université virtuelle Burkina Faso

Sujet	Apprentissage par Renforcement
Professeur	Issouf OUATTARA
Titre	Résumé d'article
Auteur	Coulibaly Cheick Ahmed

Présentation du problème

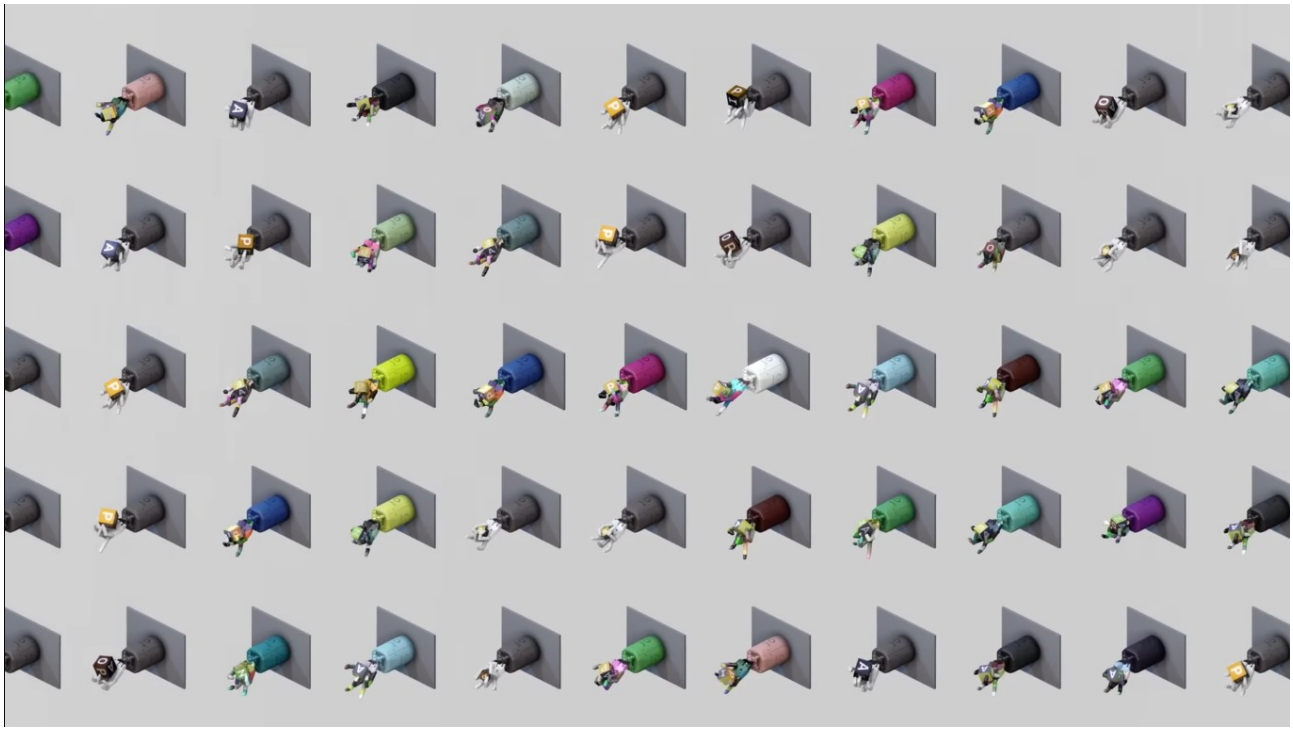
l'article cherche à résoudre le problème de la manipulation complexe d'objets avec une main robotique (la *Shadow Dexterous Hand*). L'objectif est d'apprendre à un robot à faire tourner un cube dans sa main jusqu'à obtenir une orientation cible, spécifiée par une couleur particulière sur une face. C'est un problème difficile car : la main est hautement redondante (beaucoup de degrés de liberté, mouvements coordonnés complexes), l'environnement réel est bruyant et imprévisible (frottements, imprécisions mécaniques, capteurs imparfaits), et l'apprentissage direct sur le robot est très coûteux en temps et en usure matérielle.

L'idée est donc de développer une méthode qui permet à la main robotique d'apprendre des stratégies de manipulation robustes en simulation, puis de les transférer avec succès dans le monde réel.

Méthode

L'approche repose sur l'apprentissage par renforcement profond (Deep RL), avec des techniques de randomisation pour combler le fossé simulation–réalité.

Randomisation : pendant l'entraînement en simulation, on introduit volontairement de l'aléatoire dans les paramètres de l'environnement, pour que le robot n'apprenne pas à dépendre d'un environnement trop "parfait", mais qu'il soit robuste aux imprécisions et variations qu'il rencontrera dans le monde réel
exemples : changer légèrement la masse du cube, la gravité, l'éclairage, couleur de fond
Sans cette randomisation, la politique apprend à manipuler le cube **dans une simulation parfaite**, mais quand on la transfère au robot réel avec un frottement légèrement différent, une caméra un peu plus sombre, le robot échouerait.



a) agent
- l'agent est une politique neuronale (réseau de neurones) qui prend en entrée des observations (vision et capteurs) et qui produit des actions (commandes des moteurs de la main),
- il est entraîné pour maximiser une récompense liée à l'alignement du cube avec l'orientation cible

b) environnement : l'environnement est une simulation physique réalisée avec *MuJoCo*.
On y modélise la main robotique, le cube et les lois physiques (contacts, frottements, gravité...). L'environnement est enrichi par de la randomisation (paramètres physiques, textures, bruits capteurs, etc.) pour que la politique soit robuste au transfert réel.

MuJoCo, abréviation de Multi-Joint Dynamics with Contact, est un moteur physique polyvalent adapté à des applications scientifiques telles que la robotique, la biomécanique et l'apprentissage automatique. Il a été décrit pour la première fois en 2012 dans un article d'Emanuel Todorov, Tom Erez et Yuval Tassa, puis commercialisé par Roboti LLC. Selon une recherche Google Scholar, en avril 2024, la publication originale avait été citée 5 329 fois et le moteur MuJoCo 9 250 fois. Il a été décrit par Zhao et Queralta dans leur revue comme l'un des simulateurs les plus utilisés dans la littérature.

MuJoCo a été acquis par Google DeepMind en octobre 2021 et mis en open source sous licence Apache 2.0 en mai 2022. Certaines parties de la suite de contrôle Deepmind sont alimentées par le moteur MuJoCo.

Wikipédia

c) Ensemble des états

Les observations (l'état vu par l'agent) incluent :
- la position et orientation du cube (via suivi visuel et marqueurs),
- les angles articulaires de la main,

- les vitesses articulaires,
- des informations tactiles indirectes (contact estimé via capteurs).
En simulation, ces états sont accessibles précisément ; dans le monde réel, ils proviennent de la vision par caméra et des capteurs internes.

d) Ensemble des actions : les actions correspondent aux commandes de couple (En robotique, le couple est l'équivalent rotationnel de la force, représentant la force qui fait tourner un objet, exprimé en Newton-mètre Nm) ou de position des moteurs de chaque articulation de la main robotique. La politique choisit à chaque pas de temps les mouvements des doigts pour manipuler le cube.

e) Fonction de renforcement (reward) : la récompense est définie pour encourager l'alignement du cube avec l'orientation cible, plus le cube est proche de la rotation désirée (mesurée par l'écart d'angle), plus la récompense est élevée, des bonus sont attribués lorsque le cube atteint effectivement la bonne orientation.

f) Entraînement et déploiement : l'entraînement est réalisé uniquement en simulation (très rapide et sans risque matériel). Pour combler l'écart entre simulation et réalité (*sim2real gap*), nous avons utilisés la randomisation de domaine : variation aléatoire des masses, frottements, délais moteurs, textures, bruit visuel, etc.
Après entraînement, la politique est directement déployée sur la vraie main robotique sans réapprentissage.
ésultat : le robot parvient à effectuer la manipulation du cube dans le monde réel, malgré les incertitudes.

Quand on entraîne un robot en simulation, tout est contrôlé : les lois physiques sont parfaites, les capteurs donnent des valeurs exactes, les objets ont toujours la même masse, le même frottement, la même texture. Mais dans le monde réel, rien n'est aussi parfait : les frottements peuvent varier (poussière, humidité...), les caméras ajoutent du bruit, les objets ne sont jamais exactement identiques. Du coup, un agent entraîné en simulation échoue souvent quand on le met directement dans le réel. Cet écart entre la simulation idéale et la réalité imparfaite, c'est le sim2real gap.

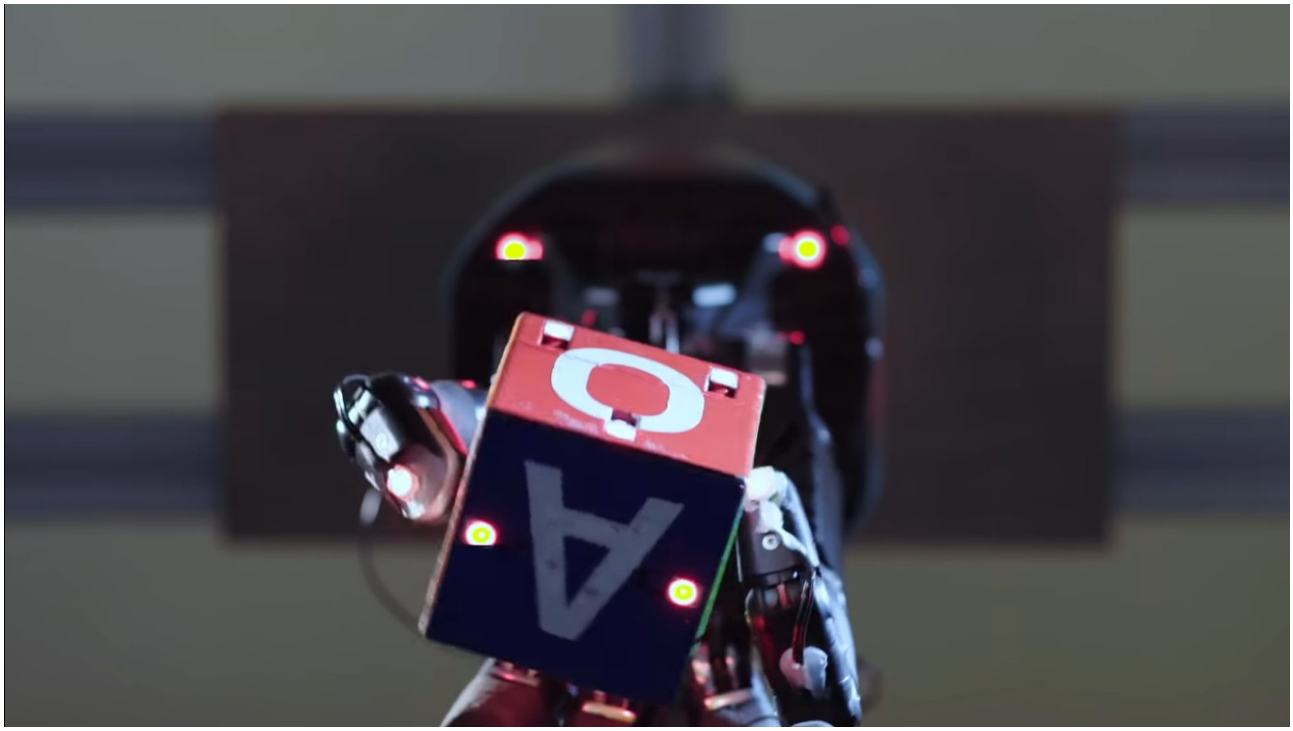
3. Expérience et résultats
Les chercheurs ont mené plusieurs expériences avec la main robotique (Shadow Dexterous Hand) pour tester la robustesse du modèle entraîné.

Tâche principale : manipulation d'un cube de 6 faces colorées, pour pouvoir tourner le cube dans la main pour que la face de couleur demandée apparaisse sur le dessus.

Résultats principaux :
- taux de réussite : environ 60 % des tentatives réussissent dans le monde réel (le robot amène la bonne face en haut) ;
- le robot est capable de réaliser des rotations continues, parfois sur plusieurs tours complets du cube ;
- le comportement appris est fluide et naturel, avec des repositionnements dynamiques du cube et de la main.

Ils ont testé la robustesse du système sous différentes perturbations :
- ajout d'objets mous dans la main → le robot compense et continue la tâche,
- déplacement du cube par un humain pendant la manipulation → le robot s'adapte et corrige,
- différentes conditions d'éclairage et de texture de cube → le système reste efficace.

Performance :
- l robot atteint la cible dans plus de la moitié des essais,
- la politique, entraînée en simulation, est directement transférée sans fine-tuning en réel,
- la politique apprend des stratégies variées : repositionner le cube, le pousser contre la paume, utiliser plusieurs doigts à la fois...



Conclusion et discussion

Forces de l'étude :
- première démonstration convaincante de manipulation d'objets complexes en main avec une politique RL entraînée en simulation,
- utilisation innovante de la randomisation de domaine pour résoudre le problème du *sim2real gap (simulation to reality gap - l'écart entre la simulation et le monde réel),*
- système robuste : le robot continue la tâche malgré des perturbations inattendues,
- contribution importante pour l'avenir de la robotique autonome et de la manipulation fine.

Limites et critiques
- taux de réussite limité (~60 %) : même si c'est impressionnant, ce n'est pas suffisant pour des applications industrielles ou médicales où il faut >95 % de fiabilité (chirurgie à distance).
- environnement simplifié
- le cube est un objet régulier, facile à suivre avec des caméras, on ne sait pas si la méthode fonctionne sur des objets mous, irréguliers ou sans marqueurs visuels, ce qui rend la généralisation limitée à d'autres formes (sphère, cylindre, outils...),
- Ressources de calcul énormes : l'entraînement a nécessité des milliers de CPU et plusieurs GPU pendant des jours → inaccessible pour la plupart des labos.