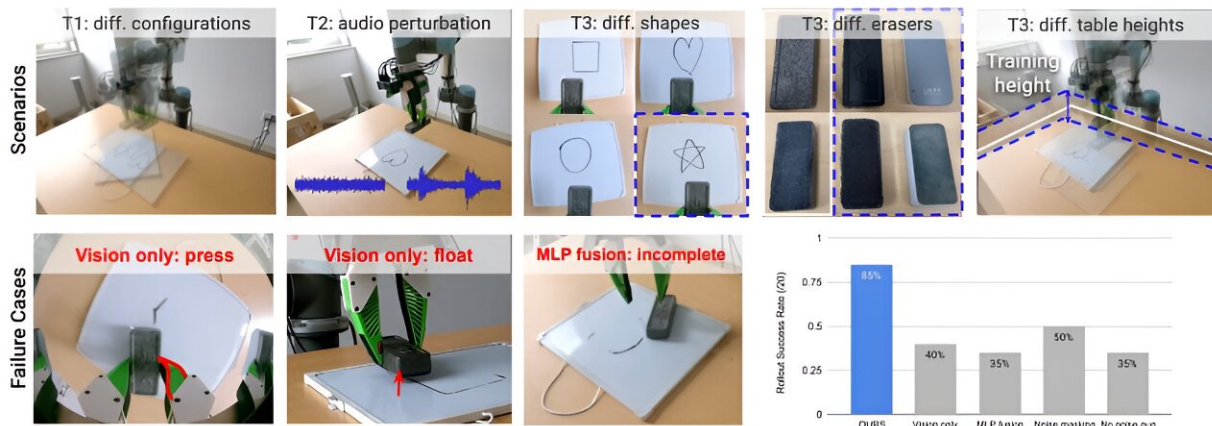


Adding audio data when training robots helps them do a better job

July 5 2024, by Bob Yirka



Wiping Evaluation. Up: Different test scenarios. Bottom: Typical failure cases and task success rate. [Vision only] policy often fails to maintain proper contact (e.g., either press too hard into the broad or float). [MLP fusion] policy often fails to fully wipe out the drawing and terminate early. Credit: *arXiv* (2024). DOI: 10.48550/arxiv.2406.19464

A combined team of roboticists from Stanford University and the Toyota Research Institute has found that adding audio data to visual data when training robots helps to improve their learning skills. The team has posted their [research](#) on the *arXiv* preprint server.

The researchers noted that virtually all training done with AI-based robots involves exposing them to a large amount of visual information,

while ignoring associated audio. They wondered if adding microphones to robots and allowing them to collect data regarding how something is supposed to sound as it is being done might help them learn a task better.

For example, if a [robot](#) is supposed to learn how to open a box of cereal and fill a bowl with it, it may be helpful to hear the sounds of a box being opened and the dryness of the cereal as it cascades down into a bowl. To find out, the team designed and carried out four robot-learning experiments.

The first experiment involved teaching a robot to turn over a bagel in a frying pan using a spatula. The second involved teaching a robot to use an eraser to erase an image on a white board. The third was pouring dice held in a cup into another cup and the fourth was to choose the correct size of tape from three available samples and to use it to tape a wire to a plastic strip.

All the experiments involved using the same robot equipped with a grasping claw. All of them were also done in two ways, using video only and using video and audio. The research team also varied teaching and performance factors such as table height, type of tape or the kind of image on the white board.

After running all their experiments, the researchers compared the results by judging how quickly and easily the robots were able to learn and carry out the tasks and also their accuracy. They found that adding audio significantly improved speed and accuracy with some tasks, but not others.

Adding audio to the task of pouring dice, for example, dramatically improved the robot's ability to figure out if there were any [dice](#) in the cup. It also helped the robot understand if it was exerting the right amount of pressure on the eraser, because of the unique sound that was

made. Adding sound did not help much, on the other hand, in determining if the bagel had been turned successfully or if all of an image had been successfully removed from a white board.

The team concludes by suggesting that their work shows that adding audio to [teaching](#) material for AI robots could provide better results for some applications.

More information: Zeyi Liu et al, ManiWAV: Learning Robot Manipulation from In-the-Wild Audio-Visual Data, *arXiv* (2024). [DOI: 10.48550/arxiv.2406.19464](https://doi.org/10.48550/arxiv.2406.19464)

Project page: mani-wav.github.io/

© 2024 Science X Network

Citation: Adding audio data when training robots helps them do a better job (2024, July 5) retrieved 13 July 2024 from <https://techxplore.com/news/2024-07-adding-audio-robots-job.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.