



به نام خدا

نام دانشکده: دانشکده برق و کامپیوتر

ترم: 4021

نام طراح: علی بزرگ‌زاد

تاریخ تحویل: 19 دی 1402

نام درس: یادگیری عمیق

تمرین عملی: مبحث حوزه شبکه‌های بازگشتی

استاد: دکتر سمانه حسینی

طبقه‌بندی ویدیو به کمک شبکه‌های عصبی بازگشتی

معرفی تمرین

در این تمرین سعی بر آن است تا بتوان به کمک شبکه‌های عصبی بازگشتی تعدادی ویدیو را طبقه‌بندی کرد. همانطور که می‌دانیم ویدیو از تعدادی تصویر یا Frame پشت سرهم تشکیل شده است، پس برای بخش تصویر به نظر می‌توان به ساختار CNN فکر کرد و برای بخش توالی و پشت سرهم بودن آن می‌توان به ساختار شبکه‌های بازگشتی فکر نمود. در هر حال مهم‌ترین مورد برای درک یک ویدیو ساختار و توالی آن است که باید سعی شود به بهترین نحو پیاده‌سازی گردد.

در طول سال‌های اخیر روش‌های متفاوتی برای طبقه‌بندی ویدیو پیشنهاد شده است که در بخش بعد به معرفی این روش‌ها خواهیم پرداخت. روش‌ها تنوع زیادی دارند ولی به هر حال پیاده‌سازی روش‌های که توالی را به نحو بهتری در نظر می‌گیرند، نتیجه بهتری خواهند داشت. شما دوستان می‌بایست از بین روش‌های موجود یکی از روش‌هایی که شامل شبکه‌های بازگشتی می‌باشد را انتخاب و پیاده‌سازی کنید (البته با توجه به محدودیت‌های مطرح شده در نکات پایینی تمرین).

Dataset

در این تمرین شما باید با Dataset معروفی به نام UCF50 کار کنید که شامل 50 کلاس از ویدیوهایی می‌باشند که از طریق سایت Youtube جمع‌آوری شده‌اند. جهت اطلاعات بیشتر می‌توانید به [سایت اصلی UCF](#) مراجعه نمایید. جهت راحتی شما دوستان عزیز این Dataset بر روی IUTBox دانشکده آپلود شده است ولی به هر حال می‌توانید از سه لینک زیر این Dataset را دانلود نمایید.

- [لینک IUTBox](#)
- [لینک اصلی](#)
- [لینک فرعی](#)

پیش پردازش

در این بخش یکسری اعمال را بر روی Dataset تعیین می‌کنیم تا همگی شما دوستان در شرایط برابری کد خود را اجرا نمایید.

- تعیین Random Seed

```
• mySeed = 42
• np.random.seed(mySeed)
• random.seed(mySeed)
• tf.random.set_seed(mySeed)
• torch.manual_seed(mySeed)
```

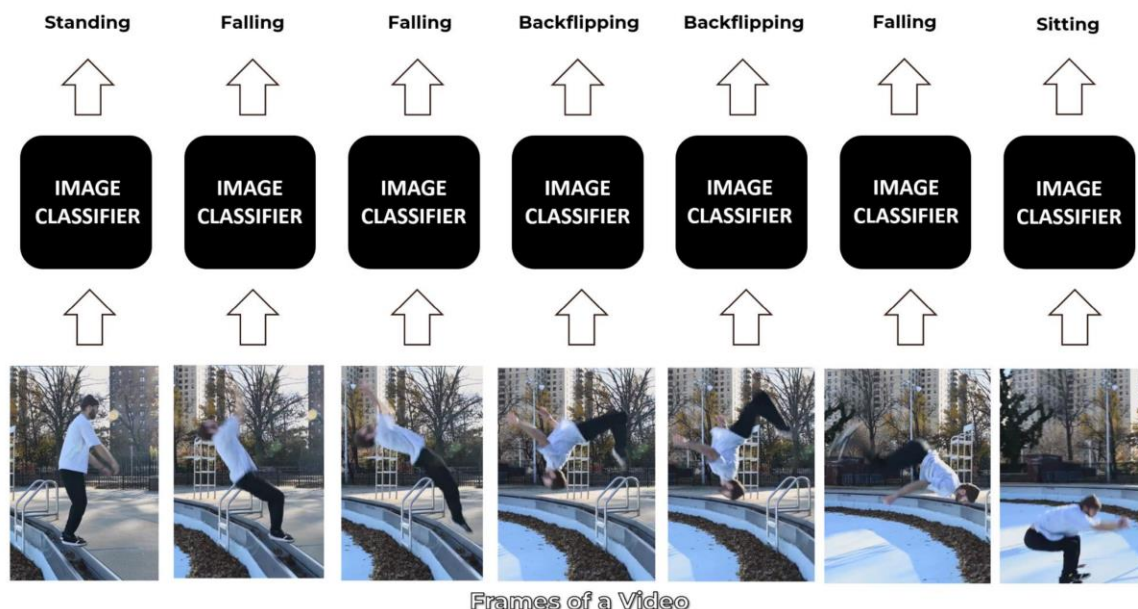
- فقط از 20 Frame ابتدایی هر ویدیو استفاده نمایید (در صورتی که مایل به استفاده از Frame‌های بیشتری می‌باشید، کد با 20 Frame بیشتر را در فایل مجزا اجرا کرده و بر روی سامانه آپلود کنید به عبارت بهتر کد با 20 Frame را حتما باید داشته باشید).

- جهت جلوگیری از پردازش سنگین و ارائه خروجی، نیاز به اجرای کد روی تمامی 50 کلاس نیست و شما باید 10 کلاس به 10 کلاس کد خود را اجرا کرده و نتیجه را به ازای هر 10 کلاس ارائه دهید، در حقیقت کلاس 0 تا 9 یک بار اجرای کد، کلاس 10 تا 19 یک بار اجرای کد الی کلاس 40 تا 49 یک بار باید کد اجرا شده و خروجی هر کدام به صورت مجزا ارائه گردد (به 5 مرتبه اجرا نیاز می‌باشد).

انواع روش‌های طبقه‌بندی تصاویر

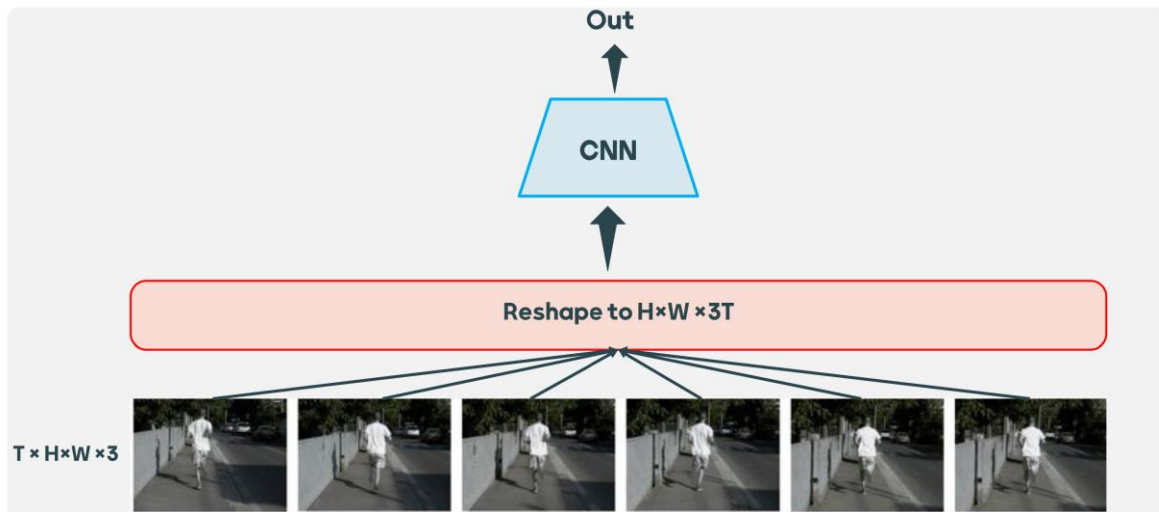
- Single Frame Classification

همانطور که در شکل زیر مشاهده می‌کنید، ابتدا باید تمامی Frame‌ها بررسی شوند و در انتها کلاسی با بیشترین تکرار تشخیص را انتخاب نماییم.



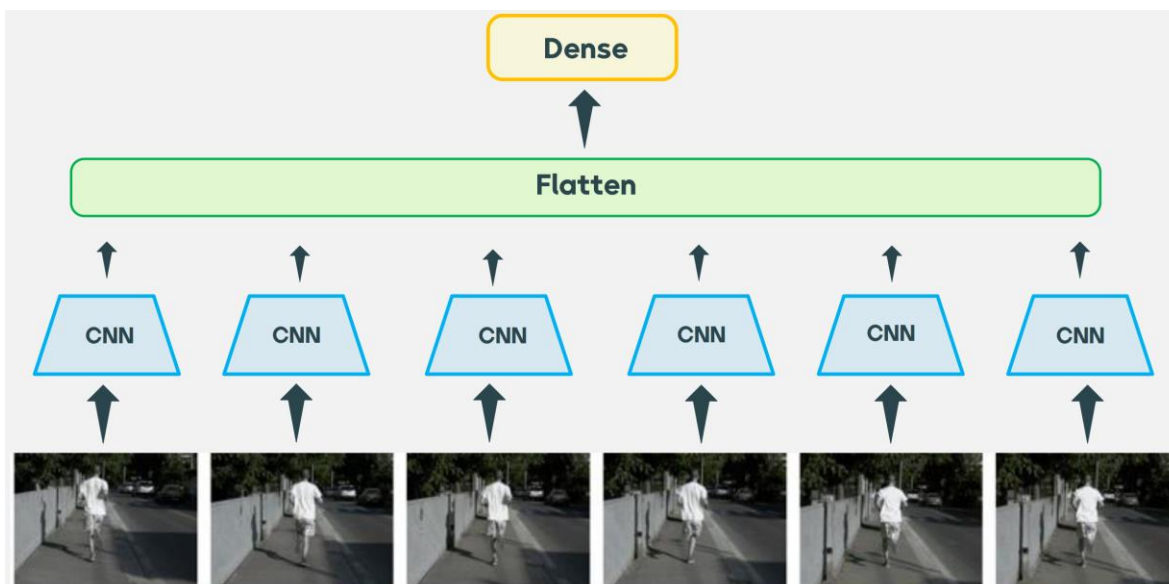
• Early Fusion

بعد زمان با بعد رنگ ترکیب شده و سپس از CNN عبور داده می شود.



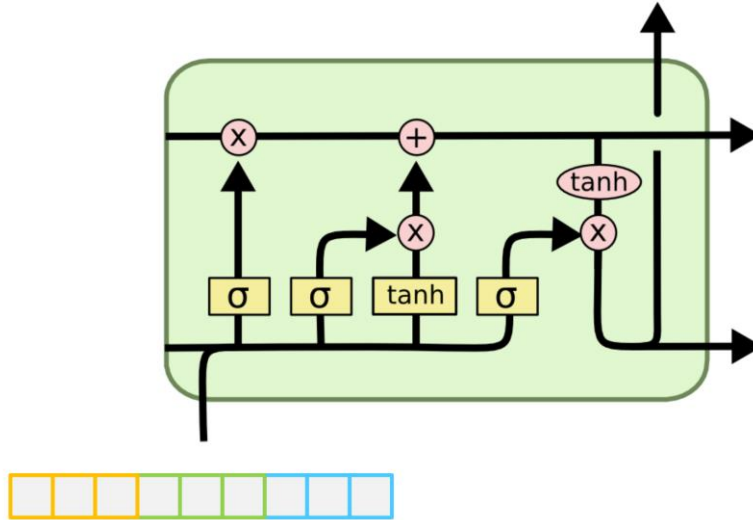
• Late Fusion (Fully Connected Layer)

ابتدا تمام تصویر به صورت مجزا از CNN عبور داده شده و در نهایت همگی تجمیع شده و به صورت یک مجموعه واحد به شبکه عصبی داده خواهند شد.



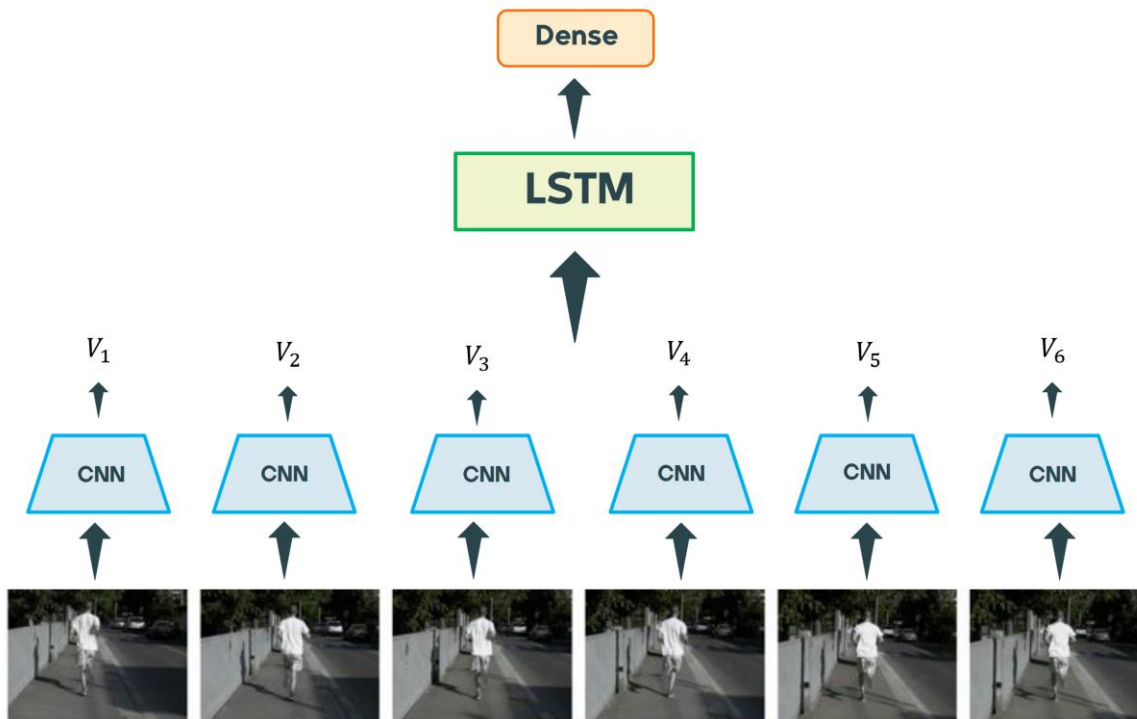
• Image Flatten with LSTM

می‌دانیم که LSTM‌ها به صورت پیش‌فرض فقط با ساختار برداری کار می‌کنند، پس در نتیجه Frame‌ها را همگی Flat کرده و هر Frame را به یک LSTM می‌دهیم. دقیقاً مشابه اینکه هر کلمه را به صورت مجزا Encode می‌کردیم و هر کدام را به عنوان ورودی یک LSTM به شبکه می‌دادیم.



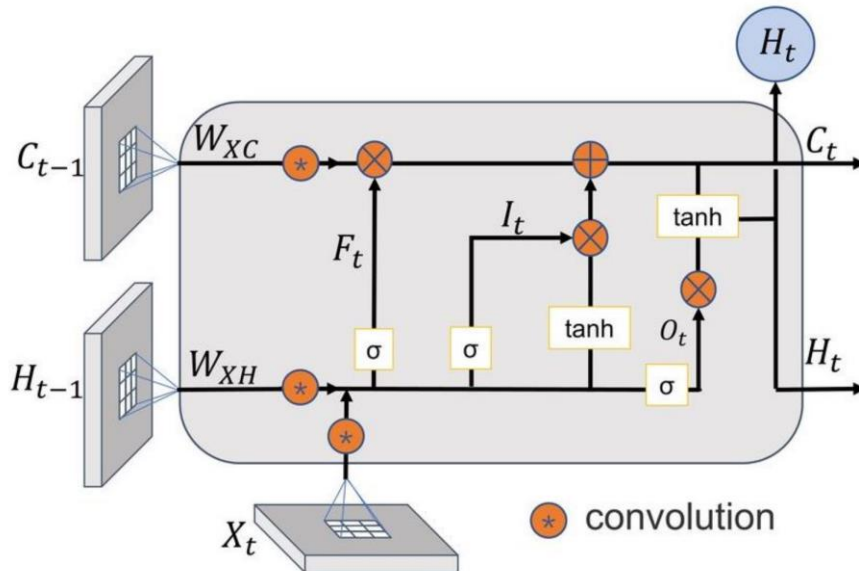
• CNN with LSTM

ابتدا هر تصویر را به صورت مجزا از یک شبکه CNN عبور داده، ویژگی‌های مطلوب آن را استخراج کرده سپس مانند روش قبل، عمل می‌کنیم (به جای Flat کردن عکس اصلی، ویژگی‌های خروجی CNN را Flat می‌کنیم و آن را به LSTM‌های مربوطه می‌دهیم).



Convolutional LSTM (Built-in Conv in LSTM)

اما روش دیگر به این صورت که چرا اصلا ساختار اصلی LSTM را تغییر ندهیم و بجای اعمال تمام عملیات به صورت عادی در داخل LSTM این بار در ساختار LSTM تغییر ایجاد کرده و تمام عملیات‌های داخل LSTM را به صورت Convolutional در نظر بگیریم.



۹- نکات بسیار کلیدی

- شما مجاز به پیاده‌سازی هر نوع ساختاری هستید و در کل می‌بایست حتماً از شبکه‌های بازگشتی استفاده کرده و محدودیت‌هایی مانند Accuracy و ... را که در طول تمرین ذکر شده است، رعایت نمایید.
- یک فیلم کوتاه تهیه نمایید و کدهای خود را به صورت خلاصه بررسی کرده و در مورد آن توضیح دهید.
- اگر پیش‌پردازش‌های گفته شده و محدودیت‌های اعمال شده را لحاظ نمایید به راحتی کد شما بر روی GoogleColab اجرا شده و دچار هیچ مشکلی نخواهید شد ولی در صورتی که بخواهید فریم‌ها را بیشتر کنید باید از نظر میزان RAM مدیریت مطلوبی داشته باشید (این مورد حتی برای داده‌های بیشتر از این Dataset نیز امکان‌پذیر است ولی در این تمرین الزامی به مدیریت مطلوب حافظه نیست).
- بهترین کد پیاده‌سازی شده شما در حالت معمول برای 5 دسته کلاس‌بندی انجام شده، می‌بایست حدوداً 75٪ Accuracy را داشته باشد.
- در صورتی که تمامی روش‌ها پیاده‌سازی شوند، 15٪ نمره اضافه برای شما لحاظ خواهد شد.
- کد شما به ازای تمامی 5 دسته کلاس باید یکسان باشد به عبارت بهتر شما فقط حق نوشتن یک کد را دارید و برای تمامی 5 مرتبه اجرا، کد شما باید ثابت باشد.

نکات تحویل تمرین:

- همانطور که قبلا هم اطلاع داده شد، شما مجاز هستید در طول ترم تا 8 روز تاخیر در تحویل کل تکالیف داشته باشید.
- دانشجویان می توانند در حل تکالیف با دوستان خود مشورت نمایند اما در نهایت هرکس موظف است تکلیف را به صورت فردی، انجام و تحویل دهد. **لذا، در صورت مشاهده تکالیف کپی بین دانشجویان، نمره تمامی افراد شرکت کننده در آن صفر خواهد بود.**
- توضیحات شما در فرآیند تصحیح از اهمیت ویژه ای برخوردار است. لطفا تمامی نکات و فرض هایی که برای پیاده سازی ها و محاسبات خود در نظر می گیرید را در گزارش ذکر کنید.
- در صورت داشتن هرگونه سوال می توانید از طریق ایمیل یا اکانت تلگرام زیر با دستیار آموزشی مربوطه در ارتباط باشید.

Email: a.bozorgzad@ec.iut.ac.ir / Telegram: [@Ali_Ai_Dev](https://t.me/Ali_Ai_Dev)

پیروز باشید