Segmentación de Superficies Viales Mediante Visión IA

Ana Cristina Medrano Lule[1], Gabriel Cruz Luna[1], Alison Guadalupe Paredes López[1], José Eleazar Peralta López[1]
[1]Instituto Tecnológico Superior de Salvatierra. at22110025@itess.edu.mx, cat2211812@itess.edu.mx, at22110361@itess.edu.mx, joperalta@itess.edu.mx.

Resumen

La detección automatizada de superficies viales a partir de secuencias de video representa un desafío en la visión por computadora debido a interferencias del entorno. En este trabajo se desarrolló un sistema de segmentación de imágenes combinando técnicas de procesamiento digital y redes neuronales deep learning para la identificación de la región vial en aplicaciones automotrices. La metodología inició con la creación de un dataset de 646 imágenes extraídas a partir de video, las cuales fueron normalizadas a 128 x 128 píxeles. Posteriormente, se aplicó un algoritmo de segmentación por umbralización de color BGR para aislar la calzada y generar las máscaras objetivo. Con estos datos se entrenó un modelo convolucional de codificador-decodificador (Encoder-Decoder) a lo largo de 60 épocas. Como resultado, la red neuronal alcanzó una precisión (Accuracy) superior al 95% y una función de pérdida reducida en validación, aislando eficazmente la superficie del camino y descartando elementos externos como vegetación y vehículos. Este flujo de trabajo establece una base sólida para sistemas de percepción visual en asistencia a la conducción y navegación autónoma

Este trabajo establece una base sólida para sistemas de percepción visual y navegación autónoma. Los resultados obtenidos demuestran que el procesamiento previo mediante la umbralización BGR optimiza significativamente la etapa de segmentación de imágenes, permitiendo la codificador-decodificador aprenda características viales de una manera eficiente a pesar de tamaño compacto de los fotogramas (128x128) pixeles. Asimismo, la estabilidad de la función de perdida a lo largo de las 60 épocas de entrenamiento confirma que la red neuronal no presenta sobreajuste, validando la precisión del modelo en la visión por computadora enfocada al área automotriz.

El hecho de que la función de perdida de convergiera sin oscilaciones durante las 60 épocas confirma estabilidad y una correcta inicialización de la red neuronal, este sistema no solo muestra un margen de error mínimo también proyecta una solución altamente viable para entornos con recursos limitados.

Esto abre una pauta a futuras implementaciones en sistemas de inteligencia artificial bajo el procesamiento de secuencias de video, que en tiempo real, mantenga los estándares internacionales de seguridad en vehículos de navegación autónoma. [3], [4].

Palabras clave: Segmentación; inteligencia artificial; superficies; visión; automotrices.

Abstract

Automated Road surface detection from video sequences poses a challenge in computer vision due to environmental interference. In this study, an image segmentation system was developed by combining digital processing techniques and deep learning neural networks to identify the road region for automotive applications. The methodology began with the creation of a dataset comprising 646 images extracted from video, which were normalized to 128 x 128 pixels. Subsequently, a BGR color-thresholding segmentation algorithm was applied to isolate the roadway and generate target masks. Using this data, an encoder-decoder convolutional model was trained over 60 epochs. As a result, the neural network achieved an accuracy exceeding 95% and a reduced validation loss, effectively isolating the road surface while filtering out external elements such as vegetation and vehicles. This workflow establishes a solid foundation for visual perception systems in driver assistance and autonomous navigation.

This work establishes a solid foundation for visual perception and autonomous navigation systems. The results demonstrate that preprocessing via BGR thresholding significantly optimizes the image segmentation stage,

enabling the encoding-decoding process to efficiently learn road features despite the compact frame size (128x128 pixels). Furthermore, the stability of the loss function across 60 training epochs confirms the absence of overfitting, validating the model's accuracy for computer vision applications in the automotive sector.

The fact that the loss function converged without oscillations over the 60 epochs confirms the stability and proper initialization of the neural network; this system not only demonstrates a minimal margin of error but also presents a highly viable solution for resource-constrained environments.

This paves the way for future implementations in artificial intelligence systems for real-time video sequence processing that uphold international safety standards for autonomous vehicles.

Keywords: Segmentación; artificial inteligencia; surfaces; visión; automotive.

Introducción

El desarrollo de asistencia al conductor (ADAS) exige una ldesarrollo de la conducción autónoma y los sistemas percepción precisa del entorno vial [1], [2]. Sin embargo, la detección de la superficie del camino representa un desafío debido a elementos del entorno como vegetación, vehículos y variaciones de iluminación que interfieren con el análisis de la escena [1].

Dentro de este contexto, la percepción visual representa un punto crítico para interpretar en tiempo real escenarios viales complejos sujetos a condiciones variables de luz y sombras [1], [2]. Para abordar esta problemática, este trabajo presenta un sistema de segmentación semántica que combina procesamiento de imágenes y aprendizaje profundo (Deep Learning) [3], [5]. La metodología contempla la construcción de una base de datos de 646 fotogramas extraídos de video [2], la generación de máscaras objetivo mediante umbralización en el espacio BGR [2] y el entrenamiento de una red neuronal convolucional de tipo Encoder-Decoder a lo largo de 60 épocas [3], [4].

Con un desempeño superior al 95% de precisión (Accuracy), el enfoque propuesto aísla eficazmente la región de la calzada [2], sentando las bases para el desarrollo de sistemas de percepción visual orientados a la navegación autónoma y la prevención de obstáculos [1]–[5].

Metodología Propuesta

A) Extracción de Fotogramas y Creación del Dataset

A partir de un recorrido en zonas aledañas a la comunidad de Puerta del Monte se obtuvieron un total de 645 frames extraídas de un archivo de video. MP4 con una resolución de (478 x 850), las cuales se escalaron a 128 x 128 para procesar la información más compacta.

B) Preprocesamiento y Segmentación por Color

Con el fin de obtener las etiquetas u objetivos (ground truth) necesarios para la supervisión de la red neuronal, se aplicó una técnica de segmentación basada en la selección de rangos de color en el espacio BGR (Blue, Green, Red). Se definieron rangos de color específicos para aislar las tonalidades características del asfalto de la carretera:

Rango mínimo BGR: [66, 66, 72]

Rango máximo BGR: [106, 106, 112]

Mediante cada imagen de entrada se convirtió en una máscara binaria donde la región vial conserva valores positivos y las áreas externas (vegetación, vehículos, cielo) son filtradas. Las imágenes resultantes se almacenaron en un directorio de salida específico.



Figura 1. Comparativa entre la captura original RGB y la máscara de segmentación binarizada de la carretera.
Nota. Comparación directa de la toma original en carretera (izquierda) frente a la máscara binarizada generada para la detección de la superficie vial (derecha).

C) Arquitectura y Entrenamiento de la Red Neuronal

La etapa final consistió en la implementación de una arquitectura de aprendizaje profundo con TensorFlow/Keras orientada a la segmentación por píxel.



Figura 2. Arquitectura de la red neuronal convolucional para la segmentación de la superficie vial.
Nota. Arquitectura de red neuronal para segmentación vial mediante imágenes RGB de 128x 128 píxeles y entrenamiento a 60 épocas.

D) Preparación y Normalización de Datos

Ajuste de Dimensión: Las imágenes originales de la carretera (entradas) y sus respectivas máscaras segmentadas (etiquetas) se reescalaron a un tamaño uniforme de 128 x 128 píxeles.

División del Dataset: Se dividieron los 646 fotogramas en un conjunto de entrenamiento (580 muestras) y un conjunto de validación/prueba (146 muestras).

Escalamiento de Píxeles: Se dividieron los valores numéricos entre 255.0 para normalizar la entrada flotante al rango.

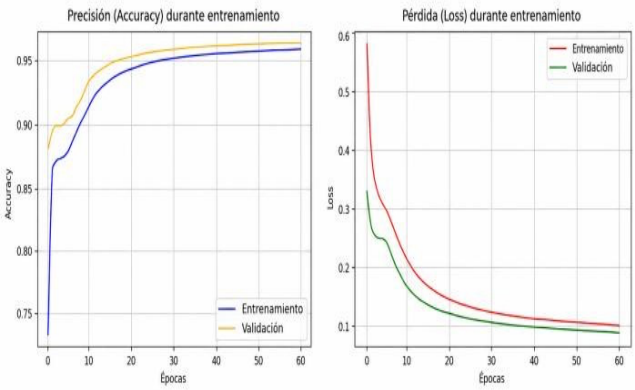


Figura 3. Curvas de aprendizaje de precisión (accuracy) y pérdida (loss) durante el entrenamiento y validación del modelo.
Nota. Curvas de rendimiento del modelo a lo largo de 60 épocas. Muestra la evolución de la precisión y la reducción de la función de pérdida para los conjuntos de entrenamiento y validación.

Resultados

Para evaluar el desempeño de la red neuronal convolucional en la segmentación vial, se monitorearon las métricas de precisión (Accuracy) y pérdida (Loss) durante el entrenamiento. Los resultados muestran una rápida convergencia inicial, alcanzando una precisión final superior al 95% y una función de pérdida reducida por debajo de 0.1 en el conjunto de validación. La trayectoria paralela y cercana entre las curvas de entrenamiento y validación evidencia una excelente capacidad de generalización del modelo, descartando la presencia de sobreajuste (overfitting) y confirmando la efectividad de la arquitectura propuesta para aislar la superficie de la calzada.

Al probar la red neuronal con una imagen nueva, el sistema nos devuelve una máscara en blanco y negro. La parte blanca muestra que la red logra reconocer y separar con éxito la carretera, mientras que la parte negra elimina la vegetación y el entorno que no nos interesa. Aunque quedan algunos puntos blancos pequeños por la resolución de la imagen, el resultado cumple con marcar correctamente la forma del camino.



Figura 4. Resultado de la máscara de segmentación binarizada obtenida por el modelo.
Nota. Máscara binaria resultante del procesamiento de la red neuronal para la identificación del área delimitada de la carretera.

Referencias

Geiger, A., Lenz, P., Stiller, C., y Urtasun, R. (2013). Vision meets robotics: The KITTI dataset. *Int. J. Robot. Res.* 32(11), 1231–1237. https://www.cvlibs.net/publications/Geiger2013IJRR.pdf

Neuhold, G., Ollmann, T,. Rota Bulò, S., y Kontschieder, P. (2017) The Mapillary Vistas dataset for semantic understanding of street scenes. roc. I*EEE Int. Conf. Comput. Vis. (ICCV),* 4990–4999. https://openaccess.thecvf.com/content_ICCV_2017/papers/Neuhold_The_Mapillary_Vistas_ICCV_2017_paper.pdf

Goodfellow, I., Bengio, Y., y Courville, A. (2016). *Deep Learning*. Cambridge, MA, EE. UU.: MIT Press.

Haykin, S. (2009). *Neural Networks and Learning Machines (3era eds.).* Pearson.

LeCun, Y,. Bengio, Y,. y Hinton, G. (2015). Deep learning. Nature, 521, 7553, 436–444. https://doi.org/10.1038/nature14539