



به نام خدا

مبانی بینایی کامپیوتر (دکتر سیفی پور)

دستیار ارشد: پارسا دربان

طراح: امیرحسام جعفری راد

تمرین سوم

نیمسال اول ۱۴۰۵-۱۴۰۴

مقدمه

در دهه اخیر، شبکه‌های کانولوشنی عمیق، بهویژه مدل‌های از پیشآموزش‌دیده (Pre-trained Models) بر روی مجموعه‌داده‌های عظیم به یکی از مهم‌ترین ابزارهای عملی در حوزه بینایی ماشین تبدیل شده‌اند. این مدل‌ها با داشتن صدها میلیون تا چند میلیارد پارامتر، قابلیت استخراج ویژگی‌های بسیار قدرتمند و عمومی از تصاویر را دارند، به‌طوری‌که حتی بدون هیچ‌گونه آموزش مجدد می‌توانند روی وظایف متعدد و تصاویر جدید عملکرد بسیار خوبی داشته باشند.

در این پژوهه، از همین قابلیت استفاده خواهد شد: با لود کردن وزن‌های از پیشآموزش‌دیده و تنها با اعمال پیش‌پردازش صحیح، سه وظیفه کلیدی بینایی ماشین را روی تصاویر واقعی انجام می‌دهید:

- طبقه‌بندی تصویر (Image Classification)
- تشخیص و محل‌سازی اشیاء (Object Detection)
- تقسیم‌بندی معنایی در سطح پیکسل (Semantic Segmentation)

هدف اصلی پژوهه، علاوه بر پیاده‌سازی فنی صحیح، درک عمیق عملکرد این مدل‌ها در دنیای واقعی است: کدام مدل در چه شرایطی موفق است؟ کجا شکست می‌خورد؟ چرا برخی اشیاء به درستی تشخیص داده نمی‌شوند؟ مرزهای تقسیم‌بندی تا چه حد دقیق هستند؟ و در نهایت، چگونه می‌توان از این مدل‌ها در کاربردهای عملی استفاده کرد؟

تصاویر ورودی

در این پژوهه، ۶ تصویر شامل صحنه‌های شهری، طبیعت و ... به شما داده شده، تا با توجه به خواسته‌های هر بخش از آن‌ها به شیوه مناسب استفاده نمایید.

بخش اول: طبقه‌بندی تصاویر (Image Classification)

در این بخش ابتدا با استفاده از قطعه کدی که در اختیارتان قرار داده شده مدل ذکر شده در نوت‌بوک را در حالت استنتاج استفاده کرده و با عملیات‌های خواسته شده به عنوان پیش‌پردازش، خروجی را به ازای تصاویر ذکر شده نمایش داده و درصد اطمینان را برای ۵ کلاس برتر را نیز ذکر کنید. همچنین پس از پاسخ به سوالات پرسیده شده در نوت‌بوک، در انتهای نتایج بدست آمده را تحلیل کنید.

بخش دوم: تشخیص اشیا (Object Detection)

در این بخش، قطعه کد داده شده را به صورتی تکمیل کنید که کادرهای تشخیص، همراه با نام کلاس و امتیاز آن روی تصویر اصلی رسم شود. مدل مربوطه را بر روی عکس‌های ذکر شده اجرا کرده و خروجی را تحلیل نمایید.

بخش سوم: بخش بندی (Segmentation)

در این قسمت، قطعه کد داده شده را طوری تکمیل نمایید که برای هر تصویر ورودی، ماسک رنگی با پالت VOC و تصویر Overlay (۶۰ درصد تصویر اصلی و ۴۰ درصد ماسک) وجود داشته باشد. مدل را بر روی تصاویر ذکر شده اجرا کرده و خروجی را تحلیل کنید.

نکات نهایی

۱. ددلاین تمرین تاریخ ۲۱ آذر ۱۴۰۴ است.
۲. تمرین دارای ارائه ۱۰ دقیقه‌ای می‌باشد و باید آمادگی ارائه آن را داشته باشید.
۳. تمرین به صورت گروههای دو نفره که در تمرین اول مشخص می‌شود، قابل انجام است. حتماً در گزارش کار نام اعضا و شماره دانشجویی نوشته شود؛ در غیر این صورت نمره برای اسامی نوشته‌نشده تعلق نمی‌گیرد.
۴. فایل کد و گزارش کار را در پوشه‌ای با نام زیر در سامانه آپلود کنید.
CV-HW3-std#1-std#2
۵. نوشتن گزارش کار الزامی است و می‌تواند در Jupyter Notebook نیز نوشته شود.
۶. در صورت پیدا شدن موارد تقلب، نمره تمرین صفر می‌شود.
۷. شما مجاز به استفاده از کتابخانه‌های رایج پایتون مانند OpenCV هستید.
۸. دقت کنید که تمامی نتایج موارد خواسته شده باید در گزارش کار یا فایل HTML آورده شده باشد.