SQuAD2.0

The Stanford Question Answering Dataset

What is SQuAD?

Stanford **Qu**estion **A**nswering **D**ataset (SQuAD) is a reading comprehension dataset, consisting of questions posed by crowdworkers on a set of Wikipedia articles, where the answer to every question is a segment of text, or *span*, from the corresponding reading passage, or the question might be unanswerable.

SQuAD2.0 combines the 100,000 questions in SQuAD1.1 with over 50,000 unanswerable questions written adversarially by crowdworkers to look similar to answerable ones. To do well on SQuAD2.0, systems must not only answer questions when possible, but also determine when no answer is supported by the paragraph and abstain from answering.

(/SQuAD-explorer/explore/v2.0/dev/)

(http://arxiv.org/abs/1806.03822)

SQuAD 1.1, the previous version of the SQuAD dataset, contains 100,000+ question-answer pairs on 500+ articles.

(/SQuAD-explorer/explore/1.1/dev/)

(http://arxiv.org/abs/1606.05250)

Getting Started

We've built a few resources to help you get started with the dataset.
Download a copy of the dataset (distributed under the CC BY-SA 4.0 (http://creativecommons.org/licenses/by-sa/4.0/legalcode) license):

Training Set v2.0 (40 MB) (/SQuAD-explorer/dataset/train-v2.0.json)

Dev Set v2.0 (4 MB) (/SQuAD-explorer/dataset/dev-v2.0.json)

To evaluate your models, we have also made available the evaluation script we will use for official evaluation, along with a sample prediction file that the script will take as input. To run the evaluation, use `python evaluate-v2.0.py <path_to_dev-v2.0> <path_to_predictions>` .

Evaluation Script v2.0
(https://worksheets.codalab.org/rest/bundles/0x6b567e1cf2e041ec80d7098f031c5c9e/conte

Sample Prediction File (on Dev v2.0)
(https://worksheets.codalab.org/bundles/0x8731effab84f41b7b874a070e40f61e2/)

Once you have a built a model that works to your expectations on the dev set, you submit it to get official scores on the dev and a hidden test set. To preserve the integrity of test results, we do not release the test set to the public. Instead, we require you to submit your model so that we can run it on the test set for you. Here's a tutorial walking you through official evaluation of your model:

Submission Tutorial
(https://worksheets.codalab.org/worksheets/0x8212d84ca41c4150b555a075b19ccc05/)

Because SQuAD is an ongoing effort, we expect the dataset to evolve.
To keep up to date with major changes to the dataset, please subscribe:

email address

Subscribe

Have Questions?

Ask us questions at our google group (https://groups.google.com/forum/#!forum/squad-stanford-qa) or at pranavsr@stanford.edu (mailto:pranavsr@stanford.edu) and robinjia@stanford.edu (mailto:robinjia@stanford.edu).

Tweet ☆ Star

Leaderboard

SQuAD2.0 tests the ability of a system to not only answer reading comprehension questions, but also abstain when presented with a question that cannot be answered based on the provided paragraph.

Rank	Model	EM	F1
	Human Performance *Stanford University* (Rajpurkar & Jia et al. '18) (http://arxiv.org/abs/1606.05250)	86.831	89.452
1 Apr 06, 2020	SA-Net on Albert (ensemble) *QIANXIN*	**90.724**	**93.011**
2 May 05, 2020	SA-Net-V2 (ensemble) *QIANXIN*	90.679	92.948
2 Apr 05, 2020	Retro-Reader (ensemble) *Shanghai Jiao Tong University* http://arxiv.org/abs/2001.09694 (http://arxiv.org/abs/2001.09694)	90.578	92.978
3 Dec 01, 2020	EntitySpanFocusV2 (ensemble) *RICOH_SRCB_DML*	90.521	92.824

3 Jul 31, 2020	ATRLP+PV (ensemble) *Hithink RoyalFlush*	90.442	92.877
3 May 04, 2020	ELECTRA+ALBERT+EntitySpanFocus (ensemble) *SRCB_DML*	90.442	92.839
4 Jun 21, 2020	ELECTRA+ALBERT+EntitySpanFocus (ensemble) *SRCB_DML*	90.420	92.799
4 Sep 11, 2020	EntitySpanFocus+AT (ensemble) *RICOH_SRCB_DML*	90.454	92.748
4 Mar 12, 2020	ALBERT + DAAF + Verifier (ensemble) *PINGAN Omni-Sinitic*	90.386	92.777
5 Jan 10, 2020	Retro-Reader on ALBERT (ensemble) *Shanghai Jiao Tong University* http://arxiv.org/abs/2001.09694 (http://arxiv.org/abs/2001.09694)	90.115	92.580
6 Nov 01, 2020	electra+nlayers+kdav (ensemble) *oppo.tensorlab*	90.002	92.497
6 Nov 06, 2019	ALBERT + DAAF + Verifier (ensemble) *PINGAN Omni-Sinitic*	90.002	92.425
7 Sep 18, 2019	ALBERT (ensemble model) *Google Research & TTIC* https://arxiv.org/abs/1909.11942 (https://arxiv.org/abs/1909.11942)	89.731	92.215
7 Feb 25, 2020	Albert_Verifier_AA_Net (ensemble) *QIANXIN*	89.743	92.180
7 Jun 28, 2020	ELECTRA+ATRLP+PV (single model) *Hithink RoyalFlush*	89.551	92.366
7 Mar 28, 2020	Retro-Reader on ELECTRA (single model) *Shanghai Jiao Tong University* http://arxiv.org/abs/2001.09694 (http://arxiv.org/abs/2001.09694)	89.562	92.052
7 Mar 27, 2020	albert+KD+transfer (ensemble) *Anonymous*	89.461	92.134
8 Nov 18, 2020	ROaD-Electra *single model*	89.449	92.118
9 Apr 21, 2020	albert+KD+transfer+twopass (single) *SPPD*	89.111	91.877
9 Apr 18, 2020	ALBERT + MTDA + SFVerifier (ensemble model) *Senseforth AI Research* https://www.senseforth.ai/ (https://www.senseforth.ai/)	89.235	91.739

10 Apr 15, 2020	ALBERT + SFVerifier (ensemble model) *Senseforth AI Research* https://www.senseforth.ai/ (https://www.senseforth.ai/)	89.133	91.666
10 Apr 23, 2020	ELECTRA+RL+EV (single model) *Hithink RoyalFlush*	89.021	91.765
11 Oct 28, 2020	electra & albert (ensemble)	88.998	91.635
11 Dec 08, 2019	ALBERT+Entailment DA (ensemble) *CloudWalk*	88.761	91.745
11 May 02, 2020	ELECTRA+EntitySpanFocus (Single model) *SRCB_DML*	88.874	91.546
12 Apr 14, 2020	SA-Net on Electra (single model) *QIANXIN*	88.851	91.486
13 Mar 06, 2020	ELECTRA (single model) *Google Brain & Stanford*	88.716	91.365
14 Aug 13, 2020	ELECTRA_ATT (single model) *Shanghai Jiao Tong University*	88.614	91.303
15 Feb 24, 2020	ALBERT (Single model) *SRCB_DML*	88.592	91.286
15 Feb 20, 2020	Tuned ALBERT (ensemble model) *Group Data & Analytics Cell | Aditya Birla Group)* https://www.adityabirla.com/About/group-data-and-analytics (https://www.adityabirla.com/About/group-data-and-analytics)	88.637	91.230
15 Jun 24, 2020	ALBERT + IG + NE (single model) *Anonymous*	88.569	91.287
16 Jun 24, 2020	ALBERT + IG (single model) *Anonymous*	88.524	91.256
16 Jan 19, 2020	Retro-Reader on ALBERT (single model) *Shanghai Jiao Tong University* http://arxiv.org/abs/2001.09694 (http://arxiv.org/abs/2001.09694)	88.107	91.419
16 Jul 22, 2019	XLNet + DAAF + Verifier (ensemble) *PINGAN Omni-Sinitic*	88.592	90.859
16 Mar 13, 2020	aanet_v2.0 (single model) *QIANXIN*	88.434	90.918
16 Dec 08, 2019	ALBERT+Entailment DA Verifier (single model) *CloudWalk*	87.847	91.265
16 Jan 07, 2020	ALBERT + SFVerifier (single model) *Senseforth AI Research* https://www.senseforth.ai/ (https://www.senseforth.ai/)	88.197	90.830

16 Sep 16, 2019	ALBERT (single model) *Google Research & TTIC* https://arxiv.org/abs/1909.11942 (https://arxiv.org/abs/1909.11942)	88.107	90.902
16 Mar 30, 2020	MTL (single model) *HAPTIK AI RESEARCH* https://haptik.ai (https://haptik.ai)	88.107	90.902
16 Jul 26, 2019	UPM (ensemble) *Anonymous*	88.231	90.713
16 Feb 10, 2020	SkERT-Large (single model) *Skelter Labs*	87.994	90.944
16 Aug 04, 2019	XLNet + SG-Net Verifier (ensemble) *Shanghai Jiao Tong University & CloudWalk* https://arxiv.org/abs/1908.05147 (https://arxiv.org/abs/1908.05147)	88.174	90.702
16 May 21, 2020	albert+KD+transfer+twopass (single) *SPPD*	87.949	90.818
16 Feb 29, 2020	ALBERT+RL (single model) *Hithink RoyalFlush*	87.870	90.823
16 May 22, 2020	albert_xxlarge (single model) *Zheyu Ye*	87.802	90.872
16 Nov 15, 2019	XLNet (single model) *Google Brain & CMU*	87.926	90.689
17 Feb 12, 2020	Tuned ALBERT (single model) *Group Data & Analytics Cell | Aditya Birla Group)* https://www.adityabirla.com/About/group-data-and-analytics (https://www.adityabirla.com/About/group-data-and-analytics)	87.847	90.532
17 Feb 10, 2020	ALBERT 1.1 (single model) *Anonymous*	87.700	90.588
18 Apr 04, 2020	LUKE (single model) *Studio Ousia & NAIST & RIKEN AIP* https://arxiv.org/abs/2010.01057 (https://arxiv.org/abs/2010.01057)	87.429	90.163
19 Aug 04, 2019	XLNet + SG-Net Verifier++ (single model) *Shanghai Jiao Tong University & CloudWalk* https://arxiv.org/abs/1908.05147 (https://arxiv.org/abs/1908.05147)	87.238	90.071
20 Jul 26, 2019	UPM (single model) *Anonymous*	87.193	89.934
20 Nov 27, 2019	RoBERTa+Verify (ensemble) *CW*	86.933	90.037

20 Mar 20, 2019	BERT + DAE + AoA (ensemble) *Joint Laboratory of HIT and iFLYTEK Research*	87.147	89.474
20 Jul 20, 2019	RoBERTa (single model) *Facebook AI*	86.820	89.795
21 Nov 12, 2019	RoBERTa+Verify (single model) *CW*	86.448	89.586
21 Mar 15, 2019	BERT + ConvLSTM + MTL + Verifier (ensemble) *Layer 6 AI*	86.730	89.286
22 Mar 05, 2019	BERT + N-Gram Masking + Synthetic Self-Training (ensemble) *Google AI Language* https://github.com/google-research/bert (https://github.com/google-research/bert)	86.673	89.147
22 May 29, 2020	Enhanced Albert+Verifier (ensemble) *Microsoft STCA AIC*	86.098	89.634
22 Oct 16, 2019	Xlnet+Verifier *single model*	86.594	89.082
23 Aug 30, 2019	Xlnet+Verifier (single model) *Ping An Life Insurance Company AI Team*	86.572	89.063
23 May 30, 2020	Enhanced Albert+Verifier3 (ensemble) *Microsoft STCA AIC*	85.827	89.778
23 Dec 09, 2019	XLNET-V2-123+ (single model) *MST/EOI* http://tia.today (http://tia.today)	86.403	89.148
24 May 21, 2019	XLNet (single model) *Google Brain & CMU*	86.346	89.133
25 May 14, 2019	SG-Net (ensemble) *Shanghai Jiao Tong University* https://arxiv.org/abs/1908.05147 (https://arxiv.org/abs/1908.05147)	86.211	88.848
25 Apr 14, 2019	SemBERT (ensemble) *Shanghai Jiao Tong University* https://arxiv.org/abs/1909.02209 (https://arxiv.org/abs/1909.02209)	86.166	88.886
25 Sep 29, 2019	BERTSP (single model) *NEUKG* http://www.techkg.cn/--please (http://www.techkg.cn/--please)	85.838	88.921
25 Sep 22, 2020	RoBERTa-Large (ensemble model) *SAIL*	85.872	88.793
25 Mar 16, 2019	BERT + DAE + AoA (single model) *Joint Laboratory of HIT and iFLYTEK Research*	85.884	88.621

25 Jul 22, 2019	SpanBERT (single model) *FAIR & UW*	85.748	88.709
26 Sep 22, 2020	RoBERTa-Large (single model) *SAIL*	85.173	88.425
26 May 14, 2019	SG-Net (single model) *Shanghai Jiao Tong University* https://arxiv.org/abs/1908.05147 (https://arxiv.org/abs/1908.05147)	85.229	87.926
26 Mar 13, 2019	BERT + ConvLSTM + MTL + Verifier (single model) *Layer 6 AI*	84.924	88.204
26 Mar 05, 2019	BERT + N-Gram Masking + Synthetic Self-Training (single model) *Google AI Language* https://github.com/google-research/bert (https://github.com/google-research/bert)	85.150	87.715
26 Jun 19, 2019	BNDVnet (single model) *PAOS*	85.003	87.833
26 Jan 15, 2019	BERT + MMFT + ADA (ensemble) *Microsoft Research Asia*	85.082	87.615
26 Apr 11, 2019	SemBERT (single model) *Shanghai Jiao Tong University* https://arxiv.org/abs/1909.02209 (https://arxiv.org/abs/1909.02209)	84.800	87.864
26 Sep 13, 2019	xlnet (single model) *VerifiedXiaoPAI*	84.642	88.000
26 Apr 16, 2019	Insight-baseline-BERT (single model) *PAII Insight Team*	84.834	87.644
27 Sep 03, 2019	Hanvon_model (single model) *Hanvon_WuHan*	84.721	87.117
28 Jan 10, 2019	BERT + Synthetic Self-Training (ensemble) *Google AI Language* https://github.com/google-research/bert (https://github.com/google-research/bert)	84.292	86.967
29 Nov 08, 2019	BERT + Multiple-CNN (ensemble) *Kyonggi University (ICL) & KISTI*	84.202	86.767
30 Jul 22, 2019	Tuned BERT-1seq Large Cased (single model) *FAIR & UW*	83.751	86.594
31 Mar 20, 2019	Bert-raw (ensemble) *None*	83.604	86.036
31 Dec 13, 2018	BERT finetune baseline (ensemble) *Anonymous*	83.536	86.096

31 Dec 21, 2018	PAML+BERT (ensemble model) *PINGAN GammaLab*	83.457	86.122
31 Dec 16, 2018	Lunet + Verifier + BERT (ensemble) *Layer 6 AI NLP Team*	83.469	86.043
32 Dec 15, 2018	Lunet + Verifier + BERT (single model) *Layer 6 AI NLP Team*	82.995	86.035
32 Jun 21, 2019	SENSEFORTH + BERT *single* https://senseforth.ai (https://senseforth.ai)	83.142	85.873
32 Jan 14, 2019	BERT + MMFT + ADA (single model) *Microsoft Research Asia*	83.040	85.892
32 May 14, 2019	ATB (single model) *Anonymous*	82.882	86.002
32 Feb 16, 2019	Bert-raw (ensemble) *None*	83.175	85.635
32 Feb 26, 2019	BERT with Something (ensemble) *Anonymous*	83.051	85.737
32 Jan 10, 2019	BERT + Synthetic Self-Training (single model) *Google AI Language* https://github.com/google-research/bert (https://github.com/google-research/bert)	82.972	85.810
32 Jul 22, 2019	Tuned BERT Large Cased (single model) *FAIR & UW*	82.803	85.863
32 Mar 11, 2019	Bert-raw (ensemble) *None*	83.119	85.510
32 Feb 15, 2019	BERT + NeurQuRI (ensemble) *2SAH*	82.803	85.703
33 Feb 28, 2019	BERT + NeurQuRI (ensemble) *2SAH*	82.713	85.584
33 May 13, 2019	BERT-Base + QA Pre-training (single model) *Anonymous*	82.724	85.491
33 Dec 16, 2018	PAML+BERT (single model) *PINGAN GammaLab*	82.577	85.603
34 Nov 16, 2018	AoA + DA + BERT (ensemble) *Joint Laboratory of HIT and iFLYTEK Research*	82.374	85.310
35 Dec 12, 2018	BERT finetune baseline (single model) *Anonymous*	82.126	84.820
35 Sep 22, 2020	BERT-Base PMI-Masking Additional Data (single model) *AI21 Labs*	82.024	84.854

36 Feb 28, 2019	BERT_s (single model) *Anonymous*	81.979	84.846
37 Feb 28, 2019	BERT-large+UBFT (single model) *anonymous*	81.573	84.535
38 Feb 15, 2019	BERT + NeurQuRI (single model) *2SAH*	81.257	84.342
38 Feb 25, 2019	BERT with Something (single model) *Anonymous*	81.110	84.386
38 Nov 16, 2018	AoA + DA + BERT (single model) *Joint Laboratory of HIT and iFLYTEK Research*	81.178	84.251
39 Mar 20, 2019	Bert-raw (single) *None*	80.693	83.922
39 Mar 07, 2019	BERT + UnAnsQ (single model) *Anonymous*	80.749	83.851
39 Sep 22, 2020	BERT-Base PMI-Masking (single model) *AI21 Labs*	80.896	83.604
40 Jan 22, 2019	BERT + NeurQuRI (single model) *2SAH*	80.591	83.391
40 Mar 12, 2019	Bert-raw (single) *None*	80.411	83.457
41 Sep 22, 2020	PMI-Masking Additional Data Random Baseline (single model) *AI21 Labs*	80.377	83.262
42 Feb 16, 2019	Bert-raw (single model) *None*	80.343	83.243
42 May 29, 2019	Bert *Single Model* https://senseforth.ai (https://senseforth.ai)	80.422	83.118
42 Sep 23, 2020	PMI-Masking Pure-PMI (single model) *AI21 Labs*	80.241	83.175
43 Apr 04, 2019	BISAN-CC (single model) *Seoul National University & Hyundai Motors*	80.208	83.149
43 Dec 03, 2018	PwP+BERT (single model) *AITRICS*	80.117	83.189
43 Jul 22, 2019	Original BERT Large Cased (single model) *FAIR & UW*	79.971	83.266
43 Feb 19, 2019	BERT + UDA (single model) *Anonymous*	80.005	83.208
44 Apr 10, 2019	bert (single model) *vinda msqjmxx*	79.971	83.184

44 Feb 28, 2019	ST_bl *single model*	80.140	82.962
44 Nov 09, 2018	BERT (single model) *Google AI Language*	80.005	83.061
45 Sep 23, 2020	PMI-Masking Additional Data Pure-PMI (single model) *AI21 Labs*	79.993	83.039
46 Feb 12, 2019	BERT + Sparse-Transformer *single model*	79.948	83.023
46 Sep 22, 2020	PMI-Masking Random Baseline (single model) *AI21 Labs*	80.038	82.796
46 Mar 07, 2019	BERT uncased (single model) *Anonymous*	79.745	83.020
46 Dec 06, 2018	NEXYS_BASE (single model) *NEXYS, DGIST R7*	79.779	82.912
47 Feb 02, 2019	{bert-finetuning} (single model) *ksai*	79.632	82.852
48 Feb 25, 2020	BERT-Large-Cased *single model*	79.610	82.692
49 Nov 09, 2018	L6Net + BERT (single model) *Layer 6 AI*	79.181	82.259
49 Mar 14, 2019	{Anonymous} (single model) *Anonymous*	78.876	82.524
50 Apr 24, 2019	BERT + WIAN (ensemble) *Infosys Limited*	78.650	81.497
51 Aug 03, 2020	AMBERT (single model) *ByteDance*	78.594	81.445
51 Mar 14, 2019	BISAN (single model) *Seoul National University & Hyundai Motors*	78.481	81.531
52 Dec 26, 2019	BERT-Large-Cased *single model*	78.357	81.500
53 Dec 14, 2018	BERT+AC (single model) *Hithink RoyalFlush*	78.052	81.174
54 Aug 03, 2020	BERT (single model) *ByteDance*	77.319	80.310
55 Nov 06, 2018	SLQA+BERT (single model) *Alibaba DAMO NLP* http://www.aclweb.org/anthology/P18-1158 (http://www.aclweb.org/anthology/P18-1158)	77.003	80.209
56 Aug 03, 2020	AMBERT-H (single model) *ByteDance*	76.710	79.659

56 Aug 03, 2020	AMBERT-S (single model) *ByteDance*	76.563	79.776
57 Jan 05, 2019	synss (single model) *bert_finetune*	76.055	79.329
58 Dec 19, 2018	ARSG-BERT (single model) *TRINITI RESEARCH LABS, Active.ai* https://active.ai (https://active.ai)	74.746	78.227
58 Aug 29, 2020	BERT-Base-V (single model) *Anonymous*	75.073	77.805
58 Nov 05, 2018	MIR-MRC(F-Net) (single model) *Kangwon National University, Natural Language Processing Lab. & ForceWin, KP Lab.*	74.791	77.988
59 Aug 06, 2020	BERT-Base-DT (single model) *Anonymous*	74.769	77.706
60 Dec 03, 2020	BERT-Base-V2 *single model*	74.656	77.404
61 Aug 14, 2020	BERT-Base-Add (single model) *Anonymous*	74.329	77.396
61 May 23, 2019	{BERTcw} (single model) *private*	74.385	77.308
62 Sep 13, 2018	nlnet (single model) *Microsoft Research Asia*	74.272	77.052
63 Jan 13, 2020	batch2 (single model) *THU*	73.742	76.858
64 Dec 29, 2018	MMIPN *Single*	73.505	76.424
65 Aug 09, 2020	BERT-Base-Baseline (single model) *Anonymous*	73.302	76.284
66 Apr 20, 2019	BERT-Base (single model) *Dining Philosophers*	73.099	76.236
67 Oct 12, 2018	YARCS (ensemble) *IBM Research AI*	72.670	75.507
67 Apr 23, 2020	BERT-base *single model*	72.072	75.513
67 Apr 25, 2020	BERTBase (single model) *Anonymous*	72.072	75.513
68 Nov 14, 2018	BERT+Answer Verifier (single model) *Pingan Tech Olatop Lab*	71.666	75.457

69 Sep 17, 2018	Unet (ensemble) *Fudan University & Liulishuo Lab* https://arxiv.org/abs/1810.06638 (https://arxiv.org/abs/1810.06638)	71.417	74.869
69 Apr 25, 2019	BERT-Base (single) *GreenflyAI* https://greenfly.ai (https://greenfly.ai)	71.699	74.430
69 Aug 15, 2018	Reinforced Mnemonic Reader + Answer Verifier (single model) *NUDT* https://arxiv.org/abs/1808.05759 (https://arxiv.org/abs/1808.05759)	71.767	74.295
69 Aug 28, 2018	SLQA+ (single model) *Alibaba DAMO NLP* http://www.aclweb.org/anthology/P18-1158 (http://www.aclweb.org/anthology/P18-1158)	71.462	74.434
69 Jan 19, 2019	{BERT-base} (single-model) *Anonymous*	70.763	74.449
69 Sep 14, 2018	SAN (ensemble model) *Microsoft Business Applications AI Research* https://arxiv.org/abs/1712.03556 (https://arxiv.org/abs/1712.03556)	71.316	73.704
70 Aug 21, 2018	FusionNet++ (ensemble) *Microsoft Business Applications Group AI Research* https://arxiv.org/abs/1711.07341 (https://arxiv.org/abs/1711.07341)	70.300	72.484
70 Sep 26, 2018	Multi-Level Attention Fusion(MLAF) (single model) *Chonbuk National University, Cognitive Computing Lab.*	69.476	72.857
71 Sep 14, 2018	Unet (single model) *Fudan University & Liulishuo Lab*	69.262	72.642
72 Dec 20, 2018	DocQA + NeurQuRI (single model) *2SAH*	68.766	71.662
73 Aug 21, 2018	SAN (single model) *Microsoft Business Applications AI Research* https://arxiv.org/abs/1712.03556 (https://arxiv.org/abs/1712.03556)	68.653	71.439
73 Sep 13, 2018	BiDAF++ with pair2vec (single model) *UW and FAIR*	68.021	71.583
73 Jun 25, 2018	KACTEIL-MRC(GFN-Net) (single model) *Kangwon National University, Natural Language Processing Lab.*	68.213	70.878
73 Jul 13, 2018	VS^3-NET (single model) *Kangwon National University in South Korea*	67.897	70.884
74 Jan 02, 2019	EBB-Net (single model) *Enliple AI*	66.610	70.303

75 Jun 25, 2018	KakaoNet2 (single model) *Kakao NLP Team*	65.719	69.381
76 Sep 13, 2018	BiDAF++ (single model) *UW and FAIR*	65.651	68.866
76 Jul 11, 2018	abcNet (single model) *Fudan University & Liulishuo AI Lab*	65.256	69.206
77 Jun 27, 2018	BSAE AddText (single model) *reciTAL.ai*	63.338	67.422
78 Aug 14, 2018	eeAttNet (single model) *BBD NLP Team* https://www.bbdservice.com (https://www.bbdservice.com)	63.327	66.633
78 May 30, 2018	BiDAF + Self Attention + ELMo (single model) *Allen Institute for Artificial Intelligence [modified by Stanford]*	63.372	66.251
79 May 30, 2018	BiDAF + Self Attention (single model) *Allen Institute for Artificial Intelligence [modified by Stanford]*	59.332	62.305
80 May 30, 2018	BiDAF-No-Answer (single model) *University of Washington [modified by Stanford]*	59.174	62.093
80 Nov 27, 2018	Tree-LSTM + BiDAF + ELMo (single model) *Carnegie Mellon University*	57.707	62.341

SQuAD1.1 Leaderboard

Here are the ExactMatch (EM) and F1 scores evaluated on the test set of SQuAD v1.1.

Rank	Model	EM	F1
	Human Performance *Stanford University* (Rajpurkar et al. '16) (http://arxiv.org/abs/1606.05250)	82.304	91.221
1 Apr 10, 2020	LUKE (single model) *Studio Ousia & NAIST & RIKEN AIP*	**90.202**	**95.379**
2 May 21, 2019	XLNet (single model) *Google Brain & CMU*	89.898	95.080
3 Dec 11, 2019	XLNET-123++ (single model) *MST/EOI* http://tia.today (http://tia.today)	89.856	94.903
3 Aug 11, 2019	XLNET-123 (single model) *MST/EOI*	89.646	94.930

4 Sep 25, 2019	BERTSP (single model) *NEUKG* http://www.techkg.cn/ (http://www.techkg.cn/)	88.912	94.584
4 Jul 21, 2019	SpanBERT (single model) *FAIR & UW*	88.839	94.635
5 Jul 03, 2019	BERT+WWM+MT (single model) *Xiaoi Research*	88.650	94.393
6 Jul 21, 2019	Tuned BERT-1seq Large Cased (single model) *FAIR & UW*	87.465	93.294
7 Oct 05, 2018	BERT (ensemble) *Google AI Language* https://arxiv.org/abs/1810.04805 (https://arxiv.org/abs/1810.04805)	87.433	93.160
8 May 14, 2019	ATB (single model) *Anonymous*	86.940	92.641
9 Jul 21, 2019	Tuned BERT Large Cased (single model) *FAIR & UW*	86.521	92.617
9 Jul 04, 2019	BERT+MT (single model) *Xiaoi Research*	86.458	92.645
10 Feb 14, 2019	KT-NET (single model) *Baidu NLP*	85.944	92.425
10 Sep 27, 2018	nlnet (ensemble) *Microsoft Research Asia*	85.954	91.677
10 Feb 28, 2019	ST_bl *single model*	85.430	91.976
11 Nov 21, 2019	EL-BERT (single model) *YeonTaek Oh*	85.335	91.807
12 Mar 14, 2019	BISAN (single model) *Seoul National University & Hyundai Motors*	85.314	91.756
12 Jun 03, 2019	DPN (single model) *Anonymous*	84.978	92.019
12 Oct 05, 2018	BERT (single model) *Google AI Language* https://arxiv.org/abs/1810.04805 (https://arxiv.org/abs/1810.04805)	85.083	91.835
12 Jul 11, 2019	BERT-uncased (single model) *Anonymous*	84.926	91.932
12 Feb 16, 2019	BERT+Sparse-Transformer *single model*	85.125	91.623
12 Sep 09, 2018	nlnet (ensemble) *Microsoft Research Asia*	85.356	91.202

12 Jul 21, 2019	Original BERT Large Cased (single model) *FAIR & UW*	84.328	91.281
12 Feb 19, 2019	WD (single model) *Anonymous*	84.402	90.561
12 Jul 11, 2018	QANet (ensemble) *Google Brain & CMU*	84.454	90.490
12 Apr 21, 2019	Common-sense Governed BERT-123 (single model) *Jerry AGI Ragtag*	83.930	90.613
13 Feb 21, 2019	WD1 (single model) *Anonymous*	83.804	90.429
13 Jul 08, 2018	r-net (ensemble) *Microsoft Research Asia*	84.003	90.147
13 May 08, 2019	Common-sense Governed BERT-123 (single model) *MST/EOI*	82.943	91.074
13 Jun 20, 2018	MARS (ensemble) *YUANFUDAO research NLP*	83.982	89.796
14 Mar 19, 2018	QANet (ensemble) *Google Brain & CMU*	83.877	89.737
14 Sep 09, 2018	nlnet (single model) *Microsoft Research Asia*	83.468	90.133
15 Sep 01, 2018	MARS (single model) *YUANFUDAO research NLP*	83.185	89.547
16 Jun 21, 2018	MARS (single model) *YUANFUDAO research NLP*	83.122	89.224
17 Mar 06, 2018	QANet (ensemble) *Google Brain & CMU*	82.744	89.045
17 Jun 20, 2018	QANet (single) *Google Brain & CMU*	82.471	89.306
17 Jan 22, 2018	Hybrid AoA Reader (ensemble) *Joint Laboratory of HIT and iFLYTEK Research*	82.482	89.281
17 Feb 19, 2018	Reinforced Mnemonic Reader + A2D (ensemble model) *Microsoft Research Asia & NUDT*	82.849	88.764
17 May 09, 2018	MARS (single model) *YUANFUDAO research NLP*	82.587	88.880
17 Jan 03, 2018	r-net+ (ensemble) *Microsoft Research Asia*	82.650	88.493
17 Jan 05, 2018	SLQA+ (ensemble) *Alibaba iDST NLP*	82.440	88.607

17 Jul 14, 2019	BERT (single model) *KTNET*	82.062	88.947
17 Feb 28, 2018	QANet (single model) *Google Brain & CMU*	82.209	88.608
17 Feb 02, 2018	Reinforced Mnemonic Reader (ensemble model) *NUDT and Fudan University* https://arxiv.org/abs/1705.02798 (https://arxiv.org/abs/1705.02798)	82.283	88.533
17 Dec 24, 2018	MMIPN *Single*	81.580	88.948
17 Dec 17, 2017	r-net (ensemble) *Microsoft Research Asia* http://aka.ms/rnet (http://aka.ms/rnet)	82.136	88.126
17 Dec 17, 2018	ARSG-BERT (single model) *TRINITI RESEARCH LABS, Active.ai* https://active.ai (https://active.ai)	81.307	88.909
17 Dec 22, 2017	AttentionReader+ (ensemble) *Tencent DPDAC NLP*	81.790	88.163
18 May 09, 2018	Reinforced Mnemonic Reader + A2D (single model) *Microsoft Research Asia & NUDT*	81.538	88.130
18 Apr 23, 2018	r-net (single model) *Microsoft Research Asia*	81.391	88.170
18 May 09, 2018	Reinforced Mnemonic Reader + A2D + DA (single model) *Microsoft Research Asia & NUDT*	81.401	88.122
18 Apr 03, 2018	KACTEIL-MRC(GF-Net+) (ensemble) *Kangwon National University, Natural Language Processing Lab.*	81.496	87.557
18 Feb 27, 2018	QANet (single model) *Google Brain & CMU*	80.929	87.773
18 Nov 17, 2017	BiDAF + Self Attention + ELMo (ensemble) *Allen Institute for Artificial Intelligence*	81.003	87.432
18 Feb 19, 2018	Reinforced Mnemonic Reader + A2D (single model) *Microsoft Research Asia & NUDT*	80.919	87.492
18 Mar 11, 2020	batch (single model) *THU*	79.859	88.263
18 Feb 12, 2018	Reinforced Mnemonic Reader + A2D (single model) *Microsoft Research Asia & NUDT*	80.489	87.454
18 Apr 12, 2018	AVIQA+ (ensemble) *aviqa team*	80.615	87.311
19 Jan 13, 2018	SLQA+ *single model*	80.436	87.021

19 Jan 04, 2018	{EAZI} (ensemble) *Yiwise NLP Group*	80.436	86.912
19 Jan 12, 2018	EAZI+ (ensemble) *Yiwise NLP Group*	80.426	86.912
19 Jan 22, 2018	Hybrid AoA Reader (single model) *Joint Laboratory of HIT and iFLYTEK Research*	80.027	87.288
19 Jan 06, 2020	BERT-INDEPENDENT-DSS-FILTERED (single model) *Brno University of Technology*	79.597	87.374
19 Mar 20, 2018	DNET (ensemble) *QA geeks*	80.164	86.721
20 Feb 13, 2018	BiDAF + Self Attention + ELMo + A2D (single model) *Microsoft Research Asia & NUDT*	79.996	86.711
21 Jan 03, 2018	r-net+ (single model) *Microsoft Research Asia*	79.901	86.536
21 Feb 23, 2018	MAMCN+ (single model) *Samsung Research*	79.692	86.727
22 Jan 29, 2018	Reinforced Mnemonic Reader (single model) *NUDT and Fudan University* https://arxiv.org/abs/1705.02798 (https://arxiv.org/abs/1705.02798)	79.545	86.654
22 Dec 05, 2017	SAN (ensemble model) *Microsoft Business AI Solutions Team* https://arxiv.org/abs/1712.03556 (https://arxiv.org/abs/1712.03556)	79.608	86.496
22 Dec 28, 2017	SLQA+ (single model) *Alibaba iDST NLP*	79.199	86.590
23 Oct 18, 2017	Interactive AoA Reader+ (ensemble) *Joint Laboratory of HIT and iFLYTEK*	79.083	86.450
23 Nov 05, 2018	KACTEIL-MRC(GF-Net+Distillation) (single model) *Kangwon National University, Natural Language Processing Lab.*	79.083	86.288
24 Jun 02, 2018	MDReader *single model*	79.031	86.006
24 Oct 24, 2017	FusionNet (ensemble) *Microsoft Business AI Solutions Team* https://arxiv.org/abs/1711.07341 (https://arxiv.org/abs/1711.07341)	78.978	86.016
25 Oct 22, 2017	DCN+ (ensemble) *Salesforce Research* https://arxiv.org/abs/1711.00106 (https://arxiv.org/abs/1711.00106)	78.852	85.996

26 Mar 30, 2018	KACTEIL-MRC(GF-Net+) (single model) *Kangwon National University, Natural Language Processing Lab.*	78.664	85.780
26 Nov 03, 2017	BiDAF + Self Attention + ELMo (single model) *Allen Institute for Artificial Intelligence*	78.580	85.833
27 May 10, 2018	KakaoNet (single model) *Kakao NLP Team*	78.401	85.724
28 Nov 30, 2017	SLQA (ensemble) *Alibaba iDST NLP*	78.328	85.682
28 Mar 19, 2018	aviqa (ensemble) *aviqa team*	78.496	85.469
28 Jan 02, 2018	Conductor-net (ensemble) *CMU* https://arxiv.org/abs/1710.10504 (https://arxiv.org/abs/1710.10504)	78.433	85.517
28 Sep 18, 2018	BiDAF++ with pair2vec (single model) *UW and FAIR*	78.223	85.535
28 Jun 01, 2018	MDReader0 *single model*	78.171	85.543
28 Jan 03, 2018	MEMEN (single model) *Zhejiang University* https://arxiv.org/abs/1707.09098 (https://arxiv.org/abs/1707.09098)	78.234	85.344
28 Jan 29, 2018	test *single*	78.087	85.348
29 Jul 26, 2017	Interactive AoA Reader (ensemble) *Joint Laboratory of HIT and iFLYTEK Research*	77.845	85.297
30 Mar 20, 2018	DNET (single model) *QA geeks*	77.646	84.905
31 Sep 18, 2018	BiDAF++ (single model) *UW and FAIR*	77.573	84.858
31 Dec 06, 2017	AttentionReader+ (single) *Tencent DPDAC NLP*	77.342	84.925
31 Dec 14, 2017	RaSoR + TR + LM (single model) *Tel-Aviv University* https://arxiv.org/abs/1712.03609 (https://arxiv.org/abs/1712.03609)	77.583	84.163
31 Dec 21, 2017	Jenga (ensemble) *Facebook AI Research*	77.237	84.466

31 Nov 06, 2017	Conductor-net (ensemble) *CMU* https://arxiv.org/abs/1710.10504 (https://arxiv.org/abs/1710.10504)	76.996	84.630
31 Jan 23, 2018	MARS (single model) *YUANFUDAO research NLP*	76.859	84.739
32 May 14, 2018	VS^3-NET (single model) *Kangwon National University in South Korea*	76.775	84.491
32 Nov 01, 2017	SAN (single model) *Microsoft Business AI Solutions Team* https://arxiv.org/abs/1712.03556 (https://arxiv.org/abs/1712.03556)	76.828	84.396
32 Sep 26, 2018	{gqa} (single model) *FAIR*	77.090	83.931
32 Dec 19, 2017	FRC (single model) *in review*	76.240	84.599
32 Oct 13, 2017	r-net (single model) *Microsoft Research Asia* http://aka.ms/rnet (http://aka.ms/rnet)	76.461	84.265
33 Oct 22, 2017	Conductor-net (ensemble) *CMU*	76.146	83.991
34 Sep 08, 2017	FusionNet (single model) *Microsoft Business AI Solutions team* https://arxiv.org/abs/1711.07341 (https://arxiv.org/abs/1711.07341)	75.968	83.900
35 Oct 22, 2017	Interactive AoA Reader+ (single model) *Joint Laboratory of HIT and iFLYTEK*	75.821	83.843
35 Oct 18, 2018	KAR (single model) *York University* https://arxiv.org/abs/1809.03449 (https://arxiv.org/abs/1809.03449)	76.125	83.538
36 Jul 14, 2017	smarnet (ensemble) *Eigen Technology & Zhejiang University*	75.989	83.475
37 Mar 15, 2018	AVIQA-v2 (single model) *aviqa team*	75.926	83.305
38 Aug 18, 2017	RaSoR + TR (single model) *Tel-Aviv University* https://arxiv.org/abs/1712.03609 (https://arxiv.org/abs/1712.03609)	75.789	83.261
38 Mar 20, 2020	Kbs (single model) *Tsinghua University*	75.034	83.405

38 Oct 23, 2017	DCN+ (single model) *Salesforce Research* https://arxiv.org/abs/1711.00106 (https://arxiv.org/abs/1711.00106)	75.087	83.081
38 Nov 01, 2017	Mixed model (ensemble) *Sean*	75.265	82.769
38 May 21, 2017	MEMEN (ensemble) *Eigen Technology & Zhejiang University* https://arxiv.org/abs/1707.09098 (https://arxiv.org/abs/1707.09098)	75.370	82.658
38 Nov 17, 2017	two-attention-self-attention (ensemble) *guotong1988*	75.223	82.716
38 Jul 10, 2017	DCN+ (single model) *Salesforce Research* https://arxiv.org/abs/1711.00106 (https://arxiv.org/abs/1711.00106)	74.866	82.806
38 Mar 09, 2017	ReasoNet (ensemble) *MSR Redmond* https://arxiv.org/abs/1609.05284 (https://arxiv.org/abs/1609.05284)	75.034	82.552
38 Oct 31, 2017	SLQA (single model) *Alibaba iDST NLP*	74.489	82.815
38 Feb 06, 2018	Jenga (single model) *Facebook AI Research*	74.373	82.845
38 Jan 02, 2018	Conductor-net (single model) *CMU* https://arxiv.org/abs/1710.10504 (https://arxiv.org/abs/1710.10504)	74.405	82.742
38 Aug 14, 2018	eeAttNet (single model) *BBD NLP Team* https://www.bbdservice.com (https://www.bbdservice.com)	74.604	82.501
39 Feb 13, 2018	SSR-BiDAF *ensemble model*	74.541	82.477
40 Jul 14, 2017	Mnemonic Reader (ensemble) *NUDT and Fudan University* https://arxiv.org/abs/1705.02798 (https://arxiv.org/abs/1705.02798)	74.268	82.371
41 Dec 23, 2017	S^3-Net (ensemble) *Kangwon National University in South Korea*	74.121	82.342
42 Jul 29, 2017	SEDT (ensemble model) *CMU* https://arxiv.org/abs/1703.00572 (https://arxiv.org/abs/1703.00572)	74.090	81.761

43 Jul 06, 2017	SSAE (ensemble) *Tsinghua University*	74.080	81.665
43 Jul 25, 2017	Interactive AoA Reader (single model) *Joint Laboratory of HIT and iFLYTEK Research*	73.639	81.931
43 Feb 22, 2017	BiDAF (ensemble) *Allen Institute for AI & University of Washington* https://arxiv.org/abs/1611.01603 (https://arxiv.org/abs/1611.01603)	73.744	81.525
43 Apr 22, 2017	SEDT+BiDAF (ensemble) *CMU* https://arxiv.org/abs/1703.00572 (https://arxiv.org/abs/1703.00572)	73.723	81.530
43 Nov 06, 2017	Conductor-net (single) *CMU* https://arxiv.org/abs/1710.10504 (https://arxiv.org/abs/1710.10504)	73.240	81.933
43 Dec 14, 2017	Jenga (single model) *Facebook AI Research*	73.303	81.754
43 Jan 24, 2017	Multi-Perspective Matching (ensemble) *IBM Research* https://arxiv.org/abs/1612.04211 (https://arxiv.org/abs/1612.04211)	73.765	81.257
43 May 01, 2017	jNet (ensemble) *USTC & National Research Council Canada & York University* https://arxiv.org/abs/1703.04617 (https://arxiv.org/abs/1703.04617)	73.010	81.517
44 Oct 22, 2017	Conductor-net (single) *CMU*	72.590	81.415
44 Apr 12, 2017	T-gating (ensemble) *Peking University*	72.758	81.001
44 Nov 16, 2017	two-attention-self-attention (single model) *guotong1988*	72.600	81.011
44 Sep 20, 2017	BiDAF + Self Attention (single model) *Allen Institute for Artificial Intelligence* https://arxiv.org/abs/1710.10723 (https://arxiv.org/abs/1710.10723)	72.139	81.048
44 Mar 03, 2018	AVIQA (single model) *aviqa team*	72.485	80.550
44 Dec 15, 2017	S^3-Net (single model) *Kangwon National University in South Korea*	71.908	81.023
45 Nov 06, 2017	attention+self-attention (single model) *guotong1988*	71.698	80.462

46 Nov 02, 2016	Dynamic Coattention Networks (ensemble) *Salesforce Research* https://arxiv.org/abs/1611.01604 (https://arxiv.org/abs/1611.01604)	71.625	80.383
46 Apr 13, 2017	QFASE *NUS*	71.898	79.989
46 Jul 14, 2017	smarnet (single model) *Eigen Technology & Zhejiang University* https://arxiv.org/abs/1710.02772 (https://arxiv.org/abs/1710.02772)	71.415	80.160
47 Jul 14, 2017	Mnemonic Reader (single model) *NUDT and Fudan University* https://arxiv.org/abs/1705.02798 (https://arxiv.org/abs/1705.02798)	70.995	80.146
47 May 23, 2018	AttReader (single) *College of Computer & Information Science, SouthWest University, Chongqing, China*	71.373	79.725
47 Apr 22, 2018	MAMCN (single model) *Samsung Research*	70.985	79.939
47 Oct 27, 2017	M-NET (single) *UFL*	71.016	79.835
48 Mar 24, 2017	jNet (single model) *USTC & National Research Council Canada & York University* https://arxiv.org/abs/1703.04617 (https://arxiv.org/abs/1703.04617)	70.607	79.821
48 Apr 02, 2017	Ruminating Reader (single model) *New York University* https://arxiv.org/abs/1704.07415 (https://arxiv.org/abs/1704.07415)	70.639	79.456
48 Mar 14, 2017	Document Reader (single model) *Facebook AI Research* https://arxiv.org/abs/1704.00051 (https://arxiv.org/abs/1704.00051)	70.733	79.353
48 Mar 08, 2017	ReasoNet (single model) *MSR Redmond* https://arxiv.org/abs/1609.05284 (https://arxiv.org/abs/1609.05284)	70.555	79.364
48 Dec 29, 2016	FastQAExt *German Research Center for Artificial Intelligence* https://arxiv.org/abs/1703.04816 (https://arxiv.org/abs/1703.04816)	70.849	78.857
48 May 13, 2017	RaSoR (single model) *Google NY, Tel-Aviv University* https://arxiv.org/abs/1611.01436 (https://arxiv.org/abs/1611.01436)	70.849	78.741

48 Apr 14, 2017	Multi-Perspective Matching (single model) *IBM Research* https://arxiv.org/abs/1612.04211 (https://arxiv.org/abs/1612.04211)	70.387	78.784
49 Aug 30, 2017	SimpleBaseline (single model) *Technical University of Vienna*	69.600	78.236
49 Feb 06, 2018	SSR-BiDAF *single model*	69.443	78.358
50 Apr 12, 2017	SEDT+BiDAF (single model) *CMU* https://arxiv.org/abs/1703.00572 (https://arxiv.org/abs/1703.00572)	68.478	77.971
51 Jun 25, 2017	PQMN (single model) *KAIST & AIBrain & Crosscert*	68.331	77.783
52 Apr 12, 2017	T-gating (single model) *Peking University*	68.132	77.569
52 Jul 29, 2017	SEDT (single model) *CMU* https://arxiv.org/abs/1703.00572 (https://arxiv.org/abs/1703.00572)	68.163	77.527
52 Dec 29, 2016	FastQA *German Research Center for Artificial Intelligence* https://arxiv.org/abs/1703.04816 (https://arxiv.org/abs/1703.04816)	68.436	77.070
52 Jan 22, 2018	FABIR *Single Model* https://arxiv.org/abs/1810.09580 (https://arxiv.org/abs/1810.09580)	67.744	77.605
52 Nov 28, 2016	BiDAF (single model) *Allen Institute for AI & University of Washington* https://arxiv.org/abs/1611.01603 (https://arxiv.org/abs/1611.01603)	67.974	77.323
53 Oct 26, 2016	Match-LSTM with Ans-Ptr (Boundary) (ensemble) *Singapore Management University* https://arxiv.org/abs/1608.07905 (https://arxiv.org/abs/1608.07905)	67.901	77.022
53 Sep 19, 2017	AllenNLP BiDAF (single model) *Allen Institute for AI* http://allennlp.org/ (http://allennlp.org/)	67.618	77.151
54 Feb 05, 2017	Iterative Co-attention Network *Fudan University*	67.502	76.786
55 Jan 03, 2018	newtest *single model*	66.527	75.787

55 Nov 02, 2016	Dynamic Coattention Networks (single model) *Salesforce Research* https://arxiv.org/abs/1611.01604 (https://arxiv.org/abs/1611.01604)	66.233	75.896
56 Oct 26, 2016	Match-LSTM with Bi-Ans-Ptr (Boundary) *Singapore Management University* https://arxiv.org/abs/1608.07905 (https://arxiv.org/abs/1608.07905)	64.744	73.743
57 Sep 21, 2017	OTF dict+spelling (single) *University of Montreal* https://arxiv.org/abs/1706.00286 (https://arxiv.org/abs/1706.00286)	64.083	73.056
57 Feb 19, 2017	Attentive CNN context with LSTM *NLPR, CASIA*	63.306	73.463
58 Nov 02, 2016	Fine-Grained Gating *Carnegie Mellon University* https://arxiv.org/abs/1611.01724 (https://arxiv.org/abs/1611.01724)	62.446	73.327
58 Sep 21, 2017	OTF spelling (single) *University of Montreal* https://arxiv.org/abs/1706.00286 (https://arxiv.org/abs/1706.00286)	62.897	72.016
59 Sep 21, 2017	OTF spelling+lemma (single) *University of Montreal* https://arxiv.org/abs/1706.00286 (https://arxiv.org/abs/1706.00286)	62.604	71.968
60 Sep 28, 2016	Dynamic Chunk Reader *IBM* https://arxiv.org/abs/1610.09996 (https://arxiv.org/abs/1610.09996)	62.499	70.956
60 Nov 15, 2019	RQA+IDR (single model) *BUAA & MSRA* https://arxiv.org/abs/2005.02925 (https://arxiv.org/abs/2005.02925)	61.145	71.389
61 Aug 27, 2016	Match-LSTM with Ans-Ptr (Boundary) *Singapore Management University* https://arxiv.org/abs/1608.07905 (https://arxiv.org/abs/1608.07905)	60.474	70.695
62 Aug 27, 2016	Match-LSTM with Ans-Ptr (Sentence) *Singapore Management University* https://arxiv.org/abs/1608.07905 (https://arxiv.org/abs/1608.07905)	54.505	67.748
62 Nov 15, 2019	RQA (single model) *BUAA & MSRA* https://arxiv.org/abs/2005.02925 (https://arxiv.org/abs/2005.02925)	55.827	65.467

63 Aug 22, 2019	UQA (single model) *Anonymous*	53.698	64.036