

Rilevamento dei pedoni

Jasmine Panariello, Giovanni Riccardi, Riccardo Romano, Mario Tranchese

Progetto ESM-Gruppo 11

(a.a 2020-2021)

Il rilevamento dei pedoni è uno dei problemi presi in considerazione nell'ambito della computer vision, utile alle applicazioni di sicurezza e sorveglianza, la robotica mobile, la guida automatica, il riconoscimento di pedoni nella folla, ed altro ancora.

Fondamentale è l'accurata cura dei sistemi del rilevamento dei pedoni che ha un impatto diretto su quest'argomento.

L'obiettivo dunque, prendendo un'immagine arbitraria, è quello di rilevare se vi sono o meno pedoni e, se presenti, bisogna restituire la posizione e l'estensione dell'immagine di ogni pedone, disegnando attorno al soggetto un riquadro di delimitazione (bounding box) e associa ad esso la classe pedone. Le difficoltà associate al rilevamento dei pedoni sono dovute a: occlusione, variazione di scala, variazione di posa, da considerare durante la creazione dell'algoritmo di rilevazione, ma negli ultimi decenni ci sono stati notevoli progressi che sono stati applicati in alcuni sistemi tipici solo per degli scenari limitati. Il successo di questi sistemi viene attribuito a 2 passaggi chiave:

- evoluzioni nel campo nella rete neurale convoluzionale che ha avuto un impatto diretto su molte attività di visione artificiale compreso il rilevamento dei pedoni;
- iniziative di raccolta dei dataset guidate da alcuni ricercatori della comunità.

1. APPROCCIO IMPLEMENTATO

Si utilizza la rete neurale Keras-Retinanet fornita dalla repository di GitHub:

<https://github.com/fizyr/keras-retinanet> [1], per effettuare le fasi di training e testing sulle immagini.

Inoltre si utilizza l'approccio riportato nel documento [2] riguardante la Focal Loss che ci permette di distinguere un soggetto che rischia di confondersi con le texture dello sfondo.

2. DATASET

Il dataset utilizzato è il WiderPerson[3] composto da un totale di 13.382 immagini con 399.786 annotazioni (si suppone che siano in media 29,87 per immagine), avendo una densità di pedoni elevata rispetto ad altri dataset, comporta così maggiori occlusioni rendendo più ardua l'individuazione e la localizzazione dei pedoni.

È composto da 5 classi: pedoni, riders, persone parzialmente visibili, folla, regioni ignote.

Il dataset viene suddiviso in 3 subset: training set, validation set e testing set.

Successivamente viene convertito in 3 file .CSV come indicato nella repository[1] [4].

Inoltre si utilizza un file di mapping (mapping.CSV) in cui è indicata la classe pedestrians considerata (es. pedestrians, 0)

[4] esempio file .CSV

WiderPerson/Images/012239.jpg,357,92,393,181,pedestrians
WiderPerson/Images/012239.jpg,332,97,337,111,pedestrians
WiderPerson/Images/012239.jpg,320,100,325,114,pedestrians
WiderPerson/Images/012239.jpg,415,103,420,120,pedestrians
WiderPerson/Images/012239.jpg,410,104,413,114,pedestrians
WiderPerson/Images/012239.jpg,407,103,410,114,pedestrians

2.1 TRAINING

Per quanto riguarda la fase di training viene utilizzato lo script "*train.py*" contenuto nella repository[1], passando alla rete i seguenti parametri:

- backbone
Rete di backbone pre-addestrata
- freeze-backbone
Utilizzata per riaddestrare o meno il backbone
- steps
Numero di steps da eseguire per epoche
- epochs
Numero di iterazioni, aumentandole il modello migliora
- batch-size
Numero di esempi contenuti in ogni batch
- learning rate
Velocità di apprendimento

2.2 TESTING

Per quanto riguarda invece la fase di testing, viene utilizzato lo script "*evaluate.py*" contenuto sempre nella repository[1], per valutare le prestazioni della rete sulle immagini contenute nel test_set, restituendo come parametro il "mAP" ossia la precisione delle predizioni sulle immagini. Maggiore sarà questo valore e più efficiente sarà stata la fase di training sulle immagini, ricordando che il mAP (Mean Average Precision) è una metrica molto diffusa utilizzata per misurare l'accuratezza dei sistemi di Object Detection ed indica il valore di precisione medio compreso tra 0 ed 1.

esempio immagini Dataset WiderPerson [3]



Tabella dei risultati ottenuti

BACKBONE	FREEZE BACKBONE	BACH-SIZE	LEARNING RATE	mAP
MOBILENET128_0.75	✓	1	1e-5	0,03
MOBILENET128_0.75	✓	10	1e-5	0,08
EFFICIENTNETB2	✓	1	1e-4	0,58
EFFICIENTNETB2	✓	4	1e-4	0,70
EFFICIENTNETB2	✗	4	1e-4	0,71

3. RISULTATI SPERIMENTALI

Dopo aver effettuato le fasi di training e testing si passa a visualizzare l'immagine con le relative Bounding Box che saranno sempre più precise all'aumentare del valore di mAP, il quale è possibile migliorare andando a modificare i parametri in ingresso.

Inizialmente infatti i risultati ottenuti utilizzando la backbone Mobilenet128_0.75 non sono stati soddisfacenti arrivando ad un valore di mAP al di sotto dello 0,1 ; andando a modificare la backbone con EfficientNetB2 nonostante un valore iniziale pari ad 1 di batch-size, i risultati sono migliorati notevolmente con un valore di 0,58.

Successivamente aumentando il batch-size a 4 il valore di mAP si è innalzato fino a 0,7 restando costante anche omettendo il parametro "Freeze-Backbone".

Possiamo inoltre dire che andando a modificare il valore di batch-size e di learning rate, la rete aumenta l'efficienza e diminuisce i tempi di esecuzione del training.

In conclusione andando ad analizzare i risultati direttamente dalle immagini, la rete sembra comportarsi meglio nel momento in cui deve visualizzare soggetti completamente visibili, occultati o parzialmente visibili, mentre trova qualche difficoltà nel momento in cui deve riconoscere un pedone che rischia di confondersi con le texture dello sfondo.

Esempio risultato ottenuto dalla rete (mAP = 0,7)



Riferimenti bibliografici

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

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

[3] S. Zhang, Y. Xie, J. Wan, H. Xia, S. Li, and G. Guo, "WiderPerson: A Diverse Dataset for Dense Pedestrian Detection in the Wild" IEEE Transactions on Multimedia, 2019. <http://www.cbsr.ia.ac.cn/users/sfzhang/WiderPerson/>