Visual Commonsense Reasoning On VCR, a model must not only answer commonsense visual questions, but also provide a rationale that explains why the answer is true.

Read our paper on arXiv»

Submitting to the leaderboard

Submission is easy! You just need to email Rowan (rowanz at cs.washington.edu) with your predictions. Formatting instructions are below:

Submit to the leaderboard»

Please include in your email 1) a name for your model, 2) your team name (including your affiliation), and optionally, 3) a github repo or paper link.

If you submit an ensemble, please tell me how many models you used in your ensemble.

I'll try to get back to you within a few days, usually sooner. Teams can only submit results from a model once every 7 days.

I reserve the right to not score any of your submissions if you cheat -- for instance, please don't make up a bunch of fake names / email addresses and send me multiple submissions under those names.

What kinds of submissions are allowed?

The only constraint is that your system must predict the **answer first**, then the rationale. (The rationales were selected to be highly relevant to the correct Q,A pair, so they leak information about the correct answer.)

- To deter this, the submission format involves submitting predictions for each possible rationale, conditioned on each possible answer.
- A simple way of setting up the experiments (used in the paper) is to consider a task with *query* and four *response* choices. For Q->A the query is the question, and the response choices are the answers. For QA->R, the query is the question and answer, concatenated together, and the response choices are the rationales.

Questions?

If it's not about something private, check out the google group below:

Get help via the google group»

VCR Leaderboard

There are two different subtasks to VCR:

- Question Answering (Q->A): In this setup, a model is provided a question, and has to pick the best answer out of four choices. Only one of the four is correct.
- **Answer Justification** (QA->R): In this setup, a model is provided a question, along with the correct answer, and it has to justify it by picking the best rationale out of four choices.

We combine the two parts with the **Q->AR** metric, in which a model only gets a question right if it answers correctly *and* picks the right rationale. Models are evaluated in terms of accuracy (%). How well will your model do?

Rank	Model Human Performance	Q- > A 91.0	QA- > R 93.0	Q- >AR 85.0
	University of Washington (Zellers et al. '18) VL-RoBERTa	84.2	86.4	72.8
December 14, 2021	Joint Laboratory of HIT and iFLYTEK Research (HFL) VLUA (single model)	82.3	87.0	72.0
June 28, 2021 3	Washou MMU	84.0	84.9	71.5
November 1, 2021 4	https://rowanzellers.com/merlo treserve	82.3	86.5	71.4
May 10, 2021		ں.۔		T
5 November 19, 2020	BLENDER (single model) WeSee Al team, Tencent	81.6	86.4	70.8
6 June 24, 2020	ERNIE-ViL-large(ensemble of 15 models) ERNIE-team - Baidu https://arxiv.org/abs/200001000	81.6	86.1	70.5
7	MMCNet (ensemble of 4 models)	80.0	83.1	66.9
October 28, 2020 8	UC Berkeley UNITER-large (ensemble of 10 models)	79.8	83.4	66.8
September 30, 2019	MS D365 AI https://arxiv.org/abs/1909.1174 0			
9 June 24, 2020	ERNIE-ViL-large(single model) ERNIE-team - Baidu https://arxiv.org/abs/2006.169	79.2	83.5	66.3
10 November 15, 2021	34 CLIP-TD anonymous	79.6	82.9	66.2
11 June 8, 2021	DELV Intel Labs Cognitive AI & MSRA NLC	79.4	82.5	65.8
12 May 22, 2020	VILLA-large (single model) MS D365 AI https://arxiv.org/pdf/2006.0619	78.9	82.8	65.7
13 May 28, 2021	5.pdf) MERLOT (single model) University of Washington / Al2	80.6	80.4	65.1
14	https://rowanzellers.com/merlo t VitsNet	78.8	81.8	64.6
October 30, 2020 15	Carnegie Mellon University gnimix anonymous	78.9	81.0	64.1
July 16, 2020 16 June 5, 2021	SEITU Anonymous	77.9	80.7	63.0
17	https://github.com/MyLittleCha nge/SEITU UNITER-large (single model)	77.3	80.8	62.8
September 23, 2019	https://arxiv.org/abs/1909.1174 0	70.3	78 7	62.6
18 November 1, 2021	الله المتلاقة المتلاق المتلاقة المتي متيان المتلاقة المتلاقة المتلاقة المتلاقة المتلاقة المتلاقة المت	⁄ પ્ ર.3	∕ ð.7	o2.6
19 June 24, 2020	ERNIE-ViL-base(single model) ERNIE-team - Baidu https://arxiv.org/abs/2006.169	77.0	80.3	62.1
20 September 10, 2021	34 Kam-net Anonymous	77.4	79.2	61.8
21 May 22, 2020	VILLA-base (single model) MS D365 AI	76.4	79.1	60.6
22	5.pdf) MMCNet UC Berkelev	76.0	79.3	60.6
October 22, 2020 23 February 17, 2021	Test_VILLA CCR	76.6	79.0	60.6
24 April 23, 2020	KVL-BERT Beijing Institute of Technology	76.4	78.6	60.3
25 August 9,2019	ViLBERT (ensemble of 10 models) Georgia Tech & Facebook Al Research	76.4	78.0	59.8
26	https://arxiv.org/abs/1908.022 65 VVT	75.3	78. <u>9</u>	59.7
∠0 March 23, 2019 27	v v I runningcat VL-BERT (single model)	75.8	, . .9 78.4	ງສ./ 59.7
September 23,2019	IVISRA & USTC https://arxiv.org/abs/1908.085 30	_	_	-
28 August 2, 2021 29	PVL (single model) PVL-team (UCLA) SGEITL	76.9 76.0	77.1 78.0	59.7 59.6
September 10, 2021 30	Anonomous vlt (single model) Anonymous	.5	77.8	58.9
January 8, 2021 31 August 9,2019	ViLBERT (ensemble of 5 models) Georgia Tech 2 Frank	75.7	77.5	58.8
	https://arxiv.org/abs/1908.022			
32 June 16, 2021	GITRL Anonomous	75.5	77.5	58.7
33 September 4,2019	Unicoder-VL (ensemble of 2 models) <i>MSRA & PKU</i> https://arxiv.org/abs/1908.060	76.0	77.1	58.6
34 September 22	66 UNITER-base (single model) <i>MS D365 AI</i>	75.0	77.2	58.2
35	https://arxiv.org/abs/1909.1174 0 TDN	75.7	76.4	58.0
November 5, 2020 36	VARMS (Sun Yat-sen University) B2T2 (ensemble of 5 models)	74.0	77.1	57.1
May 13, 2019	Google Research http://arxiv.org/abs/1908.0505 4			
37 September 11, 2021	YTX Harbin Institute of Technology B2T2 (cincle	73.8 70 f	74.8	55.4
ວັ່ວ May 13, 2019	ے ہے۔ Google Research http://arxiv.org/abs/1908.0505 4	, 2.6	v ט.7	JO.O
39 August 27,2019	Unicoder-VL (single model) <i>MSRA & PKU</i> https://arxiv.org/abs/1908.060	73.4	74.5	54.9
40 July 30,2019	66 ViLBERT (single model) Georgia Tech & Facebook Al Research	73.3	74.6	54.8
/11	https://arxiv.org/abs/1908.022 65 Vil REPT CONT	7° ^	7/ ″	54 0
41 May 26, 2021 42	Beijing Institute of Technology	73.3 72.8	74.4	J4.6 54.3
April 7, 2021 43 March 4	רווא אס אס און און איז	72.9	73.8	54.1
44 November 14, 2019	HGL Sun Yat-sen University	72.2	73.4	53.2
45 May 22, 2019	TNet (ensemble of 5) FlyingPig (Sun Yat-sen University)	72.7	72.6	53.0
46 March 31, 2021	CMR flying melon	72.3	72.8	52.8
47 July 5,2019	VisualBERT UCLA & AI2 & PKU https://arxiv.org/abs/1908.035 57	71.6	73.2	52.4
48 March 8,2021	SAC SAC	71.7	72.8	52.2
49 January 5, 2022	GTEHG MIC-Tongji University	71.2	72.4	51.7 E1
50 April 13, 2021 51	A3 Net A3 Net MKDN	/1.2 70.7	71.5 ⁷	51.4 50.5
September 4, 2019 52	Peking University TAB-VCR UIUC	70.4	71.7	50.5
October 25, 2019	https://arxiv.org/abs/1910.1467 1 RobustCl	70 ″	71 -	50 5
ວ3 June 3, 2020	RODUSTCL NeurIPS 2020 submission ID1218	, ∪.4 ,	- I.5	ວປ.5
54 October 30, 2021 55	YXY Peking University TNet (single model)	70.6	71.2	50.5 50.4
СО Мау 22, 2019 БС	FlyingPig (Sun Yat-sen University)	J.J 70 0	J.U 71 ^	50 1
March 26, 2021	WalkingDog from Qilu University of Technology	J.J 70 -	- 1.J 70 -	- J.4 F.0 -
57 November 29, 2019 58	WWR-Net Tianjin University transformer-r2c	70.8 70.7	/0.7 70.5	50.2 50.0
October 13, 2019 59	งารบ NEW BANNER (Sun Yat- sen University) HGL	70.1	70.8	49.8
May 13, 2019	HCP https://arxiv.org/abs/1910.1147 5			
60 September 4,2019 61	CAR Sun Yat-sen University SIA V1	70.5 69.6	70.4	49.8 49.8
01 December 6,2020 62	SFW1	ي. ت. 70.2	́л.З 70.7	.ు.ర 49.7
September 3, 2020 63 September 9	Jiangsu University SFW2 Jiangsu University	69.6	71.1	49.6
64 May 3, 2021	RKB Hefei University of Technology	69.6	70.7	49.3
65 June 10, 2021	PUV (Pretrain UNITER by VC feature) Anonymous	69.8	70.2	49.3
66 February 3, 2021	vlb (single model) Anonymous	69.8	69.8	48.9
67 June 10, 2021 68	BLU Anonomous MRCNet	69.3 68.4	70.1	48.9 48 <i>.</i> 4
May 16, 2019	MILAB (Seoul National University)	_J.4 6° -	J.Ə 70 -	. J.4
09 May 18, 2019 70	Anonymous MUGRN	68.2	69.4	۰.4 47.5
May 17, 2019 71 May 13, 2015	SGRE UTS	67.5	69.7	46.9
72	https://github.com/AmingWu/M ulti-modal-Circulant-Fusion/ FAIR	65.7	70.1	46.3
Feb 19, 2019 73	Facebook AI Research DAF Beijing Institute of Technology	66.9	68.7	46.0
November 28, 2019 74 Feb 25, 2019	CKRE Peking University	66.9	68.2	45.9
75 October 30, 2021	MIE flying melon	66.2	68.8	45.5
76 May 14, 2019	emnet DCP	66.6	68.0	45.4
77 Nov 28, 2018	Recognition to Cognition Networks University of Washington https://github.com/rowanz/r?c	65.1	67.3	44.0
78 May 20, 2019	DVD SL	66.3	65.0	43.3
79 March 27, 2019	GS Reasoning UC San Diego	65.7	61.0	41.1
80 May 17, 2019 81	R2R (text only) Anonymous R2CC	58.4 60.7	69.1 61.7	40.5 37.6
May 28, 2021 82	Arjun Singh BERT-Base Goodle Al Longue	53.9	64.5	.5 35.0
Nov 28, 2018	(experiment by Rowan) https://github.com/google- research/bert			
83 June 18, 2021	MUR Anonymous	46.5	46.0	25.6
84 September 15, 2021	SG-QA-model KDDI&SNAP collab (KDDI Research)	70.9	24.9	17.7
85 Nov 28, 2018	MLB Seoul National University (experiment by Rowan) https://github.com/inbudices/htt	46.2	36.8	17.2
86 Octob	ILowBiVQA BERT-base-vc-ft LUKA-Axe	58.6	24.9	14.5
October 23, 2021 87 November 15, 2019	Visual-Lang-base (ensemble) <i>Tiny-Group</i>	17.7	74.1	13.3
	Random Performance	25.0	25.0	6.2