MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21d/2023/12.21.15.45-TDI

MACHINE LEARNING E HASHING PARA IDENTIFICAÇÃO DE IMAGENS DE SENSORIAMENTO REMOTO BASEADA EM CONTEÚDO

Marcos Lima Rodrigues

Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Thales Sehn Körting, e Gilberto Ribeiro de Queiroz, aprovada em 15 de dezembro de 2023.

URL do documento original:
<http://urlib.net/8JMKD3MGP3W34T/4ADRCA2>

INPE
São José dos Campos
2023

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE
Coordenação de Ensino, Pesquisa e Extensão (COEPE)
Divisão de Biblioteca (DIBIB)
CEP 12.227-010
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/7348
E-mail: pubtc@inpe.br

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21d/2023/12.21.15.45-TDI

MACHINE LEARNING E HASHING PARA IDENTIFICAÇÃO DE IMAGENS DE SENSORIAMENTO REMOTO BASEADA EM CONTEÚDO

Marcos Lima Rodrigues

Tese de Doutorado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Thales Sehn Körting, e Gilberto Ribeiro de Queiroz, aprovada em 15 de dezembro de 2023.

URL do documento original:
<http://urlib.net/8JMKD3MGP3W34T/4ADRCA2>

INPE
São José dos Campos
2023

Dados Internacionais de Catalogação na Publicação (CIP)

Rodrigues, Marcos Lima.

R618m Machine learning e hashing para identificação de imagens de sensoriamento remoto baseada em conteúdo / Marcos Lima Rodrigues. – São José dos Campos : INPE, 2023.

xxiv + 108 p. ; (sid.inpe.br/mtc-m21d/2023/12.21.15.45-TDI)

Tese (Doutorado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2023.

Orientadores : Drs. Thales Sehn Körting, e Gilberto Ribeiro de Queiroz.

1. Recuperação de imagens baseada em conteúdo. 2. Redes neurais convolucionais. 3. EuroSAT. 4. Uso e Cobertura da terra. 5. Cerrado. I.Título.

CDU 004.62

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

DEFESA FINAL DE TESE MARCOS LIMA RODRIGUES
BANCA Nº 308/2023, REGISTRO 101214/2019

No dia 15 de dezembro de 2023, às 09:00h, no Auditório da OBT , o(a) aluno(a) mencionado(a) acima defendeu seu trabalho final (apresentação oral seguida de arguição) perante uma Banca Examinadora, cujos membros estão listados abaixo. O(A) aluno(a) foi APROVADO(A) pela Banca Examinadora, por unanimidade, em cumprimento ao requisito exigido para obtenção do Título de Doutor em Computação Aplicada, com a exigência de que o trabalho final a ser publicado deverá incorporar as correções sugeridas pela Banca Examinadora, com revisão pelo(s) orientador(es).

Título: "Machine Learning e Hashing para Identificação de Imagens de Sensoriamento Remoto Baseada em Conteúdo".

Membros da Banca:

Dra. Karine Reis Ferreira Gomes – Presidente – INPE
Dr. Thales Sehn Körting – Orientador (a) – INPE
Dr. Gilberto Ribeiro de Queiroz – Orientador (a) – INPE
Dr. Rogério Galante Negri – Membro Externo – UNESP
Dr. Alexandre Noma – Membro Externo – UFABC

Documento assinado eletronicamente por **Karine Reis Ferreira Gomes**, **Tecnologista**, em 19/12/2023, às 11:19 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Gilberto Ribeiro de Queiroz**, **Tecnologista**, em 19/12/2023, às 11:59 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Rogério Galante Negri (E)**, **Usuário Externo**, em 19/12/2023, às 12:00 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Alexandre Noma (E)**, **Usuário Externo**, em 19/12/2023, às 12:58 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Thales Sehn Korting**, **Pesquisador**, em 20/12/2023, às 12:53 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.mcti.gov.br/verifica.html, informando o código verificador **11599023** e o código CRC **25837546**.

Referência: Processo nº 01340.010462/2023-41 SEI nº 11599023

"Ninguém é tão sábio que não tenha algo para aprender e nem tão tolo que não tenha algo para ensinar".

Blaise Pascal

A meus pais ***Nelson*** *e* ***Fátima*** *(in memorian), a meus irmãos* ***Alessandra, Daiana e Brendon****, e a minha amada esposa* ***Cidinha***

AGRADECIMENTOS

Agradeço a Deus fonte de toda força e verdade, aos meus orientadores Dr. Thales Sehn Körting e Dr. Gilberto Ribeiro de Queiroz pela dedicação e paciência. A todos os amigos e colegas alunos da pós-graduação com quem tive a oportunidade de estudar e trabalhar ao longo desses últimos anos, são verdadeiros exemplos de que com esforço e perseverança grandes coisas são possíveis. A toda a equipe e corpo docente da pós-graduação em Computação Aplicada (CAP) do Instituto Nacional de Pesquisas Espaciais (INPE).

Pela infraestrutura computacional e de dados utilizados neste trabalho, agradeço ao INPE, ao subprojeto *Brazil Data Cube* que faz parte do Projeto de Monitoramento Ambiental dos Biomas Brasileiros, financiado pelo Fundo Amazônia, por meio da colaboração financeira BNDES e FUNCATE nº 17.2.0536.1 e ao projeto *Development of systems to prevent forest fires and monitor vegetation cover in the Brazilian Cerrado* financiado pelo *The World Bank* #P143185 através do *Forest Investment Program* (FIP).

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

RESUMO

Neste trabalho é apresentado o desenvolvimento de uma solução (*framework*) para busca e recuperação de imagens de satélite baseadas em conteúdo, com potencial para aplicação no escopo de grandes conjuntos de dados. A área de sensoriamento remoto (SR) para observação da Terra tem experimentado um grande desenvolvimento na última década, dando origem a era do *Remote Sensing Big Data* (RSBD), tornando desafiadora a tarefa de recuperar imagens úteis nesse grande volume de dados, por exemplo, que possam ser usadas para estudos de uso e cobertura da terra no Cerrado brasileiro. Nesse contexto, o desenvolvimento de sistemas baseados em *Content-Based Image Retrieval* (CBIR) apoiado por métodos de *Deep Learning* como as *Convolutional Neural Networks* (CNNs), têm sido empregados com sucesso a dados multifontes e multiespectrais (MS). As arquiteturas *Deep Hashing Neural Networks* (DHNNs) empregam CNNs para extração de atributos de imagens e conversão desses atributos em códigos binários (*hash codes*) para criação de um espaço métrico otimizado para CBIR no escopo do RSBD. A *Metric-Learning-Based Deep Hashing Network* (MiLaN) representa o estado da arte desse tipo de arquitetura, baseada na combinação de três funções de perda que permitem o aprendizado de um espaço métrico ideal para a recuperação de imagens baseada em conteúdo (*Semantic-Based Metric Space*). Originalmente a rede MiLaN adotou como módulo de extração de características das imagens (*backbone*) a rede Inception V3 pré-treinada com dados fora do domínio do SR (ImageNet), isso implica em limitações devido a diferenças típicas entre as imagens como a resolução espacial e influência da atmosfera nas imagens orbitais. O *framework* proposto possibilitou avanços em relação à abordagem original da MiLaN ao adotar um novo *backbone* baseado na ResNet-50 e realizar o processo de ajuste dessas arquiteturas (MiLaN+ResNet-50) através do *fine-tuning* baseado em imagens satelitais MS. Esta afirmação é evidenciada pelos resultados expressivos alcançados para tarefa CBIR medidos através da métrica *mean Average Precision - mAP*, o desempenho global baseado nas 100 primeiras imagens recuperadas (mAP@100) foi de 99,8873% para o conjunto EuroSAT MS (Sentinel 2 - 13 bandas). De maneira particular foi demonstrado que os dados MS fornecem informações semânticas de qualidade durante o processo de extração de características usando a ResNet-50, contribuindo assim para correção de erros em relação à discriminação de imagens que apresentam padrões geométricos (Áreas Industriais/Residenciais) e de textura (Floresta, Pastagem e Culturas Permanente) similares quando utilizado somente as bandas RGB das imagens de média resolução do conjunto EuroSAT. O desempenho para o conjunto EuroSAT MS superou o apresentado por outros métodos do estado da arte para realização de CBIR, inclusive utilizando imagens aéreas de alta resolução espacial do conjunto *Aerial Image Dataset* (AID).

Palavras-chave: Recuperação de Imagens Baseada em Conteúdo. Redes Neurais Convolucionais. EuroSAT. Uso e Cobertura da terra. Cerrado.

MACHINE LEARNING AND HASHING FOR CONTENT-BASED IMAGE RETRIEVAL (CBIR) OF REMOTE SENSING IMAGES

ABSTRACT

This work presents the development of a framework for searching and retrieving content-based satellite images, with potential for application in the scope of large datasets. The area of remote sensing (RS) for Earth observation has experienced great development in the last decade, giving rise to the era of Remote Sensing Big Data (RSBD), making the task of retrieving useful images from this large volume of data challenging, for example, that can be used for studies of land use and land cover in the Brazilian Cerrado. In this context, the development of systems based on Content-Based Image Retrieval (CBIR) supported by Deep Learning methods such as Convolutional Neural Networks (CNNs), have been successfully applied to multisource and multispectral (MS) data. Deep Hashing Neural Networks (DHNNs) architectures employ CNNs to extract image attributes and convert these attributes into binary codes (hash codes) to create a metric space optimized for CBIR within the scope of RSBD. The Metric-Learning-Based Deep Hashing Network (MiLaN) represents the state of the art of this type of architecture, based on the combination of three loss functions that allow the learning of a space ideal metric for CBIR (Semantic-Based Metric Space). Originally, the MiLaN network adopted the Inception V3 network pre-trained with data outside the RS domain (ImageNet) as an image feature extraction module (backbone), this implies limitations due to typical differences between images such as the spatial resolution and influence of the atmosphere on orbital images. The proposed framework enabled advances in the original MiLaN approach by adopting a new backbone based on ResNet-50 and carrying out the adjustment process of these architectures (MiLaN+ResNet-50) through fine-tuning based on MS satellite images. This statement is evidenced by the expressive results achieved for the CBIR task measured using the mean Average Precision (mAP) metric, the global performance based on the top-100 recovered images (mAP@100) was 99.8873% for the set EuroSAT MS (Sentinel 2 - 13 bands). In particular, it was demonstrated that MS data provides quality semantic information during the feature extraction process using ResNet-50, thus contributing to error correction concerning the discrimination of images that present geometric patterns (Industrial/Residential Areas) and texture (Forest, Pasture and Permanent Crops) similar when using only the RGB bands of the medium resolution images from the EuroSAT set. The performance for the EuroSAT MS dataset surpassed that presented by other state-of-the-art methods for carrying out CBIR, including using high spatial resolution aerial images from the Aerial Image Dataset (AID).

Keywords: Content-Based Image Retrieval (CBIR). Deep Hashing Neural Network (DHNN). EuroSAT. Land Use and Land Cover (LULC). The Brazilian Savanna (Cerrado).

LISTA DE FIGURAS

Pág.

1.1 Imagem contendo cicatriz de queimada na região Amazônica com focos ativos detectados em 06/08/2023 às 13:08 UTC pelo sensor MODIS do satélite Terra. . . . 3

2.1 Componentes básicos de um sistema CBIR. . . . 11
2.2 Representação do espaço de Hamming através de um cubo com vértices correspondentes a 3 bits. . . . 15
2.3 Ilustração do uso de *hash codes* para CBIR. . . . 16
2.4 Exemplo qualitativo da *hash table* com *hash code* com 2 bits de comprimento. . . . 18
2.5 Rede convolucional para identificação de caracteres (processamento de imagem). . . . 21
2.6 Representação da conexão entre a camada oculta e o campo receptivo local (a), além da disposição tridimensional dos mapas de atributos (b). . . 22
2.7 Representação da camada de subamostragem com filtro 2x2, *stride* de 2 e função de *Pooling*. . . . 24
2.8 Representação das componentes de uma rede DHNN, incluindo as unidades de extração de atributos (DFLNN) e aprendizagem *hashing* (HLNN). 27
2.9 Aprendizado semântico baseado nas redes Inception V3 e MHCLN para construção de um espaço métrico adequado para CBIR utilizando *hash codes*. . . . 29
2.10 Funções de custo otimizadas para aprendizado semântico utilizadas pela MHCLN. . . . 30
2.11 Projeção bidimensional do espaço de parâmetros (espaço métrico) de k-dimensões, sendo $k = 20$, referente ao comprimento do vetor de *hash codes* gerados a partir de imagens do conjunto UCMD[3]. . . . 32

3.1 Amostra de imagens dos tipos de uso e cobertura da terra providos pelo conjunto EuroSAT. . . . 34
3.2 Amostra de imagens dos tipos de uso e cobertura da terra providos pelo conjunto AID. . . . 35
3.3 Área de estudo para caracterização de uso e cobertura da terra localizada na região do MATOPIBA dentro do bioma Cerrado (esquerda). Destaque para composição RGB do tile 089097[1] do cubo de dados Sentinel 2-16D (composição temporal de 16 dias) com início em 19 dezembro 2018 (direita). 37

3.4 Fluxograma do WLTS para coleta de amostras de uso e cobertura da terra. 38
3.5 Esquema baseado no serviço WLTS para coleta de amostras de uso e cobertura da terra para os *patches* da área de estudo. 39
3.6 Distribuição das amostras de uso e cobertura da terra adquiridas através do serviço WLTS para a área de estudo. 40
3.7 Treinamento ResNet-152 com inicialização aleatória dos pesos (a) e pré-treinada com o conjunto ImageNet (b). 45
3.8 *Framework* utilizado para o treinamento dos *backbones*, aprendizado do espaço métricos (MiLaN) e recuperação de imagens baseada na medida de similaridade (distância de Hamming). 48
3.9 Exemplo de imagens do conjunto AID com *data augmentation*. 50
3.10 Exemplos de imagens EuroSAT corrigidas com o CLAHE. 51
3.11 Construção do espaço métrico (4^a etapa) e tarefa CBIR (5^a etapa) usando a rede MiLaN. 52
3.12 Exemplo de identificação de uso e cobertura da terra no Cerrado usando CBIR (MiLaN). 53
3.13 Concordância entre os tipos de uso e cobertura da terra identificados nas imagens do conjunto EuroSAT e o mapeamento feito pelo IBGE. 54

4.1 Diagrama experimental das etapas necessárias para treinamento, teste e avaliação da MiLaN com os conjuntos AID e EuroSAT utilizando os *backbones* Inception V3 e ResNet-50. 56
4.2 Resultados da recuperação de imagens com a rede MiLaN para as classes *Permanent Crop* e *Highway* do conjunto EuroSAT, ordenados por grau de similaridade. 61
4.3 Projeção do espaço métrico criado pela rede MiLaN para os dados dos conjuntos AID e EuroSAT (RGB/MS). 65
4.3 LULC para o tile 089097 usando Inception V3 e *patches* RGB de 128×128 pixels do cubo de dados Sentinel identificada com base na máxima similaridade em relação a dados do conjunto EuroSAT. 70
4.4 Detalhe da identificação de áreas industriais no tile 089097 usando Inception V3 e *patches* RGB de 128×128 pixels do cubo de dados Sentinel. 71
4.5 Matriz de confusão resultado da identificação de LULC no Cerrado usando CBIR (MiLaN+ResNet-50) a partir de *patches* com 64×64 pixels multiespectrais. 72

5.1 Visão geral do *framework Improved Metadata from Remote Sensing Images* (IMRSI) para busca e recuperação de imagens baseadas em conteúdo. 77

5.2 Alvos identificados pelo ReSIIM em uma cena Landsat 8 órbita/ponto 227/058 de 18/07/2017. 80

A.1 Amostra de imagens do conjunto BigEarthNet com múltiplos rótulos de uso e cobertura da terra. 97
A.2 Comparação do desempenho da classificação de imagens do conjunto BigEarthNet realizada por vários modelos de *Deep Learning*. 99
A.3 Amostras de imagens do subconjunto BigEarthNet (*single label*). 101
A.4 Treinamento da rede MiLaN com o subconjunto BigEarthNet (*single label*). 102
A.5 Exemplos de imagens BigEarthNet com problemas de rótulo e ruidosas. . 103
A.6 Teste para recuperação de imagens do tipo *Mixed Forest* do conjunto BigEarthNet com a rede MiLaN (pior desempenho). 105
A.7 Teste para recuperação de imagens do tipo (*Pastures*) do conjunto BigEarthNet com a rede MiLaN (melhor desempenho). 106

LISTA DE TABELAS

Pág.

3.1 Matriz de confusão de duas classes (**P**ositivo/**N**egativo). 41

4.1 Desempenho do CBIR (MiLaN+*backbones*) de imagens aéreas (AID) RGB, considerando k imagens de cada classe recuperadas pela menor distância de Hamming sendo $k = \{20, 50, 100\}$. 57

4.2 Desempenho do CBIR (MiLaN+*backbones*) de imagens satelitais EuroSAT RGB, considerando k imagens de cada classe recuperadas pela menor distância de Hamming sendo $k = \{20, 50, 100\}$. 58

4.3 Teste de Wilcoxon para avaliar a significância estatística da diferença entre o desempenho do CBIR utilizando como *backbones* as redes Inception V3 e ResNet-50 para ambos os conjuntos de imagens AID/EuroSAT RGB. 59

4.4 Desempenho para classificação de imagens dos conjuntos EuroSAT (RGB/MS) utilizando a ResNet-50 pré-treinada com ImageNet. 62

4.5 Desempenho do CBIR de imagens aéreas (AID) e satelitais RGB e Multiespectrais (EuroSAT)*, considerando k imagens de cada classe recuperadas pela menor distância de Hamming sendo $k = 20, 50, 100$. 63

4.6 Desempenho global (mAP) do CBIR de imagens aéreas (AID) e satelitais RGB/Multiespectral (EuroSAT), considerando k imagens recuperadas pela menor distância de Hamming sendo $k = 20, 50, 100$. 64

4.7 Desempenho para identificação de LULC no Cerrado baseado em CBIR utilizando combinações da MiLaN+Inception V3*/ResNet-50, *patches* com 64×64/128×128 pixels e dados RGB/MS. 69

4.8 Comparação do desempenho para identificação de LULC no Cerrado baseado em CBIR e classificação. 73

A.1 Desempenho da recuperação de imagens global para os conjuntos UCMD, AID e BigEarthNet, considerando $k = 20, 50, 100$. 103

A.2 Desempenho da recuperação de imagens por tipo de uso e cobertura da terra do conjunto BigEarthNet, considerando $k = 20, 50, 100$. 104

LISTA DE ABREVIATURAS E SIGLAS

AHE	–	Adaptive Histogram Equalization
AID	–	Aerial Image Dataset
BDC	–	Brazil Data Cube
CBIR	–	Content-Based Image Retrieval
CHT	–	Circular Hough Transform
CLAHE	–	Contrast Limited Adaptive Histogram Equalization
CLC	–	CORINE Land Cover
COG	–	Cloud Optimized GeoTIFF
CORINE	–	Coordination of Information on the Environment
DA	–	Data Augmentation
DFLNN	–	Deep Feature Learning Neural Network
DHN	–	Deep Hashing Network
DHNN	–	Deep Hashing Neural Network
DL	–	Deep Learning
DPSH	–	Deep Pairwise-Supervised Hashing
DSH	–	Deep Supervised Hashing
HLNN	–	Hashing Learning Neural Network
HS	–	Hough Space
HT	–	Hough Transform
IBGE	–	Instituto Brasileiro de Geografia e Estatística
IMRSI	–	Improved Metadata from Remote Sensing Images
KSLSH	–	Kernel-based Supervised Hashing
KULSH	–	Kernel-based Unsupervised Hashing
LRAP	–	Label Ranking Average Precision
LSH	–	Locality-Sensitive Hash
LSTM	–	Long-Short Term Memory
LULC	–	Land Use and Land Cover
MHCLN	–	Metric and Hash-Code Learning Network
MiLaN	–	Metric-Learning-Based Deep Hashing Network
ML	–	Machine Learning
MS	–	Multiespectral
MSI	–	MultiSpectral Instrument
RF	–	Random Forest
RGB	–	Red Green Blue
RR	–	Retroalimentação de Relevância
RS	–	Remote Sensing
RSBD	–	Remote Sensing Big Data
SIFT	–	Scale-Invariant Feature Transform
SR	–	Sensoriamento Remoto
STAC	–	SpatioTemporal Asset Catalog
SVM	–	Support Vector Machine
UCMD	–	University of California Merced Land Use Dataset
WLTS	–	Web Land Trajectory Service

SUMÁRIO

Pág.

1 INTRODUÇÃO 1
1.1 Problema 2
1.2 Hipótese 3
1.3 Objetivos 4
1.4 Contribuições 5

2 FUNDAMENTAÇÃO TEÓRICA 9
2.1 Visão geral sobre CBIR 9
2.1.1 Funções *hash* para CBIR aplicadas a RSBD 12
2.1.2 CBIR para solução do problema de busca k-NN usando funções hash . 16
2.2 Introdução às redes neurais convolucionais 19
2.2.1 Arquitetura 20
2.2.1.1 Camada de convolução 20
2.2.1.2 Camada de subamostragem 23
2.2.1.3 Camada totalmente conectada 24
2.3 Big data CBIR com redes neurais convolucionais e hashing 24
2.3.1 Deep hashing neural networks: estado da arte para CBIR de RSBD . . 25

3 MATERIAIS E MÉTODOS 33
3.1 Conjuntos de imagens de observação da Terra 33
3.1.1 EuroSAT 33
3.1.2 *Aerial Image Dataset* (AID) 34
3.2 Uso e cobertura da terra no Cerrado brasileiro 35
3.2.1 Área de estudo 36
3.2.2 Dados de uso e cobertura da terra no Cerrado 38
3.3 Recursos computacionais utilizados nos experimentos 40
3.4 Métricas utilizadas 41
3.5 Modelos de *Deep Learning* aplicados ao SR 44
3.5.1 Classificação de imagens 44
3.5.2 *Content-Based Image Retrieval* (CBIR) 46
3.5.3 *Framework* para o CBIR de imagens de SR 47
3.5.3.1 Preparação dos dados, classificação e CBIR 49
3.5.4 Identificação de uso e cobertura da terra baseada em CBIR 52

4 RESULTADOS 55
4.1 CBIR de imagens satelitais EuroSAT 56
4.2 Fine-tuning usando dados multiespectrais do conjunto EuroSAT 62
4.3 CBIR de imagens de SR com equalização dos dados 63
4.4 Uso de CBIR para identificação de uso e cobertura da terra no Cerrado . 68

5 CONCLUSÕES 75
5.1 Trabalhos futuros 76
5.1.1 *Improved Metadata from Remote Sensing Images* (IMRSI) 76
5.1.1.1 Módulo de processamento de imagens 78
5.1.1.2 Módulo de recuperação de imagens 81

REFERÊNCIAS BIBLIOGRÁFICAS 85

APÊNDICE A - *DEEP LEARNING* PARA CBIR APLICADO AO CONJUNTO BIGEARTHNET 97
A.1 Conjunto BigEarthNet 97
A.2 CBIR de imagens do conjunto BigEarthNet com a *Metric-Learning-Based Deep Hashing Network* (MiLaN) 100

ANEXO A - PRODUÇÃO CIENTÍFICA NO DOUTORADO 107
A.1 Vinculada ao tema da tese 107
A.2 Colaboração com outros grupos 108

1 INTRODUÇÃO

O Sensoriamento Remoto (SR) para observação da Terra tem experimentado um grande desenvolvimento nas últimas décadas. O aumento considerável do número de sensores orbitais, assim como a evolução tecnológica empregada nesses equipamentos (BELWARD; SKØIEN, 2015), resultou no incremento expressivo do volume de dados gerados devido à melhoria nas resoluções espaciais, espectrais e temporais (APTOULA, 2014). Esse desenvolvimento apresenta um grande potencial para o avanço da área, embora a extração de informações úteis nesse caso seja ainda mais desafiadora em termos de tempo e custo computacional (DEMIR; BRUZZONE, 2016).

A necessidade de processar grandes volumes de dados de SR (SAJJAD; KUMAR, 2019), tem impulsionado a pesquisa em busca de soluções capazes de lidar com os desafios da era do *Remote Sensing Big Data* (RSBD) (MA et al., 2015; CHI et al., 2016). Uma das primeiras abordagens empregada foi a adaptação de algoritmos tradicionais à uma infraestrutura de processamento massivamente paralelo, baseada principalmente em bancos de dados distribuídos com representação matriciais de imagens. Com esse tipo de adaptação foi possível, por exemplo, reaproveitar a expertise já desenvolvida em áreas como classificação do uso da terra (CAMARA et al., 2016) e monitoramento da qualidade do ar (SEMLALI et al., 2019). Atualmente, o armazenamento e computação em nuvem têm sido apontados como soluções eficientes para o processamento de RSBD (XU et al., 2022).

Com o crescimento expressivo do número de imagens, uma questão chave passou a ser como recuperar imagens úteis a uma determinada aplicação, por exemplo, a seleção de imagens para a pesquisa de queimadas na floresta Amazônica brasileira (PLETSCH; KÖRTING, 2018).

As imagens de observação da Terra são distribuídas pelos provedores através de plataformas online denominadas comumente pelo termo "catálogos de imagens". Em geral, nessas plataformas as imagens podem ser selecionadas a partir de características como satélite, sensor, data, posição geográfica e nebulosidade (metadados comuns). O fato de os catálogos de imagens não possuírem parâmetros de busca que incluam o conteúdo presente nas imagens, implica na subutilização de todo esse volume de dados. Por exemplo, o INPE possui um catálogo[1] que está em constante

[1] O INPE foi a primeira instituição no mundo a disponibilizar imagens de satélite de forma gratuita na Internet desde 2004. Disponível em <http://www.dgi.inpe.br/catalogo/explore>. Acesso 6 setembro 2023.

crescimento pela inclusão das novas imagens geradas e adição de novos satélites ao acervo, tendo superado o volume de 120 TB em 2016. Entretanto, este volume de dados nem sempre é aproveitado de maneira integral pelo fato de o catálogo não permitir aos pesquisadores a busca por cenas que contenham alvos de interesse como: áreas de vegetação, rios, lagos ou desmatamento (KÖRTING, 2018).

Nesse sentido, uma alternativa a ser explorada é a busca e recuperação de imagens baseadas em conteúdo (*Content-Based Image Retrieval* - CBIR). O CBIR foi proposto inicialmente por Kato (1992) para recuperação de imagens em um banco de dados a partir de características como texturas, cores e formas. O interesse pelo CBIR continua ativo levando a um grande esforço por parte dos pesquisadores no desenvolvimento e adaptação para o contexto atual de dados multifontes e multiespectrais (CHI et al., 2016).

De maneira geral, os sistemas que empregam CBIR são baseados no cálculo de similaridades entre imagens através do vetor de características (atributos), que são representações numéricas extraídas das imagens com objetivo de descrever seu conteúdo visual. Por esta razão o foco das pesquisas tem sido o desenvolvimento de métodos para extração de atributos de baixa dimensão e complexidade, de maneira a facilitar à indexação e recuperação automática de imagens (DEMIR; BRUZZONE, 2015).

Nesse contexto, o *Deep Learning* (DL) tem sido empregado para o desenvolvimento de aplicações que permitam a busca e recuperação de imagens baseadas em conteúdo, especialmente no escopo de grandes conjuntos de dados (RSBD) (LIU et al., 2016; ZHU et al., 2016; LI et al., 2018; LI et al., 2021b). O DL ganhou destaque a partir de resultados obtidos por Krizhevsky et al. (2012) e LeCun et al. (2015), que utilizaram com sucesso redes neurais convolucionais para extração de características de imagens para tarefas de classificação, reconhecimento de dígitos e objetos. As redes convolucionais são um tipo de aplicação de DL que estendem as redes neurais clássicas de múltiplas camadas. Sua principal aplicação é no reconhecimento de formas e feições em imagens (LECUN et al., 1990).

1.1 Problema

A partir da observação da Terra provida por sensores multiespectrais de média resolução espacial (10 m) é possível propor uma solução para busca e recuperação de imagens de interesse científico (Figura 1.1) baseada em conteúdo?

Figura 1.1 - Imagem contendo cicatriz de queimada na região Amazônica com focos ativos detectados em 06/08/2023 às 13:08 UTC pelo sensor MODIS do satélite Terra.

Fonte: Adaptada de <https://worldview.earthdata.nasa.gov>.

Essa questão implica em dois desafios principais: i) adaptar métodos de CBIR para torná-los capazes de trabalhar com imagens de SR no escopo de grandes volumes de dados da era do RSBD; ii) produzir representações simplificadas das imagens para o uso de estruturas de indexação que levem em conta a semântica dos alvos.

1.2 Hipótese

Técnicas de *Deep Learning*, particularmente as redes convolucionais, fornecem meios para a criação de um espaço de atributos com representação semântica das imagens, ideal para utilização em aplicações de CBIR no contexto de grandes conjuntos de dados de sensoriamento remoto.

1.3 Objetivos

Dentro do contexto exposto, esta tese propõe uma solução que permite a busca e recuperação de imagens de satélites baseadas em conteúdo. Com base na revisão fundamentada da literatura, apresenta-se a adaptação de métodos de DL para criação de um espaço de atributos adequado para CBIR com representação simplificada a partir de dados multiespectrais. Adicionalmente, foi explorado o emprego do CBIR para a identificação de uso e cobertura da terra em uma área do Cerrado brasileiro.

A seguir são enumerados os principais objetivos alcançados com o desenvolvimento deste trabalho:

a) Adaptação, melhoria e combinação de técnicas para busca e recuperação de imagens de média resolução espacial baseadas em conteúdo;

b) Desenvolvimento de uma aplicação com adaptação de domínio baseada em CBIR para identificação de uso e cobertura da terra;

c) Identificação do tamanho para particionamento de imagens (*patches*) que propicie a melhor representação do conteúdo buscado em imagens providas pelo satélite Sentinel 2;

d) Idealização de uma aplicação (*framework*) para geração de metadados adicionais e descritores de características a partir de imagens disponíveis em catálogos, permitindo assim a busca e recuperação por palavras-chaves e/ou por amostras de imagens.

1.4 Contribuições

As principais contribuições desse trabalho são:

a) Revisão fundamentada da literatura para identificação de técnicas do estado da arte empregadas para busca e recuperação de imagens de sensoriamento remoto baseadas em conteúdo no contexto da observação da Terra;

b) Testes e análises de arquiteturas de DL com e sem pré-treinamento para aplicação como módulo de extração de características de imagens como base para aplicação em métodos de CBIR;

c) Testes e análises do uso de CBIR para grandes conjuntos de imagens de SR por satélite: BigEarthNet (ver APÊNDICE A) e EuroSAT;

d) Análise e comparação do CBIR com imagens de alta e média resolução espacial ([0,5 - 10] m) multiespectrais (até 13 bandas) de observação da Terra;

e) Desenvolvimento de pacote Python para criação de *patches* de imagens baseada em cubos de dados de satélite disponíveis através do projeto *Brazil Data Cube*

O período de pesquisa para elaboração deste trabalho contribuiu para uma série de realizações ligadas direta ou indiretamente ao tema principal da tese. Por exemplo, adaptação de métodos de DL para CBIR de imagens multiespectrais de satélite, uso de *Machine Learning* (ML) e processamento digital para identificação de pivôs de irrigação, entre outras.

A seguir são elencadas por ordem cronológica reversa algumas dessas realizações, que incluem artigos publicados em simpósio, conferência, revistas e desenvolvimento de software:

a) **RODRIGUES**, M. L.; KÖRTING, T. S.; QUEIROZ, G. R. (2023). Comparative Analysis of Content-Based Image Retrieval from Aerial and Satellite Multispectral Images. Manuscrito submetido à revista Transactions on Geoscience and Remote Sensing.

b) ARANTES FILHO, L. R.; **RODRIGUES**, M. L.; ROSA, R. R.; GUIMARÃES, L. N. F. (2022). Predicting COVID-19 cases in various scenarios using RNN-LSTM models aided by adaptive linear regression to identify data anomalies. Anais Da Academia Brasileira de Ciências, 94 (suppl 3). <https://doi.org/10.1590/0001-3765202220210921>.

c) **RODRIGUES**, M. L.; KÖRTING, T. S.; QUEIROZ, G. R. (2021). A Framework to Automatic Detect Center Pivots Using Land Use and Land Cover Data. Revista Brasileira de Cartografia, 73(4), 1048–1070. <https://doi.org/10.14393/rbcv73n4-60553>.

d) **RODRIGUES**, M. L.; KÖRTING, T. S.; QUEIROZ, G. R.. Automatic Detection of Center Pivots Using Circular Hough Transform, Balanced Random Forest and Land Use and Land Cover Data. XXI WORKSHOP DO CURSO DE COMPUTAÇÃO APLICADA DO INPE (WORCAP), 2021. Resumos... São José dos Campos: INPE, 2021. On-line. IBI: <8JMKD3MGPDW34P/45U7R38>. Disponível em <<http://urlib.net/ibi/8JMKD3MGPDW34P/45U7R38>>.

e) **RODRIGUES**, M. L.; KÖRTING, T. S.; QUEIROZ, G. R.; SALES, C.; SILVA, L.. Detecting Center Pivots in MATOPIBA using Hough Transform and Web Time Series Service. Proceedings of the IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS), 2020.

f) **RODRIGUES**, M. L.; KÖRTING, T. S.; QUEIROZ, G. R. (2020). Circular Hough Transform and Balanced Random Forest to Detect Center Pivots. Proceedings of the XXI Brazilian Symposium on GeoInformatics (GEOINFO), 106–117. <http://urlib.net/rep/8JMKD3MGPDW34P/43PLCP5>.

g) EIRAS, D. M. D. A.; PLETSCH, M. A. J. S.; **RODRIGUES**, M. L.; FERREIRA, K. R.; KÖRTING, T. S. (2020). Identificação de pivôs centrais usando composições de bandas e um método rápido de *Deep Learning*. Proceedings of the XXI Brazilian Symposium on GeoInformatics (GEOINFO), 180–185. <http://urlib.net/rep/8JMKD3MGPDW34P/43PR2H2>.

h) **RODRIGUES**, M. L.; KÖRTING, T. S.; QUEIROZ, G. R. *Deep Learning* e Funções Hash para Content-Based Image Retrieval (CBIR) de Imagens de Sensoriamento Remoto. XX Workshop do Curso de Computação Aplicada do INPE (WORCAP), 2020. Vídeos... São José dos Campos: INPE, 2020. On-line. (16 min). IBI: <8JMKD3MGPDW34P/43HC4NE>. Disponível em <<http://urlib.net/ibi/8JMKD3MGPDW34P/43HC4NE>>.

Como mencionado, uma das contribuições desse trabalho foi o desenvolvimento de um pacote baseado em Python para criação de *patches* de imagens disponíveis no projeto *Brazil Data Cube*, aproveitando a infraestrutura criada baseada na *SpatioTemporal Asset Catalog* (STAC)[2] para o armazenamento de itens e coleções de imagens com foco na geração de cubo de dados. O `patch-builder package`[3] foi desenvolvido em Python e utiliza *multithreading* para execução concorrente do download de imagens e criação dos *patches*. Os dados do projeto BDC estão disponíveis no formato *Cloud Optimized GeoTIFF* (COG)[4] permitindo a criação dos *patches* de imagens diretamente por acesso remoto sem a necessidade de download prévio de toda a cena.

[2] *SpatioTemporal Asset Catalog*, que é uma especificação para organização e disponibilização de dados espaço-temporais de observação da Terra. Visa facilitar a tarefa de provedores dados, como por exemplo o projeto BDC, e dos usuários através da descrição padronizada de vários tipos de dados e de *Application Programming Interface* (APIs) para recuperação dos mesmos de maneira unificada e simples. Disponível em <https://stacspec.org>

[3] Repositório do patch-builder package disponível no Github <https://github.com/marcosmlr/patch-builder>.

[4] As coleções de imagens e cubos de dados no contexto do projeto BDC são distribuídos como arquivos COG, com uma organização interna que permite acesso eficiente aos dados nos ambientes distribuídos e de alto desempenho em nuvem (FERREIRA et al., 2020).

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Visão geral sobre CBIR

Conceitualmente, podemos definir o CBIR como um problema de busca linear de imagens em um banco de dados por cálculo de similaridade entre uma imagem de consulta e o resto da coleção. Entretanto, o conjunto pode conter milhões de imagens, cada imagem descrita por um vetor de atributo com várias dimensões, tornando inviável essa busca. Esse problema pode ser superado através do uso de estruturas de dados como árvores e tabelas *hash*, que permitem a indexação dos objetos no banco de dados. O processo de *hashing* descreve o uso de funções para o mapeamento de vetores de atributos de alta dimensão para representações de baixa dimensão denominadas *hash values* ou *hash codes*. Além de simplificar a representação dos objetos, essa operação permite agrupar objetos semelhantes na mesma posição da tabela (*hash bucket*) (SLANEY; CASEY, 2008).

Na última década a área de SR tem passado por grandes mudanças, dentre as quais podemos destacar o aumento considerável do número de sistemas de observação da Terra lançados por vários países. Além disso, a evolução tecnológica empregada nos sensores atuais permitiu a melhoria das resoluções espaciais, temporais e espectrais. Esses fatores culminaram no aumento expressivo do volume de dados gerados, superando 1 EB globalmente (MA et al., 2015), inaugurando a era do *Remote Sensing Big Data* (RBSD) (LIU et al., 2018).

Nesse contexto, passou a ser de suma importância abordagens que permitam recuperar imagens úteis, por exemplo, para aplicação em atividades como o monitoramento de florestas ou dos mares a partir desse grande volume de dados. Por isso, houve um grande esforço da comunidade científica para tentar adaptar métodos já estabelecidos para a recuperação de imagens baseada em conteúdo (CBIR) (APTOULA, 2014).

Inicialmente, as abordagens desenvolvidas para CBIR na área de SR dependiam principalmente de rótulos definidos manualmente, algo que se mostra inviável no paradigma atual de RSBD. De fato, a correta representação do conteúdo visual das imagens através de descritores de características (atributos) é mais relevante para o CBIR que os rótulos manuais (LI et al., 2018).

Como exemplo de métodos utilizados para criar esses descritores, podemos citar: histograma de cores, filtro de Gabor, matriz de coocorrência, *Scale-invariant fea-*

ture transform (SIFT), entre outros. Entretanto, esses métodos se mostraram inadequados para o emprego no escopo do RSBD devido grande volume de imagens e informações em outras faixas do espectro além do visível. Deste modo, técnicas de aprendizagem inteligentes tornam-se essenciais para lidar de forma eficiente com acervos de imagens que crescem de forma acelerada, além de apresentarem melhor representação semântica das imagens, ou seja, o contexto é considerado para representar uma imagem não somente os valores dos pixels de alguns pontos da imagem (APTOULA, 2014).

Os descritores de atributos de imagens podem ser classificados basicamente em dois tipos: *hand-crafted* e *learned*. Atributos do tipo *hand-crafted*, são descritores extraídos empiricamente através de um algoritmo baseado no conhecimento de um especialista. Geralmente esses descritores possuem representação complexa, o que dificulta a recuperação de imagens. Já os atributos do tipo *learned* são gerados através de métodos de ML que fazem a extração automática de características de imagens seja por meio do aprendizado não supervisionado, ou supervisionado, segundo exemplos previamente rotulados (NAPOLETANO, 2018).

De maneira geral, os sistemas desenvolvidos para o gerenciamento de imagens de SR exploram descritores de características ou metadados para indexação dessas imagens. O foco das pesquisas atuais tem sido o desenvolvimento de métodos que buscam explorar a extração de atributos de baixa dimensão e complexidade, de maneira a permitir a indexação e recuperação automática de imagens (DEMIR; BRUZZONE, 2015).

De acordo com Napoletano (2018), um sistema típico de CBIR é formado por quatro elementos básicos (Figura 2.1):

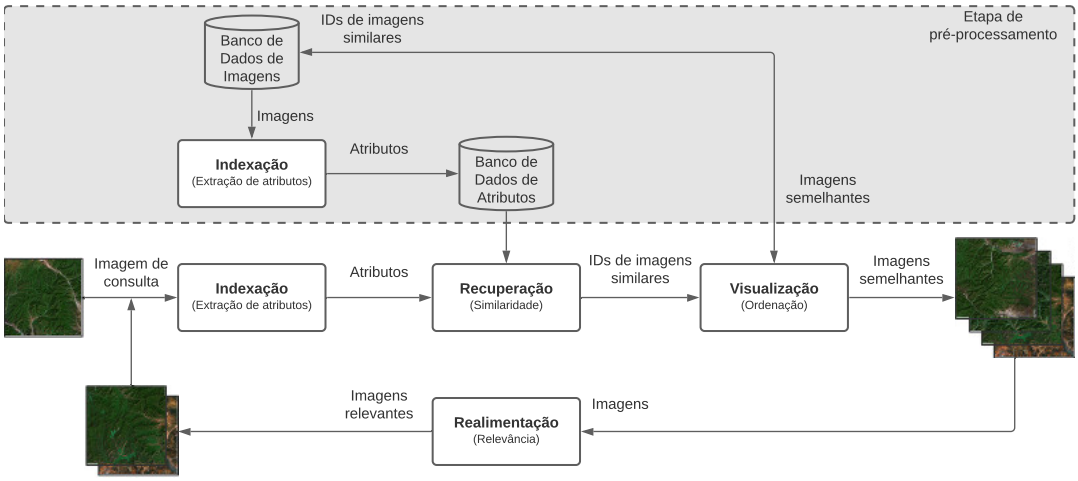
a) Indexação - módulo responsável pela extração de atributos, onde são obtidos descritores visuais para caracterização do conteúdo de imagens. Geralmente, esses descritores são pré-calculados e armazenados em uma estrutura que pode ser de arquivos, árvores ou banco de dados;

b) Recuperação - módulo que visa à recuperação de imagens que apresentem semelhanças com uma imagem alvo, baseada nos descritores visuais correspondentes;

c) Visualização - elemento responsável pela exibição do resultado da etapa anterior (recuperação), a ordenação das imagens encontradas é feita através do grau de similaridade;

d) Realimentação - módulo responsável pela seleção de imagens relevantes a partir de um subconjunto de imagens inicialmente recuperadas, essa seleção pode ser feita manualmente pelo usuário ou automaticamente por algoritmos.

Figura 2.1 - Componentes básicos de um sistema CBIR.



Fonte: Adaptada de Napoletano (2018).

O desempenho desses sistemas depende da simplicidade e eficácia dos recursos utilizados para a representação do conteúdo das imagens, sendo que as imagens de SR apresentam conteúdo muito heterogêneo, com variação de texturas (finas/grosseiras) e/ou contendo alvos específicos (objetos na imagem). Dessa forma, não é trivial a escolha de descritores que atendam com eficiência a variabilidade apresentada por imagens desse tipo (GORISSE et al., 2012).

O cenário atual apresenta uma tendência de crescimento ainda maior das bases de dados de imagens de SR, com a expansão do número de satélites para exploração comercial de serviços de observação da Terra. Desta maneira, a área de CBIR continuará sendo objeto de intensa pesquisa e desenvolvimento com foco na solução das limitações ainda apresentadas.

2.1.1 Funções *hash* para CBIR aplicadas a RSBD

O desenvolvimento de aplicações para busca e recuperação de imagens de SR em grande escala (RSBD) baseadas em CBIR tem se apresentado como uma das tarefas mais desafiadoras atualmente, atraindo a atenção de muitos pesquisadores (KAPOOR et al., 2021). Particularmente devido à rápida evolução tecnológica de sistemas satelitais e aéreos (LI et al., 2018), que implicou no crescimento acentuado do volume de imagens geradas por esses sistemas. Consequentemente, há necessidade de buscar métodos que sejam rápidos e precisos para CBIR e que permitam escalabilidade para aplicações operacionais. Uma vez que a busca exaustiva por meio de varredura linear dessas imagens seria proibitiva (SLANEY; CASEY, 2008).

Uma abordagem típica em CBIR é a consulta de imagens num conjunto P de dados baseada no algoritmo k-vizinhos mais próximos (*k-Nearest Neighbors*), que retorna k imagens mais semelhantes à imagem de referência com base no cálculo de similaridade em função da distância no espaço de atributos, quanto menor a distância maior a similaridade. Todavia, esse cálculo apresenta alto custo computacional quando P é muito grande, algo conhecido como problema de larga escala em CBIR (ANDONI; INDYK, 2017).

O sistema CBIR também pode ser modelado como um problema de classificação, nesse contexto qualquer método baseado em aprendizagem supervisionada poderia ser utilizado. A técnica *Support Vector Machine* (SVM) tem sido muito empregada, sobretudo por sua capacidade de resolver problemas não lineares complexos (BURGES, 1998). Entretanto, ela apresenta complexidade computacional linearmente proporcional ao número de imagens do conjunto e ao número de vetores de suporte obtidos durante a fase de aprendizagem do classificador, podendo superar o custo computacional do k-NN linear, inviabilizando o seu uso para aplicações que necessitem de respostas rápidas (DEMIR; BRUZZONE, 2016).

Diversos métodos de indexação têm sido propostos com o objetivo de superar as dificuldades impostas pela complexidade computacional da SVM, a maioria deles baseados em algoritmos de agrupamento ou de árvores, que visam dividir o espaço de dados para gerar subespaços em estruturas de árvores. Uma limitação dessa abordagem é que a busca com o algoritmo *kd-tree* (FRIEDMAN et al., 1977) em um espaço multidimensional tende a se tornar uma busca exaustiva, pois praticamente todos os nós do conjunto acabam sendo testados elevando a complexidade (MUJA; LOWE, 2009). Outro fator limitante do uso de métodos de indexação baseado em

árvores é que as estruturas das árvores geralmente são maiores que os dados originais, evidenciando que essa estratégia de indexação não é adequada para problemas de CBIR aplicado a RSBD (LI et al., 2018).

Alternativamente, podemos utilizar o método de indexação baseado na construção de tabelas *hash*. Esse tipo de indexação permite consultas em tempo constante e uso de memória proporcional ao número de entradas na tabela (SLANEY; CASEY, 2008). Um dos métodos mais populares para busca de vizinhança em tabelas *hash* é o método *Locality-Sensitive Hash* (LSH). Esse algoritmo utiliza uma série de funções *hash* de maneira a garantir que haja maior probabilidade de colisão para pontos próximos uns dos outros (semelhantes). Dessa forma, é possível recuperar elementos vizinhos ao ponto consultado que ocupam uma mesma posição na tabela (*hash bucket*). O LSH pertence à classe de algoritmos aleatórios, que garantem uma resposta com alta probabilidade de exatidão (SHAKHNAROVICH et al., 2005).

Os últimos anos apresentaram um crescimento no uso de *hash* em sistemas de recuperação de imagens de sensoriamento remoto em larga escala. Entretanto, a maioria desses métodos é baseada em atributos *hand-crafted*, principalmente derivados de características locais das imagens como covariância e texturas. Esses atributos possuem alta dimensão, não sendo perfeitamente compatíveis com o procedimento de *hash*. Nesse sentido, têm ganhado destaque métodos de *hash* profundo, ou seja, métodos que empregam redes neurais profundas para executar a extração de atributos e geração de códigos *hash* para representação do conteúdo de imagens (LI et al., 2016; ZHU et al., 2016).

A maioria dos métodos de *hash* profundo é supervisionada. Basicamente, durante o processo de aprendizagem os métodos mapeiam vetores de atributos de alta dimensão das imagens para um espaço de atributos de baixa dimensão, por exemplo, o *Hamming Space*, onde os atributos das imagens são representados por *hash codes* (vetores binários) que reduzem tanto complexidade computacional quanto o espaço de memória necessário para representação do conteúdo das imagens (LI; REN, 2017; LI et al., 2018). Goodfellow et al. (2016, p. 526-527) afirma que o uso de *hash codes* melhora o desempenho de aplicações baseadas em recuperação de informações a partir de um banco de dados ou conjunto de imagens. Uma vez que explora a redução de dimensionalidade e aumenta a eficiência dessas tarefas com uma representação semântica de conteúdo.

Definição 1. *Hamming Space* recebeu esse nome pois é baseado no modelo geométrico de códigos binários para identificação e correção de erros na transmissão de sinais, proposto pelo matemático Richard Hamming (HAMMING, 1950): Dado um cubo n-dimensional com vértices representados por sequências de 0 (zeros) e l (uns) de tamanho fixo, temos que o conjunto de pontos rotulados como x, y, z, ..., formam um subconjunto do conjunto de todos os vértices do cubo. A distância entre dois pontos neste espaço define uma métrica que ficou conhecida como *Hamming Distance*.

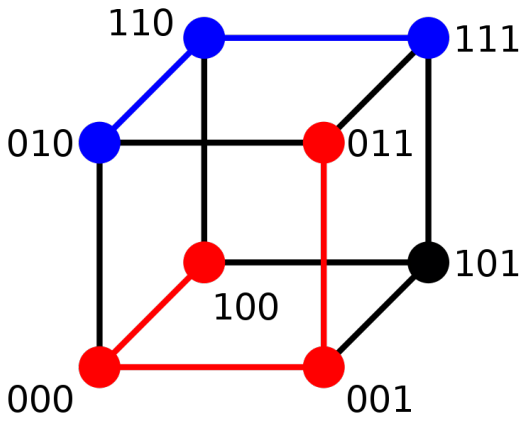
A definição dessa métrica foi baseada na observação de que a alteração de um único valor na sequência de bits, altera uma coordenada da representação geométrica, dois valores alteram duas coordenadas, e generalizando, d valores produzem uma diferença em d coordenadas. Desta maneira, podemos definir a distância $D(x, y)$ entre dois pontos x e y como o número de coordenadas para as quais x e y são diferentes. Isso possui o mesmo sentido de afirmar que a distância é equivalente ao menor número de arestas que devem ser percorridas para ir de x a y.

Essa função distância satisfaz a três condições para uma métrica:

$$D(x, y) = 0 \; \forall \; x = y$$
$$D(x, y) = D(y, x) \geq 0 \; se \; x \neq y$$
$$D(x, y) + D(z, y) + \geq D(x, z) \; (desigualdade \; triangular).$$

A Figura 2.2 ilustra o modelo geométrico proposto por Hamming, nela as cores identificam o cálculo das distâncias entre os vértices: $100_2 \rightarrow 011_2$ D=3 (caminho vermelho); $010_2 \rightarrow 111_2$ D=2 (caminho azul). Preservando a linguagem geométrica podemos definir uma esfera de raio r sobre um ponto x formada pelo conjunto de pontos que estão a uma distância r de x. Portanto, tomando um subconjunto dos vértices presente no cubo {001, 010, 100, 111}, temos que os três primeiros pontos estão na superfície de uma esfera de $r = 2$ em relação ao vértice (111). Neste exemplo qualquer que seja o ponto escolhido como centro, todos os outros estarão ao alcance de uma esfera de raio igual a 2 (*Hamming Radius*) (HAMMING, 1950).

Figura 2.2 - Representação do espaço de Hamming através de um cubo com vértices correspondentes a 3 bits.



Fonte: Burnett (2006).

O espaço métrico *Hamming Space* foi originalmente definido para a identificação e correção de erros na transmissão de sequências binárias no escopo das telecomunicações. Uma das principais características associada a esse espaço é que ele permite o cálculo eficiente de distância entre dois vetores de atributos binários (*hash codes*), utilizando uma simples operação XOR e a soma de bits (FLORES, 2015).

Basicamente os métodos de *hashing* definem funções *hash* que são aplicadas a cada imagem do conjunto de maneira a produzir os códigos binários, que irão compor a tabela *hash* onde imagens similares ocupam uma mesma posição de entrada da tabela, como demonstrado na Figura 2.4. A complexidade de armazenamento da tabela *hash* é de $O(Pb)$, onde b é o número de bits, isto é, o comprimento do vetor de *hash codes* e P o total de imagens do conjunto. As mesmas funções *hash* são utilizadas para obtenção do *hash code* da imagem de consulta.

De acordo com Demir e Bruzzone (2016), existem duas estratégias principais para recuperação de imagens baseadas em indexação por tabela *hash*, são elas:

a) *Hash lookup* - que explora a tabela *hash* para recuperar todas as imagens que estão a certa distância ou raio no *Hamming Space*. As buscas realizadas por redes neurais que utilizam esse critério são realizadas a tempo constante $O(1)$ independentemente do tamanho do conjunto de dados;

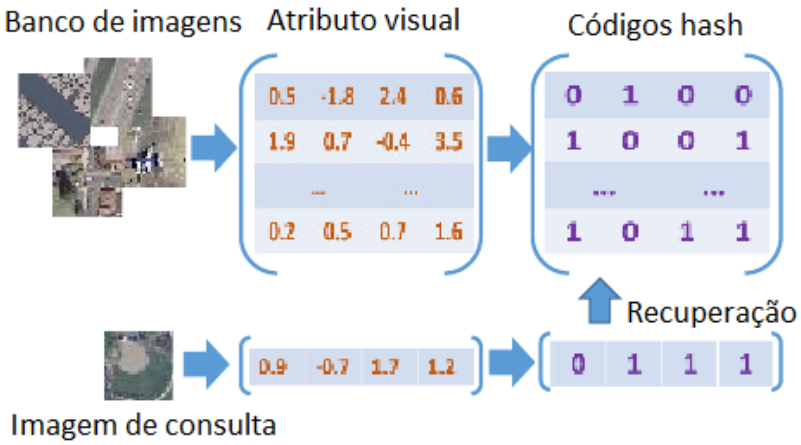
b) *Hamming ranking* - estima a *Hamming distance* entre a imagem pesquisada e todas as outras do conjunto de maneira a recuperar aquelas que possuem as menores distâncias até certa posição do ranque. As buscas utilizando essa estratégia possuem tempo linear $O(P)$, entretanto como a comparação utiliza códigos binários, a resposta é agilizada.

2.1.2 CBIR para solução do problema de busca k-NN usando funções hash

Demir e Bruzzone (2016) apresentam o CBIR como uma aplicação para a solução do problema de busca k-NN utilizando funções *hash*:

Definição 2. Sendo $X = [X_1, X_2, ..., X_P]$ um conjunto de P imagens, onde $X_i = \{x_i^1, x_i^2, ..., x_i^L\},\ i = 1, ..., P$ é a i-ésima imagem do conjunto e $x_i^l,\ l = 1, ..., L$ é o l-ésimo atributo que caracteriza o conteúdo da i-ésima imagem em X. Dada uma imagem Xq de interesse (*Query image*), temos que $\{x_q^1, x_q^2, ..., x_q^L\}$ define o conjunto de atributos desta imagem. Portanto, o objetivo é encontrar as imagens com maior similaridade a X_q, com menores tempos e requisitos de armazenamento se comparado ao método tradicional de verificação linear.

Figura 2.3 - Ilustração do uso de *hash codes* para CBIR.



Fonte: Adaptada de Li e Ren (2017).

Como demonstrado por alguns autores (RAHUL, 2014; WANG et al., 2018), o ganho de desempenho do CBIR utilizando *hashing* se deve ao fato que as funções *hash*

realizam a transformação das características extraídas das imagens (vetor de atributos) em uma representação de baixa dimensionalidade como uma sequência de bits (b) denominada *hash codes*, como ilustrado na Figura 2.3. Os métodos para geração das funções *hash* podem ser supervisionados ou não supervisionados, os métodos supervisionados possuem maior acurácia, porém, maior custo computacional (LI et al., 2016).

A solução para o problema de busca aproximada do vizinho mais próximo utilizando o LSH (não supervisionado) pode ser definida como:

Definição 3. c-vizinhos mais próximos: Dado um conjunto P de pontos em um espaço d-dimensional $\mathbb{R}^d$, construa uma estrutura de dados que, dado qualquer ponto de consulta q, relata qualquer ponto dentro da distância no máximo c vezes a distância de q a p, onde p é o ponto em P mais próximo de q (SHAKHNAROVICH et al., 2005).

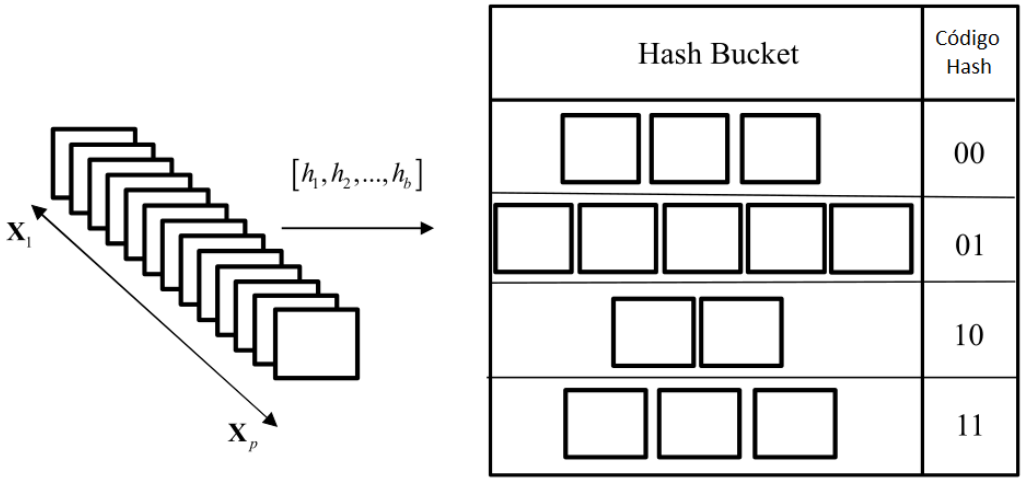
Dessa forma, temos a seguinte Equação 2.1 derivada da definição do LSH para aplicação no CBIR.

$$h_r(X_i) = \begin{cases} 1 & Se\ \nu_r^T X_i \geq 0 \\ 0 & \text{caso contrário} \end{cases}, \ r = 1, 2, ..., b \tag{2.1}$$

onde ν_r é um vetor randômico gerado a partir de uma distribuição Gaussiana multivariada com média igual a zero e uma matriz de covariância (identidade) de mesma dimensão da imagem de entrada (X_i) (SLANEY; CASEY, 2008; PAULEVÉ et al., 2010).

Ela estabelece o r-ésimo código *hash* com base na r-ésima função h_r aplicada à imagem X_i, onde $h_r, r = 1, ..., b$ define a r-ésima função *hash*. Resultando em b bits de *hash codes* para cada imagem X_i do conjunto X. A Figura 2.4 ilustra o resultado obtido após a geração da tabela *hash*, com *hash codes* de 2 bits de comprimento que agrupam imagens similares na mesma posição da tabela (*Hash bucket*). Do ponto de vista geométrico, isso define um hiperplano no espaço de atributos (DEMIR; BRUZZONE, 2016). A escolha do vetor ν_r randômico é justificada por se tratar de uma abordagem independente dos dados (SLANEY; CASEY, 2008).

Figura 2.4 - Exemplo qualitativo da *hash table* com *hash code* com 2 bits de comprimento.



Fonte: Adaptada de Demir e Bruzzone (2016).

O processo de busca e recuperação de imagens nesse caso, utiliza as mesmas funções *hash* para o computo dos *hash codes* $H_{X_q} = [h_1(X_q), h_2(X_q), ..., h_b(X_q)]$ de uma imagem X_q de interesse. Isso permite que a busca pelos vizinhos mais próximos na tabela *hash* seja realizada pelos esquemas *Hash lookup* ou *Hamming ranking* já previamente apresentados.

Apesar de o método LSH agilizar muito o processo de CBIR, sua eficiência na prática está limitada ao uso de códigos *hash* longos, o que acaba elevando os requisitos de armazenamento e o tempo de consulta. Existem alternativas ao LSH clássico, por exemplo, baseadas no uso da distância χ^2 (GORISSE et al., 2012), funções *kernel* para criação de funções *hash* para dados não linearmente separáveis como a *Kernel-based Supervised Hashing* (KSLSH) (LIU et al., 2012) e *Kernel-based Unsupervised Hashing* (KULSH) (KULIS; GRAUMAN, 2012).

Os métodos de *hashing* citados são eficazes para definição de funções *hash* de transformação dos descritores de alta dimensão para baixa dimensão, reduzindo significativamente o tempo de consulta se comparados à pesquisa de varredura linear. Entretanto, sua aplicação depende de atributos do tipo *hand-crafted*, os quais não representam com precisão o conteúdo semântico de imagens SR, levando assim à recuperações imprecisas. Nesse contexto, é cada vez maior o emprego de redes convolucionais para extração de atributos de imagens e aprendizagem conjunta para geração de códigos *hash*. Apontadas por vários autores como a melhor forma de

obter uma melhor representação semântica das imagens, assim como, para construir um espaço métrico adequado para CBIR no escopo do RSBD (LI et al., 2018; ROY et al., 2018; ROY et al., 2020).

2.2 Introdução às redes neurais convolucionais

A rede de múltiplas camadas (*Multi-Layer Perceptron* - MLP) clássica possui capacidade para solucionar problemas não linearmente separáveis, complexos a partir de um conjunto grande de dados de treinamento, que a torna fortemente indicada para aplicações de reconhecimento de imagens e de fala (RIPLEY, 1996 apud LECUN; BENGIO, 1995). Entretanto, a necessidade da extração de atributos que identifique informações relevantes a priori torna essa abordagem pouco atrativa (LECUN; BENGIO, 1995).

A rede convolucional é um caso particular do uso de rede MLP aplicada ao reconhecimento de formas ou feições, diretamente a partir de imagens brutas sem a necessidade de pré-processamento ou extração de atributos das imagens, processos que geralmente são complexos e custosos computacionalmente. A aprendizagem da rede é baseada no tradicional algoritmo de retropropagação do erro (*back-propagation* - BP), assim como em outras redes neurais, a principal diferença está na arquitetura que apresenta camadas que permitem realizar redução de escala e amostragem de imagens para a extração de características elementares como bordas, cantos, segmentos parciais entre outras a partir de informações brutas (pixels da imagem) (LECUN et al., 1990).

As aplicações de redes neurais convolucionais (*Convolutional Neural Networks* - CNNs) tiveram seu início em meados da década de 90, principalmente na área de visão computacional para o reconhecimento de padrões, mais especificamente para identificação de dígitos escritos a mão ou impressos (LECUN et al., 1990).

Recentemente as CNNs ganharam muita popularidade e têm sido empregadas nas mais diversas tarefas dentro da área de computação aplicada, por exemplo: Classificação de sentenças para processamento de linguagem natural (KIM, arXiv:1408.5882, 2014); Identificação de imagens com base no treinamento de redes profundas (*Deep Learning* - DL) utilizando funções residuais aplicadas a camadas intermediárias para facilitar o treinamento e evitar o problema do desaparecimento/explosão da resposta da função gradiente (saturação) (HE et al., 2015).

Grande parte da popularidade atribuída a esses modelos de DL se deve principalmente aos resultados obtidos pela CNN para classificação de imagens frente a outros métodos do estado da arte no desafio denominado "*ImageNet Large Scale Visual Recognition Challenge*" (KRIZHEVSKY et al., 2012) e também aos resultados notáveis obtidos por redes *Long-Short Term Memory* (LSTM) para o reconhecimento de fala (GRAVES et al., 2013).

A vantagem das redes CNNs em relação as MLPs é que elas possuem tolerância a variância de deslocamento apresentada pelos dados de entrada, por exemplo, a escrita manual pode conter variações de tamanho, inclinação e posição. Nesse caso a variância é obtida automaticamente devido à replicação de configurações de peso no espaço, isso é fundamental pois imagens e a representação da fala possuem estrutura local fortemente correlacionada espacialmente (2D). O fato das CNNs realizarem a extração de atributos locais através dos campos receptivos locais (*receptive fields*) é o que determina sua capacidade para a identificação espacial de alvos (LECUN; BENGIO, 1995).

2.2.1 Arquitetura

As redes convolucionais podem reconhecer padrões com extrema variabilidade, por exemplo caracteres escritos a mão, com tolerância a ruídos e transformações geométricas simples (LECUN et al., 1990), essa capacidade deriva de três características implementadas em sua arquitetura: i) *receptive fields*; ii) pesos compartilhados (replicação de peso) nas camadas de convolução e subamostragem (*subsampling*); iii) *pooling* espacial ou temporal (LECUN; BENGIO, 1995).

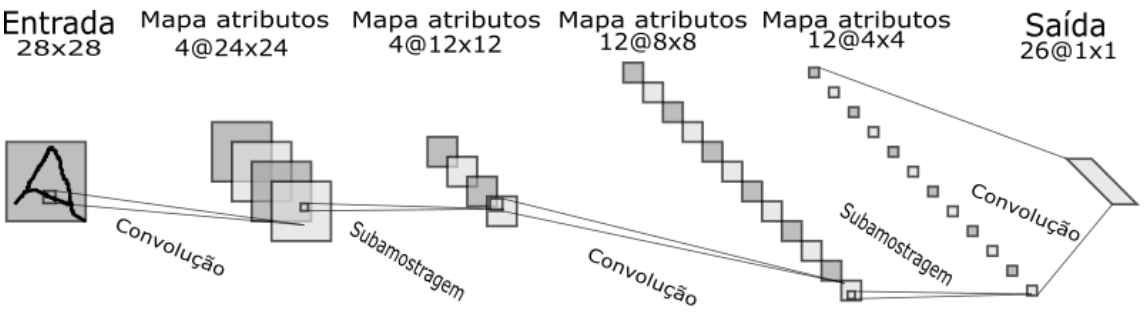
2.2.1.1 Camada de convolução

A Figura 2.5 apresenta uma rede convolucional típica baseada na abordagem proposta por LeCun et al. (1990) para o reconhecimento de caracteres. De maneira geral essa rede possui características típicas de redes neurais clássicas como mencionado anteriormente, são constituídas por neurônios que possuem pesos e *bias* ajustáveis os quais controlam o efeito da não linearidade, por exemplo, se o peso é pequeno o neurônio opera em um modo quase linear (LECUN; BENGIO, 1995).

Ao lidar com imagens é inviável conectar neurônios a todos os neurônios do volume anterior, por isso cada neurônio recebe entradas de um conjunto de unidades localizadas em uma pequena vizinhança na camada anterior delimitada pelo campo receptivo local (*receptive field*) do neurônio (equivalente ao tamanho do filtro), o conjunto

de entradas associadas a um produto escalar permite a extração de características visuais elementares (bordas e cantos) além de outras feições, essas características são combinadas em camadas posteriores sendo úteis para o compartilhamento de pesos que ajudam a identificar imagens que apresentem distorções ou variações na entrada (LECUN; BENGIO, 1995).

Figura 2.5 - Rede convolucional para identificação de caracteres (processamento de imagem).

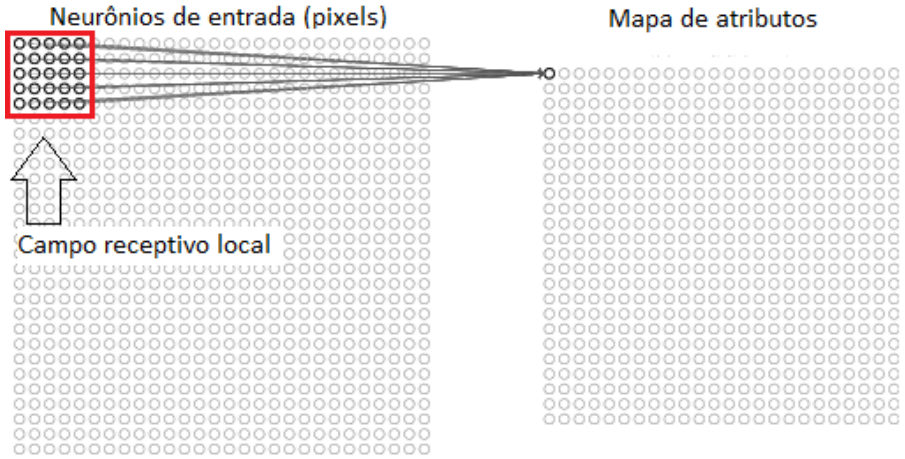


Fonte: Adaptada de LeCun e Bengio (1995).

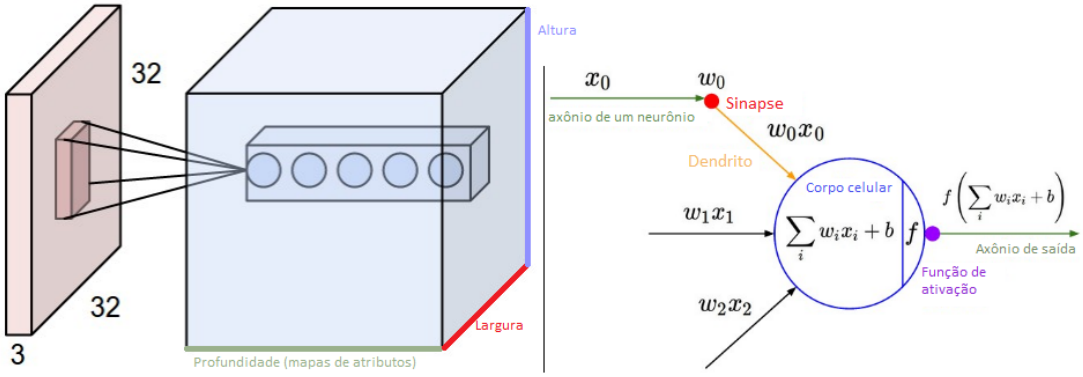
A rede completa expressa o resultado de uma função de ranqueamento tendo como base os pixels da imagem bruta em uma extremidade e os ranques da classe na outra. Além disso, é comum a utilização de uma função de ativação, por exemplo, *Softmax* na última camada totalmente conectada (KARPATHY, 2019).

A rede representada na Figura 2.5, possui 4 camadas sendo a 1ª e 3ª de extração de atributos de pesos compartilhados, enquanto a 2ª e 4ª camadas de subamostragem. O compartilhamento de pesos vem do reforço na aprendizagem do mesmo tipo de feição dependendo do filtro aplicado (*kernel*) a um conjunto de neurônios, cujos campos receptivos estão localizados em diferentes locais da imagem. As saídas desse conjunto de neurônios constituem um mapa de atributos. Esse processo de aprendizagem apresenta como principal vantagem a redução do número de parâmetros (pesos e *bias*) necessários, pois cada neurônio da camada oculta está conectado somente a uma parte das unidades de entrada (pixel), por exemplo, uma região de 5×5 correspondendo a 25 pixels de entrada. Essa região na imagem de entrada é chamada de campo receptivo local para o neurônio oculto (Figura 2.6(a)).

Figura 2.6 - Representação da conexão entre a camada oculta e o campo receptivo local (a), além da disposição tridimensional dos mapas de atributos (b).



(a)



(b)

Fonte: Adaptada de Nielsen (2015).

Nesse contexto, cada conexão $(x_0, ..., x_{24})$ ajusta um peso $(\omega_0, ..., \omega_{24})$, além disso o neurônio oculto possui um *bias* (b) geral (Figura 2.6(a)). Dessa maneira, o neurônio aprende a analisar seu campo receptivo local específico armazenando os estados desse neurônio em locais correspondentes no mapa de atributos assim como representado na Figura 2.6(b). Nela temos a representação de um volume de entrada (imagem) com dimensões 32×32×3 e um exemplo de volume de neurônios na primeira camada convolucional, onde cada neurônio (5 neste exemplo) está conectado apenas a uma região local de entrada, mas a toda a profundidade (3 canais). Importante mencionar que os neurônios aqui exemplificados são iguais aos das redes neurais clássicas, que

calculam um produto escalar de seus pesos com a entrada seguida por uma função de ativação não linear (f), por exemplo, Sigmóide (NIELSEN, 2015).

Na sequência, o campo receptivo local é movido de maneira a escanear toda a imagem de entrada armazenando os estados do neurônio em locais correspondentes no mapa de atributos. Este processo é correspondente a uma convolução com um *kernel* (filtro). O movimento do campo receptivo pode ser realizado uma unidade (pixel) por vez ou mais, sendo esse passo definido pelo hiperparâmetro *stride*. Além disso, outros dois hiperparâmetros definem o volume de saída na camada de convolução: *depth* que define o número de mapas de atributos e *padding* ou *zero-padding* que define o preenchimento com zeros na borda do volume de entrada. Isso é conveniente pois permite reservar exatamente o tamanho espacial do volume de entrada, de modo que a largura e altura de entrada e saída sejam iguais (KARPATHY, 2019).

2.2.1.2 Camada de subamostragem

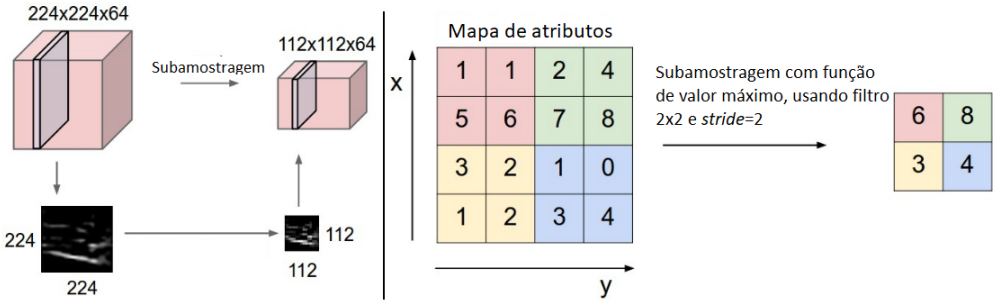
Segundo LeCun e Bengio (1995), uma vez que uma feição é aprendida, sua exata localização pode ser desconsiderada, desde que a proporcionalidade em relação a outras feições seja preservada. Dessa forma, o processo de subamostragem permite diminuir o tamanho dos mapas de atributos, reduzindo assim a sensibilidade em relação a mudanças e distorções das camadas anteriores (convolução). Camadas sucessivas de convolução e subamostragem são alternadas, resultando em uma estrutura de pirâmide bidimensional de maneira que o número de mapas de atributos é aumentado à medida que a resolução espacial diminui. A combinação convolução/subamostragem foi inspirada nas noções de células "simples" e "complexas" de Hubel e Wiesel, implementadas no modelo Neocognitron (FUKUSHIMA, 1995 apud LECUN; BENGIO, 1995).

Além da redução de dimensionalidade progressiva das camadas, a subamostragem reduz o número de parâmetros e o custo computacional da rede, auxiliando assim na diminuição do sobreajuste (*overfitting*). Isso ocorre quando a rede se especializa demais no conjunto de treinamento, mas não apresenta o mesmo desempenho para outros exemplos fora do desse conjunto, ocorrendo o problema de generalização (KARPATHY, 2019).

A Figura 2.7 ilustra a operação de subamostragem utilizando a função de valor máximo de forma independente em cada mapa de atributo (profundidade) da camada anterior, redimensionando o volume inicial de 224×224×64 para 112×112×64, utilizando um filtro 2×2 com passo de (*stride*) 2 pixels. A dimensão da profundi-

dade permanece inalterada, assim como as distâncias proporcionais entre as feições aprendidas são preservadas. Existem ainda outras funções para a operação de subamostragem, como média e normalização L2 (raiz quadrada da soma dos quadrados das ativações na região 2×2). Note que a função média foi preterida em relação a função de valor máximo por apresentar piores resultados (KARPATHY, 2019).

Figura 2.7 - Representação da camada de subamostragem com filtro 2x2, *stride* de 2 e função de *Pooling*.



Fonte: Adaptada de Karpathy (2019).

2.2.1.3 Camada totalmente conectada

De maneira geral, a arquitetura de rede convolucional pode incluir uma ou duas camadas finais totalmente conectadas. Ou seja, essa camada conecta todos os neurônios da camada de subamostragem a cada um dos 26 neurônios de saída (Figura 2.5). Nessas unidades de saída, é utilizada uma função de ativação para se obter a probabilidade de uma dada entrada pertencer a uma classe. Neste ponto, é empregado o algoritmo de retropropagação do erro, de forma que o erro obtido nesta camada seja propagado para que os pesos dos filtros das camadas convolucionais sejam ajustados. Sendo assim, os valores dos pesos compartilhados são aprendidos ao longo do treinamento (KARPATHY, 2019). Basicamente, essa camada apresenta um nível mais abstrato dos padrões aprendidos, integrando informações globais de toda a imagem. Esse é um padrão comum em redes neurais convolucionais (NIELSEN, 2015).

2.3 Big data CBIR com redes neurais convolucionais e hashing

Diversos autores apontam a vantagem do uso de redes convolucionais para extração de atributos e representação semântica de imagens através de mapas de característi-

cas, especialmente com o uso de redes pré-treinadas quando o conjunto de dados não possui tamanho suficiente (LI et al., 2018; ROY et al., 2018; ROY et al., 2020). Os sistemas CBIR desenvolvidos com esse tipo de rede têm apresentado desempenho superior àqueles que são baseados em descritores do tipo *hand-crafted* (NAPOLETANO, 2018).

Como mencionado anteriormente, os métodos de *hashing* associados a redes CNNs têm ganhado destaque na área de recuperação de imagens por conteúdo devido a sua capacidade de mapear vetores de atributos (descritores de imagens) de alta dimensão para um espaço de baixa dimensão, reduzindo assim o custo computacional para o cálculo de similaridade entre as imagens do conjunto de dados e a imagem de consulta. Geralmente, a resposta obtida é uma lista ordenada por grau de similaridade, a qual pode ser obtida de várias formas, tais como, Distância Euclidiana, Semelhança de Cosseno, Distância Manhattan, entre outras (ZHU et al., 2014).

2.3.1 Deep hashing neural networks: estado da arte para CBIR de RSBD

Um dos requisitos essenciais para o CBIR, especialmente no contexto de imagens de sensoriamento remoto em grande escala (RSBD), é a rápida pesquisa de similaridade (ROY et al., 2018). Nesse contexto, o uso de arquiteturas de DL permite o aprendizado semântico de um espaço métrico onde os atributos baseados em códigos *hash* são otimizados para tarefa de recuperação de imagens (ROY et al., 2021). Essas arquiteturas emulam a geração de funções de *hashing* que permitem mapear vetores de alta dimensão para vetores binários de baixa dimensão. O objetivo dessas funções é aprender um mapeamento intermediário das características das imagens para um espaço métrico, que é semanticamente significativo para a tarefa específica de recuperação de imagens de SR (LI et al., 2018).

O aprendizado para construção de um espaço métrico pode ser realizado através da função de perda *Triplet Loss* (SCHROFF et al., 2015), tal que a distância euclidiana entre dois pontos neste espaço correspondem fielmente a semelhança visual entre o par correspondente de imagens no espaço de pixels (ROY et al., 2021). O conceito por trás dessa função é baseado na busca do vizinho mais próximo, que busca garantir que uma imagem x (âncora) esteja mais próxima de todas as outras imagens similares a x (exemplos positivos) de um conjunto $\mathbb{U}$ do que qualquer imagem não similar (exemplos negativos), ou seja, que não possuem o mesmo rótulo (SCHROFF et al., 2015).

A seguir é apresentada à definição formal de um espaço métrico no contexto da busca do vizinho mais próximo:

Definição 4. Espaço métrico: Dados um conjunto de pontos $\mathbb{U}$ em um espaço métrico, tal que D é uma medida de distância em $\mathbb{U}$ (uma função que recebe pares de elementos de $\mathbb{U}$ e retorna um número real não negativo) e S um subconjunto de $\mathbb{U}$, temos que o problema de busca do vizinho mais próximo é construir uma estrutura de dados para S, tal que, para um ponto de consulta q, o ponto $s \in S$ minimize $D(s, q)$. $\mathbb{U}$ e D têm muitas propriedades que podem ser usadas para obter soluções para esse problema. A principal delas é que ($\mathbb{U}$, D) é um **espaço métrico** que satisfaz as seguintes condições, para todo x, y, $z \in \mathbb{U}$ (SHAKHNAROVICH et al., 2005):

1. não negatividade: $D(x, y) \geq 0$;
2. identidade: $D(x, x) = 0$;
3. simetria: $D(x, y) = D(y, x)$;
4. desigualdade triangular: $D(x, z) \leq D(x, y) + D(y, z)$.

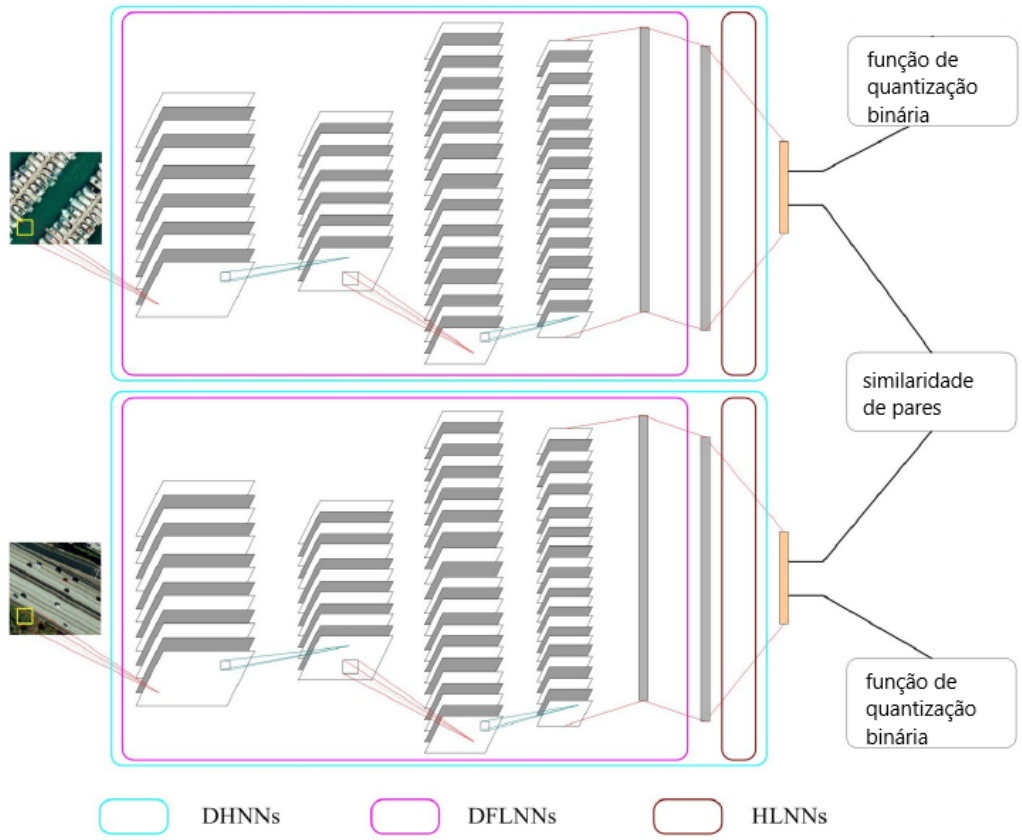
Algumas abordagens como a *Deep Pairwise-Supervised Hashing* (DPSH) (LI et al., 2016), *Deep Supervised Hashing* (DSH) (LIU et al., 2016) e *Deep Hashing Network* (DHN) (ZHU et al., 2016) visam combinar o benefício das representações semânticas geradas pelas redes CNNs com arquiteturas mais simples como as *Fully Connected Layers* para emular funções de *hashing* e permitir à busca otimizada de imagens baseada em conteúdo.

Essa combinação deu origem as redes do tipo *Deep Hashing Neural Network* (DHNN) especializadas para CBIR. Elas são compostas pelas *Deep Feature Learning Neural Networks* (DFLNNs) formadas por camadas convolucionais e *Fully connected* responsáveis pela extração de atributos das imagens, e as *Hashing Learning Neural Networks* (HLNNs) formadas por uma camada *Fully Connected* que realiza a redução de complexidade desses atributos através do aprendizado para geração de códigos *hash*. Li et al. (2018) apresentaram uma das primeiras abordagens desse tipo aplicada a dados no escopo RSBD (Figura 2.8).

O objetivo da DHNN é produzir um conjunto de vetores de atributos simplificados (*hash codes*), com baixo custo para armazenamento, que permitam a recuperação de imagens de SR em grandes conjuntos. As redes que compõem a DHNN são ajustadas

com base nas seguintes restrições: uma função de custo para quantização binária e a similaridade de pares. Basicamente, a função de custo visa mapear cada elemento representado na camada final da rede (vetores de alta dimensão) para vetores binários de baixa dimensão, e a similaridade entre pares objetiva fazer com que os atributos gerados pela DHNN concordem com as semelhanças reais identificadas pelo rótulo das imagens (LI et al., 2018).

Figura 2.8 - Representação das componentes de uma rede DHNN, incluindo as unidades de extração de atributos (DFLNN) e aprendizagem *hashing* (HLNN).



Fonte: Adaptada de Li et al. (2018).

Embora a DHNN tenha representado um grande avanço para o CBIR no cenário do RSBD, o uso da função *cross-entropy*[1] não atende aos requisitos para construção de

[1] A função de custo entropia cruzada serve para descrever a diferença entre duas distribuições de probabilidades, permitindo medir o erro entre os valores previstos e os valores esperados (rótulo dos dados). Ela é principalmente utilizada como alternativa à função de custo de erro médio quadrático pois tende a permitir a correção dos pesos mesmo quando houver saturação de alguns dos neurônios da rede (DSA,).

um espaço métrico que agrupe imagens semelhantes, algo fundamental para CBIR. Isso ocorre pela ausência de um limiar entre amostras positivas e negativas como resposta da função, o que leva a uma generalização ruim. Como consequência, são necessários vetores de *hash* longos e uma grande quantidade de imagens de treinamento rotuladas, geralmente difíceis de obter no caso de imagens de SR (ROY et al., 2020).

De maneira a resolver as dificuldades supracitadas, algumas soluções foram propostas por Roy et al. (2018) e Roy et al. (2020). A primeira abordagem parte do conceito de que o treinamento realizado para o ajuste dos pesos de redes convolucionais profundas permite transferir os ganhos de qualidade da solução de problemas da área de Visão Computacional (classificação) para outros domínios (CBIR).

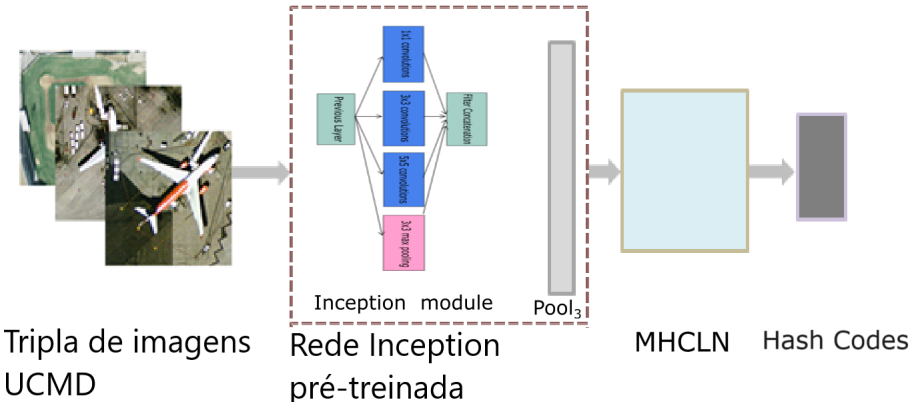
Partindo desse princípio, a *Metric and Hash-Code Learning Network* (MHCLN) é uma arquitetura baseada em camadas do tipo totalmente conectadas treinada para gerar representações simplificadas das imagens (*hash codes*) utilizadas para construção de um espaço métrico adequado para CBIR. Ela utiliza como módulo de extração de características das imagens (*backbone*) a rede convolucional profunda *Inception Net* (SZEGEDY et al., 2016) pré-treinada com o conjunto de imagens ImageNet[2], a combinação dessas arquiteturas (Figura 2.9) permite a melhoria da representação semântica das imagens e a recuperação baseada em similaridade em tempo quase real (ROY et al., 2018).

O treinamento da rede MHCLN, conforme definição a seguir, visa mapear cada vetor de atributos de alta dimensão correspondente as 2048 ativações de saída da *Inception* em um espaço métrico significativo $f : \mathbb{R}^{2048} \rightarrow \mathbb{R}^K$, a fim de permitir quantização do vetor de atributos binários através de funções *hash*.

Definição 5. Seja $I = \{X_1, ..., X_P\}$ um conjunto de imagens de SR para treinamento, onde X_i está associada a um rótulo de classe $y_i \in Y = \{y_1, y_2, ...\}$. O objetivo da MHCLN é aprender funções de *hashing* $h : I \rightarrow \{0, 1\}^K$ que mapeiam os atributos das imagens para *hash codes* binários com comprimento K, de forma que esses códigos incorporem a semântica das imagens correspondentes. Isso permite recuperar de forma eficiente uma imagem de consulta X_q comparando bit a bit os seus códigos binários com os vetores de referência.

[2]ImageNet é um conjunto de imagens muito utilizado na área de Visão Computacional para realizar a avaliação de modelos de DL. Possui muitas classes de imagens, sendo que cada classe pode conter de centenas a milhares de imagens (DENG et al., 2009; RUSSAKOVSKY et al., 2015).

Figura 2.9 - Aprendizado semântico baseado nas redes Inception V3 e MHCLN para construção de um espaço métrico adequado para CBIR utilizando *hash codes*.



Fonte: Adaptada de Roy et al. (2018).

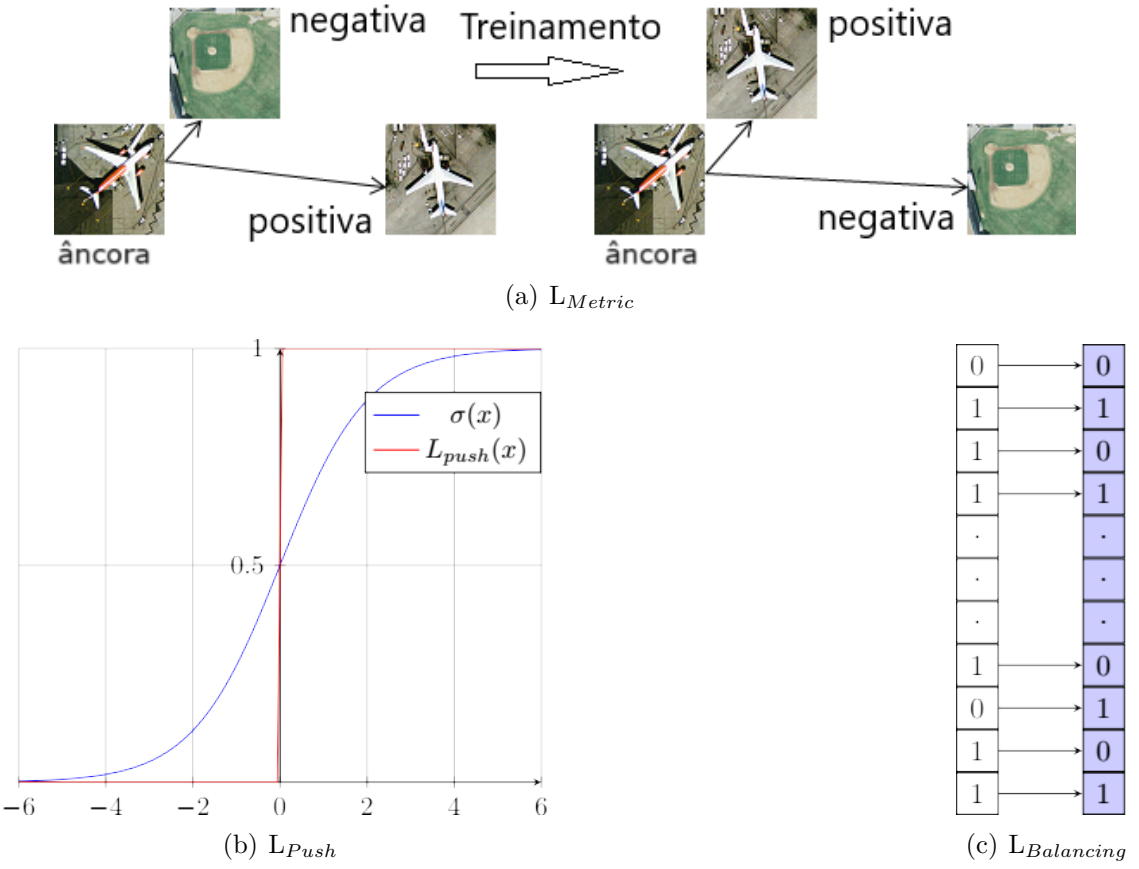
O aprendizado do espaço métrico proposto é baseado no uso combinado de três funções de custo a L_{Metric}, L_{Push} e $L_{Balancing}$ (Equação 2.2).

$$L = L_{Metric} + \lambda_1 L_{Push} + \lambda_2 L_{Balancing}, \tag{2.2}$$

onde $\lambda_1 = 0,001$ e $\lambda_2 = 1$ são hiperparâmetros que ponderam a importância relativa das funções L_{Push} e $L_{Balancing}$, determinados com base no método de validação cruzada (ROY et al., 2018).

A função de custo final L permite o aprendizado semântico de um espaço métrico adequado para a realização da busca e recuperação de imagens baseadas em conteúdo (CBIR), isso se deve a combinação do efeito das três funções de custo (Figura 2.10). A intuição por traz da função L_{Metric} é baseada no treinamento usando tripla de imagens para minimização da *Triplet-loss*, que permite que imagens similares (âncora/positiva) sejam agrupadas próximas no espaço de parâmetros em detrimento a imagens não similares (negativas). Além disso, duas outras funções são utilizadas, a L_{Push} penaliza a representação da sigmoide de maneira a empurrar as ativações da última camada da rede para serem binárias, já a L_{Push} é a função que busca balancear o número de 0 (zeros) e 1 (uns) produzidos para garantir que todos os *hash codes* gerados sejam efetivamente úteis para a tarefa de CBIR (ROY et al., 2018).

Figura 2.10 - Funções de custo otimizadas para aprendizado semântico utilizadas pela MHCLN.



(a) L_{Metric}

(b) L_{Push}

(c) $L_{Balancing}$

A L_{Metric} é baseada no conceito da minimização da função *Triplet-loss*, que permite o agrupamento de imagens similares (âncora/positiva) em detrimento de imagens não similares (negativa). A L_{Push} penaliza as saídas da última camada da rede baseada na função sigmoide para forçar valores binários. A $L_{Balancing}$ é a função de balanceamento de bits, para garantir que os *hash codes* produzidos sejam úteis para tarefa CBIR.

Fonte: Figura (a) adaptada de Roy et al. (2018), demais próprio autor.

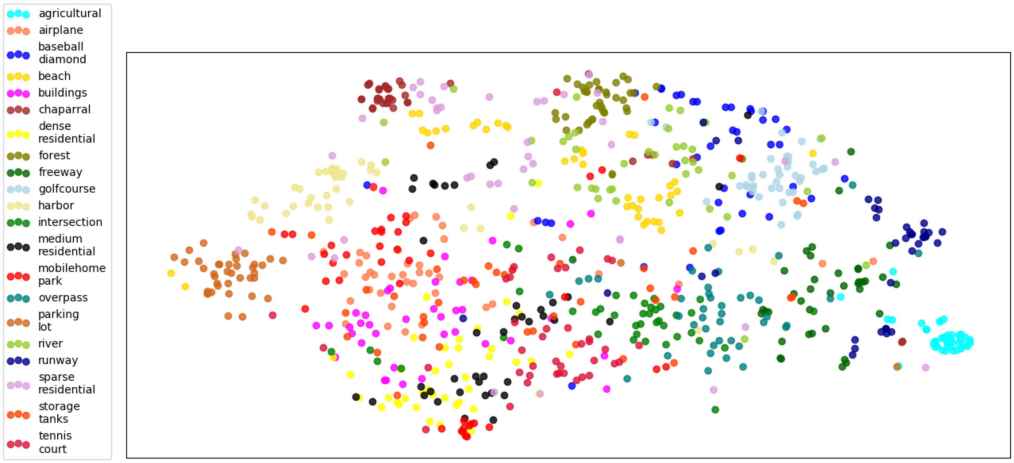
Em seu trabalho complementar, Roy et al. (2020) utilizam a mesma ideia principal da combinação tripla de funções de custo para o aprendizado e representação semântica do espaço métrico com códigos *hash*. A principal diferença é que nesse estudo eles apresentam uma avaliação dos efeitos obtidos com a variação de alguns parâmetros (*ablation study*) da rede *Metric-Learning-Based Deep Hashing Network* (MiLaN), como o comprimento dos vetores de *hash codes* e o percentual do conjunto de imagens de treinamento.

A análise realizada demonstrou ganho de desempenho à medida que o comprimento do vetor de *hash* é aumentado. Além disso, evidenciou a superioridade da rede MiLaN sobre métodos do estado da arte como o *Kernel-based Supervised* LSH (DEMIR; BRUZZONE, 2016) em mapear de forma eficiente a informação semântica extraída das imagens pela rede Inception em códigos *hash* discriminativos permitindo a construção de um espaço métrico adequado para a tarefa CBIR (ROY et al., 2020).

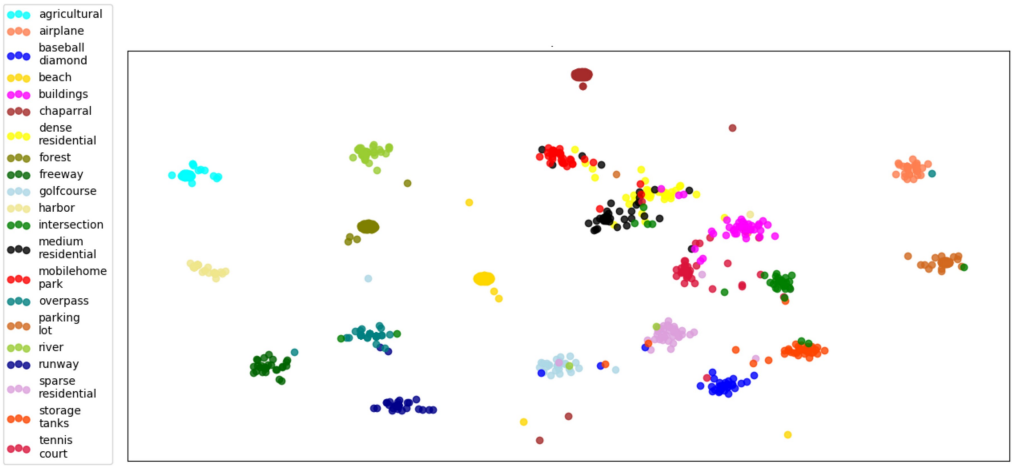
A Figura 2.11 ilustra de forma qualitativa a superioridade da MiLaN sobre o método *Kernel-based Supervised* LSH ao produzir um espaço de parâmetros (espaço métrico) com separabilidade entre as imagens de classes distintas. Em termos quantitativos, por exemplo, utilizando apenas 30% do total de imagens para treinamento a rede MiLaN alcançou uma precisão de $\approx 73\%$ na tarefa de CBIR, excedendo a precisão de $\approx 58\%$ obtido pelo outro método utilizando 80% do conjunto para treinamento, empregando vetores de *hash codes* de 32 bits ($K = 32$) em ambos os casos.

As performances obtidas tanto em precisão quanto em tempo necessário para realizar a recuperação de imagens, evidenciam o sucesso das inovações adotadas por essa abordagem: i) Aprendizado semântico baseado na representação intermediária produzida pela rede Inception, importante para evitar o risco de *overfitting* quando o número de imagens rotuladas é pequeno; ii) Uso da função de custo tripla que efetivamente aprende a produzir códigos *hash* binários discriminativos adequados para construção de um espaço métrico otimizado para a tarefa CBIR (ROY et al., 2020).

Figura 2.11 - Projeção bidimensional do espaço de parâmetros (espaço métrico) de k-dimensões, sendo $k = 20$, referente ao comprimento do vetor de *hash codes* gerados a partir de imagens do conjunto UCMD[3].



(a) *Kernel-based Supervised* LSH



(b) MiLaN

Projeção utilizando o método *t-distributed Stochastic Neighbor Embedding*[4].

Fonte: Roy et al. (2020).

[3]A proposta da rede MiLaN utilizou o conjunto de imagens *The University of California Merced Land Use Dataset* (UCMD) - conjunto de ortoimagens aéreas extraídas da coleção de mapeamento de áreas urbanas dos Estados Unidos (YANG; NEWSAM, 2010).

[4]O método *t-distributed Stochastic Neighbor Embedding* (t-SNE) serve para visualização de dados de alta dimensão. Ele converte semelhanças entre pontos de dados em probabilidades conjuntas e minimiza a divergência entre as probabilidades dos dados de baixa e alta dimensão. A função de custo utilizada pelo método não é convexa, ou seja, diferentes inicializações refletem resultados diferentes (estocástico). Isso significa que a projeção final indica com êxito a distribuição espacial dos dados, mas pode ser diferente a cada execução (MAATEN; HINTON, 2008).

3 MATERIAIS E MÉTODOS

3.1 Conjuntos de imagens de observação da Terra

A busca e recuperação de imagens baseadas em conteúdo no contexto do sensoriamento remoto para observação da Terra representa um grande desafio para comunidade científica, sendo um tema com contínuo interesse especialmente para o desenvolvimento de métodos baseados em técnicas de DL (ZHOU et al., 2023).

Um dos requisitos para o emprego dessas técnicas nesse contexto é a disponibilidade de conjuntos de dados de SR com número e variabilidade de imagens suficientes que permitam aos modelos de DL alcançarem a generalização. Por exemplo, UCMD, WHU-RS19, RSSCN7, SIRI-WHU, AID, NWPU-RESISC45, RSI-CB, EuroSat, PatternNet e BigEarthNet são *datasets* de SR empregados com sucesso para classificação de imagens (HELBER et al., 2019; SUMBUL et al., 2019; YASSINE et al., 2021) e CBIR (DEMIR; BRUZZONE, 2016; NAPOLETANO, 2018; LI et al., 2018; LI et al., 2021a; KAPOOR et al., 2021) utilizando métodos de DL. Entre eles merece destaque o conjunto AID pela resolução espacial ([0,5 - 8] m) e os conjuntos EuroSAT e BigEarthNet pela quantidade de amostras de imagens disponíveis ($\geq$ 27 mil) com informação multiespectral.

O conjunto BigEarthNet é apontado como o melhor e mais completo conjunto de imagens para o treinamento de modelos para aprendizagem profunda, especialmente para aplicações que envolvem identificação de alvos com múltiplos rótulos (*multilabel classification*) (SUMBUL et al., 2021). Entretanto, pelas limitações e problemas relativos a erros na identificação de uso e cobertura da terra em algumas amostras de imagens detalhadas neste documento (APÊNDICE A), neste trabalho foram adotados os conjuntos AID e EuroSAT para propor e avaliar inovações que permitam a melhoria da tarefa CBIR aplicada à imagens de SR. Além disso, foi desenvolvida uma aplicação baseada em CBIR para identificação do uso e cobertura da terra em uma área do Cerrado usando imagens do cubo de dados Sentinel disponíveis no projeto *Brazil Data Cube* (BDC).

3.1.1 EuroSAT

Esse foi um dos primeiros conjuntos de imagens de SR por satélite indicado para modelos de DL, demonstrando à importância e o impacto do uso do dado multiespectral (MS) na classificação de imagens. O emprego desse dado visa fornecer informações complementares quanto as diversas respostas espectrais dos alvos na superfície pos-

sibilitando a melhoria no processo de classificação de imagens de Observação da Terra de média resolução espacial (10 m) (HELBER et al., 2019).

Ele possui 27.000 *patches* de imagens com 64×64 pixels e 13 bandas espectrais imageadas com o sensor *MultiSpectral Instrument* (MSI) embarcado nos satélites Sentinel-2 (2A/2B).

Os *patches* foram rotulados com 10 tipos de uso e cobertura da terra (Figura 3.1), contendo diferentes usos de terras agrícolas tais como Culturas Anuais (*Anual Crop*), Culturas Permanentes (*Permanent Crop*, por exemplo, pomares de frutas, Vinhas e Olivais) e Pastagens (*Pasture*). O conjunto também possui áreas construídas discriminadas: Rodovias (*Highway*), Edifícios Residenciais (*Residential*) e Edifícios Industriais (*Industrial*). As imagens de corpos d'água são exemplos com associação de Mar e Lagos (*Sea & Lake*) além de Rios (*River*). Completam os dados exemplos de vegetação dos tipos Floresta (*Forest*) e Vegetação Herbácea (*Herbaceous Vegetation*). O número de amostras de cada tipo varia entre 2.000 e 3.000 *patches*.

Figura 3.1 - Amostra de imagens dos tipos de uso e cobertura da terra providos pelo conjunto EuroSAT.

Fonte: Próprio autor.

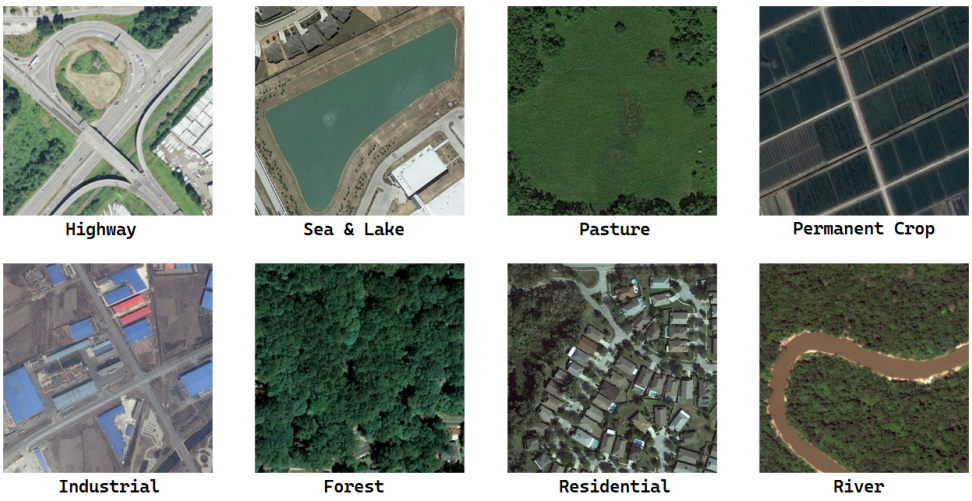
3.1.2 *Aerial Image Dataset* (AID)

Esse conjunto de imagens foi criado para prover exemplos de variados tipos de uso e cobertura da terra suficientes para o treinamento de modelos de DL evitando *overfitting* (XIA et al., 2017). Sua construção se deu através da extração de 10.000

imagens aéreas adquiridas com múltiplos sensores disponíveis no Google Earth com dimensão de 600×600 pixels e resolução espacial variando de 0,5 a 8 m.

Originalmente o conjunto possui 30 tipos de uso e cobertura da terra rotulados por especialistas da área de SR. Todavia, para permitir a intercomparação com o conjunto EuroSAT, somente as seguintes classes que possuem equivalência foram utilizadas (Figura 3.2): Viadutos (*Viaduct*≡*Highway*), Lagos (*Pond*≡*Sea & Lake*), Prados (*Meadow*≡*Pasture*), Terras Agrícolas (*Farmland*≡*Permanent Crop*), Edifícios Industriais (*Industrial*), Floresta (*Forest*), Edifícios Residenciais (*Dense+Medium Residential*≡*Residential*) e Rios (*River*). A quantidade de amostras de cada classe varia de 220 a 420 imagens.

Figura 3.2 - Amostra de imagens dos tipos de uso e cobertura da terra providos pelo conjunto AID.



Equivalência entre as classes do AID e EuroSAT: Viadutos (*Highway*), Lagos (*Sea & Lake*), Prados (*Pasture*), Terras Agrícolas (*Permanent Crop*), Edifícios Industriais (*Industrial*), Floresta (*Forest*), Edifícios Residenciais (*Residential*) e Rios (*River*).

Fonte: Próprio autor.

3.2 Uso e cobertura da terra no Cerrado brasileiro

O Cerrado é o segundo maior bioma brasileiro, cobrindo cerca de 24% do território nacional e considerado a maior savana da América do Sul. Sua vegetação típica é caracterizada por plantas lenhosas com caules grossos, de tom escuro e retor-

cidos (RIBEIRO; WALTER, 2008). Algumas das principais bacias hidrográficas do país estão localizadas nesse bioma, por exemplo, a bacia dos rios Araguaia-Tocantins e do rio São Francisco, além de afluentes dos rios Amazonas e Prata.

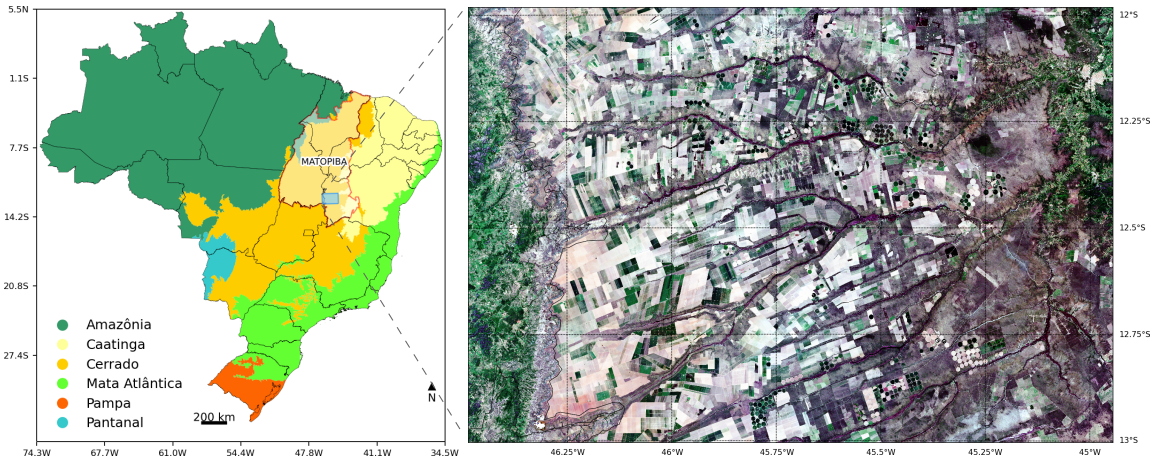
Nas últimas décadas, o Cerrado tem passado por forte transformação devido a expansão agrícola em seu território, já tendo perdido mais de 50% de sua vegetação nativa (MARRIS, 2005). Em um movimento que teve início com o desenvolvimento de novas tecnologias e a criação de programas governamentais de incentivo à ocupação, implementados a partir da década de 1970 (BRASIL, 1974) e se intensificou a partir de 2006 com a Moratória da Soja na Amazônia (RUDORFF et al., 2011).

Essa transformação alterou aspectos socioeconômicos regionais e impulsionou a produtividade desta nova e importante fronteira agrícola brasileira (SANO et al., 2019), com destaque para região composta pelos estados do Maranhão, Tocantins, Piauí e Bahia, conhecida como MATOPIBA, responsável por grande parte da produção brasileira de grãos, com uma área de cerca de 73 milhões de hectares (EMBRAPA, 2019), tornando o Brasil um dos principais produtores mundiais de *commodities* agrícolas.

3.2.1 Área de estudo

A área de estudo escolhida (Figura 3.3) está localizada na região do MATOPIBA e inclui alguns municípios do Oeste da Bahia com destacada atividade agrícola (Barreiras, Luís Eduardo Magalhães e São Desidério), essa região foi objeto de estudos para identificação de círculos de pivôs de irrigação em imagens de satélite, que antecederam este trabalho (RODRIGUES et al., 2020a; RODRIGUES et al., 2021). Com base no mapeamento de uso e cobertura da terra feito pelo projeto TerraClass Cerrado (INPE, 2018), a região possui uma ocupação de solo heterogênea com predominância de vegetação natural (nativa), culturas agrícolas de um ciclo e pastagens herbáceas, além de algumas áreas com cultivo de mais de um ciclo, vegetação natural secundária (regeneração após corte raso), agricultura perene, áreas urbanizadas, silvicultura, desmatamento, corpos d'água e pouquíssimas áreas de mineração.

Figura 3.3 - Área de estudo para caracterização de uso e cobertura da terra localizada na região do MATOPIBA dentro do bioma Cerrado (esquerda). Destaque para composição RGB do tile 089097[1] do cubo de dados Sentinel 2-16D (composição temporal de 16 dias) com início em 19 dezembro 2018 (direita).



Fonte: Próprio autor.

O projeto BDC produz dados prontos para análise para todo o território brasileiro (FERREIRA et al., 2020), o cubo de dados do Sentinel 2-16D corresponde a imagens de reflectância à superfície compostas temporalmente no intervalo de 16 dias para escolha do melhor pixel (remoção de nuvens, sombra de nuvens e dados inválidos) (MARUJO et al., 2022), projetadas utilizando a grade definida no escopo do projeto[1] resultando em rasters com dimensões 16806 (colunas) e 10986 (linhas) para o tile 089097 com 12 bandas espectrais (exceto banda 10, absorção de vapor d'água) e resolução espacial de 10 metros.

O experimento foi realizado com conjuntos de *patches* criados a partir desse raster utilizando o pacote em Python desenvolvido no escopo deste trabalho[2]. O primeiro conjunto é formado por 11352 *patches* com 128×128 pixels, já o segundo contém

[1]Os cubos de dados no escopo do projeto BDC utilizam um esquema com 3 grades hierárquicas. O tamanho dos blocos da grade (*tiles*) varia de acordo com a versão da grade, no caso do dado Sentinel 2 de 2018 a versão utilizada era a v1 com as seguintes características, $1{,}5^o \times 1^o$ (*Small Grid*, SM) baseada na projeção de Albers (BDC, 2002).

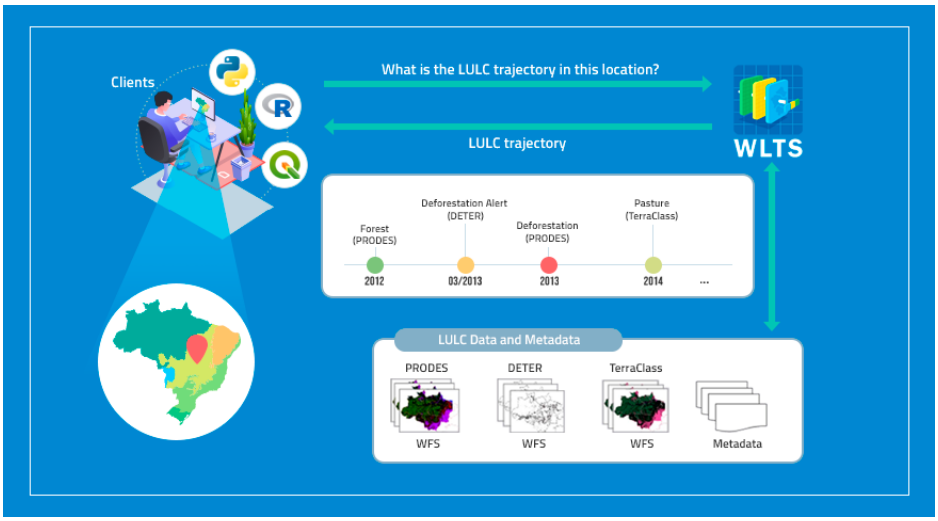
[2]O pacote path-builder foi desenvolvido em Python e explora o paralelismo por meio de *threads* para criação dos *patches* de imagens de coleções disponíveis através de catálogos no padrão SpatialTemporal Asset Catalog (STAC), como mencionado anteriormente o STAC é uma especificação para organização e disponibilização de dados geoespaciais com interface de consulta facilitada através de APIs. O pacote fornece opções para criação dos *patches* localmente (download das cenas previamente) ou remotamente (*cloud processing*), além disso ele implementa uma nova opção de pesquisa e recuperação das imagens, baseada no número do *tile*. O código fonte do pacote está disponível no Github <https://github.com/marcosmlr/patch-builder>.

45236 *patches* com 64×64 pixels, importante mencionar que os conjuntos de imagens preservam a projeção original. A variação dos tamanhos dos *patches* seguiu padrões adotados tradicionalmente por outros conjuntos de dados de ML (MNIST, ImageNet, EuroSAT, etc.) com objetivo de verificar qual tamanho seria mais adequado semanticamente para aplicação com CBIR, uma vez que a informação de vizinhança é importante para identificação de alguns tipos de uso e cobertura como estradas e rios.

3.2.2 Dados de uso e cobertura da terra no Cerrado

O *Web Land Trajectory Service* (WLTS) é um serviço disponível no projeto BDC para consulta de amostras (pontos com geolocalizações) de uso e cobertura da terra para todo o território nacional. Permitindo a pesquisadores análises baseadas em uma abstração de alto nível denominada "Trajetórias de Uso e Cobertura da Terra" que integra e apresenta ao longo do tempo dados de uso e cobertura produzidos e disponibilizados de forma livre por vários projetos como Deter, Prodes, TerraClass entre outros (ZIOTI et al., 2022). As trajetórias podem ser consultadas através de uma API padrão que é baseada na notação OpenAPI 3.0[3], o serviço provê ainda clientes desenvolvidos em R e Python além de um *plugin* para consulta de trajetórias usando o Quantum GIS (Figura 3.4).

Figura 3.4 - Fluxograma do WLTS para coleta de amostras de uso e cobertura da terra.



Fonte: Zioti et al. (2022).

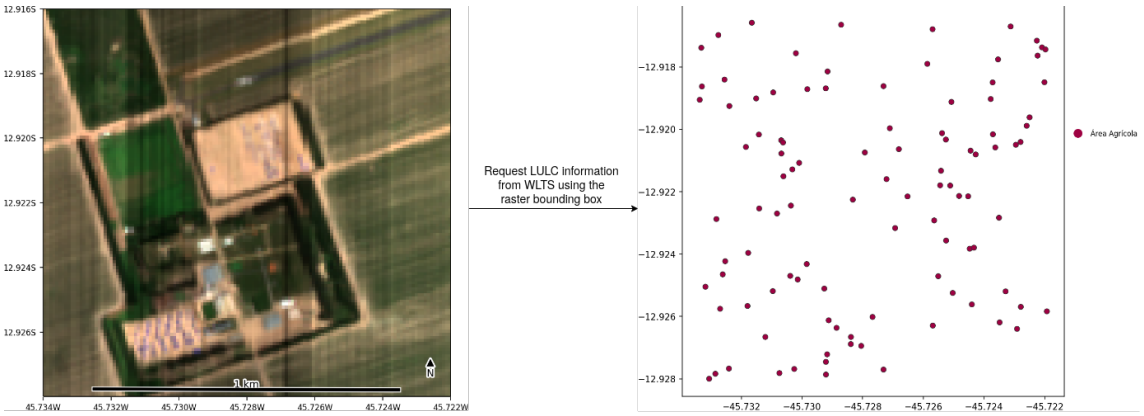
[3]A especificação da API do WLTS pode ser encontrada através do repositório disponível no Github <https://github.com/brazil-data-cube/wlts>.

Uma das principais aplicações do WLTS é o seu uso para validação de amostras de uso e cobertura do solo, empregadas para o treinamento de modelos de ML para produção de novos mapas de classificação. No contexto desse trabalho, o WLTS foi utilizado para validação da identificação de uso e cobertura da terra em *patches* de imagens na região do Cerrado (área de estudo) realizada através do uso de CBIR (Seção 3.5.4).

Dessa maneira, foi utilizado o cliente em Python do WLTS[4] para consulta e download de amostras de uso e cobertura da terra referentes a cada um dos 11352 *patches* de imagem criados para área de estudo (Figura 3.5).

Explorando paralelismo com *threads* em Python, foi criado um algoritmo para iterar sobre as imagens e determinar as coordenadas de *bounding box* necessárias para consulta ao WLTS. Para cada imagem de 128×128 pixels foram sorteados aleatoriamente 101 pontos de forma a determinar a predominância do tipo de uso e cobertura da terra naquela área.

Figura 3.5 - Esquema baseado no serviço WLTS para coleta de amostras de uso e cobertura da terra para os *patches* da área de estudo.



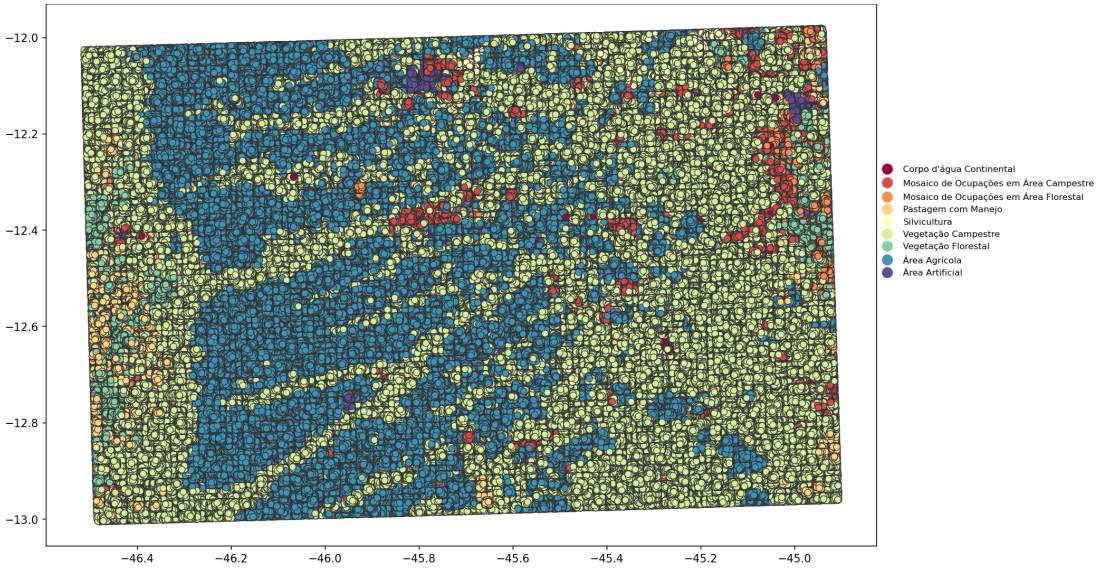
A partir do *bounding box* do raster (128×128 pixels) são sorteados aleatoriamente 101 pontos para consulta ao serviço WLTS.

Fonte: Próprio autor.

[4]Cliente desenvolvido em Python para acesso aos dados de trajetórias de uso e cobertura da terra do serviço WLTS. Disponível em <https://github.com/brazil-data-cube/wlts.py>.

A partir do WLTS foi possível recuperar e armazenar[5] 1146552 amostras (Figura 3.6) de uso e cobertura da terra provido pelo Instituto Brasileiro de Geografia e Estatística (IBGE) para o ano de 2018 (IBGE, 2020).

Figura 3.6 - Distribuição das amostras de uso e cobertura da terra adquiridas através do serviço WLTS para a área de estudo.



Fonte: Próprio autor.

3.3 Recursos computacionais utilizados nos experimentos

Diversas são as ferramentas utilizadas para o desenvolvimento de modelos de aprendizagem profunda, dentre elas podemos destacar os *frameworks* Theano, PyTorch e Tensorflow, criados para facilitar a construção de redes neurais através de linguagens de programação de alto nível.

O Tensorflow baseado em Python foi escolhido para implementar os modelos deste trabalho por possuir métodos e funções que facilitam à prototipagem de várias arquiteturas de DL do estado da arte utilizando módulos extensíveis. Seu processamento baseado no fluxo de dados através de grafos permite escalabilidade e adaptabilidade, sendo possível executá-lo de forma distribuída em vários computadores e tipos de unidades de processamento como: CPUs, GPUs e TPUs (ABADI et al., 2015).

[5]As amostras de uso e cobertura da terra adquiridas via WLTS para área de estudo foram armazenadas em um arquivo shapefile contendo o provedor do mapeamento, identificação do *patch* de imagem e o tipo uso e cobertura mapeado.

O ambiente computacional utilizado para aquisição de dados, treinamento e avaliação dos modelos de DL utilizados nos experimentos aqui descritos é composto por um sistema baseado na arquitetura 64 bits (Ubuntu 18.04 LTS) instalado em um servidor Supermicro 4028GR-TVRT com 2 Intel Xeon E5-2630L v4s 2.9GHz Deca core e cache de 25MB, 1TB RAM DDR4, 8 GPUs NVIDIA Tesla V100 SXM2 16GB.

Os resultados alcançados atestam que a infraestrutura computacional se mostrou adequada e robusta para o desenvolvimento e análise de dados através de modelos de aprendizagem profunda com grande número de camadas (> 150).

3.4 Métricas utilizadas

Medidas de desempenho para classificação no escopo do aprendizado de máquina são realizadas tomando como base número de exemplos reais e previstos (algoritmo de ML) de cada classe anotados na forma de uma matriz de confusão (Tabela 3.1).

Tabela 3.1 - Matriz de confusão de duas classes (**P**ositivo/**N**egativo).

	Real	
Predição	**P**	**N**
p	VP	FP
n	FN	VN

Fonte: Adaptada de Matos et al. (2009).

Nesse contexto, *Precision* mede a taxa de acerto dos casos verdadeiros positivos (Equação 3.1), enquanto *Recall* mede quanto casos verdadeiros positivos foram recuperados (Equação 3.2), indica a relevância da classificação dos verdadeiros positivos. A média harmônica ponderada de *Precision* e *Recall* também conhecida como *F-Measure*, permite avaliar eficácia de ambas métricas a depender do valor atribuído ao peso β (Equação 3.3). Valores típicos de β são: $\beta = 2$ (*Recall* tem o dobro do peso em relação a *Precision*) e $\beta = 0,5$ (*Precision* tem o dobro do peso em relação a *Recall*). Quando ambas têm o mesmo peso ($\beta = 1$) temos a medida conhecida como *F-Measure* tradicional (Equação 3.3) ou *F-Score* balanceada (*F1-Score*) (MATOS et al., 2009).

$$Precision = \frac{VP}{VP + FP} \tag{3.1}$$

$$Recall = \frac{VP}{VP + FN} \tag{3.2}$$

$$F\text{-}Measure = \frac{(1+\beta) \times (Precision \times Recall)}{(\beta \times Precision + Recall)}, onde\ \beta = \frac{1-\alpha}{\alpha} \tag{3.3}$$

Entretanto, a tarefa de busca e recuperação de imagens (CBIR) é baseada no conceito de recuperação de informação, nesse caso, *precision* (referenciada aqui como $Precision_{CBIR}$) é definida como a razão entre o número de documentos recuperados que são relevantes para a consulta do usuário e o total de documentos recuperados (NIST, 2023):

$$Precision_{CBIR} = \frac{\text{number of relevant itens retrieved}}{\text{total number of itens retrieved}} \tag{3.4}$$

Por definição, a equação acima leva em consideração todos os documentos recuperados, mas existe a possibilidade de avaliar um determinado número k de documentos recuperados (*cut-off rank*), então temos a medida $Precision_{CBIR}$ *at* k ou $P@k$.

Em uma aplicação típica de CBIR, o usuário fornece uma imagem de consulta para recuperar imagens similares (positivas/mesmo rótulo) a partir de um conjunto de imagens, dessa maneira a posição dessas imagens na lista de imagens recuperadas é importante para medida de desempenho do sistema (TAN, 2019). Para o cálculo da métrica $Precision_{CBIR}$ nesse contexto, pode-se definir as seguintes variáveis:

- $X_q \in X$ é a imagem consultada
- X conjunto de imagens rotuladas
- d medida de similaridade entre imagens (Por exemplo, distância de Hamming entre vetores de atributos extraídos das imagens (*hash codes*))
- R subconjunto recuperado de X ordenado com base na medida de similaridade
- k é o índice para R

Considerando a hipótese de que o conjunto X possui 3 imagens como mesmo rótulo que X_q (documentos relevantes ou *Ground Truth Positives - GTP*). Depois de

calcular d para cada uma das imagens de X em relação a X_q, pode-se ranquear X e obter R. Assumindo que o modelo recuperou as imagens relevantes na seguinte ordem (*ranks*) $k = 1$, $k = 4$ e $k = 5$.

Usando a Equação 3.4 para n arquivos, obtém-se:

$$
\begin{aligned}
P@1 &= \frac{1}{1} = 1 \\
P@2 &= \frac{1}{2} = 0,5 \\
P@3 &= \frac{1}{3} = 0,33 \\
P@4 &= \frac{2}{4} = 0,5 \\
P@5 &= \frac{3}{5} = 0,6 \\
&\ddots \\
P@n &= \frac{3}{n}
\end{aligned}
$$

De forma complementar a métrica $Precision_{CBIR}$ utiliza-se o cálculo da precisão média (*Average Precision - AP*) para avaliar melhor a capacidade de um modelo de recuperar imagens similares, ou seja, que o subconjunto R possua nas primeiras posições somente imagens com mesmo rótulo. Para uma única consulta ela é calculada tomando a média dos valores de $Precision_{CBIR}$ obtidos após a recuperação de cada imagem relevante, com as imagens relevantes que não são recuperadas recebendo o valor de $Precision_{CBIR}$ igual a zero. Ela é considerada a medida eficácia (*Effectiveness*) do sistema (modelo) utilizado para recuperação informação (imagens) (TURPIN; SCHOLER, 2006).

Tan (2019) expressa a seguinte equação para o cálculo de AP:

$$
AP@n = \frac{1}{GTP} \sum_{k=1}^{n} P@k \times rel@k, \tag{3.5}
$$

onde GTP se refere ao número total de imagens de mesmo rótulo que X_q, n é o número de arquivos de interesse, $P@k$ e $rel@k$ são respectivamente o valor de $Precision_{CBIR}$ e *relevance* at k. A função de relevância ($rel@k$) é uma função indicadora igual a 1 se a imagem recuperada no índice k é positiva e 0 caso contrário.

Assumindo o conjunto de imagens X com $GTP = 3$, podemos calcular a AP geral considerando nossa imagem de consulta X_q:

$$\text{Overall AP} = \frac{1}{3}\left(\frac{1}{1} \times 1 + \frac{1}{2} \times 0 + \frac{1}{3} \times 0 + \frac{2}{4} \times 1 + \frac{3}{5} \times 1 + 0... + 0\right) = 0,7$$

Como resultado tem-se que o AP geral para esta consulta é 0,7. Sabendo-se que existem apenas 3 casos de imagens positivas recuperadas até o índice $k = 5$, implica que o AP@5 é igual ao AP geral. Essa métrica permite quantificar a qualidade do modelo em relação a capacidade de recuperar imagens relevantes baseada na medida de similaridade d, uma vez que penaliza modelos que não conseguem ordenar adequadamente o subconjunto R com relação aos casos de verdadeiros positivos (TAN, 2019).

Para cada imagem de consulta $X_q \in X$, pode-se calcular o AP correspondente. A métrica *mean Average Precision* (mAP) é a média de todas as consultas realizadas (Equação 3.6). Essa métrica é uma das mais populares para avalização de sistemas de recuperação de informação, como CBIR, apresentando estabilidade para diferentes tamanhos de conjuntos de consulta e variações do julgamentos de relevância (TURPIN; SCHOLER, 2006).

$$mAP = \frac{1}{|X|}\sum_{i=1}^{|X|} AP_i, \tag{3.6}$$

$|X|$ indica o número total de imagens do conjunto de dados.

A métrica $P@k$ representa a porcentagem de imagens relevantes recuperadas entre as k primeiras imagens recuperadas. Dessa maneira, $mAP@k$ indica a média da precisão alcançada para todas as consultas considerando $P@k$.

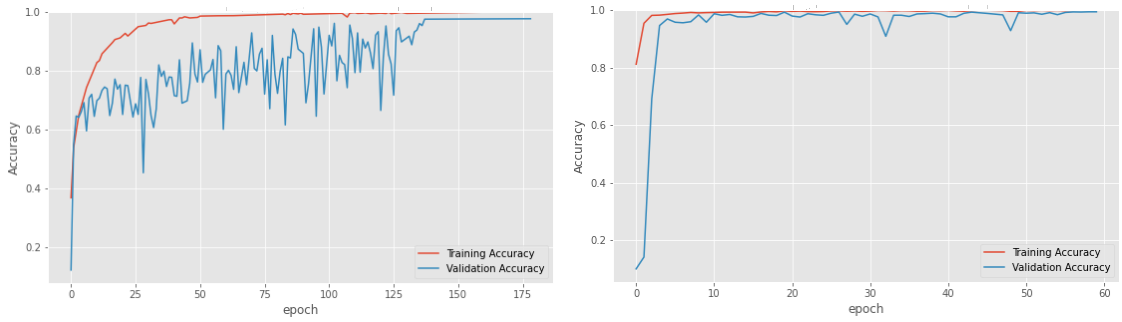
3.5 Modelos de *Deep Learning* aplicados ao SR

3.5.1 Classificação de imagens

As redes DL podem ser treinadas do zero (inicialização dos pesos de forma aleatória) ou ajustadas a partir de uma rede pré-treinada (*fine-tuning*). Com objetivo de atestar à efetividade do processo de *fine-tuning* para a tarefa de classificação de imagens de SR por satélite usando imagens de outros domínios (conjunto ImageNet), foram realizados experimentos com as redes ResNet-50 e ResNet-152.

Esses modelos possuem ótimos resultados para classificação de imagens de SR por satélites (HELBER et al., 2019; SUMBUL et al., 2020). Isso se deve principalmente a técnica de aprendizado residual baseada em blocos que contém conexões de atalho (*Residual Block*) para evitar o problema de degradação da precisão do treinamento em redes profundas (HE et al., 2016).

Figura 3.7 - Treinamento ResNet-152 com inicialização aleatória dos pesos (a) e pré-treinada com o conjunto ImageNet (b).



(a) Inicialização aleatória dos pesos (b) Pré-treinada com ImageNet

Fonte: Próprio autor.

Efetivamente, o uso de redes pré-treinadas com imagens fora do domínio do SR quando comparadas com inicialização aleatória dos pesos (Figura 3.7), propiciou melhorias tanto no tempo necessário para convergência do modelo (60 contra 140 épocas[6]) quanto em performance (0,9926 contra 0,9726 acurácia global). Foi possível determinar também que a rede ResNet-50 levou menos tempo de execução (≈ 12s contra 26s por época) e acertou mais que a ResNet-152 (0,9956 contra 0,9926 acurácia global) a classificação de uso e cobertura da terra baseada no conjunto de dados multiespectrais EuroSAT.

Os experimentos aqui descritos, visam indicar a partir de modelos "clássicos" já explorados pela comunidade de SR, aquele com maior potencial para extração semântica de atributos de imagens (*backbone*)[7] para melhoria do processo de CBIR.

[6]Cada época corresponde a uma varredura completa de todo o conjunto de treinamento (DSA,).

[7]No contexto do ML, um *backbone* é descrito como uma rede padrão normalmente empregada para classificação de imagens, mas sem a camada final de classificação. Esse arranjo permite utilizar o potencial dessa rede para extração de características das imagens para aplicação em outros tipos de tarefas como detecção de objetos, CBIR entre outras, normalmente com adição de algumas camadas auxiliares ao *backbone* (ELHARROUSS et al., arXiv:2206.08016, 2022).

Existem muitos outros modelos propostos na literatura para classificação de imagens de SR, além daqueles baseados em aprendizagem residual (YASSINE et al., 2021). Porém, as *Residual Neural Networks* (ResNets) são frequentemente mencionadas com desempenho superior em comparação com outros modelos de classificação (HELBER et al., 2019; SUMBUL et al., 2021). A opção pela rede ResNet-50 neste trabalho teve como base características como velocidade, precisão e tamanho da rede, considerando o *trade-off* entre desempenho e custo computacional.

3.5.2 *Content-Based Image Retrieval* (CBIR)

A *Metric-Learning-Based Deep Hashing Network* (MiLaN) é considerada o estado da arte para busca e recuperação de imagens baseada em conteúdo (CBIR) no escopo do SR (KAPOOR et al., 2021). Ela utiliza a rede Inception Net (SZEGEDY et al., 2016) pré-treinada com o conjunto ImageNet como módulo intermediário para extração de características das imagens (*backbone*). O vetor de atributos fornecido pelo *backbone* é utilizado para treinamento da rede MiLaN[8] resultando na construção de um espaço métrico otimizado para tarefa de CBIR através do aprendizado profundo de uma função de *hashing*. Esse processo permite mapear o descritor de imagem de alta dimensão para um vetor de códigos binários, o que melhora significativamente o desempenho para recuperação de imagens tanto em tempo quanto em precisão além de reduzir o custo necessário para armazenamento do espaço métrico (ROY et al., 2021).

Contudo, o uso de imagens fora do domínio do SR por satélite limita a representação semântica dessas imagens, uma vez que possuem uma série de características distintas, como influência da atmosfera, resolução espacial, entre outras. Como consequência a tarefa de CBIR também fica prejudicada nesse caso.

Esse trabalho apresenta uma solução para esse problema treinando diferentes arquiteturas de DL (Inception Net e ResNet) com imagens de SR (AID e EuroSAT) identificando assim um *backbone* adequado para extração de recursos intermediários de imagens de satélite usados para ajustar a rede MiLaN otimizando a tarefa de CBIR.

O emprego de imagens de outro domínio limita, mas não prejudica a extração de características pelos *backbones*, por isso foram utilizadas redes pré-treinadas com o

[8] A rede MiLaN é composta por 3 camadas totalmente conectadas com 1024, 512 e K neurônios cada, sendo K o comprimento do vetor de *hash codes* desejado. Utiliza a função LeakyReLU nas duas camadas ocultas para permitir valores negativos na retropropagação de erros e uma ativação sigmoide na camada final para restringir a saída a valores [0, 1] (ROY et al., 2021).

conjunto ImageNet. Este processo produz melhores resultados para o reconhecimento de padrões específicos nas imagens (texturas, bordas, etc) gerando assim abstrações de qualidade das imagens (vetor de atributos) que servem para o ajuste fino da rede MiLaN.

Uma vez treinada, a rede MiLaN é utilizada para criar um espaço métrico otimizado para CBIR baseado em códigos *hash* binários para recuperação rápida e precisa de imagens com base no cálculo de similaridade (distância de Hamming) entre uma imagem de consulta e as outras imagens nesse espaço.

Esse trabalho adiciona mais uma contribuição para melhoria do processo de CBIR no escopo do SR por satélite, através da agregação das informações multiespectrais (13 bandas) do Sentinel. Isso permite o aperfeiçoamento na identificação de imagens de uso e cobertura com padrões e formas semelhantes, mas com respostas espectrais diferentes, por exemplo, rodovias e rios.

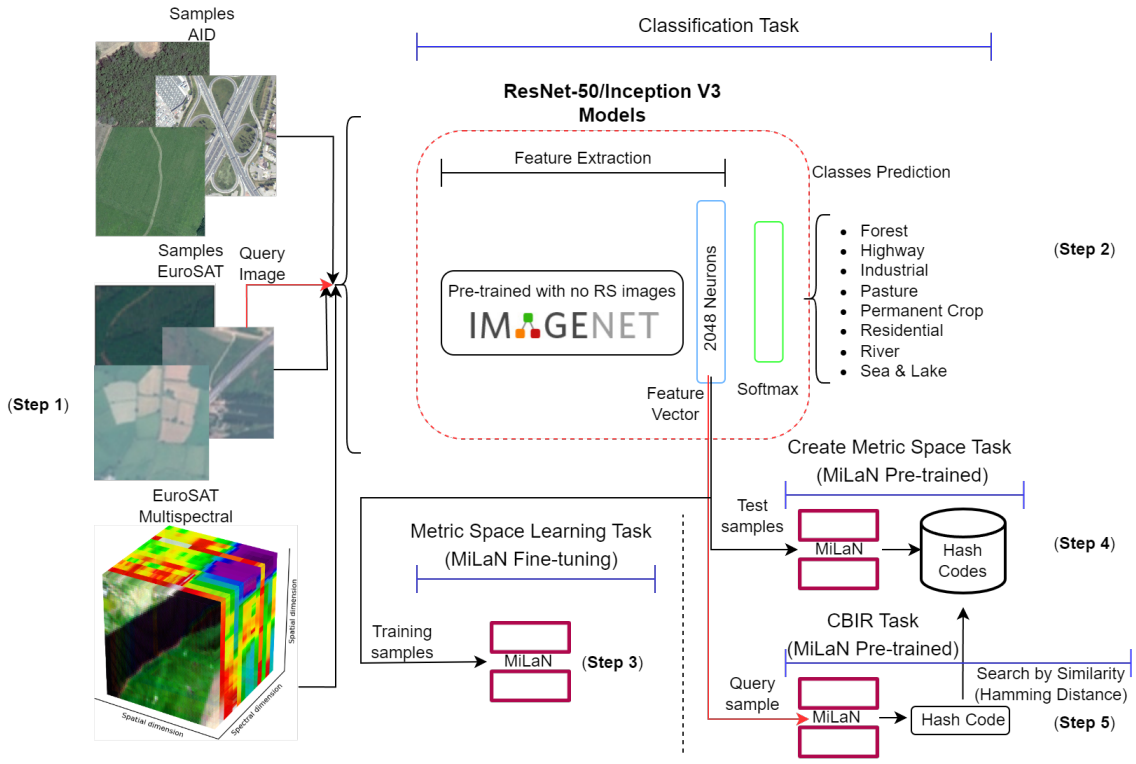
3.5.3 *Framework* para o CBIR de imagens de SR

Os seguintes modelos (Inception Net V3/ResNet-50) foram utilizados para comparação e avaliação dos *backbones* quanto a capacidade de gerar melhores abstrações de imagens úteis para o treinamento da MiLaN. A rede Inception Net foi usada inicialmente pelos autores da MiLaN, mas sem o *fine-tuning* utilizando imagens de SR por satélite como implementado aqui. Por outro lado, a ResNet é frequentemente mencionada com desempenho superior em comparação com outros modelos de classificação (HELBER et al., 2019; SUMBUL et al., 2021).

O *backbone* usa imagens quadradas de 600x600 (AID) ou 128x128 (EuroSAT) pixels com 3 canais espectrais como entrada (RGB) e 13 canais para o conjunto de dados multiespectral EuroSAT. As imagens EuroSAT de 64×64 pixels foram redimensionadas com o método do vizinho mais próximo para adaptar a entrada do modelo Inception Net V3[9]. Todos os modelos possuem um mapa de características finais de 1×1 pixel com 2.048 neurônios que possibilitam a previsão final de classes usando a função Softmax. Esta camada representa todo o conhecimento e abstrações que a rede extraiu da imagem e codificou em informações relevantes e aprimoradas servindo de base para o ajuste da rede MiLaN (Figura 3.8).

[9] A Inception Net V3 possui restrição quanto ao tamanho mínimo de entrada, somente imagens com no mínimo 75×75 pixels são aceitas (SZEGEDY et al., 2016).

Figura 3.8 - *Framework* utilizado para o treinamento dos *backbones*, aprendizado do espaço métricos (MiLaN) e recuperação de imagens baseada na medida de similaridade (distância de Hamming).



Fonte: Próprio autor.

A Figura 3.8 ilustra as seguintes etapas necessárias para construção do *framework* que permitiu a avaliação do potencial de utilização de imagens aéreas e de satélite multiespectrais (além de RGB) para aplicações em CBIR:

1. Preparação dos dados: criação de conjuntos de dados balanceados, redimensionados, aumentados e aprimorados;

2. Tarefa de classificação: treinamento das redes a serem utilizadas como *backbones* para MiLaN;

3. Tarefa de aprendizagem de espaço métrico: usar recursos extraídos pelos *backbones* pré-treinados para ajustar a rede MiLaN através do aprendizado do espaço métrico baseado na informação semântica das imagens, onde os recursos são otimizados para tarefa de CBIR;

4. Criação do espaço métrico: usa a MiLaN ajustada para gerar códigos *hash* binários compactos (64 bits) que são armazenados propiciando a pesquisa e recuperação ágeis de imagens baseadas no conteúdo;

5. Tarefa CBIR: usa a MiLaN ajustada para gerar o vetor de *hash codes* de uma imagem de consulta permitindo assim recuperar imagens similares com base na distância de Hamming.

3.5.3.1 Preparação dos dados, classificação e CBIR

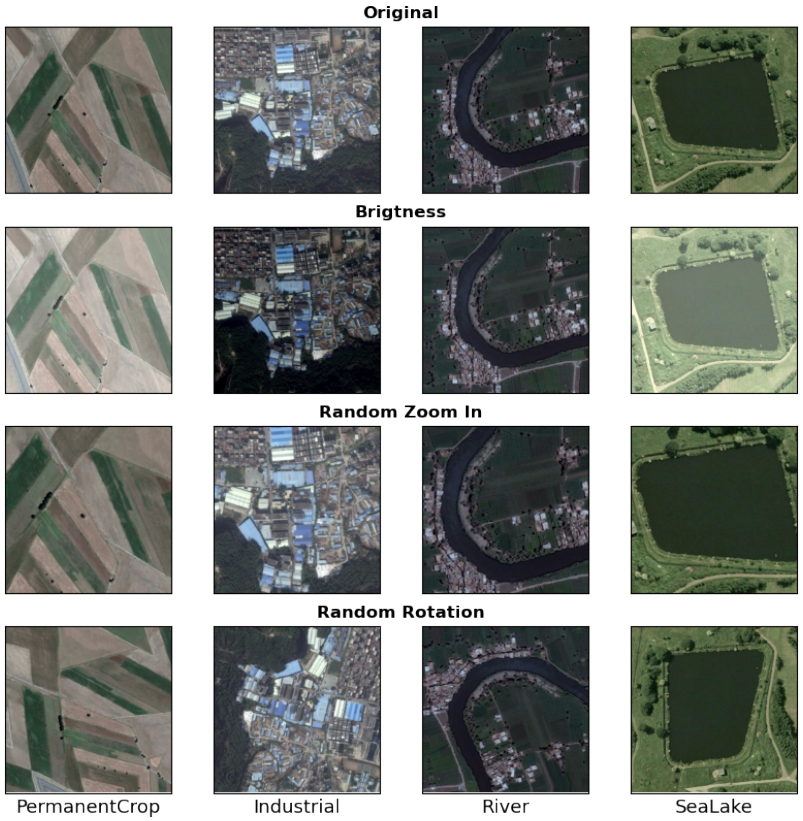
Neste trabalho são apresentadas algumas inovações em relação a abordagem original da MiLaN: i) uso de conjuntos de imagens suborbitais (AID) e orbitais (EuroSAT) explorando diferentes características de resolução espacial e espectrais (RGB/MS); ii) avaliação de novos *backbones* treinados com imagens de SR para observação da Terra com informação multiespectral para melhoria do processo de CBIR.

Conforme já descrito na Seção 3.1.2, o conjunto de dados AID possui uma série de imagens rotuladas que apresentam semelhança com as imagens do conjunto de dados EuroSAT. Um fator limitante para a intercomparação dos resultados é o número restrito de amostras de algumas classes, por exemplo, a classe *Forest* possui apenas 250 amostras no AID. O desbalanceamento no conjunto de dados afeta o aprendizado e as previsões dos modelos de ML (LEMAÎTRE et al., 2017).

Para superar o desequilíbrio de classes e aumentar os conjuntos de treinamento foram empregadas 7 técnicas de *data augmentation* que incluem inversão horizontal/vertical, rotações, *zooms* e alterações do fator de brilho[10] (Figura 3.9).

[10]De acordo com Chollet (2017), o *data augmentation* (DA) gera mais amostras para treinamento a partir dos dados preexistentes, de modo que no momento do treinamento os modelos de classificação nunca vejam exatamente a mesma amostra duas vezes, expondo o modelo a mais aspectos dos dados permitindo assim uma melhor generalização. No entanto, as entradas ainda estão fortemente correlacionadas, porque provêm de um pequeno número de imagens originais. Com DA não há informações novas, é apenas um remix de informações existentes.

Figura 3.9 - Exemplo de imagens do conjunto AID com *data augmentation*.



Fonte: Próprio autor.

Como resultado ambos os conjuntos AID/EuroSAT foram definidos com 16.000 amostras de imagens com 2.000 para cada classe: *Viaduct*, *Sea & Lake*, *Pasture*, *Permanent Crop*, *Industrial*, *Forest*, *Residential* e *River*.

Os dados foram divididos em 80%/10%/10% conjuntos de treinamento/validação/teste para a tarefa de Classificação (2^a etapa - ajuste fino dos *backbones*), enquanto para a tarefa de Aprendizagem do Espaço Métrico (3^a etapa - ajuste fino da MiLaN) a divisão foi de 80%/20% treinamento/teste. Os treinamentos foram realizados com a opção de interrupção caso não houvesse melhoria dado um número discreto de iterações (*Early Stopping Callback Function*), 9.000 iterações ($\approx$ 40 épocas) na 2^a etapa e 1.000 iterações na 3^a etapa.

O conjunto de dados EuroSAT, especialmente as imagens MS, apresentam alguns problemas como imagens com pixels saturados (amostras escuras) e opacas (amos-

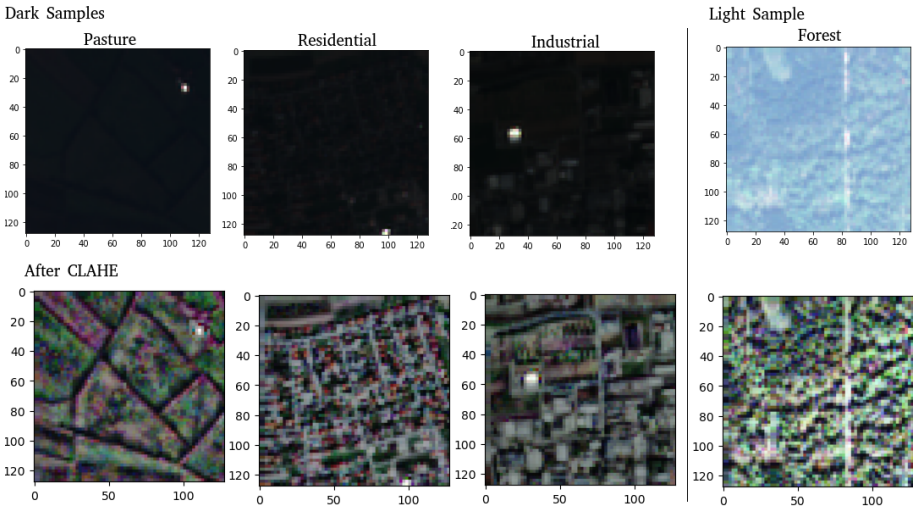
tras claras)[11] que afetam a correta identificação do uso e cobertura da terra. Esses problemas geralmente decorrem do mau funcionamento dos detectores presentes nos sensores e da falta de correção atmosférica (HELBER et al., 2019).

Para lidar com esses problemas foi empregada a técnica *Contrast Limited Adaptive Histogram Equalization* (CLAHE), amplamente utilizada para o aprimoramento de contraste, especialmente em imagens médicas (ZUIDERVELD, 1994).

A CLAHE utiliza o processo adaptativo de equalização de histograma baseado em pequenos blocos de 8×8 pixels, em vez de uma equalização de contraste global. Além disso, antes de aplicar a equalização do histograma para correção do contraste é empregado um limiar que refaz a distribuição dos valores de modo a evitar a possibilidade de amplificação de ruídos nessas pequenas áreas, algo que ocorria na técnica *Adaptive Histogram Equalization* (AHE).

Tomando como base o espalhamento Rayleigh para a radiação medida na vizinhança de um pixel, foi adotado o limiar de 0,01 para imagens MS do EuroSAT (VIDHYA; RAMESH, 2017). Isso permitiu a construção de um conjunto de imagens com informações enriquecidas (Figura 3.10), contribuindo assim para melhoria do processo de classificação e por consequência do processo de CBIR.

Figura 3.10 - Exemplos de imagens EuroSAT corrigidas com o CLAHE.



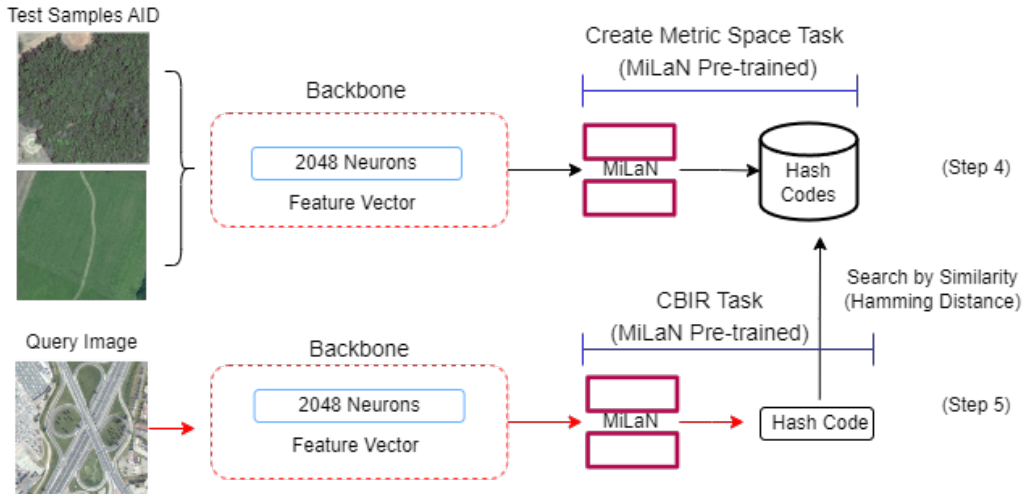
Fonte: Próprio autor.

[11] Possivelmente afetadas pelo fenômeno atmosférico *Haze*.

As etapas seguintes, 4^a etapa - Construção do espaço métrico e 5^a etapa - CBIR, são realizadas utilizando a rede MiLaN ajustada na etapa anterior com imagens do conjunto AID e EuroSAT (RGB/MS).

O espaço métrico otimizado para tarefa CBIR é construído com o conjunto de imagens de teste (20% do conjunto total para cada classe). A partir dos vetores de atributos extraídos pelo *backbone*, a rede MiLaN treinada produz um conjunto de vetores com *hash codes* os quais são armazenados para busca e recuperação de imagens por similaridade. Ao apresentar uma imagem de interesse (*Query image*) a rede MiLaN, é possível gerar a sua representação em *hash codes* para o cálculo de similaridade usando a distância de Hamming (Figura 3.11).

Figura 3.11 - Construção do espaço métrico (4^a etapa) e tarefa CBIR (5^a etapa) usando a rede MiLaN.



Fonte: Próprio autor.

3.5.4 Identificação de uso e cobertura da terra baseada em CBIR

Neste experimento foi explorado o uso do *framework* proposto para CBIR de imagens de satélite de maneira a viabilizar à identificação de tipos de uso e cobertura da terra em imagens da região do Cerrado, com base em modelos treinados com dados do conjunto EuroSAT.

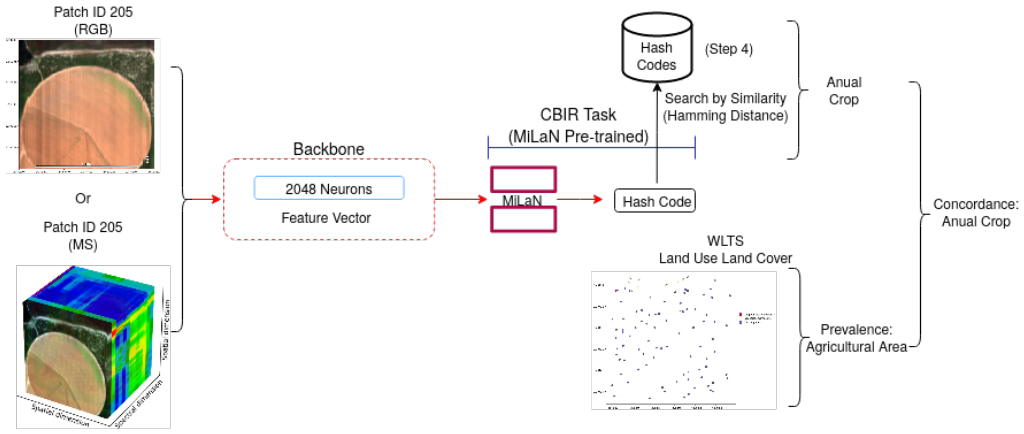
A transferência de conhecimento do aprendizado produzido com esses dados que representam áreas com coberturas e respostas espectrais de diferentes regiões (vegetação e construções no continente Europeu) para identificação de uso e cobertura na região do Cerrado brasileiro, corresponde ao que é definido no *Machine Learning* como *transfer learning* para adaptação de domínio (SARKAR; BALI, 2022).

Foram utilizados dois tamanhos de *patches* de imagens (128x128 e 64x64) de reflectância à superfície gerados a partir do *tile* 089097 do cubo de dados Sentinel 2-16D de 19 de dezembro de 2018 (ver Seção 3.2.1), com objetivo de avaliar a influência dessa característica (tamanho da imagem de entrada) na correta representação semântica dos alvos e melhoria da identificação dos padrões de uso e cobertura.

A validação foi realizada com base nas amostras de uso e cobertura da terra providas pelo mapeamento realizado pelo IBGE para todo o território nacional referente ao ano de 2018 (IBGE, 2020) adquiridas através do serviço WLTS do projeto BDC.

Empregando redes MiLaN pré-treinadas com dados EuroSAT (RGB/Multiespectrais) foi possível atribuir, a cada *patch* de imagem da área de estudo, o tipo de uso e cobertura da terra baseada na maior similaridade medida pela menor distância de Hamming em relação as imagens de teste do conjunto EuroSAT (Figura 3.12).

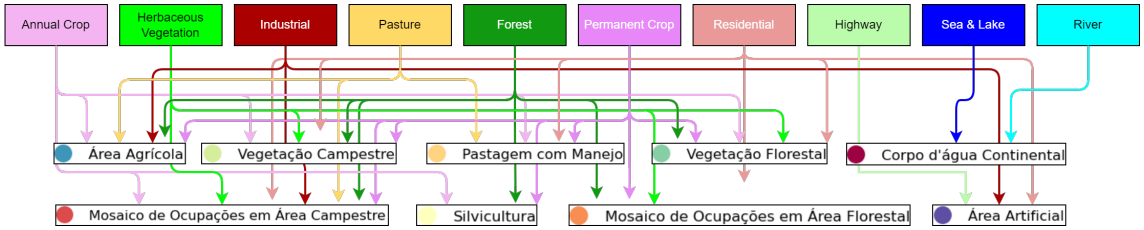
Figura 3.12 - Exemplo de identificação de uso e cobertura da terra no Cerrado usando CBIR (MiLaN).



Fonte: Próprio autor.

As áreas mapeadas pelo IBGE possuem semelhanças com o tipo de uso e cobertura identificados no conjunto EuroSAT, mas com certas particularidades mais bem descritas no documento fornecido pelo instituto[12]. Dessa maneira para intercomparação entre o resultado do processo de identificação feito pela rede MiLaN e os dados do IBGE foram necessárias a adoção da seguintes considerações de concordâncias (Figura 3.13).

Figura 3.13 - Concordância entre os tipos de uso e cobertura da terra identificados nas imagens do conjunto EuroSAT e o mapeamento feito pelo IBGE.



Tipos de usos e cobertura da terra do EuroSAT identificados pelos retângulos na parte superior e do IBGE pelos círculos na parte de baixo.

Fonte: Próprio autor.

As características de algumas áreas presentes no mapeamento feito pelo IBGE são comuns à características encontradas em mais de um tipo de uso e cobertura identificado nas imagens EuroSAT, por essa razão essas amostras foram atribuídas a mais de um tipo de classe. Por exemplo, "Área Agrícola ... inclui todas as áreas cultivadas, inclusive as que estão em pousio ou localizadas em terrenos alagáveis. Pode ser representada por zonas agrícolas heterogêneas... (IBGE, 2020)", não sendo incomum encontrar nessas áreas plantio de Florestas. Outra associação que chama atenção é o de Áreas Campestre e Agrícola a *Industrial*, isso se deve principalmente a identificação de galpões de Agroindústria cercados por áreas de cultivo.

[12]Descrição das características do mapeamento do uso e cobertura da terra pelo IBGE. Disponível em <https://geoftp.ibge.gov.br/informacoes_ambientais/cobertura_e_uso_da_terra/monitoramento/grade_estatistica/serie_revisada_2022/vetores_compactados/simbologia_cobertura.zip>, acesso 6 jan. 2022.

4 RESULTADOS

Este capítulo apresenta os resultados alcançados para busca e recuperação de imagens baseadas em conteúdo com emprego de métodos de DL (Inception V3/ResNet-50) para extração de informação semântica das imagens e geração de códigos *hash* (MiLaN) para indexação dessa informação de maneira a possibilitar a recuperação ágil.

Foi possível verificar, de maneira quantitativa e qualitativa, o impacto das inovações propostas em relação a abordagem original da MiLaN. O emprego da rede com aprendizado residual permitiu melhores abstrações das imagens, impactando a tarefa de recuperação de imagens baseada em conteúdo. Além disso, o processo de *fine-tuning* com dados do escopo do sensoriamento remoto por satélite, combinada a adição de outras bandas espectrais, garantiu a rede MiLaN o potencial para criação de um espaço métrico otimizado para recuperação de imagens nesse escopo.

Adicionalmente, foi explorado o uso do *framework* proposto neste trabalho para a identificação de uso e cobertura da terra (*Land Use Land Cover* - LULC) em uma área do Cerrado brasileiro. Esse experimento teve o objetivo de testar a transferência de aprendizado para identificação de uso e cobertura similares, mas de regiões distintas (Continente Europeu - EuroSAT).

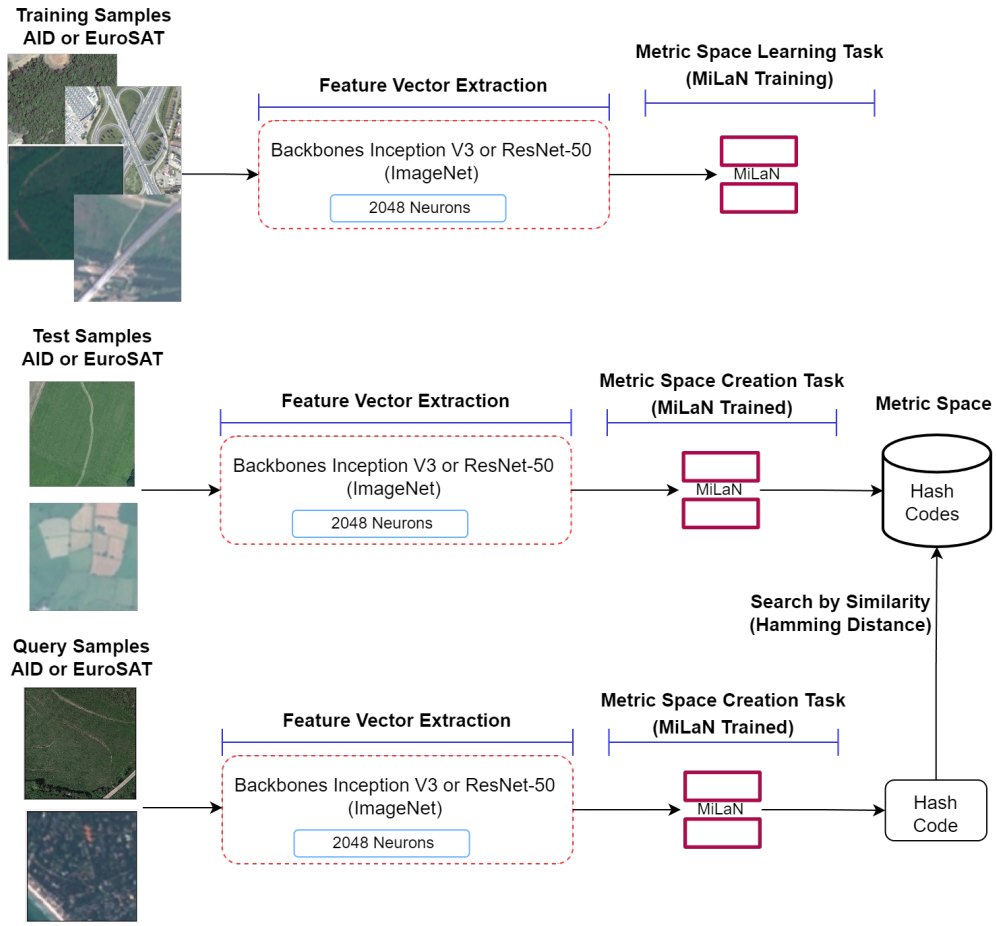
Os resultados estão organizados de acordo com os seguintes experimentos:

- Experimento 1 (seção 4.1) - buscou verificar o potencial da rede MiLaN original na tarefa de CBIR com imagens satelitais de média resolução espacial (EuroSAT) comparando com resultados obtidos com imagens de alta resolução espacial (AID). Além disso, testou também a adoção de um novo *backbone* baseado na rede ResNet-50;
- Experimento 2 (seções 4.2 e 4.3) - foi baseado no *framework* proposto neste trabalho, que utiliza a complementação (bandas multiespectrais) e equalização dos dados (Data Augmentation/CLAHE) para melhoria do processo de CBIR;
- Experimento 3 (seção 4.4) - usou o conceito de *transfer learning* para adaptação de domínio, empregando *framework* proposto para identificação de uso e cobertura da terra em uma área do Cerrado.

4.1 CBIR de imagens satelitais EuroSAT

A criação do espaço métrico para a tarefa de CBIR, envolve o processo de conversão de atributos de alta dimensão extraídos das imagens para atributos de baixa dimensão (*hash codes* - códigos binários) através da técnica de *Hashing* através da MiLaN. O comprimento dos *hash codes*, deve levar em conta o compromisso entre precisão e eficiência. Neste trabalho, foi adotado o comprimento de 64 bits, visando boa precisão e baixa demanda por armazenamento, conforme demonstrado em Roy et al. (2021). Além disso, foi incorporado o uso do *backbone* baseado em aprendizado residual (ResNet-50), solução proposta neste trabalho para melhoria do processo de CBIR com DL (Figura 4.1).

Figura 4.1 - Diagrama experimental das etapas necessárias para treinamento, teste e avaliação da MiLaN com os conjuntos AID e EuroSAT utilizando os *backbones* Inception V3 e ResNet-50.



Fonte: Próprio autor.

Os resultados deste experimento demonstram o potencial da rede MiLaN para construção de um espaço métrico ideal para CBIR, utilizando imagens de satélite do conjunto EuroSAT (RGB/media resolução 10 m). A validação dos resultados foi realizada comparando o desempenho com imagens aéreas do conjunto AID (RGB/alta resolução [0,5 - 8] m). A medida de desempenho foi realizada através da métrica *mean Average Precision* (mAP), principal métrica utilizada para avaliação de sistemas de recuperação de informações, como é o caso da tarefa de busca e recuperação de imagens baseadas em conteúdo (TURPIN; SCHOLER, 2006).

As avaliações baseadas em mAP *at* $k = 20$ (mAP@20), $k = 50$ (mAP@50) e $k = 100$ (mAP@100), ou seja, considerando as k primeiras imagens recuperadas a partir do conjunto de testes, seguiu o padrão adotado pelos autores da MiLaN, muito popular para esse tipo de aplicação (ROY et al., 2021). Com objetivo de facilitar identificação dos melhores desempenhos, são destacados nas tabelas os maiores valores alcançados levando-se em conta o conjunto de dados e as arquiteturas empregadas.

Como demonstrado nas Tabelas 4.1 e 4.2, o desempenho do processo de CBIR foi inferior para o conjunto EuroSAT em ambos os casos (MiLaN+Inception V3 e MiLaN+ResNet-50), embora o emprego da ResNet-50 tenha demonstrado um ganho de desempenho para quase todos os tipos de LULC nos dois conjuntos de dados.

Tabela 4.1 - Desempenho do CBIR (MiLaN+*backbones*) de imagens aéreas (AID) RGB, considerando k imagens de cada classe recuperadas pela menor distância de Hamming sendo $k = \{20, 50, 100\}$.

	AID Inception V3			**AID ResNet-50**		
LULC	**mAP@20**	**mAP@50**	**mAP@100**	**mAP@20**	**mAP@50**	**mAP@100**
Permanent Crop	0,9822	0,9813	0,9803	**0,9932**	**0,9923**	**0,9910**
Forest	0,9778	0,9761	0,9729	**0,9978**	**0,9961**	**0,9929**
Industrial	0,8543	0,8411	0,8273	**0,9943**	**0,9911**	**0,9903**
Pasture	0,9837	0,9789	0,9705	**0,9937**	**0,9889**	**0,9805**
Sea & Lake	0,9219	0,9234	0,9221	**0,9999**	**0,9911**	**0,9855**
River	0,9497	0,9552	0,9582	**0,9993**	**0,9966**	**0,9908**
Residential	0,9402	0,9363	0,9305	**0,9902**	**0,9863**	**0,9805**
Highway	0,9812	0,9755	0,9711	**0,9978**	**0,9970**	**0,9950**

Para verificar se o ganho de desempenho obtido com a abordagem utilizando a ResNet-50 sobre a abordagem utilizando a Inception V3 é estatisticamente significativo segundo algum nível de confiança, foram realizados testes pareados de Wilcoxon comparando as medidas $mAP@20$, $mAP@50$ e $mAP@100$ para cada conjunto de imagem.

Tabela 4.2 - Desempenho do CBIR (MiLaN+*backbones*) de imagens satelitais EuroSAT RGB, considerando k imagens de cada classe recuperadas pela menor distância de Hamming sendo $k = \{20, 50, 100\}$.

	EuroSAT Inception V3			**EuroSAT ResNet-50**		
LULC	**mAP@20**	**mAP@50**	**mAP@100**	**mAP@20**	**mAP@50**	**mAP@100**
Permanent Crop	0,9020	0,8941	0,8861	**0,9275**	**0,9285**	**0,9306**
Forest	0,9740	0,9715	0,9706	**0,9768**	**0,9766**	**0,9772**
Industrial	0,9631	0,9613	0,9606	**0,9887**	**0,9895**	**0,9902**
Pasture	**0,9380**	**0,9380**	**0,9399**	0,8643	0,8919	0,9078
Sea & Lake	0,9714	0,9669	0,9643	**0,9816**	**0,9815**	**0,9818**
River	0,8894	0,8824	0,8797	**0,9717**	**0,9723**	**0,9734**
Residential	0,9700	0,9682	0,9666	**0,9885**	**0,9895**	**0,9904**
Highway	0,8975	0,8975	0,8988	**0,9704**	**0,9552**	**0,9583**
Annual Crop	0,9239	0,9177	0,9140	**0,9482**	**0,9481**	**0,9488**
Herbaceous Vegetation	**0,9346**	**0,9317**	**0,9309**	0,9184	0,9185	0,9211

Basicamente, o teste utiliza métodos de ranqueamento para substituir os valores numéricos dos conjuntos testados, para obter uma rápida aproximação da significância estatística das diferenças entre as amostras de valores desse tipo de experimento (pareado) (WILCOXON, 1992).

Foi utilizado o teste *Wilcoxon signed-rank* implementado em Python no pacote Scipy[1]. Ele avalia a hipótese nula de que duas amostras pareadas relacionadas (mesmo LULC) vêm da mesma distribuição. Em particular, testa se a distribuição das diferenças $d = x - y$ é simétrica em relação a zero. O método fornece como resultado dois valores: *statistic* - a soma dos valores de ranque das diferenças acima ou abaixo de zero, o que for menor (*two-sided*). Caso contrário, a soma dos valores de ranque das diferenças acima de zero (*greater*); *p-valor* - a probabilidade de ocorrência de um certo valor de ranqueamento, define o grau de significância da hipótese testada (WILCOXON, 1992).

O desempenho do CBIR medido através dos valores de métrica mAP utilizando o *backbone* ResNet-50 parece superior ao alcançado com a rede Inception V3. Para testar a hipótese nula de que não há diferença entre os desempenhos obtidos, foi definida a hipótese alternativa *two-sided*, ou seja, que a distribuição de d não é simétrica em relação a zero.

Com base nos valores observados na Tabela 4.3, é possível aceitar a hipótese nula com um nível de significância de 5%, somente para o conjunto EuroSAT considerando

[1]SciPy é um software de código aberto para matemática, ciências e engenharia. Documentação para o teste *Wilcoxon signed-rank* está disponível em <https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.wilcoxon.html#scipy.stats.wilcoxon>. Acesso em: 21 jan. 2024.

as medidas mAP@20 e mAP@50, concluindo que nesses casos, não existe diferença estatisticamente significativa quando utilizada a ResNet-50 ao invés da Inception V3. Por outro lado, para todos os outros casos com mesmo nível de significância, pode-se rejeitar a hipótese nula assumindo que há diferença de desempenho entre as duas abordagens.

Sendo assim, para confirmar que a mediana das diferenças de desempenho com a ResNet-50 em relação a Inception V3 pode ser assumida como positiva (superior), foi definida a hipótese alternativa *greater* que significa que a distribuição de d é estocasticamente maior do que uma distribuição simétrica em relação a zero. Novamente recorrendo aos valores da Tabela 4.3, pode-se rejeitar a hipótese nula de que a mediana é negativa a um nível de significância de 5% em favor da alternativa de que a mediana é maior que zero. Portanto, para o conjunto AID (todas as medidas) e EuroSAT (mAP@100) a diferença de desempenho quando utilizada ResNet-50 é estatisticamente significante com 95% de confiança.

Tabela 4.3 - Teste de Wilcoxon para avaliar a significância estatística da diferença entre o desempenho do CBIR utilizando como *backbones* as redes Inception V3 e ResNet-50 para ambos os conjuntos de imagens AID/EuroSAT RGB.

	AID				**EuroSAT**			
	Alternative two-sided		**Alternative greater**		**Alternative two-sided**		**Alternative greater**	
	statistic	**p-value**	***statistic***	**p-value**	***statistic***	**p-value**	***statistic***	**p-value**
mAP@20	0	0,0078	36	0,0039	12	0,1308	43	0,0654
mAP@50	0	0,0078	36	0,0039	10	0,0840	45	0,0420
mAP@100	0	0,0078	36	0,0039	8	0,0488	47	0,0244

Na comparação dos resultados com o conjunto EuroSAT usando as duas arquiteturas, vale a pena destacar que ResNet-50 foi capaz de produzir informações semanticamente suficientes para que a rede MiLaN pudesse recuperar com maior precisão as imagens de Rios e Estradas, por outro lado houve uma confusão maior entre Pastagem e Vegetação Herbácea.

Esse primeiro experimento permitiu algumas descobertas: i) *backbones* pré-treinados com o conjunto ImageNet não são ideais, mas suficientes para a melhoria do processo de CBIR de imagens de SR, quando comparado com a inicialização dos pesos de forma aleatória e corroborado pelo desempenho elevado com as imagens do conjunto AID; ii) a resolução espacial das imagens resultou em um desempenho inferior para o conjunto EuroSAT.

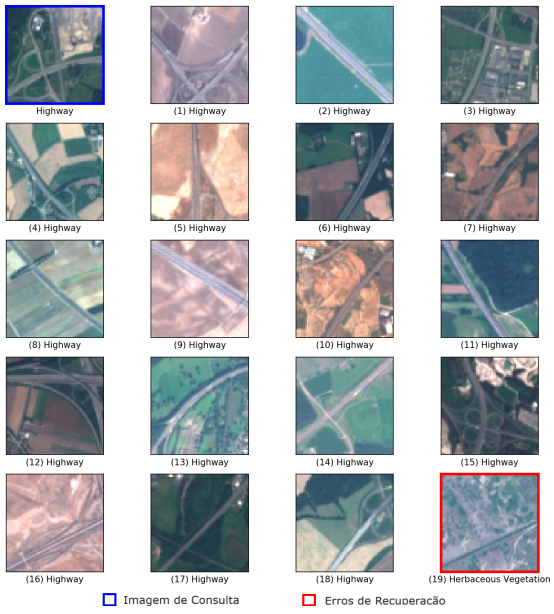
Além da resolução espacial, outro fator que impactou os resultados do conjunto EuroSAT são imagens que apresentam padrão de cobertura do uso da terra e geometria dos alvos similares (Figura 4.2). Nesses casos os atributos extraídos das imagens RGB de média resolução não foram suficientemente discriminativos para algumas classes (*Pemanent Crop, Annual Crop, Highway* e *Pasture*). O uso da arquitetura ResNet-50 como módulo de extração de características ao invés da Inception V3, melhorou os resultados para a maioria das classes de LULC, tanto para o conjunto AID quanto para o conjunto EuroSAT.

Apesar da melhoria alcançada com a ResNet-50, ainda persistem deficiências para a correta identificação de alguns tipos de uso e cobertura terra. Esses erros possivelmente estão ligados às áreas que recebem um rótulo ("correto"), mas que possuem a presença de outros tipos de uso e cobertura, como é o caso de algumas imagens contendo Rios com muita vegetação no entorno (Pastagem ou Vegetação Herbácea), como citado pelos autores do conjunto EuroSAT "[...] é mostrado que o classificador às vezes confunde as classes de terras agrícolas, bem como as classes Rodovia e Rio[...]" (HELBER et al., 2019, tradução nossa).

Figura 4.2 - Resultados da recuperação de imagens com a rede MiLaN para as classes *Permanent Crop* e *Highway* do conjunto EuroSAT, ordenados por grau de similaridade.

(a) CBIR EuroSAT MiLaN+Inception V3.

(b) CBIR EuroSAT MiLaN+Inception V3.



(c) CBIR EuroSAT MiLaN+ResNet-50.

As imagens estão ordenadas da 1^a a 19^a por grau de similaridade com a imagem de consulta.

Fonte: Próprio autor.

4.2 Fine-tuning usando dados multiespectrais do conjunto EuroSAT

Esse experimento teve como objetivo apresentar uma solução para os problemas identificados no CBIR de imagens de satélite de média resolução do conjunto EuroSAT, baseado na premissa que o uso da informação multiespectral contribui para a melhoria do processo de classificação (HELBER et al., 2019; SUMBUL; DEMIR, 2020) e que essa melhoria possivelmente impactaria no processo de CBIR propiciando uma busca e recuperação de imagens mais precisa para padrões de cobertura similares com resposta espectral distintas como é o caso de Estradas e Rios.

Foram realizados testes para ajuste dos *backbones* para classificação de imagens de LULC do conjunto EuroSAT utilizando as 13 bandas espectrais do sensor MSI do Sentinel 2 e posteriormente adoção destes para a tarefa de CBIR através da rede MiLaN também ajustada utilizando o dado multiespectral. A acurácia alcançada com a rede ResNet-50 foi de 99,56%, superior ao valor 98,57% relatado no artigo de referência (HELBER et al., 2019).

A Tabela 4.4 apresenta o relatório de desempenho baseado nas métricas Precisão, Revocação e F1-Score para classificação de imagens quanto ao tipo de LULC de imagens RGB e multiespectrais do conjunto EuroSAT, evidenciando a superioridade alcançada com uso complementar de outras bandas além das bandas RGB.

Tabela 4.4 - Desempenho para classificação de imagens dos conjuntos EuroSAT (RGB/MS) utilizando a ResNet-50 pré-treinada com ImageNet.

	EuroSAT RGB			**EuroSAT MS**		
LULC	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**
Highway	1	**0,9959**	**0,9979**	**1**	0,9924	0,9962
SeaLake	1	1	1	**1**	**1**	**1**
Pasture	0,9798	0,9898	0,9848	**0,9953**	**0,9906**	**0,9929**
PermanentCrop	**0,9959**	0,9759	0,9858	0,9922	**0,9846**	**0,9884**
Industrial	0,9886	**0,9962**	**0,9924**	**0,9918**	0,9918	0,9918
Forest	0,9965	0,9965	0,9965	**1**	**1**	**1**
Residential	**0,9931**	0,9965	0,9948	0,9928	**1**	**0,9964**
AnnualCrop	0,9897	0,9897	0,9897	**1**	**0,9967**	**0,9984**
HerbaceousVegetation	**0,9873**	0,9905	0,9889	0,9835	**0,9967**	**0,9900**
River	0,9957	0,9957	0,9957	**1**	**1**	**1**

A Tabela 4.5 apresenta a comparação dos resultados do uso combinado de dados RGB (AID/EuroSAT) e multiespectrais (EuroSAT) com aprendizado residual (ResNet-50). O emprego da informação multiespectral permitiu melhorar o processo de construção do espaço métrico, otimizando assim os resultados para o CBIR com esses dados. O desempenho da MiLaN com esse conjunto de imagens de média resolução passou a ser equivalente e em alguns casos até superior ao desempenho de imagens de alta resolução (AID).

Tabela 4.5 - Desempenho do CBIR de imagens aéreas (AID) e satelitais RGB e Multiespectrais (EuroSAT)*, considerando k imagens de cada classe recuperadas pela menor distância de Hamming sendo $k = 20, 50, 100$.

	AID ResNet-50			**EuroSAT RGB ResNet-50**			EuroSAT MS ResNet-50		
LULC	**mAP@20**	**mAP@50**	**mAP@100**	**mAP@20**	**mAP@50**	**mAP@100**	**mAP@20**	**mAP@50**	**mAP@100**
Permanent Crop	**0,9932**	**0,9923**	**0,9910**	0,9275	0,9285	0,9306	0,9914	0,9789	0,9780
Forest	**0,9978**	0,9961	0,9929	0,9768	0,9766	0,9772	0,9962	**0,9966**	**0,9969**
Industrial	0,9943	0,9911	0,9903	0,9887	0,9895	0,9902	**0,9949**	**0,9955**	**0,9959**
Pasture	**0,9937**	0,9889	0,9805	0,8643	0,8919	0,9078	0,9904	**0,9913**	**0,9918**
Sea & Lake	**0,9999**	0,9911	0,9855	0,9816	0,9815	0,9818	0,9989	**0,9989**	**0,9989**
River	**0,9993**	**0,9966**	0,9908	0,9717	0,9723	0,9734	0,9941	0,9943	**0,9944**
Residential	0,9902	0,9863	0,9805	0,9885	0,9895	0,9904	**0,9972**	**0,9973**	**0,9974**
Highway	**0,9978**	**0,9970**	**0,9950**	0,9704	0,9552	0,9583	0,9908	0,9913	0,9915
Annual Crop	-	-	-	0,9482	0,9481	0,9488	0,9881	0,9887	0,9892
Herbaceous Vegetation	-	-	-	0,9184	0,9185	0,9211	0,9806	0,9821	0,9827

* As classes *Annual Crop* e *Herbaceus Vegetation* do conjunto EuroSAT não possuem equivalentes no conjunto AID.

4.3 CBIR de imagens de SR com equalização dos dados

Tomando como base a definição de *framework* como sendo um conjunto de ideias, informações e princípios que formam um plano[2], foi apresentado no Capítulo Materiais e Métodos (3) as etapas que compõem o *framework* proposto neste trabalho para o desenvolvimento de um sistema CBIR.

Essas etapas buscam criar um conjunto de dados padronizados ("equalizados") tanto na quantidade de amostras de cada classe de LULC quanto nas dimensões, a fim de que possa haver comparações de desempenho mais adequadas entre os conjuntos de imagens. Além disso, foi utilizada a técnica CLAHE para melhoria do contraste e correção das inconsistências quanto a distribuição dos valores registrados nas imagens multiespectrais.

[2] "The ideas, information, and principles that form the structure of an organization or plan". Disponível em <https://dictionary.cambridge.org/us/dictionary/english/framework>. Acesso em: 01 Setembro 2023.

A Tabela 4.6 sumariza os resultados alcançados nesse experimento, ratificando a efetividade do emprego da informação multiespectral nessa abordagem, potencializada pela melhoria do contraste (CLAHE).

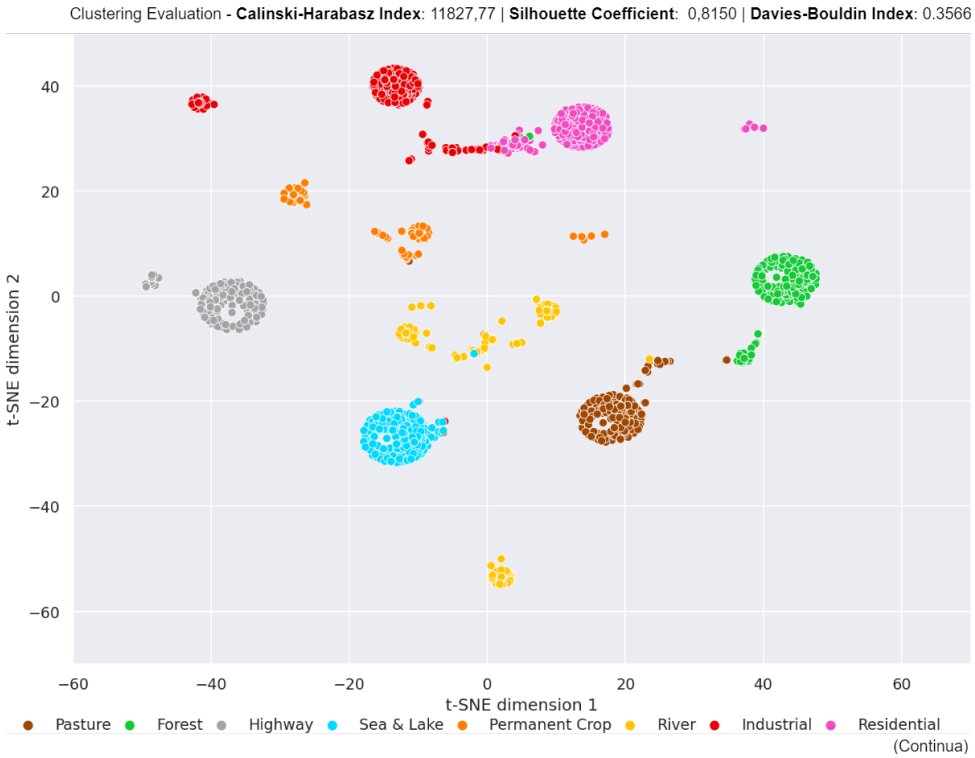
Tabela 4.6 - Desempenho global (mAP) do CBIR de imagens aéreas (AID) e satelitais RGB/Multiespectral (EuroSAT), considerando k imagens recuperadas pela menor distância de Hamming sendo $k = 20, 50, 100$.

	AID		**EuroSAT RGB**		**EuroSAT MS**	
Metrics	**InceptionV3**	**ResNET50**	**InceptionV3**	**ResNET50**	**InceptionV3**	**ResNET50**
mAP@20	99,1414	98,9876	92,1577	94,6250	99,8585	**99,8868**
mAP@50	99,1536	98,9871	91,8063	94,3257	99,8637	**99,8868**
mAP@100	99,1814	98,9734	91,5558	90,7167	99,8515	**99,8873**

Esses resultados permitem fazer algumas afirmações: i) Em geral, a arquitetura ResNet-50 superou a Inception-V3, demonstrando o potencial da aprendizagem residual para gerar informações semânticas também a partir de imagens de média resolução espacial; ii) A utilização de informação multiespectral (EuroSAT MS) de imagens de média resolução espacial contribuiu para melhorar o desempenho da tarefa de recuperação de imagens, superando os resultados alcançados com imagens de alta resolução espacial do conjunto de dados AID, provando que bandas extras são úteis para fornecer informação durante a extração de recursos em CNNs; iii) As informações multiespectrais, no caso do conjunto de dados EuroSAT, também foram importantes para resolver ambiguidades que certos tipos de LULC apresentavam, como entre Floresta, Pastagem e Cultura Permanente (padrão de cobertura similares), entre Rodovia e Rio e entre Área Industrial e Área Residencial (padrões geométricos similares).

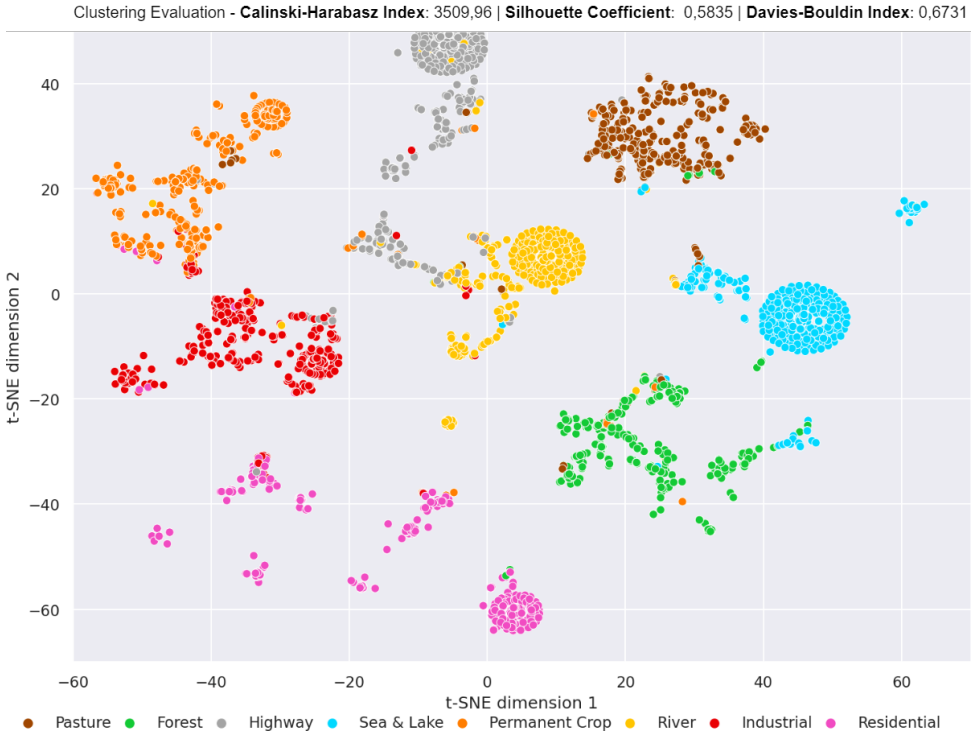
A análise das projeções bidimensionais dos espaços métricos produzidos pela rede MiLaN (Figura 4.3), confirma de forma qualitativa o seu potencial para produção de espaços métricos otimizados para a tarefa de busca e recuperação de imagens baseadas em conteúdo (CBIR) utilizando o processo de *Hashing*. Os conjuntos de códigos binários K-dimensionais com $K = 64$ correspondendo ao comprimento do vetor de atributos em bits, foram produzidos a partir das imagens de teste utilizando a MiLaN+ResNet-50.

Figura 4.3 - Projeção do espaço métrico criado pela rede MiLaN para os dados dos conjuntos AID e EuroSAT (RGB/MS).



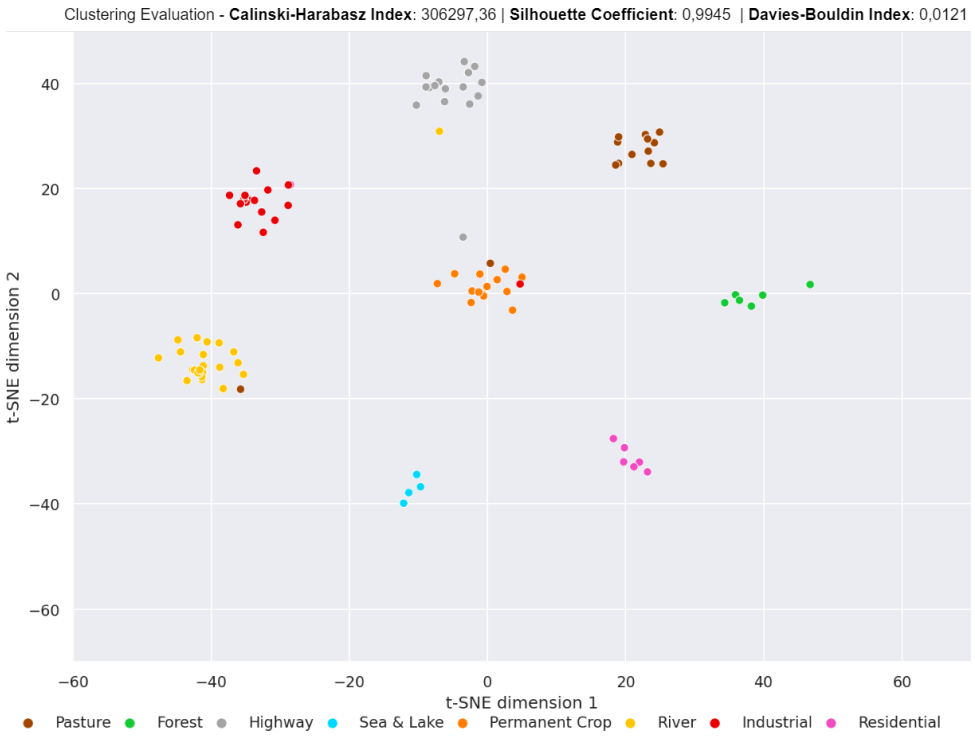
(Continua)

(a) t-SNE para AID



(b) t-SNE para EuroSAT

Figura 4.3 - Conclusão.



(c) t-SNE para EuroSAT (MS)

Projeção bidimensional com o método t-SNE do espaço com K dimensões representando o conjunto de *hash codes*, onde $K = 64$, produzidos com a rede MiLaN combinada com a ResNet-50 para as imagens de teste desses conjuntos.

Fonte: Próprio autor.

Na Figura 4.3(a) são representados os *hash codes* produzidos a partir das imagens do conjunto AID, nesse caso as amostras apresentam boa separação com proximidade natural entre as classes que possuem certa similaridade, como por exemplo, Áreas Industriais e Residenciais (construções) além de Floresta, Pastagem e Culturas Permanentes (vegetação natural). Enquanto a projeção do espaço métrico produzido a partir de dados RGB do conjunto EuroSAT (Figura 4.3(b)) demonstra uma separação mínima interclasses, indicando que estes dados não são suficientemente discriminativos para a realização de uma boa tarefa CBIR. Finalmente, a Figura 4.3(c) ilustra o melhor desempenho alcançado com o *framework*. Nela é representado o espaço métrico ("otimizado") produzido a partir de dados multiespectrais do conjunto EuroSAT tratados com a técnica CLAHE, no qual as classes dos diversos tipos de LULC possuem seus *clusters* bem definidos.

De forma complementar, foram realizadas avaliações quantitativas dos espaços métricos utilizando as seguintes métricas: *Calinski-Harabasz Index*, *Silhouette Coefficient* e *Davies-Bouldin Index*. Elas servem para indicar a qualidade do agrupamento de dados semelhantes que satisfaçam algum critério para definição de membros pertencentes à mesma classe através de alguma medida de similaridade (Figura 4.3). As métricas foram calculadas a partir do agrupamento produzido pelo algoritmo Kmeans[3], adotando 8 como o número de agrupamentos esperados a partir dos vetores de *hash codes* de cada conjunto de imagem.

O índice *Calinski-Harabasz* ou critério de razão de variância, mede a coesão (similaridade) intra-agrupamento com base nas distâncias dos pontos ao centroide de um agrupamento e a dissimilaridade baseada na distância dos centroides do agrupamento ao centroide global (CALINSKI; HARABASZ, 1974). Valores maiores indicam agrupamentos mais densos e bem separados, nesse caso, o desempenho do agrupamento produzido com os *hash codes* gerados pela MiLaN+ResNet-50 foi superior com dados EuroSAT MS (306297,36), seguido pelos conjuntos AID (11827,77) e EuroSAT (3509,96).

A medida dada pelo coeficiente *Silhouette* é definida para cada amostra como:

$$s = \frac{b - a}{max(a, b)}$$

onde a é a distância média entre a amostra e todos os outros pontos da mesma classe e b a distância média entre a amostra e todos os outros pontos do agrupamento mais próximo (ROUSSEEUW, 1987).

Valores típicos para esse coeficiente variam entre -1 (pior) e 1 (melhor), "e correspondem ao" desempenho do resultado do agrupamento, sendo que valores próximos a zero indicam sobreposição de agrupamentos. Esses foram os valores alcançados com os dados EuroSAT MS (0,9945), AID (0,8150) e EuroSAT (0,5835), ratificando os resultados obtidos com índice anterior.

Para concluir a avaliação dos espaços métricos, foi calculado o índice Davies-Bouldin, o qual é definido como uma medida da similaridade média entre agrupamentos. Esse índice é computado de forma simples e similar ao coeficiente *Silhouette*, sendo que a medida de similaridade é dada pela comparação da distância entre os agrupamen-

[3]Tanto as métricas quanto o algoritmo Kmeans utilizados nessa avaliação fazem parte do pacote Scikit-learn desenvolvido em Python. Documentação disponível em <https://scikit-learn.org/stable/modules/clustering.html#k-means>. Acesso em: 24 jan. 2024.

tos e o diâmetro dos mesmos (DAVIES; BOULDIN, 1979). Nesse contexto, quanto mais próximo de zero melhor é o resultado do agrupamento realizado. Novamente o desempenho foi superior com os dados EuroSAT MS (0,0121) com AID (0,3566) e EuroSAT (0,6731) na sequência.

O desempenho alcançado pelo *framework* proposto foi superior ao de outros métodos do estado da arte (KAPOOR et al., 2021), demonstrando que o uso combinado da aprendizagem residual, espaço métrico e informação multiespectral aprimorada pode ser potencialmente empregado como uma solução para aplicação de CBIR no contexto RSBD.

4.4 Uso de CBIR para identificação de uso e cobertura da terra no Cerrado

O Cerrado brasileiro tem sofrido grande pressão devido à expansão agrícola em seu território principalmente após a Moratória da Soja na Amazônia Legal. Por esse motivo, existem várias iniciativas que buscam mapear o uso da terra no Cerrado, por exemplo, TerraClass Cerrado (INPE, 2018) e MapBiomas (SOUZA et al., 2020), com objetivo de monitorar o desmatamento e permitir medidas de mitigação desse problema pelos órgãos responsáveis.

Dada a evidente importância dessa região, esse experimento teve o objetivo de testar o potencial do uso de CBIR para identificação de LULC em uma área do Cerrado no Oeste da Bahia na região MATOPIBA, correspondente ao *tile* 089097 do cubo de dados do Sentinel 2 disponível no catálogo do projeto BDC (Seção 3.2.1).

A Tabela 4.7 sumariza o desempenho do *framework* para identificação de LULC para os 11352 (128×128 pixels) e 45236 (64×64 pixels) *patches* de imagens de reflectância à superfície usando as bandas RGB e MS (12 bandas - exceto banda 10, absorção de vapor d'água).

Por causa da restrição quanto ao tamanho da imagem de entrada da Inception V3 optou-se por analisar os *patches* de menor tamanho somente com a ResNet-50 e de maior tamanho com a Inception V3.

Tabela 4.7 - Desempenho para identificação de LULC no Cerrado baseado em CBIR utilizando combinações da MiLaN+Inception V3*/ResNet-50, *patches* com 64×64/128×128 pixels e dados RGB/MS.

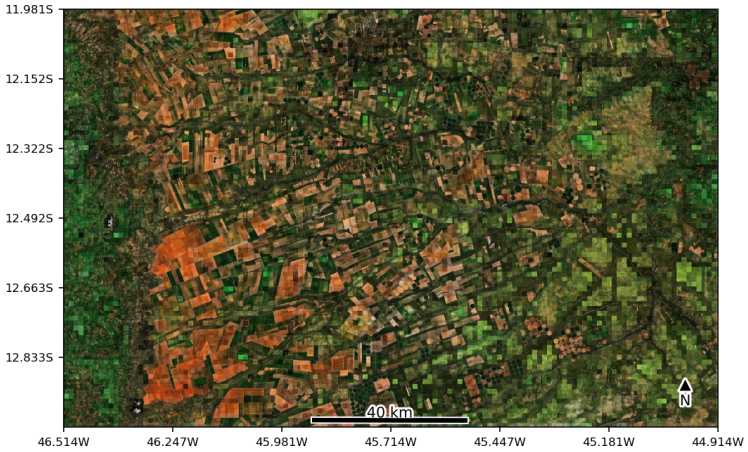
	Inception V3						ResNet-50					
	128×128 RGB			128×128 MS			64×64 RGB			64×64 MS		
LULC	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**
Annual Crop	1	1	1	**1**	**1**	**1**	0,9911	1	0,9955	0,9993	1	0,9997
Forest	0,9955	1	0,9977	**1**	**1**	**1**	0,9961	1	0,9980	1	1	1
Herbaceous Vegetation	**0,9522**	**1**	**0,9755**	0,9138	1	0,9550	0,8762	1	0,9340	0,9076	1	0,9516
Highway	**0,0244**	**1**	**0,0476**	0,0093	1	0,0183	0,0036	1	0,0071	0,0160	1	0,0316
Industrial	**1**	**1**	**1**	0,5374	1	0,6991	0,3140	1	0,4780	0,4563	1	0,6267
Pasture	0,9908	1	0,9954	**1**	**1**	**1**	0,8561	1	0,9225	0,8990	1	0,9468
Permanent Crop	1	1	1	**1**	**1**	**1**	0,9995	1	0,9998	0,9995	1	0,9997
Residential	**1**	**1**	**1**	0,9620	1	0,9806	0,9781	1	0,9889	0,9410	1	0,9696
River	**0,0526**	**1**	**0,1000**	0,0207	1	0,0405	0,0136	1	0,0269	0,0788	1	0,1461
Sea Lake	**0,0295**	**1**	**0,0573**	0,0243	1	0,0475	0,0064	1	0,0127	0,0082	1	0,0162

* Somente imagens com no mínimo 75×75 pixels são aceitas para entrada na rede Inception V3 (SZEGEDY et al., 2016).

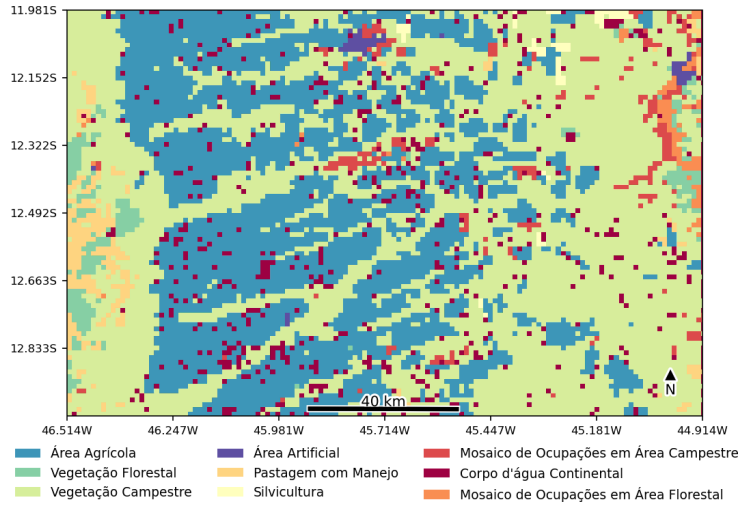
A partir dos resultados apresentados na Tabela 4.7 e ilustrados na Figura 4.3 é possível afirmar que o processo *labeling* baseado em CBIR através do *transfer learning* de dados do conjunto EuroSAT foi capaz de identificar nessa área do Cerrado imagens com os seguintes tipos de LULC: Áreas Industriais e Residenciais, Cultura Anual, Cultura Permanente, Floresta, Corpos Hídricos (Rios/Oceano e Lagos), Rodovias, Vegetação Herbácea e Pastagens.

De acordo com a análise visual do resultado de identificação de LULC realizado através da combinação MiLaN+Inception V3, foi possível verificar que os *patches* de 128×128 pixels são mais adequados para recuperação semântica da circunvizinhança, possibilitando inclusive a identificação de esparsas áreas industriais nessa região (Figura 4.4). Por outro lado, o uso de *patches* menores (64×64 pixels) representou a perda dessa informação, resultando na identificação incorreta de imagens com tipo de uso *Highway* associadas às áreas agricultáveis com vegetação campestre ou florestais como demonstrado na matriz de confusão (Figura 4.5), isso ocorre devido à presença de estradas de acesso as áreas de plantio (talhões) que têm sua geometria bem mais destacada nessas imagens de entrada, perdendo contexto da vizinhança.

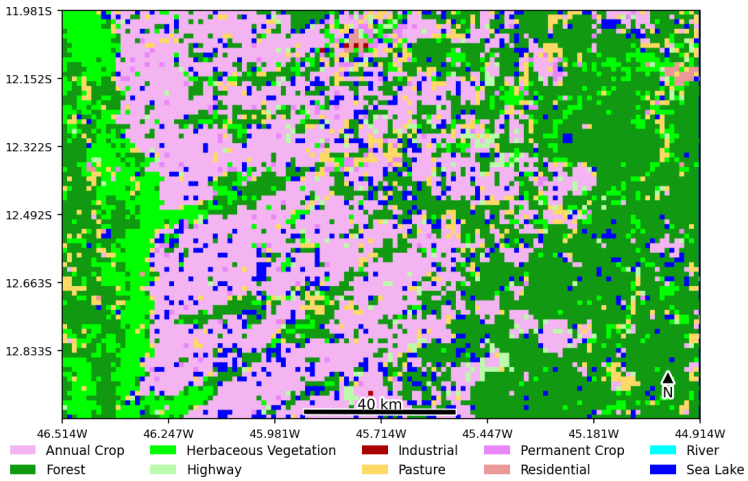
Figura 4.3 - LULC para o tile 089097 usando Inception V3 e *patches* RGB de 128×128 pixels do cubo de dados Sentinel identificada com base na máxima similaridade em relação a dados do conjunto EuroSAT.



(a) Composição RGB para os 11352 patches.



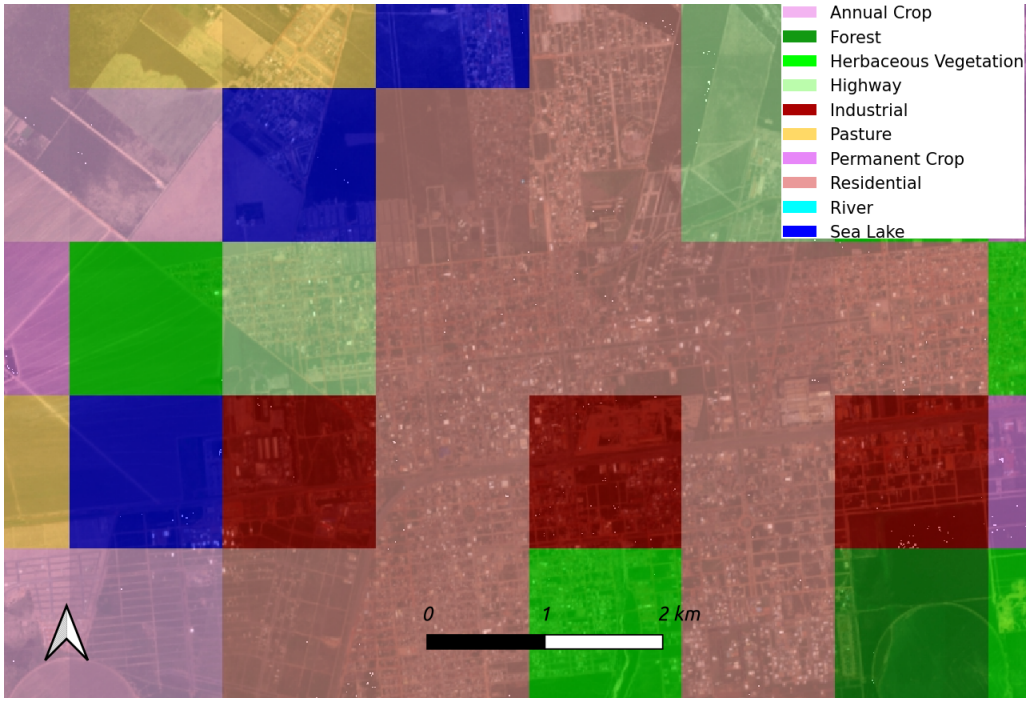
(b) LULC mapeado pelo IBGE (WLTS).



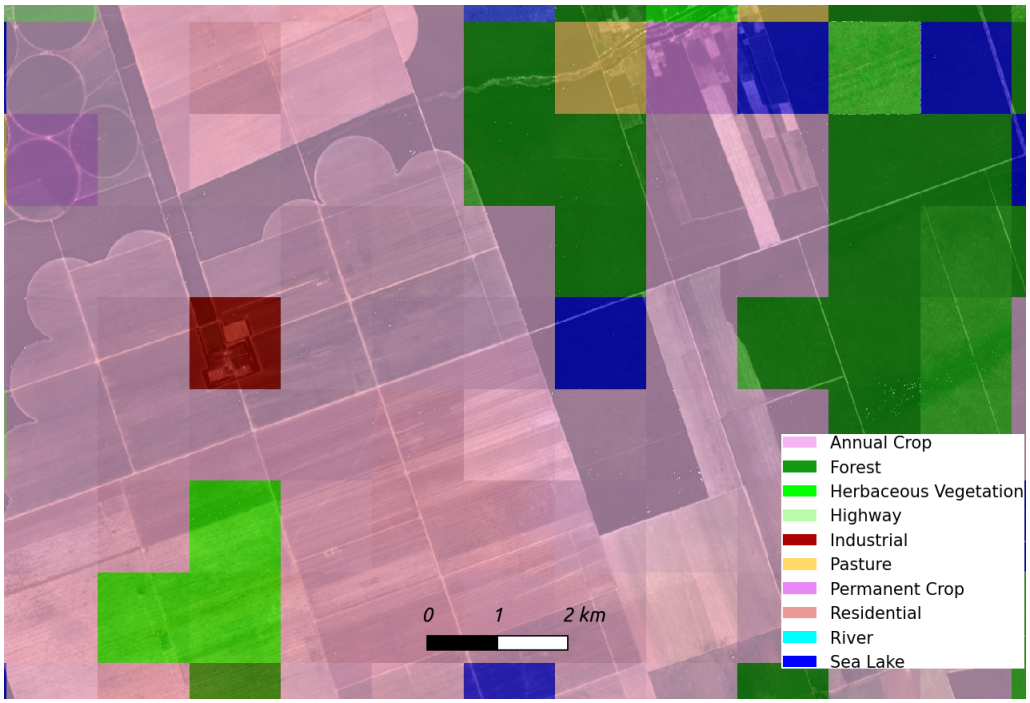
(c) LULC identificado através de CBIR.

Fonte: Próprio autor.

Figura 4.4 - Detalhe da identificação de áreas industriais no tile 089097 usando Inception V3 e *patches* RGB de 128×128 pixels do cubo de dados Sentinel.

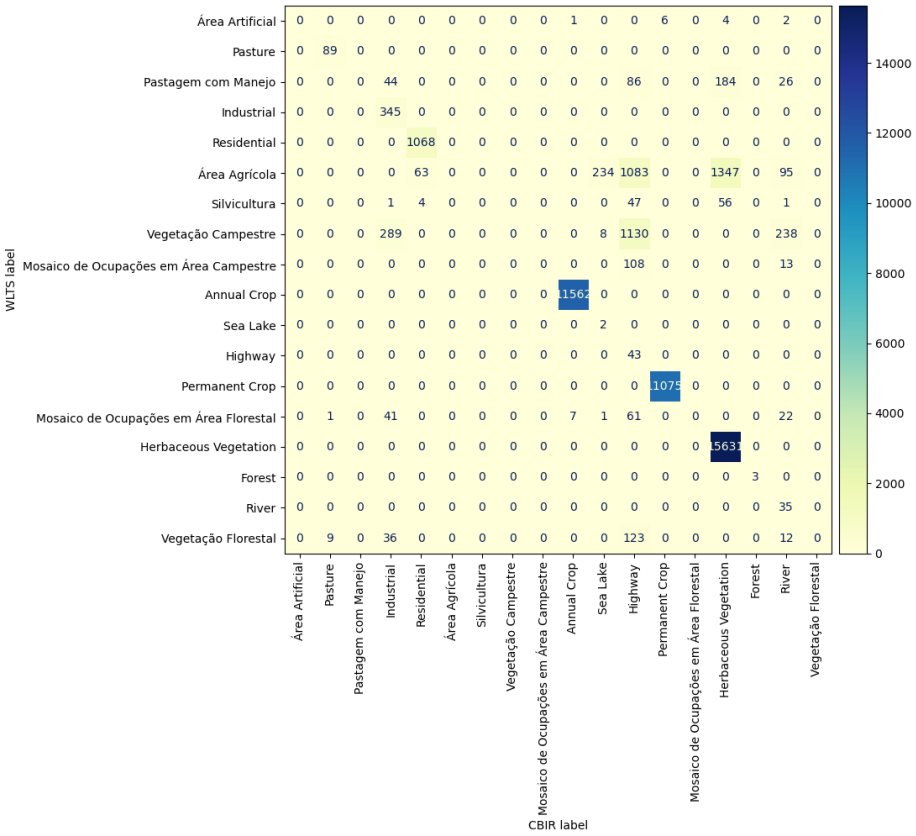


(a) Áreas Industriais identificadas na porção Norte-Central do tile 089097.



(b) Área Industrial (Agroindústria) identificada na porção Sul-Central do tile 089097.
Fonte: Próprio autor.

Figura 4.5 - Matriz de confusão resultado da identificação de LULC no Cerrado usando CBIR (MiLaN+ResNet-50) a partir de *patches* com 64×64 pixels multiespectrais.



Alguns dos usos e coberturas que aparecem na matriz de confusão são característicos exclusivamente do mapeamento feito pelo IBGE, e persistem após a análise de concordância não encontrar equivalência entre as classes do conjunto EuroSAT.

Fonte: Próprio autor.

De maneira geral, o desempenho da identificação de uso e cobertura da terra no Cerrado com CBIR foi superior ao obtido utilizando o método de classificação (Tabela 4.8), alcançando uma acurácia global[4] de 90,5% com a Inception V3+MiLaN. Entretanto, existem muitas áreas incorretamente associadas a corpos hídricos, a maior parte desses erros são de Áreas Agrícolas, Pastagem e Vegetação Herbácea. No conjunto EuroSAT, é comum a presença de corpos hídricos associados a esses tipos de usos, associação que foi incorretamente herdada nessa abordagem.

[4] A função *accuracy_score* calcula o percentual de previsões corretas, definido como a somatória das previsões corretas sobre o número total de amostras. Disponível em <https://scikit-learn.org/stable/modules/model_evaluation.html#accuracy-score>. Acesso em: 27 jan. 2024.

Tabela 4.8 - Comparação do desempenho para identificação de LULC no Cerrado baseado em CBIR e classificação.

	Classification			CBIR			Precision
LULC	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**	**Improvements/Worsens (%)**
Annual Crop	0,9993	1	0,9997	**0,9993**	**1**	**0,9997**	-0,0013
Forest	1	1	1	**1**	**1**	**1**	0
Herbaceous Vegetation	0,8984	1	0,9465	**0,9076**	**1**	**0,9516**	1,0297
Highway	0,0133	1	0,0263	**0,0160**	**1**	**0,0316**	20,4467
Industrial	0,3913	1	0,5625	**0,4563**	**1**	**0,6267**	16,6108
Pasture	0,8718	1	0,9315	**0,8990**	**1**	**0,9468**	3,1194
Permanent Crop	**0,9996**	1	**0,9998**	0,9994	**1**	0,9997	-0,0167
Residential	0,9387	1	0,9684	**0,9410**	**1**	**0,9696**	0,2394
River	0,0653	1	0,1226	**0,0788**	**1**	**0,1461**	20,7430
Sea Lake	0,0070	1	0,0139	**0,0082**	**1**	**0,0162**	16,7309

5 CONCLUSÕES

Neste trabalho foi analisada a influência da informação multiespectral para a melhoria do processo de CBIR de imagens satelitais de média resolução espacial (10 m), explorando o processo de *Hashing* através de aprendizagem profunda. Além disso, foi proposta uma aplicação para identificação do uso e cobertura da terra no Cerrado, aproveitando o potencial do *framework* desenvolvido nesta tese e o conceito de adaptação de domínio, quando o conhecimento gerado para uma tarefa pode ser adaptado para outra (SARKAR; BALI, 2022).

As análises aqui realizadas com imagens de alta e média resolução espacial (AID/EuroSAT), indicam que o uso combinado de modelos de *Deep Learning* (DL) para extração de características das imagens propiciam informações semânticas ideais para a criação de um espaço métrico otimizado para tarefa CBIR.

Nesse contexto, algumas arquiteturas e opções de inicialização de pesos das redes neurais foram testadas, apontando a ResNet-50 pré-treinada com imagens do conjunto ImageNet como módulo ideal para extração automática de características de imagens (*backbone*), reduzindo a complexidade e diminuindo o problema comumente conhecido como "lacuna semântica"[1] desses vetores de atributos. A melhoria da representação semântica das imagens indicada pelo uso da ResNet-50 é evidenciada pelo ganho de desempenho medido tanto para classificação de imagens, quanto para aplicação em CBIR.

Esta tese apresentou uma série etapas descritas como um *framework* para evolução do processo de CBIR baseado no método de *Hashing* e construção de um espaço métrico utilizando a rede neural MiLaN. Como demonstrado através dos resultados e análises, todas etapas contribuíram para melhoria da recuperação de imagens de satélite baseada no conteúdo. Entretanto, a etapa de *fine-tuning* com imagens de observação da Terra e o uso da informação multiespectral corrigida (CLAHE) foram determinantes para superar os desafios impostos pela resolução espacial das imagens do conjunto EuroSAT, permitindo a discriminação entre padrões de cobertura e geometria similares nas imagens.

O avanço em relação a abordagem original da rede MiLaN, permitiu a busca e recuperação de imagens satelitais de média resolução espacial com desempenho superior

[1] A conhecida questão da lacuna semântica representa o problema associado a uma representação limitada de imagens por recursos de baixo nível capturados por máquinas a partir de pixels de imagem em comparação com conceitos semânticos de alto nível reconhecidos por humanos (KAPOOR et al., 2021).

ao alcançado com imagens de alta resolução espacial do conjunto AID. Contribuindo para o desenvolvimento de uma aplicação potencialmente inovadora para identificação do uso e cobertura da terra em uma área do Cerrado brasileiro.

Dentre as diversas contribuições que este trabalho proporcionou relacionadas ao uso de métodos de DL para recuperação de imagens de satélite, cabe destacar as seguintes realizações:

a) Revisão fundamentada da literatura, delimitando o ciclo completo de um sistema baseado em CBIR, desde a geração de atributos do tipo *hand-crafted* dependentes de um especialista a métodos automáticos baseados em aprendizagem profunda para criação de um espaço métrico através do processo de *Hashing*;

b) Testes e análises de arquiteturas de DL para identificação de um *backbone* adequado para extração de informações semânticas ideais para o processo de CBIR;

c) Adoção da informação multiespectral com correção de contraste (CLAHE) para melhoria do processo de CBIR através da criação de um espaço métrico otimizado com a rede neural MiLaN, alcançando resultados superiores aos do processo de CBIR com imagens de alta resolução espacial;

d) Identificação de uso e cobertura da terra para uma área do Cerrado baseada no *framework* proposto para CBIR com dados multiespectrais e *Hashing* usando DL.

5.1 Trabalhos futuros

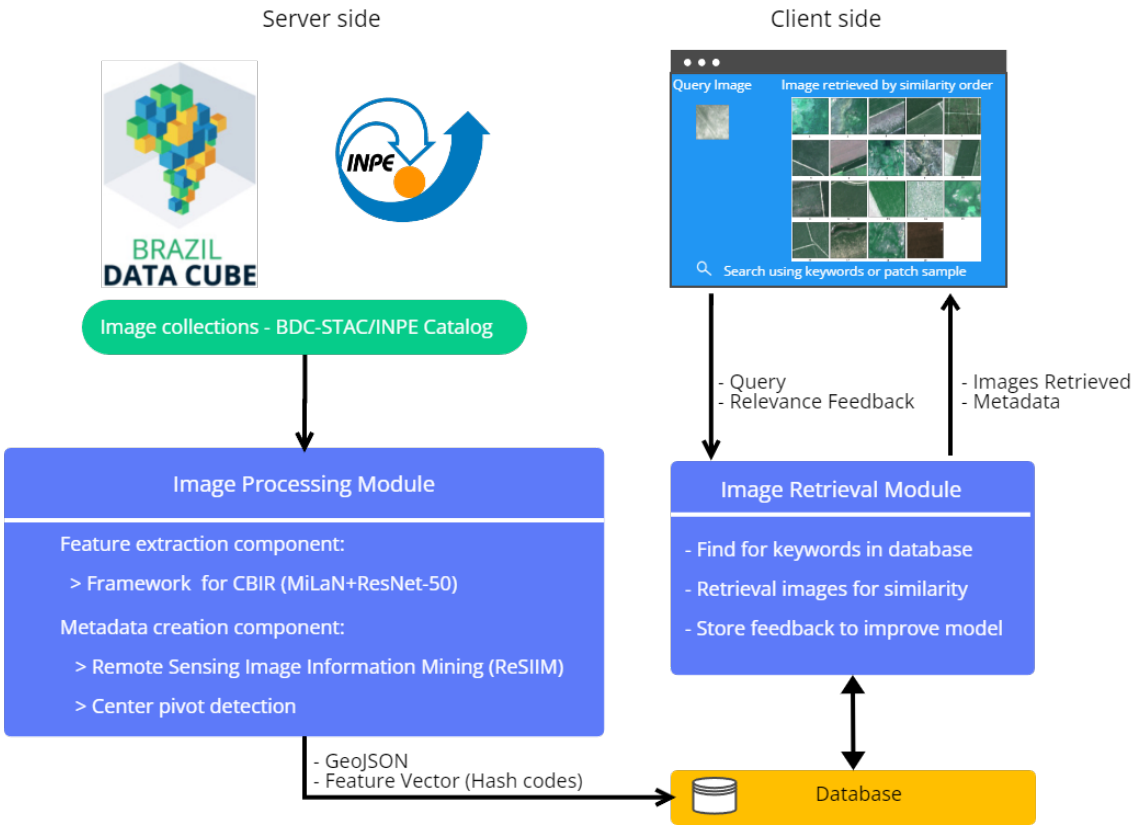
5.1.1 *Improved Metadata from Remote Sensing Images* (IMRSI)

A ideia principal desta pesquisa de doutorado foi a busca por métodos que visem facilitar à recuperação de imagens de interesse científico, por exemplo, para monitoramento do uso da terra no Cerrado. Essa busca teve como objetivos principais propor soluções para lidar com o grande volume de dados na era do *Remote Sensing Big Data* e limitações impostas pela recuperação de imagens em catálogos oficiais baseada em metadados comuns como: data de aquisição, sensor, taxa de cobertura por nuvens, etc.

Nesse sentido, com o conhecimento adquirido foi idealizada uma proposta para o desenvolvimento de uma solução que combina métodos de ML para extração de características de imagens e identificação de alvos de maneira a permitir uma recuperação eficiente de imagens em grandes conjuntos de dados baseada em conteúdo.

O *framework* de metadados aprimorados de imagens de sensoriamento remoto (*Improved Metadata from Remote Sensing Images* - IMRSI) visa aproveitar a infraestrutura criada no contexto do projeto BDC para disponibilização de imagens via catálogos STAC[2] e também a infraestrutura utilizada pelo catálogo de imagens do INPE.

Figura 5.1 - Visão geral do *framework Improved Metadata from Remote Sensing Images* (IMRSI) para busca e recuperação de imagens baseadas em conteúdo.



Fonte: Próprio autor.

[2] *SpatioTemporal Asset Catalog* (STAC) é uma especificação utilizada para disponibilização de dados espaço-temporais de observação da Terra, que visa facilitar a descrição de vários tipos de dados permitindo a recuperação dos mesmos de maneira unificada e simples. Disponível em: <https://stacspec.org>. Acesso em: 12 novembro 2020.

A Figura 5.1 apresenta uma visão geral da construção do IMRSI, a solução possui dois aspectos principais de funcionamento:

1. Módulo de processamento de imagem — responsável por acessar e processar o conjunto de imagens providas (BDC/STAC) ou (INPE/catálogo);
2. Módulo de recuperação de imagem — responsável por receber as requisições dos usuários e também por fazer a retroalimentação de relevância (*Relevance Feedback*) dos resultados das consultas geradas.

5.1.1.1 Módulo de processamento de imagens

O uso de arquiteturas de aprendizagem profunda para representação semântica de imagens apresenta melhor desempenho quando empregado o método supervisionado (LI et al., 2018). Esse tipo de método requer o treinamento com um grande número de imagens rotuladas para permitir a modelagem eficaz do conteúdo de imagens de SR.

Essa é uma questão bem desafiadora, pois no escopo do SR os conjuntos disponíveis normalmente contêm número reduzido de imagens rotuladas, geralmente com poucas bandas espectrais como é o caso do UC Merced Land Use (UCMD) (YANG; NEWSAM, 2010). Mesmo quando o número de imagens é consideravelmente maior, como no caso dos conjuntos SAT4 (500×10^3) e SAT6 (405×10^3) (BASU et al., 2015), existem perdas em relação à diversidade de classes apresentadas. Além disso, a maioria desses conjuntos derivam de imagens geradas por sensores aerotransportados, os quais apresentam condições de imageamento bem distintas das condições verificadas por sensores orbitais, com destaque para questão do espalhamento atmosférico.

Outro fator limitante, é o aprendizado realizado com conjuntos de imagens de SR que apresentam rótulos de uma única classe, porém normalmente as imagens de SR contêm vários tipos de cobertura de solo, dessa maneira uma única imagem pode ser associada simultaneamente a diferentes rótulos (*multi-labels*).

Com objetivo de superar esses problemas, planeja-se adaptar a rede neural MiLaN para construção de espaço métrico baseado na aprendizagem semântica de imagens *multi-labels* do conjunto BigEarthNet-MM[3] proposto por Sumbul et al. (2021).

[3]BigEarthNet é considerado o primeiro conjunto em grande escala de imagens *multi-labels* (43 classes), multiespectral (12 bandas do sensor MSI) formado por mais de 590×10^3 imagens satelitais. Originalmente proposto por Sumbul et al. (2019), foi criado a partir da definição de *patches* de

O BigEarthNet será fundamental para o treinamento e testes das arquiteturas de DL escolhidas para a criação de espaço métrico e representação de imagens em códigos *hash*. Tanto a classificação quanto o processo de CBIR realizados com base no BigEarthNet criarão o arcabouço necessário para rotular as imagens disponíveis no BDC, bem como para a identificação de classes em amostras utilizadas para recuperação de imagens semelhantes.

Componente para extração de atributos de imagem — será baseada no *framework* proposto neste trabalho para o CBIR de imagens satelitais de média resolução baseada em arquiteturas de DL (MiLaN+ResNet-50) para extração de atributos de imagens convertidos para *Hash codes*, formando assim um espaço métrico ideal para recuperação de imagens por similaridade pelo cálculo da distância de *Hamming* a partir de amostras de imagens utilizadas para consulta pelo usuário (*Client side*). A proposta é criar ou adaptar uma extensão em C customizada para cálculo de similaridade diretamente no banco de dados (PostresSQL).

Como resultado tem-se *patches* de imagens previamente rotuladas para buscas futuras (palavras chaves), assim como vetores de atributos para recuperação de imagens através do cálculo de similaridade entre imagem de consulta e os vetores armazenados no banco de dados.

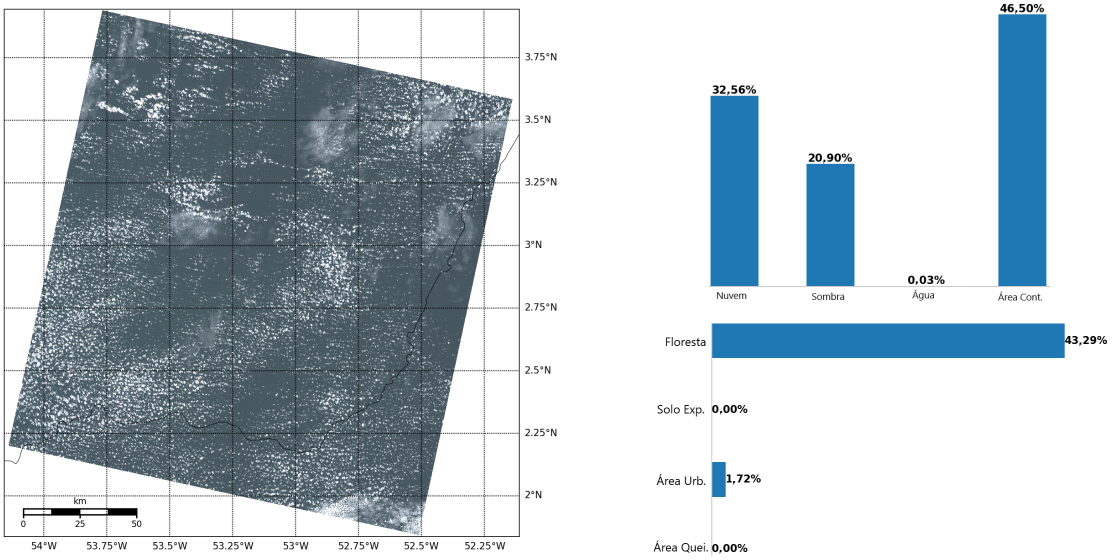
Componente para geração de metadados adicionais — combinará métodos e algoritmos especializados para a geração de metadados adicionais (*improved metadata*) que subsidiem a identificação de imagens a partir de informações que levem em conta o seu conteúdo, dessa forma pretende-se nessa componente implementar duas abordagens. A primeira baseada no protótipo *Remote Sensing Image Information Mining* (ReSIIM) descrito por Pletsch e Körting (2018), que utiliza aritmética de bandas e índices espectrais para descrever o conteúdo de imagens. A segunda será uma versão híbrida de abordagens desenvolvidas durante o doutorado para detecção de pivôs em áreas do território brasileiro (RODRIGUES et al., 2020a; RODRIGUES et al., 2020b), baseadas na caracterização da resposta máxima da vegetação, uso de séries temporais para identificação de ciclos de culturas e a classificação de alvos com base no algoritmo *Random Forests* (RF) para dados desbalanceados.

imagens Sentinel 2 com 1,2×1,2 km, adquiridas entre junho/2017 e julho/2018 e rotuladas com base no *CORINE Land Cover database* (CLC) do ano de 2018. O CLC é um inventário realizado pelos centros nacionais de referência sobre cobertura do solo (Eionet) sob a coordenação da Agência Europeia de Meio Ambiente, para a identificação e avaliação de classes de cobertura do solo, por meio de textura, padrão e densidade de alvos em imagens de SR.

O ReSIIM foi originalmente proposto para usar diferentes combinações de bandas dos satélites Landsat 5 e 8 para obtenção de índices espectrais que permitam a identificação de pixels característicos de floresta, solo exposto, área urbana e área queimada. A busca de imagens baseada na abordagem do ReSIIM, demonstrou seu potencial ao recuperar com sucesso a partir de um grande conjunto de dados, imagens úteis para o mapeamento de cicatrizes de queimadas na região Amazônica (PLETSCH; KÖRTING, 2018).

A Figura 5.2, apresenta um exemplo da identificação de alvos feita através do ReSIIM. Os metadados resultantes dessa identificação, sintetizam o conteúdo da cena através do percentual de cada classe, essas informações são exportadas para um arquivo GeoJSON e depois armazenadas no banco de dados.

Figura 5.2 - Alvos identificados pelo ReSIIM em uma cena Landsat 8 órbita/ponto 227/058 de 18/07/2017.



(a) Imagem RGB cor verdadeira. (b) Percentual de cada classe alvo.

Fonte: Próprio autor.

Detecção de pivôs — para viabilizar essa abordagem, serão realizados estudos e testes para adaptação dos métodos previamente desenvolvidos para identificação de pivôs baseados no uso combinado de índices de vegetação, séries temporais (RODRIGUES et al., 2020a) e RF (RODRIGUES et al., 2020b).

Inicialmente será realizada a identificação da máxima reposta vegetativa através da composição temporal de imagens de índice de vegetação com mínimo de cobertura de nuvens possível (*Greenest Pixel Composition*) (MUDELE; GAMBA, 2019). Essa imagem de máxima reposta vegetativa facilita a delimitação via detector de bordas das áreas com vegetação fotossinteticamente ativas, inclusas aí as áreas circulares de culturas irrigadas por pivôs identificadas pelo método *Circular Hough Transform* (CHT)[4].

Adicionalmente, será utilizada a técnica *Time-Weighted Dynamic Time Warping* (TWDTW) (MAUS et al., 2016) para classificação de séries temporais de índices de vegetação extraídas de áreas identificadas pelo CHT. O TWDTW permite identificar padrões da dinâmica da vegetação através de séries temporais para classificação de uso e cobertura da terra. Essa classificação tem o objetivo de filtrar os casos de falso positivo para áreas não agrícolas.

Esses são alguns exemplos de metadados produzidos por essa componente que poderão ser incluídos no banco de dados: localização geográfica de pivôs de irrigação, percentual de nuvens, sombra de nuvens, corpos hídricos, solo exposto, área urbana, área queimada, tipo LULC baseada na similaridade com imagens do conjunto EuroSAT.

5.1.1.2 Módulo de recuperação de imagens

Esse módulo irá viabilizar a consulta e recuperação de imagens, a partir de duas entradas possíveis: (i) pesquisa por palavras-chave em um dicionário conhecido, por exemplo, consulta de imagens com “pivôs”; (ii) pesquisa através de uma imagem de amostra.

A ideia com IMRSI é apresentar para o usuário uma interface que permita fazer a busca por múltiplos parâmetros de maneira combinada, para a busca de imagens de interesse científico. Um exemplo prático seria a busca por imagens que contemplem um determinado bioma e que possuam pivôs ao longo de uma janela de tempo. Essa consulta poderia facilmente ser resolvida através de uma busca espaço-temporal no banco de metadados, retornando à referência das imagens relacionadas diretamente

[4]O CHT é uma técnica de extração de características derivada da ideia de espaço de parâmetros ou *Hough Space* (HS) originalmente definido pela representação paramétrica usada para descrever linhas no plano da imagem usando *Hough Transform* (HT) (DUDA; HART, 1972). Ela é amplamente utilizada no processamento digital de imagens para detecção de círculos em imagens de baixa qualidade, devido à sua robustez na presença de ruído, oclusão e iluminação variável (YUEN et al., 1990; RIZON et al., 2005; DEMBELE, 2010).

da infraestrutura do BDC. Esse tipo de aplicação, poderia indicar, por exemplo, o surgimento de uma nova fronteira agrícola ou a consolidação da mecanização em uma dada área do país.

Embora a consulta por múltiplos parâmetros combinados a palavras-chaves represente um avanço em relação à busca baseada em metadados comuns, essa consulta está limitada a um conjunto de critérios predeterminados. Com objetivo de estender a funcionalidade do *framework*, será implementada também a busca e recuperação de imagens baseada no cálculo de similaridade usando a distância de *Hamming* entre o vetor de atributo (*Hash codes*) de uma amostra de imagem e vetor de atributos das imagens dos catálogos armazenas no banco (espaço métrico).

Ainda que arquiteturas profundas como a combinação das redes neurais MiLaN e ResNet-50 permitam o aprendizado semântico das imagens através de transformações não lineares, a representação simplificada do vetor de atributos conduz ao problema da lacuna semântica. Esse problema está relacionado à lacuna entre a percepção humana e a descrição da imagem com base em características, uma vez que o conteúdo da imagem é muito subjetivo e difícil de descrever analiticamente. O tratamento desse problema se dá por meio da Retroalimentação de Relevância (RR).

A forma trivial de implementar a RR é através de um processo interativo no qual o usuário faz uma busca e recebe como resposta um conjunto de imagens, a partir desse conjunto ele indica quais são as imagens relevantes em relação a sua pesquisa. Esse processo define 1 ciclo, o qual permite identificar quais características descritivas das imagens possuem maior peso e devem ser levadas em conta para o refinamento da pesquisa e oferta ao usuário de um novo conjunto de imagens possivelmente relevantes (SILVA et al., 2009).

O processo de RR é muito efetivo para melhoria do CBIR, especialmente quando apoiado por algum método que consiga aprender a vontade do usuário e evoluir esse aprendizado, como é o caso da RR baseada em algoritmos genéticos (BARCELOS et al., 2009). Todavia, após algumas repetições desse processo, o usuário tende a se cansar e passar dar um retorno sem relevância prejudicando o resultado da pesquisa. Alguns autores afirmam que o limite seria de no máximo seis repetições (MAJI; BOSE, 2020).

Alternativamente, é possível explorar abordagens que realizem a RR de maneira automática ou com mínimo de interação do usuário. Esse tipo de abordagem é conhecido como pseudo RR ou RR às cegas. Diversas técnicas podem ser empregadas, por exemplo, SVM, Árvores de decisão, Agrupamento, RF, CNNs entre outras (PUTZU et al., 2020).

Dessa maneira, serão testadas várias abordagens automáticas de RR para compará-las com as informações de relevância dadas por usuários do IMRSI, a fim de identificar a melhor no escopo de imagens de SR.

O *framework* IMRSI apresenta-se como uma ferramenta promissora para operacionalização da busca e recuperação de imagens baseadas em conteúdo para grandes conjuntos de dados, por exemplo, cubos de dados satelitais para o Brasil fornecidos pelo projeto BDC.

REFERÊNCIAS BIBLIOGRÁFICAS

ABADI, M.; AGARWAL, A.; BARHAM, P.; BREVDO, E.; CHEN, Z.; CITRO, C.; CORRADO, G. S.; DAVIS, A.; DEAN, J.; DEVIN, M.; GHEMAWAT, S.; GOODFELLOW, I.; HARP, A.; IRVING, G.; ISARD, M.; JIA, Y.; JOZEFOWICZ, R.; KAISER, L.; KUDLUR, M.; LEVENBERG, J.; MANé, D.; MONGA, R.; MOORE, S.; MURRAY, D.; OLAH, C.; SCHUSTER, M.; SHLENS, J.; STEINER, B.; SUTSKEVER, I.; TALWAR, K.; TUCKER, P.; VANHOUCKE, V.; VASUDEVAN, V.; VIéGAS, F.; VINYALS, O.; WARDEN, P.; WATTENBERG, M.; WICKE, M.; YU, Y.; ZHENG, X. **TensorFlow: large-scale machine learning on heterogeneous systems**. 2015. Disponível em: <<https://www.tensorflow.org/>>. 40

ANDONI, A.; INDYK, P. Nearest neighbors in high-dimensional spaces. In: GOODMAN, J.; O'ROURKE, J.; TÓTH, C. D. (Ed.). **Handbook of discrete and computational geometry**. 3. ed. Boca Raton, FL: CRC Press, 2017. cap. 43, p. 1135–1155. ISBN 9781498711425. Disponível em: <<https://www.csun.edu/~ctoth/Handbook/chap43.pdf>>. 12

APTOULA, E. Remote sensing image retrieval with global morphological texture descriptors. **IEEE Transactions on Geoscience and Remote Sensing**, v. 52, n. 5, p. 3023–3034, May 2014. ISSN 1558-0644. 1, 9, 10

BARCELOS, E. Z. et al. **Recuperaçao de imagens por conteúdo: uma abordagem multidimensional de modelagem de similaridade e realimentaçao de relevância**. 175 p. Dissertação (Mestrado em Engenharias) — Faculdade de Engenharia Elétrica, Universidade Federal do Uberlândia, Uberlândia, 2009. Disponível em: <<https://repositorio.ufu.br/handle/123456789/14416>>. 82

BASU, S.; GANGULY, S.; MUKHOPADHYAY, S.; DIBIANO, R.; KARKI, M.; NEMANI, R. DeepSat. In: SIGSPATIAL INTERNATIONAL CONFERENCE ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS, 23., 2015. **Proceedings...** New York: ACM Press, 2015. p. 1–10. 78

BELWARD, A. S.; SKØIEN, J. O. Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 103, p. 115–128, 2015. ISSN 0924-2716. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0924271614000720>>. 1

BRASIL. MINISTÉRIO DA ECONOMIA. **II Plano Nacional de Desenvolvimento**: 1975 - 1979. Brasília, DF: Presidência da República, 1974. 155 p. Disponível em: <<https://bibliotecadigital.economia.gov.br/handle/777/24>>. Acesso em: 10 maio 2021. 36

BRAZIL DATA CUBE (BDC). **Projection used by BDC**. 2002. Disponível em: <<https://brazil-data-cube.github.io/products/specifications/bdc-grid.html>>. Acesso em: 10 dez. 2022. 37

BROWNLEE, J. **Random oversampling and undersampling for imbalanced classification**. 2021. Disponível em: <<https://machinelearningmastery.com/random-oversampling-and-undersampling-for-imbalanced-classification>>. Acesso em: 10 jan. 2021. 100

BURGES, C. J. A tutorial on support vector machines for pattern recognition. **Data Mining and Knowledge Discovery**, v. 2, n. 2, p. 121–167, 1998. 12

BURNETT, C. M. L. **Hamming distance cube for 3-bit binary numbers**. dezembro 2006. Disponível em: <<https://pt.wikipedia.org/wiki/Ficheiro:Hamming_distance_3_bit_binary_example.svg>>. Acesso em: 31 dez. 2006. 15

CALINSKI, T.; HARABASZ, J. A dendrite method for cluster analysis. **Communications in Statistics - Theory and Methods**, v. 3, n. 1, p. 1–27, 1974. ISSN 0361-0926. Disponível em: <<http://dx.doi.org/10.1080/03610927408827101>>. 67

CAMARA, G.; ASSIS, L. F.; RIBEIRO, G.; FERREIRA, K. R.; LLAPA, E.; VINHAS, L. Big Earth observation data analytics: matching requirements to system architectures. In: SIGSPATIAL INTERNATIONAL WORKSHOP ON ANALYTICS FOR BIG GEOSPATIAL DATA, 5., 2016. **Proceedings...** California: ACM Press, 2016. p. 1–6. 1

CHI, M.; PLAZA, A.; BENEDIKTSSON, J. A.; SUN, Z.; SHEN, J.; ZHU, Y. Big data for remote sensing: challenges and opportunities. **Proceedings of the IEEE**, v. 104, n. 11, p. 2207–2219, 2016. 1, 2

CHOLLET, F. **Deep Learning with Python**. Manning Publications Company, 2017. ISBN 9781617294433. Disponível em: <<https://books.google.com.br/books?id=Yo3CAQAACAAJ>>. 49

DATA SCIENCE ACADEMY. **Deep learning book**. Disponível em: <<https://www.deeplearningbook.com.br>>. Acesso em: 01 nov. 2023. 27, 45

DAVIES, D. L.; BOULDIN, D. W. A cluster separation measure. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, PAMI-1, n. 2, p. 224–227, abr. 1979. ISSN 2160-9292. Disponível em: <<http://dx.doi.org/10.1109/TPAMI.1979.4766909>>. 68

DEMBELE, F. **Object detection using circular hough transform: introduction to the hough transform**. Michigan: Michigan State University, 2010. Disponível em: <<https://www.egr.msu.edu/classes/ece480/capstone/fall10/group03/ece480_dt3_application_note_dembelef.pdf>>. 81

DEMIR, B.; BRUZZONE, L. A novel active learning method in relevance feedback for content-based remote sensing image retrieval. **IEEE Transactions on**

Geoscience and Remote Sensing, v. 53, n. 5, p. 2323–2334, May 2015. ISSN 1558-0644. 2, 10

_____. Hashing-based scalable remote sensing image search and retrieval in large archives. **IEEE Transactions on Geoscience and Remote Sensing**, v. 54, n. 2, p. 892–904, feb 2016. ISSN 0196-2892. Disponível em: <<http://ieeexplore.ieee.org/document/7270304/>>. 1, 12, 15, 16, 17, 18, 31, 33

DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: a large-scale hierarchical image database. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION. **Proceedings...** Los Alamitos, CA, USA: IEEE Computer Society, 2009. p. 248–255. Disponível em: <<https://doi.ieeecomputersociety.org/10.1109/CVPR.2009.5206848>>. 28

DUDA, R. O.; HART, P. E. Use of the Hough transformation to detect lines and curves in pictures. **Communications of the ACM**, v. 15, n. 1, jan 1972. ISSN 00010782. Disponível em: <<doi.org/10.1145/361237.361242>>. 81

ELHARROUSS, O.; AKBARI, Y.; ALMAADEED, N.; AL-MAADEED, S. **Backbones-review: feature extraction networks for deep learning and deep reinforcement learning approaches**. arXiv:2206.08016, 2022. 45

EMPRESA BRASILEIRA DE PESQUISA AGROPECUáRIA (EMBRAPA). **MATOPIBA**. 2019. Disponível em: <<https://www.embrapa.br/tema-matopiba>>. Acesso em: 26 ago. 2019. 36

FERREIRA, K. R.; QUEIROZ, G. R.; VINHAS, L.; MARUJO, R. F. B.; SIMOES, R. E. O.; PICOLI, M. C. A.; CAMARA, G.; CARTAXO, R.; GOMES, V. C. F.; SANTOS, L. A.; SANCHEZ, A. H.; ARCANJO, J. S.; FRONZA, J. G.; NORONHA, C. A.; COSTA, R. W.; ZAGLIA, M. C.; ZIOTI, F.; KORTING, T. S.; SOARES, A. R.; CHAVES, M. E. D.; FONSECA, L. M. G. Earth observation data cubes for Brazil: requirements, methodology and products. **Remote Sensing**, v. 12, n. 24, 2020. ISSN 2072-4292. Disponível em: <<https://www.mdpi.com/2072-4292/12/24/4033>>. 7, 37

FLORES, P. J. H. **Compact features for mobile visual search**. [s.n.], 2015. 74 p. Dissertação (Mestrado em Engenharia de Computação) - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação, Campinas, SP. Disponível em: <<http://repositorio.unicamp.br/jspui/handle/REPOSIP/258941>>. 15

FRIEDMAN, J. H.; BENTLEY, J. L.; FINKEL, R. A. An algorithm for finding best matches in logarithmic expected time. **ACM Transactions on Mathematical Software (TOMS)**, v. 3, n. 3, p. 209–226, sep 1977. ISSN 15577295. Disponível em: <<http://dl.acm.org/doi/10.1145/355744.355745>>. 12

FUKUSHIMA, K. Neocognitron: a model for visual pattern recognition. In: ARBIB, M. A. (Ed.). **Handbook of brain theory and neural networks**. [S.l.]: The MIT Press, 1995. p. 613–617. 23

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. [S.l.]: MIT Press, 2016. ISSN 1548-7091. ISBN 9780521835688. 13

GORISSE, D.; CORD, M.; PRECIOSO, F. Locality-sensitive hashing for chi2 distance. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 34, n. 2, p. 402–409, 2012. ISSN 01628828. 11, 18

GRAVES, A.; MOHAMED, A.; HINTON, G. Speech recognition with deep recurrent neural networks. In: IEEE INTERNATIONAL CONFERENCE ON ACOUSTIC, SPEECH AND SIGNAL PROCESSING, 38., 2013, Vancouver, Canada. **Proceedings...** Canada, 2013. p. 6645–6649. 20

HAMMING, R. W. Error detecting and error correcting codes. **The Bell System Technical Journal**, v. 29, n. 2, p. 147–160, 1950. 14

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. **CoRR**, abs/1512.03385, 2015. Disponível em: <<http://arxiv.org/abs/1512.03385>>. 19

_____. _____. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2016, Las Vegas, USA. **Proceedings...** [S.l.]: IEEE, 2016. p. 770–778. 45

HELBER, P.; BISCHKE, B.; DENGEL, A.; BORTH, D. Eurosat: a novel dataset and deep learning benchmark for land use and land cover classification. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 12, n. 7, p. 2217–2226, 2019. ISSN 21511535. 33, 34, 45, 46, 47, 51, 60, 62

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Monitoramento da cobertura e uso da terra do Brasil**: 2016 - 2018. Rio de Janeiro, 2020. 26 p. Disponível em: <<https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2101703>>. Acesso em: 10 maio 2023. 40, 53, 54

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). **TerraClass - Bioma Cerrado**. 2018. Disponível em: <<https://www.terraclass.gov.br/helpers/terraclass_data4download/docs/NotaTecnicaTerraClassCerrado2018.pdf>>. Acesso em: 23 set. 2023. 36, 68

KAPOOR, R.; SHARMA, D.; GULATI, T. State of the art content based image retrieval techniques using deep learning: a survey. **Multimedia Tools and Applications**, v. 80, p. 29561–29583, 8 2021. ISSN 1380-7501. Disponível em: <<https://link.springer.com/10.1007/s11042-021-11045-1>>. 12, 33, 46, 68, 75

KARPATHY, A. **Convolutional Neural Networks (CNNs/ConvNets)**. Stanford-USA: Computer Science Department of Stanford University, 2019. Disponível em: <<http://cs231n.github.io/convolutional-networks/>>. 21, 23, 24

KATO, T. Database architecture for content-based image retrieval. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS, 1992, San Jose, USA. **Proceedings...** SPIE, 1992. v. 1662, p. 112 – 123. Disponível em: <<https://doi.org/10.1117/12.58497>>. 2

KIM, Y. **Convolutional neural networks for sentence classification**. arXiv:1408.5882, 2014. 19

KÖRTING, T. S. **Gerenciamento de metadados de grandes volumes de dados de sensoriamento remoto**. [S.l.]: Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), 2018. 2

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: PEREIRA, F.; BURGES, C. J. C.; BOTTOU, L.; WEINBERGER, K. Q. (Ed.). **Advances in neural information processing systems**. Curran Associates, 2012. p. 1097–1105. Disponível em: <<http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>>. 2, 20

KULIS, B.; GRAUMAN, K. Kernelized locality-sensitive hashing. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 34, n. 6, p. 1092–1104, 2012. 18

LECUN, Y.; BENGIO, Y. Convolutional networks for images, speech, and time series. In: ARBIB, M. A. (Ed.). **Handbook of brain theory and neural networks**. The MIT Press, 1995. p. 255–258. Disponível em: <<https://www.researchgate.net/profile/Yann_Lecun/publication/2453996_Convolutional_Networks_for_Images_Speech_and_Time-Series/links/0deec519dfa2325502000000.pdf>>. 19, 20, 21, 23

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 2015. ISSN 14764687. 2

LECUN, Y.; BOSER, B. E.; DENKER, J. S.; HENDERSON, D.; HOWARD, R. E.; HUBBARD, W. E.; JACKEL, L. D. Handwritten digit recognition with a back-propagation network. In: LIPPMANN, R. P.; MOODY, J.; TOURETZKY, D. (Ed.). **Advances in neural information processing systems**. [S.l.: s.n.], 1990. p. 396–404. 2, 19, 20

LEMAÎTRE, G.; NOGUEIRA, F.; ARIDAS, C. K. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. **Journal of Machine Learning Research**, v. 18, n. 17, p. 1–5, 2017. Disponível em: <<http://jmlr.org/papers/v18/16-365.html>>. 49

LI, P.; REN, P. Partial randomness hashing for large-scale remote sensing image retrieval. **IEEE Geoscience and Remote Sensing Letters**, v. 14, n. 3, p. 464–468, mar 2017. ISSN 1545-598X. Disponível em: <<http://ieeexplore.ieee.org/document/7835693/>>. 13, 16

LI, W. J.; WANG, S.; KANG, W. C. Feature learning based deep supervised hashing with pairwise labels. In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICAL INTELLIGENCE, 25., 2016, New York, USA. **Proceedings...** New York: AAAI Press, 2016. p. 1711–1717. ISSN 10450823. 13, 17, 26

LI, X.; YANG, J.; MA, J. Recent developments of content-based image retrieval (cbir). **Neurocomputing**, v. 452, p. 675–689, 9 2021. ISSN 0925-2312. 33

LI, Y.; MA, J.; ZHANG, Y. Image retrieval from remote sensing big data: a survey. **Information Fusion**, v. 67, p. 94–115, mar 2021. Disponível em: <<https://doi.org/10.1016/j.inffus.2020.10.008>>. 2

LI, Y.; ZHANG, Y.; HUANG, X.; ZHU, H.; MA, J. Large-scale remote sensing image retrieval by deep hashing neural networks. **IEEE Transactions on Geoscience and Remote Sensing**, v. 56, n. 2, p. 950–965, 2018. ISSN 01962892. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/8067633/>>. 2, 9, 12, 13, 19, 25, 26, 27, 33, 78

LIU, H.; WANG, R.; SHAN, S.; CHEN, X. Deep supervised hashing for fast image retrieval. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2016, Las Vegas, USA. **Proceedings...** Las Vegas, 2016. p. 2064–2072. ISSN 1063-6919. 2, 26

LIU, P.; DI, L.; DU, Q.; WANG, L. **Remote sensing big data: theory, methods and applications**. [S.l.]: MDPI, 2018. 711 p. 9

LIU, W.; WANG, J.; JI, R.; JIANG, Y.-G.; CHANG, S.-F. Supervised hashing with kernels. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2012, Providence, USA. **Proceedings...** Providence: IEEE, 2012. p. 2074–2081. 18

MA, Y.; WU, H.; WANG, L.; HUANG, B.; RANJAN, R.; ZOMAYA, A.; JIE, W. Remote sensing big data computing: challenges and opportunities. **Future Generation Computer Systems**, v. 51, p. 47–60, 2015. ISSN 0167739X. Disponível em: <<http://dx.doi.org/10.1016/j.future.2014.10.029>>. 1, 9

MAATEN, L. Van der; HINTON, G. Visualizing data using t-SNE. **Journal of Machine Learning Research**, v. 9, n. 11, 2008. 32

MAJI, S.; BOSE, S. An improved relevance feedback in cbir. **arXiv preprint arXiv:2006.11821**, 2020. 82

MARRIS, E. The forgotten ecosystem. **Nature**, v. 437, n. 7061, p. 944–945, oct 2005. ISSN 0028-0836. Disponível em: <<http://www.nature.com/articles/437944a>>. 36

MARUJO, R. F. B.; FERREIRA, K. R.; QUEIROZ, G. R.; COSTA, R. W.; ARCANJO, J. S.; SOUZA, R. C. M. Generating analysis ready data collections for Brazil. In: IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 2022, Kuala Lumpur, Malaysia. **Proceedings...** Kuala Lumpur: IEEE, 2022. p. 6844–6847. 37

MATOS, P.; LOMBARDI, L.; CIFERRI, R.; PARDO, T.; CIFERRI, C.; VIEIRA, M. **Relatório técnico "métricas de avaliaçao"**. São Carlos: USP, 2009. 17 p. 41

MAUS, V.; CÂMARA, G.; CARTAXO, R.; SANCHEZ, A.; RAMOS, F. M.; QUEIROZ, G. R. de. A Time-Weighted Dynamic Time Warping method for land use and land cover mapping. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 9, n. 8, p. 3729–3739, 2016. 81

MUDELE, O.; GAMBA, P. Mapping vegetation in urban areas using sentinel-2. In: IEEE JOINT URBAN REMOTE SENSING EVENT, 2019, Vannes, France. **Proceedings...** Vannes: IEEE, 2019. p. 1–4. 81

MUJA, M.; LOWE, D. G. Fast approximate nearest neighbors with automatic algorithm configuration. In: INTERNATIONAL CONFERENCE ON COMPUTER VISION THEORY AND APPLICATIONS, 4., 2009, Lisboa, Portugal. **Proceedings...** Lisboa: Springer, 2009. p. 331–340. ISBN 9789898111692. 12

NAPOLETANO, P. Visual descriptors for content-based retrieval of remote-sensing images. **International Journal of Remote Sensing**, v. 39, n. 5, p. 1343–1376, 2018. Disponível em: <<https://doi.org/10.1080/01431161.2017.1399472>>. 10, 11, 25, 33

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST). **TREC-2004 common evaluation measures**. 2023. Disponível em: <<https://trec.nist.gov/pubs/trec14/appendices/CE.MEASURES05.pdf>>. Acesso em: 07 jan. 2024. 42

NIELSEN, M. A. **Neural networks and deep learning**. Determination Press, 2015. Disponível em: <<http://neuralnetworksanddeeplearning.com/chap6.html>>. 22, 23, 24

PAULEVÉ, L.; JÉGOU, H.; AMSALEG, L. Locality sensitive hashing: a comparison of hash function types and querying mechanisms. **Pattern Recognition Letters**, v. 31, n. 11, p. 1348–1358, 2010. ISSN 01678655. Disponível em: <<http://dx.doi.org/10.1016/j.patrec.2010.04.004>>. 17

PLETSCH, M. A. J. S.; KÖRTING, T. S. Information mining for automatic search in remote sensing image catalogs. **Revista Brasileira de Cartografia**, v. 70, n. 5, p. 1860–1884, dec 2018. ISSN 18080936. Disponível em: <<http://www.seer.ufu.br/index.php/revistabrasileiracartografia/article/view/47413>>. 1, 79, 80

PUTZU, L.; PIRAS, L.; GIACINTO, G. Convolutional neural networks for relevance feedback in content based image retrieval: a content based image retrieval system that exploits convolutional neural networks both for feature extraction and for relevance feedback. **Multimedia Tools and Applications**, v. 79, n. 37-38, p. 26995–27021, 2020. ISSN 15737721. 83

RAHUL, D. J. S. A comprehensive survey on image search using binary hash codes. **International Journal of Advanced Research in Computer and Communication Engineering**, v. 3, n. 11, p. 8459–8463, nov 2014. 16

RIBEIRO, J. F.; WALTER, B. M. T. As principais fitofisionomias do bioma Cerrado. In: SANO, S. M.; ALMEIDA, S. P. d.; RIBEIRO, J. F. (Ed.). **Cerrado: ecologia e flora**. Brasília, Brasil: EMBRAPA, 2008. p. 152–212. 36

RIPLEY, B. D. **Pattern recognition and neural networks**. [S.l.]: Cambridge: Cambridge Press, 1996. 19

RIZON, M.; YAZID, H.; SAAD, P.; MD SHAKAFF, A. Y.; SAAD, A. R.; SUGISAKA, M.; YAACOB, S.; MAMAT, M.; KARTHIGAYA, M. Object detection using circular hough transform. **American Journal of Applied Sciences**, v. 2, n. 12, p. 1606–1609, 2005. ISSN 15469239. Disponível em: <<doi.org/10.3844/ajassp.2005.1606.1609>>. 81

RODRIGUES, M.; KÖRTING, T.; QUEIROZ, G.; SALES, C.; SILVA, L. Detecting center pivots in MATOPIBA using hough transform and web time series service. In: IEEE LATIN AMERICA GRSSS & ISPRS REMOTE SENSING CONFERENCE, 2020, Santiago, Chile. **Proceedings...** Santiago, 2020. p. 189–194. 36, 79, 80

RODRIGUES, M. L.; KÖRTING, T. S.; QUEIROZ, G. R. Circular hough transform and balanced random forest to detect center pivots. In: BRAZILIAN SYMPOSIUM ON GEOINFORMATICS (GEOINFO), 2020, São José dos Campos, Brasil. **Proceedings...** São José dos Campos, 2020. p. 106–117. 79, 80

RODRIGUES, M. L.; KÖRTING, T. S.; QUEIROZ, G. R. de. A framework to automatic detect center pivots using land use and land cover data. **Revista Brasileira de Cartografia**, v. 73, n. 4, p. 1048–1070, oct 2021. ISSN 1808-0936. Disponível em: <<http://www.seer.ufu.br/index.php/revistabrasileiracartografia/article/view/60553>>. 36

ROUSSEEUW, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, v. 20, p. 53–65, nov. 1987. ISSN 0377-0427. Disponível em: <<http://dx.doi.org/10.1016/0377-0427(87)90125-7>>. 67

ROY, S.; SANGINETO, E.; DEMIR, B.; SEBE, N. Deep metric and hash-code learning for content-based retrieval of remote sensing images. In: INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 2018, Valencia, Spain. **Proceedings...** Valencia: IEEE, 2018. p. 4539–4542. Disponível em: <<https://ieeexplore.ieee.org/document/8518381/>>. 19, 25, 28, 29, 30

_____. Metric-learning-based deep hashing network for content-based retrieval of remote sensing images. **IEEE Geoscience and Remote Sensing Letters**, p.

226–230, 2020. ISSN 1558-0571. Disponível em: <<http://dx.doi.org/10.1109/LGRS.2020.2974629>>. 19, 25, 28, 30, 31, 32, 100

_____. _____. **IEEE Geoscience and Remote Sensing Letters**, v. 18, n. 2, p. 226–230, 2021. 25, 46, 56, 57

RUDORFF, B. F. T.; ADAMI, M.; AGUIAR, D. A.; MOREIRA, M. A.; MELLO, M. P.; FABIANI, L.; AMARAL, D. F.; PIRES, B. M. The soy moratorium in the amazon biome monitored by remote sensing images. **Remote Sensing**, v. 3, n. 1, p. 185–202, 2011. ISSN 2072-4292. Disponível em: <<https://www.mdpi.com/2072-4292/3/1/185>>. 36

RUSSAKOVSKY, O.; DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S.; MA, S.; HUANG, Z.; KARPATHY, A.; KHOSLA, A.; BERNSTEIN, M.; BERG, A. C.; FEI-FEI, L. Imagenet large scale visual recognition challenge. **International Journal of Computer Vision (IJCV)**, v. 115, n. 3, p. 211–252, 2015. 28

SAJJAD, H.; KUMAR, P. Future challenges and perspective of remote sensing technology. In: KUMAR, P.; RANI, M.; PANDEY PREMAND SAJJAD, H. C.; CHAUDHARY, B. S. (Ed.). **Applications and challenges of geospatial technology**. Cham: Springer, 2019. p. 275–277. ISBN 978-3-319-99882-4. Disponível em: <<https://doi.org/10.1007/978-3-319-99882-4_16>>. 1

SANO, E. E.; ROSA, R.; SCARAMUZZA, C. A. d. M.; ADAMI, M.; BOLFE, E. L.; COUTINHO, A. C.; ESQUERDO, J. C. D. M.; MAURANO, L. E. P.; NARVAES, I. d. S.; Oliveira Filho, F. J. B. de; SILVA, E. B. da; VICTORIA, D. d. C.; FERREIRA, L. G.; BRITO, J. L. S.; BAYMA, A. P.; OLIVEIRA, G. H. de; BAYMA-SILVA, G. Land use dynamics in the Brazilian Cerrado in the period from 2002 to 2013. **Pesquisa Agropecuária Brasileira**, v. 54, apr 2019. ISSN 1678-3921. Disponível em: <<http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-204X2019000103700&tlng=en>>. 36

SARKAR, D.; BALI, R. **Transfer learning in action**. jan 2022. Disponível em: <<https://livebook.manning.com/book/transfer-learning-in-action>>. 53, 75

SCHROFF, F.; KALENICHENKO, D.; PHILBIN, J. Facenet: A unified embedding for face recognition and clustering. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2015. 25

SEMLALI, B.-E. B.; AMRANI, C. E.; ORTIZ, G. Adopting the hadoop architecture to process satellite pollution big data. **International Journal of Technology and Engineering Studies**, v. 5, n. 2, p. 30–39, 2019. 1

SHAKHNAROVICH, G.; DARRELL, T.; INDYK, P. **Nearest-neighbor methods in learning and vision**: theory and practice. [S.l.]: Massachusetts Institute, 2005. 13, 17, 26

SILVA, A. T. da; MAGALHÃES, L. P.; FALCÃO, A. X. Recuperação de imagens por realimentação de relevância. In: ENCONTRO DOS ALUNOS E DOCENTES DO DEPARTAMENTO DE ENGENHARIA DE COMPUTAÇÃO E AUTOMAÇÃO INDUSTRIAL. **Anais...** [S.l.], 2009. 82

SLANEY, M.; CASEY, M. Locality-sensitive hashing for finding nearest neighbors. **IEEE Signal Processing Magazine**, v. 25, n. 2, p. 128–131, 2008. ISSN 10535888. 9, 12, 13, 17

SOUZA, C. M.; SHIMBO, J. Z.; ROSA, M. R.; PARENTE, L. L.; ALENCAR, A. A.; RUDORFF, B. F. T.; HASENACK, H.; MATSUMOTO, M.; FERREIRA, L. G.; SOUZA-FILHO, P. W. M.; OLIVEIRA, S. W. de; ROCHA, W. F.; FONSECA, A. V.; MARQUES, C. B.; DINIZ, C. G.; COSTA, D.; MONTEIRO, D.; ROSA, E. R.; VéLEZ-MARTIN, E.; WEBER, E. J.; LENTI, F. E. B.; PATERNOST, F. F.; PAREYN, F. G. C.; SIQUEIRA, J. V.; VIERA, J. L.; NETO, L. C. F.; SARAIVA, M. M.; SALES, M. H.; SALGADO, M. P. G.; VASCONCELOS, R.; GALANO, S.; MESQUITA, V. V.; AZEVEDO, T. Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine. **Remote Sensing**, v. 12, p. 2735, 8 2020. ISSN 20724292. Disponível em: <<https://www.mdpi.com/2072-4292/12/17/2735>>. 68

SUMBUL, G.; CHARFUELAN, M.; DEMIR, B.; MARKL, V. BigEarthNet: a large-scale benchmark archive for remote sensing image understanding. In: IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 2019, Yokohama, Japan. **Proceedings...** Yokohama: IEEE, 2019. p. 5901–5904. ISBN 978-1-5386-9154-0. Disponível em: <<https://ieeexplore.ieee.org/document/8900532/>>. 33, 78, 97

SUMBUL, G.; DEMIR, B. A deep multi-attention driven approach for multi-label remote sensing image classification. **IEEE Access**, v. 8, p. 95934–95946, 2020. ISSN 21693536. 62

SUMBUL, G.; KANG, J.; KREUZIGER, T.; MARCELINO, F.; COSTA, H.; BENEVIDES, P.; CAETANO, M.; DEMIR, B. BigEarthNet dataset with a new class-nomenclature for remote sensing image understanding. **arXiv e-prints**, p. arXiv:2001.06372, 2020. 45, 98, 99

SUMBUL, G.; WALL, A. de; KREUZIGER, T.; MARCELINO, F.; COSTA, H.; BENEVIDES, P.; CAETANO, M.; DEMIR, B.; MARKL, V. BigEarthNet-MM: a large-scale, multimodal, multilabel benchmark archive for remote sensing image classification and retrieval [software and data sets]. **IEEE Geoscience and Remote Sensing Magazine**, v. 9, n. 3, p. 174–180, set 2021. Disponível em: <<https://doi.org/10.1109/mgrs.2021.3089174>>. 33, 46, 47, 78, 97

SZEGEDY, C.; VANHOUCKE, V.; IOFFE, S.; SHLENS, J.; WOJNA, Z. Rethinking the inception architecture for computer vision. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION, 2016, Las Vegas, USA. **Proceedings...** Las Vegas: IEEE, 2016. p. 2818–2826. 28, 46, 47, 69

TAN, R. J. **Breaking down mean average precision (map)**: another metric for your data science toolkit. 2019. Disponível em: <<https://towardsdatascience.com/breaking-down-mean-average-precision-map-ae462f623a52>>. Acesso em: 19 mar. 2020. 42, 43, 44

TURPIN, A.; SCHOLER, F. User performance versus precision measures for simple search tasks. In: INTERNATIONAL ACM SIGIR CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL, 29., 2006, Seattle, USA. **Proceedings...** Seattle: Association for Computing Machinery, 2006. p. 11–18. ISBN 1595933697. 43, 44, 57

VIDHYA, G. R.; RAMESH, H. Effectiveness of contrast limited adaptive histogram equalization technique on multispectral satellite imagery. In: INTERNATIONAL CONFERENCE ON VIDEO AND IMAGE PROCESSING, Singapore, Singapore. **Proceedings...** Singapore: Association for Computing Machinery, 2017. p. 234–239. ISBN 9781450353830. 51

WANG, J.; ZHANG, T.; SONG j.; SEBE, N.; SHEN, H. T. A survey on learning to hash. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 40, n. 4, p. 769–790, April 2018. ISSN 1939-3539. 16

WILCOXON, F. Individual comparisons by ranking methods. In: KOTZ, S.; JOHNSON, N. L. (Ed.). **Breakthroughs in Statistics**. New York: Springer, 1992. p. 196–202. Springer Series in Statistics. 58

XIA, G. S.; HU, J.; HU, F.; SHI, B.; BAI, X.; ZHONG, Y.; ZHANG, L.; LU, X. AID: A benchmark data set for performance evaluation of aerial scene classification. **IEEE Transactions on Geoscience and Remote Sensing**, v. 55, n. 7, p. 3965–3981, 2017. ISSN 01962892. 34, 102

XU, C.; DU, X.; FAN, X.; GIULIANI, G.; HU, Z.; WANG, W.; LIU, J.; WANG, T.; YAN, Z.; ZHU, J.; JIANG, T.; GUO, H. Cloud-based storage and computing for remote sensing big data: a technical review. **International Journal of Digital Earth**, v. 15, n. 1, p. 1417–1445, 2022. 1

YANG, Y.; NEWSAM, S. Bag-of-visual-words and spatial extensions for land-use classification. In: SIGSPATIAL INTERNATIONAL CONFERENCE ON ADVANCES IN GEOGRAPHIC INFORMATION SYSTEMS, 18., San Jose, USA. **Proceedings...** San Jose: Association for Computing Machinery, 2010. p. 270–279. 32, 78, 102

YASSINE, H.; TOUT, K.; JABER, M. Improving lulc classification from satellite imagery using deep learning - eurosat dataset. **International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives**, v. 43, p. 369–376, 2021. ISSN 16821750. 33, 46

YUEN, H.; PRINCEN, J.; ILLINGWORTH, J.; KITTLER, J. Comparative study of Hough Transform methods for circle finding. **Image and Vision Computing**, v. 8, n. 1, p. 71–77, feb 1990. ISSN 02628856. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/026288569090059E>>. 81

ZHOU, W.; GUAN, H.; LI, Z.; SHAO, Z.; DELAVAR, M. R. Remote sensing image retrieval in the past decade: achievements, challenges, and future directions. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 16, p. 1447–1473, 2023. ISSN 21511535. 33

ZHU, H.; LONG, M.; WANG, J.; CAO, Y. Deep hashing network for efficient similarity retrieval. In: AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE, 30., 2016, Phoenix, USA. **Proceedings...** Phoenix: AAAI Press, 2016. p. 2415–2421. ISBN 9781577357605. 2, 13, 26

ZHU, X.; ZHANG, L.; HUANG, Z. A sparse embedding and least variance encoding approach to hashing. **IEEE Transactions on Image Processing**, v. 23, n. 9, p. 3737–3750, 2014. ISSN 10577149. 25

ZIOTI, F.; FERREIRA, K. R.; QUEIROZ, G. R.; NEVES, A. K.; CARLOS, F. M.; SOUZA, F. C.; SANTOS, L. A.; SIMOES, R. E. A platform for land use and land cover data integration and trajectory analysis. **International Journal of Applied Earth Observation and Geoinformation**, v. 106, 2022. ISSN 15698432. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0303243421003627>>. 38

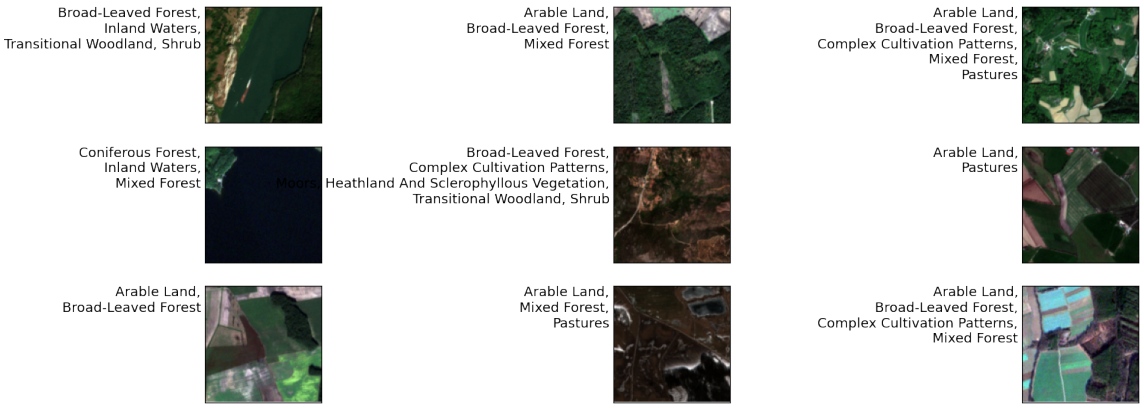
ZUIDERVELD, K. Contrast limited adaptive histogram equalization. In: HECBKERT, P. S. (Ed.). **Graphic Gems**. [S.l.]: Academic Press, 1994. p. 474–485. 51, 105

APÊNDICE A - *DEEP LEARNING* PARA CBIR APLICADO AO CONJUNTO BIGEARTHNET

A.1 Conjunto BigEarthNet

O conjunto BigEarthNet (v1.0-beta)[1] é formado por 590.326 *patches* de imagens do Sentinel-2 totalizando 125 cenas adquiridas entre junho de 2017 e maio de 2018 sobre 10 países da Europa (Áustria, Bélgica, Finlândia, Irlanda, Kosovo, Lituânia, Luxemburgo, Portugal, Sérvia, Suíça). As cenas foram corrigidas atmosfericamente (*Level 2A*) utilizando a ferramenta sen2cor. Os *patches* representam áreas sem sobreposição e foram rotulados com até 19 classes de uso e cobertura da terra identificados com base no *CORINE Land Cover inventory*[2] de 2018. O diferencial desse *dataset* é justamente o grande número de imagens multiespectrais de sensoriamento remoto (SR) por satélite rotuladas com múltiplos rótulos (*labels*), quando um *patch* pode possuir mais de um tipo de uso e cobertura associado (Figura A.1), abrindo à possibilidade de pesquisas promissoras em análise de imagens de sensoriamento remoto em grande escala (SUMBUL et al., 2019).

Figura A.1 - Amostra de imagens do conjunto BigEarthNet com múltiplos rótulos de uso e cobertura da terra.



Fonte: Próprio autor.

[1] A versão atual desse conjunto denominada BigEarthNet-MM, foi enriquecida com dados do radar de abertura sintética banda C presente no satélite Sentinel 1 (SUMBUL et al., 2021).

[2] O programa *Coordination of Information on the Environment* (CORINE), fornece a cada seis anos mapas de uso e cobertura da terra, parâmetros Biogeofísicos e de qualidade do ar para todo continente Europeu. Disponível em <https://land.copernicus.eu/en/products/corine-land-cover>, acesso setembro 2023.

De maneira a subsidiar o uso desse conjunto pela comunidade de sensoriamento remoto, os autores forneceram também várias arquiteturas de redes de *deep learning* pré-treinadas com suas imagens. Tornado possível empregá-las nas mais diversas aplicações sem a necessidade de recursos avançados normalmente necessários para o treinamento dessas arquiteturas[3].

Como demonstrado ao longo deste trabalho o uso de *backbones* (redes convolucionais profundas) pré-treinados com imagens de SR por satélite, apresentam vantagens para aplicação na tarefa de CBIR em relação a modelos pré-treinados com imagens de outros domínios. Portanto, com intuito de caracterizar e indicar a melhor arquitetura para emprego nessa tarefa, foi realizada uma avaliação de desempenho dos modelos pré-treinados com o conjunto BigEarthnet para tarefa de classificação, uma vez que esses modelos servem como módulo de extração de características quando empregados numa abordagem de busca e recuperação de imagens baseada em conteúdo (CBIR). A Figura A.2 sumariza o desempenho dos seguintes modelos: K-Branch, VGG16, VGG19, ResNet50, ResNet101 e ResNet152, baseado nas seguintes métricas de classificação: *Accuracy*, *F-measures Score*, *Hamming Loss*, *Precision* e *Recall*. Além disso, foram utilizadas as seguintes métricas de ranqueamento: *Label Ranking average precision* (LRAP), *One Error* e *Ranking Loss*.

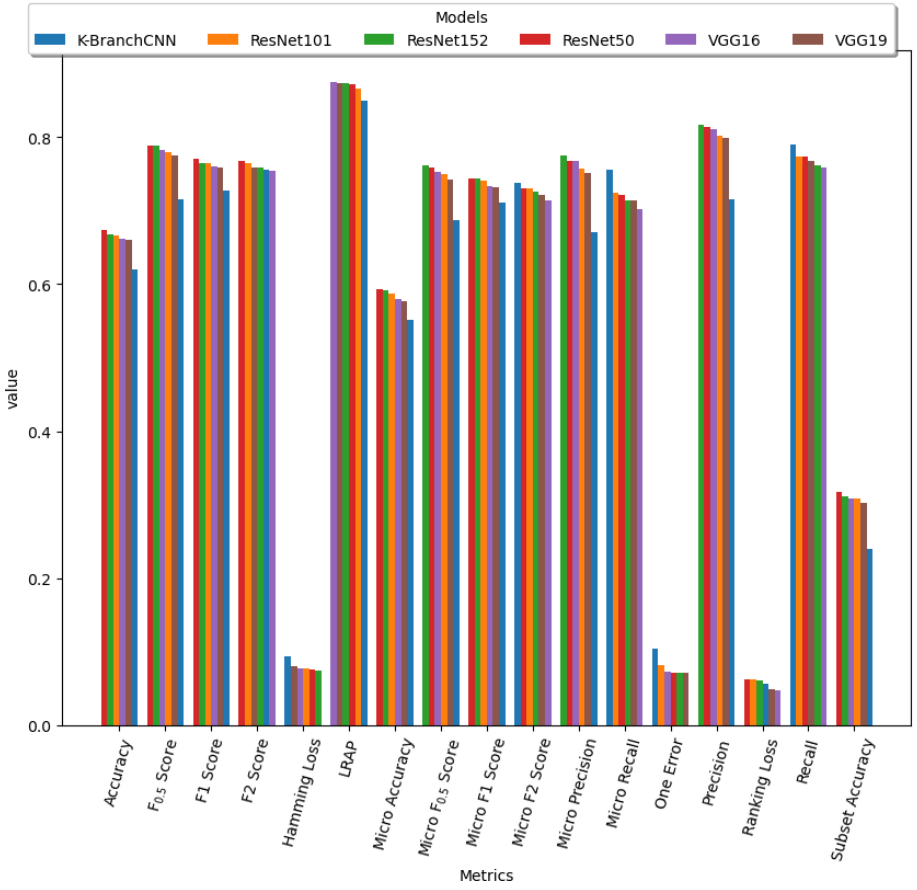
As métricas de classificação são baseadas na lista de classes previstas, enquanto as métricas de ranqueamento consideram também a lista de probabilidades de ocorrência (organizadas do maior para o menor valor) para todas as classes (SUMBUL et al., 2020). Na classificação *multilabel*, o cálculo da precisão para o subconjunto (*Subset Accuracy*) será igual a 1 caso todo o conjunto de rótulos previstos para uma amostra corresponder estritamente ao conjunto verdadeiro de rótulos, caso contrário será 0, indicando a porcentagem de amostras que têm todos os seus rótulos classificados corretamente. Para as outras métricas, o cálculo é realizado para a média de cada instância, enquanto para condição *micro* o cálculo é global baseado no total de casos verdadeiros positivos, falsos negativos e falsos positivos[4].

Especificamente a métrica *Hamming Loss* é a média da distância de *Hamming* calculada entre os rótulos previstos e os rótulos de referência. Enquanto as métricas LRAP, *One Error* e *Ranking Loss* são calculadas baseadas no ranqueamento das pro-

[3]Repositório com modelos de *deep learning* pré-treinados com o conjunto BigEarthNet. Disponível em <https://git.tu-berlin.de/rsim/BigEarthNet-S2_19-classes_models>, acesso setembro 2019.

[4]Documentação do Scikit-learn para métricas de avaliação de modelos. Disponível em <https://scikit-learn.org/stable/modules/model_evaluation.html#classification-metrics>, acesso setembro 2023

Figura A.2 - Comparação do desempenho da classificação de imagens do conjunto BigEarthNet realizada por vários modelos de *Deep Learning*.



Fonte: Próprio autor.

babilidades de ocorrência de um determinado rótulo. Em suma a LRAP é avaliada considerando a taxa de rótulos de referência com classificação mais alta em relação a cada rótulo de referência, já *One Error* calcula a taxa de imagens, cujo rótulo de referência associado não inclui o rótulo previsto em primeiro lugar, enquanto que a métrica *Ranking Loss* calcula o custo de pares de rótulos ordenados incorretamente, ou seja, a probabilidade de ocorrência de um rótulo irrelevante para a imagem, ser maior do que um rótulo de referência (SUMBUL et al., 2020).

Baseado no desempenho geral das métricas é possível inferir que a arquitetura ResNet-50 foi superior as outras arquiteturas (prevalência superior a 47%), constituindo assim uma ótima candidata para uso como *backbone*, ou seja, módulo utilizado para extração de características de imagens para o processo de criação de um espaço métrico de recuperação de imagens baseada no conteúdo. Essa avaliação

assim como outras realizadas ao longo do desenvolvimento deste trabalho culminaram com a escolha da ResNet-50 como alternativa a Inception V3 originalmente adotada na abordagem de DL para CBIR proposta Roy et al. (2020) (rede MiLaN). Entretanto, para efeito de comparação o experimento com o conjunto BigEarthNet utilizou o *backbone* da versão original.

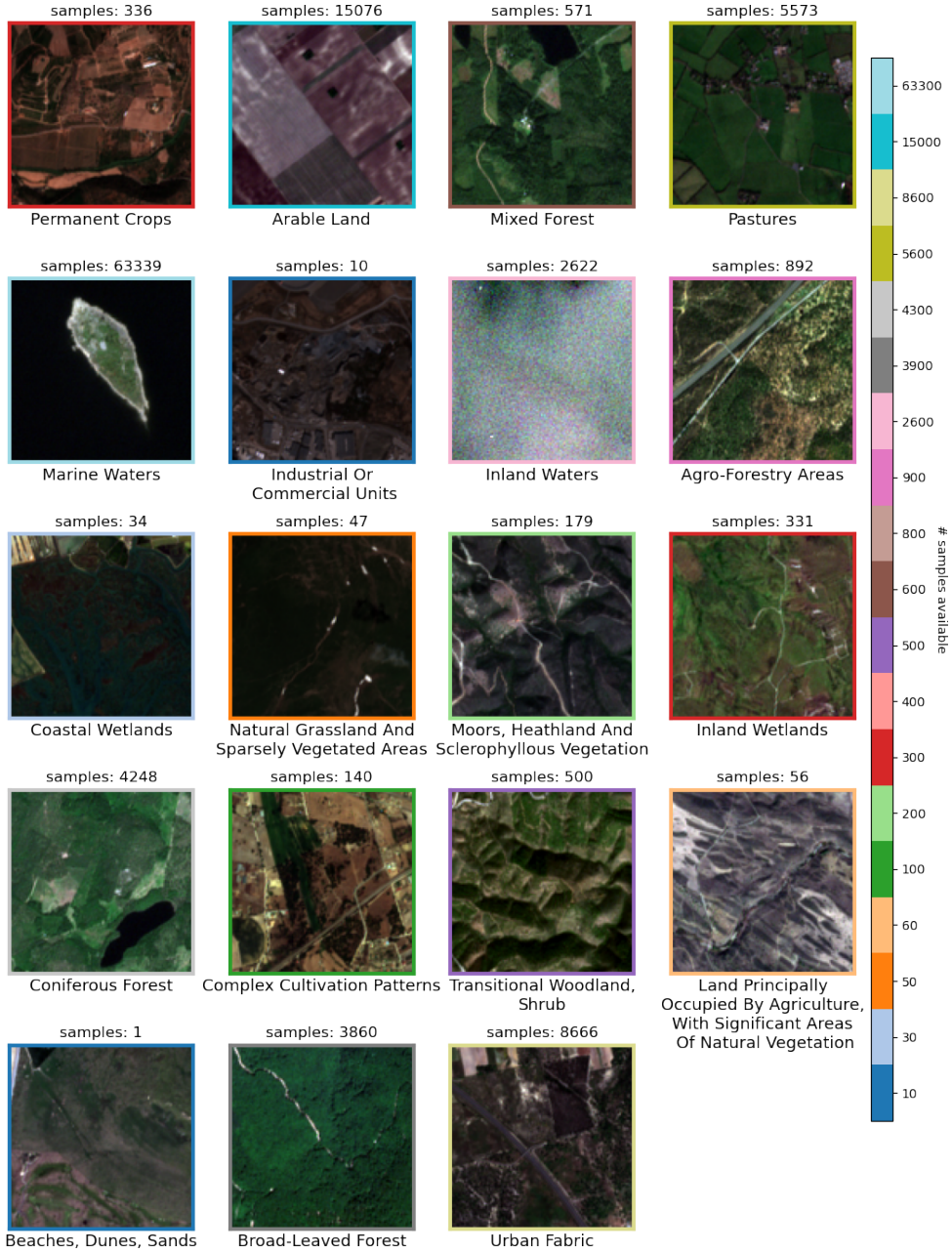
A.2 CBIR de imagens do conjunto BigEarthNet com a *Metric-Learning-Based Deep Hashing Network* (MiLaN)

Como já exposto, a abordagem da rede MiLaN é baseada em imagens de rótulo único para estimar a taxa de perda durante o aprendizado por similaridade baseado na *triplet* imagem de referência (âncora), imagens semelhantes (casos positivos, mesmo rótulo) e imagens distintas (casos negativos, rótulos diferentes) para criar o espaço métrico ideal para tarefa CBIR. Dessa forma, para realização dos testes com o conjunto BigEarthNet, foi necessário realizar uma subamostragem do conjunto para selecionar somente imagens que pudessem ser treinadas com a rede MiLaN (*single label*), como demostrando na Figura A.3, esse processo deu origem a um subconjunto de dados desbalanceado com variação de 1 a mais de 60 mil amostras a depender da classe, totalizando 106.481 *patches*.

Para corrigir o desbalanceamento do subconjunto de imagens, restringiu-se o uso de classes com no mínimo 50 amostras, além da aplicação das seguintes técnicas *Random Undersampling* e *Random Oversampling* (BROWNLEE, 2021). A primeira foi usada para reduzir o número de amostras das classes dominantes até o limite de 500 imagens, e a segunda técnica para criar novas exemplos das classes minoritárias até atingir o balanceamento. Importante mencionar que a criação de novos exemplos a partir de classes minoritárias é feita através da escolha aleatória de imagens disponíveis, ou seja, não é criada nenhuma nova informação. Como resultado temos um conjunto de 7500 imagens com os seguintes tipos de uso e cobertura da terra: *Urban fabric, Arable land, Permanent crops, Pastures, Complex cultivation patterns, (Land principally occupied by agriculture, with significant areas of natural vegetation), Agro-forestry areas, Broad-leaved forest, Coniferous forest, Mixed forest, (Moors, heathland and sclerophyllous vegetation), (Transitional woodland, shrub), Inland wetlands, Inland water* e *Marine waters*.

Como demonstrado anteriormente (Capítulo MATERIAIS E MÉTODOS 3), a rede MiLaN utiliza características como cores e formas extraídas das imagens por uma rede convolucional profunda (*backbone*) para construção de um espaço métrico ideal para tarefa de CBIR, essas informações são compiladas em um vetor que serve de

Figura A.3 - Amostras de imagens do subconjunto BigEarthNet (*single label*).



Fonte: Próprio autor.

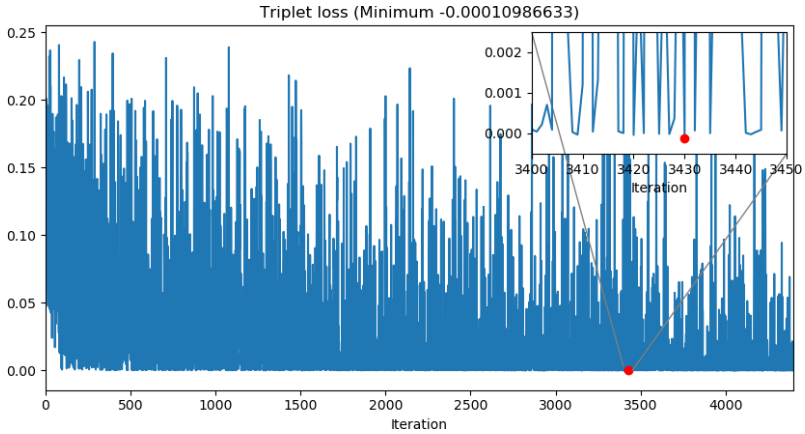
base para o treinamento da rede. Esse vetor representa todo o conhecimento extraído de forma semântica das imagens pelo *backbone*.

O experimento com o conjunto BigEarthNet foi realizado utilizando somente as bandas RGB e a rede Inception V3 pré-treinada com o conjunto ImageNet como

backbone, para permitir a comparação com os resultados obtidos originalmente com a rede MiLaN utilizando outros conjuntos de imagens[5] e reproduzidos aqui com base no código fonte disponível[6].

O subconjunto de imagens (*single label*) aplicado a rede Inception, dá origem a um conjunto de 7500 vetores de características com 2800 valores cada (camada neurônios imediatamente anterior a camada final (classificação) do *backbone*) que compilam o conhecimento da rede sobre as imagens. Esses vetores foram divididos em 70%/30% treinamento e teste da MiLaN, utilizando o método *CallBackStop* para interromper o processo de treinamento caso o decaimento não apresente melhoria ao longo das iterações, o parâmetro para interromper a otimização (*patient*) foi definido como 1.000 iterações (Figura A.4).

Figura A.4 - Treinamento da rede MiLaN com o subconjunto BigEarthNet (*single label*).



Fonte: Próprio autor.

Os experimentos incluíram a replicação dos testes e resultados com as imagens dos conjuntos UCMD, AID além do BigEarthNet utilizando a métrica *mean Average Precision* (*mAP*) para avaliar o desempenho da recuperação de imagens com a rede MiLaN. A Tabela A.1 apresenta a comparação dos resultados globais para recuperação de imagens baseada no conteúdo (CBIR), o desempenho com o conjunto BigEarthNet foi inferior em todos os casos. Alguns fatores elencados a seguir ex-

[5] A proposta da rede MiLaN utilizou os seguintes conjuntos de imagens: *The University of California Merced Land Use Dataset* (UCMD) - conjunto de ortoimagens aéreas extraídas da coleção de mapeamento de áreas urbanas dos Estados Unidos (YANG; NEWSAM, 2010) e *Aerial Image Dataset* (AID) - imagens aéreas coletadas do Google Earth Engine (XIA et al., 2017).

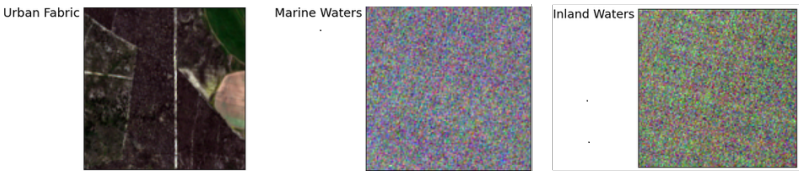
[6] Código fonte da rede MiLaN disponível em <https://github.com/MLEnthusiast/MHCLN>

plicam esse fato: i) *backbone* treinado com imagens de domínio diferente do SR por satélite; ii) resolução espacial; iii) uso limitado da informação disponível (somente bandas RGB); iv) imagens com uso e cobertura incorretamente identificados ou ruidosas (Figura A.5).

Tabela A.1 - Desempenho da recuperação de imagens global para os conjuntos UCMD, AID e BigEarthNet, considerando $k = 20, 50, 100$.

Dataset	**mAP@20**	**mAP@50**	**mAP@100**
UCMD	0,9185	0,8992	0,8873
AID	**0,9281**	**0,9230**	**0,9147**
BigEArthNet	0,7830	0,7616	0,7409

Figura A.5 - Exemplos de imagens BigEarthNet com problemas de rótulo e ruidosas.



Fonte: Próprio autor.

A Tabela A.2 sumariza os resultados obtidos para cada tipo de uso e cobertura do conjunto de imagens BigEarthNet. O desempenho para algumas das classes (valores em destaque) foi impactado pelo reuso de amostras de imagens durante o processo de balanceamento, imagens com menos de 500 exemplos. Entretanto, dois tipos de uso e cobertura realmente se destacaram alcançando mais de 90% de precisão: *Arable land* ($mAP@20 = 0,9010$) e *Pastures* ($mAP@20 = 0,9638$), por outro lado a classe *Mixed forest* ($mAP@20 = 0,4937$) teve pior desempenho principalmente pela similaridade com o tipo *Coniferous forest*.

As Figuras A.6 e A.7 ilustram respectivamente o pior e melhor desempenho para recuperação de imagens entre as classes de uso e cobertura da terra no conjunto BigEarthNet desconsiderando o efeito do processo de balanceamento.

O experimento com o conjunto de imagens BigEarthNet foi uma indicação apresentada na proposta desta tese como possibilidade para o avanço da aplicação da

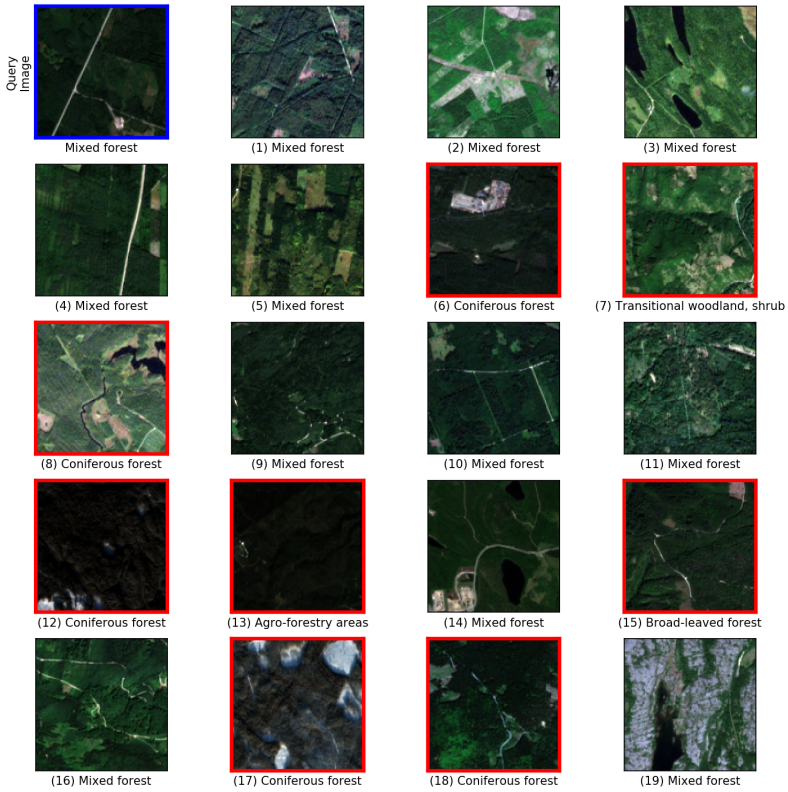
Tabela A.2 - Desempenho da recuperação de imagens por tipo de uso e cobertura da terra do conjunto BigEarthNet, considerando $k = 20, 50, 100$.

Classe	**mAP@20**	**mAP@50**	**mAP@100**
Mixed forest	0,4937	0,4292	0,3928
Coniferous forest	0,5524	0,5147	0,4932
Urban fabric	0,5998	0,5555	0,5215
Transitional woodland, shrub	0,6760	0,6463	0,6113
Broad-leaved forest	0,7194	0,7012	0,6792
Agro-forestry areas	0,7399	0,7220	0,7006
Marine waters	0,7564	0,7381	0,7220
Inland waters	0,8011	0,7890	0,7787
Permanent crops	**0,8431**	0,8257	0,7940
Inland wetlands	**0,8942**	0,8840	0,8637
Complex cultivation patterns	**0,9010**	0,8807	0,8510
Arable land	0,9010	0,8937	0,8892
Moors, heathland and sclerophyllous vegetation	**0,9032**	0,8801	0,8541
Pastures	0,9638	0,9640	0,9635
Land principally occupied by agriculture, with significant areas of natural vegetation	**1**	0,9994	0,9984

técnica de recuperação de imagens baseada no conteúdo, tendo como paradigma o *Big Data* em sensoriamento remoto por satélite. Os resultados obtidos permitiram identificar pontos importantes que podem limitar o desempenho da recuperação de imagens nesse contexto.

a) *Backbones* pré-treinados exclusivamente com imagens de outros domínios que não o do SR por satélite são insuficientes para criar representações semânticas das imagens para a identificação do seu conteúdo. Isso se deve principalmente a diferença entre as resoluções espaciais das mesmas. Para lidar com essa questão, uma das inovações proposta nessa tese foi o *fine-tuning* das arquiteturas de DL com o conjunto de imagens satélite além de testes e adoção de uma nova arquitetura baseada no aprendizado residual (ResNet-50).

Figura A.6 - Teste para recuperação de imagens do tipo *Mixed Forest* do conjunto BigEarthNet com a rede MiLaN (pior desempenho).



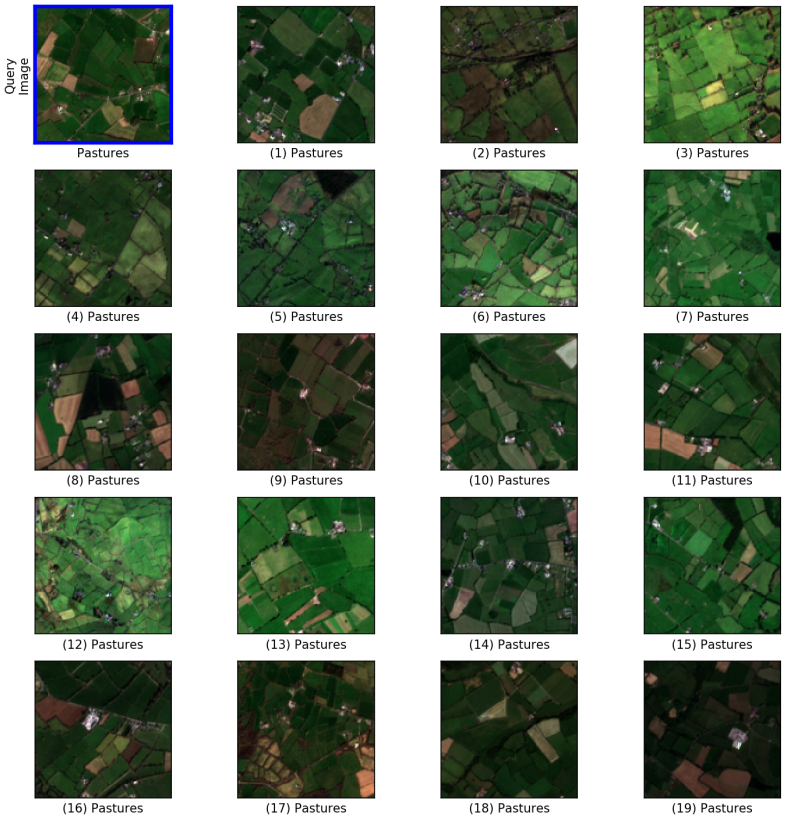
Obs.: Quadrados vermelhos indicam erros na recuperação.

Fonte: Próprio autor.

b) O uso limitado da informação espectral (RGB) torna difícil a diferenciação de imagens que possuam alvos com formas geométricas semelhantes e também que apresentem padrões de uso e cobertura análogos, por exemplo, variações de tipos de cobertura vegetal. Durante a tese foi evidenciado a superioridade alcançada para recuperação de imagens baseada no conteúdo quando utilizada as informações multiespectrais. Além disso, foi explorado o uso da técnica *Contrast Limited Adaptive Histogram Equalization* (CLAHE) (ZUIDERVELD, 1994) para lidar com amostras de imagens com problemas de contraste especialmente devido a falta da correção atmosférica.

Dessa maneira, apesar de o conjunto BigEartNet possuir um enorme potencial para aplicações especialmente com foco nos grandes conjuntos de dados da era do *Re-*

Figura A.7 - Teste para recuperação de imagens do tipo (*Pastures*) do conjunto BigEarthNet com a rede MiLaN (melhor desempenho).



Fonte: Próprio autor.

mote Sensing Big Data, pelos problemas encontrados (rótulos errados/amostras com ruído) e também pela limitação representada pela uso de amostras *single label*, sua adoção foi desencorajada neste trabalho. Possivelmente esse conjunto será alvo de trabalhos futuros para adaptação do *framework* proposto nesta tese (Seção 3.5.3) com objetivo de suportar também o treinamento para identificação de imagens com múltiplos tipos de uso e cobertura da terra (*multilabel*) disponível na nova versão desse conjunto de dados denominado BigEarthNet-MM.

ANEXO A - PRODUÇÃO CIENTÍFICA NO DOUTORADO

A.1 Vinculada ao tema da tese

Periódicos

- **Rodrigues, M. L.**, Körting, T. S., & Queiroz, G. R. de. (2023). Comparative Analysis of Content-Based Image Retrieval from Aerial and Satellite Multispectral Images. IEEE Transactions on Geoscience and Remote Sensing. Submetido.

- **Rodrigues, M. L.**, Körting, T. S., & Queiroz, G. R. de. (2021). A Framework to Automatic Detect Center Pivots Using Land Use and Land Cover Data. Revista Brasileira de Cartografia, 73(4), 1048–1070. https://doi.org/10.14393/rbcv73n4-60553

Simpósios e Conferências

- **M. L. Rodrigues**, T. S. Körting, G. R. de Queiroz, C. P. Sales and L. A. R. d. Silva, Detecting Center Pivots In Matopiba Using Hough Transform And Web Time Series Service, 2020 IEEE Latin American GRSS & ISPRS Remote Sensing Conference (LAGIRS), Santiago, Chile, 2020, pp. 189-194, doi: 10.1109/LAGIRS48042.2020.9165648.

- **Rodrigues, M. L.**, Körting, T. S., & Queiroz, G. R. (2020). Circular Hough Transform and Balanced Random Forest to Detect Center Pivots. Proceedings of the XXI Brazilian Symposium on GeoInformatics (GEOINFO), 106–117. http://urlib.net/rep/8JMKD3MGPDW34P/43PLCP5

- Eiras, D. M. D. A., Pletsch, M. A. J. S., **Rodrigues, M. L.**, & Karine, R. (2020). Identificação de pivôs centrais usando composições de bandas e um método rápido de Deep Learning. Proceedings of the XXI Brazilian Symposium on GeoInformatics (GEOINFO), 180–185. http://urlib.net/rep/8JMKD3MGPDW34P/43PR2H2

Workshops

- **Rodrigues, M. L.**, Körting, T. S., & Queiroz, G. R. Automatic detection of center pivots using circular hough transform, balanced random forest and land use and land cover data. In: WORKSHOP DO

CURSO DE COMPUTAÇÃO APLICADA DO INPE, 21. (WORCAP), 2021, São José dos Campos. Resumos... São José dos Campos: INPE, 2021. On-line. IBI: <8JMKD3MGPDW34P/45U7R38>. Disponível em: <http://urlib.net/ibi/8JMKD3MGPDW34P/45U7R38>.

- **RODRIGUES, M. L.**, ARANTES FILHO, L. R. Tree species classification using spectral samples and geographically weighted variables from Aster sensor onboard Terra satellite. In: WORKSHOP DO CURSO DE COMPUTAÇÃO APLICADA DO INPE, 20. (WORCAP), 2021, São José dos Campos. Participação em desafio (Hackathon).

- **RODRIGUES, M. L.**, Körting, T. S., & Queiroz, G. R. Deep learning e hashing para Content-Based Image Retrieval (CBIR) de imagens de sensoriamento remoto. In: WORKSHOP DO CURSO DE COMPUTAÇÃO APLICADA DO INPE, 20. (WORCAP), 2020, São José dos Campos. Vídeos... São José dos Campos: INPE, 2020. On-line. (16 min). IBI: <8JMKD3MGPDW34P/43HC4NE>. Disponível em: <http://urlib.net/ibi/8JMKD3MGPDW34P/43HC4NE>.

A.2 Colaboração com outros grupos

Periódicos

- ARANTES FILHO, L. R., **RODRIGUES, M. L.**, ROSA, R. R., & GUIMARÃES, L. N. F. (2022). Predicting COVID-19 cases in various scenarios using RNN-LSTM models aided by adaptive linear regression to identify data anomalies. Anais Da Academia Brasileira de Ciências, 94(suppl 3). https://doi.org/10.1590/0001-3765202220210921

Workshops

- ARANTES FILHO, L. R., GUIMARÃES, L. N. F., ROSA, R. R., & **RODRIGUES, M. L.** (2020). DEEP LEARNING APPROACH TO RETRIEVE IMAGE FEATURES OF RADIO SUPERNOVAE REMNANTS. The Radio Universe Workshop's Proceeding. Aceito para publicação.

PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE

Teses e Dissertações (TDI)

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

Manuais Técnicos (MAN)

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

Notas Técnico-Científicas (NTC)

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programas de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

Relatórios de Pesquisa (RPQ)

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

Propostas e Relatórios de Projetos (PRP)

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

Publicações Didáticas (PUD)

Incluem apostilas, notas de aula e manuais didáticos.

Publicações Seriadas

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Constam destas publicações o Internacional Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

Programas de Computador (PDC)

São a seqüência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. Aceitam-se tanto programas fonte quanto os executáveis.

Pré-publicações (PRE)

Todos os artigos publicados em periódicos, anais e como capítulos de livros.