

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
SANTIAGO - CHILE



**“TRATAMIENTO DE SOMBRAS PARA DETECTAR
DETERIOROS EN PAVIMENTOS CON MACHINE
LEARNING”**

FRANCISCA JAVIERA GONZÁLEZ BASTIAS

**MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN INFORMÁTICA**

Profesor Co-Guía: Alelí Osorio Lird
Profesor Co-Referente: José Luis Martí Lara

Agosto - 2025



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒ Memoria o trabajo de título; ☐ Tesis de Postgrado;

Título del trabajo: TRATAMIENTO DE SOMBRAS PARA DETECTAR DETERIOROS EN PAVIMENTOS CON MACHINE LEARNING

Nombre del candidato(a): Francisca Javiera González Bastias

Carrera / Grado: Ingeniería Civil Informática

Campus: Santiago San Joaquín ; **Departamento:** Informática

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Aleli Osorio Lird, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL

☒ El trabajo **NO contiene información que amerite confidencialidad** y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (embargo) por:

☐ 6 meses; ☐ 12 meses; ☐ 2 años; ☐ 3 años; ☐ 5 años; ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 26/09/2025

; Firma:

Estudiante o Candidato(a):

Fecha: 26/09/2025

; Firma:

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

DEDICATORIA

Considerando la importancia de este trabajo para los alumnos, este apartado es para que el autor entregue palabras personales para dedicar este documento. La extensión puede ser de máximo una hoja y se deben mantener este formato, tipo y tamaño de letra.

AGRADECIMIENTOS

Considerando la importancia de este trabajo para los alumnos, este apartado se podrá incluir en el caso de que el autor desee agradecer a las personas que facilitaron alguna ayuda relevante en su trabajo para la realización de este documento. La extensión puede ser de máximo una hoja y se deben mantener este formato, tipo y tamaño de letra.

RESUMEN

Resumen—En una metodología para la evaluación de deterioros en imágenes de alta resolución de pavimentos urbanos se busca corregir errores de detección causados por la presencia de sombras, para ello se compararon seis métodos de eliminación de sombras aplicados a tres *datasets* representativos de distintos pasos de la etapa de preprocesamiento. A continuación, las imágenes resultantes se procesaron con YOLOv5, el algoritmo detector de deterioros, para evaluar si hay mejoras en su precisión. De los dos métodos de procesamiento de imágenes y los cuatro basados en *Deep Learning*, se comprueba que *DC-ShadowNet*, un modelo basado en aprendizaje no supervisado, al aplicarse en las imágenes de pavimentos antes de la etapa de normalización, evitó que YOLOv5 confundiera las sombras con deterioros.

Palabras Clave—Eliminación de Sombras; Pavimentos Urbanos; Imágenes 2D de Alta Resolución; Procesamiento de Imágenes; *Deep Learning*

ABSTRACT

Abstract—A methodology for assessing deterioration in high-resolution images of urban pavements seeks to correct detection errors caused by the presence of shadows. To this end, six shadow removal methods were compared, applied to three representative datasets from different steps of the preprocessing stage. The resulting images were then processed with YOLOv5, the deterioration detection algorithm, to assess whether there were improvements in accuracy. Of the two image processing methods and the four based on Deep Learning, it was found that *DC-ShadowNet*, a model based on unsupervised learning, when applied to pavement images before the normalization stage, prevented YOLOv5 from mistaking shadows for deterioration.

Keywords—*Shadow Removal; Urban Pavements; High-resolution 2D Images; Image Processing; Deep Learning*

INDICE DE CONTENIDOS

RESUMEN.....	4
ABSTRACT	4
INDICE DE FIGURAS.....	8
INTRODUCCIÓN	10
CAPÍTULO 1: DEFINICIÓN DEL PROBLEMA	12
1.1 GESTIÓN DE INFRAESTRUCTURA VIAL	12
1.2 AUSCULTACIÓN DE PAVIMENTOS.....	12
1.2.1 CICLO DE VIDA DE PAVIMENTOS URBANOS	12
1.2.2 AUSCULTACIÓN MANUAL.....	14
1.2.3 AUSCULTACIÓN AUTOMÁTICA	14
1.3 PROYECTO FONDECYT 11201150.....	15
1.4 ADQUISICIÓN Y PROCESAMIENTO DE IMÁGENES	16
1.4.1 SISTEMA DE ADQUISICIÓN DE IMÁGENES.....	16
1.4.2 PROCESAMIENTO DE IMÁGENES.....	16
1.4.3 MEJORAMIENTO DE NITIDEZ	17
1.4.4 MEJORAMIENTO DE CONTRASTE	18
1.5 ENTRENAMIENTO Y EVALUACIÓN.....	19
1.5.1 DEFINICIÓN DE CONJUNTOS Y CLASES	19
1.5.2 ANÁLISIS DE LOS DATASETS.....	20
1.5.3 ANÁLISIS CON MÉTRICAS DE DESEMPEÑO	21
1.6 MEJORAS Y RECOMENDACIONES.....	22
1.7 ÁRBOL.....	23

1.8 OBJETIVO GENERAL	23
1.9 OBJETIVOS ESPECÍFICOS	23
CAPÍTULO 2: MARCO CONCEPTUAL	24
2.1 TIPOS DE SOMBRAS	24
2.2 SOMBRAS EN IMÁGENES DE PAVIMENTOS URBANOS	25
2.3 MÉTODOS ESTUDIADOS	26
2.3.1 BASADOS EN PROCESAMIENTO DE IMÁGENES	26
2.3.2 BASADOS EN DEEP LEARNING	34
CAPÍTULO 3: PROPUESTA DE SOLUCION	50
3.1 MÉTODOS PROBADOS	50
3.2 DATASETS DE PRUEBA	50
3.3 DETALLES DE IMPLEMENTACIÓN	51
CAPÍTULO 4: VALIDACIÓN DE LA SOLUCIÓN	53
4.1 RESULTADOS DE TRATAMIENTO DE SOMBRA	53
4.1.1 RESULTADOS DE DATASET 1	53
4.1.2 RESULTADOS DE DATASET 2	57
4.1.3 RESULTADOS DE DATASET 3	60
4.2 RESULTADOS EN YOLOV5	62
4.2.1 RESULTADOS DE DATASET 1	63
4.2.2 RESULTADOS DE DATASET 2	63
4.2.3 RESULTADOS DE DATASET 3	64
CAPÍTULO 5: CONCLUSIONES	66
5.1 ALCANCES Y LIMITACIONES DE LA PROPUESTA	66
5.2 RESULTADOS	66

5.3 VALIDACIÓN DE OBJETIVOS	67
5.4 CONTRIBUCIONES DEL TRABAJO	67
5.5 RECOMENDACIONES	67
REFERENCIAS BIBLIOGRÁFICAS	68

INDICE DE FIGURAS

Figura 1: Progresión del pavimento en el tiempo. Fuente: Schliessler y Bull (1992).....	13
Figura 2: Mantenimiento del pavimento. Fuente: Peshkin et al. (2007).	13
Figura 3: Auscultación manual de pavimentos. Fuente: Castro (2024).	14
Figura 4: Tecnología Laser Crack Measurement System. Fuente: Laurent et al. (2012).....	15
Figura 5: Equipo para adquisición de imágenes. Fuente: Venegas (2022).....	16
Figura 6: Imagen extraída del video. Fuente: Almendras (2023).	17
Figura 7: Imagen en sección de 10 [m] (volteada). Fuente: Almendras (2023).	17
Figura 8: Imágenes de pavimento binarizadas. Fuente: Almendras (2023).	18
Figura 9: Árbol del problema. Fuente: Elaboración Propia.	23
Figura 10: (a) Sombra según superficie en que se genera. (b) Sombra según fuente de luz y cercanía de objeto a superficie. Fuente: Guo et al. (2024).	24
Figura 11: Diagrama resumen de la metodología para detectar deterioros. Fuente: Elaboración Propia.....	25
Figura 12: Imagen de pavimento con sombras y grietas. Fuente: Elaboración propia.....	25
Figura 13: Imagen de pavimento con perspectiva corregida. Fuente: Elaboración propia.	26
Figura 14: Clasificación de filtros. Fuente: Bisht et al. (2018).	27
Figura 15: (a) Imagen ruidosa. (b) Histograma de los tonos de gris de la imagen. (c) Imagen segmentada usando <i>global thresholding</i> . Fuente: Gonzalez y Woods (1977).....	29
Figura 16: Esquema de matriz de confusión. Fuente: Polamuri.	33
Figura 17: Arquitectura de DC-ShadowNet. Fuente: Jin et al. (2021).	34
Figura 18: Arquitectura de LG-ShadowNet. Fuente: Liu et al. (2021).	38
Figura 19: Arquitectura de DMTN. Fuente: Fan et al. (2023).	41
Figura 20: Arquitectura de TBRNet. Fuente: Fan et al. (2023).	44
Figura 21: Módulo TBR. Fuente: Fan et al. (2023).	45
Figura 22: Ejemplo de <i>dataset</i> ISTD: (a): imagen con sombra, (b): máscara binaria y (c): imagen sin sombra. Fuente: Elaboración Propia.	46
Figura 23: Ejemplo de <i>dataset</i> ISTD+: (a): imagen con sombra, (b): máscara binaria y (c): imagen sin sombra con iluminación ajustada. Fuente: Elaboración Propia.	47
Figura 24: Ejemplo de <i>dataset</i> SRD: (a): imagen con sombra y (b): imagen sin sombra. Fuente: Elaboración Propia.	47
Figura 25: Ejemplos de <i>dataset</i> 1: (a): sólo sombras, (b): grieta parcialmente oculta y (c): grieta completamente oculta. Fuente: Elaboración Propia.	51
Figura 26: Ejemplos de <i>dataset</i> 2: (a): sólo sombras, (b): grieta parcialmente oculta y (c): grieta completamente oculta. Fuente: Elaboración Propia.	51
Figura 27: Resultados en <i>dataset</i> 1. Fuente: Elaboración Propia.	54
Figura 28: Resultados en <i>dataset</i> 2. Fuente: Elaboración Propia.	58
Figura 29: Resultados en <i>dataset</i> 3. Fuente: Elaboración Propia.	61
Figura 30: Resultados de YOLOv5 con el <i>dataset</i> 1. Fuente: Elaboración Propia.	63
Figura 31: Resultados de YOLOv5 con el <i>dataset</i> 2. Fuente: Elaboración Propia.	64
Figura 32: Resultados de YOLOv5 con <i>dataset</i> 3. Fuente: Elaboración Propia.	65

INDICE DE TABLAS

Tabla 1: Clases usadas para el entrenamiento de los algoritmos de Deep Learning. Fuente: Castro (2024).	20
Tabla 2: Cantidad de imágenes usadas para la implementación de los algoritmos. Fuente: Castro (2024).	21
Tabla 3: Métricas de desempeño para los <i>dataset</i> en el algoritmo YOLOv5. Fuente: Castro (2024).	21
Tabla 4: Métricas de desempeño para los <i>dataset</i> en el algoritmo YOLOv7. Fuente: Castro (2024).	22
Tabla 5: Resultados cuantitativos en las pruebas del <i>dataset</i> ISTD. Fuente: Elaboración Propia.	48
Tabla 6: Resultados cuantitativos en las pruebas del <i>dataset</i> ISTD+. Fuente: Elaboración Propia.	48
Tabla 7: Resultados cuantitativos en las pruebas del <i>dataset</i> SRD. Fuente: Elaboración Propia.	48
Tabla 8: Resumen de los métodos estudiados. Fuente: Elaboración Propia.	49
Tabla 9: Rendimiento de cada método probado en <i>dataset</i> 1. Fuente: Elaboración Propia.	57
Tabla 10: Rendimiento de cada método probado en <i>dataset</i> 2. Fuente: Elaboración Propia.	60
Tabla 11: Rendimiento de cada método probado en <i>dataset</i> 3. Fuente: Elaboración Propia.	62

INTRODUCCIÓN

La presente memoria aborda el desafío de mejorar la detección automática de deterioros en pavimentos urbanos mediante el tratamiento de sombras en imágenes de alta resolución. En los últimos años, el uso de algoritmos de visión por computador como YOLOv5 ha permitido identificar grietas y fisuras con gran rapidez; sin embargo, las variaciones de iluminación y la presencia de sombras continúan generando falsos positivos que comprometen la fiabilidad de los sistemas de inspección. Para dar solución a este problema, se propone integrar un módulo de eliminación de sombras basado en Deep Learning—específicamente DC-ShadowNet—dentro del flujo de preprocesamiento de imágenes.

La metodología seguida en este trabajo incluye la generación de tres conjuntos de datos representativos de distintas fases de adquisición y filtrado (sin tratamiento óptico, con filtro ND32 y en escala de grises), la aplicación de seis métodos de eliminación de sombras, dos de procesamiento de imágenes y cuatro de *Deep Learning*, y la evaluación de su impacto sobre la precisión de YOLOv5. Cada etapa se diseñó para garantizar la reproducibilidad: desde la captura y filtrado de imágenes hasta el preprocesamiento automatizado y la inferencia del detector. Asimismo, se realizó un análisis cualitativo y cuantitativo de resultados para identificar cuál de los métodos proporciona un equilibrio óptimo entre eliminación de sombras y conservación de detalles.

La estructura de la memoria se organiza en cinco capítulos. En el Capítulo 1 se describe la arquitectura de la metodología original y se definen los requisitos de datos necesarios para el proyecto. El Capítulo 2 introduce los conceptos fundamentales y revisa el estado del arte en detección de deterioros y métodos de eliminación de sombras. En el Capítulo 3 se detalla el diseño experimental: se explican las pruebas realizadas, los requisitos de hardware y las condiciones de ejecución, así como los tres conjuntos de datos empleados (sin filtro óptico, con filtro ND32 y en escala de grises) y el tipo de análisis aplicado a los resultados de eliminación de sombras. El Capítulo 4 presenta los resultados de los seis métodos de eliminación de sombras y muestra cómo estas imágenes preprocesadas se someten posteriormente a la inferencia con YOLOv5 para medir las mejoras en precisión. Finalmente, el Capítulo 5 discute los hallazgos, valida los objetivos, expone las contribuciones y propone recomendaciones para futuras investigaciones.

La propuesta desarrollada ha demostrado que es posible integrar de forma eficiente un módulo de eliminación de sombras en el flujo de procesamiento de imágenes de pavimentos urbanos de alta resolución, lo que permite a YOLOv5 distinguir con mayor fiabilidad las grietas y fisuras reales de las simples variaciones de iluminación. Gracias a esta

incorporación, el sistema logra ignorar las sombras sin confundirlas con deterioros, mejorando de manera notable la precisión de la detección. Aunque los métodos tradicionales basados en reglas fijas eliminan sombras y conservan la resolución original, suelen borrar o atenuar detalles críticos, como grietas finas, y provocar cambios bruscos en el color y la textura del pavimento. Por su parte, los enfoques de *Deep Learning* ofrecen una mayor capacidad de adaptación al aprender directamente de los datos, atenuando las sombras mientras preservan los rasgos ocultos, aunque a costa de una ligera variación en la resolución que puede corregirse mediante un entrenamiento específico.

En las pruebas realizadas, *DC-ShadowNet* preentrenado en SRD se reveló como la solución más eficaz: aplicado antes de la normalización de las imágenes, evitó que YOLOv5 confundiera las sombras con daños en la superficie y mantuvo intacta la textura del pavimento. Este resultado valida el objetivo general de diseñar un algoritmo de corrección de sombras y confirma la idoneidad de la metodología para elaborar conjuntos de datos especializados y comparar múltiples modelos de Deep Learning. Asimismo, aporta evidencia empírica de la integración exitosa de este módulo en el pipeline de detección de deterioros y sienta las bases para futuras mejoras.

Para maximizar los beneficios del método, resulta esencial capturar las imágenes con filtro ND32—a pesar de reducir ligeramente el contraste—ya que este paso facilita que el detector ignore las sombras. Además, entrenar *DC-ShadowNet* con datos propios del proyecto incrementará su robustez y permitirá adaptar el modelo a las condiciones específicas de cada zona urbana, sin necesidad de contar con pares de imágenes con y sin sombra coincidentes. Finalmente, la aplicación del módulo justo después de extraer los fotogramas y antes de cualquier normalización asegura que se preserve la fidelidad del espacio RGB, garantizando un preprocesamiento óptimo y una detección de deterioros más precisa.

CAPÍTULO 1: DEFINICIÓN DEL PROBLEMA

La memoria se plantea en el desarrollo de las recomendaciones formuladas tras la finalización del proyecto FONDECYT 11201150 "Evaluación y modelación del comportamiento de pavimentos urbanos usando machine learning" (Osorio, 2020), el cual tiene por objetivo la evaluación de deterioros en pavimentos urbanos mediante el uso de inteligencia artificial y equipos de bajo costo, junto con la modelación del desempeño y condición del pavimento urbano a través del tiempo. Dichas recomendaciones se basan en investigar y analizar el uso de algoritmos de inteligencia artificial para reducir el contraste luz y sombra de imágenes 2D de pavimentos urbanos preprocesadas, e investigar sobre nuevos algoritmos de inteligencia artificial para mejorar el desempeño de la detección automática de deterioros.

1.1 GESTIÓN DE INFRAESTRUCTURA VIAL

La infraestructura vial es fundamental para el desarrollo económico y social de las ciudades, permitiendo el transporte de personas, bienes y vehículos, promoviendo la comunicación, el comercio y el acceso a servicios. La gestión de infraestructura vial involucra la planificación, diseño, construcción, mantenimiento y operación de la red vial, tomando en cuenta aspectos técnicos, económicos, ambientales y sociales (Castro, 2024).

En Chile, desde el año 2006 mediante una modificación de la Ley N° 8.946 (actual 21.111) de Pavimentación Comunal, los Gobiernos Regionales, son los encargados del financiamiento de obras de pavimentación, reposición y administración, a excepción de la comuna de Santiago (Biblioteca del Congreso Nacional, s. f.). Sin embargo, son los municipios los encargados de elevar las necesidades de conservación, lo cual implica inversión de recursos en evaluación de la condición de pavimentos y herramientas para el diagnóstico de las necesidades.

1.2 AUSCULTACIÓN DE PAVIMENTOS

La auscultación de pavimentos es una técnica para evaluar el estado actual de la red vial mediante la recolección y análisis de datos generados por el tránsito. Esta técnica permite mantener la seguridad vial, analizar e identificar fallas para realizar mantenimiento, optimizar la asignación de recursos, planificar proyectos de construcción y reducir el impacto ambiental al prolongar la vida útil de la infraestructura.

A continuación, se explica cuál es el ciclo de vida de los pavimentos urbanos y qué métodos existen en la actualidad para llevar a cabo la auscultación.

1.2.1 CICLO DE VIDA DE PAVIMENTOS URBANOS

Los pavimentos urbanos son parte crucial de las infraestructuras en las ciudades, permitiendo la movilidad de vehículos y peatones. Sin embargo, debido a la constante exposición a las cargas del tráfico y los efectos climáticos, estos pavimentos se deterioran con el tiempo. Como se ve en la Figura 1, el deterioro de los pavimentos en los primeros años es más lento, pero al avanzar los años de vida útil del pavimento se puede observar un

TRATAMIENTO DE SOMBRAS PARA DETECTAR DETERIOROS EN PAVIMENTOS CON MACHINE LEARNING

umbral donde el deterioro es más acelerado, generando un término más rápido de la vida útil. (Almendras, 2023).

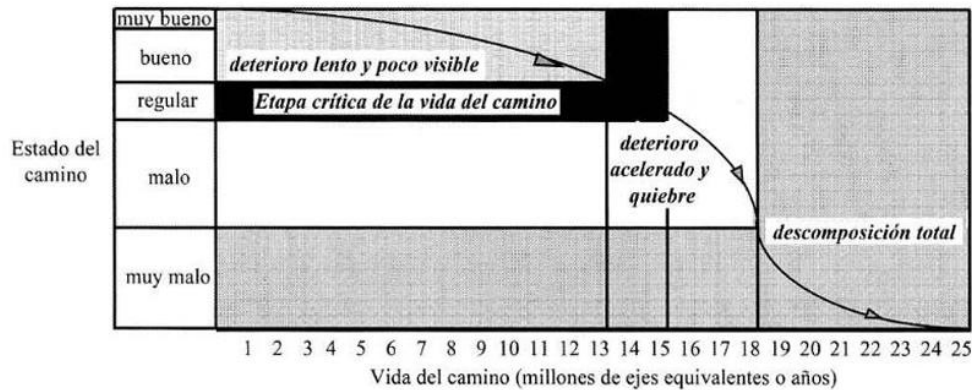


Figura 1: Progresión del pavimento en el tiempo.
Fuente: Schliessler y Bull (1992).

La detección temprana de los deterioros en los pavimentos es vital para mantener la seguridad y serviciabilidad de la red vial, como para planificar y ejecutar acciones de mantenimiento.

El mantenimiento de los pavimentos urbanos se divide en dos categorías; preventivo y correctivo. La Figura 2 muestra ambas categorías y su aplicación depende de la fase en la que se encuentra el pavimento según su ciclo de vida.

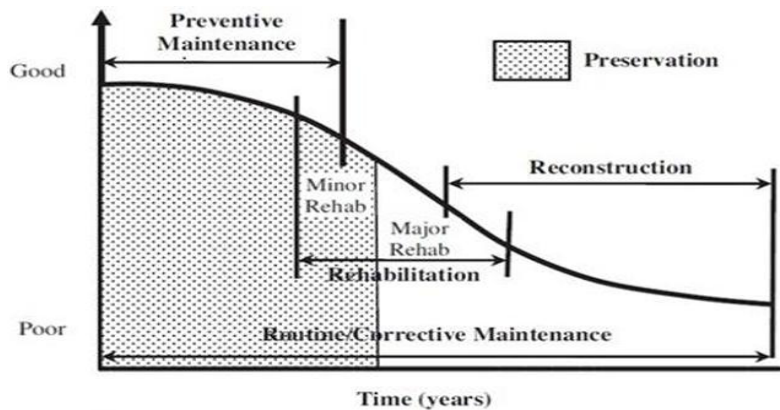


Figura 2: Mantenimiento del pavimento.
Fuente: Peshkin et al. (2007).

El primero consiste en inspecciones periódicas para detectar cualquier signo de deterioro y cuando existe algún problema, realizar trabajos para prevenir el agravamiento del pavimento. El mantenimiento correctivo en cambio se emplea cuando el pavimento presenta un deterioro demasiado grave que requiere una reparación más exhaustiva. El objetivo es priorizar un mantenimiento preventivo, dado que conlleva un menor costo y tiempo de trabajo, lo que también implica interferir menos con el normal funcionamiento de las rutas.

Para detectar los deterioros en los pavimentos se pueden aplicar dos métodos; la auscultación manual y la auscultación automática.

1.2.2 AUSCULTACIÓN MANUAL

La auscultación manual consiste en una evaluación visual llevada a cabo por inspectores, quienes toman mediciones y registro de los deterioros. En el procedimiento simplificado se realizan muestreos estadísticos a intervalos de 25 metros, de los cuales se examinan 6 metros de longitud, extendiéndose a lo ancho de la calzada.

Sin embargo, este método presenta aspectos desfavorables:

- Representatividad del Porcentaje de Auscultación:** el grado de representatividad del método es insuficiente al tomar secciones limitadas de muestreo.
- Consumo de tiempo:** el proceso requiere de un equipo de profesionales y de la interrupción del flujo vehicular en la zona a evaluar.
- Subjetividad:** la inspección visual del pavimento acarrea un grado de subjetividad por parte de los inspectores, ya que puede haber variaciones según la percepción de quien observa, lo que puede entregar resultados inconsistentes.

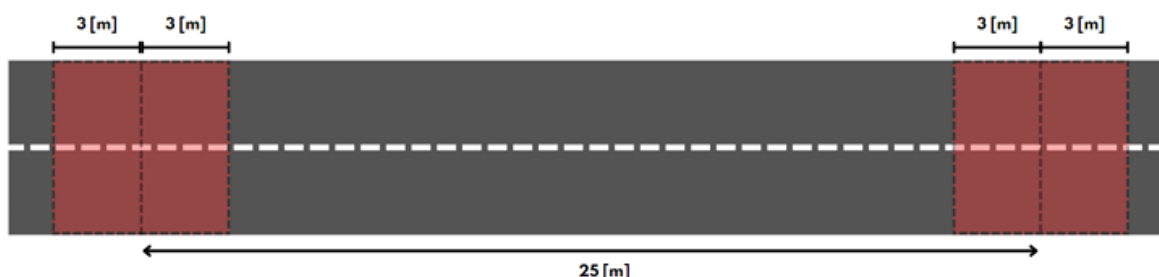


Figura 3: Auscultación manual de pavimentos.
Fuente: Castro (2024).

1.2.3 AUSCULTACIÓN AUTOMÁTICA

La auscultación automática implica el uso de sensores o herramientas para detectar los deterioros de manera más precisa sin involucrar directamente a los inspectores. El Sistema de Medición de Grietas con Láser, abreviado LCMS, es una tecnología láser que cuantifica el ancho y la longitud de fisuras presentes en los pavimentos. Su funcionamiento radica en la emisión de un haz láser sobre la superficie del pavimento. Este haz se refleja en la fisura y se mide el tiempo que tarda en recorrer la distancia desde el emisor hasta la fisura y de regreso, lo que lo vuelve un sistema de alta precisión. No obstante, presenta un aspecto desfavorable:

- Costo:** su uso implica un gasto significativo por parte de las entidades municipales, debido a la falta de recursos adecuados para su implementación.

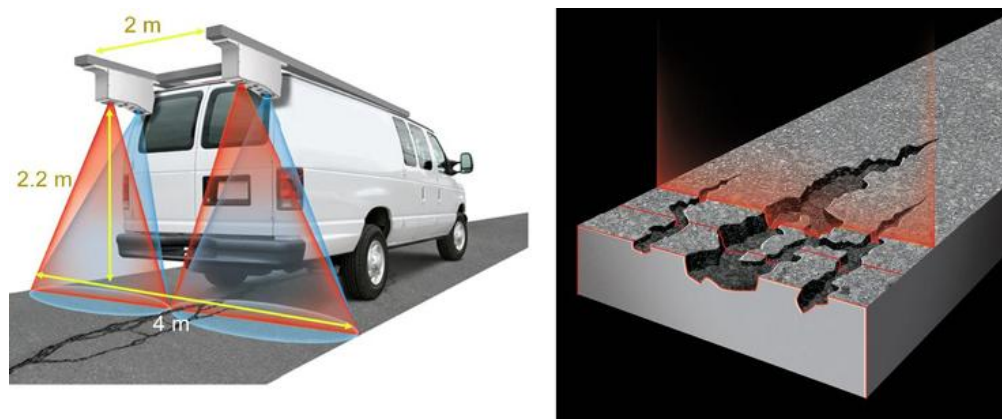


Figura 4: Tecnología Laser Crack Measurement System.
Fuente: Laurent et al. (2012).

1.3 PROYECTO FONDECYT 11201150

Debido a esto, el proyecto FONDECYT 11201150 propuso una metodología para la detección automática de deterioros, utilizando equipo de bajo costo. Titulado como “Evaluación y modelación del desempeño de pavimentos urbanos utilizando machine learning”, cuyo objetivo general es:

“Mejorar la eficiencia de la evaluación y la precisión de los modelos de desempeño para la condición del pavimento urbano, utilizando equipos de bajo costo para la recopilación de datos de campo y herramientas de aprendizaje automático para el procesamiento y análisis de datos”.

Mientras que los objetivos específicos son:

- Aplicar y validar una metodología para la recopilación de datos de campo de deterioros de pavimentos urbanos utilizando imágenes de bajo costo.
- Aplicar y validar algoritmos de *Deep Learning* para la detección automática de deterioros en el pavimento urbano a partir de imágenes recolectadas en terreno.
- Desarrollar y validar modelos de condición de pavimentos urbanos utilizando algoritmos de inteligencia artificial, considerando una combinación de herramientas para la predicción a corto y largo plazo.

Para lograr cada uno de estos objetivos es que el proyecto se divide en las siguientes tres etapas:

- Desarrollo y validación de una metodología para la recolección de datos de campo de deterioros de pavimentos urbanos con imágenes de bajo costo.
- Análisis del estado del pavimento urbano mediante la utilización de técnicas de *Deep Learning*.
- Modelado del desempeño del pavimento urbano mediante la utilización de técnicas de Inteligencia Artificial.

Aunque el proyecto finalizó, se encuentra en una nueva etapa, en la cual se busca hacer mejoras a la metodología incluyendo nuevas herramientas y tecnologías para obtener una mayor precisión en las mediciones.

1.4 ADQUISICIÓN Y PROCESAMIENTO DE IMÁGENES

1.4.1 SISTEMA DE ADQUISICIÓN DE IMÁGENES

El sistema de adquisición de imágenes consiste en una cámara GoPro que se estabiliza sobre un *monopad*, el cual se asegura en un portabicicletas junto con dos láseres que delimitan la zona a auscultar tras procesar la imagen (Venegas, 2022). El montaje en cuestión se muestra en la Figura 5.

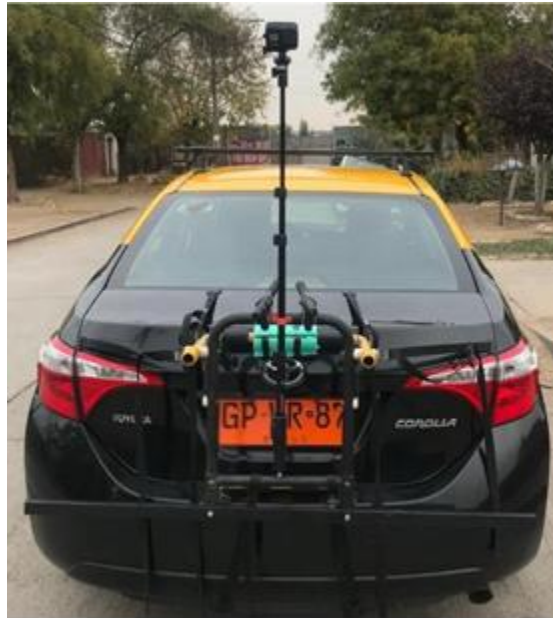


Figura 5: Equipo para adquisición de imágenes.
Fuente: Venegas (2022).

1.4.2 PROCESAMIENTO DE IMÁGENES

La cámara GoPro graba un video de la calle que será auscultada, a partir de la telemetría del video en los tiempos necesarios, se extraen imágenes de las secciones requeridas como se ve en la figura 6.



Figura 6: Imagen extraída del video.
Fuente: Almendras (2023).

Estas imágenes se redimensionan, normalizan y acoplan en lotes de 5 para formar una imagen que representa 10 metros de la vía a auscultar.

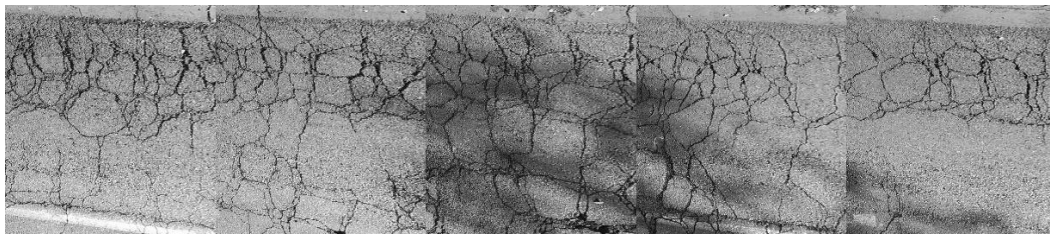


Figura 7: Imagen en sección de 10 [m] (volteada).
Fuente: Almendras (2023).

Sin embargo, los algoritmos de *Deep Learning* presentan distintos inconvenientes al analizar estas imágenes, como la clasificación parcial, errónea y/o nula de los deterioros. La clasificación parcial se relaciona a la nitidez de la imagen, al no verse con claridad los deterioros, la detección entrega dimensiones que se alejan de las dimensiones reales del deterioro. Mientras que la clasificación errónea o nula de los deterioros se relaciona con el contraste de luces en la imagen y la presencia de sombras que puede opacar deterioros. En base a esto, se analiza la variación de la nitidez al grabar videos de 24, 40 y 60 fotogramas y la variación del contraste al utilizar filtro polarizador circular (CPL) y filtros de densidad neutra (ND8, ND16 y ND32).

1.4.3 MEJORAMIENTO DE NITIDEZ

Para mejorar los resultados de la red neuronal se analizan las imágenes capturadas a 24, 30 y 60 fotogramas por segundo, y se llevan a cabo 3 tipos de evaluación; Binarización, Detección de deterioros con el proceso de inferencia a mejorar y Etiquetado manual de deterioros.

La binarización se utiliza para evaluar la nitidez con que la imagen detecta los deterioros, donde una imagen capturada a 24 FPS presenta mejores resultados de las grietas. En la figura 8 se ve la binarización de la misma imagen con distinta cantidad de FPS.

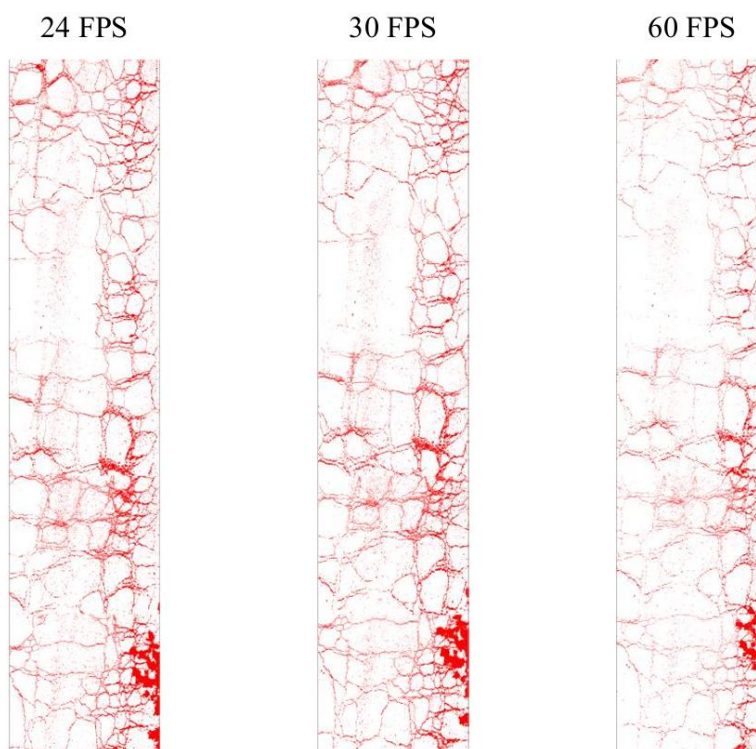


Figura 8: Imágenes de pavimento binarizadas.
Fuente: Almendras (2023).

Posteriormente, las imágenes se someten al proceso de inferencia con la red neuronal y se comparan con un etiquetado manual de los deterioros hecho en la imagen. Se concluye que las imágenes capturadas a 24 FPS entregan resultados más precisos acerca de los deterioros presentes en las imágenes. Aunque el aumento de FPS mejora la fluidez del video, la información relevante para la detección está contenida en cada fotograma, los cuales pierden calidad si se incrementa su cantidad.

1.4.4 MEJORAMIENTO DE CONTRASTE

En cuanto al contraste entre luces y sombras presentes en las imágenes, se estudia el uso de diferentes filtros ópticos para disminuir la diferencia de luz generada por sombras de árboles, postes de luz, y otros elementos cercanos al pavimento que no permiten identificar de manera clara los deterioros presentes.

Entre los filtros, se utilizan:

- a. Filtro polarizador circular (CPL)
- b. Filtros de densidad neutra (ND8, ND16 y ND32)

Tras aplicar los filtros sobre la misma imagen y someterlas al proceso de inferencia con la red neuronal, se obtienen mejores resultados con el filtro ND32. Sin embargo, este filtro no es capaz de eliminar completamente las sombras presentes en las imágenes, solo disminuye la diferencia de luz.

1.5 ENTRENAMIENTO Y EVALUACIÓN

1.5.1 DEFINICIÓN DE CONJUNTOS Y CLASES

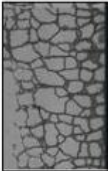
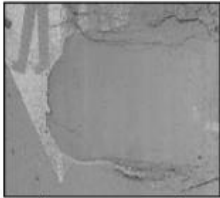
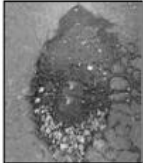
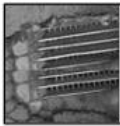
Una vez procesadas las imágenes, estas se almacenan en una base de datos que son utilizadas para entrenar y probar los algoritmos de *Deep Learning*. La cantidad de imágenes se divide en 3 conjuntos, los cuales no comparten elementos entre sí:

- a. **Conjunto de Entrenamiento:** este conjunto es la porción más grande y se utiliza para entrenar el modelo de aprendizaje automático. Aquí, el modelo aprende patrones y características de los deterioros en pavimentos urbanos. Se utiliza el 80% de las imágenes para el entrenamiento.
- b. **Conjunto de Validación (10%):** se utiliza para ajustar los hiperparámetros del modelo, tales como la tasa de aprendizaje o la arquitectura de la red, y evaluar su rendimiento durante el entrenamiento. Se utiliza el 10% de las imágenes para la validación.
- c. **Conjunto de Pruebas (10%):** este conjunto evalúa el rendimiento final del modelo luego de ser entrenado y validado. Se utiliza el 10% de las imágenes restantes que no han sido parte del conjunto de entrenamiento ni del conjunto de validación.

Para la detección de los deterioros se opta por usar dos algoritmos de *Deep Learning*; YOLOv5 y YOLOv7. Se definen 11 clases para entrenar a cada algoritmo y se compara el desempeño de ambos para la clasificación de objetos. Las clases usadas se muestran junto a un ejemplo en la Tabla 1.

TRATAMIENTO DE SOMBRAS PARA DETECTAR DETERIOROS EN PAVIMENTOS CON
MACHINE LEARNING

Tabla 1: Clases usadas para el entrenamiento de los algoritmos de Deep Learning.
Fuente: Castro (2024).

N°	Clase	Ejemplo	N°	Clase	Ejemplo
0	Fatiga Media		6	Parche Media	
1	Fatiga Alta		7	Parche Alta	
2	Transversal Media		8	Bache	
3	Transversal Alta		9	Tapa de Alcantarilla	
4	Longitudinal Media		10	Sumidero	
5	Longitudinal Alta		11	Testigo	

1.5.2 ANÁLISIS DE LOS DATASETS

Para generar los conjuntos de entrenamiento, validación y prueba, desde la base de datos generada principalmente con las imágenes obtenidas del proceso de adquisición de imágenes, se crean los siguientes datasets:

TRATAMIENTO DE SOMBRAS PARA DETECTAR DETERIOROS EN PAVIMENTOS CON MACHINE LEARNING

Tabla 2: Cantidad de imágenes usadas para la implementación de los algoritmos.
Fuente: Castro (2024).

Dataset	Dataset Completo	Dataset Reducido N°1	Dataset Reducido N°2
Imágenes Entrenamiento	22215	1275	2034
Imágenes Validación	832	195	342
Imágenes Test	833	196	340

El *dataset* completo está compuesto por 23880 imágenes, siendo algunas reales, obtenidas del proceso de adquisición de imágenes mencionado, y otras creadas artificialmente para equilibrar las clases y mejorar el proceso de entrenamiento de los algoritmos.

Por otro lado, los *datasets* reducidos se diferencian en la cantidad de imágenes empleadas. El *dataset* reducido N°1 utiliza un 7% del *dataset* completo aproximadamente, mientras que el *dataset* reducido N°2 utiliza alrededor de un 11% del *dataset* completo.

1.5.3 ANÁLISIS CON MÉTRICAS DE DESEMPEÑO

En las tablas 3 y 4, se muestra el desempeño obtenido por YOLOv5 y YOLOv7 luego del entrenamiento y validación en los *datasets* reducidos N°1 y N°2, y el promedio de diez *splits* del *dataset* completo.

Tabla 3: Métricas de desempeño para los *dataset* en el algoritmo YOLOv5.
Fuente: Castro (2024).

Parámetro	Métricas de Desempeño YOLOv5 [%]		
	Dataset Completo	Dataset Reducido	
	Promedio 10 Splits	N°1	N°2
Precisión	59,5	52,8	75,4
Recall	53,7	41,6	65,6
F1 -Score	56,4	46,5	70,2
mAP@50	54,4	39,7	67,9
mAP@50-95	30,5	20,7	42,8

TRATAMIENTO DE SOMBRAS PARA DETECTAR DETERIOROS EN PAVIMENTOS CON MACHINE LEARNING

Tabla 4: Métricas de desempeño para los *dataset* en el algoritmo YOLOv7.
Fuente: Castro (2024).

Parámetro	Métricas de Desempeño YOLOv7 [%]		
	Dataset Completo	Dataset Reducido	
	Promedio 10 Splits	Nº1	Nº2
Precisión	43,7	35,7	80,2
Recall	45,1	53,4	69,5
F1 -Score	44,4	42,8	74,5
mAP@50	40,4	37,6	72,1
mAP@50-95	22,9	21,6	51,0

De las métricas de desempeño, se puede concluir que el *dataset* reducido N°2 exhibe mejores resultados, mientras que el *dataset* completo presenta peores resultados junto con el *dataset* reducido N°1, el cual es el menos favorable. Se concluye que el uso de *datasets* con una gran cantidad de imágenes y otro con un número reducido de imágenes, afecta negativamente a las métricas, es decir, la capacidad de detectar deterioros en pavimentos urbanos.

Por otro lado, se evidencia que el rendimiento de YOLOv7 es inferior al del algoritmo YOLOv5, debido al uso escaso de imágenes o el uso de una gran cantidad de imágenes generadas artificialmente que no permiten al algoritmo predecir deterioros en las imágenes reales.

1.6 MEJORAS Y RECOMENDACIONES

De los estudios hechos para mejorar la metodología actual, acerca del mejoramiento de la nitidez y el contraste de las imágenes y la evaluación del desempeño de los algoritmos para la detección de deterioros, se hicieron las siguientes recomendaciones para futuros trabajos:

- Uso de inteligencia artificial para la eliminación de sombras:** se recomienda estudiar algoritmos de inteligencia artificial capaces de eliminar la iluminación desigual bajo el contexto específico de los pavimentos con presencia de grietas y sombras a la vez.
- Exploración de nuevos algoritmos de *Deep Learning* para la detección de deterioros:** Dada la rápida evolución en el campo de la detección de objetos en imágenes, se sugiere investigar y evaluar el rendimiento de nuevos algoritmos de *Deep Learning* que hayan demostrado eficacia en la detección de objetos. La exploración de nuevas técnicas puede ofrecer perspectivas valiosas y contribuir a la mejora continua de los métodos empleados en la detección automatizada de deterioros en pavimentos urbanos.

1.7 ÁRBOL

Por último, se presenta el árbol del problema, el cual sintetiza la problemática a resolver, sus causas y sus efectos.

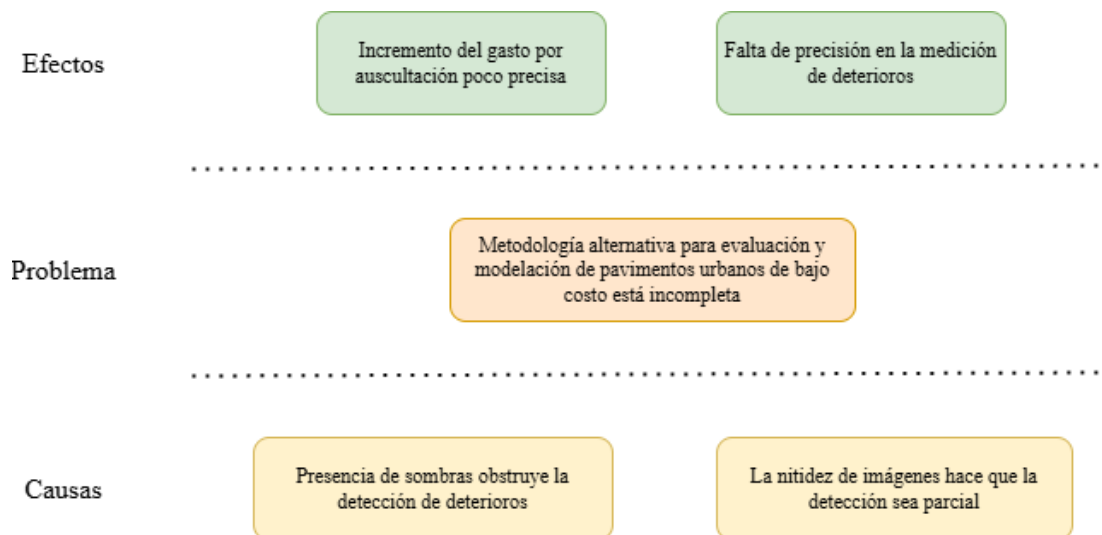


Figura 9: Árbol del problema.

Fuente: Elaboración Propia.

Como se mencionó la problemática en cuestión es que la metodología para la evaluación y modelación de pavimentos urbanos de bajo costo está incompleta, específicamente en su etapa de detección de deterioros, ya que estos no se perciben debido a la presencia de sombras que los ocultan o las sombras se detectan como deterioros erróneamente.

1.8 OBJETIVO GENERAL

Diseñar e implementar un algoritmo de aprendizaje automático para la corrección de sombras en imágenes de pavimentos urbanos 2D, que permita mejorar el desempeño de una metodología actual para la detección de deterioros en dichas imágenes.

1.9 OBJETIVOS ESPECÍFICOS

- Elaborar un *dataset* de imágenes de pavimentos urbanos 2D que presenten sombras de distintos objetos que ocluyan deterioros de manera parcial o completa.
- Seleccionar algoritmos que presenten un mejor desempeño en imágenes de pavimentos urbanos.
- Evaluar el impacto de la corrección de sombras y nitidez en la metodología de evaluación de bajo costo de pavimentos urbanos.

CAPÍTULO 2: MARCO CONCEPTUAL

La presencia de sombras en las imágenes de pavimentos dificulta la detección y clasificación de deterioros, debido a que ocuyen parte o la totalidad del deterioro o bien puede ser clasificadas erróneamente como deterioros. Para entender en profundidad el problema, se presentan algunos conceptos teóricos asociados, el tipo de imágenes con los que se trabaja y los métodos que se han implementado a lo largo del tiempo.

2.1 TIPOS DE SOMBRAS

La sombra es un fenómeno natural que se manifiesta cuando un objeto bloquea el paso de la luz hacia una superficie u objeto generando una menor iluminación en dicha área (Guo et al., 2024). Las sombras se pueden clasificar por el tipo de superficie en que se proyectan. Cuando una sombra se genera en la superficie del objeto que obstruye el paso de la luz, se conoce como *self-shadow*, pero si la sombra se genera en una superficie u otro objeto se conoce como *cast shadow*, tal como se muestra en la figura 10(a). Esta última se divide en dos secciones; umbra y penumbra.

La umbra corresponde al área interna de la sombra, siendo más oscura y alta densidad. La penumbra es el área externa de la sombra, es de baja densidad y es mucho más difuso, mezclándose con el área sin sombra circundante.

Una *cast shadow*, o sombra proyectada, se puede clasificar, además, en dos tipos de sombras; *soft shadow* y *hard shadow*. *Soft shadow* es un tipo de sombra más tenue, la cual se produce cuando la fuente de luz no es puntual y el tamaño de su penumbra puede variar según la distancia entre el objeto que produce la sombra y la superficie en que se genera, siendo mayor cuando el objeto se encuentra más alejado de la superficie. Mientras que *hard shadow* es un tipo de sombra con una umbra más intensa y sin penumbra, por lo que destaca su límite nítido respecto al fondo, esta clasificación se muestra en la figura 10(b).

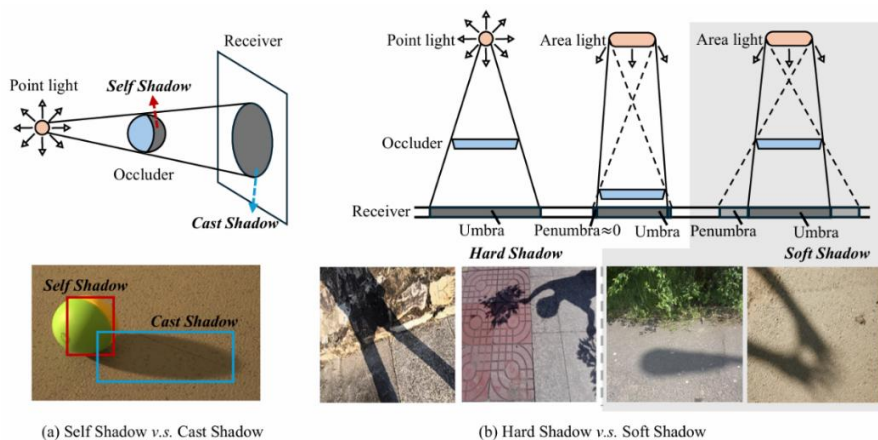


Figura 10: (a) Sombra según superficie en que se genera. (b) Sombra según fuente de luz y cercanía de objeto a superficie.

Fuente: Guo et al. (2024).

2.2 SOMBRAS EN IMÁGENES DE PAVIMENTOS URBANOS

En la figura se resume la metodología actual del proyecto.

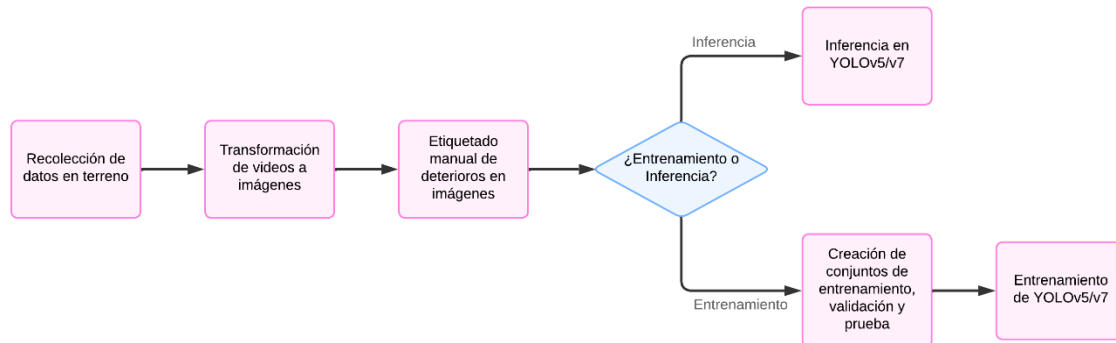


Figura 11: Diagrama resumen de la metodología para detectar deterioros.

Fuente: Elaboración Propia.

La eliminación de sombras se espera incorporar dentro de la metodología actual como parte del preprocesamiento de imagen, es decir, la etapa de Transformación de videos a imágenes. Sin embargo, dentro del preprocesamiento, se desconoce de qué manera es óptima la incorporación de la eliminación, por lo que se realizarán pruebas en las imágenes en distintas fases del preprocesamiento, las imágenes con la perspectiva corregida y las secciones de 5 imágenes concatenadas.

Las imágenes con la perspectiva corregida corresponden a fotogramas extraídos de los videos en base a su telemetría. Los fragmentos corresponden a instantes específicos de tiempo en los que se han recorrido 2,71 metros, de modo que puedan simular una secuencia continua del recorrido realizado. En la figura 12, se muestra una imagen original extraída de un video usando el código de la metodología.



Figura 12: Imagen de pavimento con sombras y grietas.

Fuente: Elaboración propia.

La resolución de estas imágenes es 3.840 x 2.160 píxeles, una vez obtenida la imagen se les ajusta la perspectiva, lo que resulta en la imagen de resolución 2.105 x 1.945 píxeles. Se muestra un ejemplo de la imagen con la perspectiva corregida en la figura 13.



Figura 13: Imagen de pavimento con perspectiva corregida.
Fuente: Elaboración propia.

Las otras imágenes son imágenes de tipo asfalto de alta resolución (de 2.105 x 9.725 píxeles), que se encuentra en escala de gris. La figura 7 muestra un ejemplo de la imagen que se utiliza para la detección de deterioros.

La característica principal de estas imágenes es el efecto de sal y pimienta que posee la textura del pavimento. El tipo de sombra que más se aprecia en las imágenes son *cast shadows* sin el objeto que opaca presente en la escena. En particular, las sombras son del tipo *soft shadow*, ya que la penumbra es grande y tenue, mezclándose con el fondo, y en algunos casos la umbra es muy oscura, ocultando grietas de una tonalidad similar.

2.3 MÉTODOS ESTUDIADOS

2.3.1 BASADOS EN PROCESAMIENTO DE IMÁGENES

Dentro de los métodos basados en procesamiento de imágenes estudiados, la mayoría se caracteriza por tener la siguiente estructura: preprocesamiento, detección de sombra y eliminación de sombra. A continuación, se definen brevemente las técnicas utilizadas en cada etapa.

2.3.1.1 PREPROCESAMIENTO

En esta etapa se hacen retoques a la imagen para remover cualquier detalle que pueda dificultar la detección de la sombra, por lo general, en esta etapa se aplican operaciones morfológicas y/o filtros.

Morfología matemática: herramienta para extraer componentes de imagen que son útiles para la representación y descripción de la forma de una región, como límites, esqueletos, etc. (Gonzalez y Woods, 1977). Las operaciones morfológicas más primitivas son erosión, dilatación, apertura y cierre. De estas operaciones nacen distintos algoritmos como *black-*

top-hat. A continuación, se definen las operaciones y algoritmos usados en los métodos tradicionales.

- **Erosión:** Barrer un elemento estructural, generalmente una matriz cuadrada con valores 1, sobre una imagen y conservar los píxeles de la imagen que corresponden al centro del elemento estructural cuando está completamente contenido por píxeles del objeto A en la imagen.

$$A \ominus B = \{z \mid (B)_z \subseteq A\}$$

- **Dilatación:** Barrer un elemento estructural sobre una imagen y si este intercepta en al menos un píxel del objeto A, se agregan los píxeles del elemento estructural en la imagen.

$$A \oplus B = \{z \mid (\hat{B})_z \cap A \neq \emptyset\}$$

- **Cierre:** Se define como la erosión de una dilatación, es decir, cierra huecos pequeños. Bi et al. (2024) lo usan para uniformar el fondo de imágenes de pavimentos eliminando grietas y pequeños huecos.

$$A \cdot B = (A \oplus B) \ominus B$$

- **Black-top-hat (bottom hat):** Algoritmo que corresponde a la resta entre el cierre de una imagen y la imagen original, de esta manera se conservan pequeños detalles oscuros. Imahayashi et al. (2023) lo emplea para conservar los caracteres en imágenes de documentos y de Azevedo et al. (2013) resalta sombras pequeñas en imágenes aéreas.

$$B_{hat}(f) = (f \cdot b) - f$$

Filtrado de imagen: Los filtros en imágenes digitales son usados para el desenfoque y la reducción de ruido, la nitidez y la detección de bordes. Los filtros pueden ser divididos en dos categorías según su dominio: espacial o de frecuencia (Dilhani, 2021). Dentro del dominio espacial, los filtros se pueden clasificar como lineales y no lineales como se muestra en la figura 12.

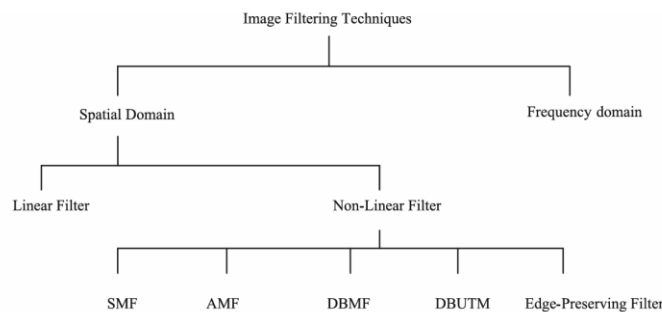


Figura 14: Clasificación de filtros.

Fuente: Bisht et al. (2018).

En los métodos tradicionales se han aplicado los siguientes filtros:

- **Filtro de mediana:** Es un filtro no lineal, reemplaza cada píxel con el valor de la mediana entre sus píxeles vecinos. Es eficiente para eliminar el ruido sal y pimienta (Dilhani, 2021). En los métodos tradicionales, por lo general se utiliza como última etapa para no perder información de objetos en la imagen (Kamaruddin et al., 2022) o para eliminar el ruido y mejorar el contraste entre grietas y fondo (Bi et al., 2024). Imahayashi et al. (2023) lo usan para eliminar los caracteres de imágenes de documentos, pero lo aplica solamente en aquellos píxeles donde hay caracteres según la máscara resultante luego de aplicar *black-top-hat*.
- **Filtro gaussiano:** Es un operador 2D convolucional, muy similar al filtro de mediana, se usa para desenfocar y suavizar la imagen, inclinando el peso de los píxeles hacia el centro de un *kernel* según una distribución gaussiana (Dilhani, 2021). Bi et al. (2024) lo utilizan para suavizar el efecto sal y pimienta que queda en las imágenes de pavimentos luego de haberles aplicado una operación de cierre morfológico.

2.3.1.2 DETECCIÓN DE SOMBRA

Según el espacio de color de la imagen, se aplican técnicas que permiten dividir la imagen en dos conjuntos; los píxeles que contiene una sombra y los que conforman el fondo de la imagen. Si se trabaja con imágenes de un solo canal (escala de gris) se puede aplicar una única segmentación, mientras que en imágenes con 3 canales (RGB) se aplican segmentaciones por cada canal en base a un tono de gris o bien un índice calculado con los valores de cada canal, como es el caso de *Shadow Detection Index* propuesto por Gao et al. (2022) o en el caso de Kamaruddin et al. (2022) que plantean un índice con información de distintos espacios de color.

Segmentación: proceso en el que se particiona una imagen R en n subregiones. Dentro de la segmentación de imagen hay una técnica conocida como *thresholding*, que consiste en la partición de la imagen en distintos niveles, divididos por uno o varios umbrales basado en los tonos de gris (Gonzalez y Woods, 1977). Sea (x,y) un punto cualquiera en una imagen $f(x,y)$, la imagen segmentada $g(x,y)$ por un umbral T se define como:

$$g(x,y) = \begin{cases} 1 & \text{si } f(x,y) > T \\ 0 & \text{si } f(x,y) \leq T \end{cases}$$

Este caso se conoce como *global thresholding*, en el cual un umbral único y constante se aplica sobre toda la imagen dividiéndola en dos grupos, aquellos píxeles que valen 1 se denominan *object pixel* mientras que el resto se conoce como *background pixel*. En la figura 11 se muestra el uso de *global thresholding*.

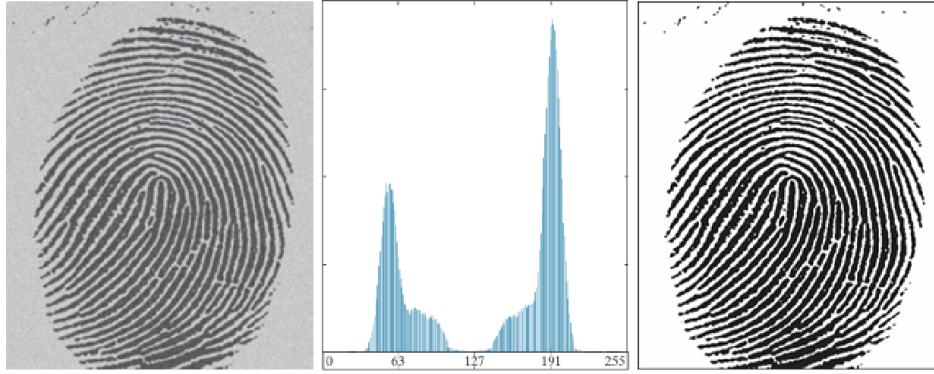


Figura 15: (a) Imagen ruidosa. (b) Histograma de los tonos de gris de la imagen. (c) Imagen segmentada usando *global thresholding*.

Fuente: Gonzalez y Woods (1977).

En los métodos tradicionales, esta técnica se utiliza para encontrar máscaras de sombras ya sea usando un umbral de valor constante (Kamaruddin et al., 2022) o bien usando métodos que optimizan el umbral para cada caso como el método de Otsu utilizado por Imahayashi et al. (2023), da Silva et al. (2013) y Gao et al. (2022).

Shadow Detection Index (SDI): índice que se calcula en base a los valores de cada canal de color de un píxel. Gao et al. (2022) proponen el siguiente índice:

$$SDI = \omega \cdot |2G - B - R| + \varepsilon \cdot G$$

En el cual, se usa la resta $2G - B - R$ ya que, en base a un análisis de las características ópticas de las distintas restas entre canales y sus correlaciones, tiene el mayor contraste entre los píxeles de objeto y píxeles de fondo, en otras palabras, permite separar mejor la sombra del resto de la imagen. Además, se destaca el canal verde debido a la gran presencia de áreas verdes en imágenes UAV. En cuanto a ω y ε , corresponden a parámetros tal que $\omega + \varepsilon = 1$.

Índice en base a múltiples espacios de color: la imagen original que se encuentra en RGB se transforma a HSV, YCbCr y LAB, de los cuales se extrae el valor de los píxeles en S, Cb y B respectivamente, y se usan en la siguiente fórmula:

$$I(x, y) = \frac{Cb(x, y) + B(x, y) + S(x, y)}{2}$$

2.3.1.3 ELIMINACIÓN DE SOMBRA

Luego de detectar en qué píxeles se encuentra la sombra, se aplican fórmulas que corrigen la iluminación de dichos píxeles o cambian su tono de gris para que se asemeje al de los píxeles que no están sombreados. Bi et al. (2024) utilizan una fórmula de compensación de brillo, Gao et al. (2022) en base a la entropía de la región sombreada aplican una relación de iluminancia o diferencia de escala de grises. Kamaruddin et al. (2022) aplica un cambio de valor a los píxeles en base a valores empíricos, Imahayashi et al. (2023) calculan el color

promedio del fondo de la imagen y aplica ese color por toda la imagen eliminando las sombras.

Compensación de brillo: Busca reemplazar el valor del píxel según valores promedios y varianzas de pixeles sombreados y no sombreados. Bi et al. (2024) proponen la siguiente fórmula para la compensación de brillo:

$$I'_{i,j} = \begin{cases} \alpha \cdot I_{i,j} + \lambda & si(i,j) \in S \\ I_{i,j} & si(i,j) \in B \end{cases}$$

Donde S corresponde a una región sombreada y B a una región no sombreada. La fórmula indica que la compensación de brillo se aplica solamente en las regiones sombreadas. Los parámetros α y λ se definen como:

$$\alpha = \frac{D_B}{D_S}, \quad \lambda = \hat{I}_B - \alpha \cdot \hat{I}_S$$

En donde D_B y D_S representan la varianza del valor de un píxel en la región no sombreada y sombreada respectivamente. Mientras que $\hat{I}_B$ e $\hat{I}_S$ corresponden al valor promedio del píxel en las regiones respectivas. Estas regiones se definen en base a niveles de brillo, siendo los niveles más bajos considerados como sombras.

Relación de iluminancia: Gao et al. (2022) definen otra fórmula basada en la reflectancia e iluminancia de un píxel en una región sombreada y no sombreada. De esta manera, definen una proporción de iluminación r que se expresa como:

$$r = \frac{I_i^{ns} - I_i^s}{I_i^s}$$

Donde I_i^{ns} e I_i^s son los valores promedio de los pixeles en la región no sombreada y sombreada respectivamente. Y este factor se aplica de la siguiente manera:

$$I_i^{Sfree} = (r + 1) \cdot I_i^s$$

Siendo I_i^{Sfree} el píxel resultante sin la sombra.

Diferencia de escala de gris: Gao et al. (2022) aplican un segundo método basado en la diferencia de escala de gris de los valores promedio que se expresa así:

$$d = \frac{1}{M} \sum_{i=1}^M I_i^{ns} - \frac{1}{N} \sum_i^N I_i^s$$

Y que se aplica:

$$I_i^{Sfree} = I_i^s + d$$

El uso de este método o el anterior se decide según el valor de entropía que tenga la región sombreada, es decir, qué tan caótica sea la distribución de los valores de los píxeles en dicha región.

Valor promedio de fondo: Imahayashi et al. (2023) calculan el color promedio G^c de los píxeles que corresponden al fondo de la imagen. Para ello, se basan en la imagen L^c que obtienen luego de aplicar el filtro de mediana y una máscara binaria L_s^c calculada con el método de Otsu en L^c .

$$G^c = \frac{\sum_{(i,j) \in \Omega} L_s^c(i,j) L^c(i,j)}{\sum_{(i,j) \in \Omega} L_s^c(i,j)}$$

Posteriormente, sobre los píxeles de la imagen original aplican esta fórmula:

$$O^c = I^c(i,j) \cdot \frac{G^c}{L^c(i,j)}$$

Esto permite obtener una imagen con un fondo de color más uniforme y conservar cualquier detalle oscuro que sea distinto a una sombra.

2.3.1.4 OTRAS TÉCNICAS USADAS

Transformada de Fourier Discreta 2D: Esta fórmula sirve para cambiar el dominio de una imagen, pasando del dominio espacial al dominio de la frecuencia. En su forma discreta, se define de esta manera:

$$F(u, v) = \frac{1}{N} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) e^{-\frac{j2\pi(ux+vy)}{N}}$$

Es usada por Li et al. (2021), para transformar imágenes de autopista a otro dominio y obtener diccionarios que describan la imagen como una de suavidad y otra de textura.

2.3.1.5 MÉTRICAS DE EVALUACIÓN

A continuación, se describen las métricas aplicadas para evaluar el rendimiento de la detección de sombra y/o la eliminación de cada método.

Mean Squared Error (MSE): Usado por Imahayashi et al. (2023), el error cuadrático medio corresponde a la distancia euclidiana que hay entre el valor obtenido y su valor real. Se utiliza para comparar el error que hay entre la imagen resultante y su *ground truth*, es decir, su referencia o resultado esperado.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Mean Structural Similarity (MSSIM): Usado por Imahayashi et al. (2023), el índice de similitud estructural cuantifica la calidad de una imagen usando otra como referencia en base a 3 componentes: iluminancia, contraste y estructura (Bovik et al., 2004).

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

El valor promedio de SSIM permite evaluar la calidad general de la imagen según el valor del índice en parches de la imagen:

$$MSSIM(X, Y) = \frac{1}{M} \sum_{j=1}^M SSIM(x_j, y_j)$$

Donde X e Y representan la imagen de referencia y la imagen distorsionada, y x_j e y_j son el contenido de la imagen en el parche j -ésimo.

Cuando MSSIM es cercano a 1, significa que las imágenes son estructuralmente similares. Esta métrica es usada por Imahayashi et al. (2023).

Color Difference (CD): Usada por Gao et al. (2022), mide la consistencia de color en una imagen basándose en la diferencia de color de las imágenes en el sistema CIE $L^*a^*b^*$, usando esta ecuación:

$$\Delta E_{lab} = [(\Delta L)^2 + (\Delta a)^2 + (\Delta b)^2]^{1/2}$$

La relación entre el valor de la diferencia de color y la percepción visual es:

- $\Delta E < 1$: la diferencia apenas se percibe.
- $1 < \Delta E \leq 2$: hay una ligera percepción en la diferencia de color.
- $2 < \Delta E \leq 3,5$: la percepción de la diferencia es moderada.
- $3.5 < \Delta E \leq 6$: la percepción de la diferencia es obvia.
- $\Delta E > 6$: la diferencia de color es muy fuerte.

Shadow Standard Index Detection (SSID): Usada por Gao et al. (2022), este índice refleja la variación en las regiones sombreadas luego de la eliminación con respecto a su región sombreada respectiva.

$$\sigma_{s-ns} = \frac{1}{B} \sum_{b=1}^B \sqrt{\frac{1}{N} \sum_{i=1}^N (F_{b,i}^s - \overline{F_b^{ns}})^2}$$

Cuando el valor del índice es bajo, significa que la región sombreada es consistente con la región no sombreada en términos de textura. Si el valor es alto, la diferencia en textura entre las regiones es muy significativo.

Gradient Similarity (GS): Usada por Gao et al. (2022), representa la similitud entre los gradientes de dos bloques de la imagen resultante, cuyo rango va entre 0 y 1.

$$g(x, y) = \frac{2g_x g_y + C}{g_x^2 + g_y^2 + C}$$

Shadow Detection Rate (R_1): Usada por Li et al. (2021), representa el éxito de la detección de píxeles sombreadas, según el total de píxeles sombreadas y los que son clasificados correctamente como píxeles sombreados.

$$R_1 = \frac{TP_s}{TP_s + FN_s}$$

Donde TP_s corresponde a todos los píxeles correctamente clasificados como píxeles sombreados y FN_s corresponde a todos los píxeles que se clasificaron como píxeles sin sombra por error. Mientras mayor es el valor de R_1 , significa que la detección fue mejor por lo que se consigue un mejor efecto en la separación de la imagen.

Matriz de confusión: Usada por Kamaruddin et al. (2022), es una matriz cuadrada que indica la relación entre la clasificación real y predicha de los píxeles de una imagen.

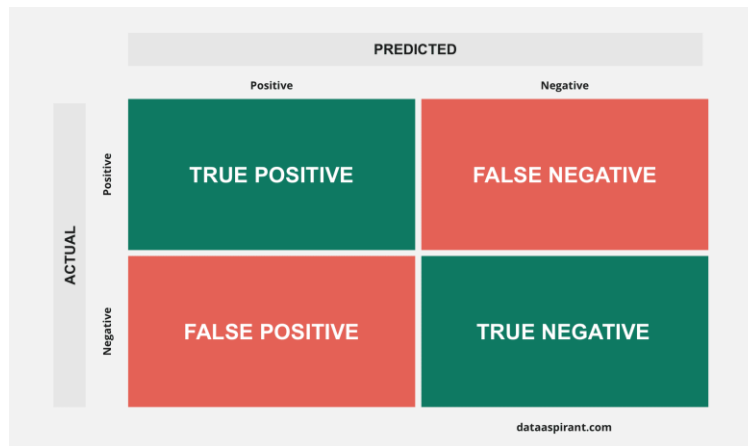


Figura 16: Esquema de matriz de confusión.
Fuente: Polamuri.

En la diagonal de la matriz se observan todos los datos que fueron bien clasificados en su respectiva clase, mientras que el resto corresponden a valores mal clasificados como falsos positivos o falsos negativos.

2.3.2 BASADOS EN DEEP LEARNING

2.3.2.1 MÉTODOS ESTUDIADOS

Guo et al. (2024) realizaron un estudio de los métodos de eliminación de sombras en imágenes basados en *Deep Learning*. Del estudio, destacan 4 métodos que pueden ser utilizados en imágenes que no poseen una máscara de las sombras, que es la condición en que se encuentran las imágenes del proyecto.

DC-SHADOWNET

El primer método estudiado es *DC-ShadowNet*, una red elaborada por Jin et al. (2021). La red en cuestión consiste principalmente en el uso de redes neuronales generativas adversarias para la generación de imágenes sin sombra, discriminando su veracidad en base a otras imágenes sin sombra, además integra un clasificador de dominio que sirve como mecanismo de atención para encasillar la imagen de entrada en uno de dos dominios: sombra (s) o sin sombra (sf). En caso de ser una imagen con sombra, el mecanismo permite enfocarse en las regiones con sombra mediante la construcción de un mapa de atención, sino el resultado de salida es la misma imagen sin sombra. La arquitectura general de *DC-ShadowNet* se muestra en la figura 16.

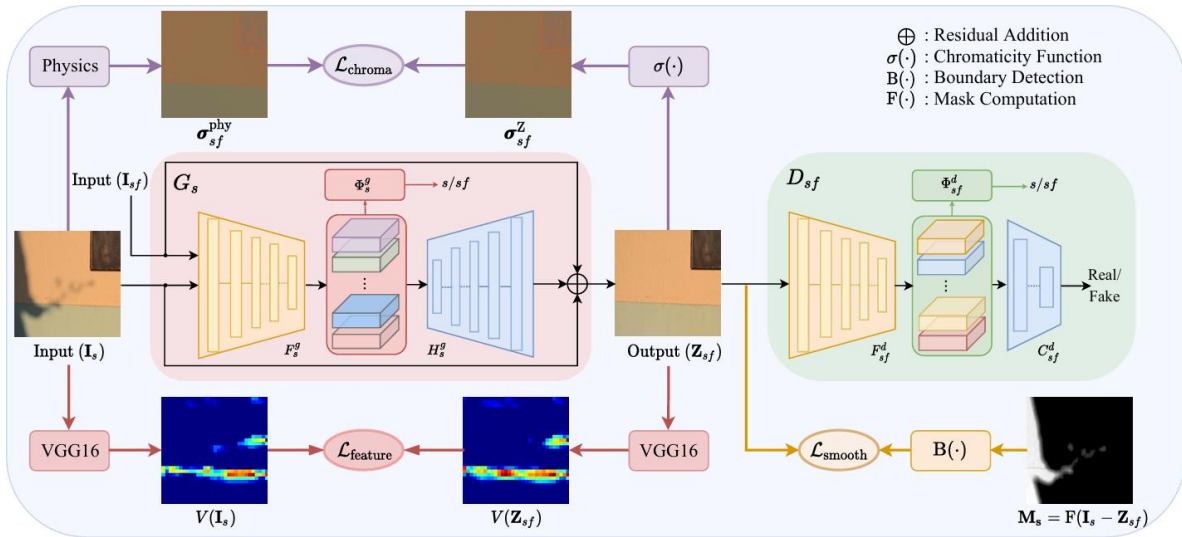


Figura 17: Arquitectura de DC-ShadowNet.

Fuente: Jin et al. (2021).

En la arquitectura destacan dos bloques: el generador (G_s) y el discriminador (D_{sf}). El generador G_s genera una imagen sin sombra Z_{sf} , usando la imagen de entrada I_s , que puede ser con o sin sombra. Consiste en un codificador F_s^g , un decodificador H_s^g y el clasificador de dominio Φ_s^g . F_s^g se encarga de extraer los mapas de características Π_s^g de la imagen de entrada, con el uso de GAP (*Global Average Pooling*), se entrena el clasificador de dominio para que aprenda los pesos w_s^g de cada mapa. En base a un promedio ponderado de los pesos de los mapas Π_s^g se obtiene el mapa de atención A_s^g , el cual realiza

las sombras de la imagen de entrada. Luego el decodificador H_s^g hace la reconstrucción de la imagen en base a los mapas de características ponderados $w_s^g \Pi_s^g$, eliminando la sombra si la imagen de entrada fue clasificada en el dominio s , o reconstruyendo la imagen de entrada si es del dominio sf .

El discriminador D_{sf} es el encargado de analizar la calidad de la imagen generada, discerniendo si es real o no. Consiste en un codificador F_{sf}^d , un clasificador C_{sf}^d y el clasificador de dominio Φ_{sf}^d . En base a los mapas de características extraídos con F_{sf}^d , la imagen que recibe D_{sf} se clasifica nuevamente usando Φ_{sf}^d , por último C_{sf}^d se encarga de discriminar si la imagen es real o no en base a la etiqueta binaria que asignó a imágenes de referencia con las que se entrenó.

Para el entrenamiento de la red se consideraron las siguientes funciones de pérdida:

- **Shadow-free Chromaticity Loss:** se basa en la cromaticidad de la imagen de entrada usando el factor σ_{sf}^{phy} , el cual se obtiene de la minimización de entropía del espacio *log-chromaticity* y se ajusta su brillo en base al 30% de los pixeles más brillantes muestreados uniformemente. Además, se obtiene el mapa de cromaticidad de la imagen generada σ_{sf}^Z como:

$$\sigma_{sf_c}^Z = \frac{Z_{sf_c}}{(Z_{sf_r} + Z_{sf_g} + Z_{sf_b})} \quad c \in [r, g, b]$$

$$\sigma_{sf}^Z = [\sigma_{sf_r}^Z, \sigma_{sf_g}^Z, \sigma_{sf_b}^Z]$$

De esta manera, la función de pérdida de cromaticidad sin sombra es:

$$\mathcal{L}_{chroma}(G_s) = \mathbb{E}_{I_s} \left[\|\sigma_{sf}^Z - \sigma_{sf}^{phy}\|_1 \right]$$

Esta función se usa para asegurar que la imagen resultante conserva la cromaticidad de la imagen original.

- **Shadow-Robust Feature Loss:** como no se posee un *ground truth* de la imagen de entrada I_s , se usan las características de esta. En base a VGG16, que es un modelo robusto para la clasificación de imágenes, se usan los mapas de características por la red preentrenada, específicamente el mapa de características generado por la capa convolucional 2-2, debido a que es menos afectado por las sombras y representa información estructural de la imagen como bordes.

Sea $V(x)$, el mapa de características de la capa convolucional 2-2 de la imagen x , la función de pérdida es:

$$\mathcal{L}_{feature}(G_s) = \mathbb{E}_{I_s} \left[\left\| V(Z_{sf}) - V(I_s) \right\|_1 \right]$$

- **Domain Classification Loss:** como se mencionó anteriormente, se agregó un clasificador como mecanismo de atención para enfocarse en las regiones con sombra tanto en el generador G_s , como en el discriminador D_{sf} . Se usan estas funciones de pérdida:

$$\mathcal{L}_{domcls}(G_s) = \mathbb{E}_{I_s} \left[-\log \left(\Phi_s^g \left(F_s^g(I_s) \right) \right) \right] + \mathbb{E}_{I_{sf}} \left[-\log \left(1 - \Phi_s^g \left(F_s^g(I_{sf}) \right) \right) \right]$$

$$\mathcal{L}_{domcls}(D_{sf}) = \mathbb{E}_{I_s} \left[-\log \left(\Phi_{sf}^d \left(F_{sf}^d(I_s) \right) \right) \right] + \mathbb{E}_{I_{sf}} \left[-\log \left(1 - \Phi_{sf}^d \left(F_{sf}^d(I_{sf}) \right) \right) \right]$$

Esta última es preentrenada y sus pesos se congelan durante el ciclo principal de entrenamiento de la red.

- **Boundary Smoothness Loss:** se usa para asegurar una transición suave en los bordes de las sombras eliminadas. La función de pérdida es:

$$\mathcal{L}_{smooth}(G_s) = \mathbb{E}_{I_s} \left[\left\| B(M_s) \cdot \nabla(Z_{sf}) \right\|_1 \right]$$

Donde $\nabla(Z_{sf})$ es el gradiente de la imagen de salida Z_{sf} , M_s es una máscara de la sombra de imagen que se define como:

$$M_s = F(I_{sc} - Z_{sf_c}) = \sum_{c \in (r,g,b)} \frac{1}{3} |N(I_{sc} - Z_{sf_c})|$$

$N(I)$ es la normalización de la imagen I . Por último, $B(x)$ es la función de robustez de ruido.

- **Adversarial Loss:** para asegurar la consistencia de reconstrucción, se usa otro generador G_{sf} asociado a su propio discriminador D_s . Por lo tanto, la pérdida adversaria de los generadores y discriminadores es:

$$\mathcal{L}_{adv}(G_s, D_{sf}) = \mathbb{E}_{I_{sf}} \left[\log \left(D_{sf}(I_{sf}) \right) \right] + \mathbb{E}_{I_s} \left[\log \left(1 - D_{sf}(G_s(I_s)) \right) \right]$$

$$\mathcal{L}_{adv}(G_{sf}, D_s) = \mathbb{E}_{I_s} \left[\log \left(D_s(I_{sf}) \right) \right] + \mathbb{E}_{I_{sf}} \left[\log \left(1 - D_s(G_{sf}(I_{sf}, M_s)) \right) \right]$$

Se busca minimizar como $\min(G_s), \max(D_{sf}) \left(\mathcal{L}_{adv}(G_s, D_{sf}) \right)$ y $\min(G_{sf}), \max(D_s) \left(\mathcal{L}_{adv}(G_{sf}, D_s) \right)$.

- **Reconstruction Consistency Loss:** Para obtener una reconstrucción más apropiada se emplea la máscara M_s , por lo que la función de pérdida para los generadores G_s y G_{sf} es:

$$\mathcal{L}_{cons}(G_s) = \mathbb{E}_{I_s} \left[\left\| G_{sf}(G_s(I_s, M_s) - I_s) \right\|_1 \right]$$

$$\mathcal{L}_{cons}(G_{sf}) = \mathbb{E}_{I_{sf}} \left[\left\| G_s(G_{sf}(I_{sf}, M_s) - I_{sf}) \right\|_1 \right]$$

- **Identity Loss:** aunque el generador G_s está implementado para eliminar las sombras de la imagen de entrada I_s , también debe ser capaz de generar la misma imagen de entrada I_s cuando esta corresponde a una imagen del dominio sf . Por lo tanto, la función de pérdida de identidad para los generadores es:

$$\mathcal{L}_{iden}(G_s) = \mathbb{E}_{I_s} \left[\left\| G_s(I_{sf}) - I_{sf} \right\|_1 \right]$$

$$\mathcal{L}_{iden}(G_{sf}) = \mathbb{E}_{I_{sf}} \left[\left\| G_{sf}(I_s, M_0) - I_s \right\|_1 \right]$$

En esta última, M_0 es una máscara donde todos sus valores son ceros.

- **Overall Loss:** corresponde a una función de pérdida generalizada que pondera las funciones de pérdida $\mathcal{L}_c$ por sus pesos λ_c y los suma.

$$\mathcal{L}_{overall}(\lambda_c, \mathcal{L}_c) = \sum_{c \in \{chroma, feat, sm, dom, adv, cons, iden\}} \lambda_c \mathcal{L}_c$$

Los pesos de cada función de pérdida $\{\lambda_{chroma}, \lambda_{feat}, \lambda_{sm}, \lambda_{iden}, \lambda_{adv}, \lambda_{cons}, \lambda_{dom}\}$ se establecieron empíricamente como $\{1, 1, 1, 10, 1, 10, 1\}$.

Este método se probó en los *datasets* ISTD+ y SRD, usando RMSE como métrica de evaluación. Este método se comparó con *Mask-ShadowGAN*, arquitectura basada en entrenamiento no supervisado propuesta por Fu et al. (2019).

En el *dataset* ISTD+, *DC-ShadowNet* obtuvo un RMSE de 4,6 que superó a *Mask-ShadowGAN* que sólo obtuvo un 5,3. Mientras que en el *dataset* SRD, *DC-ShadowNet* obtuvo un 4,66, un resultado mucho menor comparado con el 6,40 obtenido por *Mask-ShadowGAN*. Así que, cuantitativamente, *DC-ShadowNet* genera imágenes más fidedignas a las originales.

Aunque el método presenta mejores resultados en los *datasets* mencionados, tiene dificultades al momento de eliminar *soft shadows*.

LG-SHADOWNET

LG-ShadowNet es un método propuesto por Lie et. al (2021), el cual se basa en el uso de dos módulos de redes neuronales convolucionales, específicamente se aplican como redes generativas adversarias, donde el generador G_f^L obtenido en el primer módulo entrenado individualmente se utiliza para complementar y guiar al generador G_f^{Lab} obtenido en el segundo módulo. De esta manera, y junto a una nueva función de pérdida $\mathcal{L}_{colour}$, se elimina la sombra en imágenes que no tienen un par de referencia necesariamente. En la figura 17 se muestra la arquitectura de *LG-ShadowNet*.

TRATAMIENTO DE SOMBRAS PARA DETECTAR DETERIOROS EN PAVIMENTOS CON MACHINE LEARNING

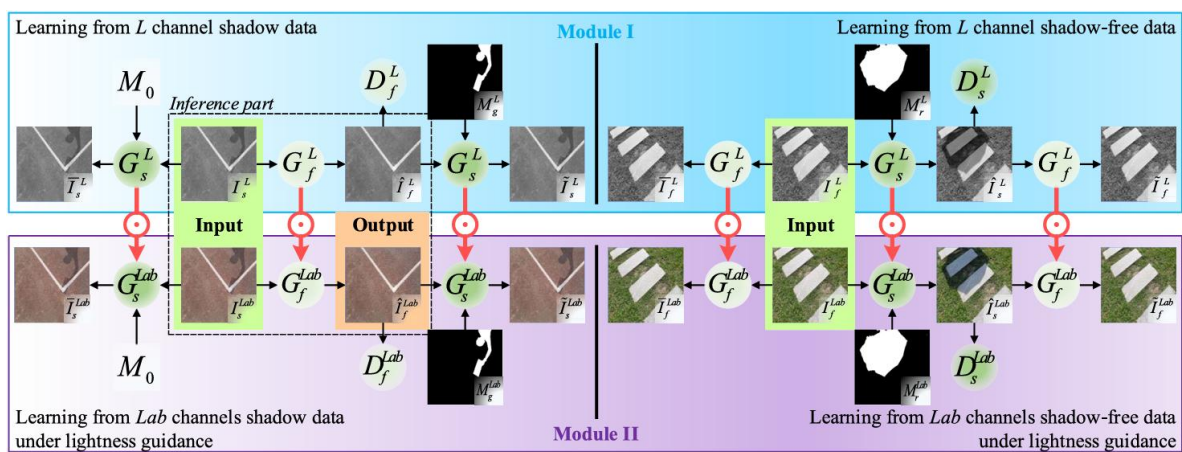


Figura 18: Arquitectura de LG-ShadowNet.

Fuente: Liu et al. (2021).

En el recuadro azul se presenta el Módulo I, cuya entrada es el canal L de la imagen con sombra en el espacio Lab, denominada I_s^L . El objetivo de este módulo es entrenar el generador G_f^L para compensar la iluminación en las regiones sombreadas, apoyándose también en el generador inverso G_s^L y en los discriminadores D_s^L y D_f^L . El entrenamiento se realiza sobre datos no pareados, diferenciando oscuridad y luminosidad en el canal L. En cuanto a la estructura, para los generadores se emplea la arquitectura descrita por Fu et al. (2019) en su método *Mask-ShadowGAN*, que se basa en el uso de capas de convolución, bloques residuales y normalización de instancia, mientras que los discriminadores se basan en *PatchGAN* para garantizar la coherencia local de la luminosidad general. Este módulo se entrena individualmente antes que el Módulo II, ya que este se encarga de ser un guía sobre la iluminación de la imagen que generada por G_f^{Lab} , el generador que se entrena en el Módulo II.

En el recuadro morado se encuentra el Módulo II, que comparte los mapas de características de los generadores del Módulo I para recibir una guía directa sobre la luminosidad en los mapas de características que crean sus generadores G_f^{Lab} y G_s^{Lab} . Es así como el Módulo II, cuya entrada es la imagen con sombra en el espacio Lab, denominada I_s^{Lab} , hace la eliminación de la sombra en la imagen. Al igual que en el Módulo I, sus generadores están basados en los que se presentan en *Mask-ShadowGAN*, mientras que los discriminadores se basan en *PatchGAN*.

El entrenamiento de este método consiste en dos pasos, en el primer paso se entrena solamente el Módulo I para ajustar sus parámetros y puede ser usado como guía de iluminación, en el segundo paso se entrena el Módulo II con guía del Módulo I usando conexiones multiplicativas para compartir los mapas de características del Módulo I.

Para el primer paso del entrenamiento, en el Módulo I se usan las siguientes funciones de pérdida:

- **Identity Loss:** se enfoca en que $\tilde{I}_s^{Lab}$ e $\tilde{I}_f^{Lab}$ sean iguales a I_s^{Lab} e I_f^{Lab} respectivamente.

$$\mathcal{L}_{identity} = \mathcal{L}_{identity}^s(G_s) + \mathcal{L}_{identity}^f(G_f)$$

Por lo que la función de pérdida es:

$$\mathcal{L}_{identity} = \mathbb{E}_{I_s \sim p(I_s)} [\|G_s(I_s, M_0), I_s^{Lab}\|_1] + \mathbb{E}_{I_f \sim p(I_f)} [\|G_f(I_f), I_f^{Lab}\|_1]$$

- **Cycle-consistency Loss:** se enfoca en que $\tilde{I}_s^{Lab}$ y $\tilde{I}_f^{Lab}$ sean iguales a I_s^{Lab} e I_f^{Lab} respectivamente.

$$\mathcal{L}_{cycle} = \mathcal{L}_{cycle}^s(G_f, G_s) + \mathcal{L}_{cycle}^f(G_s, G_f)$$

Por lo que la función de pérdida es:

$$\mathcal{L}_{identity} = \mathbb{E}_{I_s \sim p(I_s)} [\|G_s(G_f(I_s), M_g), I_s^{Lab}\|_1] + \mathbb{E}_{I_f \sim p(I_f)} [\|G_f(G_s(I_f), M_r), I_f^{Lab}\|_1]$$

- **Adversarial Loss:** se enfoca en que la distribución de datos en las imágenes reales coincida con la distribución de datos en las imágenes generadas.

$$\mathcal{L}_{GAN}(G_s, G_f, D_s^{Lab}, D_f^{Lab}) = \mathcal{L}_{GAN}^s(G_s, D_s^{Lab}) + \mathcal{L}_{GAN}^f(G_f, D_f^{Lab})$$

$$\begin{aligned} \mathcal{L}_{GAN}(G_s, G_f, D_s^{Lab}, D_f^{Lab}) &= \mathbb{E}_{I_s^{Lab} \sim p(I_s)} [\log(D_s^{Lab}(I_s^{Lab}))] + \mathbb{E}_{I_f \sim p(I_f)} [\log(1 - D_s^{Lab}(G_s(I_f, M_r)))] \\ &+ \mathbb{E}_{I_f^{Lab} \sim p(I_f)} [\log(D_f^{Lab}(I_f^{Lab}))] + \mathbb{E}_{I_s \sim p(I_s)} [\log(1 - D_f^{Lab}(G_f(I_s)))] \end{aligned}$$

Para el segundo paso del entrenamiento, en el Módulo II, se usan las funciones de pérdidas descritas anteriormente junto a la siguiente función de pérdida:

- **Colour Loss:** se enfoca en que el color en $\tilde{I}_s^{Lab}$ e $\tilde{I}_f^{Lab}$ sea el mismo que en I_s^{Lab} e I_f^{Lab} respectivamente.

$$\mathcal{L}_{colour} = \mathcal{L}_{colour}^s(G_f, G_s) + \mathcal{L}_{colour}^f(G_s, G_f)$$

Por lo tanto, la función de pérdida es:

$$\mathcal{L}_{colour} = \sum_p (J - \cos\langle (G_s(G_f(I_s), M_g))_p, (I_s^{Lab})_p \rangle) + \sum_p (J - \cos\langle (G_f(G_s(I_f, M_r)))_p, (I_f^{Lab})_p \rangle)$$

Estas 4 funciones de pérdida se condensan en esta ecuación:

$$\mathcal{L}(G_s, G_f, D_s^{Lab}, D_f^{Lab}) = \omega_1 \mathcal{L}_{identity} + \omega_2 \mathcal{L}_{cycle} + \omega_3 \mathcal{L}_{GAN} + \omega_4 \mathcal{L}_{colour}$$

Donde cada peso ω es un hiperparámetro, los primeros tres pesos se ajustan en base a los pesos usados en *Mask-ShadowGAN* (Fu et al., 2019) y son 5, 10 y 1 respectivamente. En cuanto al peso para la función de pérdida propuesta, este peso toma el valor de 10 en base a dos consideraciones. Primero, el peso ω_4 tiene un rol similar a ω_2 , ya que ambos se enfocan en que $\tilde{I}_s^{Lab}$ e $\tilde{I}_f^{Lab}$ sean iguales a I_s^{Lab} e I_f^{Lab} respectivamente. Segundo, empíricamente al usar $\omega_4 = 10$ en *LG-ShadowNet*, este obtuvo un mejor rendimiento.

Este método se probó en los *datasets* ISTD e ISTD+, y se evaluó su rendimiento cuantitativamente con RMSE, SSIM y PSNR. Las pruebas duraron 5.6 minutos para ambos *datasets*. Este método se evaluó con *Mask-ShadowGAN*.

En el *dataset* ISTD, *LG-ShadowNet* obtuvo un RMSE de 6,67 para las imágenes en Lab completas, que superó el resultado obtenido por *Mask-ShadowGAN* en sus dos variantes de entrenamiento, con imágenes RGB obtuvo 7,41 y con imágenes Lab obtuvo 7,32. En cuanto al PSNR y SSIM superó los resultados de *Mask-ShadowGAN*, obteniendo 25,92 y 0,909 respectivamente. En el *dataset* ISTD+, *LG-ShadowNet* también superó a *Mask-ShadowGAN* en las tres métricas, para RMSE obtuvo 5,02, para PSNR obtuvo 28,07 y para SSIM obtuvo 0,920.

Otras ventajas del método son que al ser no supervisado permite una mayor recolección de datos, sus tiempos de ejecución son menores, requiere menos parámetros y sus operaciones son más rápidas que en *Mask-ShadowGAN*.

Aunque no se mencionen limitaciones explícitas, se puede entender que existe una dependencia entre el Módulo I y el Módulo II, como el primero sirve como guía de iluminación para el segundo, una estimación incorrecta o ambigua de la iluminación podría disminuir el rendimiento del Módulo II.

DMTN

DMTN es un método propuesto por Fan et al. (2023), el cual se basa en que la iluminación de las imágenes se rige por esta ecuación:

$$L = L_a + L_d$$

Donde L es la iluminación, L_a es la iluminación ambiental y L_d es la iluminación directa. El objetivo es restaurar L_d en aquellas regiones sombreadas, por lo que el método consiste en el mapeo de la imagen con sombra a una sin sombra considerando que existe un *matte* de la sombra:

$$I_f = I_s + I_m$$

El método en cuestión realiza la eliminación de sombra de una imagen I_s mediante el uso de mapas de características obtenidos de una red VGG19 preentrenada que son refinados y desacoplados para realizar tres tareas simultáneas que dan como resultado la imagen del *matte* de la sombra I_m , la imagen con sombra original I_s y la imagen sin sombra I_f . En la figura se muestra la arquitectura del método.

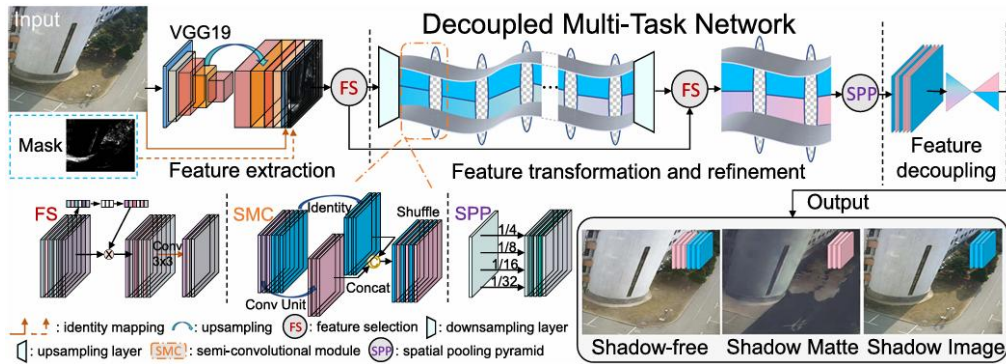


Figura 19: Arquitectura de DMTN.

Fuente: Fan et al. (2023).

Este método se compone de tres etapas principales; extracción de características, refinamiento y transformación de características, y desacoplamiento de características.

La primera etapa consiste en la extracción de características de la imagen usando una red VGG19 preentrenada. De esta etapa se extraen 64 mapas de características de la imagen de entrada I_s .

En la segunda etapa, algunas características son refinadas para la estimación del *matte* mientras que otras se conservan para la reconstrucción de la imagen con sombra. Para lograr esto, se aplica un módulo de semi convolución (SMC) que procesa la mitad de los mapas de características y conserva la otra mitad. Además, se utiliza una convolución adicional para modelar la compensación de iluminación directa y representar I_m .

En la tercera etapa, aún no se sabe qué mapas de características corresponden a cada imagen que se busca generar. Por ello, se aplica una capa de desacoplamiento restringida

(CFDL) para separar los mapas de características en base a pesos de tal manera que los mapas para la reconstrucción de la sombra y la construcción del *matte* sean independientes. En cuanto a la imagen sin sombra I_f , esta se representa como la suma de I_m y I_s .

Durante el entrenamiento, se tiene como entrada tanto la imagen con sombra I_s como la imagen sin sombra I_f , esta última es entrada para el discriminador D solamente. Se usaron las siguientes funciones de pérdida:

- **Función de pérdida para I_s :** se enfoca en que la imagen con sombra resultante $\tilde{I}_s$ sea idéntica a la imagen con sombra original I_s , la función es la siguiente:

$$\mathcal{L}_s = \mathbb{E}_{\tilde{I}_s, I_s} [\lambda_s \|\tilde{I}_s - I_s\|_1]$$

- **Función de pérdida para I_m :** se enfoca en que el matte de sombra resultante $\tilde{I}_m$ sea idéntico al *matte* original I_m , este último corresponde a la diferencia entre I_f e I_s , siendo la función:

$$\mathcal{L}_m = \mathbb{E}_{\tilde{I}_m, I_m} [\lambda_m \|\tilde{I}_m - I_m\|_1]$$

- **Función de pérdida para I_f :** se enfoca en que la imagen sin sombra resultante $\tilde{I}_f$ tenga una reconstrucción mucho más fidedigna a la imagen original I_f , para ello se usa ℓ_1 , $\mathcal{L}_{adv}$ y $\mathcal{L}_{perc}$ de la siguiente manera:

$$\mathcal{L}_f = \lambda_1 \mathcal{L}_{\ell_1} + \lambda_2 \mathcal{L}_{adv} + \lambda_3 \mathcal{L}_{perc}$$

- **ℓ_1 Loss:** se enfoca en que la imagen resultante logre capturar las frecuencias bajas, es decir, detalles como el fondo, formas grandes, entre otros. Se basa en la diferencia absoluta entre la imagen resultante $\tilde{I}_f$ y la imagen original I_f , a nivel de píxeles.
- **Adversarial Loss:** se enfoca en que la imagen resultante $\tilde{I}_f$ sea lo más real posible en cuanto a detalles finos, por lo que se entrega a un discriminador D , el cual compara la imagen resultante con la imagen original I_f . Esta es su función de pérdida:

$$\mathcal{L}_{adv} = \mathbb{E}_{I_f, I_s} [\log[D(I_f, I_s)]] + \mathbb{E}_{\tilde{I}_f, I_s} [\log[1 - D(\tilde{I}_f, I_s)]]$$

- **Perceptual Loss:** usando la red preentrenada VGG19, especializada en reconocimiento de imágenes, se espera que la imagen resultante $\tilde{I}_f$ logre activar de manera similar las capas de la red a como lo hace la imagen original I_f . La función de pérdida es la siguiente:

$$\mathcal{L}_{perc} = \sum_{k=1}^5 \mathbb{E}_{\tilde{I}_f, I_f} [\|\Phi_k(\tilde{I}_f) - \Phi_k(I_f)\|_1]$$

El objetivo de este método es que el generador G produzca imágenes reales que, en el caso de la imagen sin sombra $\tilde{I}_f$, puedan engañar al discriminador D , lo que se expresa de esta manera:

$$\arg \min_G \max_D (\mathcal{L}_s + \mathcal{L}_m + \mathcal{L}_f)$$

En cuanto a los parámetros λ se definen empíricamente como $\lambda_s = 2550$, $\lambda_m = 255$, $\lambda_1 = 2550$, $\lambda_2 = 1$ y $\lambda_3 = 1000$.

Este método se evaluó en los *datasets* ISTD, ISTD+ y SRD, y los resultados se evaluaron con las métricas RMSE, SSIM y PSNR. Se comparó con otros métodos como los mencionados anteriormente *DC-ShadowNet* y *LG-ShadowNet*.

En el *dataset* ISTD, DMTN logra reducir el RMSE en 1,24 en la imagen completa obteniendo 5,43 comparado al 6,67 que obtuvo *LG-ShadowNet*. Además, aumentó el SSIM en 0,022 y el PSNR en 2,42, demostrando que las imágenes resultantes se asemejan más a las originales. En el *dataset* ISTD+, *DC-ShadowNet* obtuvo un RMSE de 4,6 y *LG-ShadowNet* obtuvo un RMSE de 4,4. DMTN superó ambos métodos con un RMSE de 3,7 en la imagen completa. En el *dataset* SRD, DMTN obtuvo un RMSE de 4,06 que superó el RMSE obtenido por *DC-ShadowNet*. DMTN es capaz de producir imágenes que cuantitativamente son superiores y más cercanas a la imagen real esperada que los métodos anteriores.

En cuanto a las limitaciones del método, en algunos casos como en el *dataset* ISTD, se presentó un cambio en los valores de píxeles de regiones no sombreadas debido a una inconsistencia en la localización de las regiones sombreadas que se puede mejorar usando una máscara de sombra como entrada adicional. En el *dataset* SRD, hubo casos donde el color en regiones sombreadas fue restaurado incorrectamente, se piensa que es por una mala distribución en el *dataset*, por lo que estos casos se presentaron menos en el entrenamiento.

TBRNet

TBRNet es otro método propuesto por Fan et al. (2023), donde se ejecutan tres tareas en tres ramas paralelas las cuales comparten información entre sí y que se basan en un modelo de iluminación lineal, cuya ecuación es:

$$I_f = I_s + I_m$$

Donde I_f es la imagen sin sombra, I_s es la imagen original con sombra y I_m es el *matte* de la sombra. En la figura se muestra la arquitectura del método.

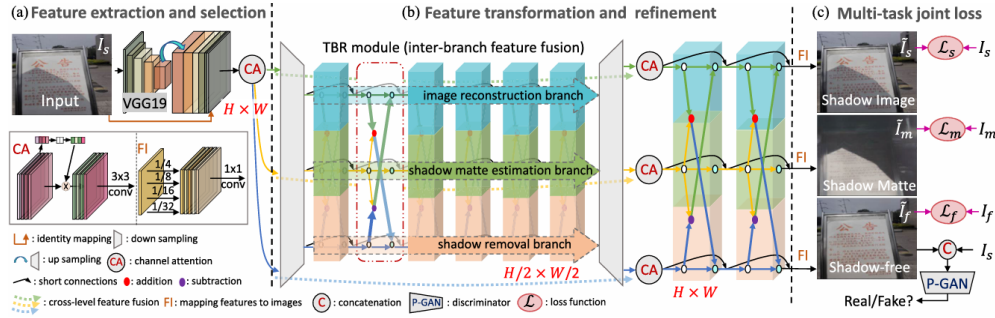


Figura 20: Arquitectura de TBRNet.

Fuente: Fan et al. (2023).

Este método, al igual que *DMTN*, consta de tres etapas principales: extracción de características, refinamiento y transformación con desacoplamiento de tareas. En la primera etapa, se extraen características de hipercolumnas utilizando una red VGG19 preentrenada. Estas características son refinadas mediante mecanismos de atención, lo que permite seleccionar información relevante para cada tarea.

La primera etapa corresponde a la extracción de características de hipercolumnas de VGG19 y se seleccionan para la siguiente etapa mediante mecanismos de atención. En la segunda etapa se implementa un módulo residual de tres ramas (TBR), como el nombre lo indica, se compone de tres ramas, cada una encargada de una tarea particular; reconstrucción de la imagen original, estimación del *matte* de sombra y eliminación de sombra. Estas ramas comparten información entre sí mediante sus mapas de características para colaborar de manera conjunta. El intercambio de mapas de características se impone bajo estas ecuaciones:

$$F_s = F_f - F_m$$

$$F_f = F_s + F_m$$

El intercambio y fusión de información entre las ramas se mejora mediante el uso de *Patch-Global Attention* (PGA), ya que resalta las regiones importantes en las que tengan que cooperar las ramas, filtra aquella información de regiones sin sombra o con poca variación.

La rama encargada de la estimación del *matte* de sombra provee información de la ubicación e intensidad de la sombra a la rama encargada de la eliminación. En la figura se muestra en detalle cómo funciona el módulo TBR.

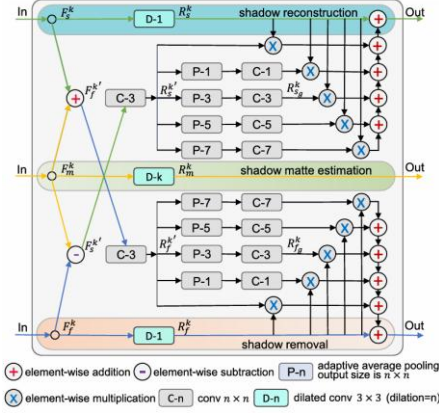


Figura 21: Módulo TBR.

Fuente: Fan et al. (2023).

Para el entrenamiento, algunas de las funciones de pérdida son similares a las presentadas en *DMTN*, como $\mathcal{L}_s$, $\mathcal{L}_m$, $\mathcal{L}_f$ y $\mathcal{L}_{perc}$. Sin embargo, las funciones de pérdida que componen a $\mathcal{L}_f$ difieren y se incluyen otras:

- **ℓ_1 Loss:**

$$\mathcal{L}_{\ell_1} = \mathbb{E}_{\tilde{I}_f, I_f} [\|\tilde{I}_f - I_f\|_1]$$

- **Adversarial Loss:**

$$\mathcal{L}_{adv} = \mathbb{E}_{\tilde{I}_f, I_s} [\log[1 - D(\tilde{I}_f, I_s)]]$$

- **Función de pérdida de generador G :**

$$\mathcal{L}_G = \mathcal{L}_s + \mathcal{L}_m + \mathcal{L}_f$$

- **Función de pérdida de discriminador D :**

$$\mathcal{L}_D = -\mathbb{E}_{I_f, I_s} [\log[D(I_f, I_s)]] - \mathbb{E}_{\tilde{I}_f, I_s} [\log[1 - D(\tilde{I}_f, I_s)]]$$

El objetivo del entrenamiento se expresa en la siguiente función:

$$\arg \min_{\theta_G, \theta_D} (\mathcal{L}_G + \mathcal{L}_D)$$

Donde θ_G y θ_D son parámetros del generador y discriminador respectivamente. Los parámetros λ se definen empíricamente como $\lambda_s = 2550$, $\lambda_m = 255$, $\lambda_1 = 2550$, $\lambda_2 = 1$ y $\lambda_3 = 1000$.

Se realizaron pruebas en los *datasets* ISTD, ISTD+ y SRD, y el rendimiento se evaluó cuantitativamente con RMSE, PSNR y SSIM. En el *dataset* ISTD+, el método fue comparado con *DC-ShadowNet*.

En el *dataset* ISTD, *TBRNet* obtuvo un RMSE de 5,59 en la imagen completa, un SSIM de 0,986 y un PSNR de 35,56, destacando como el mejor en las tres métricas. En el *dataset* ISTD+, obtuvo un RMSE en la imagen completa de 3,8 que supera al obtenido por *DC-ShadowNet* que fue de 4,6. En el *dataset* SRD, obtuvo un RMSE de 5,62 siendo el mejor resultado al comparar con otros métodos.

Respecto a las limitaciones, aunque tenga mejores resultados cuantitativos, *TBRNet* es más lento que otros métodos y presenta problemas al momento de eliminar *hard shadows* muy oscuras y para *soft shadows* existen casos en que no las elimina o compensa parcialmente.

2.3.2.2 DATASETS DE ENTRENAMIENTO

En el estudio de Guo et al. (2024) se describen los *datasets* utilizados por los métodos para el entrenamiento y las pruebas. De estos se destaca el uso de ISTD, ISTD+ y SRD.

ISTD: El *dataset* ISTD se compone de 1330 tripletes de entrenamiento y 540 para las pruebas, los cuales consisten en imágenes con sombra, su máscara etiquetada manualmente y las imágenes sin sombra. Este *dataset* contiene principalmente escenas en el exterior y predominan las sombras fuertes. Las sombras son generadas por varios objetos que no están presentes en la escena, y para obtener la correspondiente imagen sin sombra, la fuente de la sombra se quita. (Guo et al., 2024).

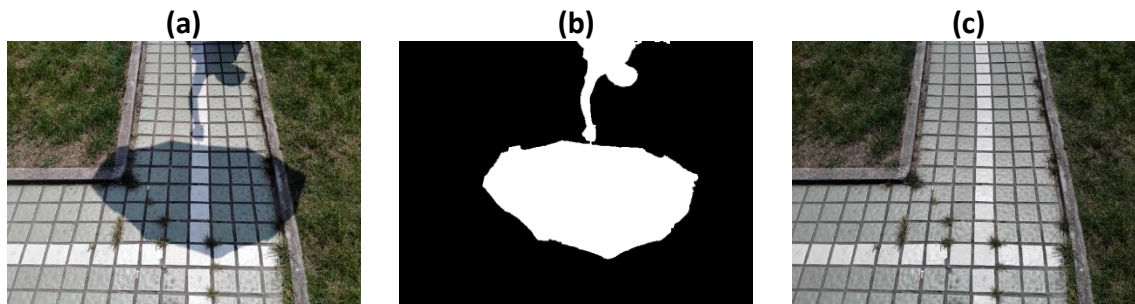


Figura 22: Ejemplo de *dataset* ISTD: (a): imagen con sombra, (b): máscara binaria y (c): imagen sin sombra.

Fuente: Elaboración Propia.

ISTD+: El *dataset Adjusted* ISTD (ISTD+) mejora las inconsistencias de iluminación entre las imágenes con sombra y sin sombra presentes en el *dataset* ISTD original. (Guo et al., 2024).

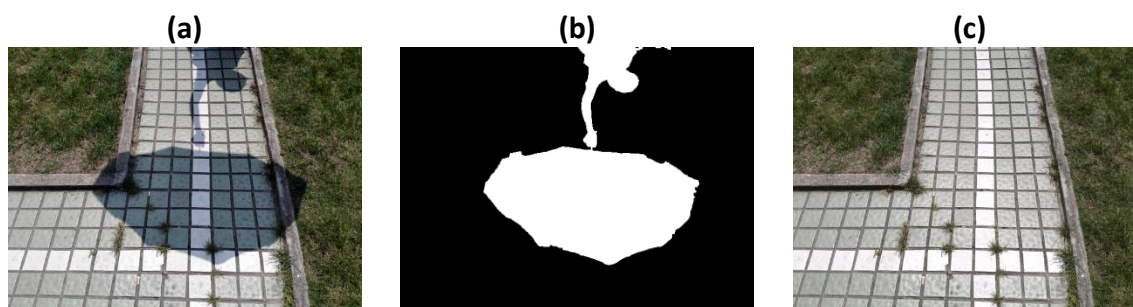


Figura 23: Ejemplo de dataset ISTD+: (a): imagen con sombra, (b): máscara binaria y (c): imagen sin sombra con iluminación ajustada.

Fuente: Elaboración Propia.

SRD: El *dataset* SRD se compone de 2680 pares de entrenamiento y 408 para las pruebas, los cuales consisten en la imagen con sombra y la imagen sin sombra. Este *dataset* no contiene máscara etiquetadas manualmente. Al igual que en ISTD, incluye sombras generadas por varios objetos y para obtener la imagen sin sombra se quita la fuente que la produce. Las imágenes con sombra se obteniendo bajo diferentes condiciones de iluminación, tales como condiciones climáticas o diferentes horas del día, para incorporar *hard shadows* y *soft shadows* en el *dataset*. (Guo et al., 2024).



Figura 24: Ejemplo de *dataset* SRD: (a): imagen con sombra y (b): imagen sin sombra.

Fuente: Elaboración Propia.

2.3.2.3 MÉTRICAS DE EVALUACIÓN

Root-Mean-Square Error (RMSE): Mide la diferencia promedio entre los valores de píxel de la imagen generada y la imagen de referencia. Un valor más bajo indica una mayor similitud. Esta métrica se expresa en unidades de nivel de intensidad de píxel y es especialmente sensible a errores grandes en regiones pequeñas. (Guo et al., 2024).

Peak Signal-to-Noise Ratio (PSNR): Calcula la relación entre la señal (imagen original) y el ruido (errores introducidos en la imagen generada). Se mide en decibelios (dB), y valores

más altos indican mejor calidad perceptual y menor distorsión visual. Es una métrica útil para evaluar qué tan "limpia" es una imagen generada respecto a su versión de referencia. (Guo et al., 2024).

Structural Similarity Index Measure (SSIM): Evalúa la similitud estructural entre dos imágenes, considerando no solo la luminancia, sino también el contraste y la estructura local. Su valor varía entre 0 y 1, donde 1 representa una similitud estructural perfecta. Esta métrica busca reflejar la percepción humana más allá de diferencias puramente numéricas. (Guo et al., 2024).

2.3.2.4 RESULTADOS CUANTITATIVOS

A continuación, se presentan los resultados obtenidos por método con las métricas de evaluación para cada *dataset*. Las flechas junto a la métrica indica si los valores son mejores cuando son más pequeños (↓) o mejores cuando son más grandes (↑) (Fan et al., 2023).

Tabla 5: Resultados cuantitativos en las pruebas del *dataset* ISTD.

Fuente: Elaboración Propia.

Método	RMSE (↓)	PSNR (↑)	SSIM (↑)
<i>DC-ShadowNet</i>	–	–	–
<i>LG-ShadowNet</i>	6,67	25,92	0,909
<i>DMTN</i>	5,43	29,04	0,958
<i>TBRNet</i>			

Tabla 6: Resultados cuantitativos en las pruebas del *dataset* ISTD+.

Fuente: Elaboración Propia.

Método	RMSE (↓)	PSNR (↑)	SSIM (↑)
<i>DC-ShadowNet</i>	4,60	–	–
<i>LG-ShadowNet</i>	5,02	28,07	0,920
<i>DMTN</i>	3,70	–	–
<i>TBRNet</i>			

Tabla 7: Resultados cuantitativos en las pruebas del *dataset* SRD.

Fuente: Elaboración Propia.

Método	RMSE (↓)	PSNR (↑)	SSIM (↑)
<i>DC-ShadowNet</i>	4,66	–	–
<i>LG-ShadowNet</i>	–	–	–
<i>DMTN</i>	4,06	–	–
<i>TBRNet</i>			

TRATAMIENTO DE SOMBRAS PARA DETECTAR DETERIOROS EN PAVIMENTOS CON MACHINE LEARNING

En la tabla 8 se resume la información de los métodos estudiados, tanto los métodos de procesamiento de imágenes como los métodos de *Deep Learning*.

Tabla 8: Resumen de los métodos estudiados.

Fuente: Elaboración Propia.

Método	Datos utilizados	Color	Técnicas/Arquitectura	Supervisión	Métricas
MCA (<i>Morphological Component Analysis</i>)	Imágenes de pavimentos urbanos	Gris	MCA usando <i>Curvelet</i> (estructura) y LDCT (textura), y separación morfológica.	-	No se especifica
Equalización local color-textura	Imágenes aéreas (UAV)	RGB	Segmentación por regiones homogéneas y compensación de brillo y textura local.	-	CD, SSID, GS
<i>Multi-Colour-Space</i> + HSV-RGB	Imágenes de caminos (carreteras)	RGB	Espacios de color HSV, RGB y YIQ para segmentar y restaurar regiones con sombras.	-	Matriz de confusión
Filtro de mediana selectivo y <i>Black-top-hat</i>	Imágenes de documentos	RGB	Filtro de mediana selectiva y transformación <i>black-top-hat</i> para fondo homogéneo.	-	MSE, MSSIM
Compensación de brillo local	Imágenes de pavimentos urbanos	Gris	Compensación de brillo en regiones sombreadas mediante niveles de brillo y valores medios de píxeles.	-	No se especifica
<i>DC-ShadowNet</i>	ISTD, ISTD+, SRD	RGB	<i>Encoder-decoder</i> , bloques de atención, clasificador de dominio, GAN.	No supervisado	
<i>LG-ShadowNet</i>	ISTD, ISTD+	RGB	Red basada en <i>Mask-ShadowGAN</i> , guía de iluminancia, discriminadores tipo <i>PatchGAN</i> .	No supervisado	
<i>DMTN</i>	ISTD, ISTD+, SRD	RGB	Red basada en VGG19, módulos SMC y CDFL, GAN.	Supervisado	
<i>TBRNet</i>	ISTD, ISTD+, SRD	RGB	Red multitarea con 3 ramas, módulo TBR, GAN.	Supervisado	

CAPÍTULO 3: PROPUESTA DE SOLUCION

3.1 MÉTODOS PROBADOS

De todos los métodos estudiados se evaluaron 2 de procesamiento de imágenes y los 4 de *Deep Learning*. La elección de los métodos consistió en lo siguiente:

Para los métodos basados en procesamiento de imágenes, se buscó aquellos métodos que se enfocasen en mantener detalles oscuros pequeños, debido a que el objetivo del tratamiento de sombras en imágenes de pavimentos es remover las sombras, pero conservar las grietas que potencialmente se encuentran ocultas bajo las sombras. Bajo esta lógica, de mantener detalles oscuros como grietas luego de la eliminación de sombras, el método desarrollado por Imahayashi et al. (2023) y Bi et al. (2024). Aunque Imahayashi et al. (2023) no buscan conservar grietas sino caracteres de un documento, esto cumple con características similares a la de imágenes de pavimentos, detalles oscuros que puede ser opacados por una sombra y que deben ser conservados. Mientras que Bi et al. (2024) proponen su método basado en el uso de compensación de brillo para ser usado en imágenes de pavimentos que sufren la misma problemática de sombras y grietas.

Para los métodos basados en *Deep Learning*, el criterio fue distinto. Muchos métodos necesitan la imagen original junto a una máscara binaria que indica la ubicación de la sombra dentro de la imagen (Guo et al., 2024). Esa información no se posee, existiendo sólo las imágenes de pavimento con sombra, por lo tanto, se buscaron aquellos métodos cuyo input para realizar la inferencia fuese sólo la imagen original. Bajo ese contexto, los métodos seleccionados son *DC-ShadowNet* (Jin et al., 2021), *LG-ShadowNet* (Liu et al., 2021), *DMTN* (Fan et al., 2023) y *TBRNet* (Fan et al., 2023), que destacan además por ser dos ejemplos de entrenamiento no supervisado y supervisado.

3.2 DATASETS DE PRUEBA

Para las pruebas se crearon tres *datasets* en base a imágenes del proyecto, estos representan etapas diferentes de la metodología, como la extracción de las imágenes desde los videos con y sin filtro, y la unión de estas en secciones largas.

Dataset 1: Corresponde a 41 imágenes en formato RGB de 2.105 x 1.945 píxeles que contienen grietas, *soft shadows* y *hard shadows*, en estas imágenes no se aplica el filtro.

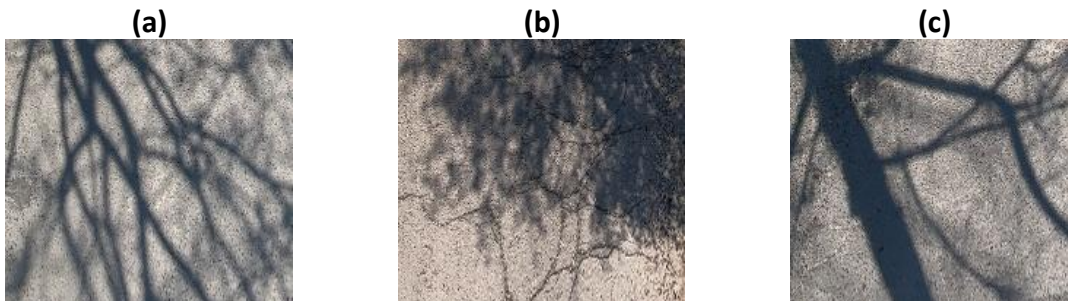


Figura 25: Ejemplos de *dataset 1*: (a): sólo sombras, (b): grieta parcialmente oculta y (c): grieta completamente oculta.

Fuente: Elaboración Propia.

Dataset 2: Contiene las mismas imágenes que en el *dataset 1*, pero el contraste entre luz y sombra está disminuido con el uso de un filtro ND32.

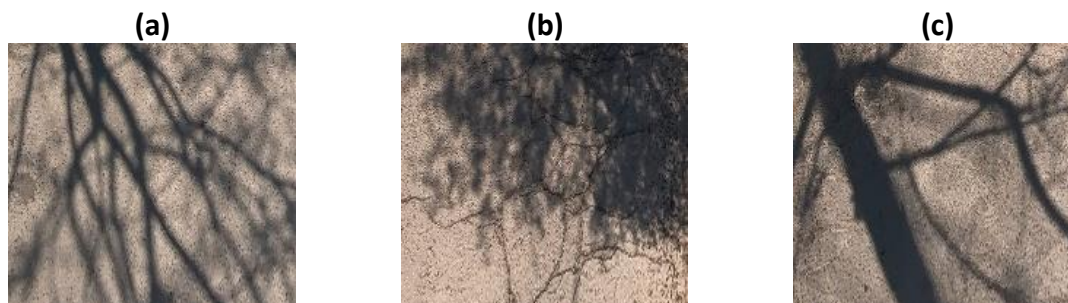


Figura 26: Ejemplos de *dataset 2*: (a): sólo sombras, (b): grieta parcialmente oculta y (c): grieta completamente oculta.

Fuente: Elaboración Propia.

Dataset 3: Corresponde a 17 imágenes en escala de gris de 2.105 x 9.725 píxeles que contienen grietas y *soft shadows*. Ambos corresponden a etapas diferentes de la metodología, de manera que se pueda analizar en cuál etapa es más eficiente usar los métodos para obtener mejores resultados. Un ejemplo del *dataset* está en la figura 7.

3.3 DETALLES DE IMPLEMENTACIÓN

Los métodos de procesamiento de imágenes se evaluaron en base a una implementación de estos usando *Python* en la versión gratuita de *Google Colab*. Para los métodos de *Deep Learning*, se usaron los modelos preentrenados disponibles para observar su rendimiento basado en el entrenamiento que se les hizo con los *datasets* ISTD, ISTD+ y SRD, y se ejecutaron de manera local con un procesador *Intel(R) Core(TM) i5-9300H*, una tarjeta gráfica *NVIDIA GeForce GTX 1650* y con 16 GB de RAM.

Por último, se probó el método de detección de deterioros, YOLOv5, con los resultados obtenidos de cada método y se comparó con las imágenes originales. De esta manera, medimos la eficiencia de cada método probado al incorporarse en la metodología. Para esta implementación también se utilizó *Python* en la versión gratuita de *Google Colab*, ya que sólo se hicieron inferencias en la red.

Se registró el tiempo de ejecución en cada método, en específico el tiempo de CPU para tener certeza de la cantidad de tiempo que utiliza la ejecución de cada método. Como no existen imágenes sin sombra originales, o *ground truths*, se realizó un análisis cualitativo en las imágenes, el cual se basa en los siguientes aspectos:

- Grado de eliminación de la sombra.

- Preservación de color, textura y resolución.
- Presencia de artefactos o errores visuales.

En el análisis cualitativo se tomaron los siguientes casos como ejemplos:

- Imagen con sombra solamente.
- Imagen con grietas parcialmente ocultas por sombras
- Imagen con grietas completamente ocultas por sombras.

CAPÍTULO 4: VALIDACIÓN DE LA SOLUCIÓN

4.1 RESULTADOS DE TRATAMIENTO DE SOMBRA

Para los primeros *datasets* se presentan tres imágenes, mostrando los resultados obtenidos en cada método que tuvo éxito. Mientras que para el último *dataset*, se presenta aquella sección que contenga las tres situaciones. Por comodidad, los métodos probados se abreviarán como:

- (1): Imahayashi et al.
- (2): Bi et al.
- (3): *DC-ShadowNet* con ISTD
- (4): *DC-ShadowNet* con AISTD
- (5): *LG-ShadowNet* con AISTD
- (6): *DC-ShadowNet* con SRD

Cabe resaltar que los métodos propuestos por Fan et al. (2023), *DMTN* y *TBRNet*, aunque se encuentran disponibles, no dieron resultados registrables por problemas en la ejecución del código. Se especula que el error fue causado por la falta de los pesos de la red preentrenada VGG19.

Se puede observar que, en general, todos los métodos afectan la imagen completa. En los métodos de procesamiento de imágenes (1) y (2), se aprecia un cambio de color general en la imagen, las sombras en cuestión son removidas, pero las grietas son parcial o completamente removidas también, por último, la resolución de la imagen se conserva. Mientras que en los métodos basados en *Deep Learning* (3), (4), (5) y (6), la sombra se atenúa, cambiando su color y textura con respecto al resto de la imagen, pero no se elimina, ya que a veces se conserva su silueta, y las grietas se conservan completa o parcialmente, en cuanto a la resolución de la imagen, esta siempre es reducida quedando como 256 x 256 para (3), (4) y (6), y 640 x 480 para (5).

4.1.1 RESULTADOS DE DATASET 1

A continuación, se muestra una figura con ejemplos del *dataset* 1 y luego se hace un análisis cualitativo detallado por cada ejemplo.

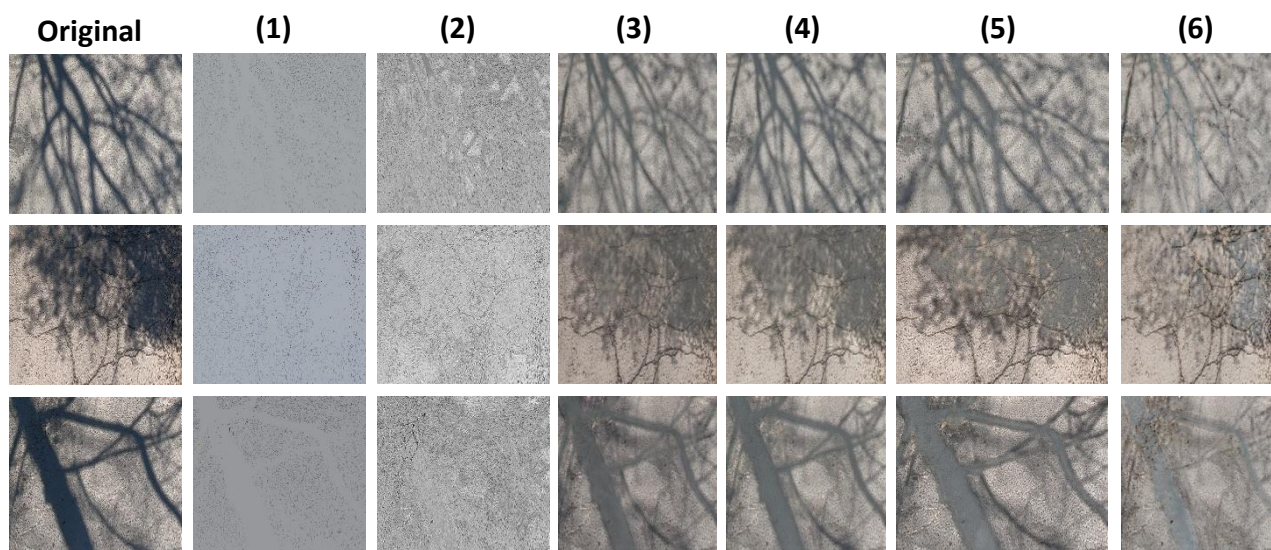


Figura 27: Resultados en *dataset 1*.

Fuente: Elaboración Propia.

4.1.1.1 ANÁLISIS CUALITATIVO DE EJEMPLO 1 (SÓLO SOMBRA)

Con (1), la sombra es eliminada completamente, sin embargo, se conserva la silueta de esta y queda sobreexpuesta. La textura del pavimento de fondo se conserva parcialmente, pero se perciben claros errores en la silueta remanente de la sombra en que la textura no es igual a la del fondo. La imagen no pierde resolución, pero hay un exceso de tonos claros, y con la pérdida de textura, la imagen a simple vista deja de tener sentido, no se percibe como pavimento.

Con (2), el primer cambio notable es que el color de la imagen cambia y pasa a estar en escala de gris, debido a que el método procesa las imágenes en este espacio de color. La sombra se elimina casi en su totalidad, pero el brillo no se iguala al del fondo siendo más oscura. Se perciben artefactos visuales, hay manchas claras en aquellas partes donde originalmente no había sombra y existen halos alrededor de la sombra que delatan que la imagen fue procesada. La textura del pavimento se conserva mejor y logra reconstruirse parcialmente en la región sombreada, la resolución de la imagen se conserva y presenta la misma calidad que la original.

Con (3), no se elimina la sombra, aunque se aclara bastante sigue siendo reconocible. Como la imagen disminuye su resolución, se detectan pocos artefactos en forma de píxeles mal distribuidos dentro de la región sombreada, la imagen se ve más borrosa y se pierde un poco la textura del pavimento y no se restaura en la región sombreada.

Con (4), no se elimina la sombra, incluso se aclara menos que en (3). Sin embargo, no se aprecian artefactos en la imagen. Como la resolución de la imagen disminuye, la textura de pavimento se difumina y no se restaura en la región sombreada.

Con (5), no se logra eliminar la sombra, sólo se aclaran las regiones más oscurecidas. Hay artefactos de píxeles en forma de tablero de ajedrez dentro de la sombra. Aunque la

resolución baja, la textura del pavimento se conserva en el fondo y logra rescatarse mejor en la región sombreada. El color y la iluminación de las regiones sin sombra se conservan.

Con (6), no se elimina la sombra, pero se reduce bastante pues se afina, su iluminación en ciertas regiones es exagerada. Hay artefactos como pequeñas manchas claras dentro de la sombra que resaltan incluso en las partes de mayor luminosidad. La resolución disminuye, se genera una pérdida de la textura del pavimento y la sombra puede confundirse con una grieta, debido a que los detalles más finos y oscuros de esta no son procesados.

4.1.1.2 ANÁLISIS CUALITATIVO DE EJEMPLO 2 (GRIETA PARCIALMENTE OCULTA)

Con (1), la sombra es eliminada por completo, sólo es visible su silueta como manchas de color uniforme, por lo que la textura de pavimento en esas regiones no fue restaurada. La grieta se elimina casi por completo, la parte que no está opacada se pierde mientras que los segmentos más oscuros de la grieta se conservan. A simple vista, la grieta parecer ser más fina. Como se conserva la resolución, la textura de pavimento se conserva en el fondo y no se ve borrosa.

Con (2), la sombra se elimina, pero en los límites con la región no sombreada, se percibe un cambio de color. Además, hay artefactos en forma de halo muy finos que rodean la sombra. La textura de pavimento se conserva en el fondo y se logra restaurar en la región sombreada. La grieta se conserva parcialmente, los segmentos más oscuros se conservan mientras que en su parte no opacada se confunden con la textura del pavimento. La resolución se conserva.

Con (3) y (4), la sombra no es eliminada, sólo son tratadas las regiones más oscuras de esta. Hay pocos artefactos visibles. La textura de pavimento es más borrosa por la disminución de resolución, pero aún se aprecia la sombra remanente. La grieta se conserva tanto en la parte no opacada como la que está bajo la sombra.

Con (5), no se elimina la sombra, las regiones más oscuras son tratadas mientras que las regiones al borde no se alteran. Casi no hay artefactos a simple vista. La textura del pavimento se conserva mejor tanto en el fondo como bajo la sombra, incluso con la disminución de resolución. La grieta se conserva bien y su parte oculta es mucho más visible.

Con (6), se elimina aún más la sombra, dejando rastros de esta con una tonalidad distinta al fondo. La textura del pavimento se pierde un poco volviéndose más borrosa por la disminución de resolución. La grieta se conserva, sin embargo, los segmentos más oscuros se sobreexponen al igual que otros detalles oscuros propios de la superficie rugosa del pavimento.

4.1.1.3 ANÁLISIS CUALITATIVO DE EJEMPLO 3 (GRIETA OCULTA)

Con (1), se elimina la sombra completamente, pero su silueta es visible debido a que no mantiene la textura del pavimento, que es parcialmente eliminada en el fondo también, aun cuando se conserva la resolución. Por ende, la silueta de la sombra se percibe como

una mancha de color uniforme. La grieta oculta en la sombra no se conserva, sólo quedan rastros de sus segmentos más oscuros.

Con (2), la sombra se elimina, pero se generan artefactos en forma de halos muy finos que rodean la región tratada por lo que delatan la antigua presencia de la sombra. La textura del pavimento se conserva en el fondo, pero no en la sombra, dónde se aprecia una restauración errónea de la textura que no coincide con el fondo. La grieta se conserva bien.

Con (3), no se elimina la sombra, sólo se aclaran las regiones más oscuras y el borde de la sombra es más visible. Existen artefactos como píxeles raros en la sombra. La textura del pavimento se conserva, pero es más difusa por la disminución de la resolución, y en la sombra no se logra restaurar. La grieta se pierde casi por completo, dejando sólo unos rastros de los segmentos más oscuros de esta.

Con (4), no se elimina la sombra, aclara las regiones oscuras y elimina las partes más tenues de la sombra, pero aún es visible. Existen artefactos como píxeles raros en la sombra. La textura del pavimento se conserva, pero es más difusa por la disminución de la resolución, y en la sombra no se logra restaurar. La grieta se pierde casi por completo, dejando sólo unos rastros de los segmentos más oscuros de esta.

Con (5), la sombra se mantiene, se aclara un poco su parte interna, pero los bordes siguen oscurecidos, por lo que se sigue viendo. Hay artefactos en forma de tablero de ajedrez en la sombra tratada. La textura de pavimento se conserva mejor incluso con la disminución de nitidez, pero en la región sombreada no se restaura por completo. La grieta se conserva parcialmente, es visible, pero se atenúa un poco.

Con (6), la sombra se elimina casi por completo, dejando rastros de una tonalidad distinta al fondo y las regiones más tenues no son alteradas. La textura del pavimento es más borrosa por la disminución de la resolución y no se restaura en la región sombreada. La grieta se conserva, pero sufre una sobreexposición por lo que se resaltan los detalles más oscuros de esta.

4.1.1.4 ANÁLISIS DE ASPECTOS TÉCNICOS

De los métodos de procesamiento de imágenes, el más rápido fue el método de Imahayashi et al. (2023), ambos métodos fueron capaces de entregar un resultado y conservar la resolución de la imagen. De los métodos de *Deep Learning*, el más rápido fue *DC-ShadowNet* preentrenado con el *dataset* SRD, aunque tuvo éxito, no pudo conservar la resolución de la imagen. Los métodos *DMTN* y *TBRNet* no tuvieron éxito en las pruebas y se especula que fue por la falta de pesos para la red VGG19 preentrenada.

En la tabla 9 están los rendimientos de cada método en términos del *dataset* con que se entrenó (para métodos de procesamiento de imágenes no aplica), el tiempo en segundos, el éxito al entregar un resultado y si conserva la resolución original o no. Las filas celestes representan los métodos por procesamiento de imágenes y las filas naranjas son los

TRATAMIENTO DE SOMBRAS PARA DETECTAR DETERIOROS EN PAVIMENTOS CON MACHINE LEARNING

métodos basados en *Deep Learning*. Se destacan en verde los mejores tiempos en cada tipo de método.

Tabla 9: Rendimiento de cada método probado en *dataset 1*.

Fuente: Elaboración Propia.

Método	<i>Dataset</i> de entrenamiento	Tiempo [s]	Resultado	Resolución
Imahayashi et al.	N/A	90	✓	✓
Bi et al.	N/A	242	✓	✓
<i>DC-ShadowNet</i>	ISTD	39	✓	×
<i>DMTN</i>		-	×	N/A
<i>TBRNet</i>		-	×	N/A
<i>DC-ShadowNet</i>	AISTD	36	✓	×
<i>LG-ShadowNet</i>		281	✓	×
<i>DMTN</i>		-	×	N/A
<i>TBRNet</i>		-	×	N/A
<i>DC-ShadowNet</i>	SRD	35	✓	×
<i>DMTN</i>		-	×	N/A
<i>TBRNet</i>		-	×	N/A

4.1.2 RESULTADOS DE DATASET 2

A continuación, se muestra una figura con ejemplos del *dataset 2* y luego se hace un análisis cualitativo detallado por cada ejemplo.

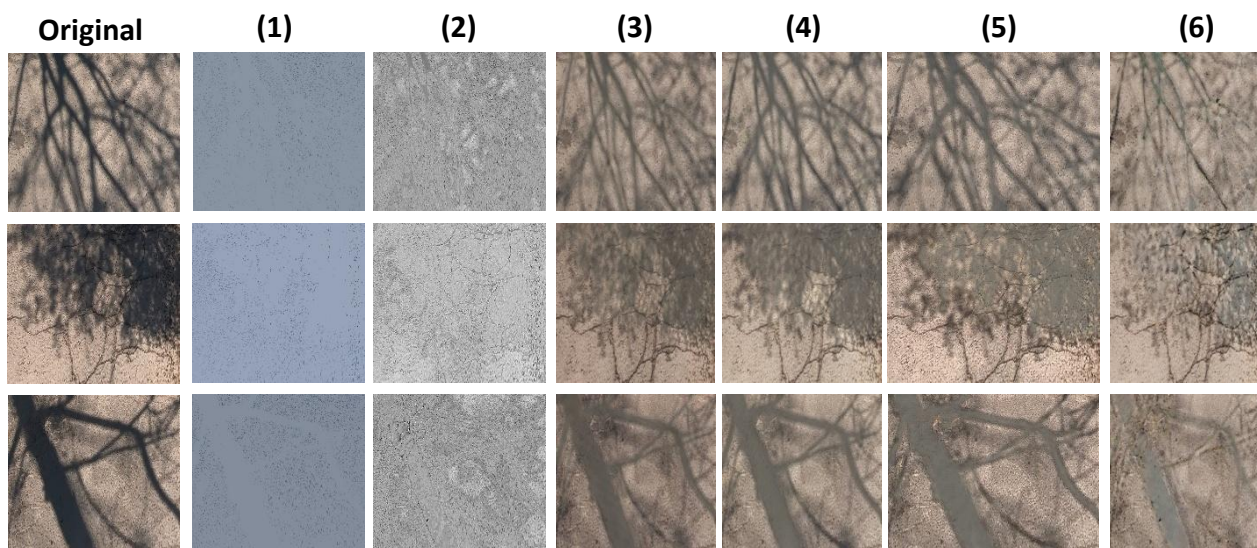


Figura 28: Resultados en *dataset 2*.

Fuente: Elaboración Propia.

4.1.2.1 ANÁLISIS CUALITATIVO DE EJEMPLO 1 (SÓLO SOMBRA)

Con (1), se elimina la sombra completamente, pero la silueta de esta se mantiene, siendo más evidente debido a la diferencia de textura con el fondo. La textura del pavimento se conserva parcialmente y el color cambia bastante. Por otro lado, la resolución de la imagen se mantiene.

Con (2), la sombra es eliminada, pero hay artefactos en formas de halos muy finos que rodean la silueta de la sombra. Además, en el interior de la sombra no se restaura la textura de pavimento, haciendo que la región que fue tratada sea más evidente a simple vista. Se perciben que regiones que no estaban con sombra aumentaron su brillo, lo que no se ajusta al brillo en otras regiones de la imagen.

Con (3) y (4), la sombra no es eliminada, sólo se atenúa y se aprecian muchos artefactos pequeños como píxeles con una distribución extraña. Debido a la disminución de la resolución, la textura del pavimento se difumina y no es restaurada en la región sombreada.

Con (5), la sombra no se elimina, sólo se atenúan las zonas más oscuras e internas de esta, bordes o regiones más suaves de la sombra no se tratan. Existe artefactos con forma de tablero de ajedrez en toda la región tratada. Aunque la resolución disminuye, la textura del pavimento se conserva relativamente bien, pero no se restaura en la sombra.

Con (6), la sombra se elimina casi por completo, quedan rastros de las regiones más oscuras y las más claras quedan con un leve cambio de tonalidad. No hay artefactos visibles. La textura del pavimento se difumina por la disminución de la resolución.

4.1.2.2 ANÁLISIS CUALITATIVO DE EJEMPLO 2 (GRIETA PARCIALMENTE OCULTA)

Con (1), se elimina la sombra, pero queda su silueta, la cual tiene una textura diferente a la del fondo. La textura del fondo se mantiene parcialmente, pero la silueta queda como una mancha de color uniforme. La grieta se elimina casi por completo tanto la parte que está oculta como la que no lo está. El color de la imagen en general cambia, pero se conserva la resolución.

Con (2), la sombra se elimina, no obstante, quedan pequeños artefactos como halos muy finos que rodean la región tratada y algunas zonas pequeñas en el límite por fuera de la sombra se oscurecen. La textura del pavimento se conserva bien en toda la imagen al igual que la grieta, que se mantiene casi por completo.

Con (3) y (4), la sombra no es eliminada, sólo se aclaran las zonas oscuras y las zonas más claras no se retocan. Se ven algunos artefactos como píxeles con una distribución extraña. La textura del pavimento se difumina, pero se puede apreciar un poco más en la región sombreada remanente. La grieta se ve en la sombra atenuada y se conserva en la región no sombreada. Además, (4) aumenta ligeramente el brillo general en la imagen.

Con (5), no se elimina la sombra, la parte más oscura de la sombra es tratada, las regiones más claras se conservan. Se ven algunos artefactos con forma de tablero de ajedrez. La textura del pavimento se mantiene bien, aun con la disminución de resolución, y en la sombra remanente se percibe ligeramente la textura. La grieta se mantiene y es reconocible en la sombra y fuera de ella.

Con (6), la eliminación de la sombra es casi completa, pero quedan unos pequeños rastros de una tonalidad diferente. La textura del pavimento se difumina, por la disminución de resolución, no obstante, en la zona dónde aún se ve sombra, se puede notar ligeramente la textura. La grieta se conserva bien tanto fuera como dentro de la sombra, pero sus segmentos más oscuros se sobreexponen.

4.1.2.3 ANÁLISIS CUALITATIVO DE EJEMPLO 3 (GRIETA OCULTA)

Con (1), la sombra se elimina, pero queda su silueta con una textura diferente a la del fondo. La textura del pavimento se conserva parcialmente, no obstante, en la región tratada no se restaura dejando una mancha de color uniforme. La grieta oculta bajo la sombra se elimina completamente.

Con (2), la sombra es eliminada, pero quedan artefactos en forma de halos finos, en los bordes de la sombra, y manchas, en el interior de la sombra. La textura del pavimento se conserva en el fondo, más no dentro de la sombra tratada. Solamente se mantienen y perciben los detalles más oscuros de la sombra, mientras que el resto se elimina.

Con (3) y (4), no se elimina la sombra, sólo se atenúa bastante, se ven artefactos en forma de diagonales finos en el interior de la región atenuada. La textura del pavimento se difumina por la disminución de resolución y no se restaura en la sombra. La grieta se elimina casi por completo, quedan rastros muy finos de esta. En (4), la iluminación general de la imagen aumenta ligeramente.

Con (5), la sombra no se elimina, sólo se atenúan las regiones más grandes, las más finas no se tratan. Se ven artefactos en forma de tablero de ajedrez dentro de la sombra. La textura del pavimento se conserva aun con la disminución de resolución, pero no se restaura en la sombra. La grieta se conserva en su mayoría, salvo en los segmentos más claros que se difuminan.

Con (6), la sombra se elimina casi por completo, quedan rastros con una tonalidad diferente y los segmentos más finos se sobreexponen. Se ven pocos artefactos. La textura del pavimento se difumina por la baja resolución y no se restaura en la sombra remanente. En la grieta, se mantienen las partes más oscuras, pero se exagera su brillo y forma.

4.1.2.4 ANÁLISIS DE ASPECTOS TÉCNICOS

En los métodos de procesamiento de imágenes, Imahayashi et al. (2023) es casi 5 veces más rápido que lo propuesto por Bi et al. (2024), pero ambos son igual de efectivos en entregar un resultado y mantener la resolución de la imagen. En los métodos basados en *Deep Learning*, *DC-ShadowNet* destacó por su rapidez tanto al usar AISTD y SRD como *datasets*

TRATAMIENTO DE SOMBRAS PARA DETECTAR DETERIOROS EN PAVIMENTOS CON MACHINE LEARNING

de entrenamiento, pero no se conservó la resolución. *DMTN* y *TBRNet* no entregaron resultados tampoco, esto se asocia a lo mencionado sobre la falta de los pesos de la red preentrenada VGG19.

En la tabla 10 están los rendimientos de cada método en términos del *dataset* con que se entrenó (para métodos de procesamiento de imágenes no aplica), el tiempo en segundos, el éxito al entregar un resultado y si conserva la resolución original o no. Las filas celestes representan los métodos por procesamiento de imágenes y las filas naranjas son los métodos basados en *Deep Learning*. Se destacan en verde los mejores tiempos en cada tipo de método.

Tabla 10: Rendimiento de cada método probado en *dataset* 2.

Fuente: Elaboración Propia.

Método	Dataset de entrenamiento	Tiempo [s]	Resultado	Resolución
Imahayashi et al.	N/A	60	✓	✓
Bi et al.	N/A	274	✓	✓
<i>DC-ShadowNet</i>	ISTD	32	✓	×
<i>DMTN</i>		-	×	N/A
<i>TBRNet</i>		-	×	N/A
<i>DC-ShadowNet</i>	AISTD	29	✓	×
<i>LG-ShadowNet</i>		188	✓	×
<i>DMTN</i>		-	×	N/A
<i>TBRNet</i>		-	×	N/A
<i>DC-ShadowNet</i>	SRD	29	✓	×
<i>DMTN</i>		-	×	N/A
<i>TBRNet</i>		-	×	N/A

4.1.3 RESULTADOS DE DATASET 3

A continuación, se muestra una figura con un ejemplo del *dataset* 3 y luego se hace un análisis cualitativo detallado por cada ejemplo.

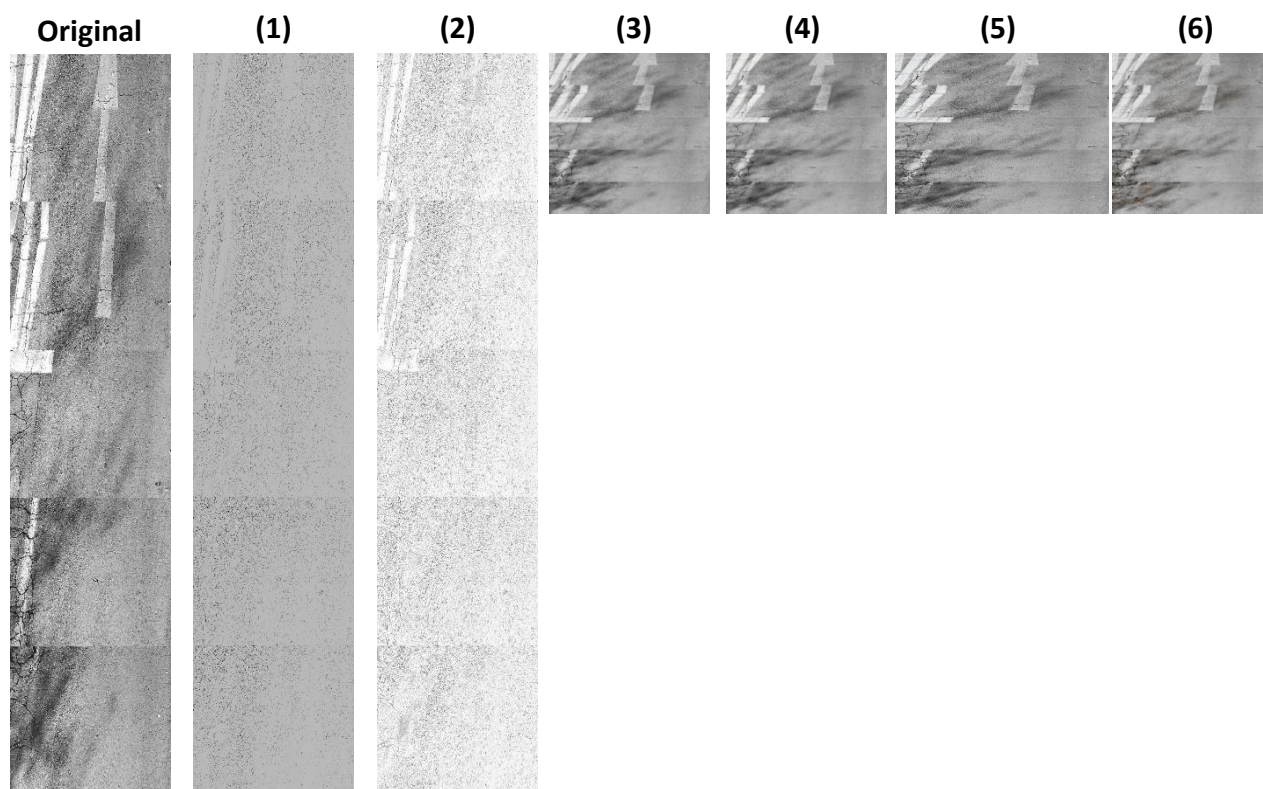


Figura 29: Resultados en *dataset 3*.

Fuente: Elaboración Propia.

4.1.2.1 ANÁLISIS CUALITATIVO DE EJEMPLO 1 (SECCIÓN LARGA)

Con (1), las sombras se eliminan por completo, pero se pierden detalles como las señales de tránsito. La textura del pavimento se conserva bien en toda la imagen, incluso donde estaban las sombras, donde sólo hay un ligero cambio de color. Las grietas se conservan, no obstante, se afinan demasiado.

Con (2), Se elimina la sombra casi por completo, salvo por pequeños rastros. La textura del pavimento se conserva bien en toda la imagen, pero se perciben algunos artefactos muy pequeños. Las grietas se mantienen, pero las partes más oscuras se aclaran y no coinciden con el color del resto de la grieta.

Con (3), (4), (5) y (6), la sombra no se elimina, la imagen no sufre más alteraciones que el cambio de resolución. Esto se debe a que los métodos basados en *Deep Learning* probados son entrenados y usados con imágenes en formato RGB, mientras que las secciones son imágenes en una escala de gris normalizada.

4.1.3.2 ANÁLISIS DE ASPECTOS TÉCNICOS

De los métodos de procesamiento de imágenes, Imahayashi et al. (2023) es mucho más rápido que Bi et al. (2024), pero ambos son capaces de entregar un resultado con una resolución igual a la original. En los métodos de *Deep Learning*, *DC-ShadowNet*

TRATAMIENTO DE SOMBRAS PARA DETECTAR DETERIOROS EN PAVIMENTOS CON MACHINE LEARNING

preentrenado con SRD es el más rápido, pero sus resultados alteran la resolución de la imagen. *DMTN* y *TBRNet* no completan la ejecución por la falta de pesos en la red preentrenada VGG19.

En la tabla 11 están los rendimientos de cada método en términos del *dataset* con que se entrenó (para métodos de procesamiento de imágenes no aplica), el tiempo en segundos, el éxito al entregar un resultado y si conserva la resolución original o no. Las filas celestes representan los métodos por procesamiento de imágenes y las filas naranjas son los métodos basados en *Deep Learning*. Se destacan en verde los mejores tiempos en cada tipo de método.

Tabla 11: Rendimiento de cada método probado en *dataset* 3.
Fuente: Elaboración Propia.

Método	Dataset de entrenamiento	Tiempo [s]	Resultado	Resolución
Imahayashi et al.	N/A	95	✓	✓
Bi et al.	N/A	540	✓	✓
<i>DC-ShadowNet</i>	ISTD	37	✓	×
<i>DMTN</i>		-	×	N/A
<i>TBRNet</i>		-	×	N/A
<i>DC-ShadowNet</i>	AISTD	36	✓	×
<i>LG-ShadowNet</i>		384	✓	×
<i>DMTN</i>		-	×	N/A
<i>TBRNet</i>		-	×	N/A
<i>DC-ShadowNet</i>	SRD	35	✓	×
<i>DMTN</i>		-	×	N/A
<i>TBRNet</i>		-	×	N/A

4.2 RESULTADOS EN YOLOV5

Por último, los resultados obtenidos con cada método por *dataset* fueron normalizados, para seguir el flujo de la metodología. En el caso del *dataset* 3, esto no fue necesario, pues la creación de las secciones largas es el paso posterior a la normalización.

Luego, se hicieron pruebas con YOLOv5, preentrenado con un archivo de pesos disponible de la metodología. Para este estudio la capacidad de clasificación de YOLOv5 no es relevante, ya que sólo se evalúa su rendimiento en detección de deterioros que estén ocultos parcial o totalmente por sombras y si es capaz de no confundir sombras como deterioros. Para ello, se compara con los resultados de YOLOv5 en las imágenes originales.

Usando los mismos ejemplos de la sección 4.1, se analizó en qué métodos se detectaron mejor las grietas ocultas y no se detectaron erróneamente sombras como deterioros.

4.2.1 RESULTADOS DE DATASET 1

En la figura 30 se muestran los resultados de las imágenes generadas con los métodos de eliminación de sombras al probarlos en YOLOv5.

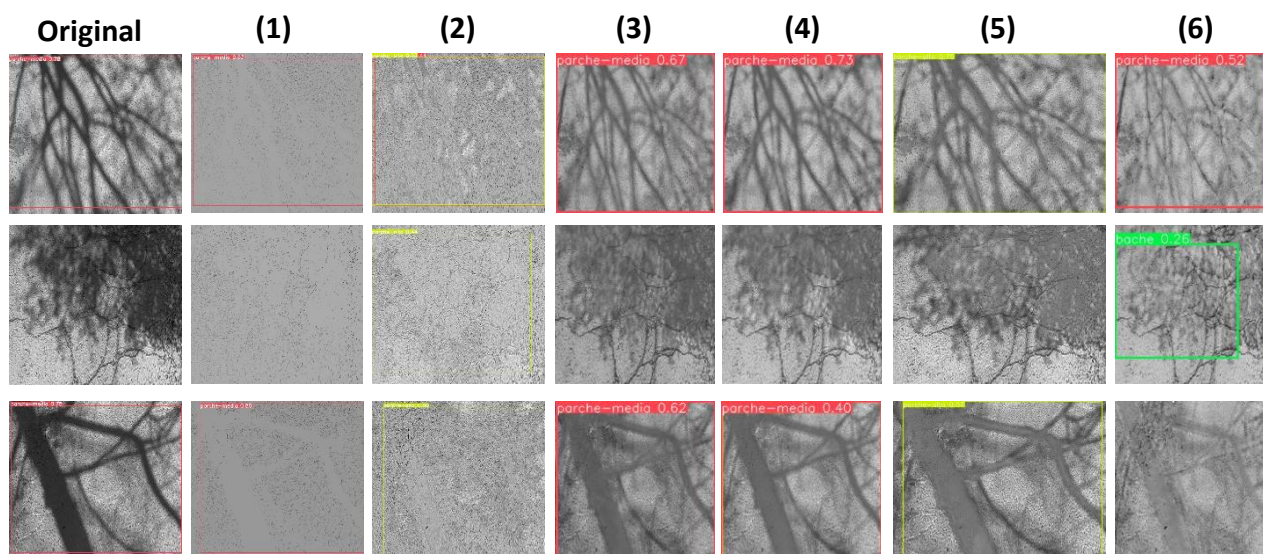


Figura 30: Resultados de YOLOv5 con el *dataset* 1.

Fuente: Elaboración Propia.

En el primer ejemplo, en la imagen original la sombra es detectada como un deterioro, esto evidencia que el problema por el que se inició el estudio aún está presente. Sin embargo, con la eliminación parcial o completa de la sombra que se hizo con los métodos, no hubo un cambio y se sigue detectando como un deterioro.

En el segundo ejemplo, en la imagen original la grieta no se detecta, ni su parte visible ni la que está oculta bajo la sombra. En el método (1), (3), (4) y (5) no se detecta la grieta, no obstante, en el método (2) y (6), independiente de la clasificación errónea, YOLOv5 logra detectar algo en la imagen resultante.

En el último ejemplo, en la imagen original la sombra se detecta como un deterioro y la grieta oculta es ignorada. Sólo el método (6) no identifica a la sombra como un deterioro, pero aun así no es capaz de detectar la grieta oculta.

4.2.2 RESULTADOS DE DATASET 2

En la figura 31 están los resultados de las imágenes generadas con los métodos de eliminación de sombras cuando se prueban con YOLOv5.

TRATAMIENTO DE SOMBRAS PARA DETECTAR DETERIOROS EN PAVIMENTOS CON MACHINE LEARNING

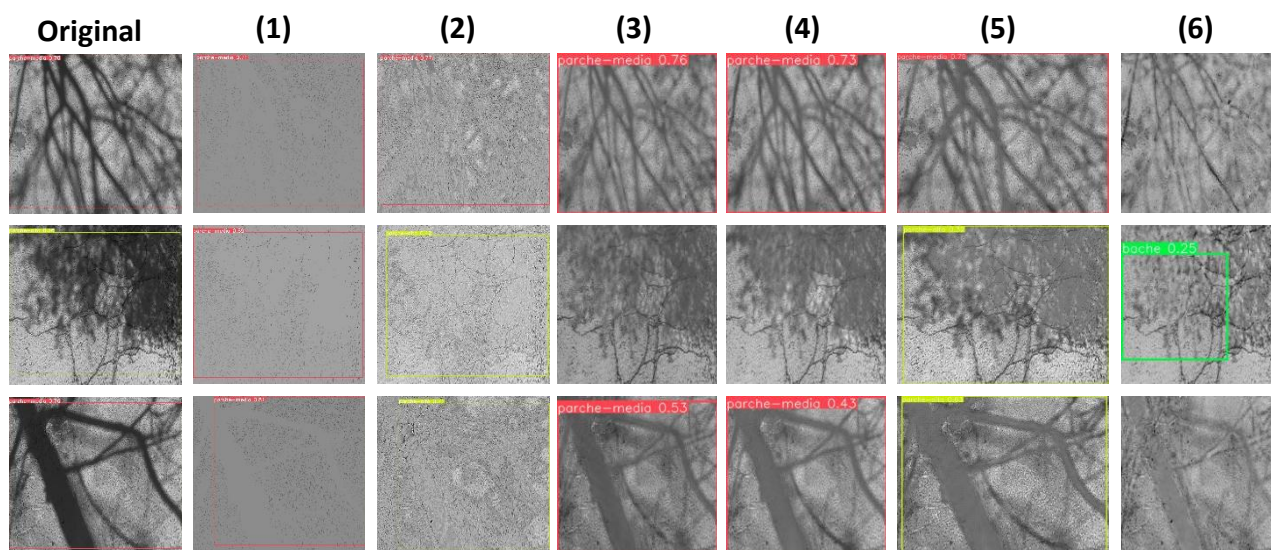


Figura 31: Resultados de YOLOv5 con el *dataset 2*.

Fuente: Elaboración Propia.

En el primer ejemplo, en la imagen original la sombra se detecta como un deterioro. El método (6) es el único que no identifica la sombra como un deterioro.

En el segundo ejemplo, en la imagen original se detecta un deterioro, pero se clasifica erróneamente como un parche de alta severidad. En los métodos (1), (2), (5) y (6), se detecta un deterioro, pero no se reporta como una grieta o fatiga, que es la verdadera naturaleza del deterioro. Mientras que los métodos (3) y (4) no se detecta nada.

En el último ejemplo, en la imagen original no se detecta la grieta oculta, en cambio se detecta la sombra como un deterioro de tipo parche de severidad media. Sólo el método (6) ignora la presencia de la sombra, pero no es capaz de detectar la grieta oculta.

4.2.3 RESULTADOS DE DATASET 3

En la figura 32 se muestran los resultados de las imágenes generadas por los métodos de eliminación de sombras al probarlas en YOLOv5.

TRATAMIENTO DE SOMBRAS PARA DETECTAR DETERIOROS EN PAVIMENTOS CON MACHINE LEARNING

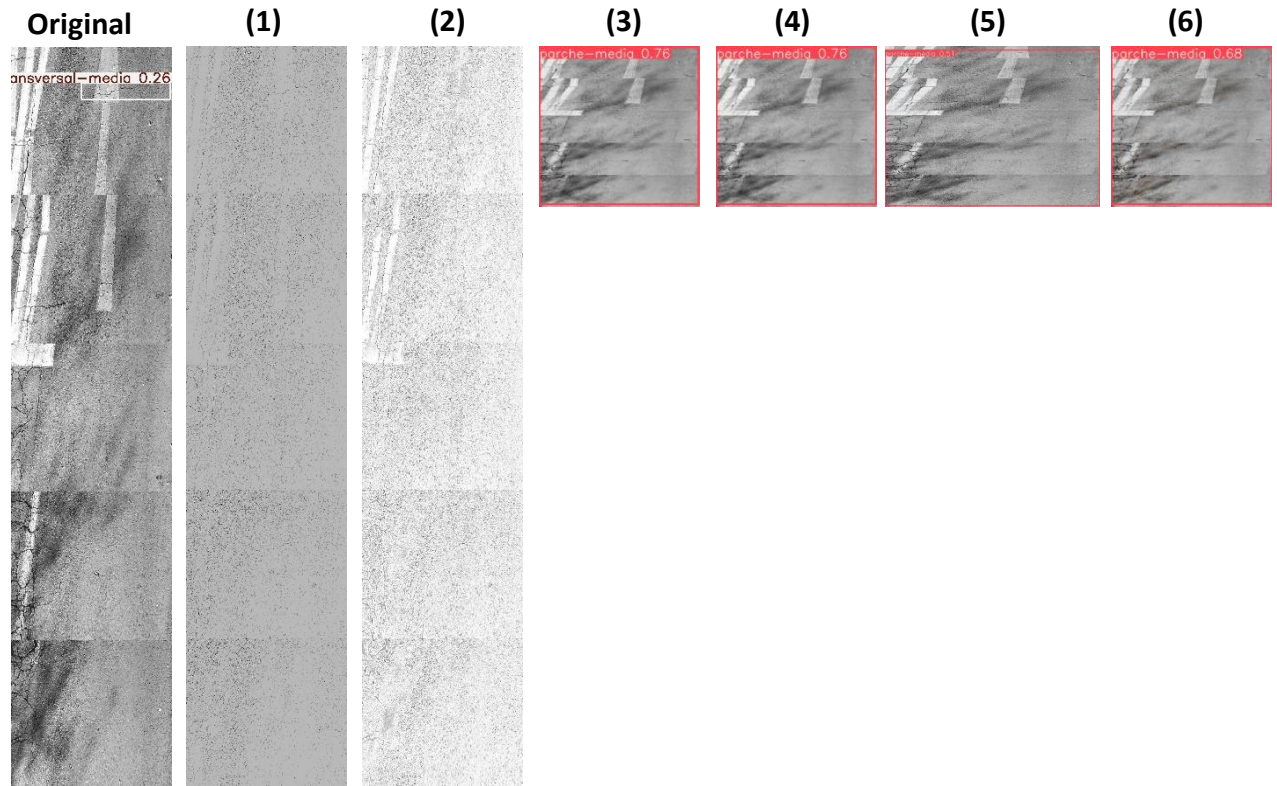


Figura 32: Resultados de YOLOv5 con *dataset* 3.

Fuente: Elaboración Propia.

En la imagen original, sólo se detecta una grieta transversal, pero se ignoran casos de fatiga y otras grietas que están más abajo. En los métodos (1) y (2), no hay sombras, pero los deterioros también fueron atenuados por lo que no se detectan ni siquiera el que estaba reportado originalmente. En el caso de los métodos (3), (4), (5) y (6), todos detectan un deterioro como un parche medio, pero esta información no es confiable, debido a que la resolución de la imagen original fue alterada de forma exagerada y la sombra no se elimina, por lo que su forma pierde sentido para el algoritmo.

CAPÍTULO 5: CONCLUSIONES

El presente trabajo ha demostrado la viabilidad de incorporar un módulo de tratamiento de sombras en la metodología de detección de deterioros en pavimentos urbanos de bajo costo.

5.1 ALCANCES Y LIMITACIONES DE LA PROPUESTA

La propuesta de solución ha demostrado una gran flexibilidad para integrar de manera eficiente un módulo de eliminación de sombras en el preprocesamiento de imágenes de pavimentos urbanos. Gracias a esta adaptabilidad, el detector de deterioros basado en YOLOv5 es capaz de ignorar las sombras sin confundirlas con grietas u otros daños en la superficie. No obstante, ninguno de los métodos evaluados logra suprimir las sombras por completo sin afectar otros atributos de la imagen, como la resolución, el color o la textura. En particular, los enfoques de *Deep Learning*, si bien ofrecen mejores resultados de eliminación, requieren un entrenamiento adicional con datos específicos de pavimentos para alcanzar la precisión y adaptabilidad deseadas.

5.2 RESULTADOS

Se hizo un análisis cualitativo en cada *dataset* de prueba acerca de la eliminación de la sombra, la preservación de textura, color y resolución de la imagen, la conservación de grietas y la restauración de detalles bajo las sombras.

Los métodos basados en procesamiento de imágenes son capaces de eliminar la sombra y conservar la resolución de la imagen, pero presentan ciertos inconvenientes como la eliminación parcial o total de grietas u otros deterioros y el cambio radical del color y textura de la imagen. Estos inconvenientes ocurren porque estos métodos, al estar diseñados con reglas fijas y supuestos concretos, no se adaptan a la diversidad de condiciones reales como variaciones de iluminación, texturas y tipos de sombra.

En cambio, los métodos basados en *Deep Learning*, aunque cambian la resolución de la imagen, lo que afecta su textura, son capaces de atenuar las sombras conservando los deterioros ocultos. Además, los inconvenientes que presentan, como el cambio de resolución, son factores que pueden ser corregidos fácilmente con el entrenamiento adecuado.

El análisis cualitativo reveló que el *dataset 2* tratado con *DC-ShadowNet* preentrenado en SRD es el que mejor ignora las sombras sin comprometer la textura del pavimento ni las grietas subyacentes. Además, al aplicar este método antes de la normalización de las imágenes, evita que YOLOv5 confunda las sombras con deterioros.

5.3 VALIDACIÓN DE OBJETIVOS

Se cumplió el objetivo general de diseñar e implementar un algoritmo de corrección de sombras, y los objetivos específicos quedaron satisfechos al (1) elaborar un dataset con imágenes sombreadas, (2) comparar distintos algoritmos de deep learning y (3) evaluar el impacto de la corrección de sombras en YOLOv5.

5.4 CONTRIBUCIONES DEL TRABAJO

A través del estudio de métodos para la eliminación de sombras, se consigue un análisis comparativo de métodos tradicionales frente a enfoques de *Deep Learning*. Además, se confirma empíricamente que *DC-ShadowNet* puede integrarse eficazmente en flujos de inferencia de detección de deterioros de la metodología, mediante las pruebas hechas sobre los tres *datasets*.

5.5 RECOMENDACIONES

Con la nueva adjudicación para continuar el proyecto, se hacen las siguientes recomendaciones para futuras investigaciones:

- Usar imágenes con filtro ND32, aunque este filtro disminuye el contraste entre el fondo y las sombras, al probar estas imágenes en YOLOv5, este último es capaz de ignorar las sombras y no detectarlas como deterioros.
- Entrenar *DC-ShadowNet* con imágenes del proyecto, ya que, en base a lo estudiado en la teoría y el análisis cualitativo, este método destaca por su rapidez en ejecución de pruebas, y es capaz de eliminar la sombra de mejor manera y conservar la textura del pavimento junto a las grietas en su mayoría. Al ser un método basado en *Deep Learning*, tiene una mayor flexibilidad para la detección de sombras, algo que no poseen los métodos de procesamiento de imágenes, debido a que no se entrenan y usan operaciones fijas que no se adaptan a los distintos casos que se puedan presentar. Además, al utilizar aprendizaje no supervisado, se puede entrenar sin necesitar pares de imágenes con y sin sombra que coincidan.
- Aplicar el método de eliminación de sombras inmediatamente después de extraer las imágenes de los fotogramas del video y antes de la etapa de normalización, ya que *DC-ShadowNet* es un método que trabaja con imágenes en el espacio RGB como se comprobó en los resultados
- Al realizar la inferencia con el algoritmo de detección de deterioros, utilizar las imágenes en sección larga, es decir, concatenar 5 imágenes. Esto se debe a que el algoritmo fue entrenado con imágenes de ese formato, por lo que se obtendrán resultados más fidedignos.

REFERENCIAS BIBLIOGRÁFICAS

1. Almendras, C. (2023). Desarrollo de una metodología de corrección de nitidez y contrastes de luz en imágenes para la detección de deterioro en pavimentos urbanos. Universidad Técnica Federico Santa María, Santiago, Chile.
2. Castro, G. (2024). Detección de deterioros superficiales en pavimentos urbanos mediante herramientas con algoritmos de deep learning: evaluación de una metodología aplicada. Universidad Técnica Federico Santa María, Santiago, Chile.
3. Osorio, A. (2020). Evaluation and modeling of urban pavements performance using machine learning.
4. Venegas, J. (2022). Desarrollo de una metodología de utilización de cámaras de bajo costo para la evaluación de pavimentos urbanos. Universidad Técnica Federico Santa María, Santiago, Chile.
5. Gao, M., Liu, X., Wei, H. y Yang, F. (2022). Shadow Removal from UAV Images Based on Color and Texture Equalization Compensation of Local Homogeneous Regions. *Remote Sensing*, 14(11), 2616. <https://doi.org/10.3390/rs14112616>
6. da Silva, E. A., de Azevedo, S. C. y Statella, T. (2013). Aplicação de Operadores Morfológicos na Detecção de Sombras em imagens QuickBird.
7. Imahayashi, S., Mukaida, M., Takeda, S. y Suetake, N. (2023). Shadow removal from document image based on background estimation employing selective median filter and black-top-hat transform. *Optical Review*, 30(3), 336-340. <https://doi.org/10.1007/s10043-023-00806-y>
8. Li, B. K., Ma, C. y Zhang, H. (2021). Shadow Separation of Pavement Images Based on Morphological Component Analysis. *Journal Of Control Science And Engineering*, 2021, 1-10. <https://doi.org/10.1155/2021/8828635>
9. Kamaruddin, M. N., Martin, A. y Sani, Z. M. (2022). Shadow detection and removal of road images with maximum extraction of road marker using multi-colour-space. *2021 International Symposium On Intelligent Signal Processing And Communication Systems (ISPACS)*, 1-4. <https://doi.org/10.1109/ispacs57703.2022.10082797>
10. Bi, C., Li, J., Liang, W., Wang, X. y Wang, X. (2024). Research on Pavement Crack Detection Based on Random Structure Forest and Density Clustering. *Automation*, 5(4), 467-483. <https://doi.org/10.3390/automation5040027>
11. Guo, L., Huang, S., Kot, A. C., Wang, C., Wang, Y., Wen, B. y Yang, W. (2024). Single-Image shadow Removal Using Deep Learning: A Comprehensive survey. *arXiv (Cornell University)*. <https://doi.org/10.48550/arxiv.2407.08865>

12. Gonzalez, R. C. y Woods, R. E. (1977). *Digital Image Processing, 4th Edition*. [https://github.com/BhanuPrakashNani/Image_Processing/blob/master/Digital Image Processing%2C 4th%20Edition-Rafael%20Gonzalez.pdf](https://github.com/BhanuPrakashNani/Image_Processing/blob/master/Digital_Image_Processing%2C%204th%20Edition-Rafael%20Gonzalez.pdf)
13. Dilhani, S. (2021). Digital Image Processing Filters. *Medium*. Recuperado de <https://medium.com/@shashikadilhani97/digital-image-processing-filters-832ec6d18a73>
14. Bisht, R., Singh, S. y Vijay, R. (2018). Comparative Analysis of Fixed Valued Impulse Noise Removal Techniques for Image Enhancement. En *Communications in computer and information science* (pp. 175-184). https://doi.org/10.1007/978-981-13-1810-8_18
15. Bovik, A., Sheikh, H., Simoncelli, E. y Wang, Z. (2004). Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions On Image Processing*, 13(4), 600-612. <https://doi.org/10.1109/tip.2003.819861>
16. Polamuri, S. (s.f.). Best Confusion Matrix Guide with Sklearn Python. *Dataaspirant*. Recuperado de <https://dataaspirant.com/confusion-matrix-sklearn-python/>
17. Jin, Y., Sharma, A., y Tan, R. T. (2021). DC-ShadowNet: Single-Image Hard and Soft Shadow Removal Using Unsupervised Domain-Classifer Guided Network. *2021 IEEE/CVF International Conference On Computer Vision (ICCV)*, 5007-5016. <https://doi.org/10.1109/iccv48922.2021.00498>
18. Liu, Z., Mi, Y., Pu, M., Wang, S. y Yin, H. (2021). Shadow Removal by a Lightness-Guided Network With Training on Unpaired Data. *IEEE Transactions On Image Processing*, 30, 1853-1865. <https://doi.org/10.1109/tip.2020.3048677>
19. Fan, H., Li, W., Liu, J., Qu, L., Tang, Y. y Wang, Q. (2023). A Decoupled Multi-Task Network for Shadow Removal. *IEEE Transactions On Multimedia*, 25, 9449-9463. <https://doi.org/10.1109/tmm.2023.3252271>
20. Fan, H., Liu, J., Tang, Y., Tian, J. y Wang, Q. (2023). A Shadow Imaging Bilinear Model and Three-Branch Residual Network for Shadow Removal. *IEEE Transactions On Neural Networks And Learning Systems*, 35(11), 15857-15871. <https://doi.org/10.1109/tnnls.2023.3290078>
21. Fu, C., Heng, P., Hu, X. & Jiang, Y. (2019). Mask-ShadowGAN: Learning to Remove Shadows From Unpaired Data. *2021 IEEE/CVF International Conference On Computer Vision (ICCV)*, 2472-2481. <https://doi.org/10.1109/iccv.2019.00256>