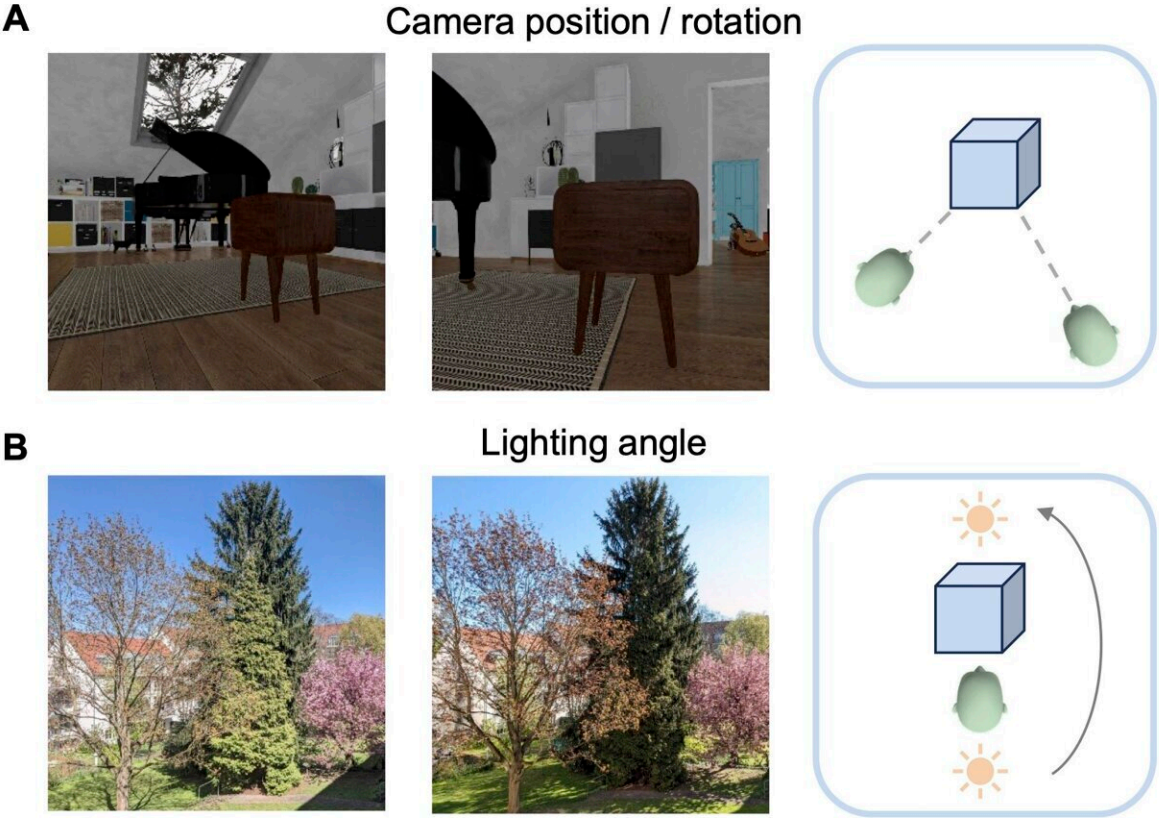


Children's visual experience may hold key to better computer vision training

May 31 2024, by Francisco Tutella and Mary Fetzer



The impact of position on the appearance, lighting, and camera distance/focal length of an image. Credit: Image provided by the research team, using tree photos (B) by Federico Adolfi and head bust photos (C) by James Z. Wang.

A novel, human-inspired approach to training artificial intelligence (AI) systems to identify objects and navigate their surroundings could set the

stage for the development of more advanced AI systems to explore extreme environments or distant worlds, according to research from an interdisciplinary team at Penn State.

In the first two years of life, children experience a somewhat narrow set of objects and faces, but with many different viewpoints and under varying [lighting conditions](#). Inspired by this developmental insight, the researchers introduced a new machine learning approach that uses information about spatial position to train AI visual systems more efficiently.

They found that AI models trained on the new method outperformed base models by up to 14.99%. They reported their findings in the May issue of the journal [Patterns](#).

"Current approaches in AI use massive sets of randomly shuffled photographs from the internet for training. In contrast, our strategy is informed by [developmental psychology](#), which studies how children perceive the world," said Lizhen Zhu, the lead author and doctoral candidate in the College of Information Sciences and Technology at Penn State.

The researchers developed a new contrastive learning algorithm, which is a type of self-supervised learning method in which an AI system learns to detect visual patterns to identify when two images are derivations of the same base image, resulting in a positive pair. These algorithms, however, often treat images of the same object taken from different perspectives as separate entities rather than as positive pairs.

Taking into account environmental data, including location, allows the AI system to overcome these challenges and detect positive pairs regardless of changes in camera position or rotation, lighting angle or condition and [focal length](#), or zoom, according to the researchers.

"We hypothesize that infants' visual learning depends on location perception. In order to generate an egocentric dataset with spatiotemporal information, we set up virtual environments in the ThreeDWorld platform, which is a high-fidelity, interactive, 3D physical simulation environment. This allowed us to manipulate and measure the location of viewing cameras as if a child was walking through a house," Zhu added.

The scientists created three simulation environments—House14K, House100K and Apartment14K, with "14K" and "100K" referring to the approximate number of sample images taken in each environment. Then they ran base contrastive learning models and models with the new algorithm through the simulations three times to see how well each classified images. The team found that models trained on their algorithm outperformed the base models on a variety of tasks.

For example, on a task of recognizing the room in the virtual apartment, the augmented model performed on average at 99.35%, a 14.99% improvement over the base model. These new datasets are available for other scientists to use in training through www.child-view.com.

"It's always hard for models to learn in a new environment with a small amount of data. Our work represents one of the first attempts at more energy-efficient and flexible AI training using visual content," said James Wang, distinguished professor of information sciences and technology and advisor of Zhu.

The research has implications for the future development of advanced AI systems meant to navigate and learn from new environments, according to the scientists.

"This approach would be particularly beneficial in situations where a team of autonomous robots with [limited resources](#) needs to learn how to

navigate in a completely unfamiliar environment," Wang said. "To pave the way for future applications, we plan to refine our model to better leverage spatial information and incorporate more diverse environments."

Collaborators from Penn State's Department of Psychology and Department of Computer Science and Engineering also contributed to this study.

More information: Lizhen Zhu et al, Incorporating simulated spatial context information improves the effectiveness of contrastive learning models, *Patterns* (2024). [DOI: 10.1016/j.patter.2024.100964](https://doi.org/10.1016/j.patter.2024.100964)

Provided by Pennsylvania State University

Citation: Children's visual experience may hold key to better computer vision training (2024, May 31) retrieved 20 June 2024 from <https://techxplore.com/news/2024-05-children-visual-key-vision.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.