

۱. پایگاه داده MSRA بزرگترین مجموعه برای توجه بصری است و نقشه‌های توجه شیء برجسته را توسط ۳ الی ۹ ناظر ارائه می‌دهد. تنوع زیادی در محتوای تصاویر وجود دارد و به طور گسترده در شناسایی شیء برجسته و بخشبندی به کار می‌رود. [پایان نامه، خانم مینا کریمیان، تشخیص برجستگی تصاویر با ورودی چند مقیاسی در یادگیری عمیق، ۱۴۰۰]

۲. مجموعه داده DUT-OMRON مجموعه‌ای از ۵۱۶۸ تصویر با کیفیت بالا، محتوای متنوع و چالش برانگیز است. برجسب‌های تمرکز چشم در سطح پیکسل توسط ۵ ناظر جمع‌آوری شده است. تصاویر این مجموعه داده دارای یک یا چند شیء برجسته هستند و در مقایسه با مجموعه‌های دیگر، پس‌زمینه شلوغ و پیچیده‌تری دارند. [https://paperswithcode.com/dataset/dut-omron]

۳. مجموعه داده‌های برجستگی صحنه‌های گسترده ECSSD حاوی صحنه‌های پیچیده‌ای است که بافت‌ها و ساختارهای مشترک تصاویر دنیای واقعی را ارائه می‌دهد. این مجموعه داده از دو مجموعه داده در دسترس عموم (PASCAL VOC و BSD) و اینترنت بدست آمده و شامل ۱۰۰۰ تصویر پیچیده با زمینه‌های متنوع است. تصاویر مبنای صحت به طور میانگین از ۵ ناظر انسانی ایجاد شده است. [پایان نامه، خانم مینا کریمیان، تشخیص برجستگی تصاویر با ورودی چند مقیاسی در یادگیری عمیق، ۱۴۰۰]

۴. مجموعه داده K5 (۵ هزار کلمه) ای IIIT از جستجوی تصویر گوگل جمع‌آوری شده است. برای جمع‌آوری تصاویر از نتایج پرس و جو مانند بیل‌بورد، تابلو، شماره خانه، پلاک نام خانه و پوستر فیلم استفاده شده. مجموعه داده شامل ۵۰۰۰ تصویر کلمه برش خورده (crop) از متن‌های صحنه و تصاویر دیجیتالی، تولید شده است. مجموعه داده به بخش آموزش و تست تقسیم می‌شود. [پایان نامه، خانم کیمیا نورعلی، استخراج متن از صحنه با سیستم پرسش و پاسخ تصویری، ۱۴۰۰]

۵. مجموعه داده SVT شامل تصاویر خیابانی جمع‌آوری شده از منظره خیابانی گوگل (Google Street View) است. بعضی از این تصاویر حاوی نویز هستند یا کیفیت پائینی دارند. SVT شامل ۲۵۷ تصویر برای آموزش و ۶۴۷ تصویر برای ارزیابی است. [پایان نامه، خانم کیمیا نورعلی، استخراج متن از صحنه با سیستم پرسش و پاسخ تصویری، ۱۴۰۰]

۶. مجموعه داده TextVQA شامل ۴۵۳۳۶ سوال پرسیده شده به وسیله افراد روی ۲۸۴۰۸ تصویر برای آموزش می‌باشد، که از مجموعه داده‌های تصاویر باز از گروه‌های مختلف شامل متن مانند: بیل‌بوردها، تابلوهای رانندگی، تابلو اعلانات است. پرسش‌ها در مجموعه داده به خواندن و استدلال درباره متن در تصویر نیاز دارد. هر جفت پرسش و تصویر ۱۰ پاسخ درستی زمینه دارد که توسط افراد پاسخ داده شده‌اند. به علاوه شامل ۳۳۵۳ تصویر و ۵۷۳۴ پرسش برای آزمایش است. [پایان نامه، خانم کیمیا نورعلی، استخراج متن از صحنه با سیستم پرسش و پاسخ تصویری، ۱۴۰۰]

۷. مجموعه داده VQA شامل ۲۰۰۰۰۰ تصویر و یک میلیون پرسش و یازده میلیون پاسخ است. [پایان نامه، خانم کیمیا نورعلی، استخراج متن از صحنه با سیستم پرسش و پاسخ تصویری، ۱۴۰۰]

۸. VQA-v2 متداول ترین مجموعه داده معیار VQA است. این دیتاست شامل جفت‌های پرسش-پاسخ‌هایی است که توسط انسان برجسبگذاری شده‌است و مربوط به تصاویر مجموعه داده MS-COCO، با ۳ سؤال در هر تصویر و ۱۰ پاسخ در هر سؤال

است. مجموعه داده به سه تقسیم می شود: آموزش ($k80$) تصویر و $k444$ جفت پرسش و پاسخ). Val (اعتبار سنجی) ($k40$) تصویر و $k14$ جفت پرسش و پاسخ؛ و تست ($k80$) تصویر و $k448$ جفت پرسش و پاسخ). [1]

۹. مجموعه داده COCO، شامل ۲۵۰۰۰۰۰ برچسب روی ۳۲۸۰۰۰ تصویر است. [پایان نامه، خانم کیمیا نورعلی، استخراج متن از صحنه با سیستم پرسش و پاسخ تصویری، ۱۴۰۰]

۱۰. مجموعه داده OpenImage شامل حدود ۹ میلیون تصویر است که با برچسب ها و جعبه های محدودکننده حاشیه نویسی شده اند. [پایان نامه، خانم کیمیا نورعلی، استخراج متن از صحنه با سیستم پرسش و پاسخ تصویری، ۱۴۰۰]

۱۱. مجموعه داده DAQUAR (مجموعه داده برای پاسخگویی به پرسش در تصاویر دنیای واقعی) اولین مجموعه داده برای وظیفه VQA بود که توسط مالدینوفسکی و فریتز معرفی شد. آنها از مجموعه داده NYU-Depth V2 استفاده کرده اند که شامل ۱۴۴۹ تصویر RGBD (۷۹۵ تصویر آموزشی و ۶۵۴ تصویر آزمایشی) است و ۱۲۴۶۸ جفت پرسش و پاسخ (۶۷۹۴ جفت برای آموزش و ۵۶۷۴ جفت برای آزمایش). [پایان نامه، مهسا رادین مهر، آگوست ۲۰۱۹]

۱۲. مجموعه داده VQA بر خلاف مجموعه داده DAQUAR که پاسخ های آن از یک دنیای بسته از پیش تعریف شده از ۱۶ رنگ اصلی یا ۸۹۴ دسته شی می آید، Antol و همکاران. در مجموعه داده VQA یک مجموعه داده پاسخگویی به سؤالات تصویری با پایان باز و آزاد ارائه کرد. از دو زیر مجموعه داده تشکیل شده است :

(آ) تصاویر دنیای واقعی: آنها از شیء مشترک میکروسافت در زمینه یا MS-COCO استفاده کرده اند که حاوی ۲۰۴۷۲۱ تصویر (۱۲۳۲۸۷ تصویر آموزشی و اعتبارسنجی و ۸۱۴۳۴ تصویر آزمایشی) است.

(ب) صحنه های انتزاعی: شامل ۵۰۰۰۰ صحنه کلیپ پارت است و برای تمرکز بر روی کار استدلالی سطح بالا ارائه شده است. دارای ۱۰۰ شی و ۳۱ حیوان در حالت های مختلف است. [پایان نامه، مهسا رادین مهر، آگوست ۲۰۱۹]

۱۳. دیتاست MNIST از ۷۰۰۰۰ عکس تشکیل شده که شامل اعداد نوشته شده توسط دانش آموزان دبیرستانی و کارمندان سازمان سرشماری آمریکا هست. هر عکس با شماره ای که توی عکس هست لیبل گذاری شده. این دیتاست انقدر مطالعه شده که بعضی اوقات بهش میگن "hello world" ماشین لرنینگ! هر وقت افراد یک الگوریتم Classification جدید پیدا میکنن، اول روی این دیتاست امتحانش میکنن. 70000 عکس وجود داره که هر عکس ۷۸۴ تا فیچر داره. این به این بخاطر هست که هر عکس ۲۸ * ۲۸ پیکسل هست و هر فیچر نشون دهنده شدت هر پیکسل هست. مقدار این شدت از ۰ (سفید) تا ۲۵۵ (سیاه) متغیر هست. [<https://vrgl.ir/6BF1c>]

۱۴. مجموعه ارقام دستنویس هدی که اولین مجموعه ی بزرگ ارقام دستنویس فارسی است، مشتمل بر ۱۰۲۳۵۳ نمونه دستنویسته سیاه سفید است. این مجموعه طی انجام یک پروژه ی کارشناسی ارشد درباره بازشناسی فرمهای دستنویس تهیه شده است. داده های این مجموعه از حدود ۱۲۰۰۰ فرم ثبت نام آزمون سراسری کارشناسی ارشد سال ۱۳۸۴ و آزمون کاردانی پیوسته ی دانشگاه جامع علمی کاربردی سال ۱۳۸۳ استخراج شده است.

تعداد نمونه ها در هر کلاس:

رقم ۰	رقم ۱	رقم ۲	رقم ۳	رقم ۴
۱۰۰۷۰	۱۰۳۳۰	۹۹۲۳	۱۰۳۳۴	۱۰۳۳۳

رقم ۵	رقم ۶	رقم ۷	رقم ۸	رقم ۹
۱۰۱۱۰	۱۰۲۵۴	۱۰۳۶۳	۱۰۲۶۴	۱۰۳۷۱

[مجموعه-ارقام-دستنویس-هدی/مجموعه-داده/ <http://farsiocr.ir>]

۱۵. CIFAR10 / CIFAR100: این دوتا دیتاست در اصل یکی هستند و تفاوت اونها در تعداد دسته های اونهاست. این دیتاست شامل تصاویر رنگی با اندازه ۳۲*۳۲ هست و شامل ۶۰ هزار تصویر میشود.

۱۶. OpenfMRI حاوی داده‌هایی از دیتاست‌های تصویربرداری است که از دستگاه‌های MRI به دست می‌آیند تا پژوهش‌های مربوطه را تقویت و به تشخیص و آموزش باکیفیت‌تر کمک کنند. این سرویس شامل ۹۵ دیتاست حاوی ۳۳۷۲ مورد بوده و همچنان که محققان داده‌های خود را منتشر می‌کنند، به داده‌های این دیتاست افزوده می‌شود. [https://hooshio.com/?p=7376]

۱۷. تصاویر پزشکی CT (پرتونگاری رایانه‌ای): این دیتاست، دیتاست کوچکی است، اما منحصراً به بحث سرطان می‌پردازد. داده‌های این دیتاست شامل تصاویری می‌شوند که بر اساس سن، حالت و تگ‌های متضاد برچسب خورده‌اند. زمانی که این تصاویر باکیفیت با داده‌های آموزشی همراه شوند، می‌توانند به کشفیات و پیشرفت‌های بزرگی منجر گردند. [https://hooshio.com/?p=7376]

۱۸. Deep Lesion: داده‌های دیتاست Deep Lesion متنوع و گسترده هستند و یکی از بزرگ‌ترین دیتاست‌های تصویری موجود در حال حاضر را تشکیل می‌دهند. تصاویر سی‌تی اسکن گرفته‌شده مؤسسات بهداشتی درمانی ملی (NIH) به منظور کمک به صحت بیشتر در تشخیص و مستندسازی جراحات در این دیتاست گردآوری شده تا مورد استفاده سایرین قرار گیرند. داده‌های این دیتاست اطلاعات مربوط به بیش از ۳۲۰۰۰ جراحی و آسیب مربوط به ۴۰۰۰ بیمار را در برمی‌گیرند. [https://hooshio.com/?p=7376]

۱۹. مجموعه داده کفش: نام کامل این مجموعه UT Zappos50K (UT-Zap50K) می‌باشد که حاصل کار گروه تحقیقاتی مهندسی داده دانشگاه تگزاس آمریکا می‌باشد.

این مجموعه داده شامل بیش از ۵۰ هزار تصویر از ۴ مدل مختلف از کفش های زنانه و مردانه از سایت Zappos.com است. تصاویر به ۴ دسته کلی تقسیم شده اند: -کفش -صندل -چکمه -دمپایی مامی عکس ها بر روی پس زمینه سفید قرار داده شده اند تا به راحتی برای سیستم های هوش مصنوعی و استخراج ویژگی ها قابل تشخیص باشند. [<https://pythoniha.ir/?p=11347>]

۲۰. JACFEE: این پایگاه داده شامل ۲۱۳ تصویر از ۷ حالت صورت (۶ حالت صورت + ۱ حالت طبیعی) است که از ۱۰ مدل زن ژاپنی ثبت شده است. هر تصویر توسط ۶۰ نفر از افراد ژاپنی به ۶ صفت حساسی رتبه بندی شده است. حالت های چهره شامل حالت طبیعی، شاد، عصبانی، انزجار، ترس، غم انگیز و تعجب است. برای هر حالت چهره از هر فرد دو نوع ثبت شده است. در مجموع ۲۱۳ عدد تصاویر از حالت چهره در این پایگاه داده وجود دارد که اندازه هر تصویر 256×256 است. [<https://onlinebme.com/facial-expression/>]

References

- [1] Z. a. Y. J. a. C. Y. a. T. D. a. T. Q. Yu, "Deep Modular Co-Attention Networks for Visual Question Answering," *yu2019mcan*, pp. 6281--6290, 2019.