

دانشگاه صنعتي امیرکبیر  
(پلی تکنیک تهران)

**دانشکده‌ی مهندسی کامپیوتر**

**گزارش کار کارآموزی**

**Video Swin Transformer**

**نگارش**

**محمدرضا قادری**

**استاد**

**دکتر صاحب الزمانی**

**تیر** 1400

**استاندارد مستندسازی راه‏اندازی کد**

فهرست مطالب

[فصل 1: مجموعه دادگان 1](#_Toc69291746)

[1-1- مقدمه 2](#_Toc69291747)

[1-2- نحوه دسترسی 2](#_Toc69291748)

[1-3- جزئیات مجموعه دادگان 2](#_Toc69291749)

[1-4- جمع‏بندی 2](#_Toc69291750)

[فصل 2: راه‏اندازی کد به صورت برخط (Online) 3](#_Toc69291751)

[2-1- مقدمه 4](#_Toc69291752)

[2-2- نصب پکیج‏ها و پیش‏نیازها 4](#_Toc69291753)

[2-3- نحوه دسترسی 4](#_Toc69291754)

[2-4- راه‏اندازی سریع 4](#_Toc69291755)

[2-5- بررسی جزئیات 5](#_Toc69291756)

[2-6- جمع‏بندی 6](#_Toc69291757)

[فصل 3: راه‏اندازی کد به صورت برون خط (Offline) 7](#_Toc69291758)

[3-1- مقدمه 8](#_Toc69291759)

[3-2- نحوه دسترسی 8](#_Toc69291760)

[3-3- نصب پیش‏نیازها به صورت Offline 8](#_Toc69291761)

[3-4- راه‏اندازی سریع 8](#_Toc69291762)

[3-5- بررسی جزئیات کد 8](#_Toc69291763)

[3-6- جمع‏بندی 9](#_Toc69291764)

[فصل 4: جمع‏بندی و پیشنهادات 12](#_Toc69291765)

[4-1- مقدمه 13](#_Toc69291766)

[4-2- جمع‏بندی کلی و پیشنهادات 13](#_Toc69291767)

[4-3- کارهای آینده 13](#_Toc69291768)

[4-4- جمع‏بندی 13](#_Toc69291769)

[مراجع 14](#_Toc69291770)

فهرست اشکال

[شکل (1-1) تصویر نمونه 2](#_Toc34822082)

**فهرست جداول**

[جدول (1-1) جدول نمونه 2](#_Toc34821840)

1. مجموعه دادگان
   1. مقدمه

در این پروژه با توجه به این که ما دنبال این هستیم که حرکات انسان رو بشناسیم و یاد بگیریم از 3 نوع دیتاست استفاده می‌کنیم که از نظر نوع داده‌های آموزشی و تستی متفاوت هستند (مقدار و نوع پخش شدن داده‌ها)

اولین دیتاستی که در مورد آن خواهیم گفت Kinetic-400 ، سپس Kinetic-600 و در نهایت  
 Something-Something v2 رو داریم.

به طور کلی Kinetic دارای یک مجموعه از بیش از 650000 ویدیو با کیفیت و اندازه مناسب است که کلاس بندی شده اند با توجه به ورژن دیتاست هست و این ویدئوها شامل فعل و انفعالات بین انسان و شیء مانند نواختن سازها ، و همچنین تعاملات انسان با انسان مانند دست دادن و در آغوش گرفتن است. که هر کدام از کلاس ها حداقل به اندازه تعداد کلاس ها ویدیو دارد، که هر ویدیو فقط یک حرکت و حدودا 10 ثانیه زمان دارد. اکثر این ویدیو ها از یوتیوب آورده شده.

* 1. نحوه دسترسی
  2. جزئیات مجموعه دادگان

[Kinetic-400:](https://arxiv.org/pdf/1705.06950.pdf)  در حال حاضر در این ورژن از دیتاست 306245 ویدیو موجود هست که هر کلاس 400 - 1150 تا ویدیو دارد که این ویدیو ها به سه قسمت تقسیم می‌شوند که یک قسمت برای یادگیری[[1]](#footnote-1) که برای هر کلاس بین 250 - 1000 ویدیو دارد ، 50 تا برای اعتبارسنجی[[2]](#footnote-2) و در نهایت 100 ویدیو برای تست[[3]](#footnote-3) برای هر کلاس موجود است.



تعداد کلیپ ها برای هر کلاس در پارتیشن های یادگیری/اعتبارسنجی/تست

در مقاله ارجاع داده شده درباره این دیتاست درباره موارد مختلفی مانند گیج کننده ترین کلاس ها و سخترین کلاس ها برای تصمیم گیری صحبت شده.(در مورد 400 کلاس که چه موضوعی هست گفته شده)

[Kinetic-600:](https://arxiv.org/pdf/1808.01340v1.pdf) در حال حاضر در این ورژن از دیتاست 495547 ویدیو موجود هست که هر کلاس 600 - 1200 تا ویدیو دارد که این ویدیو ها به سه قسمت تقسیم می‌شوند که یک قسمت برای یادگیری که برای هر کلاس بین 450 - 1000 ویدیو دارد ، 50 تا برای اعتبارسنجی و در نهایت 100 ویدیو برای تست و یک مجموعه تست اضافی حدودا 50تایی برای فعالیت های خاص در این ورژن در نظر میگیریم برای هر کلاس موجود است. نسبت به kinetic-400 از 368 تا از کلاس‌های آن استفاده شده است، برای بقیه 32 کلاس به دلیل اینکه تقاطع زیادی با کلاس های دیگر داشت و شاید برای تشخیص دچار مشکل میشدیم تغییر کرد.



تعداد کلیپ ها برای هر کلاس در پارتیشن های یادگیری/اعتبارسنجی/تست (مقایسه با 400)

در مقاله ارجاع داده شده در بالا درباره این دیتاست درباره موارد مختلفی مانند گیج کننده ترین کلاس ها و سخترین کلاس ها برای تصمیم گیری صحبت شده.(کلاس های اضافه شده نسبت به 400 در انتهای مقاله ارجاع داده شده صحبت شده)

جمع‏بندی

متن.

1. راه‏اندازی کد به صورت برخط (Online)
   1. مقدمه

متن.

* 1. نصب پکیج‏ها و پیش‏نیازها

متن.

* 1. نحوه دسترسی

متن.

* 1. راه‏اندازی سریع

متن.

نمونه کد:

args = [

        '--name', \_\_file\_\_.split('/')[-1].split('.')[0],  # name is filename

        '--print-freq', '1',

        '--data', 'D:/Charades/Charades\_v1\_rgb/Charades\_v1\_rgb',

        '--dataset', 'charadesrgb',

        '--cache-dir','./cache/',

        '--pretrained',

        #'--evaluate',

    ]

نمونه خطا:

|  |  |
| --- | --- |
| خطای 1: خطای عدم وجود کتابخانه | نحوه رفع این خطا |
|  |  |

* 1. بررسی جزئیات

متن.

.

* 1. جمع‏بندی

متن.

1. راه‏اندازی کد به صورت برون خط (Offline)
   1. مقدمه

متن.

* 1. نحوه دسترسی

متن.

* 1. نصب پیش‏نیازها به صورت Offline

متن.

* 1. راه‏اندازی سریع

متن.

* 1. بررسی جزئیات کد

متن.

* 1. جمع‏بندی

متن.

1. جمع‏بندی و پیشنهادات
   1. مقدمه

متن.

* 1. جمع‏بندی کلی و پیشنهادات

متن.

* 1. کارهای آینده

متن.

* 1. جمع‏بندی

متن.

مراجع

[1] Q. Liu, A. H. Sung, and M. Qiao, "Video steganalysis based on the expanded markov and joint distribution on the transform domains detecting msu stegovideo," in *Machine Learning and Applications, 2008. ICMLA'08. Seventh International Conference on*, 2008: IEEE, pp. 671-674.

1. Training [↑](#footnote-ref-1)
2. Validations [↑](#footnote-ref-2)
3. Test [↑](#footnote-ref-3)