

Detección de residuos urbanos mal ubicados mediante visión por ordenador e IA

Frutos, V.^a, Pérez, B.^a, Campos, S.^{a,*}, Barrera, A.^a, García, O.^b, García, F.^a, Al-Kaff, A.^a

^aDepartamento de Ingeniería de Sistemas y Automática, Universidad Carlos III de Madrid, Leganés, 28911, Madrid, España.

^bNOEDGE SL, Villaviciosa de Odón, 28670, Madrid, España.

To cite this article: Frutos, V., Pérez, B., Campos, S., Barrera, A., García, O., García, F., Al-Kaff, A. 2023. Urban anomaly detection based on computer vision and AI - Waste Management. XLIV Jornadas de Automática. 855-860. <https://doi.org/10.17979/spudc.9788497498609.855>

Resumen

La gestión de los residuos en las zonas urbanas es un aspecto esencial para la sostenibilidad y el mantenimiento del medio ambiente, así como de la salud pública. Además, el rápido crecimiento de las ciudades y de la población ha provocado un aumento de la generación de residuos que supone un reto global ya que en muchas ciudades del mundo la gestión de estos residuos se realiza de forma manual con las limitaciones típicas asociadas como el error humano, la fatiga ante tareas repetitivas o el coste económico. En este artículo se propone un sistema basado en visión por computador e inteligencia artificial con el fin de automatizar la detección de anomalías para la gestión de residuos en zonas urbanas y mejorar la eficacia en la etapa de recogida. También se expone el estudio de las pruebas realizadas en entornos reales así como un análisis de los resultados obtenidos y el proceso de mejora en las detecciones.

Palabras clave: Redes neuronales, Percepción y detección, Repercusiones de la automatización en el medio ambiente, la movilidad, la energía, la salud y la seguridad.

Detection of misplaced urban waste using computer vision and AI

Abstract

Waste management in urban areas is an essential aspect of environmental sustainability and maintenance as well as public health. In addition, the rapid growth of cities and population has led to an increase in waste generation which is a global challenge since in many cities around the world waste management is performed manually with the typical associated limitations such as human error, repetitive task fatigue or economic cost. This paper proposes a system based on computer vision and artificial intelligence in order to automate the detection of anomalies for waste management in urban areas and improve efficiency in the collection stage. The study of the tests carried out in real environments is also presented, as well as an analysis of the results obtained and the process of improvement in the detections.

Keywords: Neural networks, Perception and sensing, Environmental, mobility, energy, health and safety implications of automation.

1. Introducción

La generación de residuos es un problema importante tanto a escala mundial como en la Unión Europea (UE) (Minelgaité and Liobikienė, 2019). Concretamente, en las ciudades, el rápido crecimiento de la población y la generación masiva de residuos sólidos urbanos (RSU) resaltan la importancia de la gestión municipal para evitar las consecuencias negativas. En

primer lugar, la presencia de basura en las calles y espacios públicos crea un entorno poco atractivo y desagradable. Los desechos arrojados de manera irresponsable generan un impacto negativo en la ciudad, lo que puede afectar la calidad de vida de los residentes.

En segundo lugar, la concentración de basura también puede ser un factor que contribuye a problemas de salud. Los dese-

*Autor para correspondencia: secampos@ing.uc3m.es
Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

chos acumulados pueden atraer plagas y vectores que transmitan enfermedades a los seres humanos y también pueden resultar en la contaminación del agua y del suelo, lo que representa un riesgo para la salud pública. Además, la presencia de productos químicos tóxicos o materiales peligrosos presentes en la basura puede causar problemas respiratorios o dermatológicos.

Por último, la generación masiva de basura proveniente de las ciudades también tiene un impacto ambiental significativo. Los vertederos abarrotados emiten gases de efecto invernadero, como metano, que contribuyen al cambio climático. La falta de prácticas de reciclaje y reducción de residuos conduce a un mayor agotamiento de los recursos naturales y una mayor demanda de energía y materias primas. Además, los desechos plásticos y otros materiales no biodegradables pueden terminar en los océanos y ríos, amenazando la vida marina y afectando los ecosistemas acuáticos. En (Gaidajis et al., 2010) los autores revisaron los problemas medioambientales y de salud pública generados por los aparatos electrónicos desechados de forma inapropiada, la generación anual de este tipo de residuos aumenta anualmente según un estudio de la ONU (Forti et al., 2020). Es por esto que en la actualidad se están investigando y desarrollando diversos proyectos para la detección y gestión de estos residuos en los núcleos urbanos. En (Zhang et al., 2019) se sigue esta corriente de investigación mediante el uso de tecnologías de *Deep Learning* calculando una estimación del nivel de limpieza de una zona a partir de imágenes tomadas desde diversos vehículos por la zona. Por otro lado, en (Majchrowska et al., 2022b), los autores realizan una revisión del estado del arte de tecnologías y conjuntos de datos utilizados para la detección de desechos con *Deep Learning*.

Por otro lado, en la actualidad no está realmente estandarizada la detección y clasificación de todos los tipos de residuos. Además, tal y como se observa en el estudio de Sylwia Majchrowska (Majchrowska et al., 2022a), todavía no existen grandes conjuntos de imágenes útiles para un uso genérico, dado que el aspecto tanto del entorno como del residuo a detectar difiere mucho entre un lugar y otro, dificultando la flexibilidad en las detecciones de estos sistemas automáticos.

El presente trabajo se engloba en esta línea de investigación presentando un sistema de detección de anomalías en las inmediaciones de contenedores para encontrar basura desechada de forma inadecuada. El objetivo principal es mejorar la eficiencia en la fase de recolección y evitar la acumulación automatizando un proceso que hoy en día se realiza manualmente.

Este artículo se estructura de la siguiente forma: la sección 2 describe las arquitecturas de redes neuronales utilizadas, también se mencionan los problemas generados por la naturaleza del proyecto en cuanto al conjunto de datos y cómo se han reentrenado las redes a consecuencia. La siguiente sección, la sección 3, desarrolla los experimentos llevados a cabo en entornos reales en zonas urbanas y los resultados obtenidos por los mismos. Finalmente, las conclusiones del artículo se presentan en la sección 4.

2. Metodología

Al inicio del proyecto, se realizó un estudio de los diferentes trabajos que abordaban este problema y los métodos que

empleaban. En algunos, se mostraba el funcionamiento de diversas arquitecturas para la detección de residuos entre los que destacaba la arquitectura *Faster R-CNN* por ser la de mejores resultados, obteniendo métricas elevadas y tiempos de ejecución bajos. Consecuentemente, se acordó el empleo de esta arquitectura para el desarrollo de este proyecto.

2.1. Arquitectura *Faster R-CNN*

En este proyecto, se empleó la librería *Detectron2* (Wu et al., 2019) desarrollada por Facebook que permite el entrenamiento y uso de redes neuronales para la detección, segmentación y clasificación de imágenes, según el objetivo requerido. Concretamente, en este trabajo se entrenó un modelo de red *Faster R-CNN* con el dataset de imágenes de residuos en las calles y alrededores de contenedores con la finalidad de obtener un modelo capaz de detectar y clasificar los restos en las calles.

Faster R-CNN se basa en la arquitectura *R-CNN*, la cual en primer lugar destaca las áreas de interés de las imágenes para que posteriormente se realice la clasificación a partir de la CNN, lo que permite solucionar el problema de las redes neuronales donde solo se conoce el aspecto del objeto en la imagen, no su posición. De esta forma, mediante *R-CNN* es posible obtener las detecciones de objetos, así como su posición en la imagen mediante el procesamiento mostrado en la Figura 1.

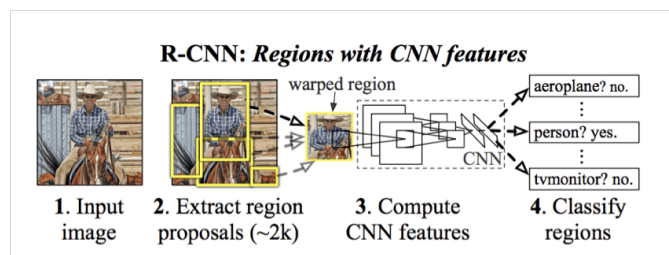


Figura 1: Procesamiento de imágenes mediante arquitectura R-CNN

No obstante, un inconveniente de esta arquitectura son los altos tiempos de ejecución, lo que la hace inviable para detecciones en tiempo real. De este modo, surgió *Faster R-CNN*, una variante que soluciona estos problemas de tiempos de ejecución. Para ello, emplea una primera convolución de las zonas de interés para posteriormente aplicar una serie de poolings para reducir el tamaño de datos y así disminuir drásticamente los tiempos de ejecución. Así, el procesamiento de la imagen mediante esta variante de *R-CNN* resultaría como se muestra en la Figura 2.

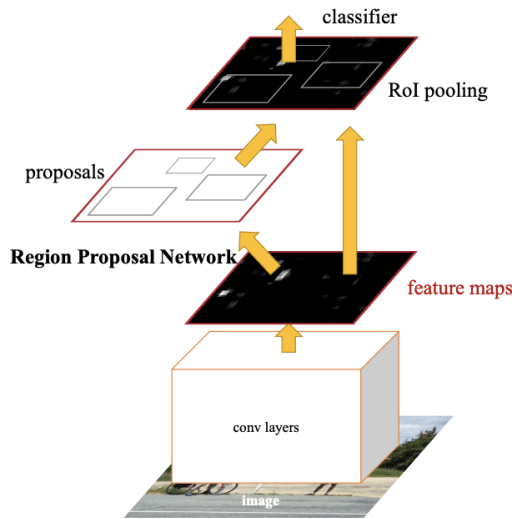


Figura 2: Arquitectura Faster R-CNN



Figura 3: Distintas imágenes del dataset creado y etiquetado manualmente.

Finalmente, se creó un código con el que se entrenó el modelo para la clasificación deseada de contenedor y basura a partir de las imágenes del dataset etiquetado.

2.2. Dataset generado

Para el desarrollo de este proyecto se creó un conjunto de imágenes propias. Primero, se obtuvieron imágenes de contenedores y basuras con una cámara de un teléfono móvil. Con los datos iniciales se preprocesaron eliminando las que no se tenía una buena visión debido a la iluminación o debido a que eran borrosas. Una vez eliminadas las imágenes borrosas, se comenzó a etiquetar las restantes de forma manual obteniendo así un conjunto inicial de 7548 imágenes para el entrenamiento, que se subdividieron en 7033 imágenes para entrenamiento, 257 imágenes para validación y 258 para test, resultando en una distribución del dataset en 93.1 % para entrenamiento, 3.4 % para validación y 3.5 % para test. Una vez que se obtuvo este dataset inicial, se procedió al primer entrenamiento de la red. Con el modelo entrenado resultante, se realizó un nuevo entrenamiento pero de manera semi supervisada y se consiguió un conjunto de datos proporcionado como se muestra en la Figura 4. Cabe destacar que, con la intención de poder realizar comparaciones entre los distintos resultados de los entrenamientos, las diversas ampliaciones de dataset generadas se introdujeron en su totalidad en el dataset de entrenamiento, manteniendo las mismas imágenes de validación y test iniciales. Además, en la Figura 4 también se puede observar el porcentaje de etiquetas finales de 'basura' y 'contenedor' en el dataset ampliado. De esta manera, se trabajó con un total de 15998 etiquetas de 'contenedor' y 17671 de 'basura' en el conjunto para entrenamiento. Así mismo, los conjuntos de validación y test en conjunto contaban con un total de 1163 y 1499 etiquetas de 'contenedor' y 'basura' respectivamente. El tipo de imágenes de este conjunto se puede observar en la Figura 3, donde se observan los residuos desde la perspectiva de un peatón.

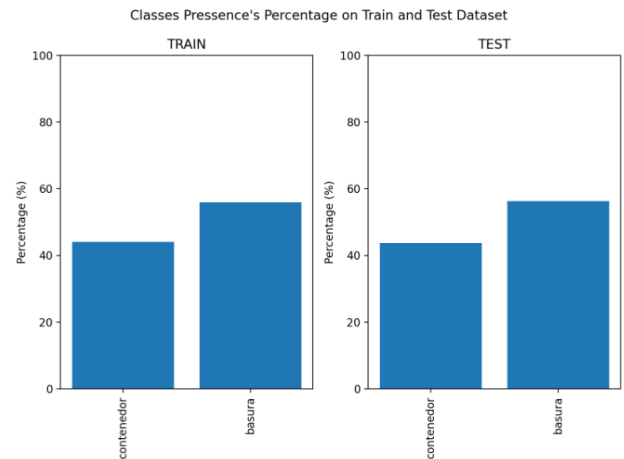


Figura 4: Porcentaje de cada clase del dataset binario

Además del conjunto obtenido anteriormente, el dataset fue ampliado con imágenes de un entorno real tomadas desde una cámara embarcada en un vehículo y con la perspectiva más similar a la aplicación final. Para ello, fue necesario implementar las siguientes restricciones y características: el vehículo no debía superar la velocidad de 20 Km/h y la frecuencia de grabación de las cámaras debía ser de 10 Hz. Por otro lado, se descartaron las imágenes que eran muy parecidas debido a que el vehículo no estaba continuamente en movimiento. Y se incluyeron imágenes de residuos desde distintas perspectivas para aumentar el volumen de datos, como se puede observar en la Figura 5. Así, el porcentaje de cada clase es similar, obteniendo un 44,24 % de detecciones de basura y un 55,76 % de detecciones de contenedores.

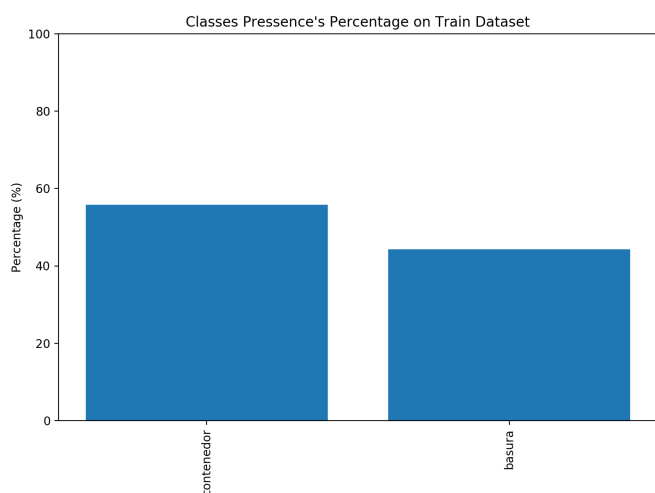


Figura 5: Porcentaje de las clases del dataset obtenido con un vehículo

Además, se consiguió un conjunto de datos más variado que incrementó la robustez del modelo frente a cambios de iluminación y mejoró la cantidad de detecciones de residuos que se encontraban ocluidos. Así, con estas ampliaciones, finalmente se trabajó con un dataset final de 1280 imágenes, distribuidas en 1031 para entrenamiento y 249 % para validación. Por último, en la Figura 6 se puede observar algunas imágenes con distinto tipos de escena de residuos.



Figura 6: Distintas imágenes del dataset creado y etiquetado manualmente.

3. Experimentación

Para el desarrollo de este proyecto, se realizaron diversas pruebas del sistema en entornos reales, empleando la plataforma *ATLAS* (de Miguel et al., 2020) para la consecución del proyecto y su uso en tiempo real con el objetivo de observar los resultados y realizar las modificaciones que fueran necesarias.

3.1. Hardware

Con respecto a los dispositivos empleados en este proyecto, se dispuso de dos cámaras *Basler AC* y un sensor *GPS novatel*

instalados en un vehículo siguiendo la distribución de la Figura 7 para la puesta en práctica del proyecto.



Figura 7: Disposición de los diferentes sensores en el vehículo

De este modo, es posible observar ambos lados de la carretera de manera simultánea, gracias a la disposición de una cámara a cada lado. Además, se diseñó y fabricó una pieza en 3D para dotar a las cámaras de un ángulo de 30° en el eje Y y de 20° en el eje Z, de modo que la perspectiva de las mismas fuera apuntando hacia el suelo próximo, que es donde se deberán detectar los residuos alrededor de los contenedores.

3.2. Escenario

En cuanto al entorno real en el que se probó el funcionamiento del sistema, se circuló por zonas urbanas de Madrid observando dónde existían contenedores y residuos en la vía para comprobar el funcionamiento de la detección. Por otro lado, también se realizaron pruebas en un entorno controlado en el que se depositaron restos en las cercanías de los contenedores de diversas formas: a simple vista, parcialmente ocluido, en zonas con mayor iluminación, zonas de sombra, entre otros, para observar el correcto funcionamiento y la robustez del proyecto en las diferentes situaciones que se pueden dar durante su empleo.

3.3. Resultados

Durante el desarrollo de este proyecto se fueron realizando pruebas y anotando los resultados obtenidos para comprobar cómo se iba progresando en el trabajo. De esta forma, con cada modelo que se entrenaba con los diferentes conjuntos de imágenes etiquetados, se usaban imágenes de validación para así comprobar su resultado utilizando matrices de confusión. Una matriz de confusión muestra las predicciones de todas las clases frente a los valores reales y permite obtener los verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos. Se puede observar en la Tabla 1, que pertenece al modelo entrenado con el conjunto inicial etiquetado manualmente, que los resultados eran buenos. Sin embargo se observa que tiene una mayor precisión en la detección de la clase *contenedor* con respecto a la clase *basura*.

Posteriormente se reentrenó el modelo con las nuevas imágenes etiquetadas. Al utilizar un conjunto más amplio de datos de entrenamiento se comprobó que los resultados del modelo mejoraron, como se puede comprobar en la Tabla 2, donde se puede apreciar un aumento del valor *FI* del modelo hasta un 80 %.

Asimismo se puede observar que debido a la diferencia en las detecciones entre ambas clases, los resultados en la precisión de la clase contenedor son mayores que en las detecciones de residuos, teniendo en los contenedores una precisión del 99 % y unos valores de recall similares. Como se puede observar en la Figura 8, la curva precisión-recall de la clase *contenedor* se asemeja mucho a una curva perfecta. En cambio, la curva de la clase *residuos* tan solo logra alcanzar un valor máximo de 85 %. Este valor se debe a que la cantidad de residuos diversos que se puede encontrar en el entorno urbano.

Con el modelo entrenado con las imágenes anteriores y el conjunto tomado desde el vehículo, se obtuvieron mejores resultados tal y como se puede observar en la Tabla 3. Además, con la incorporación de estas nuevas imágenes al dataset de entrenamiento, se consiguió entrenar un modelo más robusto.

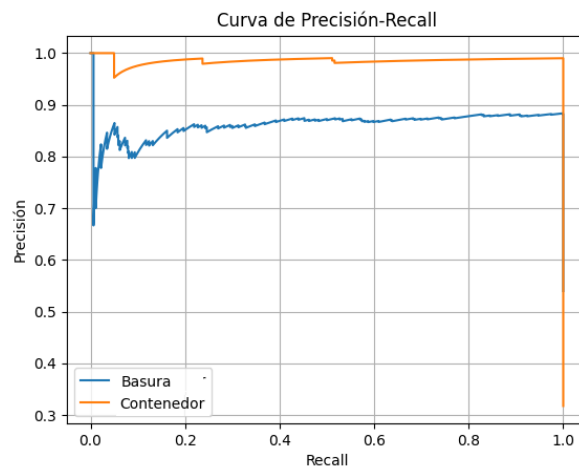


Figura 8: Curva PR del modelo binario

De este modo, en la puesta en práctica con el último modelo entrenado, se obtuvieron buenos resultados en las detecciones como los que se pueden observar en la Figura 3.3. Además, es posible comprobar la robustez del sistema tanto a cambios de iluminación como a posibles oclusiones, incluso adaptándose a cualquier tamaño de residuo.

		Valores Reales		
		Contenedor	Basura	Ninguna
Predic.	Contenedor	979	4	0
	Basura	3	921	578
	Ninguna	172	416	–

Tabla 1: Matriz de confusión con el dataset inicial

		Valores Reales		
		Contenedor	Basura	Ninguna
Predic.	Contenedor	1024	3	0
	Basura	7	979	529
	Ninguna	129	377	–

Tabla 2: Matriz de confusión con nuevas imágenes manual y automáticamente etiquetadas

		Etiquetas		
		Contenedor	Basura	Ninguna
Predic.	Contenedor	952	2	0
	Basura	30	648	851
	Ninguna	188	536	–

Tabla 3: Matriz de confusión del dataset final con imágenes desde la perspectiva final

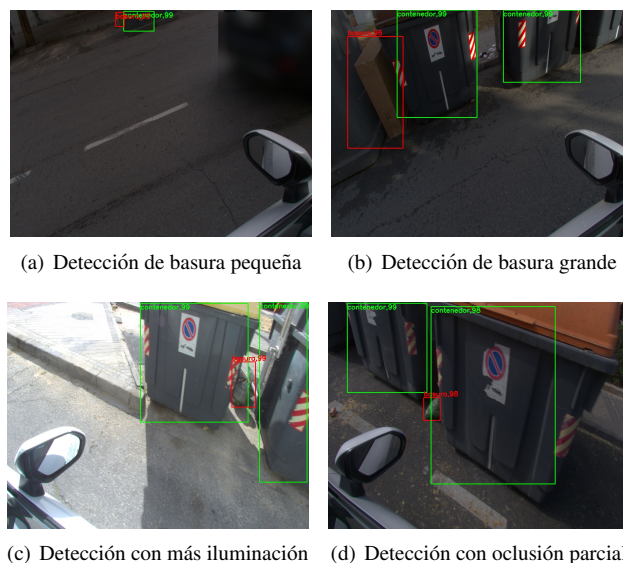


Figura 9: Distintas detecciones con el modelo entrenado final

Respecto al tiempo de inferencia de la red es 70,1 milisegundos en cada frame de modo que es posible implementar todo el modelo en tiempo real. Debido a este tiempo de ejecución y a los valores obtenidos en las métricas se observó que la arquitectura *Faster R-CNN* tiene un rendimiento en tiempo real óptimo.

4. Conclusiones

Finalmente, en este proyecto se ha logrado desarrollar un sistema de detección de residuos funcional basado en una red neuronal *Faster R-CNN* y puesto en práctica en un entorno real. De esta forma, esta herramienta podría ser empleada por las instituciones públicas en un futuro para la gestión de residuos en los entornos urbanos y el cuidado del medio ambiente. Además, siguiendo los resultados obtenidos en este proyecto, se podría desarrollar una aplicación a mayor escala para su aplicación en una ciudad, de manera que se lograra optimizar la recogida de residuos en entornos urbanos de mayor magnitud.

Con respecto a las métricas obtenidas, se obtuvieron buenos resultados, aunque existen algunos residuos no detectados por el modelo o casos de falsos positivos con objetos. Para solventar estos problemas, se podría aumentar el conjunto de datos con nuevas detecciones de residuos más diversos y en entornos más variados. No obstante, a pesar de que es posible obtener ciertas mejoras mediante este aumento en el número de imágenes de entrenamiento, es posible que las mejoras que se obtendrían en la precisión no sean realmente significativas debido a diversas limitaciones. Otra de las posibles mejoras directas en la arquitectura podrían ser la implementación de una etapa de clasificación para los diferentes tipos de residuos o un algoritmo de seguimiento o *tracking* para mejorar la robustez del sistema.

Otro de los problemas observados en el desarrollo del proyecto está relacionado con la colocación de los residuos alrededor de los contenedores. Comúnmente, las personas depositan la basura en el lado de la acera, por lo que desde la carretera por donde circula el vehículo con el sistema integrado, se producen oclusiones parciales o incluso totales del residuo desde la visión del vehículo y en consecuencia, el sistema es incapaz de detectarlo. Con el objetivo de evitar este problema, se podría hacer uso de otros sistemas o sensores para suplir esta falta de visión, como la inserción de un sensor LIDAR o la implementación de la arquitectura en sistemas con un punto de vista diferente como drones o cámaras colocadas en infraestructura.

Agradecimientos

Acción financiada por la Comunidad de Madrid a través de la línea "Estímulo a la Investigación de Jóvenes Doctores/as" del Convenio Plurianual con la UC3M (APBI-CM-UC3M), en el marco del V PRICIT (V Plan Regional de Investigación Científica e Innovación Tecnológica), y el Gobierno español a través de los proyectos ID2021-128327OA-I00 y TED2021-129374A-I00 financiados por MCIN/AEI/10.13039/501100011033.



Referencias

- de Miguel, M. Á., Moreno, F. M., Marín-Plaza, P., Al-Kaff, A., Palos, M., Martín, D., Encinar-Martín, R., García, F., 2020. A research platform for autonomous vehicles technologies research in the insurance sector. *Applied Sciences* 10 (16), 5655.
- Forti, V., Baldé, C. P., Kuehr, R., Bel, G., 2020. The global e-waste monitor 2020. United Nations University (UNU), International Telecommunication Union (ITU) & International Solid Waste Association (ISWA), Bonn/Geneva/Rotterdam 120.
- Gaidajis, G., Angelakoglou, K., Aktsoğlu, D., 2010. E-waste: environmental problems and current management. *Journal of Engineering Science and Technology Review* 3 (1), 193–199.
- Majchrowska, S., Mikołajczyk, A., Ferlin, M., Klawikowska, Z., Plantykowski, M. A., Kwasigroch, A., Majek, K., 2022a. Deep learning-based waste detection in natural and urban environments. *Waste Management* 138, 274–284.
- Majchrowska, S., Mikołajczyk, A., Ferlin, M., Klawikowska, Z., Plantykowski, M. A., Kwasigroch, A., Majek, K., 2022b. Deep learning-based waste detection in natural and urban environments. *Waste Management* 138, 274–284. DOI: <https://doi.org/10.1016/j.wasman.2021.12.001>
- Minelgaitė, A., Liobikienė, G., 2019. Waste problem in european union and its influence on waste management behaviours. *Science of The Total Environment* 667, 86–93. DOI: <https://doi.org/10.1016/j.scitotenv.2019.02.313>
- Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y., Girshick, R., 2019. Detectron2. <https://github.com/facebookresearch/detectron2>.
- Zhang, P., Zhao, Q., Gao, J., Li, W., Lu, J., 2019. Urban street cleanliness assessment using mobile edge computing and deep learning. *IEEE Access* 7, 63550–63563.