

به نام خدا

گزارش پروژه پایان ترم
پردازش تصویر

استاد: آقای دکتر کیانی

نویسنده: علی ناظری
شماره دانشجویی: ۹۸۳۶۱۳۰۵۶

تاریخ: ۱۴۰۲/۳/۲۹

مقدمه

در پروژه میان ترم با استفاده از استخراج ویژگی و مدل یادگیری ماشین SVM یک مدل برای تشخیص لبخند بدست آمد. در این قسمت پروژه باید با استفاده از یک مدل یادگیری عمیق تشخیص لبخند ساخته شود و سپس با استفاده از آن و یک دوربین لبخند افراد تشخیص داده شود. در اینجا مدل یادگیری عمیق هم ویژگی های مناسب خود را پیدا می کند و هم مدلی طبق آنها برای تشخیص لبخند بدست می آورد.

یادگیری عمیق

در ابتدای کار نیاز است تا تمامی عکس های مورد استفاده برای مدل تماما به یک اندازه باشند. همچنین تمام چهره های موجود در تصاویر در فاز قبلی پروژه با استفاده از یک مدل تشخیص چهره و بعضی از آنها به صورت دستی جدا شدند و ۴۰۰۰ تصویر از صورت افراد موجود می باشد در فاز قبلی برای بدست آوردن بعضی ویژگی ها سایز تصاویر تغییر داده شد سایز نهایی آنها به صورت میانگین سایز تمامی تصاویر در نظر گرفته شد که حدود ۱۰۰ در ۱۰۰ پیکسل بود و به همین عدد رند شد. در قسمت خواندن تصاویر تمامی تصاویر یک به یک خوانده شدند سایز آن ها به ۱۰۰، ۱۰۰ تغییر داده شد و همچنین چون تمامی مقادیر ورودی برای یک مدل یادگیری عمیق باید نرمال سازی شده باشد پس مقادیر موجود در تصاویر به ۲۵۵ تقسیم شدند تا بین ۰ و ۱ آورده شوند. سپس به یک آرایه به نام imgs اضافه شدند. پس از اضافه کردن تمامی تصاویر به فرمت یک آرایه np.array تبدیل شد.

در بلاک کد بعدی label های تصاویر خوانده شدند.

سپس ساختار shape برای ورودی بررسی شد که باید به این صورت باشد که عدد اول تعداد عکس ها و دو عدد بعدی سایز و عدد آخر تعداد کانال ها باید باشد که نیاز به تغییر ندارد و مناسب می باشد.

حال که موارد ورودی آماده هستند باید آن ها به سه دسته train, test, validation تقسیم شوند. Validation برای اینکه از overfit جلوگیری شود استفاده می شود. Validation بخشی از داده های train می باشد. با کمک گرفتن از train_test_split کتابخانه sklearn، اول داده ها به یک قسمت یادگیری و تست تقسیم می شود و این نسبت ۰٫۲ است سپس داده های یادگیری دوباره با همین تابع به دو قسمت با همین نسبت برای اعتبار سنجی تقسیم می شوند.

برای جلوگیری از overfit شدن و یادگیری با چالش و پر قدرت تر با استفاده از ImageDataGenerator که در keras موجود است تغییراتی در عکس بوجود می آوریم. این تغییرات به صورت تصادفی اعمال می شوند. تغییرات که روی تصاویر اعمال می شوند می تواند چرخش عکس باشد به اندازه ۵ درجه و جابجایی تصویر به بالا پایین و چپ راست با اندازه ۱۰ درصد و زوم داخل و خارج و چرخش افقی عکس ها می باشد. همچنین این تابع به صورت خودکار مقادیری از عکس را که با این تغییرات وجود ندارد با مقادیر همسایه های آن جاگذاری می کند.

سپس حالا باید لایه های مدل CNN تایین شود. از سه لایه Conv2D همراه با MaxPool2D ساخته شده است که در Conv2D لایه ها همه از تابع فعال سازی relu استفاده می کنند و لایه اول با ۳۲ فیلتر و در بقیه ۶۴ فیلتر استفاده شده است همچنین اندازه kernel آن ها ۳ در ۳ می باشد. برای MaxPool2D اندازه ۲ در ۲ در نظر گرفته شده است همراه با stride برابر ۲ که به این معنی است که هر بار که عملیات maxpooling انجام می شود پنجره بعدی برای انجام این عملیات چقدر جابجا شود.

Stride کمک می کند که سرعت یادگیری بالاتر برود و از overfit جلوگیری شود. چهار لایه آخر به ترتیب flatten سپس dropout بعد یک لایه dense و در آخر یک لایه sigmoid استفاده شده است. Dropout بعضی از نورون ها را در نظر نمی گیرد و موجب جلوگیری از overfit و بهبود عملکرد و جلوگیری از نویز می شود. Sigmoid برای تشخیص خروجی که شامل دو کلاس می باشد استفاده می شود.

در آخر با ورودی های تغییر یافته و داده های اعتبار سنجی مدل fit می شود. در پایان این مدل ها با کتابخانه pickle ذخیره شدند.

ابتدا با ۲۰ و ۳۰ epoch یادگیری انجام شد ولی همچنان روند تغییر و بهبود در دقت وجود داشت پس این مقدار زیاد تر و به ۴۰ و ۵۰ تنظیم شد که ۴۰ بهترین تعداد دوره می باشد در ۵۰ دوره مقدار حساسیت بالا رفت و دقت کاهش پیدا کرد و به نظر می رسید که مدل overfit شده است. عملکرد هر کدام از مدل های بدست آمده از هر دوره با استفاده از وبکم بررسی شد.

تشخیص لبخند با وبکم

در نهایت در کد موجود در فایل camera.py ورودی وبکم خوانده می شود و هر فریم را به تابعی می دهد و این تابع با استفاده از مدل آماده haarcascade_frontalface_default.xml چهره تشخیص داده می شود و سپس با استفاده از مدل predict می شود و سپس طبق خروجی آن یک حد در نظر گرفته شده است که اگر مقدار از ۰,۱ بیشتر باشد لبخند در نظر بگیرد اگر کمتر باشد لبخند در نظر نمی گیرد. نتیجه تشخیص در بالای چهره افراد که با یک مستطیل مشخص شده است نمایش داده می شود. پنجره با کلید ESC بسته می شود.