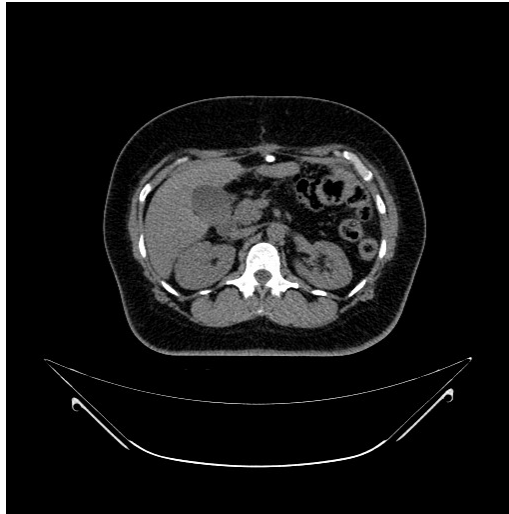


# Dicom Detect with YOLOv7

Luiz Castiglioni

04/10/2022



Trabalho de conclusão de curso do aluno Luiz Castiglioni no curso do MBA - Data Science and Analytics da USP / ESALQ. Trabalho desenvolvido em parceria com o departamento de AI da Escola Paulista de Medicina UNIFESP.

## Especificações do Projeto

### Data info

Tomographic images and Metadata of Medical Exams in DICOM Standards Format

- total files: 165371
- total size: 89.9GB
- format: “.dcm”

## Links

- Dicom Medicine Standards <https://www.dicomstandard.org/ai>
- YOLO <https://pjreddie.com/darknet/yolo/>
- Github [https://github.com/luizcast/DCM\\_TCC](https://github.com/luizcast/DCM_TCC)

## Arquivos do projeto

```
list.files(include.dirs = TRUE)
```

```
## [1] "1PRIMEIRO_SCRIPT.R"      "2SEGUNDO_SCRIPT.R"
## [3] "3TERCEIRO_SCRIPT.R"      "4QUARTO_SCRIPT.R"
## [5] "5QUINTO_SCRIPT.R"        "Anot_IMGS.R"
## [7] "dcm_tools.Rproj"         "IMGs"
## [9] "keras.R"                 "markdown_dicom_yolo_files"
## [11] "markdown_dicom_yolo.docx" "markdown_dicom_yolo.html"
## [13] "markdown_dicom_yolo.pdf"  "markdown_dicom_yolo.Rmd"
## [15] "rascunho.R"              "README.md"
## [17] "TABLEs"                  "yolo_dicom_v1.pt"
```

## Configurações da Máquina

```
neofetch --stdout
```

```
## luiz@box
## -----
## OS: Linux Mint 21 x86_64
## Host: Dell G15 5515 1.7.1
## Kernel: 5.15.0-48-generic
## Uptime: 3 hours, 1 min
## Packages: 3375 (dpkg), 9 (flatpak), 11 (snap)
## Shell: bash 5.1.16
## Resolution: 1920x1080
## DE: Cinnamon
## WM: Mutter
## WM Theme: Mint-Y
## Theme: Mint-Y-Dark-Purple [GTK2/3]
## Icons: Mint-Y-Dark-Blue [GTK2/3]
## Terminal: R
## CPU: AMD Ryzen 7 5800H with Radeon Graphics (16) @ 4.463GHz
## GPU: NVIDIA GeForce RTX 3060 Mobile / Max-Q
## GPU: AMD ATI 06:00.0 Cezanne
## Memory: 5253MiB / 15342MiB
```

```
nvidia-smi
```

```
## Tue Oct  4 14:43:30 2022
## +-----+
## | NVIDIA-SMI 515.65.01    Driver Version: 515.65.01    CUDA Version: 11.7    |
```

```
## |-----+-----+-----+
## | GPU Name Persistence-M| Bus-Id Disp.A | Volatile Uncorr. ECC |
## | Fan Temp Perf Pwr:Usage/Cap| Memory-Usage | GPU-Util Compute M. |
## | | | | MIG M. |
## |=====+=====+=====+
## | 0 NVIDIA GeForce ... Off | 00000000:01:00.0 Off | N/A |
## | N/A 52C P5 18W / N/A | 656MiB / 6144MiB | 56% Default |
## | | | | N/A |
## +-----+-----+-----+
##
## +-----+-----+-----+
## | Processes: |
## | GPU GI CI PID Type Process name GPU Memory |
## | ID ID | | | | Usage |
## |=====+=====+=====+
## | 0 N/A N/A 1334 G /usr/lib/xorg/Xorg 210MiB |
## | 0 N/A N/A 2608 G cinnamon 45MiB |
## | 0 N/A N/A 14319 G ...335803002064937364,131072 21MiB |
## | 0 N/A N/A 22295 G rstudio 60MiB |
## | 0 N/A N/A 22347 C .../lib/rstudio/bin/rsession 273MiB |
## | 0 N/A N/A 35423 G ...974436258714513412,131072 38MiB |
## +-----+-----+-----+
```

## Configurações do RStudio

```
#RStudio 2022.07.2+576 "Spotted Wakerobin"
#Release (e7373ef832b49b2a9b88162cfe7eac5f22c40b34, 2022-09-06) for Ubuntu Bionic
version
```

```
##
## platform x86_64-pc-linux-gnu
## arch x86_64
## os linux-gnu
## system x86_64, linux-gnu
## status
## major 4
## minor 2.1
## year 2022
## month 06
## day 23
## svn rev 82513
## language R
## version.string R version 4.2.1 (2022-06-23)
## nickname Funny-Looking Kid
```

## Pacotes utilizados no projeto

- oro.dicom - <https://cran.r-project.org/web/packages/oro.dicom/index.html>
- dcmkt - <https://dicom.offis.de/dcmkt.php.en>
- tidyverse - <https://www.tidyverse.org/>
- tensorflow - <https://www.tensorflow.org/>

- keras - <https://keras.io/>

```
pacotes <- c("keras", "tensorflow", "tidyverse", "oro.dicom", "dcmtk", "readxl")
if(sum(as.numeric(!pacotes %in% installed.packages())) != 0){
  instalador <- pacotes[!pacotes %in% installed.packages()]
  for(i in 1:length(instalador)) {
    install.packages(instalador, dependencies = T)
    break()}
  sapply(pacotes, require, character = T)
} else {
  sapply(pacotes, require, character = T)
}
```

```
## Loading required package: keras
```

```
## Loading required package: tensorflow
```

```
## Loading required package: tidyverse
```

```
## -- Attaching packages ----- tidyverse 1.3.2 --
```

```
## v ggplot2 3.3.6      v purrr 0.3.4
```

```
## v tibble 3.1.8      v dplyr 1.0.10
```

```
## v tidyr 1.2.0      v stringr 1.4.0
```

```
## v readr 2.1.2      v forcats 0.5.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag() masks stats::lag()
```

```
## Loading required package: oro.dicom
```

```
##
```

```
## oro.dicom 0.5.3
```

```
##
```

```
## Loading required package: dcmtk
```

```
##
```

```
## Loading required package: readxl
```

```
##      keras tensorflow tidyverse oro.dicom      dcmtk      readxl
```

```
##      TRUE      TRUE      TRUE      TRUE      TRUE      TRUE
```

```
tf_gpu_configured()
```

```
## Loaded Tensorflow version 2.9.2
```

```
## TensorFlow built with CUDA: TRUE
```

```
## GPU device name: /device:GPU:0
```

```
## [1] TRUE
```

## Reconhecendo Padrões em imagens de Tomografia de Abdômen no formato DICOM.

Foram selecionados 200 exames de tomografia de abdômen de 200 pacientes diferentes. Desses 200 exames foram geradas anotações pelo grupo de médicos e residentes da UNIFESP sobre a presença do órgão vesícula ou a presença do clipe cirúrgico em casos de pacientes que tiveram a vesícula retirada. Surgiram 4 classes:

- Pacientes com Vesícula Normal
- Pacientes com Vesícula Hipodistendida
- Pacientes sem Vesícula e com Clipe Cirúrgico
- Pacientes sem Vesícula e sem Clipe Cirúrgico

Destas anotações surgiu a tabela **annot\_clean**

```
read_excel("~/Desktop/USP/TCC/projeto_TCC/TABLEs/annot_clean.xlsx") %>% as_tibble() -> annot_clean
annot_clean
```

```
## # A tibble: 200 x 10
##   Acession Num-1 Patie-2 Corte-3 Corte-4 Hipod-5 Densi-6 Sem c-7 Arter-8 Portal
##   <dbl> <dbl> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl>
## 1 5499993 3861927 Não se~ Não se~ S Não se~ 1 1 1
## 2 5493468 823083 Não se~ Não se~ C Não se~ 1 1 1
## 3 5480343 3586593 Não se~ Não se~ S Não se~ 0 1 1
## 4 5481425 2863787 Não se~ Não se~ C Não se~ 0 0 1
## 5 5480051 3880103 Não se~ Não se~ C Não se~ 0 0 1
## 6 5459110 3825619 Não se~ Não se~ S Não se~ 0 0 1
## 7 5465667 3132902 Não se~ Não se~ C Não se~ 1 1 1
## 8 5466142 3872936 Não se~ Não se~ C Não se~ 1 0 1
## 9 5435830 3227017 Não se~ Não se~ C Não se~ 1 1 1
## 10 5433748 870140 Não se~ Não se~ S Não se~ 0 0 1
## # ... with 190 more rows, 1 more variable: Equilibrio <dbl>, and abbreviated
## # variable names 1: 'Acession Number', 2: 'Patient ID',
## # 3: 'Corte INICIAL vesícula', 4: 'Corte FINAL vesícula',
## # 5: 'Hipodistendida/Vesícula normal/Clipe/sem clipe (HVSC)',
## # 6: 'Densidade não habitual (sim=1, não=0)',
## # 7: 'Sem contraste (sim = 1; não = 0)', 8: Arterial
```

```
colnames(annot_clean)
```

```
## [1] "Acession Number"
## [2] "Patient ID"
## [3] "Corte INICIAL vesícula"
## [4] "Corte FINAL vesícula"
## [5] "Hipodistendida/Vesícula normal/Clipe/sem clipe (HVSC)"
## [6] "Densidade não habitual (sim=1, não=0)"
## [7] "Sem contraste (sim = 1; não = 0)"
## [8] "Arterial"
## [9] "Portal"
## [10] "Equilibrio"
```

## Preparando a Base Annot\_Clean

```
gsub(pattern = "Não se aplica", replacement = "NA",
      annot_clean$`Densidade não habitual (sim=1, não=0)` ->
      annot_clean$`Densidade não habitual (sim=1, não=0)`
gsub(pattern = "Não se aplica", replacement = "NA",
      annot_clean$`Corte INICIAL vesícula` ->
      annot_clean$`Corte INICIAL vesícula`
gsub(pattern = "Não se aplica", replacement = "NA",
      annot_clean$`Corte FINAL vesícula` ->
      annot_clean$`Corte FINAL vesícula`
annot_clean$`Hipodistendida/Vesícula normal/Clipe/sem clipe (HVSC)` <-
  as.factor(annot_clean$`Hipodistendida/Vesícula normal/Clipe/sem clipe (HVSC)`)
annot_clean$`Corte FINAL vesícula` <-
  as.numeric(annot_clean$`Corte FINAL vesícula`)
annot_clean$`Corte INICIAL vesícula` <-
  as.numeric(annot_clean$`Corte INICIAL vesícula`)
annot_clean$`Densidade não habitual (sim=1, não=0)` <-
  as.numeric(annot_clean$`Densidade não habitual (sim=1, não=0)`)
annot_clean
```

```
## # A tibble: 200 x 10
##   Acession Num~1 Patie~2 Corte~3 Corte~4 Hipod~5 Densi~6 Sem c~7 Arter~8 Portal
##         <dbl>   <dbl>   <dbl>   <dbl> <fct>     <dbl>   <dbl>   <dbl>   <dbl>
## 1       5499993 3861927     NA     NA S         NA       1       1       1
## 2       5493468 823083     NA     NA C         NA       1       1       1
## 3       5480343 3586593     NA     NA S         NA       0       1       1
## 4       5481425 2863787     NA     NA C         NA       0       0       1
## 5       5480051 3880103     NA     NA C         NA       0       0       1
## 6       5459110 3825619     NA     NA S         NA       0       0       1
## 7       5465667 3132902     NA     NA C         NA       1       1       1
## 8       5466142 3872936     NA     NA C         NA       1       0       1
## 9       5435830 3227017     NA     NA C         NA       1       1       1
## 10      5433748 870140     NA     NA S         NA       0       0       1
## # ... with 190 more rows, 1 more variable: Equilibrio <dbl>, and abbreviated
## #   variable names 1: 'Acession Number', 2: 'Patient ID',
## #   3: 'Corte INICIAL vesícula', 4: 'Corte FINAL vesícula',
## #   5: 'Hipodistendida/Vesícula normal/Clipe/sem clipe (HVSC)',
## #   6: 'Densidade não habitual (sim=1, não=0)',
## #   7: 'Sem contraste (sim = 1; não = 0)', 8: Arterial
```

## Criando Dataset

```
path = "/home/luiz/Desktop/DATA-SET_TESTE/"
list.files(path) %>% as_tibble() -> array_arquivos_dicom_tratados
separate(array_arquivos_dicom_tratados, col = value, sep = "_", into = c("PatientID", "SerieNumber", "InstanceNumber"))
index_dicom$InstanceNumber %>% str_sub(end=-5) -> index_dicom$InstanceNumber
bind_cols(array_arquivos_dicom_tratados, index_dicom) -> index_dicom
names(index_dicom) <- c("FileName", "PatientID", "SerieNumber", "InstanceNumber")
index_dicom$PatientID <- as.numeric(index_dicom$PatientID)
index_dicom$SerieNumber = as.numeric(index_dicom$SerieNumber)
```

```
index_dicom$InstanceNumber = as.numeric(index_dicom$InstanceNumber)
glimpse(index_dicom)
```

```
## Rows: 153,435
## Columns: 4
## $ FileName      <chr> "0000121029_2_1.dcm", "0000121029_2_10.dcm", "000012102~
## $ PatientID     <dbl> 121029, 121029, 121029, 121029, 121029, 121029, 121029,~
## $ SerieNumber   <dbl> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2~
## $ InstanceNumber <dbl> 1, 10, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109~
```

## Join das tabelas

```
index_dicom %>% group_by(PatientID) %>%
  left_join(annot_clean, by = c("PatientID" = "Patient ID")) %>%
  ungroup() -> index_dicom
index_dicom$FilePath <- paste(path, index_dicom$FileName, sep='')
index_dicom %>% dplyr::select(1:9) -> base_tratada
base_tratada %>% dplyr::select(-5) -> base_tratada
base_tratada$FilePath <- paste(path, base_tratada$FileName, sep='')
base_tratada
```

```
## # A tibble: 165,794 x 9
##   FileName      Patie~1 Serie~2 Insta~3 Corte~4 Corte~5 Hipod~6 Densi~7 FileP~8
##   <chr>          <dbl>   <dbl>   <dbl>   <dbl>   <dbl> <fct>   <dbl> <chr>
## 1 0000121029_2~ 121029     2       1      NA      NA S      NA /home/~
## 2 0000121029_2~ 121029     2      10      NA      NA S      NA /home/~
## 3 0000121029_2~ 121029     2     100      NA      NA S      NA /home/~
## 4 0000121029_2~ 121029     2     101      NA      NA S      NA /home/~
## 5 0000121029_2~ 121029     2     102      NA      NA S      NA /home/~
## 6 0000121029_2~ 121029     2     103      NA      NA S      NA /home/~
## 7 0000121029_2~ 121029     2     104      NA      NA S      NA /home/~
## 8 0000121029_2~ 121029     2     105      NA      NA S      NA /home/~
## 9 0000121029_2~ 121029     2     106      NA      NA S      NA /home/~
## 10 0000121029_2~ 121029     2     107      NA      NA S      NA /home/~
## # ... with 165,784 more rows, and abbreviated variable names 1: PatientID,
## # 2: SerieNumber, 3: InstanceNumber, 4: 'Corte INICIAL vesícula',
## # 5: 'Corte FINAL vesícula',
## # 6: 'Hipodistendida/Vesícula normal/Clipe/sem clipe (HVSC)',
## # 7: 'Densidade não habitual (sim=1, não=0)', 8: FilePath
```

## Extraindo e Comparando Tabelas

```
#TABELA FREQUENCIA
table(annot_clean$`Hipodistendida/Vesícula normal/Clipe/sem clipe (HVSC)`)
```

```
##
## C H S V
## 73 10 27 90
```

#### #N DE IMAGENS POR SERIENUMBER

```
base_tratada %>% group_by(PatientID) %>% count(SerieNumber) %>% ungroup(PatientID)
```

```
## # A tibble: 520 x 3
##   PatientID SerieNumber     n
##   <dbl>         <dbl> <int>
## 1 121029           2    150
## 2 121029           3    448
## 3 483240           4    285
## 4 483240           5    160
## 5 483240           6    478
## 6 548545           2    162
## 7 548545           3    484
## 8 604664           2    155
## 9 604664           8    461
## 10 604664          9    155
## # ... with 510 more rows
```

#### #TABELA SEM VESICULA COM CLIPE

```
base_clipe <- base_tratada %>% filter(`Hipodistendida/Vesícula normal/Clipe/sem clipe (HVSC)` == "C")
glimpse(base_clipe)
```

```
## Rows: 55,309
## Columns: 9
## $ FileName                <chr> "0000548545_2_~
## $ PatientID               <dbl> 548545, 548545~
## $ SerieNumber             <dbl> 2, 2, 2, 2, 2,~
## $ InstanceNumber          <dbl> 1, 10, 100, 10~
## $ 'Corte INICIAL vesícula' <dbl> NA, NA, NA, NA~
## $ 'Corte FINAL vesícula'  <dbl> NA, NA, NA, NA~
## $ 'Hipodistendida/Vesícula normal/Clipe/sem clipe (HVSC)' <fct> C, C, C, C, C,~
## $ 'Densidade não habitual (sim=1, não=0)' <dbl> NA, NA, NA, NA~
## $ FilePath                <chr> "/home/luiz/De~
```

#### #TABELA SEM VESICULA SEM CLIPE

```
base_sem_clipe <- base_tratada %>% filter(`Hipodistendida/Vesícula normal/Clipe/sem clipe (HVSC)` == "S")
glimpse(base_sem_clipe)
```

```
## Rows: 27,421
## Columns: 9
## $ FileName                <chr> "0000121029_2_~
## $ PatientID               <dbl> 121029, 121029~
## $ SerieNumber             <dbl> 2, 2, 2, 2, 2,~
## $ InstanceNumber          <dbl> 1, 10, 100, 10~
## $ 'Corte INICIAL vesícula' <dbl> NA, NA, NA, NA~
## $ 'Corte FINAL vesícula'  <dbl> NA, NA, NA, NA~
## $ 'Hipodistendida/Vesícula normal/Clipe/sem clipe (HVSC)' <fct> S, S, S, S, S,~
## $ 'Densidade não habitual (sim=1, não=0)' <dbl> NA, NA, NA, NA~
## $ FilePath                <chr> "/home/luiz/De~
```



#### #TABELA COM VESICULA HIPOESTENDIDA

```
base_vesicula_hipodistendida <- base_tratada %>% filter(`Hipodistendida/Vesícula normal/Clipe/sem clipe`  
glimpse(base_vesicula_hipodistendida)
```

```
## Rows: 8,353  
## Columns: 9  
## $ FileName                <chr> "0002073649_3_~  
## $ PatientID               <dbl> 2073649, 20736~  
## $ SerieNumber             <dbl> 3, 3, 3, 3, 3,~  
## $ InstanceNumber          <dbl> 1, 10, 100, 10~  
## $ 'Corte INICIAL vesícula' <dbl> 48, 48, 48, 48~  
## $ 'Corte FINAL vesícula'  <dbl> 60, 60, 60, 60~  
## $ 'Hipodistendida/Vesícula normal/Clipe/sem clipe (HVSC)' <fct> H, H, H, H, H,~  
## $ 'Densidade não habitual (sim=1, não=0)' <dbl> 1, 1, 1, 1, 1,~  
## $ FilePath                <chr> "/home/luiz/De~
```

#### #TABELA COM VESICULA NORMAL

```
base_vesicula_normal <- base_tratada %>% filter(`Hipodistendida/Vesícula normal/Clipe/sem clipe (HVSC)`  
glimpse(base_vesicula_normal)
```

```
## Rows: 71,244  
## Columns: 9  
## $ FileName                <chr> "0000483240_4_~  
## $ PatientID               <dbl> 483240, 483240~  
## $ SerieNumber             <dbl> 4, 4, 4, 4, 4,~  
## $ InstanceNumber          <dbl> 1, 10, 100, 10~  
## $ 'Corte INICIAL vesícula' <dbl> 40, 40, 40, 40~  
## $ 'Corte FINAL vesícula'  <dbl> 56, 56, 56, 56~  
## $ 'Hipodistendida/Vesícula normal/Clipe/sem clipe (HVSC)' <fct> V, V, V, V, V,~  
## $ 'Densidade não habitual (sim=1, não=0)' <dbl> 1, 1, 1, 1, 1,~  
## $ FilePath                <chr> "/home/luiz/De~
```

#### #TABELA COM AS IMAGENS DA VESICULA

```
base_vesicula_normal %>% mutate(posicao_interesse = ifelse(c(InstanceNumber >= `Corte INICIAL vesícula`  
base_frames_vesicula_normal %>% filter(posicao_interesse == 1) -> base_frames_vesicula_normal  
glimpse(base_frames_vesicula_normal)
```

```
## Rows: 4,110  
## Columns: 10  
## $ FileName                <chr> "0000483240_4_~  
## $ PatientID               <dbl> 483240, 483240~  
## $ SerieNumber             <dbl> 4, 4, 4, 4, 4,~  
## $ InstanceNumber          <dbl> 40, 41, 42, 43~  
## $ 'Corte INICIAL vesícula' <dbl> 40, 40, 40, 40~  
## $ 'Corte FINAL vesícula'  <dbl> 56, 56, 56, 56~  
## $ 'Hipodistendida/Vesícula normal/Clipe/sem clipe (HVSC)' <fct> V, V, V, V, V,~  
## $ 'Densidade não habitual (sim=1, não=0)' <dbl> 1, 1, 1, 1, 1,~  
## $ FilePath                <chr> "/home/luiz/De~  
## $ posicao_interesse        <dbl> 1, 1, 1, 1, 1,~
```