ČESKÝ INSTITUT
INFORMATIKY
ROBOTIKY
A KYBERNETIKY
ČVUT V PRAZE

Velké Modely pro Robotickou Manipulaci

Georgy Ponimatkin
IMPACT CIIRC ČVUT

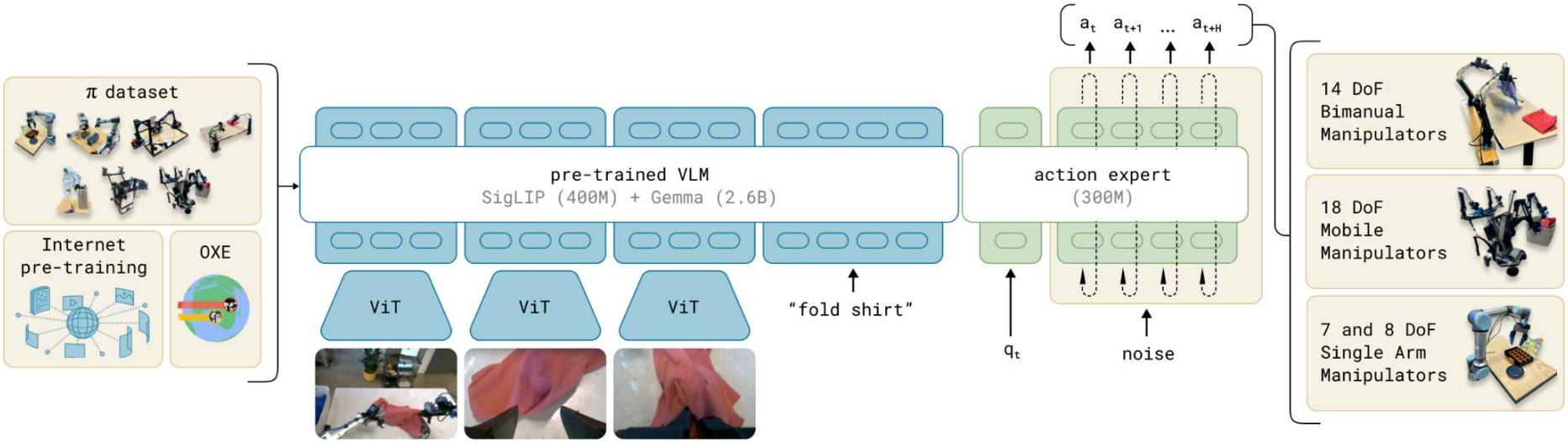
Velké Modely - Nové Příležitosti v Robotice

- Nové příležitosti pro inteligentní robotiku
- Modely jsou robustní a mají dobrou generalizaci
  - Pro některé úlohy dobrý výkon bez potřeby adaptace
  - Začínají generalizovat k neviděným úlohám
- Hlavní problémy pro širokou aplikaci
  - Hlavním limitem jsou trénovací data
  - Robotická data - velmi drahá a náročná na sběr
  - Internet - “levná” data ale jak je použít pro robotiku?

Příklad robotických dat - DROID dataset

A. Khazatsky, K. Pertsch et al.; DROID: A Large-Scale In-The-Wild Robot Manipulation Dataset, RSS 2024

State of the Art - Modely π od Physical Intelligence



K. Black et al.; π_0: A Vision-Language-Action Flow Model for General Robot Control, arxiv:2410.24164v1

State of the Art - Modely π od Physical Intelligence

“Put banana into the box”

ČVUT
CIIRC

Jak získat kvalitní alternativní data?

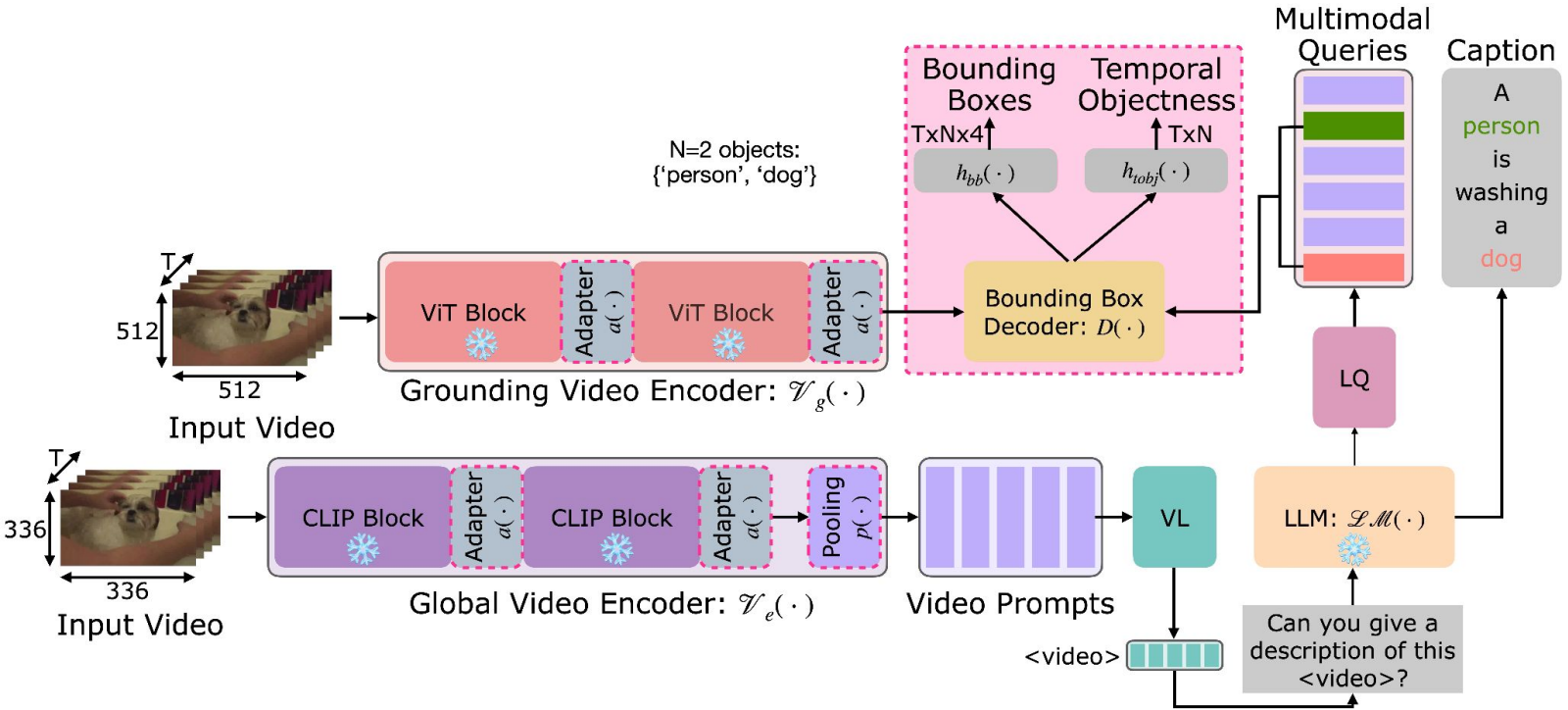
GROVE - Visual Grounding přes YouTube

“Jak spolehlivě najít správná videa ve velkém objemu dat?”

- Metoda vychází z HowTo100M datasetu (100M clipu z YouTube)
  - Automatická anotace s pomocí velkých modelů
- Nový model GROVE
  - Generuje textové popisky, lokalizuje objekty
  - Nejlepší model pro tyto úlohy na základě 5 benchmarků
- Aplikace pro robotiku
  - Inference na velkém souborů videí z YouTube
  - Následné vyhledávání na základě klíčových slov a poskytnutí lokalizace

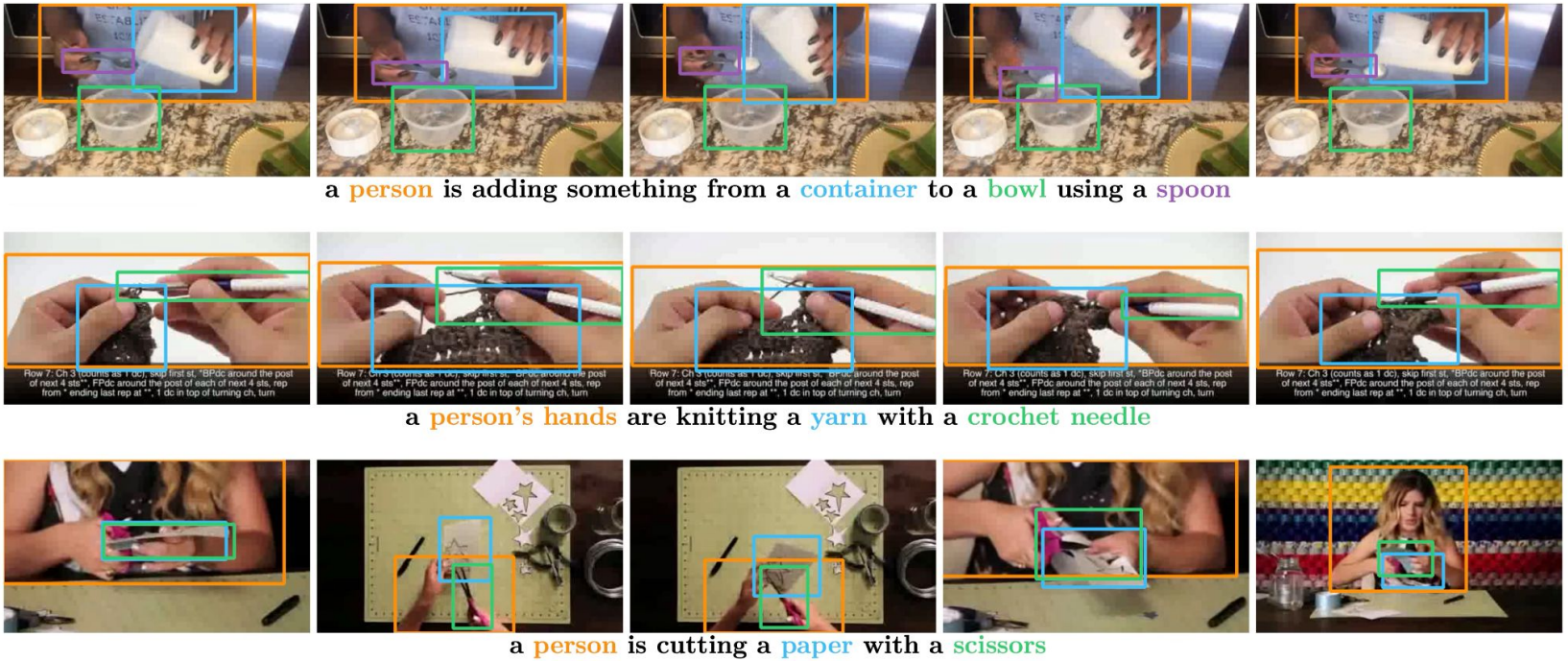
Evangelos Kazakos, Cordelia Schmid, Josef Sivic; Large-scale Pre-training for Grounded Video Caption Generation; ICCV 2025

GROVE - Visual Grounding přes YouTube



Evangelos Kazakos, Cordelia Schmid, Josef Sivic; Large-scale Pre-training for Grounded Video Caption Generation; ICCV 2025

GROVE - Visual Grounding přes YouTube



Evangelos Kazakos, Cordelia Schmid, Josef Sivic; Large-scale Pre-training for Grounded Video Caption Generation; ICCV 2025

FreePose - Přibližné 3D Trajektorie z Videa

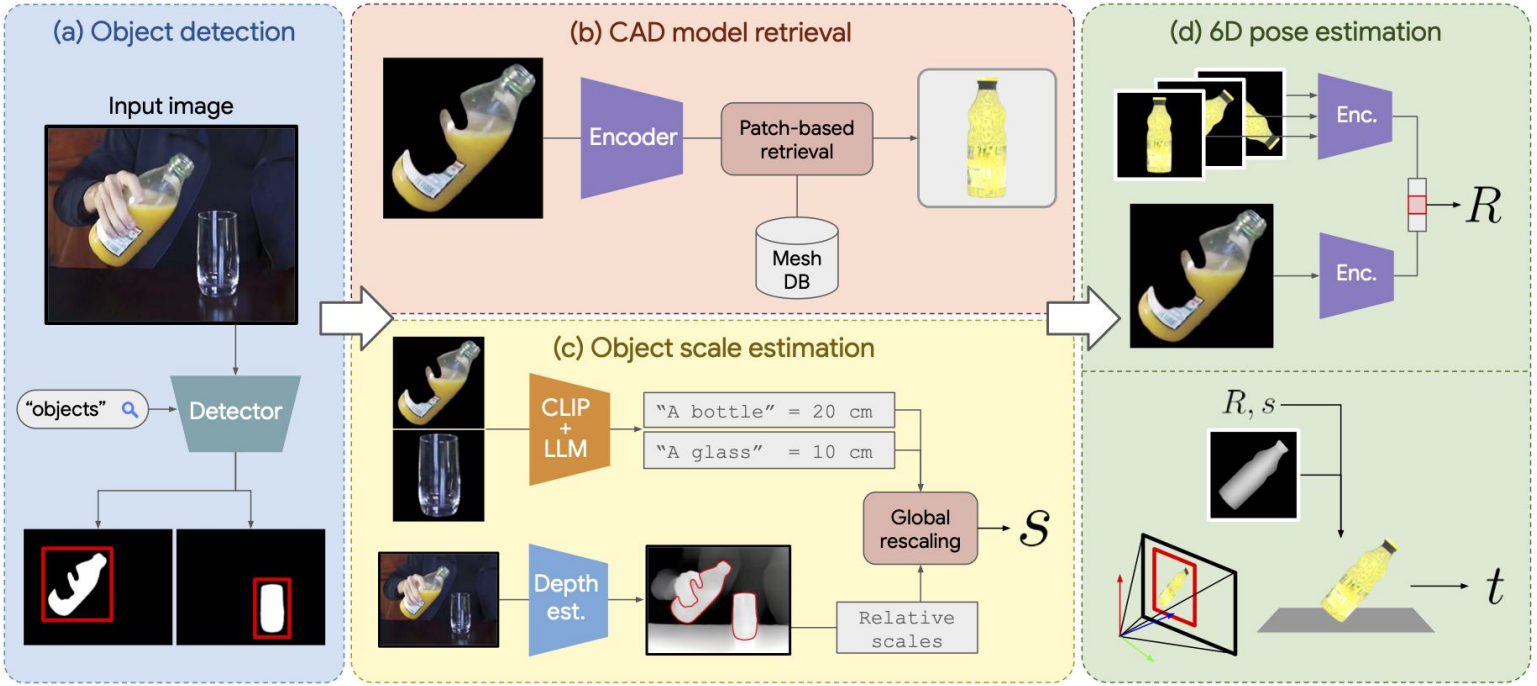
“Jak dostat přibližný 3D pohyb z videa?”

- Kombinace DINOv2 ViT a velké (~12TB) databáze 3D objektů
- Extrakce chybějící 3D informace (rozměry objektů) s pomocí LLM
- Metoda nepotřebuje žádné trénování, pouze jednorázový preprocessing

(a) input video (b) retrieved mesh (c) 6D pose trajectory (d) robot trajectory

G. Ponimatkin*, M. Cífka*, T. Souček, M. Fourmy, Y. Labbé, V. Petrík, J. Sivic; 6D Object Pose Tracking in Internet Videos for Robotic Manipulation; ICLR 2025

FreePose - Přibližné 3D Trajektorie z Videa



G. Ponimatkin*, M. Cífka*, T. Souček, M. Fourmy, Y. Labbé, V. Petrík, J. Sivic; 6D Object Pose Tracking in Internet Videos for Robotic Manipulation; ICLR 2025

FreePose + GROVE = 3D Trajektorie z YouTube

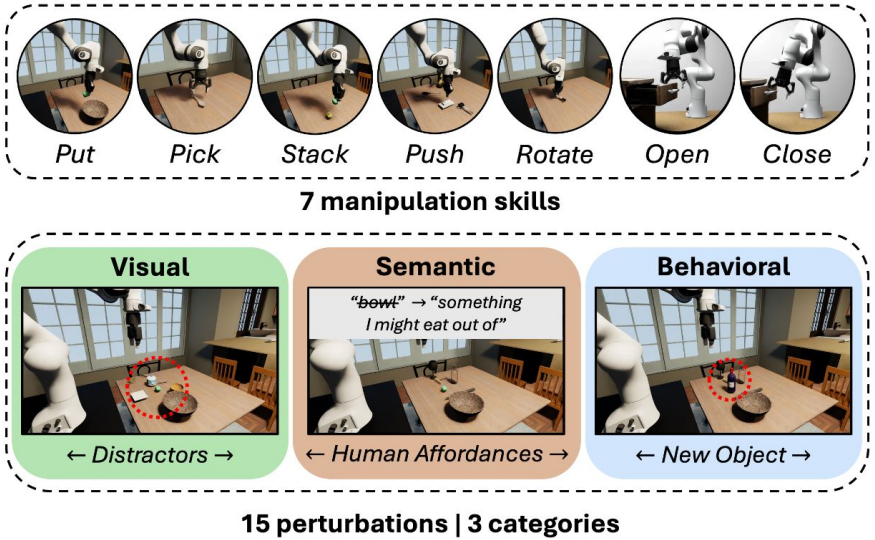
G. Ponimatkin*, M. Cífka*, T. Souček, M. Fourmy, Y. Labbé, V. Petrík, J. Sivic; 6D Object Pose Tracking in Internet Videos for Robotic Manipulation; ICLR 2025

FreePose + GROVE = 3D Trajektorie z YouTube

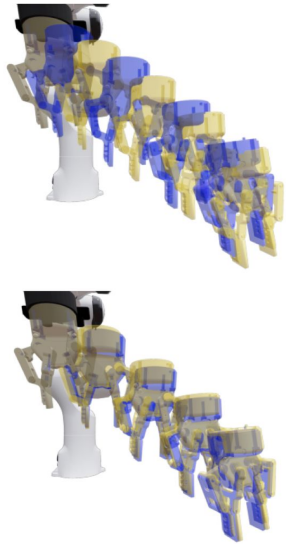
- Kombinací dvou metod dokážeme získat přibližné 3D trajektorie
- Jak to ale použít v praxi pro training VLA modelů? **Work in Progress!**

REALM - Robustní Simulátor pro Evaluaci Modelů

- Aligned simulátor pro evaluaci velkých robotických modelů
- Umožňuje reprodukovatelně testovat výkon velkých modelů pro robotiku

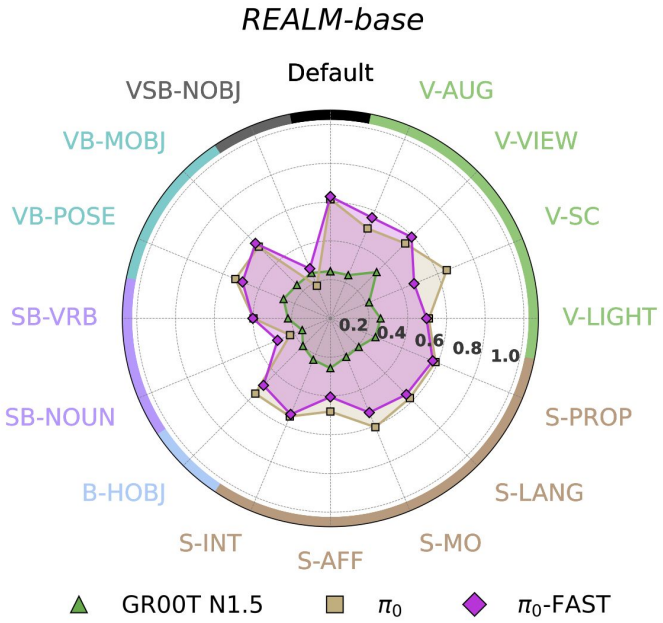


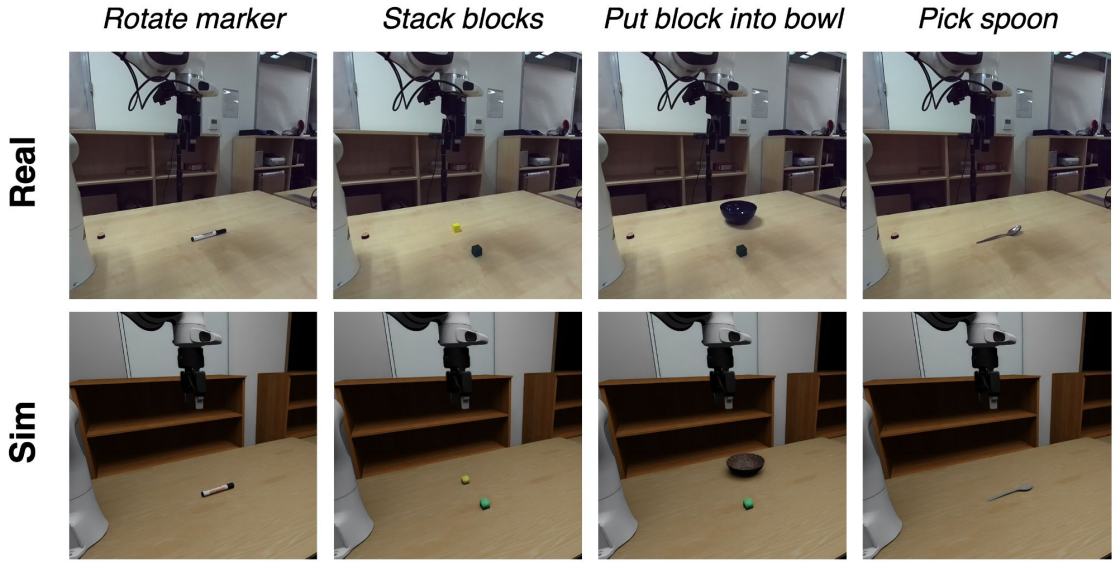
real-to-sim aligned



M. Sedlacek*, P. Yefanov*, **G. Ponimatkin** et al.; REALM: A Real-to-Sim Aligned Benchmark for Generalization in Robotic Manipulation

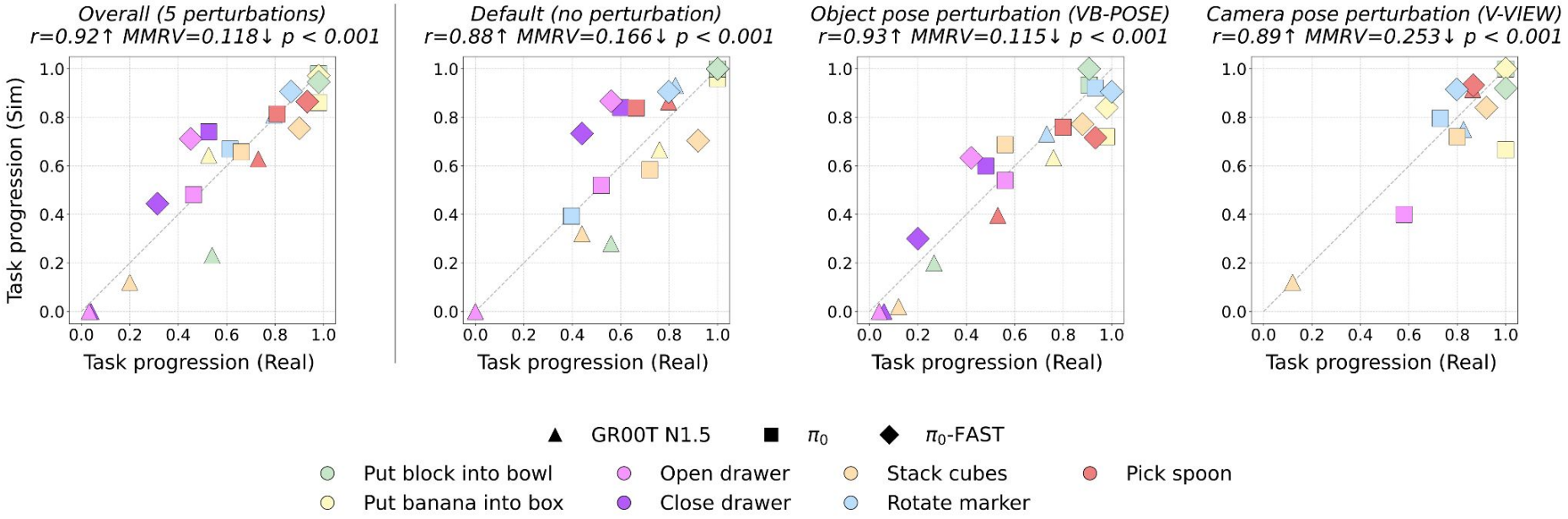
REALM - Robustní Simulátor pro Evaluaci Modelů





M. Sedlacek*, P. Yefanov*, **G. Ponimatkin** et al.; REALM: A Real-to-Sim Aligned Benchmark for Generalization in Robotic Manipulation

REALM - Robustní Simulátor pro Evaluaci Modelů



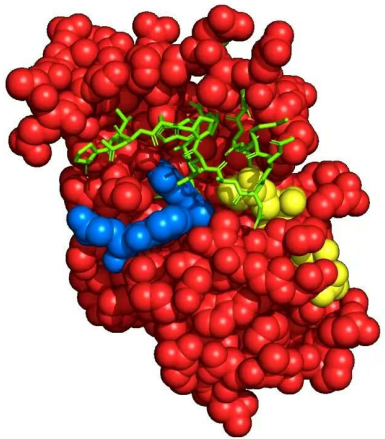
M. Sedlacek*, P. Yefanov*, **G. Ponimatkin** et al.; REALM: A Real-to-Sim Aligned Benchmark for Generalization in Robotic Manipulation

ČVUT
CIIRC

Některé výsledky z ML pro biologii

Flexpert - predikce flexibility proteinů

Flexibilita je důležitá vlastnost proteinů, často zprostředkující jejich funkci.

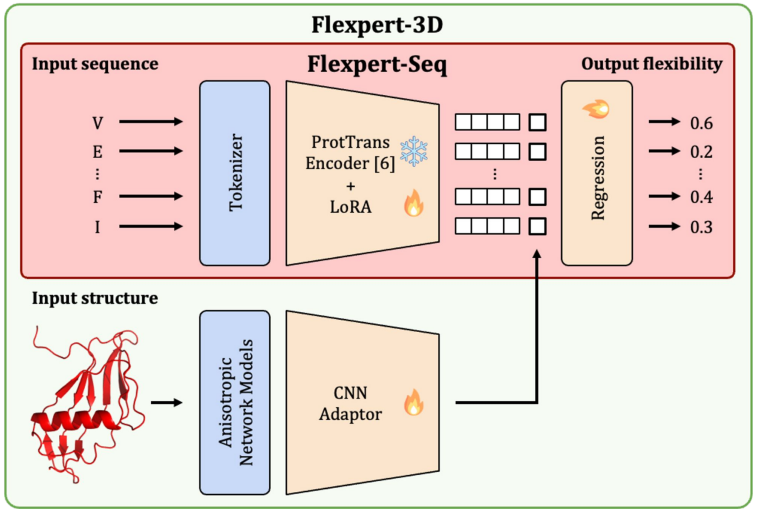


P. Kouba, J. Planas-Iglesias, J Damborsky, J Sedlar, S Mazurenko, J. Sivic; Learning to engineer protein flexibility; ICLR 2025

Flexpert - predikce flexibility proteinů

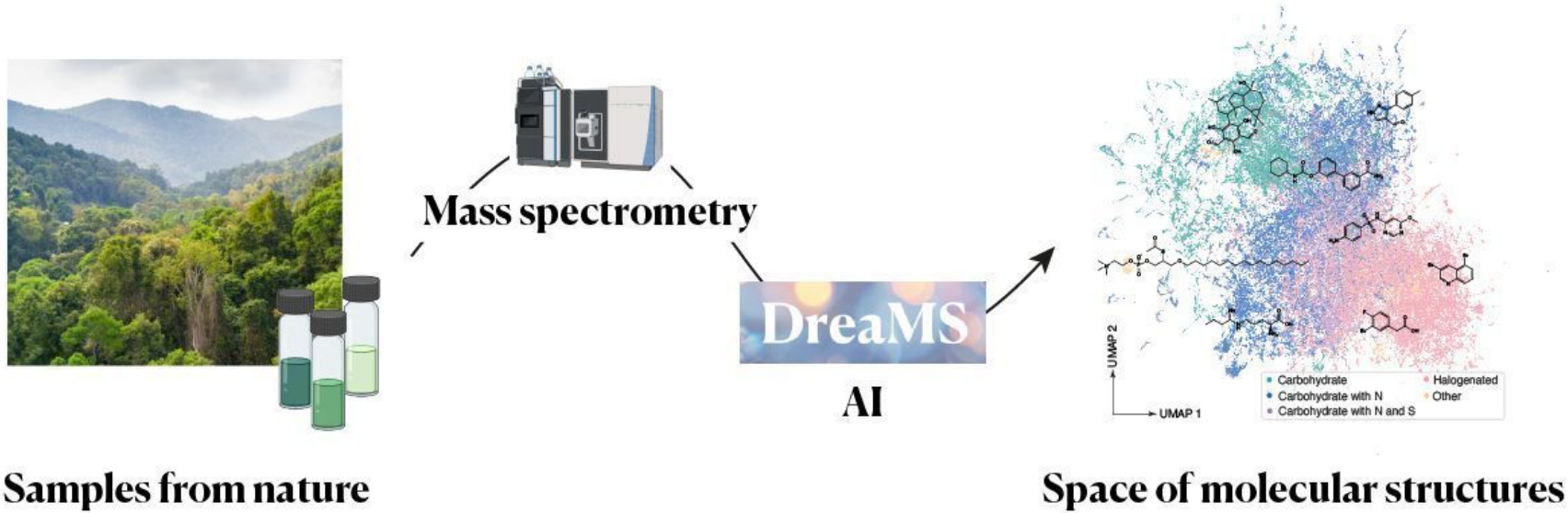
Tradiční metody pro studium proteinové flexibility, jsou nesmírně výpočetně a časově náročné.

Naše řešení: Flexpert, rychlý prediktor flexibility využívající jazykový model předtrénovaný na proteinových sekvencích:



P. Kouba, J. Planas-Iglesias, J Damborsky, J Sedlar, S Mazurenko, J. Sivic; Learning to engineer protein flexibility; ICLR 2025

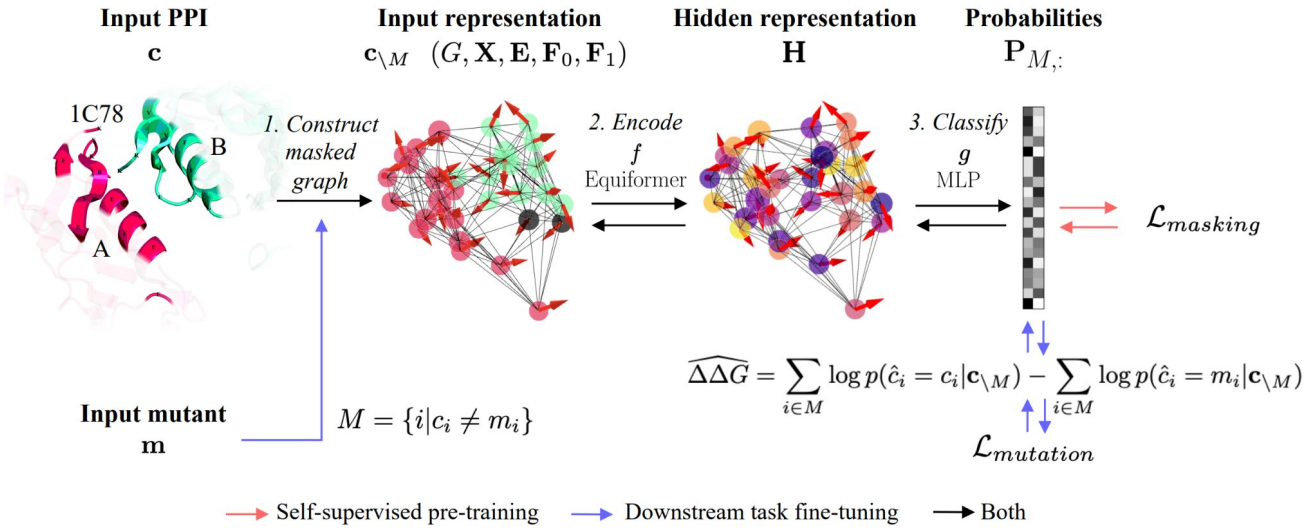
DreaMS – velký jazykový model pro objevování nových molekul z přírody (např. kandidátů na léčiva z rostlin)



R. Bushuiev*, A. Bushuiev*, R. Samusevich, C. Brungs, J. Sivic, T. Pluskal; Self-supervised learning of molecular representations from millions of tandem mass spectra using DreaMS; Nature Biotechnology 2025

PPIformer – geometrický transformer pro predikci účinků mutací na protein-proteinové interakce

Vstup: dva proteiny a seznam jejich mutací



Výstup: pravděpodobnost, že mutace zlepšuje interakci mezi proteiny.

A. Bushuiev*, R. Bushuiev*, …, J. Sivic; Learning to design protein-protein interactions with enhanced generalization; ICLR 2024

ČESKÝ INSTITUT
INFORMATIKY
ROBOTIKY
A KYBERNETIKY
ČVUT V PRAZE

Děkuji za pozornost!