

گزارش کد

این پروژه از ترکیب متن و تصویر برای درک دقیق‌تر احساسات (به‌ویژه در موارد پیچیده مانند کنایه) استفاده می‌کند.

۱. ساختار کلی پروژه

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

تولیدکننده (**Generator**): ساخت تصاویر مرتبط با متن .

اصلاحکننده (**Refiner**): متمرکز کردن تصویر بر روی جنبه (**Aspect**) مورد نظر .

طبقه‌بندی‌کننده (**Classifier**): ترکیب ویژگی‌های متنی و بصری برای تشخیص نهایی احساس .

۲. بررسی ماژول‌های فنی

الف) تولید تصویر مصنوعی (**generator.py**)

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

هدف: شبیه‌سازی داده‌های چندرسانه‌ای 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، بخش‌های اضافی را حذف کرده و تمرکز مدل را بر روی سوژه اصلی (**Aspect**) قرار می‌دهد .

ج) مدل نهایی (**classifier.py**)

این هسته اصلی سیستم است که از معماری عصبی چندرسانه‌ای بهره می‌برد:

- بخش متنی: از مدل **BERT** برای استخراج ویژگی‌های زبان طبیعی استفاده می‌کند .

- بخش بصری (az مدل) 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 Twitter-15 است:

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

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 به صورت نمودار میله‌ای نمایش داده می‌شوند تا بهبود دقت (Accuracy) و معیار F1-Score در حالت چندرشته‌ای نسبت به حالت تک‌رشته‌ای (فقط متن) اثبات شود.

۴. نیازمندی‌های سیستم

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