Avaliação Comparativa de Modelos de Segmentação Semântica de Nuvens em Imagens de Satélite

Gregory Filipe Lira da Silva

CENTRO DE INFORMÁTICA
UNIVERSIDADE FEDERAL DA PARAÍBA

João Pessoa, PB

2024

Gregory Filipe Lira da Silva

Avaliação Comparativa de Modelos de Segmentação Semântica de Nuvens em Imagens de Satélite

Monografia apresentada ao curso
Ciência de Dados e Inteligência Artificial
do Centro de Informática, da Universidade Federal da Paraíba,
como requisito para a obtenção do grau de bacharel

Orientador: Leonardo Vidal Batista

Maio de 2024

Catalogação na publicação
Seção de Catalogação e Classificação

S586a Silva, Gregory Filipe Lira da.
Avaliação comparativa de modelos de segmentação semântica de nuvens em imagens de satélite / Gregory Filipe Lira da Silva. - João Pessoa, 2024.
51 f. : il.

Orientação: Leonardo Batista.
TCC (Graduação) - UFPB/CI.

1. Segmentação semântica. 2. Sensoriamento remoto. 3. Aprendizado profundo. 4. 38-cloud. 5. Avaliação comparativa. I. Batista, Leonardo. II. Título.

UFPB/CI CDU 004.8

Elaborado por Michelle de Kássia Fonseca Barbosa - CRB-738

CENTRO DE INFORMÁTICA
UNIVERSIDADE FEDERAL DA PARAÍBA

Trabalho de Conclusão de Curso de Ciência de Dados e Inteligencia Artificialintitulado ***Avaliação Comparativa de Modelos de Segmentação Semântica de Nuvens em Imagens de Satélite*** de autoria de Gregory Filipe Lira da Silva, aprovada pela banca examinadora constituída pelos seguintes professores:

Prof. Dr. Celso Augusto Guimarães Santos
Universidade Federal da Paraíba

Prof. Dr. Richarde Marques da Silva
Universidade Federal da Paraíba

João Pessoa, 21 de maio de 2024

Centro de Informática, Universidade Federal da Paraíba
Rua dos Escoteiros, Mangabeira VII, João Pessoa, Paraíba, Brasil CEP: 58058-600
Fone: +55 (83) 3216 7093 / Fax: +55 (83) 3216 7117

"A única maneira de fazer um ótimo trabalho é amar o que você faz.- Steve Jobs

AGRADECIMENTOS

Gostaria de expressar meus sinceros agradecimentos a todas as pessoas que contribuíram para a realização desta Monografia. Em primeiro lugar, sou imensamente grato ao meu orientador, Leonardo Vidal Batista, pela orientação, paciência e apoio ao longo de todo o processo.

Também gostaria de agradecer à minha família pelo amor incondicional, incentivo e compreensão durante os momentos dedicados a este projeto. Suas palavras de encorajamento foram um impulso essencial para superar os desafios encontrados.

Agradeço igualmente aos meus amigos e colegas que estiveram ao meu lado, compartilhando ideias, oferecendo suporte e compreensão nos momentos de pressão e dúvida.

RESUMO

A presença de nuvens em imagens de satélite apresenta desafios significativos em diversas áreas, incluindo meteorologia, agricultura e monitoramento ambiental, devido à necessidade de distinguir as nuvens de outros elementos presentes nas imagens. O avanço tecnológico no sensoriamento remoto e a ampla disponibilidade de imagens de satélite têm proporcionado percepções valiosas sobre o ambiente terrestre. A detecção precisa e a compreensão subsequente da presença de nuvens em imagens de satélite são cruciais não apenas para a pesquisa científica, mas também para aplicações práticas que dependem de análises precisas do ambiente. O objetivo principal deste trabalho é desenvolver uma análise comparativa entre modelos baseados em limiar e modelos baseados em aprendizado profundo para a segmentação semântica de nuvens em imagens de satélite, com o intuito de identificar a abordagem mais eficaz para mapear e compreender a distribuição espacial das nuvens para ajudar por exemplo no uso de técnicas para remoção. As nuvens podem obstruir objetos na superfície terrestre e causar dificuldades em diversas aplicações de sensoriamento remoto. A escolha deste trabalho foi explorar tanto uma abordagem baseada em limiar, utilizando um modelo onde são aplicados vários limiares em múltiplas bandas das imagens, dependendo do satélite selecionado, quanto modelos baseados em aprendizagem profunda. Foram selecionados e analisados quatro modelos de aprendizado profundo - U-Net, FPN-Net e YOLO - os quais foram treinados sem os pesos, além de um modelo como o Cloud-net com seus pesos pré-treinados, e um modelo baseado em limiar conhecido como Fmask. Além disso, é importante mencionar que para o treinamento dos modelos foi utilizada a base de dados 38-Cloud. Devido à alta quantidade de imagens que possuíam pixels nulos, foram propostas algumas modificações, incluindo a remoção destas imagens e a aplicação de técnicas de aumento de dados. Outra etapa do treinamento incluiu a utilização de parada precoce. Foram realizados dois experimentos neste trabalho: um deles envolvendo os modelos treinados e avaliados utilizando dois conjuntos de dados públicos: 38-Cloud e 95-Cloud, empregando métricas como Interseção sobre União (IoU), precisão, revocação e acurácia. O outro experimento envolveu o uso das imagens de teste do 38-Cloud, que foram utilizadas nos testes do Cloud-net. Após a aplicação das imagens no modelo, todas as imagens que foram cortadas da imagem original foram recolocadas no lugar, e as métricas foram aplicadas na máscara gerada. Os resultados indicam que os modelos de rede neural profunda superaram as técnicas baseadas em limiar, destacando o desempenho notável do Cloud-Net, que obteve um IoU de 78.50% e uma acurácia de 96.48%. Esta análise ressalta a importância fundamental do emprego de modelos de aprendizado profundo na segmentação de nuvens em imagens de satélite, fornecendo compreensões valiosas e aplicáveis em uma ampla gama de áreas de interesse.

Palavras-chave: <Segmentação Semântica>, <Sensoriamento Remoto>, <Aprendizado Profundo>, <38-cloud>, <Avaliação Comparativa>.

ABSTRACT

The presence of clouds in satellite images poses significant challenges in various fields, including meteorology, agriculture, and environmental monitoring, due to the need to distinguish clouds from other elements present in the images. Technological advancements in remote sensing and the widespread availability of satellite images have provided valuable insights into the terrestrial environment. Accurate detection and subsequent understanding of cloud presence in satellite images are crucial not only for scientific research but also for practical applications that rely on precise environmental analyses. The main objective of this work is to develop a comparative analysis between threshold-based models and deep learning-based models for semantic segmentation of clouds in satellite images, aiming to identify the most effective approach to map and understand the spatial distribution of clouds to aid, for example, in the use of techniques for removal. Clouds can obscure objects on the Earth's surface and cause difficulties in various remote sensing applications. The choice of this work was to explore both a threshold-based approach, using a model where multiple thresholds are applied in multiple bands of the images, depending on the selected satellite, and deep learning-based models. Four deep learning models - U-Net, FPN-Net, and YOLO - were selected and analyzed, which were trained without weights, along with a model like Cloud-net with its pretrained weights, and a threshold-based model known as Fmask. Furthermore, it is important to mention that the 38-Cloud database was used for training the models. Due to the high number of images with null pixels, some modifications were proposed, including the removal of these images and the application of data augmentation techniques. Another training step included the use of early stopping. Two experiments were conducted in this work: one involving the trained and evaluated models using two public datasets: 38-Cloud and 95-Cloud, employing metrics such as Intersection over Union (IoU), precision, recall, and accuracy. The other experiment involved the use of test images from 38-Cloud, which were used in the Cloud-net tests. After applying the images to the model, all images that were cropped from the original image were placed back, and the metrics were applied to the generated mask. The results indicate that deep neural network models outperformed threshold-based techniques, highlighting the remarkable performance of Cloud-Net, which achieved an IoU of 78.50% and an accuracy of 96.48%. This analysis underscores the fundamental importance of employing deep learning models in cloud segmentation in satellite images, providing valuable and applicable insights across a wide range of areas of interest.

Key-words: <Semantic Segmentation>, <Remote Sense>, <Deep Learning>, <38-cloud>, <Comparative Evaluation>

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS

PDI – Processamento Digital de Imagem

FPN – Feature Pyramid Network

YOLO - You only look once

Fmask - Function Mask

BoW - Bag of Words

SVM - Máquinas de Vetores de Suporte

SUMÁRIO

1 INTRODUÇÃO

A análise minuciosa de imagens de satélite desempenha um papel crucial em diversas áreas de pesquisa, desde visão computacional até meteorologia. Nesse contexto, a segmentação semântica de nuvens em imagens de satélite representa uma tarefa desafiadora e de grande importância. Este trabalho propõe uma avaliação comparativa de modelos de segmentação semântica de nuvens, com foco na otimização dos resultados dos modelos selecionados e na aplicabilidade prática dessas técnicas.

O avanço tecnológico no sensoriamento remoto e a ampla disponibilidade de imagens de satélite têm proporcionado percepções valiosas sobre o ambiente terrestre. Além disso, a análise de nuvens também contribui significativamente em diversas áreas de pesquisa, estendendo-se para além da superfície terrestre. Por exemplo, é fundamental para entender melhor os padrões climáticos globais, estudar os efeitos das mudanças climáticas, monitorar a cobertura de nuvens em escalas regionais e globais, entre outros aspectos. Essa compreensão mais profunda das nuvens não apenas amplia nosso conhecimento sobre a atmosfera terrestre, mas também é essencial para o desenvolvimento de estratégias mais eficazes em áreas como agricultura, gestão de recursos naturais e prevenção de desastres naturais.

A detecção precisa e a compreensão subsequente da presença de nuvens em imagens de satélite são cruciais não apenas para a pesquisa científica, mas também para aplicações práticas que dependem de análises precisas do ambiente. Assim, este estudo não se limita apenas à análise teórica, mas busca oferecer uma abordagem prática e aplicada na segmentação de nuvens, fornecendo modelos leves com bom resultado ate modelos com maior numero de parâmetros com o melhor resultado.

1.1 Objetivo geral

O objetivo principal deste trabalho é desenvolver uma analise comparativa entre modelos tradicionais e modelos baseados em aprendizado profundo para a segmentação semântica de nuvens em imagens de satélite, com o intuito de promover avanços práticos em sensoriamento remoto e análise de dados. Além de proporcionar soluções concretas para os desafios decorrentes da presença de nuvens, este estudo busca contribuir para o conhecimento acadêmico ao apresentar uma abordagem diferenciada e eficaz. Espera-se que os resultados obtidos possam melhorar significativamente a aplicabilidade das técnicas de processamento de imagens de satélite em diversas áreas.

1.2 Objetivos específicos

A seguir, são apresentados os objetivos específicos para alcançar a resolução do trabalho:

1. Implementar os modelos selecionados para a segmentação, incluindo U-Net, YOLO, Cloud-Net, FPN-Net e Fmask.
2. Adquirir conjuntos de dados adequados para a segmentação de nuvens, tais como 38-Cloud, 95-Cloud.
3. Preparar o conjunto de treinamento e aplicar técnicas de aumento de dados, quando viável.
4. Treinar os modelos com os conjuntos de dados preparados.
5. Avaliar os modelos e técnicas utilizando métricas comparativas.

1.3 Estrutura da monografia

Este documento está estruturado em cinco seções principais: introdução, revisão de conceitos e literatura, metodologia, apresentação e análise dos dados obtidos, e conclusões e perspectivas para pesquisas futuras. Na seção inicial, apresenta a problemática investigada, o objetivo geral e os objetivos específicos da pesquisa. A revisão de conceitos e literatura abrange os fundamentos teóricos essenciais para a compreensão do estudo. Na seção de metodologia, detalha os procedimentos necessários para a realização dos experimentos. Em seguida, a seção de apresentação e análise dos dados apresenta os resultados obtidos conforme descrito nos métodos. Finalmente, concluí-se com a seção de conclusões e perspectivas futuras, na qual discute as conclusões derivadas dos resultados obtidos e sugere direções para investigações subsequentes.

2 CONCEITOS GERAIS E REVISÃO DA LITERATURA

Este capítulo é estruturado em cinco subseções principais e cada uma delas aborda um aspecto que proporciona uma visão detalhada das teorias, modelos e métricas utilizadas, fornecendo o necessário para o entendimento e compreensão do trabalho. Nesta seção, serão abordados os tipos de segmentação, as técnicas utilizadas, as métricas empregadas na avaliação e os trabalhos relacionados.

2.1 Segmentação Semântica e de Nuvens

A segmentação é definida como o processo pelo qual um rótulo de classe é atribuído a cada pixel de uma imagem, exigindo previsões de alta precisão em nível de pixel (THISANKE et al., 2023). Modelos de aprendizagem dedicados à segmentação de imagens permitem que as máquinas interpretem informações visuais de maneira semelhante ao cérebro humano. Embora os modelos de segmentação de imagem compartilhem algumas aplicações com os modelos de detecção de objetos, eles se distinguem por uma característica essencial: identificam diferentes entidades presentes em uma imagem no nível do pixel, ao invés de delinear essas entidades com uma caixa delimitadora (IBM, 2024).

A segmentação semântica é uma forma avançada de segmentação, na qual cada pixel em uma imagem é categorizado de acordo com o objeto ao qual pertence. Diferentemente da segmentação tradicional, que atribui um rótulo de classe a cada pixel, a segmentação semântica atribui rótulos baseados no significado dos objetos na imagem. Modelos de aprendizagem dedicados à segmentação semântica permitem que as máquinas compreendam informações visuais de maneira semelhante ao cérebro humano.

Para alcançar esse objetivo, os modelos de segmentação semântica empregam redes neurais complexas, capazes de agrupar pixels relacionados em máscaras de segmentação de forma precisa e reconhecer corretamente a classe semântica real de cada agrupamento de pixels, também referido como segmento (IBM, 2024). A Figura 1 ilustra visualmente o processo de segmentação semântica, conforme apresentado por LI (2024).

O principal desafio do sensoriamento remoto reside na oclusão de informações cruciais sobre a superfície terrestre, o que afeta negativamente a análise de fenômenos tais como uso do solo, cobertura vegetal e recursos hídricos, conforme evidenciado nas referências Shi et al. (2017) e Oliveira (2020).

A alta qualidade das imagens e a precisão na criação de máscaras são essenciais para o treinamento de modelos eficazes. Imagens com elevada presença de nuvens ou erros humanos durante a produção das imagens rotuladas podem induzir falhas no algoritmo, as quais, por sua vez, seriam transmitidas para a classificação das imagens projetadas (OLIVEIRA, 2020).

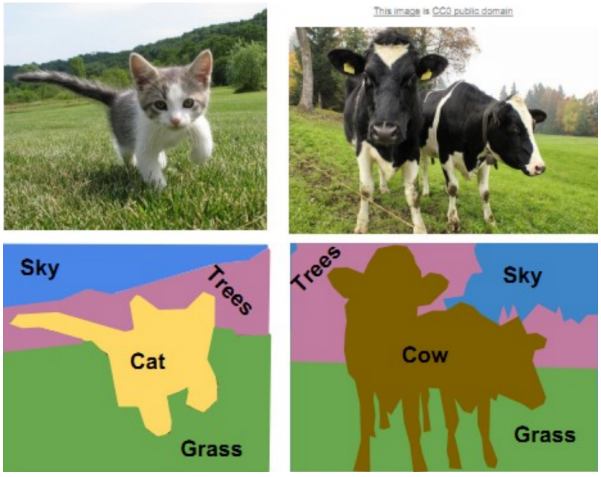

Figura 1: Segmentação semântica LI (2024)

2.2 Segmentação Semântica de Nuvens

A identificação precisa e a medição da cobertura de nuvens são etapas fundamentais na análise de imagens de satélite. As nuvens podem obstruir objetos na superfície terrestre e causar dificuldades em diversas aplicações de sensoriamento remoto, incluindo detecção de mudanças, estimativa de parâmetros geofísicos e rastreamento de objetos (MOHAJERANI; SAEEDI, 2019; JIN et al., 2008; TAYEBI et al., 2023).

Por outro lado presença extensiva de nuvens pode fornecer informações valiosas sobre parâmetros climáticos e desastres naturais, como furacões e erupções vulcânicas, ressaltando a importância da análise de imagens de satélite com alta cobertura de nuvens (TAYEBI et al., 2023; ICHIPRO, 2024; REDDY et al., 2017).

Os pesquisadores têm se dedicado à busca da melhor abordagem para detecção e segmentação de nuvens. De acordo com Mohajerani e Saeedi (2019), existem três categorias principais para este propósito. A primeira, baseada em limiar, utiliza as bandas de satélite específicas e emprega técnicas de aplicação de limiar em diferentes bandas da imagem para segmentar as nuvens, conforme destacado por Irish et al. (2006), Zhu e Woodcock (2012), Zhu, Wang e Woodcock (2015), Qiu, Zhu e He (2019), como ilustrado nas Figuras 2 e 3 os fluxogramas que estes modelos de segmentação por limiar utiliza.

Modelos baseados em contexto, conforme discutidos nos trabalhos de Yuan e Hu (2015) e Zhang, Guindon e Cihlar (2002), são desenvolvidos levando em consideração as particularidades do ambiente e das situações dos dados coletados. Um exemplo desse enfoque é o modelo de *Bag of Words (BoW)*, introduzido por Yuan e Hu (2015), que utiliza essa abordagem para extrair características dos segmentos de imagem, empregando, em seguida, Máquinas de Vetores de Suporte (SVM) para distinguir entre regiões de nuvem e não nuvem.

Os modelos baseados em redes neurais profundas têm se destacado na área de

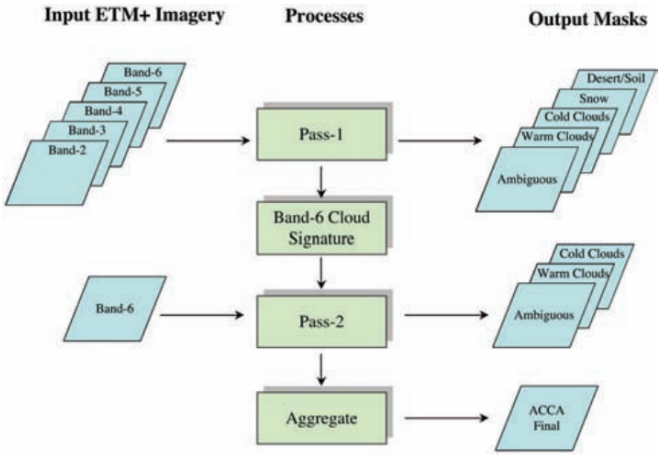

Figura 2: Visão geral da estimativa de cobertura de nuvem automatizada utilizando limiar. por (IRISH et al., 2006), onde recebe os inputs e no processos vai aplicando os limiares.

processamento de imagens (MOHAJERANI; SAEEDI, 2019; XIE et al., 2017; MOHAJERANI; KRAMMER; SAEEDI, 2018a). Nestes modelos, são empregados filtros convolucionais para detectar padrões específicos em uma imagem, como bordas, texturas ou características distintas. Cada filtro consiste em uma matriz de pesos que determina como a convolução é realizada. Durante esse processo, o filtro é aplicado a cada região da imagem e é realizada a multiplicação elemento a elemento entre os valores dos pixels da região da imagem e os valores correspondentes do filtro. Isso resulta em uma saída com um canal de cor em tons de cinza, conforme ilustrado na Figura 4.

Um exemplo proeminente desses modelos é a U-Net, que comprime a imagem em um vetor de características e, em seguida, a expande para realizar a segmentação em escala de cinza. Uma representação desse processo é apresentada por Mohajerani, Krammer e Saeedi (2018a), como mostrado na Figura 5.

A Figura 5 mostra a rede proposta para detectar nuvens. A profundidade do mapa de características no caminho de codificação é aumentada de 4 (os canais de entrada são RGBNir) para 1024. Essa profundidade é então reduzida de 1024 para 1 (mapa de probabilidade em escala de cinza) no caminho de decodificação. Enquanto isso, o tamanho espacial do mapa de características é reduzido de 192 $\times$ 192 para 6 $\times$ 6 no caminho de codificação e, em seguida, é aumentado para 192 $\times$ 192 no caminho de decodificação. A camada de cópia entre o bloco de codificação i e o bloco de decodificação j concatena a saída da segunda camada de convolução no bloco de codificação i com a saída da camada convolucional transposta no bloco de decodificação j.

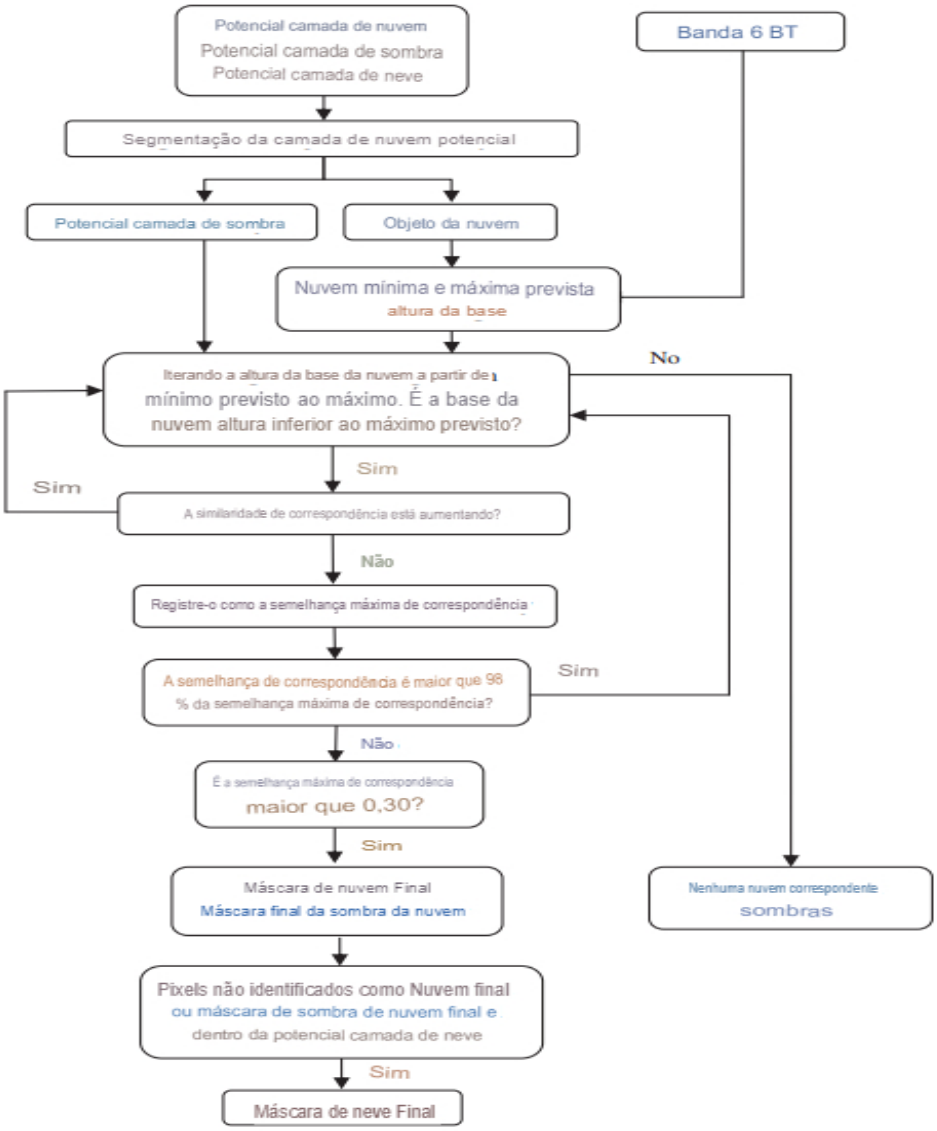

Figura 3: Detecção de nuvem e sombra de nuvem baseada em objetos em imagens Landsat, Fmask por Zhu e Woodcock (2012)

3 METODOLOGIA

Neste capítulo, serão apresentado a metodologia empregada na condução dos experimentos. Para a análise de dados, é utilizado a base de dados 38-Cloud, inicialmente introduzida por Mohajerani, Krammer e Saeedi (2018b). É utilizados modelos de segmentação semântica, como U-net, Cloud-net, FPN, e um que utiliza detecção e segmentação juntas, chamado Yolo, além da Fmask, técnica de segmentação baseadas em limiar, também é apresentado os critérios de avaliação.

Neste contexto, descreve como as etapas do experimento foram executadas, desde a seleção das bases de dados até o treinamento e comparação dos modelos e técnicas.

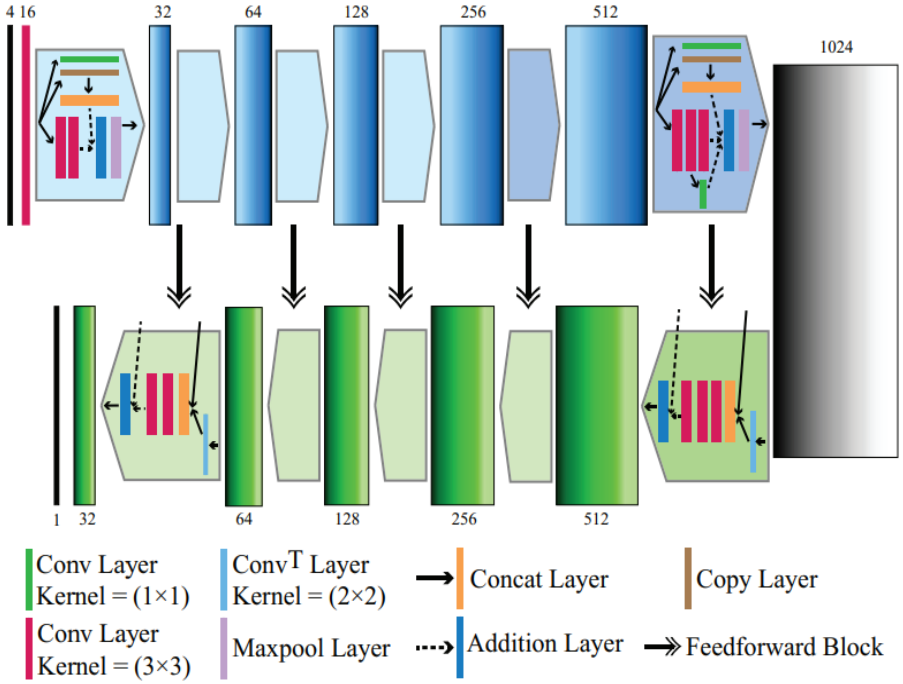

Figura 4: Arquitetura do modelo implementado com Aprendizado Profundo utilizado na CloudNet por Mohajerani e Saeedi (2019)

3.1 Coleta das Bases de Dados

A seleção das bases de dados para o treinamento dos modelos foi embasada em artigos relacionados à segmentação de nuvens. Optou-se por dois conjuntos de dados para este estudo: o 38-Cloud (MOHAJERANI, 2019a) e sua extensão, o 95-Cloud (MOHAJERANI, 2019b). Para conduzir os testes planejados, foram utilizadas imagens do conjunto de dados 95-Cloud.

Nas subseções a seguir, serão apresentadas informações detalhadas sobre cada uma dessas bases de dados, incluindo características, origem e qualquer relevância específica para o escopo deste trabalho.

3.1.1 38-Cloud

O conjunto de dados 38-Cloud consiste em 38 imagens de cena Landsat 8 e suas respectivas segmentações manuais de nuvem em nível de pixel. Este conjunto de dados foi introduzido originalmente por Mohajerani e Saeedi (2019), sendo uma modificação de um conjunto de dados anterior (MOHAJERANI; KRAMMER; SAEEDI, 2018c). As imagens inteiras dessas cenas foram recortadas em sub imagens de 384x384 denominados de *patches*, para serem adequados para algoritmos de segmentação semântica baseados em aprendizagem profunda. Há 8400 *patches* para treinamento e 9201 *patches* para teste. Cada *patches* possui 4 canais espectrais correspondentes: Vermelho (banda 4), Verde

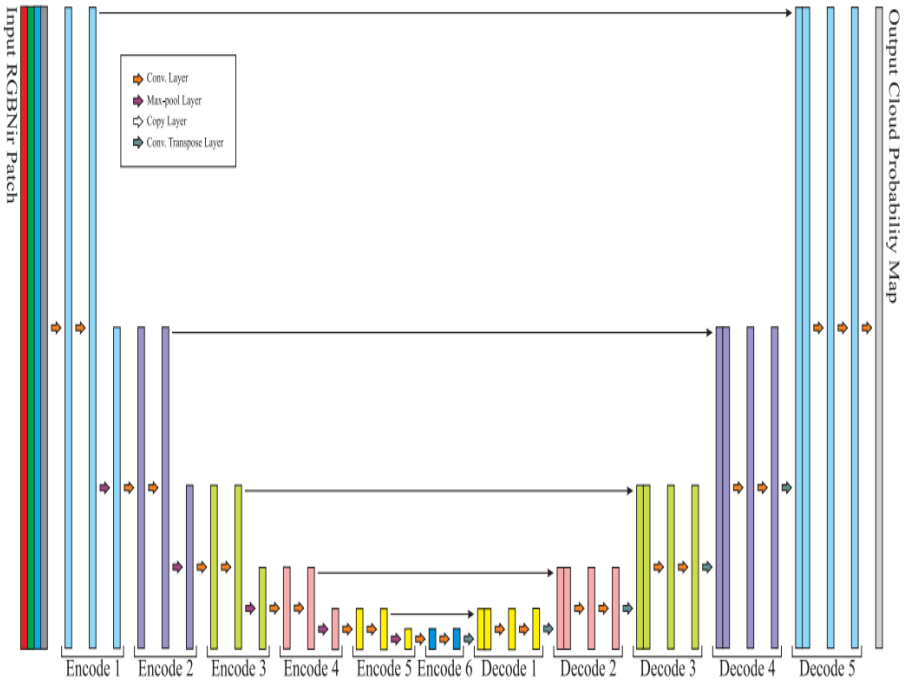

Figura 5: Arquitetura de modelo similar à U-Net, conforme apresentado em Mohajerani, Krammer e Saeedi (2018a).

(banda 3), Azul (banda 2) e Infravermelho Próximo (banda 5). Ao contrário de outras imagens de visão computacional, esses canais não são combinados. Em vez disso, eles estão em seus diretórios correspondentes. (MOHAJERANI, 2019a)

3.1.2 95-Cloud

O conjunto de dados 95-Cloud, uma extensão do conjunto de dados 38-Cloud. Ele consiste em 34.701 *patches* de 384x384 para treinamento. O conjunto de testes do 95-Cloud é exatamente o mesmo do 38-Cloud. Os *patches* de treinamento são extraídos de 75 cenas do Landsat 8 Collection 1 Level-1, localizadas principalmente na América do Norte. O conjunto de testes do 95-Cloud inclui 9201 *patches* de 20 cenas. No entanto, suas imagens de treinamento são diferentes. O 95-Cloud possui 57 cenas a mais do que o 38-Cloud para treinamento. (MOHAJERANI, 2019b)

3.2 Tratamento dos Dados

Os dados do 38-cloud e do 95-cloud apresentam um número considerável de imagens com uma porcentagem alta de pixeis nulos, exigindo um tratamento adequado. Uma das etapas de pré-processamento propostas neste trabalho é a remoção das imagens com pixeis nulos dos conjuntos que não contêm dados relevantes, como ilustrado na Figura 7.

3.2.1 Remoção de imagens com alta porcentagem de pixeis nulos

Para realizar a remoção das imagens, adotou-se a seguinte técnica: após a montagem da base de dados, as imagens são submetidas a uma função que identifica aquelas com alto percentual de pixels nulos. Esta técnica verifica se há mais de 80% dos pixels nulos em todas as bandas (vermelho, verde, azul e infravermelho próximo) que como mostrado na Figura 6 e 7 estas são as porcentagem com o maior numero de imagens com pixels nulos. Caso essa condição seja satisfeita, a imagem é removida do conjunto de dados. Isso contribui para garantir que somente imagens com conteúdo relevante sejam utilizadas durante o treinamento e a avaliação do modelo. A Figura 8 ilustra um exemplo de imagem com bordas removidas após a aplicação dessa técnica.

Essa técnica de remoção de bordas contribui para aprimorar a qualidade dos dados de entrada fornecidos ao modelo durante o treinamento, reduzindo o ruído e aumentando a eficácia do aprendizado. Essa abordagem é crucial para garantir que o modelo seja capaz de aprender padrões significativos nos dados e realizar previsões precisas em novos exemplos.

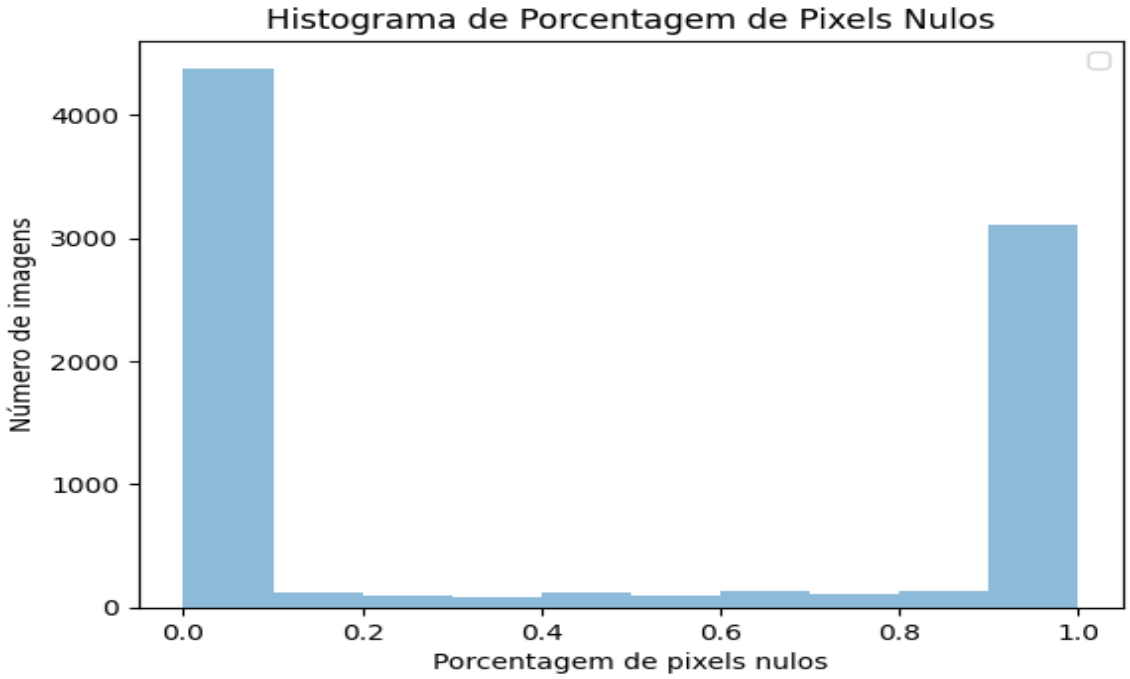

Figura 6: Histograma de porcentagem de pixeis nulos para a 38-Cloud. Fonte: Autoria Própria.

3.2.2 Aumento dos dados

Após removidas imagens com percentual elevado de pixeis nulos, a base de dados foi reduzida para 5155 imagens para treinamento. Para evitar o superajuste, é viável realizar um pré-processamento das imagens aplicando técnicas de aumento de dados.

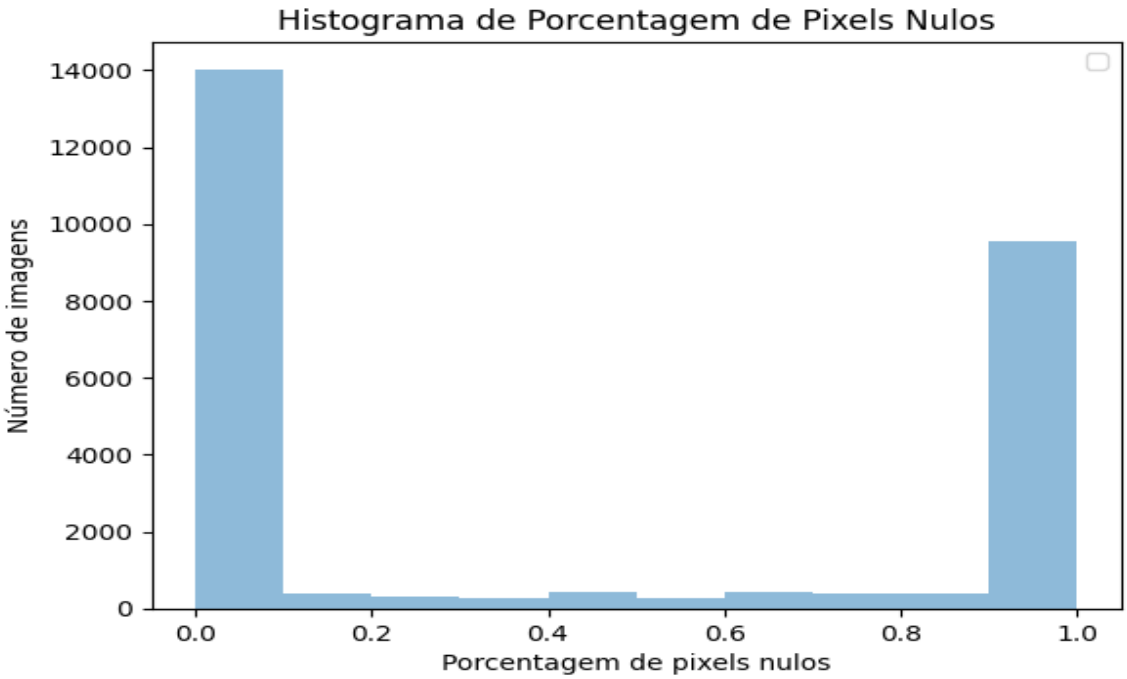

Figura 7: Histograma de porcentagem de pixeis nulos para a 95-Cloud. Fonte: Autoria Própria.

Modelos de aprendizagem profunda demandam grandes quantidades de dados para um aprendizado eficaz. No caso da base de dados 38-Cloud, com um número relativamente baixo de imagens, o aumento de dados torna-se essencial.

O método de aumento de dados proposto por (MOHAJERANI; SAEEDI, 2019) foi adotado, e além das transformações propostas por eles, também foi implementado um ajuste adicional de brilho e contraste. As seguintes transformações foram aplicadas:

1. **Rotação:** Rotaciona aleatoriamente a imagem em ±90 graus ilustrado na Figura 9.

2. **Espelhamento:** A ordem dos elementos na imagem é invertida por meio de um processo de espelhamento aleatório nos eixos horizontal e vertical, como demonstrado na Figura 10.

3. **Transformação de Escala, Translação e Rotação:** Esta transformação aplica uma combinação de deslocamento (translação), redimensionamento (zoom) e rotação fina. O limite de deslocamento é de 5%, o limite de redimensionamento é também de 5% (para aumentar ou diminuir a escala), e o limite de rotação é de ±15°. Esta operação pode ser vista como uma transformação mais sutil e detalhada que ajusta a posição, escala e orientação da imagem de forma leve, ilustrada na Figura 11.

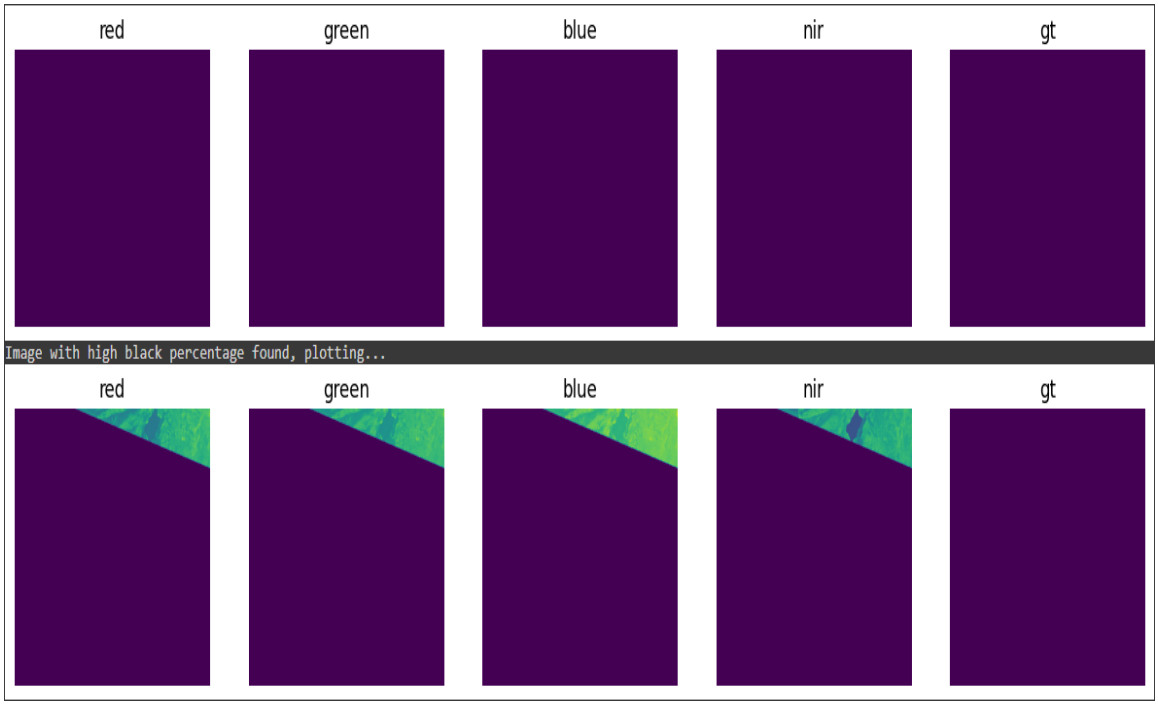

Figura 8: Exemplo de imagem com alta porcentagem de pixels nulos. Fonte: Autoria Própria.

4. **Ajuste de Brilho e Contraste:** Ajusta o brilho e o contraste da imagem de forma aleatória. Esse ajuste pode fazer com que a imagem se torne temporariamente mais clara ou mais escura, ou que o contraste seja aumentado ou diminuído, conforme exemplificado na Figura 12. O limite de ajuste para o brilho e o contraste é de 0.2, proporcionando variações sutis nos níveis de brilho e contraste.

Essas transformações serão aplicadas durante o treinamento dos modelos.

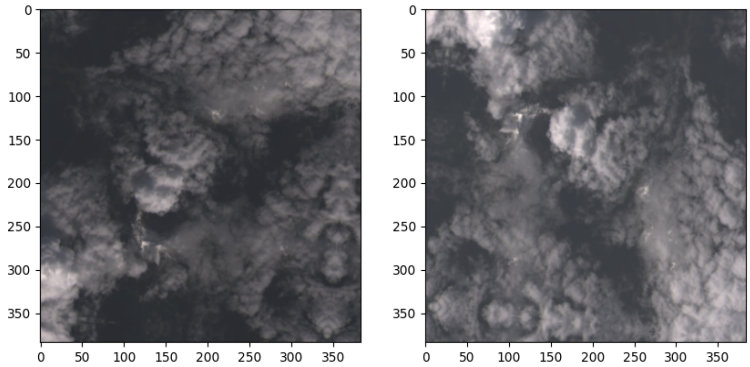

Figura 9: Rotação da imagem. Fonte: Autoria Própria

3.3 Modelos e Técnicas Selecionados

Na abordagem da segmentação semântica de nuvens, a escolha foi explorar tanto uma abordagem tradicional quanto modelos baseados em aprendizagem profunda. No

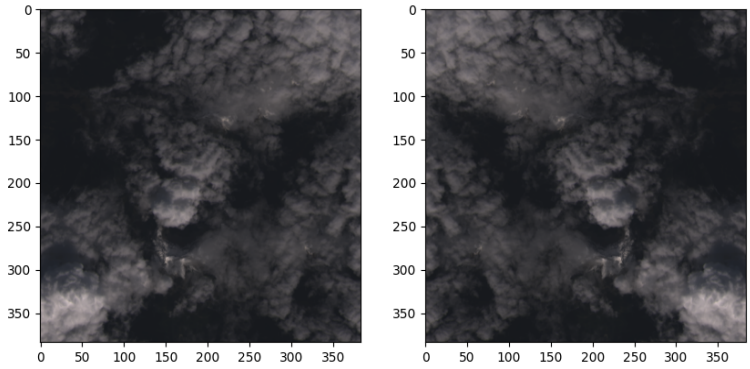

Figura 10: Espelhamento da imagem. Fonte: Autoria Própria

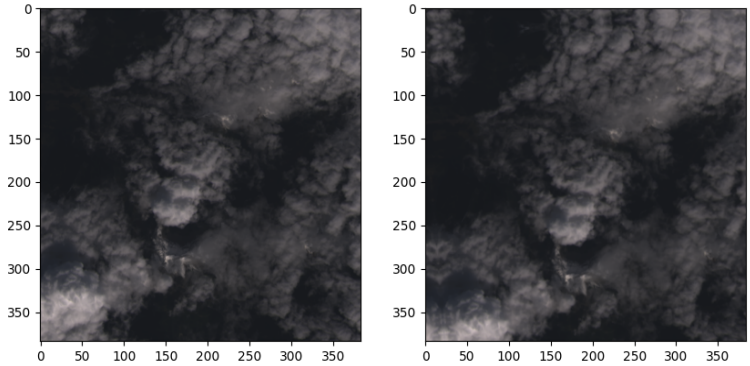

Figura 11: Transformação de Escala, Translação e Rotação. Fonte: Autoria Própria

entanto, para os modelos de rede neural utilizados nas medidas comparativas. Os modelos selecionados são:

3.3.1 Modelo baseado em limiar

O método Fmask foi a abordagem tradicional selecionada, sendo um método de segmentação reconhecido e utilizado no processamento de imagens de sensoriamento remoto, especialmente para a identificação e separação de nuvens e suas sombras nas imagens de satélite.

O Fmask é uma ferramenta criada especificamente para lidar com imagens de satélite de alta qualidade, como aquelas que vêm dos satélites Landsat ou Sentinel. Essas imagens são bastante usadas em várias áreas, como monitoramento ambiental, mapeamento de uso da terra e agricultura, entre outras. É essencial poder detectar com precisão as nuvens e as sombras nessas imagens, porque a presença delas pode atrapalhar bastante a visualização da superfície da Terra. Isso pode levar a interpretações incorretas dos dados que estão sendo analisados. (ZHU; WOODCOCK, 2012).

1. **Limiar de Temperatura:** O Fmask inicialmente utiliza a temperatura da super-

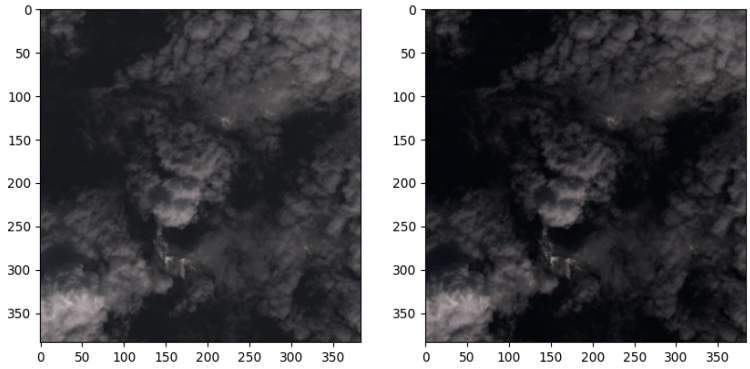

Figura 12: Ajuste de Brilho e Contraste. Fonte: Autoria Própria

fície, derivada do canal térmico da imagem de satélite, para identificar os pixels de nuvem. As nuvens, em geral, têm temperaturas mais frias na superfície em comparação com a terra ou a água em condições normais. Aplicando um limiar específico de temperatura, o algoritmo pode separar eficientemente a maioria dos pixels de nuvem. Além disso, esse método é fundamental para a detecção inicial das nuvens.

Exemplo da Matriz de Limiar de Temperatura:

$$\begin{bmatrix} \text{Temperatura} & \text{Classificação} \\ \text{Pixel 1 (15°C)} & \text{Nuvem} \\ \text{Pixel 2 (10°C)} & \text{Nuvem} \\ \text{Pixel 3 (25°C)} & \text{Terra} \\ \text{Pixel 4 (5°C)} & \text{Nuvem} \\ \cdots & \cdots \end{bmatrix}$$

2. **Detecção de Sombra:** Após a identificação das nuvens, o Fmask procura por sombras correspondentes. Isso é feito considerando a posição do sol no momento em que a imagem foi capturada e projetando a possível localização da sombra com base na altura estimada da nuvem. Este passo é crucial, pois as sombras podem afetar a interpretação dos dados subjacentes da mesma forma que as nuvens.

   Exemplo da Matriz de Detecção de Sombra:

$$\begin{bmatrix} \text{Posição Solar} & \text{Classificação} \\ \text{Pixel 1 (Sudeste)} & \text{Sombra} \\ \text{Pixel 2 (Sudoeste)} & \text{Sombra} \\ \text{Pixel 3 (Norte)} & \text{Nuvem} \\ \text{Pixel 4 (Nordeste)} & \text{Sombra} \\ \cdots & \cdots \end{bmatrix}$$

3. **Refinamento através de Índices Espectrais:** O refinamento por índices es-

pectrais é uma ferramenta poderosa para melhorar a precisão da segmentação de nuvens e sombras em imagens de satélite. Ao compreender a definição e aplicação dos índices, como NDVI e NDWI, pode aprimorar a análise de imagens e obter informações mais precisas sobre a cobertura do solo e suas características.

Exemplo da Matriz de Refinamento por Índices Espectrais:

$$\begin{bmatrix} \text{NDVI} & \text{NDWI} & \text{Classificação} \\ \text{Pixel 1 (0.8)} & \text{Pixel 1 (0.1)} & \text{Nuvem} \\ \text{Pixel 2 (0.6)} & \text{Pixel 2 (0.3)} & \text{Terra} \\ \text{Pixel 3 (0.2)} & \text{Pixel 3 (0.7)} & \text{Terra} \\ \text{Pixel 4 (0.7)} & \text{Pixel 4 (0.2)} & \text{Nuvem} \\ \cdots & \cdots & \cdots \end{bmatrix}$$

4. **Correção e Ajustes Finais:** Por fim, o Fmask aplica uma série de correções para ajustar os limites de nuvens e sombras, considerando aspectos como o espalhamento de luz e o albedo da superfície. Essas correções ajudam a minimizar a classificação incorreta de superfícies brilhantes (como corpos d'água ou telhados) como nuvens. Essa fase final é essencial para aprimorar a precisão e a confiabilidade da segmentação.

3.3.2 Modelos de Aprendizagem Profunda

U-net: A U-net é reconhecida como um modelo amplamente utilizado em tarefas de segmentação semântica em imagens, especialmente pela sua habilidade de preservar detalhes finos durante o processo de segmentação, o que a torna particularmente eficaz em cenários onde a precisão nos contornos dos objetos é crucial. A escolha da U-net para este trabalho se fundamenta em sua eficácia comprovada na segmentação de imagens, aliada à sua capacidade de lidar com imagens de alta resolução e segmentar objetos de diferentes tamanhos e formas com precisão. (RONNEBERGER; FISCHER; BROX, 2015)

A arquitetura da U-net segue um paradigma de codificador-decodificador, conforme demonstrado na Figura 5. O codificador, localizado à esquerda da figura, realiza uma redução gradual da informação da imagem de entrada em escala e uma ampliação em complexidade por meio de camadas de convolução e pooling, capturando características relevantes da imagem em diferentes níveis de abstração. Por outro lado, o decodificador, posicionado à direita da figura, constrói a máscara de segmentação a partir das características codificadas, utilizando camadas de convolução transposta para restaurar a resolução da imagem e preservar detalhes finos (RONNEBERGER; FISCHER; BROX, 2015). Um exemplo desse processo de segmentação pode ser visto na Figura 13.

A simetria na arquitetura de codificador-decodificador permite que a U-net capture

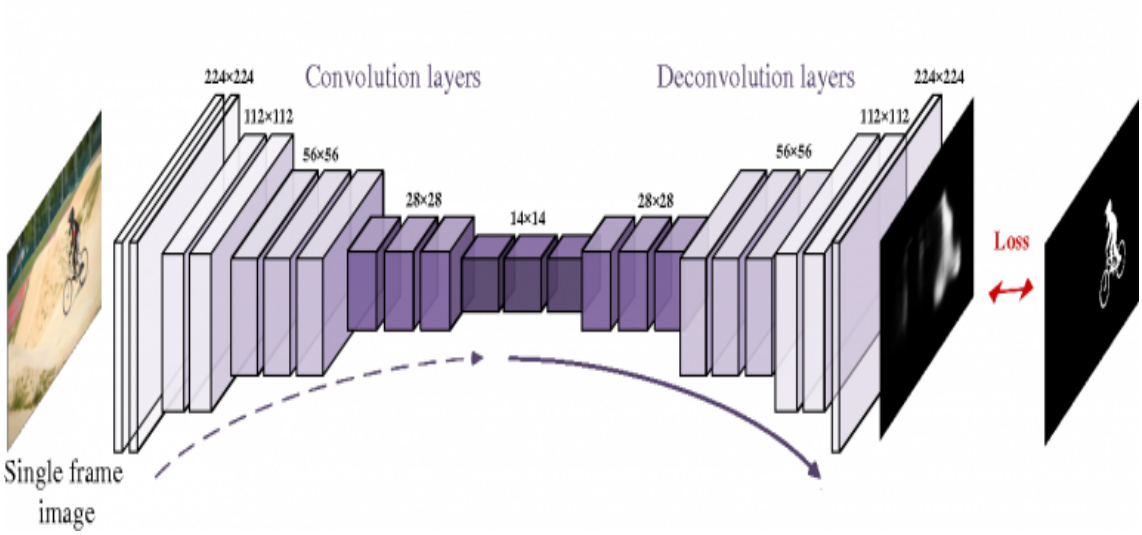

Figura 13: Exemplo de segmentação de imagem comprimida e descomprimida por um modelo. Fonte: (WANGENHEIM, 2018)

eficientemente informações contextuais em diferentes escalas, resultando em segmentações detalhadas e precisas, características essenciais para a tarefa de segmentação de nuvens.

FPN (*Feature Pyramid Network*): O FPN (*Feature Pyramid Network*) destaca-se como outro modelo de aprendizagem profunda amplamente empregado para a segmentação semântica e detecção de objetos em imagens. Em contraste com a abordagem da U-net, que se baseia em uma arquitetura de codificador-decodificador, o FPN utiliza pirâmides de características para aprimorar a precisão na segmentação de objetos em diversas escalas, o que o torna particularmente eficiente em cenários nos quais os objetos possuem tamanhos variados como ilustrado na Figura 14. (LIN et al., 2017)

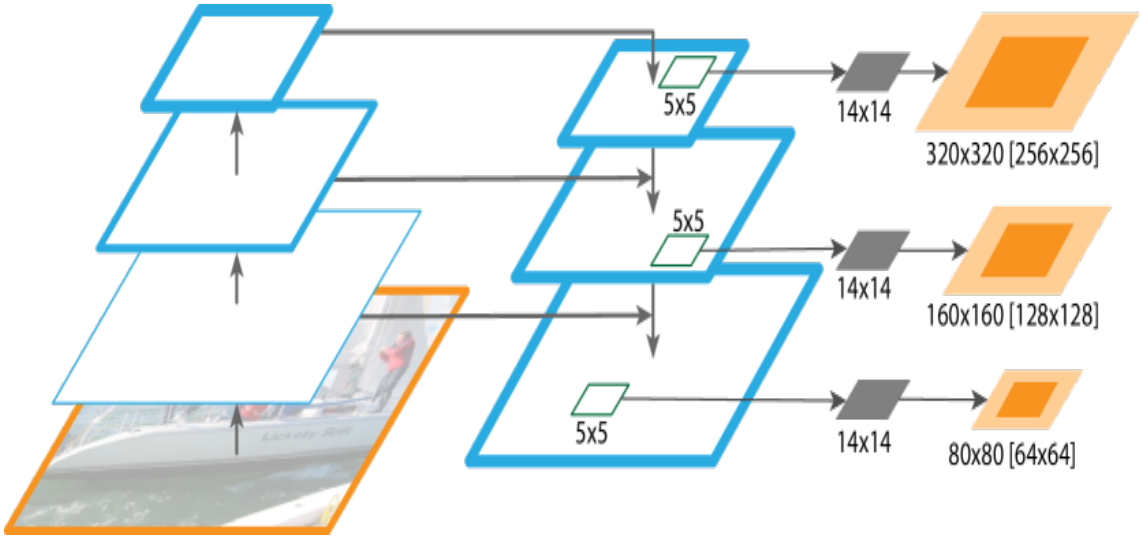

Figura 14: Representação da FPN-net em segmentação. Fonte: (LIN et al., 2017)

A explicação da Figura 14 por Lin et al. (2017) diz que a pirâmide de características

é construída com a mesma estrutura que para detecção de objetos. Aplica-se um pequeno MLP (*Multilayer Perceptron*) em janelas de 5×5 para gerar segmentos densos de objetos com uma dimensão de saída de 14×14. As regiões da imagem correspondentes à máscara para cada nível da pirâmide são mostradas em laranja (neste caso, os níveis 3 a 5). Tanto o tamanho correspondente da região da imagem (laranja claro) quanto o tamanho canônico do objeto (laranja escuro) são exibidos.

Na implementação do FPN, é comum integrá-lo a *backbones* (a ser definido no próximo paragrafo) de redes neurais convolucionais pré-treinadas, como a ResNet, para a extração de características de diferentes níveis de abstração da imagem. Posteriormente, uma cabeça FPN é incorporada para mesclar e refinar essas características em várias escalas, resultando em uma segmentação precisa. (LIN et al., 2017)

O termo *backbone* refere-se aos primeiros nós de uma rede neural encarregado de extrair características das imagens. Nos modelos de aprendizagem profunda voltados para visão computacional, o *backbone* geralmente é composto por uma sequência de camadas convolucionais, cujo propósito é capturar diversos aspectos visuais das imagens, desde bordas e texturas até características mais complexas, à medida que as camadas vão se aprofundando. A escolha do *backbone* pode influenciar significativamente o desempenho da rede neural em tarefas específicas, como segmentação de imagens e detecção de objetos, uma vez que uma representação adequada das características visuais é essencial para essas aplicações.

A cabeça FPN desempenha um papel crucial, agindo como sua principal unidade de processamento de características. Ela recebe características extraídas de diferentes níveis da rede ResNet18 e, por meio de uma combinação de convoluções 1x1 e interpolações, realiza a fusão e integração dessas características em várias escalas espaciais. Essa abordagem permite que a FPNHead crie uma pirâmide de características que captura informações contextuais em diferentes resoluções, possibilitando uma representação mais abrangente e detalhada da cena visual. Essa pirâmide de características é essencial para tarefas de visão computacional, como detecção de objetos e segmentação semântica, pois permite que a rede localize e reconheça objetos de diferentes tamanhos e em diferentes contextos.

No contexto deste trabalho, a ResNet-18 foi selecionada como *backbone*. A ResNet-18 é uma versão mais leve da ResNet (HE et al., 2015), composta por 18 camadas profundas e que utiliza conexões residuais para facilitar o treinamento, permitindo o fluxo de gradientes por caminhos alternativos durante o processo de retro-propagação. Essa escolha proporciona uma base sólida para a extração de características em múltiplas escalas. A combinação da FPN com a ResNet-18 possibilita uma segmentação eficaz em diferentes níveis de granularidade, aproveitando as características extraídas pelo *backbone* para construir uma pirâmide de características que facilita a detecção e segmentação de

objetos em várias escalas. Esta abordagem é particularmente vantajosa em aplicações nas quais os objetos de interesse podem variar significativamente em tamanho, como é o caso da segmentação de nuvens.

YOLOv8: YOLOv8 é uma evolução da família de arquiteturas YOLO, que se destaca como uma das abordagens mais eficientes para detecção de objetos em tempo real como mostra o gráfico na Figura 15. Além disso, pode ser adaptada para tarefas de segmentação semântica, oferecendo uma solução integrada para ambas as tarefas em um único modelo (JOCHER; CHAURASIA; QIU, 2023). Dividida em *Nano*(n), *Small*(s), *Medium*(m), *Large*(l), *Extra Large*(x).

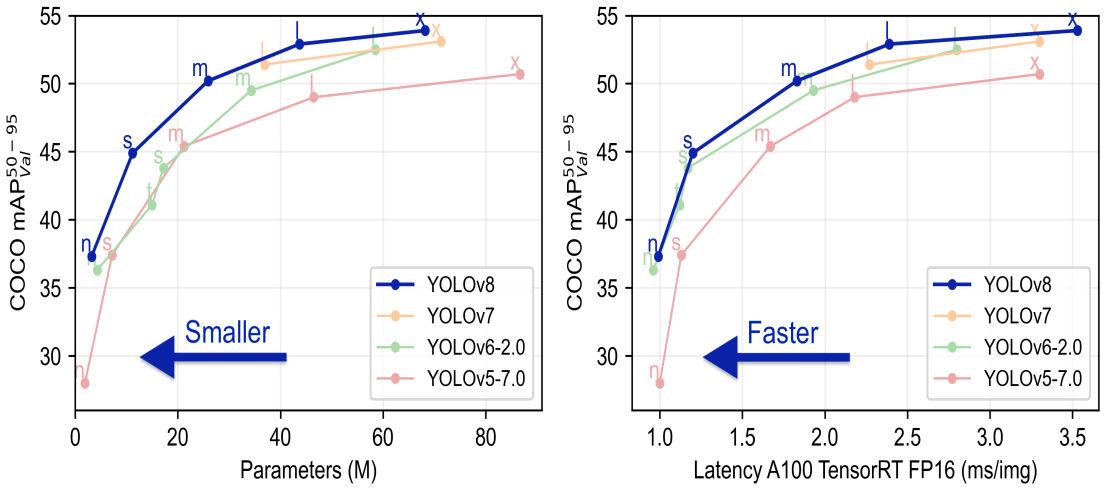

Figura 15: Comparação de modelos YOLO. Fonte: (JOCHER; CHAURASIA; QIU, 2023)

A arquitetura YOLOv8 (Figura 16) apresenta uma abordagem única conhecida como detecção de objetos em uma etapa. Diferentemente de abordagens tradicionais que dividem a detecção em várias etapas, como a extração de características seguida pela predição de caixas delimitadoras e classificação, o YOLOv8 aborda a detecção de objetos como um problema de regressão direta, onde um único modelo prediz diretamente as caixas delimitadoras e as classes dos objetos em uma única passagem pela rede. Essa abordagem proporciona tempos de inferência rápidos, tornando-o ideal para aplicações em tempo real (JOCHER; CHAURASIA; QIU, 2023).

A arquitetura YOLOv8 é composta por várias camadas convolucionais, seguidas por camadas de detecção que dividem a imagem em uma grade e fazem previsões para cada célula dessa grade. Essas previsões incluem caixas delimitadoras que definem a localização e o tamanho dos objetos detectados, bem como as probabilidades associadas a cada classe de objeto. A utilização de uma única rede para prever todas essas informações simplifica o processo de detecção e permite uma inferência rápida, sem comprometer a precisão.

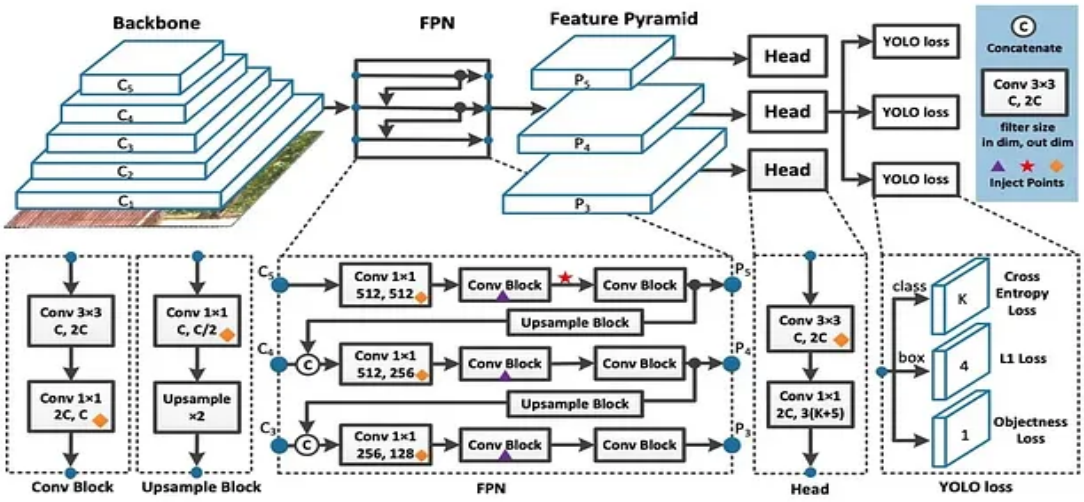

Figura 16: Arquitetura yolo v8. Fonte: (ALI, 2023)

Ademais, o YOLOv8 pode ser adaptado para tarefas de segmentação semântica por meio de ajustes na saída da rede e na forma como as previsões são interpretadas. Em vez de prever caixas delimitadoras e classes, o modelo pode ser treinado para prever máscaras de segmentação para cada classe de objeto, fornecendo uma solução abrangente para tarefas de detecção e segmentação.

Essa abordagem integrada oferecida pelo YOLOv8 é especialmente vantajosa em cenários nos quais a eficiência computacional e a rapidez na inferência são prioritárias, como em sistemas de vigilância em tempo real e aplicações de monitoramento de tráfego. Além disso, a capacidade de adaptar o YOLOv8 para tarefas de segmentação semântica amplia sua aplicabilidade em uma variedade de cenários, desde a contagem de veículos em estradas até a detecção e classificação de objetos em imagens médicas.

CloudNet: O Cloud-Net é um algoritmo de detecção de nuvens em imagens de satélite que se destaca por sua abordagem de aprendizagem profunda. Diferentemente de métodos tradicionais, o Cloud-Net é capaz de aprender características locais e globais das nuvens de forma integrada, sem a necessidade de etapas de pré-processamento complexas, como segmentação de superpixels. Além disso, o Cloud-Net utiliza blocos de convolução sofisticados para extrair características e gerar máscaras de nuvens mais precisas, resultando em um desempenho superior em comparação com métodos existentes. (MOHAJERANI; SAEEDI, 2019)

A arquitetura do Cloud-Net (Figura 4) segue um paradigma de codificador-decodificador, inspirado na U-Net. O codificador do Cloud-Net extrai características da imagem de entrada em diferentes níveis de abstração, enquanto o decodificador reconstrói a máscara de segmentação com base nessas características, preservando detalhes finos e gerando uma mapa de probabilidade de nuvens em que cada pixel representa a probabilidade de

pertencer à classe de nuvem. (MOHAJERANI; SAEEDI, 2019)

Além disso, o Cloud-Net é capaz de capturar informações contextuais em diferentes escalas, o que resulta em segmentações detalhadas e precisas, características essenciais para a detecção de nuvens em imagens de satélite. A eficácia do Cloud-Net é evidenciada por sua capacidade de superar métodos de ponta existentes, proporcionando uma detecção de nuvens mais precisa e confiável. (MOHAJERANI; SAEEDI, 2019)

3.4 Critérios de Avaliação

A escolha dos critérios de avaliação desempenha um papel crucial na comparação eficiente dos modelos selecionados. Para isso, serão utilizadas as seguintes métricas:

3.4.1 IoU (Interseção sobre União)

A métrica de IoU, também conhecida como Índice de Jaccard, é uma medida que avalia a sobreposição entre a área predita e a área verdadeira, fornecendo uma medida de quão bem o modelo de segmentação está desempenhando. O cálculo do IoU é feito da seguinte maneira:

$$\text{IoU} = \frac{\text{Verdadeiros Positivos}}{\text{Verdadeiros Positivos} + \text{Falsos Positivos} + \text{Falsos Negativos}}$$

Onde:

- **Verdadeiros Positivos (VP)**: Os exemplos corretamente classificados como nuvens.
- **Falsos Positivos (FP)**: Os exemplos erroneamente classificados como nuvens.
- **Falsos Negativos (FN)**: O número de casos em que o modelo classificou incorretamente a classe não nuvem quando na verdade era nuvem.

O resultado do IoU varia de 0 a 1, onde 0 indica nenhuma sobreposição e 1 indica uma sobreposição perfeita entre as duas máscaras. Portanto, quanto mais próximo o valor do IoU estiver de 1, melhor será a precisão da segmentação. Um exemplo que pode acontecer é o seguinte:

- Verdadeiros Positivos (VP) = 100
- Falsos Positivos (FP) = 20

- Falsos Negativos (FN) = 30

pode representar esses dados em uma matriz de confusão da seguinte forma:

–	Nuvem Predita	Não Nuvem Predita
Nuvem Verdadeira	100	20
Não Nuvem Verdadeira	30	TN

Agora, calculando o IoU usando a fórmula:

$$\text{IoU} = \frac{\text{VP}}{\text{VP} + \text{FP} + \text{FN}} = \frac{100}{100 + 20 + 30} = \frac{100}{150} = 0.6667$$

Portanto, o resultado do IoU para esses dados é aproximadamente 0.6667. Isso indica que há uma sobreposição substancial entre a área predita e a área verdadeira, o que sugere uma precisão razoável na segmentação.

3.4.2 Acurácia

A acurácia é uma métrica fundamental para avaliar a precisão geral do modelo na tarefa de segmentação. Ela mede a proporção de pixels classificados corretamente em relação ao total de pixels na imagem. A fórmula para calcular a acurácia é dada por:

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN}$$

Onde:

- **Verdadeiros Negativos (VN)**: Os exemplos corretamente classificados como não nuvem.
- O numerador (VP + VN) representa o número total de previsões corretas feitas pelo modelo.
- O denominador (VP + VN + FP + FN) representa o número total de previsões feitas pelo modelo, independentemente de estarem corretas ou não.

A acurácia fornece uma medida geral da qualidade da segmentação, indicando a capacidade do modelo de distinguir corretamente entre as classes de interesse e as áreas de fundo. Quanto mais alta for a acurácia, melhor será o desempenho do modelo na segmentação das imagens. Para mostrar como funciona este calculo segue o seguinte exemplo:

- Verdadeiros Positivos (VP) = 100
- Verdadeiros Negativos (VN) = 200
- Falsos Positivos (FP) = 20
- Falsos Negativos (FN) = 30

Pode representar esses dados em uma matriz de confusão da seguinte forma:

	Nuvem Predita	Não Nuvem Predita
Nuvem Verdadeira	100	20
Não Nuvem Verdadeira	30	200

Agora, calculando a acurácia usando a fórmula fornecida:

$$\text{Acurácia} = \frac{VP + VN}{VP + VN + FP + FN} = \frac{100 + 200}{100 + 200 + 20 + 30} = \frac{300}{350} = 0.8571$$

Portanto, o resultado da acurácia para esses dados é aproximadamente 0.8571. Isso indica que aproximadamente 85.71% dos pixels foram classificados corretamente pelo modelo em relação ao total de pixels na imagem. Quanto mais próximo o valor da acurácia estiver de 1, melhor será o desempenho do modelo na segmentação das imagens.

No contexto de segmentação semântica, a acurácia é uma métrica importante, juntamente com o IoU, fornecendo uma avaliação completa do desempenho do modelo em diferentes aspectos da segmentação de imagens.

3.4.3 Precisão e Revocação

Precisão e Revocação são métricas importantes na avaliação de modelos de classificação, especialmente em problemas onde o desequilíbrio de classes é comum.

Precisão Precisão é a medida da precisão das previsões positivas do modelo. Isso significa que mede a proporção de exemplos que o modelo classificou corretamente como positivos (VP) em relação ao total de exemplos que ele classificou como positivos, incluindo os verdadeiros e os falsos positivos (VP + FP). A precisão mostra a habilidade do modelo em identificar corretamente os exemplos positivos.

$$\textit{Precisão} = \frac{\text{VP}}{\text{VP} + \text{FP}}$$

Revocação Mede a sensibilidade do modelo para encontrar todos os exemplos positivos (nuvem) na base de dados. Essa métrica é a proporção de exemplos de nuvens que foram corretamente identificados pelo modelo em relação ao total de exemplos de nuvens na base de dados, incluindo os verdadeiros positivos e os falsos negativos. Em outras palavras, a revocação mostra a capacidade do modelo em recuperar todos os exemplos positivos presentes nos dados.

$$Revocação = \frac{\text{VP}}{\text{VP} + \text{FN}}$$

Essas métricas são cruciais para entender como o modelo está em relação às classes de interesse e podem ajudar a ajustar os modelos para maximizar a eficácia na identificação dos exemplos relevantes.

3.4.4 Perda de Entropia Cruzada

A entropia cruzada (ou *cross-entropy* em inglês) é uma métrica comumente utilizada para avaliar a disparidade entre duas distribuições de probabilidade. É frequentemente empregada em problemas de classificação no campo do aprendizado de máquina, principalmente em contextos de aprendizado supervisionado (CECCON, 2019).

A entropia cruzada entre duas distribuições de probabilidade y e $\hat{y}$ é calculada pela seguinte fórmula:

$$H(y, \hat{y}) = -\sum_{i} y_i \log(\hat{y}_i)$$

Onde:

- y_i representa a probabilidade real da classe i (ou seja, o rótulo real);
- $\hat{y}_i$ indica a probabilidade estimada pelo modelo para a classe i.

Basicamente, a entropia cruzada quantifica o quão bem a distribuição de probabilidade estimada $\hat{y}$ se assemelha à distribuição de probabilidade real y. Quando a

distribuição estimada se aproxima da real, a entropia cruzada é baixa. Em contrapartida, se a distribuição estimada difere consideravelmente da real, a entropia cruzada é alta.

3.4.5 Perda de Segmentação

A perda de segmentação é uma forma simples de entropia cruzada binária, calculada pixel a pixel, e é aplicada após a máscara estimada ser redimensionada e ajustada para coincidir com as coordenadas da máscara verdadeira. Vale ressaltar que apenas a máscara estimada para a classe correta é considerada no cálculo da perda, enquanto quaisquer outras máscaras associadas a classes incorretas são desconsideradas. Esta abordagem é citada no livro de Lakshmanan, Görner e Gillard (2021).

3.4.6 Perda Focal de Distribuição

A Perda Focal de Distribuição é uma técnica desenvolvida no contexto de detecção de objetos densamente distribuídos, visando aprimorar a precisão na estimativa da localização de caixas delimitadoras. Esta perda modela as localizações das caixas delimitadoras como distribuições gerais, em vez de adotar uma abordagem rígida ou simplificada.

Esta perda é projetada para direcionar rapidamente as redes neurais para aprender as probabilidades de valores próximos às coordenadas-alvo das caixas delimitadoras. Isso é alcançado otimizando a forma da distribuição de probabilidade para enfatizar as altas probabilidades de valores próximos ao alvo desejado, tornando as estimativas de localização mais precisas e informativas (LI et al., 2020).

$$DFL(S_i, S_{i+1}) = -\left((y_{i+1} - y)\log(S_i) + (y - y_i)\log(S_{i+1})\right)$$

Nesta fórmula:

- S_i e S_{i+1} representam as probabilidades estimadas para os valores y_i e y_{i+1}, que são os valores mais próximos do valor alvo y nas coordenadas da caixa delimitadora.
- y é o valor alvo desejado para a localização da caixa delimitadora.
- y_i e y_{i+1} são os valores mais próximos de y (os valores ao redor do alvo).
- O termo $(y_{i+1} - y)\log(S_i)$ penaliza a diferença entre a estimativa S_i e o valor alvo y quando y_{i+1} é maior que y.

- O termo $(y - y_i)\log(S_{i+1})$ penaliza a diferença entre a estimativa S_{i+1} e o valor alvo y quando y é maior que y_i.

3.4.7 mAP (Precisão média média)

O mAP 50 e o mAP50-95 são métricas de desempenho utilizadas na avaliação de modelos de detecção de objetos, como a YOLO. Eles medem a precisão média (AP) do modelo em diferentes níveis de sobreposição entre a caixa delimitadora prevista e a caixa delimitadora real do objeto. (JOCHER; CHAURASIA; QIU, 2023)

- mAP 50: Mede a AP em um único limiar de IoU (Intersection over Union) de 0,5. Indica a precisão do modelo quando a máscara segmentada pelo modelo se sobrepõe à máscara real em pelo menos 50
- mAP50-95: Mede a AP em vários limiares de IoU, variando de 0,5 a 0,95. Oferece uma visão mais abrangente do desempenho do modelo em diferentes níveis de granularidade da segmentação.

3.5 Treinamento dos Modelos

Após a preparação, que incluiu a remoção de imagens com alta porcentagem de pixeis nulos, os dados foram carregados e utilizou-se um procedimento específico para aplicar as transformações de aumento de dados durante o treinamento onde cada imagem teria probabilidade de 50% de ser aplicado as transformação. Em seguida, as bandas R, G e B foram selecionadas da base de dados, e os dados foram divididos em lotes apropriados para o treinamento dos modelos. Com essa preparação concluída, procedeu-se ao treinamento dos modelos selecionados, utilizando um processador Intel(R) Xeon(R) CPU @ 2.30GHz e uma GPU (placa gráfica) Tesla T4, para o modelo da Cloud-net será utilizado os pesos pre-treinados disponibilizado por Mohajerani (2019c). As especifidades de treinamento dos modelos U-Net, FPN-net e YOLO, detalhadas nas sessões seguintes.

3.5.1 U-Net e FPN-Net

Para o treinamento da U-Net e da FPN-Net, foi aplicado o mesmo procedimento de preparação dos dados e o mesmo sistema de treinamento.

1. **Configuração do Treinamento:** Foi adotada a função de perda entropia cruzada, e foi utilizado o otimizador Adam. Além disso, foram ajustados os hiper-parâmetros, incluindo uma taxa de aprendizado de 0.001, um número de épocas de 160 e um critério de parada precoce de 35 épocas, com base em experimentações preliminares.

2. **Treinamento do Modelo:** O modelo foi treinado utilizando o conjunto de dados preparado, iterando sobre os lotes de dados e ajustando os pesos da rede neural com base na função de perda calculada.

3. **Avaliação do Desempenho:** Durante o treinamento, foram monitoradas métricas de desempenho, como o IoU e a acurácia, utilizando um conjunto de validação separado. Isso permitiu avaliar o progresso do modelo e identificar possíveis problemas de sobre-ajuste ou subajuste.

U-Net: O treinamento da U-Net foi relativamente suave, conforme mostrado no gráfico de perda na Figura 17. Em alguns pontos o modelo não demonstrou grande aprendizado com o passar das épocas, diferentemente da FPN-Net, a U-Net não conseguiu aprender além de 125 épocas com essa configuração de hiper-parâmetros caindo no parâmetro de parada precoce.

Os parâmetros de IoU na validação apresentaram os resultados mostrados na Figura 18. É perceptível que, à medida que a perda diminui, as métricas de IoU aumentam. Ao longo das épocas, o modelo demonstrou capacidade de aprendizado, alcançando um pico de 0.90 de IoU.

FPN-Net: O treinamento da FPN-Net foi mais lento, conforme evidenciado pelo gráfico de perda, onde o modelo teve oscilações frequentes, como mostrado na Figura 19. No entanto, diferente da U-Net, o modelo completou o número determinado de épocas, indicando que ainda estava aprendendo, embora de maneira mais lenta que a U-Net, com uma diferença de mais de 0.2, como mostrado na Figura 20. Quanto às métricas de IoU, foram muito semelhantes às da U-Net, chegando a 0.90 de IoU na validação.

3.5.2 Yolo

1. **Base de Dados:** Para o treinamento do Yolo, a base de dados precisou passar por uma fase adicional de processamento, pois foi utilizada a biblioteca ultralitcs para o treinamento. Dado que o modelo não compreende máscaras do tipo imagens de 0 e 1, apenas polígonos em arquivos .txt. A seleção automática dos melhores aumentos de dados ocorre automaticamente. Neste caso, foram selecionados os aumentos de *Blur* e *MedianBlur* de 0.01, *toGray* de 0.01 e *clip_limit* entre 1 a 4.0.

2. **Configuração do Treinamento:** Os hiper-parâmetros utilizados incluíram a função de otimização SGD (descida de gradiente estocástica) com uma taxa de aprendizado de 0.01 e *momentum* de 0.9, lotes (*batch size*) de 32, 150 épocas e 35 para parada precoce.

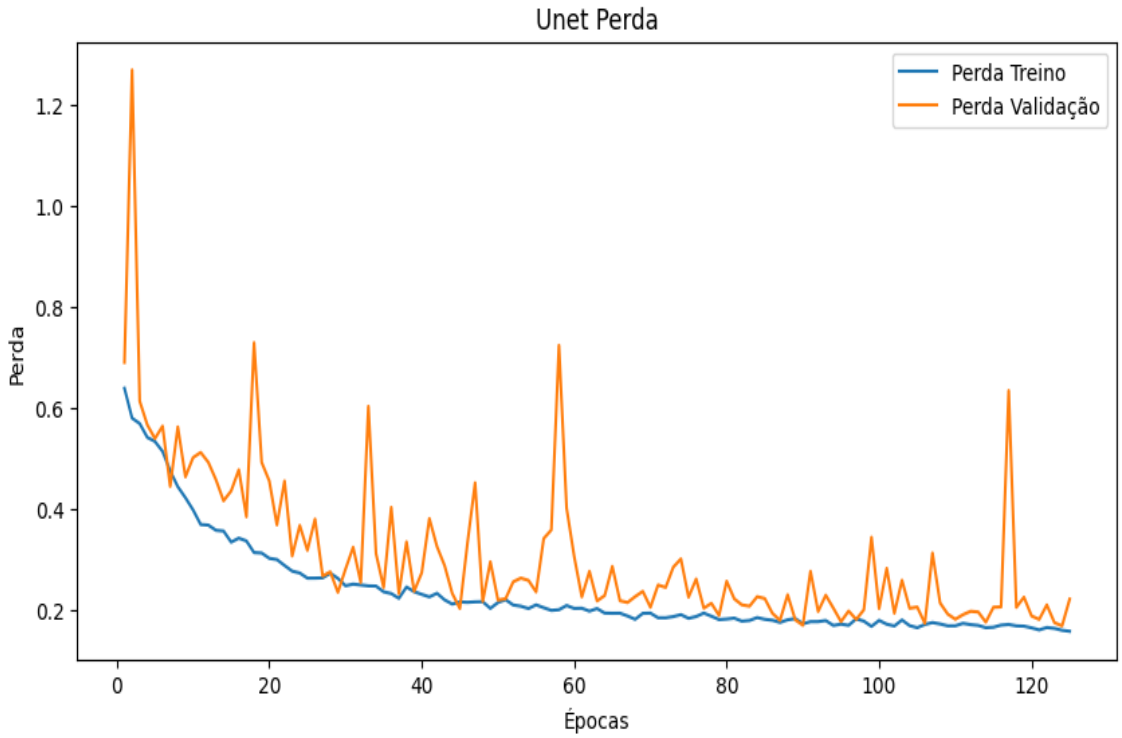

Figura 17: Gráfico de perda da U-Net. Fonte: Autoria Própria

3. **Avaliação do Desempenho:** Durante o treinamento, foram monitoradas métricas de desempenho, como mAP50, mAP50-95 e funções de perda. Diferentemente da U-Net e FPN-Net, aqui foi utilizada a DFL (Distribution Focal Loss) e a perda de segmentação.

Durante o treinamento do Yolo, observa-se que ele leva consideravelmente mais tempo para aprender do que os outros modelos, conforme indicado pela função de perda de segmentação. No entanto, pela função de perda DFL, é possível notar que ele poderia ter uma queda maior, como ilustrado na Figura 22.

3.6 Experimentos

Os experimentos foram divididos em dois estudos. O primeiro estudo aborda apenas os modelos de rede neural, utilizando as bases de dados 38-cloud e 95-cloud. O segundo experimento consiste na reprodução dos experimentos propostos por Mohajerani e Saeedi (2019).

No primeiro experimento, os modelos treinados na base de dados 38-cloud serão testados na base 95-cloud com as modificações de remoção de pixels nulos propostas na base de dados. Para esta avaliação, devido a considerações de custo computacional, optou-se por utilizar uma amostra representativa, composta por 30% dos dados da base 95-cloud escolhidas aleatoriamente com uma semente para utilizar as mesmas imagens.

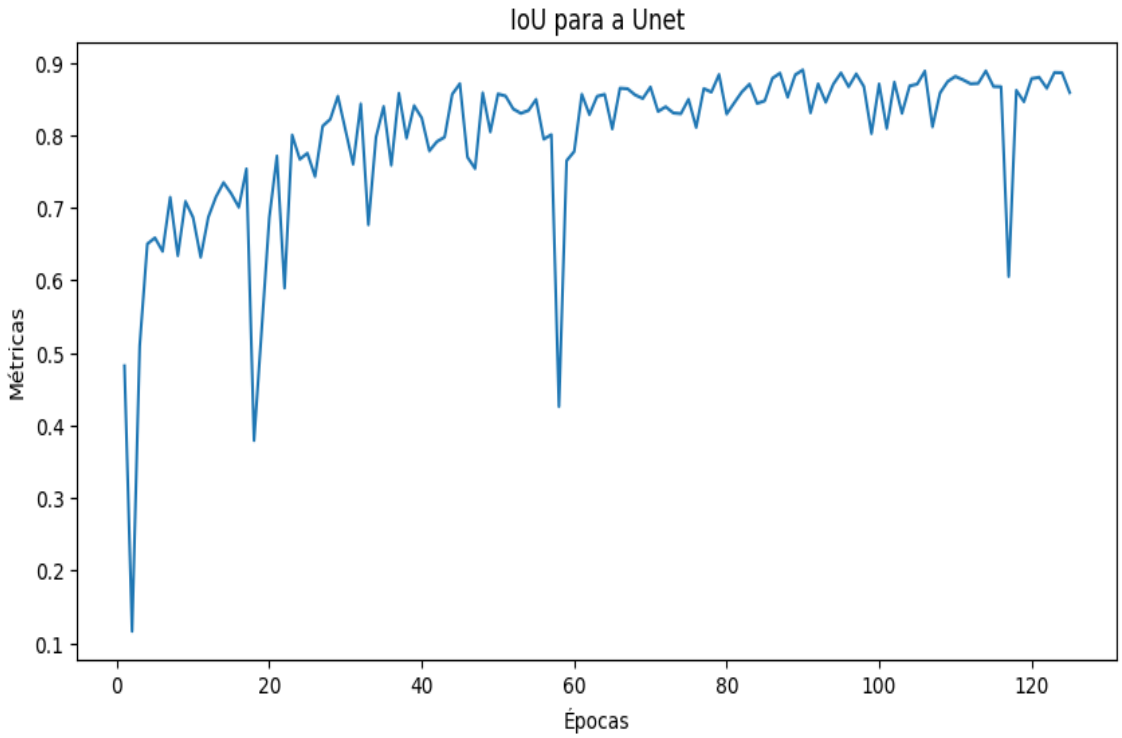

Figura 18: IoU para a U-Net. Fonte: Autoria Própria

O principal objetivo é analisar a capacidade de generalização dos modelos e verificar sua performance em relação à métrica de IoU, fundamental para determinar a concordância entre as máscaras de segmentação produzidas e as máscaras verdadeiras. Este experimento permitirá comparar os modelos de aprendizado profundo entre si.

Para o segundo experimento, utilizando os dados de teste fornecidos por Mohajerani e Saeedi (2019), será aplicado o mesmo sistema. Primeiramente, a imagem é dividida em vários *patches* não sobrepostos de 384 × 384 pixels. Em seguida, cada *patch* é redimensionado para 192 pixels (para se adequar à entrada da rede neural). Posteriormente, é redimensionado novamente para 384 × 384 pixels. Uma vez que as máscaras previstas de todos os *patches* são produzidas, elas são unidas para criar a máscara final de nuvem da imagem de teste original. (MOHAJERANI; SAEEDI, 2019)

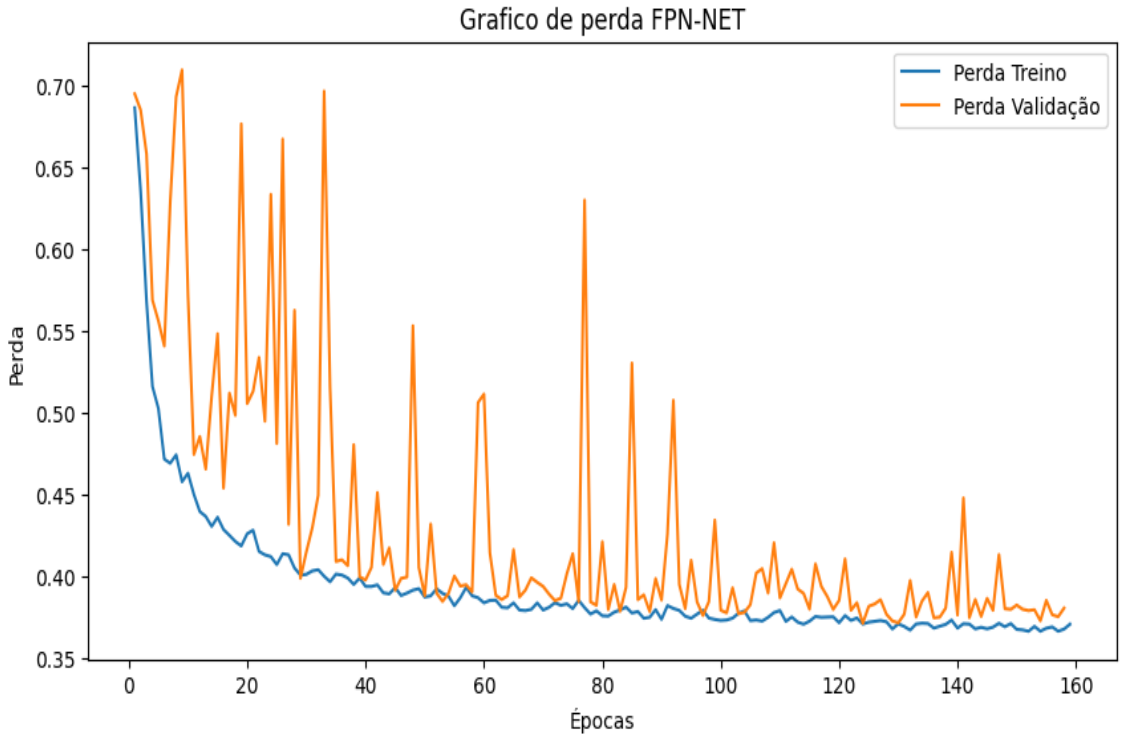

Figura 19: Gráfico de perda da FPN-Net. Fonte: Autoria Própria

4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Nesta seção, serão apresentados e analisados os resultados dos experimentos, que tiveram como objetivo avaliar a eficácia de diferentes modelos de segmentação semântica aplicados às bases de dados 38-cloud e 95-cloud. Os modelos foram testados quanto à sua capacidade de segmentar nuvens com base em imagens de satélite, utilizando métricas como o IoU, Precisão, Revocação e Acurácia Geral.

4.1 Avaliação dos Modelos

Na Tabela 1, apresentam-se os resultados obtidos pelos diferentes modelos avaliados. Destaca-se que o modelo Cloud-Net dentre os modelos de aprendizado profundo obteve o melhor desempenho, com um IoU de 73.99%, seguido pelo FPN-Net com 68.75%, U-Net com 65.04%, e YOLO com 60.91%. Esses resultados sugerem que o treinamento na base de dados 38-cloud foi capaz de gerar um modelo Cloud-Net eficiente na segmentação semântica de nuvens.

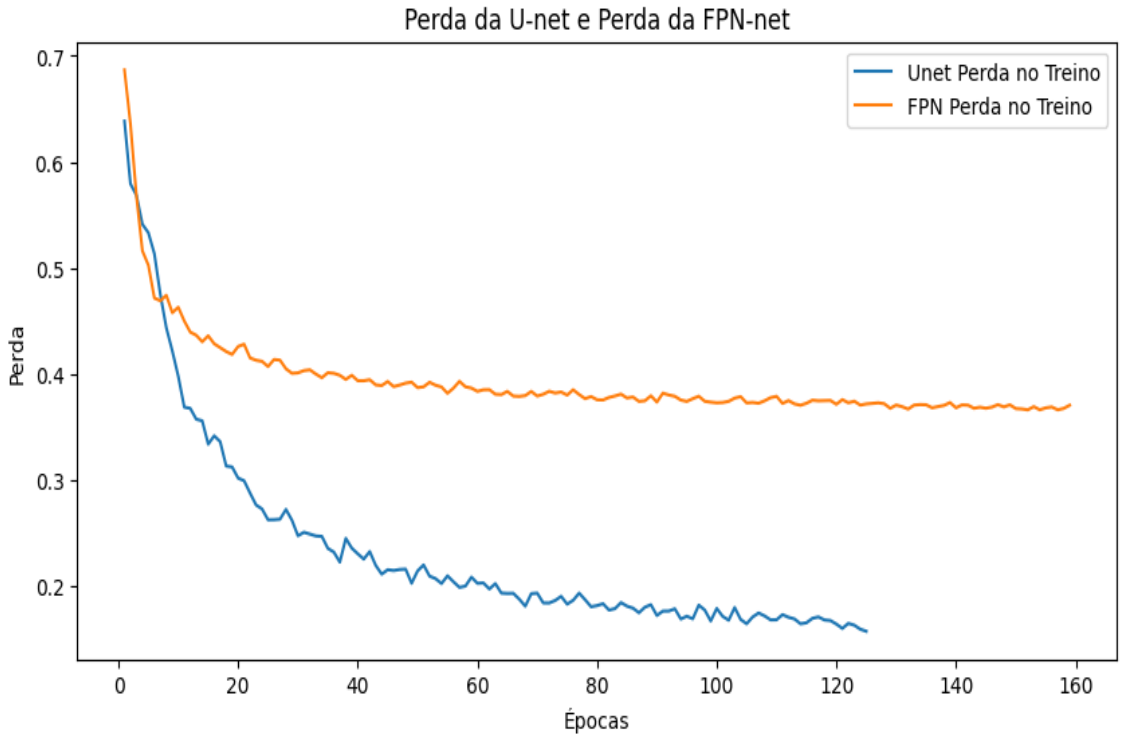

Figura 20: Perda da U-Net e Perda da FPN-Net. Fonte: Autoria Própria

Modelo	IoU (%)	Precisão (%)	Revocação (%)	Acurácia (%)	N. Parametros
YOLO	60.91	91.69	67.31	75.77	3.263.811
U-Net	65.04	82.0	75.83	82.47	537.366
FPN-Net	68.75	84.10	78.99	84.47	11.699.724
Cloud-Net	**73.99**	77.64	78.52	**97.73**	36.465.793

Tabela 1: Resultados do Experimento 1. Fonte: Autoria Própria

Na Tabela 2, são apresentados os resultados dos modelos nos testes propostos. Os resultados deste experimento para a Cloud-net e Fmask foram extraídos de Mohajerani e Saeedi (2019), visto que o Fmask requer mais bandas além das RGB e NIR (infravermelho próximo), as quais não estavam disponíveis para teste no conjunto de dados 38-cloud. A Figura 24 exibe uma imagem gerada pela Cloud-net à esquerda, enquanto à direita está a verdade terrestre dos testes do experimento 2, com um IoU de 91.68.

Modelo	IoU (%)	Precisão (%)	Revocação (%)	Acurácia (%)
YOLO	60.72	88.26	69.34	82.99
U-Net	65.63	83.0	75.64	89.00
FPN-Net	69.48	85.05	79.00	90.47
Fmask	75.16	77.71	97.22	94.89
Cloud-Net	**78.50**	91.23	84.85	**96.48**

Tabela 2: Resultados do Experimento 2. Fonte: Autoria Própria

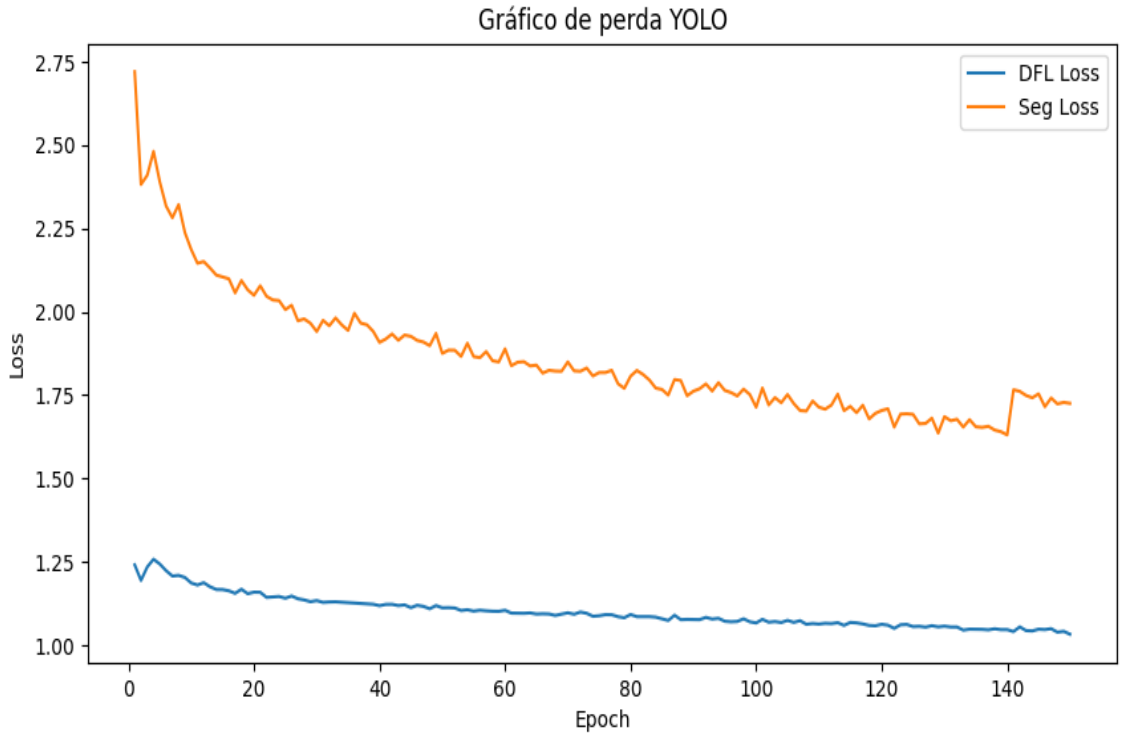

Figura 21: Gráfico de perda do Yolo. Fonte: Autoria Própria

Além dos resultados quantitativos, é importante destacar algumas observações qualitativas. A Figura 23 mostra exemplos de predições dos modelos em imagens da base 38-cloud. É possível observar que, mesmo apresentando uma media de IoU mais baixo nos testes, a YOLO tende a se assemelhar mais à paisagem terrestre em alguns casos, embora ocasionalmente falhe na segmentação precisa de algumas nuvens. Neste exemplo a Yolo se sai melhor que alguns modelos com um IoU para esta imagem de 72.64%, a U-net e Fpn-net apresentaram um IoU de 0.67 e 0.55 respectivamente, a Cloud-net teve um IoU de 91.90

4.2 Análise dos Resultados

A análise dos resultados revela ideias importantes sobre o desempenho dos diferentes modelos de segmentação. Utilizando o IoU como métrica principal na Tabela 1, mostra que o modelo Cloud-Net obteve a melhor pontuação, seguido pela Fmask, FPN-Net, U-Net e YOLO, respectivamente. Esses resultados sugerem que o treinamento na base de dados 38-cloud permitiu ao modelo Cloud-Net uma excelente capacidade de generalização para a base 95-cloud, indicando eficácia na segmentação semântica de nuvens.

Outro aspecto relevante é a discrepância no processo de treinamento entre os modelos. Enquanto a U-Net e a FPN-Net demonstraram treinamentos mais consistentes e semelhantes, a YOLO exigiu uma etapa adicional durante o treinamento, o que pode ter introduzido erros adicionais por ter que utilizar uma função para encontrar contornos do

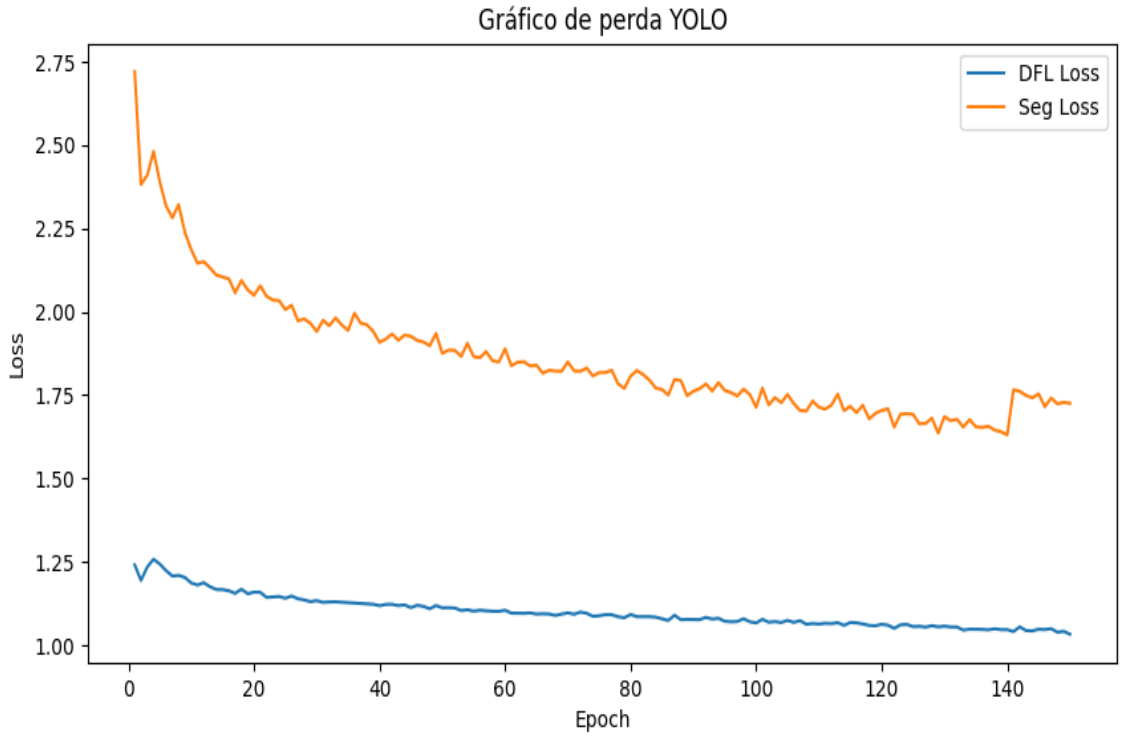

Figura 22: Grafico de perda da Yolo. Fonte: Autoria Própria

OpenCV que pode não ter encontrando algumas nuvens na verdade terrestre. Isso destaca a importância de selecionar cuidadosamente o modelo mais adequado para a tarefa específica, considerando não apenas a precisão final, mas também a eficiência, tempo de processamento e confiabilidade durante o treinamento.

É crucial considerar a eficiência computacional dos modelos avaliados. Um aspecto importante a ser destacado é o número de parâmetros de cada modelo, pois isso está diretamente relacionado aos recursos computacionais necessários para treinamento e inferência.

Ao analisar os números de parâmetros dos modelos avaliados, observa-se que a Cloud-Net possui um número significativamente maior de parâmetros em comparação com os outros modelos, totalizando 36.465.793 parâmetros. Em contraste, a U-Net, a FPN-Net e a YOLO oferecem alternativas mais leves, com 537.366, 11.699.724 e 3.263.811 parâmetros, respectivamente. Essa discrepância na complexidade dos modelos tem implicações diretas na eficiência computacional, especialmente em cenários onde os recursos são limitados.

A escolha do modelo ideal não deve se basear apenas no desempenho em termos de métricas de avaliação, mas também na eficiência computacional. Modelos mais leves, como a U-Net, a FPN-Net e a YOLO, podem ser mais adequados em cenários onde os recursos computacionais são limitados, enquanto a Cloud-Net, apesar de sua eficácia, pode exigir um investimento computacional substancial.

Essa relação entre desempenho e complexidade computacional fornece ideias valiosas para a seleção e implementação de modelos em aplicações práticas, onde a eficiência é uma consideração crucial.

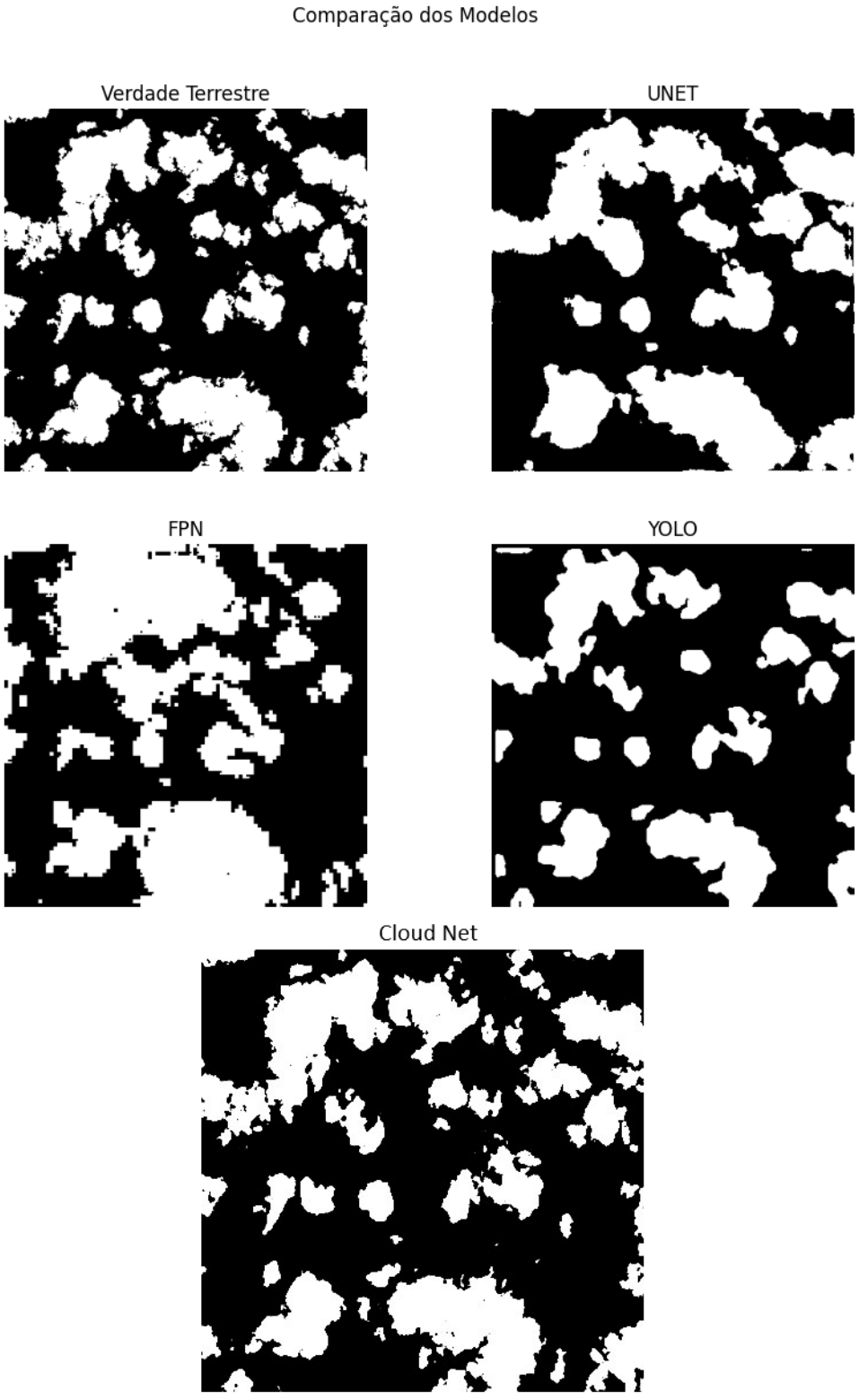

Figura 23: Exemplo de predições dos modelos em imagem 38-Cloud. Fonte: Autoria Propria

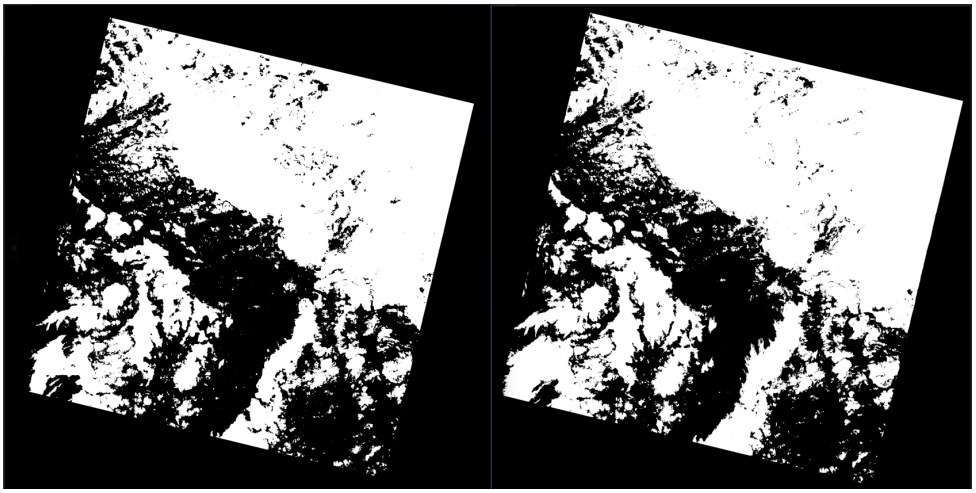

Figura 24: Exemplo de cena predita do experimento 2, a esquerda a Cloud-net e a direita a verdade terrestre. Fonte: Autoria Própria

5 CONCLUSÕES E TRABALHOS FUTUROS

Com base nos resultados do experimento e na análise detalhada dos diferentes modelos de segmentação semântica de nuvens, emerge uma compreensão mais ampla sobre a relação entre desempenho e complexidade computacional. A distinção entre os modelos avaliados destaca não apenas suas pontuações de IoU e outras métricas de avaliação, mas também a importância crítica de considerar a eficiência computacional na escolha do modelo ideal.

Enquanto a Cloud-Net se destaca com seu desempenho superior de IoU, é essencial reconhecer que sua eficácia vem acompanhada de uma complexidade computacional significativa, representada por um número substancialmente maior de parâmetros em comparação com outros modelos avaliados. Em contraste, modelos mais leves, como a U-Net, a FPN-Net, a YOLO e ate modelos baseados em limiar como a Fmask, oferecem alternativas viáveis, especialmente em cenários onde os recursos computacionais são limitados.

Essa análise detalhada proporciona uma compreensão mais abrangente das considerações práticas envolvidas na seleção e otimização de modelos de segmentação semântica em aplicações relacionadas à análise de imagens de satélite. Ao equilibrar o desempenho do modelo com sua eficiência computacional, os profissionais podem tomar decisões informadas que garantam a implementação bem-sucedida e eficaz desses modelos em ambientes do mundo real. Essas considerações são fundamentais para maximizar o desempenho e a eficiência dos sistemas de segmentação semântica em cenários práticos.

Além disso, este estudo aponta para possíveis direções futuras de pesquisa, como a exploração de técnicas de otimização e ajuste fino dos modelos para melhorar ainda mais seu desempenho em segmentação de nuvens, bem como a investigação de abordagens híbridas que combinem as vantagens de diferentes modelos para alcançar resultados ainda mais robustos e precisos. Outro ponto a ser estudado será a exploração de modelos para outros satélites como o sentinel-2 com uma base como cloudsen-12 onde tem imagem de todas as bandas do sentinel-2, e segmentar não apenas as nuvens mas as sombras de nuvens, e explorar mais as técnicas baseadas em limiar.

Portanto, para aplicações relacionadas à análise de imagens de satélite, é fundamental considerar cuidadosamente as características específicas de cada modelo e sua adequação para a tarefa em questão, visando maximizar tanto o desempenho quanto a eficiência em cenários práticos.

REFERÊNCIAS

ALI, S. Z. Principles of yolov8. **Medium**, Oct 2023. Disponível em: <https://medium.com/@syedzahidali969/principles-of-yolov8-6a90564e16c3>.

CECCON, D. **Fundamentos de ML: funções de custo para problemas de classificação (I)**. 2019. Acesso em 31 de março de 2024. Disponível em: <https://iaexpert.academy/2019/08/26/fundamentos-de-ml-funcoes-de-custo-para-problemas-de-classificacao-i/>.

HE, K.; ZHANG, X.; REN, S.; SUN, J. **Deep Residual Learning for Image Recognition**. 2015.

IBM. **O que é segmentação semântica?** 2024. <https://www.ibm.com/br-pt/topics/semantic-segmentation>. Acesso em: 07 Fev. 2024.

ICHIPRO. **Segmentação de nuvens em imagens de satélite usando aprendizado profundo**. 2024. Acessado em: 23 de Fevereiro de 2024. Disponível em: <https://ichi.pro/pt/segmentacao-de-nuvens-em-imagens-de-satelite-usando-aprendizado-profundo-186871578273853>.

IRISH, R. R.; BARKER, J. L.; GOWARD, S. N.; ARVIDSON, T. Characterization of the landsat-7 etm automated cloud-cover assessment (acca) algorithm. **Photogrammetric Engineering and Remote Sensing**, v. 72, p. 1179–1188, 2006. Disponível em: <https://api.semanticscholar.org/CorpusID:14451366>.

JIN, X.; LI, J.; SCHMIT, T.; LI, J.; GOLDBERG, M.; GURKA, J. Retrieving clear-sky atmospheric parameters from seviri and abi infrared radiances. **Journal of Geophysical Research**, v. 113, 08 2008.

JOCHER, G.; CHAURASIA, A.; QIU, J. **Ultralytics YOLOv8**. 2023. Disponível em: <https://github.com/ultralytics/ultralytics>.

LAKSHMANAN, V.; GöRNER, M.; GILLARD, R. **Practical Machine Learning for Computer Vision**. [S.l.]: O'Reilly Media, Inc., 2021. ISBN 9781098102364.

LI, F.-F. **Lecture 11: Detection and Segmentation**. 2024. <http://cs231n.stanford.edu/slides/2017/cs231n_2017_lecture11.pdf>. Acesso em: 07 Fev. 2024.

LI, X.; WANG, W.; WU, L.; CHEN, S.; HU, X.; LI, J.; TANG, J.; YANG, J. **Generalized Focal Loss: Learning Qualified and Distributed Bounding Boxes for Dense Object Detection**. 2020.

LIN, T.-Y.; DOLLáR, P.; GIRSHICK, R.; HE, K.; HARIHARAN, B.; BELONGIE, S. **Feature Pyramid Networks for Object Detection**. 2017.

MOHAJERANI, S. **38-Cloud-A-Cloud-Segmentation-Dataset**. 2019. Disponível em: <https://github.com/SorourMo/38-Cloud-A-Cloud-Segmentation-Dataset>.

_____. **95-Cloud-An-Extension-to-38-Cloud-Dataset**. 2019. Disponível em: <https://github.com/SorourMo/95-Cloud-An-Extension-to-38-Cloud-Dataset>.

_____. **Cloud-Net: A semantic segmentation CNN for cloud detection**. 2019. Disponível em: <https://github.com/SorourMo/Cloud-Net-A-semantic-segmentation-CNN-for-cloud-detection/tree/master>.

MOHAJERANI, S.; KRAMMER, T. A.; SAEEDI, P. A cloud detection algorithm for remote sensing images using fully convolutional neural networks. In: **2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP)**. [S.l.: s.n.], 2018. p. 1–5.

_____. **Cloud Detection Algorithm for Remote Sensing Images Using Fully Convolutional Neural Networks**. 2018.

_____. A cloud detection algorithm for remote sensing images using fully convolutional neural networks. In: **2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP)**. [S.l.: s.n.], 2018. p. 1–5.

MOHAJERANI, S.; SAEEDI, P. **Cloud-Net: An end-to-end Cloud Detection Algorithm for Landsat 8 Imagery**. 2019.

OLIVEIRA, P. d. **Deep Learning na Segmentação Automática de Imagens de Satélite**. Tese (Trabalho de Conclusão de Curso (Engenharia Agrícola e Ambiental)) — Universidade Federal de Viçosa, Viçosa, 2020.

QIU, S.; ZHU, Z.; HE, B. Fmask 4.0: Improved cloud and cloud shadow detection in landsats 4-8 and sentinel-2 imagery. **Remote Sensing of Environment**, v. 231, p. 111205, 09 2019.

REDDY, R. S.; LU, D.; TULURI, F.; FADAVI, M. Simulation and prediction of hurricane lili during landfall over the central gulf states using mm5 modeling system and satellite data. In: **2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)**. [S.l.: s.n.], 2017. p. 36–39.

RONNEBERGER, O.; FISCHER, P.; BROX, T. **U-Net: Convolutional Networks for Biomedical Image Segmentation**. 2015.

SHI, X.; GAO, Z.; LAUSEN, L.; WANG, H.; YEUNG, D.-Y.; WONG, W. kin; WOO, W. chun. **Deep Learning for Precipitation Nowcasting: A Benchmark and A New Model**. 2017.

TAYEBI, A.; KASMAEEYAZDI, S.; TINTI, F.; BRUNO, R. Contributions from experimental geostatistical analyses for solving the cloud-cover problem in remote sensing data. **International Journal of Applied Earth Observation and Geoinformation**, v. 118, p. 103236, 2023. ISSN 1569-8432. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1569843223000584>.

THISANKE, H.; DESHAN, C.; CHAMITH, K.; SENEVIRATNE, S.; VIDANAARACHCHI, R.; HERATH, D. **Semantic Segmentation using Vision Transformers: A survey**. 2023.

WANGENHEIM, A. von. Deep learning::segmentação con cnns. 2018. Acessado em: 14 de Maio de 2024. Disponível em: <https://lapix.ufsc.br/ensino/visao/visao-computacionaldeep-learning/deep-learningsegmentacao-semantica/>.

XIE, F.; SHI, M.; SHI, Z.; YIN, J.; ZHAO, D. Multilevel cloud detection in remote sensing images based on deep learning. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 10, n. 8, p. 3631–3640, 2017.

YUAN, Y.; HU, X. Bag-of-words and object-based classification for cloud extraction from satellite imagery. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 8, n. 8, p. 4197–4205, 2015.

ZHANG, Y.; GUINDON, B.; CIHLAR, J. An image transform to characterize and compensate for spatial variations in thin cloud contamination of landsat images. **Remote Sensing of Environment**, v. 82, n. 2, p. 173–187, 2002. ISSN 0034-4257. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0034425702000342>.

ZHU, Z.; WANG, S.; WOODCOCK, C. Improvement and expansion of the fmask algorithm: Cloud, cloud shadow, and snow detection for landsats 4-7, 8, and sentinel 2 images. **Remote Sensing of Environment**, v. 159, 01 2015.

ZHU, Z.; WOODCOCK, C. Object-based cloud and cloud shadow detection in landsat imagery. **Remote Sensing of Environment**, v. 118, p. 83–94, 03 2012.