

Jornadas de Automática

Detección jerárquica de jeroglíficos egipcios frecuentes mediante YOLO11 y RF-DETR

Alfonso Moreno-Martínez ^{a,*}, Pedro Javier Herrera ^{a,*}

^a Dpto. de Ingeniería de Software y Sistemas Informáticos, Universidad Nacional de Educación a Distancia, C/ Juan del Rosal, 16, 28040, Madrid, España.

To cite this article: Moreno-Martínez, A., Herrera, P.J. 2026. Hierarchical detection of frequent Egyptian hieroglyphs using YOLO11 and RF-DETR. *Jornadas de Automática*, 47.
<https://doi.org/10.17979/ja-cea.2026.47.13668>

Resumen

La interpretación automática de jeroglíficos en estelas de alta densidad es un problema complejo debido a múltiples factores como la variabilidad lingüística, la escasez de datos y el desgaste natural de los soportes. Con el fin de superar las limitaciones de los métodos tradicionales y reducir al máximo el costoso proceso de etiquetado manual de cada símbolo, se ha diseñado una estrategia de detección jerárquica. Esta consiste en analizar y localizar primero los signos más frecuentes que proporcionan la mayor cobertura en la estela, como las aves y glifos estructurales individuales agrupados, utilizándolos posteriormente como puntos de anclaje para un modelo capaz de aprovechar el contexto espacial para ubicar los jeroglíficos desconocidos bajo una etiqueta genérica. Así, combinando dos arquitecturas de vanguardia como YOLO11 y RF-DETR, se propone una estrategia unificada para aislar y etiquetar los signos, posibilitando su reconocimiento y traducción mediante redes de clasificación expertas.

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

Hierarchical detection of frequent Egyptian hieroglyphs using YOLO11 and RF-DETR

Abstract

The automatic interpretation of hieroglyphs on high-density stelae is a complex problem due to multiple factors such as linguistic variability, data scarcity, and the natural wear of the surfaces. To overcome the limitations of traditional methods and minimize the costly manual labeling of each symbol, a hierarchical detection strategy has been designed. This strategy involves first analyzing and locating the most frequent signs that provide the greatest coverage on the stela, such as birds and clusters of individual structural glyphs. These are then used as anchor points for a model capable of leveraging spatial context to place unknown hieroglyphs under a generic label. Thus, by combining two cutting-edge architectures, YOLO11 and RF-DETR, a unified strategy is proposed for isolating and labelling signs, enabling their recognition and translation by expert classification networks.

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

1. Introducción

La detección de glifos en estelas egipcias constituye uno de los problemas más exigentes a nivel computacional: signos minúsculos, muy agrupados y sobre un soporte naturalmente degradado por el paso del tiempo. Al principio, los trabajos pioneros de Franken y van Gemert (2013) demostraron la viabilidad de la automatización mediante descriptores visuales.

Pero el verdadero salto de calidad llegó con la evolución hacia un cambio de paradigma marcado por el aprendizaje profundo. Autores como Barucci *et al.* (2021) confirmaron que las redes neuronales convolucionales (CNN) superaban con creces a la visión tradicional a la hora de clasificar los signos.

Los jeroglíficos suelen aparecer en proporciones variables y no siempre tienen los bordes definidos, por lo que la tendencia actual es el uso de detectores libres de anclas. Tian

et al. (2026) confirman la evolución de YOLO hacia modelos sin anclajes y atención de *Transformers*, ahorrando cálculos redundantes y mejorando el rendimiento notablemente en escenas difíciles. Otro gran problema es la superposición de los signos. Para solucionarlo, Shehzadi et al. (2026) recomiendan usar redes de la familia DETR de extremo a extremo. La gran ventaja de este enfoque es que permite suprimir filtros manuales como la Supresión de No Máximos (NMS), lo cual facilita el aprendizaje de invarianzas y reduce la pérdida por superposición en objetos agrupados. Autores como Zhang et al. (2025), señalan que combinar detectores sin anclas con *Transformers* es ideal para entrenar modelos con muy pocos ejemplos (*Few-Shot*). En epigrafía esto es vital, ya que permite enseñar al sistema a detectar jeroglíficos raros sin necesidad de tener miles de imágenes etiquetadas.

Para abordar la alta degradación de los soportes, se han perfeccionado técnicas de segmentación de instancias mediante marcos como Mask R-CNN (Guidi et al., 2023) o se han combinado métodos de visión artificial clásicos con técnicas de segmentación híbridas (Fuentes-Ferrer et al., 2025). Recientemente se han implementado arquitecturas optimizadas basadas en CNN como ResNet50 para sistemas de traducción automática (Nasser et al., 2025), o enfoques de alta precisión (Fuentes-Jiménez et al., 2025) e incluso sistemas completos de traducción (Sobhy et al., 2023).

La epigrafía computacional ha evolucionado así desde los métodos basados en descriptores visuales clásicos hasta sofisticados sistemas de aprendizaje profundo. Sin embargo, este proceso siempre ha implicado localizar y aislar el símbolo, clasificarlo y traducirlo. Nuestro trabajo plantea un cambio de enfoque: en lugar de partir de esa base ideal, se aborda el problema justo en la fase crítica y anterior del proceso, localizando directamente el glifo. Aunque el alto coste del etiquetado manual de las cajas delimitadoras (*bounding boxes*) ha llevado a descartar detectores como YOLO en epigrafía, este trabajo evalúa el uso de YOLO11 (Khanam y Hussain, 2024) y RF-DETR (Robinson et al., 2026) mediante una estrategia diseñada para minimizar dicha carga de trabajo.

La familia de arquitecturas RT-DETR y, en concreto, la versión RF-DETR empleada en este trabajo, destaca por su capacidad para comprender el contexto global de una imagen. Sin embargo, presenta dificultades al detectar objetos diminutos frente a la solvencia contrastada de YOLO (Zhao et al., 2024; Asensi-González et al., 2026). Por tanto, el objetivo final es verificar si la comprensión del contexto visual global de RF-DETR compensa su menor precisión en símbolos pequeños al enfrentarse a inscripciones muy deterioradas, o si la extracción de características locales de YOLO11 sigue siendo el enfoque predominante.

Para mitigar la baja precisión característica de la detección de objetos pequeños (SOD), en este trabajo se ha implementado una solución práctica basada en el uso de anclajes lingüísticos. Primero se le enseña a la red a buscar los signos más comunes y reconocibles. Una vez que el modelo ubica estos elementos clave, comprende mucho mejor el contexto de la estela y le resulta más fácil deducir dónde están los jeroglíficos desconocidos o casi borrados.

Asegurar la detección correcta de glifos es el paso previo y fundamental para que la fase de clasificación se realice de manera adecuada. La principal ventaja del enfoque propuesto

radica en operar directamente sobre la imagen en bruto (*raw image to bounding box*), lo que independiza al modelo de los métodos clásicos de segmentación. De este modo, factores como las fisuras en la piedra no comprometen la detección, evitando la propagación de errores típica de técnicas de umbralización como Otsu o de la detección de bordes de Canny. La arquitectura RF-DETR aprovecha el contexto global mediante mecanismos de atención para predecir conjuntos de forma nativa. Esta base técnica permite implementar un proceso jerárquico capaz de aislar glifos desconocidos bajo una etiqueta genérica, optimizando su posterior revisión manual.

El resto del documento se organiza como sigue: la sección 2 detalla la metodología propuesta, la comparativa entre las arquitecturas YOLO11 y RF-DETR, y el diseño de la estrategia combinada; la sección 3 presenta el análisis de los resultados y la evaluación pormenorizada de los modelos, validando el funcionamiento del proceso de detección; finalmente, la sección 4 resume las conclusiones principales y plantea las futuras líneas de investigación derivadas de este estudio.

2. Solución propuesta

Para sortear el cuello de botella que supone el etiquetado del *dataset* epigráfico se propone una solución incremental. El proceso comienza identificando los signos más frecuentes que maximizan la cobertura en las estelas. Una vez determinados, se etiqueta el *dataset* agrupando los glifos por categorías. Estos elementos se integran en un *dataset* genérico donde los signos conocidos actúan como anclajes, proporcionando a la red el contexto espacial y lingüístico necesario para mejorar la detección. De esta forma, aplicando esta estrategia de detección jerárquica, se logra optimizar la reutilización de los modelos. Para validar esta metodología, se contrastó el enfoque de predicción densa YOLO11 frente al de predicción de conjuntos con atención global RF-DETR.

2.1. Análisis de datos

La frecuencia de aparición de los signos no es uniforme, varía dependiendo de la época y si tratamos con un texto religioso, funerario o administrativo. Este trabajo ha centrado el análisis en el corpus AED-TEI (Schweitzer, 2019), el cual contiene más de 48.000 apariciones de jeroglíficos. Tomando como referencia la lista clásica de signos propuesta por Gardiner (1957), que divide los jeroglíficos en 26 categorías, los datos son muy reveladores. Las 10 categorías más frecuentes acaparan más del 81% de los textos. Es decir, un número reducido de tipos de signos sostiene casi toda la estructura de la escritura egipcia. Al analizar los símbolos de forma individual, el contraste es aún mayor. De los 1.624 signos únicos registrados en el corpus, apenas 10 jeroglíficos conforman casi el 32% de todo el texto escrito (ver Tabla 1).

A partir del análisis de estos datos, el objetivo es entrenar un sistema que, con el mínimo número de signos, sea capaz de cubrir la mayor parte del texto, respetando la complejidad gramatical. El reto es, por tanto, optimizar los recursos y el etiquetado manual para alcanzar una cobertura mínima que permita, en una fase posterior, realizar una clasificación experta detallada.

Tabla 1: Frecuencia de jeroglíficos individuales

Id.	Img	Total	Cobert.	Descripción
X1		3.885	8,07%	Pan (t)
D21		2.072	4,30%	Boca (r)
N35		1.684	3,50%	Onda de agua (n)
Z1		1.639	3,41%	Trazo (ideograma)
M17		1.594	3,31%	Junco (i)
G7		984	2,04%	Halcón (det. divino)
Z2		950	1,97%	Trazos (plural)
S29		916	1,90%	Tela doblada (s)
Z7		865	1,80%	Espiral (w)
D36		746	1,55%	Brazo (a)

2.2. Modelos expertos

El primer paso de la estrategia jerárquica consistió en entrenar las arquitecturas YOLO y RF-DETR para la detección de glifos distintivos y frecuentes en estelas. Se seleccionó la categoría «Aves» (código G de Gardiner) como caso de estudio inicial debido a que representa por sí sola el 8,24% del corpus (incluyendo signos como el G7 o el G17). Una vez validada esta metodología, el sistema se extendió a los glifos de la clase «Top4» que agrupa los códigos X1, D21, N35 y M17, un conjunto de glifos estructurales que resultan prioritarios ya que, en conjunto, representan el 19,18% de la cobertura total. Se excluyó deliberadamente al glifo Z1 porque su morfología extremadamente simple requiere un tratamiento especializado diferente al del resto de jeroglíficos.

2.3. Modelo genérico contextual

Combinar estos dos subsistemas expertos (especializados en las clases «Aves» y «Top4») implica que el sistema es capaz de localizar casi un tercio de los glifos presentes en cualquier estela. Para lograrlo, se generó un nuevo *dataset* utilizando los fragmentos etiquetados por los modelos expertos. Los jeroglíficos desconocidos se agruparon y etiquetaron bajo la clase genérica «999». De esta forma, se consiguió entrenar un modelo unificado que utiliza las detecciones previas («Aves» y «Top4») para proporcionar el contexto necesario a la red, permitiéndole detectar jeroglíficos desconocidos en inscripciones que nunca ha evaluado.

2.4 Proceso de localización

Mediante la implementación de un sistema de ventanas deslizantes, el proceso jerárquico propuesto se materializó tal y como se detalla en la Figura 1. En primer lugar, con los modelos RF-DETR se identificaron 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 1), otorgándoles la prioridad 1, y los genéricos un umbral de 0,35, asignados con la prioridad 3. Estas detecciones se complementaron con las inferencias de los modelos YOLO11 configurados para capturar elementos de la clase «Aves» en umbrales medios (25%–40%, en color magenta) y glifos de la clase «Top4» en umbrales bajos ($\leq 25\%$, en color azul), con prioridad 2. 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 1.

3. Análisis y evaluación de resultados

3.1. Métricas

Para evaluar los modelos YOLO11 y RF-DETR se extrajeron las coordenadas absolutas y los niveles de confianza de las predicciones sobre el conjunto de prueba o test y, mediante un script propio, se midió 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 calcularon la Precisión, el *Recall*, el F1-Score y el mAP@0.5.

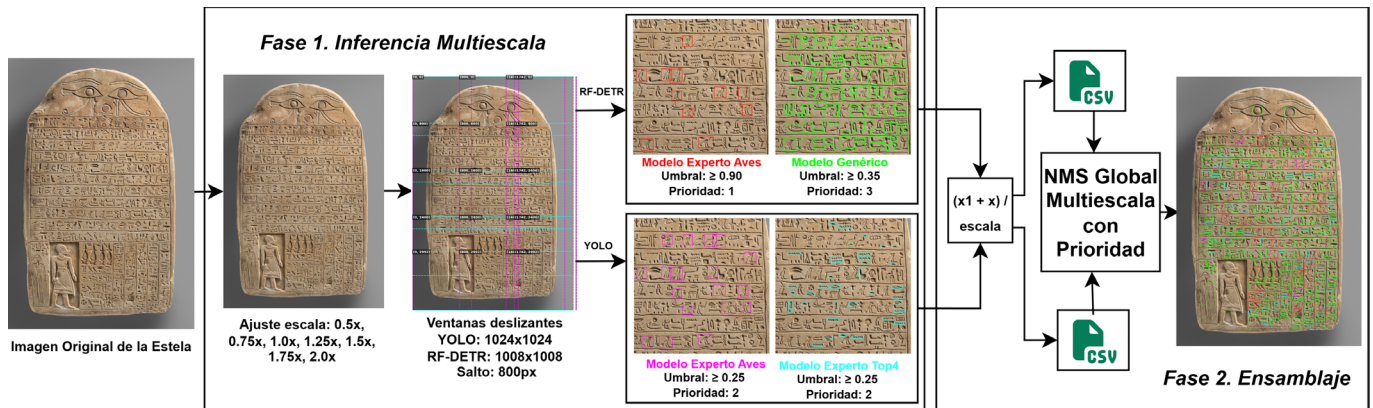


Figura 1: Proceso híbrido de localización. Estela de Horemkhauf.

3.2. 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. El límite físico de esta memoria resultó ser una restricción técnica mucho más determinante que el volumen teórico de GFLOPs. Con el objetivo de preservar los detalles epigráficos, se configuraron YOLO11 y RF-DETR (S) a la máxima resolución que permitía el equipo (1024×1024 y 1008×1008 píxeles, respectivamente). De esta forma, la comparativa evalúa el rendimiento de ambas arquitecturas al llevarlas al límite bajo las mismas condiciones de hardware.

Durante el entrenamiento, se permitió a cada *framework* utilizar sus heurísticas y valores de optimización por defecto (como el optimizador AdamW y la precisión mixta en YOLO11, o el decaimiento por pasos en RF-DETR), modificando exclusivamente los hiperparámetros necesarios para adaptar los modelos a las restricciones de hardware y asegurar las mismas condiciones base. La limitación de memoria de vídeo (VRAM) del entorno local condicionó el tamaño de lote (*batch size*). En YOLO11 se estableció un máximo de 8 imágenes por iteración. Por su parte, RF-DETR (S) emplea un sistema que combina el lote físico con la acumulación de gradientes, por lo que, para no otorgarle la ventaja de un lote efectivo mayor, se configuró con un tamaño de lote físico de 4 y 2 pasos de acumulación de gradientes, igualando así el lote efectivo de 8 (Tabla 2).

Tabla 2: Comparativa de hiperparámetros YOLO11 vs RF-DETR (S)

Hiperparámetros	YOLO11	RF-DETR (S)
imgs / resolution	1024 / 512	1008 / 508
max. epochs	200	200
patience	30	30
batch_size	8	4
grad_accum_steps	–	2
batch efectivo	8	8
scheduler	cos_lr=True	Step Decay
lr	Automático	$1e^{-4}$

3.3. Evaluación de los 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 «Top4» agrupa cuatro signos morfológicamente heterogéneos (X1, D21, N35 y M17) en 233 fragmentos de 98 estelas. Los datos se dividieron en tres grupos disjuntos a nivel de estela completa, en una proporción de 70, 20 y 10, correspondientes a los conjuntos de entrenamiento, validación y prueba, respectivamente. Por último, se generaron formatos de anotación adaptados a cada *framework*: texto plano con coordenadas normalizadas para YOLO y COCO en JSON para RF-DETR.

La Figura 2 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.

Los resultados del F1-Score detallados en las Tablas 3 y 4 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.

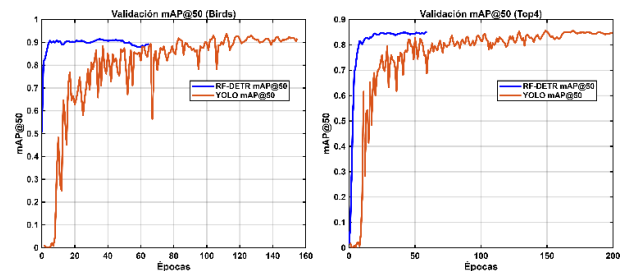


Figura 2: Comparativa Validación mAP@50 (clases «Aves» y «Top4»).

Tabla 3: Comparativa métricas 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 4: Comparativa métricas YOLO vs RF-DETR (Clase «Top4»)

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

Estos resultados se ven respaldados por el análisis de las matrices de confusión. YOLO11 detecta mejor los glifos reales, alcanzando una tasa de acierto del 85% en la categoría «Aves» y del 83% en el conjunto «Top4», situándose por encima del 77% y 71%, respectivamente, de RF-DETR. YOLO11 reduce los errores de clasificación al confundir menos el fondo con signos reales; las falsas detecciones caen a un 14% en «Aves» y a un 7% en el «Top4». A la vista de estos resultados, se constató que la arquitectura de YOLO generaliza mejor en escenarios de alta complejidad visual.

Al revisar cualitativamente los resultados obtenidos en el conjunto de prueba, estas métricas se confirmaron, pero surgen matices importantes para la práctica epigráfica. En estelas muy densas, YOLO11 prioriza la cobertura y logra un mayor volumen de detecciones totales. Sin embargo, en inscripciones con relieves difíciles o sometidas a un desgaste

severo, RF-DETR logra rescatar más signos manteniendo un alto nivel de confianza.

3.4. Evaluación del modelo genérico

Una vez detectados los glifos de alta frecuencia, el objetivo fue la detección del resto del corpus. El *dataset* partió de 53 fragmentos de alta resolución (1024×1024 píxeles) procedentes de 14 estelas, donde todos los glifos se agruparon inicialmente bajo la clase genérica «999». Sobre este conjunto se aplicó un proceso de reetiquetado (*relabeling*) con las posiciones exactas de los signos pertenecientes a las clases «Aves» y «Top4», realizando recortes de 512×512 píxeles centrados en cada glifo. Para duplicar el volumen de la muestra sin distorsionar la información semántica, se aplicó un volteo horizontal (*flip*), alcanzando un total de 1168 muestras de entrenamiento. La clase genérica «999» englobó signos con una alta variabilidad morfológica. Ante esta diversidad visual, el rendimiento de YOLO decae de forma acusada no logrando identificar patrones comunes ni definir una textura clara. Como muestra la Figura 3, este problema genera una curva de aprendizaje muy inestable y una menor capacidad para minimizar el error de localización espacial (*Box Loss*), terminando el entrenamiento de forma prematura.

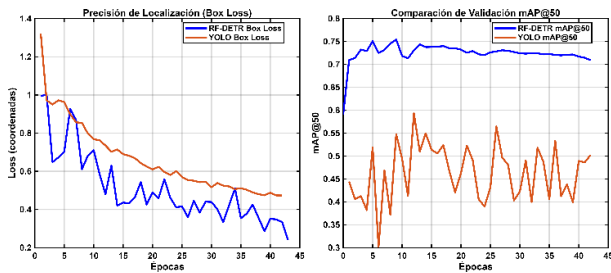


Figura 3: Comparativa entrenamiento YOLO11 vs RF-DETR (S) (clase «999»).

RF-DETR ofrece, en cambio, resultados muy superiores alcanzando un mAP@0.5 de 0,8596 frente al 0,6301 obtenido por YOLO. Este salto de rendimiento indica que la arquitectura *Transformer* no requiere memorizar la forma exacta de los cientos de símbolos distintos que componen esta clase. En lugar de eso, deduce su ubicación aprovechando la estructura gramatical y espacial que proporcionan los anclajes lingüísticos previos, relativos a las clases «Aves» y «Top4». En la misma línea, los resultados de la Tabla 5 muestran una inversión en el rendimiento observado en las pruebas detalladas en la sección 3.3 y donde RF-DETR, en umbrales de máxima confianza, es capaz de recuperar cerca del 25% de los glifos.

Tabla 5: Comparativa métricas 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

El análisis del F1-Score confirma el punto de funcionamiento óptimo de este modelo. Al establecer un umbral de 0,5, RF-DETR (S) alcanza un índice de 0,8517 frente al 0,6606 de YOLO11, demostrando que la arquitectura basada en transformadores recupera un alto volumen de glifos sin penalizar de forma severa la precisión del sistema.

Tal y como refleja la matriz de confusión (Figura 4), el modelo RF-DETR (S) 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». Lejos de suponer una limitación crítica, este comportamiento es coherente con el propósito del modelo. El objetivo no es actuar como un clasificador experto, sino como un extractor espacial orientado a dar contexto. Las clases específicas «Aves» y «Top4» se incluyeron en su entrenamiento para mejorar la comprensión global de la escena, pero su verdadero valor reside en su capacidad de recuperación masiva, logrando retener el 80% de los glifos bajo la etiqueta genérica «999».

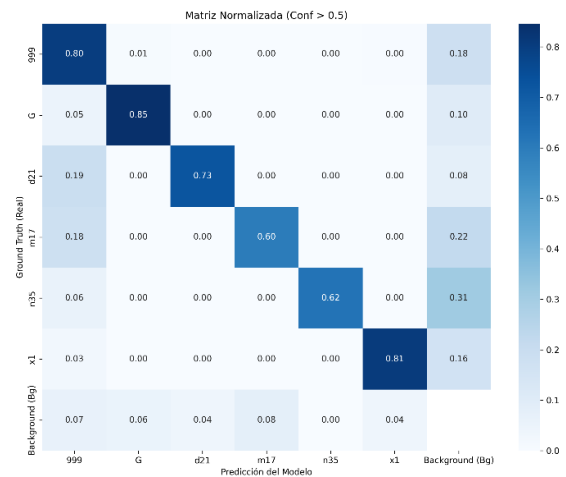


Figura 4: Matriz de confusión RF-DETR (clase «999»).

3.5. Resultados del proceso de detección

A la 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 (Figura 1). Así, el modelo híbrido YOLO11/RF-DETR, ajustado bajo sus configuraciones óptimas, mejoró la fiabilidad en la detección y, al generar una preclasificación en distintos niveles de precisión, facilitó la aplicación posterior de modelos expertos sobre los recortes obtenidos.

La Figura 5(a) muestra que, incluso en estelas con alta densidad de texto y glifos minúsculos, el sistema identifica la mayoría de los signos desconocidos. Ahora bien, el proceso multiescala puede generar ruido al detectar elementos decorativos. Para mitigarlo, se fusionan las imágenes reescaladas con más detecciones de prioridad 1 y 2, optimizando el NMS y descartando resoluciones redundantes.

En la Figura 5(b) se puede apreciar la capacidad de generalización del modelo. A pesar de haber sido entrenado exclusivamente con estelas en piedra de los Reinos Antiguo y Medio, el sistema logra identificar signos en soportes de

papiro. Este resultado es significativo, ya que en dichos materiales la escritura solía ser más pragmática y abstracta.

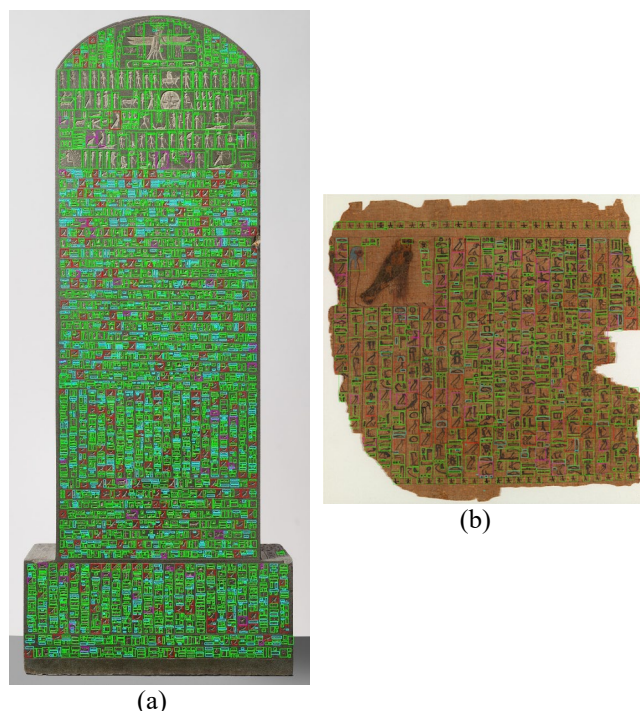


Figura 5: (a) Resultados del enfoque combinado propuesto sobre piedra (Estela Mágica, periodo tardío, 360–343 a. C., Resolución original: 3000×4000 píxeles); (b) Resultados del enfoque combinado propuesto sobre papiro (Hoja del papiro de Amenhotep, Imperio Nuevo, ca. 1427–1390 a. C., Resolución original: 3899×3700 píxeles).

4. Conclusiones

El enfoque propuesto en este trabajo, que combina los modelos YOLO11 y RF-DETR, permite aislar los jeroglíficos y generar recortes fiables que faciliten el posterior proceso de clasificación. En pruebas sobre diversos soportes, como piedra y papiro, se obtuvieron resultados cualitativos sólidos sobre un conjunto representativo, considerando la complejidad de las estelas y los niveles habituales de degradación. En concreto, YOLO11 ofreció una mayor estabilidad al identificar glifos frecuentes, mientras que RF-DETR resultó más eficaz ante signos desconocidos o en zonas deterioradas. Si bien el NMS multiescala maximiza la recuperación de signos, también provoca un aumento en la cantidad de decoración identificada como glifo, aunque la propuesta de máximos minimiza este problema. Se propone un filtrado de este ruido ornamental en una fase de clasificación posterior, basándose en las dimensiones y características de cada elemento.

La evaluación cuantitativa *end-to-end* del sistema completo se plantea como trabajo futuro, condicionada actualmente por la carencia de conjuntos de datos públicos anotados a nivel de estela completa en alta resolución. También se propone desarrollar nuevos *datasets* que optimicen el proceso de manera continua, integrándolos en un modelo capaz de identificar y reconocer jeroglíficos en textos egipcios. El fin último es proporcionar una herramienta de apoyo a los egiptólogos en la transliteración y traducción de inscripciones, y su extensión a caracteres en

otras lenguas de corpus limitado y diferentes soportes epigráficos.

Agradecimientos

Este trabajo ha sido realizado gracias al apoyo del proyecto GO-HYBRID (PID2024-156427OB-I00), financiado por MICIU/AEI/10.13039/501100011033/ FEDER, UE.

Referencias

- Asensi-González, R., Herrera, P.J., Gómez, M., 2026. Identificación facial automática en secuencias de vídeo empleando redes neuronales y aprendizaje profundo. *Revista Iberoamericana de Automática e Informática Industrial*. DOI: 10.4995/riai.2026.25454
- 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. DOI: 10.1109/ACCESS.2021.3110082
- Franken, M., Van Gemert, J.C., 2013. Automatic Egyptian hieroglyph recognition by retrieving images as texts, in: *Proceedings of the 21st ACM International Conference on Multimedia*. Presented at the MM '13: ACM Multimedia Conference, ACM, Barcelona Spain, pp. 765–768. DOI: 10.1145/2502081.2502199
- Fuentes-Ferrer, R., Duque-Domingo, J., Herrera, P.J., 2025. Recognition of Egyptian hieroglyphic texts through focused generic segmentation and cross-validation voting. *Applied Soft Computing* 171, 112793. DOI: 10.1016/j.asoc.2025.112793
- Fuentes-Jiménez, 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. DOI: 10.48550/arXiv.2512.24197
- Gardiner, A., 1957. *Egyptian grammar: Being an introduction to the study of hieroglyphics*. Griffith Institute, Ashmolean Museum.
- Guidi, T., Python, L., Forasassi, M., Cucci, C., Franci, M., Argenti, F., Barucci, A., 2023. Egyptian Hieroglyphs Segmentation with Convolutional Neural Networks. *Algorithms* 16, 79. DOI: 10.3390/a16020079
- Khanam, R., Hussain, M., 2024. YOLOv11: An Overview of the Key Architectural Enhancements. DOI: 10.48550/arXiv.2410.17725
- 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. DOI: 10.48550/arXiv.2512.03817
- Robinson, I., Robicheaux, P., Popov, M., Ramanan, D., Peri, N., 2026. RF-DETR: Neural Architecture Search for Real-Time Detection Transformers. DOI: 10.48550/arXiv.2511.09554
- Schweitzer, S. D., 2019. *Ancient Egyptian Dictionary. Corpus of Egyptian Texts for the AED (Versión 1.1.0-dev)* [Conjunto de datos]. Zenodo. DOI: 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, 310. DOI: 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. DOI: 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. DOI: 10.1016/j.neucom.2025.132436
- Zhang, J., Liu, L., Silván, O., Pietikäinen, M., Hu, D., 2025. Few-Shot Class-Incremental Learning for Classification and Object Detection: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 47, 2924–2945. DOI: 10.1109/TPAMI.2025.3529038
- 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, in: *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Presented at the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Seattle, WA, USA, pp. 16965–16974. DOI: 10.1109/CVPR52733.2024.0160