Avaliação de *Filter Augmented Input* para Detecção de Pistas de Pouso em Imagens SAR do Sentinel-1 com YOLO11x

David Kennedy de Sousa dos Santos[1], Mateus Habermann[1]e Lester de Abreu Faria[1]
[1]Instituto Tecnológico de Aeronáutica (ITA), São José dos Campos/SP – Brasil

***Resumo*— A detecção automática de pistas de pouso em regiões remotas da Amazônia é relevante para aplicações de monitoramento territorial e segurança, mas permanece desafiadora devido às características radiométricas das imagens de Radar de Abertura Sintética (SAR). Este trabalho avalia uma adaptação da estratégia de *Filter Augmented Input* para a detecção de pistas em imagens SAR do Sentinel-1 com a arquitetura YOLO11x. Foram comparadas duas entradas diferentes na rede neural, formada pela replicação da imagem SAR nos três canais, e uma configuração aumentada, composta pela imagem SAR original, pela *Wavelet Scattering Transform* (WST) e pelo *Histogram of Oriented Gradients* (HOG) da imagem original. Os experimentos foram realizados no conjunto *S1-AAD: Sentinel-1 Amazon Airstrip Dataset*, com 832 imagens de treinamento e 208 de validação, mantendo-se as mesmas partições, anotações e configurações de treinamento em três sementes aleatórias. A entrada SAR+WST+HOG apresentou ganhos médios de 9,37 pontos percentuais em *Precision*, 5,04 em *Recall*, 6,53 em mAP50 e 0,52 em mAP50-95. Os resultados indicam que representações multiescalares e direcionais podem fornecer informações complementares para a detecção de estruturas lineares em imagens SAR, com benefícios mais pronunciados na identificação dos alvos do que na precisão de suas localizações.**

***Palavras-Chave*— SAR, YOLO, Sentinel-1, detecção de objetos, pistas de pouso, *Filter Augmented Input*.**

I. Introdução

O domínio do espaço aéreo sempre esteve associado a capacidades estratégicas, tanto em aspectos psicossociais quanto operacionais. Nos últimos anos, a crescente relevância do espaço aéreo e espacial tem impulsionado a adoção de tecnologias capazes de ampliar a capacidade de monitoramento, vigilância e tomada de decisão em cenários de interesse estratégico [1].

Nesse contexto, tecnologias como os Radares de Abertura Sintética (SAR) são especialmente úteis em aplicações de monitoramento e defesa, pois permitem a aquisição de imagens independentemente da iluminação solar e mesmo sob condições climáticas adversas [2]. Essa característica é particularmente relevante em regiões tropicais, como a Amazônia, onde a cobertura persistente de nuvens pode limitar o uso de imagens ópticas. Assim, o uso de imagens SAR associado a algoritmos de visão computacional tem ganhado destaque em tarefas de detecção automática de alvos.

Entre as abordagens mais utilizadas em detecção de objetos, destaca-se a família de modelos *You Only Look Once* (YOLO), proposta originalmente por [3]. A YOLO é um método baseado em redes neurais convolucionais cuja principal vantagem é realizar a detecção em uma única passagem pela rede. Conforme sugere o nome, o modelo processa a imagem inteira de forma conjunta, estimando simultaneamente as classes e as localizações dos objetos presentes na cena. Essa característica diferencia a YOLO de métodos baseados em classificação de regiões, nos quais áreas de interesse são inicialmente selecionadas e posteriormente classificadas, tornando o processo mais lento [4].

A aplicação de redes neurais na detecção de alvos em imagens de satélite tem sido investigada em diferentes contextos [5], tanto em imagens SAR quanto no espectro visível. Algumas pesquisas concentram-se na detecção de embarcações em imagens SAR, enquanto outras abordam a identificação de pistas de pouso em regiões remotas em imagens SAR [6] ou ópticas [7]. Esses estudos demonstram o potencial das redes neurais para aplicações de sensoriamento remoto, mas também evidenciam desafios relacionados à generalização, à variabilidade dos alvos e às características específicas de cada domínio de imagem.

Em linha com esses avanços, [8] investigou o uso do YOLOv4 para detecção de embarcações de pequeno porte em imagens SAR da constelação ICEYE. O modelo apresentou bom desempenho em regiões de teste conhecidas, atingindo *F1-score* de 84%. No entanto, sua capacidade de generalização foi limitada: em áreas marítimas fora do conjunto de treinamento, a taxa de falsos positivos ultrapassou 20%, o que destaca a importância da diversidade da base de dados utilizada no treinamento.

O Brasil é um país de dimensões continentais, o que impõe desafios significativos para o monitoramento e a gestão de suas vastas áreas. Um exemplo dessa complexidade é a Amazônia Legal, cuja cobertura territorial abrange mais de 5 milhões de km², correspondendo a cerca de 58,93% do território brasileiro [9]. Essas características demandam iniciativas inovadoras para identificar e monitorar infraestruturas de interesse estratégico, como pistas de pouso não registradas.

Para lidar com esse cenário desafiador, destacam-se iniciativas como o MapBiomas [10], cuja estrutura colaborativa reúne ONGs, universidades, laboratórios e *startups* de tecnologia. Sua principal atividade consiste no levantamento anual da cobertura e uso da terra. Ao empregar a plataforma *Google Earth Engine* (GEE) [11], torna-se viável realizar o pré-processamento e a aquisição de dados de forma automatizada, possibilitando o mapeamento de áreas de interesse, incluindo regiões remotas que abrigam pistas de pouso catalogadas pelo MapBiomas [10].

Embora trabalhos anteriores tenham demonstrado a aplicabilidade de redes neurais à detecção de alvos em imagens SAR

D. K. de Sousa dos Santos, daviddkss@ita.br; M. Habermann, habermann@ita.br; L. de Abreu Faria, lester@ita.br.

e à identificação de pistas de pouso [12], [8], [6], [7], ainda não está claramente estabelecido, no contexto considerado, se descritores derivados por filtros fornecem informações complementares capazes de melhorar a detecção de pistas em imagens SAR do Sentinel-1.

Para investigar essa questão, este trabalho adapta a estratégia de *Filter Augmented Input*, apresentada no SARDet-100K como parte do *Multi-Stage with Filter Augmentation* (MSFA) [13]. Os resultados desse estudo indicaram que *Wavelet Scattering Transform* (WST) e *Histogram of Oriented Gradients* (HOG), quando avaliados individualmente como representações auxiliares, proporcionaram ganhos em relação à entrada baseada apenas na imagem SAR, motivando a seleção desses descritores. Diferentemente da formulação completa do MSFA, o presente trabalho investiga especificamente uma entrada de três canais, composta pela imagem SAR original e pelas representações WST e HOG, aplicada à arquitetura YOLO11x.

O objetivo é avaliar se a combinação das informações multiescalares fornecidas pela WST com as características de forma e orientação extraídas pelo HOG melhora a detecção de pistas de pouso, caracterizadas por sua geometria alongada e predominantemente linear, em comparação com a replicação convencional da imagem SAR nos três canais de entrada da rede YOLO.

A principal contribuição consiste em avaliar experimentalmente essa representação em comparação com a replicação convencional da imagem SAR, mantendo constantes as imagens, anotações, partições e configurações de treinamento. Os experimentos utilizam o conjunto *S1-AAD: Sentinel-1 Amazon AirstripDataset* [14], cuja distribuição geográfica é apresentada na Figura 1.

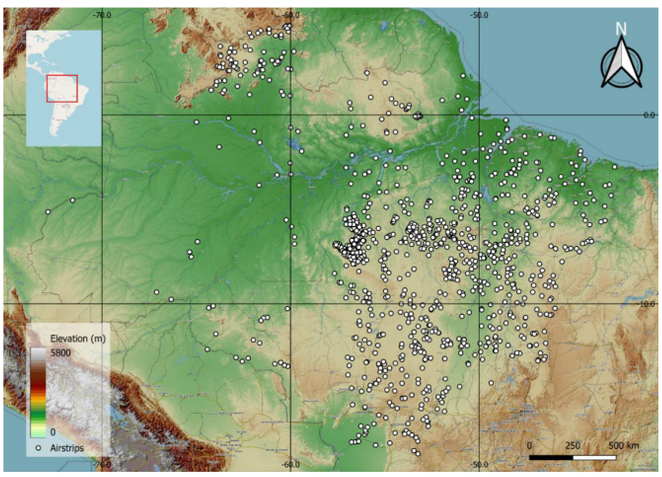


Fig. 1: Distribuição geográfica das 1.040 pistas de pouso disponíveis no conjunto de dados S1-AAD. Fonte: adaptado de [15].

A seção II descreve os métodos utilizados. Na seção III é apresentada a metodologia, bem como as características dos dados e o processo de construção das representações de entrada. Os resultados e discussões são tratados na seção IV. Por fim, na seção V são sintetizadas as principais conclusões do estudo.

II. Métodos

A. You Only Look Once

A detecção de objetos em imagens é uma tarefa central da visão computacional. Entre as abordagens existentes que utilizam redes neurais convolucionais, o método YOLO tem se destacado por realizar, de forma simultânea, a localização e a classificação de objetos a partir de uma única análise da imagem. Em vez de segmentar a imagem em diversas regiões para análise posterior, o YOLO processa a imagem como um todo, o que reduz o tempo de execução e permite a detecção em imagens de forma mais rápida que os métodos tradicionais. Essa característica é particularmente relevante quando se trabalha com grandes volumes de dados, como no monitoramento de áreas extensas por imagens de satélite.

A arquitetura do YOLO é estruturada em três componentes principais. A primeira parte, conhecida como *backbone*, é responsável por extrair as características visuais mais relevantes da imagem. Em seguida, o *neck* combina essas informações para facilitar sua interpretação. Por fim, o *head* realiza as predições, isto é, determina a localização e a classe dos objetos presentes na imagem. Essa arquitetura tem sido refinada ao longo das diferentes versões do YOLO, e cada nova versão introduz mudanças voltadas à melhoria da precisão das detecções e da capacidade de generalização dos modelos [16].

A versão YOLO11x, desenvolvida pela *Ultralytics*, mantém os princípios fundamentais da arquitetura YOLO, mas apresenta avanços técnicos que ampliam sua capacidade de detectar objetos em cenários complexos. Entre as principais atualizações estão a incorporação de mecanismos de atenção que otimizam o foco da rede nas regiões mais informativas da imagem e o refinamento dos blocos de detecção. Essas mudanças adaptam o modelo YOLOv11, cuja arquitetura é capaz de lidar com tarefas mais exigentes e conjuntos de dados com maior variabilidade [17].

B. Filter Augmented Input

Neste trabalho, foi adotada uma estratégia de *Filter Augmented Input* baseada no método proposto no SARDet-100K [13]. No referido trabalho, os autores apresentam o *framework* MSFA, cujo objetivo é reduzir a lacuna de domínio entre imagens naturais, imagens ópticas de sensoriamento remoto e imagens SAR. Uma das etapas centrais do MSFA é o *Filter Augmented Input*, em que a imagem SAR original é concatenada com representações auxiliares extraídas por filtros ou descritores manuais.

De forma geral, dada uma imagem SAR de entrada x, define-se uma transformação $T_i(x)$ responsável por gerar uma representação filtrada M_x^i. Em seguida, a entrada baseada em filtros é formada pela concatenação entre a imagem original e essa representação filtrada, conforme:

$$Inp = concat(x, M_x^i), \tag{1}$$

em que Inp representa a entrada baseada em filtros utilizada pelo detector, x corresponde à imagem SAR original e M_x^i representa a característica extraída por um filtro específico, como HOG ou WST.

C. Wavelet Scattering Transform

O *Wavelet Scattering Transform* é uma técnica de extração de características baseada em transformadas *wavelet*, proposta

por [18]. O método tem como objetivo produzir representações estáveis a pequenas deformações e capazes de preservar informações estruturais da imagem em diferentes escalas.

O WST é formado por uma sequência de operações de convolução com filtros *wavelet*, aplicação de módulo não linear e agregação por suavização. Inicialmente, a imagem é convoluída com um conjunto de *wavelets* em diferentes escalas e orientações. Essas *wavelets* permitem decompor a imagem em componentes que representam padrões locais, como bordas, texturas e variações espaciais. Em seguida, aplica-se o operador módulo, que preserva a energia das respostas e reduz a sensibilidade a pequenas translações. Por fim, uma operação de suavização gera coeficientes de espalhamento, que compõem a representação final da imagem.

Uma característica importante é sua capacidade de extrair informações multiescalares. Isso significa que o método consegue representar tanto estruturas finas quanto padrões mais amplos presentes na imagem. Essa propriedade é especialmente relevante para imagens SAR, que apresentam ruído *speckle* e variações de retroespalhamento. Ao utilizar decomposições *wavelet*, o WST pode preservar informações geométricas e texturais importantes, ao mesmo tempo em que reduz a influência de variações locais indesejadas.

D. Histogram of Oriented Gradients

O *Histogram of Oriented Gradients* é um descritor de características baseado na distribuição local das orientações dos gradientes da imagem. O método foi proposto por [19] com o objetivo de representar a forma e a estrutura de objetos a partir de informações de borda e contraste local.

Inicialmente, a imagem é dividida em pequenas regiões chamadas células. Para cada célula, são calculados os gradientes locais dos pixels, geralmente nas direções horizontal e vertical. A partir desses gradientes, obtêm-se a magnitude e a orientação de cada ponto da imagem. Em seguida, as orientações são agrupadas em histogramas, nos quais cada intervalo angular acumula contribuições ponderadas pela magnitude do gradiente. Dessa forma, regiões com bordas ou transições intensas contribuem mais para a representação final.

Após a construção dos histogramas nas células, o HOG realiza uma etapa de normalização em blocos, compostos por grupos de células vizinhas. Essa normalização reduz a influência de variações locais de iluminação, contraste e intensidade, tornando o descritor mais robusto. O vetor final de características é formado pela concatenação dos histogramas normalizados ao longo da imagem.

E. Métricas de avaliação

Para avaliar o desempenho dos modelos, foram utilizadas as métricas de *Precision* (P), *Recall* (R) e *mean Average Precision* (mAP), amplamente adotadas em tarefas de detecção de objetos. Essas métricas se baseiam na detecção ou não do alvo, considerando *True Positive* (TP), *False Positive* (FP) e *False Negative* (FN) [20], [21].

- *Precision* – Indica a proporção de predições positivas que foram, de fato, corretas. Em outras palavras, mede a exatidão das predições positivas do classificador [22], [23].
- *Recall* – Também chamada de sensibilidade ou taxa de verdadeiros positivos, mede a capacidade do modelo de identificar corretamente todas as instâncias positivas presentes [22], [23].
- *Mean Average Precision* – É uma métrica amplamente utilizada para avaliar modelos de detecção de objetos, pois considera tanto a precisão da classificação quanto a localização espacial dos objetos detectados. Para que uma predição seja considerada correta, é necessário que o modelo acerte a classe e que a *Intersection over Union* (IoU) entre a área prevista e a área real ultrapasse um limiar estabelecido. Essa abordagem considera não apenas se o objeto foi detectado, mas também a precisão de sua localização [21], [22], [23].

III. Metodologia

A. Base de Dados

Neste trabalho, utilizou-se o *Sentinel-1 Amazon Airstrip Dataset* [14], [15], uma base de dados SAR voltada para tarefas de detecção, segmentação semântica e detecção de mudanças em pistas de pouso localizadas na Amazônia brasileira. A base foi construída a partir de coordenadas de pistas catalogadas pelo projeto MapBiomas [10]. Do total de 2.869 pistas inicialmente catalogadas, foram selecionadas 1.040 localizações com disponibilidade de imagens Sentinel-1 adequadas na plataforma GEE.

As imagens foram obtidas a partir do sensor SAR em banda C do satélite Sentinel-1, utilizando produtos *Ground Range Detected* (GRD), modo de aquisição *Interferometric Wide* (IW) e polarização VV. A maior parte das imagens corresponde ao período de 2021 a 2024. Para cada coordenada de pista, foi extraído um recorte espacial de 200×200 pixels, correspondente a uma área aproximada de 4 km^2 centralizada na localização da pista. Essa dimensão foi adotada pelos autores da base para garantir que pistas com extensão inferior a aproximadamente 2 km estivessem contidas no recorte.

Segundo [15], as imagens Sentinel-1 disponíveis no GEE já passam por uma cadeia de pré-processamento padrão para produtos GRD, incluindo aplicação do arquivo de órbita, remoção de ruído de borda e ruído térmico, calibração radiométrica e correção de terreno com ortorretificação. Dessa forma, os dados utilizados preservam informações radiométricas e geométricas adequadas para aplicações de sensoriamento remoto com imagens SAR.

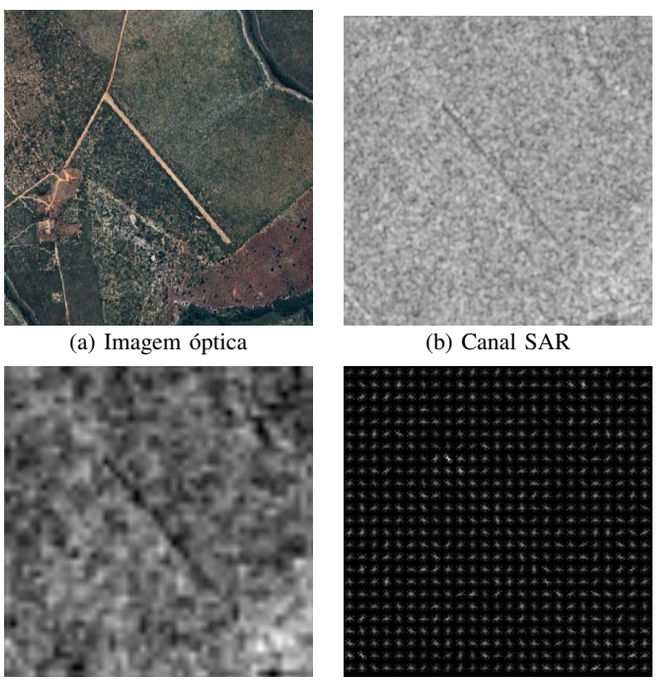
A base é disponibilizada em diferentes formatos, contemplando imagens SAR, máscaras de segmentação, anotações para detecção de objetos e arquivos geoespaciais auxiliares. O conjunto principal está no formato GeoTIFF, com uma única banda, tipo *Float64*, e dimensões de 200×200 pixels. Esses arquivos representam a fonte radiométrica, pois preservam os valores originais de retroespalhamento σ^0. Além disso, a base também disponibiliza versões das mesmas imagens em PNG de 8 bits, formato adaptado para o uso em redes neurais de detecção de objetos, como a família YOLO.

B. Preparação dos Dados para os Experimentos

Embora o S1-AAD disponibilize imagens SAR em GeoTIFF e PNG, neste experimento foi utilizada uma versão processada a partir dos dados GeoTIFF para compor a entrada da rede neural. A inspeção dos arquivos GeoTIFF indicou que os valores de retroespalhamento já estavam representados em escala logarítmica.

Para reduzir a influência de valores extremos, os valores em dB foram limitados entre os percentis 0,1% e 99,95% da distribuição dos pixels do conjunto de treinamento. Após esse recorte, os valores foram normalizados para o intervalo de 0 a 255 e convertidos para PNG de 8 bits. Como a imagem SAR resultante possui apenas um canal, esse mesmo canal foi replicado nos canais R, G e B, produzindo uma entrada compatível com a YOLO. Dessa forma, a imagem utilizada pela rede neural não corresponde a uma imagem óptica RGB, mas a uma imagem SAR, normalizada por percentis, convertida para PNG de 8 bits e replicada em três canais.

A partir dessa imagem SAR, foi gerada uma segunda condição experimental inspirada na estratégia de *Filter Augmented Input* do SARDet-100K. A imagem SAR foi mantida como primeiro canal de entrada, enquanto os outros dois canais foram substituídos por representações derivadas da própria imagem SAR: uma obtida pelo *Wavelet Scattering Transform* e outra pelo *Histogram of Oriented Gradients*. A Figura 2 apresenta um exemplo visual da diferença entre uma imagem óptica, SAR e as representações derivadas utilizadas na composição *Filter Augmented Input* proposta.



(a) Imagem óptica (b) Canal SAR

(c) Canal WST (d) Canal HOG

Fig. 2: Representações da mesma região contendo uma pista de pouso: (a) imagem óptica; (b) imagem SAR; (c) representação obtida pela WST; e (d) representação obtida pelo HOG. Fonte: imagem óptica adaptada de [15]; demais representações elaboradas pelos autores a partir dos dados do S1-AAD.

As caixas delimitadoras originais foram preservadas sem qualquer alteração geométrica. Como as transformações WST e HOG modificam apenas a representação radiométrica da imagem, e não sua geometria espacial, as coordenadas normalizadas das anotações YOLO permanecem válidas. Dessa forma, os dois conjuntos experimentais possuem exatamente os mesmos identificadores de imagens, os mesmos subconjuntos de treinamento e validação e as mesmas anotações, diferindo apenas na composição dos canais de entrada fornecidos à rede.

O conjunto de dados foi dividido em dois subconjuntos mutuamente exclusivos: treinamento e validação. Foi mantida uma proporção de aproximadamente 80% e 20%, respectivamente. Desse modo, o conjunto de treinamento contém 832 imagens e o conjunto de validação contém 208 imagens. Para preservar mais amostras no treinamento e na validação, não foi reservado um subconjunto independente de teste nem utilizado um conjunto externo. A mesma divisão foi preservada em todos os experimentos, evitando vazamento de dados e garantindo que as comparações entre as diferentes representações de entrada fossem realizadas sobre as mesmas amostras.

C. Treinamento – YOLO11x

O treinamento das redes foi conduzido utilizando Python 3.12.3, PyTorch 2.5.1 com suporte CUDA 12.1 e a biblioteca *Ultralytics* 8.4.56. A infraestrutura computacional utilizada dispõe de duas GPUs NVIDIA L40S. Neste estudo, foi adotada a arquitetura YOLO11x para a tarefa de detecção de objetos, utilizando pesos pré-treinados diponibilizados no arquivo `yolo11x.pt` [17].

A estratégia de *transfer learning* foi empregada com o objetivo de aproveitar representações iniciais aprendidas em larga escala, especialmente padrões de bordas, texturas e formas, que podem ser úteis mesmo em domínios distintos, como imagens SAR [24].

O experimento foi formulado exclusivamente como uma tarefa de detecção, utilizando as *bounding boxes* disponíveis no conjunto de dados. Para assegurar uma comparação controlada, os mesmos hiperparâmetros foram utilizados em todos os treinamentos. O treinamento foi configurado para um máximo de 300 épocas, com *early stopping* definido por `patience = 100`. Foram utilizadas três *seeds* diferentes, `42`, `123` e `321`, permitindo avaliar a estabilidade dos resultados frente à variabilidade estocástica do treinamento. Além disso, o parâmetro `deterministic = True` foi adotado para favorecer a reprodutibilidade dos experimentos.

A resolução de entrada foi definida como `imgsz = 224`, valor próximo à dimensão original das imagens do conjunto de dados, aproximadamente 200×200 pixels, e compatível com o processamento da arquitetura YOLO. As imagens foram tratadas como problema de classe única por meio de `single_cls = True`, uma vez que o objetivo do estudo é detectar apenas pistas de pouso. As métricas de desempenho foram calculadas no modo padrão de validação da YOLO, sem fixação manual do limiar de confiança durante o cálculo do mAP.

Os principais hiperparâmetros utilizados durante o treinamento são apresentados na Tabela I, enquanto as augmentações geométricas adotadas são detalhadas na Tabela II. As augmentações de cor e de composição de imagens foram explicitamente desativadas para preservar a interpretação radiométrica das imagens SAR e isolar o efeito das representações de entrada avaliadas.

IV. Resultados e Discussões

Nesta seção, são apresentados os resultados da comparação entre a entrada SAR replicada em três canais e a abordagem de *Filter Augmented Input* com SAR+WST+HOG. A Tabela III

TABELA I: Configuração geral dos hiperparâmetros utilizados no treinamento da YOLO11x.

Parâmetro	Valor	Descrição
`model`	`yolo11x.pt`	Pesos pré-treinados.
`task`	`detect`	Detecção de objetos.
`epochs`	300	Número máximo de épocas.
`patience`	100	Critério de *early stopping*.
`imgsz`	224	Resolução de entrada.
`batch`	16	Imagens por lote.
`single_cls`	`True`	Classe única.
`deterministic`	`True`	Reprodutibilidade.
`pretrained`	`True`	Pesos pré-treinados.
`seed`	42, 123, 321	Sementes fixas.
`workers`	8	Carregamento dos dados.

TABELA II: Augmentações geométricas utilizadas no treinamento da YOLO11x.

Parâmetro	Valor	Descrição
`degrees`	10.0	Aplica rotações aleatórias de até ± 10 graus, simulando variações de orientação das pistas nas imagens SAR.
`fliplr`	0.5	Aplica inversão horizontal com probabilidade de 50%.
`flipud`	0.5	Aplica inversão vertical com probabilidade de 50%.

apresenta, para cada treinamento, os resultados do *checkpoint* correspondente à época com maior mAP50 no conjunto de validação. São reportados o número total de épocas executadas, a época selecionada e as respectivas métricas *Precision*, *Recall*, mAP50 e mAP50-95. Como o mesmo conjunto de validação foi utilizado para selecionar o *checkpoint* e calcular as métricas reportadas, os resultados devem ser interpretados como uma comparação interna entre as configurações experimentais, e não como estimativas independentes da capacidade de generalização.

TABELA III: Melhores resultados por *seed*. *Filter Augmented Input* (SAR+WST+HOG); T/M: número total de épocas/melhor época.

Entrada	Semente	Ép. T/M	P	R	mAP50	mAP50-95
SAR	42	228/128	0,414	0,402	**0,359**	**0,117**
	123	198/98	0,314	0,321	0,283	0,097
	321	300/228	0,318	**0,388**	0,266	0,085
WST+HOG+SAR	42	300/167	**0,423**	**0,411**	0,355	0,088
	123	213/184	**0,423**	**0,491**	**0,413**	**0,132**
	321	300/191	**0,481**	0,360	**0,335**	**0,096**

Conforme apresentado na Tabela III, a entrada SAR+WST+HOG obteve mAP50 superior em duas das três sementes avaliadas. Nas sementes `123` e `321`, essa métrica aumentou, respectivamente, de 0,283 para 0,413 e de 0,266 para 0,335. Na semente `42`, os desempenhos foram semelhantes, com valores de 0,359 para a entrada SAR e 0,355 para a entrada aumentada. A tabela também mostra que o maior mAP50 da configuração SAR+WST+HOG foi atingido nas épocas 167, 184 e 191, enquanto a referência alcançou seus melhores valores nas épocas 128, 98 e 228. Assim, a entrada aumentada atingiu seu melhor desempenho mais tarde em duas das três sementes, com média aproximada de 181 épocas, em comparação com 151 da referência. Esse comportamento pode estar relacionado à necessidade de uma adaptação mais gradual da rede a canais com distribuições e conteúdos distintos, em contraste com os três canais idênticos da entrada SAR.

TABELA IV: Resumo médio das métricas obtidas para as *seeds* 42, 123 e 321.

Método	Precision	Recall	mAP50	mAP50-95
SAR	0,348	0,370	0,302	0,099
WST+HOG+SAR	**0,442**	**0,420**	**0,367**	**0,105**

Para sintetizar o comportamento das configurações experimentais, a Tabela IV apresenta as médias das métricas obtidas nas três sementes. A entrada SAR+WST+HOG apresentou médias superiores em todas as métricas. Os ganhos foram de 9,37 pontos percentuais em *Precision*, 5,04 em *Recall*, 6,53 em mAP50 e 0,52 em mAP50-95. Os aumentos mais expressivos em *Precision* e mAP50 sugerem que as representações auxiliares favoreceram principalmente a diferenciação entre as pistas e o fundo da imagem, possivelmente reduzindo detecções falsas e aumentando a identificação dos alvos no limiar de IoU igual a 0,5.

Uma possível explicação para esses ganhos é a complementaridade entre os canais empregados. A WST fornece informações estruturais em diferentes escalas e pode reduzir a influência de variações locais e do ruído *speckle*, enquanto o HOG enfatiza contornos e orientações de gradientes, características relevantes para a geometria alongada das pistas de pouso. Entretanto, o ganho de apenas 0,52 ponto percentual em mAP50-95, substancialmente menor que o observado em mAP50, indica que a melhoria foi limitada quando se exigiu maior precisão na localização das caixas delimitadoras.

Para tornar a comparação entre as duas configurações de entrada mais explícita, a Tabela V apresenta a diferença pareada entre o *Filter Augmented Input* e a imagem SAR para cada *seed*. Cada valor corresponde à diferença entre o resultado obtido com o *Filter Augmented Input* e aquele obtido com a imagem SAR para a mesma semente. Dessa forma, valores positivos indicam vantagem do *Filter Augmented Input*, enquanto valores negativos indicam vantagem da imagem SAR.

TABELA V: Diferença pareada entre o *Filter Augmented Input* e a imagem SAR por *seed*, em pontos percentuais.

Seed	Δ Precision	Δ Recall	Δ mAP50	Δ mAP50-95
42	**+0,85**	**+0,93**	-0,37	-2,92
123	**+10,94**	**+16,98**	**+13,01**	**+3,42**
321	**+16,32**	-2,79	**+6,93**	**+1,07**
Média	**+9,37**	**+5,04**	**+6,53**	**+0,52**

A análise pareada da Tabela V confirma que os ganhos médios foram positivos, mas também evidencia sua variabilidade. Na semente `123`, a entrada aumentada apresentou ganhos em todas as métricas, incluindo 13,01 pontos percentuais em mAP50 e 3,42 em mAP50-95. Na semente `321`, o aumento de 6,93 pontos percentuais em mAP50 foi acompanhado por uma redução de 2,79 em *Recall*, sugerindo detecções mais seletivas. Na semente `42`, por outro lado, foram observadas reduções de 0,37 ponto percentual em mAP50 e 2,92 em mAP50-95.

Em relação ao SARDet-100K, os resultados são consistentes com a conclusão geral de que representações produzidas por descritores manuais podem beneficiar a detecção em imagens SAR [13]. Naquele estudo, as entradas SAR+WST e SAR+HOG superaram individualmente a entrada baseada apenas em SAR, embora a combinação de vários descritores não tenha produzido ganho significativo sobre o WST isolado.

O presente trabalho complementa essas evidências ao aplicar o princípio de *Filter Augmented Input* à YOLO11x e à detecção de pistas de pouso. Contudo, como WST e HOG foram avaliados somente em conjunto, então não foi possível determinar se os ganhos observados decorrem de um dos descritores ou da interação entre ambos. Portanto, os resultados sustentam a utilidade da entrada adaptada neste problema específico.

V. Conclusões

Este trabalho avaliou uma estratégia de *Filter Augmented Input* para detectar pistas de pouso em imagens SAR do Sentinel-1 com a YOLO11x. Foram comparadas uma imagem SAR replicada nos três canais e uma entrada composta por SAR, WST e HOG, mantendo-se as mesmas configurações de treinamento e *seeds* experimentais.

O *Filter Augmented Input* apresentou médias superiores em *Precision*, *Recall*, mAP50 e mAP50-95, embora a vantagem não tenha sido uniforme entre as *seeds*. Os valores absolutos sugerem que a tarefa permanece desafiadora, possivelmente em razão da resolução espacial, do ruído *speckle* e da semelhança entre estruturas lineares. Assim, os resultados constituem evidência interna de melhoria relativa, e não uma validação mais abrangente.

Entre as limitações, destacam-se o tamanho reduzido do conjunto de dados, a centralização das pistas nos recortes e a ausência de exemplos negativos. Essas características podem restringir a generalização, introduzir viés espacial e limitar a avaliação de falsos positivos. Como não foi realizado teste em um conjunto independente, as métricas não estimam diretamente o desempenho em áreas não observadas durante o desenvolvimento.

Como trabalhos futuros, recomenda-se ampliar e diversificar o conjunto de dados, incorporando pistas em diferentes posições nos recortes e exemplos negativos potencialmente confundíveis, como clareiras, estradas e rios. Também deverão ser investigadas outras arquiteturas e combinações de filtros, além de técnicas de redução de *speckle*, ajustes radiométricos e estratégias de pré-treinamento em bases SAR de maior escala. Por fim, avaliações em conjuntos externos e análises estatísticas mais abrangentes poderão fornecer evidências mais robustas sobre a generalização da abordagem.

Referências

[1] Força Aérea Brasileira, "Diretriz de Implantação do Programa Estratégico de Sistemas Espaciais," Força Aérea Brasileira, Brasília, Brasil, Tech. Rep. DCA 358-1, 2020.

[2] K. Ouchi, "Recent Trend and Advance of Synthetic Aperture Radar with Selected Topics," *Remote Sensing*, vol. 5, pp. 716–807, 2013.

[3] R. G. Joseph Redmon, Santosh Divvala and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *arXiv*, jun 2015. [Online]. Available: https://doi.org/10.48550/arXiv.1506.02640

[4] D. H. C, V. R. Shetty, and G. N. Srinivasan, "An Overview of You Only Look Once: Unified, Real-Time Object Detection," *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, vol. 8, no. VI, jun 2020, available at www.ijraset.com; ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429.

[5] L. E. Falqueto, R. L. Paes, and A. Passaro, "KNN e Rede Neural Convolucional para o Reconhecimento de Plataformas de Petróleo em Imagens SAR do Sentinel-1," *Spectrum – Aplicações Operacionais em Áreas de Defesa*, vol. 1, pp. 29–33, 2023.

[6] L. S. Gomes, E. H. Shigemori, T. N. Kuck, and D. I. Alves, "Evaluation of a YOLOv8-Based Method for Detecting Unauthorized Airstrips in the Amazon Rainforest using SAR Imagery," *Anais do XLI Brazilian Symposium on Telecommunications and Signal Processing (SBrT)*, oct 2024.

[7] G. R. Pardini, P. M. Tasinaffo, E. H. Shigemori, T. N. Kuck, M. R. Q. A. Maximo, and W. R. Gyotoku, "Improved Algorithm to Detect Clandestine Airstrips in Amazon RainForest," *Algorithms*, vol. 18, p. 102, 2025.

[8] H. C. T. de Oliveira, "Detecção automática de pequenas embarcações em imagens SAR da constelação ICEYE com uso de uma rede neural convolucional," São José dos Campos, Brasil, p. 84, 2021.

[9] IBGE. (2025) Amazônia Legal: o que é. [Online]. Available: https://www.ibge.gov.br/geociencias/organizacao-do-territorio/estrutura-territorial/15819-amazonia-legal.html?=&t=o-que-e

[10] MapBiomas. (2025) O Projeto. [Online]. Available: https://brasil.mapbiomas.org/o-projeto/

[11] Google Earth Engine. (2025) Google Earth Engine. [Online]. Available: https://earthengine.google.com/

[12] L. Pang, B. Li, F. Zhang, X. Meng, and L. Zhang, "A Lightweight YOLOv5-MNE Algorithm for SAR Ship Detection," *Sensors*, vol. 22, p. 7088, 2022.

[13] Y. Li, X. Li, W. Li, Q. Hou, L. Liu, M.-M. Cheng, and J. Yang, "SARDet-100K: Towards open-source benchmark and toolkit for large-scale SAR object detection," in *Proceedings of the 38th Conference on Neural Information Processing Systems*. Neural Information Processing Systems, 2024. [Online]. Available: https://arxiv.org/abs/2403.06534

[14] L. d. S. Gomes, G. H. d. Q. Stabile, T. N. Kuck, F. A. P. d. Figueiredo, E. H. Shiguemori, and D. I. Alves, "S1-aad: Sentinel-1 amazon airstrip dataset," 2025. [Online]. Available: https://data.mendeley.com/datasets/x7rn78ymtn/1

[15] ——, "A sentinel-1 sar imagery dataset for airstrips detection and segmentation in the brazilian amazon rainforest," *Data in Brief*, vol. 65, p. 112472, 2026. [Online]. Available: https://doi.org/10.1016/j.dib.2026.112472

[16] P. Hidayatullah, N. Syakrani, M. R. Sholahuddin, T. Gelar, and R. Tubagus, "YOLOv8 to YOLO11: A Comprehensive Architecture In-depth Comparative Review," Preprint, Kab. Bandung Barat and Kota Cimahi, Indonesia, 2025.

[17] Ultralytics. (2025) YOLOv11 Model Reference. [Online]. Available: https://docs.ultralytics.com/models/yolo11/

[18] S. Mallat, "Group invariant scattering," *Communications on Pure and Applied Mathematics*, vol. 65, no. 10, pp. 1331–1398, 2012. [Online]. Available: https://doi.org/10.1002/cpa.21413

[19] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 2005, pp. 886–893. [Online]. Available: https://doi.org/10.1109/CVPR.2005.177

[20] C. H. D. Felice, E. H. Shigemori, R. L. Paes, and R. Suterio, "Análise da interação entre o tamanho do alvo e desempenho na detecção automática de embarcações em imagens SAR utilizando YOLOv8 modificada," *Instituto Tecnológico de Aeronáutica*, 2024.

[21] J. Terven, D. M. Cordova-Esparza, J. A. Romero-González *et al.*, "A Comprehensive Survey of Loss Functions and Metrics in Deep Learning," *Artificial Intelligence Review*, vol. 58, p. 195, 2025, accepted: 13 March 2025; Published: 11 April 2025.

[22] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, Inc., 2019, second edition; printed in Canada.

[23] R. Padilla, W. L. Passos, T. L. B. Dias, S. L. Netto, and E. A. B. da Silva, "A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit," *Electronics*, vol. 10, p. 279, 2021.

[24] L. Shao, F. Zhu, and X. Li, "Transfer Learning for Visual Categorization: A Survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 5, may 2015.