INTERNATIONAL DOCTORAL
SCHOOL OF THE USC

Álvaro
Acción Montes

PhD Thesis

Techniques for the extraction of spatial and spectral information in the supervised classification of hyperspectral imagery for land-cover applications

Santiago de Compostela, 2022

Doctoral Programme in Information Technology Research

TESE DE DOUTORAMENTO

TECHNIQUES FOR THE EXTRACTION OF SPATIAL AND SPECTRAL INFORMATION IN THE SUPERVISED CLASSIFICATION OF HYPERSPECTRAL IMAGERY FOR LAND-COVER APPLICATIONS

Álvaro Acción Montes

ESCOLA DE DOUTORAMENTO INTERNACIONAL DA UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

PROGRAMA DE DOUTORAMENTO EN INVESTIGACIÓN EN TECNOLOXÍAS DA INFORMACIÓN

SANTIAGO DE COMPOSTELA
2023

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Declaración do autor da tese

Don Álvaro Acción Montes

Título da tese: Techniques for the extraction of spatial and spectral information in the supervised classification of hyperspectral imagery for land-cover applications

Presento a miña tese, seguindo o procedemento adecuado ao Regulamento, e declaro que:

1. A tese abarca os resultados da elaboración do meu traballo.
2. De ser o caso, na tese faise referencia ás colaboracións que tivo este traballo.
3. Confirmo que a tese non incorre en ningún tipo de plaxio doutros autores nin de traballos presentados por min para a obtención doutros títulos.
4. A tese é a versión definitiva presentada para a súa defensa e coincide a versión impresa coa presentada en formato electrónico.

E comprométome a presentar o Compromiso Documental de Supervisión no caso de que o orixinal non estea na Escola.

En Manila, 21 de decembro de 2022

Asdo. Álvaro Acción Montes

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Autorización do director/titor da tese

Techniques for the extraction of spatial and spectral information in the supervised classification of hyperspectral imagery for land-cover applications

Dona Dora Blanco Heras, Profesor Titular da Área de Arquitectura e Tecnoloxía de Computadores da Universidade de Santiago de Compostela

Don Francisco Santiago Argüello Pedreira, Profesor Titular da Área de Arquitectura e Tecnoloxía de Computadores da Universidade de Santiago de Compostela

INFORMAN:

Que a presente tese correspóndese co traballo realizado por Don Álvaro Acción Montes, baixo a nosa dirección/titorización, e autorizamos a súa presentación, considerando que reúne os requisitos esixidos no Regulamento de Estudos de Doutoramento da USC, e que como directores/titores desta non incorre nas causas de abstención establecidas na Lei 40/2015.

De acordo co indicado no Regulamento de Estudos de Doutoramento, declaramos tamén que a presente tese de doutoramento é idónea para ser defendida en base á modalidade de COMPENDIO DE PUBLICACIÓNS, nos que a participación do/a doutorando/a foi decisiva para a súa elaboración e as publicacións se axustan ao Plan de Investigación.

En Santiago de Compostela, 21 de decembro de 2022

Asdo. Dora Blanco Heras
Directora tese

Asdo. Francisco Santiago Argüello Pedreira
Director tese

To my family and friends

Thank you very much for your support and understanding.
I would also like to extend my gratitude to my supervisors and colleagues, whose help has made this thesis possible.

Be strong enough to stand alone, smart enough to know when you need help, and brave enough to ask for it

Ziad K. Abdelnour

Acronyms

AA	Average Accuracy
ADP	Anisotropic Diffusion Profile
AE	Autoencoder
AOS	Additive Operator Splitting
AP	Attribute Profile
CNN	Convolutional Neural Networks
CPU	Central Processing Unit
CRF	Conditional Random Field
CUDA	Compute Unified Device Architecture
DBN	Deep Belief Network
DL	Deep Learning
DMA	Direct Memory Access

DMP	Differential Morphological Profile
DWS	Dual-Window Superpixel
EADP	Extended Anisotropic Diffusion Profile
EEP	Extended Extinction Profile
EF	Extinction Filter
ELU	Exponential Linear Unit
EMP	Extended Morphological Profile
EP	Extinction Profile
FE	Feature Extraction
FED	Fast Explicit Diffusion
GAN	Generative Adversarial Networks
GPU	Graphics Processing Unit
HSI	Hyperspectral Image
ICA	Independent Component Analysis
JL	Joint Learning
LULC	Land Use/Land Cover

ML	Machine Learning
MP	Morphological Profile
MSI	Multispectral Image
NADAM	Nesterov-accelerated Adaptive Movement Estimation
NVCC	NVIDIA CUDA Compiler
OA	Overall Accuracy
OpenMP	Open Multi-Processing
PCA	Principal Component Analysis
ReLU	Rectified Linear Unit
RF	Random Forest
RGB	Red, green, and blue
RO	Random Occlusion
SE	Structuring Element
SIMT	Single-Instruction Multiple-Thread
SLIC	Simple Linear Iterative Clustering
SM	Streaming Multiprocessor
SPMD	Single-Program Multiple-Data

Spp.	Superpixels
SR	Simplification Ratio
SSA	Singular Spectrum Analysis
SSIM	Structural Similarity Index Metric
SSRN	Spectral-Spatial Residual Networks
SVD	Singular Value Decomposition
SVM	Support Vector Machine
UAV	Unmanned Aerial Vehicle
UMA	Uniform Memory Access

Resumo

O obxectivo desta tese de doutoramento é o desenvolvemento de técnicas de extracción de información espacial-espectral para tarefas de clasificación supervisada, tanto mediante modelos clásicos como baseados na aprendizaxe profunda, destinadas ao seu uso na clasificación de imaxes multi e hiperespectrais de uso do solo ou cobertura terrestre (LULC ou *Land Use/Land Cover*) obtidas mediante teledetección. A finalidade principal que se persegue é a aplicación eficiente destas técnicas, de forma que sexan capaces de obter resultados de clasificación satisfactorios cun baixo uso de recursos computacionais e baixo tempo de execución. Nesta tese desenvólvense dúas liñas de investigación principais: Unha primeira liña orientada ao desenvolvemento de técnicas de extracción de información deseñadas para o seu uso con modelos clásicos, xunto coa súa adaptación a arquitecturas multinúcleo e GPU (*Graphics Processing Unit*) de consumo. Unha segunda liña orientada ao desenvolvemento de técnicas de extracción de información destinadas a aumentar os datos para modelos baseados na aprendizaxe profunda.

A teledetección pode definirse como a adquisición de información dun determinado obxecto ou, máis xenericamente, escena, sen un contacto físico directo. Esta tarefa é comúnmente realizada por unha variedade de sensores especializados que capturan información do espectro electromagnético. Dependendo da súa finalidade, estes sensores diferéncianse, entre outras cousas, no rango do espectro electromagnético sobre o que operan e a información que son capaces de adquirir. Os avances tecnolóxicos na captura de imaxes multidimensionais melloraron a capacidade de adquisición de datos dos sensores utilizados tanto en satélites como, máis recentemente, en vehículos aéreos non tripulados (UAVs ou Unmanned Aerial Vehicles). As imaxes hiperespectrais e multiespectrais véñense empregando con éxito para realizar diversas tarefas como estudos de biodiversidade [158], xestión de inventarios de vexetación [124], etc., de xeito semiautomatizado ou para aumentar a eficiencia dos procesos de produción

en campos como a agricultura de precisión [75, 146]. A principal característica distintiva deste tipo de imaxes reside no rango electromagnético e no número de frecuencias nas que se capta a información. A información captada pode incluír partes do espectro infravermello e ultravioleta mentres que o número de bandas pode ir de tres a algo máis de dez, no caso das imaxes multiespectrais, ata decenas ou centos de bandas no caso de imaxes hiperespectrais. Esta gran cantidade de información permite unha maior separabilidade [67] entre os distintos materiais presentes nunha escena en comparación coas imaxes RGB, o que fixo que este tipo de imaxes se empreguen habitualmente en tarefas de clasificación no campo da teledetección. As imaxes multiespectrais e hiperespectrais son análogas a un cubo de datos tridimensional, con dúas dimensións espaciais correspondentes á anchura e á altura, e unha dimensión espectral formada polas diferentes bandas espectrais.

Dependendo da composición química dunha substancia, esta presenta interaccións específicas co espectro electromagnético. Na espectroscopía de imaxes, cada píxel dunha imaxe está composto por un vector de bandas no que cada un dos seus elementos representa un valor de intensidade para unha determinada lonxitude de onda [56]. Esta intensidade está relacionada coas características do material [59], onde diferentes materiais presentan distintos perfís de emisión, absorción, reflectancia, fluorescencia, etc. para diferentes frecuencias. Grazas a este principio, cada un dos materiais presentes nunha escena pode caracterizarse polo conxunto de valores de intensidade en cada unha das bandas para os píxeles onde se sitúa. Esta característica, tamén chamada sinatura espectral [104], é específica de cada material e permite a súa identificación estudando a súa semellanza con outras sinaturas de materiais coñecidas. Na práctica, o proceso de identificación de materiais mediante a súa sinatura espectral vese afectado polo fenómeno da variabilidade espectral [31], onde factores como a iluminación, os efectos atmosféricos ou o ruído nas medicións [123, 21] modifican as sinaturas espectrais.

A clasificación de imaxes de alta dimensionalidade é un dos campos científicos máis activos das últimas décadas [138, 168, 157, 127, 45]. Este proceso de clasificación consiste en asignar etiquetas de clase aos píxeles presentes nunha imaxe nun proceso que pode ser non supervisado ou supervisado [6]:

- Nun proceso de clasificación non supervisado, os píxeles agrúpanse en conxuntos similares sen o uso de información previa. Neste proceso non supervisado, a imaxe de teledetección divídese en varios grupos de píxeles mediante a aplicación de algoritmos de *clustering*.

- No proceso de clasificación supervisada emprégase un subconxunto de datos que se considera representativo do conxunto orixinal, no cal existen pares de valores compostos por unha mostra e a etiqueta de clase á que esta pertence, para adestrar un modelo de aprendizaxe automática (para a aprendizaxe baseada en modelos) ou para a comparación de mostras cruzadas (para a aprendizaxe baseada en instancias). Esta información empregarase de seguido para predicir a clase do resto do conxunto orixinal.

A aprendizaxe automática tivo un papel fundamental no avance das aplicacións para as imaxes de alta dimensionalidade como ferramenta no ámbito civil. Os avances en este campo están relacionados coas melloras tecnolóxicas que permitiron o incremento nas prestacións dos sensores empregados para a captura de datos e os sistemas de computación requiridos para o seu procesamento. Un esquema de clasificación tradicional para esta tarefa está composto de catro etapas fundamentalmente, que son: preprocesado, onde un grupo de transformacións iniciais se aplica aos datos de entrada para mellorar os datos ou preparalo para a clasificación; extracción de características, onde os datos son transformados para extraer as características relevantes; clasificación, onde os datos de adestramento son enviados ao modelo seleccionado para a extracción de información que permita asignar as etiquetas de clase ás mostras e, finalmente, postprocesado, onde a saída do modelo é mellorada.

Son moitas as técnicas propostas para a clasificación supervisada deste tipo de imaxes. Ditas técnicas foron evolucionando dende o uso da información espectral, exclusivamente, ó uso combinado de información espacial-espectral a través de modelos tradicionais de aprendizaxe automática, ata finalmente chegar á aplicación de aprendizaxe mediante redes neuronais profundas. Distinguiremos tres grupos, principalmente:

- Técnicas baseadas na información espectral, onde a imaxe é considerada como un conxunto unidimensional de vectores espectrais. Este tipo de técnicas realizan a clasificación exclusivamente mediante o uso de información espectral sen ningún tipo de estrutura espacial. Exemplos disto poden ser o uso de clasificadores como *Random Forest* (RF) [63, 73, 17] ou máquinas de soporte vectorial (SVM ou *Support Vector Machines*) [61, 48, 30, 132], que non fan uso da información de contexto. Existen múltiples limitacións relacionadas co uso de información só espectral, entre as que destaca que en determinadas circunstancias a separabilidade entre clases pode ser baixa debido á variabilidade espectral para cada material e a similitude entre as sinaturas dos diferentes materiais. Esta situación pode darse, por exemplo, no caso da clasificación de varias

especies de vexetación, onde a información das texturas e as súas métricas pode ser crucial para a separabilidade entre as clases [15]. Ademais disto, para resolucións espaciais baixas, un só píxel pode conter unha combinación de información dos espectros de varias clases. Para corrixir este problema, lévase a cabo un proceso de *hyperespectral unmixing* [109] no que se identifican as achegas dos distintos materiais para cada píxel da escena. Así mesmo, o gran número de bandas altamente correlacionadas xunto cun número relativamente baixo de mostras e a variabilidade espectral poden provocar a aparición do denominado fenómeno Hughes [78] ou maldición da dimensionalidade. Para evitalo, o uso de técnicas de redución da dimensionalidade como PCA [89] ou ICA [144] permiten extraer un número reducido de bandas máis relevantes para a súa posterior análise.

- Técnicas baseadas na información espacial-espectral, nas que a clasificación fai uso das interdependencias ou relacións estruturais dos píxeles dunha imaxe. As formas e texturas xeradas polos píxeles que forman a imaxe poden permitir un aumento da precisión obtida polos clasificadores [50]. Dentro deste grupo atópanse técnicas como a regularización espacial [165], os diferentes perfís morfolóxicos [19, 18, 44, 57] ou a aprendizaxe conxunta (JL ou *Joint Learning*). [114]. A regularización espacial baséase na aplicación dun modelo con información espacial aos resultados da clasificación de datos espectrais. Hai varias formas de conseguilo, xa sexa aplicando métodos non supervisados como a segmentación [91], ou ben métodos supervisados como [155] campos de Markov ou campos aleatorios condicionais (CRF ou *Conditional Random Fields*).) [169]. Os Perfís Morfolóxicos (MP ou *Morphological Profiles*), dos que existen numerosas variantes, fan uso da morfoloxía matemática e, en particular, das operacións de apertura e peche para crear varias versións da imaxe orixinal con diferente granularidade que se agrupan no que se denomina perfil morfolóxico e despois son procesados por un clasificador. Entre as variantes máis populares dos perfís morfolóxicos podemos atopar perfís de atributos (APs ou *Attribute Profiles*) [44] ou perfís de extinción (EPs ou *Extinction Profiles*) [57]. Finalmente, a aprendizaxe conxunta consiste na aprendizaxe simultánea de características espaciais e espectrais mediante o uso de *kernels* específicos. Isto pódese ver, por exemplo, no uso de tales *kernels* en combinación con SVM para a análise de veciñanza de píxeles en tarefas de clasificación de [49].

- Técnicas baseadas na aprendizaxe profunda, que foron introducidas máis recentemente para a clasificación de datos de teledetección e, en particular, para imaxes multi e hiper-

espectrais [33, 159, 12, 85, 108]. Os resultados obtidos polos métodos de deep learning foron superiores aos obtidos polos clasificadores tradicionais e permitiron avanzar en diversas frontes, como a extracción de características (FE ou *Feature Extraction*) ao incorporar *autoencorders* ou xeración de datos sintéticos procedentes do uso de redes xerativas adversarias (GANs ou *Generative Adversarial Networks*) [164, 147]. As redes neuronais convolucionais (CNN ou *Convolutional Neural Networks*) [80] utilizáronse amplamente para a clasificación supervisada de imaxes de alta dimensionalidade [81]. Outras arquitecturas frecuentes inclúen *Deep Belief Networks* (DBN) [33], *Autoencoders* (AE) [156], ou redes residuais espectrais-espaciais (SSRN ou Spectral-Spatial Residual Networks) [170].

En xeral, un maior rango electromagnético e unha maior resolución, tanto espectral (número de bandas separables contiguas) como espacial (densidade de píxeles por unidade de área), dan como resultado unha identificación máis precisa dos materiais nunha escena. O uso de imaxes de alta dimensionalidade presenta uns retos [120, 94, 20] que teñen que ver co enorme volume de datos xerados e o custo computacional do seu procesamento, o que fai que a creación de algoritmos e técnicas eficientes aplicables aos esquemas de clasificación adquira especial importancia ante o uso de unha cantidade de datos cada vez maior. O uso da información espacial-espectral, xunto con técnicas baseadas na aprendizaxe profunda, presentan mellores resultados a costa dun maior uso dos recursos computacionais [29]. No caso destes últimos, as enormes necesidades de datos necesarias para conseguir unha boa xeneralización do modelo supoñen un problema nun ámbito no que a obtención de conxuntos de datos etiquetados pode resultar custosa debido á necesidade de intervención humana no proceso de etiquetado. Xunto a isto, a tendencia actual ao aumento da complexidade dos modelos xera unha explosión no número de parámetros que entrou no campo dos rendementos decrecentes [134].

Co obxectivo de democratizar o acceso ao procesamento das imaxes multi e hiperespectrais e, potencialmente nun futuro, levar a cabo a transición á computación móbil, resulta de especial relevancia a obtención de algoritmos que permitan a realización destas tarefas en hardware de propósito xeral. O anteriormente exposto implica a necesidade de técnicas de extracción de información espacial-espectral que sexan eficientes e capaces de obter resultados de clasificación satisfactorios nun hardware con recursos máis reducidos. Por este motivo, defínense como parte desta tese as seguintes hipóteses:

- **H1**. A aplicación de técnicas de preprocesado baseadas na extracción de información espacial-espectral, como a segmentación ou a construción de perfís baseados en operacións morfolóxicas, atributos ou difusión, permiten unha mellora na precisión da clasificación de imaxes multi e hiperespectrais de cobertura do solo. Esto é válido tanto para técnicas consideradas clásicas como para as baseadas en redes de aprendizaxe profunda.

- **H2**. O desenvolvemento de algoritmos para arquitecturas altamente paralelas como GPUs pode xerar ganancias significativas nos tempos de execución sobre as súas contrapartidas para CPU no procesamento de imaxes multi e hiperespectrais.

- **H3**. O preprocesado baseado na segmentación multidimensional das imaxes permite reducir a complexidade do adestramento dos modelos baseados na aprendizaxe profunda, conseguindo melloras significativas nos tempos de execución en escenas de gran tamaño.

- **H4**. As técnicas de aumentado de datos aplicadas á aprendizaxe profunda permiten un aumento da precisión de clasificación en situacións con escaso número de mostras de adestramento.

Das hipóteses anteriores extraéronse os seguintes obxectivos específicos:

- **O1**. Desenvolvemento de técnicas de extracción de información espacial-espectral para modelos de clasificación clásicos. Análise do estado da arte da clasificación de imaxes multi e hiperespectrais de cobertura do terrestre e desenvolvemento de novas técnicas eficientes para a extracción de información espacial-espectral.

- **O2**. Desenvolvemento de técnicas de extracción de información espacial-espectral para modelos baseados na aprendizaxe profunda. Creación de esquemas de clasificación eficientes orientados a imaxes multi e hiperespectrais de cobertura do solo de alta e moi alta resolución.

- **O3**. Proxección dos métodos desenvolvidos en arquitecturas paralelas. Consideraranse arquitecturas multinúcleo e GPU.

Dado o requirimento da obtención de algoritmos eficientes orientados a hardware de consumo, as propostas desenvolvidas foron implementadas sobre C++, empregando OpenMP e CUDA para as versións CPU e GPU, respectivamente. Numerosas optimizacións específicas

foron aplicadas orientadas a maximizar o rendemento en GPU, entre as que destacan optimizacións dedicadas ao acceso á memoria, optimizacións relativas ás instrucións para determinadas operacións, e optimizacións orientadas ao aumento do uso de recursos dispoñibles na GPU mediante a selección de tamaños de bloque. O código para as propostas de métodos de aumentado de datos foi escrito en Python, empregando a biblioteca de código aberto Tensorflow, debido ao total dominio que esta linguaxe ten no campo das redes neuronais e *machine learning* en xeral.

A validación experimental for realizada empregando tres conxuntos de datos con características moi diferentes entre eles. Un primeiro conxunto de datos hiperespectral con imaxes comunmente empregadas para *benchmarking*, de baixa resolución espacial capturadas polos sensores ROSIS-03 e AVIRIS con escenas de vexetación e urbanas. Un segundo conxunto de datos, denominado *Galicia dataset*, formado imaxes multiespectrais de vexetación próxima a cuncas de ríos, capturadas en varias rexións de Galicia mediante un sensor montado en un UAV. Estas imaxes presentan un tamaño considerablemente superior ás do primeiro conxunto e teñen unha gran resolución espacial, o que fai especialmente custoso o seu procesamento. Por último, un conxunto de datos multiespectral composto por 18 imaxes obtido polo satélite Gaofen-2 con escenas de diferentes rexións de China.

A validación das hipóteses como parte do traballo desta tese para acadar os obxectivos anteriormente descritos deu lugar ás seguintes achegas:

1. Proposta dun método de extracción de información espacial-espectral baseado na aplicación da difusión non lineal. Esta proposta deu lugar aos chamados perfís de difusión anisotrópica estendida (EADP ou *Extended Anisotropic Diffusion Profiles*, inspirados en perfís morfolóxicos. O EADP constrúese mediante a concatenación dunha serie de instancias de difusión non lineal aplicadas sobre os compoñentes principais extraídos dunha imaxe hiperespectral, o que xera novas compoñentes con un nivel de detalle decrecente. A introdución destas compoñentes permite unha redución da variabilidade espectral das sinaturas dos diferentes materiais, o que redunda nun incremento no rendemento na clasificación. Propuxéronse implementacións optimizadas para CPU multinúcleo desenvolvidas mediante OpenMP e para GPU, mediante CUDA, orientadas a hardware de consumo. As capacidades de procesamento paralelo en GPU permitiron acadar unha mellora en tempos de procesamento de arredor de 10× grazas ao emprego de algoritmos eficientes adaptados a estas arquitecturas. Finalmente, detállase unha caracterización dos perfís de difusión e o seu rendemento de clasificación mediante

un modelo SVM con outros métodos existentes na literatura como perfís morfolóxicos estendidos (EMP ou *Extended Morphological Profiles*), perfís de atributos estendidos (EAP ou *Extended Attribute Profiles*), perfís de extinción estendidos (EEP ou *Extended Extinction Profiles)*, etc.

2. Proposta para un método de aumentado de datos baseado na composición de transformacións xeométricas denominadas aumentado de datos de superpíxeles de dobre xanela (DWS ou *Dual-Window Superpixel*). Esta técnica de aumento, orientada a esquemas de clasificación de aprendizaxe profunda baseados en redes neuronais convolucionais, baséase en particionar os parches de datos de entrada á rede en dúas zonas independentes nas que se aplican diferentes transformacións xeométricas convencionais. O método parte da hipótese de que a zona central contén os datos máis relevantes para a clasificación. Os parches de entrada pódense obter mediante unha xanela deslizante, procesando píxel a píxel a imaxe e extraendo un subconxunto de píxeles coas dimensións da xanela ou, a diferencia destos, tomarse arredor dun píxel central para rexións independentes da imaxe. Estas rexións independentes corresponderanse cos superpíxeles obtidos polo algoritmo SLIC (*Simple Linear Iterative Clustering*) [4] e a etiqueta deste píxel central será asignada a todos os seus píxeles. A técnica está dividida en tres etapas diferentes: extracción de píxeles baseada en superpíxeles, subdivisión de parches de entrada, e transformación dos parches. Usar esta técnica de aumento xunto coa segmentación de superpíxeles reduce significativamente os custos de procesamento de imaxes, permitindo que se use en imaxes grandes con tempos de execución razoables. Esta proposta obtén unha maior precisión global que as transformacións de espello e rotación cando estas se aplican á rexión interior en lugar de a todo o parche. Ademais, a aplicación das transformacións sobre ambas rexións permite a xeración dunha maior cantidade de datos, o que mellora o rendemento da clasificación. O método proposto foi empregado xunto cunha arquitectura CNN, onde consegue obter rendementos superiores a métodos de aumentado de datos comparables presentes na literatura.

3. Proposta para un método de aumentado de datos [125] baseado na imputación de datos [141]. A premisa consiste na substitución de determinados datos da zona exterior dos parches de entrada obtidos mediante o uso dun algoritmo de segmentación de superpíxeles. Estes datos, en zonas de bordes irregulares nas estruturas da escena, poden pertencer a clases diferentes da clase real do parche e a súa eliminación permite mellorar a precisión da clasificación en esquemas de aprendizaxe profunda baseados en redes

neuronais convolucionais. Para xerar estes datos sintéticos utilízanse diversas técnicas de imputación de datos, dando lugar a un novo parche aumentado que se envía á rede neuronal xunto co parche orixinal. O uso da segmentación en superpíxeles permite reducir os custos de procesamento de imaxes, como sucedía con DWS. A novidade introducida neste traballo está na eliminación e o reemprazo dos píxeles dos parches de entrada mediante o uso dunha combinación da información da segmentación e imputación de datos, o que redunda en parches de entrada máis homoxéneos e permite a mellora da precisión da clasificación. Os resultados da proposta obtiveron a maior precisión en 16 das 17 escenas do conxunto de datos Gaofen-2 e constitúe o primeiro exemplo do uso de imputación para o aumentado de datos en imaxes multi e hiperespectrais.

Para a avaliación dos resultados, empregáronse diversas métricas comunmente presentes na literatura, como a precisión total, precisión media, ou *Kappa*. O custo computacional dos algoritmos foi medido en termos do tempo de execución, ou tempo de reloxo requirido entre a primeira etapa dun algoritmo e a etapa final. Dita avaliación dos resultados obtidos polas diferentes propostas apoiou as hipóteses expostas e permitiu determinar o cumprimento satisfactorio dos obxectivos marcados ao comezo da tese. Os métodos de extracción de información que se presentan son aplicables a unha variedade de problemas relevantes no ámbito da clasificación de imaxes multi e hiperespectrais.

Como traballo futuro, contémplase a posibilidade de aplicar os EADP a outros problemas, como poden ser aqueles relacionados co rexistrado, a detección de cambios ou a detección de anomalías. Do mesmo xeito, as propostas de aumentado de datos non están limitadas a unha arquitectura concreta de rede e podería estudarse o seu uso con outras arquitecturas diferentes. Adicionalmente, a inclusión de técnicas para a mitigación do desbalanceo de clases podería demostrar ser de interese para a mellora dos resultados de clasificación.

Resumen

El objetivo de la presente tesis doctoral es el desarrollo de técnicas de extracción información espacial-espectral para tareas de clasificación supervisada, tanto utilizando modelos clásicos como los basados en aprendizaje profundo, destinados a su uso en la clasificación de imágenes multi e hiperespectrales de uso de suelo o cobertura terrestre (LULC o *Land Use/Land Cover*) obtenidas mediante teledetección. El principal propósito que se persigue es la aplicación de dichas técnicas de forma eficiente, con el fin de que sean capaces de obtener resultados de clasificación satisfactorios con un bajo uso de recursos computacionales y bajo tiempo de ejecución. En esta tesis se desarrollan dos líneas de investigación principales: Una primera línea orientada al desarrollo de técnicas de extracción de información pensadas para su uso con modelos clásicos, junto a su adaptación a arquitecturas multinúcleo y GPU (*Graphics Processing Unit*) de consumo. Una segunda línea orientada al desarrollo de técnicas de extracción de información destinadas al aumentado de datos para modelos basados en aprendizaje profundo.

La teledetección puede definirse como la adquisición de información de un determinado objeto o, más genéricamente, escena, sin un contacto físico directo. Esta tarea es comúnmente realizada por una variedad de sensores especializados que capturan información del espectro electromagnético. Dependiendo de su finalidad, estos sensores se diferencian, entre otras cosas, en el rango del espectro electromagnético sobre el que operan y la información que son capaces de adquirir. Los avances tecnológicos en la captura de imágenes multidimensionales han mejorado la capacidad de adquisición de datos de los sensores utilizados tanto en satélites como, más recientemente, en vehículos aéreos no tripulados (UAVs o Unmanned Aerial Vehicles). Las imágenes hiperespectrales y multiespectrales se han utilizado con éxito para realizar diversas tareas como estudios de biodiversidad [158], gestión de inventarios de vegetación [124], etc., de forma semiautomatizada o para aumentar la eficiencia de los procesos productivos en campos como la agricultura de precisión [75, 146]. La principal

característica distintiva de este tipo de imágenes radica en el rango electromagnético y el número de frecuencias en las que se captura información. La información capturada puede incluir partes del espectro infrarrojo y ultravioleta, mientas que la cantidad de bandas puede ir desde tres a algo más de diez en el caso de las imágenes multiespectrales, a decenas o cientos de bandas en el caso de las imágenes hiperespectrales. Esta gran cantidad de información permite una mayor separabilidad [67] entre los distintos materiales presentes en una escena si se compara con imágenes RGB, lo que ha hecho que este tipo de imágenes se usen de forma habitual en tareas de clasificación en el campo de la teledetección. Las imágenes multi e hiperespectrales son análogas a un cubo de datos de tres dimensiones, con dos dimensiones espaciales correspondientes al ancho y alto y una dimensión espectral formada por las diferentes bandas espectrales.

Dependiendo de la composición química de una substancia, ésta presenta unas interacciones concretas con el espectro electromagnético. En la espectroscopia de imágenes, cada píxel de una imagen está compuesto por un vector de bandas en el que cada uno de sus elementos representa un valor de intensidad para una determinada longitud de onda [56]. Dicha intensidad está relacionada con las características del material [59], con diferentes materiales presentando diferentes perfiles de emisión, absorción, reflectancia, florescencia, etc. para diferentes frecuencias. Gracias a este principio, cada uno de los materiales presentes un una escena puede caracterizarse por el conjunto de valores de intensidad en cada una de las bandas para los píxeles donde está situado. Esta característica, también denominada firma espectral [104], es específica de cada material y posibilita su identificación mediante el estudio de su similitud con otros firmas de materiales ya conocidas. En la práctica, el proceso de identificación de los materiales mediante su firma espectral se ve afectado por el fenómeno de la variabilidad espectral [31], donde factores como la iluminación, efectos atmosféricos o el ruido en las mediciones [123, 21] modifican las firmas espectrales.

La clasificación de imágenes de alta dimensionalidad es uno de los campos con mayor actividad científica de las últimas décadas [138, 168, 157, 127, 45]. Este proceso de clasificación consiste en la asignación de etiquetas de clase a los píxeles presentes en una imagen en un proceso que puede ser no supervisado o supervisado [6]:

- En un proceso de clasificación no supervisado, los píxeles son agrupados en conjuntos similares sin el uso de información previa. En este proceso no supervisado, la imagen obtenida mediante teledetección es particionada en un número de grupos de píxeles mediante la aplicación de algoritmos de *clustering*.

- En el proceso de clasificación supervisado, un subconjunto de datos que se considera representativo del conjunto original, del que se tienen pares de valores compuestos por una muestra y la etiqueta de clase a la que ésta pertenece, es utilizado para el entrenamiento de un modelo de aprendizaje automático (para aprendizaje basado en modelos) o para la comparación entre muestras (para aprendizaje basado en instancias). Dicha información será luego utilizada para predecir la clase del resto del conjunto original.

El aprendizaje automático tuvo un papel fundamental en el avance de las aplicaciones para las imágenes de alta dimensionalidad como herramienta en el ámbito civil. Los avances en este campo están relacionados con las mejoras tecnológicas que permitieron el incremento en las prestaciones de los sensores empleados para la captura de datos y los sistemas de computación requeridos para su procesamiento. Un esquema de clasificación tradicional para esta tarea está compuesto de cuatro etapas fundamentalmente, que son: preprocesado, donde un grupo de transformaciones iniciales se aplica a los datos de entrada para mejorar los datos o prepararlo para la clasificación; extracción de características, donde los datos son transformados para extraer las características relevantes; clasificación, donde los datos de entrenamiento son enviados al modelo seleccionado para la extracción de información que permita asignar las etiquetas de clase a las muestras y, finalmente, postprocesado, donde la salida del modelo es mejorada.

Existen multitud de técnicas propuestas para la clasificación supervisada de este tipo de imágenes. Dichas técnicas han ido evolucionando desde el uso de la información espectral, exclusivamente, al uso combinado de información espacial-espectral mediante modelos de aprendizaje automático tradicionales, para terminal por la aplicación de aprendizaje profundo y sofisticadas redes neuronales. Distinguiremos tres grupos, principalmente:

- Técnicas basadas en información espectral, donde la imagen es considerada como un conjunto unidimensional de vectores espectrales. Este tipo de técnicas realizan la clasificación exclusivamente mediante el uso de la información espectral sin ningún tipo de estructura espacial. Ejemplos de esto pueden ser el uso de clasificadores tales como RF (*Random Forest*) [63, 73, 17] o máquinas de soporte vectorial (SVM o *Support Vector Machine*) [61, 48, 30, 132], que no usan información de contexto. Existen múltiples limitaciones relativas al uso de únicamente la información espectral, entre las que destaca que en determinadas circunstancias la separabilidad entre clases puede

ser baja debido a la variabilidad espectral para cada material y la similitud entre las firmas de los diferentes materiales. Esta situación puede darse, por ejemplo, en el caso de la clasificación de diversas especies de vegetación, donde la información de las texturas y sus métricas puede resultar crucial para la separabilidad entre clases [15]. Junto a esto, para bajas resoluciones espaciales, un único pixel puede contener una combinación de información de los espectros de múltiples clases. Para corregir este problema, se lleva a cabo un proceso de *hyperespectral unmixing* [109] en el que se identifican las contribuciones de los diferentes materiales para cada pixel en la escena. Así mismo, el gran número de bandas altamente correlacionadas junto a un número de muestras relativamente bajo y la variabilidad espectral pueden causar la aparición del denominado como fenómeno de Hughes [78] o la maldición de la dimensionalidad. Para evitar esto, el uso de técnicas de reducción de la dimensionalidad como PCA [89] o ICA [144] permiten la extracción de un número reducido de bandas más relevantes para su posterior análisis.

- Técnicas basadas en información espacial-espectral, en las que la clasificación hace uso de las interdependencias o relaciones estructurales de los píxeles en una imagen. Las formas y texturas que generan los pixeles que componen la imagen pueden permitir un incremento en las precisiones obtenidas por los clasificadores [50]. Dentro de este grupo se enmarcan técnicas tales como la regularización espacial [165], los diferentes perfiles morfológicos [19, 18, 44, 57] o el aprendizaje conjunto (JL o *Joint Learning*). [114]. La regularización espacial se basa en la aplicación de un modelo con información espacial a los resultados de la clasificación de los datos espectrales. Existen múltiples formas de conseguir esto, bien sea mediante la aplicación de métodos no supervisados como la segmentación [91], o bien métodos supervisados como campos de Markov [155] o los campos condicionales aleatorios (CRFs o *Conditional Random Fields*) [169]. Los perfiles morfológicos (MPs o *Morphological Profiles*), de los que existen numerosas variantes, hacen uso de la morfología matemática y, en concreto, de las operaciones de apertura y cierre para crear múltiples versiones con distinta granularidad de la imagen original que se agrupan en lo que se denomina perfil morfológico y luego son procesadas por un clasificador. Entre las variantes más populares de los perfiles morfológicos podemos encontrar los perfiles de atributos (APs o *Attribute Profiles*) [44] o los perfiles de extinción (EPs o *Extinction Profiles*) [57]. Por último, el aprendizaje conjunto consiste en el aprendizaje simultáneo de características espaciales y espectrales

mediante el uso de *kernels* específicos. Esto puede verse, por ejemplo, en el uso de dichos *kernels* en combinación con SVM para el análisis de vecindad de píxeles en tareas de clasificación [49].

- Técnicas basadas en aprendizaje profundo, que han sido introducidas de forma más reciente para la clasificación de datos de teledetección y, en particular, para imágenes multi e hiperespectrales [33, 159, 12, 85, 108]. Los resultados obtenidos por los métodos de aprendizaje profundo han resultado superiores a los obtenidos mediante clasificadores tradicionales y han permitido avanzar en diversos frentes, como la extracción de características (FE o *Feature Extraction*) mediante la incorporación de *autoencorders* o generación de datos sintéticos a partir del uso de redes generativas adversarias (GANs o *Generative Adversarial Networks*) [164, 147]. Las redes neuronales convolucionales (CNNs o *Convolutional Neural Networks*) [80] han sido usadas de forma extensiva para la clasificación supervisada de imágenes de alta dimensionalidad [81]. Otras arquitecturas frecuentes incluyen *Deep Belief Networks* (DBN) [33], *Autoencoders* (AE) [156], o redes residuales espectrales-espaciales (SSRN o Spectral-Spatial Residual Networks) [170].

Por lo general, un mayor rango electromagnético y una mayor resolución, tanto espectral (número de bandas contiguas separables) como espacial (densidad de píxeles por unidad de superficie), redundan en una mayor precisión en la identificación de los materiales en una escena. El uso de imágenes de alta dimensionalidad presenta algunos retos [120, 94, 20] que tienen que ver con el ingente volumen de datos generado y el coste computacional de su procesamiento, lo que hace que la creación de algoritmos y técnicas eficientes aplicables a los esquemas de clasificación tomen especial importancia de cara al uso de cantidades cada vez mayores de datos. El uso de información espacial-espectral, junto con las técnicas basadas en aprendizaje profundo presentan mejores resultados a costa de un mayor uso de recursos computacionales [29]. Respecto a esto último, las ingentes necesidades de datos necesarias para conseguir una buena generalización del modelo suponen un problema en un ámbito en el que la obtención de conjuntos de datos etiquetados puede resultar costesa debido a la necesidad de intervención humana en el proceso de etiquetado. Junto a esto, la tendencia actual a un aumento de la complejidad de los modelos genera una explosión del número de parámetros que ha entrado en el terreno de los retornos decrecientes [134].

Con el objetivo de democratizar el acceso al procesamiento de las imágenes multi e

hiperespectrales y, potencialmente en un futuro, llevar a cabo la transición a la computación móvil, resulta de especial relevancia la obtención de algoritmos que permitan la realización de estas tareas en hardware de propósito general. Lo anteriormente expuesto pone de manifiesto la necesidad de técnicas de extracción de información espacial-espectral eficientes y capaces de obtener resultados de clasificación satisfactorios en un hardware con recursos reducidos. Por este motivo, se definen las siguientes hipótesis como parte de esta tesis:

- **H1**. La aplicación de técnicas de preprocesamiento basadas en la extracción de información espacial-espectral, como la segmentación o la construcción de perfiles basados en operaciones morfológicas, atributos o difusión, permiten una mejora en la precisión de la clasificación de imágenes multi e hiperespectrales de cobertura terrestre. Esto es válido tanto para técnicas consideradas clásicas como para las basadas en redes de aprendizaje profundo.

- **H2**. El desarrollo de algoritmos para arquitecturas altamente paralelas como GPUs puede generar ganancias significativas en los tiempos de ejecución sobre sus contrapartidas para CPU en el procesamiento de imágenes multi e hiperespectrales.

- **H3**. El preprocesamiento basado en la segmentación multidimensional de las imágenes permite reducir la complejidad del entrenamiento de modelos basados en aprendizaje profundo, consiguiendo mejoras significativas en tiempos de ejecución sobre escenas de gran tamaño.

- **H4**. Las técnicas de aumentado de datos aplicadas al aprendizaje profundo permiten un incremento de la precisión de clasificación en situaciones con un número bajo de muestras de entrenamiento.

De las anteriores hipótesis, los siguientes objetivos específicos fueron extraídos:

- **O1**. Desarrollo de técnicas de extracción de información espacial-espectral para modelos de clasificación clásicos. Análisis del estado del arte de clasificación de imágenes multi e hiperespectrales de cobertura terrestre y desarrollo de nuevas técnicas eficientes de extracción de información espacial-espectral.

- **O2**. Desarrollo de técnicas de extracción de información espacial-espectral para modelos basados en aprendizaje profundo. Creación de esquemas de clasificación eficiente

orientados a imágenes multi e hiperespectrales de cobertura terrestre de alta y muy alta resolución.

- **O3**. Proyección de los métodos desarrollados en arquitecturas paralelas. Se considerarán las arquitecturas multi-núcleo y GPU.

Dado el requerimiento de la obtención de algoritmos eficientes orientados a hardware de consumo, las propuestas desarrolladas fueron implementadas sobre C++, empleando OpenMP y CUDA para las versiones CPU y GPU, respectivamente. Numerosas optimizaciones específicas fueron aplicadas orientadas a maximizar el rendimiento en GPU, entre las que destacan optimizaciones dedicadas al acceso a la memoria, optimizaciones relativas a las instrucciones para determinadas operaciones, y optimizaciones orientadas al aumento del uso de recursos disponibles en la GPU mediante la selección de tamaños de bloque. El código para las propuestas de métodos de aumentado de datos fue escrito en Python, empleando la biblioteca de código abierto Tensorflow, debido al total dominio que este lenguaje tiene en el campo de las redes neuronales y *machine learning* en general.

La validación experimental fue realizada usando tres conjuntos de datos con características muy diferentes entre ellos. Un primer conjunto de datos hiperespectral con imágenes comúnmente empleadas para *benchmarking*, de baja resolución espacial capturadas por los sensores ROSIS-03 y AVIRIS con escenas de vegetación y urbanas. Un segundo conjunto de datos, denominado *Galicia dataset*, formado imágenes multiespectrales de vegetación próxima a cuencas de ríos, capturadas en varias regiones de Galicia mediante un sensor montado en un UAV. Estas imágenes presentan un tamaño considerablemente superior a las del primer conjunto y tienen una gran resolución espacial, lo que hace especialmente costoso su procesamiento. Por último, un conjunto de datos multiespectral compuesto por 18 imágenes obtenido por el satélite Gaofen-2 con escenas de diferentes regiones de China.

La validación de las hipótesis como parte del trabajo de esta tesis para alcanzar los objetivos anteriormente descritos dio lugar a las siguientes aportaciones:

1. Propuesta de un método de extracción de información espacial-espectral basada en la aplicación de difusión no lineal. Dicha propuesta dio lugar a los denominados perfiles extendidos de difusión anisotrópica (EADP, *Extended Anisotropic Diffusion Profiles*), inspirados en los perfiles morfológicos. El EADP se construye mediante la concatenación de una serie de instancias de difusión no lineal aplicadas sobre las componentes principales extraídas de una imagen hiperespectral, lo que genera nuevas componentes

con niveles de detalle decrecientes. La introducción de estas componentes permite una reducción de la variabilidad espectral de las firmas de los diferentes materiales, lo que redunda en un incremento en el rendimiento en la clasificación. Se han propuesto implementaciones optimizadas para CPU multinúcleo desarrolladas mediante OpenMP y para GPU, utilizando CUDA, orientadas a hardware de consumo. Las capacidades de procesamiento paralelo en GPU permitieron alcanzar una mejora en tiempos de procesamiento de alrededor de 10× gracias al empleo de algoritmos eficientes adaptados a estas arquitecturas. Para finalizar, se detalla una caracterización de los perfiles de difusión y su rendimiento en la clasificación mediante un modelo SVM, comparándolo con otros métodos existentes en la literatura tales como perfiles morfológicos extendidos (EMP o *Extended Morphological Profiles*), perfiles de atributos extendidos (EAP o *Extended Attribute Profiles*), perfiles de extinción extendidos (EEP o *Extended Extinction Profiles*, etc.

2. Propuesta de método de aumentado de datos basado en la composición de transformaciones geométricas denominado aumentado de datos de superpíxeles de doble ventana (DWS o Dual-Window Superpixel). Esta técnica de aumentado, orientada a esquemas de clasificación de aprendizaje profundo basados en redes neuronales convolucionales se basa en un particionado de los parches de datos de entrada a la red en dos zonas independientes sobre las que se aplican diferentes transformaciones geométricas convencionales. El método parte de la hipótesis de que la zona central contiene los datos más relevantes para la clasificación. Los parches de entrada se pueden obtener mediante una ventana deslizante, procesando la imagen píxel a píxel y extrayendo un subconjunto de píxeles con las dimensiones de la ventana o, a diferencia del método anterior, tomarse alrededor de un píxel central para regiones independientes de la imagen. Estas regiones independientes se corresponden con los superpíxeles obtenidos mediante el algoritmo SLIC (*Simple Linear Iterative Clustering*) [4] y la etiqueta del píxel central será asignada a todos sus píxeles. El uso de esta técnica de aumentado junto a la segmentación en superpíxeles reduce de forma significativa los costes de procesamiento de las imágenes, lo que permite su uso en imágenes de gran tamaño con tiempos de ejecución razonables. Esta propuesta obtiene una mayor precisión global que las transformaciones de espejo y rotación cuando éstas se aplican a la región interior en lugar de a todo el parche. Además, la aplicación de las transformaciones sobre ambas regiones de forma independiente permite generar una mayor cantidad de datos, lo que mejora el rendimiento de

la clasificación. El método propuesto fue empleado junto con una arquitectura CNN, donde consigue obtener rendimientos superiores a métodos de aumentado de datos comparables presentes en la literatura.

3. Propuesta de método de aumentado de datos [125] basado en la imputación de datos [141]. La premisa consiste en el reemplazo de ciertos datos de la zona exterior de los parches de entrada obtenida mediante el uso de un algoritmo de segmentación en superpíxeles. Estos datos, en zonas de fronteras irregulares en las estructuras de la escena, pueden pertenecer a clases diferentes a la clase real del parche y su eliminación permite una mejoría de las precisiones de clasificación en esquemas de aprendizaje profundo basados en redes neuronales convolucionales. Diversas técnicas de imputación de datos son utilizadas para la generación de estos datos sintéticos, dando lugar a un nuevo parche aumentado que es enviado a la red neuronal junto con el parche original. El uso de la segmentación en superpíxeles permite reducir los costes de procesamiento de las imágenes, al igual que sucedía con DWS. La novedad introducida en este trabajo está en la eliminación y el reemplazo de los píxeles de los parches de entrada mediante el uso de una combinación de la información de la segmentación y la imputación de datos, lo que redunda en parches de entrada más homogéneos y permite una mejora en la precisión de la clasificación. Los resultados de la propuesta obtuvieron la mayor precisión en 16 de las 17 escenas del conjunto de datos Gaofen-2 y constituyen el primer ejemplo del uso de imputación para el aumentado de datos en imágenes multi e hiperespectrales.

Para la evaluación de los resultados, se emplearon diversas métricas comúnmente presentes en la literatura, como la precisión total, precisión media, o *Kappa*. El coste computacional de los algoritmos fue medido en términos del tiempo de ejecución, o tiempo de reloj requerido entre la primera etapa de un algoritmo y la etapa final. Dicha evaluación de los resultados obtenidos por las diferentes propuestas apoyó las hipótesis expuestas y permitió determinar el cumplimiento satisfactorio de los objetivos marcados al inicio de la tesis. Los métodos de extracción de información que se presentan son aplicables la una variedad de problemas relevantes en el ámbito de la clasificación de imágenes multi e hiperespectrales.

Como trabajo futuro, se contempla la posibilidad de aplicar los EADP la otros problemas, como pueden ser aquellos relacionados con el registrado, la detección de cambios o la detección de anomalías. Del mismo modo, las propuestas de aumentado de datos no están limitadas a

una arquitectura concreta de red y podría estudiarse su uso con otras arquitecturas diferentes. Adicionalmente, la inclusión de técnicas para la mitigación del desbalanceo de clases podría demostrar ser de interés para la mejora de los resultados de clasificación.

Summary

The objective of this doctoral thesis is the development of spatial-spectral information extraction techniques for supervised classification tasks, using both classical and deep learning-based models, intended for the classification of remotely sensed multi and hyperspectral, land use, and land cover (LULC) images. The main objective is to apply these techniques in an efficient way so that they are able to obtain satisfactory classification results with low use of computational resources and execution time. Two research lines are developed as part of this thesis: A first research line oriented towards the development of information extraction techniques designed to be used with classical models, as well as their adaptation to multi-core architectures and consumer GPUs (Graphics Processing Unit). A second research line oriented to the development of information extraction techniques aimed at data augmentation for models based on deep learning.

Remote sensing can be defined as the acquisition of information from a given object or, more generically, scene, without direct physical contact. This task is commonly performed by a variety of specialized sensors that capture information from the electromagnetic spectrum. Depending on their purpose, these sensors differ, among other things, in the range of the electromagnetic spectrum over which they operate and the information they are capable of acquiring. Technological advances in multidimensional imaging have improved the data acquisition capabilities of sensors used in both satellites and, more recently, unmanned aerial vehicles (UAVs). Hyperspectral and multispectral images have been successfully used to perform various tasks such as biodiversity studies [158], and vegetation inventory management [124], among others. in a semi-automated way or to increase the efficiency of production processes in fields such as precision agriculture [75, 146]. The main distinguishing feature of this type of imaging lies in the electromagnetic range and the number of frequencies at which the information is captured. The information captured can include parts of the infrared and

ultraviolet spectrum, while the number of bands can range from three to just over ten, in the case of multispectral images, to tens or hundreds of bands in the case of hyperspectral images. This large amount of information allows for a greater separability [67] between the different materials present in a scene compared to RGB images, which has led to this type of images being commonly used in classification tasks in the field of remote sensing. Multispectral and hyperspectral images are analogous to a three-dimensional data cube, with two spatial dimensions corresponding to width and height and a spectral dimension consisting of the different spectral bands.

Depending on the chemical composition of a substance, it has specific interactions with the electromagnetic spectrum. In imaging spectroscopy, each pixel of an image is composed of a vector of bands in which each of its elements represents an intensity value for a given wavelength [56]. This intensity is related to the characteristics of the material [59] since different materials have different emission, absorption, reflectance, fluorescence, etc. profiles for different frequencies. Thanks to this principle, each of the materials present in a scene can be characterized by the set of intensity values of the bands for the pixels where it is located. This characteristic, also called spectral signature [104], is specific to each material and makes it possible to identify it by studying its similarity to other known material signatures. In practice, the process of identifying materials by their spectral signature is affected by the phenomenon of spectral variability [31], where factors such as illumination, atmospheric effects, or noise in the measurements [123, 21] alter the spectral signatures.

High-dimensional image classification is one of the most scientifically active fields in recent decades [138, 168, 157, 127, 45]. This classification process consists of assigning class labels to the pixels present in an image in a process that can be either unsupervised or supervised [6]:

- In an unsupervised classification process, pixels are grouped into similar sets without the use of prior information. In this unsupervised process, the remote sensing image is partitioned into several pixel groups by applying clustering algorithms.

- In the supervised classification process, a subset of data that is considered representative of the original set is used for training a machine learning model (for model-based learning) or for inter-sample comparison (for instance-based learning). The elements of this subset are pairs of values composed of a sample and its corresponding class label. This information will then be used to predict the classes of the rest of the original set.

Machine learning has played a fundamental role in the evolution of applications for high-dimensional images as a tool in the civil sector. Advances in this field are related to the technological improvements that allowed the increase in the performance of the sensors used for data capture and the computer systems required for their processing. A traditional classification scheme for this task is composed of four fundamental stages, which are: preprocessing, where a set of initial transformations is applied to the input data to improve the data or prepare it for classification; feature extraction, where the data is transformed to extract the relevant features; classification, where the training data is sent to the selected model for the extraction of information to assign class labels to the samples and, finally, postprocessing, where the model output is enhanced.

There is a multitude of techniques proposed for the supervised classification of this type of image. These have have evolved from the use of spectral information, exclusively, to the combined use of spatial-spectral information by means of traditional machine learning models, and finally by the application of deep learning and sophisticated neural networks. We will distinguish three main groups:

- Techniques based on spectral information, where the image is considered as a one-dimensional set of spectral vectors. In this type of technique, the classification is performed by using spectral information exclusively, without any spatial structure. Examples of this can be the use of classifiers such as Random Forest (RF) [63, 73, 17] or Support Vector Machines (SVM) [61, 48, 30, 132], which do not use context information. There are multiple limitations regarding the use of only spectral information, among which is that in certain circumstances the separability between classes may be low due to spectral variability for each material and similarity among different material signatures. This situation can occur, for example, in the case of the classification of various vegetation species, where texture information and its metrics can be crucial for the separability between classes [15]. Along with this, for low spatial resolutions, a single pixel may contain a combination of information from the spectra of multiple classes. To correct this problem, a process called hyperspectral unmixing [109], in which the contributions of the different materials are identified for each pixel in the scene, is performed. Likewise, the large number of highly correlated bands together with a relatively low number of samples and spectral variability can cause the occurrence of the so-called Hughes' phenomenon [78] or the curse of dimensionality. To avoid this, the use of dimensionality reduction techniques such as PCA [89] or ICA [144] allow the

extraction of a reduced number of the most relevant bands for further analysis.

- Techniques based on spatial-spectral information, in which classification makes use of the inter-dependencies or structural relationships of the pixels in an image. The shapes and textures generated by the pixels that make up the image can allow an increase in the accuracy obtained by the classifiers [50]. This group includes techniques such as spatial regularization [165], different morphological profiles [19, 18, 44, 57] or Joint Learning (JL). [114]. Spatial regularization is based on the application of a model with spatial information to the classification results of spectral data. There are multiple ways to achieve this, either by applying unsupervised methods such as segmentation [91], or supervised methods such as Markov fields [155] or Conditional Random Fields (CRFs) [169]. Morphological Profiles (MPs), of which there are numerous variants, make use of mathematical morphology and, in particular, open and close operations to clear multiple versions with a different granularity of the original image that are grouped into what is called a morphological profile and then processed by a classifier. Among the most popular variants of morphological profiles we can find Attribute Profiles (APs) [44] or Extinction Profiles (EPs) [57]. Finally, ensemble learning consists of simultaneous learning of spatial and spectral features by using specific kernels. This can be seen, for example, in the use of such kernels in combination with SVM for pixel neighborhood analysis in classification tasks [49].

- Deep learning-based techniques, which have been introduced more recently for classification of remotely sensed data and, in particular, for multi- and hyperspectral imaging [33, 159, 12, 85, 108]. The results obtained by deep learning methods have been superior to those obtained by traditional classifiers and have allowed progress on several fronts, such as feature extraction (FE) by incorporating autoencoders or synthetic data generated from the use of generative adversarial networks (GANs). Convolutional Neural Networks (CNNs) [80] have been used extensively for supervised high-dimensional image classification [81]. Other common network architectures include Deep Belief Networks (DBN) [33], Autoencoders (AE) [156], or Spectral-Spatial Residual Networks (SSRN) [170].

Generally, a broader electromagnetic range and higher resolution, both spectrally (number of contiguous separable bands) and spatially (pixel density per unit area), result in more accurate identification of materials in a scene. The use of high-dimensional images presents some

challenges [120, 94, 20] related to the sheer volume of data generated and the computational cost of processing them, which makes the development of efficient algorithms and techniques applicable to classification schemes particularly important for the use of increasing amounts of data. The use of spatial-spectral information, together with techniques based on deep learning, presents better results at the cost of greater use of computational resources [29]. In the case of the latter, the huge data requirements necessary to achieve good model generalization pose a problem in a domain where obtaining labeled datasets can be costly due to the need for human intervention in the labeling process. Alongside this, the current trend toward increased model complexity generates an explosion in the number of parameters that has entered the realm of diminishing returns [134].

In order to democratize access to the processing of multi and hyperspectral images and, potentially in the future, to make the transition to mobile computing, it is of particular relevance to obtain algorithms that allow performing these tasks on general-purpose hardware. The previous explanation highlights the need for efficient spatial-spectral information extraction techniques capable of obtaining satisfactory classification results on hardware with reduced resources. For this reason, the following hypotheses are defined as part of this thesis:

- **H1**. The application of preprocessing methods based on the extraction of spatial-spectral information, such as segmentation or the construction of profiles based on morphological operations, attributes, or diffusion, yields an improvement in the classification accuracy of multi and hyperspectral land cover images. This is true for both traditional methods, and those based on deep learning networks.

- **H2**. Efficient algorithms developed for highly parallel architectures such as GPUs can provide significant speedups over their CPU counterparts for multi and hyperspectral image processing.

- **H3**. The preprocessing based on multidimensional segmentation of the images reduces the complexity of training models based on deep learning, achieving significant improvements in execution times on large scenes.

- **H4**. Data augmentation techniques applied to deep learning increase the classification accuracy in situations with a low number of training samples.

From the above hypotheses, the following specific objectives were extracted:

- **O1**. Development of spatial-spectral information extraction techniques for classical classification models. Analysis of the state of the art of multi and hyperspectral LULC image classification and development of new efficient spatial-spectral information extraction techniques.
- **O2**. Development of spatial-spectral information extraction techniques for deep learning-based models. Creation of efficient classification schemes oriented to multi and hyperspectral LULC images of high and very high resolution.
- **O3**. Projection of the developed methods on parallel architectures. Multi-core and GPU architectures will be considered.

Given the requirement of obtaining efficient algorithms oriented to consumer hardware, the proposals developed were implemented in C++, using OpenMP and CUDA for the CPU and GPU versions, respectively. Numerous specific optimizations were applied aimed at maximizing GPU performance, including optimizations dedicated to memory access, optimizations related to the instructions for certain operations, and optimizations aimed at increasing the use of available resources on the GPU through the selection of block sizes. The code for the proposed data augmentation methods was written in Python, using the open source library Tensorflow, due to the dominance of this language in the field of neural networks and machine learning in general.

Experimental validation was performed using three datasets with very different characteristics. A first set of hyperspectral data with images commonly used for benchmarking, of low spatial resolution captured by ROSIS-03 and AVIRIS sensors with vegetation and urban scenes. A second dataset, called *Galicia dataset*, consists of multispectral images of vegetation in the vicinity of river basins, captured in several regions of Galicia using a sensor mounted on a UAV. These images are considerably larger than the first set and have a high spatial resolution, which makes their processing particularly costly. Finally, a multispectral dataset composed of 18 images obtained by the Gaofen-2 satellite with scenes from different regions of China.

The validation of the hypotheses as part of the work of this thesis to achieve the objectives described above resulted in the following contributions:

1. Proposal of a spatial-spectral information extraction method based on the application of nonlinear diffusion. This proposal gave rise to the so-called Extended Anisotropic Diffusion Profiles (EADP), inspired by morphological profiles. The EADP is constructed by

concatenating a series of nonlinear diffusion instances applied to the principal components extracted from a hyperspectral image, generating new components with decreasing levels of detail. The introduction of these components reduces the spectral variability of the spectral signatures of the materials in the scene, increasing the classification performance. Multi-core optimized implementations have been proposed using OpenMP and GPU-optimized implementations using CUDA have been developed using commodity hardware. The parallel processing capabilities of GPUs enabled an improvement in processing times of around 10× thanks to the use of efficient algorithms adapted to these architectures. To conclude, a characterization of diffusion profiles is detailed and their classification performance using an SVM model is compared with other existing methods in the literature such as Extended Morphological Profiles (EMP), Extended Attribute Profiles (EAP), Extended Extinction Profiles (EEP), and so on.

2. A proposed data enhancement method based on the composition of geometric transformations called Dual-Window Superpixel (DWS) data enhancement. This augmentation technique, oriented to deep learning classification schemes based on CNNs, is based on a partitioning of the input data patches to the network in two independent regions on which different conventional geometric transformations are applied. The method assumes that the central zone contains the most relevant data for classification. The input patches can be obtained using a sliding window, processing the image one pixel at a time or, unlike the previous method, be taken from the central pixel of non-overlapping regions of the image. These regions correspond to the superpixels obtained by the SLIC (Simple Linear Iterative Clustering) algorithm [4] and the label of the central pixel will then be assigned to all pixels. The use of this augmentation technique together with superpixel segmentation significantly reduces image processing costs, allowing its use on large images with reasonable run times. This proposal obtains higher overall accuracy than the flip and rotation transforms when these are applied to the inner region as opposed to the entire patch. Additionally, the application of the transforms over both regions independently allows the generation of larger amounts of synthetic data, which increases classification performance. The proposed method was employed in combination with a CNN architecture, where it achieves superior performance compared to other data augmentation methods in the literature.

3. Proposed data augmentation method [125] based on data imputation [141]. The premise consists of the replacement of certain data from the outer zone of the input patches

obtained by using a superpixel segmentation algorithm. These pixels, in areas of irregular boundaries in the scene structures, may belong to classes that are different from the actual class label of the patch and their removal improves the classification accuracy in deep learning schemes based on CNNs. Various data imputation techniques are used for the generation of this synthetic data, resulting in a new augmented patch that is sent to the neural network together with the original patch. The use of superpixel segmentation reduces the image processing costs, as was the case with DWS. The novelty introduced in this work lies in the removal and replacement of pixels from the input patches by using a combination of segmentation information and data imputation, which results in more homogeneous input patches and yields an improved classification accuracy. The results of the proposal achieved the highest accuracy in 16 of the 17 scenes in the Gaofen-2 dataset and constitute the first example of the use of imputation for data augmentation in multi- and hyperspectral imagery.

For the evaluation of the results, several metrics commonly present in the literature were used, such as total accuracy, average accuracy, or Kappa. The computational cost of the algorithms was measured in terms of the execution time, or wall time elapsed between the first stage of an algorithm and the final stage. This evaluation of the results obtained by the different proposals supported the hypotheses that were previously detailed and made it possible to determine the satisfactory fulfillment of the objectives set at the beginning of the thesis. The information extraction methods presented are applicable to a variety of relevant problems in the field of multi and hyperspectral image classification.

As future work, the possibility of applying the EADP to other problems, such as those related to registration, change detection or anomaly detection, is contemplated. Similarly, the data augmentation proposals are not limited to a specific network architecture and their use with other architectures could be studied. Additionally, the inclusion of techniques for class imbalance mitigation could prove to be of interest as a means to improve the classification results.

Thesis outline

The present thesis is divided in 6 chapters. Chapter 1 introduces the theoretical framework and fundamental concepts related to the different topics presented as part of this thesis, as well as the hypothesis, objectives, and methodology used to obtain the experimental results that support the research lines that were developed. These are mainly related to the concepts of information extraction, classification of LULC multi and hyperspectral images, data augmentation applied to DL, as well as any considerations related to computational performance associated to those tasks.

In Chapter 2 presents an in-depth discussion of the main contributions introduced to the field of multi and hyperspectral image classification developed as part of this thesis, mostly referring to the work described in the following chapters, related to spatial-spectral information extraction methods. An overview of their advantages and disadvantages as well as their work principles are introduced as a means to compare the different proposals that are detailed.

Chapter 3 presents a spatial-spectral information extraction method denominated Extended Anisotropic Diffusion Profile (EADP) that is based on the application of anisotropic diffusion to generate and extended profile. It provides a multilevel characterization of the images that has interesting properties for several tasks, including supervised classification.

Chapter 4 presents a new data augmentation scheme based on a combination of superpixel segmentation for patch extraction and geometric transformations called dual-window superpixel (DWS). In DWS, patches are divided into two regions and geometric transformations are applied independently to each of them. The augmented dataset is then sent, along the original patches, to a CNN classifier.

Chapter 5 presents a new data augmentation scheme based on leveraging the properties of segment-based classification in combination with data imputation methods. Undesirable information inside the input patches is removed and then replaced using a data imputation

algorithm. The new dataset, along with the original patches, is then used as the input of a CNN model.

Chapter 6 offers some remarks on the contributions and conclusions reached as part of the work completed for this thesis. These show that it is possible to achieve the goals of developing new spatial-spectral information extraction methods and new data augmentation schemes for multi and hyperspectral image classification with both traditional and DL models in mind. These methods obtain results in line with other proposals from the relevant literature. The use of GPU implementations has also shown significant advantages over CPU implementations, given their parallel data processing capabilities.

Contents

Chapter 1

Introduction

With the advent of new technological advances, the development of specific tools made the process of information capture go from requiring physical contact with the object of interest to not being bound by that restriction and being achievable at a distance. These advances led to what is currently referred to as remote sensing [43].

Remote sensing can be defined as the acquisition of information about an object through the electromagnetic spectrum without requiring physical contact. In this thesis, we are going to refer to remote sensing as the observation of the Earth's surface by means of the reflected or emitted electromagnetic energy [26]. This process is usually performed by specialized sensors mounted on a variety of platforms, including satellites, airplanes, or certain other types of aircraft.

The first satellites that mounted sensors capable of obtaining images from the Earth's surface were launched during the decade of 1960. The decade of 1970 saw a tremendous development of a multitude of different sensors and associated data processing capabilities. In 1972, Landsat I [154], or Earth Resources Technology Satellite (ERTS), was launched and became the first Earth-orbiting satellite created with the specific intent of studying the planet's landmasses, as well as the first multispectral satellite. During the 1980s, the first hyperspectral sensors were developed by NASA's Jet Propulsion Laboratory. Remote sensing reached a global scale in the decade of 1990 and it has been gaining an increasing number of uses in civil applications ever since. Nowadays, a plethora of satellite constellations from different missions provide a wide range of different sensors and capabilities, among which we can find panchromatic sensors (GeoEye-1, Landsat-8, Landsat-9, etc.), multispectral

sensors (Sentinel-2A, Gaofen-1, Gaofen-2, etc.), hyperspectral sensors (PRISMA, Gaofen-5), and other microwave sensors such as SAR (Sentinel-1A, RADARSAT-2). The information captured by these sensors is useful for a wide rage of applications, including atmospheric studies [133], forestry and agriculture [62, 54], natural hazard studies [121, 82], or resource exploration [13], among others.

1.1 Spectral imaging

Numerous different sensors have been developed in the field of remote sensing [136], each of those aimed at capturing data from different regions of the electromagnetic spectrum and with different uses in mind. In this thesis, the focus will be placed on highly dimensional images and, more precisely, multispectral and hyperspectral images. The differentiating characteristics between sensors created with the purpose of acquiring high dimensional images that set them apart from other remote sensing instruments are their high spectral and spatial resolutions. These images have their origin in image spectroscopy [110], which can be defined as the science that studies or examines the spectrum of energy arriving at a sensor in a detailed manner. This energy will depend on the chemical composition of a substance and its interactions with the electromagnetic spectrum.

Spectroscopy can be traced back to the experiments performed by Isaac Newton and his analysis of light. Newton showed that light could be decomposed into a continuous spectrum of colors. Later on, W Herschel and J.W. Ritter proved that the sun's radiation extended beyond the visible range, showing observations that went into the infrared and ultraviolet, respectively. Joseph von Fraunhofer noticed that the light, when sufficiently dispersed, had a large number of black lines, or absorption spectra [105], crossing it. It would be Neils Bohr, however, the scientist responsible for identifying that the origin of the lines was related to the atomic structure of gases. These pioneering works set the basis for the modern quantitative analysis of the electromagnetic spectrum. Similarly to the experiments carried out by these scientists, modern hyperspectral sensors work by leveraging a set of lenses meant to disperse the light into spectral regions for its analysis.

In imaging spectroscopy, each pixel of an image is composed of a vector of bands in which each of its elements represents an intensity value for a given wavelength [56]. This intensity is related to the characteristics of the material [59], with different materials presenting different profiles of emission, absorption, reflectance, fluorescence, etc. for different frequencies.

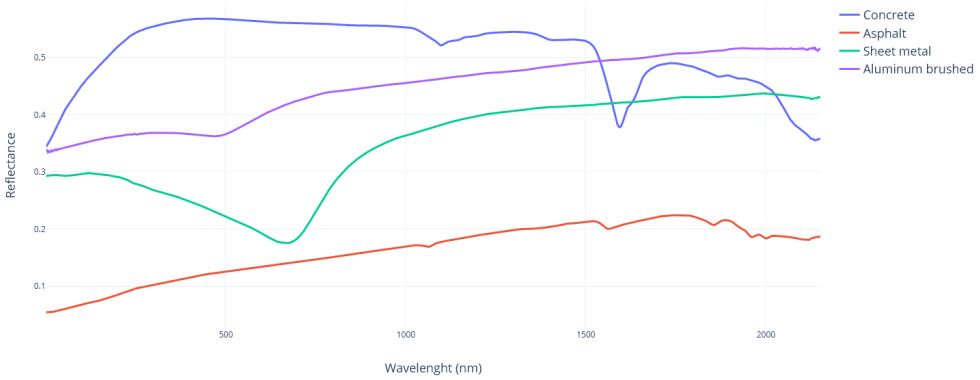


Figure 1.1: Sample spectral signatures of several artificial materials, as provided by (74)

Thanks to this principle, the materials present in a scene can be characterized by the set of intensity values in each of the bands for the pixels where it is located. This characteristic, also called spectral signature, is specific to each material and makes it possible to identify it by studying its similarity to other known material signatures. Figure 1.1 shows example spectral signatures of several artificial materials. In practice, the process of identifying materials by means of their spectral signature is affected by the phenomenon of spectral variability [31], where factors such as illumination, atmospheric effects, or noise in the measurements [123, 21] alter the spectral signatures.

Unlike traditional photography, information captured by multispectral and hyperspectral sensors can include parts of the infrared and ultraviolet spectrum, typically between 0.4µm and 2.45 µm [26], while the number of bands can range from three to just over ten in the case of multispectral images, to tens or hundreds of bands in the case of hyperspectral images. This is also referred to as spectral resolution and is given by the minimum distance between two wavelengths required to distinguish two different lines of the spectrum [99]. Multispectral and hyperspectral images are analogous to a three-dimensional data cube, with two spatial dimensions corresponding to width and height and a spectral dimension consisting of the different spectral bands. Figure 1.2 displays an example of a hyperspectral cube and its major components. The introduction of multispectral and hyperspectral sensors greatly advanced the field of remote sensing, with new concepts and methods that lead to numerous applications stemming from the large amount of information acquired. Among these, the fine-grained spectral information has led to these types of images becoming commonly used in

classification tasks in the field of remote sensing due to the greater separability [67] between the different materials present in a scene when compared to RGB images.

There are numerous sources for hyperspectral images, among which satellite-mounted sensors are the most significant contributors to data acquisition. Examples of these sensors are Hyperion [53], CHRIS (Compact High-Resolution Imaging Spectrometer) [14], PRISMA (PRecursore IperSpettrale Della Missione Applicativa) [92], or Hyperspectral and Multispectral Imager (HISUI) [72]. These sensors have specific technical capabilities that determine parameters such as the spectral range they are able to capture, spatial and spectral resolution, or other low-level parameters. Additionally, in the last years, a multitude of airborne hyperspectral platforms have been developed for commercial use, democratizing access to the technology and greatly boosting its potential applications.

1.2 Land use and land cover classification

During most of the 20th-century urban expansion, agriculture, and other changes caused by human activities have significantly altered the landscape around the world. These phenomena can affect the environment and its natural processes in significant ways. Knowing more about these changes and their impact is paramount to better manage the available resources for human development. Land use and land cover (LULC) maps are two of the most common tools used to characterize the Earth's surface and they are extensively used in many areas such as urban management and planning, monitoring biodiversity, etc. Even though these concepts are related, and often used interchangeably [51], their purpose is different. Land use usually refers to the socioeconomic uses land has, whereas land cover refers to the composition of the Earth's surface when observed directly. With the launch of Landsat in 1972, the availability of remotely sensed data expanded dramatically thanks to an instrument that allowed researchers to obtain images consistently at predictable intervals [42]. These multispectral images accelerated the development of new research areas related to LULC maps, such as data fusion or change detection [52, 95, 35, 115]. Hyperspectral data suffered from a late adoption due to the computational requirements and sensor manufacturing costs associated with it but it has recently acquired a much more prominent role in the Earth Observation community [58].

We can define the classification of images, generically, as a process that tries to assign each pixel or region of the image a precise nominal class label out of a group of mutually exclusive, available labels. This classification process can be performed in several ways. On top of the

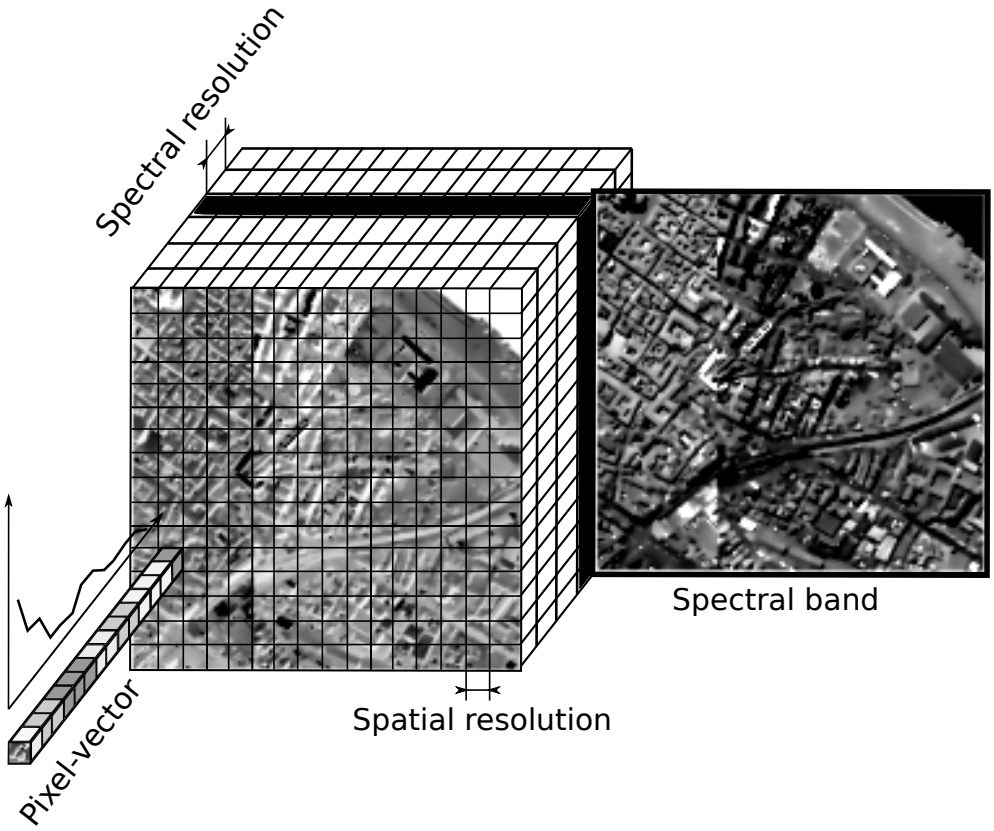


Figure 1.2: Hyperspectral data cube

traditional supervised and unsupervised approaches, semisupervised [150] training has gained traction in recent years associated with Deep Learning (DL) models [93]. In the supervised classification process, pairs of values composed of a sample and a corresponding, manually annotated class label belonging to a subset of data that is considered representative of the original set, are used for training a machine learning (ML) model (for model-based learning) or for inter-sample comparison (for instance-based learning). This information will then be used to predict the class of the rest of the original set. In the unsupervised classification process, pixels are grouped into similar sets without the use of prior information. In this unsupervised process, the remote sensing image is usually partitioned into a number of pixel groups by applying clustering algorithms. In the semisupervised classification process, both

approaches are mixed and information extracted in an unsupervised way is then used to enhance a supervised classification process.

1.2.1 Machine learning in LULC

With the launch of the first satellite-mounted sensors, the possibility of obtaining images from anywhere in the world in a continuous fashion with no costs associated with the displacement to the area of interest made remote sensing truly global. It was at this point that the need to increase the efficiency of the analysis of the gathered data became more evident. ML has had a major role in the advances of remotely sensed data applications geared towards civil uses as a means to automate this task with reduced human intervention. As such, the research interest in the field made it to become one of the most popular topics in the literature [96].

The most basic ML classification pipeline for LULC classification was traditionally comprised of four different stages [140]. These were preprocessing, feature engineering, classification, and post-processing. The preprocessing stage encompasses the initial set of transformations applied to a dataset in order to enhance it. Denoising, data fusion, and the removal of outliers or inconsistent values are performed during this stage. Segmentation is often used as a preprocessing technique in computer vision. It is capable of simplifying images, reducing them to meaningful, independent regions of pixels called segments. These properties make it useful to reduce the complexity of subsequent processing tasks.

The feature engineering step takes care of transforming the data to extract the most relevant features present in it. Feature extraction (FE) is one of the most common feature engineering operations that became increasingly relevant with the expansion in the dimensionality of the spectral information of HSIs and the large number of highly correlated bands in these images. With more available information the number of separable classes that could be classified grew accordingly but also did the requirements regarding the amount of data needed to establish the statistical integrity during classification, causing the algorithms previously applied to multispectral images to start to exhibit the so-called Hughes phenomenon [78]. Also called the curse of dimensionality, this behavior can be observed during hyperspectral image classification in the form of a slow increase in accuracy as the number of bands in an image increases, followed by a sharp fall as the number of bands reaches a certain threshold. Feature extraction algorithms such as Principal Component Analysis (PCA) [89], Independent Component Analysis (ICA) [144] or Singular Spectrum Analysis (SSA) [162] have been successfully applied to HSIs to solve this problem. During the training phase, the machine

learning model of our choice is fed the training data in order to learn to identify the set of class labels from the available features in the data. This model will be later used to predict the class labels for the remaining samples.

Lastly, the postprocessing stage is used to enhance the outputs of our model. This is typically done by applying context information such as spatial boundaries resulting from a segmentation algorithm. One of the most widely used postprocessing methods is the majority voting [77], an approach that is used to determine the class label for a pixel or group of pixels based on the most frequent result obtained from several predictions. This is done to, for example, unify the labels from a segmented region, or to reduce the salt and pepper effect after predicting the labels of a scene.

The simplest classification method would correspond to per-pixel, or spectral-based classification, where the characteristics of the neighborhood around the pixel to be classified are not taken into account. Classification is performed exclusively by relying on the spectral information of the image. Examples of this are classifiers such as random forests (RF) [63, 73, 17] or support vector machines (SVM) [61, 48, 30, 132], where no context information is exploited. There are obvious limitations to this method that may cause the models to yield poor results, such as low separability between classes due to the similarity and variability of the spectral signatures. This is one of the common problems that arise when classifying different vegetation classes [15]. In these situations, information related to vegetation textures can prove critical and provide an important boost to classification accuracy. Another common pitfall affecting classification tasks for low-resolution images is mixed pixels, or pixels with a mixture of spectral information from several different materials. Hyperspectral unmixing [109] is the process that tries to decompose the mixed pixels into constituents and measure their contributions.

A common approach used to increase classification accuracy is to also exploit the spatial information contained in HSIs. Spatial-spectral methods, in which the classification makes use of the inter-dependencies or structural relationships of the pixels in an image, have routinely proved to have many advantages over traditional spectral-based methods and obtain better accuracy [50]. This group includes algorithms such as spatial regularization [165], different morphological profiles [19, 18, 44, 57] or joint learning (JL) [114]. Spatial regularization is based on the application of a model with spatial information to the classification results of spectral data. There are multiple ways to achieve this, either by applying unsupervised methods such as segmentation [91], supervised methods such as Markov fields [155], or conditional

random fields (CRFs) [169]. Mathematical morphology [112] became one of the most common approaches used to extract spatial-spectral information from HSIs. Morphological profiles (MPs), of which there are numerous variants, make use of mathematical morphology and, in particular, opening and closing operations to create multiple versions of the original image with different granularity that are grouped into what is called a morphological profile and then processed by a classifier. Among the most popular variants of morphological profiles we can find attribute profiles (APs) [44], and extinction profiles (EPs) [57]. Finally, joint learning consists of simultaneous learning of spatial and spectral features by using specific kernels. This can be seen, for example, in the use of such kernels in combination with SVM for pixel neighborhood analysis in [49] classification tasks.

Artificial Neural Networks (ANNs) [5] try to mimic the behavior of the biological systems in the animal brain and they have become widely sought after as an information processing paradigm due to their learning and generalization capabilities. They have gained huge popularity in recent years, especially in the computer vision field, due to their ability to handle non-structured data. Convolutional Neural Networks (CNNs) [80], in particular, have been successfully used to solve problems requiring multi-class, multi-label classification involving Feature Extraction (FE) [32] from images. Deep Learning (DL) is a broad concept that can be described as a set of approaches using multiple layers of ANNs to automate the FE process. The adjective ”deep” refers to the number of layers that form a model and can perform transformations over the input data, albeit there is no universally accepted definition of a frontier that determines if a network is deemed shallow or deep. CNNs operate over small cubes of data called patches and can exploit both spatial and spectral information. The patches are centered around a pixel of the image and extracted from a sliding window of a certain size. Usually, one patch is extracted for every pixel of the image using this procedure.

Regardless of the model being used, three goals have to be considered to achieve a successful classification. The model complexity, associated computational cost of training, and classification performance need to be carefully balanced. In general, more complex models can perform better than simple models, at the cost of reduced training efficiency and, thus, higher computational cost. Excessively complex models may exhibit worse performance on unseen data compared to training data in a phenomenon called overfitting. In order to reduce overfitting, an alternative would be to reduce model complexity, which may not be ideal as it can also sacrifice classification performance. A bigger training dataset is usually the desired solution but this may not be feasible in every domain due to the lack of acquired or labeled data.

In these situations, data augmentation can be leveraged to obtain a more generalized model. DL models are notorious for their complexity and data requirements, with large amounts of data needed during the training stage.

Data augmentation synthetically generates new samples by applying a set of domain-specific transformations over the original input data to generate new samples. Numerous proposals taken from image recognition literature have been applied to MSI and HSI images [125], as well as other specific approaches. Among the latter, some notable examples include grouping pixels in blocks and block pairs that serve as the input to a CNN [88, 86], shifting samples along their first principal component based on the average of each band [103], or using generative adversarial networks (GANs) to generate new augmented samples [9, 11, 171].

1.3 Hypothesis and objectives

This thesis attempts to answer the question of whether the application of efficient spatial-spectral information extraction methods is able to successfully improve the performance of both traditional and DL-based classification schemes. This question was formulated in terms of four different hypotheses:

- **H1**. The application of preprocessing methods based on the extraction of spatial-spectral information, such as segmentation or the construction of profiles based on morphological operations, attributes, or diffusion, yields an improvement in the classification accuracy of multi and hyperspectral land cover images. This is true for both traditional methods, and those based on deep learning networks.
- **H2**. Efficient algorithms developed for highly parallel architectures such as GPUs can provide significant speedups over their CPU counterparts for multi and hyperspectral image processing.
- **H3**. The preprocessing based on multidimensional segmentation of the images reduces the complexity of training models based on deep learning, achieving significant improvements in execution times on large scenes.
- **H4**. Data augmentation techniques applied to deep learning increase the classification accuracy in situations with a low number of training samples.

To test the above-mentioned hypotheses, three specific objectives were defined as part of the final goal of the thesis, which can be summarized as the development of efficient spatial-spectral information extraction methods to use in the supervised classification of hyperspectral imagery for land-cover applications. These objectives are as follows:

- **O1**. Development of spatial-spectral information extraction techniques for classical classification models. Analysis of the state of the art of multi and hyperspectral LULC image classification and development of new efficient spatial-spectral information extraction techniques.
- **O2**. Development of spatial-spectral information extraction techniques for deep learning-based models. Creation of efficient classification schemes oriented to multi and hyperspectral LULC images of high and very high resolution.

- **O3**. Projection of the developed methods on parallel architectures. Multi-core and GPU architectures will be considered.

The validation of the hypotheses as part of the work of this thesis in pursuit of the objectives described above has resulted in the following contributions:

1. Proposal of a spatial-spectral information extraction method based on the application of nonlinear diffusion. This proposal gave rise to the so-called Extended Anisotropic Diffusion Profiles (EADP), inspired by morphological profiles. The EADP is constructed by concatenating a series of nonlinear diffusion instances applied to the principal components extracted from a hyperspectral image, generating new components with decreasing levels of detail. The introduction of these components reduces the spectral variability of the spectral signatures of the materials in the scene, increasing the classification performance. Multi-core optimized implementations have been proposed using OpenMP and GPU-optimized implementations using CUDA have been developed using commodity hardware. The parallel processing capabilities of GPUs enabled an improvement in processing times of around 10× thanks to the use of efficient algorithms adapted to these architectures. To conclude, a characterization of diffusion profiles is detailed and their classification performance using an SVM model is compared with other existing methods in the literature such as Extended Morphological Profiles (EMP), Extended Attribute Profiles (EAP), Extended Extinction Profiles (EEP), and so on.

2. A proposed data enhancement method based on the composition of geometric transformations called Dual-Window Superpixel (DWS) data enhancement. This augmentation technique, oriented to deep learning classification schemes based on CNNs, is based on a partitioning of the input data patches to the network in two independent regions on which different conventional geometric transformations are applied. The method assumes that the central zone contains the most relevant data for classification. The input patches can be obtained using a sliding window, processing the image one pixel at a time or, unlike the previous method, be taken from the central pixel of non-overlapping regions of the image. These regions correspond to the superpixels obtained by the SLIC (Simple Linear Iterative Clustering) algorithm [4] and the label of the central pixel will then be assigned to all pixels. The use of this augmentation technique together with superpixel segmentation significantly reduces image processing costs, allowing its use on large images with reasonable run times. This proposal obtains higher overall

accuracy than the flip and rotation transforms when these are applied to the inner region as opposed to the entire patch. Additionally, the application of the transforms over both regions independently allows the generation of larger amounts of synthetic data, which increases classification performance. The proposed method was employed in combination with a CNN architecture, where it achieves superior performance compared to other data augmentation methods in the literature.

3. Proposed data augmentation method [125] based on data imputation [141]. The premise consists of the replacement of certain data from the outer zone of the input patches obtained by using a superpixel segmentation algorithm. These pixels, in areas of irregular boundaries in the scene structures, may belong to classes that are different from the actual class label of the patch and their removal improves the classification accuracy in deep learning schemes based on CNNs. Various data imputation techniques are used for the generation of this synthetic data, resulting in a new augmented patch that is sent to the neural network together with the original patch. The use of superpixel segmentation reduces the image processing costs, as was the case with DWS. The novelty introduced in this work lies in the removal and replacement of pixels from the input patches by using a combination of segmentation information and data imputation, which results in more homogeneous input patches and yields an improved classification accuracy. The results of the proposal achieved the highest accuracy in 16 of the 17 scenes in the Gaofen-2 dataset and constitute the first example of the use of imputation for data augmentation in multi- and hyperspectral imagery.

1.4 Methodological tools

The objectives of this thesis include the creation of efficient versions of the spatial-spectral information extraction algorithms for their execution in consumer-grade multi-core systems and GPUs. Given the task, languages meant for system programming capable of producing high-performance executable code were favored over other alternatives. C++ was the main language used to write the implementations of the algorithms resulting from this work, along with OpenMP for the multi-core CPU versions, and CUDA for the GPU implementations. Python has become the *de facto* language for experimentation with neural networks, with most frameworks using it as the main language. As such, it was the language chosen for the implementation of the deep learning proposals.

1.4.1 CPU parallel programming

The early 2000s marked the end of a CPU innovation model based mainly on the increase of CPU frequencies due to the physical limitations related to the reduction in the size of transistors. Despite Moore's law [122] remaining valid, the increase in the number of transistors started to yield diminishing returns due to thermal constraints. The arguable engineering failure of the Pentium 4 favored the end of the single-threaded computing era and the release of the Pentium D, which consisted of two single-core dies on the same packaging, signaled the start of a new paradigm in computing where increases in CPU performance would rely heavily on the number of cores of a system. Unlike early multiprocessor systems, these new multi-core processors conformed to a Uniform Memory Access (UMA) [10] parallel architecture, where all the cores have access to the same physical memory. In particular, the Symmetric Multiprocessor (SMP) [139] architecture, where identical chips are linked sharing the same memory, became the most popular architecture for commercial processors. These architectures are able to run single-threaded software at nearly identical performance as a single-core processor but offer extra cores capable of running multiple tasks in parallel. The use of these cores, however, requires specific software support at multiple levels in order to take advantage of the extra resources. This caused the development of parallel processing paradigms and tools to simplify software development.

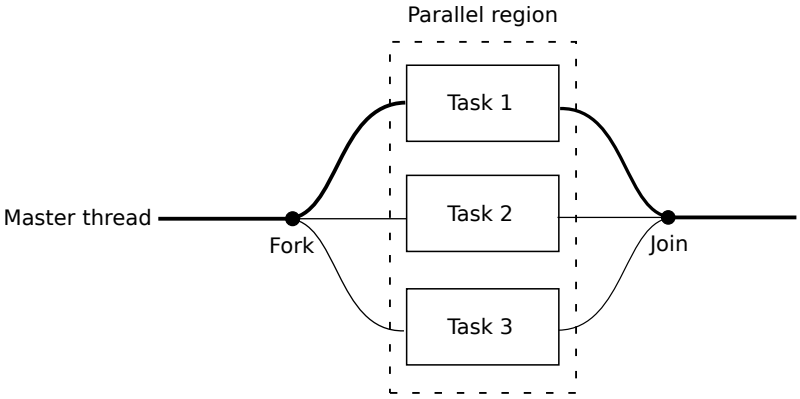


Figure 1.3: OpenMP's fork-join model

OpenMP

OpenMP [22] is a portable application programming interface (API) created to ease the development of parallel software targeting SMP architectures, available in several programming languages and platforms. OpenMP consists of a set of compiler directives, run-time libraries, and environment variables that can be used to implement explicit parallelism in sequential software via task or data parallelism constructs.

OpenMP uses a fork-join [98] model of parallel execution. In an OpenMP program, the main thread of a program is considered the master thread. This thread exists during the lifetime of the program and is responsible for spawning new slave threads for their use inside the parallel regions. When the master thread reaches a parallel construct, it will create a team composed of itself and an additional number of slave threads. These slave threads will all execute the same code in parallel and will join the master at the end of the parallel region. The master thread will then exit the region and continue the execution normally. Figure 1.3 shows a diagram illustrating the fork-join model of OpenMP.

OpenMP offers several directives to control the behavior of the parallel blocks of code, including determining how to parallelize regions, control the scope of variables, distribute the load, synchronize access, or even vectorize operations. The maximum speedup or performance gain, achievable by OpenMP is related to Ahmdal's law, which states that the overall performance improvement gained by optimizing a single part of a system is limited by the fraction of time that the improved part is used. Or, in other words, the performance improvement will depend on the fraction of parallelized code in a program, and the amount of time that

code is used compared to the total execution time. Additionally, a certain level of overhead exists when using parallel constructs. This overhead includes the creation of new threads, synchronization features, or data access. A careful understanding of the code and the correct use of these features will increase the speedup obtained by OpenMP.

1.4.2 GPU parallel programming

General purpose GPU (GPGPU) has emerged as one of the strongest trends of the modern era in computing. This paradigm, where GPU devices are used as general-purpose co-processors to perform massively parallel work, is not something new. Even though GPUs have traditionally been hardware specialized in graphics processing, barring some exceptions, early works [69, 79] from some researchers already highlighted the possibilities for GPUS from the late 90s to perform general computations, albeit with a very restricted programming model. This changed with the introduction of Compute Unified Device Architecture (CUDA) [36] in 2007, simplifying the programmer's work and greatly reducing the entry barrier to GPGPU programming.

GPUs constitute an example of a manycore architecture [23], or architecture with a very high number of cores. Although GPUs contain a much higher number of cores than even the most recent CPUs, these are significantly different and also have different purposes. CPU cores are large and contain many complex circuits tuned to achieve high throughput in sequential programs. GPU cores are relatively lightweight, with little logic, and geared towards achieving high throughput in parallel programs. The domain-specific trade-offs from GPU design made these more suited for exploiting massive data parallelism compared to traditional CPUs, more focused on performance for serial workloads. This property has made GPUs popular in high-performance computing applications. Since the last decade, computers have progressively switched to using other processing elements, with GPU being the most prevalent co-processor. These systems are built to exploit each co-processor to perform the task they are the most suited to.

The flexibility, cost, and throughput for parallel workloads have made GPUs the computing platform of choice for a wide range of tasks in the remote sensing field, including hyperspectral image processing [34, 161]. The literature contains several examples of the use of GPUs for classification [152], registration [106], segmentation [118], etc.

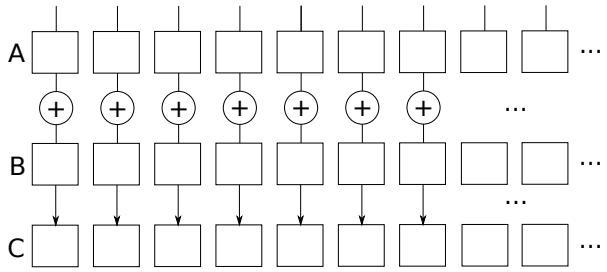


Figure 1.4: Basic vector addition implementation in a GPGPU platform. Each thread operates on a different element of the vector.

CUDA

CUDA is a parallel computing platform and programming model developed by NVIDIA and available through a variety of programming languages, including C, C++, and Fortran. CUDA provides specific tools to allow developers to perform general computing on GPUs. GPU architectures are geared towards a set of specific tasks that have very specific characteristics. Among those, we can list the large computational requirements, with thousands of operations that must be performed over several millions of elements before an image can be rendered; high data parallelism and the greater importance of throughput over latency [107].

GPUs use the single-program multiple-data (SPMD) programming model, a subcategory of multiple-instruction multiple-data (MIMD) where tasks are split in a way that causes multiple elements to be processed in parallel by the same code running in multiple processing elements simultaneously. GPGPU programs are usually divided into blocks of threads, where grids of threads of arbitrary sizes operate. Inside those grids, blocks of threads execute the same SPMD program and compute an individual result each, as shown in Figure 1.4. This result is obtained by reading data from global memory, performing a set of computations using a relatively rich set of instructions, and storing the value back into global memory. CUDA's architecture also follows this model, mapped to CUDA-specific terminology.

The Streaming Multiprocessor (SM) constitutes the basic building block of the CUDA architecture. These are small processors composed of a certain number of execution cores, different caches, thread schedulers, and registers. Figure 1.5 shows the elements inside a SM from a NVIDIA GTX 1080. SMs can execute groups of threads in parallel, called thread blocks, which are, in turn, organized in grids. This is illustrated in Figure 1.6. Thread block sizes have to be defined by the programmer and tuning them based on the hardware's

Table 1.1: CPU hardware specification

CPU model	# of cores	# of cores	Core clock (MHz)	RAM (GB)	L1 (KB)	L2 (KB)	L3 (KB)
Intel Core i5 8400	6	6	2800	32/48	384	1536	9216

characteristics and compute capability is required to maximize its utilization.

Thread blocks are scheduled into SMs by the hardware, and threads from the same block must reside in the same SM. Thread blocks are isolated from each other and can only communicate using the device memory, or global memory. All the threads inside a block run the same program, called a kernel. Thread blocks are split into groups of 32 threads called warps at the hardware level, with all of them being issued an instruction simultaneously in a programming model denominated single-instruction multiple-thread (SIMT). Warps are assigned to warp schedulers inside the SMs, capable of stopping and swapping between different warps based on their execution state. When a warp is not ready, a different warp is scheduled and run in a strategy aimed at hiding hardware latency.

1.4.3 Hardware

The algorithms detailed in this thesis were developed using several consumer-grade computing platforms. With the increase in the capabilities of modern CPUs and GPUs, tasks that were previously relegated to supercomputers are now manageable in commodity hardware. The algorithms developed targeted mid-range devices and can run efficiently in the platforms developed in this section. These, however, should still perform adequately in similar devices as long as the required feature set is supported.

The test system used for this thesis was a custom-built computer with an Intel Core i5 8400 CPU at its core. Two GPGPU devices were used in the chosen build, the GeForce GTX 1060 6GB (Pascal [37]), and GeForce GTX 3070 Ti (Turing [40]). Tables 1.1, and 1.2 show the hardware specifications for the aforementioned CPU and GPUs, respectively.

1.4.4 Software

The algorithms developed as part of this thesis were developed in C++ and Python [142]. All experiments were run under a 64-bit Ubuntu Linux distribution.

Figure 1.5: GP104 SM diagram (38)

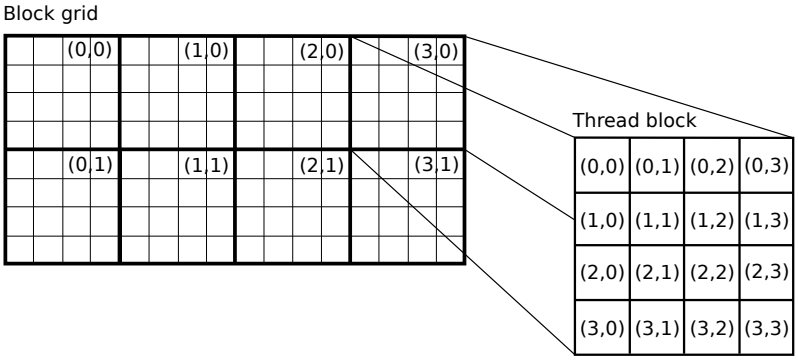


Figure 1.6: Illustration of a thread grid and one of its thread blocks in CUDA

Table 1.2: GPU hardware specification

	GeForce GTX 1060 6GB	GeForce GTX 3070 Ti
Architecture	Pascal	Turing
Compute Capability	6.1	8.6
Core clock	1506 MHz	1575
Memory clock	2002 MHz	1188
Streaming Multiprocessors (SMs)	10	48
CUDA cores per SM	128	128
Tensor cores per SM	-	4
RT cores per SM	-	1
Device memory	6 GB	8 GB
Shared memory	96 KB	100 KB
L1 cache	48 KB	128 KB
L2 cache	1536 KB	4096 KB
Threads per warp	32	32
Maximum thread block size	1024	1024
Maximum threads per SM	2048	1536
Maximum blocks per SM	32	16
Max warps per SM	64	48

C++ codes were built using GNU GCC [1] using the highest optimization level available -O3. This also applies to the CUDA code, built using NVCC. The code targeting CUDA relies on several high-performance libraries that are part of the official CUDA Toolkit [41].

Classification schemes based on DL were written mostly in Python, with some wrappers around C++ code in specific cases. The neural network models were implemented using Google's TensorFlow [2]. The official Tensorflow Docker images with GPU support were used to perform the training and prediction, of the neural networks.

1.4.5 Datasets

The schemes developed as part of this thesis were tested in three image datasets comprising both hyperspectral and multispectral images, and vastly different spatial and spectral resolutions.

The first dataset is composed of widely available hyperspectral benchmarking images obtained by the ROSIS-03 and AVIRIS sensors. Figures 1.7 and 1.8 display the false-color composite and reference data for the scenes.

1. Houston University (Houston) [3]: The aerial view of Houston University was obtained by the CASI sensor. The image has a spatial resolution of 2.5 meters/pixel and it covers a spectral range from 380 to 1050 nm. The size of the image is 1905 x 349 pixels and 144 spectral bands. The reference information classification contains fifteen classes.

2. Indian Pines (IndianP) [70]: The mixed vegetation area of IndianP was obtained by NASA's AVIRIS sensor over the Indian Pines test site in North-western Indiana. The spatial resolution is 20 meters/pixel and it covers a spectral range from 400 to 2500 nm. IndianP dataset consists of 145 x 145 pixels and 220 spectral bands. The ground truth is divided into sixteen classes.

3. Salinas valley (Salinas): Mixed vegetation scene in California. It was obtained by the NASA AVIRIS sensor with a spatial resolution of 3.7 m/pixel, covering a spectral range from 400 to 2500 nm. The image is 512×217 pixels and has 220 spectral bands. The reference information for classification contains sixteen classes.

4. Pavia University (PaviaU) [71]: acquired by the ROSIS-03 sensor over the city of Pavia, Italy. Its spatial resolution is 2.6 meters/pixel and covers the spectral range from 430 to 860 nm. PaviaU consists of 610 x 340 pixels and 103 spectral bands. The ground truth contains nine classes.

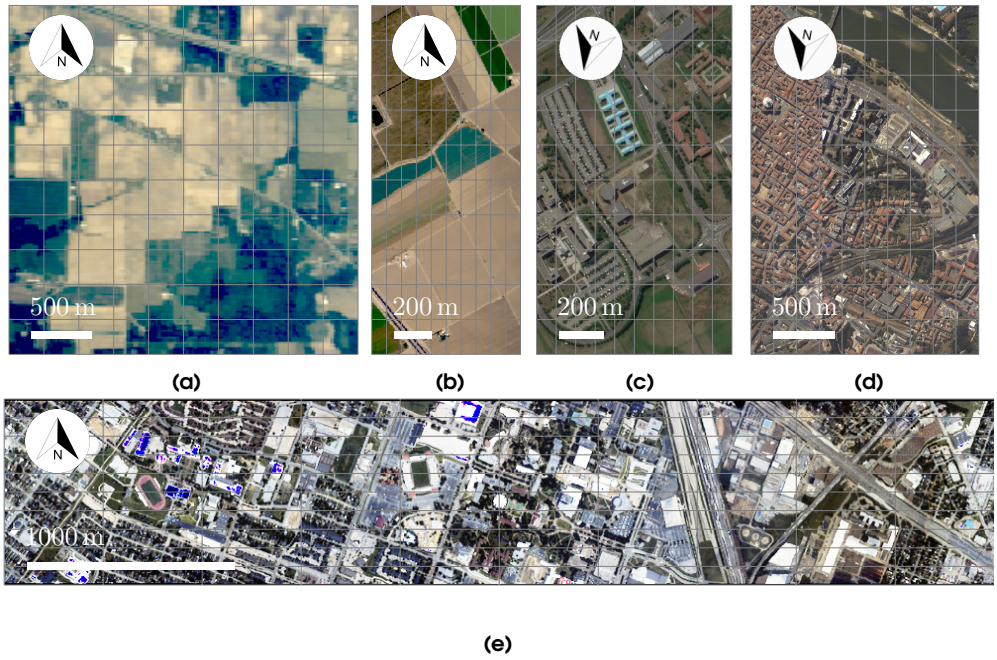


Figure 1.7: False color composite for the selected images from the stardard dataset: (**a**) IndianP, (**b**) Salinas, (**c**) PaviaU, (**d**) PaviaC, and (e) Houston.

5. Pavia University (PaviaU) [71]: Urban scene acquired by the ROSIS-03 sensor over the city of Pavia, Italy. Its spatial resolution is 2.6 m/pixel and covers a spectral range from 430 to 860 nm. The image is 610×340 pixels and has 103 spectral bands. The ground truth contains nine classes.

6. Pavia Centre (PaviaC): Urban scene acquired by the ROSIS-03 sensor over the city of Pavia, Italy. Its spatial resolution is 2.6 m/pixel and covers a spectral range from 430 to 860 nm. The image is 1096×715 pixels and has 103 spectral bands. The reference information for classification contains nine classes.

The second dataset, also referred to as Galicia dataset [15] has four scenes from river basins captured at an altitude of 120 m by a UAV mounting a MicaSense RedEdge multispectral camera [101]. The sensor resolution is 8.2 cm/pixel and it covers a spectral range from 475 to 840 nm, with blue (475 nm), green (560 nm), red (668 nm), red-edge (717 nm), and NIR (840 nm) channels available. The flights were conducted during the summer months of 2018,

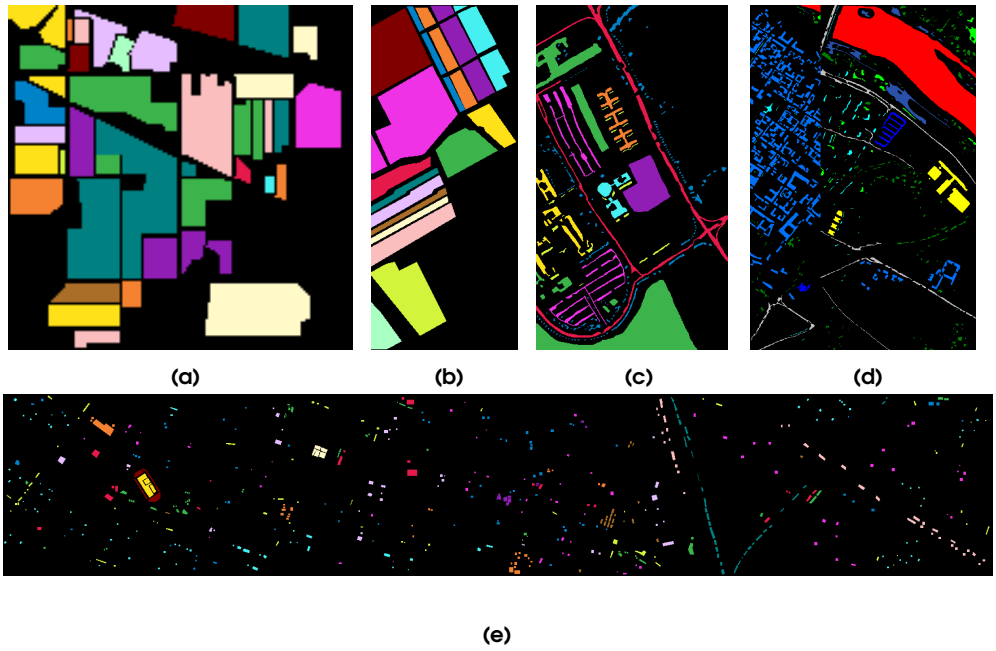


Figure 1.8: Reference data for the selected images from the stardard dataset: (**a**) Salinas, (**b**) PaviaU and (**c**) PaviaC.

2019, and 2020. Registration, geometric and light corrections were performed using the Pix4D software suite. Figures 1.9 and 1.10 display the false-color composite and reference data for the scenes.

1. River Oitavén (Oitaven): Multispectral vegetation scene of the Oitaven river from Pontevedra, Spain. The image is 6689×6722 pixels and has 5 spectral bands. The ground truth contains eleven classes.

2. Creek Ermidas (Ermidas): Multispectral vegetation scene showing the point where Creek Ermidas and River Oitavén meet, from Pontevedra, Spain. The image is 11,924 $\times$ 18,972 pixels and has 5 spectral bands. The ground truth contains ten classes.

3. Eiras Dam (Eiras): Multispectral vegetation scene showing the reservoir that supplies running water to the town of Vigo from Pontevedra, Spain. The scene is $5176 \times 18{,}224$ pixels and has 5 spectral bands. The ground truth contains ten classes.

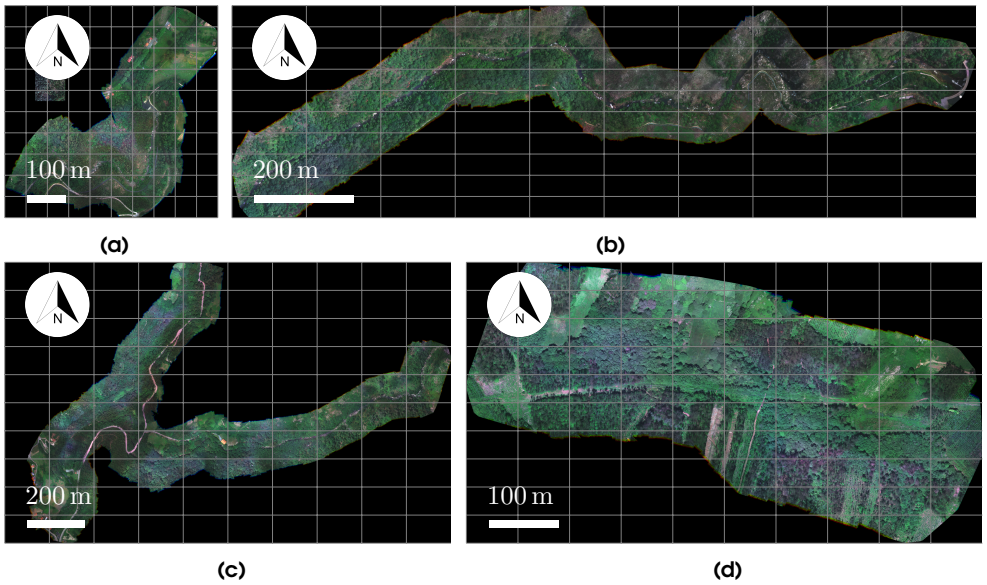


Figure 1.9: False color composite for images from the Galicia dataset: (**a**) Oitaven, (**b**) Eiras, (**c**) Ermidas and (**d**) Mestas.

4. River Mestas (Mestas): Multispectral vegetation scene showing the River Mestas from Pontevedra, Spain. The image is 4915×9040 pixels and has 5 spectral bands. The ground truth contains four classes.

The last dataset contains 17 images from the GID [135]. These images were obtained by the Gaofen-2 (GF-2) satellite of the China National Space Administration. The GF-2 multispectral sensor has a resolution of 4 m per pixel and covers a spectral range from 450 to 890 nm. divided into two groups. The first group contains the first seven images from the 5-class large-scale classification set. The second group contains ten images from the 15-class large-scale classification set. For simplicity's sake, we will refer to the images in the first group as GF2-5A to GF2-5G, and the images of the second group as GF2-15A to GF2-15J. Figures 1.13 and 1.14 display the false color composite and reference data for the images of the first group. Figures 1.11 and 1.12 display the false-color composite and reference data for the scenes of the second group.

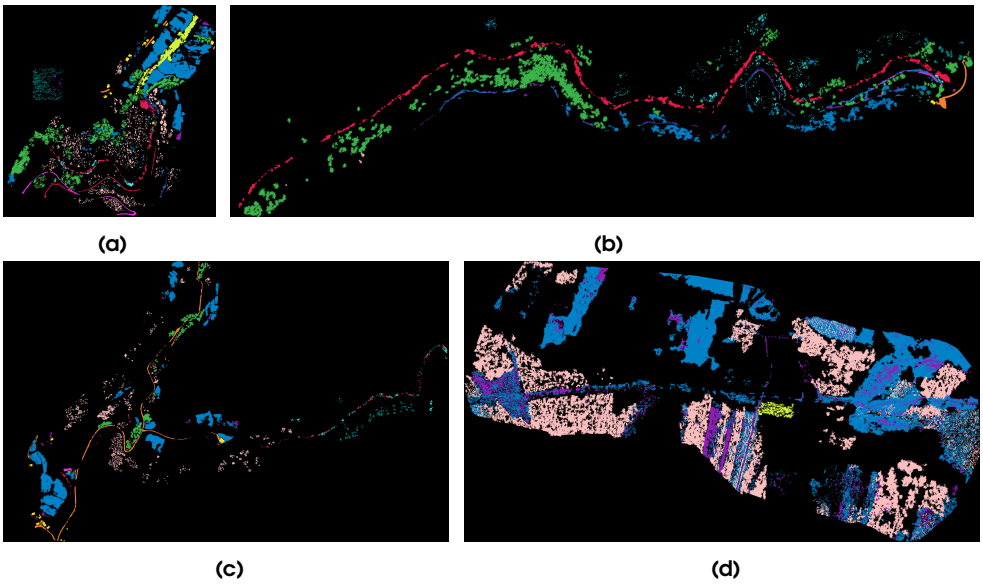


Figure 1.10: Ground truth for images from the Galicia dataset: (**a**) Oitaven, (**b**) Eiras, (**c**) Ermidas and (**d**) Mestas.

1.4.6 Performance measures

The present section describes the performance measures used in this thesis to characterize the performance of the different proposals in terms of classification performance and computational cost.

The supervised classification process, as previously defined, attempts to assign each pixel or region of the image a precise nominal class label out of a group of mutually exclusive, available labels. These predicted pixels then have to be compared against the original set of labels assigned to them in the reference data to determine the classification performance. This is done through the application of different metrics, each of which tries to quantify the achieved results under a specific set of constraints:

- Overall Accuracy (OA) [65]: Percentage of overall pixels correctly predicted.

$$OA = \frac{\sum_{i=1}^{k} P_{i,i}}{\sum_{i=1}^{k} T_i}, \tag{1.1}$$

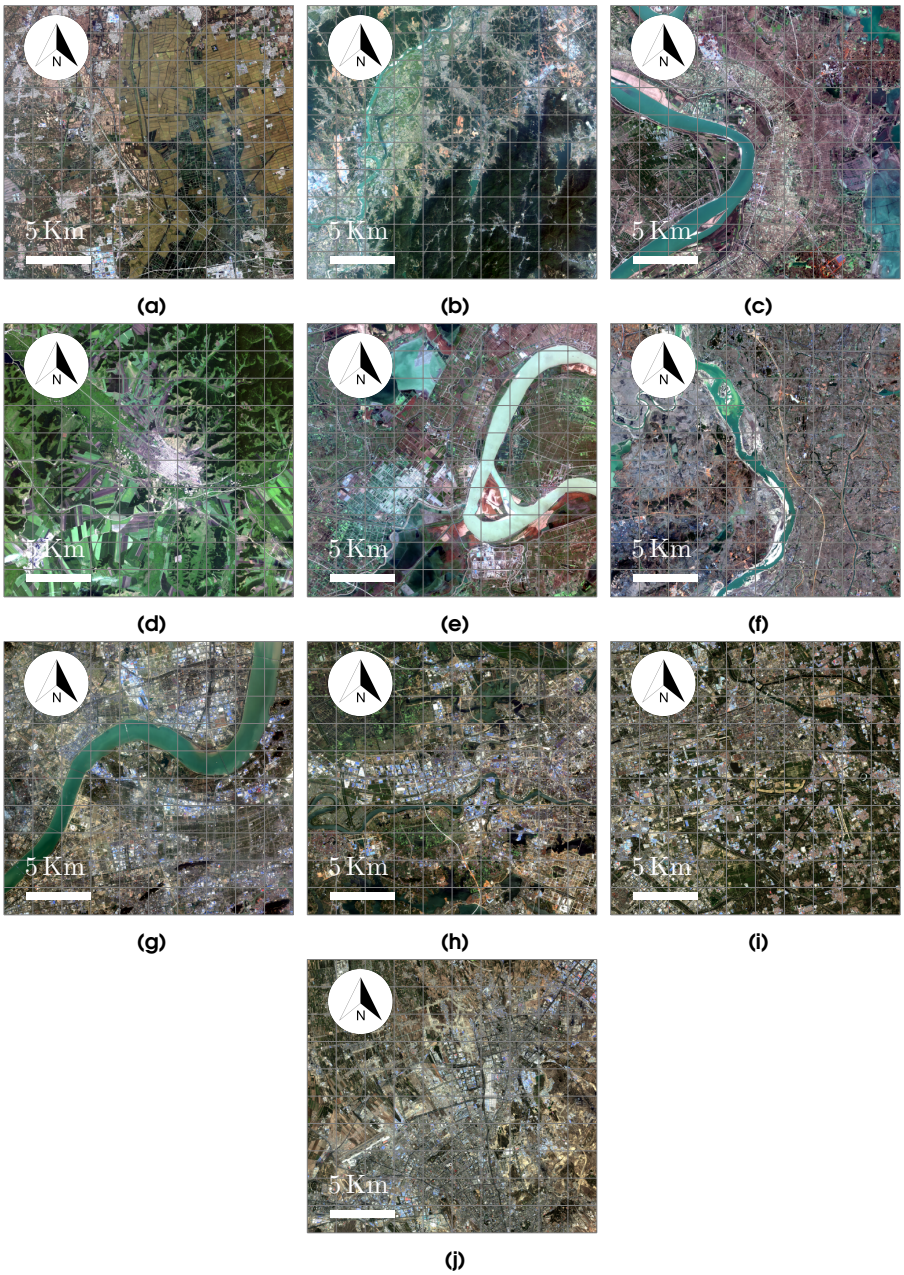


Figure 1.11: False color composite for scenes from the GID: (**a**) GF2-15A, (**b**) GF2-15B, (**c**) GF2-15C, (**d**) GF2-15D, (**e**) GF2-15E, (**f**) GF2-15F, (**g**) GF2-15G, (**h**) GF2-15H, (**i**) GF2-15I, and (**j**) GF2-15J.

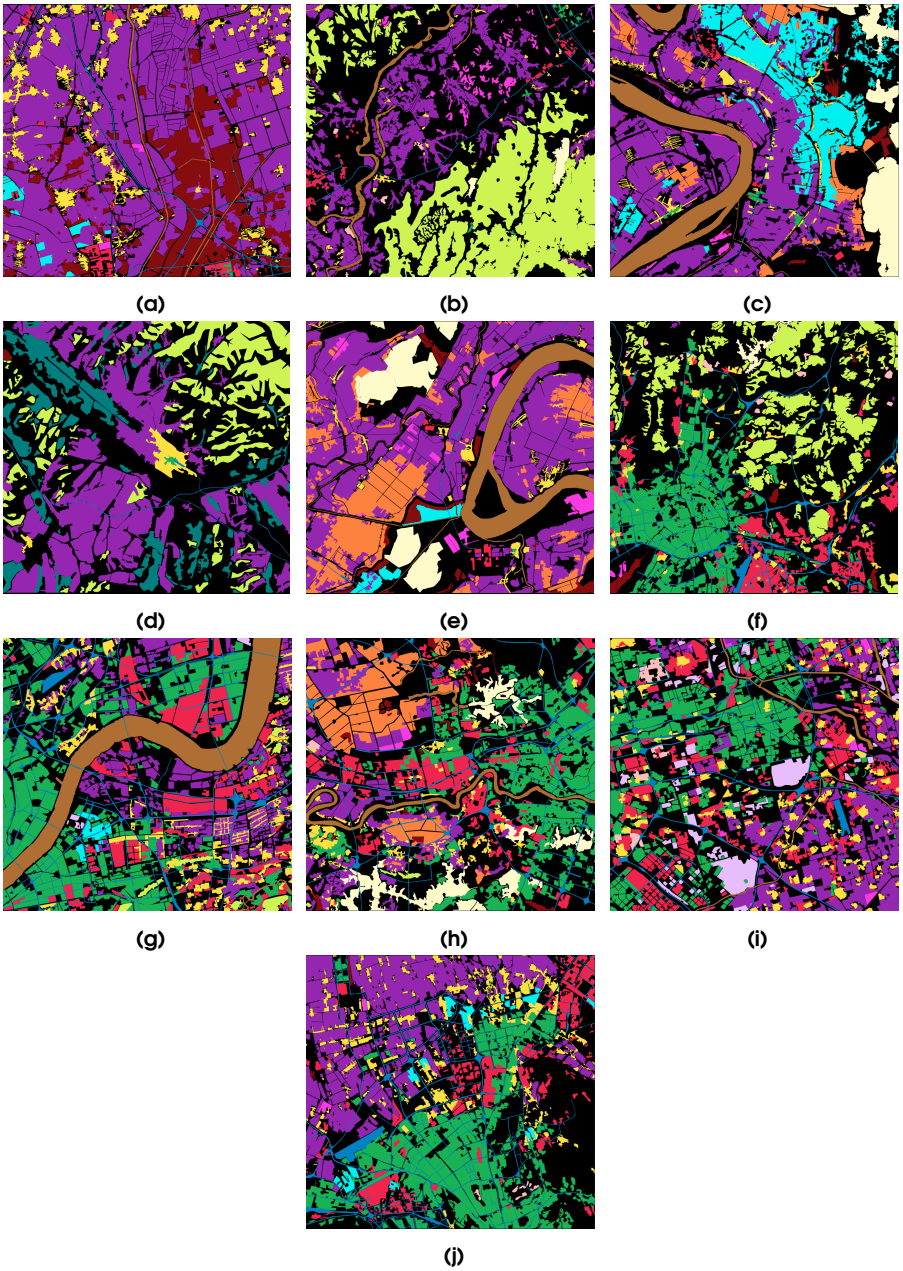


Figure 1.12: Ground truth for scenes from the GID: (**a**) GF2-15A, (**b**) GF2-15B, (**c**) GF2-15C, (**d**) GF2-15D, (**e**) GF2-15E, (**f**) GF2-15F, (**g**) GF2-15G, (**h**) GF2-15H, (**i**) GF2-15I, and (**j**) GF2-15J.

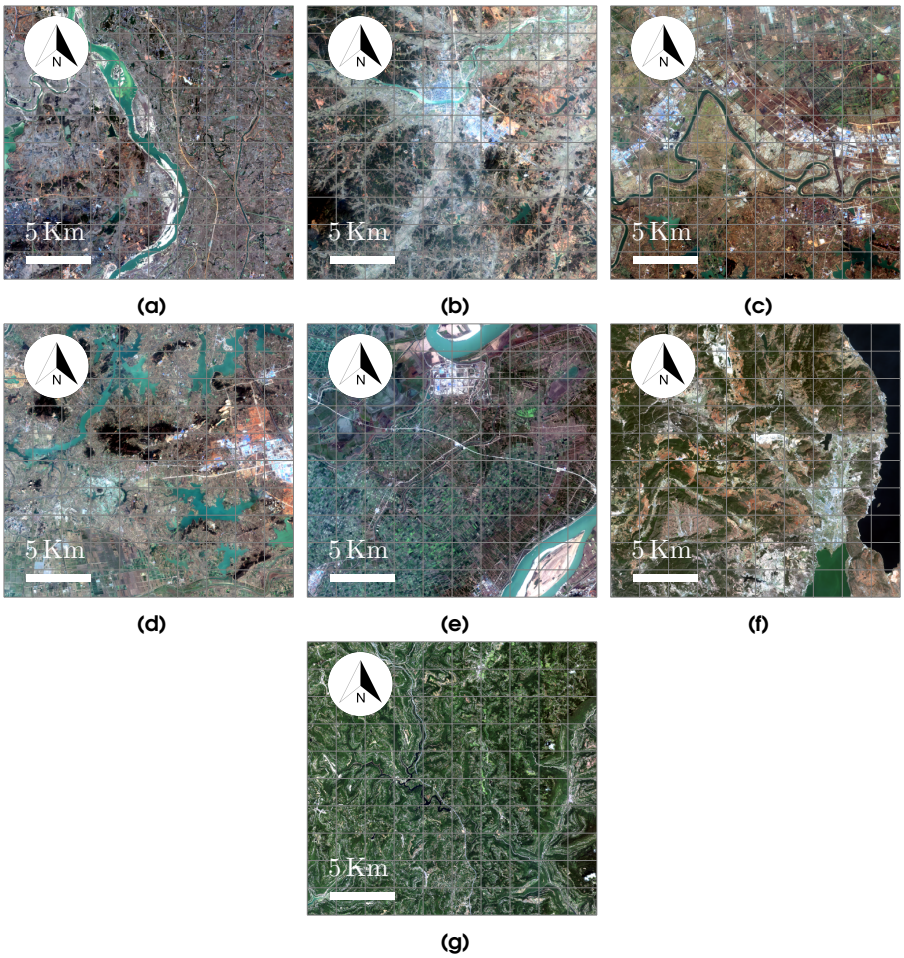


Figure 1.13: False color composite for scenes from the GID: (**a**) GF2-5A, (**b**) GF2-5B, (**c**) GF2-5C, (**d**) GF2-5D, (**e**) GF2-5E, (**f**) GF2-5F, and (**g**) GF2-5G.

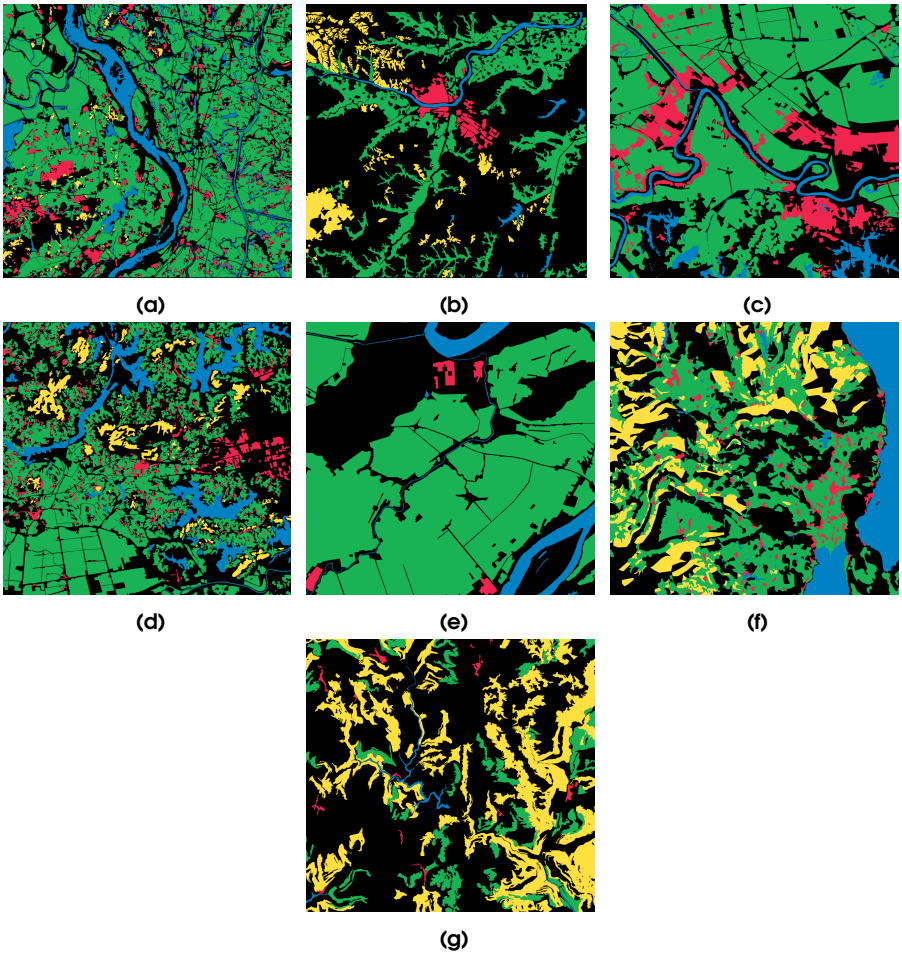


Figure 1.14: Ground truth for scenes from the GID: (**a**) GF2-5A, (**b**) GF2-5B, (**c**) GF2-5C, (**d**) GF2-5D, (**e**) GF2-5E, (**f**) GF2-5F, and (**g**) GF2-5G.

where k represents the number of class labels, $P_{i,i}$ corresponds to the correctly predicted pixels belonging to class i, and T_i the number of truth values from class i.

- Average Accuracy (AA) [65]: Mean of correctly predicted pixels per class.

  where k represents the number of class labels, $P_{i,i}$ corresponds to the correctly predicted pixels belonging to class i, and T_i the number of truth values from class i.

$$OA = \sum_{i=1}^{k} \frac{P_{i,i}}{T_i k}, \tag{1.2}$$

- Kappa (κ) [119]: Agreement between pixel predictions considering occurrences attributed to chance.

$$\kappa = \frac{N \sum_{i=1}^{n} P_{i,i} - \sum_{i=1}^{n} (P_i T_i)}{N^2 - \sum_{i=1}^{n} (P_i T_i)}, \tag{1.3}$$

  where N represents the number of predicted pixels, k is the number of class labels, $P_{i,i}$ corresponds to the correctly predicted pixels belonging to class i, P_i is the number of predicted pixels belonging to class i , and T_i is the number of truth values from class i.

The computational cost of the proposals from the present thesis is measured in terms of execution time. This execution time is defined as the elapsed wall time from the start of the computations of the first stage of a scheme until the completion of the last stage. Execution times do not take into account the initial read time of the data and the final write of the result, as it is assumed those are readily available in RAM. CPU and GPU implementation performance is measured in terms of achieved speedup, or CPU execution time relative to GPU execution time.

1.5 Publications

The following is the list of publications that originated from the work of this thesis:

1.5.1 International journals

Acción, Álvaro, Francisco Argüello, and Dora B. Heras. "Extended anisotropic diffusion profiles in GPU for hyperspectral imagery." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12, no. 12 (2019): 4964-4976.

- Impact factor (JCR 2019): 3.827. Q1
- Category:
  - Remote Sensing. Rank 11/30.

The author of the thesis is the main contributor to the article. This includes the scientific research, comprising the characterization of anisotropic diffusion and possible application to the problem; creation of the scientific software used to generate the anisotropic diffusion profile, process and analyze the result; validation of the research methodology; experimentation and selection of optimal parameters, and writing of the final article.

Acción, Álvaro, Francisco Argüello, and Dora B. Heras. "Dual-window superpixel data augmentation for hyperspectral image classification." Applied Sciences 10, no. 24 (2020): 8833.

- Impact factor (JCR 2020): 2.679. Q2.
- Category:
  - Engineering, Multidisciplinary. Rank 38/90.

The author of the thesis is the main contributor to the article. This includes the conceptualization of the dual-window superpixel data augmentation; data pre-processing; creation of the scientific software, both for the classification of the input scenes and the subsequent analysis; validation of the research methodology; experimentation with different CNN architectures to determine whether the proposed data augmentation framework was suitable for its application, and writing of the final article.

Acción, Álvaro, Francisco Argüello, and Dora B. Heras. "A new multispectral data augmentation technique based on data imputation." Remote Sensing 13, no. 23 (2021): 4875.

- Impact factor (JCR 2020): 4.848. Q1.
- Category:
  - Remote Sensing. Rank 12/32.

The author of the thesis is the main contributor to the article. This includes the scientific research related to the potential use of data imputation as a data augmentation mechanism to increase the homogeneity of input patches, the writing and adaptation of existing software in order to apply this method to a CNN classification pipeline, experimentation with different data imputation algorithms and pipeline variations, and writing of the final article.

1.5.2 International conferences

Montes, Álvaro Acción, Dora B. Heras, and Francisco Argüello. "A new data augmentation technique for the CNN-based classification of hyperspectral imagery." Image and Signal Processing for Remote Sensing XXVII. Vol. 11862. SPIE, 2021.

1.5.3 National conferences

Montes, Álvaro Acción, Dora B. Heras, and Francisco Argüello. "Una nueva técnica de aumentado de datos para la clasificación de imágenes hiperespectrales mediante CNN". Actas de las jornadas SARTECO 20/21, 49–55.
Montes, Álvaro Acción, Dora B. Heras, and Francisco Argüello. "Computación eficiente de perfiles de difusión para la extracción de información espectral-espacial". Actas de las jornadas SARTECO 2019, 6–11.

CHAPTER 2

DISCUSSION

The literature on multi and hyperspectral classification is relatively mature and dates back to the launch of the first widely available spectral, satellite-mounted sensors. The approaches used to perform this task have varied with time, with recent advances focusing on the extraction of spatial-spectral information using traditional shallow models, and, more recently, deep learning [7]. These two advances have been, arguably, the most important breakthroughs of the last two decades.

The first pixel-based classification methods developed rely on the application of shallow classifiers to the spectral information of the scenes [61, 100]. This information has certain characteristics that pose several challenges during the training of the classifiers. First, pixels are processed as isolated entities, with no information about their surroundings. The information contained in a single pixel may belong to several different materials, increasing the spectral variability of the materials and reducing their separability. Different acquisition conditions may also cause an additional issue, given that the light captured by the sensor may come from several sources and originate from the reflection and scattering from the surroundings [76], or be affected by sensor noise [163]. A second limitation of spectral classifiers is given by the high requirements in terms of training samples due to the high-dimensional spectra and the large number of free parameters of the models, resulting in overfitting and reduced generalization capability. This made the classification of pixels based exclusively on spectral information to be limited in terms of performance, which manifested itself strongly in certain scenes, such as those involving different vegetation species. Class unbalances [84], a fundamental problem in these scenes, further compound this issue due to the large difference between the most

common class labels and the least common ones.

Given the nature of multi and hyperspectral images, the spatial dependencies generated by the structures present in a scene can also be used as part of the classification process [65]. This spatial dependency can be one of two kinds. The first one is related to the neighboring relationship between different pixels and the structures generated by these [129]. Pixels from these structures, often clustered in homogeneous regions, tend to belong to the same material. The second one is given by the correlation in the class labels assigned to neighboring pixels in the image, making it more likely for labels that are together to belong to the same class. The inclusion of spatial features during the classification process has been proven to increase the performance of the models in several studies [27, 113].

Typically, spatial-spectral classification schemes are split into three main categories. The first group is comprised of different spatial regularization algorithms [165], usually applied as a post-processing step to refine the classification maps obtained by the pixel-based classifiers. Numerous proposals have made use of watershed segmentation [68], expectation-maximization segmentation [55], hierarchical segmentation [130], or superpixel segmentation [66] to divide an image in a set of homogeneous, non-overlapping regions. This step can either be done by taking a dimensionally-reduced image as the input to the segmentation algorithm or by applying adapted algorithms that can make use of both the spectral and spatial information of the original image [131, 118]. The class labels from those regions, previously predicted by a pixel-wise classifier, are replaced by new labels determined by majority voting in an attempt to reduce miss-classified pixels. It has been observed that higher-quality segmentation algorithms usually yield better classification accuracy, which is related to the desired properties of intra-class dissimilarity, inter-class similarity, and boundary adherence. A second category is comprised of algorithms that exploit the spatial dependency to refine the values of the pixels of the scene with the application of spatial filters. This includes proposals such as 2D Gabor filters [87], morphological profiles [19], etc. The third and last category is comprised of models, also called joint learning, that try to simultaneously learn both spatial and spectral characteristics [49].

This chapter discusses the different spatial-spectral information extraction methods developed in this thesis. The work focuses on the first two groups, with spatial regularization used along geometric transforms and imputation as a means to perform data augmentation in DL classification schemes, and the use of spatial filters through the application of anisotropic diffusion to extract spatial-spectral information. Additionally, and as described in Section

1.2.1, special care is dedicated to the development of efficient algorithms for consumer-grade hardware.

2.1 Extended Anisotropic Diffusion Profiles

A scale-space is an image representation at a continuum of scales, embedding gradually simplified versions of the image, provided that it fulfills certain requirements [148]. Its first known application dates to 1959 in a Japanese paper written by Taizo Iijima. Witkin [149] proposed, in 1983, a similar concept based on the application of a Gaussian kernel to create progressively smoother versions of an image.

Convolving an image with a Gaussian kernel is the equivalent of solving the heat diffusion Partial Differential Equation (PDE) with a constant diffusion coefficient c with the image as its initial condition:

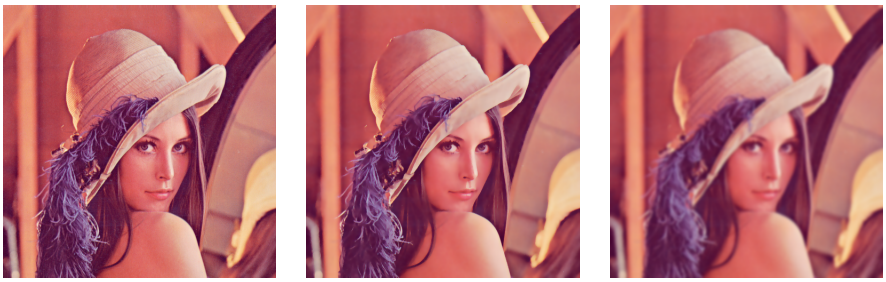
$$\frac{\delta L}{\delta t} = \mathrm{div}(c(x, y, t) \cdot \nabla L). \tag{2.1}$$

This application, although successful in reducing the variability of the pixels of the image, carries the drawback of blurring the edges of the objects present on it as well. To prevent this, Perona and Malik [111] formulated a nonlinear diffusion coefficient based on the image gradient, capable of preserving the natural boundaries occurring in the image:

$$c(x, y, t) = g(|\nabla L_\sigma(x, y, t)|), \tag{2.2}$$

The difference between both approaches can be seen in Fig. 2.1.

The diffusion equation is one of the possibilities to build a linear scale-space [8] from an image, a concept that has been successfully applied to hyperspectral image classification in the past. Examples of this can be seen in [24, 46, 143]. Anisotropic diffusion, in particular, has been used for the generation of extended profiles for hyperspectral image classification [102]. The working principle of this proposal is related to the reduction in the variability of the spectral signatures of the pixels in an image, which in turn can increase the accuracy of the applied classification models. In this work, the extended profile for the different hyperspectral scenes is generated by applying a diffusion equation a certain number of times to each of the principal components of the image. For each of the iterations over a principal component,



(a) (b) (c)

Figure 2.1: Anisotropic diffusion vs Gaussian filtering: (**a**) Lena image, (**b**) Anisotropic diffusion applied, and (**c**) Gaussian filtering applied.

the diffusion equation will generate a new image with a decreasing level of detail as one of the input parameters, called process time, is progressively increased. Once all the instances of diffusion are applied to a principal component, the output constitutes an ADP (Anisotropic Diffusion Profile). The ADPs for all the principal components are, then, stacked on top of each other, along with the principal component itself, and generate the so-called EADP (Extended Anisotropic Diffusion Profile).

The proposal described in this thesis is comprised of eight different stages, summarized below:

- Setup: The initial stage, where the memory is allocated and all memory transfers required to start the computation of the EADP are performed. This includes copying the memory contents from the host memory to the device memory.

- Gaussian: An initial convolution with a Gaussian kernel is performed to smooth the noise present in the principal component and avoid affecting the calculation of the gradient in the next step.

- Scharr: The horizontal and vertical derivatives of the principal component are obtained using the Scharr operator [83]. What this effectively does is perform edge detection on the smoothed principal component in a similar fashion to the well-known Sobel operator. The main difference, however, lies in the fact that the Scharr operator matrix is anisotropic, unlike Sobel's. Afterward, these derivatives will be used as part of the computation of the diffusivity coefficients.

- Contrast: The contrast parameter is calculated from the smoothed image and used as part of the nonlinear diffusivity matrix.
- Diffusivity: The diffusivity coefficients are obtained, taking the shape of a matrix with the dimensions of the principal component. These will control how the diffusion process operates over each of the regions of the image and are responsible for preserving the borders of the image while performing the intra-region smoothing.
- FED τ: This stage obtains the number of inner step sizes (also denoted by τ) for the iterations required to apply the diffusion equation using the FED method. These step sizes are obtained from the initial parameters passed to the diffusion algorithm.
- FED: At this stage, the FED process is run iteratively, updating the image on each step. After all the iterations are completed, the diffused principal component will become one of the components of the Anisotropic Diffusion Profile.
- Cleanup: During the cleanup stage, the resulting EADP is copied to memory and all the allocated memory is freed.

The parameters of the profile have been selected in a way that obtains the highest possible accuracy with the smallest number of components. A study on this topic is detailed in Chapter 3.

2.1.1 GPU implementation

GPUs have been extensively used in the scientific community to perform data-intensive computations efficiently for nearly two decades. GPU-accelerated algorithms for hyperspectral image classification are commonly found in the relevant literature, including extraction of spatial-spectral information [117, 16], classifiers [128, 151, 153], and others.

As part of the work developed for this thesis, an efficient GPU implementation of the EADP algorithm using CUDA has been proposed in Chapter 3. During the development process of the GPU implementation, a series of optimizations focused on different areas were applied based on the recommendations found in [39], and the information obtained during the profiling of the different kernels that take part in the computations.

The following is a list of the optimization techniques used to improve the performance of the CUDA code:

Memory optimizations

1. Minimization of memory allocations and deallocations, given their high cost. Allocated memory should be reused whenever possible to reduce the impact of these operations on the performance of the algorithm.

2. Minimization of memory usage by performing some computations in place, thus lowering the required memory, and reducing the access to memory, which has a relatively low performance. As a rule of thumb, the higher the arithmetic load of a kernel, the greater the speedup of the GPU implementation compared to a CPU implementation.

3. Minimization of data transfers between host and device. Memory transfers are limited by the bandwidth of the PCIe bus and these should be avoided. Careful consideration must be exercised to determine if moving data between devices must be carried out in the event certain steps in an algorithm run faster on the CPU compared to the CPU.

4. Ensuring memory access coherence by adjacent threads in order to allow access to several memory items to be coalesced into a single operation. When the memory access pattern does not allow coalescing of memory accesses or the data locality is too low, performance will be degraded due to the higher cost of read/write operations

5. Use of pinned memory for data transfers between CPU and GPU to achieve the highest possible bandwidth. Pinned memory is the denomination given to virtual memory pages that have been page-locked and can make use of Direct Memory Access (DMA) to perform the data transfer.

6. Data packing in order to reuse memory elements over multiple operations. Choosing the right data representation can have a significant impact on the performance of a kernel. This is, however, completely dependent on the specific computations performed by the kernel.

Instruction optimizations

1. Using vectorial instructions to maximize instruction parallelism and optimize memory bandwidth. This is especially important in kernels that are bandwidth-bound but has some drawbacks such as an increased register pressure and reduced parallelism.

2. Reduction operations using warp-level primitives to avoid shared memory latency. CUDA 9.0 introduced new primitives to exchange data between the threads of a warp, thus reducing the need to access shared memory. This offers better performance compared to the load and store instructions required in the latter case.

3. Use of shared memory atomics vs global memory ones due to specific hardware support since Maxwell given the faster nature of this memory.

Other optimizations

1. Manual tuning of kernel block sizes to ensure optimal hardware occupancy and resource usage. This includes maximizing instruction parallelism per thread to achieve higher performance at lower occupancy based on data dependencies [145].

Additionally, GPU-optimized libraries were used whenever possible. These libraries have been extensively optimized and support the capabilities of several CUDA hardware versions for the best possible performance.

2.1.2 Experimental results

The GPU implementation was compared to an OpenMP-accelerated CPU implementation that acted as our baseline. These implementations rely on Fast Explicit Diffusion (FED) schemes [60], more computationally efficient than the previously proposed Additive Operator Splitting (AOS) schemes, and well suited for parallelization using GPUs. The classification performance achieved by both methods showed no statistical differences, with identical diffusion outputs in all tests.

The GPU implementation is capable of outperforming the CPU counterpart on every stage but the Setup and Cleanup ones. These two stages are responsible for the initial memory allocation and copy of the data into the temporary buffers where it will be processed. It is a well-known fact that GPU memory allocation and data transfers between the host memory and device memory are slow. The execution results display the entire dimension of this difference between both implementations, with the CPU being several orders of magnitude faster. The highest speedups are achieved in the Contrast and Scharr stages, with values that reach over $100\times$ and $30\times$, respectively, in the Houston scene for an EADP with 63 components. The results obtained for IndianP and PaviaU also show a similar dynamic, with slight differences given by the dimensions and aspect ratio differences across scenes.

The overall speedup across all the scenes is slightly above $10\times$. It's worth noting that the achieved speedup was obtained on a consumer-grade GPU with an artificially low cap to double precision arithmetic, with a performance that is 1/32th that of single precision compared to the expected 1/2.

A comparison against other methods from the literature, namely EMP, $\mathrm{EAP_a}$, $\mathrm{EEA_a}$, EMAP, and EMEP shows that the obtained performance of an EADP of 63 components is comparable in terms of overall accuracy for the Houston, IndianP, and PaviaU scenes. The performance increase obtained by EADP is especially notable for the scenes of Houston and PaviaU.

The contributions from this section have been published in an article detailed on Chapter 3.

2.2 Dual-Window Superpixel Data Augmentation (DWS)

Since its inception, Deep Learning has become the most prominent family of learning algorithms applied to a multitude of machine learning problems due to its superior performance. This can also be observed in the hyperspectral image classification literature, where DL methods have overcome the popularity of traditional classifiers such as SVM or RF [85]. Despite the interest received by the scientific community, Deep Learning has some shortcomings that may not make it adequate for all problem types. Some of these problems are the lack of sufficient labeled data and the high computational cost of the models, which become especially relevant in the multispectral and hyperspectral LULC classification tasks. To overcome these limitations, multiple proposals have been made related to data augmentation [126, 166, 160, 64], and superpixel segmentation [66, 47, 116]. Data augmentation is commonly used as a means to increase the capability of a model to generalize, while superpixel segmentation can be used as a means to reduce the computational cost associated with DL models [90, 28]. Despite this, the majority of the data augmentation algorithms in the multispectral and hyperspectral literature are based on per-pixel or pixel-based classification schemes. In contrast to this, the proposal described in Chapter 4 presents a data augmentation method that can be used with both pixel and segment-based DL classification schemes.

As part of this thesis, two DL classification schemes were proposed for the application of the proposed data augmentation method. The first one is a traditional pixel-based schema, where three-dimensional regions or patches are extracted from the original hyperspectral image

using a sliding window of fixed dimensions. The second one is a superpixel-based scheme, where the original image is first segmented and the resulting segments are used for patch extraction. This is achieved by taking a single patch from the center of each of the superpixels, effectively reducing the computational load by taking advantage of the high similarity of the pixels that belong to the same region. Then, data augmentation using DWS is applied to the patches. Once the new augmented patches are generated, they are finally fed to a CNN to be processed.

2.2.1 DWS framework

The data augmentation method proposed, called DWS, is based on the composition of multiple geometric transformations that do not need manual tuning and can act as a convenient replacement for the traditional rotation and mirroring operations, providing better classification performance with the same amount of generated data. In contrast to other data augmentation methods relying on geometric transforms such as [166, 160, 64], our proposal subdivides an input patch into two different regions, external and internal, which can then be used to stack the geometric transform of choice. The subdivision of the patch is based on the assumption that the most relevant features for classification are present in the inner region and a combination of the original patch information along with a newly created set of augmented patches can generate enough data to improve the generalization capabilities of the resulting model.

The idea of subdividing a patch into regions has already been exploited in [167], where different, predefined sub-regions are generated to better exploit the existing spatial information. Other works following a similar strategy to exploit spatial correlation are [88], and [86]. In the first publication, the authors define the so-called pixel-pair features, which are pairs of pixels that are then used as inputs to a CNN. In the second, pixel-block pairs are built as an extension of the original pixel-pair features. These block pairs incorporate spatial features to the data augmentation method to further improve classification accuracy. Apart from these, other relevant data augmentation methods can be seen in [126], where samples are synthesized from a multivariate normal distribution, and [11, 9], making use of GANs to generate new augmented samples.

The DWS data augmentation framework consists of three different stages, with the first stage being only present in the superpixel-based schema:

1. Superpixel-based patch extraction: The first stage consists of an extraction method that

decides the strategy to choose when selecting the subsets of data that will be used as the input patches. This stage makes use of the properties of image segmentation to notably reduce the computational cost of performing classification over large scenes. This is achieved thanks to the strategy chosen, in which only one patch will be selected for each of the segments, taken around its central pixel.

2. Patch subdivision: In this stage, patches are divided into two different regions. Contrary to how data augmentation operations are usually applied, affecting the entire patch, this approach is applied only to the desired region. This allows for the generation of a larger number of augmented patches by making it possible to only apply the operations to each region independently.

3. Patch transformation: Geometric operations are applied to either of the subdivisions of the patch obtained in the previous step. A variety of different transformations can be applied, based on the chosen operation and region of application. A total of 6 variations were proposed and tested.

2.2.2 Experimental results

In this thesis, we compare different variants of DWS to a mixture of existing augmentation methods for both pixel and superpixel-based classification schemes in a systematic manner. This setup was chosen in an attempt to accurately compare the performance of the different algorithms in an environment that allows for reliable and repeatable results to be obtained. All the experiments were performed with the same network architecture and under the same set of hyperparameters. Additionally, a variety of scenes, both multi and hyperspectral were considered for the experiments. These scenes cover a wide range of resolutions, from a few hundreds of pixels on each dimension to nearly twenty thousand. A total of 12 data augmentation methods were chosen: the traditional Rotate and Flip transforms, with 4× and 16× variants; Inner-Flip 4×, a method based on DWS that applies the Flip transform to the inner region of the patch; Dual-Rotate, and Dual-Flip, both based on DWS, with 4× and 16× variants; PVSA(+/-), a data augmentation method described in [103], with 4× and 16× variants, and Random Occlusion, described in [64], with 4× and 16× variants as well. Each variant describes the number of new samples that are obtained after applying the selected method. A value of 4× represents a method where four patches are generated from the original input patch. Consequently, 16× represents the generation of 16 synthetic samples

for each input patch. These two settings were considered to evaluate the methods across the different number of generated samples. Despite a potentially higher classification accuracy, the downside of having a higher number of samples is the higher computational cost associated with the training.

The experimental results obtained as part of this work comprise two different scenarios. One scenario with a pixel-based classification scheme, where all pixels of the image are individually processed, and one scenario with a superpixel-based classification scheme, where a single patch per segment is extracted from the central pixel. In this later scenario, the cost of performing the prediction for the entire image is greatly reduced. Prediction times remain mostly stable, barring small differences due to the overhead of the classification pipeline, as data augmentation is not applied during the prediction stage. There is a very significant difference in the run times of the prediction stage between the superpixel and pixel pipelines, with the former being three orders of magnitude faster. Prediction times are directly related to the number of pixels present in an image. Each extracted pixel becomes a three-dimensional patch that needs to be processed, which in turn determines the total run-time for this stage. In this regard, the difference in the prediction times closely matches the average superpixel size selected for the segmentation algorithm. It was observed that the training times scale linearly with the number of generated patches for all the variants that were tested, which matches, with an additional overhead also present for the DWS-based methods.

Regarding the classification performance, the results obtained for the pixel-based scenario show that the methods based on DWS achieve the highest classification performance in two out of the three images from the standard dataset. Dual-Flip 16×, and Inner-Flip 4× achieve an OA of 93.4% in Salinas and 88.17% in PaviaU, respectively. For the PaviaU and PaviaC scenes, those with the most complex structures, it can be observed that the increase in the number of generated synthetic samples does not yield an increase in OA, with values decreasing for Dual-Flip 16×, PVS(+/-) 16×, and Random Occlusion 16× compared to the 4× counterparts.

For the superpixel-based scenario, traditional data augmentation methods obtain the highest OA in the PaviaU and PaviaC scenes. Dual-Rotate 16× yields the best accuracy for the Salinas scene. The results of the Dual-Rotate 4× and Dual-Flip 4× data augmentation methods prove to be systematically worse than the baseline. Contrary to this, the methods based on DWS obtain the highest classification performance for all the scenes of the Galicia dataset. In this set of tests, for nearly all instances, an increase in the number of generated samples redound to an increase in OA. In addition to the increase in accuracy, the standard deviation across all

experiments presents lower values when the DWS framework is applied. Random Occlusion 4× does not yield good results when applied to the images of the Galicia dataset, with an average improvement in OA of -0.19.

The experimental data shows DWS obtains satisfactory results across a variety of scenes in both pixel and superpixel-based classification schemes. Despite this, DWS may not be of application to all image types due to the nature of the operations it performs. The patch subdivision stage can alter the original structures in the scene, which may cause worse results to be obtained for regions with small, irregular structures. One possible line of future work could be the study of the possibility of adding some parametrization to the augmentation methods to increase the adaptability to different scene types and mitigate the current shortcomings.

The contributions from this section have been published in an article detailed on Chapter 4.

2.3 A new multispectral data augmentation technique based on data imputation

The proposal detailed in this work introduces a new data augmentation method for segment-based classification schemes based on a combination of superpixel segmentation and data imputation. The method focuses on increasing DL models' generalization capabilities in scenarios where the computational costs of the training and prediction stages for a scene are prohibitive due to the large amount of data contained in multi or hyperspectral images. A superpixel segmentation [4] algorithm is applied to these images in order to reduce the cost of applying the chosen model. In this situation where a low number of samples are available, our data augmentation constructs a newly imputed dataset that is then fed to the model and serves as the input during training. There are many data augmentation methods in the literature developed with multi and hyperspectral images in mind. These methods are usually based on geometric transforms and created with per-pixel or pixel-based classification in mind [166, 160, 64]. A previous contribution from this thesis, described in Section 2.2, expands this idea by subdividing input patches into two independent regions where one of several geometric transforms is applied independently. Other works such as [126, 11, 9] focus on different alternatives, such as creating new samples by drawing those from a multivariate normal distribution or making use of GANs.

The use of segmentation is not a new concept in the hyperspectral classification field.

There are several examples of the use of segmentation as a means to enhance classification results as a post-processing step that relies on the spatial dependencies between pixels, such as [90]. These works make use of spatial regularization to extract information that will later be combined with the classification results to generate a more accurate and uniform classification map with the information from the segments. These spatial regularization approaches have also been applied in combination with DL models, such as CNNs, in [28]. DWS, described in the previous section, already introduced the idea of using superpixel segmentation as a means to reduce the computational requirements of processing multi and hyperspectral images in combination with a CNN model. In contrast with these previous works, the current proposal does not only rely on the information obtained by the superpixel segmentation as a means to reduce the computational complexity or perform spatial regularization but builds upon it to develop a completely new data augmentation method that can be applied to superpixel-based classification schemes. In this new method, part of the information contained in the original input patches will be removed based on the resulting segmentation and replaced using a data imputation algorithm. This imputation algorithm generates new pixel vectors that will then yield a new set of synthetic patches that will be sent to the classifier along with the original set. Several data augmentation algorithms have been analyzed for this task as part of this work. To the best of our knowledge, this work constitutes the first time that data imputation algorithms are used to perform data augmentation for either multispectral or hyperspectral images.

2.3.1 Data imputation

This work introduces a novel data augmentation method for multi and hyperspectral images based on superpixel segmentation and data imputation. Our data augmentation method works by removing and replacing undesirable information contained inside the input patches. This information refers to the pixels not belonging to the segment the patch was extracted from. The removed information is then replaced by newly obtained information generated by different imputation methods from the contents of the original patch.

The proposed data augmentation method is divided into four stages.

1. Segmentation: A superpixel segmentation algorithm is applied to the original image with two goals in mind. The first one would be the simplification of the scene for the training and prediction stages. The second one is to use the spatial information from the

superpixels to determine what regions of the input patch to replace in order to generate new samples.

2. Patch extraction: Once the segments have been generated, the next step is to extract the patches. One patch will be extracted from each segment. The dimensions of the patch should be adequate for the segment size obtained by the superpixel segmentation algorithm, as segments that are too big or too small in proportion to the patch size won't generate new samples.

3. Patch erasure: Input patches may contain information from one or more segments. This implies that the information present in a patch may belong to several different class labels. To mitigate this phenomenon and prevent a negative impact on classification performance, certain information about the patch has to be replaced. In this stage, the areas of the patch that do not fall within the segment that originated it are removed.

4. Patch imputation: In this final stage, the information removed from the input patches during the previous stage is replaced by newly generated information belonging to the pixels that remain in the patch. This is done by applying data imputation to the erased patch. To achieve this, patches are flattened from 3D cubes of data to 2D matrices with each of the rows corresponding to a pixel vector. After this step, the data imputation algorithm of choice is applied. Finally, the 2D matrix is again reshaped into a 3D cube of data. The result of this process is a set of new, synthetic patches.

A total of five data imputation algorithms with different complexity levels have been tested: Constant imputation, KNNimpute [137], MICE [25], SoftImpute [97] and SVDimpute [137]. Any data imputation algorithm can be used, as long as they are capable of converging with relatively low amounts of data. In this work, the patch size was chosen to be 25×25, and imputation was performed as long as the patches had 10% of the pixels left after the patch erasure stage. Even though it is possible to perform data imputation in those circumstances, the results would be of low quality. A balance between the remaining amount of data and the number of pixels should be maintained. As a rule of thumb, aiming for a 50% of missing pixels should be the preferred strategy when selecting the parameters for the segmentation stage.

2.3.2 Experimental results

The classification scheme in combination with the data augmentation method developed as part of this work was able to achieve an increase in the classification performance for all datasets regardless of the nature of the image. KNNimpute achieves the highest OA for two out of the four scenes of the Galicia dataset, while Dual-Flip 2× does the same for the remaining two. The increase in accuracy obtained by the augmentation methods was relatively modest, with the largest increase being observed in the Oitaven scene, at 2.14%, for Dual-Flip 2×.

When it comes to the large-scale multispectral scenes from the GID, MICE consistently obtains the highest OA across 16 out of the 17 images. The remaining image has a baseline performance of over 99.5%, which makes it exceedingly hard to obtain performance improvements from data augmentation methods. The results achieved for the 5-class large scene classification set tend to be lower than those of the 15-class set due to the higher baseline OA of the former. The scenes from the 5-class large scene classification set have more irregular structures distributed across the image, making it a better fit for the proposed data augmentation method. The OA increase achieved in the 5-class large scene classification set reaches up to 2.57% for the GF2-5G scene. The OA increase achieved in the 15-class large scene classification set reaches up to 5.61% for the GF2-15I scene.

The execution times obtained for the 15-class large-scale classification set of the GID show the computational cost of the prediction stage stays almost unchanged, as expected as data augmentation is applied only during the training stage. The small differences that can be observed are caused by the overhead of the pipeline setup used for data augmentation. The runtime of the prediction stage is affected by the chosen imputation algorithm and is related to its computational complexity. Constant imputation offers the fastest execution times, as we would expect from its linear complexity. MICE is, by a significant margin, the slowest data imputation algorithm. In general terms, the complexity of the data augmentation method introduced during the training stage causes an increase of roughly 2× to 5× in the execution time. Dual-Flip 2× has a lower execution time than the proposed data augmentation methods based on data imputation due to the simplicity of the geometric transformations it's built upon. In spite of this advantage, the results for the scenes from the GID lag behind even the simple Constant imputation in many cases.

The contributions from this section have been published in an article detailed in Chapter 5.

Chapter 3

Extended Anisotropic Diffusion Profiles

The current chapter is a verbatim reproduction of the following article:

International journal

Acción, Álvaro, Francisco Argüello, and Dora B. Heras. "Extended anisotropic diffusion profiles in GPU for hyperspectral imagery." IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 12, no. 12 (2019): 4964-4976.
DOI: 10.1109/JSTARS.2019.2939857

CHAPTER 4

DUAL-WINDOW SUPERPIXEL DATA AUGMENTATION

The current chapter is a verbatim reproduction of the following article:

International journal

Acción, Álvaro, Francisco Argüello, and Dora B. Heras. "Dual-window superpixel data augmentation for hyperspectral image classification." Applied Sciences 10, no. 24 (2020): 8833.
DOI: 10.3390/app10248833

CHAPTER 5

DATA AUGMENTATION BASED ON DATA IMPUTATION

The current chapter is a verbatim reproduction of the following article:

International journal

Acción, Álvaro, Francisco Argüello, and Dora B. Heras. "A new multispectral data augmentation technique based on data imputation." Remote Sensing 13, no. 23 (2021): 4875.

DOI: 10.3390/rs13234875

Chapter 6

Conclusions

This thesis addresses the application of efficient spatial-spectral information extraction methods as a means to improve the classification accuracy of multi and hyperspectral images.

Classification, defined as the assignation of a class label to each of the pixels of a scene, of multispectral and hyperspectral remote sensing images is a complex problem. One of the reasons is the small amount of labeled information that can be used to train supervised classifiers, such as the ones this research is focused on. Different techniques allow us to tackle this problem, such as different machine learning approaches, both classical and those based on deep learning. These rely on extracting as much spatial and spectral information as possible from the bands captured by the sensors, exploiting the spatial and spectral resolution of the images.

More specifically, in this thesis, spatial and spectral information extraction methods based on the extraction of profiles using anisotropic diffusion and data augmentation have been proposed. Two data augmentation methods have been detailed. The first one uses a composition of geometric transforms over the input patches. These are 3D cubes of data of dimensions $width \times height \times bands$, extracted from the superpixels of the image. These patches are subdivided into two regions where these transforms can be applied independently. The second one uses a combination of superpixel segmentation and data imputation to replace information on the input patches that belongs to a superpixel other than the one which the patch was extracted from. These information extraction methods were validated with a variety of multispectral and hyperspectral images captured from sensors mounted on satellites and unmanned aerial vehicles. The methods developed generate new synthetic information that can be used during

the training of several supervised classification models. The increase in the amount of data to be processed made it essential to devote special attention to ensure the efficiency in terms of computational requirements.

The following list summarizes the contributions of this thesis:

1. *An efficient spatial-spectral information extraction method denominated EADP has been proposed.* This approach, named extended anisotropic diffusion profiles, or EADP, uses similar foundations as morphological profiles. The construction of morphological profiles is performed by applying a series of opening and closing operations with a structuring element of increasing size over the components of an image. This generates a new, larger image that can be several times the size of the original one. Several different types of profiles have been proposed in the literature, including attribute profiles, extinction profiles, etc. The construction of profiles increases the size of the original data, as the profiles are usually stacked on top of the original image information in order to be processed. As a result, it is particularly important to carefully consider the profile creation algorithms and make sure to extract as much parallelism as possible from them.

   In this thesis, we propose the application of the nonlinear diffusion equation for the creation of profiles from hyperspectral images. The process begins by applying a principal component analysis algorithm is applied to the original hyperspectral image. The EADP is constructed by concatenating a series of nonlinear diffusion instances applied to the principal components, generating new components with decreasing levels of detail. The introduction of these components reduces the spectral variability of the spectral signatures of the materials in the scene, increasing the classification performance. The nonlinear diffusion equation is implemented using a FED scheme, which makes it a good candidate for its implementation in parallel architectures since each pixel in the image can be processed independently, generating significant data parallelism. An efficient GPU implementation capable of achieving a speedup of nearly 10× was developed for commodity hardware using CUDA. The proposed method achieves a significant increase in the classification performance for several benchmarking datasets, most notably in the hyperspectral Houston scene with a 98.82% overall accuracy, improving the results from comparable methods from the bibliography.

2. *Different data augmentation methods based on subdividing an input patch into multiple regions and applying geometric operations, and data replacement using data imputation*

have been proposed. In situations where a limited number of samples are available, a model may become more prone to overfitting due to the limited capability to generalize based on the training data. Data augmentation methods were created to solve this problem by generating new synthetic samples based on the information of the existing ones. This is especially relevant in deep learning models, where large amounts of data are required. Many different approaches have been developed over the years, most of them related to image processing and RGB images in particular. Among these, geometric transforms are one of the most common means to perform data augmentation. These transforms are commonly applied to an input patch as a whole during the augmentation process of multispectral and hyperspectral images and used in classification schemes where pixel-based classification is performed. These schemes have very large computational requirements that may render them unusable when processing large images. Our proposal includes data augmentation methods that can be used for both these pixel-based schemes, as well as in superpixel-based ones, where superpixel segmentation is applied and a subset of the pixels from each segment is taken as representative of the whole region. In this thesis, we propose two data augmentation methods that have been shown to improve the results over other comparable data augmentation methods present in the literature:

a) *A new data augmentation method called dual-window superpixel, or DWS, has been proposed.* It relies on the subdivision of the input patches into two separate regions. First, an inner region is assumed to contain the most relevant information to the classification process, as it is closer to the pixel that determines the class label, and second, an outer region of lesser importance. Several different geometric transforms can be applied to each region independently. The application of these transforms to the inner region exclusively generates better classification results than applying it to the entire patch. The composition of transforms, applying them independently to both regions generates more data and is capable of further increasing the classification performance.

   The proposed augmentation method can be applied to a variety of classification schemes based on deep learning. In our tests, DWS was applied to CNN networks, using both pixel and superpixel-based classification schemes. The results applying DWS to a dataset of multispectral vegetation scenes captured by a drone-mounted sensor show that the proposal achieves up to a 98.56% accuracy for the Ermidas

scene, higher than other state-of-the-art data augmentation methods from the literature.

b) *A new data augmentation method based on data imputation has been proposed.* Apart from using geometric transforms, such as in the case of contribution 2.a that we have just described, other types of operations can also be used to perform data augmentation. More specifically, the experiments performed have shown that the replacement of pixels using data imputation methods can also obtain good results. As part of this thesis, a data augmentation method based on data imputation after performing superpixel segmentation of the original image has been developed. This data augmentation method was created for its use with superpixel-based classification schemes. The original image goes through a segmentation process. The resulting segments are used to both simplify the computational complexity of the processing as well as to remove and replace some information in the input patches with the data obtained by an imputation algorithm. Several imputation algorithms have been tested, including SoftImpute, SVDImpute, and MICE, among others. This increases the homogeneity of the patches and prevents some undesirable cases where patches next to the natural frontiers of the structures in the images may contain information belonging to other class labels. To achieve this, the pixels in the input patches that do not belong to the segment they were extracted from are removed. An imputation algorithm is then applied to the remaining pixels to generate the missing data. The augmented dataset is then fed to a CNN.

   The experimental results were obtained after applying the SLIC superpixel segmentation algorithm to the input images, using a CNN model. Two multispectral datasets were used; a vegetation dataset obtained by the authors using a drone-mounted sensor, and a second dataset captured by the Gaofen-2 satellite. The results obtained show that the proposed data augmentation method obtains the highest overall classification accuracy on 16 out of the 17 scenes of the Gaofen-2 dataset. This work constitutes, to the best of our knowledge, the first time data imputation is used as a means to perform data augmentation on multi and hyperspectral images.

The results presented as part of the findings that lead to the aforementioned contributions show that all the objectives of this thesis have been successfully completed. The methods

developed as part of the thesis can be applied to a variety of problems that are very relevant in the field of multi and hyperspectral image processing. The research lines that were explored open two main directions regarding hypothetical future work. First, the data augmentation methods can be applied to a variety of classification networks and to a variety of multi and hyperspectral images, not only to images from the land cover domain, as is the case in this thesis. Additionally, another very relevant problem in the field of classification of remote sensing multi and hyperspectral images is the imbalance of the classes. The number of pixels that belong to each class label for which reference information is available can show large differences depending on the class itself. For example, in vegetation images it is common for certain vegetation species to be less abundant than others, which can cause the learning process to become more difficult. In our data augmentation proposals, no specific actions are being taken to mitigate the class imbalance typically associated with hyperspectral and multispectral multi-class classification problems. A more targeted data augmentation approach that reduced the imbalance would probably improve the convergence of the models. The methods developed as part of the thesis apply data augmentation to all classes indistinctly.

The extraction of spatial and spectral information using extended anisotropic diffusion profiles can also be applied to other problems, not exclusively classification-related ones. This is the case with change detection and anomaly detection. Anomaly detection for high-dimensional images tries to identify, without prior knowledge, elements in a scene that show significant enough differences compared to the background information. At the time of the writing, an initial attempt to develop an anomaly detection method that could be applied to multispectral and hyperspectral images relying on extended anisotropic diffusion profiles was being approached.

Bibliography

[1] GCC , the GNU compiler collection. https://gcc.gnu.org, 2022. Last accessed on 20/05/2022.

[2] Tensorflow. https://www.tensorflow.org/, 2022. Last accessed on 20/05/2022.

[3] 2013 IEEE GRSS data fusion contest – fusion of hyperspectral and LiDAR data. https://hyperspectral.ee.uh.edu/?page_id=459. Last accessed on 20/11/2022.

[4] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. SLIC superpixels. Technical report, EPFL, 2010.

[5] S Agatonovic-Kustrin and Rosemary Beresford. Basic concepts of artificial neural network (ann) modeling and its application in pharmaceutical research. *Journal of pharmaceutical and biomedical analysis*, 22(5):717–727, 2000.

[6] Asmala Ahmad and Shaun Quegan. Comparative analysis of supervised and unsupervised classification on multispectral data. *Applied Mathematical Sciences*, 7(74):3681–3694, 2013.

[7] Muhammad Ahmad, Sidrah Shabbir, Swalpa Kumar Roy, Danfeng Hong, Xin Wu, Jing Yao, Adil Mehmood Khan, Manuel Mazzara, Salvatore Distefano, and Jocelyn Chanussot. Hyperspectral image classification—traditional to deep models: A survey for future prospects. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:968–999, 2021.

[8] Pablo Fernández Alcantarilla, Adrien Bartoli, and Andrew J Davison. KAZE features. In *European Conference on Computer Vision*, pages 214–227. Springer, 2012.

[9] T Alipourfard and H Arefi. Virtual training sample generation by generative adversarial networks for hyperspectral images classification. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, 42:63–69, 2019.

[10] Bowen Alpern, Larry Carter, Ephraim Feig, and Ted Selker. The uniform memory hierarchy model of computation. *Algorithmica*, 12(2):72–109, 1994.

[11] Nicolas Audebert, Bertrand Le Saux, and Sébastien Lefèvre. Generative adversarial networks for realistic synthesis of hyperspectral samples. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 4359–4362. IEEE, 2018.

[12] Nicolas Audebert, Bertrand Le Saux, and Sébastien Lefèvre. Deep learning for classification of hyperspectral data: A comparative review. *IEEE Geosci. Remote Sens. Mag.*, 7(2):159–173, 2019.

[13] Ram Avtar, Netrananda Sahu, Ashwani Kumar Aggarwal, Shamik Chakraborty, Ali Kharrazi, Ali P Yunus, Jie Dou, and Tonni Agustiono Kurniawan. Exploring renewable energy resources using remote sensing and GIS—A review. *Resources*, 8(3):149, 2019.

[14] Michael J Barnsley, Jeff J Settle, Mike A Cutter, Dan R Lobb, and Frederic Teston. The PROBA/CHRIS mission: A low-cost smallsat for hyperspectral multiangle observations of the earth surface and atmosphere. *IEEE Transactions on Geoscience and Remote Sensing*, 42(7):1512–1520, 2004.

[15] Pedro G Bascoy, Alberto S Garea, Dora B Heras, Francisco Argüello, and Alvaro Ordóñez. Texture-based analysis of hydrographical basins with multispectral imagery. In *Remote Sensing for Agriculture, Ecosystems, and Hydrology XXI*, volume 11149, page 111490Q. International Society for Optics and Photonics, 2019.

[16] Pedro G Bascoy, Pablo Quesada-Barriuso, Dora B Heras, Francisco Argüello, Begüm Demir, and Lorenzo Bruzzone. Extended attribute profiles on gpu applied to hyperspectral image classification. *The Journal of Supercomputing*, 75(3):1565–1579, 2019.

[17] Mariana Belgiu and Lucian Drăguţ. Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, 114:24–31, 2016.

[18] Jón Atli Benediktsson, Jón Aevar Palmason, and Johannes R Sveinsson. Classification of hyperspectral data from urban areas based on extended morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):480–491, 2005.

[19] Jon Atli Benediktsson, Martino Pesaresi, and Kolbeinn Amason. Classification and feature extraction for remote sensing images from urban areas based on morphological transformations. *IEEE Transactions on Geoscience and Remote Sensing*, 41(9):1940–1949, 2003.

[20] Ujwala Bhangale, Surya Durbha, Abhishek Potnis, and Rajat Shinde. Rapid earthquake damage detection using deep learning from vhr remote sensing images. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 2654–2657. IEEE, 2019.

[21] José M Bioucas-Dias, Antonio Plaza, Gustavo Camps-Valls, Paul Scheunders, Nasser Nasrabadi, and Jocelyn Chanussot. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and remote sensing magazine*, 1(2):6–36, 2013.

[22] OpenMP Architecture Review Board. OpenMP application programming interface. `https://www.openmp.org/wp-content/uploads/OpenMP-API-Specification-5-2.pdf`, 2021. Last accessed on 20/05/2022.

[23] Michael Boyer, David Tarjan, Scott T Acton, and Kevin Skadron. Accelerating leukocyte tracking using CUDA: A case study in leveraging manycore coprocessors. In *2009 IEEE international symposium on parallel & distributed processing*, pages 1–12. IEEE, 2009.

[24] Lori Mann Bruce and Jiang Li. Wavelets for computationally efficient hyperspectral derivative analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7):1540–1546, 2001.

[25] S van Buuren and Karin Groothuis-Oudshoorn. mice: Multivariate imputation by chained equations in r. *Journal of statistical software*, pages 1–68, 2010.

[26] James B Campbell and Randolph H Wynne. *Introduction to remote sensing*. Guilford Press, 2011.

[27] Gustavo Camps-Valls, Devis Tuia, Lorenzo Bruzzone, and Jon Atli Benediktsson. Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. *IEEE signal processing magazine*, 31(1):45–54, 2013.

[28] Jiayan Cao, Zhao Chen, and Bin Wang. Deep convolutional networks with superpixel segmentation for hyperspectral image classification. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 3310–3313. IEEE, 2016.

[29] Gabriele Cavallaro, Dora B Heras, Zebin Wu, Manil Maskey, Sebastian Lopez, Piotr Gawron, Mihai Coca, and Mihai Datcu. High-performance and disruptive computing in remote sensing: HDCRS-A new working group of the GRSS earth science informatics technical committee. *IEEE Geoscience and Remote Sensing Magazine*, 2022.

[30] Xavier Ceamanos, Björn Waske, Jon Atli Benediktsson, Jocelyn Chanussot, Mathieu Fauvel, and Johannes R Sveinsson. A classifier ensemble based on fusion of support vector machines for classifying hyperspectral data. *International Journal of Image and Data Fusion*, 1(4):293–307, 2010.

[31] 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.

[32] 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 Trans. Geosci. Remote Sens.*, 54(10):6232–6251, 2016.

[33] Yushi Chen, Xing Zhao, and Xiuping Jia. Spectral–spatial classification of hyperspectral data based on deep belief network. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 8(6):2381–2392, 2015.

[34] Emmanuel Christophe, Julien Michel, and Jordi Inglada. Remote sensing processing: From multicore to GPU. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(3):643–652, 2011.

[35] Russell G Congalton. A review of assessing the accuracy of classifications of remotely sensed data. *Remote sensing of environment*, 37(1):35–46, 1991.

[36] NVIDIA Corporation. NVIDIA developer. https://developer.nvidia.com/, 2007. Last accessed on 20/05/2022.

[37] NVIDIA Corporation. NVIDIA tesla p100 whitepaper. https://images.nvidia.com/content/pdf/tesla/whitepaper/pascal-architecture-whitepaper.pdf, 2016. Last accessed on 20/05/2022.

[38] NVIDIA Corporation. Whitepaper NVIDIA geforce GTX 1080. http://international.download.nvidia.com/geforce-com/international/pdfs/GeForce_GTX_1080_Whitepaper_FINAL.pdf, 2016. Last accessed on 20/05/2022.

[39] NVIDIA Corporation. CUDA C best practices guide. https://docs.nvidia.com/cuda/pdf/CUDA_C_Best_Practices_Guide.pdf, 2018. Last accessed on 20/05/2022.

[40] NVIDIA Corporation. NVIDIA TURING GPU ARCHITECTURE. https://images.nvidia.com/aem-dam/en-zz/Solutions/design-visualization/technologies/turing-architecture/NVIDIA-Turing-Architecture-Whitepaper.pdf, 2018. Last accessed on 20/05/2022.

[41] NVIDIA Corporation. CUDA TOOLKIT. https://developer.nvidia.com/cuda-toolkit, 2022. Last accessed on 20/05/2022.

[42] Lewis M Cowardin. *Classification of wetlands and deepwater habitats of the United States*. Fish and Wildlife Service, US Department of the Interior, 1979.

[43] Arthur P Cracknell. *Introduction to remote sensing*. CRC press, 2007.

[44] Mauro Dalla Mura, Jón Atli Benediktsson, Björn Waske, and Lorenzo Bruzzone. Morphological attribute profiles for the analysis of very high resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, 48(10):3747–3762, 2010.

[45] Bharath Bhushan Damodaran and Rama Rao Nidamanuri. Dynamic linear classifier system for hyperspectral image classification for land cover mapping. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):2080–2093, 2014.

[46] Julio M Duarte-Carvajalino, Paul E Castillo, and Miguel Velez-Reyes. Comparative study of semi-implicit schemes for nonlinear diffusion in hyperspectral imagery. *IEEE Transactions on Image Processing*, 16(5):1303–1314, 2007.

[47] Leyuan Fang, Shutao Li, Wuhui Duan, Jinchang Ren, and Jón Atli Benediktsson. Classification of hyperspectral images by exploiting spectral–spatial information of superpixel via multiple kernels. *IEEE Trans. Geosci. Remote Sens.*, 53(12):6663–6674, 2015.

[48] Mathieu Fauvel, Jón Atli Benediktsson, Jocelyn Chanussot, and Johannes R Sveinsson. Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, 46(11):3804–3814, 2008.

[49] Mathieu Fauvel, Jocelyn Chanussot, and Jon Atli Benediktsson. A spatial–spectral kernel-based approach for the classification of remote-sensing images. *Pattern Recognition*, 45(1):381–392, 2012.

[50] Mathieu Fauvel, Yuliya Tarabalka, Jon Atli Benediktsson, Jocelyn Chanussot, and James C Tilton. Advances in spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE*, 101(3):652–675, 2013.

[51] Peter Fisher, Alexis J Comber, and Richard Wadsworth. Land use and land cover: contradiction or complement. *Re-presenting GIS*, pages 85–98, 2005.

[52] Michael D Fleming, JS Berkebile, and RM Hoffer. Computer-aided analysis of landsat-1 MSS data: A comparison of three approaches. 1975.

[53] Mark A Folkman, Jay Pearlman, Lushalan B Liao, and Peter J Jarecke. EO-1/hyperion hyperspectral imager design, development, characterization, and calibration. *Hyperspectral Remote Sensing of the Land and Atmosphere*, 4151:40–51, 2001.

[54] Max Gerhards, Martin Schlerf, Kaniska Mallick, and Thomas Udelhoven. Challenges and future perspectives of multi-/hyperspectral thermal infrared remote sensing for crop water-stress detection: A review. *Remote Sensing*, 11(10):1240, 2019.

[55] Pedram Ghamisi, Jon Atli Benediktsson, and Magnus O Ulfarsson. The spectral-spatial classification of hyperspectral images based on hidden markov random field and its expectation-maximization. In *2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS*, pages 1107–1110. IEEE, 2013.

[56] Pedram Ghamisi, Javier Plaza, Yushi Chen, Jun Li, and Antonio J Plaza. Advanced spectral classifiers for hyperspectral images: A review. *IEEE Geoscience and Remote Sensing Magazine*, 5(1):8–32, 2017.

[57] Pedram Ghamisi, Roberto Souza, Jon Atli Benediktsson, Xiao Xiang Zhu, Leticia Rittner, and Roberto A Lotufo. Extinction profiles for the classification of remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 54(10):5631–5645, 2016.

[58] Pedram Ghamisi, Naoto Yokoya, Jun Li, Wenzhi Liao, Sicong Liu, Javier Plaza, Behnood Rasti, and Antonio Plaza. Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):37–78, 2017.

[59] Alexander FH Goetz, Gregg Vane, Jerry E Solomon, and Barrett N Rock. Imaging spectrometry for earth remote sensing. *science*, 228(4704):1147–1153, 1985.

[60] Sven Grewenig, Joachim Weickert, and Andrés Bruhn. From box filtering to fast explicit diffusion. In *Joint Pattern Recognition Symposium*, pages 533–542. Springer, 2010.

[61] JA Gualtieri and S Chettri. Support vector machines for classification of hyperspectral data. In *IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment. Proceedings (Cat. No. 00CH37120)*, volume 2, pages 813–815. IEEE, 2000.

[62] Driss Haboudane, John R Miller, Elizabeth Pattey, Pablo J Zarco-Tejada, and Ian B Strachan. Hyperspectral vegetation indices and novel algorithms for predicting green

LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote sensing of environment*, 90(3):337–352, 2004.

[63] Jisoo Ham, Yangchi Chen, Melba M Crawford, and Joydeep Ghosh. Investigation of the random forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):492–501, 2005.

[64] Juan Mario Haut, Mercedes E Paoletti, Javier Plaza, Antonio Plaza, and Jun Li. Hyperspectral image classification using random occlusion data augmentation. *IEEE Geosci. Remote Sens. Lett.*, 16(11):1751–1755, 2019.

[65] Lin He, Jun Li, Chenying Liu, and Shutao Li. Recent advances on spectral–spatial hyperspectral image classification: An overview and new guidelines. *IEEE Trans. Geosci. Remote Sens.*, 56(3):1579–1597, 2017.

[66] Zhi He, Yue Shen, Miao Zhang, Qiang Wang, Yan Wang, and Renlong Yu. Spectral-spatial hyperspectral image classification via SVM and superpixel segmentation. In *2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, pages 422–427. IEEE, 2014.

[67] Uta Heiden, Karl Segl, Sigrid Roessner, and Hermann Kaufmann. Determination of robust spectral features for identification of urban surface materials in hyperspectral remote sensing data. *Remote Sensing of Environment*, 111(4):537–552, 2007.

[68] Dora B Heras, Francisco Argüello, and Pablo Quesada-Barriuso. Exploring ELM-based spatial–spectral classification of hyperspectral images. *International Journal of Remote Sensing*, 35(2):401–423, 2014.

[69] Kenneth E Hoff III, John Keyser, Ming Lin, Dinesh Manocha, and Tim Culver. Fast computation of generalized voronoi diagrams using graphics hardware. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 277–286, 1999.

[70] Hyperspectral images. `https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html`. Last accessed on 20/11/2022.

[71] Hyperspectral remote sensing scenes. `https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes#Pavia_Centre_and_University`. Last accessed on 20/11/2022.

[72] Akira Iwasaki, Nagamitsu Ohgi, Jun Tanii, Takahiro Kawashima, and Hitomi Inada. Hyperspectral imager suite (HISUI)-japanese hyper-multi spectral radiometer. In *2011 IEEE International Geoscience and Remote Sensing Symposium*, pages 1025–1028. IEEE, 2011.

[73] Sveinn R Joelsson, Jon Atli Benediktsson, and Johannes R Sveinsson. Random forest classifiers for hyperspectral data. In *Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS'05.*, volume 1, pages 4–pp. IEEE, 2005.

[74] RF Kokaly, RN Clark, GA Swayze, KE Livo, TM Hoefen, NC Pearson, RA Wise, WM Benzel, HA Lowers, RL Driscoll, et al. Usgs spectral library version 7 data: Us geological survey data release. *United States Geological Survey (USGS): Reston, VA, USA*, 2017.

[75] Marek Kovar, Marian Brestic, Oksana Sytar, Viliam Barek, Pavol Hauptvogel, and Marek Zivcak. Evaluation of hyperspectral reflectance parameters to assess the leaf water content in soybean. *Water*, 11(3):443, 2019.

[76] VV Kozoderov, TV Kondranin, and EV Dmitriev. Recognition of natural and man-made objects in airborne hyperspectral images. *Izvestiya, Atmospheric and Oceanic Physics*, 50(9):878–886, 2014.

[77] Louisa Lam and SY Suen. Application of majority voting to pattern recognition: an analysis of its behavior and performance. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 27(5):553–568, 1997.

[78] David A Landgrebe. *Signal theory methods in multispectral remote sensing*, volume 24. John Wiley & Sons, 2003.

[79] E Scott Larsen and David McAllister. Fast matrix multiplies using graphics hardware. In *Proceedings of the 2001 ACM/IEEE Conference on Supercomputing*, pages 55–55, 2001.

[80] Steve Lawrence, C Lee Giles, Ah Chung Tsoi, and Andrew D Back. Face recognition: A convolutional neural-network approach. *IEEE Trans. Neural Netw.*, 8(1):98–113, 1997.

[81] Hyungtae Lee and Heesung Kwon. Going deeper with contextual CNN for hyperspectral image classification. *IEEE Trans. Image Process.*, 26(10):4843–4855, 2017.

[82] Leigh B Lentile, Zachary A Holden, Alistair MS Smith, Michael J Falkowski, Andrew T Hudak, Penelope Morgan, Sarah A Lewis, Paul E Gessler, and Nate C Benson. Remote sensing techniques to assess active fire characteristics and post-fire effects. *International Journal of Wildland Fire*, 15(3):319–345, 2006.

[83] G Levkine. Prewitt, sobel and scharr gradient 5x5 convolution matrices. *Image Process. Articles, Second Draft*, 2012.

[84] Jiaojiao Li, Qian Du, Yunsong Li, and Wei Li. Hyperspectral image classification with imbalanced data based on orthogonal complement subspace projection. *IEEE Transactions on Geoscience and Remote Sensing*, 56(7):3838–3851, 2018.

[85] Shutao Li, Weiwei Song, Leyuan Fang, Yushi Chen, Pedram Ghamisi, and Jon Atli Benediktsson. Deep learning for hyperspectral image classification: An overview. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):6690–6709, 2019.

[86] Wei Li, Chen Chen, Mengmeng Zhang, Hengchao Li, and Qian Du. Data augmentation for hyperspectral image classification with deep CNN. *IEEE Geosci. Remote Sens. Lett.*, 16(4):593–597, 2018.

[87] Wei Li and Qian Du. Gabor-filtering-based nearest regularized subspace for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(4):1012–1022, 2014.

[88] Wei Li, Guodong Wu, Fan Zhang, and Qian Du. Hyperspectral image classification using deep pixel-pair features. *IEEE Trans. Geosci. Remote Sens.*, 55(2):844–853, 2016.

[89] Giorgio Licciardi, Prashanth Reddy Marpu, Jocelyn Chanussot, and Jon Atli Benediktsson. Linear versus nonlinear PCA for the classification of hyperspectral data based on the extended morphological profiles. *IEEE Geoscience and Remote Sensing Letters*, 9(3):447–451, 2012.

[90] Yazhou Liu, Guo Cao, Quansen Sun, and Mel Siegel. Hyperspectral classification via deep networks and superpixel segmentation. *Int. J. Remote Sens.*, 36(13):3459–3482, 2015.

[91] Javier López-Fandiño, Pablo Quesada-Barriuso, Dora B Heras, and Francisco Argüello. Efficient ELM-based techniques for the classification of hyperspectral remote sensing images on commodity GPUs. *IEEE Journal of Selected topics in applied earth observations and remote sensing*, 8(6):2884–2893, 2015.

[92] Ettore Lopinto and Cristina Ananasso. The prisma hyperspectral mission. In *Proceedings of the 33rd EARSeL Symposium, Towards Horizon*, 2020.

[93] Wenjing Lv and Xiaofei Wang. Overview of hyperspectral image classification. *Journal of Sensors*, 2020, 2020.

[94] Javier Marcello, Dionisio Rodriguez-Esparragon, and Daniel Moreno. Comparison of land cover maps using high resolution multispectral and hyperspectral imagery. In *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, pages 7312–7315. IEEE, 2018.

[95] George A Maul and Howard R Gordon. On the use of the earth resources technology satellite (LANDSAT-1) in optical oceanography. *Remote Sensing of Environment*, 4:95–128, 1975.

[96] Aaron E Maxwell, Timothy A Warner, and Fang Fang. Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, 39(9):2784–2817, 2018.

[97] Rahul Mazumder, Trevor Hastie, and Robert Tibshirani. Spectral regularization algorithms for learning large incomplete matrices. *The Journal of Machine Learning Research*, 11:2287–2322, 2010.

[98] Michael McCool, James Reinders, and Arch Robison. *Structured parallel programming: patterns for efficient computation*. Elsevier, 2012.

[99] Alan D McNaught, Andrew Wilkinson, et al. *Compendium of chemical terminology*, volume 1669. Blackwell Science Oxford, 1997.

[100] Grégoire Mercier and Marc Lennon. Support vector machines for hyperspectral image classification with spectral-based kernels. In *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No. 03CH37477)*, volume 1, pages 288–290. IEEE, 2003.

[101] Micasense rededge mx multispectral camera. https://micasense.com/rededge-mx/. Last accessed on 13/10/2020.

[102] Fardin Mirzapour and Hassan Ghassemian. Hyperspectral image classification using profiles based on partial differential equations. In *Electrical Engineering (ICEE), 2015 23rd Iranian Conference on*, pages 288–292. IEEE, 2015.

[103] Jakub Nalepa, Michal Myller, and Michal Kawulok. Training-and test-time data augmentation for hyperspectral image segmentation. *IEEE Geosci. Remote Sens. Lett.*, 2019.

[104] Nasser M Nasrabadi. Hyperspectral target detection: An overview of current and future challenges. *IEEE Signal Processing Magazine*, 31(1):34–44, 2013.

[105] Gerald R North, John A Pyle, and Fuqing Zhang. *Encyclopedia of atmospheric sciences*, volume 1. Elsevier, 2014.

[106] Álvaro Ordóñez, Francisco Argüello, and Dora B Heras. GPU accelerated FFT-based registration of hyperspectral scenes. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(11):4869–4878, 2017.

[107] John D Owens, Mike Houston, David Luebke, Simon Green, John E Stone, and James C Phillips. GPU computing. *Proceedings of the IEEE*, 96(5):879–899, 2008.

[108] ME Paoletti, JM Haut, J Plaza, and A Plaza. Deep learning classifiers for hyperspectral imaging: A review. *ISPRS J. Photogramm. Remote Sens.*, 158:279–317, 2019.

[109] Lucas Parra, Clay Spence, Paul Sajda, Andreas Ziehe, and Klaus-Robert Müller. Unmixing hyperspectral data. *Advances in neural information processing systems*, 12, 1999.

[110] Donald L Pavia, Gary M Lampman, George S Kriz, and James A Vyvyan. *Introduction to spectroscopy*. Cengage learning, 2014.

[111] Pietro Perona and Jitendra Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.

[112] Martino Pesaresi and Jon Atli Benediktsson. A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(2):309–320, 2001.

[113] Antonio Plaza, Jon Atli Benediktsson, Joseph W Boardman, Jason Brazile, Lorenzo Bruzzone, Gustavo Camps-Valls, Jocelyn Chanussot, Mathieu Fauvel, Paolo Gamba, Anthony Gualtieri, et al. Recent advances in techniques for hyperspectral image processing. *Remote sensing of environment*, 113:S110–S122, 2009.

[114] Antonio Plaza, Pablo Martinez, Rosa Pérez, and Javier Plaza. Spatial/spectral endmember extraction by multidimensional morphological operations. *IEEE transactions on geoscience and remote sensing*, 40(9):2025–2041, 2002.

[115] Cle Pohl and John L Van Genderen. Review article multisensor image fusion in remote sensing: concepts, methods and applications. *International journal of remote sensing*, 19(5):823–854, 1998.

[116] Tanu Priya, Saurabh Prasad, and Hao Wu. Superpixels for spatially reinforced Bayesian classification of hyperspectral images. *IEEE Geosci. Remote Sens. Lett.*, 12(5):1071–1075, 2015.

[117] Pablo Quesada-Barriuso, Francisco Argüello, and Dora B Heras. Spectral–spatial classification of hyperspectral images using wavelets and extended morphological profiles. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(4):1177–1185, 2014.

[118] Pablo Quesada-Barriuso, Dora Blanco Heras, and Francisco Argüello. GPU accelerated waterpixel algorithm for superpixel segmentation of hyperspectral images. *The Journal of Supercomputing*, 77(9):10040–10052, 2021.

[119] John A Richards and JA Richards. *Remote sensing digital image analysis*, volume 3. Springer, 1999.

[120] Bahram Salehi, Y Ming Zhong, and V Dey. A review of the effectiveness of spatial information used in urban land cover classification of vhr imagery. *International Journal of Geoinformatics*, 8(2):35, 2012.

[121] Joy Sanyal and Xi Xi Lu. Application of remote sensing in flood management with special reference to monsoon asia: a review. *Natural Hazards*, 33(2):283–301, 2004.

[122] Robert R Schaller. Moore's law: past, present and future. *IEEE spectrum*, 34(6):52–59, 1997.

[123] Gary Shaw and Dimitris Manolakis. Signal processing for hyperspectral image exploitation. *IEEE Signal processing magazine*, 19(1):12–16, 2002.

[124] Caileigh Shoot, Hans-Erik Andersen, L Monika Moskal, Chad Babcock, Bruce D Cook, and Douglas C Morton. Classifying forest type in the national forest inventory context with airborne hyperspectral and lidar data. *Remote Sensing*, 13(10):1863, 2021.

[125] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *J. Big Data*, 6(1):60, 2019.

[126] Viktor Slavkovikj, Steven Verstockt, Wesley De Neve, Sofie Van Hoecke, and Rik Van de Walle. Hyperspectral image classification with convolutional neural networks. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 1159–1162, 2015.

[127] Dimitris G Stavrakoudis, Georgia N Galidaki, Ioannis Z Gitas, and John B Theocharis. A genetic fuzzy-rule-based classifier for land cover classification from hyperspectral imagery. *IEEE Transactions on Geoscience and Remote sensing*, 50(1):130–148, 2011.

[128] Kun Tan, Junpeng Zhang, Qian Du, and Xuesong Wang. GPU parallel implementation of support vector machines for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(10):4647–4656, 2015.

[129] Yuliya Tarabalka, Jón Atli Benediktsson, and Jocelyn Chanussot. Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques. *IEEE transactions on geoscience and remote sensing*, 47(8):2973–2987, 2009.

[130] Yuliya Tarabalka, Jón Atli Benediktsson, Jocelyn Chanussot, and James C Tilton. Multiple spectral–spatial classification approach for hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):4122–4132, 2010.

[131] Yuliya Tarabalka, Jocelyn Chanussot, and Jon Atli Benediktsson. Segmentation and classification of hyperspectral images using watershed transformation. *Pattern Recognition*, 43(7):2367–2379, 2010.

[132] Yuliya Tarabalka, Mathieu Fauvel, Jocelyn Chanussot, and Jón Atli Benediktsson. SVM-and MRF-based method for accurate classification of hyperspectral images. *IEEE Geoscience and Remote Sensing Letters*, 7(4):736–740, 2010.

[133] Benjamin Thomas, Grégory David, Christophe Anselmo, Jean-Pierre Cariou, Alain Miffre, and Patrick Rairoux. Remote sensing of atmospheric gases with optical correlation spectroscopy and lidar: first experimental results on water vapor profile measurements. *Applied Physics B*, 113(2):265–275, 2013.

[134] Neil C Thompson, Kristjan Greenewald, Keeheon Lee, and Gabriel F Manso. Deep learning's diminishing returns: The cost of improvement is becoming unsustainable. *IEEE Spectrum*, 58(10):50–55, 2021.

[135] Xin-Yi Tong, Gui-Song Xia, Qikai Lu, Huangfeng Shen, Shengyang Li, Shucheng You, and Liangpei Zhang. Land-cover classification with high-resolution remote sensing images using transferable deep models. *Remote Sensing of Environment, doi: 10.1016/j.rse.2019.111322*, 2020.

[136] Charles Toth and Grzegorz Jóźków. Remote sensing platforms and sensors: A survey. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115:22–36, 2016.

[137] Olga Troyanskaya, Michael Cantor, Gavin Sherlock, Pat Brown, Trevor Hastie, Robert Tibshirani, David Botstein, and Russ B Altman. Missing value estimation methods for DNA microarrays. *Bioinformatics*, 17(6):520–525, 2001.

[138] Fuan Tsai and William D Philpot. A derivative-aided hyperspectral image analysis system for land-cover classification. *IEEE Transactions on Geoscience and Remote Sensing*, 40(2):416–425, 2002.

[139] Dean M Tullsen, Susan J Eggers, and Henry M Levy. Simultaneous multithreading: maximizing on-chip parallelism. In *25 years of the international symposia on Computer architecture (selected papers)*, pages 533–544, 1998.

[140] Ava Vali, Sara Comai, and Matteo Matteucci. Deep learning for land use and land cover classification based on hyperspectral and multispectral earth observation data: A review. *Remote Sensing*, 12(15):2495, 2020.

[141] S. van Buuren. *Flexible Imputation of Missing Data*. Chapman & Hall/CRC Interdisciplinary Statistics. CRC Press LLC, 2018.

[142] Guido Van Rossum and Fred L Drake Jr. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.

[143] Santiago Velasco-Forero and Vidya Manian. Improving hyperspectral image classification using spatial preprocessing. *IEEE Geoscience and Remote Sensing Letters*, 6(2):297–301, 2009.

[144] Alberto Villa, Jón Atli Benediktsson, Jocelyn Chanussot, and Christian Jutten. Hyperspectral image classification with independent component discriminant analysis. *IEEE transactions on Geoscience and remote sensing*, 49(12):4865–4876, 2011.

[145] Vasily Volkov. Better performance at lower occupancy. In *Proceedings of the GPU technology conference, GTC*, volume 10, page 16. San Jose, CA, 2010.

[146] Chunying Wang, Baohua Liu, Lipeng Liu, Yanjun Zhu, Jialin Hou, Ping Liu, and Xiang Li. A review of deep learning used in the hyperspectral image analysis for agriculture. *Artificial Intelligence Review*, 54(7):5205–5253, 2021.

[147] Xue Wang, Kun Tan, Qian Du, Yu Chen, and Peijun Du. Caps-triplegan: Gan-assisted capsnet for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):7232–7245, 2019.

[148] Joachim Weickert. *Anisotropic diffusion in image processing*, volume 1. Teubner Stuttgart, 1998.

[149] Andrew P Witkin. Scale-space filtering. In *Readings in Computer Vision*, pages 329–332. Elsevier, 1987.

[150] Hao Wu and Saurabh Prasad. Semi-supervised deep learning using pseudo labels for hyperspectral image classification. *IEEE Transactions on Image Processing*, 27(3):1259–1270, 2017.

[151] Zebin Wu, Jiafu Liu, Antonio Plaza, Jun Li, and Zhihui Wei. GPU implementation of composite kernels for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters*, 12(9):1973–1977, 2015.

[152] Zebin Wu, Linlin Shi, Jun Li, Qicong Wang, Le Sun, Zhihui Wei, Javier Plaza, and Antonio Plaza. GPU parallel implementation of spatially adaptive hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(4):1131–1143, 2017.

[153] Zebin Wu, Qicong Wang, Antonio Plaza, Jun Li, Le Sun, and Zhihui Wei. Parallel spatial–spectral hyperspectral image classification with sparse representation and markov random fields on GPUs. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(6):2926–2938, 2015.

[154] Michael A Wulder, Thomas R Loveland, David P Roy, Christopher J Crawford, Jeffrey G Masek, Curtis E Woodcock, Richard G Allen, Martha C Anderson, Alan S Belward, Warren B Cohen, et al. Current status of landsat program, science, and applications. *Remote sensing of environment*, 225:127–147, 2019.

[155] Junshi Xia, Jocelyn Chanussot, Peijun Du, and Xiyan He. Spectral–spatial classification for hyperspectral data using rotation forests with local feature extraction and markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5):2532–2546, 2014.

[156] Chen Xing, Li Ma, and Xiaoquan Yang. Stacked denoise autoencoder based feature extraction and classification for hyperspectral images. *Journal of Sensors*, 2016, 2016.

[157] Bing Xu and Peng Gong. Land-use/land-cover classification with multispectral and hyperspectral eo-1 data. *Photogrammetric Engineering & Remote Sensing*, 73(8):955–965, 2007.

[158] Yan Xu, Congling Zhang, Ruijiao Jiang, Zifei Wang, Mengchen Zhu, and Guochun Shen. Uav-based hyperspectral images and monitoring of canopy tree diversity. *Biodiversity Science*, 29(5):647, 2021.

[159] Xiaofei Yang, Yunming Ye, Xutao Li, Raymond YK Lau, Xiaofeng Zhang, and Xiaohui Huang. Hyperspectral image classification with deep learning models. *IEEE Transactions on Geoscience and Remote Sensing*, 56(9):5408–5423, 2018.

[160] Xingrui Yu, Xiaomin Wu, Chunbo Luo, and Peng Ren. Deep learning in remote sensing scene classification: A data augmentation enhanced convolutional neural network framework. *Giscience & Remote Sens.*, 54(5):741–758, 2017.

[161] Ayomide Yusuf and Shadi Alawneh. A survey of GPU implementations for hyperspectral image classification in remote sensing. *Canadian Journal of Remote Sensing*, 44(5):532–550, 2018.

[162] Jaime Zabalza, Jinchang Ren, Zheng Wang, Stephen Marshall, and Jun Wang. Singular spectrum analysis for effective feature extraction in hyperspectral imaging. *IEEE Geoscience and Remote Sensing Letters*, 11(11):1886–1890, 2014.

[163] Alina Zare and KC Ho. Endmember variability in hyperspectral analysis: Addressing spectral variability during spectral unmixing. *IEEE Signal Processing Magazine*, 31(1):95–104, 2013.

[164] Ying Zhan, Dan Hu, Yuntao Wang, and Xianchuan Yu. Semisupervised hyperspectral image classification based on generative adversarial networks. *IEEE Geoscience and Remote Sensing Letters*, 15(2):212–216, 2017.

[165] Bing Zhang, Shanshan Li, Xiuping Jia, Lianru Gao, and Man Peng. Adaptive markov random field approach for classification of hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 8(5):973–977, 2011.

[166] Haokui Zhang, Ying Li, Yuzhu Zhang, and Qiang Shen. Spectral-spatial classification of hyperspectral imagery using a dual-channel convolutional neural network. *Remote Sens. Lett.*, 8(5):438–447, 2017.

[167] Mengmeng Zhang, Wei Li, and Qian Du. Diverse region-based CNN for hyperspectral image classification. *IEEE Trans. Image Process.*, 27(6):2623–2634, 2018.

[168] Wendy W Zhang and Shobha Sriharan. Using hyperspectral remote sensing for land cover classification. In *Multispectral and Hyperspectral Remote Sensing Instruments and Applications II*, volume 5655, pages 261–270. SPIE, 2005.

[169] Ping Zhong and Runsheng Wang. Jointly learning the hybrid crf and mlr model for simultaneous denoising and classification of hyperspectral imagery. *IEEE Transactions on Neural Networks and Learning Systems*, 25(7):1319–1334, 2014.

[170] 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, 2017.

[171] Lin Zhu, Yushi Chen, Pedram Ghamisi, and Jón Atli Benediktsson. Generative adversarial networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.*, 56(9):5046–5063, 2018.

List of Figures

List of Tables

The objective of this PhD thesis is the development of spatial-spectral information extraction techniques for supervised classification tasks, both by means of classical models and those based on deep learning, to be used in the classification of land use or land cover (LULC) multi- and hyper-spectral images obtained by remote sensing. The main goal is the efficient application of these techniques, so that they are able to obtain satisfactory classification results with a low use of computational resources and low execution time.