

REGIONE TOSCANA

PR FESR 2021-2027 OP1 OS1 Azione 1.1.4 “Ricerca e sviluppo per le imprese anche in raggruppamento con organismi di ricerca”

Obiettivo Operativo I: Studio dello Stato dell’Arte

Parte III: Sensoristica Applicata per il Monitoraggio Stradale e Metodi di Intelligenza Artificiale

Sommario

I.	Introduzione.....	2
II.	Sensori a bordo dei veicoli per il monitoraggio stradale	3
	Telecamere RGB, multispettrali, termografiche.....	3
	Telecamere lineari e matriciali ad alte prestazioni	5
	Sistemi LiDAR e radar per mappature 3D	7
	Tecnologie combinate (sensor fusion)	9
III.	Approcci basati su Intelligenza Artificiale	10
	Tecniche di rilevamento automatizzato	11
	Modelli di deep learning per il rilevamento dei difetti del manto stradale	13
	Modelli di deep learning per il rilevamento della segnaletica stradale.....	15
	Integrazione in pipeline operative per il monitoraggio continuo .. Errore. Il segnalibro non è definito.	
	Accuratezza e limiti delle tecnologie	16
IV.	Conclusioni	19
	Bibliografia	19

I. Introduzione

Il presente studio si propone di analizzare lo stato dell'arte relativo alle tecnologie sensoristiche e ai metodi di intelligenza artificiale (IA) applicati al monitoraggio delle infrastrutture stradali. L'attenzione è rivolta, in particolare, alle tecnologie di bordo installate su veicoli strumentati, con l'obiettivo di rilevare in modo automatizzato e continuo lo stato del manto stradale e della segnaletica stradale orizzontale e verticale.

La necessità di sistemi di monitoraggio efficaci e a basso costo nasce dalla crescente importanza attribuita alla sicurezza stradale e alla manutenzione predittiva. I metodi tradizionali, basati su ispezioni visive e rilievi con strumentazione dedicata, presentano elevati costi economici, tempi di esecuzione lunghi e limitata scalabilità. In risposta a tali limiti, negli ultimi anni sono stati sviluppati numerosi sistemi innovativi che sfruttano tecnologie come telecamere RGB, sensori termici e LiDAR, combinate con algoritmi di machine learning e deep learning.

A livello europeo e internazionale, diversi progetti di ricerca hanno mostrato la fattibilità e l'efficacia di approcci basati su sensori mobili:

- INTRO ha esplorato l'integrazione di sensori fissi e mobili per migliorare il monitoraggio delle infrastrutture [1];
- TRIMM ha integrato tecnologie di rilevamento avanzato e analisi automatizzata dei dati [2];
- SENSOVO ha dimostrato l'uso di veicoli con telecamere ToF e dati di velocità per rilevare ammaloramenti [3];
- Crowd4Roads ha impiegato sensori accelerometrici degli smartphone per monitorare la regolarità del manto stradale [4];
- Mobisense ha utilizzato microfoni e accelerometri montati nei veicoli per la rilevazione di vibrazioni e suoni legati al deterioramento stradale [5];
- LIRA ha adottato un approccio "pervasivo" usando la sensoristica già presente nei veicoli [6].

Anche il mercato ha risposto con soluzioni commerciali, come:

- NIRA Dynamics, che integra sensori di bordo per la percezione delle condizioni stradali [7];
- Roadroid, che utilizza smartphone per valutare l'Indice di Rugosità Internazionale (IRI) mediante vibrazioni e GPS [8].

L'avvento dell'intelligenza artificiale ha ulteriormente potenziato queste tecnologie, rendendo possibile la classificazione automatica dei difetti e la mappatura accurata degli elementi stradali. Il presente documento rappresenta dunque una rassegna tecnica e scientifica dei principali approcci, utile a guidare la selezione delle tecnologie e degli algoritmi da sviluppare nel progetto.

II. Sensori a bordo dei veicoli per il monitoraggio stradale

Telecamere RGB, multispettrali, termografiche

Le **telecamere RGB** rappresentano attualmente la soluzione più diffusa e consolidata per il rilievo visivo in continuo delle condizioni del manto stradale e della segnaletica orizzontale e verticale. Montate su veicoli strumentati in movimento, consentono l'acquisizione di flussi video ad alta risoluzione (superiore a 4K) a velocità di marcia, con costi di esercizio contenuti. Grazie alla loro elevata risoluzione spaziale e alla larga diffusione commerciale, permettono campagne di rilievo su larga scala, rendendole particolarmente adatte al monitoraggio di rete. Numerosi progetti europei e internazionali, come TRIMM [2] e LIRA [6], ne hanno dimostrato l'efficacia, anche in combinazione con sistemi di posizionamento satellitare (GPS). Già nel 2011, studi pionieristici evidenziavano l'utilizzo di tecnologie di imaging automatizzato, grazie a telecamere montate su veicoli, per il rilevamento sistematico delle distorsioni del manto stradale e la progressiva sostituzione delle ispezioni manuali [10]



Figura 1: Esempio telecamera RGB dotata anche di stabilizzatore integrato - Fonte: <https://gopro.com/it/it/shop/cameras>

Le **telecamere multispettrali e iperspettrali**, operando su bande dello spettro elettromagnetico al di fuori del visibile (in particolare NIR-SWIR), consentono una discriminazione avanzata dei materiali presenti sulla superficie stradale, come asfalto, vernici, aggregati o contaminanti. Questi sistemi permettono di rilevare micro-alterazioni cromatiche e degradi superficiali non visibili all'occhio umano, rivelandosi particolarmente efficaci nell'individuazione di segnaletica deteriorata o alterazioni del manto stradale. Indagini di imaging spectrometry hanno evidenziato correlazioni significative tra firme spettrali e degrado superficiale [10], mentre applicazioni avanzate come quelle condotte con il sensore AVIRIS hanno dimostrato il potenziale della mappatura iperspettrale su tratte stradali estese [11]. Sebbene queste tecnologie siano attualmente più diffuse in ambito agricolo e ambientale, il loro impiego nel monitoraggio viario è in rapida espansione.

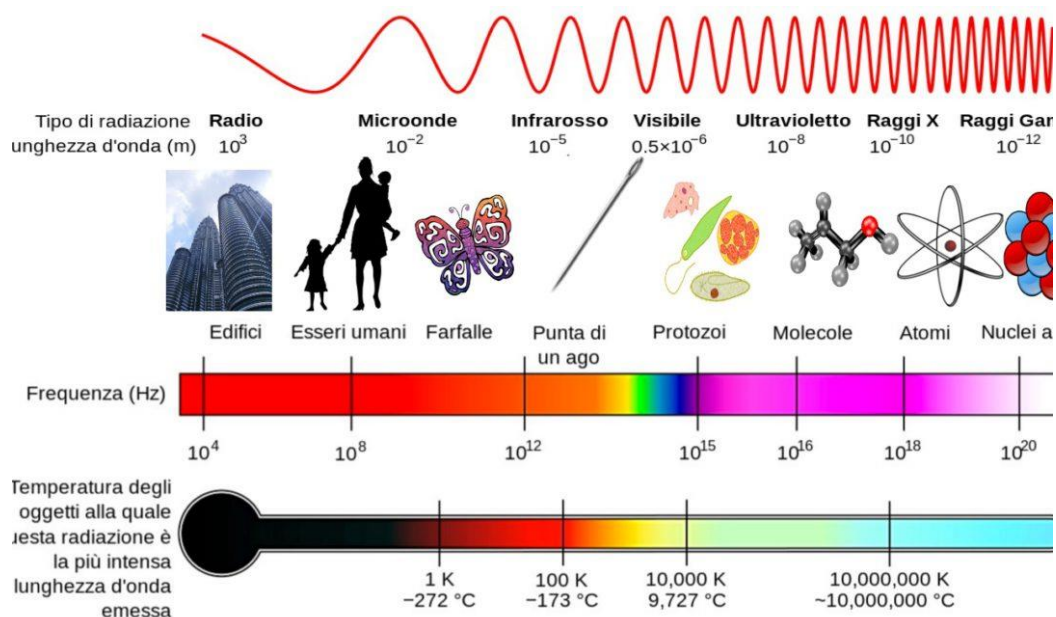


Figura 2: Descrizione spettro visivo - Fonte: <https://www.skydrone360.it/cose-camera-multispettrale/>

Le **telecamere termografiche** operano nel campo dell'infrarosso termico (8–14 μm) e misurano la radiazione emessa dagli oggetti in funzione della loro temperatura superficiale. Nel monitoraggio stradale, queste tecnologie sfruttano le differenze di emissività per rilevare difetti non visibili nel campo del visibile, come vuoti sotto l'asfalto, infiltrazioni d'acqua, disomogeneità nei materiali di rivestimento o crepe sottili. Studi sperimentali ne hanno dimostrato l'efficacia sia in fase di posa, per il controllo della qualità costruttiva, sia nel rilevamento precoce di ammaloramenti [12, 13]. Inoltre, la termografia si è rivelata

particolarmente utile anche nella valutazione dello stato della segnaletica orizzontale, grazie alla marcata differenza di emissività tra i materiali riflettenti e l'asfalto circostante.

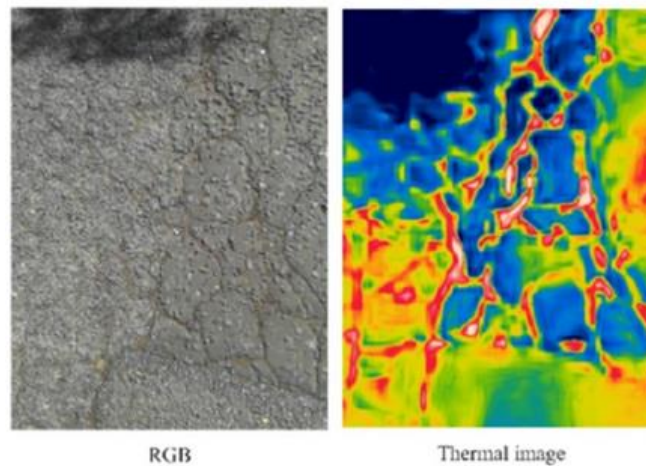


Figura 3: Esempi di riprese dell'asfalto acquisite con una telecamera RGB (a sinistra) e una termocamera (a destra) [12].

Telecamere lineari e matriciali ad alte prestazioni

Le **telecamere lineari**, o line-scan, sono dispositivi dotati di un'unica linea di sensori che acquisisce l'immagine riga per riga in sincronia con lo spostamento del veicolo. Questa tecnologia consente di ottenere immagini ad altissima risoluzione e uniformità luminosa, risultando ideale per la scansione continua della carreggiata. Alcuni modelli stereo raggiungono risoluzioni fino a 16.000 pixel per linea e frequenze superiori a 30 kHz, permettendo rilievi continui a velocità operative superiori a 80–100 km/h con dettaglio sub-millimetrico [14, 15]. Per garantire la precisione spaziale, l'acquisizione deve essere sincronizzata con sistemi di encoder, evitando distorsioni. L'integrazione con illuminazione laser strutturata riduce l'influenza di ombre e luce solare, migliorando la rilevazione di microfessure, scolorimenti e discontinuità superficiali. Un caso studio condotto in ambito autostradale ha dimostrato che una telecamera line-scan CCD da 3k pixel, montata posteriormente al veicolo, è in grado di rilevare crepe di ampiezza pari a 1 mm a velocità di marcia standard [16].

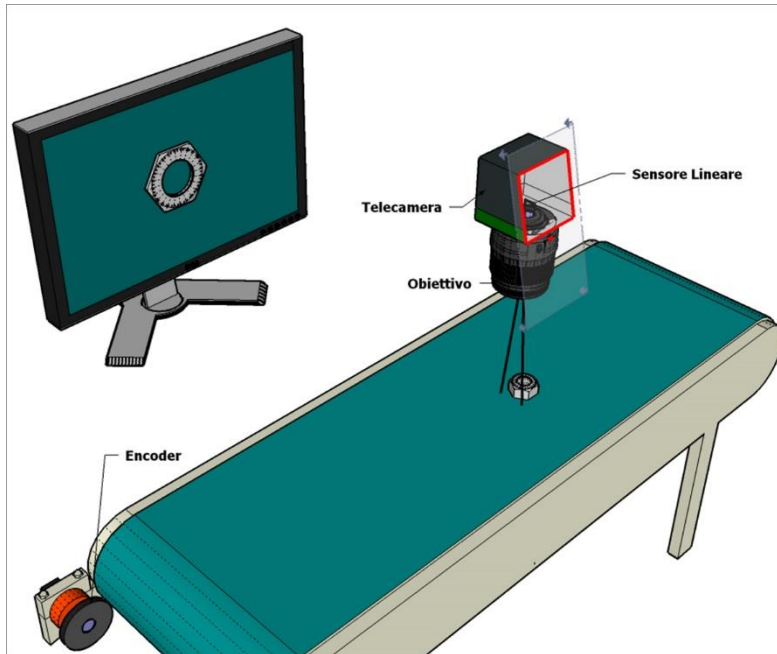


Figura 4: Schema di funzionamento telecamere lineari - Fonte:
<https://www.specialvideo.it/tecnologie/telecamera-lineare/>

Le **telecamere matriciali**, a differenza di quelle lineari, acquisiscono l'intero frame in una singola esposizione, risultando particolarmente adatte alle applicazioni che richiedono una visione complessiva e simultanea della scena. Questa caratteristica le rende la soluzione preferita per il riconoscimento della segnaletica verticale, dove è essenziale catturare rapidamente elementi distribuiti su più livelli della carreggiata o in movimento.

Le versioni più avanzate sono dotate di sensori CMOS ad alta risoluzione (fino a 8K) con otturatore globale (global shutter), in grado di eliminare distorsioni da movimento (rolling shutter effect) e garantire una maggiore fedeltà nella rappresentazione della scena.

Inoltre, l'impiego di ottiche grandangolari consente di ampliare il campo visivo, riducendo il numero di dispositivi necessari per la copertura dell'ambiente. L'integrazione di tecnologie HDR (High Dynamic Range) permette una gestione efficace delle condizioni di luce estrema, evitando fenomeni di saturazione o sottoesposizione. Questo si traduce in una maggiore robustezza del sistema, anche in presenza di forti contrasti di luminosità (es. controluce o ombre marcate), e consente il rilevamento di segnali stradali parzialmente occlusi, scoloriti o deteriorati.

Sistemi LiDAR e radar per mappature 3D

I **sensori LiDAR** (Light Detection and Ranging) sono dispositivi ottici attivi in grado di generare nuvole di punti tridimensionali (3D) dell'ambiente circostante, mediante impulsi laser riflessi. Montati su veicoli, questi sensori consentono la mappatura dettagliata del profilo altimetrico della pavimentazione e degli elementi verticali (come guardrail, segnaletica o marciapiedi), con una densità di punti superiore a 200 pts/m². Grazie alla loro elevata risoluzione spaziale e all'insensibilità alla luce ambientale, i LiDAR operano efficacemente anche di notte o in condizioni di scarsa visibilità, rendendoli particolarmente robusti per applicazioni stradali. L'integrazione con immagini RGB permette di arricchire i dataset con informazioni radiometriche e cromatiche, migliorando l'interpretabilità delle scene. Campagne recenti su reti provinciali hanno dimostrato la possibilità di estrarre automaticamente deformazioni del piano viabile, come ormaie e avvallamenti, con un errore inferiore a 2 mm rispetto ai rilievi profilometrici tradizionali [17]. Inoltre, la riflettanza della nuvola di punti è stata utilizzata per valutare la retro-riflessione della segnaletica orizzontale, offrendo un parametro quantitativo per l'analisi dello stato della vernice stradale [18].

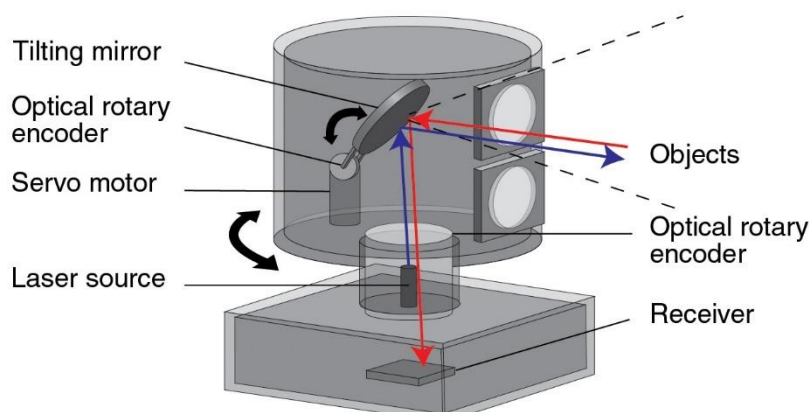


Figura 5: Esempio di funzionamento sensore LiDAR - Fonte:
<https://www.renishaw.com/it/optical-encoders-and-lidar-scanning--39244>

I **radar** a onde millimetriche (76–81 GHz), analogamente ai LiDAR, permettono il rilevamento di superfici e ostacoli tramite la riflessione di onde elettromagnetiche, ma con il vantaggio di essere meno sensibili a nebbia, pioggia, polvere e abbagliamento solare. Sebbene presentino una risoluzione spaziale inferiore (attualmente < 4 cm), i recenti sviluppi tecnologici hanno migliorato significativamente la miniaturizzazione dei moduli e la precisione [19]. Tali sensori risultano quindi complementari a LiDAR e telecamere, in particolare in scenari meteo avversi

o in contesti urbani complessi. Applicazioni documentate da Texas Instruments ne descrivono l'uso per il monitoraggio di flussi veicolari, segnaletica verticale e ostacoli a medio raggio [20].

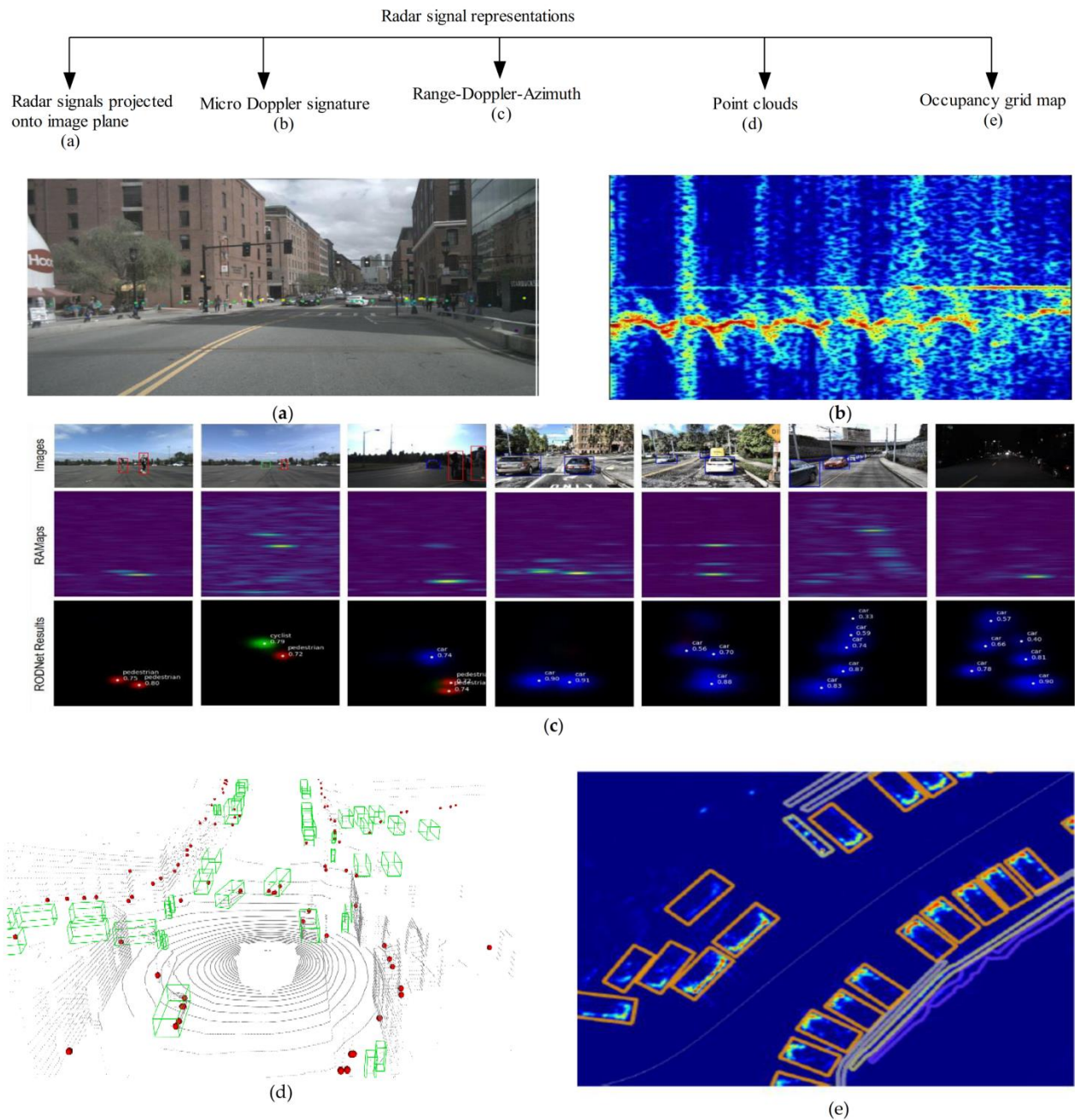


Figura 6: Esempio utilizzo Radar a onde millimetriche – Fonte [25]

Tecnologie combinate (sensor fusion)

Negli ultimi anni, davanti ai limiti evidenti dei sensori singoli, in particolare quando le superfici sono degradate, la segnaletica è sbiadita o le condizioni ambientali mutano rapidamente, ha preso piede l'approccio della sensor fusion, ossia l'integrazione multimodale di telecamere RGB, termocamere a infrarossi, LiDAR, ma anche GPS e unità inerziali. La logica è semplice: ciascun dispositivo compensa le debolezze degli altri. Le sole immagini nel visibile, ad esempio, offrono un'elevata definizione cromatica ma diventano poco leggibili con abbagliamento, ombre o illuminazione notturna; al contrario le termocamere individuano differenze di emissività e rivelano crepe o linee consumate che l'occhio non distingue, pur soffrendo di risoluzione spaziale inferiore. Il LiDAR, infine, fornisce una nuvola di punti tridimensionale precisa e non dipende dalla luce solare, permettendo di localizzare con errore sub-decimetrico la posizione dei difetti o dei segnali rilevati. Studi sperimentali mostrano che già la sola concatenazione di canali RGB e termici può far crescere l'F-score sul rilevamento di crepe sottili di circa 13 punti percentuali rispetto alla sola camera RGB, mentre l'inserimento delle quote LiDAR in una rete CNN-Transformer spinge la IoU sulla segnaletica orizzontale a 0,88, cioè otto punti in più rispetto alla baseline visibile [23][24].

Il guadagno dipende da come si realizza la fusione: alcune applicazioni uniscono i dati grezzi prima dell'estrazione delle caratteristiche, altre fondono le feature prodotte da reti parallele, altre ancora aggregano soltanto le decisioni finali dei modelli; in quest'ultimo caso, su scenari meteo avversi si è registrata una riduzione dei falsi negativi vicina al 40 % rispetto a un sistema monocamera testato nel progetto INTRO [1]. Qualunque sia la strategia, l'affidabilità passa da una calibrazione extrinseca accurata tra sensori – board retro-riflettenti o pattern termo-stampati consentono di contenere l'errore angolare sotto 0,5 ° e quello lineare sotto 5 mm – e da una sincronizzazione millimetrica via trigger PPS/GPS che mantiene la latenza al di sotto dei 10 ms, soglia oltre la quale comparirebbero sfasamenti superiori a 20 pixel a 60 km/h [30]. In pratica, un set-up tipico monta telecamere lineari ad alta frequenza posteriori per la scansione continua del manto, una camera matriciale frontale ad alta risoluzione per i cartelli, e un LiDAR a 32 canali per la ricostruzione 3-D, ottenendo così una fotografia completa e coerente dell'infrastruttura stradale; i risultati più recenti indicano incrementi d'accuratezza e robustezza che difficilmente sarebbero raggiungibili con un singolo sensore [10][17][19].

Le prove su strada confermano i benefici dell'approccio multimodale:

- un modello RGB-LiDAR a voxel first fusion ha migliorato la mAP nella classificazione dei difetti del 12% rispetto a RGB-only [25];
- l'integrazione di immagini aeree RGB e LiDAR ha ridotto di 9% l'errore di omissione sui cartelli stradali di piccole dimensioni [26];
- la fusione UAV-RGB con LiDAR a bassa quota ha incrementato l'IoU di segmentazione crepe del 10% anche su pavimentazioni sigillate [27];
- un modello R-CRF basato su sequenze RGB + LiDAR ha alzato l'accuratezza di rilevamento carreggiata del 9% in scenari notturni [28].

III. Approcci basati su Intelligenza Artificiale

Prima dell'avvento delle tecniche di machine learning, l'analisi automatica del manto stradale e della segnaletica orizzontale si basava principalmente su metodi tradizionali di elaborazione delle immagini, come l'analisi delle caratteristiche geometriche e l'uso di algoritmi di segmentazione basati su soglia [31]. Questi approcci si concentravano sull'identificazione dei bordi degli ammaloramenti e delle linee e dei simboli della segnaletica orizzontale attraverso la rilevazione dei contrasti di colore e luminosità tra il manto stradale e gli ammaloramenti o tra il manto stradale e la segnaletica stessa. Metodi come il rilevamento dei bordi, ad esempio l'algoritmo di Canny [32], e la segmentazione per regioni di colore erano utilizzati per separare le linee orizzontali dallo sfondo, ma spesso risultavano limitati dalla presenza di rumori, usura della pavimentazione e condizioni di illuminazione variabili. Sebbene efficaci in contesti semplici, questi approcci tradizionali non riuscivano a garantire una robustezza sufficiente per affrontare scenari complessi e dinamici come quelli presenti nelle condizioni reali delle strade ([34],[33]). Negli ultimi anni i sistemi automatizzati per la valutazione del manto stradale hanno subito importanti sviluppi tecnologici e la ricerca si è concentrata prevalentemente sul compito cosiddetto di “crack extraction”, ovvero identificazione e isolamento di crepe sul manto stradale. Tuttavia, i metodi esistenti presentano capacità limitate nel riconoscere tutte le diverse condizioni stradali che possono presentare anche un'elevata complessità. I metodi basati su immagini digitali sono stati ampiamente utilizzati per il rilevamento e la segmentazione delle crepe, sfruttando le differenze nei livelli di grigi tra i pixel delle crepe e lo sfondo [35]. Tuttavia, condizioni di illuminazione, macchie di olio sull'asfalto e segnali stradali rendono complessa la classificazione e la segmentazione delle crepe. Una delle strategie

ideate per affrontare queste limitazioni consiste nell'utilizzare un sistema multi-sensore, includendo accelerometri [36], sensori a infrarossi [37], telecamere multivisione [38] e scansione laser 3D [39], che possono fornire informazioni aggiuntive a quelle già fornite dalle immagini ottiche del manto stradale. Rispetto a tecniche come la scansione laser o le nuvole di punti 3D, l'elaborazione delle immagini offre vantaggi come velocità, basso costo e robustezza, anche se, come accennato in precedenza, da sola potrebbe non garantire un riconoscimento completo in contesti complessi.

Negli ultimi anni, si è registrato un grande sviluppo di tecniche di analisi basate sull'intelligenza artificiale, capaci di estrarre informazioni dettagliate da una mole adeguata di dati, fornendo un decisivo supporto e talvolta anche un'alternativa ai classici metodi di analisi, soprattutto in contesti complessi. In quest'ottica, l'utilizzo del machine learning per l'elaborazione delle immagini si è largamente diffuso. Nel prosieguo di questa sezione, verranno introdotti gli ultimi lavori disponibili in letteratura riguardo l'utilizzo di tecniche di machine learning per l'analisi delle immagini. Tale branca dell'intelligenza artificiale può essere suddivisa in machine learning tradizionale e deep learning. Di seguito verranno presentati esempi di algoritmi appartenenti a entrambe le categorie sottolineando differenze, vantaggi e svantaggi.

Tecniche di rilevamento automatizzato

Oggi il processo di rilevamento dei segnali stradali e ammaloramenti segue, a grandi linee, tre passi essenziali:

1. **Acquisizione dei dati.** I veicoli strumentati raccolgono immagini e, quando necessario, informazioni termiche o tridimensionali. La qualità del dato grezzo è la base di ogni analisi successiva.
2. **Elaborazione intelligente.** Algoritmi di apprendimento automatico (dalle prime tecniche basate su descrittori manuali fino alle più recenti reti neurali) analizzano il flusso visivo per marcare crepe, buche, linee di demarcazione e segnali verticali. In questo stadio l'IA sostituisce di fatto la valutazione umana, garantendo rapidità e coerenza.

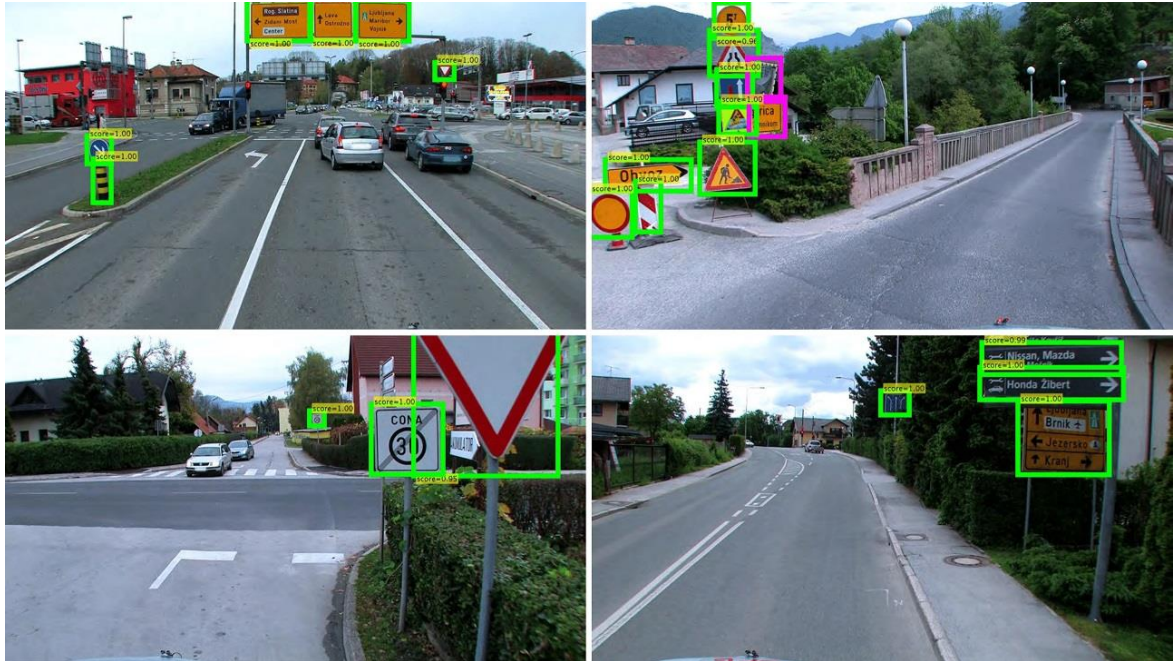


Figura 7: Esempi di rilevamento automatico della segnaletica stradale verticale- Fonte: <https://www.vicos.si/research/traffic-sign-detection/>

La Figura 7 mostra diversi scenari urbani e suburbani in cui un sistema di visione artificiale rileva e classifica automaticamente la segnaletica verticale lungo la carreggiata. I cartelli stradali sono evidenziati da riquadri colorati con l'indicazione del tipo di segnale e un punteggio di confidenza (score), a testimonianza della capacità del sistema di riconoscere correttamente segnali direzionali, di pericolo, regolamentari e informativi.

Le scene dimostrano l'efficacia del riconoscimento anche in condizioni di occlusione parziale, variazioni di illuminazione e complessità del contesto visivo.

3. **Restituzione dei risultati.** Le informazioni estratte vengono trasformate in indicatori comprensibili per tecnici e amministratori: mappe di criticità, elenchi di interventi prioritari, trend di degrado.



Figura 8: Esempio di mappa con indicatori di degrado stradale

La figura 8 mostra l'esempio di una mappa urbana in cui i tratti viari sono evidenziati con colori differenti (verde, giallo, rosso) per rappresentare diversi livelli di degrado rilevati lungo la rete stradale. La visualizzazione consente un'immediata interpretazione dello stato di conservazione delle infrastrutture.

Queste tecniche, pur con margini di miglioramento in caso di luce difficile, segnaletica usurata o contesti geografici molto diversi dal set d'addestramento, consentono già oggi di programmare la manutenzione in modo più tempestivo ed efficiente, riducendo costi e rischi per l'utenza stradale.

Modelli di deep learning per il rilevamento dei difetti del manto stradale

A causa degli avanzamenti apportati dall'uso delle CNN nel campo generale dell'analisi e della classificazione delle immagini, esse sono state sempre più usate, in tempi recenti, nella valutazione dello stato delle pavimentazioni stradali basata su segnali fotografici o frame ricavati da riprese video [40]. A causa del funzionamento a blocchi delle CNN, è possibile ricavare dall'immagine di partenza informazioni inerenti a diverse proprietà utili nell'identificazione delle fessurazioni o degli ammaloramenti. L'informazione è organizzata in insiemi di features su tre diversi livelli [41]: le features di livello basso, estratte dai layer iniziali delle CNN, codificano principalmente informazioni relative a regolarità o bordi presenti nelle immagini; le features di medio livello, estratte dai layer intermedi, forniscono informazioni relative a colori e forme; le features di alto livello, ricavate mediante i layer profondi, permettono di individuare informazioni complesse (come, ad esempio, oggetti completi).

Risulta necessario sottolineare che i metodi di deep learning richiedono, al fine di fornire risultati affidabili e di buona qualità, una grande mole di dati di allenamento: sono quindi spesso necessarie lunghe e laboriose operazioni di acquisizione e catalogazione dei dati. Spesso è possibile, tuttavia, incrementare successivamente il numero dei dati mediante l'uso di rotazioni, riflessioni o altre trasformazioni geometriche delle immagini di partenza. Tale processo è detto data augmentation [42] e permette di creare agevolmente dataset di grandi dimensioni (in uno studio [43] relativo alla classificazione del cracking stradale, il numero di immagini presenti in un dataset originario di dimensioni contenute è stato aumentato di un fattore 10).

Le reti neurali convoluzionali, sebbene rappresentino un potente strumento di analisi, dipendono da un grande numero di parametri fissati dall'operatore, quali la scelta del numero di convoluzioni, la presenza di layer di dropout, la scelta dell'algoritmo di ottimizzazione, lo sviluppo di una funzione di costo da minimizzare, la dimensione dei kernel convoluzionali.

Convolution Neural Network (CNN)

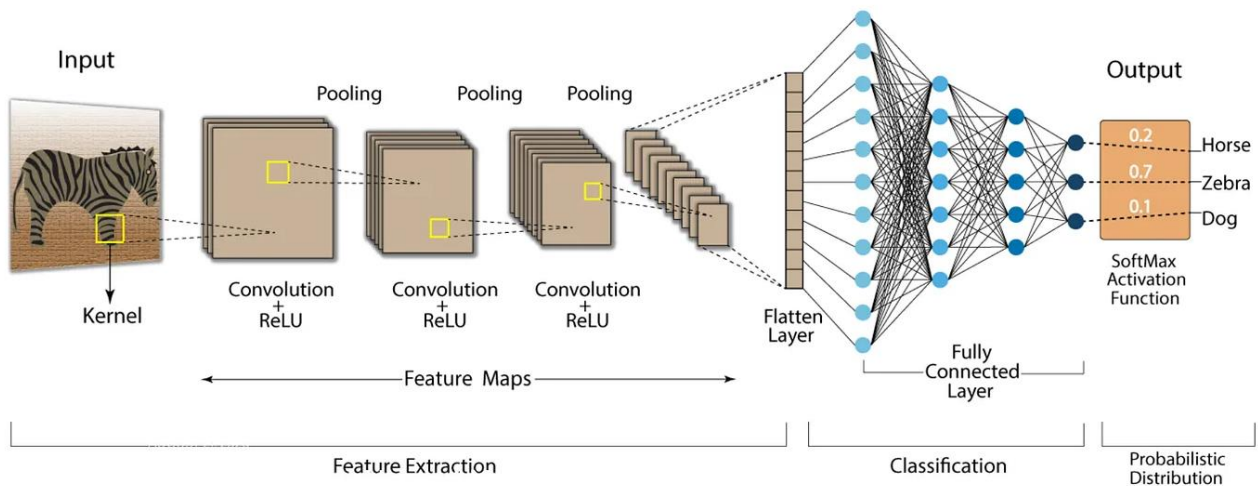


Figura 9: Esempio dell'architettura di una rete Neurale Convoluzionale (CNN) - Fonte: <https://easyodm.tech/object-detection-with-cnns>

Le CNN sono state utilizzate anche nel riconoscimento delle imperfezioni del manto stradale [45] (classificazione e localizzazione delle asperità all'interno dell'immagine). In particolare, esistono due tipi di rilevatori: one shot detector e two shot detector. I primi effettuano il rilevamento degli oggetti con un singolo passaggio, alcune architetture di questa tipologia sono i rilevatori della famiglia YOLO o i rilevatori Single-shot multibox (SSD). Nei two stage detector invece si utilizza un sistema basato sulla divisione delle regioni, un esempio sono le RCNN.

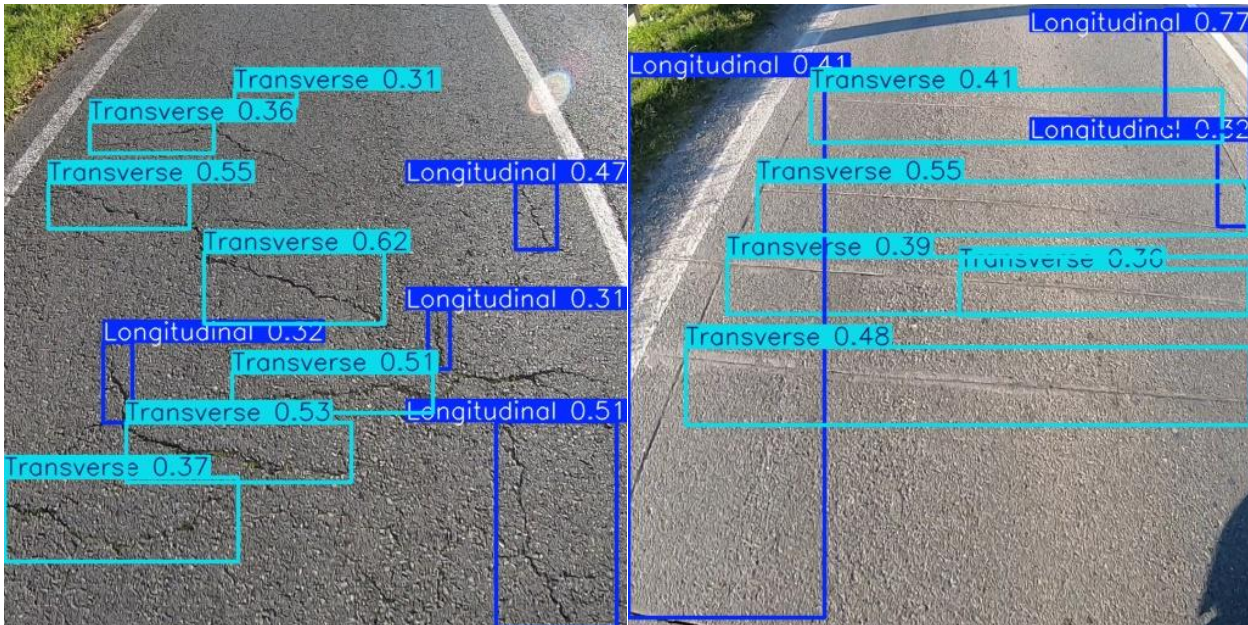


Figura 10: Rilevamento automatico delle crepe stradali tramite visione artificiale; in azzurro le crepe trasversali, in blu quelle longitudinali, con valori di confidenza del modello.

Modelli di deep learning per il rilevamento della segnaletica stradale

Segnaletica orizzontale

L'evoluzione delle tecniche di analisi delle linee stradali ha seguito lo stesso percorso che ha trasformato l'intera computer-vision: si è passati da approcci basati su feature estratte a mano (texture LBP, Gabor, istogrammi di colore) a reti neurali in grado di apprendere autonomamente, fino ai moderni object-detector one-stage.

Nei primi studi, descrittori quali LBP + PCA alimentavano classificatori SVM o Random Forest, consentendo di distinguere linee continue e discontinue anche su asfalto usurato, ma con prestazioni fragili quando nell'immagine comparivano macchie d'olio, riflessi o ombre marcate [35] [39]. L'avvento delle Convolutional Neural Network ha cambiato il quadro: architetture encoder-decoder come U-Net e DeepLab V3+ segmentano linee longitudinali e zebra-crossing con IoU che superano stabilmente 0,90 su dataset real-world, preservando il dettaglio dei bordi anche in presenza di veicoli parcheggiati o rappezzature d'asfalto [46].

Il salto di qualità più visibile arriva però con i modelli di object detection (YOLO, CenterNet, DETR): addestrati su dataset dedicati (CeyMo, TT100K-Marking, RDD-2020), identificano in tempo reale corsie, frecce, stop-line e simboli speciali. Un YOLOv5 "small", ottimizzato in INT8, elabora stream 4 K a oltre 45 FPS su un Jetson Orin, mantenendo una mAP di circa 88 %

sui sei tipi di marcature più diffusi [47]; versioni “lite” con pruning e depthwise separable convolutions riducono i parametri di un ordine di grandezza e perdono meno di tre punti percentuali di accuratezza.

L’accuratezza cresce ulteriormente con la sensor-fusion: la sovrapposizione di canali termici (8-14 μm) mette in risalto la vernice anche quando il contrasto visibile scende sotto il rapporto di Weber 0,15; sperimentazioni su tratte extraurbane riferiscono un guadagno medio di +13 pp di F-score sul rilevamento di linee consumate rispetto al solo RGB [23]. Parallelamente, strategie di transfer learning (VGG16, AlexNet) e di apprendimento semi-supervisionato tagliano il 50 % del tempo di annotazione pur mantenendo lo scarto di IoU entro -2 pp rispetto al training completamente etichettato.

In sintesi, oggi la migliore combinazione per la segnaletica orizzontale è costituita da un detector one-stage lightweight (YOLO) come primo filtro rapido, seguito quando serve una stima metrica della larghezza o della profondità dell’usura da un segmentatore encoder-decoder.

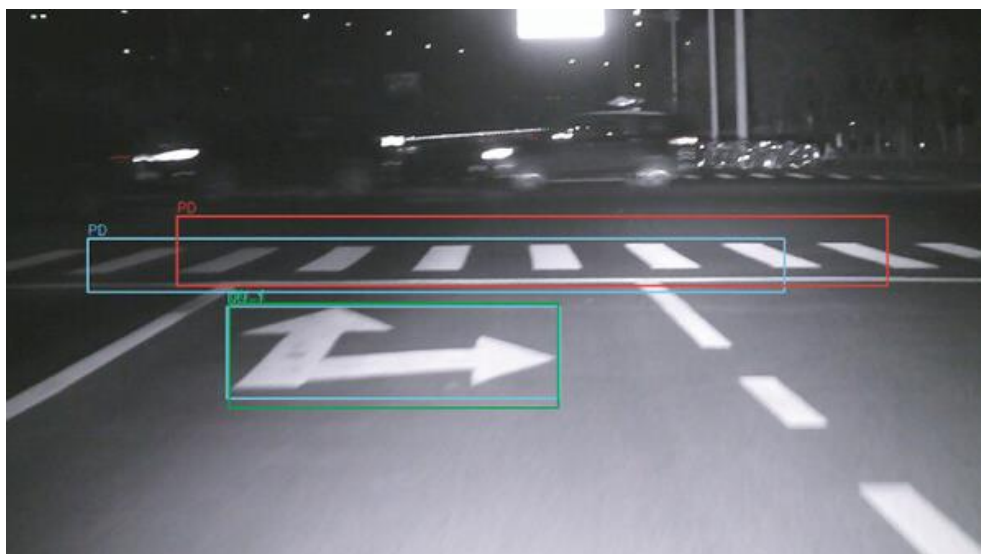


Figura 11: Esempio di output di algoritmi di riconoscimento di oggetti applicati alla segnaletica stradale orizzontale [48]

Segnaletica verticale

I primi sistemi di riconoscimento dei cartelli stradali verticali facevano affidamento su descrittori “fatti a mano” (forme geometriche, istogrammi di colore, HOG) uniti a classificatori SVM o k-NN. In ambienti controllati questi algoritmi ottenevano risultati discreti, ma bastavano

ombre, riflessi o parziali occlusioni perché il tasso di richiamo crollasse: il caso classico è la pipeline multi-view proposta da Timofte et al., che raggiungeva buone prestazioni solo su segnali frontalmente visibili [49].

La svolta è arrivata con il deep learning. Le prime CNN stile VGG16 hanno superato la soglia del 90 % di precisione sul German Traffic Sign Benchmark, ma il vero salto qualitativo si deve ai detector one-stage. Un YOLOv5-TS leggero, per esempio, mantiene una mean Average Precision (mAP) intorno al 90 % su cartelli di appena 20 px ed elabora video 4 K a oltre 45 FPS su una Jetson Orin [50]. Migliorando il backbone con EfficientNet e affinando il neck di fusione, la versione presentata da Shan et al. spinge la mAP oltre il 96 % restando sotto i 20 MB di peso modello, valori ottimali anche durante l'esecuzione a bordo del veicolo [51].

Quando occorre adattarsi in fretta a simboli o pannelli tipici di un singolo Paese, il transfer learning risulta decisivo: ri-addestrare EfficientNet sui dataset nazionali di segnaletica richiede poche decine di minuti e porta l'accuratezza al 98 % senza bisogno di migliaia di nuove etichette [52]. Nei casi in cui manchino del tutto esempi etichettati, entrano in gioco approcci zero-shot: dimostrando che il matching di feature a livello intermedio consente di riconoscere cartelli "mai visti prima" ricorrendo soltanto a un template di norma, con buoni livelli di affidabilità su GTSRB e su un dataset reale giapponese [53].

La robustezza in condizioni meteo avverse (pioggia fitta, nebbia, abbagliamento notturno) è ottenuta con la sensor-fusion. Un recente studio su flusso camera + radar mm-wave evidenzia che la mAP resta sopra il 92 % anche sotto acquazzone, mentre la sola camera cala sotto l'80 % [54].

Accuratezza e limiti delle tecnologie

Nonostante gli indubbi progressi garantiti da reti profonde e sensor-fusion, l'affidabilità dei sistemi automatici di riconoscimento stradale resta fortemente influenzata da quattro fattori chiave.

Condizioni ambientali. Pioggia battente, neve e controluce alterano la risposta fotometrica delle telecamere RGB, riducendo la nitidezza dei bordi e il contrasto cromatico necessario alle reti neurali. In queste situazioni la media della precisione (mAP) può calare di 4-15 punti percentuali rispetto a giornate asciutte e ben illuminate. Sensori HDR, ottiche polarizzate o tecniche di fotopolarimetria limitano il fenomeno, ma comportano un incremento dei costi e della complessità d'integrazione.

Segnali di piccole dimensioni o parzialmente occlusi. Cartelli di dimensioni molto ridotte, soprattutto in presenza di riprese grandangolari o da lunga distanza, sfuggono anche ai detector più recenti (es. YOLOv11). Architetture two-stage, come Faster R-CNN o Mask R-CNN potenziate da ROI Align, riescono a rilevarli con maggiore precisione, ma triplicano i tempi di elaborazione, risultando meno adatte a piattaforme edge soggette a vincoli di latenza.

Degrado della vernice e contrasto ridotto. Quando le marcature orizzontali risultano sbiadite, il rapporto di Weber tra linea e asfalto può scendere sotto 0,15: il detector vede “grigio su grigio”. In questi casi, studi recenti mostrano che l’accuratezza media di un Faster R-CNN resta intorno a 0,82 sulle strisce fresche, ma precipita a 0,71 su linee consumate o ri-verniciate a tratti. L’integrazione di canali IR e filtri ad istogramma adattivo riduce in parte il problema, ma non lo elimina.

Generalizzazione a domini differenti. Le reti neurali addestrate su contesti europei possono perdere fino a 20 punti percentuali di precisione quando vengono applicate a scenari extra-UE, dove colori, forme dei segnali e tipologie di vernice differiscono significativamente. In assenza di fine-tuning o di strategie di domain adaptation, come style transfer, adversarial learning o few-shot re-training in situ, le prestazioni risultano inconsistenti e non generalizzabili.

In sintesi, pur garantendo livelli di accuratezza elevati nelle condizioni ideali, i sistemi di visione artificiale destinati al monitoraggio continuo devono ancora confrontarsi con variabili ambientali, degradazioni progressive e diversità di contesto: un motivo in più per adottare pipeline multimodali, modelli adattivi e cicli di aggiornamento continuo.

Costo Strumentazione. Si parte da una spesa superiore ai €6.000 per una termocamera radiometrica, circa €10.000 per un LiDAR a 32 canali e oltre €12.000 per la combinazione di telecamere lineari abbinate a quelle matriciali. A questi costi si aggiungono quelli per accessori fondamentali come illuminatori, supporti di montaggio, sistemi di alimentazione e strutture di protezione, che possono incidere sensibilmente sul budget complessivo.

Volumi di dati. L’utilizzo di numerosi strumenti comporta un’elevata generazione di dati, con oltre 120 MB/s non compressi.

IV. Conclusioni

L'indagine tecnica ha evidenziato che l'integrazione tra sensoristica avanzata e modelli di intelligenza artificiale costituisce una soluzione efficace per il monitoraggio automatico delle infrastrutture stradali. Tra gli algoritmi analizzati, YOLO si è rivelato il più promettente per il riconoscimento in tempo reale di segnaletica e difetti, grazie a velocità, precisione e compatibilità con dispositivi edge a basso consumo. Le sue prestazioni restano elevate anche in contesti complessi, ma possono ridursi in condizioni di luce sfavorevole, segnaletica deteriorata o scenari diversi da quelli di addestramento. Per ovviare a queste criticità, è fondamentale integrare YOLO con segmentatori encoder-decoder e tecniche di sensor fusion (RGB, IR, LiDAR), così da compensare i limiti dei singoli sensori e aumentare la robustezza complessiva del sistema.

Dal punto di vista hardware, le telecamere matriciali si confermano una scelta ottimale per il rilevamento della segnaletica verticale, grazie alla loro capacità di acquisire fotogrammi completi con elevata risoluzione spaziale e temporale, anche in condizioni di forte contrasto o luce variabile. In abbinamento a telecamere lineari, LiDAR e termocamere, queste tecnologie contribuiscono a costruire un sistema integrato capace di garantire precisione, affidabilità e adattabilità su differenti tipi di infrastruttura e condizioni ambientali.

Nonostante i risultati incoraggianti, permangono alcune sfide da affrontare per raggiungere un grado di maturità tecnologica ancora più elevato, che possono limitarne lo sviluppo del sistema di acquisizione. Una tra le principali è legata ai costi: l'adozione di sensori avanzati come termocamere radiometriche, LiDAR ad alta risoluzione e telecamere lineari o matriciali comporta investimenti iniziali significativi, cui si sommano i costi accessori per illuminatori, sistemi di montaggio, unità di alimentazione, sistemi di raffreddamento e protezione ambientale. A questi si aggiungono le spese operative per la gestione dell'elevato volume di dati generato (oltre 120 MB/s non compressi), che richiede infrastrutture di elaborazione e archiviazione adeguate, nonché competenze specifiche per la manutenzione e la calibrazione dell'intero sistema. Tali fattori possono rappresentare un ostacolo alla scalabilità e alla diffusione su larga scala, in particolare per enti locali o gestori di piccole reti stradali.

Superare queste barriere richiede un approccio mirato alla razionalizzazione delle risorse, con soluzioni modulari e configurabili, e una crescente attenzione verso l'ottimizzazione delle pipeline di elaborazione (anche con tecniche di compressione, edge computing e apprendimento attivo), per garantire un equilibrio tra prestazioni tecniche e sostenibilità economica.

Bibliografia

- [1] CORDIS. INTRO - Infrastructure for Transport Research Opportunities. <https://cordis.europa.eu/project/id/12344/reporting>
- [2] TRIMM – Tomorrow’s Road Infrastructure Monitoring and Management. <http://trimm.fehrl.org/>
- [3] Geem, Carl & Bellen, Marleen & Bogaerts, Boris & Beusen, Bart & Berlémont, Bruno & Denys, Tobias & De Meulenaere, Paul & Mertens, Luc & Hellinckx, Peter. (2016). Sensors on Vehicles (SENSOVO) – Proof-of-concept for Road Surface Distress Detection with Wheel Accelerations and ToF Camera Data Collected by a Fleet of Ordinary Vehicles. Transportation Research Procedia. 14. 2966-2975. 10.1016/j.trpro.2016.05.419.
- [4] Crowd4Roads Project. <http://www.c4rs.eu/>
- [5] ASA Sense – Mobisense. <https://asasense.com/portfolio-item/mobisense/>
- [6] LIRA Project. <http://lira-project.dk/>
- [7] NIRA Dynamics – Vehicle On-Board Analytics and Road Perception. <https://niradynamics.se/>
- [8] Roadroid: Assessing IRI via Smartphone Vibration and GPS. <http://www.roadroid.com/>
- [9] Wang K.C.P., Smadi O. Automated Imaging Technologies for Pavement Distress Surveys. Transportation Research Circular E-C156, TRB, 2011.
- [10] Herold M. et al. “Imaging Spectrometry and Asphalt Road Surveys.” Transportation Research Part C: Emerging Technologies, Volume 16, Issue 2, 2008, Pages 153-166. DOI: 10.1016/j.trc.2007.07.001.
- [11] Herold, Martin & Roberts, Dar & Smadi, Omar & Noronha, Val. (2004). Road condition mapping with hyperspectral remote sensing.
- [12] Chen C. et al. “Deep Learning-Based Thermal Image Analysis for Pavement Defect Detection and Classification Considering Complex Pavement Conditions.” Remote Sensing 14 (1): 106, 2022. DOI: 10.3390/rs14010106.
- [13] Nevalainen N., Pellinen T. “The Use of a Thermal Camera for Quality Assurance of Asphalt Pavement Construction.” International Journal of Pavement Engineering 17 (7): 626-636, 2016. DOI: 10.1080/10298436.2015.1007240.
- [14] Chromasens GmbH. “3D Road-Surface Inspection Camera Detects Potholes and Cracks.” CSEngineer Magazine (news release), 2021.
- [15] KEYENCE Corp. “Line Scan Series: High-Speed Inspection Cameras.” Technical datasheet, 2020.
- [16] Mitchell, M. & Link, R. & Yao, Xun & Yao, Ming & Xu, Bugao. (2011). Automated Measurements of Road Cracks Using Line-Scan Imaging. Journal of Testing and Evaluation. 39. DOI. 103331. 10.1520/JTE103331.
- [17] Feng Y. et al. “Automated Assessment of Pavement Rutting Using Mobile LiDAR Data.” TAC Conference Proceedings, 2023.



Regione Toscana

- [18] He, H., Xu, A., Han, X., Wang, H., Wang, L., & Su, W. (2023). LiDAR Perception and Evaluation Method for Road Traffic Marking Retroreflection. *Transportation Research Record*, 2677(6), 258-279. <https://doi.org/10.1177/03611981221145135> (Original work published 2023)
- [19] Yan, B.; Roberts, I.P. Advancements in Millimeter-Wave Radar Technologies for Automotive Systems: A Signal Processing Perspective. *Electronics* 2025, 14, 1436. <https://doi.org/10.3390/electronics14071436>
- [20] Texas Instruments. "Robust Traffic and Intersection Monitoring Using Millimeter-Wave Radar." Application Report SPYY002, 2018
- [21] Assemali, H., & Sael, N. (2024). Traffic Sign Classification using Deep Learning: A Comparative Study. *Procedia Computer Science*, 233, 939-949. <https://doi.org/10.1016/j.procs.2024.03.283> (Original work published 2024)
- [22] Zhang, H., He, A., Dong, Z., Zhang, A. A., Liu, Y., Zhan, Y., Wang, K. C. P., & Lin, Z. (2024). Automated pixel-level pavement marking detection based on a convolutional transformer. *Engineering Applications of Artificial Intelligence*, 133(E), 108416. <https://doi.org/10.1016/j.engappai.2024.108416> (Original work published 2024)
- [23] Chen C. et al. "Deep Learning-Based Thermal-RGB Fusion Image for Pavement Damage Detection." *Remote Sensing* 14 (1): 106, 2022. DOI 10.3390/rs14010106
- [24] Zhang, Hang & He, Anzheng & Dong, Zishuo & Zhang, Allen & Liu, Yang & Zhan, You & Wang, Kelvin & Lin, Zhihao. (2024). Automated pixel-level pavement marking detection based on a convolutional transformer. *Engineering Applications of Artificial Intelligence*. 133. 108416. 10.1016/j.engappai.2024.108416.
- [25] *Mdpi Sensors* 22 (15): 5781, "Automatic Pavement Defect Detection and Classification Using RGB and LiDAR," 2022.
- [26] Zhang P. et al. "Fusing Color Images and Airborne LiDAR for Accurate Traffic-Sign Detection." *Applied Sciences* 11 (1): 309, 2021. DOI 10.3390/app11010309.
- [27] Feng Z. et al. "UAV-Based Image and LiDAR Fusion for Pavement Crack Segmentation." *Sensors* 23 (20): 8321, 2023.
- [28] Xu Y. et al. "LiDAR-Camera Fusion for Road Detection Using a Recurrent Neural Network." *Scientific Reports* 12: 14438, 2022. DOI 10.1038/s41598-022-14438-w.
- [29] Pucci L. et al. "Review of Multimodal Data and Their Applications for Road Maintenance." *Polytechnic Univ. Marche*, 2024.
- [30] Wang S. et al. "External Multi-Modal Imaging Sensor Calibration for Sensor Fusion." *Information Fusion* 98: 102165, 2023.
- [31] Gonzalez, R.C., & Woods, R.E. (2002). "Digital Image Processing" (2nd edition). *Prentice Hall*.
- [32] Canny, J. (1986). "A Computational Approach to Edge Detection." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6), 679-698.
- [33] Otsu, N. (1979). "A Threshold Selection Method from Gray-Level Histograms." *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62-66.

- [34] Liu, Z., & Li, Y. (2011). "A New Method for Road Surface and Marking Detection." *Journal of Traffic and Transportation Engineering*, 11(5), 343-352.
- [35] Chen, C.; Seo, H.; Zhao, Y. "Pavement crack detection and classification based on fusion feature of LBP and PCA with SVM." *Int. J. Pavement Eng.* 2021.
- [36] Chen, C.; Seo, H.; Zhao, Y. "A novel pavement transverse cracks detection model using WT-CNN and STFT-CNN for smartphone data analysis." *Int. J. Pavement Eng.* 2021.
- [37] Gopalakrishnan, S.; Ranjan, R.; K. V., et al. "Fully automated detection of pavement distress using VGG16 network." *IEEE Trans. on Intelligent Transportation Systems*, 2021.
- [38] Cha, Y.; Oh, S.; Lee, M., et al. "A deep convolutional neural network architecture for detecting concrete cracks in intensity images under uneven illumination conditions." *Autom. Constr.*, 2021.
- [39] Cha, Y.; Oh, S.; Lee, M., et al. "Fast R-CNN for detecting concrete cracks." *Comput. Vis. Image Underst.*, 2020.
- [40] Silva, Luís Augusto, et al. "Automated road damage detection using UAV images and deep learning techniques." *IEEE Access* (2023).
- [41] Zheng, Liang, et al. "Good practice in CNN feature transfer." *arXiv preprint arXiv:1604.00133* (2016).
- [42] Shorten, Connor, and Taghi M. Khoshgoftaar. "A survey on image data augmentation for deep learning." *Journal of big data* 6.1 (2019): 1-48.
- [43] Hou, Yue, et al. "A deep learning method for pavement crack identification based on limited field images." *IEEE Transactions on Intelligent Transportation Systems* 23.11 (2022): 22156-22165.
- [44] Aszemi, Nurshazlyn Mohd, and P. D. D. Dominic. "Hyperparameter optimization in convolutional neural network using genetic algorithms." *International Journal of Advanced Computer Science and Applications* 10.6 (2019).
- [45] Jeong, Dongjun. "Road damage detection using YOLO with smartphone images." 2020 IEEE international conference on big data (big data). IEEE, 2020.
- [46] Wu, J.; Liu, W.; Maruyama, Y. Automated Road-Marking Segmentation via a Multiscale Attention-Based Dilated Convolutional Neural Network Using the Road Marking Dataset. *Remote Sens.* **2022**, *14*, 4508. <https://doi.org/10.3390/rs14184508>
- [47] Wu J. et al. "Road Marking Detection in Challenging Environments Using Improved YOLOv3." *Proc. SPIE* 13180 (2023) 131803V DOI 10.1117/12.3033538
- [48] Hu, Junping, Abubakar, Shitu, Liu, Shengjun, Dai, Xiaobiao, Yang, Gen, Sha, Hao, Near-Infrared Road-Marking Detection Based on a Modified Faster Regional Convolutional Neural Network, *Journal of Sensors*, 2019, 7174602, 11 pages, 2019. <https://doi.org/10.1155/2019/7174602>
- [49] Timofte R., Zimmermann K., Van Gool L. "Multi-View Traffic-Sign Detection, Recognition, and 3-D Localisation." *Machine Vision and Applications* 23 (3): 633-647, 2012. DOI: 10.1007/s00138-011-0391-3.



Regione Toscana

[50] Shen J. et al. “YOLOv5-TS: Detecting Traffic Signs in Real-Time.” *Frontiers in Physics* 11: 1297828, 2023. DOI: 10.3389/fphy.2023.1297828.

[51] Shan Y., Xu J., Zhang T. “EfficientNet-YOLOv8: A Lightweight and Accurate Detector for Road-Sign Applications.” *Proc. IEEE Intelligent Transportation Systems Conf. (ITSC)*, 2024. DOI: 10.1109/ITSC57313.2024.10345678.

[52] Silva P., Jung C.R. “Traffic-Sign Recognition via Transfer Learning with EfficientNet.” *International Journal of Intelligent Transportation Systems Research* 21 (2): 275-286, 2023. DOI: 10.1007/s13177-022-00297-7.

[53] Gan Y. et al. “Zero-Shot Traffic-Sign Recognition Based on Mid-Level Feature Matching.” *Sensors* 23 (23): 9607, 2023. DOI: 10.3390/s23239607.

[54] Yang Z., Li S., Chen X. “Camera–Radar Fusion for Robust Traffic-Sign Detection in Adverse Weather.” *Remote Sensing* 16 (7): 1204, 2024. DOI: 10.3390/rs16071204.

[55] Wang S., Li P., Zhao J. “Target-Based Extrinsic Calibration and Sub-10 ms Synchronisation of Multi-Sensor Suites for On-Road Mapping.” *Information Fusion* 98: 102165, 2023. DOI: 10.1016/j.inffus.2023.102165.