



CBEB 2024

XXIX CONGRESSO BRASILEIRO
DE ENGENHARIA BIOMÉDICA

Deep Learning for Medical Image Analysis: tips, tricks and traps

Profa. Dra. Leticia Rittner

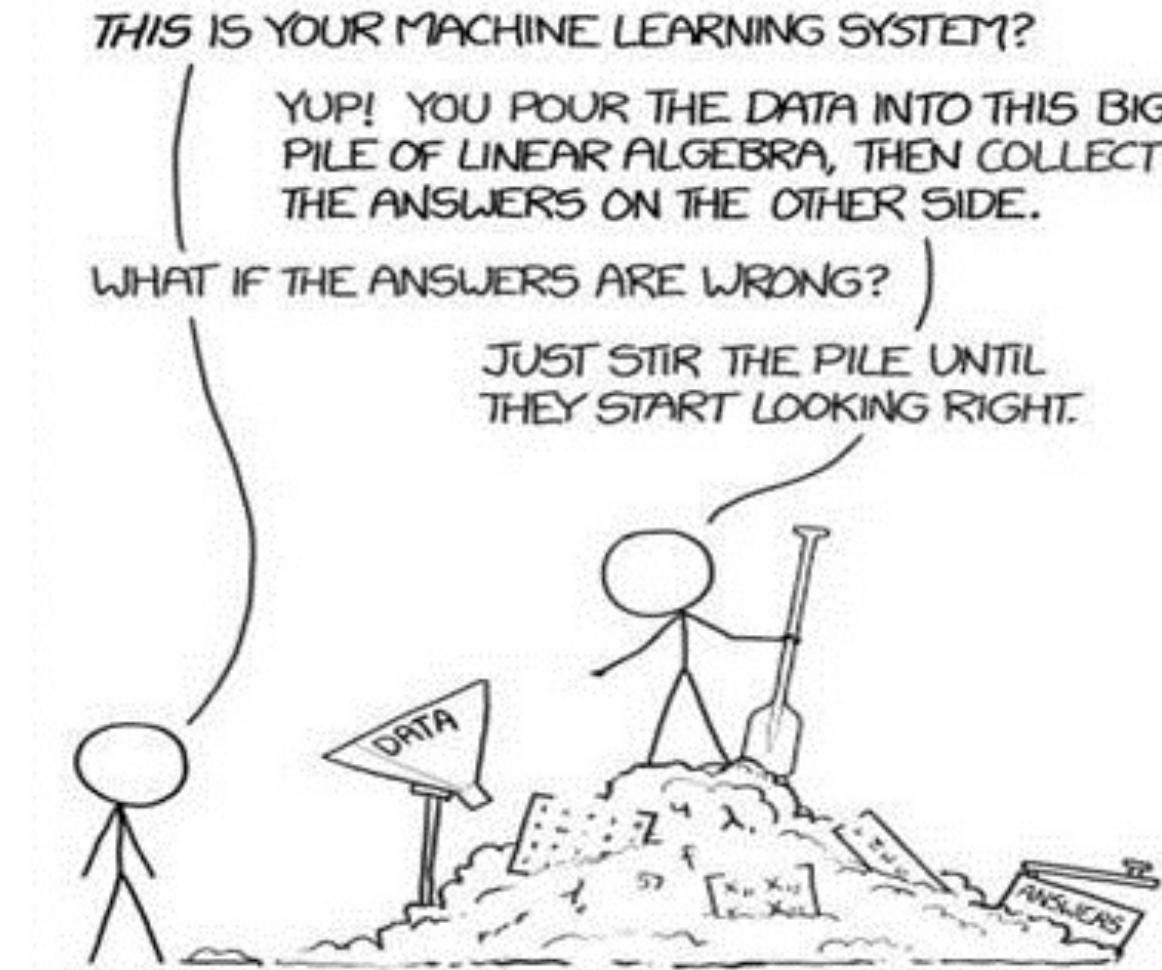
Dr. Diedre do Carmo

Medical Image Computing Lab. (MICLab)

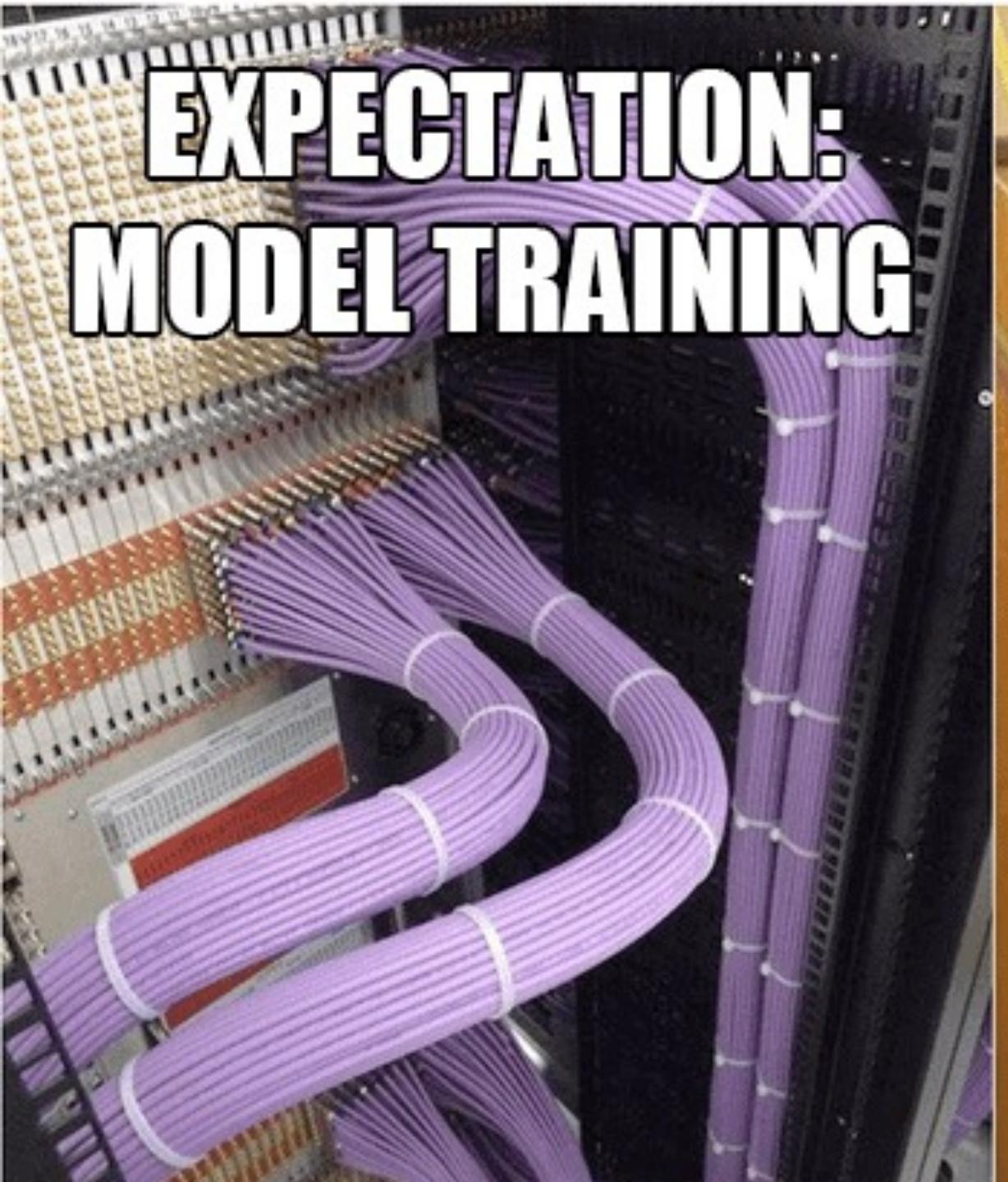
Faculdade de Engenharia Elétrica e de Computação - Unicamp



Garbage in, Garbage out



**EXPECTATION:
MODEL TRAINING**



**REALITY:
DATA CLEANING**



POST

The Future of Healthcare AI Depends on Data

February 25, 2022

Dmytro Romanchenko
March 19, 2024

CREATING AI SOFTWARE: THE CRITICAL ROLE OF DATA COLLECTION

DIGITAL TRANSFORMATION

Home » Business Topics » Data Strategist

AI's Kryptonite: Data Quality



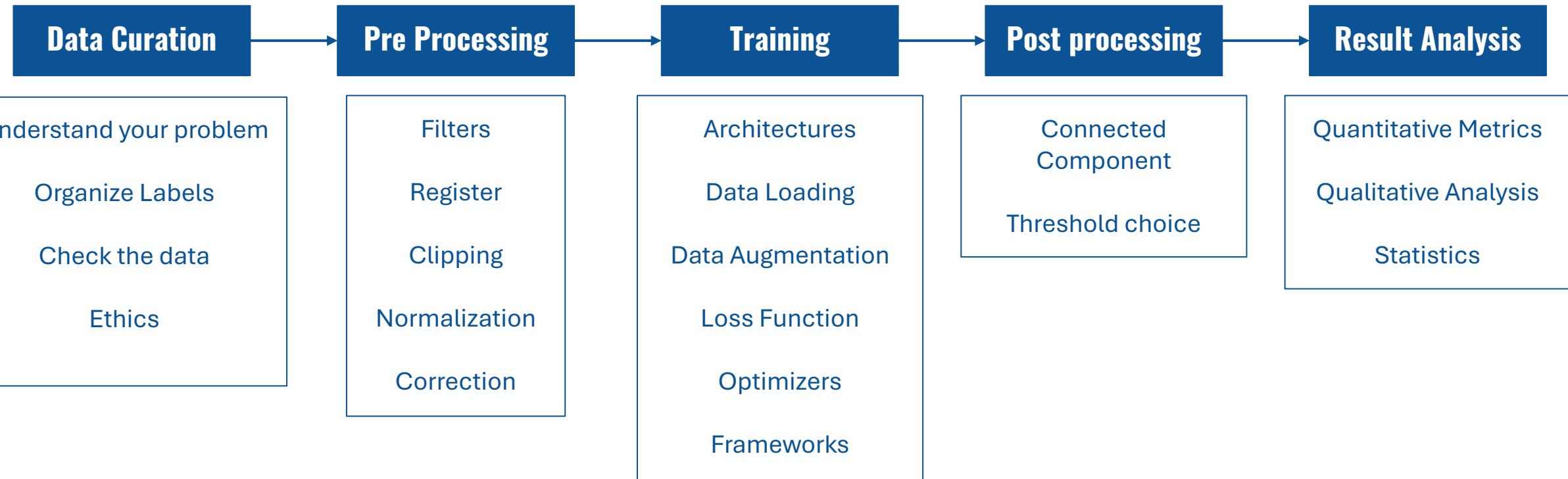
Bill Schmarzo | October 14, 2023 at 2:53 pm



AI's
Kryptonite:
Data
Qualiyt

g Data, LLC - Bill Schmarzo. All rights reserved.

Pipeline típico de análise



Pipeline típico de análise



1. Entendendo o problema/tarefa
2. Curadoria dos dados
3. Questões éticas envolvendo os dados

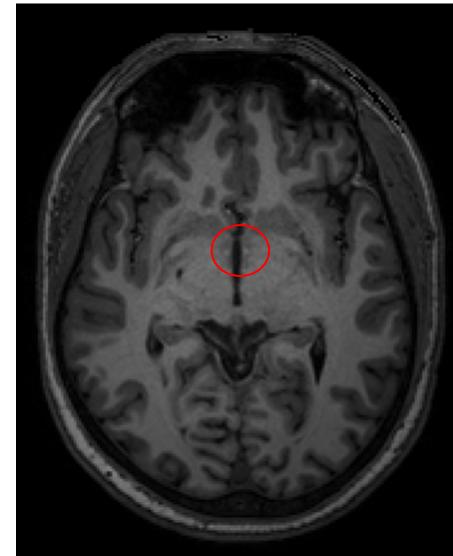
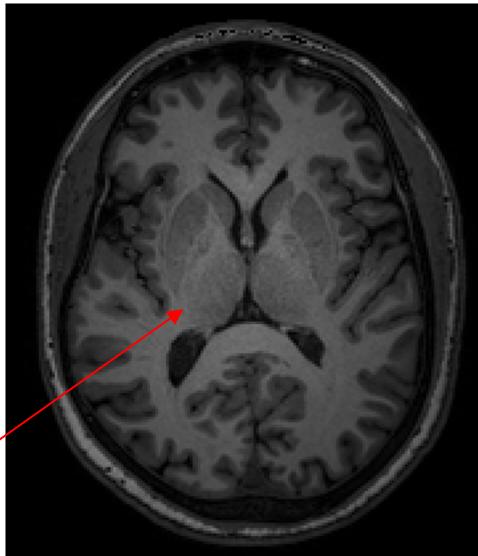
Pipeline típico de análise



1. Entendendo o problema/tarefa
2. Curadoria dos dados
3. Questões éticas envolvendo os dados

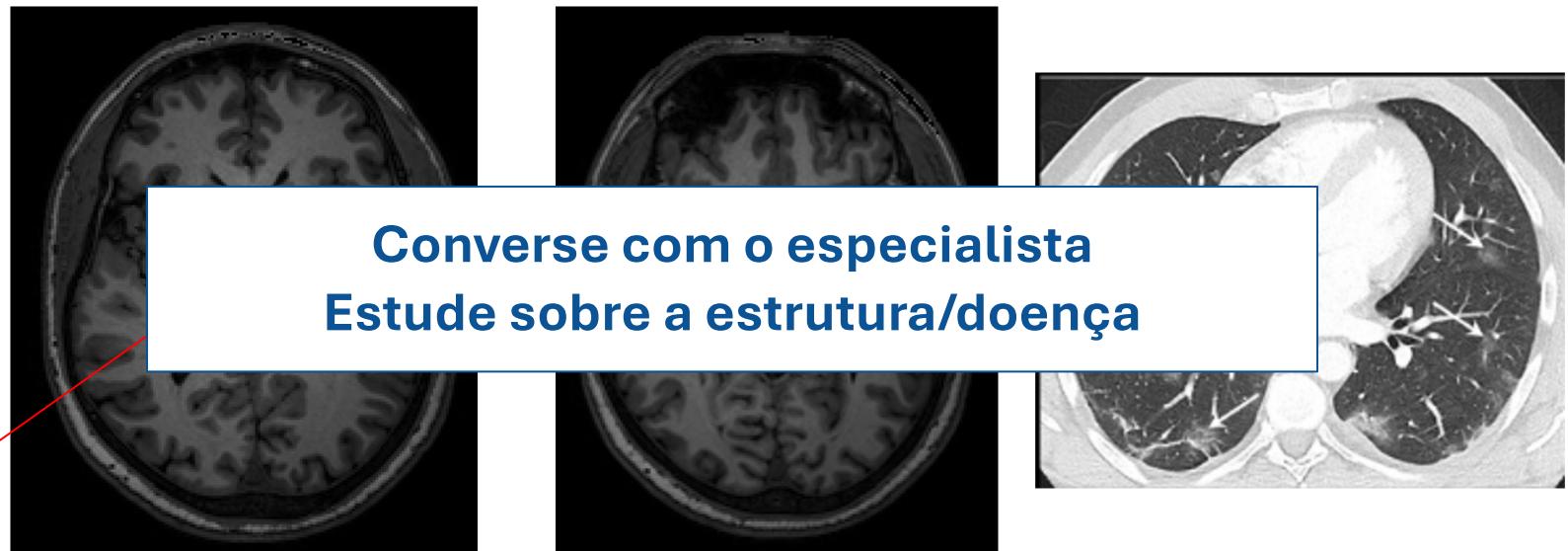
Entendendo o problema a ser resolvido

- ✓ Quais as características da sua imagem?
- ✓ Quais os atributos da estrutura de interesse ou da doença sendo estudada?
- ✓ Quais os maiores desafios?



Entendendo o problema a ser resolvido

- ✓ Quais as características da sua imagem?
- ✓ Quais os atributos da estrutura de interesse ou da doença sendo estudada?
- ✓ Quais os maiores desafios?



**Converse com o especialista
Estude sobre a estrutura/doença**

Pipeline típico de análise



1. Entendendo o problema/tarefa
2. Curadoria dos dados
3. Questões éticas envolvendo os dados

Curadoria dos dados

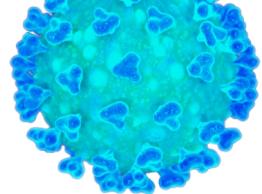
Dados brutos e rótulos

Dados e rótulos processados

Dados resultantes



Exemplo 1: Ideiagov Challenge



TASK 1

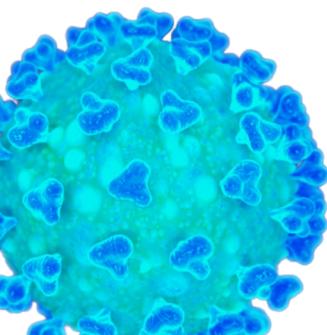
To predict COVID-19 probability from chest **X-ray** images

TASK 2

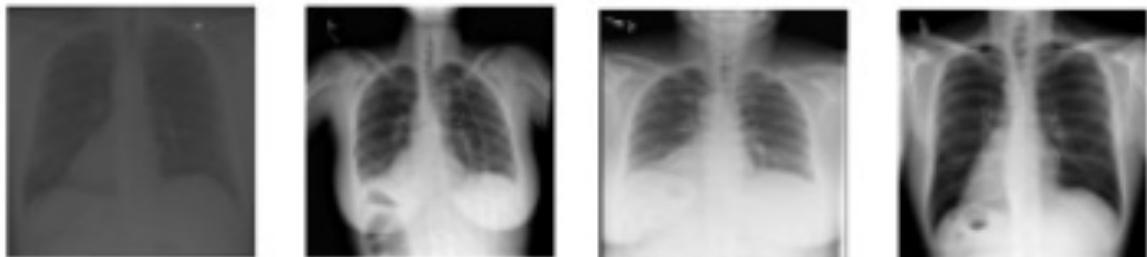
To predict COVID-19 probability from **CT** scans and to quantify findings in the lung

Data from 48 hospitals

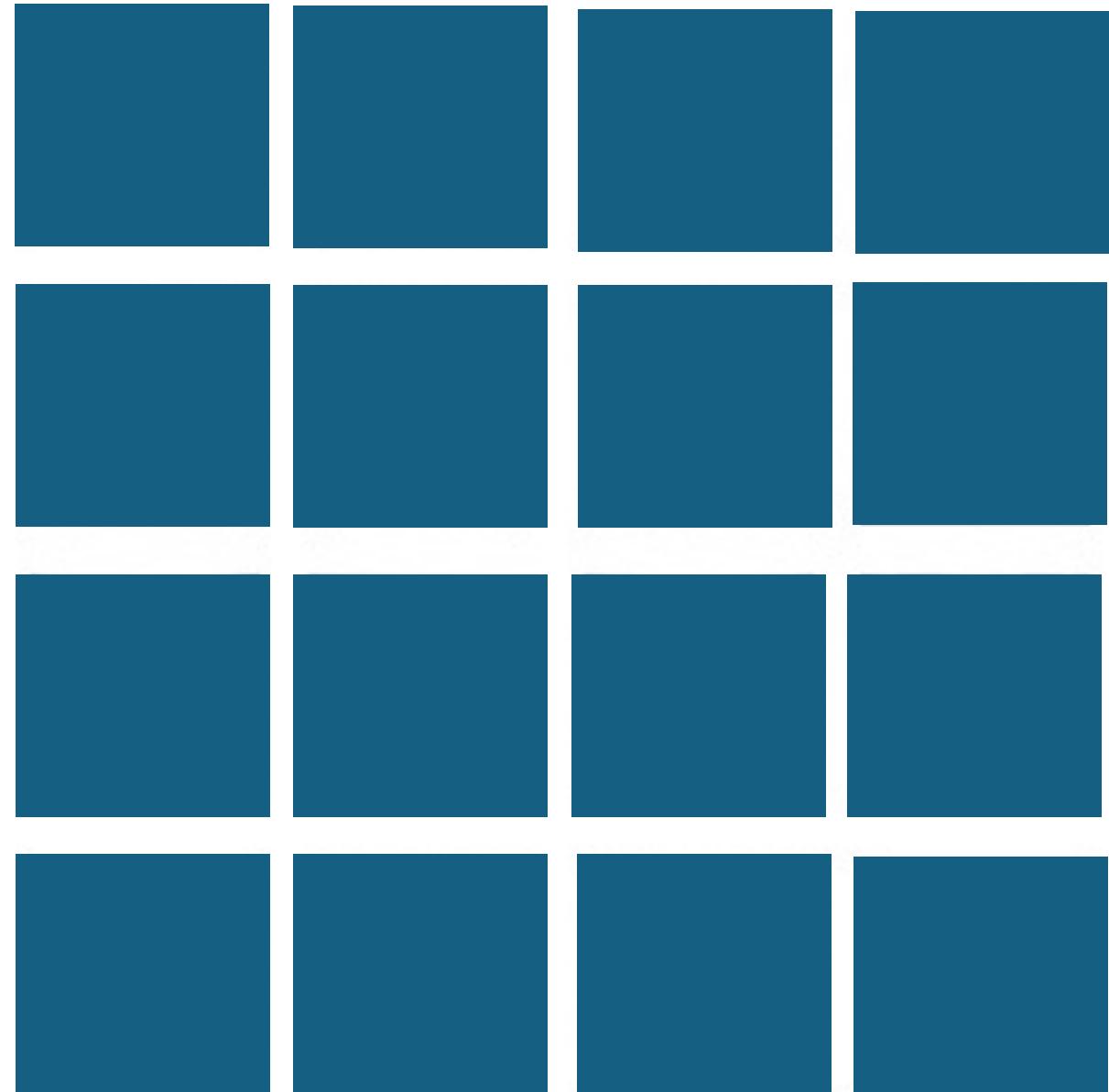
Real time operation



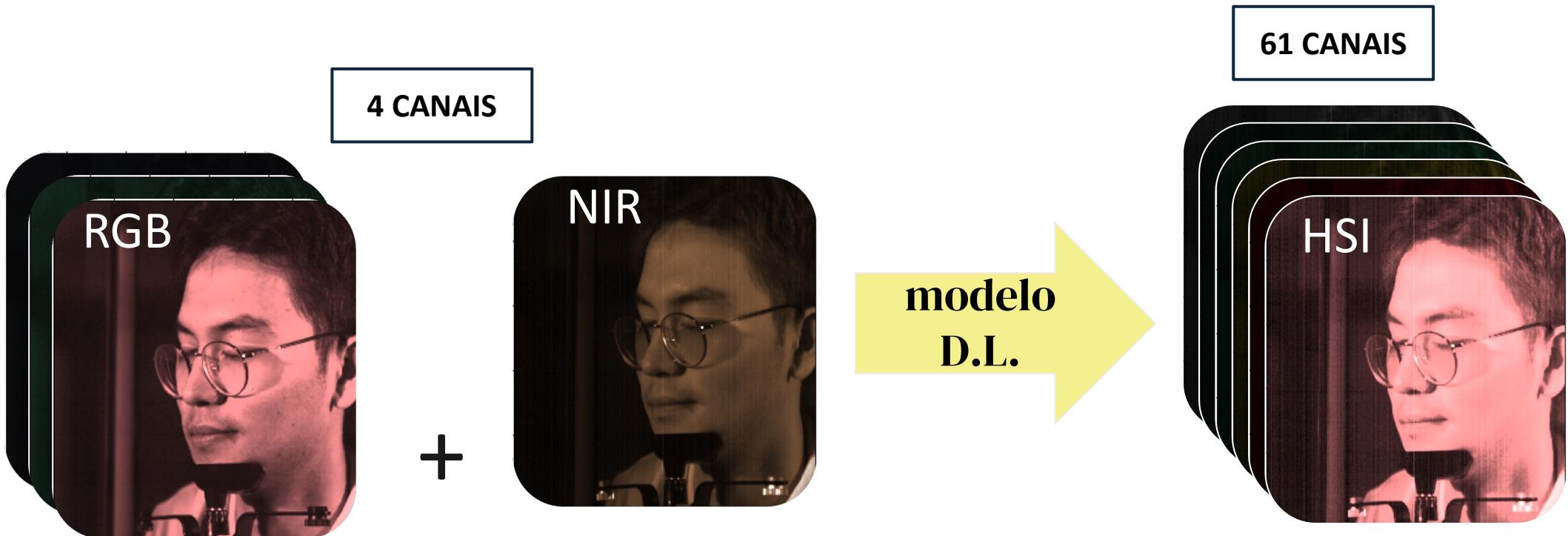
HIGH VARIABILITY



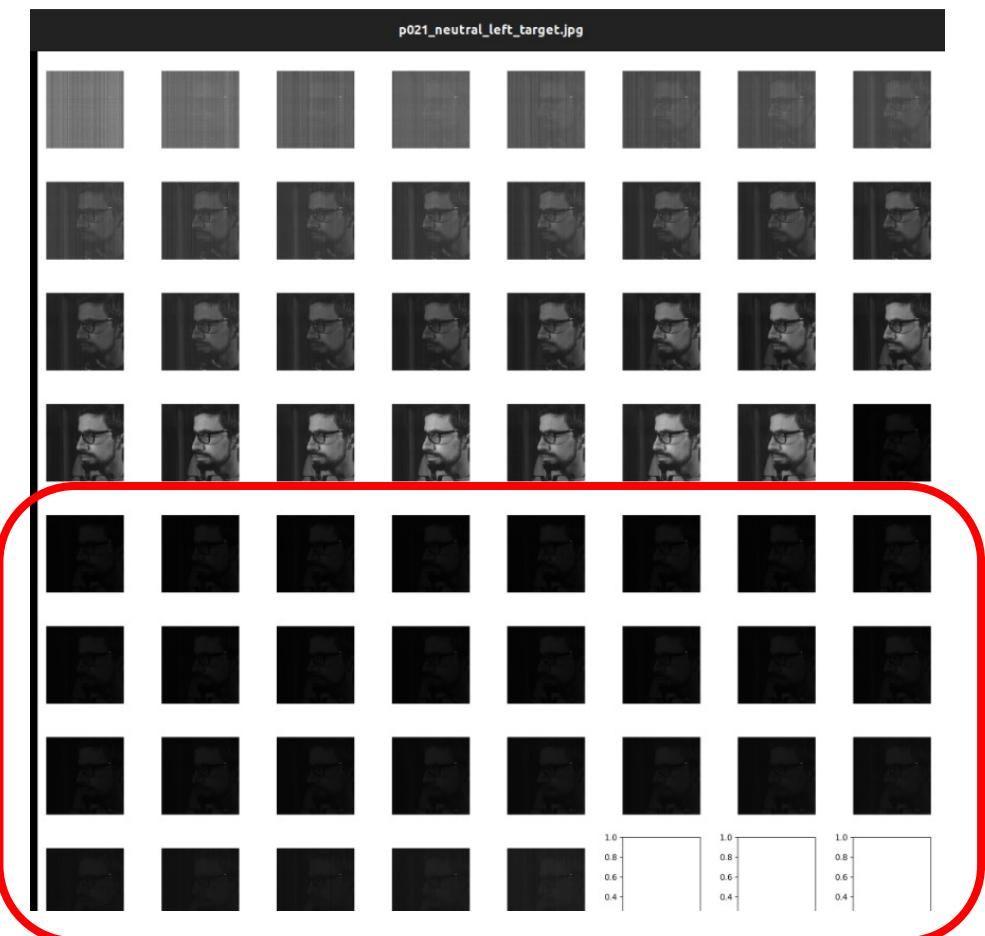
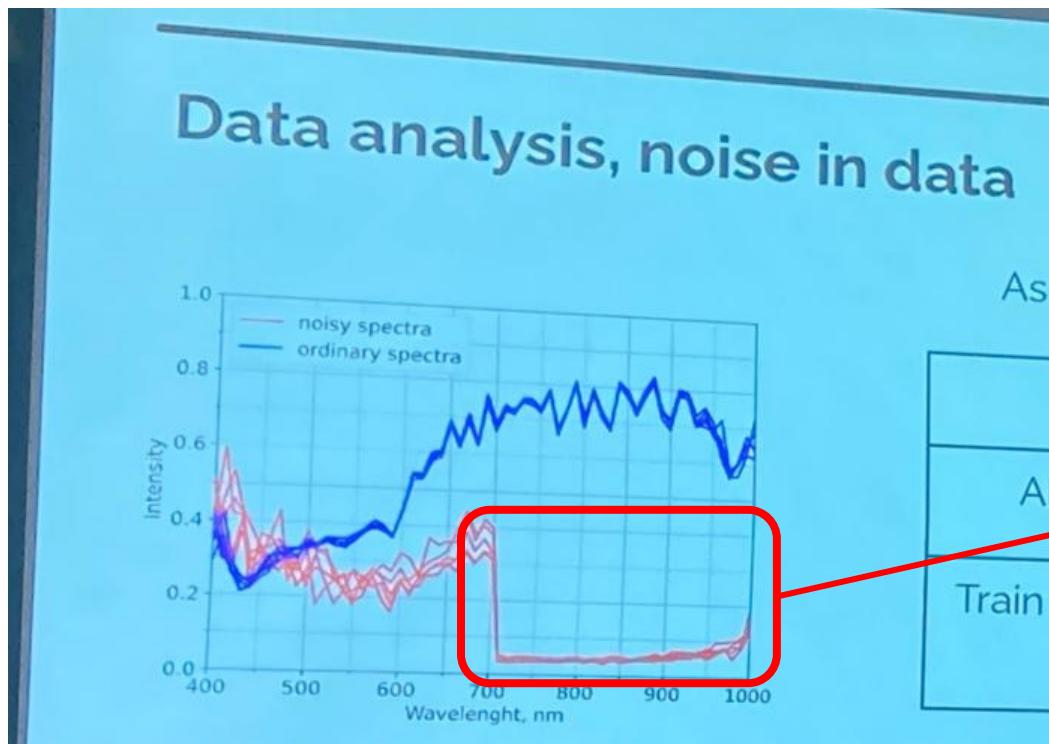
BAD DATA



Exemplo 2: Grand Challenge on Hyperspectral Skin Vision



The difference between these networks was in the training data. For *MST*-- training we used the whole training set and did neither augmentations nor input scaling. For *No-Noise-MST*-- training we dropped 18 pairs with very noisy 700..1000 nm range from training set. Examples of these noisy spectra



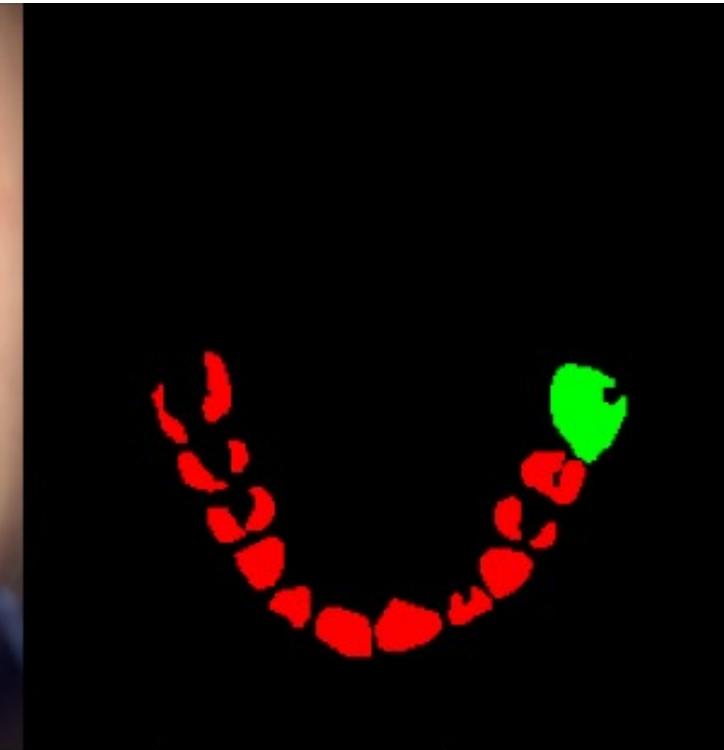
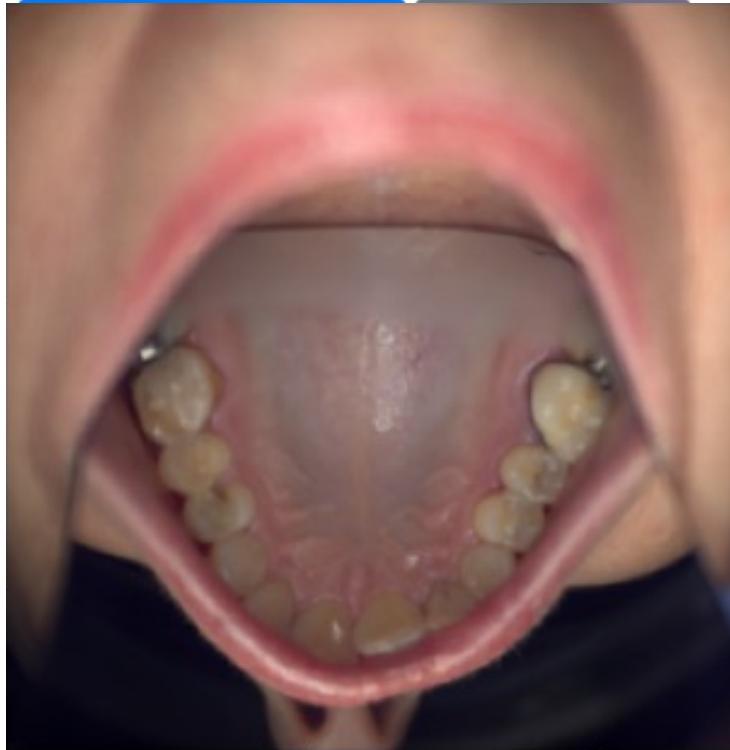


bda7a267_bottom_1

- Attached gingiva
- Calculus
- Dentine caries
- Enamel
- Lip
- Oral mucosa
- Out of focus area
- Plastic
- Skin
- Specular reflection
- Tongue

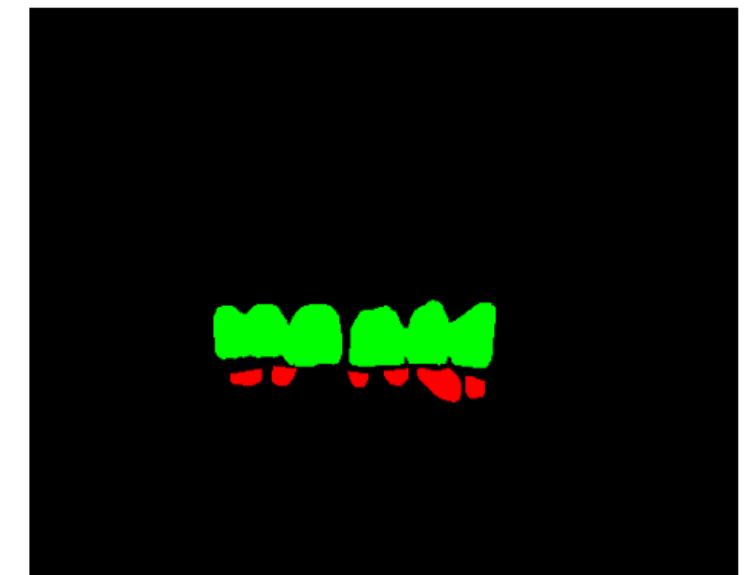
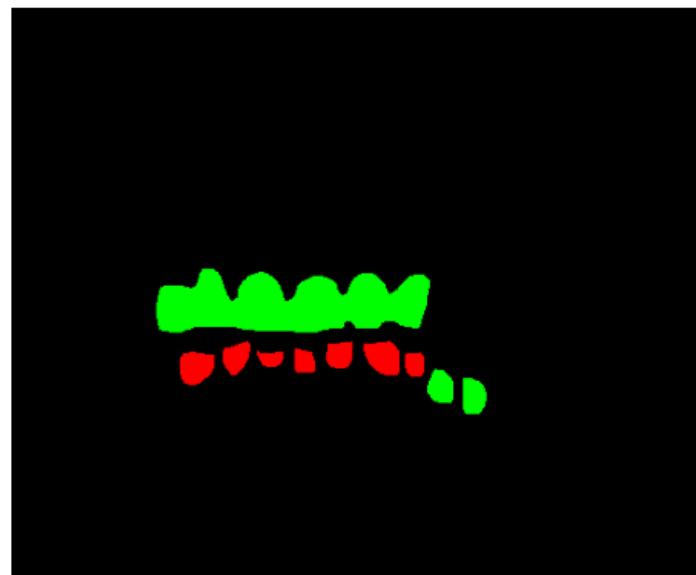


Exemplo 3: ODSI-DB Oral and Dental Spectral Image Database



Prótese

Enamel



■ Prótese ■ Enamel

Pipeline típico de análise



1. Entendendo o problema/tarefa
2. Curadoria dos dados
3. Questões éticas envolvendo os dados

Questões éticas



IMAGENS MÉDICAS: DADOS
SENSÍVEIS



APROVAÇÃO DO COMITÊ DE
ÉTICA



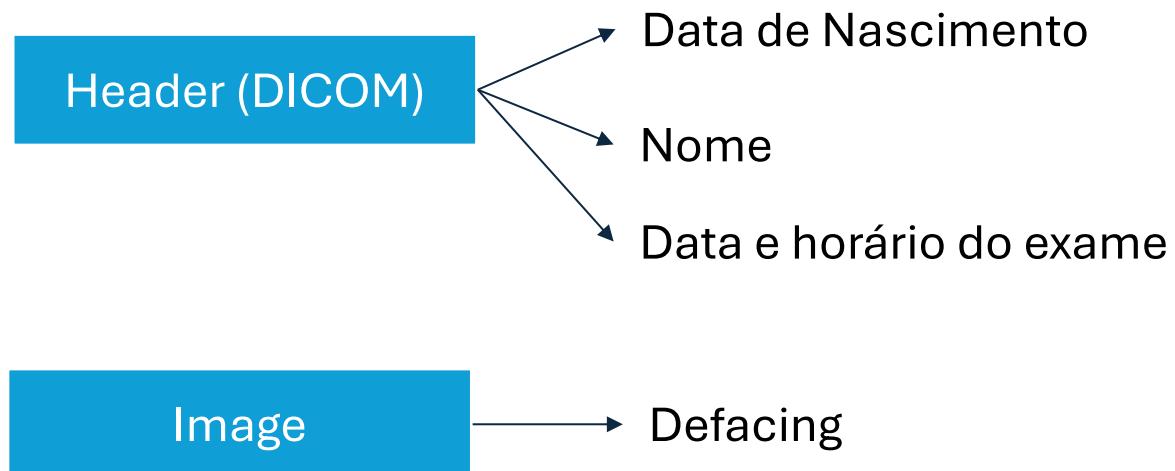
PROTEÇÃO DA PRIVACIDADE
DO PACIENTE

Questões éticas

Anonlização



Excluir as informações
que podem identificar o
sujeito

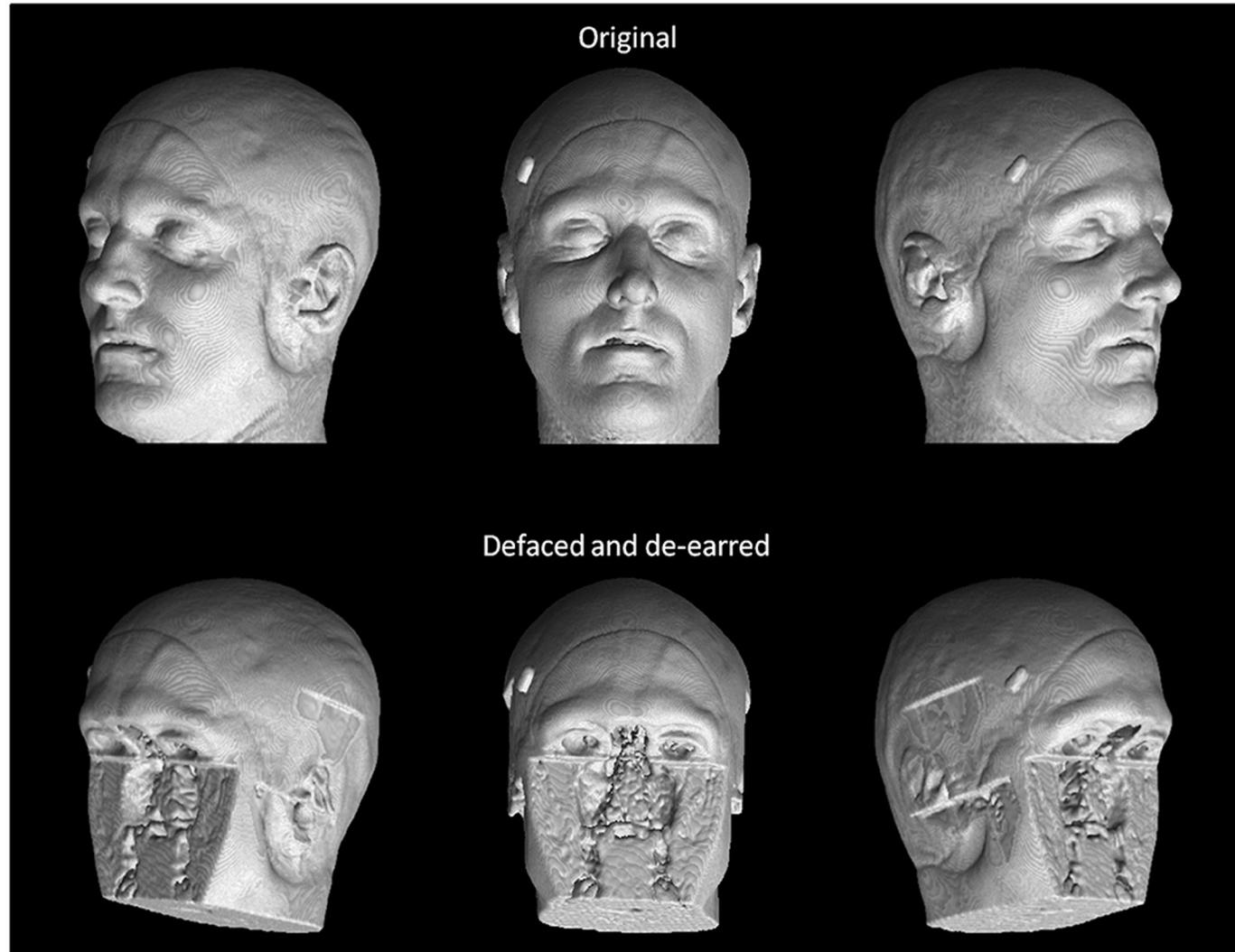


Questões éticas

Anonimização



Excluir as informações
que podem identificar o
sujeito



Pipeline típico de análise



Architectures

Data loading

Data augmentation

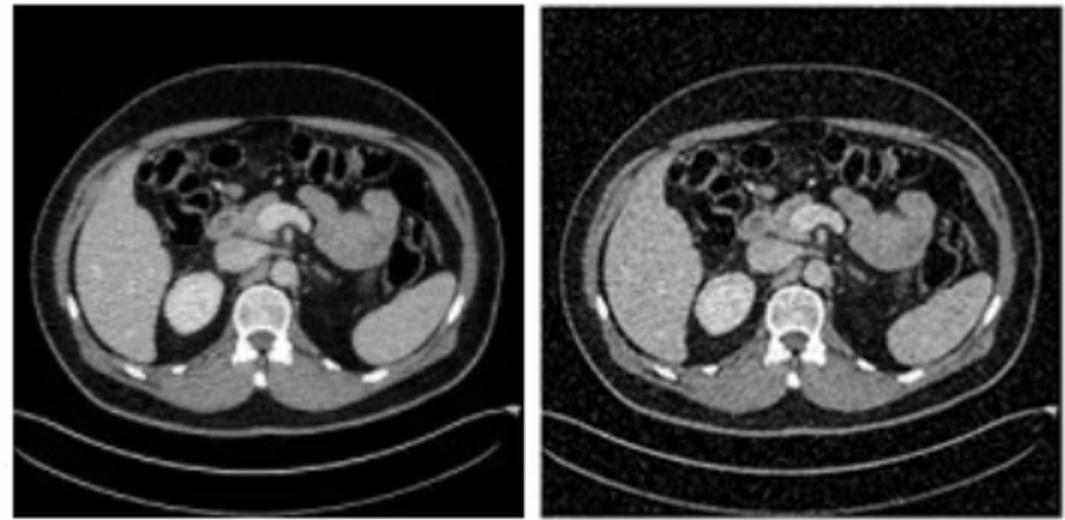
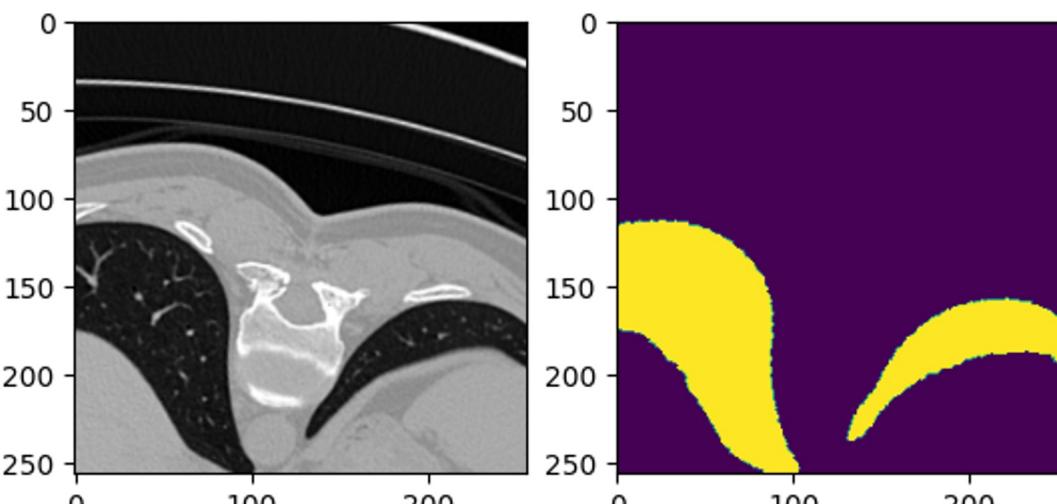
Loss Function

Optimizers

Frameworks

Aumento de dados durante o treinamento

Aumenta a variabilidade dos dados criando artificialmente casos de treinamento em tempo real (durante o treinamento)



Pre-processamento != aumento de dados

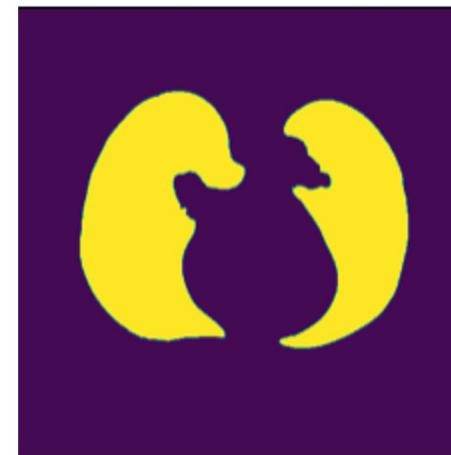
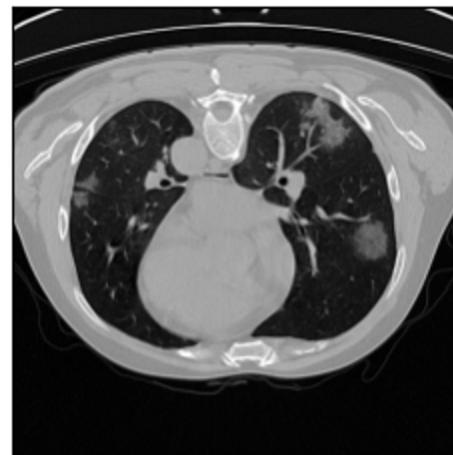
Processamento de dados fixos antes do treinamento
VS
Aumento de dados em tempo real na memória

Aumento de dados durante o treinamento

Certifique-se de visualizer se o dado entrando no seu modelo está correto!

Bugs em dados são difíceis de detectar.

A rede sempre vai tentar aprender, mesmo a partir de dados errados!



Você consegue perceber o que está errado com esse par?

Aumento de dados durante o treinamento

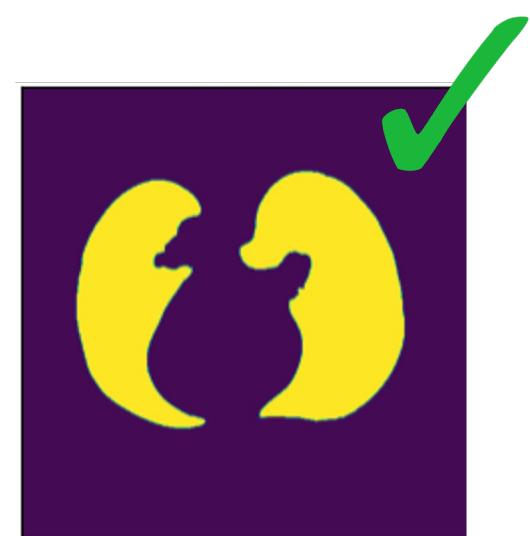
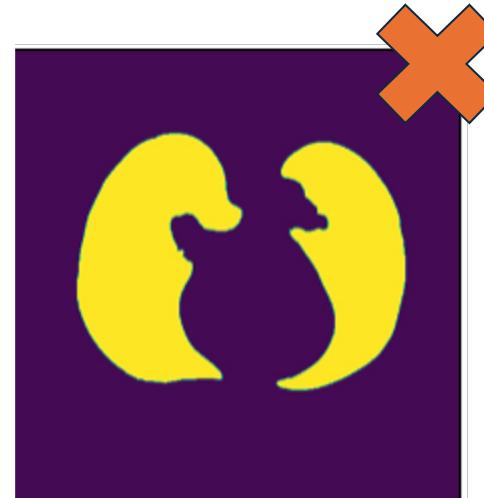
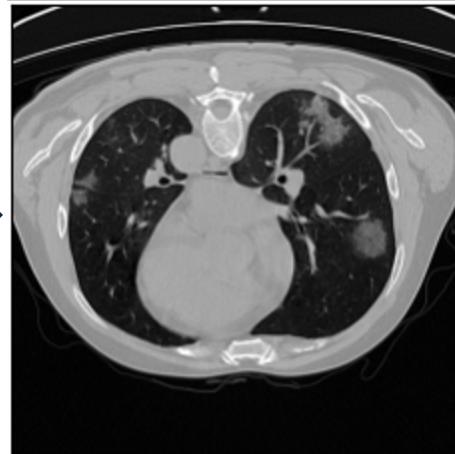
Certifique-se de visualizer se o dado entrando no seu modelo está correto!

Bugs em dados são difíceis de detectar.

A rede sempre vai tentar aprender, mesmo a partir de dados errados!



Flip
→

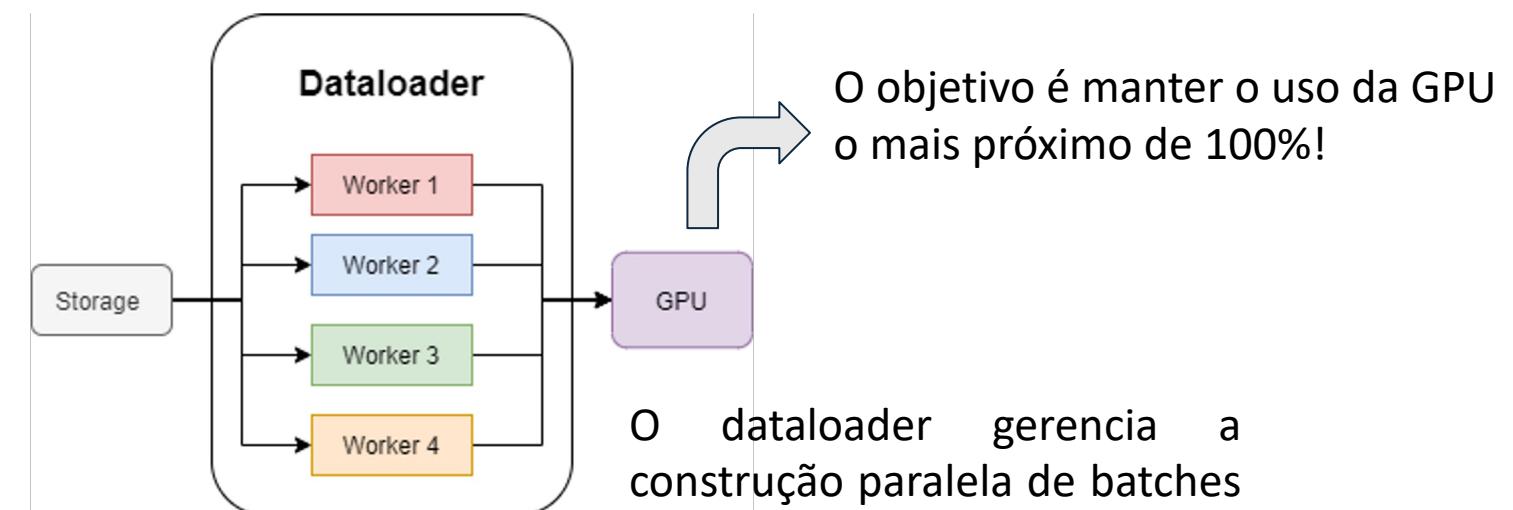
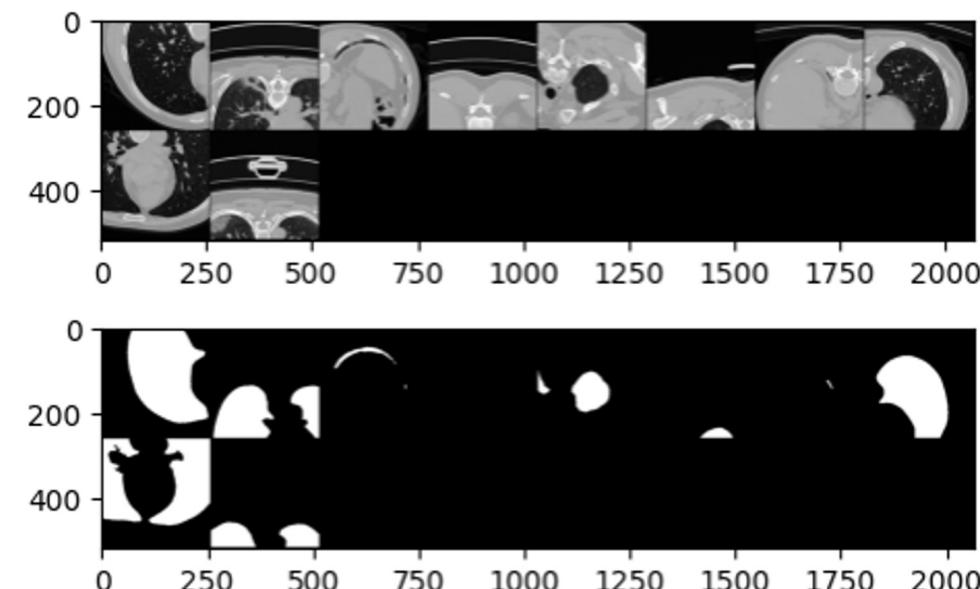


Apenas a imagem foi invertida horizontalmente!

Transformações espaciais precisam ser aplicadas às imagens e aos rótulos também.

Aumento de dados durante o treinamento

CNNs consomem lotes (batches) de imagens que são produzidos dataloaders.



O objetivo é manter o uso da GPU
o mais próximo de 100%!

O dataloader gerencia a
construção paralela de batches
de dados do armazenamento
para entregar para a GPU para
treinamento

Exemplo de um batch de 10 fatias de CT do pulmão com aumento de dados e suas máscaras binárias.

Agenda do curso



8h00 – Motivação: o uso de DL na análise de imagens médicas



9h00 – *Garbage in, Garbage out*

Diretrizes para o tratamento adequado dos dados
Parte prática: pré-processamento e uso dos dados



10h00 – Escolhendo o Modelo: o melhor? Depende...

Como funcionam as CNNs? E os Transformers?
Parte prática: Modelos multimodais envolvendo imagens e texto



11h00 – Avaliando os resultados: do's and dont's

Métricas são mais complicadas do que parecem
Parte prática: Dice não é tudo