

پروژه پایانی درس بینایی کامپیوتر
اعضای گروه:
ملیکا محمدی فخار - ۹۹۵۲۲۰۸۶
ستاره باباجانی - ۹۹۵۲۱۱۰۹

۱. Dataset

برای دیتاست، از یکی از دیتاست‌های پیشنهادی در داکيومنت پروژه (CASIA-FASD) استفاده کردیم. از آنجا که حجم داده زیاد بود و آپلود هرباره آن میسر نبود، آن را در Google Drive ذخیره کردیم. داده‌ی آموزشی شامل ۱۲ ویدیو از ۲۰ شخص متفاوت می‌باشد. از این ۱۲ ویدیو ۴ تای آن‌ها ویدیوهای زنده و ۸ تای آن‌ها غیرزنده می‌باشند. از آنجا که موفق به پیدا کردن داده‌ی برچسب‌خورده نشدیم، تکه کدی نوشتیم تا به ویدیوهای زنده برچسب ۱ و به ویدیوهای غیرزنده برچسب ۰ اختصاص دهد. سپس از هریک از ویدیوها، با نرخ ۱ فریم در ۳ ثانیه، فریم استخراج کردیم و بر اساس برچسب ویدیو، به هر فریم نیز برچسب مناسب اختصاص دادیم.

۲. Preprocessing

ابتدا هر تصویر (فریم) را ریسایز کردیم تا ابعاد همه تصاویر برابر شوند. سپس هیستوگرام تصویر را متعادل کردیم تا وضوح تصویر با توجه به شرایط نورپردازی آن افزایش یابد. پس از آن با استفاده از یک کرنل 5×5 تابع gaussian blur را بر روی تصویر اعمال کردیم تا نویزهای احتمالی کاهش یابد. از آنجا که حجم داده کافی نبود، از augmentation استفاده کردیم، که شامل flip، چرخش ۹۰ درجه‌ای، چرخش ۱۸۰ درجه‌ای، چرخش ۲۷۰ درجه‌ای می‌باشد.

۳. Models

• Feature Model

در کلاس این مدل، یک تابع برای استخراج ویژگی تعریف شده است که ویژگی‌های بافتی (texture)، فرکانسی (frequency)، ماتریس هم‌روی سطح خاکستری (glcm) از تصویر ورودی استخراج می‌شوند. همچنین از SVC به عنوان Classifier میان دو کلاس استفاده کردیم که بتواند داده‌ها را به دو دسته جدا کند به طوری که فاصله بین نزدیک‌ترین نقاط هر دسته به هاپرپلین بیشترین مقدار ممکن باشد.

• Deep Learning Model

در این بخش، از مدل پیش‌آمورخته‌ی vgg۱۶ استفاده کردیم که وزن‌های اولیه آن برابر وزن‌های شبکه imagenet می‌باشد. در ادامه یک Global Pooling، Dense Layer، Dropout اضافه کردیم، در انتها از یک Dense Layer با تابع فعال‌سازی sigmoid استفاده کردیم تا خروجی نهایی در قالب یک احتمال به دست آید. برای بهینه‌سازی از بهینه‌ساز adam و تابع ضرر binary crossentropy استفاده کردیم.

۴. Training

برای آموزش، به Feature Model ویژگی‌های استخراج شده از هر فریم که پیش‌تر توضیح داده شد به عنوان ورودی داده شدند.

به مدل Deep Learning نیز، تصاویر پیش‌پردازش شده در ۱۰ ایپوک ورودی داده شد که دقت نهایی به حدود ۹۲ درصد رسید.

۵. Evaluation on CASIA Test Data

داده‌های تست دیتاست Casia-fasd را لود کرده و عملکرد دو مدل را بر روی آن‌ها ارزیابی کردیم. دقت مدل Feature حدود ۷۴ درصد و دقت مدل Deep Learning حدود ۹۰ درصد می‌باشد.

۶. Load Our Test Data

برای تولید داده‌های آزمایشی، ۱۰ ویدیو ضبط کردیم، که شامل ۵ ویدیوی زنده و ۵ ویدیوی غیرزنده است. سپس مراحل برچسب‌زنی، پیش‌پردازش تصاویر، اعمال augmentation مشابه داده‌های آموزشی بر روی داده‌های آزمایشی نیز اعمال شد.

۷. Evaluation

- ارزیابی با تصویر کامل به عنوان ورودی
با ارزیابی عملکرد دو مدل با داده‌های برش نخورده توانستیم به دقت‌های زیر دست یابیم:
مدل Feature: حدود ۶۰ درصد، دقت مدل Deep Learning: حدود ۷۶ درصد
همانطور که مشاهده می‌شود، دقت مدل‌ها بر روی داده‌های تولید شده توسط ما نسبت به داده‌های آزمایشی CASIA-FASD حدود ۱۴ درصد کاهش یافته که علت احتمالی آن، دامنه متفاوت داده‌های تولید شده توسط ما به دلیل جنسیت، حجاب اسلامی و ... می‌باشد.

- ارزیابی با تصویر برش خورده به عنوان ورودی
ابتدا از یک face detector پیش‌آمخته به نام haarcascade استفاده کردیم تا چهره شخص را در هر تصویر پیدا کرده و برش بزنند. در این بخش تصاویر برش خورده به عنوان ورودی به دو مدل داده شدند.

- ارزیابی با ویژگی‌های فرکانسی استخراج شده از تصاویر

ابتدا ویژگی‌های فرکانسی را با تابعی مشابه تابعی که در **feature model** داشتیم با گرفتن اطلاعات از تبدیل فوریه تصاویر، استخراج کردیم. سپس برای یکسان کردن ابعاد ویژگی‌های فرکانسی با ابعاد مورد انتظار مدل‌ها از **zero padding** استفاده کردیم.

۸. Conclusion

با استفاده از هر یک از دو مدل، و به ازای سه نوع ورودی داده شده برای هر فریم مقدار احتمال زنده بودن محاسبه شد. سپس برای تخمین احتمال زنده بودن ویدیو، میان احتمالات فریم‌های آن میانگین گرفتیم. نتایج ذخیره نتایج در فایل‌های **CSV** بدین صورت است:

processed_deep_predictions

Video	Deep Full Image	Deep Cropped Face	Deep Frequency
live_1.mp4	0.9799948334693910	0.9441207051277160	0.9187396754633720
live_2.mp4	0.8753222107887270	0.7493106842041020	0.8643915889892890
dead_1.mp4	0.00013242998702160900	0.14551694508936300	0.8359577580365890
dead_2.mp4	7.81866304350842E-05	0.07853370399143390	0.43035118277013400
dead_3.mp4	0.5361511513590810	0.3352015785872940	0.9010896658819750
live_3.mp4	0.7229750990867620	0.8875286102294920	0.924127442665905
live_4.mp4	0.6209064520895480	0.9277768492698670	0.9472239955365010
dead_4.mp4	0.8677674293518070	0.8594944834709170	0.9937826224116220
dead_5.mp4	0.48459483236074400	0.31250537633895900	0.9629794505183410
live_5.mp4	0.17415261939168000	0.5329391628503800	0.9882280473385680

processed_feature_predictions

Video	Full Image	Cropped Face	Frequency
live_1.mp4	0.891909517619627	0.855676411097156	0.9187396754633720
live_2.mp4	0.8075353051269200	0.6168533930858000	0.8643915889892890
dead_1.mp4	0.7660604037118580	0.64953621328303	0.8359577580365890
dead_2.mp4	0.34226059954826000	0.5388996554905340	0.43035118277013400
dead_3.mp4	0.8376150502710510	0.8910385600968510	0.9010896658819750
live_3.mp4	0.912153309444166	0.9250107710267230	0.924127442665905
live_4.mp4	0.9369809855373990	0.918511529732257	0.9472239955365010
dead_4.mp4	0.9923565361457520	0.993065274433196	0.9937826224116220
dead_5.mp4	0.9517159758430060	0.9541535971256940	0.9629794505183410
live_5.mp4	0.9864534950183480	0.9575375237724790	0.9882280473385680