

Melhoria da Segmentação de Imagens em Condições Adversas no Monitoramento de Infraestrutura Costeira

Pedro V. Rubinstein¹, Ricardo L. Kauer¹,
Alexandre G. Cardeman², Marcelo Abelheira², Pedro Cruz¹,
Rodrigo S. Couto¹, Luís Henrique M. K. Costa¹ *

¹GTA – Universidade Federal do Rio de Janeiro (UFRJ)
Caixa Postal 68.504 – 21.941-972 – Rio de Janeiro – RJ – Brazil

{prubinstein,rickauer,cruz,rodrigo,luish}@gta.ufrj.br

²Centro de Operações e Resiliência (COR) - Rio de Janeiro - RJ - Brazil

{acardeman,marceloabelheira}@gmail.com

Abstract. *Smart cities increasingly rely on AI-driven solutions. In coastal areas like Rio de Janeiro, maritime risks present significant challenges, as exemplified by the collapse of the Tim Maia Bike Path. This paper proposes a method for monitoring coastal infrastructure using a custom segmentation model based on YOLO segmentation, reducing the need for constant human supervision. However, external camera placement introduces challenges making segmentation difficult. To address these challenges, we investigate the effects of data augmentation on AI model performance. As a case study, we apply this method to develop a system for the Tim Maia Bike Path, with models achieving 98% mAP50-95 throughout the day.*

Resumo. *Cidades inteligentes dependem cada vez mais de soluções baseadas em inteligência artificial. Em áreas costeiras, como o Rio de Janeiro, riscos marítimos são críticos, exemplificados pelo colapso da Ciclovía Tim Maia em 2016. Este artigo propõe um método para monitorar a infraestrutura costeira usando um modelo de segmentação baseado no YOLO, reduzindo a necessidade de supervisão humana. No entanto, o posicionamento externo das câmeras traz desafios, dificultando a segmentação. Investigamos os efeitos de técnicas de aumento de dados no desempenho do modelo de IA. Como caso de estudo, aplica-se o método para desenvolver um sistema para monitorar a Ciclovía Tim Maia, com modelos alcançando 98% de mAP50-95 ao longo do dia inteiro.*

1. Introdução

A crescente urbanização e os desafios relacionados à segurança pública têm impulsionado o desenvolvimento de diferentes tecnologias para cidades inteligentes. Em

*O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior Brasil (CAPES) - Código de Financiamento 001. O trabalho também foi financiado pelo CNPq (SEI:01300.010794/2023-1, 408255/2023-4 e 309304/2021-0), FAPERJ (E-26/204.562/2024 e E-26/204.122/2024) e pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), auxílios n°. 23/00673-7 e 23/00811-0.

particular, a utilização de sistemas de visão computacional para monitoramento de infraestruturas críticas tem se mostrado essencial para a prevenção de desastres e a garantia de respostas rápidas em situações de emergência. Um exemplo emblemático dessa necessidade ocorreu em abril de 2016 na cidade do Rio de Janeiro, quando o desabamento da Ciclovía Tim Maia, causado por fortes ondas que atingiram a parte inferior da estrutura, evidenciou um exemplo de vulnerabilidade de ambientes urbanos costeiros e a urgência por sistemas de monitoramento mais precisos e contínuos [Rio, 2016]. Em resposta a esse cenário, a prefeitura instalou uma câmera estrategicamente voltada para o costão rochoso ao lado da ciclovía, permitindo o acompanhamento em tempo real das condições marítimas.

Este trabalho concentra-se em um problema específico: a elaboração de um método para a detecção automática de riscos marítimos em infraestruturas urbanas, utilizando como estudo de caso a Ciclovía Tim Maia, a partir das imagens capturadas pelas câmeras instaladas para monitoramento. Assim, desenvolve-se um sistema de visão computacional em parceria com o Centro de Operações e Resiliência (COR) para, utilizando técnicas de segmentação de imagens, inferir a altura das ondas na encosta e gerar alertas de risco. No entanto, a variação das condições de iluminação, diurna e noturna, impõe desafios significativos aos modelos de segmentação de imagens, sobretudo em decorrência das dificuldades na anotação manual e da presença de ruídos e interferências visuais. A precisão desses modelos é crucial para interpretar corretamente os sinais de alerta e, consequentemente, aumentar a segurança dos usuários da ciclovía e da população em geral.

Este artigo mostra que a aplicação de técnicas de aumento de dados¹, como o ajuste de brilho e a aplicação de *blur* (isto é, “embaçamento” da imagem), podem melhorar significativamente a generalização dos modelos de segmentação, permitindo uma melhor adaptação entre os domínios diurno e noturno. Assim, neste trabalho foi conduzida uma análise experimental utilizando conjuntos de dados específicos para cada condição de iluminação, bem como um conjunto de dados combinado, validando os resultados por meio de validação cruzada. Entre as principais contribuições deste trabalho, destacam-se: o método de monitoramento e o desenvolvimento do sistema; a análise detalhada dos efeitos das variações de iluminação na performance dos modelos de segmentação aplicados à detecção de riscos na ciclovía; a contestação de que técnicas de aumento de brilho são fundamentais para reduzir a disparidade de desempenho entre imagens capturadas durante o dia e à noite; uma abordagem experimental que combina diferentes estratégias de aumento de dados para promover um modelo mais confiável.

Diferentemente de outros estudos que exploram aplicações do YOLO [Jocher et al., 2023] e outras redes neurais convolucionais (*Convolutional Neural Network* – CNN) em contextos variados, desde a detecção de veículos e peixes até o monitoramento de fenômenos costeiros [Midigudla et al., 2025, Wang e Zhao, 2024, Muhadi et al., 2021], o trabalho se diferencia ao investigar o impacto das condições de iluminação no desempenho de modelos de segmentação em um contexto de dados limitados. Ao utilizar conjuntos de dados específicos para imagens diurnas, noturnas e combinadas, e ao aplicar técnicas de aumento de dados como ajuste de brilho e *blur*, busca-se aprofundar o entendimento sobre como adaptar os modelos ao caso do monitoramento de encosta. Essa abordagem comparativa, que avalia a robustez dos

¹*Data augmentation* é o termo em inglês frequentemente utilizado na literatura.

modelos quando expostos a domínios distintos, possibilita uma melhor compreensão sobre a generalização dos modelos ao longo do dia, além de possibilitar uma melhor administração dos esforços de melhoria dos sistemas. Como resultado, obtém-se diversos modelos com alto grau de precisão, alcançando 98% ao longo de todo o dia.

Este artigo está organizado da seguinte forma. A Seção 2 discute os trabalhos relacionados. A Seção 3 discute o cenário de utilização dos modelos e o estudo de caso com a ciclovía Tim Maia. A Seção 4 discute o processo de obtenção do conjunto de dados, além de descrever a metodologia dos experimentos deste trabalho. A Seção 5 apresenta e discute os resultados dos modelos. Por fim, a Seção 6 apresenta as conclusões e os trabalhos futuros.

2. Trabalhos Relacionados

A detecção e a segmentação de objetos em tempo real têm sido amplamente utilizadas no contexto de cidades inteligentes [Midigudla et al., 2025, Herath e Mittal, 2022, Zaib Ullah e Gagliardi, 2020], possibilitando a identificação e o rastreamento de elementos em imagens e vídeos. Nesse cenário, abordagens baseadas em redes neurais convolucionais (CNNs) se destacam, oferecendo alto desempenho na análise visual. Diversos modelos são utilizados na literatura, como o YOLO, MobileNet e o Detec-tron [Midigudla et al., 2025, Yuldashev et al., 2023, Talaat e ZainEldin, 2023]. Neste trabalho, foi adotado o YOLO, devido à sua capacidade de processar imagens com alta precisão e eficiência, além da facilidade de implementação. Aplicações desses métodos incluem a identificação de veículos, detecção de ações suspeitas e monitoramento de fenômenos naturais [Liu, 2024, Kaur et al., 2023, Talaat e ZainEldin, 2023].

Scardino *et al.* [Scardino et al., 2022] propuseram um mecanismo de monitoramento de vídeo com o emprego de redes convolucionais aliadas a técnicas de fluxo óptico, que permitem estimar a velocidade aparente de objetos em uma sequência de imagens, para a extração de parâmetros relacionados a ondas e marés. O estudo destaca o crescente uso de técnicas de visão computacional para automatizar essa análise, reduzindo a necessidade de intervenção humana. A abordagem emprega câmeras para capturar imagens do ambiente costeiro, facilitando a detecção de padrões de comportamento das ondas. Diferentemente deste trabalho, cujo foco está na proteção e no monitoramento de infraestruturas urbanas, o estudo se concentra na obtenção de parâmetros meteorológicos oceânicos.

Muhadi *et al.* [Muhadi et al., 2021] implementam um sistema baseado em redes neurais convolucionais para detectar o nível da água em rios por meio de câmeras. A metodologia utiliza dois tipos de redes neurais convolucionais: *DeepLabV3+* e *SegNet*. Ao usar a segmentação de imagens, são utilizadas duas classes: água e fundo. O sistema classifica os níveis de risco em diferentes categorias, permitindo a antecipação de eventos críticos e auxiliando na mitigação de desastres ambientais. A avaliação dos modelos é feita por meio das métricas de acurácia, interseção sobre união e pontuação de contorno F1. Para este trabalho, optamos pela utilização do YOLOv8 que se destaca pela facilidade de implementação devido ao ecossistema fornecido pela Ultralytics [Jocher et al., 2023].

Wang e Zhao [Wang e Zhao, 2024] empregam o algoritmo YOLOv8-MSS, uma versão modificada do YOLOv8, para a detecção de objetos na superfície aquática. O modelo conta com ajustes no processamento dos dados e na função de perda para refinar a precisão das detecções, tendo seu desempenho avaliado por meio da precisão média (*mean*

Average Precision – mAP) em comparação com o YOLOv8 original e outras arquiteturas de redes neurais convolucionais. Martinho *et al.* [Martinho et al., 2023] utilizam o YOLO em conjunto com técnicas de restauração de imagem para detectar peixes em ambientes marítimos. O estudo aborda desafios em um ambiente de baixa visibilidade, utilizando técnicas de restauração de imagem para melhorar a precisão das detecções. Apesar de investigarem aspectos similares, estes trabalhos possuem um contexto de aplicação diferente. Alam *et al.* [Alam et al., 2025] exploram o uso do YOLOv5 para detecção de veículos em cruzamentos urbanos em tempo real. A metodologia de treinamento adotada envolve a separação do conjunto de dados em três categorias: imagens diurnas, noturnas e mistas, permitindo uma adaptação mais eficaz do modelo a diferentes condições de iluminação. De forma semelhante, este trabalho também emprega a divisão dos conjuntos de dados em categorias diurnas e noturnas, para avaliar o impacto das variações de iluminação na performance dos modelos de segmentação aplicados à detecção de riscos na ciclovía.

Os estudos apresentados demonstram a aplicação de técnicas baseadas em CNNs para detecção e segmentação de objetos, variando em contexto e aplicação. Enquanto alguns trabalhos focam ambientes naturais, como o monitoramento subaquático, outros abordam cenários urbanos e a análise de riscos ambientais. No entanto, não foram encontrados estudos que explorem o contexto específico deste trabalho, o monitoramento das ondas para a proteção da infraestrutura em encostas litorâneas. Neste trabalho, apresenta-se um método de inferência do risco das ondas, de acordo com a altura das ondas que atingem a encosta. O foco é a melhoria do desempenho do sistema desenvolvido a partir de desafios encontrados durante o estudo de caso da ciclovía, para diferentes condições climáticas. O método adotado busca superar desafios impostos por meio de aumento de dados por variações de iluminação e simulação das condições ambientais adversas, utilizando ajustes finos nas imagens a partir de *Gaussian blur* e de ajuste de brilho. Em particular, o conjunto de dados é separado entre um subconjunto diurno, um subconjunto noturno e um subconjunto combinado, que contém ambos os tipos de imagens. Essa divisão permite avaliar o impacto das variações de iluminação no desempenho dos modelos e investigar quais estratégias de pré-processamento e treinamento são mais eficazes para cada cenário. Além disso, analisa-se a generalização dos modelos treinados em um conjunto específico quando testados em outro, buscando compreender a robustez das diferentes abordagens utilizadas ao longo do dia inteiro.

3. Cenário de Aplicação

A proximidade de avenidas costeiras ao mar pode representar riscos, especialmente durante períodos de ressaca, nos quais as ondas podem atingir parte da infraestrutura urbana. Neste contexto, este trabalho apresenta um sistema de monitoramento baseado em visão computacional para detectar e avaliar o nível de risco associado à presença de ondas próximas a infraestruturas públicas. O objetivo principal do sistema é gerar alertas sobre o nível de risco associado à altura das ondas que chegam à costa.

3.1. Descrição Geral

O sistema de visão computacional desenvolvido utiliza imagens de determinada estrutura na encosta utilizando uma câmera posicionada estrategicamente, como na Figura 1. A partir dessas imagens, um modelo de segmentação baseado no YOLO recorta a região que contém o mar. Com base em uma marcação manual de pontos de referência, o sistema



Figura 1. Imagem da câmera de monitoramento nº 4118 do Centro de Operações do Rio de Janeiro, utilizada para monitorar a região costeira próxima ao local do acidente ocorrido em 2016.

desenvolvido é capaz de estimar se o nível do mar está elevado e, consequentemente, se há risco iminente para a infraestrutura urbana.

A Figura 2 ilustra um diagrama de funcionamento do sistema, implementado em Python 3.9. No sistema, o vídeo é capturado em tempo real de uma URL fornecida pela prefeitura e enviado ao módulo de Segmentação de Objetos, no qual o processamento é realizado a uma taxa de 15 quadros por segundo. Assim, o coração do sistema é o modelo de aprendizado de máquina responsável pela tarefa de segmentação, processo que identifica áreas de interesse dentro de uma imagem a partir de uma máscara. No contexto deste trabalho, a segmentação consiste em separar a região correspondente ao mar do restante da cena. Isso é feito por meio de um modelo treinado para reconhecer padrões visuais associados ao mar como um todo. Os pontos que delimitam o polígono gerado pela máscara de segmentação servem como base para as estimativas de altura feitas.

O modelo de segmentação utilizado é o YOLOv8n, treinado a partir de imagens coletadas das câmeras de prefeitura. O processo de anotação e geração dos conjuntos de dados foi realizado por meio da ferramenta de código aberto CVAT [CVAT.ai Corporation, 2023], escolhida pela facilidade de alteração dinâmica dos parâmetros das imagens durante o processo de anotação. A escolha do YOLOv8n como modelo de segmentação se deve ao baixo volume de dados anotados, além da necessidade de execução eficiente na infraestrutura em nuvem da prefeitura.

Para estimar a altura das ondas, o sistema se baseia na máscara de segmentação gerada dentro de uma área de interesse (no caso, encosta e ciclovía). São definidos dois contornos principais: uma curva base, que representa a linha de referência do nível do mar, e uma curva superior, que delimita a ciclovía. O sistema ajusta dinamicamente uma reta entre essas duas curvas, considerando a inclinação natural da encosta como exposto na figura 4. A altura da onda é então expressa como um percentual da cobertura total da encosta. Esse percentual é utilizado como critério para a classificação do nível de risco. A partir desse nível de risco, regras são aplicadas para o envio de alertas. Assim, o módulo de envio de mensagens é implementado utilizando a API pública de mensagens

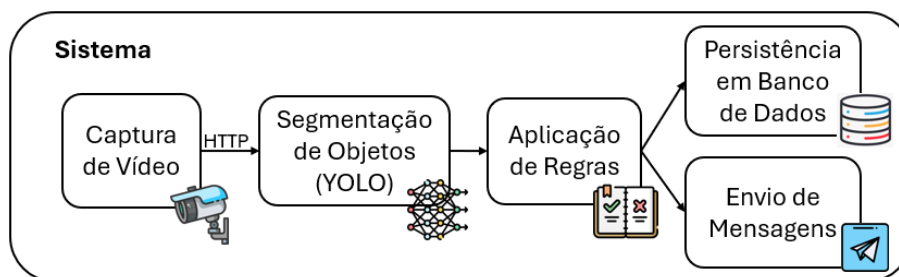


Figura 2. Diagrama do funcionamento completo do sistema desenvolvido.

do Telegram, a partir da biblioteca Aiogram [Onufriichuk, 2025]. Para fins de registro, todas as informações de altura das ondas e alertas são persistidas em um banco de dados.

3.2. Desafios

A tarefa de segmentação da água apresenta desafios significativos, especialmente devido ao ambiente externo no qual a câmera se encontra. Os desafios encontrados são o ponto de partida para as propostas de aumento de dados investigadas ao longo do trabalho. Alguns dos fatores principais que afetam a precisão dos modelos são:

- **Iluminação:** Mudanças bruscas de iluminação são causadas pelos prédios, veículos e pela infraestrutura urbana, dificultando a generalização do modelo. Além disso, especialmente durante a noite, é muito difícil distinguir a borda do corpo de água.
- **Obstrução de parte da lente:** Partículas de poeira, além de gotas de chuva, podem comprometer a visibilidade da câmera, manchando regiões específicas da imagem.
- **Embaçamento:** Em regiões costeiras, a alta umidade e a condensação frequente nas lentes da câmera podem gerar embaçamento, efetivamente reduzindo a resolução útil da imagem.

3.3. Estudo de Caso: Ciclovía Tim Maia

Como estudo de caso, o método proposto é aplicado ao sistema desenvolvido para fazer o monitoramento da Ciclovía Tim Maia. A segurança da ciclovía está diretamente ligada à proximidade do mar, principalmente durante ressacas. Nesse contexto, são enfrentados os desafios mencionados na subseção anterior, sendo o principal a dificuldade de anotação e o desempenho do sistema no período noturno, devido à complexa identificação da borda do mar. Exemplos reais das situações descritas podem ser encontrados na Figura 3.

O sistema obtido a partir deste trabalho, em funcionamento desde maio de 2024, foi calibrado para analisar a cobertura percentual da encosta pelas ondas. O sistema classifica as ondas em distintos níveis de risco de acordo com o percentual de cobertura da encosta pelo mar. Atualmente, os seguintes limiares são utilizados, escolhidos empiricamente a partir das discussões realizadas com a equipe do COR da prefeitura do Rio:

- **Muito alto:** A onda atingiu parte da ciclovía;
- **Alto:** A onda cobriu ao menos 80% da encosta;
- **Médio:** A onda cobriu de 60% a 80% da encosta;
- **Baixo:** A onda cobriu até 60% da encosta.



Figura 3. Imagens dos desafios encontrados no estudo de caso da câmera da ciclovia Tim Maia. Todas as imagens fazem parte dos conjuntos de dados capturados e utilizados neste trabalho.

Durante a operação, o sistema registra os vídeos correspondentes aos alertas disparados, além de armazenar falsos positivos e a série histórica das alturas das ondas. Sempre que um alerta é identificado, uma notificação é enviada à prefeitura por meio do aplicativo de mensagens Telegram [Telegram, 2025], conforme ilustrado na Figura 5. Desde o início do funcionamento do sistema, enviam-se alertas quando o nível de risco é considerado “alto” ou “muito alto”. Além dos alertas em tempo real, a aplicação gera relatórios periódicos: mensagens de controle e resumos a cada hora, bem como um relatório diário contendo a série de alturas registradas e os respectivos níveis de risco ao longo do período de monitoramento. A Figura 6 mostra um exemplo de gráfico diário gerado pelo sistema.



Figura 4. Exemplo de retas utilizadas durante o cálculo de altura da onda em função das curvas delimitadas.

4. Metodologia

A construção e anotação dos conjuntos de dados representam um dos maiores desafios para a realização precisa da tarefa de segmentação no sistema de visão computacional desenvolvido para o monitoramento das ondas. As dificuldades apresentadas na Seção 3 fazem com que, até mesmo para um observador humano, seja difícil identificar



Figura 5. Exemplo de alerta do sistema desenvolvido, gerado por uma onda de risco médio. A imagem enviada consiste em uma captura de tela do sistema. Em roxo, está sobreposta a máscara de segmentação do modelo. No meio, a duração do evento de risco em segundos. O número à direita indica a altura da onda em porcentagem de encosta coberta.

o contorno do mar precisamente naquelas situações. Assim, obter máscaras de qualidade durante o processo de anotação exige um trabalho minucioso que, por sua dificuldade, torna indispensável o uso de ferramentas auxiliares. Para enfrentar essa dificuldade, este trabalho adota análises e ajustes manuais nos parâmetros das imagens durante o processo de anotação. Embora algumas imagens em condições pouco desafiadoras sejam anotadas rapidamente, a maioria exige um tempo considerável para finalização.

Os conjuntos de dados utilizados neste trabalho são compostos por imagens capturadas a partir da câmera direcionada para o costão da Av. Niemeyer, na cidade do Rio de Janeiro. No presente trabalho, foram utilizadas imagens armazenadas entre janeiro e agosto de 2024. Foram construídos dois conjuntos de dados principais, abrangendo diversas condições climáticas, como chuva, vento, sol e céu nublado, além de variações na coloração do mar e na quantidade de espuma formada. Durante a fase inicial de testes, foram anotadas 500 imagens diurnas para validação do método, nas quais o modelo de produção atualmente utilizado no sistema foi treinado. Para este trabalho, um segundo conjunto de dados de tamanho equivalente foi construído a partir do mesmo processo, porém com foco em imagens noturnas. Cada conjunto de dados é dividido em 70% para treinamento, 10% para validação e 20% para teste. Em virtude da baixa disponibilidade de dados, utiliza-se a versão “nano” do YOLOv8 (isto é, YOLOv8n), para evitar problemas de sobreajuste no sistema real.

Para melhorar o desempenho dos modelos, são aplicadas técnicas de aumento de dados. Os principais desafios que puderam ser observados com o sistema em operação estavam relacionados ao embaçamento e às variações de iluminação. Para mitigar esses problemas, são aplicadas duas técnicas de aumento de dados: um *Gaussian blur* aleatório, exemplificado na Figura 7, que borra a imagem em até 2,5 pixels, e um ajuste de brilho, que altera o brilho de forma aleatória, aumentando-o ou diminuindo-o em até 25%. Ademais, foram testadas três estratégias distintas para o aumento de dados:

- **Apenas desfoque:** aplicação exclusiva do efeito de *blur*;
- **Apenas ajuste de brilho:** aplicação isolada do ajuste de brilho;
- **Desfoque e ajuste de brilho:** aplicação combinada das duas técnicas.

Para cada abordagem, apenas o conjunto de dados de treinamento foi aumentado.

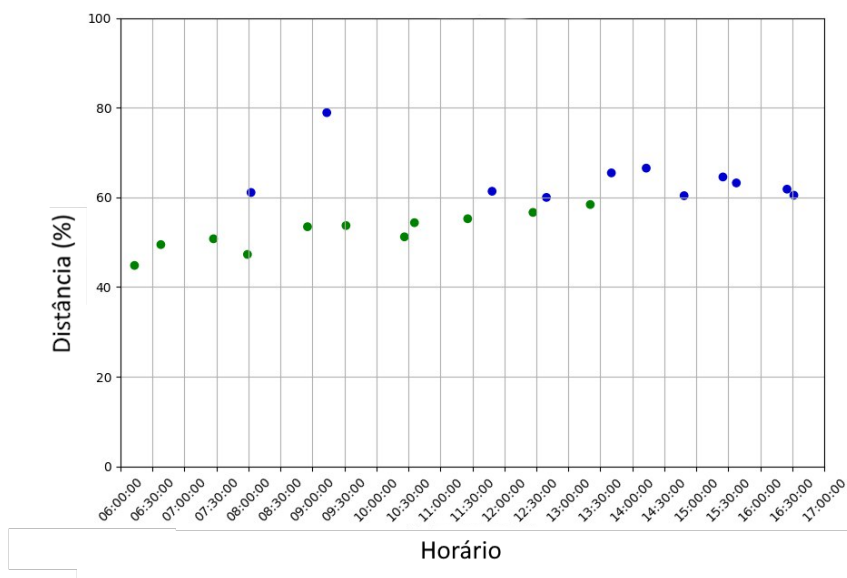
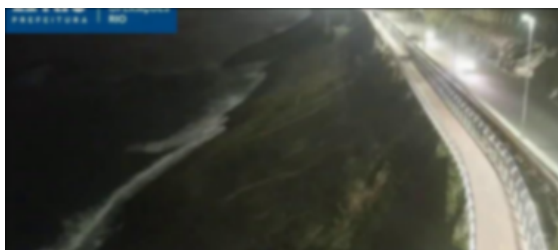
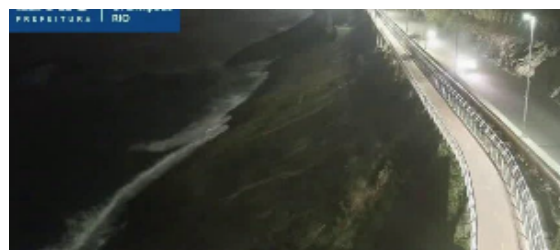


Figura 6. Exemplo de gráfico diário do sistema, coletado em 13/11/24. Cada ponto representa a maior altura detectada em percentual da costa dentro de janelas de 30 minutos. As cores indicam o nível de alerta gerado pela onda sendo verde e azul, baixo e médio respectivamente.

As técnicas foram aplicadas duas ou três vezes nos conjuntos de dados, efetivamente triplicando ou quadruplicando o conjunto de treinamento original, dado que, para cada imagem original, é gerada uma nova imagem para cada técnica de aumento de dados aplicada.



(a) Imagem da câmera com *blur* de 2,5 pixels.



(b) Imagem da câmera normal.

Figura 7. Imagens capturadas pela câmera com e sem o efeito de *Gaussian blur*.

A separação dos conjuntos de dados em categorias diurna e noturna tem o intuito de investigar o impacto das condições de luminosidade no desempenho dos modelos dado um contexto limitado de dados. Para fins de comparação, é criada uma abordagem adicional que envolve a construção de um conjunto de dados completo, composto por 250 imagens do conjunto de dados diurno e outras 250 do conjunto de dados noturno. O número total de imagens é mantido constante para garantir a consistência na comparação entre os modelos treinados. Utilizou-se validação cruzada *k-fold*, que é uma técnica que divide o conjunto de dados em múltiplos *folds* (subconjuntos), onde, a cada iteração, $k - 1$ subconjuntos são utilizados para treinamento e o restante para validação. Esse processo permite treinar e avaliar o modelo em diferentes combinações desses subconjuntos, reduzindo o risco de sobreajuste e proporcionando uma estimativa mais confiável do seu desempenho. Para este

experimento, adotou-se k igual a 5 [Pachouly et al., 2022], resultado em cada subconjunto contendo 70 imagens.

Tabela 1. mAP50-95 por fold (Dia). Em negrito, os folds selecionados para compor o dataset.

Fold	mAP50-95 (%)
Fold 1	97,84
Fold 2	98,87
Fold 3	97,21
Fold 4	96,63
Fold 5	99,32

Tabela 2. mAP50-95 por fold (Noite). Em negrito, os folds selecionados para compor o dataset.

Fold	mAP50-95 (%)
Fold 1	74,491
Fold 2	87,503
Fold 3	90,437
Fold 4	80,446
Fold 5	72,578

A seleção das imagens do conjunto de dados completo (isto é, o composto por imagens diurnas e noturnas) é feita com base nos resultados de validação obtidos para cada *fold*. A avaliação da qualidade das máscaras obtidas através dos modelos é realizada utilizando o mAP. No contexto da segmentação, o mAP é a precisão calculada a partir da interseção sobre a união (IoU) entre os pixels da máscara prevista pelo modelo e a máscara real. Uma máscara é considerada correta quando o IoU atinge ou supera um limiar pré-definido. Quanto maior o mAP, melhor a capacidade do modelo de prever com precisão os contornos das regiões de interesse. São selecionados os conjuntos de validação correspondentes aos dois modelos com pior desempenho, observados nas Tabelas 1 e 2. Essa escolha se fundamenta na observação de que as imagens presentes nesses conjuntos são consideradas essenciais para o treinamento, visto que, quando incluídas no conjunto de treinamento em outros *folds*, os modelos apresentaram desempenho superior. Em outras palavras, tais imagens representam padrões que os modelos treinados com os demais dados não conseguem aprender adequadamente. Adicionalmente, é escolhido o terceiro modelo com desempenho inferior e, deste, utiliza-se metade do conjunto de validação. O conjunto total dessas imagens selecionadas (dos dois piores modelos mais metade do terceiro) constituirá 50% do conjunto de treinamento do conjunto de dados completo.

Além da seleção de imagens baseada nos conjuntos de validação, a validação cruzada também é aplicada durante o treinamento nos diferentes conjuntos de dados, sendo essencial para identificar a variabilidade de desempenho dos modelos e, portanto, a sua capacidade de generalização. O desvio padrão entre os resultados de cada modelo obtidos através da validação cruzada é utilizado para avaliar a robustez da generalização.

5. Resultados

Os experimentos realizados envolveram o treinamento de 21 modelos, distribuídos em três grupos (diurno, noturno e completo) e submetidos a sete estratégias de aumento de dados (Puro, *blur* em dois níveis, Brilho em dois níveis e as combinações de *blur* + Brilho). Os gráficos nas Figuras 8, 9 e 10 apresentam as médias da métrica mAP50-95 obtidas por meio de validação cruzada das diferentes combinações avaliadas nos conjunto de testes para cada conjunto de dados, com imagens não utilizadas na validação cruzada. Além disso, a Tabela 3 apresenta os desvios padrão entre os modelos de cada *fold* avaliados em cada cenário.

Os resultados indicam que os modelos treinados exclusivamente com imagens diurnas apresentam alto desempenho para esse domínio (Figura 8), diurno, sendo o maior mAP de 98,08%, mas sofrem uma queda considerável ao serem avaliados no domínio noturno (Figura 9), no qual o mAP cai para 53,16%. De forma semelhante, os modelos treinados apenas com imagens noturnas alcançam bons resultados para o domínio noturno (máximo de 91,33%), porém apresentam desempenho reduzido em condições diurnas (máximo de 76,22%). Essa diferença evidencia a importância de treinar modelos específicos para cada domínio ou adotar estratégias que promovam maior generalização, sendo o aumento de dados uma solução viável para mitigar essa limitação.

Em particular, os modelos treinados com a técnica de aumento de Brilho apresentaram melhor desempenho de mAP50-95 ao serem testados no conjunto oposto (isto é, modelos diurnos testados no conjunto noturno e vice-versa). Isso sugere que a simulação de condições distintas de iluminação durante o treinamento contribui para a capacidade do modelo de lidar com variações de luz e, consequentemente, reduz a necessidade de grandes volumes de dados reais para alcançar robustez nas previsões.

O grupo de modelos treinados com o conjunto completo apresentou resultados mais equilibrados, atingindo os melhores resultados de mAP em todos os conjuntos de teste, na maioria deles com uma boa margem. Em particular, as técnicas de ajuste de brilho (especialmente *Brilho* em escala de 3x com mAP de 98,33%) se destacaram, evidenciando que a expansão do domínio através de aumento de dados pode mitigar, ao menos parcialmente, os problemas relacionados às variações de iluminação. Assim, em cenários operacionais com condições de iluminação variáveis, a utilização de um conjunto combinado, aliada a estratégias adequadas de aumento de dados durante o treinamento dos modelos, pode otimizar o desempenho do sistema sem a necessidade de coleta massiva de novos dados.

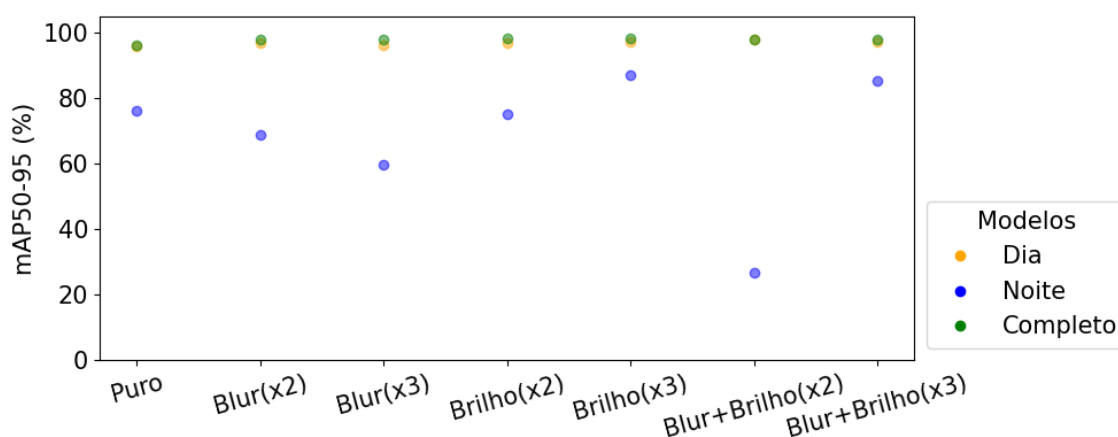


Figura 8. Gráfico contendo a média de mAP50-95 dos modelos treinados com validação cruzada no conjunto de teste do conjunto de dados diurno.

A Tabela 3 revela padrões distintos de estabilidade preditiva conforme o domínio e as técnicas de aumento de dados. Modelos treinados com o conjunto de dados diurno apresentam baixa variabilidade intra-domínio, mas grande instabilidade em testes noturnos, especialmente ao utilizar *Blur+Brilho*, conforme indicado pelos altos desvios padrão. Por outro lado, os modelos noturnos demonstram grande volatilidade quando aplicados a

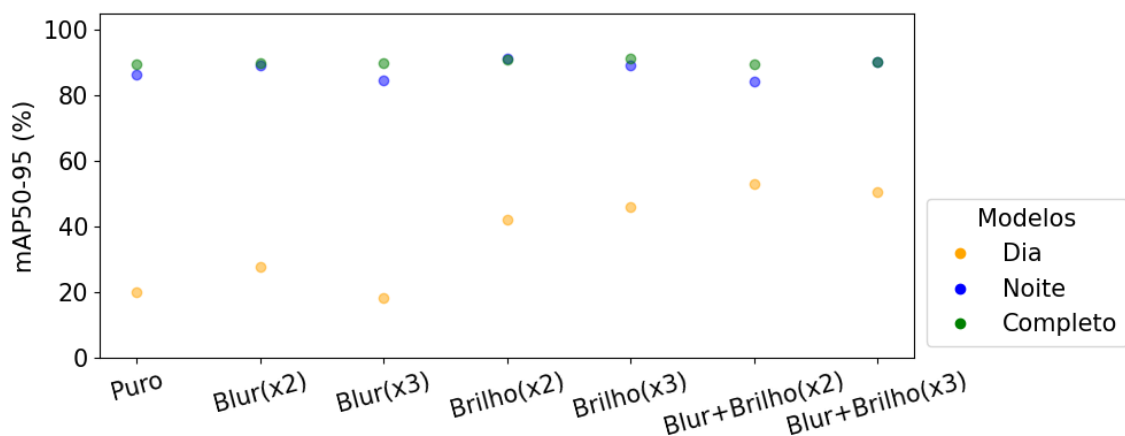


Figura 9. Gráfico contendo a média de mAP50-95 dos modelos treinados com validação cruzada no conjunto de teste do conjunto de dados noturno.

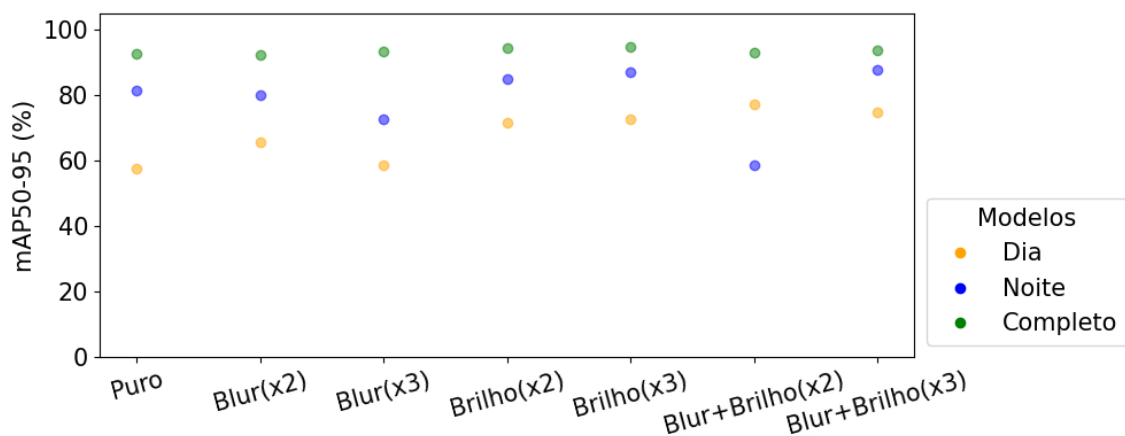


Figura 10. Gráfico contendo a média de mAP50-95 dos modelos treinados com validação cruzada no conjunto de teste do conjunto de dados completo.

cenários diurnos, atingindo $\sigma = 29,17$ com *Blur* (x3), o que sugere que borramentos muito intensos omitem características essenciais para a generalização. Já os modelos treinados no conjunto completo mostraram-se notavelmente robustos, com desvios padrão significativamente menores em todos os domínios avaliados, conforme se pode verificar nas últimas linhas da Tabela 3.

Em geral, os resultados indicam que o treinamento em um conjunto de dados combinado, aliado a estratégias de aumento de dados eficazes, podem reduzir a variabilidade das previsões e melhorar a estabilidade do modelo em cenários com iluminação variável. Além disso, a utilização de técnicas de aumento eficazes permite compensar a limitação imposta pelo baixo volume de dados disponíveis, ampliando a diversidade do conjunto de treinamento, evitando a necessidade de aquisição de novas amostras.

6. Conclusão

Este trabalho apresentou um método de monitoramento baseado em visão computacional para detectar e avaliar o nível de risco associado à presença de ondas próximas a infraestruturas públicas. Dada a dificuldade de obtenção de máscaras confiáveis para

Tabela 3. Tabela contendo o desvio padrão dos modelos treinados com validação cruzada no conjunto de teste.

Modelo	Transformação	Dia	Noite	Completo
Dia	Puro	1,0766	1,7602	1,9779
	<i>Blur</i> (x2)	1,0459	5,5409	4,060
	<i>Blur</i> (x3)	0,8800	5,8229	3,0992
	<i>Brilho</i> (x2)	0,9810	14,7440	8,4161
	<i>Brilho</i> (x3)	0,5699	12,6112	5,5970
	<i>Blur + Brilho</i> (x2)	0,3940	17,0630	7,1284
	<i>Blur + Brilho</i> (x3)	1,2265	19,2736	9,6371
Noite	Puro	9,4856	4,4367	5,9646
	<i>Blur</i> (x2)	15,0832	4,0400	6,6491
	<i>Blur</i> (x3)	29,1712	9,1313	19,5986
	<i>Brilho</i> (x2)	11,4704	1,1234	4,3062
	<i>Brilho</i> (x3)	1,4265	3,1448	4,0940
	<i>Blur + Brilho</i> (x2)	10,8416	8,0250	8,5890
	<i>Blur + Brilho</i> (x3)	2,6044	2,0080	1,3738
Completo	Puro	1,9130	1,1160	1,1185
	<i>Blur</i> (x2)	0,3841	4,4660	3,8592
	<i>Blur</i> (x3)	0,7763	1,4022	0,8480
	<i>Brilho</i> (x2)	0,4063	1,4888	0,7260
	<i>Brilho</i> (x3)	0,8747	1,0344	1,0220
	<i>Blur + Brilho</i> (x2)	0,3735	2,8581	1,1969
	<i>Blur + Brilho</i> (x3)	0,4635	2,3042	1,2501

treinar os modelos, investigou-se o impacto das condições de iluminação na precisão e generalização dos modelos de visão computacional aplicados a um sistema de monitoramento de risco na Ciclovía Tim Maia. Em particular, analisou-se como diferentes meios de aumentar o conjunto de dados impactam a precisão final ao longo de diferentes períodos do dia. Observou-se, ao menos para o tamanho atual dos conjuntos de dados, não se encontram justificativas para chavear entre modelos especializados durante a operação do sistema, uma vez que o modelo treinado com o conjunto de dados completo demonstrou melhor desempenho e maior consistência, chegando a um resultado de 98% de mAP50-95 no conjunto de testes.

Trabalhos futuros devem focar a ampliação do conjunto de dados e na implementação de novas técnicas de expansão de domínio para aumentar a robustez do modelo. Além disso, a investigação de arquiteturas alternativas ao YOLO pode fornecer ganhos adicionais em precisão e eficiência. Também pretende-se aprimorar o sistema desenvolvido com a detecção automática da ciclovía e do nível do mar, visando maior automação e confiabilidade na emissão de alertas.

Referências

Alam, M. J., Hasan, M. Z., Rahman, M. M., Rahman, M. A., Sarker, N. H., Azad, S., Islam, T. N., Paul, B., Anjum, T., Halder, B. e Fattah, S. A. (2025). An optimized yolov5 based approach for real-time vehicle detection at road intersections using fisheye cameras.

- CVAT.ai Corporation (2023). Computer Vision Annotation Tool (CVAT).
- Herath, H. e Mittal, M. (2022). Adoption of artificial intelligence in smart cities: A comprehensive review. *International Journal of Information Management Data Insights Volume 2, Issue 1, April 2022, 100076*.
- Jocher, G., Qiu, J. e Chaurasia, A. (2023). Ultralytics YOLO.
- Kaur, R., Roul, R. K. e Batra, S. (2023). A hybrid deep learning cnn-elm approach for parking space detection in smart cities. *Neural Computing and Applications*, 35(18):13665–13683.
- Liu, X. (2024). The application of cnn in smart city management and supervision under the background of smart city. *Measurement: Sensors*, 33:101227.
- Martinho, L., Neto, O., Calvalcanti, J., Pio, J. e Oliveira, F. (2023). An approach for fish detection in underwater images. Em *Anais do XVIII Workshop de Visão Computacional*, p. 6–11, Porto Alegre, RS, Brasil. SBC.
- Midigudla, R. S., Dichpally, T., Vallabhaneni, U., Wutla, Y., Sundaram, D. M. e Jayachandran, S. (2025). A comparative analysis of deep learning models for waste segregation: YOLOv8, efficientdet, and detectron 2. *Multimedia Tools and Applications*.
- Muhadi, N. A., Abdullah, A. F., Bejo, S. K., Mahadi, M. R. e Mijic, A. (2021). Deep learning semantic segmentation for water level estimation using surveillance camera. *Applied Sciences*, 11(20).
- Onufriichuk, O. (2025). aiogram.
- Pachouly, J., Ahirrao, S., Kotecha, K., Selvachandran, G. e Abraham, A. (2022). A systematic literature review on software defect prediction using artificial intelligence: Datasets, data validation methods, approaches, and tools. *Engineering Applications of Artificial Intelligence*, 111:104773.
- Rio, G. (2016). Parte da ciclovia desaba em são conrado, zona sul do rio.
- Scardino, G., Scicchitano, G., Chirivì, M., Costa, P. J. M., Luparelli, A. e Mastronuzzi, G. (2022). Convolutional neural network and optical flow for the assessment of wave and tide parameters from video analysis (leucotea): An innovative tool for coastal monitoring. *Remote Sensing*, 14(13).
- Talaat, F. M. e ZainEldin, H. (2023). An improved fire detection approach based on yolo-v8 for smart cities. *Neural Computing and Applications*, 35(28):20939–20954.
- Telegram (2025). Telegram.
- Wang, J. e Zhao, H. (2024). Improved yolo-v8 algorithm for water surface object detection. *Sensors*, 24(15).
- Yuldashev, Y., Mukhiddinov, M., Abdusalomov, A. B., Nasimov, R. e Cho, J. (2023). Parking lot occupancy detection with improved mobilenetv3. *Sensors*, 23(17).
- Zaib Ullah, Fadi Al-Turjman, L. M. e Gagliardi, R. (2020). Applications of artificial intelligence and machine learning in smart cities. *Computer Communications Volume 154, 15 March 2020, Pages 313-323*.