

به نام خدا

لیست دیتاست های تصویر:

1- VisualQA: در میان مجموعه داده های تصویری، VisualQA یا VQA مجموعه داده جدیدی است که حاوی سوالات باز درباره تصاویر است. پاسخ به این سوالات نیازمند درک بینش، زبان و دانش عام است.

- 265016 تصویر (صحنه های انتزاعی)
- حداقل 3 سوال (به طور متوسط 5.4 سوال) در هر تصویر
- 10 پاسخ حقیقت پایه در هر سوال
- 3 پاسخ قابل قبول (اما به احتمال زیاد نادرست) در هر سوال
- معیار ارزیابی خودکار

Who is wearing glasses?

man



woman



2- CompCars: مجموعه داده جامع ماشین ها (CompCars) حاوی داده هایی از دو سناریو است، از جمله تصاویری از طبیعت وب و طبیعت نظارت. داده های طبیعت وب شامل 163 خودرو با 1716 مدل خودرو است. در مجموع 136726 تصویر از کل خودروها و 27618 تصویر از قطعات خودرو ثبت شده است. تصاویر کامل خودرو با جعبه های مرزی و دیدگاه ها برچسب گذاری شده اند. هر مدل خودرو دارای پنج ویژگی است که شامل حداکثر سرعت، جابجایی، تعداد درها، تعداد صندلی ها و نوع خودرو می شود. داده های نظارتی-طبیعت شامل 50000 تصویر خودرو است که در نمای جلو گرفته شده است.

مجموعه داده را می توان برای وظایف زیر استفاده کرد:

- طبقه بندی ریزدانه
- پیش بینی صفت
- تایید مدل ماشین

این مجموعه داده همچنین می تواند برای کارهای دیگری مانند رتبه بندی تصویر، یادگیری چند وظیفه ای و بازسازی سه بعدی استفاده شود.



3- مجموعه داده تصاویر حیوانات خانگی Oxford-IIIT: این مجموعه داده تصاویر حیوان خانگی دارای 37 دسته با 200 تصویر برای هر کلاس است. تصاویر بر اساس مقیاس، ژست و نور متفاوت هستند، و دارای حاشیه نویسی حقیقت زمینی مربوط به نژاد، بازگشت سرمایه سر، و تقسیم بندی trimap در سطح پیکسل هستند.

4- CIFAR-10: مجموعه داده CIFAR-10 شامل 60000 تصویر رنگی پایین (32x32) در 10 کلاس مختلف است. [3] 10 کلاس مختلف نشان دهنده هواپیما، اتومبیل، پرنده، گربه، گوزن، سگ، قورباغه، اسب، کشتی و کامیون است. 6000 تصویر از هر کلاس وجود دارد. از آنجایی که تصاویر در CIFAR-10 با وضوح پایین (32x32) هستند، این مجموعه داده می تواند به محققان اجازه دهد تا الگوریتم های مختلف را به سرعت امتحان کنند تا ببینند چه چیزی کار می کند.

5- تشخیص صحنه داخلی یا scene recognition dataset: تشخیص صحنه داخلی یک مشکل باز چالش برانگیز در دید سطح بالا است. اکثر مدل های تشخیص صحنه که برای صحنه های بیرونی به خوبی کار می کنند، در حوزه داخلی عملکرد ضعیفی دارند. مشکل اصلی این است که در حالی که برخی از صحنه های داخلی (مانند راهروها) را می توان به خوبی با ویژگی های فضایی جهانی مشخص کرد، برخی دیگر (مثلاً کتابفروشی ها) با اشیایی که در خود دارند بهتر مشخص می شوند. به طور کلی، برای رسیدگی به مشکل تشخیص صحنه های داخلی، به مدلی نیاز داریم که بتواند از اطلاعات تبعیض آمیز محلی و جهانی بهره برداری کند. پایگاه داده شامل 67 دسته بندی داخلی و در مجموع 15620 تصویر است. تعداد تصاویر در دسته ها متفاوت است، اما حداقل 100 تصویر در هر دسته وجود دارد. تمامی تصاویر با فرمت jpg هستند. تصاویر ارائه شده در اینجا فقط برای اهداف تحقیقاتی است.

6- اشیاء خانگی: شامل اشیایی که معمولاً از اطراف خانه پیدا می شوند. این مجموعه داده شامل دو مجموعه عکس از اشیاء خانگی است برای توسعه سیستم های تشخیص اشیاء برای یک ربات خانگی. اولین مجموعه داده با اشیاء از یک سوپرمارکت محلی ایجاد شد. برندهای محصول از برزیل معمولی هستند. مجموعه داده دوم از اشیاء مسابقه RoboCup @ Home 2018 OPL تشکیل شده است.

- مجموعه 1: 166 تصویر حاشیه نویسی شده با 1028 شیء از 13 کلاس مقابل: غلات - شیر کاکائو - هاینکن - مردآهنی-پزشکی- بطری شیر - جعبه شیر - هیولا - آب بنفش -آبمیوه قرمز -شامپو -جعبه چای - آبمیوه زرد و همچنین 28 ویدیو برای آزمایش وجود دارد که با چندین گوشی هوشمند گرفته شده است.

- مجموعه 2: 388 تصویر حاشیه نویسی شده با 1737 شیء از 20 کلاس زیر: سیب - سید - غلات - شکلات نوشیدنی - cloth_opl - کک - کراکرها - آب انگور - رشته فرنگی - نارنجی - آب پرتقال - فلفل قرمز - چپیس سیب زمینی - پرینگ ها سوسیس و کالباس - اسکرابی - sponge_opl - جین - سینی و همچنین یک ویدیوی طولانی و 398 تصویر بدون حاشیه برای آزمایش وجود دارد.

7- Celebfaces: مجموعه داده‌های ویژگی‌های CelebFaces (CelebA) یک مجموعه داده ویژگی‌های چهره در مقیاس بزرگ با بیش از 200 هزار عکس افراد مشهور است که هر کدام دارای 40 حاشیه‌نویسی ویژگی است. تصاویر موجود در این مجموعه داده، تغییرات ژست‌های بزرگ و درهم ریختگی پس زمینه را پوشش می‌دهند. CelebA دارای تنوع زیاد، مقادیر زیاد و حاشیه نویسی‌های غنی از جمله است. 10177 تعداد هویت، 202599 تعداد تصویر چهره و 5 مکان برجسته، 40 حاشیه نویسی ویژگی‌های باینری در هر تصویر. مجموعه داده را می‌توان به عنوان مجموعه‌های آموزشی و آزمایشی برای وظایف بینایی رایانه زیر به کار برد: تشخیص ویژگی چهره، تشخیص چهره، تشخیص چهره، محلی سازی نقطه عطف (یا قسمت چهره)، و ویرایش و ترکیب چهره.

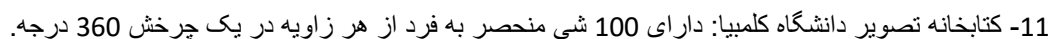


8- مجموعه داده سگ‌های استنفورد یا Stanford Dogs Dataset: مجموعه داده Stanford Dogs شامل تصاویری از 120 نژاد سگ از سراسر جهان است. این مجموعه داده با استفاده از تصاویر و حاشیه نویسی از ImageNet برای طبقه‌بندی تصاویر ریز دانه ساخته شده است. محتویات این مجموعه داده:

- تعداد دسته‌ها: 120
- تعداد تصاویر: 20580
- حاشیه نویسی: برچسب‌های کلاس، جعبه‌های محدود کننده

9- مجموعه داده تصاویر باز Fishnet: برای آموزش الگوریتم‌های تشخیص چهره ایده‌آل، مجموعه داده‌های تصاویر باز Fishnet دارای 35000 تصویر ماهیگیری است که هر کدام شامل 5 جعبه محدودکننده است.

balcony, stairs, facade, iron, door, interior design, gate, architecture, handrail, baluster, window, arch

[illegible]

- تقسیم بندی شی
- شناخت در زمینه
- تقسیم بندی چیزهای سوپر پیکسلی

- تصاویر K330 (بیش از 200 هزار برچسب دار)
- 1.5 میلیون نمونه شی
- 80 دسته شی
- 91 دسته موارد
- 5 زیرنویس در هر تصویر
- 250000 نفر با نکات کلیدی

Dataset examples



13- Lego Bricks: این مجموعه داده تصویری شامل 12700 تصویر از آجرهای لگو است که هر کدام قبلاً طبقه بندی و با استفاده از آنها ارائه شده است.

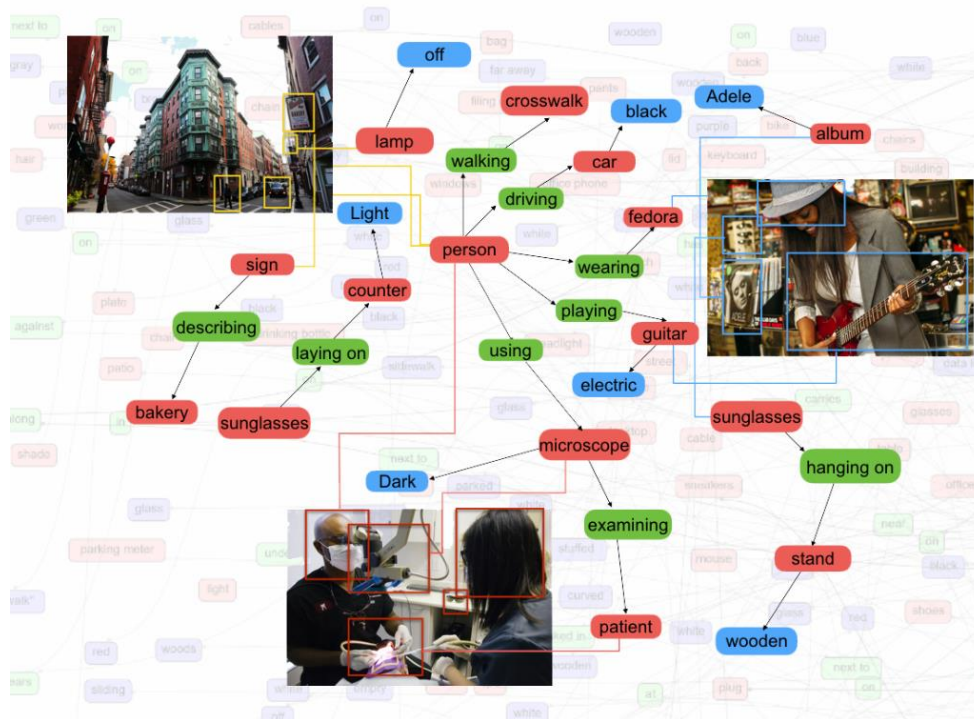
14- Labelme: یکی از مجموعه داده‌های تصویری علوم کامپیوتر MIT که در ارتباط با آزمایشگاه هوش مصنوعی (CSAIL) ایجاد شده است، این یکی دارای 187240 تصویر، 62197 تصویر قبلاً مشروح شده در 658992 شی برچسبگذاری شده است.

15- ImageNet: سازماندهی شده مطابق با سلسله مراتب WordNet، ImageNet یکی از مجموعه داده‌های تصویری برای همه الگوریتم‌های جدید است. هر گره در سلسله مراتب WordNet در صدها هزار تصویر به تصویر کشیده شده است.



16- VisualGenome: Visual Genome یک مجموعه داده، یک پایگاه دانش، تلاشی مداوم برای اتصال مفاهیم تصویر ساختاریافته به زبان است.

- 108,077 تصاویر
- 5.4 میلیون توصیف منطقه
- 1.7 میلیون پاسخ سوال تصویری
- 3.8 میلیون نمونه شی
- 2.8 میلیون ویژگی
- 2.3 میلیون رابطه
- همه چیز در Wordnet Synsets نگاشت شده است



17- Youtube-8M: این مجموعه داده در مقیاس بزرگ با میلیون ها شناسه ویدیوی YouTube، همراه با حاشیه نویسی بیش از 3800 موجودیت بصری برچسب گذاری شده است. موجودیت هایی که مانند فیلم ها یا سریال های تلویزیونی قابل بومی سازی نیستند، مستثنی هستند.

- 237 هزار برچسب بخش تأیید شده توسط انسان
- 1000 کلاس
- 5.0 میانگین بخش ها / ویدئو

18- FERET: FERET (پایگاه داده فناوری تشخیص چهره). پایگاه داده FERET که شامل 14126 تصویر از صورت 1199 فرد است. بخش توسعه پایگاه داده FERET به بیش از 100 گروه خارج از برنامه اصلی FERET توزیع شده است.

19- چهره‌های برجسب‌گذاری شده در طبیعت: مجموعه‌ای از تصاویر با عنوان بیش از حد مناسب، چهره‌های برجسب‌گذاری‌شده در طبیعت دارای ۱۳۰۰۰ تصویر برجسب‌گذاری شده از چهره‌های انسان است. مخصوصاً برای تشخیص چهره مفید است.

20- مکان‌ها: این مجموعه داده تصویر صحنه محور شامل 205 دسته صحنه منحصر به فرد با 2.5 میلیون تصویر است که بر اساس یک دسته برچسب‌گذاری شده‌اند.

21- گل‌ها: این مجموعه داده تصویری با گل‌هایی که معمولاً در سراسر بریتانیا یافت می‌شوند، شامل بیش از 102 دسته مختلف است که هر گل از حالت‌ها و تغییرات نوری متفاوت دیده می‌شود.

