Learning Generalizable Robotic Reward Functions from “In-The-Wild” Human Videos

Annie S. Chen, Suraj Nair, Chelsea Finn
Stanford University

Contact: asc8@stanford.edu
Project Website + Code: https://sites.google.com/view/dvd-human-videos
Want: Robots that can generalize broadly across tasks, objects, and environments

Tasks

Objects

Environments
Want: Robots that can generalize broadly across tasks, objects, and environments

Requires metric of task success or reward signal
Want: Robots that can generalize broadly across tasks, objects, and environments

Requires metric of task success or reward signal

Key: Need reward functions to generalize broadly
Key: Need reward functions that generalize broadly

One promising direction: Learning from large and diverse datasets

Collecting lots of diverse robotic data is difficult
What about “In the wild” human videos?
What about “In the wild” human videos?

Challenges:

• Domain shift

• No shared action space

• “In-the-wild” human videos are noisy

How do we leverage such data for a reward function?

**Key idea:** *Domain-agnostic Video Discriminator (DVD)*

Input two videos $\rightarrow$ output whether doing the same task
How do we leverage such data for a reward function?

**Key idea:** Domain-agnostic Video Discriminator (DVD)
Input two videos → output whether doing the same task

![Label = 1](image1)
“Closing Something” = “Closing Something”

![Label = 0](image2)
“Moving Something Away” ≠ “Closing Something”

![Label = 1](image3)
“Closing Something” = “Closing Something”
Can DVD generalize to new environments using human videos?
Can DVD generalize to new environments using human videos?

Human Video

Planning w.r.t. DVD reward

Task Completion
Can DVD generalize to new environments using human videos?
Can DVD generalize to **new environments** using human videos?

**Human Video**

**Planning w.r.t. DVD reward**

**Task Completion**
Can DVD generalize to new environments using human videos?

Training w/ human videos improves environment generalization by over 20%
Does DVD work on a real robot?

Training
- Closing something
- Pushing something from left to right

Testing
- Closing the door
- Pushing the tissue box to the left
Takeaways & Limitations

- DVD is a simple, scalable approach to task specification w/ videos
- By leveraging diverse human data in the robot learning process, DVD yields
  - improved environment generalization
  - Improved task generalization
Takeaways & Limitations

• DVD is a simple, scalable approach to task specification w/ videos
• By leveraging diverse human data in the robot learning process, DVD yields
  • improved environment generalization
  • Improved task generalization
• Limitations
  • DVD focuses only on reward functions → can we get a generalizable policy directly?
  • Extend to more fine-grained tasks
  • Reward is non-Markovian