

Object detection nell'infrarosso per sistemi di guida autonoma

My Kieu, Andrew D. Bagdanov, Marco Bertini, Alberto Del Bimbo

Università degli Studi di Firenze - MICC

[nome.cognome]@unifi.it

Abstract

Il sistema presentato, sviluppato nell'ambito di un progetto di ricerca per conto di RFI, effettua il riconoscimento di oggetti in video ripresi nel campo dell'infrarosso a scopo di guida autonoma di mezzi ferroviari e video sorveglianza. Lo scopo è quello di riconoscere persone, veicoli ed animali usando sensori termici, così da poter funzionare in condizioni di ripresa difficili, prive di illuminazione e con condizioni atmosferiche avverse.

1 Introduzione

In molti casi applicativi di guida autonoma risulta necessario usare telecamere termiche dato che il segnale ottenibile nel visibile non è in grado di fornire evidenze utilizzabili per sistemi di visione artificiale. Un esempio è la notte in casi di mancanza di illuminazione o comunque con illuminazione insufficiente, ma anche condizioni meteo avverse come nebbia o pioggia forte risultano difficoltose anche di giorno. In queste condizioni le telecamere termiche sono in grado di fornire immagini utilizzabili, per es. per riconoscere la presenza di persone, animali o veicoli. Per questo motivo sono spesso utilizzate in applicazioni di sorveglianza o anche di guida automatica.

Nel caso applicativo del sistema presentato in questo lavoro, ci si trova ad operare di notte in condizioni di mancanza di illuminazione, in qualsiasi condizione meteo possibile e con necessità di riconoscere la presenza di persone, animali e veicoli nei pressi di linee ferroviarie. L'obiettivo è quello di sviluppare un sistema di sorveglianza a bordo di un veicolo ferroviario che sia in grado di rilevare questi oggetti in tempo reale, mentre il mezzo si muove sui binari.

2 Il sistema

Per affrontare il problema si è partiti dall'architettura della rete YOLOv2 [Redmon *et al.*, 2016], per la detection di oggetti, esaminando diversi approcci di addestramento e fine-tuning, valutando quale metodo portasse alla migliore performance nel caso d'uso specifico di visione notturna usando solo il segnale termico.

In particolare, si è partiti da una rete YOLOv2 addestrata su PASCAL VOC per adattarla alle condizioni di un caso

applicativo vicino a quello del problema, ovvero quello della guida autonoma di auto (ADAS). Infatti è da notare come non esistano ancora dataset pubblici in ambito ferroviario. Sono state quindi testate tre tipologie di fine tuning della rete per valutarne il funzionamento

- fine tuning su dataset ADAS usando il segnale RGB di una telecamera nel visibile;
- fine tuning su dataset ADAS usando il segnale termico;
- fine tuning su dataset ADAS usando sia il segnale RGB nel visibile che il termico.

3 Risultati sperimentali

3.1 Dataset

Il dataset utilizzato per gli esperimenti è il KAIST Multi-spectral Pedestrian Dataset [Hwang *et al.*, 2015], che consiste di 95k coppie di immagini colore-termico (di dimensione 640×512) riprese da un veicolo. Tutte le coppie sono annotate manualmente indicando la presenza di persone (singole e in gruppo) e ciclisti per un totale di 103.128 annotazioni dense. Un esempio di immagini del dataset, riprese sia di giorno che di notte, è mostrato in Fig. 1. I migliori risultati presentati in letteratura usano la fusione del segnale visibile e termico.

3.2 Risultati

La comparazione con un sistema basato su fusione dei sensori (ACF+T+THOG [Hwang *et al.*, 2015]), la versione di YOLOv2 allenata su RGB, una versione allenata sul termico in combinazione con la luminanza del visibile (YOLO_TLV, [Vandersteegen *et al.*, 2018]) e le tre versioni di fine tuning proposte, mostra come la nostra proposta di addestrare su visibile e termico porti a risultati migliori in termini sia di miss rate che di precisione, come mostrato in Fig. 2 e 3.

In particolare si può notare come la fusione del segnale visibile nel metodo ACF+T+THOG si comporti peggio di reti addestrate specificatamente per il solo segnale termico, sebbene le condizioni operative prevedano una certa illuminazione, dato che le riprese sono state fatte in città. La comparazione con YOLO_TLV mostra comunque come la rete benefici durante l'addestramento del segnale RGB nel visibile, sebbene questo non venga più usato in fase di test.



Figura 1: Esempi di fotogrammi del dataset KAIST: ripresa nel visibile e nel termico di una scena diurna ed una notturna.

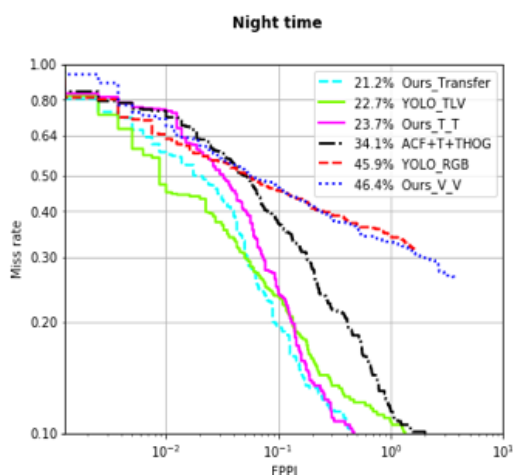


Figura 2: Comparazione tra i metodi proposti e approcci stato dell'arte su immagini notturne del dataset KAIST in termini di miss rate (valori più bassi sono migliori).

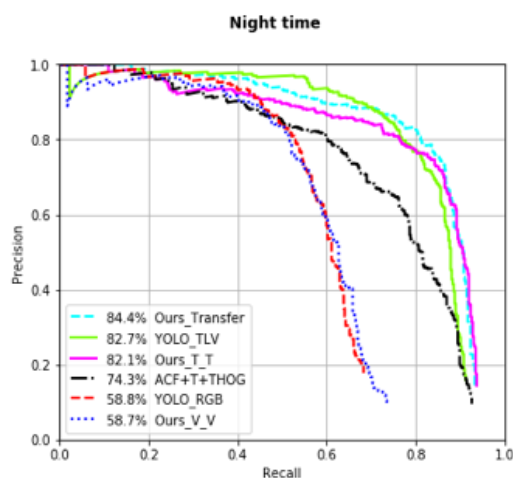


Figura 3: Comparazione tra i metodi proposti e approcci stato dell'arte su immagini notturne del dataset KAIST in termini di precision (valori più alti sono migliori).

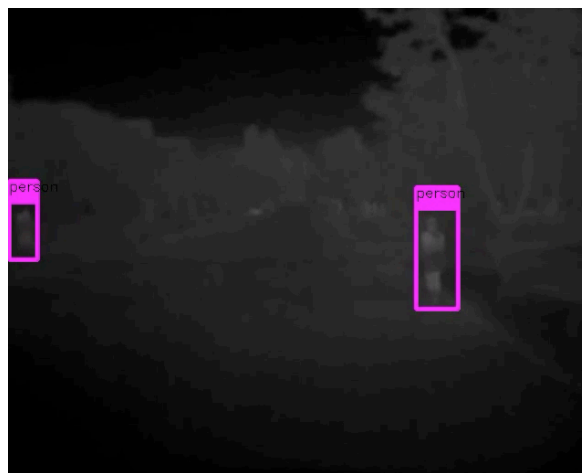


Figura 4: Esempio di riconoscimento della presenza di persone nel termico, usando il classificatore presentato, nel dataset KAIST.

4 Conclusioni

Il lavoro futuro prevede la creazione di un dataset di video di riprese da treni nel termico, lo studio di tecniche per l'adattamento non supervisionato e l'uso della coerenza temporale dei video, per migliorare ulteriormente la performance di detection.

Ringraziamenti Il lavoro è supportato in parte dal progetto "Progettazione e Realizzazione del Sistema di Visione Artificiale per l'Ispezione a Lunga e Breve Distanza da Mezzi a Guida Automatica".

Riferimenti bibliografici

- [Hwang *et al.*, 2015] Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, e In So Kweon. Multispectral pedestrian detection: Benchmark dataset and baselines. In *Proc. of CVPR*, 2015.
- [Redmon *et al.*, 2016] J. Redmon, S. Divvala, R. Girshick, e A. Farhadi. You only look once: Unified, real-time object detection. In *In Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [Vandersteegen *et al.*, 2018] Maarten Vandersteegen, Kristof Van Beeck, e Toon Goedemé. Real-time multispectral pedestrian detection with a single-pass deep neural network. In *Proc. of Image Analysis and Recognition*, 2018.