

PREFACE

What datasets are used in this research work?

During training, our TAG is supervised with the binary cross-entropy loss applied using the originally annotated QA pairs and adapts to generate novel QA pairs during generation. During the QA pairs generation process, we pass an input answer, each of which is selected from the extracted OCR tokens, into the TAG module and generate the corresponding question accordingly. In this way, the generated QA pairs cover a diverse set of scene text which was not directly exploited in the original annotation set. For answer selection, we perform a simple yet efficient strategy that is feeding the OCR token with the largest bounding box as the answer candidate to the proposed TAG. The intuition behind this design is that the scene text with the largest bounding box region is likely to encode semantically meaningful information for scene text-based understanding and reasoning. Also, scene text with a larger font size has a higher chance to be detected correctly without recognition error in general. As we illustrate in our experiments, our simple design facilitates a better understanding of the visual content and provides promising Text-VQA performance. Note that, more high-quality QA pairs could be continuously augmented with a more sophisticated answer-candidate selection strategy. We leave this direction as future work.

STUDENT, PROF, COLLABORATOR: BMVC AUTHOR GUIDELINES

7


4 Experiments

We evaluate TAG both qualitatively and quantitatively on the TextVQA [41] and the ST-VQA [8] datasets. We first present a brief overview of the datasets and implementation details. Then, we empirically validate the effectiveness of our proposed method by comparing it with the existing Text-VQA approaches. Our framework clearly outperforms previous work by a significant margin on both datasets.

4.1 Datasets and Evaluation Metrics

TextVQA dataset [41] is a widely used benchmark for the Text-VQA task. It consists of 28,408 images sourced from the Open Images dataset [31], with human-annotated questions that require reasoning over text in the images. We follow the standard split on the training, validation and test sets [22, 50]. For each question, the answer prediction is evaluated based on the soft-voting accuracy of 10 human-annotated answers [18, 22, 50].

ST-VQA dataset [8] is another popular dataset for the Text-VQA task. It contains 23,038 images from multiple sources including ICDAR 2013 [27], ICDAR 2015 [28], ImageNet [13], VizWiz [20], IIIT STR [35], Visual Genome [30], and COCO-Text [44]. The standard evaluation protocol on the ST-VQA dataset consists of accuracy and Average Normalized Levenshtein Similarity (ANLS) [8].



Can seeing documents improve Data efficiency and Domain adaptability in extractive Document Question Answering?

Venkat Mohit Sornapudi

Vrije Universiteit Amsterdam | v.m.sornapudi@student.vu.nl

Deloitte | vsornapudi@deloitte.nl



AGENDA

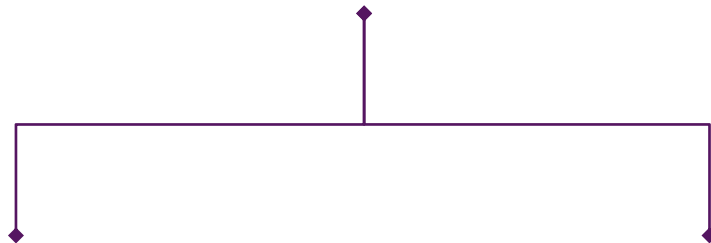
1. Introduction
2. Related work
3. Methodology
4. Experimental setup
5. Results
6. Next steps
7. Questions



INTRODUCTION

Task subcategories?

Document Question Answering Task



Extractive Question Answering

Answers only from given data

Generative Question Answering

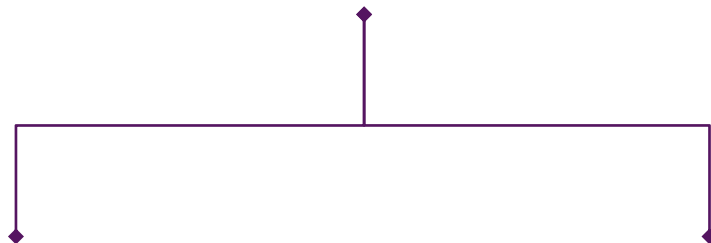
Answers creatively based on given data



INTRODUCTION

Task subcategories?

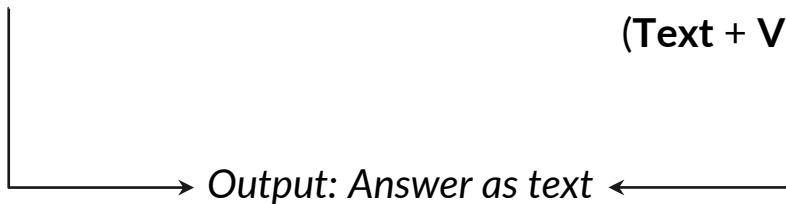
Document Question Answering Task



Textual Question Answering

Multimodal Question Answering

(Text + Vision)





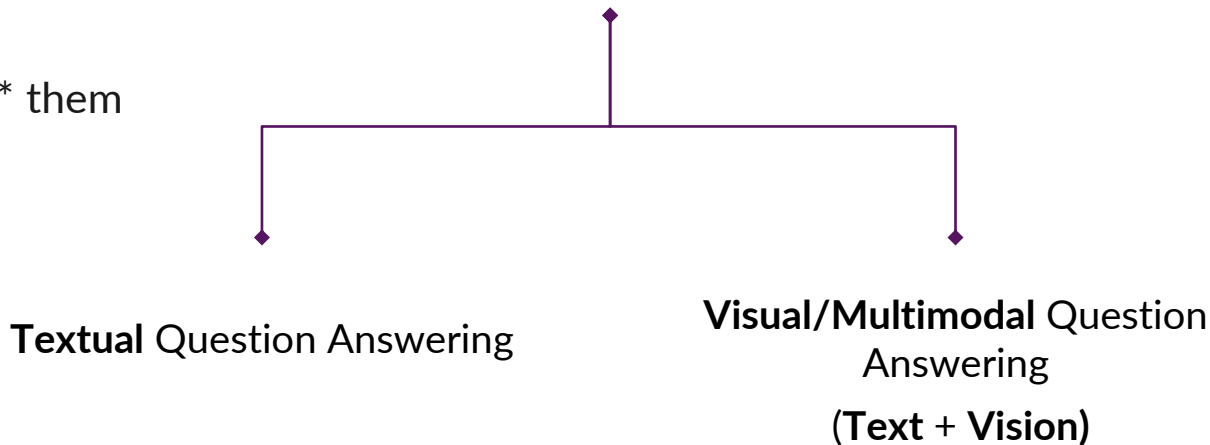
INTRODUCTION

How to choose?

We need to compare* them

* *In a **fair** manner*

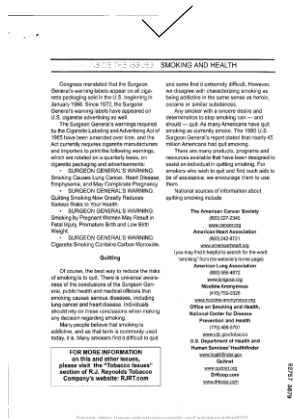
Document Question Answering Task




INTRODUCTION

Why and How to do fair comparison?

Text-only documents



Source: <https://www.industrydocuments.ucsf.edu/docs/bfck002>



Agenda for Menopausal Health Publication/Menopausal Meeting
 September 23, 2003, 9:00 AM-10:30 AM, CVCB302

Please note that the status of the presentation/publications that are bolded and not bolded has changed since the last meeting. The status of the abstracts shown is as follows:

- I. Action Items from August's Meeting
- II. Study Tracking Update: Premenopausal Factors of Products (August 19, 2003; September 25, 2003)
- A. Medical and Scientific Meeting Presentations

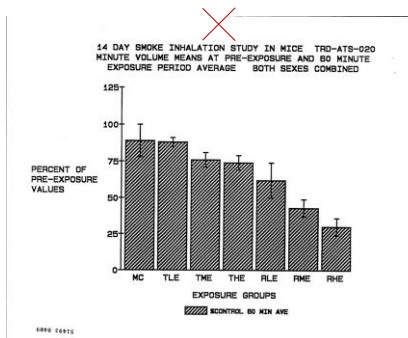
Status	Meeting	Investigator	Study/Title	Source
PRESENTED	American Chemical Society, 272nd National Meeting, New York, NY September 2-11, 2003	Collier	Novel peptide-containing estrogen modulators Design & synthesis of selective estrogen receptor beta selective ligands	W0101
		Malamas	Selective estrogen receptor β agonists are potent anti-inflammatory agents	W0102
PRESENTED	Third International Meeting on Estrogenic and Steroid Hormones	Biochemistry/Endocrinology	Rapid estradiol signaling mechanisms and function in the classical hormone progesterone receptor (ER-estral presentation)	W0103
	Fluoride, Rab September 12-14, 2003			
PRESENTED	American Society for Bone and Mineral Research, 75th Annual Meeting, Minneapolis, MN September 25-29, 2003	Zhao	Inhibition of pyruvate carboxylase (PC) reduces bone formation in a mouse calvarial model (O&P)	W0104
		Bhai	Functional modulation of ER α activity by strong repressors in LRP5-BL/PLP1 (O&P)	W0105
		Khalade	Aluminum to bone density: mice of low to heterogeneous resistance to ER α (Phenotype study)	W0106

DESIGN137023

What questions should be excluded?

Questions based on:

- Tables
- Images (infographics/photos)
- Numerical reasoning





INTRODUCTION

Research gap:

No dataset exists to do **fair** comparisons of ...

Textual and Multimodal QA models

Courtesy: <https://paperswithcode.com/task/question-answering>

include [SQuAD](#), [HotPotQA](#), [bAbI](#), [TriviaQA](#), [WikiQA](#), and typically evaluated on metrics like EM and F1. Some recent top performing models are [T5](#) and [GPT-4](#).

(Image credit: [SQuAD](#))

Benchmarks

These leaderboards are used to track progress in Question Answering

[Add a Result](#)

Trend	Dataset	Best Model	Paper	Code	Compare



INTRODUCTION

Abstract:

*This thesis aims to address the **lack of a fair comparison** between Visual Question Answering (**VQA**) models and Textual Question Answering (**TextualQA**) models on text-only Document QA datasets, where both models can excel. To facilitate this comparison, **a new text-only Document QA dataset** was generated using the existing SQuAD 2.0 dataset. Subsequently, selected VQA and TextualQA models are **compared in terms of their data efficiency and domain adaptability**. The results of this study can discover potential benefits of incorporating visual information into the Document Question Answering task.*



INTRODUCTION

Why to compare Data efficiency?

To get best results even while having less training data



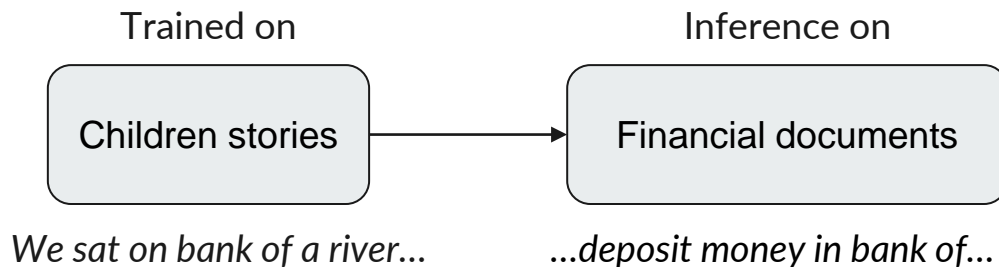
INTRODUCTION

What is the significance Domain adaptability?

1. Different interpretations

e.g, interpretations of “bank”

2. no/very few interpretation

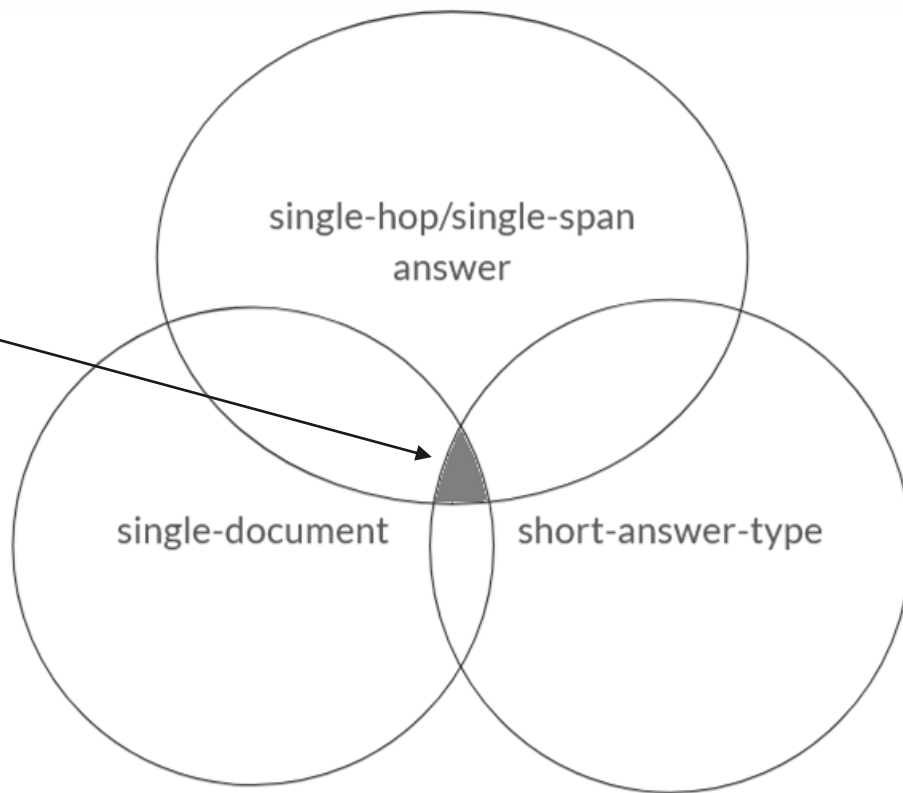


Not all models have the same ability to adapt to different domains. Hence we need to compare.



INTRODUCTION

Scope of the project



Related work

Extractive Textual QA models:

BERT, RoBERTa, XLM, Longformer, T5...

Extractive Visual QA models:

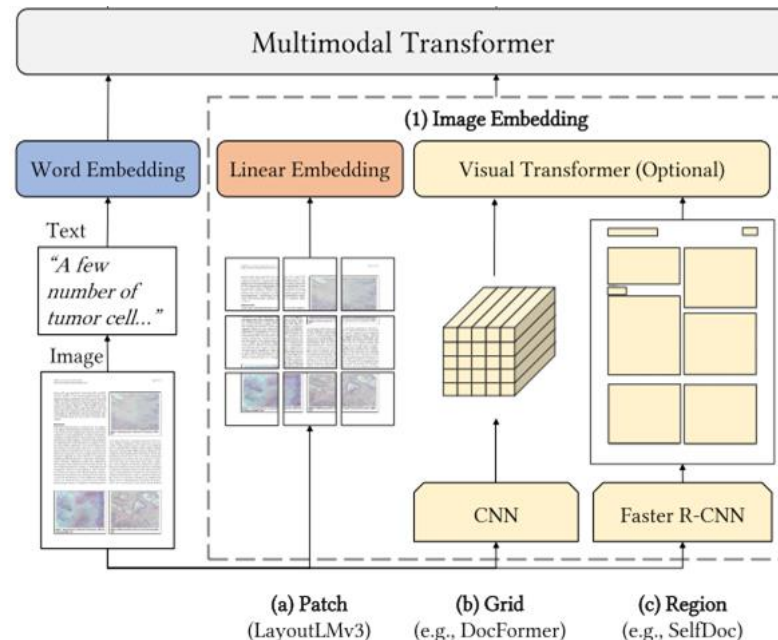
VisualBERT, LayoutLMv3, LayoutXML, Docformer, Hi-VT5...

All are Encoder models: Seq2seq Extractive models

Inputs:

1. Textual: context, question, answer position
2. Visual: image, word list, bounding boxes, question, answer position

How do they encode information?



Courtesy: Huang, Y., Lv, T., Cui, L., Lu, Y., & Wei, F. (2022, July 19). Layoutlmv3: Pre-training for document AI with unified text and image masking. arXiv.org. <https://arxiv.org/abs/2204.08387>

Related work

Which ones of the available datasets is/are suited for the project?

Datasets:

Webpage vs Scanned Documents

Datasets	Lang.	Text-only QA	Multi-hop/ Multi-span	Multi-page	Multi-doc	Needs image understanding	Comments
WebSRC [9]	En	Yes	No	No	No	No	HTML source code, screenshots and metadata; answer is either a text span or yes/no
SQuAD2.0 [16]	En	Yes	No	No	No	No	web-page can be extracted using wikipedia title
DocVQA [4]	En	No	No	No	No	No	mix of printed, typewritten and handwritten content; extractive answers
VisualMRC [5]	En	No	No	No	No	No	web-pages; mostly short sentences/phrases; abstractive answers
NewsQA [12]	En	–	Yes	No	No	No	HTML source code; short answers
Natural Questions [13]	En	–	Yes	No	No	No	web-page; short + long answers
DuReader _{vis} [10]	Zh	–	Yes	No	No	No	noisy texts; answers contain long answers such as multi-span texts, lists, and tables
Insurance VQA [19]	Zh	–	–	–	–	–	scanned documents of insurance scenarios (for example: medical bills)
MP-DocVQA [6]	En	–	Yes	Yes	No	No	has sections, paragraphs, diagrams, table
TAT-DQA [7]	En	–	Yes	Yes	No	No	+tables; requires numerical reasoning
MultiModalQA [14]	En	–	Yes	No	No	Yes	web-page can be extracted using wikipedia url
InfographicVQA [8]	En	–	Yes	No	No	Yes	has infographics; requires numerical reasoning; extractive answers
SlideVQA [17]	En	–	Yes	Yes	No	Yes	requires numerical reasoning
BioASQ [11]	En	–	Yes	Yes	Yes	No	need to refer web-pages (docs + concepts)
DocCVQA [18]	En	–	Yes	Yes	Yes	No	only 20 questions

Related work

What are the promising datasets?

Vehicle Highlights
Fuel Economy: 24 mpg City, 34 mpg Hwy
Engine: 2.4 L Premium Unleaded I-4, 201 HP
Transmission: Automatic
[View More Features and Specifications](#)

(a)

Displays 6.5 inch (1720x1080) Processor 10th Core Front Camera 8MP
 Rear Camera 12MP + 3MP RAM 8GB Storage 128GB
 CDR Software Capacity 350GBs OS Android 10

(b)

2013 Acura TSX 4dr Sdn I4 Auto
 Color: White VIN: 1HCUJ44200001713 **Price \$14,500**
 Interior: Leather Engine: 2.4L DOHC 16v Value: 19702 34 engine Mileage: 16,214 Stock #: 01733

2010 Audi A5 3dr Cpe Auto quattro 2.0L Premium
 Color: Black VIN: WNUCAF8A00007920 **Price \$13,600**
 Interior: Leather Engine: 2.0L DOHC 1931 4-cyl engine Mileage: 52,870 Stock #: 70730

(c)

Season	Team	GP	GS
2016-17	BOS	78	20
2017-18	BOS	70	70
2018-19	BOS	74	25
2019-20	BOS	57	57
2020-21	BOS	58	58
CAREER		337	230

(d)

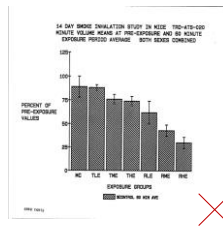
WebSRC

Chen, X., Zhao, Z., Chen, L., Zhang, D., Ji, J., Luo, A., Xiong, Y., & Yu, K. (2021, November 8). WebSRC: A dataset for web-based structural reading comprehension. *arXiv.org*. <https://arxiv.org/abs/2101.09465>

THE WINSTON "NO BULL" INTERCEPT PROCESS

Booth Agents
 Booth Agents will remain in the booth until 15 minutes prior to the start of the event. Booth Agents will remain in the booth until 15 minutes prior to the start of the event. Booth Agents will remain in the booth until 15 minutes prior to the start of the event.

Security Guards
 Security Guards must arrive at least 15 minutes prior to the start of the event. Security Guards must arrive at least 15 minutes prior to the start of the event. Security Guards must arrive at least 15 minutes prior to the start of the event.



Appendix A: Personnel Supply Management Reporting
 Worksheet: 2019-2020

Personnel Supply Management Reporting
 Worksheet: 2019-2020

Position	Current	Projected	Recruitment	Retention	Training
Police Officer	100	105	10	95	5
Police Sergeant	50	55	5	90	5
Police Lieutenant	20	25	2	85	3
Police Captain	10	15	1	80	2
Police Major	5	10	0.5	75	1
Police Chief	2	5	0.2	70	0.5

DocVQA

Mathew, M., Karatzas, D., & Jawahar, C. V. (2021, January 5). DocVQA: A dataset for VQA on Document Images. *arXiv.org*. <https://arxiv.org/abs/2007.00398>

East page:

Other relevant activity is known to:

Signature page:

Trustee's Annual Report for the period

Section 1: Financial Statements

Section 2: Governance

Section 3: Environmental and Social

Section 4: Other

STATEMENT OF FINANCIAL ACTIVITIES

For the year ended 31 December 2017 comprising an interim and retrospective annual

	2017	2016	2015	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005	2004	2003	2002	2001	2000	1999	1998	1997	1996	1995	1994	1993	1992	1991	1990	1989	1988	1987	1986	1985	1984	1983	1982	1981	1980	1979	1978	1977	1976	1975	1974	1973	1972	1971	1970	1969	1968	1967	1966	1965	1964	1963	1962	1961	1960	1959	1958	1957	1956	1955	1954	1953	1952	1951	1950	1949	1948	1947	1946	1945	1944	1943	1942	1941	1940	1939	1938	1937	1936	1935	1934	1933	1932	1931	1930	1929	1928	1927	1926	1925	1924	1923	1922	1921	1920	1919	1918	1917	1916	1915	1914	1913	1912	1911	1910	1909	1908	1907	1906	1905	1904	1903	1902	1901	1900	1899	1898	1897	1896	1895	1894	1893	1892	1891	1890	1889	1888	1887	1886	1885	1884	1883	1882	1881	1880	1879	1878	1877	1876	1875	1874	1873	1872	1871	1870	1869	1868	1867	1866	1865	1864	1863	1862	1861	1860	1859	1858	1857	1856	1855	1854	1853	1852	1851	1850	1849	1848	1847	1846	1845	1844	1843	1842	1841	1840	1839	1838	1837	1836	1835	1834	1833	1832	1831	1830	1829	1828	1827	1826	1825	1824	1823	1822	1821	1820	1819	1818	1817	1816	1815	1814	1813	1812	1811	1810	1809	1808	1807	1806	1805	1804	1803	1802	1801	1800	1799	1798	1797	1796	1795	1794	1793	1792	1791	1790	1789	1788	1787	1786	1785	1784	1783	1782	1781	1780	1779	1778	1777	1776	1775	1774	1773	1772	1771	1770	1769	1768	1767	1766	1765	1764	1763	1762	1761	1760	1759	1758	1757	1756	1755	1754	1753	1752	1751	1750	1749	1748	1747	1746	1745	1744	1743	1742	1741	1740	1739	1738	1737	1736	1735	1734	1733	1732	1731	1730	1729	1728	1727	1726	1725	1724	1723	1722	1721	1720	1719	1718	1717	1716	1715	1714	1713	1712	1711	1710	1709	1708	1707	1706	1705	1704	1703	1702	1701	1700	1699	1698	1697	1696	1695	1694	1693	1692	1691	1690	1689	1688	1687	1686	1685	1684	1683	1682	1681	1680	1679	1678	1677	1676	1675	1674	1673	1672	1671	1670	1669	1668	1667	1666	1665	1664	1663	1662	1661	1660	1659	1658	1657	1656	1655	1654	1653	1652	1651	1650	1649	1648	1647	1646	1645	1644	1643	1642	1641	1640	1639	1638	1637	1636	1635	1634	1633	1632	1631	1630	1629	1628	1627	1626	1625	1624	1623	1622	1621	1620	1619	1618	1617	1616	1615	1614	1613	1612	1611	1610	1609	1608	1607	1606	1605	1604	1603	1602	1601	1600	1599	1598	1597	1596	1595	1594	1593	1592	1591	1590	1589	1588	1587	1586	1585	1584	1583	1582	1581	1580	1579	1578	1577	1576	1575	1574	1573	1572	1571	1570	1569	1568	1567	1566	1565	1564	1563	1562	1561	1560	1559	1558	1557	1556	1555	1554	1553	1552	1551	1550	1549	1548	1547	1546	1545	1544	1543	1542	1541	1540	1539	1538	1537	1536	1535	1534	1533	1532	1531	1530	1529	1528	1527	1526	1525	1524	1523	1522	1521	1520	1519	1518	1517	1516	1515	1514	1513	1512	1511	1510	1509	1508	1507	1506	1505	1504	1503	1502	1501	1500	1499	1498	1497	1496	1495	1494	1493	1492	1491	1490	1489	1488	1487	1486	1485	1484	1483	1482	1481	1480	1479	1478	1477	1476	1475	1474	1473	1472	1471	1470	1469	1468	1467	1466	1465	1464	1463	1462	1461	1460	1459	1458	1457	1456	1455	1454	1453	1452	1451	1450	1449	1448	1447	1446	1445	1444	1443	1442	1441	1440	1439	1438	1437	1436	1435	1434	1433	1432	1431	1430	1429	1428	1427	1426	1425	1424	1423	1422	1421	1420	1419	1418	1417	1416	1415	1414	1413	1412	1411	1410	1409	1408	1407	1406	1405	1404	1403	1402	1401	1400	1399	1398	1397	1396	1395	1394	1393	1392	1391	1390	1389	1388	1387	1386	1385	1384	1383	1382	1381	1380	1379	1378	1377	1376	1375	1374	1373	1372	1371	1370	1369	1368	1367	1366	1365	1364	1363	1362	1361	1360	1359	1358	1357	1356	1355	1354	1353	1352	1351	1350	1349	1348	1347	1346	1345	1344	1343	1342	1341	1340	1339	1338	1337	1336	1335	1334	1333	1332	1331	1330	1329	1328	1327	1326	1325	1324	1323	1322	1321	1320	1319	1318	1317	1316	1315	1314	1313	1312	1311	1310	1309	1308	1307	1306	1305	1304	1303	1302	1301	1300	1299	1298	1297	1296	1295	1294	1293	1292	1291	1290	1289	1288	1287	1286	1285	1284	1283	1282	1281	1280	1279	1278	1277	1276	1275	1274	1273	1272	1271	1270	1269	1268	1267	1266	1265	1264	1263	1262	1261	1260	1259	1258	1257	1256	1255	1254	1253	1252	1251	1250	1249	1248	1247	1246	1245	1244	1243	1242	1241	1240	1239	1238	1237	1236	1235	1234	1233	1232	1231	1230	1229	1228	1227	1226	1225	1224	1223	1222	1221	1220	1219	1218	1217	1216	1215	1214	1213	1212	1211	1210	1209	1208	1207	1206	1205	1204	1203	1202	1201	1200	1199	1198	1197	1196	1195	1194	1193	1192	1191	1190	1189	1188	1187	1186	1185	1184	1183	1182	1181	1180	1179	1178	1177	1176	1175	1174	1173	1172	1171	1170	1169	1168	1167	1166	1165	1164	1163	1162	1161	1160	1159	1158	1157	1156	1155	1154	1153	1152	1151	1150	1149	1148	1147	1146	1145	1144	1143	1142	1141	1140	1139	1138	1137	1136	1135	1134	1133	1132	1131	1130	1129	1128	1127	1126	1125	1124	1123	1122	1121	1120	1119	1118	1117	1116	1115	1114	1113	1112	1111	1110	1109	1108	1107	1106	1105	1104	1103	1102	1101	1100	1099	1098	1097	1096	1095	1094	1093	1092	1091	1090	1089	1088	1087	1086	1085	1084	1083	1082	1081	1080	1079	1078	1077	1076	1075	1074	1073	1072	1071	1070	1069	1068	1067	1066	1065	1064	1063	1062	1061	1060	1059	1058	1057	1056	1055	1054	1053	1052	1051	1050	1049	1048	1047	1046	1045	1044	1043	1042	1041	1040	1039	1038	1037	1036	1035	1034	1033	1032	1031	1030	1029	1028	1027	1026	1025	1024	1023	1022	1021	1020	1019	1018	1017	1016	1015	1014	1013	1012	1011	1010	1009	1008	1007	1006	1005	1004	1003	1002	1001	1000	999	998	997	996	995	994	993	992	991	990	989	988	987	986	985	984	983	982	981	980	979	978	977	976	975	974	973	972	971	970	969	968	967	966	965	964	963	962	961	960	959	958	957	956	955	954	953	952	951	950	949	948	947	946	945	944	943	942	941	940	939	938	937	936	935	934	933	932	931	930	929	928	927	926	925	924	923	922	921	920	919	918	917	916	915	914	913	912	911	910	909	908	907	906	905	904	903	902	901	900	899	898	897	896	895	894	893	892	891	890	889	888	887	886	885	884	883	882	881	880	879	878	877	876	875	874	873	872	871	870	869	868	867	866	865	864	863	862	861	860	859	858	857	856	855	854	853	852	851	850	849	848	847	846	845	844	843	842	841	840	839	838	837	836	835	834	833	832	831	830	829	828	827	826	825	824	823	822	821	820	819	818	817	816	815	814	813	812	811	810	809	808	807	806	805	804	803	802	801	800	799	798	797	796	795	794	793	792	791	790	789	788	787	786	785	784	783	782	781	780	779	778	777	776	775	774	773	772	771	770	769	768	767	766	765	764	763	762	761	760	759	758	757	756	755	754	753	752	751	750	749	748	747	746	745	744	743	742	741	740	739	738	737	736	735	734	733	732	731	730	729	728	727	726	725	724	723	722	721	720	719	718	717	716	715	714	713	712	711	710	709	708	707	706	705	704	703	702	701	700	699	698	697	696	695	694	693	692	691	690	689	688	687	686	685	684	683	682	681	680	679	678	677	676	675	674	673	672	671	670	669	668	667	666	665	664	663	662	661	660	659	658	657	656	655	654	653	652	651	650	649	648	647	646	645	644	643	642	641	640	639	638	637	636	635	634	633	632	631	630	629	628	627	626	625	624	623	622	621	620	619	618	617	616	615	614	613	612	611	610	609	608	607	606	605	604	6
--	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	---



Methodology

SQuAD 2.0

- Questions based on individual paras
- SQuAD 2.0 was made in 2018
- SQuAD2.0 combines the 100,000 questions (on 500+ articles) in SQuAD1.1 with over 50,000 unanswerable questions
- SQuAD1.1 was made in 2016

New Visual dataset

Scraping wikipedia and gather visual assets
(PDF, images, word bounding boxes)

```
{
  "version": "v2.0",
  "data": [{
    "title": "Normans",
    "paragraphs": [{
      "context": "The Normans (Norman: Nourmands; French: Normands; ...",
      "qas": [{
        "question": "In what country is Normandy located?",
        "id": "56dde6b9a695914005b9628",
        "answers": [{
          "text": "France",
          "answer_start": 159
        }, ...
      ],
      "is_impossible": false
    }, ...
  ], ...
}]
}
```


SQuAD 2.0 data




Methodology

Selected models

Textual model	Multimodal model	Similarities	Differences
BERT	VisualBERT (Faster-RCNN)	Same Transformer, Word tokenizer	Inputs, Training objectives
RoBERTa	LayoutLMv3 (Linear Image transformation)	Same Transformer, Word tokenizer	Inputs, Training objectives
XLM	LayoutXLM (Linear Image transformation)	Same Transformer, Word tokenizer	Inputs, Training objectives



Same transformer, but trained on different objectives and have different tokenizers



Same transformer, trained on different objectives and have different tokenizers

Methodology

How can the models be ran on long docs?

1. Para level (complete/broken)
2. Page/document level: Needs loss balancing

Data efficiency: By increasing trainset size

Domain adaptation: Using metadata

Fine tuned not a training from scratch

This article is about the game. For the character referred to as the "Twilight Princess", see Midna.

The Legend of Zelda: Twilight Princess (Japanese: 伝説の勇者の伝説; *Hyōka Yūmei no Densetsu*; *Tsumetaru Yūmei no Densetsu*) is an action-adventure game developed and published by Nintendo for the Wii and GameCube. It is the seventh mainline in the *The Legend of Zelda* series. Originally planned for release on the GameCube in November 2005, *Twilight Princess* was delayed by Nintendo to allow its developers to refine the game, add more content, and port it to the Wii. The Wii version was released alongside the console in North America in November 2006, and in Japan, Europe, and Australia the following month. The GameCube version was also released worldwide in December 2006.

The story focuses on series protagonist Link, who tries to prevent Hyrule from being engulfed by a corrupted parallel dimension known as the Twilight Realm. To do so, he takes the form of both a Hylian and a wolf, and is assisted by a mysterious creature named Midna. The game takes place hundreds of years after *Chronicle of Time* and *Majora's Mask*, an alternate timeline from *The Wind Waker*.

At the time of its release, *Twilight Princess* was critically acclaimed, receiving several Game of the Year awards. As of September 2013, 3.65 million copies of the game have been sold worldwide, making it the best-selling title in the series. In 2011, the Wii version was nominated under the Nintendo Selects label. A high-definition remaster for the Wii U, *The Legend of Zelda: Twilight Princess HD*, was released in March 2016.

Gameplay

See also: *The Legend of Zelda 3 Gameplay*

The Legend of Zelda: Twilight Princess is an action-adventure game focused on combat, exploration, and puzzle-solving. It uses the basic control scheme introduced in *Chronicle of Time*, including context-sensitive action buttons and *L*-triggering (L-triggering on the Wii, a system that allows the player to keep Link's view focused on an enemy or important object while moving, and attacking. Link can walk, run, and attack, and will automatically jump when running off of or reaching for a ledge. Link uses a sword and shield in combat, complemented with secondary weapons and items, including a bow and arrows, a boomerang, and bombs. While L-triggering, projectile-based weapons can be fired at a target without the need for manual aiming.

The context-sensitive button mechanics allow one button to serve a variety of functions, such as taking, opening doors, and pushing, pulling, and throwing objects. The on-screen display shows what action, if any, the button will trigger, determined by the situation. For example, if Link is holding a rock, the context-sensitive button will cause Link to throw the rock if he is moving or targeting an object or enemy, or place the rock on the ground if he is standing still. The GameCube and Wii versions feature several other differences in their controls. The Wii version of the game makes use of the motion sensors and built-in speaker of the Wii Remote. The speaker emits the sound of Link's breathing when charging an arrow. Midna's high when she gives advice to Link, and the series' trademark "Yume" when discovering secrets. The player controls Link's sword by using the Wii Remote. Other actions are triggered using similar gestures with the "Nunchuk". Unique to the GameCube version is the ability for the player to control the camera freely without entering a special "lock-on" mode required by the Wii. However, in the GameCube version, only one of Link's secondary weapons can be equipped at a time, as opposed to four in the Wii version.

The game features nine dungeons—large, contained areas where Link battles enemies, collects items, and solves puzzles. Link requires these dungeons and fulfills a quest or the end of the world to obtain better offensive weapons for the plot. The dungeons are connected by a large overworld, an area which Link can travel on foot, on his horse, *Epona*, or by teleporting with Midna's assistance.

While Link enters the Twilight Realm, he used that corrupts parts of Hyrule, he transforms into a wolf. He is eventually able to transform between his Hylian and wolf forms at will. As a wolf, Link loses the ability to use his sword, shield, or any secondary items. He instead attacks by biting and defends primarily by dodging attacks. However, "Wolf Link" gains several key advantages in combat: he moves faster than he does as a human (though using Epona as a wolf allows and also helps to create more passages and access barred areas), has less exposed senses, including the ability to follow scent trails. He also carries Midna, a small creature who gives him hints, uses an energy ball to attack enemies, helps him jump long distances, and eventually allows him to "surf" to any of several preset locations throughout the overworld. Using Link's wolf senses, the player can see and hear the "whispering spirits" often affected by the Twilight, as well as hear for enemy secret items.

The artificial intelligence (AI) of enemies in *Twilight Princess* is more advanced than that of enemies in *The Wind Waker*. Enemies react to defeated companions and to arrows or slingshot pellets that pass by, and can detect Link from a greater distance than was possible in previous games.

There is very little voice acting in the game, as is the case in most *The Legend of Zelda* titles to date. Link remains silent in conversation, but grunts when attacking or injured and grunts when surprised. His emotions and responses are largely indicated visually by masks and facial expressions. Other characters have similar language-independent vocalizations, including laughter, surprise or fearful exclamations, and screams. *Midna* is voiced by Midna, her most voice acting—her on-screen dialogue is often accompanied by a bubble of pseudo-speech, which was produced by synthesizing English phrases sampled by Japanese voice actress *Adria Katsuno*.

Plot

Twilight Princess takes place three several centuries after *Chronicle of Time* and *Majora's Mask*. The game begins with a youth named Link, who is working as a ranch hand in Ordon Village. One day, the village is attacked by *Midna*, who carry off the village's children with Link in pursuit before he encounters a wolf of Twilight. A Shadow Beast path has beyond the wolf and the Twilight-dimension forest, where he is transformed into a wolf and imprisoned. Link is soon freed by an up-like Twilight creature named Midna, who offers to help him by the city for her occasionally. She guides him to *Twilight Princess*, *Zelda*, who explains that she, the King of the Twilight, without *Twilight's* wolf and forced her to surrender. The corrupted kingdom was overpowered in Twilight, rendering all of inhabitants besides Link and *Zelda* spirits. In order to save Hyrule, Link must first free the Link Spirit by entering the Twilight-realm region and, as a wolf, recovering the Spirit's light from the Twilight before the Link is made. Once released, each Spirit returns Link to his Hylian form.

During his time, Link also helps Midna acquire the Forest Shadow, fragments of a tree containing powerful dark magic. In return, she aids Link in saving Ordon Village's children and restoring the country of Hyrule, the *Castle of Twilight*, and the *Castle of Lamps*. After restoring the Light Spirits and obtaining the Forest Shadow, Link and Midna are ambushed by Zant, who reduces Midna of the fragments. She releases him for



Methodology

Metrics

1. EM: Exact match of answer
2. F1-score:

```
precision = n_common_tokens / len(pred_tokens)
recall = n_common_tokens / len(truth_tokens)
f1_score = 2 * (precision * recall) / (precision + recall)
```

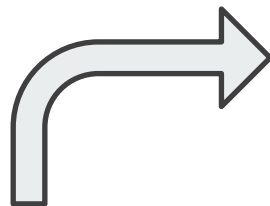
<https://torchmetrics.readthedocs.io/en/stable/text/squad.html>

Experimental setup

Dataset

```
{
  "version": "v2.0",
  "data": [
    {
      "title": "Normans",
      "paragraphs": [
        {
          "context": "The Normans (Norman: Nourmands; French: Normands; ...",
          "qas": [
            {
              "question": "In what country is Normandy located?",
              "id": "56dde6b9a695914005b9628",
              "answers": [
                {
                  "text": "France",
                  "answer_start": 159
                },
                {
                  "text": "...",
                  "answer_start": 159
                }
              ],
              "is_impossible": false
            },
            {
              "question": "The Normans (Norman: Nourmands; French: Normands; ...",
              "id": "56dde6b9a695914005b9628",
              "answers": [
                {
                  "text": "France",
                  "answer_start": 159
                },
                {
                  "text": "...",
                  "answer_start": 159
                }
              ],
              "is_impossible": false
            }
          ]
        }
      ]
    }
  ]
}
```

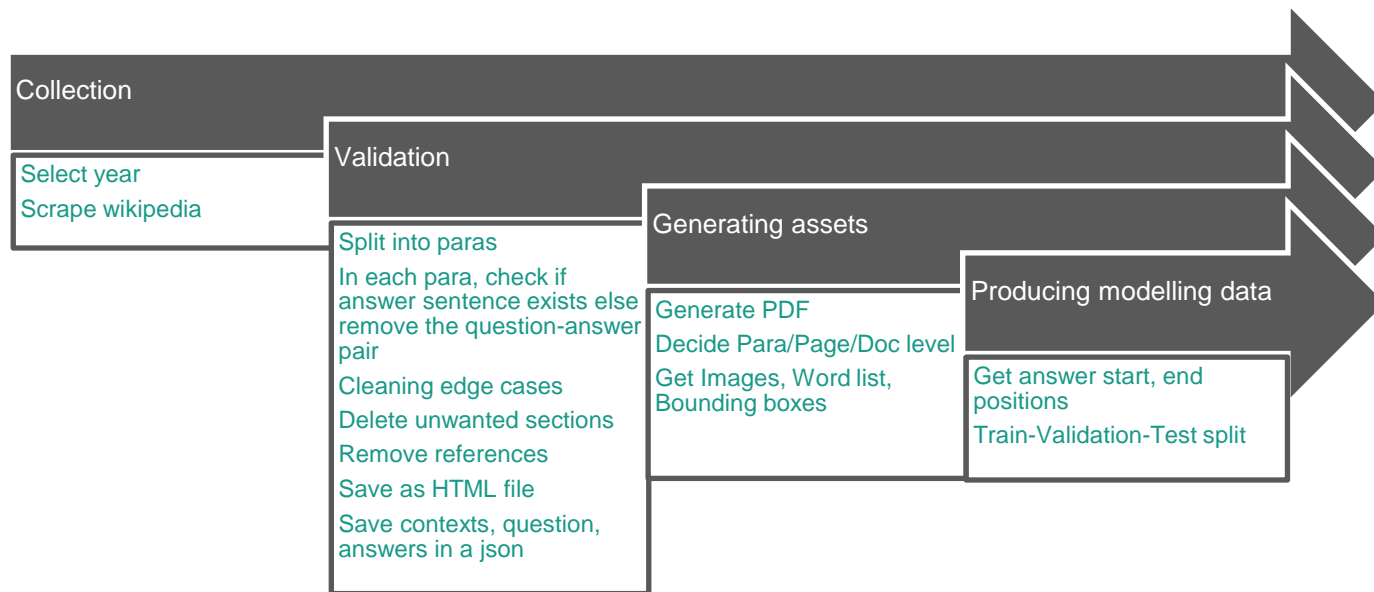
SQuAD 2.0 data



Required Visual data

Experimental setup

Dataset





Experimental setup

Data efficiency

1. Slice the train and validation sets into 1%, 2%, 5%, 10%, 50%, 100%
2. Train, validate models on these datasets
3. Test them on the same test set.
4. Generate and visualize the metrics.



Experimental setup

Domain adaptation

1. Gather categories of each wiki article
2. Encode categories (including title) as BERT embeddings
3. Cluster wiki titles using these embeddings (Distance between 2 points = cosine similarity between 2 embeddings)
4. Split domains to train, validation and test sets (e.g., train:valid:test = 5:1:2)
5. Train, validate and test each model

Categories: Beyoncé | 1981 births | Living people | 20th-century African-American businesspeople
20th-century African-American women | 20th-century American businesspeople | 20th-century American businesswomen
20th-century American singers | 20th-century American women singers | 21st-century African-American women
21st-century American actresses | 21st-century American businesspeople | 21st-century American businesswomen
21st-century American singers | 21st-century American women singers | Actresses from Houston | African-American actresses
African-American artists | African-American choreographers | African-American dancers | African-American fashion designers
American fashion designers | African-American female dancers | African-American women rappers | American women rappers
African-American women singers | African-American feminists | American feminists | African-American Methodists
African-American record producers | African-American women in business | African-American women writers
American choreographers | American contemporary R&B singers | American cosmetics businesspeople
American fashion businesspeople | American women pop singers | American film actresses | American mezzo-sopranos
American music publishers (people) | American music video directors | American people of Acadian descent
American people of Creole descent | American retail chief executives | American television actresses | American United Methodists
American voice actresses | American women business executives | American women philanthropists
American women record producers | Black Lives Matter people | Brit Award winners | Businesspeople from Houston
Businesspeople from Texas | Columbia Records artists | Dance-pop musicians | Destiny's Child members
Female music video directors | Feminist musicians | Gold Star Records artists | Grammy Award winners
Grammy Award winners for rap music | High School for the Performing and Visual Arts alumni | Ivor Novello Award winners | Jay-Z
Solange Knowles | Louisiana Creole people | MTV Europe Music Award winners | MTV Video Music Award winners
Music video codirectors | Musicians from Houston | Musicians from Texas | NME Awards winners | Parkwood Entertainment artists
Record producers from Texas | Shoe designers | Singers from Texas | Singers with a four-octave vocal range | Texas Democrats
Women who experienced pregnancy loss | World Music Awards winners | Writers from Houston

<https://en.wikipedia.org/wiki/Beyonc%C3%A9>



Results

Data link:

<https://drive.google.com/file/d/1gCF-AyRA2tsT1lWqAZpKSaXmyAXxTGC-/view?usp=sharing>

Visual SQuAD dataset stats

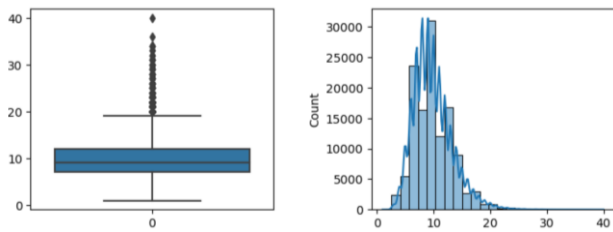
- Number of documents: 457
- Number of questions: 124827
- Number of unique questions: 124692
- Number of distinct words in questions: 74689
- Number of answerable questions: 77510
- Number of answers: 81073
- Number of distinct words in answers: 51010
- Number of paragraphs in dataset: 18850

Visual SQuAD Dataset EDA

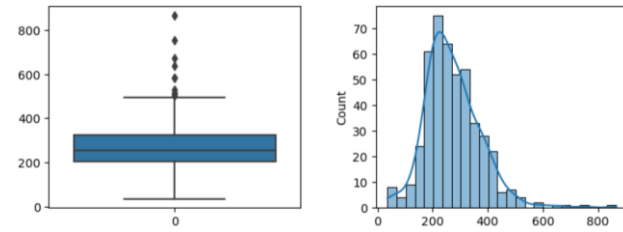
First 2 words of questions



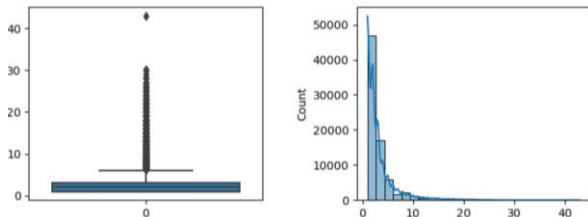
Question lengths



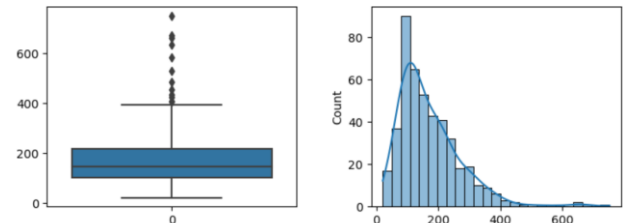
Questions per document



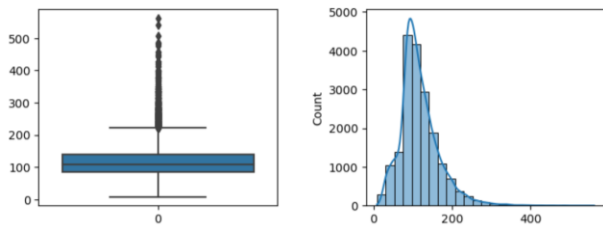
Answer lengths



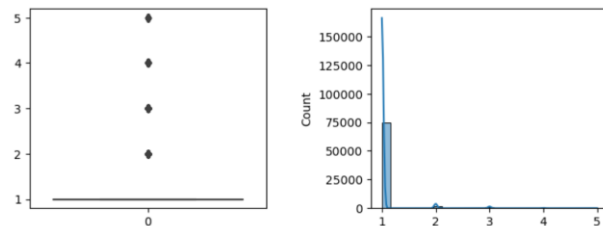
Answerable questions per document



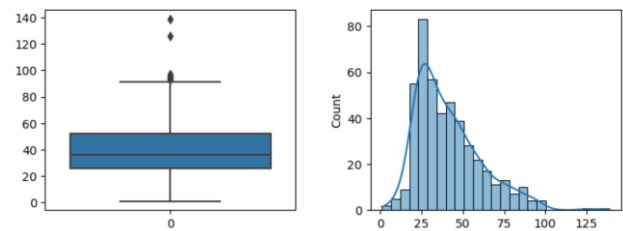
Paragraph lengths



Answers per question



Paragraphs per doc

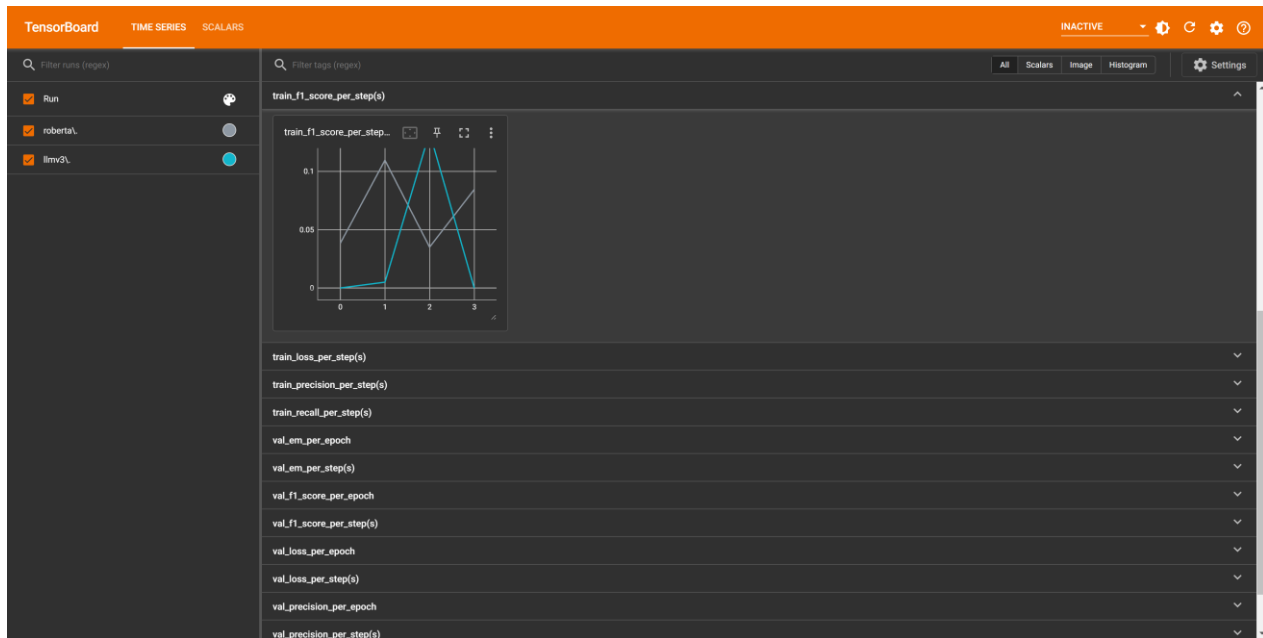


Results

Data efficiency

Currently working on
para level.

Code for RoBERTa,
LLMv3 tests is ready.





Next steps

Data efficiency

Run experiments

1. Now
2. Add impossible questions
3. Doc level
4. Add extra models

Domain adaptation

1. Create domains
2. Select domains after EDA
3. Run experiments



Major contributions

1. New dataset for extractive Document QA
2. Fair comparisons of Textual and Multimodal QA models



Questions?