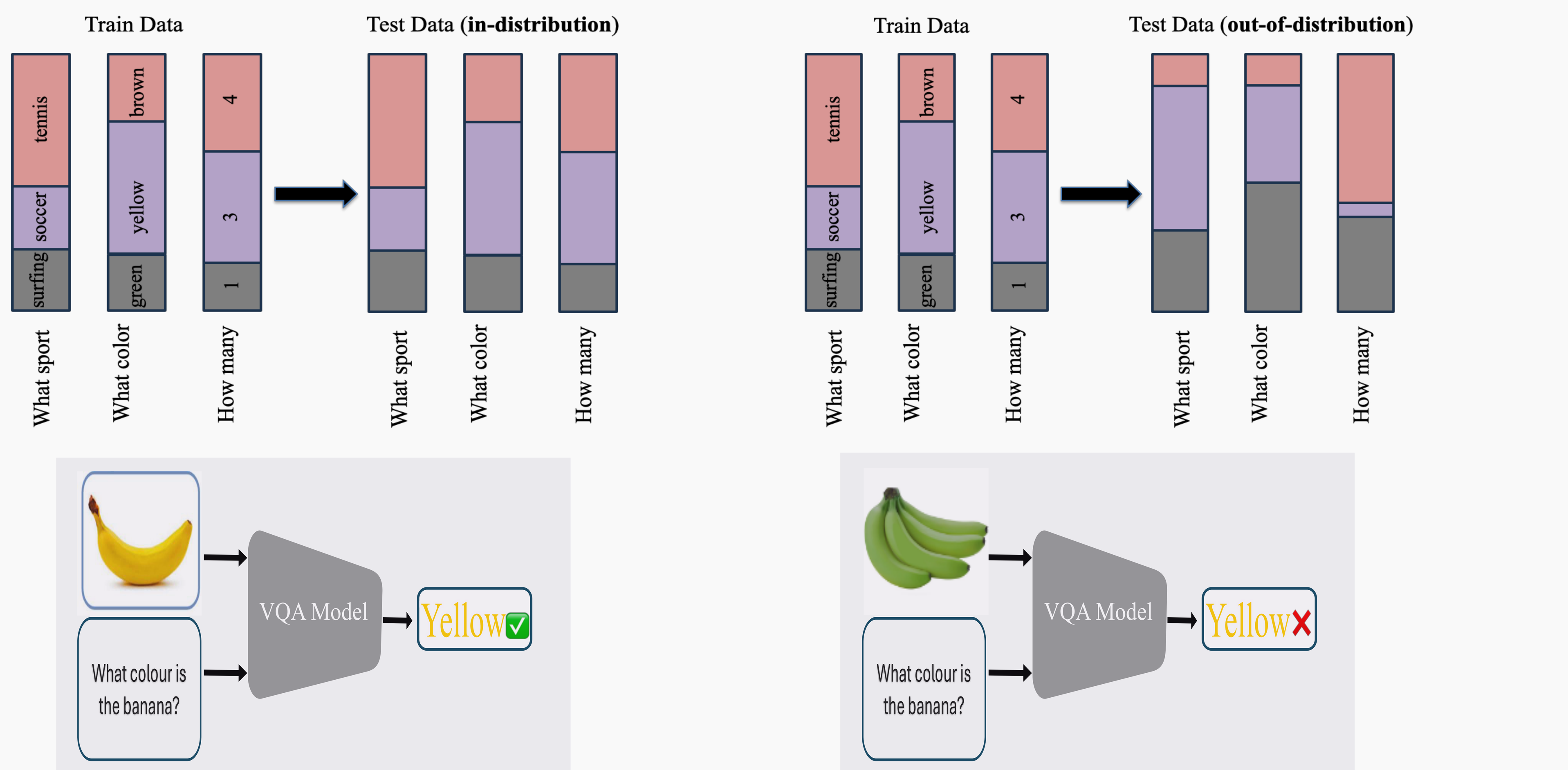
Task Progressive Curriculum Learning for Robust Visual Question Answering

Ahmed Akl, Abdelwahed Khamis, Zhe Wang
Ali Cheraghian, Sara Khalifa, Kewen Wang

BMVC 2025

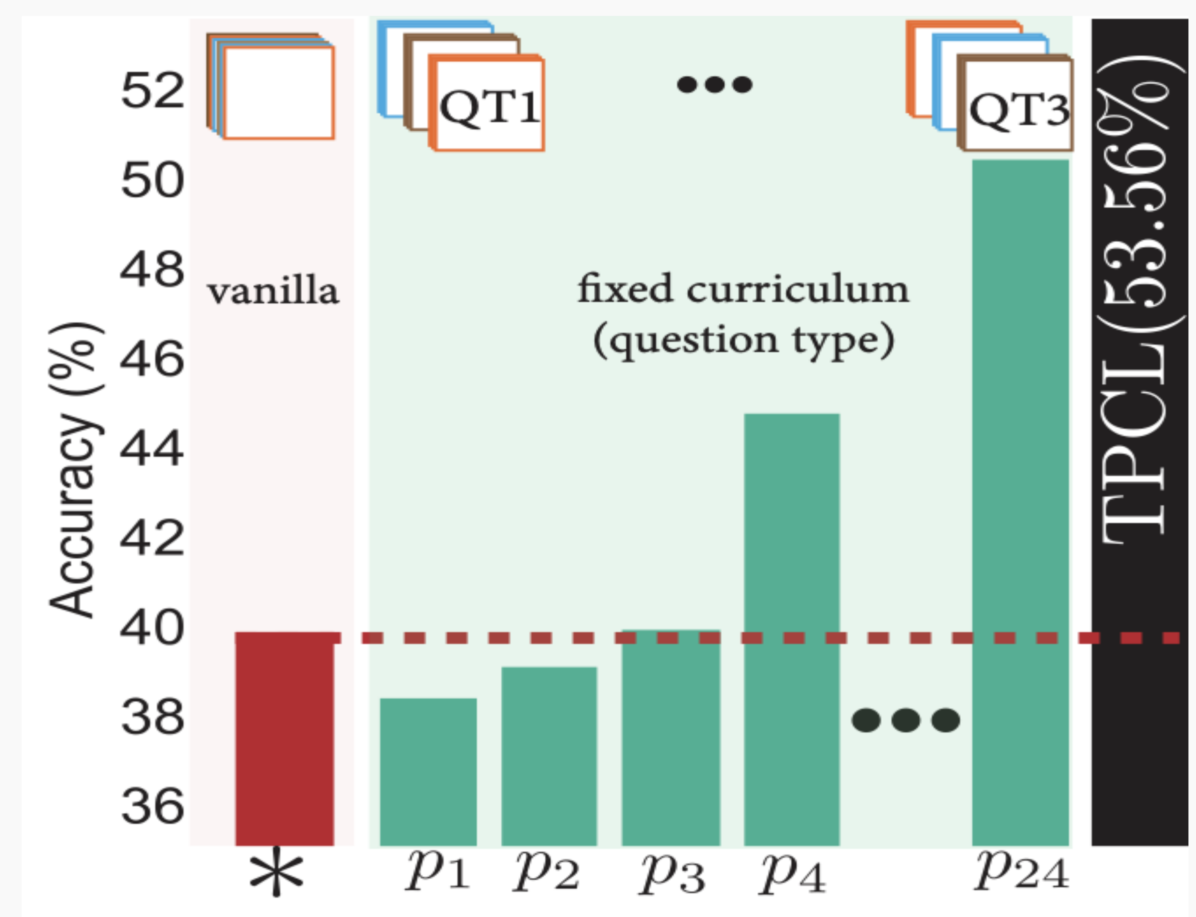
Problem Definition

Visual Question Answering are notorious for their poor performance in out of distribution.



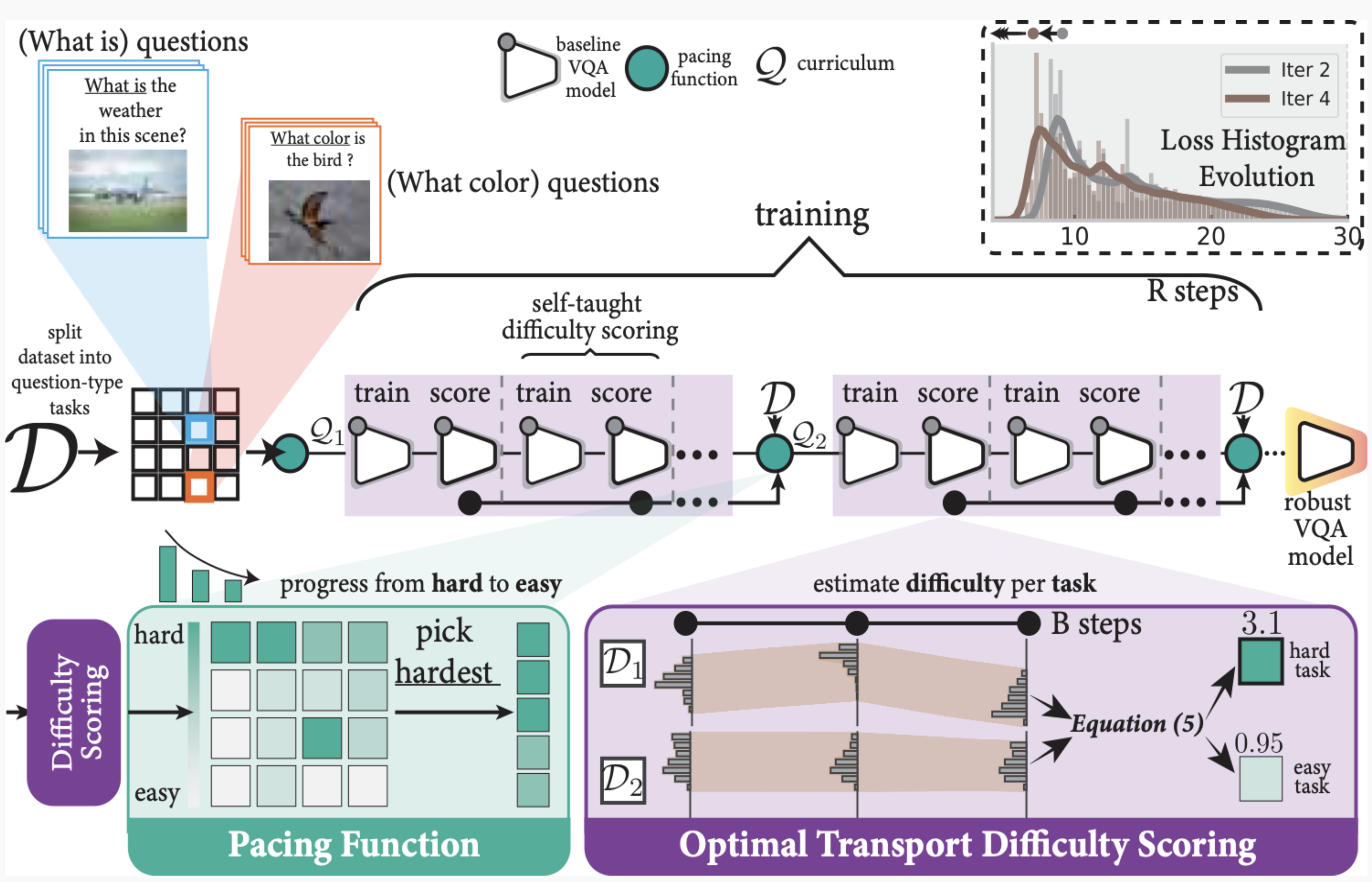
Motivation

Inspired by Psycholinguistics, we design a training strategy that enhance the VQA model performance in OOD.



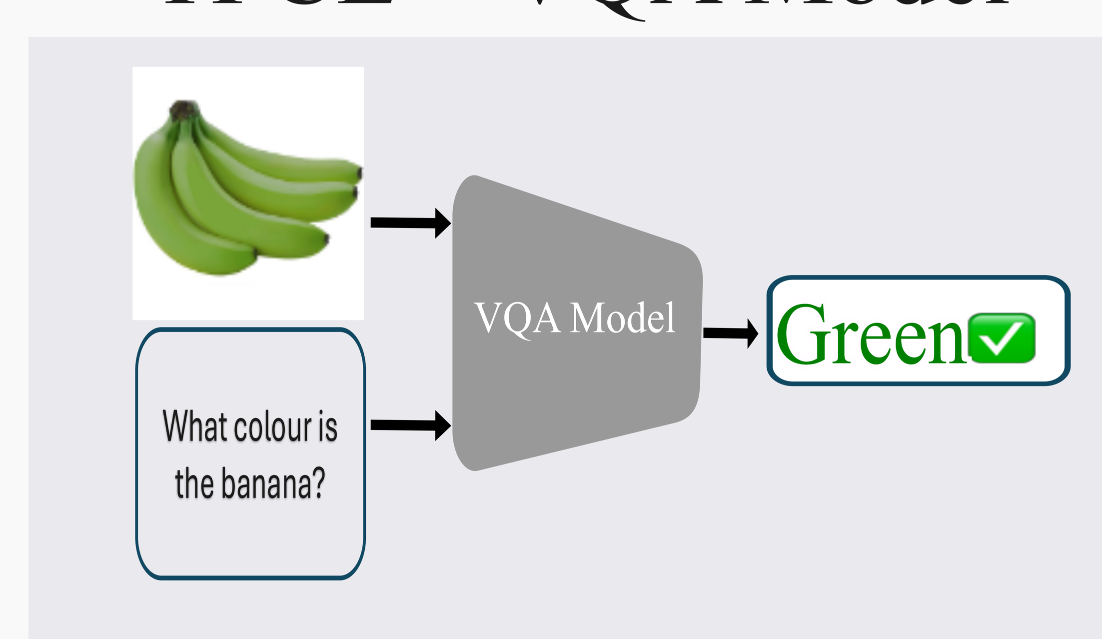
Proposed Method

Our approach builds a robust VQA model, we design a task-based curriculum that can be used to train a baseline back-bone and enable out-of-distribution generalisation.



Robust VQA Model

TPCL + VQA Model

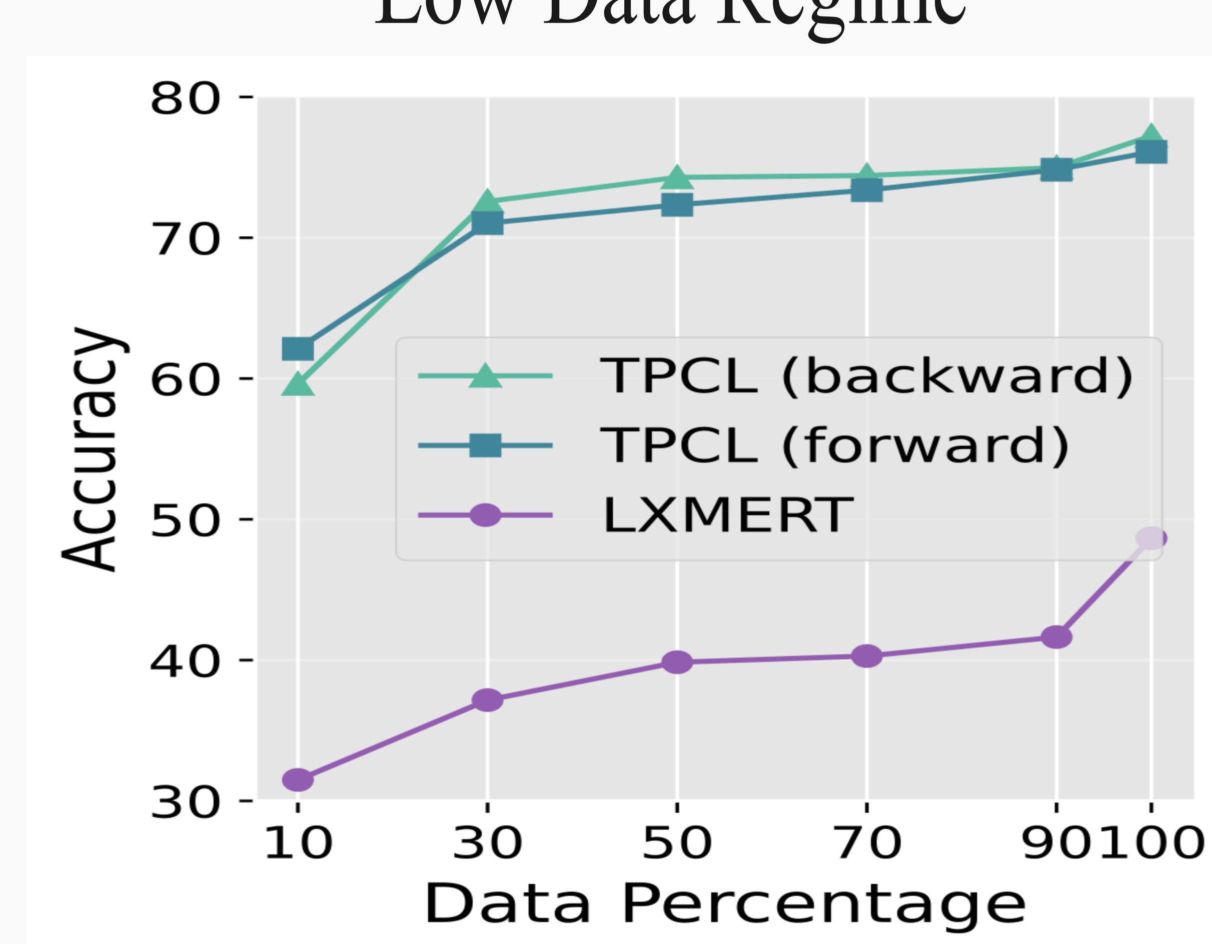


Results

OOD and ID datasets

Methods	VQA-CP v2				VQA v2
	Overall	Y/N	Num	Others	Overall
LXMERT	48.66	47.49	22.24	56.52	73.06
D-VQA	69.75	80.43	58.57	67.23	-
SIMPLEAUG	62.24	69.72	53.63	60.69	<u>74.98</u>
BILI	71.18	<u>92.18</u>	64.90	61.90	-
FAN-VQA	<u>72.18</u>	84.76	<u>65.98</u>	<u>67.29</u>	-
TPCL$_{Fix\uparrow}$	**75.83**	91.55	**68.49**	**69.61**	**78.42**
TPCL$_{Dyn\uparrow}$	**77.23**	**93.10**	**72.00**	**70.34**	**78.03**

Low Data Regime



Baseline Agnostic Approach