



UNIVERSIDADE FEDERAL DO MARANHÃO
Curso de Graduação em Ciência da Computação

Pedro Thiago Cutrim dos Santos

**Aprimoramento da Detecção de Áreas de
Garimpo na Região do Tapajós através de
Redes Adversárias de Super-Resolução**

São Luís - MA

2021

Pedro Thiago Cutrim dos Santos

Aprimoramento da Detecção de Áreas de Garimpo na Região do Tapajós através de Redes Adversárias de Super-Resolução

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão, como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Graduação em Ciência da Computação

Universidade Federal do Maranhão

Orientador: Prof. Dr. Geraldo Braz Junior

São Luís - MA

2021

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

Cutrim dos Santos, Pedro Thiago.

Aprimoramento da Detecção de Áreas de Garimpo na Região do Tapajós através de Redes Adversárias de Super-Resolução / Pedro Thiago Cutrim dos Santos. - 2021.

57 f.

Orientador(a): Geraldo Braz Junior.

Monografia (Graduação) - Curso de Ciência da Computação, Universidade Federal do Maranhão, Universidade Federal do Maranhão, 2021.

1. Garimpo. 2. Landsat 8. 3. Redes Adversárias de Super-resolução. 4. Sensoriamento Remoto. 5. U-Net. I. Braz Junior, Geraldo. II. Título.

Pedro Thiago Cutrim dos Santos

Aprimoramento da Detecção de Áreas de Garimpo na Região do Tapajós através de Redes Adversárias de Super-Resolução

Monografia apresentada ao curso de Ciência da Computação da Universidade Federal do Maranhão, como parte dos requisitos necessários para obtenção do grau de Bacharel em Ciência da Computação.

Trabalho _____. São Luís - MA, 30 de abril de 2021:

Prof. Dr. Geraldo Braz Junior
Orientador
Universidade Federal do Maranhão

**Prof. Dr. Darlan Bruno Pontes
Quintanilha**
Universidade Federal do Maranhão

Prof. Dr. Tiago Bonini Borchartt
Universidade Federal do Maranhão

São Luís - MA
2021

Agradecimentos

Agradeço primeiramente as minhas tias, que sempre me deram todo o suporte e cuidado e são responsáveis pela pessoa que sou hoje.

Ao meu orientador, Geraldo Braz Junior, quem sempre me acompanhou e aconselhou em minha vida acadêmica e despertou meu encanto pela ciência.

A meus bons amigos, Matheus Boaro e Arthur Serra, por me darem suporte e contribuírem na revisão deste trabalho.

Também agradeço a minha melhor amiga Fernanda Ferreira e a meus amigos Venicius Garcia, Carlos Vinicios, Vitor França e Pedro Henrique por me ajudarem no momento difícil que eu me encontrava enquanto escrevia este trabalho.

Por fim, sou grato ao Núcleo de Computação Aplicada da Universidade Federal do Maranhão por dar todos os recursos necessários para a realização deste trabalho.

"Agora não há tempo para pensar o que você não tem. Pense no que pode fazer com o que tem."

(Ernest Hemingway)

Resumo

A mineração ilegal é um dos principais fatores que causam degradação ambiental no Brasil. A Região de Integração do Tapajós é responsável por mais de 60% do desmatamento causado pelo garimpo ilegal. Atualmente, a identificação de garimpos é feita usando imagens de satélite e outras técnicas de sensoriamento remoto, devido a seu difícil acesso na mata densa. Este trabalho propõe o uso de redes adversárias de super-resolução para o melhoramento de imagens de satélite de baixa resolução em conjunto com técnicas de aprendizado de máquina para detecção de regiões de garimpo. Para as redes de super-resolução são usadas a SRGAN e ESRGAN. Na etapa de detecção usa-se a rede convolucional U-Net. Foram obtidos ganhos de até 15% na detecção com o uso das imagens melhoradas, sendo o melhor resultado com os valores de 86% de acurácia, 80% de sensibilidade, 94% de precisão e 86% de *F1-Score*.

Palavras-chave: Redes Adversárias de Super-resolução, Sensoriamento Remoto, Garimpo, Landsat 8, RapidEye, U-Net, Segmentação, Detecção

Abstract

Illegal mining is one of the main factors that cause environmental degradation in Brazil. The Tapajós Integration Region is responsible for more than 60% of the deforestation caused by this type of mining. Currently, the identification of gold mining is done using satellite images and other remote sensing techniques, due to its difficult access in the dense forest. This work proposes the use of super-resolution generative adversarial networks to improve low-resolution satellite images with machine learning techniques for detecting gold mining regions. For super-resolution networks, SRGAN and ESRGAN are used. In the detection stage, we used the U-Net convolutional network. Gains of up to 15% in detection we go with the use of improved images, the best result being the values of 86% accuracy, 80% sensitivity, 94% accuracy and 86% F1-Score.

Keywords: Super-Resolution Generative Adversarial Networks, Remote Sensing, Gold Mining, Landsat 8, RapidEye, U-Net, Segmentation, Detection

Lista de ilustrações

Figura 1 – Região de Garimpo.	14
Figura 2 – Imagens de satélite sobre mesma área com resoluções espaciais diferentes (quanto menor a resolução nominal maior é a capacidade de resolução do satélite).	20
Figura 3 – Modelo de um neurônio artificial.	23
Figura 4 – Topologia de uma ANN.	24
Figura 5 – Exemplo de uma convolução. A região azul escura indica os neurônios da região filtrada. A região verde escura indica o valor de saída do neurônio para qual toda a ativação de entrada foi computada.	25
Figura 6 – Usando preenchimento com zeros para evitar que a dimensionalidade da imagem diminua e manter o seu tamanho original.	25
Figura 7 – Exemplo de convolução nas três primeiras seções da imagem usando preenchimento com zeros e um <i>stride</i> de 2 em conjunto com um filtro 3x3.	26
Figura 8 – Exemplo de <i>pooling</i> usando a técnica do maior valor em áreas 3x3 da imagem.	26
Figura 9 – Arquitetura da U-Net.	27
Figura 10 – Bloco residual.	28
Figura 11 – Funcionamento das GANs.	29
Figura 12 – Arquiteturas das redes geradora e discriminadora com seus correspondentes tamanho de filtros (<i>k</i>), quantidade de mapas de características (<i>n</i>) e tamanho de <i>stride</i> (<i>s</i>).	31
Figura 13 – Reconstrução da resolução da imagem em 4x usando a SRGAN e ESRGAN. Pode-se observar que a imagem da ESRGAN possui mais nitidez e detalhes do que a da SRGAN.	33
Figura 14 – Representação dos mapas de características antes e depois da ativação. Quanto mais profunda é a rede, mais informações são perdidas após a ativação, enquanto antes da ativação há características mais ricas.	34
Figura 15 – Metodologia do Estudo.	35
Figura 16 – Mapa da região de garimpo e outros limites regionais.	37
Figura 17 – Composição de falsa cor RGB nos satélites RapidEye e Landsat 8.	38
Figura 18 – Área antropisada. (a) <i>Aerial View (Bing Maps)</i> (MICROSOFT, 2021) (b) RapidEye e (c) Landsat 8.	38
Figura 19 – Marcação de um garimpo. (a) visualização em falsa cor RGB (b) validação visual e (c) marcação.	39
Figura 20 – Região de garimpo visualizada através de satélite. (a) Landsat 8 e (b) RapidEye.	40

Figura 21 – Base de dados convencional.	41
Figura 22 – Base de dados de transformação.	42
Figura 23 – Arquitetura da U-Net.	43
Figura 24 – Resultado do melhoramento de uma imagem Landsat utilizando os modelos treinados.	46
Figura 25 – Região de garimpo com presença de nuvens, nota-se que (c) consegue eliminar melhor a presença das nuvens. (a) Imagem Original, (b) ESRGAN (BS1) e (c) ESRGAN (BS2).	47
Figura 26 – Garimpo com segmentação difícil. Nota-se que os modelos BS2 se aproximaram melhor do <i>ground truth</i>	48
Figura 27 – Exemplo de garimpos detectados pela abordagem BS2 mas não por BS1. (a) e (b) garimpos parcialmente cortados pela imagem e (c) garimpo capturado com poucos detalhes pelo satélite.	50
Figura 28 – Imagem de Região de garimpo antes e depois da reconstrução (abordagem BS1) com super-resolução. (a) imagem original e (b) imagem de super-resolução.	50
Figura 29 – Fonte: Imagem de Região de garimpo antes e depois da transformação (abordagem BS2) com super-resolução. (a) imagem original Landsat 8 (b) imagem de super-resolução e (c) imagem de referencia RapidEye	51
Figura 30 – Fonte: Imagem de Região de garimpo com alta presença de nuvens antes e depois da transformação (abordagem BS2) com super-resolução. (a) imagem original e (b) imagem de super-resolução.	51

Lista de tabelas

Tabela 1 – Principais Divisões do Espectro Eletromagnético	20
Tabela 2 – Especificações Técnicas do Landsat 8	22
Tabela 3 – Especificações Técnicas do RapidEye	22
Tabela 4 – Panorama do Janelamento	40
Tabela 5 – Bases para treinamento de super-resolução.	45
Tabela 6 – Resultados da Segmentação.	47
Tabela 7 – Resultados da Detecção.	49

Lista de abreviaturas e siglas

ACC	<i>Acurácia</i>
ANN	<i>Rede Neural Artificial</i>
BN	<i>Camada de Normalização de Lotes</i>
CNN	<i>Rede Neural Convolutacional</i>
DICE	<i>Coefficiente de Dice</i>
ESRGAN	<i>Rede Adversária de Super-resolução Aprimorada</i>
FN	<i>Falso Negativo</i>
FP	<i>Falso Positivo</i>
GAN	<i>Rede Adversária Geradora</i>
MMA	<i>Ministério do Meio Ambiente</i>
MME	<i>Ministério de Minas e Energia</i>
MSE	<i>Erro Quadrático Médio</i>
PNSR	<i>Relação Sinal-ruído de Pico</i>
PREC	<i>Precisão</i>
RAISG	<i>Rede Amazônica de Informação Socioambiental</i>
ResNet	<i>Redes Neurais Residuais Profundas</i>
RRDB	<i>Blocos Densos Residual-in-Residual</i>
SENS	<i>Sensibilidade</i>
SIGMINE	<i>Sistema de Informação Geográfica de Mineração</i>
SPEC	<i>Especificidade</i>
SRGAN	<i>Rede Adversária de Super-resolução</i>
USGS	<i>Serviço Geológico dos Estados Unidos</i>
VN	<i>Verdadeiro Negativo</i>
VP	<i>Verdadeiro Positivo</i>

Sumário

1	INTRODUÇÃO	14
1.1	Objetivos	15
1.1.1	Objetivos Específicos	16
1.2	Organização do Trabalho	16
2	TRABALHOS RELACIONADOS	17
3	FUNDAMENTAÇÃO TEÓRICA	19
3.1	Sensoriamento Remoto	19
3.2	Satélite Landsat 8	21
3.3	Satélites RapidEye	21
3.4	Redes Neurais Artificiais	22
3.5	Redes Neurais Convolucionais	24
3.6	U-Net	26
3.7	Redes Neurais Residuais Profundas	28
3.8	Redes Adversárias Geradoras	28
3.9	GANs de Super-resolução	30
3.9.1	Arquitetura	30
3.9.2	Função de Custo Perceptual	31
3.10	GANs de Super-resolução Aprimoradas	32
3.10.1	Arquitetura da Rede Geradora	32
3.10.2	Arquitetura da Rede Discriminadora	33
3.10.3	Função de Custo Perceptual	33
4	METODOLOGIA	35
4.1	Aquisição de Dados	35
4.2	Pré-processamento dos Dados	36
4.2.1	Reorganização de Bandas	37
4.2.2	Marcação dos Garimpos	37
4.2.3	Janelamento de Imagens e Conversão dos Dados	39
4.3	Melhoramento de Resolução	40
4.3.1	Base de Dados Convencional	41
4.3.2	Treinamento de Transformação	41
4.4	Segmentação e Detecção	42
4.4.1	Avaliação	43
4.5	Considerações finais	44

5	RESULTADOS	45
5.1	Super-Resolução	45
5.2	Segmentação	46
5.3	Detecção	48
5.4	Estudos de Caso	49
5.5	Considerações Finais	52
6	CONCLUSÃO	53
	REFERÊNCIAS	55

1 Introdução

A mineração ilegal é um dos principais fatores que causam degradação ambiental no Brasil. Estima-se que entre janeiro de 2020 e abril de 2021, cerca de 110 km^2 foram desmatados na Amazônia Legal devido a atividades de mineração, sendo a Região de Integração do Tapajós responsável por mais de 60% desse desmatamento (ASSIS et al., 2019). A atividade de mineração irregular mais notável é a do garimpo ilegal.

O garimpo consiste na extração, de forma industrial ou manual, de metais e pedras preciosas da natureza. O metal precioso mais explorado em garimpos é o ouro. A garimpagem pode ser realizada em terra firme ou nos leitos dos cursos d'água. Em terra firme normalmente ocorre o desmonte de margens e encostas com fortes jatos d'água em conjunto com o fracionamento do minério lavrado utilizando moinhos de martelos e centrífuga, onde o solo é revolvido sem nenhum cuidado de preservá-lo. Utiliza-se o mercúrio para formar uma amálgama com as partículas de ouro presentes e isolá-los de outros elementos presentes (CAHETÉ, 2008). A Figura 1 mostra a aparência de uma região de garimpo.

Figura 1 – Região de Garimpo.



Fonte: Retirado de Greenpeace (2020)

Um dos principais impactos ambientais causados pelo garimpo é a poluição mercurial. O mercúrio metálico lançado no ambiente, é volátil, podendo ser oxidado e metilado resultando em uma substância mais tóxica que o mercúrio comum, o metil-mercúrio, podendo ser incorporado aos organismos vivos presentes como diferentes tipos peixes. O contato humano ou de animais com o mercúrio pode ocasionar em sérios danos a saúde, como atrofia muscular e redução das funções cognitivas (DE-PAULA; LAMAS-CORRÊA; TUTUNJI, 2006).

O garimpo ilegal também impacta a população indígena. Em 2020, cerca de 500 hectares foram desmatados por garimpeiros na Terra Indígena Yanomami. Comparado com anos anteriores, é um avanço de 30% no desmatamento causado diretamente pela prática de garimpo nessa terra (SOCIOAMBIENTAL, 2020). Qualquer atividade garimpeira realizada em território indígena por não-indígenas é estritamente proibida como previsto na Constituição federal (art. 231, § 7º) (CURI, 2007).

O impacto na população indígena é mais agravante com a pandemia do Sars-CoV-2. A presença de garimpeiros coloca em risco a saúde de comunidades indígenas isoladas, que são mais vulneráveis a enfermidades. Um eventual contato direto ou indireto, arrisca desencadear num trágico episódio de genocídio dessas comunidades (SOCIOAMBIENTAL, 2020). Até fevereiro de 2021, pelo menos 9 crianças morreram com sintomas do vírus em duas comunidades na Terra Indígena Yanomami.¹

As áreas de garimpos habitualmente se encontram em regiões de difícil acesso por serem cobertas pela mata densa, o que dificulta a fiscalização em solo (FERNANDES; FILHO, 2016). Por isso, para identificá-los são usadas técnicas de sensoriamento remoto, como imagens de satélite. Entretanto, seu uso não é perfeito, dado que a resolução espacial do satélite ou a presença de nuvens podem dificultar na identificação de garimpos menores (MINERACAO, 2020).

A identificação de áreas de garimpos é importante para que medidas legais e políticas públicas sejam tomadas. Este trabalho propõe o uso de redes adversárias geradores de super-resolução para aprimorar a qualidade de imagens de satélite com o objetivo de melhorar a detecção e segmentação de regiões de garimpos usando redes neurais convolucionais.

1.1 Objetivos

O objetivo deste trabalho consiste em desenvolver uma metodologia usando redes adversárias geradoras para melhoramento de imagens de satélite com o objetivo de melhorar o desempenho de segmentação e detecção de garimpos usando redes neurais profundas.

¹ <https://g1.globo.com/rr/roraima/noticia/2021/01/28/conselho-de-saude-indigena-relata-em-oficio-mortes-de-criancas-yanomami-com-sintomas-de-covid-em-roraima.ghtml>

1.1.1 Objetivos Específicos

Especificamente, este trabalho busca os seguintes objetivos aplicados ao problema de melhoramento em imagens de satélite e detecção de áreas de garimpo:

- Desenvolver uma metodologia usando redes adversárias geradoras de super-resolução para melhoramento de imagens de satélite de baixa resolução espacial.
- Analisar o impacto da resolução das imagens de satélite na segmentação semântica de áreas de garimpo usando redes neurais convolucionais.
- Analisar o impacto da resolução das imagens de satélite na detecção de áreas de garimpo usando redes neurais convolucionais.
- Contribuir com um método baseado em redes adversárias de super-resolução para aprimoramento de imagens de satélite em regiões de garimpo.

1.2 Organização do Trabalho

Este trabalho está organizado da seguinte maneira:

- O Capítulo 2 descreve trabalhos relacionados a super-resolução em sensoriamento remoto e o uso de super-resolução para aprimoramento de classificação usando redes neurais convolucionais.
- O Capítulo 3 mostra a fundamentação teórica necessária para o desenvolvimento do trabalho. Os conceitos apresentados são sensoriamento remoto, satélites, redes neurais convolucionais e redes adversárias geradoras de super-resolução.
- O Capítulo 4 demonstra as etapas da metodologia deste trabalho. São elas: aquisição dos dados, pré-processamento, melhoramento de resolução, segmentação e detecção.
- O Capítulo 5 exhibe os resultados deste estudo. São apresentados os resultados em relação ao melhoramento de imagem, segmentação e detecção. Também são mostrados alguns estudos de caso.
- O Capítulo 6 fala sobre as considerações finais deste trabalho. É mostrada qual a contribuição e os objetivos futuros.

2 Trabalhos Relacionados

A aplicação de redes adversárias de super-resolução para imagens de satélite já é um tema abordado pela literatura recente, assim como o uso de super-resolução para aprimorar a classificação de imagens usando aprendizado de máquina. Este capítulo irá mostrar e discutir os trabalhos relacionados ao tema deste estudo.

[Bosch, Gifford e Rodriguez \(2018\)](#) propôs o uso de redes adversárias de super-resolução em imagens aéreas e de satélite. A arquitetura da rede é baseada no uso de redes convolucionais densas (*DenseNets*) para melhorar as imagens em um fator 8. O uso de redes densas serve para obter características descritivas mais ricas dos elementos visuais presentes nas imagens. A avaliação dos resultados foi realizada em cima de 4 bases de dados diferentes com imagens de resolução de 256x256 e 344x344 píxeis. Todas as imagens mostraram uma boa relação sinal-ruído de pico (PSNR) com as imagens originais de alta-resolução.

Partindo do princípio em que características hierárquicas de uma imagem não compartilham uma equação de mapeamento unificada, [Jiang et al. \(2019\)](#) propõe um *framework* de mapeamento multinível com treinamento adversário para reconstrução de imagens de baixa resolução, chamado de MMGAM. Para a extração de características da imagem, são usados os *feature extraction and tuning block* (FETB). Seu funcionamento baseia-se na conexão de uma camada com suas subsequentes, para melhor inferência de características locais e globais. A base de dados consiste em imagens do satélite comercial chinês *Jilin-1* cortadas em blocos de 96x96 píxeis. Os resultados foram promissores, mostrando um PSNR acima de outras redes adversárias de super-resolução como a SRGAN ([LEDIG et al., 2017](#)).

Usando um modelo baseado em redes adversárias geradoras *Wassertein* (WGAN) e uma técnica de treinamento melhorada ([GULRAJANI et al., 2017](#)), o trabalho de [Luo, Lu e Jia \(2019\)](#) propõe o melhoramento de imagens de nuvens do satélite chinês *FY-3D*. As cenas do satélite são cortadas em blocos de imagens com tamanho de 256x256 píxeis (para servir como imagem de alta resolução) e depois para 64x64 píxeis (imagem de baixa resolução para reconstrução). A WGAN possui o treinamento mais estável devido a menor necessidade no ajuste de parâmetro, reduzindo o tempo de preparação e de treinamento das redes. A reconstrução da imagem de satélite usando a WGAN se mostrou superior às técnicas de interpolação do vizinho mais próximo e também da reconstrução usando redes neurais convolucionais.

Fora do tema de imagens de satélite, [Mahapatra et al. \(2017\)](#) propôs a utilização de super-resolução para melhorar a qualidade da segmentação de imagens de fundo de olho.

A base contém cerca de 5000 imagens geradas a partir de técnicas de aumento dos dados originais, como translação e rotação. As imagens originais possuem resolução de 1024x1024 píxeis, as imagens de baixa resolução são geradas usando um fator de redução de escala 4x e 8x. Usando a SRGAN, o trabalho conseguiu um aumento de até 11% na acurácia de segmentação quando comparado aos resultados das imagens de baixa resolução.

Ainda no tema de imagens médicas, [Tan et al. \(2021\)](#) propõe o uso de super-resolução para melhorar a qualidade de imagens de tomografia do tórax para classificar a ocorrência de pneumonia causada pelo coronavírus. A base conta com 470 imagens, sendo 275 delas com ocorrência de pneumonia. As imagens são melhoradas usando o modelo original pré-treinado da SRGAN e VGG16, após o melhoramento é realizada a classificação das imagens através de redes neurais convolucionais. Os resultados mostraram um melhoramento de até 13% na acurácia da classificação quando comparado com outro estudo usando a mesma base de dados.

Dentre os trabalhos que apresentam o uso de imagens de satélite, todos exibem resultados somente baseados em alguma métrica de qualidade da imagem, como o PSNR. Entretanto, não há avaliação do impacto do uso de super-resolução nos resultados de classificação/segmentação usando algoritmos de aprendizado de máquina, como abordam os trabalhos de [Mahapatra et al. \(2017\)](#) e [Tan et al. \(2021\)](#). Este estudo visa implementar uma metodologia baseada no uso de redes adversárias de super-resolução para melhoramento de imagens de satélite de baixa resolução e avaliar o impacto na segmentação e classificação de garimpos na região dos Tapajós situada na Amazônia Legal.

3 Fundamentação Teórica

Este capítulo apresenta o embasamento teórico do estudo. São abordados os conceitos de sensoriamento remoto, satélites, redes neurais convolucionais e redes adversárias geradoras de super-resolução.

3.1 Sensoriamento Remoto

Em todo processo científico é necessária uma etapa de coleta de dados/informações do objeto de estudo, tal coleta geralmente é feita *in situ* pelo próprio cientista utilizando algum dispositivo. Entretanto, ocasionalmente não é viável coletar os dados desta maneira, seja pela interferência causada pelo instrumento dentro da própria área de análise, pelos possíveis riscos da proximidade do objeto ou até pelo tamanho da área de estudo (JENSEN, 2007). Quando algum desses problemas são presentes, os dados podem ser coletados usando alguma técnica de sensoriamento remoto.

O sensoriamento remoto é definido como a medição ou aquisição da informação sobre as propriedades de algum objeto, ou fenômeno através de algum dispositivo com sensor que não esteja em contato físico e nem relativamente próximo do mesmo (JENSEN, 2007). A distância do objeto de estudo é relativa de acordo com a área específica em que se utiliza do sensoriamento remoto, em escala microscópica, por exemplo, essa distância pode ser de centímetros. Para o estudo, visa-se analisar áreas da superfície terrestre, estas informações são obtidas através do uso plataformas orbitais (satélites) ou sub-orbitais (aviões e drones).

Os dados coletados através do sensoriamento remoto são representações da quantidade de energia eletromagnética refletidas pelo objeto de estudo, em bandas ou frequências específicas. A quantidade de frequências diferentes do espectro eletromagnético que um instrumento de sensoriamento remoto pode capturar é chamada de *resolução espectral*, dispositivos que são capazes de observar múltiplas bandas são classificados como *multiespectrais*. A Tabela 1 mostra exemplos de frequências mais comumente utilizadas.

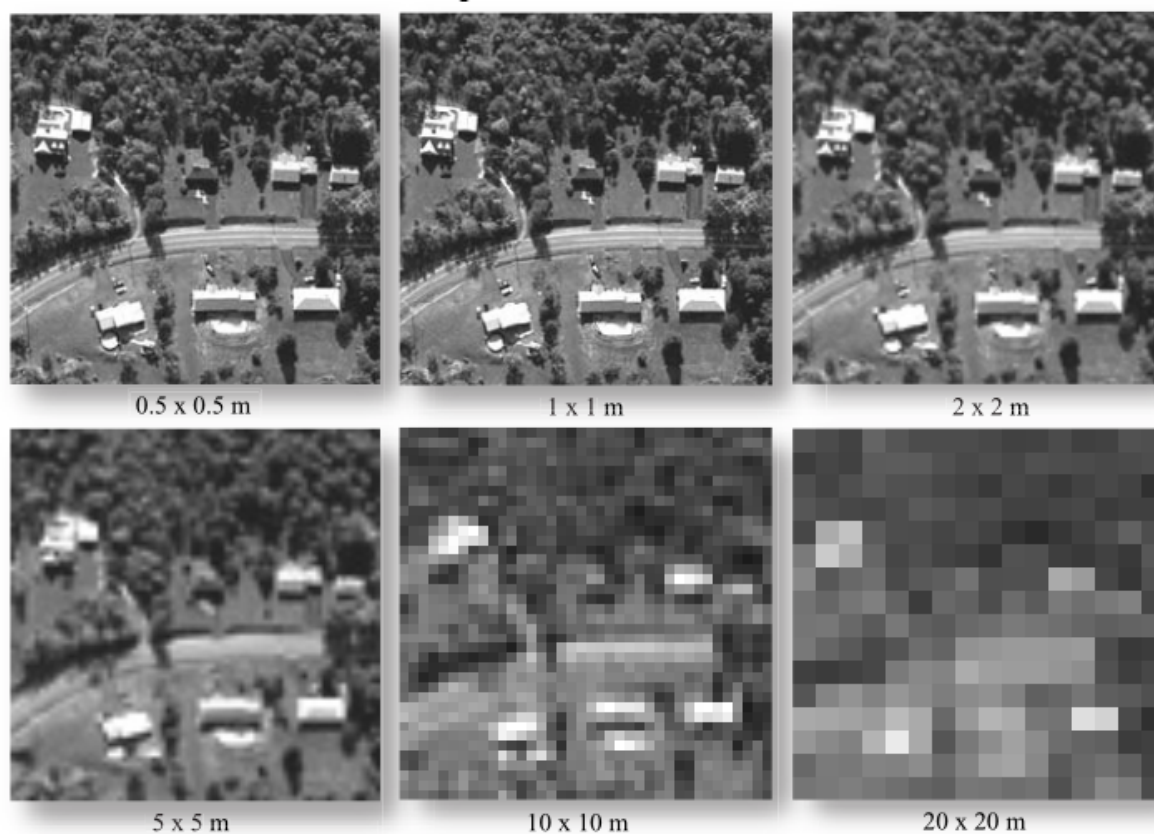
Além da quantidade de bandas que um instrumento pode capturar, outro fator importante é sua capacidade de detectar os objetos na superfície analisada, chamado de *resolução espacial*. A Figura 2 mostra alguns exemplos de imagens de satélite com resoluções espaciais diferentes, quanto menor a resolução nominal melhor é a capacidade do satélite de capturar objetos na superfície.

O intervalo entre as aquisições de uma mesma área por um satélite é chamado de *resolução temporal*, quanto menor o intervalo maior é a capacidade de análise na

Tabela 1 – Principais Divisões do Espectro Eletromagnético

Divisão	Limites
Raios Gamma	≥ 0.03 nm
Raios-X	0.03-300 nm
Radiação ultravioleta	0.3-0.38 μm
Luz visível	0.38-0.72 μm
Infravermelho Próximo	0.72-1.30 μm
Infravermelho Médio	1.30-3.00 μm
Infravermelho Distante	7.0-1,000 μm (1 mm)
Radiação Microondas	1 mm-30 cm
Radio	≤ 30 cm

Figura 2 – Imagens de satélite sobre mesma área com resoluções espaciais diferentes (quanto menor a resolução nominal maior é a capacidade de resolução do satélite).



Fonte: Retirado de Jensen (2007)

transformação de uma área específica da superfície terrestre. Uma baixa resolução temporal geralmente é necessária para casos como respostas emergenciais (queimadas, desastres ambientais) enquanto resoluções temporais mais altas são utilizadas em análises de clima, tempo, floresta e agricultura (NOVO, 2001).

Entre o meio espectral, as faixas mais comumente exploradas para fins do sensoriamento

remoto dos recursos naturais são: luz visível, infravermelho próximo, infravermelho médio e micro-ondas (NOVO, 2001). Para esse estudo são utilizados as faixas espectrais do infravermelho próximo e a luz visível.

3.2 Satélite Landsat 8

O Landsat 8 é um satélite de observação da terra que está sobre o controle do Serviço Geológico dos Estados Unidos (USGS) (US, 2020). É o oitavo da série de satélites do programa Landsat (*Landsat Data Continuity Mission*). Sua missão é de fornecer imagens visíveis e infravermelhas de alta qualidade em tempo hábil de todas as áreas costeiras e massas da Terra, atualizando continuamente um banco de dados Landsat existente (IHLEN, 2019). Os objetivos gerais de sua missão são:

- Fornecer continuidade de dados com os Landsat 4, 5 e 7;
- Oferecer cobertura terrestre repetitivamente a cada 16 dias;
- Criar e atualizar periodicamente um repositório global de imagens de terrenos iluminadas pela luz solar e substancialmente livre de nuvens.

Entre uma de suas principais características está o uso do sensor *Operational Land Imager* (OLI), que coleta dados de imagem para 9 bandas espectrais de ondas curtas em uma faixa de 190 km com uma resolução espacial de 30 metros para todas as bandas, exceto a banda pancromática que possui uma resolução espacial de 15 metros (IHLEN, 2019). Outro sensor presente é o *Thermal Infrared Sensor* (TIRS), que mede a temperatura da superfície da terra em duas bandas térmicas usando uma tecnologia que aplica a física quântica para detectar calor (NASA, 2012). A Tabela 2 mostra as especificações técnicas do satélite.

Neste estudo são utilizadas as bandas vermelho (B4), infravermelho próximo (B5) e verde (B3) que possuem resolução espacial de 30 metros e uma resolução temporal de 16 dias.

3.3 Satélites RapidEye

O RapidEye trata-se de uma constelação formada por 5 microsatélites multiespectrais, seu controle é mantido por uma empresa privada alemã. Sua proposta é oferecer uma fonte de dados contendo uma combinação excepcional de cobertura de grande área, intervalos de revisita frequentes, alta resolução e recursos multiespectrais (INPE, 2013).

Tabela 2 – Especificações Técnicas do Landsat 8

Sensor	Bandas Espectrais	Resolução Espectral	Resolução Espacial	Resolução Temporal
OLI	(B1) Azul Costeiro	0.43 - 0.45 μm	30 m	16 dias
	(B2) Azul	0.45 - 0.51 μm		
	(B3) Verde	0.53 - 0.59 μm		
	(B4) Vermelho	0.64 - 0.67 μm		
	(B5) Infravermelho Próximo	0.85 - 0.88 μm		
	(B6) Infravermelho Médio	1.57 - 1.65 μm		
	(B7) Infravermelho Médio	2.11 - 2.29 μm		
	(B8) Pancromática	0.50 - 0.68 μm	15 m	
	(B9) Cirrus	1.36 - 1.38 μm	30 m	
TIRS	(B10) Infravermelho Termal	10.6 - 11.19 μm	100 m	
	(B11) Infravermelho Termal	11.5 - 12.51 μm		

Devido a sua alta resolução espacial e temporal, o satélite RapidEye é usado em aplicações da indústria como agricultura, floresta, exploração de petróleo, cartografia e mineração (RAPIDEYE, 2013). A Tabela 3 mostra as especificações técnicas do satélite.

Tabela 3 – Especificações Técnicas do RapidEye

Bandas Espectrais	Resolução Espectral	Resolução Espacial	Resolução Temporal
(B1) Azul	440 - 510 nm	5 m	A cada 1 ou 5.5 dias
(B2) Verde	520 - 590 nm		
(B3) Vermelho	630 - 685 nm		
(B4) Vermelho (Borda)	690 - 730 nm		
(B5) Infravermelho Próximo	760 - 850 nm		

Para este estudo são utilizadas as bandas vermelho (B3), infravermelho próximo (B5) e verde (B2) de forma análoga as bandas retiradas do Landsat como foi mostrado na Seção 3.2.

3.4 Redes Neurais Artificiais

Uma Rede Neural Artificial (ANN) é inspirada na arquitetura de neurônios biológicos, por exemplo, o cérebro humano. Em uma ANN cada célula de neurônio pode realizar uma simples tarefa de resposta a um sinal de entrada, entretanto, quando há uma rede de neurônios conectados, estes podem executar tarefas mais complexas como reconhecimento de texto, fala ou imagem (ZOU; HAN; SO, 2008).

Toda ANN é composta por três componentes críticos:

- **Características dos Neurônios Artificiais:** determina a quantidade de entrada

e saída relacionada aos neurônios artificiais, como também os pesos e a função de ativação;

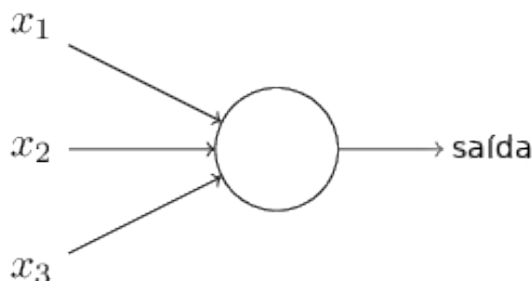
- **Topologia da Rede:** determina a forma em como os nós da rede são organizados e conectados;
- **Regras de Aprendizado:** determina com os pesos são inicializados e ajustados.

A Figura 3 mostra o modelo básico do neurônio artificial (também conhecido como nó) de uma ANN. Cada neurônio recebe múltiplos sinais de entradas uns dos outros através de conexões, estas associadas a pesos. Quando a soma ponderada dos sinais de entrada excedem um valor de limiar o nó é ativado e passa o seu sinal de ativação para os nós vizinhos (NIELSEN, 2015). Esse processo pode ser expresso matematicamente pela Equação 3.1:

$$y = f\left(\sum_{i=0}^n w_i x_i - T\right) \quad (3.1)$$

onde y é o valor de saída do nó, f é a função de ativação, w_i é o peso atrelado ao valor de entrada x_i e T é o limiar.

Figura 3 – Modelo de um neurônio artificial.

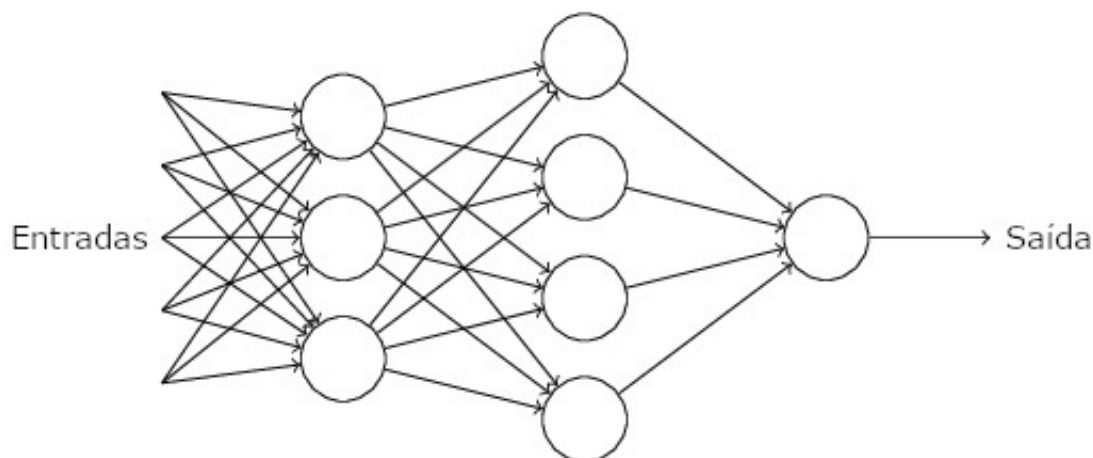


Fonte: Adaptado de Nielsen (2015)

Uma ANN possui seus nós organizados em camadas, que podem ser divididas entre camadas de entrada, ocultas e de saída. A topologia de uma rede envolve a determinação da quantidade de nós por camada, o número de camadas na rede e a forma em como os nós são conectados. Normalmente, uma ANN deve ter pelo menos uma camada de entrada e saída, o número de camadas ocultas é arbitrário. A Figura 4 mostra um exemplo de topologia de uma ANN.

Toda ANN utiliza alguma regra de aprendizado para treinar sua rede de neurônios artificiais, é no treinamento que os pesos w são ajustados para os valores desejados. Em um aprendizado supervisionado, um conjunto de dados é provido junto com seus respectivos rótulos. Os pesos w são ajustados de forma que o erro entre a saída da rede e o rótulo correto do dado seja minimizado.

Figura 4 – Topologia de uma ANN.



Fonte: Adaptado de [Nielsen \(2015\)](#)

Normalmente, um algoritmo de retropropagação ([RUMELHART; HINTON; WILLIAMS, 1986](#)) é utilizado no treinamento da maioria das ANNs. Inicialmente o dado de entrada é propagado pela rede e é gerado uma saída, e então através de uma função de custo calcula-se o erro entre a saída gerada e o rótulo correto que é propagado para trás desde a saída até a entrada para o ajuste dos pesos ([NIELSEN, 2015](#)).

A maioria das técnicas aplicadas neste estudo utilizam algum conceito da ANN.

3.5 Redes Neurais Convolucionais

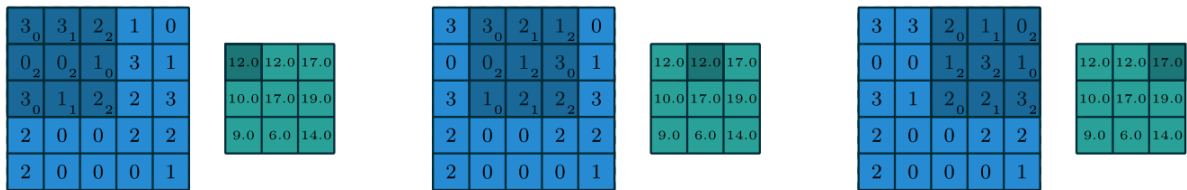
Na tarefa de classificação de imagens, as ANNs não possuem um desempenho tão bom quanto outros domínios de dados, por exemplo texto e voz. Isso se dá ao fato de que a arquitetura de uma ANN não considera a estrutura espacial das imagens ([NIELSEN, 2015](#)). [LeCun et al. \(1989\)](#) propôs então uma rede neural capaz de obter melhor desempenho na tarefa de classificação de imagens chamada de Rede Neural Convolutiva (CNN). Uma CNN é composta de três ideias básicas: campos receptivos locais, pesos compartilhados e *pooling*.

No contexto de uma ANN cada pixel de uma imagem é representado como um valor de entrada da rede e, sendo assim, todos possuem conexões com as camadas seguintes. No caso das CNNs, são extraídas ativações em campos da imagem ao invés de individualmente em cada valor de pixel, essa região analisada da imagem é chamada de campo receptivo local para um neurônio escondido ([NIELSEN, 2015](#)).

Considere uma imagem de entrada de tamanho 25x25 píxeis, para extrair informações estruturais locais são usados filtros de tamanho 5x5. Ao usar esse filtro como um campo

receptivo local o número de conexões da rede diminui em conjunto com os pesos, isso reduz o tempo computacional de treinamento e uso da rede. Além da otimização, o uso de filtros reduz a chance da sobre-ajuste (MURPHY, 2016). A Figura 5 mostra um exemplo de convolução nas primeiras três seções de uma imagem.

Figura 5 – Exemplo de uma convolução. A região azul escura indica os neurônios da região filtrada. A região verde escura indica o valor de saída do neurônio para qual toda a ativação de entrada foi computada.



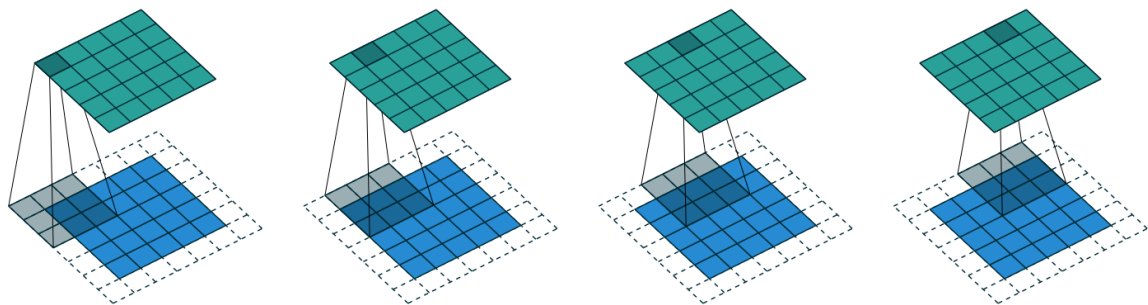
Fonte: Retirado de Murphy (2016)

A maneira de como o filtro é convolvido na imagem depende de vários parâmetros, são estes: o tamanho do filtro, o preenchimento e tamanho de *stride*.

O tamanho do filtro é comumente usado em formatos quadrangulares, este em conjunto com o tamanho da imagem determinam quais características podem ser detectadas pelo filtro.

O preenchimento é a quantidade de valores em zero que serão colocados ao redor da imagem. O preenchimento é usado para evitar que ocorra uma grande redução na dimensionalidade da imagem de entrada, já que um pouco da área da imagem é perdida a cada processo de convolução. A Figura 6 mostra como é realizado o preenchimento.

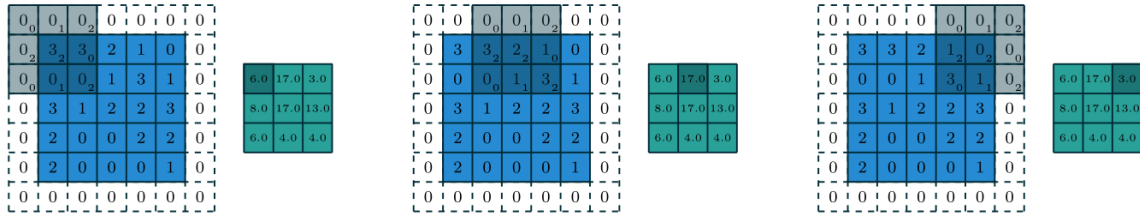
Figura 6 – Usando preenchimento com zeros para evitar que a dimensionalidade da imagem diminua e manter o seu tamanho original.



Fonte: Retirado de Murphy (2016)

O *stride* é a quantidade de pixels que o filtro irá se deslocar na convolução. A Figura 7 mostra um exemplo de *stride* aplicado em conjunto com preenchimento.

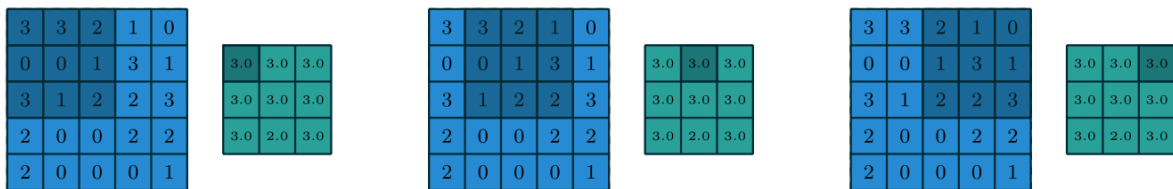
Figura 7 – Exemplo de convolução nas três primeiras seções da imagem usando preenchimento com zeros e um *stride* de 2 em conjunto com um filtro 3x3.



Fonte: Retirado de [Murphy \(2016\)](#)

Outra característica importante das CNNs é a presença das camadas de *pooling*. Estas camadas geralmente são postas logo após uma camada de convolução. O seu papel é simplificar a informação recebida da camada de convolução. Isso é feito simplificando a representação do vetor de características recebido. Esta simplificação geralmente é executada de duas formas: utilizando o maior valor de intensidade do pixel para representar a região analisada ou utilizar o valor médio ([NIELSEN, 2015](#)). A Figura demonstra como é realizada uma operação de *pooling*.

Figura 8 – Exemplo de *pooling* usando a técnica do maior valor em áreas 3x3 da imagem.



Fonte: Retirado de [Murphy \(2016\)](#)

Normalmente a saída de uma CNN é uma camada totalmente conectada a cada neurônio da camada anterior. A quantidade de classes presentes no problema são representadas pela quantidade de neurônios nessa camada ([NIELSEN, 2015](#)).

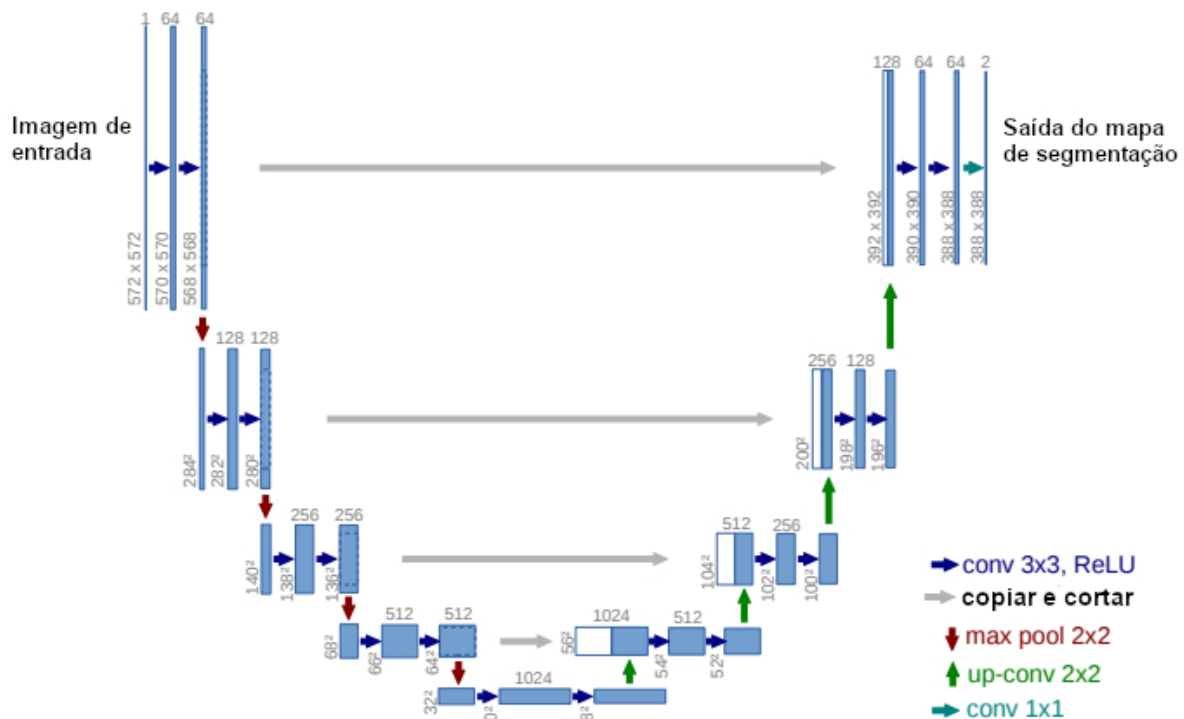
3.6 U-Net

Entre os fatores que influenciam no desempenho de uma CNN estão a sua profundidade e a quantidade de dados para treino disponível. Tal performance geralmente é avaliada através da capacidade da rede de classificar a imagem em um único rótulo correto. Entretanto, para alguns problemas, como na área médica, a segmentação da imagem é muito importante para dar um contexto de localização do que está sendo analisado. Pensando nisso, ([RONNEBERGER; FISCHER; BROX, 2015](#)) propôs a U-Net, uma CNN

capaz de realizar segmentação semântica em imagens de alta complexidade, especificamente focada na área médica.

A arquitetura da U-Net está ilustrada na Figura 9, em seu total existem 23 camadas convolucionais. Sua arquitetura é dividida entre uma fase de contração (no lado esquerdo) e uma fase de expansão (no lado direito). A fase de contração da rede possui uma arquitetura tipicamente igual a de uma rede convolucional, é formada por repetidas convoluções 3x3 (sem preenchimento) seguidas por uma ativação *ReLU* (XU et al., 2015) e uma camada de *maxpooling* 2x2 para realizar a contração da imagem. A cada contração, o número de mapas de características é duplicado.

Figura 9 – Arquitetura da U-Net.



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

No lado expansivo da rede, são usadas as operações chamadas de *upsampling* para expandir os mapas de características, estas são seguidas por convoluções 2x2 para diminuir pela metade a quantidade dos mapas. Como mostrado na Figura 9, nessas camadas são concatenadas cópias dos mapas de características extraídos na fase de contração da rede, isso é necessário devido à perda dos valores de píxeis na borda gerada pela convolução.

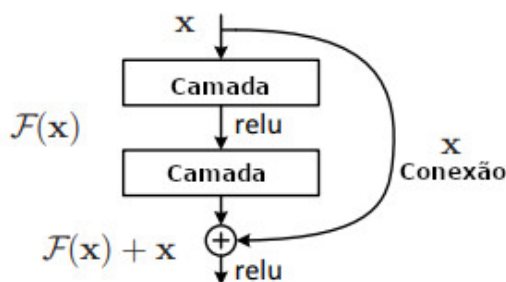
Por fim, a U-Net possui duas camadas de convolução 3x3 seguidas por ativações *ReLU*. Uma camada final de convolução 1x1 é usada para mapear cada um dos 64 mapas de características para o número desejado de classes do problema.

3.7 Redes Neurais Residuais Profundas

As redes neurais realizam a extração características de baixo, médio e alto nível em suas camadas. Em redes neurais muito profundas, a quantidade de camadas presentes pode enriquecer a qualidade dessas características extraídas, melhorando a performance. Entretanto, redes muito profundas são mais difíceis de treinar devido ao problema de dissipação do gradiente (BENGIO; SIMARD; FRASCONI, 1994). Visando esse problema, He et al. (2016) propôs as Redes Neurais Residuais Profundas (ResNet).

A principal característica da ResNet são os chamados *blocos residuais*, que usam uma espécie de curto-circuito para pular uma ou mais camadas da rede neural. Caso um grupo de camadas não consiga se adequar a um determinado mapeamento, o bloco residual adéqua estas explicitamente a um mapa residual, substituindo o mapeamento original por $F(x) + x$ (HE et al., 2016). A Figura 10 demonstra um bloco residual.

Figura 10 – Bloco residual.



Fonte: Adaptado de He et al. (2016)

A arquitetura da ResNet é inspirada nas redes neurais VGG (SIMONYAN; ZISSERMAN, 2014). As camadas convolucionais em sua grande maioria possuem filtros de tamanho 3×3 e possuem um *stride* de 2. Devido ao *stride*, a quantidade de mapas de características é diminuída pela metade, para preservar a complexidade das camadas a quantidade de filtros é sempre duplicada. No total existem 34 camadas, compostas por 4 blocos, estes contendo 3, 4, 6 e 3 blocos residuais, respectivamente.

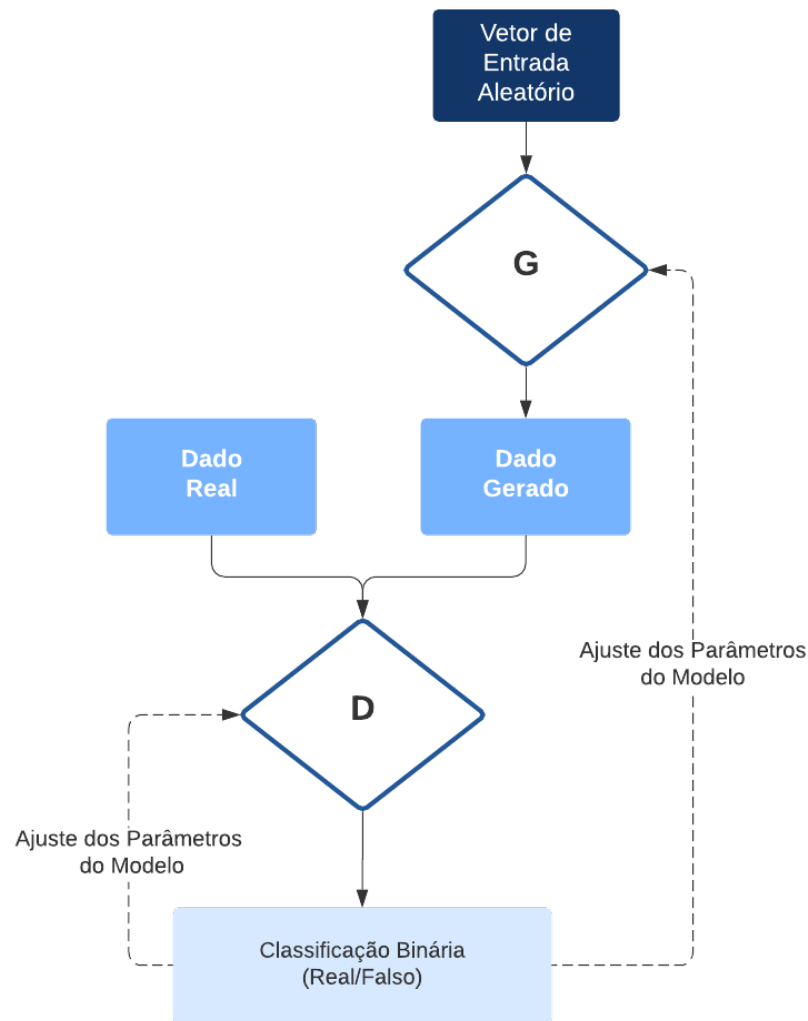
3.8 Redes Adversárias Geradoras

As Redes Adversárias Geradoras (GANs) tratam-se de uma técnica de aprendizado de máquina, de forma semi supervisionada ou não supervisionada, que fornece uma maneira de aprender representações complexas sem possuir dados de treinamento extensivamente catalogados (CRESWELL et al., 2018).

Para dados visuais, como imagens, as GANs podem ser caracterizadas como um par de redes neurais que estão sempre em competição uma com a outra. Uma rede

geradora G é treinada para criar imagens falsas enquanto outra rede discriminadora D é simultaneamente treinada para diferenciar imagens falsas de imagens reais (CRESWELL et al., 2018). A competição entre as redes faz com que ambas melhorem seus parâmetros e que as imagens falsas se aproximem do indistinguível em relação as reais. A Figura 11 mostra o funcionamento das GANs.

Figura 11 – Funcionamento das GANs.



Fonte: Criado pelo autor

A rede discriminadora D é treinada com o intuito de maximizar a probabilidade de atribuir o rótulo correto para os dados de treinamento, tanto para exemplos reais quanto para amostras geradas pela rede geradora G . Simultaneamente, G é treinada com o objetivo de minimizar $\log(1 - D(G(z)))$ (GOODFELLOW et al., 2014). A função objetivo

V do treinamento adversário é definida na Equação 3.2:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{dado}}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (3.2)$$

onde p_{dado} é a distribuição original dos dados x e p_z é a distribuição aprendida por G no treinamento.

Para este estudo, são usadas GANs para refinar a qualidade de imagens de satélites de baixa resolução com o objetivo de melhorar o desempenho na classificação automática de garimpos ilegais.

3.9 GANs de Super-resolução

O desafio de problemas de super-resolução consiste em estimar uma imagem de alta resolução (HR) a partir de uma imagem de baixa resolução (LR). Atualmente, o estado da arte para a resolução deste problema consiste no uso de GANs para o treinamento de uma rede capaz de gerar uma imagem de super-resolução (SR) que chegue bem próximo à qualidade de HR (LEDIG et al., 2017).

Uma Rede Adversária de Super-resolução (SRGAN) combina o uso de redes neurais convolucionais profundas com redes adversárias para produzir imagens de alta resolução. A SRGAN utiliza uma função de custo perceptual que consiste nos custos da GAN e da perda de conteúdo da imagem (LEDIG et al., 2017).

Dado uma rede discriminadora D_{θ_D} que possui o objetivo de discriminar entre uma imagem HR de outra imagem SR gerada pela rede geradora G_{θ_G} , a definição matemática da SRGAN análoga a mostrada na Equação 3.2 é expressa na Equação 3.3:

$$\min_{\theta_G} \max_{\theta_D} \mathbb{E}_{I^{HR} \sim p_{HR}(I^{HR})} [\log D_{\theta_D}(I^{HR})] + \mathbb{E}_{I^{LR} \sim p_{SR}(I^{LR})} [\log(1 - D_{\theta_D}(G_{\theta_G}(I^{LR})))] \quad (3.3)$$

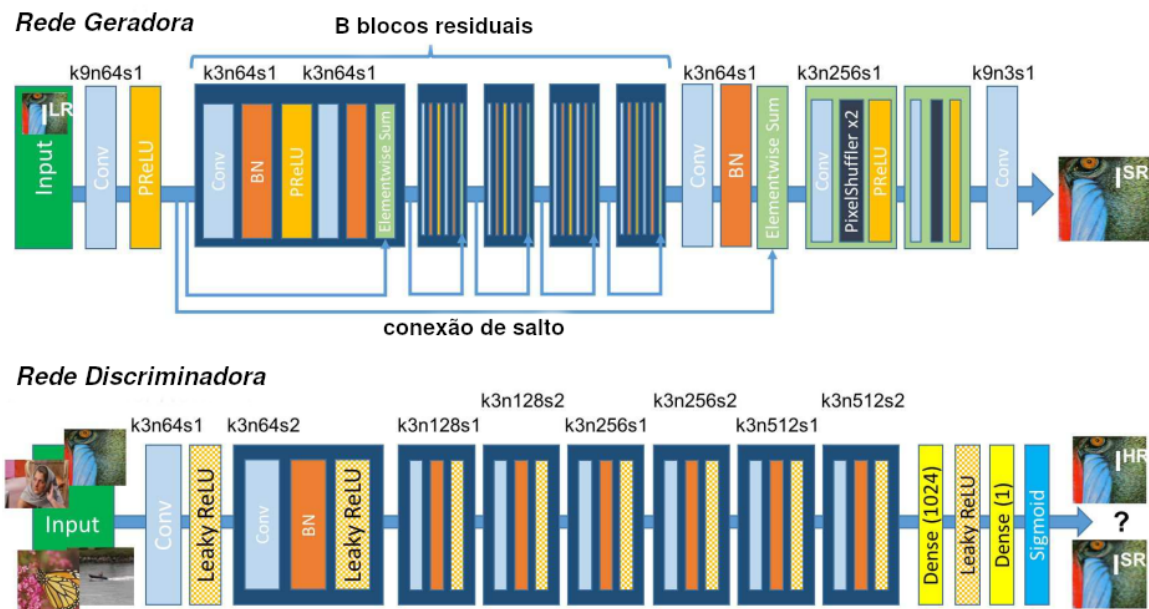
onde p_{HR} é distribuição original das imagens HR, p_{SR} é a distribuição das imagens SR, I^{HR} é uma imagem de alta resolução original e I^{LR} é a sua imagem equivalente em baixa resolução.

3.9.1 Arquitetura

Na arquitetura de G existem B blocos residuais iguais como mostrado no exemplo da Figura 12. Cada bloco contém duas camadas convolucionais com filtros 3x3 e 64 mapas de características, seguidos de camadas de normalização de lote (IOFFE; SZEGEDY, 2015) e uma *ParametricReLU* (HE et al., 2015) como função de ativação. A resolução da imagem é aumentada usando duas camadas de convolução de sub-pixel treinadas.

Para discriminar entre imagens HR de imagens SR a rede D contém 8 camadas convolucionais com um número crescente de filtros 3×3 , estes são aumentados por um fator 2 de 64 até 512 filtros tal como é feito para redes neurais VGG (SIMONYAN; ZISSERMAN, 2014). Não há camadas de *pooling* e a função de ativação utilizada é a *LeakyReLU* (RADFORD; METZ; CHINTALA, 2015). São usadas convoluções com *stride* toda vez que a quantidade de características chega ao dobro, e então, os 512 mapas de características resultantes são enviados para duas camadas densas e o resultado é gerado por uma função de ativação *sigmoid*. A Figura 12 mostra a arquitetura de D .

Figura 12 – Arquiteturas das redes geradora e discriminadora com seus correspondentes tamanho de filtros (k), quantidade de mapas de características (n) e tamanho de *stride* (s).



Fonte: Adaptado de Ledig et al. (2017)

3.9.2 Função de Custo Perceptual

A função de custo perceptual l^{SR} é crucial para o funcionamento da SRGAN. Apesar de ser muito comum usar funções de custo baseadas no erro quadrático médio (MSE) em trabalhos de super-resolução (DONG et al., 2015) a função de custo da SRGAN utiliza meios diferente, em l^{SR} , são consideradas as características perceptuais relevantes, baseando-se nos trabalhos de Johnson, Alahi e Fei-Fei (2016) e Bruna, Sprechmann e LeCun (2015). A função de custo perceptual é expressa matematicamente pela Equação 3.4:

$$l^{SR} = l_{VGG}^{SR} + 10^{-3} l_{Gen}^{SR} \quad (3.4)$$

A fórmula l_{VGG}^{SR} mostrada na Equação 3.4 é definida como o custo da VGG baseado na distância euclidiana entre as representações das características da imagem reconstruída $G_{\theta_G}(I^{LR})$ e a imagem original I^{HR} , a sua fórmula é expressa na Equação 3.5:

$$l_{VGG/i,j}^{SR} = \frac{1}{W_{i,j}H_{i,j}} \sum_{x=1}^{W_{i,j}} \sum_{y=1}^{H_{i,j}} (\phi_{i,j}(I^{HR})_{x,y} - \phi_{i,j}(G_{\theta_G}(I^{LR}))_{x,y})^2 \quad (3.5)$$

sendo $\phi_{i,j}$ o mapa de características obtido pela j -ésima convolução (após ativação) antes do i -ésima camada de *maxpooling* na VGG19 e $W_{i,j}$ e $H_{i,j}$ as dimensões desses respectivos mapas.

A segunda parte da Equação 3.4 é ligada ao componente gerador da rede e é definida na Equação 3.6. A probabilidade de que a imagem reconstruída $G_{\theta_G}(I^{LR})$ seja a imagem natural HR é definido como $D_{\theta_D}(G_{\theta_G}(I^{LR}))$. Para um melhor comportamento de gradiente tudo é normalizado pelo negativo do logaritmo.

$$l_{Gen}^{SR} = \sum_{n=1}^N -\log D_{\theta_D}(G_{\theta_G}(I^{LR})) \quad (3.6)$$

3.10 GANs de Super-resolução Aprimoradas

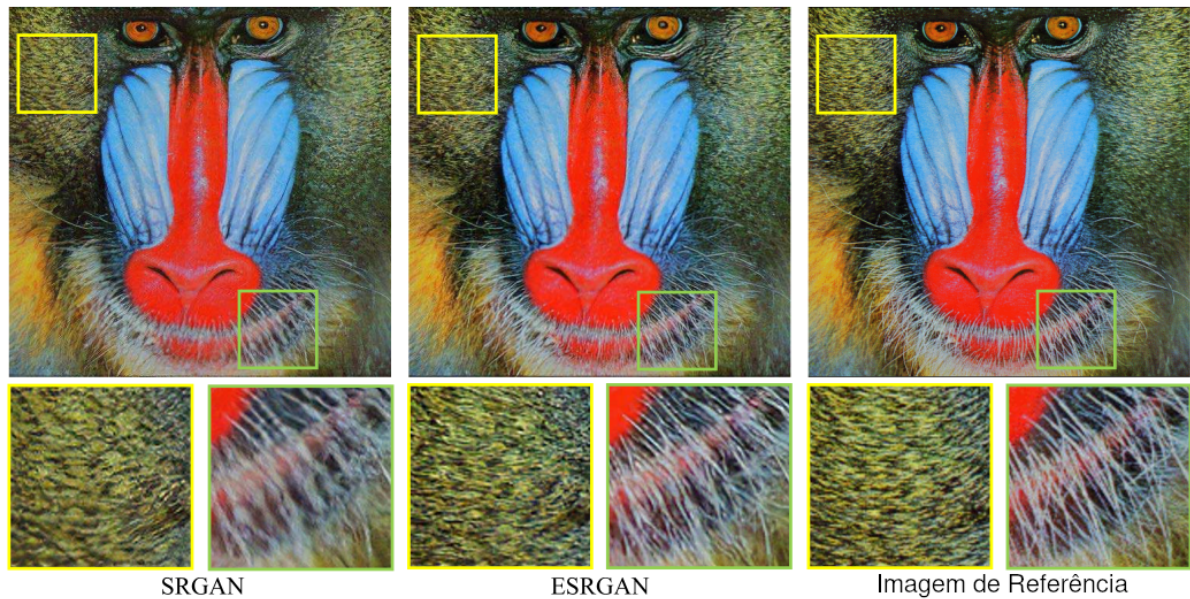
Um dos principais problemas no uso da SRGAN proposta na Seção 3.9 é que, apesar de reconstruir imagens de super-resolução com muitos detalhes e texturas realísticas, a presença de artefatos desagradáveis nessas imagens também é comum (WANG et al., 2018). Com a proposta de melhorar a qualidade visual das imagens, Wang et al. (2018) propôs a GAN de Super-resolução Aprimorada (ESRGAN), uma rede com arquitetura similar a da SRGAN com mudanças nas função de custo perceptual e remoção das camadas de normalização de lotes. A Figura 13 demonstra uma comparação na qualidade das imagens da SRGAN e ESRGAN.

3.10.1 Arquitetura da Rede Geradora

A arquitetura da ESRGAN para a rede geradora G difere da SRGAN em dois pontos principais, são estes: (1) remoção de todas as camadas de normalização de lotes (BN) e (2) substituição de todos blocos básicos por Blocos Densos *Residual-in-Residual* (RRDB) (LIM et al., 2017), que combinam redes residuais multinível com conexões densas.

A remoção das camadas de normalização de lotes melhoram a performance e reduzem a complexidade computacional em tarefas de super-resolução. Esses problemas de performance são mais visíveis quando há variância entre os conjuntos de treino e teste, o que acaba causando a presença de artefatos desagradáveis (WANG et al., 2018).

Figura 13 – Reconstrução da resolução da imagem em 4x usando a SRGAN e ESRGAN. Pode-se observar que a imagem da ESRGAN possui mais nitidez e detalhes do que a da SRGAN.



Fonte: Adaptado de Wang et al. (2018)

Visto que o uso de mais camadas e conexões em uma rede neural pode aumentar sua performance, são usadas camadas RRDB para dar uma estrutura mais profunda e complexa que a SRGAN.

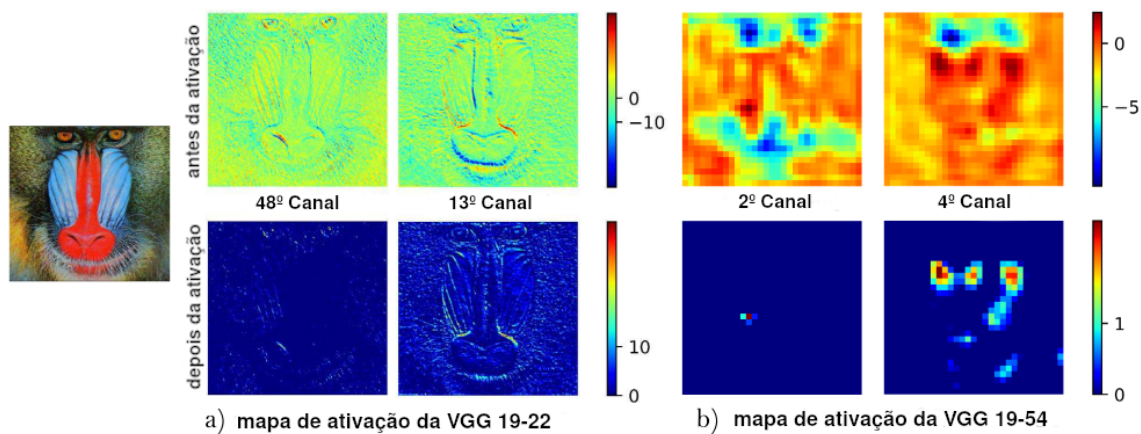
3.10.2 Arquitetura da Rede Discriminadora

A rede discriminadora D da ESRGAN usa a proposta do discriminador relativístico proposto por Jolicoeur-Martineau (2018). Na SRGAN D tenta estimar a probabilidade que uma imagem x possa ser a real, já o discriminador relativístico tenta prever a probabilidade de que uma imagem real x_r seja relativamente mais realística que uma imagem falsa x_f .

3.10.3 Função de Custo Perceptual

A função de custo perceptual da SRGAN utiliza o mapa de características da VGG após cada ativação, estes mapas tendem a ficar muito esparsos quanto mais profunda for a rede. Mapas de características esparsos possuem menos características relevantes e podem ocasionar uma performance inferior no treinamento da rede (LIM et al., 2017). Para contornar isso, a ESRGAN propõe utilizar o mapa de características $\phi_{i,j}$ mostrado na Equação 3.5 antes da ativação da j -ésima convolução da rede. A Figura 14 mostra um exemplo desses mapas antes e depois da ativação.

Figura 14 – Representação dos mapas de características antes e depois da ativação. Quanto mais profunda é a rede, mais informações são perdidas após a ativação, enquanto antes da ativação há características mais ricas.

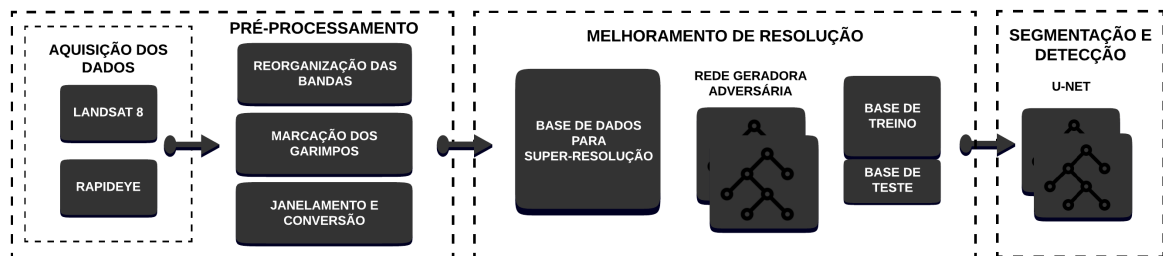


Fonte: Adaptado de Wang et al. (2018)

4 Metodologia

Neste capítulo é apresentada a metodologia proposta neste estudo, mostrada no fluxograma da Figura 15. Inicialmente, é realizada a aquisição dos dados provenientes dos satélites alvo, depois é feito um pré-processamento dos dados adquiridos, seguido do melhoramento da resolução das imagens através de GANs de super-resolução. Por fim, é realizada a segmentação e detecção dos garimpos usando a U-Net e a avaliação dos resultados obtidos.

Figura 15 – Metodologia do Estudo.



Fonte: Acervo do autor

4.1 Aquisição de Dados

Os dados utilizados neste trabalho foram adquiridos usando bancos de imagens do satélite *RapidEye* (RAPIDEYE, 2013) e *Landsat 8* (IHLEN, 2019). Isso se dá ao fato de que é necessário para a metodologia possuir uma imagem de satélite em alta resolução espacial e outra equivalente com resolução espacial inferior.

As imagens do RapidEye foram adquiridas de forma gratuita através do GEO Catálogo,¹ projeto realizado pelo Ministério do Meio Ambiente (MMA) (FEDERAL, 2020). Este satélite foi escolhido devido a sua alta resolução espacial (5 m) e riqueza de detalhes em suas imagens. O GEO Catálogo possui imagens adquiridas datadas entre os anos de 2011 e 2014, foram selecionadas as imagens mais recentes no ano de 2014.

Para obter as imagens do Landsat 8, utilizou-se o *EarthExplorer*,² sistema que é disponibilizado pelo USGS para busca e aquisição gratuita de imagens de sensoriamento remoto. Sua escolha foi realizada devido a sua resolução espacial inferior (30 m) e alta disponibilidade de imagens em seu banco de dados. As imagens do Landsat 8 estão datadas

¹ <http://geocatalogo.mma.gov.br/>

² <https://earthexplorer.usgs.gov/>

entre o ano de 2013 até a data presente deste trabalho. Foram escolhidas apenas imagens datadas no ano de 2014 devido à indisponibilidade gratuita de imagens do RapidEye mais recentes.

As imagens de satélite obtidas são referentes a região da área de proteção ambiental do Tapajós (AMBIENTE, 2020) em conjunto com parte da área da Floresta Nacional do Crepori (AGRICULTURA, 2020). Estas áreas estão localizadas no estado do Pará, nos municípios de Itaituba, Jacaraecanga, Trairão e Novo Progresso. Sua criação foi realizada em 13 de fevereiro de 2006 através de uma alteração no decreto N^o 4.340, seu objetivo é de proteger a diversidade biológica, disciplinar o progresso de ocupação e assegurar a sustentabilidade do uso de recursos naturais (CIVIL, 2002).

Essa região foi selecionada devido sua alta presença de garimpo e existência de áreas de floresta nacional como a Floresta Nacional do Jamanxin (8.628,95 km^2) e Floresta Nacional de Itaituba I (2.131,05 km^2). Há também a presença de áreas indígenas que costumam ser afetadas pelo garimpo ilegal, por exemplo, a reserva indígena do Munduruku.³

Os locais de garimpo na região de estudo foram identificados utilizando informações georreferenciadas fornecidas pelo Sistema de Informações Geográficas de Mineração (SIGMINE), disponibilizado pelo Ministério de Minas e Energia (MME), e por um conjunto de marcações compiladas pela Rede Amazônica de Informação Socioambiental (RAISG). Para maior confiabilidade, estes locais também foram inspecionados visualmente utilizando a ferramenta *Google Earth* (GORELICK et al., 2017). A Figura 16 ilustra o mapa da área de garimpos utilizada no estudo.

O Landsat possui imagens disponibilizadas em blocos de 7601x7741 píxeis no formato ".tif", contendo todas as bandas espectrais mostradas na Tabela 2. Cada pixel representa uma área de 900 m^2 . Foram selecionados 2 blocos de imagens que cobrem uma área de 66.600 km^2 .

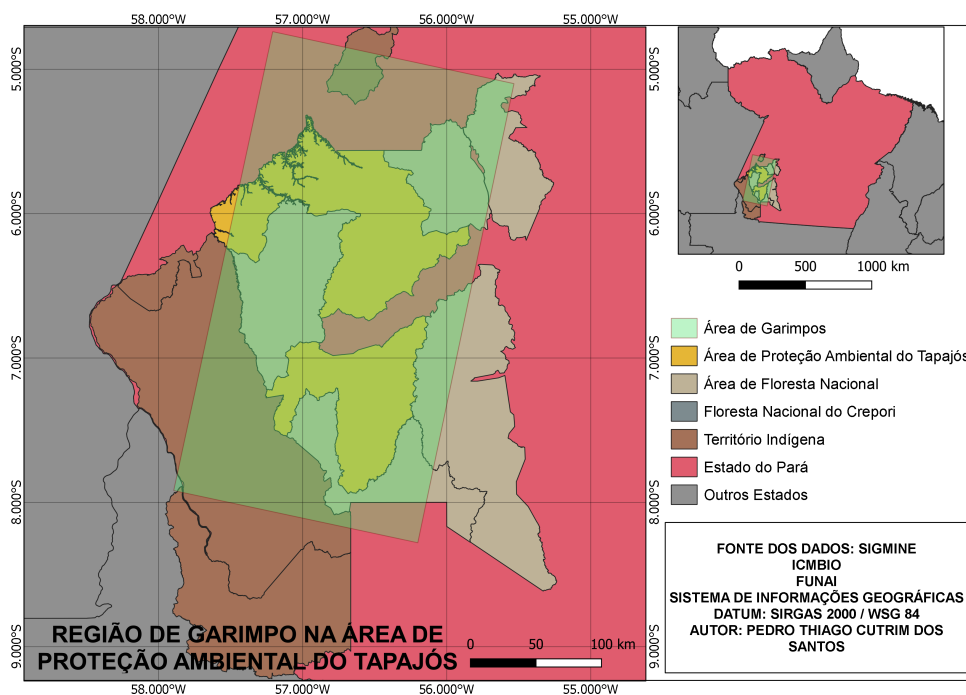
As imagens do RapidEye são disponibilizadas em blocos de 5000x5000 píxeis no formato ".img", contendo todas as bandas espectrais mostradas na Tabela 3. Cada pixel representa uma área de 25 m^2 . Foram selecionados 31 blocos de imagens que cobrem uma área de 19.375 km^2 .

4.2 Pré-processamento dos Dados

Essa seção mostra a etapa de pré-processamento dos dados deste estudo, está dividida em: reorganização de bandas, marcação de garimpos, janelamento de imagens e conversão dos dados.

³ <https://oglobo.globo.com/sociedade/um-so-planeta/garimpo-ignora-pandemia-avanca-30-na-terra-indigena-yanomami-em-2020-24939963>

Figura 16 – Mapa da região de garimpo e outros limites regionais.



Fonte: Acervo do autor

4.2.1 Reorganização de Bandas

Uma característica das imagens dos dois satélites obtidas neste estudo é a presença de múltiplas bandas, o Landsat 8 possui 11 bandas diferentes enquanto o RapidEye possui 5. É recomendado utilizar apenas as faixas espectrais relevantes para o domínio do problema devido ao ruído ou complexidade que o uso de muitas bandas pode causar (NOVO, 2001).

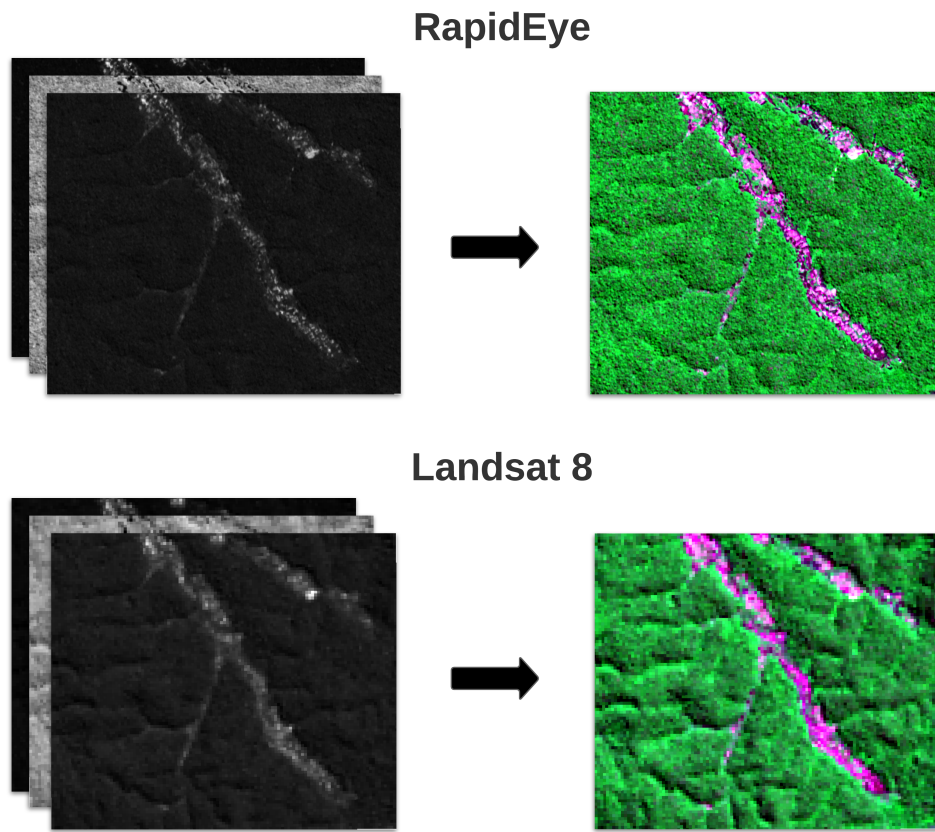
Para este estudo, foram utilizadas as bandas referentes as faixas espectrais da luz visível (vermelho e verde) e o infravermelho (infravermelho próximo) para montar imagens com composição de falsa cor RGB (DEMIRKESEN et al., 2007). A escolha dessas bandas foi baseando-se no estudo de Simionato et al. (2019), que mostra a efetividade dessa composição para regiões de garimpo. A Figura 17 mostra imagens do RapidEye e Landsat 8 utilizando esses 3 canais.

Através dessa composição, as áreas com maior atividade humana (Ex: áreas desmatadas e garimpos) em conjunto com sistema hidrográfico foram ressaltadas em relação a regiões apenas de mata preservada. A Figura 18 mostra um exemplo de área antropizada que foi realçada.

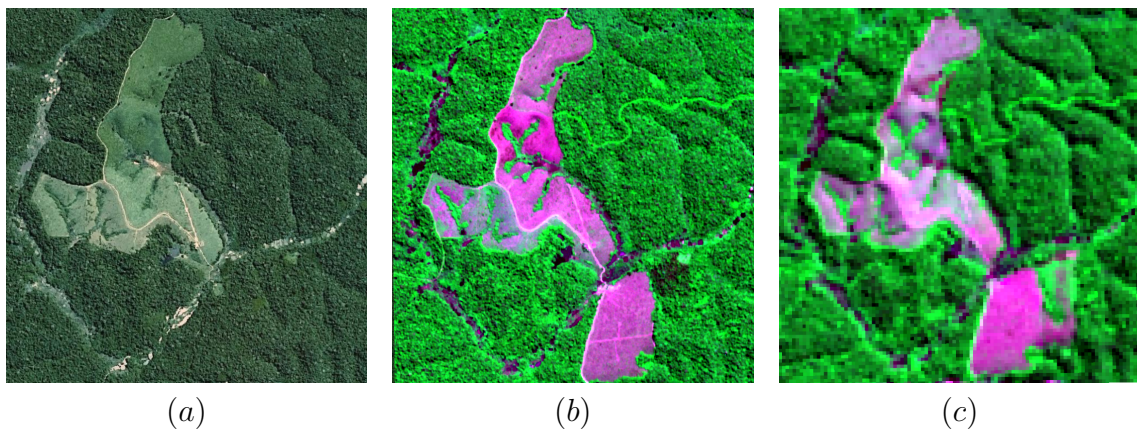
4.2.2 Marcação dos Garimpos

Para a etapa de treinamento é necessário rotular as regiões de garimpo nas imagens de satélite. Para criar as marcações, utilizou-se o *software* QGIS (QGIS, 2021). Através

Figura 17 – Composição de falsa cor RGB nos satélites RapidEye e Landsat 8.



Fonte: Acervo do autor

Figura 18 – Área antropizada. (a) *Aerial View (Bing Maps)* (MICROSOFT, 2021) (b) RapidEye e (c) Landsat 8.

Fonte: Acervo do autor.

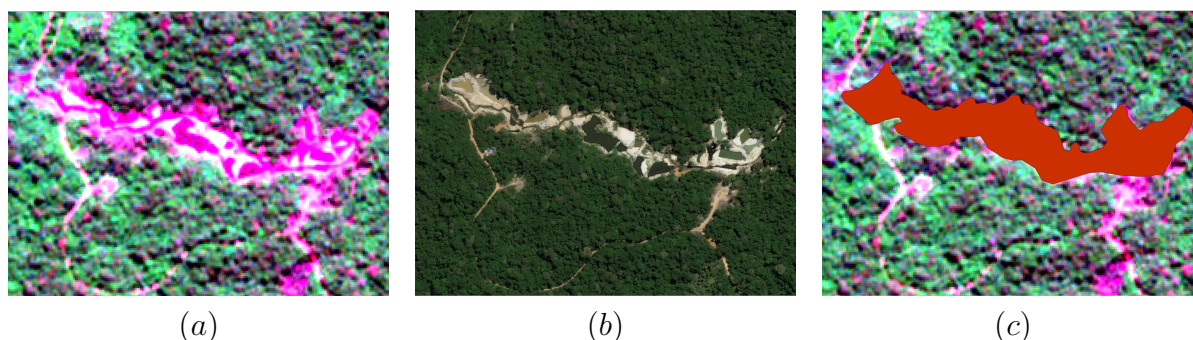
do QGIS foi possível carregar as camadas das imagens de satélite e criar marcações dos garimpos de forma georreferenciada para que seja possível aproveitar em ambos os satélites.

A demarcação foi realizada de maneira manual utilizando as imagens do RapidEye e Landsat em conjunto com uma validação visual feita através de imagens geográficas de

vista aérea fornecida pelo *Bing Maps* (MICROSOFT, 2021).

Na heurística utilizada para realizar as marcações, foram consideradas regiões que possuem removidas toda sua cobertura vegetal das jazidas, solo revolvido e degradação da qualidade das águas superficiais, características predominantes em garimpos (EICKHOFF, 2011). Após a identificação, a área é validada visualmente e marcada. A Figura 19 mostra um exemplo de como é feito a marcação.

Figura 19 – Marcação de um garimpo. (a) visualização em falsa cor RGB (b) validação visual e (c) marcação.



Fonte: Acervo do autor.

Após completa, as marcações são exportadas em formato *".shp"* e são criadas máscaras binárias usando as bibliotecas *GDAL* (WARMERDAM, 2008) e *GeoPandas* (JORDAHL, 2014) desenvolvidas na linguagem *Python 3* (ROSSUM; DRAKE, 2009).

4.2.3 Janelamento de Imagens e Conversão dos Dados

O treinamento de uma rede neural utilizando as resoluções originais das imagens (7601x7741 Landsat 8 e 5000x5000 RapidEye) é inviável devido a limitações no *hardware* em relação a grande quantidade de parâmetros que seriam gerados durante o treinamento (SZE et al., 2017). Para solucionar este problema cada imagem foi subdividida em frações iguais.

Em imagens do Landsat 8, o janelamento foi realizado em subdivisões de 128x128 píxeis. Para manter as subdivisões georreferenciadas entre os satélites, as imagens do RapidEye foram subdivididas em 768x768 píxeis, isso foi necessário pelo fato de sua resolução espacial ser seis vezes maior.

Após a realização do janelamento, as novas imagens são convertidas para o formato *".png"* com o objetivo de facilitar o uso das bibliotecas de aprendizado de máquina, utilizadas neste estudo. Utilizando as marcações dos garimpos, foram identificadas quais imagens possuem atividade garimpeira e quais não possuem. A Tabela 4 mostra um panorama da nova base de imagens criada.

Tabela 4 – Panorama do Janelamento

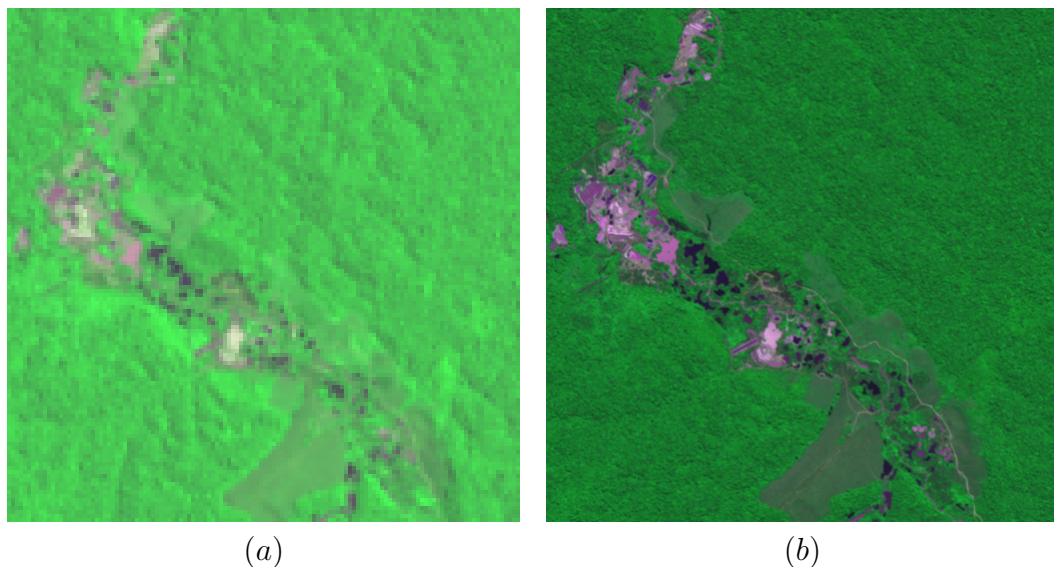
Satélite	c/ Garimpo	s/ Garimpo	Resolução	Área p/ Imagem	Formato
Landsat 8	286	830	128x128	14,75 km ²	PNG
RapidEye	286	830	768x768		

4.3 Melhoria de Resolução

Esta seção mostra a etapa de melhoria de resolução das imagens de satélite Landsat usando GANs de super-resolução e abordagens diferentes no treinamento.

Por possuir uma resolução espacial inferior, as imagens do Landsat possuem um nível menor de detalhamento em relação ao RapidEye. A Figura 20 mostra uma comparação de uma região de garimpo vista pelos dois satélites.

Figura 20 – Região de garimpo visualizada através de satélite. (a) Landsat 8 e (b) RapidEye.



Fonte: Acervo do autor.

A ausência de certos detalhes importantes em uma imagem pode interferir na acurácia de uma CNN (DODGE; KARAM, 2016). Para melhorar a qualidade em imagens Landsat, foram utilizadas redes de super-resolução treinadas com a resolução espacial do RapidEye como referência.

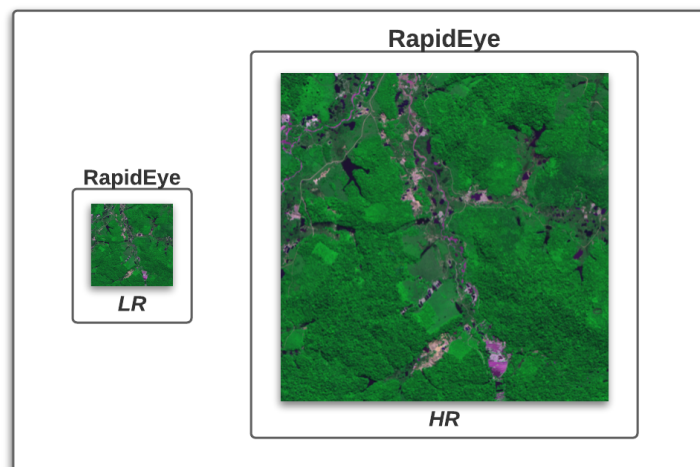
As GANs de super-resolução utilizadas neste estudo foram a SRGAN (LEDIG et al., 2017) e a ESRGAN (WANG et al., 2018). Para treinamento das GANs foram utilizados duas abordagens diferentes em relação a criação da base de dados. A primeira abordagem visa o treinamento das GANs de forma convencional, as imagens de baixa resolução, LR, correspondendo as de alta resolução, HR. Para a segunda técnica, as GANs são treinadas com o objetivo de aprender uma relação entre imagens LR Landsat e imagens HR do

Rapideye. As abordagens são descritas na subseções seguintes.

4.3.1 Base de Dados Convencional

Como mostrado na Seção 3.9, uma base de treinamento para GANs de super-resolução consiste normalmente em um conjunto de imagens em alta resolução (HR) com suas respectivas imagens de baixa resolução (LR), estas normalmente geradas usando alguma técnica de redimensionamento de imagem. Para imagens HR foram utilizadas as imagens do RapidEye criadas na Seção 4.2, as imagens LR referentes foram geradas a partir do redimensionamento de HR utilizando uma interpolação bicúbica (FADNAVIS, 2014). A Figura 21 exemplifica a estrutura a base de dados.

Figura 21 – Base de dados convencional.



Fonte: Acervo do autor

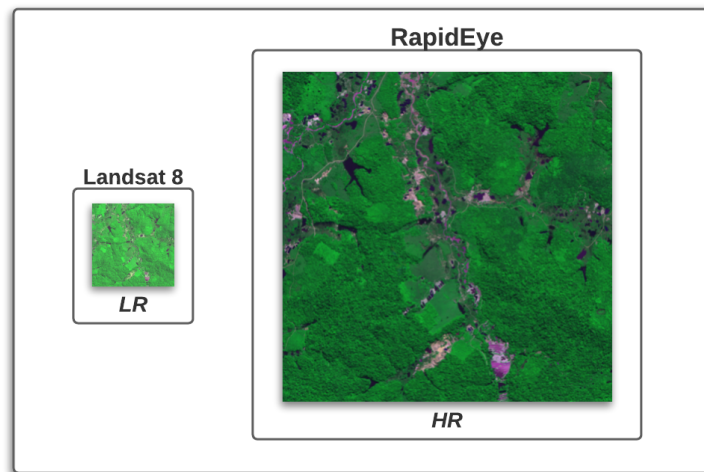
A base consiste em 720 imagens HR com resolução 512x512 e 720 respectivas imagens LR com resolução 128x128. A resolução das imagens HR foi diminuída de 768x768 para 512x512 utilizando interpolação bicúbica, isso foi necessário para manter um fator de escala 4x em relação as imagens do Landsat e também para diminuir o tempo de treinamento e a a quantidade de parâmetros das redes utilizadas neste estudo.

4.3.2 Treinamento de Transformação

Nesta abordagem, os indivíduos LR da base consistem nas imagens de resolução 128x128 do Landsat. Para HR, os indivíduos continuam sendo as imagens 512x512 do RapidEye, essa resolução é usada pelos motivos apresentados na Subseção 4.3.1. A Figura 22 mostra um exemplo da estrutura da base de dados.

A estratégia na utilização dessa base de dados é conseguir treinar uma GAN de super-resolução capaz de realizar uma transformação de uma imagem Landsat, com

Figura 22 – Base de dados de transformação.



Fonte: Acervo do autor

resolução espacial 30 *m*, em uma do RapidEye, que possui resolução espacial de 5 *m*. A quantidade de elementos na base é a mesma mostrada na Subseção 4.3.1.

4.4 Segmentação e Detecção

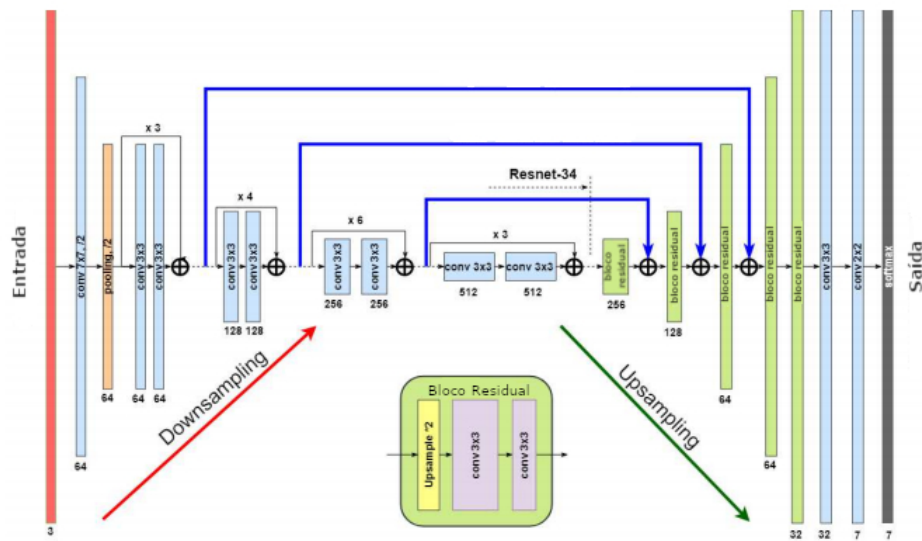
Nesta seção é apresentado o treinamento de uma rede neural para segmentação semântica e como é feita a avaliação dos resultados das bases de dados criadas neste estudo.

Para esta etapa, usou-se a U-Net para realizar a segmentação e detecção dos garimpos. Por mostrar bons resultados na classificação usando imagens de satélite de cobertura de terra, a arquitetura usada é a mostrada no trabalho de [Rakhlin, Davydov e Nikolenko \(2018\)](#). Para uma melhor comparação entre as bases de dados criadas no estudo, são usados os mesmos parâmetros de treinamento e arquitetura. A Figura 23 mostra a arquitetura utilizada.

No processo de divisão das bases, foi utilizado a técnica de validação cruzada para particionar os dados em subconjuntos mutuamente exclusivos. O objetivo é ter uma maior precisão ao avaliar a capacidade de generalização dos modelos em cima dos dados.

A função de custo utilizada é a *Cross Dice* ([ISENSEE et al., 2018](#)), que tem capacidade em fugir de mínimos locais devido a uma combinação entre as funções de entropia cruzada e o coeficiente *dice*. O otimizador utilizado é o *Adadelta* ([ZEILER, 2012](#)), uma robusta técnica de descida de gradiente estocástico que não precisa *tuning* da taxa de aprendizado durante o treinamento.

Figura 23 – Arquitetura da U-Net.



Fonte: Adaptado de [Rakhlín, Davydov e Nikolenko \(2018\)](#)

4.4.1 Avaliação

Para avaliar os resultados, foram usadas métricas para medir o desempenho da segmentação e detecção. Para calcular as métricas que serão apresentadas a seguir, são usados os conceitos de verdadeiro positivo, verdadeiro negativo, falso positivo e falso negativo:

- **Verdadeiro Positivo (VP):** quantidade de indivíduos da área de interesse que foram classificados corretamente.
- **Verdadeiro Negativo (VN):** quantidade de indivíduos fora da área de interesse que foram classificados corretamente.
- **Falso Positivo (FP):** quantidade de indivíduos classificados incorretamente como pertencentes a área de interesse.
- **Falso Negativo (FN):** quantidade de indivíduos classificados incorretamente como não pertencentes a área de interesse.

A métrica usada para avaliar a segmentação foi o coeficiente de *dice*. Para o cálculo, é levado em consideração a classificação de cada píxel na imagem. Essa métrica é definida como:

- **Coeficiente de Dice (DICE) ou F1-Score:** similaridade entre a região de interesse real e a região de interesse gerada pela rede neural. Definido pela Equação

4.1:

$$\frac{2VP}{2VP + FN + FP} \quad (4.1)$$

Para avaliar a detecção, também são utilizadas as métrica de acurácia, sensibilidade e precisão. Para o cálculo das métricas de detecção, cada imagem é considerada como um indivíduo. Foi considerado um limiar de 0.5 em relação ao valor de DICE para definir se uma imagem deve ser considerado como uma área de garimpo ou não. São definidas como:

- **Acurácia (ACC):** definida pela Equação 4.2, significa o acerto geral do classificador.

$$\frac{VP + VN}{VP + FN + VN + FP} \quad (4.2)$$

- **Sensibilidade (SENS):** desempenho da rede em classificar corretamente a área de interesse. Definida pela Equação 4.3:

$$\frac{VP}{VP + FN} \quad (4.3)$$

- **Precisão (PREC):** proporção dos indivíduos da área de interesse da imagem classificados corretamente em relação a todos os indivíduos classificados como pertencentes a mesma. Definido pela Equação 4.4:

$$\frac{VP}{VP + FP} \quad (4.4)$$

4.5 Considerações finais

Este capítulo mostrou de forma detalhada as etapas realizadas na metodologia do estudo. As etapas foram: aquisição dos dados, pré-processamento, melhoramento de resolução, segmentação e detecção. No próximo capítulo são apresentados e discutidos os resultados obtidos.

5 Resultados

Essa seção apresenta os resultados obtidos após a execução dos experimentos propostos. Todos os experimentos foram executados na plataforma *Google Colab* (BISONG, 2019). O *hardware* é composto por 13 *gigabytes* de memória RAM e um processador Intel(R) Xeon(R) @ 2.00GHz. A GPU disponibilizada pode variar entre uma NVIDIA Tesla T4 ou P100, ambas com 14 *gigabytes* de memória.

5.1 Super-Resolução

As GANs de super-resolução são treinadas usando duas bases de dados, criadas baseando-se nas abordagens mostradas na Seção 4.3. Para gerar essas bases, foram usadas 726 imagens do total das 1116 presentes no estudo, 10% dos indivíduos são imagens com presença de garimpo. Esse balanceamento foi escolhido devido às imagens que possuem áreas de garimpo também serem cobertas por cerca de 90% de mata, o que infere que o aprendizado das características de mata é de grande importância para a reconstrução da imagem de alta resolução. No total são realizados quatro treinamentos diferentes, dois na SRGAN e outros dois na ESRGAN. A Tabela 5 mostra informações das bases de treinamento.

Tabela 5 – Bases para treinamento de super-resolução.

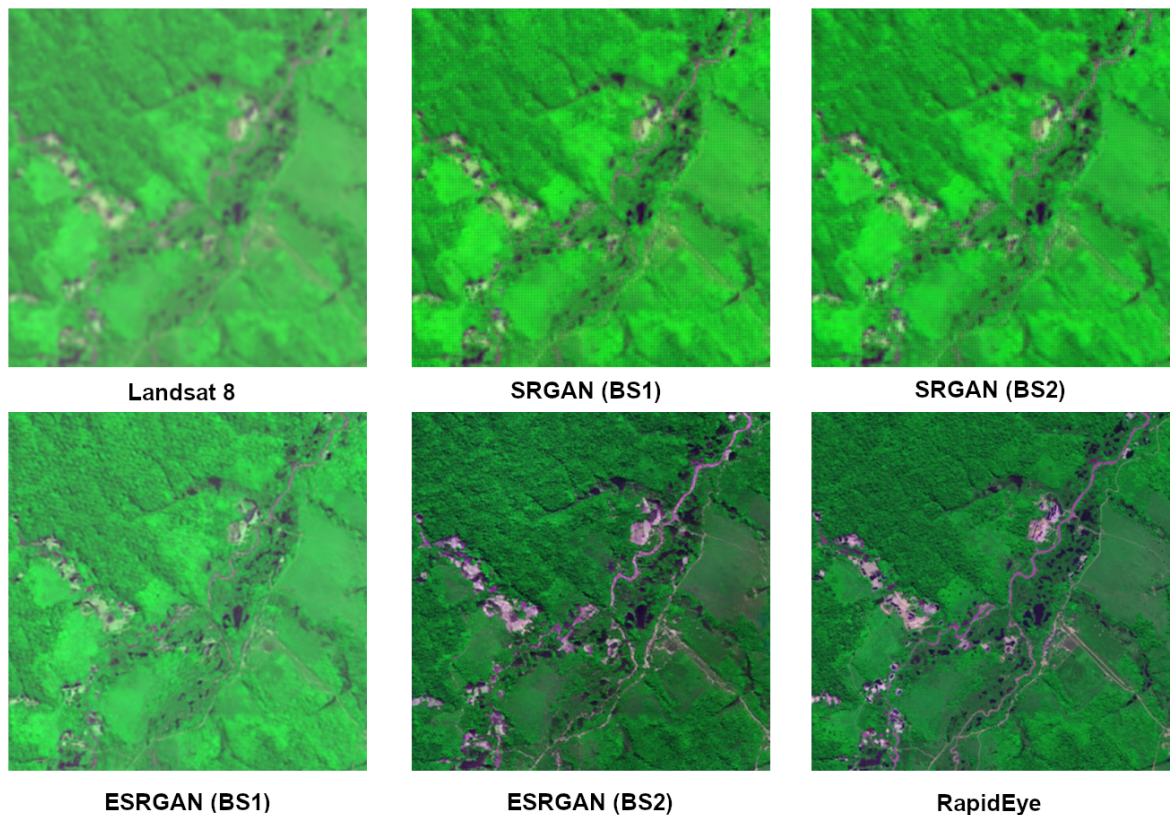
Nomenclatura	Abordagem	Qtd. HR	Qtd. LR	Total de Imagens
BS1	Subseção 4.3.1.	720	720	1.440
BS2	Subseção 4.3.2.	720	720	1.440

Para treinamento da SRGAN, foram utilizados a mesma arquitetura e parâmetros propostos em Ledig et al. (2017). A taxa de aprendizado do treinamento adversário utilizada foi de 0.1. O número de épocas usadas foi de 2000 e o treinamento foi realizado através da biblioteca *Tensorlayer* (DONG et al., 2017), demorando cerca de 25 horas.

A ESRGAN foi treinada com mesma arquitetura e parâmetros propostos em Wang et al. (2018), através da biblioteca *BasicSR* (WANG et al., 2020). O número de épocas foi de 2000 e a taxa de aprendizado de 0.01. O treinamento durou cerca de 30 horas.

Como observado na Figura 24, todos os modelos conseguiram gerar imagens com um detalhamento perceptualmente superior à imagem original Landsat. O treinamento de transformação (BS2) da ESRGAN conseguiu se aproximar de forma razoável da imagem RapidEye, mostrando ser possível o aprendizado usando essa abordagem.

Figura 24 – Resultado do melhoramento de uma imagem Landsat utilizando os modelos treinados.



Fonte: Acervo do autor

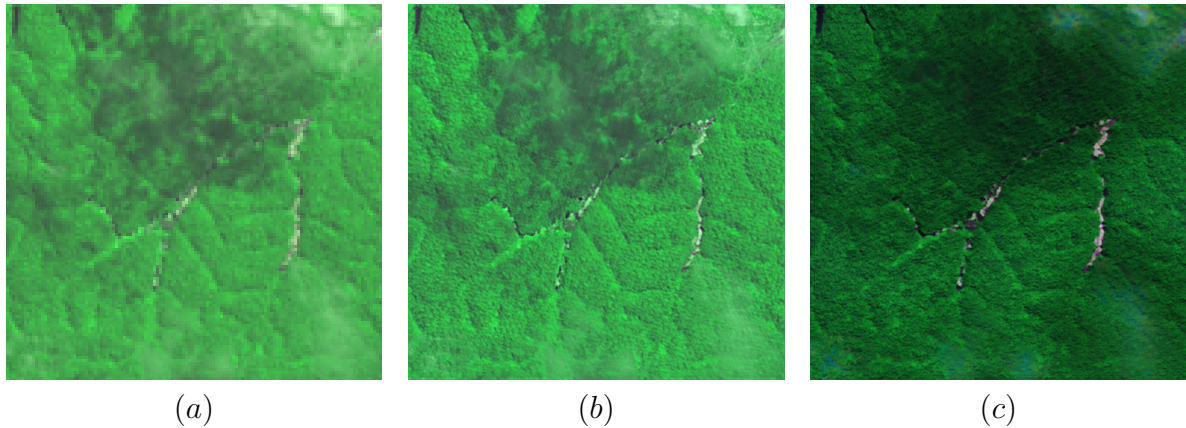
Notou-se que os modelos treinados com a BS2 conseguiram minimizar a presença de nuvens na imagem, o que pode melhorar a visualização do garimpo, tal exemplo pode ser visto na Figura 25. Isso pode explicado pelo fato da GAN ser treinada com o objetivo de aprender um mapeamento entre uma imagem Landsat 8 (que normalmente tem a presença de nuvens) para uma imagem RapidEye (que possui baixa presença de nuvens nos dados usados neste estudo).

As GANs treinadas nesta etapa são usadas para melhoramento das imagens utilizadas nas etapas de segmentação e detecção.

5.2 Segmentação

Nos experimentos de segmentação, é usada a U-Net com arquitetura mostrada na Seção 3.6. Das 360 imagens restantes, são usadas 252 imagens de garimpos para gerar as bases de treinamento. Essas bases são criadas usando os modelos de GANs de super-resolução mostrados na Subseção 5.1, são elas:

Figura 25 – Região de garimpo com presença de nuvens, nota-se que (c) consegue eliminar melhor a presença das nuvens. (a) Imagem Original, (b) ESRGAN (BS1) e (c) ESRGAN (BS2).



Fonte: Acervo do autor.

- **128x128 Original:** composta de 252 imagens de garimpo Landsat de resolução 128x128.
- **512x512 SRGAN (BS1):** composta de 252 imagens de garimpo Landsat de resolução 512x512 geradas pelo modelo SRGAN (BS1).
- **512x512 SRGAN (BS2):** composta de 252 imagens de garimpo Landsat de resolução 512x512 geradas pelo modelo SRGAN (BS2).
- **512x512 ESRGAN (BS1):** composta de 252 imagens de garimpo Landsat de resolução 512x512 geradas pelo modelo ESRGAN (BS1).
- **512x512 ESRGAN (BS2):** composta de 252 imagens Landsat de resolução 512x512 geradas pelo modelo ESRGAN (BS2).

Para maior confiabilidade no treinamento, foi usada a técnica de validação cruzada com um número de *folds* igual a 10. Cada *fold* foi treinado por 200 épocas. Os resultados são apresentados na Tabela 6.

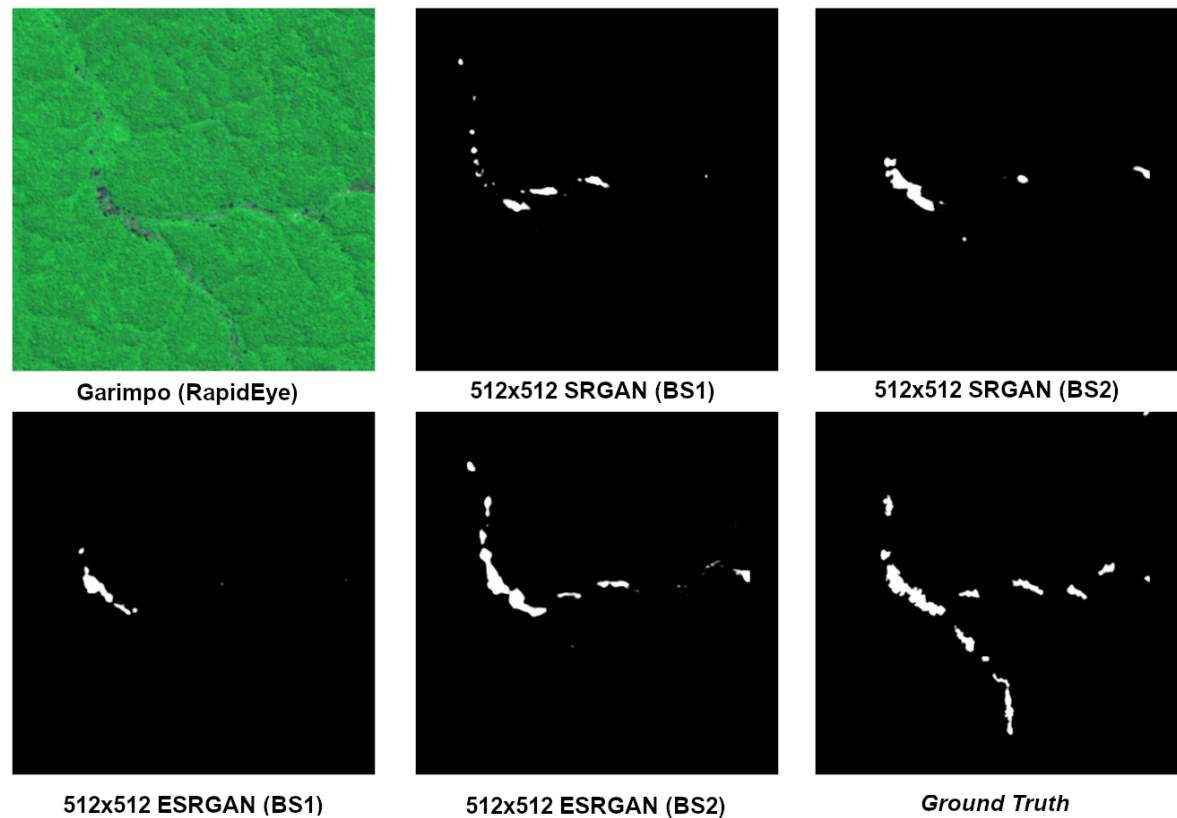
Tabela 6 – Resultados da Segmentação.

Base de Dados	DICE
128x128 Original	0.40 ± 0.08
512x512 SRGAN (BS1)	0.50 ± 0.07
512x512 SRGAN (BS2)	0.44 ± 0.04
512x512 ESRGAN (BS1)	0.49 ± 0.07
512x512 ESRGAN (BS2)	0.47 ± 0.03

Como observado na Tabela 6, todas as bases geradas através dos modelos de super-resolução tiveram um desempenho superior à base original. A abordagem BS1 alcançou os

maiores níveis de DICE, isso pode ser relacionado ao fato de que essa técnica preserva e realça melhor os traços da imagem original, ao contrário da técnica BS2, que tenta melhorar a imagem transformando esses mesmos traços, o que pode aumentar a probabilidade de criar artefatos não desejados.

Figura 26 – Garimpo com segmentação difícil. Nota-se que os modelos BS2 se aproximaram melhor do *ground truth*.



Fonte: Acervo do autor

Os modelos treinados com a BS2 demonstraram uma maior capacidade de generalização, como pode ser observado em seu menor desvio padrão. Isso pode ser devido à estrutura da imagem de BS2 facilitar em segmentações mais difíceis (devido à visibilidade ruim, formato mais complexo e/ou presença de nuvens), como pode ser visto na Figura 26.

5.3 Detecção

Para os experimentos de detecção, são usados os modelos U-Net treinados na Seção 5.2. A detecção é feita em cima das 108 imagens restantes, sendo 20 delas de garimpos. É usado um limiar de 0.5 em relação ao valor de DICE para considerar se uma imagem é área de garimpo ou não. Os resultados obtidos podem ser vistos na Tabela 7.

Tabela 7 – Resultados da Detecção.

Base de Dados	ACC	SENS	PREC	FSCORE
128x128 Original	0.78	0.65	0.93	0.76
512x512 SRGAN (BS1)	0.83	0.75	0.94	0.83
512x512 SRGAN (BS2)	0.81	0.70	0.93	0.80
512x512 ESRGAN (BS1)	0.81	0.70	0.93	0.80
512x512 ESRGAN (BS2)	0.86	0.80	0.94	0.86

Todas as abordagens obtiveram um aumento nas métricas calculadas em relação as obtidas utilizando a base de dados original. Isso reforça que o uso de super-resolução em imagens de satélite pode contribuir positivamente no treinamento de redes neurais convolucionais, em específico para detecção de áreas de garimpo.

O aprimoramento na qualidade das imagens pode estar ligado a influência positiva nas métricas de sensibilidade dos experimentos, maior ganho obtido por todas as abordagens. Vale apontar que a precisão, que originalmente possui um valor elevado, se manteve usando as bases melhoradas. Isso mostra que não há um aparente *trade-off* nos resultados usando a metodologia deste estudo.

O modelo ESRGAN treinado usando a abordagem BS2 obteve o melhor resultado em relação as outras abordagens. O maior ganho foi na sensibilidade de detecção dos garimpos, aumentando, em média, 15% através do melhoramento da imagem. Esse aumento afetou o *F1 Score*, que cresceu, no melhor caso, cerca de 10%. No acerto geral, ocorreu um aumento médio de 5% usando o melhoramento. Todas os modelos tiveram um boa capacidade em evitar a ocorrência de falso positivos, como mostra os resultados de precisão.

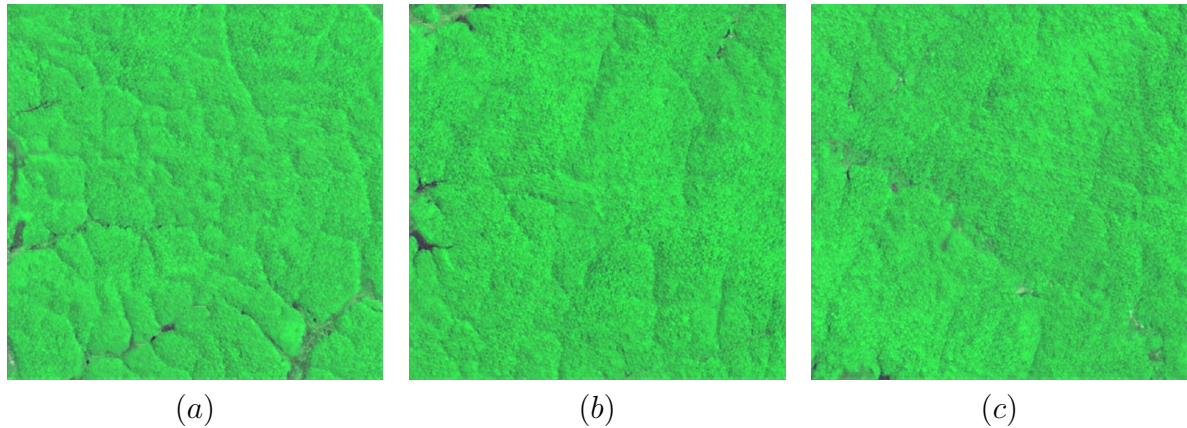
Os motivos para a abordagem BS2 possuir desempenho superior na detecção também pode estar relacionado a melhor detecção de garimpos difíceis, mostrado na Seção 5.2. A Figura 27 mostra alguns exemplo de regiões que foram detectadas por BS2 mas não por BS1.

5.4 Estudos de Caso

Para uma melhor confiabilidade do resultados, é necessário realizar uma análise rigorosa sobre os pontos positivos e negativos do estudo. A identificação de falhas ajuda na evolução do método proposto contribuindo no avanço de seu campo de pesquisa. Os estudos de caso foram realizado usando os melhores modelos de super-resolução em cada uma das abordagens.

A análise é feita utilizando as imagens originais Landsat 8 com as respectivas imagens de super-resolução construídas usando os modelos SRGAN e ESRGAN. Desta maneira, é apresentado um caso de sucesso para abordagem BS1 na Figura 28, onde a

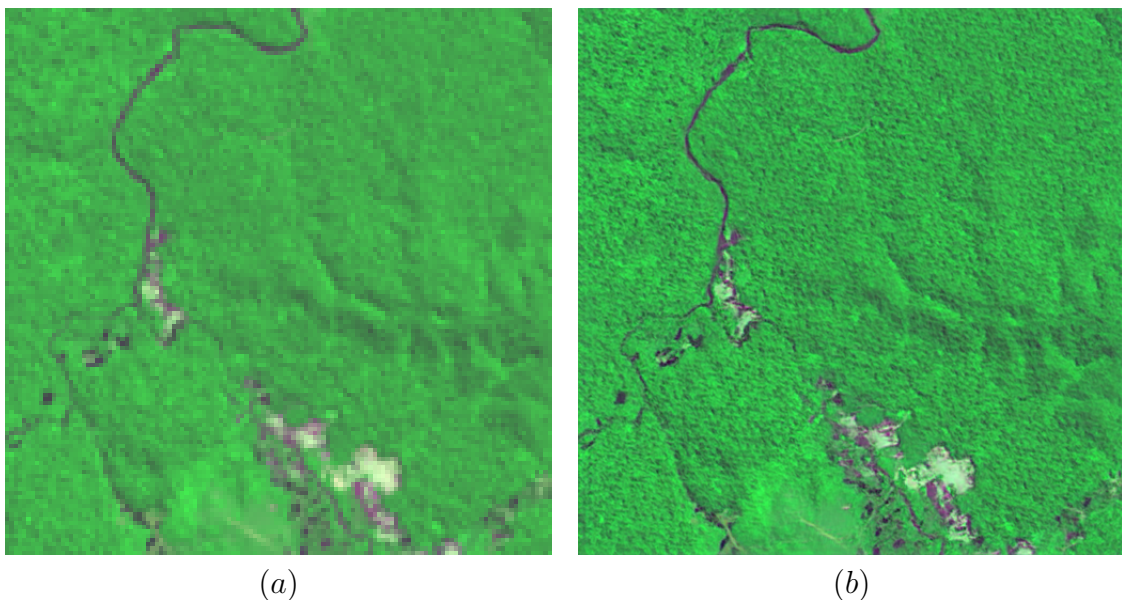
Figura 27 – Exemplo de garimpos detectados pela abordagem BS2 mas não por BS1. (a) e (b) garimpos parcialmente cortados pela imagem e (c) garimpo capturado com poucos detalhes pelo satélite.



Fonte: Acervo do autor.

região de garimpo encontra-se perto do leito de um rio. Nota-se que (b) possui resolução melhor e mantém todas as características de (a) sem introduzir muitos artefatos.

Figura 28 – Imagem de Região de garimpo antes e depois da reconstrução (abordagem BS1) com super-resolução. (a) imagem original e (b) imagem de super-resolução.

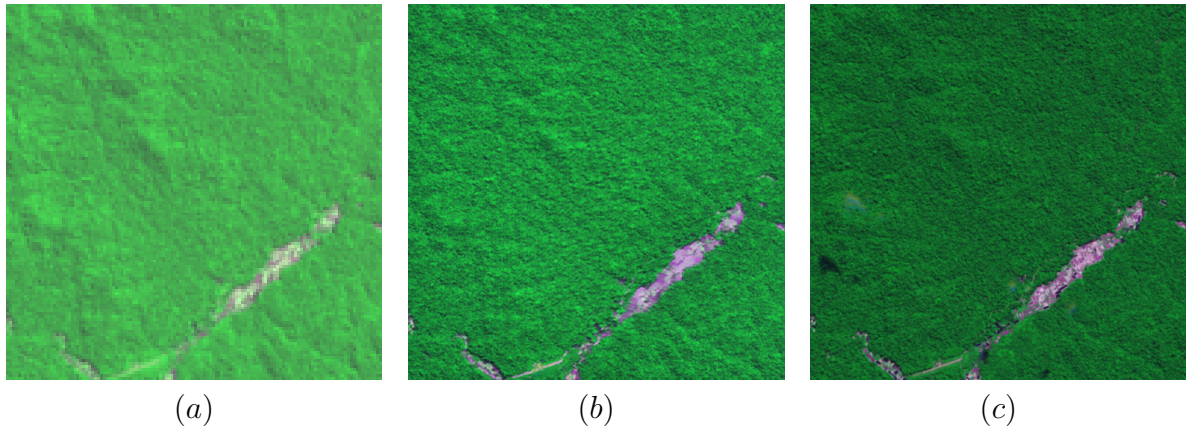


Fonte: Acervo do autor.

Um caso de sucesso para BS2 pode ser visto na Figura 29. Percebe-se que a imagem reconstruída (b) consegue ser uma boa transformação da imagem (a) para a cena do RapidEye mostrada em (c). A presença de alguns artefatos não desejados que já era esperada, se mostrou mínima.

Como um caso de erro para a abordagem BS2, a Figura 30 mostra a reconstrução

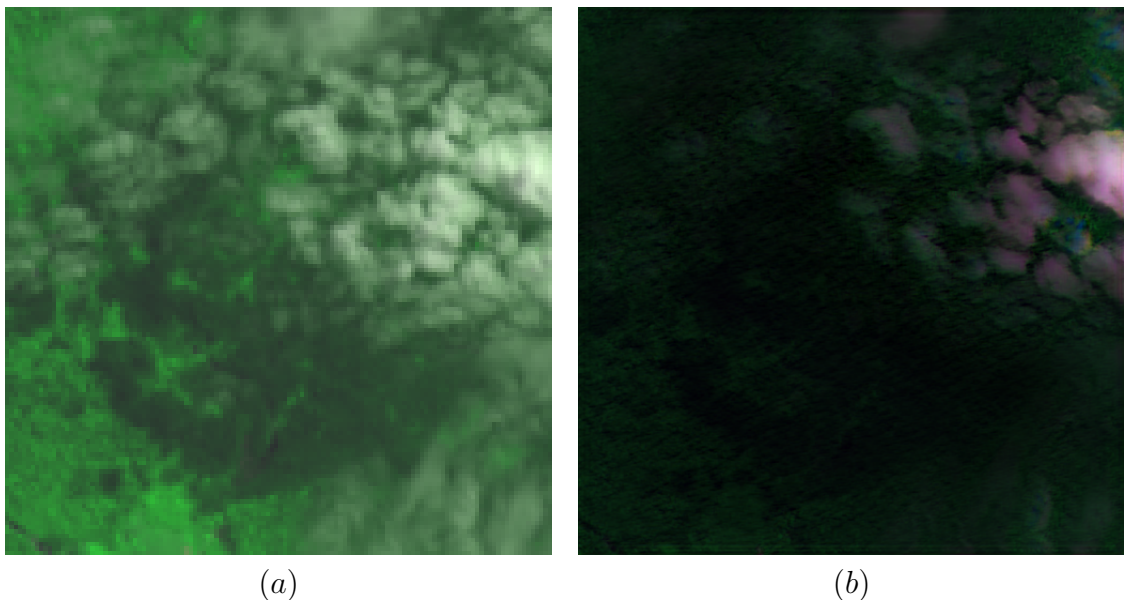
Figura 29 – Fonte: Imagem de Região de garimpo antes e depois da transformação (abordagem BS2) com super-resolução. (a) imagem original Landsat 8 (b) imagem de super-resolução e (c) imagem de referencia RapidEye



Fonte: Acervo do autor.

errada de uma imagem Landsat 8. A alta presença de nuvens, que pode ser frequente em imagens de satélite, pode ter confundido a rede, introduzindo artefatos indesejados e um erro na iluminação da imagem (b).

Figura 30 – Fonte: Imagem de Região de garimpo com alta presença de nuvens antes e depois da transformação (abordagem BS2) com super-resolução. (a) imagem original e (b) imagem de super-resolução.



Fonte: Acervo do autor.

5.5 Considerações Finais

Neste capítulo foram mostrados e discutidos os resultados obtidos neste estudo. No próximo capítulo são discutidos os objetivos que foram alcançados, limitações e trabalhos futuros.

6 Conclusão

Este trabalho propõe o desenvolvimento e avaliação de uma metodologia baseada em redes adversárias de super-resolução para aprimoramento da detecção de garimpos em imagens do satélite Landsat 8 situadas na Região de Integração do Tapajós.

Para o melhoramento da qualidade das imagens Landsat 8, são usadas duas redes adversárias de super-resolução: SRGAN e ESRGAN. Para a criação das bases são utilizadas duas abordagens diferentes. O treinamento é feito utilizando 1.440 imagens, sendo 10% de regiões de garimpo. Usando os modelos de super-resolução treinados, é criada uma base contendo 252 imagens com qualidade aprimorada para avaliação dos resultados de segmentação de regiões de garimpo. É utilizada a rede neural convolucional U-Net para gerar a segmentação semântica das regiões de interesse, os resultados são obtidos usando validação cruzada com 10 *folds*. A detecção é feita em cima de 108 novas imagens usando o modelo da U-Net treinado na etapa anterior, para avaliar os resultados são extraídas múltiplas métricas.

O melhor resultado obtido para a tarefa de segmentação foi usando a abordagem BS1 em conjunto com a rede SRGAN, obtendo um aumento de 10% no valor do coeficiente de *dice*. Na detecção, o melhor resultado alcançado foi utilizando a abordagem BS2 com a ESRGAN. Com um aumento médio de 8.5% em todas as métricas, foi alcançado 86% de acurácia, 80% de sensibilidade, 94% de precisão e 86% de *F1-Score*.

Entre algumas limitações encontradas, a do erro de melhoramento de imagens com altos índices de nuvens usando a abordagem BS2 foi a mais evidente em relação ao método. Outro problema se diz a respeito do melhoramento e segmentação de garimpos que não foram muito bem capturados devido à resolução do Landsat 8. A falta de informações precisas sobre regiões de garimpos também foi uma dificuldade, deixando a cargo do autor a tarefa de validação dessas áreas. Portanto, a quantidade de imagens usadas nesta metodologia foi limitada devido ao tempo levado para validar cada região de garimpo. Outro fator limitante é que parte das imagens utilizadas neste estudo são de propriedade privada licenciadas para instituições públicas, inviabilizando a divulgação de uma base de dados completa, o que pode dificultar na reprodução do estudo.

Como contribuição, desenvolveu-se uma técnica usando redes adversárias de super-resolução para o melhoramento da qualidade de imagens do satélite estadunidense Landsat 8 e aprimoramento de detecção das regiões de garimpo usando a rede convolucional U-Net. O uso desse método pode contribuir positivamente em estudos usando sensoriamento remoto voltados a áreas de mineração no Brasil, principalmente na região da Amazônia Legal.

Para trabalhos futuros, propõe-se utilizar alguma técnica de pré-processamento para reduzir a presença de nuvens nas imagens de satélite com o objetivo de reduzir os erros encontrados neste caso específico. Realizar um treinamento mais robusto das GANs de super-resolução, com um número bem maior de imagens e gerações, visando melhorar os resultados. Utilizar o método em imagens do recém-lançado satélite nacional Amazonia 1,¹ que possui uma resolução espacial inferior a do Landsat 8, para avaliar o impacto do método nessas imagens de satélite.

¹ <http://www.inpe.br/amazonia1/>

Referências

- AGRICULTURA, P. e. A. Ministério da. *Floresta Nacional do Crepori (PA)*. 2020. Disponível em: <<https://www.florestal.gov.br/florestas-sob-concessao/92-concessoes-florestais/florestas-sob-concessao/105-floresta-nacional-do-crepori-pa>>. Citado na página 36.
- AMBIENTE, M. do M. *APA do Tapajós*. 2020. Disponível em: <<https://www.florestal.gov.br/florestas-sob-concessao/92-concessoes-florestais/florestas-sob-concessao/105-floresta-nacional-do-crepori-pa>>. Citado na página 36.
- ASSIS, L. F. F.; FERREIRA, K. R.; VINHAS, L.; MAURANO, L.; ALMEIDA, C.; CARVALHO, A.; RODRIGUES, J.; MACIEL, A.; CAMARGO, C. Terrabrasilis: A spatial data analytics infrastructure for large-scale thematic mapping. *ISPRS International Journal of Geo-Information*, Multidisciplinary Digital Publishing Institute, v. 8, n. 11, p. 513, 2019. Citado na página 14.
- BENGIO, Y.; SIMARD, P.; FRASCONI, P. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, IEEE, v. 5, n. 2, p. 157–166, 1994. Citado na página 28.
- BISONG, E. Google colabory. In: *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. London, UK: Springer, 2019. p. 59–64. Citado na página 45.
- BOSCH, M.; GIFFORD, C. M.; RODRIGUEZ, P. A. Super-resolution for overhead imagery using densenets and adversarial learning. In: IEEE. *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*. New York, NY, 2018. p. 1414–1422. Citado na página 17.
- BRUNA, J.; SPRECHMANN, P.; LECUN, Y. Super-resolution with deep convolutional sufficient statistics. *arXiv preprint arXiv:1511.05666*, 2015. Citado na página 31.
- CAHETÉ, F. S. A extração do ouro na amazônia e implicações para o meio ambiente. *Novos Cadernos NAEA*, v. 1, n. 2, 2008. Citado na página 14.
- CIVIL, P. d. R. C. *DECRETO Nº 4.340, DE 22 DE AGOSTO DE 2002*. 2002. Disponível em: <http://www.planalto.gov.br/ccivil_03/decreto/2002/d4340.htm>. Citado na página 36.
- CRESWELL, A.; WHITE, T.; DUMOULIN, V.; ARULKUMARAN, K.; SENGUPTA, B.; BHARATH, A. A. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, IEEE, v. 35, n. 1, p. 53–65, 2018. Citado 2 vezes nas páginas 28 e 29.
- CURI, M. V. Aspectos legais da mineração em terras indígenas. *Revista de Estudos e Pesquisas, FUNAI*, v. 4, n. 2, p. 221–252, 2007. Citado na página 15.
- DE-PAULA, V. G.; LAMAS-CORRÊA, R.; TUTUNJI, V. L. Garimpo e mercúrio: impactos ambientais e saúde humana. *Universitas: Ciências da Saúde*, v. 4, n. 1, p. 101–110, 2006. Citado na página 15.

DEMIRKESEN, A.; EVRENDILEK, F.; BERBEROGLU, S.; KILIC, S. Coastal flood risk analysis using landsat-7 etm+ imagery and srtm dem: A case study of izmir, turkey. *Environmental monitoring and assessment*, Springer, v. 131, n. 1, p. 293–300, 2007. Citado na página 37.

DODGE, S.; KARAM, L. Understanding how image quality affects deep neural networks. In: IEEE. *2016 eighth international conference on quality of multimedia experience (QoMEX)*. New York, NY, 2016. p. 1–6. Citado na página 40.

DONG, C.; LOY, C. C.; HE, K.; TANG, X. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 38, n. 2, p. 295–307, 2015. Citado na página 31.

DONG, H.; SUPRATAK, A.; MAI, L.; LIU, F.; OEHMICHEN, A.; YU, S.; GUO, Y. TensorLayer: A Versatile Library for Efficient Deep Learning Development. *ACM Multimedia*, 2017. Disponível em: <<http://tensorlayer.org>>. Citado na página 45.

EICKHOFF, A. P. D. N. *ELEMENTOS-TRAÇO EM PASSIVOS AMBIENTAIS DE GARIMPO DE OURO: UMA AVALIAÇÃO DA POSSIBILIDADE DE CONTAMINAÇÃO DE SOLOS, SEDIMENTOS, PLANTAS E PEIXES EM ÁREAS DE PISCICULTURAS EM POCONÉ E ALTO PARAGUAI-MT*. Tese (Doutorado) — Universidade do Estado de Mato Grosso, 2011. Citado na página 39.

FADNAVIS, S. Image interpolation techniques in digital image processing: an overview. *International Journal of Engineering Research and Applications*, v. 4, n. 10, p. 70–73, 2014. Citado na página 41.

FEDERAL, G. 2020. Disponível em: <<https://www.gov.br/mma/pt-br>>. Citado na página 35.

FERNANDES, S. S.; FILHO, P. M. Educação ambiental enquanto instrumento para redução de impactos ambientais em garimpos. *Revista Práxis: saberes da extensão*, v. 4, n. 6, p. 59–66, 2016. Citado na página 15.

GOODFELLOW, I. J.; POUGET-ABADIE, J.; MIRZA, M.; XU, B.; WARDE-FARLEY, D.; OZAI, S.; COURVILLE, A.; BENGIO, Y. Generative adversarial networks. *arXiv preprint arXiv:1406.2661*, 2014. Citado na página 29.

GORELICK, N.; HANCHER, M.; DIXON, M.; ILYUSHCHENKO, S.; THAU, D.; MOORE, R. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, Elsevier, v. 202, p. 18–27, 2017. Citado na página 36.

GREENPEACE, B. *Destruição sem fim*. 2020. Disponível em: <<https://www.greenpeace.org/brasil/blog/destruicao-sem-fim/>>. Citado na página 14.

GULRAJANI, I.; AHMED, F.; ARJOVSKY, M.; DUMOULIN, V.; COURVILLE, A. Improved training of wasserstein gans. *arXiv preprint arXiv:1704.00028*, 2017. Citado na página 17.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: *Proceedings of the IEEE international conference on computer vision*. New York, NY: IEEE, 2015. p. 1026–1034. Citado na página 30.

- HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. New York, NY: IEEE, 2016. p. 770–778. Citado na página 28.
- IHLEN, V. *Landsat 8 (L8): Data Users Handbook*. Sioux Falls, South Dakota: EROS, 2019. Citado 2 vezes nas páginas 21 e 35.
- INPE. *RAPIDEYE*. 2013. Disponível em: <<http://www.dgi.inpe.br/documentacao/satelites/rapideye>>. Citado na página 21.
- IOFFE, S.; SZEGEDY, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: PMLR. *International conference on machine learning*. Lille, FR, 2015. p. 448–456. Citado na página 30.
- ISENSEE, F.; PETERSEN, J.; KLEIN, A.; ZIMMERER, D.; JAEGER, P. F.; KOHL, S.; WASSERTHAL, J.; KOEHLER, G.; NORAJITRA, T.; WIRKERT, S. et al. nnu-net: Self-adapting framework for u-net-based medical image segmentation. *arXiv preprint arXiv:1809.10486*, 2018. Citado na página 42.
- JENSEN, J. *Remote sensing of the environment : an earth resource perspective*. Upper Saddle River, NJ: Pearson Prentice Hall, 2007. ISBN 0131889508. Citado 2 vezes nas páginas 19 e 20.
- JIANG, K.; WANG, Z.; YI, P.; JIANG, J.; WANG, G.; HAN, Z.; LU, T. Gan-based multi-level mapping network for satellite imagery super-resolution. In: IEEE. *2019 IEEE International Conference on Multimedia and Expo (ICME)*. New York, NY, 2019. p. 526–531. Citado na página 17.
- JOHNSON, J.; ALAHI, A.; FEI-FEI, L. Perceptual losses for real-time style transfer and super-resolution. In: SPRINGER. *European conference on computer vision*. New York, NY, 2016. p. 694–711. Citado na página 31.
- JOLICOEUR-MARTINEAU, A. The relativistic discriminator: a key element missing from standard gan. *arXiv preprint arXiv:1807.00734*, 2018. Citado na página 33.
- JORDAHL, K. Geopandas: Python tools for geographic data. URL: <https://github.com/geopandas/geopandas>, 2014. Citado na página 39.
- LECUN, Y. et al. Generalization and network design strategies. *Connectionism in perspective*, Elsevier Zurich, Switzerland, v. 19, p. 143–155, 1989. Citado na página 24.
- LEDIG, C.; THEIS, L.; HUSZÁR, F.; CABALLERO, J.; CUNNINGHAM, A.; ACOSTA, A.; AITKEN, A.; TEJANI, A.; TOTZ, J.; WANG, Z. et al. Photo-realistic single image super-resolution using a generative adversarial network. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. New York, NY: IEEE, 2017. p. 4681–4690. Citado 5 vezes nas páginas 17, 30, 31, 40 e 45.
- LIM, B.; SON, S.; KIM, H.; NAH, S.; LEE, K. M. Enhanced deep residual networks for single image super-resolution. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. New York, NY: IEEE, 2017. p. 136–144. Citado 2 vezes nas páginas 32 e 33.

- LUO, Y.-Y.; LU, H.-G.; JIA, N. Super-resolution algorithm of satellite cloud image based on wgan-gp. In: IEEE. *2019 International Conference on Meteorology Observations (ICMO)*. New York, NY, 2019. p. 1–4. Citado na página 17.
- MAHAPATRA, D.; BOZORGTABAR, B.; HEWAVITHARANAGE, S.; GARNAVI, R. Image super resolution using generative adversarial networks and local saliency maps for retinal image analysis. In: SPRINGER. *International conference on medical image computing and computer-assisted intervention*. New York, NY, 2017. p. 382–390. Citado 2 vezes nas páginas 17 e 18.
- MICROSOFT. Microsoft, 2021. Disponível em: <<https://www.bing.com/maps>>. Citado 3 vezes nas páginas 8, 38 e 39.
- MINERACAO, O. da. *Amazônia tem 4,5 mil locais de garimpo ilegal identificados, mais da metade no Brasil - Observatório da Mineração*. 2020. <<https://observatoriodamineracao.com.br/amazonia-tem-45-mil-locais-de-garimpo-ilegal-identificados-mais-da-metade-no-brasil/>>. (Accessed on 04/24/2021). Citado na página 15.
- MURPHY, J. An overview of convolutional neural network architectures for deep learning. *Microway Inc*, 2016. Citado 2 vezes nas páginas 25 e 26.
- NASA, U. *Thermal Infrared Sensor (TIRS)*. NASA, 2012. Disponível em: <<https://landsat.gsfc.nasa.gov/landsat-8/thermal-infrared-sensor-tirs>>. Citado na página 21.
- NIELSEN, M. A. *Neural networks and deep learning*. San Francisco, CA: Determination press, 2015. v. 25. Citado 3 vezes nas páginas 23, 24 e 26.
- NOVO, E. *Introdução ao sensoriamento remoto*. São José dos Campos - São Paulo: Instituto Nacional de Pesquisas Espaciais, 2001. Citado 3 vezes nas páginas 20, 21 e 37.
- QGIS. 2021. Disponível em: <<https://github.com/qgis/QGIS/releases>>. Citado na página 37.
- RADFORD, A.; METZ, L.; CHINTALA, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015. Citado na página 31.
- RAKHLIN, A.; DAVYDOW, A.; NIKOLENKO, S. Land cover classification from satellite imagery with u-net and lovasz-softmax loss. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. New York, NY: IEEE, 2018. p. 262–266. Citado 2 vezes nas páginas 42 e 43.
- RAPIDEYE. *RapidEye Satellite Sensors*. 2013. Disponível em: <<https://www.satimagingcorp.com/satellite-sensors/other-satellite-sensors/rapideye/>>. Citado 2 vezes nas páginas 22 e 35.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. *International Conference on Medical image computing and computer-assisted intervention*. New York, NY, 2015. p. 234–241. Citado 2 vezes nas páginas 26 e 27.

- ROSSUM, G. V.; DRAKE, F. L. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009. ISBN 1441412697. Citado na página 39.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *nature*, Nature Publishing Group, v. 323, n. 6088, p. 533–536, 1986. Citado na página 24.
- SIMIONATO, J. et al. Identificação de áreas de garimpo na floresta nacional do crepori (pará) através de análise de imagens baseada em objetos geográficos (geobia) e mineração de dados. Florianópolis, SC, 2019. Citado na página 37.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. Citado 2 vezes nas páginas 28 e 31.
- SOCIOAMBIENTAL, I. *Cicatrizes na floresta: garimpo avançou 30% na Terra Indígena Yanomami em 2020 | ISA - Instituto Socioambiental*. 2020. <<https://www.socioambiental.org/pt-br/noticias-socioambientais/cicatrizes-na-floresta-garimpo-avancou-30-na-terra-indigena-yanomami-em-2020>>. (Accessed on 04/23/2021). Citado na página 15.
- SZE, V.; CHEN, Y.-H.; YANG, T.-J.; EMER, J. S. Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE*, Ieee, v. 105, n. 12, p. 2295–2329, 2017. Citado na página 39.
- TAN, W.; LIU, P.; LI, X.; LIU, Y.; ZHOU, Q.; CHEN, C.; GONG, Z.; YIN, X.; ZHANG, Y. Classification of covid-19 pneumonia from chest ct images based on reconstructed super-resolution images and vgg neural network. *Health Information Science and Systems*, Springer, v. 9, n. 1, p. 1–12, 2021. Citado na página 18.
- US, F. G. 2020. Disponível em: <<https://www.usgs.gov/>>. Citado na página 21.
- WANG, X.; YU, K.; CHAN, K. C.; DONG, C.; LOY, C. C. *BasicSR*. 2020. <<https://github.com/xinntao/BasicSR>>. Citado na página 45.
- WANG, X.; YU, K.; WU, S.; GU, J.; LIU, Y.; DONG, C.; QIAO, Y.; LOY, C. C. Esrgan: Enhanced super-resolution generative adversarial networks. In: *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*. Glasgow, SCO: Springer, 2018. p. 0–0. Citado 5 vezes nas páginas 32, 33, 34, 40 e 45.
- WARMERDAM, F. The geospatial data abstraction library. In: *Open source approaches in spatial data handling*. New York, NY: Springer, 2008. p. 87–104. Citado na página 39.
- XU, B.; WANG, N.; CHEN, T.; LI, M. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*, 2015. Citado na página 27.
- ZEILER, M. D. Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012. Citado na página 42.
- ZOU, J.; HAN, Y.; SO, S.-S. Overview of artificial neural networks. *Artificial Neural Networks*, Springer, p. 14–22, 2008. Citado na página 22.