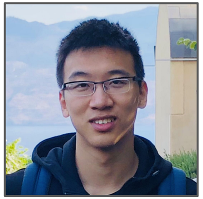
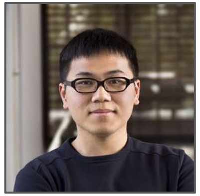
BabyWalk: Going Farther in Vision-and-Language Navigation by Taking Baby Steps

(Paper Id:158)



Wang Zhu*
SFU

Vihan Jain
Google



Hexiang Hu*
USC

Eugene Ie
Google

Jiacheng Chen
USC

Fei Sha
Google

Zhiwei Deng
Princeton

(*: authors contributed equally)

Embodied AI: a motivating application

Example application:

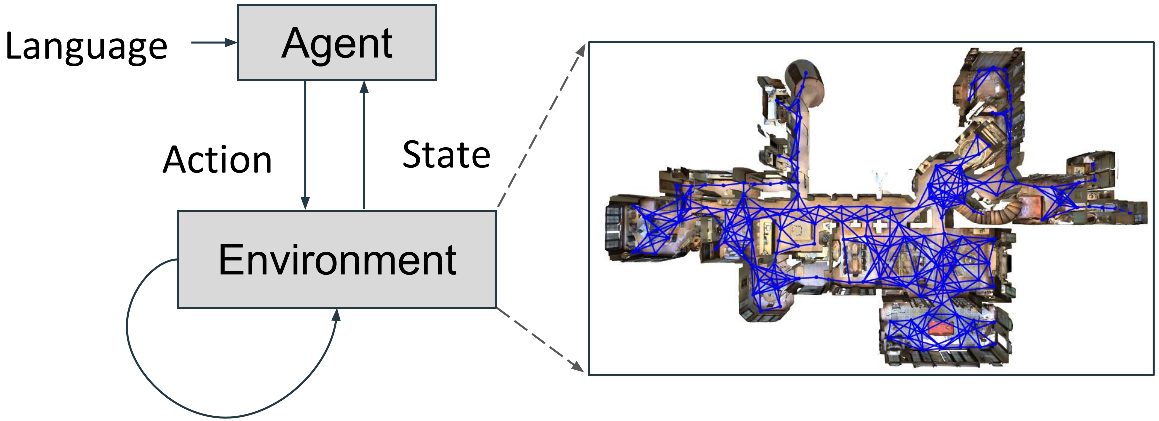
Household robots to understand human *language* and execute accordingly, in an unconstrained setting.

Fig. Example of Room2Room

Demo credit: Anderson et. al. CVPR 2018

Vision and Language Navigation **(VLN)**

(Anderson et. al. CVPR 2018)



In VLN, an agent follows human annotated language instructions in a photo-realistic simulator .

VLN interested the community, and inspires a large body of follow-up works.

[Fried et. al. NeurIPS 2019, Wang et. al. CVPR 2019, Tan et. al. NAACL 2019, Jain et. al. ACL 2019, etc..]

Challenges

How much data to train models?

Need a large amount of parallel data.

Supplement with high-fidelity simulation.

How well models generalize?

Variability across perception and environments, & language instructions.

Discrepancy between simulation and real-physical world.

Outline

Generalization

BabyWalk

Conclusion

Generalization

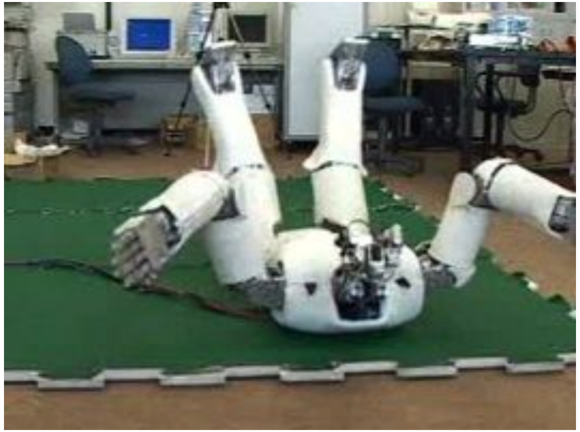
Key observations

- Learn skills in small space (home, nursery) with simple language instructions
  - Transferable to bigger space
  - Transferable to complex language instructions

Key hypothesis

- Follow “baby steps”
  - Break down long navigation tasks to shorter ones
  - Follow instructions by small pieces

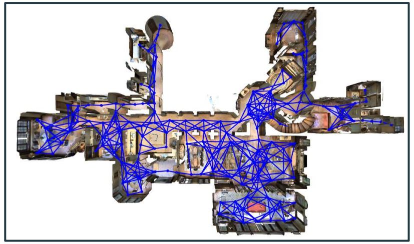
But can robot do as well?



VLN Datasets

Make navigation tasks longer.

Original Room2Room
(Anderson et. al. CVPR 2018)



Task Horizon	
Avg Words	29.4
Avg Path Len	6.0

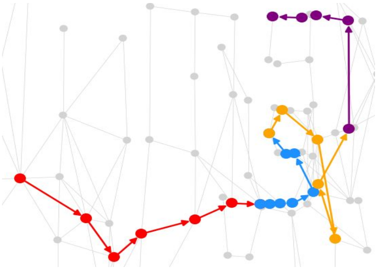
Room4Room
(Jain et. al. ACL 2019)

Task Horizon	
Avg Words	58.4
Avg Path Len	11.1

Room6Room
(Ours)

Task Horizon	
Avg Words	**91.2**
Avg Path Len	**16.5**

Room8Room
(Ours)

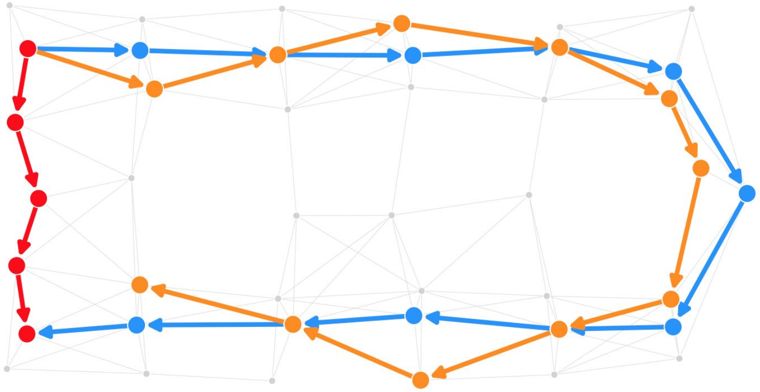


Task Horizon	
Avg Words	**121.6**
Avg Path Len	**21.6**

Models trained on R2R do not follow instruction!

Previous models trained on R2R

- Cares only about **reaching the goal**
- Take shortcut (**Red path)**
- Ignore instructions (**Blue Path**)
- Penalize instruction-observing (**Orange path**)

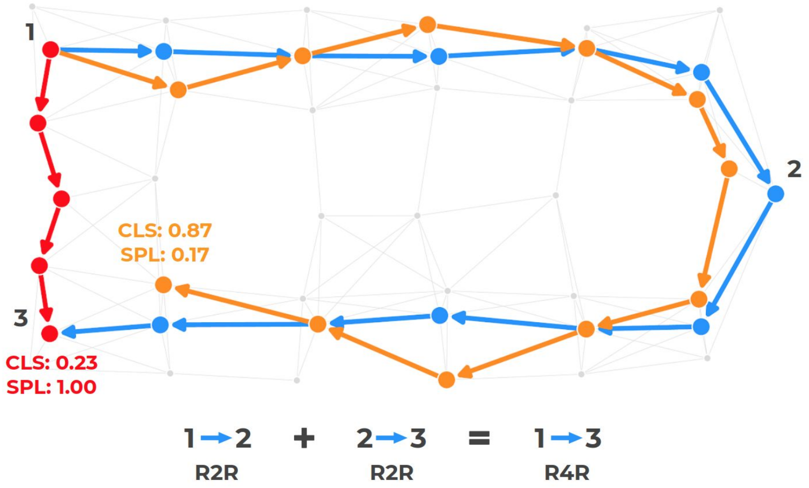


(Jain et. al. ACL 2019)

Existing approaches for better generalization

Train on longer horizon navigation tasks

Room4Room *(Jain et. al. ACL 2019)* was created partially for that purpose.



Optimizing the right reward

RL with FIDELITY reward

Better metric

Favor instruction-observing paths

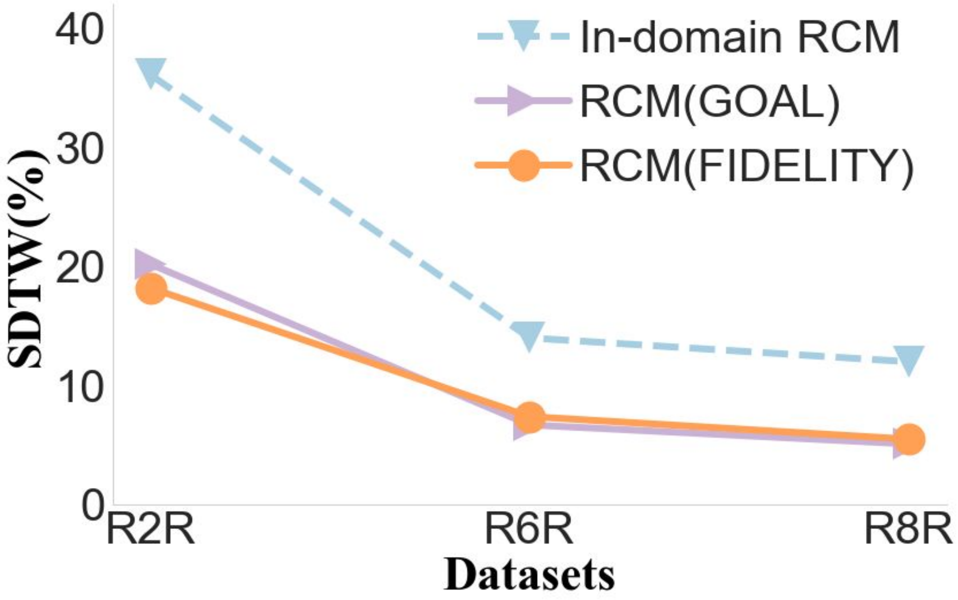
Penalize pure short-cuts for goal reaching

Perhaps models trained on R4R generalize well?

Trained on	VLN Data w/ **a Predetermined** Horizon Length (Ex: the seen split in R4R)
Traditional Evaluation	VLN Task w/ **the Given** Horizon Length (Ex: unseen R4R)
Transfer Evaluation **(Our Proposal)**	VLN Task w/ **the Unseen** Horizon Lengths

No, training on R4R do not generalize well

R4R trained model performs poorly on R2R, R6R, R8R

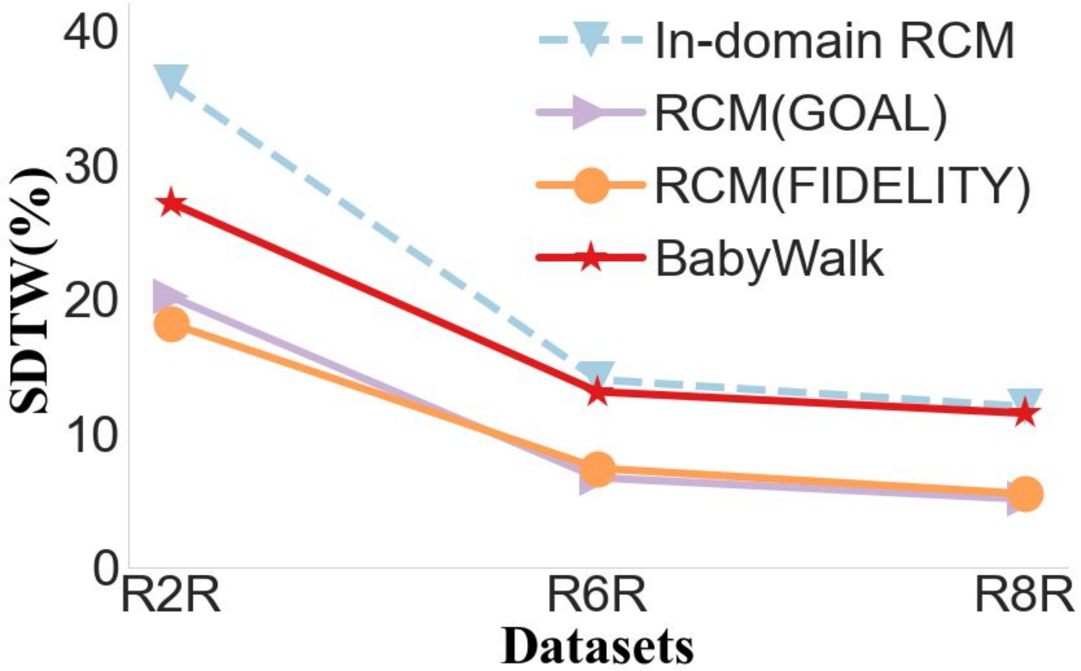


(Success by Dynamic Time Warping (**SDTW**) is a recently proposed metric, which *aligns best with human judgement*.)

How do we make them generalize well?

Babywalk (our approach) generalizes!

As a final result, babywalk trained on R4R generalize significantly better



Outline

Generalization

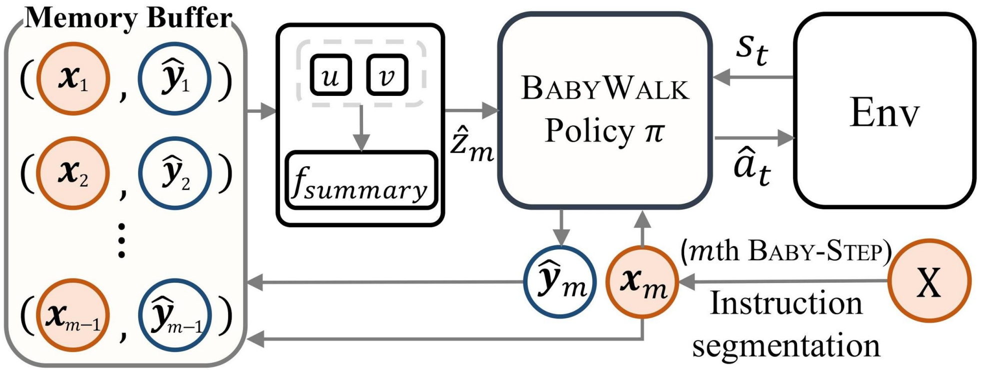
BabyWalk

Conclusion

BabyWalk: Main ideas

- Subtask (BabyStep) based Navigation Agent (BabyWalk)
  - Babywalk is associated with external memory of sub-tasks history
- BabyStep Imitation Learning
  - Decompose long navigation tasks into short BabySteps
  - Imitation learning to follow BabySteps
- Curriculum Reinforcement Learning
  - Reinforcement learning to improve Babywalk on longer task horizons
  - Gradually Increase difficulty (ie, path lengths to execute)

BabyWalk: Overall Navigation Agent



The BabyWalk agent predict the ***t*-th** action of ***m*-th** task depends on:

Input

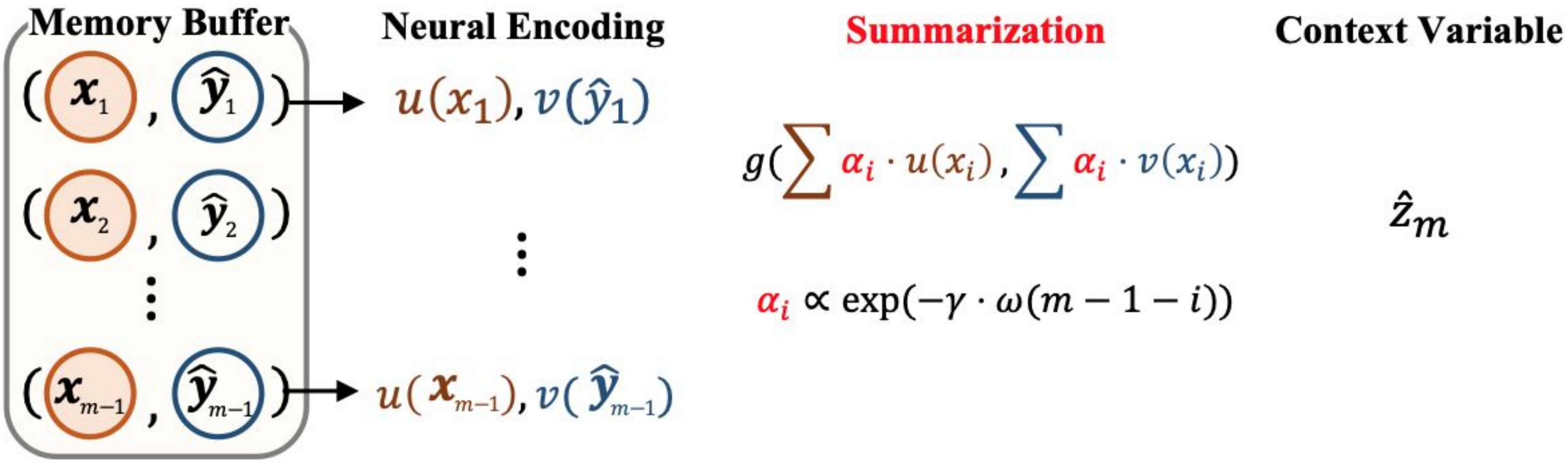
History Context	$\hat{\boldsymbol{z}}_m$
Instruction Vector	$\boldsymbol{x}_m$
State Feature	$\boldsymbol{s}_t$

Output

Action (index)	$\hat{\boldsymbol{a}}_t$
Trajectory	$\hat{\boldsymbol{y}}_m$

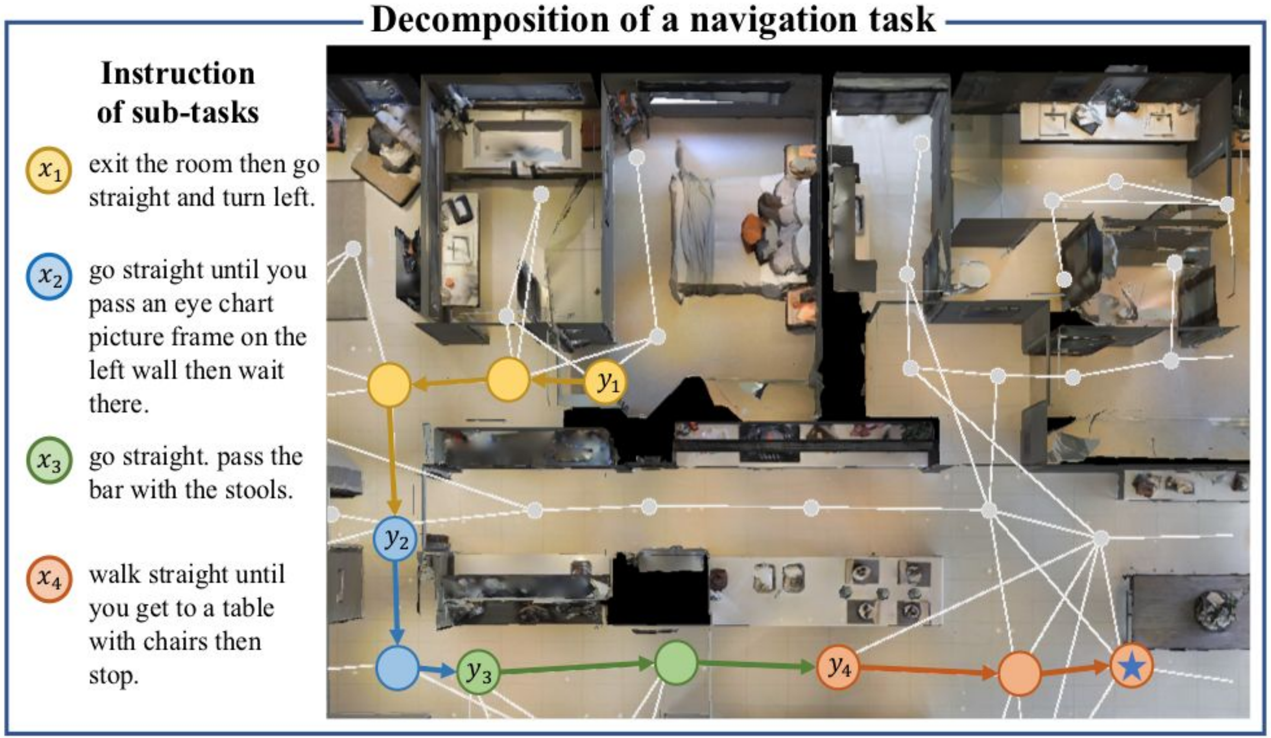
BabyWalk: summarize history as context variable

We use an **external memories** to store the history, and **summarize** them into a context variable using **an temporally decaying weighting**:



Stage 1: Baby-step imitation learning

Instruction segmentation. Template based sentence segmentation.



We use a set of heuristic rules to identify all the executable **baby-step instructions** from a long instruction.

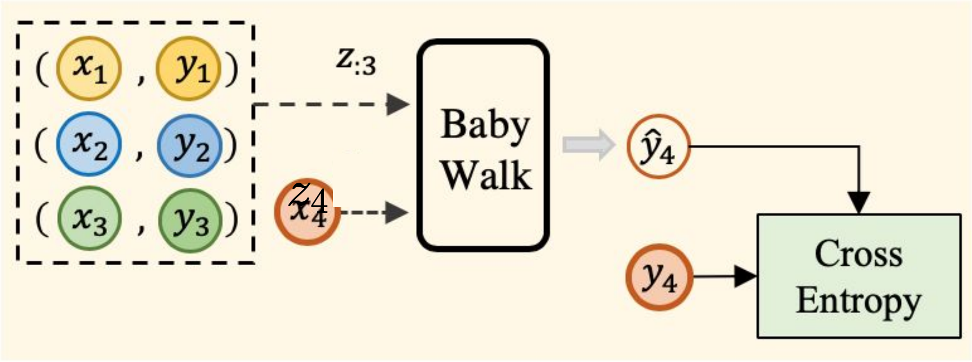
(details in the paper)

Stage 1: Baby-step imitation learning

Data Alignment. Align trajectories to baby-step instructions via dynamic programming with a weakly supervised visual classifier (without extra annotation).

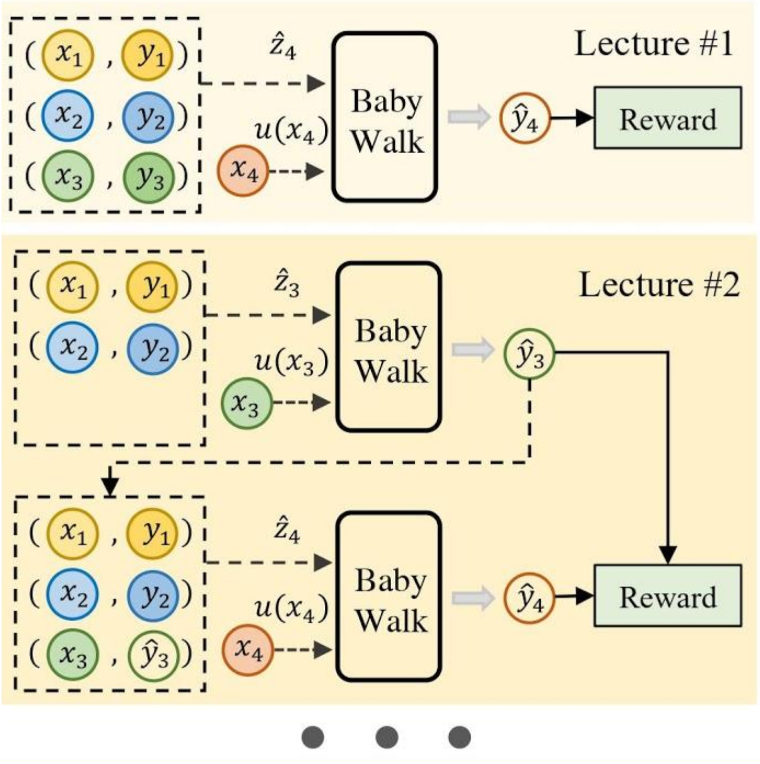
Stage 1: Baby-step imitation learning

Imitation learning. Given the true history context variable $\boldsymbol{z}_m$, and one baby-step instruction $\boldsymbol{x}_m$, minimize imitation loss with aligned baby-step trajectory.



Stage 2: Curriculum reinforcement learning

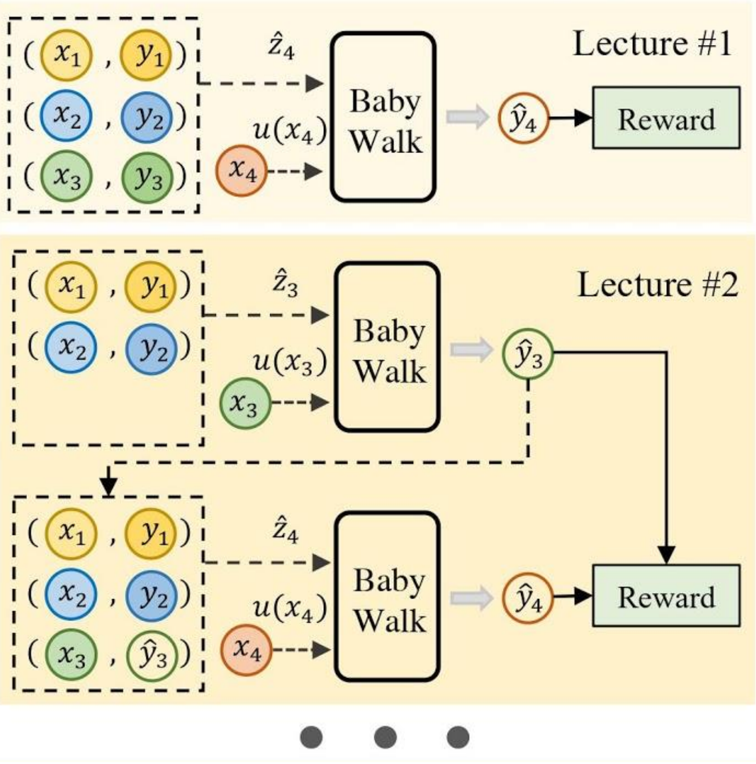
Intuition. Make an agent learning to gradually navigate with longer task-horizon.



Stage 2: Curriculum reinforcement learning

Intuition. Make an agent learning to gradually navigate with longer task-horizon.

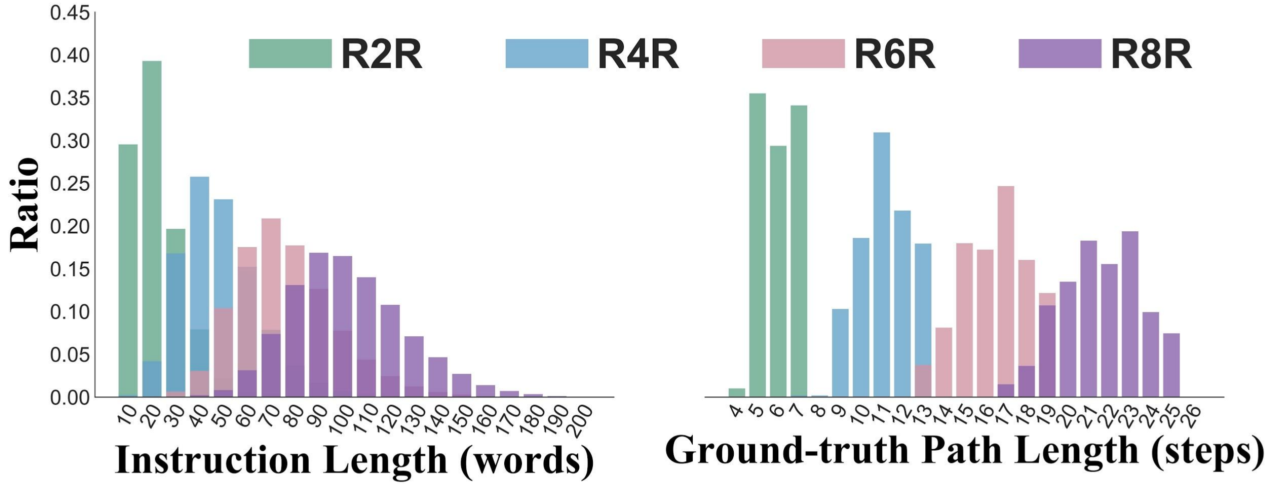
Curriculum Design. Suppose that there are **M** steps in total, **at the lecture 2**, an babywalk agent is given **(M - 2)** steps of "ground-truth" history and asked to learn executing **2** steps of baby-step instruction (with REINFORCE).



Datasets and Setups

Datasets

- Training Set:
  - *R4R training dataset on **61 Seen Scenes***
- Evaluation Set:
  - *R2R, R4R, R6R, R8R datasets on **11 Unseen Scenes***



Datasets and setups

Evaluation Metrics

- Success Rate (**SR**)
- Coverage by Length Score (**CLS**) [Jain et. al. 2019]
  - Treat the generated path and ground-truth path as two sets of nodes and evaluates the **Node Coverage**, weighted by a Path Length Score**.**
- Success weighted Dynamic Time Warping (**SDTW**) [Ilharco et. al. 2019]
  - Treat the generated path and ground-truth path as two **Time Series** to evaluate their similarity, weighted by the Success Rate. *Best correlates to human.*

In-Domain results

- Evaluated **in-domain**, babywalk **works the best in instruction following**

Setting	In-domain			Generalization to other datasets											
	R4R → R4R			R4R → R2R			R4R → R6R			R4R → R8R			Average		
Metrics	SR↑	CLS↑	SDTW↑	SR↑	CLS↑	SDTW↑	SR↑	CLS↑	SDTW↑	SR↑	CLS↑	SDTW↑	SR↑	CLS↑	SDTW↑
SEQ2SEQ	25.7	20.7	9.0												
SF+	24.9	23.6	9.2												
RCM(GOAL)+	28.7	36.3	13.2												
RCM(FIDELITY)+	24.7	39.2	13.7												
REGRETFUL+*	30.1	34.1	13.5												
FAST+*	**36.2**	34.0	15.5												
BABYWALK	29.6	47.8	**18.1**												
BABYWALK+	27.3	**49.4**	17.3												

(+: pre-trained with data augmentation, *: reimplemented or adapted from the open sourced code release)

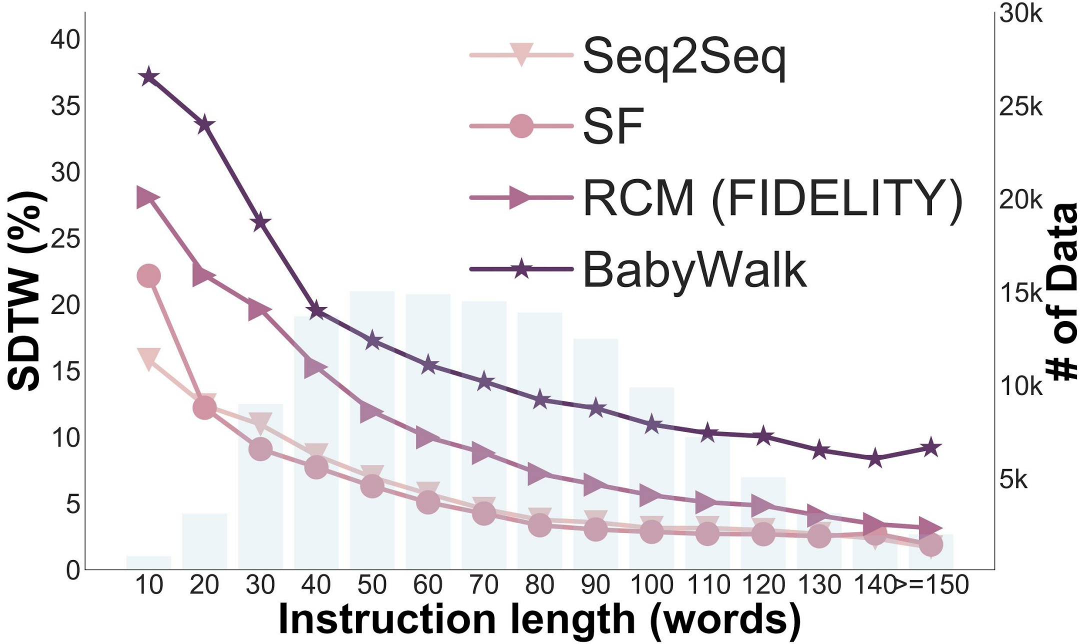
Cross dataset (horizon) generalization results

- Acrossing different horizons, **babywalk consistently wins in all metrics**

Setting	In-domain R4R → R4R			Generalization to other datasets R4R → R2R			R4R → R6R			R4R → R8R			Average		
Metrics	SR↑	CLS↑	SDTW↑	SR↑	CLS↑	SDTW↑	SR↑	CLS↑	SDTW↑	SR↑	CLS↑	SDTW↑	SR↑	CLS↑	SDTW↑
SEQ2SEQ	25.7	20.7	9.0	16.3	27.1	10.6	14.4	17.7	4.6	20.7	15.0	4.7	17.1	19.9	6.6
SF+	24.9	23.6	9.2	22.5	29.5	14.8	15.5	20.4	5.2	21.6	17.2	5.0	19.9	22.4	8.3
RCM(GOAL)+	28.7	36.3	13.2	25.9	44.2	20.2	19.3	31.8	7.3	22.8	27.6	5.1	22.7	34.5	10.9
RCM(FIDELITY)+	24.7	39.2	13.7	29.1	34.3	18.3	20.5	38.3	7.9	20.9	34.6	6.1	23.5	35.7	10.8
REGRETFUL+*	30.1	34.1	13.5	22.8	32.6	13.4	18.0	31.7	7.5	18.7	29.3	5.6	19.8	31.2	8.8
FAST+*	**36.2**	34.0	15.5	25.1	33.9	14.2	22.1	31.5	7.7	**27.7**	29.6	6.3	25.0	31.7	9.4
BABYWALK	29.6	47.8	**18.1**	**35.2**	48.5	27.2	**26.4**	44.9	13.1	26.3	44.7	**11.5**	**29.3**	46.0	17.3
BABYWALK+	27.3	**49.4**	17.3	34.1	**50.4**	**27.8**	25.5	**47.2**	**13.6**	23.1	**46.0**	11.1	27.6	**47.9**	**17.5**

(+: pre-trained with data augmentation, *: reimplemented or adapted from the open sourced code release)

Babywalk works better especially w/ long instructions



- Babywalk works better than previous methods, particularly on long instructions
- As the total length of instruction grows, the performance of Babywalk decreases slower

How useful are various learning strategies?

(**<u>Average performances</u>** on R2R ~ R8R)

Data Splits	Overall SR↑	Overall CLS↑	Overall SDTW↑
BabyWalk (IL+RL)	25.0	44.2	14.0
BabyWalk (IL)	24.3	26.3	10.4
$+1^{st}$ CRL	24.1	43.5	13.6
$+2^{nd}$ CRL	26.3	44.3	14.9
$+3^{rd}$ CRL	27.0	45.9	16.5
$+4^{th}$ CRL	27.5	48.3	17.5

- Babywalk w/ **Curriculum RL** improves over its **IL** and **IL + vanilla RL** variants significantly
- Babywalk w/ Curriculum RL improves as the number of lectures increases

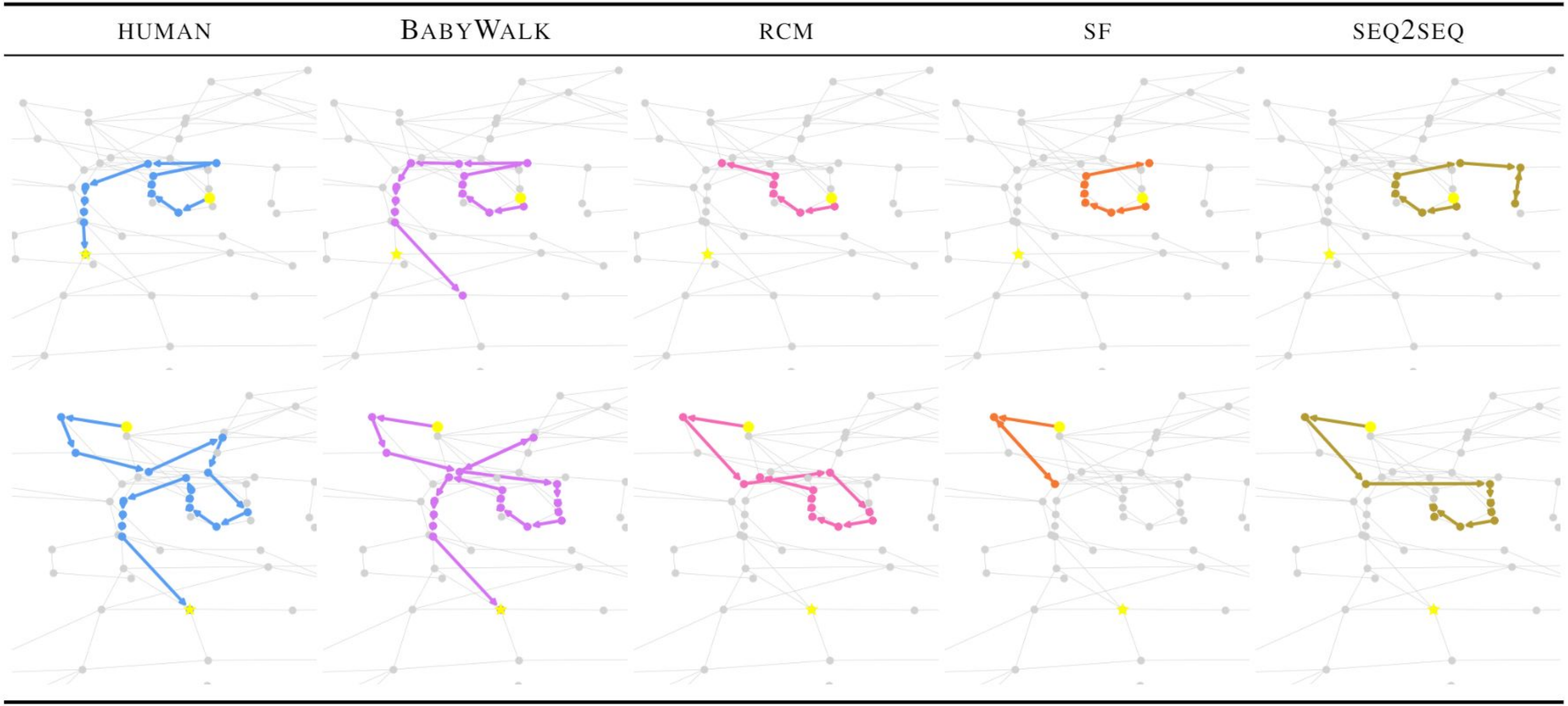
How useful is the summary of the histories?

Setting	R4R $\rightarrow$ R4R			R4R $\rightarrow$ others		
Metrics	SR$\uparrow$	CLS$\uparrow$	SDTW $\uparrow$	SR$\uparrow$	CLS$\uparrow$	SDTW $\uparrow$
$f_{\text{SUMMARY}} =$						
NULL	18.9	43.1	9.9	17.1	42.3	9.6
LSTM$(\cdot)$	25.8	44.0	14.4	25.7	42.1	14.3
$f_{\text{SUMMARY}} = \sum_{i=1}^{m-1} \alpha_i \cdot (\cdot)$, *i.e.*, eqs. (2,3)						
$\gamma = 5$	27.5	46.8	15.8	26.7	44.4	14.9
$\gamma = 0.5$	27.3	**49.4**	**17.3**	**27.6**	**47.9**	**17.5**
$\gamma = 0.05$	**27.5**	47.7	16.2	26.0	45.5	15.2
$\gamma = 0$	26.1	46.6	15.1	25.1	44.3	14.4

- The proposed history summary mechanism outperforms the various baselines, i.e. averaging and LSTM, by a margin.

Qualitative visualization of the path babywalk takes

- Qualitatively, **babywalk** generates trajectory that is more human-like.



Revisit Room2Room

Our Model (**BabyWalk)** trained on **Room2Room** can **transfer comparably well** to counterpart trained on **Room4Room**.

Eval	$\rightarrow$ R6R			$\rightarrow$ R8R		
Training	SR↑	CLS↑	SDTW↑	SR↑	CLS↑	SDTW↑
R2R	21.7	**49.0**	11.2	20.7	**48.7**	9.8
R4R	**25.5**	47.2	**13.6**	**23.1**	46.0	**11.1**

Outline

Generalization

BabyWalk

Conclusion

Summary

- **Take-home message**
  - Transfer is crucial for agents on “small” datasets with limited variability
  - Evaluating the generalizations across different task horizons helps measuring such transfer.
  - **Subtask-based IL** followed by **curriculum RL** is a promising learning approach to this purpose.
- **Future directions**
  - Better subtask segmentation
  - More Real-world scenarios
    - More diverse visual environments
    - More linguistic variabilities in instructions

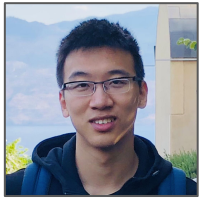
Thank you for watching!

For more details, please visit our live Q&A session at:

1. Monday July 6, 2020 Session 4B - 18:00 UTC+0 (11:00 PDT)

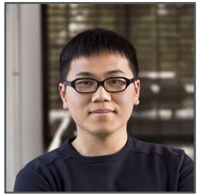
2. Monday July 6, 2020 Session 5B - 21:00 UTC+0 (14:00 PDT)

Our **code** is publically available at https://github.com/Sha-Lab/babywalk



Wang Zhu*
SFU

Vihan Jain
Google



Hexiang Hu*
USC

Eugene Ie
Google

Jiacheng Chen
USC

Fei Sha
Google

Zhiwei Deng
Princeton

(*: authors contributed equally)