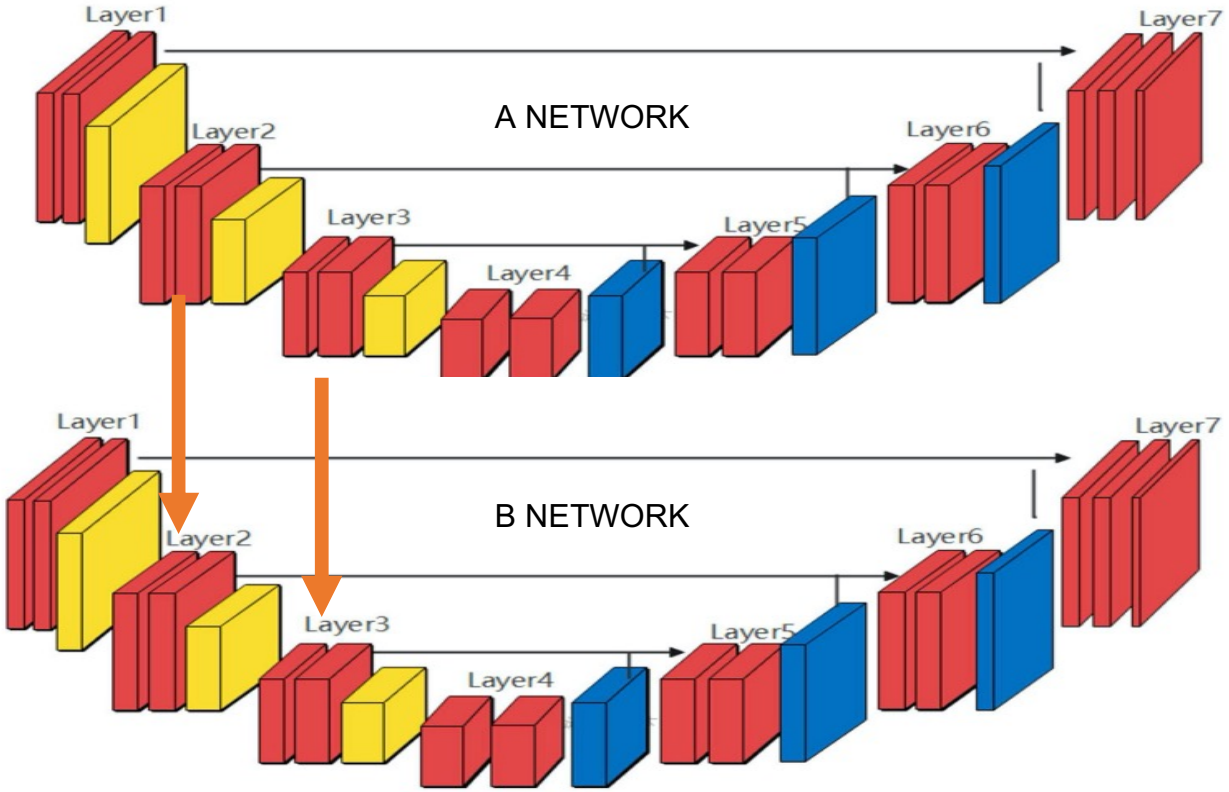
Conceptos previos: Transferencia de Estilo



PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

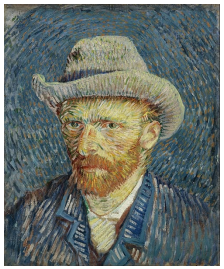
Conceptos previos: Transferencia de Estilo

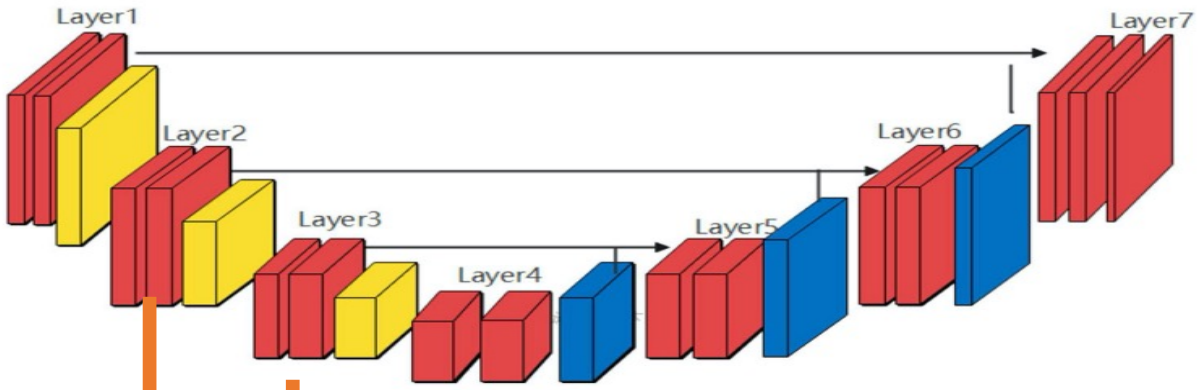




PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Conceptos previos: Transferencia de Estilo





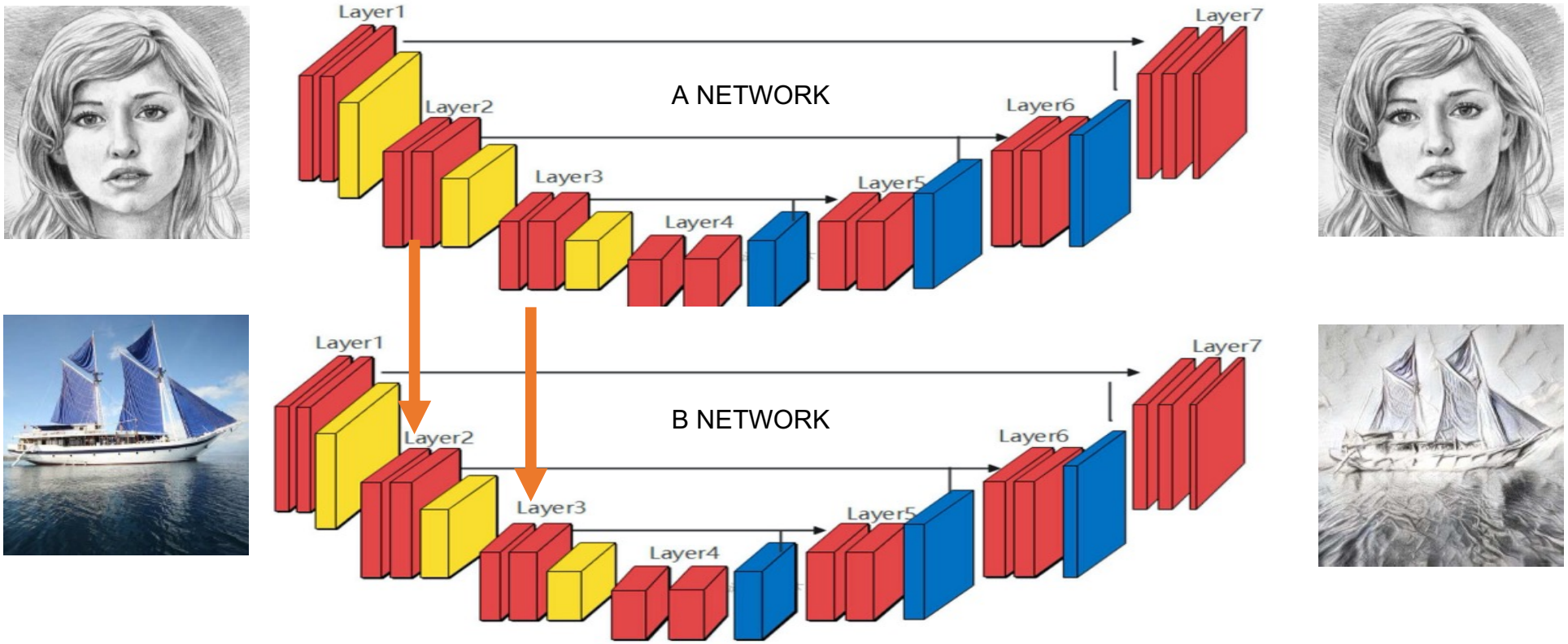
$$\mathrm{AdaIN}(\mathbf{x}_i, \mathbf{y}) = \mathbf{y}_{s,i}\frac{\mathbf{x}_i - \mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)} + \mathbf{y}_{b,i},$$

TRANSFERING THE STYLE:

1. COMPUTE MEAN AND VARIANCE OF SOME LAYER CHANNELS
2. MAKE AN AFFINE TRANSFORMATION TO THE COEFFICIENTS OF B NETWORK TO FORCE THAT THE SELECTED LAYERS HAVE THE SAME MEAN AND VARIANCE OF THE A NETWORK

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Conceptos previos: Transferencia de Estilo

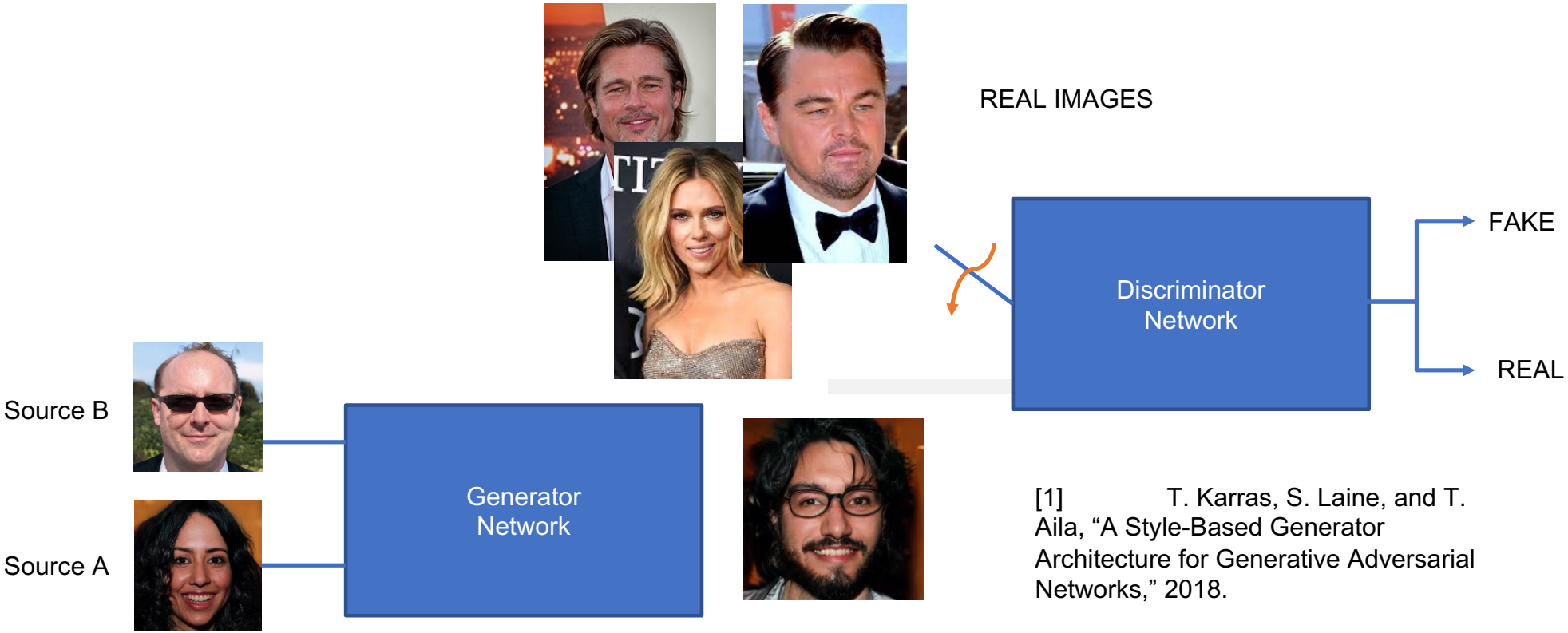


PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Conceptos previos: Transferencia de Estilo

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

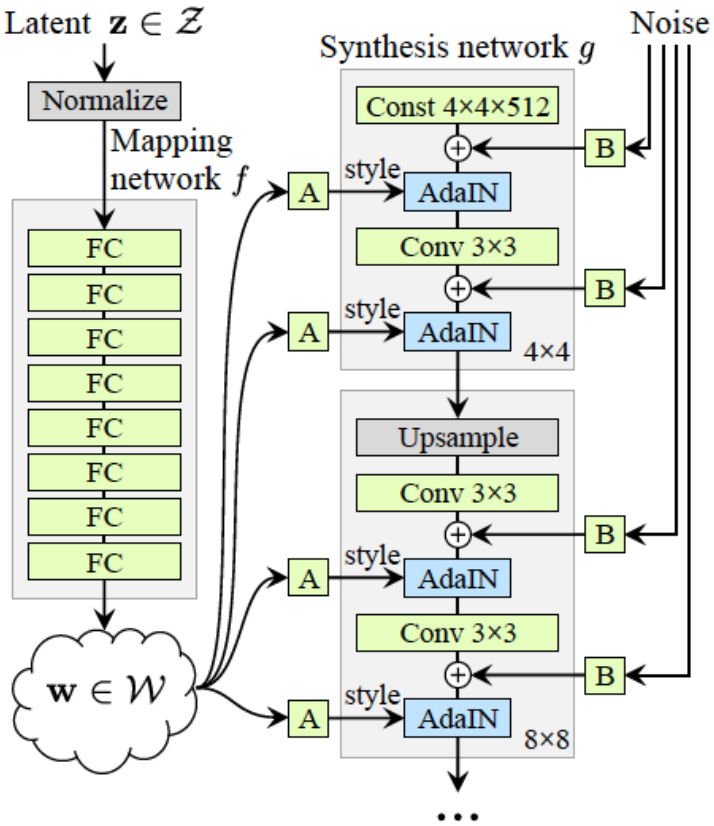
StyleGAN



[1] T. Karras, S. Laine, and T. Aila, “A Style-Based Generator Architecture for Generative Adversarial Networks,” 2018.

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

StyleGAN



Se calculan los pesos para dos imágenes de referencia:

Latent Za, Latent Zb

Se selecciona una capa y en todas las capas anteriores se transfiere el estilo de Za mientras que en las superiores se transfiere el estilo de Zb

Se introduce un ruido blanco para que la red pueda tener suficientes grados de libertad para generar caras originales.

[1] T. Karras, S. Laine, and T. Aila, “A Style-Based Generator Architecture for Generative Adversarial Networks,” 2018.

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

StyleGAN

Globales: Pose, forma cara, edad gafas from B … detalles color, pelo, etc from A

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

StyleGAN2

PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

StyleGAN2

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Ejemplos de generación StyleGAN

This person does not exist: Genera una nueva cara cada vez que refrescamos la URL

https://www.thispersondoesnotexist.com

StyleGAN NADA Playground

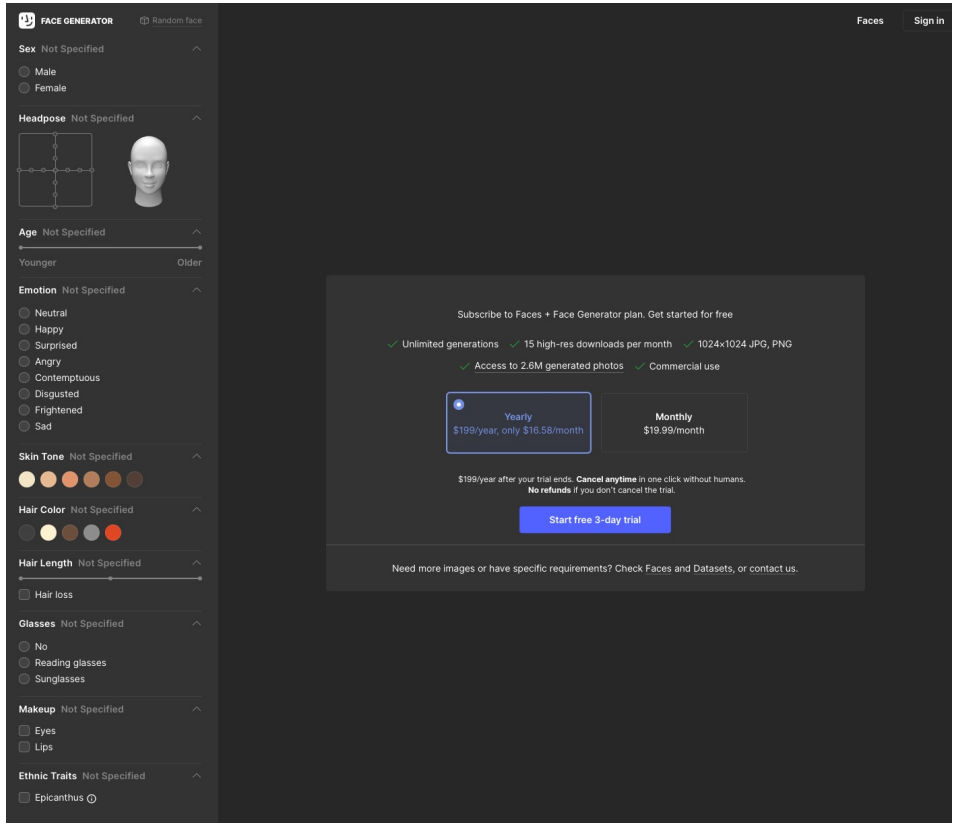
https://www.kaggle.com/code/ratthachat/stylegan-nada-playground

Generated Photos

https://generated.photos/face-generator/new#

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Ejemplos de generación StyleGAN

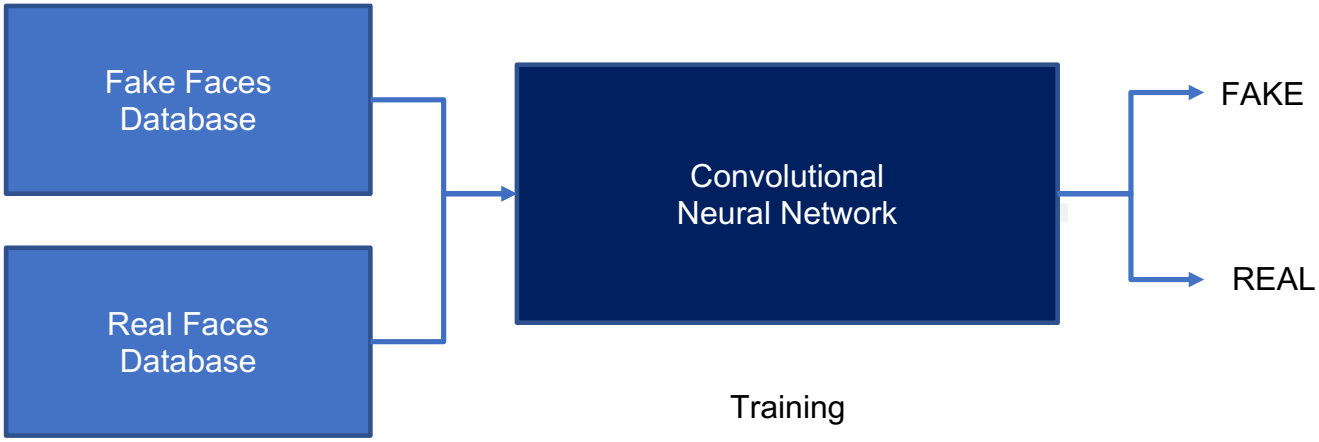


PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Detección de Fakes

Entire Face Synthesis Detection

Every GAN imprints a fingerprint in the image. The detection system has to detect that fingerprint.
DRAWBACK: Detection methods are specialized on specific software (methods, generation models)

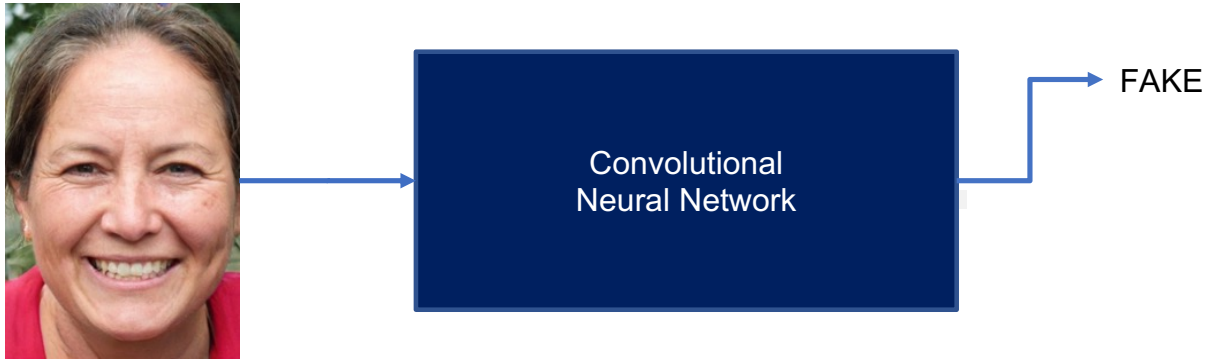


PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Detección de Fakes

Entire Face Synthesis Detection

Every GAN imprints a fingerprint in the image. The detection system has to detect that fingerprint.
DRAWBACK: Detection methods are specialized on specific software (methods, generation models)



Predicting

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Detección de Fakes

Algunas técnicas para la detección de fakes.

1. Redes neuronales especificamente entrenadas para detectar las imágenes falsas a partir de bases de datos de ejemplos.

2. Uso de transformadas tipo FFY, Wavelet para detectar patrones en frecuencia propios de las GAN

3. PRNU (métodos clásicos). Detectar patrones de ruido típicos en las cámaras que no existen en las imágenes generadas de forma artificial.

4. Geometría, Perspectiva, patrones de iluminación.

5. Software específico: Deepware, Sensity AI, DeepFake-o-meter, Microsoft Video Authenticator.

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Manipulación de imágenes (Deep Fake)

Técnicas más habituales

1. Generación de caras ’nuevas’. Avatares
2. Generación con control de características.
3. Face Reenactment
4. Face Swapping
5. Talking Face
6. Text-driven generative images
7. In-Painting & Out-Painting Text-driven generative images

PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Face Reenactment & Face Swapping

Face Reenactment

Modificar las expresiones faciales de una persona en un vídeo utilizando las expresiones de otra persona

El proceso básico es:

1) Detección y seguimiento facial (puntos clavel), 2) Captura expresiones, 3) Transferencia de Expresiones, 4) Renderizado.

Aplicaciones: Cine y animación, Mejora interacción digital, Revivir personajes históricos, entretenimiento

Mantiene la identidad de la persona pero cambia sus expresiones faciales

Face Swapping

Intercambiar los rostros de dos personas en una imagen o vídeo, de modo que cada persona tenga el rostro de la otra

Se reemplazan los rostros manteniendo el resto del cuerpo y el entorno intactos.

Cambia completamente el rostro de una persona por el de otra, manteniendo sus expresiones faciales

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Face Reenactment & Face Swapping

Face Reenactment

- Manipular expresiones. Alterar expresiones en vídeos de figuras públicas. Manipular la percepción pública y alterar la interpretación de eventos.
- Genera desinformación.
- Alterar vídeos de testimonios.
- Extorsionar a personas manipulando videos en situaciones embarazosas o comprometedoras.

Face Swapping

- Falsificación de identidad. Persona puede estar realizando acciones que en realidad no realizó.
- Generación de contenido engañoso en el ámbito político
- Violación de privacidad (vídeos pornográficos)

PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Examples Face Reenactment

Face2Face project and software.

Proyecto de investigación en la Universidad de Erlangen

https://youtu.be/KUjn6SrNbSo

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Examples Face Swapping

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

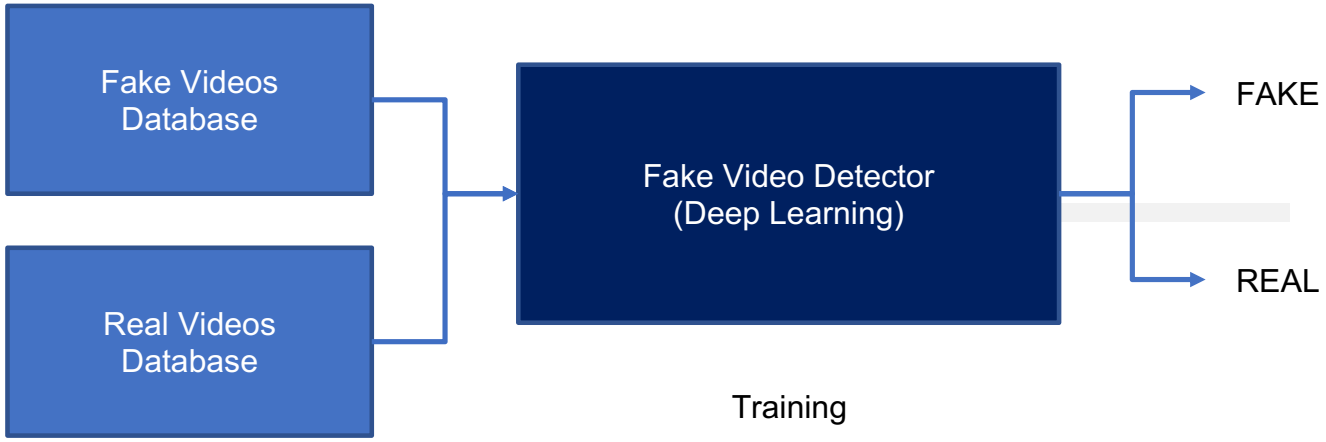
Examples Face Swapping

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Face Swapping & Reenactment

Deep Fake Detection Challenge Database

Creada por AWS, Facebook, Microsoft, U Munich, U Naples, MIT, Cornell Tech, Berkeley, U Maryland, U Albany, etc.
Generada usando actores profesionales con 8 tecnologias diferentes. Se crearon un total de 115.000 videos falsos.
El challenge tenía un premio de 1M$.
Se presentaron un total de 35109 modelos de detección de fake distintos



En el proyecto se han seleccionado dos de las propuestas que han dado mejores resultados

- Deepware Face Detector
- Polimi Fake Detector.

https://youtu.be/KUjn6SrNbSo

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Face Swapping & Reenactment

Deep Fake Detection Challenge Database (FAKE Samples)

PROYECTO iveres

UPC

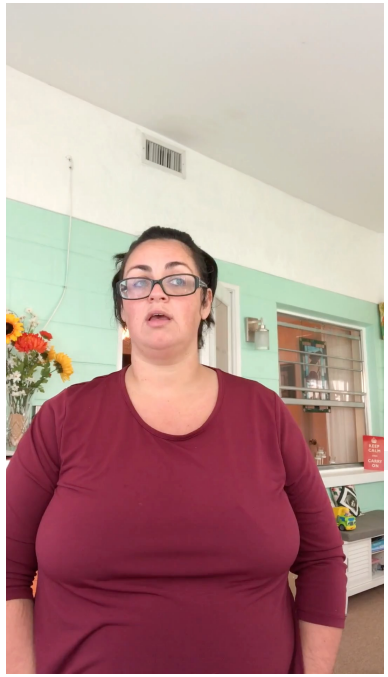
rtve

Plan de Recuperación, Transformación y Resiliencia

Financiado por la Unión Europea NextGenerationEU

Face Swapping & Reenactment

Deep Fake Detection Challenge Database (REAL Samples)



PROYECTO iveres

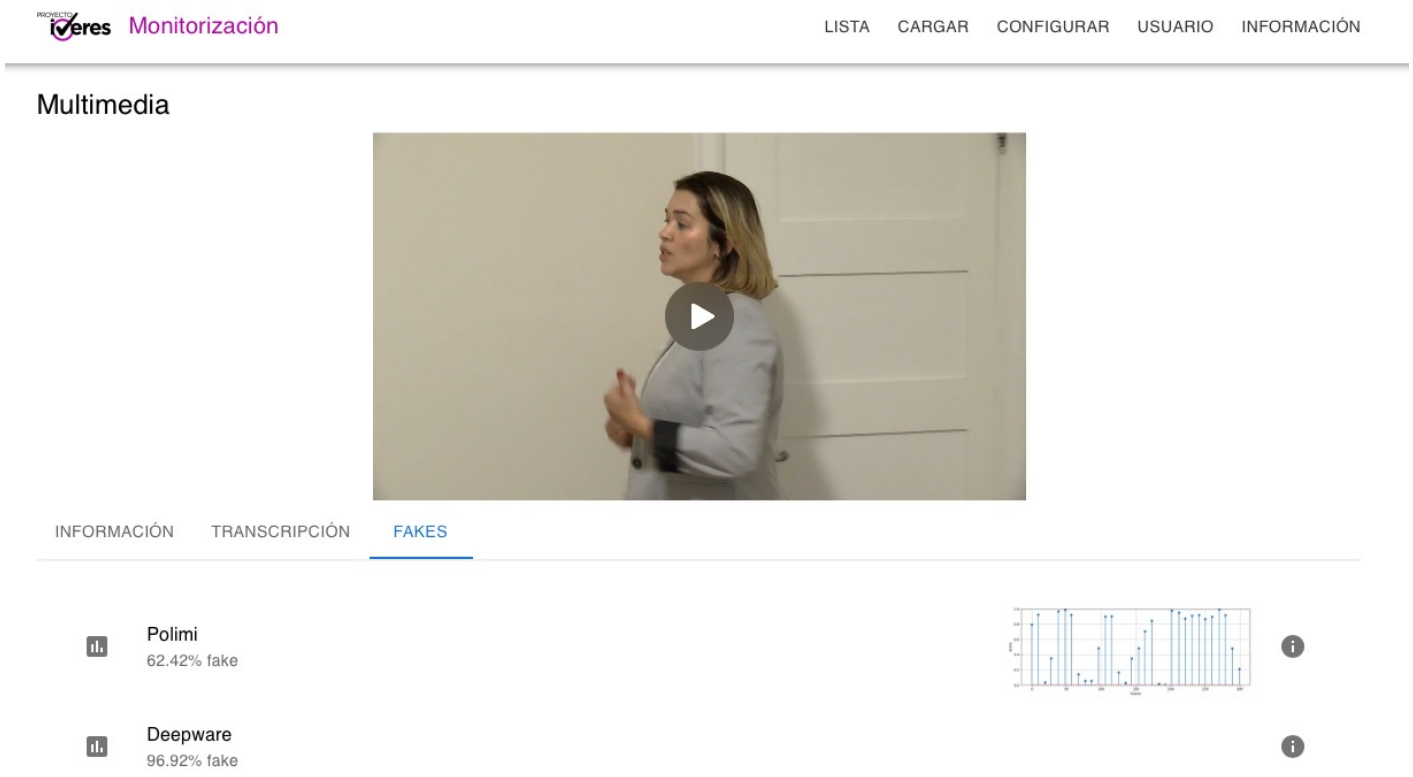
UPC

rtve

Plan de Recuperación, Transformación y Resiliencia

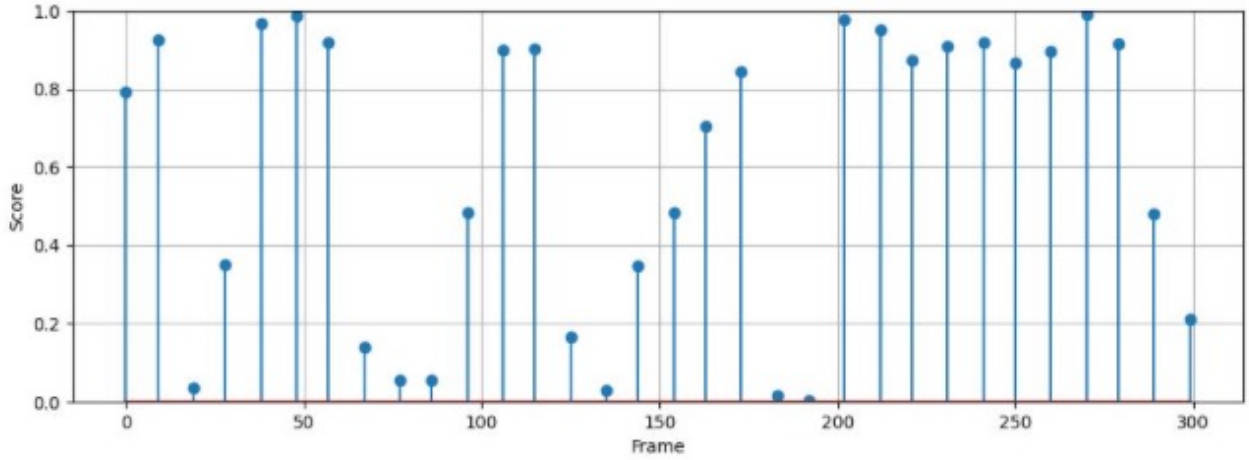
Financiado por la Unión Europea NextGenerationEU

Resultados



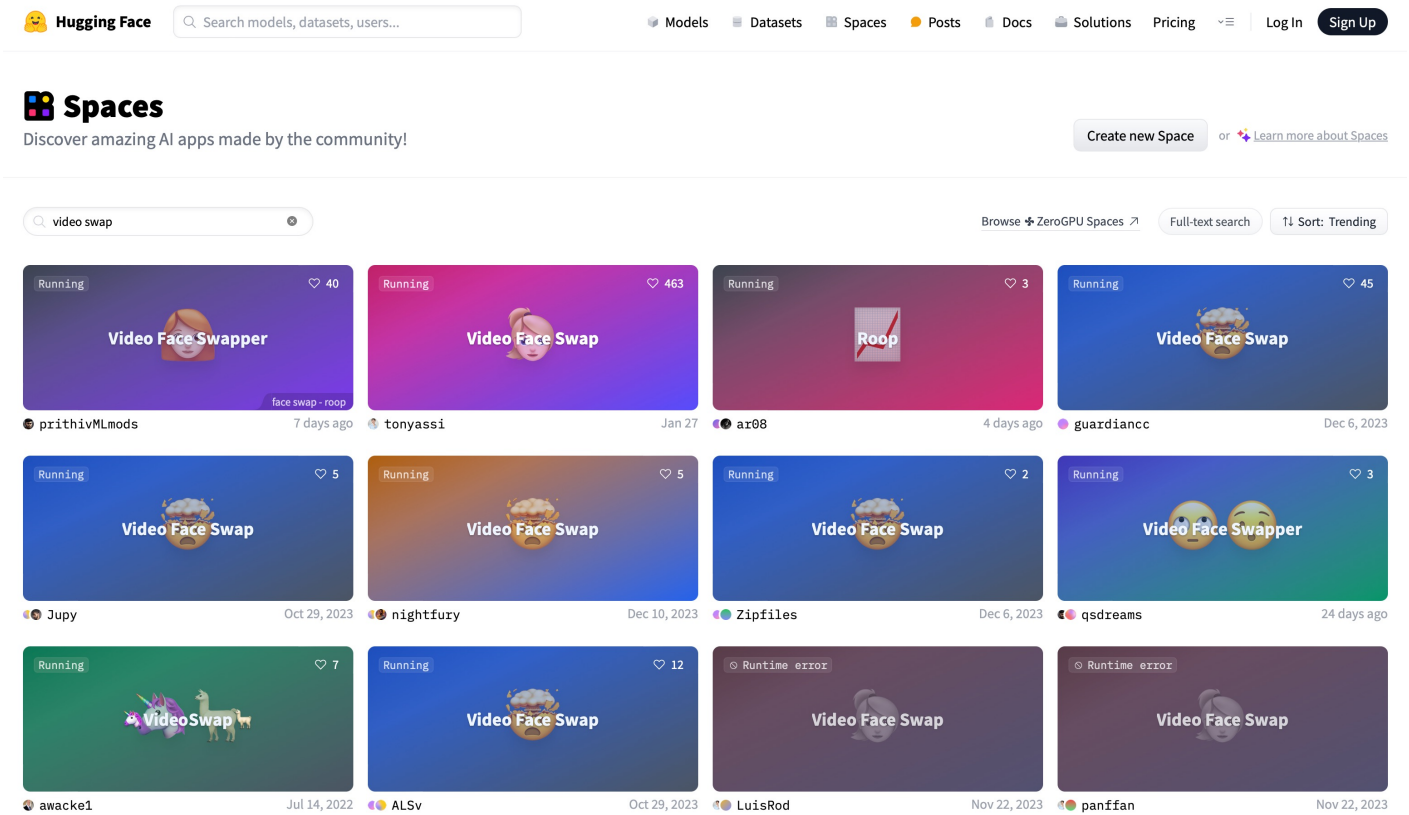
PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Resultados



PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Generar Face Swapping



https://huggingface.co/spaces

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Generar Face Swapping

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

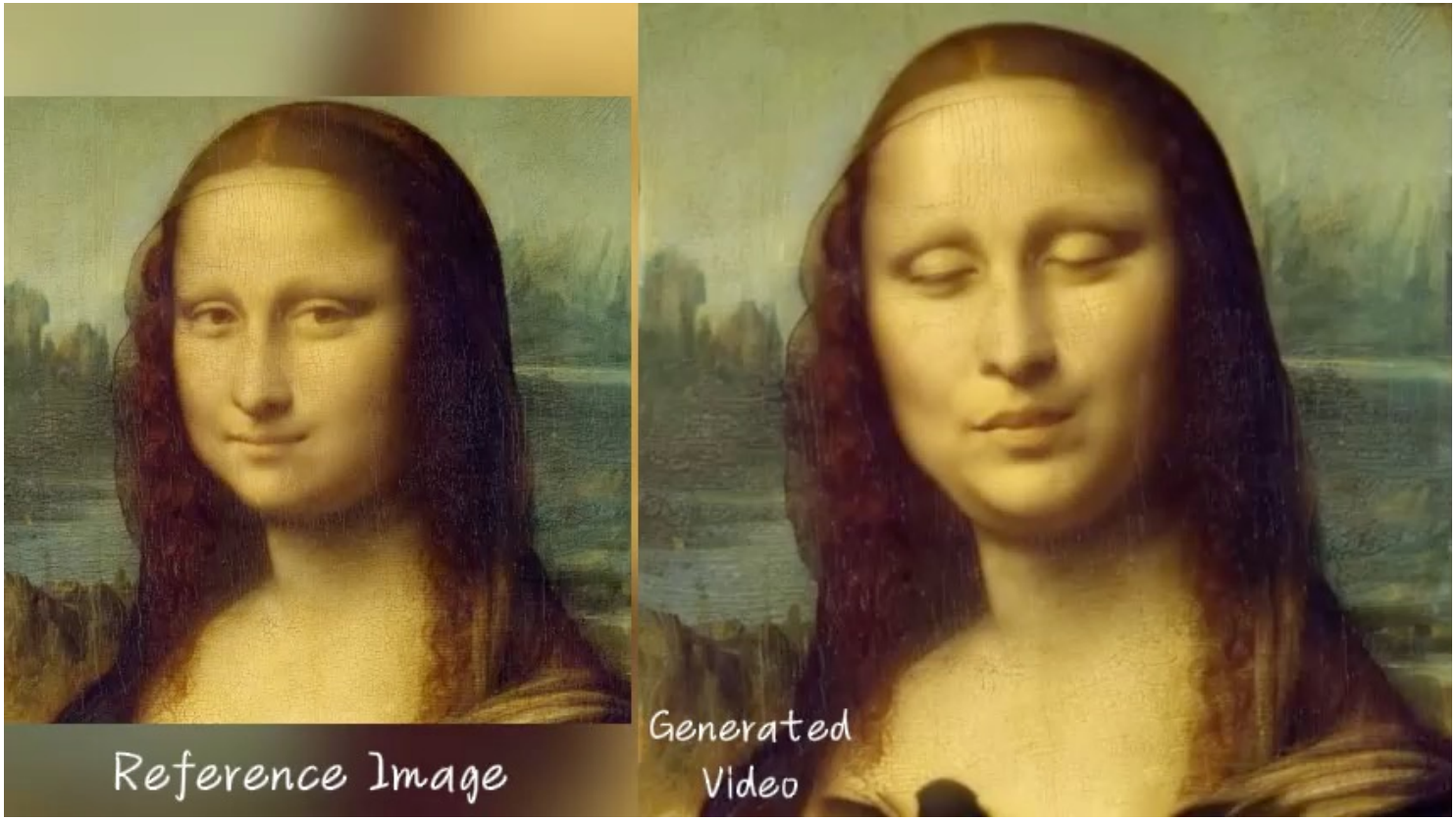
Talking Faces

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Singing Faces

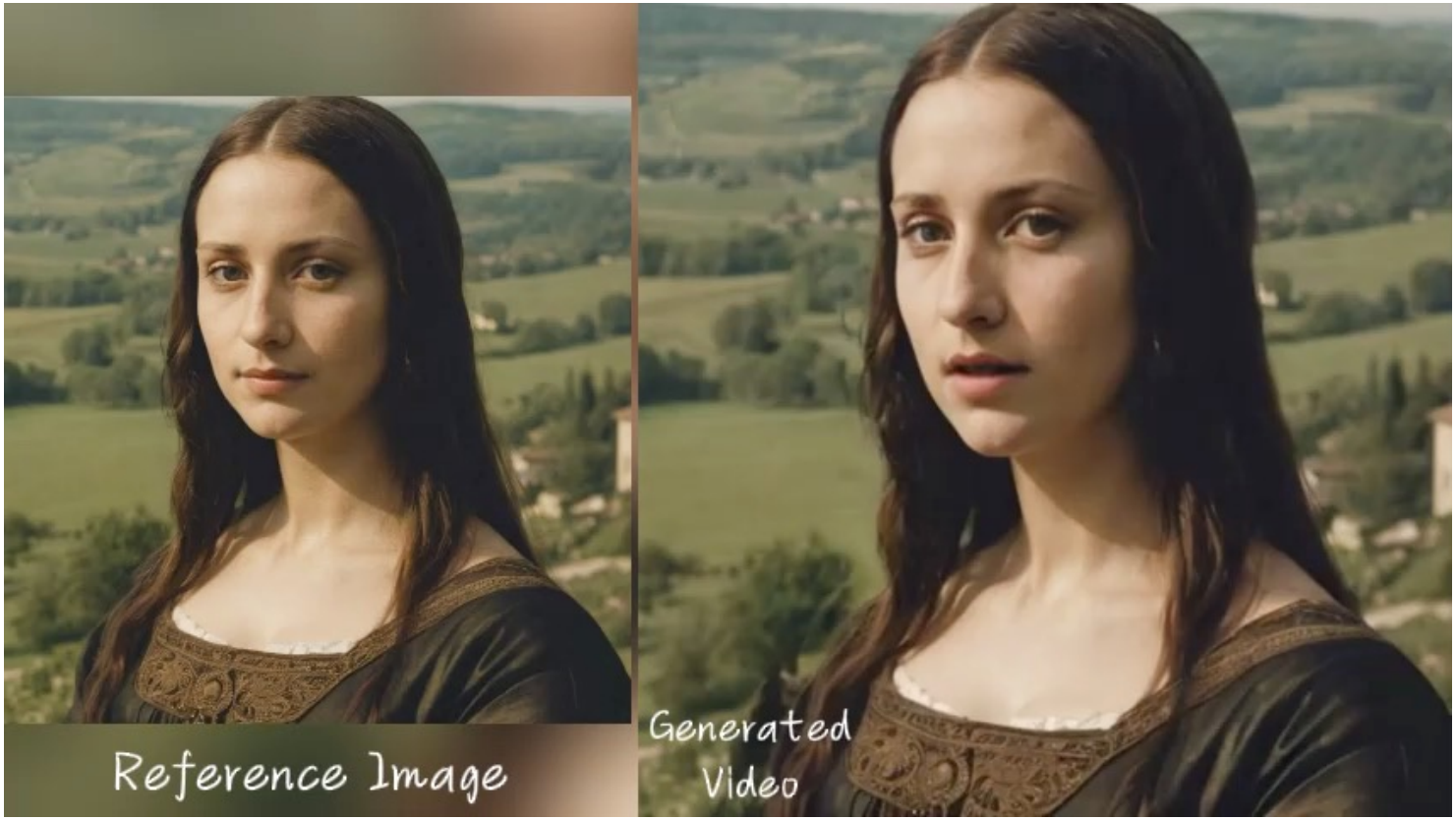
PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Talking Faces



PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Singing Faces



PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Difussion Models Generated Images

PROYECTO iveres

UPC

rtve

Plan de Recuperación, Transformación y Resiliencia

Financiado por la Unión Europea NextGenerationEU

Difussion Models Generated Images

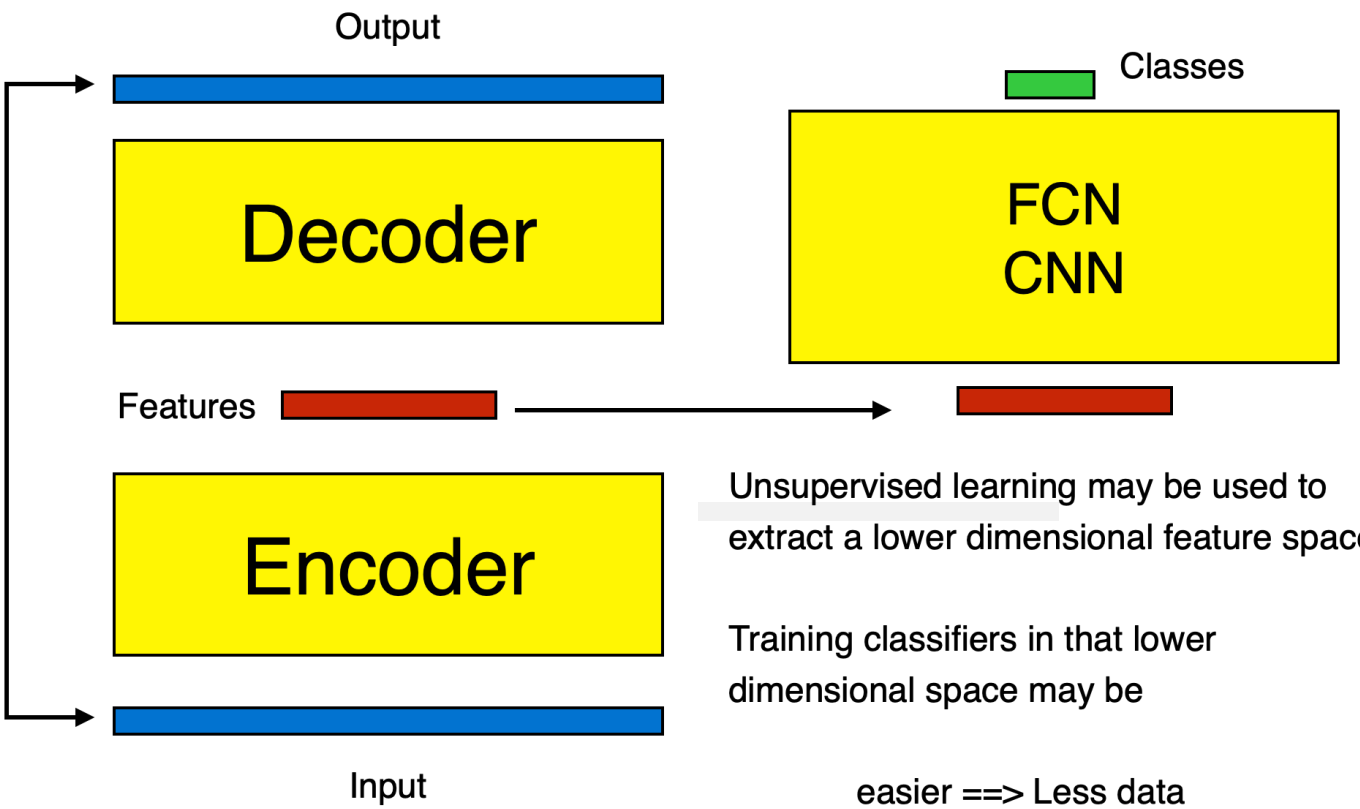
PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Difussion Models Generated Images

Jason Allen's A.I.-generated work, "Théâtre D'opéra Spatial," took first place in the digital category at the Colorado State Fair. via Jason Allen

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Simple Techniques to Text-to-Image



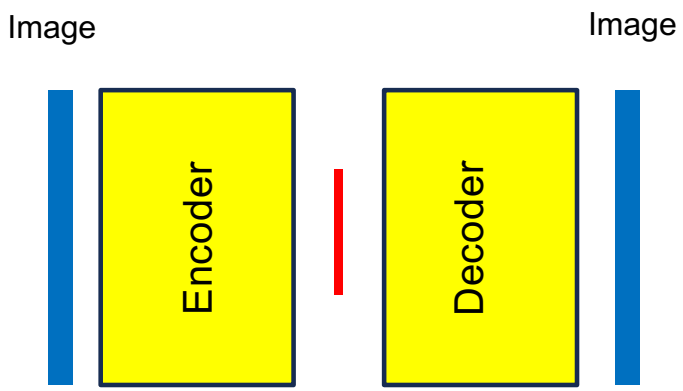
Unsupervised learning may be used to extract a lower dimensional feature space.

Training classifiers in that lower dimensional space may be

easier ==> Less data

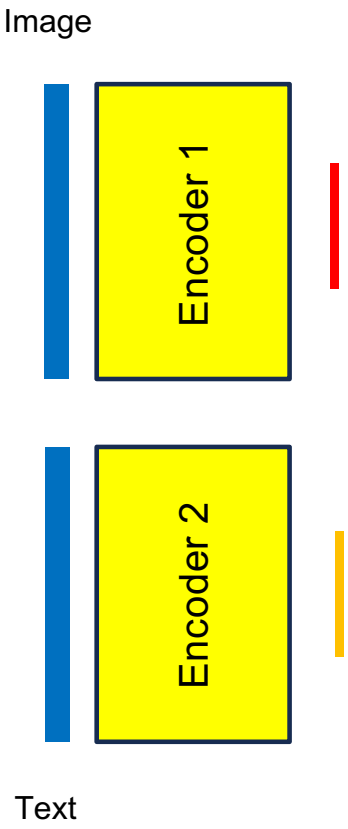
PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Simple Techniques to Text-to-Image



PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Simple Techniques to Text-to-Image

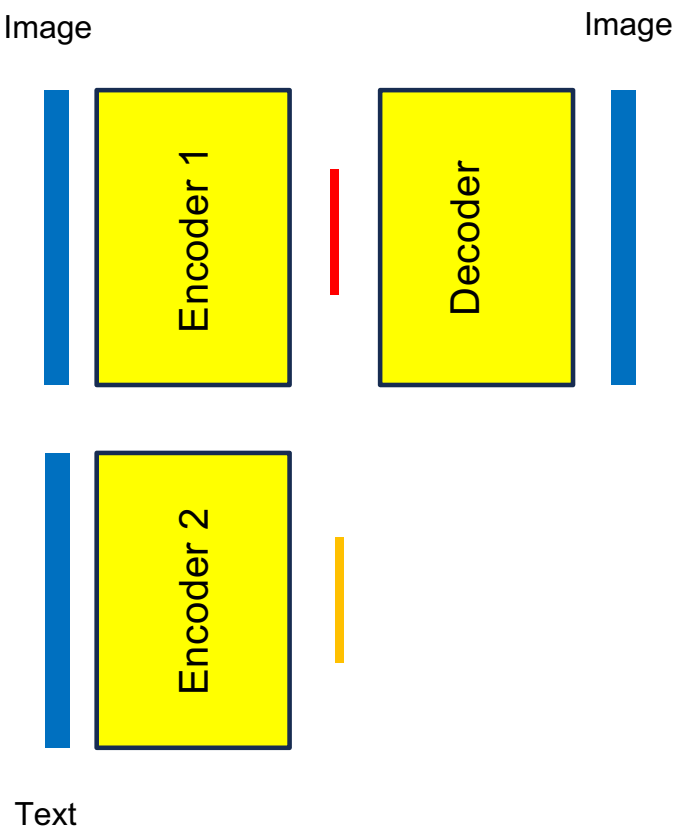


Training Database (400 M images + Descriptive Text)

Train the encoders to have the maximum similarity in the latent space when Text and Images correspond and maximum disimilarity when do not corresponde

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Simple Techniques to Text-to-Image

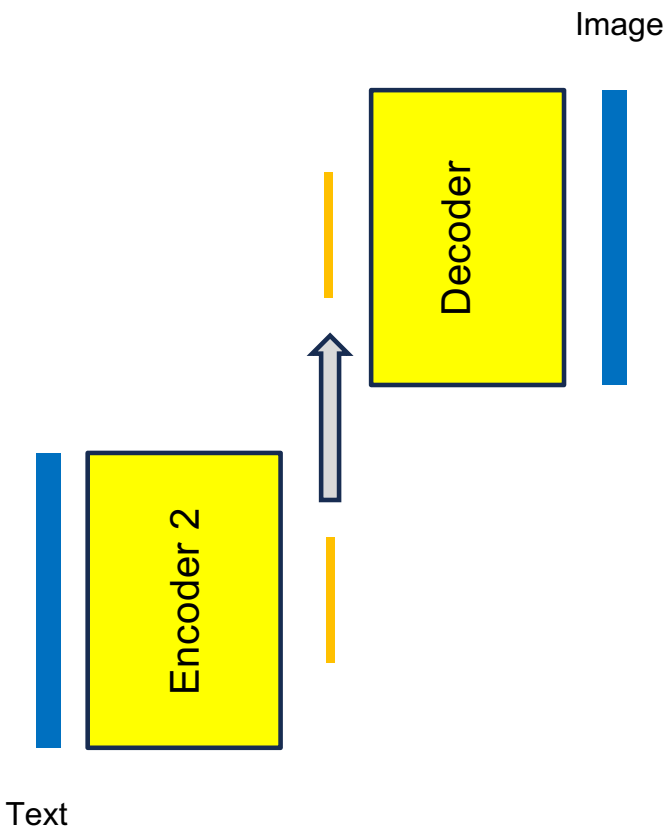


Congelar los 2 encoders

Entrenar el Decoder para que produzca en la salida la misma imagen que en la entrada

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

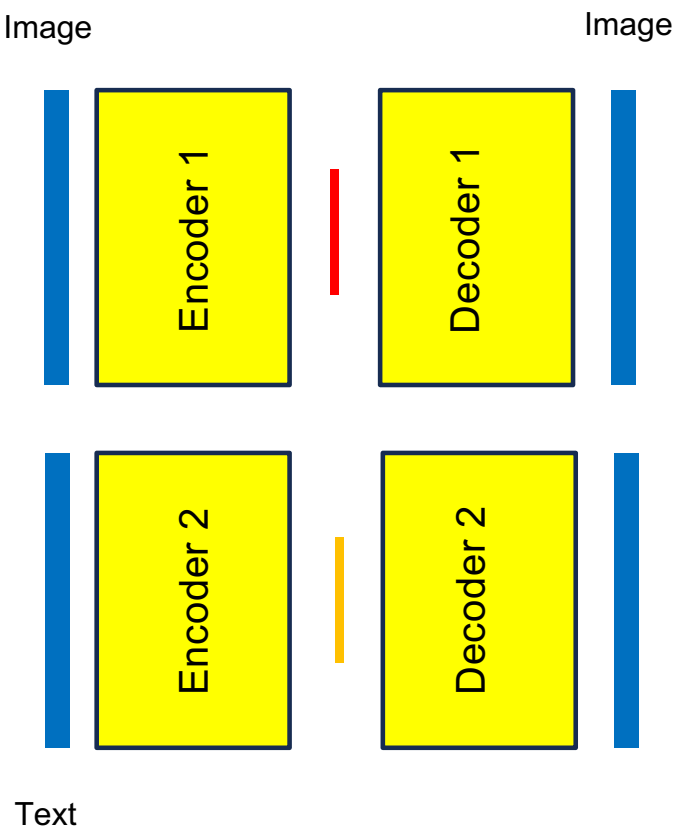
Simple Techniques to Text-to-Image



Simple Text-to-Image Generator

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Simple Techniques to Text-to-Image



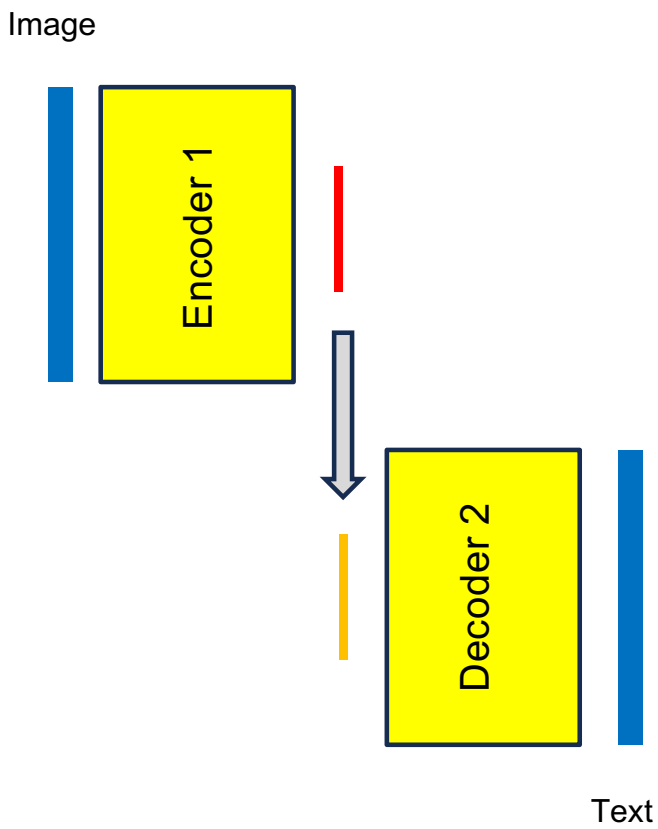
Congelar los 2 encoders

Entrenar Decoder 1 para que produzca en la salida la misma imagen que en la entrada

Entrenar Decoder 2 para que produzca en la salida el mismo texto que en la entrada

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Simple Techniques to Text-to-Image



Congelar los 2 encoders

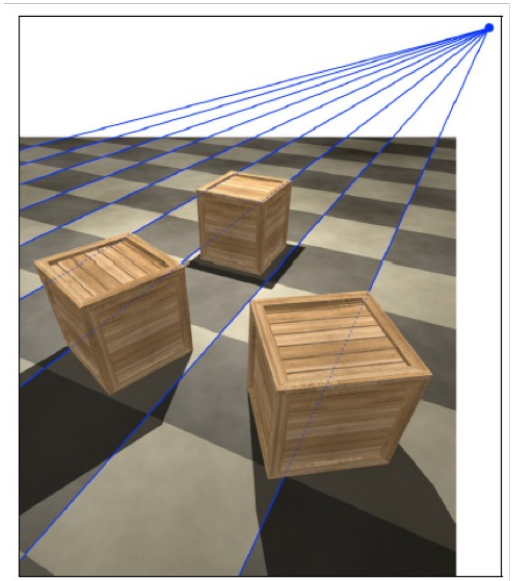
Entrenar Decoder 1 para que produzca en la salida la misma imagen que en la entrada

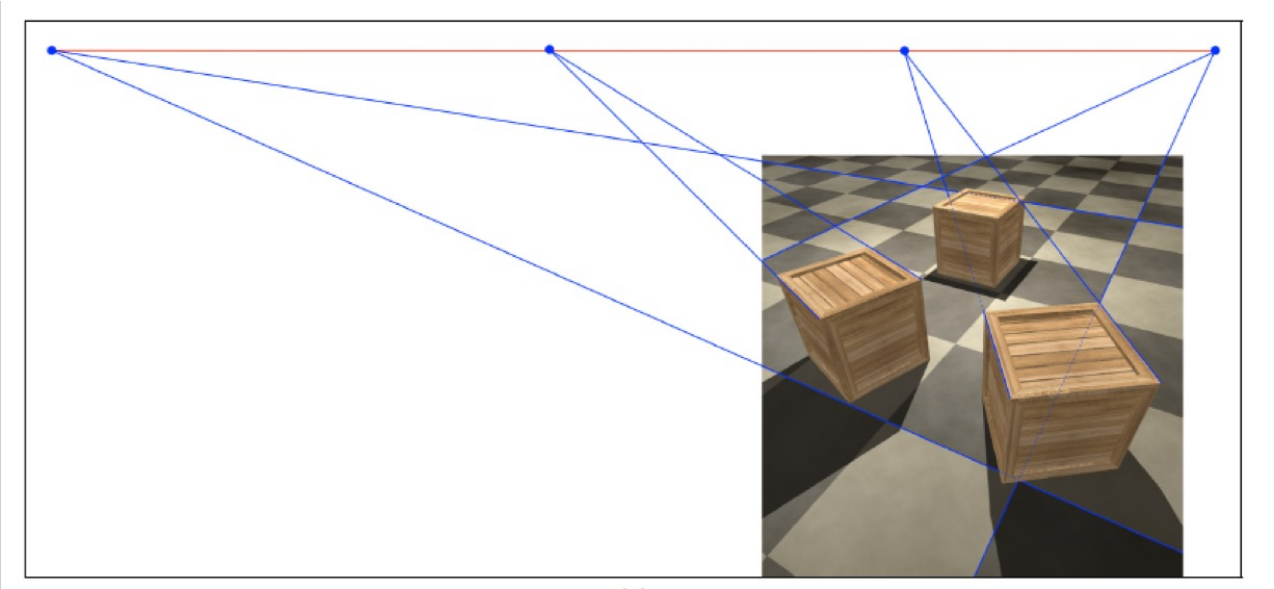
Entrenar Decoder 2 para que produzca en la salida el mismo texto que en la entrada

CLIP (main idea). Los modelos existentes en codificadores y decodificadores son complejos y generalmente basados en transformers

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Detecting Fakes: Perspective Analysis

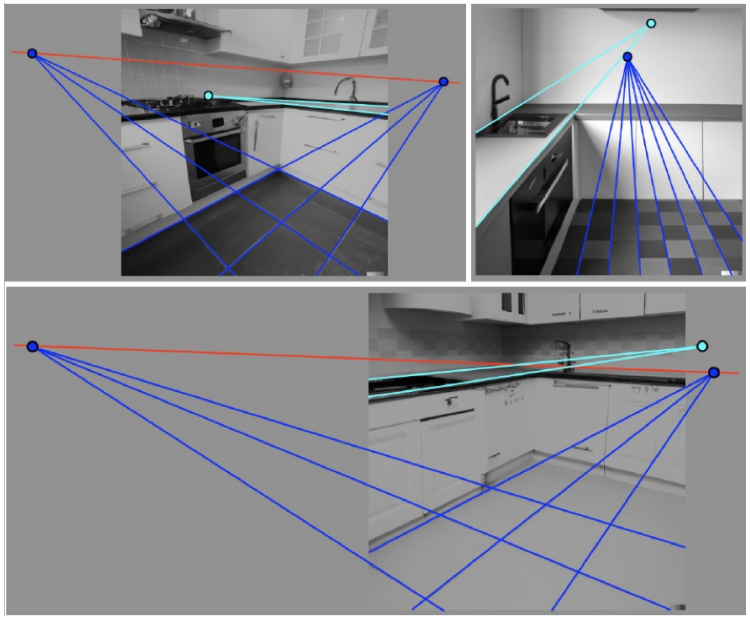




1.H. Farid. Perspective (In)consistency of Paint by Text. *arXiv:2206.14617*, 2022. [paper]

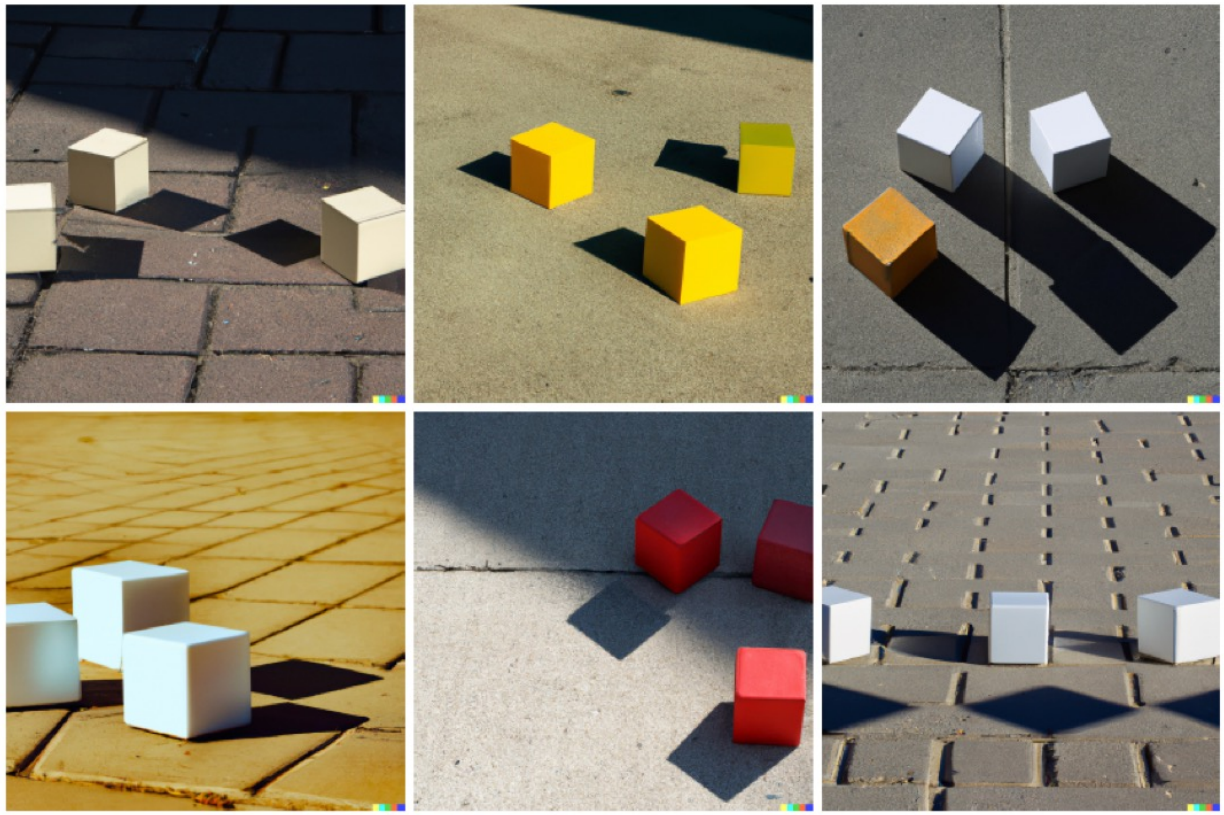
PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Detecting Fakes: Perspective Analysis



PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Detecting Fakes: Shadow Analysis



PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Detecting Fakes: Reflection Analysis

PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Detecting Fakes: Hands Analysis

PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

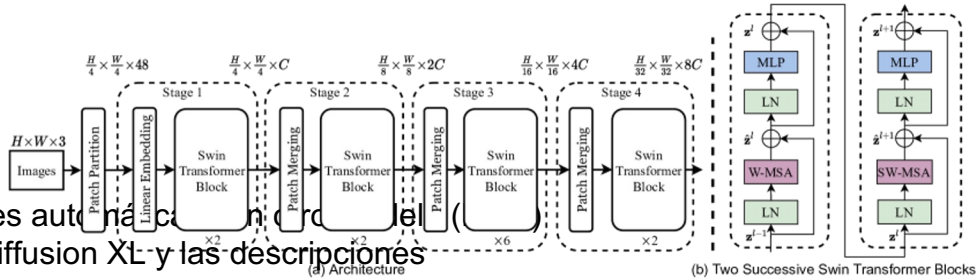
Detección de fakes

- La mayoría de métodos actuales de detección están basados en redes de aprendizaje profundo
- Se preentrenan de manera no supervisada para que aprendan a extraer patrones útiles y posteriormente se ajustan con un finetunning con etqietas supervisadas
- Las arquitecturas más utilizadas son redes convolucionales (CNN) o transformers para visión (ViT)
- En la mayoría de casos se plantea la detección como un problema clasificación de imágen entre las classes real vs fake o en algunos casos los métodos de generación (real, stable diffusion, dall-e, …)

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

SDXL-Detector

- https://huggingface.co/Organika/sdxl-detector
- Método actualmente implementado en Iveres
- Arquitectura
  - Swin Transformer
  - Transformer de visión jerárquico
- Datos de entrenamiento
  - Preentrenado con 1.28M imágenes (ImageNet)
  - Finetunning
    - Imágenes reales extraídas de Wikimedia
    - Para las imágenes reales se generan descripciones automáticamente con otro modelo (
    - Imágenes fake generadas con el modelo Stable Diffusion XL y las descripciones

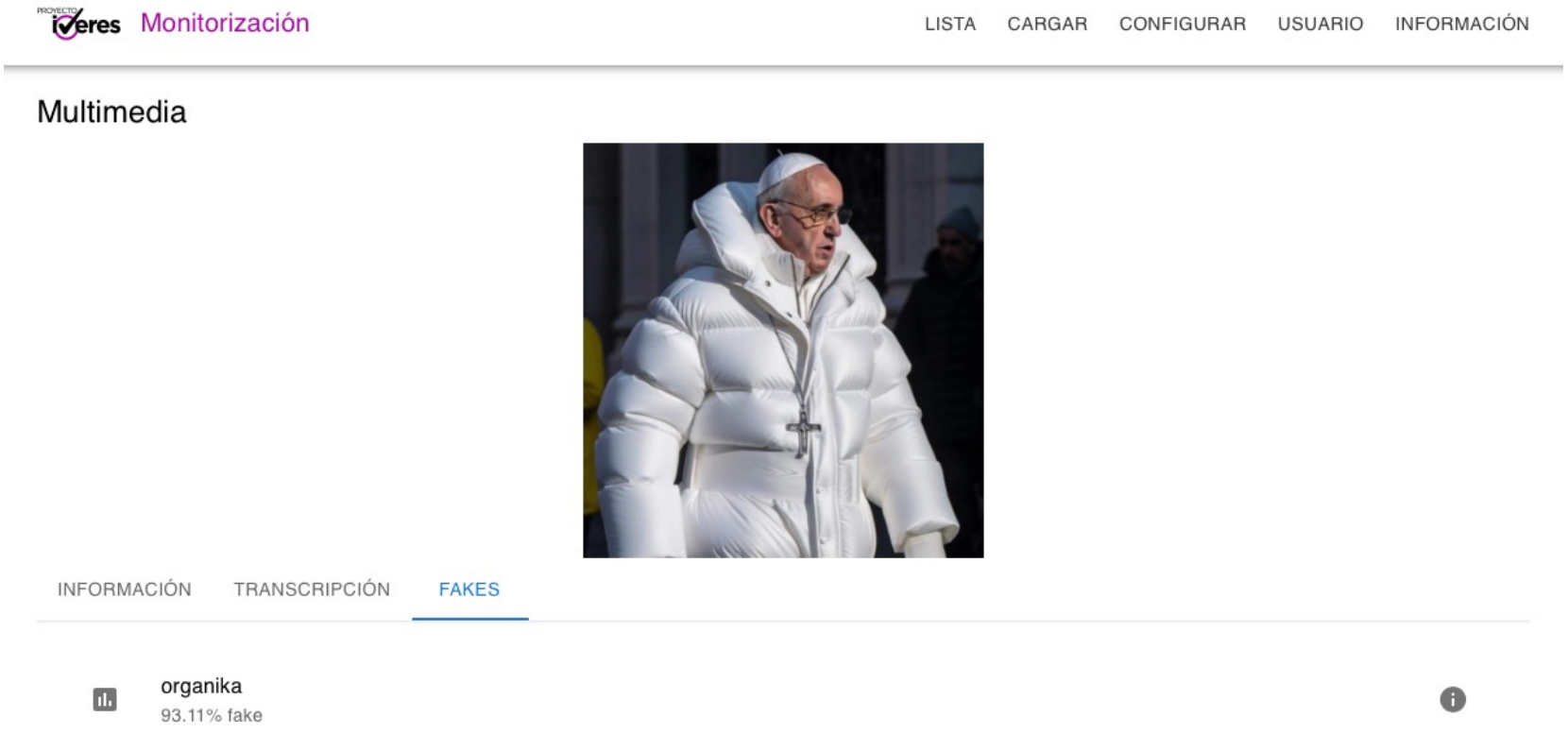


SDXL-Detector

- Puntos fuertes
  - Funciona bien tanto para imágenes artísticas como realistas / fotografías
  - Funciona muy bien para imágenes generadas con Stable Diffusion XL
  - Funciona bien para otros modelos recientes
- Puntos débiles
  - No funciona bien para imágenes generadas con modelos anteriores como GANs
  - No funciona bien para imágenes editadas en regiones pequeñas (inpaint, outpaint)

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Text-Generated Image Detection



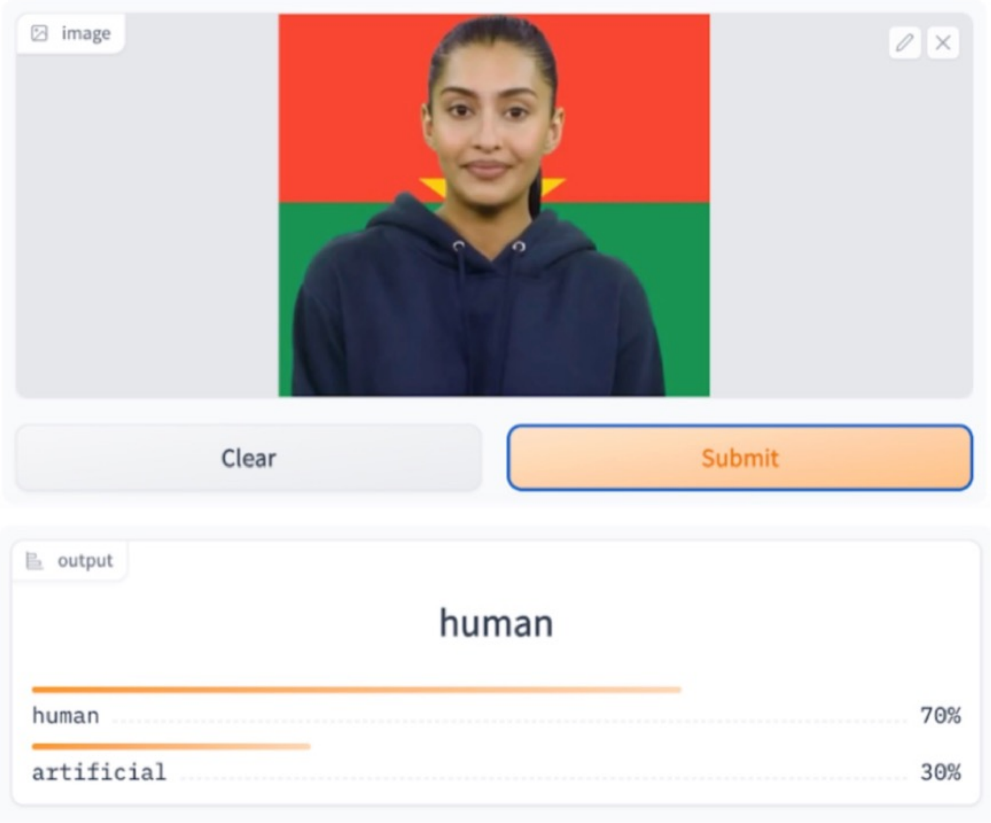
PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Text-Generated Image Detection

An AI detection tool analyses a famous clip of President Obama's deepfake. Whenthe resolution is lowered and the end is edited out, the tool judges it as "not a deepfake".

PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Text-Generated Image Detection



A screenshot with an AI avatar from one of the clips was detected as 70% human.

PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Text-Generated Image Detection

Los móviles actuales realizan mucho procesado de la imagen: Portrait, smart HDR, diffusion, night mode, etc.

Muchas imágenes producidas por móviles son interpretadas como fakes

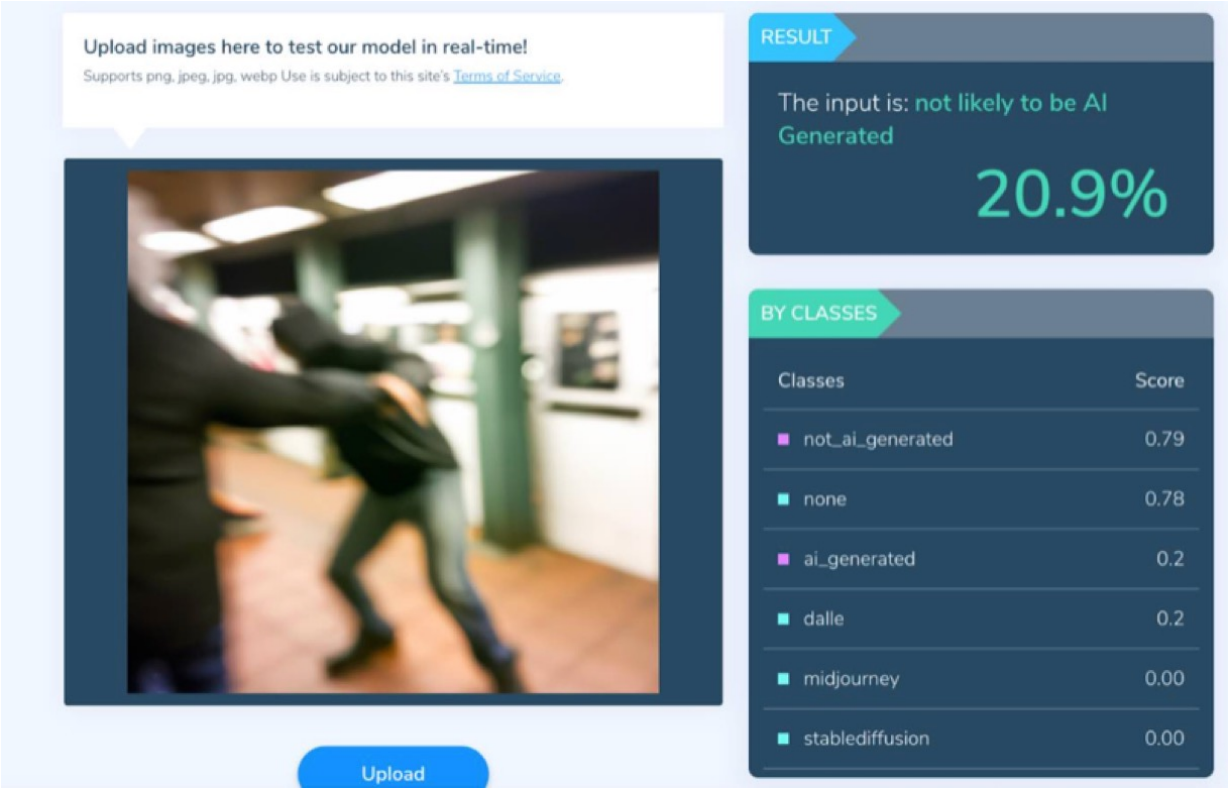
Los entrenamientos para la detección de imágenes generadas a partir de texto se obtienen con un set de ‘prompts’.

Si se modifican los ‘prompts’ el sistema de detección puede fallar.

- No existen todavía estrategias para trabajar con imágenes in-painting o out-painting

PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

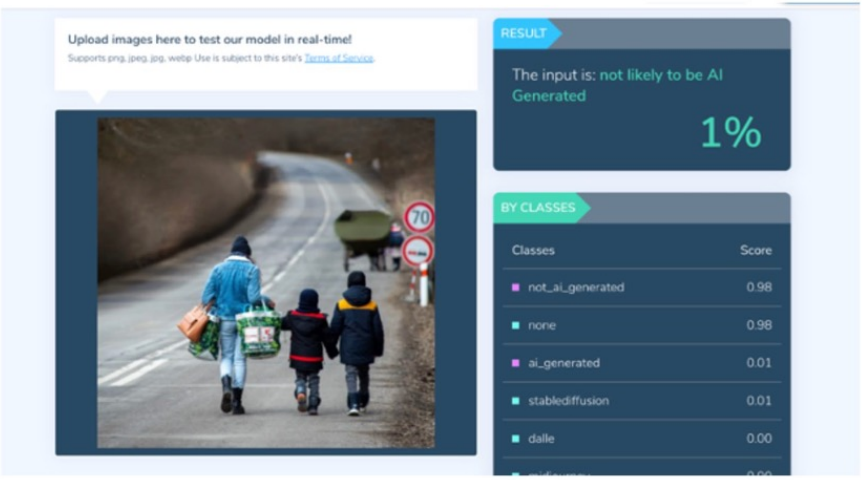
Text-Generated Image Detection



PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Text-Generated Image Detection

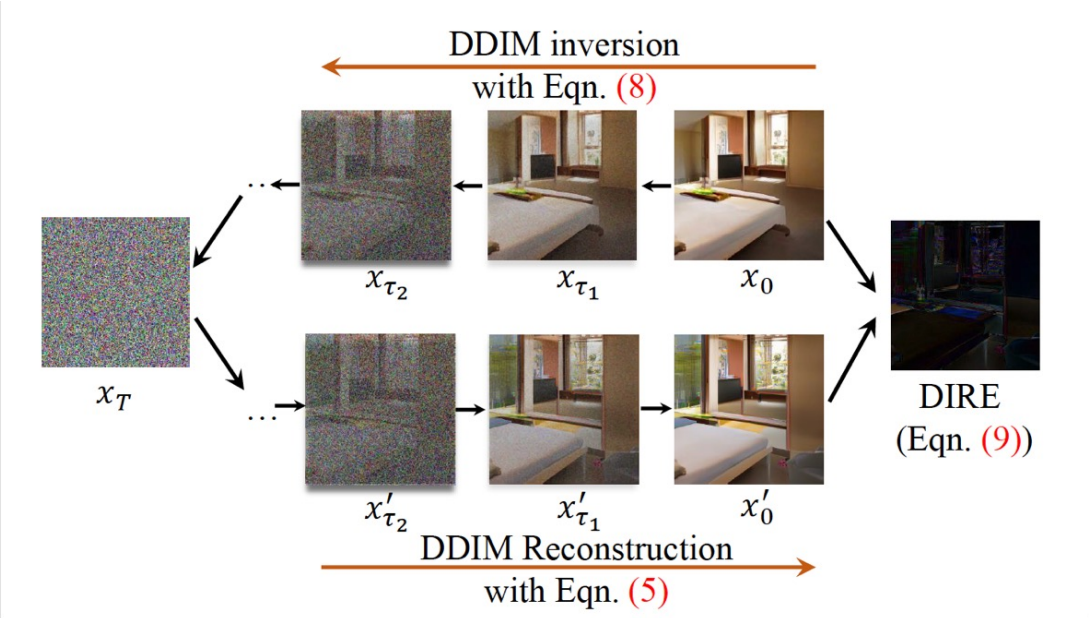
In-painting example using a real image showing Ukrainian refugees, adding a tank with DALL-E 2. Original photograph by Peter Lazar/AFP/Getty Images.



The image with the in-painted tank was detected as "likely not to be AI-generated".

PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Text-Generated Image Detection

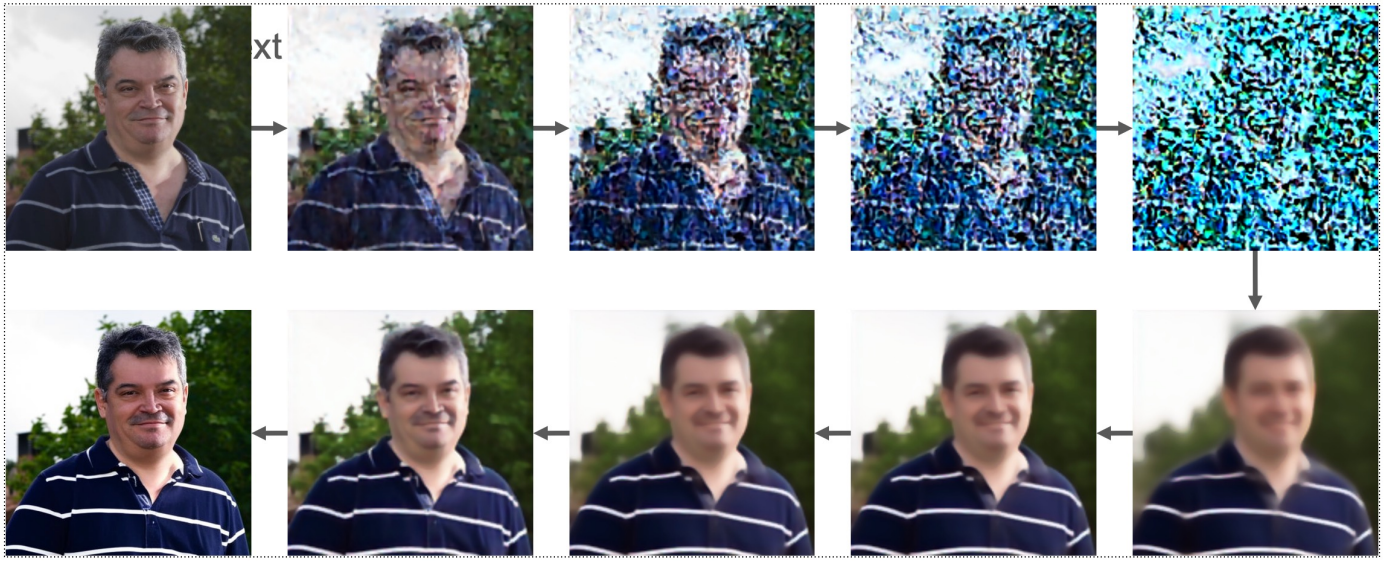


[1] Z. Wang *et al.*, "DIRE for Diffusion-Generated Image Detection," https://arxiv.org/abs/2303.09295 2023.

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Text-Generated Image Detection

Test inicial de reconstrucción



PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Text-Generated Image Detection

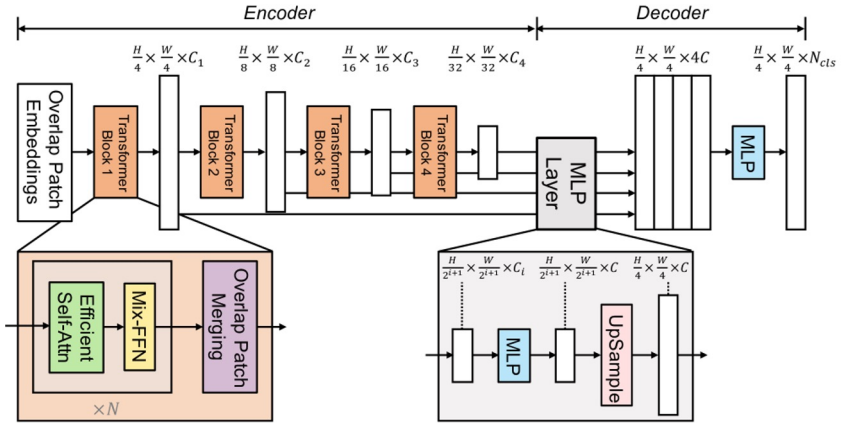
Reconstrucción imagen fake

a woman with a large hoop earrings and a white tank top is looking at the camera and has her eyes closed, full round face, photorealism, Emma Ríos, a portrait

PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Modelo segmentación

- En desarrollo
- Arquitectura
  - Segformer
  - Transformer jerárquico
  - Optimizado para segmentación semántica
- Datos de entrenamiento
  - Preentrenado con 1.28M imágenes (ImageNet)
  - Finetunning
    - Imágenes reales extraídas de ImageNet
    - Generación de máscaras aleatorias
    - Imágenes fake generadas a partir de las máscaras con los modelos Stable Diffusion 1.5 y XL



PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Máscaras aleatorias

- Transiciones graduales para mejor integración de regiones real - fake
- Formas aleatorias basadas en ruido para evitar que los modelos memoricen patrones
- Escalas desde pequeñas hasta imágen completa para detectar también imágenes totalmente generadas

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Generaciones

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Generaciones

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Generaciones

PROYECTO iveres

UPC

rtve

Plan de Recuperación, Transformación y Resiliencia

Financiado por la Unión Europea NextGenerationEU

Ejemplos test

Generación

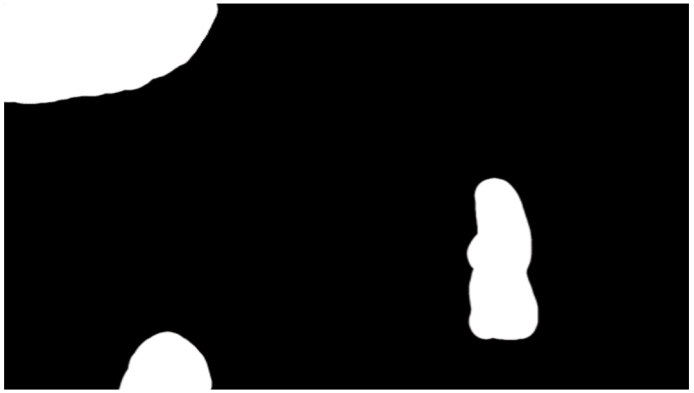
PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Ejemplos test

Segmentación:

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Ejemplos test



Generación

PROYECTO iveres

UPC

rtve

Plan de Recuperación, Transformación y Resiliencia

Financiado por la Unión Europea NextGenerationEU

Ejemplos test

Segmentación:

- Comete errores
- Falsos positivos
- Trabajando en esto

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Open AI SORA

SORA: creating video from text (15 Feb 2024)

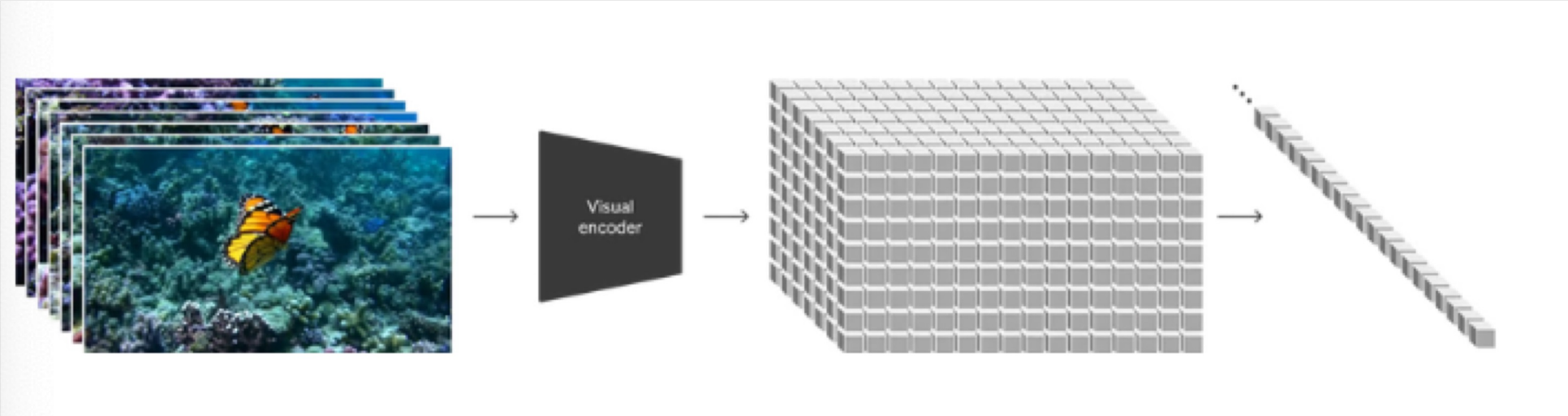
Generation models as world simulators

Technology:

- Text-conditional diffusion models on videos and images of variable durations, resolutions and aspect ratios
- Transformer architecture that operates on spacetime patches of video and image latent codes.
- Generation of 1 minute of high fidelity video.
- Video generators may be viewed as general purpose simulators of the physical world

Open AI SORA

Using visual patches



Video is compressed into a latent space with lower dimensionality and decomposed into time-space patches

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Open AI SORA

Prompt: A stylish woman walks down a Tokyo street filled with warm glowing neon and animated city signage. She wears a black leather jacket, a long red dress, and black boots, and carries a black purse. She wears sunglasses and red lipstick. She walks confidently and casually. The street is damp and reflective, creating a mirror effect of the colorful lights. Many pedestrians walk about.

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Open AI SORA

Variable durations, resolutions, aspect ratios

The data is used in training at the original resolution (not converted to 4s, 256x256)

Sora can work up to 1920x1080 resolution in horizontal o vertical and everything in between.

It can be used to create content for different devices.

Prototype contents at lower resolutions before applying full resolution

Training on original data improves composition and framing

Open AI SORA

Language understanding

Training requieres a large amount of videos with corresponding text captions.

The recaptioning technique applied in Dall-e 3 is introduced to videos.

A highly descriptive captioner model is trained to produce text captions.

This model is used to generate descriptions for all the videos in the dataset

Highly descriptive prompts improves the quality of the generated image.

GPT turns short user prompts into longer detailed captions

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Open AI SORA

Prompting with images and videos

Sora can be used with image or video prompts (not text).

Image generated with Dall-e 3

Video Generated from Dall-e 3

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Open AI SORA

Prompting with images and videos

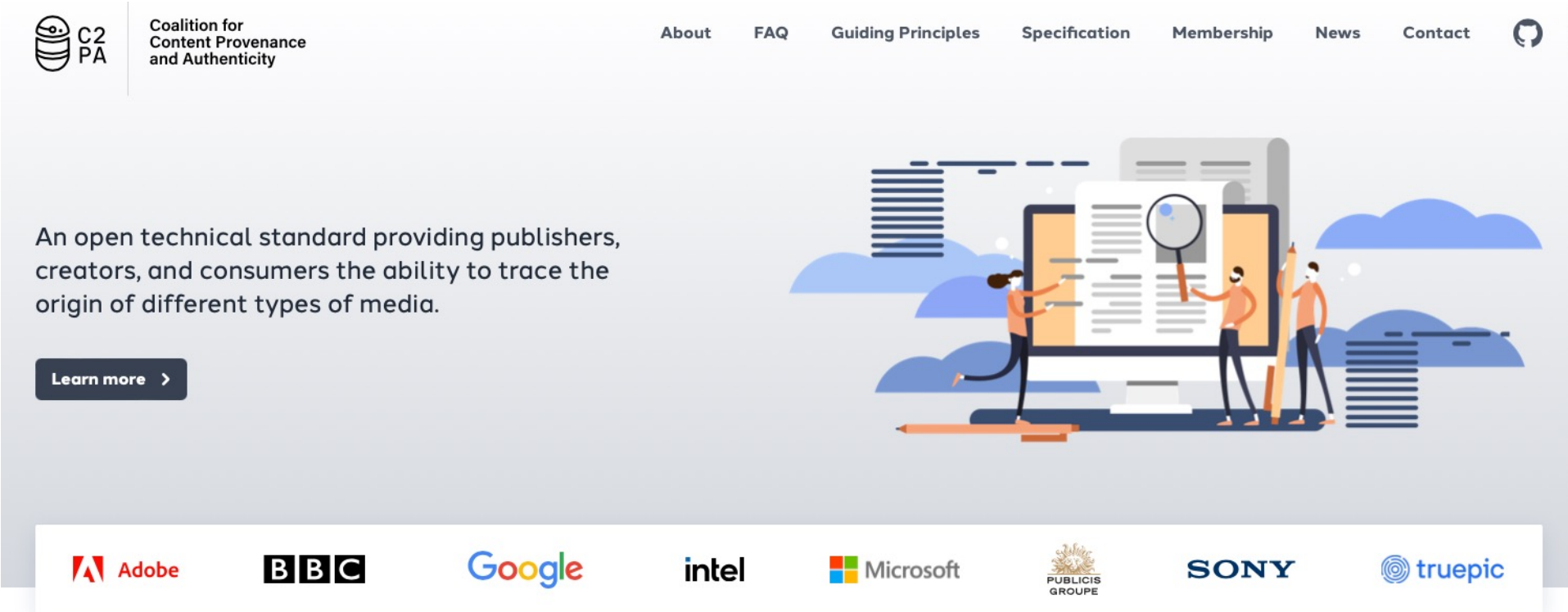
Sora can be used with image or video prompts (not text).

Image generated with Dall-e 3

Video Generated from Dall-e 3 input

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Coalition for Content Provenance and Authenticity (C2PA)



PROYECTO iveres

UPC

rtve

Plan de Recuperación, Transformación y Resiliencia

Financiado por la Unión Europea NextGenerationEU

Coalition for Content Provenance and Authenticity (C2PA)

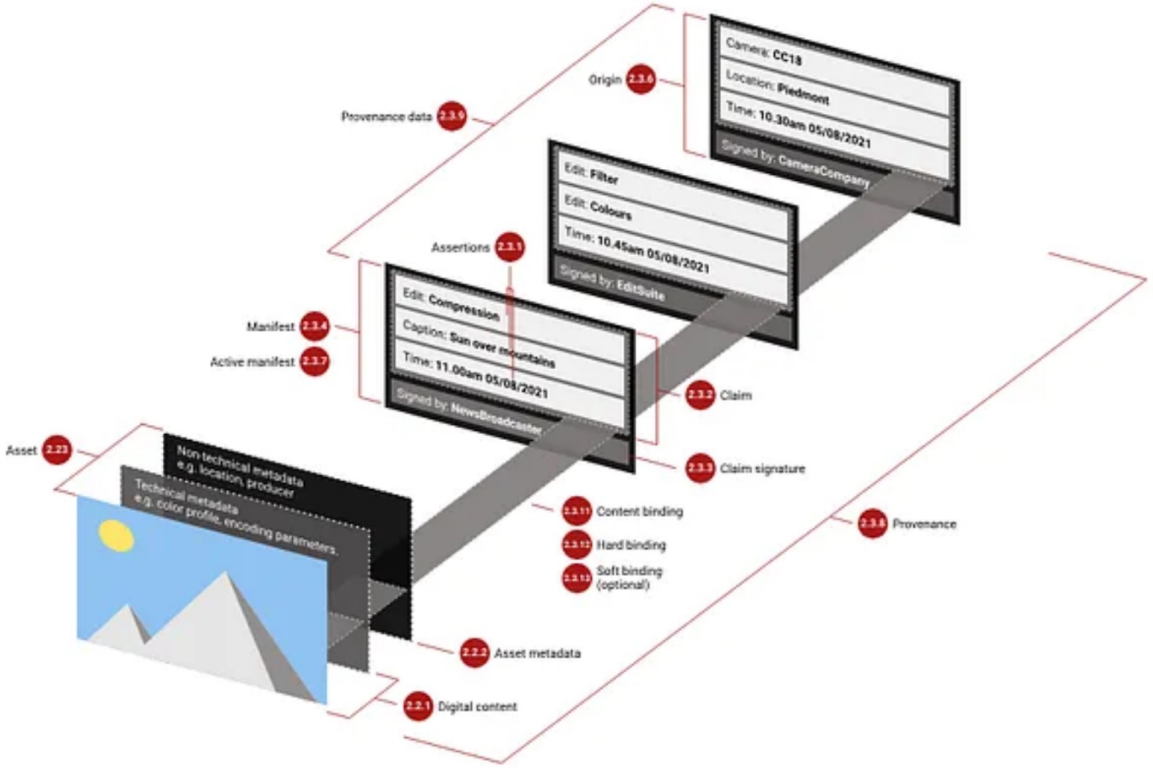
General Members

PROYECTO iveres

Financiado por la Unión Europea NextGenerationEU

Plan de Recuperación, Transformación y Resiliencia

Coalition for Content Provenance and Authenticity (C2PA)



PROYECTO iVeres

UPC

rtve

Plan de Recuperación, Transformación y Resiliencia

Financiado por la Unión Europea NextGenerationEU

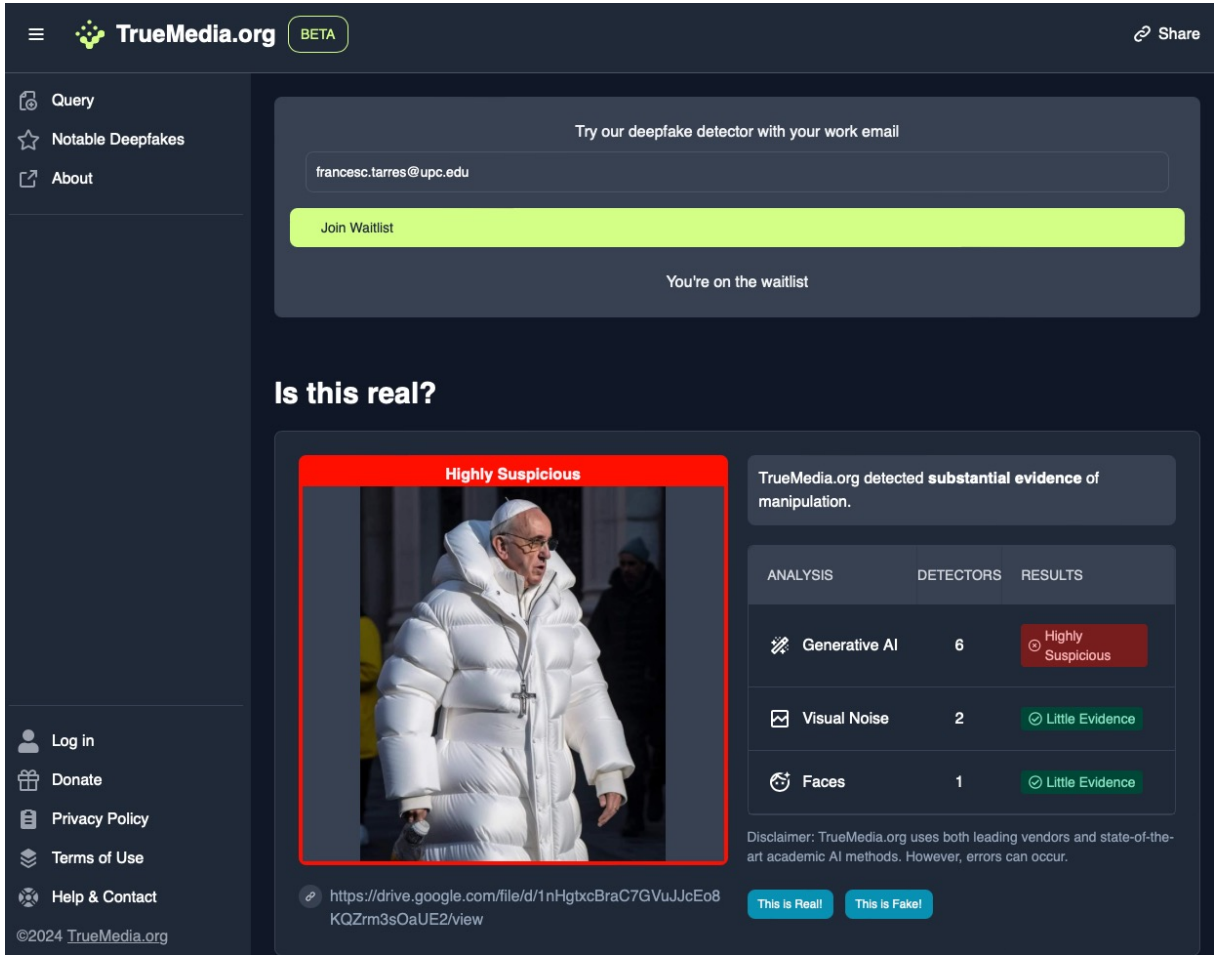
Coalition for Content Provenance and Authenticity (C2PA)

C2PA Facts

- Microsoft / OpenAI / Google / Sony / Adobe / Nikon are very favourable to the initiative.
- Dall-e 3 is including metadata in the C2PA format to AI generated images since Feb 14, 2024
- SORA is on standby for quality improvement and for fake detection.
- The first version of the C2PA API is already available for partners in the consortium since Feb, 2024

Otros generadores de vídeo

- Gen3Alpha	https://runwayml.com/ai-tools/gen-3-alpha/
- Dream Machine	https://lumalabs.ai/dream-machine
- Kling Video Model	https://klingai.co
- Pika	https://pika.art/login



https://detect.truemedia.org/media/analysis?id=jp3AvXVwXkbtmKYjxwLkpqMIS38.jpg

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

DeepFake-o-Meter

DeepFake-O-Meter
MDFL at UB

HOME ACCOUNT MY SUBMISSION CONTACT LOG OUT

The likelihood score derived from the method indicates the probability of the input being fake. This score reflects the extent to which the input shares characteristics with the real or fake data labeled in the training set of the method, as reported in the original study, supported by correlation statistics. However, it is important to note that this score should not be considered as providing deterministic result.

Detector	Result	Details
Univ. Buffalo - video detector (2024)	99.6% AI-Generated Likelihood	Completed

Upload Another File | Check Submission History

https://humanaigc.github.io/emote-portrit-alive/

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

DeepFake-O-Meter
MDFL at UB

HOME ACCOUNT MY SUBMISSION CONTACT LOG OUT

DEEPFAKE-O-METER

An Open Platform Integrating State-Of-The-Art Algorithms for DeepFake Image, Video, and Audio Detection

3.3 MB
DEEP FAKE MACRON.mp4

Detector Selection

- UNIV. BUFFALO - VIDEO DETECTOR (2022)
- UNIV. BUFFALO - VIDEO DETECTOR (2024)
- THE UNIVERSITY OF TOKYO - VIDEO DETECTOR (2022)
- UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA - VIDEO DETECTOR (2023)
- UA - VIDEO DETECTOR (2019)
- XIAMEN UNIVERSITY - VIDEO DETECTOR (2021)
- CUHK - VIDEO DETECTOR (2024)

Do you think it's real or fake?
I don't know Real Fake

Enter any additional information here, e.g., data source.

SUBMIT

Do you want to share data with us?
Yes No

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

DeepFake-O-Meter
MDFL at UB

HOME ACCOUNT MY SUBMISSION CONTACT LOG OUT

DEEPFAKE-O-METER

An Open Platform Integrating State-Of-The-Art Algorithms for DeepFake Image, Video, and Audio Detection

Detector Selection

- UNIV. FEDERICO II OF NAPLES - IMAGE DETECTOR (2021)
- UNIV. BUFFALO - IMAGE DETECTOR (2024)
- UNIV. FEDERICO II OF NAPLES - IMAGE DETECTOR (2023)
- MICHIGAN STATE UNIVERSITY - IMAGE DETECTOR (2023)
- UNIV. WISCONSIN-MADISON - IMAGE DETECTOR (2023)
- UA - IMAGE DETECTOR (2019)
- BEIJING JIAOTONG UNIVERSITY - IMAGE DETECTOR (2024)

Do you think it's real or fake?

I don't know Real Fake

Enter any additional information here, e.g., data source.

SUBMIT

Do you want to share data with us?

Yes No

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

DeepFake-O-Meter
MDFL at UB

HOME ACCOUNT MY SUBMISSION CONTACT LOG OUT

The likelihood score derived from the method indicates the probability of the input being fake. This score reflects the extent to which the input shares characteristics with the real or fake data labeled in the training set of the method, as reported in the original study, supported by correlation statistics. However, it is important to note that this score should not be considered as providing deterministic result.

Detector	Result	Details
Univ. Buffalo - image detector (2024)	9.7% AI-Generated Likelihood	Completed

Upload Another File | Check Submission History

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Spaces | umm-maybe / AI-image-detector | like 105 | Running | App | Files | Community 5

Maybe's AI Art Detector

This app is a proof-of-concept demonstration of using a ViT model to predict whether an artistic image was generated using AI.

It was created in October 2022, and as such, the training data did not include any samples generated by Midjourney 5, SDXL, or DALLE-3. It still may be able to correctly identify samples from these more recent models due to being trained on outputs of their predecessors.

Furthermore the intended scope of this tool is artistic images; that is to say, it is not a deepfake photo detector, and general computer imagery (webcams, screenshots, etc.) may throw it off.

In general, this tool can only serve as one of many potential indicators that an image was AI-generated. Images scoring as very probably artificial (e.g. 90% or higher) could be referred to a human expert for further investigation, if needed.

For more information please see the blog post describing this project at: https://medium.com/@matthewmaybe/can-an-ai-learn-to-identify-ai-art-545d9d6af226

image

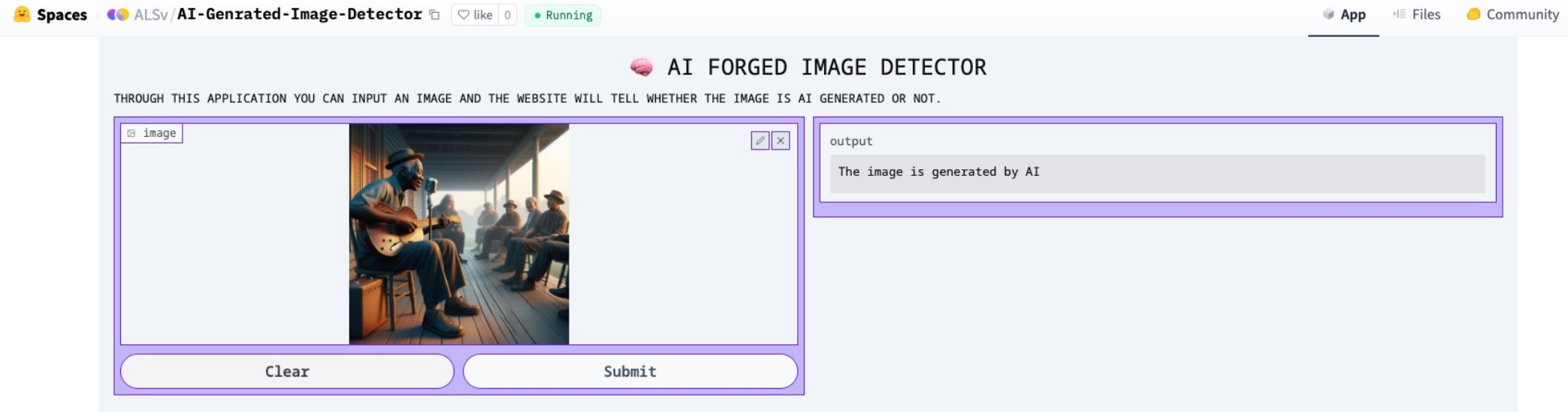
Clear | Submit

output

human

human	91%
artificial	9%

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU



PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Deepfakes en 2024

2.298

TOOLS FOR AI FACE SWAP, LIP SYNC, FACE REENACTMENT, AI-AVATARS

10.206

TOOLS FOR AI IMAGE GENERATION

1.018

TOOLS FOR AI VOICE GENERATION, VOICE CLONING

47

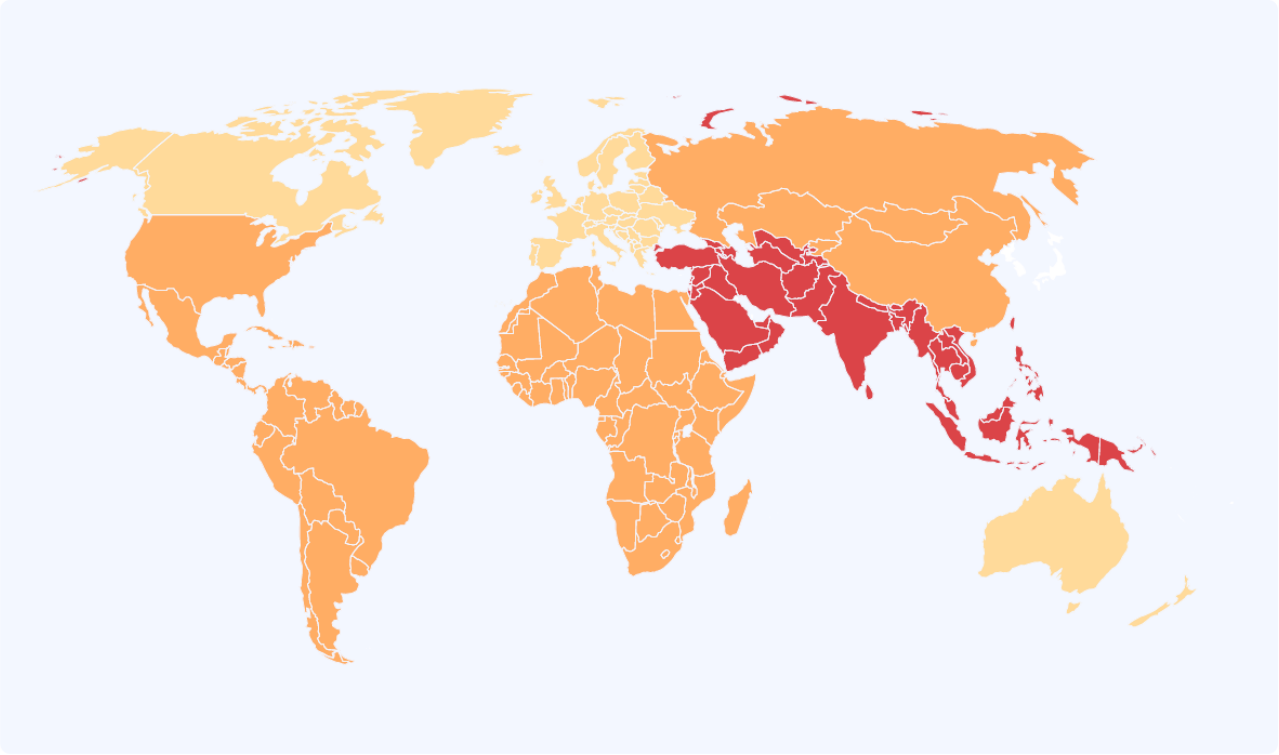
TOOLS FOR DEEPFAKE KYC INJECTION

Source: The state of Deepfakes 2024. SensityAI report

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Deepfakes en 2024

Influence Campaign



PROYECTO iVeres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Deepfakes en 2024

Influence Campaign

PROYECTO iVeres

UPC

rtve

Plan de Recuperación, Transformación y Resiliencia

Financiado por la Unión Europea NextGenerationEU

Deepfakes en 2024

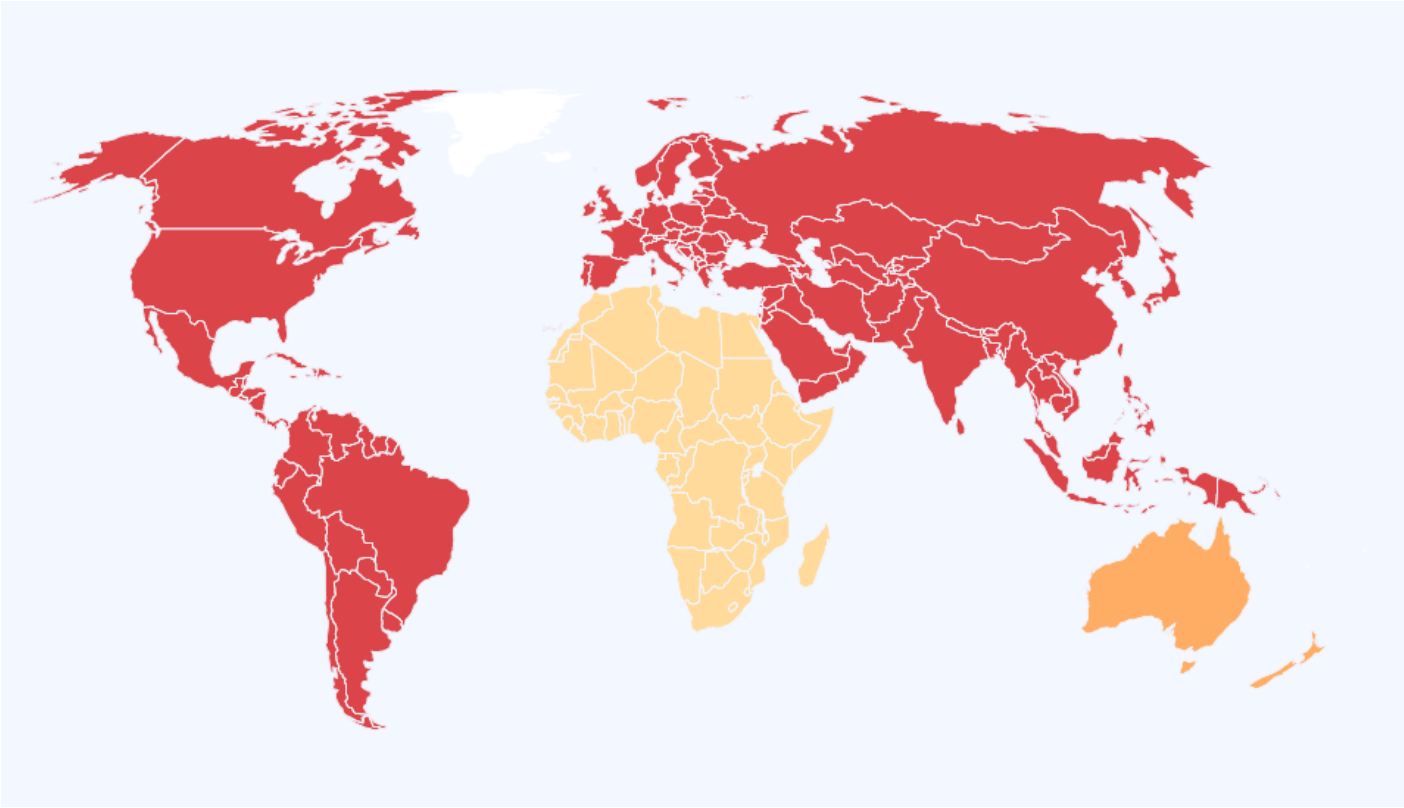
Influence Campaign (Ron deSantis)

https://www.dailymotion.com/playlist/x8e4t4

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Deepfakes en 2024

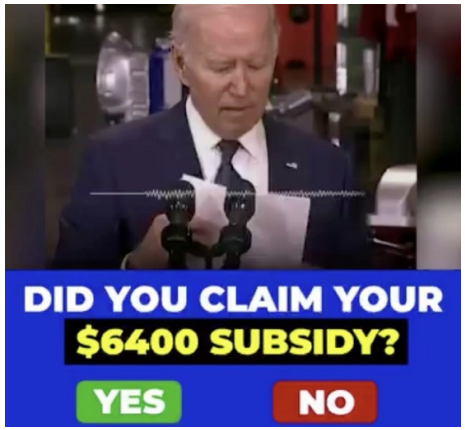
Deep Fake Scams



PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Deepfakes en 2024

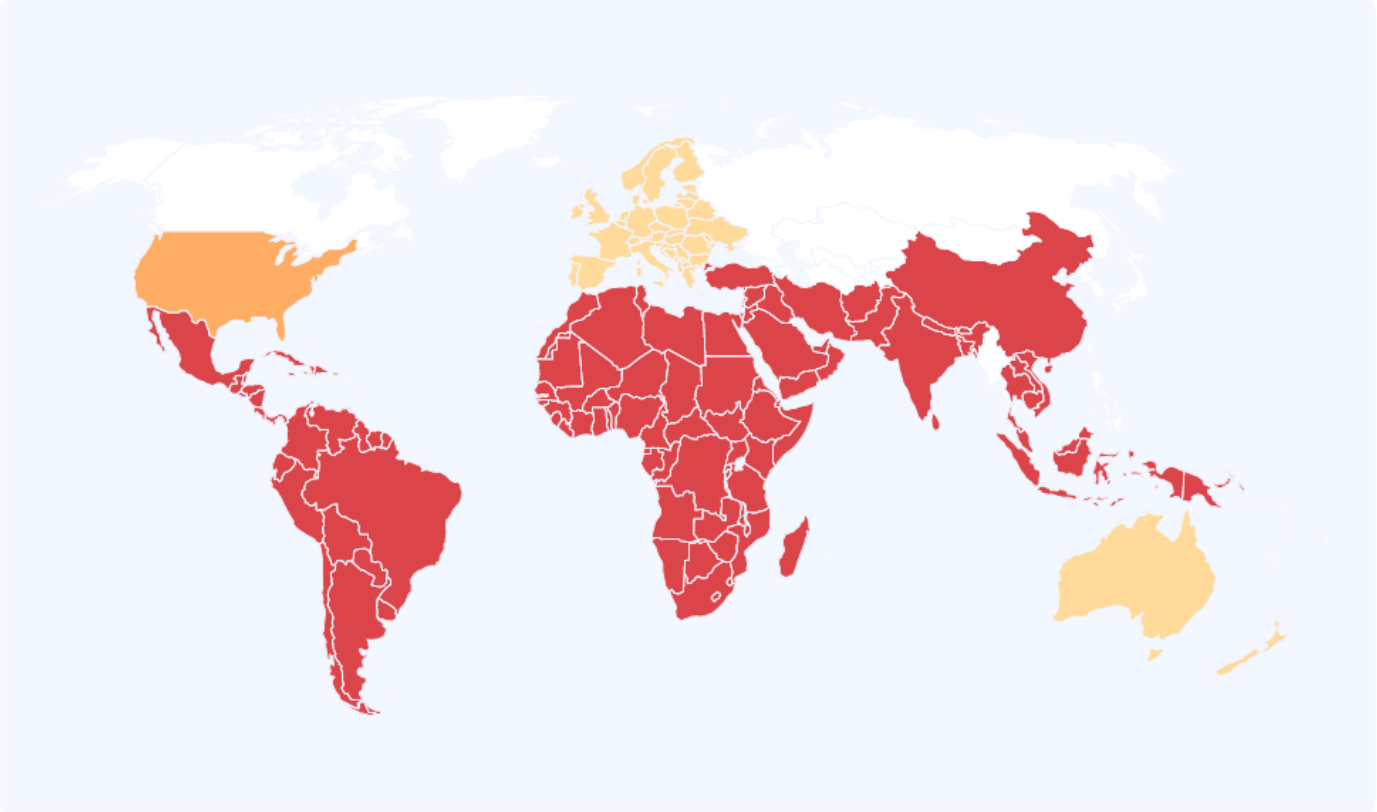
Deep Fake Scams



PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea
NextGenerationEU

Deepfakes en 2024

KYC Frauds

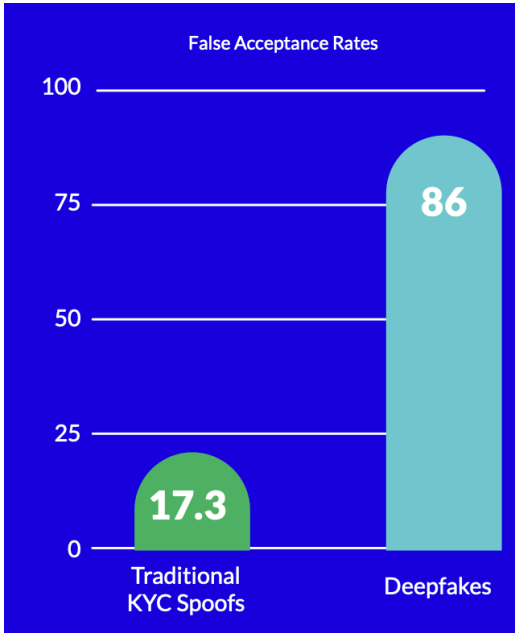


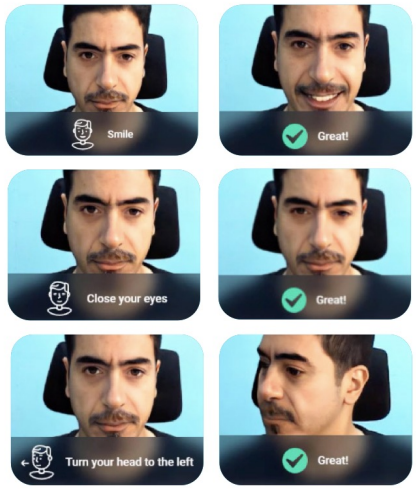
PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Deepfakes en 2024

KYC Frauds

Los sistemas de reconocimiento facial más conocidos son vulnerables a los deepfake (95% False Acceptance Rate).
Ex: Google, Azure, AWS face recognition FAR 80% - 99%





Active Liveness

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU

Conclusiones y Evolución Futura.

- **Los algoritmos de generación de fakes van por delante de los algoritmos de detección si hay cambios sustanciales en el algoritmo.**

- **Los algoritmos de generación evolucionan y mejoran de modo constante generando imágenes de mayor calidad.**

- **La generación de fakes combina no solo técnicas de imagen/video sino también de generación de voz y de texto.**

- **La aparición de herramientas como: Whisper, ChatGPT3/4, Generación de avatares, etc.; está revolucionando el proceso de generación y valoración de noticias.**

- **El potencial de estas herramientas aumentará considerablemente en los próximos años, a medida que se vaya incorporando la corrección humana al proceso de generación. Por ejemplo, ChatGPT4 está disponible en bing/word, puede redactar automáticamente el borrador de un artículo, que posteriormente será corregido por el autor … realimentando la corrección para realizar nuevos entrenamientos.**

- **Estamos experimentando un 'boom' de la inteligencia artificial usando métodos estadísticos, con aprendizaje supervisado y modelos de neuronas que datan de 1950. En paralelo se están desarrollando algoritmos conceptualmente diferentes, con modelos y arquitecturas neuronales más avanzados, con autoaprendizaje, sin necesidad de supervisión que posiblemente evolucionaran de forma considerable, incluso poco previsible, todas estas herramientas.**

PROYECTO iveres
UPC
rtve
Plan de Recuperación, Transformación y Resiliencia
Financiado por la Unión Europea NextGenerationEU