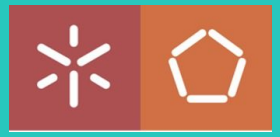
PhageScan

An automated viral particle classifier



Rui Sousa | PG21019
Mestrado em Bioinformática
Escola de Engenharia - Departamento de Engenharia Biológica

Introduction to Bacteriophages

Universidade do Minho
Escola de Engenharia

Overview

- Viruses that thrive by infecting and replicating within bacterial cells
- Species-specific
- Maintain life cycle by co-evolving with target-species[1]
- Three different life styles:
    - Lysogenic
    - Lytic
    - Chronic
- Valuable biological assets with therapeutical interest[2]

1 Koskella, 2014 (https://doi.org/10.1111/1574-6976.12072)

2 Sulakvelidze, 2001 (https://doi.org/10.1128/aac.45.3.649-659.2001)

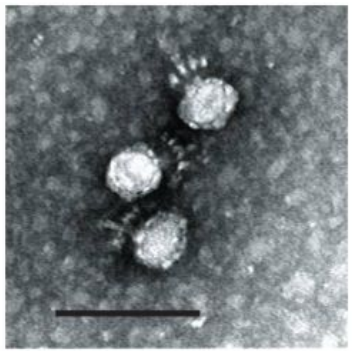
Universidade do Minho
Escola de Engenharia

Phage morphology

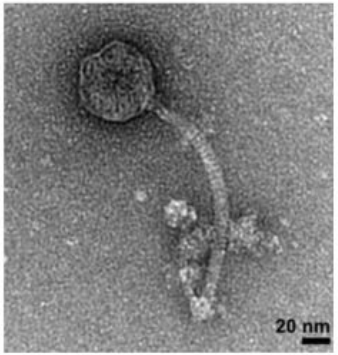
- Tailed phages are composed by:
  - Icosahedral capsid (genetic material storage)
  - Tail (to attach to and infect host)
- Tailed phages exhibit 3 different morphotypes:
  - Podoviruses: short non-contractile tail
  - Siphovirus: long, flexible, non-contractile tail
  - Myoviruses: long contractile tail with a base plate and spikes

Universidade do Minho
Escola de Engenharia

Phage morphology



(a) Podovirus



(b) Siphovirus

(c) Myovirus

Morphotypes of tailed phages [1, 2, 3]

1 Volozhantsev, 2012 - https://doi.org/10.1016/j.ijfoodmicro.2021.109446
2 Demo, 2021 - https://doi.org/10.1371/journal.pone.0248418
3 Mann, 2005 - https://doi.org/10.1371/journal.pbio.0030182

Automated classification of viral particles

Universidade do Minho
Escola de Engenharia

Overview

- Morphology retains visual classification potential!
- Manual identification is laborious and fastidious
- Automated viral classification is growing in popularity
- Several research efforts to develop novel tools

Definition of a "valuable" automated classification tool:

- >75% accuracy[1]

1 Glaeser, 2004 (https://doi.org/10.1016/j.jsb.2003.09.005)

Universidade do Minho
Escola de Engenharia

Application of machine learning techniques

- Gaussian classifiers
- Multi-layer perceptrons
- Naive bayes
- Random forests
- Convolutional Neural Networks

Universidade do Minho
Escola de Engenharia

Goal of this project

Develop a tool to infer virus' taxonomy from TEM micrographs

- Streamline taxonomical classification
- Rough estimation of phage genome size based on TEM micrographs

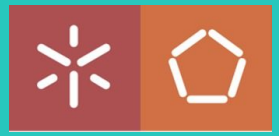
Apply Convolutional Neural Networks (CNNs) to classify morphotypes

- Specifically YOLO (You Only Look Once) algorithms

Why CNNs?

- Versatile models specialized in processing grid-like data (i.e. images)
- Fast, flexible and easy to parametrize
- Strong capability to discern patterns → popular object classifier

Teuwen and Moriakov, 2020 (https://doi.org/10.1016/B978-0-12-816176-0.00025-9)



Introduction to Convolutional Neural Networks

Universidade do Minho
Escola de Engenharia

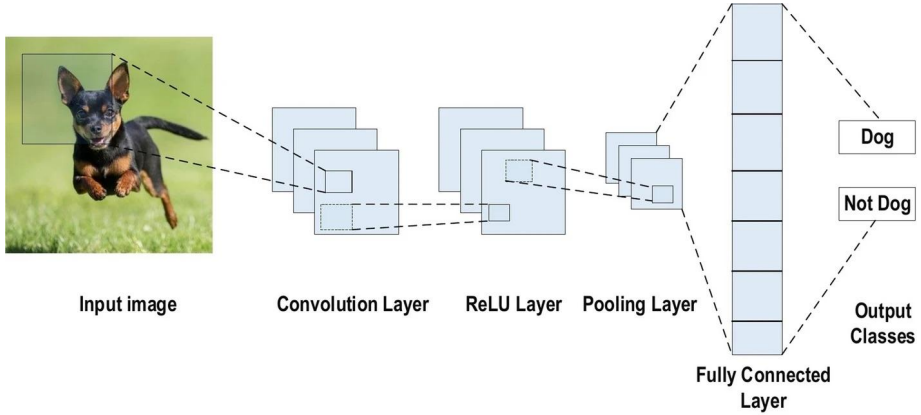
Introduction to CNNs

Constituted by:

- Input layer (receives raw input data)
- Hidden layers:
  - Convolutional layers (mathematical processing)
  - Pooling layers (downsampling operations)
  - Fully connected layers (vectorial computation)
- Output layer (produces final output)

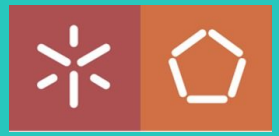
Universidade do Minho
Escola de Engenharia

Phage morphology



Schematic representation of a CNN

Source: Alzubaidi et al, 2021 (https://doi.org/10.1186/s40537-021-00444-8)



Methods

Universidade do Minho
Escola de Engenharia

Data collection

- Data collected from PhagesDB and CEB Phage Research Lab
  - 300+ phage TEM micrographs

Noteworthy:

- Vast majority of micrographs are of siphoviruses
- Myoviruses in “relaxed state” and siphoviruses are similar
- Noisy micrographs

Universidade do Minho
Escola de Engenharia

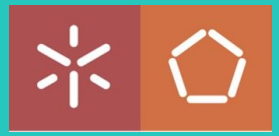
Data preprocessing

- Resizing images to fixed size (640 x 640 px)
- Annotation using bounding boxes and polygons
  - 1 class: “siphovirus”
- Data set augmentation:
  - Random rotations
  - Flips

Universidade do Minho
Escola de Engenharia

Model training parameters

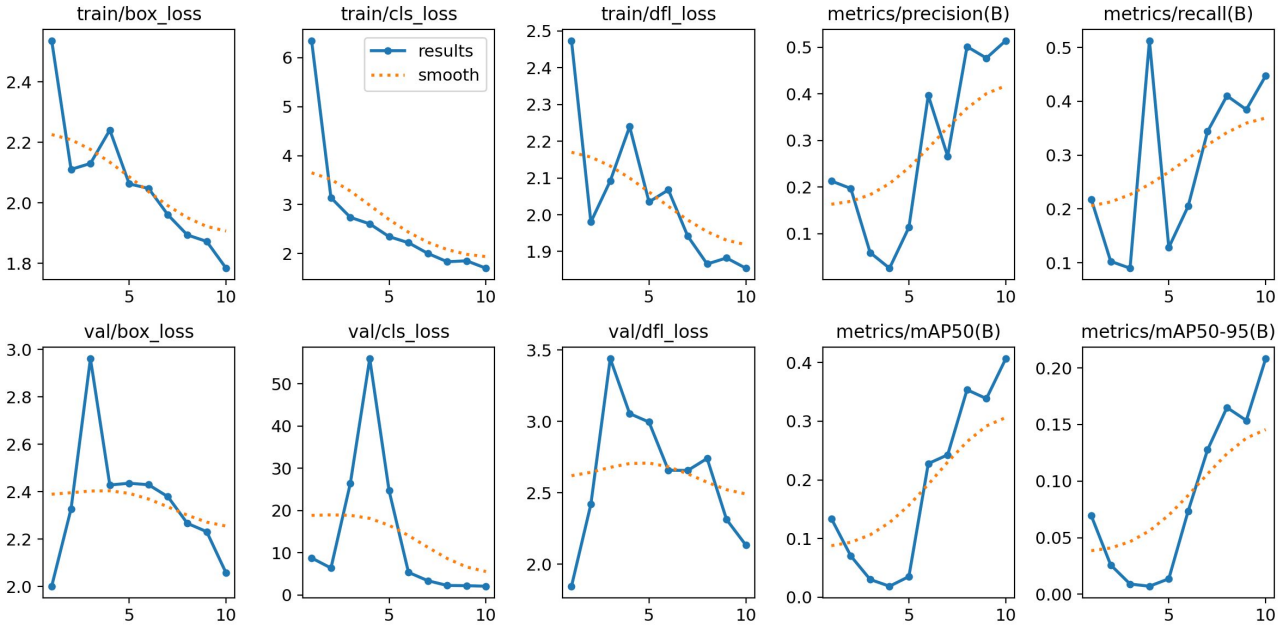
- Epochs: 10; 25; 50
- Batch size: 16
- Activation function: SiLU (YOLO default)
- Learning rate: 10e-3 (Adam - YOLO default)
- Data set splitting - 80:15:5 (Train:Validate:Test)



Results

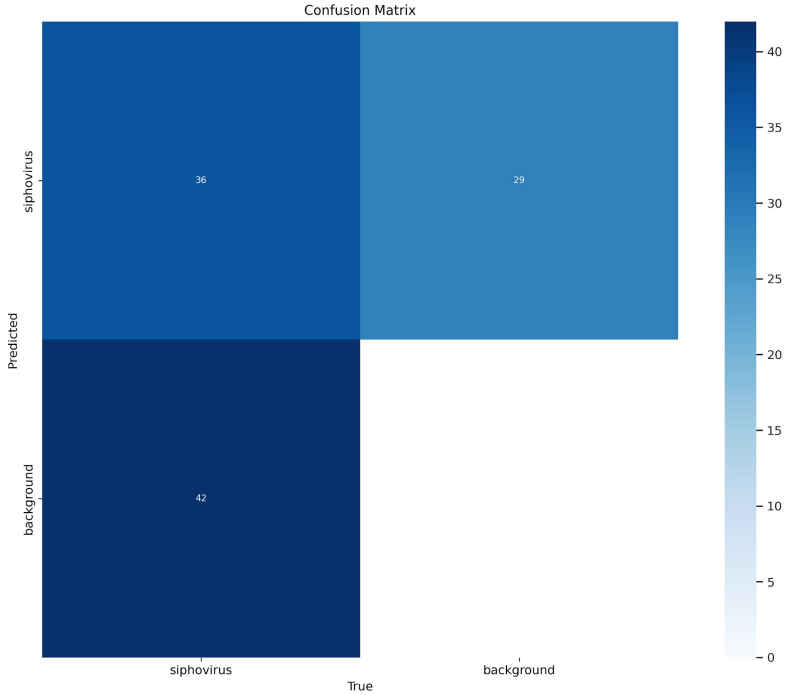
Universidade do Minho
Escola de Engenharia

Training results: 10 epochs



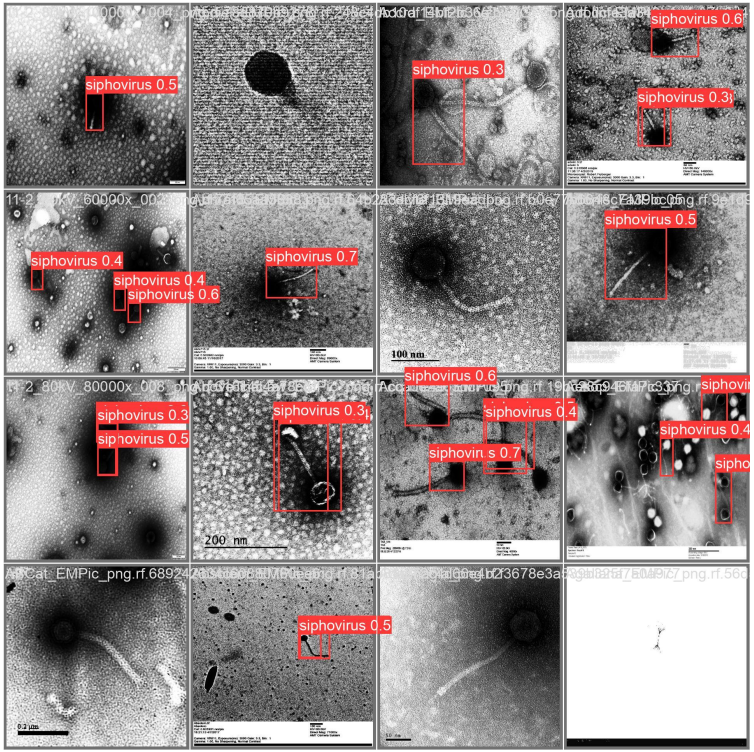
Universidade do Minho
Escola de Engenharia

Training results: 10 epochs



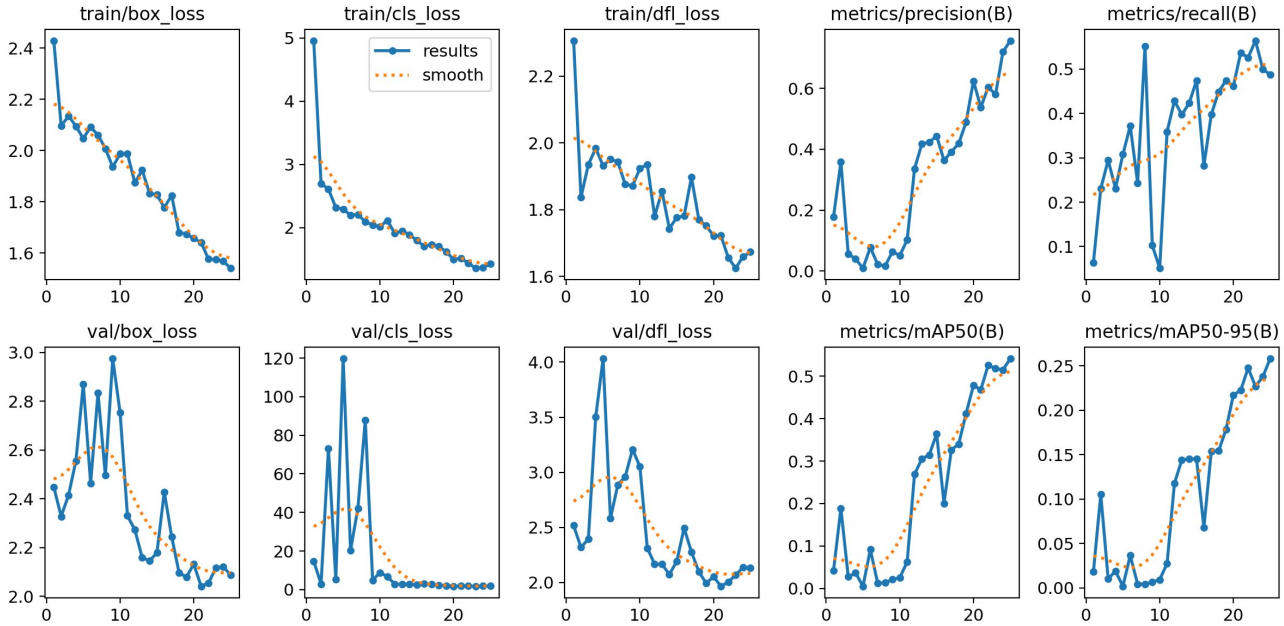
Universidade do Minho
Escola de Engenharia

Training results: 10 epochs



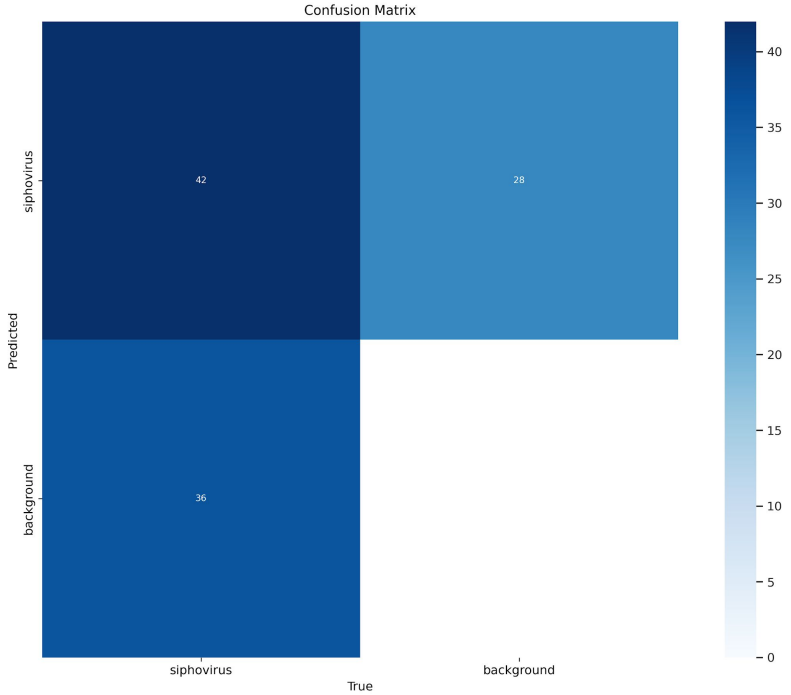
Universidade do Minho
Escola de Engenharia

Training results: 25 epochs



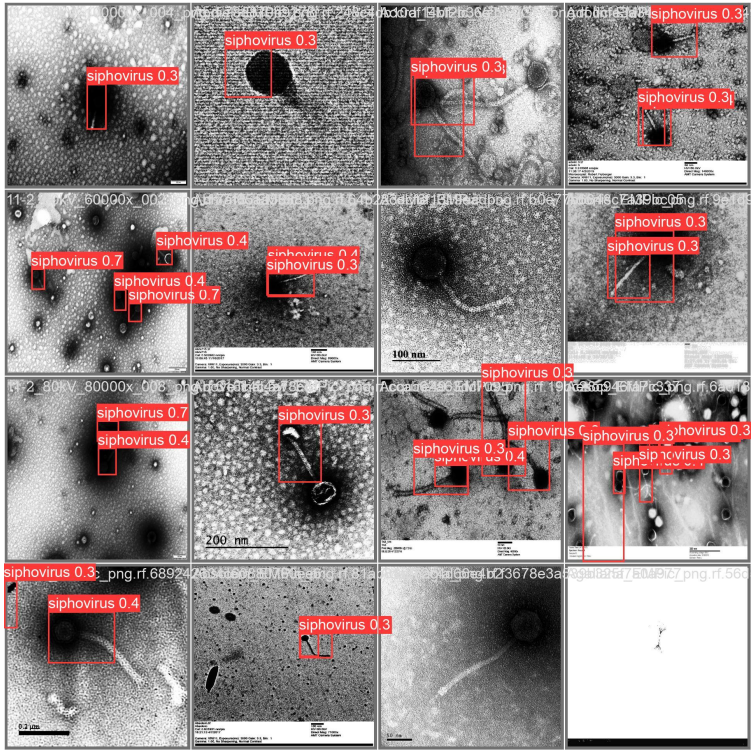
Universidade do Minho
Escola de Engenharia

Training results: 25 epochs



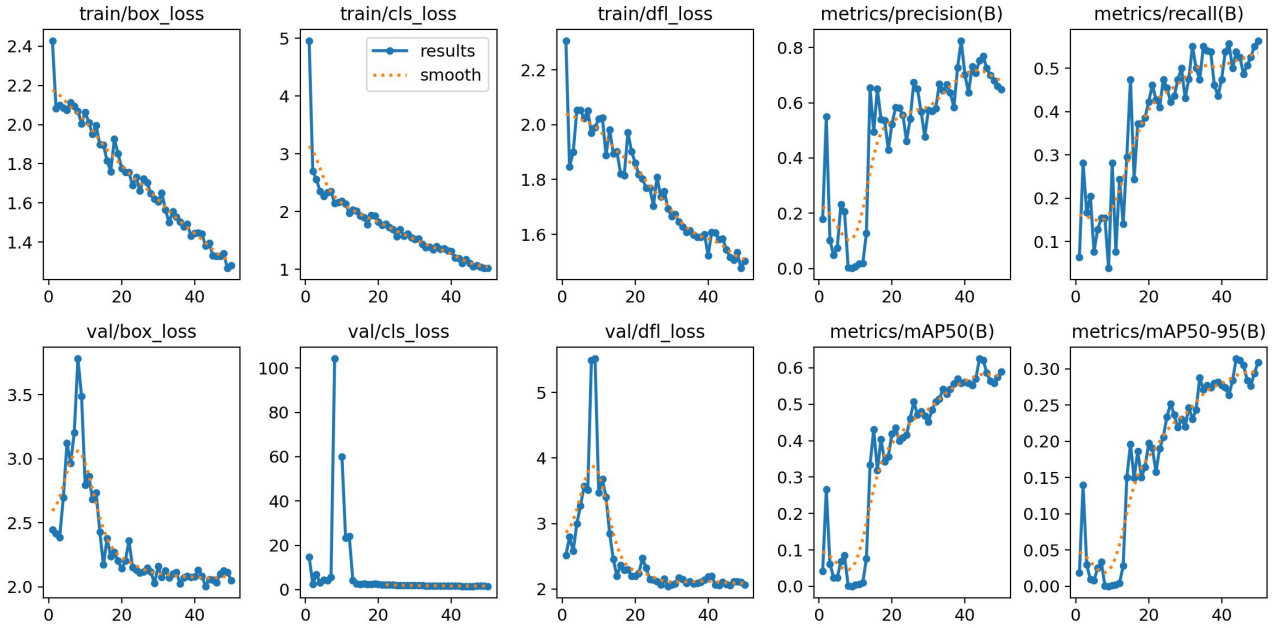
Universidade do Minho
Escola de Engenharia

Training results: 25 epochs



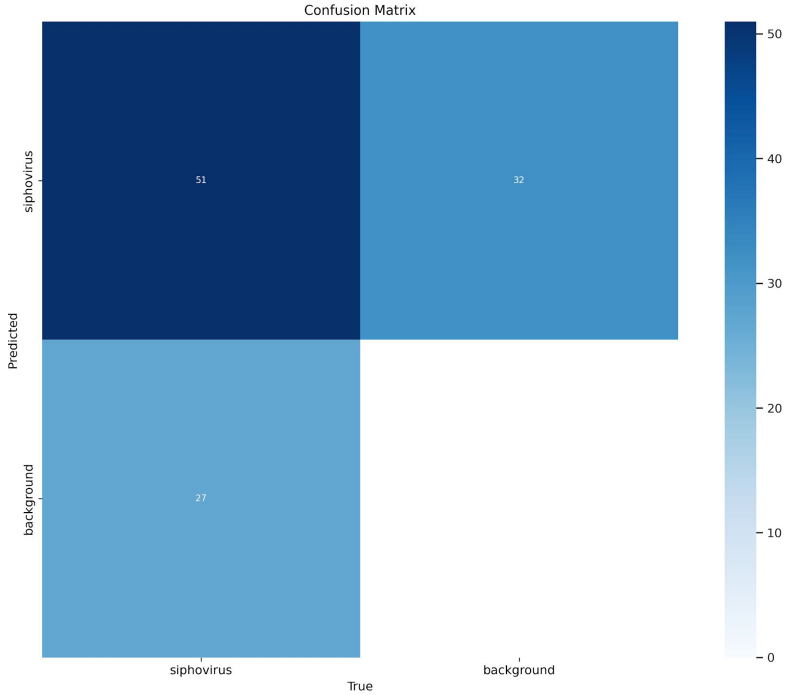
Universidade do Minho
Escola de Engenharia

Training results: 50 epochs



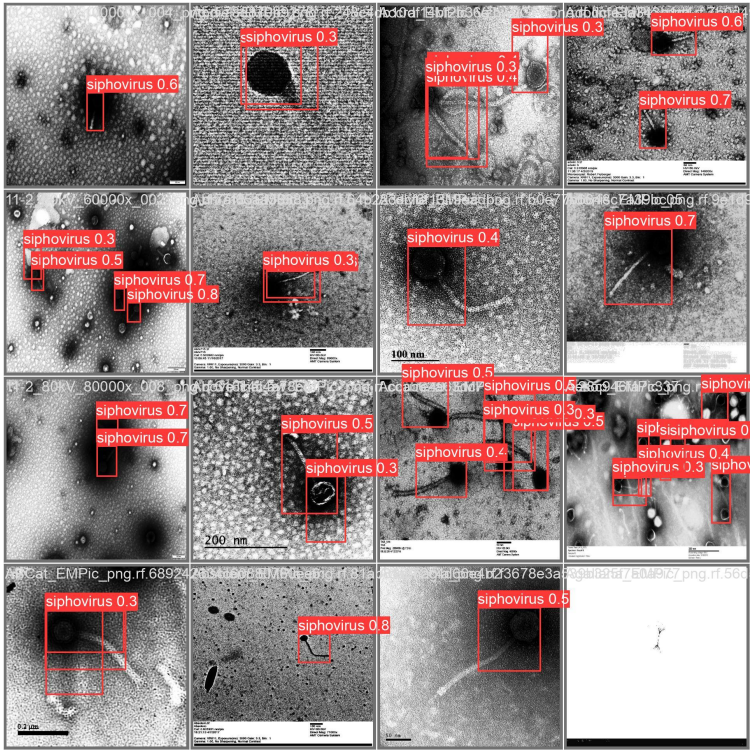
Universidade do Minho
Escola de Engenharia

Training results: 50 epochs



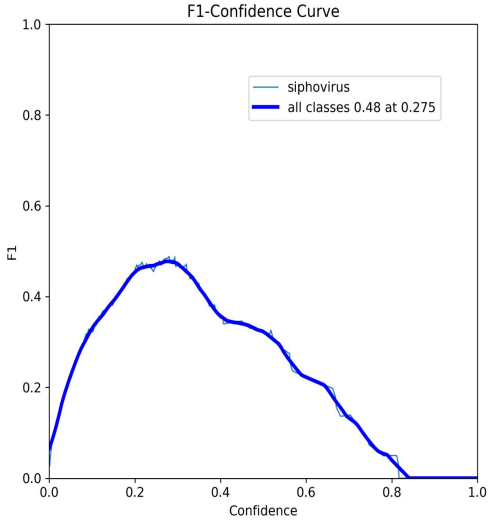
Universidade do Minho
Escola de Engenharia

Training results: 50 epochs

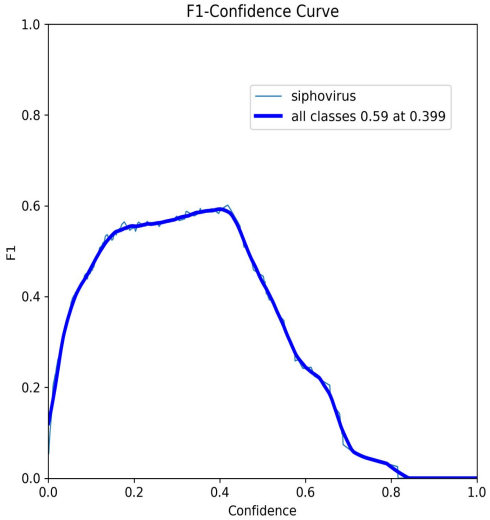


Universidade do Minho
Escola de Engenharia

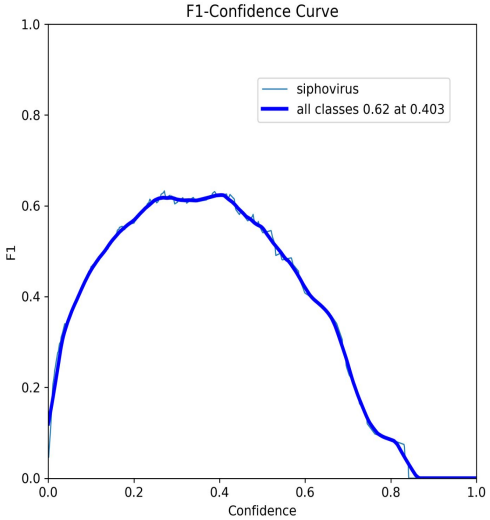
Comparison



10 Epochs



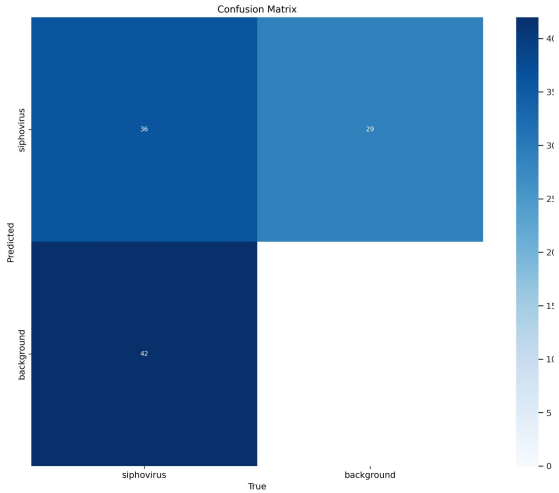
25 Epochs



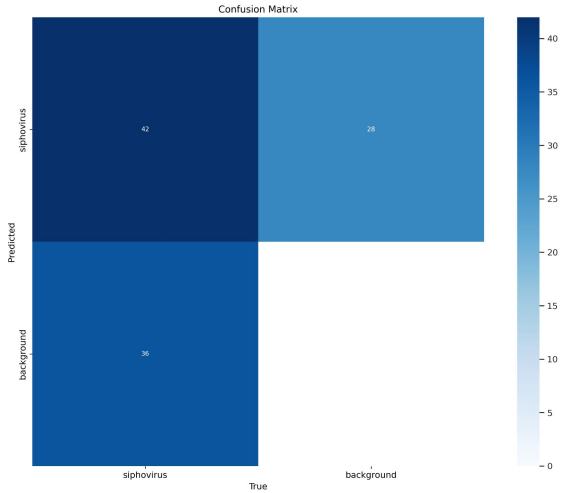
50 Epochs

Universidade do Minho
Escola de Engenharia

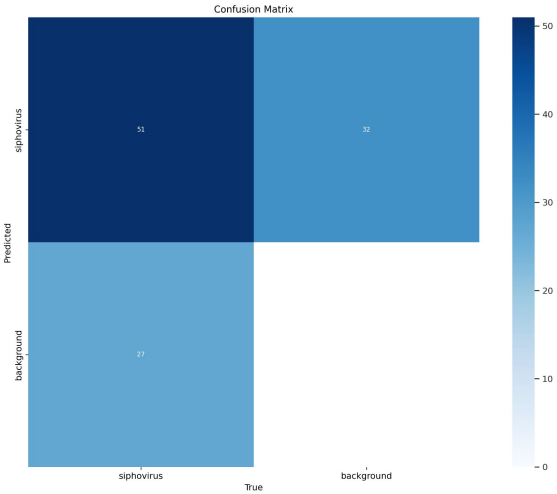
Comparison



10 Epochs



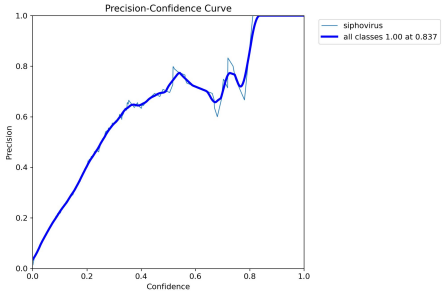
25 Epochs

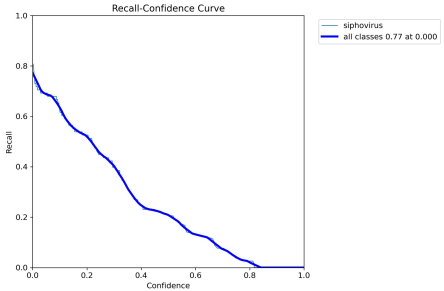


50 Epochs

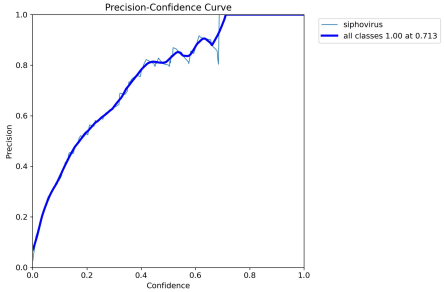
Universidade do Minho
Escola de Engenharia

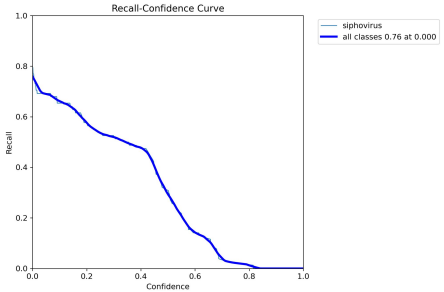
Comparison



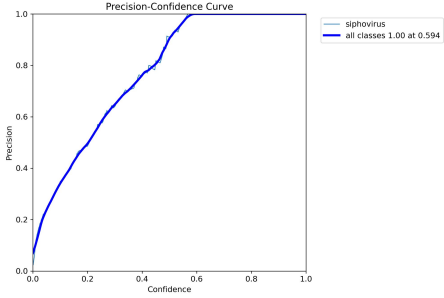


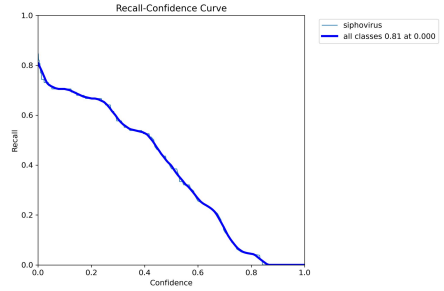
10 Epochs



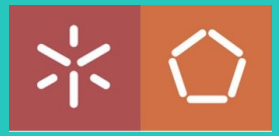


25 Epochs





50 Epochs

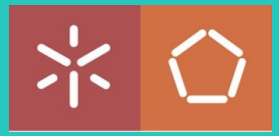


Conclusions

Universidade do Minho
Escola de Engenharia

Conclusions

- Reasonable F1 score for 0.6 confidence threshold;
- Model improves through epochs, being the optimized around 40 epochs;
- Overfitting is visible (a problem when using single classes), especially for high confidence thresholds;

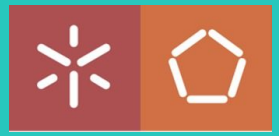


Future work

Universidade do Minho
Escola de Engenharia

Future work

- Train the model with more classes;
- Increase data variety and quality to improve results;
- Add automated measurement features (exploratory):
  - implement OCR for scale reading;
  - edge detection to isolate capsid;
  - estimate measurements;



Thank you!