Vision & Language Models **struggle to answer** questions composed with **logic operators**

LREC 2022

Fine-tuning vs From Scratch: Do Vision & Language Models Have Similar Capabilities on Out-of-Distribution Visual Question Answering?

Jensen, Kristian Nørgaard and Plank, Barbara

INTRO

- Many new transformer-based Vision & Language models (22)
- Minimal work to analyse the performance benefit of these new models

METHODS

1. Testing BAN and LXMERT
2. Using VQAv2 as In-domain data
3. Using VQA-LOL, VQA-Rephrasings and VQA-Introspect as out-of-distribution data

RESULTS

	Out-of-distribution				In-domain
Model↓/Data→	C	S	R	I	VQA v2
LXMERT	49.51	46.61	**68.64**	**76.90**	**71.96**
BAN	**51.63**	**52.39**	61.43	70.05	66.27

DISCUSSION

- Both models struggle with the OR (∨) operator
- BAN best on combinations of logic operators
- LXMERT best at visual reasoning and consistency
- Dataset imperfections (skewed answer distributions, computer generated question) might harm performance of model

Model	Compose		Supplement	
	Single	Multi	Single	Multi
LXMERT	**57.33**	41.68	**55.32**	43.35
BAN	56.95	**46.30**	54.41	**51.63**

Table 6: **Accuracy for single vs multiple boolean operations.** Included are results for both VQA-Compose and VQA-Supplement.

Model	Reasoning	Perception
LXMERT	86.14	75.79
BAN	81.50	68.42

Table 7: **Accuracy on the VQA-Introspect data set.**

Model	Main + Sub	Only Main	Only Sub	Neither
LXMERT	**67.60**	**19.20**	**8.20**	**5.00**
BAN	58.34	24.05	10.08	7.53
Pythia	50.05	19.73	17.40	12.83

Table 8: **Quadrants of the VQA-Introspect data set.** Higher score on Main + Sub and lower score on other quadrants is better performance. Pythia scores reported from (Selvaraju et al., 2020).

Model	K				Accuracy	
	1	2	3	4	ORI	REP
LXMERT	**75.11**	**68.20**	**63.91**	**60.69**	**80.83**	**73.18**
BAN	67.37	58.54	53.34	49.77	74.75	64.91

Table 9: **Consensus score on VQA-Rephrasings.** The reaults are presented along with the accuracy on the original questions (ORI), and the rephrasings (REP).

IT-UNIVERSITETET I KØBENHAVN