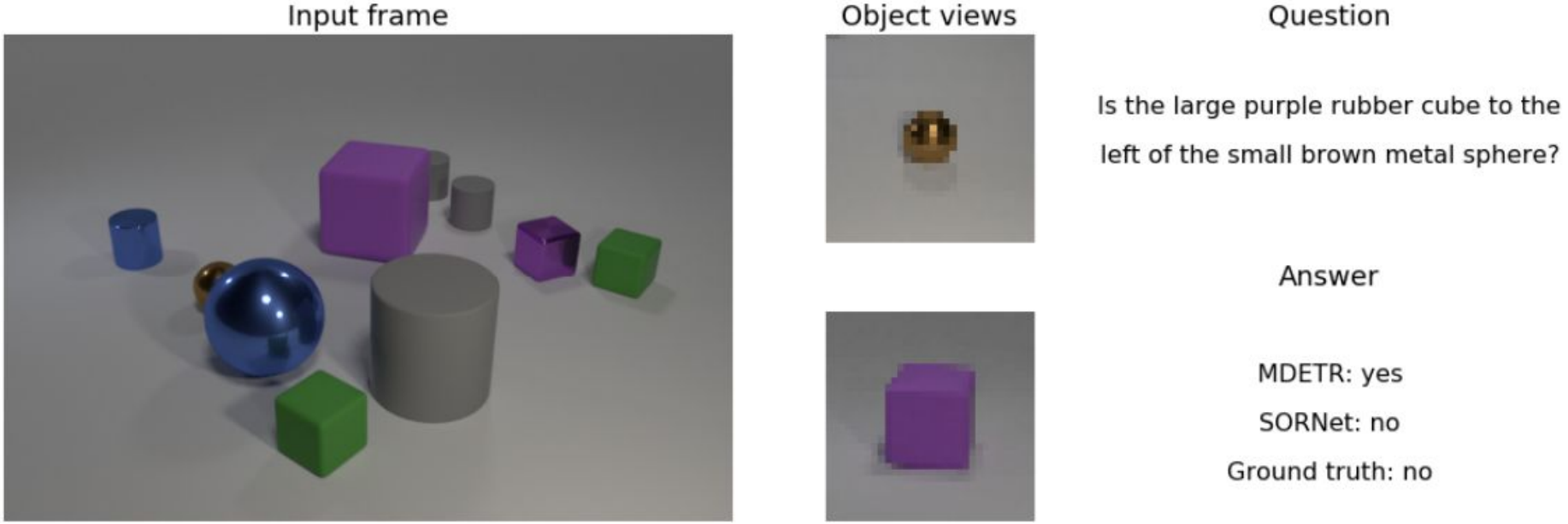
Авторы

NYU Center for Data Science; Facebook AI Research; NYU Courant Institute

- Aishwarya Kamath:
  PhD student at NYU CS, previously Facebook AI Research
- Mannat Singh:
  Facebook AI Research
- Yann LeCun:
  Chief AI Scientist at Facebook & Silver Professor at the Courant Institute, NYU
- Gabriel Synnaeve:
  Research scientist, Facebook AI Research
- Ishan Misra:
  Research Scientist, Facebook AI Research
- Nicolas Carion:
  Post-Doc at NYU Courant Institute

Контекст работы

- Pre-published on April 2021, Oral Presentation at ICCV2021/CVF
- Опорная статья **End-to-End Object Detection with Transformers (2020)**
- 18 цитирований
  - Visual Grounding, Referring Task
  - **SORNet: Spatial Object-Centric Representations for Sequential Manipulation (09/2021)**



	MDETR [34]	MDETR-oracle [34]	**SORNet(ours)**
ValA Accuracy	84.950	97.944	**99.006**
ValB Accuracy	59.627	98.052	**98.222**

Table 1: Zero-shot relation classification accuracy on CLEVR-CoGenT [24]. The MDETR-oracle model has seen all the objects during training, where as MDETR and SORNet have only see objects in condition A. SORNet takes canonical views as queries whereas MDETR takes text queries.

Дальнейшее направление

Исследование:

- эксперименты с текстовыми моделями
- добавление генеративных способностей

Применение:

- Referring Understanding Tasks
- Visual Grounding
- Visual Question Answering

Ссылки и источники

- End-to-End Object Detection with Transformers (2020)
- SORNet: Spatial Object-Centric Representations for Sequential Manipulation
- ICCV Daily 2021 - Tuesday
- W&B Paper Reading Group: MDETR with author Aishwarya Kamath