

Generación de datos virtuales de objetos de cocina para entrenamiento de redes neuronales



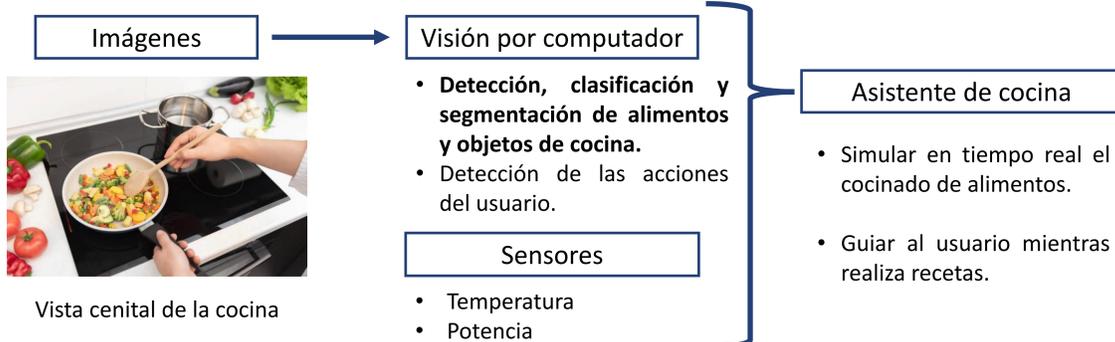
Universidad Zaragoza

Javier Fañanás-Anaya, Gonzalo López-Nicolás, Carlos Sagüés
 javierfa@unizar.es, gonlopez@unizar.es, csagues@unizar.es
 Junio - 2023

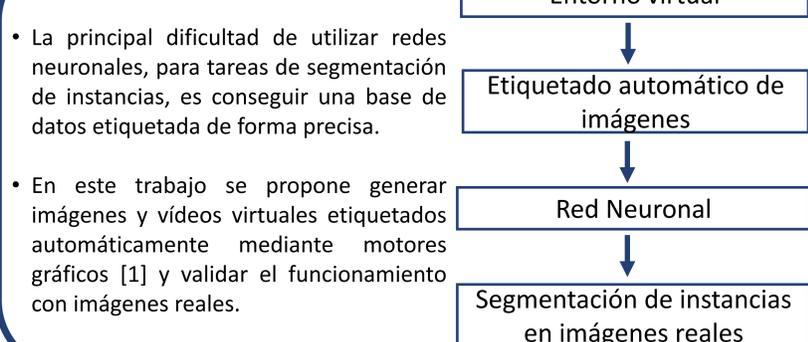


Instituto Universitario de Investigación de Ingeniería de Aragón
 Universidad Zaragoza

Asistente de cocina basado en IA



Problema



Entorno virtual

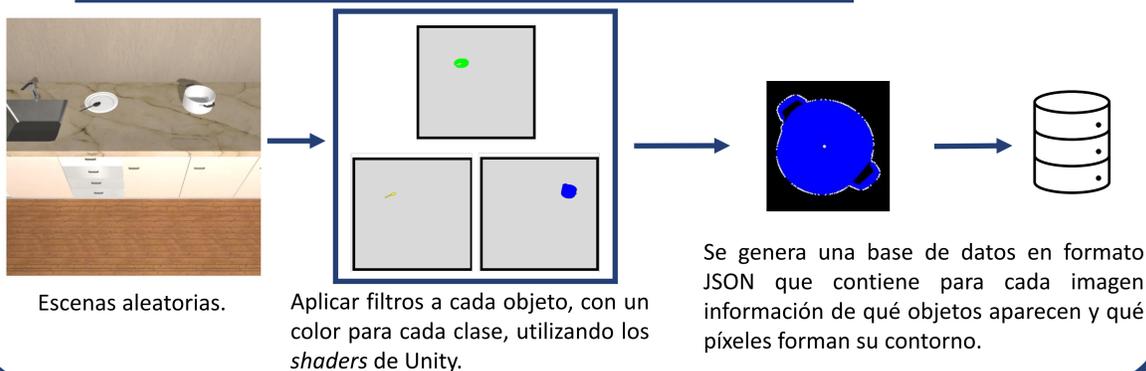


Se ha utilizado Unity para crear un entorno virtual realista de una cocina. Incluyendo detalles como la iluminación exterior e interior, así como los reflejos de la encimera y vitrocerámica.

El problema se ha centrado en la segmentación de instancias de 5 clases: Plato, cubierto, vaso, cazuela y sartén.

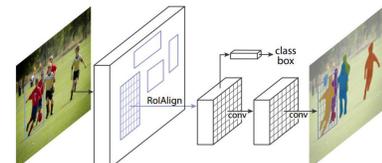
Generación de escenas aleatorias: Varía la intensidad de la iluminación, los objetos que aparecen y su posición, y la posición/ángulo de la cámara.

Etiquetado automático de imágenes



Red Neuronal

- Se ha utilizado la red neuronal convolucional Mask R-CNN [2].
- La arquitectura de la red está orientada para la segmentación de instancias.
- La red se ha entrenado con el algoritmo de 3 fases propuesto en [2], y haciendo *transfer learning*.



Base de datos: 5500 imágenes

- Entrenamiento: 4000
- Validación: 1000
- Test: 500

Validación experimental

Validación a partir de las 500 imágenes virtuales de test

Imágenes virtuales



Para evaluar de forma cualitativa cómo funciona este modelo con imágenes reales se han utilizado distintas distribuciones, incluyendo vídeos de EPIC-KITCHENS [3].

Conclusiones:

- Las métricas obtenidas con los datos virtuales de test son muy positivas. La mayoría de los falsos negativos son provocados por objetos que aparecen en los bordes de la imagen, donde solo ocupan unos pocos píxeles. Los falsos positivos han mejorado incluyendo en el entorno virtual objetos diferentes a los que se están clasificando.
- Los resultados positivos de los experimentos realizados con imágenes reales demuestran el potencial del uso de entornos virtuales para generar datos etiquetados y entrenar redes neuronales a partir de estos.

Trabajo futuro:

- Extender el entorno desarrollado, añadiendo más escenarios, objetos y clases. Generar más datos.
- Obtener un conjunto de imágenes reales etiquetadas, en el contexto de la cocina. Evaluar cuantitativamente cómo funciona el modelo con estas imágenes y comparar el rendimiento de redes entrenadas a partir de datos virtuales, con redes entrenadas a partir de datos reales.

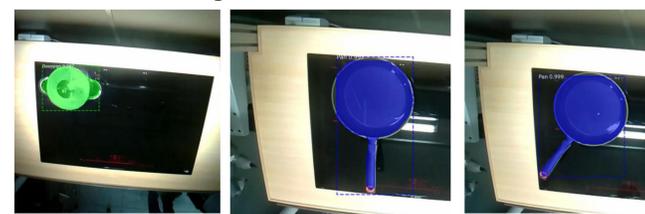
Imágenes reales



Videos de EPIC-KITCHENS Dataset



Imágenes reales: Plano cenital



	Plato	Sartén	Cazuela	Vaso	Cubierto	FP
Predicciones	345	0	1	1	0	2
	1	364	2	1	0	2
	0	0	366	1	0	9
	0	0	0	213	0	19
	0	4	0	0	325	18
	6	3	11	1	17	0
	Plato	Sartén	Cazuela	Vaso	Cubierto	FP

	Precisión	Recall
	96 %	97 %

Referencias

- [1] GAIDON, A., et al., 2016. VirtualWorlds as Proxy for Multi-object Tracking Analysis. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4340-4349. DOI 10.1109/CVPR.2016.470.
- [2] HE, K., et al., 2017. Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV). pp. 2980-2988. DOI 10.1109/ICCV.2017.322.
- [3] DAMEN, D., et al., 2022. Rescaling Egocentric Vision: Collection, Pipeline and Challenges for EPIC-KITCHENS-100. International Journal of Computer Vision, vol. 130, no. 1, ISSN 1573-1405. DOI 10.1007/s11263-021-01531-2.

Agradecimientos

Trabajo financiado por el Gobierno de Aragón, grupo T45_23R, y por proyectos CPP2021-008938, PID2021-124137OB-I00 y TED2021-130224B-I00, financiado por MCIN/AEI/10.13039/501100011033 y por la Unión Europea-NextGenerationEU/PRTR.