

Detección De Concentraciones de Agua en Caminos Usando YOLO

Water-puddle Detection on Roads Using YOLO

Trujillo Lopez, Miguel; Ramos Arreguin, Juan Manuel; Pedraza Ortega, Jesús Carlos; Tovar Arriaga, Saul; Aceves Fernández, Marco Antonio; Ramírez Arriaga, Karen Andrea.

Facultad de Ingeniería

Universidad Autónoma De Querétaro

Querétaro, México

miguetrulo@gmail.com; ORCID ([https://orcid.org/0009-0003-3807-4136])

juan.ramos@uaq.edu.mx; ORCID ([https://orcid.org/0000-0002-2604-9692])

caryoko@yahoo.com; ORCID ([https://orcid.org/0000-0001-5125-8907])

saul.tovar@uaq.mx; ORCID ([https://orcid.org/0000-0002-2695-1934])

marco.aceves@uaq.edu.mx; ORCID ([https://orcid.org/0000-0002-5455-0329])

andrea.r.arriaga@gmail.com ORCID ([https://orcid.org/0000-0003-4910-6298])

Resumen: La presencia de cúmulos de agua en caminos o carreteras puede ser una situación peligrosa para los automóviles, estos pueden ocultar de la vista del conductor ciertos obstáculos los cuales pueden causar un accidente. Tan solo en México el número de accidentes de tráfico aumenta considerablemente en temporada de lluvias. El desarrollo de tecnologías que permitan la detección de este tipo de fenómeno puede ayudar a la asistencia en la conducción. En este trabajo se presenta la implementación de un modelo de inteligencia artificial basado en el método YOLO el cual es capaz de detectar las concentraciones de agua con una alta precisión (78%) bajo diversas condiciones climatológicas, de luz, en distintos tipos de condiciones de camino, estructurado (carreteras o caminos pavimentado) y no estructurado (caminos sin pavimento) y en una topología de México (Querétaro-Guanajuato).

Palabras clave: Detección de cúmulos de agua; Visión por computadora; YOLO; Aprendizaje profundo; Detección de objetos.

Abstract: The presence of water-puddle on roads or highways can be a dangerous situation for cars. These can obscure certain obstacles from the driver's view, potentially causing an accident. In Mexico alone, the number of traffic accidents increases considerably during the rainy season. The development of technologies that allow the detection of this type of phenomenon can aid in driver assistance. This paper presents the implementation of an artificial intelligence model based on the YOLO method, which is capable of detecting water concentrations with high precision (78%) under various weather and light conditions, in different types of road conditions, structured (roads or paved roads) and unstructured (unpaved roads) and in a topology of Mexico (Querétaro-Guanajuato).

Keywords: Water-puddle detection; Computer Vision; YOLO; Deep Learning; Object detection.

1. Introducción

La conducción se ha convertido en una de las actividades principales y más frecuentes dentro de nuestra sociedad. Millones de personas viajan o conducen un vehículo automotor, por lo que el desarrollo de tecnologías que mejoren o den asistencia al conductor son de beneficio para sus usuarios. A través de los años los vehículos han sido construidos con un mayor número de herramientas que ayuden a la conducción y seguridad del mismo.

La lluvia es un evento climatológico el cual afecta muchas veces a varios sectores de la ciudad, de entre los cuales se encuentra la conducción y el tráfico. Las afectaciones causadas por la lluvia son diversas, estas

afectan la infraestructura de las calles cuando se llenan de agua, llegando así a provocar diversos percances a los conductores. Debido a que el pavimento mojado puede causar un mal funcionamiento del vehículo o provocar concentraciones de agua (charcos) las cuales ocultan a la vista del piloto posibles obstáculos como hoyos u objetos (ramas, topes,) que afecten al vehículo y sus tripulantes.

Solamente en México entre el año 2021 y 2022 los accidentes vehiculares en temporada de lluvias llegaron a aumentar hasta un 13%, mientras que en los últimos 5 años alrededor del 25% de los accidentes registrados se dan en condiciones de lluvia [1]. En Estados Unidos se reporta que alrededor de 3 billones de dólares se gasta de forma anual debido a daños causados por baches [2]. Manejar en medio de cúmulos de agua o zonas inundadas puede causar

daños al vehículo y sus tripulantes, esto debido a que puede ocurrir el fenómeno llamado hidropneumático, en el cual como resultado a la acumulación de agua en los caminos el neumático pierde contacto con la superficie (carretera) y rebote sobre el agua llegando a provocar la pérdida de control del automóvil, mientras más profunda sea el cumulo de agua, con mayor facilidad se puede perder el control. En la figura 1 se puede observar una representación de este fenómeno. Se reportó que el 21% de 5,891,000 accidentes de vehículos en Estados Unidos son relacionados con la climatología, el 75% de estos ocurren en pavimento mojado y 47% durante la lluvia [3].



Figura 1. Hidropneumático [4].

Lo que se busca es desarrollar un algoritmo el cual sea capaz de detectar concentraciones de agua en caminos con una topografía de México en imágenes, para una futura asistencia en la conducción autónoma de vehículos.

Una de las principales problemáticas que se tiene al momento de detectar cúmulos de agua es la naturaleza del charco, esto se debe a que no es sencillo definirlo. El comportamiento de éste según el caso puede ser como un espejo, reflejando los objetos que tiene alrededor (figura 2) o la luz que le incide generando la apariencia de que se ilumina el segmento de agua. Esto depende mucho del tamaño del charco, el tipo de luz que le incide y el ángulo en que se percibe el charco. Los cúmulos de agua tienden a ser áreas con un alto brillo y una baja saturación.



Figura 2. Reflexión del cúmulo de agua.

También es complejo determinar los límites de un charco, ya que tiende a tener zonas con una clara

presencia de agua mientras que en los bordes tiende a difuminarse, y no tener una apariencia tan clara, esto dependerá del tipo de suelo en el que se encuentre. En la figura 3 se muestra el mismo charco, pero se muestra que pueden abarcar diferentes áreas, aunque es el mismo objeto de análisis.

Para tratar el problema de la detección de cúmulos de agua se han desarrollado diversas metodologías, algunas tienen su desarrollo centrado en el uso de sensores y periféricos [5], [6], [7], mientras que otros tienen una metodología en la cual usan cámaras y algoritmos de inteligencia artificial [8], [9], [10], [11], [12], [13].

El desarrollo de sensores especializados para detectar acumulaciones de agua requiere del funcionamiento de diversos periféricos y hardware, lo que eleva el costo de desarrollo y producción. En cambio, las metodologías que constan de cámaras y algoritmos tienden a ser más económicas y fáciles de probar e implementar, debido a que en la actualidad los automóviles ya cuentan con diversos sensores y cámaras distribuidos en el automóvil.

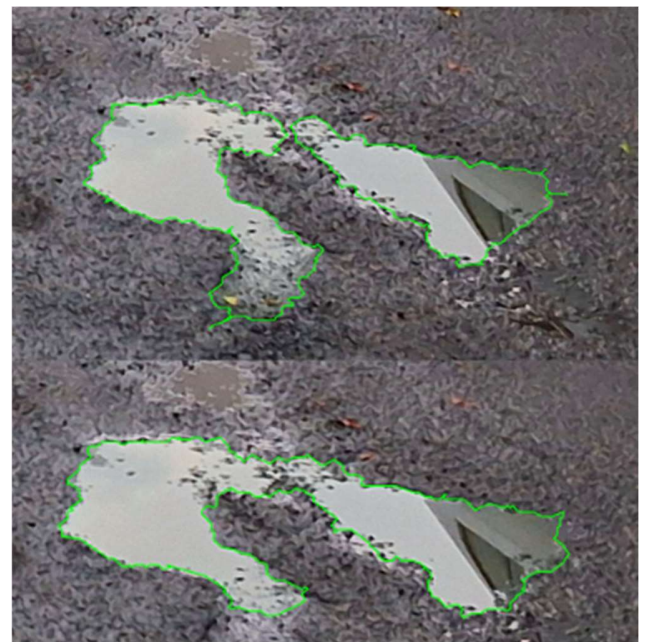


Figura 3. Variabilidad del área de un charco.

Se han desarrollado diversas técnicas de detección de objetos en el área de visión por computadora, los más destacados son basados en técnicas de Aprendizaje Profundo[14], YOLO (You Only Look Once)[15], SSD (Single Shot Detector) [16], Faster R-CNN (Faster Region-Based Convolutional Neural Network) [17]entre otros.

YOLO es un modelo enfocado en una respuesta rápida. Debido a esto, es ampliamente utilizado en

proyectos relacionados con la conducción autónoma. Su adaptabilidad y gran precisión en diversas condiciones climatológicas como días despejados, lluvia, atardecer entre otros, lo hacen adecuado para este tipo de problemáticas.[8], [18]

YOLO unificó los pasos de detección de objetos al detectar todos los cuadros delimitadores de forma simultánea. Para lograr esto YOLO divide la imagen de entrada en una cuadrícula $S \times S$ y predice B número de cajas delimitadoras de la misma clase, junto con la probabilidad de que cada recuadro pertenezca a una de las C clases posibles para cada celda de la cuadrícula, figura 4. [15]

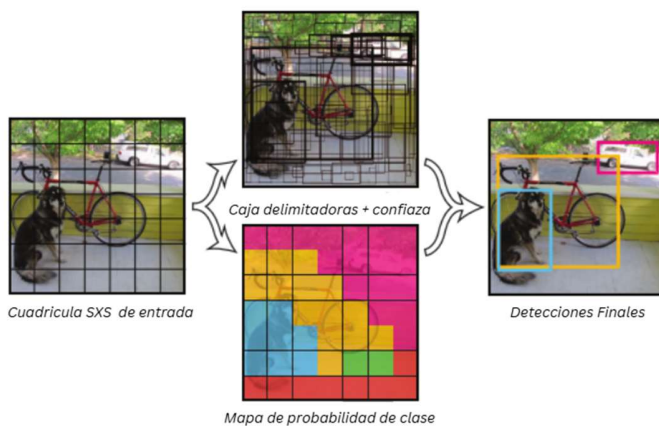


Figura 4. Funcionamiento de YOLO[15].

La arquitectura del modelo de YOLO se muestra en la figura 5, las primeras capas convolucionales extraen las características de la imagen mientras que las capas completamente conectadas predicen las probabilidades y las coordenadas. La red tiene 24 capas convolucionales seguidas de dos capas completamente conectadas, se usa una capa de reducción de 1×1 seguido de una capa convolucional de 3×3 . [15]

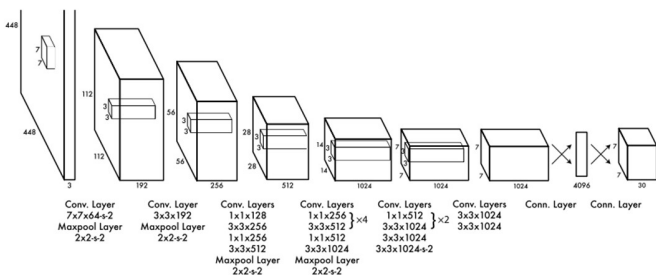


Figura 5. Arquitectura de YOLO [15].

Para poder cuantificar que tan bien un modelo realiza la tarea de detección, se utilizan métricas para

poder medir su desempeño, como la métrica mAP (Mean Average Precision). Para poder entender la métrica mAP es necesario conocer los siguientes conceptos.

La precisión es la habilidad del modelo de identificar objetos relevantes, y se da como la relación entre las detecciones correctas (Verdaderos positivos) y el total de predicciones positivas (Verdaderos Positivos, Falsos Positivos), figura 6.[19]

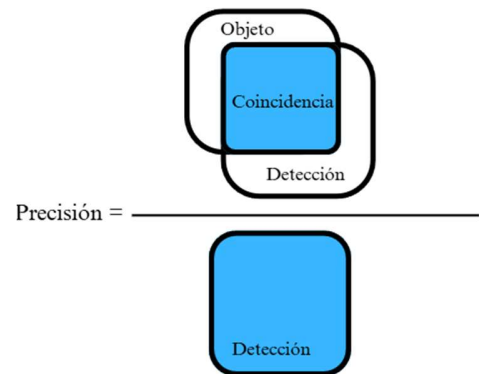


Figura 6. Precisión.

La recuperación es la capacidad del modelo de recuperar todos los casos relevantes, la relación entre las predicciones correctas (Verdaderos Positivos) y el total de las instancias (Verdaderos positivo, Falsos negativos), figura 7. [19]

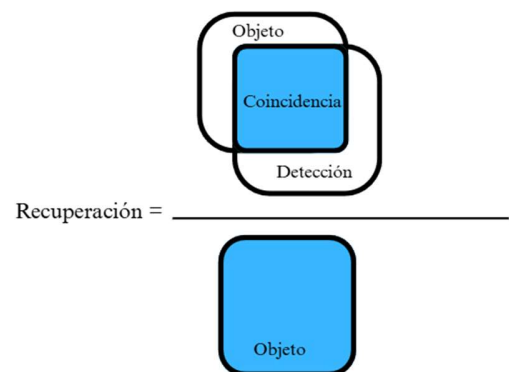


Figura 7. Recuperación

La Intersección sobre la unión (IoU), figura 8, es una métrica basada en el índice de Jaccard, que es un coeficiente de similitud entre dos conjuntos de datos (etiquetas/anotaciones). El IoU corresponde al área de superposición (intersección) entre el área predicha B_p y la anotación verdadera B_{pt} y dividido entre el área de su unión. Cuando el valor de $\text{IoU} = 1$, se dice considera que es una detección perfecta, mientras que es cuando $\text{IoU} = 0$ es una detección nula. Por lo cual cuanto más cercano se IoU a 1 la detección es mejor.[19]

La precisión media (mAP) junta tanto la precisión como la recuperación, para las diversas clases, dependiendo del tipo de evaluación mAP puede reportarse con varios tipos de umbrales los más comunes es mAP.50, el cual se calcula con un umbral de IoU de 0.5, en el cual se busca que el modelo detecte al menos el 50% del área del objeto, mientras que mAP50-90 es más estricto y únicamente muestra aquellas detecciones la cuales tienen una localización muy exacta [19].

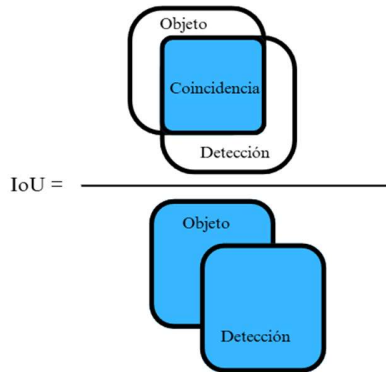


Figura 8. Intersección Sobre la Unión

2. Materiales y métodos

La metodología propuesta para dar solución a la problemática será la siguiente mostrada en la figura 9, posteriormente se describirá cada uno de los pasos.

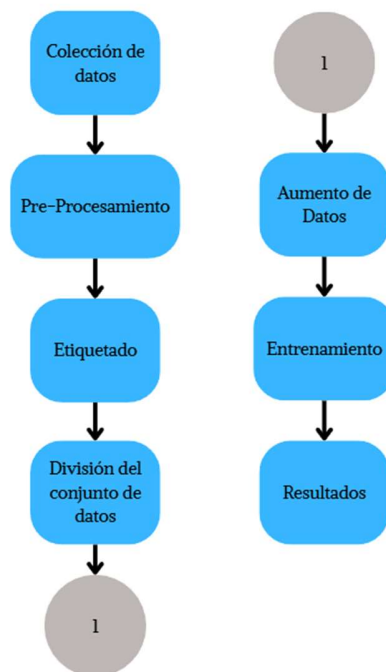


Figura 9. Metodología.

2.1 Colección de los datos

Para la construcción del conjunto de datos se buscó obtener muestras lo más cercanas a condiciones reales. Para esto, se montó una cámara en el tablero de un automóvil (figura 10) en movimiento y se capturó diversos segmentos de ruta, considerando diversas condiciones del clima y de luz. Las zonas en donde se realizó la captura de estas muestras fueron en la ciudad de Querétaro y Guanajuato. En la tabla 1 se muestran las características técnicas de la cámara usada para la adquisición de los datos.



Figura 10. Dispositivo de captura.

Características	Especificación
Sensor de imagen	1/2.9" CMOS, ~2MP pixeles efectivos
Resolución	(1280x720) px (Full HD, 16:9)
Apertura	f/2.0
Campo de visión	130 diagonal (110° horizontal)
Lentes	6 elementos de lente de vidrio
Rango de frames	30fps
Formato de video	AVI

Tabla 1. Especificaciones técnicas de la cámara.

2.2 Pre-procesamiento

Una vez que se capturaron los segmentos de ruta se extrajeron imágenes en el que aparecen cúmulos de agua (muestras positivas) y aquellas en las que hubiera ausencia de los cúmulos (muestras negativas). Se cuenta con imágenes en caminos estructurados (calles con pavimento) y en caminos no estructurados (calles sin pavimentar o en malas condiciones), de la misma forma hay imágenes que contienen cúmulos de agua y otras que

no contienen. En la figura 11 se muestran ejemplos de las diversas condiciones climatológicas en las cuales se capturaron las imágenes que servirán para entrenar el modelo.

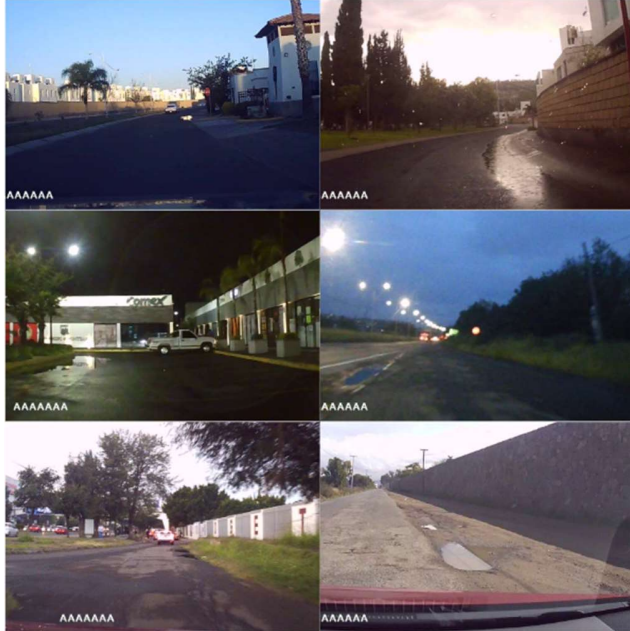


Figura 11. Condiciones de luz en el conjunto de datos.

2.2 Etiquetado

Para el etiquetado de las imágenes se utilizó la herramienta CVAT la cual permite hacer anotaciones y exportar las etiquetas en el formato requerido por el modelo. En la figura 12 se muestra cómo se realiza el etiquetado de las imágenes, se genera un cuadro delimitador en el cual se encierra el charco para obtener las coordenadas del lugar del charco.



Figura 12. Etiquetado de las imágenes.

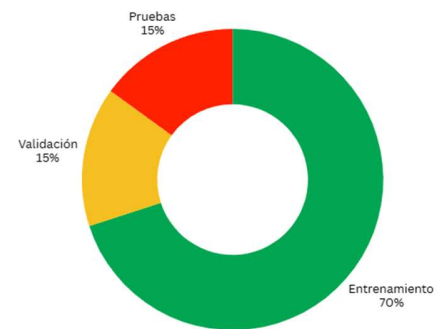


Figura 13. División de los Datos

2.3 Conjunto de datos

El conjunto de datos será dividido en los siguientes conjuntos: Entrenamiento, validación y prueba (figura 13). La división para estos será un 70:15:15 donde el 70% de las imágenes se utilizarán para entrenamiento, 15% para validación y el 15% restante para pruebas. El total de imágenes usadas es de 1112 para entrenamiento del modelo, 238 imágenes para la validación y 239 para realizar pruebas.

2.4 Aumento de datos

Para lograr que el modelo sea lo suficientemente robusto y pueda tener un buen desempeño en los diversos escenarios en los que se pueda encontrar, se usará la técnica de aumento de datos, la cual busca mitigar el sobreajuste. Se aumenta de forma artificial los datos

logrando que el modelo pueda aprender y generalizar de mejor forma debido a la diversidad de datos y así obtener una mejor precisión en las detecciones.

En la tabla 2 se resume el conjunto de técnicas usadas para el aumento de datos y en la figura 14 se muestra un ejemplo de cómo se ve el aumento de datos en una muestra, estas corresponden a la variabilidad de espacio de color (cambios de tono, saturación y brillo) y variabilidad geométrica (rotación, traslación, escalado).

Parámetro	Valor
hsv_h	0.1
hsv_s	0.7
hsv_v	0.4
Degrees	30
Translate	0.2
Scale	0.5
Shear	5.0
Fliplr	0.5

Tabla 2. Aumento de datos



Figura 14. Aumento de datos.

2.5 Entrenamiento

Para comenzar el entrenamiento del modelo es necesario ajustar ciertos valores de hiperparámetros los cuales ayudan a que el modelo minimice el error y tenga una mayor precisión en las detecciones. La tabla 3 muestra el conjunto de hiperparámetros usados para el entrenamiento.

Parámetro	Valor
Épocas	250
Optimizador	AdamW
Tasa de aprendizaje	0.0001
Momento	0.9
Lote	32
Tamaño de imagen	640

Tabla 3. Hiperparámetros usados para el entrenamiento.

3. Resultados

Los resultados obtenidos en las métricas se presentarán a continuación en la tabla 4. En la figura 15, 16 y 17 se muestra imágenes que forman parte del conjunto de pruebas donde el modelo realiza la detección de cúmulos de agua y se compara con diferentes versiones del modelo de YOLO. Se muestra como el modelo tiene un buen desempeño y es capaz de detectar la presencia de cúmulos de agua en diversas condiciones de luz y diversos tipos de caminos.

Modelo	Precisión	Recall	mAP@.50
YOLOV5	70.0	41.7	56.8
YOLOV8	78.9	45.1	64.0
YOLOV11	76.9	48.5	64.0

Tabla 4. Resultados de los modelos entrenados



Figura 15. Detección de YOLOV5



Figura 16. Detección de YOLOV8.



Figura 17. Detección de YOLOV11.

4. Discusión

La detección de cúmulos de agua es una tarea la cual es compleja debido a que el charco cambia dependiendo la luz que le incida y el ángulo en que se observa por lo cual es necesario tener un mayor número de muestras para poder abarcar la mayor cantidad de situaciones posibles, para así tener un modelo el cual pueda generalizar en el mayor número de escenario posibles. Si bien el modelo presentado tiene un buen rendimiento en diversas condiciones de luz, tiene ciertas fallas en las cuales no logra detectar ciertas acumulaciones de agua, en la figura 18 se muestra como el modelo falla al no detectar algunas de estas acumulaciones.



Figura 18. No detecciones.

De las tres versiones de YOLO la que obtuvo una mayor precisión fue la versión 8. Sin embargo, la que

tuvo menos detecciones fallidas, es decir, mayor recuperación, fue la versión 11, mientras que mAP tanto la versión 8 y 11 obtuvieron el mismo resultado. La versión 3 fue la que peor rendimiento obtuvo de los tres modelos. Esto se puede observar en la figura 19 donde tanto la versión 8 y 11 es capaz de detectar los dos cúmulos con una alta confianza figura 19 (a) y (b). Por el contrario, la versión 3 únicamente logra detectar uno con poca confianza figura 19 (c), en rojo la detección no lograda. También la versión 3 suele dividir los cúmulos en segmentos más pequeños figura 19 (f) mientras que las versiones superiores logran detectar los cúmulos grandes en una sola caja figura 19 (d) y (e).



Figura 19. Comparación entre versiones

En la página oficial de YOLO[20], se menciona que la versión 8 tiene, a lo largo del tiempo, un mayor soporte debido a la colaboración de terceros, en su implementación en entornos de producción. Por otro lado, El modelo Yolo V11 es más nuevo, pero tiene una menor integración de terceros en comparación a YOLO V8.

YOLOv8 supera a YOLOv5 debido a que la V8 introduce mejoras arquitectónicas, funcionales y de entrenamiento que incrementan la precisión (mAP) y la velocidad de inferencia, sin aumentar significativamente el tamaño del modelo.

5. Conclusión

La tarea de detectar cúmulos de agua en imágenes es compleja debido a la naturaleza del problema, la variabilidad de la luz y las diversas condiciones climáticas hacen que la detección en este tipo de tareas sea complicada. En este trabajo, el modelo implementado fue YOLO, en sus versiones 5, 8 y 11, para esta tarea. El mejor desempeño se tuvo con la

versión 8, considerando las diferentes condiciones de luz, de clima y de ambiente.

Se entrenó el modelo con una base de datos propia, para poder realizar el reconocimiento de las acumulaciones de agua usando la base de datos mencionada en el trabajo y con una topología característica de México.

Para el entrenamiento y pruebas realizadas, se usaron las tres versiones de YOLO, bajo las mismas condiciones. En general, todas tuvieron buen desempeño, pero con la versión 8 se obtuvo la mejor precisión del 78%, como se presenta en la tabla 3. Esta precisión es comparable con las reportadas por otros autores en el estado del arte. La diferencia principal es que las bases de datos usadas por ellos no son accesibles, por lo que no se pudo hacer pruebas con esas bases de datos, para tener una comparación confiable.

Finalmente, es posible la detección de cúmulos de agua en caminos con el uso de algoritmos de inteligencia artificial con un gran nivel de precisión. Si bien, aún el modelo presenta predicciones faltantes, es posible mejorar el desempeño del modelo, si se expande el tamaño y diversidad del conjunto de datos, añadiendo más muestras. Estas muestras deben ser obtenidas en diversos lapsos del día y bajo diversas condiciones climáticas.

Este trabajo abre la posibilidad de que en un futuro se realice la implementación de YOLO en vehículos autónomos terrestres con la ayuda de sistemas embebidos como NVIDIA Jetson o Raspberry Pi para poder evaluar la eficacia del modelo entrenado en tiempo real. Esto es gracias a que YOLO cuenta con versiones ligeras (lite) que permiten su implementación en este tipo de sistemas embebidos.

Referencias bibliográficas

- [1] EXPANSION, “Los siniestros de autos repuntan en México por las lluvias,” *EXPANSION*, México, Jul. 2023.
- [2] American Automobile Association, “Pothole Damage Costs U.S. Drivers \$3 Billion Annually,” <https://info.oregon.aaa.com/pothole-damage-costs-u-s-drivers-3-billion-annually/>.
- [3] Department of Transportation Federal Highway Administration, “How do Weather Events Impact Roads.”
- [4] Den, “Aquaplaning: qué es el hidropneumático, cuándo se produce el aquaplaning y qué hacer en ese caso,” [https://club.autodoc.es/magazin/aquaplaning-que-es-como-ocurre-y-como-evitarlo](https://club.autodoc.es/magazin/aquaplaning-que-es-como-ocurre-y-como-evitarlohttps://club.autodoc.es/magazin/aquaplaning-que-es-como-ocurre-y-como-evitarlo).
- [5] M. Lee, J. C. Kim, M. H. Cha, H. Lee, and S. Lee, “Identifying Puddles based on Intensity Measurement using LiDAR,” *Journal of Sensor Science and Technology*, vol. 32, no. 5, pp. 267–274, Sep. 2023, doi: 10.46670/JSST.2023.32.5.267.
- [6] N. Long, H. Yan, L. Wang, H. Li, and Q. Yang, “Unifying Obstacle Detection, Recognition, and Fusion Based on the Polarization Color Stereo Camera and LiDAR for the ADAS,” *Sensors*, vol. 22, no. 7, Apr. 2022, doi: 10.3390/s22072453.
- [7] H. Tahara, I. Ikegami, K. Takakura, T. Kato, and M. Nagata, “Puddle Detection for Avoidance Path Planning of Wheeled Mobile Robot Using Laser Reflection Intensity,” *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society*, vol. 1, pp. 699–704, Dec. 2019.
- [8] M. Jakubec, E. Lieskovská, B. Bučko, and K. Záborská, “Comparison of CNN-Based Models for Pothole Detection in Real-World Adverse Conditions: Overview and Evaluation,” *Applied Sciences (Switzerland)*, vol. 13, no. 9, May 2023, doi: 10.3390/app13095810.
- [9] J. J. Qiao, X. Wu, J. Y. He, W. Li, and Q. Peng, “SWNet: A Deep Learning Based Approach for Splashed Water Detection on Road,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 4, pp. 3012–3025, Apr. 2022, doi: 10.1109/TITS.2020.3029006.
- [10] A. Kumar and A. Choudhary, “Water-Puddle Segmentation Using Deep Learning in Unstructured Environments,” in *Proceedings of the 17th IEEE International Conference on Service Operations and Logistics, and Informatics, SOLI 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/SOLI60636.2023.10425657.
- [11] L. Juntao, N. Chuong, and Y. Shaodi, “Temporal 3D fully connected network for water-hazard detection,” *Digital Image Computing: Techniques and Applications (DICTA)*, pp. 1–5, 2019, doi: <https://ieeexplore.ieee.org/abstract/document/8945849>.
- [12] R. Zhang *et al.*, “ABCDWaveNet: Advancing Robust Road Ponding Detection in Fog through Dynamic Frequency-Spatial Synergy.”
- [13] M. S. C. Cruz, L. A. Ong, and A. N. Yumang, “Road Wetness Estimation Using Deep Learning Model,” *Engineering Proceedings*, vol. 92, no. 1, 2025, doi: 10.3390/engproc2025092051.
- [14] M. Maity, S. Banerjee, and S. Sinha Chaudhuri, “Faster R-CNN and YOLO based Vehicle detection: A Survey,” in *Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021*, Institute of Electrical

- and Electronics Engineers Inc., Apr. 2021, pp. 1442–1447. doi: 10.1109/ICCMC51019.2021.9418274.
- [15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Dec. 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.
- [16] W. Liu *et al.*, “SSD: Single Shot MultiBox Detector,” Dec. 2016, doi: 10.1007/978-3-319-46448-0_2.
- [17] R. Girshick, “Fast R-CNN.” [Online]. Available: <https://github.com/rbgirshick/>
- [18] Y. Zhao, B. Shi, X. Duan, W. Zhu, L. Ren, and C. Liao, “Research on road surface damage detection based on SEA-YOLO v8,” *PLoS One*, vol. 20, no. 6 JUNE, Jun. 2025, doi: 10.1371/journal.pone.0324439.
- [19] R. Padilla, S. L. Netto, and E. A. B. da Silva, “A Survey on Performance Metrics for Object-Detection Algorithms,” *International Conference on Systems, Signals and Image Processing (IWSSIP)*, pp. 237–242, 2020.
- [20] ultralytics, “YOLOv8 vs YOLO11: Una comparación técnica detallada,” <https://docs.ultralytics.com/es/compare/yolov8-vs-yolo11/#conclusion-and-recommendation>.