

Progetto 5

Identificazione del testo nelle immagini

Obiettivo di questo progetto è l'identificazione del testo nelle immagini (Fig.1). A tal fine userete il dataset MSRA Text Detection 500 (MSRA-TD500) [1] costituito da circa 500 immagini e l'architettura RetinaNet proposta in [2]. Questa architettura si basa su una struttura encoder-decoder dove la parte di encoder, denominata anche *backbone*, viene realizzata attraverso una rete pre-addestrata per il riconoscimento di oggetti. In particolare, utilizzerete come *backbone* la rete EfficientNetB2. Per il training e il test utilizzerete l'implementazione in keras di RetinaNet scaricabile alla repository:

<https://github.com/fizyr/keras-retinanet>

In questo progetto i passi da seguire sono:

1. **Download dei dati.** Scaricare dal sito ufficiale [1] il dataset MSRA.
2. **Preparazione dei dati.** Dividete le immagini nei tre set di training (50%), validazione (10%) and test (40%). Preparate per ogni set, il file CSV con la lista delle immagini secondo le indicazioni della repository.
3. **Addestramento.** Utilizzate lo script `train.py` della repository per eseguire il training. Allo script di training è possibile passare diversi parametri come: la rete di backbone (`--backbone EfficientNetB2`), se riaddestrare o meno il backbone (`--freeze-backbone`), il batch-size (`--batch-size 1`), il numero di epoche (`--epochs 50`), il learning-rate (`--lr 1e-5`). Utilizzate le prestazioni sul set di validation per selezionare i migliori valori per questi parametri.
4. **Valutazione delle prestazioni.** Per valutare le prestazioni utilizzate lo script `evaluate.py` sul test set e calcolate il mean average precision (mAP). Provate a cambiare il backbone e vedete come variano le prestazioni.

References

- [1] C. Yao, X. Bai, W. Liu, Y. Ma, and Z. Tu., "Detecting Texts of Arbitrary Orientations in Natural Images" Computer Vision and Pattern Recognition (CVPR), 2012 [http://www.iapr-tc11.org/mediawiki/index.php/MSRA_Text_Detection_500_Database_\(MSRA-TD500\)](http://www.iapr-tc11.org/mediawiki/index.php/MSRA_Text_Detection_500_Database_(MSRA-TD500))



Figure 1: Esempi di identificazione del testo.

- [2] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal Loss for Dense Object Detection” International Conference on Computer Vision (ICCV), 2017.