



**UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
CENTRO DE ENGENHARIA ELÉTRICA E INFORMÁTICA
UNIDADE ACADÊMICA DE SISTEMAS E COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

ANTÔNIO GERMINEO LIMA ESTEVES

**RECONHECIMENTO DE PADRÕES DE CRESCIMENTO URBANO EM IMAGENS
DE SATÉLITE UTILIZANDO APRENDIZAGEM PROFUNDA E TEORIA DOS
VALORES EXTREMOS**

**CAMPINA GRANDE - PB
2022**

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Coordenação de Pós-Graduação em Ciência da Computação

Reconhecimento de padrões de crescimento urbano
em imagens de satélite utilizando Aprendizagem
Profunda e Teoria dos Valores Extremos

Antônio Germineo Lima Esteves

Dissertação submetida à Coordenação do Curso de Pós-Graduação em
Ciências da Computação da Universidade Federal de Campina Grande -
Campus I como parte dos requisitos necessários para obtenção do grau
de Mestre em Ciência da Computação.

Área de Concentração: Ciências da Computação

Linha de Pesquisa: Visão Computacional

Orientador: Prof. Dr. Eanes Torres Pereira

Campina Grande, Paraíba, Brasil

Antônio Germineo Lima Esteves, 21/02/2022

E79r

Esteves, Antônio Germineo Lima.

Reconhecimento de padrões de crescimento urbano em imagens de satélite utilizando aprendizagem profunda e teoria dos valores extremos / Antônio Germineo Lima Esteves. – Campina Grande, 2022.

71 f.

Dissertação (Mestrado em Ciência da Computação) – Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2022.

"Orientação: Prof. Dr. Eanes Torres Pereira".

Referências.

1. Sensoriamento Remoto. 2. Teoria dos Valores Extremos. 3. Aprendizagem Profunda. 4. Detecção de Anomalias. 5. *Peaks Over Threshold* (POT). I. Pereira, Eanes Torres. II. Título.

CDU 528.8(043)



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE CAMPINA GRANDE
POS-GRADUACAO CIENCIAS DA COMPUTACAO
Rua Aprigio Veloso, 882, - Bairro Universitario, Campina Grande/PB, CEP 58429-900

FOLHA DE ASSINATURA PARA TESES E DISSERTAÇÕES

ANTÔNIO GERMINEO LIMA ESTEVES

RECONHECIMENTO DE PADRÕES DE CRESCIMENTO URBANO EM IMAGENS DE SATÉLITE
UTILIZANDO APRENDIZAGEM PROFUNDA E TEORIA DOS VALORES EXTREMOS

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação como pré-requisito para obtenção do título de Mestre em Ciência da Computação.

Aprovada em: 21/02/2022

Prof. Dr. EANES TORRES PEREIRA, Orientador, UFCG

Prof. Dr. HERMAN MARTINS GOMES, Examinador Interno, UFCG

Profa. Dra. JOSEANA MACÊDO FECHINE RÉGIS DE ARAÚJO, Examinadora Interna, UFCG

Prof. Dr. MARCO ANTONIO ALVAREZ VEGA, Examinador Externo, UNIVERSITY OF RHODE ISLAND



Documento assinado eletronicamente por **EANES TORRES PEREIRA, PROFESSOR 3 GRAU**, em 21/02/2022, às 19:16, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



Documento assinado eletronicamente por **JOSEANA MACEDO FECHINE, PROFESSOR DO MAGISTERIO SUPERIOR**, em 21/02/2022, às 20:04, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



Documento assinado eletronicamente por **HERMAN MARTINS GOMES, PROFESSOR 3 GRAU**, em 21/02/2022, às 20:50, conforme horário oficial de Brasília, com fundamento no art. 8º, caput, da [Portaria SEI nº 002, de 25 de outubro de 2018](#).



A autenticidade deste documento pode ser conferida no site <https://sei.ufcg.edu.br/autenticidade>, informando o código verificador **2126809** e o código CRC **DBAF12AF**.

Referência: Processo nº 23096.006361/2022-67

SEI nº 2126809

Resumo

O reconhecimento de padrões de crescimento urbano em imagens de satélite apresenta aplicações que variam desde o entendimento da dinâmica da urbanização, até inferir a expansão urbana futura. Tanto a disponibilidade de inventários globais de uso da superfície terrestre, baseados em sensoriamento remoto, quanto os avanços nos métodos de aprendizado profundo, oferecem uma oportunidade para impulsionar o estado da arte dos modelos existentes para este fim. Essa tarefa tem amplas implicações para a preparação para desastres, meio ambiente, desenvolvimento de infraestrutura e prevenção de epidemias, além de desenvolver novos métodos de visão computacional para dados de séries temporais. Inspirados por modelos sequenciais, esse trabalho propõe um método para a detecção de anomalias, ou alterações, utilizando o algoritmo *Peaks Over Threshold* (POT), uma abordagem probabilística paramétrica com base na Teoria dos Valores Extremos que não requer limites definidos manualmente e não pressupõe a distribuição dos dados. O algoritmo foi aplicado às representações obtidas por uma rede neural convolucional (arquitetura U-Net) de modo a reconhecer e detectar possíveis alterações na geografia das regiões, tirando proveito de uma sequência temporal de imagens de sensoriamento remoto extraída do conjunto de dados SpaceNet. Os resultados mostram que, apesar da resolução moderada dos dados, foi possível rastrear identificadores de alterações na superfície terrestre temporalmente. Os resultados validam a eficácia do método proposto na detecção de novidades com resultados de 91,34% e 85% para F-score e revocação, respectivamente, bem como um F-beta de 87,42%, pontuação que representa a média harmônica ponderada de precisão e revocação. Em comparação com a literatura, os resultados reportados por três pesquisas correlatas são de 90%, 71,16% e 69% para o F-score, exclusivamente, sem registro de métricas adicionais. O uso do algoritmo POT em conjunto com uma rede convolucional de arquitetura U-Net, aplicados ao conjunto de dados SpaceNet trouxe evidências experimentais de que se trata de uma abordagem promissora para fornecer automaticamente detecções de mudanças ou alterações espaço-temporais na superfície terrestre em aplicações práticas.

Abstract

The recognition of urban growth patterns in satellite images has applications that range from understanding the dynamics of urbanization to inferring future urban expansion. Both the availability of global inventories of land use, based on remote sensing, and advances in deep learning methods, offer an opportunity to boost the state of the art of existing models for this purpose. This task has broad implications for disaster preparedness, the environment, infrastructure development, and epidemic prevention, as well as developing new computer vision methods for time series data. Inspired by sequential models, this work proposes a method for detecting novelties, or anomalies, using the *Peaks Over Threshold* (POT) algorithm, a parametric probabilistic approach based on the Theory of Extreme Values that does not require manually defined thresholds. and does not presuppose data distribution. The algorithm was applied to representations obtained by a convolutional neural network (U-Net architecture) in order to recognize and detect possible changes in the geography of the regions, taking advantage of a temporal sequence of remote sensing images extracted from the SpaceNet dataset. The results show that, despite the moderate resolution of the data, it was possible to track identifiers of changes in the Earth's surface temporally. The results validate the effectiveness of the proposed method in detecting anomalies with results of 91.34% and 85% for F-score and recall respectively, as well as an F-beta of 87.42%, a score that represents the average precision weighted harmonica and recall. In comparison with the literature, the reported results are 90%, 71.16% and 69% for the F-score, exclusively, without recording additional metrics. The use of the POT algorithm in conjunction with a convolutional network of U-Net architecture, applied to the SpaceNet dataset, has brought experimental evidence that it is a promising approach to automatically provide detections of spatio-temporal changes or alterations on the Earth's surface in practical applications.

Agradecimentos

Agradeço a minha família pelo apoio e dedicação de integral, ao meu orientador que pacientemente me apoiou e fez encontrar o melhor jeito de lidar com os desafios encontrados no decorrer da pesquisa e aos meus colegas de grupo de pesquisa que trouxeram força e sabedoria nos momentos de dificuldade.

Conteúdo

1	Introdução	1
1.1	Problema	3
1.2	Relevância	5
1.3	Escopo	6
1.4	Contribuições da Pesquisa	7
2	Objetivos	8
2.1	Objetivo Geral	8
2.2	Objetivos Específicos	8
3	Fundamentação Teórica	10
3.1	Introdução	10
3.2	Sensoriamento Remoto	10
3.3	Redes Neurais Convolucionais	15
3.4	Detecção de Anomalias	21
3.4.1	Abordagem Probabilística Paramétrica	23
3.4.2	Distribuição de Valores Extremos	24
3.4.3	<i>Peaks-Over-Threshold</i> (POT)	25
3.4.4	Desafios na Detecção de Anomalias	27
4	Revisão de Literatura	29
4.1	Detecção de Anomalias em Imagens de Satélite	29
4.2	Detecção de Anomalias em Imagens Multitemporais	30
4.3	Detecção de Anomalias Não-Supervisionadas	31
4.4	Detecção de Anomalias com Aprendizado Profundo	32

5	Abordagem para Detecção de Anomalias em Imagens de Satélite	36
5.1	Introdução	36
5.2	Critério para Determinação de Anomalia	39
6	Resultados	42
6.1	Conjunto de Dados	42
6.2	Limiar para Detecção de Anomalias	43
6.3	Classificação de Anomalias	46
6.4	Treinamento e Teste do Modelo	47
6.5	Resultados e Validação	49
6.5.1	Segmentação em Imagens de Satélite	49
6.5.2	Identificação e Classificação de Anomalias	50
7	Considerações Finais	57
7.1	Conclusões	57
7.2	Trabalhos Futuros	58
A	Arquitetura da Rede Convolutacional U-net	69
B	Algoritmo POT	71

Lista de Símbolos

Adam - *Adaptive Moment Estimation*

ANN - *Artificial Neural Network*

AVIRIS - *Airbone Visible/Infrared Image Spectrometer*

AOI - *Area of interest*

AUC - *Area Under Curve*

CNN - *Convolutional Neural Network*

DBN - *Deep Believe Network*

EFB - *Exclusive Feature Bundling*

EVD - *Extreme Value Distribution*

EVT - *Extreme Value Threory*

FCN - *Fully Convolutional Networks*

GBDT - *Gradient boosting decision tree*

GMM - *Gaussian Mixture Model*

GPD - *General Pareto Distribution*

GOSS - *Gradient-based One-side Sampling*

LoF - *Local Outlier Factor*

LULC - *Land Use and Land Cover*

MODIS - *Moderate Resolution Imaging Spectroradiometer*

MUDS - *Multi Temporal Urban Development SpaceNet*

POT - *Peaks-Over-Threshold*

ReLU - *Rectified Linear Unit*

Lista de Figuras

3.1	Espectro Eletromagnético.	11
3.2	Mudanças sazonais em imagens de satélite.	12
3.3	Skip Connection.	18
3.4	Arquitetura U-net	20
3.5	Cauda da função de distribuição F	26
5.1	Abordagem proposta.	39
5.2	Limiar de Detecção de Anomalia.	40
6.1	Pontos de anomalia para a característica 02 das 1024 características monito- radas.	44
6.2	Superfície terrestre no mês de Outubro de 2019.	45
6.3	Superfície terrestre no mês de Novembro de 2019.	45
6.4	Inferência da rede U-net.	49
6.5	Desempenho do treinamento da rede convolucional.	50
6.6	Pontos de anomalia ampliados para a característica 02.	51
6.7	Curva ROC.	53
6.8	Precisão x Revocação.	53
A.1	Arquitetura U-net.	70

Lista de Tabelas

3.1	Estatísticas do Conjunto de dados SN7.	15
3.2	Relação da distribuição com γ	25
4.1	Sumário dos resultados da Literatura Analisada.	34
6.1	Hiperparâmetros U-net.	49
6.2	Matriz de Confusão.	52
6.3	Relatório de Classificação.	52
6.4	Métricas do Modelo.	55
6.5	Sumário dos resultados da Literatura Analisada e resultados do método pro- posto nesta pesquisa.	56

Capítulo 1

Introdução

O reconhecimento de padrões de crescimento urbano em imagens de satélite apresenta aplicações que variam desde o entendimento da dinâmica de urbanização, até a inferência da expansão urbana futura. Tanto a disponibilidade de inventários globais de uso da superfície terrestre, baseados em sensoriamento remoto, quanto os avanços nos métodos de aprendizado profundo, oferecem oportunidades para impulsionar o estado da arte dos modelos existentes para esse fim. O crescimento urbano é um processo espacial e demográfico e refere-se a importância crescente das cidades como locais de concentrações da população numa economia ou sociedade particular. Logo, o crescimento populacional é um dos principais fatores responsáveis pela expansão urbana. Este crescimento está diretamente ligado a fatores de atração, de ordem econômica, em especial a oferta de empregos. O aumento das atividades econômicas no meio urbano gera aglomeração humana na cidade que, por sua vez, provoca tanto mudanças na intensidade de ocupação do espaço urbano, como a expansão física da cidade.

O processamento de imagens de sensoriamento remoto atraiu atenção generalizada devido ao seu valor especial no extenso cenário de aplicativos. Uma imagem de satélite é uma importante ferramenta de previsão do tempo, além da sua capacidade de indicar a evolução dos sistemas climáticos. Outro importante fator é sua utilização como uma fonte de informação útil na elaboração de previsões meteorológicas precisas.

Imagens de sensoriamento remoto conseguem fornecer quantidades massivas de dados, permitindo acompanhar a evolução dos sistemas climáticos em larga escala. Por meios da captura dessas imagens, é possível tomar decisões e até tornar as previsões climatológicas

mais precisas. Além disso, a utilização dessas imagens tem alcançado novos desafios, sua análise levanta questionamentos exclusivos que se traduzem em novas e desafiadoras questões científicas como as seguintes:

- Dados de sensoriamento remoto geralmente são multimodais. Podemos citar dados provenientes dos sensores ópticos e dos radares de abertura sintética (SAR — *Synthetic Aperture Radar*), nesses sensores, a geometria da imagem e o conteúdo são completamente diferentes, visto que são transferidos por diversas bandas do espectro eletromagnético. Antes da extração conjunta das informações, uma etapa crucial é desenvolver novas arquiteturas para combinar imagens obtidas de diferentes perspectivas e até diferentes modos. Além das decisões de fusão convencionais, outra alternativa seria investigar a aplicação da transferência de conhecimento de redes neurais previamente treinadas para outros modos de imagem.
- O georreferenciamento presente nos dados de sensoriamento remoto garante que cada pixel corresponda a uma coordenada espacial. O objetivo é promover a fusão das informações do pixel com outras fontes de dados, como camadas de sistemas de informações geográficas (GIS — *Geographic Information System*), imagens de mídia social com marcação geográfica ou simplesmente outros sensores (multi espectrais, hiperespectrais, SAR). Assim é possível combinar informações de modalidades de dados não tradicionais, além de possibilitar a criação de novos aplicativos, como localização de imagens, serviços baseados em localização ou realidade aumentada.
- Programas de observação terrestre, como o *Copernicus*¹, garantem a aquisição contínua de dados por décadas, assim o fator temporal é uma característica intrínseca em dados oriundos do sensoriamento remoto. A exemplo disso existe o programa *Sentinel*², que mapeia toda a superfície terrestre a cada cinco dias, bem como o programa *Spacenet*³, que oferece acesso a dados geoespaciais de alta qualidade para desenvolvedores, pesquisadores e startups. Assim, o fator temporal presente em dados de sensoriamento remoto apresenta maiores possibilidades na análise de imagens, devido à amplitude de informações presentes da análise de uma única imagem isoladamente

¹<https://www.copernicus.eu/en> — Acesso em: 21 mar 2022

²<http://www.engesat.com.br/sentinel-2/> — Acesso em: 21 mar 2022

³<https://spacenet.ai/> — Acesso em: 21 mar 2022

para o processamento de imagens em séries temporais. Novas arquiteturas de redes neurais artificiais podem ser construídas para otimizar o uso de dados temporais em combinação com informações espaciais e espectrais.

- O sensoriamento remoto visa também a recuperação de quantidades geofísicas ou bioquímicas, em vez de apenas rastrear ou identificar objetos. Isso inclui níveis de movimento de massa, composição mineral do solo, elementos da água, níveis de biomassa, altitude, concentrações de gases e traços ambientais.

Os sensores de observação da superfície terrestre fornecem dados em intervalos diários ou semanais. Os crescentes recursos temporais dos sensores atuais permitem o uso temporal, além de recursos espectrais e espaciais. No entanto, a maioria das abordagens de uso e classificação da cobertura da superfície terrestre (*LULC - Land Use and Land Cover*), é projetada para observações sem nuvens e unitemporais.

Tipicamente, a tarefa de mapeamento da superfície terrestre, é uma atividade baseada em pesquisas e realizada por profissionais treinados. Como tal, essa tarefa é trabalhosa, pouco frequente, lenta e dispendiosa. Entender o desenvolvimento da superfície terrestre é um insumo importante para aplicações que variam entre planejamento urbano, zoneamento, emissão de permissões de negócios, construção, avaliação imobiliária e desenvolvimento de infraestrutura.

1.1 Problema

As imagens obtidas por sensoriamento remoto temporalmente são esparsas e não uniformes, o que dificulta a identificação de padrões, não apenas devido à presença de nuvens, como também à sua ampla área de observação e riqueza de detalhes na superfície terrestre. A implantação de novos satélites e sensores pode resolver esses problemas, mas possuem um alto custo. Além disso, há um atraso considerável na transmissão de imagens de satélite de volta à terra, devido à limitação da largura de banda e dos modos de transmissão.

De modo a identificar padrões urbanísticos e de adensamento utilizando imagens de satélite, é relevante estimar cerca do desenvolvimento terrestre em uma dada região. Devido aos crescentes recursos temporais dos sensores atuais, é possível tirar proveito do fator temporal,

bem como de recursos espectrais e espaciais das imagens de sensoriamento remoto. Embora estudos tenham sido desenvolvidos para detectar alterações na superfície terrestre utilizando séries temporais em imagens de satélite [29; 58; 67], geralmente esses estudos enfatizam a detecção de alterações anuais ou abruptas da cobertura da superfície terrestre, mas não se concentram na detecção de mudanças espaçotemporais menores e com intervalos contínuos de tempo.

A análise de imagens de sensoriamento remoto é uma área de pesquisa altamente demandada, isso é consequência do alto volume de informações incorporadas nessa categoria de imagem que é utilizada em muitas aplicações, incluindo agricultura, silvicultura, monitoramento urbano, gerenciamento de desastres, navegação de robôs, etc. Esse interesse se deve à alta qualidade da informação que pode ser derivada dessas imagens usando algoritmos de aprendizado de máquina adequados. Os avanços em sensores de imagem e tecnologias de satélite conseguem fornecer imagens de altíssima resolução de plataformas aéreas e de satélite. Essas imagens, incluindo dados de sensores aéreos e espaciais, variam em resoluções espaciais, radiométricas, espectrais e temporais. A necessidade do usuário determina a natureza da classificação e a escala da área de estudo, afetando assim a seleção de resolução espacial adequada de dados de sensoriamento remoto. Para acomodar esse interesse crescente, há uma demanda contínua por métodos algorítmicos sofisticados para o mapeamento preciso de imagens de sensoriamento remoto de alta resolução.

Imagens de satélite são caracterizadas por sua capacidade de registrar informações detalhadas sobre a distribuição espectral da luz recebida. Sensores de imagem multiespectral normalmente medem a energia da luz recebida em dezenas ou centenas de bandas espectrais estreitas em cada posição espacial na imagem, de modo que cada pixel em uma imagem de satélite possa ser representado como um vetor de alta dimensão contendo o espectro amostrado. Uma vez que diferentes texturas exibem diferentes assinaturas digitais, a imagem de satélite é uma tecnologia adequada para inúmeras aplicações de sensoriamento remoto, incluindo detecção de anomalias ou novidades. Quando nenhuma informação sobre a assinatura digital de determinado conjunto de pixels está disponível, uma abordagem popular para a detecção de anomalias é procurar objetos ou regiões que se desviem das características de textura e superfície típicas na imagem. Esta abordagem é comumente referida como detecção de anomalias [71] e está relacionada ao que é frequentemente chamado detecção

de *outlier* na estatística.

Anomalias são padrões nos dados que não estão em conformidade com o comportamento esperado, também chamados de *outliers*, exceções, peculiaridades ou surpresas [13]. Regiões anômalas capturadas por imagens de satélite podem revelar alterações ou distúrbios na cobertura da superfície terrestre que ocorrem em todo o mundo.

Séries temporais no sensoriamento remoto possuem um potencial significativo para monitorar as mudanças nessas regiões, em escalas regional e global, devido à observação sinóptica e frequente [33]. A identificação de regiões que apresentam anomalias em séries temporais é importante para os pesquisadores estudarem os processos dinâmicos de mudanças na cobertura do solo, bem como para os gerentes de recursos monitorarem a dinâmica terrestre em grandes áreas, especialmente onde o acesso é difícil ou perigoso [38]. Além disso, séries temporais de imagens de satélite registram continuamente o status variável da cobertura do solo, incluindo os padrões sazonais conduzidos por interações anuais de temperatura e precipitação, bem como as anomalias causadas por fatores naturais ou antropogênicos [79].

Considerando os estudos anteriormente citados, abre-se espaço para novas pesquisas e para a criação de novas abordagens, que sejam ainda mais eficientes na identificação de padrões e anomalias na superfície terrestre, aproveitando todos os recursos existentes em imagens de satélite. Logo, nesta pesquisa será abordado o problema de identificação de padrões de crescimento de urbano, por intermédio do sensoriamento remoto, utilizando séries temporais de imagens de satélite, aplicando uma técnica probabilística paramétrica de detecção de anomalias e redes neurais convolucionais.

1.2 Relevância

A urbanização tem se mostrado uma tendência fundamental dos últimos séculos e uma força essencial no desenvolvimento do mundo moderno. Somente nos próximos anos a população mundial deve crescer de 7,7 bilhões em 2019 para 8,5 bilhões em 2030 (aumento de 10%), e ainda para 9,7 bilhões em 2050 (26%) e para 10,9 bilhões em 2100 (42%) [57]. A menos que guiada por políticas e planejamento apropriado, a velocidade e a escala da urbanização nos países em desenvolvimento pode resultar em assentamentos precários [43], falha no fornecimento de serviços essenciais ao público [18] e, conseqüentemente, cidades incapazes

de competir em um ambiente globalizado [76].

Assim, a evolução da distribuição espacial do ambiente urbano construído e sua dependência de fatores socioeconômicos são de alta relevância para o planejamento urbano e para os investimentos em infraestrutura. No entanto, as ferramentas para estudar a forma urbana em escala global até o momento foram amplamente baseadas em modelos simplificados [64], que não conseguem capturar a complexidade observada em dados reais e produzir previsões realistas da dinâmica espaço-temporal do uso da superfície terrestre. Outras limitações dos modelos tradicionais de desenvolvimento urbano são: i) sua incapacidade de alavancar efetivamente a vasta e crescente quantidade de dados observacionais sobre áreas urbanas construídas; e ii) sua dependência de dados socioeconômicos locais detalhados que não estão disponíveis na grande maioria das cidades.

Nesse contexto, a identificação de padrões utilizando imagens de sensoriamento remoto, pode trazer benefícios importantes. Tanto o conhecimento dos recursos naturais quanto das estruturas criadas pelo homem tornam-se fundamentais na tomada de decisão. Diversas variáveis definem a saúde do ecossistema urbano, ao apresentarem padrões nas mudanças de longo e curto prazo. É possível, então, utilizar tais variáveis como indicadores precoces de desastres naturais, mudanças climáticas e desenvolvimento urbano total.

Essas mudanças podem ter impactos socioeconômicos profundos, daí a importância de se desenvolver recursos para identificação, análise e resposta a essas mudanças em tempo hábil. Para entender completamente a dinâmica terrestre, é preciso conseguir analisar vários conjuntos de dados de satélites, sensores de solo e modelos. Uma imagem de satélite capturada possui uma observação diferente ou capacidade preditiva particular. Portanto, uma análise global sobre uma combinação de imagens, fornece melhores resultados que estudar esses recursos de forma isolada.

1.3 Escopo

Esta pesquisa se propõe a identificar padrões por meio de uma abordagem probabilística paramétrica de detecção de anomalias, baseando-se na teoria dos valores extremos. Logo, no desenvolvimento desse estudo, utilizamos dados de imagens multi-espectrais fornecidos pelo programa *SpaceNet*. O *SpaceNet* oferece acesso a dados geoespaciais de alta qualidade para

desenvolvedores, pesquisadores e *startups*. Antes do *SpaceNet*, os pesquisadores de visão computacional tinham opções mínimas para obter imagens de satélite gratuitas, rotuladas com precisão e de alta resolução. O programa *SpaceNet* se concentra em quatro pilares-chave de código aberto: dados, desafios, algoritmos e ferramentas [1].

1.4 Contribuições da Pesquisa

A pesquisa de mestrado apresentada possui contribuições interdisciplinares, envolvendo as áreas de análise de anomalias, aprendizado profundo e identificação de padrões de crescimento urbano. A seguir, apontamos as contribuições da realização deste estudo:

1. Aplicação de uma abordagem probabilística paramétrica de detecção de padrões, aplicando a teoria dos valores extremos em características extraídas de imagens de satélite;
2. Identificar a capacidade preditiva da abordagem em identificar anomalias em imagens de satélite.
3. Identificar o quanto redes convolucionais de segmentação podem evidenciar o fator discriminativo de anomalias em imagens de satélite.
4. Mensurar o quão precisa uma abordagem probabilística paramétrica pode identificar anomalias a partir de características extraídas de imagens de satélite.
5. Contribuição com a pesquisa na identificação de soluções para tomada de decisão no que se refere ao desenvolvimento urbano;
6. Mapeamento e análise de mudanças temporais em regiões específicas da superfície terrestre no período de 2018 a 2020;
7. Desenvolvimento de um método para a detecção de padrões de alterações na superfície terrestre utilizando: imagens de satélites, teoria dos valores extremos e redes neurais convolucionais.

Capítulo 2

Objetivos

Nesta seção são apresentados os objetivos gerais e específicos para esta a pesquisa. Esses objetivos delineiam e norteiam nossa pesquisa, além de explicitar as contribuições esperadas.

2.1 Objetivo Geral

Esta pesquisa visa, identificar padrões de crescimento urbano na superfície terrestre, aplicando visão computacional as imagens de satélite, a partir da teoria dos valores extremos, aplicada às representações de características de imagens extraídas por uma arquitetura de rede convolucional (U-Net).

2.2 Objetivos Específicos

De forma a alcançar o objetivo geral definido na seção anterior, o presente estudo pode ser subdividido nos seguintes objetivos específicos:

- Objetivo Específico I - Evidenciar o fator discriminativo das representações de características geradas por uma arquitetura de rede convolucional (U-Net).
 1. Mais especificamente pretende-se extrair características de imagens de satélite utilizando uma rede convolucional e utilizar as representações do espaço latente para identificar padrões de anomalias na superfície terrestre.

- Objetivo Específico II - Mensurar o potencial de precisão na identificação de padrões anômalos utilizando uma abordagem probabilística paramétrica
 1. Mais especificamente pretende-se mensurar e validar a precisão com que uma abordagem de detecção de anomalias, que utiliza a teoria dos valores extremos, consegue identificar e possivelmente inferir padrões de alterações na superfície terrestre a partir de representações de características de imagens de satélite.

Capítulo 3

Fundamentação Teórica

3.1 Introdução

Nesta seção são apresentados conceitos e fundamentos de temas abordados durante a pesquisa. Inicia-se com a definição de sensoriamento remoto, apresenta-se em seguida uma explanação sobre redes convolucionais com ênfase na arquitetura de rede utilizada nesta pesquisa. Por fim, é apresentado o conceito de detecção de anomalias.

3.2 Sensoriamento Remoto

O sensoriamento remoto, em geral, refere-se a qualquer técnica pela qual o espaço do objeto disposto na superfície terrestre possa ser observado [46]. Tradicionalmente, o termo sensoriamento remoto era utilizado apenas para plataformas aéreas e via satélite, que adquirem dados normalmente por sensores ópticos e de radar [11; 66]. No entanto, qualquer método de aquisição de imagens e dados espaciais, incluindo levantamento aéreo e fotogrametria é considerado sensoriamento remoto, embora os termos sensoriamento remoto via satélite e mapeamento móvel, para plataformas terrestres, ainda são usadas ocasionalmente [54]. Além do agrupamento primário, baseado em plataformas, a tecnologia de sensoriamento remoto é caracterizada pelo sensor de imagem utilizado, como uma câmera aérea de grande formato, e pela área de aplicação, como a fotogrametria de curto alcance [23; 53]. Posteriormente, o termo sensoriamento remoto passou a ser utilizado de maneira inclusiva para incluir todas as tecnologias tradicionais de aquisição de dados.

O objetivo de qualquer tecnologia de sensoriamento remoto, é fornecer a observação de algum parâmetro físico em um quadro de mapeamento, limitado a um determinado momento ou período. O espaço físico terrestre é vasto e rico em objetos tornando-o amplamente definido e, a partir da forma da terra, é possível incluir objetos artificiais, vegetação, além de parâmetros atmosféricos.

Os sensores de imagem são o componente principal de qualquer sistema de sensoriamento remoto e vêm com uma ampla variedade de resoluções espaciais, temporais e espectrais. A Figura 3.1 apresenta o espectro eletromagnético subdividido entre as bandas de frequência mais importantes usadas no sensoriamento remoto, especialmente em sensores hiper-espectrais.

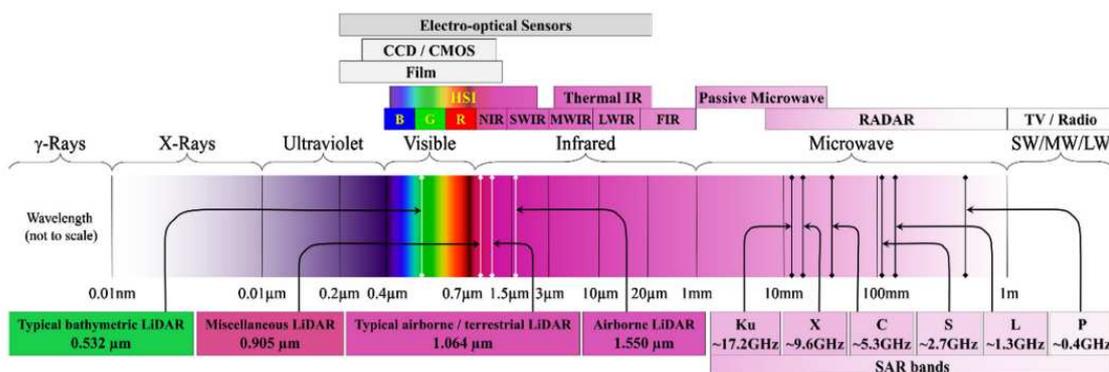


Figura 3.1: Espectro Eletromagnético.
Fonte: [74]

Devido à quantidade de bandas disponíveis no espectro eletromagnético, a análise de séries temporais de imagens de satélite é uma atividade desafiadora no campo de visão computacional, com muitas aplicações de desenvolvimento humano. Além de sua relevância para resposta a desastres, preparação para doenças e monitoramento ambiental, a análise de séries temporais de imagens de satélite apresenta desafios técnicos únicos, muitas vezes não resolvidos pelos métodos existentes.

Até o momento dessa pesquisa, poucos conjuntos de dados existentes ofereceram uma série temporal de imagens de satélite contemplando intervalos temporais. Vários trabalhos anteriores estudaram a extração de edifícios a partir de imagens de satélite ([75], [19], [80], [68]), mas esses conjuntos de dados eram estáticos. A comparação mais próxima é o desafio *xView2* e o conjunto de dados *xBD* [30], que examinou danos em edifícios em pares de imagens de satélite adquiridos antes e após desastres naturais em menos de 20 locais, no entanto,

essa tarefa não aborda as complexidades e oportunidades apresentadas pela análise de dados de séries temporais contemplando longos intervalos temporais, como vegetação sazonal e mudanças de iluminação, ou rastreamento consistente de objetos em escala global. Outras competições exploraram dados de séries temporais na forma de vídeo, por exemplo, tarefas de detecção de objetos [20] e segmentação [10]. Existem várias diferenças significativas entre esses desafios e a tarefa descrita no presente trabalho. Em primeiro lugar, a variação quadro a quadro é muito pequena em conjuntos de dados de vídeo.

Ao comparar o tamanho das instâncias anotadas no conjunto de dados COCO [49], os objetos menores presentes no conjunto de dados, intrinsecamente fornecem menos informações, pois compreendem menos pixels, tornando sua identificação uma tarefa mais difícil. Por outro lado, a aparência das imagens de satélite pode mudar drasticamente de mês para mês devido a diferenças de clima, iluminação e efeitos sazonais no solo, conforme mostrado na Figura 3.2.

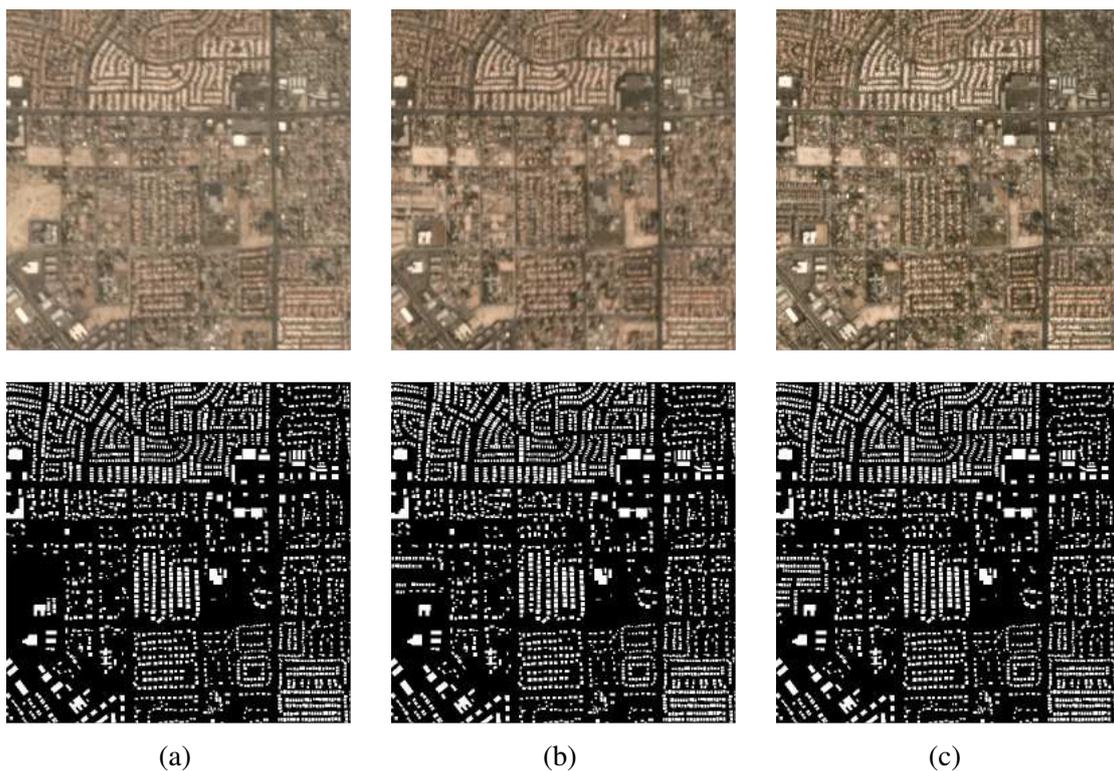


Figura 3.2: Mudanças sazonais em imagens de satélite.

Fonte: Pesquisa do Autor.

Além dos desafios mencionados anteriormente, a atividade de identificar mudanças ou anomalias a partir de imagens de satélite considera que não existe conhecimento prévio do

domínio e que esse conhecimento pode ainda depender do comportamento do domínio no tempo, ou seja, alterações nas estatísticas dos pixels ao longo do tempo. Logo, um dos principais desafios de identificar mudanças ou padrões em imagens de sensoriamento remoto é rotular corretamente essas imagens com suas classes semânticas correspondentes e em conformidade com seu conteúdo, por exemplo, categorizando uma imagem de sensoriamento remoto de área urbana em residencial, comercial ou industrial.

De modo geral, uma imagem de sensoriamento remoto contém uma variedade de objetos no solo. Por exemplo, estradas, árvores e edifícios podem ser incluídos em uma cena industrial. Diferente de um processo de classificação aplicada a objetos, a identificação de mudanças ou cenários anômalos em imagens de satélite é um problema consideravelmente desafiador devido a variações e distribuições espaciais complexas de objetos de solo existentes nas cenas.

Os desafios da classificação de cenas de imagens de sensoriamento remoto podem apresentar:

- Grande diversidade intraclasse (árvores, campos, construções, etc.);
- Alta similaridade interclasse (também conhecida como baixa separabilidade entre classes);
- Alta variação na escala de objetos/cenas;
- Coexistência de múltiplos objetos no solo, conforme mostrado nas Figuras 3.3, 3.4 e 3.4.

Em termos de diversidade dentro de uma classe específica, o desafio origina-se principalmente das grandes variações nas aparências dos objetos de solo dentro da mesma classe semântica. Objetos de solo geralmente variam em estilo, forma, escala e distribuição, o que dificulta classificar corretamente as imagens da cena. Além disso, quando as plataformas aéreas ou espaciais captam imagens de sensoriamento remoto, pode haver grandes diferenças na cor e na intensidade da radiação aparecendo dentro da mesma classe semântica, devido às condições de imagem, que podem ser influenciadas por fatores como clima, nuvem, névoa, etc. As variações na iluminação da cena também podem causar diversidade na classe.

Para a similaridade entre classes, o desafio é causado principalmente pela presença dos mesmos objetos em diferentes classes de cena ou pela alta sobreposição semântica entre as essas categorias de classes. Além disso, as definições ambíguas das classes presentes em uma imagem degeneram a dissimilaridade entre classes. Algumas cenas complexas também são semelhantes entre si em termos de conteúdo visual. Portanto, pode ser extremamente difícil distinguir essas classes de cena.

Além disso, devido à distribuição complexa e diversa dos objetos na superfície terrestre bem como à ampla perspectiva panorâmica dos equipamentos de imagem de sensoriamento remoto, é bastante comum que vários objetos no solo apareçam em uma única imagem. Diante da situação, é difícil para a classificação de imagens de sensoriamento remoto estabelecer uma compreensão profunda embasada na categoria de controle utilizado para a captura das imagens.

Como mencionado anteriormente, para a pesquisa proposta, foram coletadas imagens de sensoriamento remoto, temporalmente organizadas, do conjunto de dados de desenvolvimento urbano multitemporal *SpaceNet* (*MUDS - Multi-Temporal Urban Development SpaceNet*), que consiste em 101 sequências rotuladas de imagens de satélite coletadas pela constelação Dove do *Planet Labs* entre 2017 e 2020, com rótulos e marcações para as construções presentes em cada imagem. As sequências de imagens são amostradas em 101 áreas distintas de interesse em todo o mundo, cobrindo seis continentes. Esses locais foram selecionados por serem geograficamente diversos e apresentarem mudanças expressivas na urbanização ao longo de um período de dois anos. O conjunto de dados MUDS é de código aberto sob uma licença CC-BY4.0 *ShareAlike International*¹ para incentivar o uso amplo. Esse conjunto de dados pode ser potencialmente útil para muitas outras tarefas de visão computacional geoespacial, além de poder, ser facilmente combinado ou aumentado com qualquer outra camada de dados disponível. Os próprios rótulos também podem ser aplicados a qualquer outro mosaico de imagem de sensoriamento remoto, como radar de abertura óptica ou sintética de alta resolução.

O tamanho de cada imagem é de 1024×1024 pixels, correspondendo a 18 km^2 , e a área total das imagens no conjunto de dados é de 41.250 km^2 . No entanto, devido a limitações técnicas (memória, espaço em disco e capacidade de processamento), para essa pesquisa as

¹<https://registry.opendata.aws/spacenet/>

imagens foram redimensionadas para 512 x 512 pixels. Estatísticas adicionais são apresentadas na Tabela 3.1 ou spacenet.ai². A série temporal contemplando intervalos temporais de 18 a até 26 meses, a depender da área de interesse. Este período captura várias estações e condições atmosféricas, bem como o início e a conclusão de vários projetos de construção (Figuras 3.2 a 3.4). Além disso, imagens contendo uma quantidade excessiva de nuvens ou neblina foram totalmente excluídas do conjunto de dados, causando pequenas lacunas temporais em algumas séries temporais.

Tabela 3.1: Estatísticas do Conjunto de dados SN7.

Categoria	Valor
Nº de áreas de interesse	101
Nº de observações	2389
Nº de construções	11.080,000
Área observada total (km ²)	41.000
Média de construções por observação	4.700
Média de área construída (m ²)	190
Média da distância da amostra do solo (m)	4,0

3.3 Redes Neurais Convolucionais

A Redes Neurais Convolucionais (*CNN - Convolutional Neural Networks*) obtiveram resultados inovadores na última década em uma variedade de campos relacionados ao reconhecimento de padrões; do processamento de imagem ao reconhecimento de voz. Um dos aspectos mais benéficos das CNNs é reduzir o número de parâmetros das redes neurais artificiais (*ANN - Artificial Neural Networks*).

As CNNs efetuam um processo denominado convolução, o qual funciona deslizando “filtros” sobre uma imagem. Cada filtro tem pesos diferentes, portanto, são treinados para reconhecer uma característica específica de uma imagem. Quanto mais filtros uma rede tem — ou mais profunda uma rede é — mais características ela pode extrair de uma imagem,

²<https://spacenet.ai>

portanto, mais padrões complexos ela pode aprender com o propósito de informar sua decisão de classificação final. No entanto, dado que cada filtro é representado por um conjunto de pesos a serem aprendidos, ter muitos filtros do mesmo tamanho da imagem de entrada original torna o treinamento de um modelo bastante caro computacionalmente.

Conforme descrito no parágrafo anterior, as características extraídas por uma CNN contêm informações semânticas de alto nível sendo utilizadas para prever, comumente, classes de imagens inteiras. A atividade de segmentação de imagens também usa essas características de alto nível extraídas por um processo de convolução, mas prevê classes em nível de pixel, o que leva ao seguinte problema: na execução de um processo de convolução, os mapas de características são posteriormente redimensionados para uma resolução menor com nível de abstração semântica crescente. Portanto, tais mapas fornecem informações de localização menos precisas, o que é necessário para uma previsão precisa de pixels.

Além disso, para prever a classe de um único pixel, o relacionamento contextual é importante. Essas informações podem ser encontradas em uma faixa de distâncias grandes e pequenas ao redor do pixel. A informação contextual depende do tamanho e continuidade do segmento semântico uniforme ao qual o pixel pertence, bem como da quantidade e densidade de segmentos vizinhos de outras classes e também do fundo da imagem. A segmentação difere da detecção de objetos porque funciona no nível do pixel para determinar os contornos dos objetos em uma imagem [50]. Portanto, a segmentação de imagens pode ser vista como um problema de contexto multiescala, mesmo quando a previsão é feita em pixels únicos. Os avanços na segmentação de imagens estão focando fortemente na solução deste problema ao explorar características de diferentes estágios na rede, preservando, ou reconstruindo a resolução mais alta resolução durante a extração e previsão de características.

Na atividade de segmentação, uma CNN objetiva produzir um mapa de classificações em pixels. Portanto, afim de construir essa saída, é ideal que a rede tenha um canal de saída codificado para cada um dos rótulos de classe possíveis. Esses mapas são então agrupados por meio de uma operação *argmax*, operação essa que considera os argumentos dos máximos em cada pixel. Assim, a atividade de segmentação de imagens visa a rotular cada pixel de uma imagem com uma classe correspondente do domínio que está sendo representado.

Ao contrário da classificação de imagens, em que uma imagem inteira é classificada de acordo com um rótulo, a atividade de segmentação de imagens envolve a detecção e a

classificação de objetos individuais contidos na imagem. A ideia geral é assumir que grandes grupos de pixels contíguos tendem a ser associados ao mesmo rótulo, além de descrever as relações probabilísticas entre pixels vizinhos [61].

Cabe ressaltar ainda que a complexidade do processo de segmentação reside no fato de que a saída deve ser alinhada com a imagem de entrada logo seguir a mesma arquitetura de redução da resolução utilizada em um processo de classificação traz consigo uma perda na qualidade dos mapas de características processadas. Apesar da redução da resolução utilizada em processos de classificação promover a eficiência computacional, a perda ocasionada no mapa de características decorrente do processo de convolução compromete a saída do processamento, dado que o tamanho e a localização das áreas de cada classe devem ser preservados. Além disso, a rede também precisa ser suficientemente profunda para aprender representações detalhadas de cada uma das classes de entrada, de forma que seja possível distingui-las.

As CNNs padrão, comumente utilizadas no processo de classificação de imagens, em sua arquitetura convencional, têm sua primeira camada totalmente conectada, ou seja, essa camada tem as mesmas dimensões da imagem de entrada; isso significa que as dimensões da entrada na primeira camada devem estar sempre alinhadas com as dimensões da imagem utilizada na entrada. Assim, a rede não se torna apenas inflexível para entradas de diferentes tamanhos, mas também utiliza informações globais (ou seja, informações de toda a imagem) para tomar sua decisão de classificação, o que não faz sentido no contexto de segmentação de imagens, em que o objetivo é atribuir rótulos de classe diferentes a regiões diferentes da imagem. Outro ponto é que as camadas convolucionais, por outro lado, são menores que a imagem de entrada para poderem deslizar sobre ela e assim operar em regiões mais específicas da entrada.

Em 2014, Long et al. [50] introduziu as Redes Convolucionais Totalmente Conectadas (*FCN - Fully Convolutional Networks*) para segmentação semântica. Diferente das CNNs padrão, as FCNs são dotadas exclusivamente de camadas convolucionais assim, as FCNs substituem as camadas totalmente conectadas presentes nas CNNs padrão, por camadas convolucionais com grandes campos receptivos.

À medida que avançamos na arquitetura de uma FCN, é possível observar que o tamanho das camadas torna-se decrescente devido à necessidade de aprendizagem de características

mais específicas (ou abstratas) de uma maneira computacionalmente eficiente — um processo conhecido como *downsampling*.

Uma técnica conhecida como *upsampling* foi utilizada por [50] para expandir o mapa de características de volta ao mesmo tamanho da imagem de entrada, e em seguida, um processo denominado *Skip Connection* é utilizado para aumentar sua resolução, porém uma técnica que usa a convolução transposta para aumentar a amostragem de mapas de características de uma forma que estas possam ser aprendidas, foi proposta em [50].

Essa técnica de convolução transposta funciona de maneira diferente da convolução tradicional: os pesos do filtro são todos multiplicados pelo valor escalar do pixel de entrada sobre o qual ele está posicionado e esses valores são projetados para o mapa de características de saída, onde as projeções do filtro no mapa de saída se sobrepõem e seus valores são adicionados.

A convolução transposta visa aumentar a amostragem do mapa de características renderizado pelas camadas de redução da amostragem da rede, de modo a traduzir sua saída imprecisa de volta para pixels que se alinham com os mesmos pixels da imagem de entrada. No entanto, simplesmente adicionar uma dessas camadas convolucionais transpostas no final das camadas de *downsampling* produz resultados espacialmente imprecisos, visto que é necessário um valor muito alto de *stride* para fazer o tamanho da saída correspondente ser idêntico ao da entrada, limitando a escala de detalhes que a camada de *upsampling* pode alcançar.

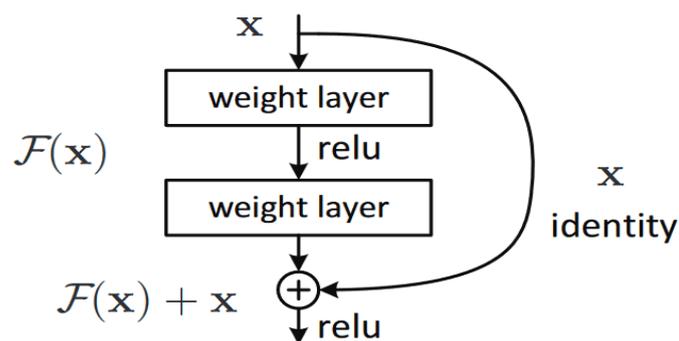


Figura 3.3: Skip Connection.

Fonte: [35].

Logo, para sanar o problema da perda de contexto espacial e a dificuldade de localização dos limiares de decisão, é possível utilizar a técnica de *Skip Connection* (Figura 3.3), que

nada mais é do que uma fusão entre camadas não adjacentes. Neste caso, tal característica é utilizado para transferir informações locais, somando mapas de características do caminho de redução da resolução (*downsampling*) com mapas de características do caminho de aumento da resolução (*upsampling*).

Intuitivamente isso faz sentido, visto que para cada camada que uma entrada percorre em uma CNN, é aplicada uma redução da resolução dessa entrada (processo de *downsampling*), as informações globais então acabam se perdendo. Tal perda é ainda acentuada logo que o processo inverso é aplicado (processo *upsampling*), ou seja, quando a rede reconstrói a entrada na saída. Assim, a medida que a imagem é reconstruída, áreas específicas da imagem passam a ter seus mapas de características mais grosseiros e imprecisos. Assim as *Skip Connections* atenuam essa perda e funcionam tal qual uma ponte a partir das camadas finais da rede, recuperando as informações armazenadas nos mapas de características de alta resolução aprendidos nas camadas anteriores da rede.

Uma aplicação das *Skip Connection* para segmentação pode ser observada na arquitetura **U-net** proposta pela primeira vez por [63]. A arquitetura **U-net** foi proposta como um modelo FCN para uso na segmentação de imagens biomédicas. A arquitetura consiste em um caminho de contração para capturar o contexto e um caminho de expansão simétrico que permite uma localização precisa, resultando em uma arquitetura em forma de **U** (Figura 3.4).

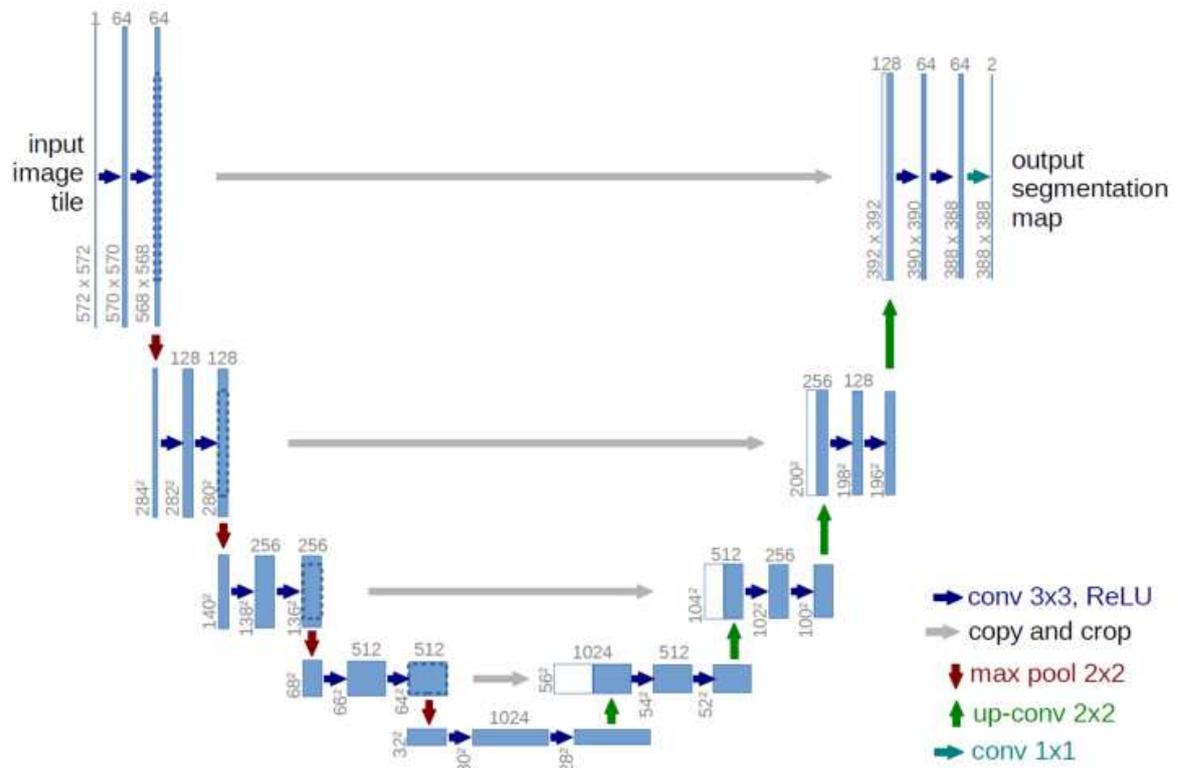


Figura 3.4: Arquitetura U-net
Fonte: [63]

É possível observar que esta arquitetura apresenta 4 *Skip Connections* — após cada convolução de transposição no caminho de *upsampling*, o mapa de características resultante é concatenado com um do caminho de *downsampling*. Além disso, vemos que os mapas de características no caminho de *upsampling* têm um número maior de canais do que na arquitetura FCN original visando a passar mais informações de contexto para camadas de resolução mais alta.

A arquitetura U-net também alcança melhor resolução nos limites de segmentação ao pré-computar um mapa de pixels para cada instância de treinamento. A função usada para calcular o mapa coloca pesos mais altos nos pixels ao longo dos limites de segmentação. Esses pesos são então fatorados na função de perda na etapa de treinamento, de modo que os pixels componham os limites ou contornos da imagem tenham maior prioridade para serem classificados corretamente.

3.4 Detecção de Anomalias

Detectar novos eventos é uma habilidade importante para qualquer esquema de classificação. Dado o fato de nunca podermos treinar um sistema de aprendizado de máquina em todas as classes de objetos possíveis cujos dados o sistema provavelmente encontrará, torna-se importante que seja possível diferenciar entre informações de objetos conhecidos e desconhecidos durante a etapa de teste. Assim, a detecção de anomalias é a identificação de dados ou sinais, novos ou desconhecidos, dos quais um sistema de aprendizado de máquina não tem conhecimento durante o treinamento [52].

Tanto a importância prática da detecção de anomalias, quanto sua natureza desafiadora permitiram que estudos propusessem abordagens para sua identificação e classificação [60]. Tais técnicas são geralmente utilizadas em conjuntos de dados onde um número muito grande de exemplos da “condição padrão” — também conhecida como exemplos positivos — está disponível e onde há dados insuficientes para explicar as “anormalidades” — também conhecidas como exemplos negativos. A presença dessas “anormalidades” traz consigo uma consequência, sendo a existência de uma grande variedade de possíveis tipos “anormais”, alguns dos quais podem não ser identificados a priori, tornando os esquemas tradicionais de classificação multi-classe inadequados para tais aplicações.

Cabe ressaltar que dois sinônimos intercambiáveis de detecção de novidades [8; 72] frequentemente usados na literatura são detecção de anomalias e detecção de *outliers* [62]. Os diferentes termos se originam de diferentes domínios de aplicação aos quais a classificação de uma classe pode ser aplicada e não há definição universalmente aceita. Anomalias e *outliers* são dois termos usados mais comumente no contexto da detecção de anomalias; às vezes de forma intercambiável [13]. Um *outlier* pode ser definido como um ponto de dados que "parece ser inconsistente com o restante desse conjunto de dados [de treinamento]" [6].

A detecção de novidades ou anomalias propõe então que uma definição de normalidade seja obtida através da construção de um modelo, onde vários exemplos que descrevem dados representativos do comportamento normal do sistema sejam utilizados. Padrões previamente não vistos são então avaliados comparando-os ao modelo de normalidade, geralmente resultando em alguma categoria de classificação de anomalia. Na detecção de novidades ou anomalias, uma descrição da normalidade é aprendida se construindo um modelo com vários

exemplos representando instâncias positivas (ou seja, dados indicativos do comportamento normal do sistema). Padrões não vistos anteriormente são então testados comparando-os com o modelo de normalidade, muitas vezes resultando em alguma forma de pontuação de anomalias. A pontuação, que pode ou não ser probabilística, é normalmente comparada a um limite de decisão, e os dados do teste são então considerados “anormais” se o limite for excedido [60].

Segundo [52], existem vários pontos importantes relacionados à detecção de anomalias. Pode-se resumi-los em termos dos seguintes princípios.

- *Princípio da robustez e do trade-off*: um método de detecção de anomalias deve conseguir ter um desempenho robusto nos dados de teste que maximiza a exclusão de novas amostras e minimiza a exclusão de amostras conhecidas. Este *trade-off* deve ser, até certo ponto, previsível e sob controle experimental.
- *Princípio da escala uniforme de dados*: para ajudar na detecção de anomalias, deve ser possível que todos os dados de teste e treinamento, após a normalização, estejam no mesmo intervalo [70].
- *Princípio da minimização de parâmetros*: um método de detecção de anomalias deve ter como objetivo minimizar o número de parâmetros definidos pelo usuário.
- *Princípio da generalização*: o sistema deve conseguir generalizar sem confundir informações generalizadas como novas [73].
- *Princípio da independência*: o método de detecção de anomalias deve ser independente do número de características e classes disponíveis e deve mostrar desempenho razoável no contexto do conjunto de dados desequilibrados, número baixo de amostras e ruído.
- *Princípio da adaptabilidade*: um sistema que reconhece novas amostras durante o teste deve conseguir usar esta informação para retreinamento.[65].
- *Princípio da complexidade computacional*: vários aplicativos de detecção de anomalias estão *on-line* portanto, a complexidade computacional de um mecanismo de detecção de anomalias deve ser a menor possível.

Em [60] são sugeridas técnicas de detecção de anomalias organizadas de acordo com às cinco categorias gerais a seguir: (i) probabilística, (ii) baseada em distância, (iii) baseada em reconstrução, (iv) baseada em domínio e (v) técnicas de teoria da informação, conforme discutido na próxima seção.

3.4.1 Abordagem Probabilística Paramétrica

As abordagens probabilísticas para a detecção de anomalias são baseadas na estimativa de uma função de densidade de probabilidade geradora dos dados. A distribuição resultante pode então ser limitada para definir os limites da normalidade no espaço de dados e testar se uma amostra de teste vem da mesma distribuição ou não [60]. Presume-se que os dados de treinamento sejam gerados a partir de alguma distribuição de probabilidade subjacente D , que pode ser estimada usando os dados de treinamento. Essa estimativa \hat{D} geralmente representa um modelo de normalidade. Um limiar de anomalia pode então ser definido usando \hat{D} de alguma maneira, de modo a ter uma interpretação probabilística.

Em termos gerais, para estimar a função de densidade de probabilidade, existem duas abordagens principais, paramétricas e não paramétricas [21]. As abordagens paramétricas assumem que os dados normais são gerados a partir de uma distribuição paramétrica subjacente com os parâmetros $\theta \in \Theta$, sendo θ finito e a função de densidade de probabilidade $p(x, \theta)$, onde x corresponde a uma observação. Os parâmetros θ são estimados a partir dos dados de treinamento fornecidos. A forma de distribuição mais usada para variáveis contínuas é a gaussiana. Na ausência de informações prévias sobre a forma da distribuição subjacente dos dados, a distribuição Gaussiana é frequentemente usada devido às suas propriedades analíticas convenientes ao determinar a localização de um limiar de anomalia.

Assim, os parâmetros são estimados a partir dos dados de treinamento fornecidos, usando estimativas de máxima verossimilhança, para as quais existe uma solução analítica de forma fechada para uma distribuição gaussiana [52]. Formas mais complexas de distribuição de dados podem ser modeladas por modelos de mistura, como os modelos de mistura gaussiana (GMM) [9; 26].

O modelo de mistura gaussiana estima a densidade de probabilidade da classe alvo (classe normal), fornecida por um conjunto de treinamento, normalmente usando menos *kernels* do que o número de padrões no conjunto de treinamento [16]. Os parâmetros do modelo são

escolhidos maximizando-se a probabilidade logarítmica dos dados de treinamento em relação ao modelo. Isso é feito usando algoritmos de otimização, como gradientes conjugados ou técnicas de reestimação, como o algoritmo de maximização de expectativa. No entanto, o GMM sofre com a maldição da dimensionalidade [15], no sentido de que, se a dimensionalidade dos dados for alta, um número muito grande de amostras será necessário para treinar o modelo.

3.4.2 Distribuição de Valores Extremos

Um dos principais problemas na detecção de anomalias é a seleção de um limite adequado. Um método baseado em princípios para definir limites para detecção de anomalias usa a teoria dos valores extremos (*EVT - Extreme Value Theory*), definida por [59]. Logo, a teoria dos valores extremos, é um ramo da estatística que lida diretamente com eventos raros ou extremos, e seu objetivo procura modelar o comportamento assintótico de eventos que se distanciam muito da mediana de uma distribuição.

Resultados anteriores [22; 27], afirmam que sob uma condição fraca, esses eventos extremos têm o mesmo tipo de distribuição, independentemente da distribuição original.

Considere, por exemplo, o conjunto $X_m = \{x_1, x_2, x_3, \dots, x_n\}$ de n variáveis independentes e identicamente distribuídas, com $x_i \in \mathbb{R}$ extraídos de uma função de distribuição \mathcal{D} . Sendo $x_m = \max(X_m)$, o maior elemento observado em X_m , provavelmente esse evento seguirá a distribuição original, a qual será definida por γ .

Essas leis extremas são chamadas distribuições de valor extremo (*Extreme Value Distributions - EVD*), e têm a seguinte forma:

$$G_\gamma, : x \mapsto \exp\left(- (1 + \gamma x)^{\frac{-1}{\gamma}}\right), \gamma \in \mathbb{R}, 1 + \gamma x > 0. \quad (3.1)$$

em que γ representando a forma, parâmetro que governa o comportamento da cauda da distribuição limite, e x a variável reduzida dada por:

$$x = \sigma_m^{-1}(x_m - \mu_m). \quad (3.2)$$

em que σ_m, μ_m denotando respectivamente a escala e o local.

Ainda segundo [22] e [27], a cauda da distribuição pode assumir uma de três distribuições

possíveis: Gumbel, Frechet ou Weibull (Tabela 3.2).

Comportamento da cauda	Domínio	Distribuição
Cauda pesada	$\gamma > 0$	Frechet, [24]
Cauda exponencial	$\gamma = 0$	Gumbel, [31]
Limitada	$\gamma < 0$	Weibull, [27]

Tabela 3.2: Relação da distribuição com γ .

Assim, é possível calcular com precisão probabilidades sem inferir a lei inicial que pode ser realmente complexa. De fato, o teorema do limite central afirma que a média das variáveis aleatórias n independentes e identicamente distribuídas converge na distribuição para uma distribuição normal. O teorema EVT indica o mesmo resultado para o valor máximo.

3.4.3 *Peaks-Over-Threshold (POT)*

A abordagem de picos acima do limite (*POT - Peaks Over Threshold*), baseia-se no teorema de Pickands-Balkema-de Haan [5; 59] (também chamado de segundo teorema na EVT ou problema de estimação da cauda), e afirma que a partir de uma amostra aleatória $X = x_1, \dots, x_n$ de uma função de distribuição desconhecida F , é possível estimar a função de distribuição condicional F_u da variável X acima de um certo limiar u .

A Figura 3.5 ilustra o termo “cauda da função de distribuição F acima de um certo limiar u ”, que é a parte do gráfico de F à direita do ponto $(u, F(u))$, bem como o valor $F^{-1}(p)$, que traduz-se como o valor que desejamos estimar para valores de p pertencentes a um intervalo próximo a 1. Cabe ressaltar que, o limiar u é inicializado aleatoriamente, porém deve permitir uma boa precisão na aproximação, bem como ser adequado ao contexto do problema que deseja-se resolver.

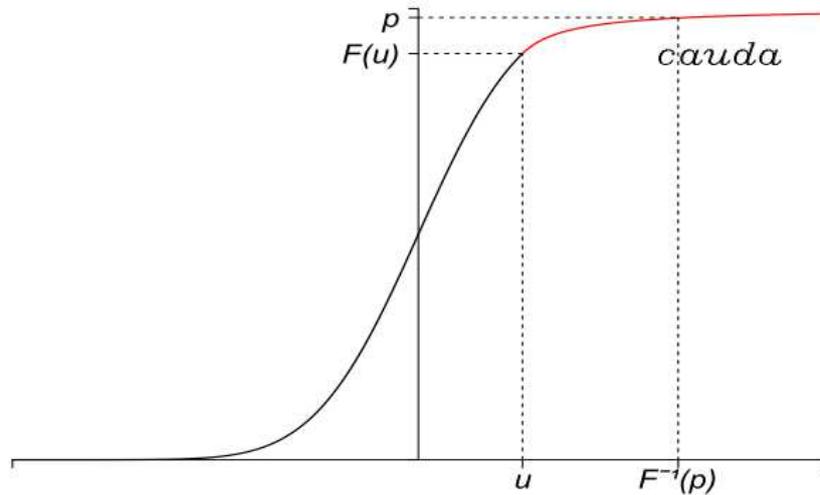


Figura 3.5: Cauda da função de distribuição F .

Fonte: [7]

Logo, de forma geral, o teorema de Pickands-Balkema-de Haan é dado por:

A função de distribuição cumulativa $F \in \mathcal{D}_\gamma^1$ se e somente se uma função σ existir, para todo $x \in \mathbb{R} \mid 1 + \gamma x > 0$:

$$\frac{F(u + \sigma(u)x)}{F(u)} \rightarrow (1 + \gamma x)^{-1/\gamma} \quad (3.3)$$

Uma visão mais clara do teorema é expresso na Equação 3.4:

$$F_u(x) = P(X - u > x \mid X > u) \left(1 + \frac{\gamma x}{\sigma(u)}\right)^{-1/\gamma} \quad (3.4)$$

Este resultado mostra que o excesso acima de um limite u , escrito $X - u$, provavelmente seguirá uma Distribuição Generalizada de Pareto (*GPD - General Pareto Distribution*) com os parâmetros γ, σ .

A ideia decorre de um resultado provado por [59], que revela a relação entre duas famílias de distribuição: a chamada *Distribuição de Valores Extremos* e a chamada *Distribuição Generalizada de Pareto*. O Teorema de Pickands esclarece que aquelas funções de distribuição, cujas caudas podem ser satisfatoriamente aproximadas pelas distribuições generalizadas de Pareto, são precisamente funções que se encontram nos domínios de atração das Distribuições de Valores Extremos.

Ainda segundo [59], a distribuição generalizada de Pareto é uma família de distribuições de probabilidade contínua, frequentemente utilizada para modelar as caudas de outras distribuições, e são definidas por três parâmetros a saber: local (μ), escala (σ) e forma (γ). Em vez de ajustar um EVD aos valores extremos de X , a abordagem POT tenta ajustar um GPD aos excessos $X - u$.

3.4.4 Desafios na Detecção de Anomalias

Segundo [13], em um nível abstrato, uma anomalia é definida como um padrão que não está consoante o comportamento normal esperado. Uma abordagem direta de detecção de anomalias, portanto, é definir uma região que representa o comportamento normal e declarar qualquer observação nos dados que não pertençam a essa região normal como uma anomalia. Contudo, vários fatores tornam essa abordagem aparentemente simples muito desafiadora:

- Definir uma região normal que englobe todos os comportamentos normais possíveis é difícil. Além disso, a fronteira entre comportamento normal e anômalo geralmente não é precisa. Assim, uma observação anômala que fica próxima à fronteira pode ser realmente normal e vice-versa.
- Quando as anomalias são o resultado de ações maliciosas, os adversários maliciosos geralmente se adaptam para fazer com que as observações anômalas pareçam normais, dificultando a tarefa de definir o comportamento normal.
- Em muitos domínios, o comportamento normal continua evoluindo e uma noção atual de comportamento normal pode não ser suficientemente representada no futuro.
- A noção exata de uma anomalia muda para diferentes domínios de aplicação. Por exemplo, no domínio médico, um pequeno desvio do normal (por exemplo, flutuações na temperatura corporal) pode ser uma anomalia, enquanto um desvio semelhante no domínio do mercado de ações (por exemplo, flutuações no valor de uma ação) pode ser considerado normal. Assim, a aplicação de uma técnica desenvolvida em um domínio para outro não é direta.
- A disponibilidade de dados rotulados para treinamento/validação de modelos usados pelas técnicas de detecção de anomalias é geralmente uma questão importante.

- Muitas vezes, os dados contêm ruído que tende a ser semelhante às anomalias reais, portanto, é difícil de distinguir e remover.

Devido a esses desafios, o problema de detecção de anomalias, em sua forma mais geral, não é fácil de resolver. De fato, a maioria das técnicas de detecção de anomalias existentes resolve uma formulação específica do problema. A formulação é induzida por vários fatores, como os dados, disponibilidade de dados rotulados, categorias de anomalias a serem detectadas e assim por diante. Geralmente, esses fatores são determinados pelo domínio do problema em que as anomalias precisam ser detectadas.

Capítulo 4

Revisão de Literatura

Os estudos relacionados discutidos nesta seção apresentam uma análise dos trabalhos disponíveis na literatura pesquisada sobre detecção de anomalias utilizando imagens de satélite. Assim, cada subseção representa trabalhos mais correlatos ao presente estudo.

4.1 Detecção de Anomalias em Imagens de Satélite

Para detectar regiões de anomalias em séries temporais de imagens de satélite, muitos estudos desenvolveram métodos particulares de identificação de regiões anômalas. Os métodos podem ser divididos em dois grupos. O primeiro grupo inclui os métodos que se encaixam em toda a série temporal com algum modelo, como modelo harmônico por partes [77; 78], modelo harmônico não linear [12] e função de autocorrelação temporal [41]. Logo, as regiões de anomalia nas séries temporais de imagens de satélite são detectadas pela comparação dos parâmetros dos modelos ajustados para diferentes partes dos dados das séries temporais.

O segundo grupo consiste nos métodos que monitoram anomalias nos dados de séries temporais utilizando algum modelo de previsão, como o Filtro Estendido de Kalman [40], Processo Gaussiano [14], modelo harmônico [79; 82; 83], teoria de Martingale e teorema do limite central de Martingale [3]. Em geral, esses métodos de detecção consistem em duas etapas principais, ajustar o modelo aos dados históricos e a detecção de anomalias, comparando as novas observações com as previsões do modelo ajustado.

Anomalias são padrões nos dados que não estão em conformidade com uma noção bem

definida de comportamento normal [13]. Se as regiões anômalas forem pequenas em comparação com o tamanho da imagem, as características espaciais na imagem serão dominadas pela região no entorno. Uma etapa importante na detecção de regiões externas é geralmente calcular uma métrica de correspondência com o entorno, que pode ser um limiar para detectar objetos que dificilmente serão objetos pertencentes a esse entorno.

4.2 Detecção de Anomalias em Imagens Multitemporais

O estudo proposto por Zhou et al. [81] cria um método para detectar regiões anômalas em séries temporais de imagens de satélite. O método é baseado na análise de autocorrelação sazonal de dados de séries temporais em imagens de satélite para remover ao máximo a sazonalidade das séries temporais originais e, então, aprimorar as informações de anomalia presentes. Assim, as anomalias são detectadas com maior confiabilidade por análise estatística. O método foi validado por um estudo de caso para detectar regiões anômalas causadas por inundações severas em séries temporais de imagens MODIS (*Moderate Resolution Imaging Spectroradiometer*) coletadas dos satélites *Landsat-7* e *Landsat-8*.

O método para detecção de região de anomalia usando análise de autocorrelação sazonal é flexível para atender a alguns requisitos de precisão de detecção. Ele não precisa de suavização de séries temporais e pode detectar com confiança regiões de anomalias em séries temporais de satélites ruidosas. No entanto, os autores sugerem fazer um pré-processamento para a remoção de nuvens e tendências dos dados de séries temporais. Assim, o uso de séries temporais de imagens de satélite com maior resolução espacial tende a aumentar o desempenho do método.

Para validar quantitativamente os resultados das regiões de anomalia detectadas, os resultados de detecção foram comparados com a região de anomalia de referência. A revocação (*Recall*) das regiões de anomalias detectadas é de 79,62%, indicando que cerca de quatro quintos dos pixels nas regiões de anomalias dos rótulos são identificados como região de anomalia. A precisão (*Precision*) das regiões de anomalias detectadas é de 90,62%, indicando que mais de noventa por cento da região de anomalia detectada é corretamente identificada nos rótulos, ou seja, apenas dez por cento são detecções errôneas quando comparadas aos rótulos de cada região. A acurácia geral das regiões de anomalias detectadas é de

88,68%, o que mostra que quase noventa por cento de toda a região da imagem é identificada corretamente, distinguindo corretamente regiões com anomalia e regiões sem anomalias.

O trabalho proposto por Huang et al. [36] foi proposto um método de detecção automática de mudanças que consiste em três componentes: 1) em primeiro lugar, os autores utilizaram em conjunto as características verticais planas para delinear as áreas de edificações recém-construídas; 2) a correção temporal baseada em objeto é aplicada para melhorar a consistência espaço-temporal das feições; e 3) finalmente, um modelo de detecção de mudanças multitemporais é usado para capturar simultaneamente as áreas de edificações recém-construídas e o tempo de mudança. O método foi aplicado em duas áreas periféricas urbanas de Pequim (7 conjuntos de imagens multi-temporais) e Xangai (7 conjuntos de imagens multi-temporais), respectivamente, cidades que estão passando por uma rápida urbanização. O método de detecção de mudança multi-temporal proposto apresenta um desempenho, com um valor de F-score de 90% na detecção de mudanças espaciais e uma acurácia geral de 92% na identificação de tempos de mudança usando uma tolerância temporal de um ano.

4.3 Detecção de Anomalias Não-Supervisionadas

A detecção de mudanças em imagens de satélite utilizando abordagens não supervisionadas, também foi avaliada, como abordado em Han et al. [32]. Nesse trabalho, foi proposta uma abordagem de otimização para melhorar o desempenho de detecção de alterações em imagens de alta resolução do conjunto de dado Worldview-3. O objetivo do estudo apresentado em [32] é analisar o efeito do desempenho do co-registro fino na detecção de alterações não supervisionadas. Para isso, foram extraídas amostras de ruído de registro, denotadas como pixels desalinhados em uma região local. Em seguida, a localização dos pontos conjugados é posicionada analisando a distribuição local das amostras de ruído extraídas. Os pontos conjugados são empregados para gerar um modelo de transformação não rígido para deformar uma imagem detectada em uma imagem de referência. Uma abordagem de análise de vetor de mudança não supervisionada é usada para validar a eficácia do desempenho do co-registro fino proposto. Para avaliar quantitativamente o impacto do desempenho do co-registro na precisão da detecção de alterações, a área sob a curva (AUC) e a acurácia geral foram calculadas com base na comparação dos resultados da detecção de alterações com um

mapa de detecção de alterações de referência, assinalando 70,00% 90,99% respectivamente.

No estudo proposto por [28], um método de detecção de anomalia baseado em *Deep Belief Networks* de peso adaptativo foi proposto para imagens hiperespectrais. Um codificador automático não supervisionado para aprender as características de alto nível do conjunto de dados HSI foi utilizado.

As características de alto nível e os erros de reconstrução são aprendidos então através da rede, de uma maneira que não esta não seja afetada pela distribuição do *background* anterior. Para reduzir a contaminação por anomalias locais, os pesos adaptativos são construídos a partir de erros de reconstrução e informações estatísticas. Logo, ao combinar a informação dos erros de reconstrução e os fatores de penalidade, pesos adaptativos são criados para medir a contribuição dos pixels vizinhos.

Ao usar a imagem de código gerada durante a inferência e modificada por pesos adaptados atualizados, uma distância euclidiana local entre os pixels sob teste e seus pixels vizinhos é usada para determinar os alvos da anomalia. O método proposto mostra um desempenho, com um valor de AUC de 99,8% na detecção de mudanças e uma curva ROC de 92% na identificação de tempos de mudança usando uma tolerância temporal de um ano.

4.4 Detecção de Anomalias com Aprendizado Profundo

As arquiteturas de redes neurais convolucionais também desempenharam papel importante na detecção de anomalias em imagens de satélite, tal qual podemos observar em Lim et al. [47]. Nessa pesquisa são propostas três redes neurais convolucionais estruturadas por codificador-decodificador para fins de detecção de alterações. O algoritmo proposto detecta mudanças nas imagens de entrada bitemporais e produz um mapa de segmentação. Para tanto, as três redes são treinadas de maneira ponta a ponta usando imagens temporais e o rótulo correspondente, que representa as regiões de mudança. Ao combinar um codificador com dois decodificadores, são projetadas três CNNs codificador-decodificador treinadas separadamente. Finalmente, obtém-se o mapa final de mudança binária, limitando a saída média das três CNNs.

Um conjunto de dados multitemporal para detecção de alterações usando imagens do Google Earth foi criado, captando imagens dos satélites Landsat 7 e 8. As imagens continham

três bandas (RGB), incluindo áreas urbanas ao redor de Seul, na Coreia do Sul. Em 10 áreas de Seul e 3 áreas de cidades ao redor de Seul, obtiveram-se imagens de satélite capturadas em diferentes momentos.

Para avaliar o desempenho, os pixels foram classificados em um mapa de detecção de alterações usando os rótulos correspondentes. Verdadeiros positivos e verdadeiros negativos denotam os números de pixels corretamente previstos como pixels alterados e inalterados, respectivamente. Alarmes falsos são o número de pixels previstos como alterados, mas inalterados, na verdade, e alarmes de falta são o número de casos inversos. As três redes têm as arquiteturas codificador-decodificador e produzem F1-score de 71,16%, F2-score de 75,21% respectivamente.

Em Li et al. [45] uma estrutura de detecção de anomalias, em que uma Rede Neural Convolutiva (*CNN - Convolutional Neural Network*) de várias camadas é treinada usando as diferenças nos valores de cinza entre pares de pixels vizinhos na imagem de referência como dados de entrada é proposta. Em seguida, na fase de teste, as anomalias são detectadas avaliando-se as diferenças entre pares de pixels vizinhos usando a CNN treinada.

Os dois dados experimentais de teste foram obtidos do sensor AVIRIS (*Airborne Visible/Infrared Imaging Spectrometer*, cobrindo o Moffett Field, Califórnia, e Cuprite, Nevada, respectivamente. Essas duas cenas consistem em 512×512 pixels com 224 bandas abrangendo cerca de 20 m de resolução espacial.

Para avaliar quantitativamente o impacto do desempenho da CNN, a área sob a curva AUCROC foi calculada com base na comparação dos resultados da detecção de alterações, assinalando respectivamente 98,28%.

Um algoritmo de agrupamento FCM foi aplicado para a classificação de imagens de sensoriamento remoto de baixo custo [44]. A técnica LoF foi aplicada para remover as anomalias. Os autores tentaram diferentes configurações de parâmetros, e a configuração acima parece ser a mais representativa para o método proposto. *Support Vector Regressor*, um método de regressão popular, foi usado para prever a taxa de alteração da função objetivo com a precisão desejada.

O conjunto de dados utilizado é o *SpaceNet*, um conjunto de dados público contendo imagens de satélite, fornecido pela *Digital Globe*, um fornecedor americano de imagens espaciais e conteúdo geoespacial. O *SpaceNet* contém mais de 17.533 imagens de sensoriamento

remoto de alta resolução (438×406 pixels) em Las Vegas, Cartum, Rio de Janeiro, Xangai, etc. O conjunto de dados inclui uma grande quantidade de informações geoespaciais relacionadas a vários casos de uso, por exemplo, mapeamento de infraestrutura e classificação da cobertura do solo.

De forma geral, o algoritmo *Local Outlier Factors* é utilizado para remover anomalias antes de ajustar a relação entre a taxa de variação da função objetivo e a acurácia, o que melhora o custo-benefício na classificação da cobertura do solo. Os resultados mostram que o framework proposto pode alcançar precisão suficiente e economizar muito custo computacional na nuvem.

A acurácia geral da região de anomalias detectada é de 85%, o que mostra que quase noventa por cento de toda a região da imagem é identificada corretamente, distinguindo corretamente regiões com anomalia e regiões sem anomalias.

Dataset	Método	ACC	F1	F2	AUC
ZY-3	. Object-based Segmentation [36]	92,00%	90,00%	-	-
SpaceNet 5 [2]	4 x ResNet50 + 4 x SE-ResNeXt50 [75]	-	69,00%	-	-
SpaceNet 5 [2]	Fuzzy C-means Clustering [44]	85,00%	-	-	-
Worldview-3	Change Vector Analysis [32]	90,99%	-	-	70,00%
Google Earth	Encoder-Decoder CNN [47]	-	71,16%	75,21%	-
AVIRIS [28]	Deep Belief Network [51]	-	-	-	99,80%
Terra/MODIS	SARIMA [81]	88,68%	-	-	-

Tabela 4.1: Sumário dos resultados da Literatura Analisada.

Em resumo, o aprendizado profundo tem sido amplamente aplicado à classificação de imagens multi/hiper-espectrais, e resultados promissores tem sido alcançados. No entanto, novos desafios a serem explorados tem emergido, e incluem aprimoramento de imagens hiper-espectrais, mistura espectral não linear, análise de séries temporais hiper-espectrais,

etc.

Embora os métodos e estudos citados anteriormente demonstrem resultados promissores na detecção de anomalias aplicada a séries temporais de imagens de satélite, comumente, tais métodos são projetados para identificar alterações bruscas na série temporal ou que decorrem apenas no intervalo de um ano para o outro.

Capítulo 5

Abordagem para Detecção de Anomalias em Imagens de Satélite

Neste capítulo, é discutida a abordagem proposta para detecção de anomalias em imagens de satélite. Como já dito na Seção 3.4, a teoria dos valores extremos é um método estatístico, baseado em princípios que definem limites para a detecção de anomalias, lidando diretamente com valores raros ou extremos e foi proposta em 1975. Assim, pode-se dizer que a teoria dos valores extremos procura modelar o comportamento assintótico de eventos que se distanciam muito da mediana de uma distribuição. Este capítulo é composto por duas seções, na Seção 5.1 é apresentada uma breve introdução da abordagem proposta e a Seção 5.2 são descritos os critérios utilizados pela abordagem para a de detecção de anomalias

5.1 Introdução

Tomando como base o conceito de detecção de anomalias proposto por Markou et al.[52], a abordagem proposta nessa dissertação visa identificar padrões em imagens de satélite e, possivelmente, inferir regiões de adensamento urbano além de mudanças geográficas em regiões específicas, utilizando modelos urbanos espacialmente explícitos. Logo, nessa pesquisa entende-se por anomalia mudanças ou alterações abruptas observadas em imagens de satélite em diferentes períodos.

O método proposto é composto de três etapas a saber: (1) redução da dimensionalidade de imagens do conjunto de dados SN7, (2) detecção de anomalias usando uma abordagem

estatística paramétrica que se baseia na teoria dos valores extremos e (3) classificação dos pontos de anomalias.

Em uma imagem multi-espectral, os pixels cujas assinaturas espectrais são significativamente diferentes dos pixels da imagem como um todo são considerados anomalias. Como o conhecimento prévio do espectro anômalo é difícil de obter na prática, a detecção de anomalias é geralmente resolvida por modelagem de *background* ou caracterização estatística para dados multi-espectrais. Assim, foram coletadas imagens de sensoriamento remoto, com base em uma sequência temporal de imagens do programa *SpaceNet*¹ de forma a compor nosso conjunto de dados.

As regiões compõem um novo conjunto de dados abertos de imagens de satélite, que inclui 24 imagens (uma por mês) cobrindo cerca de 100 regiões geográficas. O conjunto de dados compreende mais de 40.000 quilômetros quadrados de imagens e rótulos poligonais denotando marcações de edifícios nas imagens, totalizando mais de 10 milhões de anotações individuais. Os locais foram selecionados por destacar mudanças geográficas significativas, embora alguns tenham sido selecionados para se sobrepor a locais anteriores do *SpaceNet* e alguns foram selecionados, devido ao interesse geopolítico.

Inspirados por essa periodicidade e de modo a extrair as características das imagens de sensoriamento remoto do conjunto de dados *SpaceNet*, as imagens foram submetidas a uma rede convolucional treinada para uma atividade de segmentação, de onde foram extraídas informações das características das imagens geográficas a partir da camada mais profunda da rede. A etapa de segmentação faz-se necessária dada a complexidade de se extrair informações e características de um domínio complexo que são as imagens de satélite. As representações obtidas são então submetidas a uma abordagem probabilística paramétrica de detecção de anomalias, baseando-se na teoria dos valores extremos, visando-se reconhecer e inferir possíveis padrões de alteração em regiões urbanas.

Apesar do conjunto de dados contar com um total de 100 regiões, algumas imagens do conjunto de dados encontravam-se com ruídos (degraus, nuvens, e regiões sobrepostas) e até tamanhos diversos, logo para essa pesquisa optou-se por trabalhar apenas com 65 regiões. Estes dados foram posteriormente separados de forma equilibrada em três conjuntos menores de forma que tivéssemos conjuntos diferentes para treino, validação e testes da rede

¹<https://spacenet.ai/> — Acesso em: 21 set 2021

convolucional.

Na etapa de pré-processamento do conjunto de dados, foi possível constatar que o tamanho ideal de recorte de imagem que apresentava um melhor desempenho do ponto de vista técnico com um baixo custo operacional, no que se refere a recursos de hardware, foi 512×512 pixels. Além disso, as imagens disponibilizadas no conjunto de dados já contavam com tratamento para remoção de nuvens, o que garantiu maior confiabilidade aos dados bem como ausência de ruídos.

Utilizamos como entrada os valores de intensidade de pixel de recortes com dimensões de 512×512 pixels, portanto $n = 262.144$. O objetivo desta etapa traduz-se como capturar as representações do espaço latente geradas pela rede convolucional U-net. Em seguida, utilizando o conjunto de dados previamente separado para a etapa de testes, agrupando as representações dos recortes em um vetor unidimensional, tomando o cuidado de organizá-los corretamente quanto a sua cronologia temporal e posteriormente servindo o algoritmo de detecção de anomalias picos acima do limite (Peaks-Over-Threshold - **POT**), o qual pode ter sua implementação visualizada no Apêndice B. Na etapa seguinte, o algoritmo mapeia as distribuições das características presentes no espaço latente extraído das imagens da rede convolucional, onde identifica representações de padrões de crescimento anômalo conforme pode ser visto no diagrama apresentado pela Figura 5.1.

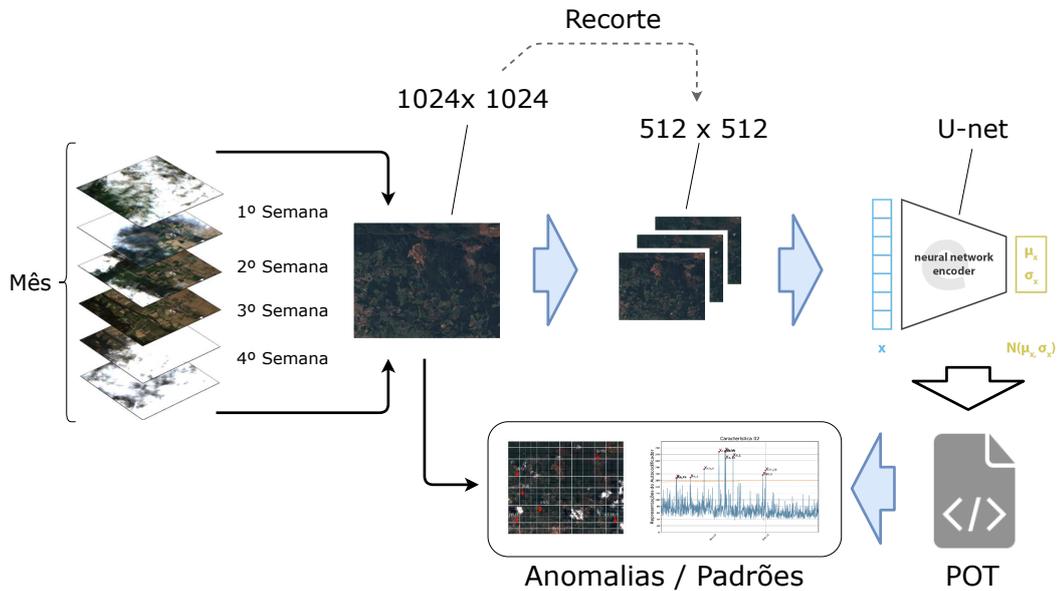


Figura 5.1: Abordagem proposta.
 Fonte: Elaborado pelo próprio autor.

5.2 Critério para Determinação de Anomalia

O método denominado Picos Acima do Limite (*POT - Peaks-Over-Threshold*) é um método estatístico que surgiu na área de hidrologia em meados dos anos 80, e que constitui uma das principais ferramentas estatísticas utilizadas na estimativa da probabilidade de eventos raros [7].

De forma prática, o método **POT** retorna um limiar z_q que usamos para definir um “limite de normalidade” (Figura 5.2). O algoritmo **POT**, cuja implementação pode ser visualizada no Apêndice B, recebe dois fluxos de dados, o primeiro serve para calibração. Uma vez, calibrado, o algoritmo utiliza então as observações presentes no segundo fluxo para detectar os pontos de anomalias e refinar o limite de anomalia z_q .

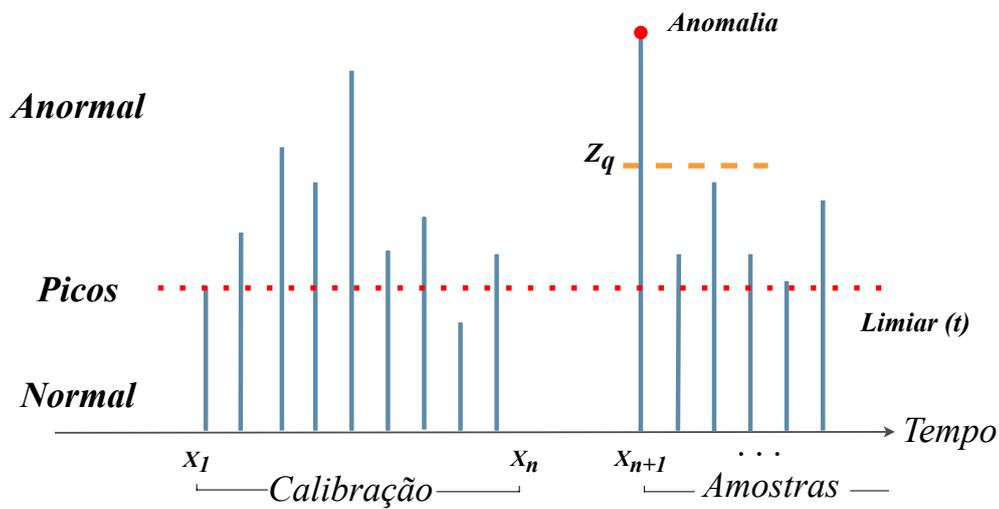


Figura 5.2: Limiar de Detecção de Anomalia.
 Fonte: Elaborada pelo autor inspirada em [69]

A estimativa gerada pelo POT não precisa armazenar toda a série temporal (apenas os picos), esse algoritmo requer pouca memória para ser utilizado em um fluxo de dados de entrada. No entanto, o fluxo deve conter valores oriundos da mesma distribuição, portanto, essa distribuição não pode ser estacionária, ou seja, permanecer inalterada no decorrer do tempo.

O algoritmo POT funciona detectando eventos anormais em um fluxo $(X_i)_{i>0}$ de forma agnóstica quanto à distribuição. Primeiramente, o algoritmo realiza uma estimativa nos n primeiros valores ($n \sim 1000$) e obtém um limiar inicial z_q (inicialização). Então, todos os próximos valores observados, podem sinalizar os eventos anormais ou atualizar o limiar (Figura 5.2). Se um valor ultrapassa o limiar z_q , é considerado anormal. As anomalias não são consideradas na atualização do modelo. Nos outros casos, ou X_i é maior que o limiar inicial (caso de pico) ou é um valor "comum" (caso normal). No caso de pico, identifica-se o pico definido e atualiza-se o limiar.

O procedimento descrito no parágrafo anterior foi aplicado individualmente a cada região de forma que a distribuição dos espaços latentes de uma região A estivesse isolada da distri-

buição de uma região B. A ideia é definir um primeiro limite z_q verificando $P(X > z_q) < q$. A Figura 5.2 ilustra o comportamento geral do algoritmo POT.

Os picos acima do limiar t (linha pontilhada vermelha), representam possíveis pontos de anomalias, ajustados a uma Distribuição Generalizada de Pareto (*GPD — General Pareto Distribution*) para, em seguida, inferir a distribuição dos valores extremos e calcular o limite z_q .

Logo, quanto maior for o limiar t , mais relevante será o ajuste do GPD (baixo viés). No entanto, se t for muito alto, os picos definidos Y_t seriam pouco preenchidos, então o modelo seria mais variável (alta variância). A única condição importante é garantir que t seja menor que z_q , o que significa que a probabilidade associada a t deve ser menor que $1 - q$. Esta etapa de inicialização é resumida no Apêndice B.

Por fim, para classificar os pontos de anomalias gerados pelo POT, optamos por validar a abordagem proposta submetendo o resultado do algoritmo não paramétrico a uma abordagem supervisionada para identificação de anomalias descrita em detalhes na Seção 6.

Capítulo 6

Resultados

Nesta seção, investiga-se a capacidade de detecção de anomalias do modelo proposto e o compara com os algoritmos de detecção de anomalias existentes na literatura examinada.

6.1 Conjunto de Dados

Como mencionado, para o estudo proposto, foram coletadas imagens de sensoriamento remoto, temporalmente organizadas, do conjunto de dados de desenvolvimento urbano multi-temporal SpaceNet (*MUDS - Multi-Temporal Urban Development SpaceNet*) que consiste em 101 sequências rotuladas de imagens de satélite coletadas pela constelação Dove do Planet Labs entre 2017 e 2020, com rótulos e marcações para as construções presentes em cada imagem. As sequências de imagens são amostradas em 101 áreas distintas de interesse em todo o mundo, cobrindo seis continentes. Esses locais foram selecionados por serem geograficamente diversos e apresentarem mudanças expressivas na urbanização ao longo de um período de dois anos. O conjunto de dados MUDS é de código aberto sob uma licença *CC-BY4.0 ShareAlike International*¹ para incentivar o uso amplo.

O tamanho de cada imagem é de 1024×1024 pixels, correspondendo a 18 km^2 , e a área total das imagens no conjunto de dados é de 41.250 km^2 . No entanto, devido a limitações técnicas, para essa pesquisa as imagens foram redimensionadas para 512×512 pixels. Consulte a Tabela 1 ou spacenet.ai para estatísticas adicionais. A série temporal contém imagens de 18 a 26 meses, dependendo da AOI (*AOI - Area of Interest*). Este período captura

¹<https://registry.opendata.aws/spacenet/>

várias estações e condições atmosféricas, bem como o início e a conclusão de vários projetos de construção (Figura 3.4). Além disso, imagens contendo uma quantidade excessiva de nuvens ou neblina foram totalmente excluídas do conjunto de dados, causando pequenas lacunas temporais em algumas séries temporais.

6.2 Limiar para Detecção de Anomalias

Na etapa de definição dos parâmetros do algoritmo **POT** definiram-se $q = 10^{-3}$ e $n = 3200$, onde q representa um limiar de risco e n representa o número de amostras utilizado como lote inicial na etapa de calibração do algoritmo, para o qual utilizamos as imagens referentes ao ano de 2018. Para a etapa de identificação das anomalias, utilizamos 3880 amostras como entrada unidimensional referente aos anos seguintes. Cabe ainda ressaltar que as etapas de calibração e execução do algoritmo ocorreram com imagens referentes à mesma região

A Figura 6.1, apresenta ao longo do eixo x a linha temporal das imagens utilizadas na etapa de validação, enquanto o eixo y apresenta as representações extraídas das imagens pela rede convolucional. Os pontos vermelhos que ultrapassam o limiar (linha pontilhada amarela) denotam os alertas de anomalias emitidos pelo **POT**.

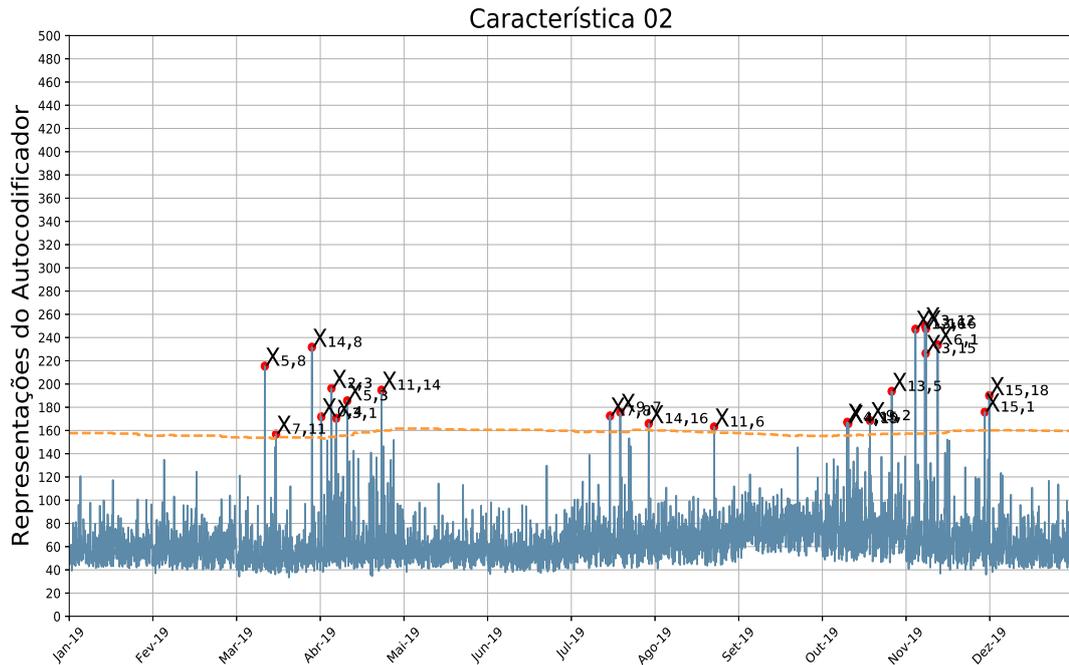


Figura 6.1: Pontos de anomalia para a característica 02 das 1024 características monitoradas.

Os resultados mostraram que os alertas emitidos pelo POT estão presentes nos meses de Março, Abril, Julho, Agosto, Outubro, Novembro e Dezembro. É possível observar ainda a presença de vários pontos ao longo do eixo y , possivelmente anômalos, representados por círculos vermelhos. O limiar então é recalculado e atualizado para as amostras seguintes.

A Figura 6.1 também apresenta de forma aproximada as possíveis regiões com anomalias para os meses de Outubro, Novembro e Dezembro. Cada ponto do vetor de dados de entrada é representado por uma legenda, que refere-se a região da imagem original a qual esse ponto pertence. Assim, cada ponto é representado na forma de uma matriz de imagens $X_{i,j}$, onde i identifica a linha que representa aquela região na imagem original e j representa a coluna referentes a imagem original de 1024×1024 .

A Figura 6.2 apresenta a imagem referente ao mês de outubro de 2019 e em seguida a Figura 6.3 apresenta a imagem referente ao mês de novembro de 2019. As regiões destacadas identificam possíveis regiões anômalas apontadas pelo POT diante das características extraídas pela rede convolucional. Baseando-se em imagens dos meses anteriores, as regiões destacadas apresentam consideráveis pontos de anomalias. Apesar do algoritmo POT mostrar certa eficiência na identificação das anomalias, a principal dificuldade reside na di-

versidade das superfícies capturadas por imagens de satélite, bem como no nível de detalhe apresentado pelas mesmas.

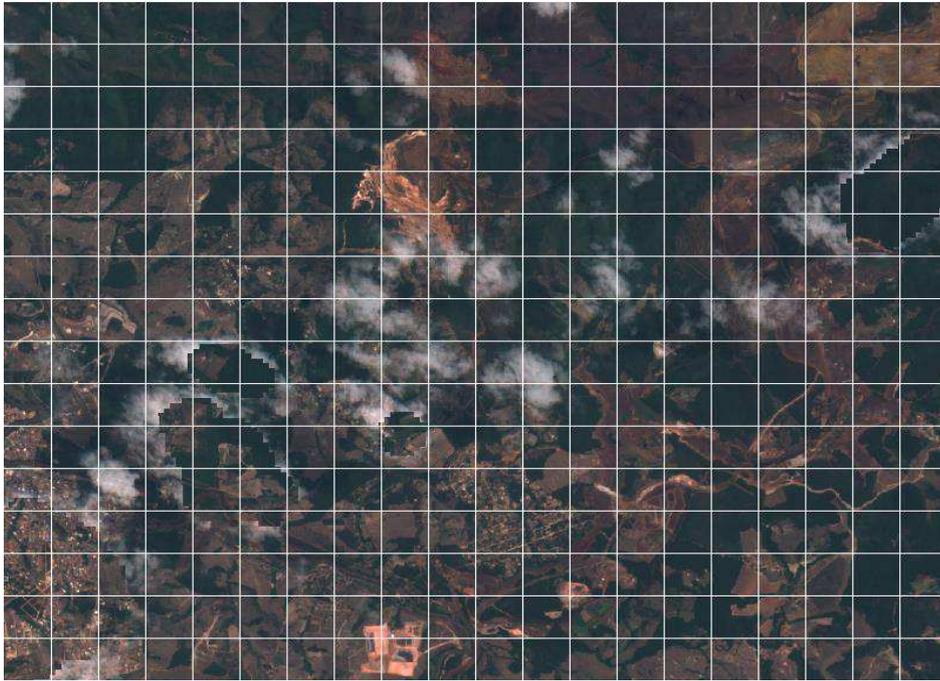


Figura 6.2: Superfície terrestre no mês de Outubro de 2019.

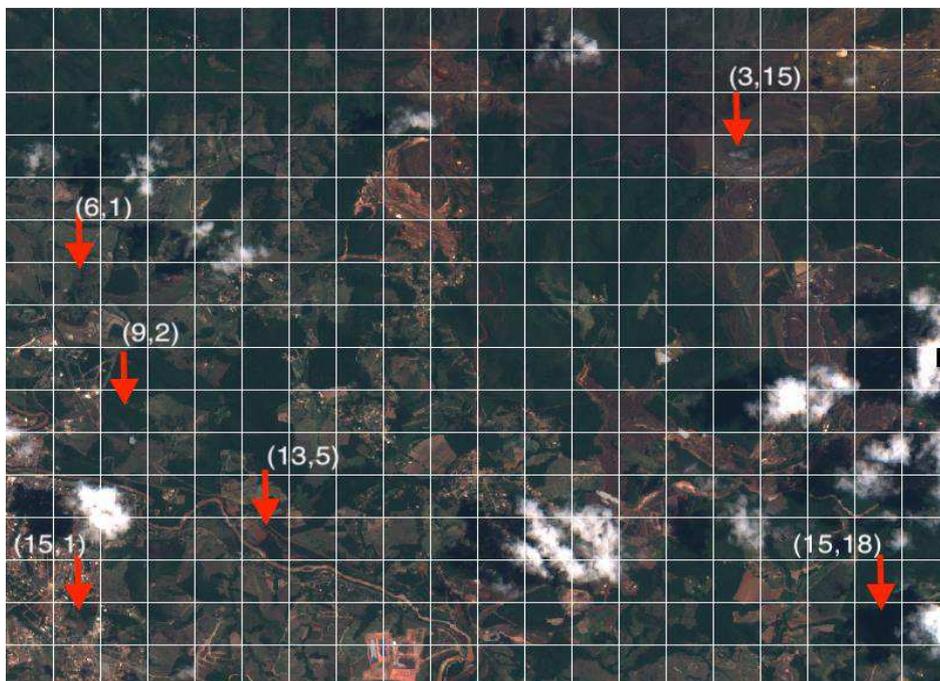


Figura 6.3: Superfície terrestre no mês de Novembro de 2019.

6.3 Classificação de Anomalias

Como já mencionado, com o propósito de definir objetivamente o limiar de anomalias, optamos por construir uma estratégia de detecção de anomalias de forma semi-supervisionada. Assim, os pontos de anomalias previamente identificados pelo **POT** foram submetidos a um algoritmo de classificação. O intuito da etapa de classificação foi validar as regiões de anomalia assinaladas pelo **POT**, bem como identificar e categorizar esses pontos em anomalias ou pontos normais.

Árvore de decisão de aumento de gradiente (*GBDT - Gradient boosting decision tree*) é um algoritmo de classificação popular [25]. Dado um conjunto de treinamento $\{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_n, y_n)\}$, tal que as amostras de dados são representadas por x e os rótulos das classes representados por y , utilizamos $F(x)$ para representar a função estimada e o objetivo de otimização do GBDT é minimizar a função de perda $L(y, F(x))$.

O GBDT é um algoritmo *ensemble* cujo classificador base é uma árvore de decisão, portanto, o principal custo é encontrar os melhores pontos de divisão ao aprender árvores de decisão. Ke et al. [37] propuseram uma árvore de decisão de aumento de gradiente altamente eficiente usando amostragem unilateral baseada em gradiente (*GOSS - gradient-based one-side sampling*) e agrupamento de características exclusivas (*EFB - Exclusive Feature Bundling*) chamado *LightGBM*. Para GBDT, o ganho de informação é geralmente obtido de cada nó. *LightGBM* usa GOSS para determinar o ponto de divisão através do cálculo do ganho de variação.

O agrupamento de características exclusivas (EFB) é um método quase sem perdas para reduzir o número de características efetivas. Em um espaço de características esparsas, muitas características são quase exclusivas, devido à sua alta dimensionalidade, o que implica que raramente assumem valores diferentes de zero simultaneamente. As características codificadas *One-hot* são um exemplo perfeito de características exclusivas. O EFB reúne essas características, reduzindo a dimensionalidade para melhorar a eficiência, mantendo um alto nível de precisão. Em outras palavras, o *LightGBM* pode agrupar características exclusivas em uma única característica e o algoritmo de varredura de características pode então ser projetado para criar os mesmos histogramas de características de modo mais otimizado a partir das características originais.

Em resumo, *LightGBM* é uma implementação eficaz de GBDT com GOSS e EFB para melhorar a eficiência computacional sem prejudicar a precisão. O GOSS é empregado para dividir o nó ótimo através do cálculo do ganho de variância. O EFB pode acelerar o processo de treinamento do GBDT através do agrupamento de muitas características exclusivas para características menos densas. *LightGBM* é um modelo *ensemble* cujo classificador base é uma árvore de decisão, que pode ser treinada em sequência ajustando os gradientes negativos da função de perda.

6.4 Treinamento e Teste do Modelo

Para o desenvolvimento desse experimento, utilizamos uma CPU AMD Ryzen 7 5800H Radeon Graphic de 3.20 GHz e 16 GB de RAM, além de uma placa gráfica NVIDIA [GeForce RTX 3060] com CUDA instalado. Para validar a abordagem proposta, optou-se por extrair um total de 1024 características das imagens de entrada, obtidas por representações do espaço latente, geradas pela rede convolucional e submetê-las a nossa abordagem de detecção de anomalias, visto que essas representações encontram-se na etapa final do processo de codificação de características sendo extraídas da camada mais profunda da rede convolucional. Logo entende-se que na camada mais profunda as características apresentem uma maior completude informacional para a detecção de anomalias

A tarefa de extração das representações de uma quantidade finita de características de imagens de sensoriamento remoto utilizou como entrada um conjunto de imagens de teste composto por 1236 amostras, com dimensões de 512×512 , compostas por 3 canais de cores (RGB) extraídas do conjunto de dados *Spacenet 7*.

A Figura 3.4 mostra a arquitetura U-net utilizada neste estudo. As camadas ocultas utilizaram a função de ativação ReLU [17] com a entropia binária (*Binary Cross entropy*) como função de custo e a função sigmoide foi utilizada na camada de saída. A dimensionalidade de entrada e saída foi de $512 \times 512 \times 3$, conforme o conjunto de dados escolhido. Todas as operações de convolução no modelo tinham um tamanho de kernel de 3×3 , um passo de 2 e um preenchimento de 0. Todas as operações de deconvolução usaram um tamanho de kernel de 3×3 , um passo de 2 e um preenchimento de 0. Cada operação de convolução e deconvolução foi seguida pela normalização em lote [48].

O codificador da arquitetura U-net consistia em cinco níveis. Nos primeiros quatro níveis, duas operações de convolução foram aplicadas, seguidas por uma operação de agrupamento máximo com um tamanho de kernel de 3×3 , um passo de 2 e um preenchimento de 0. Na quinta camada (*Skip Connection*), duas operações de convolução foram aplicadas, mas nenhuma operação de *pooling* máximo foi aplicada. O decodificador tinha quatro níveis, cada um consistindo em uma operação de deconvolução, seguida por duas operações de convolução. O nível final do decodificador teve uma terceira operação de convolução para gerar o número correto de canais para a saída. O número de kernels no primeiro nível do codificador foi de 64. Esse número dobrou a cada nível subsequente, atingindo um total de 1024 na última camada. O número de kernels foi reduzido pela metade em cada nível do decodificador, resultando em 64 no último nível, antes que a operação de convolução final fosse aplicada. A operação de convolução final reduziu o número de canais para 3.

Para o propósito deste estudo, a variação da estimativa de momento adaptativo (*Adam* — *adaptive moment estimation*) [39] do algoritmo gradiente descendente foi utilizada para treinar o modelo. A abordagem Adam foi escolhida porque exigia a menor quantidade de ajuste de parâmetros. O modelo foi treinado para realizar segmentação semântica de imagens com tamanho de mini-lote de 64. O modelo foi treinado por 100 épocas com uma taxa de aprendizado de 0,0003. Observe que o método de treinamento é irrelevante para o método de detecção de mudanças proposto. Os parâmetros listados acima foram escolhidos com base em seu desempenho empiricamente adequado. Otimizar ainda mais os parâmetros do algoritmo de treinamento pode melhorar o desempenho do método de detecção de anomalias proposto sendo deixado para pesquisas futuras.

De forma geral, a Tabela 6.8 apresenta os valores dos hiperparâmetros definidos para a rede.

Hiperparâmetro	Valor
Função de custo	Binary Cross Entropy
Algoritmo de otimização	Adam
Número de épocas	100
Tamanho do lote de dados	64

Tabela 6.1: Hiperparâmetros U-net.

6.5 Resultados e Validação

Nesta seção, investiga-se a capacidade de detecção de anomalias do modelo proposto e compara-se com os algoritmos de detecção de anomalias existentes na literatura levantada.

6.5.1 Segmentação em Imagens de Satélite

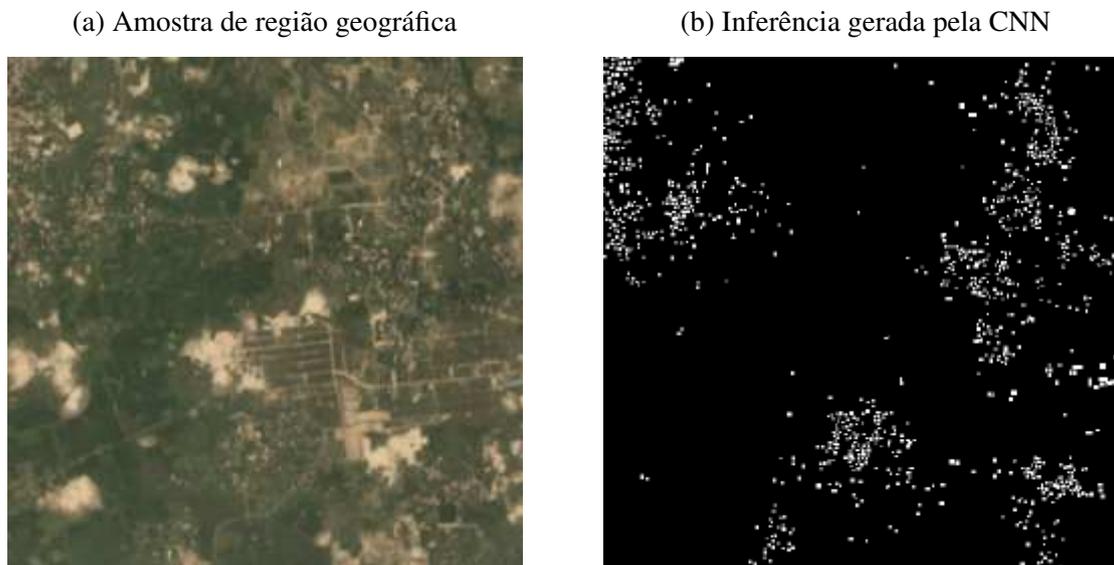


Figura 6.4: Inferência da rede U-net.

A esquerda (a) uma amostra de região geográfica utilizada como rótulo para a inferência da rede convolucional e a direita (b) o resultado da inferência gerado como saída.

A Figura 6.4 mostra um exemplo de entrada para a rede convolucional, com a saída correspondente produzida pelo modelo treinado. Dado que os valores dos pixels foram nor-

malizados, houve uma mudança de cores nas imagens de entrada, como pode ser visto na Figura 6.4. A Figura 6.4 também mostra a saída correspondente, onde a cor de cada pixel é uma representação da confiança do modelo naquele pixel pertencente a uma das classes onde existiram alterações. Analisando-se qualitativamente, é possível perceber que o modelo foi suficientemente eficaz em reconhecer de forma distinta os limites das edificações presentes nas imagens, bem como lidar com sombras. O modelo conseguiu detectar a localização da mudança e classificá-la para a classe semântica correta para a maioria dos pares de teste.

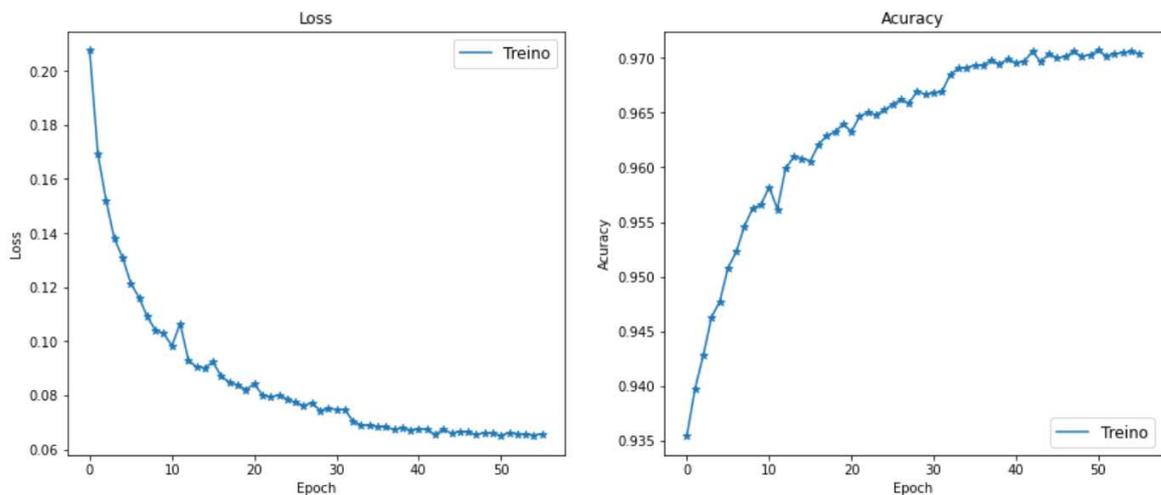


Figura 6.5: Desempenho do treinamento da rede convolucional.

A arquitetura geral da rede convolucional utilizada pode ser verificada no Apêndice A. O desempenho da rede ao longo de 100 épocas convergiu para um valor de perda de 0.0657% e uma acurácia geral de 97,04% (Figura 6.5).

6.5.2 Identificação e Classificação de Anomalias

As características extraídas desse conjunto de dados foram então submetidas à abordagem proposta para de detecção de anomalias. De forma geral, é possível observar que o algoritmo POT captura bem os pontos de anomalia nas representações extraídas das imagens.

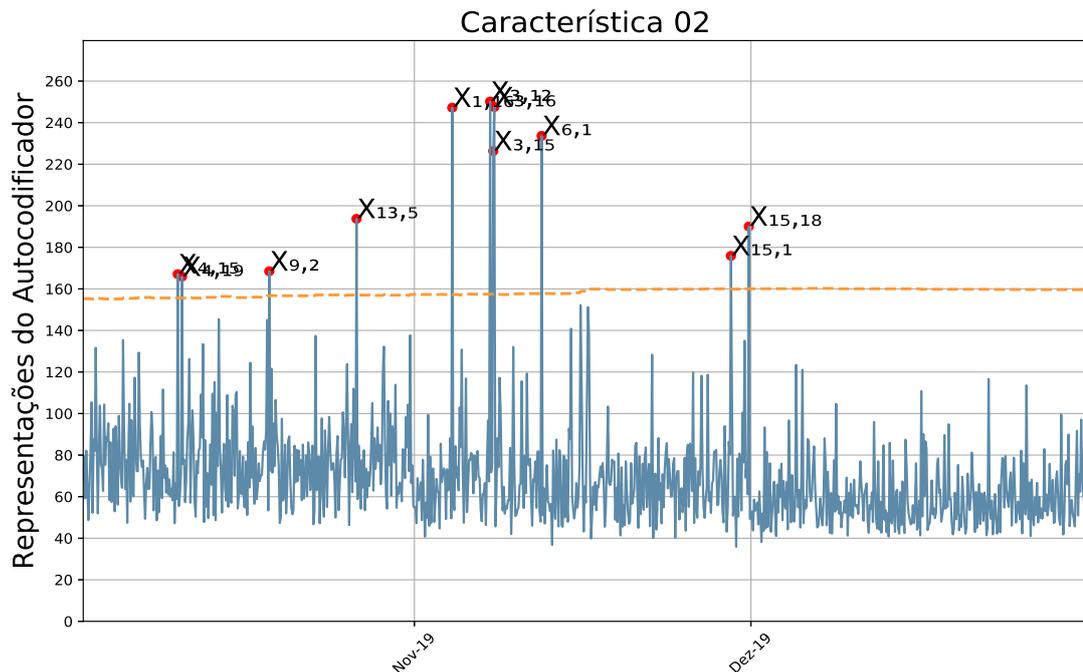


Figura 6.6: Pontos de anomalia ampliados para a característica 02.

Observa-se ainda que quando as linhas temporais de representações são dispostas em ordem cronológica, os pontos de anomalias apresentam correlação entre os meses passados e futuros (Figura 6.6). Isto denota prováveis alterações daquela região, porém em meses diferentes, o que nos leva entender a capacidade do algoritmo em capturar as alterações geográficas.

Para a avaliação de desempenho da abordagem proposta, submeteram-se as representações extraídas pelo POT, um total de 1048 amostras, a um algoritmo de classificação utilizando validação cruzada, que é uma técnica para avaliar a capacidade de generalização de um modelo [42], a partir de um conjunto de dados. Cabe mencionar ainda que não foi preciso efetuar o balanceamento dos dados, visto que o algoritmo utilizado na etapa de classificação baseia-se em *ensembles* que são uma categoria de algoritmos de aprendizado de máquina resilientes a conjuntos de dados desbalanceados. Classificaram-se as amostras extraídas pelo POT para detectar mudanças usando o rótulo correto de cada imagem e apresentando os mesmos em uma matriz de confusão. Verdadeiro Positivo (TP) e Verdadeiro Negativo (TN) denotam o número de imagens corretamente inferidos como imagens com anomalias e sem anomalias, respectivamente. Falso Positivo (FP) é o número de imagens previstas como ano-

malias, mas que, na verdade, não possuem anomalias, e Falso Negativo (FN) é o número de casos que em verdade possuem anomalias, mas que foram classificadas como amostras normais. A Tabela 6.2 apresenta o resumo do quantitativo de classes por categoria alcançada nesse estudo.

	Classe Normal	Classe Anomalia
Classe Normal	155 (67%)	27 (33%)
Classe Anomalia	2 (1%)	153 (99%)

Tabela 6.2: Matriz de Confusão.

Dos dados submetidos ao algoritmo de classificação proposto, observamos que, de forma eficaz, conseguimos identificar 99% das amostras anômalas contra apenas 65% de amostras que pertencem à classe normal. Intuitivamente, apesar do contraste entre a classificação das classes normais e anômalas, entende-se que em um cenário de detecção de anomalias a atividade de identificar precisamente a classe minoritária (anomalias) faz-se mais importante.

	Precisão	Revocação	F1
Classe Normal	96,49%)	67,07%	79,14%
Classe Anomalia	85,00%	98,71%	91,34%

Tabela 6.3: Relatório de Classificação.

Outro ponto que vale ainda mencionar é que quando confrontadas as taxas de precisão e revocação observa-se que a capacidade da abordagem de evitar falsos positivos é de 96,49% enquanto a capacidade de detecção das classes anômalas é de 98,71%. Isso significa que, apesar de termos uma quantidade menor de classes de anomalia, quando comparadas com a classe normal, o algoritmo proposto consegue classificar de forma eficiente e respeitar a separabilidade das classes. Isso pode ser ratificado ainda pelas curvas ROC apresentada nas Figuras 6.7 e 6.8.

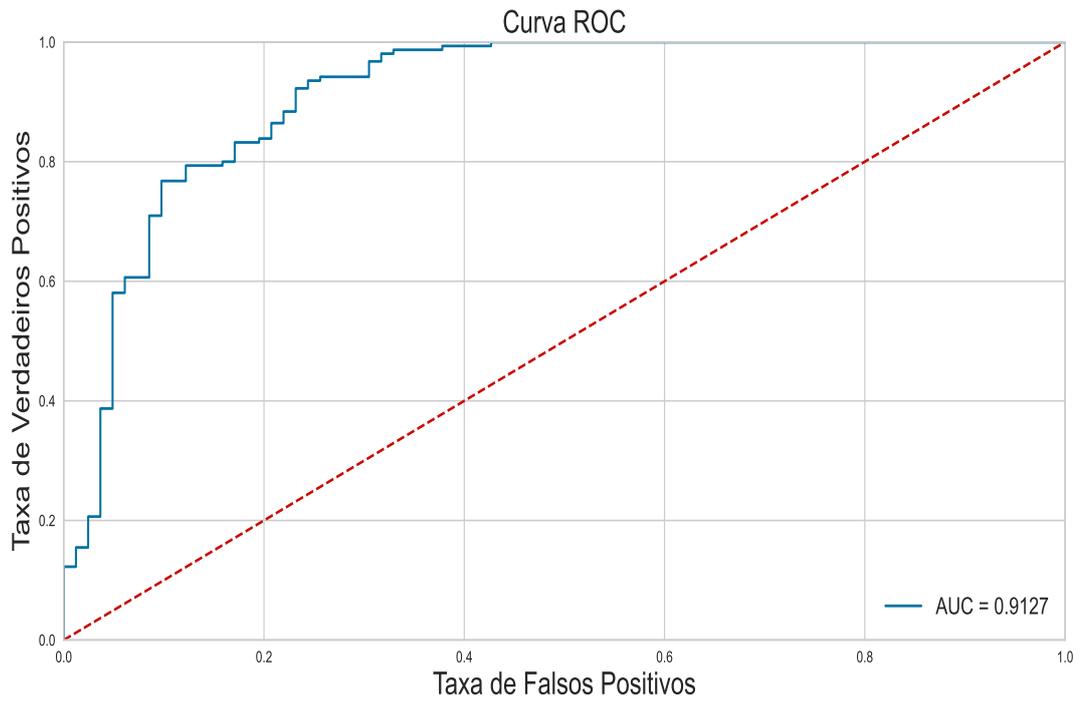


Figura 6.7: Curva ROC.

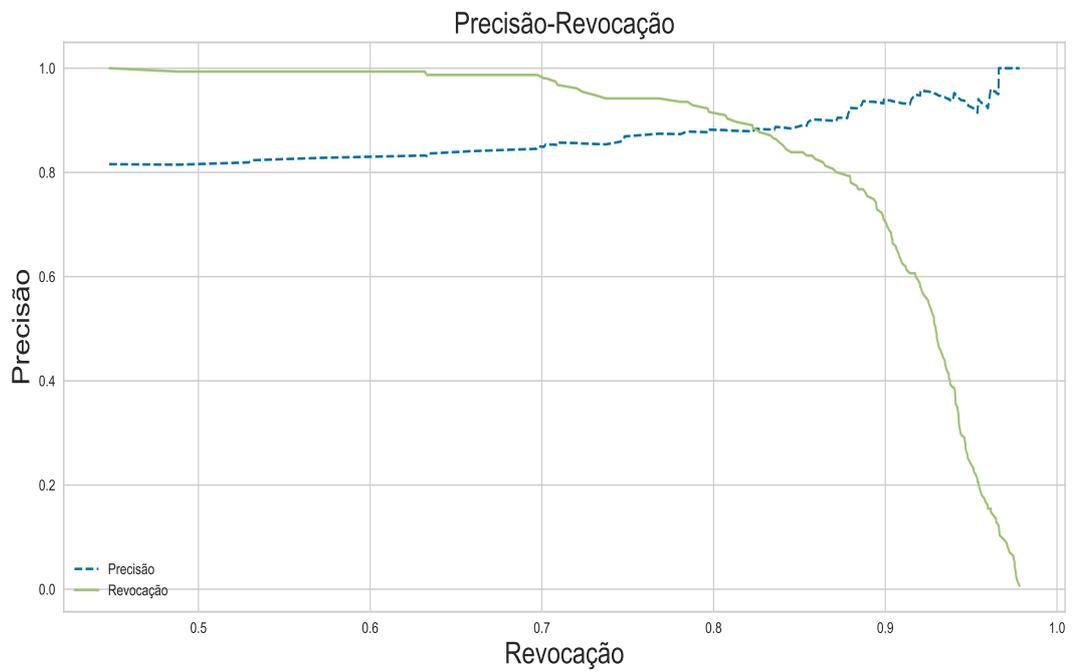


Figura 6.8: Precisão x Revocação.

Ainda para validar nossos resultados e deixar nossas previsões mais assertivas, foram calculadas outras métricas apresentadas na Tabela 4.1 que entendemos serem relevantes para a pesquisa, visto que tais métricas propiciam mais fidedignidade e completude ao procedimento de detecção de anomalias. As métricas apresentam o resultado imediato do algoritmo de classificação (coluna **Resultado**) e posteriormente, cada uma foi otimizada para obtermos o ponto ótimo na detecção de anomalias.

A otimização das métricas ocorreu identificando-se o melhor limiar de classificação dentro de um intervalo de valores de limiar possíveis. Ajustar ou mudar o limiar de decisão para abranger os requisitos mais amplos em um problema de classificação é geralmente referido como “movimentação de limiar”, “ajuste de limiar”. Foi afirmado que tentar outros métodos, como amostragem, sem tentar simplesmente definir o limite pode ser enganoso. O método de movimentação de limite usa o conjunto de treinamento original para treinar [um modelo] e, em seguida, move o limite de decisão de modo que os exemplos de classe minoritária sejam mais fáceis de serem previstos corretamente [34].

De forma geral, o processo envolve primeiro ajustar o modelo em um conjunto de dados de treinamento e fazer previsões em um conjunto de dados de teste. As previsões estão na forma de probabilidades normalizadas ou pontuações transformadas em probabilidades normalizadas. Diferentes valores de limiar são então testados e os rótulos definidos são avaliados usando a métrica de avaliação escolhida. O limiar que atinge a melhor métrica de avaliação é então adotado para o modelo.

Métrica	Resultado	Resultado - Limiar Ótimo
Recall	85,00%	100,00%
F1-Score	91,34%	89,85%
F-Beta Score (<i>beta=2</i>)	87,42%	95,67%
Matthew Corr. Coef.	73,21%	68,38%
Balanced Accuracy Score	90,74%	78,65%
Brier Score	12,23%	14,76%
Kappa Coefficient	70,87%	70,87%
F-Beta (Macro) (<i>beta=2</i>)	88,07%	79,17%
F-Beta (Weight) (<i>beta=2</i>)	87,74%	84,26%

Tabela 6.4: Métricas do Modelo.

A acurácia geral da abordagem assinalou 87,76% enquanto a área sobre a curva ROC foi de 91,27%. Comparados com os resultados obtidos na Tabela 6.4, nossos resultados se demonstram superiores ratificando a abordagem proposta.

Dataset	Método	ACC	F1	F2	AUC
SpaceNet 7 [1]	. Abordagem Proposta	87,76%	91,34%	100%	91.27%
ZY-3	. Object-based Segmentation [36]	92,00%	90,00%	-	-
SpaceNet 5 [2]	4 x ResNet50 + 4 x SE-ResNeXt50 [75]	-	69,00%	-	-
SpaceNet 5 [2]	Fuzzy C-means Clustering [44]	85,00%	-	-	-
Worldview-3	Change Vector Analysis [32]	90,99%	-	-	70,00%
Google Earth	Encoder-Decoder CNN [47]	-	71,16%	75,21%	-
AVIRIS [28]	Deep Belief Network [51]	-	-	-	99,80%
Terra/MODIS	SARIMA [81]	88,68%	-	-	-

Tabela 6.5: Sumário dos resultados da Literatura Analisada e resultados do método proposto nesta pesquisa.

A importância de calcular as métricas considerando o limiar ótimo deve-se ao fato de que estamos lidando com um conjunto de dados desbalanceado e considerando a perspectiva de detecção de anomalias, a identificação da classe minoritária faz-se mais importante que a detecção das classes normais. Logo, métricas como revocação (*recall*) e o cálculo da pontuação F-beta, que é a média harmônica ponderada de precisão e *recall*, faz-se tão importante nesse cenários

Capítulo 7

Considerações Finais

Neste capítulo, serão apresentadas as conclusões, trabalhos futuros e produção científica desta dissertação.

7.1 Conclusões

Detectar mudanças nas imagens aéreas é uma tarefa difícil, muito complicada pelos efeitos sazonais, atmosféricos e de iluminação. No entanto, a capacidade de localizar e rastrear as mudanças temporalmente é uma faceta importante em uma série de aplicações, desde a resposta a desastres até a preparação para doenças e o monitoramento ambiental, além de ser uma tarefa desafiadora para a comunidade de visão computacional.

Esta pesquisa propôs um método semi-supervisionado eficiente em complexidade computacional e precisão para detecção de alterações usando as informações do mapa de características extraídas de pares de imagens relacionadas usando uma CNN treinada para segmentação semântica de imagens. O estudo empírico confirmou que a utilização do algoritmo (POT) resulta na capacidade de detectar alterações com precisão de até 87,76% e uma curva ROC de 91,27%. Assim, a abordagem proposta consegue evidenciar o fator discriminativo das características extraídas por uma CNN identificando posteriormente anomalias através de uma abordagem probabilística paramétrica. Logo, a utilização de informações de representações de imagens extraídas de uma CNN para a detecção de alterações oferece vários benefícios. Em primeiro lugar, o método proposto é semi-supervisionado, oferecendo uma flexibilidade tanto no seu uso quanto para a base de dados que se possa utilizar. Em segundo

lugar, o uso de um modelo CNN preestabelecido oferece uma solução computacionalmente eficiente em comparação com as técnicas clássicas de detecção de alterações. A capacidade da arquitetura CNN de extrair características essenciais das imagens de satélite torna a abordagem proposta robusta à detecção de mudanças. A precisão da detecção de alterações depende em grande parte da eficácia da segmentação semântica do modelo subjacente, portanto, uma melhor segmentação de imagem CNN pode produzir melhores resultados de detecção de alterações.

7.2 Trabalhos Futuros

A pesquisa sobre a detecção de anomalias em imagens de satélite é uma área desafiadora, porém com amplos benefícios para a comunidade de forma geral, conseqüentemente existem algumas questões que não foram totalmente exploradas, portanto, merecem atenção e continuidade da pesquisa, entre as quais:

- **Comparação com outras abordagens de detecção de anomalias em séries temporais.**

Faz-se relevante a comparação do método **POT** utilizado nesta pesquisa com outros métodos já consolidados pela comunidade (SARIMA, SARIMAX, etc.), além de uma melhor seleção de variáveis.

- **Inspeção e Causalidade**

O estudo da causalidade em aprendizado profundo tem ganho considerável atenção da comunidade acadêmica [4; 55; 56]. Identificar as possíveis causas da classificação de uma amostra específica como anomalia, pode contribuir para a interpretabilidade na identificação de anomalias em séries temporais, além de permitir replicar, perceber e interpretar fatores exógenos e endógenos às séries temporais utilizando imagens de satélite.

- **Aplicação de outras arquiteturas de rede na extração de características**

Comparar e entender como que outras arquiteturas de rede já consolidadas pela comunidade (HRNet, FastFCN, Gated-SCNN, etc.) conseguem evidenciar o fator discriminativo das representações de características extraídas das imagens de satélite.

- **Realização de testes estatísticos e intervalos de confiança**

Faz-se relevante validar a significância estatística da abordagem. A significância estatística ajuda a quantificar se o resultado da abordagem proposta é devido ao acaso ou a algum fator de interesse.

Bibliografia

- [1] The spacenet partners, “sn7: Multi-temporal urban development challenge ”, <https://spacenet.ai/sn7-challenge/>. Acessado: 21 mar 2022.
- [2] The spacenet partners, “spacenet5: Automated road network extraction and route travel time estimation from satellite imagery”, <https://spacenet.ai/sn5-challenge/>. Acessado: 17 set 2022.
- [3] Asim Anees and Jagannath Aryal. A statistical framework for near-real time detection of beetle infestation in pine forests using modis data. *IEEE Geoscience and Remote Sensing Letters*, 11(10):1717–1721, 2014.
- [4] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion*, 58:82–115, 2020.
- [5] August Aimé Balkema and Laurens De Haan. Residual life time at great age. *The Annals of probability*, pages 792–804, 1974.
- [6] Vic Barnett and Toby Lewis. *Outliers in statistical data*. Wiley, 1974.
- [7] Vladimir Belitsky and Francisco Martins Moreira. A thorough yet simple exposition of the peaks-over-threshold method and its employment for. 2012.
- [8] Christopher Michael Bishop. Novelty detection and neural network validation. *IEE Proceedings-Vision, Image and Signal processing*, 141(4):217–222, 1994.

-
- [9] Christopher Michael Bishop. *Pattern recognition and machine learning*. springer, 2006.
- [10] Sergi Caelles, Jordi Pont-Tuset, Federico Perazzi, Alberto Montes, Kevis-Kokitsi Marinis, and Luc Van Gool. The 2019 davis challenge on vos: Unsupervised multi-object segmentation. *arXiv preprint arXiv:1905.00737*, 2019.
- [11] James B. Campbell and Randolph H. Wynne. *Introduction to remote sensing*. Guilford Press, 2011.
- [12] Hugo Carrão, Paulo Gonçalves, and Mario Caetano. A nonlinear harmonic model for fitting satellite image time series: Analysis and prediction of land cover dynamics. *IEEE Transactions on Geoscience and Remote Sensing*, 48(4):1919–1930, 2009.
- [13] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15, 2009.
- [14] Varun Chandola and Ranga Raju Vatsavai. A gaussian process based online change detection algorithm for monitoring periodic time series. In *Proceedings of the 2011 SIAM International Conference on Data Mining*, pages 95–106. SIAM, 2011.
- [15] Lei Chen. *Curse of Dimensionality*, pages 545–546. Springer US, Boston, MA, 2009.
- [16] C. Chow. On optimum recognition error and reject tradeoff. *IEEE Transactions on information theory*, 16(1):41–46, 1970.
- [17] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). *arXiv preprint arXiv:1511.07289*, 2015.
- [18] Barney Cohen. Urbanization in developing countries: Current trends, future projections, and key challenges for sustainability. *Technology in society*, 28(1-2):63–80, 2006.
- [19] Ilke Demir, Krzysztof Koperski, David Lindenbaum, Guan Pang, Jing Huang, Saikat Basu, Forest Hughes, Devis Tuia, and Ramesh Raskar. Deepglobe 2018: A challenge to parse the earth through satellite images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 172–181, 2018.

- [20] Patrick Dendorfer, Hamid Rezatofghi, Anton Milan, Javen Shi, Daniel Cremers, Ian Reid, Stefan Roth, Konrad Schindler, and Laura Leal-Taixé. Mot20: A benchmark for multi object tracking in crowded scenes. *arXiv preprint arXiv:2003.09003*, 2020.
- [21] M.J. Desforges, P.J. Jacob, and J.E. Cooper. Applications of probability density estimation to the detection of abnormal conditions in engineering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, 212(8):687–703, 1998.
- [22] Ronald Aylmer Fisher and Leonard Henry Caleb Tippett. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge University Press, 1928.
- [23] Clive S Fraser, Alex Woods, and Danny Brizzi. Hyper redundancy for accuracy enhancement in automated close range photogrammetry. *The Photogrammetric Record*, 20(111):205–217, 2005.
- [24] Maurice Fréchet. Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Math. Polon.*, 6:93–116, 1927.
- [25] Jerome Harold Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.
- [26] D.N. Geary. Mixture models: Inference and applications to clustering. *Journal of the Royal Statistical Society Series A*, 152(1):126–127, 1989.
- [27] Boris Gnedenko. Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of mathematics*, pages 423–453, 1943.
- [28] Robert O. Green, Michael L. Eastwood, Charles M. Sarture, Thomas G. Chrien, Mikael Aronsson, Bruce J. Chippendale, Jessica A. Faust, Betina E. Pavri, Christopher J. Chovit, Manuel Solis, et al. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (aviris). *Remote sensing of environment*, 65(3):227–248, 1998.

- [29] Veronika Gstaiger, Jiaojiao Tian, Ralph Kiefl, and Franz Kurz. 2d vs. 3d change detection using aerial imagery to support crisis management of large-scale events. *Remote Sensing*, 10(12):2054, 2018.
- [30] Ritwik Gupta, Bryce Goodman, Nirav Patel, Ricky Hosfelt, Sandra Sajejev, Eric Heim, Jigar Doshi, Keane Lucas, Howie Choset, and Matthew Gaston. Creating xbd: A dataset for assessing building damage from satellite imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 10–17, 2019.
- [31] Laurens Haan. A form of regular variation and its application to the domain of attraction of the double exponential distribution. *Probability Theory and Related Fields*, 17(3):241–258, 1971.
- [32] Youkyung Han, Sejung Jung, Sicong Liu, and Junho Yeom. Effect analysis in the fine co-registration of very-high-resolution satellite images for unsupervised change detection. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pages 1558–1561. IEEE, 2019.
- [33] Matthew C. Hansen and Thomas R. Loveland. A review of large area monitoring of land cover change using landsat data. *Remote sensing of Environment*, 122:66–74, 2012.
- [34] Haibo He and Yunqian Ma. *Imbalanced learning: foundations, algorithms, and applications*. Wiley-IEEE Press, 2013.
- [35] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [36] Xin Huang, Yinxia Cao, and Jiayi Li. An automatic change detection method for monitoring newly constructed building areas using time-series multi-view high-resolution optical satellite images. *Remote Sensing of Environment*, 244:111802, 2020.
- [37] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei

- Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30:3146–3154, 2017.
- [38] Robert E. Kennedy, Philip A. Townsend, John E. Gross, Warren B. Cohen, Paul Bolsstad, YQ Wang, and Phyllis Adams. Remote sensing change detection tools for natural resource managers: Understanding concepts and tradeoffs in the design of landscape monitoring projects. *Remote sensing of environment*, 113(7):1382–1396, 2009.
- [39] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [40] Waldo Kleynhans, Jan C. Olivier, Konrad J. Wessels, Brian P. Salmon, Frans Van den Bergh, and K Steenkamp. Detecting land cover change using an extended kalman filter on modis ndvi time-series data. *IEEE Geoscience and Remote Sensing Letters*, 8(3):507–511, 2010.
- [41] Waldo Kleynhans, Brian P. Salmon, Jan C. Olivier, Frans van den Bergh, Konrad J. Wessels, and T. Grobler. Detecting land cover change using a sliding window temporal autocorrelation approach. In *2012 IEEE International Geoscience and Remote Sensing Symposium*, pages 6765–6768. IEEE, 2012.
- [42] Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada, 1995.
- [43] Caitlin Kontgis, Annemarie Schneider, Jefferson Fox, Sumeet Saksena, James H. Spencer, and Miguel Castrence. Monitoring peri-urbanization in the greater ho chi minh city metropolitan area. *Applied Geography*, 53:377–388, 2014.
- [44] Dongwei Li, Shuliang Wang, Nan Gao, Qiang He, and Yun Yang. Cost-effective land cover classification for remote sensing images. *arXiv preprint arXiv:2107.12016*, 2021.
- [45] Wei Li, Guodong Wu, and Qian Du. Transferred deep learning for anomaly detection in hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters*, 14(5):597–601, 2017.
- [46] Thomas Lillesand, Ralph W. Kiefer, and Jonathan Chipman. *Remote sensing and image interpretation*. John Wiley & Sons, 2015.

- [47] Kyungsun Lim, Dongkwon Jin, and Chang-Su Kim. Change detection in high resolution satellite images using an ensemble of convolutional neural networks. In *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 509–515. IEEE, 2018.
- [48] Guosheng Lin, Chunhua Shen, Anton Van Den Hengel, and Ian Reid. Efficient piecewise training of deep structured models for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3194–3203, 2016.
- [49] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.
- [50] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [51] Ning Ma, Yu Peng, Shaojun Wang, and Philip H.W. Leong. An unsupervised deep hyperspectral anomaly detector. *Sensors*, 18(3):693, 2018.
- [52] Markos Markou and Sameer Singh. Novelty detection: a review—part 1: statistical approaches. *Signal processing*, 83(12):2481–2497, 2003.
- [53] Santiago Martínez, Juan Ortiz, Ma Luz Gil, and Ma Teresa Rego. Recording complex structures using close range photogrammetry: The cathedral of santiago de compostela. *The Photogrammetric Record*, 28(144):375–395, 2013.
- [54] Edward M. Mikhail, James S. Bethel, and J. Chris McGlone. Introduction to modern photogrammetry. *New York*, 2001.
- [55] Shane T. Mueller, Robert R. Hoffman, William Clancey, Abigail Emrey, and Gary Klein. Explanation in human-ai systems: A literature meta-review, synopsis of key ideas and publications, and bibliography for explainable ai. *arXiv preprint arXiv:1902.01876*, 2019.

- [56] Tanmayee Narendra, Anush Sankaran, Deepak Vijaykeerthy, and Senthil Mani. Explaining deep learning models using causal inference. *arXiv preprint arXiv:1811.04376*, 2018.
- [57] United Nations. *World Population Prospects 2019: Highlights*. 2019.
- [58] Charlotte Pelletier, Geoffrey I. Webb, and François Petitjean. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing*, 11(5):523, 2019.
- [59] James Pickands III et al. Statistical inference using extreme order statistics. *the Annals of Statistics*, 3(1):119–131, 1975.
- [60] Marco A.F. Pimentel, David A. Clifton, Lei Clifton, and Lionel Tarassenko. A review of novelty detection. *Signal Processing*, 99:215–249, 2014.
- [61] P. K. S. Prakash and Achyutuni Sri Krishna Rao. *R Deep Learning Cookbook: Solve Complex Neural Net Problems with TensorFlow, H2O and MXNet*. Packt Publishing, 2017.
- [62] Gunter Ritter and María Teresa Gallegos. Outliers in statistical pattern recognition and an application to automatic chromosome classification. *Pattern Recognition Letters*, 18(6):525–539, 1997.
- [63] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [64] Diego Rybski, Anselmo Garcia Cantu Ros, and Jürgen P Kropp. Distance-weighted city growth. *Physical Review E*, 87(4):042114, 2013.
- [65] Rob Saunders and John S. Gero. The importance of being emergent. In *Proceedings of Artificial Intelligence in Design*, 2000.
- [66] Robert A. Schowengerdt. *Remote sensing: models and methods for image processing*. Elsevier, 2006.

- [67] Atharva Sharma, Xiuwen Liu, and Xiaojun Yang. Land cover classification from multi-temporal, multi-spectral remotely sensed imagery using patch-based recurrent neural networks. *Neural Networks*, 105:346–355, 2018.
- [68] Jacob Shermeyer, Daniel Hogan, Jason Brown, Adam Van Etten, Nicholas Weir, Fabio Pacifici, Ronny Hansch, Alexei Bastidas, Scott Soenen, Todd Bacastow, et al. Spacenet 6: Multi-sensor all weather mapping dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 196–197, 2020.
- [69] Alban Siffer, Pierre-Alain Fouque, Alexandre Termier, and Christine Largouet. Anomaly detection in streams with extreme value theory. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1067–1075. ACM, 2017.
- [70] Sameer Singh and Markos Markou. An approach to novelty detection applied to the classification of image regions. *IEEE Transactions on Knowledge and Data Engineering*, 16(4):396–407, 2004.
- [71] David W.J. Stein, Scott G. Beaven, Lawrence E. Hoff, Edwin M. Winter, Alan P. Schaum, and Alan D. Stocker. Anomaly detection from hyperspectral imagery. *IEEE signal processing magazine*, 19(1):58–69, 2002.
- [72] Lionel Tarassenko, Paul Hayton, Nicholas Cerneaz, and Michael Brady. Novelty detection for the identification of masses in mammograms. 1995.
- [73] David M.J. Tax and Robert P.W. Duin. Outlier detection using classifier instability. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, pages 593–601. Springer, 1998.
- [74] Charles Toth and Grzegorz Józków. Remote sensing platforms and sensors: A survey. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115:22–36, 2016.
- [75] Adam Van Etten, Dave Lindenbaum, and Todd M. Bacastow. Spacenet: A remote sensing dataset and challenge series. *arXiv preprint arXiv:1807.01232*, 2018.

- [76] Anthony J. Venables. *Breaking into tradables: Urban form and urban function in a developing city*. The World Bank, 2017.
- [77] Jan Verbesselt, Rob Hyndman, Glenn Newnham, and Darius Culvenor. Detecting trend and seasonal changes in satellite image time series. *Remote sensing of Environment*, 114(1):106–115, 2010.
- [78] Jan Verbesselt, Rob Hyndman, Achim Zeileis, and Darius Culvenor. Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. *Remote Sensing of Environment*, 114(12):2970–2980, 2010.
- [79] Jan Verbesselt, Achim Zeileis, and Martin Herold. Near real-time disturbance detection using satellite image time series. *Remote Sensing of Environment*, 123:98–108, 2012.
- [80] Nicholas Weir, David Lindenbaum, Alexei Bastidas, Adam Van Etten, Sean McPherson, Jacob Shermeyer, Varun Kumar, and Hanlin Tang. Spacenet mvoi: A multi-view overhead imagery dataset. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 992–1001, 2019.
- [81] Zeng-Guang Zhou, Ping Tang, and M. Zhou. Detecting anomaly regions in satellite image time series based on seasonal autocorrelation analysis. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3:303, 2016.
- [82] Zengguang Zhou, Ping Tang, and Zheng Zhang. A method for monitoring land-cover disturbance using satellite time series images. In *Land Surface Remote Sensing II*, volume 9260, page 926038. International Society for Optics and Photonics, 2014.
- [83] Zhe Zhu, Curtis E. Woodcock, and Pontus Olofsson. Continuous monitoring of forest disturbance using all available landsat imagery. *Remote sensing of environment*, 122:75–91, 2012.

Apêndice A

Arquitetura da Rede Convolutacional

U-net

A U-net foi desenvolvida por Olaf Ronneberger et al. [63] para a segmentação de imagens biomédicas. A arquitetura de rede é ilustrada na Figura 1. Ela consiste em um caminho de codificação (lado esquerdo) e um caminho decodificação (lado direito). O caminho de codificação segue a arquitetura típica de uma rede convolutacional. Consiste na aplicação repetida de duas convoluções 3x3 (convoluções sem preenchimento), cada uma seguida por uma unidade linear retificada (ReLU) e uma operação de *max pooling* 2x2 com passo 2 para *downsampling*. Em cada etapa de redução da amostra, dobra-se o número de canais de características.

Cada passo no caminho expansivo consiste em um *upsampling* do mapa de características seguido por uma convolução 2x2 (“convolução ascendente”) que reduz pela metade o número de canais de características, uma concatenação com o mapa de características correspondentemente recortado do caminho de codificação e duas convoluções 3x3, cada uma seguida por uma ReLU. O recorte é necessário devido à perda de pixels da borda em cada convolução. Na camada final, uma convolução 1x1 é usada para mapear cada vetor de características de 64 componentes para o número desejado de classes. No total, a rede possui 23 camadas convolucionais. Assim, é uma rede totalmente convolutacional (FCN) de ponta a ponta, ou seja, contém apenas camadas convolucionais e não contém nenhuma camada densa.

O diagrama apresentado na Figura A.1 representa a estrutura completa da rede U-net.

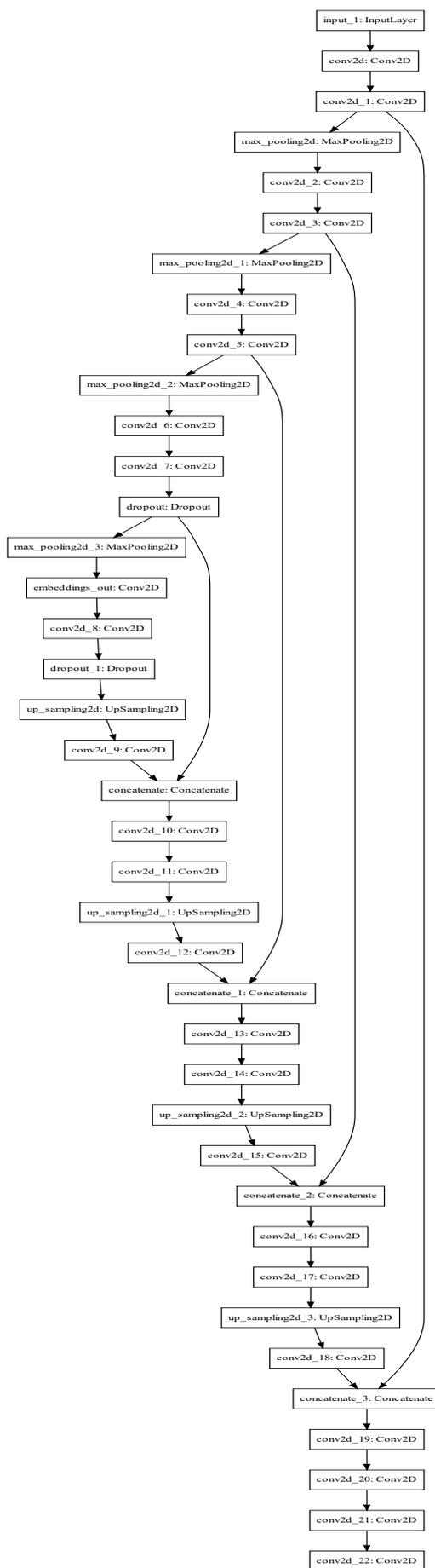


Figura A.1: Arquitetura U-net.
Fonte: Pesquisa do Autor.

Apêndice B

Algoritmo POT

Em termos práticos, o objetivo principal do algoritmo é receber n observações X_1, \dots, X_n , e uma variável de risco fixo representada por q , cuja função é controlar o número de falsos positivos. O objetivo é calcular um primeiro limite z_q verificando $P(X > z_q) < q$. A Figura 5.2 representa graficamente o que é feito neste lote inicial (calibração). A ideia básica é estabelecer um limiar inicial u (linha 2), recuperar os picos (os excessos que ultrapassam u). O conjunto Y_u é o conjunto de picos em que armazenamos os excessos observados em u (linha 3), posteriormente ajusta-se uma GPD para esses conjuntos de picos (linha 4). Por fim, para inferir-se a distribuição dos valores extremos e realiza-se o cálculo do limite z_q (linha 5). O ajuste da GPD é realizado com o truque de Grimshaw, sendo z_q calculado conforme a Equação 4.

Algoritmo 1 POT - (Peaks-over-Threshold)

```
1: procedure POT( $X_1, \dots, X_n, q$ )
2:    $u \leftarrow \text{SetInitialThreshold}(X_1, \dots, X_n, q)$ 
3:    $Y_u \leftarrow \{X_i - u \mid X_i > u\}$ 
4:    $\hat{\gamma} \hat{\sigma} \leftarrow \text{Grimshaw}(Y_u)$ 
5:    $z_q \leftarrow \text{CalcThreshold}(q, \hat{\gamma}, \hat{\sigma}, n, N_u, u)$ 
6:   return  $z_q, u$ 
```
