La Consapevolezza del Terreno nel Framework RLOC per la Locomozione Quadrupedale: Dettagli e Logica

1. Introduzione: Contesto della Locomozione Quadrupedale e RLOC

1.1 Le Sfide della Locomozione su Terreni Irregolari per Robot Quadrupedi

La locomozione con robot dotati di arti su terreni variabili rappresenta una sfida intrinseca e complessa, che richiede una percezione estremamente precisa sia dello stato interno del robot (propriocezione) sia dell'ambiente circostante (visione).[1] I metodi di controllo tradizionali, pur efficaci in contesti specifici, tendono a incrementare la loro complessità in modo significativo man mano che vengono considerati scenari più diversificati di terreno irregolare, evidenziando la necessità di soluzioni più adattabili.[2] Il vero potenziale della locomozione con arti risiede nella capacità di un robot di navigare in ambienti complessi con movimenti dinamici, superando le limitazioni imposte dalle assunzioni di stabilità statica.[3] Storicamente, molti approcci si sono basati sull'accesso a mappe di elevazione locali (heightmaps) o su parametri del terreno pre-ingegnerizzati. Tuttavia, tali dipendenze possono comportare costi computazionali elevati per la post-elaborazione dei dati o limitare la capacità di pianificazione reattiva su terreni sconosciuti e diversificati.[4] La geometria del terreno nel mondo reale è intrinsecamente non liscia, non lineare, non convessa e, se percepita da un'unità visiva robot-centrica, appare spesso parzialmente occlusa e rumorosa.[3] Questa complessità intrinseca ha spinto la ricerca verso metodi più adattivi e basati sui dati, capaci di operare con percezione e decisione in tempo reale, piuttosto che affidarsi a modelli perfetti o a mappe predefinite. La transizione verso tali approcci è motivata dalla necessità di superare le limitazioni dei sistemi che richiedono una conoscenza a priori dettagliata dell'ambiente, consentendo ai robot di generalizzare e operare efficacemente in contesti non strutturati e imprevedibili.

1.2 RLOC: Un Approccio Unificato Basato su Modello e Dati

Il lavoro di ricerca intitolato "RLOC: Terrain-Aware Legged Locomotion using Reinforcement Learning and Optimal Control" (arXiv:2012.03094) propone un approccio unificato che combina metodologie basate su modello e basate sui dati per la pianificazione e il controllo di robot quadrupedi, al fine di ottenere una locomozione dinamica su terreni irregolari.[5]
L'obiettivo primario di RLOC è consentire a robot quadrupedi, in particolare i modelli ANYmal B e C, di muoversi in modo robusto attraverso terreni complessi, privilegiando la stabilità rispetto a una locomozione eccessivamente aggressiva.[5]
La denominazione "unificato" per l'approccio di RLOC è significativa, poiché indica una combinazione strategica dei punti di forza di due paradigmi distinti nel controllo robotico. Da un lato, si sfrutta la Reinforcement Learning (RL), una metodologia data-driven che offre capacità di adattamento e apprendimento in ambienti complessi e incerti. Dall'altro lato, si integra il controllo basato su modello, che fornisce precisione e garanzie di stabilità nell'esecuzione a basso livello. Questa fusione è progettata per capitalizzare il meglio di entrambi i mondi: la capacità dell'RL di navigare in ambienti sconosciuti e dinamici, e la prevedibilità e l'affidabilità del controllo basato su modelli fisici. La combinazione di queste tecniche permette al sistema di apprendere strategie di alto livello per la pianificazione del movimento e, contemporaneamente, di eseguire tali piani con la precisione e la robustezza necessarie per l'interazione fisica con il terreno.

2. La "Mappatura del Terreno" nell'Approccio RLOC: Una Prospettiva Implicita

2.1 Distinzione tra Mappatura Esplicita e Consapevolezza Implicita del Terreno in RLOC

A differenza di alcuni metodi che costruiscono esplicitamente mappe geometriche del terreno, come le heightmaps, utilizzate per l'ottimizzazione dei punti di appoggio (ad esempio, come descritto in [3] che tratta l'ottimizzazione della posa della base e dei punti di appoggio rispetto a una heightmap), la "consapevolezza del terreno" in RLOC è ottenuta in modo differente. Il framework RLOC non genera esplicitamente né si affida a un modulo separato per la rappresentazione geometrica del terreno. Invece, la sua policy centrale di Reinforcement Learning (RL) "mappa direttamente le informazioni sensoriali e i comandi di velocità desiderati della base in piani di appoggio".[5]
Questo implica che la policy RL apprende implicitamente a interpretare le caratteristiche del terreno direttamente dai dati sensoriali grezzi durante il suo processo di addestramento,

piuttosto che elaborare questi dati in una rappresentazione di mappa strutturata in anticipo. La natura implicita della mappatura del terreno in RLOC segna un allontanamento dalle tradizionali pipeline robotiche "sense-plan-act" (percepisci-pianifica-agisci), dove la percezione (mappatura) è un modulo distinto e a monte. In RLOC, la policy RL integra la percezione direttamente nel processo decisionale, configurandosi come un sistema "sense-act" (percepisci-agisci), in cui la comprensione del terreno è incorporata all'interno della policy di controllo appresa. Questo approccio può portare a tempi di reazione più rapidi e potenzialmente a comportamenti più robusti in ambienti dinamici e incerti, poiché evita la latenza e i potenziali errori associati alla costruzione esplicita di mappe. La capacità di apprendere direttamente la relazione tra input sensoriali e azioni efficaci sul terreno, senza la necessità di una rappresentazione intermedia, rende il sistema più diretto e potenzialmente più efficiente.

2.2 Come la Policy RL Apprende a Interpretare le Informazioni Sensoriali

La policy RL viene addestrata in simulazione su una "vasta gamma di terreni generati proceduralmente".[5] Questo addestramento estensivo consente alla policy di apprendere complesse correlazioni tra i vari input sensoriali e piani di appoggio efficaci su diverse superfici irregolari. La policy stessa è approssimata da un "multilayer perceptron (MLP)" [6], che funge da funzione che traduce direttamente gli stati osservati (inclusi i dati sensoriali relativi al terreno) in distribuzioni di probabilità sulle azioni (piani di appoggio).
La generazione procedurale di terreni non è semplicemente una comodità per la raccolta dei dati; è un fattore abilitante fondamentale per l'approccio di mappatura implicita. Esponendo l'agente RL a una varietà quasi infinita di configurazioni del terreno durante l'addestramento, si costringe la policy ad apprendere caratteristiche generalizzabili e strategie robuste, piuttosto che memorizzare specifiche configurazioni del terreno. Ciò affronta direttamente la sfida della generalizzazione a nuovi ambienti non visti nel mondo reale. Se la policy non costruisce una mappa esplicita, ha bisogno di imparare modelli robusti da input diversi. La generazione procedurale fornisce questa diversità su vasta scala. Ciò significa che la policy impara a "riconoscere" implicitamente caratteristiche come pendenze, gradini o fosse attraverso le loro firme sensoriali, senza la necessità di una rappresentazione geometrica. Questo è essenziale per l'implementazione nel mondo reale, dove ogni terreno è unico e imprevedibile. L'MLP, come approssimatore di policy, è in grado di apprendere questa complessa funzione non lineare che collega direttamente le firme sensoriali grezze del terreno a sequenze ottimali di passi.

3. Input Sensoriali per la Consapevolezza del Terreno

3.1 Dettaglio del Feedback Sensoriale Utilizzato in RLOC

RLOC dichiara esplicitamente di utilizzare "feedback propriocettivo ed esterocettivo a bordo".[5] Questa combinazione di input sensoriali è fondamentale per consentire al robot di comprendere il proprio stato e l'ambiente circostante in modo completo.

- **Stati Propriocezione:** Questi forniscono misurazioni precise dello stato attuale del robot per una reazione immediata.[9] Tipicamente includono:
    - La posa del robot (posizione e orientamento della base).
    - Letture dell'Unità di Misura Inerziale (IMU), che forniscono dati di accelerazione e velocità angolare.
    - Rotazioni locali delle giunture, che comprendono le posizioni e le velocità delle articolazioni.[4] Sebbene [4] descrivano sensori specifici (encoder digitale a effetto Hall per lo stato delle giunture, IMU LORD Microstrain per lo stato del corpo, filtro di Kalman per la stima) provenienti da un altro studio (tesi di Margolis), questi rappresentano sensori propriocettivi comuni nei robot con arti e sono concettualmente simili a quelli che RLOC utilizzerebbe.
- **Feedback Esterocettivo (Visione):** Questo tipo di input aiuta l'agente a pianificare le manovre su terreni irregolari e intorno a grandi ostacoli, anticipando i cambiamenti nell'ambiente con molti passi di anticipo.[9]
    - Gli input visivi da una telecamera di profondità sono cruciali per percepire gli ostacoli a distanza e per regolare dinamicamente le traiettorie.[9] Analogamente alla propriocezione[4] menzionano esplicitamente la telecamera di profondità Intel RealSense D435 come input visivo, un sensore standard per tali compiti.
- **Comandi di Velocità Desiderati della Base:** Questi comandi vengono anch'essi forniti alla policy RL, indicando l'obiettivo di alto livello o il movimento desiderato del robot.[5]

La combinazione di feedback propriocettivo ed esterocettivo costituisce un input sensoriale multimodale essenziale per una locomozione intelligente. La propriocezione fornisce le informazioni "qui e ora" per un controllo reattivo immediato, mentre l'esterocezione fornisce le informazioni "cosa c'è davanti" per una pianificazione proattiva e l'evitamento degli ostacoli. Questa dualità è fondamentale per una navigazione robusta in ambienti complessi e dinamici.[9] Permette al robot di essere sia agile nelle risposte immediate che lungimirante nella pianificazione del percorso, caratteristiche indispensabili per la locomozione dinamica su terreni accidentati.

3.2 Preprocessing e Rappresentazione degli Input Sensoriali (Contesto Generale e Riferimenti Esterni)

Sebbene gli estratti di RLOC non forniscano dettagli specifici sui passaggi di pre-elaborazione per i suoi input sensoriali, lavori correlati offrono una visione delle pratiche comuni. Le

immagini di profondità grezze possono essere utilizzate direttamente come input per i controller.[4] Per ridurre il rumore e il disallineamento, ad esempio tra i dati di simulazione e quelli del mondo reale, vengono comunemente applicati filtri di pre-elaborazione alle immagini di profondità, come il "Threshold Filter" (che taglia la profondità dei punti nell'immagine tra 0.1m e 1.0m) e l'"Hole-Filling Filter" (un filtro spaziale che rimuove euristicamente artefatti di discontinuità minori).[4]
Le mappe di elevazione del terreno centrate sul corpo (heightmaps) sono spesso utilizzate per addestrare "policy esperte" in simulazione. Questo approccio è preferito per la sua efficienza nel campionamento e per le prestazioni superiori che consente, specialmente per policy altamente dinamiche e adattive al passo.[4] Questo evidenzia una strategia comune per il trasferimento dalla simulazione al mondo reale (sim-to-real), dove una policy addestrata con dati ideali in simulazione viene poi adattata per operare con dati sensoriali più realistici. La fusione dei dati avviene nell'architettura della rete neurale della policy di alto livello, che concatena le caratteristiche di output del modulo di percezione (derivanti dall'osservazione del terreno) con gli input propriocettivi e l'azione precedente.[4] Questo processo illustra come le diverse modalità sensoriali siano combinate per informare il processo decisionale.
La scelta tra dati sensoriali grezzi e rappresentazioni elaborate (come le heightmaps) per la mappatura del terreno rappresenta un compromesso fondamentale nella progettazione. Mentre i dati grezzi offrono immediatezza e evitano la perdita di informazioni, le rappresentazioni elaborate possono semplificare il compito di apprendimento e migliorare le prestazioni in simulazione. La sfida successiva è colmare questa lacuna per l'implementazione nel mondo reale, spesso rendendo necessarie tecniche come il behavioral cloning o l'adattamento del dominio. Ciò implica che la "mappatura" non riguarda solo la raccolta dei dati, ma anche il modo in cui vengono rappresentati e trasformati per l'algoritmo di apprendimento.

Tabella 1: Input Sensoriali e il loro Ruolo nella Policy RL di RLOC

Tipo di Sensore/Input	Dati Forniti (Esempi)	Ruolo nella "Mappatura" Implicita e Controllo	Riferimento
Propriocezione	Stato del robot (pose, velocità), letture IMU (accelerazione, velocità angolare), rotazioni/velocità delle giunture	Misurazione precisa dello stato corrente del robot per reazioni immediate; fornisce feedback sull'interazione fisica con il terreno.	[4]
Esterocettività	Immagini di profondità (da telecamera di profondità)	Percezione di ostacoli e cambiamenti nel terreno a distanza; consente la pianificazione proattiva	[4]

		e l'anticipazione delle modifiche del terreno prima del contatto.	
Comandi di Velocità Desiderati	Velocità lineare e angolare desiderate del corpo base del robot	Indicano l'obiettivo di alto livello del movimento, condizionando la policy RL per generare piani di passo che raggiungano tale obiettivo sul terreno percepito.	[5]

4. La Logica della Policy di Reinforcement Learning per la Pianificazione dei Passi

4.1 Formulazione del Problema come Processo Decisionale di Markov (MDP)

La locomozione quadrupedale è comunemente formulata come un problema di Reinforcement Learning (RL) all'interno del framework dei Processi Decisionali di Markov (MDP).[2] Un MDP è formalmente definito dalla tupla
$M:=(S,A,R,P,\rho 0,\gamma)$.[2]

- **S (Spazio degli Stati):** Rappresenta l'insieme di tutti i possibili stati in cui il robot può trovarsi. In RLOC, questo include le informazioni sensoriali propriocettive ed esterocettive, insieme ai comandi di velocità desiderati della base.[5]
- **A (Spazio delle Azioni):** Rappresenta l'insieme delle azioni che il robot può intraprendere. Per RLOC, la policy RL genera "piani di appoggio".[5] Questi piani specificano probabilmente le posizioni di contatto desiderate dei piedi, la tempistica e, potenzialmente, le forze.
- **R (Funzione di Ricompensa):** $R:S\times A\rightarrow R$ è una funzione di ricompensa scalare che guida il processo di apprendimento. La policy apprende a prendere decisioni per ottenere il massimo beneficio basandosi sulla ricompensa ricevuta dall'ambiente.[2] Per la locomozione, questa include tipicamente termini per il progresso in avanti, la stabilità, l'evitamento di collisioni e, potenzialmente, il raggiungimento di un obiettivo (ad esempio, raggiungere la cima di una montagna).[9]
- **P(s'|s, a) (Probabilità di Transizione di Stato):** Definisce la probabilità di transizione a

un nuovo stato s′ dato lo stato corrente s e l'azione a. Nelle complesse simulazioni fisiche, questa transizione è spesso stocastica.

- **ρ0 (Distribuzione dello Stato Iniziale):** Specifica la distribuzione degli stati di partenza per gli episodi (a volte indicata come μ [6]).
- **γ (Fattore di Sconto):** $\gamma \in [0,1)$ è un fattore di sconto che determina il valore attuale delle ricompense future. Assicura che le ricompense ricevute prima abbiano un valore maggiore rispetto a quelle ricevute più tardi.[11]

L'obiettivo dell'RL è trovare i parametri θ della policy $\pi\theta$ che massimizza il ritorno atteso.[11] Il ritorno atteso per una traiettoria
τ è tipicamente definito come la somma delle ricompense scontate: $Gt=\Sigma k=0T\gamma krt+k+1$.
Il framework MDP fornisce il rigore matematico per la mappatura implicita del terreno di RLOC. Definendo stati che comprendono la percezione del terreno e azioni come piani di appoggio, la funzione di ricompensa può codificare implicitamente un comportamento di attraversamento del terreno "buono". L'algoritmo RL ottimizza quindi la policy per scoprire questa mappatura dalla percezione all'azione, senza la necessità esplicita di un modello geometrico del terreno. Questo costituisce il nucleo dell'implementazione della logica del sistema.

4.2 La Definizione della Policy RL e il suo Addestramento

La policy RL è definita come $\pi:S\rightarrow P(A)$, una funzione che mappa gli stati a distribuzioni di probabilità sulle azioni, tale che $\pi(a|s)$ denota la probabilità di selezionare l'azione a nello stato s.[6] In RLOC, questa policy RL è approssimata utilizzando un "multilayer perceptron (MLP)".[6] Gli MLP sono approssimatori di funzione universali, capaci di apprendere complesse relazioni non lineari tra input ad alta dimensionalità (dati sensoriali) e output (piani di appoggio).
La policy viene addestrata in simulazione su una "vasta gamma di terreni generati proceduralmente".[5] Questa generazione procedurale, che include la variazione di ostacoli, irregolarità, pendenze e dislivelli, espone la policy a una vasta diversità di caratteristiche del terreno, promuovendo la generalizzazione.[9] La scelta dell'MLP come approssimatore di policy, combinata con l'estesa generazione procedurale del terreno, è fondamentale per la capacità di RLOC di "mappare" implicitamente il terreno. L'MLP apprende una funzione non lineare ad alta dimensionalità che traduce direttamente le firme sensoriali grezze del terreno in sequenze ottimali di passi. Questo approccio bypassa la necessità di un'ingegneria esplicita delle caratteristiche o di una ricostruzione geometrica, rendendo il sistema altamente adattabile ma anche fortemente dipendente dalla qualità e dalla diversità dei dati di addestramento della simulazione. La policy non crea una rappresentazione visiva del terreno, ma piuttosto una relazione funzionale tra l'input sensoriale e l'output motorio per l'attraversamento del terreno.
Tabella 3: Formulazione del Processo Decisionale di Markov (MDP) in RLOC

Componente MDP	Descrizione Generale	Interpretazione nel	Riferimento

		Contesto di RLOC	
S (Stato)	Insieme di tutti i possibili stati dell'ambiente e dell'agente.	Include il feedback propriocettivo (pose, velocità, giunture), il feedback esterocettivo (immagini di profondità del terreno) e i comandi di velocità desiderati del robot.	[2]
A (Azione)	Insieme di tutte le azioni che l'agente può intraprendere.	I "footstep plans" generati dalla policy RL, che specificano dove e quando il robot dovrebbe posizionare i suoi piedi.	[5]
R (Ricompensa)	Funzione che assegna un valore numerico a ogni transizione stato-azione.	Progettata per incentivare la locomozione in avanti, la stabilità, l'evitamento di ostacoli e il raggiungimento di obiettivi (es. raggiungere la cima di una montagna).	[2]
P (Probabilità di Transizione)	Probabilità di passare a un nuovo stato s′ dato lo stato attuale s e l'azione a.	Le dinamiche del robot e dell'ambiente simulato, che determinano il prossimo stato del robot dopo l'esecuzione di un piano di passo.	[2]
γ (Fattore di Sconto)	Valore tra 0 e 1 che sconta le ricompense future.	Bilancia l'importanza delle ricompense immediate rispetto a quelle future, incoraggiando la policy a considerare il successo a lungo termine.	[11]

5. Integrazione con il Controllo Basato su Modello

5.1 Il Ruolo del Controller di Movimento Basato su Modello

RLOC adotta un approccio ibrido in cui la policy RL genera piani di appoggio di alto livello, e un "controller di movimento basato su modello" viene utilizzato per tracciare questi piani generati quando il sistema è in funzione online.[5] Questa combinazione sfrutta i punti di forza di entrambi i paradigmi: la capacità della policy RL di adattarsi a ambienti complessi e incerti (approccio data-driven) e la precisione e le garanzie di stabilità del controller basato su modello per l'esecuzione a basso livello.[12]
La struttura di controllo gerarchica in RLOC, che impiega l'RL per la pianificazione e un controllo basato su modello per l'esecuzione, rappresenta un modello di progettazione robusto in robotica. Questa architettura consente alla policy RL di concentrarsi sul compito complesso e guidato dalla percezione del *cosa* fare (pianificazione dei passi basata sul terreno), delegando al contempo il compito preciso e ad alta frequenza del *come* farlo (attuazione delle giunture per realizzare i passi) a un controller basato su modello ben compreso. Questa modularità migliora sia le prestazioni complessive che la sicurezza del sistema. La policy RL gestisce l'aspetto "intelligente" dell'adattamento al terreno, mentre il controller basato su modello si occupa dell'esecuzione fisica con precisione e prevedibilità.

5.2 Policy RL Ausiliarie per Robustezza e Recupero

Oltre alla policy primaria di pianificazione dei passi, RLOC introduce "due policy RL ausiliarie per il tracciamento correttivo del movimento dell'intero corpo e il controllo di recupero".[5] Queste policy ausiliarie sono progettate per gestire "cambiamenti nei parametri fisici e perturbazioni esterne".[5] Un tipo di policy ausiliaria è una "policy residuale" che adatta le traiettorie dall'ottimizzazione della traiettoria (TO), come citato in recensioni generali di DRL (Gangapurwala et al., 2020, che è RLOC stesso).[12] Ciò suggerisce che queste policy potrebbero affinare o correggere l'output del controller basato su modello in base al feedback in tempo reale. Il controller di recupero ha dimostrato di rispondere stabilmente a forze esterne significativamente maggiori rispetto a quelle che il solo controller basato su modello avrebbe potuto gestire.[6]
L'inclusione di policy RL ausiliarie dimostra l'impegno di RLOC verso la robustezza nel mondo reale, andando oltre le condizioni ideali. Mentre la policy RL primaria fornisce la "consapevolezza del terreno" per la locomozione nominale, queste policy secondarie fungono da "rete di sicurezza" o strato adattivo, consentendo al robot di mantenere la stabilità e di recuperare da disturbi inattesi o dinamiche non modellate. Questa capacità è cruciale per colmare il divario tra simulazione e realtà e per garantire prestazioni affidabili in ambienti non

strutturati. La presenza di queste policy aggiuntive rivela un livello più profondo di progettazione orientata alla robustezza, riconoscendo che, anche con una buona pianificazione, la fisica del mondo reale è complessa e imprevedibile.

Tabella 2: Componenti Chiave del Framework RLOC

Componente	Funzione Principale	Metodologia/Tecnologia	Ruolo nella "Mappatura" del Terreno	Riferimento
Policy RL per la Pianificazione dei Passi	Mappa le informazioni sensoriali e i comandi di velocità desiderati in piani di passo.	Reinforcement Learning (RL), Multilayer Perceptron (MLP) come approssimatore di policy.	Il "cuore" della consapevolezza implicita del terreno, apprende direttamente come agire sul terreno.	[5]
Controller di Movimento Basato su Modello	Traccia i piani di passo generati dalla policy RL.	Controllo basato su modello, probabilmente Controllo Ottimale (Optimal Control).	Esegue fisicamente i piani di passo generati in base alla percezione del terreno.	[5]
Policy RL Ausiliarie (Tracciamento e Recupero)	Forniscono correzioni per il tracciamento del movimento del corpo intero e gestiscono perturbazioni esterne.	Reinforcement Learning (RL), Policy residuale.	Migliorano la robustezza e la stabilità, adattandosi a cambiamenti dinamici e imprecisioni nella percezione del terreno o nel modello.	[5]

6. Sfide e Robustezza nell'Implementazione

6.1 Il Gap Sim-to-Real e le sue Implicazioni per la Mappatura del Terreno

Una sfida significativa nella locomozione basata sull'apprendimento è il "gap sim-to-real",

ovvero la discrepanza tra gli ambienti simulati e quelli del mondo reale.[1] Le fonti di questo disallineamento includono dinamiche degli attuatori non modellate, rilevamento impreciso dei contatti, ritardo tra rilevamento, decisione e attuazione, rendering visivo irrealistico e rumore dei sensori non modellato.[4] Per policy altamente dinamiche e adattive al passo, l'addestramento con immagini di profondità grezze può portare a prestazioni significativamente inferiori rispetto all'addestramento con mappe di elevazione del terreno ideali in simulazione. Si ipotizza che ciò sia dovuto al "sostanziale movimento di beccheggio del corpo durante l'esecuzione di salti attraverso grandi fessure", che rende le immagini di profondità grezze meno affidabili per queste manovre dinamiche.[4]
Il gap sim-to-real non è solo un ostacolo tecnico, ma una limitazione fondamentale per qualsiasi sistema basato sull'apprendimento che si affidi alla simulazione. Per la mappatura del terreno, ciò significa che anche se una policy apprende implicitamente a interpretare perfettamente il terreno simulato, il trasferimento di tale "comprensione" al mondo reale, rumoroso e imprevedibile, richiede strategie sofisticate. La sfida è amplificata per i movimenti dinamici, dove i dati dei sensori possono subire distorsioni significative. Questo implica che il "mapping" implicito appreso in simulazione potrebbe non tradursi fedelmente nei dati sensoriali del mondo reale a causa del rumore, delle differenze di rendering e della fisica non modellata.

6.2 Strategie di RLOC per la Robustezza e il Trasferimento Sim-to-Real

RLOC è stato valutato per la sua "robustezza... su una vasta varietà di terreni complessi" e mostra comportamenti che "privilegiano la stabilità rispetto a una locomozione aggressiva".[5] Il framework dimostra "trasferibilità a un robot più grande e pesante, ANYmal C, senza la necessità di riaddestramento".[5] Ha anche mostrato "trasferibilità al robot ANYmal B reale".[6]
Sebbene gli estratti di RLOC non descrivano in dettaglio il suo algoritmo *specifico* di trasferimento sim-to-real, i metodi comuni per il DRL nei robot con arti includono la randomizzazione del dominio, l'adattamento del dominio, l'apprendimento per imitazione, il meta-apprendimento e la distillazione della conoscenza.[2] Una tecnica come il Behavioral Cloning (DAgger) combinata con una Rete Neurale Ricorrente (RNN) può essere utilizzata per trasferire policy addestrate con mappe di elevazione del terreno ground-truth in simulazione a robot reali utilizzando immagini di profondità grezze. La RNN aiuta a integrare i dati di profondità locali nel tempo per imitare la comprensione del terreno da parte dell'esperto.[4] Questa è una forte candidata per il modo in cui RLOC potrebbe ottenere la sua trasferibilità per la consapevolezza del terreno basata sulla visione.
La trasferibilità e la robustezza dimostrate da RLOC, nonostante la natura implicita della sua mappatura del terreno, suggeriscono che la sua metodologia di addestramento (ad esempio, l'estesa generazione procedurale e potenzialmente la randomizzazione del dominio) crea efficacemente policy che sono invarianti a piccole discrepanze tra simulazione e realtà. La priorità data alla stabilità [5] indica inoltre una scelta di progettazione che mitiga intrinsecamente i rischi associati a una comprensione imperfetta del terreno

nell'implementazione nel mondo reale. Ciò dimostra che la "mappatura" non riguarda solo la precisione, ma anche la robustezza e la tolleranza agli errori di fronte alle incertezze del mondo reale.

7. Conclusioni

7.1 Sintesi dell'Approccio RLOC alla Mappatura del Terreno

RLOC definisce la "consapevolezza del terreno" non attraverso una mappatura geometrica esplicita, ma come una capacità implicita appresa da una policy di Reinforcement Learning. Questa policy traduce direttamente gli input sensoriali propriocettivi ed esterocettivi grezzi, insieme ai comandi di velocità desiderati, in piani di appoggio dinamici.[5] La logica centrale è radicata nel framework del Processo Decisionale di Markov (MDP), dove una policy basata su MLP viene addestrata in simulazione su terreni generati proceduralmente per massimizzare una funzione di ricompensa che incoraggia una locomozione stabile ed efficace.[2] Il sistema impiega un'architettura ibrida robusta, combinando la policy RL per la pianificazione di alto livello con un controller basato su modello per un'esecuzione precisa a basso livello, aumentata da policy RL ausiliarie per una maggiore robustezza contro le perturbazioni.[5]

7.2 Implicazioni e Contributo di RLOC nel Campo della Locomozione Robotica

Il successo di RLOC su complessi sistemi quadrupedali (ANYmal B e C) e la sua dimostrata trasferibilità dalla simulazione al mondo reale [5] evidenziano l'efficacia del suo approccio unificato basato su modello e dati per la locomozione dinamica su terreni irregolari. L'apprendimento implicito delle caratteristiche del terreno da parte della policy RL rappresenta un paradigma potente per gestire la complessità ambientale senza la necessità di una mappatura esplicita computazionalmente intensiva o di caratteristiche ingegnerizzate manualmente.
RLOC contribuisce alla crescente tendenza dell'apprendimento "end-to-end" o "percettivo-motorio" in robotica, dove comportamenti complessi emergono dall'apprendimento diretto sui dati sensoriali piuttosto che basarsi su una rigorosa decomposizione modulare. Questo approccio, sebbene impegnativo a causa del gap sim-to-real, offre il potenziale per una locomozione altamente adattiva e agile in ambienti reali non strutturati. La mappatura implicita è una caratteristica chiave di questa tendenza, posizionando RLOC come un passo significativo verso robot con arti veramente autonomi e

adattabili, spingendo i confini di ciò che questi sistemi possono realizzare.

Bibliografia

1. World Model-based Perception for Visual Legged Locomotion - arXiv, accesso eseguito il giorno luglio 20, 2025, https://arxiv.org/html/2409.16784v1
2. Deepreinforcementlearningforreal-world quadrupedal locomotion: a comprehensive review, accesso eseguito il giorno luglio 20, 2025, https://f.oaes.cc/xmlpdf/12bbd333-926b-4682-b8bc-e9945c452687/5115_down.pdf?v=17
3. TAMOLS: Terrain-Aware Motion Optimization for Legged Systems - Research Collection, accesso eseguito il giorno luglio 20, 2025, https://www.research-collection.ethz.ch/bitstream/handle/20.500.11850/558163/1/T_RO21_TAMOLS.pdf
4. Learning Robust Terrain-Aware Locomotion Gabriel ... - DSpace@MIT, accesso eseguito il giorno luglio 20, 2025, https://dspace.mit.edu/bitstream/handle/1721.1/139325/Margolis-gmargo-meng-eecs-2021-thesis.pdf?sequence=1&isAllowed=y
5. arxiv.org, accesso eseguito il giorno luglio 20, 2025, https://arxiv.org/abs/2012.03094
6. (PDF) RLOC: Terrain-Aware Legged Locomotion using Reinforcement Learning and Optimal Control - ResearchGate, accesso eseguito il giorno luglio 20, 2025, https://www.researchgate.net/publication/346700885_RLOC_Terrain-Aware_Legged_Locomotion_using_Reinforcement_Learning_and_Optimal_Control
7. RLOC: Terrain-Aware Legged Locomotion using Reinforcement Learning and Optimal Control - شمرا أكاديميا, accesso eseguito il giorno luglio 20, 2025, https://www.shamra-academia.com/show/3a4436d0e9be5a
8. RLOC: Terrain-Aware Legged Locomotion Using Reinforcement Learning and Optimal Control | Request PDF - ResearchGate, accesso eseguito il giorno luglio 20, 2025, https://www.researchgate.net/publication/360834470_RLOC_Terrain-Aware_Legged_Locomotion_Using_Reinforcement_Learning_and_Optimal_Control
9. [2107.03996] Learning Vision-Guided Quadrupedal Locomotion End-to-End with Cross-Modal Transformers - ar5iv, accesso eseguito il giorno luglio 20, 2025, https://ar5iv.labs.arxiv.org/html/2107.03996
10. Visual-Locomotion: Learning to Walk on Complex Terrains with Vision, accesso eseguito il giorno luglio 20, 2025, https://proceedings.mlr.press/v164/yu22a/yu22a.pdf
11. Reference Free Platform Adaptive Locomotion for Quadrupedal Robots using a Dynamics Conditioned Policy - arXiv, accesso eseguito il giorno luglio 20, 2025, https://www.arxiv.org/pdf/2505.16042
12. Model-free reinforcement learning for robust locomotion using demonstrations from trajectory optimization - Frontiers, accesso eseguito il giorno luglio 20, 2025, https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2022.854212/epub

13. Model-free reinforcement learning for robust locomotion using demonstrations from trajectory optimization - PMC, accesso eseguito il giorno luglio 20, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC9484268/
14. Marco Hutter | Papers With Code, accesso eseguito il giorno luglio 20, 2025, https://paperswithcode.com/author/marco-hutter
15. RLOC: Terrain-Aware Legged Locomotion using Reinforcement Learning and Optimal Control - YouTube, accesso eseguito il giorno luglio 20, 2025, https://www.youtube.com/watch?v=GTI-0gl6Hg0