

گزارش کد

این پروژه از ترکیب متن و تصویر برای درک دقیق تر احساسات (به ویژه در موارد پیچیده مانند کنایه) استفاده می کند.

۱. ساختار کلی پروژه

کد ارائه شده شامل سه بخش اصلی (ماژول) است که هر کدام وظیفه مشخصی در چرخه پردازش داده دارند:

تولیدکننده (Generator): ساخت تصاویر مرتبط با متن .

اصلاح کننده (Refiner): متمرکز کردن تصویر بر روی جنبه (Aspect) مورد نظر .

طبقه بندی کننده (Classifier): ترکیب ویژگی های متنی و بصری برای تشخیص نهایی احساس .

۲. بررسی ماژول های فنی

الف) تولید تصویر مصنوعی (generator.py)

این بخش از مدل **Stable Diffusion** استفاده می کند تا بر اساس متن ورودی، جنبه مورد نظر و نوع احساس، یک تصویر متناظر تولید کند .

هدف: شبیه سازی داده های چندرسانه ای در زمانی که تصویر واقعی در دسترس نیست .

ب) پالایش بصری (refiner.py)

در این مرحله از مدل **SAM (Segment Anything Model)** و **CLIP** استفاده شده است .

عملکرد: مدل SAM با ایجاد یک ماسک روی تصویر، بخش های اضافی را حذف کرده و تمرکز مدل را بر روی سوژه اصلی (Aspect) قرار می دهد .

ج) مدل نهایی (classifier.py)

این هسته اصلی سیستم است که از معماری عصبی چندرسانه ای بهره می برد:

- **بخش متنی:** از مدل **BERT** برای استخراج ویژگی های زبان طبیعی استفاده می کند .

- بخش بصری: از مدل (CLIP) محصول OpenAI برای تبدیل تصویر به بردارهای معنایی استفاده می‌کند.

- ادغام: (Fusion) بردار متن و تصویر با هم ترکیب (torch.cat) شده و به یک لایه تمام‌متصل (FC) برای پیش‌بینی دسته‌بندی احساس (مثبت، منفی، خنثی) ارسال می‌شوند.

۳. تحلیل داده‌ها و ارزیابی

کد شامل یک سناریوی آزمایشی با داده‌های فرضی (مشابه فرمت استاندارد (۲۰۱۵-Twitter است:

چالش فنی	احساس واقعی	جنبه (Aspect)	متن نمونه
کنایه (Sarcasm)؛ متن مثبت است اما تصویر شکست را نشان می‌دهد.	منفی (Negative)	Move	"That was a smart move"

مقایسه عملکرد (Evaluation)

سیستم در پایان دو حالت را با هم مقایسه می‌کند:

۱. Text-Only: پیش‌بینی فقط بر اساس متن (با صفر کردن اثر تصویر).

۲. Multimodal: پیش‌بینی با استفاده از هر دو ورودی متن و تصویر.

نتایج به‌صورت نمودار میله‌ای نمایش داده می‌شوند تا بهبود دقت (Accuracy) و معیار F1-Score در حالت چندرسانه‌ای نسبت به حالت تک‌رسانه‌ای (فقط متن) اثبات شود.

۴. نیازمندی‌های سیستم

برای اجرای صحیح این کد، علاوه بر کتابخانه‌های نصب شده در ابتدای فایل (torch, transformers, diffusers, sam, clip)، فایل وزن‌های مدل SAM با نام sam_vit_h_4b8939.pth باید دانلود شده باشد.