

Képalapú navigációs modellek összehasonlító vizsgálata önjáró robotok esetében

Vision-based Navigation Models for Autonomous Rovers: Experimental Analysis and Comparison

MOLNÁR Szilárd, TAMÁS Levente

Kolozsvári Műszaki Egyetem
Kolozsvár, str. Memorandumului 28, 400114, <https://utcluj.ro/>
{szilard.molnar,levente.tamas}@aut.utcluj.ro

Abstract

Traditional navigation algorithms rely heavily on the fusion of heterogeneous data from different sensor modalities for precise localization and navigation tasks. However, these approaches lack generalization and interaction with human operators, particularly in navigating unknown environments and human language-based task specifications. In contrast, the foundation models trained on internet-scale data have shown higher generalization capabilities and show an emerging trend in zero-shot learning for navigation tasks. Furthermore, these generic models can close the perception-planning loop through common sense reasoning, applicable to both language-based tasks and open-vocabulary visual recognition. However, due to the scarcity of robotics data for these vision-language-action-based approaches and the lack of clear evaluation protocols for these models, including safety guarantees, it is hard to evaluate the real benefits of these models. In this work, we analyze three different image-based navigation models: ViNT, GNM, and NoMaD. We propose specific metrics to quantitatively evaluate the visual-action-based navigation methods in indoor, outdoor, and simulation environments. In each scenario, these methods are adapted to a different robotic setup than the one provided in the original papers, providing an opportunity to benchmark the generalizability of these methods.

Keywords: autonomous navigation, vision-language-action-models

Kivonat

A hagyományos navigációs algoritmusok nagymértékben támaszkodnak a különböző érzékelőkből származó heterogén adatok fúziójára a pontos lokalizációs és navigációs feladatokhoz. Ezek a megközelítések azonban nem rendelkeznek általánosítással és interakcióval az emberi operátorokkal, különösen az ismeretlen környezetekben való navigálás és az emberi nyelven alapuló feladat-specifikációk feldolgozása során. Ezzel szemben az internetes szintű adatokon betanított alapmodellek nagyobb általánosítási képességeket mutattak és nagymértékű potenciállal rendelkeznek az előre nem látott környezetekben való navigációs feladatokhoz. Továbbá ezek az általános modellek az érvelésükön keresztül előre be nem programozható útvonalakat tervezhetnek meg. Azonban ezen – látáson, nyelven és cselekvésen alapuló – megközelítések esetében a robotikai adatok szűkössége, valamint az egyértelmű értékelési protokollok (beleértve a biztonsági garanciákat is) hiánya miatt nehéz felmérni e modellek valódi előnyeit. Jelen munkában három különböző képalapú navigációs modellt – a ViNT, a GNM és a NoMaD – vizsgálunk. Konkrét mérőszámokat javasolunk a vizuális-cselekvési alapú navigációs módszerek kvantitatív értékelésére beltéri, kültéri és szimulációs környezetben. Az egyes forgatókönyvek során a módszereket az eredeti tanulmányokban szereplőtől eltérő robotikai konfigurációhoz igazítjuk, lehetőséget teremtve ezzel a módszerek általánosíthatóságának összehasonlító vizsgálatára.

Kulcsszavak: önjáró robotok, vizuális és nyelvi modellek

1. FELHASZNÁLT ROBOTOK

A dolgozat célja a képalapú navigációs módszerek összehasonlítása teljesen ismeretlen robotokon és teljesen ismeretlen környezetben, tehát a modellek mindenféle finomhangolása nélkül. Három modellt hasonlítunk össze: ViNT [1], GNM [2] és NoMaD [3]. Mindegyik modell alapja egy referenciaútvonal: először az

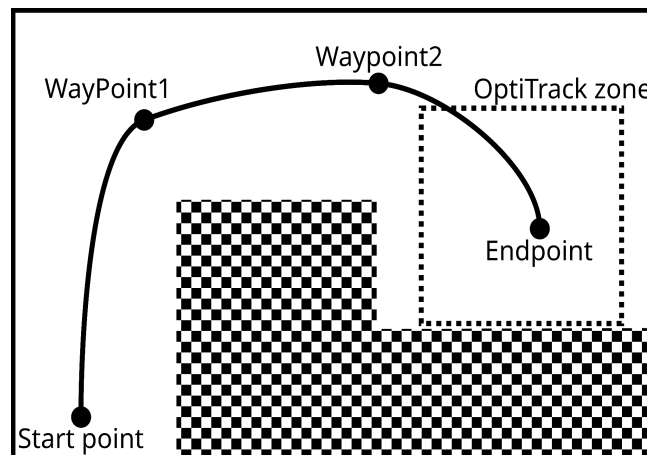
adott robotot kézívezérléssel elkörmányozzuk egy megjelölt útvonalon, miközben periódikusan képeket mentünk el. Ezután minden modell számára a referenciaútvonalat egy sorozatnyi kép reprezentálja; ebben a sorozatban kell mindegyik modellnek megtalálnia az adott pillanatban elfoglalt helyét, illetve kiszámítania azt a mozdulatot, ami a sorozatban következő kép felé navigálja a robotot. Mindegyik környezetben, mindegyik robottal, minden modellt 10 alkalommal futtatunk.

Ehhez három különböző környezetben teszteltünk négy különböző robotot. A környezetek és a hozzájuk tartozó robotok a következők: beltér (Pioneer 3-AT, AD-R1M), kültér (Husarion Panther) és szimulált raktár az Isaac Simben (Clearpath Jackal). A robotokat és környezeteket az 1. ábra mutatja be. Mindegyik roboton egy 1080p felbontású USB kamera szolgáltatja a képet (kivéve a szimulált környezetben, ahol a felbontás 720p-re volt beállítva), illetve minden esetben a modellek egy Jetson Xavier AGX számítógépen futnak (beleértve a szimulált környezetet is).



1. ábra. A felhasznált robotplatformok: Pioneer 3-AT, AD-R1M, Clearpath Jackal, Husarion Panther

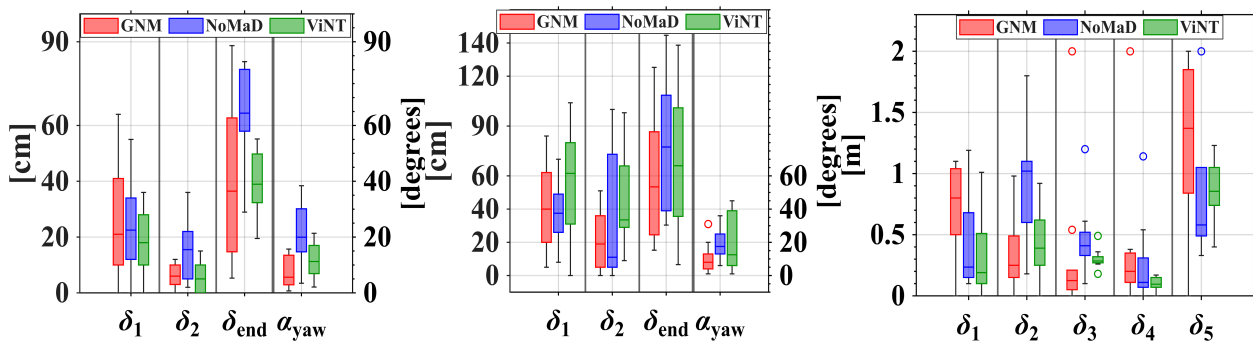
A betéri tesztek során egy körülbelül 7 méter hosszú útvonalat jelöltünk ki, ahol összesen 3 mérföldkő esetében végzünk méréseket (a térkép a 2. ábrán látható): az első kettőnél megmérjük, hogy mennyire közel haladt el a robot az adott mérföldkőhöz, az utolsónál azt pontosan vizsgáljuk, ahol a modell megállította a robotot (mivel úgy ítélte, hogy megérkezett a végső mérföldkőhöz). Az első két mérföldkő esetében egyszerű mérőszalagot használunk, az utolsó pontnál a távolságot és az elfordulást is egy Optitrack rendszerrel mérjük le. A kültéri tesztben a referenciútvonal hossza körülbelül 30 méter, amin 5 mérföldkőhöz képest mérjük a robot elhaladásának távolságát mérőszalaggal. A szimulált környezetben körülbelül 70 méter hosszú a megtett útvonal, amin mérföldkövek helyett az útvonalak között számoljuk ki a távolságot.



2. ábra. A beltéri tesztek referenciatérképe

2. KIÉRTÉKELÉS

A beltéri környezetben két robotplatformmal 1-1 tesztet végeztünk el, míg a kültéri környezetben csak 1 robottal kísérleteztünk. A mérföldkövekhez viszonyított távolságok eloszlását a 3. ábra mutatja be. Ebben látszik, hogy a ViNT és a GNM modellek többnyire közelebb vannak a referenciapontokhoz, mint a NoMaD. Ráadásul a NoMaD esetében nagy probléma volt a balesetek gyakorisága, amikor a robotot kézzel a helyes irányba kellett fordítani, nehogy beleütközzön valamibe. A kültéri és beltéri tesztekben több mint az esetek 70%-ában kellett a NoMaD módszer navigálásába kívülről beleszólni, míg a másik két módszernél csak 30%-ban.

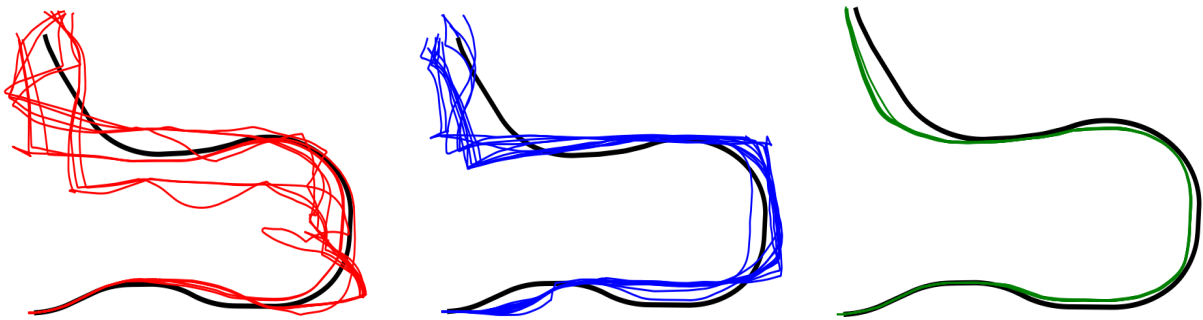


3. ábra. A hibák eloszlása a különböző mérőldkövek esetében.

Bal: a beltéri teszt a Pioneer 3-AT robottal. Közép: a beltéri teszt az AD-RIM robottal.

Jobb: A kültéri teszt a Husarion Panther robottal.

A szimulált környezetben valamelyest megváltozik az erőviszony; az útvonalakat a 4. ábrán rajzoltuk meg. A ViNT modell a valós környezetben elfogadható eredményt ért el, a szimulációban pedig szinte tökéletesen követi a referenciaútvonalat, és mindegyik teszt futtatása megegyezik. A NoMaD a valós tesztekhez hasonlóan bizonytalan. A GNM a valós tesztekben elfogadható volt, a szimulációban viszont a legrosszabb teljesítményt nyújtotta. Se a GNM, se a NoMaD nem tudott végigmenni a pályán beavatkozás nélkül. Az útvonalakat összehasonlítva a referenciával, a ViNT eltérése 0,82 méter, a NoMaD-é 2,03 méter, a GNM-é pedig 2,44 méter.



4. ábra. Megtett útvonalak a szimulált környezetben.

Fekete – referencia, Zöld – ViNT [1], Kék – NoMaD [3], Piros – GNM [2].

Mindegyik modellt ugyanazon a Jetson Xavier AGX beágyazott számítógépen futtattuk, ahol a Robotic Operating System (ROS) segítségével irányítottuk a robotokat. Megmértük mindegyik modell memóriaigényét, teljesítményét és futási idejét, amelyeket az 1. táblázatban mutatunk be. Ebből az tűnik ki, hogy a legenergiatakarékosabb modell a GNM, utána a ViNT, majd a NoMaD.

A modellek teljesítménybeli összehasonlítása.

1. táblázat

	Memória [GB]	Futási idő [mp]	Teljesítmény [W]
GNM	0.63	0.28	11.4
NoMaD	1.29	1.16	12.7
ViNT	0.95	0.65	12.1

3. ÖSSZEFOGLALÁS

Ebben a dolgozatban három látásalapú navigációs módszert hasonlítottunk össze négy különféle platformon három különböző környezetben. A ViNT [1] modell bizonyult a legmegbízhatóbbnak, amelyik minden környezetben elfogadhatóan működik, közepes energiafelhasználással. A GNM [2] jól és hatékonyan működik

valós környezetekben, a szimulált környezetben viszont használhatatlan. A NoMaD [3] általánosan nemcsak megbízhatatlan, hanem lassú és nagy memóriaigényű is. Az eredményekről egy bővebb leírás található a [4] dolgozatban.

KÖSZÖNETNYILVÁNÍTÁS

A dolgozat készítői hálásak a Analog Devices International cégnek, a felajánlott Nvidia eszközökért. Ezt a munkát a Domus Alapítvány is támogatta.

IRODALMI HIVATKOZÁSOK

- [1] Shah D., Sridhar A., Dashora N., Stachowicz K. et al., ViNT: A Foundation Model for Visual Navigation, Conference on Robot Learning, CoRL 2023, 6-9 November 2023, Atlanta, GA, USA, ser. Proceedings of Machine Learning Research, J. Tan, M. Toussaint, and K. Darvish, Eds., vol. 229. PMLR, 2023, pp. 711–733.
- [2] Shah D., Sridhar A., Bhorkar A., Hirose N., and Levine S., GNM: A General Navigation Model to Drive Any Robot, IEEE International Conference on Robotics and Automation, ICRA 2023, London, UK, May 29 - June 2, 2023. IEEE, 2023, pp. 7226–7233.
- [3] Sridhar A., Shah D., Glossop C., and Levine S., NoMaD: Goal Masked Diffusion Policies for Navigation and Exploration, IEEE International Conference on Robotics and Automation, ICRA 2024, Yokohama, Japan, May 13- 17, 2024. IEEE, 2024, pp. 63–70.
- [4] Molnar S. and Tamas T., Vision-based Navigation Models for Autonomous Rovers: Experimental Analysis and Comparison, *2026 12th International Conference on Control, Decision and Information Technologies (CoDIT)*, Bari, Italy, 2026, pp. 2223-2228.