

# A simpler method to teach robots new skills

June 17 2024, by Ingrid Fadelli

---



A high-level overview of the Render and Diffuse process in the context of a putting the toilet seat down task. Rendered robot grippers are updated iteratively using a learnt denoising process until they represent actions needed to complete the task. Different colors represent different robot actions in the future. Credit: Vosylius et al.

While roboticists have introduced increasingly sophisticated systems over the past decades, teaching these systems to successfully and reliably tackle new tasks has often proved challenging. Part of this training entails mapping high-dimensional data, such as images collected by on-board RGB cameras, to goal-oriented robotic actions.

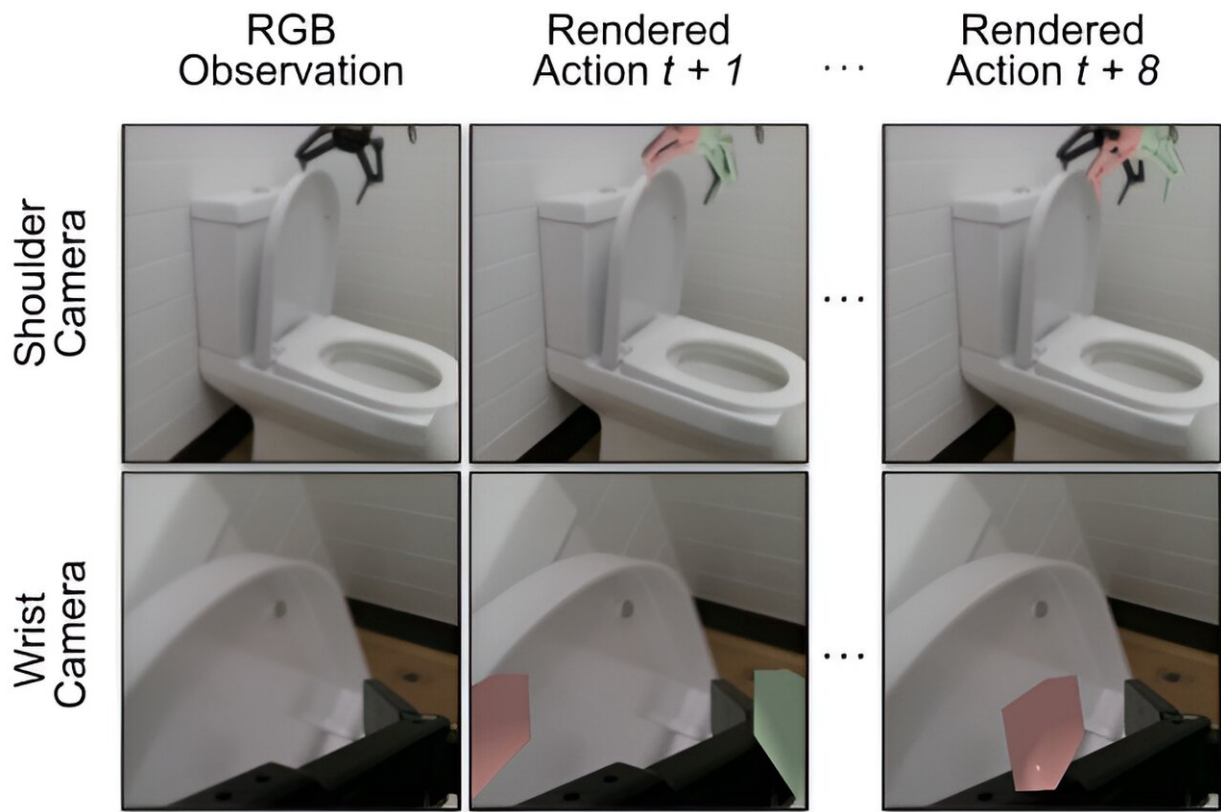
Researchers at Imperial College London and the Dyson Robot Learning Lab recently introduced Render and Diffuse (R&D), a method that unifies low-level robot actions and RGB images using virtual 3D renders of a robotic system. This method, introduced in a paper [published](#) on the *arXiv* preprint server, could ultimately facilitate the process of teaching robots new skills, reducing the vast amount of human demonstrations required by many existing approaches.

"Our recent paper was driven by the goal of enabling humans to teach robots new skills efficiently, without the need for extensive demonstrations," said Vitalis Vosylius, final year Ph.D. student at Imperial College London and lead author. "Existing techniques are data-intensive and struggle with spatial generalization, performing poorly when objects are positioned differently from the demonstrations. This is because predicting precise actions as a sequence of numbers from RGB images is extremely challenging when data is limited."

During an internship at Dyson Robot Learning, Vosylius worked on a project that culminated in the development of R&D. This project aimed to simplify the learning problem for robots, enabling them to more efficiently predict actions that will allow them to complete various tasks.

In contrast with most robotic systems, while learning new manual skills, humans do not perform extensive calculations to determine how much they should move their limbs. Instead, they typically try to imagine how their hands should move to tackle a [specific task](#) effectively.

"Our method, Render and Diffuse, allows robots to do something similar: 'imagine' their actions within the image using virtual renders of their own embodiment," Vosylius explained. "Representing robot actions and observations together as RGB images enables us to teach robots various tasks with fewer demonstrations and do so with improved spatial generalization capabilities."



Visualization of the RGB observations from two different cameras and rendered action representation at time steps  $t + 1$  and  $t + 8$  for the task of putting the toilet seat down. Credit: Vosylius et al.

For a robot to learn to complete a new task, it first needs to predict the

actions it should perform based on the images captured by its sensors. The R&D method essentially allows robots to learn this mapping between images and actions more efficiently.

"As hinted by its name, our method has two main components," Vosylius said. "First, we use virtual renders of the robot, allowing the robot to 'imagine' its actions in the same way it sees the environment. We do so by rendering the robot in the configuration it would end up in if it were to take certain actions.

"Second, we use a learned diffusion process that iteratively refines these imagined actions, ultimately resulting in a sequence of actions the robot needs to take to complete the task."

Using widely available 3D models of robots and rendering techniques, R&D can greatly simplify the acquisition of new skills while also significantly reducing training data requirements. The researchers evaluated their method in a series of simulations and found that it improved the generalization capabilities of robotic policies.

They also showcased their method's capabilities in effectively tackling six everyday tasks using a real robot. These tasks included putting down the toilet seat, sweeping a cupboard, opening a box, placing an apple in a drawer, and opening and closing a drawer.

"The fact that using virtual renders of the robot to represent its actions leads to increased data efficiency is really exciting," Vosylius said. "This means that by cleverly representing robot actions, we can significantly reduce the data required to train robots, ultimately reducing the labor-intensive need to collect extensive amounts of demonstrations."

In the future, the method introduced by this team of researchers could be tested further and applied to other tasks that robots could tackle. In

addition, the researchers' promising results could inspire the development of similar approaches to simplify the training of algorithms for robotics applications.

"The ability to represent robot actions within images opens exciting possibilities for future research," Vosylius added. "I am particularly excited about combining this approach with powerful image foundation models trained on massive internet data. This could allow robots to leverage the general knowledge captured by these models while still being able to reason about low-level [robot](#) actions."

**More information:** Vitalis Vosylius et al, Render and Diffuse: Aligning Image and Action Spaces for Diffusion-based Behaviour Cloning, *arXiv* (2024). [DOI: 10.48550/arxiv.2405.18196](https://doi.org/10.48550/arxiv.2405.18196)

© 2024 Science X Network

Citation: A simpler method to teach robots new skills (2024, June 17) retrieved 26 June 2024 from <https://techxplore.com/news/2024-06-simpler-method-robots-skills.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.