Detección Fuera de Distribución Jerárquica para Reconocimiento de Estilos Arquitectónicos en Escenarios de Mundo Abierto

Juan Luis Suárez-Díaz[1 [0000−0001−8587−4345]], Ignacio Antequera-Sánchez[1 [0009−0006−5843−3468]], Rosana Montes[2 [0000−0002−0183−044X]], and Francisco Herrera[1 [0000−0002−7283−312X]]

[1] Departamento de Ciencias de la Computación e Inteligencia Artificial (DECSAI), Instituto Andaluz Interuniversitario en Ciencia de Datos e Inteligencia Computacional (DaSCI), Universidad de Granada, Granada, España
jlsuarezdiaz@ugr.es, ignacioantequera@ugr.es, herrera@decsai.ugr.es
[2] Departamento de Lenguajes y Sistemas Informáticos (LSI), Instituto Andaluz Interuniversitario en Ciencia de Datos e Inteligencia Computacional (DaSCI), Universidad de Granada, Granada, España
rosana@ugr.es

Resumen La detección de entradas fuera de distribución es esencial para desplegar sistemas de visión por computador en escenarios reales de mundo abierto. Este trabajo aborda este problema en MonuMAI, una aplicación de ciencia ciudadana orientada al reconocimiento de estilos arquitectónicos en imágenes de monumentos. Se propone una metodología jerárquica basada en una cascada de decisiones semánticas que filtra progresivamente imágenes alejadas del dominio visual esperado, imágenes no arquitectónicas, edificios sin relevancia monumental y monumentos pertenecientes a estilos no contemplados por el sistema. La evaluación experimental sobre un dataset compuesto por imágenes de usuarios y recursos externos muestra que la metodología reduce de forma clara la aceptación errónea de entradas fuera del dominio respecto al filtrado original de MonuMAI. Además, cada rechazo queda asociado a una causa semántica interpretable, lo que facilita la curación de datos, el análisis de errores y la ampliación progresiva de la taxonomía arquitectónica mediante revisión experta y aprendizaje activo.

Keywords: Detección Fuera de Distribución · Mundo Abierto · Patrimonio Cultural

1. Introducción

La detección de datos fuera de distribución (out-of-distribution, OOD) [14] se ha convertido en un proceso crucial en inteligencia artificial, especialmente en aplicaciones de alto riesgo, donde entradas no fiables pueden inducir salidas impredecibles con consecuencias graves.

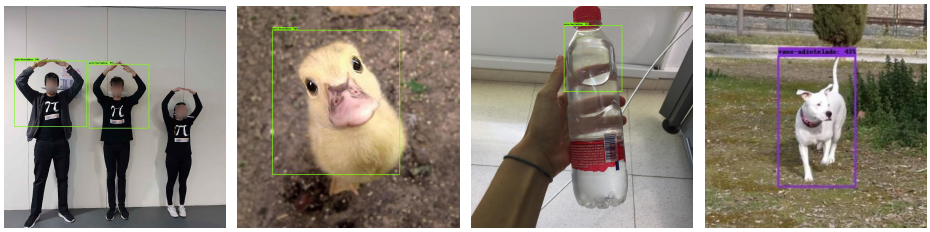


Figura 1: Ejemplos de clasificación errónea de imágenes OOD como monumentos en la aplicación de MonuMAI.

Su relevancia abarca desde datos tabulares hasta entradas heterogéneas o multimodales. En visión por computador, identificar imágenes ajenas a la distribución esperada permite prevenir fallos críticos en modelos desplegados bajo la hipótesis de mundo abierto [2]. Estas entradas pueden aparecer de forma accidental o intencionada, degradar el rendimiento del sistema y presentar naturalezas distintas, por lo que su detección y caracterización resultan especialmente relevantes. Esta necesidad aparece en ámbitos como el diagnóstico médico por imagen [16], la conducción autónoma [12] o las aplicaciones de usuario basadas en modelos de visión [9], expuestas a contenido no relacionado, ataques adversarios, clases no reconocidas o imágenes semánticamente ambiguas.

Este trabajo se centra en MonuMAI [6], una aplicación orientada al patrimonio cultural que proporciona información sobre estilos arquitectónicos y elementos relevantes de monumentos a partir de imágenes. Por su naturaleza abierta, así como por los datos curados y modelos desplegados que proporciona, MonuMAI constituye un banco de pruebas adecuado para estudiar la detección OOD en entornos reales orientados al usuario.

Durante más de cinco años de funcionamiento, MonuMAI ha recibido miles de imágenes diversas, incluyendo contenido no relacionado con monumentos, capturas de baja calidad, edificios con relevancia arquitectónica limitada, monumentos de estilos no soportados y entradas deliberadamente engañosas. La Figura 1 muestra ejemplos en los que MonuMAI no detecta correctamente entradas OOD proporcionadas por usuarios. Dado que el modelo subyacente fue entrenado con cuatro estilos arquitectónicos, hispanomusulmán, gótico, renacentista y barroco, las imágenes de estilos no observados tienden a generar clasificaciones erróneas con alta confianza. Por ello, resulta importante disponer de mecanismos capaces de detectar estilos aún no reconocidos, facilitando la ampliación progresiva del sistema y estrategias de aprendizaje activo.

En aplicaciones como MonuMAI, filtrar distintos tipos de datos OOD permite gestionarlos según sus características. Para ello, el espacio de entrada puede organizarse mediante puntos de control semánticos sucesivos, en lugar de plantear una partición plana entre datos dentro y fuera de distribución. Cada punto de control captura una distinción progresivamente más específica, siguiendo el razonamiento de un experto que descarta contenido no plausible antes de asig-

nar una interpretación concreta. Esta organización jerárquica favorece criterios especializados en cada etapa, limita la propagación de errores y proporciona una salida diagnóstica útil para monitorización, curación de datos y adquisición de nuevas muestras.

En este estudio proponemos, sobre MonuMAI, una metodología de detección OOD basada en una cascada semántica de decisiones dicotómicas. La jerarquía incorpora conocimiento experto alineado con el razonamiento humano y permite separar progresivamente distintos tipos de entradas fuera de distribución en un entorno abierto. En cada nivel, se aprovechan las capacidades representacionales de modelos de inteligencia artificial para discriminar las dicotomías definidas. Así, el enfoque combina conocimiento experto y aprendizaje automático para mejorar la fiabilidad de MonuMAI ante entradas OOD.

A partir de un conjunto de datos compuesto por imágenes subidas por usuarios de MonuMAI y complementado con imágenes de conjuntos públicos de distintos dominios, se compara la metodología propuesta con el modelo de aprendizaje profundo original de la aplicación. Los resultados muestran que esta metodología reduce significativamente las falsas detecciones de monumentos, preservando al mismo tiempo las imágenes dentro de distribución. Asimismo, se analiza el sistema capa por capa para evaluar la contribución de cada etapa y caracterizar los tipos de imágenes OOD filtrados y priorizar para revisión experta aquellas muestras rechazadas con mayor interés semántico, habilitando un ciclo de aprendizaje activo para la ampliación progresiva de la taxonomía arquitectónica.

El resto del artículo se organiza como sigue. La Sección 2 presenta los conceptos necesarios, la Sección 3 describe la metodología propuesta, la Sección 4 expone los experimentos y resultados, y la Sección 5 recoge las conclusiones.

2. Preliminares

En esta sección se introducen los conceptos relevantes para la aplicación de la metodología propuesta: la detección OOD y MonuMAI.

2.1. Detección fuera de distribución

La detección OOD es fundamental para mejorar la robustez, seguridad y confiabilidad de los sistemas de aprendizaje automático [14]. En escenarios reales, los modelos pueden recibir entradas alejadas de la distribución de entrenamiento, procedentes de dominios, condiciones de adquisición, categorías semánticas o manipulaciones adversarias no contempladas. Si no se detectan, estas muestras pueden clasificarse con alta confianza y producir errores graves en ámbitos críticos como conducción autónoma, medicina o sistemas financieros.

La detección OOD actúa como mecanismo de salvaguarda ante situaciones no previstas. Por ejemplo, en conducción autónoma, permite alertar ante entornos o condiciones diferentes a las observadas durante el entrenamiento [12]; en imagen médica, ayuda a identificar enfermedades raras o variantes no representadas en

los datos [16]; y, en procesamiento del lenguaje natural, resulta cada vez más relevante por el uso de modelos fundacionales y agentes autónomos [15]. Así, su objetivo es evitar predicciones erróneas, favorecer despliegues seguros y señalar casos inesperados para revisión experta [1].

Este problema es complejo por la heterogeneidad de los cambios distribucionales, que suelen dividirse en cambio covariante y cambio semántico. El primero modifica la distribución de entrada, pero mantiene el espacio de etiquetas; el segundo introduce clases no vistas durante el entrenamiento y exige mecanismos explícitos de rechazo, por lo que constituye el foco principal de la detección OOD.

Los métodos existentes se agrupan principalmente en enfoques basados en clasificación, densidad, distancia y reconstrucción. Estos emplean, respectivamente, salidas o representaciones internas de la red [8], estimaciones de probabilidad de los datos dentro de distribución [7], distancias a prototipos o vecinos en el espacio de características [11], y errores de reconstrucción mediante autoencoders o modelos generativos [3]. Aunque cada familia presenta limitaciones, su combinación ha permitido mejorar la robustez de los detectores OOD.

2.2. MonuMAI

El framework MonuMAI (Monument with Mathematics and Artificial Intelligence) [6] integra visión por computador, conocimiento experto en historia del arte y ciencia ciudadana para el reconocimiento automático de estilos y elementos arquitectónicos en patrimonio monumental. Su objetivo es democratizar el acceso al conocimiento cultural mediante modelos de aprendizaje profundo capaces de analizar imágenes y aproximar el razonamiento experto de los historiadores de la arquitectura.

MonuMAI aborda retos propios del análisis arquitectónico, como el solapamiento entre estilos, el deterioro de los edificios y las distorsiones o variaciones de perspectiva en imágenes reales, frente a métodos previos con menor escalabilidad y robustez [5,17]. La iniciativa, presentada en la Noche Europea de los Investigadores de 2018 en Granada, fue desarrollada por la Universidad de Granada y la Fundación Descubre, combinando investigación académica y participación ciudadana.

El núcleo del sistema es una taxonomía jerárquica definida con expertos, que relaciona elementos arquitectónicos con cuatro estilos europeos de los siglos VIII al XVIII: hispanomusulmán, gótico, renacentista y barroco. A partir de esta taxonomía se construyó el conjunto de datos MonuMAI, formado por 1514 imágenes de fachadas monumentales etiquetadas por estilo y 6650 instancias de quince elementos arquitectónicos anotadas con cajas delimitadoras. Esta doble anotación permite abordar tareas de clasificación, detección y explicabilidad, mientras que el desbalance natural entre elementos introduce un escenario experimental realista.

El pipeline combina dos modelos: MonuNet, una red convolucional ligera para clasificación eficiente de estilos [13], y MonuMAI-KED, un detector basado en Faster R-CNN [10] con ResNet-101 [4] para localizar y clasificar elementos

arquitectónicos. Estos modelos alimentan una plataforma móvil y web de ciencia ciudadana, disponible en iOS y Android, que permite subir imágenes de monumentos y recibir visualizaciones anotadas con el estilo y los elementos detectados. Además de difundir conocimiento cultural, la plataforma facilita la expansión colaborativa del conjunto de datos mediante imágenes revisadas por expertos. Hasta la fecha, ha reunido más de 3500 participantes y más de 11700 fotografías, apoyando el refinamiento continuo de los modelos y la investigación en patrimonio monumental. La información adicional del proyecto está disponible en el sitio web de MonuMAI[3].

3. Detección OOD con jerarquía semántica para MonuMAI

Como se ha comentado en secciones anteriores, MonuMAI puede recibir imágenes OOD de distinta naturaleza, ya sean imágenes no relacionadas subidas intencionalmente o por error, imágenes con ruido o imágenes que, aunque muestren edificios con valor arquitectónico, no contienen ninguno de los estilos que reconoce actualmente MonuMAI. En esta sección se presenta la metodología propuesta para abordar y procesar adecuadamente la detección OOD en MonuMAI. Esta metodología consiste en establecer una jerarquía semántica de distintas categorías o mundos OOD que da lugar a una serie de capas de filtrado a través de las cuales se van encadenando dicotomías para determinar si una imagen subida a la aplicación es OOD y, en caso afirmativo, de qué tipo es.

3.1. Descripción de la metodología

Las entradas fuera de distribución se definen como muestras que difieren significativamente de la distribución de entrenamiento, aunque en visión por computador este concepto resulta especialmente complejo por la diversidad de casos posibles: entradas maliciosas o accidentales, imágenes degradadas, clases no observadas y otros patrones anómalos capaces de comprometer el comportamiento del modelo.

Para abordar esta heterogeneidad, las entradas pueden filtrarse mediante una jerarquía semántica de capas, donde cada nivel impone una frontera de decisión específica y descarta progresivamente muestras alejadas del dominio de interés. Cada capa puede estar gestionada por un detector experto entrenado para discriminar una discrepancia semántica concreta, permitiendo una clasificación más refinada mediante la integración secuencial de decisiones.

De forma general, esta jerarquía parte del *mundo lejano*, compuesto por ejemplos *far-OOD* sin relación semántica con el problema, y continúa con el *mundo cercano*, que incluye muestras ajenas a la tarea pero con características similares al dominio objetivo. Posteriormente, se incorporan superclases semánticas cada vez más específicas hasta alcanzar el *mundo completo* del problema, que engloba

[3] https://monumai.ugr.es/

todo el dominio relevante, y finalmente el *mundo conocido*, correspondiente a los datos en distribución utilizados durante el entrenamiento.

Por ejemplo, en un modelo entrenado con ciertas razas de perros, el mundo lejano incluiría números o texturas, el mundo cercano imágenes naturales como las de ImageNet, y la jerarquía semántica podría avanzar desde animales hasta mamíferos, cánidos y, finalmente, todos los perros. El mundo conocido quedaría restringido a las razas observadas por el modelo durante el entrenamiento. La siguiente sección describe la instanciación concreta de esta metodología en MonuMAI.

3.2. Descripción e implementación de las capas de filtrado

Para el caso de MonuMAI, proponemos como mundo lejano todo aquello exterior a ImageNet, como mundo cercano el propio ImageNet, y una jerarquía de clases que empieza por edificios, pasa por monumentos y termina con los monumentos de estilos conocidos por el modelo de MonuMAI. Estos mundos definen una jerarquía progresiva de filtros OOD. La implementación de cada capa en el sistema se describe a continuación.

Capa 0: filtrado del mundo lejano La capa inicial actúa como un filtro de detección OOD destinado a descartar entradas pertenecientes al mundo lejano y permitir únicamente el paso de imágenes compatibles con el mundo cercano de MonuMAI, entendido como el dominio de fotografías en color de escenas reales. Para ello, se utiliza ImageNet como referencia en distribución y un detector AdaScale-A con una arquitectura RegNet-Y-16GF preentrenada, optimizado para escenarios far-OOD. El conjunto OOD se construye con datos como MNIST y DTD, que representan entradas alejadas del dominio visual esperado. Esta capa prioriza una detección tolerante, de modo que las muestras visualmente próximas a ImageNet puedan ser refinadas posteriormente por capas más específicas.

Capa 1: edificios frente a no edificios Tras el filtrado inicial, la capa 1 opera sobre imágenes del mundo cercano y plantea una tarea de clasificación binaria entre edificios y no edificios. La clase no edificio se obtiene a partir de categorías de ImageNet no relacionadas con arquitectura, mientras que la clase edificio combina imágenes arquitectónicas de ImageNet y escenas exteriores de `Places365` [18]. Con este conjunto equilibrado, se entrena un modelo EfficientNet_v2 reducido para identificar si una imagen pertenece al dominio arquitectónico relevante para MonuMAI.

Capa 2: monumentos frente a edificios sin valor arquitectónico La capa 2 discrimina, dentro del subconjunto de edificios, entre monumentos arquitectónicamente relevantes y edificios genéricos. Para construir el conjunto de entrenamiento, se recopilan imágenes de Wikimedia Commons mediante exploración de categorías, limpieza automática, eliminación de duplicados y filtrado

semántico basado en CLIP con múltiples prompts. Las imágenes aceptadas se organizan en una tarea binaria entre monumentos y edificios no monumentales, preservando diversidad geográfica y estilística. Sobre este corpus curado se entrena una ResNet-50 con aprendizaje por transferencia para separar monumentos de construcciones sin relevancia arquitectónica específica.

Capa 3: estilos conocidos frente a desconocidos La capa final se aplica únicamente a imágenes de monumentos y detecta si pertenecen a alguno de los estilos arquitectónicos conocidos por MonuMAI. Esta etapa se formula como detección near-OOD sobre el propio modelo MonuMAI. El modelo se ajusta mediante outlier exposure usando monumentos globales no presentes en el dominio en distribución, combinando una función de pérdida de clasificación para estilos conocidos y una pérdida OOD que favorece distribuciones uniformes ante estilos desconocidos. En inferencia, la detección se basa en la entropía: predicciones de baja entropía indican estilos conocidos, mientras que alta entropía sugiere estilos fuera del dominio de MonuMAI. Las imágenes rechazadas en esta última capa no se descartan como errores sin utilidad, sino que se almacenan como candidatos prioritarios para revisión experta. De este modo, el sistema puede identificar monumentos plausibles pertenecientes a estilos no cubiertos por la taxonomía actual, proporcionando un mecanismo natural de aprendizaje activo para ampliar progresivamente el conjunto de estilos reconocidos.

La Figura 2 representa el esquema de funcionamiento del sistema de detección OOD jerárquico propuesto sobre MonuMAI.

4. Análisis experimental

Esta sección evalúa la metodología jerárquica propuesta para la detección OOD en MonuMAI. El objetivo principal es comprobar si una cascada jerárquica de filtros semánticos permite rechazar entradas fuera de distribución sin degradar excesivamente la aceptación de monumentos pertenecientes a estilos conocidos.

4.1. Configuración experimental

Para la evaluación se construyó el conjunto *MonuTest-OOD*, compuesto por 6354 imágenes recopiladas y procesadas manualmente a partir de imágenes subidas al servidor de MonuMAI y de recursos adicionales procedentes de Wikimedia. El conjunto incluye tanto imágenes pertenecientes al dominio objetivo como distintos niveles de datos fuera de distribución: imágenes claramente alejadas del dominio, imágenes naturales no arquitectónicas, edificios sin relevancia monumental y monumentos de estilos no reconocidos por MonuMAI.

La evaluación se realiza tanto por capas como de forma global. Cada capa resuelve un problema binario: la capa 0 filtra imágenes del mundo lejano; la capa 1 distingue edificios frente a no edificios; la capa 2 separa monumentos de edificios no monumentales; y la capa 3 decide si un monumento pertenece a un estilo conocido o desconocido. Para cada una se reportan accuracy, precision, recall,

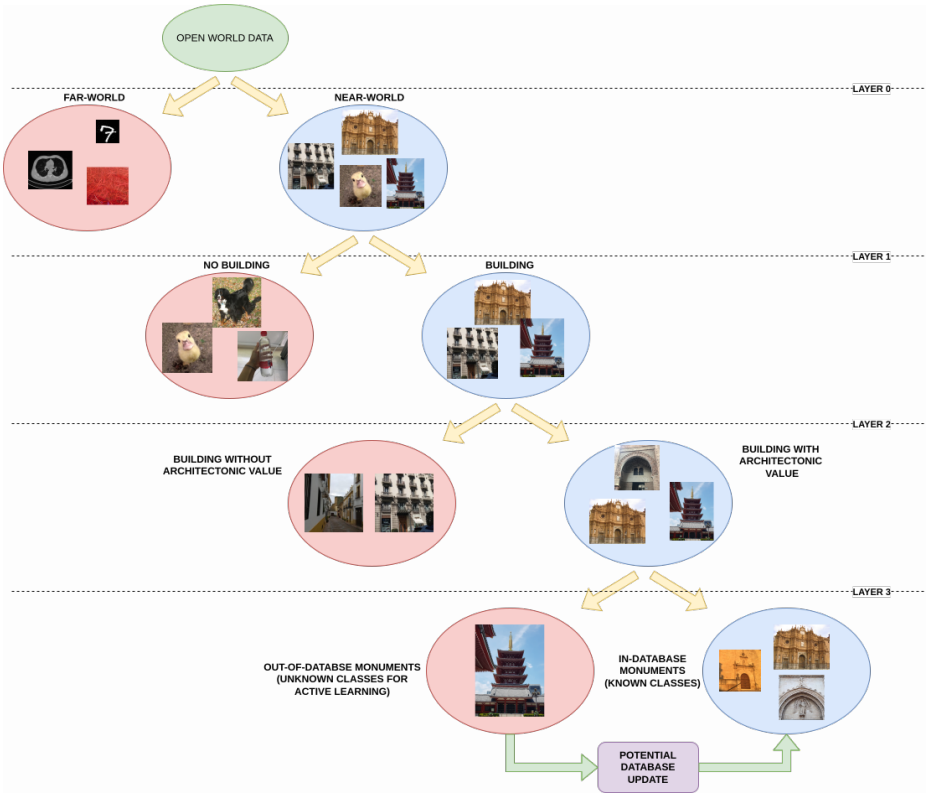


Figura 2: Cascada de dicotomías semánticas para la detección OOD en MonuMAI.

Cuadro 1: Distribución de imágenes en *MonuTest-OOD*.

Categoría	**Número de imágenes**
Mundo lejano (OOD)	1550
Mundo cercano (OOD)	1518
Edificios (OOD)	1467
Monumentos no conocidos (OOD)	781
Monumentos conocidos (en distribución)	1038
Total	**6354**

F1 y AUROC. En la evaluación global, la clase positiva corresponde únicamente a los monumentos de estilo conocido, mientras que todas las demás categorías se consideran OOD.

Además, los resultados se comparan con el mecanismo OOD original de MonuMAI. Este sistema rechaza una imagen cuando MonuMAI-KED no detecta

elementos arquitectónicos. Sin embargo, este criterio resulta limitado, ya que ciertos patrones visuales no monumentales pueden ser interpretados erróneamente como elementos arquitectónicos, permitiendo la aceptación de imágenes fuera del dominio.

4.2. Resultados cuantitativos

El Cuadro 2 resume el rendimiento de cada capa. En conjunto, todas las etapas muestran una separación sólida entre las clases correspondientes, con valores de AUROC superiores a 0.95 en todas las capas. La capa 3 presenta el problema más difícil, ya que debe discriminar entre monumentos de estilos conocidos y monumentos estilísticamente cercanos pero no contemplados por el sistema.

Cuadro 2: Resultados por capa de la metodología propuesta sobre MonuMAI.

Capa	**Decisión**	**ACC**	**PREC**	**REC**	**F1**	**AUROC**
0	Mundo cercano frente a mundo lejano	0.9397	0.9076	0.9776	0.9413	0.9601
1	Edificio frente a no edificio	0.9266	0.8878	0.9796	0.9314	0.9774
2	Monumento frente a edificio	0.9394	0.9576	0.9318	0.9446	0.9502
3	Estilo conocido frente a desconocido	0.9245	0.7851	0.8120	0.7983	0.9660

Los resultados muestran que la cascada reduce progresivamente el espacio de entrada hasta llegar al dominio realmente relevante para MonuMAI. Las dos primeras capas eliminan contenido claramente ajeno al dominio arquitectónico; la tercera descarta edificios sin relevancia monumental; y la última capa identifica monumentos cuyo estilo no forma parte de la taxonomía reconocida. Esta descomposición evita que una única frontera OOD deba resolver simultáneamente problemas visuales muy diferentes.

La comparación global con el sistema original de MonuMAI se muestra en el Cuadro 3. La metodología jerárquica propuesta obtiene una mejora clara en Accuracy, Precision, F1 y AUROC. La principal diferencia se observa en Precision, que pasa de 0.4130 a 0.8823. Esto indica que las imágenes aceptadas por la cascada como monumentos de estilo conocido son mucho más fiables. El sistema original mantiene un recall superior, pero a costa de aceptar una gran cantidad de contenido OOD como si perteneciera al dominio objetivo.

Estos resultados evidencian un comportamiento conservador. La metodología sacrifica parte del Recall para reducir de forma sustancial los falsos positivos, lo cual resulta adecuado en un contexto de patrimonio cultural, donde aceptar imágenes no monumentales o estilos no soportados puede degradar la fiabilidad del sistema. En consecuencia, no solo se mejora la detección OOD, sino que también se producen decisiones más interpretables, ya que cada rechazo queda asociado a una frontera semántica concreta: imagen no natural, no edificio, edificio no monumental o monumento de estilo desconocido.

Cuadro 3: Resultados globales de detección OOD en *MonuTest-OOD*. La clase positiva corresponde a monumentos de estilo conocido.

Modelo	**ACC**	**PREC**	**REC**	**F1**	**AUROC**
Propuesta	**0.9417**	**0.8823**	0.7625	**0.8180**	**0.9642**
MonuMAI original	0.7748	0.4130	**0.8991**	0.5660	0.9268

4.3. Análisis cualitativo

El análisis visual de los errores confirma que las equivocaciones restantes no son arbitrarias, sino que se concentran en casos frontera. En la capa 0, algunas texturas repetitivas pueden compartir estadísticas de bajo nivel con fotografías naturales. En la capa 1, fuentes, esculturas o composiciones con estructuras similares a fachadas pueden ser confundidas con edificios. En la capa 2, ciertos edificios genéricos con simetría, portales u ornamentación pueden parecer monumentos. Finalmente, en la capa 3, los errores aparecen principalmente en monumentos con señales estilísticas mixtas, vistas degradadas o rasgos cercanos a estilos conocidos.

Este comportamiento es coherente con el diseño de la cascada. Las decisiones ambiguas no se fuerzan en las capas externas, sino que se trasladan a filtros más específicos. Así, cada capa actúa sobre el nivel semántico que le corresponde y los errores pueden abordarse mediante mejoras localizadas: ampliar negativos difíciles para la capa 1, enriquecer el conjunto de edificios no monumentales para la capa 2, o incorporar más ejemplos de estilos desconocidos y casos híbridos para la capa 3.

4.4. Conclusión experimental

La experimentación demuestra que la metodología propuesta mejora de forma significativa la robustez de MonuMAI frente a entradas fuera de distribución. Frente al filtrado original, basado únicamente en la ausencia de detecciones arquitectónicas, la cascada jerárquica ofrece una separación más precisa, interpretable y mantenible del dominio operativo. Su principal ventaja es la reducción drástica de falsos positivos, reflejada en el aumento de Precision de 0.4130 a 0.8823 y en la mejora del F1 global de 0.5660 a 0.8180. Aunque el Recall disminuye, esta pérdida responde a una estrategia conservadora que prioriza la fiabilidad de las imágenes aceptadas. Además, las imágenes rechazadas en la última capa constituyen un conjunto útil para revisión experta y futura ampliación de la taxonomía de estilos.

5. Conclusiones

En este trabajo se ha presentado una metodología jerárquica para la detección de entradas fuera de distribución en MonuMAI, basada en una cascada de

decisiones semánticas progresivas. En lugar de abordar la detección OOD como una decisión única entre muestras dentro y fuera de distribución, la propuesta descompone el problema en varios niveles especializados, permitiendo distinguir entre imágenes alejadas del dominio, imágenes no arquitectónicas, edificios sin relevancia monumental y monumentos pertenecientes a estilos no contemplados por el sistema.

Los resultados experimentales muestran que esta estrategia mejora de forma notable la fiabilidad del pipeline original de MonuMAI. La cascada reduce sustancialmente la aceptación errónea de imágenes fuera del dominio, incrementando la precisión global y el valor de F1 frente al mecanismo original de filtrado. Aunque esta mejora implica una reducción del *recall*, el comportamiento resultante es coherente con un escenario de patrimonio cultural, donde resulta preferible rechazar casos dudosos antes que proporcionar clasificaciones potencialmente incorrectas con alta confianza.

Además del rendimiento cuantitativo, la metodología aporta ventajas en términos de interpretabilidad y mantenibilidad. Cada rechazo queda asociado a una frontera semántica concreta, lo que permite identificar la causa aproximada por la que una imagen no pertenece al dominio operativo previsto. Esta propiedad facilita el análisis de errores, la curación de datos y la mejora localizada de cada capa, especialmente en casos frontera.

Como líneas futuras, será necesario ampliar y diversificar los conjuntos de entrenamiento de cada filtro, incorporar más ejemplos difíciles y estudiar mecanismos de calibración adaptativa que mantengan la estabilidad ante cambios en la distribución de datos. Asimismo, las imágenes rechazadas en la última capa constituyen una fuente valiosa para revisión experta y aprendizaje activo, permitiendo extender progresivamente la taxonomía de estilos arquitectónicos reconocidos por MonuMAI. Los resultados obtenidos indican que la descomposición jerárquica de la detección OOD en decisiones semánticamente significativas constituye una vía prometedora para construir sistemas de visión más robustos, interpretables y adecuados para escenarios reales de mundo abierto.

Agradecimientos

Esta investigación es parte del Proyecto TSI-100927-2023-1, financiado por Plan de Recuperación, Transformación y Resiliencia Europeo, de la Unión Europea-Next Generation EU a través del Ministerio para la Transformación Digital y de la Función Pública y está parcialmente financiada por el Ministerio de Ciencia y Tecnología de España en el marco del proyecto PID2023-150070NB-I00, financiado por MICIU/AEI/10.13039/501100011033 y por FEDER, UE.

Referencias

1. Azizmalayeri, M., Soltani Moakhar, A., Zarei, A., Zohrabi, R., Manzuri, M., Rohban, M.H.: Your out-of-distribution detection method is not robust! Advances in Neural Information Processing Systems **35**, 4887–4901 (2022)

2. Barcina-Blanco, M., Lobo, J.L., Garcia-Bringas, P., Del Ser, J.: Managing the unknown: a survey on open set recognition and tangential areas. arXiv preprint arXiv:2312.08785v2 (2023). https://doi.org/10.48550/arXiv.2312.08785
3. Denouden, T., Salay, R., Czarnecki, K., Abdelzad, V., Phan, B., Vernekar, S.: Improving reconstruction autoencoder out-of-distribution detection with mahalanobis distance. arXiv preprint arXiv:1812.02765 (2018)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
5. Hesham, S., Khaled, R., Yasser, D., Refaat, S., Shorim, N., Ismail, F.H.: Monuments recognition using deep learning vs machine learning. In: 2021 IEEE 11th annual computing and communication workshop and conference (CCWC). pp. 0258–0263. IEEE (2021)
6. Lamas, A., Tabik, S., Cruz, P., Montes, R., Martínez-Sevilla, Á., Cruz, T., Herrera, F.: Monumai: Dataset, deep learning pipeline and citizen science based app for monumental heritage taxonomy and classification. Neurocomputing **420**, 266–280 (2021)
7. Liu, W., Wang, X., Owens, J., Li, Y.: Energy-based out-of-distribution detection. Advances in neural information processing systems **33**, 21464–21475 (2020)
8. Liu, X., Lochman, Y., Zach, C.: Gen: Pushing the limits of softmax-based out-of-distribution detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 23946–23955 (2023)
9. Pertusa, A., Gallego, A.J., Bernabeu, M.: Mirbot: A collaborative object recognition system for smartphones using convolutional neural networks. Neurocomputing **293**, 87–99 (2018)
10. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems **28** (2015)
11. Sevillano-García, I., Luengo, J., Herrera, F.: Stood-x methodology: using statistical nonparametric test for ood detection large-scale datasets enhanced with explainability. arXiv preprint arXiv:2504.02685 (2025)
12. Shoeb, Y., Nowzad, A., Gottschalk, H.: Out-of-distribution segmentation in autonomous driving: Problems and state of the art. arXiv preprint arXiv:2503.08695 (2025)
13. Wiatowski, T., Bölcskei, H.: A mathematical theory of deep convolutional neural networks for feature extraction. IEEE Transactions on Information Theory **64**(3), 1845–1866 (2017)
14. Yang, J., Zhou, K., Li, Y., Liu, Z.: Generalized out-of-distribution detection: A survey. International Journal of Computer Vision **132**(12), 5635–5662 (2024)
15. Yuan, L., Chen, Y., Cui, G., Gao, H., Zou, F., Cheng, X., Ji, H., Liu, Z., Sun, M.: Revisiting out-of-distribution robustness in nlp: Benchmarks, analysis, and llms evaluations. Advances in Neural Information Processing Systems **36**, 58478–58507 (2023)
16. Zamzmi, G., Venkatesh, K., Nelson, B., Prathapan, S., Yi, P., Sahiner, B., Delfino, J.G.: Out-of-distribution detection and radiological data monitoring using statistical process control. Journal of Imaging Informatics in Medicine pp. 1–19 (2024)
17. Zhao, P., Miao, Q., Song, J., Qi, Y., Liu, R., Ge, D.: Architectural style classification based on feature extraction module. Ieee Access **6**, 52598–52606 (2018)
18. Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A.: Places: A 10 million image database for scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence (2017)