



CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS AVANZADOS
DEL INSTITUTO POLITÉCNICO NACIONAL

UNIDAD ZACATENCO
DEPARTAMENTO DE COMPUTACIÓN

Segmentación en superpíxeles sin reducción de dimensionalidad para la clasificación semi-supervisada de imágenes hiperespectrales

Tesis que presenta

Ángel Alonso Galarza Chávez

para obtener el Grado de

**Maestro en Ciencias
en Computación**

Director de Tesis

Dra. Xiaou Li

Resumen

La clasificación de imágenes hiperespectrales mediante aprendizaje profundo enfrenta la escasez de muestras etiquetadas, dado el alto costo de etiquetar píxeles. El aprendizaje semi-supervisado sobre redes convolucionales de grafos sortea esta restricción propagando la información de pocos nodos etiquetados a través de un grafo de superpíxeles. Sin embargo, los algoritmos de segmentación del estado del arte exigen una reducción de dimensionalidad previa que suaviza las firmas espectrales que distinguen clases similares, por lo que el grafo hereda nodos que mezclan clases, y bajo escasez de etiquetas la GCN no puede corregirlos. Esta tesis propone PCNN-SAM, un algoritmo de segmentación bio-inspirado que incorpora la métrica *Spectral Angle Mapper* en la función de enlace de una red neuronal de pulsos acoplados, sustituyendo el peso sináptico fijo del modelo clásico por una compuerta espectral: un píxel se agrega a una región solo si su divergencia angular respecto a ella es suficientemente pequeña. Los superpíxeles se generan a partir del vector espectral completo, sin reducción de dimensionalidad previa, y se acoplan como nodos de un clasificador GCN semi-supervisado, dando lugar a la arquitectura híbrida PCNN-GCN. Evaluada con cinco muestras etiquetadas por clase sobre seis conjuntos de datos de referencia frente a SLIC, LSC, ERS y SEEDS, PCNN-GCN ocupa el segundo lugar en exactitud global (70.75 % frente a 72.80 % de SLIC). El resultado clave es otro: ninguna estrategia domina en todas las escenas, y PCNN-GCN obtiene su mayor ventaja en los dos conjuntos con mayor correlación espectral entre clases y mayor fragmentación espacial, pues la varianza explicada predice la exactitud casi el doble de bien que la exactitud de segmentación alcanzable. La elección del algoritmo depende, por tanto, de si la dificultad de la escena es geométrica o espectral.

Palabras clave: imágenes hiperespectrales, segmentación en superpíxeles, redes neuronales de pulsos acoplados, *Spectral Angle Mapper*, redes convolucionales de grafos, aprendizaje semi-supervisado, escasez de etiquetas.

Abstract

Deep learning classification of hyperspectral images is limited by the high cost of labelling pixels, leaving very few annotated samples available in practice. Semi-supervised graph convolutional networks address this by propagating information from a handful of labelled nodes across a superpixel graph. However, state-of-the-art segmentation algorithms require a prior dimensionality reduction that smooths away the subtle spectral signatures distinguishing similar classes, so the graph inherits class-mixing nodes that the GCN cannot correct under label scarcity. This thesis proposes PCNN-SAM, a bio-inspired segmentation algorithm that embeds the Spectral Angle Mapper metric into the linking function of a pulse-coupled neural network, replacing its fixed synaptic weight with a spectral gate: a pixel joins a region only if its angular divergence with respect to it is sufficiently small. Superpixels are generated from the complete spectral vector, without prior dimensionality reduction, and coupled as nodes of a semi-supervised GCN, yielding the hybrid PCNN-GCN architecture. Evaluated with five labelled samples per class on six benchmark datasets against SLIC, LSC, ERS and SEEDS, PCNN-GCN ranks second overall (70.75 % accuracy versus 72.80 % for SLIC), so no general superiority holds. Crucially, no single strategy dominates across all scenes: PCNN-GCN attains its largest advantage, the widest margin in the study, on the two datasets with the highest spectral correlation between classes and the greatest spatial fragmentation, this is because explained variance, which measures how much spectral information survives within a superpixel, predicts classification accuracy nearly twice as well as achievable segmentation accuracy, which measures geometric purity. The optimal segmentation algorithm therefore depends on whether scene difficulty is geometric or spectral in nature.

Keywords: hyperspectral images, superpixel segmentation, pulse-coupled neural networks, Spectral Angle Mapper, graph convolutional networks, semi-supervised learning, label scarcity.

Agradecimientos

Agradezco a SECIHTI por el apoyo económico brindado a través de la beca de posgrado, que me permitió dedicarme de tiempo completo a esta investigación y concluir mis estudios de maestría.

Agradezco también a los investigadores y al personal del Departamento de Computación por sus enseñanzas y su respaldo constante a lo largo de mi estancia en la maestría.

A mis padres y hermanos, por impulsarme siempre a seguir estudiando; a mis amigos, presenciales y virtuales, y a mis compañeros del Departamento de Computación, por hacer de este camino una experiencia grata y compartida.

Por último, y con especial cariño, agradezco a mi asesora, la Dra. Xiaou Li, por darme la oportunidad de trabajar a su lado y confiar en mí desde el primer momento, así como por sus enseñanzas tanto académicas como de vida que hicieron posible este trabajo de investigación.

Índice general

Resumen	i
Abstract	iii
Agradecimientos	v
1. Introducción	1
1.1. Antecedentes	1
1.2. Motivación	2
1.3. Planteamiento del problema	3
1.3.1. Hipótesis	5
1.3.2. Objetivos	5
1.4. Contribuciones	6
1.5. Organización del documento	7
2. Teledetección y segmentación en superpíxeles	9
2.1. Adquisición de imágenes de teledetección	9
2.1.1. Reflectancia espectral	11
2.1.2. Plataformas y sensores de adquisición	12
2.1.3. Imágenes térmicas	13
2.1.4. Imágenes multiespectrales	13
2.2. Imagen hiperespectral y similitud espectral	14
2.2.1. Representación de la imagen hiperespectral	14
2.2.2. Similitud espectral	15
2.3. Tipos de imágenes digitales	16
2.4. Corrección de errores de adquisición	17
2.5. Reducción de dimensionalidad	18
2.6. Segmentación de imágenes	18
2.6.1. Superpíxeles	19
2.6.2. Algoritmos de superpíxeles	20
2.7. Segmentación bio-inspirada mediante PCNN	22
2.8. Métricas de evaluación de la segmentación	25

3. Clasificación de imágenes hiperespectrales	29
3.1. Aprendizaje no supervisado	30
3.2. Aprendizaje supervisado	31
3.2.1. Métodos clásicos	31
3.2.2. Redes neuronales convolucionales	32
3.3. Aprendizaje semi-supervisado	35
3.3.1. Redes neuronales de grafos	36
3.3.2. Redes convolucionales de grafos	38
3.4. Métricas de evaluación de la clasificación	43
4. Clasificación semi-supervisada mediante PCNN-GCN	45
4.1. Planteamiento experimental	45
4.2. Método propuesto	46
4.2.1. Visión general del método	46
4.2.2. Conjunto de datos	48
4.2.3. Normalización espectral	52
4.2.4. Segmentación mediante PCNN-SAM	53
4.2.5. Métodos de segmentación de referencia	58
4.2.6. De superpíxeles a grafo	58
4.2.7. Clasificador GCN semi-supervisado sobre superpíxeles	59
4.3. Configuración experimental	59
4.3.1. Protocolo de muestreo semi-supervisado	60
4.3.2. Métricas y protocolo de reporte	62
4.4. Selección de la escala de superpíxeles (K)	62
4.5. Resultados y análisis	64
4.5.1. Calidad de la segmentación por método	65
4.5.2. Desempeño de clasificación semi-supervisada	71
4.5.3. Relación entre segmentación y clasificación	75
4.5.4. Análisis visual de los mapas de clasificación	77
4.6. Discusión general	82
4.6.1. Interpretación integrada	82
5. Conclusiones y trabajo futuro	87
5.1. Conclusiones	87
5.2. Discusión	90
5.3. Trabajo futuro	92

Índice de figuras

1.1. Origen del problema abordado en esta tesis	4
1.2. Organización del documento	8
2.1. Interacción de la energía electromagnética incidente con la superficie de un objeto: fracciones reflejada, absorbida y transmitida.	10
2.2. Curvas de reflectancia espectral características de vegetación, suelo, nieve, agua y otros materiales en función de la longitud de onda.	11
2.3. Representación de una imagen hiperespectral como un hipercubo $\mathbf{X} \in \mathbb{R}^{B \times (N \times M)}$ y la firma espectral $\mathbf{x}_i \in \mathbb{R}^B$ de un píxel individual.	15
2.4. Interpretación geométrica de la métrica SAM como el ángulo α que forman dos vectores espectrales.	16
2.5. Arquitectura de una neurona PCNN	24
3.1. Los tres paradigmas de aprendizaje según la supervisión disponible	30
3.2. Estrategias de adaptación de las CNN a datos hiperespectrales	34
3.3. Vecindario de un nodo y radio de propagación en una GNN	37
3.4. Actualización del estado de un nodo en una capa de GNN	37
3.5. Aprendizaje semi-supervisado transductivo sobre un grafo	40
4.1. Flujo de trabajo propuesto: de la normalización del cubo hiperespectral a la clasificación semi-supervisada mediante GCN.	47
4.2. Conjunto de datos Indian Pines: composición en color e imagen de referencia (<i>ground truth</i>) con sus 16 clases.	48
4.3. Conjunto de datos Salinas: composición en color e imagen de referencia (<i>ground truth</i>) con sus 16 clases.	49
4.4. Conjunto de datos Pavia University: composición en color e imagen de referencia (<i>ground truth</i>) con sus 9 clases.	49
4.5. Conjunto de datos WHU-Hi-LongKou: composición en color e imagen de referencia (<i>ground truth</i>) con sus 9 clases.	50
4.6. Conjunto de datos WHU-Hi-HanChuan: composición en color e imagen de referencia (<i>ground truth</i>) con sus 16 clases.	51

4.7. Conjunto de datos WHU-Hi-HongHu: composición en color e imagen de referencia (<i>ground truth</i>) con sus 22 clases.	52
4.8. Firmas espectrales de un píxel representativo por cada clase: (a) perfiles originales sin transformación; (b) perfiles tras la normalización <i>min-max</i> independiente por banda (Ecuación (4.1)).	53
4.9. Ejemplo de la matriz de tiempo de disparo T_{ij} : el orden de activación de los píxeles delimita la formación de las regiones de los superpíxeles.	55
4.10. Dinámica de una neurona en PCNN-SAM	57
4.11. Correspondencia entre superpíxeles y nodos del grafo: cada región se abstracta como un nodo cuya característica es la firma espectral media, en 10 componentes principales, de los píxeles que la componen.	59
4.12. Evolución de la exactitud global (OA) promedio, sobre los seis conjuntos de datos y ambas arquitecturas de GCN, en función de la escala de superpíxeles K	63
4.13. Tiempo de ejecución promedio de la etapa de segmentación en función de la escala de superpíxeles K	64
4.14. Partición en superpíxeles sobre Indian Pines con los métodos a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM.	66
4.15. Partición en superpíxeles sobre Salinas con los métodos a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM.	67
4.16. Partición en superpíxeles sobre Pavia University con los métodos a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM.	68
4.17. Partición en superpíxeles sobre WHU-Hi-LongKou con los métodos a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM.	68
4.18. Partición en superpíxeles sobre WHU-Hi-HanChuan con los métodos a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM.	69
4.19. Partición en superpíxeles sobre WHU-Hi-HongHu con los métodos a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM.	70
4.20. Exactitud global por conjunto de datos y método de segmentación	71
4.21. Relación entre las métricas de segmentación y la exactitud de clasificación	77
4.22. Mapas de clasificación sobre Indian Pines: a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM, f) referencia (<i>ground truth</i>).	78
4.23. Mapas de clasificación sobre Salinas: a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM, f) referencia (<i>ground truth</i>).	78
4.24. Mapas de clasificación sobre Pavia University: a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM, f) referencia (<i>ground truth</i>).	79
4.25. Mapas de clasificación sobre WHU-Hi-LongKou: a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM, f) referencia (<i>ground truth</i>).	80
4.26. Mapas de clasificación sobre WHU-Hi-HanChuan: a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM, f) referencia (<i>ground truth</i>).	81

4.27. Mapas de clasificación sobre WHU-Hi-HongHu: a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM, f) referencia (<i>ground truth</i>).	81
4.28. Los seis conjuntos de datos situados por tipo de dificultad	83

Índice de tablas

1.1. Correspondencia entre las contribuciones de esta tesis y las secciones en las que se desarrollan y se evalúan.	7
2.1. Plataformas y dispositivos de adquisición de HSI	13
2.2. Comparación conceptual de los algoritmos de superpíxel evaluados en el Capítulo 4	22
3.1. Métodos clásicos de clasificación supervisada y sus limitaciones sobre HSI .	32
3.2. Compromiso entre información explotada y costo computacional en las cuatro estrategias de CNN para HSI	35
3.3. CNN frente a GNN/GCN: supuestos sobre la estructura de los datos	39
3.4. Arquitecturas GCN semi-supervisadas para HSI, clasificadas según la etapa del flujo de trabajo en la que intervienen	43
4.1. Propiedades de los conjuntos de datos hiperespectrales utilizados	48
4.2. Correspondencia entre las cuatro modificaciones introducidas por PCNN-SAM, las ecuaciones que las formalizan y las líneas del Algoritmo 4.1 en las que se aplican	57
4.3. Parámetros de PCNN-SAM empleados en la evaluación experimental	61
4.4. Configuración del clasificador GCN semi-supervisado	62
4.5. Métricas de segmentación consolidadas (ASA, BR, EV) por conjunto de datos y método (K=500)	65
4.6. Métricas de clasificación consolidadas (OA, AA, κ) por conjunto de datos y método (K=500)	72
4.7. Resumen del desempeño de clasificación: conjuntos de datos ganados por métrica y promedio global sobre los seis conjuntos	73
4.8. Correlación de Pearson entre las métricas de segmentación y la exactitud global (OA), agregada sobre los 30 pares dataset $\times$ método (n=30)	75
4.9. Correlación de Pearson entre ASA y OA, desagregada por conjunto de datos (n=5 por dataset)	76
4.10. Los cuatro regímenes identificados: método ganador, correlación ASA-OA y mecanismo dominante	84

Capítulo 1

Introducción

1.1. Antecedentes

La teledetección mediante imágenes hiperespectrales (*hyperspectral images*, HSI) ha emergido como una herramienta esencial, ya que su alta resolución espectral facilita la identificación detallada de la composición de los materiales en la superficie terrestre [1]. Históricamente, la clasificación de estas imágenes se realizaba mediante algoritmos de aprendizaje automático clásico, tales como k -vecinos más cercanos (K-NN), máquinas de vectores de soporte (SVM) y bosques aleatorios [2]. Sin embargo, en la última década, los algoritmos de aprendizaje profundo, particularmente las redes neuronales convolucionales (*convolutional neural networks*, CNN), han demostrado resultados superiores gracias a su capacidad para extraer y explotar jerarquías de características espacial-espectrales [3].

Pese a estos avances, la principal limitación operativa no reside en la capacidad de los modelos sino en la disponibilidad de datos: la recolección de las imágenes y, especialmente, su etiquetado a nivel de píxel constituyen una tarea intrínsecamente costosa en recursos y tiempo, lo que ocasiona una escasez estructural de datos etiquetados [3]. Esta escasez es el hilo conductor que articula las decisiones metodológicas del resto de esta tesis.

Para reducir el número de elementos sobre los que debe operar un clasificador, los métodos del estado del arte emplean técnicas de sobresegmentación que agrupan píxeles en superpíxeles, sustituyendo la retícula original por un grafo de regiones. No obstante, algoritmos ampliamente utilizados como *simple linear iterative clustering* (SLIC) [4], *linear spectral clustering* (LSC) [5], *entropy rate superpixel* (ERS) [6] o *superpixels extracted via energy-driven sampling* (SEEDS) [7] operan típicamente mediante distancias euclidianas $\|\mathbf{x}_i - \mathbf{x}_j\|_2$ sobre espacios de color de baja dimensión. Al aplicarlos a HSI se requiere una reducción de dimensionalidad previa, como el análisis de componentes principales (*principal component analysis*, PCA) [8] o el análisis de discriminante lineal (*linear discriminant analysis*, LDA) [9], para mantener la viabilidad del algoritmo. Esta compresión altera las firmas espectrales sutiles y genera superpíxeles con impureza semántica, es decir, regiones que mezclan varias clases en un mismo nodo. La literatura reciente sobre segmentación de

HSI documenta con detalle esta dependencia [10], y explora alternativas que la atenúan mediante perfiles estructurales multiescala [11] o representaciones bi-estocásticas a nivel de superpíxel [12].

Como alternativa a los algoritmos basados en gradientes, las redes neuronales de pulsos acoplados (*pulse-coupled neural networks*, PCNN) ofrecen un mecanismo de crecimiento de regiones biológicamente inspirado y libre de entrenamiento por retropropagación [13]. Si bien las PCNN estándar operan sobre imágenes monocromáticas, su potencial para HSI radica en que su función de enlace y su función de alimentación son modificables [14, 15], lo que abre la posibilidad de incorporar mecanismos específicos para el análisis espectral dentro de la propia dinámica de la red.

Garantizar la pureza de los superpíxeles resulta crítico precisamente por la escasez de etiquetas señalada arriba [3]. En respuesta a esta limitación, el paradigma de aprendizaje semi-supervisado mediante redes convolucionales de grafos (*graph convolutional networks*, GCN) ha ganado popularidad. Trabajos recientes proponen arquitecturas como X-GPN para extraer dependencias topológicas multiescala [16], mientras que otras investigaciones integran capas prototípicas para mejorar la exactitud de clasificación [17]. La combinación de superpíxeles como nodos y redes de grafos se ha consolidado desde entonces en múltiples variantes: grafos dispersos construidos sobre una segmentación por cuencas [18], esquemas que prescinden deliberadamente del grafo de superpíxeles por considerar que una región puede degradar la discriminación entre clases [19], y arquitecturas que combinan grafos regionales con atención global [20].

1.2. Motivación

Las HSI se han consolidado como una herramienta de uso extendido en teledetección gracias a su capacidad para capturar información espectral detallada de la superficie terrestre. Sus aplicaciones abarcan ámbitos tan diversos como el monitoreo ambiental, la agricultura de precisión, el análisis astronómico y la industria alimentaria [21], y en la actualidad se explotan principalmente mediante técnicas de aprendizaje automático y aprendizaje profundo [22].

En el contexto ambiental, la información espectral y espacial conjunta que proporcionan las HSI resulta esencial para el diagnóstico y monitoreo de fenómenos naturales y antropogénicos: la clasificación automática de estas imágenes permite la detección temprana de zonas contaminadas, alteraciones en ecosistemas o anomalías que indiquen riesgos potenciales. El valor práctico de estas aplicaciones depende, sin embargo, de que la clasificación sea confiable en condiciones realistas, es decir, cuando solo se dispone de unas pocas muestras etiquetadas por clase.

La extracción de características profundas a nivel de píxel resulta computacionalmente prohibitiva debido al volumen de información espectral. Por ello, los métodos del estado del arte agrupan píxeles espacial y espectralmente similares en superpíxeles, lo cual reduce

drásticamente la redundancia de los datos y la complejidad computacional. El problema es que estos algoritmos de segmentación fueron diseñados originalmente para imágenes tricromáticas convencionales, de modo que su aplicación a HSI obliga a comprimir la dimensión espectral antes de segmentar [4, 6, 7]. La reducción de dimensionalidad sigue siendo un campo activo por sí mismo, y los estudios que la evalúan de manera controlada muestran que el criterio de compresión afecta de forma medible a la exactitud posterior [23].

Esta observación motiva la propuesta de esta tesis: desarrollar un método de segmentación basado en PCNN que integre una evaluación de vectores espectrales dentro de su comportamiento de sincronización, de manera que los superpíxeles se generen a partir de la firma espectral nativa en toda su dimensionalidad. Al estructurar estas regiones como nodos de entrada de una GCN, se busca mejorar la exactitud de clasificación sin renunciar a la reducción del costo computacional que aporta la sobresegmentación.

1.3. Planteamiento del problema

La clasificación de HSI mediante aprendizaje profundo ha demostrado un rendimiento superior en la caracterización de la superficie terrestre, pero su adopción operativa enfrenta un cuello de botella fundamental: la escasez de datos etiquetados. La anotación a nivel de píxel es una tarea manual, lenta y costosa, y esta carencia entra en conflicto directo con las arquitecturas profundas tradicionales, que exigen grandes volúmenes de datos etiquetados y son propensas al sobreajuste en escenarios de escasez [17].

Para mitigar esta limitante, el aprendizaje semi-supervisado mediante GCN se ha posicionado como la estrategia dominante [24, 16, 25, 26]. El interés por operar con muy pocas muestras anotadas no se restringe a los modelos de grafos: también se aborda mediante pseudoetiquetado sobre redes convolucionales multiescala [27] y mediante fusión ponderada de características guiada por superpíxeles [28]. Estos modelos compensan la falta de etiquetas propagando información a través de una estructura topológica construida a partir de superpíxeles. Ahora bien, el éxito de una GCN semi-supervisada es estrictamente dependiente de la pureza espectral de los nodos que conforman su grafo base: si un nodo mezcla dos clases, ninguna cantidad de propagación posterior puede separarlas.

Aquí surge el conflicto que da origen a esta tesis. Los algoritmos de segmentación del estado del arte no pueden operar directamente en espacios de alta dimensionalidad y requieren una reducción de dimensionalidad previa que elimina características espectrales sutiles. En consecuencia, al aplicar métodos de segmentación tradicionales sobre HSI se generan topologías de grafo defectuosas; cuando estas estructuras se introducen en una GCN bajo escasez de etiquetas, el error de subsegmentación se amplifica y se propaga durante el entrenamiento, degradando de manera significativa la exactitud global del clasificador. La Figura 1.1 contrasta esquemáticamente este flujo con el que propone esta tesis, y localiza el punto exacto en el que se pierde la información espectral.

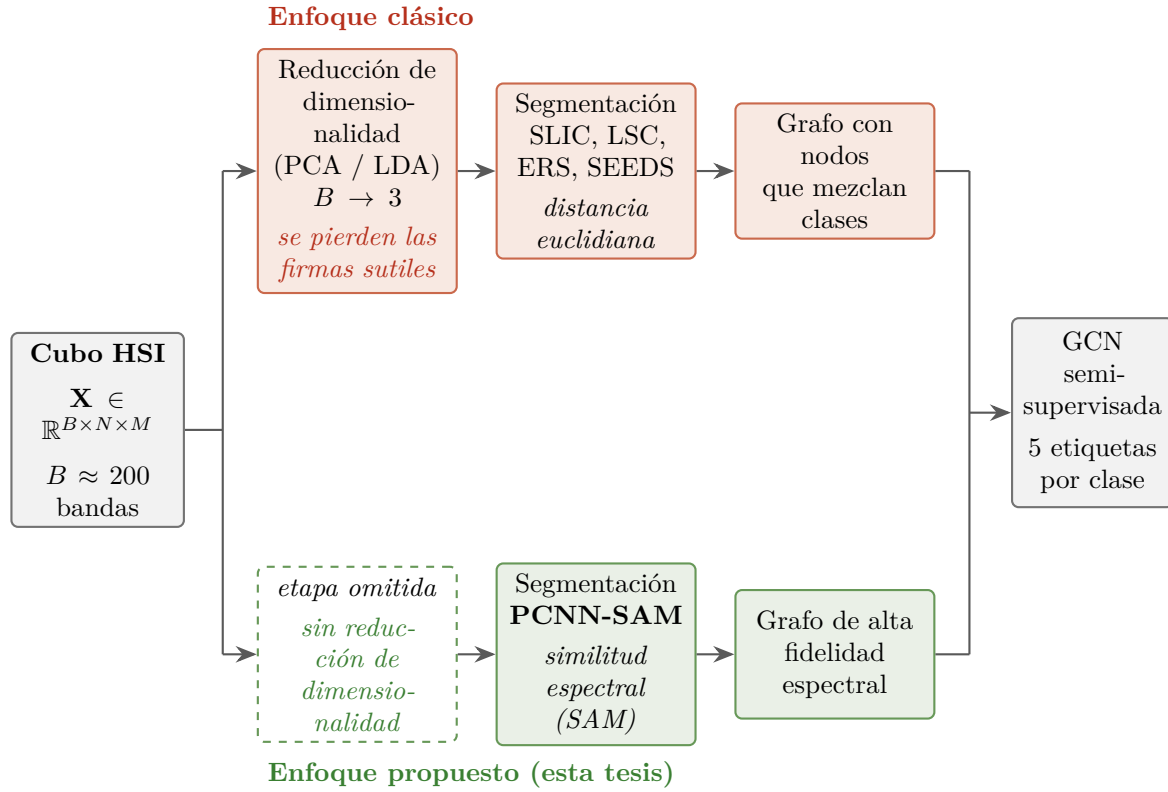


Figura 1.1: Origen del problema abordado en esta tesis. En el flujo clásico, la reducción de dimensionalidad que exigen los algoritmos de superpíxeles para operar sobre HSI degrada las firmas espectrales *antes* de segmentar, de modo que el grafo hereda nodos que mezclan clases. PCNN-SAM suprime esa etapa y segmenta sobre el vector espectral completo, delegando cualquier compresión a un momento posterior a la construcción de la topología.

1.3.1. Hipótesis

Ante la problemática descrita, esta tesis plantea la siguiente hipótesis de trabajo.

La integración de un mecanismo de evaluación de vectores espectrales como comportamiento de enlace dentro de una red neuronal de pulsos acoplados permitirá generar superpíxeles con mayor pureza espacial-espectral al operar directamente sobre la imagen hiperespectral. Al estructurar estos superpíxeles de alta fidelidad como nodos topológicos en una red convolucional de grafos, se incrementará significativamente la exactitud global de clasificación en escenarios de aprendizaje semi-supervisado con escasez de datos etiquetados, superando el rendimiento de los enfoques de segmentación del estado del arte basados en métricas euclidianas.

Con el fin de que esta hipótesis sea contrastable de manera inequívoca, se descompone en dos predicciones verificables por separado:

- P₁** *Superioridad general.* PCNN-GCN superará a SLIC, LSC, ERS y SEEDS en exactitud global promediada sobre el conjunto completo de escenas evaluadas.
- P₂** *Ventaja condicional.* La ventaja de PCNN-GCN será mayor en las escenas cuya dificultad sea de naturaleza espectral —alta correlación entre las firmas de clases distintas— que en aquellas cuya dificultad sea de naturaleza geométrica.

Esta distinción es deliberada: como se documenta en el Capítulo 4 y se discute en el Capítulo 5, la evidencia experimental obtenida no respalda **P₁** pero sí respalda **P₂**, y es precisamente ese contraste el que constituye el hallazgo central de este trabajo.

1.3.2. Objetivos

Objetivo general

Diseñar, implementar y evaluar una arquitectura híbrida que acople un algoritmo de segmentación en superpíxeles guiado por similitud espectral, basado en redes neuronales de pulsos acoplados, con un clasificador semi-supervisado basado en redes convolucionales de grafos, con el fin de determinar bajo qué condiciones espaciales y espectrales la preservación del vector espectral completo durante la segmentación mejora la clasificación de imágenes hiperespectrales con escasez de datos etiquetados.

Objetivos específicos

1. Modificar la dinámica de acoplamiento de una PCNN para que la propagación de la activación entre neuronas vecinas quede condicionada por la similitud espectral medida con SAM, de manera que la segmentación opere sobre el vector espectral completo sin reducción de dimensionalidad previa.

2. Formular un mecanismo de competencia entre regiones en crecimiento y un procedimiento de consolidación morfológica que garanticen una partición de la imagen en superpíxeles conexos y sin píxeles sin asignar.
3. Construir, a partir de las regiones segmentadas, un grafo de adyacencia espacial cuyos nodos se caractericen mediante la firma espectral media de cada región.
4. Implementar y entrenar un clasificador GCN semi-supervisado transductivo sobre dicho grafo bajo un régimen de cinco muestras etiquetadas por clase.
5. Determinar empíricamente la escala de superpíxeles que ofrece el mejor balance entre exactitud de clasificación y costo computacional.
6. Comparar la arquitectura propuesta frente a SLIC, LSC, ERS y SEEDS sobre seis conjuntos de datos de referencia, evaluando tanto la calidad de la segmentación (ASA, BR, EV) como la del clasificador (OA, AA, κ).
7. Cuantificar la relación estadística entre las métricas de calidad de segmentación y la exactitud de clasificación final, con el fin de identificar qué propiedad de un superpíxel resulta determinante para la propagación semi-supervisada.

1.4. Contribuciones

Para poner a prueba la hipótesis anterior, este trabajo diseña y evalúa una arquitectura híbrida que integra una PCNN guiada por similitud espectral, para la generación de superpíxeles, con una GCN semi-supervisada, para la clasificación bajo escasez de etiquetas. Esta arquitectura se concreta en cuatro contribuciones.

En primer lugar, se desarrolla **PCNN-SAM**, un algoritmo de segmentación espacial-espectral basado en el comportamiento de sincronización de las PCNN. Se trata de una formulación bio-inspirada que incorpora la comparación de vectores espectrales completos dentro de la dinámica de acoplamiento de la red, lo cual permite extraer superpíxeles directamente de la imagen original sin recurrir a una reducción de dimensionalidad previa.

En segundo lugar, las regiones segmentadas se modelan como estructuras topológicas: a partir de las firmas espectrales medias de los superpíxeles se inicializan tanto la matriz de características como la matriz de adyacencia del grafo resultante.

En tercer lugar, se implementa y entrena un clasificador GCN semi-supervisado que emplea dicha topología para propagar etiquetas en escenarios con datos de entrenamiento limitados, acoplando el algoritmo de segmentación bio-inspirado con el clasificador de grafos en una única arquitectura híbrida, **PCNN-GCN**.

Finalmente, la arquitectura resultante se somete a un estudio comparativo frente a cuatro métodos de segmentación del estado del arte sobre seis conjuntos de datos hiperespectrales de referencia, seleccionados para cubrir un espectro amplio de condiciones espaciales y

espectrales. De este estudio se desprende una contribución adicional de carácter metodológico: la evidencia de que las métricas geométricas de segmentación de uso habitual no son un buen predictor del desempeño de un clasificador basado en grafos, y de que la retención de información espectral lo predice considerablemente mejor.

La Tabla 1.1 indica dónde se desarrolla y se evalúa cada una de estas contribuciones.

Tabla 1.1: Correspondencia entre las contribuciones de esta tesis y las secciones en las que se desarrollan y se evalúan.

#	Contribución	Se desarrolla en	Se evalúa en
C ₁	Algoritmo de segmentación PCNN-SAM	Sec. 4.2.4	Sec. 4.5.1
C ₂	Construcción del grafo de superpíxeles	Sec. 4.2.6	Sec. 4.5.2
C ₃	Clasificador GCN semi-supervisado	Sec. 4.2.7	Sec. 4.5.2
C ₄	Estudio comparativo y análisis por régimen	Sec. 4.3	Secs. 4.5.3–4.6

1.5. Organización del documento

El resto de este documento se organiza como sigue. El Capítulo 2 presenta los fundamentos de la teledetección y de las imágenes hiperespectrales, y aborda el procesamiento y la segmentación de imágenes, describiendo los algoritmos de superpíxeles del estado del arte (SLIC, LSC, ERS y SEEDS) y los fundamentos de las redes neuronales acopladas por pulsos. El Capítulo 3 presenta los fundamentos de clasificación de HSI organizados según el tipo de supervisión disponible durante el entrenamiento —aprendizaje no supervisado, supervisado y semi-supervisado—, con énfasis en las redes neuronales convolucionales y en los fundamentos de las redes neuronales de grafos sobre los que se construye el esquema semi-supervisado adoptado en esta tesis. El Capítulo 4 describe la arquitectura híbrida propuesta, la configuración experimental y los resultados obtenidos sobre los seis conjuntos de datos evaluados, junto con su análisis y discusión. Finalmente, el Capítulo 5 presenta las conclusiones del trabajo y las líneas de trabajo futuro derivadas de los hallazgos obtenidos. La Figura 1.2 resume esta estructura y las dependencias entre capítulos.

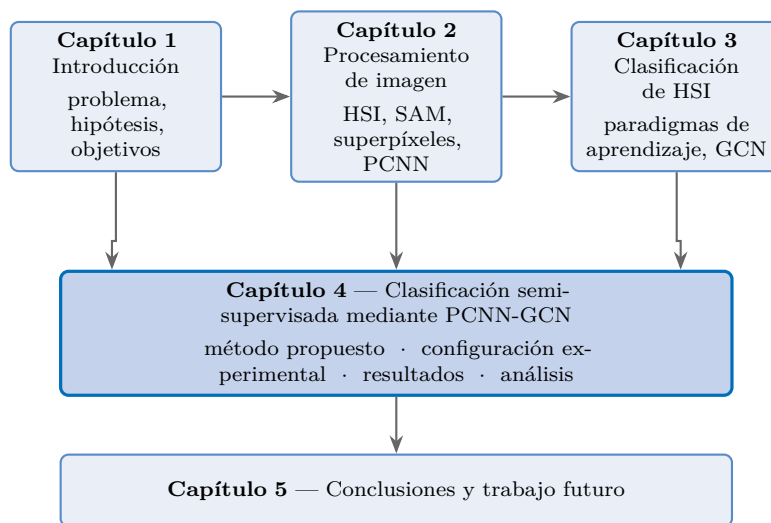


Figura 1.2: Organización del documento y dependencias entre capítulos. Los Capítulos 2 y 3 aportan, respectivamente, los fundamentos de segmentación y de clasificación sobre los que se construye la propuesta del Capítulo 4; el Capítulo 5 retoma las predicciones P_1 y P_2 formuladas en el presente capítulo.

Capítulo 2

Fundamentos de teledetección y segmentación en superpíxeles

Este capítulo reúne los fundamentos necesarios para la propuesta de segmentación presentada en el Capítulo 4: por un lado, los principios de la teledetección que explican cómo se adquiere una imagen hiperespectral y qué información codifica; por otro, las técnicas de procesamiento y segmentación de imágenes —clásicas y bio-inspiradas— mediante las cuales dicha imagen se estructura en superpíxeles.

El capítulo se organiza en dos partes. La primera (Sección 2.1 y Sección 2.2) presenta los fundamentos de teledetección: el principio físico de la reflectancia espectral, las plataformas y sensores de adquisición, los distintos tipos de imagen de teledetección y, en particular, la imagen hiperespectral y la métrica de similitud espectral sobre la que se construye la propuesta de esta tesis. La segunda parte (Sección 2.3 a Sección 2.8) aborda el procesamiento de imágenes propiamente dicho: la corrección de errores comunes en HSI, la reducción de dimensionalidad, los algoritmos de segmentación por superpíxeles del estado del arte, y los fundamentos de la segmentación bio-inspirada mediante PCNN sobre los que se construye PCNN-SAM.

2.1. Adquisición de imágenes de teledetección

La teledetección es la ciencia de obtener información sobre un objeto, área o fenómeno mediante el análisis de datos adquiridos por dispositivos que no están en contacto físico con dicho objeto, área o fenómeno [1]. Esta información puede obtenerse de distintas formas, como variaciones en las distribuciones de fuerza, ondas acústicas o distribuciones de energía electromagnética; esta última es la que se emplea en la adquisición de imágenes multiespectrales e hiperespectrales, y es en la que se enfoca esta sección [29].

Un proceso generalizado de teledetección electromagnética de recursos terrestres involucra dos etapas generales: la adquisición y el análisis de los datos. La etapa de adquisición comprende la captura de información de la superficie terrestre —las fuentes de energía,

su propagación a través de la atmósfera y su interacción con las características de la superficie— mediante sensores que se operan desde plataformas aéreas y espaciales para asistir en el inventario, mapeo y monitoreo de los recursos terrestres [1, 29]. Posteriormente, en la etapa de análisis, la información recolectada se examina mediante diversos dispositivos y métodos de observación para su interpretación final [29].

Cuando la energía electromagnética incide sobre un objeto de la superficie terrestre, son posibles tres interacciones fundamentales entre la energía y dicho objeto: una fracción de la energía incidente es reflejada, otra es absorbida y una tercera es transmitida a través del objeto. Denotando estas tres cantidades como $E_R(\lambda)$, $E_A(\lambda)$ y $E_T(\lambda)$ respectivamente —cada una función de la longitud de onda λ — y aplicando el principio de conservación de la energía, la energía total incidente $E_I(\lambda)$ sobre un cuerpo puede expresarse mediante la conservación de la energía radiante [29]:

$$E_I(\lambda) = E_R(\lambda) + E_A(\lambda) + E_T(\lambda). \quad (2.1)$$

La proporción relativa de energía reflejada, absorbida y transmitida no es constante: varía en función de la longitud de onda y del material sobre el que incide, y es precisamente esta variación diferenciada la que constituye el fundamento físico que permite distinguir un material de otro mediante teledetección. La Figura 2.1 ilustra estas tres interacciones de manera esquemática.

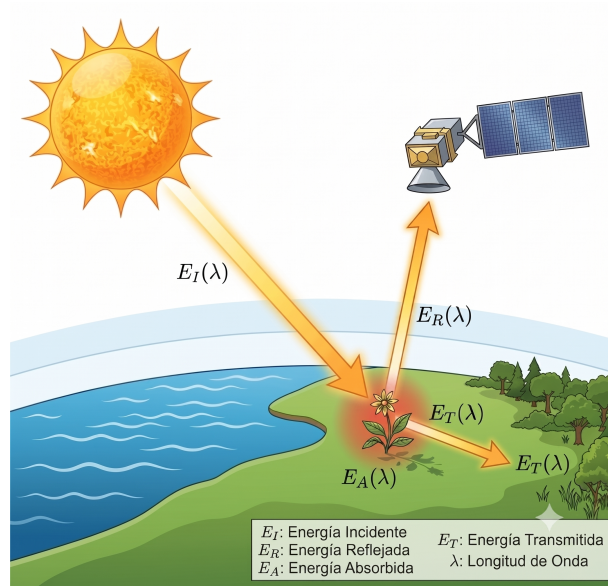


Figura 2.1: Interacción de la energía electromagnética incidente con la superficie de un objeto: fracciones reflejada, absorbida y transmitida.

2.1.1. Reflectancia espectral

De las tres interacciones energéticas descritas en la Ecuación (2.1), la reflexión es la de mayor interés para la mayoría de los sistemas de teledetección óptica, ya que es la energía reflejada la que finalmente es captada por el sensor [29]. Despejando la energía reflejada de la Ecuación (2.1) se obtiene:

$$E_R(\lambda) = E_I(\lambda) - [E_A(\lambda) + E_T(\lambda)]. \quad (2.2)$$

Es decir, la energía reflejada por un material en una longitud de onda dada es igual a la energía incidente en esa misma longitud de onda, menos la porción que dicho material absorbe y transmite. Cuanto mayor sea la capacidad de un material de absorber o transmitir energía en una longitud de onda particular, menor será la energía que refleje en ella, y viceversa.

Para poder comparar la respuesta de distintos materiales de forma independiente de la intensidad de la fuente de iluminación, se define la reflectancia espectral ρ_λ como el cociente entre la energía reflejada y la energía incidente en cada longitud de onda, expresada como porcentaje [29]:

$$\rho_\lambda = \frac{E_R(\lambda)}{E_I(\lambda)} \times 100. \quad (2.3)$$

Al evaluar ρ_λ sobre un rango continuo de longitudes de onda para un mismo material, se obtiene su curva de reflectancia espectral. La forma de esta curva —los rangos de longitud de onda en los que un material refleja más o menos energía— es característica de su composición física y química, y es precisamente esta variación la que permite distinguir un material de otro a partir de datos de teledetección [29]. La Figura 2.2 ilustra este comportamiento para tipos de cobertura frecuentes en aplicaciones HSI.

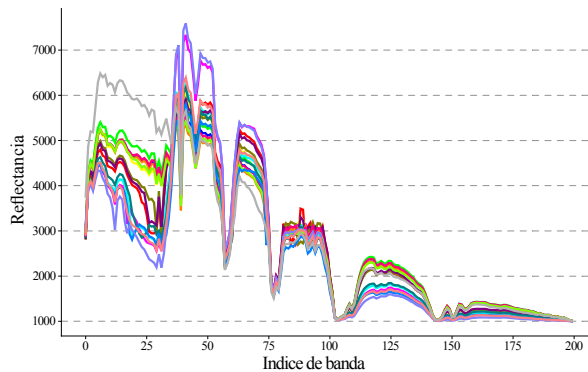


Figura 2.2: Curvas de reflectancia espectral características de vegetación, suelo, nieve, agua y otros materiales en función de la longitud de onda.

Algunos patrones característicos de reflectancia espectral, relevantes para la interpretación de HSI en aplicaciones de monitoreo ambiental, son los siguientes [29]:

- **Vegetación:** la vegetación sana presenta una fuerte absorción de energía en las regiones azul y roja del espectro visible —debido a la clorofila—, lo que produce un valle de reflectancia visible característico, seguido de un incremento abrupto en el infrarrojo cercano.
- **Suelo:** el patrón de reflectancia del suelo es, en general, más plano que el de la vegetación y tiende a incrementarse de forma gradual con la longitud de onda; la presencia de humedad y materia orgánica reduce significativamente su reflectancia.
- **Nieve:** la nieve exhibe una reflectancia muy alta en las longitudes de onda visibles e infrarrojas cercanas, pero esta cae drásticamente en el infrarrojo medio debido a una fuerte absorción.
- **Agua:** el agua presenta el patrón más distintivo, con una fuerte absorción de energía a partir del infrarrojo cercano; como resultado, los cuerpos de agua aparecen prácticamente negros en imágenes capturadas en esas bandas.

Esta variación diferenciada de reflectancia entre materiales es, en última instancia, el principio físico que hace posible construir la firma espectral de un píxel —introducida formalmente en la Sección 2.2.1— y, con ella, evaluar la similitud espectral entre píxeles mediante métricas como la que se presenta en la Sección 2.2.2.

2.1.2. Plataformas y sensores de adquisición

La adquisición de imágenes multiespectrales e hiperespectrales requiere del uso de plataformas y sensores especializados, cuya elección determina directamente la resolución espacial, espectral y temporal disponible para una aplicación dada [30]. La Tabla 2.1 resume las cuatro categorías principales de plataformas de adquisición, los sensores representativos de cada una y sus resoluciones espaciales típicas [31, 32].

Tabla 2.1: Plataformas y dispositivos de adquisición de HSI

Plataforma	Sensores representativos	Resolución espacial
Satélite	EO-1 Hyperion, PROBA-CHRIS	Media (30 m)
	DESI, HISUI, PRISMA, EnMAP, SHALOM	Media a baja (10–30 m)
Avión / helicóptero	AVIRIS, CASI, HyMap, AISA Eagle	Media a alta (4–20 m)
	Probe-1, RDACS-H4, AHS-160, PHI	Submétrica
UAV (dron)	Multirrotores, ala fija Sensores ligeros (p. ej. Headwall)	Alta (metro a centímetro)
Corto alcance	Sistemas de laboratorio o de tierra	Muy alta (centímetro a subcentímetro)
	Sensores en andamios, camiones y etapas lineales	

Entre estas plataformas destacan los sensores aerotransportados AVIRIS y ROSIS, así como los sensores montados en UAV de la familia Headwall Nano-Hyperspec [33], ya que son los que capturaron los seis conjuntos de datos utilizados en la evaluación experimental de esta tesis.

2.1.3. Imágenes térmicas

A diferencia de las imágenes multiespectrales e hiperespectrales, que se basan en la radiación solar reflejada por la superficie, la obtención de imágenes infrarrojas térmicas se centra en la detección de la radiación que los propios materiales emiten [1, 34, 35]. Debido a que no depende de la iluminación solar, este tipo de sensores puede operarse en cualquier momento del día o de la noche [36].

Las imágenes térmicas se han utilizado con éxito en diversas aplicaciones de monitoreo ambiental y geológico, entre las que se incluyen la determinación del tipo y estructura de rocas, la localización de fallas geológicas, el mapeo de tipos de suelo y de humedad, la localización de fugas en canales de riego, y el estudio de las características térmicas de cuerpos volcánicos y de la evapotranspiración de la vegetación [1]. La mayoría de estas aplicaciones son de naturaleza cualitativa, aunque algunas requieren de un análisis cuantitativo para determinar temperaturas absolutas [1].

2.1.4. Imágenes multiespectrales

Las imágenes multiespectrales capturan la reflectancia de la superficie terrestre en un número reducido de bandas espectrales —típicamente entre 3 y 10—, cada una de las cuales integra un rango relativamente amplio del espectro electromagnético [29, 34]. Sensores multiespectrales de amplio uso, como los de las familias Landsat y Sentinel-2, han sido la base de aplicaciones de monitoreo ambiental, agricultura y cartografía de uso de suelo durante varias décadas [1].

Si bien las imágenes multiespectrales resultan suficientes para distinguir clases de cobertura terrestre con firmas espectrales considerablemente distintas entre sí (por ejemplo, agua, vegetación y suelo desnudo), su capacidad se ve limitada al intentar diferenciar materiales espectralmente similares entre sí, como distintas especies vegetales o distintos minerales, ya que el reducido número de bandas no captura las variaciones sutiles de reflectancia que estos materiales presentan a lo largo del espectro. Esta limitación es, precisamente, la que motiva el uso de imágenes hiperespectrales, presentadas en la siguiente sección.

2.2. La imagen hiperespectral y la similitud espectral

2.2.1. Representación de la imagen hiperespectral

Las HSI extienden el principio de las imágenes multiespectrales llevándolo a su expresión más detallada: en lugar de un puñado de bandas anchas, una HSI recolecta información a través de cientos de bandas espectrales, angostas y contiguas, lo que permite construir un espectro de reflectancia prácticamente continuo para cada píxel de la imagen [1, 33, 30, 32]. Esta riqueza espectral es la que permite identificar materiales que, bajo un sensor multiespectral, resultarían indistinguibles entre sí [29].

Una imagen hiperespectral se representa formalmente como un cubo tridimensional, o hipercubo, denotado como $\mathbf{X} \in \mathbb{R}^{B \times (N \times M)}$, donde B es el número de bandas espectrales y N, M son las dimensiones espaciales de la imagen (alto y ancho, respectivamente) [37]. Bajo esta representación, cada píxel de la imagen corresponde a un vector espectral $\mathbf{x}_i \in \mathbb{R}^B$, cuyas B componentes describen la reflectancia del material presente en ese píxel a lo largo de todo el rango espectral capturado por el sensor; a este vector se le denomina la firma espectral del píxel. La Figura 2.3 ilustra esta estructura.

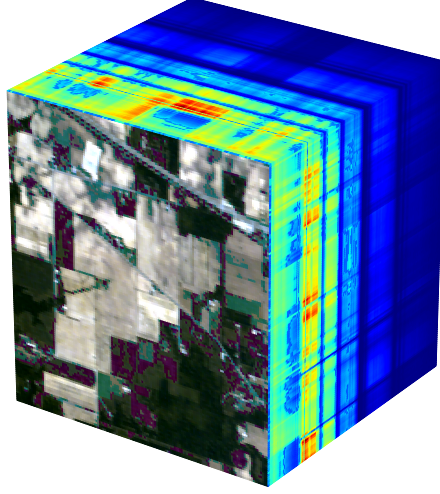


Figura 2.3: Representación de una imagen hiperespectral como un hipercubo $\mathbf{X} \in \mathbb{R}^{B \times (N \times M)}$ y la firma espectral $\mathbf{x}_i \in \mathbb{R}^B$ de un píxel individual.

2.2.2. Similitud espectral

Dado que la firma espectral $\mathbf{x}_i \in \mathbb{R}^B$ de un píxel describe de manera prácticamente única la composición del material que representa, es posible cuantificar qué tan parecidos son dos píxeles evaluando la similitud entre sus firmas espectrales, en lugar de recurrir únicamente a distancias geométricas en un espacio de color de baja dimensión. Esta noción de similitud espectral es central para los algoritmos de segmentación de HSI que operan directamente sobre el vector espectral completo.

Una de las métricas de similitud espectral más utilizadas es el *Spectral Angle Mapper* (SAM), que parte de la idea de que dos firmas espectrales que pertenecen a un mismo material son, en esencia, versiones escaladas una de la otra —la escala refleja diferencias de iluminación o de geometría de captura, no de composición—, por lo que dos vectores espectrales $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^B$ pueden compararse evaluando el ángulo que forman entre sí, en lugar de la distancia euclidiana entre ellos [38, 39]:

$$\text{SAM}(\mathbf{x}_i, \mathbf{x}_j) = \arccos \left(\frac{\mathbf{x}_i^T \mathbf{x}_j}{\|\mathbf{x}_i\| \|\mathbf{x}_j\|} \right). \quad (2.4)$$

El valor de $\text{SAM}(\mathbf{x}_i, \mathbf{x}_j)$ es un ángulo, en radianes, entre los dos vectores espectrales: cuanto más cercano a cero, mayor es la similitud espectral entre los píxeles evaluados, independientemente de las diferencias de magnitud (intensidad) entre ambos vectores. Esta invariancia a la magnitud es la propiedad que hace de SAM una métrica particularmente

robusta frente a variaciones de iluminación dentro de una misma escena [38]. La Figura 2.4 ilustra geoméricamente esta interpretación.

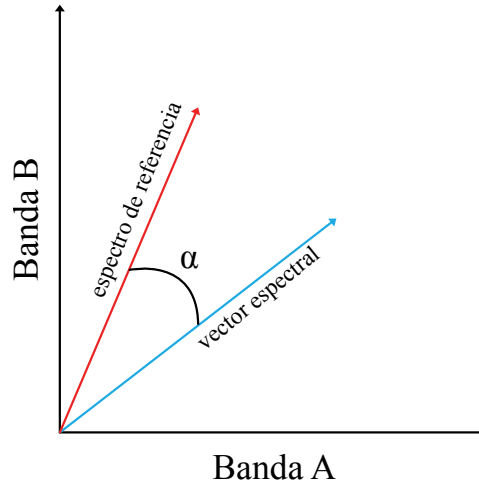


Figura 2.4: Interpretación geométrica de la métrica SAM como el ángulo α que forman dos vectores espectrales.

SAM no es la única métrica de similitud espectral reportada en la literatura; existen alternativas como la Divergencia de información espectral (SID) y esquemas híbridos que combinan ambas [40]. Sin embargo, es SAM la métrica que se adopta en esta tesis como mecanismo de acoplamiento espectral dentro del algoritmo de segmentación PCNN-SAM, descrito en la Sección 4.2.4.

2.3. Tipos de imágenes digitales

El procesamiento de imágenes comprende, a su vez, el conjunto de técnicas computacionales aplicadas a una imagen digital con el fin de mejorar su calidad, corregir defectos introducidos durante su adquisición, o extraer de ella información estructurada y utilizable por etapas posteriores de análisis. Dentro de este campo se distinguen, entre otras, tareas de restauración y realce (mejorar la calidad visual o corregir defectos de la imagen), corrección geométrica y radiométrica (alineal y calibrar la imagen respecto a la realidad física que representa), y segmentación (particionar la imagen en regiones que faciliten su interpretación) [41]. El resto de este capítulo se centra en tres de estas tareas: la corrección de errores comunes en la adquisición de HSI, la reducción de dimensionalidad, y la segmentación de imágenes mediante superpíxeles, incluyendo tanto los algoritmos clásicos del estado del arte como los fundamentos de la segmentación bio-inspirada mediante redes neuronales acopladas por pulsos sobre los que se construye la propuesta de esta tesis.

Desde el punto de vista de su representación digital, una imagen es una función discretizada $I(x, y)$ que asigna un valor de intensidad a cada posición espacial (x, y) de una

rejilla regular de píxeles [41]. De acuerdo con el número de valores que esta función toma en cada posición —es decir, el número de bandas o canales que describen a cada píxel—, las imágenes digitales pueden clasificarse en las siguientes categorías:

- **Imágenes binarias:** cada píxel toma únicamente uno de dos valores posibles (0 o 1); se utilizan típicamente para representar máscaras o el resultado de una segmentación.
- **Imágenes en escala de grises:** cada píxel se describe mediante un único valor de intensidad, generalmente cuantizado en 8 bits (256 niveles).
- **Imágenes a color:** cada píxel se describe mediante tres valores de intensidad, correspondientes a las bandas roja, verde y azul del espectro visible.
- **Imágenes multiespectrales e hiperespectrales:** cada píxel se describe mediante decenas o cientos de valores de intensidad, uno por cada banda espectral capturada por el sensor, tal como se presentó en las Sección 2.1.4 y Sección 2.2.1.

Esta clasificación es relevante para el procesamiento de imágenes porque muchas de las técnicas de corrección, realce y segmentación fueron desarrolladas originalmente para imágenes en escala de grises o a color, y su extensión a imágenes multiespectrales e hiperespectrales —con decenas o cientos de bandas en lugar de una o tres— no siempre es directa.

2.4. Corrección de errores de adquisición

La adquisición de imágenes de teledetección está sujeta a diversas fuentes de error que deben corregirse antes de que la imagen pueda utilizarse de manera confiable en tareas de análisis o clasificación. A continuación se describen dos de las correcciones más comunes en el preprocesamiento de HSI.

Corrección geométrica

Las distorsiones geométricas ocurren cuando la posición de un objeto en la imagen capturada no corresponde con precisión a su posición real sobre la superficie terrestre, debido a factores como el movimiento de la plataforma, el relieve del terreno o la geometría de la óptica del sensor. Una forma habitual de corregir este error es mediante el uso de puntos de control terrestre, ubicaciones de coordenadas conocidas con precisión que se utilizan como referencia para transformar geoméricamente la imagen y alinearla con la superficie real [29, 2].

Eliminación de nubes

La presencia de nubes durante la captura contamina la información de reflectancia en las bandas afectadas. Existen diversas estrategias para mitigar este problema, desde sensores que detectan directamente la presencia de nubes y su reflectancia asociada, hasta técnicas de análisis que identifican las bandas afectadas para su corrección o descarte, o el uso de múltiples adquisiciones de una misma zona para reconstruir la información terrestre correcta [2].

2.5. Reducción de dimensionalidad y pérdida de información espectral

Como se estableció en el planteamiento del problema, gran parte de los algoritmos de segmentación del estado del arte no pueden operar directamente sobre el espacio hiperespectral completo, debido tanto a su costo computacional como a que fueron diseñados originalmente para espacios de color de baja dimensión. Para conciliar ambas necesidades, es práctica común aplicar una técnica de reducción de dimensionalidad sobre la HSI antes de segmentarla, generando una representación compacta que aproxima la información espectral original en un número reducido de componentes.

La técnica más utilizada con este fin es el PCA. El cual busca una transformación lineal del espacio original $\mathbb{R}^B$ a un nuevo espacio $\mathbb{R}^{B'}$, con $B' \ll B$, cuyos ejes (componentes principales) están ordenados de manera que el primero captura la mayor varianza posible de los datos, el segundo captura la mayor varianza posible bajo la restricción de ser ortogonal al primero, y así sucesivamente [8]. Al proyectar cada píxel sobre los primeros B' componentes principales se conserva la mayor parte de la varianza total del conjunto de datos utilizando muchas menos dimensiones que el espacio espectral original.

Una alternativa es el LDA, el cual, a diferencia de PCA, no busca maximizar la varianza total de los datos sino la separabilidad entre las clases conocidas, lo cual requiere disponer de información de clase durante la reducción [9].

Si bien esta compresión permite que los algoritmos de segmentación clásicos operen de manera computacionalmente viable sobre HSI, tiene un costo: al conservar únicamente las primeras componentes se descartan variaciones espectrales más sutiles, que pueden ser precisamente las que distinguen entre clases espectralmente similares.

2.6. Segmentación de imágenes

La segmentación de imagen es el proceso de particionar una imagen en un conjunto de regiones —o, en el caso de las técnicas de sobresegmentación descritas en la Sección 2.6.1, en un conjunto de superpíxeles— que sean internamente homogéneas respecto a algún criterio de similitud (intensidad, color, textura o, en el caso de HSI, firma espectral), y que

en conjunto cubran la totalidad de la imagen sin traslaparse [41]. Esta partición reduce la complejidad estructural de la escena y facilita tareas posteriores como la clasificación, al permitir operar sobre un número reducido de regiones en lugar de sobre cada píxel de manera individual.

Los algoritmos de segmentación pueden agruparse, de manera general, en cuatro paradigmas según el criterio que utilizan para definir las fronteras entre regiones [41]:

- **Basados en bordes**, que identifican discontinuidades abruptas de intensidad y las utilizan como fronteras entre regiones.
- **Basados en regiones**, que agrupan iterativamente píxeles vecinos que satisfacen un criterio de homogeneidad, haciendo crecer regiones a partir de semillas iniciales o refinando progresivamente sus fronteras.
- **Basados en agrupamiento**, que tratan la segmentación como un problema de agrupamiento de puntos en un espacio de características que combina información de color/espectro y posición espacial.
- **Basados en grafos**, que representan la imagen como un grafo de píxeles y formulan la segmentación como un problema de optimización sobre dicho grafo.

Como se detalla en la Sección 2.6.2, de los algoritmos de superpíxeles evaluados en esta tesis, SLIC y LSC pertenecen al paradigma de agrupamiento, SEEDS al de regiones, y ERS al basado en grafos. La segmentación bio-inspirada presentada en la Sección 2.7 constituye un paradigma adicional, ajeno a esta clasificación, basado en la dinámica de sincronización de redes neuronales pulsantes.

2.6.1. Superpíxeles

Los superpíxeles son el resultado de aplicar una sobresegmentación a una imagen: en lugar de dividirla en un pequeño número de regiones semánticamente significativas —como haría una segmentación completa—, se agrupan los píxeles vecinos y perceptualmente similares en un número relativamente grande de regiones compactas, cada una de las cuales contiene típicamente información redundante de un solo material o clase [4]. Esta sustitución de la rejilla de píxeles individuales por un conjunto de superpíxeles reduce drásticamente el número de elementos primitivos sobre los que debe operar un algoritmo posterior —como un clasificador—, disminuyendo el costo computacional y, al promediar la información dentro de cada región, atenuando el ruido presente a nivel de píxel individual [4, 42].

En el contexto de esta tesis, los superpíxeles cumplen además un segundo propósito: constituyen la unidad topológica —el nodo— sobre la que se construye el grafo de entrada al clasificador basado en GCN presentado en el Capítulo 3. La calidad de esta sobresegmentación —qué tan bien cada superpíxel respeta las fronteras reales entre clases y cuánta información espectral conserva— determina directamente la calidad de dicho grafo.

2.6.2. Algoritmos de superpíxeles del estado del arte

Esta sección describe los cuatro algoritmos de superpíxeles frente a los que se compara experimentalmente la propuesta de esta tesis en el Capítulo 4: SLIC, ERS, LSC y SEEDS. Los cuatro comparten una característica metodológicamente relevante: fueron diseñados originalmente para operar sobre imágenes de color de tres bandas o su equivalente en un espacio de color perceptual como CIELAB, por lo que su aplicación directa sobre HSI requiere primero de una reducción de dimensionalidad como la descrita en la Sección 2.5.

Simple Linear Iterative Clustering (SLIC)

SLIC es una adaptación del algoritmo K-means para la generación de superpíxeles, que combina en una sola medida de distancia ponderada la proximidad en el espacio de color CIELAB y la proximidad espacial [4]. El algoritmo inicia posicionando k centros de clúster $C_i = [l_i, a_i, b_i, x_i, y_i]^T$ sobre una rejilla regular espaciada por un intervalo

$$S = \sqrt{N/k}, \quad (2.5)$$

donde N es el número total de píxeles de la imagen; estos centros se reposicionan posteriormente al píxel de menor gradiente dentro de una vecindad de 3×3 , con el fin de evitar inicializarlos sobre un borde o un píxel ruidoso.

A diferencia de K-means estándar, SLIC restringe la búsqueda del vecino más cercano de cada píxel a una región de $2S \times 2S$ centrada en cada centro de clúster, lo cual reduce considerablemente la complejidad del algoritmo respecto a una búsqueda global. Dentro de esa región, la asignación de cada píxel al clúster más cercano se realiza mediante la distancia combinada [4]:

$$D_s = d_{lab} + \frac{m}{S} d_{xy}, \quad (2.6)$$

donde $d_{lab} = \sqrt{(l_k - l_i)^2 + (a_k - a_i)^2 + (b_k - b_i)^2}$ es la distancia de color en el espacio CIELAB, $d_{xy} = \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2}$ es la distancia espacial euclidiana, y m es un parámetro de compacidad que pondera el peso relativo de ambos términos: valores de m mayores favorecen superpíxeles más compactos y regulares, a costa de una menor adherencia a los bordes reales de la imagen. Tras cada asignación, los centros de clúster se actualizan como el promedio de los píxeles asignados, y el proceso se repite iterativamente hasta la convergencia.

Entropy Rate Superpixel (ERS)

ERS formula la generación de superpíxeles como un problema de selección de un subconjunto de aristas sobre un grafo, en lugar de como un problema de agrupamiento en un espacio de características [6]. La imagen se representa como un grafo $G = (V, E)$, donde

cada píxel es un vértice y las aristas conectan píxeles vecinos con un peso proporcional a su similitud. El algoritmo busca el subconjunto de aristas $A \subseteq E$ que maximiza la función objetivo

$$E(A) = H(A) + \lambda B(A), \quad (2.7)$$

sujeto a la restricción de que el subgrafo inducido por A tenga exactamente K componentes conexas. El primer término, $H(A)$, es la tasa de entropía de un paseo aleatorio sobre el grafo inducido por A , y favorece la formación de clústeres compactos y homogéneos; el segundo término, $B(A)$, es un término de equilibrio que favorece clústeres de tamaño similar entre sí [6]. Debido a que $E(A)$ es una función submodular y monótona, ERS la optimiza mediante un algoritmo voraz (*greedy*) que, partiendo de un conjunto vacío de aristas, añade en cada iteración la arista que produce la mayor ganancia en $E(A)$, hasta alcanzar el número deseado de componentes conexas [6].

Linear Spectral Clustering (LSC)

LSC formula la generación de superpíxeles como una instancia del problema de *normalized cuts*: una partición del grafo de la imagen en k componentes que minimiza el costo de las aristas cortadas entre componentes, normalizado por el tamaño de cada componente [5]. Resolver este problema de manera exacta requiere una descomposición en autovalores computacionalmente costosa; la contribución de LSC consiste en evitar dicha descomposición mediante una función núcleo (*kernel*) que proyecta explícitamente el color y la posición de cada píxel a un espacio de características de alta dimensión, en el cual se demuestra que el óptimo de un K-means ponderado coincide con el óptimo de *normalized cuts* en el espacio original [5]. De esta manera, LSC aproxima una solución de calidad comparable a *normalized cuts*, pero mediante iteraciones de K-means ponderado, lo que reduce su complejidad computacional a un orden lineal respecto al número de píxeles.

Superpixels Extracted via Energy-Driven Sampling (SEEDS)

A diferencia de los tres algoritmos anteriores, SEEDS no construye ni actualiza explícitamente centros de clúster; en su lugar, refina iterativamente las fronteras de una partición inicial mediante una optimización de ascenso de colina (*hill-climbing*) [7]. La partición inicial consiste en bloques rectangulares dispuestos sobre una rejilla regular. En cada iteración, el algoritmo evalúa mover un píxel —o, en las primeras iteraciones, un bloque completo— de la frontera entre dos superpíxeles vecinos de uno a otro; el movimiento se acepta únicamente si incrementa una función de energía que favorece la similitud entre el histograma de color de cada superpíxel y los píxeles en su frontera, junto con un término adicional que penaliza fronteras irregulares [7]. Al operar mediante bloques que se refinan progresivamente a nivel de píxel individual, y al requerir solo una actualización local del histograma en cada movimiento evaluado, SEEDS logra un costo computacional

considerablemente menor que los métodos que reconstruyen sus clústeres por completo en cada iteración.

La Tabla 2.2 resume los cuatro algoritmos descritos.

Tabla 2.2: Comparación conceptual de los algoritmos de superpíxel evaluados en el Capítulo 4

Algoritmo	Paradigma	Mecanismo central	Parámetro principal
SLIC	Agrupamiento (K-means adaptado)	Distancia ponderada color-espacio con búsqueda restringida a $2S \times 2S$	Compacidad m
ERS	Basado en grafos	Maximización voraz de la tasa de entropía de un paseo aleatorio más un término de equilibrio	Peso del equilibrio λ
LSC	Agrupamiento (aproximación de <i>normalized cuts</i>)	K-means ponderado en un espacio de características de alta dimensión	Peso espacial (compacidad)
SEEDS	Regiones (refinamiento de fronteras)	Ascenso de colina sobre histogramas de color, por bloque y por píxel	Número de niveles jerárquicos

2.7. Segmentación bio-inspirada mediante PCNN

Los algoritmos descritos en la Sección 2.6.2 representan el paradigma dominante para la generación de superpíxeles, pero requieren de una reducción de dimensionalidad previa para poder aplicarse sobre HSI. Esta sección presenta un paradigma alternativo, bio-inspirado y libre de entrenamiento por retropropagación [13], sobre el cual se construye la propuesta de esta tesis.

La PCNN es un modelo bio-inspirado basado en el estudio de los fenómenos de ráfagas neuronales sincronizadas en la corteza visual de los mamíferos [43, 14]. Pertenece a la tercera generación de redes neuronales artificiales y se fundamenta en la observación de experimentos de emisión de pulsos en la corteza cerebral [15].

El modelo PCNN contiene dos compartimentos principales: alimentación y acoplamiento. La representación de estos compartimentos, Alimentación (F) y Acoplamiento (L), se determina mediante las siguientes ecuaciones [44]:

$$F_{ij}[n] = e^{-\alpha_F \delta_n} F_{ij}[n-1] + S_{ij} + V_F \sum_{kl} M_{ij,kl} Y_{kl}[n-1], \quad (2.8)$$

$$L_{ij}[n] = e^{-\alpha_L \delta_n} L_{ij}[n-1] + V_L \sum_{kl} W_{ij,kl} Y_{kl}[n-1]. \quad (2.9)$$

Donde n es el número de la iteración, e i, j son los índices de la posición de la neurona en el arreglo bidimensional. F_{ij} y L_{ij} son los compartimentos de alimentación y acoplamiento respectivamente, S_{ij} representa el estímulo externo (intensidad del píxel), Y_{kl} es la salida binaria de las neuronas vecinas, y M, W corresponden a los pesos sinápticos de conexión. Los términos exponenciales $e^{-\alpha\delta_n}$ representan el decaimiento o memoria del sistema, mientras que V_F y V_L son constantes de normalización.

El estado de estos componentes se combina mediante una modulación multiplicativa para crear el estado interno de la neurona, U . Esta combinación es controlada por la fuerza de acoplamiento β [44]:

$$U_{ij}[n] = F_{ij}[n](1 + \beta L_{ij}[n]). \quad (2.10)$$

El estado interno se compara con un umbral dinámico Θ . La matriz $Y[n]$ denota la salida binaria de la PCNN en cada iteración, donde la salida Y_{ij} se produce según [15, 44]:

$$Y_{ij}[n] = \begin{cases} 1 & \text{Si } U_{ij}[n] > \Theta_{ij}[n-1] \\ 0 & \text{En otro caso} \end{cases}. \quad (2.11)$$

El umbral es dinámico; cuando la neurona dispara, el umbral incrementa significativamente su valor para simular el periodo refractario. La progresión del umbral es controlada por [44]:

$$\Theta_{ij}[n] = e^{-\alpha_\Theta \delta_n} \Theta_{ij}[n-1] + V_\Theta Y_{ij}[n]. \quad (2.12)$$

Donde V_Θ es una constante de magnitud elevada y α_Θ controla el decaimiento del umbral [15]. En el procesamiento de imágenes, cada neurona corresponde a un píxel. La neurona (i, j) emitirá un pulso siempre que U_{ij} supere a Θ_{ij} . Tras el disparo, el umbral aumenta al valor V_Θ ; de lo contrario, disminuye gradualmente conforme avanzan las iteraciones [15].

La Figura 2.5 resume la arquitectura completa de una neurona PCNN e ilustra cómo se articulan entre sí los cinco elementos descritos: los dos compartimentos de entrada, la modulación multiplicativa, el comparador y el lazo de realimentación del umbral.

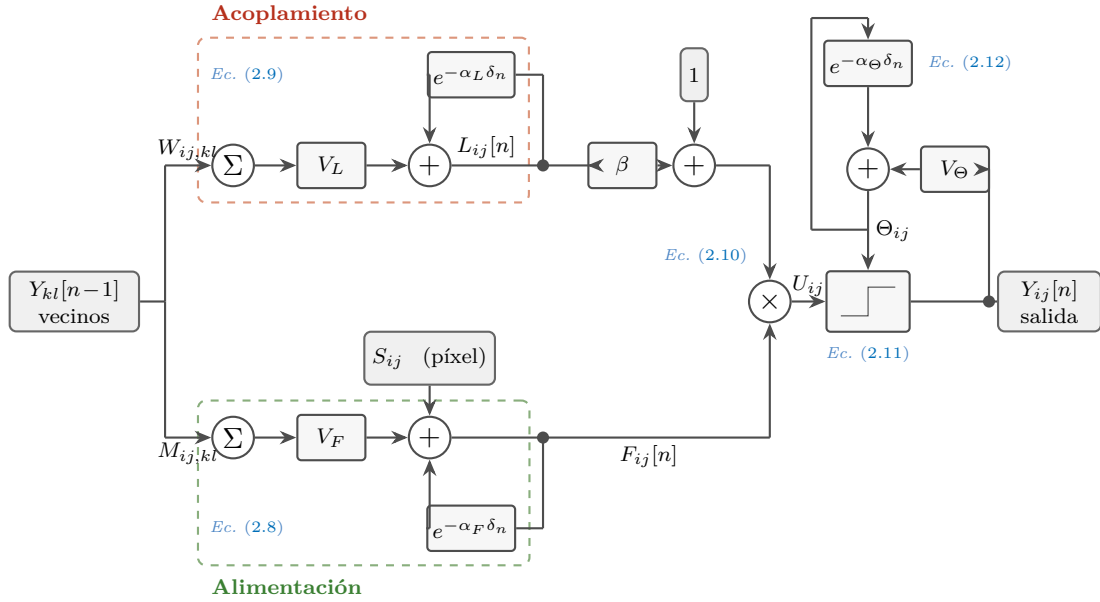


Figura 2.5: Arquitectura de una neurona PCNN. El compartimento de acoplamiento (L) y el de alimentación (F) integran las salidas $Y_{kl}[n-1]$ de las neuronas vecinas ponderadas por $W_{ij,kl}$ y $M_{ij,kl}$ respectivamente; solo el compartimento de alimentación recibe el estímulo externo S_{ij} . Ambos se combinan mediante la modulación multiplicativa $U_{ij} = F_{ij}(1 + \beta L_{ij})$, cuyo resultado se compara contra el umbral dinámico Θ_{ij} para producir la salida binaria $Y_{ij}[n]$. Los tres lazos de realimentación con decaimiento exponencial implementan la memoria del sistema, y el lazo del umbral reproduce el periodo refractario.

El Algoritmo 2.1 formaliza esta dinámica como un procedimiento iterativo. Conviene detenerse en la línea 6, porque es exactamente ahí donde interviene la propuesta de esta tesis: la contribución de cada neurona vecina al compartimento de acoplamiento está ponderada únicamente por el peso sináptico $W_{ij,kl}$, que depende de la distancia espacial entre neuronas pero no de la naturaleza del material que cada una representa. Dicho de otro modo, un vecino que ya disparó influye sobre la neurona (i, j) con el mismo peso sin importar cuán distinta sea su firma espectral. Sobre imágenes monocromáticas esta indiferencia es inocua, ya que el único atributo disponible es la intensidad; sobre una HSI, en cambio, implica que una región sincronizada puede crecer libremente a través de la frontera entre dos materiales espectralmente distintos. El algoritmo PCNN-SAM, descrito en el Capítulo 4, sustituye ese peso fijo por una condición de afinidad espectral evaluada mediante SAM (Sección 2.2.2).

Algoritmo 2.1 Dinámica de una PCNN clásica aplicada a la segmentación de una imagen

Entrada: Estímulo normalizado $S \in [0, 1]^{N \times M}$ (intensidad del píxel); parámetros de decaimiento $\alpha_F, \alpha_L, \alpha_\Theta$; constantes V_F, V_L, V_Θ ; fuerza de acoplamiento β ; núcleos sinápticos M y W ; número de iteraciones $N_{\text{máx}}$

Notación: $(K * Y)_{ij} = \sum_{kl} K_{ij,kl} Y_{kl}$ denota la suma ponderada sobre la vecindad de (i, j) .

Salida: Mapa de tiempos de disparo T , cuyas regiones de valor constante constituyen los grupos de neuronas sincronizadas

```
1:  $F \leftarrow \mathbf{0}, L \leftarrow \mathbf{0}, Y \leftarrow \mathbf{0}, \Theta \leftarrow \mathbf{1}$   $\triangleright$  estados iniciales de la red
2:  $T_{ij} \leftarrow 0$  para todo  $(i, j)$   $\triangleright$  ninguna neurona ha disparado aún

3: para  $n = 1$  hasta  $N_{\text{máx}}$  hacer
4:   para cada neurona  $(i, j)$  en paralelo hacer
5:      $F_{ij}[n] \leftarrow e^{-\alpha_F \delta_n} F_{ij}[n-1] + S_{ij} + V_F (M * Y[n-1])_{ij}$   $\triangleright$  Ec. (2.8)
6:      $L_{ij}[n] \leftarrow e^{-\alpha_L \delta_n} L_{ij}[n-1] + V_L (W * Y[n-1])_{ij}$   $\triangleright$  Ec. (2.9)
7:      $U_{ij}[n] \leftarrow F_{ij}[n] (1 + \beta L_{ij}[n])$   $\triangleright$  Ec. (2.10)
8:     si  $U_{ij}[n] > \Theta_{ij}[n-1]$  entonces
9:        $Y_{ij}[n] \leftarrow 1$   $\triangleright$  la neurona emite un pulso
10:      si  $T_{ij} = 0$  entonces
11:         $T_{ij} \leftarrow n$   $\triangleright$  se registra el primer disparo
12:      fin si
13:    si no
14:       $Y_{ij}[n] \leftarrow 0$ 
15:    fin si
16:     $\Theta_{ij}[n] \leftarrow e^{-\alpha_\Theta \delta_n} \Theta_{ij}[n-1] + V_\Theta Y_{ij}[n]$   $\triangleright$  Ec. (2.12)
17:  fin para
18: fin para

19: devolver  $T$ 
```

Aunque el modelo clásico presentado en esta sección sigue siendo la base teórica dominante, la PCNN continúa siendo objeto de extensiones activas en la literatura reciente, particularmente para reducir su dependencia del ajuste manual de parámetros y para adaptar su dinámica a tareas específicas de teledetección, como la fusión de imágenes multiespectrales [45] y el procesamiento eficiente de imágenes en dispositivos con recursos limitados [46].

2.8. Métricas de evaluación de la segmentación

Así como la Sección 3.4 del Capítulo 3 define las métricas con las que se evalúa la etapa de clasificación, esta sección define las tres métricas con las que se evalúa la calidad de

una segmentación por superpíxeles a lo largo de esta tesis: la exactitud de segmentación alcanzable (ASA), la tasa de recuperación de bordes (BR) y la varianza explicada (EV). Las tres se calculan respecto a una segmentación de referencia (*ground truth*) $\mathcal{G} = \{g_1, \dots, g_L\}$, obtenida de las etiquetas de clase disponibles para cada conjunto de datos.

Exactitud de segmentación alcanzable

La exactitud de segmentación alcanzable mide la pureza semántica de los superpíxeles: representa la exactitud de clasificación máxima que podría alcanzarse si a cada superpíxel se le asignara la etiqueta de la clase de referencia con la que más se traslapa [6]. Formalmente,

$$\text{ASA}(\mathcal{S}) = \frac{\sum_k \max_i |s_k \cap g_i|}{\sum_i |g_i|}, \quad (2.13)$$

donde $\mathcal{S} = \{s_1, \dots, s_K\}$ es el conjunto de superpíxeles evaluado, g_i es un segmento de la referencia, y $|\cdot|$ denota el número de píxeles de un conjunto. Un valor de ASA cercano a 1 indica que prácticamente ningún superpíxel traslapa dos clases distintas —es decir, que la sobresegmentación es semánticamente pura—, mientras que un valor bajo indica que los superpíxeles mezclan sistemáticamente píxeles de clases diferentes, lo cual introduce ruido en la matriz de características del grafo (Sección 2.6.1).

Tasa de recuperación de bordes

La métrica BR mide, en cambio, la adherencia geométrica de los superpíxeles a los bordes reales entre clases, en lugar de la pureza de su contenido:

$$\text{BR}(\mathcal{S}) = \frac{\sum_{p \in \mathcal{B}(\mathcal{G})} \mathbf{1} \left[\min_{q \in \mathcal{B}(\mathcal{S})} \|p - q\| < \epsilon \right]}{|\mathcal{B}(\mathcal{G})|}, \quad (2.14)$$

donde $\mathcal{B}(\mathcal{G})$ y $\mathcal{B}(\mathcal{S})$ son los conjuntos de píxeles de borde de la referencia y de la segmentación evaluada, respectivamente, y ϵ es una tolerancia de distancia en píxeles [7]. BR cuantifica, por lo tanto, qué fracción de los bordes reales de la escena tiene un borde de superpíxel suficientemente cercano; a diferencia de ASA, es sensible a la forma de los superpíxeles y no a su contenido espectral.

Varianza explicada

La varianza explicada cuantifica qué fracción de la varianza espectral total de la imagen se conserva al representar cada superpíxel mediante su firma espectral media, en lugar de por sus píxeles individuales [47]:

$$\text{EV}(\mathcal{S}) = \frac{\sum_k |s_k| (\bar{x}_{s_k} - \bar{x})^2}{\sum_p (x_p - \bar{x})^2}, \quad (2.15)$$

donde $\bar{x}_{s_k}$ es la firma espectral media del superpíxel s_k , y $\bar{x}$ es la firma espectral media de toda la imagen. A diferencia de ASA y BR, que evalúan propiedades geométricas de la segmentación, EV evalúa directamente cuánta información espectral sobrevive a la sobresegmentación. Esta distinción es central para esta tesis: como se cuantifica en el Capítulo 4, EV resulta ser un predictor considerablemente más confiable del desempeño de clasificación posterior que ASA, consistente con el mecanismo que motiva a PCNN-SAM.

El modelo PCNN presentado en la sección anterior constituye la base sobre la cual se construye la propuesta de segmentación PCNN-SAM, que modifica su dinámica de acoplamiento incorporando la métrica de similitud espectral SAM definida en la Sección 2.2.2, tal como se detalla en el Capítulo 4. Los superpíxeles resultantes de cualquiera de los cinco algoritmos presentados en este capítulo —los cuatro métodos clásicos y PCNN-SAM— se evalúan mediante ASA, BR y EV, y se estructuran, en la etapa de clasificación, como los nodos de un grafo procesado por una Red Convolutiva de Grafos, cuyos fundamentos se presentan en el Capítulo 3.

Capítulo 3

Clasificación de imágenes hiperespectrales

El objetivo de la clasificación de imágenes es construir una interpretación visual de los datos de forma automática para la categorización de clases sobre los píxeles de una escena [29]. La clasificación involucra el análisis de múltiples bandas y la aplicación de reglas de decisión basadas en estadísticas para determinar la identidad de un píxel. Esta categorización se utiliza para construir regiones temáticas sobre la superficie cubierta presente en una imagen [29].

En el contexto de HSI, la clasificación consiste en asignar una etiqueta a cada píxel de la escena. Formalmente, dado el hipercubo $\mathbf{X} \in \mathbb{R}^{B \times (N \times M)}$ introducido en la Sección 2.2.1, la tarea consiste en asociar a cada firma espectral $\mathbf{x}_i \in \mathbb{R}^B$ una etiqueta $y_i \in \{1, \dots, C\}$, donde C es el número de clases temáticas presentes en la escena [37].

Los algoritmos que resuelven esta tarea suelen agruparse, de acuerdo con la cantidad y el tipo de supervisión disponible durante el entrenamiento, en tres paradigmas de aprendizaje [48], recorridos en este capítulo en el siguiente orden: no supervisado, que prescinde por completo de etiquetas; supervisado, desde los algoritmos clásicos hasta las redes neuronales convolucionales, la arquitectura de aprendizaje profundo dominante para datos con estructura de rejilla; y semi-supervisado sobre grafos, el paradigma que esta tesis adopta para enfrentar la escasez de datos etiquetados, construido sobre los fundamentos de las redes neuronales de grafos y convolucionales de grafos. El capítulo cierra con las métricas utilizadas para evaluar la etapa de clasificación.

La Figura 3.1 contrasta los tres paradigmas en función de la cantidad de etiquetas disponibles durante el entrenamiento.

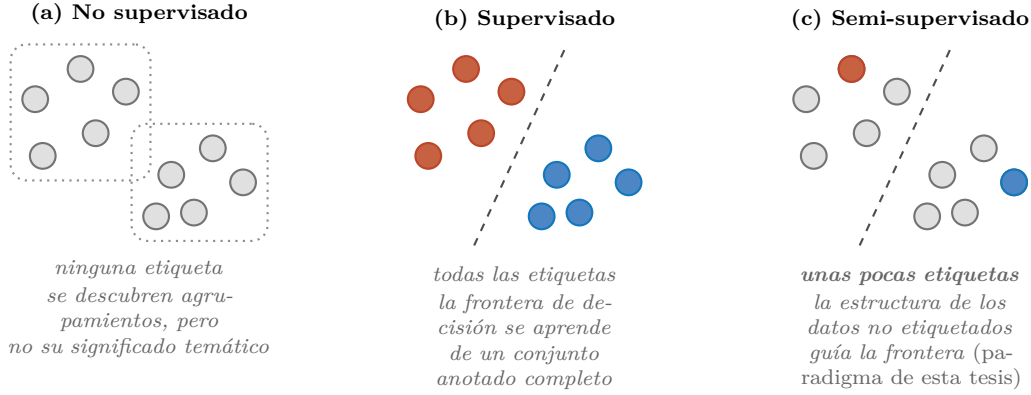


Figura 3.1: Los tres paradigmas de aprendizaje según la supervisión disponible durante el entrenamiento. El aprendizaje no supervisado descubre agrupamientos pero no su significado temático; el supervisado exige anotar la totalidad de las muestras; el semi-supervisado, adoptado en esta tesis, aprovecha la estructura de los datos no etiquetados para propagar la información de unas pocas muestras anotadas.

3.1. Aprendizaje no supervisado

El aprendizaje no supervisado busca descubrir estructura —agrupamientos, patrones de similitud o representaciones compactas— directamente en un conjunto de datos que no cuenta con ninguna etiqueta de clase [48]. En el contexto de HSI, este paradigma resulta atractivo precisamente porque evita por completo el costo de anotación, a costa de que las agrupaciones resultantes no necesariamente coinciden con las clases temáticas de interés para el intérprete humano.

El enfoque clásico consiste en aplicar un algoritmo de agrupamiento directamente sobre los píxeles de la imagen, generalmente después de una reducción de dimensionalidad. Un esquema representativo combina el análisis de componentes principales con agrupamiento K-means, asignando posteriormente una etiqueta temática a cada grupo mediante un clasificador supervisado entrenado sobre los centroides [49]. Variantes más recientes reemplazan la distancia euclidiana entre píxeles por formulaciones que explotan la estructura espacial de la escena: por ejemplo, un esquema de variación total no local combinado con un algoritmo primal-dual de gradiente híbrido permite agrupar píxeles espectralmente similares mientras se preserva la coherencia espacial de las regiones resultantes, superando a los métodos de agrupamiento puramente espectrales sobre varios conjuntos de datos hiperespectrales de referencia [50].

El aprendizaje profundo también se ha explorado en un régimen completamente no supervisado, principalmente mediante autocodificadores: redes que aprenden a reconstruir su propia entrada a través de una representación intermedia de menor dimensionalidad, sin requerir ninguna etiqueta durante el entrenamiento. Un autocodificador convolucio-

nal tridimensional (3D-CAE), compuesto únicamente por operaciones 3D —convolución, agrupamiento y normalización por lotes—, permite aprender conjuntamente características espaciales y espectrales de una HSI, entrenando tanto el codificador como un decodificador convolucional complementario únicamente a partir del error de reconstrucción [51]. Trabajos más recientes combinan esta idea con aprendizaje contrastivo: ContrastNet emplea dos autocodificadores como funciones de aumento de datos dentro de un esquema de aprendizaje contrastivo prototípico, de manera que las características aprendidas de forma no supervisada resultan, en los experimentos reportados por sus autores, competitivas incluso frente a métodos supervisados [52].

Pese a estos avances, ninguno de estos esquemas resuelve directamente el problema de esta tesis: al no incorporar ninguna etiqueta durante el entrenamiento, las agrupaciones obtenidas deben mapearse posteriormente a las clases temáticas de interés, lo cual introduce una fuente adicional de error que es difícil de controlar cuando las clases son espectralmente similares entre sí. Esta limitación motiva recurrir, en algún grado, a la información de clase disponible, ya sea de manera completa (aprendizaje supervisado) o parcial (aprendizaje semi-supervisado).

3.2. Aprendizaje supervisado

A diferencia del paradigma anterior, el aprendizaje supervisado entrena un modelo a partir de un conjunto de pares entrada-etiqueta completamente anotado, con el objetivo de aprender una función que generalice correctamente sobre datos no observados durante el entrenamiento [48]. Es, históricamente, el paradigma dominante en la clasificación de HSI, y se presenta aquí en dos etapas: primero los algoritmos de aprendizaje automático clásico, que dependen de características diseñadas manualmente, y después las redes neuronales convolucionales, que extraen dichas características automáticamente a partir de los datos.

3.2.1. Métodos clásicos de aprendizaje automático

Antes de la adopción generalizada del aprendizaje profundo, la clasificación de HSI se abordaba principalmente mediante algoritmos de aprendizaje automático clásico, los cuales continúan utilizándose como referencia comparativa en la literatura actual [2]:

- **Máxima Verosimilitud Gaussiana:** evalúa la varianza y covarianza de las bandas espectrales, asumiendo una distribución normal para describir la clase de un píxel desconocido [29].
- **K-Vecinos Cercanos:** asigna a un píxel desconocido la clase mayoritaria entre sus k vecinos más cercanos en el espacio de características [2].

- **Máquina de Vectores de Soporte:** clasificador supervisado que, a partir de los datos de entrenamiento, construye un hiperplano que maximiza el margen entre clases, definiendo así los límites de decisión utilizados para clasificar píxeles desconocidos [2].
- **Bosques Aleatorios:** combina las predicciones de un conjunto de árboles de decisión, cada uno entrenado sobre un subconjunto aleatorio de los datos y de las características, reduciendo el riesgo de sobreajuste respecto a un único árbol [2].

La Tabla 3.1 resume estos cuatro métodos y la razón por la que cada uno pierde eficacia conforme crece la dimensionalidad espectral.

Tabla 3.1: Métodos clásicos de clasificación supervisada y sus limitaciones sobre HSI

Método	Principio	Limitación al aumentar B
Máxima verosimilitud gaussiana	Modelo paramétrico por clase	Requiere estimar una covarianza $B \times B$ por clase, inviable con pocas muestras etiquetadas
k -vecinos cercanos	Voto mayoritario entre los k vecinos más próximos	La distancia euclidiana pierde poder discriminante en dimensión alta
Máquina de vectores de soporte	Hiperplano de margen máximo	Robusta en dimensión alta, pero no explota la estructura espacial de la escena
Bosques aleatorios	Ensamble de árboles de decisión	Trata cada banda por separado y desaprovecha la correlación entre bandas contiguas

Estos métodos, si bien efectivos en diversos escenarios, dependen de características diseñadas manualmente o de suposiciones estadísticas simples sobre la distribución de los datos, y su desempeño se degrada conforme aumenta la dimensionalidad espectral de las HSI modernas. Esta limitación motivó la adopción de algoritmos de aprendizaje profundo, presentados a continuación.

3.2.2. Redes neuronales convolucionales

El aprendizaje profundo es una rama de la inteligencia artificial y el aprendizaje automático que utiliza una jerarquía de conceptos para permitir que las computadoras aprendan de la experiencia, construyendo representaciones complejas a partir de conceptos más simples [53]. A diferencia de los métodos de aprendizaje automático clásico, que dependen de características diseñadas manualmente, los modelos de aprendizaje profundo extraen automáticamente las características relevantes directamente de los datos mediante arquitecturas de neuronas organizadas en múltiples capas [53]. Las CNN son la arquitectura de aprendizaje profundo dominante para el procesamiento de imágenes, incluyendo la

clasificación de HSI [53]. Una CNN extrae características de manera jerárquica mediante la aplicación sucesiva de operaciones de convolución —filtros que se desplazan sobre la imagen detectando patrones locales, como bordes o texturas, en las primeras capas, y patrones cada vez más abstractos en las capas subsecuentes— intercaladas con operaciones de agrupamiento que reducen progresivamente la resolución espacial de los mapas de características [53].

A diferencia de una imagen RGB convencional, una HSI añade una tercera dimensión de gran tamaño —la banda espectral—, lo que ha dado lugar a al menos cuatro estrategias distintas para adaptar las CNN a este dato, cada una con una relación de compromiso distinta entre información explotada y costo computacional.

La estrategia más simple aplica una CNN puramente espacial (2D) sobre una representación comprimida de la HSI: tras reducir la dimensionalidad espectral mediante PCA, las primeras componentes principales se apilan como si fueran los canales de una imagen convencional, y una CNN 2D estándar extrae características espaciales sobre dicha representación [54]. Esta estrategia es computacionalmente económica, pero al comprimir la dimensión espectral antes de clasificar, descarta gran parte de la información espectral fina que distingue a las HSI de una imagen RGB —la misma limitación que motiva a esta tesis a evitar la reducción de dimensionalidad en la etapa de segmentación.

Para explotar directamente la información espectral sin comprimirla, una segunda estrategia reemplaza las convoluciones 2D por convoluciones 3D, que se desplazan simultáneamente a lo largo de las dos dimensiones espaciales y de la dimensión espectral de un cubo de HSI, extrayendo así características espacio-espectrales conjuntas desde la primera capa de la red [55]. Esta capacidad tiene un costo: el número de parámetros y el costo computacional de una CNN 3D crecen considerablemente respecto a su contraparte 2D, lo cual, combinado con la escasez de datos etiquetados propia de HSI, agrava el riesgo de sobreajuste al apilar muchas capas 3D.

Una tercera estrategia aborda precisamente esta dificultad incorporando conexiones residuales —enlaces que permiten a una capa sumar su entrada directamente a su salida— dentro de una arquitectura 3D profunda: la Red Residual Espacial-Espectral (SSRN) apila bloques residuales espectrales y espaciales de manera consecutiva, tomando el cubo de HSI en bruto como entrada sin ingeniería de características previa, y mitiga así el problema de degradación que de otro modo limitaría la profundidad —y por tanto la capacidad discriminativa— de una CNN 3D convencional [56].

Finalmente, una cuarta estrategia busca un punto intermedio entre la riqueza espacio-espectral de una CNN 3D y la eficiencia computacional de una CNN 2D: HybridSN apila primero un bloque de convoluciones 3D, que extrae conjuntamente características espaciales y espectrales de bajo nivel a partir de una vecindad reducida de bandas, seguido de un bloque de convoluciones 2D que refina dichas características a un nivel espacial más abstracto, reduciendo así la complejidad del modelo respecto a una CNN 3D pura sin renunciar por completo a la información espectral conjunta [57]. Esta familia de arquitecturas híbridas ilustra un principio recurrente en el diseño de CNN para HSI —y directamente relevante

para esta tesis—: la información espacial y la espectral rara vez se explotan de manera igualmente eficiente por un mismo tipo de operación, lo cual motiva arquitecturas que combinan explícitamente ambos tipos de procesamiento en lugar de tratar a la HSI como una imagen convencional con más canales. El desarrollo de variantes cada vez más ligeras y adaptables de este esquema continúa activo: trabajos recientes incorporan convoluciones deformables —que ajustan la posición de muestreo del filtro a la geometría local de la escena, en lugar de a una rejilla fija— para reducir el costo computacional de la extracción espacial-espectral en escenarios con recursos limitados, como la clasificación de HSI capturadas por UAV [58]. Los avances más recientes en el área, incluyendo arquitecturas basadas en mecanismos de atención y transformadores, se documentan en [37].

La Figura 3.2 contrasta gráficamente las cuatro estrategias y la Tabla 3.2 resume el compromiso que asume cada una.

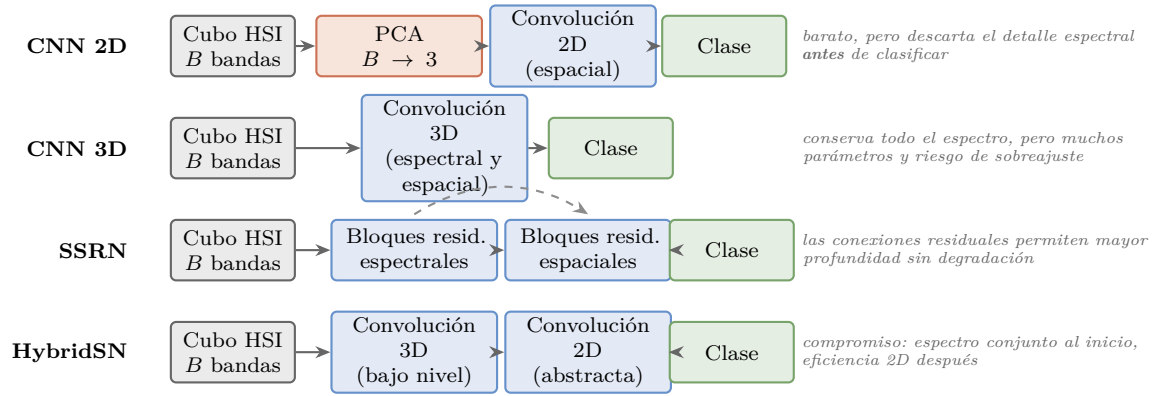


Figura 3.2: Estrategias de adaptación de las CNN a datos hiperespectrales. La diferencia esencial entre ellas reside en *cuándo* —y si acaso— se comprime la dimensión espectral: la CNN 2D lo hace antes de aprender, mientras que las tres restantes conservan el cubo completo a la entrada.

Tabla 3.2: Compromiso entre información explotada y costo computacional en las cuatro estrategias de CNN para HSI

Estrategia	Entrada	Información explotada	Costo y riesgo
CNN 2D sobre PCA [54]	Primeras componentes principales	Solo espacial; el detalle espectral se pierde antes de clasificar	Bajo
CNN 3D [55]	Cubo completo	Espacial y espectral de forma conjunta desde la primera capa	Alto; propenso al sobreajuste con pocas etiquetas
SSRN [56]	Cubo completo	Espectral y espacial en bloques residuales consecutivos	Alto, pero las conexiones residuales evitan la degradación
HybridSN [57]	Cubo completo	3D de bajo nivel seguida de 2D abstracta	Intermedio

Sin embargo, esta misma fortaleza —operar sobre una rejilla regular de vecinos, es decir, sobre un dato euclidiano, ya sea en dos o en tres dimensiones— es también su principal limitación para los fines de esta tesis: una CNN asume que cada elemento de entrada tiene el mismo número de vecinos, dispuestos siempre en la misma configuración espacial relativa, lo cual dificulta modelar de manera flexible las relaciones de vecindad irregulares que existen entre superpíxeles de tamaño y forma variable. Esta limitación motiva el uso de una representación de datos no euclidiana, presentada a continuación.

3.3. Aprendizaje semi-supervisado

El aprendizaje semi-supervisado combina un conjunto reducido de datos etiquetados con un conjunto —típicamente mucho más grande— de datos no etiquetados durante el entrenamiento de un modelo, con el objetivo de mejorar su desempeño respecto a entrenar únicamente con los datos etiquetados disponibles [24, 59]. A diferencia del aprendizaje no supervisado, sí incorpora información de clase; y a diferencia del aprendizaje supervisado, no requiere que dicha información esté disponible para la totalidad de las muestras. Este paradigma resulta particularmente relevante en la clasificación de HSI, donde, como se estableció en el Planteamiento del problema (Sección 1.3), la escasez de datos etiquetados constituye la principal limitante operativa.

Las redes neuronales de grafos son la herramienta que esta tesis emplea para instanciar este paradigma sobre HSI. A continuación se presentan primero sus fundamentos generales, para después describir el esquema de entrenamiento transductivo que las combina con el aprendizaje semi-supervisado.

3.3.1. Redes neuronales de grafos

Las Redes neuronales de grafos (GNN) generalizan el aprendizaje profundo a datos representados en espacios no euclidianos [60, 61, 62], como el que resulta de modelar los superpíxeles de una HSI y sus relaciones de vecindad como un grafo.

Definición de grafo

Un grafo se define como $G = (\mathcal{V}, \mathcal{E})$, donde $\mathcal{V}$ es el conjunto de nodos o vértices y $\mathcal{E}$ es el conjunto de aristas que conectan un nodo $u \in \mathcal{V}$ con un nodo $v \in \mathcal{V}$, tal que $(u, v) \in \mathcal{E}$. La forma estándar de representar esta estructura es mediante la Matriz de Adyacencia $\mathbf{A} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$, definida como:

$$A_{ij} = \begin{cases} 1 & \text{si } (v_i, v_j) \in \mathcal{E} \\ 0 & \text{en caso contrario} \end{cases}. \quad (3.1)$$

Para el caso de grafos no dirigidos, la matriz de adyacencia $\mathbf{A}$ es simétrica [60].

Actualización del estado de los nodos: el paso de mensajes

Las redes neuronales de grafos emplean un método basado en la actualización iterativa para describir un nodo, obteniendo información de sus vecinos y de la estructura del grafo. El objetivo es aprender las características del estado de un nodo integrando la información de su vecindario para generar un estado actualizado [60]. La actualización de un nodo sigue la ecuación general:

$$\mathbf{h}_v^{(k)} = \text{UPDATE} \left(\mathbf{h}_v^{(k-1)}, \text{AGGREGATE} \left(\left\{ \mathbf{h}_u^{(k-1)} : u \in \mathcal{N}(v) \right\} \right) \right). \quad (3.2)$$

Donde $\mathbf{h}_v^{(k)}$ es el vector de características del nodo v en la capa k , y $\mathcal{N}(v)$ es el conjunto del vecindario de todos los nodos adyacentes a v . Las funciones UPDATE y AGGREGATE son funciones diferenciables arbitrarias. La función UPDATE genera el nuevo estado combinando la información anterior con el resultado de la función AGGREGATE, la cual combina la información del vecindario en una representación invariante a la permutación. La Figura 3.3 ilustra esta operación sobre el vecindario de un nodo, junto con su consecuencia más relevante para esta tesis: dado que cada capa incorpora únicamente un salto adicional, el número de capas determina el radio del vecindario que influye sobre la representación de un nodo.

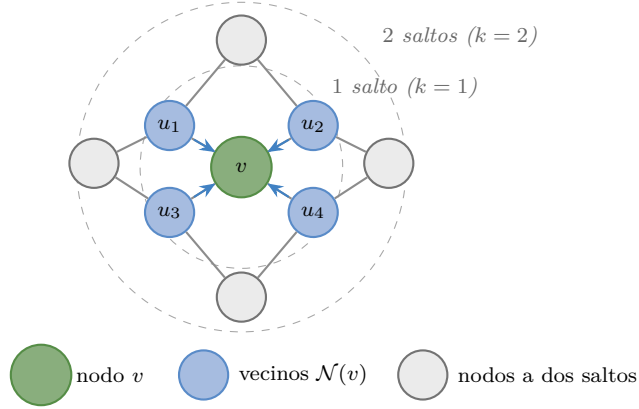


Figura 3.3: Vecindario de un nodo y radio de propagación en una GNN. En cada capa, las salidas de los nodos vecinos $u \in \mathcal{N}(v)$ fluyen hacia v . Los nodos situados a dos saltos (en gris) solo influyen sobre la representación de v si se apila una segunda capa: apilar k capas equivale a extender el radio del vecindario a k saltos, lo cual explica tanto la capacidad de propagación de una GCN bajo escasez de etiquetas como el fenómeno de sobresuavizado que aparece al aumentar excesivamente la profundidad.

Este proceso puede descomponerse mediante una función de mensaje:

$$\mathbf{m}_v^{(k)} = \sum_{u \in \mathcal{N}(v)} \text{MESSAGE} \left(\mathbf{h}_u^{(k-1)} \right). \quad (3.3)$$

Donde la función MESSAGE transforma las características de los vecinos antes de ser combinadas. Finalmente, el estado se actualiza como:

$$\mathbf{h}_v^{(k)} = \text{UPDATE} \left(\mathbf{h}_v^{(k-1)}, \mathbf{m}_v^{(k)} \right). \quad (3.4)$$

Después de K iteraciones de este proceso de actualización, la salida de la última capa se utiliza para definir las características finales de cada nodo. La Figura 3.4 resume el encadenamiento de las tres funciones involucradas en una sola capa.

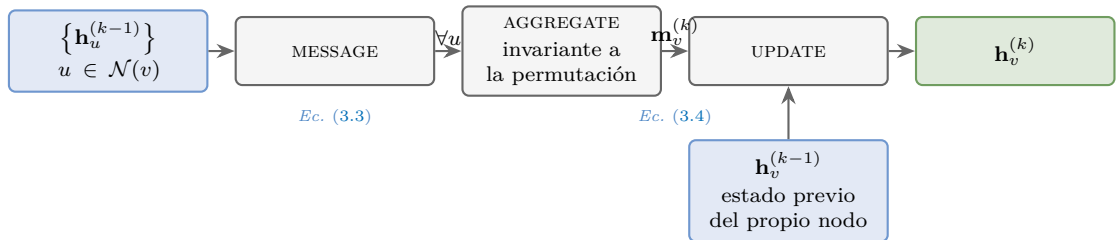


Figura 3.4: Actualización del estado de un nodo en una sola capa de GNN. La función MESSAGE transforma la característica de cada vecino, AGGREGATE las combina en una representación invariante a la permutación y UPDATE integra ese resultado con el estado previo del propio nodo para producir $\mathbf{h}_v^{(k)}$.

Modelo básico de red neuronal de grafos

La forma básica de una red neuronal de grafos se obtiene instanciando las funciones mencionadas. Una simplificación común utilizada en algoritmos de aprendizaje profundo es [60]:

$$\mathbf{h}_v^{(k)} = \sigma \left(\mathbf{W}_{\text{self}}^{(k)} \mathbf{h}_v^{(k-1)} + \mathbf{W}_{\text{neigh}}^{(k)} \sum_{u \in \mathcal{N}(v)} \mathbf{h}_u^{(k-1)} + \mathbf{b}^{(k)} \right). \quad (3.5)$$

3.3.2. Redes convolucionales de grafos

Las GCN son algoritmos de aprendizaje profundo que, a diferencia de las CNN, utilizan como entrada la estructura de un grafo en lugar de una rejilla regular. El modelo propuesto por Kipf y Welling [63] representa una simplificación y mejora de la GNN básica de la Ecuación (3.5), resultando en un método más eficiente y estable, y es el que se adopta como clasificador en esta tesis (Capítulo 4).

La fórmula de propagación de estados se define como:

$$\mathbf{H}^{(k+1)} = \sigma \left(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{H}^{(k)} \mathbf{W}^{(k)} \right). \quad (3.6)$$

Donde $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_N$ es la matriz de adyacencia a la que se le ha añadido la matriz identidad $\mathbf{I}_N$ para incluir las conexiones propias de cada nodo. El término $\tilde{\mathbf{D}}$ es la matriz de grado diagonal de $\tilde{\mathbf{A}}$, donde $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$. El componente $(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}})$ corresponde a la matriz de adyacencia normalizada simétricamente, y actúa como la función de agregación (AGGREGATE), realizando un promedio ponderado y estable de las representaciones de la capa anterior $\mathbf{H}^{(k)}$ dentro del vecindario de cada nodo. Finalmente, el resultado de esta agregación es transformado linealmente por una matriz de pesos $\mathbf{W}^{(k)}$, que cumple las funciones de actualización (UPDATE) y mensaje (MESSAGE), y se procesa mediante una función de activación no lineal σ para generar la matriz de estados de la siguiente capa, $\mathbf{H}^{(k+1)}$.

Dado que cada capa de una GCN agrega información únicamente del vecindario inmediato de un nodo, apilar K capas permite que la información se propague hasta K saltos de distancia dentro del grafo. Sin embargo, incrementar excesivamente la profundidad de una GCN no mejora indefinidamente su capacidad discriminativa: cuando el número de capas es demasiado alto, las representaciones de todos los nodos de una misma componente conexa tienden a converger a un valor similar entre sí, un fenómeno conocido como sobresuavizado (*over-smoothing*), que desvanece la capacidad del modelo para distinguir entre nodos [64]. Por esta razón, las arquitecturas GCN utilizadas en la práctica —incluyendo la presentada en el Capítulo 4— suelen restringirse a un número reducido de capas, típicamente entre una y tres.

La Tabla 3.3 resume las diferencias entre CNN y GNN/GCN relevantes para esta tesis.

Tabla 3.3: CNN frente a GNN/GCN: supuestos sobre la estructura de los datos

	CNN	GNN / GCN
Estructura de datos	Rejilla regular (euclidiana)	Grafo (no euclidiano)
Vecindad	Fija: misma cantidad y disposición para todo elemento	Variable: definida por la topología del grafo
Unidad de entrada en esta tesis	Píxel individual (no se emplea como clasificador final)	Superpíxel, como nodo del grafo

Con los fundamentos de GNN y GCN establecidos, el resto describe cómo se combinan específicamente con el aprendizaje semi-supervisado para instanciar el clasificador de esta tesis.

Las GCN se prestan de manera natural al aprendizaje semi-supervisado transductivo: dado un único grafo que contiene tanto nodos etiquetados como no etiquetados, el conjunto de nodos $\mathcal{V}$ se particiona en un subconjunto etiquetado $\mathcal{V}_L$ y un subconjunto no etiquetado $\mathcal{V}_U$, y el modelo se entrena minimizando una función de pérdida calculada únicamente sobre $\mathcal{V}_L$. Gracias a las capas de propagación descritas anteriormente, la información fluye desde los nodos etiquetados hacia sus vecinos no etiquetados a través de la estructura del grafo, permitiendo inferir las etiquetas de $\mathcal{V}_U$ sin haber observado directamente ninguna etiqueta para ellos [63]. La Figura 3.5 ilustra este mecanismo. Es este esquema el que instancia, en la práctica, el clasificador semi-supervisado presentado en el Capítulo 4.

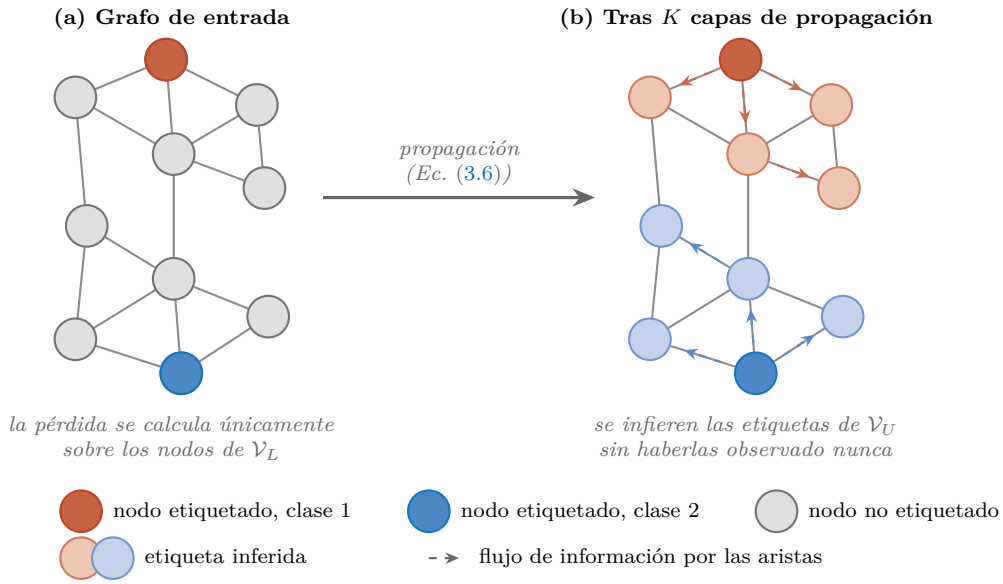


Figura 3.5: Aprendizaje semi-supervisado transductivo sobre un grafo. En (a) solo dos nodos poseen etiqueta y son los únicos que contribuyen a la función de pérdida. En (b), tras K capas de propagación, la información ha fluido por las aristas hasta los nodos de $\mathcal{V}_U$, cuyas etiquetas se infieren sin haber sido observadas en ningún momento del entrenamiento. El resultado depende por completo de que las aristas conecten nodos que efectivamente pertenecen a la misma clase.

Algoritmo 3.1 Entrenamiento transductivo de una GCN semi-supervisada

Entrada: Grafo $G = (\mathcal{V}, \mathcal{E})$ con matriz de adyacencia $\mathbf{A}$; matriz de características $\mathbf{X} \in \mathbb{R}^{|\mathcal{V}| \times d}$; subconjunto etiquetado $\mathcal{V}_L \subset \mathcal{V}$ con etiquetas $\mathbf{Y}$; número de capas K ; épocas T ; tasa de aprendizaje η

Salida: Etiqueta inferida $\hat{y}_v$ para cada nodo no etiquetado $v \in \mathcal{V}_U = \mathcal{V} \setminus \mathcal{V}_L$

- 1: $\tilde{\mathbf{A}} \leftarrow \mathbf{A} + \mathbf{I}_N$ $\triangleright$ se añaden las conexiones propias
 - 2: $\tilde{D}_{ii} \leftarrow \sum_j \tilde{A}_{ij}$ $\triangleright$ matriz de grado
 - 3: $\hat{\mathbf{A}} \leftarrow \tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-1/2}$ $\triangleright$ normalización simétrica, Ec. (3.6)
 - 4: Inicializar aleatoriamente los pesos $\mathbf{W}^{(0)}, \dots, \mathbf{W}^{(K-1)}$

 - 5: **para** $t = 1$ **hasta** T **hacer**
 - 6: $\mathbf{H}^{(0)} \leftarrow \mathbf{X}$
 - 7: **para** $k = 0$ **hasta** $K - 2$ **hacer**
 - 8: $\mathbf{H}^{(k+1)} \leftarrow \text{ReLU}(\hat{\mathbf{A}} \mathbf{H}^{(k)} \mathbf{W}^{(k)})$ $\triangleright$ propagación por el vecindario
 - 9: **fin para**
 - 10: $\mathbf{Z} \leftarrow \hat{\mathbf{A}} \mathbf{H}^{(K-1)} \mathbf{W}^{(K-1)}$ $\triangleright$ capa de salida, sin activación
 - 11: $\hat{\mathbf{Y}} \leftarrow \text{softmax}(\mathbf{Z})$
 - 12: $\mathcal{L} \leftarrow - \sum_{v \in \mathcal{V}_L} \sum_{c=1}^C Y_{vc} \ln \hat{Y}_{vc}$ $\triangleright$ solo los nodos de $\mathcal{V}_L$ aportan gradiente
 - 13: $\mathbf{W}^{(k)} \leftarrow \mathbf{W}^{(k)} - \eta \nabla_{\mathbf{W}^{(k)}} \mathcal{L}$ para $k = 0, \dots, K - 1$
 - 14: **fin para**

 - 15: **para cada** $v \in \mathcal{V}_U$ **hacer**
 - 16: $\hat{y}_v \leftarrow \arg \max_c \hat{Y}_{vc}$ $\triangleright$ inferencia transductiva
 - 17: **fin para**
 - 18: **devolver** $\{\hat{y}_v : v \in \mathcal{V}_U\}$
-

El Algoritmo 3.1 formaliza este esquema, y dos de sus líneas concentran tanto su carácter semi-supervisado como la vulnerabilidad que motiva esta tesis. La línea 12 evalúa la pérdida exclusivamente sobre $\mathcal{V}_L$: bajo el régimen de cinco muestras etiquetadas por clase adoptado en el Capítulo 4, el gradiente proviene de apenas un puñado de nodos. La línea 8, en cambio, mezcla las características de cada nodo con las de todos sus vecinos a través de $\hat{\mathbf{A}}$, que no es un parámetro aprendido sino una entrada fija heredada de la etapa de segmentación. En consecuencia, si dos superpíxeles adyacentes pertenecen a clases distintas pero una segmentación deficiente los unió en el grafo, la red promediará sus representaciones en cada capa sin disponer de evidencia etiquetada suficiente para revertir ese error. Esta dependencia es precisamente la razón por la que el Capítulo 4 mantiene fija la arquitectura del clasificador y hace variar únicamente el algoritmo que produce $\hat{\mathbf{A}}$.

La aplicación de GCN a la clasificación semi-supervisada de HSI ha recibido una atención creciente en años recientes [65]. Diversas arquitecturas han explorado variantes de este esquema. Por ejemplo, X-GPN incorpora una capa prototípica y matrices de adyacencia multiescala para mejorar la propagación semi-supervisada [16]; otros trabajos integran

mecanismos de compuerta entre ramas convolucionales y de grafos [25], o exploran esquemas de propagación dinámica y multiescala [66].

Otro grupo de trabajos ha enriquecido específicamente el mecanismo de propagación mediante mecanismos de atención: la red de atención de grafos con agrupamiento espacial [67] y la red de fusión multivista con atención de doble salto [68] incorporan capas de atención para ponderar de manera diferenciada la contribución de cada nodo vecino; otros integran aprendizaje contrastivo dentro de la propagación convolucional de grafos para obtener representaciones más discriminativas [69], o combinan una rama convolucional 3D-2D con un mecanismo de atención de grafos en una arquitectura híbrida [70]. A diferencia de estos trabajos, que se enfocan en enriquecer la arquitectura del clasificador, esta tesis mantiene deliberadamente la formulación de GCN más simple (la de Kipf y Welling, Ecuación (3.6)) para aislar y evaluar el efecto de la calidad de la segmentación —la variable de interés de esta investigación— sin introducir factores de confusión provenientes de la complejidad del clasificador.

Un desafío común a todas estas arquitecturas, y central para esta tesis, es que su desempeño depende críticamente de la calidad de la topología del grafo sobre la que operan y se evalúa experimentalmente. Este desafío ha sido reconocido explícitamente en la literatura reciente, aunque abordado desde estrategias distintas a la de esta tesis. LAGAN propone una segmentación en superpíxeles alineada e inducida espectralmente (SIAS), que fusiona información espectral cruda y características espectrales profundas abstractas para generar, desde el inicio, un grafo más discriminativo [71]; MSGCN-CRF, en cambio, no interviene en la etapa de segmentación, sino que corrige de manera posterior los errores de clasificación introducidos por una segmentación deficiente, mediante una agregación de similitud mixta seguida de un refinamiento final basado en campos aleatorios condicionales (CRF) [72]. Ambas estrategias —mejorar la segmentación de entrada o corregir sus errores después de clasificar— contrastan con el enfoque de esta tesis, que ataca el problema en su origen: evaluar y comparar, entre varios algoritmos de segmentación bio-inspirados y clásicos, cuál produce la topología de grafo más adecuada antes de que el proceso de clasificación siquiera comience.

La Tabla 3.4 sintetiza estas propuestas y sitúa la de esta tesis respecto a ellas, atendiendo a la etapa del flujo de trabajo en la que cada una interviene.

Tabla 3.4: Arquitecturas GCN semi-supervisadas para HSI, clasificadas según la etapa del flujo de trabajo en la que intervienen

Trabajo	Estrategia principal	Etapa de intervención
X-GPN [16]	Capa prototípica y adyacencia multiescala	Clasificador
Compuerta CNN–GCN [25]	Fusión por compuerta entre ramas convolucional y de grafos	Clasificador
Propagación dinámica [66]	Propagación multiescala adaptativa	Clasificador
GAT con agrupamiento espacial [67]	Atención sobre los nodos vecinos	Clasificador
Fusión multivista [68]	Atención de doble salto	Clasificador
GCN contrastiva [69]	Aprendizaje contrastivo en la propagación	Representación
Rama 3D–2D con atención [70]	Arquitectura híbrida convolucional y de grafos	Clasificador
LAGAN [71]	Superpíxeles alineados e inducidos espectralmente (SIAS)	Segmentación
MSGCN-CRF [72]	Refinamiento posterior mediante CRF	Corrección posterior
MFSGCN [73]	Búsqueda de arquitectura (NAS) sobre múltiples escalas	Escala del superpíxel
Esta tesis	Comparación de algoritmos de segmentación manteniendo fijo el clasificador	Segmentación

La selección de la escala del superpíxel —el número de superpíxeles K generados sobre una escena— constituye otro punto de comparación relevante. Mientras que MFSGCN aborda este problema de manera automática, empleando búsqueda de arquitectura neuronal (NAS) para ponderar y combinar características de superpíxeles extraídas simultáneamente en múltiples escalas [73], esta tesis opta por una selección empírica de un único valor de K por conjunto de datos, evaluada experimentalmente en el Capítulo 4, priorizando la interpretabilidad del procedimiento sobre la automatización de la búsqueda de escala.

3.4. Métricas de evaluación de la clasificación

Es esencial definir métricas de evaluación precisas para medir el desempeño de los clasificadores. Las matrices de confusión comparan, categoría por categoría, la relación entre los datos de referencia conocidos y los resultados producidos por el algoritmo [29]. A partir de esta matriz se obtienen las tres métricas utilizadas en esta tesis para evaluar la etapa de clasificación.

Exactitud global y exactitud promedio

La exactitud global (OA) es la proporción de píxeles correctamente clasificados respecto al total:

$$OA = \frac{\text{Total de predicciones correctas}}{\text{Total de predicciones}}. \quad (3.7)$$

La exactitud promedio (AA) complementa a la OA promediando, en lugar de sumar directamente, la exactitud obtenida en cada clase de manera individual:

$$AA = \frac{1}{C} \sum_{c=1}^C \frac{x_{cc}}{x_{c+}}, \quad (3.8)$$

donde C es el número de clases, x_{cc} es el número de aciertos de la clase c , y x_{c+} es el total de muestras de referencia de dicha clase. A diferencia de OA, que puede estar dominada por las clases con mayor número de muestras, AA pondera por igual a todas las clases independientemente de su tamaño, por lo que resulta más sensible al desempeño sobre clases minoritarias, una consideración relevante en conjuntos de datos con fuerte desbalance de clases como los descritos en la Sección 4.2.2.

Coefficiente kappa

Finalmente, el estadístico Kappa ($\hat{\kappa}$) mide la diferencia entre el acuerdo real (datos de referencia frente al clasificador) y el acuerdo que podría ocurrir por puro azar, sirviendo como un indicador de qué parte de la exactitud reportada se debe a un acuerdo verdadero [29]:

$$\hat{\kappa} = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \cdot x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \cdot x_{+i})}, \quad (3.9)$$

donde N es el número total de observaciones, r es el número de filas de la matriz de confusión, x_{ii} son las observaciones en la diagonal principal (predicciones correctas), y x_{i+} , x_{+i} son los totales marginales de las filas y columnas, respectivamente, utilizados para calcular la concordancia esperada por azar [29].

Estas tres métricas —OA, AA y $\hat{\kappa}$ — son las que se reportan de manera consistente a lo largo de la evaluación experimental presentada en el Capítulo 4.

Capítulo 4

Clasificación semi-supervisada mediante PCNN-GCN

4.1. Planteamiento experimental

El Capítulo 1 estableció que el principal cuello de botella para la clasificación de imágenes hiperespectrales mediante arquitecturas profundas es la escasez de datos etiquetados, y que las GCN ofrecen una vía para sortear esta limitación al propagar la información espectral a través de la estructura topológica de superpíxeles. Sin embargo, el rendimiento de este esquema semi-supervisado depende críticamente de la pureza espectral y la integridad espacial del grafo base: una segmentación defectuosa introduce errores que la GCN no puede corregir cuando dispone de apenas un puñado de etiquetas por clase.

A partir de esta premisa, se planteó la hipótesis de que una arquitectura híbrida PCNN-GCN alcanzaría una exactitud de clasificación significativamente superior a los métodos de segmentación del estado del arte bajo condiciones de etiquetado extremadamente limitado. Este capítulo reporta el diseño experimental construido para poner a prueba dicha hipótesis, el diseño del algoritmo de segmentación PCNN-SAM, el entrenamiento del clasificador GCN semi-supervisado, la evaluación mediante métricas estandarizadas y validación cruzada, y la implementación conjunta de la arquitectura híbrida.

En este marco, el capítulo busca responder cuatro preguntas concretas:

- ¿Cuál es la granularidad de superpíxel (K) que ofrece el mejor balance entre exactitud de clasificación y costo computacional bajo el régimen de 5 muestras por clase?
- ¿Qué estrategia de segmentación genera la topología de grafo más favorable para la propagación semi-supervisada, y esta respuesta depende de la naturaleza espacial y espectral de la escena?
- ¿Ofrece la preservación del vector espectral completo, sin reducción de dimensionalidad previa, una ventaja medible frente a los métodos clásicos, y bajo qué condiciones se

manifiesta?

El resto del capítulo se organiza como sigue: primero se describe la arquitectura híbrida propuesta, incluyendo los conjuntos de datos sobre los que se evalúa; después se detalla la configuración experimental —protocolo de muestreo semi-supervisado y parámetros de entrenamiento—; a continuación se determina la escala óptima de superpíxeles; posteriormente se presentan los resultados de segmentación y clasificación de manera comparativa; y finalmente se discuten los hallazgos encontrados.

4.2. Método propuesto

Esta sección describe la arquitectura híbrida PCNN-GCN. El método consta de dos componentes acoplados: (i) un algoritmo de segmentación bio-inspirado, PCNN-SAM, que transforma la retícula de píxeles en un conjunto de superpíxeles espectral y espacialmente coherentes, y (ii) un clasificador GCN semi-supervisado que propaga una cantidad extremadamente limitada de etiquetas a través del grafo construido sobre dichos superpíxeles. Los detalles de instanciación se presentan en las subsecciones siguientes; aquí se presenta la formulación general del método.

4.2.1. Visión general del método

El flujo de trabajo propuesto, ilustrado en la Figura 4.1, consta de cinco etapas: (1) normalización del cubo hiperspectral, (2) generación de superpíxeles mediante PCNN-SAM, (3) construcción de un grafo sobre las regiones resultantes, (4) extracción de características por nodo, y (5) clasificación semi-supervisada mediante GCN.

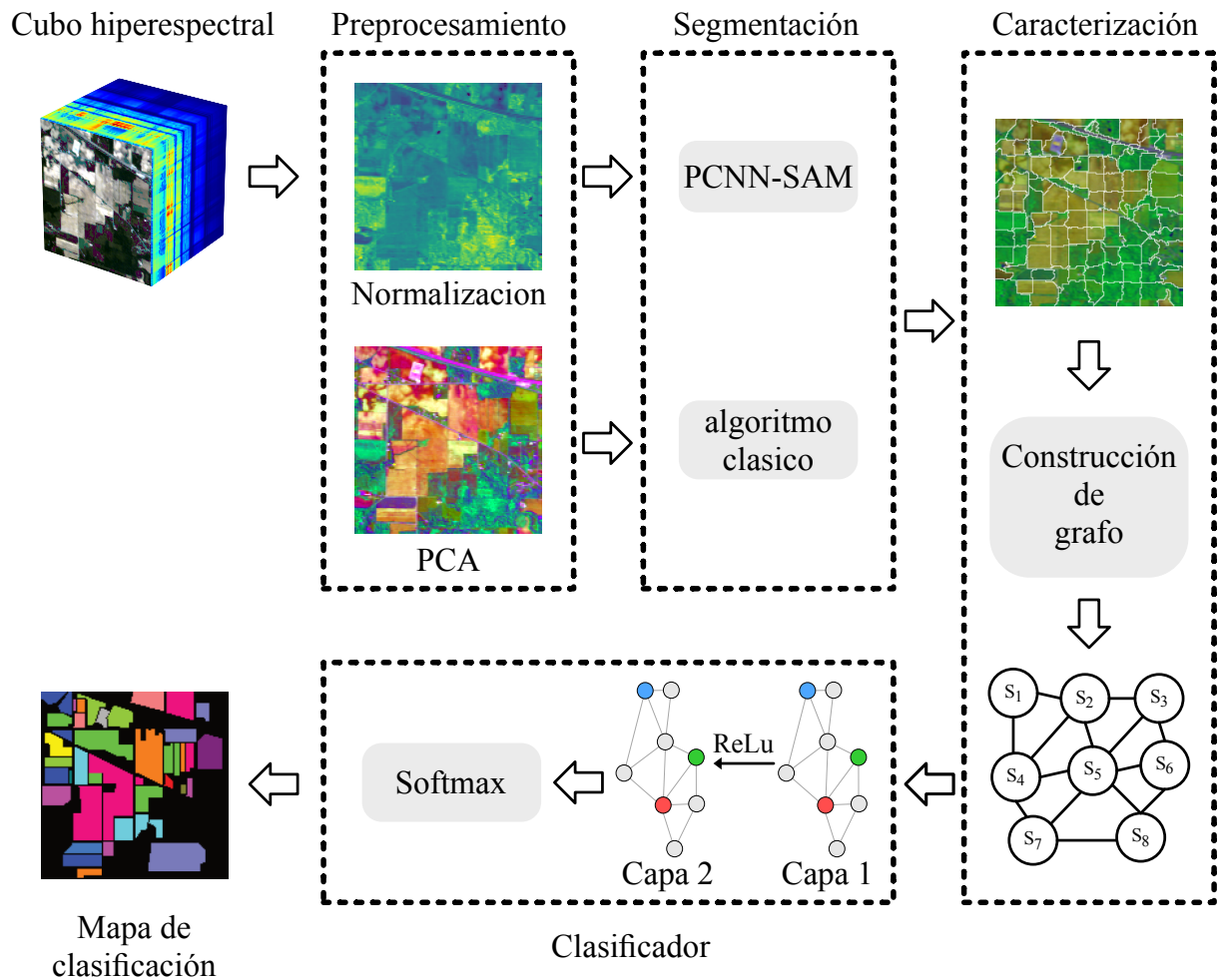


Figura 4.1: Flujo de trabajo propuesto: de la normalización del cubo hiperespectral a la clasificación semi-supervisada mediante GCN.

Esta secuencia responde directamente al problema de la escasez extrema de etiquetas: bajo esta condición, la GCN carece de evidencia suficiente para corregir errores heredados de una topología de grafo defectuosa. En consecuencia, la etapa de segmentación no se trata como un preprocesamiento intercambiable, sino como el componente que determina si la información espectral sutil sobrevive hasta la fase de clasificación. PCNN-SAM se diseñó bajo esa premisa: en lugar de reducir la dimensionalidad espectral antes de segmentar como exigen los algoritmos clásicos de superpíxeles, opera directamente sobre el vector espectral completo, delegando la reducción de dimensionalidad a una etapa posterior y exclusivamente para la construcción de características del grafo.

4.2.2. Conjunto de datos

La validación se realiza sobre seis conjuntos de datos hiperespectrales de uso extendido en la literatura de clasificación semi-supervisada.

Tabla 4.1: Propiedades de los conjuntos de datos hiperespectrales utilizados

Conjunto	Sensor	Dimensiones	Bandas	Rango espectral	Resolución espacial	Clases
Indian Pines	AVIRIS	145×145	200	0.4–2.5 μm	20 m/píxel	16
Salinas	AVIRIS	512×217	200	0.4–2.5 μm	3.7 m/píxel	16
Pavia University	ROSIS	610×340	103	430–860 nm	1.3 m/píxel	9
WHU-Hi-LongKou	Headwall Nano-Hyperspec	550×400	270	400–1000 nm	0.463 m/píxel	9
WHU-Hi-HanChuan	Headwall Nano-Hyperspec	1217×303	274	400–1000 nm	N/D	16
WHU-Hi-HongHu	Headwall Nano-Hyperspec	940×475	270	400–1000 nm	0.043 m/píxel	22

Indian Pines fue capturado con el sensor satelital AVIRIS sobre parcelas agrícolas en distintos estados de labranza, acompañadas de vegetación natural y una fracción menor de estructuras humanas (Figura 4.2). Es uno de los benchmarks más citados en la literatura de clasificación de HSI precisamente por combinar baja resolución espacial —lo que produce una alta proporción de píxeles mixtos en los bordes de parcela— con un desbalance severo entre clases.

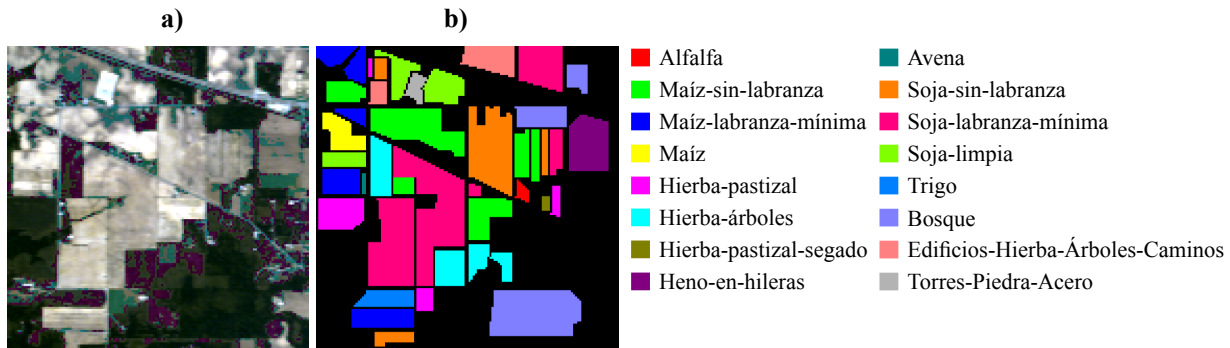


Figura 4.2: Conjunto de datos Indian Pines: composición en color e imagen de referencia (*ground truth*) con sus 16 clases.

Salinas corresponde a una escena AVIRIS sobre el valle de Salinas, California, con cultivos hortícolas en distintas etapas de desarrollo (Figura 4.3). A diferencia de Indian Pines, ofrece parcelas extensas, contiguas y geométricamente bien delimitadas, con una resolución espacial considerablemente más fina.

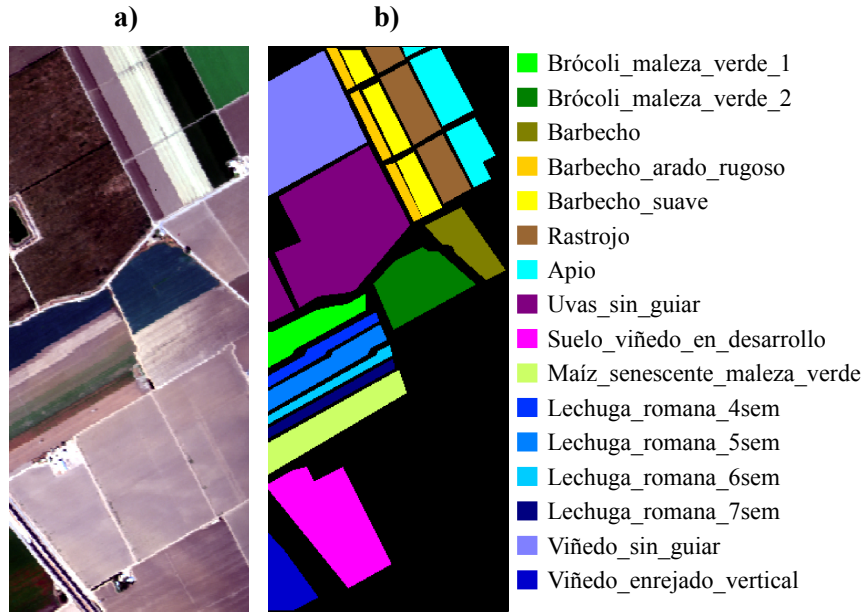


Figura 4.3: Conjunto de datos Salinas: composición en color e imagen de referencia (*ground truth*) con sus 16 clases.

Pavia University fue capturado por el sensor aerotransportado ROSIS sobre un entorno urbano construido en el norte de Italia (Figura 4.4). Introduce un régimen de dificultad distinto al de los conjuntos agrícolas: fronteras entre clases geoméricamente abruptas y ambigüedad espectral entre materiales artificiales de composición similar.

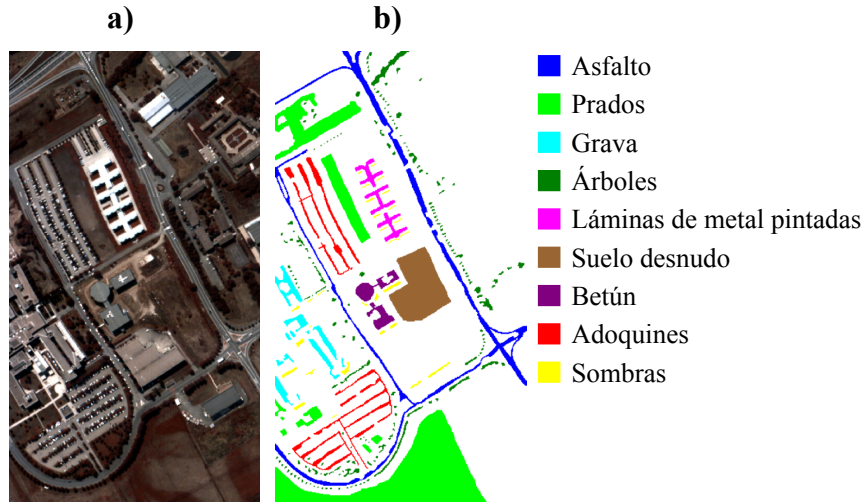


Figura 4.4: Conjunto de datos Pavia University: composición en color e imagen de referencia (*ground truth*) con sus 9 clases.

WHU-Hi-LongKou fue adquirido mediante un sensor Headwall Nano-Hyperspec montado

en UAV sobre una zona agrícola en Hubei, China, con seis cultivos dominantes (Figura 4.5). Su inclusión permite contrastar el desempeño del método sobre datos de plataforma UAV de muy alta resolución espacial, en un escenario con parcelas extensas y homogéneas [74].

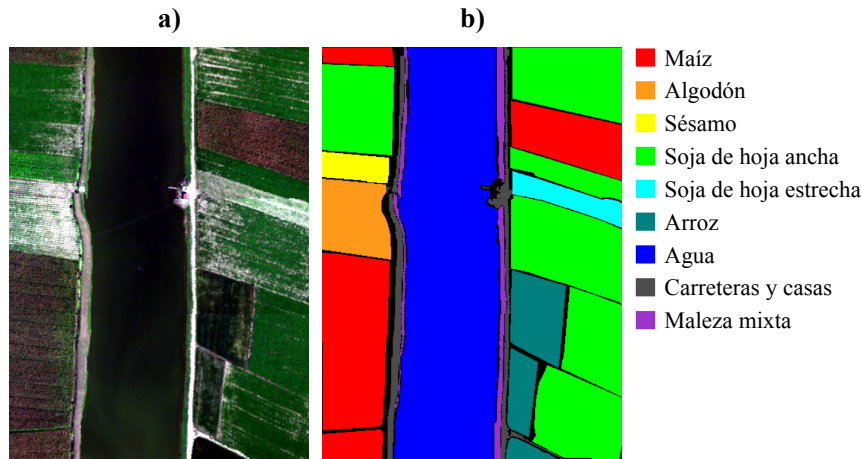


Figura 4.5: Conjunto de datos WHU-Hi-LongKou: composición en color e imagen de referencia (*ground truth*) con sus 9 clases.

WHU-Hi-HanChuan, capturado con la misma plataforma sobre una franja de transición rural-urbana, combina siete tipos de cultivo con infraestructura y cuerpos de agua (Figura 4.6). La captura se realizó con ángulo de elevación solar bajo, generando zonas extensas de sombra sobre la escena [74].

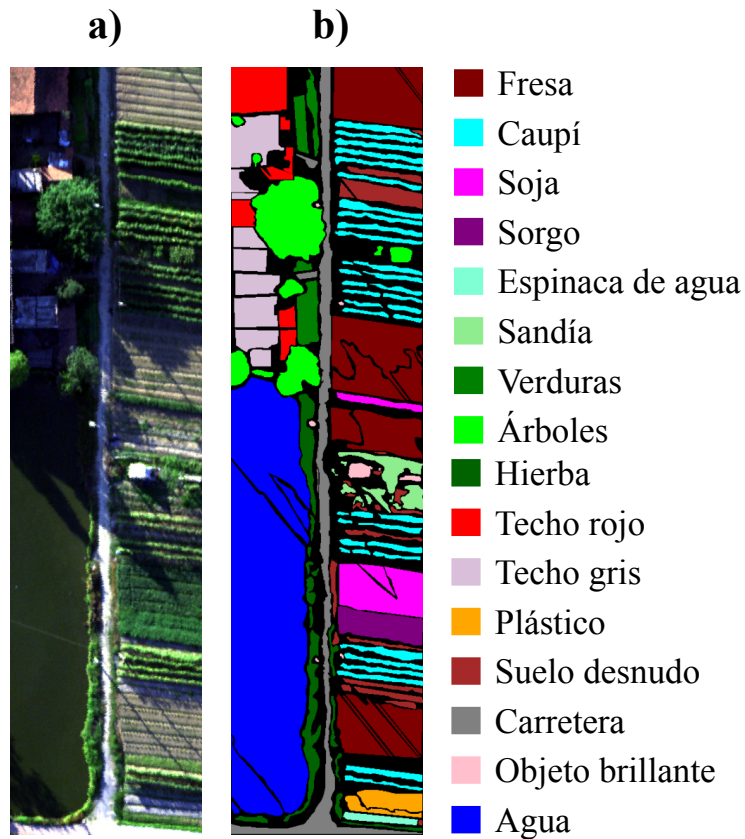


Figura 4.6: Conjunto de datos WHU-Hi-HanChuan: composición en color e imagen de referencia (*ground truth*) con sus 16 clases.

WHU-Hi-HongHu es el conjunto con mayor diversidad taxonómica del estudio: sus 22 clases incluyen múltiples cultivares de una misma familia vegetal —distintas variedades de col, mostaza y lechuga— con firmas espectrales altamente correlacionadas entre sí [74] (Figura 4.7).

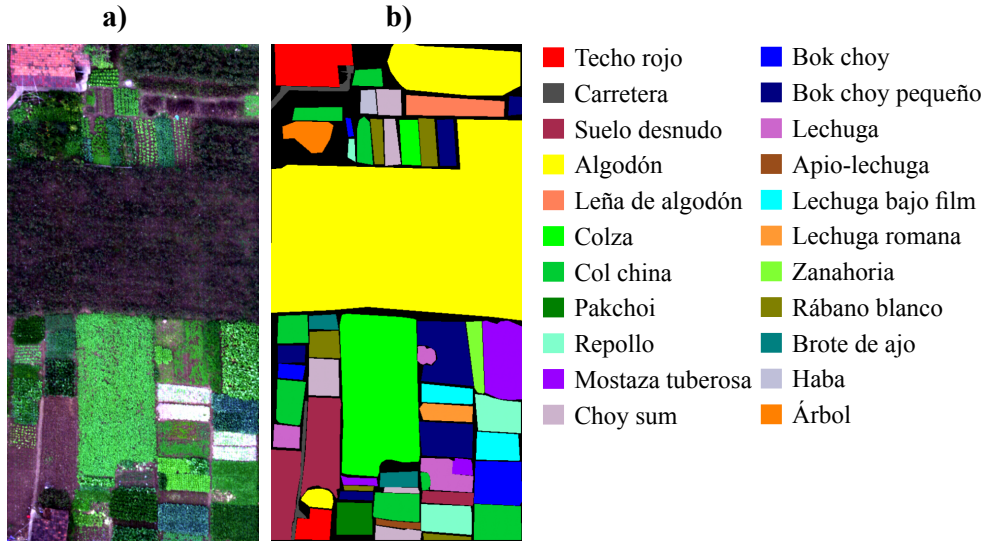


Figura 4.7: Conjunto de datos WHU-Hi-HongHu: composición en color e imagen de referencia (*ground truth*) con sus 22 clases.

4.2.3. Normalización espectral

Sea I el cubo hiperespectral original. Para su procesamiento en la red neuronal se deriva un tensor de estímulo F mediante una normalización *min-max* aplicada de manera independiente en cada banda espectral $b \in \{1, 2, \dots, B\}$:

$$F_{i,j,b} = \frac{I_{i,j,b} - \min(I_b)}{\max(I_b) - \min(I_b) + \epsilon}, \quad (4.1)$$

donde $\min(I_b)$ y $\max(I_b)$ son los valores extremos observados en la banda b sobre toda la imagen, y ϵ es un término de regularización que evita divisiones por cero. Esta normalización por banda —en contraste con una normalización global del cubo— es la que garantiza que ninguna banda con magnitudes de reflectancia atípicamente altas domine el estímulo F_{ij} de la Ecuación (4.1) y desestabilice el umbral de disparo dinámico de la red. La Figura 4.8 contrasta las firmas espectrales antes y después de esta transformación, mostrando que la morfología distintiva de cada clase se preserva pese al reescalamiento.

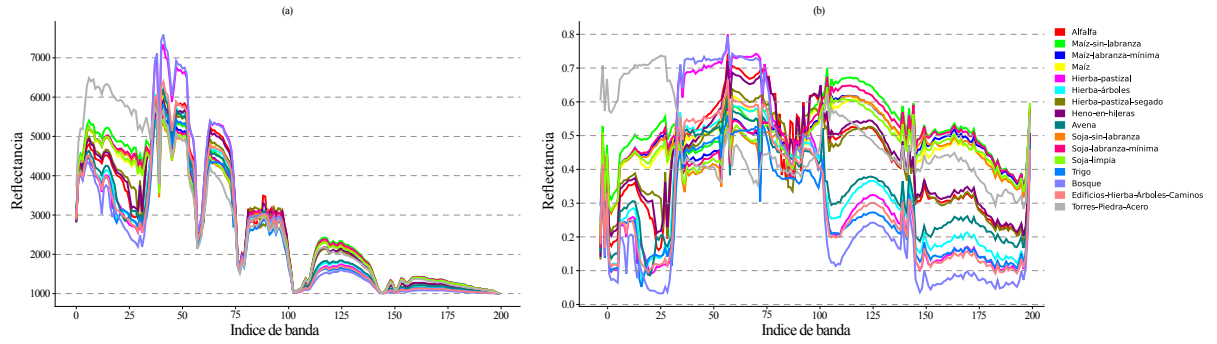


Figura 4.8: Firmas espectrales de un píxel representativo por cada clase: (a) perfiles originales sin transformación; (b) perfiles tras la normalización *min-max* independiente por banda (Ecuación (4.1)).

4.2.4. Segmentación mediante PCNN-SAM

El modelo base de la PCNN define su dinámica mediante los compartimentos de alimentación y acoplamiento (Ecs. Ecuación (2.8)–Ecuación (2.9)), el estado interno (Ec. Ecuación (2.10)), la regla de disparo (Ec. Ecuación (2.11)) y el umbral dinámico (Ec. Ecuación (2.12)). Ese modelo, sin embargo, no fue concebido para producir regiones espacialmente delimitadas ni para discriminar entre vecinos espectralmente similares y disímiles: cualquier píxel vecino que dispare influye sobre sus adyacentes con el mismo peso sináptico, sin importar cuán distinta sea su firma espectral.

La contribución de este trabajo consiste en modificar esa dinámica en cuatro aspectos: (a) sustituir el disparo espontáneo pixel a pixel por una inicialización de semillas sobre una retícula regular, (b) restringir el acoplamiento sináptico a vecinos espectralmente afines mediante la métrica SAM, (c) resolver la competencia entre superpíxeles colindantes mediante una función de energía híbrida espectral-espacial, y (d) consolidar morfológicamente las regiones resultantes. Cada uno se desarrolla a continuación.

Inicialización espacial de semillas

Siguiendo la lógica de inicialización de SLIC, el paso espacial entre centroides se define como

$$S = \max \left(\sqrt{\frac{H \cdot W}{K}}, 1 \right), \quad (4.2)$$

donde H y W son las dimensiones de la imagen y K el número deseado de superpíxeles. Las semillas se posicionan en el centro geométrico de cada parcela $S \times S$ y actúan como los núcleos desde los que se propaga la actividad de la red; su influencia local se evalúa sobre una vecindad de 3×3 .

Propagación restringida por similitud espectral

Sea $x_{ij} \in \mathbb{R}^B$ el vector espectral del píxel (i, j) . La divergencia angular respecto a un vecino se mide mediante el Mapeador de Ángulo Espectral:

$$\text{SAM}(x_{ij}, x_{mn}) = \arccos \left(\frac{x_{ij}^T x_{mn}}{\|x_{ij}\| \|x_{mn}\|} \right). \quad (4.3)$$

A diferencia del campo de acoplamiento genérico de la Ecuación (2.9), donde cualquier vecino activo $Y_{kl}[n-1]$ contribuye ponderado por un peso sináptico fijo $W_{ij,kl}$, el campo de enlace propuesto sustituye esa ponderación por una compuerta binaria condicionada a la similitud espectral:

$$L_{ij}[t] = L_{ij}[t-1]e^{-\alpha_i} + \sum_{(m,n) \in \mathcal{N}(i,j)} Y_{mn}[t-1] \cdot \mathbf{1}(\text{SAM}(x_{ij}, x_{mn}) < \tau). \quad (4.4)$$

Es decir, un vecino solo contribuye a la activación de (i, j) si disparó en la iteración anterior y su divergencia espectral respecto al píxel evaluado es inferior al umbral de tolerancia τ . Adicionalmente, se impone una restricción de pertenencia topológica: el estímulo solo se acumula si el vecino activado pertenece al superpíxel candidato que se encuentra en expansión durante esa iteración, evitando que regiones en crecimiento simultáneo se contaminen mutuamente antes de la fase de competencia.

El estado interno y la regla de disparo conservan la forma funcional de las Ecuaciones (2.10)–(2.11), evaluadas ahora sobre el campo de enlace restringido:

$$U_{ij}[t] = F_{ij}(1 + \beta L_{ij}[t]), \quad Y_{ij}[t] = \begin{cases} 1, & U_{ij}[t] > \Theta_{ij}[t] \\ 0, & \text{en otro caso} \end{cases}. \quad (4.5)$$

Competencia entre regiones y energía híbrida

Cuando múltiples superpíxeles intentan asimilar el mismo píxel en una misma iteración, el conflicto se resuelve mediante una función de energía que penaliza simultáneamente la divergencia espectral respecto al prototipo del superpíxel k y la distancia espacial respecto a su centroide:

$$E_k = \text{SAM}(x_{ij}, p_k) + w_{\text{spa}} \cdot d_{\text{spa}}((i, j), c_k). \quad (4.6)$$

El píxel se asigna al superpíxel que minimiza E_k , y w_{spa} actúa como ponderador de compacidad. Tras cada iteración, el umbral dinámico se actualiza según la misma forma de decaimiento exponencial de la Ecuación (2.12), y la matriz de tiempo de disparo $T_{ij} = \min\{t : Y_{ij}[t] = 1\}$ registra el instante de activación de cada píxel, sirviendo como bitácora del orden de propagación (Figura 4.9).

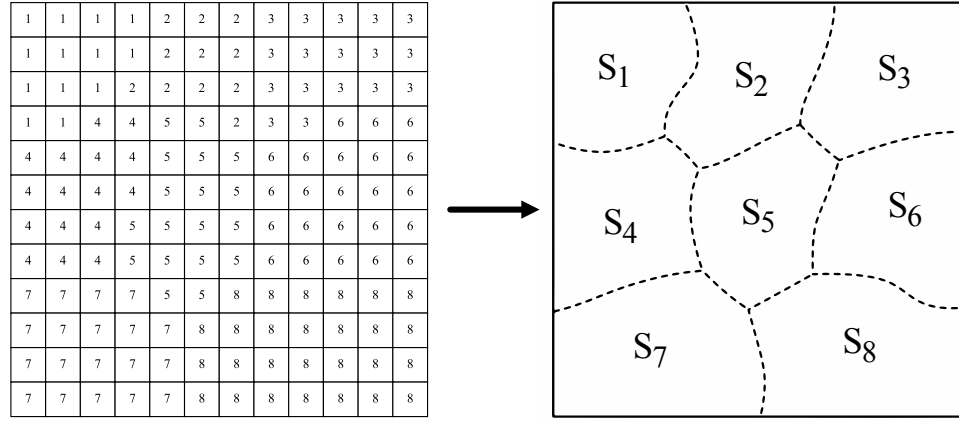


Figura 4.9: Ejemplo de la matriz de tiempo de disparo T_{ij} : el orden de activación de los píxeles delimita la formación de las regiones de los superpíxeles.

Refinamiento morfológico y consolidación

Los píxeles que no logran ser asignados durante la propagación —típicamente en fronteras entre materiales— se resuelven reevaluando la energía híbrida E_k (Ecuación (4.6)) sobre los superpíxeles ya consolidados en su vecindad. Las regiones resultantes por debajo de un umbral mínimo de área (microsuperpíxeles) se fusionan con el vecino cuya firma espectral media minimiza la distancia euclidiana en el espacio de características. Finalmente, las regiones se reetiquetan como componentes conexas del mapa T , produciendo el conjunto consolidado $\mathcal{S} = \{S_1, S_2, \dots, S_M\}$. La Figura 4.10 resume esta dinámica completa en forma de diagrama de bloques. El Algoritmo 4.1 presenta el procedimiento completo, y la Tabla 4.2 indica en qué líneas se materializa cada una de las cuatro modificaciones respecto al modelo clásico del Algoritmo 2.1.

Algoritmo 4.1 PCNN-SAM: segmentación en superpíxeles guiada por similitud espectral

Entrada: Cubo normalizado F (Ec. (4.1)); número objetivo de superpíxeles K ; tolerancia angular τ ; acoplamiento β ; decaimientos α_L, α_Θ ; peso de compacidad w_{spa} ; área mínima $A_{\text{mín}}$; iteraciones $N_{\text{máx}}$

Salida: Partición en superpíxeles $\mathcal{S} = \{S_1, \dots, S_M\}$

(a) Inicialización espacial de semillas

- 1: $S \leftarrow \text{máx}(\sqrt{H \cdot W / K}, 1)$ $\triangleright$ paso de la retícula, Ec. (4.2)
- 2: Colocar una semilla en el centro de cada celda $S \times S$ y asignarle una etiqueta k distinta
- 3: $\ell_{ij} \leftarrow 0$; $L \leftarrow \mathbf{0}$; $Y \leftarrow \mathbf{0}$; $T_{ij} \leftarrow 0$ para todo (i, j)

4: **para** $t = 1$ **hasta** $N_{\text{máx}}$ **hacer**

5: **para cada** píxel no asignado (i, j) adyacente a alguna región en expansión **hacer**

(b) Acoplamiento restringido por similitud espectral

- 6: $\mathcal{C} \leftarrow \{k : \exists (m, n) \in \mathcal{N}(i, j) \text{ con } \ell_{mn} = k, Y_{mn}[t-1] = 1, \text{SAM}(\mathbf{x}_{ij}, \mathbf{x}_{mn}) < \tau\}$
- 7: **si** $\mathcal{C} = \emptyset$ **entonces**
- 8: **continuar** $\triangleright$ ningún vecino activo es espectralmente afín
- 9: **fin si**
- 10: $L_{ij}[t] \leftarrow L_{ij}[t-1] e^{-\alpha_L} + |\{(m, n) \text{ que superan la compuerta}\}|$ $\triangleright$ Ec. (4.4)
- 11: $U_{ij}[t] \leftarrow F_{ij} (1 + \beta L_{ij}[t])$ $\triangleright$ Ec. (4.5)

(c) Competencia por energía híbrida

- 12: **si** $U_{ij}[t] > \Theta_{ij}[t]$ **entonces**
- 13: $k^* \leftarrow \arg \min_{k \in \mathcal{C}} [\text{SAM}(\mathbf{x}_{ij}, \mathbf{p}_k) + w_{\text{spa}} d_{\text{spa}}((i, j), \mathbf{c}_k)]$ $\triangleright$ Ec. (4.6)
- 14: $\ell_{ij} \leftarrow k^*$; $Y_{ij}[t] \leftarrow 1$; $T_{ij} \leftarrow t$
- 15: Actualizar el prototipo $\mathbf{p}_{k^*}$ y el centroide $\mathbf{c}_{k^*}$
- 16: **fin si**
- 17: $\Theta_{ij}[t+1] \leftarrow e^{-\alpha_\Theta} \Theta_{ij}[t] + V_\Theta Y_{ij}[t]$ $\triangleright$ Ec. (2.12)
- 18: **fin para**
- 19: **fin para**

(d) Refinamiento morfológico y consolidación

- 20: **para cada** píxel (i, j) con $\ell_{ij} = 0$ **hacer**
 - 21: $\ell_{ij} \leftarrow \arg \min_k E_k$ sobre los superpíxeles ya consolidados de su vecindad
 - 22: **fin para**
 - 23: Fusionar toda región con $|S_k| < A_{\text{mín}}$ con el vecino cuya firma espectral media esté más próxima
 - 24: Reetiquetar las componentes conexas de ℓ para obtener $\mathcal{S} = \{S_1, \dots, S_M\}$
 - 25: **devolver** $\mathcal{S}$
-

Tabla 4.2: Correspondencia entre las cuatro modificaciones introducidas por PCNN-SAM, las ecuaciones que las formalizan y las líneas del Algoritmo 4.1 en las que se aplican

Modificación	Efecto sobre la dinámica de la PCNN	Ecuación	Líneas
(a) Inicialización por semillas	Sustituye el disparo espontáneo píxel a píxel por núcleos de crecimiento distribuidos regularmente	(4.2)	1–3
(b) Compuerta espectral SAM	Un vecino activo solo contribuye al acoplamiento si su divergencia angular es inferior a τ	(4.3), (4.4)	6–10
(c) Energía híbrida espectral-espacial	Resuelve a qué superpíxel se asigna un píxel disputado por varias regiones	(4.6)	13
(d) Consolidación morfológica	Garantiza una partición conexa y sin píxeles sin asignar	—	20–24

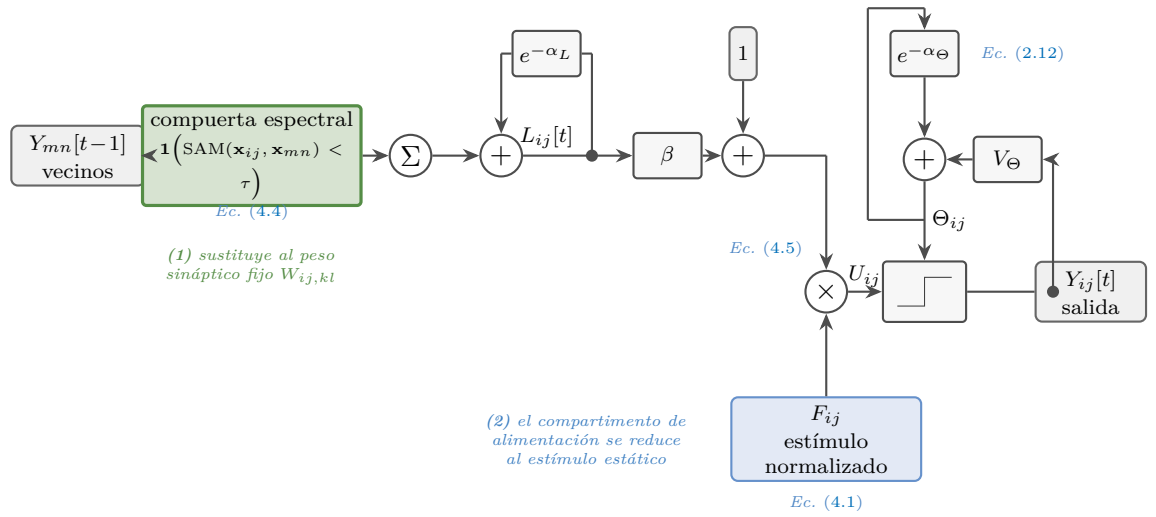


Figura 4.10: Dinámica de una neurona en PCNN-SAM, presentada en la misma notación que la Figura 2.5 para facilitar el contraste. Dos elementos cambian respecto al modelo clásico. (1) El peso sináptico fijo $W_{ij,kl}$ se sustituye por una compuerta binaria que solo deja pasar la contribución de los vecinos cuya divergencia angular respecto al píxel evaluado es inferior a τ . (2) El compartimento dinámico de alimentación se reduce al estímulo estático F_{ij} obtenido de la normalización por banda. El resto de la dinámica —modulación multiplicativa, comparación contra el umbral y realimentación exponencial del umbral— se conserva sin cambios.

4.2.5. Métodos de segmentación de referencia

PCNN-SAM se evalúa frente a cuatro algoritmos clásicos de generación de superpíxeles: SLIC, ERS, LSC y SEEDS. Estos cuatro métodos comparten la característica de operar sobre una representación de dimensionalidad reducida del cubo hiperspectral, lo cual constituye el contraste metodológico central de este estudio frente a PCNN-SAM.

4.2.6. De superpíxeles a grafo

Cada superpíxel $S_i \in \mathcal{S}$ se representa como un nodo de un grafo $G = (V, E)$. Dos nodos se conectan si sus regiones son espacialmente adyacentes:

$$A_{ij} = \begin{cases} 1, & \text{si } S_i \text{ y } S_j \text{ son espacialmente adyacentes} \\ 0, & \text{en otro caso} \end{cases}, \quad (4.7)$$

con matriz de adyacencia normalizada $\hat{A} = \bar{D}^{-1/2}(A + I)\bar{D}^{-1/2}$.

Para la matriz de características del grafo se introduce una asimetría deliberada respecto a la etapa de segmentación: mientras que PCNN-SAM opera sobre el vector espectral completo para preservar matices espectrales sutiles, los nodos del grafo se representan mediante los $d = 10$ componentes principales de dicho vector, promediados sobre los píxeles de cada región,

$$x_i = \frac{1}{|S_i|} \sum_{p \in S_i} z_p, \quad z_p \in \mathbb{R}^{10}. \quad (4.8)$$

Esta decisión de diseño —segmentar con información espectral completa pero clasificar con una representación comprimida— se retoma más adelante en la discusión general, donde se examina si la ventaja del método reside precisamente en que la pérdida de información ocurre *después* de que la topología del grafo ya capturó la pureza espectral, y no antes, como ocurre en los métodos clásicos de superpíxeles. La Figura 4.11 ilustra la correspondencia entre regiones y nodos.

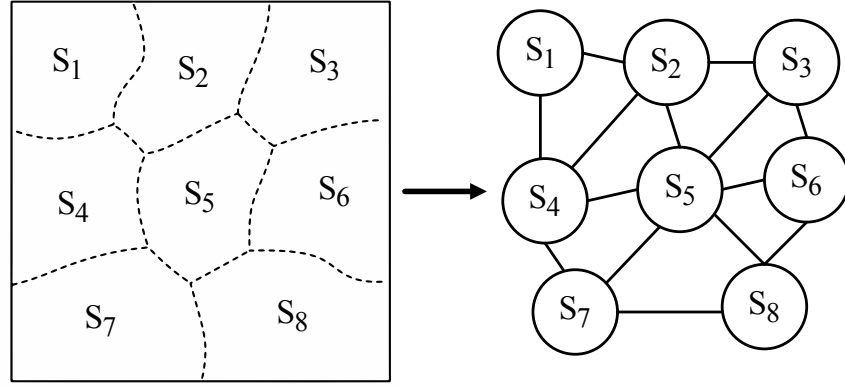


Figura 4.11: Correspondencia entre superpíxeles y nodos del grafo: cada región se abstrae como un nodo cuya característica es la firma espectral media, en 10 componentes principales, de los píxeles que la componen.

4.2.7. Clasificador GCN semi-supervisado sobre superpíxeles

Sobre el grafo construido se aplica el modelo de propagación de Kipf y Welling (Ecuación (3.6)), instanciado aquí como una red de dos capas. El conjunto de nodos V se particiona en un subconjunto etiquetado V_L —limitado a un número reducido de superpíxeles por clase— y un subconjunto no etiquetado V_U , sobre el cual se inferen las etiquetas por propagación, siguiendo un esquema semi-supervisado transductivo:

$$H^{(1)} = \sigma(\hat{A}XW^{(0)}), \quad Z_s = \hat{A}H^{(1)}W^{(1)}, \quad (4.9)$$

con salida $\hat{Y} = \text{softmax}(Z_s)$. El entrenamiento minimiza la entropía cruzada evaluada exclusivamente sobre V_L :

$$\mathcal{L} = - \sum_{i \in V_L} \sum_{c=1}^C Y_{ic} \ln \hat{Y}_{ic}. \quad (4.10)$$

Es este último punto el que materializa la hipótesis de este trabajo: dado que el gradiente solo se calcula sobre los nodos etiquetados, la calidad con la que $\hat{A}$ —es decir, la topología heredada de PCNN-SAM— conecta esos nodos con el resto del grafo determina por completo la capacidad del modelo para generalizar hacia V_U . La arquitectura concreta (número de capas, dimensión oculta, regularización) y el procedimiento de entrenamiento se especifican más adelante, como parte de la configuración experimental.

4.3. Configuración experimental

Esta sección especifica cómo se operacionalizó el método propuesto para su evaluación empírica: el protocolo de muestreo semi-supervisado, la arquitectura concreta del clasificador

GCN y las métricas de reporte.

4.3.1. Protocolo de muestreo semi-supervisado

La etapa de clasificación se evalúa bajo un régimen de aprendizaje semi-supervisado extremo: para cada conjunto de datos se seleccionan aleatoriamente únicamente 5 muestras (superpíxeles) por clase para conformar el conjunto de entrenamiento V_L , reservando la totalidad de los superpíxeles restantes como conjunto de prueba V_U . Este umbral se fijó deliberadamente por debajo del rango típicamente reportado en la literatura de clasificación semi-supervisada de HSI, con el fin de exponer las diferencias entre topologías de grafo bajo la condición más adversa posible: cuando prácticamente no hay margen para que la red corrija errores heredados de una segmentación defectuosa.

La construcción del grafo sigue el procedimiento descrito anteriormente: cada superpíxel se representa mediante el promedio de los 10 primeros componentes principales de los píxeles que lo componen, y las aristas conectan superpíxeles espacialmente adyacentes de forma bidireccional. Con el fin de controlar la varianza introducida por la selección aleatoria de las 5 muestras por clase, cada configuración experimental (dataset $\times$ método de segmentación $\times$ escala K) se ejecuta diez veces de manera independiente, reportándose los estadísticos correspondientes según se detalla a continuación.

Parámetros de PCNN-SAM

La Tabla 4.3 consolida los valores concretos con los que se instanció PCNN-SAM para generar los resultados reportados en este capítulo. La fuerza de acoplamiento β se aplica de manera uniforme sobre todas las bandas espectrales del cubo normalizado: cada banda mantiene su propio compartimento de acoplamiento $L_{ij}^{(b)}$, pero comparte el mismo valor de β y el mismo umbral dinámico Θ_{ij} , de modo que la neurona dispara en cuanto el estado interno de al menos una banda supera dicho umbral. Salvo K —cuya selección se justifica por separado en la Sección 4.4—, el resto de los parámetros se fijó mediante exploración manual sobre un subconjunto de los datos y se mantuvo constante a través de los seis conjuntos evaluados.

Tabla 4.3: Parámetros de PCNN-SAM empleados en la evaluación experimental

Símbolo	Descripción	Valor	Referencia
K	Número objetivo de superpíxeles	500	Sección 4.4
τ	Tolerancia angular de la compuerta SAM (rad)	0.15	Ec. (4.4)
β	Fuerza de acoplamiento	1.0	Ec. (4.5)
α_L	Decaimiento del compartimento de acoplamiento	0.5	Ec. (4.4)
α_Θ	Decaimiento del umbral dinámico	0.1	Ec. (2.12)
V_Θ	Incremento del umbral tras el disparo	10.0	Ec. (2.12)
Θ_0	Umbral inicial, antes del primer decaimiento	0.5	—
w_{spa}	Ponderador de compacidad espacial	0.15	Ec. (4.6)
$A_{\text{mín}}$	Área mínima antes de fusionar (píxeles)	8	Sección 4.2.4
$N_{\text{máx}}$	Número máximo de iteraciones	15	Sección 4.2.4
$\mathcal{N}$	Vecindad de evaluación	3×3	Sección 4.2.4

Arquitectura y entrenamiento de la GCN

Se implementaron y evaluaron dos variantes del clasificador descrito anteriormente: una arquitectura de una sola capa convolucional y una de dos capas. Esta restricción en la profundidad obedece a que, en redes convolucionales de grafos, un número excesivo de capas induce el fenómeno de sobresuavizado, que desvanece la capacidad discriminante de las firmas espectrales de los nodos.

En la variante de dos capas —utilizada para los resultados reportados más adelante en este capítulo—, la primera capa proyecta las 10 características de entrada de cada nodo a un espacio oculto de 64 dimensiones, seguida de una activación ReLU y una capa de regularización *Dropout* ($p = 0.5$). La segunda capa convolucional mapea dicho espacio oculto al número de clases del conjunto de datos correspondiente, con una activación *log-softmax* en la salida.

El entrenamiento se ejecuta durante 200 épocas mediante el optimizador Adam, con una tasa de aprendizaje de 0.01 y un decaimiento de pesos de 5×10^{-4} . La función de pérdida —probabilidad logarítmica negativa— se calcula y retropropaga exclusivamente sobre el subconjunto etiquetado V_L (Ecuación (4.10)). La Tabla 4.4 reúne la configuración completa del clasificador. Concluido el entrenamiento, la red infiere las etiquetas probabilísticas de la totalidad de los nodos del grafo; las predicciones a nivel de superpíxel se proyectan de regreso al espacio bidimensional de la imagen original, asignando a cada píxel la clase inferida de su superpíxel matriz y descartando las regiones correspondientes al fondo (clase cero).

Tabla 4.4: Configuración del clasificador GCN semi-supervisado

Componente	Valor	Justificación
Capas convolucionales de grafo	2	Evita el sobreajuste (Sección 3.3.2)
Dimensión de entrada por nodo	$d = 10$ componentes principales	Ec. (4.8)
Dimensión oculta	64	—
Dimensión de salida	C (clases del conjunto)	—
Activación intermedia	ReLU	—
Regularización	<i>Dropout</i> , $p = 0.5$	—
Activación de salida	<i>log-softmax</i>	—
Función de pérdida	Log-verosimilitud negativa sobre $\mathcal{V}_L$	Ec. (4.10)
Optimizador	Adam	—
Tasa de aprendizaje	0.01	—
Decaimiento de pesos	5×10^{-4}	—
Épocas	200	—
Muestras etiquetadas por clase	5	Sección 4.3.1
Ejecuciones independientes	10	Control de la varianza de muestreo

4.3.2. Métricas y protocolo de reporte

El desempeño se evalúa en dos niveles, correspondientes a las dos etapas del método propuesto:

- Calidad de segmentación, mediante las métricas ASA, BR y EV, definidas en el Capítulo 2.
- Calidad de clasificación, mediante las métricas OA, AA y κ , definidos en el Capítulo 3.

Adicionalmente, se registra el costo computacional de la etapa de segmentación, medido en segundos de ejecución, con el fin de contrastarlo contra la ganancia en exactitud al variar la escala K . Todas las métricas se reportan como el promedio sobre las diez ejecuciones independientes descritas anteriormente.

4.4. Selección de la escala de superpíxeles (K)

Antes de realizar el análisis comparativo detallado entre métodos de segmentación, es necesario fijar un valor único de granularidad K —el número de superpíxeles objetivo— bajo el cual dicha comparación sea justa para los cinco algoritmos evaluados. Esta sección responde a la primera de las preguntas planteadas al inicio del capítulo: ¿qué escala de superpíxel ofrece el mejor balance entre exactitud de clasificación y costo computacional bajo el régimen de 5 muestras por clase?

Para ello se evaluó el desempeño de los cinco métodos de segmentación —SLIC, LSC, ERS, SEEDS y PCNN-SAM— sobre los seis conjuntos de datos descritos anteriormente, variando la escala objetivo $K \in \{100, 250, 500, 1000\}$. Con el fin de que la escala seleccionada fuera robusta frente a la profundidad del clasificador y no estuviera sesgada hacia la arquitectura que finalmente se reportaría más adelante, este barrido se evaluó promediando los resultados de las dos variantes de GCN descritas anteriormente (una y dos capas convolucionales). Siguiendo el protocolo de muestreo descrito anteriormente, cada combinación de conjunto de datos, método, escala y arquitectura se ejecutó de manera independiente diez veces. Los valores de exactitud global reportados en esta sección corresponden, por lo tanto, a un promedio agregado sobre ambas arquitecturas; los resultados de clasificación presentados más adelante —donde se fija la arquitectura de dos capas por ofrecer el mejor desempeño observado— son consecuentemente superiores en magnitud absoluta a los aquí discutidos, aunque la tendencia respecto a K se mantiene consistente entre ambas arquitecturas.

Exactitud de clasificación. Como se observa en la Figura 4.12, incrementar la cantidad de superpíxeles mejora inicialmente el OA promedio: el rendimiento asciende de 62.90 % con $K = 100$ a 65.20 % con $K = 250$, y continúa una tendencia positiva hasta alcanzar su punto máximo en 66.68 % para $K = 500$. Al aumentar la resolución a $K = 1000$, sin embargo, la mejora no solo se estanca sino que retrocede ligeramente, a 66.57 %. Este comportamiento es consistente con la hipótesis de que una sobresegmentación excesiva fragmenta la coherencia estructural de las parcelas, introduciendo ruido topológico en el grafo que perjudica la propagación de etiquetas durante la convolución.

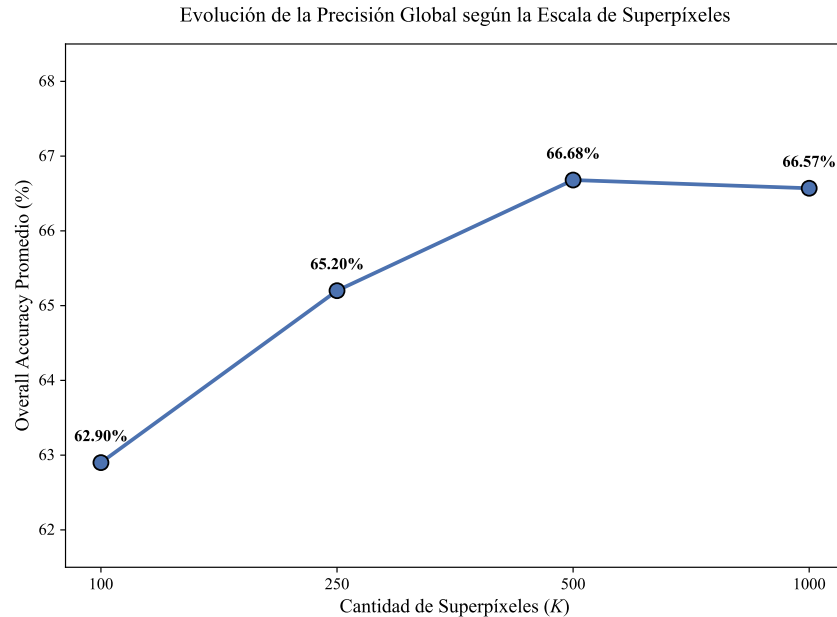


Figura 4.12: Evolución de la exactitud global (OA) promedio, sobre los seis conjuntos de datos y ambas arquitecturas de GCN, en función de la escala de superpíxeles K .

Costo computacional. El análisis del tiempo de ejecución (Figura 4.13) revela un incremento monótono conforme crece K : el procesamiento de segmentación requiere en promedio 7.92 segundos con $K = 100$, asciende a 9.94 segundos con $K = 250$, continúa hasta 11.57 segundos con $K = 500$, y se prolonga hasta 13.11 segundos con $K = 1000$ —es decir, casi el doble del costo de la configuración más gruesa para una ganancia de exactitud que, como se mostró arriba, ya se había revertido.

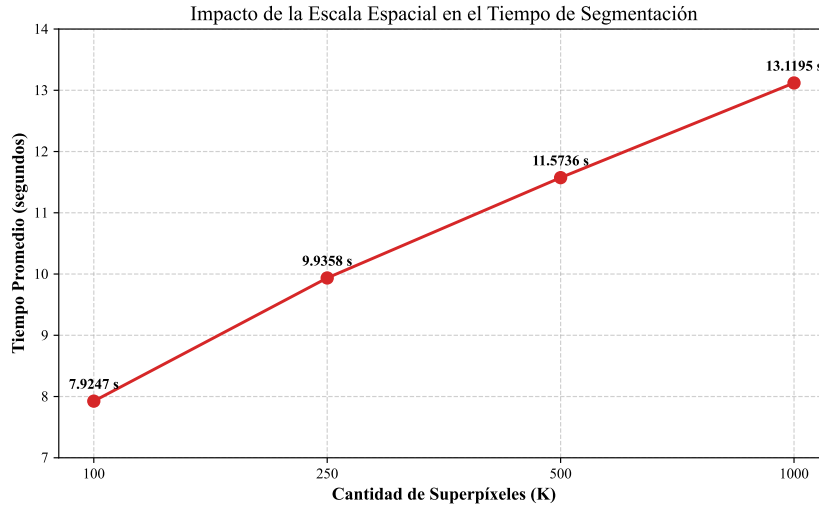


Figura 4.13: Tiempo de ejecución promedio de la etapa de segmentación en función de la escala de superpíxeles K .

Selección de la escala óptima. Al contrastar ambas curvas, $K = 500$ domina la decisión en los dos ejes de interés: ofrece la mayor exactitud de clasificación observada en el barrido completo, sin incurrir en el costo computacional adicional que exige $K = 1000$ a cambio de una ligera pérdida de desempeño. Por esta razón, $K = 500$ se fija como la escala de superpíxeles para el desglose comparativo detallado entre métodos de segmentación que se presenta a continuación.

4.5. Resultados y análisis

Esta sección presenta la evidencia empírica correspondiente a las preguntas 2 y 3 planteadas al inicio del capítulo: qué estrategia de segmentación produce la topología de grafo más favorable para la propagación semi-supervisada, y si la preservación del vector espectral completo por parte de PCNN-SAM constituye una ventaja medible. Los resultados se presentan primero de manera consolidada por etapa (segmentación y clasificación), después se cuantifica la relación entre ambas, y finalmente se examinan los mapas de clasificación resultantes.

4.5.1. Calidad de la segmentación por método

La Tabla 4.5 consolida las métricas de segmentación —ASA, BR y EV, definidas en el Capítulo 2— para los cinco métodos evaluados sobre los seis conjuntos de datos, a la escala $K = 500$ seleccionada anteriormente.

Tabla 4.5: Métricas de segmentación consolidadas (ASA, BR, EV) por conjunto de datos y método (K=500)

Conjunto de datos	Método	ASA	BR	EV
Indian Pines	SLIC	97.83	0.9871	0.8528
	LSC	96.13	0.9948	0.8429
	ERS	96.64	0.9956	0.8645
	SEEDS	95.36	0.9682	0.8414
	PCNN-SAM	97.83	0.9944	0.8654
Salinas	SLIC	99.45	0.7099	0.8751
	LSC	98.66	0.6311	0.9237
	ERS	98.01	0.7916	0.9114
	SEEDS	97.99	0.7224	0.9040
	PCNN-SAM	97.21	0.6493	0.8845
Pavia University	SLIC	94.07	0.3857	0.3572
	LSC	97.30	0.4742	0.5517
	ERS	98.05	0.5866	0.5665
	SEEDS	96.41	0.4810	0.4741
	PCNN-SAM	95.33	0.4954	0.4105
LongKou	SLIC	95.33	0.6976	0.9278
	LSC	96.76	0.6627	0.9464
	ERS	98.78	0.8039	0.9434
	SEEDS	96.81	0.6821	0.9358
	PCNN-SAM	94.29	0.5710	0.9171
HanChuan	SLIC	90.43	0.4416	0.5564
	LSC	92.57	0.3870	0.6770
	ERS	91.19	0.5682	0.6630
	SEEDS	93.01	0.4292	0.7280
	PCNN-SAM	93.36	0.4374	0.6057
HongHu	SLIC	91.92	0.5938	0.7738
	LSC	95.00	0.5083	0.8356
	ERS	94.52	0.6324	0.7817
	SEEDS	94.75	0.5377	0.8212
	PCNN-SAM	92.67	0.4194	0.7462

Ningún método domina las tres métricas de manera consistente. SLIC y ERS se alternan el liderazgo en ASA dependiendo de la regularidad geométrica de la escena (SLIC en Indian Pines y Salinas; ERS en Pavia University y LongKou), mientras que ERS es sistemáticamente el método con mayor adherencia a bordes, consistente con su formulación

con base en optimización de entropía sobre la estructura del grafo. PCNN-SAM, por su parte, únicamente lidera en ASA sobre HanChuan y empata con SLIC en Indian Pines; en el resto de los conjuntos su pureza de segmentación es intermedia o la más baja del grupo. Esta observación es importante porque, como se cuantifica más adelante, la ventaja de PCNN-SAM en la etapa de clasificación no se explica por una superioridad en estas métricas geométricas clásicas.

A continuación se detalla, dataset por dataset, la evidencia detrás de esta tabla consolidada.

Indian Pines

Las métricas de segmentación para *Indian Pines* reflejan el desafío de procesar una imagen con baja resolución espacial. Se observa que SLIC y PCNN logran la mayor pureza de segmentación ($ASA = 97.83\%$), mientras que ERS destaca por su excelente adherencia a los bordes reales de las parcelas ($BR = 0.9956$). Es notable que la PCNN obtenga la mayor Varianza Explicada ($EV = 0.8654$); esto es un reflejo directo de su capacidad para procesar el vector espectral completo, capturando toda la varianza de los datos. Sin embargo, en imágenes ruidosas esta retención total del espectro puede introducir interferencias que dificultan la propagación de etiquetas en la fase de clasificación, favoreciendo en la práctica a métodos que imponen una mayor regularidad geométrica y compacidad, como SLIC. La Figura 4.14 muestra la partición en superpíxeles producida por cada método sobre esta escena.

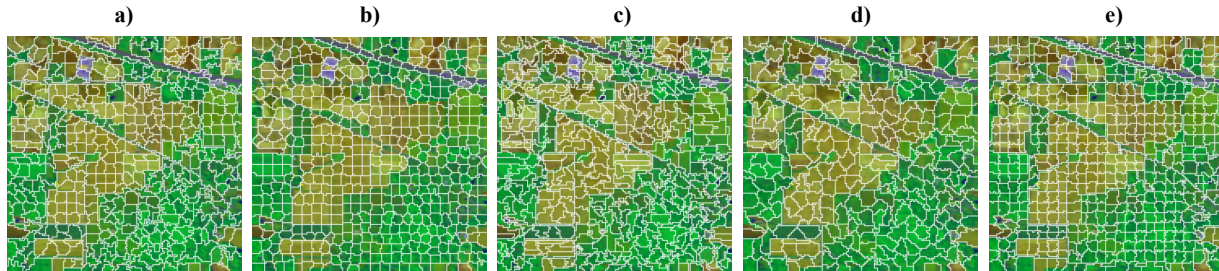


Figura 4.14: Partición en superpíxeles sobre Indian Pines con los métodos a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM.

Salinas

El conjunto *Salinas*, caracterizado por regiones extensas y homogéneas, encuentra en SLIC su método de segmentación más óptimo a nivel de pureza espacial. SLIC roza la perfección en la métrica ASA (99.45%), garantizando agrupaciones de alta fidelidad semántica. Esta alta homogeneidad interna y regularidad geométrica son las responsables de que la GCN logre su máximo rendimiento en la etapa de clasificación con este método. Por su parte, LSC demuestra ser el algoritmo con mayor retención de la varianza local

($EV = 0.9237$), mientras que ERS se consolida como el más eficaz para ajustarse a los bordes estructurales del terreno ($BR = 0.7916$). La Figura 4.15 muestra la partición en superpíxeles producida por cada método sobre esta escena.

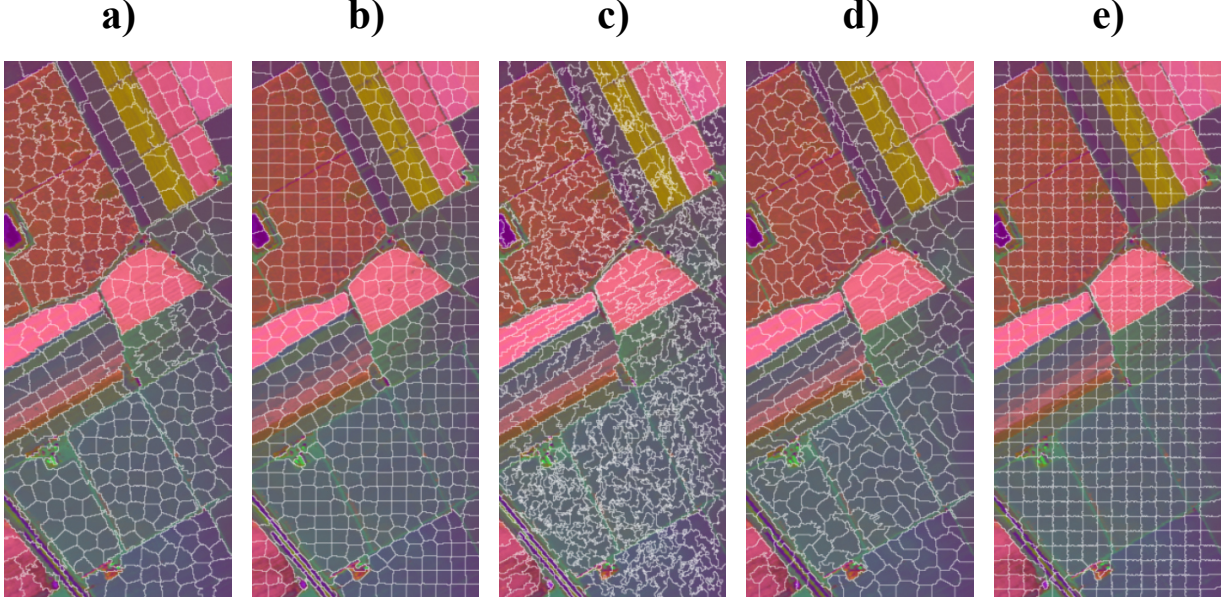


Figura 4.15: Partición en superpíxeles sobre Salinas con los métodos a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM.

Pavia University

En el entorno urbano de *Pavia University*, la delimitación exacta de las estructuras es crítica. El método ERS sobresale contundentemente en las métricas de pureza ($ASA = 98.05\%$), recuperación de bordes ($BR = 0.5866$) y varianza explicada ($EV = 0.5665$). En una escena donde los límites entre asfalto, metal y sombras son abruptos, el alto índice de BR de ERS garantiza que los superpíxeles se ajusten firmemente a las fronteras de los objetos, reduciendo la contaminación cruzada inicial. Sin embargo, resulta revelador que, aunque ERS entrega la mejor integridad estructural en esta etapa (lo que le permite alcanzar posteriormente la mejor exactitud promedio), la GCN logró su máximo OA capitalizando la distribución rígidamente regular de SLIC. Esto evidencia un complejo equilibrio en el aprendizaje basado en grafos: la adherencia perfecta a los bordes no siempre supera la robustez que ofrece la compacidad geométrica frente a la escasez extrema de etiquetas. La Figura 4.16 muestra la partición en superpíxeles producida por cada método sobre esta escena.

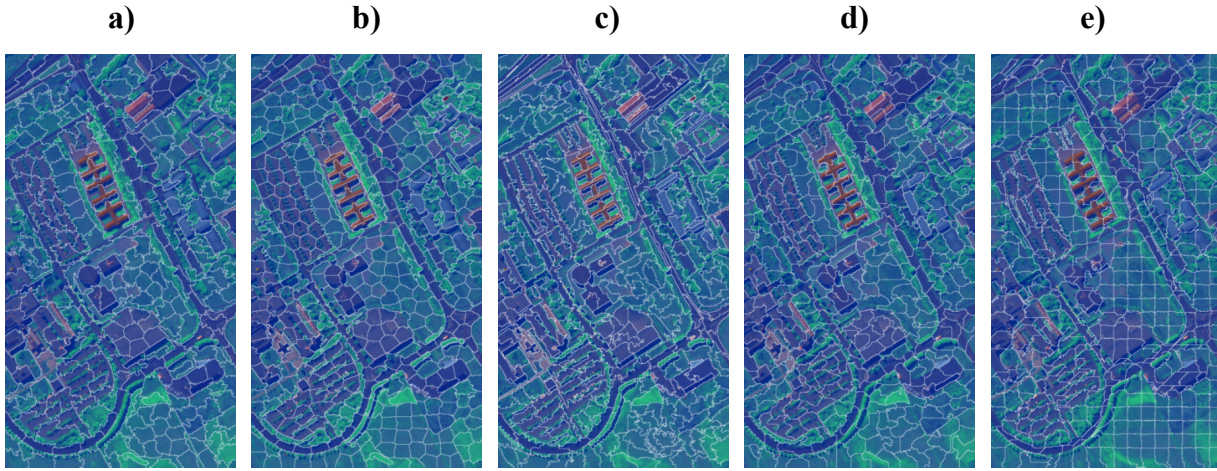


Figura 4.16: Partición en superpíxeles sobre Pavia University con los métodos a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM.

LongKou

Para el dataset *LongKou*, el algoritmo ERS vuelve a demostrar su superioridad en la recuperación de límites espaciales ($BR = 0.8039$) y en la pureza de los segmentos ($ASA = 98.78\%$). La formulación de ERS, basada en la optimización de la tasa de entropía, resulta ideal para delinear las complejas texturas presentes en la escena. De manera complementaria, LSC retiene el mayor porcentaje de información estructural y varianza local ($EV = 0.9464$). Ambas propiedades, la alta adherencia a bordes por parte de ERS y la óptima retención de varianza por LSC, se tradujeron directamente en la generación de grafos altamente coherentes, permitiendo a la GCN alcanzar exactitudes de clasificación superiores al 85% en ambos métodos, con ERS consolidándose como el ganador indiscutible en este conjunto. La Figura 4.17 muestra la partición en superpíxeles producida por cada método sobre esta escena.

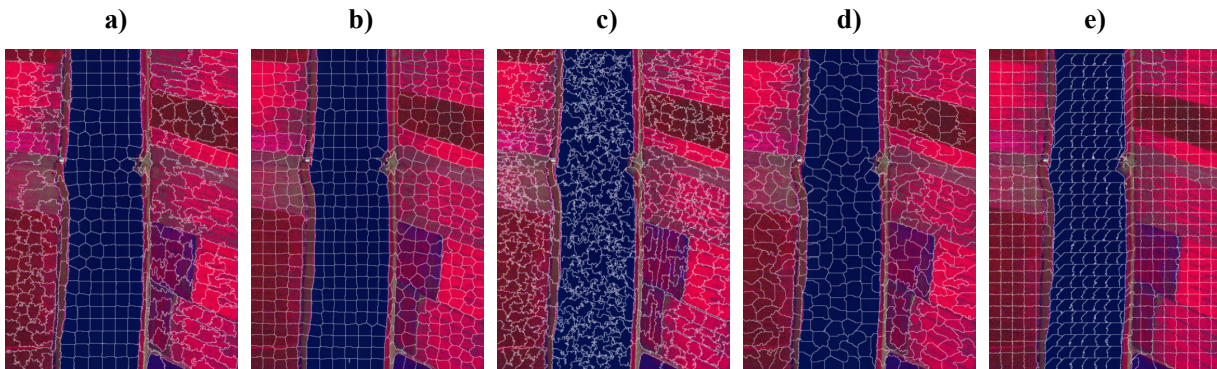


Figura 4.17: Partición en superpíxeles sobre WHU-Hi-LongKou con los métodos a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM.

HanChuan

Las métricas de segmentación en *HanChuan* muestran una distribución de rendimiento muy fragmentada, lo cual es indicativo de la alta complejidad espectral y la mezcla intrincada de píxeles en la imagen. Mientras que la arquitectura bio-inspirada PCNN obtiene la mayor pureza de segmentación ($ASA = 93.36 \%$), SEEDS logra maximizar la retención de varianza local ($EV = 0.7280$), y ERS mantiene su característico liderazgo en la recuperación de los bordes estructurales ($BR = 0.5682$). Esta dispersión en la calidad geométrica y espectral explica las dificultades generales experimentadas en la posterior fase de clasificación (donde los valores de exactitud global apenas rozan el 60%). Como ningún algoritmo clásico logró generar un grafo que combinara simultáneamente alta pureza, retención de varianza y precisión de bordes, el aprendizaje semi-supervisado de la GCN se vio limitado. Sin embargo, resulta destacable que la superioridad de la PCNN en la métrica fundamental de pureza semántica ASA fue el factor determinante que le permitió, posteriormente, liderar la métrica OA del modelo conjunto. La Figura 4.18 muestra la partición en superpíxeles producida por cada método sobre esta escena.

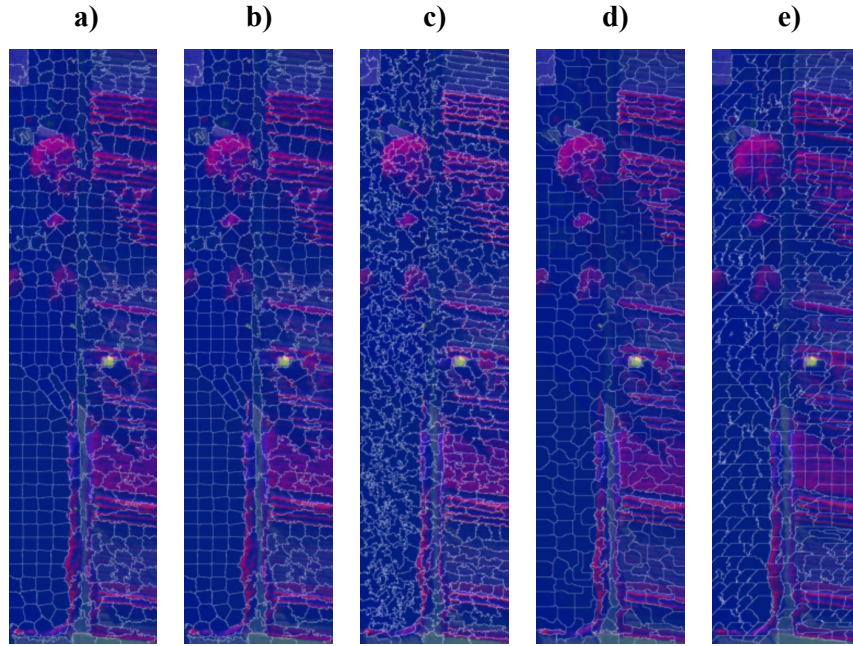


Figura 4.18: Partición en superpíxeles sobre WHU-Hi-HanChuan con los métodos a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM.

HongHu

El análisis del conjunto *HongHu* revela una disonancia sumamente interesante entre las métricas geométricas tradicionales y la utilidad real de los superpíxeles para la GCN. Cuantitativamente, LSC y ERS dominan las métricas espaciales previas: LSC alcanza

la mejor pureza de segmentación ($ASA = 95.00\%$) y la máxima retención de varianza local ($EV = 0.8356$), mientras que ERS lidera en el ajuste a los bordes ($BR = 0.6324$). Sin embargo, fue la arquitectura PCNN la que obtuvo los mejores resultados en la etapa final de clasificación para este dataset. Esto demuestra empíricamente que las métricas de segmentación clásicas no logran capturar la sutil riqueza espectral que la PCNN preserva al procesar todo el vector espectral sin un paso de reducción de dimensionalidad previo. En imágenes con múltiples clases de coberturas y cultivos altamente correlacionados, la fidelidad espectral profunda de la PCNN superó con creces a la regularidad y perfección geométrica de LSC o ERS, dotando a la red GCN de características discriminativas muy superiores frente a la extrema escasez de etiquetas. La Figura 4.19 muestra la partición en superpíxeles producida por cada método sobre esta escena.

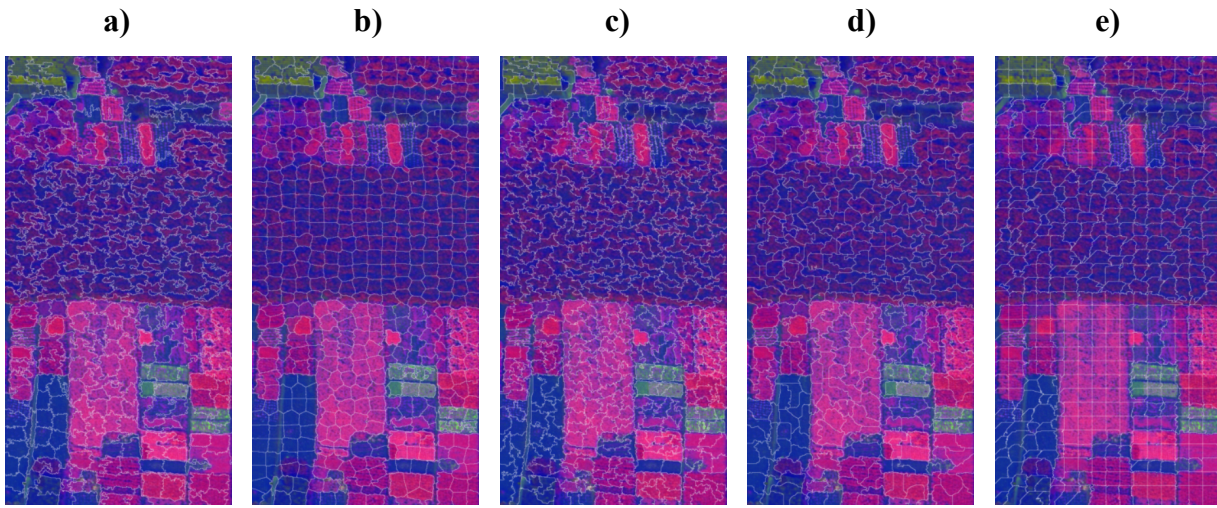


Figura 4.19: Partición en superpíxeles sobre WHU-Hi-HongHu con los métodos a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM.

La evaluación exhaustiva de los métodos de segmentación de superpíxeles evidencia que la construcción de la topología del grafo es el factor más determinante para el éxito de una GCN en un entorno semi-supervisado con escasez extrema de etiquetas (5 muestras por clase). Los resultados demuestran que no existe un algoritmo universalmente superior, sino que la elección del método depende de la naturaleza geométrica y espectral de la imagen hiperespectral. Métodos basados en restricción espacial iterativa y optimización de entropía (como SLIC y ERS) resultan idóneos cuando la imagen exige alta compacidad en áreas homogéneas (como en Salinas y Pavia University, dominados por SLIC) o una fuerte adherencia a fronteras estructurales complejas (como en LongKou, dominado por ERS). Estos algoritmos garantizan nodos regulares y evitan la contaminación espacial de características durante la convolución del grafo. Por otro lado, la integración de modelos biológicos como la PCNN introduce un paradigma diferente: al prescindir de la compresión espectral previa y procesar el vector original completo, se prioriza la preservación de

anomalías, pureza semántica y matices espectrales por encima de la regularidad geométrica. Esta característica demostró ser la clave del éxito en conjuntos de datos con una gran diversidad de clases altamente correlacionadas (como HongHu y HanChuan), compensando con riqueza espectral lo que las métricas tradicionales de evaluación espacial estricta (como BR) no logran medir.

4.5.2. Desempeño de clasificación semi-supervisada

La Tabla 4.6 consolida los resultados de clasificación —OA, AA y kappa bajo el protocolo de 5 muestras por clase descrito anteriormente.

La Tabla 4.7 resume el número de conjuntos de datos en que cada método obtiene el mejor valor por métrica, y el promedio global de cada métrica sobre los seis conjuntos de datos.

La Figura 4.20 presenta los mismos valores de OA de forma comparativa por conjunto de datos.

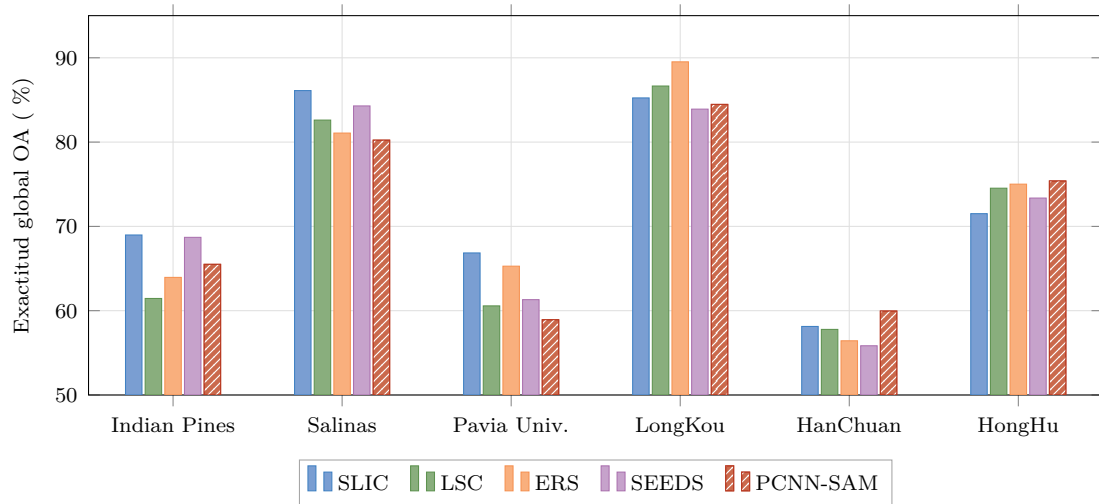


Figura 4.20: Exactitud global (OA) por conjunto de datos y método de segmentación, con $K = 500$ y cinco muestras etiquetadas por clase. PCNN-SAM (barra rayada) no domina el promedio, pero es el método más alto en HanChuan y HongHu, los dos conjuntos con mayor correlación espectral entre clases.

Dos lecturas se desprenden de esta tabla y conviene distinguirlas con cuidado. Por número de conjuntos ganados, SLIC domina (3 de 6 en OA y en kappa), seguido de PCNN-SAM (2 de 6) y ERS (1 de 6); en AA, sin embargo, SLIC y ERS empatan con 3 conjuntos ganados cada uno. Por promedio global las diferencias entre los cinco métodos son modestas —apenas 2.2 puntos porcentuales de OA separan al mejor (SLIC, 72.80 %) del peor (LSC, 70.60 %)—, y PCNN-SAM no ocupa ni el primer ni el último lugar en esta métrica agregada. Esto confirma cuantitativamente lo que ya se observaba cualitativamente

Tabla 4.6: Métricas de clasificación consolidadas (OA, AA, κ) por conjunto de datos y método (K=500)

Conjunto de datos	Método	OA (%)	AA (%)	$\hat{\kappa}$
Indian Pines	SLIC + GCN	68.98	77.17	0.6527
	LSC + GCN	61.45	75.02	0.5702
	ERS + GCN	63.95	70.86	0.5967
	SEEDS + GCN	68.70	69.10	0.6485
	PCNN-SAM + GCN	65.50	71.55	0.6145
Salinas	SLIC + GCN	86.11	89.92	0.8462
	LSC + GCN	82.61	87.67	0.8076
	ERS + GCN	81.07	86.18	0.7914
	SEEDS + GCN	84.29	89.22	0.8262
	PCNN-SAM + GCN	80.23	84.55	0.7819
Pavia University	SLIC + GCN	66.85	63.57	0.5790
	LSC + GCN	60.57	62.18	0.5126
	ERS + GCN	65.28	66.55	0.5663
	SEEDS + GCN	61.31	60.80	0.5151
	PCNN-SAM + GCN	58.93	59.24	0.4925
LongKou	SLIC + GCN	85.24	71.54	0.8096
	LSC + GCN	86.65	81.30	0.8286
	ERS + GCN	89.52	87.50	0.8649
	SEEDS + GCN	83.91	73.91	0.7938
	PCNN-SAM + GCN	84.47	78.92	0.8025
HanChuan	SLIC + GCN	58.13	58.40	0.5252
	LSC + GCN	57.78	57.42	0.5207
	ERS + GCN	56.43	54.19	0.5054
	SEEDS + GCN	55.84	54.11	0.4989
	PCNN-SAM + GCN	59.96	54.49	0.5431
HongHu	SLIC + GCN	71.51	64.62	0.6547
	LSC + GCN	74.53	68.48	0.6889
	ERS + GCN	75.01	72.35	0.6932
	SEEDS + GCN	73.36	66.78	0.6740
	PCNN-SAM + GCN	75.39	65.10	0.6966

Tabla 4.7: Resumen del desempeño de clasificación: conjuntos de datos ganados por métrica y promedio global sobre los seis conjuntos

Método	Conjuntos ganados			Promedio global		
	OA	AA	$\hat{\kappa}$	OA (%)	AA (%)	$\hat{\kappa}$
SLIC + GCN	3	3	3	72.80	70.87	0.6779
LSC + GCN	0	0	0	70.60	72.01	0.6548
ERS + GCN	1	3	1	71.88	72.94	0.6697
SEEDS + GCN	0	0	0	71.24	68.99	0.6594
PCNN-SAM + GCN	<u>2</u>	0	<u>2</u>	70.75	68.98	0.6552

Negritas: mejor valor. Subrayado: segundo lugar en número de conjuntos ganados.

en el análisis dataset por dataset: no existe un método universalmente superior; cada uno gana en un régimen distinto de condiciones espaciales y espectrales, y PCNN-SAM en particular gana exactamente en los dos conjuntos con mayor complejidad espectral y fragmentación espacial del estudio (HanChuan y HongHu), no en el promedio general. Esta distinción —ganador por régimen, no ganador promedio— es el argumento central que se retoma en la discusión general.

A continuación se detalla, dataset por dataset, la evidencia detrás de esta tabla consolidada, junto con los mapas de clasificación correspondientes.

Como se observa en la Tabla 4.6 y en los mapas de clasificación correspondientes, el conjunto de datos *Indian Pines* representa un escenario desafiante debido a su baja resolución espacial, el alto nivel de ruido y la mezcla de sus parcelas agrícolas. En este contexto, el método SLIC acoplado a la GCN demuestra un rendimiento superior absoluto, alcanzando la mejor exactitud global (OA = 68.98 %), la máxima exactitud promedio (AA = 77.17 %) y el mayor coeficiente Kappa ($\kappa = 0.6527$). Esto sugiere que, frente a un escenario ruidoso, la estrategia de agrupar píxeles forzando una alta compacidad y una regularidad espacial rígida (SLIC) mitiga eficazmente las interferencias y favorece una propagación de etiquetas más estable en el grafo, protegiendo incluso la clasificación de las clases minoritarias. SEEDS muestra también un comportamiento competitivo, quedando muy cerca en OA. En contraste, LSC presenta la exactitud global más baja (OA = 61.45 %). Por su parte, la PCNN exhibe un rendimiento intermedio (OA = 65.50 %); al utilizar el vector espectral original completo sin una etapa previa de filtrado de ruido o reducción de dimensionalidad (como PCA), el enfoque bio-inspirado resulta susceptible a las interferencias inherentes de este sensor específico, lo que introduce ruido en las conexiones iniciales del grafo y limita la eficacia de la red frente a los métodos de restricción espacial estricta.

Los resultados para el conjunto *Salinas* ilustran el comportamiento de los algoritmos en un escenario de alta homogeneidad espacial, caracterizado por extensas parcelas agrícolas de contornos bien definidos. Bajo estas condiciones, el algoritmo SLIC domina consistentemente la clasificación, logrando los valores máximos en OA (86.11 %), AA (89.92 %) y κ (0.8462).

La naturaleza iterativa de SLIC, que restringe la búsqueda espacial y penaliza las formas irregulares, genera superpíxeles altamente compactos y puros que coinciden con la morfología del terreno. Esto permite que, a pesar de contar con únicamente 5 muestras por clase, la GCN propague las etiquetas con alta certidumbre. SEEDS se posiciona como el segundo método más robusto, aunque no logra igualar la uniformidad estructural proporcionada por SLIC en este tipo de topología.

El conjunto *Pavia University* introduce la complejidad de un entorno urbano, donde coexisten estructuras fuertemente delimitadas como edificios, asfalto y vegetación. La Tabla 4.6 evidencia un comportamiento interesante en la etapa de clasificación: aunque el método ERS sigue siendo altamente efectivo para delinear infraestructuras y proteger las clases minoritarias (obteniendo la mejor exactitud promedio, $AA = 66.55\%$), es el algoritmo SLIC el que alcanza la mejor exactitud global ($OA = 66.85\%$) y el mejor coeficiente kappa ($\kappa = 0.5790$). Esto indica que, frente a la extrema escasez de etiquetas, la compacidad geométrica impuesta por SLIC proporciona una topología de grafo más robusta globalmente. En contraste, el modelo PCNN presenta el rendimiento más bajo en este conjunto ($OA = 58.93\%$). Este resultado demuestra empíricamente que, en escenarios urbanos con bordes artificiales abruptos y variaciones de iluminación (sombras), prescindir de restricciones espaciales rígidas y procesar el vector espectral original completo vuelve al modelo susceptible a la heterogeneidad intrínseca de los materiales, dificultando la separación de clases espacialmente contiguas durante la convolución en el grafo.

En el análisis del conjunto *LongKou*, se observa un rendimiento global consistentemente alto para todos los métodos, con exactitudes que se acercan o superan el 84 %. El algoritmo ERS destaca como el método óptimo, alcanzando la máxima exactitud global ($OA = 89.52\%$), exactitud promedio ($AA = 87.50\%$) y coeficiente Kappa ($\kappa = 0.8649$). Este dataset, al igual que Salinas, es de naturaleza agrícola, pero las intrincadas texturas de sus cultivos interactúan mucho mejor con la formulación basada en grafos y optimización de entropía de ERS, permitiendo segmentaciones puras que se traducen en una clasificación superior. Por su parte, LSC y SLIC se mantienen como alternativas altamente competitivas (con un OA de 86.65 % y 85.24 %, respectivamente), confirmando empíricamente que los enfoques que logran equilibrar la compacidad geométrica con una alta adherencia a los bordes son los más adecuados para este escenario espectral.

La Tabla 4.6 presenta los resultados para *HanChuan*, el cual resulta ser uno de los escenarios más complejos para la propagación semi-supervisada debido a la intrincada mezcla de cultivos y estructuras. En este entorno adverso, las métricas de clasificación sufren una caída generalizada, sin lograr superar la barrera del 60 % de exactitud global. Sin embargo, en este contexto de alta complejidad, la arquitectura bio-inspirada PCNN acoplada a la GCN logra destacar, obteniendo la mejor exactitud global ($OA = 59.96\%$) y el mayor coeficiente kappa ($\kappa = 0.5431$). Esto reafirma que, ante firmas espectrales muy correlacionadas y mezcladas, retener el vector espectral original completo provee características más discriminativas que la regularidad espacial forzada. Por otro lado, SLIC preserva mejor el equilibrio de clasificación entre las distintas clases, alcanzando la

máxima exactitud promedio ($AA = 58.40\%$). Esta fragmentación en el liderazgo de los resultados indica que, ante una heterogeneidad espacial tan severa y con un tamaño de muestra sumamente reducido, la red GCN encuentra un cuello de botella para propagar el aprendizaje, haciendo que ningún método logre una dominancia absoluta en todas las métricas de evaluación.

Finalmente, la Tabla 4.6 ilustra un hallazgo fundamental de la experimentación a través del dataset *HongHu*. Esta imagen hiperespectral es particularmente compleja por su gran cantidad de clases (22 cultivos distintos) con firmas espectrales altamente correlacionadas y sutiles variaciones. Aquí, el acoplamiento PCNN + GCN supera a todos los enfoques tradicionales, logrando el mejor desempeño con una exactitud global (OA) del 75.39 % y un coeficiente kappa (κ) de 0.6966. Este resultado valida empíricamente la principal ventaja de la PCNN: al prescindir del paso de reducción de dimensionalidad (PCA) en la etapa de delimitación espacial, el algoritmo evalúa el vector espectral original completo. Esta riqueza informativa permite a la PCNN agrupar regiones con una fidelidad superior, otorgando a la GCN los nodos puros necesarios para discriminar clases de vegetación críticas a partir de solo 5 muestras etiquetadas por clase. Por su parte, ERS compensa como el mejor método en exactitud promedio ($AA = 72.35\%$), equilibrando de manera efectiva la clasificación de las clases menos representadas.

4.5.3. Relación entre calidad de segmentación y desempeño de clasificación

Para cuantificar si una segmentación geoméricamente superior se traduce de manera confiable en una mejor clasificación, se calculó el coeficiente de correlación de Pearson entre cada métrica de segmentación y la exactitud global (OA), sobre los 30 pares dataset $\times$ método reportados en las Tabla 4.5 y Tabla 4.6. La Tabla 4.8 resume los resultados.

Tabla 4.8: Correlación de Pearson entre las métricas de segmentación y la exactitud global (OA), agregada sobre los 30 pares dataset $\times$ método (n=30)

Métrica	r	r^2 (% varianza explicada)	p
ASA	+0.515	26.6 %	0.0036
BR	+0.250	6.2 %	0.184
EV	+0.722	52.2 %	< 0.0001

Tres hallazgos se desprenden de este análisis. Primero, la exactitud de segmentación alcanzable (ASA) —la métrica más frecuentemente citada como indicador de calidad de segmentación— muestra una correlación positiva y estadísticamente significativa con OA, pero moderada: explica apenas el 26.6 % de la varianza observada en la exactitud de clasificación final. Segundo, la métrica EV —la métrica que específicamente cuantifica cuánta información espectral retiene cada superpíxel respecto a sus píxeles constituyentes—

es un predictor considerablemente más fuerte, explicando el 52.2 % de la varianza en OA. Esto es consistente con el mecanismo que motiva a PCNN-SAM: lo que determina el éxito de la propagación semi-supervisada no es solo la pureza geométrica del superpíxel, sino cuánta información espectral sobrevive dentro de él. Tercero, la métrica BR —la métrica en la que ERS domina de manera consistente (Tabla 4.5)— no muestra una correlación estadísticamente significativa con OA ($p = 0.184$). Esto explica cuantitativamente un patrón que ya se observaba cualitativamente en el análisis dataset por dataset: ERS obtiene sistemáticamente el mejor BR, pero solo traduce esa ventaja en la mejor OA en uno de los seis conjuntos (LongKou).

Al desagregar la relación ASA–OA por conjunto de datos (Tabla 4.9, $n=5$ por celda, por lo que estos valores son descriptivos y no deben interpretarse con la misma confianza estadística que la correlación agregada) se observa una variación considerable:

Tabla 4.9: Correlación de Pearson entre ASA y OA, desagregada por conjunto de datos ($n=5$ por dataset)

Conjunto de datos	r	p
Salinas	+0.823	0.087
LongKou	+0.774	0.125
HongHu	+0.444	0.453
HanChuan	+0.203	0.743
Indian Pines	+0.144	0.818
Pavia University	−0.162	0.795

En Salinas y LongKou —los dos conjuntos con parcelas más extensas y homogéneas— una mejor segmentación geométrica predice de manera confiable una mejor clasificación. En el extremo opuesto, Pavia University exhibe una correlación débil y de signo negativo: el método con mejor ASA (ERS, 98.05 %) no es el que produce la mejor OA (SLIC, 66.85 %), lo que es consistente con la naturaleza de bordes artificiales abruptos de esta escena, discutida anteriormente. Esta gradación —de correlación fuerte en escenas homogéneas a correlación nula o negativa en escenas urbanas o de alta correlación espectral— sugiere que la utilidad de las métricas de segmentación clásicas como predictor de desempeño depende de la naturaleza de la escena, y motiva por qué una arquitectura como PCNN-SAM, que optimiza explícitamente por fidelidad espectral en lugar de regularidad geométrica, puede resultar ventajosa precisamente en los regímenes donde ASA deja de ser un predictor confiable.

La Figura 4.21 muestra los 30 pares en ambos planos. El contraste entre los dos paneles resume el hallazgo metodológico de este capítulo: la nube de puntos frente a ASA está considerablemente más dispersa que la nube frente a EV, y las excepciones al patrón agregado —en particular los cinco puntos de Pavia University, que descienden en OA conforme aumenta ASA— resultan visibles directamente en el panel (a).

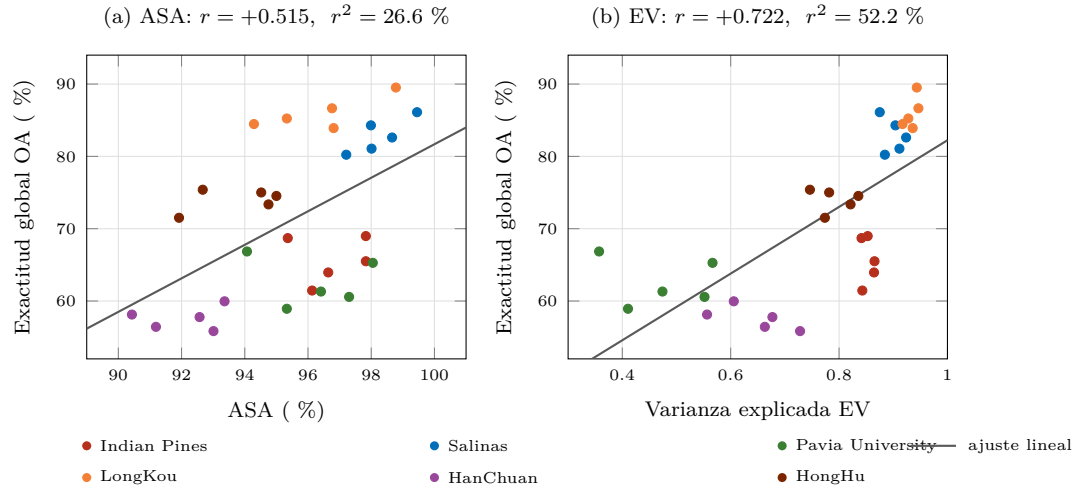


Figura 4.21: Relación entre las métricas de segmentación y la exactitud de clasificación sobre los 30 pares conjunto de datos $\times$ método. (a) frente a ASA; (b) frente a EV. La recta corresponde al ajuste lineal por mínimos cuadrados sobre los 30 puntos. La retención de información espectral que mide EV predice la OA casi el doble de bien que la pureza geométrica que mide ASA.

4.5.4. Análisis visual de los mapas de clasificación

Las Figuras 4.22 a 4.27 muestran los mapas de clasificación producidos por cada método sobre los seis conjuntos de datos, junto con la referencia de verdad de terreno. Su valor reside en localizar espacialmente las diferencias cuantitativas de la Tabla 4.6, y en particular en distinguir si una ventaja en OA proviene de las clases dominantes por superficie o de las clases minoritarias. El contraste entre OA y AA permite anticipar cuál de los dos casos aplica en cada conjunto.

En **Indian Pines** (Figura 4.22), la brecha entre SLIC (OA = 68.98 %) y LSC (OA = 61.45 %) es de 7.5 puntos, mientras que en AA la diferencia se reduce a 2.2 puntos (77.17 % frente a 75.02 %). Esta asimetría indica que el déficit de LSC no se concentra en las clases minoritarias sino en las clases dominantes por superficie, que son las que gobiernan la OA. SEEDS acompaña a SLIC en OA (68.70 %) pero es el método con menor AA del conjunto (69.10 %), de modo que su competitividad global se sostiene sobre las parcelas grandes.

En **Salinas** (Figura 4.23), todos los métodos superan el 80 % de OA y el rango completo abarca apenas 5.9 puntos, por lo que las diferencias entre paneles son sutiles. Se concentran en las fronteras entre las cuatro etapas de crecimiento de la lechuga romana, que son las cuatro clases espectralmente más próximas entre sí del conjunto.

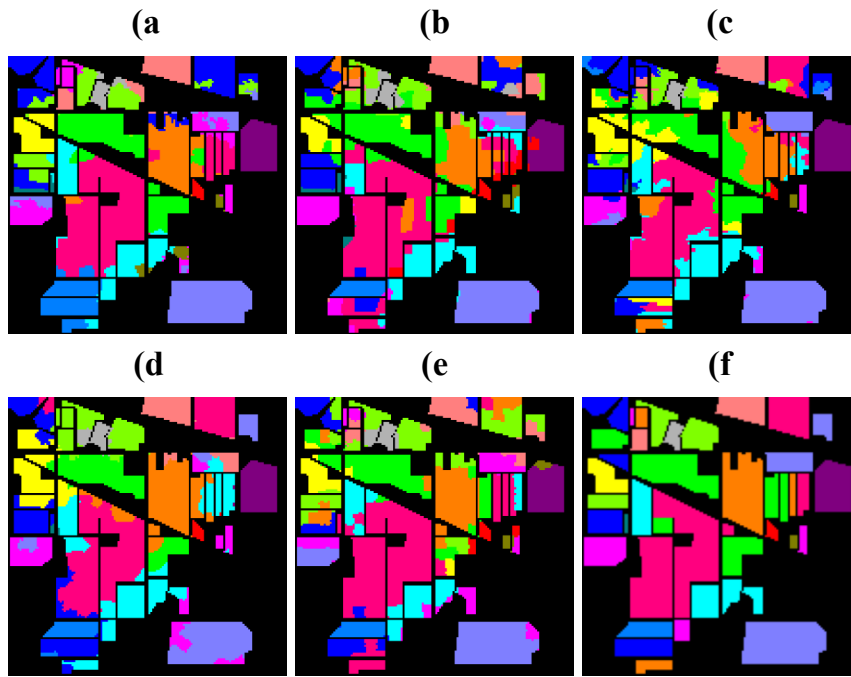


Figura 4.22: Mapas de clasificación sobre Indian Pines: a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM, f) referencia (*ground truth*).

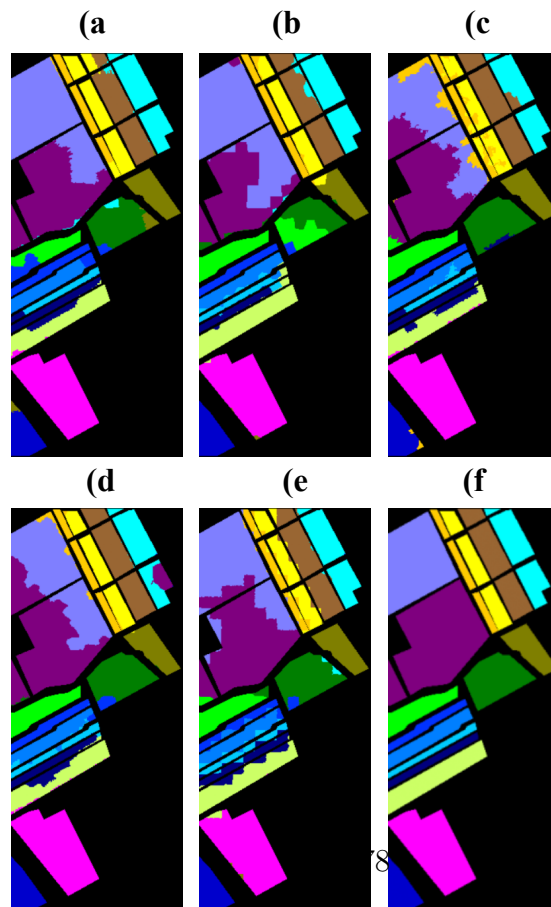


Figura 4.23: Mapas de clasificación sobre Salinas: a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM, f) referencia (*ground truth*).

En **Pavia University** (Figura 4.24), se produce la única inversión del estudio entre OA y AA: ERS obtiene la mejor AA (66.55 %) pero SLIC la mejor OA (66.85 %) y el mejor $\hat{\kappa}$ (0.5790). La lectura es que ERS acierta más en las clases de menor extensión —betún, láminas de metal pintadas—, mientras que SLIC clasifica mejor las clases dominantes por superficie —asfalto y prados—, que pesan más en la OA. PCNN-SAM registra aquí su peor resultado del estudio (58.93 %).

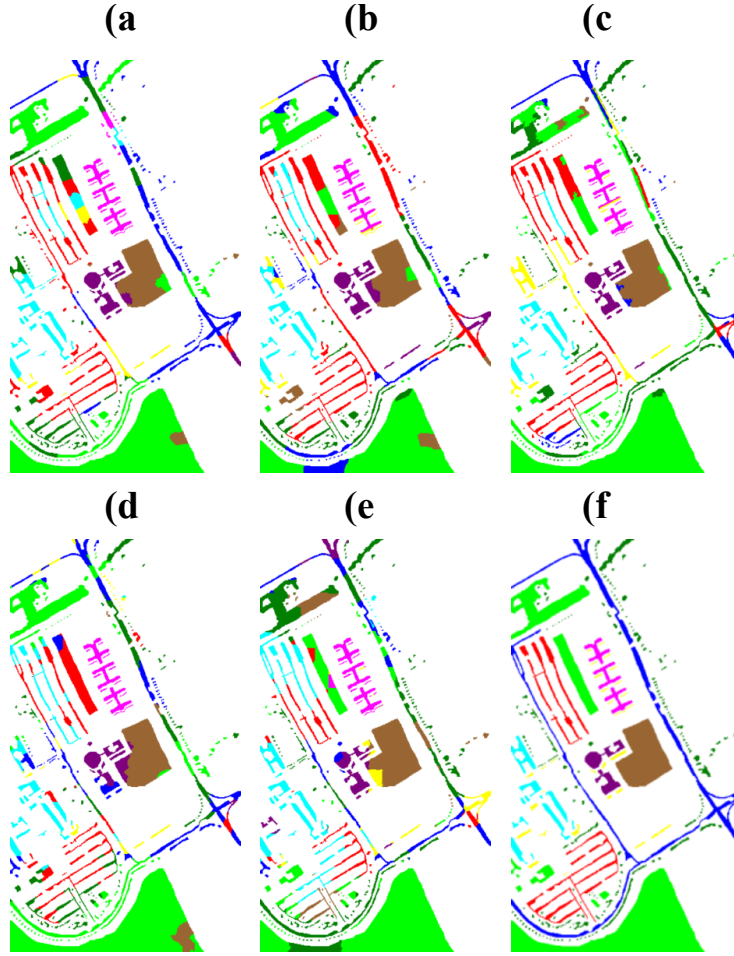


Figura 4.24: Mapas de clasificación sobre Pavia University: a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM, f) referencia (*ground truth*).

En **LongKou** (Figura 4.25), ERS lidera simultáneamente las tres métricas (OA = 89.52 %, AA = 87.50 %, $\hat{\kappa}$ = 0.8649), un caso de dominancia sin ambigüedad que se corresponde con su mejor adherencia a bordes (BR = 0.8039) sobre las fronteras entre los seis cultivos y el canal de agua central.

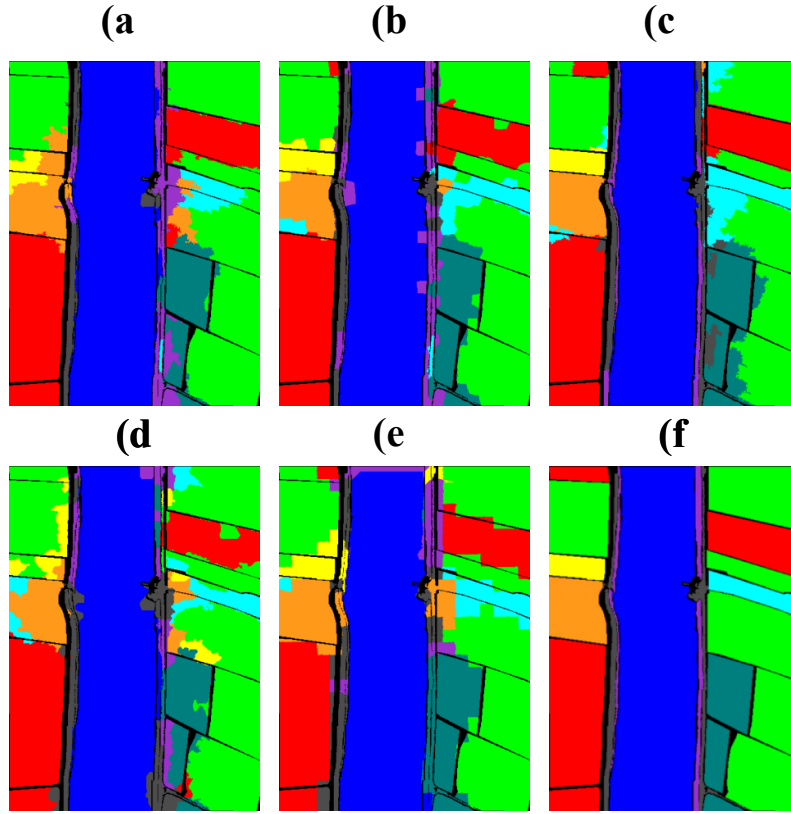


Figura 4.25: Mapas de clasificación sobre WHU-Hi-LongKou: a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM, f) referencia (*ground truth*).

En **HanChuan** (Figura 4.26), ningún método alcanza el 60 % de OA. PCNN-SAM obtiene la mejor OA (59.96 %) y el mejor $\hat{\kappa}$ (0.5431), pero la peor AA del grupo junto con SEEDS (54.49 %), mientras que SLIC lidera la AA (58.40 %). El patrón es el inverso al de Pavia University: la ventaja de PCNN-SAM proviene de las clases extensas, no de un tratamiento más equilibrado de las minoritarias.

En **HongHu** (Figura 4.27), se repite ese mismo patrón: PCNN-SAM alcanza la mejor OA (75.39 %) y el mejor $\hat{\kappa}$ (0.6966), mientras que ERS obtiene la mejor AA (72.35 %). Dado que este conjunto reúne veintidós clases, varias de ellas cultivares de una misma familia vegetal, el panel correspondiente a PCNN-SAM es el más relevante para sustentar el hallazgo de la Sección 4.5.3: las diferencias frente a los demás métodos deberían concentrarse en los grupos de clases espectralmente más correlacionados entre sí —las variedades de col (*bok choy*, *bok choy* pequeño, repollo) y de lechuga (lechuga, lechuga bajo film, lechuga romana).

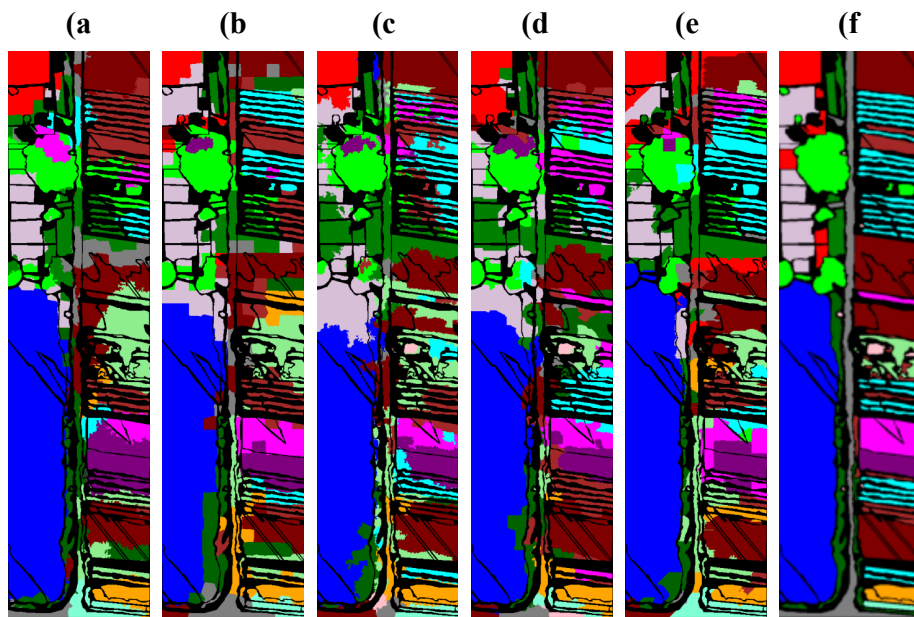


Figura 4.26: Mapas de clasificación sobre WHU-Hi-HanChuan: a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM, f) referencia (*ground truth*).

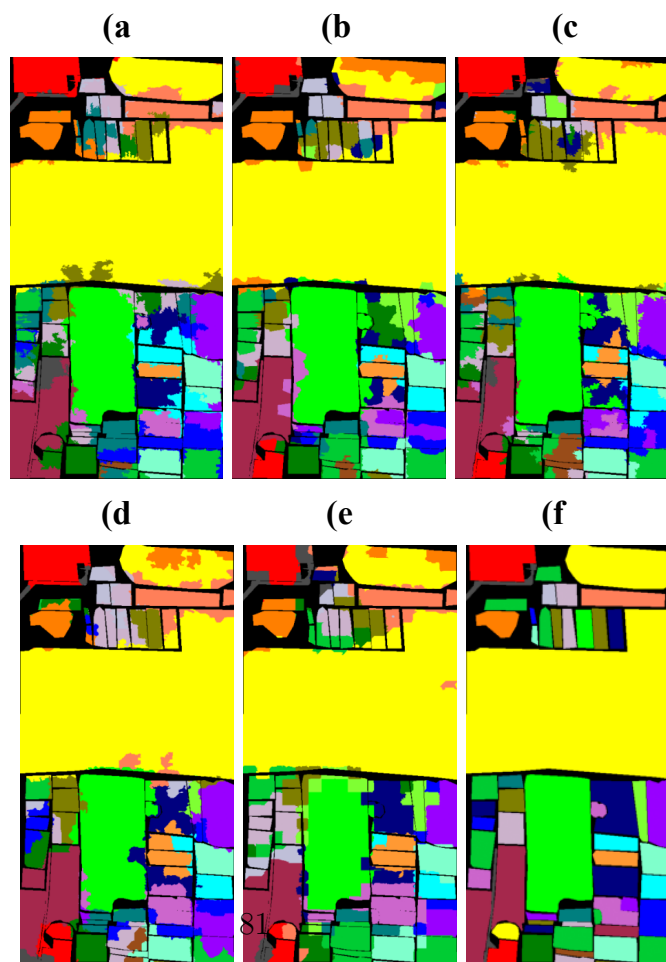


Figura 4.27: Mapas de clasificación sobre WHU-Hi-HongHu: a) SLIC, b) LSC, c) ERS, d) SEEDS, e) PCNN-SAM, f) referencia (*ground truth*).

4.6. Discusión general

Esta sección cierra el capítulo regresando a las cuatro preguntas planteadas al inicio del capítulo. La primera —qué escala de superpíxeles ofrece el mejor balance costo-precisión— quedó resuelta anteriormente ($K = 500$). Las tres restantes —qué estrategia de segmentación resulta más favorable y si esto depende de la naturaleza de la escena, si la preservación del espectro completo ofrece una ventaja medible, y si las diferencias observadas son estadísticamente significativas— se abordan a continuación.

Revisión de la hipótesis a la luz de la evidencia

La hipótesis de este trabajo planteaba que la arquitectura híbrida PCNN-GCN lograría una exactitud de clasificación significativamente superior a los métodos de segmentación del estado del arte, bajo condiciones de etiquetado extremadamente limitado. La evidencia reunida anteriormente obliga a matizar esta afirmación en dos direcciones.

Por un lado, la hipótesis no se sostiene en su lectura más fuerte, la de una superioridad general y consistente: la Tabla 4.7 muestra que PCNN-GCN obtiene el segundo lugar, no el primero, tanto en número de conjuntos de datos ganados (2 de 6, frente a los 3 de SLIC) como en OA promedio (70.75 %, por debajo del 72.80 % de SLIC y del 71.88 % de ERS). Si la hipótesis se interpreta como “PCNN-GCN superará a los métodos clásicos en cualquier escenario”, los resultados la refutan.

Por otro lado, la hipótesis sí se sostiene en una lectura más precisa y, en última instancia, más interesante: PCNN-GCN no gana en promedio, pero gana exactamente en los dos conjuntos de datos con mayor correlación espectral entre clases y mayor fragmentación espacial del estudio —HanChuan y HongHu—, y lo hace además con el margen relativo más amplio observado en todo el experimento sobre HanChuan (donde ningún otro método supera el 60 % de OA). Esto sugiere que la ventaja de PCNN-SAM no es general sino condicional: se manifiesta específicamente cuando la discriminación entre clases depende de matices espectrales sutiles que un paso previo de reducción de dimensionalidad tiende a suavizar. El mecanismo detrás de esta condicionalidad se desarrolla a continuación.

4.6.1. Interpretación integrada: qué método domina y bajo qué condiciones

Al integrar los resultados de segmentación, clasificación y la relación cuantitativa entre ambos, emerge un patrón consistente organizado en cuatro regímenes: La Figura 4.28 sitúa los seis conjuntos de datos en el plano definido por ambos tipos de dificultad, y la Tabla 4.10 resume el mecanismo dominante en cada régimen.

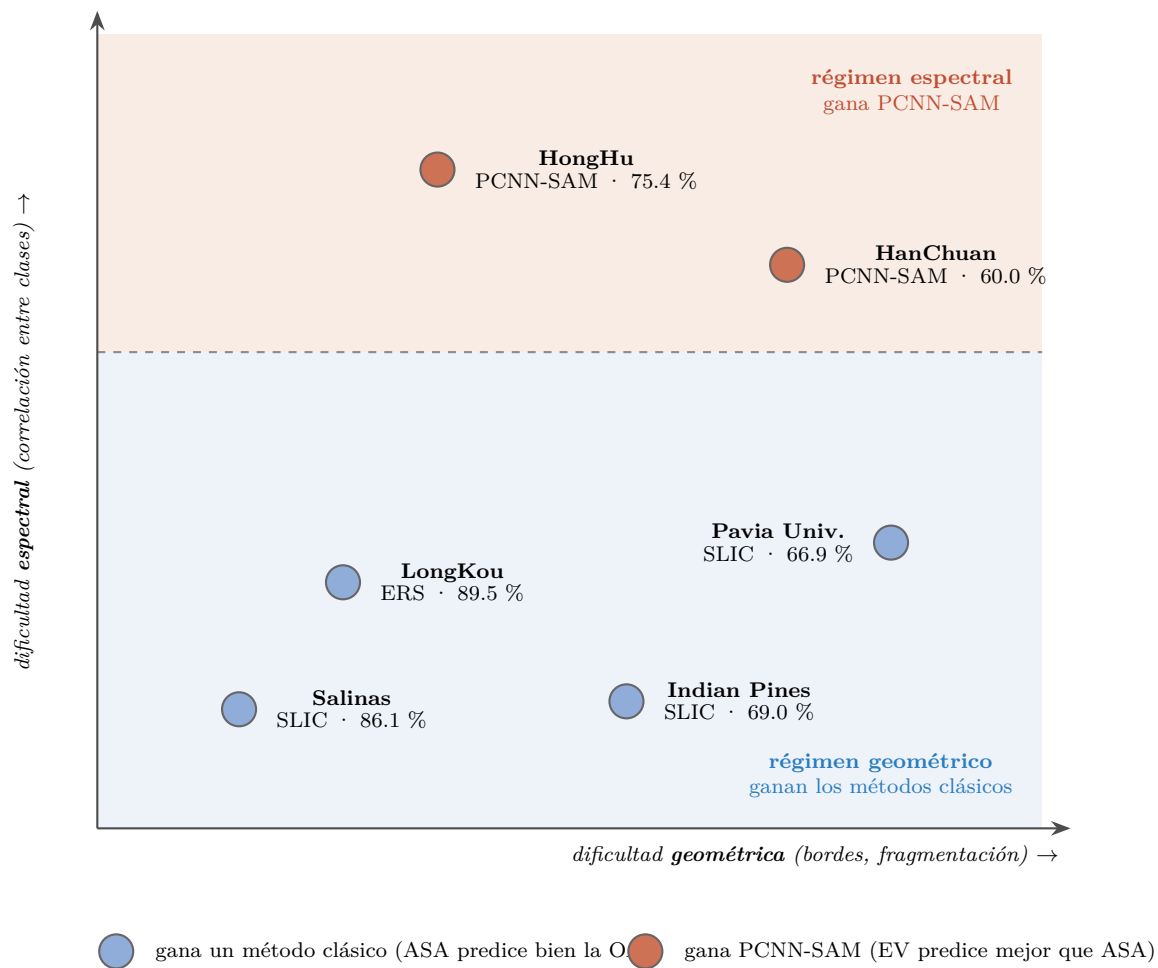


Figura 4.28: Los seis conjuntos de datos situados según el tipo de dificultad que plantean, con el método ganador y su exactitud global. La ventaja de PCNN-SAM no se distribuye al azar: se concentra en la banda superior, donde la discriminación entre clases depende de matices espectrales y no de la geometría de las fronteras.

Tabla 4.10: Los cuatro regímenes identificados: método ganador, correlación ASA–OA y mecanismo dominante

Régimen	Conjuntos	Método ganador	$r(\text{ASA}, \text{OA})$	Mecanismo dominante
Homogeneidad espacial	Salinas, LongKou	SLIC, ERS	+0.823, +0.774	La pureza geométrica predice de forma confiable la utilidad del superpíxel
Ruido de baja resolución	Indian Pines	SLIC	+0.144	La compacidad forzada actúa como filtro frente a los píxeles mixtos de borde
Bordes urbanos abruptos	Pavia University	SLIC	−0.162	Sin explicación mecánica a partir de las métricas disponibles
Alta correlación espectral	HanChuan, HongHu	PCNN-SAM	+0.203, +0.444	Lo determinante es la retención espectral (EV), no la regularidad geométrica

Homogeneidad espacial (Salinas, LongKou). En estos dos conjuntos, la correlación entre ASA y OA es la más fuerte observada en todo el estudio ($r = +0.823$ y $r = +0.774$ respectivamente, Tabla 4.9). Cuando las parcelas son extensas, contiguas y bien delimitadas, la pureza geométrica de un superpíxel predice de manera confiable su utilidad para la GCN, y los métodos que maximizan esa pureza —SLIC en Salinas, ERS en LongKou— dominan también la clasificación. PCNN-SAM no tiene ventaja que ofrecer aquí: la información espectral adicional que preserva es redundante cuando la geometría ya es informativa por sí sola.

Ruido de baja resolución espacial (Indian Pines). Aquí, la compacidad forzada de SLIC actúa como un filtro protector contra la mezcla espectral en los bordes de parcela característica de este conjunto. La correlación ASA-OA es débil ($r = +0.144$), lo que indica que factores distintos a la pureza geométrica dominan el resultado; PCNN-SAM, al carecer de esa restricción de compacidad, resulta más vulnerable al ruido del sensor y obtiene un desempeño intermedio.

Bordes urbanos abruptos (Pavia University). Este es el único conjunto donde la correlación ASA-OA es negativa ($r = -0.162$): el método con mejor segmentación (ERS, ASA = 98.05 %) no es el de mejor clasificación (SLIC, OA = 66.85 %). A diferencia de los tres regímenes anteriores, este resultado no tiene una explicación mecánica clara a partir de las métricas disponibles, y se reconoce explícitamente como una limitación de este estudio en lugar de forzar una narrativa post-hoc.

Alta correlación espectral entre clases (HanChuan, HongHu). Es en este régimen donde se concentra la ventaja de PCNN-SAM, y la relación cuantitativa presentada anteriormente

ofrece una explicación de por qué: mientras que ASA explica solo el 26.6 % de la varianza en OA a través de los 30 pares dataset-método, la métrica EV explica el 52.2 %. PCNN-SAM no fue diseñado para maximizar ASA ni BR; fue diseñado para maximizar precisamente lo que EV mide. Cuando la escena exige distinguir variedades espectralmente casi idénticas —como los cultivos de col y lechuga en HongHu—, esa es la propiedad que determina el éxito de la propagación semi-supervisada, no la regularidad geométrica del superpíxel.

Esta lectura por régimen, más que la comparación agregada de la Tabla 4.7, es la que mejor responde a la segunda pregunta planteada al inicio del capítulo: no existe una estrategia de segmentación universalmente superior, y la elección óptima depende de si la dificultad de la escena es de naturaleza geométrica (bordes, compacidad) o espectral (correlación entre clases). Las conclusiones generales de esta tesis, y las líneas de trabajo futuro que se desprenden de este patrón, se desarrollan en el Capítulo 5.

Capítulo 5

Conclusiones y trabajo futuro

5.1. Conclusiones

Este trabajo desarrolló y evaluó PCNN-SAM, un algoritmo de segmentación bio-inspirado que incorpora la métrica SAM en la dinámica de acoplamiento de una red neuronal de pulsos, y lo acopló a un clasificador semi-supervisado GCN. La evaluación se realizó bajo un régimen de escasez extrema de etiquetas —cinco muestras por clase— sobre seis conjuntos de datos hiperespectrales seleccionados deliberadamente por cubrir un espectro amplio de condiciones espaciales y espectrales, y se contrastó frente a cuatro algoritmos de segmentación del estado del arte: SLIC, LSC, ERS y SEEDS. El diseño experimental comprendió 30 pares conjunto de datos $\times$ método, cada uno ejecutado diez veces de manera independiente, para un total de 300 ejecuciones de entrenamiento y evaluación.

De la evidencia reunida en el Capítulo 4 se desprenden siete conclusiones, ordenadas para responder, en secuencia, a cada uno de los siete objetivos específicos planteados en la Sección 1.3.2: **C₁** a **C₃** corresponden a los tres objetivos de diseño del algoritmo de segmentación (objetivos 1–3); **C₄** corresponde a la implementación del clasificador bajo escasez de etiquetas (objetivo 4); **C₅** corresponde a la selección de la escala de superpíxeles (objetivo 5); **C₆** corresponde a la comparación experimental frente a los métodos de referencia (objetivo 6); y **C₇** corresponde a la cuantificación de la relación estadística entre segmentación y clasificación (objetivo 7).

C₁ *Es posible condicionar la propagación de una PCNN a la similitud espectral sin recurrir a una reducción de dimensionalidad previa.* A diferencia de los cuatro métodos de referencia, que comprimen necesariamente el cubo hiperespectral a un espacio de tres o diez componentes antes de segmentar (Sección 2.5), la compuerta espectral de PCNN-SAM (Ecuación (4.4)) evalúa la afinidad entre vecinos directamente sobre el vector espectral completo de B bandas, sin proyección ni selección de bandas previa. Que esta preservación espectral se traduzca en una ventaja medible de clasificación es, como muestran las conclusiones siguientes,

condicional al régimen espectral de la escena.

- C₂** *El mecanismo de competencia entre regiones y el refinamiento morfológico garantizan, por construcción, una partición completa y conexa.* La resolución de píxeles en disputa mediante la energía híbrida E_k (Ecuación (4.6)) y el reetiquetado final de componentes conexas del mapa de disparo T (Sección 4.2.4) eliminan, respectivamente, los píxeles sin asignar y los fragmentos inconexos. Esta garantía es la que permite calcular sin excepción las métricas ASA, BR y EV reportadas en la Tabla 4.5 sobre los seis conjuntos de datos: ninguna de ellas admite como entrada una segmentación parcial o desconectada.
- C₃** *Caracterizar cada nodo por la firma espectral media de su superpíxel es suficiente para sostener la propagación semi-supervisada.* El grafo construido a partir de las regiones segmentadas —con matriz de adyacencia espacial (Ecuación (4.7)) y vector de características promediado sobre $d = 10$ componentes principales por nodo (Ecuación (4.8))— es la única representación de entrada que recibe la GCN. Que esta representación, deliberadamente simple, haya bastado para alcanzar los niveles de exactitud reportados en el Capítulo 4 bajo un régimen de apenas cinco etiquetas por clase confirma que es la topología heredada de la segmentación la que determina el resultado final, en línea con **C₄**.
- C₄** *La topología del grafo inicial determina el desempeño de la clasificación semi-supervisada.* Bajo cinco etiquetas por clase, la GCN carece de evidencia suficiente para corregir los errores que hereda de la etapa de segmentación. En consecuencia, la elección del algoritmo que construye el grafo no es un detalle de preprocesamiento, sino el factor que gobierna el resultado final. Esta conclusión justifica el diseño experimental adoptado, que mantiene fijo el clasificador y hace variar únicamente el algoritmo de segmentación.
- C₅** *Existe una escala de superpíxel óptima, y no es la más fina posible.* El barrido sobre $K \in \{100, 250, 500, 1000\}$ (Sección 4.4) muestra que la exactitud global asciende de manera consistente hasta $K = 500$ (66.68 %) y retrocede ligeramente en $K = 1000$ (66.57 %), mientras que el costo computacional de la segmentación crece de forma monótona en todo el rango, casi duplicándose entre los extremos evaluados. Sobresegmentar no es, por tanto, una estrategia dominante: más allá de cierto punto, fragmentar en exceso introduce ruido topológico en el grafo sin compensación en exactitud, lo cual justifica empíricamente la escala $K = 500$ empleada en el resto del análisis comparativo.
- C₆** *No existe un algoritmo de segmentación universalmente superior: el método que gana depende del régimen de la escena.* Los métodos clásicos dominan en los conjuntos con homogeneidad espacial o bordes estructurales bien definidos (Salinas, Pavia University e Indian Pines), y ganan tres de los seis conjuntos tanto en

exactitud global como en el coeficiente $\hat{\kappa}$ (Sección 4.5.2); la regularidad geométrica que estos algoritmos imponen sobre el superpíxel resulta, en esos escenarios, más valiosa para la GCN que la fidelidad espectral adicional que aporta PCNN-SAM. PCNN-SAM, en cambio, gana en HanChuan y HongHu con el margen más amplio observado en todo el experimento; que la ventaja aparezca justamente en los dos conjuntos que el mecanismo del método predice, y no distribuida al azar entre los seis, es la evidencia que sostiene esta conclusión.

- C₇** *La retención de información espectral predice el desempeño mejor que la pureza geométrica.* Sobre los treinta pares conjunto de datos $\times$ método, la métrica EV predice la exactitud global casi el doble de bien que la ASA: 52.2 % frente a 26.6 % de EV (Sección 4.5.3). La correlación EV–OA es estadísticamente significativa ($r = 0.722$, $n = 30$, $p < 0.001$), mientras que la correlación BR–OA no lo es ($p = 0.184$). Esta conclusión, que constituye el hallazgo principal del trabajo, confirma con evidencia estadística —y no solo por diseño— que PCNN-SAM gana precisamente en el tipo de escenario para el que fue concebido: aquel donde discriminar entre clases exige preservar matices espectrales que una reducción de dimensionalidad previa tiende a suavizar.

Contraste de las predicciones

La hipótesis formulada en la Sección 1.3.1 se descompuso allí en dos predicciones verificables por separado, precisamente para que la evidencia pudiera resolverlas de manera independiente. El experimento las resuelve en sentidos opuestos, y ese contraste es el resultado central de la tesis.

- P₁** *Superioridad general: refutada.* PCNN-GCN ocupa el segundo lugar, no el primero, tanto en número de conjuntos ganados (dos de seis, frente a los tres de SLIC) como en exactitud global promedio: 70.75 % frente al 72.80 % de SLIC y al 71.88 % de ERS. La predicción de dominancia en cualquier escenario queda descartada por la evidencia.
- P₂** *Ventaja condicional: confirmada.* La ventaja de PCNN-SAM se concentra en las escenas cuya dificultad es de naturaleza espectral y no geométrica, tal como predecía **P₂**, y el análisis de correlación de la conclusión **C₇** aporta el mecanismo que explica por qué, con respaldo estadístico.

Refutar **P₁** y confirmar **P₂** es un resultado más informativo que haber confirmado ambas. Una superioridad general habría dejado abierta la pregunta de por qué el método funciona; la ventaja condicional, en cambio, viene acompañada de la explicación cuantitativa de **C₇** y delimita con precisión cuándo conviene emplear cada estrategia. Esta distinción —*ganador por régimen* y no *ganador promedio*— constituye la contribución central de esta tesis: la

elección óptima de un algoritmo de segmentación para la clasificación semi-supervisada de HSI depende de si la dificultad de la escena es de naturaleza geométrica o espectral, y no existe una respuesta única válida para todas las escenas.

En conjunto, C_1 a C_7 dan respuesta al objetivo general planteado en la Sección 1.3.2: la arquitectura híbrida PCNN-SAM + GCN se diseñó, implementó y evaluó de manera sistemática, y la evidencia reunida permite determinar con precisión bajo qué condiciones espaciales y espectrales la preservación del vector espectral completo durante la segmentación mejora la clasificación de imágenes hiperespectrales con escasez de datos etiquetados: en escenas donde la dificultad de discriminación entre clases es de naturaleza espectral, y no meramente geométrica.

5.2. Discusión

Aportación metodológica

Más allá de la comparación entre métodos, este trabajo aporta una observación de alcance más general para la evaluación de segmentaciones destinadas a un clasificador basado en grafos: las métricas geométricas de uso habitual —ASA y BR, ambas centradas en qué tan bien un superpíxel respeta fronteras y formas— no son un indicador suficiente de la utilidad real de una segmentación para la propagación semi-supervisada. La diferencia entre el 26.6 % de ASA y el 52.2 % de EV (Sección 4.5.3) muestra que una métrica centrada en la fidelidad espectral captura algo que la pureza geométrica no captura por sí sola, y el caso de BR es aún más claro: es la métrica en la que ERS domina de manera sistemática, y sin embargo su correlación con la exactitud final no es estadísticamente distinguible de cero.

Esta observación tiene una implicación práctica directa: al diseñar o seleccionar un algoritmo de segmentación para HSI destinado a un clasificador de grafos conviene priorizar explícitamente la retención espectral, y no únicamente la regularidad geométrica, sobre todo en escenas con clases espectralmente correlacionadas. La recomendación es utilizable de inmediato, ya que EV se calcula sobre la segmentación sin necesidad de entrenar el clasificador.

Costo computacional

Esta ventaja no está libre de costo. PCNN-SAM opera sobre el vector espectral completo en cada comparación de vecindad, en contraste con los métodos clásicos, que operan sobre una representación previamente reducida a unas pocas dimensiones. El comportamiento observado durante la experimentación es consistente con un tiempo de segmentación mayor que el de sus contrapartes clásicas, lo cual resulta esperable dado el volumen de operaciones involucrado. Se trata de un costo que se paga una sola vez, en la etapa de preprocesamiento, y que no afecta al tiempo de entrenamiento ni de inferencia del clasificador.

Alcance y delimitación del estudio

Las conclusiones anteriores se sostienen dentro del marco experimental definido en el Capítulo 4. Las decisiones que delimitan ese marco fueron deliberadas y se justifican a continuación, junto con la razón por la que ninguna de ellas compromete las conclusiones obtenidas.

- **Un único clasificador, mantenido fijo.** Se adoptó la formulación de GCN más simple —la de dos capas de Kipf y Welling— de manera deliberada. Emplear una arquitectura con mecanismos de atención o propagación multiescala habría introducido factores de confusión que impedirían atribuir las diferencias observadas a la etapa de segmentación, que es la variable de interés de esta investigación. El control de esa variable es una condición necesaria para que las conclusiones basadas en la comparación empírica entre métodos de segmentación —**C₄**, **C₆** y **C₇**— sean interpretables.
- **Una única escala de superpíxeles.** El valor $K = 500$ no se fijó de manera arbitraria, sino que resultó de un barrido sobre cuatro escalas ($K \in \{100, 250, 500, 1000\}$), evaluado sobre los seis conjuntos de datos y las dos arquitecturas de GCN consideradas (Sección 4.4). Se seleccionó por ofrecer la mayor exactitud del barrido sin incurrir en el costo adicional de la escala más fina. Mantener la escala constante entre métodos es, además, condición de equidad en la comparación.
- **Comparación por régimen y no prueba de superioridad puntual.** El estudio se diseñó para responder en qué condiciones domina cada estrategia, no para establecer la superioridad de un método sobre otro en un conjunto de datos aislado. Por ello, el respaldo estadístico se concentra en el hallazgo principal —la correlación EV–OA sobre los treinta pares, significativa con $p < 0.001$ — y las comparaciones por conjunto de datos se reportan como promedios sobre diez ejecuciones independientes, con carácter descriptivo. Esta elección es consistente con la conclusión **C₆**: dado que ningún método domina universalmente, una prueba de superioridad global no sería la herramienta adecuada para la pregunta planteada.
- **El comportamiento de Pavia University queda abierto.** Es el único conjunto donde la correlación entre ASA y OA resulta negativa, y ese comportamiento no se explica a partir de las métricas empleadas en este trabajo. Se reporta como observación abierta en lugar de construir una interpretación *post hoc* no sustentada por la evidencia disponible, y se retoma como línea de trabajo futuro.

5.3. Trabajo futuro

De la experiencia de desarrollar y evaluar PCNN-SAM se desprenden cuatro líneas de trabajo futuro.

- TF₁** *Reducir el costo del acoplamiento espectral.* Se conjetura que el componente responsable del mayor tiempo de segmentación es precisamente SAM: a diferencia de una distancia euclidiana sobre tres o diez dimensiones, evaluar SAM entre vecinos exige calcular un producto punto y un arcocoseno sobre el vector espectral completo, en cada iteración y para cada par de píxeles vecinos activos. Un desglose formal del tiempo de ejecución por componente permitiría confirmarlo y, a partir de ahí, explorar mecanismos de acoplamiento espectral alternativos que conserven la propiedad que hace valiosa a SAM —la invariancia a la magnitud, que permite comparar firmas espectrales con independencia de la iluminación— a un costo inferior: por ejemplo, aproximaciones precalculadas de la similitud espectral o métricas de menor complejidad que preserven la misma noción de afinidad angular.
- TF₂** *Generar pseudoetiquetas a partir del tiempo de disparo.* Existe una similitud conceptual entre la dinámica de sincronización de la PCNN y los algoritmos de propagación de etiquetas (*label propagation*): en ambos casos la información se propaga de un punto a sus vecinos en función de su afinidad, y el orden de esa propagación —el tiempo de disparo T_{ij} de la Sección 4.2.4— codifica implícitamente qué tan fuertemente conectado está un píxel a su vecindario. Esto sugiere emplear la dinámica conjunta de PCNN y SAM no solo para segmentar, sino para asignar a cada píxel no etiquetado una etiqueta candidata y un grado de confianza derivados de su cercanía de sincronización con superpíxeles ya etiquetados. De resultar viable, este mecanismo permitiría enriquecer el conjunto $\mathcal{V}_L$ más allá de las cinco muestras por clase evaluadas aquí, mitigando aún más la escasez de etiquetas que motiva este trabajo desde su planteamiento (Sección 1.3).
- TF₃** *Verificar la generalidad del hallazgo sobre EV.* La conclusión **C₇** se obtuvo con una GCN de dos capas deliberadamente simple, elegida para aislar el efecto de la segmentación. Comprobar si la varianza explicada sigue siendo mejor predictor que la exactitud de segmentación alcanzable bajo arquitecturas con mecanismos de atención, propagación multiescala o refinamiento posterior determinaría si se trata de una propiedad general del aprendizaje semi-supervisado sobre superpíxeles o de un efecto propio de la formulación empleada en esta tesis.
- TF₄** *Caracterizar el régimen de bordes urbanos.* El comportamiento de Pavia University, donde la mejor segmentación geométrica no produce la mejor clasificación, sugiere que en escenas dominadas por fronteras artificiales abruptas interviene un factor que ni ASA ni EV capturan. Identificar una métrica que lo describa completaría el mapa de regímenes propuesto en la Sección 4.6.1.

Bibliografía

- [1] Aman Srivastava and Shubham Jain. *Remote Sensing for Environment Assessment: Multispectral, Hyperspectral, and Thermal Imaging Applications*, pages 1–31. Springer Nature Singapore, Singapore, 2025.
- [2] Xiuping Jia. *Field Guide to Hyperspectral/Multispectral Image Processing*, volume FG52 of *SPIE Field Guides*. SPIE Press, Bellingham, Washington, 2022.
- [3] Vinod Kumar, Ravi Shankar Singh, Medara Rambabu, and Yaman Dua. Deep learning for hyperspectral image classification: A survey. *Computer Science Review*, 53:100658, 2024.
- [4] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2274–2282, 2012.
- [5] Jiansheng Chen, Zhengqin Li, and Bo Huang. Linear spectral clustering superpixel. *IEEE Transactions on Image Processing*, 26(7):3317–3330, 2017.
- [6] Ming-Yu Liu, Oncel Tuzel, Srikumar Ramalingam, and Rama Chellappa. Entropy rate superpixel segmentation. In *CVPR 2011*, pages 2097–2104, 2011.
- [7] Michael Van den Bergh, Xavier Boix, Gemma Roig, and Luc Van Gool. Seeds: Superpixels extracted via energy-driven sampling, 2013.
- [8] Chao Yao, Junrui Gu, Zehua Guo, Miao Ma, Qingrui Guo, and Gong Cheng. Collaborative superpixelwise pca for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18:2589–2601, 2025.
- [9] Cheng-Hsuan Li, Hui-Shan Chu, Bor-Chen Kuo, and Chin-Teng Lin. Hyperspectral image classification using spectral and spatial information based linear discriminant analysis. In *2011 IEEE International Geoscience and Remote Sensing Symposium*, pages 1716–1719, 2011.

- [10] Reaya Grewal, Singara Singh Kasana, and Geeta Kasana. Hyperspectral image segmentation: a comprehensive survey. *Multimedia Tools and Applications*, 82(14):20819–20872, 2023.
- [11] Nanlan Wang, Xiaoyong Zeng, Yanjun Duan, Bin Deng, Yan Mo, Zhuojun Xie, and Puhong Duan. Multi-scale superpixel-guided structural profiles for hyperspectral image classification. *Sensors*, 22(21):8502, 2022.
- [12] Cheng Chen, Nian Wang, Shengming Wang, Jiping Cao, Tao Wang, Zhigao Cui, and Yanzhao Su. Spectral–spatial superpixel bi-stochastic graph learning for large-scale and high-dimensional hyperspectral image clustering. *Remote Sensing*, 17(23):3799, 2025.
- [13] Nurul Rafi and Pablo Rivas. A review of pulse-coupled neural network applications in computer vision and image processing, 2024.
- [14] Puhong Duan, Xudong Kang, Shutao Li, and Pedram Ghamisi. Multichannel pulse-coupled neural network-based hyperspectral image visualization. *IEEE Transactions on Geoscience and Remote Sensing*, 58(4):2444–2456, 2020.
- [15] Xinyu Xu, Xiaojun Li, Yikun Li, Lu Kang, and Junfei Ge. A novel adaptively optimized pcnn model for hyperspectral image sharpening. *Remote Sensing*, 15(17), 2023.
- [16] Bobo Xi, Jiaojiao Li, Yunsong Li, Rui Song, Yuchao Xiao, Qian Du, and Jocelyn Chanussot. Semisupervised cross-scale graph prototypical network for hyperspectral image classification. *IEEE Transactions on Neural Networks and Learning Systems*, 34(11):9337–9351, 2023.
- [17] Bobo Xi, Jiaojiao Li, Yunsong Li, and Qian Du. Semi-supervised graph prototypical networks for hyperspectral image classification. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 2851–2854, 2021.
- [18] Peng Chen, Xin Li, Yong Peng, Xing Fan, and Qiang Li. WSSGCN: Hyperspectral forest image classification via watershed superpixel segmentation and sparse graph convolutional networks. *Forests*, 16(5):827, 2025.
- [19] Maryam Imani and Daniele Cerra. Triple graph convolutional network for hyperspectral image feature fusion and classification. *Remote Sensing*, 17(9):1623, 2025.
- [20] Hao Shi, Zhen Luo, Yan Ma, Gang Zhu, and Xin Dai. SSGTN: Spectral–spatial graph transformer network for hyperspectral image classification. *Remote Sensing*, 18(2):199, 2026.

- [21] William Michael Laprade, Pawel Pieta, Svetlana Kutuzova, Jesper Cairo Westergaard, Mads Nielsen, Svend Christensen, and Anders Bjorholm Dahl. Hyperleaf2024 – a hyperspectral imaging dataset for classification and regression of wheat leaves. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1234–1243, 2024.
- [22] M.E. Paoletti, J.M. Haut, J. Plaza, and A. Plaza. Deep learning classifiers for hyperspectral imaging: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158:279–317, 2019.
- [23] Wolfgang Kurz, Kun Wang, Furkan Bektas, Changyan Zhu, Emre Kariper, Xingchen Dong, Michael Kurz, Martin Jakobi, Danny Baranes, and Alexander W. Koch. Dimensionality reduction in hyperspectral imaging using standard deviation-based band selection for efficient classification. *Scientific Reports*, 15:34478, 2025.
- [24] Philip Sellars, Angelica I. Aviles-Rivero, and Carola-Bibiane Schönlieb. Superpixel contracted graph-based learning for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 58(6):4180–4193, 2020.
- [25] Zhen Zhang, Lehao Huang, Bo-Hui Tang, Qingwang Wang, Zhongxi Ge, and Linhuan Jiang. Non-euclidean spectral-spatial feature mining network with gated gcnn for hyperspectral image classification. *Expert Systems with Applications*, 272:126811, 2025.
- [26] Huaze Xie, Xiaobin Zhao, Junjun Yin, Jiangtao Peng, Qingwang Wang, and Gemine Vivone. Cross-architecture contrastive learning for few-shot hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 64:1–15, 2026.
- [27] Chen Yang, Zizhuo Liu, Renchu Guan, and Haishi Zhao. A semi-supervised multi-scale convolutional neural network for hyperspectral image classification with limited labeled samples. *Remote Sensing*, 17(19):3273, 2025.
- [28] Yao Li, Liyi Zhang, Long Chen, et al. Superpixel guided spectral-spatial feature extraction and weighted feature fusion for hyperspectral image classification with limited training samples. *Scientific Reports*, 15:3473, 2025.
- [29] Thomas M. Lillesand, Ralph W. Kiefer, and Jonathan W. Chipman. *Remote Sensing and Image Interpretation*. John Wiley & Sons, Hoboken, NJ, 7 edition, 2015.
- [30] Billy G. Ram, Peter Oduor, C. Igathinathane, Kirk Howatt, and Xin Sun. A systematic review of hyperspectral imaging in precision agriculture: Analysis of its current state and future prospects. *Computers and Electronics in Agriculture*, 222:109037, 2024.

- [31] Bing Lu, Phuong D. Dao, Jiangui Liu, Yuhong He, and Jiali Shang. Recent advances of hyperspectral imaging technology and applications in agriculture. *Remote Sensing*, 12(16), 2020.
- [32] Ivan Podlesnykh, Michael Kovalev, and Pavel Platonov. Towards the future of ubiquitous hyperspectral imaging: Innovations in sensor configurations and cost reduction for widespread applicability. *Technologies*, 12(11), 2024.
- [33] Zhen Zhang, Lehao Huang, Qingwang Wang, Linhuan Jiang, Yemao Qi, Shunyuan Wang, Tao Shen, Bo-Hui Tang, and Yanfeng Gu. Uav hyperspectral remote sensing image classification: A systematic review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18:3099–3124, 2025.
- [34] Andrea Caroppo, Giovanni Diraco, and Alessandro Leone. A systematic review of available multispectral uav image datasets for precision agriculture applications. *Remote Sensing*, 18(4), 2026.
- [35] Lili Zhao, Xuncheng Fan, and Tao Hong. Urban heat island effect: Remote sensing monitoring and assessment—methods, applications, and future directions. *Atmosphere*, 16(7), 2025.
- [36] Magdalena Smigaj, Avinash Agarwal, Harm Bartholomeus, Mathieu Decuyper, Ahmed Elsherif, Arjen de Jonge, and Lammert Kooistra. Thermal infrared remote sensing of stress responses in forest environments: a review of developments, challenges, and opportunities. *Current Forestry Reports*, 10(1):56–76, February 2024.
- [37] Muhammad Ahmad, Salvatore Distefano, Adil Mehmood Khan, Manuel Mazzara, Chenyu Li, Hao Li, Jagannath Aryal, Yao Ding, Gemine Vivone, and Danfeng Hong. A comprehensive survey for hyperspectral image classification: The evolution from conventional to transformers and mamba models. *Neurocomputing*, 644:130428, 2025.
- [38] Sujata Chakravarty, Bijay Kumar Paikaray, Rutuparnna Mishra, and Satyabrata Dash. Hyperspectral image classification using spectral angle mapper. In *2021 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*, pages 87–90, 2021.
- [39] Lin Wei, Huiyun Ma, Yuping Yin, and Chao Geng. Kmeans-cm algorithm with spectral angle mapper for hyperspectral image classification. *IEEE Access*, 11:26566–26576, 2023.
- [40] Chein-I Chang. An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis. *IEEE Transactions on Information Theory*, 46(5):1927–1932, 2000.
- [41] R.C. Gonzalez and R.E. Woods. *Digital Image Processing*. Pearson, 2018.

- [42] Isabela Borlido Barcelos, Felipe De Castro Belém, Leonardo De Melo João, Zenilton K. G. Do Patrocínio, Alexandre Xavier Falcão, and Silvio Jamil Ferzoli Guimarães. A comprehensive review and new taxonomy on superpixel segmentation. *ACM Computing Surveys*, 56(8):1–39, April 2024.
- [43] R. Eckhorn, H. J. Reitboeck, M. Arndt, and P. Dicke. Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex. *Neural Computation*, 2(3):293–307, 1990.
- [44] Thomas Lindblad and Jason M. Kinser. *Image Processing Using Pulse-Coupled Neural Networks*. Springer-Verlag Berlin Heidelberg, 2013.
- [45] Zhaoyang Hou, Kaiyun Lv, Xunqiang Gong, and Yuting Wan. A remote sensing image fusion method combining low-level visual features and parameter-adaptive dual-channel pulse-coupled neural network. *Remote Sensing*, 15(2), 2023.
- [46] Rong Ma, Zhen Zhang, Yide Ma, Xiping Hu, Edith C.H. Ngai, and Victor C.M. Leung. An improved pulse coupled neural networks model for semantic iot. *Digital Communications and Networks*, 10(3):557–567, 2024.
- [47] Alastair P. Moore, Simon J. D. Prince, Jonathan Warrell, Umar Mohammed, and Graham Jones. Superpixel lattices. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.
- [48] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, New York, 2006.
- [49] Sameer Ranjan, Deepak Ranjan Nayak, Kallepalli Satish Kumar, Ratnakar Dash, and Banshidhar Majhi. Hyperspectral image classification: A k-means clustering based approach. In *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)*, pages 1–7, 2017.
- [50] Wei Zhu, Victoria Chayes, Alexandre Tiard, Stephanie Sanchez, Devin Dahlberg, Andrea L. Bertozzi, Stanley Osher, Dominique Zosso, and Da Kuang. Unsupervised classification in hyperspectral imagery with nonlocal total variation and primal-dual hybrid gradient algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 55(5):2786–2798, 2017.
- [51] Shaohui Mei, Jingyu Ji, Yunhao Geng, Zhi Zhang, Xu Li, and Qian Du. Unsupervised spatial-spectral feature learning by 3d convolutional autoencoder for hyperspectral classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):6808–6820, 2019.

- [52] Zeyu Cao, Xiaorun Li, Yueming Feng, Shuhan Chen, Chaoqun Xia, and Liaoying Zhao. Contrastnet: Unsupervised feature learning by autoencoder and prototypical contrastive learning for hyperspectral imagery classification. *Neurocomputing*, 460:71–83, 2021.
- [53] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [54] Konstantinos Makantasis, Konstantinos Karantzas, Anastasios Doulamis, and Nikolaos Doulamis. Deep supervised learning for hyperspectral data classification through convolutional neural networks. In *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 4959–4962, 2015.
- [55] Yushi Chen, Hanlu Jiang, Chunyang Li, Xiuping Jia, and Pedram Ghamisi. Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10):6232–6251, 2016.
- [56] Zilong Zhong, Jonathan Li, Zhiming Luo, and Michael Chapman. Spectral-spatial residual network for hyperspectral image classification: A 3-d deep learning framework. *IEEE Transactions on Geoscience and Remote Sensing*, 56(2):847–858, 2018.
- [57] Swalpa Kumar Roy, Gopal Krishna, Shiv Ram Dubey, and Bidyut B. Chaudhuri. Hybridsn: Exploring 3-d-2-d cnn feature hierarchy for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 17(2):277–281, 2020.
- [58] Xiaohu Ma, Mengmeng Zhang, Zheng Kan, Yunhao Gao, and Wei Li. A lightweight spatial-spectral deformable cnn for uav hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 63:1–13, 2025.
- [59] Xiangli Yang, Zixing Song, Irwin King, and Zenglin Xu. A survey on deep semi-supervised learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(9):8934–8954, 2023.
- [60] William L. Hamilton. Graph representation learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 14(3):1–159, 2020.
- [61] Xiaofeng Zhao, Junyi Ma, Lei Wang, Zhili Zhang, Yao Ding, and Xiongwu Xiao. A review of hyperspectral image classification based on graph neural networks. *Artificial Intelligence Review*, 58(6):172, mar 2025.
- [62] Yu Zhou, Haixia Zheng, Xin Huang, Shufeng Hao, Dengao Li, and Jumin Zhao. Graph neural networks: Taxonomy, advances, and trends. *ACM Transactions on Intelligent Systems and Technology*, 13(1):1–54, January 2022.

- [63] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks, 2017.
- [64] Weigang Lu, Yibing Zhan, Binbin Lin, Ziyu Guan, Liu Liu, Baosheng Yu, Wei Zhao, Yaming Yang, and Dacheng Tao. Skipnode: On alleviating performance degradation for deep graph convolutional networks. *IEEE Transactions on Knowledge and Data Engineering*, 36(11):7030–7043, 2024.
- [65] Guoyong Wu, Mohammed A.A. Al-qaness, Dalal Al-Alimi, Abdelghani Dahou, Mohamed Abd Elaziz, and Ahmed A. Ewees. Hyperspectral image classification using graph convolutional network: A comprehensive review. *Expert Systems with Applications*, 257:125106, 2024.
- [66] Yuqun Yang, Xu Tang, Xiangrong Zhang, Jingjing Ma, Fang Liu, Xiuping Jia, and Licheng Jiao. Semi-supervised multiscale dynamic graph convolution network for hyperspectral image classification. *IEEE Transactions on Neural Networks and Learning Systems*, 35(5):6806–6820, 2024.
- [67] Qi Diao, Yaping Dai, Jiacheng Wang, Xiaoxue Feng, Feng Pan, and Ce Zhang. Spatial-pooling-based graph attention u-net for hyperspectral image classification. *Remote Sensing*, 16(6), 2024.
- [68] Ying Cui, Li Luo, Lu Wang, Liwei Chen, Shan Gao, Chunhui Zhao, and Cheng Tang. Hyperspectral image classification based on double-hop graph attention multiview fusion network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:20080–20097, 2024.
- [69] Wentao Yu, Sheng Wan, Guangyu Li, Jian Yang, and Chen Gong. Hyperspectral image classification with contrastive graph convolutional network. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–15, 2023.
- [70] Hui Zhang, Kaiping Tu, Huanhuan Lv, and Ruiqin Wang. Hyperspectral image classification based on 3d–2d hybrid convolution and graph attention mechanism. *Neural Processing Letters*, 56(2):117, March 2024.
- [71] Zhonghao Chen, Guoyong Wu, Hongmin Gao, Yao Ding, Danfeng Hong, and Bing Zhang. Local aggregation and global attention network for hyperspectral image classification with spectral-induced aligned superpixel segmentation. *Expert Systems with Applications*, 232:120828, 2023.
- [72] Ronghua Shang, Keyao Zhu, Huidong Chang, Weitong Zhang, Jie Feng, and Songhua Xu. Hyperspectral image classification based on mixed similarity graph convolutional network and pixel refinement. *Applied Soft Computing*, 170:112657, 2025.

- [73] Ke Wu, Yanting Zhan, Ying An, and Suyi Li. Multiscale feature search-based graph convolutional network for hyperspectral image classification. *Remote Sensing*, 16(13), 2024.
- [74] Yanfei Zhong, Xin Hu, Chang Luo, Xinyu Wang, Ji Zhao, and Liangpei Zhang. Whu-hi: Uav-borne hyperspectral with high spatial resolution (h2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with crf. *Remote Sensing of Environment*, 250:112012, 2020.