

Equipe 4 - Detecção de fratura óssea

Integrantes:

José Arthur Soares Bezerra

Isabella Guimarães Aguiar de Faria

Júlia Vassimon da Silva

Sumário

- Introdução
- Metodologia
- Resultados
- Conclusões

Introdução

Resumo do problema

- Detecção automatizada de fraturas ósseas em imagens de raio-X.
- Desenvolver algoritmos de visão computacional capazes de identificar e classificar diferentes tipos de fraturas em imagens radiográficas das extremidades superiores do corpo humano.

Porque esse problema?

- A detecção manual de fraturas em imagens de raio-X é demorada e sujeita a erros humanos.
- Pode acelerar o diagnóstico, reduzir a carga de trabalho dos radiologistas e melhorar a precisão na detecção de fraturas.

Introdução

Objetivo Geral

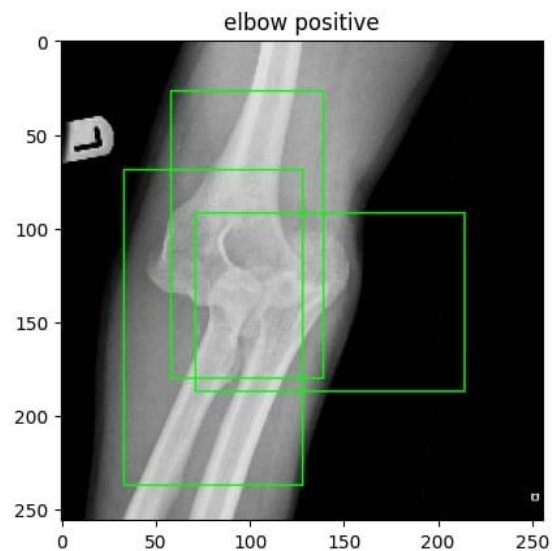
- Desenvolver algoritmos de visão computacional robustos e precisos que possam detectar automaticamente fraturas ósseas em imagens de raio-x das extremidades superiores, como cotovelo, dedos, antebraço, úmero, ombro e pulso

Justificativa

- Intenção de fornecer uma base sólida para o treinamento e avaliação de algoritmos.
- Esses modelos têm o potencial de melhorar a eficiência dos serviços de saúde, pois tem como objetivo auxiliar no diagnóstico precoce.

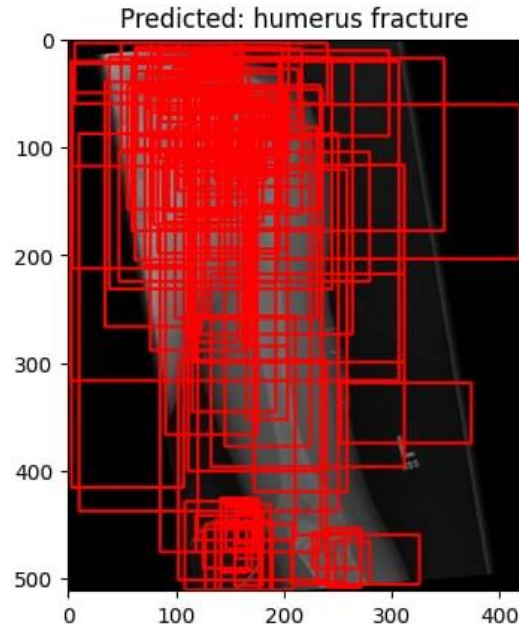
Metodologia

- Método Faster R-CNN



Metodologia

Processada apenas uma vez



Metodologia

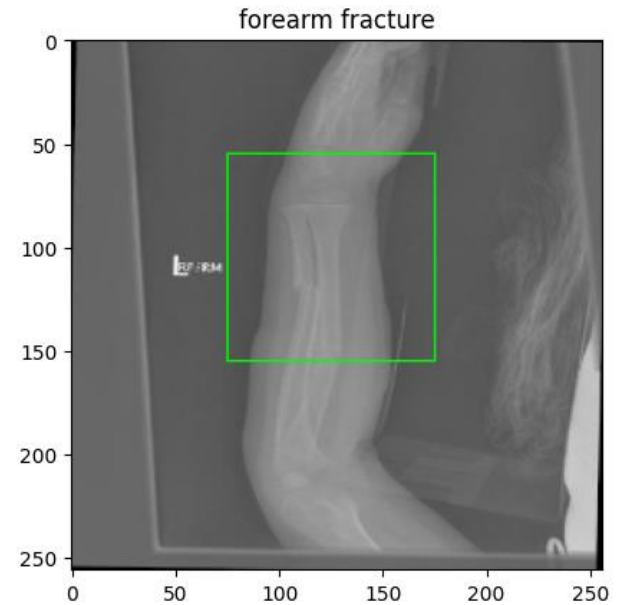
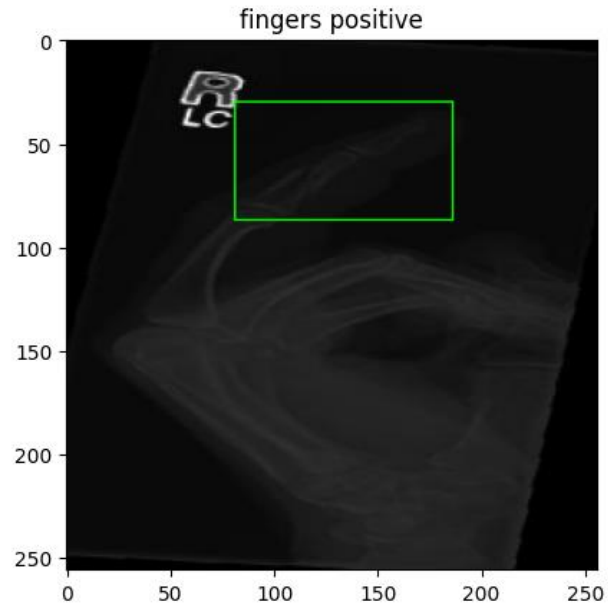
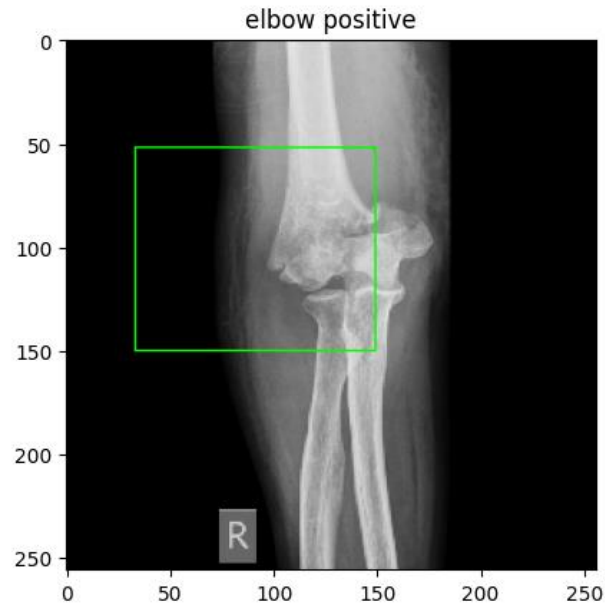
- Base de dados;

Classe	Quantidade de Imagens
Elbow positive	385
Fingers positive	606
Forearm fracture	373
Humerus fracture	3
Humerus	362
Shoulder fracture	397
Wrist positive	262

- Extraída do Kaggle
- A base de dados é composta por 3631 imagens de treino e 348 imagens de validação.

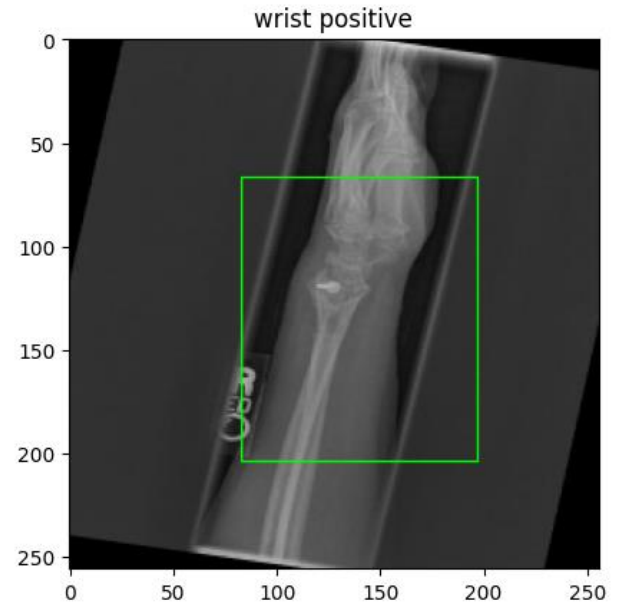
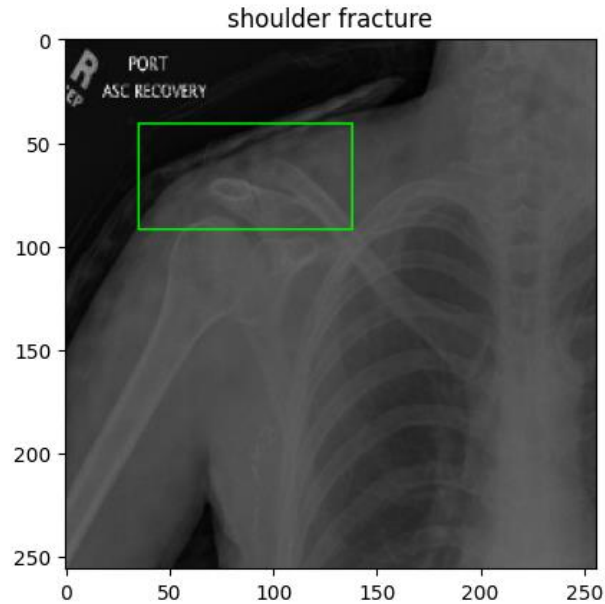
Metodologia

- Base de dados;



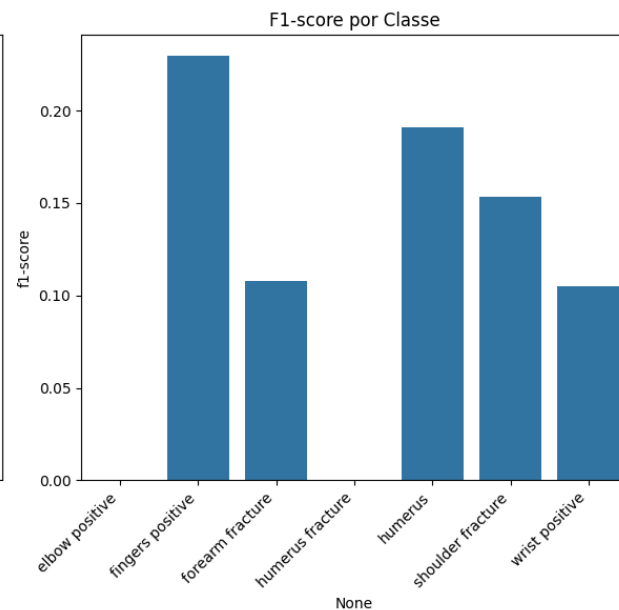
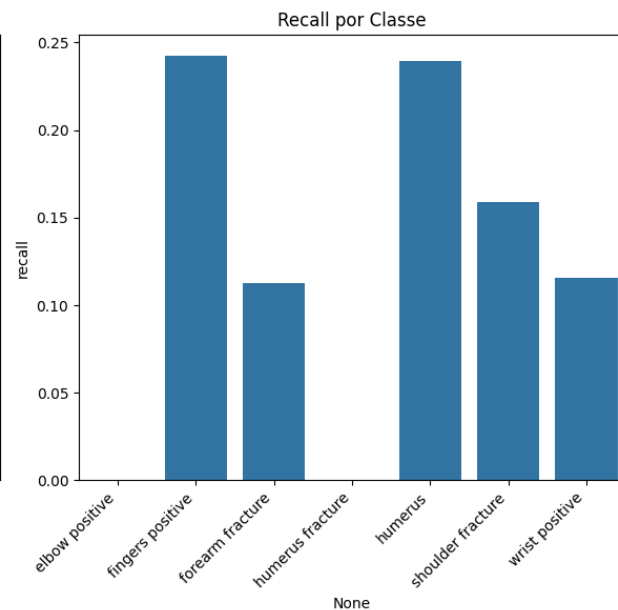
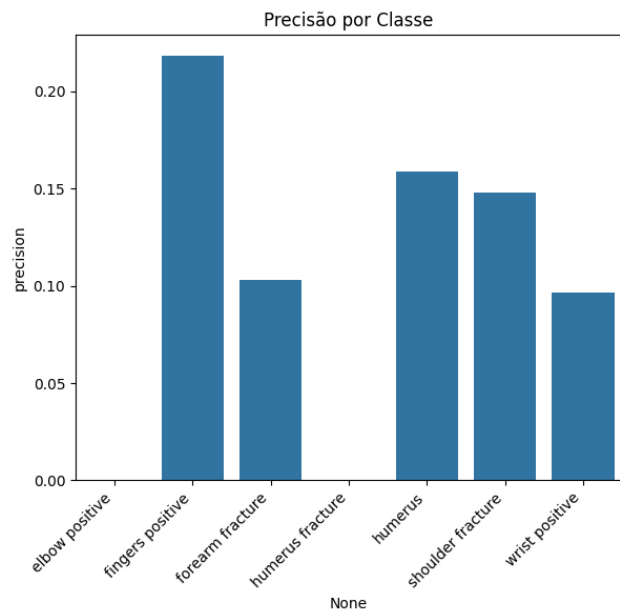
Metodologia

- Base de dados;



Resultados

Métricas de avaliação por classe



Resultados

Precisão por classe: mede a proporção de verdadeiros positivos em relação ao total de previsões positivas feitas por uma classe. Quando a precisão é alta, o modelo comete poucos erros ao classificar exemplos positivos para uma classe.

- Classe com maior precisão: fingers positive;
- Classe com menor precisão: wrist positive.

Recall por classe: mede a proporção de verdadeiros positivos em relação ao total de exemplos reais da classe. Alto recall indica que o modelo consegue identificar a maioria dos exemplos positivos da classe. Baixo recall indica que o modelo perde muitos exemplos positivos.

- Classe com alto recall: fingers positive;
- Classe com baixo recall: wrist positive.

F1 score por classe: mostra o balanço entre precisão e recall. Alto recall indica um bom equilíbrio entre as duas métricas.

- Alto F1 score: fingers positive;
- Baixo F1 score: wrist positive.

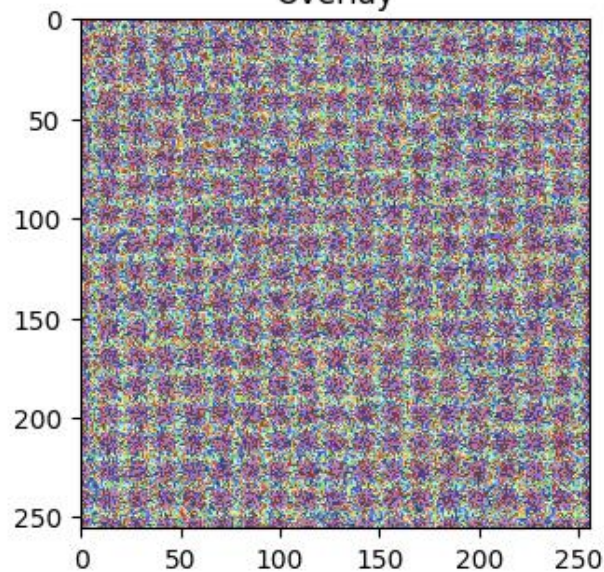
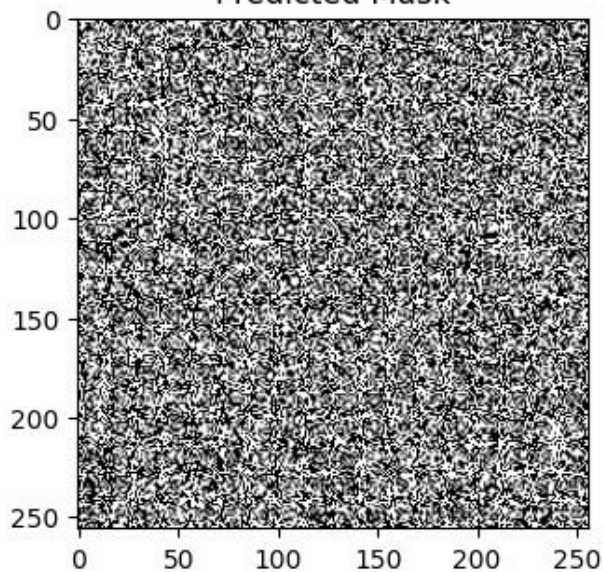
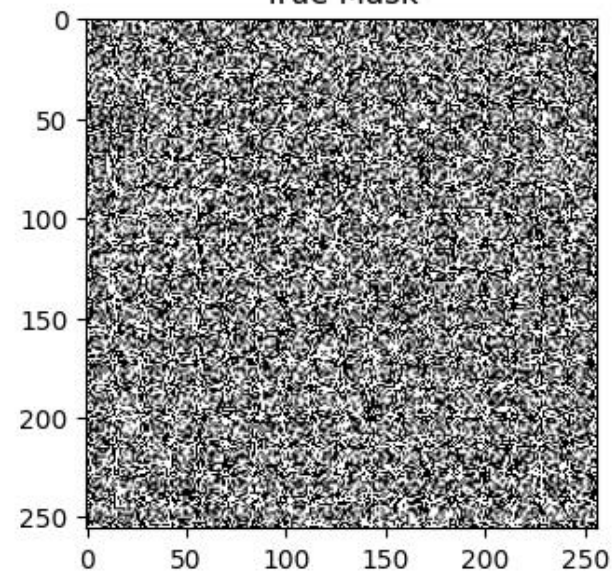
Resultados

Segmentação

True Mask

Predicted Mask

Overlay



Mean Dice Coefficient: 0.4996

Resultados

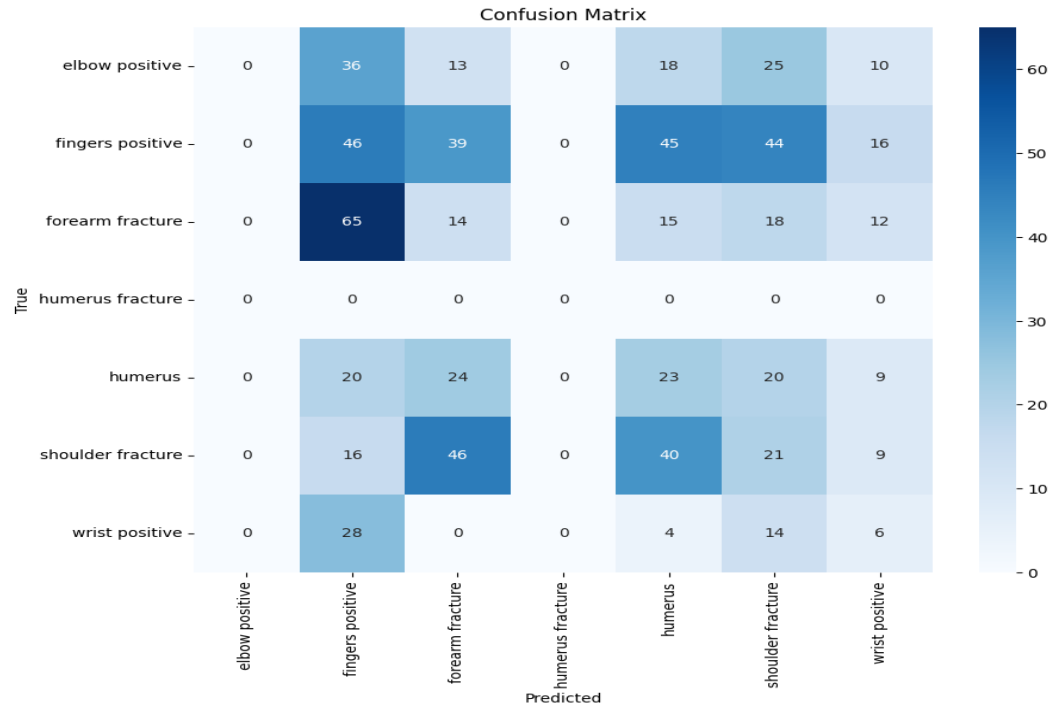
O **overlay** entre a true mask e a predicted mask facilita a visualização do verdadeiro positivo, do falso positivo, do falso negativo e do verdadeiro negativo. Para isso, escolhe-se uma cor para cada um desses componentes.

O **DICE** é a métrica usada para quantificar esse overlay (a similaridade entre as duas amostras). Essa métrica varia entre 0 e 1:

- 0: nenhuma sobreposição;
- 1: sobreposição perfeita.

Resultados

Detecção/Classificação



Accuracy: 0.1580

Resultados

A **matriz de confusão** ajuda a ter mais detalhes de como o modelo está performando para cada classe

A **acurácia** representa a proporção de previsões corretas em relação ao total de previsões. É importante considerar a quantidade de imagens em cada classe, uma vez que essa métrica não possui sensibilidade em relação às classes desbalanceadas.

Conclusões

Resultados numéricos

- Em relação às métricas de avaliação por classe (precisão, recall e F1-score), observou-se uma relação entre os resultados: a classe com maior precisão, também foi a com maior recall e, consequentemente, maior F1-score.
- O valor do DICE indica uma sobreposição de aproximadamente 50% entre a máscara verdadeira e a máscara predita, o que indica que o modelo não possui bom desempenho. Esse valor para essa métrica indica que há uma quantidade significativa de erros (falsos positivos e falsos negativos).
- O valor da acurácia para esse modelo, indica que apenas 15,8% das previsões do modelo estão corretas. A baixa acurácia pode significar que o modelo utilizado está inadequado, que os dados de treinamento são insuficientes ou desbalanceados ou que o pré-processamento de dados está inadequado.

Conclusões

Contribuição do estudo

- Aumento da velocidade de diagnóstico;
- Com a melhora do modelo, ele pode ser utilizado como ferramenta para treinar novos radiologistas;
- Os resultados podem ser usados para comparação com os resultados de outras arquiteturas de machine learning.

Conclusões

Possibilidades de melhorias/novos estudos

- Balanceamento de dados;
- Ajuste de hiperparâmetros;
- Realizar data augmentation;
- Reavaliação da arquitetura do modelo;
- Reavaliação do pré-processamento de dados.

Referências Bibliográficas

ENTENDENDO o que é Matriz de Confusão com Python: Venha aprender o que são matrizes de confusão, seu uso e confira uma implementação em Python. [S. l.], 5 mar. 2019. Disponível em: <https://medium.com/data-hackers/entendendo-o-que-%C3%A9-matriz-de-confus%C3%A3o-com-python-114e683ec509>. Acesso em: 24 maio 2024.

GILNEY: Junior. In: Qualidade de Agrupamentos (Ciência de Dados). [S. l.], 19 out. 2019. Disponível em: <https://medium.com/@gilneyjnr/qualidade-de-agrupamentos-ci%C3%Aancia-de-dados-4b1176bef5e5>. Acesso em: 24 maio 2024.