



data.life.
making life better

Machine Learning aplicado a Imagens Médicas

Apresentação



**Fernando
Camargo**

- Mestre em Engenharia de Computação pela UFG
- Atua como Engenheiro de Machine Learning na Data H
- 10 anos de experiência como Desenvolvedor
- Conhecimento em algoritmos de Machine Learning (especialmente Deep Learning), Análise de Dados, ETL, etc.



<https://www.linkedin.com/in/fernando-camargo/>

<https://github.com/fernandocamargoti>

Roteiro

- Formato de imagens médicas: DICOM
- Projetos:
 - RSNA Challenge 2018 - Object Detection de Pneumonia
 - Classificação de Pneumonia - Fleury
 - Classificação de Endometriose - BP
 - Classificação de Câncer de Mama - Hospital de Amor
 - SIIM Challenge 2019 - Segmentação de Pneumotórax
 - RSNA Challenge 2019 - Classificação de Hemorragia Cerebral

Formato de imagens médicas: DICOM

- Image + Metadados

Formato de imagens médicas: DICOM

- Image + Metadados
- Imagem não é diretamente compatível com PNG -> conversão

Formato de imagens médicas: DICOM

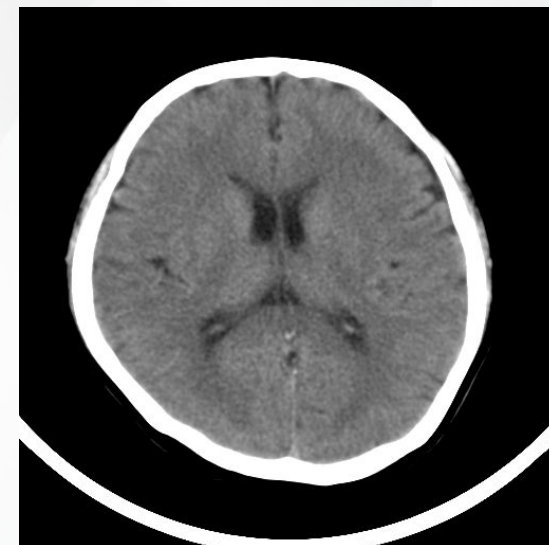
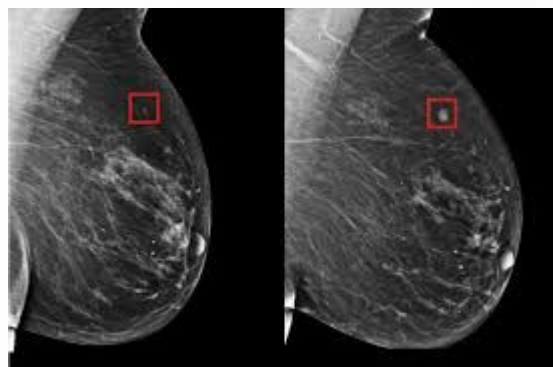
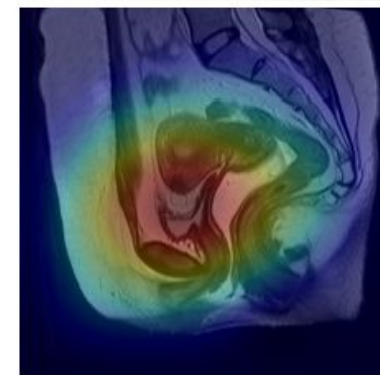
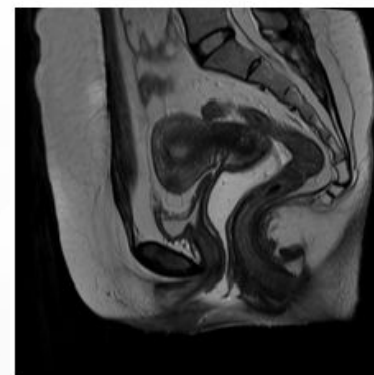
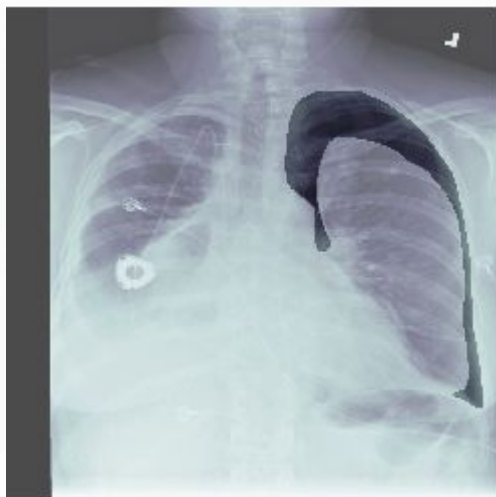
- Image + Metadados
- Imagem não é diretamente compatível com PNG -> conversão
- Metadados possuem informações sensíveis -> anonimização

Formato de imagens médicas: DICOM

- Image + Metadados
- Imagem não é diretamente compatível com PNG -> conversão
- Metadados possuem informações sensíveis -> anonimização
- Sistema de Comunicação e Arquivamento de Imagens: PACS

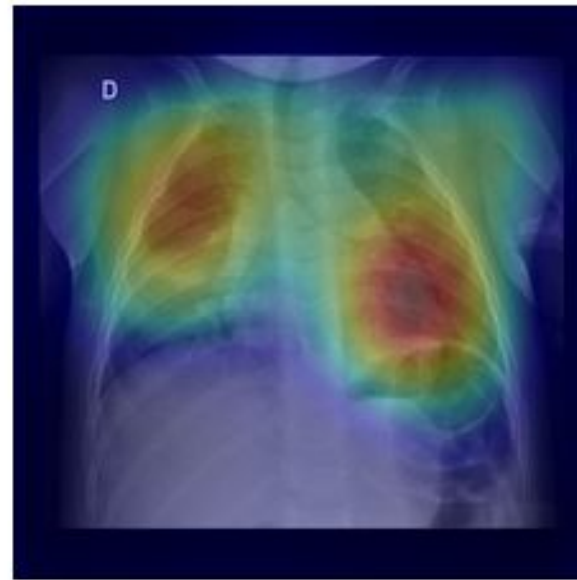
Projetos

RSNA[®]



RSNA Challenge 2018 - Object Detection de Pneumonia

RSNA®



RSNA Challenge 2018 - Object Detection de Pneumonia

Dificuldades:

- Imagens escaneadas (baixa qualidade)
- Descordância entre radiologistas
- Má formulação do problema (object detection)
- Equipe não familiarizada com Object Detection

Resultados:

- 9º lugar (porém, desclassificados)
- Primeiro aprendizado em imagens médicas e Object Detection
- Chamado para a Data H
- Criação da DataLife com primeiro projeto: Classificação de Pneumonia

Classificação de Pneumonia - Fleury

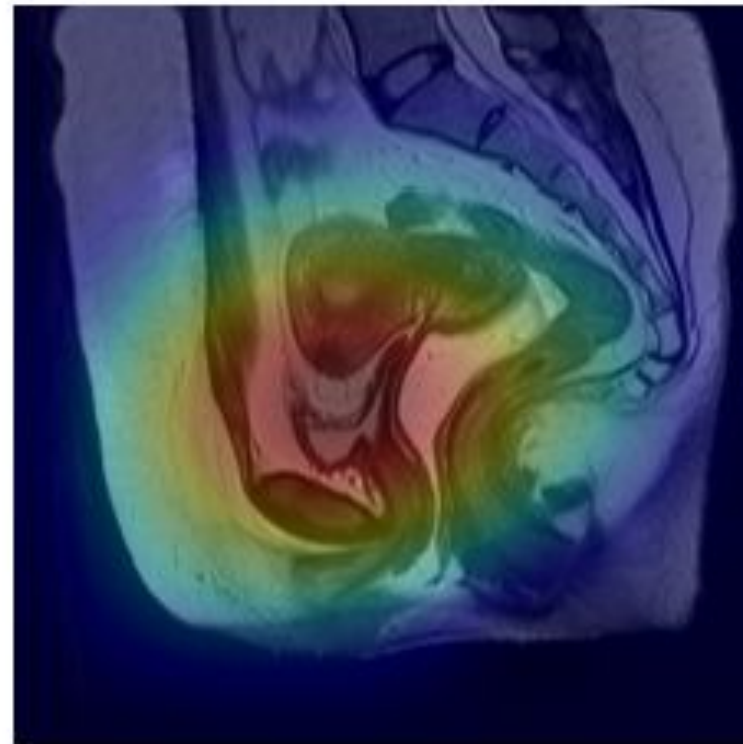
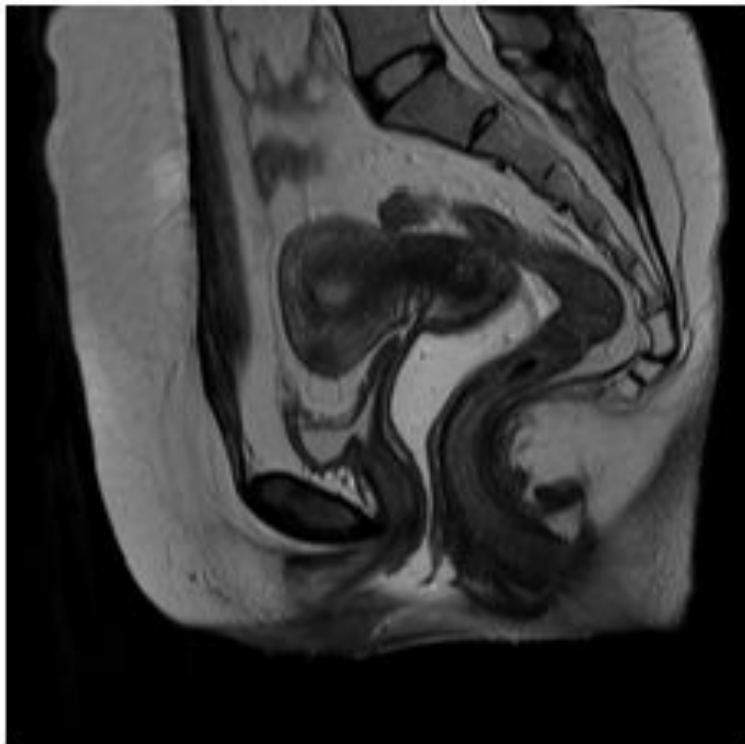
Dificuldades:

- Dataset de treinamento vindo do RSNA 2018 + alguns outros públicos
- Dataset de validação vindo da Fleury pequeno e desbalanceado
- Anonimização e comunicação com PACS para colocar em produção

Resultados:

- Projeto em fase de testes na Fleury

Classificação de Endometriose - BP



Classificação de Endometriose - BP

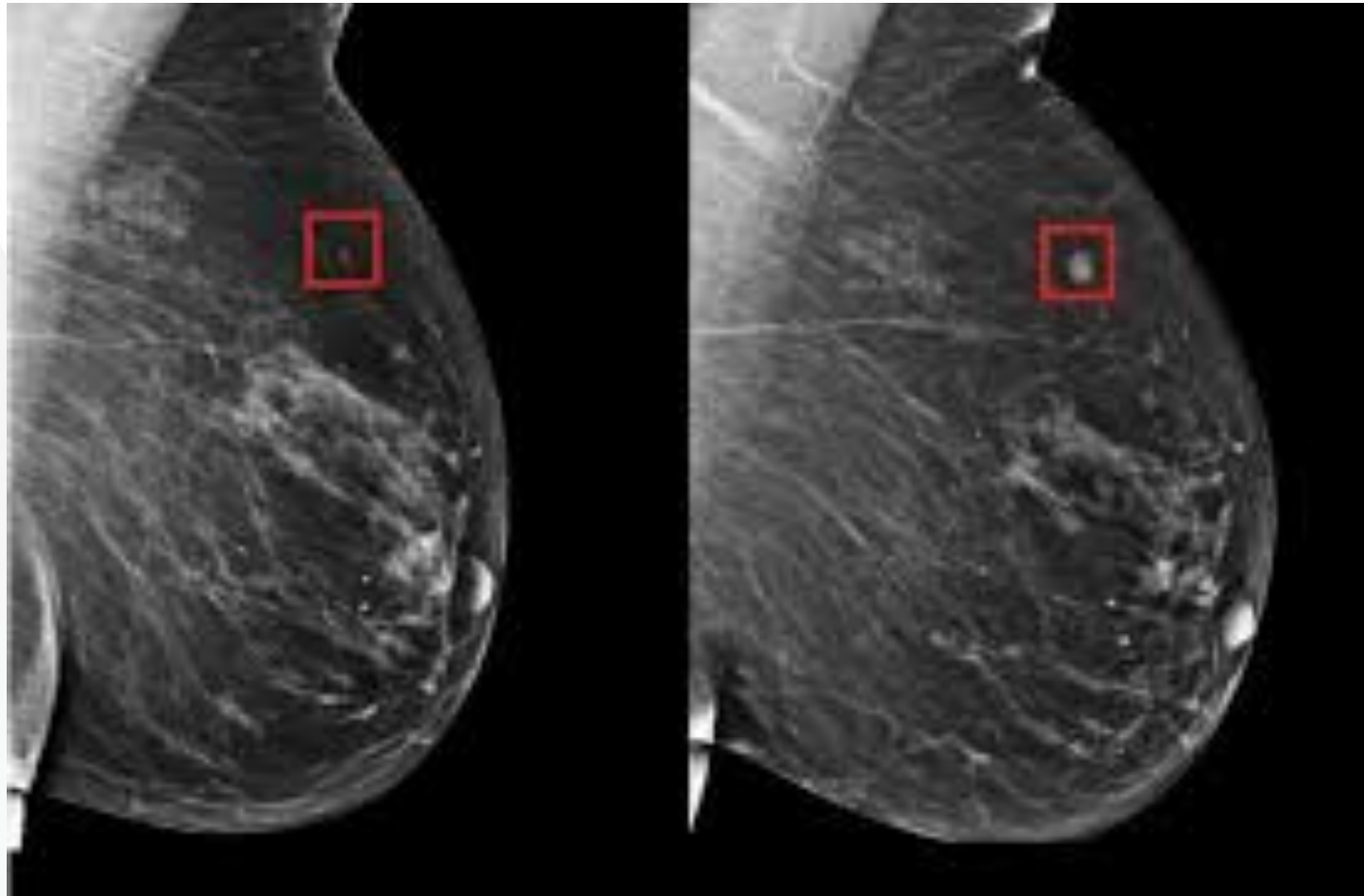
Dificuldades:

- Problema muito complexo, até mesmo para radiologistas experientes
- Primeiro dataset com erro de formulação
- Data Leakage no split do dataset
- Problemas de comunicação com os médicos
- Dataset final muito pequeno para resolver o problema

Resultados:

- É possível resolver, mas é necessário um dataset maior

Classificação de Câncer de Mama - Hospital de Amor



Classificação de Câncer de Mama - Hospital de Amor

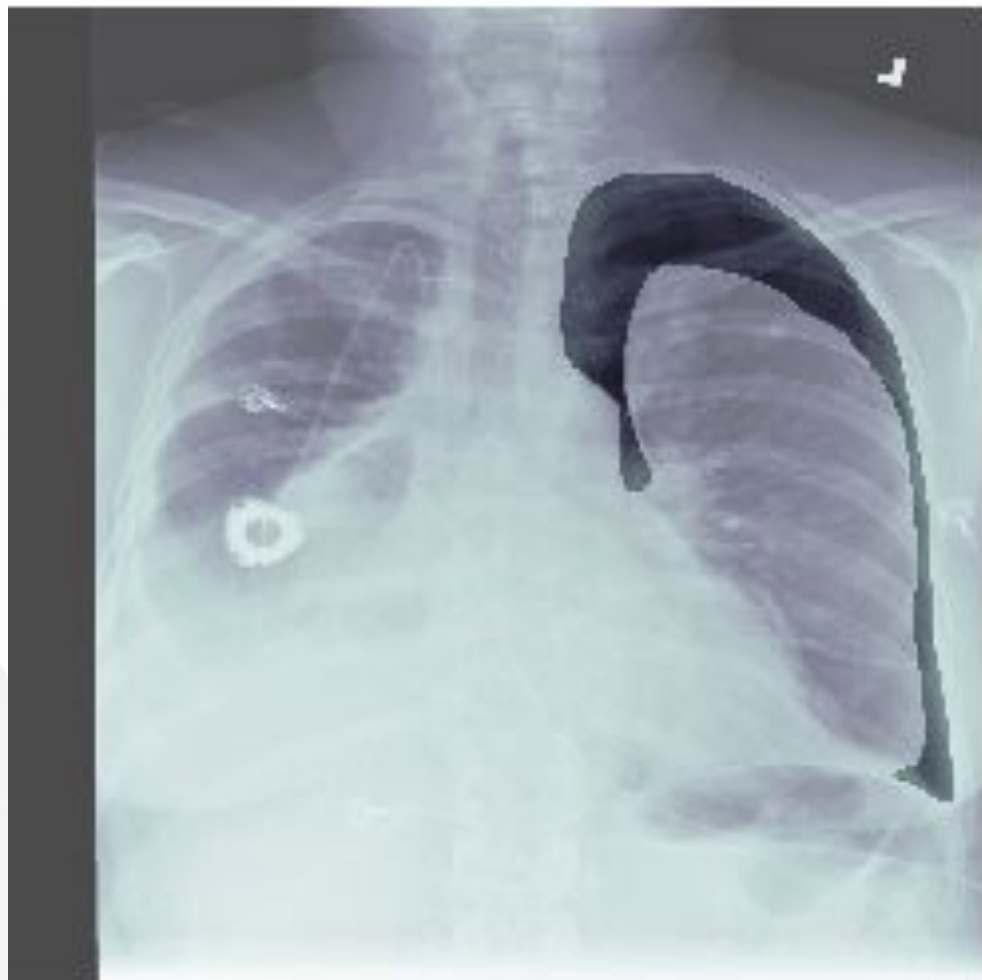
Dificuldades:

- Os nódulos são muito pequenos, precisando de alta resolução das imagens
- Labels disponíveis por exame (4 imagens)
- Falta de tempo para iniciar efetivamente o projeto

Resultados:

- Treinamento ainda não iniciado

SIIM Challenge 2019 - Segmentação de Pneumotórax



SIIM Challenge 2019 - Segmentação de Pneumotórax

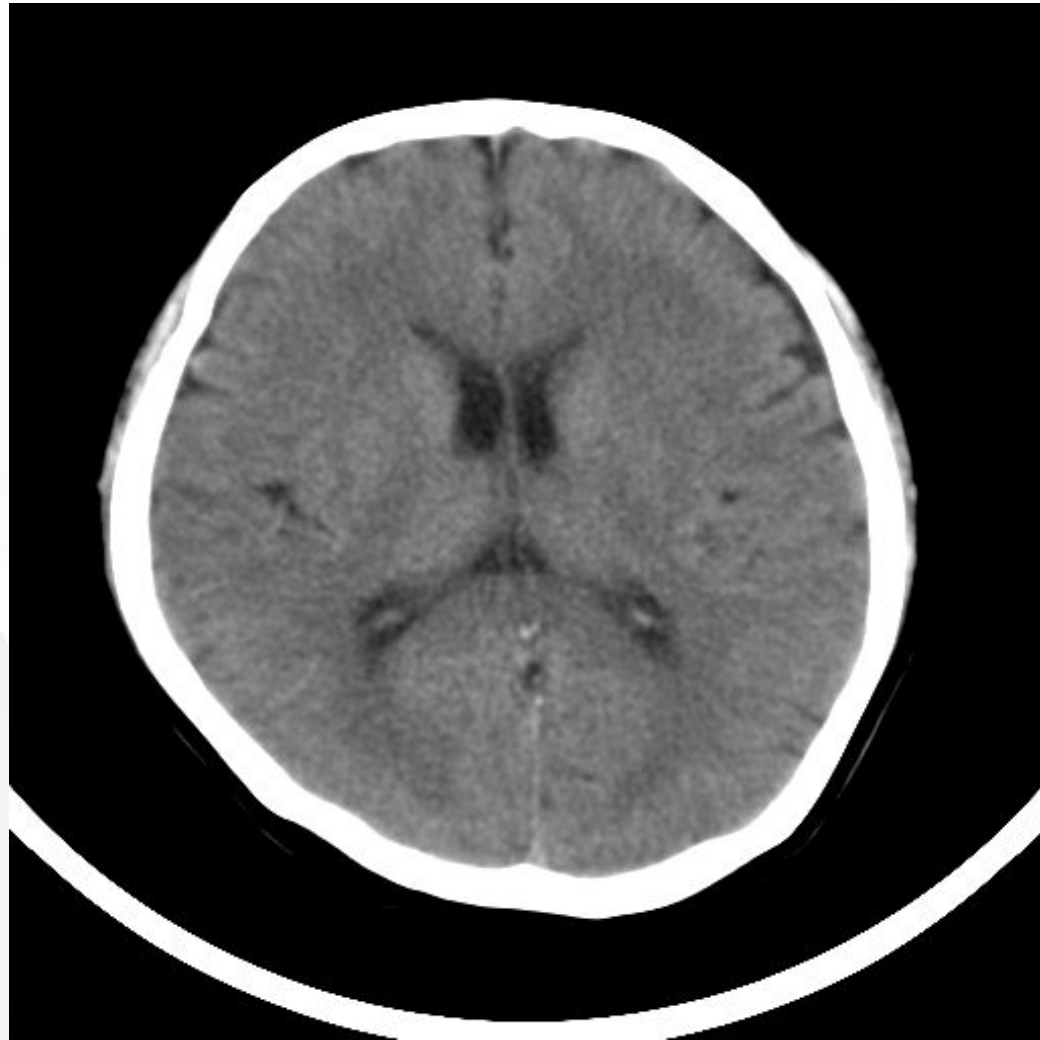
Dificuldades:

- Primeira vez trabalhando com segmentação
- Falta de tempo para focar na competição

Resultados:

- Há poucas posições de uma medalha no Kaggle
- Aprendizado na área de segmentação

RSNA Challenge 2019 - Classificação de Hemorragia Cerebral



RSNA Challenge 2019 - Classificação de Hemorragia Cerebral

Dificuldades:

- Muitas imagens 460k+ -> longo período de treinamento
- Problema complexo (múltiplos slices)
- Multi label: presença de hemorragia e seus sub tipos
- Diferente tratamento das imagens
- Métrica da competição (weighted binary crossentropy)

Resultados:

- Oportunidade de aplicação de novas técnicas (multi task learning, por exemplo)
- Começamos entre os líderes (mas estamos patinando um pouco no momento)

Obrigado!



**Fernando
Camargo**

- Mestre em Engenharia de Computação pela UFG
- Atua como Engenheiro de Machine Learning na Data H
- 10 anos de experiência como Desenvolvedor
- Conhecimento em algoritmos de Machine Learning (especialmente Deep Learning), Análise de Dados, ETL, etc.



<https://www.linkedin.com/in/fernando-camargo/>

<https://github.com/fernandocamargoti>