






# APLICACIÓN DE MODELOS U-NET PARA SEGMENTACIÓN SEMÁNTICA DE DEFECTOS EN PANELES FOTOVOLTAICOS

## U-NET–BASED SEMANTIC SEGMENTATION OF DEFECTS IN PHOTOVOLTAIC PANELS

Franklin Gómez-López<sup>1</sup> , Danny Ochoa-Correa<sup>2,\*</sup> , Isabel Cabrera-Carrera<sup>1</sup> 

Recibido: 16-06-2025, Recibido tras revisión: 18-11-2025, Aceptado: 04-12-2025, Publicado: 01-01-2026

### Resumen


Este artículo presenta un estudio sobre la segmentación semántica de defectos en celdas fotovoltaicas de silicio cristalino mediante modelos basados en U-Net, entrenados con imágenes de electroluminiscencia (EL). Se empleó un conjunto de datos compuesto por imágenes adquiridas en laboratorio y un prototipo público de pruebas, ambos con anotaciones manuales de grietas, zonas oscuras y discontinuidades en barras colectoras. Se entrenaron ocho versiones del modelo, incorporando variaciones controladas en la resolución, la profundidad del codificador y las estrategias de regularización. La evaluación incluyó métricas clase a clase (precisión, *recall* y *F1-score*), análisis visual mediante mapas de calor y superposiciones, así como validación por expertos. Si bien la segmentación fue consistente en defectos de morfología clara, como zonas oscuras y barras colectoras, las grietas presentaron mayores dificultades debido a su baja densidad de píxeles y geometría irregular. Asimismo, se analizaron arquitecturas alternativas (U-Net++ y MAU-Net), sin evidenciar mejoras relevantes frente a la configuración optimizada de U-Net. Los resultados respaldan el uso de este enfoque en tareas de inspección automatizada bajo condiciones controladas, y se proponen extensiones para su aplicación en contextos operativos más diversos.

**Palabras clave:** electroluminiscencia, mantenimiento predictivo, paneles solares, segmentación semántica, U-Net

### Abstract

This article presents a study on the semantic segmentation of defects in crystalline-silicon photovoltaic cells using U-Net–based models trained on electroluminescence (EL) images. The dataset combines laboratory-acquired images with a publicly available benchmark, both manually annotated to identify cracks, dark zones, and collector-bar discontinuities. Eight model variants were trained with controlled variations in input resolution, encoder depth, and regularization strategies. Performance was assessed using per-class precision, recall, and F1-score, complemented by visual inspection through heatmaps and overlays and by expert validation. Segmentation was robust for defects with well-defined morphology, such as dark zones and busbars; however, cracks remained more difficult to detect due to their sparse pixel representation and irregular geometry. Alternative architectures (U-Net++ and MAU-Net) were also evaluated but did not yield meaningful improvements over the optimized U-Net configuration. Overall, the results support the use of this approach for automated inspection under controlled conditions, while highlighting the need for future adaptation to more diverse operational scenarios.

**Keywords:** Electroluminescence, predictive maintenance, photovoltaic panels, semantic segmentation, U-Net

<sup>1,\*</sup>Facultad de Ingeniería, Universidad de Cuenca, Ecuador. 

<sup>2</sup>Departamento de Ingeniería Eléctrica, Electrónica y de Telecomunicaciones, Universidad de Cuenca, Ecuador.   
 Autor para correspondencia ✉: [danny.ochoac@ucuenca.edu.ec](mailto:danny.ochoac@ucuenca.edu.ec)

Forma sugerida de citación: F. Gómez-López, D. Ochoa-Correa y I. Cabrera-Carrera, “Aplicación de modelos U-Net para segmentación semántica de defectos en paneles fotovoltaicos,” *Ingenius, Revista de Ciencia y Tecnología*, N.º 35, pp. 110-121, 2026. DOI: <https://doi.org/10.17163/ings.n35.2026.08>.

## 1. Introducción

Los sistemas de energía fotovoltaica (PV) se han convertido en un pilar fundamental de los esfuerzos globales de transición energética, particularmente a través de la generación distribuida y la electrificación descentralizada. A medida que las instalaciones fotovoltaicas continúan expandiéndose, mantener su integridad operativa mediante métodos diagnósticos precisos resulta cada vez más crucial.

La imagen de electroluminiscencia (EL) se utiliza ampliamente para detectar anomalías estructurales en los módulos fotovoltaicos, incluidas microfisuras, dedos rotos y áreas eléctricamente inactivas [1, 2]. Aunque estos defectos suelen ser imperceptibles durante las inspecciones visuales, pueden degradar progresivamente el rendimiento del módulo con el tiempo.

En comparación con la termografía infrarroja y la fluorescencia ultravioleta, la imagen EL ofrece una resolución espacial superior para la detección temprana de defectos, sin depender de la irradiancia solar ni de gradientes de temperatura [3, 4]. No obstante, las imágenes EL sin procesar suelen presentar ruido e iluminación desigual, lo que dificulta la identificación de patrones sutiles de defectos. Estas limitaciones subrayan la necesidad de técnicas avanzadas de análisis de imágenes capaces de aislar de manera fiable características de defectos finos y heterogéneos.

La segmentación semántica, que clasifica los píxeles individuales en categorías predefinidas, permite una localización precisa de defectos en imágenes de electroluminiscencia (EL) [5, 6]. Entre los enfoques de aprendizaje profundo, las redes neuronales convolucionales (CNN) han demostrado un desempeño sólido en tareas de segmentación. En particular, la arquitectura U-Net es ampliamente adoptada para la detección de defectos en módulos fotovoltaicos [7].

Desarrollada originalmente para el análisis de imágenes biomédicas, U-Net integra de manera eficaz la extracción de características a múltiples escalas con la preservación de la resolución espacial, lo que favorece la detección de estructuras estrechas y localizadas, como microfisuras e inconsistencias de soldadura en celdas fotovoltaicas [8, 9].

Investigaciones recientes han adaptado y refinado variantes de U-Net para la segmentación de defectos en sistemas fotovoltaicos mediante la incorporación de mecanismos de atención, conexiones residuales y funciones de pérdida híbridas [10, 11]. Estas modificaciones mejoran la robustez del modelo bajo condiciones variables de iluminación y aumentan la precisión de la segmentación, dada la variabilidad entre celdas presente en los conjuntos de datos de módulos fotovoltaicos.

Además, los conjuntos de datos de imágenes EL de acceso público, incluidos aquellos publicados por Fraunhofer ISE y Sandia National Laboratories, respaldan la evaluación comparativa y la reproducibilidad de los

resultados de segmentación [12, 13]. En este estudio se implementan y evalúan múltiples arquitecturas U-Net entrenadas con imágenes EL de celdas solares de silicio cristalino. Mediante el uso de conjuntos de datos tanto propietarios como públicos, se analiza la precisión de la segmentación en distintos tipos de defectos.

La evaluación incorpora estrategias de optimización de parámetros y comparaciones cuantitativas con el fin de determinar la idoneidad de los métodos basados en U-Net para respaldar la inspección automatizada bajo condiciones controladas de laboratorio.

## 2. Materiales y métodos

### 2.1. Conjuntos de datos

#### 2.1.1. Conjunto de datos de electroluminiscencia interno

El conjunto de datos principal fue recopilado en el Laboratorio de Micro-Red de la Universidad de Cuenca. Este comprende 180 celdas solares extraídas de módulos fotovoltaicos monocristalinos y policristalinos fuera de servicio. Cada celda fue capturada bajo condiciones controladas utilizando una cámara OWL 640 M SWIR (Raptor Photonics), equipada con un sensor InGaAs sensible al rango espectral de 900–1700 nm. Este intervalo proporciona un contraste mejorado para la identificación de defectos en silicio cristalino en comparación con sensores que operan en el espectro visible.

Para garantizar una iluminación uniforme, las celdas fueron polarizadas mediante una fuente de alimentación programable Chroma DC. La corriente aplicada se varió entre  $\frac{1}{6}I_{SC}$  y  $I_{SC}$ , siguiendo la norma IEC TS 60904-13, con el fin de examinar cómo la intensidad de excitación influye en las características de la luminiscencia. La adquisición de imágenes se gestionó mediante la plataforma XCAP-Std, lo que posibilitó el ajuste en tiempo real del tiempo de exposición, la ganancia y la velocidad de fotogramas. Los parámetros óptimos de adquisición fueron tiempos de exposición de 30 y 50 ms, un factor de ganancia de 2.5 y una corriente de polarización de aproximadamente 6 A, correspondiente a cerca de  $\frac{2}{3}I_{SC}$ .

Para validar la calidad del sistema de adquisición, la relación señal-ruido (SNR) se cuantificó de acuerdo con la norma IEC TS 60904-13 [14] y se definió como se ve en la ecuación (1):

$$SNR_{IEC} = \frac{\sum\{0.5 \cdot (I_1 + I_2) - I_{BG}\}}{\sum[\sqrt{0.5 \cdot |I_1 - I_2| \cdot (\frac{2}{\pi})^{-0.5}}]} \quad (1)$$

Donde  $I_1$  e  $I_2$  son capturas EL consecutivas e  $I_{BG}$  representa el ruido de fondo. El arreglo experimental alcanzó un valor de  $SNR_{IEC} > 45$ , cumpliendo los criterios establecidos para mediciones de laboratorio.

Posteriormente, las imágenes EL sin procesar fueron sometidas a un preprocesamiento que incluyó sustracción de fondo, corrección de perspectiva y mejora del contraste local mediante la ecualización de histograma adaptativa con límite de contraste (CLAHE), configurada con un límite de recorte de 2.0 y un tamaño de cuadrícula de  $8 \times 8$ . Esta cadena de preprocesamiento mejoró de manera significativa la calidad de las imágenes para las tareas de segmentación.

### 2.1.2. Conjunto de datos de referencia público

Además del conjunto de datos propietario, se utilizó para la validación el conjunto de datos público ELPV, desarrollado por el grupo ZAE Bayern. Este conjunto comprende imágenes de electroluminiscencia (EL) y sus correspondientes máscaras binarias de grietas para 2624 celdas solares, obtenidas en condiciones de fábrica.

A pesar de las diferencias en la resolución y en la estructura de las máscaras en comparación con el conjunto de datos interno, se emplearon scripts de preprocesamiento para unificar las dimensiones de las imágenes y los formatos de codificación.

Al integrar tanto los conjuntos de datos internos como externos, el conjunto combinado respalda la generalización del modelo a través de diferentes tipos de paneles y configuraciones de adquisición. Estudios previos que han adoptado estrategias similares basadas en conjuntos de datos híbridos han reportado mejoras tanto en la precisión de la segmentación como en la robustez entre dominios.

### 2.1.3. Integración y preprocesamiento de datos

Dado que el conjunto de datos público ELPV proporcionaba únicamente máscaras binarias de grietas, se llevó a cabo un proceso de reanotación utilizando la plataforma Supervisely [15], con el fin de generar manualmente las máscaras de referencia correspondientes a las clases no anotadas originalmente.

Esto garantizó que los conjuntos de datos interno y público compartieran un esquema de anotación multiclase coherente, compuesto por fondo y tres clases de defectos.

La Tabla 1 resume la distribución de las instancias de defectos en los conjuntos de entrenamiento y validación. El predominio de la clase *busbar* es esperado, ya que este componente estructural está presente en cada celda.

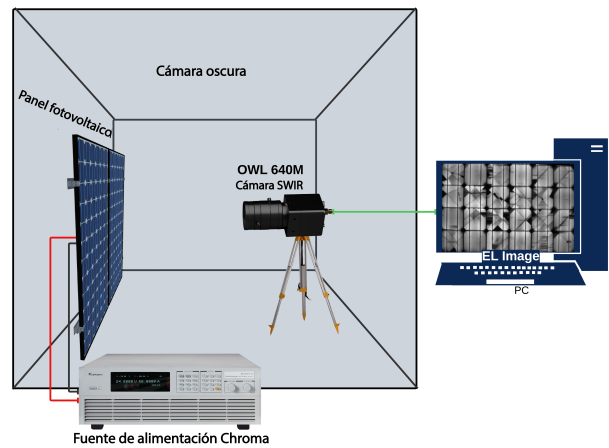
Este desbalance de clases se abordó durante el entrenamiento mediante la cadena de aumento de datos descrita en la sección siguiente.

**Tabla 1.** Distribución de instancias de defectos en los conjuntos de entrenamiento y validación

Clase de defectos	Instancias de entrenamiento	Instancias de validación
Barras colectoras	1453	96
Grietas	608	92
Zonas oscuras	425	84

## 2.2. Configuración y flujo de adquisición

El montaje completo del sistema de adquisición se muestra en la Figura 1. La cámara SWIR se posicionó ortogonalmente sobre la celda, manteniendo una distancia fija de 50 cm entre la lente y el objeto. La captura de imágenes se realizó en una cámara oscurecida con el fin de eliminar la interferencia de la luz ambiental. Asimismo, se aplicó la corrección de no uniformidad (NUC) mediante un método de tres puntos (Offset + Gain + Dark), con el objetivo de minimizar el ruido de patrón fijo.



**Figura 1.** Montaje experimental que muestra la cámara OWL 640 M SWIR y el módulo fotovoltaico bajo prueba.

Todos los pasos de procesamiento de imágenes se implementaron en Python mediante las bibliotecas OpenCV y NumPy. Las métricas de evaluación, incluidas la relación de mejora de contraste (CIR), la relación pico-a-bajo (PL) y los mapas de calor, se extrajeron de cada imagen preprocesada.

## 2.3. Coherencia de la anotación

La concordancia entre anotadores se evaluó periódicamente utilizando el coeficiente de Dice, ver ecuación (2):

$$\text{Dice}(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad (2)$$

Donde A y B representan conjuntos de píxeles anotados de forma independiente por distintos evaluadores.

A lo largo de todo el proceso de anotación se mantuvo de manera consistente un valor medio del coeficiente de Dice superior a 0.85.

Estos conjuntos de datos sirvieron como base para el entrenamiento y la validación de los modelos de segmentación descritos en la sección siguiente.

## 2.4. Arquitectura del modelo

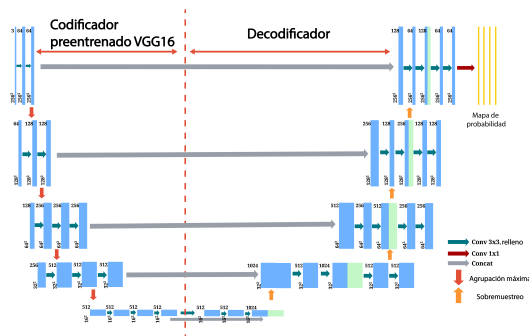
### 2.4.1. U-Net estándar con codificador VGG16

La arquitectura empleada en este estudio se basa en U-Net, desarrollada originalmente para la segmentación de imágenes biomédicas [7]. Su estructura de tipo codificador-decodificador permite la clasificación a nivel de píxel mediante la integración de características multiescala, al tiempo que preserva la resolución espacial. En esta implementación, la ruta de contracción se sustituye por un codificador VGG16 preentrenado en el conjunto de datos ImageNet [16].

Dado que la arquitectura VGG16 estándar espera una entrada de tres canales (RGB) y las imágenes EL son de un solo canal (escala de grises), la entrada se adaptó durante el preprocesamiento replicando el canal único tres veces, con el fin de coincidir con las dimensiones tensoriales esperadas sin modificar los pesos preentrenados. Esta configuración favorece una representación más rica de características en la etapa de entrada y contribuye a una convergencia más estable durante el entrenamiento [17].

Cada etapa de muestreo descendente en el codificador incluye dos capas convolucionales de  $3 \times 3$ , seguidas de una operación de *max-pooling* de  $2 \times 2$ . El codificador genera cinco mapas de características de profundidad creciente. Para ampliar el campo receptivo sin comprometer las dimensiones espaciales, el *bottleneck* emplea convoluciones dilatadas con una tasa de dilatación fija de  $r = 2$ .

La Figura 2 ilustra la variante de U-Net adoptada en este estudio, que incorpora convoluciones dilatadas en el *bottleneck* con el objetivo de ampliar el campo receptivo sin incrementar el número de parámetros del modelo [8, 18].



**Figura 2.** Arquitectura U-Net con codificador VGG16 y *bottleneck* de convoluciones dilatadas.

### 2.4.2. Configuración de capas, funciones de activación y salida multiclase

Todas las capas convolucionales emplean funciones de activación ReLU y normalización por lotes con el fin de mejorar la estabilidad del entrenamiento. Se aplica una tasa de dropout de 0.2 en el *bottleneck* y en las capas del decodificador para mitigar el sobreajuste.

El decodificador realiza el muestreo ascendente de los mapas de características mediante convoluciones transpuestas, incorporando conexiones de salto que vinculan cada bloque de muestreo ascendente con la etapa correspondiente del codificador. La capa de salida consiste en una convolución de  $1 \times 1$  seguida de una activación softmax, con el propósito de generar una distribución de probabilidad por píxel sobre  $C$  clases de defectos, ver ecuación (3):

$$\text{Salida}_{i,j,c} = \frac{\exp(z_{i,j,c})}{\sum_{k=1}^C \exp(z_{i,j,k})} \quad (3)$$

Donde  $z_{i,j,c}$  es la puntuación sin procesar para la clase  $c$  en el píxel  $(i, j)$ , y  $C = 3$  representa las clases de defectos: grietas, zonas oscuras y *busbars*.

Las imágenes de entrada corresponden a imágenes de electroluminiscencia (EL) en escala de grises, redimensionadas a una resolución de  $256 \times 256$  píxeles. Las máscaras de segmentación de salida preservan la misma resolución. Toda la cadena de procesamiento y entrenamiento se implementó utilizando el entorno PyTorch.

Estudios previos indican que la combinación de conexiones de salto, una arquitectura simétrica y salidas multiclase resulta adecuada para la detección de patrones sutiles de defectos en imágenes EL [10, 19].

Asimismo, se examinaron arquitecturas alternativas, como MAU-Net con mecanismos de atención espacial y de canal, con fines comparativos; sus resultados se presentan y discuten en la sección de Resultados y discusión.

## 2.5. Proceso de entrenamiento

### 2.5.1. Preprocesamiento y aumento de datos

Antes del entrenamiento, todas las imágenes de electroluminiscencia (EL) se redimensionaron a una resolución estándar de  $256 \times 256$  píxeles. Para distinguir de forma adecuada las características continuas de la imagen de las etiquetas categóricas durante el redimensionamiento, se aplicó interpolación bilineal a las imágenes EL en escala de grises, mientras que se utilizó interpolación por vecino más cercano para las máscaras, con el fin de preservar los valores discretos de las clases. Posteriormente, los valores de los píxeles se normalizaron al rango  $[0, 1]$  para garantizar un formato de entrada coherente y mejorar la estabilidad del modelo. Cada imagen de entrada en escala de grises se emparejó con una máscara codificada en one-hot que

representaba las tres categorías de defectos: grietas, zonas oscuras y *busbars* (barras colectoras).

Con el objetivo de ampliar la variabilidad de las muestras de entrenamiento, se aplicó aumento de datos *on-the-fly*. Las transformaciones incluyeron volteos horizontales y verticales, rotaciones aleatorias de hasta  $\pm 15^\circ$ , escalamiento por zoom en el rango de  $0.9\times$  a  $1.1\times$ , y deformaciones elásticas. Estos aumentos se aplicaron de manera probabilística mediante la biblioteca *Albumentations*, garantizando que la integridad estructural de las máscaras se mantuviera.

Este enfoque de aumento de datos mejora la capacidad de generalización, particularmente en modelos entrenados con conjuntos de datos EL pequeños u homogéneos. Estrategias similares han sido empleadas para abordar el desbalance de clases y aumentar la sensibilidad a patrones de defectos poco frecuentes.

### 2.5.2. Configuración de hiperparámetros y programador de tasa de aprendizaje

El entrenamiento del modelo se realizó utilizando el optimizador Adam, que combina la estimación adaptativa del gradiente con momentum. Los hiperparámetros seleccionados fueron los siguientes:

- Tasa de aprendizaje:  $lr = 0.0018$
- Tasa de decaimiento del primer momento:  
 $\beta_1 = 0.9$
- Tasa de decaimiento del segundo momento:  
 $\beta_2 = 0.999$
- Término de estabilidad numérica:  $\epsilon = 1 \times 10^{-8}$

El entrenamiento se guio mediante la función de pérdida estándar de entropía cruzada categórica, ver ecuación (4):

$$\mathcal{L} = - \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(\hat{y}_{i,c}) \quad (4)$$

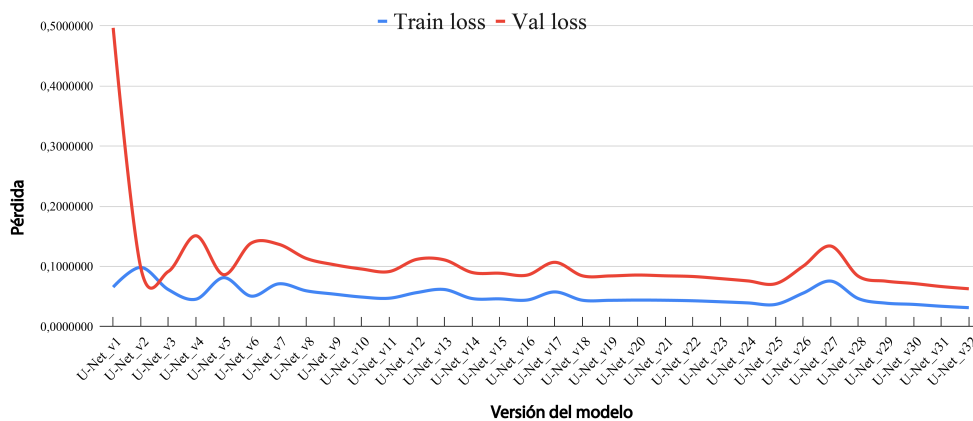
Donde  $y_{i,c}$  representa la etiqueta verdadera (*codificada en one-hot*)  $\hat{y}_{i,c}$  corresponde a la probabilidad predicha para la clase  $c$  en el píxel  $i$ . Aunque el desbalance de clases es inherente a la detección de defectos en imágenes de electroluminiscencia (EL), este se abordó principalmente mediante la estrategia agresiva de aumento de datos descrita en la sección anterior, en lugar de aplicar ponderaciones de clase en la función de pérdida.

Se utilizó un programador de tasa de aprendizaje basado en pasos con el fin de mejorar la convergencia y evitar mínimos locales. La tasa de aprendizaje se redujo por un factor  $\gamma = 0.05$  cada ocho épocas. Este decaimiento gradual facilitó el ajuste fino durante las fases posteriores del entrenamiento. Asimismo, se habilitó *early stopping* para finalizar el entrenamiento si la pérdida de validación no mostraba mejoras durante 10 épocas consecutivas.

El proceso de entrenamiento fue iterativo. Cada versión posterior del modelo se inicializó con los pesos de la versión previa de mejor desempeño, lo que permitió un refinamiento progresivo sin reinicialización completa. La versión final del modelo (U-Net\_v32) se entrenó durante 48 épocas, observándose la menor pérdida de validación en la época 26. Los valores correspondientes de pérdida de entrenamiento y validación se muestran en la ecuación (5):

$$\text{Train loss} = 0.0313 \quad \text{Val loss} = 0.0627 \quad (5)$$

La Figura 3 muestra la evolución de los valores de pérdida a lo largo de las distintas versiones del modelo. Estas visualizaciones permitieron monitorizar el comportamiento del entrenamiento y validar las mejoras incrementales obtenidas en cada iteración.



**Figura 3.** Evolución de la pérdida de entrenamiento y validación durante el proceso iterativo de optimización del modelo U-Net.

Con el fin de garantizar la reproducibilidad, todos los experimentos se realizaron utilizando el entorno PyTorch (v2.3.1) en Google Colab Pro, equipado con una GPU NVIDIA T4. Se fijó una semilla aleatoria de 42 para la inicialización de los pesos y la partición de los datos. El entrenamiento se llevó a cabo con un tamaño de lote de 16 durante 48 épocas, empleando el optimizador Adam con una tasa de aprendizaje inicial de 0.0018 y un programador de decaimiento basado en pasos (*step-based*), con  $step\_size = 8$  y  $\gamma = 0.05$ . Los mejores pesos del modelo se almacenaron en función de la mínima pérdida de validación, que típicamente convergía alrededor de la época 26.

El modelo U-Net final presenta un tamaño de almacenamiento aproximado de 233 MB. Durante la fase de inferencia, el procesamiento de un lote de 16 imágenes ( $256 \times 256$  píxeles) requirió en promedio 0.52 s, lo que corresponde a una latencia de inferencia de aproximadamente 32.6 ms por imagen, equivalente a cerca de 30 FPS.

## 2.6. Métricas de evaluación

### 2.6.1. Precisión, *recall* y *F1-score*

Para evaluar el desempeño del modelo a nivel de píxel, se calcularon tres métricas complementarias: precisión, *recall* y *F1-score*. Estas métricas se estimaron de manera independiente para cada clase de defecto —grietas, zonas oscuras y *busbars* (barras colectoras)—, tal como se recomienda en estudios de segmentación orientados a la detección de defectos en sistemas fotovoltaicos [8, 18].

La precisión cuantifica la proporción de píxeles de defecto correctamente identificados entre todos los píxeles clasificados como defectuosos, ver ecuación (6):

$$\text{Precisión} = \frac{TP}{TP + FP} \quad (6)$$

Donde TP representa los verdaderos positivos (píxeles de defectos correctamente predichos) y FP denota los falsos positivos (píxeles sin defectos incorrectamente clasificados como defectuosos).

El *recall*, también denominado sensibilidad, mide la proporción de píxeles defectuosos reales que son identificados correctamente, ver ecuación (7):

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

Donde FN indica los falsos negativos (píxeles defectuosos que el modelo no logró detectar).

El *F1-score*, calculado como la media armónica de la precisión y el *recall*, proporciona una métrica equilibrada para evaluar la exactitud de la clasificación, ver ecuación (8):

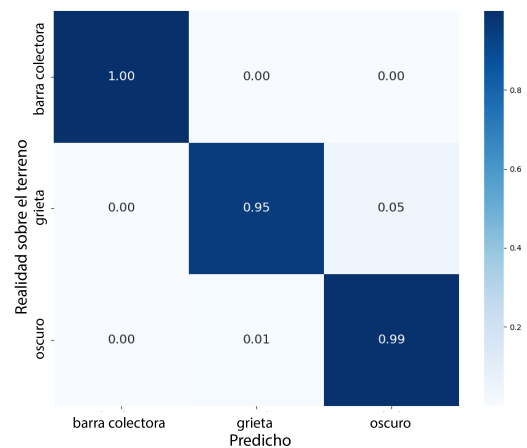
$$\text{F1-score} = \frac{2TP}{2TP + FP + FN} \quad (8)$$

Estas métricas resultan particularmente relevantes en tareas de segmentación multiclase que involucran desbalance de clases y características morfológicas sutiles, como las microgrietas [10].

### 2.6.2. Estrategia de validación y matriz de confusión

Para evaluar el desempeño del modelo y monitorizar la convergencia durante el entrenamiento iterativo de 32 versiones del modelo, se empleó una estrategia de validación *hold-out* fija, en lugar de validación cruzada *k-fold*. El conjunto de datos se dividió estrictamente en tres subconjuntos independientes: entrenamiento (1453 imágenes, incluidas las generadas mediante aumento de datos), validación (96 imágenes originales) y prueba. Este enfoque garantizó que las métricas de evaluación permanecieran comparables a lo largo de las distintas épocas de entrenamiento y ajustes de hiperparámetros, al tiempo que evitó la fuga de datos derivada del proceso de aumento.

Se generó una matriz de confusión utilizando el modelo de mejor desempeño (U-Net\_v32) sobre el conjunto de validación, con el fin de visualizar los errores de predicción específicos por clase. La Figura 4 presenta la matriz de confusión normalizada. Se observó una mayor tasa de clasificación errónea en la clase de grietas, lo cual se atribuye a la distribución espacial dispersa y fragmentada de este tipo de defecto, en comparación con las características más distintivas de los *busbars* (barras colectoras) y las zonas oscuras.



**Figura 4.** Matriz de confusión normalizada obtenida para el modelo U-Net\_v32 en el conjunto de validación, que muestra el desempeño de clasificación por clase.

Este protocolo de validación posibilita una evaluación clara de la capacidad de generalización del modelo en datos no vistos, con especial énfasis en la identificación correcta de las clases de defectos minoritarios.

### 3. Resultados y discusión

#### 3.1. Desempeño por clase de defecto

##### 3.1.1. Segmentación de grietas, zonas oscuras y busbars

El desempeño de segmentación de los modelos U-Net entrenados varió entre las distintas clases de defectos. Como se muestra en la Tabla 2, la puntuación F1 más alta se obtuvo para la clase de zonas oscuras, las cuales suelen ocupar regiones más extensas y homogéneas. Las discontinuidades en los *busbars* (barras colectoras) también exhibieron un desempeño de segmentación consistente, particularmente en aquellos casos en los que existía un contraste claro con el fondo de la celda.

En contraste, las grietas plantearon mayores desafíos para el modelo. Este tipo de defectos se caracteriza por estructuras estrechas e irregulares que, con frecuencia, se superponen con texturas de bajo contraste. De manera consistente con estudios previos de Pratt *et al.* [8] y Deitsch *et al.* [18], la clase de grietas presentó valores inferiores de precisión y *recall*, atribuibles tanto a su dispersión a nivel de píxel como a su elevada variabilidad morfológica.

**Tabla 2.** Métricas medias de segmentación por clase (U-Net\_v32)

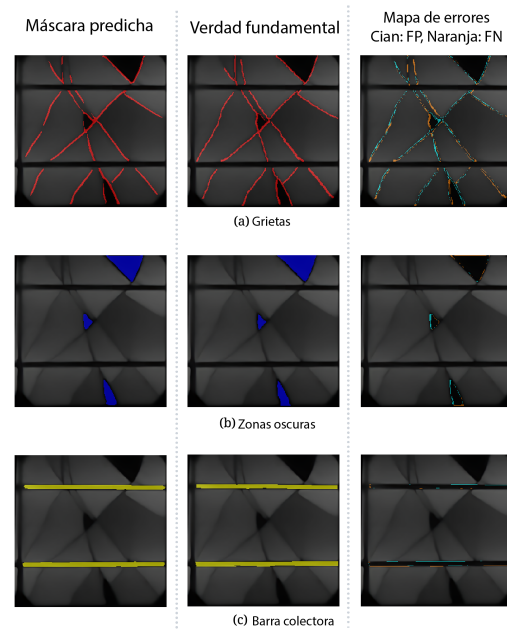
Clase de defecto	Precisión	Recall	F1-score
Grietas	0.6794	0.6465	0.6557
Zonas oscuras	0.8714	0.9066	0.9064
Barras colectoras	0.9151	0.9031	0.9083

La inspección cualitativa indicó, además, que las grietas ubicadas cerca de los bordes de la celda o parcialmente ocultas por la metalización de los dedos fueron omitidas con mayor frecuencia. Estas observaciones son coherentes con los resultados de evaluaciones comparativas obtenidos en conjuntos de datos de electroluminiscencia, como ELPV [19].

La Figura 5 ilustra resultados representativos de segmentación y su comparación con la verdad de terreno para los tres tipos de defectos analizados. Para cada clase, la figura presenta la imagen EL original, la máscara predicha superpuesta sobre la imagen, la máscara de verdad de terreno correspondiente y un mapa de errores que resalta las discrepancias entre ambas. En dichos mapas, el color cian indica falsos positivos (píxeles incorrectamente clasificados como defectuosos), mientras que el color naranja señala falsos negativos (píxeles defectuosos no detectados por el modelo).

En el caso de las grietas, los mapas de errores revelan bordes fragmentados, particularmente en regiones con luminiscencia débil, lo que corrobora las limitaciones identificadas en la evaluación cuantitativa. Por el contrario, para las zonas oscuras y los *busbars* (barras

colectoras), las regiones asociadas a falsos positivos y falsos negativos son mínimas, en concordancia con sus puntuaciones F1 comparativamente más elevadas.



**Figura 5.** Ejemplos de máscaras de segmentación predichas para (a) grietas, (b) zonas oscuras y (c) *busbars* (barras colectoras). Cada fila muestra la imagen EL original, la segmentación predicha superpuesta sobre la imagen, la máscara de verdad de terreno superpuesta y un mapa de errores, en el que el color cian indica falsos positivos (FP) y el color naranja indica falsos negativos (FN).

##### 3.1.2. Análisis de errores y desbalance de clases

Un desafío persistente durante el entrenamiento fue el desbalance de clases. Los píxeles de fondo representaron más del 85 % del conjunto de datos, mientras que las grietas constituyeron menos del 5 %. Este desbalance afectó tanto el proceso de optimización como la interpretación de las métricas, dado que la función de pérdida tendía a priorizar las clases mayoritarias.

Para mitigar este problema, se emplearon funciones de pérdida ponderadas por clase, asignando penalizaciones más altas a los errores asociados a las clases subrepresentadas. Además, el aumento de datos se orientó específicamente a incrementar el número de muestras que contenían grietas. En conjunto, estos ajustes mejoraron la sensibilidad del modelo hacia las clases minoritarias sin comprometer la estabilidad del entrenamiento.

Como se ilustra en la matriz de confusión normalizada presentada en la Figura 4, las tasas más altas de clasificación errónea se observaron en la clase de grietas, particularmente en casos en los que defectos

difusos o de bajo contraste fueron confundidos con zonas oscuras.

Los falsos positivos en la clase *busbar* (barras colectoras) aparecieron ocasionalmente cerca de los bordes de la celda. Estos errores se atribuyeron a variaciones abruptas de brillo, causadas por iluminación no uniforme o por reflexiones ópticas durante la adquisición de las imágenes.

### 3.2. Análisis comparativo entre versiones del modelo

#### 3.2.1. Evaluación iterativa de U-Net\_v24 a U-Net\_v32

Se llevó a cabo un proceso de evaluación iterativo a través de las versiones del modelo comprendidas entre U-Net\_v24 y U-Net\_v32. En cada iteración se introdujeron modificaciones controladas con el objetivo de mejorar el desempeño bajo condiciones limitadas de datos y hardware. Estos ajustes incluyeron el aumento de la resolución de entrada de  $128 \times 128$  a  $256 \times 256$  píxeles, la variación de las tasas de dropout en el rango de 0.15 a 0.3 y la reducción del tamaño de lote para acomodar entradas de mayor resolución dentro de la memoria GPU disponible.

La Tabla 3 resume la evolución de los puntajes F1 para las clases de grietas y *busbars*.

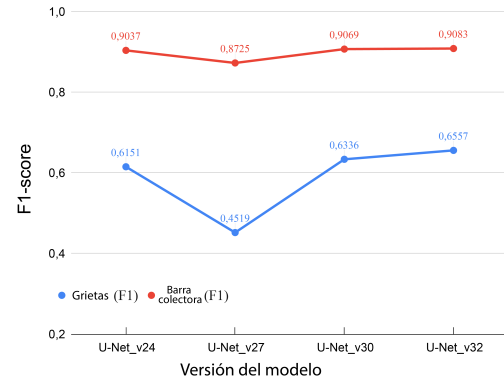
**Tabla 3.** Evolución del *F1-score* para la segmentación de grietas y *busbars*

Versión del modelo	Grietas (F1)	Barras colectoras (F1)
U-Net_v24	0.6151	0.9037
U-Net_v27	0.4519	0.8725
U-Net_v30	0.6336	0.9069
U-Net_v32	0.6557	0.9083

Las mejoras fueron más evidentes en la clase de grietas, para la cual las primeras versiones del modelo no lograban detectar patrones lineales tenues, particularmente en las proximidades de los límites de la celda. El uso de una mayor resolución de entrada, junto con configuraciones de codificador más profundas en U-Net\_v30 y U-Net\_v32, mejoró la capacidad de la red para preservar características detalladas durante el proceso de muestreo descendente.

La Figura 6 ilustra la progresión de los puntajes F1 a lo largo de las distintas versiones del modelo.

Estos resultados son coherentes con observaciones previas que indican que el refinamiento gradual de los hiperparámetros y de la estructura de entrada, más que la introducción de cambios arquitectónicos radicales, tiende a producir mejoras más fiables en modelos de segmentación entrenados con conjuntos de datos limitados o específicos del dominio.



**Figura 6.** Progresión del *F1-score* a lo largo de las versiones del modelo para las clases de grietas y *busbars*.

#### 3.2.2. Discusión sobre arquitecturas alternativas: U-Net++, MAU-Net

Se evaluaron dos variantes de U-Net, U-Net++ y MAU-Net, bajo las mismas condiciones de entrenamiento. U-Net++ incorpora conexiones de salto anidadas con el objetivo de mejorar el flujo del gradiente y la fusión semántica, mientras que MAU-Net introduce mecanismos de atención espacial y de canal para priorizar características relevantes y suprimir el ruido.

Ambas variantes mostraron una ligera mejora en el *recall* para la clase de grietas en regiones de bajo contraste; sin embargo, la ganancia global en *F1-score* fue marginal (inferior al 1.5 %) en comparación con U-Net\_v32. En particular, MAU-Net implicó, además, un incremento aproximado del 35 % en el tiempo de inferencia.

Dadas estas compensaciones, se optó por la U-Net base, ya que ofrece una mayor estabilidad, menor complejidad computacional y un desempeño más consistente entre los distintos tipos de defectos.

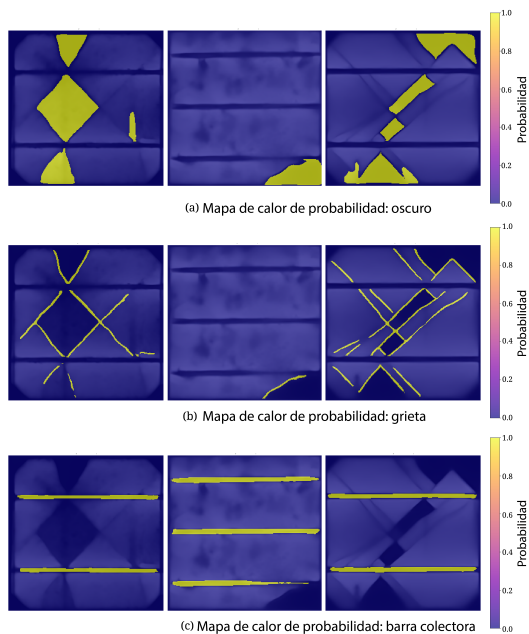
Dadas estas compensaciones, se optó por la U-Net base, ya que ofrece una mayor estabilidad, menor complejidad computacional y un desempeño más consistente entre los distintos tipos de defectos.

### 3.3. Validación visual y cualitativa

#### 3.3.1. Mapas de calor y superposiciones en celdas individuales

Además de las métricas cuantitativas, la inspección visual resulta esencial para comprender el comportamiento del modelo bajo condiciones diagnósticas realistas. Con este propósito, se generaron mapas de calor a partir de las probabilidades de salida softmax para cada clase de defecto, los cuales se superpusieron sobre las imágenes de electroluminiscencia (EL) originales con el fin de visualizar el grado de confianza asociado a las predicciones del modelo.

Como se ilustra en la Figura 7, las zonas oscuras fueron identificadas de manera consistente, atribuible a sus límites bien definidos y a sus características de contraste. En comparación, las predicciones correspondientes a las grietas omitieron ocasionalmente secciones fragmentadas, especialmente en regiones cercanas a los bordes de la celda o bajo la metalización de los dedos, donde los niveles de luminiscencia eran menores. Estas observaciones son coherentes con los hallazgos reportados en evaluaciones previas basadas en conjuntos de datos de electroluminiscencia, como ELPV [19], así como con enfoques que emplean superposiciones visuales para la verificación de resultados de segmentación [8].



**Figura 7.** Ejemplos de superposición de mapas de calor sobre imágenes EL: (a) zonas oscuras correctamente predichas; (b) detección parcial de grietas; y (c) sobrepredicción en el área del *busbar* (barra colectora).

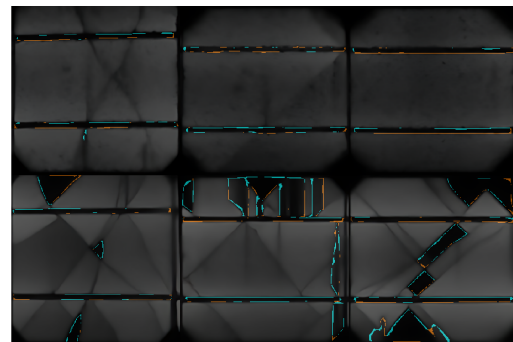
### 3.3.2. Evaluación de expertos y visualización de predicciones

Para complementar las métricas automáticas de desempeño, se llevó a cabo una revisión cualitativa por parte de tres expertos en diagnóstico fotovoltaico. Cada experto evaluó de manera independiente un conjunto de 30 máscaras de segmentación predichas, con énfasis en la exactitud de los contornos, la consistencia espacial y la interpretabilidad de los resultados. Los criterios de evaluación se alinearon con estándares de inspección visual adaptados de la norma IEC TS 60904-13 para el análisis de imágenes de electroluminiscencia (EL) [14].

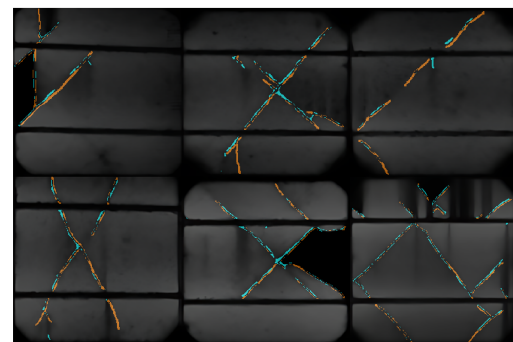
La mayoría de las predicciones correspondientes a zonas oscuras y *busbars* (barras colectoras) se consideraron altamente consistentes con las expectativas diag-

nósticas, presentando discrepancias de segmentación mínimas. La Figura 8 muestra mapas de errores representativos para estas clases, donde los píxeles en color cian indican falsos positivos y los píxeles en color naranja representan falsos negativos. En la mayoría de los casos, únicamente se observaron desalineamientos menores, que se manifestaron como trazos delgados asociados a ligeras diferencias en el grosor de los bordes entre las máscaras predichas y las anotaciones de la verdad de terreno.

En contraste, la segmentación de grietas resultó más desafiante. Algunas máscaras de predicción presentaron contornos irregulares o sobreestimaron las regiones defectuosas en áreas de bajo contraste. La Figura 9 proporciona ejemplos representativos de estos casos de error. En particular, trazos naranjas prominentes señalan segmentos de grietas no detectados (falsos negativos), mientras que las regiones en cian corresponden a predicciones de falsos positivos en zonas ambiguas. Estos patrones coincidieron con frecuencia con gradientes de electroluminiscencia inducidos por corrientes de polarización inconsistentes o por leves desalineamientos durante la adquisición de las imágenes.



**Figura 8.** Mapas de error representativos para zonas oscuras y *busbars* segmentados correctamente. El color cian indica falsos positivos y el color naranja denota falsos negativos.



**Figura 9.** Ejemplos de errores de segmentación. Las grietas no detectadas se muestran en color naranja, mientras que los falsos positivos en áreas de bajo contraste se indican en color cian.

Este análisis cualitativo reveló patrones útiles para orientar el refinamiento del modelo. Implementaciones futuras podrían beneficiarse de estrategias de posprocesamiento selectivo o de la incorporación de módulos basados en atención, orientados a reducir falsos positivos y a mejorar la resolución espacial en regiones donde los límites de los defectos se encuentran pobremente definidos.

En conjunto, la combinación de superposiciones visuales, mapas de calor basados en probabilidades y la evaluación por expertos ofreció una perspectiva multidimensional del desempeño del modelo, que complementó de manera efectiva métricas estándar como el *F1-score* y las matrices de confusión.

## 4. Conclusiones

Este estudio examinó arquitecturas convolucionales basadas en U-Net para la segmentación semántica de defectos en celdas fotovoltaicas de silicio cristalino utilizando imágenes de electroluminiscencia (EL). Se empleó un conjunto de datos de doble fuente, que incluyó imágenes EL internas y públicas, tanto para el entrenamiento como para la evaluación del modelo.

Los modelos demostraron un desempeño consistente en defectos con morfología clara y alto contraste, como las zonas oscuras y las discontinuidades en los *busbars* (barras colectoras). Las grietas permanecieron como la clase más desafiante de segmentar debido a su distribución espacial dispersa y a su geometría irregular, en concordancia con hallazgos reportados previamente [8, 18]. Desde una perspectiva práctica, los resultados obtenidos están sujetos a varias limitaciones.

En particular, muchas grietas se manifiestan como trayectorias delgadas e irregulares, con una densidad de píxeles muy baja y un contraste sutil respecto al material circundante, lo que conduce a la omisión de segmentos cortos o tenues y, en algunos casos, a la confusión entre fragmentos de grietas y texturas de fondo o artefactos de adquisición.

Adicionalmente, el redimensionamiento obligatorio de las imágenes EL de celdas individuales a una resolución de  $256 \times 256$  píxeles, requerido por el codificador basado en VGG16, puede comprimir las estructuras más delgadas en un número reducido de píxeles, limitando así la sensibilidad del modelo frente a defectos de muy pequeño tamaño.

Asimismo, el tamaño y la diversidad del conjunto de datos constituyen una restricción, dado que todas las imágenes fueron adquiridas bajo condiciones controladas de laboratorio y no incluyen escenarios de campo con corrientes de polarización variables, contaminación superficial o no uniformidades ópticas. Estos factores deben considerarse al extrapolar el desempeño reportado a otros montajes experimentales o condiciones

operativas. Ocho versiones del modelo fueron evaluadas mediante un flujo de entrenamiento estructurado que incorporó modificaciones progresivas en la profundidad del codificador, la resolución de entrada, la regularización y las estrategias de aprendizaje.

La configuración final, U-Net\_v32, ofreció el mejor compromiso entre exactitud de segmentación y eficiencia computacional. Arquitecturas alternativas, como U-Net++ y MAU-Net, proporcionaron únicamente mejoras marginales en el desempeño, a costa de un incremento en el tiempo de inferencia [11].

Las métricas cuantitativas, las visualizaciones de salida y las evaluaciones cualitativas realizadas por expertos se emplearon de manera complementaria para analizar el comportamiento del modelo. Esta validación integrada confirmó la utilidad de los modelos basados en U-Net para la localización de defectos en imágenes EL bajo condiciones controladas.

Como líneas de trabajo futuro, se propone (1) ampliar el conjunto de datos mediante la incorporación de imágenes EL adquiridas en condiciones de campo que reflejen la variabilidad del entorno real y (2) desarrollar e integrar técnicas de posprocesamiento orientadas a reducir falsos positivos y a mejorar la definición de contornos en regiones de bajo contraste. Estos avances contribuirían a incrementar la aplicabilidad de los modelos de segmentación en diagnósticos fotovoltaicos operativos y en procesos de control de calidad durante la fabricación.

## Agradecimientos

Los autores expresan su agradecimiento al Laboratorio de Micro-Red de la Universidad de Cuenca por proporcionar las instalaciones y el apoyo técnico necesarios para el desarrollo de este estudio. Asimismo, se reconoce de manera especial al personal del laboratorio, Vinicio Íñiguez, Edisson Villa y Pablo J. Delgado, por su valiosa asistencia durante la fase experimental.

## Rol de autores

- **Franklin Gómez-López:** conceptualización, curación de datos, análisis formal, investigación, metodología, desarrollo de software, validación, visualización, escritura – borrador original.
- **Danny Ochoa-Correa:** análisis formal, adquisición de financiación, investigación, administración de proyecto, recursos, supervisión, escritura – revisión y edición.
- **Isabel Cabrera-Carrera:** Curación de datos, análisis formal, investigación, metodología, supervisión, escritura – revisión y edición.

## Referencias

- [1] K. G. Bedrich, M. Bliss, T. R. Betts, and R. Gottschalg, “Electroluminescence imaging of PV devices: Determining the image quality,” in *2015 IEEE 42nd Photovoltaic Specialist Conference (PVSC)*. IEEE, Jun. 2015, pp. 1–5. [Online]. Available: <https://doi.org/10.1109/PVSC.2015.7356011>
- [2] T. Fuyuki and A. Kitiyanan, “Photographic diagnosis of crystalline silicon solar cells utilizing electroluminescence,” *Applied Physics A*, vol. 96, no. 1, pp. 189–196, Dec. 2008. [Online]. Available: <http://doi.org/10.1007/s00339-008-4986-0>
- [3] M. Akram and J. Bai, “Defect detection in photovoltaic modules based on image-to-image generation and deep learning,” *Sustainable Energy Technologies and Assessments*, vol. 82, p. 104441, Oct. 2025. [Online]. Available: <http://doi.org/http://doi.org/>
- [4] J. Wang, L. Bi, P. Sun, X. Jiao, X. Ma, X. Lei, and Y. Luo, “Deep-learning-based automatic detection of photovoltaic cell defects in electroluminescence images,” *Sensors*, vol. 23, no. 1, p. 297, Dec. 2022. [Online]. Available: <http://doi.org/10.3390/s23010297>
- [5] Q. Liu, M. Liu, C. Wang, and Q. J. Wu, “An efficient CNN-based detector for photovoltaic module cells defect detection in electroluminescence images,” *Solar Energy*, vol. 267, p. 112245, Jan. 2024. [Online]. Available: <http://doi.org/10.1016/j.solener.2023.112245>
- [6] J. Fioresi, D. J. Colvin, R. Frota, R. Gupta, M. Li, H. P. Seigneur, S. Vyas, S. Oliveira, M. Shah, and K. O. Davis, “Automated defect detection and localization in photovoltaic cells using semantic segmentation of electroluminescence images,” *IEEE Journal of Photovoltaics*, vol. 12, no. 1, pp. 53–61, Jan. 2022. [Online]. Available: <http://doi.org/10.1109/jphotov.2021.3131059>
- [7] O. Ronneberger, P. Fischer, and T. Brox, *U-Net: Convolutional Networks for Biomedical Image Segmentation*. Springer International Publishing, 2015, pp. 234–241. [Online]. Available: [http://doi.org/10.1007/978-3-319-24574-4\\_28](http://doi.org/10.1007/978-3-319-24574-4_28)
- [8] L. Pratt, D. Govender, and R. Klein, “Defect detection and quantification in electroluminescence images of solar PV modules using U-net semantic segmentation,” *Renewable Energy*, vol. 178, pp. 1211–1222, Nov. 2021. [Online]. Available: <http://doi.org/10.1016/j.renene.2021.06.086>
- [9] H. Eesaar, S. Joe, M. U. Rehman, Y. Jang, and K. T. Chong, “SEiPV-net: An efficient deep learning framework for autonomous multi-defect segmentation in electroluminescence images of solar photovoltaic modules,” *Energies*, vol. 16, no. 23, p. 7726, Nov. 2023. [Online]. Available: <https://doi.org/10.3390/en16237726>
- [10] M. R. U. Rahman and H. Chen, “Defects inspection in polycrystalline solar cells electroluminescence images using deep learning,” *IEEE Access*, vol. 8, pp. 40 547–40 558, 2020. [Online]. Available: <http://doi.org/10.1109/access.2020.2976843>
- [11] R. A. Mamun Rudro, K. Nur, F. A. Al Sohan, M. Mridha, S. Alfarhood, M. Safran, and K. Kanagarathinam, “SPF-Net: Solar panel fault detection using U-net based deep learning image classification,” *Energy Reports*, vol. 12, pp. 1580–1594, Dec. 2024. [Online]. Available: <http://doi.org/10.1016/j.egyr.2024.07.044>
- [12] S. Deitsch, V. Christlein, S. Berger, C. Buerhop-Lutz, A. Maier, F. Gallwitz, and C. Riess, “Automatic classification of defective photovoltaic module cells in electroluminescence images,” *Solar Energy*, vol. 185, pp. 455–468, Jun. 2019. [Online]. Available: <http://doi.org/10.1016/j.solener.2019.02.067>
- [13] U. Hijjawi, S. Lakshminarayana, T. Xu, and M. Rahman, “A benchmarking study of instance segmentation methods for photovoltaic cell defect detection using electroluminescence images,” *Solar Energy*, vol. 303, p. 114083, Jan. 2026.
- [14] IEC, *IEC Technical Specification 60904-13. Photovoltaic devices - Part 13: Electroluminescence of photovoltaic modules*. International Electrotechnical Commission, 2018. [Online]. Available: <http://doi.org/10.4229/35thEUPVSEC20182018-5CV.3.15>
- [15] Supervisely. (2023) Supervisely computer vision platform. Supervisely OÜ. [Online]. Available: <https://upsalesiana.ec/ing35ar8r2>
- [16] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014.
- [17] A. Sohail, N. Ul Islam, A. Ul Haq, S. Ul Islam, I. Shafi, and J. Park, “Fault detection and computation of power in PV cells under faulty conditions using deep-learning,” *Energy Reports*, vol. 9, pp. 4325–4336, Dec. 2023.
- [18] S. Deitsch, C. Buerhop-Lutz, E. Sovetkin, A. Steiland, A. Maier, F. Gallwitz, and C. Riess, “Segmentation of photovoltaic module cells in uncali-

- brated electroluminescence images,” *Machine Vision and Applications*, vol. 32, no. 4, May 2021.
- [19] C. Buerhop-Lutz, S. Deitsch, A. Maier, F. Gallwitz, S. Berger, B. Doll, J. Hauch, C. Camus, and C. Brabec, “A benchmark for visual identification of defective solar cells in electroluminescence imagery,” *35th European Photovoltaic Solar Energy Conference and Exhibition; 1287-1289*, 2018. [Online]. Available: <http://doi.org/10.4229/35THEUPVSEC20182018-5CV.3.15>