بسم الله الرحمن الرحیم

شرکت مهندسی نرم افزاری هلو

گزارش مربوط به تولید مجموعه های داده-ای مربوط به Action Recognition

کاری از امیرعلی نسیمی

- **مقدمه**

با توجه به اهمیت پردازش فایل های ویدئویی مربوط به شناسایی Command، ایده‌ی شناسایی Action در فایل های ویدئویی مورد بررسی قرار گرفت. با توجه به این مسئله، یکسری از مشهورترین مجموعه‌های داده‌ای در قالب فایل ویدئویی در این گزارش معرفی شده است.

- **UCF101 (UCF101 Human Actions dataset)**

مجموعه داده UCF101 توسعه ای از UCF50 بوده و شامل ۱۳۳۲۰ کلیپ ویدیویی است که در ۱۰۱ دسته طبقه بندی شده‌اند. این ۱۰۱ دسته را می توان به ۵ نوع (حرکت بدن، تعامل انسان و انسان، تعامل انسان و شی، نواختن آلات موسیقی و ورزش) طبقه بندی کرد. طول زمانی کل کلیپ های ویدیویی مذکور به بیش از ۲۷ ساعت ضبط شده است. تمامی ویدیوها از یوتیوب جمع آوری شده و دارای نرخ فریم ثابت ۲۵ فریم در ثانیه با رزولوشن ۳۲۰ × ۲۴۰ هستند.

- **Kinetics (Kinetics Human Action Video Dataset)**

مجموعه داده‌ای Kinetics یک مجموعه داده با مقیاس بزرگ و با کیفیت بالا برای تشخیص اقدامات انسانی در انواع ویدیو می‌باشد. مجموعه داده‌ای مذکور شامل حدود ۵۰۰۰۰۰ کلیپ ویدیویی است که ۶۰۰ کلاس عمل انسانی را با حداقل ۶۰۰ کلیپ ویدیویی برای هر کلاس اکشن پوشش می دهد. هر کلیپ ویدیویی حدود ۱۰ ثانیه بوده و با یک کلاس اکشن برچسب گذاری شده است. فیلم ها از یوتیوب جمع آوری شده‌اند.

- **HMDB51**

مجموعه داده HMDB51 مجموعه بزرگی از ویدیوهای واقعی از منابع مختلف، از جمله فیلم‌ها و ویدیوهای موجود در اینترنت است. مجموعه داده مذکور از ۶۷۶۶ کلیپ ویدیویی بصورت ۵۱ دسته اکشن انسانی (مانند "پرش"، "بوسه" و "خنده") تشکیل شده است که هر دسته شامل حداقل ۱۰۱ کلیپ می‌باشد. جهت ارزیابی نسبت ۷۰ به ۳۰ برای هر کلاس رعایت شده است.

- **ActivityNet**

مجموعه داده ActivityNet شامل ۲۰۰ نوع فعالیت مختلف و در مجموع ۸۴۹ ساعت ویدیو جمع آوری شده از YouTube است. ActivityNet بزرگترین معیار برای تشخیص فعالیت-های زمانی تا به امروز - از نظر تعداد دسته های فعالیت و تعداد ویدیوها بوده که این کار را به ویژه چالش برانگیز می کند. نسخه ۱.۳ مجموعه داده در مجموع شامل ۱۹۹۹۴ ویدیو بدون ترمیم شده است و به سه زیرمجموعه مجزا، آموزش، اعتبارسنجی و آزمایش با نسبت ۲:۱:۱ تقسیم شده است. به طور متوسط، هر دسته فعالیت دارای ۱۳۷ ویدیوی بدون برش است. هر ویدیو به طور متوسط دارای ۱.۴۱ فعالیت است که با مرزهای زمانی حاشیه نویسی شده است. حاشیه‌نویسی‌های واقعی ویدیوهای آزمایشی عمومی نیستند.

- **Charades**

مجموعه داده Charades از ۹۸۴۸ ویدیو از فعالیت‌های روزانه در داخل خانه با تعداد فریم-های ۳۰ در ۱ ثانیه تشکیل شده است که شامل ۴۶ کلاس شی در ۱۵ نوع صحنه داخلی و حاوی برچسب‌هایی از ۳۰ عمل است که منجر به ۱۵۷ کلاس اکشن می‌شود. هر ویدیو در این مجموعه داده با توضیحات متن آزاد متعدد، برچسب‌های کنش، فواصل عمل و کلاس‌های اشیاء متقابل حاشیه‌نویسی می‌شود. به ۲۶۷ کاربر مختلف جمله ای ارائه شد که شامل اشیاء و اعمال از یک

واژگان ثابت است و آنها فیلمی را با اجرای این جمله ضبط کردند. در مجموع، مجموعه داده شامل ۶۶۵۰۰ حاشیه نویسی زمانی برای ۱۵۷ کلاس عمل، ۴۱۱۰۴ برچسب برای ۴۶ کلاس شی و ۲۷۸۴۷ شرح متنی ویدیوها است. در تقسیم استاندارد ۷۹۸۶ فیلم آموزشی و ۱۸۶۳ ویدیوی تأیید وجود دارد.

- **MPII (MPII Human Pose)**

مجموعه داده‌های MPII Human Pose برای تخمین ژست تک نفره از حدود ۲۵ هزار تصویر متشکل از ۱۵هزار نمونه آموزشی، ۳ هزار نمونه‌ی Validation و ۷ هزار نمونه‌ی آزمایشی می‌باشد (دسترسی به برچسب ها محدود می‌باشد). تصاویر مربوطه از ویدئوهای یوتیوب گرفته شده است که ۴۱۰ فعالیت مختلف انسان را پوشش داده و ژست ها به صورت دستی با حداکثر ۱۶ مفصل بدن برچسب زده شده‌اند.

- **NTU RGB+D**

NTU RGB+D یک مجموعه داده در مقیاس بزرگ برای تشخیص اقدامات انسانی است. این مجموعه شامل ۵۶۸۸۰ نمونه از ۶۰ کلاس بوده که از ۴۰ نفر جمع آوری شده است. این اعمال را می توان به طور کلی به سه دسته تقسیم کرد: ۴۰ عمل روزانه (مانند نوشیدن، خوردن، خواندن)، ۹ عمل مرتبط با سلامتی (مانند عطسه، تلوتلو خوردن، افتادن) و ۱۱ عمل متقابل (مانند مشت، لگد، در آغوش گرفتن). این اقدامات تحت ۱۷ شرایط صحنه مختلف مربوط به ۱۷ دنباله ویدئو (به عنوان مثال، S001-S017) انجام می شود. اقدامات با استفاده از سه دوربین با نماهای تصویربرداری افقی مختلف گرفته شد. اطلاعات چند وجهی

مختلفی از قبیل نقشه های عمق، موقعیت مفصل اسکلت سه بعدی، فریم های RGB و دنباله های مادون قرمز برای توصیف عملکرد ارائه شده است.

- **KTH (KTH Action dataset)**

تلاش ها برای ایجاد یک مجموعه داده غیر پیش پا افتاده و در دسترس عموم برای شناسایی کنش ها در مؤسسه فناوری KTH در سال ۲۰۰۴ آغاز شد. مجموعه داده KTH یکی از استانداردترین مجموعه داده ها است که شامل شش عمل است: راه رفتن، دویدن، راه رفتن، جعبه، موج دست، و کف زدن دست. برای در نظر گرفتن تفاوت های ظریف عملکرد، هر عمل توسط ۲۵ فرد مختلف انجام شده و در نهایت تنظیمات به طور سیستماتیک برای هر عمل به ازای هر بازیگر تغییر می کند. تغییرات تنظیمات عبارتند از: فضای باز (s1)، فضای باز با تغییر مقیاس (s2)، فضای باز با لباس های مختلف (s3)، و فضای داخلی (s4). این تغییرات توانایی هر الگوریتم را برای شناسایی اقدامات مستقل از پس زمینه، ظاهر بازیگران و مقیاس بازیگران آزمایش می کند.

- **Sports-1M**

مجموعه داده Sports-1M شامل بیش از یک میلیون ویدیو از YouTube است. ویدیوهای موجود در مجموعه داده را می توان از طریق URL YouTube مشخص شده توسط نویسندگان به دست آورد. تقریباً ۷٪ (از سال ۲۰۱۶) از ویدیوها توسط آپلودکنندگان YouTube از زمان گردآوری مجموعه داده حذف شده است. با این حال، هنوز بیش از یک میلیون ویدیو در مجموعه داده با ۴۸۷ دسته مرتبط با ورزش با ۱۰۰۰ تا ۳۰۰۰ ویدیو در هر دسته وجود دارد. ویدیوها به طور خودکار با ۴۸۷

کلاس ورزشی با استفاده از YouTube Topics API با تجزیه و تحلیل فراداده متنی مرتبط با ویدیوها (به عنوان مثال برچسب ها، توضیحات) برچسب گذاری می شوند. تقریباً ۵٪ از ویدیوها با بیش از یک کلاس حاشیه نویسی شده اند.

- **تحلیل مجموعه‌های داده**

مجموعه داده	تعداد کل فایل‌ها	تعداد کل کلاس ها	حوزه	ساختار دقیق برچسب ها	نمونه	امکان دانلود	میزان وفق پذیری با پروژه
UCF101	۱۳۳۲۰ کلیپ ویدیویی	۱۰۱	۱. حرکت بدن ۲. تعامل انسان و انسان ۳. تعامل انسان و شی ۴. نواختن آلات موسیقی ۵. ورزش	لینک	لینک	**لینک**	**متوسط**
Kinetics-700	۶۵۰۰۰۰ کلیپ ویدئویی	۷۰۰	تعامل انسان	بغل کردن، دست دادن و ..	لینک	**لینک**	**متوسط**
Kinetics-600		۶۰۰				**لینک**	
Kinetics-400		۴۰۰				**لینک**	
HMBD51	۷۰۰۰ کلیپ ویدئو	۵۱	۱. رفتار های چهره	لینک ۱ و ۲	لینک ۱ و ۲	**لینک**	**متوسط – بخشی**

				۲. رفتار های چهره با شی ۳. رفتار های اندام ۴. تعامل انسان با انسان			از نمونه‌ها لازم نیست
ActivityNet	۲۰۰۰۰ ویدئو کلیپ	-	لینک	لینک	لینک	لینک ۱ و ۲ – هر چند دانلود این مجموعه بسیار مشکل می‌باشد.	**مناسب**
Charades	۹۸۴۸ ویدئو کلیپ	-	تعامل انسان	۱۵۷ برچسب	لینک		**عدم امکان دانلود**
MPII	۲۵۰۰۰ کلیپ ویدئویی	۴۱۰	لینک ۱ و ۲ و ۳	لینک	لینک	لینک ۱ و .. حدود ۲۵ فایل	**مناسب**
NTU RGB+D	۵۷۰۰۰ کلیپ ویدئو	۶۰	**به دلیل عدم کار وبسایت مربوطه، اطلاعاتی در این خصوص وجود ندارد**				

KTH	2391 sequences	۶	۱. راه رفتن ۲. دویدن ۳. راه رفتن ۴. جعبه ۵. موج دست ۶. کف زدن دست	لینک	لینک	لینک ۱ و ... حدود ۱۲	**نامناسب**
Sports-1M	۱,۱۳۳,۱۵۸	۴۸۷	ورزش	لینک	لینک لینک	**Git clone**	**نامناسب**