

Deep Learning per Videosorveglianza Aerea in Smart Cities

Danilo Avola¹, Luigi Cinque¹, Gian Luca Foresti²,
Cristiano Massaroni¹, Daniele Pannone¹, Claudio Piciarelli²

¹Dipartimento di Informatica, Sapienza Università di Roma, Via Salaria 113, 00198, Roma

²Dipartimento di Matematica, Informatica, e Fisica, Università di Udine, Via delle Scienze 206, 33100, Udine
{avola, cinque, massaroni, pannone}@di.uniroma1.it
{gianluca.foresti, claudio.piciarelli}@uniud.it

Abstract

Nella videosorveglianza aerea effettuata tramite multicotteri di piccola dimensione, ossia Unmanned Aerial Vehicles (UAVs), il monitoraggio delle aree d'interesse rappresenta una sfida estremamente complessa. Questo in quanto alcuni tasks, come il mosaicking, la change detection e l'object detection, devono essere effettuati in tempo reale e in maniera estremamente accurata. In questo contributo viene proposto un sistema di visione basato su multicottero UAV e tecniche di deep learning per il monitoraggio intelligente di zone d'interesse. I test effettuati utilizzando il dataset UAV Mosaicking and Change Detection (UMCD) hanno evidenziato ottimi risultati nei tasks sopraelencati.

1 Introduzione

Nell'ultimo decennio, i sistemi di visione basati su multicotteri UAVs (d'ora in avanti, semplicemente UAVs) hanno guadagnato sempre più notorietà, in quanto rendono possibile l'esecuzione di determinati task che in molti casi risulterebbero impossibili da compiere per sistemi simili basati su altre tipologie di velivoli. Ciò è dovuto alle caratteristiche fisiche di questi UAVs, le quali consentono, a questi ultimi, di avvicinarsi notevolmente a determinate aree, altrimenti irraggiungibili da aeromobili più voluminosi. In base a queste premesse, in questo contributo viene presentato un sistema di visione basato su UAV in grado di monitorare ampie aree in tempo reale, rilevando e classificando possibili cambiamenti sul terreno. Il sistema presenta due differenti modalità operative.

Nella prima modalità, l'UAV esegue uno o più voli sulla zona interessata, creando un mosaico di riferimento. Ogni volo è guidato da un insieme di coordinate GPS che specifica l'area in cui eseguire l'operazione di mosaicking. Tutti gli elementi rilevati sul terreno (la cui posizione sarà segnalata all'interno del mosaico attraverso le coordinate GPS) sono classificati per mezzo di una rete neurale di tipo Faster R-CNN [Ren *et al.*, 2017] precedentemente addestrata, la cui Region Proposal Network (RPN) è stata ridisegnata per aumentarne l'accuratezza in termini di classificazione. Successivamente, sorvolando nuovamente l'area di interesse, e tramite un algoritmo di change detection, è possibile identificare i cambiamenti avvenuti all'interno dell'area stessa. Inol-

tre, sempre utilizzando la rete neurale, sarà anche possibile identificare e classificare i cambiamenti avvenuti (il tutto, in tempo reale). Tramite la seconda modalità, il sistema può essere utilizzato per operazioni critiche quali, ad esempio, il Search and Rescue (SAR), dove l'elaborazione viene focalizzata nella fase di riconoscimento e classificazione in tempo reale degli elementi presenti nell'area di interesse. A causa della criticità di questi scenari applicativi, le fasi di mosaicking e change detection non sono eseguite, in quanto si rende necessario intervenire nel minor tempo possibile.

Il resto del contributo è strutturato come segue. Nella Sezione 2 è presentata l'architettura del sistema. Nella Sezione 3 sono illustrati gli esperimenti effettuati e i risultati ottenuti. Infine, la Sezione 4 conclude questo contributo.

2 Architettura del Sistema Proposto

Il sistema proposto (Figura 1) è composto da 3 moduli: mosaicking, change detection e object detection. Il primo è il modulo che si occupa di creare una mappa georeferenziata dell'area sorvolata dall'UAV tramite i frames inviati dallo stesso, implementato seguendo la pipeline proposta in [Avola *et al.*, 2017b]. Il secondo è il modulo che si occupa, da un lato, di confrontare i frames ricevuti in tempo reale dall'UAV con un mosaico generato precedentemente, dall'altro, di trovare i cambiamenti in quest'ultimo. Per questo modulo, è stata implementata la pipeline descritta in [Avola *et al.*, 2017a]. Infine, l'ultimo modulo, nonché cuore del sistema proposto, è il quello riguardante l'object detection. Tale modulo è utilizzato sia per il mosaicking e la change detection sia in modalità stand-alone. In dettaglio, per il sistema proposto è stata utilizzata una rete Faster-RCNN, ben nota in letteratura, la quale è stata modificata al fine di ottenere risultati migliori in termini di detection degli elementi, riduzione del numero di regioni generate durante la fase di estrazione delle stesse e, infine, velocità nella fase di testing. Tale rete è composta, a sua volta, da due ulteriori reti: una Convolutional Neural Network (CNN), utilizzata per il task di classificazione, e una Region Proposal Network (RPN), utilizzata per generare le bounding box (i.e., le regioni contenenti gli elementi da classificare). Una prima miglioria è stata quella di sostituire la rete CNN utilizzata nella Faster R-CNN con una ResNet [He *et al.*, 2016] a 50 livelli, per via degli ottimi risultati ottenuti nella competizione ImageNet 2015. Successivamente, abbiamo migliorato la RPN sostituendo la rete della Faster R-CNN con

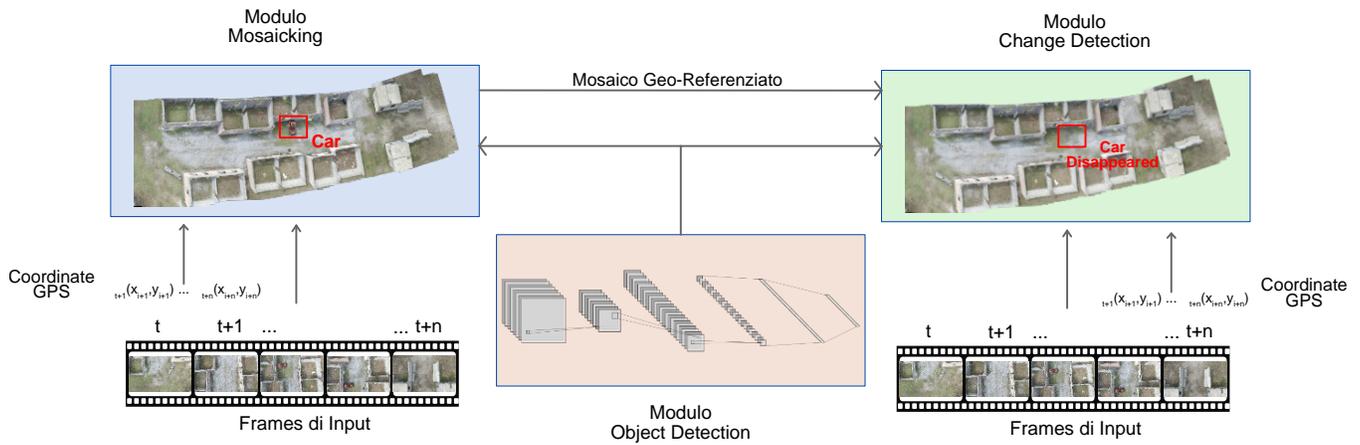


Figura 1: Architettura logica del sistema proposto.

Tabella 1: Valori mAP ottenuti attraverso la Faster R-CNN standard e la versione da noi proposta.

Modello	mAP
Faster R-CNN Standard	0,693004
Faster R-CNN Proposta	0,768859

una rete Inception [Szegedy *et al.*, 2015] da noi ridisegnata. La modifica della rete Inception consiste nell'utilizzare uno stack di 9 moduli Inception, il quale ha come input le features estratte dalla rete CNN, e come output un collegamento a un livello di regressione e un collegamento a un livello di classificazione.

3 Risultati Sperimentali

In questa sezione vengono riportati i risultati ottenuti negli esperimenti effettuati. Tutti i moduli del sistema sono stati testati utilizzando il dataset UMCD [Avola *et al.*, 2018], ma in questo contributo ci focalizzeremo sul modulo di object detection. La metrica utilizzata per misurare l'efficacia del modello proposto è la Mean Average Precision (mAP). Negli esperimenti effettuati, abbiamo utilizzato 3 classi: persona, veicolo e scatola, in quanto le più ricorrenti all'interno del dataset utilizzato. Al fine di testare l'efficacia della rete proposta, abbiamo effettuato il training sia di quest'ultima sia della Faster R-CNN classica. In Tabella 1, sono riportati i valori mAP ottenuti con le due reti.

4 Conclusioni

In questo contributo, è stato presentato un sistema basato su droni UAV in grado di effettuare il monitoraggio in modo automatico di una zona d'interesse. Il sistema funziona in due modalità. Nella prima, viene effettuato un mosaico dell'area di interesse e, nei voli successivi, vengono identificati i cambiamenti in tale zona tramite un algoritmo di change detection. Sia nella fase di mosaicking sia in quella di change detection, vengono classificati gli elementi (e i cambiamenti) all'interno dell'area osservata. Nella seconda modalità, si effettua direttamente la detection degli elementi all'interno

dell'area osservata, senza utilizzare un mosaico della stessa. Questa modalità è utile per le operazioni da effettuare in tempo reale, come ad esempio le SAR. In entrambe le modalità, viene utilizzato un modello di deep learning da noi proposto. I risultati ottenuti superano l'attuale stato dell'arte.

Riferimenti bibliografici

- [Avola *et al.*, 2017a] D. Avola, G. L. Foresti, N. Martinel, C. Micheloni, D. Pannone, e C. Piciarelli. Aerial video surveillance system for small-scale uav environment monitoring. In *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6, 2017.
- [Avola *et al.*, 2017b] D. Avola, G. L. Foresti, N. Martinel, C. Micheloni, D. Pannone, e C. Piciarelli. Real-time incremental and geo-referenced mosaicking by small-scale uavs. In Sebastiano Battiato, Giovanni Gallo, Raimondo Schettini, e Filippo Stanco, editors, *Image Analysis and Processing - ICIAP 2017*, pages 694–705, Cham, 2017. Springer International Publishing.
- [Avola *et al.*, 2018] D. Avola, L. Cinque, G. L. Foresti, N. Martinel, D. Pannone, e C. Piciarelli. A uav video dataset for mosaicking and change detection from low-altitude flights. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pages 1–11, 2018.
- [He *et al.*, 2016] K. He, X. Zhang, S. Ren, e J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [Ren *et al.*, 2017] S. Ren, K. He, R. Girshick, e J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149, 2017.
- [Szegedy *et al.*, 2015] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, e A. Rabinovich. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015.