



**MÁSTER UNIVERSITARIO DE INVESTIGACIÓN EN
INGENIERÍA DE SOFTWARE Y SISTEMAS INFORMÁTICOS**

**31105151
TRABAJO FIN DE MÁSTER EN INGENIERÍA DE SOFTWARE Y
SISTEMAS INFORMÁTICOS**

**Detección jerárquica y reconocimiento de jeroglíficos egipcios
mediante visión artificial y redes neuronales profundas**

ALFONSO MORENO MARTÍNEZ
DIRECTOR: PEDRO JAVIER HERRERA CARO
Curso: 2025/2026
Convocatoria: Junio

**Máster Universitario de Investigación en Ingeniería
de Software y Sistemas Informáticos**

Código: 31105151 - Trabajo Fin de Máster en Ingeniería
de Software y Sistemas Informáticos

Título: Detección jerárquica y reconocimiento de jeroglíficos egipcios
mediante visión artificial y redes neuronales profundas

Línea de Trabajo: Sistemas de Percepción Visual. Tipo A

Estudiante: Alfonso Moreno Martínez

Director: Pedro Javier Herrera Caro

RESUMEN

La automatización en la detección y clasificación de jeroglíficos sobre estelas de alta densidad encuentra importantes obstáculos en factores como el desigual estado de conservación y degradación de la piedra, la similitud morfológica de los signos y la limitada disponibilidad de datos etiquetados a nivel global. El uso de redes neuronales profundas está lastrado por el prohibitivo coste temporal que supone la anotación manual de un conjunto masivo de datos de entrenamiento. Por ello, esta investigación descarta las arquitecturas monolíticas de extremo a extremo y propone, en su lugar, un sistema modular híbrido basado en una aproximación jerárquica.

El análisis epigráfico computacional opera en fases consecutivas, donde cada etapa condiciona el rendimiento de la siguiente. Este trabajo se centra en las dos primeras fases sobre estelas completas, donde primero ejecuta una localización gruesa para aislar los signos frecuentes de una categoría o conjunto de clases y, seguidamente, aplica una clasificación fina que asigna cada carácter a un código Gardiner.

La etapa inicial de localización compara el rendimiento de la red convolucional YOLO11 y el modelo de transformadores RF-DETR. Tomando como base el análisis de frecuencias del diccionario AED-TEI, se entrenan modelos específicos sobre recortes de alta resolución procedentes de distintas estelas. Uno se centra en las aves y el otro en los glifos de mayor cobertura del corpus. Tras afianzar estos localizadores expertos, un tercer modelo genérico pseudetiquetado con las detecciones de estos modelos aprovecha la atención global de RF-DETR para inferir, utilizando el contexto, la posición de los signos desconocidos. La hibridación de los modelos más eficientes de cada arquitectura permite procesar estelas completas mediante un sistema de ventanas deslizantes. En este flujo se incorpora un algoritmo global de supresión de no máximos condicionado por prioridades para filtrar las detecciones redundantes producidas por el solapamiento.

La posterior clasificación de grano fino se confía a modelos ConvNeXt, una arquitectura validada por su resistencia ante la erosión en relieves. Para gestionar la incertidumbre, esta fase aplica umbrales de confianza estrictos e implementa clases de rechazo, derivando los símbolos dudosos al modelo genérico para su revisión morfológica. Para resolver el déficit de muestras en las categorías minoritarias, el entrenamiento se refuerza generando datos sintéticos mediante redes cGAN condicionales y BicycleGAN, que preservan la fidelidad del trazo.

Esta integración de localización jerárquica con la clasificación experta proporciona una base técnica sólida para el procesamiento epigráfico digital. El sistema constituye una herramienta escalable y prepara el terreno para las futuras fases de transliteración y traducción automática de los textos.

Palabras clave: Procesamiento de imágenes, Redes neuronales, Aprendizaje máquina, Técnicas de inteligencia artificial, Visión por computador.

ABSTRACT

The automation of hieroglyph detection and classification on high-density stelae faces significant challenges, including the uneven state of stone preservation and degradation, the morphological similarity of the signs, and the limited availability of globally annotated data. The use of deep neural networks is hampered by the prohibitive time cost involved in the manual annotation of a large training dataset. Consequently, this research rejects monolithic end-to-end architectures and instead proposes a hybrid modular system based on a hierarchical approach.

Computational epigraphic analysis operates in consecutive phases, where the performance of each stage conditions the next. This work focuses on the first two phases concerning complete stelae: first, it executes a coarse localization to isolate frequent signs from a specific category or set of classes. Subsequently, it applies a fine-grained classification that assigns each character to a Gardiner code. The initial localization stage compares the performance of the YOLO11 convolutional network and the RF-DETR transformer model. Relying on frequency analysis from the AED-TEI dictionary, specific models are trained on high-resolution crops sourced from different stelae. One model focuses on birds, while the other targets the glyphs with the highest coverage in the corpus. After consolidating these expert localizers, a third generic model pseudo-labeled using the detections from these prior models leverages the global attention of RF-DETR to infer the position of unknown signs using context. The hybridization of the most efficient models from each architecture enables the processing of complete stelae via a sliding window system. Into this pipeline, a priority-conditioned global non-maximum suppression algorithm is incorporated to filter out redundant detections caused by overlapping.

The subsequent fine-grained classification is entrusted to ConvNeXt models, an architecture validated for its robustness against erosion in reliefs. This phase handles uncertainty by applying strict confidence thresholds and implements rejection classes, which route doubtful symbols to the generic model for morphological review. To overcome the shortage of samples in minority categories, training is augmented by generating synthetic data through conditional cGANs and BicycleGAN, which preserve stroke fidelity.

This integration of hierarchical localization with expert classification provides a solid technical foundation for digital epigraphic processing. The system constitutes a scalable tool and lays the groundwork for future phases of transliteration and automated translation of the texts.

Keywords: Image processing, Neural networks, Machine learning, Artificial intelligence techniques, Computer vision.

AGRADECIMIENTOS

A Javier, por confiar en mí y darme la oportunidad de dar forma a este proyecto bajo tu dirección. Gracias por tu paciencia, apoyo y consejos, sin los cuales este trabajo no habría sido posible.

De igual manera, quiero dar las gracias a Raúl. Su extraordinario trabajo previo y el conjunto de datos que ha recopilado han sido el punto de partida y el motor que ha permitido articular toda esta investigación.

ÍNDICE

1	Introducción.....	1
1.1	Motivación y Justificación.....	1
1.2	Objetivos.....	2
1.3	Organización de la memoria.....	3
2	Estado del arte	5
2.1	Evolución de las técnicas de visión artificial aplicadas a la epigrafía egipcia	5
2.2	Detección de objetos.....	6
2.3	Aumento de datos	7
2.4	Clasificadores de grano fino	8
2.5	Conclusiones del estado del arte.....	9
3	Solución propuesta	11
3.1	Análisis y preparación de datos	11
3.2	Modelos expertos y modelo genérico contextual.....	12
3.3	Clasificadores y aumento de datos.....	13
3.4	Arquitectura del sistema	17
3.4.1	Fase de localización.....	17
3.4.2	Fase de clasificación.....	19
4	Resultados obtenidos.....	21
4.1	Métricas	21
4.2	Configuración	22
4.3	Evaluación de modelos expertos.....	24
4.4	Evaluación del modelo genérico.....	28
4.5	Resultados del proceso de detección.....	29
4.6	Clasificador de «Aves»	33
4.6.1	Modelo 1: Frecuencia	36
4.6.2	Modelo 2: Aislamiento semántico.....	38
4.6.3	Modelo 3: Agrupación.....	39
4.6.4	Evolución comparada.....	41
4.7	Clasificador «Top 4».....	43
4.8	Clasificador genérico «999»	45
4.9	Resultados del proceso global de clasificación.....	47
4.9.1	Recortes.....	49
5	Conclusiones y trabajo futuro.....	51
	Bibliografía	53
	Apéndice A: <i>Datasets</i> y configuraciones de entrenamiento.....	57
	A.1 <i>Datasets</i>	57
	A2. Configuraciones de entrenamiento	63
	Apéndice B: Estelas y papiros analizados	67
	B1. Stela of Nya	67

B2. Round-topped stele of Ipuy	71
B3. Stela dedicated to the goddess Renenutet by Mutneferet and Iyneferet	75
B4. Stela of Neferabu	79
B5. Stela of Intef and Nesumontu	81
B6. False-door stela of Harbes	85
B7. Magical stela or cippus of Horus (Reverso)	89
B8. Stela of the lector priest of Amun Siamun.....	93
B9. Stela of the temple servant of Amun Irthareru	97
B10. Stela of Horemkhauf.....	101
B11. Magical Stela with Horus the Child (Anverso).....	105
B12. Magical Stela with Horus the Child (Reverso).....	109
B13. Sheet from the Papyrus of Amenhotep	113
B14. Heqanahkt Letter II.....	117

ÍNDICE DE FIGURAS

FIGURA 1. FRECUENCIA DE LOS PRINCIPALES JEROGLÍFICOS DEL CORPUS AED-TEI	2
FIGURA 2. SIMILITUD MORFOLÓGICA: (A) G1 (ALIMOCHÉ), FUENTE NEWGARDINER G1; (B) G4 (BUSARDO), FUENTE NEWGARDINER G4 ...	13
FIGURA 3. PIPELINE AUMENTO DE DATOS CGAN+	15
FIGURA 4. PROCESO HÍBRIDO DE LOCALIZACIÓN. ESTELA DE HOREMKHAUEF.	17
FIGURA 5. FASE DE CLASIFICACIÓN.....	20
FIGURA 6. COMPARATIVA VALIDACIÓN MAP@50 (CLASES «AVES» Y «TOP4»).	25
FIGURA 7. MATRICES DE CONFUSIÓN NORMALIZADAS DEL MODELO «AVES» SOBRE EL CONJUNTO DE TEST. COMPARATIVA ENTRE LOS RESULTADOS DE: (A) YOLO11 Y (B) RF-DETR.	26
FIGURA 8. MATRICES DE CONFUSIÓN NORMALIZADAS DEL MODELO «TOP 4» SOBRE EL CONJUNTO DE TEST. COMPARATIVA ENTRE LOS RESULTADOS DE: (A) YOLO11 Y (B) RF-DETR.	26
FIGURA 9. COMPARATIVA ENTRE LOS RESULTADOS DE DETECCIÓN DEL MODELO «AVES» YOLO11 (IZQUIERDA) Y RF-DETR (DERECHA) CON UMBRAL DE CONFIANZA AL 0,25. ADAPTADO DE FRAGMENT OF HENENU STELA. THE METROPOLITAN MUSEUM OF ART (THE MET).	27
FIGURA 10. COMPARATIVA ENTRE LOS RESULTADOS DE DETECCIÓN DE MODELO «TOP 4». YOLO11 (IZQUIERDA) Y RF-DETR (DERECHA) CON UMBRAL DE CONFIANZA AL 0,25. ADAPTADO DE TOMB STELA OF HETEPSI. THE MET.	27
FIGURA 11. COMPARATIVA ENTRE LOS RESULTADOS DE DETECCIÓN DE MODELO «AVES». YOLO11 (IZQUIERDA) Y RF-DETR (DERECHA) CON UMBRAL DE CONFIANZA AL 0,50. ADAPTADO DE ESTELA FUNERARIA DEL SACERDOTE GEM. POR TM, 2021, FLICKR.	27
FIGURA 12. COMPARATIVA ENTRENAMIENTO YOLO VS RF-DETR (CLASE «999»).	28
FIGURA 13. MATRIZ DE CONFUSIÓN RF-DETR (CLASE «999»).	29
FIGURA 14. COMPARATIVA ENTRE LOS RESULTADOS DE DETECCIÓN DEL MODELO «999». YOLO11 (IZQUIERDA) Y RF-DETR (DERECHA) CON UMBRAL DE CONFIANZA AL 0.25. ADAPTADO DE STELA DEDICATED BY SMEN TO RE-HORAKHTY. MUSEO EGIZIO DI TORINO.	30
FIGURA 15. RESULTADO DEL PROCESO DE DETECCIÓN SOBRE MAGICAL STELA WITH HORUS THE CHILD (ESTELA MÁGICA).	32
FIGURA 16. SELECCIÓN DE ESCALA POR MÁX. DETECCIONES.	33
FIGURA 17. DETALLE ANVERSO DE MAGICAL STELA WITH HORUS THE CHILD (ESTELA MÁGICA): (A) NMS MULTIESCLA; (B) NMS GLOBAL ...	33
FIGURA 18. HOJA DEL PAPIRO DE AMENHOTEP.	34
FIGURA 19. TODAS LAS MUESTRAS ORIGINALES DE G7.	35
FIGURA 20. EJEMPLOS DE MUESTRAS GENERADAS SINTÉTICAMENTE PARA LA CLASE G7	36
FIGURA 21. CURVAS DE ENTRENAMIENTO DEL MODELO 1	37
FIGURA 22. MATRIZ DE CONFUSIÓN DEL MODELO 1	37
FIGURA 23. CURVA DE ENTRENAMIENTO DEL MODELO 2	38
FIGURA 24. MATRIZ DE CONFUSIÓN DEL MODELO 2	39
FIGURA 25. CURVA DE ENTRENAMIENTO DEL MODELO 3	40
FIGURA 26. MATRIZ DE CONFUSIÓN DEL MODELO 3	40
FIGURA 27. HALCÓN G5 COMO RESTO_AVES.....	42
FIGURA 28. RESOLUCIÓN DE AMBIGÜEDAD G36/G37	42
FIGURA 29. RECUPERACIÓN POR SUPERCLASE.	42
FIGURA 30. SIMILITUD MORFOLÓGICA: (A) D4 (OJO), F. NEWGARDINER D4; (B) D21 (BOCA), F. NEWGARDINER D24	43
FIGURA 31. CURVA DE ENTRENAMIENTO «TOP 4»	44
FIGURA 32. MATRIZ DE CONFUSIÓN «TOP 4»	44
FIGURA 33. CURVA DE ENTRENAMIENTO DEL CLASIFICADOR GENÉRICO «999».....	46
FIGURA 34. MATRIZ DE CONFUSIÓN DEL CLASIFICADOR GENÉRICO «999».....	47
FIGURA 35. RESULTADOS DEL PROCESO GLOBAL DE CLASIFICACIÓN SOBRE UN EXTRACTO DE ESTELA DE HOREMKHAUEF.	48
FIGURA 36. RECORTES DE RESTO DE AVES.	49
FIGURA 37. D21 CLASIFICADOS CORRECTAMENTE, PERO CON ALGÚN D4.	49
FIGURA 38. EJEMPLOS CLASIFICADOS COMO A1 O PARECIDOS A A1.	49
FIGURA B. 1 – B. 56.	67-119
▪ DETECCIONES POR ESCALA RF-DETR (AZUL) Y YOLO (NARANJA)	
▪ DETECCIONES SOBRE ESTELA CON NMS GLOBAL	
▪ DETECCIONES SOBRE ESTELA UTILIZANDO NMS CON FILTRADO DE PRIORIDAD.	
▪ CLASIFICACIÓN SOBRE DETECCIONES.	

ÍNDICE DE TABLAS

TABLA 1. TRABAJOS RELEVANTES EN LA EVOLUCIÓN DE LA EPIGRAFÍA EGIPCIA DIGITAL	5
TABLA 2. EVOLUCIÓN DE YOLO Y DETR.....	7
TABLA 3. FRECUENCIAS CATEGORÍAS GARDINER EN EL CORPUS AED-TEI.....	11
TABLA 4. FRECUENCIA DE JEROGLÍFICOS INDIVIDUALES EN EL CORPUS AED-TEI	12
TABLA 5. COMPARATIVA DE HIPERPARÁMETROS YOLO11L VS RF-DETR (S).....	23
TABLA 6. COMPARATIVA DE MÉTRICAS DE YOLO11 VS RF-DETR (CLASE «AVES»)	25
TABLA 7. COMPARATIVA DE MÉTRICAS DE YOLO11 VS RF-DETR (CLASE «TOP 4»).....	25
TABLA 8. COMPARATIVA DE MÉTRICAS DE YOLO11 VS RF-DETR (CLASE «999»)	29
TABLA 9. EJEMPLOS DE LOCALIZACIÓN SOBRE ESTELAS	31
TABLA 10. JEROGLÍFICOS CATEGORÍA G	35
TABLA 11. DISTRIBUCIÓN DATASET MODELO DE FRECUENCIA.	36
TABLA 12. MÉTRICAS POR CLASE DEL MODELO 1	37
TABLA 13. DISTRIBUCIÓN DEL DATASET DEL MODELO DE AISLAMIENTO SEMÁNTICO.	38
TABLA 14. MÉTRICAS POR CLASE DEL MODELO 2	39
TABLA 15. DISTRIBUCIÓN DATASET DE ENTRENAMIENTO DEL MODELO AGRUPADO.	39
TABLA 16. MÉTRICAS POR CLASE DEL MODELO 3	40
TABLA 17. DISTRIBUCIÓN DATASET MODELO «TOP 4».	43
TABLA 18. MÉTRICAS DEL CLASIFICADOR «TOP 4»	44
TABLA 19. DISTRIBUCIÓN DEL DATASET DEL CLASIFICADOR GENÉRICO «999»	45
TABLA 20. MÉTRICAS DEL CLASIFICADOR GENÉRICO «999».....	46
TABLA 21. CLASIFICACIÓN DEL CONJUNTO DE EJEMPLOS.	48
TABLA A. 1. LISTADO DE FRAGMENTOS UTILIZADOS EN LOS DATASETS DE LOCALIZACIÓN	57
TABLA B. 1 – B. 28.	67 -117
▪ RESULTADOS DE DETECCIÓN GLOBAL Y OPTIMIZACIÓN DE ESCALAS	
▪ RESULTADOS DE CLASIFICACIÓN EXPERTA (CONVNeXT)	

GLOSARIO

Epigrafía: Disciplina arqueológica centrada en el estudio, desciframiento e interpretación de inscripciones antiguas.

Estela: Soporte físico que contiene los relieves epigráficos analizados.

Glifo (Jeroglífico): Unidad gráfica mínima de la escritura del antiguo Egipto y objeto central que las redes neuronales deben aislar y clasificar.

Recorte (Tile): Extracto de una fotografía original de alta resolución mediante la técnica de ventanas deslizantes.

Categoría de Gardiner: Agrupación temática, identificada con una letra mayúscula, en la que el egiptólogo Alan Gardiner dividió funcionalmente los símbolos jeroglíficos.

Clase o Código Gardiner: Signo jeroglífico individual. Se designa mediante un código alfanumérico que combina la letra de su categoría temática seguida de un identificador numérico secuencial.

Distribución de larga cola (Long-tailed distribution): Desequilibrio estadístico del corpus epigráfico, donde un grupo muy reducido de glifos acapara la gran mayoría de las muestras, frente a una inmensa cantidad de signos con presencia marginal.

Aumento de datos (Data Augmentation): Generación de transformaciones sintéticas sobre imágenes originales que permite multiplicar las muestras de entrenamiento para compensar las asimetrías entre clases.

Etiquetado de referencia (Ground truth): Anotaciones manuales que definen la solución correcta. Incluyen tanto las coordenadas espaciales exactas de la caja delimitadora como la asignación de su clase Gardiner correspondiente.

Contaminación de datos (Data Leakage): Defecto metodológico que ocurre cuando información del conjunto de validación o prueba se filtra hacia la fase de entrenamiento, inflando irrealmente las métricas de rendimiento.

Sobreajuste (Overfitting): Fenómeno provocado cuando la red neuronal memoriza las muestras originales en lugar de aprender sus rasgos morfológicos, anulando su capacidad para detectar jeroglíficos en estelas no vistas.

Transferencia de aprendizaje (Transfer Learning): Inicialización de los pesos de la red importando el conocimiento previo de modelos entrenados sobre conjuntos masivos (ImageNet, COCO) para adaptarlos después al contexto epigráfico.

Relleno (Zero-padding): Técnica de preprocesamiento que añade píxeles nulos en los bordes de la imagen para homogeneizar las dimensiones de los recortes antes de introducirlos en la red.

Espacio latente: Representación matemática comprimida que utilizan las redes generativas para mapear y procesar la información visual de forma eficiente.

Modelo de difusión latente (Latent Diffusion Model): Arquitectura generativa que añade y elimina ruido gradualmente dentro del espacio latente comprimido, para evitar la carga computacional de operar píxel a píxel.

Adaptadores de bajo rango (LoRA): Estrategia de ajuste fino que adapta grandes topologías preentrenadas a tareas muy concretas sin necesidad de recalcular todos los parámetros originales del modelo base.

LISTADO DE ACRÓNIMOS

CNN: Redes Neuronales Convolucionales (Convolutional Neural Networks).

DETR: Transformador de Detección (DEtection TRansformer).

YOLO: You Only Look Once.

RT-DETR: Transformadores de Detección en Tiempo Real (Real-Time Detection Transformers).

RF-DETR: Roboflow Detection Transformer.

GAN: Redes Generativas Antagónicas (Generative Adversarial Networks).

cGAN: Redes Generativas Antagónicas Condicionales (Conditional Generative Adversarial Networks).

HOG: Histograma de Gradientes Orientados (Histogram of Oriented Gradients).

NMS: Supresión de No Máximos (Non-Maximum Suppression).

NST: Transferencia de Estilo Neuronal (Neural Style Transfer).

SAHI: Inferencia Asistida por Fragmentación (Slicing Aided Hyper Inference).

GPU: Unidad de Procesamiento Gráfico (Graphics Processing Unit).

VRAM: Memoria de Vídeo de Acceso Aleatorio (Video Random Access Memory).

GFLOPs: Gigaoperaciones de Punto Flotante (Giga Floating-Point Operations).

1 Introducción

El estudio y la traducción de los jeroglíficos egipcios plantean dificultades cognitivas y técnicas complejas, derivadas de la tensión entre la naturaleza flexible de esta lengua y la necesidad metodológica actual de establecer reglas estandarizadas. A diferencia de las clasificaciones rígidas de los sistemas de información modernos, la escritura del antiguo Egipto operaba como un sistema dinámico donde el arte figurativo y la sintaxis gramatical resultaban inseparables (Gracia Zamacona, 2018). Los signos modificaban su forma, género y atributos según el contexto semántico de la oración, lo que impide un análisis puramente estático.

A esta variabilidad morfológica se suma la profunda tradición gráfica del idioma. Como señaló Gardiner (1957), la temprana estandarización de los caracteres provocó que cientos de símbolos se mantuvieran en uso milenios después de la desaparición de los objetos reales que representaban. Esta distancia entre el símbolo original y la realidad de los escribas tardíos generó fosilizaciones, errores de copia y ambigüedades que dificultan la taxonomía científica actual y dejan glifos sin clasificar. Por consiguiente, los esquemas computacionales rígidos resultan ineficaces ante un corpus histórico tan dependiente del contexto visual, lo que exige el desarrollo de enfoques tecnológicos capaces de operar dentro de estos márgenes de incertidumbre.

1.1 Motivación y Justificación

Para evaluar las distintas arquitecturas de visión artificial, se ha recurrido al Egyptian Hieroglyphic Dataset (Fuentes-Ferrer *et al.*, 2025a), dando así continuidad a los experimentos previos (Fuentes-Ferrer *et al.*, 2025b). Sus más de 13.000 muestras repartidas en 310 clases proporcionan un volumen de datos lo suficientemente amplio para entrenar modelos de aprendizaje profundo frente a los problemas reales que presentan los soportes epigráficos, tales como el desgaste de la piedra, el ruido de fondo o los cambios drásticos de luz.

Si bien este planteamiento aprovecha los principios de segmentación y posterior clasificación propuestos por los citados autores, el avance en las arquitecturas de detección de una sola etapa abre la puerta a una reorientación metodológica. El desarrollo actual de las redes neuronales convolucionales y la consolidación de los modelos basados en Transformers aplicados a la visión artificial permiten comprobar si las herramientas vigentes son capaces de superar las limitaciones técnicas que, en su momento, obligaron a descartar la detección directa sobre soportes complejos. Bajo esta premisa, el análisis de la frecuencia de los jeroglíficos en el corpus AED-TEI (Schweitzer y Ancient Egyptian Dictionary, 2019) sirve para identificar las clases con mayor cobertura (Figura 1) y articular una estrategia de detección adaptada a las particularidades lingüísticas del periodo.

Estas técnicas se complementan con la transferencia de aprendizaje y el aumento de datos sintéticos con el fin de compensar el desbalance en las clases prioritarias, con el fin de lograr clasificadores más robustos.

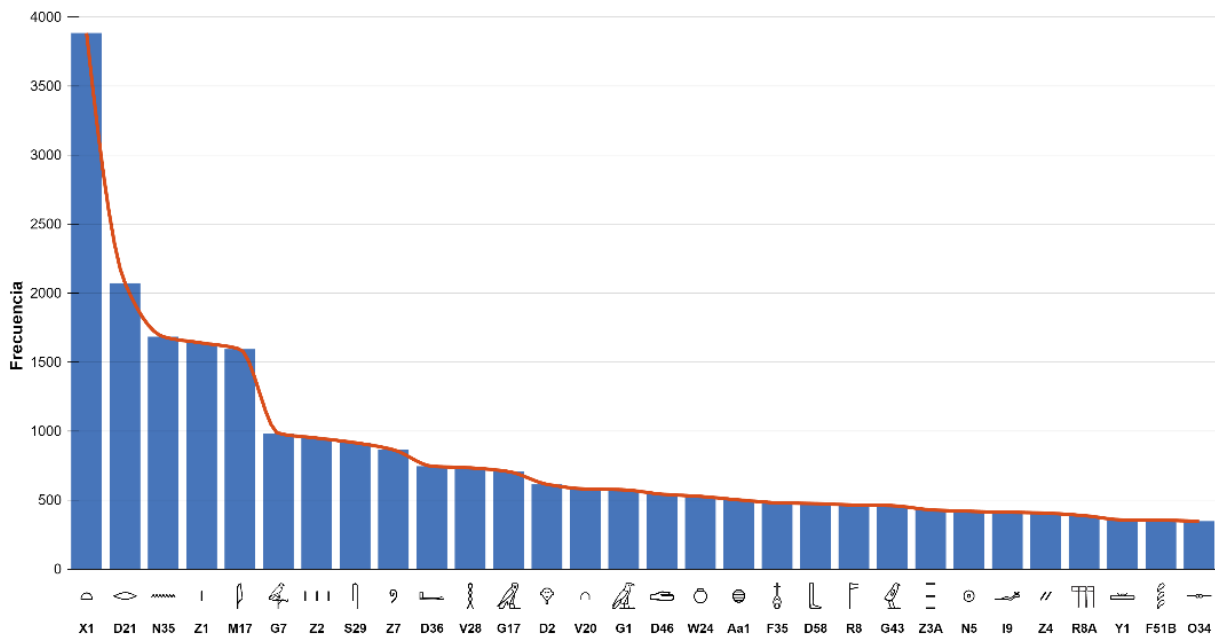


Figura 1. Frecuencia de los principales jeroglíficos del corpus AED-TEI

La primera fase de la metodología consiste en recopilar estelas en alta resolución desde diversos repositorios (Apéndice A) y dividir las en recortes regulares (tiles). Este paso facilita el entrenamiento de un modelo inicial centrado en un subconjunto de glifos específicos, cuya selección responde a su alta frecuencia de aparición en los textos de los Reinos Antiguo y Medio del corpus AED-TEI. Al priorizar estos signos mayoritarios, el entrenamiento garantiza una alta cobertura y se enfoca directamente su entrenamiento en la base estructural del lenguaje.

La estrategia de detección jerárquica se articula en varios frentes complementarios. Comienza por el etiquetado y entrenamiento de modelos YOLO11 (Khanam y Hussain, 2024) y RF-DETR (Robinson *et al.*, 2026; Zhao *et al.*, 2024a) sobre los glifos de alta frecuencia para obtener su localización exacta en la estela. Estos modelos permiten realizar una primera identificación gruesa y procesar los recortes con modelos expertos de grano fino que utilizan ConvNeXt (Woo *et al.*, 2023), empleando el conjunto de datos de Fuentes-Ferrer *et al.* (2025a) y mitigando la escasez de las clases minoritarias mediante la aplicación de redes generativas antagónicas (Isola *et al.*, 2017). Esta arquitectura comparativa permite evaluar el rendimiento entre los paradigmas convolucional y de transformadores, prestando especial atención a la precisión de detección sobre soportes fuertemente erosionados.

La integración definitiva del sistema utiliza los modelos entrenados con los jeroglíficos más frecuentes como puntos de anclaje contextual para estimar la posición del resto de los signos desconocidos. De este modo se consolida una cadena de procesamiento integral capaz de localizar y clasificar signos en inscripciones completas, apoyándose en la técnica de inferencia asistida por fragmentación, (Akyon *et al.*, 2022).

1.2 Objetivos

Para abordar esta problemática, la investigación plantea varios objetivos interrelacionados. En primer lugar, se evalúa si arquitecturas avanzadas como YOLO11 y RF-DETR logran establecer un nuevo estándar de rendimiento en el reconocimiento de la epigrafía de este periodo, superando las barreras

de complejidad visual documentadas previamente (Fuentes-Ferrer *et al.*, 2025b). En el plano metodológico, se busca determinar si una estrategia de grueso a fino resulta más eficaz que los métodos directos al separar la localización espacial del glifo de su clasificación final. De forma complementaria, se analiza la viabilidad del aumento de datos híbrido (geométrico y generativo) para facilitar la transferencia de aprendizaje en las clases críticas que sufren escasez de muestras. Como objetivo final, se pretende comprobar de forma empírica si la capacidad de atención global, característica de modelos como RF-DETR, favorece la detección de jeroglíficos desconocidos. Todo este diseño está orientado a consolidar un sistema de apoyo al egiptólogo que agilice la labor de transcripción y transliteración frente a estelas complejas.

1.3 Organización de la memoria

En cuanto a la organización del artículo: en el Capítulo 2 se realiza una revisión histórica de las técnicas de visión artificial aplicadas a la epigrafía, analizando los retos de la microdetección y los avances logrados por las arquitecturas libres de anclas y los modelos basados en transformadores, así como las técnicas avanzadas de aumento de datos y los clasificadores de grano fino. En el Capítulo 3 se expone el diseño metodológico de la solución. Inicialmente, las arquitecturas YOLO11 y RF-DETR se entrenan de forma progresiva basándose en un análisis de frecuencias, empleando pseudoetiquetado para detectar glifos inéditos y estructurales. Seguidamente, la fase de clasificación se delega en redes ConvNeXt, cuyo entrenamiento se ha estabilizado mediante estrategias de muestreo selectivo y aumento de datos. Finalmente, tras exponer el marco de evaluación, definir las métricas de control y validar el rendimiento de toda la cadena de procesamiento frente a escenarios de referencia en el Capítulo 4, el trabajo concluye en el Capítulo 5 con la discusión de los resultados obtenidos y la apertura de futuras líneas de investigación.

2 Estado del arte

2.1 Evolución de las técnicas de visión artificial aplicadas a la epigrafía egipcia

La aplicación de la visión artificial a la epigrafía egipcia ha evolucionado desde los primeros modelos de descriptores visuales en 2013 hasta los sistemas actuales de aprendizaje profundo (*deep learning*). El análisis computacional de estas inscripciones exige un proceso en cadena: aislar el glifo del ruido del soporte, clasificarlo, transliterar su valor fonético y traducirlo. Al tratarse de un proceso secuencial, la precisión de cada fase determina directamente el éxito de las posteriores.

La revisión bibliográfica de la última década (Tabla 1) analiza cómo los desarrollos en visión artificial se han orientado prioritariamente hacia la localización, aislamiento y clasificación de signos en soportes con diversos estados de conservación.

Tabla 1. Trabajos relevantes en la evolución de la epigrafía egipcia digital

Año	Autor(es)	Título del Artículo	Alcance
2013	Franken y van Gemert	Automatic Egyptian hieroglyph recognition by retrieving images as texts	Identificación y Clasificación
2021	Barucci <i>et al.</i>	A Deep Learning Approach to Ancient Egyptian Hieroglyphs Classification	Clasificación
2023	Guidi <i>et al.</i>	Egyptian Hieroglyphs Segmentation with Convolutional Neural Networks	Identificación/Segm. y Clasificación
2023	Sobhy <i>et al.</i>	An AI Based Automatic Translator for Ancient Hieroglyphic Language...	Identificación, Clasificación y Traducción
2025	Fuentes-Jimenez <i>et al.</i>	The OCR-PT-CT Project: Semi-Automatic Recognition of Ancient Egyptian Hieroglyphs Based on Metric Learning	Clasificación
2025	Nasser <i>et al.</i>	HieroGlyphTranslator: Automatic Recognition and Translation of Egyptian Hieroglyphs to English	Identificación, Clasificación y Traducción
2025	Fuentes-Ferrer <i>et al.</i>	Recognition of Egyptian hieroglyphic texts through focused generic segmentation and cross-validation voting	Identificación/Segm. y Clasificación

Los primeros pasos en la automatización de la lectura corresponden a Franken y van Gemert (2013), quienes demostraron la viabilidad del proceso utilizando el *dataset* de la Pirámide de Unas y descriptores clásicos como HOG. Posteriormente, la adopción de redes neuronales supuso un salto cualitativo. Barucci *et al.* (2021) confirmaron la superioridad del aprendizaje profundo sobre la visión clásica con su red dedicada Glyphnet, mientras que Guidi *et al.* (2023) abordaron el problema crítico de los soportes deteriorados empleando técnicas de segmentación de instancias con la arquitectura Mask R-CNN.

En paralelo a la clasificación aislada, diversos autores han buscado soluciones integrales (*end-to-end*). Sobhy *et al.* (2023) propusieron combinar detección, clasificación mediante redes siamesas y

traducción al inglés en un mismo marco de trabajo. En esta misma línea, el sistema *HieroGlyphTranslator* (Nasser *et al.*, 2025) propone una arquitectura mixta que extrae los códigos de Gardiner mediante redes ResNet50 y delega la traducción directa en modelos basados en transformadores.

Para mitigar el impacto del desbalance de datos en la fase de clasificación, Fuentes-Jimenez *et al.* (2025) recurren al *Deep Metric Learning*. Al mapear las imágenes por similitud dentro de un espacio de características, este enfoque reduce el sesgo hacia las clases dominantes y permite identificar glifos excepcionales mediante el aprendizaje de pocas muestras (*few-shot learning*). El avance metodológico más significativo en este ámbito, es el conseguido por Fuentes-Ferrer *et al.* (2025b) combinando métodos híbridos de segmentación con modelos ConvNeXt y validación cruzada por votación, generando en el proceso corpus masivo que permite elevar la precisión de referencia actual.

2.2 Detección de objetos

La localización de jeroglíficos en estelas egipcias plantea un desafío complejo debido a la alta densidad de los signos, su tamaño reducido y la frecuente degradación del soporte. Como señalan Sayyad y Attarde (2026), los métodos convencionales de visión artificial resultan insuficientes en escenarios de baja claridad y alta aglomeración, lo que justifica el empleo de sistemas híbridos y redes neuronales de última generación capaces de extraer características morfológicas finas.

Con el fin de buscar una solución a la detección de objetos pequeños, la literatura reciente se divide en dos enfoques principales. El primero prescinde de las cajas ancla (*anchor-free*). Dado que los glifos menores presentan proporciones variables y carecen de rasgos geométricos estables, la predicción directa de puntos clave combinada con información contextual dota de mayor estabilidad a la detección. Al evitar las restricciones de los mapas de anclaje tradicionales, se mitigan los errores por desalineación. En este sentido, Tian *et al.* (2026) constatan que la evolución de la arquitectura YOLO hacia diseños libres de anclajes, complementados con mecanismos de atención, reduce la redundancia computacional y mejora el rendimiento en escenarios visualmente complejos.

El segundo enfoque se basa en los modelos basados en Transformers de extremo a extremo (*end-to-end*), cuya ventaja principal radica en la gestión de caracteres superpuestos. Shehzadi *et al.* (2026) recomiendan el uso de la familia DETR por su capacidad para suprimir componentes heurísticos manuales como la supresión de no máximos. La omisión de este filtrado favorece el aprendizaje de invarianzas y minimiza la pérdida de información en zonas con alta densidad de signos (González-Hernández *et al.*, 2025). En entornos de aprendizaje con pocos ejemplos (Zhang *et al.*, 2025), la hibridación de detectores libres de anclas y transformadores ofrece resultados óptimos. En el ámbito epigráfico, esta combinación facilita la identificación de jeroglíficos de baja frecuencia e incorpora nuevas clases sin requerir reentrenamientos masivos ni incurrir en el olvido catastrófico de la red.

En cuanto al rendimiento global, arquitecturas basadas en transformadores como RT-DETR demuestran que los mecanismos de autoatención y atención cruzada logran modelar el contexto espacial con gran eficacia (Madan y Reich, 2025). Las evaluaciones recientes de Robinson *et al.* (2026) con su modelo RF-DETR demuestran que las variantes optimizadas de los transformadores logran establecer el estado del arte en precisión bruta, siendo los primeros en superar la métrica de 60 mAP en COCO en tiempo real. Sin embargo, más allá de la precisión bruta, el principal debate se centra en el coste computacional de implementar estos sistemas. En este sentido, Sapkota y Karkee (2026) señalan que las versiones más avanzadas de la familia YOLO, resultan opciones mucho más viables para operar en entornos de bajos recursos. Al no exigir el enorme volumen de operaciones propio de los transformadores, estas arquitecturas convolucionales garantizan una exportación más sencilla y una latencia

significativamente menor. La Tabla 2 recoge la evolución y los principales avances técnicos de las arquitecturas YOLO y DETR.

Tabla 2. Evolución de YOLO y DETR.

Año	Evolución YOLO	Evolución DETR/Transformers
2016	YOLOv1: You Only Look Once: Unified, Real-Time Object Detection (Redmon <i>et al.</i>)	-
2017	-	Attention Is All You Need (Vaswani <i>et al.</i>)
2018	YOLOv3: An Incremental Improvement (Redmon & Farhadi)	-
2020	YOLOv4: Optimal Speed and Accuracy of Object Detection (Bochkovskiy <i>et al.</i>)	DETR: End-to-End Object Detection with Transformers (Carion <i>et al.</i>)
	YOLOv5 (Jocher <i>et al.</i> / Ultralytics)	
2023	YOLOv8 (Jocher <i>et al.</i> / Ultralytics)	RT-DETR: DETRs Beat YOLOs on Real-time Object Detection (Zhao <i>et al.</i> / Baidu)
2024	YOLOv10: Real-Time End-to-End Object Detection (Wang <i>et al.</i>)	RT-DETR (Publicación oficial): (Zhao <i>et al.</i>)
	YOLOv11 (Jocher <i>et al.</i> / Ultralytics)	
2025	YOLO26 (Sapkota <i>et al.</i> / Ultralytics)	
2026	-	RF-DETR: Neural Architecture Search for Real-Time Detection Transformers (Robinson <i>et al.</i>)

Se puede comprobar como la evolución de ambas vertientes presenta trayectorias distintas. Mientras que YOLO ha avanzado de forma incremental hacia la integración de tareas complejas como la segmentación y el seguimiento, los detectores basados en Transformers derivan del mecanismo de auto-atención introducido por Vaswani *et al.* (2017). Esta base permitió el desarrollo posterior de RT-DETR (Zhao *et al.*, 2024), el primer modelo de esta línea capaz de operar en tiempo real y competir de forma directa con YOLO al eliminar el procesamiento de supresión de no máximos, un avance optimizado recientemente mediante búsqueda de arquitectura neuronal en modelos como RF-DETR (Robinson et a., 2026).

2.3 Aumento de datos

La escasez y el desequilibrio de datos en epigrafía obligan a recurrir a técnicas de aumento que superen las limitaciones de las transformaciones geométricas clásicas. Entre estas técnicas destacan las Redes Generativas Antagónicas (GAN). En concreto, arquitecturas como Pix2Pix (Isola *et al.*, 2017) y CycleGAN (Zhu, Park, *et al.*, 2017) establecieron las bases para la traducción entre dominios de imagen. A partir de estas, modelos como BicycleGAN (Zhu, Zhang, *et al.*, 2018) consiguen variabilidad al evitar el colapso de modo, el cual demuestra su utilidad práctica en la generación de caracteres antiguos (Liu *et al.*, 2019). Recientemente, se han desarrollado enfoques más específicos centrados en jeroglíficos, como el trabajo de Creed (2025), quien emplea la transferencia de estilo neuronal (NST) para proyectar texturas, relieves y pátinas de fotografías arqueológicas reales sobre glifos sintéticos procedentes de tipografías digitales.

Actualmente, los modelos de difusión latente (Rombach *et al.*, 2022) superan a las GANs gracias a su rendimiento con texturas complejas, aunque su principal desafío sigue siendo el control geométrico de

la generación. Soluciones como ControlNet (Zhang *et al.*, 2023) o el condicionamiento local de Zhao *et al.* (2024b) resuelven este problema mediante la inyección de mapas de bordes que guían la estructura formal sin desvirtuar el entorno del objeto. Además, la integración de adaptadores de bajo rango como LoRA (Hu *et al.*, 2021) permite optimizar estas redes con arquitecturas de aprendizaje de pocos ejemplos.

A pesar de esta potencia visual, la aplicación de modelos de difusión en epigrafía presenta claroscuros. Cao *et al.* (2026) advierten de la brecha existente entre el realismo estético de las imágenes y la fidelidad de grano fino requerida. Dado que la disciplina exige una precisión morfológica absoluta, la pérdida de información de alta frecuencia al decodificar desde el espacio latente compromete la fiabilidad de los signos sintéticos generados.

El problema no se limita solo a la generación de muestras, influye también el desequilibrio estructural inherente a los corpus epigráficos, caracterizados por distribuciones de larga cola (*Long-tailed distribution*). Para mitigar la pérdida de rendimiento provocada por la desigualdad de clases, estudios como los de Kang *et al.* (2020) y Xiong *et al.* (2026) recomiendan aislar el aprendizaje de representaciones. Bajo este planteamiento, se considera que el extractor visual es resistente al desbalanceo, lo que permite derivar todo el esfuerzo de compensación estadística (mediante ajuste de pesos o remuestreo) directamente al clasificador. De este modo se estabilizan las fronteras de decisión y se corrige la tendencia del modelo a favorecer los jeroglíficos mayoritarios.

2.4 Clasificadores de grano fino

El sistema propuesto por Fuentes-Ferrer *et al.* (2025b) fundamenta la extracción de características y la categorización de glifos en la arquitectura ConvNeXt. Este diseño surge para actualizar las redes convolucionales clásicas y mantener su competitividad frente al avance de los Vision Transformers (Liu *et al.*, 2022). La red ConvNeXt evoluciona la estructura tradicional de ResNet asimilando mecánicas de atención, destacando la ampliación de sus núcleos y el uso de convoluciones en profundidad. Esta arquitectura se optimiza en la variante ConvNeXt V2 (Woo *et al.*, 2023), que perfecciona este diseño integrando autoencoders de máscara para estabilizar y potenciar el aprendizaje de las representaciones visuales.

La elección de ConvNeXt para el análisis epigráfico responde a su alta tolerancia frente a la degradación física del soporte. Como extractores de características, estos bloques han demostrado previamente su capacidad para aislar elementos diminutos sobre fondos ruidosos o de muy baja visibilidad. Una capacidad de resistencia al ruido que ya ha sido documentada en la detección de defectos milimétricos y en termografía infrarroja (Liao *et al.*, 2022; Zhou *et al.*, 2023). Precisamente, esta eficacia en condiciones adversas ha motivado la reciente integración de módulos ConvNeXt dentro de las arquitecturas basadas en transformadores. Por ejemplo, Zhou *et al.* (2025) sustituyeron el extractor base de RT-DETR por un esquema Composite-ConvNeXtV2 con el fin de potenciar la percepción de bordes y la fusión multiescala en superficies texturizadas. De manera similar, el modelo NeXt-DETR (Choi y Lee, 2026) aprovecha la normalización por capa (Layer Normalization) en lugar de la normalización por lote (Batch Normalization). Esta sustitución asegura un entrenamiento estable y eficiente en sistemas con restricciones de memoria gráfica y altas resoluciones (1024x1024), lo que evita las distorsiones asociadas al microloteado.

Frente a alternativas sólidas dentro de la clasificación de grano fino, como los Swin Transformers (Liu *et al.*, 2021) o InternImage (Wang *et al.*, 2023), ConvNeXt ofrece ventajas específicas para la epigrafía egipcia. Mientras que los primeros computan la atención de forma jerárquica mediante ventanas

desplazadas, logrando un buen modelado de las dependencias espaciales, pero a costa de una complejidad algorítmica y un consumo de memoria superiores a las redes neuronales convolucionales modernas. InternImage recurre a convoluciones deformables para generar campos receptivos adaptativos. En este escenario, ConvNeXt representa una solución más equilibrada, ya que hereda la eficiencia computacional de las convoluciones estándar y ofrece un rendimiento equiparable al de arquitecturas más complejas sin incrementar el coste técnico.

2.5 Conclusiones del estado del arte

El análisis del estado del arte demuestra que ninguna arquitectura de aprendizaje profundo resuelve por sí sola todos los problemas relacionados con el procesamiento de las estelas egipcia. Ante esta limitación, en este trabajo de fin de máster se descarta la búsqueda de una red estandarizada en favor de un diseño modular. Aunque una arquitectura de extremo a extremo resultaría ideal, un diseño híbrido de aproximación grueso a fino representa una alternativa más realista. Por esta razón, la comparativa entre YOLO y RF-DETR para evaluar cuál proporciona mejores resultados en la detección general se alinea con los marcos de trabajo actuales con mayor probabilidad de éxito en esta tarea, teniendo en cuenta la importancia del contexto y del detalle epigráfico. No obstante, presenta un desafío técnico importante, como es el etiquetado manual masivo. La literatura señala que este proceso requiere una inversión de tiempo y recursos difícilmente asumible, por lo que resulta necesario definir una estrategia de selección de glifos específicamente diseñada para minimizar dicha carga de trabajo.

Para la clasificación fina, se ha comprobado que la arquitectura ConvNeXt iguala el rendimiento de los modelos *Transformer* frente a la degradación de la piedra, pero con una mayor eficiencia computacional. Como este trabajo da continuidad a la investigación de Fuentes-Ferrer *et al.* (2025b), se ha decidido mantener este mismo clasificador para poder establecer comparaciones válidas y garantizar la coherencia metodológica. En esta misma línea, se propone utilizar como recurso de clasificación el conjunto de datos generado por los mismos autores. Sin embargo, debido a la naturaleza de bajo recurso de la lengua egipcia, se produce un desequilibrio de muestras entre las clases minoritarias, lo que obliga a utilizar técnicas avanzadas de aumento de datos. En este contexto, es recomendable descartar el uso de modelos de difusión recientes, ya que la bibliografía advierte sobre la pérdida de detalles de alta frecuencia al decodificar la imagen, un efecto inaceptable al trabajar con jeroglíficos. En su lugar, parece razonable explorar arquitecturas cGAN, capaces de sintetizar nuevas muestras conservando la fidelidad morfológica.

La metodología propuesta integra así la precisión de localización, el análisis contextual, la capacidad clasificatoria demostrada y la fidelidad generativa, estableciendo una base viable para la identificación y clasificación de jeroglíficos mediante una estrategia jerárquica.

3 Solución propuesta

Este trabajo plantea el diseño de un flujo de procesamiento epigráfico vertebrado por arquitecturas de detección profunda, empleando tanto modelos de base convolucional como basados en transformadores. Este detector servirá como fuente de datos de entrada para un sistema de clasificación entrenado con los signos de mayor frecuencia. Apoyado en técnicas realistas de aumento de datos (*data augmentation*), el objetivo final es identificar y clasificar un conjunto de elementos suficientemente representativo para sentar las bases necesarias que permitan avanzar hacia una futura traducción automatizada.

3.1 Análisis y preparación de datos

La frecuencia de aparición de los signos no es uniforme a lo largo del tiempo y depende también del tipo de documento: religioso, funerario o administrativo. Para determinar la distribución y cobertura de los símbolos del Reino Antiguo y Medio, se realiza un análisis del corpus AED-TEI, que recoge más de 48.000 apariciones de jeroglíficos.

Al clasificar estos datos según el estándar de codificación de Gardiner (1957), que distribuye los jeroglíficos en 26 categorías principales, comprobamos que hay diez clases dominantes que concentran más del 81 % del volumen total de inscripciones. Esta concentración evidencia una marcada asimetría en la estructura de la escritura egipcia, tal como se detalla en la Tabla 3:

Tabla 3. Frecuencias categorías Gardiner en el corpus AED-TEI

Categoría	Descripción	Frecuencia	Cobertura
D	Partes del cuerpo humano	6.115	12,70%
Z	Trazos y determinativos	5.135	10,67%
N	Cielo, tierra, agua	4.305	8,94%
X	Panes y pasteles	4.270	8,87%
G	Aves	3.968	8,24%
F	Partes de mamíferos	3.321	6,90%
M	Árboles y plantas	2.811	5,84%
A	El hombre y sus ocupaciones	2.729	5,67%
V	Cuerdas, cestas, sacos	2.688	5,58%
O	Edificios y partes	2.440	5,07%

Respecto a la distribución de signos individuales, se observa que de un total de 1.624 caracteres únicos registrados en el corpus AED-TEI, un grupo de tan solo 10 jeroglíficos genera el 31,86% de la totalidad del texto escrito. Este subconjunto se compone mayoritariamente de signos unilíteros o fonogramas básicos, junto con marcadores gramaticales y determinativos estructurales, tal como se detalla en la Tabla 4.

A la vista de esta distribución estadística, el objetivo se centra en entrenar un sistema capaz de maximizar la cobertura global identificando un grupo reducido de clases clave, respetando siempre la carga semántica del egipcio. Aunque la selección ideal de clases para el modelo de localización se basaría en la frecuencia pura, teniendo en cuenta el esfuerzo y la dificultad que supone el etiquetado manual, es necesario encontrar un compromiso entre la selección de categorías o conjuntos de glifos que optimicen la cobertura y el coste de etiquetado. En la fase de clasificación, este mismo análisis estadístico nos advierte de los retos que se deberán solventar ante la escasez de muestras en clases prioritarias o en conjuntos de glifos estructurales.

Tabla 4. Frecuencia de jeroglíficos individuales en el corpus AED-TEI

Clase	Representación	Total	Cobert.	Descripción
X1		3.885	8,07%	Ideograma/Fonograma Pan (t)
D21		2.072	4,30%	Fonograma de Boca (r)
N35		1.684	3,50%	Fonograma de Agua (n)
Z1		1.639	3,41%	Ideograma/Determinativo Trazo vertical
M17		1.594	3,31%	Ideograma/Fonograma Junco (i)
G7		984	2,04%	Determinativo/Ideograma Halcón sobre estandarte
Z2		950	1,97%	Determinativo Tres trazos verticales (plural)
S29		916	1,90%	Ideograma/Determinativo/Fonograma de Tela doblada (s)
Z7		865	1,80%	Fonograma Espiral (w)
D36		746	1,55%	Ideograma/Determinativo Brazo (a)

3.2 Modelos expertos y modelo genérico contextual

A partir del estado del arte, se ha optado por contrastar el rendimiento de YOLO y RF-DETR. Esta elección permite evaluar los dos paradigmas de detección en tiempo real más avanzados de la actualidad mediante dos lógicas contrapuestas. Como modelo de referencia se establece YOLO11, una arquitectura de una sola etapa altamente optimizada que combina bloques convolucionales con mecanismos de atención para maximizar la extracción jerárquica de características locales. Frente a esta predicción densa se sitúa el enfoque de predicción de conjuntos característico de la familia DETR. El modelo RT-DETR (Zhao *et al.*, 2024a) introdujo la capacidad de operar en tiempo real al prescindir de algoritmos heurísticos de postprocesamiento como la supresión de no máximos. Para esta investigación, la variante implementada (RF-DETR; Robinson *et al.*, 2026) evoluciona dicho esquema al incorporar el modelo fundacional DINOv2 como extractor base, lo que facilita la captura del contexto global de la imagen desde las etapas iniciales mediante autoatención.

La microdetección densa de jeroglíficos mediante arquitecturas de transformadores constituye un desafío técnico, dado que estos modelos presentan limitaciones estructurales ante objetos diminutos frente a la eficacia demostrada por las redes convolucionales. Zhao *et al.* (2024a) advierten de esta carencia al señalar que, si bien su modelo RT-DETR alcanza niveles óptimos de velocidad y precisión, hereda el déficit de la familia DETR para la localización de elementos a pequeña escala respecto a los detectores más avanzados del momento. En este contexto, la incorporación de RF-DETR al estudio permite evaluar si su mayor comprensión del contexto visual global logra mitigar dicha deficiencia al operar en entornos epigráficos de alta incertidumbre, marcados por la degradación de la piedra o el bajo contraste de los relieves.

Para realizar esta comparativa, se ha diseñado una estrategia de entrenamiento jerárquica dividida en fases consecutivas. El proceso comienza con una etapa inicial centrada en la categoría de las aves (clase G de Gardiner), un grupo que constituye el 8,24% del corpus analizado e incluye signos

esenciales como el halcón (G7) o la lechuza (G17). La elección de este conjunto como caso de estudio inicial se debe a su morfología distintiva. Esta característica agiliza el etiquetado manual y facilita la transferencia de aprendizaje desde modelos preentrenados sobre MS COCO. Además, la orientación de estas figuras determina el sentido de lectura, una cualidad que, si bien no se explota en este trabajo, puede ser determinante en una fase futura de traducción.

Una vez validada la viabilidad de entrenar un modelo a partir de una categoría, se busca ampliar la cobertura mediante la integración de algunos de los glifos de mayor frecuencia: X1, D21, N35 y M17. Donde se excluye deliberadamente al glifo Z1 por su morfología extremadamente simple que requiere un tratamiento especializado diferente al del resto de jeroglíficos. El conjunto resultante, denominado clase «*Top 4*», representa el 19,18% de la cobertura total y ejerce un papel vertebrador en la oración como preposiciones, sufijos o fonogramas básicos. Al tratarse de clases con alta disponibilidad de muestras, se garantiza la robustez de los clasificadores.

La combinación de ambos subsistemas expertos permite localizar casi un tercio de los glifos de cualquier estela. Con el fin de explotar este conocimiento en un modelo genérico, se genera un *dataset* unificado donde el resto de los jeroglíficos no integrados en las fases previas se agrupan bajo la etiqueta de la clase «999» (desconocido). El modelo resultante utiliza las detecciones de alta confianza de las fases anteriores como puntos de anclaje contextuales, lo que permite a la red para inferir la presencia de caracteres desconocidos o no evaluados previamente en nuevas inscripciones.

3.3 Clasificadores y aumento de datos

En la aplicación de clasificadores a la epigrafía, el rendimiento de los modelos de aprendizaje profundo se ve limitado por la degradación natural del soporte, la similitud morfológica entre signos de una misma categoría (Figura 2) y la escasez de muestras para conformar el *dataset*. En el conjunto de referencia de Fuentes-Ferrer (2025a), el subconjunto de aves presenta un fuerte desequilibrio en ciertas clases de alta frecuencia que son indispensables para el entrenamiento. Aunque la solución óptima pasaría por ampliar manualmente estas clases deficitarias con imágenes reales, el experimento propuesto en la sección 4.6 tiene como objetivo evaluar si es posible obtener resultados aceptables mediante técnicas de aumento de datos, determinando el límite inferior de muestras necesarias y las opciones de agrupación.

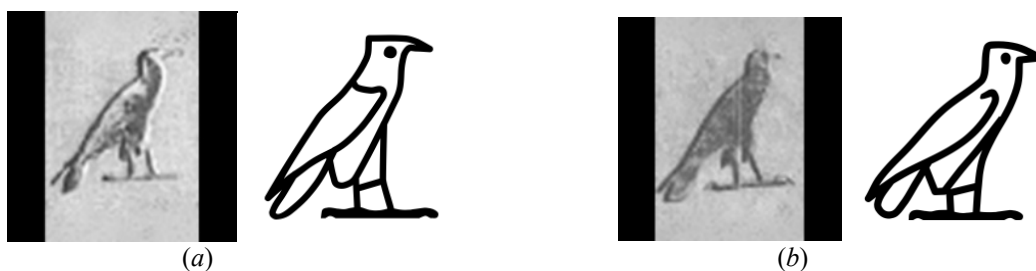


Figura 2. Similitud Morfológica: (a) G1 (Alimoche), Fuente NewGardiner G1; (b) G4 (Busardo), Fuente NewGardiner G4

La similitud intraclase y el desequilibrio en las categorías más frecuentes son dos problemas inseparables que marcan la concepción del sistema. Tal como se expone en la sección de resultados, el modelo de aves se utiliza como base para comprobar si es posible entrenar un sistema fiable a partir de conjuntos de datos muy desbalanceados. Esta misma técnica de compensación para estabilizar el número de muestras se aplica también al clasificador de «*Top 4*» y al genérico multiclase.

Para mitigar la falta de muestras, se valoró la implementación de técnicas avanzadas de síntesis de imágenes. En un contexto similar, Creed (2025) demostró la eficacia del NST para transferir la textura de soportes históricos a glifos sintéticos. Sin embargo, en esta investigación se evaluaron arquitecturas con mayor capacidad generativa, como los modelos de difusión y las Redes Generativas Antagónicas (GANs). Se descartaron los modelos de difusión debido a su tendencia a introducir artefactos imaginarios o alucinaciones visuales. Dado que en la epigrafía egipcia la precisión morfológica del trazo es un factor crítico, la aplicación de arquitecturas basadas en GANs condicionales y BicycleGAN ofrece mayores garantías de fidelidad estructural y control sobre el soporte, un nivel que los modelos de difusión aún no han consolidado.

En este trabajo fin de máster se ha implementado un flujo de aumento de datos multietapa fundamentado en redes generativas antagónicas. La generación combina el uso de cGANs, como Pix2pix y CycleGAN, para generar diversas siluetas en combinación con técnicas de visión clásica. Una red convolucional EfficientNet, evalúa las características visuales de la imagen para determinar el generador más adecuado. La reconstrucción y generación de imágenes variadas se realiza mediante el modelo BicycleGAN, un algoritmo con capacidad para forzar la variabilidad en los resultados y evitar el habitual colapso de modo que sufren los generadores tradicionales. Trabajos como FontGAN (Liu *et al.* 2019), enfocados en estilización y desestilización, demuestran la importancia de preservar el estilo visual.

La investigación experimental desarrollada en este trabajo se estructuró en torno a la integración de diferentes modelos generativos. Inicialmente se emplearon Pix2Pix (Isola *et al.*, 2017), una GAN condicional diseñada para la traducción de imagen a imagen, y CycleGAN (Zhu, Park, *et al.*, 2017), que permite la traducción entre dominios sin necesidad de pares alineados. Sin embargo, estas arquitecturas no lograban aportar variabilidad en la reconstrucción de las imágenes desde el boceto hasta el glifo, por lo que se integró BicycleGAN (Zhu *et al.*, 2017). Esta variante permitió solventar dicha limitación, superando problemas críticos de sus predecesoras y aportando la diversidad necesaria en las imágenes resultantes.

La combinación de estas técnicas da lugar a un *pipeline* completo de aumento de datos, diseñado para extraer siluetas a partir de imágenes reales y, posteriormente, generar un conjunto variado de imágenes sintéticas utilizando modelos BicycleGAN. En la Figura 3 se detalla el flujo completo del generador.

El proceso se inicia tomando como punto de partida el directorio *1_glifos*, donde se almacenan los glifos originales seleccionados del conjunto de datos (Fuentes-Ferrer, 2025a). Todo el conjunto original fue procesado manualmente para asegurar que las imágenes mantuvieran la misma orientación, facilitando el control del volteo (*flip*) horizontal. A continuación, las imágenes son evaluadas por una CNN basada en EfficientNet, entrenada previamente con un subconjunto de 3.500 imágenes etiquetadas (del cual se descartaron unas 800 por no presentar un relieve claro). Esta red clasifica automáticamente qué generador (G1, G2 o G3) debe devolver la silueta más acertada según el tipo de relieve, almacenando los resultados en la carpeta *2_clasificados*.

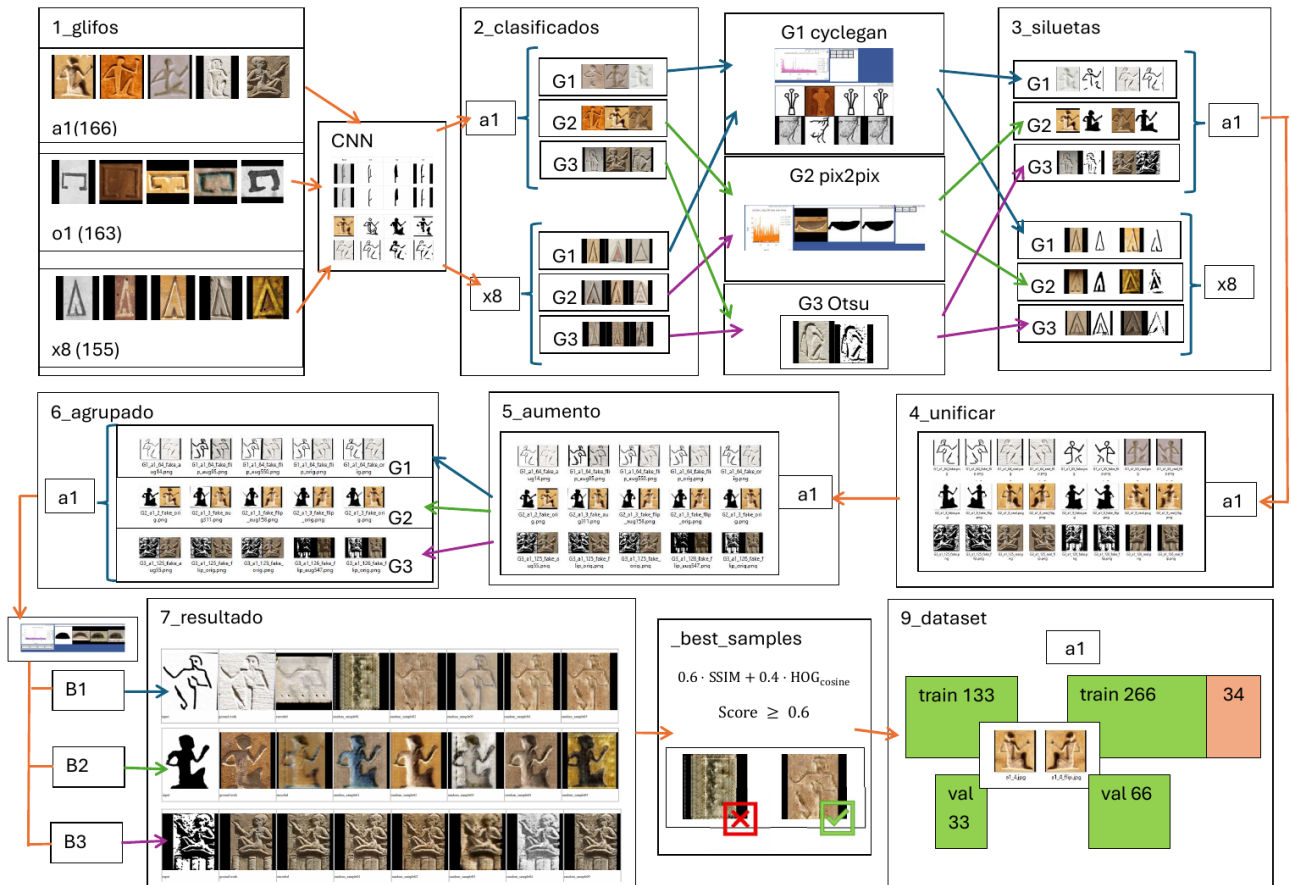


Figura 3. Pipeline Aumento de Datos CGAN+

Las imágenes categorizadas son procesadas por el generador de siluetas correspondiente, almacenando los resultados en *3_siluetas*. Tras el procesamiento completo del *dataset*, la distribución fue la siguiente:

- Generador 1 (G1): Basado en CycleGAN y entrenado con pares de representaciones ideales (fuente NewGardiner de la University of St Andrews, s. f.) frente a reales. Está especializado en grabado fino (líneas). Procesó un total de 3.000 imágenes.
- Generador 2 (G2): Basado en Pix2Pix y especializado en bajo relieve (o hueco relieve). Para obtener las siluetas con la precisión necesaria mediante ajuste fino, se utilizaron las siluetas generadas previamente por G1 como máscaras centroides en SAM2. Procesó un total de 6.500 imágenes.
- Generador 3 (G3): Basado en un algoritmo de Otsu +10% con erosión. Resulta especialmente efectivo para alto relieve, donde los modelos anteriores presentaban dificultades. Procesó un total de 3.000 imágenes.

Las siluetas se trasladan a una carpeta común (*4_unificar*) para facilitar la introducción de variaciones morfológicas. Sobre cada una de ellas se generan 5 nuevas siluetas aplicando transformaciones aleatorias (aumento y disminución del grosor de línea, rotación ligera del 5% y volteo horizontal), guardando los resultados en *5_aumento*. Seguidamente, los datos se vuelven a separar en la carpeta *6_agrupado* respetando sus categorías originales (G1, G2, G3).

Las siluetas aumentadas son introducidas en tres modelos BicycleGAN (cada uno especializado en uno de los tipos de relieve) para aplicar la textura y generar la imagen final, almacenándose en 7 *resultados*. A estos resultados, se le aplica un filtro basado en una métrica de similitud para discriminar imágenes que no se ajustan a la estructura esperada, cada imagen se compara con su *ground truth* y se obtiene un score formado por un 60% de peso para el Índice de Similitud Estructural (SSIM) y un 40% para la similitud del coseno de los descriptores HOG (*Histogram of Oriented Gradients*). Solo se seleccionan las muestras que superan un umbral de 0,6.

Una vez consolidado el pipeline de generación de imágenes mediante BicycleGAN, el siguiente paso fue determinar la proporción óptima de datos sintéticos a introducir en el conjunto de entrenamiento. Considerando investigaciones previas sobre el entrenamiento de clasificadores bajo desbalance extremo Kang *et al.* (2020) y Xiong *et al.* (2026), y para evitar la saturación del espacio de características que distorsiona las fronteras de decisión, se determinó que la proporción de datos artificiales incorporada debía restringirse al mínimo indispensable.

En el último paso, se combinan las imágenes validadas hasta alcanzar el objetivo definido por clase. Se aplica un *flip* a la carpeta de glifos originales para reservar el conjunto de validación (compuesto exclusivamente por muestras reales). El conjunto de entrenamiento se completa con las imágenes sintéticas, añadiendo además un margen del 10% extra de muestras sintéticas para permitir una revisión manual y un ajuste final de las imágenes.

Los *dataset* generados se procesan con ConvNext. Esta decisión de utilizar ConvNeXt como backbone para el clasificador de grano fino busca utilizar la misma red base validada por la investigación de Fuentes-Ferrer *et al.* (2025b), que garantiza que las métricas de rendimiento deriven exclusivamente de la capacidad de detección (YOLO11 vs. RF-DETR) y no de discrepancias en la calidad del extractor de características.

El conjunto de modelos clasificadores entrenados basados en la arquitectura ConvNeXt, están diseñados para operar sobre subconjuntos de clases específicos. Este conjunto está compuesto por tres redes especializadas.

El primer modelo, está centrado en la categoría de aves, utilizando una estrategia condicionada por la frecuencia de aparición. El fuerte desequilibrio entre clases disponibles y frecuentes necesarias para el entrenamiento exige la aplicación de técnicas de aumento de datos avanzadas. Para garantizar la estabilidad del aprendizaje, el conjunto de entrenamiento se ha balanceado a 300 muestras por categoría. Este modelo discrimina signos específicos (como, por ejemplo, G17 o G1) y agrupa bajo una clase residual («*RESTO_AVES*») a aquellos glifos minoritarios que carecen del volumen de imágenes suficiente para constituir una categoría independiente.

El segundo modelo, denominado «*Top 4*», se especializa en la identificación de los signos de mayor frecuencia (X1, D21, N35 y M17). Para detectar falsos positivos del modelo localizador, se incorpora una clase de desambiguación («*OTROS*»), conformada por glifos que presentan similitudes morfológicas capaces de inducir falsos positivos, como ocurre entre la morfología del signo D4 y la del D21. Como en estas clases no existe déficit de muestras, el conjunto de entrenamiento se equilibra al alza fijando un umbral de 900 imágenes por clase, aplicando técnicas de generación sintética solo en aquellas categorías que requieren un ajuste menor.

El tercer modelo («*999*») utiliza las clases restantes del *dataset* de Fuentes-Ferrer (2025a) que superan un umbral de 150 muestras reales, 19 clases en total. Estas clases se equilibran en sus respectivos

conjuntos hasta alcanzar las 300 imágenes mediante un aumento de datos moderado. A diferencia del «Top 4», esta red carece de una categoría específica de desambiguación, ya que su propósito es forzar una clasificación por proximidad morfológica. De este modo, si los signos A1, A2 y B4 comparten el patrón visual de un hombre sentado, es de esperar que el modelo posea la capacidad de generalización suficiente para que, aun reconociendo solo la clase principal A1, logre asociar correctamente las variantes morfológicas A2 y B4 a dicha categoría. En cualquier caso, en el desarrollo de nuevos *datasets* expertos, las clases no reconocidas que generan confusión serán candidatas a conformar la clase negativa del modelo.

3.4 Arquitectura del sistema

3.4.1 Fase de localización

Mediante la implementación de un sistema de ventanas deslizantes y el ajuste de la resolución de cada fragmento a la resolución requerida por cada conjunto de modelos: 1008×1008 píxeles para RF-DETR (utiliza resoluciones múltiplos de 56) y 1024x1024 píxeles para YOLO11. El proceso jerárquico propuesto se desarrolla tal y como se ilustra en la Figura 4. En primer lugar, se identifican los signos pertenecientes a la clase «Aves» con un nivel de certeza elevado (esto es, umbral $\geq 0,90$ y resaltados en color rojo en la Figura 4), complementando la inferencia con modelos secundarios configurados para capturar elementos de la clase «Aves» en umbrales medios (25%–40%, en color magenta) y glifos de la clase «Top 4» en umbrales bajos ($\leq 25\%$, en color azul). Para resolver el solapamiento generado por las ventanas deslizantes y los múltiples modelos, se aplicó un algoritmo NMS personalizado y basado en prioridades. Esto permitió generar un conjunto de recortes fiables donde los signos restantes correspondientes a la clase «999» se resaltan en color verde en la Figura 4.

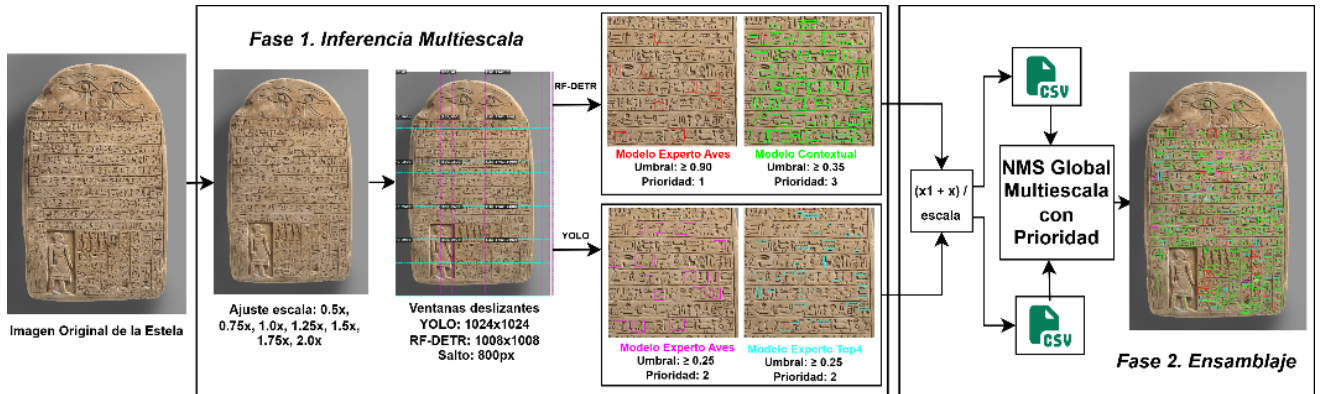


Figura 4. Proceso híbrido de localización. Estela de Horemkhauef.

Aunque la revisión del estado del arte destaca que el principal avance arquitectónico de modelos como RT-DETR es la eliminación nativa del algoritmo NMS, su reintroducción en este trabajo resulta estrictamente necesaria. Esta decisión técnica se debe a dos factores fundamentales del sistema propuesto: la fragmentación de las imágenes de alta resolución generada por las ventanas deslizantes y la integración de las inferencias conjuntas procedentes de los modelos YOLO11 y RF-DETR. Ambas técnicas generan solapamientos que obligan a aplicar un filtrado espacial global para unificar las detecciones. Al combinar múltiples escalas, el algoritmo resultante gana en complejidad, ya que su implementación se concreta en el script que se describe a continuación:

- Para integrar correctamente ambos modelos predictivos $M = \{M_{RF}, M_{YOLO}\}$, se define un conjunto discreto de escalas $S = \{0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 2.25, 2.5, 2.75, 3.0\}$. Al operar

sobre recortes locales, resulta imprescindible proyectar cada predicción al espacio de coordenadas absoluto de la imagen original. Este ajuste se realiza sumando el desplazamiento espacial (x, y) a la caja delimitadora (B_{local}) y dividiendo el resultado por el factor de escala s correspondiente:

$$B_{orig} = \frac{B_{local} + (x, y)}{s} \quad (1)$$

- Mediante una agregación global, se unifican todas las detecciones generadas del conjunto de detecciones brutas, denotado como $D_{raw}(s)$, en todas las escalas de trabajo S :

$$D_{global} = \bigcup_{s \in S} D_{raw}(s) \quad (2)$$

- Sobre este conjunto que agrupa alto nivel de ruido y redundancias, se aplica una supresión global unificada. Esto implica un único filtrado NMS para extraer el vector definitivo de detecciones:

$$D_{final} = \text{NMS}(D_{global}, 0.25) \quad (3)$$

El algoritmo NMS se utiliza en el filtrado espacial. Buscar una solución analítica exacta supone enfrentarse a un problema NP-Hard de complejidad exponencial $O(2^N)$, que paralizaría el sistema. Este método opta por una vía algorítmica más pragmática. El proceso ordena los elementos por prioridad, con una complejidad $O(N \log N)$, y descarta de forma iterativa aquellas cajas cuya superposición supere el umbral de 0,25 de IoU. Mediante este enfoque, la fase de filtrado posterior reduce el tiempo de procesamiento a un escenario prácticamente lineal $O(N)$.

El postprocesamiento NMS multiescala resulta lo bastante eficiente para ejecutarse en una CPU estándar en un tiempo razonable. Sin embargo, la gran cantidad de elementos a procesar en diferentes escalas aumenta tanto el número de artefactos decorativos detectados como el número de objetos identificados como glifo prioritario, lo que genera un ruido innecesario.

Es por esto por lo que se propone una mejora optimizada del script anterior, que intenta maximizar la precisión y minimizar el ruido de decoración y cajas fantasma utilizando los datos de detecciones de confianza como referencia para identificar la escala de mejor rendimiento. El script se modifica:

- En lugar de realizar una agregación global, iterar sobre toda la imagen para generar un conjunto de detecciones brutas para cada escala analizada, denotado como $D_{raw}(s)$. Aplicando un algoritmo NMS sobre cada nivel de escala, se descartan aquellas cajas cuyo nivel de solapamiento supera un umbral de 0.25 al compararse con detecciones de mayor prioridad:

$$D_{clean}(s) = \text{NMS}(D_{raw}(s), 0.25) \quad (4)$$

- Se contabilizan tanto las detecciones aportadas por el modelo RF-DETR «Aves» (asignado con prioridad 1) como las aportadas por YOLO «Aves» + «Top 4» (prioridad 2) en cada escala s :

$$C_{RF}(s) = |\{d \in D_{clean}(s) \mid \text{prio}(d) = 1\}| \quad (5)$$

$$C_{YOLO}(s) = |\{d \in D_{clean}(s) \mid \text{prio}(d) = 2\}| \quad (6)$$

- Se identifica la escala óptima (s_{opt}) en la que cada arquitectura que maximiza su volumen de detecciones válidas. Identificadas las resoluciones de mayor rendimiento, se unifican en un único conjunto global. La operación de unión descarta las redundancias en caso de que ambos modelos alcancen su máximo en el mismo factor de escala, obteniendo un vector que agrupa exclusivamente las predicciones limpias:

$$D_{global} = \bigcup_{s \in S_{final}} D_{clean}(s) \quad (7)$$

- Para solventar posibles solapamientos entre las inferencias conjuntas de RF-DETR y YOLO, se aplica un último filtrado espacial sobre la agrupación global, consiguiendo así el vector definitivo de detecciones:

$$D_{final} = \text{NMS}(D_{global}, 0.25) \quad (8)$$

La elección entre la primera formulación y esta variante depende de las propiedades físicas del soporte analizado. Al restringir la extracción a las escalas de mejor rendimiento, el método prioriza la precisión y reduce los falsos positivos, siendo óptimo para imágenes estables o con un nivel de ruido controlado. Pero cuando el nivel de erosión es muy elevado, suele ser mejor opción no descartar ninguna escala de antemano, de este modo, el sistema maximiza la tasa de recuperación, capturando fragmentos que el método alternativo podría ignorar en primera instancia.

3.4.2 Fase de clasificación

Los datos obtenidos en la fase de localización contienen las coordenadas espaciales y una etiqueta preliminar asignada (*glifo_bird*, *glifo_top* o *999*), lo que nos permiten recortar y procesar la estela, como se muestra en la Figura 5. Antes de la inferencia, se ajusta el tamaño del recorte a 224 píxeles manteniendo la relación de aspecto original mediante interpolación bilineal (rellenando el espacio restante con *zero-padding*) y se normaliza según el estándar de ImageNet.

Para optimizar la precisión, se implementa una arquitectura de inferencia en cascada, si el modelo experto especializado en el «*Top 4*» clasifica un símbolo dentro de su clase de rechazo («*OTROS*»), el sistema transfiere el recorte al modelo «*999*» para una segunda opinión. De este modo, la clase predicha C_{final} para un recorte de imagen (x) en función de una etiqueta preliminar (l_i), obtenida durante la fase de localización, se obtiene de la siguiente manera:

$$C_{final} = \begin{cases} M_{birds}(x) & \text{si } l_i \in \{rf_bird, yo_bird\} \\ M_{top4}(x) & \text{si } l_i = yo_top4 \text{ y } M_{top4}(x) \neq \text{"OTROS"} \\ M_{999}(x) & \text{si } l_i = rf_999 \text{ o } (l_i = yo_top4 \text{ y } M_{top4}(x) = \text{"OTROS"}) \end{cases} \quad (9)$$

rf_bird: Detección del glifo RF-DETR «Aves», filtrada con un nivel de confianza ≥ 90 %.

yo_birds: Detección del glifo YOLO11 «Aves», filtrada con un nivel de confianza ≥ 25 %.

yo_top4: Detección del glifo YOLO11 «Top 4», filtrada con un nivel de confianza ≥ 25 %.

rf_bird: Detección del glifo RF-DETR «999», filtrada con un nivel de confianza ≥ 35 %.

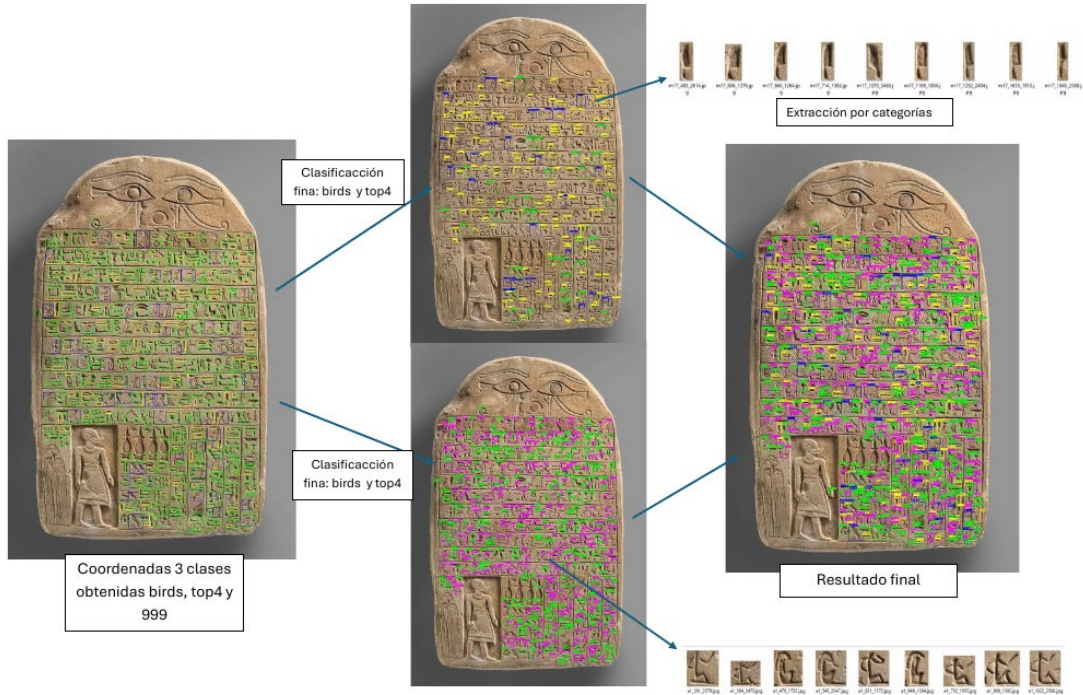


Figura 5. Fase de Clasificación

La red experta procesa la imagen (omitiendo el cálculo de gradientes para optimizar el consumo de memoria VRAM) y devuelve la clase predicha junto a su nivel de confianza. Para obtener esta métrica de certeza, el vector de características Z devuelto se transforma mediante la función Softmax. La confianza de la predicción (p_{max}) se corresponde con el valor máximo de dicha distribución sobre el total de K clases:

$$p_{max} = \max \text{Softmax}(Z) \quad (10)$$

Para reducir los falsos positivos, se aplican umbrales mínimos de confianza ajustados empíricamente: 0,60 para «Aves», 0,80 para el «Top 4» y 0.85 para «999». Si la predicción no supera su límite, la instancia se clasifica como dudosa.

Como salida complementaria, el sistema aísla cada recorte clasificado y lo almacena emparejado con su representación idealizada, con el objetivo de facilitar la construcción de nuevos *datasets*. Finalmente, el sistema exporta un nuevo archivo CSV con las etiquetas definitivas y los valores de confianza detallados y genera una imagen de validación visual, donde se dibujan cajas delimitadoras cuyo código de color indica si la detección ha superado o no los umbrales de confianza requeridos, quedando en verde las cajas que no han conseguido superar el umbral, pero mostrando en la etiqueta la predicción más cercana.

4 Resultados obtenidos

4.1 Métricas

Para evaluar y comparar el rendimiento tanto de las arquitecturas YOLO11 y RF-DETR, como de los modelos ConvNext, se ha empleado un conjunto de métricas estándar en la detección de objetos. Para determinar la mejor arquitectura para este problema se presentan los resultados junto a una comparativa YOLO11 vs RFDETR. Aunque los *frameworks* utilizados en este estudio (Ultralytics para YOLO11 y la implementación nativa para RF-DETR) proporcionan gráficas y métricas de evaluación autogeneradas durante el proceso de entrenamiento y validación, en esta investigación se ha decidido prescindir de su uso directo para el análisis comparativo final. Esto es debido a que cada *framework* actúa como una "caja negra" que aplica heurísticas dispares. La falta de consenso en las implementaciones de evaluación genera resultados distintos dependiendo del tipo de interpolación matemática utilizada o de cómo se resuelven los empates en los niveles de confianza (Padilla *et al.*, 2020).

Por lo que, para garantizar la transparencia y asegurar que los algoritmos se evalúan verdaderamente bajo las mismas reglas matemáticas (Padilla *et al.*, 2020), se ha procedido a extraer las predicciones en bruto (coordenadas absolutas de las cajas delimitadoras y niveles de confianza) generadas por ambos modelos sobre el conjunto de prueba. Mediante un script propio, se mide el rendimiento fijando el umbral IoU (Intersection over Union) en 0,5. El uso de este umbral obedece a las condiciones reales de las estelas, en las que los límites de cada jeroglífico son imprecisos y el etiquetado de referencia solo puede ser aproximado. Si se impusiera un ajuste espacial más rígido, el sistema rechazaría aciertos reales y la tasa de recuperación caería drásticamente. En cuanto al umbral de confianza (*confidence threshold*), debido a que cada arquitectura calibra su certeza de forma distinta, fue necesario analizar la relación precisión-exhaustividad (Recall) en diferentes valores para evitar penalizaciones por excesos de confianza inherentes a cada modelo y determinar el punto de funcionamiento óptimo para ambos casos. Sobre el conjunto de datos homogeneizado, se han calculado la Precisión, el Recall, el F1-score y el mAP@0.5 aplicando las ecuaciones (11) a (15) definidas en este apartado en concordancia con las formulaciones estándar de evaluación de detección de objetos.

Precisión: Mide la proporción de verdaderos positivos sobre el total de predicciones positivas en términos de cajas delimitadoras. Refleja la exactitud del modelo. Una mayor precisión significa menos falsos positivos, lo que indica que el modelo es más preciso al identificar glifos verdaderos

$$\text{Precisión} = \frac{TP}{(TP+FP)} \quad (11)$$

TP: Caja delimitadora correctamente estimada.

FP: Caja delimitadora incorrectamente estimada.

Recall (Tasa de verdaderos positivos): Mide la proporción de cajas delimitadoras estimadas correctamente entre todas las detecciones reales (*ground truths*). Un Recall más alto significa menos falsos negativos, lo que indica que el modelo puede detectar exitosamente la mayoría de los glifos presentes

$$\text{Recall} = \frac{TP}{(TP + FN)} \quad (12)$$

TP: Verdaderos positivos.

FN: Falsos negativos.

F1-score: Proporciona una evaluación equilibrada del rendimiento del modelo. Combina la Precisión y el Recall en un solo número, calculándose la media armónica de estos. Un F1-score más alto indica una mejor capacidad del modelo de detección para encontrar la ubicación del glifo.

$$\text{F1 - Score} = \frac{2 \times (\text{Precisión} \times \text{Recall})}{(\text{Precisión} + \text{Recall})} \quad (13)$$

mAP@0.5 (*mean Average Precision* con IoU = 0.5): Indica la precisión promedio del modelo a diferentes tasas de *Recall*. Se calcula como el promedio de las Precisiones Promedio (AP) de todas las categorías calculado utilizando un umbral de IoU de 0,5. Este umbral mide la superposición entre la caja delimitadora predicha y la caja delimitadora real (*ground truth*).

$$AP_i = \int_0^1 P_i(R) dR \quad (14)$$

AP_i (*Average Precision*): Es la integral del área bajo la curva de Precisión vs. Recall.

$P_i(R)$: Es la precisión como función del recall R para la clase i .

$$\text{mAP@0.5} = \frac{1}{N} \sum_{i=1}^N AP_i \quad (15)$$

N : número de clases.

En mAP@0.5, se considera correcta una predicción si el IoU ≥ 0.5 .

La evaluación de las detecciones se rige por el criterio de Intersección sobre Unión, fijando un umbral de aceptación del 50 %. Para evaluar correctamente las inferencias, las predicciones se ordenan por confianza descendente y se vinculan a la anotación de referencia (*ground truth*) que presente mayor IoU. Como cada objetivo real restringe su emparejamiento a una única predicción para evitar redundancias, cualquier intento de asociación posterior sobre una etiqueta ya cubierta se penaliza registrándose como Falso Positivo (FP). Las anotaciones no detectadas se añaden al recuento de Falsos Negativos (FN). A partir de la matriz de confusión resultante de este filtrado, se derivan las métricas finales de Precisión, Recall y la media armónica F1-score.

La clasificación de las detecciones y el cálculo de la Precisión Promedio (AP) siguen el estándar de evaluación de PASCAL VOC (Everingham *et al.*, 2015), aplicando la interpolación de todos los puntos descrita por Padilla *et al.* (2020). En el script de cálculo de métricas, en lugar de promediar un número fijo de intervalos, este método calcula el área exacta bajo la curva de Precisión-Recall. Para eliminar las oscilaciones propias de la clasificación, la curva original se suaviza trazando una envolvente monótona decreciente, es decir, a cada nivel de recall se le asigna la precisión máxima alcanzada en cualquier punto posterior, obteniendo finalmente el área total (AP) mediante la suma de los bloques rectangulares que conforman la nueva gráfica escalonada.

4.2 Configuración

Para la ejecución y aceleración del entrenamiento de los modelos, se utilizó una GPU NVIDIA GeForce RTX 5060 Ti con 16 GB de VRAM. Debido a las restricciones de memoria de vídeo (VRAM) en el entorno local y la diferente gestión de mapas de características entre arquitecturas, los tamaños de lote y las resoluciones se ajustaron para operar en el límite de saturación del hardware, asegurando la máxima preservación del detalle espacial necesario para la microdetección epigráfica, el tamaño de lote físico (*batch size*) máximo permitido fue de 8 imágenes por iteración en el caso de YOLO (véase Tabla 5). En el caso de RF-DETR, la arquitectura permite combinar el tamaño de lote físico con la

acumulación de gradientes. Su configuración le permitía un tamaño de lote de 4 y 4 pasos de acumulación (*grad_accum_steps*), lo que arrojaba un lote efectivo de 16. Para evitar que esta capacidad técnica supusiera una ventaja injusta frente al otro modelo, se ajustó el tamaño del lote, reduciendo la acumulación a 2, logrando un lote efectivo final de 8, para garantizar así una comparativa equitativa entre ambas.

Para adaptar las resoluciones de entrada al techo operativo de cada modelo, YOLO11l operó con configuraciones de 1024×1024 y 512×512 píxeles, alcanzando un coste computacional aproximado de 222,7 y 55,7 GFLOPs, respectivamente. Por el contrario, las restricciones arquitectónicas del backbone DINOv2 en RF-DETR S obligaron a encajar las dimensiones en múltiplos exactos de 56. Esto fijó sus resoluciones en 1008×1008 y 508×508 píxeles, con un consumo de 154,8 y 39,3 GFLOPs. Reconocer esta asimetría resulta indispensable para llevar ambas topologías al límite de su capacidad gráfica compartiendo las mismas restricciones de hardware.

Con el entorno homogeneizado, se ajustaron las herramientas de optimización para sacar el máximo rendimiento de cada *framework*. En el caso de YOLO11l, aunque se mantuvieron las heurísticas automáticas para el optimizador (AdamW), las fases de *warmup* y los aumentos por mosaico, se forzó explícitamente la configuración de un planificador de tasa de aprendizaje por coseno (*Cosine Annealing*) para suavizar la convergencia. Para RF-DETR, la convergencia se gestionó de forma íntegramente manual mediante una tasa de aprendizaje base de $1 \cdot 10^{-4}$, un factor multiplicador de 15 para el ajuste fino del codificador, un planificador por pasos (*Step Decay*) y una penalización de pesos (*Weight Decay*) adaptada a sus funciones de pérdida específicas.

Tabla 5. Comparativa de hiperparámetros YOLO11l vs RF-DETR (S)

Hiperparámetros	YOLO11l	RF-DETR (S)
Épocas máximas	200 / 30	200 / 30
Early Stopping (Paciencia)	30	30
Lote (Batch size) físico /Efectivo	8/8	4 x 2 acum. / 8
Resolución expertos	1024×1024	1008×1008
Resolución genéricos	512×512	508×508
Codificador (Backbone)	Propio	DINOv2 (ViT)
Número de Queries	—	300
Optimizador	AdamW (M 0,937)	Manual
Weight Decay	0,0005	1,0
Tasa de aprendizaje (LR)	0,01 (Auto)	$1 \cdot 10^{-4}$
Planificador	Coseno (Forzado)	Step Decay
Coeficientes de pérdida	—	L1: 5,0 GIoU: 2,0 Clasif: 1,0
Aumento de datos	Mosaico (100%) Volteo H (50%)	Interno

Para los clasificadores se emplea la arquitectura ConvNeXt de Meta (versión preentrenada en ImageNet-22K), disponible a través de la librería `timm` (`timm`, s. f.) e integrada mediante `torchvision` en `PyTorch`. En todos los casos se aplica aprendizaje por transferencia (*transfer learning*) partiendo de los pesos preentrenados en el conjunto de datos COCO. Con el objetivo de realizar una evaluación objetiva, se mantuvieron las configuraciones de hiperparámetros por defecto sugeridas por los autores de cada arquitectura, limitando los ajustes únicamente a las restricciones de memoria del hardware.

4.3 Evaluación de modelos expertos

Para llevar a cabo la comparativa inicial de rendimiento entre YOLO11 y RF-DETR, se construyeron dos *datasets* específicos con imágenes en alta resolución (1024×1024 píxeles) priorizando la máxima diversidad visual posible. El *dataset* de «*Aves*» cuenta con 285 fragmentos procedentes de 115 estelas, mientras que el «*Top 4*» agrupa cuatro signos morfológicamente heterogéneos (X1, D21, N35 y M17) en 233 fragmentos de 98 estelas. Los datos se dividieron en tres grupos disjuntos en una proporción de 70, 20 y 10, correspondientes a los conjuntos de entrenamiento, validación y prueba, respectivamente. La separación no se ha realizado de forma aleatoria. Las imágenes del *dataset* son tiles que pueden provenir de la misma fotografía o estela original, una asignación aleatoria correría el riesgo de sufrir contaminación de datos (*Data Leakage*). El modelo podría memorizar la textura, la iluminación o el patrón de ruido de una estela durante el entrenamiento y, posteriormente, ser evaluado sobre un fragmento adyacente de esa misma estela en los conjuntos de validación o prueba, inflando artificialmente las métricas de rendimiento.

Para evitarlo y garantizar una evaluación realista de la capacidad de generalización de ambos modelos, se ha aplicado una separación a nivel de imagen, garantizado que ningún tile perteneciente a una estela utilizada en el conjunto de entrenamiento forme parte de los conjuntos de validación o test (véase Apéndice A). Además, para asegurar que la comparativa sea justa, se ha utilizado exactamente la misma distribución de imágenes para ambos modelos. Se han generado dos *datasets* simétricos, en los cuales el único parámetro que varía es el formato de las anotaciones, adaptándose a los requisitos de cada *framework*: YOLO con archivos de texto plano con coordenadas normalizadas para YOLO11 y archivo formato COCO en json para RF-DETR.

La Figura 6 muestra las gráficas de validación, en donde se observa una rápida convergencia de RF-DETR durante las primeras épocas. En cambio, al intentar agrupar formas morfológicamente distintas, situación que se da en el conjunto «*Top4*», el aprendizaje se estanca de forma prematura. YOLO comienza con una alta inestabilidad, pero acaba consolidando mejor los resultados a medida que avanza el entrenamiento. De este modo, logra detectar las clases concretas con mayor fiabilidad, obteniendo un $mAP@0.5$ de 0,8769, superando el 0,8169 de RF-DETR.

La Tabla 6 revela que RF-DETR es sumamente dependiente del umbral de confianza seleccionado. Un ejemplo claro se observa en la categoría «*Aves*», donde aumentando el umbral a 0,50 mejora la precisión a 0,78 a expensas de sacrificar la capacidad de recuperación. Aunque YOLO11 presenta un comportamiento mucho más estable en configuraciones estándar, RF-DETR demuestra una superioridad inesperada ante una exigencia de confianza del 0,90 y es capaz de recuperar un 40 % de los glifos con total seguridad, mientras que YOLO11 cae hasta el 2 %. Esta ventaja desaparece en «*Top4*», donde el rendimiento de RF-DETR sufre un descenso drástico, como se muestra en la Tabla 7. Los resultados del F1-score confirman que YOLO11 ofrece el mejor rendimiento para las clases específicas. Al fijar un umbral de 0,5, esta red supera claramente a RF-DETR en la categoría «*Aves*», logrando un resultado de 0,8557 frente a 0,7761. Esta superioridad se repite en el grupo «*Top4*», donde alcanza un 0,8773 frente a 0,7793, lo que nos indica que la arquitectura convolucional se consolida como la herramienta más precisa para aislar los glifos frecuentes.

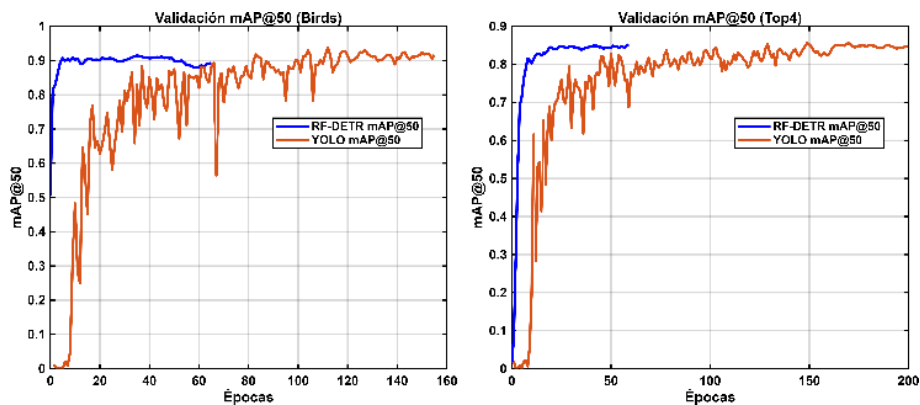


Figura 6. Comparativa Validación mAP@50 (clases «Aves» y «Top4»).

Tabla 6. Comparativa de métricas de YOLO11 vs RF-DETR (Clase «Aves»)

Modelo	Conf.	Precisión	Recall	F1 Score
YOLO11	0,25	0,8598	0,9109	0,8846
RF-DETR	0,25	0,6496	0,8812	0,7479
YOLO11	0,5	0,8600	0,8515	0,8557
RF-DETR	0,5	0,7800	0,7723	0,7761
YOLO11	0,9	1,0000	0,0198	0,0388
RF-DETR	0,9	1,0000	0,4059	0,5775

Tabla 7. Comparativa de métricas de YOLO11 vs RF-DETR (Clase «Top 4»)

Modelo	Conf.	Precisión	Recall	F1 Score
YOLO11	0,25	0,9077	0,8762	0,8917
RF-DETR	0,25	0,6968	0,8193	0,7531
YOLO11	0,5	0,9282	0,8317	0,8773
RF-DETR	0,5	0,8667	0,7079	0,7793
YOLO11	0,9	1,0000	0,0050	0,0099
RF-DETR	0,9	1,0000	0,0842	0,1553

El análisis de las matrices de confusión (véase Figuras 7 y 8) respalda esta diferencia de rendimiento. Los datos demuestran que YOLO11 es mucho más preciso aislando los glifos reales. Al comparar las métricas, consigue un 85 % de acierto en «Aves» y un 83 % en «Top 4», dejando atrás a RF-DETR (77 % y 71 %). Esta superioridad probablemente se debe a la capacidad del modelo para discriminar el fondo del soporte, evitando confundir el ruido o la erosión con caracteres epigráficos. Como consecuencia, las falsas detecciones descienden al 14 % en las aves y apenas al 7 % en los signos más frecuentes, confirmando la capacidad de generalización de esta arquitectura frente a estelas muy deterioradas. Además, RF-DETR demuestra ser algorítmicamente más conservador, descartando de forma nativa cualquier inferencia con menos del 50 % de confianza, el modelo ignora directamente los glifos dudosos que YOLO11 sí logra recuperar.

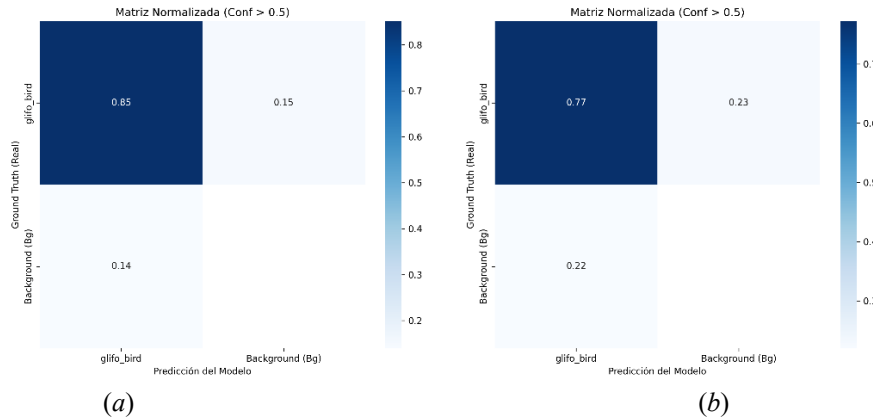


Figura 7. Matrices de confusión normalizadas del modelo «Aves» sobre el conjunto de test. Comparativa entre los resultados de: (a) YOLO11 y (b) RF-DETR.

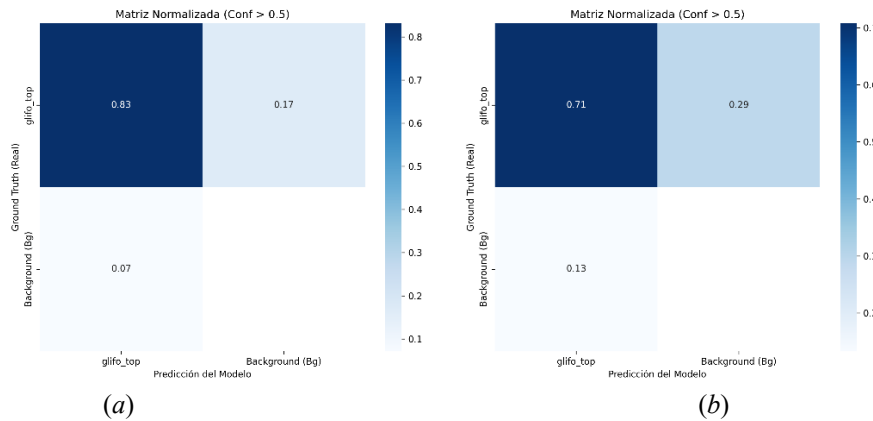


Figura 8. Matrices de confusión normalizadas del modelo «Top 4» sobre el conjunto de test. Comparativa entre los resultados de: (a) YOLO11 y (b) RF-DETR.

El análisis cualitativo sobre el conjunto de prueba confirma las métricas obtenidas y revela comportamientos diferenciados relevantes para la práctica epigráfica. La inspección visual de los resultados (Figuras 9 y 10) ilustra dos comportamientos bien diferenciados. En estelas con alta densidad de caracteres, YOLO11 tiende a saturar la imagen de detecciones para asegurar la cobertura. Al rebajar el umbral de confianza a 0.25, el algoritmo dispara su tasa de localización y logra capturar multitud de candidatos, incluyendo glifos fragmentados. Por su parte, la respuesta de RF-DETR resulta mucho más selectiva a la erosión. Aunque recupere un menor volumen total, aísla los caracteres severamente desgastados conservando niveles de confianza superiores a los de YOLO11. En la Figura 10, se aprecia que, ante situaciones de complejidad similar, ambos modelos presentan un comportamiento casi equivalente.

Al aumentar a 0,50 el umbral, YOLO experimenta una caída notable en su fiabilidad frente a jeroglíficos erosionados o en contextos de borrosidad. Frente a esto, RF-DETR muestra una mayor capacidad de inferencia, conservando un grado de certeza elevado incluso cuando la calidad visual de la estela es baja, como muestra la Figura 11 (imagen no profesional).

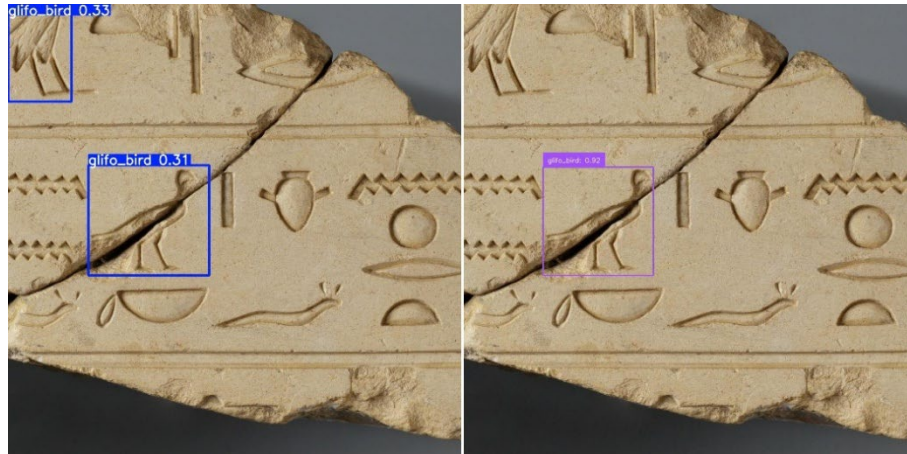


Figura 9. Comparativa entre los resultados de detección del modelo «Aves» YOLO11 (izquierda) y RF-DETR (derecha) con umbral de confianza al 0,25. Adaptado de Fragment of Henenu Stela. The Metropolitan Museum of Art (The Met).

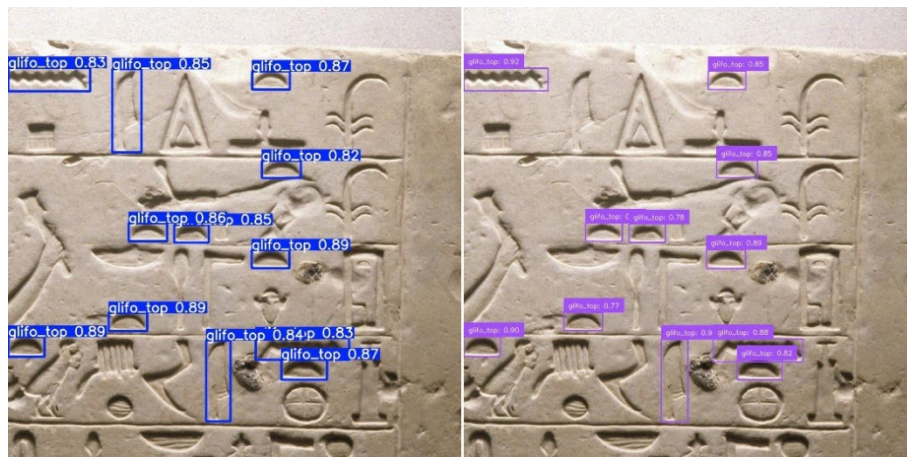


Figura 10. Comparativa entre los resultados de detección de modelo «Top 4». YOLO11 (izquierda) y RF-DETR (derecha) con umbral de confianza al 0,25. Adaptado de Tomb Stela of Hetepsi. The Met.

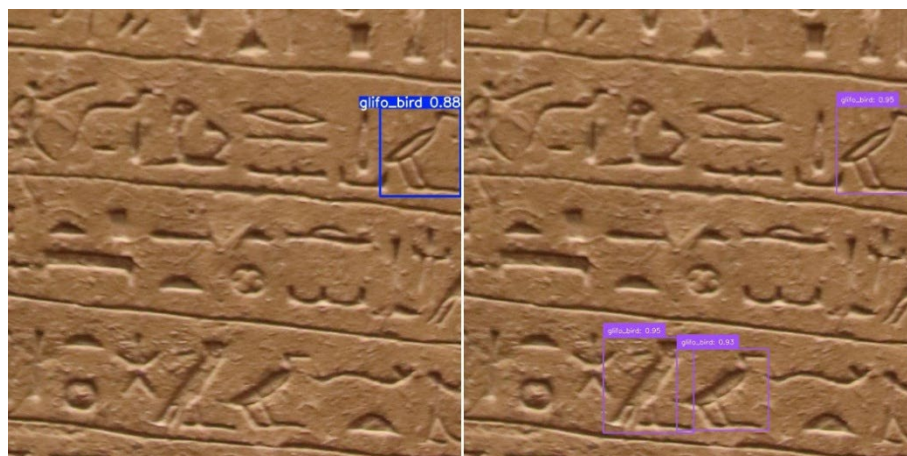


Figura 11. Comparativa entre los resultados de detección de modelo «Aves». YOLO11 (izquierda) y RF-DETR (derecha) con umbral de confianza al 0,50. Adaptado de Estela funeraria del sacerdote Gem. por Tm, 2021, Flickr.

4.4 Evaluación del modelo genérico

Una vez localizados los glifos de alta frecuencia, el esfuerzo se dirigió a detectar el resto del corpus aplicando un etiquetado genérico. Con este propósito, se conformó un conjunto de datos con 53 recortes en alta resolución de 1024×1024 píxeles, procedentes de 14 estelas. Durante este aprendizaje, las categorías «Aves» y «Top 4» se mantuvieron para que actuaran como anclajes visuales durante el aprendizaje, agrupando el resto de los glifos no identificados bajo la etiqueta genérica «999». Sobre este conjunto, primero se aplicó una clasificación de los símbolos «Top 4» (véase 4.7) y, después, un proceso de reetiquetado (*relabeling*) con las posiciones exactas de los signos pertenecientes a «Aves» y los jeroglíficos individuales que conforman «Top 4», realizando recortes de 512×512 píxeles centrados en cada glifo detectado por los modelos expertos. Para duplicar el volumen de la muestra sin distorsionar la información semántica, se empleó un volteo horizontal, consiguiendo un total de 1168 imágenes de entrenamiento. Debido a la alta variabilidad morfológica de la clase «999», el rendimiento de YOLO11 disminuye notablemente al no identificar patrones comunes ni definir una textura homogénea. Esta dificultad provoca una curva de aprendizaje inestable, compromete la minimización del error de localización espacial (*box loss*) y causa una parada prematura del entrenamiento (Figura 12).

RF-DETR ofrece, en cambio, resultados muy superiores alcanzando un $\text{mAP}@0.5$ de 0,8596 frente al 0,6301 de YOLO11. Este comportamiento indica que la arquitectura transformer no requiere memorizar la morfología exacta de los cientos de símbolos que componen esta clase híbrida. En su lugar, infiere su ubicación apoyándose en la estructura espacial de los anclajes lingüísticos previos («Aves» y «Top 4»).

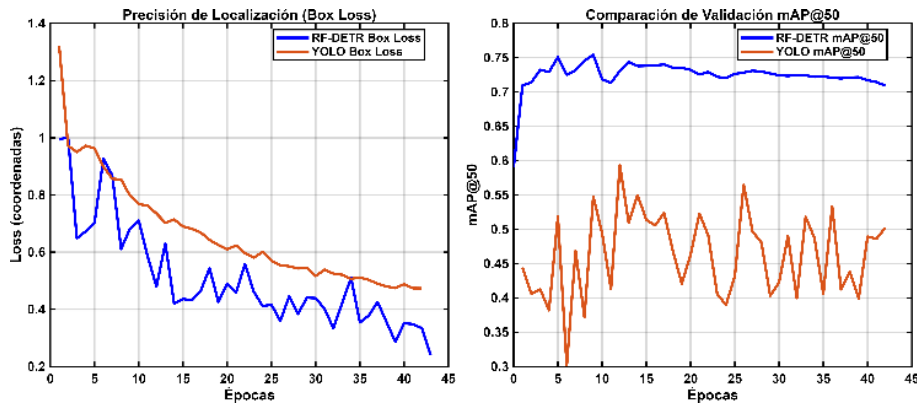


Figura 12. Comparativa entrenamiento YOLO vs RF-DETR (clase «999»).

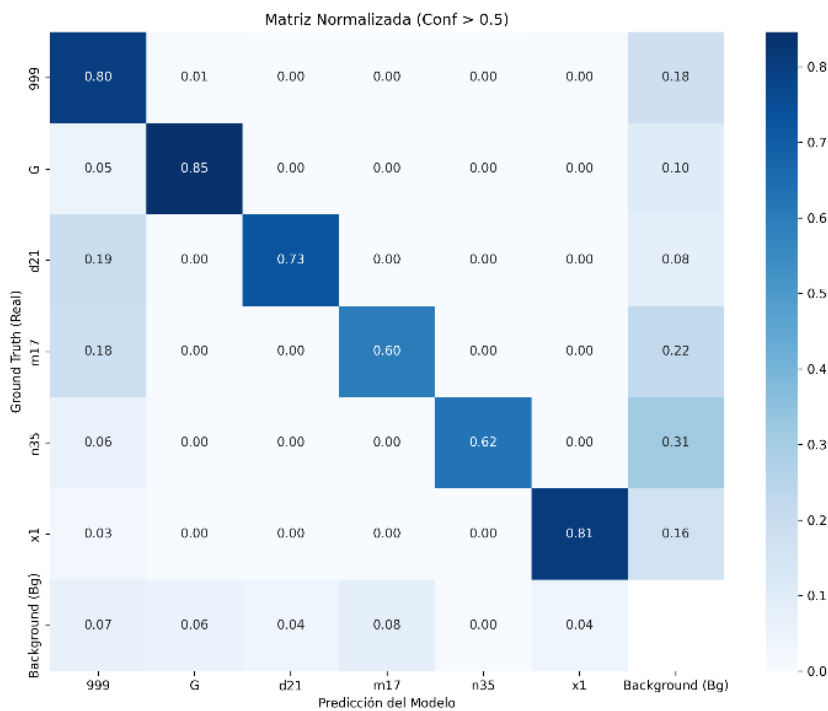
Las métricas de la Tabla 8 reflejan una inversión en la tendencia observada en los experimentos previos (Sección 4.3). Mientras que YOLO11 mantenía la estabilidad con clases específicas, su capacidad de respuesta se reduce significativamente ante el modelo genérico. En cambio, RF-DETR conserva su capacidad de recuperación incluso en umbrales de máxima confianza (0,9), donde logra rescatar cerca del 25 % de los glifos (recall de 0,2453).

Tal y como refleja la matriz de confusión (Figura 13), el modelo RF-DETR muestra una precisión moderada en glifos específicos como M17 (60%) o N35 (62%), tendiendo a confundirlos con el fondo o a relegarlos hacia la categoría genérica «999». Este comportamiento no representa un fallo del sistema, ya que este modelo no se concibió como un clasificador exacto, sino como un localizador espacial. Aunque la inclusión de las clases «Aves» y «Top 4» sirve para aportar contexto geométrico a la inferencia, la utilidad real de esta red reside en su alta tasa de recuperación general, ya que consigue aislar el 80 % de los caracteres bajo la etiqueta «999».

Tabla 8. Comparativa de métricas de YOLO11 vs RF-DETR (Clase «999»)

Modelo	Conf.	Precisión	Recall	F1-score
YOLO11	0,25	0,6520	0,6761	0,6639
RF-DETR	0,25	0,7405	0,8913	0,8090
YOLO11	0,5	0,7797	0,5730	0,6606
RF-DETR	0,5	0,9114	0,7994	0,8517
YOLO11	0,9	1,0000	0,0722	0,1346
RF-DETR	0,9	0,9983	0,2453	0,3938

Al revisar cualitativamente los resultados en el conjunto de prueba, se observan claramente los resultados esperados. En la Figura 14 se observa un recorte centrado en el signo D21 que contiene, además, el jeroglífico D36 (catalogado entre los signos frecuentes de la Tabla 4). Mientras que YOLO11 no logra localizar este glifo desconocido, RF-DETR lo detecta y lo delimita con confianza elevada.

**Figura 13.** Matriz de confusión RF-DETR (clase «999»).

4.5 Resultados del proceso de detección

En vista del rendimiento obtenido por cada uno de los modelos por separado, y donde cada uno destacó en determinados escenarios, se planteó el enfoque combinado propuesto en este trabajo. El modelo híbrido YOLO11/RF-DETR, ajustado bajo sus configuraciones óptimas, mejoró la fiabilidad en la detección. El generar una preclasificación en distintos niveles de precisión, facilita la aplicación posterior de modelos expertos sobre los recortes obtenidos. La Tabla 9 detalla los resultados de aplicar NMS utilizando máximos sobre un conjunto representativo de estelas. En las columnas P1 y P2 se especifican las escalas de la fusión de imágenes redimensionadas de prioridad 1 y prioridad 2. La información ampliada correspondiente a este análisis se encuentra disponible en el Apéndice B.

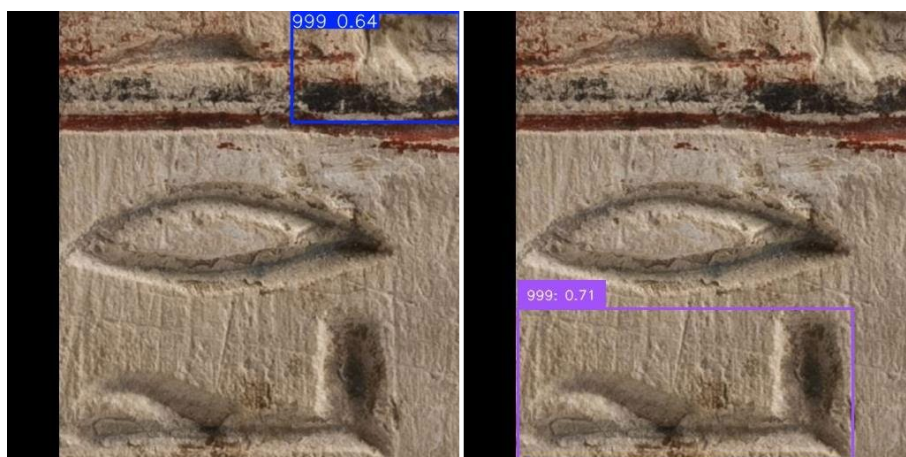


Figura 14. Comparativa entre los resultados de detección del modelo «999». YOLO11 (izquierda) y RF-DETR (derecha) con umbral de confianza al 0.25. Adaptado de Stela dedicated by Smen to Re-Horakhty. Museo Egizio di Torino.

La Figura 15 muestra la estela con mayor densidad de signos procesada en este análisis. Como se observa en la Tabla 8, tanto en el anverso como en el reverso de la Magical Stela with Horus the Child (Estela Mágica), la prioridad 1 (caja roja), que corresponde a las detecciones de alta confianza de RF-DETR sobre el modelo de «Aves», se establece en la máxima escala (3x) del conjunto *S*. Esto se debe a que, al ser una estela de grandes dimensiones, las ventanas deslizantes no tienen suficiente detalle para obtener una imagen clara del glifo aviar. La redimensión compensa este problema. En contraste, la prioridad 2 (cajas magenta «Aves» y azul «Top 4»), formada por la suma de las detecciones de confianza de los modelos YOLO11, no necesita el escalado para obtener los mejores resultados, confirmando el comportamiento detectado durante la fase de evaluación de ambos modelos.

A modo de ejemplo, al analizar el reverso de la estela de la Figura 15, el algoritmo NMS debe procesar 60.751 cajas delimitadoras en bruto si opera sobre todas las escalas. Sin embargo, al seleccionar únicamente las escalas que concentran el número máximo de detecciones (Figura 16) para elevar la precisión y reducir el ruido de fondo, la carga computacional se reduce. Al priorizar las predicciones de los modelos expertos, las aves del modelo RF-DETR con prioridad 1 y las detecciones YOLO con prioridad 2, el procesamiento NMS se reduce a 6.851 cajas en la escala 3x y 4.300 en la escala 1x.

Como se puede observar en el detalle de la Figura 17, trabajar con máximos minimiza el ruido que se genera especialmente en la decoración. En la izquierda de la imagen se puede apreciar cómo el NMS multiescala, que en este caso ha procesado 44.166 cajas totales en la estela, reconoce más elementos de decoración que el procesado por máximos, el cual ha escogido la escala 3x para la prioridad 1 con 4.768 cajas y la escala 1x con 3.001 detecciones para la prioridad 2. Sin embargo, este resultado no descarta la utilización del NMS global, ya que, en estelas fuertemente degradadas, este modelo puede conseguir resultados donde el filtro de máximos ignora glifos.

Tabla 9. Ejemplos de localización sobre estelas

Estela	Escala		Localizador			
	P1	P2	RF-DETR «Aves»	YOLO11 «Aves»	YOLO11 «Top 4»	RF-DETR «999»
Stela of Nya	1,0	0,5	P1: 10	P1: 3	P1: 27	P1: 435
			P2: 6	P2: 10	P2: 31	P2: 363
			NMS: 11	NMS: 6	NMS: 39	NMS: 449
Round-topped stele of Ipuy	2,25	1,75	P1: 7	P1: 1	P1: 25	P1: 153
			P2: 3	P2: 4	P2: 32	P2: 147
			NMS: 9	NMS: 2	NMS: 37	NMS: 174
Stela dedicated to the goddess Renenutet by Mutneferet and Iyneferet	0,75	0,75	P1: 3	P1: 0	P1: 15	P1: 26
			P2: 3	P2: 0	P2: 15	P2: 26
			NMS: 3	NMS: 0	NMS: 15	NMS: 26
Stela of Neferabu	0,75	0,5	P1: 35	P1: 6	P1: 59	P1: 375
			P2: 26	P2: 17	P2: 59	P2: 353
			NMS: 35	NMS: 10	NMS: 75	NMS: 402
Stela of Intef and Nesumontu	1,0	0,5	P1: 5	P1: 3	P1: 21	P1: 57
			P2: 3	P2: 6	P2: 19	P2: 58
			NMS: 5	NMS: 5	NMS: 21	NMS: 57
False-door stela of Harbes	0,5	2,5	P1: 0	P1: 1	P1: 15	P1: 56
			P2: 0	P2: 0	P2: 18	P2: 56
			NMS: 0	NMS: 1	NMS: 18	NMS: 58
Magical stela or cippus of Ho- rus	1,5	0,5	P1: 34	P1: 3	P1: 104	P1: 441
			P2: 14	P2: 35	P2: 140	P2: 286
			NMS: 35	NMS: 16	NMS: 150	NMS: 438
Stela of the lector priest of Amun Siamun...	0,5	1,5	P1: 7	P1: 0	P1: 27	P1: 104
			P2: 5	P2: 3	P2: 28	P2: 80
			NMS: 7	NMS: 1	NMS: 31	NMS: 107
Stela of the temple servant of Amun Irthareru	1,75	2,0	P1: 14	P1: 6	P1: 31	P1: 295
			P2: 8	P2: 10	P2: 35	P2: 294
			NMS: 14	NMS: 7	NMS: 39	NMS: 344
Stela of Horemkhauef	1,25	2,75	P1: 46	P1: 17	P1: 133	P1: 482
			P2: 19	P2: 16	P2: 160	P2: 391
			NMS: 48	NMS: 15	NMS: 170	NMS: 480
Magical Stela with Horus the Child (Anverso)	3,0	1,0	P1: 151	P1: 36	P1: 547	P1: 1938
			P2: 40	P2: 152	P2: 579	P2: 885
			NMS: 152	NMS: 54	NMS: 626	NMS: 1917
Magical Stela with Horus the Child (Reverso)	3,0	1,0	P1: 208	P1: 38	P1: 971	P1: 2580
			P2: 44	P2: 212	P2: 971	P2: 1073
			NMS: 211	NMS: 49	NMS: 1024	NMS: 2601
Sheet from the Papyrus of Amenhotep	1,0	0,5	P1: 39	P1: 26	P1: 74	P1: 643
			P2: 28	P2: 60	P2: 71	P2: 470
			NMS: 46	NMS: 48	NMS: 93	NMS: 608
Heqanahkt Letter II	2,25	0,5	P1: 11	P1: 12	P1: 24	P1: 1319
			P2: 2	P2: 56	P2: 54	P2: 625
			NMS: 12	NMS: 61	NMS: 70	NMS: 1270



Figura 15. Resultado del proceso de detección sobre Magical Stela with Horus the Child (Estela Mágica).

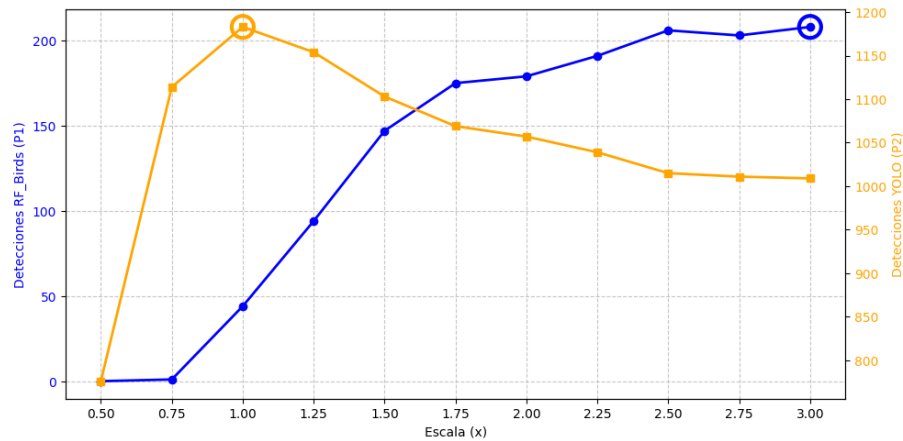


Figura 16. Selección de escala por máx. detecciones.



Figura 17. Detalle anverso de Magical Stela with Horus the Child (Estela Mágica): (a) NMS Multiescla; (b) NMS global.

Además de valorar el rendimiento de detección, también cabe destacar la capacidad de generalización de los modelos. A pesar de haber sido entrenados exclusivamente con estelas de los Reinos Antiguo y Medio, el sistema logra identificar signos de periodos posteriores, como los de la Estela Mágica (Fig. 15) del Periodo Tardío (360–343 a. C.), e incluso en soportes de papiro. Este resultado es aún más significativo, ya que en dichos materiales la escritura solía ser más pragmática y abstracta. La Figura 18 corresponde a la Hoja del papiro de Amenhotep, redactada en jeroglífico cursivo, que presenta un nivel de simplificación mayor que el de la piedra. A pesar de que, como se puede apreciar en los datos de la Tabla 8 y en la Figura 18, los modelos expertos no dan garantía de identificación de las clases de «Aves» y «Top 4» en este papiro, el modelo genérico es capaz de detectar la mayor parte de los glifos de la estela correctamente.

4.6 Clasificador de «Aves»

Para evaluar el impacto de la distribución de clases sobre el rendimiento de la red ConvNeXt, se diseñó un análisis comparativo entre tres estrategias de estructuración de datos aplicadas al subgrupo de aves (categoría G).

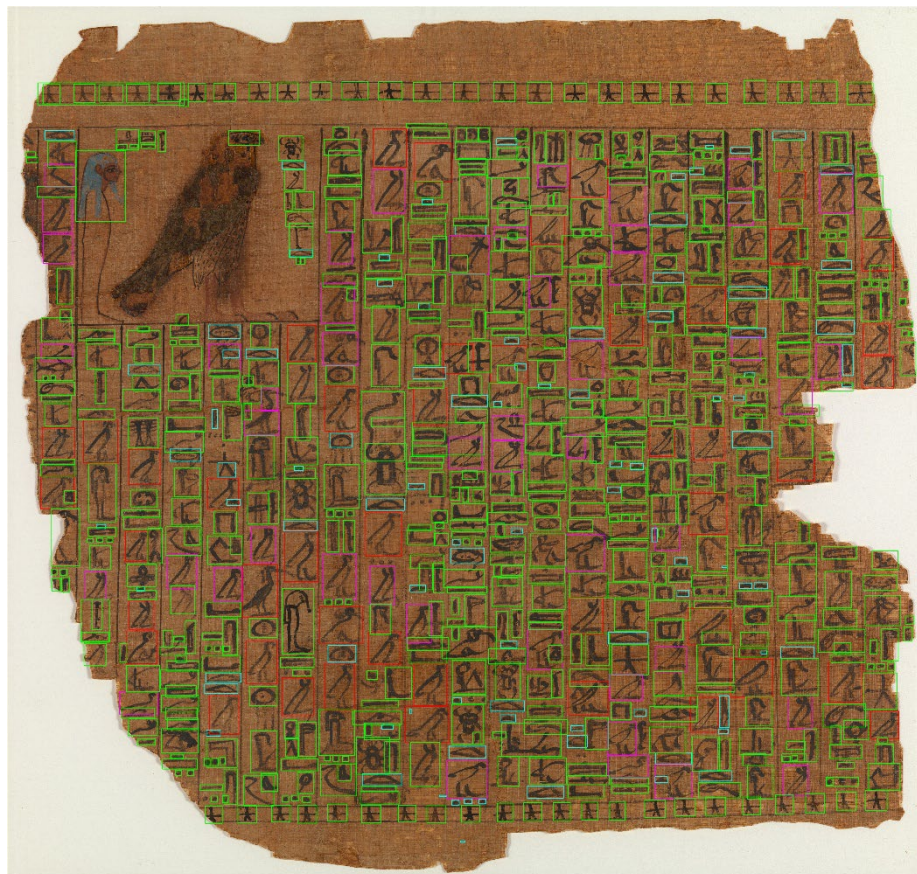


Figura 18. Hoja del papiro de Amenhotep.

El análisis inicial de este subgrupo revela una discrepancia evidente entre la disponibilidad de imágenes y su frecuencia de aparición real (Tabla 10). El *dataset* consta de 1060 imágenes distribuidas en 26 clases (comprendidas entre G1 y G50). Aunque el corpus de referencia (AED) incluye hasta 106 variantes, estas 26 conforman el 82,76 % de los registros de la categoría. Esto representa un 6,82 % del corpus total (sobre el 8,24 % que suman todas las aves), garantizando así la representatividad estadística de la muestra empleada.

Una primera estrategia de clasificación planteaba aislar las seis categorías principales y relegar las menos frecuentes a una clase genérica. Aunque teóricamente esta decisión maximiza la cobertura sin saturar el modelo, en la práctica introduce un grave problema de ruido semántico al agrupar bajo una misma etiqueta signos morfológicamente casi idénticos, como los pares G5-G7 o G4-G1. Esta pérdida de discriminación visual exige plantear estrategias de agrupación alternativas. A este problema se suma el extremo desequilibrio del *dataset* en clases de detección prioritaria, ya que la diferencia entre signos muy escasos (el G7 cuenta con solo 11 muestras) y clases con un número suficiente de muestras (el G43 acumula 373) distorsiona el cálculo del gradiente y desploma la precisión sobre las categorías minoritarias.

A primera vista, una configuración lógica consistiría en limitar la clasificación a las seis categorías principales, desplazando las menos frecuentes a una clase resto. Esta aproximación busca maximizar la cobertura sin sobrecargar el modelo con clases infrarrepresentadas. Pero el problema de este enfoque es que agrupa signos morfológicamente muy similares (como G5 frente a G7, o G4 frente a G1) dentro de una misma categoría. Esta mezcla introduce ruido semántico que degrada drásticamente la capacidad de discriminación del sistema, obligando a plantear estrategias de agrupación alternativas. Además, la extrema disparidad de muestras entre signos muy frecuentes (como G7, con 11 muestras) y

signos mayoritarios en el *dataset* (como G43, con 373 muestras) genera un desequilibrio perjudicial. Esta asimetría penaliza la precisión de las clases minoritarias debido a su escasa representatividad durante el cálculo del gradiente. Para corregir esta desviación, se estableció el objetivo de alcanzar un mínimo de 300 muestras por clase mediante técnicas avanzadas de aumento de datos. Como se observa en la Figura 19, las imágenes originales de la clase G7 no solo eran escasas, sino también monocromáticas y de baja variabilidad. Mediante la aplicación del pipeline generativo basado en BicycleGAN (Figura 20), se logró sintetizar un volumen suficiente de muestras integrando diferentes texturas y tipologías de trazado para completar la clase.

Tabla 10. Jeroglíficos categoría G

Clase	Representación	Imágenes	Frecuencia	Cobertura
g7	 (Halcón sobre estandarte)	11	984	2,04%
g17	 (Lechuza)	334	707	1,47%
g1	 (Alimoche)	41	575	1,19%
g43	 (Polluelo de codorniz)	373	460	0,96%
g37	 (Gorrión)	14	110	0,23%
g36	 (Golondrina)	52	94	0,20%
g4	 (Busardo)	8	60	0,12%
g39	 (Halcón)	21	48	0,10%
g29	 (Halcón)	3	48	0,10%
g5	 (Halcón)	40	40	0,08%
g41	 (Halcón)	4	38	0,08%
g25	 (Halcón)	32	28	0,06%
g40	 (Halcón)	27	26	0,05%
g14	 (Halcón)	21	20	0,04%
g47	 (Halcón)	2	12	0,02%
g20	 (Halcón)	1	10	0,02%
g21	 (Halcón)	2	9	0,02%
g38	 (Halcón)	28	5	0,01%
g35	 (Halcón)	38	3	< 0,01%
Resto	 ...	3 - 1	3 - 1	< 0,01%

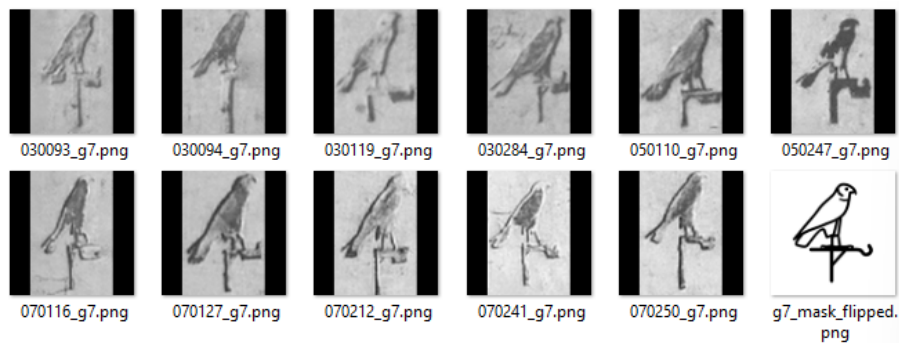


Figura 19. Todas las muestras originales de g7.

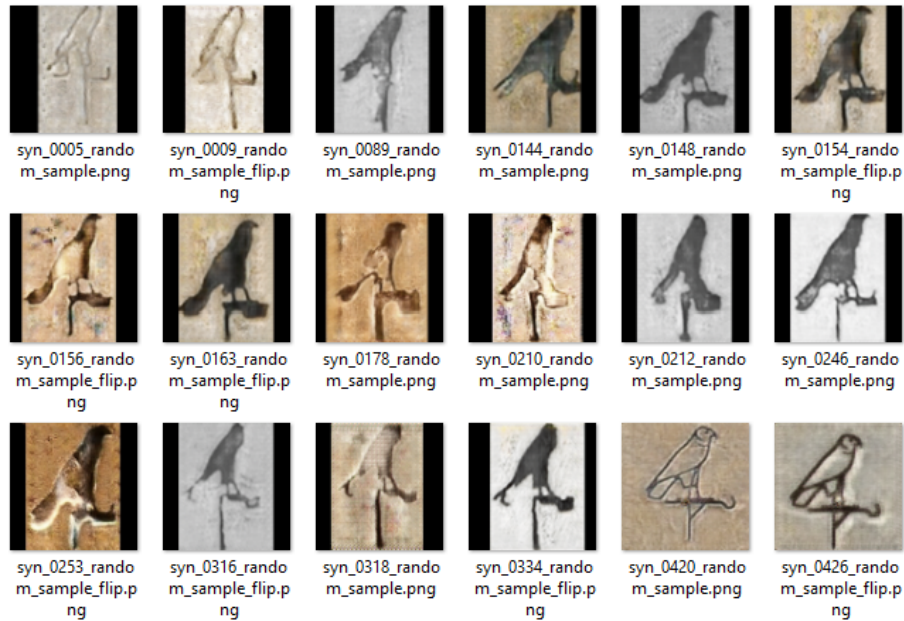


Figura 20. Ejemplos de muestras generadas sintéticamente para la clase g7

Al equiparar el volumen de las categorías minoritarias con las dominantes, se balancea físicamente su contribución al descenso del gradiente. Con ello se logra estabilizar las fronteras de decisión de la red sin necesidad de manipular de forma artificial los hiperparámetros de la función de coste.

4.6.1 Modelo 1: Frecuencia

En el enfoque inicial, nos basamos exclusivamente en la frecuencia de aparición de los signos. Para ello, se conservaron como clases independientes los glifos mayoritarios y se agruparon los minoritarios en una única categoría denominada «*RESTO_AVES*». Esta estrategia de agrupación incluye en un mismo conjunto signos morfológicamente conflictivos, como el G4 (busardo) y el G5 (halcón). La Tabla 11 detalla la distribución del *dataset* generado.

Tabla 11. Distribución *dataset* modelo de frecuencia.

Clase	Originales	Train real	Val real	Train final	Val final	Sintéticos
g7	12	10	2	300	4	140
g17	335	268	67	536	134	0
g1	42	34	8	300	16	116
g43	374	299	75	598	150	0
g37	15	12	3	300	6	138
g36	53	42	11	300	22	108
RESTO AVES	236	189	47	378	94	0

Como parte de la estrategia de configuración de datos, se reservaron imágenes originales exclusivamente para la fase de validación. Asimismo, para alcanzar el volumen requerido en el conjunto de entrenamiento (*train*), se aplicó un volteo horizontal (*flip* horizontal) a la totalidad de las muestras, tanto reales como sintéticas.

Durante el entrenamiento, la red experimenta una rápida convergencia, el conjunto de entrenamiento alcanza valores cercanos a 1 en precisión, mientras que la curva de validación se estanca entre 0,85 y

0,90 (Figura 21). Esta divergencia evidencia un claro sobreajuste, que no logra mitigarse pese a la inyección de datos sintéticos para el balanceo.

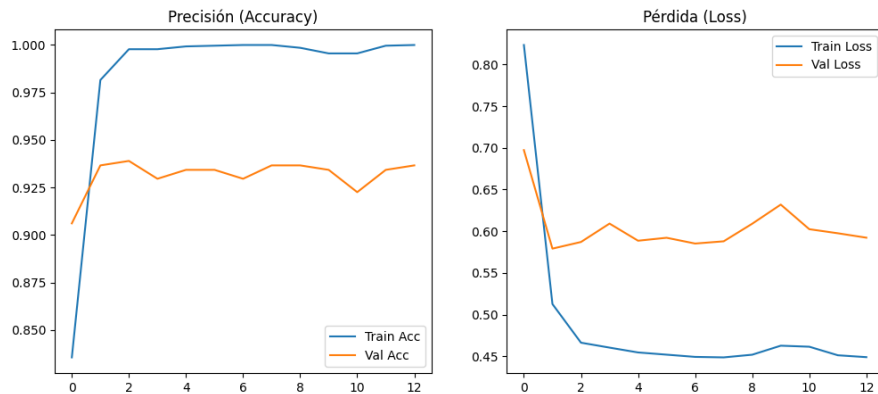


Figura 21. Curvas de entrenamiento del Modelo 1

Aunque las métricas globales sugieren inicialmente una alta precisión, este valor se encuentra distorsionado por el peso de las clases mayoritarias (G17 y G43). El desglose de las métricas (Tabla 12) y la matriz de confusión (Figura 22) demuestran las carencias estructurales de agrupar los signos infrecuentes.

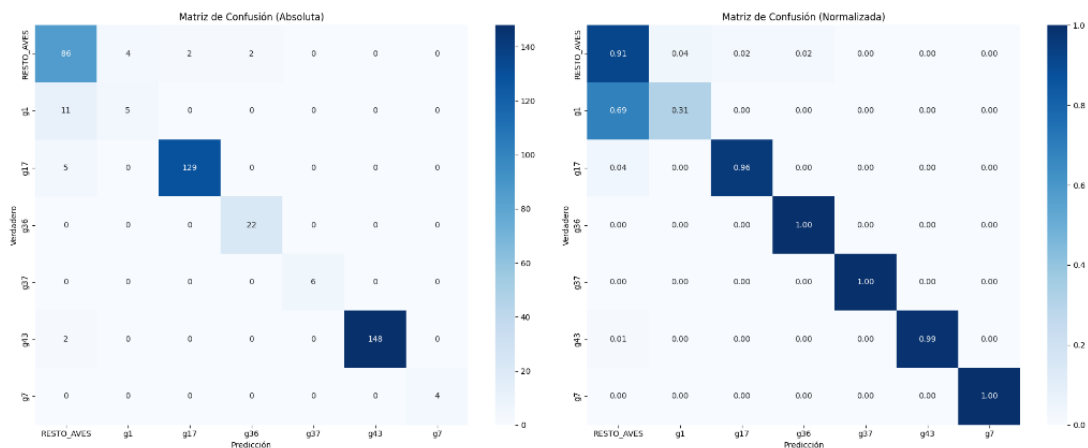


Figura 22. Matriz de confusión del Modelo 1

Tabla 12. Métricas por clase del Modelo 1

Clase	Precision	Recall	F1-score	Muestras
g7	1.0	1.0	1.0	4
g17	0.9847	0.9627	0.9736	134
g1	0.5556	0.3125	0.4000	16
g43	1.0	0.9867	0.9933	150
g37	1.0	1.0	1.0	6
g36	0.9167	1.0	0.9565	22
RESTO AVES	0.8269	0.9149	0.8687	94

El mayor impacto se observa en la clase G1 (alimoche), cuya precisión cae al 55 % con una recuperación de apenas el 31%. La red confunde sistemáticamente este signo con la categoría «*RESTO_AVES*», un error de clasificación provocado de forma directa por haber incluido el glifo G4 (busardo) en dicho grupo genérico. Las clases G7 y G37 muestran métricas perfectas, un resultado condicionado por el bajo número de muestras. Aunque los datos cuantitativos no lo reflejen explícitamente, la similitud

morfológica crítica entre G5 y G7 (que representan la misma ave diferenciada únicamente por el estandarte) constituye un riesgo identificable gracias al conocimiento del dominio aplicado al diseño del modelo.

4.6.2 Modelo 2: Aislamiento semántico

Para corregir la confusión semántica detectada en el primer experimento, se extrajeron los signos G4 y G5 de la clase «*RESTO_AVES*». y se configuraron como categorías independientes. De este modo, se evita que la red asocie rasgos morfológicos propios de aves específicas a un grupo genérico de ruido. La Tabla 13 detalla esta redistribución de muestras con respecto a la Tabla 12, mostrando las nuevas clases segregadas y la consiguiente reducción de volumen en «*RESTO_AVES*».

Tabla 13. Distribución del *dataset* del modelo de aislamiento semántico.

Clase	Originales	Train real	Val real	Train final	Val final	Sintéticos
g4	9	7	2	300	4	143
g5	41	33	8	300	16	117
RESTO_AVES	186	149	37	298	74	0

Durante el entrenamiento (Figura 23), la separación de estas clases conflictivas no desestabiliza la convergencia de la red. Al contrario, la curva de validación (val) presenta una menor volatilidad y alcanza una asíntota superior, estabilizándose por encima del 94 % de precisión (*accuracy*). La eliminación de esta confusión semántica en la clase genérica facilita al modelo la definición de fronteras de decisión mucho más claras. La matriz de confusión (Figura 24) permite confirmar la eficacia de la estrategia de aislamiento y, además, identificar sus límites.

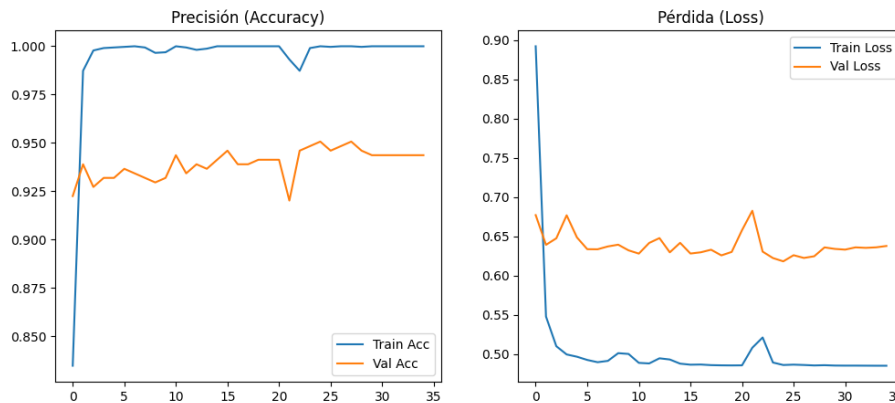


Figura 23. Curva de entrenamiento del Modelo 2

Como se puede apreciar en la Tabla 14, al extraer las clases conflictivas de la categoría genérica, el sistema logra aislar correctamente la morfología del G1 (alimoche), elevando su tasa de recuperación del 31,25 % al 81,25 %. Al estabilizarse las inferencias del G1 en la diagonal de aciertos, se logra contener la fuga de predicciones hacia el bloque «*RESTO_AVES*». Sin embargo, este experimento confirma la inoperancia de entrenar categorías infrarrepresentadas. La detección del busardo (G4) falla totalmente al registrar un F1-score de 0,00, siendo absorbida por el G1. También se constata un claro empeoramiento en la clasificación de la golondrina (G36) y el gorrión (G37) respecto al Modelo 1, lo que indica que el margen de distinción morfológica entre ambas aves resulta demasiado difuso para la red neuronal.

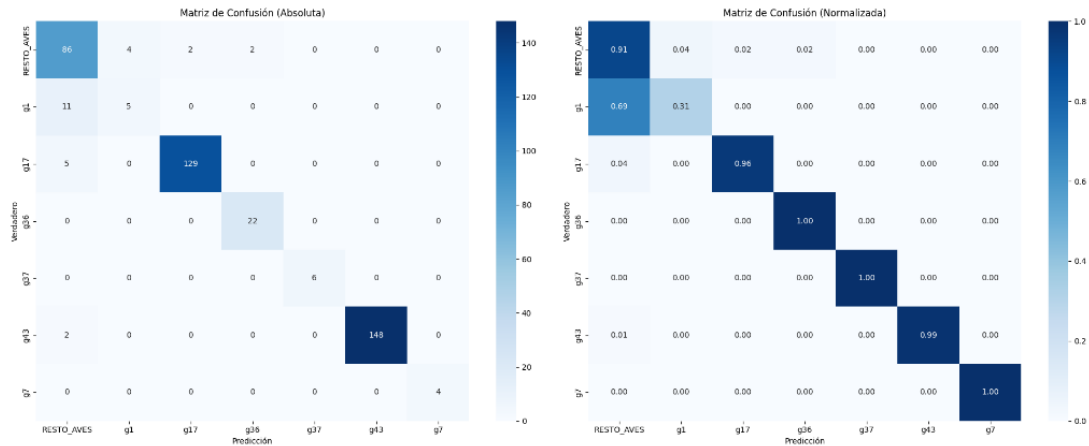


Figura 24. Matriz de confusión del Modelo 2

Tabla 14. Métricas por clase del Modelo 2

Clase	Precision	Recall	F1-score	Muestras
g7	1	1	1	4
g17	0.9851	0.9851	0.9851	134
g1	0.6190	0.8125	0.7027	16
g43	1	1	1	150
g37	0.8000	0.6667	0.7273	6
g36	0.8095	0.7727	0.7907	22
g4	0	0	0	4
g5	0.8421	1	0.9143	16
RESTO_AVES	0.9583	0.9324	0.9452	74

En contraste, el comportamiento de la clase G5 (halcón) demuestra que el aislamiento semántico es una estrategia viable cuando se dispone de un volumen de muestras mínimo (41 imágenes originales). En este caso, la red alcanza un recall del 100 % y un F1-score del 91,43 %, logrando diferenciar al halcón tanto del G7 (halcón sobre estandarte) como de la clase resto.

4.6.3 Modelo 3: Agrupación

En este modelo, se aplicó el conocimiento experto del dominio con el fin de redefinir la distribución del *dataset*, fusionando aquellas clases que presentan una alta ambigüedad visual (G36/G37) o que sufren un déficit crítico de muestras (caso G4). Para evitar que la escasez de muestras penalice la distinción de detalles visuales muy sutiles, la estrategia fusiona estas aves en las superclases G1_G4 y G36_G37. Esto facilita el aprendizaje de características morfológicas compartidas. La clase G5 se mantiene segregada, ya que, como se ha validado en el modelo anterior, esto garantiza su correcta separación del fondo. Como resultado de esta reestructuración, las etiquetas individuales G4, G36 y G37 desaparecen del conjunto de datos, cuya nueva distribución queda detallada en la Tabla 15.

Tabla 15. Distribución *dataset* de entrenamiento del modelo agrupado.

Clase	Originales	Train real	Val real	Train final	Val final	Sintéticos
g7	12	10	2	300	4	140
g17	335	268	67	536	134	0
g1_g4	51	41	10	300	20	109
g43	374	299	75	598	150	0
g36_g37	68	54	14	300	28	96
g5	41	33	8	300	16	117
RESTO_AVES	186	149	37	298	74	0

Las curvas de aprendizaje de este modelo (Figura 25) muestran unos datos de convergencia muy similares al de los escenarios precedentes, alcanzando rápidamente la saturación en el conjunto de entrenamiento. La curva de validación se mantiene estable por encima del 94 % de precisión, sin mostrar signos de degradación a pesar de la reducción del número total de etiquetas. Esto indica que la red neuronal se adapta correctamente a la nueva topología, encontrando patrones discriminativos válidos para las agrupaciones.

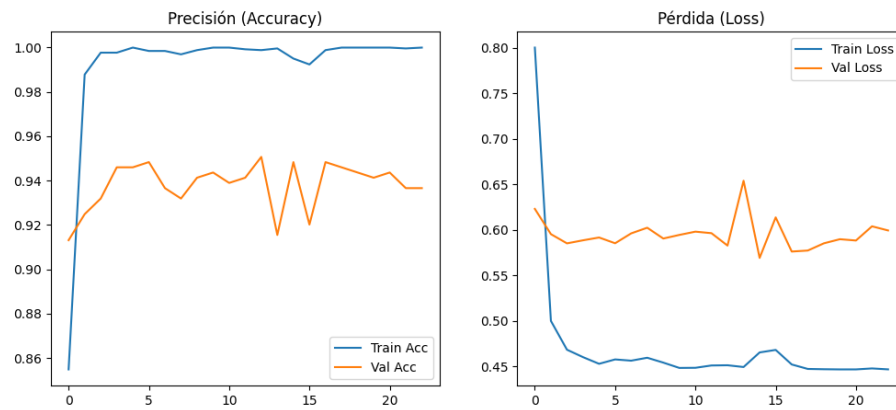


Figura 25. Curva de entrenamiento del Modelo 3

El desglose de métricas por clase (Tabla 16) y el análisis de la distribución de errores visualizada en la matriz de confusión (Figura 26), ofrece una imagen precisa de la eficacia de la estrategia de agrupación.

Tabla 16. Métricas por clase del Modelo 3

Clase	Precision	Recall	F1-score	Muestras
g7	1.0	1.0	1.0	4
g17	0.9781	1.0	0.9889	134
g1_g4	0.7143	0.50	0.5882	20
g43	0.9804	1.0	0.9901	150
g36_g37	0.9032	1.0	0.9492	28
g5	1.0	0.8125	0.8966	16
RESTO AVES	0.8919	0.8919	0.8919	74

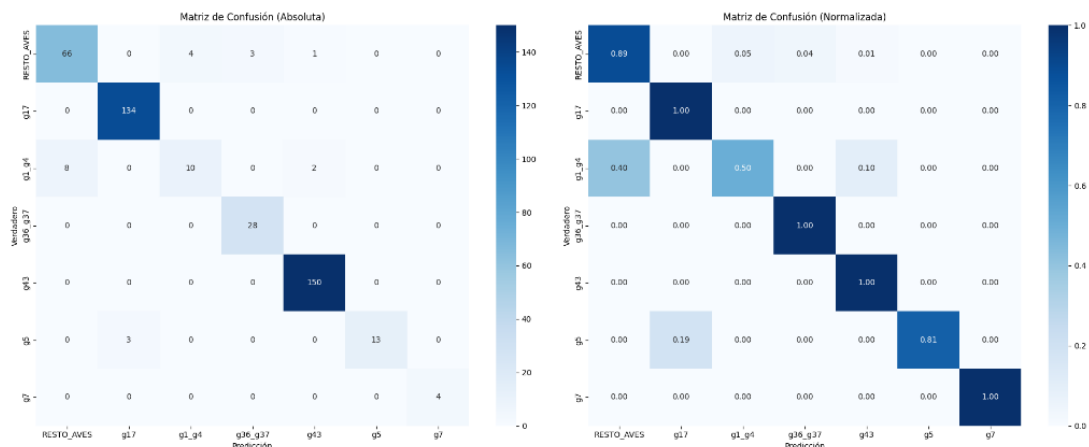


Figura 26. Matriz de confusión del Modelo 3

La agrupación de golondrinas (G36) y gorriones (G37) demuestra ser una estrategia efectiva, logrando la nueva superclase G36_G37 una tasa de recuperación del 100 % y un F1-score de 0,9492.

Sin embargo, la fusión del alimoche (G1) y el busardo (G4) arroja resultados deficientes. Aunque esta unión evita el fallo absoluto registrado en el G4 de forma aislada, el rendimiento conjunto de la superclase G1_G4 se limita a un F1-score de 0,5882 y un recall del 50 %. La matriz de confusión revela como se desplaza el error de una confusión interna a una externa, tendiendo el modelo a clasificar las instancias de esta superclase como «*RESTO_AVES*» o G17.

Por su parte, las categorías con alto soporte (G17, G43) mantienen su estabilidad. La clase G7 conserva métricas perfectas, lo que reitera que su morfología distintiva (el estandarte) constituye un rasgo discriminativo lo suficientemente robusto para la red, incluso partiendo de un volumen de muestras reducido.

4.6.4 Evolución comparada

La creación de diferentes modelos para esta tarea pone de manifiesto la dificultad de identificar correctamente las clases de similitud morfológica. Gracia Zamacona (2017) ya advierte de casos de confusión como el G1 y G4, pero en un problema de visión artificial es fácil caer en el autoengaño si se consideran únicamente las métricas, como en el caso del Modelo 1. Así, el análisis de la matriz de confusión y la validación visual resultan indispensables para evaluar la eficacia real del sistema.

En la transición hacia el Modelo 3 hubo que tomar la decisión de fusionar las clases G1 (alimoche) y G4 (busardo). El Modelo 2 lograba aislar la clase G1 con un F1-score de 0,70, un rendimiento que desciende a 0,58 tras su integración en el Modelo 3. Esta caída pone de manifiesto que sumar las imágenes del G4 contamina el espacio representacional del G1.

A pesar de esta penalización, mantener la superclase G1_G4 resulta indispensable. Suprimir la categoría G4 mejoraría artificialmente las métricas, pero dejaría al sistema sin capacidad para detectar esta clase en inscripciones reales, clasificándolos como ruido o falsos positivos. Por lo que hay es necesario asumir un coste de generalización en donde localizar la familia alimoche/busardo con menor confianza es estratégicamente preferible a perder un glifo estructural.

Es por esto, que el Modelo 3, aunque mejorable mediante nuevas muestras reales, se consolida como la arquitectura más viable para el despliegue inicial. Su diseño no solo resuelve la ambigüedad visual entre golondrinas y gorriones (G36_G37), sino que mitiga la escasez de datos que inhabilitaba la detección individual del busardo. La evaluación visual de ejemplos de inferencia, que se muestra en las imágenes con el formato Modelo 1 | Modelo 2 | Modelo 3, confirma este comportamiento.

En la Figura 27, un halcón (G5) inicialmente descartado por el Modelo 1 como «*RESTO_AVES*», es recuperado por el resto de modelos gracias a la estrategia de aislamiento semántico. La Figura 28 muestra como ante glifos de la misma familia presentes en la misma inscripción, el Modelo 1 y el Modelo 2 emiten predicciones contrarias. El Modelo 3 estabiliza la inferencia englobando ambas detecciones en una única superclase. En la Figura 29 se reflejan los límites del entrenamiento con bajo soporte, mostrando cómo el Modelo 2 fracasa ante una muestra aislada de G4 y descarta la predicción a la categoría de ruido.



Figura 27. Halcón G5 como RESTO_AVES

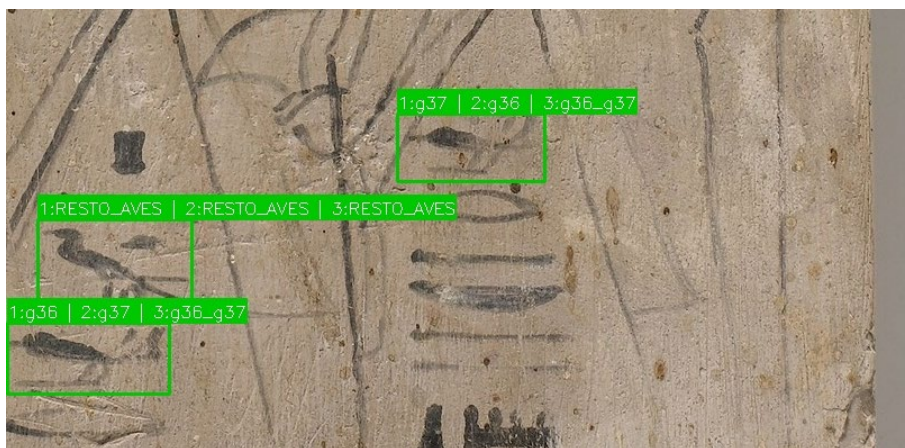


Figura 28. Resolución de ambigüedad G36/G37



Figura 29. Recuperación por superclase.

Este experimento demuestra que, incluso partiendo de un número mínimo de muestras originales, es factible construir un clasificador competente mediante el uso estratégico de datos sintéticos. Pero hay que tener en cuenta otros factores como la correcta definición de clases para evitar la contaminación cruzada, la flexibilidad para adaptar las etiquetas a la realidad visual del soporte, y el compromiso entre la precisión y la cobertura esperadas por el modelo. De todas formas, en futuras revisiones se

debería ajustar el modelo con imágenes reales de las clases deficitarias, si bien el desarrollo actual puede servir de apoyo para el etiquetado en nuevos *datasets*.

4.7 Clasificador «Top 4»

El segundo modelo experto, denominado «Top 4», está diseñado para la identificación de cuatro de los signos de mayor frecuencia de aparición en los textos: X1, D21, N35 y M17.

Como estos glifos son omnipresentes, el localizador puede generar falsos positivos con signos de morfología afín, como la fácil confusión entre el signo D4 y el D21 visible en la Figura 30. Para mitigar este riesgo, la estrategia adoptada incluye la incorporación de una clase de desambiguación o rechazo denominada «OTROS».

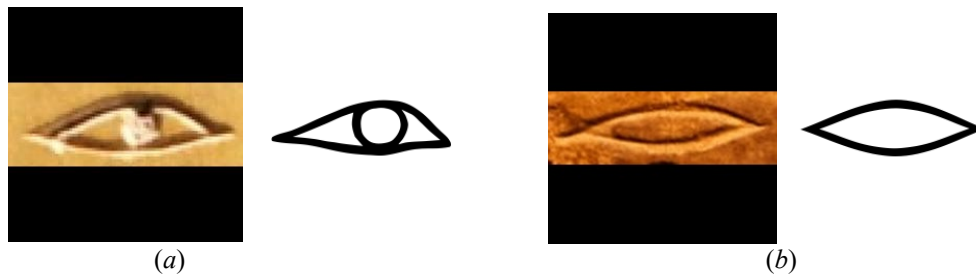


Figura 30. Similitud Morfológica: (a) D4 (Ojo), F. NewGardiner D4; (b) D21 (Boca), F. NewGardiner D24

A diferencia del clasificador de aves, en este conjunto no existe un déficit de muestras originales. Por lo tanto, la estrategia de balanceo se proyectó al alza, estableciendo un objetivo de 900 muestras por clase para el conjunto de entrenamiento. La inyección de imágenes sintéticas generadas mediante el pipeline de aumento de datos (cGANs y BicycleGAN) se aplicó de forma restrictiva.

Como se observa en la Tabla 17, la clase M17 alcanzó el umbral empleando exclusivamente sus 675 muestras reales de entrenamiento (apoyadas con técnicas de aumento clásico), mientras que categorías como D21 y X1 requirieron la integración de 467 y 362 muestras sintéticas, respectivamente

Tabla 17. Distribución *dataset* modelo «Top 4».

Clase	Originales	Train real	Val real	Train final	Val final	Sintéticos
d21	541	433	108	900	108	467
m17	844	675	169	900	169	0
n35	988	790	198	900	198	110
x1	672	538	134	900	134	362

Las curvas de aprendizaje de la Figura 31 muestran una convergencia rápida y estable. A diferencia del sobreajuste que penalizó las primeras fases del clasificador de «Aves», en este caso la curva de validación se asienta pronto en valores de precisión superiores al 98 %, confirmando que el modelo discrimina las clases correctamente.

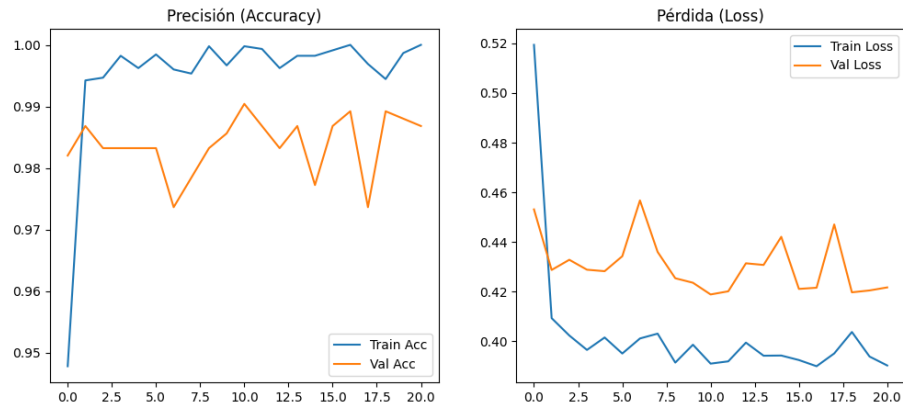


Figura 31. Curva de entrenamiento «Top 4»

La matriz de confusión (Figura 32), muestra una diagonal principal sólida y libre de ruido. Por otra parte, la efectividad de la clase «OTROS» queda demostrada al absorber correctamente 220 instancias anómalas, registrando fugas marginales (apenas 3 muestras confundidas con M17 y 2 con X1).

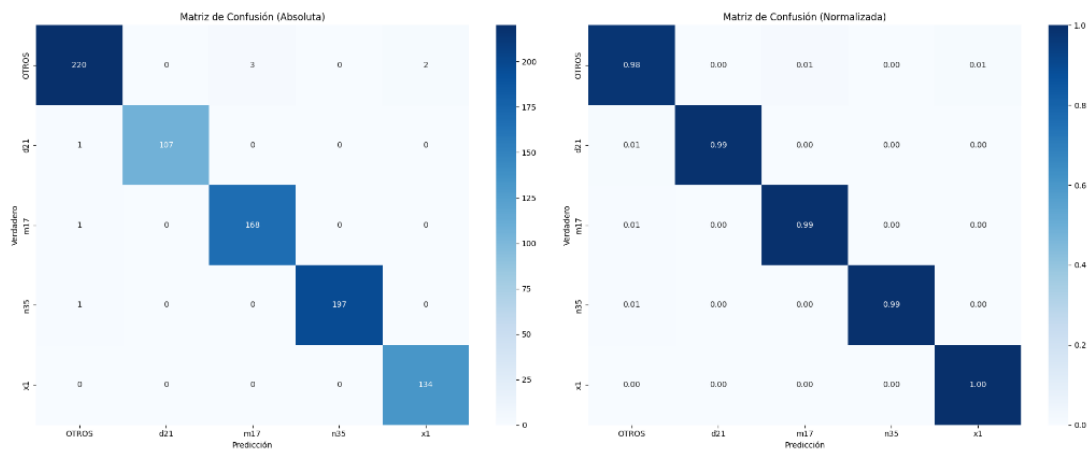


Figura 32. Matriz de confusión «Top 4»

Los valores de la Tabla 18 constatan el alto rendimiento cuantitativo del modelo. La totalidad de las clases evaluadas supera un F1-score de 0,98, destacando el desempeño casi perfecto de las clases N35 (0,9975), D21 (0,9953) y X1 (0,9926). Este grado de exactitud responde a que los glifos del top-4 poseen características estructurales muy distintivas y factores de forma singulares. Al combinar estas propiedades con un volumen de entrenamiento masivo, el modelo minimiza drásticamente el margen de error.

Tabla 18. Métricas del clasificador «Top 4»

Clase	Precision	Recall	F1-score	Muestras
x1	0.9853	1.0000	0.9926	134
n35	1.0000	0.9949	0.9975	198
m17	0.9825	0.9941	0.9882	169
d21	1.0000	0.9907	0.9953	108
OTROS	0.9865	0.9778	0.9821	225

Esta arquitectura confirma que, al disponer de un soporte de datos masivo y equilibrado sobre glifos con marcadas diferencias morfológicas, el extractor de características ConvNeXt genera un clasificador altamente preciso. Se puede confirmar que el diseño de modelos expertos basados en agrupaciones de glifos es una estrategia empírica sólida para asegurar la correcta clasificación de los signos estructurales más frecuentes.

4.8 Clasificador genérico «999»

Este clasificador se encarga de evaluar las 19 categorías restantes del *dataset* que lograron superar el umbral de viabilidad de 150 muestras reales. A diferencia del clasificador «*Top 4*», esta red omite deliberadamente la clase de rechazo. El objetivo de este diseño es forzar a la red a realizar una inferencia basada en la proximidad morfológica, lo que garantiza que los signos visualmente emparentados (como las variaciones de figuras humanas A1, A2 y B4) puedan generalizarse correctamente hacia su categoría matriz.

Para estabilizar el entrenamiento y asegurar la equidad en la actualización de pesos durante el descenso del gradiente, el conjunto de entrenamiento se balanceó hasta alcanzar un objetivo de 300 muestras por clase. Como detalla la Tabla 19, se aplicó una inyección moderada de imágenes sintéticas para complementar el volumen de aquellas categorías que no alcanzaban dicha cantidad de muestras mediante sus recortes originales.

Tabla 19. Distribución del *dataset* del clasificador genérico «999»

Clase	Originales	Train real	Val real	Train final	Val final	Sintéticos
a1	166	133	33	306	33	20
d36	223	178	45	300	45	0
d4	276	221	55	300	55	0
d58	157	126	31	309	31	29
e34	158	126	32	309	32	29
f35	170	136	34	332	34	196
i9	474	379	95	300	95	0
m23	209	167	42	300	42	0
n33	177	142	35	331	35	189
o1	163	130	33	334	33	204
q1	175	140	35	304	35	12
q3	218	174	44	325	44	151
r8	236	189	47	300	47	0
s29	473	378	95	300	95	0
v13	163	130	33	308	33	24
v28	208	166	42	326	42	160
v30	285	228	57	314	57	86
v31	267	214	53	300	53	0
x8	155	124	31	335	31	211

La evolución del entrenamiento (Figura 33) muestra una adaptabilidad excelente a la variabilidad de las 19 clases. Las curvas de pérdida (*loss*) y precisión (*accuracy*) indican una reducción drástica del error en las primeras cinco épocas. Posteriormente, la curva de validación se mantiene en una franja de precisión estable que oscila entre el 97 % y el 99 %.

El rendimiento de este clasificador de grano fino demuestra que, tras derivar la mayor parte de la ambigüedad visual a los modelos expertos previos («*Aves*» y «*Top 4*»), las categorías restantes poseen una varianza interclase lo suficientemente amplia como para ser discriminadas con facilidad por el modelo.

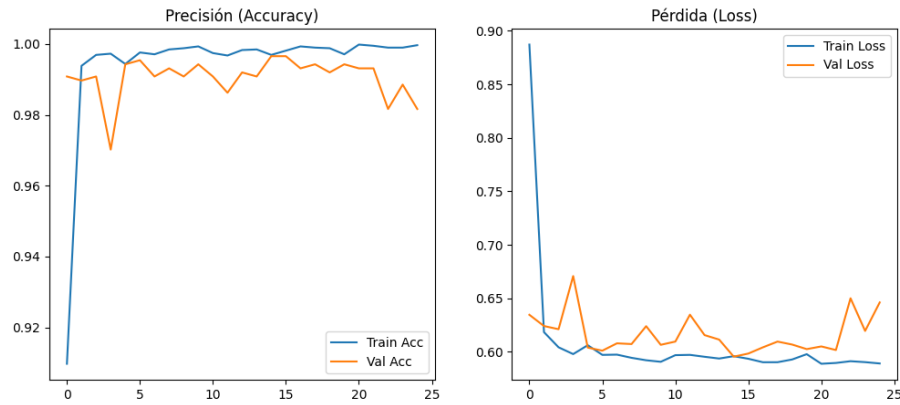


Figura 33. Curva de entrenamiento del clasificador genérico «999»

El análisis de métricas de la Tabla 20, revela un F1-score perfecto en la gran mayoría de los signos. Las desviaciones en la matriz de confusión (Figura 34) son anecdóticas y no representan un patrón de confusión estructural, únicamente se reporta un falso positivo cruzado entre D58 y Q3, y una mínima desviación de I9 hacia E34.

Estas métricas ratifican la fiabilidad de la arquitectura en cascada propuesta, cuando un recorte se somete a este clasificador genérico (tras pasar o ser descartado por las etapas previas), el modelo es capaz de emitir una inferencia con un alto nivel de certeza, culminando exitosamente la fase de identificación del *pipeline*.

Tabla 20. Métricas del clasificador genérico «999»

Clase	Precision	Recall	F1-score	Muestras
a1	1.0	1.0	1.0	33
d36	1.0	1.0	1.0	45
d4	1.0	1.0	1.0	55
d58	1.0	0.9677	0.9836	31
e34	0.9697	1.0000	0.9846	32
f35	1.0	1.0	1.0	34
i9	1.0	0.9895	0.9947	95
m23	1.0	1.0	1.0	42
n33	1.0	1.0	1.0	35
o1	1.0	1.0	1.0	33
q1	1.0	1.0	1.0	35
q3	0.9778	1.0	0.9888	44
r8	1.0	1.0	1.0	47
s29	1.0	1.0	1.0	95
v13	1.0	1.0	1.0	33
v28	1.0	1.0	1.0	42
v30	1.0	0.9825	0.9912	57
v31	0.9815	1.0	0.9907	53
x8	1.0	1.0	1.0	31

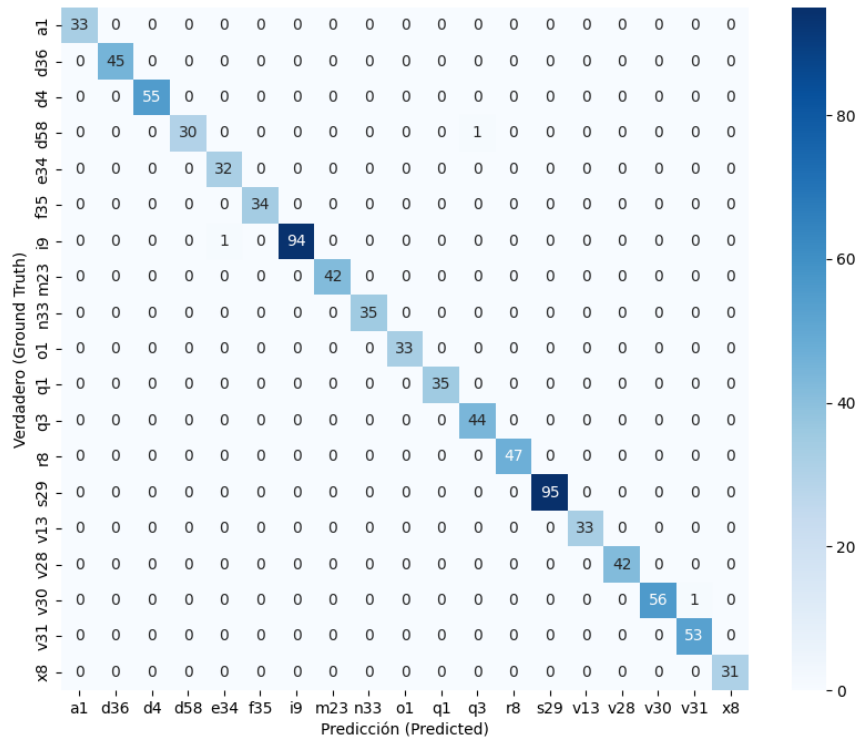


Figura 34. Matriz de confusión del clasificador genérico «999»

4.9 Resultados del proceso global de clasificación

El proceso global de clasificación diseñado consigue gestionar de manera eficiente la ambigüedad visual gracias a un proceso de adaptación a la realidad epigráfica y a la consecución de un sistema modular con inferencia en cascada. En la Tabla 21, se pueden observar los resultados obtenidos por cada clasificador en el conjunto de estelas de la Tabla 9, separando los datos que superan la confianza mínima exigida (C=Confiables) de los que no alcanzan el umbral (D=Dudosos). Los datos extendidos se encuentran en el Apéndice B.

En el flujo de clasificación final (véase Figura 5), se aplican estos clasificadores para obtener los resultados que superan el umbral de confianza y los dudosos, que se compilan en un archivo CSV y se plasman en una imagen de validación con cajas delimitadoras codificadas por color. Como se puede observar en la Figura 35, los glifos clasificados deben superar los umbrales de confianza mínimos definidos: 0,60 para el modelo de «Aves» con caja azul, 0,80 para el «Top 4» con caja amarilla y 0,85 para el clasificador genérico «999» con caja magenta. Los que no llegan a estos umbrales, se marcan en verde mostrando en su etiqueta la clase con mayor confianza a la que corresponde la predicción y acertando en muchos casos, lo que hace difícil cuantificar el número real de aciertos conseguido.

4.9.1 Recortes

Las inferencias generadas facilitan el etiquetado de nuevos conjuntos de datos, abriendo la vía a un proceso de especialización y mejora continua. Por ejemplo, las instancias identificadas como restos (véase Figura 36) pueden integrarse en futuras fases de entrenamiento para reforzar la representatividad estadística de las categorías en las que hay un menor número de muestras.



Figura 36. Recortes de resto de aves.

La revisión de los recortes, tanto los clasificados correctamente como los dudosos, permite ampliar la clase «*OTROS*» para posteriores ajustes finos. En la Figura 37, se puede apreciar un error puntual del clasificador donde el tercer glifo corresponde en realidad a un ojo (D4) y no a una boca (D21).



Figura 37. D21 clasificados correctamente, pero con algún D4.

Pero sin duda, el clasificador con más posibilidades es el genérico. La agrupación automatizada en directorios de glifos morfológicamente afines (Figura 38), facilita tanto la creación de nuevas categorías independientes como la detección de signos confusos, elementos fundamentales para generar una clase de ruido bien definida para futuras iteraciones del modelo.



Figura 38. Ejemplos clasificados como A1 o parecidos a A1.

5 Conclusiones y trabajo futuro

Una estrategia de etiquetado basada en la frecuencia y una evolución jerárquica del desarrollo ha vertebrado todo este estudio epigráfico. La combinación de modelos como YOLO11 y RF-DETR ha permitido generar una base sólida para aislar los jeroglíficos y producir recortes fiables que faciliten el posterior proceso de clasificación. De manera complementaria, el uso de un sistema de ventanas deslizantes ha permitido el procesamiento de estelas de grandes dimensiones, conservando la resolución estructural. Esta estrategia, al combinarse con técnicas de reescalado, permite optimizar la tasa de recuperación de los glifos. En las pruebas sobre diversos soportes, como piedra y papiro, se obtuvieron resultados cualitativos sólidos en un conjunto representativo, considerando la complejidad de las estelas y sus niveles habituales de degradación. Sin embargo, la validación cuantitativa de este estudio se centra en la evaluación independiente de las arquitecturas sobre recortes individuales y, aunque habría sido necesaria una evaluación *end-to-end* del sistema completo, la carencia de conjuntos de datos públicos anotados a nivel de estela completa en alta resolución es un condicionante insalvable actualmente, lo que hace necesario plantearlo como trabajo futuro. A pesar de ello, los resultados obtenidos indican que YOLO11 ofreció una mayor estabilidad en la identificación de glifos frecuentes, mientras que RF-DETR resultó más eficaz ante signos desconocidos o en zonas deterioradas.

En la fase de postprocesado, si bien el NMS multiescala maximiza la recuperación de signos, se observa que también provoca un aumento en la cantidad de decoración identificada como glifo, aunque la variante de máximos minimiza este problema. Se propone como trabajo futuro un filtrado de este ruido ornamental utilizando los datos de clasificación obtenidos, basándose en las dimensiones y características de cada elemento, así como en su concentración y distribución espacial.

En cuanto a la fase de clasificación, se ha intentado aprovechar al máximo el *dataset* desarrollado por Fuentes-Ferrer *et al.* (2025a), sin el cual no habría sido posible siquiera intentar una clasificación aproximada debido a la cantidad y variabilidad de clases. La distribución de muestras por clase en este conjunto de datos ha condicionado el desarrollo de los modelos de clasificación. La dificultad del etiquetado manual en los *datasets* de los modelos de localización también ha influido en su diseño. Por ejemplo, el signo Z1 (trazo vertical, con un 3,41 % de cobertura), excluido deliberadamente del clasificador «*Top 4*», a pesar de ser el cuarto glifo más frecuente en el corpus. Debido a su reducido tamaño y alta variabilidad, es un elemento propenso a la confusión y especialmente complejo de delimitar. Por su función estructural como marcador numérico, el signo Z1 forma parte de un trabajo futuro orientado al desarrollo de un modelo especializado en la detección de trazos simples. Mediante técnicas de análisis de distancias y agrupación geométrica, un sistema de este tipo podría inferir no solo el ideograma aislado (Z1), sino también sus variantes plurales (Z2 y Z3).

Superada la clasificación por frecuencia, el análisis debe orientarse hacia la funcionalidad semántica aislando jeroglíficos vertebradores como el M23. Al actuar como marcador de inicio sistemático en fórmulas de alabanza, su correcta identificación aporta un valor estructural decisivo para la traducción. Al estar ya integrado en el clasificador genérico, resulta viable extraer un subconjunto de datos para entrenar un futuro analizador de contexto. Un tratamiento similar presenta los sellos, cuya densidad compositiva confunde a los modelos de localización y exige un procesamiento aislado.

Este trabajo confirma que es posible unificar la localización y clasificación de jeroglíficos en un entorno operativo estable que, sin ser perfecto ni aspirar a ello, es suficiente para sentar las bases de una clasificación experta en profundidad, con posibilidad de mejora, y de una futura transliteración y desambiguación semántica que sirva de apoyo a los egiptólogos.

Bibliografía

- Akyon, F. C., Onur Altinuc, S., & Temizel, A. (2022). Slicing Aided Hyper Inference and Fine-Tuning for Small Object Detection. *2022 IEEE International Conference on Image Processing (ICIP)*, 966-970. <https://doi.org/10.1109/ICIP46576.2022.9897990>
- Barucci, A., Cucci, C., Franci, M., Loschiavo, M., & Argenti, F. (2021). A Deep Learning Approach to Ancient Egyptian Hieroglyphs Classification. *IEEE Access*, 9, 123438-123447. <https://doi.org/10.1109/ACCESS.2021.3110082>
- Cao, P., Zhou, F., Song, Q., & Yang, L. (2026). Controllable Generation with Text-to-Image Diffusion Models: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(4), 4771-4791. <https://doi.org/10.1109/TPAMI.2025.3646548>
- Creed, L. M. (2025). Neural style transfer for synthesising a dataset of ancient Egyptian hieroglyphs. arXiv. <https://doi.org/10.48550/arXiv.2504.02163>
- Choi, C.-Y., y Lee, S.-W. (2026). NeXt-DETR: A scalable and efficient transformer-based detector for resource-constrained systems. *Information Sciences*, 731, Artículo 122913. <https://doi.org/10.1016/j.ins.2025.122913>
- Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2015). The Pascal Visual Object Classes Challenge: A Retrospective. *International Journal of Computer Vision*, 111(1), 98-136.
- Franken, M., & Van Gemert, J. C. (2013). Automatic Egyptian hieroglyph recognition by retrieving images as texts. *Proceedings of the 21st ACM International Conference on Multimedia*, 765-768. <https://doi.org/10.1145/2502081.2502199>
- Fuentes-Ferrer, R. *et al.* (2025a). *EgyptianHieroglyphicText* [Conjunto de datos]. GitHub. <https://github.com/rfuentesfe/EgyptianHieroglyphicText>
- Fuentes-Ferrer, R., Duque-Domingo, J., y Herrera, P. J. (2025b). Recognition of Egyptian Hieroglyphic Texts through Focused Generic Segmentation and Cross-Validation Voting. *Applied Soft Computing*. <https://doi.org/10.1016/j.asoc.2025.112793>
- Fuentes-Jimenez, D., Pizarro, D., Hernández, Á., Bartoli, A., Méndez, C. G., Diego-Otón, L. de, Palazuelos-Cagigas, S., & Zamacona, C. G. (2025). *The OCR-PT-CT Project: Semi-Automatic Recognition of Ancient Egyptian Hieroglyphs Based on Metric Learning* (arXiv:2512.24197). arXiv. <https://doi.org/10.48550/arXiv.2512.24197>
- Gardiner, A. (1957). *Egyptian grammar: Being an introduction to the study of hieroglyphics*. Griffith Institute, Ashmolean Museum. http://web.ff.cuni.cz/ustavy/egyptologie/pdf/Gardiner_signlist.pdf
- González Hernández, A. J., Sánchez Hernández, J. P., Hernández Rabadán, D. L., Frausto Solis, J., & González Barbosa, J. (2025). An analysis of YOLO models versus RT-DETR applied to multi-object detection in images. *International Journal of Combinatorial Optimization Problems and Informatics*, 16(3), 360-377. <https://doi.org/10.61467/2007.1558.2025.v16i3.778>

- Gracia, Zamacona, Carlos. Manual de Egiptio Medio (segunda Edición), Archaeopress, 2018. ProQuest Ebook Central, <http://ebookcentral.proquest.com/lib/unedbiblioteca-ebooks/detail.action?docID=5845673>.
- Guidi, T., Python, L., Forasassi, M., Cucci, C., Franci, M., Argenti, F., & Barucci, A. (2023). Egyptian Hieroglyphs Segmentation with Convolutional Neural Networks. *Algorithms*, 16(2), 79. <https://doi.org/10.3390/a16020079>
- Hu, E. J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., & Chen, W. (2021). *LoRA: Low-Rank Adaptation of Large Language Models* (arXiv:2106.09685). arXiv. <https://doi.org/10.48550/arXiv.2106.09685>
- Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-Image Translation with Conditional Adversarial Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5967-5976. <https://doi.org/10.1109/CVPR.2017.632>
- Kang, B., Xie, S., Rohrbach, M., Yan, Z., Gordo, A., Feng, J., & Kalantidis, Y. (2020). *Decoupling Representation and Classifier for Long-Tailed Recognition* (arXiv:1910.09217). arXiv. <https://doi.org/10.48550/arXiv.1910.09217>
- Khanam, R., & Hussain, M. (2024). *YOLOv11: An Overview of the Key Architectural Enhancements* (arXiv:2410.17725). arXiv. <https://doi.org/10.48550/arXiv.2410.17725>
- Liao, S., Huang, C., Liang, Y., Zhang, H., y Liu, S. (2022). Solder joint defect inspection method based on ConvNeXt-YOLOX. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 12(11), 1890-1898. <https://doi.org/10.1109/TCPMT.2022.3224997>
- Liu, X., Meng, G., Xiang, S., y Pan, C. (2019). FontGAN: A unified generative framework for Chinese character stylization and de-stylization. arXiv. <https://doi.org/10.48550/arXiv.1910.12604>
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 9992-10002. <https://doi.org/10.1109/ICCV48922.2021.00986>
- Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). *A ConvNet for the 2020s* (arXiv:2201.03545). arXiv. <https://doi.org/10.48550/arXiv.2201.03545>
- Madan, M., & Reich, C. (2025). Strengthening Small Object Detection in Adapted RT-DETR Through Robust Enhancements. *Electronics*, 14(19), 3830. <https://doi.org/10.3390/electronics14193830>
- Nasser, A., Mohamed, M., Sherif, A., Mahmoud, B., Yehia, S., Saad, A., El-Rahmany, M. S., & Mohamed, E. H. (2025). *HieroGlyphTranslator: Automatic Recognition and Translation of Egyptian Hieroglyphs to English* (arXiv:2512.03817). arXiv. <https://doi.org/10.48550/arXiv.2512.03817>
- Padilla, R., Netto, S. L., & Da Silva, E. A. B. (2020). A Survey on Performance Metrics for Object-Detection Algorithms. *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 237-242. <https://doi.org/10.1109/IWSSIP48289.2020.9145130>

- Robinson, I., Robicheaux, P., Popov, M., Ramanan, D., & Peri, N. (2026). *RF-DETR: Neural Architecture Search for Real-Time Detection Transformers* (arXiv:2511.09554). arXiv. <https://doi.org/10.48550/arXiv.2511.09554>
- Roboflow. (s. f.). RF-DETR [Documentación de Modelo]. Recuperado <https://rfdetr.roboflow.com/develop/learn/install/>
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10674-10685. <https://doi.org/10.1109/CVPR52688.2022.01042>
- Sapkota, R., & Karkee, M. (2025). *Ultralytics YOLO Evolution: An Overview of YOLO26, YOLO11, YOLOv8 and YOLOv5 Object Detectors for Computer Vision and Pattern Recognition* (arXiv:2510.09653). arXiv. <https://doi.org/10.48550/arXiv.2510.09653>
- Sayyad, J., & Attarde, K. (2026). A systematic literature review on deep learning approaches for small object detection. *Array*, 29, 100615. <https://doi.org/10.1016/j.array.2025.100615>
- Schweitzer, S. D., y Ancient Egyptian Dictionary. (2019). *Corpus of Egyptian Texts for the AED – Ancient Egyptian Dictionary* (Versión 1.1.0-dev) [Conjunto de datos]. Zenodo. <https://doi.org/10.5281/zenodo.3580939>
- Shehzadi, T., Ifza, I., Liwicki, M., Stricker, D., & Afzal, M. Z. (2026). Semi-Supervised Object Detection: A Survey on Progress from CNN to Transformer. *Sensors*, 26(1), 310. <https://doi.org/10.3390/s26010310>
- Sobhy, A., Helmy, M., Khalil, M., Elmasry, S., Boules, Y., & Negied, N. (2023). An AI Based Automatic Translator for Ancient Hieroglyphic Language—From Scanned Images to English Text. *IEEE Access*, 11, 38796-38804. <https://doi.org/10.1109/access.2023.3267981>
- Tian, Y., Xu, W., Yang, B., Yang, X., Guo, H., Wang, G., & Yu, H. (2026). Development and evolution of YOLO in object detection: A survey. *Neurocomputing*, 669, 132436. <https://doi.org/10.1016/j.neucom.2025.132436>
- timm. (s. f.). *Convnext_base.fb_in22k_ft_in1k* [Modelo de visión artificial]. Recuperado https://huggingface.co/timm/convnext_base.fb_in22k_ft_in1k
- Ultralytics. (s. f.). *YOLO11* [Documentación de Modelo]. Recuperado <https://docs.ultralytics.com/es/models/yolo11/>
- University of St Andrews. (s. f.). *New Gardiner Egyptian Font*. <https://mjn.host.cs.st-andrews.ac.uk/egyptian/fonts/newgardiner.html>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention Is All You Need* (arXiv:1706.03762). arXiv. <https://doi.org/10.48550/arXiv.1706.03762>

- Wang, W., Dai, J., Chen, Z., Huang, Z., Li, Z., Zhu, X., Hu, X., Lu, T., Lu, L., Li, H., Wang, X., & Qiao, Y. (2023). *InternImage: Exploring Large-Scale Vision Foundation Models with Deformable Convolutions* (arXiv:2211.05778). arXiv. <https://doi.org/10.48550/arXiv.2211.05778>
- Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I. S., & Xie, S. (2023). ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16133-16142. <https://doi.org/10.1109/CVPR52729.2023.01548>
- Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I. S., & Xie, S. (2023). *ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders* (arXiv:2301.00808). arXiv. <https://doi.org/10.48550/arXiv.2301.00808>
- Xiong, H., Man, D., Lv, J., Xu, C., Zeng, F., Shi, Y., Lai, M., & Yang, W. (2026). Decoupling representation learning and classifier for long-tailed adversarial training. *Pattern Recognition*, 172, 112607. <https://doi.org/10.1016/j.patcog.2025.112607>
- Zhang, J., Liu, L., Silvén, O., Pietikäinen, M., & Hu, D. (2025). Few-Shot Class-Incremental Learning for Classification and Object Detection: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(4), 2924-2945. <https://doi.org/10.1109/TPAMI.2025.3529038>
- Zhang, L., Rao, A., & Agrawala, M. (2023). *Adding Conditional Control to Text-to-Image Diffusion Models* (arXiv:2302.05543). arXiv. <https://doi.org/10.48550/arXiv.2302.05543>
- Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y., & Chen, J. (2024). DETRs Beat YOLOs on Real-time Object Detection. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16965-16974. <https://doi.org/10.1109/CVPR52733.2024.01605>
- Zhao, Y., Peng, L., Yang, Y., Luo, Z., Li, H., Chen, Y., Yang, Z., He, X., Zhao, W., lu, qinglin, Wu, B., & Liu, W. (2024). *Local Conditional Controlling for Text-to-Image Diffusion Models* (arXiv:2312.08768). arXiv. <https://doi.org/10.48550/arXiv.2312.08768>
- Zhou, J., Zhang, B., Yuan, X., Lian, C., Ji, L., Zhang, Q., y Yue, J. (2023). YOLO-CIR: The network based on YOLO and ConvNeXt for infrared object detection. *Infrared Physics & Technology*, 131, Artículo 104703. <https://doi.org/10.1016/j.infrared.2023.104703>
- Zhou, S., Cai, Y., Zhang, Z., y Yin, J. (2025). MESC-DETR: An improved RT-DETR algorithm for steel surface defect detection. *Electronics*, 14(11), Artículo 2232. <https://doi.org/10.3390/electronics14112232>
- Zhu, J.-Y., Park, T., Isola, P., y Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. En *2017 IEEE International Conference on Computer Vision (ICCV)* (pp. 2242-2251). IEEE. <https://doi.org/10.1109/ICCV.2017.244>
- Zhu, J.-Y., Zhang, R., Pathak, D., Darrell, T., Efros, A. A., Wang, O., & Shechtman, E. (2018). *Toward Multimodal Image-to-Image Translation* (arXiv:1711.11586). arXiv. <https://doi.org/10.48550/arXiv.1711.11586>

Apéndice A: *Datasets* y configuraciones de entrenamiento

A.1 *Datasets*

Para el entrenamiento, validación y evaluación de las arquitecturas de detección (YOLO11 y RF-DETR), los datos se dividieron en tres conjuntos independientes atendiendo a los objetivos de la estrategia jerárquica: la clase de las «*Aves*», el núcleo de signos funcionales «*Top 4*» y la categoría genérica de caracteres desconocidos «*999*». Con un *split* de 70/20/10 aproximadamente, las imágenes originales se procesaron mediante una estrategia de fragmentación en mosaicos (tiles).

El subconjunto dedicado a las aves se estructuró a partir de 115 imágenes originales que dieron lugar a un total de 285 fragmentos. Los tiles de 1024×1024 px para YOLO y 1008×1008 px para RF-DETR se distribuyeron en 214 fragmentos (80 imágenes) para el conjunto de entrenamiento, 39 fragmentos (17 imágenes) para el de validación y 32 fragmentos (18 imágenes) para la fase de prueba.

El conjunto enfocado en la clase «*Top 4*» agrupó 98 imágenes fraccionadas en 233 tiles de 1024×1024 y 1008×1008 px, de los cuales 170 fragmentos (68 imágenes) se destinaron a entrenamiento, 32 fragmentos (15 imágenes) a validación y 31 fragmentos (15 imágenes) al test definitivo.

La categoría «*999*» se configuró a partir de 14 imágenes originales, divididas en tiles de 512×512 y 504×504 px, generando 1168 fragmentos, repartidos en 980 fragmentos (9 imágenes) para entrenamiento, 108 fragmentos (2 imágenes) para validación y 80 fragmentos (3 imágenes) para la evaluación de prueba.

La Tabla A.1 pormenoriza los fragmentos a nivel de imagen original, detallando la asignación y el número de muestras aportadas por cada estela a las distintas particiones del sistema.

Tabla A. 1. Listado de fragmentos utilizados en los datasets de localización

Nombre del Archivo	« <i>Aves</i> »			« <i>Top 4</i> »			« <i>999</i> »		
	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
Ancient_egyptian_stela_Senuserk III.jpg	1			1					
Basalt triumphal stela of Ramesses II... excavated at Beth Shean, 1290-1224 BCE (19th Dynasty) Egypt Penn Museum 02.jpg	2								
BCM H487 Stela of Meni.jpg	1						26		
Block statue of Mersuptah.jpg	1								
BM EA1262 Stela of Meni (Petrie).jpg	1			1					
Cippus (magical stela) MET 1989.234 EGDP017 779.jpg	2								
CUR.07.422 wwg8.jpg	2								
Decree of Pepi II.jpg		4		4					

Nombre del Archivo	«Aves»			«Top 4»			«999»		
	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
Donation Stela of Osorkon I dated to year 6 MET DP-12500-003.jpg		3		1				42	
Donation Stela of Ptolemy I.png			4	4					
Egypt Stela of Nacht-Mahes-eru.jpg	1			1					
Egyptian Limestone Stela (35799531723).jpg	2								
Egyptian Limestone Stela (36437605482).jpg	3			3					
Egyptian Limestone Stela (36561962736).jpg	5			4					
Egyptian Memorial Stela (7163624092).jpg	6			5					
Egyptian Painted Limestone Stela, 18th Dynasty (36437615082).jpg		3		3					
Egyptian Painted Limestone Stela, Deir el-Medina, 19th Dynasty (36368252212).jpg			2			2			
Egyptian Painted Wooden Stela of Neswy, Thebes, Ptolemaic (36437609542).jpg	2								
Egyptian Stela - British Museum.jpg			1	1					
Egyptian Stela (35701891104).jpg	2				1				
Egyptian Stela (35701893624).jpg	6				6				
Egyptian Stela (35727799283).jpg	2								
Egyptian Stela (35727806743).jpg	6					6			
Egyptian Stela (35727858233).jpg	1			1					
Egyptian Stela (36368414792).jpg	1			1					
Egyptian Stela (36368419572).jpg	6			6					
Egyptian Stela (36399315121).jpg	3				3				
Egyptian Stela (36490530346).jpg	2								
Egyptian Stela (36536788335).jpg		4		4					
Egyptian Stela (36536865995).jpg		5		4					
Egyptian Stela (48464726186).jpg	1			1					
Egyptian Stela (48464795481).jpg			6	5					
Egyptian Stela (48464822826).jpg		2							
Egyptian Stela (48464876702).jpg	3			3					
Egyptian Stela (48464880842).jpg	3			1					
Egyptian Stela (48464986242).jpg	2			2					

Nombre del Archivo	«Aves»			«Top 4»			«999»		
	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
Egyptian Stela Dedicated by Priest Neferrenpet to God Ptah, Deir-el-Medina, Dynasty 20, 1196-1070 BC (48464807786).jpg		1		1					
Egyptian stela of Hetep and Khnumu, World Museum Liverpool.jpg		1		1					
False Door of the Royal Sealer Neferiu.jpg	1			1					
Fragment of Henenu Stela.jpg			1	1					
Fragment of the stela of Horira.jpg		1		1					
Funerary stela of Thutirtis, born of Kerbet.jpg	3			2					
Funerary stela of Thutirtis, born of Kerbet () MET DP241029.jpg	1			1			64		
Inscribed Lintel from the Tomb of the Overseer of Priests and Keeper of the Sacred Cattle Mereri.jpg		3		3					
Inventory Stela (1872).png			1		1				
Limestone stela of Sobekhotep.jpg	10			7					
Magical Stela (Cippus of Horus) MET EG546.jpg	8			6					
Metternich stela 1.jpg			2		2				
Miniature Stela of Ahmose.jpg			1			1			
Painted stela of Djedbastet MET 2 2.3.34 EGDP021580.jpg		1		1					
Ptolemaic limestone funerary stela from Thebes at the Gregorian Egyptian Museum.jpg	6				6				
Relief from an Offering Niche.jpg	1			1					
Small Magical Stela with Shed dedicated by Nesamenopet, son of Djedkhonsuifankh MET 44.4.54 EGDP015274.jpg			1		1				40
Stela of Djemi.jpg	4			4					
Stela of Four Women.jpg	2			2					
Stela of Inamennavefnebu.jpg	1			1					
Stela of Itubaal and Masutu.jpg			2	2					
Stela of Ptahmose.jpg	3			3					
Stela of Saiah.jpg	1			1					
Stela of the Steward Mentuwser.jpg	1					1			

Nombre del Archivo	«Aves»			«Top 4»			«999»		
	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
Stela of Tjanetiset.jpg	1			1					
Stela dedicated by Huy to Amun-Re, limestone - Museo Egizio, Turin C 1607 p01.jpg	1					1	46		
Stela dedicated by Pendua to Meretseger, limestone - Museo Egizio Turin C 1564 p01.jpg		1		1					
Stela dedicated by Smen to Re-Horakhty, limestone - Museo Egizio, Turin C 1648 p01.jpg	1								4
Stela in the Shape of a Py-lon MET LC-2016 77 EGD029832.jpg									
Stela Manchester 1781 Petrie.jpg			1	1					
Stela of Aafenmut MET DP370296.jpg	2			1					
Stela of Aafenmut MET eg28.3.35.jpg	2			1					
Stela of Aamtju MET 19.3.32 EGD022605.jpg	1				1			66	
Stela of Ahmose MET 20.2.4 front.jpg	4			3					
Stela of Amen-nakht MET 65.269 EGD022604.jpg	1				1				
Stela of Dedu MET 16.10.333.jpg	1				1				
Stela of Djehutynefer, scribe of the treasury, limestone - Museo Egizio Turin C 1456 p02.jpg			1		1				
Stela of Four Women MET LC-36 3 270 EGD029178.jpg	9			9			78		
Stela of Hay, limestone - Museo Egizio, Turin C 1606 p01.jpg		2		2					
Stela of Hay, limestone - Museo Egizio, Turin C 1606 p04.jpg	6					5			
Stela of Hetepiqr () and his wife, limestone - Museo Egizio, Turin S 1277 p01.jpg	1				1				
Stela of Horem-khauf MET 101714.jpg	2								
Stela of Huy, limestone - Museo Egizio Turin C 1608 p01.jpg	3			3					
Stela of Huy, limestone - Museo Egizio, Turin C 1463 p01.jpg		1							
Stela of Intef and Nesumontu MET 16.10.327.jpg	4			2					

Nombre del Archivo	«Aves»			«Top 4»			«999»		
	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
Stela of Kemes MET DP340537.jpg	3			3					
Stela of Kemes MET DP340554.jpg	4				4				
Stela of Kemes MET DP340556.jpg	1								
Stela of Kemes MET DP340559.jpg	2			1					
Stela of Kemes MET DP340563.jpg	2					1			
Stela of Khonsu MET 21.2.69 EG DP022530.jpg			1						
Stela of Maya, limestone - Museo Egizio, Turin C 1579 p01.jpg		3		3					
Stela of Merneptah MET 04.2.533 EGD022601.jpg	6			6					
Stela of Mery, limestone - Museo Egizio, Turin S 1269 p01.jpg			1						
Stela of Nakht MET LC-11 155 1 EGD029179.jpg	1			1					
Stela of Neferabu, limestone - Museo Egizio, Turin C 1593 p01.jpg	3					3			
Stela of Neferrenpet and Huinfret, limestone - Museo Egizio, Turin C 1592 p01.jpg	3					3			
Stela of Nemtiui.jpg			2	2					
Stela of Ptahmose MET LC-67 3 EGD028267.jpg	2					2			
Stela of Rehirdjersen MET EG41.jpg	3			3					
Stela of Reniseneb MET 63.154.jpg	1					1			
Stela of Senu Adoring Osiris MET 18.2.5 05.jpg	1			1			34		
Stela of Senu MET 12.182.39 01.jpg	1				1		22		
Stela of Shuamay Abydos 1292-1190 BCE Dynasty 19 Egypt Penn Museum 02.jpg	3			3					
Stela of the bowman Khuiui, limestone - Museo Egizio, Turin S 1276 p01.jpg	3			3					
Stela of the Gatekeeper Maati MET DP352181.jpg			2						
Stela of the Gatekeeper Maati MET DT292973.jpg	5			5					

Nombre del Archivo	«Aves»			«Top 4»			«999»		
	Train	Valid	Test	Train	Valid	Test	Train	Valid	Test
Stela of the God's-Father of Amun Pakeshi MET 90.6.30 EGD015368.jpg	1			1			76		
Stela of the Overseer of the Fortress Intef MET DP322110.jpg	2			2			626		
Stela of the sculptor Qen, limestone - Museo Egizio, Turin C 1635 p01.jpg		2				2			
Stela of the standard-bearer Maienhekau, limestone - Museo Egizio, Turin C 1459 p01.jpg	2					1			
Stela of the Steward Mentuwer MET DP322064.jpg	8			8					
Stela of the Steward Mentuwer MET DP322066.jpg	3			3					
Stela of the Steward Mentuwer MET DT527.jpg	2				2				
Stela Shabaqo Met.jpg	2			2					
Stela.jpg			2	1					
Tomb Stela of Hetepsi.jpg			1	1					
Tomb Stela of Hetepsi MET 98.4.6_01.jpg	1					1			36
Unfinished Stela to Amun-Re MET DP234728.jpg		2		2			8		
Votive stela of Userhat MET DP226728.jpg	1					1			

A2. Configuraciones de entrenamiento

YOLO

Para los modelos localizadores, se estableció la configuración de hiperparámetros mediante el comando de entrenamiento estándar de la librería ultralytics:

«Aves»

```
yolo task=detect mode=train model=yolo11l.pt data=dataset_yolo_birds_corregido/data.yaml epochs=200 patience=30 batch=8
imgsz=1024 device=0 workers=0 cos_lr=true name=train_birds
```

«Top 4»

```
yolo task=detect mode=train model=yolo11l.pt data=dataset_yolo_top4_corregido/data.yaml epochs=200 patience=30 batch=8
imgsz=1024 device=0 workers=0 cos_lr=true name=train_top4
```

«999»

```
yolo task=detect mode=train model=yolo11l.pt data=dataset_yolo_999_corregido/data.yaml epochs=200 patience=30 batch=32
imgsz=512 device=0 workers=0 cos_lr=true name=train_999
```

RFDETR «Aves»/«Top 4»

```
model.train(
    dataset_dir="dataset_rtdetr_birds",
    epochs=200,
    num_workers = 0,
    image_size=1024,
    resolution=1008, #multiplo de 56
    batch_size=4,
    grad_accum_steps=4,
    lr=1e-4,
    output_dir="runs",
    early_stopping=True,
    early_stopping_patience=30
)
```

RFDETR «999»

```
model.train(
    dataset_dir="dataset_rfdetr_999",
    epochs=200,
    num_workers = 0,
    image_size=512,
```

```

        resolution=504,
        batch_size=4,
        grad_accum_steps=8,
        lr=1e-4,
        output_dir="runs_rfdetr_999_b4_g8",
        early_stopping=True,
        early_stopping_patience=30
        #tensorboard=True
    )

```

AUMENTO DE DATOS

CycleGAN

```

#cycleGAN G1
python train.py --dataroot datasets/cycleGAN_AB_sketch2glyph --name
cycleGAN_s2g --model cycle_GAN --dataset_mode aligned --ngf 64 --ndf
64 --netG resnet_9blocks --netD basic --direction AtoB --gpu_ids 0
#--batch_size 2 --niter 100 --niter_decay 100 --GAN_mode lsgan --lr
0.0002 --beta1 0.5

```

Pix2Pix

```

#pix2pix G2
python train.py --dataroot datasets/pix2pix_sketch2glyph --name
pix2pix_s2g_AB --model pix2pix --netG unet_256 --direction BtoA --
netD n_layers --n_layers_D 4 --norm instance

```

BicycleGAN

```

#B1_G1
python train.py --dataroot datasets/G1_dataset --name G1 --model bi-
cycle_GAN --netG unet_256 --nz 32 --niter 100 --niter_decay 100 --
batch_size 2 --netD basic_256_multi --gpu_ids 0 --lambda_L1 10

```

```

#B2_G2
python train.py --dataroot datasets/G2_dataset --name G2 --model bi-
cycle_GAN --netG unet_256 --nz 32 --niter 100 --niter_decay 100 --
batch_size 2 --netD basic_256_multi --gpu_ids 0 --lambda_L1 10

```

```

#B3_G3
python train.py --dataroot datasets/G3_dataset --name G3 --model bi-
cycle_GAN --netG unet_256 --nz 32 --niter 100 --niter_decay 100 --
batch_size 2 --netD basic_256_multi --gpu_ids 0 --lambda_L1 10

```

CLASIFICADORES

ConvNeXT «Aves»/«Top 4»/«999»

```
CONFIG = {  
    'data_dir': '9_datasets_birds',  
    'model_name': 'convnext_base.fb_in22k_ft_in1k',  
    'batch_size': 16,  
    'img_size': 224,  
    'epochs': 100,  
    'learning_rate': 5e-5,  
    'weight_decay': 1e-3,  
    'num_workers': 2,  
    'device': 'cuda',  
    'patience': 10  
}
```


Apéndice B: Estelas y papiros analizados

Este apéndice contiene los resultados visuales y cuantitativos pormenorizados obtenidos por el *pipeline*, generados a partir de las imágenes de dominio público.

B1. Stela of Nya

Objeto: Estela de Nya.

Referencia: Museo Egizio de Turín (Cat. 1585).

URL: https://collezioni.museoegizio.it/en-GB/material/Cat_1585

Características: Dinastía XIX (ca. 1292–1191 a. C.). Piedra caliza (35 × 67 × 11 cm).

Imagen: 00135N_68F0E49E946C1C322D5E864BEDC706A3_full.jpg (4912 × 6747 px).

Tabla B 1. Resultados de detección global y optimización de escalas.

Modelo	NMS Global	Escala 1.0x (Prioridad 1)	Escala 0.5x (Prioridad 2)	NMS Escala
(RF) «Aves»	15	10	6	11
(YO) «Aves»	8	3	10	6
(YO) «Top4»	75	27	31	39
(RF) «999»	614	435	363	449
Total	712	475	410	505

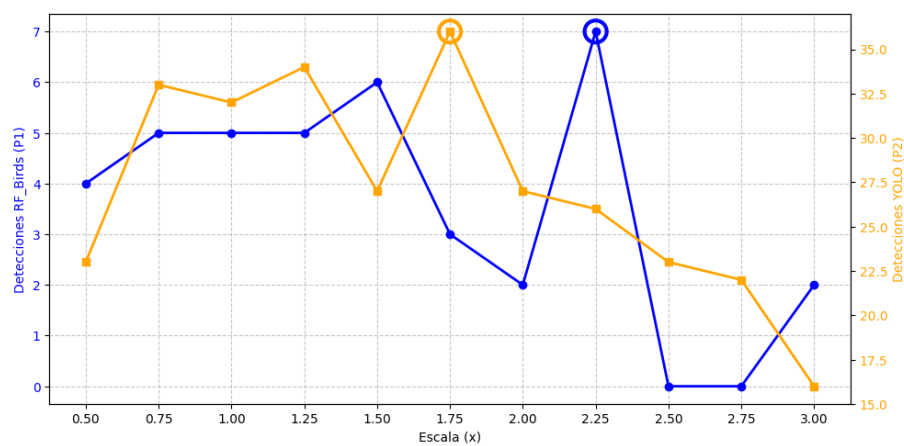


Figura B. 1. Detecciones por escala RF-DETR (azul) y YOLO (naranja).

Tabla B 2. Resultados de clasificación experta (ConvNeXt)

Modelo	Etiquetado superando umbral de confianza	Etiquetado dudoso
(CN) «Aves»	8	9
(CN) «Top4»	30	3
(CN) «999»	162	293



Figura B. 2. Detecciones sobre estela con NMS global.



Figura B. 3. Detecciones sobre estela utilizando NMS con filtrado de prioridad.



Figura B. 4. Clasificación sobre detecciones.

B2. Round-topped stele of Ipuý

Objeto: Estela de Ipuý con remate redondeado.

Referencia: Round-topped stele of Ipuý (shown below with wife and child). Above Hathor, Osiris, Re-Harakhte and King. Museo Egizio de Turín (Cat. 7357).

URL: https://collezioni.museoegizio.it/en-GB/material/Cat_7357

Características: Dinastía XIX (ca. 1292–1190 a. C.). Piedra caliza (38.4 x 57 x 7.4 cm).

Imagen: 00155B_A5E26C3412D62F4A0EDAA6A00756F7ED_full.jpg (4999 × 6694 px).

Tabla B 3. Resultados de detección global y optimización de escalas.

Modelo	NMS Global	Escala 2.25x (Prioridad 1)	Escala 1.75x (Prioridad 2)	NMS Escala
(RF) « <i>Aves</i> »	11	7	3	9
(YO) « <i>Aves</i> »	8	1	4	2
(YO) « <i>Top4</i> »	63	25	32	37
(RF) « <i>999</i> »	242	153	147	174
Total	324			222

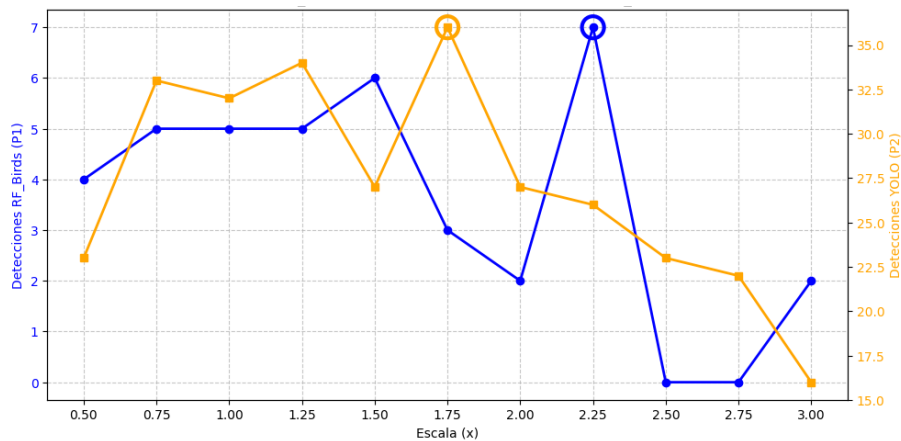


Figura B. 5. Detecciones por escala RF-DETR (azul) y YOLO (naranja).

Tabla B 4. Resultados de clasificación experta (ConvNeXt)

Modelo	Etiquetado superando umbral de confianza	Etiquetado dudoso
(CN) « <i>Aves</i> »	8	3
(CN) « <i>Top4</i> »	27	5
(CN) « <i>999</i> »	58	121



Figura B. 6. Detecciones sobre estela con NMS global.



Figura B. 7. Detecciones sobre estela utilizando NMS con filtrado de prioridad



Figura B. 8. Clasificación sobre detecciones.

B3. Stela dedicated to the goddess Renenutet by Mutneferet and Iyneferet

Objeto: Estela dedicada a la diosa Renenutet por Mutneferet e Iyneferet.

Referencia: Stela dedicated to the goddess Renenutet by Mutneferet and Iyneferet. Museo Egizio de Turín (Cat. 6138). https://collezioni.museoegizio.it/en-GB/material/S_6138

Características: Dinastía XIX (ca. 1292–1190 a. C.). Piedra caliza (16 x 23.5 x 4 cm).

Imagen: 0015H9_7E7E5123EC4A25C9A4B8D1A754B1F045_full.jpg (4271 × 5482 px).

Tabla B 5. Resultados de detección global y optimización de escalas.

Modelo	NMS Global	Escala 0.75x (Prioridad 1)	Escala 0.75x (Prioridad 2)	NMS Escala
(RF) «Aves»	3	3	3	3
(YO) «Aves»	1	0	0	0
(YO) «Top4»	29	15	15	15
(RF) «999»	58	26	26	26
Total	91	44	44	44

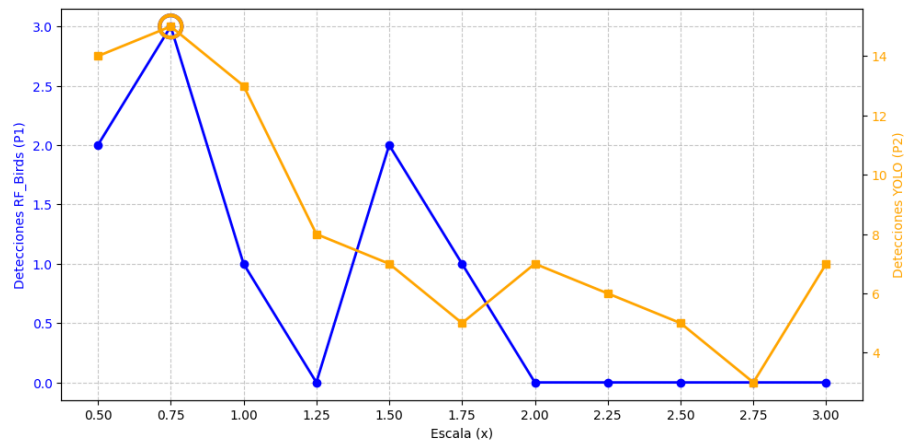


Figura B. 9. Detecciones por escala RF-DETR (azul) y YOLO (naranja).

Tabla B 6. Resultados de clasificación experta (ConvNeXt)

Modelo	Etiquetado superando umbral de confianza	Etiquetado dudoso
(CN) «Aves»	3	0
(CN) «Top4»	11	2
(CN) «999»	12	16



Figura B. 10. Detecciones sobre estela con NMS global.



Figura B. 11. Detecciones sobre estela utilizando NMS con filtrado de prioridad



Figura B. 12. Clasificación sobre detecciones.

B4. Stela of Neferabu

Objeto: Estela de Neferabu.

Referencia: Stela of Neferabu. Museo Egizio de Turín (Cat. 1593). https://collezioni.museoegizio.it/en-GB/material/Cat_1593

Características: Dinastía XIX (ca. 1292–1191 a. C.). Piedra caliza (54 x 21 x 6.2 cm).

Imagen: 0015HL_609A3EFA50B402788C12D6469B761004_full.jpg (7096 × 3047 px).

Tabla B 7. Resultados de detección global y optimización de escalas.

Modelo	NMS Global	Escala 0.75x (Prioridad 1)	Escala 0.5x (Prioridad 2)	NMS Escala
(RF) «Aves»	40	35	26	35
(YO) «Aves»	14	6	17	10
(YO) «Top4»	108	59	59	75
(RF) «999»	504	375	353	402
Total	666	475	455	522

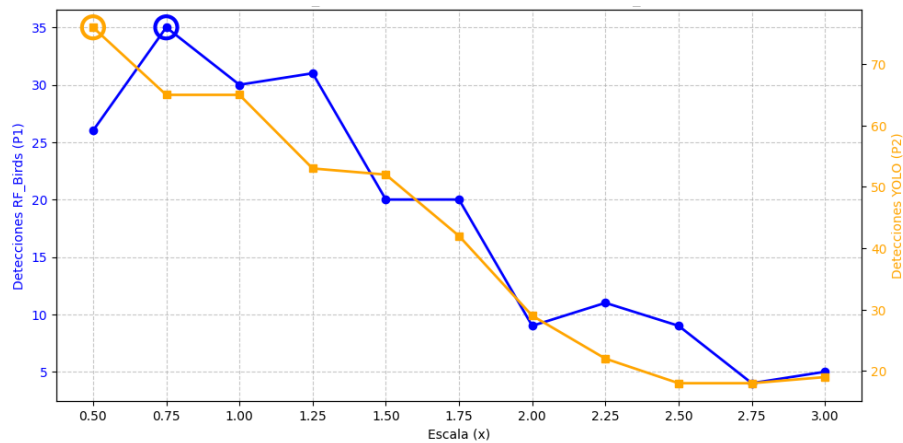


Figura B. 13. Detecciones por escala RF-DETR (azul) y YOLO (naranja).

Tabla B 8. Resultados de clasificación experta (ConvNeXt)

Modelo	Etiquetado superando umbral de confianza	Etiquetado dudoso
(CN) «Aves»	33	12
(CN) «Top4»	65	2
(CN) «999»	146	264



Figura B. 14. Detecciones sobre estela con NMS global.



Figura B. 15. Detecciones sobre estela utilizando NMS con filtrado de prioridad



Figura B. 16. Clasificación sobre detecciones.

B5. Stela of Intef and Nesumontu

Objeto: Estela de Intef y Nesumontu.

Referencia: Stela of Intef and Nesumontu. Met Museum (Cat. 546944). <https://www.met-museum.org/art/collection/search/546944>

Características: Dinastía XII (ca. 1961–1917 a. C.). Piedra caliza/Pintura (32 x 27 cm).

Imagen: 16.10.327.jpg (2546 × 3371 px).

Tabla B 9. Resultados de detección global y optimización de escalas.

Modelo	NMS Global			NMS Escala
(RF) «Aves»	7	5	3	5
(YO) «Aves»	4	3	6	5
(YO) «Top4»	29	21	19	21
(RF) «999»	73	57	58	57
Total	113	86	83	88

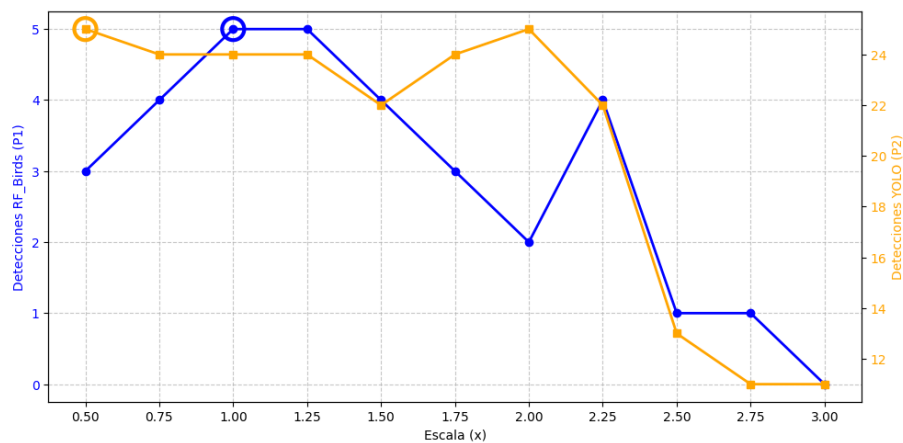


Figura B. 17. Detecciones por escala RF-DETR (azul) y YOLO (naranja).

Tabla B 10. Resultados de clasificación experta (ConvNeXt)

Modelo	Etiquetado superando umbral de confianza	Etiquetado dudoso
(CN) «Aves»	3	7
(CN) «Top4»	20	0
(CN) «999»	27	31



Figura B. 18. Detecciones sobre estela con NMS global.



Figura B. 19. Detecciones sobre estela utilizando NMS con filtrado de prioridad



Figura B. 20. Clasificación sobre detecciones.

B6. False-door stela of Harbes

Objeto: Estela de la falsa puerta de Harbes.

Referencia: False-door stela of Harbes. Museo Egizio de Turín (Cat. 17161). https://collezioni.museoegizio.it/en-GB/material/S_17161

Características: Dinastía XXV (ca. 722–525 a. C.). Piedra caliza (71.3 x 92.5 x 10 cm).

Imagen: 001347_3E9C9A5F1E308D34DE40E775ABAD843C_full.jpg (1008 × 1377 px).

Tabla B 11. Resultados de detección global y optimización de escalas.

Modelo	NMS Global	Escala 0.5x (Prioridad 1)	Escala 2.5x (Prioridad 2)	NMS Escala
(RF) «Aves»	0	0	0	0
(YO) «Aves»	1	1	0	1
(YO) «Top4»	20	15	18	18
(RF) «999»	67	56	56	58
Total	88	72	74	77

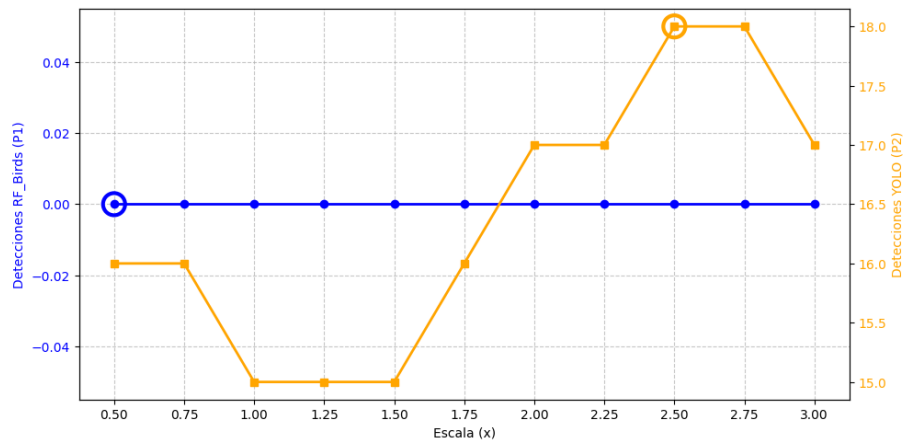


Figura B. 21. Detecciones por escala RF-DETR (azul) y YOLO (naranja).

Tabla B 12. Resultados de clasificación experta (ConvNeXt)

Modelo	Etiquetado superando umbral de confianza	Etiquetado dudoso
(CN) «Aves»	1	0
(CN) «Top4»	16	0
(CN) «999»	29	31



Figura B. 22. Detecciones sobre estela con NMS global.



Figura B. 23. Detecciones sobre estela utilizando NMS con filtrado de prioridad

Figura B. 24. Clasificación sobre detecciones.

B7. Magical stela or cippus of Horus (Reverso)

Objeto: Estela mágica o cipo de Horus.

Referencia: Magical stela or cippus of Horus (Cat. 545766). <https://www.metmuseum.org/art/collection/search/545766>

Características: Período ptolemaico (ca. 332–280 a. C.). Esquisto de clorita (20.5 x 12 x 4.5 cm).

Imagen: DP243458.jpg (3001 × 4000 px).

Tabla B 13. Resultados de detección global y optimización de escalas.

Modelo	NMS Global			NMS Escala
(RF) «Aves»	46	34	14	35
(YO) «Aves»	12	3	35	16
(YO) «Top4»	183	104	140	150
(RF) «999»	518	441	286	438
Total	759	582	475	639

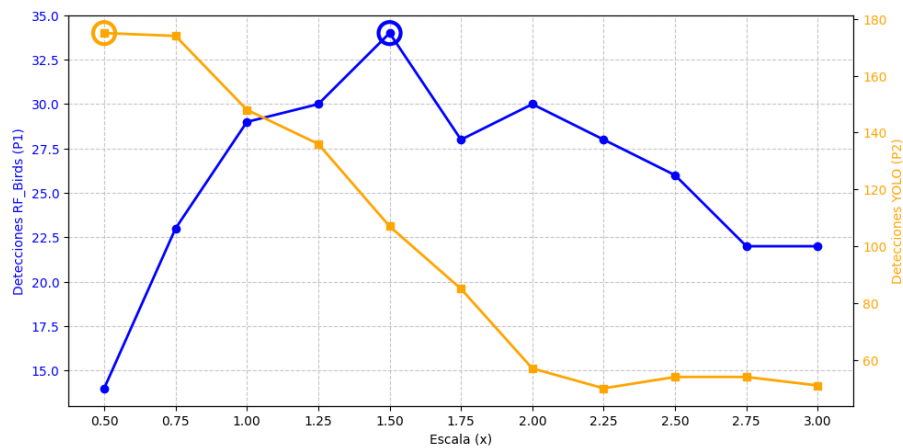


Figura B. 25. Detecciones por escala RF-DETR (azul) y YOLO (naranja).

Tabla B 14. Resultados de clasificación experta (ConvNeXt)

Modelo	Etiquetado superando umbral de confianza	Etiquetado dudoso
(CN) «Aves»	29	22
(CN) «Top4»	142	5
(CN) «999»	198	243



Figura B. 26. Detecciones sobre estela con NMS global.



Figura B. 27. Detecciones sobre estela utilizando NMS con filtrado de prioridad

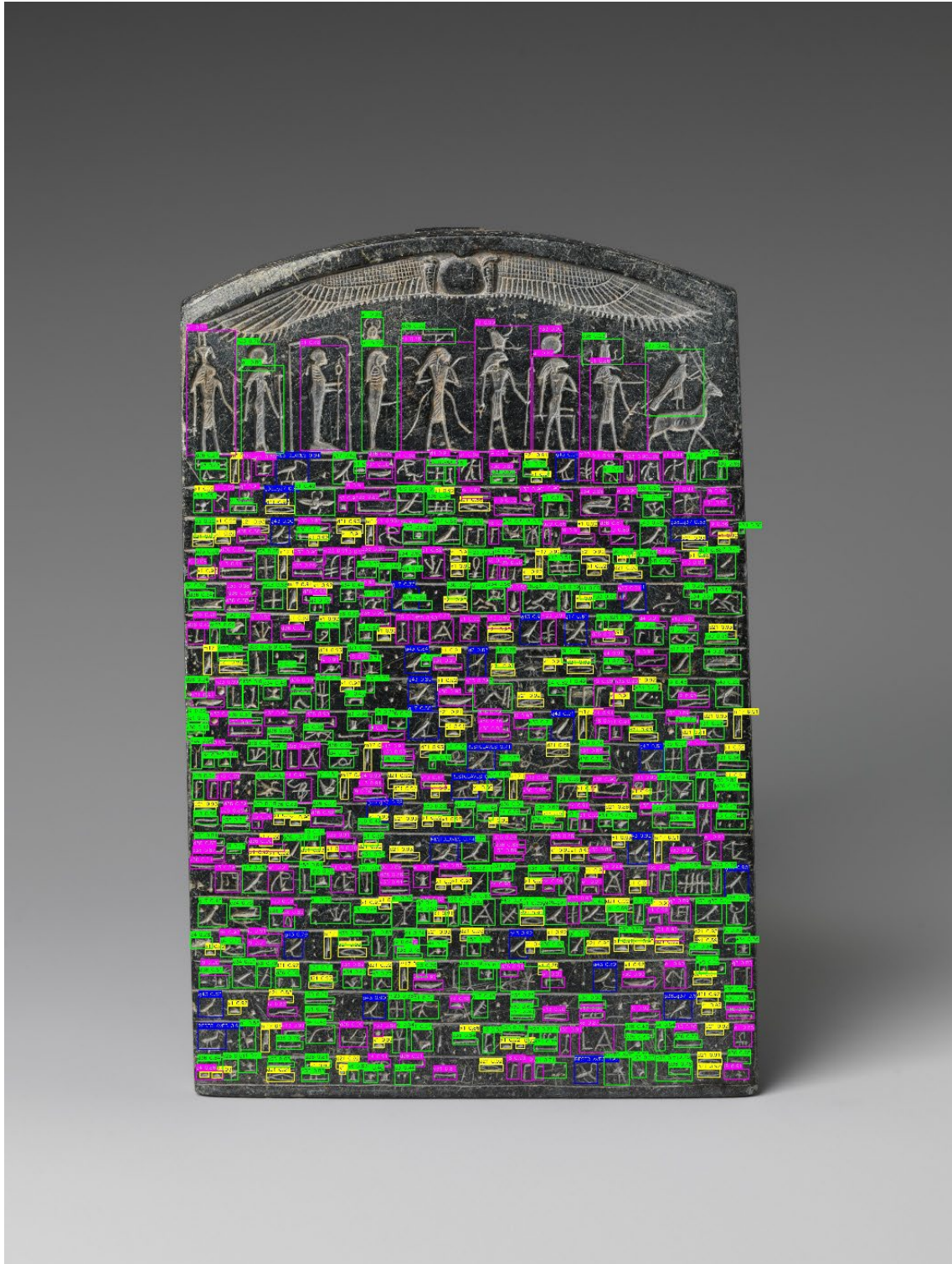


Figura B. 28. Clasificación sobre detecciones.

B8. Stela of the lector priest of Amun Siamun

Objeto: Estela del sacerdote lector de Amón, Siamón.

Referencia: Stela of the lector priest of Amun Siamun and his mother the singer Amenhotep (Cat. 555406). <https://www.metmuseum.org/art/collection/search/555406>

Características: Dinastía XVIII (ca. 1420 a. C.). Piedra caliza/Pintura (32.5 x 20.5 cm).

Imagen: DT2928.jpg (2978 × 3722 px).

Tabla B 15. Resultados de detección global y optimización de escalas.

Modelo	NMS Global	Escala 0.5x (Prioridad 1)	Escala 1.5x (Prioridad 2)	NMS Escala
(RF) «Aves»	8	7	5	7
(YO) «Aves»	2	0	3	1
(YO) «Top4»	42	27	28	31
(RF) «999»	135	104	80	107
Total	187	138	116	146

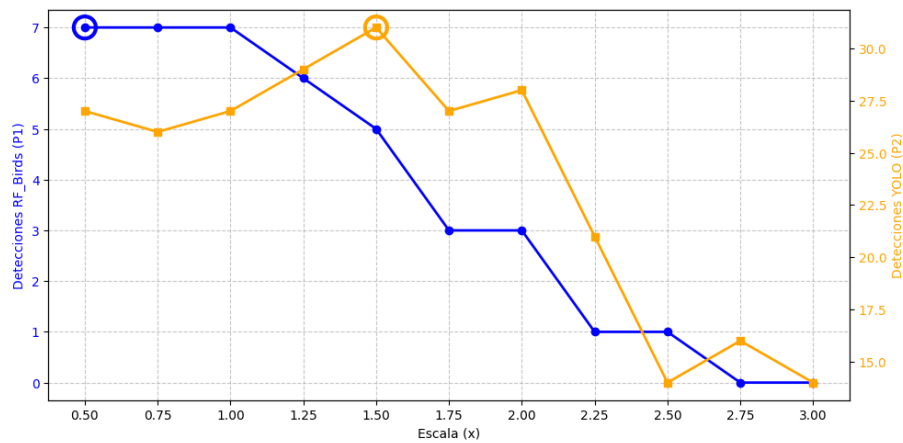


Figura B. 29. Detecciones por escala RF-DETR (azul) y YOLO (naranja).

Tabla B 16. Resultados de clasificación experta (ConvNeXt)

Modelo	Etiquetado superando umbral de confianza	Etiquetado dudoso
(CN) «Aves»	7	1
(CN) «Top4»	29	0
(CN) «999»	32	77



Figura B. 30. Detecciones sobre estela con NMS global.



Figura B. 31. Detecciones sobre estela utilizando NMS con filtrado de prioridad



Figura B. 32. Clasificación sobre detecciones.

B9. Stela of the temple servant of Amun Irthareru

Objeto: Estela del sirviente del templo de Amón, Irthareru

Referencia: Stela of the temple servant of Amun Irthareru adoring Re-Harakhty and Atum (Cat. 550792). <https://www.metmuseum.org/art/collection/search/550792>

Características: Dinastía XXV-XXVI (ca. 750–525 a. C.). Madera/Yeso/Pintura (35 x 24 x 2 cm).

Imagen: DP-34997-001.jpg (2978 × 3722 px).

Tabla B 17. Resultados de detección global y optimización de escalas.

Modelo	NMS Global	Escala 1.75x (Prioridad 1)	Escala 2.0x (Prio- ridad 2)	NMS Escala
(RF) «Aves»	21	14	8	14
(YO) «Aves»	6	6	10	7
(YO) «Top4»	64	31	35	39
(RF) «999»	402	295	294	344
Total	493	346	347	404

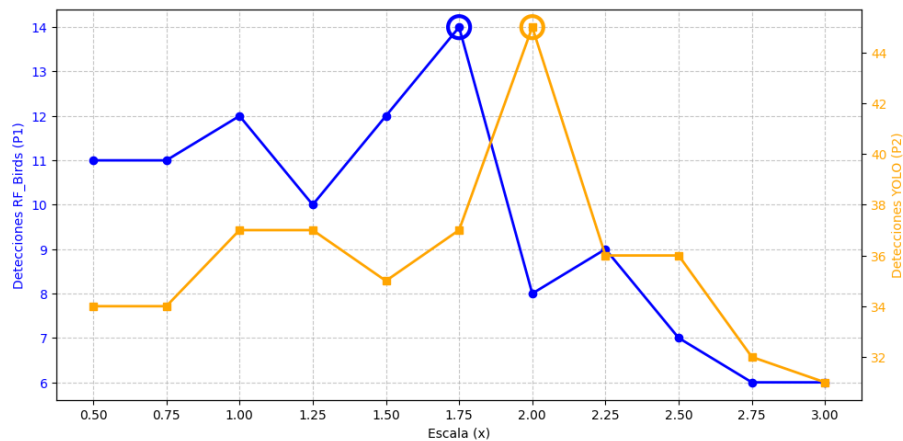


Figura B. 33. Detecciones por escala RF-DETR (azul) y YOLO (naranja).

Tabla B 18. Resultados de clasificación experta (ConvNeXt)

Modelo	Etiquetado superando umbral de confianza	Etiquetado dudoso
(CN) «Aves»	11	10
(CN) «Top4»	24	11
(CN) «999»	183	165



Figura B. 34. Detecciones sobre estela con NMS global.



Figura B. 35. Detecciones sobre estela utilizando NMS con filtrado de prioridad



Figura B. 36. Clasificación sobre detecciones.

B10. Stela of Horemkhauf

Objeto: Estela de Horemjauef.

Referencia: Stela of Horemkhauf (Cat. 544364). <https://www.metmuseum.org/art/collection/search/550792>

Características: Dinastía XIII (ca. 1802–1640 a. C.). Piedra caliza (58 x 35 x 15 cm).

Imagen: DP-34999-001.jpg (2750 × 4000 px).

Tabla B 19. Resultados de detección global y optimización de escalas.

Modelo	NMS Global	Escala 1.25x (Prioridad 1)	Escala 2.75x (Prioridad 2)	NMS Escala
(RF) «Aves»	60	46	19	48
(YO) «Aves»	13	17	16	15
(YO) «Top4»	206	133	160	170
(RF) «999»	515	482	391	480
Total	794	678	586	713

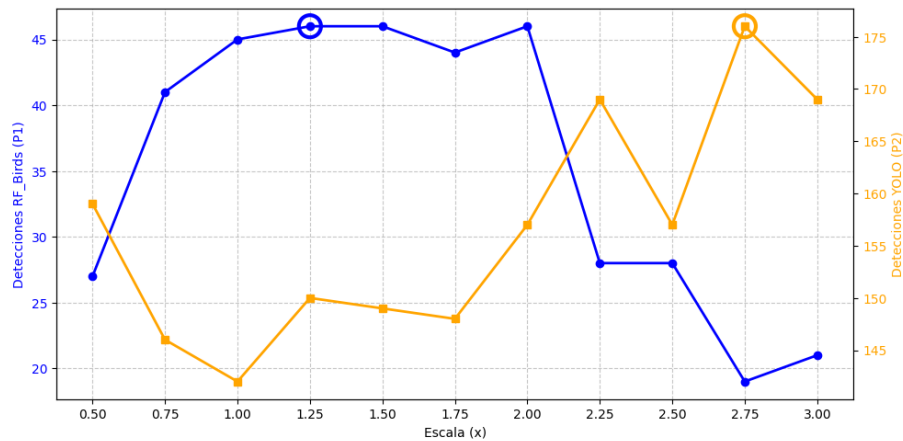


Figura B. 37. Detecciones por escala RF-DETR (azul) y YOLO (naranja).

Tabla B 20. Resultados de clasificación experta (ConvNeXt)

Modelo	Etiquetado superando umbral de confianza	Etiquetado dudoso
(CN) «Aves»	48	15
(CN) «Top4»	141	4
(CN) «999»	270	235



Figura B. 38. Detecciones sobre estela con NMS global.



Figura B. 39. Detecciones sobre estela utilizando NMS con filtrado de prioridad



Figura B. 40. Clasificación sobre detecciones.

B11. Magical Stela with Horus the Child (Anverso)

Objeto: Estela mágica con Horus Niño (Anverso).

Referencia: Magical Stela with Horus the Child (Cat. 546037). <https://www.met-museum.org/es/art/collection/search/546037>

Características: Dinastía XXX (ca. 360–343 a. C.). Metagrauvaca (a. total 83,5 cm).

Imagen: DP-38634-001.jpg (3000 × 4000 px).

Tabla B 21. Resultados de detección global y optimización de escalas.

Modelo	NMS Global			NMS Escala
(RF) «Aves»	170	151	40	152
(YO) «Aves»	55	36	152	54
(YO) «Top4»	760	547	579	626
(RF) «999»	2067	1938	885	1917
Total	3052	2672	1656	2749

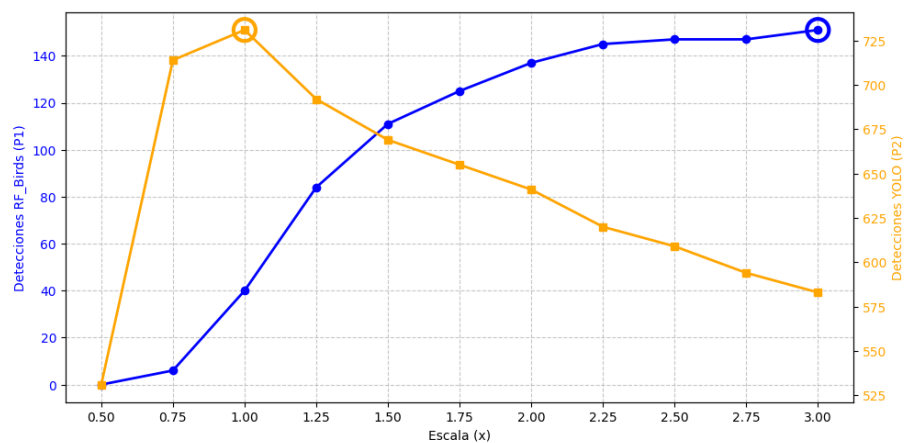


Figura B. 41. Detecciones por escala RF-DETR (azul) y YOLO (naranja).

Tabla B 22. Resultados de clasificación experta (ConvNeXt)

Modelo	Etiquetado superando umbral de confianza	Etiquetado dudoso
(CN) «Aves»	130	76
(CN) «Top4»	534	68
(CN) «999»	759	1182



Figura B. 42. Detecciones sobre estela con NMS global.



Figura B. 43. Detecciones sobre estela utilizando NMS con filtrado de prioridad



Figura B. 44. Clasificación sobre detecciones.

B12. Magical Stela with Horus the Child (Reverso)

Objeto: Estela mágica con Horus Niño (Reverso).

Referencia: Magical Stela with Horus the Child (Cat. 546037). <https://www.met-museum.org/es/art/collection/search/546037>

Características: Dinastía XXX (ca. 360–343 a. C.). Metagrauvaca (a. total 83,5 cm).

Imagen: DP-38634-004.jpg (3000 × 4000 px).

Tabla B 23. Resultados de detección global y optimización de escalas.

Modelo	NMS Global	Escala 3.0x (Prioridad 1)	Escala 1.0x (Prioridad 2)	NMS Escala
(RF) «Aves»	231	208	44	211
(YO) «Aves»	53	38	212	49
(YO) «Top4»	1184	971	971	1024
(RF) «999»	2735	2580	1073	2601
Total	4203	3797	2300	3885

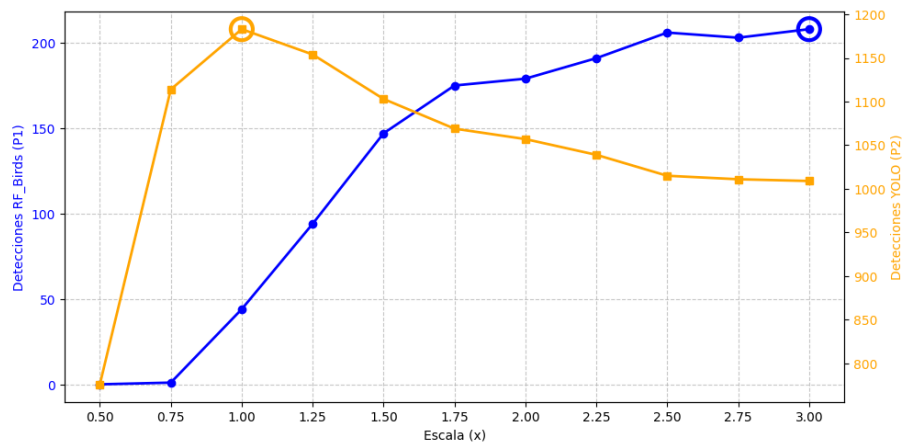


Figura B. 45. Detecciones por escala RF-DETR (azul) y YOLO (naranja).

Tabla B 24. Resultados de clasificación experta (ConvNeXt)

Modelo	Etiquetado superando umbral de confianza	Etiquetado dudoso
(CN) «Aves»	184	66
(CN) «Top4»	981	32
(CN) «999»	1142	1470



Figura B. 46. Detecciones sobre estela con NMS global.



Figura B. 47. Detecciones sobre estela utilizando NMS con filtrado de prioridad

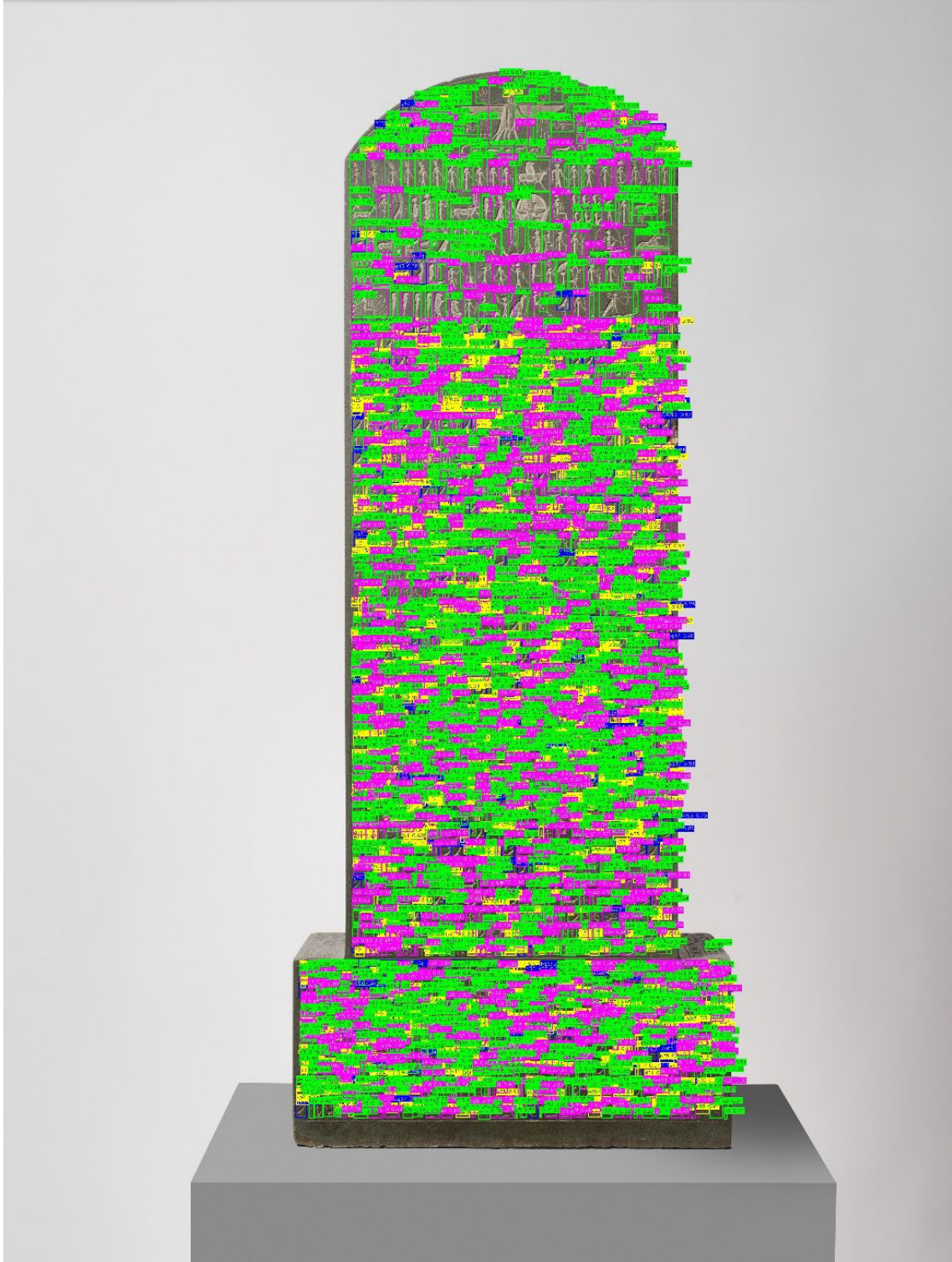


Figura B. 48. Clasificación sobre detecciones.

B13. Sheet from the Papyrus of Amenhotep

Objeto: Hoja del papiro de Amenhotep.

Referencia: Sheet from the Papyrus of Amenhotep (Cat. 554677). <https://www.met-museum.org/art/collection/search/554677>

Características: Dinastía XVIII (ca. 1427–1390 a. C.). Papiro (33,8 x 35 cm).

Imagen: DP292923.jpg (3899 × 3700 px).

Tabla B 25. Resultados de detección global y optimización de escalas.

Modelo	NMS Global	Escala 1.0x (Prioridad 1)	Escala 0.5x (Prioridad 2)	NMS Escala
(RF) «Aves»	87	39	28	46
(YO) «Aves»	51	26	60	48
(YO) «Top4»	157	74	71	93
(RF) «999»	796	643	470	608
Total	1091	782	629	895

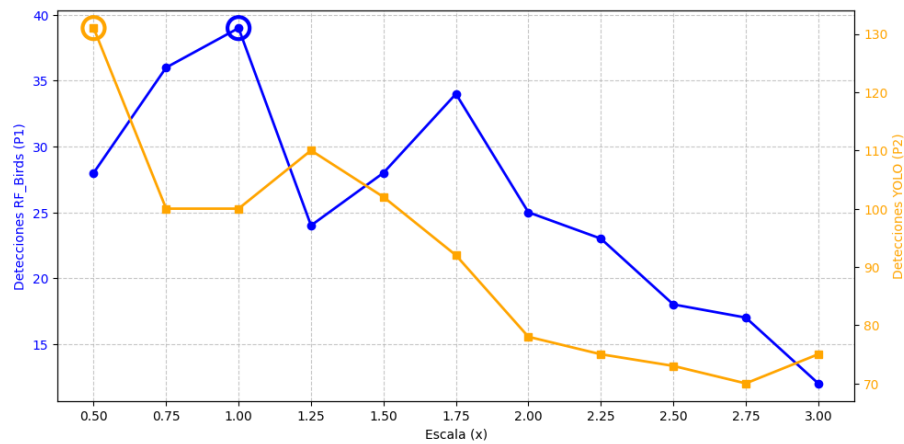


Figura B. 49. Detecciones por escala RF-DETR (azul) y YOLO (naranja).

Tabla B 26. Resultados de clasificación experta (ConvNeXt)

Modelo	Etiquetado superando umbral de confianza	Etiquetado dudoso
(CN) «Aves»	65	29
(CN) «Top4»	72	5
(CN) «999»	223	401



Figura B. 50. Detecciones sobre estela con NMS global.



Figura B. 51. Detecciones sobre estela utilizando NMS con filtrado de prioridad

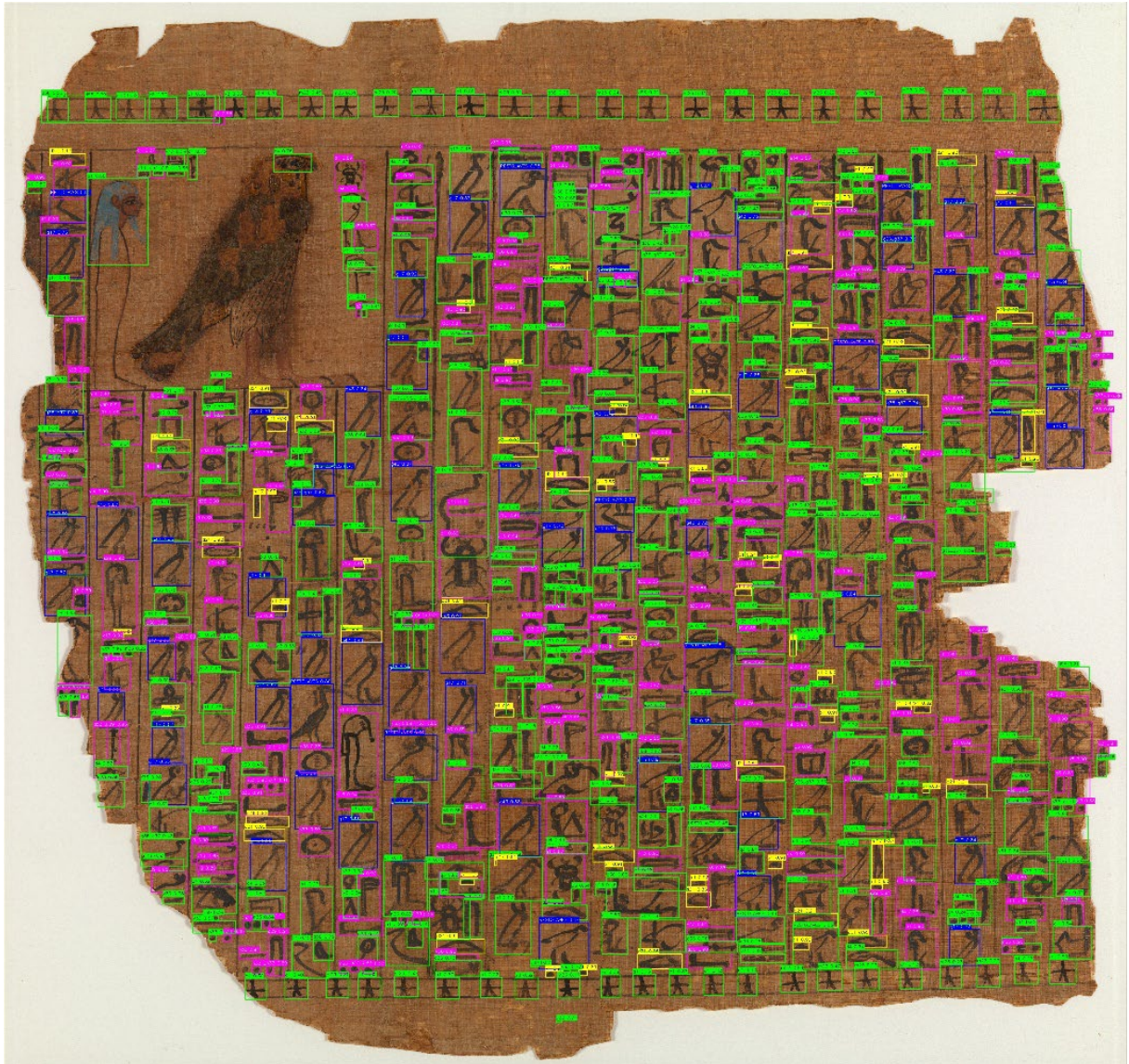


Figura B. 52. Clasificación sobre detecciones.

B14. Heqanahkt Letter II

Objeto: Carta II de Hekanajt.

Referencia: Heqanahkt Letter II (Cat. 545446). <https://www.metmuseum.org/art/collection/search/545446>

Características: Dinastía XII (ca. 1961–1917 a. C.). Papiro (27,9 x 40,8 cm). Hierático.

Imagen: DP-30871-001.jpg (4000 × 2997 px).

Tabla B 27. Resultados de detección global y optimización de escalas.

Modelo	NMS Global	Escala 2.25x (Prioridad 1)	Escala 0.5x (Prioridad 2)	NMS Escala
(RF) «Aves»	87	11	2	12
(YO) «Aves»	51	12	56	61
(YO) «Top4»	157	24	54	70
(RF) «999»	796	1319	625	1270
Total	1091	1366	737	1413

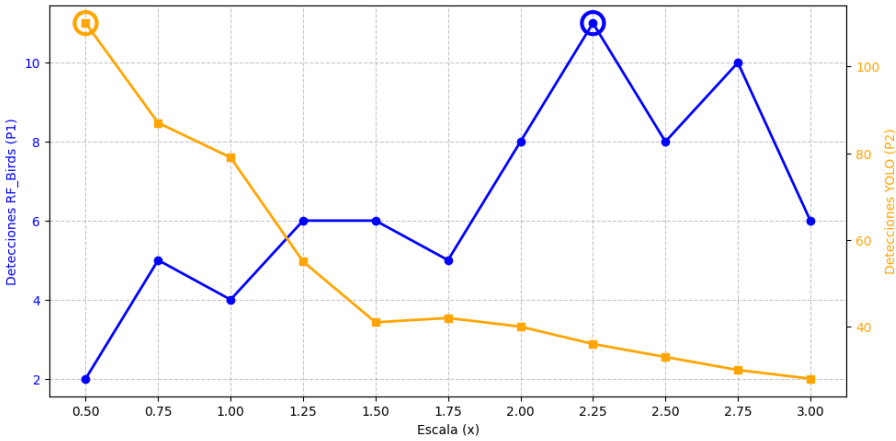


Figura B. 53. Detecciones por escala RF-DETR (azul) y YOLO (naranja).

Tabla B 28. Resultados de clasificación experta (ConvNeXt)

Modelo	Etiquetado superando umbral de confianza	Etiquetado dudoso
(CN) «Aves»	17	56
(CN) «Top4»	44	12
(CN) «999»	298	986



Figura B. 54. Detecciones sobre estela con NMS global.



Figura B. 55. Detecciones sobre estela utilizando NMS con filtrado de prioridad

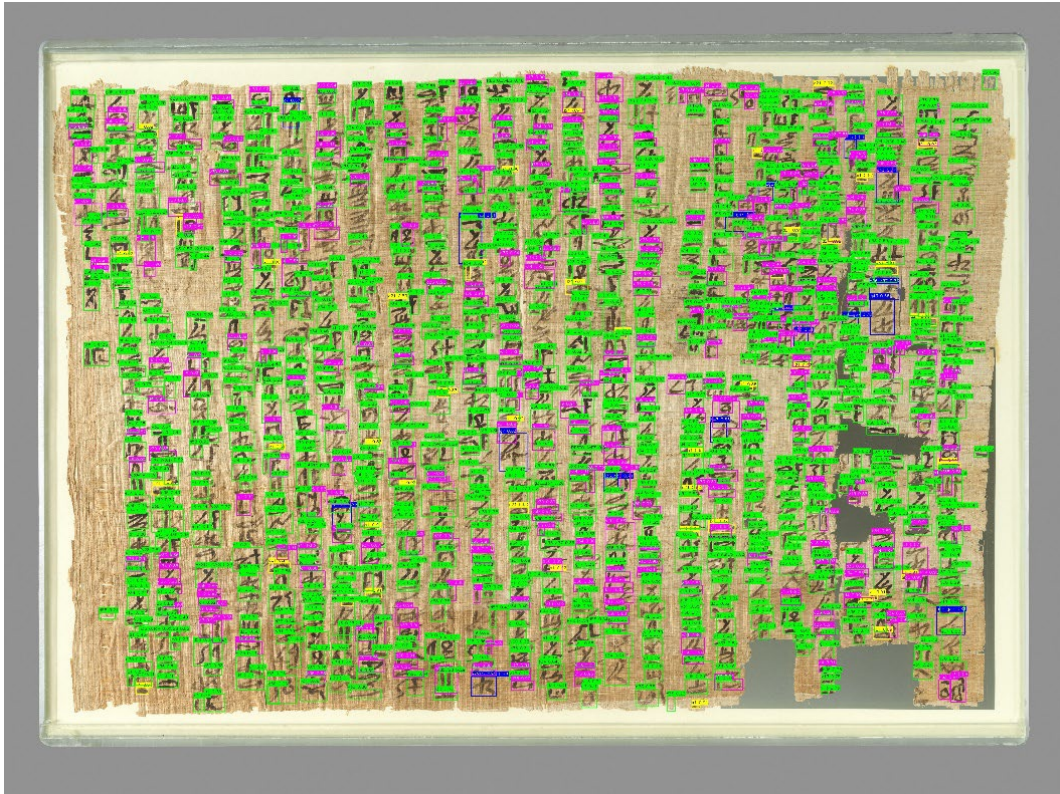


Figura B. 56. Clasificación sobre detecciones.