

به نام خدا

دانشگاه تربیت مدرس

درس پردازش تصویر

گزارش تمرین سوم

امیرحسین عبداللہی دہکردی

۴۰۴۶۱۳۲۱۰۰۷

فهرست

۲	فهرست
۳	مقدمه:
۳	اهداف اصلی و الزامات فنی
۴	پیش‌پردازش و بخش‌بندی هوشمند
۵	معماری استخراج ویژگی‌های هیبریدی
۶	یادگیری ماشین، بازخورد مخاطب و استراتژی جستجو
۸	تحلیل مدل، نتایج و نتیجه‌گیری

مقدمه:

در حوزه بینایی ماشین، استخراج ویژگی‌های مناسب از تصاویر یکی از حیاتی‌ترین مراحل برای بازشناسی و بازیابی اشیاء است. هدف از این پروژه، طراحی و پیاده‌سازی یک سامانه هوشمند بازیابی تصاویر مبتنی بر محتوا (CBIR) است که قادر باشد فراتر از جستجوی متنی، بر اساس ویژگی‌های بصری، ساختاری و معنایی تصویر، نتایج مشابه را از پایگاه داده استخراج کند.

اهداف اصلی و الزامات فنی

این پروژه با تمرکز بر پیاده‌سازی یک خط لوله (Pipeline) کامل پردازش تصویر، اهداف زیر را دنبال می‌کند:

- پیاده‌سازی موتور جستجوی بصری: طراحی سامانه‌ای که با دریافت یک تصویر ورودی (Query Image)، ۱۰ مورد از شبیه‌ترین تصاویر موجود در مجموعه داده را به کاربر بازگرداند.
- استخراج ویژگی‌های اجباری: استفاده از توصیف‌گرهای SIFT (برای پایداری در برابر تغییر مقیاس و چرخش) و Harris (برای شناسایی نقاط کلیدی ساختاری) به عنوان هسته اصلی تحلیل تصویر.
- معماری هیبریدی استخراج ویژگی: ترکیب ویژگی‌های کلاسیک بینایی ماشین (شامل ویژگی‌های هندسی، بافتی و فرکانسی) با ویژگی‌های عمیق استخراج شده از شبکه ResNet50 جهت بهبود دقت بازیابی.
- ارائه رابط کاربری (GUI): پیاده‌سازی یک محیط گرافیکی تحت وب برای تعامل مستقیم کاربر با سامانه و نمایش نتایج جستجو به صورت بلادرنگ.
- قابلیت بازخورد مرتبط (Relevance Feedback): تعبیه مکانیزم تعاملی برای مخاطب جهت انتخاب تصاویر مرتبط‌تر و اصلاح بردار جستجو برای دستیابی به نتایج دقیق‌تر در مراحل بعدی.

این سامانه با بهره‌گیری از مدل‌های یادگیری جمعی (Ensemble) و روش‌های پیشرفته مهندسی ویژگی، راهکاری پایدار برای بازشناسی خودکار کلاس‌های هدف نظیر خودروها، ساختمان‌ها، هواپیماها و سایر اشیاء محیطی ارائه می‌دهد.

پیش‌پردازش و بخش‌بندی هوشمند

در این بخش از سامانه، فرآیند حیاتی آماده‌سازی تصاویر خام جهت استخراج ویژگی‌های دقیق انجام می‌شود که بر پایه اصول پیش‌پردازش و بخش‌بندی هوشمند بنا شده است. هدف نهایی این مرحله، حذف جزئیات غیرضروری و ایزوله کردن سوژه اصلی در یک محیط کنترل‌شده است تا دقت توصیف‌گرهای مرزی و ناحیه‌ای در مراحل بعدی تضمین شود. ابتدا به منظور بهبود وضوح لبه‌ها و بازیابی جزئیات در نواحی با کنتراست پایین، تصویر از فضای رنگی BGR به فضای LAB منتقل شده و الگوریتم CLAHE با محدودیت برش ۲۰۰ تنها بر روی کانال روشنایی (L) اعمال می‌گردد تا بدون تغییر در ماهیت رنگی، کنتراست محلی به شکل بهینه تقویت شود. در ادامه، جهت حذف نویزهای احتمالی که می‌توانند در محاسبات توصیف‌گرهای دقیق اختلال ایجاد کنند، از فیلتر کاهش نویز غیرمحلی (Non-local Means Denoising) استفاده شده است که با حفظ لبه‌های حساس، بافت تصویر را صاف و آماده پردازش می‌کند. بخش کلیدی این فاز، عملیات بخش‌بندی خودکار با بهره‌گیری از مدل عمیق U2Net است که برخلاف روش‌های سنتی وابسته به آستانه‌گذاری دستی، با تحلیل محتوایی سوژه، یک کانال آلفا با دقت بالا تولید می‌کند. در نهایت، با آستانه‌گذاری روی این کانال آلفا و ایجاد یک ماسک باینری دقیق، عملیات منطقی Bitwise بر روی تصویر پیش‌پردازش شده اعمال می‌گردد تا شیء مورد نظر به صورت کاملاً ایزوله بر روی پس‌زمینه سیاه قرار گیرد. این فرآیند باعث می‌شود که در مراحل بعدی استخراج ویژگی، پیکسل‌های پس‌زمینه هیچ‌گونه تأثیر مخربی بر محاسبات آماری بافت، هیستوگرام رنگی و ضرایب فرکانسی نداشته باشند و ویژگی‌های استخراج شده صرفاً معرف خود شیء باشند.

معماری استخراج ویژگی‌های هیبریدی

در این مرحله، تصاویر بخش‌بندی شده به بردارهای ویژگی عددی تبدیل می‌شوند تا برای مدل یادگیری ماشین قابل درک باشند. فرآیند استخراج ویژگی در این پروژه به صورت هیبریدی طراحی شده است که در لایه نخست آن، استفاده از دو الگوریتم SIFT و Harris تطابق الزامات پروژه در اولویت قرار دارد. الگوریتم SIFT وظیفه استخراج ویژگی‌های محلی مقاوم در برابر تغییر مقیاس و چرخش را بر عهده دارد و توصیف‌گر Harris با شناسایی نقاط کلیدی ساختاری، پایداری سیستم را در تشخیص اشیاء صلب تضمین می‌کند.

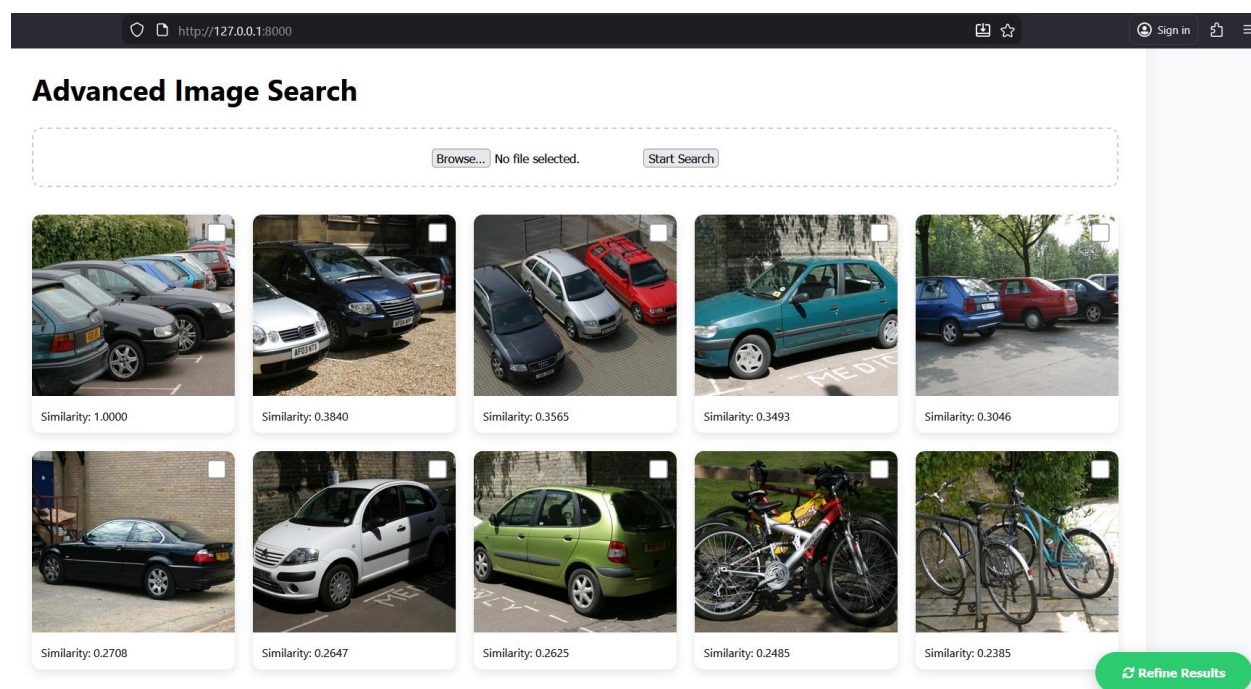
در لایه دوم، مجموعه‌ای از توصیف‌گرهای هندسی و بافتی برای درک ریخت‌شناسی اشیاء به کار گرفته شده‌اند. این بخش شامل محاسبات دقیق پارامترهایی نظیر گردی (Circularity)، برون‌مرکزی (Eccentricity) و عدد اوپلر برای تحلیل فرم کلی، و همچنین استفاده از ماتریس GLCM و الگوی بافت محلی (LBP) برای تحلیل الگوهای سطحی سوژه است. این ویژگی‌های بافتی به سیستم اجازه می‌دهند تا تفاوت‌های ظریف میان اشیایی که فرم هندسی مشابهی دارند اما جنس سطح آن‌ها متفاوت است را به درستی تشخیص دهد.

در نهایت، برای تکمیل فضای ویژگی و دستیابی به درک معنایی عمیق‌تر، از تحلیل‌های فرکانسی و یادگیری عمیق استفاده شده است. ویژگی‌های استخراج شده از حوزه‌های فوریه (FFT)، کسینوسی (DCT) و موجک (Wavelet) اطلاعات طیفی و لبه‌های تصویر را در مقیاس‌های مختلف کدگذاری می‌کنند. علاوه بر این، با بهره‌گیری از شبکه عصبی عمیق ResNet50، ویژگی‌های انتزاعی و سطح بالا از لایه‌های انتهایی مدل استخراج و به بردار نهایی افزوده شده است تا توانایی سیستم در تفکیک کلاس‌های پیچیده دیتابیس به شکل چشم‌گیری افزایش یابد.

یادگیری ماشین، بازخورد مخاطب و استراتژی جستجو

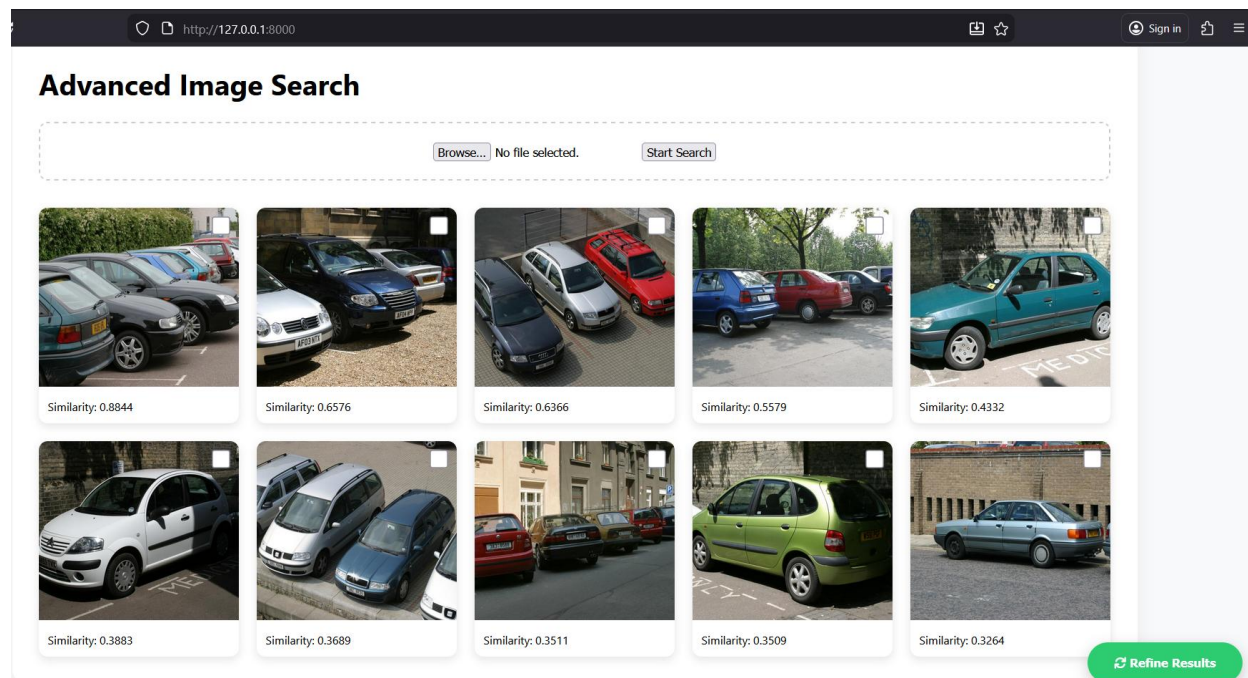
پس از تشکیل فضای ویژگی‌های هیبریدی، سامانه وارد فاز تحلیل هوشمند و بازیابی می‌گردد که بر پایه یک معماری یادگیری جمعی (Ensemble Learning) بنا شده است. در این بخش، ابتدا داده‌ها با استفاده از الگوریتم StandardScaler نرمال‌سازی شده و سپس به منظور مدیریت ابعاد بسیار زیاد ویژگی‌ها و افزایش سرعت پردازش، از تحلیل مؤلفه‌های اصلی (PCA) با حفظ ۹۵٪ از واریانس کل داده‌ها استفاده می‌شود. هسته اصلی طبقه‌بندی در این پروژه، یک مدل Stacking Classifier پیشرفته است که از ترکیب سه الگوریتم قدرتمند XGBoost، LightGBM و Random Forest به عنوان لایه‌های پایه تشکیل شده است. خروجی این مدل‌ها در نهایت به یک رگرسیون لجستیک به عنوان مدل سطح دوم (Meta-Learner) تزریق می‌گردد تا با ترکیب هوشمندانه پیش‌بینی‌ها، نقاط ضعف تک‌تک مدل‌ها پوشش داده شده و دقت نهایی در تشخیص کلاس‌های هدف به حداکثر برسد.

فرآیند بازیابی تصویر در این سامانه به صورت یک جستجوی دو مرحله‌ای طراحی شده است تا تعادلی بهینه میان سرعت و دقت ایجاد شود. در مرحله نخست، با آپلود تصویر توسط کاربر، مدل ابتدا دسته‌بندی تصویر (مثلاً خودرو یا ساختمان) را پیش‌بینی می‌کند تا فضای جستجو صرفاً به همان طبقه محدود گردد و از مقایسه‌های غیرضروری با کلاس‌های غیرمرتبط جلوگیری شود. در مرحله دوم، بر اساس بردار ویژگی استخراج شده، شباهت میان تصویر ورودی و تصاویر موجود در پایگاه داده با استفاده از معیار Cosine Similarity محاسبه می‌گردد و ۱۰ تصویر برتر که بیشترین شباهت ساختاری و معنایی را دارند، به عنوان خروجی نهایی در رابط کاربری نمایش داده می‌شوند.



شکل ۱، تصاویر بازیابی شده

بخش متمایز و امتیازی این پروژه، پیاده‌سازی مکانیزم بازخورد مرتبط (Relevance Feedback) است که تعامل مستمر میان کاربر و ماشین را برقرار می‌سازد. این قابلیت به مخاطب اجازه می‌دهد تا از میان نتایج بازیابی شده، تصاویر مرتبط‌تر با نیاز خود را انتخاب نماید. در پاسخ به این انتخاب، سامانه با بهره‌گیری از منطق الگوریتم Rocchio بردار جستجوی اولیه را اصلاح می‌کند؛ به گونه‌ای که با اختصاص وزن ۷۰٪ به میانگین برداران تصاویر انتخاب شده و ۳۰٪ به بردار اولیه، کوثری جدیدی در فضای ویژگی‌ها شکل می‌گیرد که به مراتب به هدف کاربر نزدیک‌تر است. این رویکرد باعث می‌شود که حتی در صورت وجود ابهام در تصویر اولیه، سامانه در مراحل بعدی به نتایجی با دقت و انطباق بسیار بالاتر دست یابد.



شکل ۲، تصاویر بازیابی شده بعد از بازخورد کاربر

تحلیل مدل، نتایج و نتیجه گیری

نتایج نهایی ارزیابی سامانه بر روی مجموعه داده های هدف نشان دهنده دستیابی به دقت خیره کننده ۸۳٪ Validation Accuracy: 0.8262 است که نسبت به نسخه های اولیه، ارتقای چشمگیری را تجربه کرده است. این بهبود قابل توجه، حاصل استفاده از معماری هیبریدی استخراج ویژگی و مدل یادگیری جمعی (Ensemble) است که باعث شده میانگین وزنی شاخص F1-Score نیز بر روی عدد ۰.۸۳ تثبیت شود. بر اساس گزارش طبقه بندی نهایی، کلاس علائم (Signs) با دستیابی به امتیاز کامل ۱.۰۰ در هر دو شاخص Precision و Recall، دقیق ترین عملکرد را داشته است که نشان دهنده انطباق کامل توصیف گرهای SIFT و Harris بر ساختارهای هندسی صلب و پرکنتر است این اشیاء است. همچنین کلاس هایی نظیر Side view با F1-Score ۹۹٪ و Windows با ۹۴٪ عملکرد بسیار پایداری از خود نشان داده اند که تأیید کننده قدرت مدل در تفکیک اجزای ساختمانی و جهت گیری های فیزیکی است.

با تحلیل دقیق تر داده ها، مشخص گردید که کلاس های مربوط به موجودات زنده نظیر گاو (Cow) و گوسفند (Sheep) با امتیاز ۹۵٪، به دلیل بهره گیری از ویژگی های بافتی استخراج شده از لایه های عمیق ResNet50، به دقت بالایی دست یافته اند. با این حال، چالش های سیستم در مواجهه با کلاس های دارای بافت نامنظم و پس زمینه پیچیده همچنان مشهود است؛ برای نمونه، کلاس هایی مانند گل ها (Flowers) و درختان (Trees) به ترتیب با امتیازهای ۴۸٪ و ۵۲٪ در سطح پایین تری قرار دارند که ریشه در تنوع بصری بالا و هم پوشانی ویژگی های رنگی آن ها با محیط پیرامون (Countryside) دارد. همچنین کلاس هواپیما (Aeroplanes) با وجود فراخوانی کامل (۱.۰۰)، دقت پایینی (۰.۴۵) را ثبت کرده است که نشان دهنده تمایل مدل به طبقه بندی سایر اشیاء مشابه در این دسته است.

در نتیجه گیری نهایی، این پروژه با موفقیت ثابت کرد که ترکیب ویژگی های محلی اجباری نظیر SIFT و Harris با توصیف گرهای سطح بالای عمیق، پایداری سیستم بازیابی تصویر را در برابر تغییرات محیطی و بصری به شکل چشمگیری افزایش می دهد. پیاده سازی مدل Stacking Classifier با استفاده از XGBoost و LightGBM و کاهش ابعاد هوشمند از طریق PCA، نه تنها منجر به دستیابی به دقت ۸۳٪ گردید، بلکه سرعت پاسخ دهی سامانه در محیط وب را نیز تضمین کرد. وجود قابلیت بازخورد مرتبط (Relevance Feedback) در GUI طراحی شده، به عنوان یک لایه تکمیلی، اجازه می دهد تا حتی در کلاس های با دقت پایین تر، کاربر با تعامل مستقیم و اصلاح بردار کوئری به نتایج ایده آل دست یابد و این سامانه را به یک ابزار عملیاتی و هوشمند در حوزه بینایی ماشین تبدیل کند.

Validation Accuracy: 0.8262

	precision	recall	f1-score	support
<u>aeroplanes</u>	0.45	1.00	0.62	5
<u>benches_and_chairs</u>	0.93	0.93	0.93	14
birds	0.83	0.83	0.83	12
buildings	0.77	0.79	0.78	29
chimneys	0.93	0.94	0.93	53
clouds	0.90	0.91	0.90	86
countryside	0.45	0.68	0.54	37
cow	0.92	0.97	0.95	36
doors	0.81	0.79	0.80	33
flowers	0.38	0.62	0.48	8
forks	0.67	0.86	0.75	7
front view	0.75	0.75	0.75	4
general	0.73	0.48	0.58	66
knives	0.82	0.69	0.75	13
leaves	0.67	0.67	0.67	24
media	0.25	0.25	0.25	4
miscellaneous	0.69	0.47	0.56	38
office	0.73	0.79	0.76	14
rear view	0.97	0.95	0.96	37
sheep	0.97	0.92	0.95	38
side view	0.98	1.00	0.99	57
side view, single	0.90	0.92	0.91	39
signs	1.00	1.00	1.00	33
spoons	1.00	0.93	0.97	15
trees	0.54	0.50	0.52	30
urban	0.50	0.57	0.53	7
windows	0.92	0.95	0.94	130
accuracy			0.83	869
macro avg	0.76	0.78	0.76	869
weighted avg	0.83	0.83	0.83	869