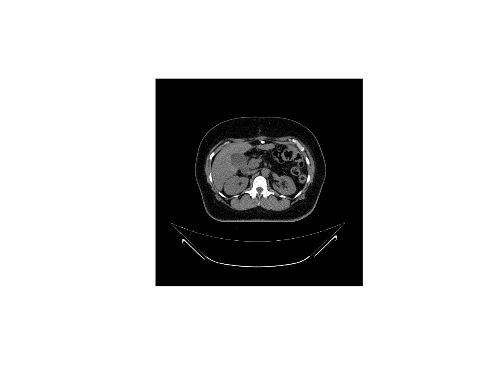
Dicom Detect with YOLOv7

Luiz Castiglioni

06/10/2022



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>

### 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, 25 mins   
## Packages: 3380 (dpkg), 13 (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: 5168MiB / 15342MiB

nvidia-smi

## Thu Oct 6 11:37:34 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 50C P5 16W / N/A | 513MiB / 6144MiB | 27% Default |  
## | | | N/A |  
## +-------------------------------+----------------------+----------------------+  
##   
## +-----------------------------------------------------------------------------+  
## | Processes: |  
## | GPU GI CI PID Type Process name GPU Memory |  
## | ID ID Usage |  
## |=============================================================================|  
## | 0 N/A N/A 1286 G /usr/lib/xorg/Xorg 246MiB |  
## | 0 N/A N/A 2544 G cinnamon 90MiB |  
## | 0 N/A N/A 11427 G ...436071394914752260,131072 10MiB |  
## | 0 N/A N/A 15891 G ...843406173015110053,131072 82MiB |  
## | 0 N/A N/A 26126 G /usr/lib/rstudio/bin/rstudio 79MiB |  
## +-----------------------------------------------------------------------------+

### 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>
* dcmtk - <https://dicom.offis.de/dcmtk.php.en>
* tidyverse - <https://www.tidyverse.org/>
* tensorflow - <https://www.tensorflow.org/>
* keras - <https://keras.io/>

library("keras")   
library("tensorflow")  
library("tidyverse")

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.2 ──  
## ✔ ggplot2 3.3.6 ✔ purrr 0.3.4   
## ✔ tibble 3.1.8 ✔ dplyr 1.0.10  
## ✔ tidyr 1.2.0 ✔ stringr 1.4.0   
## ✔ readr 2.1.2 ✔ forcats 0.5.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()

library("oro.dicom")

## oro.dicom 0.5.3

library("dcmtk")  
library("readxl")

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

Foram selecionados 200 exames de tomografia de abdomen 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 × 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 × 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  
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, 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 × 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 × 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 (HVSC)` == "H")  
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)` == "V")  
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` & InstanceNumber <= `Corte FINAL vesícula`), 1, 0)) -> base\_frames\_vesicula\_normal  
base\_frames\_vesicula\_normal %>% filter(posicao\_interesse == 1) -> base\_frames\_vesicula\_normal  
base\_frames\_vesicula\_normal

## # A tibble: 4,110 × 10  
## 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 0000483240\_4… 483240 4 40 40 56 V 1 /home/…  
## 2 0000483240\_4… 483240 4 41 40 56 V 1 /home/…  
## 3 0000483240\_4… 483240 4 42 40 56 V 1 /home/…  
## 4 0000483240\_4… 483240 4 43 40 56 V 1 /home/…  
## 5 0000483240\_4… 483240 4 44 40 56 V 1 /home/…  
## 6 0000483240\_4… 483240 4 45 40 56 V 1 /home/…  
## 7 0000483240\_4… 483240 4 46 40 56 V 1 /home/…  
## 8 0000483240\_4… 483240 4 47 40 56 V 1 /home/…  
## 9 0000483240\_4… 483240 4 48 40 56 V 1 /home/…  
## 10 0000483240\_4… 483240 4 49 40 56 V 1 /home/…  
## # … with 4,100 more rows, 1 more variable: posicao\_interesse <dbl>, 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