Detección imágenes o vídeos modificados mediante Redes Neuronales.

# Indicación del nivel de seguimiento de la planificación prevista para el desarrollo del TFG, y de los ajustes efectuados, junto con su justificación en este último caso.

En la actualidad, nos encontramos en una fase avanzada del proyecto, en la cual hemos logrado exitosamente el objetivo de desarrollar un modelo de inteligencia artificial basado en redes neuronales convolucionales. Este modelo ha demostrado obtener resultados razonables y de calidad.

En relación con la planificación inicial del proyecto, nos aproximamos al final de la etapa de implementación, que consistía en llevar a cabo la implementación del proyecto mediante la búsqueda e identificación de variables e indicadores a medir, realizar un análisis preliminar de los resultados y corregir las desviaciones surgidas durante la implementación.

Es importante destacar que en esta fase hemos enfrentado desafíos y problemas principales relacionados con la desviación del objetivo principal del proyecto, que era desarrollar un detector de imágenes modificadas. Nos hemos enfocado en lograr un resultado final que no solo clasifique la imagen en cuestión, sino también identifique las áreas modificadas de la imagen. Sin embargo, al centrarnos en esta tarea, nos hemos desviado del objetivo inicial. Para abordar este desafío, hemos recurrido a la búsqueda y creación de nuevos conjuntos de datos (datasets) con el fin de desarrollar modelos adicionales que nos permitan detectar las zonas modificadas en la imagen.

Por otro lado, nos habíamos propuesto como meta adicional la identificación del tipo de modificación realizada en las imágenes alteradas. Esta meta ha requerido una considerable inversión de tiempo, ya que, si bien agrega valor a nuestro trabajo, se aleja de la idea principal del proyecto.

En esta etapa, nos hemos vuelto a enfocar en el diseño y desarrollo del modelo con el objetivo de buscar mejoras. Por ejemplo, hemos realizado la búsqueda de hiperparámetros utilizando la biblioteca Optuna y hemos explorado la carga de datos en lotes, además de probar nuevos parámetros y construir un modelo con una arquitectura más compleja.

Finalmente, con el propósito de aportar un mayor valor al proyecto, hemos investigado y desarrollado técnicas de detección de alteraciones en imágenes, incluyendo la distribución de píxeles, el histograma de gradientes orientados, el gradiente local, la detección de sombras, el análisis de color y la inconsistencia de iluminación.

En conclusión, podemos afirmar que estamos cumpliendo con la planificación inicial y que el TFG se está desarrollando de manera adecuada.

# Explicación general de la metodología que se ha seguido finalmente y de los cambios respecto a la propuesta inicial.

En la metodología inicial, se planteó la utilización de un tablero de kanban para marcar y hacer seguimiento de las tareas propuestas para el TFG. Sin embargo, durante el desarrollo del proyecto, se ha optado por seguir la metodología del diagrama de Gantt, que ha permitido establecer un cronograma y cumplir con las tareas dentro de los plazos establecidos.

En relación con la propuesta inicial, no se han realizado cambios sustanciales, ya que el objetivo principal del proyecto sigue siendo el mismo: desarrollar un sistema para la identificación de imágenes modificadas o alteradas. No obstante, se han identificado retos importantes en esta etapa, especialmente en lo que respecta a la generalización del modelo actual, con el fin de lograr mejoras significativas para la presentación final.

Para abordar estos desafíos, se ha seguido una metodología que se basa en la investigación exhaustiva de técnicas y algoritmos relacionados con la detección de imágenes modificadas. Se han empleado redes neuronales convolucionales como base del modelo, aprovechando su capacidad para extraer características relevantes de las imágenes. Además, se ha trabajado en la recopilación y preparación de conjuntos de datos adecuados, incluyendo imágenes con diferentes tipos de modificaciones y alteraciones.

En cuanto a los cambios respecto a la propuesta inicial, se ha dedicado más tiempo y recursos al ajuste y refinamiento del modelo, con el fin de mejorar su capacidad de generalización y obtener resultados más precisos. Se han realizado experimentos con diferentes arquitecturas de redes neuronales, así como la exploración de técnicas de aumento de datos para ampliar la diversidad del conjunto de entrenamiento.

Además, se ha incorporado un enfoque en la interpretación de los resultados obtenidos, mediante el análisis de las áreas destacadas por el modelo como posibles modificaciones. Esto permite comprender mejor el funcionamiento del sistema y brindar explicaciones claras y justificadas sobre las decisiones tomadas por el modelo.

En resumen, la metodología seguida finalmente se ha basado en el diagrama de Gantt para el seguimiento y cumplimiento de tareas, y se han realizado ajustes y refinamientos en el modelo con el objetivo de mejorar la capacidad de generalización y obtener resultados más precisos. A su vez, se ha puesto énfasis en la interpretación de los resultados obtenidos para brindar una explicación clara y justificada del sistema desarrollado.

# Exposición de los resultados

En esta etapa del proyecto, nos hemos centrado en la búsqueda de hiperparámetros que mejoren el modelo actual, con el objetivo de aumentar su precisión y lograr una mejor generalización de este. Para alcanzar estos resultados, hemos llevado a cabo experimentos exhaustivos y detallados.

En primer lugar, hemos utilizado técnicas de búsqueda de hiperparámetros, como la optimización bayesiana utilizando la biblioteca Optuna. Esta metodología nos ha permitido explorar y encontrar configuraciones óptimas de parámetros para nuestro modelo de inteligencia artificial basado en redes neuronales convolucionales.

A través de la optimización bayesiana, hemos realizado múltiples iteraciones, evaluando diferentes combinaciones de hiperparámetros, como la tasa de aprendizaje, el tamaño del lote (batch size), la cantidad de capas y filtros convolucionales, entre otros. El objetivo ha sido encontrar la configuración que maximice la precisión del modelo y minimice el riesgo de sobreajuste (overfitting) o subajuste (underfitting).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **N\_iteration** | **filters** | **dropout** | **learning\_rate** | **accuracy** |
| 14 | 32 | 0,337851872 | 0,000436237 | 0,931249976 |
| 17 | 32 | 0,335981099 | 0,000101042 | 0,930000007 |
| 6 | 32 | 0,32737358 | 0,00011386 | 0,92750001 |
| 18 | 32 | 0,41901214 | 0,000237771 | 0,926249981 |
| 13 | 32 | 0,352835171 | 0,000472356 | 0,926249981 |
| 12 | 16 | 0,36716443 | 0,000109433 | 0,922500014 |
| 2 | 16 | 0,470238655 | 0,000163346 | 0,91624999 |
| 4 | 16 | 0,412211671 | 0,000226372 | 0,915000021 |
| 16 | 32 | 0,285562407 | 0,000587016 | 0,911249995 |
| 15 | 32 | 0,264748223 | 0,000445889 | 0,911249995 |
| 19 | 32 | 0,345584422 | 0,00074962 | 0,908749998 |
| 9 | 16 | 0,22617134 | 0,001233481 | 0,908749998 |
| 10 | 32 | 0,229985454 | 0,000115427 | 0,904999971 |
| 11 | 32 | 0,383809772 | 0,000323638 | 0,903750002 |
| 7 | 16 | 0,477499231 | 0,011103526 | 0,507499993 |
| 3 | 64 | 0,4112627 | 0,040741447 | 0,507499993 |
| 1 | 64 | 0,485465104 | 0,001413594 | 0,492500007 |
| 8 | 64 | 0,124090189 | 0,010006914 | 0,492500007 |
| 5 | 64 | 0,308739329 | 0,001753823 | 0,492500007 |

Los resultados del modelo son los siguiente con respecto al rendimiento:

|  |  |
| --- | --- |
|  |  |
|  |  |

La estructura del modelo actual es la siguiente:

Diagrama, Tabla

Descripción generada automáticamente Diagrama, Tabla

Descripción generada automáticamente Diagrama, Tabla

Descripción generada automáticamente

A continuación, presentaremos los resultados obtenidos en nuestro proyecto, primero, mostraremos las predicciones realizadas por el modelo en las imágenes de prueba. Luego, discutiremos los resultados de otras técnicas utilizadas, como la distribución de píxeles, el gradiente local y la detección de sombras.

**Imágenes modificadas**

|  |  |
| --- | --- |
| **Imagen Original**  Una roca en el agua  Descripción generada automáticamente | **Imagen Modificada**  Una roca en el agua  Descripción generada automáticamente |
| **Mascara de la zona modificada**  Imagen en blanco y negro  Descripción generada automáticamente con confianza media | **Imagen de extracción de zona**  Un barco en el agua cerca de un edificio  Descripción generada automáticamente |

**Resultado del modelo:**

Imagen de la pantalla de un video juego

Descripción generada automáticamente con confianza baja

|  |  |
| --- | --- |
| **Imagen Original**  Animal de cuatro patas  Descripción generada automáticamente con confianza media | **Imagen Modificada** |
| **Mascara de la zona modificada**  Dibujo en blanco y negro  Descripción generada automáticamente con confianza media | **Imagen de extracción de zona**  Cara de un animal marino  Descripción generada automáticamente con confianza baja |

**Resultado del modelo:**

Interfaz de usuario gráfica, Gráfico

Descripción generada automáticamente

**Matriz de confusión**

**Gráfico

Descripción generada automáticamente**

**ROC curve**

**Gráfico, Gráfico de líneas

Descripción generada automáticamente**

# Conclusiones provisionales

1. El objetivo principal del proyecto consiste en desarrollar un sistema para la identificación de imágenes modificadas o alteradas. Hasta ahora, hemos logrado avanzar en el desarrollo de un modelo de inteligencia artificial basado en redes neuronales convolucionales, el cual ha mostrado resultados razonables y prometedores.
2. Durante la implementación del proyecto, hemos seguido una planificación inicial que ha experimentado algunos ajustes a lo largo del camino. Estos ajustes han estado relacionados principalmente con la búsqueda de nuevas técnicas y la generación de nuevos conjuntos de datos para mejorar la detección de las zonas modificadas en las imágenes.
3. Se han enfrentado desafíos significativos en la consecución del objetivo principal del proyecto. En particular, la detección de las zonas modificadas ha requerido una mayor atención y enfoque, lo cual ha implicado una desviación momentánea del objetivo inicial. Sin embargo, estas desviaciones se consideran necesarias para lograr un resultado final más completo y preciso.
4. Además del objetivo principal, también nos hemos planteado metas adicionales, como la identificación del tipo de modificación en las imágenes alteradas. Si bien estas metas añaden valor al proyecto, también han consumido un tiempo considerable y se han alejado ligeramente de la idea principal del TFG.
5. Durante el desarrollo del proyecto, hemos realizado diversas mejoras en el modelo, tales como la búsqueda de hiperparámetros óptimos utilizando técnicas como el estudio con la librería Optuna, la optimización de la carga de datos mediante lotes y la construcción de una arquitectura más compleja.

En conclusión, hasta el momento, se ha logrado un avance significativo en el desarrollo del TFG, con resultados razonables y ajustes realizados para mejorar la detección de imágenes modificadas. Sin embargo, se requiere seguir trabajando en la generalización del modelo y en la mejora de la precisión, así como en la evaluación exhaustiva de las técnicas empleadas, con el fin de alcanzar los objetivos establecidos y obtener un resultado final satisfactorio y de calidad para la presentación y defensa del TFG.

# Fuentes de información consultadas

1. "Digital Image Forensics” - Hany Farid.
2. Exposing Photo Manipulation From User-Guided 3-D Lighting Analysis Tiago Carvalhoa , Hany Faridb and Eric Keec aCemaden, S˜ao Jos´e dos Campos, SP, Brazil bDartmouth College, Hanover, NH, USA cColumbia University, New York, NY, USA.
3. Exposing Photo Manipulation with Inconsistent Reflections James F. O’Brien University of California, Berkeley and Hany Farid Dartmouth College.
4. Exposing Photo Manipulation with Inconsistent Shadows Eric Kee, Dartmouth College James O’Brien, University of California, Berkeley and Hany Farid, Dartmouth College.
5. How to Detect Faked Photos - A reprint from American Scientist the magazine of Sigma Xi, The Scientific Research Society.
6. Image Forgery Identification using Convolution Neural Network N. Hema Rajini.
7. PROTECTING PRESIDENT ZELENSKYY AGAINST DEEP FAKES A PREPRINT Matyáš Bohácek ,Hany Farid.
8. Detecting Deep-Fake Videos from Aural and Oral Dynamics Shruti Agarwal and Hany Farid University of California, Berkeley Berkeley, CA USA.
9. Detecting Deep-Fake Videos from Phoneme-Viseme Mismatches Shruti Agarwal and Hany Farid University of California, Berkeley Berkeley, CA USA.
10. Detecting Deep-Fake Videos from Appearance and Behavior Shruti Agarwal1 , Tarek El-Gaaly2 , Hany Farid1 , and Ser-Nam Lim2.
11. Exposing Digital Forgeries in Interlaced and De-Interlaced Video Weihong Wang, Student Member, IEEE, and Hany Farid, Member, IEEE.
12. Exposing Digital Forgeries in Video by Detecting Duplication Weihong Wang, Hany Farid.
13. Detecting Real-Time Deep-Fake Videos Using Active Illumination -Candice R. Gerstner Hany Farid.
14. Advances in Computer Vision and Pattern Recognition - Husrev Taha Sencar Luisa Verdoliva Nasir Memon.
15. A Coarse-to-fine Deep Convolutional Neural Network Framework for Frame Duplication Detection and Localization in Forged Videos Chengjiang Long Arslan Basharat Anthony Hoogs.
16. DEEPFAKE DETECTION: CURRENT CHALLENGES AND NEXT STEPS Siwei Lyu
17. Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics Yuezun Li1 , Xin Yang1 , Pu Sun2 , Honggang Qi2 and Siwei Lyu1 1 University at Albany, State University of New York, USA 2 University of Chinese Academy of Sciences, China.
18. DETECTION OF REAL-TIME DEEPFAKES IN VIDEO CONFERENCING WITH ACTIVE PROBING AND CORNEAL REFLECTION Hui Guo, Xin Wang, Siwei Lyu.
19. ELA (Error Level Analysis) - <https://fotoforensics.com/tutorial-ela.php>
20. Sobel Filter - <https://spec.oneapi.io/oneipl/0.5/filtering/sobel.html>
21. An efficient post-processing adaptive filtering technique to rectifying the flickering effects - <https://www.researchgate.net/publication/351462926_An_efficient_post-processing_adaptive_filtering_technique_to_rectifying_the_flickering_effects/figures?lo=1>

# Otra documentación acordada con la persona tutora, dependiendo del tipo de proyecto, para poder validar el correcto desarrollo del trabajo.

**Resultados de otras técnicas utilizadas**

Además de la predicción del modelo, hemos explorado y aplicado otras técnicas para detectar alteraciones en las imágenes. Estas técnicas incluyen la distribución de píxeles, el gradiente local y la detección de sombras.

La *distribución de píxeles* nos permite analizar los patrones de distribución y cambios en los valores de los píxeles de una imagen. Al detectar variaciones significativas en esta distribución, podemos identificar áreas que podrían haber sido modificadas.

Interfaz de usuario gráfica

Descripción generada automáticamente

Interfaz de usuario gráfica, Gráfico

Descripción generada automáticamente

La *inconsistencia de luz* se refiere a las discrepancias en la iluminación dentro de una imagen. Al analizar estas inconsistencias, podemos identificar áreas que han sido alteradas o manipuladas en términos de su iluminación.

Gráfico

Descripción generada automáticamente

Interfaz de usuario gráfica, Gráfico, Aplicación, PowerPoint

Descripción generada automáticamente

Interfaz de usuario gráfica, Aplicación

Descripción generada automáticamente

El *Histograma de Gradientes* Orientados (HOG) es una técnica que se utiliza para describir y detectar patrones locales en una imagen. Al calcular los gradientes de los píxeles en diferentes regiones de la imagen, podemos identificar características distintivas que pueden indicar modificaciones o alteraciones.

Gráfico

Descripción generada automáticamente con confianza media Interfaz de usuario gráfica

Descripción generada automáticamente con confianza baja

Los resultados obtenidos con estas técnicas adicionales han complementado las predicciones del modelo y han proporcionado información adicional sobre las áreas modificadas en las imágenes. Esta combinación de enfoques ha demostrado ser efectiva para identificar y analizar las alteraciones presentes en las imágenes evaluadas.