

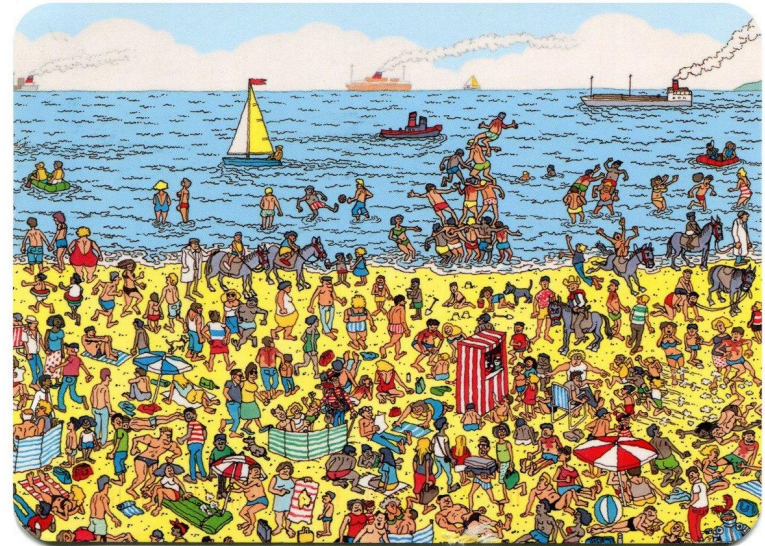
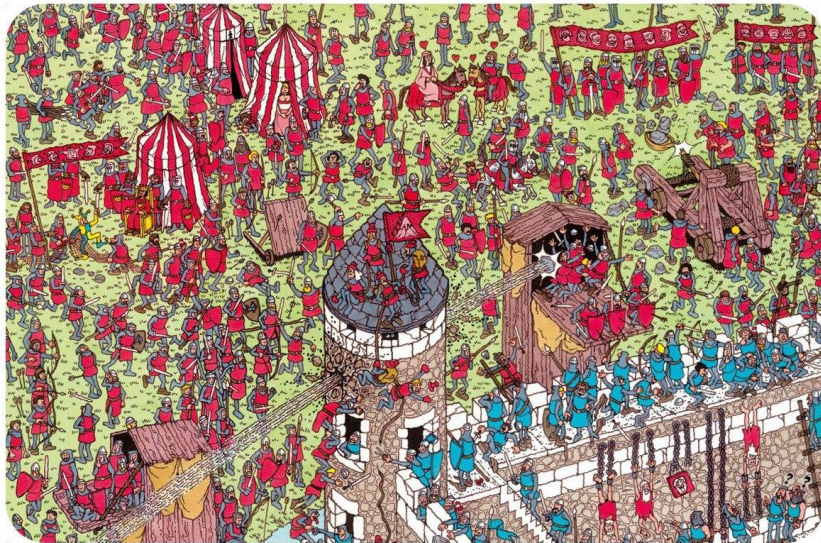


WALL-Y

Progetto di Intelligenza Artificiale - A.A. 2018/19
Anna Avena, Giulia Cantini, Matteo Del Vecchio, Simone Preite



Obiettivo del progetto





Difficoltà

- > Wally appare nelle immagini una volta sola → esiste una sola bounding box che è **ground truth**;
- > Le immagini hanno risoluzioni diverse → problema di **scaling**;
- > Wally, o la sua faccia, non appaiono quasi mai per intero → problema di **occlusion**;
- > Le immagini sono caratterizzate da molto **noise**.

Object Detection

Insieme di due task diversi:

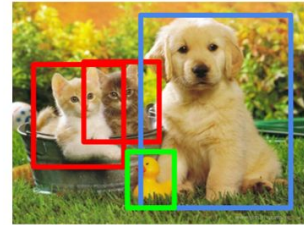
- *Object Classification* → classificazione dell'oggetto secondo una specifica label
- *Object Localization* → individuazione delle coordinate nell'immagine (regressione)

Classification

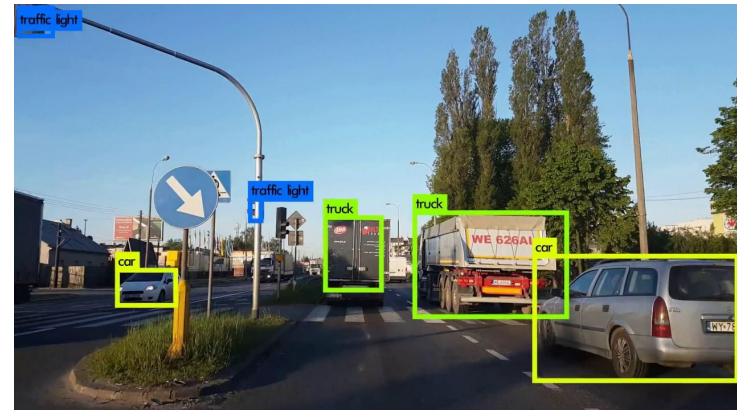


CAT

Object Detection



CAT, DOG, DUCK





Modello

Sono tre i modelli più famosi:

1. R-CNN
2. Fast R-CNN
3. Faster R-CNN

Per questo progetto abbiamo utilizzato la **FASTER R-CNN RESNET 101** che separa il task di classificazione da quello di localizzazione e ci restituisce una tupla con i valori relativi all'offset del bounding box.

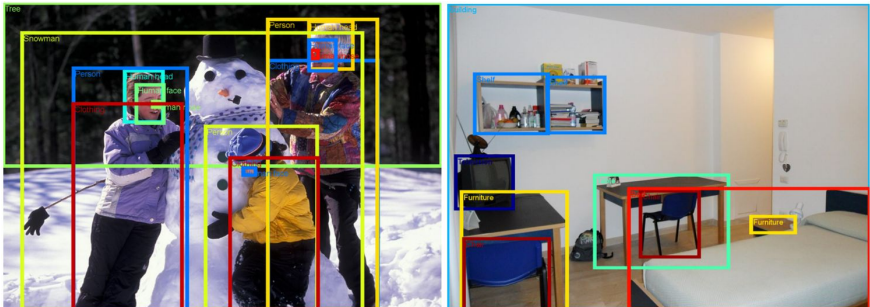
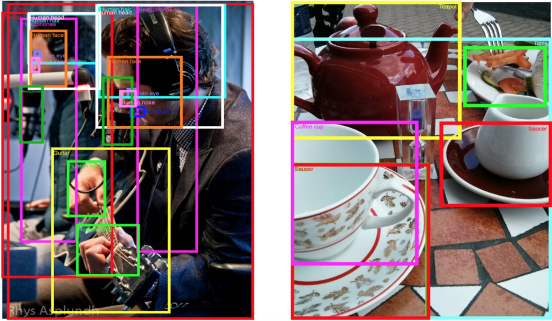
Wally = (x offset, y offset, width, height)

/coco:

- Contiene 91 categorie di oggetti comuni, ciascuna categoria con in media 5.000 istanze etichettate;
- Presta particolare attenzione al contesto e alla comprensione della scena;
- È il dataset che Tensorflow usa per pre-allenare i suoi modelli.

WALDO.

- Dataset customizzato COCO-like;
- 90% training set, 10% test set.



Preprocessing

- **Labeling** per preparare le annotazioni sulle immagini
- **Data augmentation** per evitare il rischio di overfitting
- **Cropping** per ridurre il problema della elevata dimensione delle immagini





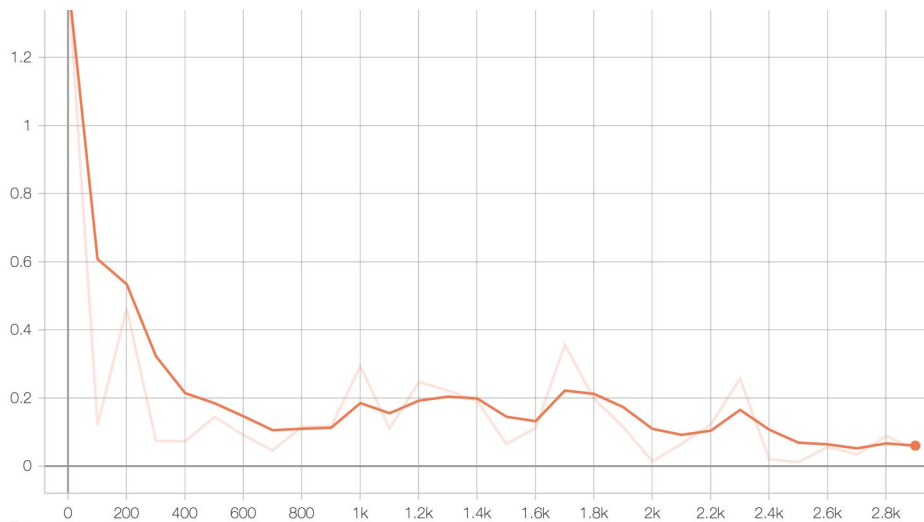
Metriche

Le metriche utilizzate per la valutazione del modello sul test set sono basate su COCO e comprendono:

- Intersection over Union (IoU)
- Precision
- Recall
- Precision-Recall Curve
- Recall-IoU Curve
- Average Precision (AP) e Average Recall (AR)

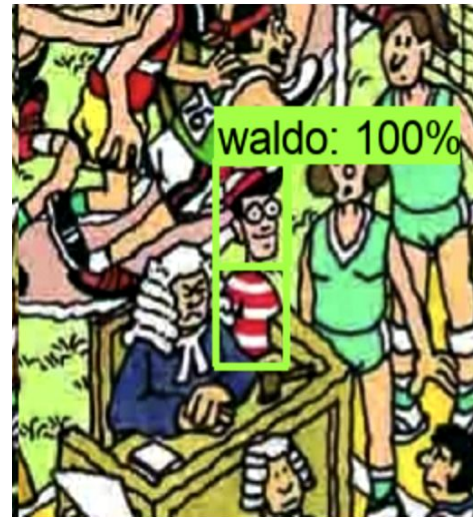
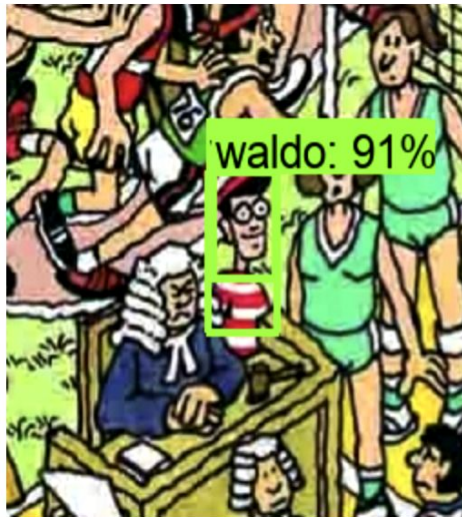
Training

- > Rete utilizzata: Faster R-CNN
- > Feature Extractor: ResNet101
- > Learning rate: $3e-4$
- > Nr. step: 3000
- > Joint Loss: Binary Cross-Entropy e L1 Loss



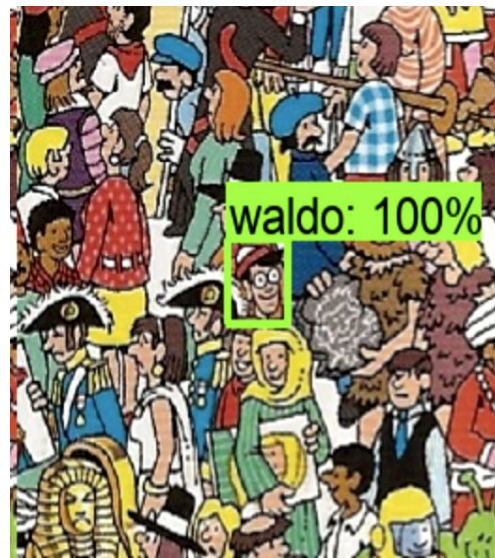
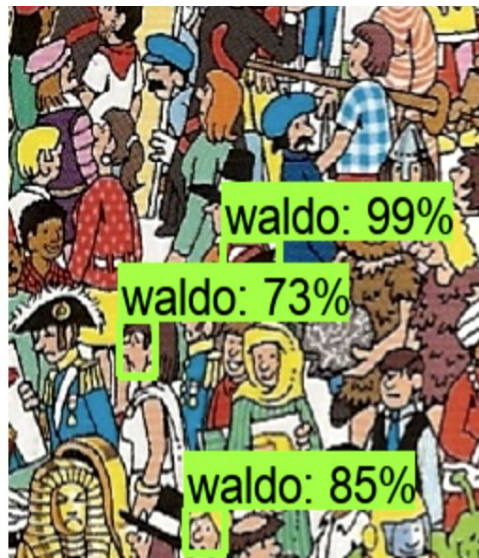
Risultati

> Inferenza



Risultati

> Inferenza





Conclusioni e sviluppi futuri

1. I risultati sono soddisfacenti, migliori di quelli che si sperava di ottenere.
2. Tentare un approccio in cui durante l'allenamento si usino immagini in cui Wally non si trova solo al centro.
3. Un approccio di “ricombinazione” delle tiles più efficiente.

Altri miglioramenti più generici: aumentare il numero di esempi nel dataset, provare tecniche automatiche di data augmentation e sperimentare tutte le tecniche con diversi modelli pretrained.

Grazie per l'attenzione

