

# TITULOOOOO

Hugo Veríssimo  
NOME DE CADEIRA ATUALIZAR24/25  
University of Aveiro  
Aveiro, Portugal  
hugoverissimo@ua.pt

João Cardoso  
Foundations of Machine Learning 24/25  
University of Aveiro  
Aveiro, Portugal  
[joaocardoso@ua.pt](mailto:joaocardoso@ua.pt)

### *Abstract*—abstract

**Keywords:** MovieLens, GroupLens, quais meter

## I. INTRODUCTION

lala intro

## II. METHODOLOGY

#### A. Data description

data

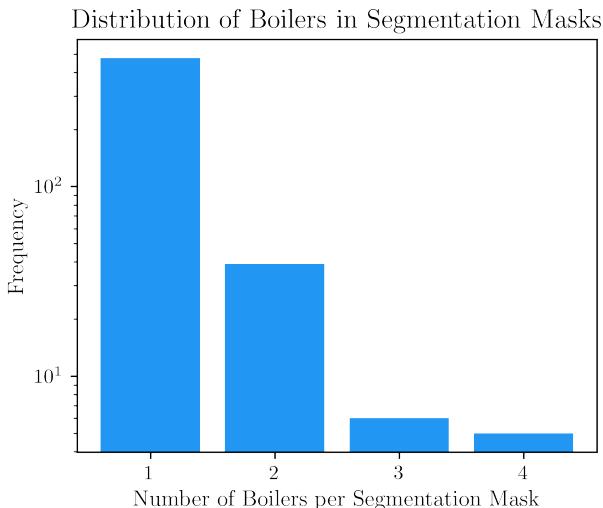


Fig. 1: CAPTION CAPTION CAPTION CAPTION CAPTION CAPTION

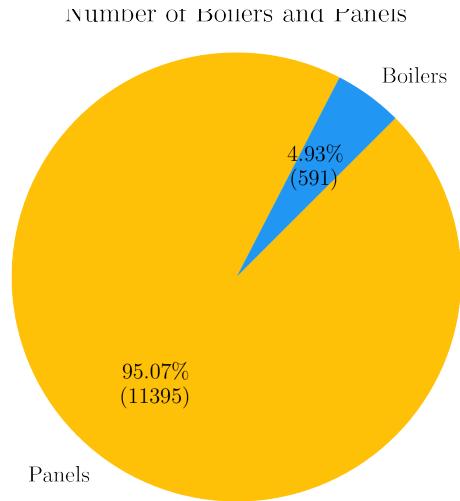


Fig. 2: CAPTION CAPTION CAPTION CAPTION CAPTION CAPTION



Fig. 3: CAPTION CAPTION CAPTION CAPTION CAPTION CAPTION

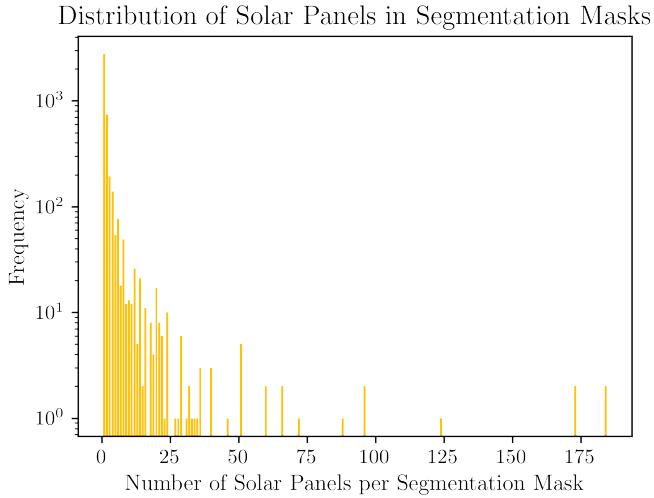


Fig. 4: CAPTION CAPTION CAPTION CAPTION CAPTION

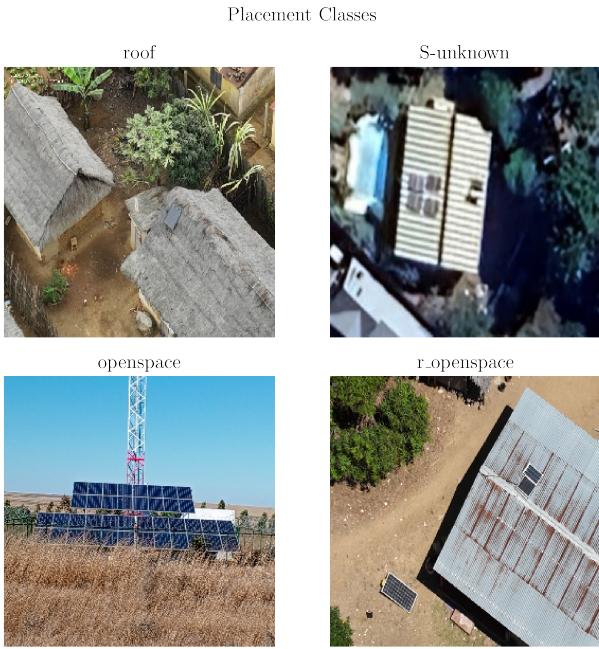


Fig. 5: CAPTION CAPTION CAPTION CAPTION CAPTION



Fig. 6: CAPTION CAPTION CAPTION CAPTION CAPTION CAPTION



Fig. 7: CAPTION CAPTION CAPTION CAPTION CAPTION CAPTION

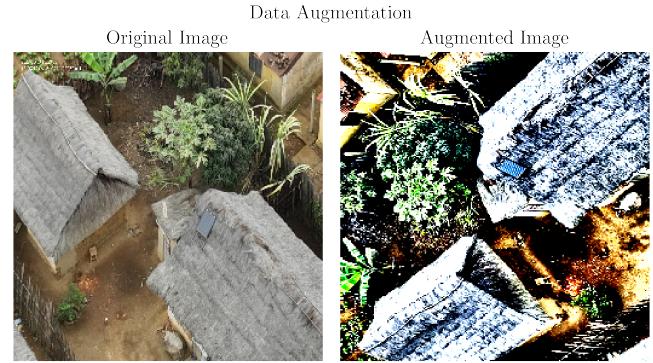


Fig. 8: CAPTION CAPTION CAPTION CAPTION CAPTION CAPTION

B. data: coisas pre processamento  
01 data v07:

- descrição dos dados, qnt e afins
- tratamento dos polígonos (str para lsita, tratar polígonos com letras no meio, juntar polígonos de imagens c o mesmo id, na mesma linha, pq antes era um polígono por linha)
- revisão manual de todas as imagens e correção de polígonos mal desenhados ou localizados nas imagens
- criação de classe python para correção dos polígonos e facilitar leitura

02 data v01:

- 263 imagens para o lixo
- separação em 80/20 do dataset em treino e validação, visto q o teste é dado pelo concurso, online
- e acho q é só

### C. Data splitting & models implemented

....

### D. Exploratory data analysis

aaa

## III. DJAISJD MODELS

llalal

### A. 03 model01 v04 - hugo

- ler ficheiro train e test pickle
- fez se data augmentation e resize das imagens para 240x240
- input foi so a imagem e output foi so nr boil, nr pan
- modelo EfficientNetV2B0, imagenet, sem mudar os pesos originais
- adicionei global avg pool, dense, dropout e dps output
- usei random search para encontrar units do dense e learning rate, se calhar devia ter usado para o dropout tbm, too late
- usei early stopping, paciente 3, restore best wiehgts
- loss Huber delta 1, compiler adamW, métrica para melhor: mae
- dps de escolher o melhor, tem uns graficos no notebook das metricas
- o modelo n aprendeu bem as features q devia, underfitting
- **public score:** 3.040397919

### B. zulo40 models type

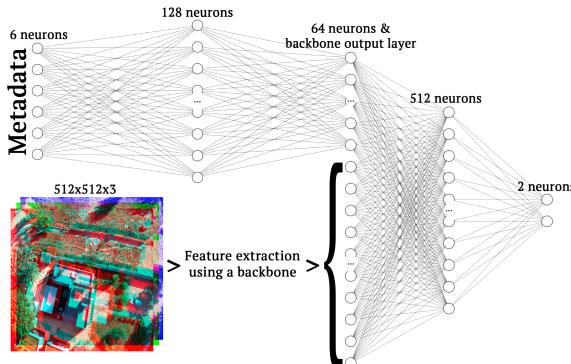


Fig. 9: CAPTION CAPTION CAPTION CAPTION CAPTION CAPTION

faz data augmentations

o cv q ele usa consiste em

- 3 folds

- para cada fold, um treino e um validação (80/20), ajustando um modelo em cada fold.

- no fim, junta os modelos (3, pq são 3 folds) e faz a média das previsões

- ou seja, usa 3 modelos como um só

ele usa Test-Time Augmentation (TTA), where the model makes multiple predictions on augmented versions of the same image (e.g., flipping, scaling, cropping). These predictions are later averaged to improve accuracy.

### C. yolo

ideia inicial era fazer segmentação de conjuntos de painéis e de boilers e posteriormente com outro modelo fazer contagem

problema foi q o modelo ao reconhecer painéis individuais, mesmo em conjuntos identificava individuais, levando ao mau desempenho do modelo

surgiu a ideia de criar 4 classes: boiler, panel, conjunto de boîleres, conjunto de panels, so q devido ao class imbalance o modelo teve resultados não satisfatórios o suficiente

### METER GRAFICO IMABALANCE

então decidiu-se rever o dataset e separar manualmente os polígonos em painéis e boîleres individuais, deixando de existir grupos de painéis

o yolo apresentou assim mt melhores resultados, apesar da redução da dimensão do dataset em termos de imagens mas aumento de amostras do que realmente são painéis e boîleres individuais

fazer ja learning curve ou correr de novo last yolo?

IDEIA: correr de novo o last yolo, mas mudar hiperparâmetros (learning rate, adamw, etc.), a ver se temos uma solução melhor

## IV. FAA02 IG

TABLE I: Error metrics for the train and test set (CF-LR), along with the number of samples for each.

Dataset	RMSE	MAE	Support
Train Set	0.58842	0.44604	80668
Test Set	1.24037	0.90651	20168

## V. DISCUSSION

### A. Performance Metrics

discussão

## VI. CONCLUSION

conc

### WORK LOAD

Both authors contributed equally to the project.

## REFERENCES