

# Análise da interação entre o tamanho do alvo e desempenho na detecção automática de embarcações em imagens SAR utilizando YOLOv8 modificada

Caio Henrique De Felice<sup>1</sup>, Elcio Hideiti Shiguemori<sup>2</sup>, Rafael Lemos Paes<sup>3</sup>, Ricardo Suterio<sup>1</sup>

<sup>1</sup>Instituto Tecnológico de Aeronáutica (ITA), São José dos Campos/SP - Brasil

<sup>2</sup> Instituto de Estudos Avançados (IEAv), São José dos Campos/SP - Brasil

<sup>3</sup>Estado Maior da Aeronáutica (EMAER), Brasília/DF - Brasil

**Resumo**— A detecção de embarcações é de suma importância para monitoramento das águas brasileiras. Entretanto, a extensão territorial faz com que tal tarefa seja repetitiva e demorada. Nesse contexto, técnicas de detecção automática baseadas em aprendizado de máquina se mostram viáveis, ainda mais quando atreladas a imagens SAR, que possibilitam imageamento a quaisquer hora e condição meteorológica. Este trabalho propõe a utilização de um design de experimento fatorial  $2^k$ , visando analisar, os impactos do tamanho do alvo e da utilização de um mecanismo de atenção em relação ao desempenho da YOLOv8 na detecção de embarcações utilizando o banco de dados HRSID. Conclui-se que o mecanismo de atenção empregado levou tanto a aumento na detecção de alvos pequenos quanto de falsos positivos para ambos os tamanhos, refletindo nas métricas de precisão, recall e F1-Score independente do nível de confiança analisado.

**Palavras-Chave**— Imagens SAR, Inteligência Artificial, Experimento Fatorial

## I. INTRODUÇÃO

A Estratégia Nacional de Defesa (END) afirma que as Forças Armadas Brasileiras devem ser capazes de monitorar as Águas Jurisdicionais Brasileiras (AJB) além do território e do espaço aéreo. Ainda, as dimensões continentais do Brasil evidenciam a necessidade de maiores eficiência e eficácia no emprego dos mais diversos meios e recursos [1], [2].

Tratando especificamente do ambiente aquático, seja marítimo ou fluvial, o sensoriamento remoto através de radares de abertura sintética (SAR) aparece como um método importante de monitoramento e detecção de embarcações, seja para aplicações civis ou militares. Essa evidência se dá pela sua capacidade de operar independente da luz do sol ou das condições meteorológicas [3].

Entretanto, a análise visual de muitas imagens, decorrentes da extensão territorial brasileira, pode se tornar inviável quando se deseja monitorar atividades de embarcações, por exemplo. Nesse sentido, a utilização técnicas de detecção automatizadas mostram-se mais viáveis, sendo o aprendizado profundo (DL, do inglês *deep learning*) largamente utilizado devido à sua capacidade de aprender características relevantes dos dados [4].

Diversas abordagens das mais diversas naturezas foram propostas, no âmbito da aplicação do DL para detecção de embarcações em imagens SAR, assim como abordado por [3]. As mais tradicionais envolvem a segmentação de terreno

e mar, detecção através de *constant false alarm rate* (CFAR) e discriminação. Entretanto, ainda segundo o autor, surgiram novas abordagens a partir da disponibilização de bancos de dados públicos. Com esse recurso, a reprodutibilidade em novas pesquisas se tornou possível, sendo o ponto de inflexão o *SAR Ship Detection Dataset* (SSDD) em 2017 [3].

A detecção de embarcações nem sempre ocorre no mesmo cenário. Cenários mais complexos (tratados neste trabalho como imagens onde não estão presentes apenas embarcações e o mar como plano de fundo), e a presença de alvos de variados tamanhos, posições e aspectos trouxeram novas dificuldades e novas propostas de aplicações como, por exemplo [5], [4], [6]. Um dos aspectos que se sobressaem diz respeito à dimensão dos alvos que se deseja detectar, mais especificamente, alvos ditos "pequenos".

Inicialmente, é necessário entender a metodologia utilizada para classificá-los dessa maneira. As dimensões de um alvo dito "pequeno" normalmente são apresentadas em função das dimensões totais da imagem na qual está presente: [7] exemplifica um alvo com dimensões 20 x 20 pixels numa imagem aérea como pequeno, [6] propõe a utilização de uma razão entre as dimensões da caixa que delimita o objeto de interesse (*bouding box*) e da imagem assim como utilizado por [8] e [9].

Aplicações relacionadas à detecção de alvos pequenos apresentam dificuldades adicionais: a incapacidade de distinguir suas características, baixa resolução, *backgrounds* complexos e informações contextuais limitadas [7], [4]. [7] fornece uma visão geral das abordagens utilizadas para lidar com alvos dessas dimensões, entretanto com um enfoque em imagens ópticas. [3] apresenta um histórico relacionado à detecção de embarcações em imagens SAR e cita a necessidade de trabalhos inovadores na detecção de alvos pequenos. [4] apresenta um algoritmo baseado numa rede YOLOv8 modificada para detecção de alvos pequenos, novamente com ênfase em imagens ópticas. [6], por sua vez, apresenta uma arquitetura combinada de redes e o emprego de um mecanismo de atenção com vistas a melhorar a detecção de embarcações de pequeno porte em imagens SAR, porém com um enfoque em desempenho e velocidade. [10] apresenta um mecanismo de atenção com potencial de melhorar a detecção de alvos pequenos. Entretanto, não são apresentadas análises diretas sobre o impacto dos mecanismos de atenção no desempenho

de acordo com o tamanho dos alvos, mas em bancos de dados como um todo.

Neste trabalho é proposto um modelo de experimento fatorial  $2^k$ , visando explorar, além dos efeitos independentes, os efeitos conjuntos de dois fatores em uma determinada resposta através da variação em dois níveis [11]. Os fatores explorados são: o tamanho do alvo e presença ou não do mecanismo de atenção ECA (*Efficient Channel Attention*), que pode melhorar a eficiência geral e desempenho da rede para detecção de alvos pequenos [10]. A arquitetura base da rede neural utilizada é a YOLOv8m (*You Only Look Once* em sua oitava versão e tamanho médio) de [12]. Emprega-se a YOLOv8 original e a arquitetura com adição de 4 módulos ECA, conforme proposto por [13].

Cada arquitetura avaliada foi treinada e avaliada objetivando analisar o impacto do ECA no desempenho de acordo com métricas relevantes à detecção de objetos como precisão, *recall* [14] e *F1-Score* [15] bem como suas significâncias estatísticas através da Análise de Variância (ANOVA) proposta em [11], constatando o impacto da variação do tamanho do alvo, da inserção do ECA na arquitetura da rede bem como a interação entre ambos.

Por fim, é realizada uma análise exploratória de diferentes níveis de confiança para as detecções buscando diminuir a quantidade de indicações de objetos que não são alvos, mas foram considerados como "falsos positivos".

Na seção II são apresentadas informações sobre o *dataset* utilizado. Já na seção III são apresentados a metodologia e materiais. Os resultados e discussões são apresentados na seção IV e, por fim, a conclusão na seção V.

## II. Dataset

A base de dados utilizada neste experimento é composta das imagens e anotações provenientes do *High Resolution SAR Images Dataset* (HRSID) de [8]. Tal escolha se justifica na pretensão futura de explorar a utilização de *datasets* públicos para aplicações com sensores cujos dados não são obtidos de maneira gratuita, como o ICEYE, por exemplo. As resoluções disponibilizadas pelo HRSID abarcam a resolução pretendida para o modo de imageamento *strip* do ICEYE [16] a ser estudado futuramente.

O banco de dados utilizado dispõe de 99 imagens provenientes do satélite Sentinel-1B, 36 provenientes do TerraSAR-X e 1 do TanDEM-X, com as polarizações variando entre VV, HH e VH e a resolução entre 0,5, 1 e 3 metros.

Em sua composição original, são feitos recortes nas imagens satelitais de maneira que cada recorte resulte em imagens menores com dimensões iguais a 800 x 800 pixels. Como resultado, totalizam 5604 imagens com 16951 embarcações. Esses dados se encontram disponíveis ao público já divididos nos grupos de treino e teste. A proporção da divisão é de 65% para treinamento e 35% para testes levando em consideração a quantidade de embarcações por dimensão [8], [3]. Os critérios relacionados ao tamanho do alvo são descritos a seguir.

Utilizando o mesmo critério de [9], [8] classifica os alvos de acordo com a área da *bounding box* em relação a dimensão da imagem, sendo os alvos pequenos (menores que 32 x 32 pixels), médios (entre 32 x 32 e 96 x 96 pixels) e grandes (maiores que 96 x 96 pixels). Nesse critério, o *dataset* possui a composição nas seguintes proporções: 54,5% embarcações

pequenas, 43,5% médias e 2% grandes, todas atribuídas a uma única classe.

Neste experimento foram utilizadas as anotações disponíveis no formato *bounding box* vertical, ou seja, estão presentes nas anotações os dados sobre a classe, coordenadas do seu centro, largura e altura. Os dados sobre coordenadas e dimensões são normalizados de acordo com as dimensões das imagens.

A Fig. 1 exibe um mosaico com quatro amostras de imagens presentes no *dataset* com fins ilustrativos.

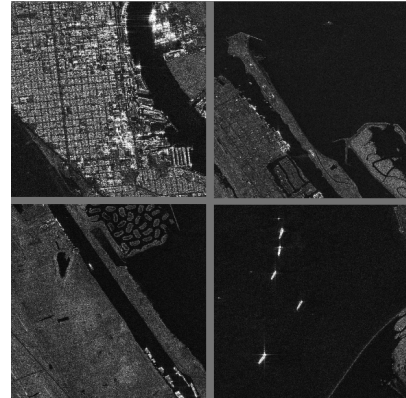


Fig. 1: Mosaico com amostras de imagens do *dataset* HRSID.

## III. MATERIAIS E EXPERIMENTO

### A. YOLOv8

A arquitetura utilizada na rede neural YOLO se mostra eficiente e precisa na detecção de objetos. Desde sua criação, em 2015, vem sofrendo melhorias e novas versões são lançadas. A ideia principal da rede consiste no emprego de *fully convolutional layers* (FCN) ao invés de *dense layers*, trazendo mais eficiência e dando mais versatilidade na utilização de diferentes tamanhos de imagem [15].

Sua versatilidade e histórico de emprego na mesma tarefa de detecção de embarcações em imagens SAR ([3], [4], [6], [5], [13]) foram os motivos que levaram à sua escolha neste experimento. A versão escolhida, oitava, era a mais atual disponível à época.

### B. Módulo ECA

Um módulo ou mecanismo de atenção faz basicamente com que a rede neural dê menos importância a determinadas informações para focar em outras mais importantes. É possível que esse mecanismo atue nos domínios do espaço, dos canais, das camadas ou uma mistura deles [3]. Um desses mecanismos é o *Efficient Channel Attention* (ECA).

O módulo ECA (Fig. 2) é focado na atenção entre os canais, visando destacar características relevantes com base nos canais vizinhos [10].

Sua atuação ocorre da seguinte forma: inicialmente é aplicado um *Global Average Pooling* (GAP) ao *input* de dimensões  $W \times H \times C$  (largura, altura e canais respectivamente). Enquanto as dimensões  $W$  e  $H$  são resultados dos redimensionamentos sofridos pela imagem nos processos anteriores da rede (convoluções e *poolings*, por exemplo),  $C$  representa a quantidade de filtros aplicada até aquele momento.

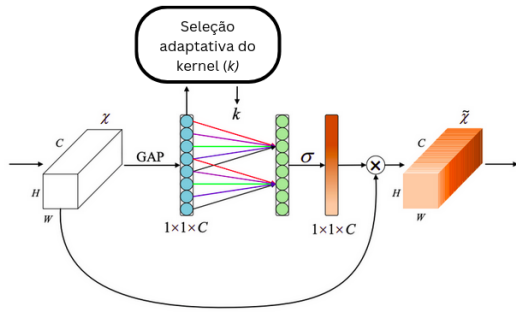


Fig. 2: Módulo ECA. Fonte: Adaptado de [10]

O GAP computa um valor médio para cada um dos mapas de feições [15]. Esse conjunto de valores, sendo da mesma quantidade de mapas ( $C$ ) evita a perda de dimensionalidade e, conseqüentemente, perda de informação [10].

A seguir, é calculada a interação entre um canal e seus  $k$  vizinhos mais próximos, por exemplo, para um  $k = 3$ , são considerados o canal em voga, seu vizinho anterior e o seguinte. O valor  $k$  é, então, ajustado de maneira adaptativa de acordo com a interação entre eles. O ajuste se faz necessário pois canais com maiores dimensões apresentam interações mais profundas e o oposto para menores dimensões. Por fim, são gerados pesos com base nas interações calculadas, que são multiplicados em cada um dos canais originais, realçando os mais importantes [10], [13]. Essa aplicação, além de tornar a rede mais eficiente, pode melhorar seu desempenho em detecção de alvos pequenos [10], melhora essa a ser explorada no presente experimento.

Os módulos de atenção, então, foram incorporados à YOLOv8 conforme exibido na Fig. 3. A arquitetura original consiste, basicamente, na mesma estrutura, porém sem os módulos ECA. O funcionamento e os detalhes de cada um dos elementos da YOLOv8 podem ser encontrados em [12]. O modelo exibido na Fig. 3 segue o padrão utilizado por [13].

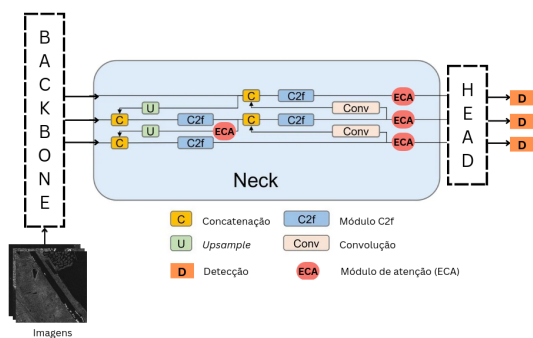


Fig. 3: Arquitetura da YOLOv8 + ECA. Fonte: Adaptado de [13]

### C. Design do experimento

Designs fatoriais são amplamente utilizados em experimentos nos quais se deseja entender os efeitos conjuntos de fatores para uma determinada resposta. Um dos casos envolve o impacto de  $k$  fatores, sejam eles quantitativos ou qualitativos, cada um em dois níveis ("alto" e "baixo", "presente" ou "ausente", por exemplo), tal caso é denominado

design fatorial  $2^k$ . Nessa modalidade, assume-se que a resposta será aproximadamente linear no intervalo entre os níveis selecionados [11].

A utilização de designs experimentais relacionados a redes neurais já se mostrou eficiente em experimentos como o realizado por [17] para otimização de hiperparâmetros.

Neste trabalho, foi empregado um design  $2^2$  com rodadas aleatórias. Foram adotados como fatores a presença ou não de módulos de atenção incorporados à arquitetura da YOLOv8 (fator qualitativo) e o tamanho dos alvos nas imagens (fator quantitativo) e a interação entre ambos, objetivando explorar os efeitos desses fatores nas métricas de desempenho (resposta).

A categorização do tamanho dos alvos se deu através da relação ( $r$ ) entre as dimensões das *bounding boxes* e as da imagem. Foi utilizada a expressão proposta por [6]:

$$r = \frac{\sqrt{W_{bbox} \times H_{bbox}}}{\sqrt{W_{img} \times H_{img}}} \quad (1)$$

tal que  $W_{bbox}$  e  $H_{bbox}$  referem-se à largura e altura, respectivamente, da *bounding box* do alvo e  $W_{img}$  e  $H_{img}$  à largura e altura da imagem (800 x 800 pixels).

Tomaram-se como alvos pequenos, aqueles nos quais  $r \leq 0,02$ , totalizando 545 alvos. A escolha do valor de  $r$  para os alvos menores se deu com base no proposto por [6], que considerou como pequenos os alvos nos quais  $r \leq 0,2$ , entretanto, visando explorar os alvos com dimensões ainda menores, empregou-se um  $r$  dez vezes menor, ou seja, os alvos considerados pequenos em [6], aqui são tratados como alvos grandes.

A quantidade de alvos, então, foi adotada como parâmetro para seleção do outro extremo (grandes), ou seja, ao selecionar os 545 maiores alvos, obteve-se um valor de  $r \geq 0,065$ . A finalidade dessa escolha se dá no intuito de comparar grupos com a mesma quantidade de alvos.

Foram feitas três rodadas de treinamentos para cada rede partindo de pesos aleatórios, ou seja, não houve *transfer learning*. O *dataset* utilizado foi o HRSID dividido na proporção 65% treinamento e 35% validação, sendo essa divisão fixada e idêntica para cada rodada, seguindo os padrões disponibilizados por [8].

Os parâmetros adaptáveis da rede, como otimizador e taxa de aprendizado, por exemplo (lista completa disponível na seção correspondente em [12]) foram mantidos nas configurações padrão, incluindo o nível de confiança padrão. A única modificação se deu na inserção dos módulos ECA, conforme a arquitetura da Fig. 3. Tal escolha se justifica no objetivo do trabalho que é, basicamente, explorar o mecanismo de atenção e o tamanho do alvo através do aumento ou não no número de detecções. A maximização de desempenho através de busca de hiperparâmetros não se encontra no escopo.

Em suma, foram treinadas 3 redes YOLOv8 e 3 redes YOLOv8 + ECA visando diminuir interferências aleatórias (ruídos). Os treinamentos foram feitos até que não fossem constatadas melhoras nas métricas de desempenho por 100 épocas. A escolha desse parâmetro se deu, apesar do custo computacional, com o objetivo de diminuir as possibilidades de comparar modelos que não tivessem atingido seu auge de desempenho. Os melhores pesos de cada modelo, tomando como base os parâmetros descritos por [12] para tal seleção, foram utilizadas nas comparações que seguem.

As métricas de resposta ao experimento a serem avaliadas envolvem a detecção ou não de alvos, ou seja, verdadeiros e falsos positivos (TP e FP) respectivamente. São elas precisão (P) e *recall* (R), ambas relevantes no contexto de detecção de objetos [14] e *F1-Score* (F1) como forma de determinar o equilíbrio entre elas. Todas são expressas conforme segue [15]:

$$P = \frac{TP}{TP + FP} \quad (2)$$

$$R = \frac{TP}{TP + FN} \quad (3)$$

e

$$F1 = 2 \cdot \frac{\text{precisao} \cdot \text{recall}}{\text{precisao} + \text{recall}} \quad (4)$$

Após feitos os treinamentos dos modelos e computadas as métricas de acordo com o tamanho dos alvos e arquitetura da rede, é proposto um modelo de regressão, conforme abordado por [11], para este tipo de experimento:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2, \quad (5)$$

de maneira que os coeficientes  $\beta$  serão determinados pelo modelo de regressão e  $x_1$  e  $x_2$  são os fatores "arquitetura" e "tamanho do alvo" respectivamente.

Os resultados obtidos são submetidos à Análise de Variância (ANOVA) visando determinar a relevância estatística dos fatores e interação, conforme abordado em [11]. A seguir, é empregado um teste de aderência do modelo, ou seja, visando verificar a normalidade dos resíduos, é empregado o teste estatístico de Shapiro-Wilk [18]. A seleção desse teste se deve devido à sua robustez e à quantidade de amostras analisada ser inferior a 50 [19]. Considerou-se um nível de significância  $\alpha = 0,05$ , conforme adotado por [11]. Os resultados são expostos na seção seguinte.

#### IV. RESULTADOS E DISCUSSÕES

A Fig. 4 apresenta um exemplo de comparação entre as detecções de cada arquitetura, após o treinamento, para uma imagem com alvos pequenos (linha inferior) e uma com alvos grandes (linha superior). O gabarito é exposto ao centro.

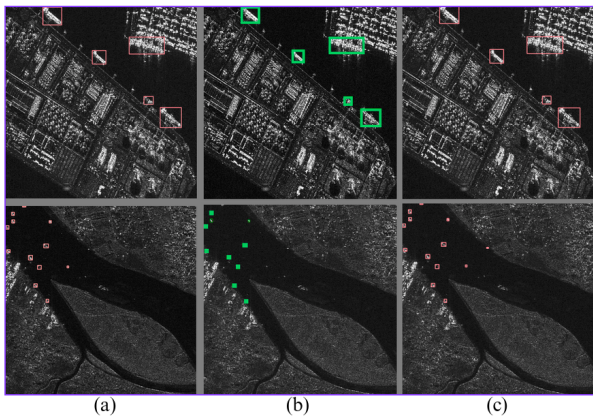


Fig. 4: Imagens com as respectivas detecções. As imagens da coluna (a) apresentam as detecções da YOLOv8, na coluna (b) tem-se o gabarito e na coluna (c) as detecções feitas pela YOLOv8 + ECA.

Realizando o experimento para os fatores e métricas propostos, pode-se observar os resultados conforme a Tabela I,

cujas numerações ao lado da arquitetura representam a rodada de treinamento, e Fig. 5. Foi considerado um *threshold* de 0,25 (parâmetro padrão da rede) de confiança, ou seja, objetos classificados como "alvos" com uma certeza acima dessa faixa eram tomados como detecções pela rede.

TABELA I: Dados do Experimento

Fatores		Métricas		
Arquitetura	Tamanho do alvo	Precisão	Recall	F1-Score
YOLOv8 + ECA (1)	Pequeno	0,805	0,908	0,853
YOLOv8 (1)	Pequeno	0,822	0,873	0,847
YOLOv8 + ECA (2)	Pequeno	0,801	0,892	0,844
YOLOv8 (2)	Pequeno	0,848	0,862	0,855
YOLOv8 + ECA (3)	Pequeno	0,758	0,897	0,822
YOLOv8 (3)	Pequeno	0,824	0,877	0,850
YOLOv8 + ECA (1)	Grande	0,681	0,890	0,772
YOLOv8 (1)	Grande	0,727	0,890	0,800
YOLOv8 + ECA (2)	Grande	0,712	0,899	0,795
YOLOv8 (2)	Grande	0,769	0,899	0,829
YOLOv8 + ECA (3)	Grande	0,699	0,899	0,787
YOLOv8 (3)	Grande	0,636	0,899	0,745

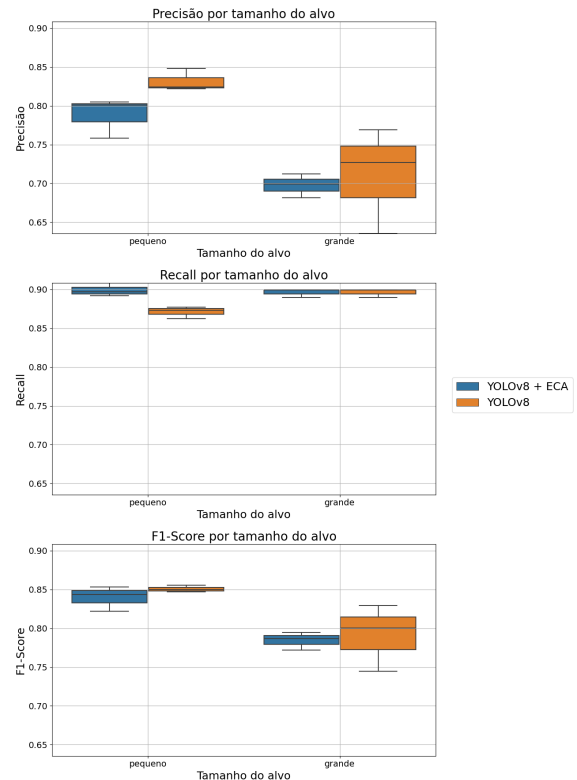


Fig. 5: Boxplot das métricas precisão, *recall* e *F1* de acordo com o tamanho do alvo.

Nota-se que os resultados variam de acordo com o tamanho do alvo e modelo empregados. Para alvos menores, é observada uma queda de precisão e aumento de *recall*, ou seja, são detectados mais alvos pequenos. Esse comportamento condiz com o esperado de acordo com [10], mas também há uma elevação de FP.

Por outro lado, para os alvos maiores, a precisão apresenta uma menor variação impactando na métrica F1 da mesma forma, mas não são observadas mudanças significativas de *recall*.

Ainda, o desempenho para alvos maiores pode ter sofrido influência da composição do banco de dados. A menor representatividade de alvos grandes no *dataset* (2% conforme

[8]) impacta a capacidade dos modelos em generalizar. Vale ressaltar que o treinamento para todas as arquiteturas foi feito utilizando a mesma distribuição.

Utilizando-se, agora, da análise estatística, a ANOVA indica que o tamanho do alvo possui contribuição significativa para as métricas de precisão e *F1-Score* em detrimento dos demais. Já para a métrica de *recall* ambos os fatores principais e a interação entre eles possuem contribuição significativa. Vale ressaltar que a análise contou com 8 graus de liberdade.

Os resíduos da ANOVA, ainda, mostraram-se normais de acordo com o teste de aderência proposto (Shapiro-Wilk), indicando a validade da análise. Para todas as análises foi considerado o nível de significância  $\alpha = 0,05$ .

Como forma de aprofundar um pouco mais a análise dos impactos do ECA e do tamanho do alvo no desempenho das redes, são exibidos na Tabela II e Fig. 6 os resultados das redes considerando o grupo de validação inteiro (sem considerar a divisão de tamanhos dos alvos):

TABELA II: Métricas para o grupo de validação inteiro

Fator	Métricas		
	Precisão	Recall	<i>F1-Score</i>
YOLOv8 + ECA (1)	0,893	0,942	0,917
YOLOv8 (1)	0,921	0,932	0,926
YOLOv8 + ECA (2)	0,891	0,939	0,914
YOLOv8 (2)	0,936	0,925	0,930
YOLOv8 + ECA (3)	0,869	0,945	0,905
YOLOv8 (3)	0,916	0,929	0,922

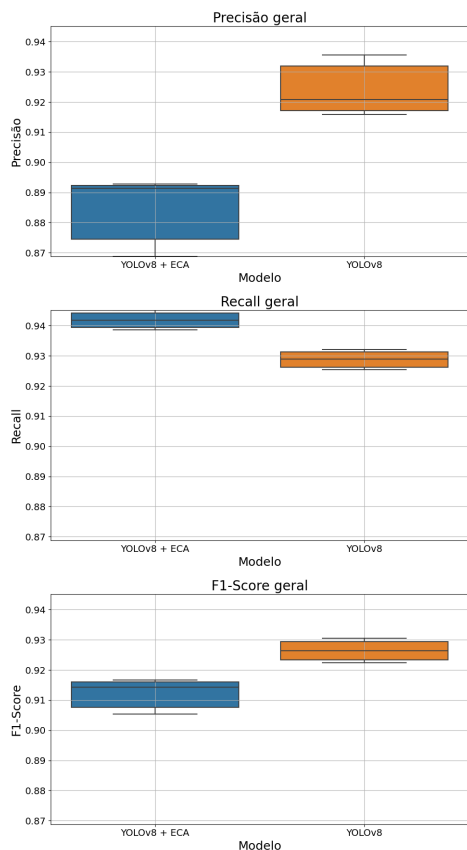


Fig. 6: Boxplot das médias das métricas precisão, *recall* e *F1* para o grupo de validação completo por modelo.

Pode-se observar pela Fig. 6 que a utilização da YOLOv8 + ECA ocasiona uma queda na precisão e aumento no *recall*

indicando, de acordo com (2) e (3), aumentos tanto nos TP quanto nos FP. A diminuição da *F1-Score* indica que o aumento dos FP é maior que o dos TP, haja vista que essa medida representa a média harmônica entre as outras duas [15].

Comparando o disposto pelas Fig. 6 e Fig. 5, observamos que o desempenho obtido para os alvos com as dimensões em ambos os extremos (pequenos e grandes) é menor do que o desempenho geral obtido pelo grupo de validação como um todo, indicando que as redes, independente da arquitetura usada, possuem melhor desempenho detectando os alvos com dimensões medianas (nesse caso, quando  $0,02 \leq r \leq 0,065$ ) para as métricas empregadas. Tal fato pode ser justificado pela catalogação original feita para o banco de dados na qual todos os alvos, independente de suas dimensões, pertenciam a mesma classe. Tal medida fez com que os valores de  $r$  variassem entre 0,001 a 0,7 para o grupo de validação.

Expressando de maneira quantitativa, para cada unidade de TP que se obtinha a mais quando utilizada a arquitetura com ECA, eram obtidos em média 2 FP a mais para alvos pequenos. Para os grandes, observou-se o acréscimo de 10 FP, em média, sem ganho de TP para a mesma arquitetura. A título de comparação, para o grupo de validação completo, a proporção entre ganho de TP por FP foi de 1:4 ainda considerando o modelo em voga.

Num contexto geral, a elevação tanto dos FP quanto dos TP indica que a arquitetura modificada (YOLOv8 + ECA) tende a ficar mais permissiva. Tal tendência, então, passou a ser analisada neste trabalho com o intuito de encontrar, caso presente, algum nível de confiança que otimizasse as métricas propostas para tal modelo. Seguindo tal lógica, foram testados diferentes níveis de confiança (*threshold*) e calculadas novamente as métricas. Os resultados obtidos para os alvos grandes e pequenos são expressos pelas Fig. 7 e Fig. 8 respectivamente.

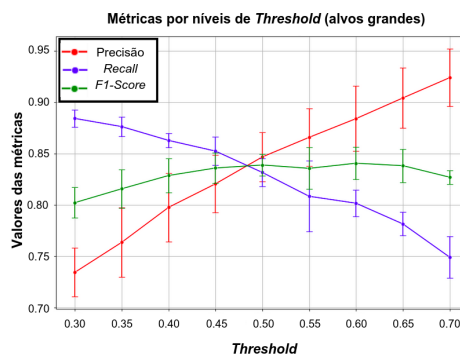


Fig. 7: Valores das métricas de acordo com diferentes níveis de *threshold* para alvos grandes.

Diante do disposto pelas Fig. 7 e Fig. 8, vê-se que não há nível de confiança no qual a queda na quantidade de FP não resulta em perdas de TP, ou seja, os alvos classificados como TP são feitos com baixo grau de confiança. Não se vislumbrou a necessidade de abordar valores maiores que 0,7 pois foi constatada uma tendência apenas de queda na métrica F1 a partir desse nível bem como o declínio acentuado no *recall*.

A análise dos diferentes níveis de confiança, ainda, mostra os pontos onde o *recall* e a precisão apresentam o melhor equilíbrio entre TP e FP, ou seja, maior F1. O evento ocorre quando os *thresholds* são selecionados em, aproximadamente,

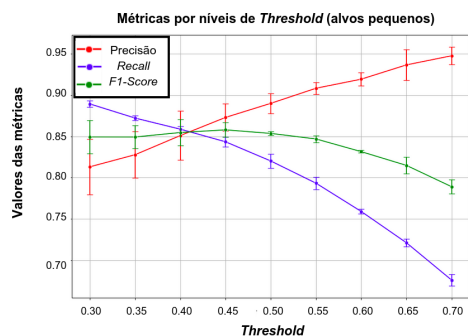


Fig. 8: Valores das métricas de acordo com diferentes níveis de *threshold* para alvos pequenos. Fonte: O autor.

0,5 para os alvos grandes e 0,45 para alvos pequenos. A seleção do valor para os alvos grandes foi tomada como a melhor neste contexto por representar uma perda menor de TP, apesar do F1 possuir o mesmo valor quando o *threshold* é selecionado em 0,6.

## V. CONCLUSÃO

O presente trabalho propôs a utilização de um design de experimento fatorial  $2^2$  visando identificar o impacto do tamanho de embarcações em imagens SAR, da implementação de um mecanismo de atenção (*Efficient Channel Attention - ECA*) e da interação entre ambos. Para tal, as arquiteturas utilizadas foram YOLOv8 tradicional e modificada.

As redes neurais foram treinadas no *dataset* HRSID sob as mesmas condições em todas as rodadas. As métricas utilizadas para comparação de desempenho foram precisão, *recall* e *F1-Score*. Comparou-se, também, essas métricas para o grupo de validação como um todo e para o conjunto dos alvos selecionados como pequenos e grandes.

A análise de cada um dos fatores e suas contribuições estatísticas foram feitas através da ANOVA e se observou que para as métricas de precisão e *F1-Score*, estatisticamente, o tamanho do alvo é mais relevante que o modelo e que a interação entre eles. Já para o *recall*, todos os fatores se mostraram relevantes. Foi considerado, em ambas análises, um nível de significância  $\alpha = 0,05$ .

O mecanismo de atenção utilizado, de fato, eleva a quantidade de verdadeiros positivos (TP) para alvos pequenos, entretanto, também torna a rede mais permissiva, fazendo com que uma maior quantidade de falsos positivos (FP) também apareça, neste caso, na proporção de 2 FP para cada 1 TP em média. Para os alvos grandes, apesar de não haver incremento nos TP, observou-se um aumento de 10 FP em média.

Em seguida foi realizada uma análise de diferentes níveis de *threshold* buscando otimizar as métricas selecionadas para a arquitetura YOLOv8 + ECA através da redução dos FP. Observou-se que não houve nível de confiança no qual a queda de FP não ocasionasse queda de TP. Os pontos com as melhores relações entre TP e FP encontrados foram 0,45 para alvos pequenos e 0,5 para alvos grandes. Nesses pontos, a métrica F1 se mostrou a maior possível.

Entretanto, quando consideradas as métricas abordadas, o desempenho geral das redes em sua arquitetura original ainda se mostra melhor. Tal fato demonstra que a utilização da arquitetura YOLO + ECA se mostra ininteressante apenas do ponto de vista de aumento no *recall*.

Como sugestões de trabalhos futuros, pode-se realizar experimentos visando otimizar os demais hiperparâmetros da arquitetura YOLOv8 através de métodos de superfícies de resposta (RSM) ou utilizar outros mecanismos de atenção, como o *attention feature fusion* utilizado em [6]. A utilização de um *dataset* contendo exclusivamente alvos de pequenas dimensões também é sugerida para aplicações envolvendo apenas alvos dessas dimensões.

## REFERÊNCIAS

- [1] L. E. Falqueto, R. Lemos, and A. Passaro, "Persistência da aplicabilidade de informações de imagens SAR orbitais em missões aéreas de vigilância e reconhecimento: uma análise por meio de simulação Monte Carlo," *Spectrum - Aplicações Operacionais em Áreas de Defesa*, vol. 1, pp. 39–43, 2022.
- [2] BRASIL, Comando da Aeronáutica, and Força Aérea Brasileira, "Concepção Estratégica Força Aérea 100," *Boletim do Comando da Aeronáutica*, p. 155, 2018. [Online]. Available: [https://www.fab.mil.br/Download/arquivos/DCA\\_11-45\\_Concepcao\\_Estrategica\\_Forca\\_Aerea\\_100.pdf](https://www.fab.mil.br/Download/arquivos/DCA_11-45_Concepcao_Estrategica_Forca_Aerea_100.pdf)
- [3] J. Li, C. Xu, H. Su, L. Gao, and T. Wang, "Deep Learning for SAR Ship Detection: Past, Present and Future," *Remote Sensing*, vol. 14, no. 11, 2022.
- [4] H. Yi, B. Liu, B. Zhao, and E. Liu, "Small Object Detection Algorithm Based on Improved YOLOv8 for Remote Sensing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 17, pp. 1734–1747, 2024.
- [5] L. Ting, Z. Baijun, Z. Yongsheng, and Y. Shun, "Ship Detection Algorithm based on Improved YOLO V5," *Proceedings - 2021 6th International Conference on Automation, Control and Robotics Engineering, CACRE 2021*, no. 3132020110, pp. 483–487, 2021.
- [6] M. Sun, Y. Li, X. Chen, Y. Zhou, J. Niu, and J. Zhu, "A Fast and Accurate Small Target Detection Algorithm Based on Feature Fusion and Cross-Layer Connection Network for the SAR Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 16, pp. 8969–8981, 2023.
- [7] Y. Liu, P. Sun, N. Wergeles, and Y. Shang, "A survey and performance evaluation of deep learning methods for small object detection," *Expert Systems with Applications*, vol. 172, no. April 2020, p. 114602, 2021. [Online]. Available: <https://doi.org/10.1016/j.eswa.2021.114602>
- [8] S. Wei, X. Zeng, Q. Qu, M. Wang, H. Su, and J. Shi, "HRSID: A High-Resolution SAR Images Dataset for Ship Detection and Instance Segmentation," *IEEE Access*, vol. 8, pp. 120 234–120 254, 2020.
- [9] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland*, pp. 740–755, 2014.
- [10] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-Net: Efficient channel attention for deep convolutional neural networks," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 11 531–11 539, 2020.
- [11] D. C. Montgomery, *Design and Analysis of Experiments*. John Wiley & Sons, Inc, 2017, vol. 106, no. 11.
- [12] Ultralytics, "Yolov8," <https://github.com/ultralytics/ultralytics>, 2023.
- [13] C.-T. Chien, R.-Y. Ju, K.-Y. Chou, E. Xieerke, and J.-S. Chiang, "YOLOv8-AM: YOLOv8 with Attention Mechanisms for Pediatric Wrist Fracture Detection," 2024. [Online]. Available: <http://arxiv.org/abs/2402.09329>
- [14] R. Padilla, W. L. Passos, T. L. Dias, S. L. Netto, and E. A. Da Silva, "A comparative analysis of object detection metrics with a companion open-source toolkit," *Electronics (Switzerland)*, vol. 10, no. 3, pp. 1–28, 2021.
- [15] A. Geron, *Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow*, 2nd ed., R. Roumeliotis, N. Tache, K. Brown, and A. Kersey, Eds. O'Reilly Media, 2019.
- [16] ICEYE, "Iceye tasking imaging modes," 2020.
- [17] A. K. Chiu, "Exploring the Use of Experimental Design Techniques for Hyperparameter Optimization in Convolutional Neural Networks," 2021.
- [18] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965. [Online]. Available: <http://www.jstor.org/stable/2333709>
- [19] N. M. Razali and Y. B. Wah, "Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests," *Journal of Statistical Modeling and Analytics*, vol. 2, no. 1, pp. 13–14, 2011.