

Intelligenza artificiale nelle analitiche video

Considerazioni sulle analitiche basate su machine learning e deep learning

Marzo 2021

Sommario

1	Sommario	3
2	Introduzione	4
3	Intelligenza artificiale, machine learning e deep learning	4
	3.1 Machine learning	5
	3.2 Deep learning	6
	3.3 Machine learning classico e deep learning a confronto	6
4	Le fasi del machine learning	7
	4.1 Raccolta e annotazione dei dati	7
	4.2 Addestramento	7
	4.3 Test	9
	4.4 Implementazione	9
5	Analitiche edge-based	9
6	Acceleratori hardware	9
7	Intelligenza artificiale: siamo solo all'inizio	10
8	Considerazioni per prestazioni analitiche ottimali	11
	8.1 Utilizzabilità dell'immagine	11
	8.2 Rilevamento delle distanze	11
	8.3 Configurazione di allarmi e registrazioni	12
	8.4 Manutenzione	13
9	Privacy e integrità personale	13
10	Appendice	15
	10.1 Reti neurali	15
	10.2 Reti neurali convoluzionali (CNN)	16

1 Sommario

Le analisi video basate su AI sono uno dei temi più dibattuti nel settore della sorveglianza. Alcune applicazioni possono velocizzare notevolmente l'analisi dei dati e automatizzare operazioni ripetitive. Tuttavia, le soluzioni odierne con intelligenza artificiale non possono sostituirsi all'esperienza dell'uomo e alle sue capacità decisionali. La vera forza sta nella combinazione tra i due metodi, ovvero nello sfruttare le soluzioni AI per aumentare l'efficienza di un operatore.

Il concetto di intelligenza artificiale comprende gli algoritmi di machine learning e deep learning. Entrambi costruiscono automaticamente un modello matematico utilizzando notevoli quantità di dati campione (*dati di addestramento*). In questo modo, il sistema impara a calcolare i risultati senza essere stato programmato specificamente. Un algoritmo AI viene sviluppato mediante un processo iterativo: si ripete un ciclo di raccolta dei dati di addestramento, etichettatura dei dati, uso dei dati etichettati per addestrare l'algoritmo e test dell'algoritmo, fino a raggiungere il livello qualitativo desiderato. Quindi, l'algoritmo è pronto per essere utilizzato in applicazioni analitiche che possono essere acquistate e utilizzate in un sito da sorvegliare. A questo punto, tutto l'addestramento è completato e l'applicazione non impara niente di nuovo.

Un compito tipico delle analitiche video basate su AI è individuare e distinguere persone e veicoli in un flusso video. Un algoritmo di *machine learning* ha imparato la combinazione di tratti visivi che definisce questi oggetti. Un algoritmo di *deep learning* è più rifinito e, se opportunamente addestrato, può rilevare oggetti molto più complessi. Tuttavia, richiede sforzi nettamente maggiori per lo sviluppo e l'addestramento, nonché più risorse di calcolo per l'uso dell'applicazione finita. Per esigenze di sorveglianza ben specifiche, dunque, è opportuno valutare se possa bastare un'applicazione con machine learning dedicata e ottimizzata.

L'evoluzione degli algoritmi e le maggiori capacità di elaborazione delle telecamere hanno permesso di eseguire analitiche video basate su AI direttamente sulla telecamera (modalità edge-based) anziché su server (modalità server-based). Questo migliora le funzionalità in tempo reale perché le applicazioni possono accedere immediatamente al materiale video non compresso. Dotando le telecamere di acceleratori hardware dedicati come MLPU (Machine Learning Processing Unit) e DLPU (Deep Learning Processing Unit), le analitiche edge-based possono essere implementate consumando meno energia rispetto a quando si utilizza una CPU o GPU (Graphics Processing Unit).

Prima di installare un'applicazione videoanalitica basata su AI, è necessario studiare e seguire attentamente le raccomandazioni dei produttori in base alle precondizioni e alle limitazioni note. Ogni sistema di sorveglianza è unico e le prestazioni delle applicazioni devono essere valutate sito per sito. Se la qualità è inferiore alle attese, occorre indagare a livello olistico e non solo sull'applicazione analitica. Le prestazioni delle analitiche video dipendono da molti fattori legati all'hardware della telecamera, alla sua configurazione, alla qualità video, alle dinamiche nella scena e all'illuminazione. In molti casi, conoscere gli effetti di questi fattori e ottimizzarli di conseguenza consente di aumentare le prestazioni delle analitiche video utilizzate nel sistema.

Poiché l'intelligenza artificiale è sempre più utilizzata nella videosorveglianza, i vantaggi in termini di efficienza operativa e in nuovi casi d'uso devono essere bilanciati valutando attentamente quando e dove applicare la tecnologia.

2 Introduzione

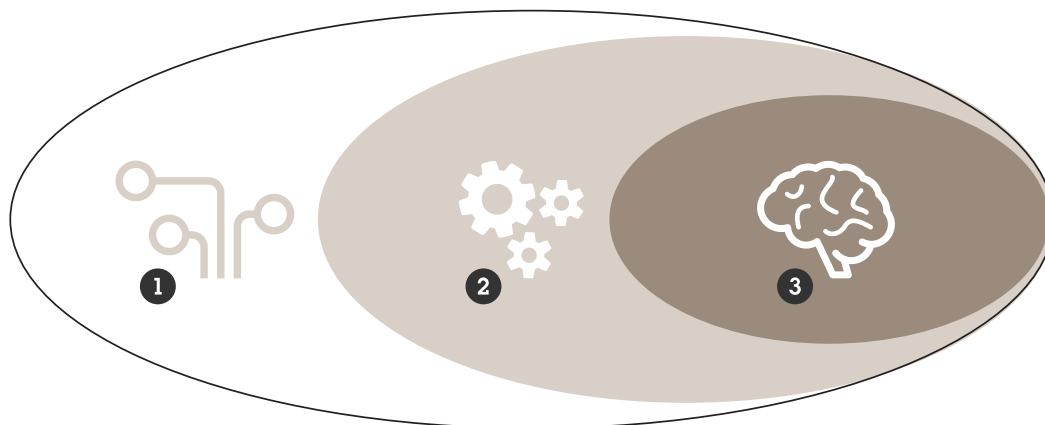
L'intelligenza artificiale (AI) viene sviluppata e discussa fin dall'invenzione dei primi computer. Anche se le sue implementazioni più rivoluzionarie non sono ancora disponibili, oggi le tecnologie basate sull'intelligenza artificiale sono ampiamente utilizzate per svolgere operazioni ben definite in applicazioni come il riconoscimento vocale, i motori di ricerca e gli assistenti virtuali. L'intelligenza artificiale è sempre più utilizzata anche nella sanità: ad esempio è utile nella diagnostica a raggi X e nell'analisi delle scansioni della retina.

Le analisi video basate su AI sono uno dei temi più dibattuti nella sorveglianza e alimentano grandi aspettative. Sul mercato esistono applicazioni che utilizzano algoritmi AI per velocizzare efficacemente l'analisi dei dati e automatizzare operazioni ripetitive. Ma in un contesto più ampio di sorveglianza, l'intelligenza artificiale di oggi e del prossimo futuro deve essere considerata semplicemente un elemento, tra i tanti, nel processo di costruzione di soluzioni accurate.

Questo documento tecnico illustra le nozioni tecniche fondamentali sugli algoritmi di machine learning e deep learning e spiega come possono essere sviluppati e applicati per le analitiche video. Quindi, descrive brevemente gli acceleratori hardware utilizzati per l'intelligenza artificiale e i pro e i contro delle analitiche basate su AI in modalità edge e su server. Infine, il documento spiega come ottimizzare le precondizioni per aumentare le prestazioni delle analitiche video basate su AI, tenendo in considerazione un ampio numero di fattori.

3 Intelligenza artificiale, machine learning e deep learning

L'intelligenza artificiale (AI) è un concetto ampio, associato alle macchine che possono risolvere operazioni complesse dimostrando tratti tipici di intelligenza. Il deep learning e il machine learning sono sottoinsiemi dell'intelligenza artificiale.



- 1 *Intelligenza artificiale*
- 2 *Machine learning*
- 3 *Deep learning*

3.1 Machine learning

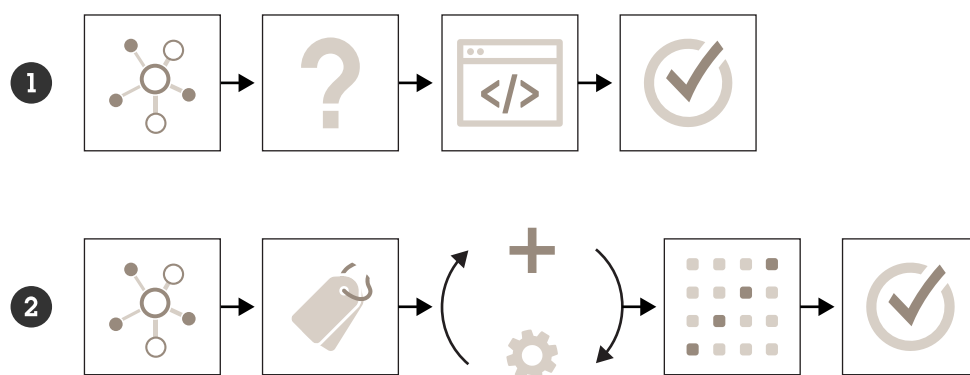
Il machine learning è un sottoinsieme dell'intelligenza artificiale che utilizza algoritmi di apprendimento statistici per costruire sistemi capaci di imparare automaticamente e migliorare con l'addestramento, senza essere programmati esplicitamente.

In questo capitolo, distingueremo la programmazione tradizionale dal machine learning nel contesto della *visione artificiale*, ovvero la disciplina che permette ai computer di capire cosa succede in una scena analizzando le immagini o i video.

La visione artificiale ottenuta con la programmazione tradizionale si basa su metodi che calcolano le *caratteristiche* di un'immagine: è il caso dei programmi per computer che cercano bordi pronunciati e punti angolari. Queste caratteristiche devono essere definite manualmente dallo sviluppatore di un algoritmo, che sa cosa è importante nei dati d'immagine. Quindi, lo sviluppatore combina le caratteristiche nell'algoritmo in modo che possa concludere ciò che si riscontra nella scena.

Gli algoritmi di machine learning costruiscono automaticamente un modello matematico utilizzando grandi quantità di dati campione, o *dati di addestramento*. In questo modo, il sistema impara a prendere decisioni calcolando i risultati senza essere stato programmato specificamente. Le caratteristiche vengono comunque definite manualmente, ma il modo di combinarle viene imparato dall'algoritmo mediante l'esposizione a grandi quantità di dati di addestramento etichettati, o meglio *annotati*. In questo documento, la tecnica che prevede l'uso di caratteristiche definite manualmente in combinazioni apprese viene detta *machine learning classico*.

In altre parole, in un'applicazione di machine learning è necessario addestrare il computer a ottenere il programma che desideriamo. I dati vengono raccolti e quindi annotati dall'uomo, che a volte aiutato con la preannotazione dai computer server. Il risultato viene inserito nel sistema e il processo continua fino a quando l'applicazione impara a sufficienza per rilevare ciò che volevamo, ad esempio uno specifico tipo di veicolo. Il modello addestrato diventa il programma. È opportuno notare che, quando il programma è terminato, il sistema non impara niente di nuovo.



- 1 *Programmazione tradizionale:*
I dati vengono raccolti. Vengono definiti i criteri del programma. Il programma viene codificato (da una persona). Fine.
- 2 *Machine learning:*
I dati vengono raccolti. I dati vengono etichettati. Il modello viene sottoposto a un processo di addestramento iterativo. Il modello addestrato finalizzato diventa il programma. Fine.

Nel realizzare un programma per la visione artificiale, il vantaggio dell'intelligenza artificiale rispetto alla programmazione tradizionale è la possibilità di elaborare grandi quantità di dati. Un computer

può processare migliaia di immagini senza deconcentrarsi, mentre un programmatore prima o poi si stancherà e perderà la concentrazione. In questo modo, l'intelligenza artificiale può rendere l'applicazione sostanzialmente più accurata. Tuttavia, più l'applicazione è complicata, più per la macchina è difficile produrre il risultato desiderato.

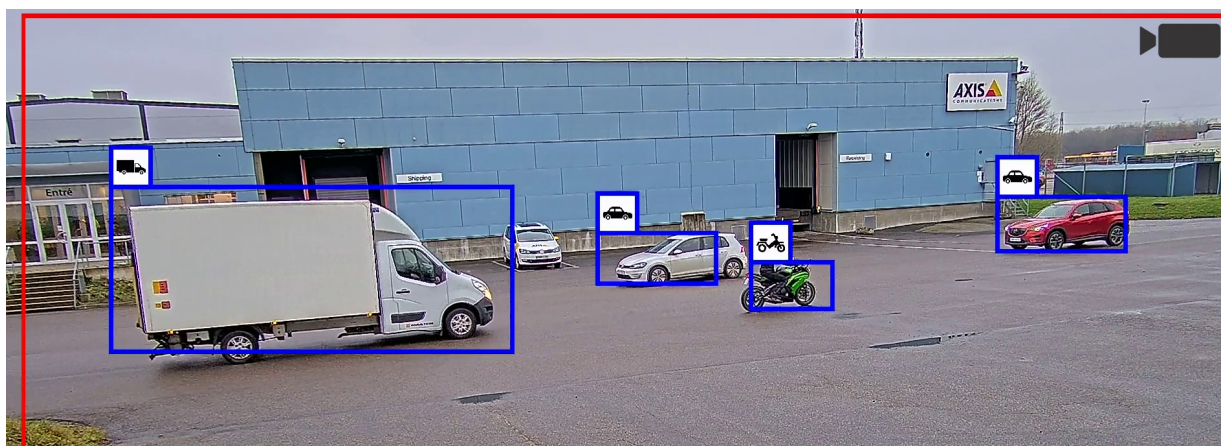
3.2 Deep learning

Il deep learning è una versione rifinita del machine learning. Sia l'estrazione delle caratteristiche (o feature) che la loro combinazione in strutture profonde di regole per produrre un output vengono apprese a partire dai dati. L'algoritmo può definire automaticamente le caratteristiche da cercare nei dati di addestramento. Inoltre, può apprendere strutture molto profonde di combinazioni di caratteristiche concatenate.

Il nucleo degli algoritmi di deep learning si ispira al funzionamento dei neuroni. Il cervello forma una conoscenza di livello superiore combinando gli output dei neuroni in una gerarchia profonda, o *rete*, di regole concatenate. Il cervello è un sistema in cui le combinazioni stesse sono formate da neuroni, eliminando la distinzione tra estrazione e combinazione delle caratteristiche e rendendole in un certo senso identiche. Queste strutture sono state simulate dai ricercatori con le cosiddette *reti neurali artificiali*, il tipo di algoritmo più utilizzato nel deep learning. Per una breve descrizione delle reti neurali si rimanda all'appendice del presente documento.

Utilizzando gli algoritmi di deep learning, è possibile creare rilevatori ottici evoluti e addestrarli automaticamente a individuare oggetti molto complessi, a prescindere da dimensioni, rotazione e altre variazioni.

Il motivo di una tale versatilità è che i sistemi di deep learning riescono a imparare da una quantità di dati molto maggiore e da dati molto più variegati rispetto ai classici sistemi di machine learning. Nella maggior parte dei casi, superano di gran lunga gli algoritmi di visione artificiale creati manualmente. Questo rende il deep learning particolarmente adatto a problemi complessi, in cui la combinazione delle caratteristiche non può essere realizzata facilmente da esperti umani: è il caso della classificazione delle immagini, dell'elaborazione linguistica e del rilevamento di oggetti.



Il rilevamento basato su deep learning è in grado di classificare oggetti complessi. In questo esempio, l'applicazione analitica riesce non solo a rilevare i veicoli, ma anche a classificarli per tipo.

3.3 Machine learning classico e deep learning a confronto

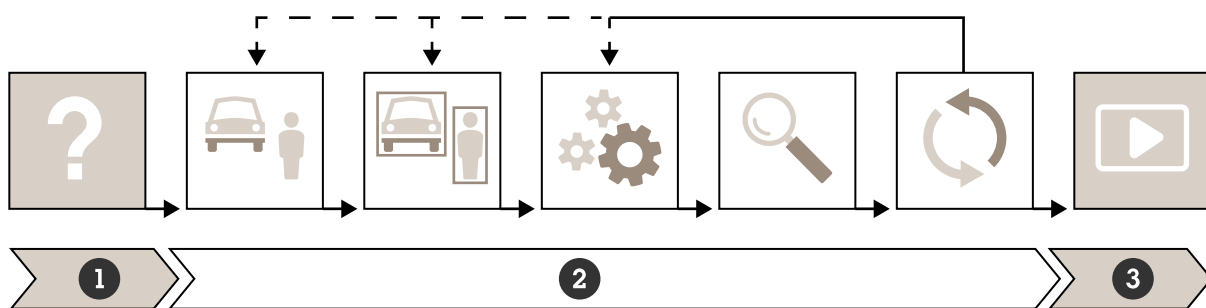
Nonostante le somiglianze, in genere un algoritmo di deep learning utilizza una serie molto più grande di combinazioni di caratteristiche apprese rispetto a un algoritmo di machine learning classico. Questo

significa che le analitiche basate su deep learning possono essere molto più versatili e, se opportunamente addestrate, imparare a svolgere operazioni molto più complesse.

Per le specifiche analitiche di sorveglianza, però, può bastare un algoritmo di machine learning classico dedicato e ottimizzato. In ambiti ben specifici, può dare risultati simili a quelli di un algoritmo di deep learning, eseguendo però meno operazioni matematiche: in questo modo si ha una maggiore efficienza energetica e in termini di costi. Inoltre, servono molti meno dati di addestramento e si riducono enormemente gli sforzi in fase di sviluppo.

4 Le fasi del machine learning

Prima di poter utilizzare l'applicazione analitica finale, lo sviluppo di un algoritmo di machine learning deve seguire una serie di passaggi e iterazioni, raffigurata schematicamente di seguito. Al centro di un'applicazione analitica ci sono uno o più algoritmi, ad esempio un rilevatore di oggetti. Nel caso delle applicazioni basate su deep learning, il nucleo dell'algoritmo è il modello di deep learning.



- 1 *Preparativi: definizione delle finalità dell'applicazione.*
- 2 *Addestramento: raccolta dei dati di addestramento. Annotazione dei dati. Addestramento del modello. Test del modello. Se la qualità è inferiore alle attese, i passaggi precedenti vengono ripetuti con un ciclo di miglioramento iterativo.*
- 3 *Implementazione: installazione e uso dell'applicazione finita.*

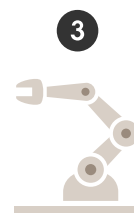
4.1 Raccolta e annotazione dei dati

Per sviluppare un'applicazione analitica basata su AI è necessario raccogliere grandi quantità di dati. Nella videosorveglianza, si tratta in genere di immagini e clip video di persone, veicoli o altri oggetti di interesse. Per rendere i dati riconoscibili da una macchina o un computer, è necessario un processo di annotazione dei dati: ovvero, gli oggetti pertinenti devono essere categorizzati ed etichettati. L'annotazione dei dati è un lavoro manuale e impegnativo. I dati preparati devono avere una varietà sufficientemente ampia e rilevante per il contesto in cui si utilizzerà l'applicazione analitica.

4.2 Addestramento

L'addestramento, o apprendimento, è la fase in cui il modello riceve dati annotati e si utilizza un framework per modificare e migliorare ciclicamente il modello, fino a ottenere la qualità desiderata. In altre parole, il

modello viene ottimizzato per assolvere il compito prestabilito. L'addestramento può essere eseguito con tre metodi principali.



- 1 *Apprendimento supervisionato: il modello impara a eseguire previsioni accurate*
- 2 *Apprendimento non supervisionato: il modello impara a identificare cluster*
- 3 *Apprendimento per rinforzo: il modello impara dagli errori*

4.2.1 Apprendimento supervisionato

L'apprendimento supervisionato è il metodo più utilizzato oggi nel machine learning. Può essere descritto come un apprendimento basato su esempi. I dati di addestramento vengono annotati chiaramente: ovvero, i dati di input sono già abbinati ai risultati di output desiderati.

In genere, l'apprendimento supervisionato richiede grandissime quantità di dati annotati e le prestazioni dell'algoritmo addestrato dipendono direttamente dalla qualità dei dati di addestramento. L'aspetto qualitativo più importante è utilizzare un set di dati che rappresenti tutti i potenziali dati di input in una situazione di utilizzo reale. Nel caso dei rilevatori di oggetti, lo sviluppatore deve sempre addestrare l'algoritmo con un'ampia varietà di immagini e con diversi esempi di oggetti, orientamenti, dimensioni, condizioni di illuminazione, sfondi e distrazioni. Solo se i dati di addestramento sono rappresentativi per il caso d'uso previsto, l'applicazione analitica finale sarà in grado di effettuare previsioni accurate anche quando elaborerà dati nuovi e non visti durante la fase di addestramento.

4.2.2 Apprendimento non supervisionato

L'apprendimento non supervisionato utilizza algoritmi per analizzare e raggruppare i set di dati non etichettati. Non si tratta di un metodo di addestramento comune nel settore della sorveglianza, perché il modello richiede molte calibrazioni e test ma la qualità può ancora risultare imprevedibile.

I set di dati devono essere rilevanti per l'applicazione analitica ma non devono essere etichettati o marcati chiaramente. Il lavoro di annotazione manuale viene eliminato, ma il numero di immagini o video necessari per l'addestramento deve aumentare di diversi ordini di grandezza. Durante la fase di addestramento, il modello, supportato dal framework, identifica le caratteristiche comuni all'interno dei set di dati. In questo modo, durante la fase di implementazione saprà raggruppare i dati in base a criteri comuni, riscontrando anche anomalie che non rientrano nei gruppi appresi.

4.2.3 Apprendimento per rinforzo

L'apprendimento per rinforzo è utilizzato ad esempio nella robotica, nell'automazione industriale e nel planning strategico aziendale; tuttavia, poiché richiede feedback in grandi quantità, ha un uso limitato nella sorveglianza di oggi. L'apprendimento per rinforzo consiste nell'adottare misure idonee per aumentare al massimo il potenziale *premio* in una situazione specifica, premio che aumenta quando il modello compie le scelte giuste. L'algoritmo non utilizza le coppie dati/etichetta per l'addestramento, ma viene invece ottimizzato testandone le decisioni attraverso l'interazione con l'ambiente mentre si misura il premio. L'obiettivo dell'algoritmo è imparare una strategia operativa che permetta di aumentare al massimo il premio.

4.3 Test

Una volta che il modello è addestrato, deve essere testato accuratamente. Questo passaggio prevede in genere una parte automatizzata, integrata da un test completo in situazioni di utilizzo concrete.

Nella parte automatizzata, l'applicazione viene valutata utilizzando nuovi set di dati, non visti dal modello durante l'addestramento. Se le valutazioni non corrispondono alle attese, il processo ricomincia daccapo: si raccolgono nuovi dati per l'addestramento, si effettuano o rfiniscono le annotazioni e si addestra nuovamente il modello.

Dopo aver raggiunto il livello qualitativo desiderato, inizia un test sul campo. In questo test, l'applicazione viene esposta a scenari reali. La quantità e le variazioni dipendono dall'ambito di utilizzo dell'applicazione. Più l'ambito è ristretto, meno è necessario testare variazioni. Se è più ampio, occorrono più test.

Quindi, i risultati vengono nuovamente confrontati e valutati. Questo passaggio può far ricominciare nuovamente il processo. Un altro potenziale esito potrebbe essere la definizione delle precondizioni, spiegando uno scenario in cui l'applicazione non è consigliata per l'uso o è consigliata solo in parte.

4.4 Implementazione

La fase di implementazione è anche detta fase di inferenza o predizione. L'*inferenza o predizione* è il processo di esecuzione di un modello di machine learning addestrato. L'algoritmo utilizza ciò che ha imparato durante la fase di addestramento per produrre l'output desiderato. Nel contesto delle analitiche di sorveglianza, la fase di inferenza è l'applicazione eseguita in un sistema di sorveglianza che monitora scene reali.

Per ottenere buone prestazioni in tempo reale quando si esegue un algoritmo di machine learning su dati audio o video, in genere occorre un acceleratore hardware specifico.

5 Analitiche edge-based

Le analitiche video ad alte prestazioni venivano utilizzate su server perché richiedevano più energia (e un maggiore raffreddamento) di quanta potesse offrirne una telecamera. Tuttavia, l'evoluzione degli algoritmi e la maggiore potenza di calcolo offerta dai dispositivi edge negli ultimi anni hanno permesso di eseguire le analitiche video basate su AI anche in modalità edge.

Le applicazioni analitiche edge offrono vantaggi evidenti: hanno accesso al materiale video non compresso con una latenza molto bassa, consentendo di eseguire applicazioni in tempo reale ed evitando i costi supplementari e le complessità del trasferimento dei dati su cloud per i calcoli. Le analitiche edge-based hanno anche costi dell'hardware e di utilizzo più bassi perché riducono le risorse necessarie sul server del sistema di sorveglianza.

Alcune applicazioni possono trarre benefici combinando l'elaborazione edge-based e server-based, con una pre-elaborazione sulla telecamera e un'elaborazione successiva sul server. Un sistema ibrido del genere può agevolare la scalabilità delle applicazioni analitiche, e quindi ridurre i costi, lavorando su diversi flussi delle telecamere.

6 Acceleratori hardware

Anche se spesso è possibile eseguire una specifica applicazione analitica su varie piattaforme, l'uso di un acceleratore hardware dedicato consente di ottenere prestazioni molto più elevate quando la potenza è limitata. Gli acceleratori hardware consentono un'implementazione a basso consumo energetico delle

applicazioni analitiche e possono eventualmente essere integrati da risorse computazionali su server e cloud.

- **GPU (Graphics Processing Unit).** Le GPU sono state sviluppate principalmente per applicazioni grafiche ma vengono utilizzate anche per l'accelerazione AI su piattaforme server e cloud. Benché a volte siano utilizzate anche in sistemi embedded (edge), le GPU non sono ottimali per l'inferenza nel machine learning perché consumano molta energia.
- **MLPU (Machine Learning Processing Unit).** Una MLPU può accelerare l'inferenza di specifici algoritmi di machine learning classico per svolgere task di visione artificiale con una efficienza energetica molto alta. Sono progettate per rilevare in tempo reale un numero limitato di oggetti simultanei, ad esempio persone e veicoli.
- **DLPU (Deep Learning Processing Unit).** Le telecamere con DLPU integrata riescono ad accelerare l'inferenza di algoritmi di deep learning generici con un'alta efficienza energetica, consentendo una classificazione degli oggetti più dettagliata.

7 Intelligenza artificiale: siamo solo all'inizio

È interessante confrontare le potenzialità di una soluzione dotata di intelligenza artificiale con ciò che può fare un essere umano. Mentre gli operatori di videosorveglianza riescono a rimanere vigili solo per un breve periodo di tempo, un computer può continuare a elaborare grandi quantità di dati in modo estremamente veloce senza mai stancarsi. Tuttavia, sarebbe un grosso equivoco ipotizzare che le soluzioni AI possano sostituire un operatore umano. La vera forza sta in una combinazione realistica, ovvero nello sfruttare le soluzioni AI per aumentare l'efficienza di un operatore.

Spesso, si sente dire che le soluzioni basate su machine learning o deep learning sono capaci di apprendere automaticamente o migliorare con l'esperienza. In realtà, i sistemi AI disponibili oggi *non* acquisiscono automaticamente nuove capacità dopo l'uso e *non* ricordano gli specifici eventi che si sono verificati. Per aumentare le prestazioni, il sistema deve essere riaddestrato con dati migliori e più precisi in sessioni di apprendimento supervisionato. In genere, l'apprendimento non supervisionato richiede molti dati per generare cluster, quindi non è utilizzato nelle applicazioni di videosorveglianza. Oggi viene utilizzato soprattutto per analizzare grandi set di dati e riscontrare anomalie, ad esempio nelle transazioni finanziarie. Nella videosorveglianza, la maggior parte degli approcci pubblicizzati come "ad apprendimento automatico" si basano sull'analisi di dati statistici e non sull'effettivo riaddestramento dei modelli di deep learning.

L'esperienza dell'uomo batte ancora molte applicazioni analitiche basate sull'intelligenza artificiale e utilizzate per la sorveglianza. Soprattutto quelle che devono eseguire operazioni molto generiche e in cui è fondamentale capire il contesto. Se addestrata specificamente, un'applicazione basata sul machine learning può individuare con successo una "persona che corre"; ma al contrario di un essere umano, che può inserire i dati in un contesto, l'applicazione non sa se la persona sta correndo per prendere l'autobus o per sfuggire a un agente. Nonostante le promesse delle aziende che utilizzano l'intelligenza artificiale nelle applicazioni analitiche per la sorveglianza, l'applicazione non riesce neanche lontanamente a capire ciò che vede con la stessa perspicacia di un essere umano.

Per lo stesso motivo, le applicazioni analitiche basate sull'intelligenza artificiale possono attivare falsi allarmi o mancarli del tutto. In genere, questo può avvenire in ambienti complessi con molto movimento. Ma potrebbe anche verificarsi, ad esempio, se una persona trasporta un grosso oggetto, che nasconde i tratti umani all'applicazione e rende meno probabile una classificazione corretta.

Oggi, le analitiche basate su AI dovrebbero essere utilizzate, ad esempio, per determinare approssimativamente quanto sia rilevante un evento prima di avvisare un operatore umano per decidere

come intervenire. In questo modo, l'intelligenza artificiale viene utilizzata per fini di scalabilità e l'operatore umano è a disposizione per valutare i potenziali eventi.

8 Considerazioni per prestazioni analitiche ottimali

Per soddisfare le aspettative di qualità di un'applicazione analitica basata sull'intelligenza artificiale, si consiglia di studiare attentamente le precondizioni e le limitazioni note, che di solito vengono indicate nella documentazione dell'applicazione.

Ogni sistema di sorveglianza è unico e le prestazioni delle applicazioni devono essere valutate sito per sito. Se la qualità è inferiore alle attese, si consiglia vivamente di non indagare solo sull'applicazione. Tutte le indagini devono essere svolte a livello olistico perché le prestazioni di un'applicazione analitica dipendono da molti fattori, molti dei quali possono essere ottimizzati conoscendone gli effetti. Tra questi figurano ad esempio l'hardware della telecamera, la qualità video, le dinamiche nella scena e il livello di illuminazione, ma anche la configurazione, la posizione e la direzione della telecamera.

8.1 Utilizzabilità dell'immagine

Spesso si dice che la qualità d'immagine dipende dall'alta risoluzione e da un'elevata sensibilità alla luce della telecamera. Anche se l'importanza di questi fattori è indubbia, ne esistono sicuramente altri che hanno la stessa importanza, come l'*utilizzabilità* effettiva di un'immagine o un video. Ad esempio, anche il miglior flusso video della telecamera di sorveglianza più costosa può risultare inutile se la scena non è sufficientemente illuminata di notte, se la telecamera viene spostata intenzionalmente o se la connessione al sistema viene interrotta.

Il posizionamento della telecamera deve essere valutato attentamente prima dell'utilizzo. Affinché le analitiche video funzionino come previsto, la telecamera deve essere posizionata in modo da avere una visuale chiara e priva di ostacoli della scena desiderata.

L'utilizzabilità dell'immagine può anche dipendere dal caso d'uso. Un video che sembra di buona qualità agli occhi di un operatore potrebbe non avere una qualità ottimale per le prestazioni di un'applicazione videoanalitica. In effetti, molti metodi di elaborazione dell'immagine comunemente utilizzati per rendere il video più gradevole all'occhio non vengono consigliati quando si utilizzano le analitiche video. Tra questi figurano la riduzione del rumore, la tecnologia Wide Dynamic Range o gli algoritmi di esposizione automatica.

Oggi, le telecamere sono spesso dotate di illuminazione a infrarossi integrata, che consente il funzionamento nell'oscurità totale. Questa soluzione è efficace perché permette di utilizzare le telecamere in siti con un'illuminazione difficile, senza dover installare un'illuminazione supplementare. Tuttavia, in caso di pioggia o neve abbondante, si consiglia di non fare affidamento sulla luce proveniente dalla telecamera o da un punto molto vicino. Le gocce di pioggia e i fiocchi di neve potrebbero riflettere una quantità eccessiva di luce direttamente verso la telecamera, compromettendo il funzionamento delle analitiche. Con la luce ambientale si hanno invece più possibilità di ottenere risultati con le analitiche, anche in condizioni meteo difficili.

8.2 Rilevamento delle distanze

È difficile determinare la distanza di rilevamento massima per un'applicazione analitica basata su AI: un valore esatto in metri o piedi indicato su una scheda tecnica non può mai rappresentare tutta la verità. La qualità d'immagine, le caratteristiche della scena, le condizioni meteorologiche e le proprietà degli oggetti (come colore e luminosità) hanno effetti significativi sulla distanza di rilevamento. È evidente, ad esempio,

che un oggetto chiaro su sfondo scuro in una giornata di sole può essere rilevato a una distanza molto maggiore rispetto un oggetto scuro in una giornata di pioggia.

La distanza di rilevamento dipende anche dalla velocità degli oggetti. Per produrre risultati accurati, un'applicazione videoanalitica deve "vedere" l'oggetto per un periodo sufficientemente lungo. La durata di tale periodo dipende dalle capacità di elaborazione (framerate) della piattaforma: minori sono le capacità di elaborazione, più l'oggetto deve essere visibile per poter essere rilevato. Se il tempo di esposizione della telecamera non è idoneo alla velocità dell'oggetto, anche la sfocatura da movimento che appare sull'immagine può ridurre la precisione di rilevamento.

Gli oggetti veloci potrebbero essere mancati più facilmente se passano vicino alla telecamera. Una persona che corre lontano dalla telecamera, ad esempio, potrebbe essere rilevata con precisione, mentre una che corre vicino alla telecamera alla stessa velocità potrebbe entrare e uscire dal campo visivo così velocemente da non attivare alcun allarme.

Nelle analitiche basate sul rilevamento del movimento, sono altrettanto problematici gli oggetti che si muovono direttamente verso la telecamera o si allontanano da essa. Il rilevamento è particolarmente difficile per oggetti che si muovono lentamente, perché causano solo minime variazioni nell'immagine rispetto al movimento nella scena.

In genere, una telecamera con una risoluzione superiore non offre una distanza di rilevamento più alta. Le capacità di calcolo necessarie per eseguire un algoritmo di machine learning sono proporzionali alle dimensioni dei dati di input. Questo significa che le capacità di calcolo necessarie per analizzare interamente l'immagine di una telecamera 4K sono almeno quattro volte superiori a quelle necessarie per una telecamera 1080p. A causa delle limitate capacità di calcolo della telecamera, è molto comune eseguire le applicazioni basate su AI a una risoluzione minore rispetto a quella che può offrire la telecamera o il flusso.

8.3 Configurazione di allarmi e registrazioni

A causa dei filtri applicati a vari livelli, le analitiche di rilevamento oggetti generano pochissimi falsi allarmi. Tuttavia, queste analitiche funzionano come previsto solo se tutte le precondizioni indicate sono soddisfatte. Negli altri casi, potrebbero perdere eventi importanti.

Se non si è assolutamente certi di soddisfare sempre tutte le condizioni, si consiglia di adottare un approccio conservativo e di configurare il sistema in modo che la classificazione di oggetti specifici non sia l'unico fattore di attivazione di un allarme. Questo causerà più falsi allarmi, ma ridurrà il rischio di perdere eventi importanti. Se gli allarmi o le attivazioni vengono inviati direttamente a una centrale di monitoraggio, ogni falso allarme sarà molto costoso. Occorre quindi una classificazione degli oggetti affidabile per filtrare gli allarmi indesiderati. Tuttavia, la soluzione di registrazione può e deve comunque essere configurata in modo da non affidarsi solo alla classificazione degli oggetti: nel caso di un allarme mancato, si osserverà la registrazione e si capirà il motivo per cui non si è attivato. Quindi, si procederà a migliorare il sistema e la configurazione a livello complessivo.

Se la classificazione degli oggetti viene eseguita sul server durante la ricerca di un evento, si consiglia di impostare il sistema sulla registrazione continua e di non filtrare affatto la registrazione iniziale. La registrazione continua occupa molto spazio, ma il consumo viene in qualche modo compensato dai moderni algoritmi di compressione come Zipstream.

8.4 Manutenzione

Un sistema di sorveglianza deve essere sottoposto a regolare manutenzione. Le ispezioni fisiche, e non solo la visione del video tramite l'interfaccia del software di gestione, sono consigliate per identificare e rimuovere tutto ciò che potrebbe disturbare od ostacolare il campo visivo. Questo è importante nei sistemi standard e dedicati solo alla registrazione, ma è ancora più importante quando si utilizzano le analitiche.

Nel caso di un'applicazione di base per il rilevamento di movimento nel video, un comune ostacolo come una ragnatela smossa dal vento potrebbe far aumentare il numero di allarmi, consumando più spazio di archiviazione del necessario. Con l'analisi degli oggetti, la ragnatela creerebbe sostanzialmente una zona di esclusione nell'area di rilevamento. I fili oscurerebbero gli oggetti, riducendo enormemente la possibilità di rilevarli e classificarli.



Le ragnatele possono disturbare il campo visivo di una telecamera di sorveglianza.

Nelle ore diurne, lo sporco sul vetro o sulla cupola della telecamera non dovrebbe causare problemi. In condizioni di scarsa illuminazione, però, una cupola sporca che viene colpita lateralmente da una luce, proveniente ad esempio dai fari di un'auto, può causare riflessi imprevisti e ridurre la precisione del rilevamento.

La manutenzione della scena ha la stessa importanza di quella della telecamera. Durante il ciclo di vita di una telecamera, la scena sorvegliata può cambiare molto. Un semplice confronto con un'immagine "prima" e "dopo" rivelerà i potenziali problemi. Com'era la scena quando la telecamera è stata utilizzata per la prima volta? Com'è oggi? È necessario regolare la zona di rilevamento? Occorre regolare il campo visivo della telecamera, oppure la telecamera deve essere spostata in un punto diverso?

9 Privacy e integrità personale

Per lavorare nella sicurezza e nella sorveglianza, è necessario trovare un equilibrio tra i diritti alla privacy e all'integrità personale e la volontà di aumentare la sicurezza prevenendo i reati o consentendo le indagini forensi. Per ogni specifico sistema e caso d'uso, questo richiede attente considerazioni etiche nonché la comprensione e l'applicazione delle leggi locali. La soluzione deve anche garantire la cybersecurity e prevenire gli accessi non intenzionali al materiale video. Allo stesso tempo, le analitiche edge-based e la generazione di metadati per finalità statistiche possono tutelare maggiormente la privacy trasmettendo solo dati anonimi per l'elaborazione successiva.

Con la maggiore applicazione delle analitiche automatizzate nei sistemi di sorveglianza, è necessario tenere in considerazione alcuni nuovi aspetti. Poiché con le applicazioni analitiche si rischiano falsi rilevamenti, è importante che il processo decisionale coinvolga un operatore o un utente esperto: si tratta dell'approccio "human in the loop", o HITL. Inoltre, è importante capire che la decisione dell'uomo può essere condizionata da come viene generato e recapitato l'allarme. Senza un opportuno addestramento e una conoscenza delle funzionalità della soluzione analitica, è possibile giungere a conclusioni errate.

Ulteriori preoccupazioni possono riguardare il modo in cui vengono sviluppati gli algoritmi di deep learning: in alcuni casi d'uso, questo richiede un'applicazione cauta della tecnologia. La qualità degli algoritmi è fondamentale legata ai set di dati, ovvero alle immagini e ai video utilizzati per addestrare l'algoritmo. I test hanno dimostrato che, se non si seleziona accuratamente il materiale, alcuni sistemi AI possono presentare deviazioni nel rilevamento per etnia e genere. Questo ha scatenato un dibattito e dato origine a limiti di legge e attività di vario tipo, per fare in modo che tali aspetti vengano sempre valutati durante lo sviluppo dei sistemi.

Poiché l'intelligenza artificiale è sempre più utilizzata nella videosorveglianza, è importante bilanciare i vantaggi in termini di efficienza operativa e nei nuovi casi d'uso valutando attentamente dove e quando applicare la tecnologia.

10 Appendice

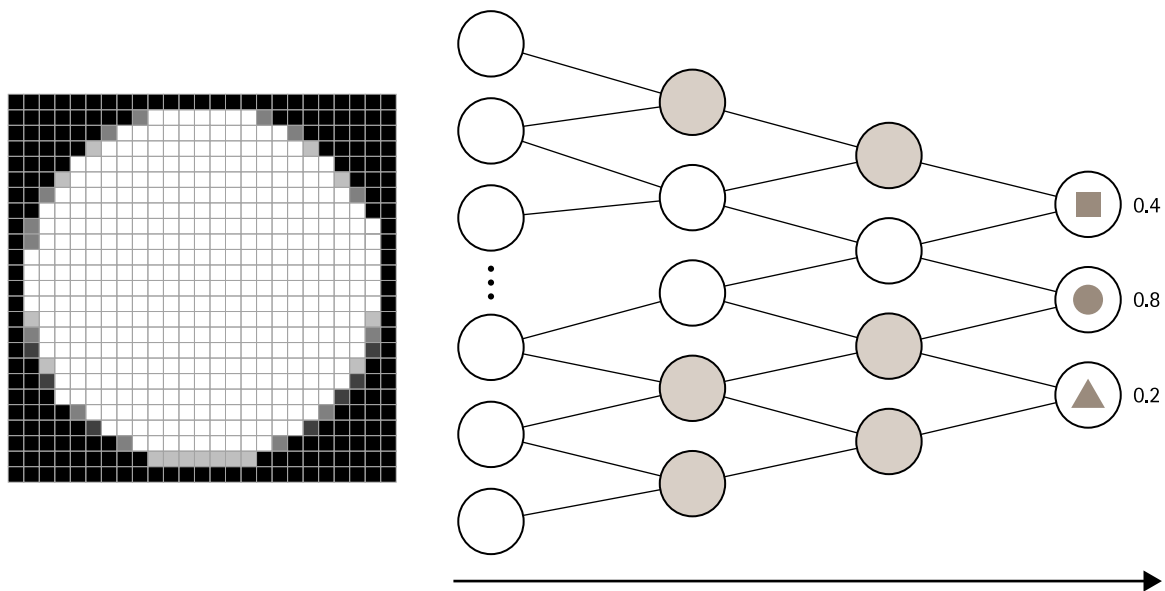
Questa appendice illustra le nozioni di base sulle reti neurali artificiali, che sono alla base del deep learning.

10.1 Reti neurali

Le reti neurali sono algoritmi utilizzati per riconoscere relazioni presenti nei set di dati con un processo piuttosto simile a quello utilizzato dal cervello. Una rete neurale è formata da una gerarchia di più layer di nodi o neuroni interconnessi. Dal layer di input, le informazioni attraversano le connessioni e la rete fino al layer di output.

Affinché le reti neurali funzionino, il presupposto è che un campione di dati di input possa essere ridotto a un insieme finito di caratteristiche, creando una rappresentazione valida dei dati in ingresso. Le caratteristiche possono quindi essere combinate e aiutano a classificare i dati di input, descrivendo ad esempio i contenuti di un'immagine.

La seguente figura mostra un esempio di rete neurale utilizzata per identificare la classe di appartenenza di un'immagine di input. Ogni pixel dell'immagine è rappresentato da un nodo di input. Tutti i nodi di input vengono abbinati ai nodi del primo layer. In questo modo vengono prodotti valori di output, che passano al secondo layer come valori di input, e così via. In ogni layer entrano in gioco anche funzioni di ponderazione, valori di scostamento e funzioni di attivazione.



Esempio di immagine di input (a sinistra) e di rete neurale (a destra). Raggiunto il layer di output, la rete definisce le probabilità che l'oggetto appartenga a ogni categoria (quadrato, cerchio o triangolo). La categoria con il valore di probabilità più alto è la forma più probabile dell'immagine di input.

Questo processo è detto *propagazione in avanti*. In caso di non corrispondenza del risultato ottenuto con la propagazione in avanti, i parametri della rete vengono modificati leggermente attraverso una *propagazione all'indietro*. Durante questo processo di addestramento iterativo, le prestazioni della rete aumentano gradualmente.

Dopo l'implementazione, in genere una rete neurale non ha memoria dei passaggi in avanti precedenti. Questo significa che non migliora nel tempo e che può rilevare solo i tipi di oggetti, o assolvere i compiti, per i quali è stata addestrata.

10.2 Reti neurali convoluzionali (CNN)

Le *reti neurali convoluzionali* (CNN) sono un sottotipo di reti neurali. Si sono dimostrate particolarmente adatte ai task di visione artificiale e sono al centro dei rapidi progressi nel deep learning. Nel caso della visione artificiale, la rete viene addestrata a cercare automaticamente tratti di immagine distintivi, come bordi, angoli e differenze di colore, identificando di fatto le forme degli oggetti.

L'operazione matematica principale per svolgere questo compito è detta *convoluzione*. Si tratta di un'operazione molto efficiente perché l'output di ogni singolo nodo dipende solo da un insieme limitato e circostante di dati in ingresso (prodotto dal layer precedente), senza utilizzare l'intero volume di dati di input. In altre parole: in una CNN, ciascun nodo non è collegato a ogni nodo del layer precedente, ma solo a un piccolo sottoinsieme. Le convoluzioni vengono completate da altre operazioni che riducono le dimensioni dei dati mantenendo le informazioni più utili. Come in una rete neurale artificiale standard, i dati diventano sempre più astratti man mano che circolano nella rete.

Durante la fase di addestramento, la CNN impara ad applicare al meglio i layer. Ovvero, come le convoluzioni devono combinare le caratteristiche del layer precedente affinché l'output della rete concordi il più possibile con le annotazioni dei dati di addestramento. Durante l'inferenza, la rete neurale convoluzionale addestrata applicherà quindi i layer delle convoluzioni risultanti dall'addestramento.

Informazioni su Axis Communications

Axis consente un mondo più intelligente e più sicuro creando soluzioni per migliorare la sicurezza e le prestazioni aziendali. Come società di tecnologie di rete e leader nel settore, Axis offre soluzioni nella videosorveglianza, controllo degli accessi, interfono e sistemi audio. Queste sono ottimizzate da applicazioni di analisi intelligente e supportate da formazione di alta qualità.

Axis ha circa 4.000 impiegati dedicati in più di 50 paesi e collabora con partner di tecnologia e integrazione di sistema in tutto il mondo per offrire soluzioni di clienti. Fondata nel 1984, Axis è con sede a Lund, in Svezia