Towards Generalist Robot Learning from Internet Video: A Survey (Abstract Reprint)

Robert McCarthy[1], **Daniel C.H. Tan**[1], **Dominik Schmidt**[2], **Fernando Acero**[1], **Nathan Herr**[1], **Yilun Du**[3], **Thomas G. Thuruthel**[1], **Zhibin Li**[1]

[1]University College London, United Kingdom
[2]Weco AI, United Kingdom
[3]Massachusetts Institute of Technology, United States of America

Abstract Reprint. This is an abstract reprint of the journal article by McCarthy, Tan, Schmidt, Acero, Herr, Du, Thuruthel, and Li (2025).

Abstract

Scaling deep learning to massive and diverse internet data has driven remarkable breakthroughs in domains such as video generation and natural language processing. Robot learning, however, has thus far failed to replicate this success and remains constrained by a scarcity of available data. Learning from Videos (LfV) methods aim to address this data bottleneck by augmenting traditional robot data with large-scale internet video. This video data provides foundational information regarding physical dynamics, behaviours, and tasks, and can be highly informative for general-purpose robots. This survey systematically examines the emerging field of LfV. We first outline essential concepts, including detailing fundamental LfV challenges such as distribution shift and missing action labels in video data. Next, we comprehensively review current methods for extracting knowledge from large-scale internet video, overcoming LfV challenges, and improving robot learning through video-informed training. The survey concludes with a critical discussion of future opportunities. Here, we emphasize the need for scalable foundation model approaches that can leverage the full range of available internet video and enhance the learning of robot policies and dynamics models. Overall, the survey aims to inform and catalyse future LfV research, driving progress towards general-purpose robots.

References

McCarthy, R.; Tan, D. C.; Schmidt, D.; Acero, F.; Herr, N.; Du, Y.; Thuruthel, T. G.; and Li, Z. 2025. Towards Generalist Robot Learning from Internet Video: A Survey. *Journal of Artificial Intelligence Research*, 83: 12:1–12:48.