



Universidade do Minho

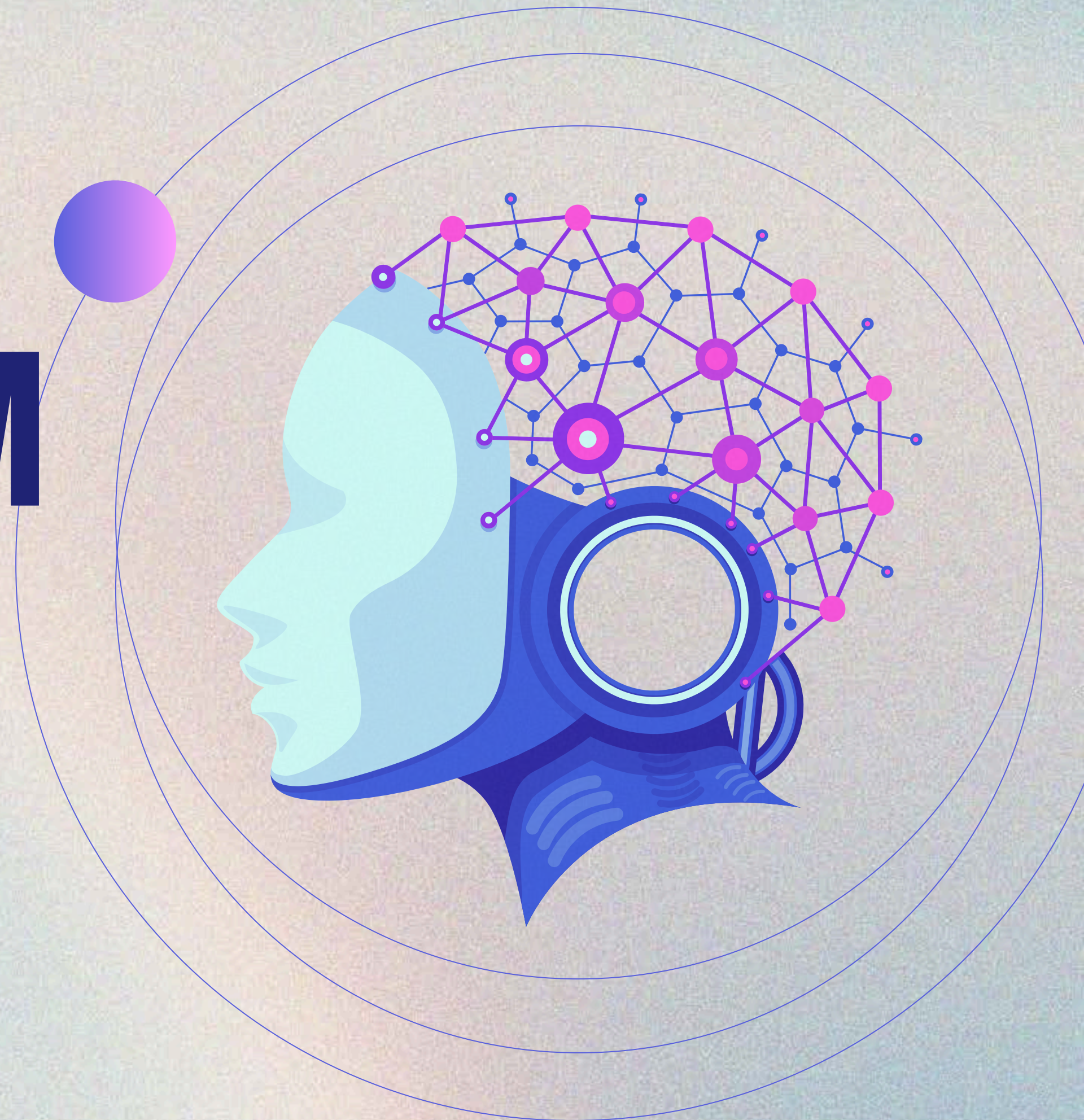
APRENDIZAGEM PROFUNDA

David Teixeira PG55929

Eduardo Cunha PG55939

Nuno Rodrigues PG55966

Tiago Rodrigues PG56013



INTRODUÇÃO

A inteligência artificial (IA) tem transformado a educação baseada em competências, permitindo feedback personalizado e avaliação de desempenho mais objetiva. Em áreas como a sutura cirúrgica, onde são exigidas competências motoras complexas, para fornecer feedback imediato e personalizado, promovendo a motivação e a autorreflexão dos alunos.

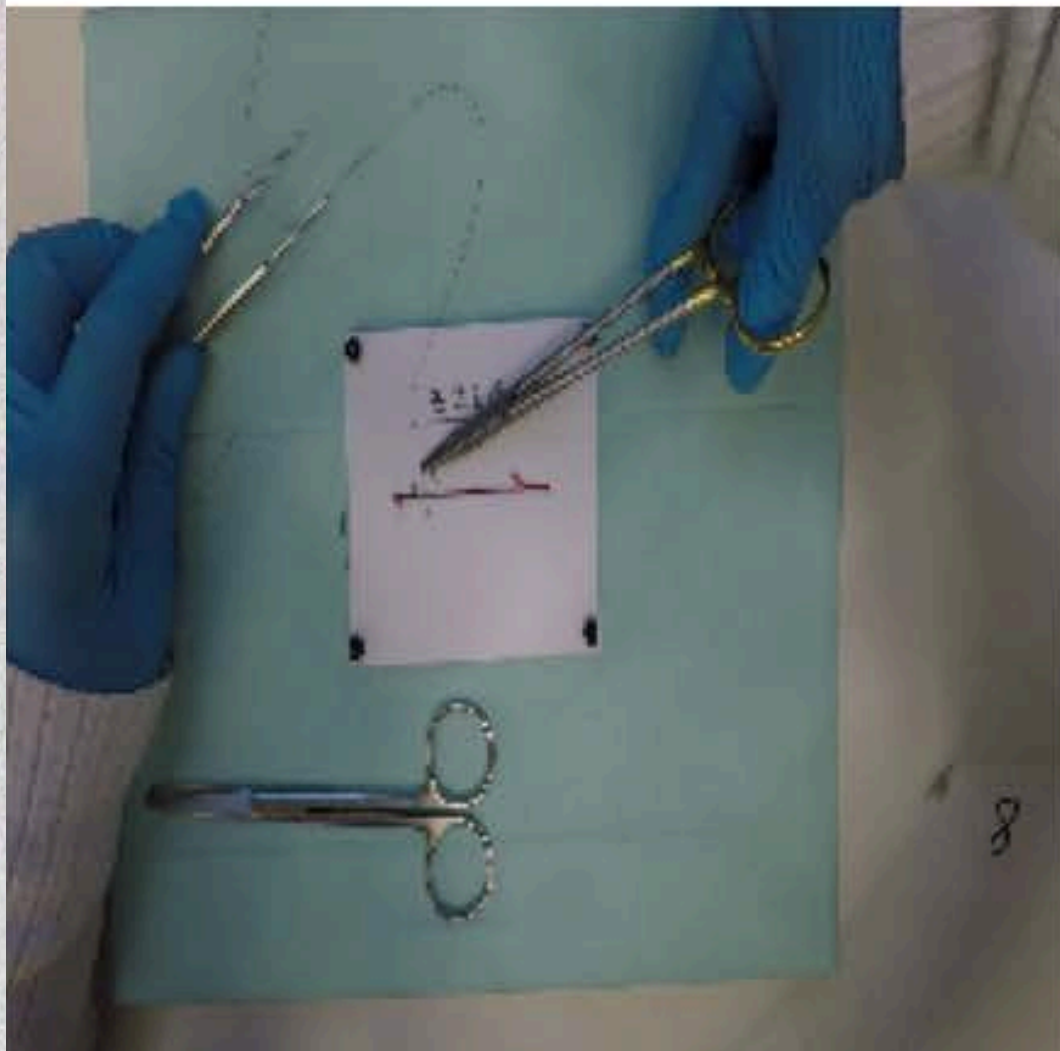
Este trabalho propõe um modelo de avaliação de competências em sutura com base em deep learning, capaz de analisar vídeos e classificar o desempenho dos alunos com elevada precisão, recorrendo a modelos visuais de última geração.

METODOLOGIA

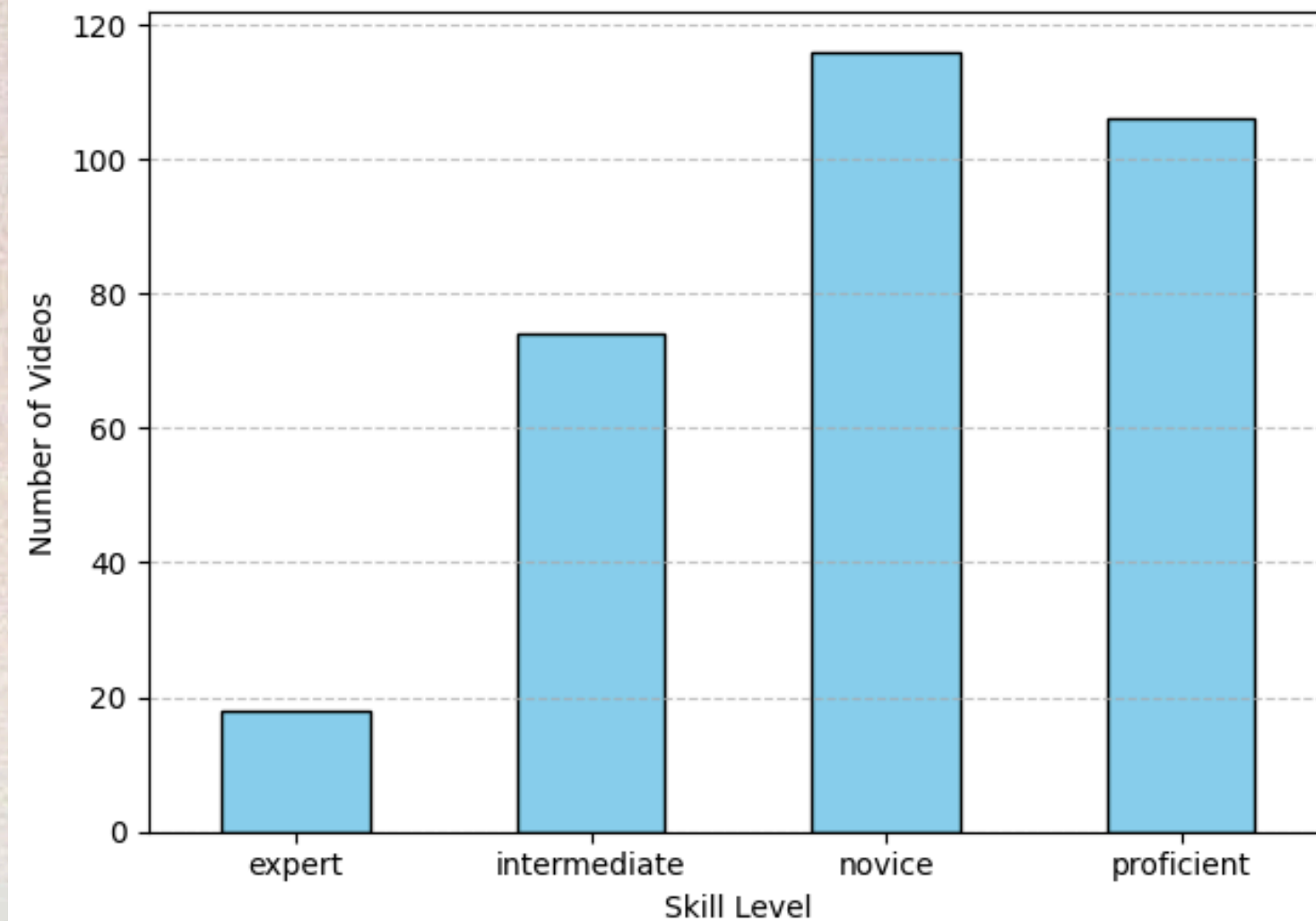


EXPLORAÇÃO

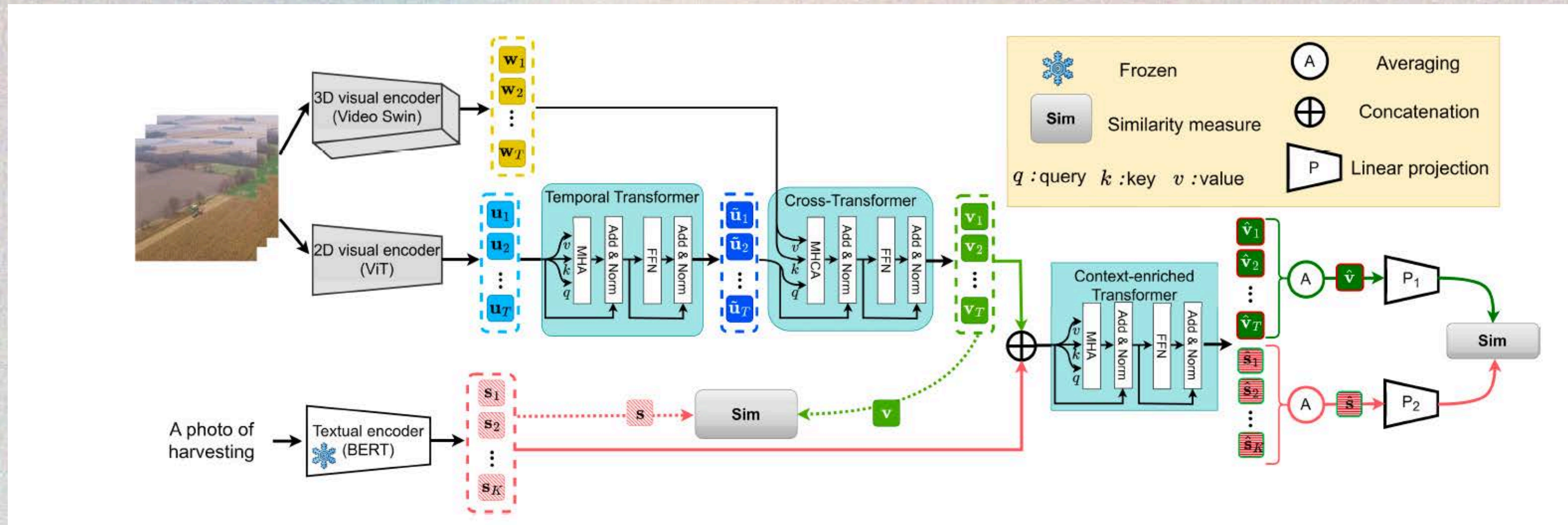
First Frame



Skill Level Distribution

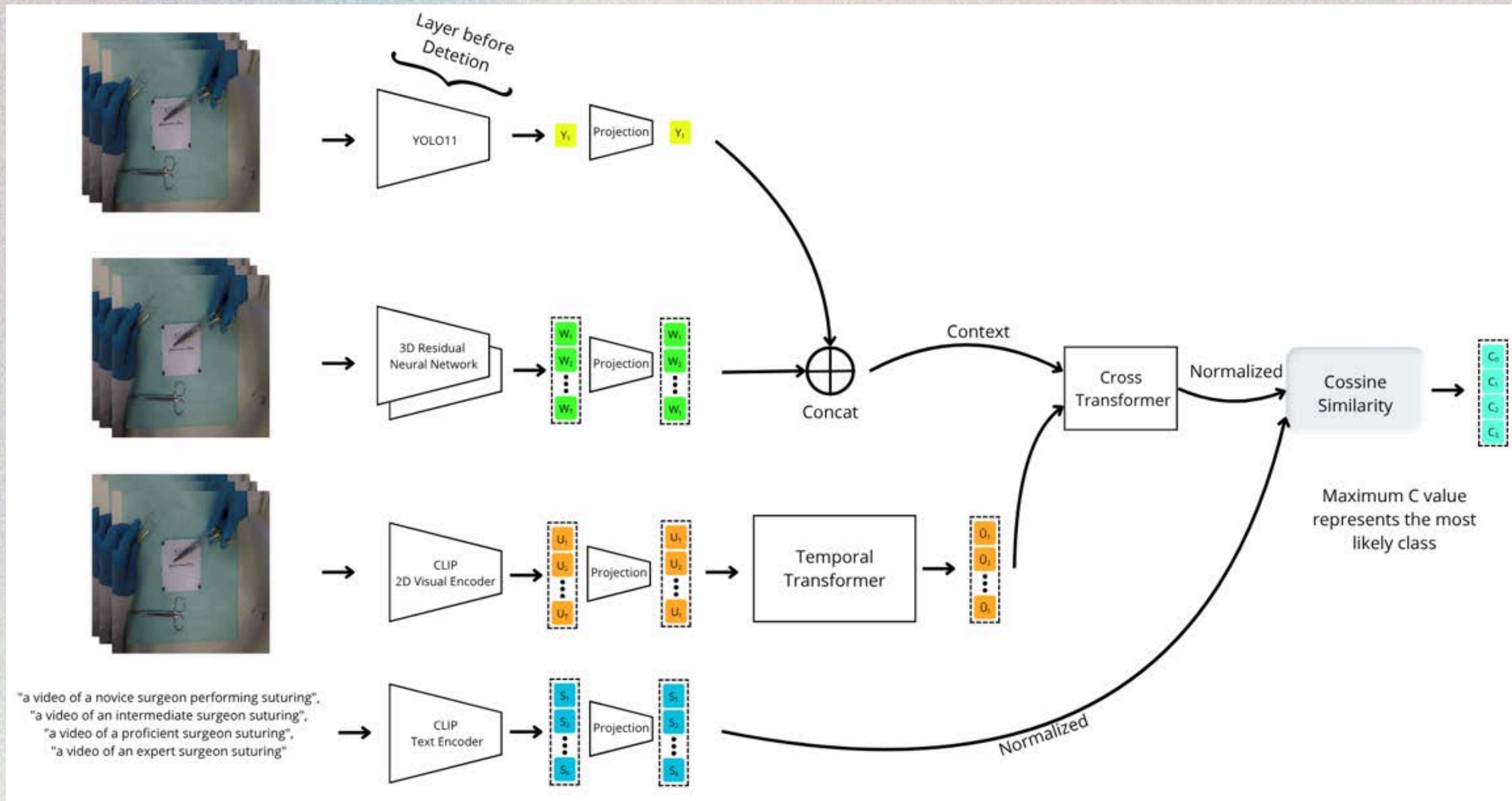


DESENVOLVIMENTO DO MODELO

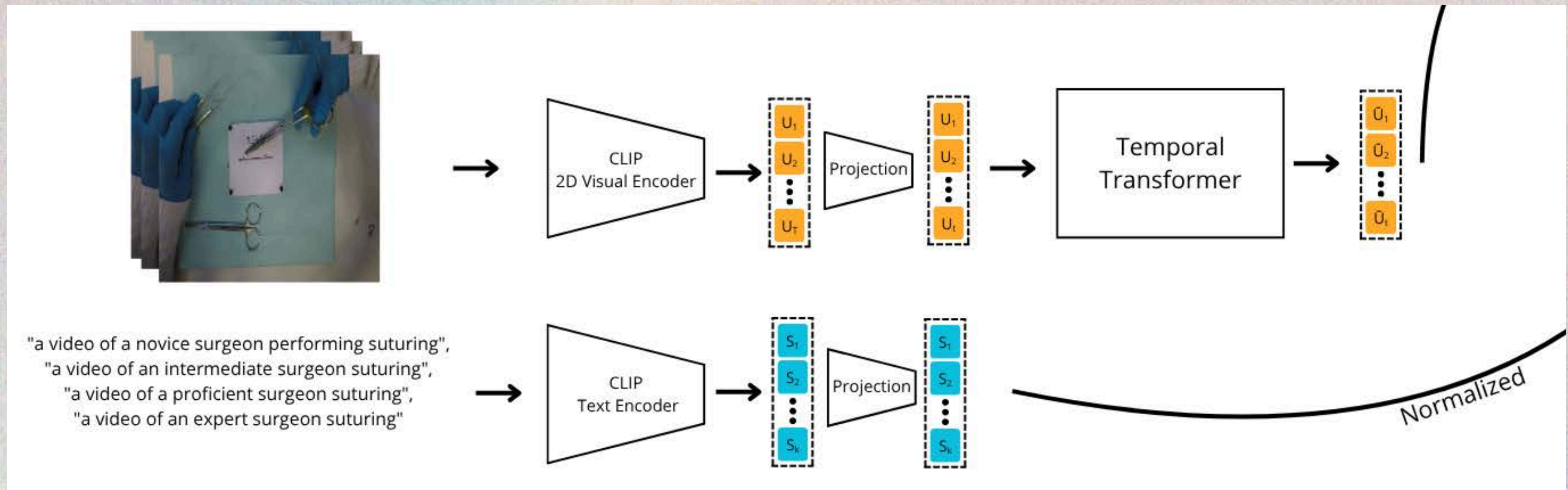


N. A. Tu and N. Aikyn, "Improving Vision-Language Models With Attention Mechanisms for Aerial Video Classification," in IEEE Geoscience and Remote Sensing Letters, vol. 22, pp. 1-5, 2025, Art no. 8000505, doi: 10.1109/LGRS.2025.3532987.

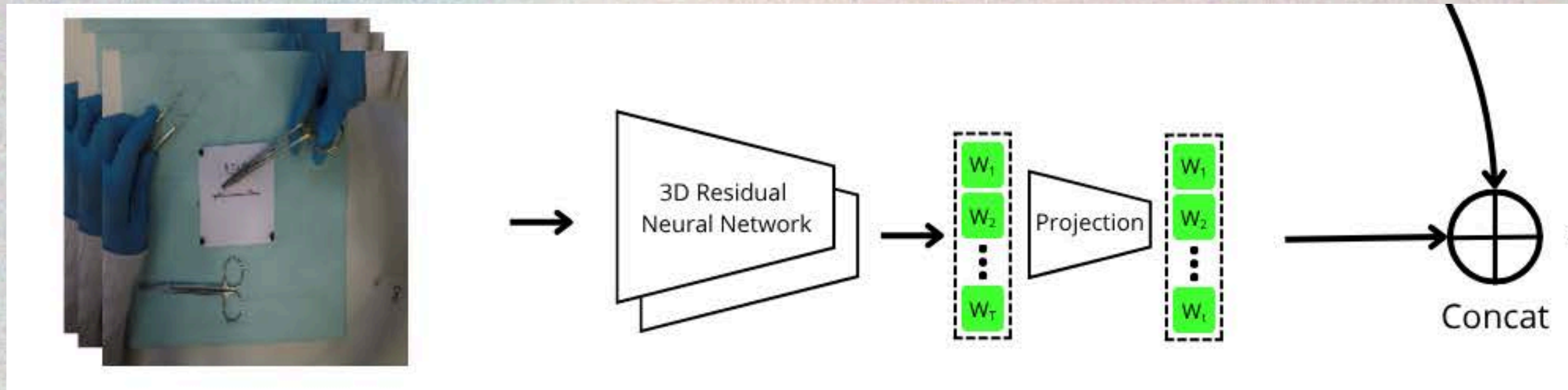
TASK 1 E 2 - ARQUITETURA



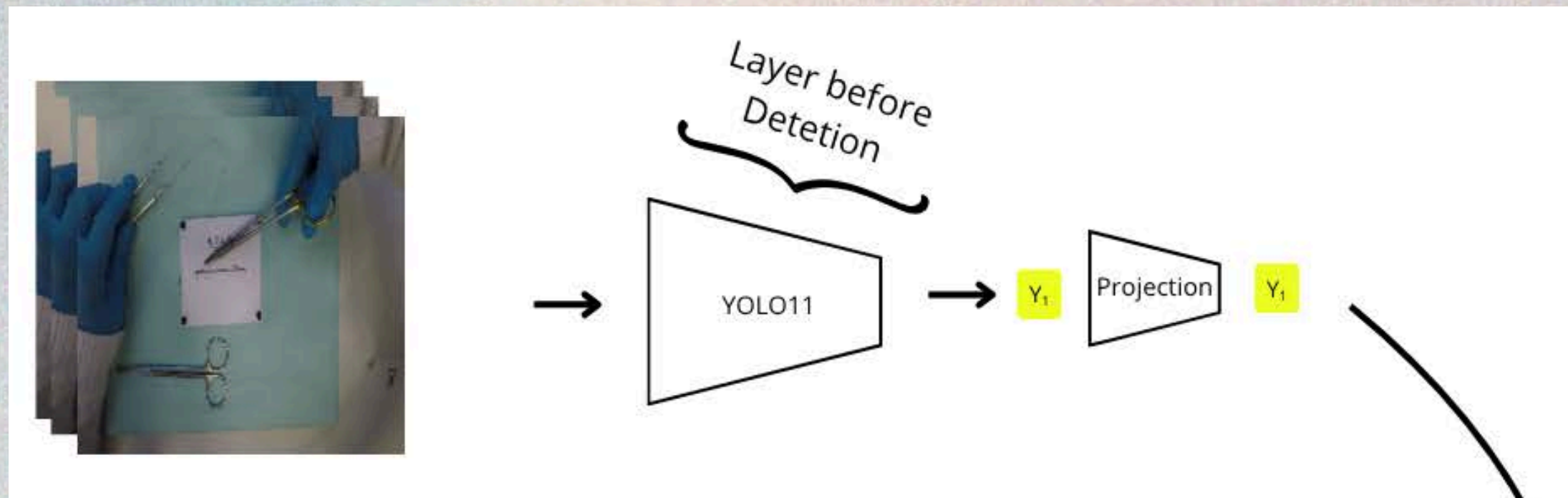
TASK 1 E 2 - ARQUITETURA



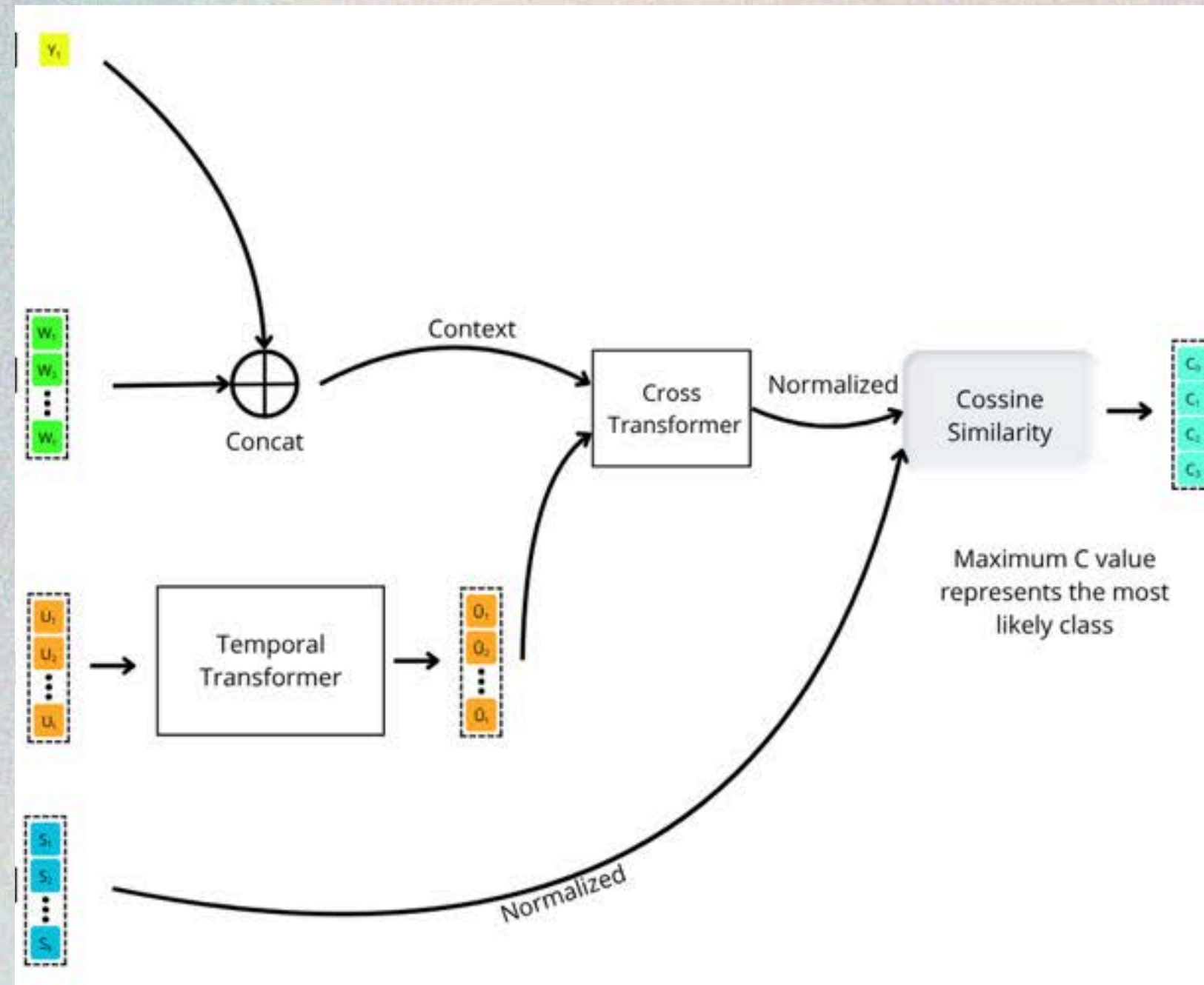
TASK 1 E 2 - ARQUITETURA



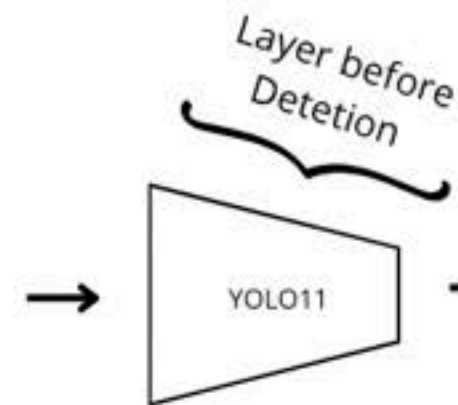
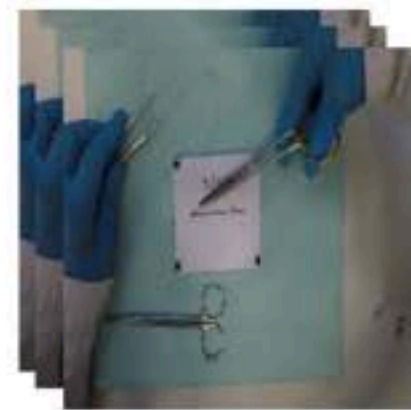
TASK 1 E 2 - ARQUITETURA



TASK 1 E 2 - ARQUITETURA



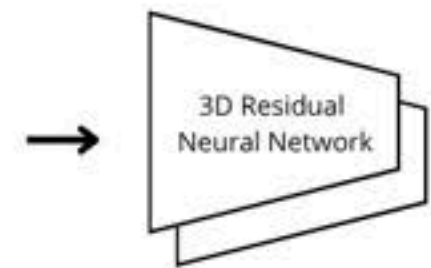
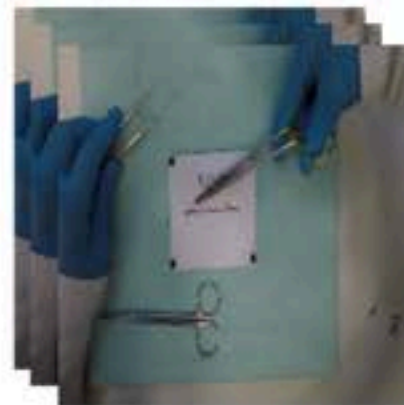
TASK 2 - ARQUITETURA



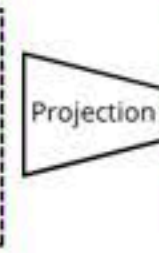
Y_1



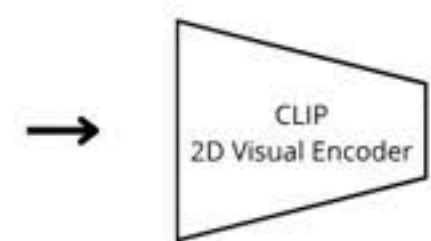
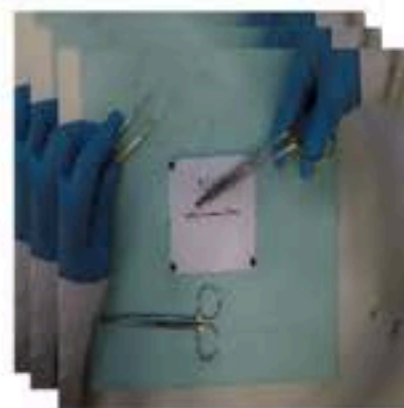
Y_1



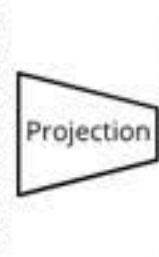
W_1
 W_2
 \vdots
 W_T



W_1
 W_2
 \vdots
 W_t



U_1
 U_2
 \vdots
 U_T



U_1
 U_2
 \vdots
 U_t



O_1
 O_2
 \vdots
 O_t

Concat

Context

Cross
Transformer

MLP
HEAD

L_1
 L_2
 L_3
 L_4
 L_5
 L_6
 L_7

TASK 3

TASK3

YOLO11 with COCO
Pre-Training

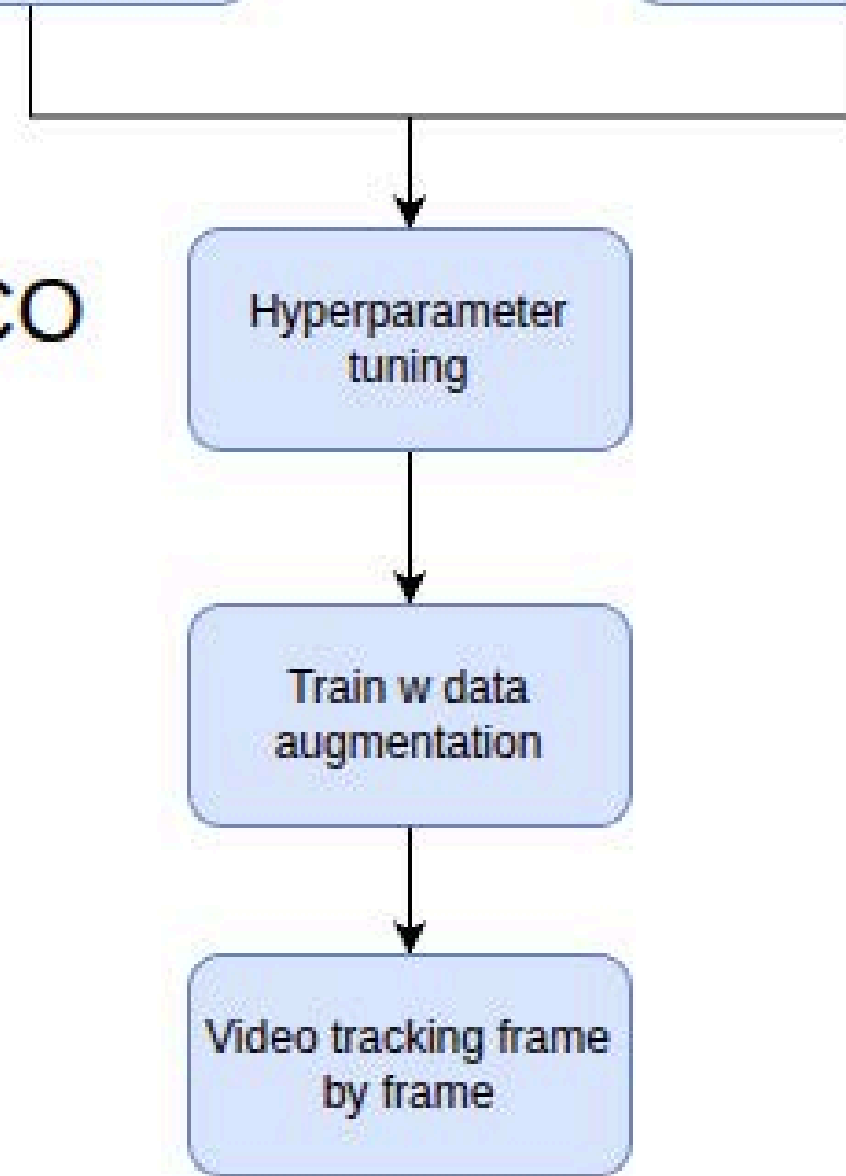
Dataset Labeled-
surgical-tools

Dataset Surgical-
hands

Hyperparameter
tuning

Train w data
augmentation

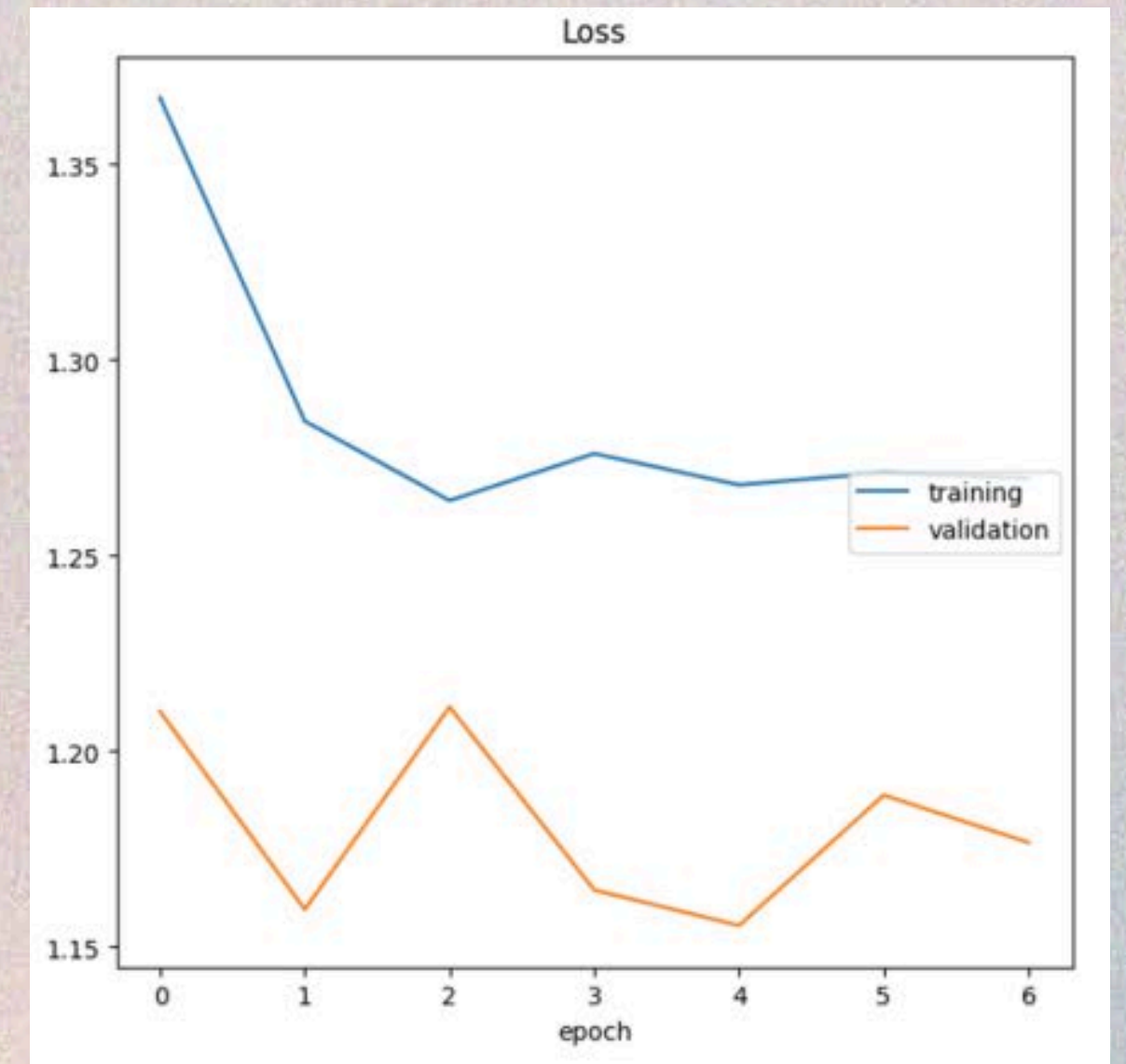
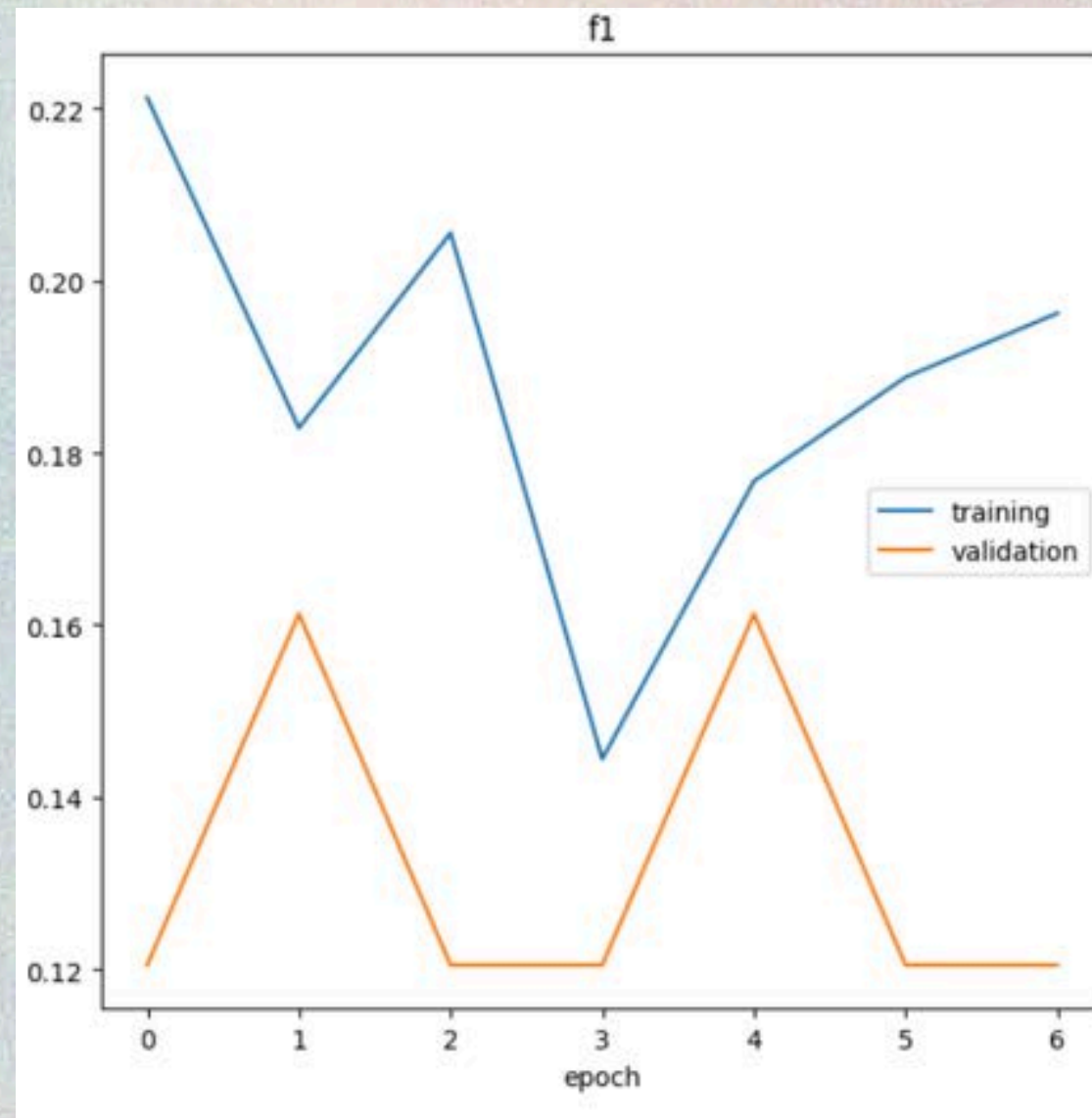
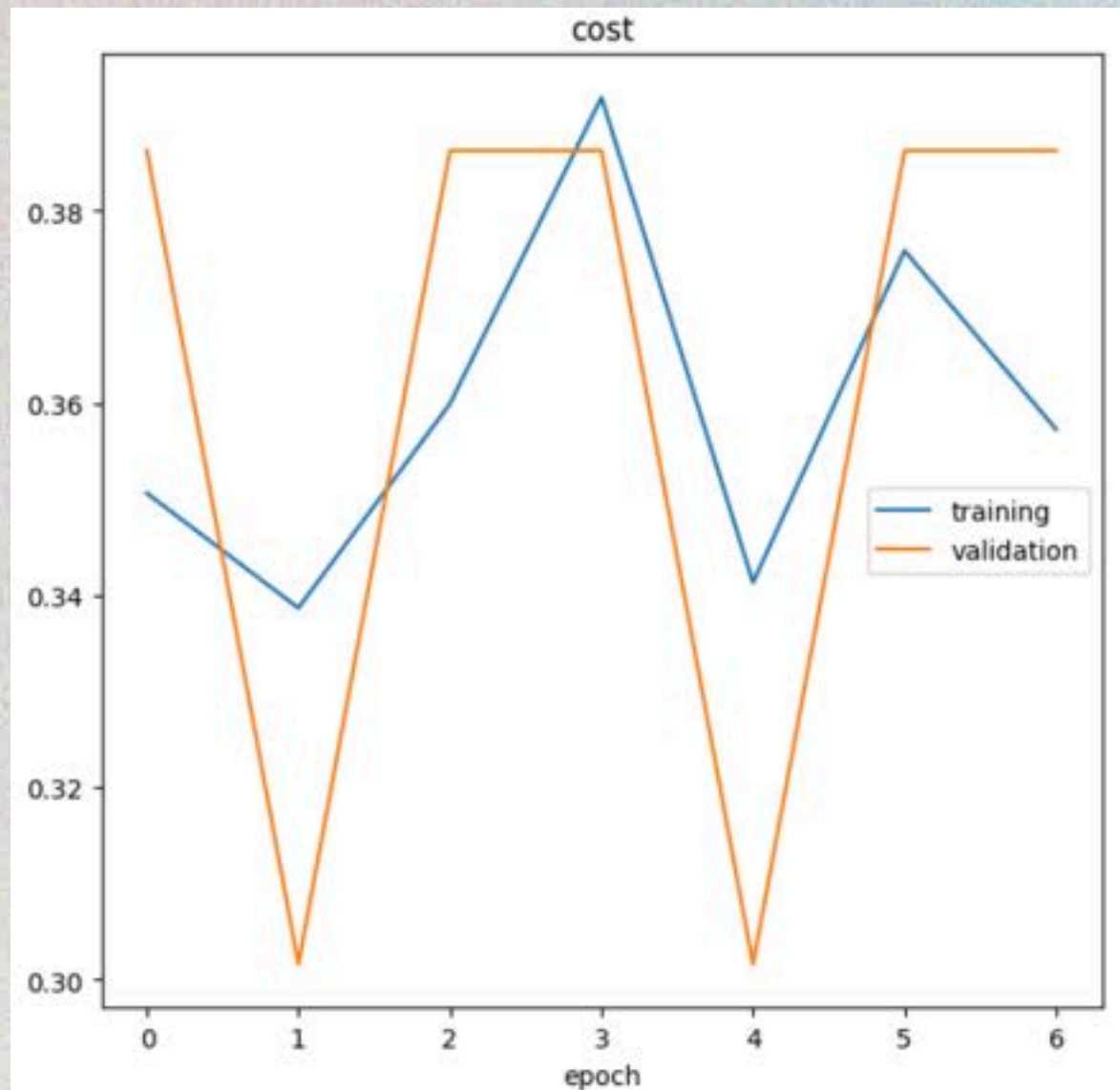
Video tracking frame
by frame



RESULTADOS - TASK 1

Designação do Modelo	F1-Score	Expected Cost	Epoch
Frozen CLIP + R3D + YOLO	0.161	0.386	6
CLIP + R3D + YOLO + MLP	0.161	0.302	2
CLIP + R3D + MLP	0.161	0.302	3
CLIP + MLP	0.161	0.302	4
CLIP + R3D + YOLO + MLP Large	0.161	0.302	1

RESULTADOS - TASK 1

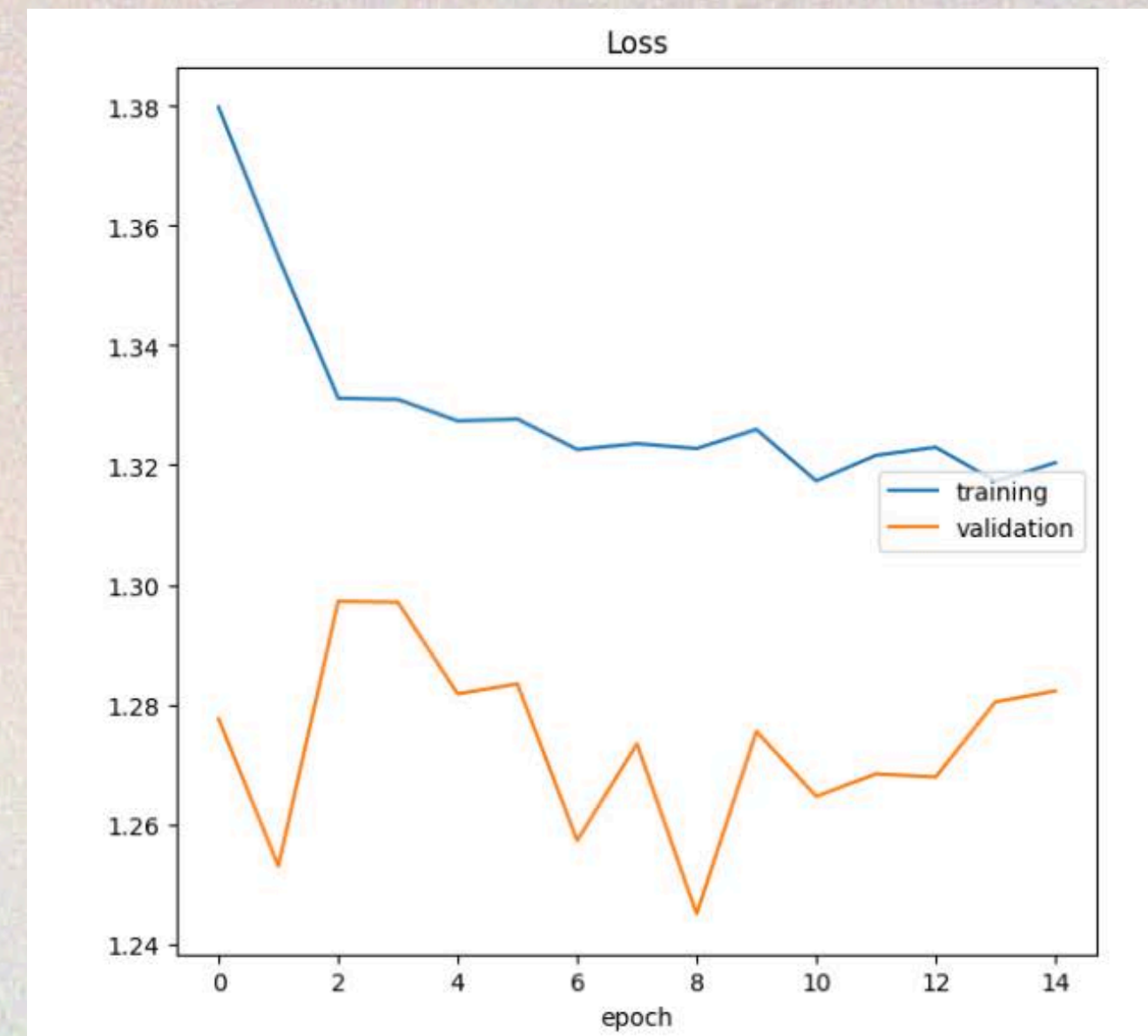
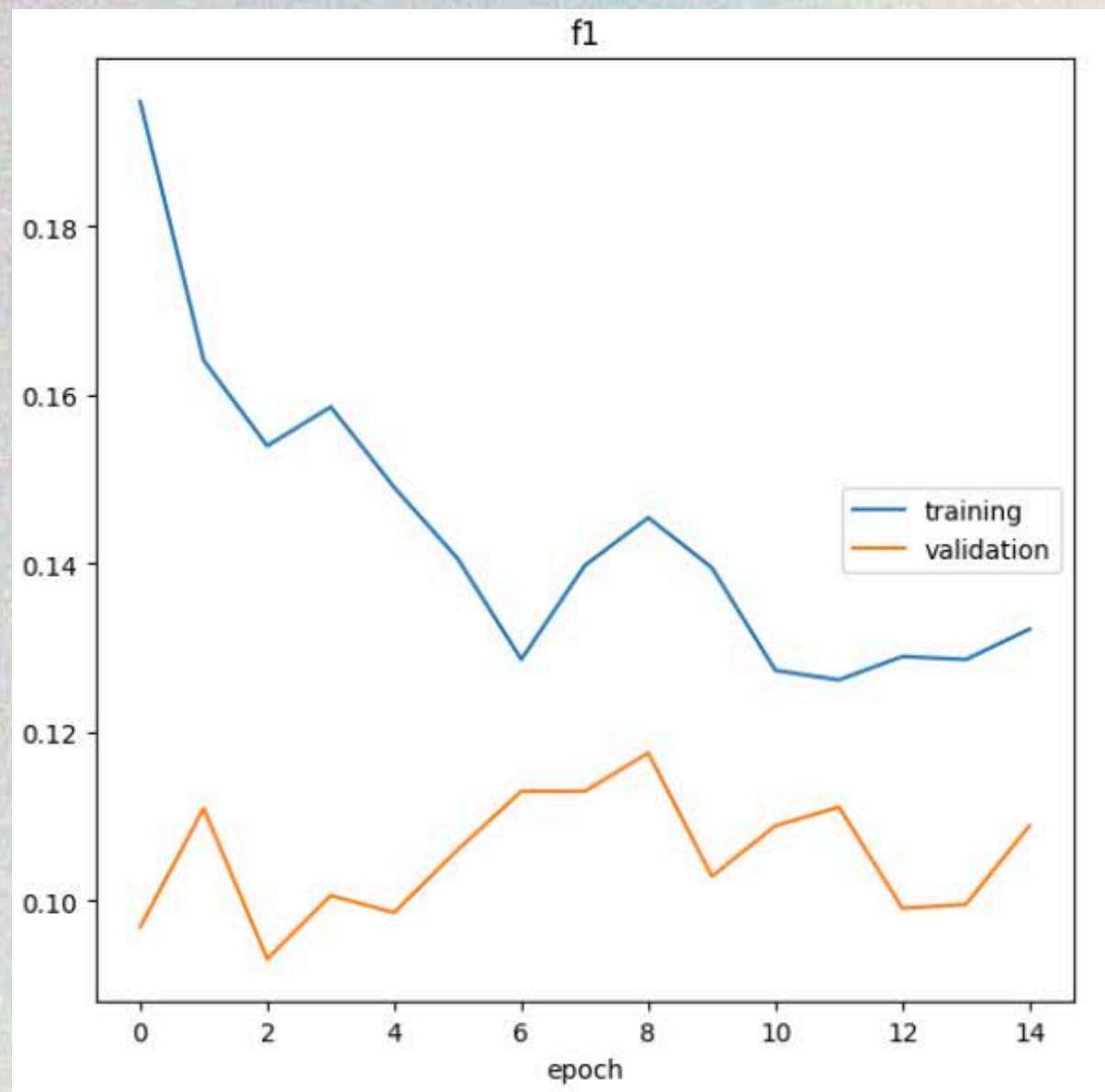


CLIP + R3D + YOLO + MLP

RESULTADOS - TASK 2

Designação do Modelo	F1-Score	Expected Cost	Epoch
Frozen CLIP + R3D + YOLO	0.118	---	9

RESULTADOS - TASK 2



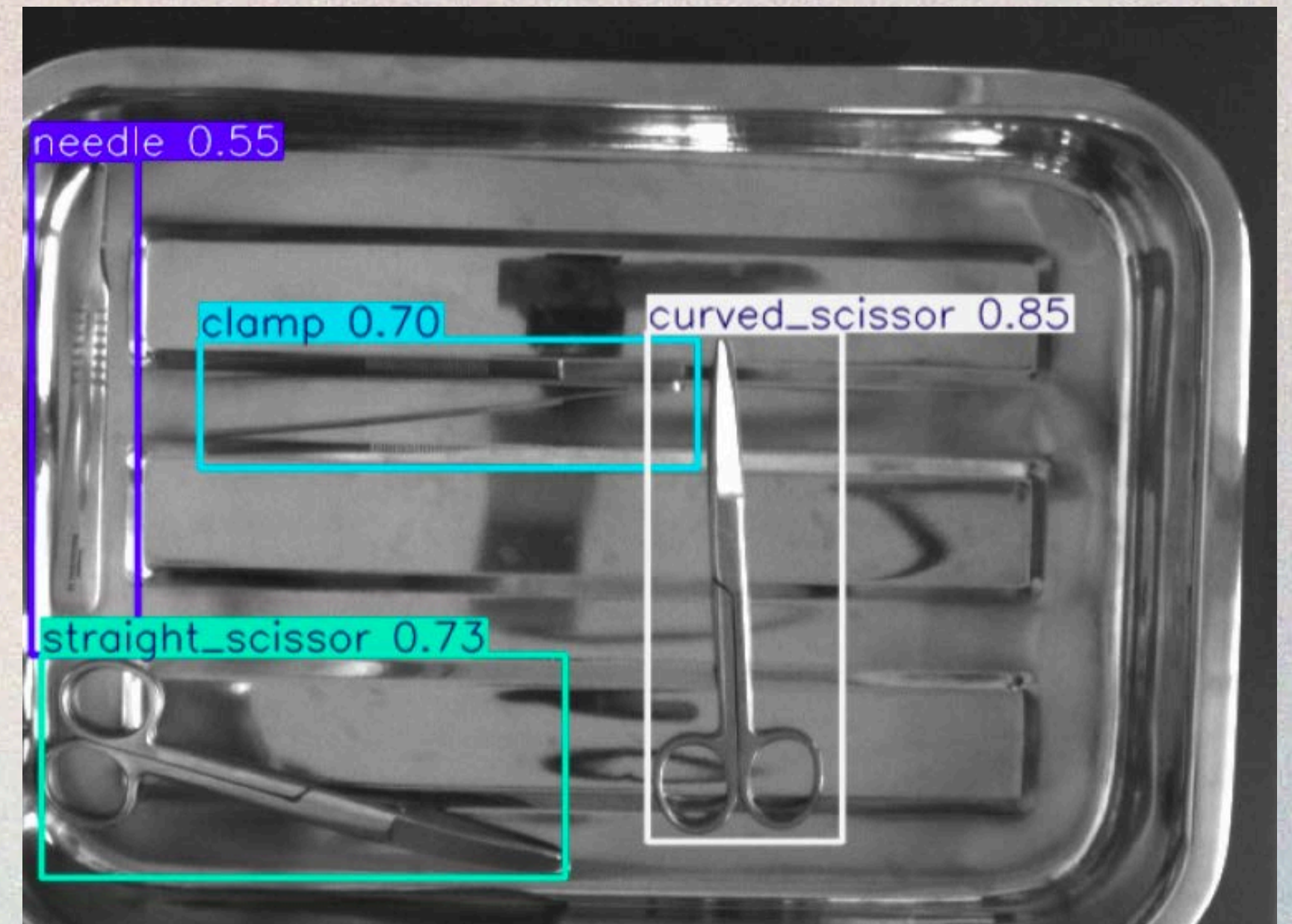
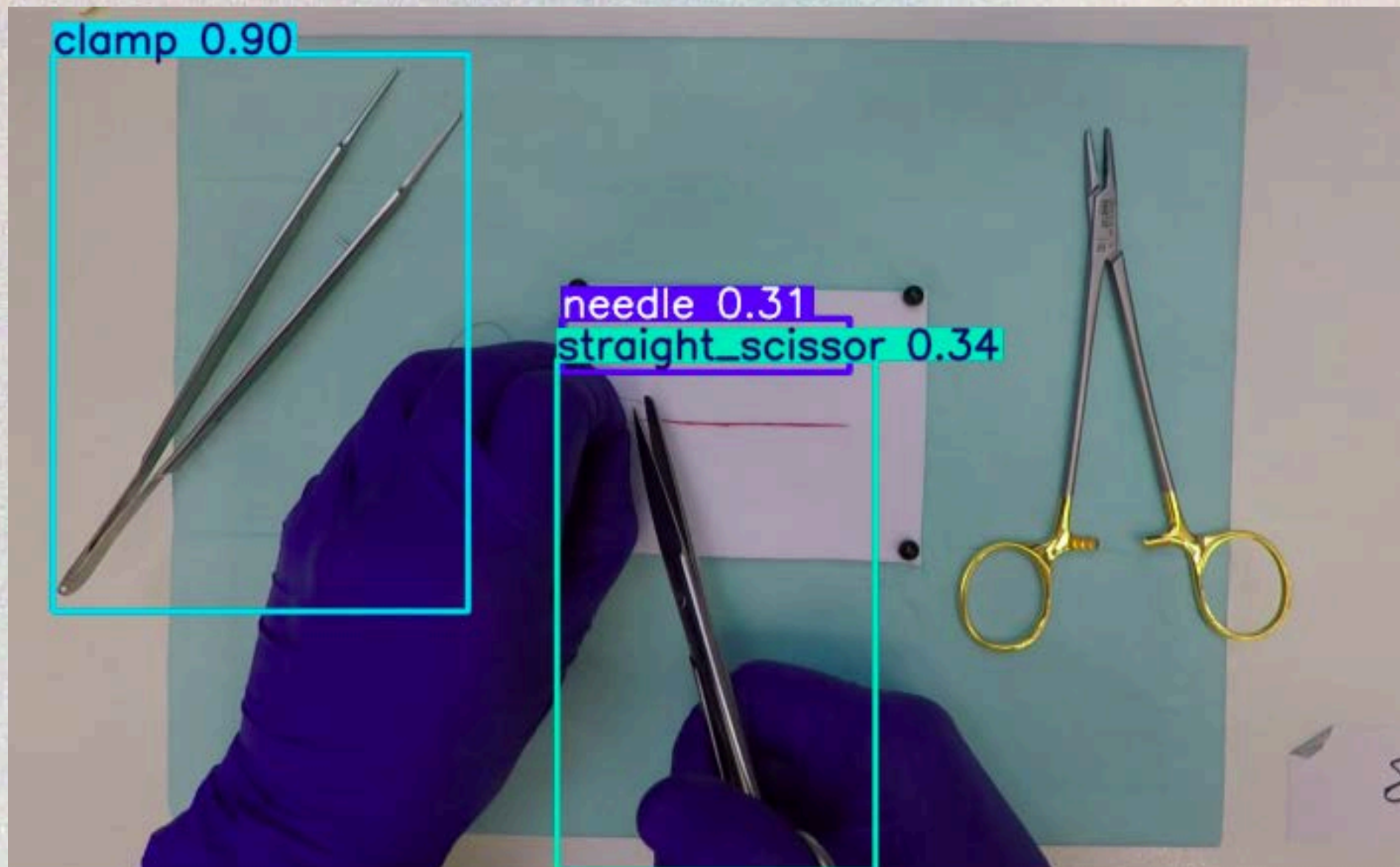
RESULTADOS - TASK 3

Designação do Modelo	mAP@0.5 geral	Velocidade de inferência
YoloFineTune	0.9788	0.9ms
YoloTrackingWithoutHands	0.8642	2.1ms
yoloTrackingWithHands	0.9897	0.8ms

HOTA (Higher Order Tracking Accuracy)

- Combina DetA (detection accuracy) e AssA (association accuracy).
- Penaliza erros de localização (ex: bounding boxes imprecisas) e trocas de IDs

RESULTADOS - TASK 3



TRABALHOS FUTUROS E CONCLUSÃO

- Mitigar classes desbalanceadas
- Especialistas alternativos
- Aumentar a capacidade da Rede
- Input Multi-frame para as Yolo



Universidade do Minho

APRENDIZAGEM PROFUNDA

David Teixeira PG55929

Eduardo Cunha PG55939

Nuno Rodrigues PG55966

Tiago Rodrigues PG56013

