UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

FABRICIO BIZOTTO

REDES NEURAIS CONVOLUCIONAIS NA SEGMENTAÇÃO SEMÂNTICA DE IMAGENS AÉREAS PARA O MAPEAMENTO DA COBERTURA DO SOLO EM ÁREAS DE PROTEÇÃO AMBIENTAL

PONTA GROSSA

2023

FABRICIO BIZOTTO

REDES NEURAIS CONVOLUCIONAIS NA SEGMENTAÇÃO SEMÂNTICA DE IMAGENS AÉREAS PARA O MAPEAMENTO DA COBERTURA DO SOLO EM ÁREAS DE PROTEÇÃO AMBIENTAL

Convolutional Neural Networks for Semantic Segmentation of Aerial Images in Land Cover Mapping of Environmental Protection Areas

Dissertação apresentada como requisito para obtenção do título de Mestre em Ciência da Computação do Programa de Pós-Graduação em Ciência da Computação da Universidade Tecnológica Federal do Paraná.

Orientadora: Profa. Dra. Mauren Louise Sguario Coelho de Andrade

PONTA GROSSA

2023

Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Ponta Grossa

FABRICIO BIZOTTO

REDES NEURAIS CONVOLUCIONAIS NA SEGMENTAÇÃO SEMÂNTICA DE IMAGENS AÉREAS PARA O MAPEAMENTO DA COBERTURA DO SOLO EM ÁREAS DE PROTEÇÃO AMBIENTAL

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestre Em Ciência Da Computação da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Sistemas E Métodos De Computação.

Data de aprovação: 21 de Novembro de 2023

Dra. Mauren Louise Sguario Coelho De Andrade, Doutorado - Universidade Tecnológica Federal do Paraná

Dra. Alaine Margarete Guimaraes, Doutorado - Universidade Estadual de Ponta Grossa (Uepg)

Dr. Gilson Antonio Giraldi, Doutorado - Laboratório Nacional de Computação Científica

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 22/11/2023.

Dedico este trabalho a minha família e aos
meus amigos, pelos momentos de ausência.

AGRADECIMENTOS

A minha família, cujo apoio incondicional sempre foi minha maior fortaleza. Vocês foram a base sólida que sustentou meus desafios e conquistas. Obrigado por sua paciência e compreensão durante os momentos intensos desta jornada acadêmica.

À minha respeitada orientadora. Sua expertise, paciência e sabedoria foram essenciais neste percurso acadêmico. Além disso, suas sugestões e críticas construtivas não apenas aprimoraram o conteúdo, mas também refinaram a abordagem metodológica, resultando em um trabalho mais robusto e significativo.

Ao professor Dr. Gilson, que contribuiu com sua expertise valiosa e insights enriquecedores, agradeço por sua orientação adicional que enriqueceu significativamente o desenvolvimento da minha dissertação. Sua generosidade em compartilhar conhecimentos foi essencial para a qualidade do trabalho.

Ao ICMBio, expresso minha gratidão pela oportunidade de colaborar e aprender com uma equipe tão dedicada à preservação ambiental. As experiências adquiridas junto a vocês enriqueceram minha compreensão prática e contextual do tema abordado na dissertação.

Cada um de vocês desempenhou um papel fundamental nesta jornada, e sou profundamente grato por ter compartilhado este desafio com indivíduos tão notáveis. Este trabalho é não apenas meu, mas também de todos vocês, que contribuíram de maneira única e significativa para o seu desenvolvimento.

Obrigado por fazerem parte deste capítulo importante da minha vida. Enfim, a todos os que de alguma forma contribuíram para a realização deste trabalho.

Primeira Lei: Um robô não pode ferir um ser humano ou, por omissão, permitir que um ser humano sofra algum mal. Segunda Lei: Um robô deve obedecer as ordens que lhe sejam dadas por seres humanos, exceto nos casos em que tais ordens contrariem a Primeira Lei. Terceira Lei: Um robô deve proteger sua própria existência desde que tal proteção não entre em conflito com a Primeira e Segunda Leis (ASIMOV, 1950).

RESUMO

No Brasil, a Área de Proteção Ambiental (APA) desempenha um papel fundamental na conservação ambiental e no equilíbrio entre o uso sustentável dos recursos naturais e o desenvolvimento socioeconômico local. A eficácia da gestão das atividades da APA requer um monitoramento abrangente do uso e cobertura do solo. O sensoriamento remoto surge como uma alternativa eficiente e econômica para o monitoramento dessas áreas. Nesse contexto, técnicas computacionais avançadas, como redes neurais convolucionais (RNCs), emergem como ferramentas promissoras. O estudo propõe uma metodologia de baixo custo para a segmentação semântica, utilizando imagens de sensoriamento remoto adquiridas na plataforma Google Earth para monitorar o uso e cobertura do solo na região da APA-Petrópolis, Rio de Janeiro, empregando RNCs. As arquiteturas SegNet e U-Net foram utilizadas para segmentação semântica, junto ao desenvolvimento de um banco de imagens aéreas para o território da APA-Petrópolis, utilizado no treinamento e teste dos modelos. Foram apresentados e discutidos quatro cenários de treinamento e teste com diferentes ajustes. Os resultados da análise indicam que o cenário 4, utilizando a rede U-NET com a função de custo *Focal Loss*, obteve a melhor acurácia global (0.87). Apesar disso, o cenário 3, que emprega a U-NET com a função de custo entropia cruzada, apresentou resultados comparáveis (0.87). Em termos de Índice de Jaccard (IoU), o cenário 3 se destacou com o melhor valor (0.72), enquanto o Cenário 4 mostrou-se próximo (0.71). Classes desafiadoras, como Solo Exposto, evidenciaram baixos índices de f1-score (0.31 a 0.52). A variação na função de custo entre cenários teve impacto limitado. Os resultados destacam a eficácia da U-NET e sugerem a necessidade de refinamentos contínuos, especialmente em classes mais desafiadoras.

Palavras-chave: rede neural convolucional; imagens aéreas; segmentação semântica; área de preservação ambiental.

ABSTRACT

In Brazil, the Environmental Protection Area (EPA) plays a key role in environmental conservation and in balancing the sustainable use of natural resources with local socio-economic development. Effective management of EPA activities requires comprehensive monitoring of land use and land cover. Remote sensing has emerged as an efficient and economical alternative for monitoring these areas. In this context, advanced computational techniques, such as convolutional neural networks (CNNs), are emerging as promising tools. This study proposes a low-cost methodology for semantic segmentation, using remote sensing images acquired on the Google Earth platform to monitor land use and land cover in the EPA-Petrópolis, Rio de Janeiro, using RNCs. The SegNet and U-Net architectures were employed for semantic segmentation, along with the development of an aerial image database for the EPA-Petrópolis region, used to training and testing of the models. Four training and testing scenarios with different settings were presented and discussed. The analysis results indicate that scenario 4, using the U-Net with the Focal Loss function, achieved the highest overall accuracy (0.87). However, scenario 3, employing the U-Net with the cross-entropy loss function, showed comparable results (0.87). In terms of the Jaccard Index (IoU), scenario 3 stood out with the best value (0.72), while Scenario 4 was close (0.71). Challenging classes, such as Exposed Soil, showed low f1-score indices (0.31 to 0.52). The variation in the loss function between scenarios had limited impact. The results highlight the effectiveness of the U-Net and suggest the need for continuous refinements, especially in more challenging classes.

Keywords: convolutional neural network; aerial images; semantic segmentation; environmental preservation area.

LISTA DE FIGURAS

LISTA DE QUADROS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

AP	Aprendizagem Profunda
APA	Área de Proteção Ambiental
ARP	Aeronave Remotamente Pilotada
CVAT	*Computer Vision Annotation Tool*
DWT	*Discrete Wavelet Transform*
EC	Entropia Cruzada
GE	*Google Earth*
ICMBio	Instituto Chico Mendes de Conservação da Biodiversidade
IoU	*Intersection Over Union*
IWT	*Inverse Discrete Wavelet Transform*
PDI	Processamento digital da imagem
PNG	*Portable Network Graphic*
ReLU	*Rectified Linear Units*
REM	Radiação Eletromagnética
RGB	*Red-Green-Blue*
RGB-IR	*Red-Green-Blue-Infrared*
RNA	Redes Neural Artificial
RNC	Rede Neural Convolucional
SGD	*Stochastic Gradient Descent*
SR	Sensoriamento Remoto
TIFF	*Tag Image File Format*
UCP	Unidade Central de Processamento
UFMS	Universidade Federal do Mato Grosso do Sul
UPG	Unidade de Processamento Gráfico

SUMÁRIO

1 INTRODUÇÃO

No Brasil, a legislação da Área de Proteção Ambiental (APA) é uma das principais ferramentas de conservação da natureza, estabelecida por lei e destinada à proteção ambiental e ao uso sustentável dos recursos naturais. Seu objetivo é conciliar a conservação do meio ambiente com o desenvolvimento econômico e social das comunidades que habitam a região (CORREIA; LEAO, 2015). As APAs são gerenciadas por órgãos governamentais, como o Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio), e contam com a participação da sociedade civil em sua gestão. As atividades humanas são regulamentadas de acordo com as características ambientais da região, permitindo apenas aquelas que não colocam em risco a biodiversidade, os recursos naturais e a qualidade de vida da população local (ECO, 2015).

A APA-Petrópolis está situada na região serrana do estado do Rio de Janeiro, cobrindo uma vasta extensão de cerca de 60.000 hectares. Inaugurada em 1982, a APA-Petrópolis tem como missão principal preservar a biodiversidade local, assegurando simultaneamente o uso sustentável dos recursos naturais e o bem-estar da comunidade circundante. A administração dessa área é conduzida pelo ICMBio, responsável pela implementação de iniciativas de conservação, fiscalização e monitoramento ambiental. Dentre as diversas atividades desempenhadas pelo ICMBio na APA-Petrópolis, destacam-se a manutenção de trilhas ecológicas, a restauração de áreas degradadas, o combate efetivo a incêndios florestais e a proteção ativa de espécies ameaçadas de extinção, conforme relatado por (BRASIL, 2023). No contexto específico desta APA, a eficiência e eficácia dessas iniciativas poderiam ser significativamente aprimoradas por meio do monitoramento contínuo da cobertura e do uso do solo na região. Essa abordagem permitiria uma gestão mais informada, possibilitando a organização e o gerenciamento otimizado das atividades, alinhados com os objetivos de conservação e sustentabilidade estabelecidos para a área.

O uso e cobertura do solo constituem dois conceitos fundamentais no campo da geografia e da ciência ambiental para monitorar a distribuição de diferentes tipos de vegetação, solo, floresta, recursos hídricos e outros elementos em uma determinada área. A cobertura do solo refere-se às coberturas físicas e biológicas da superfície terrestre, tais como florestas, pastos, zonas úmidas e áreas urbanas, enquanto que o uso do solo se refere às atividades humanas que têm lugar no solo, tais como a agricultura, silvicultura, mineração e desenvolvimento urbano. Compreender a relação entre cobertura e uso do solo é essencial para gerir os recursos naturais, mitigar as alterações climáticas e proteger a biodiversidade (ZIN; LIN, 2018).

O desenvolvimento socioeconômico dos seres humanos tem sido fortemente apoiado pelo uso do solo, sendo a urbanização um dos principais exemplos de mudanças de ocupação do solo em todo o mundo (LIU, 2018). Durante esse processo, as atividades humanas são as principais responsáveis pela mudança no uso do solo, assim como pela elaboração de políticas associadas a essa questão. Essas mudanças antropogênicas abrangem o desflorestamento, desocupação, urbanização, alterações nos tipos de cultivo e adaptações nas práticas utilizadas

em cada uso do solo, tais como, técnicas de plantio e sistemas de rotação de cultura florestal (PETERSON *et al.*, 2014). Assim, as políticas que regulamentam a utilização e administração do solo assumem uma função importante no processo de identificação, avaliação e acompanhamento das transformações na dinâmica da paisagem, além de atuarem na redução dos danos causados pela degradação do solo (GERLAK, 2014).

Entretanto, o monitoramento e a análise da utilização do solo não é uma tarefa simples. Entre os principais desafios estão a necessidade de grandes equipes de trabalho especializada, o deslocamento à regiões de difícil acesso, o alto custo para manutenção das equipes, além de perigos associados as características de fauna e flora de cada região. Portanto, faz-se necessário a utilização de ferramentas tecnológicas que possam contribuir para o monitoramento destas regiões, como por exemplo, os Sistemas de Informações Geográficas (GIS). O GIS engloba diversas técnicas para mapeamento da cobertura do solo. Nesse contexto, a classificação de imagens da região de interesse pode ser realizada através da interpretação visual, fundamentada no conhecimento local. Contudo, é importante destacar que esse procedimento frequentemente demanda um investimento considerável em licenças de *software* e um alto nível de proficiência em sua utilização (HUTH *et al.*, 2012).

Diante destes desafios, a área de sensoriamento remoto surge como uma alternativa eficiente e de baixo custo para auxiliar o monitoramento das APAs. Dentre as principais vantagens estão o acesso gratuito a base de dados de imagens de diferentes regiões, o acompanhamento temporal das áreas de estudo, além da cobertura de extensas áreas, bem como o acesso às áreas que não são possíveis via solo.

Diversas fontes de dados de sensoriamento remoto oferecem imagens valiosas, como satélites, câmeras em Aeronave Remotamente Pilotada (ARP) e plataformas online, contudo, as imagens de satélite gratuitas frequentemente apresentam baixa resolução. Por outro lado, o uso de ARP está sujeito às condições meteorológicas para a aquisição de imagens. Uma alternativa viável pode ser a utilização de imagens disponíveis na internet, como as imagens aéreas em RGB (*Red-Green-Blue*) fornecidas pela plataforma Google Earth. Essa abordagem pode representar uma estratégia mais acessível para aplicações de baixo custo.

Lançado em 2005, o GE se tornou um dos globos virtuais mais populares, amplamente utilizado para ensino e pesquisa na ciência geográfica, especialmente em estudos de formas e processos relativos a paisagem (BOARDMAN, 2016).

A medida que novas imagens de sensoriamento remoto são disponibilizadas, novas ferramentas computacionais são desenvolvidas a fim de lidar com o mapeamento de uso e cobertura do solo. Entre elas, técnicas de aprendizagem profunda (AP, do inglês *Deep Learning*) começaram a ser amplamente utilizadas pela comunidade de sensoriamento remoto (DUHL; GUENTHER; HELMIG, 2012; TORRES-SáNCHEZ *et al.*, 2013). Rede Neural Convolucional (RNC) surge como uma ferramenta promissora para a análise de imagens de sensoriamento remoto (CHEN *et al.*, 2021). As RNCs são um tipo de algoritmo de aprendizagem profunda que pode aprender a reconhecer padrões em grandes conjuntos de dados, tornando-as bem ade-

quadas para tarefas de classificação de imagens, como o mapeamento da cobertura do solo. Vários estudos exploraram a aplicação das RNC na análise da cobertura do solo com resultados promissores (HU *et al.*, 2013; LI *et al.*, 2020).

Neste contexto, este trabalho propõe o desenvolvimento de uma metodologia de baixo custo para a segmentação semântica, utilizando imagens de sensoriamento remoto adquiridas na plataforma GE a fim de monitorar o uso e cobertura do solo na região da APA-Petrópolis, Rio de Janeiro, por meio de RNC. Para tanto, as redes do tipo SegNet e do tipo U-NET serão utilizadas para avaliar a eficácia da segmentação semântica em imagens RGB de baixa resolução, juntamente com o desenvolvimento de um banco de dados com imagens da APA-Petrópolis. Assim, será possível analisar como a segmentação semântica pode oferecer informações confiáveis que indique como o solo está sendo utilizado, a fim de contribuir na tomada de decisão no que se refere à gestão adequada dos recursos naturais envolvidos. Isso permitirá a criação de orientações e estratégias baseadas em informações espaciais de uma determinada região, fundamentando o planejamento, a gestão e a governança necessários para promover o desenvolvimento sustentável. Além disso, possibilitará que sejam elaboradas novas recomendações para a construção de infraestrutura urbana adequada, levando em conta as particularidades do ecossistema envolvido, a fim de incentivar a recuperação, conservação e preservação de seu equilíbrio ecológico. Os resultados deste estudo permitirão avaliar o potencial desta estratégia para mapeamento do uso e cobertura do solo para enfrentar desafios ambientais.

1.1 Objetivos

Objetivos a serem alcançados foram definidos para a resolução do problema deste estudo, divididos em um objetivo geral e três objetivos específicos.

1.1.1 Objetivo geral

O objetivo geral deste estudo consiste em desenvolver uma metodologia de baixo custo para a segmentação semântica de imagens de sensoriamento remoto do *Google Earth* na região da APA-Petrópolis, no estado do Rio de Janeiro, por meio de Redes Neurais Convolutivas.

1.1.2 Objetivos específicos

Para atender o objetivo geral deste trabalho, foram propostos os seguintes objetivos específicos:

- Capturar e rotular imagens de sensoriamento remoto da área de proteção ambiental de Petrópolis, Rio de Janeiro;

- Efetuar uma análise do desempenho dos modelos, utilizando métricas apropriadas;
- Examinar as abordagens adotadas nos resultados obtidos por estudos correlatos.

1.2 Contribuições do Trabalho

As principais contribuições deste trabalho para a literatura científica são:

1. Desenvolvimento de um conjunto de dados com imagens aéreas rotuladas da APA-Petrópolis, obtidas por meio da plataforma GE.
2. Implementação de uma abordagem de baixo custo para o monitoramento do uso e cobertura do solo em áreas de proteção ambiental usando apenas imagens aéreas RGB.
3. Aplicação de técnicas de balanceamento de classes no contexto de sensoriamento remoto, um tópico pouco explorado na literatura.
4. Avaliação de cenários de treinamento e teste por meio da comparação de duas redes neurais convolucionais com diferentes configurações.

Essas contribuições não apenas ampliam o conhecimento existente na área, mas também abrem novas perspectivas para a utilização de dados de sensoriamento remoto e técnicas de processamento de imagens em áreas de proteção ambiental. Além disso, oferece uma base sólida para futuras investigações e aplicações práticas.

2 REFERENCIAL TEÓRICO

Este capítulo descreve os principais conceitos teóricos que fundamentam o desenvolvimento deste trabalho. A seção 2.1 apresenta os conceitos fundamentais sobre o processamento de imagem e suas etapas. A seção 2.1.2 introduz conceitos de segmentação de imagem. A seção 2.3 conceitua os principais tipos de aprendizagem de máquina. A seção 2.4 faz uma breve introdução às redes neurais artificiais. Nas seções 2.5 e 2.6, são descritos os conceitos fundamentais sobre aprendizagem profunda e redes neurais convolucionais. Por fim, as seções 2.7 e 2.8 introduz matematicamente as métricas usadas na etapa de treinamento e teste dos modelos para avaliação dos resultados.

2.1 Processamento Digital de Imagem

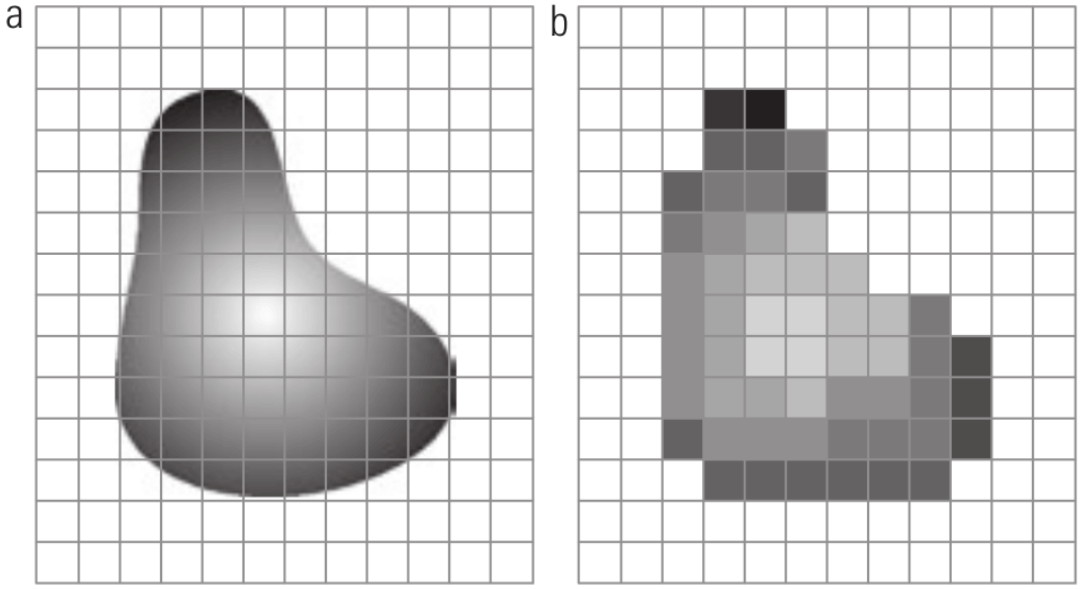
Uma imagem digital é criada através do processo conhecido como digitalização, que envolve dois passos principais: amostragem e quantização. A amostragem refere-se à resolução espacial da imagem, determinada pela quantidade de linhas e colunas nas direções x e y, resultando em uma matriz de amostras de tamanho $M \times N$. Por sua vez, a quantização consiste em atribuir valores discretos de cor ou nível de cinza a cada ponto da imagem. Essa quantidade de níveis de cinza vai definir a profundidade de bits da imagem digital (PEDRINI; SCHWARTZ, 2008).

Para Gonzalez e Woods (2010), imagem digital pode ser definida como uma função bidimensional $f(x,y)$, sendo x e y coordenadas do plano e a amplitude de f é chamada de intensidade, para qualquer par de coordenadas. Esta é composta por números finitos e discretos em suas coordenadas (amostragem) e por valores de intensidade (quantização). Uma imagem digital possui ainda número finito de elementos denominados *Picture Elements* ou, em sua forma abreviada apenas pixels. Em outras palavras, trata-se da discretização de uma imagem em formato analógico em uma matriz bidimensional.

Para melhor entendimento, a Figura 1 (a) mostra a imagem de maneira contínua enquanto que a Figura 1 (b) ilustra a amostragem e quantização da imagem. Cada quadrado na imagem é um pixel.

O Processamento digital da imagem (PDI) consiste em um conjunto de técnicas para capturar, representar e transformar imagens digitais com o auxílio de computador (PEDRINI; SCHWARTZ, 2008). A utilização de técnicas de processamento digital de imagens vem crescendo desde a década de 60, período em que computadores começaram a ter processamento suficiente para realizar tarefas de processamento de imagem. Nessa época, as primeiras imagens da Lua transmitidas por uma sonda foram processadas para corrigir distorções na imagem (GONZALEZ; WOODS, 2010). Nos anos seguintes, técnicas de processamento de imagem começaram a ser aplicadas também na medicina e em programas espaciais. A computação é utilizada para realçar o contraste, melhorar imagens radiográficas, processar e restaurar ima-

Figura 1 – Imagem projetada em uma matriz bidimensional.



Fonte: Gonzalez e Woods (2010).

gens de objetos arqueológicos, estudar padrões de poluição através de imagens aéreas e de satélite, entre outras aplicações (GONZALEZ; WOODS, 2010). O processamento de imagens aéreas auxilia na previsão do tempo e na avaliação ambiental, problemas típicos da percepção por máquina, ou seja, que não dependem da interpretação humana (GONZALEZ; WOODS, 2010).

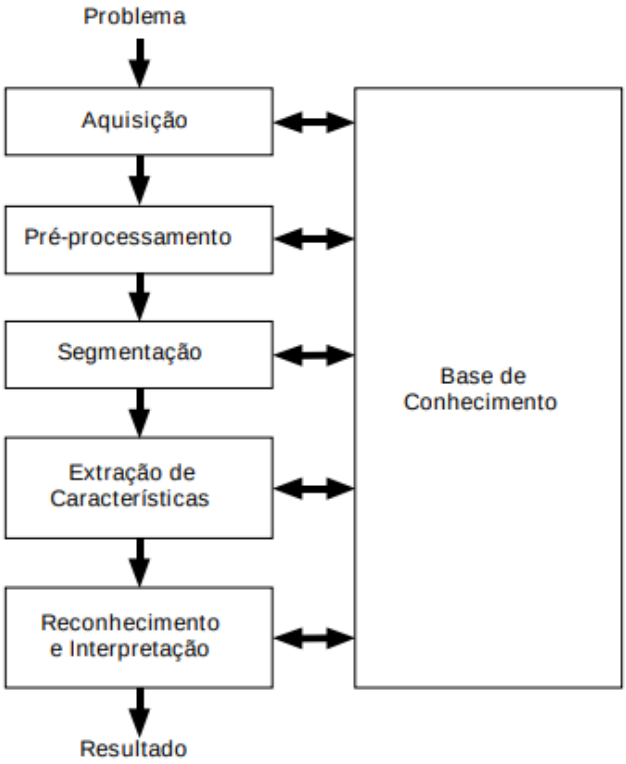
O PDI pode ser dividido em dois métodos: baixo e alto nível. Os métodos de baixo nível são utilizados em imagens digitais para redução de ruído, ajuste do contraste, extração de bordas e compressão de imagens. Os métodos de alto nível são utilizados na segmentação da imagem, auxiliando no reconhecimento ou classificação de regiões ou objetos de interesse (PEDRINI; SCHWARTZ, 2008).

2.1.1 Etapas de um Sistema de Processamento de Imagens

Um sistema de processamento de imagens passa por cinco etapas, desde o domínio do problema até o resultado (PEDRINI; SCHWARTZ, 2008).

Conforme ilustrado na Figura 2, a primeira etapa de um sistema de processamento de imagens é a aquisição. Nessa etapa, a captura das imagens é feita por meio de algum dispositivo ou sensor, como por exemplo um satélite. Leva-se em consideração o número de camadas da imagem digitalizada. Na etapa de pré-processamento, técnicas são aplicadas para melhorar, suavizar ou remover ruídos das imagens. A segmentação faz a extração de áreas de interesse, detectando regiões ou bordas. Essa segmentação é representada e descrita na etapa seguinte, onde objetos de interesse extraídos da imagem formam um vetor de características. Por fim, a etapa de reconhecimento ou classificação atribui rótulos aos objetos encontrados no vetor de

Figura 2 – Etapas do Sistema de Processamento de Imagens.



Fonte: Pedrini e Schwartz (2008).

características. A partir disso, a interpretação atribui um significado ao que for reconhecido e gera um resultado (PEDRINI; SCHWARTZ, 2008).

2.1.2 Segmentação da Imagem

Conforme descrito na seção 2.1.1, Pedrini e Schwartz (2008) colocam a segmentação como uma das etapas de um sistema de processamento de imagens. Gonzalez e Woods (2010) explicam que os procedimentos de segmentação subdividem e identificam regiões e objetos de uma imagem. A segmentação de imagens não é uma tarefa trivial pois tenta traduzir para o computador um processo cognitivo humano. Porém, quando realizada corretamente, possui maior probabilidade de reconhecer e identificar objetos.

Existem duas abordagens convencionais utilizadas para segmentação de imagens, sendo que ambas baseiam-se basicamente nos níveis de cinza da imagem: a descontinuidades e a similaridade. A primeira busca encontrar descontinuidades na imagem através da busca por mudanças abruptas nos níveis de cinza, como retas ou bordas. A segunda busca agrupar pontos de similaridade na imagem de acordo com critérios predefinidos. A intensidade de cinza, cor, semântica ou textura são propriedades que podem caracterizar uma região de interesse (PEDRINI; SCHWARTZ, 2008).

Na segmentação semântica, a cada pixel da imagem é atribuida uma classe correspondente ao objeto ou regiao a qual o pixel pertence, como exemplificado na Figura 3 (b). Em outras palavras, a segmentação semântica tem como objetivo dividir uma imagem em subconjuntos mutuamente exclusivos, nos quais cada subconjunto representa uma região significativa da imagem original (HAO; ZHOU; GUO, 2020).

Figura 3 – Aplicação da segmentação semântica.

(a) Imagem (b) Segmentação Semântica

Fonte: Adaptado de Hao, Zhou e Guo (2020).

2.2 Sensoriamento Remoto

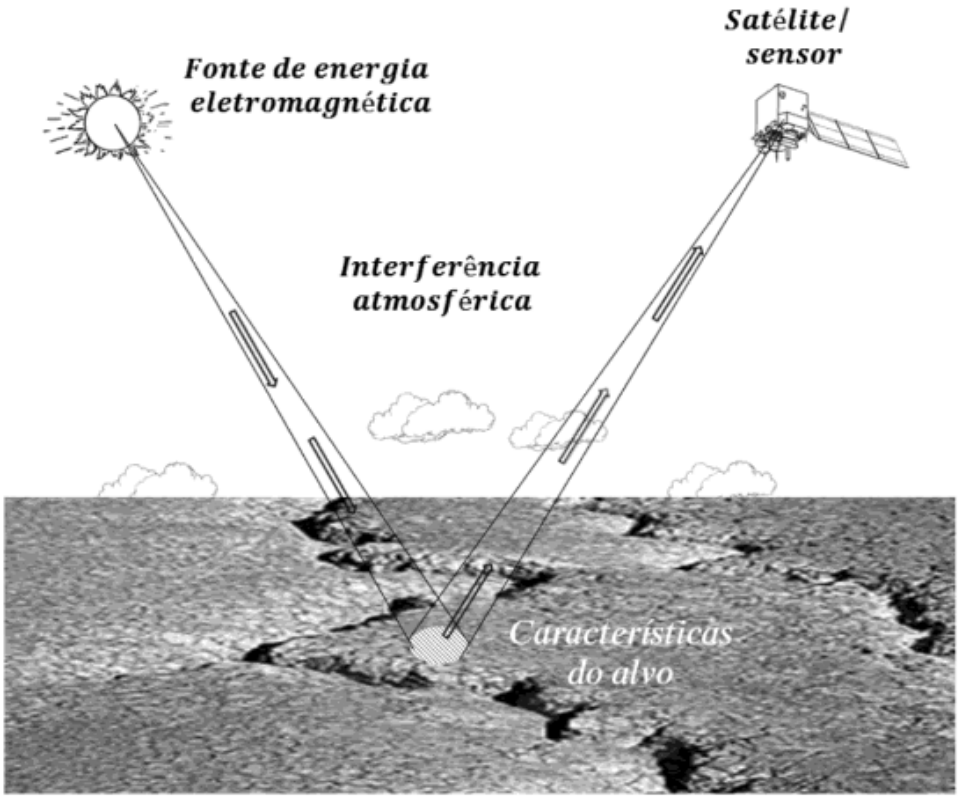
O termo Sensoriamento Remoto (SR) apareceu pela primeira vez na literatura científica em 1960, e desde então esse termo tem abrigado tecnologia e conhecimentos complexos derivados de diferentes campos que vão desde a física até a cartografia (NOVO; PONZONI, 2001).

SR é a área que combina sensores, equipamentos de processamento e transmissão de dados em aeronaves, espaçonaves ou outras plataformas para estudar eventos, fenômenos e processos que ocorrem na superfície terrestre. As técnicas em SR permitem registrar e analisar as interações entre a Radiação Eletromagnética (REM) e as substâncias presentes na terra (NOVO, 2010). Segundo Kumar *et al.* (2012), a radiação eletromagnética ao incidir sobre a superfície de um material, parte dela será refletida por esta superfície, parte será absorvida e parte pode ser transmitida caso a matéria possua alguma transparência. A soma desses três componentes será igual à intensidade da energia incidente.

De acordo com Lorenzzetti (2015), o sensoriamento remoto pode ser aplicado para monitoramento ambiental e militar. Dentre as áreas de aplicação do sensoriamento remoto ambiental está o monitoramento da superfície do solo que envolve diversas áreas, tais como, geologia, agronomia, florestas e cobertura e ocupação do solo. A Figura 4 ilustra o esquema que envolve a aquisição de informação por meio da técnica de SR destacando quatro fatores principais: 1) A energia que incide sobre o alvo; 2) As características do meio em que ocorre a propagação da energia (ex: atmosfera); 3) As propriedades do alvo, visto que a energia incidente pode sofrer interferências, como reflexão, refração ou absorção, dependendo das características do objeto

ou superfície; 4) e as próprias características do sensor que captura os dados para extrair informações relevantes, como a refletância, temperatura ou índices de vegetação. De maneira geral, a Figura 4 demonstra que a energia, neste caso proveniente do sol, é refletida pelo alvo e chega ao sensor no satélite.

Figura 4 – Esquema simplificado dos principais componentes presentes em SR.



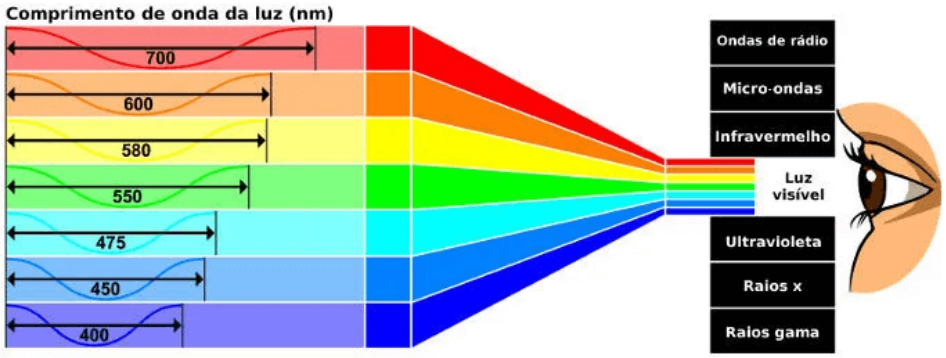
Fonte: Lorenzzetti (2015).

A Figura 5 destaca o comprimento de onda referente a luz visível. Essa é a região do espectro magnético com a mais alta intensidade de fluxo radiante e onde há a melhor janela atmosférica, bastante transparente, deixando passar uma grande quantidade de radiação, razão pela qual é muito utilizada na área de SR. O problema dessa faixa espectral é o alto espalhamento da radiação solar incidente pelos gases atmosféricos, que pode reduzir o contraste da reflectância dos alvos terrestres (MENESES *et al.*, 2012).

2.2.1 Imagens do Google Earth

O aplicativo do GE disponibiliza ao usuário um globo virtual que é constituído por imagens de satélite ou fotografias aéreas de todo o mundo. O usuário pode explorar virtualmente o planeta através de imagens em alta resolução. Desde o seu lançamento em 2005, o GE vem evoluindo e ganhando popularidade devido ao seu acesso gratuito, interface fácil de usar e ampla cobertura global, oferecendo uma visão realista e imersiva da superfície do planeta. Com o aumento da sua popularidade, um número crescente de pesquisadores recentemente começou

Figura 5 – Faixas espectrais com foco na luz visível.



Fonte: Helerbrock (2023).

a utilizar imagens do GE em diversas aplicações em projetos técnicos e científicos, incluindo estudos sobre uso e cobetura do solo (NGUYEN-KHANH; NGUYEN-NGOC-YEN; DINH-QUOC, 2021) e outras áreas de aplicação (WATANABE; SUMI; ISE, 2020; XING *et al.*, 2020).

Crowder (2007) explica que a qualidade das imagens do GE pode variar em diferentes regiões, devido à sua dependência em relação a vários fornecedores externos para obter imagens aéreas e de satélite. As áreas rurais tendem a ter menos detalhes, pois não são tão frequentemente fotografadas do espaço. Isso não é uma limitação do GE em si, mas sim um reflexo do estado atual dos dados disponíveis. Geralmente, quanto mais valioso o imóvel, maior a probabilidade de ter sido objeto de análise detalhada por satélite. Embora o GE se baseie em imagens de satélite capturadas nos últimos três anos, ele não é apenas uma coleção estática de imagens de diferentes fontes. Em vez disso, o serviço é continuamente atualizado. Essa atenção aos detalhes é uma das razões pelas quais usuários casuais e profissionais confiam no GE.

Para ilustrar a diferença de qualidade das imagens do GE entre regiões, quatro imagens foram capturadas e agrupadas, duas da cidade de São Paulo (SP) e duas da cidade de Petrópolis (RJ). A Figura 6 demonstra a diferença de qualidade visual das imagens obtidas pelo GE nas áreas urbana e rural. Percebe-se que a Figura 6 (a)(b) da região de São Paulo possui maior qualidade visual em comparação com a Figura 6 (c)(d) da cidade de Petrópolis, principalmente na área rural.

2.3 Aprendizagem de Máquina

A aprendizagem de máquina (AM) é uma subárea da inteligência artificial que permite ao sistema aprender através de experiências e ficar mais inteligente ao passar do tempo de maneira autônoma, sem ajuda do ser humano (SHARMA; SHARMA; JINDAL, 2021).

De acordo com Sharma, Sharma e Jindal (2021), essa área é dividida em três categorias: aprendizagem supervisionada, não supervisionada e por reforço. A Figura 7 ilustra essas abordagens.

Figura 6 – Qualidade visual das imagens obtidas pelo GE.

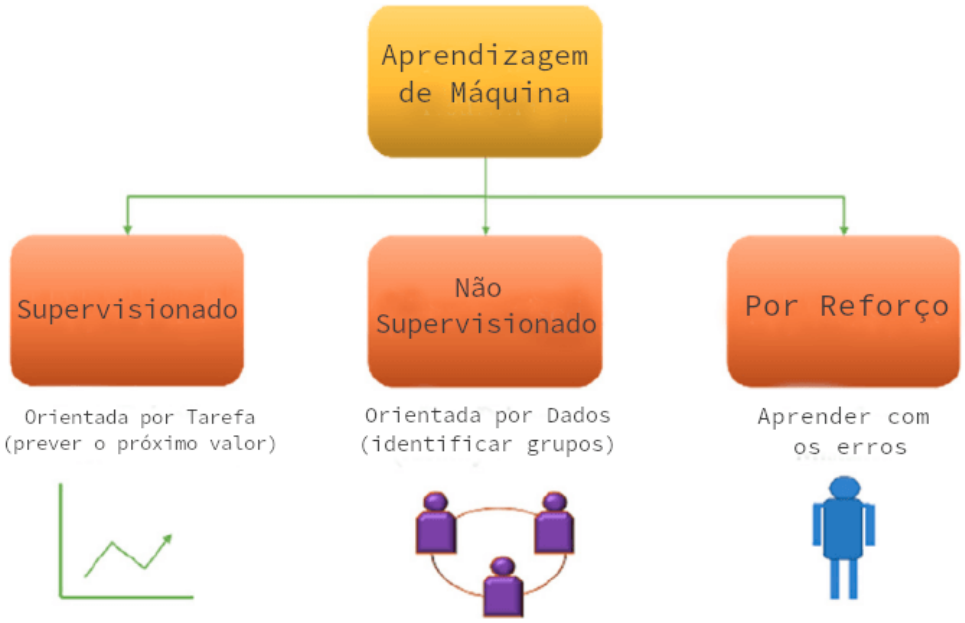
Fonte: Adaptado de Google (2023).

Cada abordagem ilustrada na Figura 7 é usada para tipos diferentes de aprendizagem. Na aprendizagem supervisionada todos os registros no conjunto de dados são rotulados e divididos em conjuntos de treinamento e teste. O algoritmo aprende a prever a saída a partir de um conjunto de dados de treinamento e avalia a precisão do modelo por meio do conjunto de testes. Essa categoria é geralmente utilizada para tarefas de classificação e regressão. Diferentemente disso, na aprendizagem não-supervisionada, os rótulos são desconhecidos e o algoritmo aprende características a partir da estrutura dos dados de entrada. Essa categoria é geralmente utilizada para redução de características e para agrupamento. Por fim, na aprendizagem por reforço, geralmente as duas estratégias anteriores são utilizadas. Uma parte dos dados são rotulados e a outra parte não. Essa estratégia é usada para estimar erros como recompensa ou penalidade e tomar uma decisão de acordo com o resultado. Quanto maior for o erro, maior será a penalidade atribuída a rede (SHARMA; SHARMA; JINDAL, 2021).

2.4 Redes Neurais Artificiais

Redes Neural Artificial (RNA), (do inglês, *Artificial Neural Network*) são modelos matemáticos que usam uma coleção de unidades computacionais simples (neurônios) interligadas em uma rede. Esses modelos são utilizados em diversas tarefas de reconhecimento de padrões, como reconhecimento de fala, reconhecimento de objetos, identificação de células cancerígenas, entre outras (HERTZ; KROGH; PALMER, 1991).

Figura 7 – Abordagens em ML.



Fonte: Sharma, Sharma e Jindal (2021).

RNA, ou simplesmente redes neurais, são máquinas baseadas no funcionamento do cérebro humano. Para isso, usam de células computacionais denominadas neurônios. São semelhantes ao cérebro humano pois o conhecimento da rede é adquirido a partir do ambiente e das conexões entre neurônios, conhecidos como pesos sinápticos (HAYKIN, 2000).

O modelo de um neurônio artificial, ilustrado na Figura 8, é composto essencialmente por três elementos fundamentais: um conjunto de sinais de entrada, um somador e uma função de ativação. Conforme destacado por Haykin (2000), a cada sinal de entrada é associado um peso sináptico. Dessa forma, cada sinal x_j na entrada da sinapse j, conectada ao neurônio k, é multiplicado pelo seu peso sináptico correspondente w_{kj}. O somatório ponderado dos sinais de entrada e seus pesos é calculado, e o resultado é submetido a uma função de ativação que controla a amplitude da saída do neurônio. Para ajustar a saída em relação à soma ponderada, é adicionado um parâmetro extra na entrada da função de ativação, conhecido como bias e representado na figura por b_k.

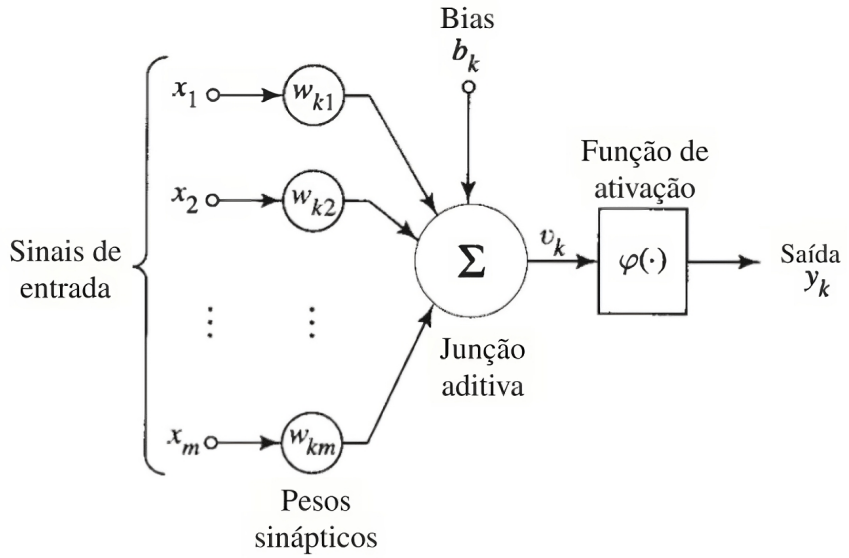
Matematicamente podemos representar um neurônio a partir das seguintes equações:

$$u_k = \sum_{j=1}^{m} w_{kj} x_j \tag{1}$$

e

$$y_k = \text{sign}(u_k + b_k) \tag{2}$$

Figura 8 – Modelo de um neurônio artificial.



Fonte: Haykin (2000).

em que x_i são os sinais de entrada; w_{ki} são os pesos sinápticos para cada sinal de entrada; u_k é a saída do somatório linear calculado a partir dos sinais de entrada e seus pesos sinápticos; b_k é o bias; $sign$ é a função de ativação; e y_k é a saída do neurônio.

As redes neurais podem ser classificadas em diferentes tipos, as quais são usadas para diferentes propósitos. A Rede Neural Perceptron foi introduzida em 1958 por Frank Rosenblatt. Esse modelo é um dos mais antigos e lida com um único neurônio, classificando o resultado de forma linear. As redes neurais *feedforward*, ou perceptron de múltiplas camadas (MLP) são compostas por uma camada de entrada, camadas ocultas e uma camada de saída. Esse é o modelo base utilizado em visão computacional já que resolve problemas não lineares, que são a maioria dos problemas presentes no mundo real (IBM, 2020).

2.4.1 Medidas de Custo

A Entropia Cruzada (EC) é uma medida comum utilizada em problemas de classificação. Essa função de custo é frequentemente utilizada quando a saída da rede neural representa uma distribuição de probabilidade sobre classes diferentes (LI *et al.*, 2019). Para problemas com múltiplas classes, é definida pela Equação 3.

$$EC = -\sum_{i=1}^{C} y_i \log(\hat{y}_i) \tag{3}$$

onde C é o número total de classes, EC é a função de entropia cruzada, $\hat{y}_i$ é a probabilidade predita pela rede para a classe i, y_i representa a probabilidade real associada à classe i.

A função de custo *Focal Loss*, uma variação da entropia cruzada, tenta resolver o problema de desbalanceamento de classes, dando maior peso às classes minoritárias (LIN *et al.*, 2017). A função é definida pela Equação 4.

$$FL(p_t) = -\sum_{i=1}^{C}(1 - p_{ti})^{\gamma} \cdot \log(p_{ti}) \tag{4}$$

onde:

- C é o número de classes.
- $p_t i$ é a probabilidade prevista da classe verdadeira.
- γ é um parâmetro de foco ajustável.
- O termo $(1 - p_t i)^{\gamma}$ reduz o peso da perda para exemplos bem classificados, focando mais nos exemplos difíceis e mal classificados.

2.4.2 Métodos de Otimização

Para o ajuste dos parâmetros durante a fase de treinamento de uma rede neural artificial, utiliza-se uma função de otimização. A Descida de Gradiente é um algoritmo de otimização. A ideia geral por trás da Descida de Gradiente é ajustar iterativamente os parâmetros θ a fim de minimizar uma função de custo, medindo o gradiente da função de custo em relação aos parâmetros. Inicialmente, os parâmetros podem ser inicializados com valores aleatórios ou por meio de inicializadores, e então os atualizamos gradualmente, um passo de cada vez. A cada passo, diminuímos a função de custo até que ela convirja para um mínimo (RUDER, 2016).

A taxa de aprendizagem é um importante hiperparâmetro da Descida de Gradiente. Se definirmos uma taxa muito baixa, o algoritmo levará muito tempo para convergir para um mínimo. Por outro lado, se a taxa for muito alta, o algoritmo de Descida de Gradiente pode não convergir. Para encontrar uma boa taxa de aprendizagem, podemos usar busca em grade ou busca aleatória (RUDER, 2016).

A descida de gradiente estocástica, (do inglês, *Stochastic Gradient Descent* - SGD) é uma variação do algoritmo Gradiente Descendente usado para otimizar modelos de aprendizagem de máquina. O *Stochastic Gradient Descent* (SGD) é um algoritmo de otimização amplamente utilizado em algoritmos de aprendizagem de máquina e treinamento de modelos, especialmente em problemas de grande escala. Dada uma função de perda $J(\theta; x, y)$, onde θ são os parâmetros do modelo, x é o conjunto de dados de entrada e y é a saída desejada, o SGD atualiza os parâmetros θ em cada etapa com base em uma amostra aleatória dos dados (RUDER, 2016). Matematicamente, a atualização dos parâmetros θ no SGD é feita da seguinte forma:

$$\theta = \theta - \eta \cdot \nabla_{\theta} J(\theta; x^{(i)}; y^{(i)}) \tag{5}$$

Onde:

- θ representa os parâmetros do modelo que estamos otimizando.
- η é a taxa de aprendizagem que controla o tamanho dos passos de atualização).
- $\nabla_\theta J(\theta; x^{(i)}; y^{(i)})$ é o gradiente da função de perda $J(\theta; x^{(i)}; y^{(i)})$ em relação aos parâmetros θ, calculado com base em uma amostra aleatória (x_i, y_i) dos dados.

Uma técnica comumente usada em algoritmos de otimização, como o SGD é o parâmetro de *momentum*. Basicamente, é um valor entre zero e um que é adicionado para acelerar a convergência e melhorar a estabilidade do treinamento de modelos de aprendizagem de máquina, especialmente em cenários nos quais a função de custo possui muitos mínimos locais, platôs ou variações abruptas.

2.4.3 Função de Ativação

Uma função de ativação decide se um neurônio deve ser ativado ou não. Isso significa que ele decidirá se a entrada do neurônio para a rede é importante ou não no processo de previsão usando operações matemáticas mais simples.

Em RNC as camadas de convolução e a camada totalmente conectada são geralmente seguidas por uma função de ativação não linear (KHAN *et al.*, 2018). Uma das funções mais indicada para esse tipo de rede é a função *Rectified Linear Units* (ReLU) (do inglês, *Rectified Linear Unit* - ReLU). Um neurônio com uma função de ativação ReLU assume quaisquer valores reais como entrada, mas só é ativado quando essas entradas são maiores que 0. Valores negativos são convertidos em 0 (GOODFELLOW; BENGIO; COURVILLE, 2016). Matematicamente, a função ReLU é dada pela Equação 6.

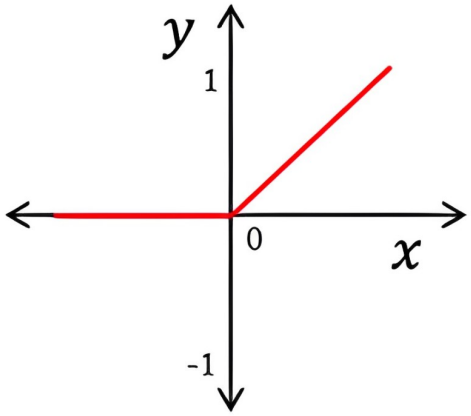
$$f(x) = max(0, x) \tag{6}$$

Esta função é aplicada para cada pixel do mapa de características de forma a retificar os valores negativos.

2.5 Aprendizagem Profunda

A Aprendizagem Profunda (AP) é uma subárea da aprendizagem de máquina que se baseia em modelos de redes neurais artificiais para realizar tarefas complexas de aprendizagem e representação de dados. Essa abordagem usa arquiteturas de redes neurais profundas, que consistem em várias camadas de neurônios interconectados. (SHARMA; SHARMA; JINDAL, 2021).

Figura 9 – Função de Ativação ReLU



Fonte: Adaptado de Khan *et al.* (2018).

De acordo com LECUN, BENGIO e HINTON (2015), a AP permite que modelos computacionais compostos de várias camadas de processamento aprendam representações de dados com vários níveis de abstração. Essa técnica está sendo utilizada cada vez mais em diversas áreas, como por exemplo no desafio de escrita manual MNIST feito por LECUN, CORTES e BURGES (2021), para avaliar o impacto de diferentes estratégias de desenvolvimento urbano na cobertura do solo e nas mudanças na qualidade ecológica Shi, Yang e Li (2023) e para identificação e mapeamento de árvores através de dados de sensoriamento remoto com aplicação no manejo florestal Onishi e Ise (2021).

RNC foram as primeiras técnicas de AP nas quais várias camadas foram treinadas de maneira eficaz. Sua arquitetura consegue reduzir o número de parâmetros que precisam ser aprendidos, otimizando o treinamento da rede (AREL; ROSE; KARNOWSKI, 2010). Segundo Hu *et al.* (2018), o sucesso da AP envolve dois fatores: a melhoria na capacidade de processamento por parte da Unidade de Processamento Gráfico (UPG) e a grande quantidade de dados disponíveis para treinamento. RNC será contextualizada na seção 2.6.

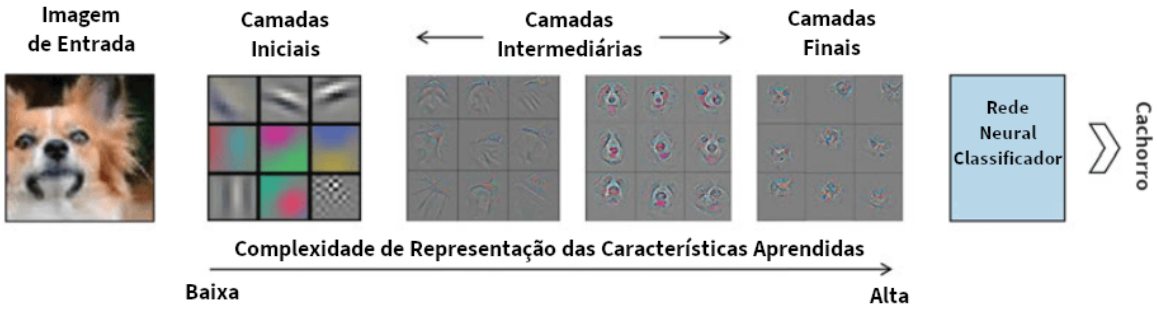
2.6 Rede Neural Convolucional

Segundo Khan *et al.* (2018), RNC (em inglês, *Convolutional Neural Network*) é uma rede neural que pertence à família das redes neurais multicamadas, especialmente projetadas para trabalhar com imagem e vídeo. RNC utilizam um filtro pelo menos bidimensional que é convolvido com a entrada da camada, essencial para o reconhecimento de padrões neste tipo de mídia.

A Figura 10 representa o processo de extração de características de uma RNC, que recebe uma imagem como entrada e tem como objetivo produzir um rótulo como saída. Nessa figura, observamos que a aprendizagem da rede começa com a extração de características de nível mais elementar nas camadas iniciais. Em seguida, avança para a extração de caracte-

rísticas mais complexas e culmina na etapa de classificação, executada por uma rede neural especializada em classificação (conforme destacado por Khan *et al.* (2018)).

Figura 10 – Representação de aprendizagem da RNC



Fonte: Adaptado de Khan *et al.* (2018).

2.6.1 Camada Convolucional

Em uma rede neural convolucional, encontramos três tipos de camadas: a camada convolucional, a camada de subamostragem (ou *pooling*) e a camada totalmente conectada (*fully connected*). A camada convolucional opera convolvendo a imagem com filtros específicos, com o objetivo de identificar e extrair as características mais relevantes e significativas da imagem (GOODFELLOW; BENGIO; COURVILLE, 2016). Matematicamente, o operador da convolução pode ser visualizado na Equação 7.

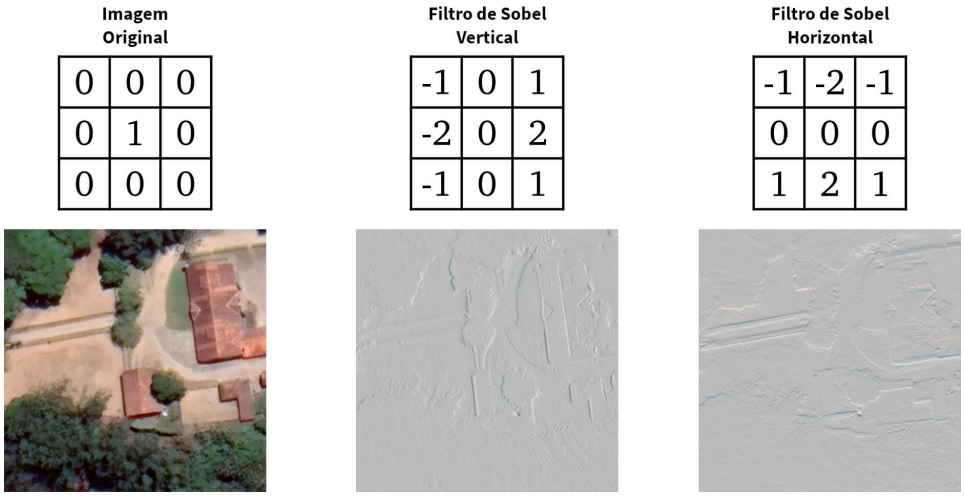
$$S(i,j) = (K * I)(i,j) = \sum_{m} \sum_{n} I(i-m,j-n)K(m,n) \tag{7}$$

em que, a imagem resultante (representada por S) é obtida a partir da convolução entre a imagem de entrada (I) com um filtro específico (K). Durante o processo de convolução, utilizam-se as variáveis i e j para iterar ao longo do eixo x e y da imagem de entrada, respectivamente, enquanto as variáveis m e n são utilizadas para iterar ao longo do eixo x e y do filtro, respectivamente. Essas iterações permitem que o filtro percorra toda a extensão da imagem de entrada, resultando na formação da imagem convolvida S.

A camada convolucional é o componente mais importante de uma RNC. Compreende um conjunto de filtros (*kernels*) que são convolvidos com uma determinada entrada para gerar um mapa de características como saída. Cada filtro possui pesos que são aprendidos durante o processo de treinamento. Os pesos, normalmente, são inicializados aleatoriamente e a profundidade da saída de uma convolução é igual a quantidade de filtros aplicados (KHAN *et al.*, 2018). Nessa operação, é possível utilizar-se de um passo (*stride*) para reduzir o tempo de computação. Um passo com tamanho 1, significa que a convolução será feita em todos os pares possíveis (HAFEMANN, 2014).

Cada camada do modelo funciona como uma espécie de detector de características, convertendo os dados em representações mais abstratas e potencialmente úteis. As camadas iniciais têm a capacidade de aprender características de baixo nível, como a detecção de bordas, enquanto as camadas subsequentes aprendem representações de nível mais avançado, como identificar formas locais mais complexas, até alcançar representações de alto nível, como o reconhecimento de uma região de interesse ou objeto (FREUDENBERG, 2019). Filtros podem ser aplicados, por exemplo, para detectar bordas com o filtro de Sobel e Feldman (1968). A Figura 11 ilustra a aplicação desse filtro destacando o resultado da convolução em uma imagem *Red-Green-Blue* (RGB). Percebe-se que a aplicação do filtro de Sobel extrai atributos significativos da imagem original (esquerda), neste caso ressaltando as bordas existentes na direção vertical (centro) e horizontal (direita).

Figura 11 – Aplicação do filtro de Sobel vertical e horizontal para destacar bordas



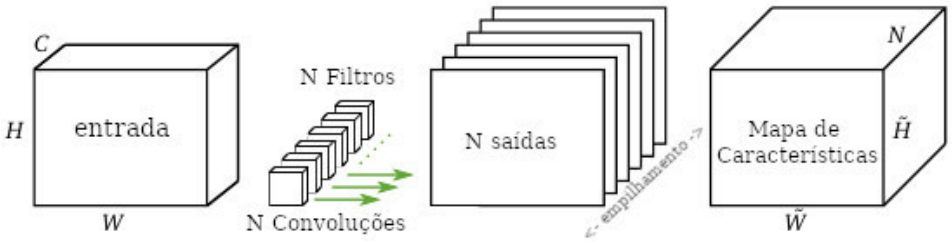
Fonte: Autoria Própria (2023).

De acordo com Freudenberg (2019), existem várias convoluções atuando em paralelo na mesma entrada, cada uma com um filtro diferente. O tipo de filtro utilizado define a característica que será capturada na operação de convolução. Como resultado intermediário, existem N imagens diferentes, mais conhecido por mapa de características. Essa operação em paralelo está ilustrada na Figura 12. A figura demonstra que N convoluções atuam na entrada com diferentes tipos de filtros, resultando em N saídas bidimensionais que são empilhadas para gerar o mapa de características final, que é a saída da camada.

A operação de convolução é ilustrada na Figura 13 que mostra uma simulação com os cálculos realizados em cada etapa, à medida que um filtro desliza sobre a imagem de entrada em passos com tamanhos previamente definidos (*strides*), captando os traços mais marcantes.

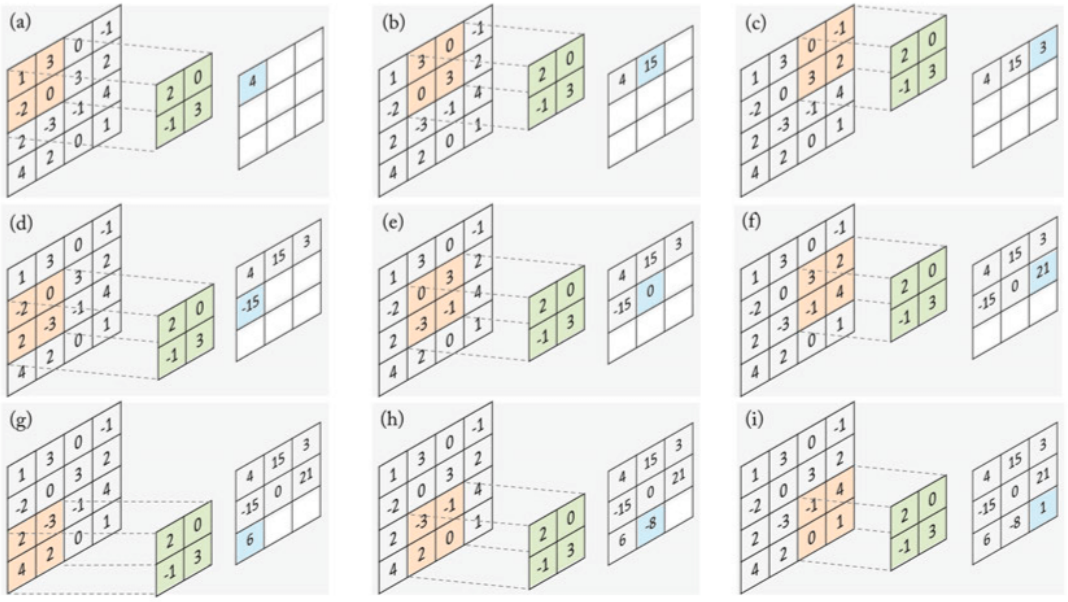
No exemplo ilustrado na Figura 13, um filtro 2x2 (em verde) é multiplicado pela região do mesmo tamanho (em laranja) dentro de um mapa de características de 4x4 entradas e os valores resultantes são somados para obter uma entrada correspondente no mapa de características de saída (em azul).

Figura 12 – Operação de Convolução



Fonte: Adaptado de Freudenberg (2019).

Figura 13 – Operação de Convolução - Simulação



Fonte: Khan *et al.* (2018).

2.6.2 Camada de Subamostragem

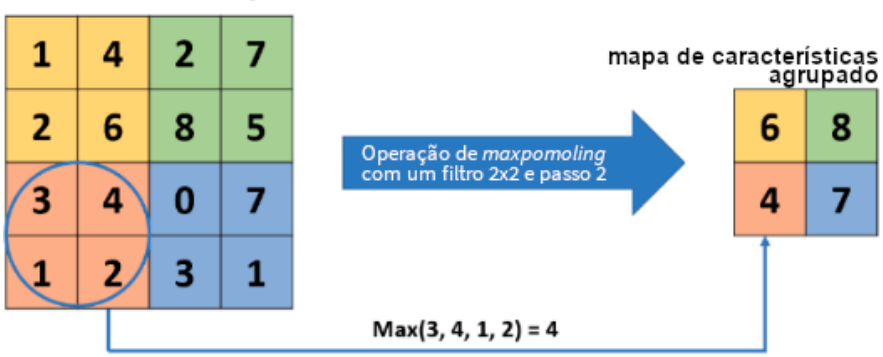
RNC utiliza várias camadas de subamostragem (em inglês, *pooling*). As camadas de *pooling* são utilizadas para reduzir a taxa de amostragem espacial das matrizes resultantes da convolução *down-sampling*. Por consequência, a operação de *pooling* reduz o custo computacional da rede e ajuda a controlar o *overfitting*[1]. Assim como na convolução, o *pooling* é responsável por resumir a informação daquela área em um único valor. O método mais utilizado é o *max-pooling* (HAFEMANN, 2014).

A Figura 14 exemplifica o resultado da aplicação do *max-pooling* em uma imagem (4x4) utilizando um *kernel* de tamanho (2x2) e *stride* 2.

Conforme ilustrado na Figura 14, a aplicação do *max-pooling* mantém o maior valor do pixel reduzindo o tamanho da imagem, mas mantendo informações importantes. Outras funções de *pooling* populares são a média (*average pooling*) e o *pooling* misto (*mixed pooling*) (GHOLAMALINEZHAD; KHOSRAVI, 2020)

[1] É quando o modelo aprende demais sobre os dados, como se tivesse decorado os dados de treino e não fosse capaz de generalizar.

Figura 14 – Exemplo da operação de *max-pooling*



Fonte: Adaptado de Gholamalinezhad e Khosravi (2020).

2.6.3 Camada Totalmente Conectada

A última etapa de uma RNC é a camada totalmente conectada (em inglês, *Fully Connected Layer*). O significado do termo totalmente conectada vem do fato que cada neurônio da camada anterior está conectado a cada neurônio da próxima camada. O número de neurônios na saída dessa camada corresponde ao número de classes desejáveis no experimento. Em tarefas de classificação a saída da rede geralmente utiliza a função de ativação *softmax*. Essa função é útil para gerar uma saída de probabilidades da imagem de entrada entre 0 e 1 para cada classe esperada. A soma da saída de probabilidades para a camada totalmente conectada é 1. (KARN, 2016). Matematicamente, a função *softmax* é representada pela Equação 8.

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{8}$$

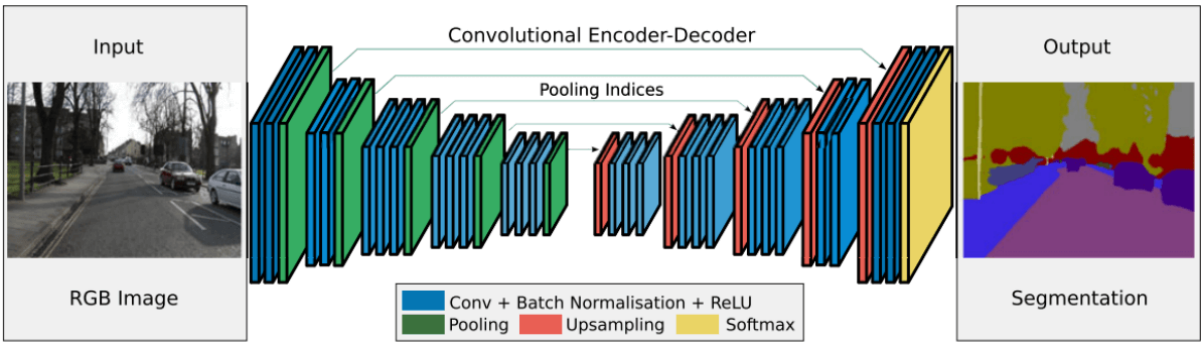
em que z é o vetor de entrada da função. Todos os valores de z_i são elementos do vetor de entrada para a função *softmax*, e podem assumir qualquer valor real, positivo, zero ou negativo. A função exponencial padrão é aplicada a cada elemento do vetor de entrada em e^{z_i}. Isso resultará em um valor positivo acima de 0. O termo K corresponde ao número de classes do classificador. O termo de baixo da fórmula é um termo de normalização e garante que a soma da saída será 1 e cada saída terá um valor de probabilidade entre 0 e 1.

2.6.4 Rede Completamente Convolucional - SegNet

Criada por Badrinarayanan, Kendall e Cipolla (2017) é uma arquitetura de rede neural profunda e totalmente convolucional para segmentação semântica denominada SegNet. O mecanismo central de segmentação treinável consiste em uma rede de codificadores, uma rede de decodificadores correspondente seguida por uma camada de classificação em pixels. A arquitetura da rede de codificadores é topologicamente idêntica às 13 camadas convolucionais da rede VGG16 (SIMONYAN; ZISSERMAN, 2015). O codificador fornece um mapa de ativação de baixa resolução que representa as características mais importantes para cada entrada. A

imagem segmentada é reconstruída pelo decodificador. A rede de decodificação é composta de camadas convolucionais e de subamostragem que utilizam os índices de agrupamento máximo *(max-pooling)* correspondentes do codificador para subamostragem do mapa de características de baixa resolução. Na última camada, um classificador *softmax* recebe o mapa de características do decodificador para classificação por pixel. A Figura 15 mostra uma ilustração da arquitetura padrão da rede SegNet.

Figura 15 – Arquitetura Padrão da Rede SegNet



Fonte: Badrinarayanan, Kendall e Cipolla (2017).

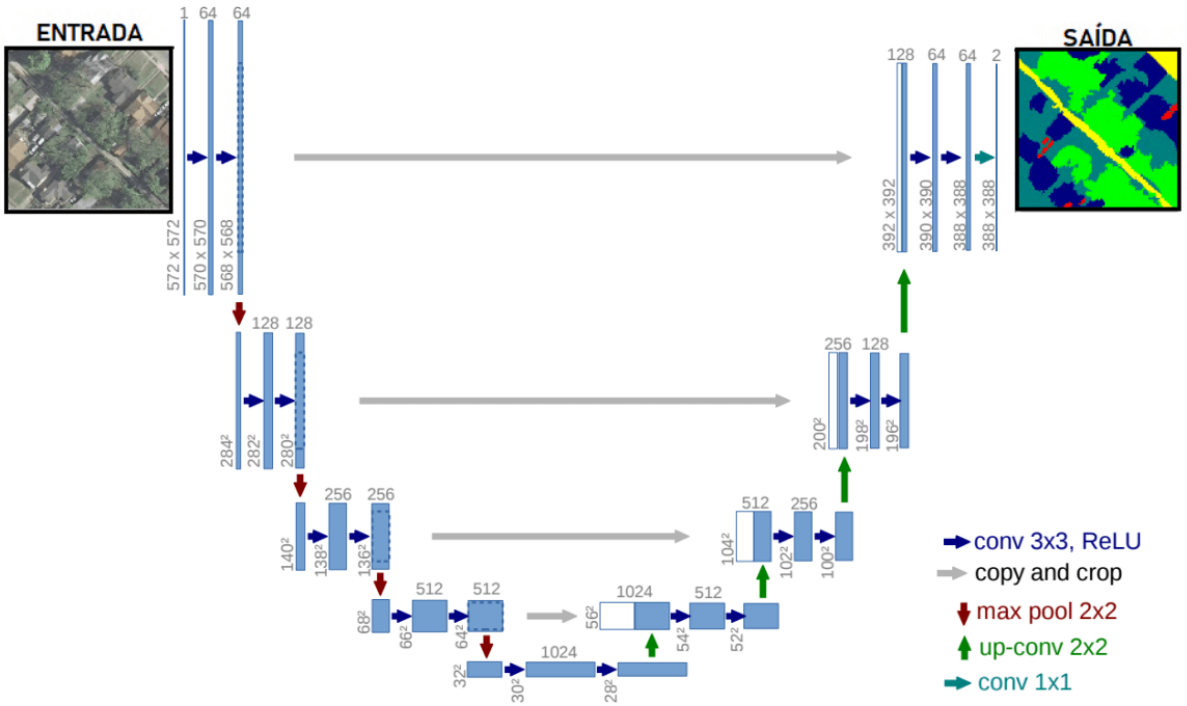
Uma das principais novidades da rede SegNet em relação a outras redes neurais convolucionais é a sua abordagem baseada em *encoder-decoder*. Outra novidade da rede SegNet é a sua capacidade de lidar com imagens de alta resolução, utilizando menos parâmetros do que outras redes convolucionais. Isso é possível graças ao uso de camadas de *pooling* invertido, que preservam a resolução espacial da imagem durante a decodificação.

2.6.5 Rede Completamente Convolucional - U-NET

Introduzida por Ronneberger, Fischer e Brox (2015), a U-Net é uma rede neural convolucional projetada para segmentação de imagens, especialmente em tarefas de segmentação de células e órgãos em imagens médicas. A arquitetura da U-Net, representada na Figura 16, tem uma forma de U, com uma etapa de codificação e uma etapa de decodificação. A etapa de codificação é composta por várias camadas convolucionais, cada uma seguida por uma camada de *pooling*. A etapa de decodificação é composta por camadas de convolução transposta, que permitem que a rede reconstrua a imagem segmentada de alta resolução a partir da imagem de baixa resolução gerada pela etapa de codificação.

De acordo com Ronneberger, Fischer e Brox (2015), a seção de codificação (lado esquerdo da rede) envolve a repetição de duas convoluções utilizando um kernel 3x3, seguidas por uma camada de ativação *ReLU* e uma operação de *max pooling* com um kernel 2x2 e passo *(stride)* 2 para diminuir o tamanho do mapa de características. Em cada etapa de *downsampling*, o número de mapas de características é duplicado. Na seção de expansão (lado direito da rede), cada passo consiste em um *upsampling* do mapa de características, seguido por uma *up-convolution* com um kernel 2x2 para reduzir pela metade a quantidade de mapas de caracte-

Figura 16 – Representação esquemática da arquitetura da rede U-NET original



Fonte: Adaptado de Ronneberger, Fischer e Brox (2015).

rísticas. Um mapa de características cortado da seção de contração (indicado pelas setas cinzas na Figura 16) é concatenado com o mapa de características atual. Duas convoluções com um kernel 3x3 seguidas por uma camada de ativação ReLU são então realizadas. Conforme explicado pelos autores, é necessário cortar o mapa de características da seção de contração devido à perda de pixels de borda durante a convolução. Na camada final, uma convolução com um kernel 1x1 e uma função de ativação Softmax é usada para realizar a classificação pixel-a-pixel da imagem de saída.

A principal estratégia que diferencia a U-Net das outras arquiteturas completamente convolucionais é a combinação entre os mapas de características do estágio de codificação (lado esquerdo) e seus correspondentes simétricos no estágio de decodificação (lado direito), permitindo a propagação de informações de contexto para os mapas de características de alta resolução. Como a rede consegue trabalhar com poucas imagens de treinamento, os autores projetaram esta arquitetura para utilizar aumento de dados (ver subseção 2.6.6).

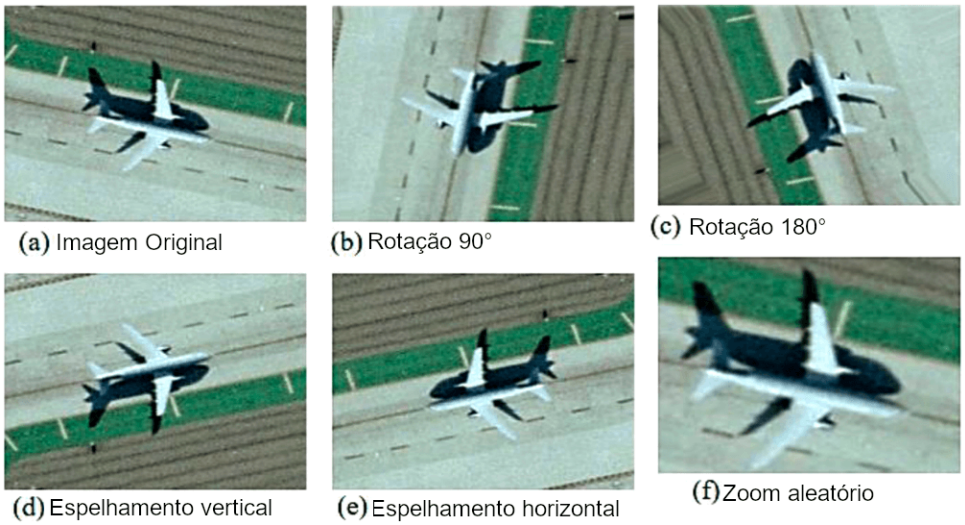
A arquitetura tem se tornado base para o desenvolvimento de adaptações em outros domínio de aplicação. No trabalho de Chen *et al.* (2021) concluíram que as características de convolução e *skip connection* da U-Net podem ter um melhor resultado no reconhecimento de objetos com bordas embaçadas, portanto o efeito de reconhecimento nas imagens do GE tende a ser mais eficiente.

2.6.6 Aumento de Dados

Redes neurais profundas precisam de muitos dados de treinamento rotulados para aumentar as chances de obter boa precisão na classificação e garantir que o modelo vai convergir de maneira estável. A estratégia de aumento dos dados (em inglês, *data augmentation*) consiste em aumentar o conjunto de dados de treinamento gerando novas imagens artificiais a partir de transformações geométricas como rotação, translação, adição de ruído, alteração na cor, brilho e contraste da imagem, recortes, entre outros, criando assim novas representações a partir das imagens originais existentes no *dataset* (LEE *et al.*, 2017).

Para Shawky *et al.* (2020), o aumento de dados em imagens permite melhorar a precisão do modelo, principalmente nas classes que estão desbalanceadas. Essa estratégia pode ser usada para gerar várias versões de uma imagem aplicando diferentes transformações no subconjunto de treinamento. A Figura 17 ilustra a aplicação de aumento de dados em uma imagem aérea.

Figura 17 – Aplicação de aumento de dados em uma imagem aérea



Fonte: Adaptado de Shawky *et al.* (2020).

Conforme pode ser visto na Figura 17, em (a) temos a imagem original, em (b)(c) temos a mesma imagem rotacionada com ângulos diferentes, em (d)(e) temos a imagem virada verticalmente e horizontalmente e em (f) a imagem foi ampliada aleatoriamente.

2.6.7 Transferência de Aprendizagem

Os seres humanos possuem intrinsecamente a habilidade de transferir conhecimento. Por exemplo, as habilidades adquiridas ao andar de bicicleta podem ser aproveitadas para aprender a conduzir uma motocicleta. A estratégia de transferência de aprendizagem, também conhecida como *transfer learning*, é uma técnica aplicada no treinamento de redes neurais profundas que permite capitalizar o conhecimento adquirido em um problema específico de

aprendizagem profunda. Essa técnica é implementada através do uso de uma rede pré-treinada extensivamente em um conjunto de dados amplo, aplicando esse conhecimento a um problema semelhante de aprendizagem profunda (SARKAR, 2018).

De acordo com Hafemann (2014), a transferência de aprendizagem é uma estratégia que utiliza o conhecimento obtido em uma tarefa genérica e o aplica em outra tarefa similar. Sarkar (2018) destaca que a motivação principal por trás do uso dessa estratégia reside no fato de que tarefas de aprendizagem profunda supervisionada abordam problemas complexos que demandam grandes volumes de dados rotulados, os quais podem ser difíceis de obter ou preparar. Por exemplo, o conjunto de dados Deng *et al.* (2009) contém milhões de imagens rotuladas e demandou anos para ser construído.

Em pesquisas anteriores, como o estudo de Carranza-Rojas *et al.* (2017), a estratégia de transferência de aprendizagem foi aplicada com sucesso na classificação de espécies vegetais. Esteva *et al.* (2017) utilizaram essa abordagem para classificação do nível de câncer de pele, enquanto Lima *et al.* (2019) aplicaram-na na classificação de diversas imagens geocientíficas. Em outro contexto, Weinstein *et al.* (2019) empregaram a transferência de aprendizagem para a localização de copas de árvores utilizando imagens RGB, e Onishi e Ise (2021) a utilizaram na identificação e mapeamento de árvores para aplicações no manejo florestal.

No estudo conduzido por Lima e Marfurt (2020), observa-se que a estratégia de transferência de aprendizagem está sendo amplamente adotada em diversas áreas, principalmente quando há escassez de dados para o treinamento do modelo. Os autores investigaram a aplicação da transferência de aprendizagem em dados de sensoriamento remoto e constataram que: a estratégia obteve excelentes resultados ao realizar o *fine-tuning* na rede neural para imagens aéreas; a transferência de aprendizagem de imagens naturais para imagens aéreas é viável; e, em comparação com RNC treinadas do zero com pesos inicializados aleatoriamente, obteve resultados superiores.

2.7 Métricas de Avaliação

As métricas de avaliação de modelos desempenham um papel fundamental na mensuração do desempenho e da eficácia de redes neurais. Elas fornecem relatórios estatísticos que avaliam a capacidade do modelo em aprender e generalizar a partir dos dados de treinamento. Embora a acurácia, que mede a porcentagem de previsões corretas, seja comumente utilizada, ao avaliar e comparar modelos com diversos parâmetros em treinamentos de múltiplas classes, é aconselhável recorrer a outras métricas para obter uma avaliação mais confiável. A matriz de confusão destaca-se como uma ferramenta valiosa na avaliação de modelos, pois quantifica as relações entre os valores reais ou esperados e as saídas previstas. Esse instrumento proporciona uma base sólida para a comparação entre diferentes cenários, conforme discutido por Grandini, Bagli e Visani (2020).

2.8 Matriz de Confusão e Métricas Relacionadas

De acordo com Grandini, Bagli e Visani (2020), a matriz de confusão é formada por uma tabela que apresenta os valores previstos pela rede neural em relação aos valores reais ou esperados. A matriz de confusão é uma estrutura composta por quatro categorias principais:

1. Verdadeiros Positivos (VP): Representam a quantidade de valores previstos corretamente pela rede neural como pertencentes à classe positiva.

2. Falsos Positivos (FP): Representam a quantidade de valores previstos incorretamente como pertencentes à classe positiva.

3. Verdadeiros Negativos (VN): Representam a quantidade de valores previstos corretamente como pertencentes à classe negativa.

4. Falsos Negativos (FN): Representam a quantidade de valores previstos incorretamente como pertencentes à classe negativa.

O Quadro 1 ilustra a matriz de confusão para um problema que envolve duas classes. A partir desses valores, é possível calcular métricas como precisão, sensibilidade, f1-score que são comumente utilizadas para avaliar a classificação em modelos de redes neurais.

Quadro 1 – Exemplificando uma matriz de confusão binária.

		Verdadeiro	
		Condição Positiva	Condição Negativa
Predito	Condição Positiva	VP	FP
	Condição Negativa	FN	VN

Fonte: Autoria Própria (2023).

Essas métricas são relevantes quando há a necessidade de realizar uma avaliação quantitativa de um ou mais modelos de classificação. No entanto, é essencial observar que todas essas métricas estão definidas especificamente para o tipo de classificação binária, ou seja, quando existem apenas duas classes possíveis: positivo ou negativo. Para um problema com mais classes, é possível adaptar a matriz de confusão, conforme ilustrado no Quadro 2. Em suma, os elementos diagonais são classificações corretas, e todos os outros, incorretas (GRANDINI; BAGLI; VISANI, 2020).

A partir da matriz de confusão é possível obter diversas outras métricas que serão apresentadas a seguir, tais como a acurácia, a precisão, a sensibilidade e a media-f.

2.8.1 Acurácia

Essa métrica define a porcentagem de amostras que foram corretamente classificadas, sejam elas positivas ou negativas. Após treinar o modelo de rede neural, o conjunto de dados

Quadro 2 – Exemplificando uma matriz de confusão para múltiplas classes.

		Verdadeiro			
Predito	$C_{(11)}$	$C_{(12)}$	$\cdots$	$C_{(1n)}$
	$C_{(21)}$	$C_{(ij)}$		
	$\vdots$		$\ddots$	
	$C_{(n1)}$			$C_{(nm)}$

Fonte: Autoria Própria (2023).

de teste é usado para fazer previsões e compará-las com as classes verdadeiras. Em seguida, a acurácia é calculada dividindo o número de previsões corretas pelo número total de exemplos de teste. O resultado pode ser multiplicado por 100 para obter uma porcentagem. Matematicamente, a acurácia é dada pela Equação 9. A equação considera no denominador a soma das amostras que foram avaliadas corretamente, verdadeiro positivo (VP), com as amostras avaliadas incorretamente, e verdadeiro negativo (VN). Estes valores estão presentes na diagonal principal da matriz de confusão. No denominador, a soma de todas as entradas da matriz de confusão (GRANDINI; BAGLI; VISANI, 2020).

$$Acuracia = \frac{VP + VN}{VP + VN + FP + FN} \tag{9}$$

Grandini, Bagli e Visani (2020) destaca que, em conjunto de dados está desbalanceado, a acurácia pode não ser a melhor métrica pois tende a esconder erros na classificação para classes menos relevantes, ou seja, com menor quantidade de amostras.

2.8.2 Precisão e Sensibilidade

As métricas de precisão (*precision*) e sensibilidade (*recall*), também conhecidas como taxa de positivos verdadeiros, são duas medidas essenciais para avaliar o desempenho de modelos de classificação.

Powers (2020) destaca que precisão avalia a capacidade do modelo de identificar corretamente as amostras positivas de cada classe em relação ao número total de amostras que o modelo classificou como positivos, ou seja, a precisão mede a proporção de amostras classificadas como positivos que são realmente positivas. Matematicamente, a sensibilidade é dada pela Equação 10.

$$Precisao = \frac{VP}{VP + FP} \tag{10}$$

Ainda segundo Powers (2020), a sensibilidade é uma métrica que mede a capacidade do modelo de identificar corretamente os exemplos positivos em relação ao número total as amostras positivas no conjunto de teste. Enquanto a acurácia fornece uma visão geral do de-

sempenho geral do modelo em todas as classes, a sensibilidade se concentra especificamente em quão bem o modelo identifica corretamente as amostras da classe positiva. Ela é especialmente relevante quando há uma classe de interesse que você deseja que o modelo identifique com alta precisão. Matematicamente, a sensibilidade é dada pela Equação 11.

$$Sensibilidade = \frac{VP}{VP + FN} \tag{11}$$

2.8.3 Medida F

Mais conhecida como *F-score*, essa medida faz uma média ponderada entre precisão e sensibilidade, em que quanto mais próximo de 1 melhor é o resultado. Ela pode ser usada tanto em situações de classificação binária quanto para multi-classe. Em problemas de classificação multi-classe, todas as entradas da matriz de confusão serão consideradas. o F1-score é comumente calculado para cada classe individualmente e depois é feita uma média ponderada (ou não) dos resultados, dependendo do tipo de média utilizada (*micro ou macro*). Sob a perspectiva Macro, os valores de performance são computados para cada classe separadamente e depois obtém-se sua média, enquanto que para a forma Micro esses valores são computados diretamente da matriz de confusão (GRANDINI; BAGLI; VISANI, 2020). Matematicamente, a sensibilidade é dada pela Equação 12.

$$fscore = 2 * \frac{precisao * sensibilidade}{precisao + sensibilidade} \tag{12}$$

2.8.4 Interseção Sobre União

A interseção sobre a união (em inglês, *Intersection Over Union* (IoU)), também conhecida como *Jaccard index* é uma métrica de avaliação usada em tarefas de visão computacional, especialmente na detecção de objetos e segmentação de imagens. É uma estatística que pode ser usada para determinar a similaridade e a diversidade de um conjunto de amostras. Ele é definido como o tamanho da interseção dividido pela união dos conjuntos de amostras, ou sejam, mede a sobreposição entre duas amostras e fornece um valor entre 0 e 1, em que 0 indica nenhuma interseção e 1 indica sobreposição completa (YANG *et al.*, 2022).

Além da detecção de objetos, a IoU também é usada em tarefas de segmentação de imagens. É comumente usado para avaliar a precisão da anotação na segmentação de imagens, detecção de objetos e localização de objetos (ISLAM *et al.*, 2022). A interseção média sobre a união (mIoU) é um índice de avaliação quantitativa usado também com frequência em estudos de segmentação de imagens (WANG *et al.*, 2022). Ele fornece uma medida da precisão geral dos resultados da segmentação, considerando a interseção e a união das amostras previstas e

da amostras verdadeiras anotadas (WANG *et al.*, 2022). Matematicamente, a IoU é dada pela Equação 13.

$$IoU = \frac{VP}{VP + FP + FN} \tag{13}$$

3 TRABALHOS RELACIONADOS

Devido à relevância do problema, recentemente alguns pesquisadores têm se dedicado ao desenvolvimento de novas abordagens utilizando modelos de RNC e segmentação semântica em imagens aéreas, alguns deles utilizaram imagens do GE em seus estudos. Neste contexto, o presente capítulo apresenta alguns destes recentes trabalhos publicados. A seleção considerou a relevância temática, a atualidade, a qualidade científica, a diversidade de perspectivas, a conexão direta com os objetivos específicos da pesquisa e a pertinência em relação ao contexto atual da área de estudo. A inclusão de trabalhos recentes de fontes confiáveis e a consideração de diversas abordagens e metodologias foram aspectos importantes na escolha dos trabalhos, visando contribuir para o embasamento teórico e metodológico da pesquisa.

O primeiro trabalho, realizado por OCHOA e GUO (2019), aborda a aplicação de RNC para a detecção e classificação de espécies de árvores em imagens aéreas de alta resolução coletadas por drones. Foram utilizados conjuntos de dados de duas localidades diferentes, Tonga e Potsdam. A arquitetura da RNC baseou-se no modelo YOLO e foi treinada para localizar e classificar árvores. A precisão de localização do modelo para a cidade de Tonga foi de 80%, enquanto a precisão de classificação foi de 98%. O segundo modelo treinado para a cidade de Potsdam teve maior dificuldade em detectar ruas devido à diferença no cenário de treinamento e teste. Essa abordagem foi eficaz para a detecção de árvores e tem potencial em situações como inventário após desastres naturais.

O segundo trabalho, de Nguyen-Khanh, Nguyen-Ngoc-Yen e Dinh-Quoc (2021), focou na construção de mapas digitais a partir de imagens de satélite, utilizando a arquitetura U-Net com EfficientNet-B0 como codificador e decodificador. A eficiência do modelo foi avaliada usando imagens de satélite do Google, mostrando que a combinação do EfficientNet-B0 com a U-Net resultou em bons resultados de segmentação semântica. No entanto, o estudo se baseou em um único conjunto de dados, e a generalização para outras situações não foi discutida de forma clara.

No terceiro trabalho, Chen *et al.* (2021) empregaram redes U-Net para segmentar estufas agrícolas de plástico em imagens de sensoriamento remoto de alta resolução. A abordagem foi dividida em três etapas: coleta e anotação de imagens, treinamento da rede U-Net e pós-processamento para remover elementos confundíveis com as estufas. A metodologia apresentada mostrou alta precisão na segmentação das estufas, demonstrando a eficácia da U-Net em imagens de sensoriamento remoto.

No quarto trabalho, Anderson-bell, Schillaci e Lipani (2021) propuseram uma rede neural híbrida multimodal para prever o risco de incêndio em construções não residenciais, usando imagens aéreas de alta resolução. A abordagem combinou uma RNC com um Perceptron Multicamadas (MLP). As imagens aéreas foram coletadas ao longo de vários anos e usadas para treinar o modelo, que demonstrou promissora capacidade de prever riscos de incêndio.

No quinto trabalho, Alam *et al.* (2021) abordam a aplicação de técnicas de aprendizagem profunda, mais especificamente Redes RNC, na tarefa de segmentação semântica de imagens obtidas por sensoriamento remoto. Duas arquiteturas de RNC, SegNet e U-net, são aprimoradas por meio da introdução de *index pooling* para melhorar essas arquiteturas, permitindo a preservação de informações espaciais cruciais durante a ampliação da resolução. Além disso, as melhorias na normalização dos dados são realizadas por meio da técnica de *Batch Normalization*. Essas arquiteturas melhoradas são então empregadas para segmentar imagens com múltiplas classes no contextos de sensoriamento remoto. Os experimentos são conduzidos usando dados de alta resolução provenientes de imagens de sensoriamento remoto na China. Para lidar com o tamanho das imagens aéreas, as imagens são cortadas aleatoriamente em pequenas partes e submetidas a várias operações de aprimoramento de dados, como rotação, espelhamento, ajuste de luz e adição de ruído. O estudo oferece contribuições significativas ao propor e avaliar modificações nas arquiteturas SegNet e U-net para segmentação semântica de imagens de sensoriamento remoto.

Finalmente, Lumnitz *et al.* (2021) apresentaram uma abordagem para a detecção automática e geolocalização de árvores urbanas usando a técnica de Region-CNN (R-CNN). O estudo destacou a eficácia da R-CNN para identificar árvores em imagens de ruas da cidade de Vancouver, mesmo com um número limitado de imagens de treinamento.

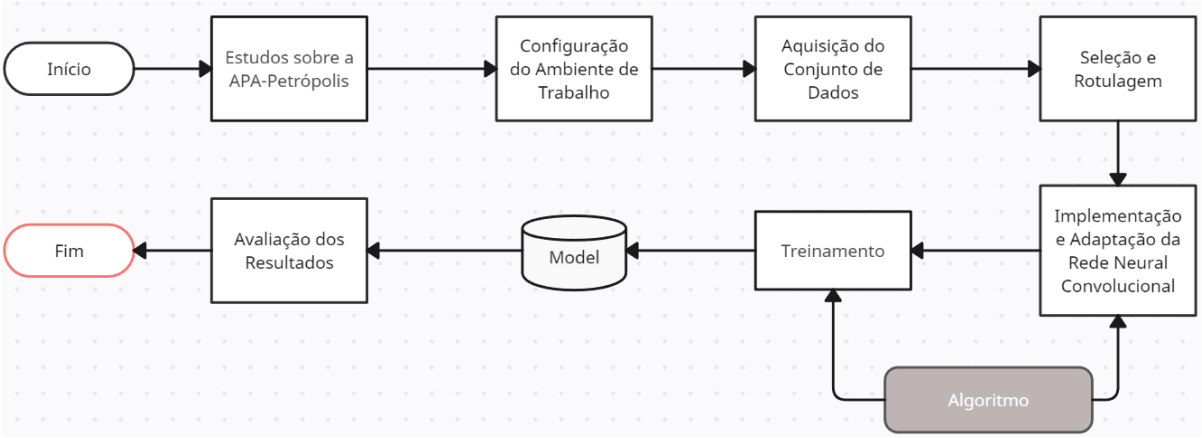
4 MATERIAIS E MÉTODOS

Este capítulo apresenta a metodologia adotada neste trabalho. Dessa maneira, a seção 4.1 aborda as etapas da metodologia proposta para a realização do trabalho. Na seção subseção 4.1.1 são abordados os detalhes sobre a área de estudo. Na seção subseção 4.1.2, serão apresentadas as principais ferramentas utilizadas para a realização dos experimentos. Na subseção 4.1.3, serão detalhados os procedimentos para obtenção do conjunto de dados. Na seção subseção 4.1.4, é descrita a etapa de pré-processamento, seleção e rotulagem realizada no conjunto de dados. Na seção subseção 4.1.6 são abordadas as adaptações feitas nas redes neurais e detalhes de implementação para segmentação de imagens aéreas. Na seção subseção 4.1.7 serão apresentados os métodos de treinamento e teste. Os resultados serão apresentados e discutidos no Capítulo 5.

4.1 Etapas da Metodologia Proposta

A metodologia utilizada para a realização dos objetivos propostos é formada pelas seguintes etapas: estudo sobre a APA-Petrópolis, configuração do ambiente de trabalho, aquisição do conjunto de dados, seleção e rotulagem, implementação do algoritmo e adaptações na rede neural para segmentação de imagens aéreas, treinamento e teste da rede e avaliação dos resultados. A Figura 18 descreve em diagrama de blocos as etapas descritas acima.

Figura 18 – Metodologia Proposta



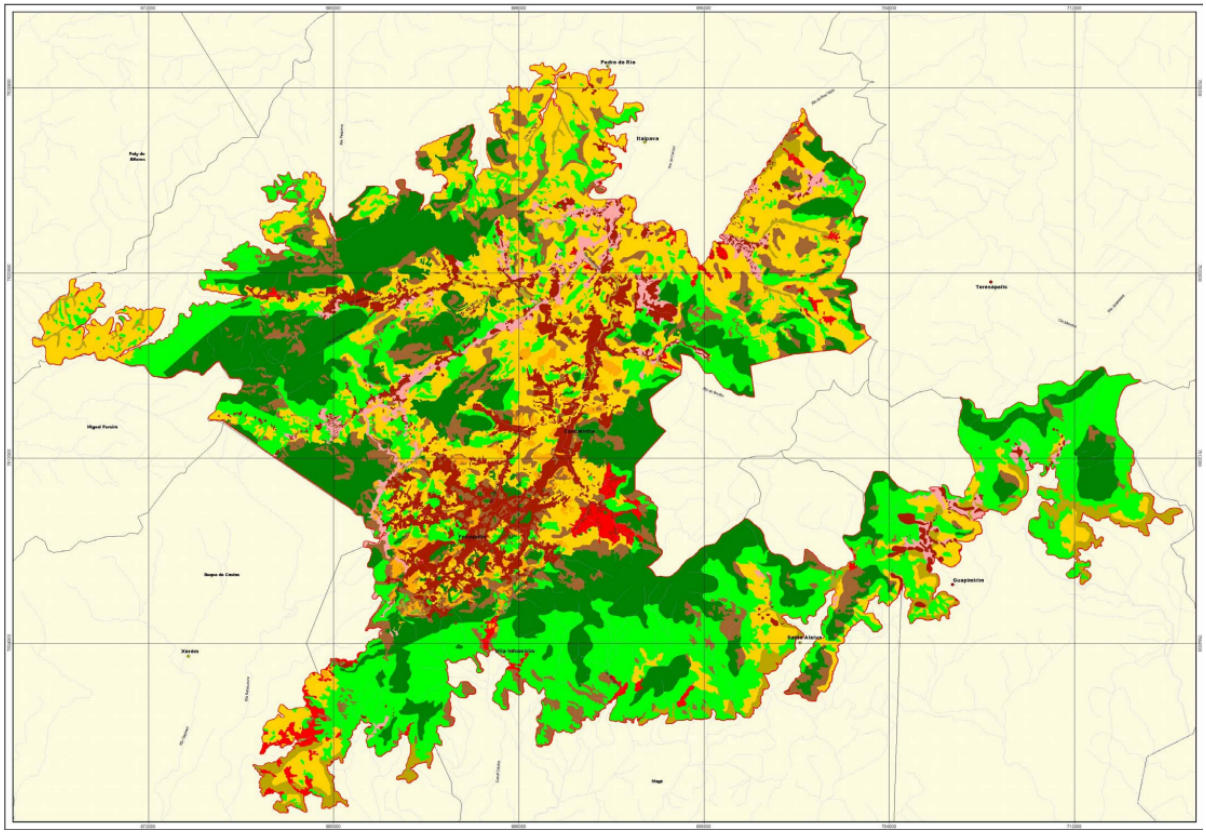
Fonte: Autoria Própria (2023).

4.1.1 Área de Estudo

A APA-Petrópolis abrange uma área de 59.618 hectares, equivalente a cerca de 5,69% das áreas protegidas da Mata Atlântica. Localizada em uma região urbanizada, contém importantes sítios de biodiversidade e endemismo. Sua gestão se dá através de um Plano de Manejo, que define ações e restrições para diferentes áreas. A APA possui um mapa de zoneamento

complexo que indica áreas com diferentes níveis de intervenção. Entretanto, o mapa de cobertura e uso do solo está desatualizado, levando a classificações incorretas das zonas, permitindo a exploração inadequada de áreas naturais. Principais atividades que afetam os recursos naturais da APA incluem agricultura extensiva, práticas agrícolas prejudiciais, expansão urbana, queimadas, mineração, extrativismo e poluição. O Zoneamento Ambiental, elaborado pelo Instituto Ecotema (2003), consiste em onze zonas com diretrizes de uso, baseadas em mapas de cobertura vegetal, suscetibilidade a fenômenos naturais e qualidade de vida. O zoneamento orienta o crescimento da cidade, recuperação ambiental e o desenvolvimento sustentável, envolvendo poder público e sociedade civil. O Plano de Manejo da APA foi desenvolvido com metodologia do IBAMA, incorporando estudos anteriores, como o Zoneamento Ambiental de 2003. A Figura 19 ilustra o mapa de zoneamento da APA-Petrópolis e seus limites entre os quatro municípios que abrangem a região, sendo que a maior parte da área pertence ao município de Petrópolis.

Figura 19 – Mapa de Zoneamento da APA-Petrópolis



Fonte: Petrópolis (2006).

4.1.2 Configuração do Ambiente de Trabalho

Para configurar o ambiente de trabalho foi preciso instalar um conjunto de bibliotecas e pacotes para a captura das imagens e a construção e execução da rede. Para obter as imagens, a ferramenta ArcGis Pro (ESRI INC., 2023) foi utilizada. Essa foi a única ferramenta utilizada que

necessita de licença para seu uso. A licença foi disponibilizada gratuitamente pela Universidade Federal do Mato Grosso do Sul (UFMS) por um período de um ano. Para o desenvolvimento do algoritmo de aprendizagem profunda, as principais bibliotecas utilizadas foram o PyTorch[1] (PASZKE *et al.*, 2017) e CUDA Toolkit[2] (NVIDIA Corporation, 2023). Todas as demais dependências foram instaladas com o auxílio da plataforma Miniconda[3] (INC, 2023). A linguagem de programação utilizada na implementação do algoritmo foi a linguagem Python[4] (ROSSUM, 2009). Para anotação, algumas imagens foram anotadas com a ferramenta ArcGis Pro e outras com a ferramenta *Computer Vision Annotation Tool* (CVAT)[5] (CVAT.AI, 2023). Mais detalhes sobre a metodologia de anotação das imagens será detalhada na subseção 4.1.4.

4.1.2.1 Configurações da Máquina

Os tópicos seguintes especificam as configurações da máquina utilizadas para a realização do trabalho:

- Processador: Intel(R) Xeon(R) CPU E5-2666 v3 @ 2.90GHz 2.90 GHz;
- Memória RAM: 32 GB;
- Placa de Vídeo: NVIDIA GIGABYTE RTX 3060 EAGLE OC – 12GB dedicada;
- Sistema Operacional: Microsoft Windows 10 PRO 64 bits.

O Quadro 3 apresenta as configurações específicas da placa de vídeo utilizada.

Quadro 3 – Configurações específicas da placa de vídeo utilizada.

Parâmetros	**Valores**
Modelo	NVIDIA RTX 3060 EAGLE OC
Interface	PCIe 4.0 x16
Total de Núcleos CUDA	3584
Clock Padrão GPU	1320Mhz
Quantidade de memória	12GB
Clock memória	1875Mhz
Interface de memória	192-bit GDDR6

Fonte: Autoria Própria (2023).

[1] https://pytorch.org/
[2] https://pytorch.org/get-started/locally/
[3] https://docs.conda.io/
[4] https://www.python.org/
[5] https://github.com/opencv/cvat

4.1.3 Aquisição do Conjunto de Dados

O conjunto de dados utilizado no trabalho foi obtido a partir de imagens RGB de satélite de alta resolução na plataforma do GE. As imagens foram obtidas entre maio e dezembro de 2022. Vale ressaltar que o procedimento realizado não permite selecionar uma data específica para obter as imagens.

As imagens foram extraídas por uma conexão feita na plataforma ArcGis Pro[6] (ESRI INC., 2023). Para obter as imagens, inicialmente foi feita uma conexão entre a plataforma ArcGIS Pro e a plataforma Google Earth utilizando o endereço `https://mt1.google.com/vt/lyrs=s&x=x&y=y&z=z`. Com isso, é possível acessar o mapa global e visualizar as imagens do Google Earth diretamente na plataforma. Ao localizar a APA-Petrópolis, várias áreas foram selecionadas aleatoriamente e exportadas. Para isso, foi necessário criar um polígono em cada região selecionada. Em seguida, cada região exportada foi importada novamente na plataforma ArcGIS para ser recortada em pedaços de 2048x2048 pixels. Essa etapa foi necessária para que, primeiro, toda a área selecionada pudesse ser exportada e, depois, a região pudesse ser recortada em pedaços menores e exportada em formato RGB.

O procedimento de exportação resultou em 322 imagens aéreas. Das 322 imagens exportadas, 214 foram selecionadas para treinamento e 42 para teste. As demais imagens foram desconsideradas pois muitas eram repetidas e poderiam aumentar o desbalanceamento entre as classes. A Tabela 1 ilustra os dados apontados anteriormente.

Tabela 1 – Separação das imagens no conjunto de dados.

Grupo	**Quantidade**
Treinamento	214 ($\approx$ 67%)
Teste	42 ($\approx$ 13%)
Desconsideradas	66 ($\approx$ 20%)
Total	**322**

Fonte: Autoria Própria (2023).

As imagens são representadas no formato *Tag Image File Format* (TIFF) e possuem uma resolução padrão de 2048 por 2048 pixels no espaço de cores RGB. A Figura 20 exibe um conjunto de oito imagens que fazem parte deste conjunto de dados.

4.1.4 Seleção e Rotulagem

Nesta etapa foi feita a seleção e rotulagem das imagens obtidas na subseção 4.1.3. O processo de rotulagem, utilizado para gerar *ground truth*, consiste em delimitar as regiões de interesse na imagem e indicar qual a sua respectiva classe ou categoria. Para a cobertura e uso

[6] Plataforma que constitui um sistema de informação geográfica

Figura 20 – Amostras do conjunto de dados

Fonte: Autoria Própria (2023).

do solo foram consideradas 8 classes: área desenvolvida, floresta, sombra, área em regeneração, solo exposto, água, rocha e agricultura. As características consideradas para cada classe estão elencadas na lista 4.1.4.

1. **Área desenvolvida**: objetos e superfícies que possuem algo construído. Por exemplo: casa, asfalto, calçada, carro, estufa;
2. **Floresta:** área verde dentro ou fora da área urbana. Por exemplo: Árvores, arbustos, grama, exceto área de mata que estão em regeneração;
3. **Sombra:** área escura que se forma no espaço atrás de um objeto que impede a passagem completa da luz. Por exemplo: sombra das árvores e das casas;
4. **Área em Regeneração:** área de mata que está em desenvolvimento;
5. **Solo Exposto:** terra exposta e estradas de terra;
6. **Água:** porções de água, tais como, rios e lagos;
7. **Rocha:** material sólido e natural composto por um ou mais minerais;
8. **Agricultura:** terra cultivada ou em preparação para o cultivo.

Algumas amostras de cada classe estão representadas na Figura 21, ordenadas por classe. A figura contém uma amostra de área desenvolvida (1), uma região de mata (2), uma área com sombra (3), uma área de mata em regeneração (4), uma estrada ou solo exposto (5), parte de um lago (6), parte de uma região rochosa (7) e uma área de plantio agrícola (8).

Figura 21 – Amostras do conjunto de dados por classe

Fonte: Autoria Própria (2023).

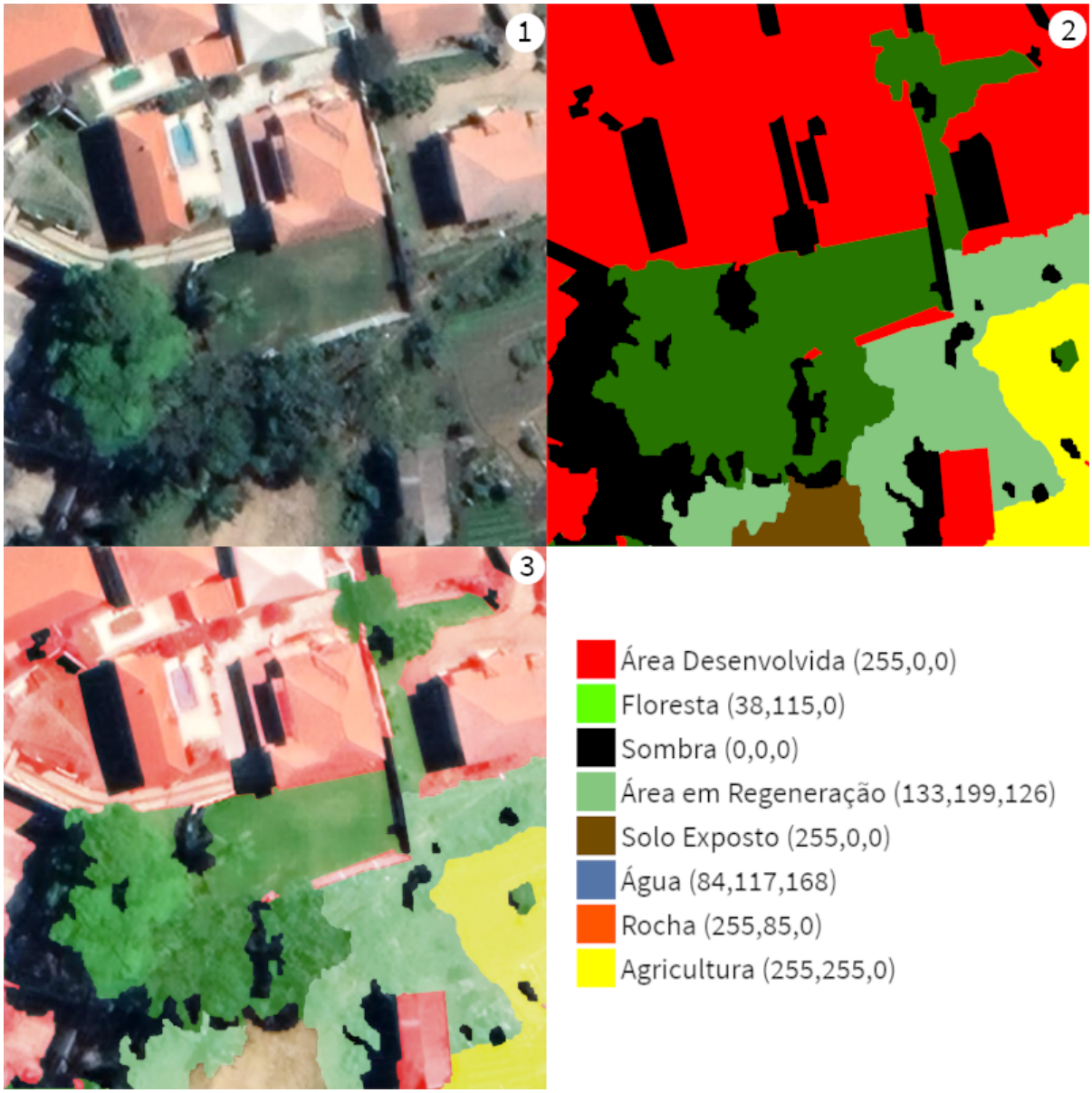
A partir da definição das classes, as imagens começaram a ser rotuladas com o auxílio de um profissional do ICMBio. Para esta rotulagem, uma nova imagem chamada de imagem de rótulos ou *label* é gerada, com o mesmo tamanho da imagem original. Nesta etapa, cada pixel da imagem é identificado utilizado o critério rígido para rotulagem das imagens, onde todos os pixels da imagem são rotuladas para alguma classe (ver 4.1.4), sem exceção, mesmo havendo algum impasse na identificação.

Uma parte das imagens foi rotulada na ferramenta ArcGis Pro e a outra parte na ferramenta livre CVAT. No ArcGis Pro, a rotulagem foi feita de maneira semi-automática. O recorte da área foi pré-segmentado pela ferramenta a partir de amostras de polígonos criados manualmente. Pela baixa qualidade das imagens, a segmentação automática resultou em muitos erros que precisaram ser corrigidos manualmente. Esse procedimento se mostrou ineficiente e impreciso. O restante das imagens foram rotuladas manualmente com a ferramenta CVAT. Nesta ferramenta, cada imagem de 2048 x 2048 pixels foi rotulada individualmente e manualmente. Tanto as imagens originais quanto as imagens rotuladas foram então exportadas em pastas separadas, respectivamente, em formato *TIFF* e *Portable Network Graphic (PNG)*. A Figura 22 ilustra uma imagem de amostra com a imagem original (1), a imagem rotulada (2), a imagem rotulada sobreposta a imagem original (3) e a legenda das classes e suas cores para facilitar a interpretação.

4.1.5 Metodologia Aplicada no Desbalanceamento de Classes

Para tratar o problema de desbalanceamento de classes, seguimos o que foi proposto no trabalho de Bressan *et al.* (2022). Para determinar o peso de cada classe, os autores utilizaram o conjunto de treinamento de acordo com a Equação 14. Quanto menor o número de pixels

Figura 22 – Amostra da imagem original e da imagem rotulada



Fonte: Autoria Própria (2023).

em uma determinada classe, maior o peso para que os filtros da camada da RNC se encaixem de forma uniforme. Quando $\varphi(c)$ é igual a 1 para todas as classes, o treinamento é realizado como tradicionalmente. É importante observar que este peso é o mesmo para todos os pixels da mesma classe C.

$$\varphi(c) = \frac{m}{C * n^c} \tag{14}$$

em que m é o número de pixels de todas as imagens de treinamento, C é o número de classes, e n^c é o número de pixels que pertencem à classe c.

Os valores calculados pela Equação 14 foram usados como pesos da função de custo de entropia cruzada para o treinamento da rede em alguns cenários. Os valores calculados

estão detalhados na Tabela 2. O resultado demonstra que a classe solo exposto tem o maior peso. Consequentemente, é a classe com menor quantidade de amostras de treinamento.

Tabela 2 – Pesos calculados para o conjunto de treinamento.

Classe	Peso
Área Desenvolvida	1,1472
Floresta	0,3832
Sombra	2,0468
Área em Regeneração	0,4436
Agricultura	1,4859
Rocha	2,4287
Solo Exposto	5,2043
Água	2,0031

Fonte: Autoria Própria (2023).

4.1.6 Adaptações na Rede SegNet e U-NET

Nas redes convolucionais a extração de características é feita de forma automática. As características são extraídas pelos filtros (em inglês, *kernels*) presentes em cada camada por meio de operações de convolução e *pooling*. Normalmente nas primeiras camadas são extraídas características mais simples e de difícil compreensão para seres humanos, como por exemplo as bordas. À medida que as camadas superiores vão sendo ativadas, características de mais alto nível são extraídas, ou seja, a tendência é que sejam obtidas características mais discriminativas. Neste trabalho, foram utilizadas duas arquiteturas de redes neurais completamente convolucionais para segmentação semântica, a SegNet e a U-NET, detalhadas a seguir.

Uma das redes utilizada para a realização da segmentação foi uma adaptação da rede original SegNet projetada por Badrinarayanan, Kendall e Cipolla (2017). Esta segue a arquitetura original descrita na subseção 2.6.4, exceto pelas alterações propostas no trabalho de Brito *et al.* (2021). Os autores propuseram uma nova estratégia de *multi-pooling*, substituindo o *max-pooling* por *Discrete Wavelet Transform (DWT)* e *unpooling* por *Inverse Discrete Wavelet Transform (IWT)*. Para os experimentos, os autores usaram imagens *Red-Green-Blue-Infrared* (RGB-IR)[7] dos conjuntos de dados das cidades de Potsdam e Vaihingen 2D Semantic Labeling Contest (2014), enquanto que o presente trabalho usou imagens RGB. Neste caso, trazendo uma contribuição a mais ao estudo inicial de Brito *et al.* (2021).

Em relação a U-NET, Nguyen-Khanh, Nguyen-Ngoc-Yen e Dinh-Quoc (2021) sugerem em seu artigo a combinação da eficiente arquitetura U-Net, que é uma combinação de EfficientNet, ou seja, EfficientNet-B0 (TAN; LE, 2019) como codificador para extrair as características com U-Net como decodificador para reconstruir o mapa de características. Os autores avaliaram

[7] Imagens que contêm informações de três bandas espectrais primárias - vermelho, verde e azul - além de uma quarta banda espectral no infravermelho próximo (NIR)

o modelo com imagens do GE e demonstram que mesmo com menor quantidade de parâmetros essa combinação obteve eficiência em termos da função de custo em relação a outras estruturas.

Os pesos de ambas as redes foram inicializados na parte codificadora usando os pesos pré-treinados na base de dados ImageNet (DENG *et al.*, 2009), que conta com 1 milhão e 200 mil imagens, rotuladas em 1000 classes, processo descrito como transferência de aprendizado.

4.1.7 Treinamento e Teste

Seguindo o protocolo utilizado no trabalho conduzido por Brito *et al.* (2021), o conjunto de dados foi dividido em dois subconjuntos distintos: treinamento e teste (ver Tabela 1). O conjunto de treinamento é responsável por permitir que a rede aprenda os padrões com base nos dados fornecidos e ajuste seus parâmetros. Por fim, o conjunto de teste é usado para avaliar o modelo final gerado após o treinamento.

Os autores implementaram um protocolo de pré-processamento de dados semelhante ao que foi adotado por Liu *et al.* (2017). Neste protocolo, janelas deslizantes com um tamanho fixo de 256 × 256 pixels são cortadas das amostras. O conjunto de imagens obtido é expandido ainda mais usando estratégias de aumento de dados compostas por espelhamento vertical e horizontal dos recortes (ver subseção 2.6.6). Este procedimento gera 10.000 imagens de treinamento, que mais tarde serão usadas para melhorar o processo de aprendizagem da rede. Para o subconjunto de teste, o mesmo foi aplicado com exceção do aumento de dados.

No subconjunto de teste, um procedimento de janela deslizante é empregado para cortar as imagens em pequenas partes para a entrada da rede. Entretanto, esta abordagem pode causar inconsistências na segmentação, especialmente nas bordas das imagens, o que pode prejudicar a precisão do modelo, como demonstrado por (LIU *et al.*, 2017). Para evitar essas bordas, seguimos Farhangfar e Rezaeian (2019). Os autores usam um passo de 32 pixels para a fase de treinamento e 16 pixels para a fase de teste. Para este trabalho, utilizou-se um passo de 32 pixels para a etapa de teste com o intuito de acelerar o teste para analisar os resultados. As múltiplas predições nessas áreas sobrepostas tornam o resultado mais suave e confiável. De acordo com Brito *et al.* (2021), esta abordagem sobreposta foi utilizada apenas na fase de testes porque os ganhos observados com a aplicação dessa técnica durante o treinamento da rede não justificavam o esforço computacional exigido por ela, especialmente em um ambiente com recursos computacionais limitados.

O treinamento da rede SegNet Modificada e da U-NET foi feito usando os pesos pré-treinados na ImageNet como pesos iniciais da rede. Isso tende a ajudar a rede a convergir mais rápido do que usar pesos aleatórios. Inicialmente um novo modelo com a arquitetura da rede é instanciado passando como parâmetro o número de classes e em qual dispositivo o modelo será gerado, possibilitando a utilização de Unidade Central de Processamento (UCP) ou UPG. Neste caso utilizou-se UPG. A arquitetura das redes foi implementada na biblioteca

PyTorch (PASZKE *et al.*, 2017). Os mesmos parâmetros foram usados em ambos os conjuntos de dados. Os seguintes hiperparâmetros foram usados no processo de aprendizado da rede: taxa de aprendizagem de 1e-2, tamanho do *batch* de 8, SGD como otimizador com parâmetro de *momentum* com valor de 0,9. A taxa de aprendizagem foi escalonada por um fator de 10 nas 25ª, 35ª e 45ª épocas. A redução da taxa de aprendizagem foi limitado a 1e-5. Os demais detalhes de configuração do ambiente de trabalho estão descritos na subseção 4.1.2. A escolha dos parâmetros considerou as escolhas feitas no trabalho realizado por Brito *et al.* (2021). As redes foram treinadas por 100 épocas. Uma época é um passo completo no treinamento, ou seja, quando toda a base de dados de treinamento é processada. Uma melhor visualização dos parâmetros utilizados no treinamento pode ser observada no Quadro 4

Quadro 4 – Hiperparâmetros da Rede

Parâmetro	**Valor**
Épocas de Treinamento	100
Tamanho do *Batch*	8
Taxa de Aprendizagem	1e-2
Momentum	0,9
Weight Decay	1e-5
Otimizador	SGD

Fonte: Autoria Própria (2023).

Outro fator considerado na etapa de treinamento foi o embaralhamento das imagens do grupo de treinamento. Durante o treinamento, é importante misturar as imagens na base de dados de treinamento. Se as imagens forem sempre apresentadas na mesma ordem, a rede neural pode se ajustar às imagens individuais em vez de aprender padrões gerais, o que resultaria na falta de capacidade de generalização da rede quando se trata de novas imagens. Ao embaralhar as imagens a cada época, garante-se que a rede neural aprenda os padrões de maneira mais robusta e generalizada.

Com base no detalhamento apresentado nesta seção, alguns cenários de treinamento e teste com diferentes configurações serão apresentados no Capítulo 5.

5 RESULTADOS

Neste capítulo são apresentados os resultados obtidos, bem como discussões sobre os principais problemas encontrados durante o treinamento e teste a partir da metodologia proposta no Capítulo 4.

5.1 Cenários

Este trabalho usou a abordagem de cenários para o treinamento e teste das redes SegNet e U-NET para apresentar os resultados. Os detalhes de implementação usados para cada cenário estão detalhados na subseção 4.1.6. A metodologia usada em cada cenário foi a mesma descrita na subseção 4.1.7, com algumas pequenas modificações que serão detalhadas em cada cenário. Assim, analisar o impacto de diferentes abordagens isoladamente torna-se mais organizado. Os cenários criados serão apresentados, comparados e discutidos neste capítulo a partir de métricas de avaliação discutidas na seção 2.7. A matriz de confusão nos fornece informações importantes sobre o desempenho do modelo de classificação para cada classe, auxiliando na identificação de classes com maior potencial de melhoria no desempenho do modelo. Além da matriz de confusão, é importante avaliar as métricas de desempenho, como precisão, sensibilidade, f1-score e IoU, para ter uma compreensão completa do desempenho geral do modelo. Essas métricas proporcionam dados adicionais sobre como o modelo lida com cada classe individualmente, considerando o suporte[1], e como equilibra o desempenho global entre todas as classes. Os quatro cenários estão elencados a seguir:

- Cenário um: SegNet Modificada, usando entropia cruzada como função de custo. O primeiro experimento usa pesos iguais (1,0) para todas as classes na função de custo, enquanto que o segundo experimento usa pesos ponderados. Vale destacar que este cenário considerou a utilização de 9 classes, incluindo a classe piscina que será desconsiderada nos cenários subsequentes.

- Cenário dois: SegNet Modificada, usando entropia cruzada como função de custo. O primeiro experimento usa aumento de dados, enquanto que o segundo experimento não utiliza. Ambos experimentos usam pesos ponderados na função de custo.

- Cenário três: U-NET, usando entropia cruzada como função de custo. Neste cenário, o primeiro experimento usa aumento de dados enquanto que o segundo experimento não utiliza.

- Cenário quatro: SegNet Modificada e U-NET. Neste cenário, será feita uma comparação entre as duas redes usando a função de custo *Focal Loss*.

[1] A métrica support refere-se ao número de amostras que pertencem a cada classe no conjunto de dados.

Para melhor visualização, as métricas que forem melhores em comparação ao experimento anterior serão apresentadas em negrito nas tabelas de resultado.

5.1.1 Cenário um: SegNet Modificada com entropia cruzada

A primeira arquitetura testada foi a SegNet com as modificações propostas no por Brito *et al.* (2021). Para este experimento, a função de custo utilizada foi a entropia cruzada (ver Equação 3). Para cada classe, a função de custo foi inicializada com pesos iguais (1.0). Os demais parâmetros são os mesmos definidos na subseção 4.1.7.

No primeiro experimento, como mostra a Tabela 3, nota-se que a rede obteve os melhores resultados de precisão para as classes Floresta, Sombra, Rocha e Agricultura. As mesmas classes também obtiveram bons resultados nas demais métricas apresentadas, exceto a classe Rocha que teve resultados ruins. O modelo obteve uma razoável acurácia, porém, o valor da métrica IoU pode ser considerado insatisfatório. Dentre os destaques negativos, as classes Água, Solo Exposto e a classe Piscina foram as que obtiveram os piores resultados.

Tabela 3 – Resultados da SegNet no cenário 1 com pesos iguais.

Classe	Precisão	Sensibilidade	F1-Score	IoU	Suporte
Área Desenvolvida	0.79	0.80	0.80	0.66	17035895
Floresta	0.89	0.87	0.88	0.78	60729653
Piscina	0.18	0.89	0.30	0.17	251264
Sombra	0.86	0.80	0.83	0.70	11518048
Área em Regeneração	0.73	0.89	0.80	0.66	35927788
Agricultura	0.82	0.83	0.82	0.70	28838187
Rocha	0.86	0.45	0.59	0.42	9374260
Solo Exposto	0.56	0.20	0.29	0.17	3166452
Água	0.45	0.06	0.11	0.05	930613
Acurácia	0.81	IoU Global		0.48	167772160

Fonte: Autoria Própria (2023).

No segundo experimento, ainda no mesmo cenário, a mesma rede SegNet foi experimentada novamente com uma pequena modificação nos pesos da função de custo. Agora, os pesos de cada classe foram ponderados na função de custo de entropia cruzada usando a estratégia descrita na subseção 4.1.5 com o objetivo de tentar minimizar o desbalanceamento entre as classes. Excepcionalmente para este cenário, os pesos foram ponderados conforme Tabela 4. Os demais cenários consideram os valores descritos na Tabela 2. Estes valores são diferentes pois a classe Piscina ainda está sendo considerada no treinamento e teste da rede. Conforme observado na Tabela 4, a classe piscina tem um peso muito grande em comparação as demais classes. Nos cenários futuros, a classe piscina foi concatenada com a classe Área Desenvolvida para tentar minimizar o desbalanceamento entre as classes.

Como mostra a Tabela 5, foram destacados em negrito as métricas que obtiveram resultados iguais ou superiores ao experimento anterior. Nota-se que a rede obteve os resultados

Tabela 4 – Pesos ponderados para a função de custo no cenário 1.

Classe	Peso
Área Desenvolvida	0,9786
Floresta	0,3264
Piscina	51,3827
Sombra	1,7031
Área em Regeneração	0,3702
Agricultura	1,2344
Rocha	2,0176
Solo Exposto	4,8396
Água	10,5859

Fonte: Autoria Própria (2023).

de precisão ligeiramente superiores para as classes Área em Regeneração e Água. A sensibilidade obtida pela classe Piscina, que é a classe mais desbalanceada, foi alta em comparação a sua precisão, já que a sensibilidade preocupa-se apenas com a classificação das amostras positivas. Para os valores de IoU e f1-score, destaque positivo para a classe Água, porém, as demais classes obtiveram valores inferiores ao experimento anterior em praticamente todas as métricas.

Tabela 5 – Resultados da SegNet no cenário 1 com pesos ponderados

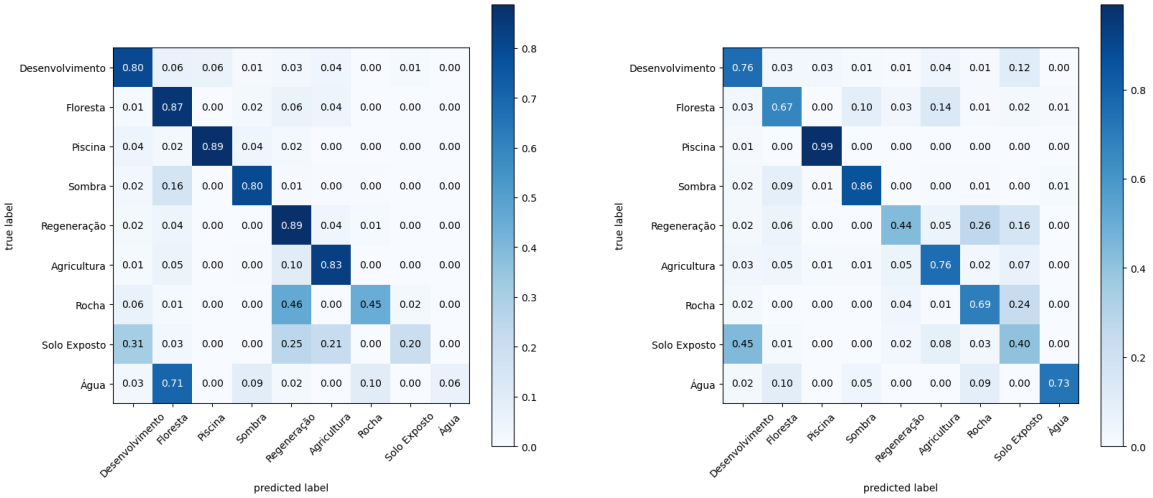
Classe	Precisão	Sensibilidade	F1-Score	IoU	Suporte
Área Desenvolvida	0.71	0.76	0.73	0.57	17035895
Floresta	0.89	0.67	0.76	0.61	60729653
Piscina	0.18	**0.99**	0.30	0.18	251264
Sombra	0.60	**0.86**	0.71	0.55	11518048
Área em Regeneração	**0.80**	0.44	0.56	0.39	35927788
Agricultura	0.66	0.76	0.70	0.54	28838187
Rocha	0.37	**0.69**	0.48	0.31	9374260
Solo Exposto	0.09	**0.40**	0.14	0.08	3166452
Água	**0.46**	**0.73**	**0.56**	**0.39**	930613
Acurácia	0.65	IoU Global		0.40	167772160

Fonte: Autoria Própria (2023).

Analisando a matriz de confusão presente na Figura 23, podemos identificar que as classes Piscina (87%), Área em Regeneração (89%) e Agricultura (83%) parecem ter o melhor desempenho, com altos valores de Verdadeiros Positivos (acertos) e baixos valores de Falsos Positivos e Falsos Negativos (erros). Por outro lado, as classes Solo Exposto (20%), Rocha (45%) e principalmente a classe Água (6%) obtiveram baixos valores de acertos, com valores relativamente altos de Falsos Positivos e Falsos Negativos, sugerindo que precisam de ajustes. A classe Água confundiu-se muito com a classe Floresta enquanto que a classe Rocha foi confundida com a classe Área em Regeneração. Com relação a matriz de confusão apresentada na Figura 23, nota-se que ponderar os pesos na função de custo equilibrou um pouco o resultado em comparação a matriz de confusão anterior. O destaque positivo foi para a classe Água com

73% de acertos enquanto que a mesma classe obteve apenas 0,6% de acertos no resultado anterior. Contudo, a classe Área em Regeneração obteve apenas 44% de acertos enquanto que no experimento anterior foram 89%.

Figura 23 – Matriz de Confusão no cenário 1: pesos iguais (à esquerda) e pesos ponderados (à direita).



Fonte: Autoria Própria (2023).

Considerando os dois treinamentos com a rede SegNet com modificações de pesos na função de custo, identificamos que o peso da classe Piscina é muito desproporcional em comparação as outras classes (ver Tabela 4), visto que é a classe que possui a menor quantidade de amostras anotadas no conjunto de dados. Para tentar minimizar o erro, os próximos cenários desconsideram esta classe. A classe Piscina será mesclada com a classe Área Desenvolvida.

5.1.2 Cenário dois: SegNet Modificada com entropia cruzada

Neste cenário, o treinamento ocorreu da mesma forma que o cenário um. A diferença entre os cenários foi a classe Piscina. As amostras da classe foram incorporada à classe Área Desenvolvida. O primeiro experimento não teve modificações nos parâmetros de treinamento, enquanto que no segundo experimento não foi aplicado aumento de dados. Ambos experimentos usam pesos ponderados na função de custo.

No primeiro experimento, como mostra a Tabela 6, nota-se que a rede obteve de maneira geral melhores resultados em comparação ao cenário 1 para todas as métricas apresentadas, principalmente para a classe Água. Contudo, a classe Solo Exposto continua a apresentar resultados insatisfatórios. O resultado demonstra que a remoção da classe com maior desbalanceamento (a classe Piscina) foi fundamental para melhorar a precisão do modelo. O f1-score maior mostra que a arquitetura melhorou em classificar a classe (classifica menos vezes, mas quando o faz é provavelmente mais correto).

Tabela 6 – Resultados da SegNet no cenário 2 com aumento de dados

Classe	Precisão	Sensibilidade	F1-Score	IoU	Suporte
Área Desenvolvida	0.83	0.77	0.80	0.66	17287159
Floresta	0.94	0.77	0.85	0.74	64497596
Sombra	0.69	0.93	0.79	0.66	11525775
Área em Regeneração	0.83	0.66	0.73	0.58	35927788
Agricultura	0.83	0.77	0.80	0.66	28838187
Rocha	0.50	0.77	0.61	0.43	9374260
Solo Exposto	0.17	0.79	0.28	0.16	3166452
Água	0.72	0.89	0.79	0.66	5543551
Acurácia	0.76	IoU Global		0.57	176160768

Fonte: Autoria Própria (2023).

No segundo experimento, ainda no mesmo cenário, a mesma rede SegNet foi experimentada novamente sem aplicar a estratégia de aumento de dados. Os resultados obtidos foram ligeiramente superiores ao treinamento anterior com aumento de dados. Como mostra a Tabela 7, foram destacados em negrito as métricas que obtiveram melhores resultados. Percebe-se que as métricas de precisão e f1-score foram superiores em todas as classes, enquanto que a métrica IoU só não foi superior para a classe Floresta. Assim como no experimento anterior, a classe Solo Exposto ainda está bem abaixo.

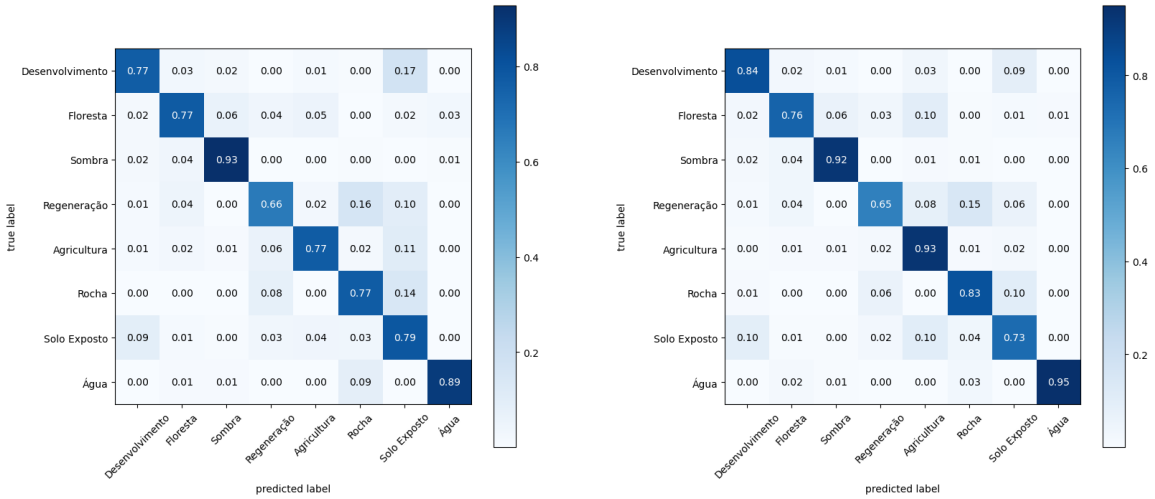
Tabela 7 – Resultados da SegNet no cenário 2 sem aumento de dados

Classe	Precisão	Sensibilidade	F1-Score	IoU	Suporte
Área Desenvolvida	**0.84**	**0.84**	**0.84**	**0.72**	17287159
Floresta	**0.95**	0.76	**0.85**	0.73	64497596
Sombra	**0.70**	0.92	**0.80**	**0.67**	11525775
Área em Regeneração	**0.88**	0.65	**0.75**	**0.60**	35927788
Agricultura	**0.72**	**0.93**	**0.81**	**0.69**	28838187
Rocha	**0.55**	**0.83**	**0.66**	**0.50**	9374260
Solo Exposto	**0.27**	0.73	**0.40**	**0.25**	3166452
Água	**0.89**	**0.95**	**0.92**	**0.85**	5543551
Acurácia	**0.79**	IoU Global		**0.62**	176160768

Fonte: Autoria Própria (2023).

Comparando os resultados a partir da matriz de confusão, percebe-se que o resultado do experimento com aumento de dados apresentado na Figura 24 não foi superior na maioria dos casos ao comparar com o resultado apresentado na Figura 24 sem aumento de dados. A classe Água e Agricultura, por exemplo, obtiveram respectivamente 95% e 93% de acertos sem aumento de dados. Em resumo, ambos os cenários mostram um bom desempenho geral do modelo, com altas taxas de classificação correta para a maioria das classes. O cenário 2 sugere apresentar pequenas melhorias em algumas classes, reduzindo as taxas de Falsos Positivos e Falsos Negativos.

Figura 24 – Matriz de Confusão no cenário 2: com aumento de dados (à esquerda) e sem aumento de dados (à direita).



Fonte: Autoria Própria (2023).

Considerando os dois treinamentos com a rede SegNet no cenário dois, um com aumento de dados e outro não, observamos que os resultados obtidos foram superiores, principalmente no segundo experimento sem aumento de dados. Apesar dos resultados terem sido melhores, as métricas das classes Rocha e Solo Exposto ainda são baixas. O baixo valor obtida com a métrica IoU evidencia erros na anotação das amostras.

5.1.3 Cenário três: U-NET com aumento de dados

Neste cenário, o treinamento ocorreu da mesma forma que os cenários anteriores, com exceção ao modelo de rede neural. Neste cenário, utilizamos a rede U-NET. Os experimentos apresentados neste cenário usaram mesma função de custo de entropia cruzada com pesos ponderados (ver Tabela 2). A diferença entre os experimentos apresentados neste cenário é o aumento de dados que não foi aplicado no segundo experimento.

No primeiro experimento, como mostra a Tabela 8, nota-se que a rede U-NET obteve de maneira geral resultados muito similares na maioria das classes em comparação a rede SegNet apresentada no cenário anterior. As classes Área Desenvolvida, Floresta, Agricultura, Solo Exposto e Água obtiveram melhores valores de precisão e a acurácia global melhorou de 79% para 81%.

No segundo experimento, como mostra a Tabela 8, nota-se que sem aumento de dados a maioria das classes obtiveram resultados melhores em comparação ao primeiro experimento deste cenário, com exceção da classe Água. A precisão e o índice IoU global também foram melhores.

Considerando a matriz de confusão para o experimento com U-NET com aumento de dados na Figura 25 e sem aumento de dados na Figura 25, percebe-se, de maneira geral, que

Tabela 8 – Resultados da U-NET no cenário 3 com aumento de dados

Classe	Precisão	Sensibilidade	F1-Score	IoU	Suporte
Área Desenvolvida	0.89	0.81	0.85	0.73	17287159
Floresta	0.96	0.80	0.87	0.77	64497596
Sombra	0.68	0.94	0.79	0.65	11525775
Área em Regeneração	0.86	0.64	0.73	0.58	35927788
Agricultura	0.80	0.96	0.87	0.77	28838187
Rocha	0.52	0.91	0.66	0.49	9374260
Solo Exposto	0.31	0.73	0.43	0.28	3166452
Água	0.96	0.93	0.95	0.90	5543551
Acurácia	0.81	IoU Global		0.65	176160768

Fonte: Autoria Própria (2023).

Tabela 9 – Resultados da U-NET no cenário 3 sem aumento de dados

Classe	Precisão	Sensibilidade	F1-Score	IoU	Suporte
Área Desenvolvida	0.89	**0.87**	**0.88**	**0.79**	17287159
Floresta	0.92	**0.89**	**0.91**	**0.83**	64497596
Sombra	**0.87**	0.83	**0.85**	**0.73**	11525775
Área em Regeneração	0.82	**0.88**	**0.85**	**0.74**	35927788
Agricultura	**0.86**	0.95	**0.90**	**0.82**	28838187
Rocha	**0.80**	0.71	**0.75**	**0.60**	9374260
Solo Exposto	**0.70**	0.42	**0.52**	**0.35**	3166452
Água	**0.98**	0.88	0.93	0.86	5543551
Acurácia	0.87	IoU Global		0.72	176160768

Fonte: Autoria Própria (2023).

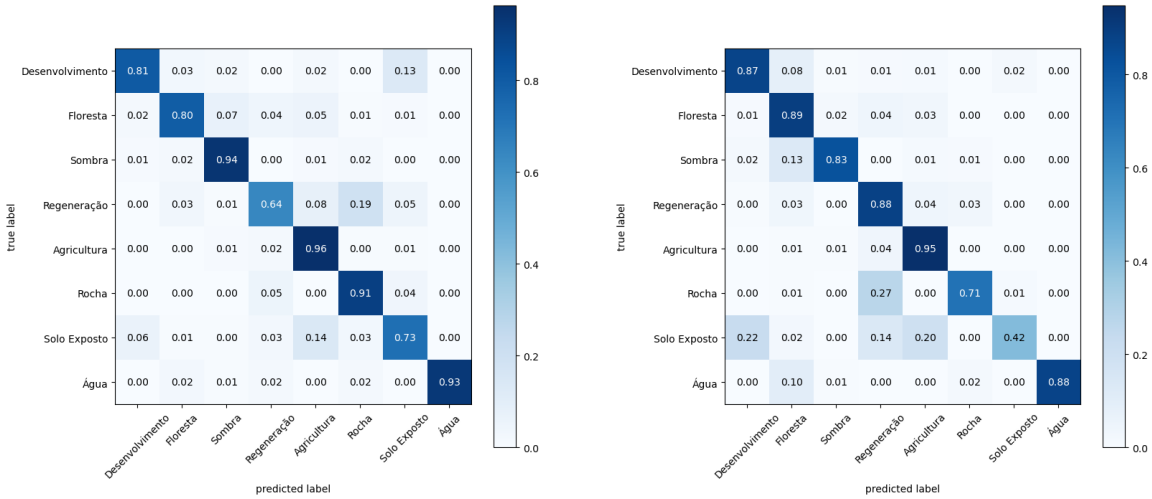
ambos apresentam desempenhos semelhantes para a maioria das classes, com altas taxas de classificação correta. No entanto, o segundo experimento sem aumento de dados, parece apresentar melhorias para as classes Desenvolvimento, Floresta e Área em Regeneração, reduzindo as taxas de Falsos Positivos. Contudo, a classe Solo Exposto ainda apresenta muitos erros, confundindo-se principalmente com a classe Desenvolvimento e Agricultura.

Considerando os dois treinamentos com a rede U-NET no cenário três, um com aumento de dados e outro não, observamos que os resultados obtidos foram mais uma vez superiores, principalmente no segundo experimento sem aumento de dados. Em comparação com o cenário dois (ver subseção 5.1.2), usando a rede Segnet, os valores obtidos com a rede U-NET foram ligeiramente superiores.

5.1.4 Cenário quatro: SegNet e U-NET com *Focal Loss*

Consideramos agora uma comparação entre a rede SegNet (usada nos cenários um e dois) e a rede U-NET (usada no cenário três). Diferentemente dos cenários anteriores, agora realizamos dois experimentos usamos a função de custo *focal loss* (ver Equação 4) com o

Figura 25 – Matriz de Confusão no cenário 3: com aumento de dados (à esquerda) e sem aumento de dados(à direita).



Fonte: Autoria Própria (2023).

hiperparâmetro γ igual a dois, ambos com aumento de dados. No primeiro experimento, como pode ser visto na Tabela 10,

Tabela 10 – Resultados da SegNet no cenário 4 com *Focal Loss*

Classe	Precisão	Sensibilidade	F1-Score	IoU	Suporte
Área Desenvolvida	0.85	0.86	0.86	0.75	17287159
Floresta	0.89	0.90	0.89	0.81	64497596
Sombra	0.89	0.74	0.81	0.68	11525775
Área em Regeneração	0.75	0.89	0.81	0.68	35927788
Agricultura	0.85	0.85	0.85	0.74	28838187
Rocha	0.61	0.53	0.57	0.40	9374260
Solo Exposto	0.52	0.22	0.31	0.18	3166452
Água	0.98	0.43	0.60	0.42	5543551
Acurácia	0.83	IoU Global		0.58	176160768

Fonte: Autoria Própria (2023).

No segundo experimento, como mostra a Tabela 11, foram destacados em negrito as métricas que obtiveram melhores valores iguais ou superiores ao experimento anterior. Percebe-se que praticamente todas as métricas foram iguais ou ligeiramente superiores para a U-NET em comparação a SegNet Modificada usando a mesma função de custo. Apesar da melhoria no resultado, o valor do índice IoU para a classe Solo Exposto ainda está muito baixo, indicando erro entre a predição do modelo e a anotação feita nas amostras.

Analisando a matriz de confusão do cenário quatro, presente na Figura 26 para a rede SegNet e na Figura 26 para a rede U-NET, percebe-se que os experimentos têm resultados similares em algumas classes. O destaque vai para a classe Água que obteve 89% de acerto no experimento com a rede U-NET. Com a rede SegNet, a classe obteve apenas 43% de acerto.

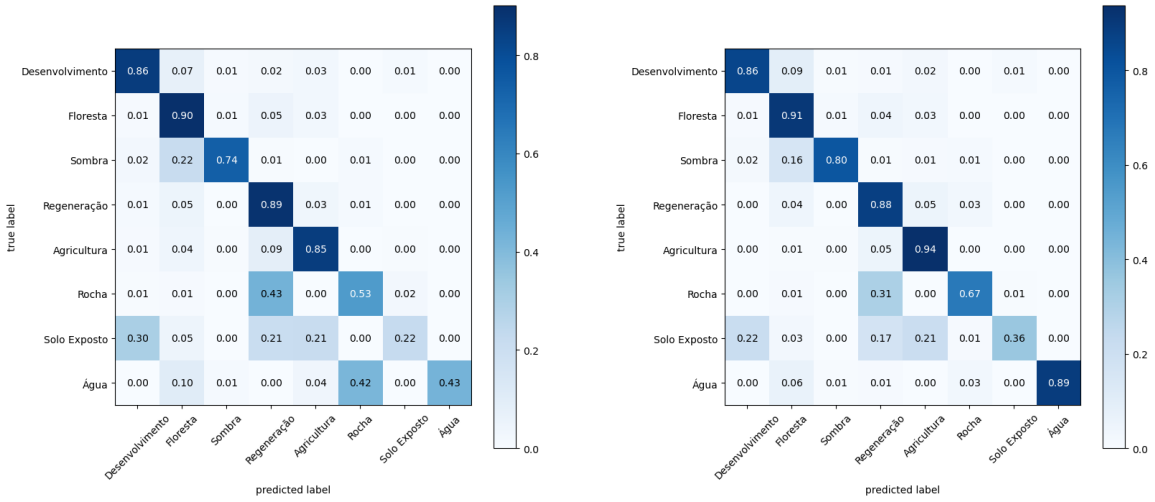
Tabela 11 – Resultados da U-NET no cenário 4 com *Focal Loss*

Classe	Precisão	Sensibilidade	F1-Score	IoU	Suporte
Área Desenvolvida	**0.90**	**0.86**	**0.88**	**0.79**	17287159
Floresta	**0.91**	**0.91**	**0.91**	**0.83**	64497596
Sombra	**0.89**	**0.80**	**0.84**	**0.73**	11525775
Área em Regeneração	**0.80**	0.88	**0.84**	**0.72**	35927788
Agricultura	**0.85**	**0.94**	**0.89**	**0.81**	28838187
Rocha	**0.81**	**0.67**	**0.74**	**0.58**	9374260
Solo Exposto	**0.75**	**0.36**	**0.49**	**0.32**	3166452
Água	**0.99**	**0.89**	**0.94**	**0.88**	5543551
Acurácia	**0.87**	IoU Global		**0.71**	176160768

Fonte: Autoria Própria (2023).

Grande parte dos Falsos Positivos para a classe Água ficaram concentradas na classe Rocha com 42%.

Figura 26 – Matriz de Confusão no cenário 4: SegNet com *Focal Loss* (à esquerda) e U-NET com *Focal Loss* (à direita).



Fonte: Autoria Própria (2023).

5.2 Análise das Métricas por Cenário

Para cada cenário, os resultados foram analisados em termos de métricas como acurácia, precisão, sensibilidade, f1-Score e IoU. Observou-se que as diferentes abordagens e variações nos parâmetros resultaram em impactos variados no desempenho das redes. Alguns cenários apresentaram resultados melhores do que outros em termos de métricas de avaliação.

Conforme descrito na seção 2.7, uma métrica bastante presente para avaliação do desempenho geral das redes é a acurácia. Ela serve para avaliar o desempenho geral da rede. A Tabela 12 contém a acurácia de cada experimento em cada cenário. Analisando a acurácia global, podemos inferir que o segundo experimento do cenário 4, que usou como parâmetro a rede

U-NET e a função de custo *Focal Loss* obteve os melhores resultados, contudo, não melhorou em comparação ao cenário 3 que utilizou a função de custo entropia cruzada. Assim, temos que no geral a melhor rede quanto à acurácia global foi a U-NET.

Tabela 12 – Comparativo de acurácia dos experimentos em cada cenário.

	Cenário 1	Cenário 2	Cenário 3	Cenário 4
Experimento 01	0.81	0.76	0.81	0.83
Experimento 02	0.65	0.79	0.87	0.87

Fonte: Autoria Própria (2023).

Além da acurácia, a métrica de IoU, descrita na subseção 2.8.4 mede a sobreposição entre a área prevista pelo modelo e a área verdadeira (anotada manualmente) da área segmentada. A Tabela 13 contém o valor de IoU de cada experimento em cada cenário. Percebe-se que o segundo experimento do cenário três obteve o melhor resultado. Este experimento usou a rede U-NET e entropia cruzada como função de custo sem aumento de dados e obteve um resultado superior ao primeiro experimento do mesmo cenário, com aumento de dados. Novamente, o resultado foi bem similar ao segundo experimento do cenário quatro.

Tabela 13 – Comparativo de IoU dos experimentos em cada cenário

	Cenário 1	Cenário 2	Cenário 3	Cenário 4
Experimento 01	0.48	0.57	0.65	0.58
Experimento 02	0.40	0.62	0.72	0.71

Fonte: Autoria Própria (2023).

A análise global feita para as medidas de acurácia e IoU demonstram que os cenários três e quatro obtiveram os melhores resultados. Sendo assim, vamos agora analisar as medias IoU e a medida f para cada classe considerando apenas estes dois cenários.

A métrica f1-score é uma medida amplamente utilizada na avaliação de modelos de segmentação de imagens. Ela combina a precisão e a sensibilidade em uma única pontuação, fornecendo uma medida geral do desempenho do modelo (ver subseção 2.8.3). A Tabela 14 mostra o desempenho do modelo em cada classe para os cenários três e quatro que obtiveram os melhores resultados conforme análise feita no início desta seção e também na seção 5.1. Após analisar e comparar os valores, os valores indicam que o modelo obteve resultados muito similares principalmente no segundo experimento de cada cenário. Assim, podemos também inferir que mudar apenas a função de custo entre os cenários não resultou em resultados melhores. Dentre as classes, o destaque negativo continua sendo para a classe Solo Exposto que obteve baixo índice de acertos, conforme pode ser observado ao analisar a matriz de confusão de cada cenário na seção 5.1. Ao analisar a medida IoU apresentada na Tabela 15, a conclusão é a mesma. Os melhores resultados foram muito similares entre o segundo experimento de cada cenário.

Tabela 14 – Resultados da medida f dos experimentos nos melhores cenários

Classe	Cenário 3		Cenário 4	
	Exp 1	Exp 2	Exp 1	Exp 2
Área Desenvolvida	0.85	0.88	0.86	0.88
Floresta	0.87	0.91	0.89	0.91
Sombra	0.79	0.85	0.81	0.84
Área em Regeneração	0.73	0.85	0.81	0.84
Agricultura	0.87	0.90	0.85	0.89
Rocha	0.66	0.75	0.57	0.74
Solo Exposto	0.43	0.52	0.31	0.49
Água	0.95	0.93	0.60	0.94

Fonte: Autoria Própria (2023).

Tabela 15 – Resultados da medida IoU dos experimentos nos melhores cenários

Classe	Cenário 3		Cenário 4	
	Exp 1	Exp 2	Exp 1	Exp 2
Área Desenvolvida	0.73	0.79	0.75	0.79
Floresta	0.77	0.83	0.81	0.83
Sombra	0.65	0.73	0.68	0.73
Área em Regeneração	0.58	0.74	0.68	0.72
Agricultura	0.77	0.82	0.74	0.81
Rocha	0.49	0.60	0.40	0.58
Solo Exposto	0.28	0.35	0.18	0.32
Água	0.90	0.86	0.42	0.88

Fonte: Autoria Própria (2023).

5.3 Análise das Imagens

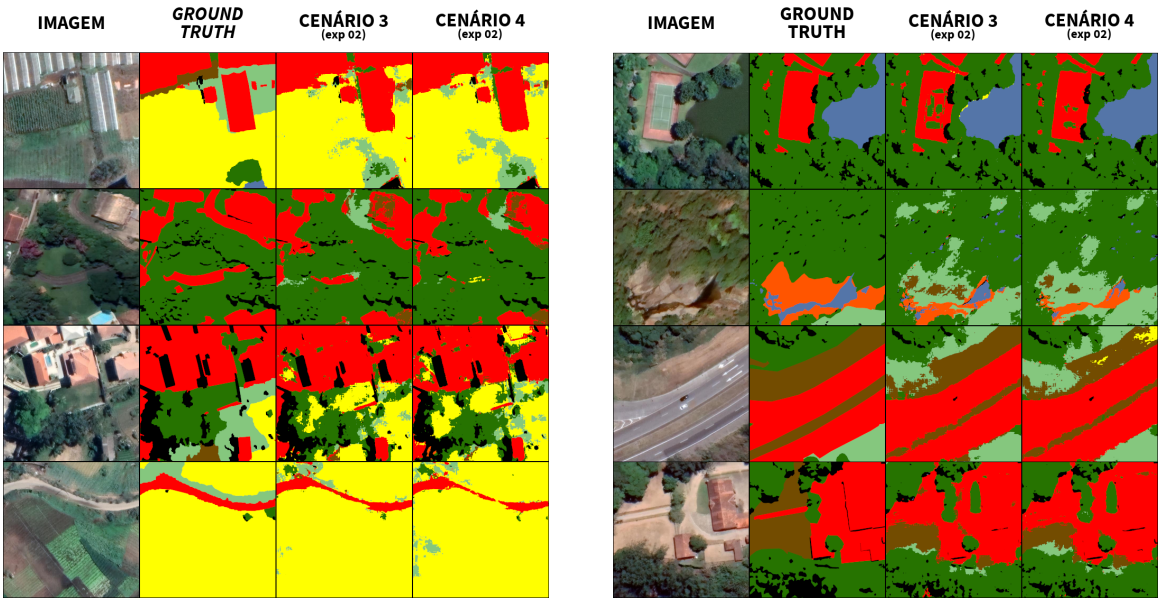
Após analisar as medidas de desempenho na seção 5.3 é possível observar que os melhores resultados foram obtidos no segundo experimento do cenário três e do cenário quatro. O segundo experimento do cenário três (subseção 5.1.3) usou a rede U-NET sem aumento de dados com a função de custo de entropia cruzada com pesos ponderados, já o segundo experimento do cenário 4 (subseção 5.1.4) usou a mesma rede U-NET mas com a função de custo *focal loss*.

Assim, esta seção vai apresentar algumas imagens resultantes da inferência nas imagens do conjunto de teste destes dois cenários para que possa ser comparada visualmente com a anotação (*ground truth*) feita manualmente.

A Figura 27 e a Figura 27 apresentam oito das 42 imagens do conjunto de teste. Essas imagens foram escolhidas para serem apresentadas aqui pois abrangem diversas classes. Da esquerda para a direita a imagem original, o *ground truth*, a inferência no cenário três e a inferência no cenário quatro. Para melhor visualização e comparação, as imagens foram redimensionadas para um tamanho de 256x256 e agrupadas.

A análise visual do resultado reforça a análise feita na seção 5.3. A classe Rocha confundiu-se com as classes Área em Regeneração e Solo Exposto. Já a classe Solo Exposto

Figura 27 – Comparativo entre as Imagens do Conjunto de Teste: 4 amostras (à esquerda) e 4 amostras (à direita).



Fonte: Autoria Própria (2023).

principalmente com Área Desenvolvida e Área em Regeneração. Outros erros também foram observados, tais como, a classificação de áreas anotadas como Floresta e Área em Regeneração que foram inferidas para a classe Agricultura.

5.3.1 Comparações com trabalhos relacionados

Este seção vai fazer uma comparação entre os resultados apresentados na seção seção 5.1 em comparação com trabalhos relacionados na Capítulo 3.

Dentre os trabalhos relacionados, o trabalho de Nguyen-Khanh, Nguyen-Ngoc-Yen e Dinh-Quoc (2021) mais se aproxima ao presente trabalho proposto. Para o conjunto de dados, os autores também usaram imagens de satélite do Google e fizeram as anotações manuais para as classes rua, árvore, água, residencial, urbano, edifícios, industrial e comercial, terreno urbano vago, floresta esparsa, parque, grama, agrícola, urbano esparso. Para anotação, os autores também usaram a ferramenta CVAT (CVAT.AI, 2023). Além disso, temos outras semelhanças apresentadas neste trabalho e no trabalho de Nguyen-Khanh, Nguyen-Ngoc-Yen e Dinh-Quoc (2021).

Ambos os trabalhos usam redes que seguem o conceito de *Encoder-Decoder* que consiste em extrair características ao mesmo tempo que reduz a sua resolução. Isso diminui o custo computacional da rede para em seguida reconstrução das imagens para seu tamanho original de entrada.

Outro ponto correlato pode ser observado a partir dos resultados obtidos. No trabalho de Nguyen-Khanh, Nguyen-Ngoc-Yen e Dinh-Quoc (2021), os autores usaram a rede U-NET e

compararam os resultados com diferentes *backbones* e funções de custo. O melhor resultado obtido foi com o *backbone* EfficientNet-B0. Em nossos resultados, o experimento que obteve os melhores resultados usou a mesma rede e *backbone*. Os autores não apresentaram métricas de avaliação já que o objetivo deles era comparar diferentes funções de custo e *backbone* para a rede U-NET.

5.3.2 Considerações Finais

Neste capítulo, apresentamos e analisamos os resultados obtidos por meio da metodologia descrita no Capítulo 4. Os resultados revelaram um desempenho superior da rede U-NET em comparação com a rede SegNet modificada, além de um tempo de treinamento e teste significativamente menor, considerando as configurações de hardware disponíveis (consulte a subseção 4.1.2.1 para detalhes). Realizamos comparações entre diferentes cenários, conduzindo dois experimentos em cada um dos quatro cenários propostos. Além disso, relatamos e comparamos os resultados dos experimentos usando métricas de avaliação e também fornecemos uma breve análise comparativa com um dos estudos relacionados mais próximos à nossa metodologia. No próximo capítulo, abordaremos as conclusões deste trabalho e apresentaremos sugestões para pesquisas futuras.

6 CONCLUSÃO

A avaliação manual do uso e cobertura do solo requer uma expertise significativa por parte dos avaliadores, bem como uma equipe qualificada e recursos substanciais. Nesse contexto, a aplicação de abordagens computacionais para o mapeamento do uso e cobertura do solo em unidades de preservação emerge como fundamental, pois permite identificar áreas prioritárias para a conservação e aquelas suscetíveis à expansão da urbanização e ocupação.

Dado a problemática, esta dissertação introduziu uma metodologia de baixo custo para o mapeamento do uso e cobertura do solo em áreas de proteção ambiental, utilizando adaptações em redes neurais convolucionais para a segmentação semântica de imagens aéreas adquiridas na plataforma GE.

A abordagem proposta envolveu o emprego de duas redes neurais para a segmentação semântica: a SegNet, com adaptações similares às apresentadas por Brito *et al.* (2021), e a U-NET, sem modificações na sua estrutura original. Os experimentos foram conduzidos em quatro cenários distintos, cada um com configurações e parâmetros diferentes, conforme detalhado na Capítulo 5. Os resultados obtidos no estudo apresentam um aumento de performance para todas as métricas de avaliação a medida que novos cenários foram sendo ajustados, treinados e testados.

Nesse sentido, a metodologia de baixo custo proposta destaca a viabilidade do mapeamento de uso e cobertura do solo utilizando exclusivamente imagens aéreas RGB de alta resolução da plataforma GE. A contribuição significativa deste trabalho é evidenciada ao lidar com a limitação do conjunto de dados, que é intrinsecamente desbalanceado e restrito em virtude das dificuldades na anotação das imagens, juntamente com a escassez de recursos financeiros e humanos. A análise comparativa revela a superioridade da rede U-NET em relação à SegNet. Destaca-se que os resultados foram otimizados, especialmente após a fusão das classes Piscina e Área Desenvolvida no segundo cenário. Os desempenhos mais destacados foram alcançados no segundo experimento do cenário três, com a aplicação de ponderação de pesos na função de custo de entropia cruzada, e no segundo experimento do cenário quatro, utilizando a função de custo *focal loss*. Essas descobertas ilustram a eficácia da abordagem proposta em cenários desafiadores com recursos limitados.

6.1 Trabalhos Futuros

Para aprimorar ainda mais o mapeamento do uso e cobertura do solo em áreas de proteção ambiental, empregando abordagens computacionais de baixo custo, algumas sugestões para futuras pesquisas incluem:

- **Aprimoramento da Anotação de Dados:** Melhorar o processo de anotação de dados, incorporando técnicas semi ou não supervisionadas. Isso pode incluir a utilização

de algoritmos de aprendizagem para otimizar o processo de rotulação das imagens, especialmente em áreas complexas.

- **Expansão do Conjunto de Dados:** Incrementar a quantidade de imagens no conjunto de dados com o objetivo de equalizar as classes e aprimorar a capacidade de generalização do modelo.
- **Otimização do Treinamento:** Identificar e implementar estratégias de aprimoramento no processo de treinamento e teste, visando aperfeiçoar os resultados do modelo e prevenir imprecisões na segmentação.
- **Integração de Redes do Tipo Transformer:** Explorar o potencial das redes do tipo Transformer, originalmente desenvolvidas para tarefas de processamento de linguagem natural, na segmentação semântica de imagens do Google Earth.

Essas diretrizes representam oportunidades promissoras para avançar na aplicação de técnicas computacionais no mapeamento do uso e cobertura do solo em áreas de proteção ambiental, contribuindo para uma gestão mais eficiente e sustentável dessas importantes regiões por parte das instituições governamentais.

REFERÊNCIAS

2D Semantic Labeling Contest. *2D Semantic Labeling Contest*. 2014. Disponível em: <https://www.isprs.org/education/benchmarks/UrbanSemLab/semantic-labeling.aspx>. Acesso em: 28 mar. 2023.

ALAM, M. *et al.* Convolutional neural network for the semantic segmentation of remote sensing images. **Mobile Networks and Applications**, v. 26, n. 1, p. 200–215, fev. 2021.

ANDERSON-BELL, J.; SCHILLACI, C.; LIPANI, A. Predicting non-residential building fire risk using geospatial information and convolutional neural networks. **Remote Sensing Applications Society and Environment**, v. 21, p. 100470, Jan 2021.

AREL, I.; ROSE, D. C.; KARNOWSKI, T. P. Deep Machine Learning: A new frontier in artificial intelligence research. **IEEE Computational Intelligence Magazine**, v. 5, p. 13–18, nov 2010.

ASIMOV, I. **I, Robot**. Dennis Dobson, 1950. (Panther science fiction). ISBN 9780451012821. Disponível em: https://books.google.com.br/books?id=MD0GAQAAIAAJ.

BADRINARAYANAN, V.; KENDALL, A.; CIPOLLA, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 39, n. 12, p. 2481–2495, 2017.

BOARDMAN, J. The value of google earth™ for erosion mapping. **CATENA**, v. 143, p. 123–127, 2016. ISSN 0341-8162. Disponível em: https://www.sciencedirect.com/science/article/pii/S0341816216301187.

BRASIL, M. D. M. A. **APA da Região Serrana de Petrópolis**. 2023. Acesso em: 24 fev 2023. Disponível em: https://www.gov.br/icmbio/pt-br/assuntos/biodiversidade/unidade-de-conservacao/unidades-de-biomas/mata-atlantica/lista-de-ucs/apa-da-regiao-serrana-de-petropolis.

BRESSAN, P. O. *et al.* Semantic segmentation with labeling uncertainty and class imbalance applied to vegetation mapping. **International Journal of Applied Earth Observation and Geoinformation**, v. 108, p. 102690, 2022. ISSN 1569-8432. Disponível em: https://www.sciencedirect.com/science/article/pii/S0303243422000162.

BRITO, A. S. *et al.* Combining max-pooling and wavelet pooling strategies for semantic image segmentation. **Expert Systems with Applications**, v. 183, p. 115403, 2021. ISSN 0957-4174. Disponível em: https://www.sciencedirect.com/science/article/pii/S0957417421008253.

CARRANZA-ROJAS, J. *et al.* Going deeper in the automatedidentification of herbarium specimens. **BMC Ecology and Evolution**, v. 17, p. 181, 2017.

CHEN, W. *et al.* Mapping agricultural plastic greenhouses using google earth images and deep learning. **Computers and Electronics in Agriculture**, v. 191, p. 106552, 2021. ISSN 0168-1699. Disponível em: https://www.sciencedirect.com/science/article/pii/S016816992100569X.

CORREIA, F. P.; LEAO, S. A. da S. GestÃo participativa de unidades de conservaÇÃo da natureza: reflexões a partir da Área de proteção ambiental da região de maracanã, são luís-ma, brasil. *In*: UFMA, U. F. do M. (Ed.). **VII JORNADA INTERNACIONAL POLÍTICAS PÚBLICAS**. São Luiz, Maranhão, Brasil: [*s.n.*], 2015.

CROWDER, D. A. **Google Earth for Dummies**. 1. ed. [*S.l.*]: For Dummies, 2007. 360 p.

CVAT.AI. **opencv/cvat: v2.4.0**. Zenodo, 2023. Disponível em: https://doi.org/10.5281/zenodo.7739965.

DENG, J. *et al.* Imagenet: A large-scale hierarchical image database. *In*: **2009 IEEE Conference on Computer Vision and Pattern Recognition**. [*S.l.*: *s.n.*], 2009. p. 248–255.

DUHL, T. R.; GUENTHER, A.; HELMIG, D. Estimating urban vegetation cover fraction using google earth® images. **Journal of Land Use Science**, Taylor Francis, v. 7, n. 3, p. 311–329, 2012.

ECO, R. **O que é uma Área de Proteção Ambiental**. 2015. Acesso em: 24 fev 2023. Disponível em: https://oeco.org.br/dicionario-ambiental/29203-o-que-e-uma-area-de-protecao-ambiental/.

ECOTEMA, I. **Zoneamento Ambiental da APA Petrópolis**. 2003. Convênio FNMA/IBAMA. Rio de Janeiro.

ESRI INC. **ArcGIS Pro**. 2023. Acesso em: 24 mar 2023. Disponível em: https://www.esri.com/pt-br/arcgis/products/arcgis-pro/overview.

ESTEVA, A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. **Nature**, v. 542, p. 115–118, 2017.

FARHANGFAR, S.; REZAEIAN, M. Semantic segmentation of aerial images using fcn-based network. *In*: **2019 27th Iranian Conference on Electrical Engineering (ICEE)**. [*S.l.*: *s.n.*], 2019. p. 1864–1868.

FREUDENBERG, M. U. **Tree Detection in Remote Sensing Imagery Baumerkennung in Fernerkundungs-Bildmaterial**. mar. 2019. 85 p. Dissertação (Mestrado) — Burckhardt Institute, Gottingen, Alemanha, mar. 2019.

GERLAK, A. K. Policy interactions in human-landscape systems. environmental management. **Environmental Management**, v. 53, p. 67–75, 2014.

GHOLAMALINEZHAD, H.; KHOSRAVI, H. Pooling methods in deep neural networks, a review. **Computer Vision and Pattern Recognition**, p. 144–157, Set 2020.

GONZALEZ, R. C.; WOODS, R. E. **Processamento Digital de Imagens**. 3. ed. São Paulo, SP: Pearson, 2010. 644 p.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge, MA, USA: MIT Press, 2016. http://www.deeplearningbook.org.

GOOGLE. **Google Maps/Google Earth**. 2023. Disponível em: https://www.google.com/permissions/geoguidelines/. Acesso em: 28 fev. 2023.

GRANDINI, M.; BAGLI, E.; VISANI, G. **Metrics for Multi-Class Classification: an Overview**. 2020.

HAFEMANN, L. G. **An analysis of deep neural networks for texture classification**. Nov 2014. 76 p. Dissertação (Mestrado) — Programa de Pós-Graduação em Informática, Setor de Ciências Exatas, Universidade Federal do Paraná, Curitiba, Nov 2014.

HAO, S.; ZHOU, Y.; GUO, Y. A brief survey on semantic segmentation with deep learning. **Neurocomputing**, v. 406, p. 302–321, 2020. ISSN 0925-2312.

HAYKIN, S. **Redes Neurais**: Princípios e prática. 2. ed. [*S.l.*]: Bookman, 2000. 898 p.

HELERBROCK, R. **Espectro Eletromagnético**: O que é, usos, cores, frequências. 2023. BRASIL ESCOLA. Disponível em: https://brasilescola.uol.com.br/fisica/espectro-eletromagnetico.htm. Acesso em: 24 ago. 2023.

HERTZ, J. A.; KROGH, A. S.; PALMER, R. G. **Introduction to the Theory of Neural Computacion**. 1. ed. [*S.l.*]: Westview Press, 1991. 327 p.

HU, F. *et al.* Recent advances and opportunities in scene classification of aerial images with deep models. **Computer Vision and Pattern Recognition**, p. 211–252, Jun 2018.

HU, Q. *et al.* Exploring the use of google earth imagery and object-based methods in land use/cover mapping. **Remote Sensing**, v. 5, n. 11, p. 6026–6042, 2013. ISSN 2072-4292. Disponível em: https://www.mdpi.com/2072-4292/5/11/6026.

HUTH, J. *et al.* Land cover and land use classification with twopac: towards automated processing for pixel- and object-based image classification. **Remote Sensing**, v. 4, n. 9, p. 2530–2553, 2012. ISSN 2072-4292. Disponível em: https://www.mdpi.com/2072-4292/4/9/2530.

IBM. **Redes Neurais**. 2020. IBM Cloud Education. Disponível em: https://www.ibm.com/br-pt/cloud/learn/neural-networks#toc-tipos-de-r-DbL0dXJo. Acesso em: 11 fev. 2022.

INC, A. **Anaconda Documentation: Release 2.0**. 2023. Acesso em: 24 mar 2023. Disponível em: https://readthedocs.com/projects/continuumio-docs/downloads/pdf/latest.

ISLAM, M. R. *et al.* Deep learning-based glaucoma detection with cropped optic cup and disc and blood vessel segmentation. **IEEE Access**, v. 10, p. 2828–2841, 2022.

KARN, U. **An Intuitive Explanation of Convolutional Neural Networks**. 2016. Acesso em: 21 jan. 2022. Disponível em: https://ujjwalkarn.me/2016/08/11/intuitive-explanation-convnets.

KHAN, S. *et al.* **A Guide to Convolutional Neural Networks for Computer Vision**. 1. ed. London, UK: Morgan Claypool, 2018. 207 p.

KUMAR, D. A. *et al.* Detecting diseased images by segmentation and classification based on semi-supervised learning. *In*: **2012 12th International Conference on Hybrid Intelligent Systems (HIS)**. [*S.l.*: *s.n.*], 2012. p. 549–554.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature Publishing Group**, v. 521, n. 7553, p. 436, 2015.

LECUN, Y.; CORTES, C.; BURGES, C. **The MNIST database of handwritten digits**. 2021. Acesso em: 17 maio 2021. Disponível em: http://yann.lecun.com/exdb/mnist.

LEE, H. *et al.* Fully automated deep learning system for bone age assessment. **Journal of digital imaging**, v. 30, n. 4, p. 427–441, Mar 2017.

LI, W. *et al.* Integrating google earth imagery with landsat data to improve 30-m resolution land cover mapping. **Remote Sensing of Environment**, v. 237, p. 111563, 2020. Disponível em: https://www.sciencedirect.com/science/article/pii/S0034425719305838.

LI, X. *et al.* Dual cross-entropy loss for small-sample fine-grained vehicle classification. **IEEE Transactions on Vehicular Technology**, v. 68, n. 5, p. 4204–4212, 2019.

LIMA, R. P. de *et al.* Deep convolutional neural networks as a geological image classification tool. **The Sedimentary Record**, v. 17, p. 4–9, 2019.

LIMA, R. P. de; MARFURT, K. Convolutional neural network for remote-sensing scene classification: Transfer learning analysis. **Remote Sensing**, v. 12, n. 1, p. 86, 2020.

LIN, T.-Y. *et al.* **Focal Loss for Dense Object Detection**. 2017.

LIU, Y. Introduction to land use and rural sustainability in china. **Land Use Policy**, v. 74, p. 1–4, 2018. ISSN 0264-8377. Land use and rural sustainability in China. Disponível em: https://www.sciencedirect.com/science/article/pii/S0264837717315636.

LIU, Y. *et al.* Hourglass-shapenetwork based semantic segmentation for high resolution aerial imagery. **Remote Sensing**, v. 9, n. 6, 2017. ISSN 2072-4292. Disponível em: https://www.mdpi.com/2072-4292/9/6/522.

LORENZZETTI, J. A. **Princípios físicos de sensoriamento remoto**. 1. ed. São Paulo, SP: Blucher, 2015. 292 p. ISBN 978-85-2120-835-8.

LUMNITZ, S. *et al.* Mapping trees along urban street networks with deep learning and street-level imagery author links open overlay panel. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 175, p. 144–157, 2021.

MENESES, P. R. *et al.* **Introdução ao processamento de imagens digitais de satélites de sensoriamento remoto**. 1. ed. Brasília, DF: Universidade de Brasília, 2012. 276 p.

NGUYEN-KHANH, L.; NGUYEN-NGOC-YEN, V.; DINH-QUOC, H. U-net semantic segmentation of digital maps using google satellite images. *In*: **2021 8th NAFOSTED Conference on Information and Computer Science (NICS)**. [*S.l.*: *s.n.*], 2021. p. 386–391.

NOVO, E. L. M.; PONZONI, F. J. **INTRODUÇÃO AO SENSORIAMENTO REMOTO**. São José dos Campos, SP: Instituto Nacional de Pesquisas Espaciais, 2001. 68 p.

NOVO, E. M. L. d. M. **Sensoriamento Remoto**: Princípios e aplicações. 4. ed. São Paulo, SP: Edgard Blücher Ltda, 2010. 387 p. ISBN 9788521205401.

NVIDIA Corporation. **CUDA Toolkit Documentation 12.1**. 2023. Acesso em: 24 mar 2023. Disponível em: https://docs.nvidia.com/cuda/.

OCHOA, K. S.; GUO, Z. A framework for the management of agricultural resources with automated aerial imagery detection. **Comput Electron Agric**, v. 162, p. 53–69, 2019.

ONISHI, M.; ISE, T. Explainable identification and mapping of trees using uav rgb image and deep learning. **Scientific Reports**, v. 11, p. 903, Jan 2021.

PASZKE, A. *et al.* Automatic differentiation in PyTorch. *In*: **NIPS Autodiff Workshop**. [*S.l.*: *s.n.*], 2017.

PEDRINI, H.; SCHWARTZ, W. **Análise de Imagens Digitais**: Princípios, algoritmos e aplicações. 1. ed. São Paulo, SP: Cengage Learning, 2008. 528 p.

PETERSON, J. M. *et al.* Economic linkages to changing landscapes. **Environmental Management**, v. 53, p. 55–56, 2014.

PETRóPOLIS, P. M. d. S. d. M. A. **APA da Região Serrana de Petrópolis**. 2006. Acesso em: 23 mar 2023. Disponível em: https://www.petropolis.rj.gov.br/sma/index.php/downloads/category/4-mapas.html.

POWERS, D. M. W. **Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation**. 2020.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. **CoRR**, abs/1505.04597, 2015. Disponível em: http://arxiv.org/abs/1505.04597.

ROSSUM, G. V. **Python 3 Reference Manual**. Scotts Valley, CA: CreateSpace, 2009. ISBN 978-1441412690.

RUDER, S. An overview of gradient descent optimization algorithms. **ArXiv**, abs/1609.04747, 2016. Disponível em: https://api.semanticscholar.org/CorpusID:17485266.

SARKAR, D. **A Comprehensive Hands-on Guide to Transfer Learning with Real-World Applications in Deep Learning**. 2018. Towards Data Science. Disponível em: https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3 Acesso em: 25 jan. 2022.

SHARMA, N.; SHARMA, R.; JINDAL, N. Machine learning and deep learning applications: A vision. *In*: PROCEEDINGS, G. T. (Ed.). Bangalor, IN: 1st International Conference on Advances in Information, Computing and Trends in Data Engineering, 2021. v. 2, p. 24–28.

SHAWKY, O. A. *et al.* Remote sensing image scene classification using cnn-mlp with data augmentation. **Optik**, v. 221, p. 165356, Jun 2020.

SHI, F.; YANG, B.; LI, M. An improved framework for assessing the impact of different urban development strategies on land cover and ecological quality changes -a case study from nanjing jiangbei new area, china. **Ecological Indicators**, v. 147, p. 109998, 2023. ISSN 1470-160X. Disponível em: https://www.sciencedirect.com/science/article/pii/S1470160X23001401.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *In*: (ICLR), I. C. O. L. R. (Ed.). **3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings**. San Diego, CA, USA: [*s.n.*], 2015. p. 1–9.

SOBEL, I.; FELDMAN, G. A 3 × 3 isotropic gradient operator for image processing. a talk at the stanford artificial project. **Pattern Classification and Scene Analysis**, p. 271–272, 1968.

TAN, M.; LE, Q. V. Efficientnet: Rethinking model scaling for convolutional neural networks. **CoRR**, abs/1905.11946, 2019. Disponível em: http://arxiv.org/abs/1905.11946.

TORRES-SáNCHEZ, J. *et al.* Configuration and specifications of an unmanned aerial vehicle (uav) for early site specific weed management. **PLOS ONE**, Public Library of Science, v. 8, n. 3, p. 1–15, 03 2013. Disponível em: https://doi.org/10.1371/journal.pone.0058210.

WANG, B. *et al.* Improved u-net fundus image segmentation algorithm integrating effective channel attention. **jist**, v. 66, p. 040408–1–040408–8, 2022.

WATANABE, S.; SUMI, K.; ISE, T. Identifying the vegetation type in google earth images using a convolutional neural network: a case study for japanese bamboo forests. **BMC Ecology**, v. 20, n. 1, p. 65, nov. 2020.

WEINSTEIN, B. *et al.* Individual tree-crown detection in rgb imagery using semi-supervised deep learning neural networks. **Remote Sens.**, v. 11, p. 1309, 2019.

XING, J. *et al.* Building extraction from google earth images. *In*: **Proceedings of the 21st International Conference on Information Integration and Web-Based Applications amp; Services**. New York, NY, USA: Association for Computing Machinery, 2020. p. 502—-511. ISBN 9781450371797. Disponível em: https://doi.org/10.1145/3366030.3368456.

YANG, L. *et al.* Towards synoptic water monitoring systems: a review of ai methods for automating water body detection and water quality monitoring using remote sensing. **Sensors**, v. 22, p. 2416, 2022.

ZIN, T. T.; LIN, J. C.-W. Big data analysis and deep learning applications. *In*: **Proceedings of the First International Conference on Big Data Analysis and Deep Learning**. Berlin, Germany: [*s.n.*], 2018. p. 3027–3031.