

Videosorveglianza e Intelligenza Artificiale (AI): quale impatto “concreto”?

L'introduzione della AI nella videosorveglianza spingerà le soluzioni “edge” invece che “cloud” per l'analitica video. Nell'articolo l'ing. Angelo Carpani, libero professionista iscritto nell'Ordine degli Ingegneri della Provincia di Como n.2368 sez.A e docente per securindex formazione, spiega perché.

Affermare che l'intelligenza artificiale avrà un impatto importante anche nel mondo della videosorveglianza è scontato e quasi banale. Scrivo volgendo lo sguardo al futuro ma, in realtà, le cose stanno già cambiando e le nuove tecnologie di AI sono già presenti nel mercato della videosorveglianza.

Cambieranno i paradigmi nella scelta di una telecamera e nell'architettura dei sistemi di videosorveglianza. Rispetto a quest'ultimo aspetto, sono già intervenuto in diverse occasioni affermando che il mondo della videosorveglianza evolve sempre di più verso l'*edge* anziché il *cloud*.

Gestire a livello centralizzato su un server (modalità *server-based*) gli algoritmi di intelligenza artificiale applicati a centinaia di telecamere richiederebbero delle prestazioni in termini di capacità computazionali della CPU o GPU (Graphics Processing Unit) estremamente elevati, anche in termini di costi e consumi di energia. L'introduzione dell'AI nel mondo della videosorveglianza spingerà quindi ancora di più verso le soluzioni in *edge* anziché in *cloud*: le analitiche video basate su AI vengono applicate direttamente sulle telecamere (modalità *edge-based*) dotando le stesse di acceleratori hardware dedicati come *MLPU (Machine Learning Processing Unit)* e *DLPU (Deep Learning Processing Unit)*, evitando la complessità e i costi per il trasferimento dell'immagini su cloud per la loro elaborazione.

Dell'intelligenza artificiale ne sentiamo parlare ogni giorno sempre di più ed è un campo vastissimo, con innumerevoli applicazioni; in questo breve articolo vorrei soffermarmi in particolare sulle implicazioni nel mondo della videosorveglianza che sono, a mio parere, due.

Della prima ne ho già accennato sopra e riguarda le implicazioni sull'architettura dei sistemi di videosorveglianza in cui si sta attuando un passaggio nell'elaborazione delle immagini dalla *modalità server-based* alla *modalità edge-based*, quindi con una elaborazione delle immagini bordo camera e non su server.



La seconda riguarda la scelta della telecamera e non è così scontato/necessario e conveniente che si debba optare per forza per telecamere che implementano algoritmi di intelligenza artificiale come *Machine Learning* o *Deep Learning* più costose. Sarebbe forse necessario illustrare prima il vocabolario dei termini e delle tecnologie che si usano ma, senza entrare in troppi dettagli tecnici (piuttosto complessi e che richiederebbero molte pagine) mi limito ad accennarne qualcuno peculiare nell'elaborazione delle immagini.

Sentiamo ad esempio parlare di reti neurali che implementano algoritmi matematici per imitare il processo decisionale del cervello umano. I neuroni del cervello vengono replicati in un modello matematico della rete neurale umana per generare un algoritmo. Ad es. per il riconoscimento di immagini è più adatta una *rete neurale convoluzionale (CNN – Convolutional Neural Network)*, mentre una *rete neurale ricorrente (RNN – Recurrent Neural Network)* è più indicata per l'elaborazione del parlato.

Dicevo: non è così scontato/necessario e conveniente che si debba optare per forza per telecamere che implementano algoritmi di intelligenza artificiale. Ad esempio, per applicazioni di videosorveglianza in cui si debbano rilevare delle intrusioni in un ambiente in cui lo sfondo è fisso (ad esempio, una persona che entra in una reception) è sufficiente l'adozione di una telecamera che implementi un classico algoritmo di video analisi come la "sottrazione dello sfondo" (*background subtraction*).

La "sottrazione dello sfondo" richiede una potenza di elaborazione molto bassa ed è mirata agli oggetti in movimento, ignorando tutti quelli statici.

L'adozione invece di telecamere che implementano algoritmi di intelligenza artificiale come *Machine Learning* (ML) e *Deep Learning* (DL) dipende dal target che si vuole rilevare.

Prima di entrare nel merito occorre sapere che le telecamere che implementano algoritmi di intelligenza artificiale, per imparare a riconoscere gli oggetti e a classificarli in determinate categorie (classi), *vanno addestrate* con una grande quantità di dati campione o *dati di addestramento*; questi ultimi vengono etichettati, o meglio *annotati*. L'addestramento può essere fatto dal costruttore con un apprendimento di tipo:

- *supervisionato*: è il metodo più utilizzato nel Machine Learning e può essere descritto come un apprendimento basato su esempi. I dati di addestramento vengono annotati (o etichettati) chiaramente: i dati di input sono già abbinati ai risultati di output desiderati;

- *per rinforzo*: l'algoritmo non utilizza coppie dati/etichetta per l'addestramento, ma impara da solo dagli errori ricevendo un premio quando compie le scelte giuste.

Ebbene, senza farla troppo complicata (perché la materia è davvero complicata), possiamo dire che:

- se si ha necessità di riconoscere all'interno di una scena un "qualsiasi" oggetto in movimento, è sufficiente l'adozione di una telecamera che implementi un classico algoritmo di video analisi di tipo "background subtraction";

- se si vogliono riconoscere solo "oggetti specifici", occorre adottare una telecamera che implementi algoritmi di AI di tipo *Machine Learning addestrata sul campo* (generalmente sono necessari 100-300 campioni e necessita di ridotte risorse hardware);

- se si vogliono riconoscere determinate "classi" di oggetti (es. automobili piuttosto che motociclette), occorre adottare una telecamera che implementi algoritmi di AI di tipo *Deep Learning addestrata dal costruttore* (generalmente sono necessari 20.000-100.000 campioni e necessita di una elevata potenza computazionale).

La scelta di una telecamera non si baserà quindi più solo sui classici paradigmi quali la risoluzione, la sensibilità del sensore, il tipo di ottica, ecc. ma anche sul tipo di AI in funzione della scena che si vorrà inquadrare (ad esempio se lo sfondo è statico come la parete di una stanza, o in continuo movimento come i veicoli che percorrono una strada con traffico intenso) e degli oggetti che si vorranno rilevare (ad esempio se sono "specifici" o se ci si vuole limitare a rilevare il tipo o la "classe").

Nella progettazione di un impianto di videosorveglianza, con l'introduzione dell'AI, bisognerà quindi fare riferimento "concretamente" a dei nuovi paradigmi che riguardano sia l'architettura degli impianti (sempre più verso l'*edge*) che la scelta delle telecamere.