Sobre la Robustez Inherente de los Detectores de Objetos de una Única Etapa frente a Datos fuera de Distribución

Aitor Martinez-Seras[1], Javier Del Ser[2,3], Aitzol Olivares-Rad[3], Alain Andres[2,4], and Pablo Garcia-Bringas[4]

[1] IDEKO, Basque Research and Technology Alliance (BRTA), 20870 Elgoibar, Spain
[2] TECNALIA, Basque Research & Technology Alliance (BRTA), 48160 Derio, Spain
[3] University of the Basque Country (UPV/EHU), 48940 Leioa, Spain
[4] University of Deusto, 48007 Bilbao, Spain
amartinezseras@ideko.es, javier.delser@tecnalia.com

Abstract. La detección de objetos en entornos abiertos es un desafío de investigación fundamental para el despliegue seguro y fiable de sistemas prácticos de visión artificial. En estos escenarios, los modelos de detección de objetos deben ser capaces no solo de reconocer objetos conocidos, sino también de identificar instancias pertenecientes a categorías no observadas durante el entrenamiento. La mayoría de las propuestas existentes dentro del paradigma *Open World Object Detection* (OWOD) requieren algoritmos de reentrenamiento específicos y se apoyan en arquitecturas de dos etapas, lo que limita su aplicabilidad práctica a modelos con diseños arquitecturales alternativos. En este trabajo, publicado recientemente en la revista Neural Networks [4], presentamos un estudio que demuestra la robustez inherente de los detectores de objetos de una sola etapa frente a datos fuera de distribución (*Out-of-Distribution*, OoD). En particular proponemos FMap, un método post-hoc basado en la caracterización de mapas de activación internos al modelo que permite detectar objetos desconocidos sin modificar ni reentrenar el modelo original. Los resultados obtenidos sobre el benchmark UOD muestran que la familia de detectores YOLO contiene información latente suficiente para discriminar objetos desconocidos, y que la combinación de FMap con técnicas clásicas de detección OoD mejora significativamente la robustez del sistema. Nuestros hallazgos abren vías para el desarrollo de sistemas de visión artificial más seguros, eficientes y adaptables a escenarios abiertos.

Keywords: Detección de Objetos · Robustez · Aprendizaje Automático en Mundo Abierto · Detección de Datos fuera de Muestra.

1 Introducción

La creciente adopción de sistemas de Inteligencia Artificial en entornos abiertos y dinámicos exige modelos capaces de operar de forma segura cuando se enfrentan a situaciones no observadas durante el entrenamiento. En tareas de visión artificial, y particularmente en detección de objetos, esta necesidad se traduce en la

capacidad de distinguir entre objetos conocidos y objetos desconocidos o fuera de distribución (*Out-of-Distribution*, OoD).

Desde la introducción formal del paradigma *Open World Object Detection* (OWOD) por parte de Joseph et al. mediante ORE [2], numerosas contribuciones han abordado el problema de la detección de objetos desconocidos mediante la incorporación explícita de una clase "unknown" y mecanismos de aprendizaje incremental. Entre ellas destacan OW-DETR [1], basado en arquitecturas Transformer y estrategias de pseudo-etiquetado; o los enfoques propuestos en [6, 5], que revisan los protocolos de evaluación existentes y exploran nuevas formas de generar propuestas de objetos desconocidos. Más recientemente, UnSniffer [3] ha demostrado que una adecuada modelización de la confianza asociada a los objetos puede mejorar significativamente la detección de instancias desconocidas. Pese a que la literatura sobre detección OoD ha explorado técnicas post-hoc originalmente desarrolladas para clasificación de imágenes, su extensión a modelos de detección de objetos ha recibido una atención comparativamente menor, especialmente en arquitecturas de una sola etapa, donde la detección y clasificación se realizan de forma conjunta. La mayor parte de los avances recientes en se han centrado en detectores de dos etapas que requieren procedimientos específicos de reentrenamiento para aprender a identificar objetos desconocidos. Pese a sus buenos resultados, estas estrategias incrementan significativamente el coste computacional y dificultan el aprovechamiento de modelos ya desplegados.

El trabajo original [4] resumido en este artículo plantea una pregunta de interés para la comunidad: ¿poseen los detectores de objetos de una sola etapa, tales como YOLO, una capacidad inherente para identificar objetos desconocidos sin necesidad de ser reentrenados? Para responder a esta cuestión, los autores proponen FMap, un nuevo detector OoD basado en la caracterización de los mapas de activación internos generados por el modelo. A diferencia de la mayoría de propuestas previas, FMap opera como un mecanismo completamente post-hoc, capaz de aprovechar detectores previamente entrenados sin modificar su arquitectura ni sus parámetros.

2 Propuesta Metodológica: Detector OoD FMap

La contribución principal del trabajo es FMap (*Feature Map Characterization Detector*), un mecanismo post-hoc que opera sobre modelos de detección previamente entrenados y que no requiere modificar ni reentrenar la arquitectura original. La idea fundamental consiste en explotar la información contenida en los mapas de características generados por el detector en inferencia. Para ello:

1. Se extraen las activaciones correspondientes a objetos correctamente detectados en el conjunto de entrenamiento.
2. Estas representaciones se agrupan mediante *clustering* para caracterizar las regiones del espacio de características asociadas a cada clase conocida.
3. Durante la inferencia, las activaciones asociadas a una nueva detección se comparan con los centroides aprendidos.

4. Si la distancia a dichas regiones excede un umbral previamente estimado, la detección se considera un posible objeto desconocido.

Además del algoritmo base arriba descrito, proponemos dos extensiones:

- *Reducción supervisada de dimensionalidad* (SDR): Las activaciones en los detectores de objetos suelen tener una dimensionalidad elevada, lo que dificulta la discriminación entre objetos conocidos y desconocidos. Para mitigar este problema, FMap incorpora una etapa opcional de reducción supervisada de dimensionalidad basada en en algoritmo IVIS, que proyecta las características a espacios latentes más compactos y discriminativos.
- *Enhanced Unknown Localization* (EUL): La segunda extensión busca mejorar la localización de objetos desconocidos mediante el análisis de mapas de activación de alta resolución. A diferencia de las técnicas tradicionales, que dependen de las predicciones emitidas por el detector, EUL genera propuestas adicionales de regiones potencialmente desconocidas a partir de información no utilizada explícitamente por el modelo durante la inferencia.

Finalmente, en [4] exploramos diferentes estrategias de fusión entre FMap y detectores OoD basados en logits, tales como MSP, Energy u ODIN, con el objetivo de aprovechar la complementariedad entre los detectores fusionados.

3 Principales Resultados Experimentales

La evaluación experimental se realizó sobre el benchmark UOD (*Unknown Object Detection Benchmark* [3]), utilizando YOLOv8 como detector base y comparando FMap con métodos clásicos de detección OoD y con algoritmos del estado del arte específicamente diseñados para el paradigma OWOD. Los resultados permitieron extraer varias conclusiones relevantes:

Robustez inherente de los detectores YOLO. El hallazgo más importante es que los detectores de una sola etapa poseen información interna suficiente para discriminar objetos desconocidos sin necesidad de procedimientos específicos de entrenamiento. Este resultado valida la hipótesis central del trabajo y abre una nueva línea de investigación en OWOD basada en análisis post-hoc.

Beneficios de la caracterización mediante mapas de activación: FMap con SDR obtiene los mejores resultados en detección de objetos desconocidos y reducen significativamente los errores de clasificación de objetos OoD como clases conocidas.

Complementariedad con métodos basados en logits: Aunque las técnicas clásicas de detección OoD basadas en logits mantienen una mejor precisión sobre clases conocidas, la combinación de FMap con estos métodos genera modelos más robustos que cualquiera de dichos métodos por separado. La fusión produce compromisos balanceados y Pareto-superiores entre la precisión sobre objetos conocidos y la capacidad de detección de objetos desconocidos.

Competitividad frente al estado del arte: Las configuraciones más avanzadas de FMap muestran resultados comparables e incluso superiores a métodos de referencia como ORE, RandBox o UnSniffer en diversas métricas de robustez, manteniendo además la ventaja de no requerir reentrenamiento.

Generalización a nuevas versiones de YOLO: Los experimentos realizados sobre YOLOv9, YOLOv10, YOLOv11 y YOLOv12 confirman que los beneficios observados no son exclusivos de una arquitectura concreta, sino que se mantienen de forma consistente a lo largo de distintas generaciones de detectores.

4 Conclusiones y Líneas Futuras

Los resultados obtenidos sugieren varias direcciones prometedoras para futuras investigaciones:

- En primer lugar, la caracterización de activaciones internas en la que se basa FMap abre la puerta al desarrollo de detectores OoD completamente independientes del proceso de entrenamiento. Este enfoque podría facilitar la incorporación de mecanismos de robustez en sistemas ya desplegados en entornos industriales, vehículos autónomos o sistemas de vigilancia.
- En segundo lugar, la combinación de información procedente de múltiples niveles de la red constituye una línea de trabajo especialmente atractiva. En el artículo planteamos la posibilidad de integrar activaciones de diferentes escalas y capas intermedias para construir representaciones más expresivas de la incertidumbre del modelo.

Finalmente, desde una perspectiva más amplia de IA fiable, este trabajo muestra que una parte importante de la robustez necesaria para operar en entornos abiertos puede obtenerse explotando mejor la información ya presente en los modelos, evitando costes adicionales de entrenamiento y facilitando una adopción más práctica de mecanismos de seguridad en sistemas reales.

References

1. Gupta, A., et al.: OW-DETR: Open-world detection transformer. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9235–9244 (2022)
2. Joseph, K., Khan, S., Khan, F.S., Balasubramanian, V.N.: Towards open world object detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5830–5840 (2021)
3. Liang, W., Xue, F., Liu, Y., Zhong, G., Ming, A.: Unknown Sniffer for Object Detection: Don't Turn a Blind Eye to Unknown Objects. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3230–3239 (2023)
4. Martinez-Seras, A., Del Ser, J., Olivares-Rad, A., Andres, A., Garcia-Bringas, P.: On the inherent robustness of one-stage object detection against out-of-distribution data. Neural Networks p. 108683 (2026)
5. Wang, Y., et al.: Random boxes are open-world object detectors. In: IEEE/CVF International Conference on Computer Vision. pp. 6233–6243 (2023)
6. Zhao, X., et al.: Revisiting open world object detection. IEEE Transactions on Circuits and Systems for Video Technology **34**(5), 3496–3509 (2024)