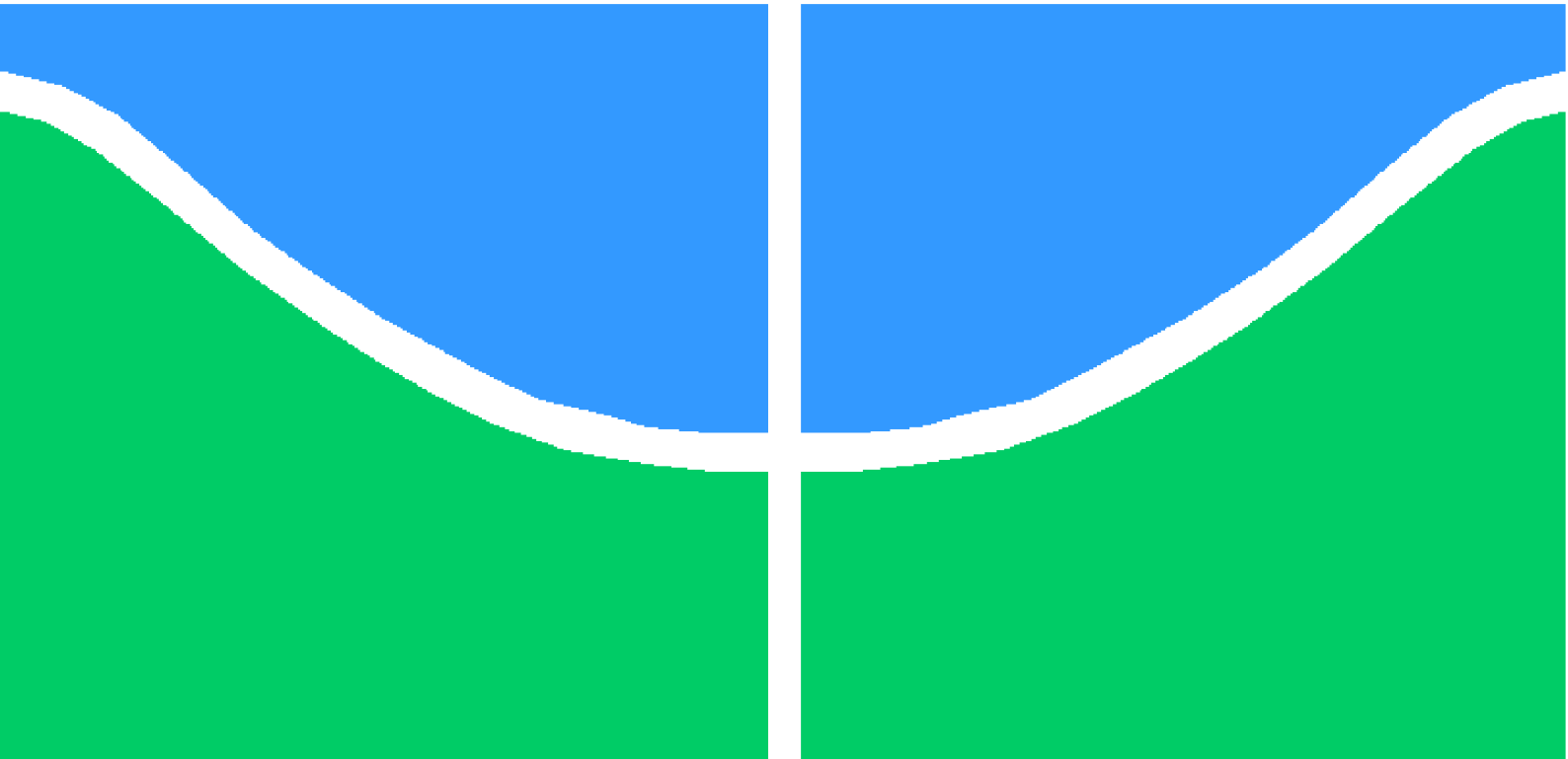




Universidade de Brasília - UnB
Faculdade de Ciências e Tecnologias em Engenharia- FCTE
Engenharia Aeroespacial

Título: Detecção de embarcações no Rio Amazonas por meio de Radar de Abertura Sintética

Autor: José Alejandro López Lorenzo
Orientador: (Dr. William Reis Silva)

Brasília, DF
2025



José Alejandro López Lorenzo

**Título: Detecção de embarcações no Rio Amazonas por
meio de Radar de Abertura Sintética**

Monografia submetida ao curso de graduação
em (Engenharia Aeroespacial) da Universi-
dade de Brasília, como requisito parcial para
obtenção do Título de Bacharel em (Enge-
nharia Aeroespacial).

Universidade de Brasília - UnB

Faculdade de Ciências e Tecnologias em Engenharia- FCTE

Orientador: (Dr. William Reis Silva)

Coorientador: (Dr. Fernando Merege)

Brasília, DF

2025

José Alejandro López Lorenzo

Título: Detecção de embarcações no Rio Amazonas por meio de Radar de Abertura Sintética/ José Alejandro López Lorenzo. – Brasília, DF, 2025-
122 p. : il. (algumas color.) ; 30 cm.

Orientador: (Dr. William Reis Silva)

Trabalho de Conclusão de Curso – Universidade de Brasília - UnB
Faculdade de Ciências e Tecnologias em Engenharia- FCTE , 2025.

1. SAR. 2. Embarcações. I. (Dr. William Reis Silva). II. Universidade de Brasília. III. Faculdade de Ciências e Tecnologias em Engenharia- FCTE. IV. Título: Detecção de embarcações no Rio Amazonas por meio de Radar de Abertura Sintética

CDU 02:141:005.6

José Alejandro López Lorenzo

Título: Detecção de embarcações no Rio Amazonas por meio de Radar de Abertura Sintética

Monografia submetida ao curso de graduação em (Engenharia Aeroespacial) da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em (Engenharia Aeroespacial).

Trabalho aprovado. Brasília, DF, 17 de Julho 2025:

(Dr. William Reis Silva)
Orientador

Dr. Ronne Toledo
Convidado 1

Dr. Lui Txai Calvoso Habl
Convidado 2

Brasília, DF
2025

A Dios y a todas las personas que me dieron una oportunidad y nunca
dejaron de creer en mi.

Agradecimientos

Agradezco infinitamente a Dios y a todas las personas que estuvieron a mi lado en esta larga jornada, principalmente a mis padres Edgar Armando López Gonzales y mi madre Rossina Emilia Lorenzo Hernández. Siempre me brindaron su apoyo y me instruyeron por el camino del bien.

Agradezco a mi novia Emely Morales, que me acompañó en cada momento, y con su amor incondicional me dió fuerzas y ánimos para nunca desistir de mi sueño.

Agradezco a mis hermanos Kevin y Allan López por estar cuando más los necesité. Agradezco también a mis primos, primas, tíos, tías y abuelos que nunca dejaron de enviarme mensajes de apoyo y por siempre creer en mí.

Agradezco a mis amigos y compañeros de curso, tanto nacionales como internacionales, que hicieron de todos estos años en la Universidad una experiencia inolvidable.

Agradezco a Marcus Vinícius Coelho, Erika Regina Prado, Jhonata Oliveira e Filipe Silveira, que entre estágio profesional y trabajo, me han brindado experiencias que día a día me forman como un profesional capaz y competente en mi área.

Por fin, agradezco a todos mis profesores, en especial a mi orientador William Reis Silva, mi co-orientador Fernando Merege, y a los profesores que despertaron mi interés por la Observación de la Tierra: Giancarlo Santilli Y Luciano Emidio Fonseca. Todos ellos, con paciencia y sabiduría, me supieron enseñar y transmitir conocimientos valiosos que me moldearon como un futuro Ingeniero Aeroespacial!

(David Goggins, Nada pode me Ferir) *“Foi quando percebi pela primeira vez que nem todas as limitações físicas e mentais são reais e que eu tinha o hábito de desistir muito cedo.*

Entendi também que seria preciso cada grama de coragem e força que eu conseguisse reunir para realizar o impossível.

Estava diante de horas, dias e semanas de sofrimento contínuo. Teria que me forçar até o limite mais extremo da minha mortalidade.

Tive que aceitar a possibilidade muito real de talvez morrer, porque dessa vez eu não iria desistir, por mais que meu coração disparasse e por maior que fosse a dor.

(David Goggins, Nada pode me Ferir)

Resumo

O sensoriamento remoto desempenha papel fundamental no monitoramento ambiental no Brasil, um país de dimensões continentais que concentra aproximadamente 50% das terras da América do Sul e 81% de seus recursos hídricos na Região Hidrográfica Amazônica. O rio Amazonas, principal via fluvial da região, conecta localidades estratégicas que vão do Peru até os estados do Pará e Amapá. A tríplice fronteira entre Brasil, Peru e Colômbia destaca-se como uma das áreas mais vulneráveis à movimentação de atividades ilícitas, majoritariamente realizadas por transporte marítimo. Nesse contexto, a detecção e o monitoramento de embarcações por meio de satélites, em especial utilizando Radar de Abertura Sintética (SAR), tornam-se ferramentas indispensáveis para reforçar as ações de controle e combate às atividades ilícitas realizadas por meio de meios aquáticos. Nesse sentido, este estudo propõe a análise do modelo YOLOv12, em suas variantes **n**, **s** e **m**, para a tarefa de detecção de embarcações. Este documento aborda tanto a elaboração de uma base de dados proveniente do satélite de radar Sentinel-1 quanto o treinamento dos modelos mencionados. A base de dados resultante compreende 359 imagens e 674 amostras de embarcações. Os resultados, avaliados por meio das métricas de desempenho, mostram que a variante YOLOv12 **s** alcançou precisão de 72,00%, revocação de 65,06% e *F1-Score* de 68,35% para a cena complexa de Manaus, utilizando um limiar de confiança de 0,50. Para essa mesma cena, com um limiar de 0,25, os valores de precisão, revocação e *F1-Score* foram de 67,89%, 68,67% e 68,26%, respectivamente. Os resultados evidenciam que a variante YOLOv12 **s** apresentou desempenho superior em comparação às variantes **n** e **m**, sendo, portanto, o modelo selecionado para a detecção de embarcações nas áreas de interesse. A partir das detecções realizadas, foram gerados arquivos vetoriais contendo informações relevantes sobre cada embarcação identificada. Por fim, conclui-se que o uso desse tipo de solução tecnológica contribui para o avanço e o fortalecimento de iniciativas espaciais voltadas à segurança pública.

Palavras-chaves: Sensoriamento Remoto. Rio Amazonas. Tríplice fronteira. SAR. Base de dados. YOLOv12. Sentinel-1.

Abstract

Remote sensing plays a fundamental role in environmental monitoring in Brazil, a country of continental dimensions that contains approximately 50% of the land area of South America and 81% of its water resources within the Amazon Hydrographic Region. The Amazon River, the main fluvial route of the region, connects strategic locations ranging from Peru to the states of Pará and Amapá. The tri-border area between Brazil, Peru, and Colombia stands out as one of the most vulnerable regions to illicit activities, which are mostly carried out through maritime transport. In this context, the detection and monitoring of vessels by means of satellites, especially using Synthetic Aperture Radar (SAR), become indispensable tools to strengthen control actions and combat illicit activities conducted through aquatic routes. In this sense, this study proposes the analysis of the YOLOv12 model, in its **n**, **s** and **m** variants, for the task of vessel detection. This document addresses both the construction of a database derived from Sentinel-1 radar satellite imagery and the training of the aforementioned models. The resulting database comprises 359 images and 674 vessel samples. The results, evaluated through performance metrics, show that the YOLOv12 **s** variant achieved a precision of 72.00%, recall of 65.06% and *F1-Score* of 68.35% for the complex scene of Manaus, using a confidence threshold of 0.50. For this same scene, but with a threshold of 0.25, the values of precision, recall and *F1-Score* reached 67.89%, 68.67% and 68.26%, respectively. The results show that the YOLOv12 **s** variant presented superior performance compared to the **n** and **m** variants, being therefore the chosen model for vessel detection in the areas of interest, generating from the detection results vector files containing important information about each vessel. Finally, it is concluded that the use of this type of technological solution contributes to the development and strengthening of space-based initiatives in favor of public security

Key-words: Remote Sensing. Amazon River. Tri-border area. SAR. Dataset. YOLOv12. Sentinel-1.

Lista de ilustrações

Figura 1 – Região Hidrográfica Amazônica (KMUSSER, 2013)	22
Figura 2 – Tríplice Fronteira Amazônica (OCHA, 2020)	23
Figura 3 – Rotas marítimas possíveis no Corredor Colômbia - Brasil. (Agência Cenarium, 2025)	24
Figura 4 – Satélite Sentinel 1. Desenvolvido pela European Space Agency (ESA). (European Space Agency, 2025)	25
Figura 5 – Componentes do Sensoriamento Remoto. (ZANNOTTA et al., 2019) . .	33
Figura 6 – Onda Eletromagnética. Campo elétrico e magnético em fase perpendiculares entre si. (ACADEMY, 2025)	33
Figura 7 – Espectro eletromagnético. Em destaque a região do visível. (DORNELLES, 2008)	35
Figura 8 – Janelas atmosféricas no Espectro eletromagnético. (GOMES, 2006) . .	36
Figura 9 – Geometria de imageamento de um sistema SAR	38
Figura 10 – A resolução em range depende exclusivamente da duração do pulso τ , ou do seu derivado comprimento de pulso $\overline{PL}$. Se a distância Δr_0 entre A e B for menor a $\overline{PL}/2$, os sinais não serão separáveis.	39
Figura 11 – Visualização clara dos parâmetros usados para os calculos das resoluções.	43
Figura 12 – Esquema que representa a organização da inteligência artificial e exemplos de produtos. Nessa representação, ANN, CNN, RNN correspondem, respectivamente, a redes neurais artificiais (<i>artificial neural networks</i>), redes neurais convolucionais (<i>convolutional neural networks</i>) e redes neurais recorrentes (<i>recurrent neural networks</i>). Adaptado de (CHENG; BRADLEY, 2023)	45
Figura 13 – Estrutura de um neurônio humano (Khan Academy, 2025).	46
Figura 14 – Arquitetura da rede Perceptron	47
Figura 15 – Representação de uma imagem de uma pessoa por meio de um sistema de aprendizado profundo (GOODFELLOW; BENGIO; COURVILLE, 2016).	48
Figura 16 – Arquitetura do modelo transformador original. Adaptado de (VASWANI et al., 2017)	49
Figura 17 – Operação de convolução (PODAREANU et al., 2019)	51
Figura 18 – Filtros por meio de convolução	52
Figura 19 – Operação <i>max pooling</i> por meio de uma janela 2x2	52
Figura 20 – Etapas para detectar objetos no modelo YOLO original (REDMON et al., 2016)	54
Figura 21 – Arquitetura do modelo Yolov12.	56

Figura 22 – Diagrama da arquitetura do bloco A2C2F , implementação do YoloV12 que aumenta a robustez na extração de características.	57
Figura 23 – Representação espacial do tensor após operação de redimensionamento e aplanamento. Extraído de (TORNERO, 2025).	58
Figura 24 – Matriz resultante de valores de atenção. Extraído de (TORNERO, 2025).	59
Figura 25 – Tensor resultante da operação de soma entre o tensor de atenção e o tensor de codificação posicional. Extraído de (TORNERO, 2025).	59
Figura 26 – Etapas para realização do trabalho.	61
Figura 27 – Exemplos das áreas de aquisição do Sentinel-1 para as regiões de interesse. (a) Tabatinga, (b) Delta do Amazonas, e (c) Rio Solimões.	63
Figura 28 – Representação das subfaixas no modo IW (ØSTERGAARD et al., 2013).	63
Figura 29 – Movimento de trás para frente do feixe (Earth Science Data Systems, NASA, 2016).	63
Figura 30 – Mapa da Região de Interesse (ROI).	64
Figura 31 – Alguns exemplos de imagens de visualização rápida (superior) e imagens SAR (inferior).	65
Figura 32 – Fluxo de pré-processamento das imagens SAR e seleção de amostras de treinamento.	68
Figura 33 – Exemplos de limiarização usando o método de Otsu. No topo, observamos uma imagem referente a Leticia, nela, foi aplicado o limiar de 0,482. Embaixo, temos uma imagem representando uma zona do Delta do Amazonas. O limiar encontrado ideal para esta imagem foi de 0,498.	70
Figura 34 – Remoção do ruído termal na polarização VH. Temos na direita cena, imagem e histograma sem remoção do ruído termal. Na esquerda, os elementos com remoção do ruído termal.	72
Figura 35 – Representação física das unidades de calibração, sendo σ_0 a mais utilizada para estudos quantitativos (GUMELAR; LIU; RIZKIYANTO, 2017).	73
Figura 36 – Distorções geométricas ilustradas em (a), (b), (c), para Encurtamento, inversão do terreno e sombreamento, respectivamente. Em (d),(e),(f), observa-se exemplos reais (PATRUNO, 2014).	75
Figura 37 – Correção do terreno. (a) Imagem sem correção do terreno. (b) Imagem com correção de terreno. Se obtém uma imagem espelhada e rotacionada, pois esse é o formato de aquisição da imagem por parte do satélite.	76
Figura 38 – Mecanismos principais de retroespalhamento em sistemas SAR (FLORES et al., 2019).	76
Figura 39 – Mecanismos presentes em alvos do tipo embarcações (WANG et al., 2022).	77
Figura 40 – Histograma de intensidade para retroespalhamento do tipo refletor de canto em polarizações VV (esquerda) e VH (direita).	78

Figura 41 – Amostras de treinamento anotadas na plataforma <i>Roboflow</i>	80
Figura 42 – Abordagem para recorte, selecção e anotação de amostras de treinamento.	80
Figura 43 – Exemplo do formato das anotações. Na imagem da embarcação, o ponto azul na posição x,y representa o centro da caixa delimitadora. O quadro vermelho tem as dimensões de largura e altura representado no texto.	81
Figura 44 – Formato dos tensores da variante <i>Nano</i>	82
Figura 45 – Gráficos de precisão em função das épocas. (a) precisão em função das épocas para a variante n , (b) precisão em função das épocas para a variante s , (c) precisão em função das épocas para a variante m	87
Figura 46 – Gráficos de revocação em função das épocas. (a) revocação em função das épocas para a variante n , (b) revocação em função das épocas para a variante s , (c) revocação em função das épocas para a variante m	88
Figura 47 – Gráficos de <i>F1-score</i> em função das épocas. (a) <i>F1-score</i> em função das épocas para a variante n , (b) <i>F1-score</i> em função das épocas para a variante s , (c) <i>F1-score</i> em função das épocas para a variante m	88
Figura 48 – Gráficos de mAP0,50 em função das épocas. (a) mAP0,50 em função das épocas para a variante n , (b) mAP0,50 em função das épocas para a variante s , (c) mAP0,50 em função das épocas para a variante m	89
Figura 49 – Gráficos de mAP0,50:0,95 em função das épocas. (a) mAP0,50:0,95 em função das épocas para a variante n , (b) mAP0,50:0,95 em função das épocas para a variante s , (c) mAP0,50:0,95 em função das épocas para a variante m	90
Figura 50 – Gráfico de precisão versus confiança. (a) precisão em função da confiança para a variante n , (b) precisão em função da confiança para a variante s , (c) precisão em função da confiança para a variante m	91
Figura 51 – Gráfico de revocação versus confiança. (a) revocação em função da confiança para a variante n , (b) revocação em função da confiança para a variante s , (c) revocação em função da confiança para a variante m	91
Figura 52 – Gráfico de <i>F1-Score</i> versus confiança. (a) <i>F1-Score</i> em função da confiança para a variante n , (b) <i>F1-Score</i> em função da confiança para a variante s , (c) <i>F1-Score</i> em função da confiança para a variante m	92
Figura 53 – Gráfico de curva precisão-revocação. (a) curva precisão-revocação para a variante n , (b) curva precisão-revocação para a variante s , (c) curva precisão-revocação para a variante m	93
Figura 54 – Matriz de confusão gerada através da inferência no teste de validação. (a) Matriz de confusão para a variante n , (b) Matriz de confusão para a variante s , (c) Matriz de confusão para a variante m	93

Figura 55 – Inferência no conjunto de validação usando os três modelos treinados. Nas imagens mostradas, os três tamanhos obtiveram detecções coerentes à verdade de campo.	95
Figura 56 – Falhas na detecção de objetos na inferência no conjunto de validação usando os três modelos treinados.	95
Figura 57 – Cena para realização de teste de generalização na área do Rio Negro, Manaus, com imagem Sentinel-1.	97
Figura 58 – Teste de generalização usando o modelo s com limiar de confiança de 0,25, na área do Rio Negro, Manaus com imagem Sentinel-1. Em (a) observa-se no detalhe as embarcações marcadas como verdadeiros positivos (TP). Em (b), alvos marcados erroneamente como embarcações, representando falsos positivos.	98
Figura 59 – Cena para realização de teste de generalização na área do Delta do Amazonas com imagem Sentinel-1.	100
Figura 60 – Teste de generalização usando o tamanho n , na área do Delta do Amazonas com imagem Sentinel-1. Em (a) observa-se no detalhe as embarcações marcadas como verdadeiros positivos (TP). Em (b), ilhas marcadas como embarcações, representando falsos positivos.	101
Figura 61 – Processo de extração de atributos na cena referente ao Delta do Amazonas. Em (a) observa-se um recorte da cena do Delta do Amazonas. Em (b), observa-se a embarcação alvo na imagem SAR. Em (c), apresenta-se a imagem binarizada pelo método de Otsu. Em (d), são projetados sobre a imagem binarizada os eixos principal (comprimento) e secundário (largura). Por fim, em (e), é exibido um possível modelo da embarcação detectada pelo algoritmo.	105
Figura 62 – Processo de extração de atributos na cena referente a Manaus. Em (a) observa-se um recorte da cena. Em (b), observa-se a embarcação alvo na imagem SAR. Em (c), apresenta-se a imagem binarizada pelo método de Otsu. Em (d), são projetados sobre a imagem binarizada os eixos principal (comprimento) e secundário (largura), além da legenda com os valores estimados de comprimento e largura, 28 metros para ambos. Por fim, em (e), é exibido um possível modelo da embarcação detectada pelo algoritmo.	106
Figura 63 – Histogramas de comprimento e largura estimados para as cenas de Manaus (a,b) e Delta do Amazonas (c-d)	107
Figura 64 – Pontos vetoriais exportados e projetados no QGIS. em (a) pontos sobrepostos sobre a cena de Manaus e (b) pontos sobre a cena de Delta do Amazonas.	107

Figura 65 – Tabela de atributos dos pontos vetoriais exportados e projetados no QGIS para a cena de Delta do Amazonas. 108

Figura 66 – Experimento final utilizando o modelo YOLOv12-s. Observa-se um desempenho consistente, com detecção de embarcações no centro e nas extremidades da cena, abrangendo diferentes tamanhos, e com a maioria das predições apresentando confiança superior a 40%. 108

Lista de tabelas

Tabela 1 – Faixas de frequência e comprimento de onda das bandas de RADAR .	37
Tabela 2 – Softwares utilizados e suas aplicações	62
Tabela 3 – Parâmetros orbitais do satélite Sentinel-1	62
Tabela 4 – Parâmetros importantes do modo IW.	64
Tabela 5 – Cenas baixadas para cada região de interesse.	66
Tabela 6 – Informações básicas sobre algumas cenas.	66
Tabela 7 – Intensidade relativa de espalhamento por polarização (FLORES et al., 2019).	77
Tabela 8 – Resumo dos tamanhos dos modelos YOLOv12.	82
Tabela 9 – Comparação de métricas de avaliação dos dados de validação entre modelos YOLOv12. Em vermelho, os valores que atingiram os máximos entre os três modelos.	94
Tabela 10 – Comparação de métricas de avaliação Precisão e Revocação entre modelos YOLOv12, com confiança 0,50 para o tamanho n , 0,30 para o tamanho s e 0,40 tamanho m . Cena extraída de uma área do Rio Negro, Manaus	98
Tabela 11 – Comparação das métricas de avaliação <i>F1-Score</i> , mAP 0,50 e mAP 0,50-0,95 entre modelos YOLOv12, com confiança 0,50 para o tamanho n , 0,30 para o tamanho s e 0,40 tamanho m . Cena extraída de uma área do Rio Negro, Manaus	99
Tabela 12 – Comparação de métricas de avaliação Precisão e Revocação entre modelos YOLOv12 com limiar de confiança fixo de 0,25 para todos os tamanhos.Cena extraída de uma área do Rio Negro, Manaus	99
Tabela 13 – Comparação das métricas de avaliação <i>F1-Score</i> , mAP 0,50 e mAP 0,50-0,95 entre modelos YOLOv12, limiar de confiança fixo de 0,25 para todos os tamanhos.Cena extraída de uma área do Rio Negro, Manaus	99
Tabela 14 – Comparação de métricas de avaliação Precisão e Revocação entre modelos YOLOv12, com confiança 0,50 para o tamanho n , 0,30 para o tamanho s e 0,40 tamanho m . Cena extraída do Delta do Amazonas .	101
Tabela 15 – Comparação das métricas de avaliação <i>F1-Score</i> , mAP 0,50 e mAP 0,50-0,95 entre modelos YOLOv12, com confiança 0,50 para o tamanho n , 0,30 para o tamanho s e 0,40 tamanho m . Cena extraída do Delta do Amazonas	101

Tabela 16 – Comparação de métricas de avaliação Precisão e Revocação entre modelos YOLOv12 com limiar de confiança fixo de 0,25 para todos os tamanhos. Cena extraída do Delta do Amazonas	102
Tabela 17 – Comparação das métricas de avaliação <i>F1-Score</i> , mAP 0,50 e mAP 0,50-0,95 entre modelos YOLOv12, limiar de confiança fixo de 0,25 para todos os tamanhos. Cena extraída do Delta do Amazonas.	102

Lista de abreviaturas e siglas

CNN	Rede Neural Convolucional
SAR	Radar de Abertura Sintética.
EO	Observação da Terra
ASPRS	American Society for Photogrammetry and Remote Sensing
RPA	Aeronave Remotamente Tripulada
RADAR	Radio Detection and Ranging
SLAR	Side Looking Airborne Radar
LFM	Modulação Linear de Frequência
SNAP	Sentinel Applications Plattform
IW	Interferometric Wide Swath Mode
GRD	Ground Range Detected
PR	Probabilidade de classe
IoU	Intersecção sobre união
TP	Verdadeiro Positivo
FP	Falso Positivo
FN	Falso Negativo
P	Precisão
R	Revocação
AP	Precisão Média
mAP50	Precisão média-média em um limiar de IoU de 0,50
mAP50:95	Precisão média-média avaliada em um limiar de IoU de 0,50 a 0,95.

Lista de símbolos

λ	comprimento de onda.
f	Frequência.
c	Velocidade da luz.
Q	Energia de um quantum.
h	Constante de Planck
G	Ganho da antena.
I	Intensidade radiante.
Ω	Ângulo sólido.
θ	Ângulo zenital.
φ	Ângulo Azimutal.
P_T	Potência total transmitida.
L	Radiância Espectral
A	Superfície.
ν	Largura de banda.
A_e	Superfície coletora.
P_r	Potência recebida pela antena.
P_i	Densidade de potência recebida por um alvo.
σ	Seção Transversal de Radar .
σ_0	Seção Transversal de Radar normalizada.
r_0	Distância entre satélite e o solo.
δ_a	Resolução em azimuth.
d_a	Diâmetro da antena.
τ	Duração de pulso do radar

Δr_0	Distância entre dois objetos.
δ_r	Resolução em Slant Range.
W	Bandwidth.
E_{Rx}	Energia do sinal recebido.
E_{Tx}	Energia do sinal transmitido.
P	Potência de pico.
K	Chirp Rate.
t	Fast time
A	Amplitude do LFM transmitido.
f_0	Frequência central de operação do radar.
$rect()$	Função retangular.
s_r	Sinal atenuado
g_r	Refletividade da superfície iluminada
$s_{recebido}$	Sinal recebido
θ_a	Largura de feixe.
L_{sa}	Comprimento da abertura sintética.
θ_{sa}	Largura de feixe em função da abertura sintética.
δ_a	Resolução em azimuth.
Q	Consulta
K	Chave
Y	Valor
I	Imagem bidimensional
K	Núcleo
$q_1(t)$	Probabilidade acumulada de ocorrência de pixel.
$P(i)$	Probabilidade de ocorrência do nível de cinza i.
$\mu(t)$	Média de classe.

Sumário

1	INTRODUÇÃO	22
1.1	Contextualização	22
1.1.1	Motivação	24
1.2	Objetivos	25
1.2.1	Objetivo geral	25
1.2.2	Objetivos específicos	26
1.3	Justificativa	26
1.4	Organização do Trabalho	27
2	REVISÃO DE LITERATURA	29
3	SENSORIAMENTO REMOTO	32
3.1	Sensoriamento Remoto: Conceitos Gerais	32
3.1.0.1	Comprimento de Onda e Frequência	33
3.1.0.2	Espectro Eletromagnético	34
3.2	Sensoriamento Remoto de Micro-ondas	36
3.3	SAR: Radar de Abertura Sintética	37
3.3.1	Princípio de Medição por Radar	38
3.3.2	Resolução Espacial	38
3.3.2.1	Resolução em Slant Range	39
3.3.2.2	Resolução em Azimuth	42
4	INTELIGÊNCIA ARTIFICIAL	44
4.1	Aprendizado de Máquina	44
4.2	Redes Neurais Artificiais	46
4.3	Aprendizado Profundo	47
4.3.1	Transformadores	48
4.3.2	Redes Neurais Convolucionais	49
4.3.3	YOLO	53
4.3.4	YOLOV12	55
5	METODOLOGIA	61
5.1	Softwares auxiliares	61
5.2	Base de dados SAR	62
5.2.1	Região de interesse e zonas de aquisição das imagens	64
5.3	Pré-processamento das imagens	66

5.3.0.1	Recortes (<i>Subset</i>)	71
5.3.1	Aplicar Arquivo de Órbita	71
5.3.2	Remover ruído termal	71
5.3.3	Calibração Radiométrica.	73
5.3.4	Correção do Terreno	74
5.3.4.1	Encurtamento (<i>Foreshortening</i>)	74
5.3.4.2	Inversão do terreno (<i>Layover</i>)	74
5.3.4.3	Sombreamento (<i>Shadowing</i>)	74
5.4	Anotação de amostras	79
5.5	Treinamento do modelo YOLOV12	81
5.6	Métricas de avaliação	83
5.6.1	Precisão	84
5.6.2	Taxa de Revocação (<i>Recall</i>)	84
5.6.3	<i>F1-Score</i>	84
5.6.4	Precisão Média	84
5.6.5	mAP ₅₀	85
5.6.6	mAP _{50:95}	85
5.7	Estimativa de atributos geométricos	85
5.8	Extração de pontos vetoriais	86
6	RESULTADOS E DISCUSSÕES	87
6.1	Resultados de métricas de avaliação do modelo	87
6.1.1	Métricas de avaliação versus épocas de treinamento	87
6.1.1.1	Gráficos Precisão vs épocas para os tamanho n, s e m	87
6.1.1.2	Gráficos Revocação vs épocas para os tamanho n, s e m	88
6.1.1.3	Gráficos <i>F1-Score</i> vs épocas para os tamanho s, n e m	88
6.1.1.4	Gráficos mAP _{0,50} vs épocas para os tamanhos n, s e m	89
6.1.1.5	Gráficos mAP _{0,50:0,95} vs épocas para os tamanhos n, s e m	89
6.1.2	Gráficos Precisão vs confiança para os tamanhos n, s e m	90
6.1.3	Gráficos Revocação vs confiança para os tamanhos n, s e m	91
6.1.4	Gráficos <i>F1-Score</i> vs confiança para os tamanhos n, s e m	92
6.1.5	Gráficos Precisão média (AP50) para os tamanhos n, s e m	92
6.1.6	Matrizes de confusão	93
6.1.6.1	Tamanho n, s e m	93
6.1.7	Desempenho geral por métrica	94
6.2	Teste de generalização: inferência no conjunto de validação	94
6.3	Testes em cenas Sentinel 1	96
6.3.1	Manaus	97
6.3.2	Delta do Amazonas	100
6.4	Estimativa de atributos	104

6.5	Geração de pontos vetoriais	105
6.6	Experimento final	107
7	IMPLEMENTAÇÃO FUTURA	110
8	CONCLUSÕES	111
9	TRABALHOS FUTUROS	113
	REFERÊNCIAS	114

1 Introdução

1.1 Contextualização

O Brasil é um país extenso, ocupando uma superfície de 8.515.767,049 quilômetros quadrados, e é responsável por 47,2 % da ocupação do continente sul-americano. Pelas suas dimensões continentais, o Brasil ocupa o quinto lugar no mundo, sendo superado apenas pelos Estados Unidos, Canadá, China e Rússia ([MINISTÉRIO DAS RELAÇÕES EXTERIORES, 2025](#)). Dentro dessas vastas terras, existem extensas redes de rios, a maioria localizada na Região Hidrográfica Amazônica, que ocupa 45 % do território brasileiro e concentra 81 % da disponibilidade de águas superficiais do país ([AGÊNCIA NACIONAL DE ÁGUAS E SANEAMENTO BÁSICO \(ANA\), 2025](#)). Na Região Hidrográfica Amazônica se estende o rio Amazonas, destacado na Fig. (1) o qual é considerado o rio mais extenso do planeta Terra. O rio Amazonas nasce na Cordilheira dos Andes, no Peru, e desemboca em sua foz no conhecido Delta do Amazonas, compartilhado pelos estados do Pará e Amapá. O rio Amazonas recebe diversos nomes desde o seu nascimento até a sua foz, sendo nomeado de rio Amazonas após seu encontro com o Rio Negro, no estado do Amazonas. Antes disso, o rio é conhecido como rio Solimões, no Brasil, iniciando na famosa Tríplice Fronteira Amazônica ([NATIONAL GEOGRAPHIC, 2025](#)).



Figura 1 – Região Hidrográfica Amazônica ([KMUSSE, 2013](#))

A Tríplice Fronteira Amazônica, representada na Fig. (2) é constituída por três países: Brasil, Colômbia e Peru. Mais especificamente, nessa fronteira complexa, três cidades se encontram: no Brasil, a cidade de Tabatinga, no Amazonas, que possui uma

área territorial de 3.260,245 km² e uma população estimada de 66.764 habitantes (IBGE, 2025). Na Colômbia, na cidade de Letícia, no Departamento do Amazonas, vivem 41.957 habitantes, em uma área territorial de 5.832,00 km² (DANE, 2025). Por último, a cidade de Santa Rosa de Yavari, no Peru, onde vivem cerca de 1 mil habitantes.

De acordo com Andrade e Granziera (2022), devido ao isolamento geográfico da região, as cidades, principalmente Letícia e Tabatinga, compartilham até serviços informais ou ilegais. Segundo Magalhães e Carmo (2022), nos anos de 1970, já havia registros de entrada de cocaína colombiana no território brasileiro via Letícia. Os grandes responsáveis por esse tráfico eram cartéis internacionais, como o Cartel de Medellín e o Cartel de Cali, ambos da Colômbia. Conforme Jacarandá (2024), a cidade de Tabatinga é considerada o segundo maior ponto de entrada de cocaína no Brasil. Ainda segundo o autor, entre 1995 e 2021, as apreensões de cocaína pela Polícia Federal (PF) somaram mais de 33.313 kg.

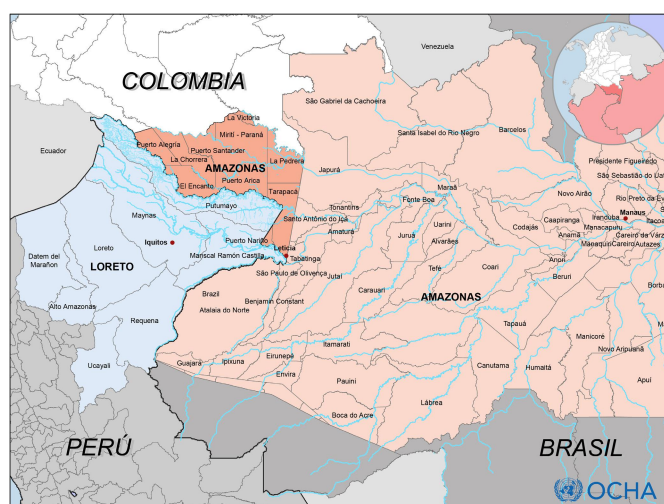


Figura 2 – Tríplice Fronteira Amazônica (OCHA, 2020)

Esse cenário é resultado de grandes disputas entre organizações criminosas pelo controle do tráfico de cocaína na região Norte do país, refletindo diretamente no aumento das taxas de homicídios nessas cidades. Dessa forma, Ribeiro (2022) afirma que cidades como Tabatinga e Letícia, no Amazonas, tornam-se entrepostos comerciais, sendo controladas e delimitadas por traficantes.

Couto (2020) discute a estrutura espacial do narcotráfico na Amazônia e descreve a dinâmica do percurso das substâncias ilícitas desde o Peru e a Colômbia até a chegada no Brasil. Em síntese, quase toda substância ilícita que entra no Brasil proveniente desses dois países chega, e é distribuído, por meio aquático. No Corredor Peru-Brasil, o produto entra pelo Acre, seguindo uma rota no rio Madeira com destino ao estado do Amazonas. A partir daí, o destino é o rio Solimões, que se conecta à capital, Manaus. Pelo rio Amazonas, a rota segue de barco pelo rio Branco até chegar a Roraima, com a finalidade de a droga atravessar a fronteira do Brasil com as Guianas e Suriname, e, assim, continuar

seu trajeto até o destino final, a Europa. Já no Corredor Colômbia-Brasil, a droga que chega de Medellín por meios aéreos até Letícia é cruzada para Tabatinga. A partir daí, utilizando a rota no rio Solimões, ilustrado na Fig. 3, o produto segue em direção a Manaus e, daí, é distribuído para o resto das conexões, sempre com a finalidade de chegar à costa atlântica, para, então, empreender viagem em direção ao mercado europeu.



Figura 3 – Rotas marítimas possíveis no Corredor Colômbia - Brasil. (Agência Cenarium, 2025)

Atualmente, já existem sistemas de monitoramento em operação no Brasil. A *Planet* fornece imagens de alta resolução ao governo brasileiro por meio do programa Brasil MAIS, o qual contempla ações voltadas à segurança pública e possibilita o acesso a imagens de satélite de alta resolução (Ministério da Justiça e Segurança Pública, 2024). Essas imagens podem ser acessadas e compartilhadas por meio de uma plataforma web, que também disponibiliza alertas de desmatamento, focos de queimadas, detecção de embarcações, entre outros (Programa Brasil MAIS, 2024). Entretanto, os sensores desses satélites operam na faixa óptica do espectro eletromagnético, a qual é fortemente afetada pela presença de nuvens e pelas condições climáticas em geral, representando uma limitação crítica na região de interesse.

1.1.1 Motivação

Com base nas informações apresentadas e considerando a dimensão continental do Brasil, bem como a necessidade de monitoramento de embarcações na região de interesse, questiona-se se é possível aplicar métodos de detecção de objetos baseados em aprendizado profundo para gerar detecções confiáveis de embarcações e, a partir delas, produzir produtos vetoriais geoespaciais que possam ser utilizados como ferramenta de monitoramento na região.

Esse questionamento será respondido ao longo deste trabalho por meio de metodologias desenvolvidas para a criação de bases de dados compostas por imagens de satélite destinadas ao treinamento de redes neurais convolucionais, especificamente do modelo YOLOv12, em suas variantes **n**, **s** e **m**.

É fundamental a implementação de soluções que possibilitem o monitoramento de grandes áreas, bem como a detecção automática de veículos aquáticos tanto embarcações legais quanto ilegais especialmente na região mais vulnerável da zona norte do país: a Tríplice Fronteira Amazônica, assim como ao longo do percurso do rio Amazonas até a costa atlântica.

Diante desse cenário, o sensoriamento remoto orbital ganha relevância. [Zannota et al. \(2019\)](#) afirma que o sensoriamento remoto orbital, realizado por meio de plataformas em órbita — geralmente satélites de Observação da Terra (EO) —, consiste na obtenção de informações da superfície terrestre a partir de imagens capturadas do espaço. Esse método utiliza a radiação eletromagnética refletida ou emitida em uma ou mais regiões do espectro eletromagnético.

Entre os satélites de Observação da Terra, destacam-se o CBERS-4A, o Landsat 9 e o Sentinel-1 (Fig. 4), sendo este último de papel fundamental ao longo do desenvolvimento deste trabalho.

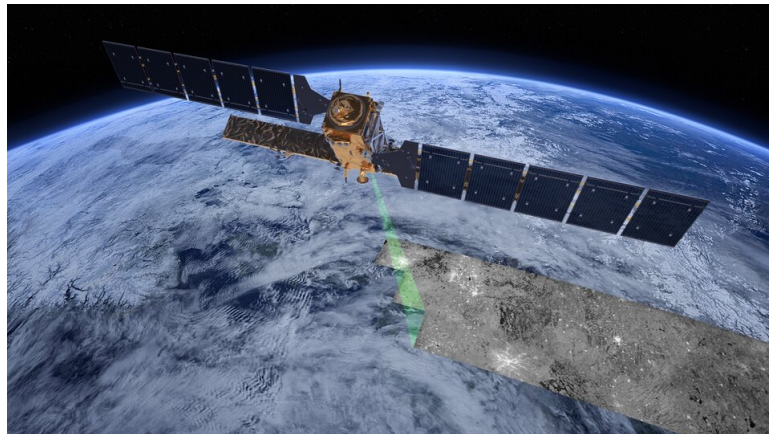


Figura 4 – Satélite Sentinel 1. Desenvolvido pela European Space Agency (ESA). ([European Space Agency, 2025](#))

1.2 Objetivos

1.2.1 Objetivo geral

O objetivo deste trabalho é aplicar uma técnica baseada em inteligência artificial, especificamente redes neurais do modelo YOLOv12, para a detecção de embarcações em

imagens de Radar de Abertura Sintética (SAR) de áreas do rio Amazonas, com especial atenção à região da Tríplice Fronteira Amazônica.

Os resultados obtidos com a aplicação dessa técnica poderão, futuramente, servir como ferramenta de apoio às medidas já existentes de combate ao tráfico de ilícitos na região.

Ao longo do trabalho, serão abordados os conceitos fundamentais de Sensoriamento Remoto e de Sistemas de Observação da Terra, bem como técnicas de processamento de imagens de satélite e a aplicação de algoritmos de aprendizado profundo para detecção de objetos em imagens de sensoriamento remoto.

1.2.2 Objetivos específicos

Revisar os conceitos fundamentais de Sensoriamento Remoto e dos Sistemas de Observação da Terra, que constituem a base teórica do presente trabalho. Definir a região de interesse e selecionar o satélite e o instrumento mais adequados para a extração das imagens que comporão o conjunto de amostras de treinamento. Processar as imagens selecionadas utilizando softwares específicos para o tratamento de produtos provenientes de sensores orbitais. Avaliar as imagens processadas e realizar a anotação das amostras de treinamento. Treinar o modelo YOLOv12 nos tamanhos **n**, **s** e **m**, utilizando as amostras previamente anotadas. Analisar o desempenho dos diferentes tamanhos do modelo por meio de métricas de avaliação e gráficos comparativos, considerando tanto o conjunto de validação quanto as imagens de generalização. Por fim, gerar arquivos vetoriais contendo informações geográficas e descritivas, como posição, comprimento, largura, data e hora da detecção e demais atributos, referentes aos navios detectados pelo modelo.

1.3 Justificativa

O desenvolvimento de pesquisas na área de sensoriamento remoto é de extrema relevância para o Brasil. O setor aeroespacial desempenha um papel fundamental na produção de plataformas espaciais, como satélites, que são essenciais para o sensoriamento remoto orbital, gerando, na maioria das vezes, imagens digitais.

Para a escolha adequada do satélite que atenderá aos objetivos da pesquisa, é indispensável que o usuário tenha um entendimento dos parâmetros orbitais (altitude, inclinação, período orbital) e dos subsistemas do satélite de sensoriamento remoto, especialmente da carga útil, que envolve características como resolução espacial e resolução radiométrica.

Os métodos aplicados ao processamento desses produtos requerem conhecimentos em diversas áreas, processamento de imagens, geoprocessamento e sistemas de informação

geográfica. Neste caso, a aplicação de técnicas de Inteligência Artificial, como aprendizado de máquina e aprendizado profundo, também se faz necessária. Por fim, os produtos gerados a partir do sensoriamento remoto são amplamente aplicados em diversas áreas, como o monitoramento, gerenciamento e gestão de recursos naturais. Essas tecnologias possibilitam, entre outras aplicações, a detecção de crimes ambientais e outras infrações, contribuindo para a proteção e preservação do território brasileiro em seus diversos aspectos.

Nesse contexto, o presente trabalho aborda técnicas de detecção de embarcações a partir de produtos de sensoriamento remoto orbital, com ênfase em dados provenientes de satélites de Radar de Abertura Sintética (SAR). Os resultados obtidos poderão servir como ferramenta de apoio para as autoridades competentes, contribuindo para a mitigação dos fluxos de atividades ilícitas na região mencionada.

1.4 Organização do Trabalho

- **Capítulo 1:** Apresenta a contextualização do tema, a motivação para o desenvolvimento da pesquisa, os objetivos geral e específicos, além da justificativa do trabalho.
- **Capítulo 2:** Neste capítulo foi realizada uma revisão da literatura focada na detecção de embarcações, trabalhos incluindo métodos tradicionais e não tradicionais, trabalhos que são fundamentais para esta pesquisa.
- **Capítulo 3:** Aborda de forma aprofundada os conceitos fundamentais de Sensoriamento Remoto, partindo de definições gerais, passando pela radiação eletromagnética e culminando na descrição detalhada dos sistemas de Radar de Abertura Sintética (SAR).
- **Capítulo 4:** Explora os principais conceitos relacionados à inteligência artificial, bem como uma explicação sobre os principais sub-ramos desta disciplina, incluindo o modelo YOLO e sua variante, YOLOV12 .
- **Capítulo 5:** Descreve a metodologia adotada no trabalho, destacando os softwares utilizados, as bases de dados selecionadas, os procedimentos de pré-processamento das imagens, o processo de anotação das amostras e todo o relacionado ao treinamento e avaliação do modelo YOLOv12.
- **Capítulo 6:** Apresenta e analisa os resultados obtidos a partir da aplicação da metodologia proposta, incluindo figuras e tabelas para melhor entendimento dos resultados.

- **Capítulo 7 e 8:** Expõe os componentes e as conclusões do trabalho, relacionando os resultados alcançados com os objetivos propostos, bem como as propostas de melhoras para serem implementadas em trabalhos futuros.

2 Revisão de literatura

Muitos trabalhos têm utilizado imagens de radar de abertura sintética em métodos de vigilância e reconhecimento marítimo. Métodos de classificação supervisionados e não supervisionados são discutidos em [GAMBA, SANO e ROCHA \(2011\)](#) para a detecção de embarcações na área de Porto de Tubarão, ES, classificando as imagens SAR em categorias como embarcações, instalações portuárias, água, vegetação e solo. Neste trabalho, classificadores como Distância Mínima, Máxima Verossimilhança e *Support Vector Machine (SVM)* demonstraram alcançar uma exatidão global superior a 80% em análises do radar aerotransportado R-99.

Outros trabalhos de classificação, como o apresentado em [Meyer, Kleynhans e Schwegmann \(2016\)](#), complementam as informações do Sistema Automático de Identificação (AIS) embarcado em embarcações. Neste trabalho os autores usam o método de classificação SVM para predecir o tamanho dos barcos contidos em uma cena Sentinel-1 EW-GRDM de baixa resolução e grande *swath*, ideal para monitoramento em oceanos.

Estudos como [Grover, Kumar e Kumar \(2018\)](#) empregam imagens SAR para detectar embarcações e/ou objetos marítimos usando o algoritmo "Búsqueda de objetos marítimos não identificados" abreviado como "SUMO". Estes objetos são identificados por meio de pixels brilhantes que apresentam alto contraste em relação ao fundo. Para isso, uma máscara de terra é aplicada à imagem para evitar alvos fora da área marítima. Logo, procede-se a detectar pixels brilhantes dentro de uma janela móvel por meio do algoritmo Taxa de Falso Alarme Constante (CFAR) com limiar adaptativo para cada região da imagem. Além disso, o algoritmo conta com remoção de ambiguidades e extração de atributos.

O método CFAR já foi amplamente utilizado para realizar a tarefa de detecção de embarcações. [Bezerra \(2021\)](#) utiliza este método para detectar alvos marítimos no Nordeste do Brasil. Descrevendo melhor essa técnica, ela consiste em buscar pixels brilhantes em comparação com a distribuição estatística do *clutter* marinho. Esta operação é feita a través das janelas do alvo, de fundo e de guarda, sendo o tamanho destas janelas escolhidas com base ao satélite utilizado. Os valores de retroespalhamento na janela de fundo são usados como entrada da função de densidade de probabilidade (PDF) para modelar estatisticamente o *clutter* marinho. O PDF, além de um valor de Probabilidade de Falso Alarme (PFA), são usados para calcular um limiar "T" que funciona como filtro entre pixels de interesse e não interesse. O autor propõe como trabalho futuro o desenvolvimento de modelos de aprendizado profunda para a realização de esta tarefa. . [Li et al. \(2019\)](#) afirma que o CFAR apresenta desvantagens como alta taxa de falsos alarmes

e baixa adaptabilidade. Isto porque o CFAR é um algoritmo de detecção em nível de pixel. Consequentemente, ignorará algumas informações estruturais importantes (distribuição da estrutura, contorno do alvo, formato das partes do alvo...), de modo que seu desempenho de generalização e sua robustez são pobres.

Zhang e Hao (2022) argumenta que ainda existem diversas limitações na tarefa de detecção de objetos em imagens SAR quando se utilizam métodos tradicionais, e que abordagens baseadas em inteligência artificial apresentam vantagens significativas. Segundo o autor, modelos de aprendizado profundo apresentam elevada capacidade de discriminação, uma vez que utilizam múltiplas camadas de rede e funções de ativação não lineares para realizar aprendizado de alto nível. Dessa forma, conseguem extrair características complexas com maior precisão e eliminar a necessidade de extração manual de atributos, reduzindo substancialmente a carga de trabalho. Além disso, quando há volume suficiente de dados para o treinamento, esses modelos apresentam maior robustez e capacidade de generalização, mantendo desempenho satisfatório mesmo em ambientes complexos. Outro ponto destacado é a forte capacidade de modelagem não linear das redes profundas, que utilizam convoluções e camadas totalmente conectadas para aprender representações estruturais e texturais de elevada complexidade.

Dentro das áreas da inteligência artificial, métodos de aprendizado profundo têm sido amplamente utilizados para a tarefa em questão. As redes neurais convolucionais (CNNs) são componentes essenciais do aprendizado profundo e são amplamente utilizadas para a detecção de objetos em imagens. Nesse contexto, Wei et al. empregaram uma versão modificada da *High-Resolution Network* para a detecção de embarcações. A abordagem inclui uma FPN (*Feature Pyramid Network*), uma RPN (*Region Proposal Network*) responsável pela geração das caixas delimitadoras candidatas das embarcações e, por fim, uma *Soft Non-Maximum Suppression*, que aprimora o desempenho da detecção ao refinar as caixas delimitadoras. Esse método apresentou uma precisão média (AP) de 63.9% quando treinado com o *backbone* HRFPN-W18, 64.5% quando treinado com O HRFPN-W32 e 64.6% quando treinado com o HRFPN-W40. Cabe mencionar que a base de dados *SAR Ship Detection Dataset* (SSDD), foi utilizada nesse trabalho. Dita base de dados é composta por 1160 imagens SAR e 2456 amostras de embarcações, geradas pelos sensores RadarSat-2, TerraSAR-X, Sentinel-1 (ZHANG et al., 2021). A pesquisa feita por Zhang e Zhang (2019) apresenta um método para detecção de embarcações (G-CNN) usando o banco de dados SSDD. Nele, a imagem SAR é separada por meio de uma grade para fazer detecções de barcos em cada uma delas. Logo, a imagem inteira é passada para o *backbone* para extração de características e logo as detecções são completadas por meio do detector. Tudo baseado em redes neurais convolucionais. Os autores mencionam que o sistema G-CNN apresentou desempenho ligeiramente inferior na detecção de embarcações pequenas e densamente agrupadas. Esta metodologia foi inspirada no modelo YOLO, principalmente no processo de separação da imagem por grades.

No mundo YOLO, [Oliveira \(2021\)](#) utilizou um modelo YOLOv4 para detectar embarcações usando imagens de radar da constelação ICEYE, em regiões como São Sebastião e Ilhabela, no litoral paulista, no estado de São Paulo. O modelo alcançou uma precisão e acurácia de 88,88% e 82,75%, respectivamente, nos testes de generalização realizados no experimento central. O autor menciona a importância da quantidade e qualidade das amostras de treinamento, além de mencionar a importância de recuperar a posição geográfica nas detecções. [Wang et al. \(2022\)](#) utilizaram uma versão modificada do YOLOX para detectar pequenas embarcações utilizando convolução multiescala (MSC), que tem como objetivo fundir informações de características em diferentes escalas para melhorar a extração de características, além do Módulo Transformador de Características (FTM), que é projetado para capturar características globais e conectá-las ao contexto, com o objetivo de otimizar a informação semântica nas camadas superiores e, assim, alcançar excelente desempenho de detecção. Este artigo utilizou as bases de dados do HRSID, que conta com 16.951 amostras de embarcações distribuídas em 6504 imagens SAR ([WEI et al., 2020](#)), para realizar testes, alcançando um AP de 59,5% com imagem de entrada de 800x800 pixels.

As bases de dados existentes, como o SSDD e o HRSID, concentram a maior parte de suas amostras em regiões costeiras e áreas de mar aberto. Com o objetivo de adaptar o treinamento às características específicas das áreas fluviais da região Norte do país e de incluir embarcações autóctones das diversas zonas do Rio Amazonas, este trabalho propõe a criação de uma nova base de dados SAR composta por imagens do satélite Sentinel-1, em banda C, obtidas ao longo do curso do Rio Amazonas.

Essa base tem como objetivo o treinamento de uma rede neural do modelo YOLOv12, visando alcançar valores altos quando avaliado por meio de métricas consagradas de detecção de objetos.

O presente estudo busca, ainda, entregar um produto vetorial confiável que indique a posição geográfica das embarcações detectadas, juntamente com atributos complementares, de modo a fornecer às autoridades competentes informações precisas e prontamente utilizáveis no contexto de monitoramento e fiscalização fluvial.

3 Sensoriamento Remoto

3.1 Sensoriamento Remoto: Conceitos Gerais

A American Society for Photogrammetry and Remote Sensing (ASPRS) define o sensoriamento remoto como “arte, ciência e tecnologia de obter informações confiáveis sobre objetos físicos e o ambiente, por meio do processo de registro, medição e interpretação de imagens e representações digitais de padrões de energia derivados de um sistema sensor sem contato” (COLWELL, 1989). Alguns anos antes, Lillesand, Kiefer e Chipman (2013) definiram o sensoriamento remoto como o processo de obtenção de informações sobre um objeto, área ou fenômeno por meio da análise de dados coletados por instrumentos que não estão em contato direto com essas entidades. Para Lorenzzetti (2022), o entendimento dessa ciência está diretamente relacionado à aplicação dos dados coletados, podendo ser voltado à identificação de feições na superfície terrestre ou à extração de informações quantitativas sobre parâmetros que influenciam sistemas físicos, biológicos, entre outros.

Observa-se, a partir dessas definições, que o sensoriamento remoto depende de instrumentos equipados com sensores capazes de captar a energia refletida pelos objetos. Gomes (2020) destaca que esses sensores podem estar embarcados em satélites (sensoriamento remoto orbital), aeronaves ou RPAs (aeronaves remotamente pilotadas), popularmente conhecidas como drones. Os sensores a bordo dessas plataformas podem ser classificados em dois tipos: passivos e ativos. Os sensores passivos dependem da radiação eletromagnética proveniente do Sol, captando essa energia quando ela é emitida ou refletida pelos alvos. Já os sensores ativos geram sua própria fonte de energia e detectam o sinal refletido a partir do retroespalhamento, permitindo a obtenção de dados independentemente da iluminação solar (LORENZZETTI, 2022).

O sensoriamento remoto depende de componentes que fazem possível a aquisição de imagens com características geométricas, espectrais e temporais. Entre esses componentes podem ser listados (ilustrados na Figura 5):

- A energia que incide sobre o alvo.
- Características do meio em que ocorre a propagação de energia.
- Propriedades do alvo a ser estudado.
- Características do sensor a ser usado.

Segundo Meneses, Almeida e Baptista (2019), a compreensão do SR fundamenta-se nos princípios físicos da interação da radiação eletromagnética com a matéria.

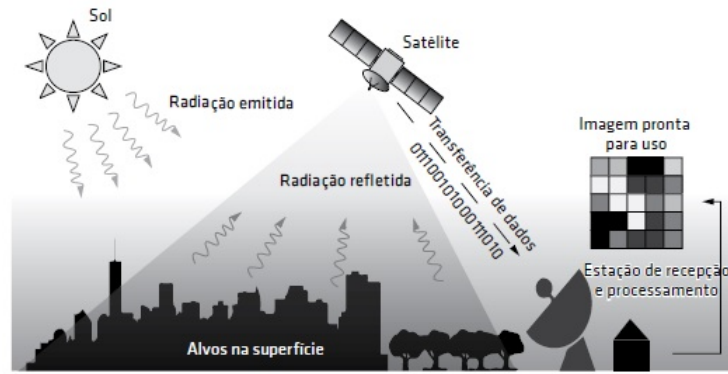


Figura 5 – Componentes do Sensoriamento Remoto. (ZANNOTA et al., 2019)

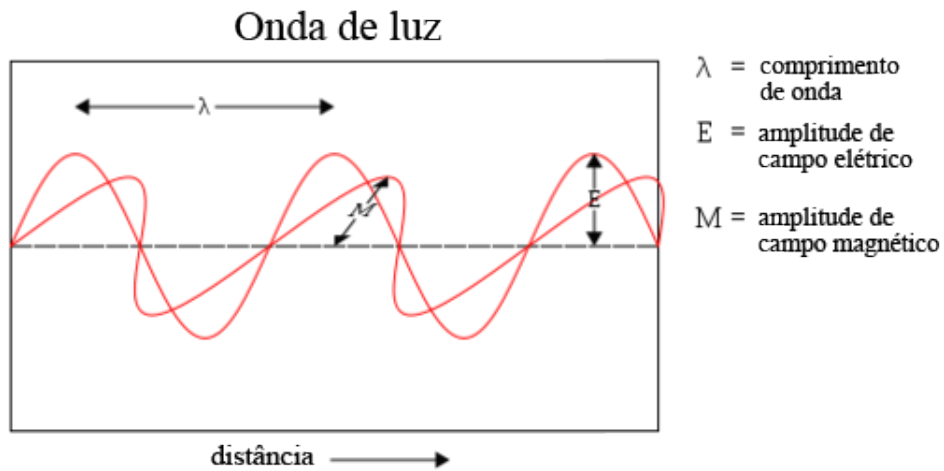


Figura 6 – Onda Eletromagnética. Campo elétrico e magnético em fase perpendiculares entre si. (ACADEMY, 2025)

3.1.0.1 Comprimento de Onda e Frequência

Quando falamos sobre sensoriamento remoto, duas características são fundamentais para entender como a radiação interage com a matéria. Por um lado, o comprimento de onda, geralmente representado por λ , é definido como a distância entre duas cristas consecutivas de um pulso de radiação (ver Figura 6). Por outro lado, a frequência f representa o número de oscilações realizadas pelo pulso de radiação em um segundo (ZANNOTA et al., 2019).

É importante notar que, à medida que o comprimento de onda aumenta, a frequência diminui. A multiplicação destas duas constantes deve resultar na constante c , a velocidade da luz. Matematicamente, o comprimento de onda e a frequência estão relacionados conforme a seguinte equação:

$$\lambda f = c \quad (3.1)$$

3.1.0.2 Espectro Eletromagnético

A luz visível é apenas uma parte do espectro eletromagnético (figura 7). As ondas de rádio, os raios ultravioleta e o infravermelho também fazem parte dele. No sensoriamento remoto, é comum agrupar a energia dentro do espectro eletromagnético de acordo com sua frequência e comprimento de onda. A unidade mais utilizada para medir o comprimento de onda é o micrômetro (μm) que equivale a 1×10^{-6} m. A porção do espectro perceptível ao olho humano é conhecida como a região do visível e se estende de $0,4 \mu m$ (azul) até $0,7 \mu m$ (vermelho) (LILLESAND; KIEFER; CHIPMAN, 2013).

Neste trabalho, será utilizada a região das micro-ondas, que compreende comprimentos de onda de 1 mm a 1 m dentro do espectro eletromagnético.

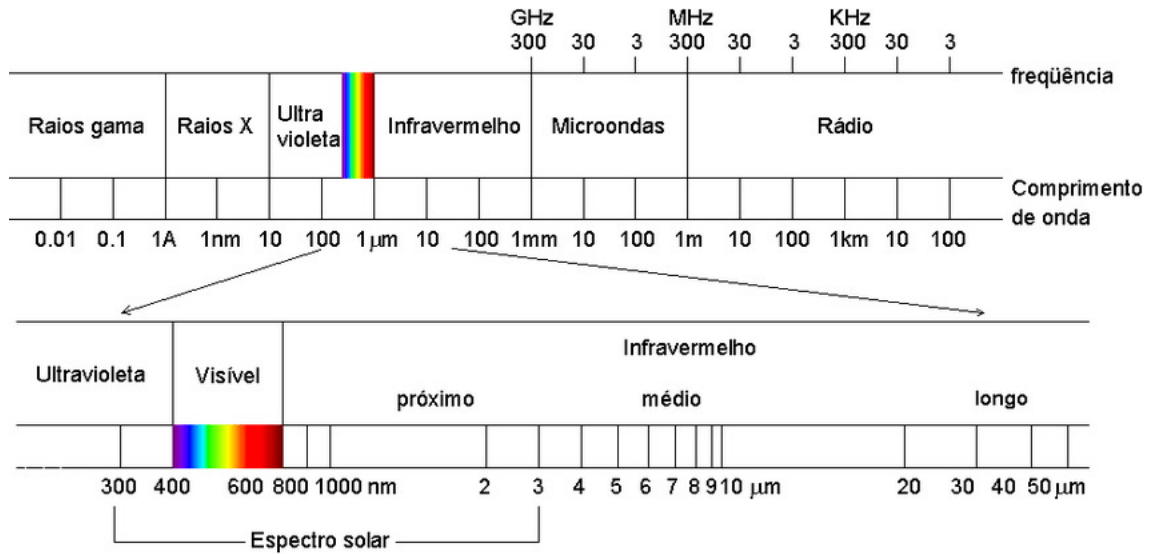


Figura 7 – Espectro eletromagnético. Em destaque a região do visível. (DORNELLES, 2008)

Para melhorar nossa compreensão da interação entre a energia e a matéria, devemos saber que a luz, além de se comportar como onda, também se comporta como partícula. A teoria do comportamento corpuscular da luz nos diz que a radiação eletromagnética é composta por *fótons* ou *quanta*. A energia de um quanta pode ser obtida por meio da seguinte relação:

$$Q = hf \quad (3.2)$$

Onde Q representa a energia de um quantum, em joules (J), h representa a constante de Planck, com valor de 6.626×10^{-34} (J x seg), e f representa a frequência.

É comprovado que, quanto maior for o comprimento de onda, menor será a energia de um quantum (LILLESAND; KIEFER; CHIPMAN, 2013). Assim, a energia de um quantum é inversamente proporcional ao comprimento de onda, ou matematicamente:

$$Q = \frac{hc}{\lambda} \quad (3.3)$$

3.2 Sensoriamento Remoto de Micro-ondas

Segundo [Lorenzzetti \(2022\)](#), o comprimento de onda que define os limites da região das micro-ondas varia de 1 centímetro a 1 metro. Nessa faixa, a transmitância atmosférica é alta; no entanto, bandas com frequências de 22,2, 60, 118,7 e 183,3 GHz são conhecidas como bandas de absorção, sendo afetadas pela presença de vapor d'água e oxigênio (ver figura 8). Os comprimentos de onda nesta faixa do espectro chegam a ser da ordem de 200.000 vezes maior que os comprimentos de onda da região do visível ([MENESES; ALMEIDA, 2012](#)).

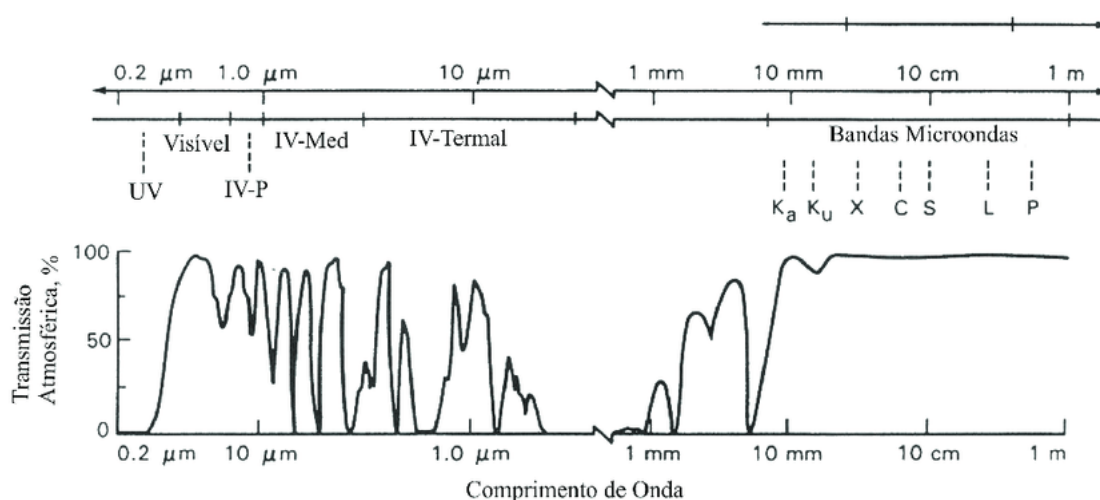


Figura 8 – Janelas atmosféricas no Espectro eletromagnético. ([GOMES, 2006](#))

Como vimos na Seção 2.1, os sensores podem ser classificados em passivos e ativos. Os sensores passivos registram a energia eletromagnética refletida e emitida pelos alvos, sendo altamente dependentes da iluminação solar. Já os sensores ativos não dependem do Sol, pois geram sua própria energia eletromagnética, que interage com o terreno e produz um sinal de retroespalhamento (backscatter), posteriormente registrado pelo sensor. Para sistemas ativos, o Retroespalhamento é equivalente à energia solar refletida e detectada por sensores passivos ([SANO; MATRICARDI; CAMARGO, 2020](#)). O instrumento ativo mais amplamente utilizado é o Radar. ([JENSEN, 2014](#)).

O RADAR (Radio Detection and Ranging), como é conhecido atualmente, foi estudado por A. H. Taylor e L. C. Young em 1922. Enviando sinais de um lado do rio Anacosta para o outro, eles observaram que, quando um navio passava pelo meio, o sinal, com comprimento de onda entre 1 e 10 metros, era interrompido. Esse foi o primeiro indício de que a radiação nesses comprimentos de onda poderia ser utilizada para medir distâncias (range) .

Rapidamente, essa nova descoberta teve aplicações militares, principalmente durante a Segunda Guerra Mundial, para navegação e localização de alvos. Radares construídos pelo Reino Unido permitiam detectar a chegada de aeronaves com um alcance máximo

de 80 quilômetros, o que foi fundamental para conter o ataque aéreo da Alemanha naquela época.

Como mencionado, o radar foi desenvolvido e aplicado em cenários de guerra. Nesse contexto, as bandas de operação utilizadas foram designadas por letras, aparentemente com o intuito de manter sigilo militar. Na Tabela 3.2, apresentam-se as principais bandas utilizadas dentro do espectro das micro-ondas. Nota-se que o parâmetro mais comum, adotado para representar essas bandas, além do comprimento de onda, é a frequência, expressa em Hertz. Em sistemas que operam na faixa das micro-ondas, essa abordagem é comum, pois, citando textualmente a [Meneses e Almeida \(2012\)](#), "quando a radiação eletromagnética atravessa materiais de diferentes densidades, a frequência não se altera, enquanto o comprimento de onda e a velocidade se alteram"([MENESES; ALMEIDA, 2012](#)).

Banda	Comprimento de onda (cm)	Frequência (GHz)
X	2,4 – 3,8	12,5 – 8
C	3,8 – 7,5	8 – 4
S	7,5 – 15	4 – 2
L	15 – 30	2 – 1
P	30 – 100	1 – 0,3

Tabela 1 – Faixas de frequência e comprimento de onda das bandas de RADAR

Apesar das grandes aplicações da tecnologia RADAR naquela época, imagens de RADAR como as conhecemos hoje ainda não estavam disponíveis.

Um marco para a geração de imagens e o mapeamento de terrenos em faixas contínuas foi o desenvolvimento do SLAR (Side-Looking Airborne Radar) na década de 1950. Esse sistema se destacou pela alta capacidade de obter imagens de grandes porções de terreno, sendo instalado no lado esquerdo ou direito de uma aeronave, permitindo a aquisição de dados laterais de longo alcance. Existem dois tipos principais de SLAR: o radar de abertura real e o radar de abertura sintética (SAR) ([JENSEN, 2014](#)), sendo este último o tipo de radar a ser utilizado para o desenvolvimento do presente trabalho.

3.3 SAR: Radar de Abertura Sintética

O Radar de Abertura Sintética (SAR) é um tipo de sensor ativo que geralmente opera na faixa de micro-ondas. Esses sensores foram desenvolvidos como uma alternativa aos sensores ópticos, pois o SAR é capaz de obter informações de forma independente das condições climáticas ou de iluminação ([SAUSEN; LACRUZ, 2015](#)). Os sistemas SAR geram imagens a partir de uma visão lateral (side-looking), representado na Figura 9, enviando pulsos de energia eletromagnética e captando o sinal refletido de volta. As pro-

priedades desse sinal (amplitude e fase) dependem dos parâmetros do sensor (ou seja, frequência) e das características físicas (como geometria e rugosidade) e dielétricas (como permissividade) do alvo. A direção de voo é conhecida como azimuth e a linha de visada é conhecida como slant range.

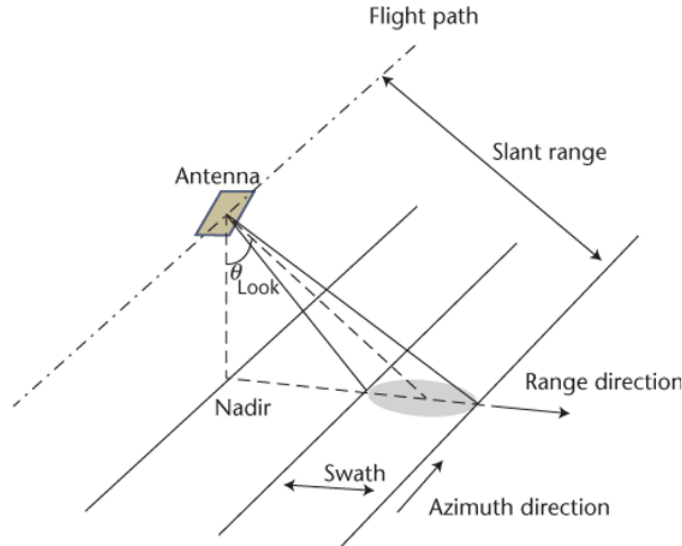


Figura 9 – Geometria de imageamento de um sistema SAR
(GOMES, 2006)

3.3.1 Princípio de Medição por Radar

A energia se propaga pelo ar aproximadamente na velocidade da luz (c), percorrendo a distância de ida e volta, a uma distância r_0 entre o satélite e o alvo. O tempo t que o sinal demorou em ir e voltar (em forma de eco, ou sinal retro-espalhado) será de:

$$t = \frac{2r_0}{c} \quad (3.4)$$

Com $c = 3 \times 10^8 m/seg$, temos, como exemplo, que um tempo de atraso de $1 \mu s$ representa uma distância de 150 metros em slant range. Este princípio de medição de distância é fundamental para o funcionamento e entendimento dos sistemas de radar (LILLESAND; KIEFER; CHIPMAN, 2013).

3.3.2 Resolução Espacial

Os radares mais antigos, que não utilizavam o princípio de Abertura Sintética, apresentavam uma resolução moderada em azimuth. A resolução em azimuth refere-se à menor separação entre dois objetos que o radar é capaz de distinguir em azimuth, equação mostrada a seguir.

$$\delta_a = \frac{\lambda}{d_a} \cdot r_0 \quad (3.5)$$

Onde λ representa o comprimento de onda, d_a representa o diâmetro da antena e r_0 representa a distância em range entre a antena e o alvo. Podemos observar claramente na equação (2.12) que a resolução em azimuth diminui linearmente com o aumento da distância entre o satélite e o solo. Como exemplo, para uma antena funcionando na banda X (0,03 m), com uma antena de diâmetro 3 m e uma distância em range de 5 km, temos que $\delta_a = 50m$, uma resolução relativamente baixa, sendo impossível usar este tipo de sistemas para aplicações orbitais. Contudo, o sistema SAR resolve isso, incrementando a resolução espacial em até 30 vezes, sendo a resolução a metade do comprimento da antena e sendo também independente da distância em Range (MOREIRA et al., 2013).

3.3.2.1 Resolução em Slant Range

Para que dois objetos sejam considerados separáveis, é necessário que o sinal retroespalhado por cada um deles seja recebido pela antena em instantes de tempo distintos. A distinção entre os objetos torna-se problemática quando a distância Δr_0 entre os dois objetos é muito pequeno, ou seja, menor que o comprimento do pulso transmitido (que é determinado pela duração do pulso τ). Esse cenário pode ser melhor visualizado na figura 12.

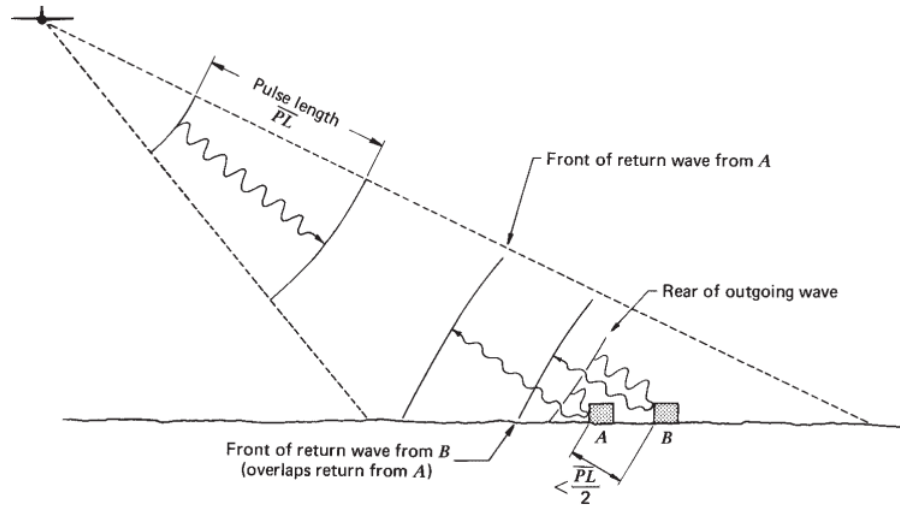


Figura 10 – A resolução em range depende exclusivamente da duração do pulso τ , ou do seu derivado comprimento de pulso $\overline{PL}$. Se a distância Δr_0 entre A e B for menor a $\overline{PL}/2$, os sinais não serão separáveis.

(LILLESAND; KIEFER; CHIPMAN, 2013)

O problema de Δr_0 ser muito pequeno reside no fato de que, consequentemente, a diferença de tempo Δt entre a chegada dos sinais retroespalhados também será reduzida, conforme a relação abaixo:

$$\Delta t = \frac{2\Delta r_0}{c} \quad (3.6)$$

Definimos a resolução em *slant range* δ_r como a distância mínima entre dois objetos que ainda podem ser distinguidos um do outro, ou seja, a distância que corresponde a uma diferença de tempo Δt igual ou maior à duração do pulso τ (German Aerospace Center (DLR), Microwaves & Radar Institute (HR), Pol-InSAR Research Group, 2024). Assim, representamos essa resolução como:

$$\delta_r = \frac{c\tau}{2} \quad (3.7)$$

Observa-se, pela equação acima, que a resolução em *range* não depende da distância entre o sensor e o alvo; depende exclusivamente da duração do pulso τ , que é inversamente proporcional à largura de banda do pulso (pulse bandwidth) W :

$$\delta_r = \frac{c\tau}{2} = \frac{c}{2W} \quad (3.8)$$

Como exemplo, temos que uma largura de banda de pulso de $W = 100$ MHz leva a uma resolução em alcance de $\delta_r = 1,5$ m (correspondente a uma duração de pulso de $\tau = 10$ ns).

Até aqui tudo certo, porém existe um grande dilema. A energia do pulso transmitido é o que define a distância máxima na qual um objeto pode ser detectado. Quando um pulso de radar é enviado por um sensor, ele percorre uma certa distância até atingir um objeto e retornar ao sensor. A distância entre o sensor e o objeto é r_0 .

A energia do sinal recebido, E_{Rx} , de um objeto localizado a uma distância r_0 , depende da energia do pulso transmitido, E_{Tx} , e decai de acordo com a quarta potência inversa da distância:

$$E_{\text{Rx}} \sim \frac{E_{\text{Tx}}}{r_0^4} \quad (3.9)$$

A energia do pulso transmitido é determinada pelo produto entre a potência de pico instantânea P e a duração do pulso τ . No entanto, a potência P é limitada pelas capacidades do hardware do sensor, principalmente em sensores embarcados no espaço. Para aumentar a energia do pulso, seria necessário aumentar sua duração.

Contudo, existe uma limitação: para obter alta resolução espacial, são necessários pulsos curtos, o que exige o uso de larguras de banda maiores. Isso cria um conflito entre o alcance e a resolução.

A solução para esse dilema é o uso de pulsos modulados em frequência, como o chamado *FM Chirp*, em conjunto com técnicas de filtragem casada (*matching filtering*), que permitem manter alta resolução mesmo com pulsos mais longos.

3.3.2.2 Resolução em Azimuth

A resolução em Azimuth, representada por δ_a , é dada pela abertura sintética L_{SA} . A largura do feixe da antena (Beamwidth) é a amplitude angular do lóbulo principal do padrão de radiação da antena, que é a região do padrão onde a maior parte da potência é irradiada. O Beamwidth é dado por:

$$\theta_a = \frac{\lambda}{d_a} \quad (3.10)$$

Onde d_a representa o comprimento da antena. Usando a equação 2.34, temos que:

$$L_{SA} = r_0 \cdot \theta_a = \frac{\lambda r_0}{d_a} \quad (3.11)$$

Podemos ver melhor r_0 e θ_a na figura 11. Em comparação com o Beamwidth de uma antena real, que depende totalmente do comprimento d_a , no radar de abertura sintética temos o comprimento L_{SA} , que atua substituindo ao comprimento físico da antena (MOREIRA et al., 2013). Assim, a relação que representa o beamwidth em um sistema SAR é:

$$\theta_{sa} = \frac{\lambda}{2L_{sa}} \quad (3.12)$$

É importante notar que na equação acima, o '2' representa o caminho de ida e volta da onda eletromagnética. Assim, nossa resolução em azimuth fica:

$$\delta_a = r_0 \theta_{sa} = r_0 \frac{\lambda}{2L_{sa}} = \frac{da}{2} \quad (3.13)$$

Ou seja, que não dependemos da distância em range e que antenas menores podem gerar uma resolução em azimuth consideravelmente boa (MOREIRA et al., 2013).

Em síntese, O radar pode medir a amplitude (a força do eco refletido) e a fase (a posição de um ponto no tempo em um ciclo de forma de onda). O radar só pode medir a parte do eco refletido de volta para a antena (retroespalhamento). Os pulsos de radar viajam à velocidade da luz. A intensidade do eco refletido é o coeficiente de retroespalhamento σ^0 e é expressa em decibéis (dB).

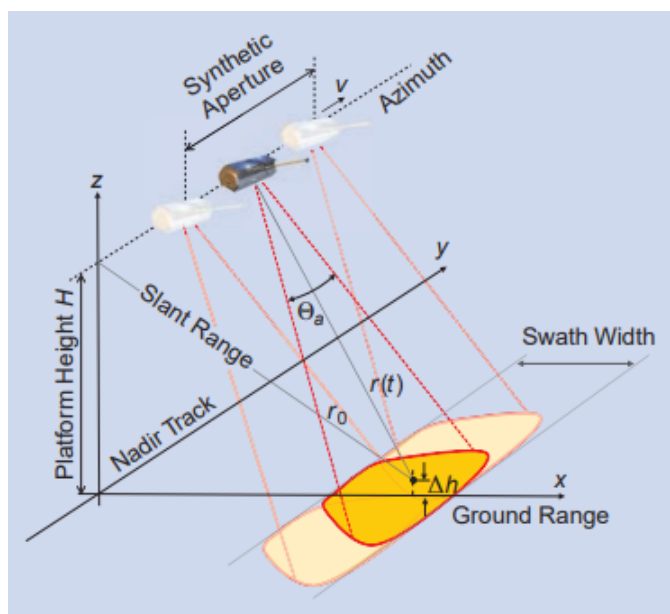


Figura 11 – Visualização clara dos parâmetros usados para os calculos das resoluções.

(MOREIRA et al., 2013)

4 Inteligência Artificial

A Inteligência Artificial (IA) foi introduzida pela primeira vez no Colégio Dartmouth, nos Estados Unidos, no ano de 1956 (DEMIR; SARI, 2020). Segundo RUSSELL e NORVIG (2021), a inteligência artificial pode ser compreendida pelos processos de pensamento e comportamento, sendo seu sucesso avaliado pela aproximação ao desempenho humano ou pelo ideal de racionalidade. Neste sentido, SALEHI e BURGUEÑO (2018) afirma que a inteligência artificial é um método de cálculo que busca simular a cognição humana. Complementando, HAENLEIN e KAPLAN (2019) definem ela como “a capacidade de um sistema de interpretar corretamente dados externos, aprender a partir desses dados e utilizar esses aprendizados para alcançar objetivos e tarefas específicos por meio de adaptação flexível”.

CHENG e BRADLEY (2023) argumentam que a inteligência artificial está se consolidando como uma das tecnologias mais relevantes da atualidade, uma vez que o aprendizado de máquina (em especial o aprendizado profundo) tem permitido que os computadores adquiram conhecimento a partir do treinamento com grandes bases de dados e desenvolvam a capacidade de aprender por meio da análise dessas informações, em vez de dependerem exclusivamente de algoritmos determinísticos previamente programados.

A inteligência artificial (IA) é tradicionalmente subdividida em áreas como o aprendizado de máquina (ML) e o aprendizado profundo (DL) (DEMIR; SARI, 2020). Um esquema para visualizar esta divisão pode ser vista na figura 12. Esses três termos são amplamente utilizados, porém, muitas vezes de forma inadequada. Embora estejam relacionados, não se referem aos mesmos conceitos e, portanto, não devem ser empregados de maneira intercambiável (PUCHI-CABRERA et al., 2023).

4.1 Aprendizado de Máquina

O aprendizado de máquina é uma subárea da inteligência artificial dedicada ao desenvolvimento e estudo de modelos e métodos de IA (CHENG; BRADLEY, 2023). O objetivo deste ramo da IA é a construção de programas que melhorem seu desempenho por meio de exemplos. Para que isso seja possível, é necessária uma grande quantidade de exemplos para gerar o conhecimento do computador, que são hipóteses geradas a partir dos dados. Quanto mais precisos forem esses dados, teremos generalizações mais precisas (LUDERMIR, 2021).

Os sistemas de aprendizado de máquina podem ser classificados de acordo com a quantidade e o tipo de supervisão recebida durante o treinamento. Alguns dos tipos

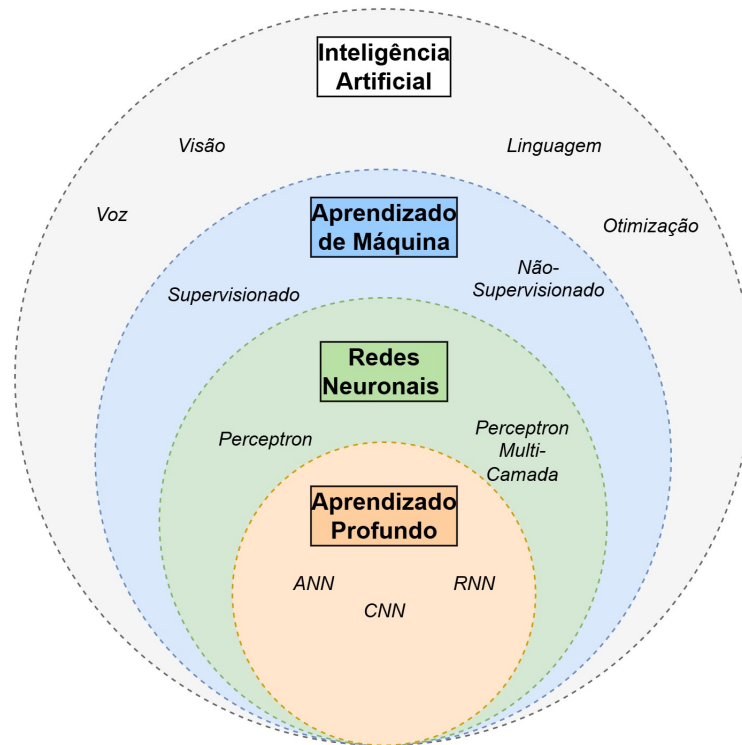


Figura 12 – Esquema que representa a organização da inteligência artificial e exemplos de produtos. Nessa representação, ANN, CNN, RNN correspondem, respectivamente, a redes neurais artificiais (*artificial neural networks*), redes neurais convolucionais (*convolutional neural networks*) e redes neurais recorrentes (*recurrent neural networks*). Adaptado de (CHENG; BRADLEY, 2023)

principais são:

- Aprendizado supervisionado, onde para cada exemplo apresentado ao algoritmo de aprendizado é necessário apresentar um rótulo informando a que classe o exemplo pertence (LUDERMIR, 2021). Uma tarefa típica de aprendizado supervisionado é a classificação (GÉRON, 2019).
- Aprendizado não supervisionado: os dados de treinamento não são rotulados. O sistema tenta aprender sem a presença de um “professor”, que no caso dos algoritmos supervisionados, são os rótulos (GÉRON, 2019). O algoritmo agrupa os exemplos pelas similaridades dos seus atributos. Estas agrupações são chamadas de *clusters* (LUDERMIR, 2021).
- Aprendizado semi-supervisionado: Dependendo da quantidade de dados, o processo de rotulação pode ser demorado e custoso. Nesse sentido, é comum haver uma grande quantidade de instâncias não rotuladas e apenas algumas instâncias rotuladas em algumas bases de dados. Alguns algoritmos são capazes de lidar com este tipo de dados, ou seja, dados parcialmente rotulados. Este tipo de algoritmos são conhecidos como algoritmos de aprendizado semi-supervisionado (GÉRON, 2019).

4.2 Redes Neurais Artificiais

As redes neuronais artificiais podem ser representadas como redes neurais *feed-forward*, compostas por diferentes camadas de nós, que são chamados também de neurônios, que produzem uma saída a partir da entrada recebida das camadas anteriores (PUCHI-CABRERA et al., 2023).

Estas redes foram inspiradas no funcionamento dos neurônios humanos. Um neurônio artificial pode ser considerado um modelo matemático influenciado por um neurônio biológico. Um neurônio é composto por um corpo celular, também conhecido como 'soma' ou 'núcleo', um ou mais dendritos, que são os responsáveis por receber sinais de outros neurônios e um axônio responsável por transmitir os sinais gerados pelo neurônio (estrutura representada na figura 13). Dizemos que um neurônio está em estado ativo quando envia sinais a outros neurônios; entretanto, encontra-se em estado inativo quando apenas está recebendo sinais. Para que o neurônio possa enviar sinais para outros neurônios, isto é, atingir o estado de ativação, é necessário alcançar o limiar de ativação, que ocorre pela acumulação de sinais no estado inativo. Todo esse processo motivou os pesquisadores a desenvolver e testar as redes neurais artificiais (ANNs) (RAZZAQUE, 2019).

As redes neurais artificiais (ANNs) permanecem como os algoritmos de aprendizado de máquina mais difundidos. A primeira rede neural artificial foi o *perceptron* que é uma rede neural de camada única e funciona como classificador binário (figura 14). Quando temos várias camadas de neurônios conectados dentro da rede, isto é, quando a profundidade da nossa rede aumenta, dizemos que estamos trabalhando com redes de aprendizado profundo (PUCHI-CABRERA et al., 2023).

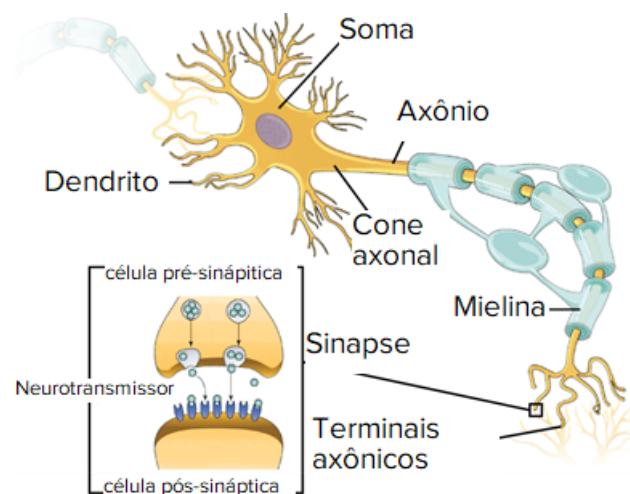


Figura 13 – Estrutura de um neurônio humano (Khan Academy, 2025).

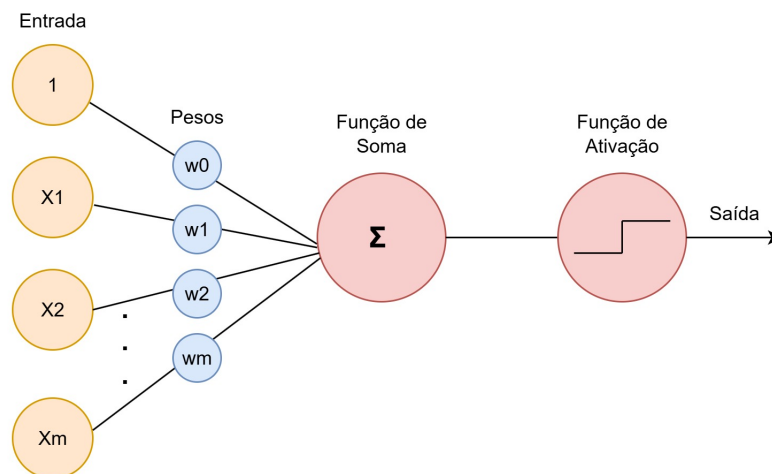


Figura 14 – Arquitetura da rede Perceptron

4.3 Aprendizado Profundo

Enquanto modelos tradicionais de aprendizado de máquina deixam de melhorar após determinado ponto, modelos de aprendizado profundo tendem a apresentar bom desempenho quando treinados com grandes volumes de dados (RAZZAQUE, 2019). De acordo com GOODFELLOW, BENGIO e COURVILLE (2016):

"o aprendizado profundo é um tipo específico de aprendizado de máquina que alcança grande poder e flexibilidade ao aprender a representar o mundo como uma hierarquia aninhada de conceitos, em que cada conceito é definido em relação a conceitos mais simples, e representações mais abstratas são computadas a partir de representações menos abstratas."

Ou seja, Ele permite que o computador construa conceitos complexos a partir de conceitos mais simples. Isso é feito por meio da utilização de redes neurais multicamadas, chamadas redes neurais profundas, para simular o poder de tomada de decisão do cérebro humano (IBM, 2025). Esta ideia está representada na figura 15 na qual é ilustrado como um sistema de aprendizado profundo, combinando conceitos simples, pode representar o conceito de uma imagem de uma pessoa. Segundo (GOODFELLOW; BENGIO; COURVILLE, 2016), no sistema representado na figura 15, temos uma entrada, chamada 'camada visível'. São os dados que passamos para o nosso sistema. No nosso caso, uma imagem, que não é nada mais que uma matriz conformada por pixels. Em seguida, temos uma série de 'camadas ocultas'. A primeira camada oculta identifica arestas, comparando o brilho de pixels vizinhos. Com base nessas informações, a seguinte camada oculta pode juntar essas arestas e identificar cantos e contornos. A partir das informações da segunda camada oculta, a terceira camada pode reconhecer partes inteiras de objetos presentes na imagem de entrada. Por fim, na camada de saída são representados os objetos detectados a partir das características extraídas de cada camada oculta.

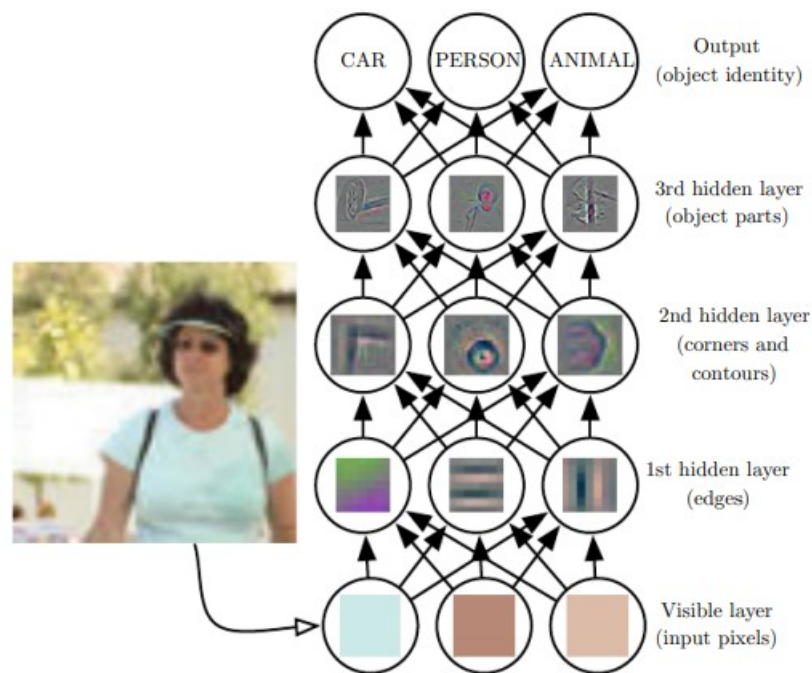


Figura 15 – Representação de uma imagem de uma pessoa por meio de um sistema de aprendizado profundo (GOODFELLOW; BENGIO; COURVILLE, 2016).

Existem diversos tipos de redes neurais artificiais, porém dois se consolidaram mais do que os demais: as redes neurais recorrentes (RNNs) e as redes neurais convolucionais (CNNs). Esta última constitui parte do objeto de estudo do presente trabalho e, por esse motivo, será aprofundada nas próximas seções. Além disso, uma breve introdução sobre *transformers* será dada.

4.3.1 Transformadores

O "Transformador" é um componente de rede neural que pode ser utilizado para aprender representações úteis de sequências ou conjuntos de pontos de dados (TURNER, 2023), ou seja, este aprende o contexto de dados sequenciais e gera novos dados a partir deles. As entradas, também conhecidas como *tokens*, passam por uma sequência de codificadores e decodificadores que trabalham paralelamente para gerar uma saída. Na parte inicial de um codificador ocorre a transformação das entradas em vetores; isso é feito pelas chamadas "incorporações de entrada" (*input embeddings*). A ordem ou sequência destas entradas é regida pela "codificação posicional" (*positional encoders*). A camada de codificação serve para transformar todas as sequências de entrada em uma representação contínua e abstrata que encapsula as informações aprendidas de toda a sequência; este bloco pode ser visto na figura 16a.

Dentro dele, como visto na figura 16b, temos a "atenção multi-cabeça" e uma rede *feed forward* para refinamento adicional, além das operações de normalização e das co-

nexões residuais. Os transformadores são projetados para compreender o contexto analisando a relação entre diferentes elementos, e dependem quase inteiramente de uma técnica matemática chamada de "atenção" para fazer isso (DataCamp, 2025).

Segundo Vaswani et al. (2017), uma função de atenção pode ser descrita como um mapeamento entre uma *query* (consulta, Q) e um conjunto de pares chave-valor (*key-value*, K , V) para uma saída, onde a *query*, as *keys*, os *values* e a saída são todos vetores. A saída é computada como uma soma ponderada dos valores, na qual o peso atribuído a cada valor é calculado por meio de uma função de compatibilidade entre a *query* e a *key* correspondente. Esta função pode ser vista na equação 4.1.

$$\text{Atenção}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) \cdot V \quad (4.1)$$

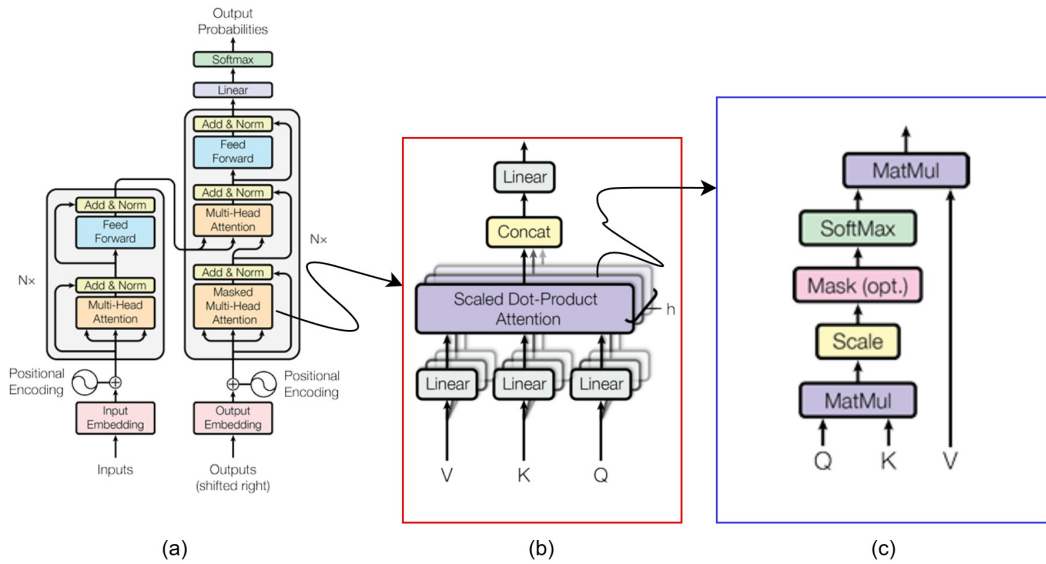


Figura 16 – Arquitetura do modelo transformador original. Adaptado de (VASWANI et al., 2017)

4.3.2 Redes Neurais Convolucionais

As redes neurais convolucionais têm sido amplamente adotadas na visão computacional, especialmente em tarefas como o reconhecimento de imagens (RAZZAQUE, 2019). O nome “rede neural convolucional” indica que a rede emprega uma operação matemática denominada convolução. As camadas de convolução geralmente vem acompanhadas de camadas de *pooling* que permite a simplificação gradual dos dados, e uma camada de normalização, que possibilita lidar com a variabilidade nas dimensões das características. A convolução é um tipo especializado de operação linear. As redes convolucionais são, portanto, redes neurais que utilizam a convolução em substituição à multiplicação matricial

geral em pelo menos uma de suas camadas (GOODFELLOW; BENGIO; COURVILLE, 2016).

Em termos gerais, a operação de convolução diz sobre como duas funções produzem uma terceira que descreve como a forma de uma é modificada pela outra (PODAREANU et al., 2019), e é representado pelo asterisco. A sua forma regular pode ser visto na equação 4.2 A operação de convolução utilizada em uma rede neural convolucional não necessariamente corresponde à definição de convolução empregada em outras áreas.

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau) g(t - \tau) d\tau \quad (4.2)$$

Em aplicações de aprendizado de máquina, a entrada normalmente é um tensor (arranjo multidimensional de dados), e o *kernel* também é um tensor de parâmetros ajustados pelo algoritmo. Como só armazenamos valores em um conjunto finito de pontos, a convolução prática é implementada como uma soma finita. Muitas vezes, essa operação ocorre em mais de um eixo; por exemplo, ao processar uma imagem bidimensional I , utiliza-se um *kernel* bidimensional K . A operação de convolução deve ser discreta, ou seja, operamos sobre valores inteiros, e comutativa, ou seja, tanto faz aplicar a imagem sobre o *kernel* ou o *kernel* sobre a imagem, o resultado será o mesmo. Por fim, temos a equação resultante destas propriedades na equação 4.3.

$$(I * K)(i, j) = \sum_{m=-M}^M \sum_{n=-N}^N I(i - m, j - n) K(m, n) \quad (4.3)$$

Um exemplo da operação de convolução é apresentado na Figura 17. Nesse caso, temos a imagem de entrada (*Source layer*) sobre a qual é aplicado um filtro (*Convolutional kernel*) de dimensão 3×3 . Em cada posição, realiza-se a operação de produto escalar entre os valores dos pixels da vizinhança da imagem de entrada e os coeficientes do filtro convolucional. O resultado dessa soma ponderada define o novo valor do pixel central, formando assim a matriz de saída (*Destination layer*).

Alguns resultados da aplicação de diferentes filtros de convolução podem ser observados na Figura 17. O objetivo desta figura é ilustrar o processo de extração de características por meio de filtros de convolução. Embora esses filtros não sejam aplicados diretamente na rede neural, eles servem como exemplo inicial do que ocorre nas primeiras camadas convolucionais dos modelos de Redes Neurais Convolucionais (CNNs). Na figura, são apresentados os filtros Sobel inferior (superior esquerdo), Sobel superior (superior central), Sobel esquerdo (superior direito), Sobel direito (inferior esquerdo), desfoque (inferior central) e contorno (inferior direito), este último apresenta o filtro de convolução com os parâmetros para gerar a imagem de saída.

Além da operação de convolução, é comum utilizar uma função de *pooling* para

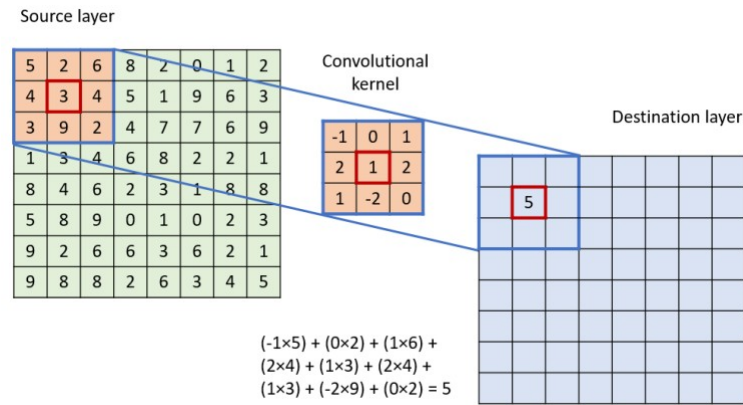


Figura 17 – Operação de convolução (PODAREANU et al., 2019)

definir uma hierarquia de informações, além disso, o *pooling* ajuda a tornar a representação aproximadamente invariante a pequenas translações da entrada. Isto é feito por meio da substituição da saída da rede em uma determinada localização por uma estatística resumida das saídas vizinhas. Um dos tipos mais utilizados é o *max pooling*, o qual retorna o valor máximo dentro de uma janela retangular (figura 19).

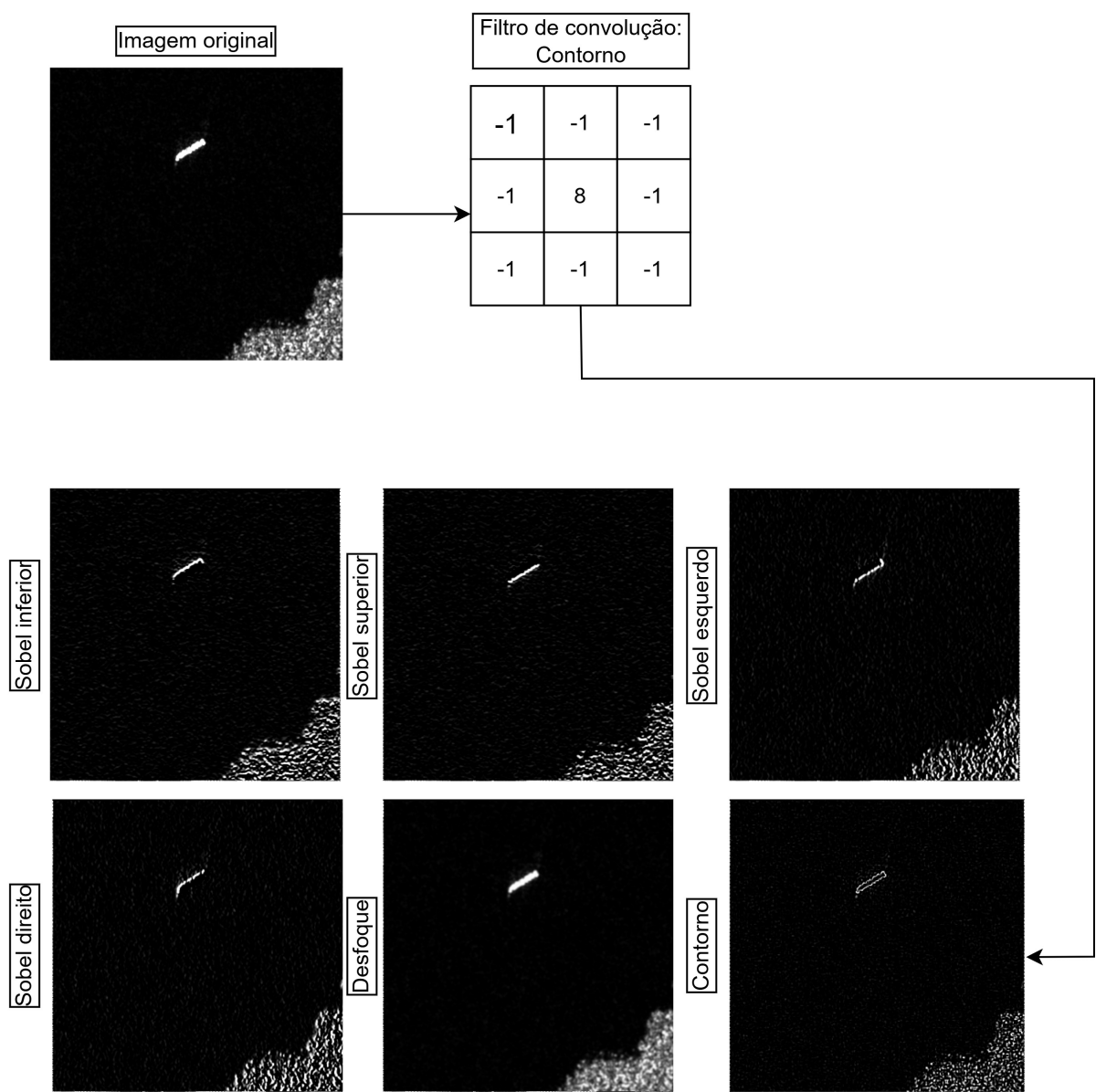


Figura 18 – Filtros por meio de convolução
(IMAGE..., 2025)

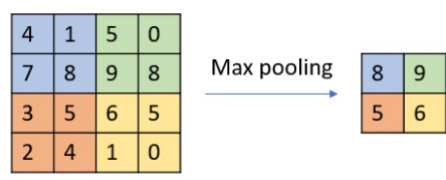


Figura 19 – Operação *max pooling* por meio de uma janela 2x2
(PODAREANU et al., 2019)

As redes neurais convolucionais são tão eficazes porque conseguem aprender de forma eficiente padrões em imagens. Isso acontece graças a três ideias principais: cada

neurônio olha apenas para uma pequena região da entrada (interações esparsas), o mesmo conjunto de filtros é usado em toda a imagem (compartilhamento de parâmetros) e a rede responde de forma consistente quando um objeto se desloca de posição (equivariância à translação). Essas características reduzem drasticamente o número de parâmetros, tornam o treinamento mais rápido e permitem reconhecer padrões recorrentes em diferentes partes da entrada.

4.3.3 YOLO

A detecção de objetos é uma técnica utilizada para identificar e localizar objetos dentro de uma imagem ou de um vídeo. A detecção de objetos desempenha um papel fundamental em sistemas de visão computacional, podendo ser aplicada em diversas áreas. Por exemplo, na segurança, é utilizada em tarefas de vigilância por vídeo, enquanto na saúde contribui para a análise de imagens médicas (GALVEZ et al., 2018). Neste sentido, o algoritmo YOLO (*You Only Look Once*) é considerado um dos mais proeminentes algoritmos nesta área (JEGHAM et al., 2024). O YOLO foi introduzido em 2015 por Joseph Redmon, Santosh Divvala, Ross Girshick e Ali Farhadi. Em resumo, YOLO funciona a partir das seguintes etapas (baseado em (REDMON et al., 2016)):

- **Divisão da imagem em células da grade:** O sistema divide a imagem de entrada em uma grade de $S \times S$.
- **Caixas delimitadoras e confiança:** Cada célula da grade prevê B caixas delimitadoras e as respectivas pontuações de confiança. Essas pontuações refletem quão confiante o modelo está de que a caixa contém um objeto e também quão precisa acredita ser a predição dessa caixa. Formalmente, define-se:

$$\text{Confiança} = \Pr(\text{objeto}) \times \text{IOU}_{\text{pred, truth}}.$$

$$\Pr(\text{objeto}) = \begin{cases} 1, & \text{se o centro do objeto estiver dentro da caixa,} \\ 0, & \text{caso contrário.} \end{cases}$$

- **Probabilidade de classe:** Cada célula da grade também prevê C probabilidades condicionais de classe:

$$\Pr(\text{Classe}_i | \text{Objeto}).$$

Essas probabilidades são condicionadas ao fato de a célula conter um objeto.

- **Probabilidade final:** Na inferência, multiplicam-se as probabilidades condicionais de classe pelas predições individuais de confiança das caixas:

$$\Pr(\text{Classe}_i | \text{Objeto}) \cdot \Pr(\text{Objeto}) \cdot \text{IOU}_{\text{pred}}^{\text{truth}} = \Pr(\text{Classe}_i) \cdot \text{IOU}_{\text{pred}}^{\text{truth}}. \quad (4.4)$$

Essa operação fornece pontuações de confiança específicas por classe para cada caixa delimitadora. Essas pontuações codificam tanto a probabilidade de aquela classe aparecer na caixa quanto o quão bem a caixa predita se ajusta ao objeto.

Cada uma destas etapas está ilustrada na figura 20:

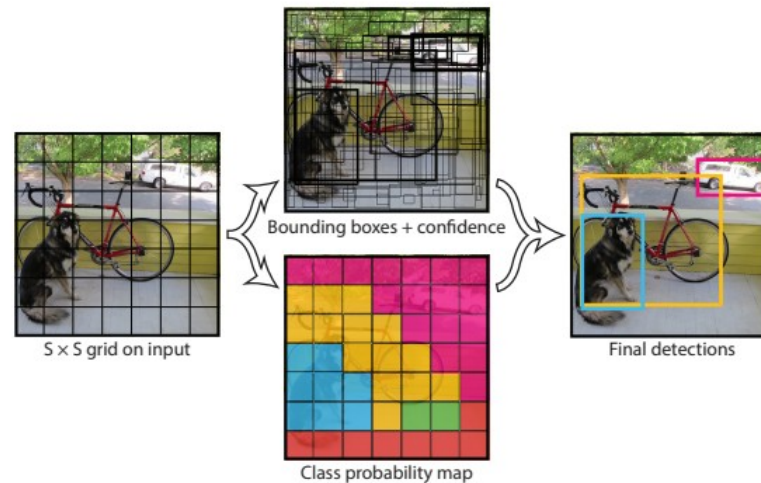


Figura 20 – Etapas para detectar objetos no modelo YOLO original (REDMON et al., 2016)

Para compreender o modelo que será utilizado no desenvolvimento deste trabalho, faz-se necessário apresentar brevemente os principais modelos antecessores. No YOLOv8 surgiu o módulo C2f (*Convolutional to Fused*). Ele funciona assim: divide os mapas de características, processa com camadas "pescoço de garrafa" e depois junta tudo de novo. Isso melhora a aprendizagem em múltiplas escalas e amplia o campo de visão da rede, sem aumentar o custo de processamento. Além disso, este modelo introduz o suporte a caixas delimitadoras orientadas (OBB). O YOLOv9 evoluiu o YOLOv8 ao introduzir o PGI, que garante melhor fluxo de gradientes e treinamento mais eficiente, e o GELAN, que melhora o aproveitamento dos parâmetros da rede, além de ampliar o suporte a OBB na variante YOLOv9u. O YOLOv10 inovou ao remover a dependência da Supressão Não Máxima (NMS) para reduzir latência e acelerar a inferência, além de otimizar a arquitetura para equilibrar eficiência e precisão; em algumas variantes, também recebeu suporte para OBB. O YOLOv11 foi o último YOLO puramente baseado em CNN, focado em velocidade e eficiência com C3k2 e convoluções separáveis em profundidade, além de oficializar o OBB, preparando o terreno para o YOLOv12, que é uma rede neural profunda baseada majoritariamente em convoluções, com módulos de atenção integrados (SINGH, 2025).

4.3.4 YOLOV12

O YOLOV12 foi lançado o 18 de fevereiro de 2025, sendo um dos modelos mais recentes na área de detecção de objetos. Este modelo foi desenvolvido por Yunjie Tian (Universidade de Buffalo, SUNY), Qixiang Ye (Universidade da Academia Chinesa de Ciências) e David Doermann (Universidade de Buffalo, SUNY) ([Ultralytics, 2025](#)).

O YOLOV12 é a primeira versão a integrar mecanismos de atenção como um elemento arquitetônico central, com o objetivo de aprimorar a detecção de objetos pequenos em ambientes complexos, mantendo ao mesmo tempo a velocidade de inferência em tempo real ([CUI; LIU; YANG, 2025a](#)).

Modelos YOLO anteriores são baseados puramente em CNN. Um desafio fundamental em detectores baseados em CNN é sua capacidade limitada de capturar dependências de longo alcance, que são cruciais para compreender relações espaciais em cenas complexas. Os Transformadores de visão ou *Vision transformers* se destacam na modelagem de características globais, porém aumentam a complexidade computacional ([KHANAM; HUSSAIN, 2025](#)).

YOLOV12 faz uma integração entre redes neurais convolucionais e Transformadores, incluindo na arquitetura YOLO mecanismos de atenção A2, dentro do bloco A2C2F, que faz agregação de características multi-escala, facilitando a fusão eficaz de características semânticas em diferentes escalas, melhorando assim o desempenho da detecção para alvos de vários tamanhos ([CUI; LIU; YANG, 2025b](#)).

A arquitetura geral do YOLOV12 está formada por três componentes principais: Espinha dorsal (para extração de características), pescoço (para fusão de características) e cabeça (para predição) conforme a figura 21.

Na espinha dorsal é onde tudo começa. Este componente é responsável por analisar a imagem de entrada e extrair padrões significativos. as duas primeiras fases correspondem a camadas convolucionais convencionais (**Conv**), que são compostas por Conv2d, normalização em lote (*batch normalization*) e funções de ativação SiLU. Estes blocos são responsáveis por reduzir as dimensões espaciais da imagem ao mesmo tempo em que capturam padrões básicos, como contornos e cantos. Em seguida, temos os blocos **C3k2**,

Os blocos **C3k2** que refinam a extração de características ao mesmo tempo em que mantêm a eficiência.

Esses módulos enfatizam tanto os padrões locais quanto os globais nas imagens, criando mapas de características hierárquicos. As camadas rasas capturam detalhes mais finos, enquanto as camadas profundas se concentram em características de alto nível.

O bloco **A2C2f** é o mecanismo de atenção de área desenhado para aprimorar a habilidade do modelo em focar nas regiões mais relevantes da imagem, partindo a ima-

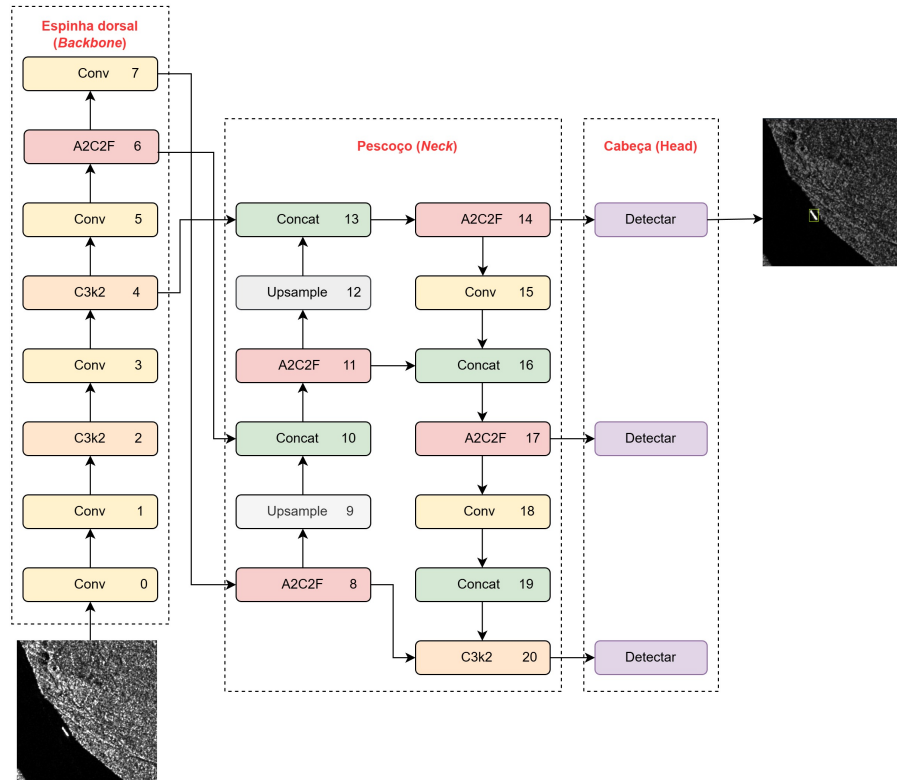


Figura 21 – Arquitetura do modelo YOLOv12.

gem em sub imagens, alcançando maior velocidade enquanto mantém o desempenho de detecção alta. Dentro deste bloco, temos a arquitetura mostrada na figura 22. Após o processamento, as saídas de todos os fluxos paralelos A2 são concatenadas ao longo da dimensão de canais formando um novo mapa de características rico em detalhes (NGUYEN, 2025). Este bloco é poderoso pois usa convoluções para transformar e misturar informações de canais, permitindo à rede combinar características ao longo dos canais sem afetar resolução espacial. Além disso, dentro dos blocos A2, aplicamos atenção de área para focar nas regiões mais relevantes da imagem (fronteiras, zonas ricas em texturas). Este bloco enlaça múltiplas transformações permitindo que cada estágio seja construído sobre o anterior, refinando a saída ainda mais para frente (TORNERO, 2025).

Os blocos A2 combinam mecanismos de atenção de área com redes *feedforward* para processar os mapas de características de entrada, mantendo um campo receptivo amplo enquanto reduzem a complexidade computacional.

O módulo de atenção de área (AAttn) processa a entrada e extrai relações espaciais. Os resultados são adicionados de volta à entrada original, formando uma conexão residual. Isso ajuda a melhorar o fluxo do gradiente e estabiliza o treinamento.

O bloco MLP (duas convoluções sequenciais) é responsável por expandir e reduzir o número de canais, mantendo sempre conexões residuais.

Indo mais profundo no módulo de atenção de área AAttn, observamos uma ar-

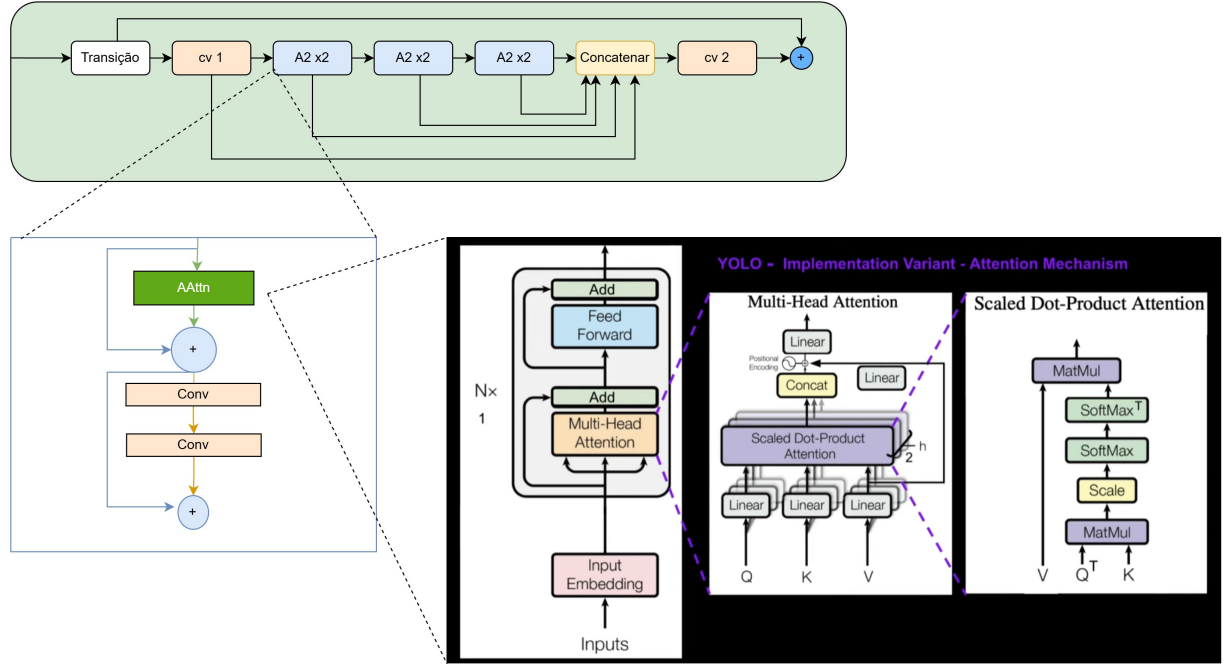


Figura 22 – Diagrama da arquitetura do bloco **A2C2F**, implementação do YoloV12 que aumenta a robustez na extração de características.

quitetura *transformer* bastante parecida à arquitetura original, mostrada na figura 16. O YOLOv12 implementa um mecanismo de atenção baseado em áreas, que processa as características de entrada de forma sensível ao espaço, tornando-o particularmente eficaz para tarefas de detecção de objetos.

Como exemplo do processo interno, tomemos uma entrada com tamanho de lote 1, 64 canais, 30 de altura e 40 de largura [1, 64, 30, 40]. Esse tensor, ao passar pelo módulo AAttn, é transformado em um espaço de 192 dimensões, produzindo um tensor denominado qkv . Em seguida, as colunas e linhas são aplanadas, resultando em uma dimensão de 1200 posições espaciais.

O próximo passo é dividir o tensor em quatro áreas, obtendo, portanto, quatro segmentos com 300 posições cada. Depois disso, as 192 dimensões são separadas em duas cabeças de atenção (padrão do YOLOv12), resultando em 96 dimensões para cada cabeça. Por fim, o tensor é dividido em seus componentes q , k e v , que compartilham as mesmas dimensões.

A visualização espacial dessas operações pode ser observada na Figura 23. A equação que unifica esses tensores é apresentada na Equação 4.5. Apesar da semelhança matemática com a Equação 4.1, as variações implementadas otimizam o desempenho em tarefas de visão computacional (TORNERO, 2025).

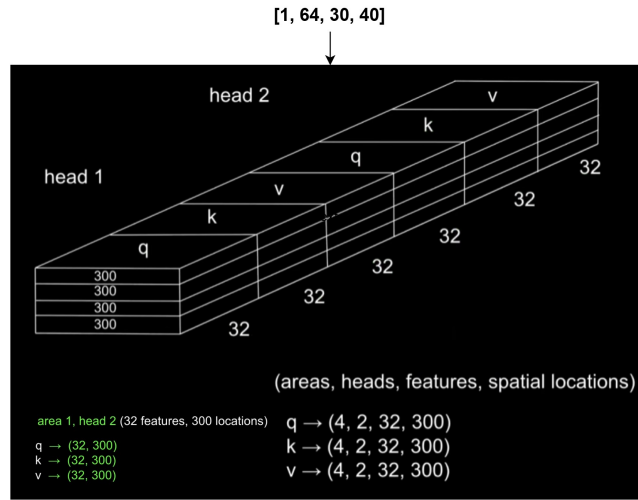


Figura 23 – Representação espacial do tensor após operação de redimensionamento e achatamento. Extraído de (TORNERO, 2025).

$$\text{Attention} = V \cdot \left(\text{softmax} \left(\frac{Q^T \cdot K}{\sqrt{d_k}} \right) \right)^T \quad (4.5)$$

Com base na equação acima, procede-se à operação de transposição na matriz q , resultando em Q^T . A matriz Q (*Query*, em inglês) representa as consultas, isto é, as características ou aspectos da imagem aos quais o modelo tenta dar foco, como bordas, texturas ou objetos.

Em seguida, realiza-se o produto ponto com a matriz k , que representa as chaves (*keys*), isto é, regiões potenciais da imagem que podem conter informações relevantes relacionadas àquelas características. Ao multiplicar essas matrizes, obtém-se a matriz de atenção, formada por pesos que capturam a similaridade entre as consultas e as chaves.

Os pesos de atenção podem ser interpretados como uma medida de quão relevante cada chave é para cada consulta. Esses valores são então normalizados para o intervalo entre 0 e 1 por meio da função *Softmax*. A matriz resultante é transposta e, por fim, multiplicada pela matriz v . A matriz de valores representa o conteúdo ou informação visual que será repassada depois dos pesos serem calculados. Estos valores contém características relevantes dos dados de entrada, os quais o modelo usa para tomar decisões. Os pesos de atenção são usados para calcular uma soma ponderada dos valores (v), na qual o peso de cada valor (v) é determinado pelo peso de atenção correspondente. Portanto, a saída da atenção é a soma ponderada dos valores (v) baseada na similaridade entre as consultas (q) e as chaves (k). Este processo é executado em paralelo para as quatro áreas e para as duas cabeças de atenção, concatenando as cabeças ao longo da dimensão correspondente. Assim, a matriz resultante de 32×300 , apresentada na Figura 24, é reorganizada para formar um tensor de dimensão $[4, 2, 32, 300]$. Em seguida, esse tensor é transformado

para a forma $x = [1, 64, 30, 40]$. O tensor de valores assume a mesma estrutura, isto é, $v = [1, 64, 30, 40]$.

O tensor v é essencial para a geração do tensor de codificação posicional (*positional encoding*), uma vez que contém características ricas em contexto extraídas da imagem, incluindo informação sobre "o quê" e "onde" estão os objetos de interesse. Os codificadores posicionais são fundamentais, pois introduzem consciência espacial em um mecanismo de atenção que, de outra forma, seria indiferente à localização das características.

A matriz de valores é então processada por meio de uma convolução de separação, utilizando um núcleo de 7×7 aplicado independentemente sobre cada canal do tensor v . O resultado dessa operação é o tensor de codificação posicional, que fornece contexto sobre onde cada característica está localizada dentro da *grid* espacial (TORNERO, 2025).

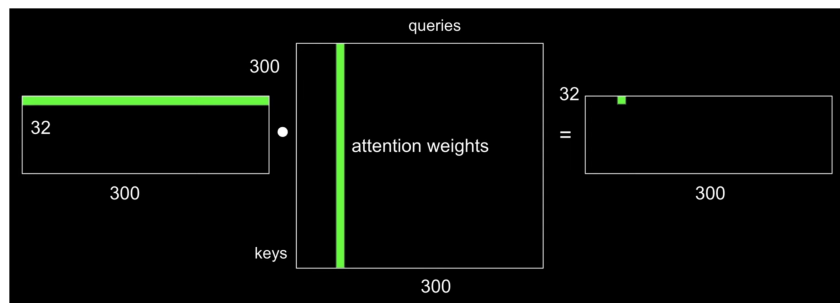


Figura 24 – Matriz resultante de valores de atenção. Extraído de (TORNERO, 2025).

Por fim, somam-se as matrizes de atenção, que contém informação sobre como cada característica se relaciona com as demais, à matriz de codificação posicional, gerando, assim, um novo tensor que contém informação sobre o que cada característica representa e onde ela se encontra espacialmente (Figura 25). Em seguida, aplica-se uma operação de linearização, correspondente ao último passo do módulo de atenção, deixando a matriz pronta para prosseguir com as operações subsequentes.

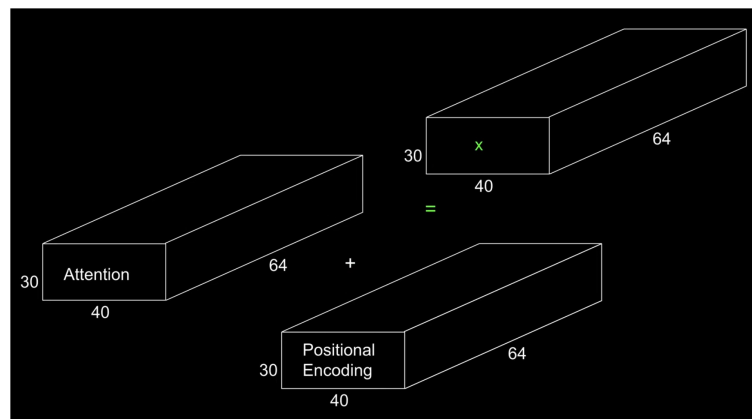


Figura 25 – Tensor resultante da operação de soma entre o tensor de atenção e o tensor de codificação posicional. Extraído de (TORNERO, 2025).

Na etapa conhecida como *pescoço* (*Neck*), utilizam-se as características extraídas pela espinha dorsal (*Backbone*) e procede-se à sua fusão e refinamento para a tarefa de detecção de objetos. Operações de superamostragem (*upsampling*) integram detalhes provenientes de estágios anteriores, combinando-os com características abstratas mais profundas por meio de camadas de concatenação. No *Neck* existem camadas **A2C2f**; entretanto, nesse estágio elas atuam de forma distinta, não empregando o mecanismo de atenção de área, funcionando de maneira mais semelhante a um bloco **C3K2**. Esses blocos realizam a combinação e o enriquecimento de características com o objetivo de melhorar o desempenho final do detector.

Na *cabeça* (*Head*), encontra-se o módulo responsável pela detecção propriamente dita, que processa as características refinadas provenientes de três escalas diferentes. Esse módulo realiza a predição das caixas delimitadoras, da classe do objeto e da pontuação de confiança, identificando de forma eficaz os objetos presentes na imagem de entrada.

É importante mencionar que essa arquitetura é utilizada em todas as variantes do YOLOv12 (nano, pequeno, médio, grande e extra-grande). Nota-se também que os tamanhos específicos dos tensores seguem fórmulas que dependem diretamente dos valores de “MC” e “W” de cada modelo. Exemplos concretos desses tamanhos são apresentados na Seção 5.5.

5 Metodologia

Este capítulo se divide em 5 grandes seções, uma sobre os Softwares e as quatro restantes, mostradas na Figura 26 referentes às tarefas de aquisição de dados, processamento dos dados, anotação de amostras e treinamento da Rede Neural. Todas estas etapas são detalhadas nas seções subsequentes.

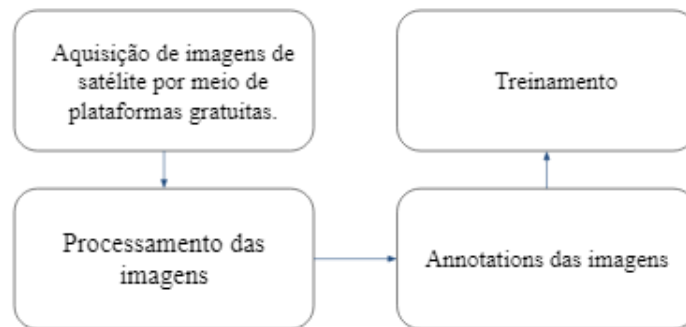


Figura 26 – Etapas para realização do trabalho.

5.1 Softwares auxiliares

Foram utilizados quatro softwares principais para realizar todo o processo de tratamento e treinamento dos dados, conforme ilustrado na Tabela 5.1. O Sentinel Applications Platform (SNAP), por meio da API python denominada *ESA SNAPPY* foi o software escolhido para o processamento de imagens. Ele é ideal para o processamento e a análise de dados de Observação da Terra, sendo capaz de lidar com informações de diferentes sensores por meio de suas caixas de ferramentas (Toolboxes). Essa plataforma é de uso gratuito e foi desenvolvida pelas empresas Brockmann Consult, Skywatch, Sensor e CS (SNAP - Earth Online, 2024). O software QGIS é um sistema de informação geográfico e foi utilizado como ferramenta para comparação e validação da existência de embarcações nas imagens SAR, por meio dos basemaps disponíveis no Plugin *World Map Services*. Para realizar as anotações das embarcações nas imagens, foi utilizado a plataforma *Robo-flow*. Por fim, para o treinamento da rede com os dados personalizados, o Google Colab foi a plataforma escolhida. Ele oferece acesso gratuito a recursos computacionais, sendo especialmente útil para aprendizado de máquina e ciência de dados(COLAB.GOOGLE,).

Software	Aplicação
SNAP (API ESA) SNAPPY)	Processamento de imagens SAR
QGIS	Comparação visual com imagens ópticas
Roboflow	Geração de arquivos XML com coordenadas de caixas delimitadoras
Google Colab	Desenvolvimento de scripts e treinamento de redes neurais

Tabela 2 – Softwares utilizados e suas aplicações

5.2 Base de dados SAR

Nesta fase do trabalho, coletam-se imagens das Regiões de Interesse com o propósito de obter amostras, que, neste caso, consistem em embarcações.

O satélite escolhido para cumprir a nossa missão foi o Sentinel-1. Na realidade, Sentinel-1 é o nome da missão que inicialmente compreendeu os satélites Sentinel-1A e Sentinel-1B, satélites gêmeos em órbita polar heliossíncrona, localizados no mesmo plano orbital e deslocados em fase de 180 graus. Cada um deles possui um instrumento SAR em banda C, disponível para operar em qualquer condição climática, tanto em períodos diurnos quanto noturnos. Infelizmente, o satélite Sentinel-1B sofreu uma falha de energia em 2021 e ficou inoperável. Em dezembro de 2024, foi lançado o Sentinel-1C, completando novamente a constelação de satélites gêmeos. A resolução temporal desses satélites é de 12 dias e 175 órbitas por ciclo para um único satélite. Informações relevantes para este satélite podem ser encontradas na tabela 3.

Parâmetro	Valor
Altitude	693 km
Inclinação	98,18°
Período	98,6 minutos
Ciclo	12 dias
Desvio do tubo de referência	± 120 m
Hora local no nó ascendente	18:00 horas

Tabela 3 – Parâmetros orbitais do satélite Sentinel-1

As imagens foram obtidas de duas fontes, a Alaska Satellite Facility (ASF), que é uma página administrada pela NASA e que oferece muitos produtos satelitais, incluindo os produtos do Sentinel 1 ([ASF – Alaska Satellite Facility, 2024](#)), e o Sentinel Hub, página para baixar exclusivamente produtos Sentinel, pertencente à Agência Espacial Europeia. A figura 27 mostra parte do processo para baixar os produtos satelitais.

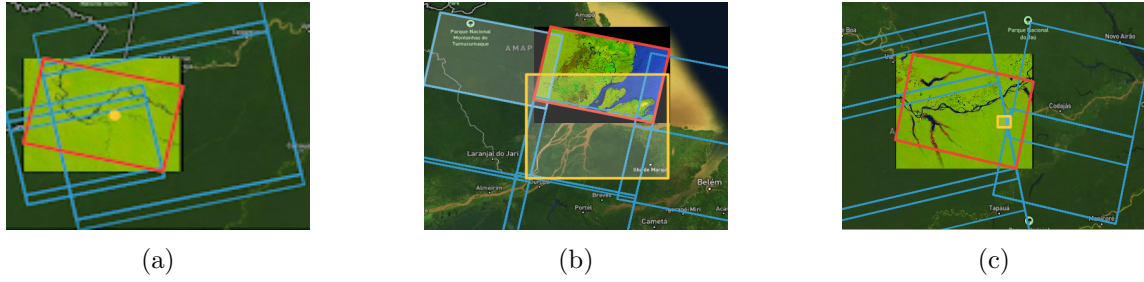


Figura 27 – Exemplos das áreas de aquisição do Sentinel-1 para as regiões de interesse. (a) Tabatinga, (b) Delta do Amazonas, e (c) Rio Solimões.

Os satélites Sentinel-1 apresentam distintos tipos de modos de aquisição de dados, dependendo da plataforma em questão. Na área de estudo, o método de aquisição mais usual é o Interferometric Wide Swath Mode (IW), portanto, todas as imagens a serem utilizadas para a criação do dataset serão neste modo, além disso se utilizaram imagens Ground Range Detected (GRD) em alta definição (GRDH). Parâmetros importantes desse modo estão contidos na tabela 4.

Para explicar o processo de aquisição de dados IW, vamos observar a Figura 28. Nela, é possível ver três subfaixas que compõem uma faixa completa, chamada de faixa larga, e têm um comprimento de 250 km (Copernicus – European Space Agency, 2024). Para este modo, se usa a técnica chamada *TOPSAR*, que consiste basicamente em alternar o feixe da antena entre as três subfaixas, e através de um sistema eletrônico, o feixe é direcionado de trás para frente na direção do voo, representado pelas setas em cor preta na Figura 29.

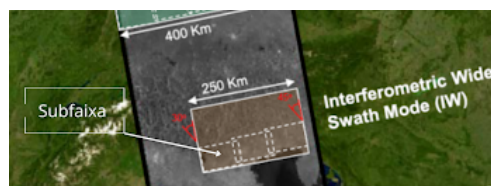


Figura 28 – Representação das subfaixas no modo IW (ØSTERGAARD et al., 2013).

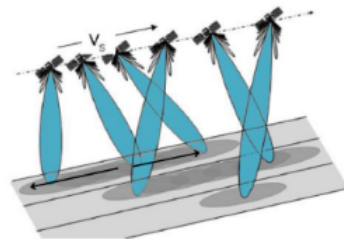


Figura 29 – Movimento de trás para frente do feixe (Earth Science Data Systems, NASA, 2016).

Parâmetro	Valor
Largura mínima da faixa terrestre	250 km
Faixa de ângulo de incidência	29,1° a 46,0°
Número de subfaixas	3
Ângulo de direção em azimuth	$\pm 0,6^\circ$
Resolução em azimuth	20 m
Resolução de alcance terrestre	5 m
Opções de polarização	Simples (HH ou VV) ou dupla (HH+HV ou VV+VH)

Tabela 4 – Parâmetros importantes do modo IW.

5.2.1 Região de interesse e zonas de aquisição das imagens

Um dos objetivos do presente trabalho é criar uma base de dados de imagens SAR, extraídas integralmente em território nacional, contendo amostras de embarcações para o treinamento de uma rede neural. O objetivo é, no futuro, estabelecer um sistema eficaz para a detecção e monitoramento de embarcações na complexa tríplice fronteira, que abrange os territórios do Brasil, Peru e Colômbia, mostrados na Figura 30. Focando especificamente na área de Tabatinga, situada no estado do Amazonas, almeja-se aprimorar a vigilância e o controle das atividades de navegação nessa região crítica do Rio Amazonas.

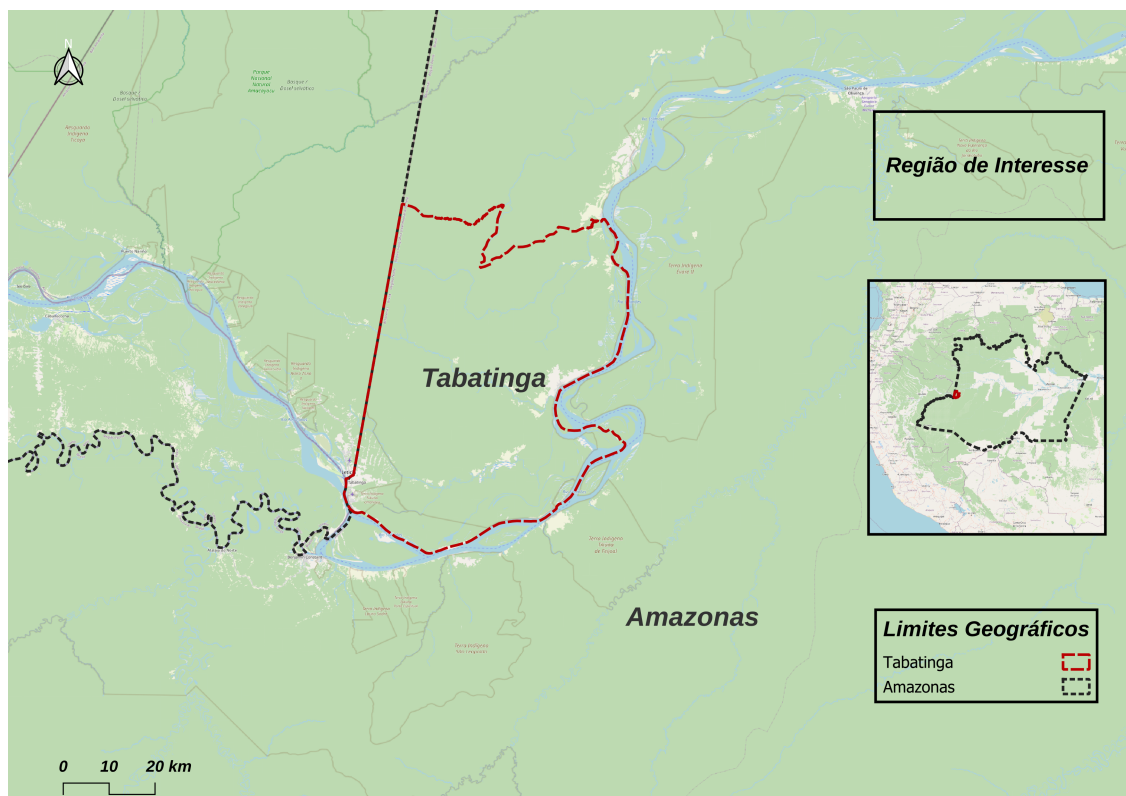


Figura 30 – Mapa da Região de Interesse (ROI).

Contudo, é relevante destacar que a metodologia empregada transcende as fronteiras geográficas estritas de Tabatinga. Para o treinamento da rede neural, nossa abordagem inclui não apenas amostras de embarcações específicas dessa localidade, mas também dados provenientes de outras duas áreas igualmente vitais e movimentadas ao longo do Rio Amazonas: o Delta do Amazonas e o Rio Solimões.

Essa escolha estratégica de incorporar amostras de diferentes regiões do Amazonas amplia significativamente a robustez e a generalização do sistema de detecção de embarcações. Compreender a variação nas características das embarcações, dependendo da região, é fundamental para garantir que nosso sistema seja capaz de lidar com uma ampla gama de situações. Na Figura 31, apresenta-se exemplos visuais dessas amostras coletadas em diferentes regiões.

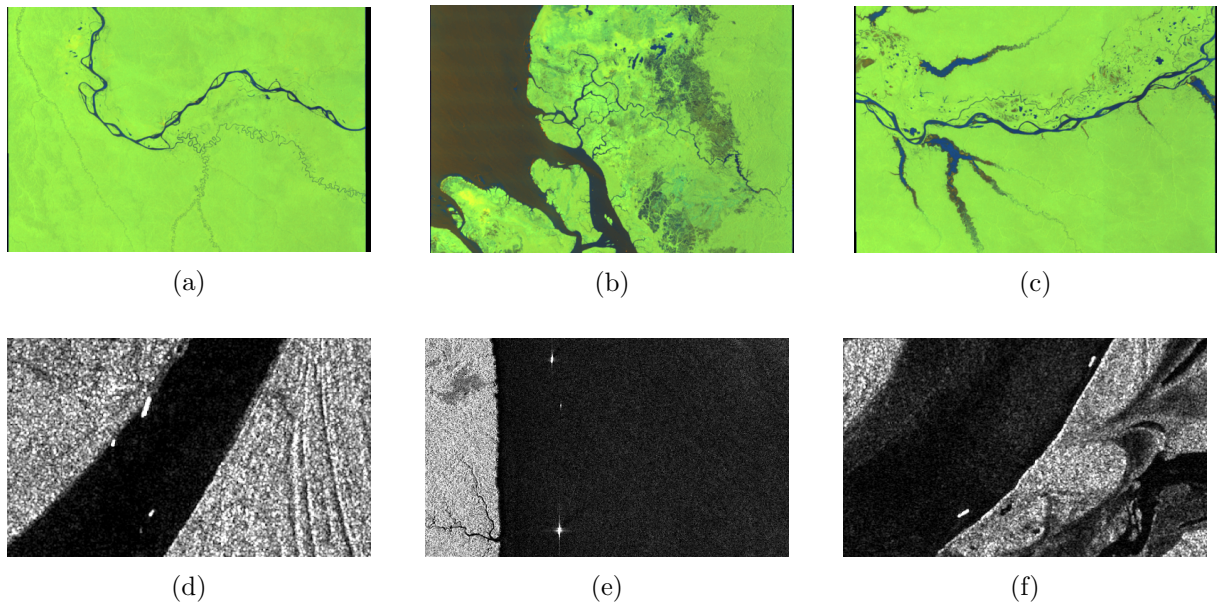


Figura 31 – Alguns exemplos de imagens de visualização rápida (superior) e imagens SAR (inferior).

Na figura acima, (a), (b) e (c) representam as imagens em cor verdadeira de cenas completas de Tabatinga, Delta do Amazonas e Rio Solimões, respectivamente. (d), (e) e (f) são exemplos de imagens de embarcações das cenas de Tabatinga, Delta do Amazonas e Rio Solimões, respectivamente, cortadas e extraídas das cenas SAR, com polarização VV (amplitude).

Para cada região, obtivemos várias cenas (tabela 5) nas quais foi possível observar aproximadamente cinco embarcações em cada uma. Essa pouca quantidade de embarcações encontradas em cada cena foi outro dos motivos pelo qual escolheu-se três áreas de interesse para as amostras. As embarcações são distinguíveis devido à sua forma alongada ou cilíndrica e ao seu brilho, chamado de retroespalhamento (*backscatter*) na teoria SAR. As informações de algumas dessas cenas estão contidas na Tabela 6. Cada cena

tem um tamanho aproximado de 1 GB, por isso, foi necessário adquirir armazenamento na nuvem, pois o Notebook no qual se estava trabalhando (Lenovo ideapad S340, 12GB de RAM DDR4 e processador Ryzen-5-3500U com disco SSD.M2 de 512 GB) não tinha a capacidade necessária para armazenar e processar essa quantidade de dados, considerando que outros arquivos e outros softwares usados em engenharia estavam instalados previamente nele. O plano adquirido para armazenamento na nuvem foi o Microsoft 365 Basic, que permite armazenar até 100 GB na nuvem, pagando uma quantidade econômica por mês. O peso total das imagens foi de 95 GB aproximadamente.

Local	Quantidade de cenas
Letícia	30
Delta Amazonas	15
Solimões	15

Tabela 5 – Cenas baixadas para cada região de interesse.

Local	Produto	Data de aquisição	Passe	Modo
Letícia	20230622T102109	22/06/2023 10:20:08	Descendente	IW
Delta A.	20230917T090546	17/09/2023 09:04:40	Descendente	IW
Solimões	20231005T095626	05/10/2023 09:55:22	Descendente	IW

Tabela 6 – Informações básicas sobre algumas cenas.

Após baixar as imagens e tê-las prontas para processamento, realizou-se o pré-processamento e adaptação das mesmas, preparando-as para a rede neural. Essas etapas serão executadas nas seções seguintes.

5.3 Pré-processamento das imagens

Inicialmente, as imagens SAR do Sentinel 1 são disponibilizadas em formato ZIP (chamado de produto) contendo informação da cena (preview, metadados, banner da ESA, etc.). No software SNAP, podemos abrir a imagem sem necessidade de descomprimir o arquivo. Os produtos Sentinel-1 são classificados em diferentes níveis, designadamente Nível 0, Nível 1 e Nível 2 (ESA, 2025). Neste trabalho, serão utilizados os produtos do Nível 1, que se subdividem em dois tipos: SLC e GRD, especificamente focando nos produtos GRD (*Ground Range Detected*).

Uma imagem GRD é o resultado de aplicar certos processamentos às imagens SLC. Cada píxel que conforma uma imagem SLC contém informação de amplitude e fase, essas informações formam uma imagem "complexa". Para obter uma imagem adequada à vista do ser humano, é imprescindível obter os valores de intensidade de cada píxel. Este valor retornará, quando aplicada a calibração radiométrica (seção 5.3.3), os valores reais de retroespalhamento da superfície imageada. Os valores de intensidade são obtidos pelos valores dos números digitais de cada célula quadrada (DN), que por sua vez, pode ser obtida por meio da raiz quadrada da soma dos quadrados das componentes reais (SLC_{real}) e imaginárias ($SLC_{imaginário}$) do sinal complexo (equação 5.1).

$$\text{Intensidade} = DN_{i,j}^2 = SLC_{real}^2 + SLC_{imaginário}^2 \quad (5.1)$$

Realizando este processo na imagem SLC, temos uma imagem Detectada, ou "*Detected*". Além do processamento "*Detected*", a imagem é projetada no solo, com correções geométricas e redução de ruído (*speckle*) ([Alaska Satellite Facility, 2025](#)).

Os produtos GRD são disponibilizados como produtos "prontos para uso". Apesar de já estarem prontos para uso, o usuário ainda pode realizar processamentos adicionais para aprimorar determinadas características, dependendo exclusivamente do tipo de estudo ou aplicação. Neste trabalho, foram aplicados os processamentos descritos na figura 32

O Pré - Processamento das imagens consiste em aplicar um recorte (*subset*), aplicar arquivo de órbita, remover ruído termal, realizar a calibração radiométrica e fazer a correção de terreno ([MARZUKI et al., 2021](#)). Porém, antes disso, temos que selecionar quadrantes que contenham uma certa porcentagem de pixels que representem corpos de água. Este pré-processamento será feito utilizando a linguagem de programação Python, além da API de Python da plataforma SNAP, chamada ESA-SNAPPY. O diagrama de blocos está ilustrado na Figura 32. Cada etapa está marcada com números, totalizando quatro etapas, desde baixar as imagens de radar, até a seleção manual de amostras de treinamento.

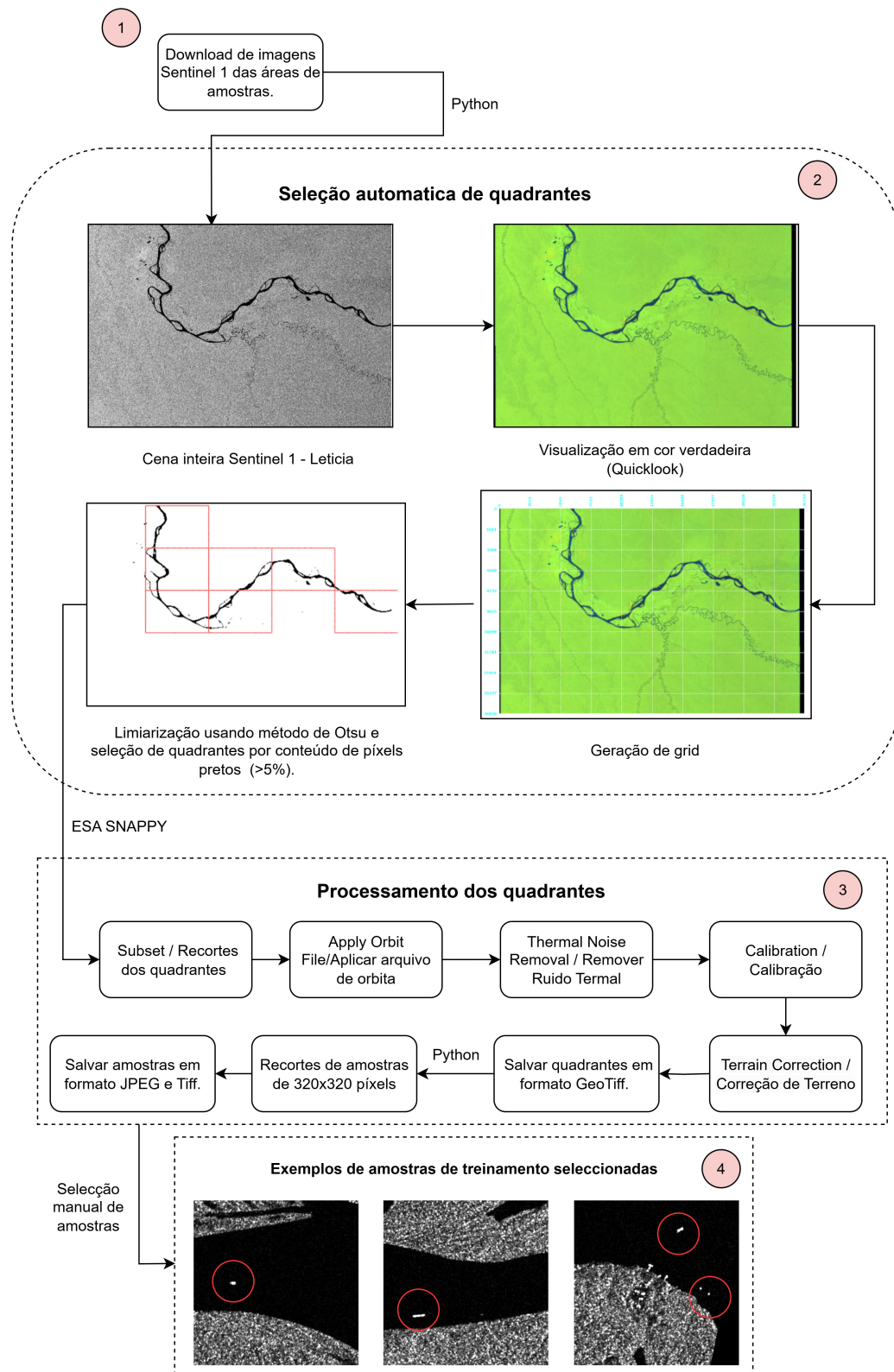


Figura 32 – Fluxo de pré-processamento das imagens SAR e seleção de amostras de treinamento.

Para o passo número um, e como foi mostrado na secção 5.2, as imagens foram baixadas de fontes gratuitas de dados. Para mais informações das cenas, consultar a figura 27 e as tabelas 5 e 6.

No segundo passo, a tarefa principal é selecionar os quadrantes de interesse, ou seja, aqueles que contenham corpos d'água. Para isso, foi utilizada principalmente a imagem de visualização rápida (*Quick-look*), em formato PNG, disponível dentro da pasta do produto da cena Sentinel-1. Essa escolha foi feita com a intenção de trabalhar com arquivos mais leves.

A imagem de visualização rápida é gerada a partir da composição de bandas da imagem SAR. Nesse processo, atribui-se a primeira polarização (VV) ao canal vermelho, a segunda polarização (VH) ao canal verde e, por fim, ao canal azul é atribuída a média dos valores absolutos de VV e VH.

Em seguida, a tarefa consiste em dividir a imagem por meio de uma *grid*. Para isso, é necessário extrair os valores de altura e largura tanto da cena SAR real quanto da imagem de visualização rápida. As dimensões da cena SAR são obtidas a partir dos metadados disponíveis dentro da pasta do produto, enquanto as dimensões da imagem de visualização rápida são extraídas com auxílio da biblioteca *Matplotlib*, que permite ler essa imagem e representá-la como uma matriz *NumPy* de três canais (RGB). A partir disso, basta obter o número de linhas (altura) e colunas (largura), em pixels, para determinar as dimensões da imagem.

Por fim, realizam-se divisões pelo número de linhas e colunas desejadas no *grid*, tanto para plotar essa malha sobre a imagem de visualização rápida quanto para obter as coordenadas reais correspondentes na cena SAR.

Tendo pronta a imagem separada por quadrantes, agora o desafio é selecionar, para processamento, unicamente os quadrantes de interesse, ou seja, os quadrantes que contenham uma porção de 'rio' dentro. Para isso, a abordagem foi binarizar a imagem de visualização rápida para separar os corpos d'água do terreno seco. A binarização é feita por meio de um limiar calculado a través do método de Otsu. Este algoritmo calcula um limiar "ótimo" ao maximizar a variância entre duas classes de pixels, que são separadas por esse limiar. De forma equivalente, esse limiar minimiza a variância intra-classe, definida como ([scikit-image developers, 2025](#)):

$$\sigma_w^2(t) = q_1(t)\sigma_1^2(t) + q_2(t)\sigma_2^2(t), \quad (5.2)$$

onde $q_1(t)$ e $q_2(t)$ representam as probabilidades acumuladas de ocorrência de pixels em cada classe, calculadas como

$$q_1(t) = \sum_{i=1}^t P(i), \quad q_2(t) = \sum_{i=t+1}^I P(i), \quad (5.3)$$

sendo $P(i)$ a probabilidade de ocorrência do nível de cinza i e I o número total de níveis de intensidade.

As médias de cada classe são definidas por

$$\mu_1(t) = \frac{\sum_{i=1}^t iP(i)}{q_1(t)}, \quad \mu_2(t) = \frac{\sum_{i=t+1}^I iP(i)}{q_2(t)}, \quad (5.4)$$

e, a partir delas, obtêm-se as variâncias intra-classe correspondentes:

$$\sigma_1^2(t) = \frac{\sum_{i=1}^t [i - \mu_1(t)]^2 P(i)}{q_1(t)}, \quad \sigma_2^2(t) = \frac{\sum_{i=t+1}^I [i - \mu_2(t)]^2 P(i)}{q_2(t)}. \quad (5.5)$$

O limiar ótimo é aquele que minimiza $\sigma_w^2(t)$, garantindo que a separação entre as duas classes maximize a distinção entre os grupos de pixels, ao mesmo tempo em que reduz a dispersão dentro de cada classe. Para isso, foi utilizada a biblioteca *scikit-image*, por meio da função *threshold_otsu*, no Python. Alguns resultados da aplicação deste método nas imagens podem ser consultados na Figura 33.

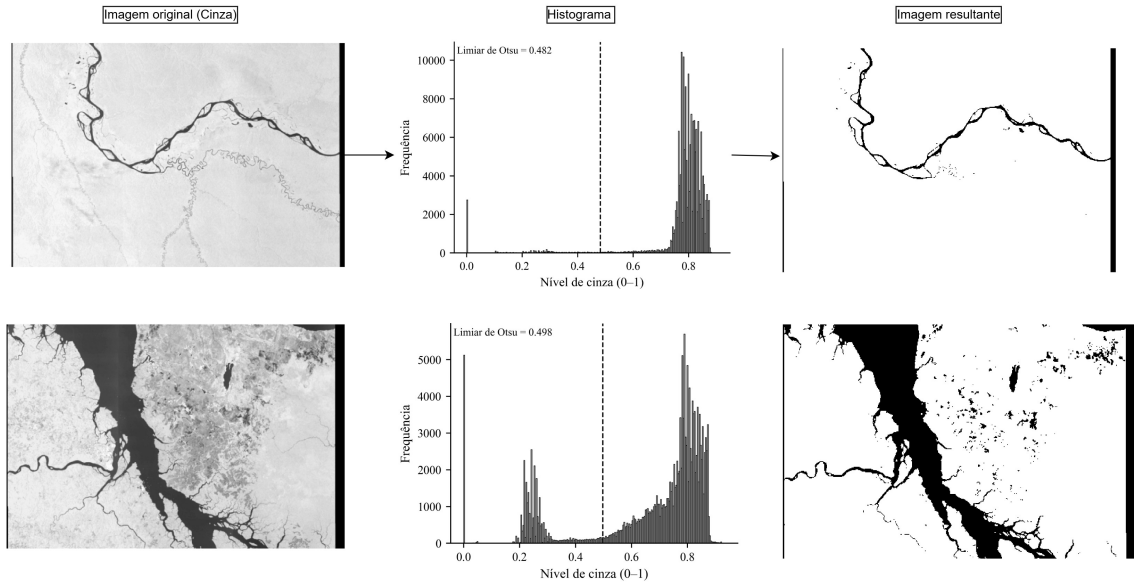


Figura 33 – Exemplos de limiarização usando o método de Otsu. No topo, observamos uma imagem referente a Leticia, nela, foi aplicado o limiar de 0,482. Embaixo, temos uma imagem representando uma zona do Delta do Amazonas. O limiar encontrado ideal para esta imagem foi de 0,498.

Por fim, após a geração da máscara binária, foi realizada uma análise por quadrantes com o objetivo de identificar regiões da cena contendo uma proporção relevante de pixels classificados como fundo (valor 0, representados em preto).

Dentro de cada quadrante, foi contabilizado o número total de pixels e, em seguida, o número de pixels pretos. A partir desses valores, calculou-se o percentual de pixels pretos em relação ao total da célula. Quando esse percentual ultrapassava um limiar pré-definido (no caso das cenas de Letícia, 5%), o quadrante era marcado como contendo predominância de fundo. As coordenadas desses quadrantes foram extraídas e armazenadas em um arquivo *CSV*. O objetivo desse procedimento é garantir que, no próximo passo (passo número três), o processamento subsequente seja realizado unicamente nos quadrantes de interesse.

O passo número três no esquema apresentado na figura 32 consiste no processamento dos quadrantes selecionados na etapa dois. Cada uma das etapas do processamento das imagens será explicada a seguir, fundamentado segundo [Filipponi \(2019\)](#) e outros autores:

5.3.0.1 Recortes (*Subset*)

Aplicar um recorte na imagem reduz significativamente o volume de dados com que se irá trabalhar. Uma imagem SAR sem recorte pode pesar mais de 800 MB. Ao fazer um recorte para as áreas de interesse, que no nosso caso são regiões pequenas contendo barcos, o peso da imagem pode diminuir até 99%, passando, por exemplo, de 800 MB para 8 MB. Isso torna o pré-processamento muito mais rápido. Estes recortes são os quadrantes selecionados no passo dois. As coordenadas de cada quadrante são extraídas do arquivo *CSV* gerado.

5.3.1 Aplicar Arquivo de Órbita

Esse passo consiste em atualizar os vetores do estado da órbita do satélite em uso ([European Space Agency, a](#)). É importante atualizar o arquivo de órbita para obter informação precisa sobre posição e velocidade do satélite, pois a geocodificação feita na última etapa do processamento (correção do terreno) utiliza estas informações para posicionar os pixels da imagem na posição correta na superfície da terra ([BRAUN, 2019](#)).

5.3.2 Remover ruído termal

No presente trabalho usamos imagens de Intensidade do Sentinel 1. Esse tipo de imagem apresenta um ruído termal, sendo muito mais evidente na polarização cruzada, como VH ou HV, e em dados SAR com baixo retroespalhamento ([European Space Agency, b](#)). A abordagem utilizada pela ESA para remover esse ruído consiste na subtração da intensidade medida no retroespalhamento de cada pixel pelos valores de ruído previamente calculados. Para isso, é gerado um arquivo *XML* que contém, para grupos de pixels em determinadas linhas da imagem, os valores de ruído equivalente. Para os pixels situados

entre esses pontos de referência, o valor de ruído é estimado por meio de interpolação bilinear, assegurando que cada pixel receba o valor correspondente de ruído a ser removido.

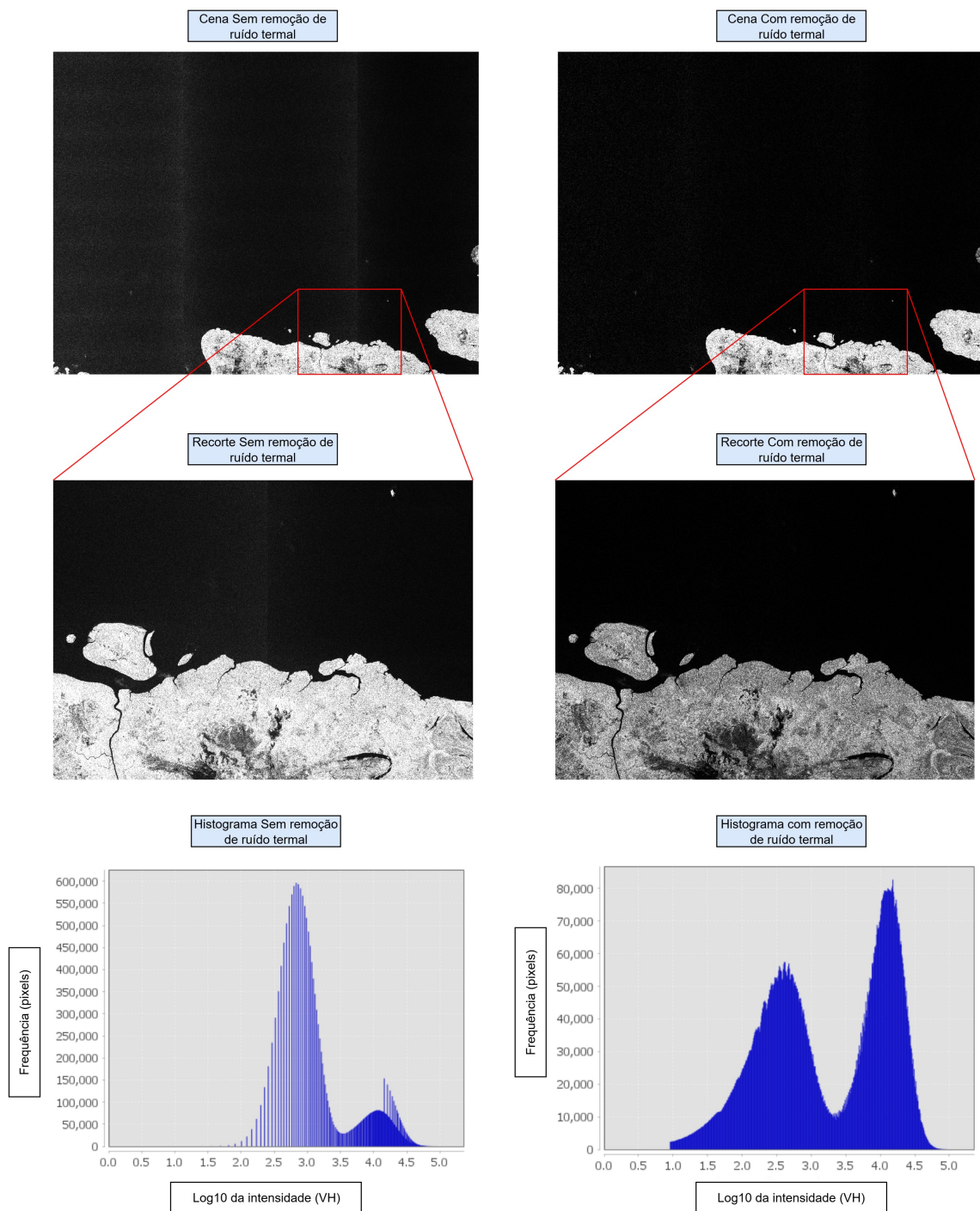


Figura 34 – Remoção do ruído termal na polarização VH. Temos na direita cena, imagem e histograma sem remoção do ruído termal. Na esquerda, os elementos com remoção do ruído termal.

A Figura 34 mostra uma comparação entre uma imagem sem remoção de ruído versus outra com remoção aplicada, incluindo os histogramas correspondentes. Observa-se que, no histograma da imagem sem a remoção do ruído térmico, há uma concentração ar-

tificial de pixels em determinados valores de intensidade (convertidos para a escala $\log_{10}$), a qual não corresponde às condições físicas reais da cena. Em contraste, no histograma da imagem corrigida pela remoção do ruído térmico, verifica-se uma distribuição mais uniforme dos pixels, representando de forma mais fidedigna a resposta de retroespalhamento do alvo.

5.3.3 Calibração Radiométrica.

A Calibração radiométrica permite relacionar diretamente os valores de cada pixel da imagem com valores de retroespalhamento da superfície. Embora para análises qualitativas a calibração não seja essencial, para análises quantitativas é fundamental. Existem 4 unidades de calibração, sendo estas β_0 , σ_0 , γ_0 , DN. O tipo de calibração mais comum é a σ_0 , pois ela está diretamente relacionada a informações do terreno, representada na Figura 35. Sigma Zero, σ_0 É a variação espacial do coeficiente de retroespalhamento pela geometria da célula de resolução e é uma medida fundamental nas imagens de radar. Ele é calculado pela razão entre uma área de seção transversal σ iluminada pelo pulso de onda e uma unidade de área (A) de reflexão perfeita (retroespalhada isotropicamente):

$$\sigma^0 = \frac{\sigma}{A} \quad (5.6)$$

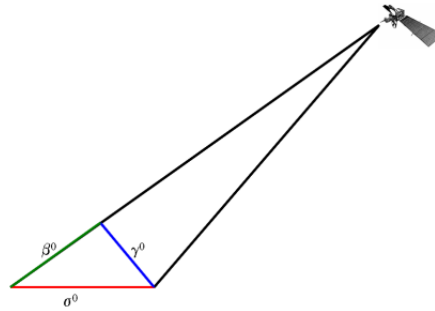


Figura 35 – Representação física das unidades de calibração, sendo σ_0 a mais utilizada para estudos quantitativos (GUMELAR; LIU; RIZKIYANTO, 2017).

5.3.4 Correção do Terreno

A correção do terreno inclui a geocodificação, a qual projeta os pixels em um sistema de referência de coordenadas e permite o uso conjunto das imagens com outros dados geoespaciais, como arquivos vetoriais ou imagens ópticas. Deve ser realizada como última etapa, pois envolve o reamostramento dos pixels.(BRAUN, 2019). Os produtos GRD são projetados para o alcance em solo utilizando o modelo de elipsoide da Terra WGS84. Enquanto a correção de elipsoide é suficiente para geocodificar uma imagem SAR em superfícies planas, os efeitos de encurtamento (*foreshortening*) e sobreposição (*layover*), precisam ser corrigidos com o uso de um modelo digital de elevação. Uma breve descrição sobre estes efeitos geométricos é apresentada a seguir:

5.3.4.1 Encurtamento (*Foreshortening*)

O tempo de iluminação dos pulsos gerados pelo radar com a base de uma montanha, por exemplo, geralmente é maior do que o tempo de contato da iluminação e o topo. Isso faz com que a pendente sofra um encurtamento. Existe uma exceção, e ela aparece quando o ângulo de incidência local θ é igual a 90° . Ela está ilustrada na Figura 36 (a), (d).

5.3.4.2 Inversão do terreno (*Layover*)

Quando o topo da montanha está mais próximo do satélite do que a base, a superfície do topo fica mais exposta à radiação electromagnética. Provocando que volte antes do resto da radiação que interage com a geometria restante. Isso provoca uma inversão do terreno ou *Layover*. Pois são mapeados primeiro os pontos mais próximos do satélite e só depois o resto de elementos. Este efeito está ilustrado na Figura 36 (b), (e).

5.3.4.3 Sombreamento (*Shadowing*)

As áreas atrás de relevos acentuados não são iluminadas adequadamente, isso acontece pela geometria de aquisição e iluminação dos sistemas SAR. Essas áreas aparecem em preto na imagem e contêm pouca ou nenhuma informação. Este efeito é mostrado na Figura 36 (c), (f).

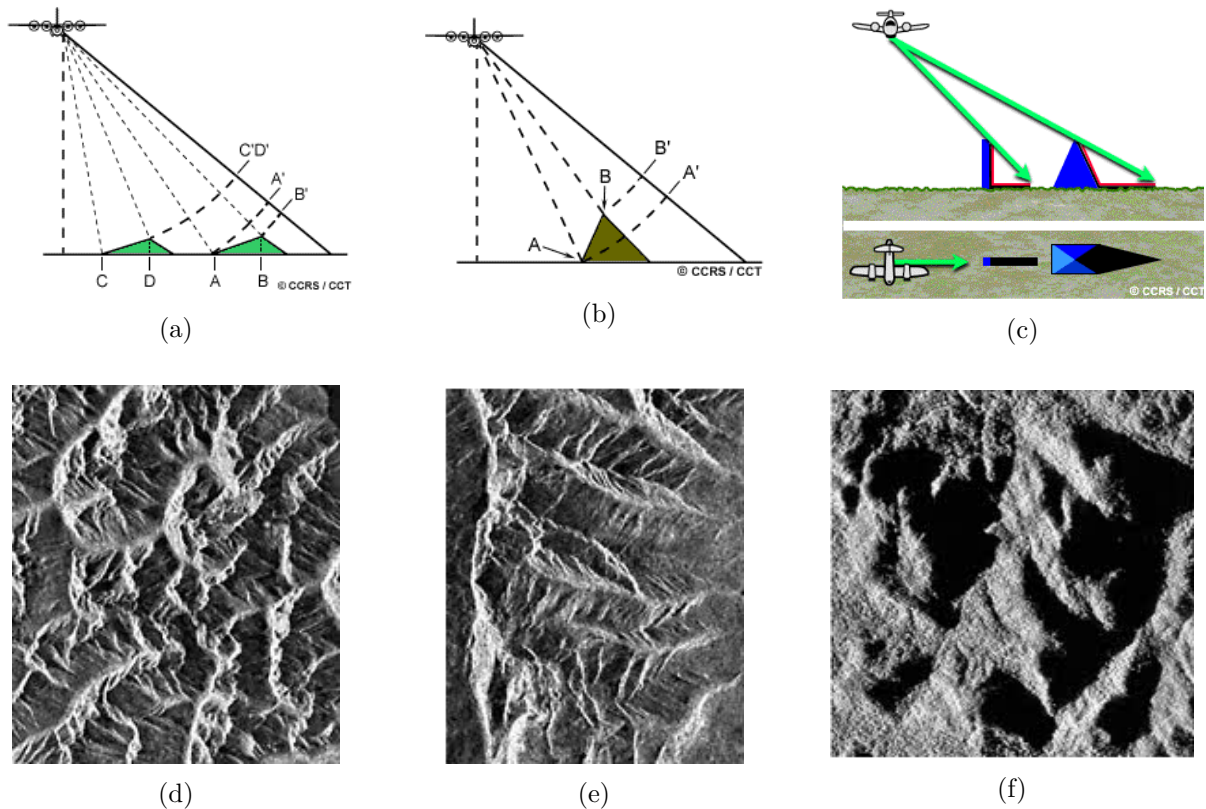


Figura 36 – Distorções geométricas ilustradas em (a), (b), (c), para Encurtamento, inversão do terreno e sombreamento, respectivamente. Em (d),(e),(f), observa-se exemplos reais (PATRUNO, 2014).

Se bem, a pasta do produto Sentinel-1 contém as imagens georreferenciadas (VV e VH, para Sentinel 1) em formato tiff,

As imagens SAR quando não corrigidas geometricamente, representam a geometria do radar (alcance, azimuth). Aplicando a correção geométrica (elipsoide ou do terreno), as coordenadas de pixels (alcance, azimuth) são convertidas em um sistema de coordenadas de mapa (latitude e longitude). Quando trabalhamos com o software *SNAP*, as bandas (isto é, as imagens VV/VH) são abertas inicialmente na grade de alcance e azimuth, correspondendo à geometria de aquisição do radar. Ao aplicarmos a correção de terreno, é gerado um novo produto no qual a imagem é reprojeta para coordenadas geográficas reais (Lat/Lon). Como resultado desse processo de geocodificação, a cena pode parecer "espelhada" ou "rotacionada" em relação à visualização original, conforme ilustrado na Figura 37.

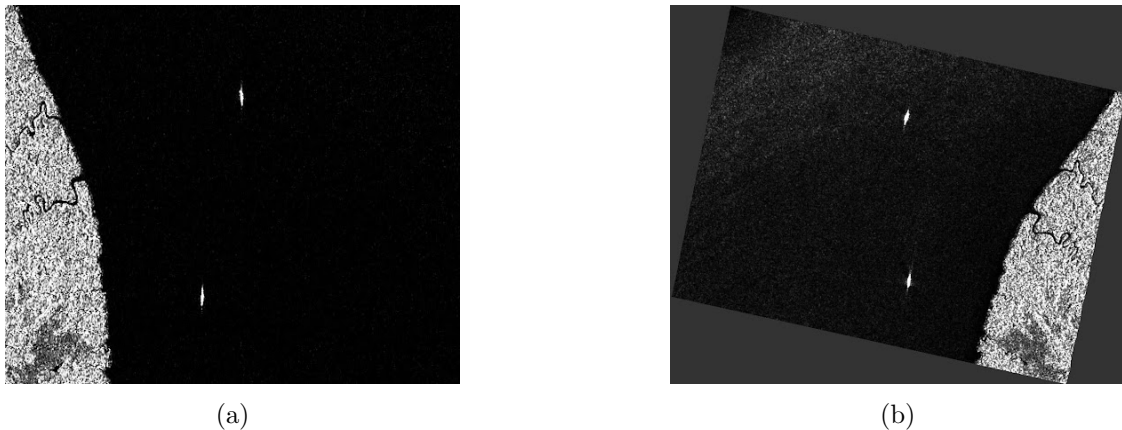


Figura 37 – Correção do terreno. (a) Imagem sem correção do terreno. (b) Imagem com correção de terreno. Se obtém uma imagem espelhada e rotacionada, pois esse é o formato de aquisição da imagem por parte do satélite.

Em resumo, temos três mecanismos principais de retroespalhamento: especular (*Rough Surface Scattering*), volumétrico (*Volume Scattering*) e de refletor de canto (*Double bounce scattering*). Estes três mecanismos estão ilustrados na figura 38.

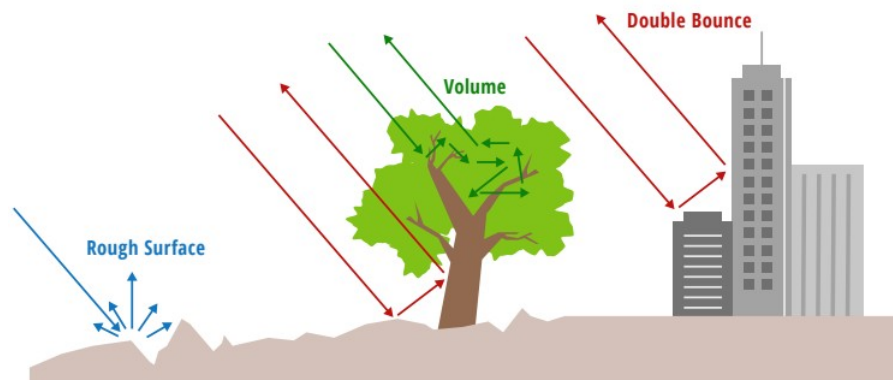


Figura 38 – Mecanismos principais de retroespalhamento em sistemas SAR (FLORES et al., 2019).

O Sentinel 1 conta com várias polarizações: HH, HV, VV e VH. Os diferentes mecanismos de retroespalhamento não contribuem igualmente para todas as polarizações. Cada polarização “favorece” determinados tipos de espalhamento, de modo que a potência retroespalhada em cada canal polarimétrico segue um comportamento característico. Existe uma predominância dos tipos de retroespalhamento em função da polarização (FLORES et al., 2019). Esta relação entre mecanismo de retroespalhamento e polarização está mostrada na tabela 7

Tipo de espalhamento	Relação de intensidade relativa
Especular	$ S_{VV} > S_{HH} > S_{HV} $ ou $ S_{VH} $
Refletor de canto	$ S_{HH} > S_{VV} > S_{HV} $ ou $ S_{VH} $
Volumétrico	Principal fonte de $ S_{HV} $ e $ S_{VH} $

Tabela 7 – Intensidade relativa de espalhamento por polarização (FLORES et al., 2019).

Na tabela acima, $|S_{VV}|$, por exemplo, representa a intensidade do retroespalhamento registrado quando o radar transmite e recebe em polarização vertical. Analisando a tabela 7, concluímos que as co-polarizações HH e VV são ideais para detectar estruturas naturais e artificiais tais como árvores e prédios. Ainda, a polarização VV é recomendada por sua eficácia para monitorar superfícies naturais como solos e corpos de água. A polarização cruzada, tanto VH quando HV apresenta eficiência para diversos estudos envolvendo áreas volumétricas.

O mecanismo de interesse principal neste trabalho é o do tipo refletor de canto. Os navios aparecem como pixels brilhantes principalmente devido a esse tipo de mecanismo (WANG et al., 2022) como visto na figura 39.

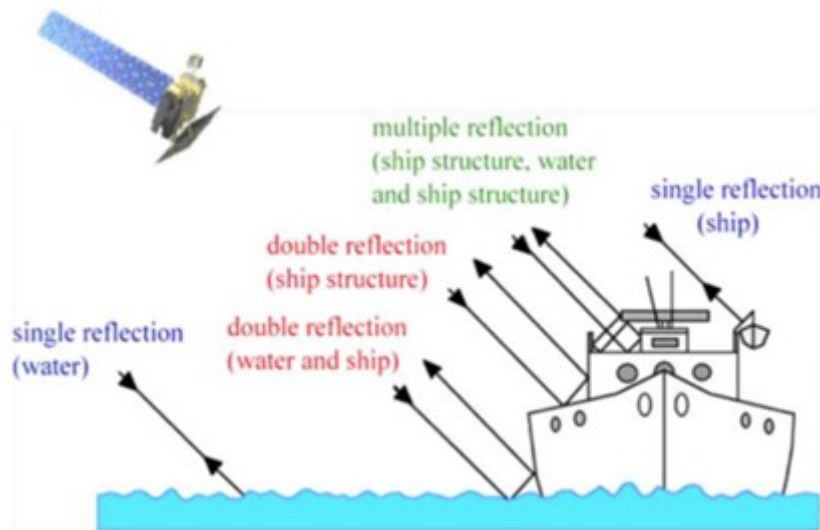


Figura 39 – Mecanismos presentes em alvos do tipo embarcações (WANG et al., 2022).

Uma comparação entre a polarização simples e cruzada em função da intensidade de retroespalhamento é apresentada na Figura 40. Esta análise foi feita por meio da ferramenta Gráfico de perfil (*profile plot*) no SNAP.

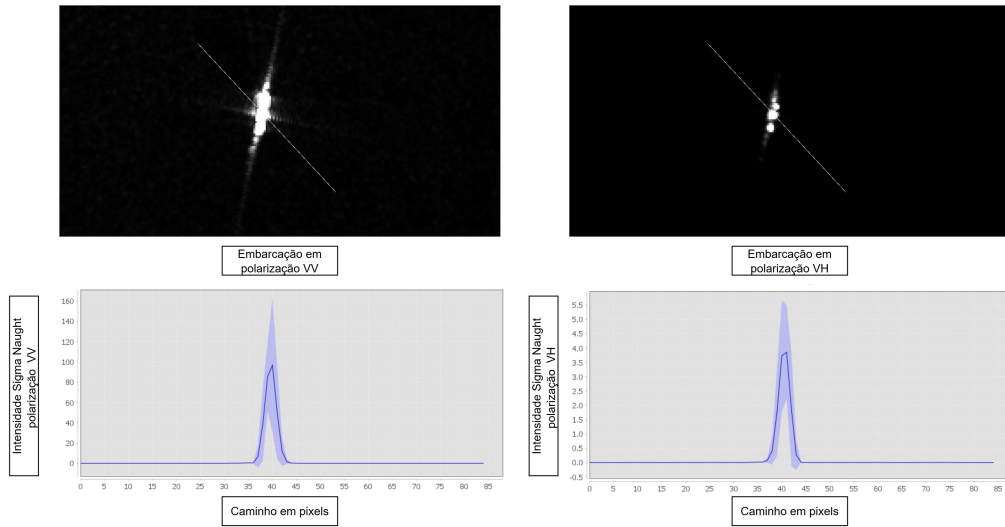


Figura 40 – Histograma de intensidade para retroespalhamento do tipo refletor de canto em polarizações VV (esquerda) e VH (direita).

Após a realização desses passos, duas pastas foram geradas: uma contendo os recortes em formato GeoTIFF e outra contendo os recortes em formato JPEG. Inicialmente, na pasta de recortes JPEG, foram selecionados visualmente os recortes com presença potencial de embarcações, estes recortes contam com uma profundidade de 8 bits. Em seguida, as embarcações identificadas nesses recortes foram validadas utilizando as imagens GeoTIFF.

As imagens GeoTIFF cumprem o papel de servir como referência geográfica, permitindo a comparação direta com as imagens do Google Satellite no QGIS, uma vez que, por estarem georreferenciadas, são posicionadas exatamente no local correspondente no mapa. Esse procedimento foi realizado para todos os recortes selecionados, que foram organizados em uma pasta específica, preparados para a etapa seguinte, referente à anotação das amostras de treinamento.

Em princípio, o *Ultralytics* oferece suporte apenas a imagens com profundidade de 8 bits, tanto para o treinamento quanto para a inferência. É possível adicionar funções no código-fonte do *YOLO*, especificamente no *dataloader*, para converter e ler imagens em outros formatos; contudo, trata-se de um processo adicional que requer a edição do código original, o que pode gerar inconsistências entre bibliotecas e etapas do processamento.

Dessa forma, neste trabalho optou-se por realizar o treinamento utilizando imagens *JPEG* de 8 bits. Além disso, a plataforma *Roboflow*, utilizada para a anotação das amostras, não possui suporte nativo para imagens no formato *TIF* ou *GeoTIFF*.

5.4 Anotação de amostras

Anotação de amostras é o processo de anexar rótulos significativos aos dados brutos para ajudar algoritmos de IA a compreender e classificar as informações, ou seja, o objetivo é criar caixas delimitadoras (*Bounding Boxes*) ao redor dos objetos de interesse. Ditas caixas são retangulares e dentro dos quatro vértices que conformam ela teremos, principalmente, as embarcações. Esta tarefa foi feita por meio da plataforma *Roboflow*, a qual fornece uma interface rápida e robusta através da qual podemos anotar as nossas imagens ([Roboflow, 2025c](#)).

A anotação das amostras foi elaborado seguindo as recomendações a seguir ([Roboflow, 2025a](#)):

- **Rotular todos os objetos:** cada instância deve ser anotada, evitando falsos negativos que reduzem o desempenho do modelo.
- **Delimitar o objeto completo:** as caixas devem englobar todo o objeto, sem incluir áreas externas desnecessárias.
- **Considerar objetos ocluídos:** mesmo parcialmente encobertos, os objetos devem ser rotulados como se estivessem totalmente visíveis.
- **Criar caixas ajustadas:** as delimitações precisam ser justas, sem margens excessivas, mas sem cortar partes do objeto.

Resultados das anotações podem ser consultadas na figura [41](#):

A anotação de amostras em imagens SAR não segue exatamente o mesmo procedimento aplicado a imagens no espectro visível, como aquelas provenientes de satélites Sentinel-2 ou Landsat-8. Nesse contexto, o uso de ferramentas de apoio, como softwares de SIG e o Google Earth Pro, torna-se essencial. A comparação entre diferentes tipos de imagens auxilia a reduzir a ocorrência de falsos positivos ([WEI et al., 2020](#)).

Por fim, na figura [42](#) temos o fluxo seguido para chegar até a anotação de amostras. Inicia-se o processo com o quadrante cortado e processado (resultado do passo 2 na figura [32](#)). Logo, seleciona-se os recortes de tamanho 320x320 pixels que contenham embarcações (passo 4 na figura [32](#)). Por fim, procede-se à anotação de amostras na plataforma *Roboflow*.

No final, foram 359 imagens, com 674 amostras anotadas (promédio de 1,9 por imagem) distribuídas em 480 para treinamento, 123 para validação e 71 para teste.

No formato YOLOv12 para PyTorch, cada imagem possui um arquivo de anotação associado em extensão `.txt`. Nesse arquivo, cada linha corresponde a uma *bounding box* e contém cinco valores:



Figura 41 – Amostras de treinamento anotadas na plataforma *Roboflow*.

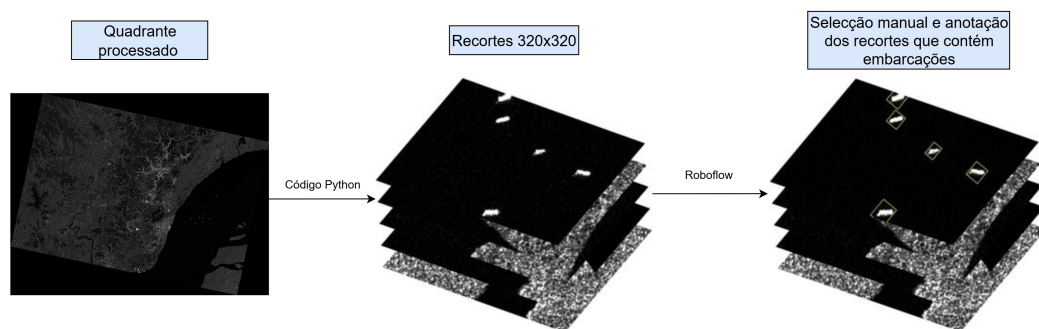


Figura 42 – Abordagem para recorte, selecção e anotação de amostras de treinamento.

```
class_id center_x center_y width height
```

Os campos são separados por espaço, sendo que as coordenadas são normalizadas no intervalo de 0 a 1. Dessa forma, os valores não estão expressos em pixels absolutos, mas em proporções relativas ao tamanho da imagem.

Para converter coordenadas em pixels para o formato normalizado ($xywh$), divida-se a posição x e a largura pela largura total da imagem, e a posição y e a altura pela altura total da imagem. O formato da anotação pode ser consultado na figura 43.

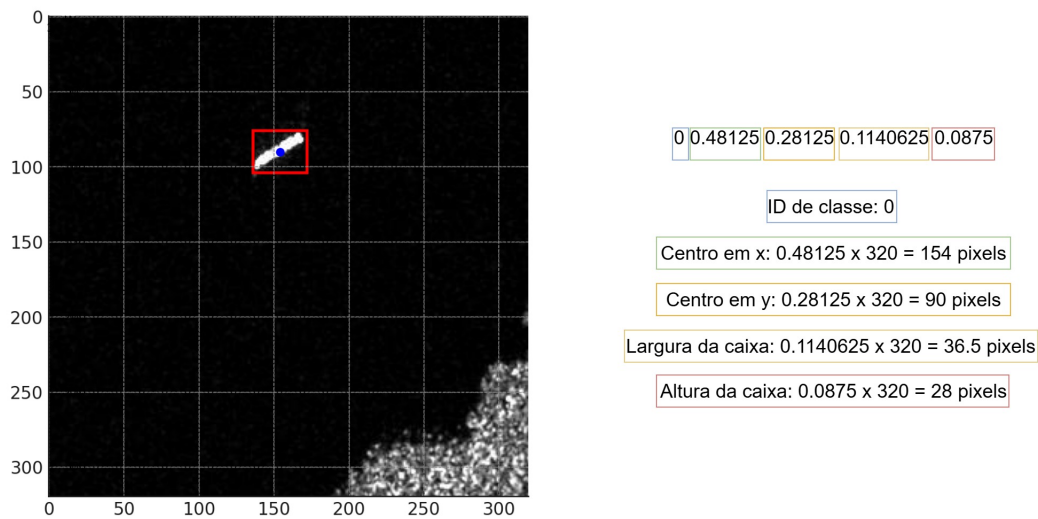


Figura 43 – Exemplo do formato das anotações. Na imagem da embarcação, o ponto azul na posição x,y representa o centro da caixa delimitadora. O quadro vermelho tem as dimensões de largura e altura representado no texto.

5.5 Treinamento do modelo YOLOV12

O treinamento do modelo YOLOv12 utilizando uma base de dados personalizada foi realizado seguindo as instruções de (Roboflow, 2025b). A página do *Roboflow* disponibiliza um conjunto de orientações para essa tarefa. Como utilizamos a base de dados de Embarcações, criada no *Roboflow*, foi necessário gerar uma chave *API* para acessar o espaço de trabalho do usuário e, assim, disponibilizar as bases de dados na plataforma.

Além disso, procedeu-se à instalação dos módulos do YOLOv12 diretamente a partir da página oficial do *GitHub* dos desenvolvedores. Após esses preparativos, entre outros passos, realizou-se o ajuste fino (*fine-tuning*) do modelo. O ajuste fino consiste em utilizar os parâmetros de um modelo previamente treinado para realizar um treinamento adicional em um conjunto de dados próprio, geralmente menor, que reflete diretamente as tarefas específicas para as quais o modelo será aplicado (IBM, 2024). Para isto, faz-se também um ajuste das configurações e hiperparâmetros a serem utilizados.

O YOLOv12 conta com cinco tamanhos de modelos: Nano(n), pequeno(s), médio(m), grande(l) e extra grande(x), sendo descritos na tabela 8. Cada tamanho do modelo difere em relação aos valores dos tensores utilizados nas operações internas da arquitetura. Um exemplo pode ser observado na Figura 44, onde são apresentados os valores dos tensores correspondentes ao modelo de tamanho **n**. Esses tensores dependem diretamente da largura “W” e da quantidade máxima de canais “MC” adotada na configuração do modelo.

Outro parâmetro importante que diferencia os tamanhos é o custo computacional: o modelo *nano* é o que menos demanda recursos, enquanto o *extra-grande* é o mais exi-

gente. Essa diferença decorre diretamente da quantidade de parâmetros presentes em cada variante, sendo aproximadamente 2,5, 9,1, 19,6, 26,5 e 59,3 milhões de parâmetros para os modelos **n**, **s**, **m**, **l** e **x**, respectivamente. Neste trabalho, será realizada uma comparação entre as variantes **n**, **s** e **m** do YOLOv12, a fim de selecionar aquela que melhor atende aos objetivos propostos. Os modelos grandes (**l** e **m**) exigem um poder computacional grande, que sobrepassa os recursos que o *Google Colab* oferece, inclusive na sua versão paga.

Tamanho	profundidade (d)	largura (w)	Canais (mc)	Camadas
n	0,50	0,25	1024	497
s	0,50	0,50	1024	497
m	0,50	1,00	512	533
l	1,00	1,00	512	895
x	1,00	1,50	512	895

Tabela 8 – Resumo dos tamanhos dos modelos YOLOv12.

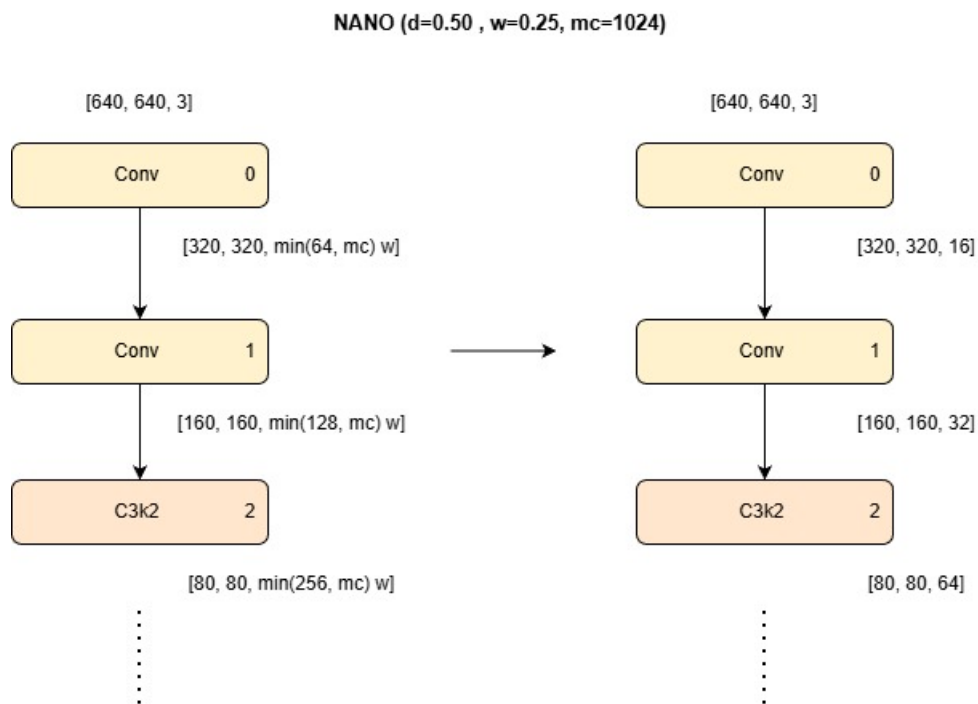


Figura 44 – Formato dos tensores da variante *Nano*

A base de dados que será utilizada pode ser considerada pequena por ter 359 imagens e 674 amostras de embarcações, assim, um número de épocas entre 100 e 300 é adequado (KUMAR, 2025). O otimizador é selecionado automaticamente quando usamos "*optimizer=Auto*" nas configurações do treinamento do nosso modelo. A partir desta escolha, o algoritmo usado foi o *AdamW*, com taxa de aprendizado de 0,002 e momentum de 0,9 (ULTRALYTICS, 2025). Outro hiperparâmetro importante é o tamanho do lote, que representa o número de amostras de treinamento processadas antes de recalcular/atualizar os parâmetros do modelo. Neste trabalho foi usado um número de lote de 16,

pois apresenta um excelente equilíbrio para uma boa generalização e utilização de recursos (AWAIS et al., 2025). O tamanho da imagem de entrada padrão do YOLOv12 é 640x640, assim, utilizou-se este tamanho de imagem como entrada para o modelo (KUMAR, 2025). A base de dados criada ao longo deste trabalho tem algumas características diferentes do que as bases de dados convencionais usadas para treinar os modelos YOLO. Para começar, estamos usando imagens de radar de abertura sintética, estas imagens no final do pre-processamento, são disponibilizadas em escala de cinza. Para se adaptar ao modelo YOLO, triplicamos este canal, então no final teremos três canais cinzas. Neste sentido, o primeiro ajuste de hiperparâmetros tem a ver com a aumentação de dados. Basicamente, o aumento de dados copia dados da base de dados já existente e, através de diversas técnicas, essas cópias são transformadas para aumentar a diversidade e a quantidade de dados. O intuito de aumentar a diversidade e quantidade dos dados é a melhora da otimização e a generalização de modelos de modelos, ou seja, pode reduzir o sobreajuste (*overfitting*) e melhorar a robustez do modelo.

Nesse contexto, foram desativadas as transformações cromáticas (matiz, saturação), por se tratarem de dados monocromáticos SAR, porém manteve-se a mudança de brilho (valor). Se desativaram também as funções de inversão de canais. Ainda, mantiveram-se rotações de 10 graus, translações de até 10%, escalonamento de 0.5 (reduzir ou aumentar o tamanho), espelhamento vertical e de esquerda à direita de 0.5 (aplicado ao 50% das imagens de treinamento), a criação de mosaicos, de modo a introduzir variabilidade espacial (ULTRALYTICS, 2025).

Após finalização do treinamento, dois modelos são salvos: *best.pt* e *last.pt*. Estes são os pesos que representam os melhores e os últimos resultados do treinamento. Para todos os testes a serem feitos na seção seguinte, usaremos o arquivo de melhores pesos *best.pt*, que é escolhido mediante a formula *fitness*, conforme a equação a seguir:

$$\text{fitness}(x) = 0,0 \cdot P + 0,0 \cdot R + 0,1 \cdot \text{mAP}_{0,50} + 0,9 \cdot \text{mAP}_{0,50:0,95} \quad (5.7)$$

Sendo P, R, e mAP métricas de avaliação explicadas a continuação.

5.6 Métricas de avaliação

Para avaliar de forma abrangente o desempenho do modelo YOLOv12 na tarefa proposta neste trabalho, adota-se diversas métricas comumente utilizadas para detecção de objetos. Estas incluem Precisão (Precision), Revocação (Recall), F1-Score.

5.6.1 Precisão

A Precisão (P) representa a razão entre os verdadeiros positivos e o número total de amostras classificadas como positivas. Ela indica quão corretas são as previsões positivas e é expressa pela Equação 5.8:

$$\text{Precisão} = \frac{TP}{TP + FP} \quad (5.8)$$

onde TP refere-se ao número de verdadeiros positivos, FP indica o número de falsos positivos. Além disso, temos FN , que indica falsos negativos e por fim TN , sendo verdadeiros negativos.

5.6.2 Taxa de Revocação (*Recall*)

A Taxa de Revocação (R), também chamada de "sensibilidade", avalia a capacidade do modelo de detectar todos os casos positivos reais. É definida como a razão entre o número de verdadeiros positivos e o número total de instâncias positivas reais, conforme a Equação 5.9:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5.9)$$

5.6.3 *F1-Score*

O *F1-Score* é a média harmônica entre Precisão (P) e Revocação (R), fornecendo uma métrica única que equilibra tanto a correção quanto a completude. Ele é definido conforme a Equação 5.10:

$$F_1 \text{ score} = 2 \times \frac{P \times R}{P + R} \quad (5.10)$$

Um valor mais alto de *F1-Score* reflete um melhor desempenho do modelo, equilibrando a precisão das predições e a capacidade de detectar alvos relevantes.

5.6.4 Precisão Média

Além do *F1 score*, a precisão média pode ser obtida combinando precisão e revocação. Ao variar o limiar de confiança, obtêm-se múltiplos pares (Precisão, Revocação), formando a curva precisão–revocação (P–R). A área sob essa curva é definida como a *Precisão Média* (AP):

$$\text{AP} = \int_0^1 p(r) dr, \quad (5.11)$$

em que $p(r)$ denota a precisão como função da revocação r . Na prática, a integral é aproximada usando um conjunto finito de níveis de revocação e os respectivos valores de precisão.

5.6.5 mAP₅₀

A precisão média-média em um limiar de IoU de 0,50, denotada por mAP₅₀, é calculada como a média da AP sobre todas as categorias:

$$\text{mAP}_{50} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i^{\text{IoU}=0.50} \quad (5.12)$$

onde N é o número de classes de objetos.

5.6.6 mAP_{50:95}

Seguindo o protocolo de avaliação do COCO, mAP_{50:95} é obtida ao se proferir a média da AP em dez limiares de IoU, de 0,50 a 0,95, com passo de 0,05:

$$\text{mAP}_{50:95} = \frac{1}{10N} \sum_{j=1}^{10} \sum_{i=1}^N \text{AP}_i^{\text{IoU}=0.45+0.05j}. \quad (5.13)$$

Essa métrica fornece uma avaliação mais rigorosa e abrangente do desempenho do detector sob diferentes condições de sobreposição (IoU).

5.7 Estimativa de atributos geométricos

A etapa de estimativa dos atributos geométricos das embarcações representa um passo importante para este trabalho. Essa etapa foi desenvolvida com base na metodologia descrita em (GREIDANUS et al., 2017), a qual permite estimar, de forma aproximada, o comprimento e a largura de cada alvo detectado.

Inicialmente, utilizam-se as caixas delimitadoras das embarcações obtidas por meio do modelo YOLOv12 para recortar a porção correspondente da imagem SAR em formato GeoTIFF. Em seguida, aplica-se o algoritmo de Otsu para realizar a binarização do recorte, procedimento também empregado durante a criação da base de dados (ver Seção 5.3).

Cada píxel do recorte binarizado é associado às suas coordenadas $\mathbf{x}$ e $\mathbf{y}$, permitindo representar a embarcação como um conjunto de pontos. Sobre esse conjunto, aplica-se o método dos mínimos quadrados para ajustar uma reta que representa o eixo principal do alvo. A partir dessa reta, o comprimento é calculado em metros considerando a distância entre os pontos extremos (mínimo e máximo) ao longo do eixo principal. Por fim, traça-se

uma reta perpendicular a esse eixo, denominada eixo secundário, utilizada para estimar a largura da embarcação em análise. Ese processo foi feito por meio de um código em *Python*.

5.8 Extração de pontos vetoriais

Os pontos vetoriais finais representam a posição geográfica das embarcações detectadas. Um centróide é gerado para cada caixa delimitadora, e em seguida as coordenadas em pixels desse ponto são convertidas em coordenadas geográficas no sistema *WGS84*. Cada ponto contém informações como classe, confiança, data e hora de início da aquisição (UTC), data e hora de término da aquisição (UTC), coordenadas (latitude e longitude), comprimento e largura. Essas informações são armazenadas em um *GeoDataFrame* da biblioteca *GeoPandas*. Por fim, os pontos com suas informações geográficas são exportados para o formato *ESRI Shapefile*, permitindo sua visualização e análise em softwares SIG, como o QGIS. Ese processo foi feito por meio de um código em *Python*.

6 Resultados e discussões

6.1 Resultados de métricas de avaliação do modelo

Nesta seção, serão apresentadas tabelas e gráficos representando os resultados do treinamento e avaliações no conjunto de validação. Estes gráficos incluem gráficos de precisão, revocação, $F1-Score$ em função das épocas, gráficos de precisão, revocação, $F1-Score$, mAP50, mAP50:95 em função da confiança, gráficos de precisão média (Precisão versus revocação) e matrizes de confusão. Cada análise é feita nos três tamanhos avaliados (**n**, **s**, **m**).

6.1.1 Métricas de avaliação versus épocas de treinamento

Os gráficos de métricas de avaliação versus épocas de treinamento mostram como o modelo evolui e estabiliza (converge) durante cada época de treinamento.

6.1.1.1 Gráficos Precisão vs épocas para os tamanho **n**, **s** e **m**

A Figura 45, apresenta as métricas de precisão em função das 300 épocas de treinamento. A precisão, conforme a Equação 5.8, representa a proporção de predições positivas corretas em relação ao total de predições positivas realizadas. Observa-se um progresso consistente da métrica ao longo das épocas, atingindo valores superiores a 93% para o modelo de tamanho **n** e acima de 95% para os tamanhos **s** e **m**. Nas três variações, nota-se uma melhora acentuada a partir da época 20, seguida de pequenas oscilações no decorrer do processo de treinamento, o que evidencia a capacidade aprimorada do modelo em reduzir falsos positivos.

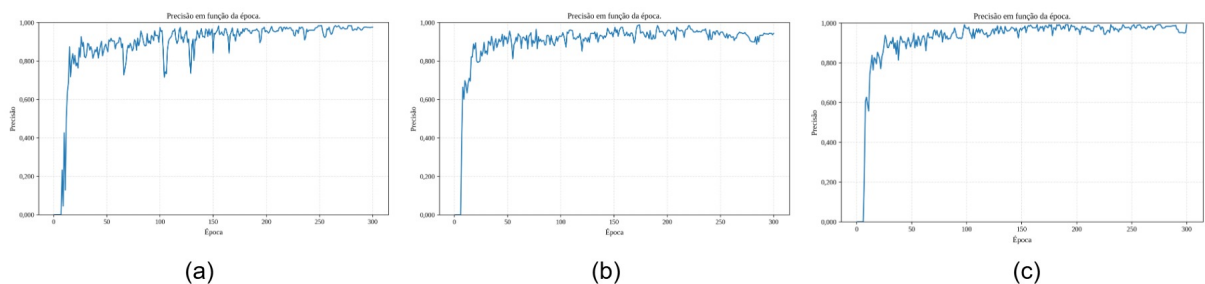


Figura 45 – Gráficos de precisão em função das épocas. (a) precisão em função das épocas para a variante **n**, (b) precisão em função das épocas para a variante **s**, (c) precisão em função das épocas para a variante **m**.

6.1.1.2 Gráficos Revocação vs épocas para os tamanho **n**, **s** e **m**

A Figura 46 apresenta as métricas de revocação em função das 300 épocas de treinamento. A revocação tem a função de medir o quão bem o modelo identifica todos os casos relevantes sem deixar de detectar nenhum (isto é, minimizando os falsos negativos). A métrica de revocação atinge mais de 95% para a variante **n**, e mais de 90% para as variantes **s** e **m**. Ao igual que nos gráficos de precisão, nota-se uma melhora nos valores de revocação a partir da época 20, aumentando minimamente até estabilizar na época 100 aproximadamente, o que demonstra que o modelo alcançou sensibilidade de detecção efetiva, isto é, habilidade para detectar a maioria de verdadeiros positivos.

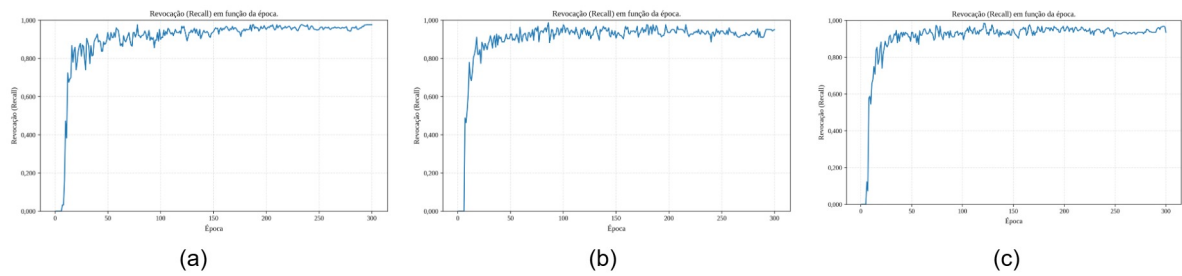


Figura 46 – Gráficos de revocação em função das épocas. (a) revocação em função das épocas para a variante **n**, (b) revocação em função das épocas para a variante **s**, (c) revocação em função das épocas para a variante **m**.

6.1.1.3 Gráficos *F1-Score* vs épocas para os tamanho **s**, **n** e **m**

A Figura 47 apresenta o *F1-score* em função das épocas, que representa a média harmônica entre precisão e revocação. Essa métrica mostra uma melhora rápida, por volta da época 15, e converge para aproximadamente 94%, para as três variantes, o que comprova que o modelo atinge um equilíbrio entre desempenho de precisão e revocação, isto é, detecções de embarcações altamente precisas e sensíveis.

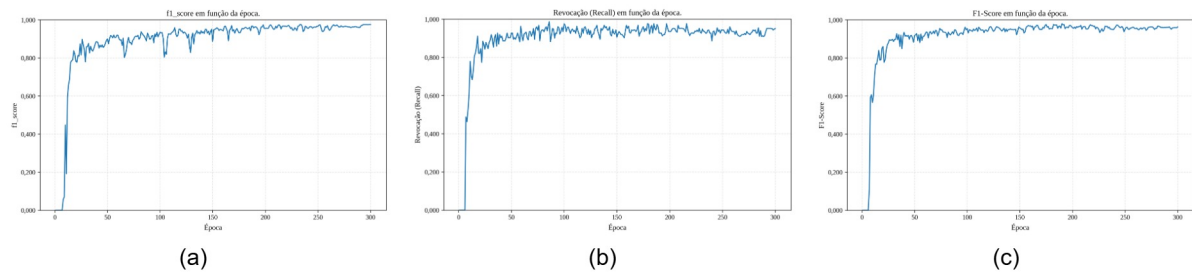


Figura 47 – Gráficos de *F1-score* em função das épocas. (a) *F1-score* em função das épocas para a variante **n**, (b) *F1-score* em função das épocas para a variante **s**, (c) *F1-score* em função das épocas para a variante **m**.

6.1.1.4 Gráficos mAP0,50 vs épocas para os tamanhos **n**, **s** e **m**

O gráfico na figura 48 apresenta o mAP0,50 dos três modelos em 0,50 de IoU. Os três modelos mostram melhora contínua ao longo de todo o treinamento, com uma particularidade no gráfico da variante **n**, que apresenta flutuações ao longo da curva. Isto é explicado pelo fato que o otimizador (no nosso caso, o **AdamW**) tenta ajustar os pesos da rede em direção a um mínimo global da função de perda. Isto é feito a cada lote usando gradientes locais, o que pode gerar pequenos aumentos ou diminuições dos valores da perda, consequentemente, das métricas derivadas dela (Precisão, Revocação, mAP0,50, etc.)

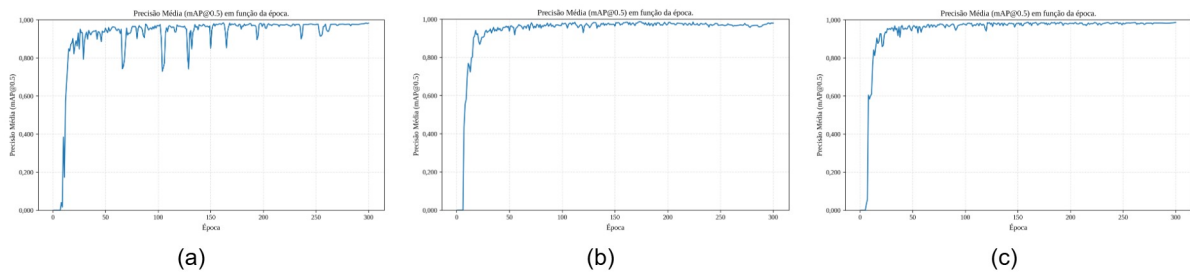


Figura 48 – Gráficos de mAP0,50 em função das épocas. (a) mAP0,50 em função das épocas para a variante **n**, (b) mAP0,50 em função das épocas para a variante **s**, (c) mAP0,50 em função das épocas para a variante **m**.

6.1.1.5 Gráficos mAP0,50:0,95 vs épocas para os tamanhos **n**, **s** e **m**

Esta métrica é O mAP calculado em vários limites de IoU, variando de 0,50 a 0,95. Ele fornece uma visão abrangente do desempenho do modelo em diferentes níveis de dificuldade de detecção. É normal que esta métrica apresente valores menores que o mAP0,50 pois é uma métrica mais rigorosa, que envolve vários limiares. Para os nossos modelos, a curva aumenta rapidamente até a época 25 aproximadamente, aonde continua aumentando porém mais sutilmente, até convergir por volta de 0,60. A variante **m** apresenta um valor maior em comparação aos modelos **n** e **s**.

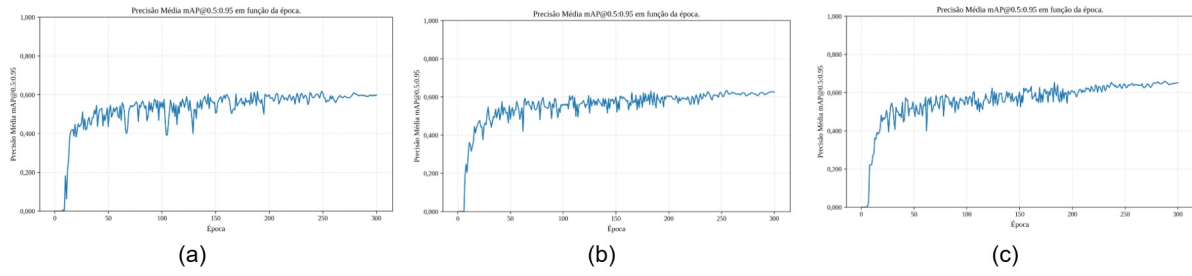


Figura 49 – Gráficos de $mAP_{0,50:0,95}$ em função das épocas. **(a)** $mAP_{0,50:0,95}$ em função das épocas para a variante **n**, **(b)** $mAP_{0,50:0,95}$ em função das épocas para a variante **s**, **(c)** $mAP_{0,50:0,95}$ em função das épocas para a variante **m**.

6.1.2 Gráficos Precisão vs confiança para os tamanhos **n**, **s** e **m**

A curva de precisão em função da confiança ilustra como a precisão varia conforme diferentes níveis de confiança. Idealmente, busca-se manter alta precisão em toda a faixa de valores de confiança. A tendência crescente da curva indica que, em níveis baixos de confiança (por exemplo, 0,00), o modelo permite muitas detecções (entre as quais diversas são Falsos Positivos) o que reduz a precisão.

De forma intuitiva, ao aumentar o limiar de confiança, as detecções menos seguras são progressivamente filtradas, restando apenas aquelas em que o modelo tem maior certeza. Entretanto, quando a confiança é excessivamente elevada, pode não haver detecções que alcancem valores tão altos, o que explica a queda observada na curva a partir de aproximadamente 0,80 nos três modelos avaliados.

Observa-se também que, em cada gráfico da Figura 50, é exibida a legenda indicando o *limiar ótimo*, definido com base no valor máximo do *F1-Score*. Esse limiar representa o ponto ideal, pelo menos quando avaliado no conjunto de validação, para realizar inferência, correspondendo aos valores de 0,50, 0,30 e 0,40 para as variantes **s**, **n** e **m**, respectivamente.

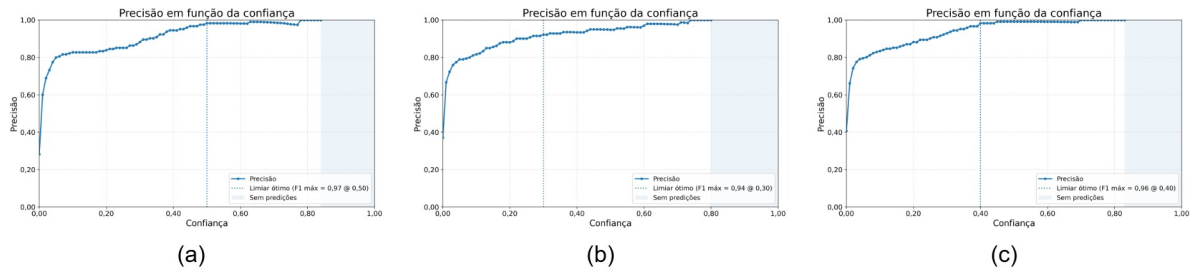


Figura 50 – Gráfico de precisão versus confiança. **(a)** precisão em função da confiança para a variante **n**, **(b)** precisão em função da confiança para a variante **s**, **(c)** precisão em função da confiança para a variante **m**.

6.1.3 Gráficos Revocação vs confiança para os tamanhos **n**, **s** e **m**

Esta curva mostra a revocação em diversos limiares de confiança. O ideal é manter valores altos de revocação em toda a faixa de confiança.

Nos três modelos, a revocação começa próxima de 1,00 em níveis baixos de confiança, o que significa que o modelo está detectando praticamente todos os alvos, ainda que algumas dessas detecções sejam pouco seguras.

No caso do modelo **s**, observa-se que a curva apresenta uma queda mais rápida à medida que a confiança aumenta. Isso ocorre porque esse modelo tende a atribuir valores de confiança mais baixos às suas detecções verdadeiras. Assim, quando o limiar de confiança é elevado, muitas dessas detecções válidas acabam sendo descartadas, reduzindo rapidamente a revocação.

Já os modelos **n** e **m** mantêm a revocação alta por um intervalo maior de confiança, indicando que suas detecções verdadeiras costumam ter valores de confiança mais altos e consistentes. Esses modelos são, portanto, mais seguros em suas predições.

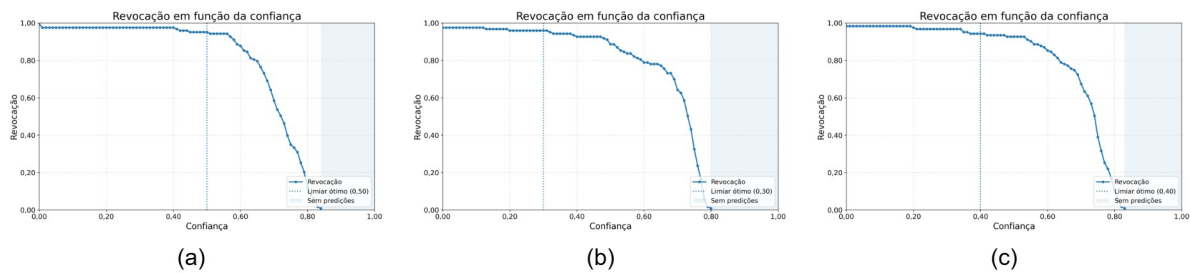


Figura 51 – Gráfico de revocação versus confiança. **(a)** revocação em função da confiança para a variante **n**, **(b)** revocação em função da confiança para a variante **s**, **(c)** revocação em função da confiança para a variante **m**.

6.1.4 Gráficos $F1$ -Score vs confiança para os tamanhos **n**, **s** e **m**

Esta curva mostra o valor do $F1$ -score (média harmônica entre precisão e revocação) em diferentes limiares de confiança. Um pico mais alto indica um desempenho melhor do modelo.

Particularmente, esta métrica é relevante pois, no contexto da detecção de embarcações, representa o equilíbrio entre identificar corretamente os alvos reais e evitar omissões. O valor máximo da curva indica o melhor desempenho de cada modelo em função do limiar de confiança adotado.

Os modelos **n** e **m** apresentaram os maiores valores, com 0,97 e 0,96, respectivamente, em limiares de 0,50 e 0,40. Nesse contexto, o modelo **m** oferece o melhor compromisso entre precisão e revocação, com estabilidade e robustez superiores; Já o modelo **s** obteve um valor de 0,94 em um limiar de 0,30, indicando que, mesmo em níveis mais baixos de confiança, o modelo já alcança um bom equilíbrio entre precisão e revocação. A partir desse ponto, contudo, observa-se uma redução progressiva do desempenho à medida que o limiar de confiança aumenta, evidenciando a sua maior sensibilidade e menor consistência nas detecções mais seguras.

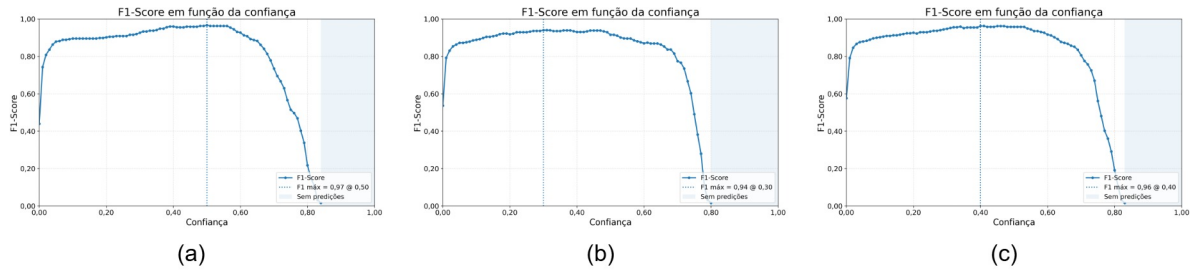


Figura 52 – Gráfico de $F1$ -Score versus confiança. (a) $F1$ -Score em função da confiança para a variante **n**, (b) $F1$ -Score em função da confiança para a variante **s**, (c) $F1$ -Score em função da confiança para a variante **m**.

6.1.5 Gráficos Precisão média (AP50) para os tamanhos **n**, **s** e **m**

Esta curva ilustra o equilíbrio entre precisão e revocação para diferentes limiares. Um modelo que se aproxima mais do canto superior direito é considerado melhor. A área sob a curva é denominada Precisão Média (*Average Precision*).

Nessa métrica, observa-se que os três modelos apresentaram valores excelentes de desempenho, com 0,98 para os modelos **n** e **m**, e 0,96 para o modelo **s**. Esses resultados reforçam que as três variantes mantêm um ótimo equilíbrio entre precisão e revocação, demonstrando elevada capacidade de detecção de embarcações e baixo índice de falsos positivos.

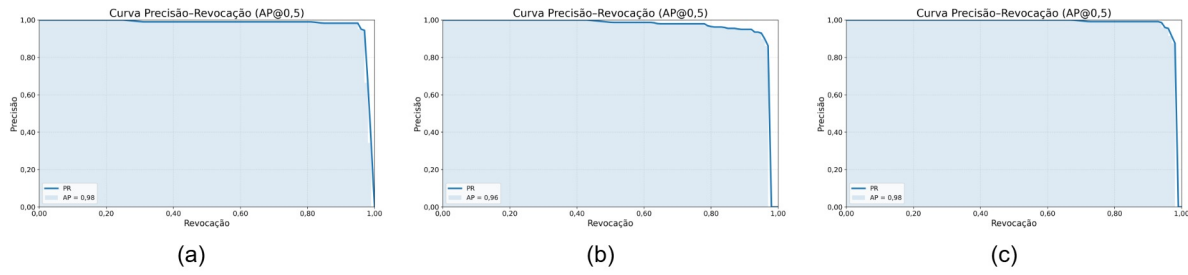


Figura 53 – Gráfico de curva precisão-revoação. **(a)** curva precisão-revoação para a variante **n**, **(b)** curva precisão-revoação para a variante **s**, **(c)** curva precisão-revoação para a variante **m**.

6.1.6 Matrizes de confusão

A matriz de confusão é uma ferramenta usada para avaliar o desempenho de um algoritmo de classificação. Uma matriz de confusão fornece uma análise detalhada de como um modelo está se comportando e mostra não apenas quando o modelo está certo, mas também como ele está errado, revelando onde reside a "confusão". Esta matriz é gerada a partir de valores fixos de confiança (0,25) e de IoU (0,45), descartando todas as detecções que não atendem a esses limiares. Ao analisar as matrizes e considerando o número total de embarcações no conjunto de validação (123), observa-se que o modelo de tamanho **n**, ilustrado na Figura 54 (a), apresentou a maior quantidade de verdadeiros positivos (TP) em relação aos demais modelos. Por outro lado, conforme mostrado na Figura 54 (c), o modelo de tamanho **m** exibiu o menor número de falsos positivos, isto é, regiões de fundo incorretamente detectadas como embarcações.

6.1.6.1 Tamanho n, s e m

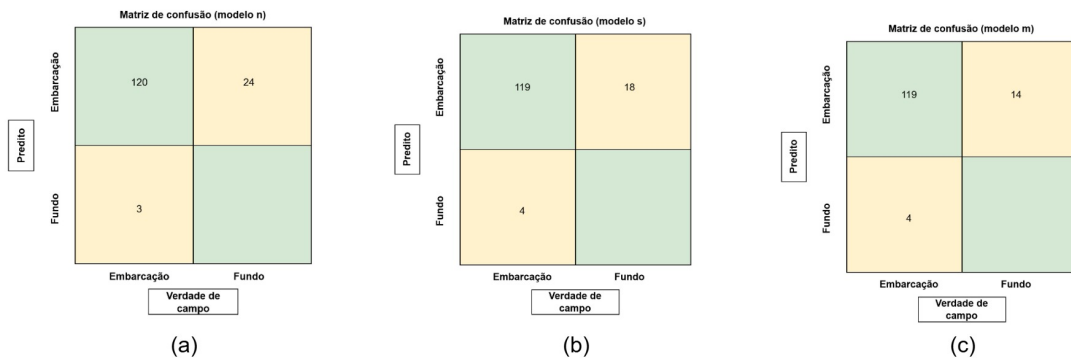


Figura 54 – Matriz de confusão gerada através da inferência no teste de validação. **(a)** Matriz de confusão para a variante **n**, **(b)** Matriz de confusão para a variante **s**, **(c)** Matriz de confusão para a variante **m**.

6.1.7 Desempenho geral por métrica

Na Tabela 9, observam-se os resultados das principais métricas de desempenho para cada tamanho do modelo YOLOv12 quando avaliados no conjunto de validação (72 imagens, 123 embarcações). Esta tabela representa os valores médios das métricas de avaliação dos modelos com melhores pesos conforme explicado na secção 5.5, usando a equação 5.7. O tamanho **n** apresentou excelente revocação (94,94 %), o que indica alta capacidade de detectar praticamente todas as embarcações presentes nas imagens. Além disso, obteve um valor de precisão de 98,32 %, demonstrando bom desempenho na hora de detectar barcos reais.

O tamanho **s** destacou-se pela alta precisão e elevado mAP 50, o que revela que suas detecções são realizadas com grande confiabilidade, embora de forma mais seletiva. Por sua vez, o tamanho **m** apresentou a maior precisão dos três modelos, atingindo um 98,59 %, além disso, obteve os maiores valores de *F1-Score*, mAP0,50 e mAP0,50:0,95. O valor alto de *F1-Score* resulta em um ótimo equilíbrio entre precisão e revocação. Esse equilíbrio indica detecções simultaneamente confiáveis e abrangentes. Por fim, o modelo de tamanho **m** apresentou o melhor desempenho geral no conjunto de validação, embora muito próximo ao modelo **n** e não tão distante do modelo **s**. Tal resultado não garante desempenho superior em cenas inéditas; entretanto, constitui forte indicativo para sua seleção em aplicações futuras.

Tamanho	Precisão (%)	Revocação (%)	F1 (%)	mAP 50 (%)	mAP 50:95 (%)
n	98,32	94,94	94,3	96,60	61,80
s	92,09	94,61	93,33	96,71	63,39
m	98,59	93,50	95,97	98,29	65,76

Tabela 9 – Comparação de métricas de avaliação dos dados de validação entre modelos YOLOv12. Em vermelho, os valores que atingiram os máximos entre os três modelos.

6.2 Teste de generalização: inferência no conjunto de validação

Em aplicações de aprendizado de máquina, aprendizado profundo e visão computacional, a inferência é o processo de extrair significado de dados visuais. Isto faz parte dos testes do modelo treinado. A finalidade dos testes é avaliar o modelo em relação à precisão e garantir que o modelo faça previsões de acordo com a finalidade do projeto. Nesse sentido, usamos, como teste inicial, o conjunto de teste da nossa base de dados para realizar inferência. Resultados qualitativos, com detecções corretas podem ser consultados na figura 55. Por outra parte, observamos, na figura 56, alguns erros de detecções por tamanho.

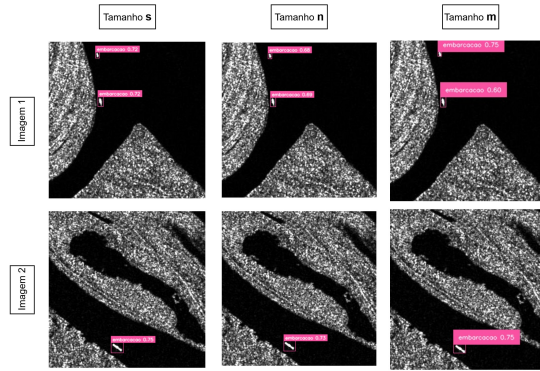


Figura 55 – Inferência no conjunto de validação usando os três modelos treinados. Nas imagens mostradas, os três tamanhos obtiveram detecções coerentes à verdade de campo.

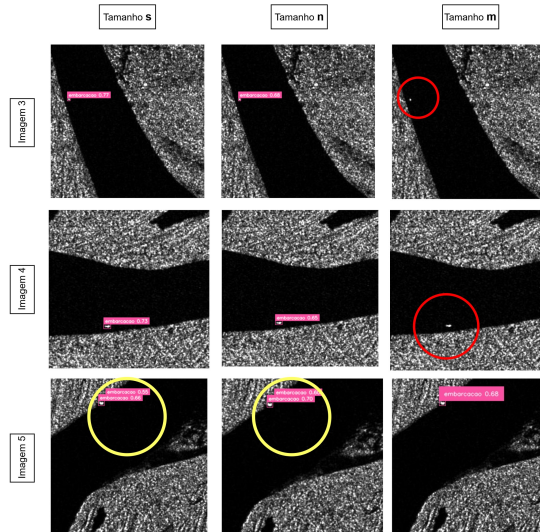


Figura 56 – Falhas na detecção de objetos na inferência no conjunto de validação usando os três modelos treinados.

Podemos ver que as caixas delimitam relativamente bem as embarcações. Observar que uma etiqueta contendo o nome da classe daquele objeto e um número entre 0 e 1 acompanha cada caixa delimitadora. Esta probabilidade, para o caso de inferência no conjunto de teste, onde temos verdade de campo, é calculado usando a equação 4.4.

Como observado na Figura 56, é possível identificar algumas falhas de detecção. Na imagem 3, para o tamanho **m**, observa-se um falso negativo, o que também ocorre na imagem 4 (primeira e segunda linhas da terceira coluna). Essa característica está associada ao comportamento mais conservador do modelo de tamanho **m**, conforme indicado pelo seu valor de revocação na Tabela 9.

Por outro lado, os tamanhos **n** e **s** apresentaram falsos positivos na imagem 5 (terceira linha, colunas 1 e 2), destacados com círculos amarelos. Esse resultado está

relacionado ao valor ligeiramente menor de precisão em comparação com o tamanho **m**, conforme mostrado na Tabela 9. É interessante notar que os valores de precisão para **n** e **m** são bastante próximos, o que indica que tais falsos positivos no modelo **n** não são recorrentes quando se considera a média das métricas.

6.3 Testes em cenas Sentinel 1

As imagens do *Sentinel-1* são disponibilizadas em formato *float32*, o que permite armazenar valores decimais referentes à intensidade do retroespalhamento. Após o pré-processamento no *ESA SNAP*, os produtos são exportados em formato *GeoTIFF*, preservando essa profundidade radiométrica de 32 bits. No entanto, o modelo YOLO requer imagens com profundidade de 8 bits para a etapa de inferência. Para contornar essa limitação, as imagens raster são normalizadas para o intervalo de 0 a 255 (equivalente a 8 bits), garantindo compatibilidade com a rede neural sem alterar as relações espaciais originais.

Após a normalização, a imagem é subdividida em blocos (tiles) de 320×320 pixels, sobrepostos em 64 pixels, de modo a evitar a perda ou o corte de embarcações localizadas nas bordas dos recortes. A inferência é executada em cada recorte individualmente, aplicando-se um limiar de confiança de 0,25, ou seja, detecções com probabilidade inferior a 25% são descartadas, e um limiar de sobreposição (*IoU*) de 50%. As detecções resultantes de todos os recortes são então combinadas e passam por uma supressão não máxima global (*Non-Maximum Suppression*), eliminando duplicidades e gerando uma imagem final anotada que representa todas as detecções consolidadas na cena original.

Como produto final, o processo gera uma imagem anotada no formato *.jpg*, contendo as caixas delimitadoras e rótulos das detecções. Este processo foi feito por meio de um código *Python*.

Testes de generalização foram realizados para duas imagens *Sentinel-1*, correspondentes à região do Rio Negro, em Manaus (Figura 57), e à região da Ilha de Santana, em Macapá, que forma parte do chamado Delta do Amazonas (Figura 59).

É importante mencionar que, a partir daqui, todos os modelos serão testados de duas formas: a primeira, no valor de confiança que retorna o *F1-Score* máximo para cada tamanho. A segunda, valor fixo de confiança de 0,25 para todos os tamanhos. Os resultados das inferências, tanto qualitativas quanto quantitativas são apresentados a seguir.

6.3.1 Manaus

A região da passagem do Rio Negro por Manaus foi escolhida devido à alta movimentação do tráfego marítimo na área, o que permite avaliar o desempenho dos modelos treinados em um ambiente com grande diversidade de tamanhos de embarcações.

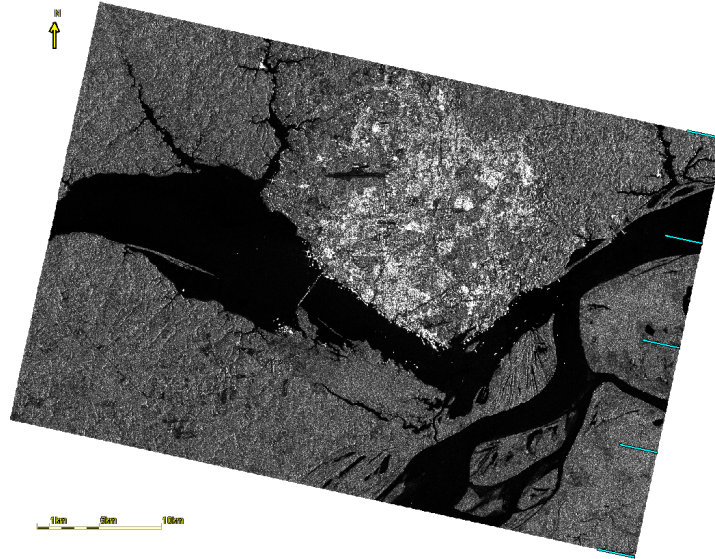


Figura 57 – Cena para realização de teste de generalização na área do Rio Negro, Manaus, com imagem Sentinel-1.

Como observado na Figura 58a, o modelo **s**, com limiar de confiança de 0,25, consegue detectar embarcações de larguras e comprimentos variados. Em contrapartida, na Figura 58b, observa-se a presença de falsos positivos, em sua maioria correspondentes a porções de terra que se assemelham a pequenas embarcações.

Essa limitação não é exclusiva do tamanho **s**, pois, conforme apresentado na Tabela 12, os tamanhos **n** e **m** também apresentam falsos positivos, totalizando 22, 27 e 25 para os modelos **n**, **s** e **m**, respectivamente.

Por outro lado, observando a Tabela 10, nota-se que a quantidade de falsos positivos diminui. Isso ocorre porque o limiar de confiança foi ajustado para os valores que maximizam o *F1-Score* de cada variante. No entanto, essa modificação também reduz o número de verdadeiros positivos e aumenta a taxa de falsos negativos.

Essa discrepância entre os resultados com diferentes limiares de confiança está relacionada à complexidade da cena analisada, que inclui diversas construções humanas, como *decks* e pontes, dificultando a distinção entre embarcações reais e estruturas artificiais.

Em relação às métricas restantes, observa-se, tanto na Tabela 11 quanto na Tabela 13, que o modelo **s** apresentou os melhores resultados para as métricas *F1-Score*, mAP 0,50 e mAP 0,50–0,95, com exceção do *F1-Score* na Tabela 13, em que foi superado pelo

modelo **n** por aproximadamente 1%.

É importante destacar que os valores de $mAP@0.50-0.95$ são consideravelmente menores em comparação aos obtidos no conjunto de validação. Esse resultado indica que o modelo apresenta dificuldades em realizar uma localização precisa dos objetos quando avaliados sob limiares de IoU mais restritivos. Em outras palavras, à medida que se aumenta o percentual de interseção sobre união exigido para considerar uma detecção correta, o modelo tende a apresentar maior imprecisão no ajuste das caixas delimitadoras em relação às caixas da verdade de campo.

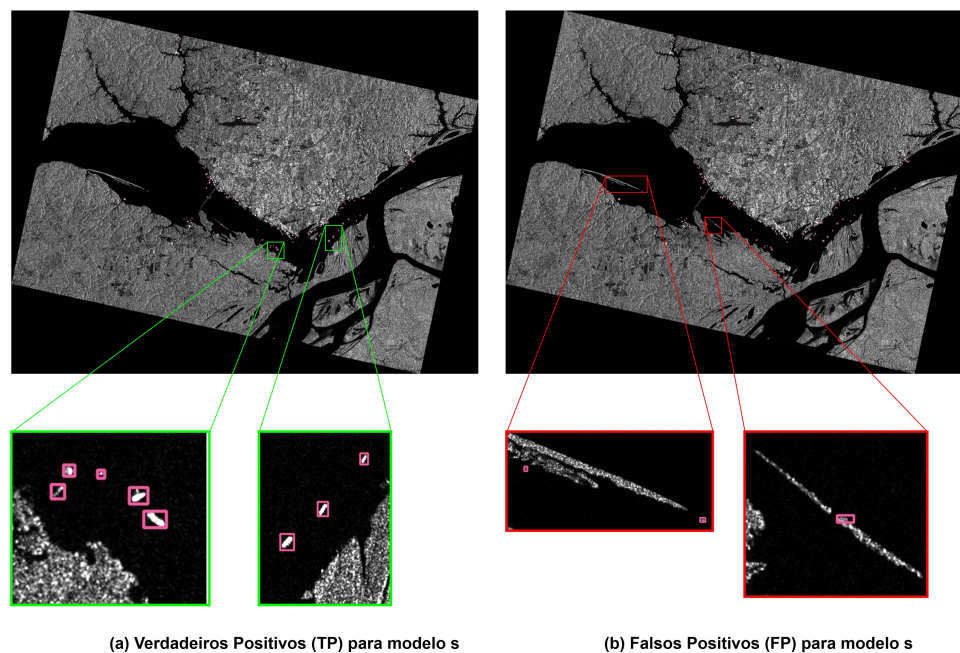


Figura 58 – Teste de generalização usando o modelo **s** com limiar de confiança de 0,25, na área do Rio Negro, Manuas com imagem Sentinel-1. Em (a) observa-se no detalhe as embarcações marcadas como verdadeiros positivos (TP). Em (b), alvos marcados erroneamente como embarcações, representando falsos positivos.

Tamanho	Verdade de campo	TP	FN	FP	Precisão (%)	Revocação (%)
n	83	28	55	4	87,50	33,73
s	83	54	29	21	72,00	65,06
m	83	45	38	13	77,59	54,22

Tabela 10 – Comparação de métricas de avaliação Precisão e Revocação entre modelos YOLOv12, com confiança 0,50 para o tamanho **n**, 0,30 para o tamanho **s** e 0,40 tamanho **m**. Cena extraída de uma área do Rio Negro, Manaus

Tamanho	F1 (%)	mAP 0,50 (%)	mAP 0,50-0,95 (%)
n	48,70	23,62	08,26
s	68,35	43,77	15,14
m	63,83	34,60	11,42

Tabela 11 – Comparação das métricas de avaliação *F1-Score*, mAP 0,50 e mAP 0,50-0,95 entre modelos YOLOv12, com confiança 0,50 para o tamanho **n**, 0,30 para o tamanho **s** e 0,40 tamanho **m**. Cena extraída de uma área do Rio Negro, Manaus

Tamanho	Verdade de campo	TP	FN	FP	Precisão (%)	Revocação (%)
n	83	56	27	22	71,79	67,47
s	83	57	26	27	67,86	68,67
m	83	51	32	25	67,11	61,45

Tabela 12 – Comparação de métricas de avaliação Precisão e Revocação entre modelos YOLOv12 com limiar de confiança fixo de 0,25 para todos os tamanhos. Cena extraída de uma área do Rio Negro, Manaus

Tamanho	F1 (%)	mAP 0,50 (%)	mAP 0,50-0,95 (%)
n	69,57	40,02	13,05
s	68,26	45,36	15,46
m	64,15	38,43	12,65

Tabela 13 – Comparação das métricas de avaliação *F1-Score*, mAP 0,50 e mAP 0,50-0,95 entre modelos YOLOv12, limiar de confiança fixo de 0,25 para todos os tamanhos. Cena extraída de uma área do Rio Negro, Manaus

6.3.2 Delta do Amazonas

Os modelos também foram testados em uma cena que corresponde a uma parte do Delta do Amazonas que passa pelo município de Macapá, próximo à Ilha de Santana, no estado de Amapá, como se observa na figura 59.

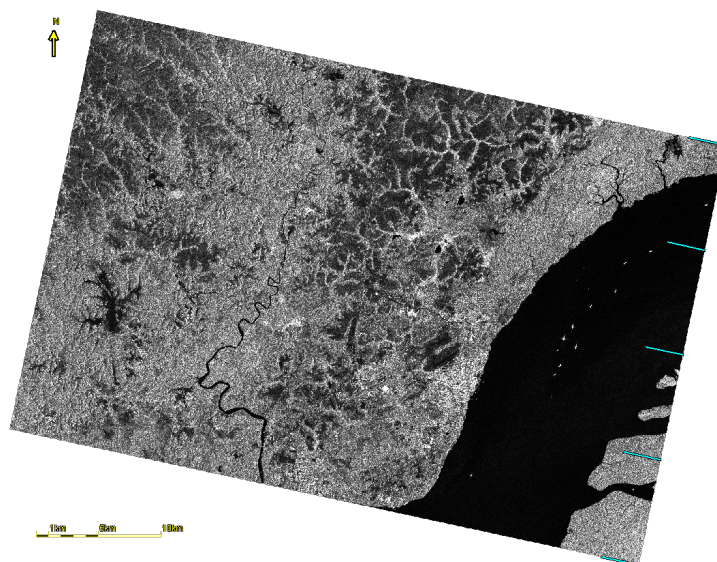


Figura 59 – Cena para realização de teste de generalização na área do Delta do Amazonas com imagem Sentinel-1.

Embora esta não seja uma cena complexa em termos de presença de construções artificiais, como ocorre na região de Manaus, os valores das métricas para a cena do Delta do Amazonas apresentaram melhora em relação à cena anterior, especialmente na Revocação.

Como mostrado na Figura 60a, o modelo **n**, com um limiar de confiança de 0,25, conseguiu detectar a maioria dos verdadeiros positivos, ainda que tenha identificado algumas pequenas ilhas como embarcações. Este é um problema recorrente em cenas *Sentinel-1*, particularmente em regiões fluviais próximas às margens, o que dificilmente ocorreria em áreas de mar aberto.

Ainda assim, conforme apresentado na Tabela 14, ao utilizar o limiar que maximiza o *F1-Score* para cada tamanho, observa-se uma redução na quantidade de falsos positivos em relação à Tabela 16.

Quanto aos verdadeiros positivos, especificamente nesta cena há uma embarcação cuja detecção é particularmente problemática. Ao aumentar os limiares de confiança, nenhum dos tamanhos conseguiu detectá-la; entretanto, para o limiar de 0,25, os modelos **n** e **m** conseguiram identificá-la, demonstrando capacidade de detecção de embarcações pequenas em áreas relativamente remotas.

Isso pode ser explicado pela melhor sensibilidade desses modelos a alvos de menor

dimensão, característica observada também em outros experimentos de generalização.

Nas tabelas 15 e 17 observa-se, de novo, uma predominância do modelo **s** em relação a mAP. Enquanto o modelo **n** teve melhor desempenho na métrica *F1-Score*.

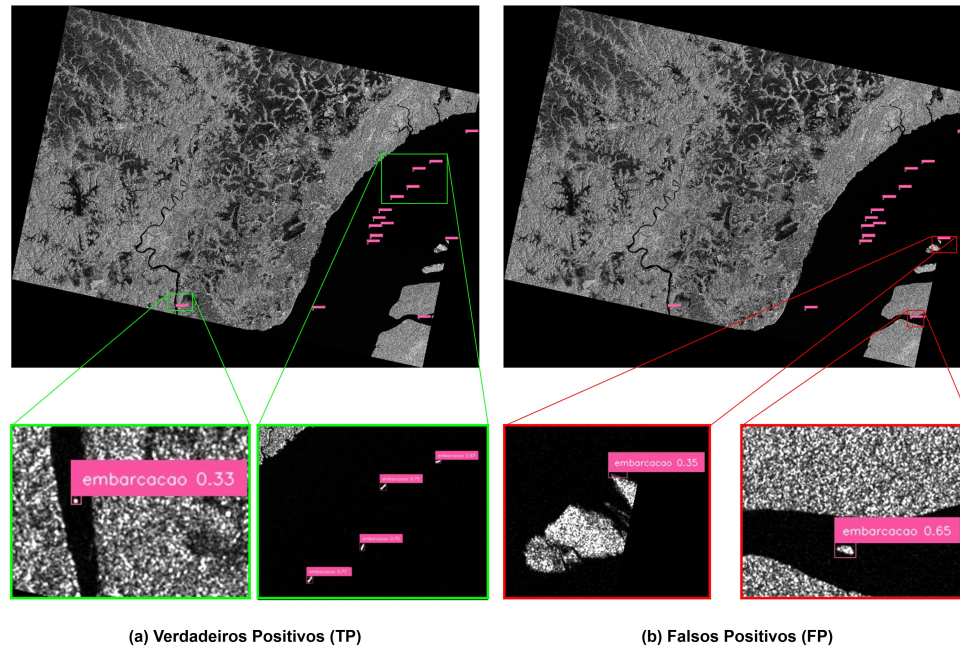


Figura 60 – Teste de generalização usando o tamanho **n**, na área do Delta do Amazonas com imagem Sentinel-1. Em (a) observa-se no detalhe as embarcações marcadas como verdadeiros positivos (TP). Em (b), ilhas marcadas como embarcações, representando falsos positivos.

Tamanho	Verdade de campo	TP	FN	FP	Precisão (%)	Revocação (%)
n	13	12	1	2	85,71	92,31
s	13	12	1	6	66,67	92,31
m	13	12	1	3	80,00	92,31

Tabela 14 – Comparação de métricas de avaliação Precisão e Revocação entre modelos YOLOv12, com confiança 0,50 para o tamanho **n**, 0,30 para o tamanho **s** e 0,40 tamanho **m**. Cena extraída do Delta do Amazonas

Tamanho	F1 (%)	mAP 0,50 (%)	mAP 0,50-0,95 (%)
n	88,89	82,26	47,92
s	77,72	87,85	49,43
m	85,71	79,44	46,81

Tabela 15 – Comparação das métricas de avaliação *F1-Score*, mAP 0,50 e mAP 0,50-0,95 entre modelos YOLOv12, com confiança 0,50 para o tamanho **n**, 0,30 para o tamanho **s** e 0,40 tamanho **m**. Cena extraída do Delta do Amazonas

Tamanho	Verdade de campo	TP	FN	FP	Precisão (%)	Revocação (%)
n	13	13	0	5	72,22	100,00
s	13	12	1	7	63,16	92,31
m	13	13	0	6	68,42	100,00

Tabela 16 – Comparação de métricas de avaliação Precisão e Revocação entre modelos YOLOv12 com limiar de confiança fixo de 0,25 para todos os tamanhos. Cena extraída do Delta do Amazonas

Tamanho	F1 (%)	mAP 0,50 (%)	mAP 0,50-0,95 (%)
n	83,87	82,26	47,92
s	75,00	87,85	49,43
m	81,25	79,44	46,81

Tabela 17 – Comparação das métricas de avaliação *F1-Score*, mAP 0,50 e mAP 0,50-0,95 entre modelos YOLOv12, limiar de confiança fixo de 0,25 para todos os tamanhos. Cena extraída do Delta do Amazonas.

A partir dos resultados quantitativos apresentados nas tabelas anteriores, procede-se à seleção do modelo que demonstrou o melhor desempenho nas duas cenas avaliadas, as quais representam de forma adequada tanto a geografia do Rio Amazonas quanto a dos demais rios da região Norte do Brasil. Entre as variantes analisadas, o modelo YOLOv12 **s** mostrou-se o mais apropriado para este trabalho, por apresentar o melhor equilíbrio entre capacidade de aprendizado e robustez frente às mudanças de cenário.

Como discutido em capítulos anteriores, o tamanho de cada variante do YOLOv12 está diretamente relacionado à quantidade de parâmetros treináveis da rede, que está relacionado com os valores de profundidade, largura e canais (ver tabela 8). Quanto maior o número de parâmetros, maior a capacidade do modelo de aprender padrões complexos, o que geralmente resulta em maior precisão (PyImageSearch, 2025). Entretanto, modelos muito grandes, como o tamanho **m**, tendem a apresentar maior propensão ao sobreajuste (*overfitting*), isto é, aprendem excessivamente as características do conjunto de treinamento. Esse comportamento leva a métricas elevadas no conjunto de validação, mas compromete o desempenho em imagens mais desafiadoras ou com distribuição estatística distinta, como é o caso das cenas de Manaus, onde as condições ambientais são mais complexas.

Por outro lado, o modelo **n**, apesar de ser mais leve e apresentar menor risco de sobreajuste, possui quase dez vezes menos parâmetros que a variante **m**, o que reduz sua capacidade de representar padrões mais sutis presentes nas imagens SAR. Consequentemente, sua precisão global tende a ser inferior.

O equilíbrio ideal é atingido pela variante **s**, que possui capacidade suficiente para aprender representações robustas, mas não é grande o bastante para memorizar o conjunto de treinamento. Isso se traduz em um desempenho mais estável e consistente nas imagens mais complexas. Dessa forma, o modelo **s** obteve os melhores resultados de generalização, especialmente na cena de Manaus, caracterizada pela presença de alvos ambíguos.

Isso não implica que as variantes **n** e **m** sejam inadequadas; ao contrário, todas apresentaram desempenho elevado na cena do Delta do Amazonas. Contudo, a cena de Manaus revelou diferenças significativas entre os modelos, desempenhando papel decisivo na escolha final. Diante desse contraste, os resultados quantitativos favorecem claramente a variante **s**, que, portanto, será adotada nas análises subsequentes.

6.4 Estimativa de atributos

A seguir, são apresentados os resultados dos cálculos dos atributos geométricos das embarcações detectadas pelos modelos. Esse procedimento foi realizado por meio da implementação de um script em *Python*, utilizando a biblioteca de visão computacional *OpenCV*. A partir das imagens binarizadas das embarcações, foram extraídas as coordenadas dos píxeis com valor igual a 1 (correspondentes à região branca da máscara binária). De posse dessa matriz de pontos, aplicou-se o método dos Mínimos Quadrados, por meio da função `cv2.fitLine()`, com o objetivo de projetar uma linha de melhor ajuste entre os pontos, utilizando a métrica de distância `DIST_L2`, que corresponde à distância euclidiana. Esse método é considerado o mais simples e eficiente para esse tipo de ajuste ([OpenCV Team, 2025](#)). Os resultados obtidos são apresentados na Figura 61.

Como mostrado na Figura 61d, o comprimento da embarcação foi estimado em 138 metros, enquanto a largura resultante foi de 25 metros. Essas dimensões são compatíveis com as embarcações que normalmente navegam na região de análise. Na Figura 61e, observa-se uma embarcação do tipo *graneleiro*. Esse tipo de navio apresenta comprimentos variando entre 100 e 300 metros, dependendo do modelo e da finalidade ([GUIDE, 2025](#)), o que corrobora as dimensões estimadas pelo algoritmo.

Uma embarcação de menor porte pode ser observada na Figura 62b. Estima-se que se trate de um iate, tipo de embarcação bastante comum na região, com dimensões compatíveis aos comprimentos detectados ([YACHTWORLD, 2025](#)). O comprimento e a largura estimados foram de aproximadamente 28 m. Embora o comprimento esteja de acordo com as dimensões típicas de um iate, a largura apresenta uma discrepância pelo formato quadrado do resultado da binarização.

Cabe destacar que a validação das dimensões estimadas das embarcações só seria possível por meio do Sistema de Identificação Automática (*AIS*). No entanto, não foi possível acessar dados históricos, pois essas informações são comercializadas por empresas privadas, devido ao seu alto valor no contexto marítimo. Contudo, são apresentados, na Figura 63, os histogramas do comprimento e da largura estimados.

A análise da cena de Manaus é mostrada nas Figuras 63a e 63b. Como pode ser observado na Figura 63a, a distribuição não é uniforme e apresenta *outliers*, que podem corresponder a falsos positivos (como *decks*, ilhas ou pontes) ou a superestimações resultantes do algoritmo de estimação de atributos. A Figura 63b mostra uma concentração de valores de largura entre 15 e 60 metros, o que reflete as dimensões típicas das embarcações que navegam pelo rio Negro, em Manaus. Ainda assim, valores superiores a 60 metros podem ser considerados *outliers*, pois é improvável que embarcações nessa região apresentem tais larguras.

As Figuras 63c e 63d apresentam os resultados obtidos para a cena correspon-

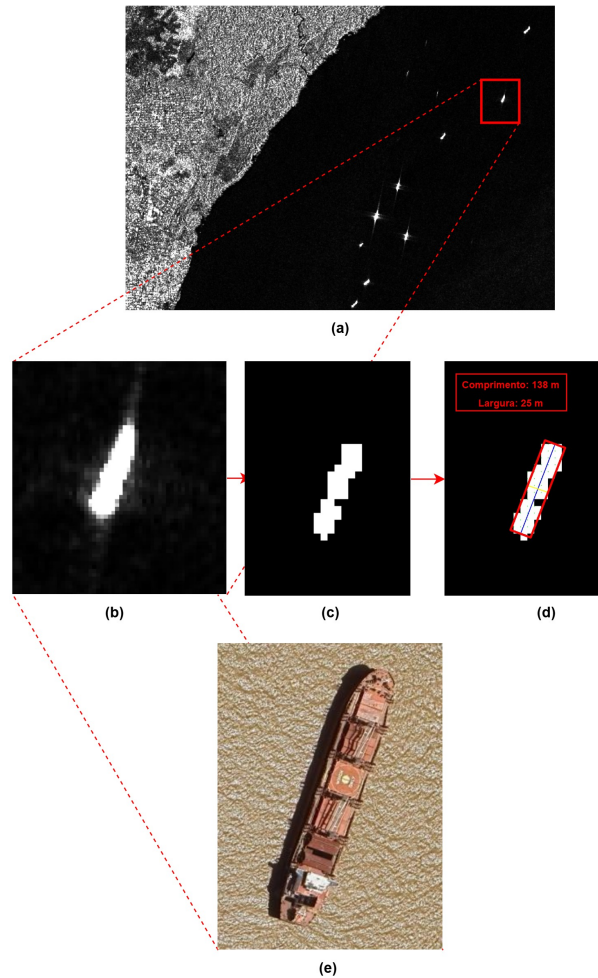


Figura 61 – Processo de extração de atributos na cena referente ao Delta do Amazonas. Em (a) observa-se um recorte da cena do Delta do Amazonas. Em (b), observa-se a embarcação alvo na imagem SAR. Em (c), apresenta-se a imagem binarizada pelo método de Otsu. Em (d), são projetados sobre a imagem binarizada os eixos principal (comprimento) e secundário (largura). Por fim, em (e), é exibido um possível modelo da embarcação detectada pelo algoritmo.

dente ao Delta do Amazonas. Na Figura 63c, observa-se que os comprimentos estimados das embarcações refletem a realidade do fluxo marítimo da região, uma vez que navios graneleiros costumam navegar por essa área. Entretanto, ainda são identificados *outliers* ou superestimações que indicam embarcações com larguras de até 100 metros, o que não é realista. Em síntese, as dimensões estimadas mostram-se coerentes com as características das regiões analisadas.

6.5 Geração de pontos vetoriais

Após a extração de atributos, um arquivo CSV é gerado contendo informações essenciais para a geração dos pontos vetoriais do tipo *Shapefile*. O procedimento consiste

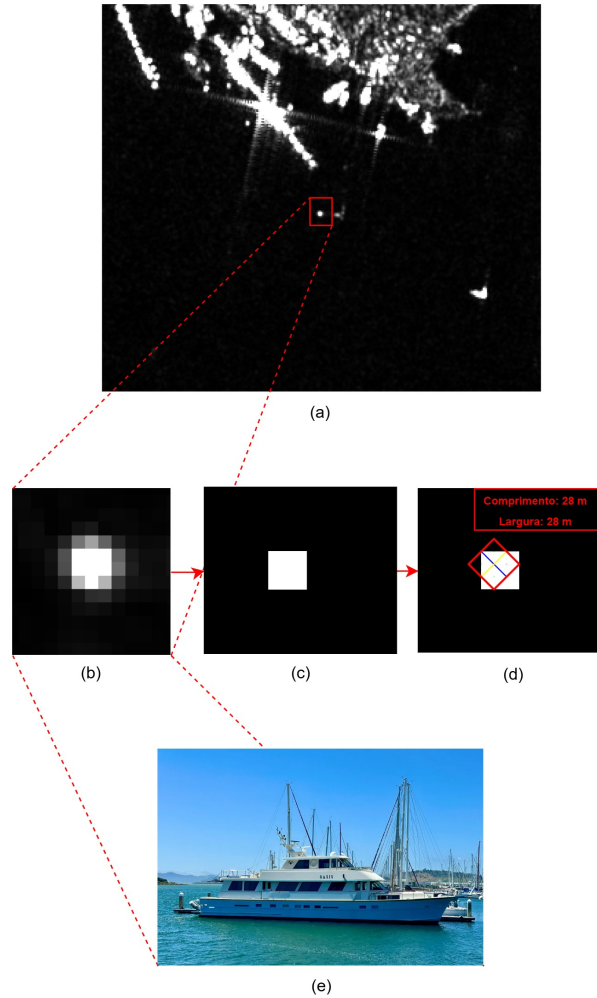


Figura 62 – Processo de extração de atributos na cena referente a Manaus. Em (a) observa-se um recorte da cena. Em (b), observa-se a embarcação alvo na imagem SAR. Em (c), apresenta-se a imagem binarizada pelo método de Otsu. Em (d), são projetados sobre a imagem binarizada os eixos principal (comprimento) e secundário (largura), além da legenda com os valores estimados de comprimento e largura, 28 metros para ambos. Por fim, em (e), é exibido um possível modelo da embarcação detectada pelo algoritmo.

em ler o arquivo de atributos e criar um *GeoDataFrame* a partir das coordenadas das detecções, utilizando o sistema de referência espacial (CRS) extraído automaticamente do raster Sentinel-1 associado. A partir do nome do arquivo de imagem, são obtidos os tempos de início e fim de aquisição (t_{0_utc} e t_{1_utc}). Os nomes dos campos são encurtados para até dez caracteres, respeitando as limitações do formato. Em seguida, os pontos são transformados em geometrias do tipo *Point*. O resultado é exportado em formato *ESRI Shapefile*, preservando as informações de posição, dimensões e confiança das detecções. As principais bibliotecas utilizadas foram **pandas**, para leitura das tabelas em formato CSV, **geopandas** e **shapely** para construção e geração geométrica dos pontos vetoriais e **rasterio** para manipulação da imagem raster. Resultados da exportação dos pontos

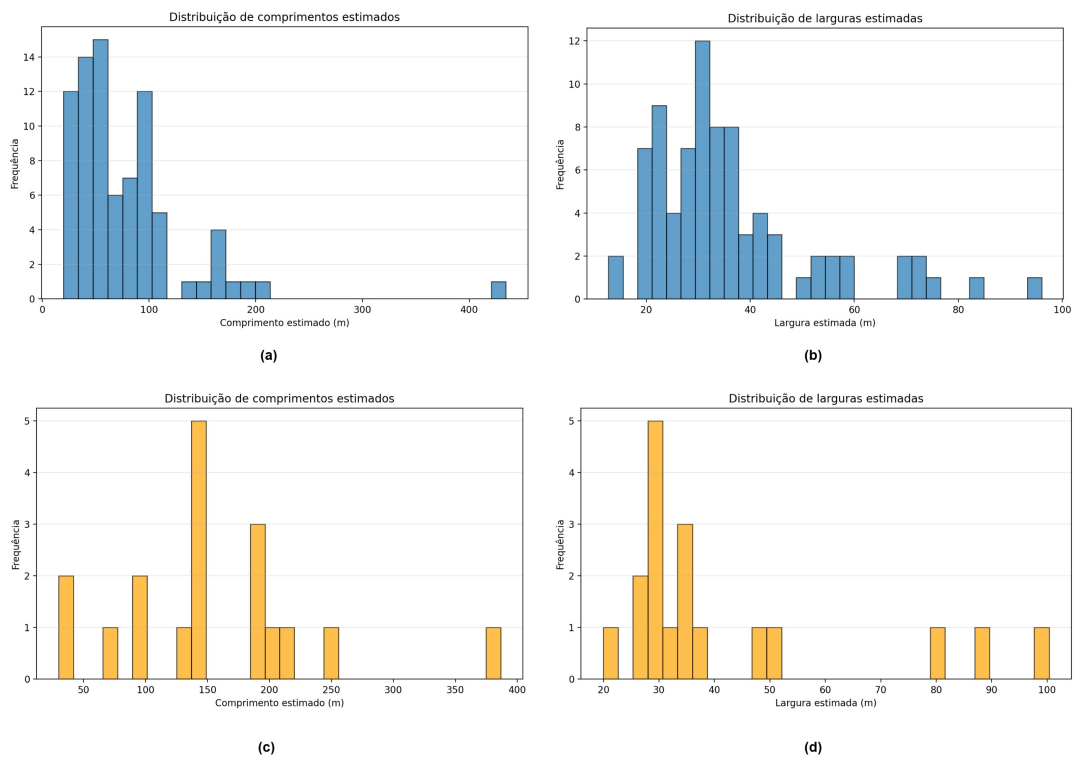


Figura 63 – Histogramas de comprimento e largura estimados para as cenas de Manaus (a,b) e Delta do Amazonas (c-d)

vetoriais podem ser observados na figura 64.

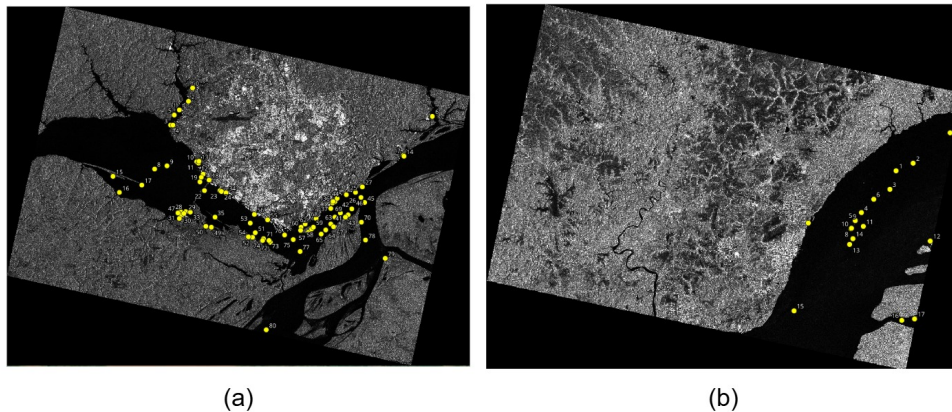


Figura 64 – Pontos vetoriais exportados e projetados no QGIS. em (a) pontos sobrepostos sobre a cena de Manaus e (b) pontos sobre a cena de Delta do Amazonas.

6.6 Experimento final

Como mencionado ao longo deste trabalho, busca-se monitorar a zona da Tríplice Fronteira entre Colômbia, Brasil e Peru. Nesse sentido, realizou-se um experimento final para avaliar o desempenho do modelo escolhido na área de interesse. Conforme discutido

detecoes_pontos—Objetos Totales: 18, Filtrados: 18, Seleccionados: 0

	id	conf	lon	lat	compr_m	larg_m	RA	t0_utc	t1_utc
1	0	0.7194	512843.700803927611560	17894.277442608552519	163.93999999999999	21.320000000000000	7.689000000000000	2025-08-25T09:05:57Z	2025-08-25T09:06:22Z
2	1	0.7314	506352.527475813054480	13292.623846717870038	205.37000000000000	23.250000000000000	8.833000000000000	2025-08-25T09:05:57Z	2025-08-25T09:06:22Z
3	2	0.7193	508401.584494134876877	14224.919862353364806	148.15999999999999	31.530000000000001	4.699000000000000	2025-08-25T09:05:57Z	2025-08-25T09:06:22Z
4	3	0.7578	505564.812125704018399	11054.116301154583198	138.87999999999999	25.859999999999999	5.370000000000000	2025-08-25T09:05:57Z	2025-08-25T09:06:22Z
5	4	0.6478	502134.760474595823325	8237.286093485854508	174.13999999999999	23.920000000000002	7.279000000000000	2025-08-25T09:05:57Z	2025-08-25T09:06:22Z
6	5	0.4044	501411.857146964059211	7339.888859184313333	88.56000000000000	25.510000000000002	3.471000000000000	2025-08-25T09:05:57Z	2025-08-25T09:06:22Z
7	6	0.7605	503610.480371002806351	9857.586655419192539	223.50000000000000	49.420000000000002	4.522000000000000	2025-08-25T09:05:57Z	2025-08-25T09:06:22Z
8	7	0.2864	495768.225651245506015	6980.929965463696135	94.57999999999999	27.969999999999999	3.382000000000000	2025-08-25T09:05:57Z	2025-08-25T09:06:22Z
9	8	0.665	500913.303127907624003	6327.824200499795552	99.28000000000000	42.250000000000000	2.350000000000000	2025-08-25T09:05:57Z	2025-08-25T09:06:22Z
10	9	0.7102	501421.828227345133200	7265.105756325850962	160.93000000000000	31.530000000000001	5.104000000000000	2025-08-25T09:05:57Z	2025-08-25T09:06:22Z
11	10	0.6807	500943.216369051020592	6352.751901452618768	120.00000000000000	42.630000000000003	2.815000000000000	2025-08-25T09:05:57Z	2025-08-25T09:06:22Z
12	11	0.6428	502394.008564505144022	6611.999991361950379	210.71999999999999	28.879999999999999	7.297000000000000	2025-08-25T09:05:57Z	2025-08-25T09:06:22Z
13	12	0.3226	510470.583673219021875	4832.16214330562685	442.87999999999999	102.170000000000002	4.335000000000000	2025-08-25T09:05:57Z	2025-08-25T09:06:22Z
14	13	0.7546	500723.85260066170474	4433.318928085431253	251.06999999999999	55.939999999999998	4.488000000000000	2025-08-25T09:05:57Z	2025-08-25T09:06:22Z
15	14	0.7378	501112.724735530209728	5131.294554764408531	218.34999999999999	28.520000000000000	7.655000000000000	2025-08-25T09:05:57Z	2025-08-25T09:06:22Z
16	15	0.4166	494003.344423785747495	-3588.415238532237709	148.12999999999999	14.100000000000000	10.506000000000000	2025-08-25T09:05:57Z	2025-08-25T09:06:22Z
17	16	0.6553	506990.676620205282234	-4730.103942171423114	215.25000000000000	84.549999999999997	2.546000000000000	2025-08-25T09:05:57Z	2025-08-25T09:06:22Z
18	17	0.4944	508561.121780232992023	-4550.624495311109058	146.30000000000000	66.790000000000006	2.190000000000000	2025-08-25T09:05:57Z	2025-08-25T09:06:22Z

Figura 65 – Tabela de atributos dos pontos vetoriais exportados e projetados no QGIS para a cena de Delta do Amazonas.

em seções anteriores, a variante **s** do YOLOv12 apresentou os melhores resultados; portanto, as inferências na cena do experimento final foram realizadas utilizando esse modelo. O resultado pode ser visto na Figura 66.

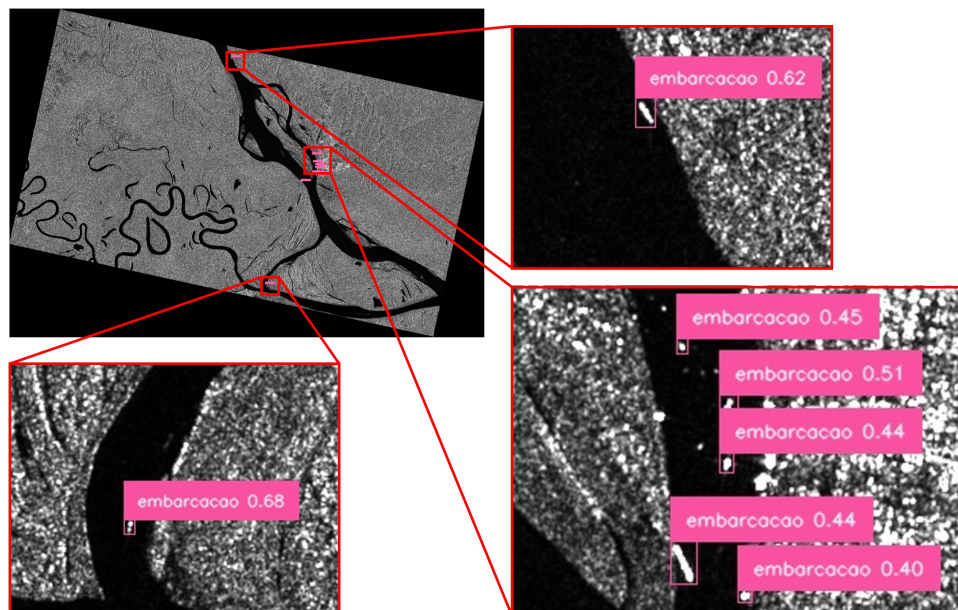


Figura 66 – Experimento final utilizando o modelo YOLOv12-s. Observa-se um desempenho consistente, com detecção de embarcações no centro e nas extremidades da cena, abrangendo diferentes tamanhos, e com a maioria das predições apresentando confiança superior a 40%.

Na Figura 66 é possível observar os resultados da aplicação do modelo YOLOv12 **s** na área da Tríplice Fronteira, apresentando elevada precisão geral. É importante destacar que o modelo consegue detectar embarcações pouco evidentes ao olho humano, como a embarcação localizada na extremidade sul da cena, bem como outra situada na extremidade norte. No centro da cena, observa-se uma maior densidade de detecções. Nesta região,

estão presentes as chamadas *casas flutuantes*, estruturas que, quando analisadas por sensores orbitais, podem se assemelhar a embarcações, gerando potenciais falsos positivos, o que representa uma possível limitação do algoritmo.

Em um primeiro momento, não é possível determinar se as embarcações detectadas são legais ou ilegais. Entretanto, conhecer sua posição, dimensões aproximadas e a data de detecção constitui o primeiro passo essencial para o desenvolvimento de um sistema de monitoramento fluvial na Tríplice Fronteira.

As imagens de treinamento, pesos para inferência e os códigos utilizados neste trabalho foram disponibilizados publicamente. A versão estável utilizada na pesquisa foi arquivada no Zenodo (DOI: 10.5281/zenodo.17686813) para garantir preservação permanente. Além disso, o repositório ativo do projeto, encontra-se disponível no GitHub: [Repositório GitHub](#)

7 Implementação futura

O modelo desenvolvido neste trabalho poderá ser implementado em situações reais de operação por meio de sua integração a plataformas geoespaciais de monitoramento de ativos. Um exemplo é a plataforma Fiscalize.IO, desenvolvida pela empresa ORN Space, que atualmente atua no monitoramento de ativos de grandes empresas dos setores mineral e elétrico, com perspectiva de expansão futura para o monitoramento de rios e áreas litorâneas no Brasil.

A partir das detecções automáticas de embarcações geradas pelo modelo proposto e, mediante acordos institucionais com os órgãos competentes de segurança pública e defesa, será possível estabelecer um intercâmbio de informações entre a empresa operadora da plataforma e as instituições governamentais responsáveis. Esse fluxo de dados pode subsidiar ações de vigilância, fiscalização e resposta rápida, contribuindo para o fortalecimento da segurança e do controle em áreas de interesse estratégico.

8 Conclusões

Este trabalho teve como objetivo a criação de uma base de dados e o treinamento do modelo YOLOv12, em suas variantes **n**, **s** e **m**, para realizar a tarefa de detecção de embarcações no rio Amazonas. Em um país de dimensões continentais, o monitoramento de zonas estratégicas voltado à segurança pública deve ser complementado por instrumentos satelitais capazes de ampliar a vigilância e a capacidade de resposta das autoridades.

A primeira parte deste estudo foi dedicada ao entendimento do problema. Nela, discutiu-se a necessidade do monitoramento de embarcações na Região de Interesse e apresentaram-se os objetivos gerais e específicos da pesquisa. Ao final, constatou-se que os resultados e as ferramentas desenvolvidas neste trabalho podem funcionar como instrumento de apoio às autoridades competentes na mitigação de fluxos ilícitos na área estudada.

Nos capítulos 2 e 3, foi realizada uma revisão da literatura focada na detecção de embarcações, apresentando os trabalhos fundamentais para esta pesquisa, bem como uma revisão sistemática dos conceitos essenciais de sensoriamento remoto, sensoriamento remoto por micro-ondas e radar de abertura sintética (SAR). Essa etapa foi de grande relevância, pois forneceu o embasamento técnico necessário para as decisões relacionadas à escolha de instrumentos, sensores e plataformas empregados no desenvolvimento do projeto.

O capítulo 4 apresentou os fundamentos de inteligência artificial e, em especial, os aspectos técnicos do modelo YOLO, incluindo as inovações introduzidas na versão YOLOv12. Essa seção é fundamental para compreender o modelo adotado como base no sistema de detecção de embarcações.

O capítulo dedicado à metodologia constituiu a espinha dorsal do trabalho. Foram identificadas e descritas as abordagens mais adequadas para alcançar os objetivos propostos. Revisaram-se os softwares utilizados, definiu-se o satélite selecionado, estabeleceu-se a plataforma de aquisição das cenas, e detalhou-se o pré-processamento das imagens, a anotação das amostras, o treinamento dos modelos, as métricas de avaliação e, por fim, o método de estimativa dos atributos geométricos das embarcações e a geração dos produtos vetoriais.

A seção de resultados e discussões apresentou uma análise abrangente, contemplando produtos qualitativos e quantitativos derivados dos experimentos realizados.

Em síntese, observou-se que o processo de criação de uma base de dados pode ser amplamente automatizado mediante o uso das ferramentas tecnológicas disponíveis, pro-

duzindo um conjunto de dados de alta qualidade, contudo, com supervisão e intervenção do analista. Verificou-se também que os resultados de treinamento podem apresentar métricas elevadas quando o conjunto de treinamento possui baixa variabilidade de cenários. Entretanto, ao generalizar para outros contextos, observou-se queda nesses indicadores, especialmente em ambientes menos ideais. Ainda assim, o melhor modelo treinado apresentou métricas de avaliação satisfatórias.

A partir desse modelo, foi desenvolvido com sucesso um método para o cálculo dos atributos geométricos das embarcações detectadas, demonstrando um desempenho compatível com o esperado, embora sua avaliação completa tenha sido limitada pela ausência de dados AIS nas cenas analisadas. Além disso, foi gerada uma camada vetorial contendo todas as informações relevantes das embarcações detectadas.

Conclui-se, portanto, que todos os objetivos específicos foram alcançados, ainda que tenham sido identificadas oportunidades claras de aprimoramento e de continuidade para trabalhos futuros.

9 Trabalhos futuros

Com base nos resultados alcançados, alguns direcionamentos podem ser considerados para dar continuidade e aprimorar esta pesquisa.

O primeiro consiste no desenvolvimento de um método para separar alvos do tipo ilha. Observou-se que a maior parte dos falsos positivos está relacionada a pequenas porções de terra confundidas com embarcações. Dessa forma, recomenda-se a criação de um procedimento específico para discriminar esses alvos, possivelmente por meio da análise de valores de retroespalhamento e do estabelecimento de limiares adequados.

Outro ponto relevante é a integração de dados AIS (*Automatic Identification System*). A incorporação dessas informações pode contribuir para a validação das detecções e dos atributos extraídos, especialmente em regiões nas quais esse sistema está disponível e em embarcações equipadas com o *transponder*, aumentando a confiabilidade do processo.

Também se destaca a possibilidade de utilizar imagens ICEYE. Considerando que a Força Aérea Brasileira opera satélites ICEYE (constelação Lessônia) para fins de monitoramento territorial, o uso dessas imagens, tanto para treinamento quanto para inferência, pode reduzir a dependência de dados estrangeiros, mesmo quando disponibilizados gratuitamente. Tal medida reforça a soberania espacial do Brasil no contexto de monitoramento e detecção marítima.

Por fim, recomenda-se o aprimoramento da estimativa de atributos por meio de métodos de aprendizado profundo. Embora a abordagem adotada neste trabalho tenha fornecido estimativas aceitáveis das dimensões das embarcações, estudos baseados em técnicas de aprendizado profundo, como o apresentado por (REN; LI; XU, 2023), demonstram desempenho superior. Assim, a investigação desses métodos constitui uma linha promissora para pesquisas futuras.

Referências

- ACADEMY, K. *Luz e o espectro eletromagnético*. 2025. Acesso em: 25 mar. 2025. Disponível em: <<https://pt.khanacademy.org/science/physics/light-waves/introduction-to-light-waves/a/light-and-the-electromagnetic-spectrum>>. Citado 2 vezes nas páginas 9 e 33.
- AGÊNCIA NACIONAL DE ÁGUAS E SANEAMENTO BÁSICO (ANA). *Região Hidrográfica Amazônica*. 2025. Disponível em: <<https://www.gov.br/ana/pt-br/assuntos/gestao-das-aguas/panorama-das-aguas/regioes-hidrograficas/regiao-hidrografica-amazonica/regiao-hidrografica-amazonica-1>>. Acesso em: 5 mar. 2025. Citado na página 22.
- Agência Cenarium. *Rios amazônicos são principais rotas para transporte de toneladas de drogas*. 2025. Disponível em: <<https://agenciacenarium.com.br/rios-amazonicos-sao-principais-rotas-para-transporte-de-toneladas-de-drogas/>>. Acesso em: 6 mar. 2025. Citado 2 vezes nas páginas 9 e 24.
- Alaska Satellite Facility. *Introduction to SAR - HyP3*. 2025. <https://hyp3-docs.asf.alaska.edu/guides/introduction_to_sar/>. Acesso em: 24 jun. 2025. Citado na página 67.
- ANDRADE, L. D. N. R.; GRANZIERA, M. L. M. Abordagens paradiplomáticas na tríplice fronteira amazônica: Brasil, colômbia e peru. *Revista Tempo do Mundo (RTM)*, v. 27, n. 27, p. 305–324, 2022. Acesso em: 5 mar. 2025. Citado na página 23.
- ASF – Alaska Satellite Facility. *ASF Data Search*. 2024. Disponível em: <<https://search.asf.alaska.edu/#/>>. Citado na página 62.
- AWAIS, C. M. et al. A survey on sar ship classification using deep learning. *arXiv*, arXiv, arXiv:2503.11906, mar 2025. Acesso em: 26 set. 2025. Disponível em: <<https://doi.org/10.48550/arXiv.2503.11906>>. Citado na página 83.
- BEZERRA, D. X. *Deteção de Alvos Marítimos a partir de dados Sentinel-1 em Região Oceânica Adjacente ao Nordeste do Brasil*. Trabalho de Conclusão de Curso, 2021. Acesso em: 9 out. 2025. Disponível em: <<https://doi.org/10.13140/RG.2.2.33991.98725>>. Citado na página 29.
- BRAUN, A. *Radar satellite imagery for humanitarian response: Bridging the gap between technology and application*. Tese (Doctoral Thesis) — University of Tübingen, Faculty of Science, Department of Geography, Tübingen, Germany, 2019. Advisor: Volker Hochschild. Oral examination: 26 Jul. 2019. Disponível em: <<http://hdl.handle.net/10900/91317>>. Citado 2 vezes nas páginas 71 e 74.
- CHENG, B.; BRADLEY, P. A. What machine learning can and cannot do for inertial confinement fusion. *Plasma*, v. 6, n. 2, p. 334–344, 2023. Disponível em: <<https://www.mdpi.com/2571-6182/6/2/23>>. Acesso em: 22 ago. 2025. Disponível em: <<https://www.mdpi.com/2571-6182/6/2/23>>. Citado 3 vezes nas páginas 9, 44 e 45.

COLAB.GOOGLE. *colab.google*. <<http://0.0.0.0:8080/>>. Acesso em: 9 jun. 2024. Citado na página 61.

COLWELL, R. N. The emergence of remote sensing as a major environmental tool. *Photogrammetric Engineering & Remote Sensing*, v. 55, n. 7, p. 979–984, jul 1989. Acesso em: 20 mar. 2025. Disponível em: <https://www.asprs.org/wp-content/uploads/pers/1989journal/jul/1989_jul_979-984.pdf>. Citado na página 32.

Copernicus – European Space Agency. *S1 Processing*. 2024. Disponível em: <<https://sentiwiki.copernicus.eu/web/s1-processing>>. Citado na página 63.

COUTO, A. C. O. Fronteiras e estrutura espacial do narcotráfico na amazônia. *Boletim Gaúcho de Geografia*, v. 47, n. 1, p. xx–xx, 2020. Citado na página 23.

CUI, B.; LIU, Z.; YANG, Q. Uav-yolo12: A multi-scale road segmentation model for uav remote sensing imagery. *Drones*, v. 9, n. 8, p. 533, 2025. Disponível em: <<https://doi.org/10.3390/drones9080533>>. Citado na página 55.

CUI, B.; LIU, Z.; YANG, Q. Uav-yolo12: A multi-scale road segmentation model for uav remote sensing imagery. *Drones*, v. 9, n. 8, 2025. ISSN 2504-446X. Disponível em: <<https://www.mdpi.com/2504-446X/9/8/533>>. Citado na página 55.

DANE. *Cuentas Nacionales Departamentales*. 2025. Disponível em: <<https://www.dane.gov.co/index.php/estadisticas-por-tema/cuentas-nacionales/cuentas-nacionales-departamentales>>. Acesso em: 5 mar. 2025. Citado na página 23.

DataCamp. *How Transformers Work*. 2025. Accessed: 2025-11-09. Disponível em: <<https://www.datacamp.com/tutorial/how-transformers-work>>. Citado na página 49.

DEMIR, H.; SARI, F. The effect of artificial intelligence and industry 4.0 on robotic systems. In: *Book Title*. [S.l.]: Publisher Name, 2020. p. 51–72. ISBN 978-625-7139-55-7. Citado na página 44.

DORNELLES, K. *Absortância Solar de Superfícies Opacas: Métodos de Determinação e Base de Dados para Tintas Látex Acrílica e PVA*. 2008. Disponível em: <<http://rgdoi.net/10.13140/RG.2.1.4498.0967>>. Acesso em: 26 mar. 2025. Disponível em: <<http://rgdoi.net/10.13140/RG.2.1.4498.0967>>. Citado 2 vezes nas páginas 9 e 35.

Earth Science Data Systems, NASA. *Create an Interferogram Using ESA's Sentinel-1 Toolbox / NASA Earthdata*. 2016. NASA Earthdata – Earth Science Data Systems. Disponível em: <<https://www.earthdata.nasa.gov/learn/data-recipes/create-interferogram-using-esas-sentinel-1-toolbox>>. Acesso em: 23 ago. 2025. Citado 2 vezes nas páginas 10 e 63.

ESA. *S1 Mission*. 2025. <<https://sentiwiki.copernicus.eu/web/s1-mission>>. Acesso em: 20 abr. 2025. Citado na página 66.

European Space Agency. *Apply Orbit File Operation*. <<https://step.esa.int/main/wp-content/help/versions/10.0.0/snap-toolboxes/eu.esa.microwavetbx.sar.io.ephemeris.ui/operators/ApplyOrbitFileOp.html>>. Acesso em: 8 dez. 2024. Citado na página 71.

European Space Agency. *Thermal Noise Removal*. <<https://step.esa.int/main/wp-content/help/versions/10.0.0/snap-toolboxes/eu.esa.microwavetbx.sar.op.calibration.ui/operators/ThermalNoiseRemoval.html>>. Acesso em: 8 dez. 2024. Citado na página 71.

European Space Agency. *Mission ends for Copernicus Sentinel-1B satellite*. 2025. Acesso em: 9 mar. 2025. Disponível em: <https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Sentinel-1/Mission_ends_for_Copernicus_Sentinel-1B_satellite>. Citado 2 vezes nas páginas 9 e 25.

FILIPPONI, F. Sentinel-1 grd preprocessing workflow. In: *3rd International Electronic Conference on Remote Sensing*. MDPI, 2019. p. 11. Disponível em: <<https://www.mdpi.com/2504-3900/18/1/11>>. Citado na página 71.

FLORES, A. et al. *Synthetic Aperture Radar (SAR) Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation*. [s.n.], 2019. Acesso em: 2 abr. 2025. Disponível em: <https://gis1.servirglobal.net/TrainingMaterials/SAR/SARHB_FullRes.pdf>. Citado 4 vezes nas páginas 10, 14, 76 e 77.

GALVEZ, R. L. et al. Object detection using convolutional neural networks. In: *Proceedings of TENCON 2018 – 2018 IEEE Region 10 Conference*. Jeju, Korea: IEEE, 2018. p. 2023–2027. Citado na página 53.

GAMBA, S. R. H.; SANO, E. E.; ROCHA, M. P. Identificação de embarcações em imagens aerotransportadas de radar de abertura sintética (r-99 sar) na área marítima do brasil. *Boletim de Ciências Geodésicas*, v. 17, n. 3, p. 458–475, 2011. Acesso em: 23 dez. 2024. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1982-21702011000300008&lng=pt&tlng=pt>. Citado na página 29.

German Aerospace Center (DLR), Microwaves & Radar Institute (HR), Pol-InSAR Research Group. *The Pol-InSAR course*. 2024. Prepared by DLR-HR's Pol-InSAR Team. Contact: kostas.papathanassiou@dlr.de, matteo.pardini@dlr.de, islam.mansour@dlr.de. Citado na página 41.

GOMES, J. V. P. *Fundamentos do sensoriamento remoto*. Curitiba, PR: Intersaberes, 2020. Citado na página 32.

GOMES, M. F. *Avaliação de dados Radarsat-1 e CBERS-2 para estimativa da estrutura do cerrado: uma abordagem utilizando dados alométricos e históricos*. 2006. Acesso em: 1 abr. 2025. Disponível em: <<http://rgdoi.net/10.13140/2.1.1241.4249>>. Citado 3 vezes nas páginas 9, 36 e 38.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. Cambridge, Massachusetts: The MIT Press, 2016. (Adaptive computation and machine learning). Citado 4 vezes nas páginas 9, 47, 48 e 50.

GREIDANUS, H. et al. The sumo ship detector algorithm for satellite radar images. *Remote Sensing*, MDPI, v. 9, n. 3, p. 246, mar 2017. Acesso em: 9 out. 2025. Disponível em: <<https://doi.org/10.3390/rs9030246>>. Citado na página 85.

GROVER, A.; KUMAR, S.; KUMAR, A. Ship detection using sentinel-1 sar data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information*

- Sciences*, IV–5, p. 317–324, 2018. Accessed: 2024-12-23. Disponível em: <<https://isprs-annals.copernicus.org/articles/IV-5/317/2018/>>. Citado na página 29.
- GUIDE, B. C. *Size Range of Bulk Carriers*. 2025. Accessed: 2025-11-08. Disponível em: <<https://bulkcarrierguide.com/size-range.html>>. Citado na página 104.
- GUMELAR, O.; LIU, D.; RIZKIYANTO, R. Polarimetric decomposition in calibrated radar image for detecting vegetation object. In: *Proceedings of the 2017 3rd International Conference on Science in Information Technology (ICSITech)*. Yogyakarta, Indonesia: IEEE, 2017. p. 499–503. Citado 2 vezes nas páginas 10 e 73.
- GÉRON, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. 2. ed. [s.l.]: O'Reilly Media, 2019. Citado na página 45.
- HAENLEIN, M.; KAPLAN, A. A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review*, v. 61, n. 4, p. 5–14, jul 2019. Disponível em: <<https://journals.sagepub.com/doi/10.1177/0008125619864925>>. Acesso em: 22 ago. 2025. Disponível em: <<https://journals.sagepub.com/doi/10.1177/0008125619864925>>. Citado na página 44.
- IBGE. *Tabatinga (AM) | Cidades e Estados*. 2025. Disponível em: <<https://www.ibge.gov.br/cidades-e-estados/am/tabatinga.html>>. Acesso em: 5 mar. 2025. Citado na página 23.
- IBM. *O que é ajuste fino?* 2024. Accessed: 2025-09-19. Disponível em: <<https://www.ibm.com/br-pt/think/topics/fine-tuning>>. Citado na página 81.
- IBM. *What Is Deep Learning?* 2025. Disponível em: <<https://www.ibm.com/think/topics/deep-learning>>. Acesso em: 22 ago. 2025. Disponível em: <<https://www.ibm.com/think/topics/deep-learning>>. Citado na página 47.
- IMAGE Kernels explained visually. 2025. Explained Visually. Disponível em: <<https://setosa.io/ev/image-kernels/>>. Acesso em: 23 ago. 2025. Citado na página 52.
- JACARANDÁ, R. A cocaína na amazônia: o tráfico de drogas e a redistribuição das redes criminais no sudoeste amazônico. *Boletim de Análise Político-Institucional*, n. 36, p. 81–90, 2024. Acesso em: 5 mar. 2025. Citado na página 23.
- JEGHAM, N. et al. Evaluating the evolution of yolo (you only look once) models: A comprehensive benchmark study of yolo11 and its predecessors. *arXiv preprint*, 2024. Disponível em: <<https://arxiv.org/abs/2411.00201>>. Citado na página 53.
- JENSEN, J. R. *Remote Sensing of the Environment: An Earth Resource Perspective*. Second edition [exclusive edition only for the benefit of students outside the united states and canada]. [S.l.]: Pearson, 2014. Citado 2 vezes nas páginas 36 e 37.
- Khan Academy. *O neurônio: estrutura e função*. 2025. Disponível em: <<https://pt.khanacademy.org/science/6-ano/vida-e-evolucao-os-sistemas-do-corpo-humano/os-neuronios/a/o-neuronio-estrutura-e-funcao>>. Acesso em: 23 ago. 2025. Disponível em: <<https://pt.khanacademy.org/science/6-ano/vida-e-evolucao-os-sistemas-do-corpo-humano/os-neuronios/a/o-neuronio-estrutura-e-funcao>>. Citado 2 vezes nas páginas 9 e 46.

- KHANAM, R.; HUSSAIN, M. *A Review of YOLOv12: Attention-Based Enhancements vs. Previous Versions*. 2025. Disponível em: <<https://arxiv.org/abs/2504.11995>>. Citado na página 55.
- KMUSSER. *Map of the Amazon River drainage basin with the Amazon River highlighted*. 2013. Disponível em: <<https://commons.wikimedia.org/wiki/File:Amazonrivermap.svg>>, na Wikipedia. Licença: CC BY-SA 3.0. Criação: 25 de fevereiro de 2013. Carregamento: 6 de agosto de 2015. Acesso em: 5 mar. 2025. Citado 2 vezes nas páginas 9 e 22.
- KUMAR, A. *Fine-Tuning YOLO for Custom Object Detection: A Complete Guide*. 2025. Medium. Acesso em: 26 set. 2025. Disponível em: <<https://medium.com/@amresh.kumar11/fine-tuning-yolo-for-custom-object-detection-a-complete-guide-6ce3724ce9f1>>. Citado 2 vezes nas páginas 82 e 83.
- LI, Y. et al. Sar ship detection based on resnet and transfer learning. In: *Proceedings of the 2019 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Yokohama, Japan: [s.n.], 2019. p. 1188–1191. Citado na página 29.
- LILLESAND, T. M.; KIEFER, R. W.; CHIPMAN, J. W. *Remote Sensing and Image Interpretation*. 6. ed. New Delhi: Wiley, 2013. Student edition. Citado 5 vezes nas páginas 32, 34, 35, 38 e 39.
- LORENZZETTI, J. A. *Princípios físicos de sensoriamento remoto*. [s.l.]: Editora Edgard Blucher, 2022. Citado 2 vezes nas páginas 32 e 36.
- LUDERMIR, T. B. Inteligência artificial e aprendizado de máquina: estado atual e tendências. *Estudos Avançados*, v. 35, n. 101, p. 85–94, 2021. Disponível em: <http://www.scielo.br/scielo.php?script=sci_abstract&pid=S0103-40142021000100085&lng=pt>. Acesso em: 22 ago. 2025. Disponível em: <>. Citado 2 vezes nas páginas 44 e 45.
- MAGALHÃES, C. A. B. d.; CARMO, S. A. d. *TRÍPLICE FRONTEIRA: ASPECTOS DO CRIME ORGANIZADO EM RORAIMA*. 2022. Disponível em: <<https://zenodo.org/record/7484534>>. Acesso em: 5 mar. 2025. Citado na página 23.
- MARZUKI, M. I. et al. Fishing boat detection using sentinel-1 validated with viirs data. *IOP Conference Series: Earth and Environmental Science*, v. 925, n. 1, p. 012058, 2021. Disponível em: <<https://iopscience.iop.org/article/10.1088/1755-1315/925/1/012058>>. Citado na página 67.
- MENESES, P. R.; ALMEIDA, T. d. *Introdução ao Processamento de Imagens de Sensoriamento Remoto*. [S.l.]: Universidade de Brasília (UnB), 2012. Citado 2 vezes nas páginas 36 e 37.
- MENESES, P. R.; ALMEIDA, T. d.; BAPTISTA, G. M. d. M. *Refletância dos materiais terrestres*. [S.l.]: Edição Português, 2019. Capa comum, Compiladores: Paulo Roberto Meneses, Tati de Almeida, Gustavo Macedo de Mello Baptista. Citado na página 32.

MEYER, R. G. V.; KLEYNHANS, W.; SCHWEGMANN, C. P. Small ships don't shine: Classification of ocean vessels from low resolution, large swath area sar acquisitions. In: *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Beijing, China: IEEE, 2016. p. 975–978. Accessed: 2024-12-23. Disponível em: <<http://ieeexplore.ieee.org/document/7729247/>>. Citado na página 29.

MINISTÉRIO DAS RELAÇÕES EXTERIORES. *Geografia*. 2025. Disponível em: <<https://www.gov.br/mre/pt-br/embaixada-bogota/o-brasil/geografia>>. Acesso em: 5 mar. 2025. Citado na página 22.

Ministério da Justiça e Segurança Pública. *Programa Brasil MAIS*. 2024. Accessed: 2024-01-04. Disponível em: <https://www.gov.br/mj/pt-br/acesso-a-informacao/acoes-e-programas/programa_brasil_mais>. Citado na página 24.

MOREIRA, A. et al. A tutorial on synthetic aperture radar. *IEEE Geoscience and Remote Sensing Magazine*, v. 1, n. 1, p. 6–43, 2013. Acesso em: 20 fev. 2025. Disponível em: <<http://ieeexplore.ieee.org/document/6504845/>>. Citado 3 vezes nas páginas 39, 42 e 43.

NATIONAL GEOGRAPHIC. *As 5 curiosidades sobre o maior rio do mundo*. 2025. Disponível em: <<https://www.nationalgeographicbrasil.com/viagem/2023/05/as-5-curiosidades-sobre-o-m>>. Acesso em: 5 mar. 2025. Citado na página 22.

NGUYEN, T. *Improve Underwater Object Detection through YOLOv12 Architecture and Physics-informed Augmentation*. 2025. Disponível em: <<https://arxiv.org/abs/2506.23505>>. Citado na página 56.

OCHA. *Plan de Acción Triple Frontera Amazónica: Colombia-Brasil-Perú (agosto 2020)*. 2020. Disponível em: <<https://www.unocha.org/publications/report/brazil/plan-de-acci-n-triple-fro>>. Acesso em: 6 mar. 2025. Citado 2 vezes nas páginas 9 e 23.

OLIVEIRA, H. C. T. d. *Detecção automática de pequenas embarcações em imagens SAR da constelação ICEYE com uso de uma rede neural convolucional*. Trabalho de Conclusão de Curso de Especialização, 2021. Curso de Especialização em Análise de Ambiente Eletromagnético. Acesso em: 9 out. 2025. Citado na página 31.

OpenCV Team. *Image Processing — Structural Analysis and Shape Descriptors*. 2025. Accessed: 2025-11-09. Disponível em: <https://docs.opencv.org/4.x/d3/dc0/group__imgproc__shape.html#gaf849da1fdafa67ee84b1e9a23b93f91f>. Citado na página 104.

PATRUNO, J. Polarimetric radarsat-2 and alos palsar multi-frequency analysis over the archaeological site of gebel barkal (sudan). *Archaeological Prospection*, v. 21, n. 2, p. 109–117, 2014. Citado 2 vezes nas páginas 10 e 75.

PODAREANU, D. et al. *Best Practice Guide - Deep Learning*. 2019. Disponível em: <<https://doi.org/10.13140/RG.2.2.31564.05769>>. Acesso em: 23 ago. 2025. Citado 4 vezes nas páginas 9, 50, 51 e 52.

Programa Brasil MAIS. *RedeMAIS: Plataforma PF-SCCON*. 2024. Accessed: 2024-01-04. Disponível em: <<https://plataforma-pf.sccon.com.br/#/>>. Citado na página 24.

PUCHI-CABRERA, E. et al. Machine learning aided nanoindentation: A review of the current state and future perspectives. *Current Opinion in Solid State and Materials Science*, v. 27, p. 101091, jul 2023. Disponível em: <<https://doi.org/10.1016/j.cossms.2023.101091>>. Acesso em: 22 ago. 2025. Disponível em: <<https://doi.org/10.1016/j.cossms.2023.101091>>. Citado 2 vezes nas páginas 44 e 46.

PyImageSearch. *Breaking the CNN Mold: YOLOv12 Brings Attention to Real-Time Object Detection*. 2025. Accessed: 2025-11-09. Disponível em: <<https://pyimagesearch.com/2025/07/07/breaking-the-cnn-mold-yolov12-brings-attention-to-real-time-object-detection/>>. Citado na página 102.

RAZZAQUE, A. *Hands-On Deep Learning for IoT: Train neural network models to develop intelligent IoT applications*. 1. ed. Birmingham: Packt Publishing Limited, 2019. Citado 3 vezes nas páginas 46, 47 e 49.

REDMON, J. et al. You only look once: Unified, real-time object detection. In: *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, 2016. p. 779–788. Disponível em: <<http://ieeexplore.ieee.org/document/7780460/>>. Citado 3 vezes nas páginas 9, 53 e 54.

REN, Y.; LI, X.; XU, H. Extracting ship's size from sar images by deep learning. In: LI, X.; WANG, F. (Ed.). *Artificial Intelligence Oceanography*. Singapore: Springer, 2023. Disponível em: <https://doi.org/10.1007/978-981-19-6375-9_15>. Citado na página 113.

RIBEIRO, L. A. *A logística do tráfico de cocaína na Bacia do Rio Amazonas: o impacto na securitização da Tríplice Fronteira entre Brasil, Colômbia e Peru*. Tese (Doutorado) — Universidade Federal de Santa Catarina, Florianópolis, 2022. Monografia (Bacharelado em Relações Internacionais), Centro Socioeconômico, Departamento de Ciências Econômicas e Relações Internacionais. Citado na página 23.

Roboflow. *Data Annotation Explained: Types, Tools, When to Use Each*. 2025. Roboflow Blog. Disponível em: <<https://blog.roboflow.com/data-annotation/>>. Acesso em: 29 ago. 2025. Citado na página 79.

Roboflow. *How to Train a YOLOv12 Object Detection Model on a Custom Dataset*. 2025. Accessed: 2025-09-19. Disponível em: <<https://blog.roboflow.com/train-yolov12-model/>>. Citado na página 81.

Roboflow. *Introduction to Roboflow Annotate / Roboflow Docs*. 2025. Roboflow Documentation. Disponível em: <<https://docs.roboflow.com/annotate/annotation-tools>>. Acesso em: 29 ago. 2025. Citado na página 79.

RUSSELL, S. J.; NORVIG, P. *Artificial intelligence: a modern approach*. 4. ed. Hoboken: Pearson, 2021. (Pearson series in artificial intelligence). Citado na página 44.

SALEHI, H.; BURGUEÑO, R. Emerging artificial intelligence methods in structural engineering. *Engineering Structures*, v. 171, p. 170–189, 2018. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0141029617335526>>. Acesso em: 22 ago. 2025. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0141029617335526>>. Citado na página 44.

SANO, E. E.; MATRICARDI, E. A. T.; CAMARGO, F. F. Estado da arte do sensoriamento remoto de radar: Fundamentos, sensores, processamento de imagens e aplicações. *Revista Brasileira de Cartografia*, 2020. Disponível em: <<https://doi.org/10.14393/revbrascartogr>>. Citado na página 36.

SAUSEN, T. M.; LACRUZ, M. S. P. (Ed.). *Sensoriamento remoto para desastres*. 1. ed. São Paulo: Oficina de Textos, 2015. 288 p. Disponível também em formato digital (eISBN 978-85-7975-178-3). ISBN 978-85-7975-175-2. Citado na página 37.

scikit-image developers. *Thresholding — scikit-image 0.25.2 documentation*. 2025. Scikit-image Documentation. Disponível em: <https://scikit-image.org/docs/0.25.x/auto_examples/segmentation/plot_thresholding.html>. Acesso em: 24 ago. 2025. Citado na página 69.

SINGH, V. *Breaking the CNN Mold: YOLOv12 Brings Attention to Real-Time Object Detection*. 2025. PyImageSearch. Disponível em: <<https://pyimagesearch.com/2025/07/07/breaking-the-cnn-mold-yolov12-brings-attention-to-real-time-object-detection/>>. Acesso em: 7 jul. 2025. Citado na página 54.

SNAP - Earth Online. *SNAP - Earth Online*. 2024. <<https://earth.esa.int/eogateway/tools/snap>>. Acesso em: 9 jun. 2024. Citado na página 61.

TORNERO, M. *YOLOv12 Architecture Explained*. 2025. YouTube. Acesso em: 29 nov. 2025. Disponível em: <<https://www.youtube.com/watch?v=Luiaac3eDuU>>. Citado 5 vezes nas páginas 10, 56, 57, 58 e 59.

TURNER, R. E. *An Introduction to Transformers*. 2023. Disponível em: <<https://doi.org/10.48550/ARXIV.2304.10557>>. Citado na página 48.

ULTRALYTICS. *Entrenar*. 2025. Ultralytics Documentation. Accedido 26 de septiembre de 2025. Disponível em: <<https://docs.ultralytics.com/es/modes/train>>. Citado 2 vezes nas páginas 82 e 83.

Ultralytics. *YOLO12 explicado: Aplicações do mundo real*. 2025. Ultralytics Blog. Disponível em: <<https://www.ultralytics.com/pt/blog/yolo12-explained-real-world-applications-and-use-ca>>. Acesso em: 8 set. 2025. Citado na página 55.

VASWANI, A. et al. *Attention Is All You Need*. 2017. Disponível em: <<https://doi.org/10.48550/ARXIV.1706.03762>>. Citado 2 vezes nas páginas 9 e 49.

WANG, S. et al. Yolo-sd: Small ship detection in sar images by multi-scale convolution and feature transformer module. *Remote Sensing*, v. 14, n. 20, p. 5268, 2022. Citado 3 vezes nas páginas 10, 31 e 77.

WEI, S. et al. Hrsid: A high-resolution sar images dataset for ship detection and instance segmentation. *IEEE Access*, v. 8, p. 120234–120254, 2020. Citado 2 vezes nas páginas 31 e 79.

YACHTWORLD. *1981 Hatteras Motoryacht*. 2025. Accessed: 2025-11-09. Disponível em: <<https://www.yachtworld.com/yacht/1981-hatteras-motoryacht-9724588/>>. Citado na página 104.

ZANNOA, D. C. et al. *Processamento de imagens de satélite*. [S.l.]: Oficina de Textos, 2019. Citado 3 vezes nas páginas 9, 25 e 33.

ZHANG, T.; ZHANG, X. High-speed ship detection in sar images based on a grid convolutional neural network. *Remote Sensing*, v. 11, n. 10, p. 1206, May 2019. Citado na página 30.

ZHANG, T. et al. Sar ship detection dataset (ssdd): Official release and comprehensive data analysis. *Remote Sensing*, v. 13, n. 18, p. 3690, 2021. Disponível em: <<https://www.mdpi.com/2072-4292/13/18/3690>>. Citado na página 30.

ZHANG, Y.; HAO, Y. A survey of sar image target detection based on convolutional neural networks. *Remote Sensing*, v. 14, p. 6240, 2022. Citado na página 30.

ØSTERGAARD, A. et al. Sentinel-1 in-orbit calibration approach. In: *Proceedings of the ESA Living Planet Symposium*. Edinburgh, United Kingdom: European Space Agency, 2013. ESA-ESTEC, EADS Astrium, DLR, Thales Alenia Space Italy. Citado 2 vezes nas páginas 10 e 63.