



دانشگاه الزهراء (س)
دانشکده فنی-مهندسی

پایان نامه کارشناسی ارشد
رشته مهندسی کامپیوتر گرایش نرم افزار

عنوان

جستجوی بصری در سیستم بازیابی تصویر بر اساس محتوا

دانشجو
سروین ساروی

استاد راهنما
دکتر مسعود ساغریچیان

مهر ماه سال ۱۴۰۲

Title: Visual Search in Content-Based Image Retrieval System	
Surname: Saravi	Name: Sarvin
First Supervisor: Dr. Masoud Sagharichian Affiliation: Department of Computer., Faculty of Engineering, Alzahra University, Tehran, Iran.	Second Supervisor: Affiliation: Department of, Faculty of,.... (University).....,(city),(country)...
First Advisor: Affiliation: Department of, Faculty of,.... (University).....,(city),(country)...	Second Advisor: Affiliation: Department of, Faculty of,.... (University).....,(city),(country)...
Degree: Master Department: Computer Faculty: Engineering Alzahra university	
Subject: Software Engineering	Field: Image Retrieval
Abstract: <p>Today, content-based image retrieval has become an active and widely used research area in many fields of computer science and computer vision. Content-based image retrieval and visual search systems allow users to find related images based on content similarity with the desired image, to obtain the best results in image retrieval by using textual or image inputs. Researches in this field are mainly focused on providing solutions to increase speed and accuracy, optimal use of resources, as well as improving the experience of searching and retrieving images, and finding the right method in this field is still a challenging and attractive thing in industrial and academic environments.</p> <p>Today, with the evolution of image recording and production tools, as well as the development of the Internet and cloud space, we have witnessed a significant growth in the number and dimensions of available image data and an increase in the demand for browsing and searching among them, and this has led to the creation of the concept of visual search. The challenges of visual search on a very high scale and dimensions have recently become the most important issue in this field, and efforts to propose solutions to deal with them have been especially noticed by researchers.</p> <p>Therefore, in this research, we decided to investigate the previous methods in the field of image recovery and research that provide optimal solutions to face the challenges in this field, and develop our proposed solution. In short, in this solution, we use Elastic Search for storage and search, and with the help of ResNet101 deep neural network and algorithm for creating alternative text for images, we perform the recovery process. We tested our approach on the INRIA Holidays and Flickr1M datasets and were able to propose a method that maximizes search speed while maintaining retrieval precision.</p>	
Keywords: Content-based Image Retrieval, Visual Search, High-dimensional data, Image indexing, Elasticsearch	

عنوان: جستجوی بصری در سیستم بازیابی تصویر مبتنی بر محتوا	
نام خانوادگی: ساروی	نام: سروین
استاد راهنمای اول: دکتر مسعود ساغریچیان وابستگی سازمانی: گروه کامپیوتر، دانشکده فنی مهندسی، دانشگاه الزهراء، تهران، ایران.	استاد راهنمای دوم: وابستگی سازمانی: گروه.....، دانشکده.....، (دانشگاه).....، (شهر).....، (کشور).....
استاد مشاور اول: وابستگی سازمانی: گروه.....، دانشکده.....، (دانشگاه).....، (شهر).....، (کشور).....	استاد مشاور دوم: وابستگی سازمانی: گروه.....، دانشکده.....، (دانشگاه).....، (شهر).....، (کشور).....
مقطع: کارشناسی ارشد گروه: مهندسی کامپیوتر دانشکده: فنی مهندسی دانشگاه الزهراء^(س)	
رشته / گرایش: مهندسی نرم افزار	زمینه پژوهش: بازیابی تصویر
<p> چکیده: بازیابی تصویر مبتنی بر محتوا امروزه به حوزه تحقیقاتی فعال و پرکاربرد در بسیاری از زمینه‌های علوم کامپیوتر و بینایی ماشین تبدیل شده است. سیستم‌های بازیابی تصویر مبتنی بر محتوا و جستجوی بصری، برای یافتن تصاویر مرتبط بر اساس مشابهت محتوایی با تصویر موردنظر، به کاربران امکان می‌دهند تا با استفاده از ورودی‌های متنی یا تصویری، بهترین نتایج را در بازیابی تصاویر به دست آورند. پژوهش‌های این حوزه، عمدتاً متمرکز بر ارائه راهکارهایی برای بالا بردن سرعت و دقت، استفاده بهینه از منابع و همچنین بهبود تجربه جستجو و بازیابی تصاویر هستند که یافتن روش مناسب در این حوزه همچنان امری چالش‌برانگیز و جذاب در محیط‌های صنعتی و دانشگاهی می‌باشد. </p> <p> امروزه با تکامل ابزارهای ثبت و تولید تصاویر و همچنین پیشرفت اینترنت و فضای ابری، شاهد رشد چشمگیر در تعداد و ابعاد داده‌های تصویری موجود و افزایش تقاضا برای مرور و جستجو در میان آن‌ها هستیم و همین امر منجر به ایجاد مفهوم جستجوی بصری شده است. چالش‌های جستجوی بصری در مقیاس و ابعاد بسیار بالا، اخیراً به مهم‌ترین مسئله در این حوزه بدل گشته و تلاش برای پیشنهاد راهکارهای مقابله با آن‌ها به طور ویژه موردتوجه محققان قرار گرفته است. </p> <p> از این رو تصمیم گرفتیم تا در این پژوهش به بررسی روش‌های پیشین در حوزه بازیابی تصویر و تحقیقاتی که به ارائه راهکارهای بهینه برای مقابله با چالش‌ها در این زمینه می‌پردازند، اقدام کنیم و به توسعه راهکار پیشنهادی خود بپردازیم. به طور خلاصه در این راهکار از الستیک سرچ برای ذخیره‌سازی و جستجو استفاده می‌کنیم و با کمک شبکه عصبی عمیق ResNet101 و الگوریتم ایجاد متن جایگزین برای تصاویر، عمل بازیابی را انجام می‌دهیم. </p> <p> ما رویکرد خود را بر روی مجموعه داده‌های INRIA Holidays و Flickr1M مورد آزمایش قرار دادیم و توانستیم روشی را پیشنهاد کنیم که با حفظ دقت بازیابی، تا حد امکان سرعت جستجو را افزایش دهد. </p>	
کلیدواژه: بازیابی تصاویر مبتنی بر محتوا، جستجوی بصری، داده با ابعاد بالا، نمایه‌سازی تصاویر، الستیک سرچ	

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه الزهراء (س)

دانشکده فنی-مهندسی

پایان نامه کارشناسی ارشد

رشته مهندسی کامپیوتر گرایش نرم افزار

عنوان

جستجوی بصری در سیستم بازیابی تصویر بر اساس محتوا

دانشجو

سروین ساروی

استاد راهنما

دکتر مسعود ساغریچیان

مهر ماه سال ۱۴۰۲



تعهد اصالت اثر دانشگاه الزهرا^(س)

اینجانب **سروین ساروی** دانش‌آموخته مقطع کارشناسی ارشد در رشته مهندسی کامپیوتر گرایش نرم افزار که در تاریخ ۱۴۰۲/۰۷/۱۵ از پایان‌نامه خود تحت عنوان:

جستجوی بصری در سیستم بازیابی تصویر بر اساس محتوا

با کسب درجه خیلی خوب دفاع نموده‌ام، متعهد می‌شوم:

۱. این پایان‌نامه/رساله دستاورد پژوهش اینجانب بوده و محتوای آن از درستی و اصالت برخوردار است.

۲. این پایان‌نامه/رساله و محتوای آن تاکنون توسط اینجانب یا فرد دیگری برای دریافت مدرک یا امتیاز در هیچ کجا ارائه نشده است.

۳. مقالات مستخرج از این پایان‌نامه/رساله کاملاً منطبق بر آن بوده و از هرگونه جعل داده و یا تغییر اطلاعات پرهیز خواهم نمود.

۴. در همه آثار مستخرج از این پایان‌نامه/رساله، نام استاد(ان) راهنما و در صورت تشخیص استاد راهنمای نخست، نام استاد(ان) مشاور و نشانی رایانامه سازمانی آنان را درج خواهم کرد.

۵. حق نشر، تکثیر و درآمدهای حاصل از این پایان‌نامه/رساله مطابق آیین‌نامه مالکیت فکری و تجاری‌سازی دانشگاه الزهرا مصوب هیات امناء دانشگاه مورخ ۱۳۹۷/۱۱/۳۰ متعلق به دانشگاه است و هرگونه بهره‌مندی و یا نشر دستاوردهای حاصل از این تحقیق اعم از چاپ کتاب، مقاله و ... چه در زمان دانشجویی و یا بعد از فراغت از تحصیل، با کسب اجازه از معاونت پژوهشی دانشگاه مجاز می‌باشد و چنانچه تدوین پایان‌نامه/رساله منجر به اختراع و یا دستیابی به دانش فنی جدید شود، مشمول احکام مرتبط با اختراع اثر و یا نوآوری، مندرج در آیین‌نامه فوق خواهد بود.

۶. حقوق معنوی همه کسانی را که در به‌دست آمدن نتایج اصلی پایان‌نامه/رساله تاثیرگذار بوده‌اند، رعایت کرده و هنگام به کار بردن دستاورد پژوهش‌های دیگران در آن، با دقت و به درستی به آن‌ها استناد نموده‌ام، درغیراین صورت برابر قوانین و مقررات مسئول بوده و دانشگاه در این خصوص مسئولیتی ندارد.

۷. در صورت اثبات تقلب در تهیه پایان‌نامه/رساله، برابر قانون پیشگیری و مقابله با تقلب در وزارت علوم با دانشجو رفتار خواهد شد.

نام و نام خانوادگی دانشجو: **سروین ساروی**

تاریخ و امضا: آبان ماه ۱۴۰۲

این پایان‌نامه/رساله با حمایت مالی انجام رسیده است.

(ویژه پایان‌نامه/رساله‌هایی که موافقت معاونت پژوهشی دانشگاه الزهرا و سازمان حمایت کننده را دارند.)



کد: GM-FR-04-00

صفحه: ۱ از ۱

صورت جلسه دفاع از پایان نامه کارشناسی ارشد

شماره: ۰۲/۱۰۲۹۸/د

تاریخ: ۱۴۰۲/۰۷/۱۲

تاریخ تهیه: ۱۴۰۱/۰۹/۲۷

بازنگری: ۰۰

جلسه دفاع از پایان نامه کارشناسی ارشد خاتم سروین ساروی به شماره دانشجویی ۹۹۲۵۰۳۱۰۱۲ رشته مهندسی کامپیوتر-ترم افزار به تعداد ۶ واحد با عنوان: جستجوی بصری در سیستم بازیابی تصویر بر اساس محتوا در تاریخ ۱۴۰۲/۰۷/۱۵ از ساعت ۱۵:۰۰ تا ۱۷:۰۰ با حضور هیات داوران تشکیل شد و پس از ارزیابی، اعتبار پایان نامه برای اخذ مدرک کارشناسی ارشد به شرح زیر مورد تایید است:

الف) قبول با درجه

ب) مردود (نمره کمتر از ۱۴)

..... عالی (نمره ۱۹ تا ۲۰)

✓ خیلی خوب (نمره ۱۸ تا ۱۸/۹۹)

..... خوب (نمره ۱۶ تا ۱۷/۹۹)

..... متوسط (نمره ۱۴ تا ۱۵/۹۹)

عنوان نهایی: شان عرفان

نام و نام خانوادگی	سمت	مرتبه علمی	دانشگاه یا موسسه	امضا
دکتر مسعود ساغری بچیان	راهنما	استادیار	دانشگاه الزهرا	
دکتر وجیهه ثابتی	داور داخلی	استادیار	دانشگاه الزهرا	
دکتر جلال الدین نصیری	داور خارجی	دانشیار	دانشگاه فردوسی مشهد	

گزارش ناظر:

الف) دفاع نامبرده با شرایط فوق و با توجه به شرح وظایف استاد ناظر مورد تایید است. ✓
ب) هرگونه مغایرت با اطلاعات مندرج در دعوتنامه استاد ناظر جلسه به شرح زیر است:

نام و نام خانوادگی استاد ناظر: وجیهه ثابتی

نام و نام خانوادگی مدیر گروه: محمد رضا کیوان پور

نام و نام خانوادگی معاون آموزشی/رییس دانشکده: نوشین ریاحی

امضا و تاریخ
۱۵/۱۱/۱۴۰۲
امضا و تاریخ
امضا و تاریخ

تقدیم به

آغوش گرم مادرم

و دست های سراسر مهر پدرم

(گرچه می دانم این صفحات ناچیز ذره ای از محبت های بی شمارشان را جبران نخواهد کرد...)

یاد بعضی نفرات روشم می دارد...

(نیما یوشیج)

نخست سپاس پروردگار یگانه را که هستی‌ام بخشید و مرا در راه تعلیم دانش رهنمون گرداند و افتخار همنشینی با رهروان این طریق را بر من ارزانی داشت.

همچنین با سپاس بی کران از پدر و مادر عزیزم که لطف و محبت و دلسوزی شایان خود را چراغ راه زندگی‌ام نمودند و بی شک بدون همراهی و همدلی و همگامی‌شان، مشعل این راه بی فروغ می ماند.

در اینجا بر خود لازم می دانم از زحمات ارزشمند استاد مبتکر و خلاق جناب آقای دکتر مسعود ساغریچیان نهایت تشکر و امتنان را به جای آورم که در طول مسیر این پژوهش از افتخار همکاری و راهنمایی‌های سازنده‌ی ایشان برخوردار بودم.

و نیز بسی شایسته است که در اینجا از تلاش و مشارکت بی دریغ دو پژوهشگر گرامی سرکار خانم مهندس منیبا رون و جناب آقای دکتر میررضا میرمعروف زبینه، کمال سپاس و قدردانی را داشته باشم.

در پایان از تمامی دوستان و عزیزانی که هر یک به نوعی مرا در ارائه هر چه بهتر این اثر یاری کرده اند، سپاسگزاری می کنم.

امیدوارم که این تلاش برای عده‌ای هر چند اندک، سودمند واقع گردد.

سروین ساروی

اما شکری بی پایان به خودم می‌یونم
به کسی که آن روزها بودم،
کسی که زندگی کرد
با این که نمی‌خواست...

برگرفته از کتاب:

اشکال ندارد که حالت خوب نیست.

نوشته‌ی:

مگان دیواین

چکیده

بازیابی تصویر مبتنی بر محتوا امروزه به حوزه تحقیقاتی فعال و پرکاربرد در بسیاری از زمینه‌های علوم کامپیوتر و بینایی ماشین تبدیل شده است. سیستم‌های بازیابی تصویر مبتنی بر محتوا و جستجوی بصری، برای یافتن تصاویر مرتبط بر اساس مشابهت محتوایی با تصویر موردنظر، به کاربران امکان می‌دهند تا با استفاده از ورودی‌های متنی یا تصویری، بهترین نتایج را در بازیابی تصاویر به دست آورند. پژوهش‌های این حوزه، عمدتاً متمرکز بر ارائه راهکارهایی برای بالا بردن سرعت و دقت، استفاده بهینه از منابع و همچنین بهبود تجربه جستجو و بازیابی تصاویر هستند که یافتن روش مناسب در این حوزه همچنان امری چالش‌برانگیز و جذاب در محیط‌های صنعتی و دانشگاهی می‌باشد.

امروزه با تکامل ابزارهای ثبت و تولید تصاویر و همچنین پیشرفت اینترنت و فضای ابری، شاهد رشد چشمگیر در تعداد و ابعاد داده‌های تصویری موجود و افزایش تقاضا برای مرور و جستجو در میان آن‌ها هستیم و همین امر منجر به ایجاد مفهوم جستجوی بصری شده است. چالش‌های جستجوی بصری در مقیاس و ابعاد بسیار بالا، اخیراً به مهم‌ترین مسئله در این حوزه بدل گشته و تلاش برای پیشنهاد راهکارهای مقابله با آن‌ها به طور ویژه موردتوجه محققان قرار گرفته است.

از این رو تصمیم گرفتیم تا در این پژوهش به بررسی روش‌های پیشین در حوزه بازیابی تصویر و تحقیقاتی که به ارائه راهکارهای بهینه برای مقابله با چالش‌ها در این زمینه می‌پردازند، اقدام کنیم و به توسعه راهکار پیشنهادی خود بپردازیم. به طور خلاصه در این راهکار از الستیک سرچ برای ذخیره‌سازی و جستجو استفاده می‌کنیم و با کمک شبکه عصبی عمیق ResNet101 و الگوریتم ایجاد متن جایگزین برای تصاویر، عمل بازیابی را انجام می‌دهیم.

ما رویکرد خود را بر روی مجموعه داده‌های INRIA Holidays و Flickr1M مورد آزمایش قرار دادیم و توانستیم روشی را پیشنهاد کنیم که با حفظ دقت بازیابی، تا حد امکان سرعت جستجو را افزایش دهد.

کلیدواژه‌ها:

بازیابی تصاویر مبتنی بر محتوا، جستجوی بصری، داده با ابعاد بالا، نمایه‌سازی تصاویر، الستیک سرچ

فهرست مطالب

فصل اول	۱۹
۱. مقدمه	۲۱
۱-۱. بیان مسئله	۲۲
بازیابی تصاویر بر اساس متن	۲۲
بازیابی تصاویر بر اساس محتوا	۲۲
۲-۱. اهمیت و ضرورت	۲۳
۳-۱. اهداف	۲۵
۴-۱. پرسش‌های پژوهشی	۲۵
گروه اول	۲۵
گروه دوم	۲۶
گروه سوم	۲۶
۵-۱. چالش‌ها	۲۷
انتخاب روش مناسب نمایش تصویر در قالب بردار ویژگی	۲۷
ابعاد بسیار بالا	۲۷
شکاف مقصود	۲۷
شکاف معنایی	۲۸
مجموعه داده	۲۸
ورودی پویا	۲۸
بهبود تعامل کاربران	۲۹
معرفی بهترین روش ایندکس و جستجو	۲۹
۶-۱. ساختار کلی پایان نامه	۲۹
فصل دوم	۳۱
۲. مروری بر پژوهش‌های انجام شده	۳۳
۱-۲. تاریخچه	۳۳
۲-۲. ساختار کلی سیستم‌های بازیابی تصویر مبتنی بر محتوا	۳۴
۳-۲. نمایش تصویر	۳۶
۴-۲. نمایه سازی	۳۹
۵-۲. جستجوی شباهت	۴۱
۶-۲. مرور کارهای مرتبط	۴۲
۷-۲. جمع بندی	۴۶
فصل سوم	۴۷
۳. روش پیشنهادی	۴۹
۱-۳. مقدمه	۴۹

۵۰	۲-۳. پیش زمینه.....
۵۱	معرفی الستیک سرچ.....
۵۳	معرفی شبکه عصبی مصنوعی عمیق RESTNET101.....
۵۴	۳-۳. سیستم پیشنهادی.....
۵۴	بخش اول - استخراج ویژگی.....
۵۵	بخش دوم - پردازش و تولید نمایش متنی.....
۵۸	بخش سوم - ذخیره در الستیک سرچ.....
۵۹	بخش چهارم - جستجو و ارزیابی.....
۶۱	۴-۳. جزییات بخش ذخیره سازی و جستجو.....
۶۲	ایده نخست - جستجوی فازی.....
۶۴	ایده دوم - جستجو عین عبارت.....
۶۵	ایده سوم - جستجو با حذف تکرار.....
۶۶	ایده چهارم - جستجو با در نظر گرفتن پیشوندها.....
۶۸	ایده پنجم - جستجو بر اساس پارتیشن بندی.....
۷۰	۵-۳. جمع بندی.....
۷۱	فصل چهارم.....
۷۳	۴. پیاده سازی و ارائه نتایج.....
۷۴	۱-۴. مشخصات سیستم مورد آزمایش.....
۷۴	۲-۴. مجموعه داده.....
۷۴	مجموعه داده تعطیلات (INRIA HOLIDAYS DATASET).....
۷۵	مجموعه داده فلیکر ۱ میلیون (FLICKR1M).....
۷۷	۳-۴. معیارهای ارزیابی.....
۷۹	۴-۴. آزمایش ها در مجموعه داده تعطیلات.....
۸۶	مقایسه با روش های دیگر.....
۸۷	۵-۴. آزمایش در مجموعه داده فلیکر.....
۸۹	۶-۴. جمع بندی.....
۹۱	فصل پنجم.....
۹۳	۵. جمع بندی.....
۹۳	۱-۵. نتیجه گیری.....
۹۴	۲-۵. پژوهش های آینده.....
۹۵	فهرست منابع و مآخذ.....
۹۹	پیوست ها.....
۱۰۱	واژه نامه توصیفی.....
۱۰۳	واژه نامه انگلیسی به فارسی.....
۱۰۵	واژه نامه فارسی به انگلیسی.....

فهرست جدول‌ها

جدول ۱-۲ : خلاصه روش‌های پیشین ۴۳

جدول ۱-۴ : مقایسه با روش‌های دیگر ۸۶

جدول ۲-۴ : مقایسه زمان برای دو روش پارتیشن بندی و حذف تکرار ۸۸

فهرست شکل‌ها

- شکل ۲-۱: مراحل کلی عملیات بازیابی تصویر مبتنی بر محتوا ۳۴
- شکل ۳-۱: چارچوب کلی روش پیشنهادی ۵۰
- شکل ۴-۱: نمونه تصاویر مجموعه داده تعطیلات ۷۵
- شکل ۴-۲: نمونه تصاویر مجموعه داده فلیکر ۱ میلیون ۷۶

فهرست پرس و جوها

- پرس و جو ۱-۳ : جستجو به روش فازی..... ۶۳
- پرس و جو ۲-۳ : جستجوی عین عبارت..... ۶۵
- پرس و جو ۳-۳ : جستجو با حذف تکرار..... ۶۶
- پرس و جو ۱-۴-۳ : جستجو با در نظر گرفتن پیشنوندها برای TEXT..... ۶۷
- پرس و جو ۲-۴-۳ : جستجو با در نظر گرفتن پیشنوندها برای TEXT و KEYWORD..... ۶۸
- پرس و جو ۵-۳ : جستجو بر اساس پارتیشن بندی..... ۶۹

فهرست نمودارها

- نمودار ۴-۱: مقایسه تاثیر متغیر FUZZINESS در جستجوی فازی..... ۸۰
- نمودار ۴-۲: مقایسه تاثیر تکرار در جستجوی ساده..... ۸۱
- نمودار ۴-۳: مقایسه تفاوت نوع داده ای در جستجو با در نظر گرفتن پیشوندها..... ۸۲
- نمودار ۴-۴: مقایسه تاثیر میزان افزایش پیشوندی..... ۸۳
- نمودار ۴-۵: مقایسه تاثیر تکرار در روش پارتیشن بندی..... ۸۴
- نمودار ۴-۶: مقایسه کلی روش ها..... ۸۵
- نمودار ۴-۷: مقایسه دقت و سرعت روش پارتیشن بندی و روش حذف تکرار..... ۸۸

فصل اول

مقدمه و کلیات پژوهش

۱. مقدمه

استفاده از تصاویر که در واقع بازتابی از واقعیت هستند، همواره یکی از روش‌های ارتباط بین انسان‌ها برای انتقال دانش و مفاهیم در طول تاریخ بوده است. جمله مشهور "یک تصویر بیشتر از هزاران کلمه ارزش دارد" را می‌توان به عنوان ضرب‌المثلی در زبان‌های مختلف دانست که بیانگر اهمیت نقش تصاویر در بیان ایده‌های پیچیده و خاص در میان اقشار گوناگون بشر در دوره‌های مختلف بوده است. ("A Picture Is Worth a Thousand Words," 2023; Mu et al., 2018)

با گذشت زمان و همراه با پیشرفت دانش و فناوری استفاده از تصاویر گسترش چشمگیری یافت. در سال‌های اخیر با توسعه دستگاه‌های دیجیتال و ابزارهای سخت‌افزاری و نرم‌افزاری متنوع برای تولید و پردازش تصویر، ثبت تصاویر از سوژه‌های موردعلاقه افراد تبدیل به امری بسیار ساده و سرگرم‌کننده شده است. همین امر سبب شده است که حجم داده‌های تصویری موجود، به صورت خیره‌کننده‌ای افزایش یابد. با پیشرفت و گسترش تکنولوژی اینترنت و امکان بارگذاری و نگهداری داده‌های بدون ساختار مانند تصاویر، به صورت آنلاین از یک سمت و همچنین پیشرفت مشترک رشته‌های علمی و زمینه‌های گوناگون فناوری‌های ارتباطی و نیز فناوری‌های ذخیره سازی از سوی دیگر، منجر به ظهور عصر داده‌های بزرگ (کلان داده‌ها) در مفاهیم مربوط به علوم کامپیوتر شد. (Camli, 2020; Y. Li et al., 2021)

امروزه با گسترش چنین داده‌هایی، نیاز به مدیریت و نگهداری داده‌های بزرگ بیش از هر زمان احساس می‌شود. در واقع این خود داده‌ها نیستند که منجر به خلق ارزش می‌شوند بلکه کسی می‌تواند ادعا کند از این داده‌ها سود برده است که قابلیت استخراج دانش در مدت زمان معین و قابل قبول از این داده‌ها را داشته باشد. (Y. Li et al., 2021) بنابراین امروزه نیاز به ذخیره سازی و جستجوی مناسب در میان این داده‌ها بیش از پیش مورد توجه قرار گرفته است.

بازیابی تصویر مبتنی بر محتوا و یا به عبارت دیگر جستجوی بصری یکی از زمینه‌های تحقیقاتی در میان علوم مربوط به هوش مصنوعی و بینایی ماشین می‌باشد که به پیدا کردن راه حل بهینه‌ای برای این نیاز فراگیر دنیای امروز می‌پردازد. اگرچه پژوهش‌ها و بهبودهای زیادی در این زمینه انجام شده است، اما مفهوم جستجوی بصری در داده‌های بزرگ هنوز یک حوزه تحقیقاتی فعال و مورد توجه چه در محیط‌های صنعتی و چه در محیط‌های دانشگاهی است. (Camli, 2020)

۱-۱. بیان مسئله

جستجوی بصری با استفاده از سیستم بازیابی تصاویر بر اساس محتوا یک حوزه بسیار مهم در عرصه تحقیقات امروزی در زمینه بینایی ماشین را شامل می‌شود. بنابراین تعریف صحیح و کاربردی این مسئله از اهمیت ویژه‌ای برخوردار است.

به طور کلی برای جستجو و بازیابی تصاویر ۲ روش عمده وجود دارد:

- بازیابی تصویر بر اساس متن (توصیف)
- بازیابی تصاویر بر اساس محتوا

بازیابی تصاویر بر اساس متن^۱

در این روش جستجوی تصاویر با استفاده از کلمات کلیدی که توسط کاربر به سیستم داده می‌شود، انجام می‌پذیرد. از این تکنیک به عنوان بازیابی تصاویر بر اساس شرح (توصیف)^۲ نیز یاد می‌شود. (Khawandi et al., 2019) در این رویکرد بازیابی در مراحل اولیه برای تعریف و شناسایی تصاویر مجموعه داده در سیستم، بیشتر تاکید بر شناسایی اشیا و استخراج و انتخاب ویژگی‌های کلی تصاویر است و از حاشیه نویسی^۳ و یا ذخیره توصیف‌های متنی اضافه به عنوان فراداده^۴ در کنار داده‌های تصویری اصلی استفاده می‌شود. در گذشته این رویکرد بیشتر مورد توجه بوده است اما امروزه به دلیل دقت کمتر در ارائه نتایج مشابه و همچنین موانع زبانی موجود در روش‌های سنتی جستجوی تصاویر، در موارد محدودتری به کار گرفته می‌شود. (Camli, 2020)

بازیابی تصاویر بر اساس محتوا

سیستم بازیابی تصاویر مبتنی بر محتوا بر یافتن تصاویر مشابه با یک تصویر پرس و جو و برگرداندن نتایج به ترتیب رتبه بندی شده تمرکز دارد که در آن مشابه‌ترین تصاویر در رتبه‌های بالاتر قرار می‌گیرند. (Camli, 2020) یا به عبارت دیگر، تکنیک بازیابی تصویر مبتنی بر محتوا یا جستجوی بصری می‌تواند تصاویر مرتبط با تصویر ورودی از یک شی یا محتوای مورد نظر ما را از یک پایگاه داده بازیابی کند. (X. Li et al., 2021)

در این پژوهش بیشتر تمرکز بر روی تکنیک بازیابی تصویر بر اساس محتوا قرار دارد و تلاش شده است تا راه حل مفیدی در این زمینه ارائه شود.

¹ Text-Based Image Retrieval

² Description Based Image Retrieval (DBIR)

³ annotation

⁴ metadata

۲-۱. اهمیت و ضرورت

همان‌طور که پیش‌تر نیز بیان شد، امروزه تصاویر به عنوان گونه‌ای از مدارک که قادر به انتقال حجم بالایی از اطلاعات هستند، مورد توجه قرار دارند. پردازش و بازیابی تصاویر حوزه‌ای است که امروزه به طور قابل توجهی شروع به رشد کرده است و چشم انداز جستجوی فعلی را هم در تحقیقات علمی-آکادمیک و هم در صنعت تغییر داده است. با توسعه فناوری اینترنت و محبوبیت دستگاه‌های دیجیتال، بازیابی تصویر مبتنی بر محتوا به طور گسترده در بسیاری از زمینه‌ها مانند پزشکی، نظامی، تجارت، هنر و غیره استفاده شده است. (Camli, 2020; X. Li et al., 2021)

استفاده گسترده شرکت‌های تجاری بزرگی همچون Google و Baidu از جستجوی بصری در موتورهای جستجوی تصاویر، نمونه‌های موفقی از کاربرد این تکنیک در زندگی روزمره افراد است. برخی از سایت‌های تجارت الکترونیک مانند Amazon، Alibaba و eBay برای جستجوی تصاویر کالاهای موردنظر کاربران از میان کالاهای موجود در پایگاه داده‌های خود از سیستم بازیابی تصویر مبتنی بر محتوا بهره می‌برند. حتی شبکه‌های اجتماعی مانند Pinterest نیز روش‌های مشابهی برای توصیه محتوای مورد علاقه کاربران دارند. به همین صورت بسیاری از نشان‌های تجاری بزرگ از جستجوی بصری در محصولات خود استفاده می‌کنند. (X. Li et al., 2021)

امروزه استفاده از تکنیک‌های بازیابی تصویر مبتنی بر محتوا در سیستم‌های سنجش از راه دور نیز گسترش یافته است. تصاویر سنجش از دور را می‌توان در برخی زمینه‌ها برای حل مسائل مهم مانند پیش بینی آب و هوا، پایش آب و هوا، برنامه ریزی شهری، تجزیه و تحلیل زمین شناسی، پایش بلایا، بررسی منابع و غیره استفاده کرد. (X. Li et al., 2021; Y. Li et al., 2021)

استفاده از سیستم بازیابی تصویر، در بازیابی داده‌های lifelog (که یک ثبت شخصی از زندگی روزمره فرد با جزئیات مختلف است و شامل تصاویر دیجیتالی، اسناد، فعالیت‌ها، بیومتریک و بسیاری از منابع داده دیگر است) می‌تواند برای اهداف مختلف مانند نظارت بر رژیم غذایی، ابزارهای دستیار حافظه، مطالعات اپیدمیولوژیک و تجزیه و تحلیل بازاریابی در نظر گرفته شود. در سال‌های اخیر، برخی از داده‌های lifelog به طور خودکار توسط فناوری پوشیدنی یا دستگاه‌های تلفن همراه قابل جمع‌آوری و نگهداری هستند. در پژوهشی برای بازیابی بخش تصاویر دیجیتالی lifelog ها، از ساختار داده‌ای گراف چشم انداز^۱ شده است. گراف چشم انداز که نخستین بار در سال ۲۰۱۵ برای بازیابی تصویر معرفی شده و از آن زمان به بعد به طور گسترده در تحقیقات به عنوان روشی برای نمایش تصاویر در نظر گرفته می‌شود. (Chang et al., 2021; Nguyen et al., 2021)

¹ scene graph

بازیابی تصویر در سطح نمونه مد^۱ به دلیل افزایش داده‌های تصویری پوشاک در اینترنت، باعث شده است که موضوع بازیابی مد به موضوع موردعلاقه بسیاری در زمینه بینایی کامپیوتری تبدیل شود. با توجه به تکنیک‌های مختلف عکس‌برداری، تصاویر گرفته شده از سوژه‌های مشابه می‌توانند از نظر نور، زاویه دید، پس زمینه و ... متفاوت باشند و این موضوع به چالشی در زمینه بازیابی تصاویر پوشاک تبدیل می‌شود. در دهه گذشته، مجموعه داده‌های زیادی برای سهولت در تحقیقات برای بازیابی تصویر حوزه پوشاک پیشنهاد شده است. از جمله این مجموعه داده‌ها می‌توان به DARN، DeepFashion، WTBI و Modanet اشاره کرد. (X. Li et al., 2021)

شناسایی مجدد شخص^۲، همچنین به عنوان بازیابی فرد^۳ نیز شناخته می‌شود، برای مطابقت با تصاویر یک فرد گرفته شده در نماهای مختلف دوربین است و معمولاً به عنوان یک مشکل فرعی در بازیابی تصویر در نظر گرفته می‌شود. این مساله به طور گسترده در زمینه‌های مختلفی کاربرد دارد از جمله آن می‌توان به مانند امنیت هوشمند، نظارت تصویری هوشمند، بازیابی هوشمند و غیره اشاره نمود. در اینجا نیز چالش‌هایی مانند تغییر زاویه عکس‌برداری و نور و پس زمینه به طور عمومی می‌تواند وجود داشته باشد و همچنین چالش‌هایی به طور خاص در این زمینه وجود دارد مانند: وجود افراد با هویت‌های مختلف و تصاویر مشابه که ممکن است موجب تداخل در تصاویر آنان شود و نیز برای یک فرد، تغییر در وضعیت و شکل و حالت‌های بدن انسان ممکن است مشکل را پیچیده‌تر کند. (X. Li et al., 2021)

امروزه استفاده از تکنیک‌های بازیابی تصویر مبتنی بر محتوا در سیستم‌های سنجش از راه دور^۴ نیز گسترش یافته است. تصاویر سنجش از دور را می‌توان در برخی زمینه‌ها برای حل مسائل مهم مانند پیش بینی آب و هوا، پایش آب و هوا، برنامه ریزی شهری، تجزیه و تحلیل زمین شناسی، پایش بلایا، بررسی منابع و غیره استفاده کرد. (X. Li et al., 2021; Y. Li et al., 2021) همچنین بازیابی تصویر مبتنی بر محتوا در زمینه‌های بازیابی ساختمان و بازیابی تصاویر پزشکی کاربرد دارد. بازیابی ساختمان^۵ از روی نقشه می‌تواند به ما کمک کند مکان یابی دقیق‌تری داشته باشیم و احتمال گم شدن را کاهش دهیم. بازیابی تصویر پزشکی می‌تواند به پزشکان کمک کند تا تشخیص پزشکی را به طور مؤثرتری انجام دهند. (X. Li et al., 2021)

¹ fashion image retrieval

² Re-ID

³ person re-identification

⁴ remote sensing

⁵ Building retrieval

۱-۳. اهداف

به طور کلی جدا کردن اطلاعات مرتبط از اطلاعات نامرتب، جزئی لاینفک از نظام بازیابی اطلاعات می باشد و هدف اصلی سیستم بازیابی تصویر مبتنی بر محتوا نیز همین امر است. (Ahmadi, 2020) هدف ما از انجام این پژوهش آن است که بتوانیم سیستمی را ارائه دهیم که با اضافه شدن تصاویر و بالارفتن حجم و ابعاد داده‌ای، بتواند به صورت مقیاس پذیر مرتبط ترین تصاویر با تصویر ورودی به عنوان پرس و جو را از یک پایگاه داده‌ای با ابعاد بالا، با سرعت بالا و دقت قابل قبول شناسایی کند و به عنوان خروجی‌های مرتب شده بر اساس بیشترین ارتباط با تصویر داده شده بازگرداند. برای حل چالش توزیع شدگی به دنبال آن هستیم که از یک سیستم همه منظوره منبع باز مانند Elasticsearch استفاده کنیم و ببینیم که چطور ویژگی‌ها را روی این سیستم نگاشت کنیم که این چالش‌ها کمرنگ تر شوند.

در واقع هدف اصلی بالابردن سرعت جستجو با حفظ دقت تا حد امکان می باشد. برای حفظ دقت، نیاز به مقایسه دقیق تصویر ورودی با تمامی داده‌های موجود در آرشیو وجود دارد و از طرفی برای بالابردن سرعت، باید بتوانیم با کمترین تعداد مقایسه و مطابقت، نزدیک ترین پاسخ (ها) به درخواست کاربر را پیدا کنیم. (Speed-Accuracy trade-off)

۱-۴. پرسش‌های پژوهشی

گروه اول

به طور کلی چه روش‌هایی برای نمایش تصاویر وجود دارد و چگونه می توان یک روش بهینه را انتخاب نمود؟

رویکردهای مختلفی برای نمایش تصاویر در قالب بردار ویژگی وجود دارد که این روش‌ها از نظر دقت، میزان فضای مورد نیاز برای ذخیره سازی و ابعاد ویژگی استخراج شده با یکدیگر متفاوتند. در این پژوهش به دنبال آن هستیم که یک دسته بندی و مقایسه بین این روش‌ها انجام دهیم (با تعیین یک سری پارامتر) و ببینیم کدام روش‌ها برای چه نوع تصاویری مناسب هستند و دقت این روش‌ها چگونه است؟

آیا امکان ترکیب این روش‌ها باهم وجود دارد و آیا این کار منجر به بهبود دقت می شود یا خیر؟ در نهایت به دنبال این هستیم که چه روش نمایشی تصویر بهینه‌ای را برای این سیستم می خواهیم انتخاب کنیم؟

گروه دوم

اگر بخواهیم فرآیند ایندکس را به صورتی مقیاس‌پذیر انجام دهیم باید از چه روشی استفاده کنیم؟ در واقع به دنبال ارائه یک روش مقیاس‌پذیر جهت انجام فرآیند نمایه‌سازی و جستجو در حجم بالایی از تصاویر هستیم. در این بخش نیز رویکردهای متعددی ارائه شده است. برخی از پژوهش‌ها از سیستم‌های خاص‌منظوره برای این کار استفاده کرده‌اند و نیز برخی دیگر رویکرد توزیع‌شده را در نظر گرفته‌اند. اما برخی روش‌ها از یک سری موتورهای جستجوی همه‌منظوره یا خاص‌منظوره استفاده کردند و آن‌ها را برای این کار شخصی‌سازی کردند.

در این پژوهش ما از یک موتور جستجوی مبتنی بر متن مانند Elasticsearch استفاده می‌کنیم که امکان جستجوی توزیع‌شده را فراهم می‌آورد. اما سؤال اساسی اینجاست که آیا استفاده از موتور جستجوی مبتنی بر متن مانند Elasticsearch برای فرآیند ایندکس مناسب است یا خیر؟ و چگونه سیستم جستجوی بصری را بر روی این موتور جستجو طراحی کنیم که هم مقیاس‌پذیر باشد و هم دقت قابل قبول و سرعت خوبی داشته باشد؟

چطور می‌توان حجم بالایی از اطلاعات را بدون این که دقت را از دست بدهیم، فشرده‌سازی کنیم و ابعاد آن را کاهش دهیم تا بتوان سرعت و میزان فضای ذخیره‌سازی را تا حد امکان بهبود بخشید؟

گروه سوم

از بین مجموعه تصاویر کاندید که به عنوان پاسخ سیستم در نظر گرفته می‌شوند، چگونه مرتبط‌ترین تصویر (تصاویر) را به درخواست کاربر پیدا کنیم و به عنوان جواب بازگردانیم؟ در این بخش به دنبال آن هستیم که بتوانیم معیارهای مشابهت را تعیین کنیم و با استفاده از آن مجموعه‌ای از تصاویر را که به تصویر پرس وجو (کوئری) کاربر نزدیک‌تر است و شباهت بیشتری دارد (بر اساس معیارهای تعیین شده) انتخاب کنیم و این مجموعه تصاویر را به عنوان مجموعه کاندید در نظر بگیریم.

در چنین سیستمی، وقتی برای یک تصویر پرس وجو یک مجموعه کاندید به عنوان پاسخ پیدا می‌کنیم، چگونه می‌توان این مجموعه را به بهترین شکل مرتب‌سازی^۱ و رتبه‌بندی^۲ نمود تا به درخواست کاربر نزدیک‌تر باشد؟

¹ sort

² rank

۵-۱. چالش‌ها

چالش‌های بسیاری در زمینه بازیابی تصویر مبتنی بر محتوا و جستجوی بصری مطرح می‌شود که به عنوان نمونه می‌توان موارد زیر را نام برد:

انتخاب روش مناسب نمایش تصویر در قالب بردار ویژگی

معمولاً روش‌های گوناگونی برای نمایش تصاویر وجود دارد که در فصل دوم به صورت دقیق‌تری به توصیف آن‌ها خواهیم پرداخت. اما به طور خلاصه می‌توان گفت که انتخاب روش مناسب برای نمایش بردار ویژگی استخراج شده از تصویر می‌تواند تأثیر مستقیمی در دقت نتایج به دست آمده از سیستم جستجوی موردنظر داشته باشد. برخی از روش‌ها می‌توانند منجر به دقت بالایی در نتیجه عمل بازیابی شوند اما به دلیل سرعت پایینی که دارند عملاً استفاده از آن‌ها در حجم بالای داده‌ها غیرممکن می‌شود. از سوی دیگر برخی روش‌ها سرعت بیشتری در استخراج و پردازش بردار ویژگی دارند ولی ممکن است دقت کافی برای حصول نتیجه مناسب را نداشته باشند.

از این رو بسیاری از مطالعات همچنان تلاش بر پیدا کردن راهکار بهینه برای نمایش تصاویر دارند زیرا که این مرحله از بازیابی همچنان چالش برانگیز است.

ابعاد بسیار بالا

این چالش در بازیابی اطلاعات به پدیده‌های گوناگونی اشاره دارد که معمولاً هنگام تجزیه و تحلیل و ساماندهی داده‌ها در فضاها با ابعاد بسیار بالا (اغلب در اندازه صدها یا هزاران بعد) روی می‌دهند که اصطلاحاً با عناوینی همچون نفرین ابعاد، مزاحمت ابعاد و یا مشقت بعدچندی از آنها یاد می‌شود. این مشکلات در حوزه‌هایی همچون محاسبات عددی، یادگیری ماشینی، داده کاوی و... ایجاد می‌شوند و گاهی از تکنیک‌هایی همچون کاهش ابعاد و انتخاب ویژگی^۱ برای مقابله با آن‌ها استفاده می‌شود. (2023, "Curse of Dimensionality")

خروجی برخی از روش‌های استخراج ویژگی که امروزه کاربرد بیشتری دارند، داده‌هایی با ابعاد بسیار بالا تولید می‌کنند که این موضوع، پردازش و بازیابی این داده‌ها را دچار مشکل می‌کند.

شکاف مقصود^۲

چالش شکاف مقصود به این مفهوم اشاره دارد که چگونه کاربر سیستم بازیابی تصویر مبتنی بر محتوا می‌تواند برای بیان دقیق‌تر منظور خود از روش مناسبی استفاده کند. به بیان دیگر این که چه

^۱ feature selection

^۲ intention gap

روش‌هایی برای معرفی دقیق‌تر قصد کاربر برای پیدا کردن شبیه‌ترین نتیجه به درک و ذهنیت درونی فرد وجود دارد و چگونه می‌توان بهترین روش را برای این کار برگزید، همچنان از اهداف چالش برانگیز بازیابی تصویر مبتنی بر محتواست. (Zhou et al., 2017)

شکاف معنایی^۱

به عقیده بسیاری از نویسندگان بزرگترین چالش در حوزه بازیابی تصویر مبتنی بر محتوا، شکاف معنایی بین معنای سطح بالای تصویر و ویژگی‌های بصری سطح پایین آن است. به عبارت دیگر شکاف معنایی ناشی از دشواری در توصیف مفهوم معنایی سطح بالا با ویژگی بصری سطح پایین است. برای کاهش این شکاف معنایی، تلاش‌های گسترده‌ای از سوی دانشگاه و صنعت صورت گرفته است. (X. Li et al., 2021; Zhou et al., 2017)

مجموعه داده

یکی دیگر از چالش‌های موجود در حوزه جستجوی بصری، مشکل کمبود داده در این زمینه است. مجموعه داده‌های موجود که در بسیاری از تحقیقات بازیابی تصویر مبتنی بر محتوا دردسترس هستند و مورد استفاده قرار می‌گیرند، بیشتر برای مسائلی همچون خوشه بندی^۲، طبقه بندی^۳، تشخیص اشیاء^۴ و... ایجاد شده‌اند و برای جستجوی بصری کاملاً مناسب نیستند. از طرفی هم مجموعه داده‌هایی که شرکت‌های بزرگ و یا نهاد های دولتی برای سیستم‌های بازیابی خود استفاده می‌کنند معمولاً دردسترس همه قرار نمی‌گیرد مگر آن که همه بتوانند به منافع و مشوق‌های مادی برای به اشتراک گذاری داده‌ها دست پیدا کنند که این خود می‌تواند خطرات دیگری در پی داشته باشد. (Camli, 2020; Y. Li et al., 2021)

ورودی پویا

امروزه سایت‌های فروش محصولات که از اصلی‌ترین مشتریان سیستم‌های بازیابی تصویر مبتنی بر محتوا هستند، با خصوصیات توصیف می‌شوند که خود می‌توانند عملیات بازیابی را با چالش مواجه کنند. این سایت‌ها و همچنین دیگر سایت‌های مشابه معمولاً دارای موجودی پویا با نرخ تغییر و گردش مالی بالا و نیز حجم بالای درخواست و پرس و جو های کاربران هستند که نیاز به انعطاف و مقیاس پذیری سیستم جستجوی بصری را دوچندان می‌کنند. (Degenhardt et al., 2021)

¹ semantic gap

² clustering

³ classification

⁴ object recognition

بهبود تعامل کاربران

اساساً موفقیت هر سیستم در گرو میزان رضایتی است که کاربران از آن سیستم دارند. این که بتوان به رویکردی دست یافت که کاربران بیشتری را برای استفاده از سیستم بازیابی ترغیب کند و نیز تجربه کاربری موجود را بهبود ببخشد، همچنان به عنوان چالشی در حوزه بازیابی تصویر مبتنی بر محتوا در نظر گرفته می‌شود. (Y. Zhang et al., 2018)

معرفی بهترین روش ایندکس و جستجو

به طور کلی هر تصویر به صورت یک بردار ویژگی با ابعاد بالایی نمایش داده می‌شود. یکی از چالش‌های موجود در این زمینه پیدا کردن نحوه مناسب جستجو بین این اطلاعات است که این روش باید بتواند توزیع شده و مقیاس پذیر باشد چون ابعاد این داده‌ای و حجم داده‌ها مدام در حال افزایش است.

۱-۶. ساختار کلی پایان نامه

سازمان دهی این گزارش بدین شرح است:

در این فصل مقدمه‌ای از این پروژه پژوهشی و اهداف و ضرورت‌های کلی پرداختن به موضوع جستجوی بصری در محیط‌های صنعتی و دانشگاهی و زندگی روزمره افراد بیان شد.

در فصل دوم این پایان نامه، به مرور برخی از مهم‌ترین و جدیدترین پژوهش‌ها و مطالعات انجام شده در زمینه بازیابی تصویر مبتنی بر محتوا می‌پردازیم و تلاش می‌کنیم در انتهای فصل، دیدگاهی کلی و مناسب از این حوزه برای دنبال کنندگان این موضوع ترسیم کنیم.

در فصل سوم ایده پیشنهادی برای بهبود این روش را بیان می‌کنیم و همچنین به توضیح نکات لازم برای پیاده سازی ایده های معرفی شده خواهیم پرداخت.

در فصل چهارم این تحقیق مجموعه داده مورد استفاده و معیارهای ارزیابی را معرفی نموده و به بررسی و ارائه نتایج آزمایش‌های انجام شده می‌پردازیم و در فصل پنجم و پایانی به جمع بندی کلی و اهداف پژوهشی آینده می‌پردازیم.

فصل دوم

مروری بر پژوهش‌های پیشین

۲. مروری بر پژوهش های انجام شده

امروزه عملیات هایی همچون جستجو، رتبه بندی و پیشنهاددهی در سیستم های مختلف کاربرد ویژه ای دارند. بسیاری از کسب و کارها برای جستجو و بازیابی تصاویر امروزه از تکنیک ها و روش های متفاوتی نسبت به روش های سنتی استفاده می کنند. در سایت های محصول (مانند eBay، Flipkart، Amazon، و Alibaba)، ویژگی های سنتی رتبه بندی صفحات وب یا وجود ندارند یا نسبت به گذشته تغییرات بسیاری کرده اند. روش ها و تغییراتی که نیازمند آنست که کشف و مطالعه آن ها را بیش از پیش مورد توجه قرار دهیم. (Degenhardt et al., 2021)

در این فصل ابتدا توضیح مختصری درباره ساختار کلی سیستم های بازیابی تصویر مبتنی بر محتوا و بخش ها و مراحل مختلف آن ارائه می کنیم. در ادامه روش های نوین نمایه سازی و جستجو که امروزه بیشتر مورد توجه هستند را مورد مرور و بررسی قرار خواهیم داد. هدف اصلی از ارائه این فصل، دستیابی به تصویری کلی از سیستم بازیابی تصویر مبتنی بر محتوا و بخش های مختلف آن است.

۲-۱. تاریخچه

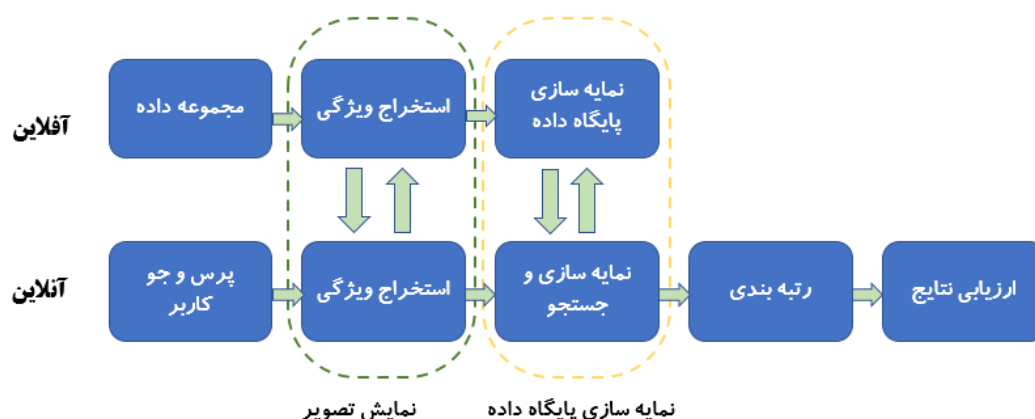
تقریباً از سال ۱۹۹۰ میلادی پژوهش پیرامون عملیات بازیابی تصویر مبتنی بر محتوا به طور گسترده شکل گرفت. در حدود سال ۲۰۰۰ میلادی با معرفی روش های پیشگامی همچون SIFT و BOW روند تحقیقاتی در این زمینه تغییرات مهمی کرد و به سمت بازیابی تصاویر از پایگاه داده های بزرگ متمایل شد. (Zhou et al., 2017)

در سال های اخیر، سیستم های بازیابی تصویر از منابع داده ای بزرگ، به دو شاخه مجزا گرایش پیدا کرده اند. یکی از شاخه ها نمونه هایی همچون FAISS (Johnson et al., 2019) را دربر می گیرند که با بهره گیری از مزایای موازی سازی های مبتنی بر GPU^۱ تلاش در غلبه بر ویژگی های با ابعاد بالا را دارند. در سمت دیگر رویکردهایی قرار دارند که با استفاده از موتورهای جستجوی مبتنی بر متن همچون Elasticsearch و Solr سعی در کاهش هزینه های نمایه سازی و جستجو به روش های قدیمی تر و بالغ تر را دارند. (Stanley et al., 2020)

^۱ Graphics Processing Unit

۲-۲. ساختار کلی سیستم‌های بازیابی تصویر مبتنی بر محتوا

مراحل عملیات بازیابی به دو بخش کلی آنالین و آفلاین تقسیم می‌شوند که در شکل ۱-۲ نمایش داده شده است. بیشتر سیستم‌ها و چارچوب‌هایی که به منظور جستجو و بازیابی تصاویر از پایگاه داده‌های کوچک یا بزرگ ایجاد شده‌اند، شامل همین دو بخش هستند. بخش آفلاین بیشتر به دنبال اهداف نمایش تصویر و سازماندهی مناسب ساختار داده‌ای به دست آمده در پایگاه داده می‌باشد و بخش آنالین معمولاً مسئول دریافت و پردازش پرس و جوهای کاربر و مقایسه آنها با داده‌های موجود و به دست آوردن نتیجه موردنظر کاربر است. (Zhou et al., 2017) در زیرسیستم آفلاین، ویژگی هر تصویر با متد مشخصی استخراج و کدگذاری می‌شود و در پایگاه داده بازیابی به روشی بهینه نمایه سازی و ذخیره می‌شود. در زیر سیستم آنالین، ابتدا پرس و جو کاربر (که می‌تواند یک تصویر، یک طراحی دستی یا موارد دیگر باشد) به سیستم وارد می‌شود و مشابه به روش نمایش در بخش آفلاین، ویژگی‌های آن استخراج و کدگذاری می‌شود. سپس این بردار ویژگی استخراج شده توسط الگوریتم جستجو تعیین شده در سیستم با سایر بردارهای موجود در پایگاه داده مقایسه و امتیازدهی می‌شود و در نهایت مواردی که امتیاز بالاتری (بر اساس یک حد آستانه) کسب کرده‌اند به عنوان لیست کاندید پاسخ‌ها بازگردانده می‌شوند و در قسمت پایانی این پاسخ‌ها مرتب سازی می‌شوند و به عنوان نتایج خروجی سیستم به کاربر فرستاده می‌شوند. (X. Li et al., 2021)



شکل ۱-۲: مراحل کلی عملیات بازیابی تصویر مبتنی بر محتوا

۲ مأموریت (وظیفه) اصلی سیستم بازیابی تصویر مبتنی بر محتوا عبارتست از: (۱) نمایش تصویر (۲) نمایه سازی و جستجو تصویر

در بخش اول این سیستم، عملیات استخراج ویژگی از تصاویر موجود در پایگاه داده انجام می‌شود. در واقع این ماژول، فایل‌های تصویری را از یک مسیر مشخص در سیستم (پایگاه داده) می‌خواند و در ادامه بردار ویژگی مناسب را از آن استخراج می‌کند. (Mezzoudj et al., 2021) تا بتوان تصویر را

قابلی که برای سیستم قابل درک باشد، نمایش داد. سپس این بردارهای ویژگی در جایگاه‌های مشخصی به نام پوشه ویژگی^۱ ذخیره می‌شوند. (Camli, 2020) برای انجام این بخش روش‌های گوناگون بسیاری وجود دارد که یکی از اهداف این پژوهش یافتن روش بهینه در نمایش تصاویر در قالب بردار ویژگی می‌باشد به طوری که این نمایش بهینه بتواند بعدها در عمل جستجو و بازیابی سریع تصاویر مفید باشد.

بخش دیگری از مطالعات در زمینه بازیابی تصویر مبتنی بر محتوا معطوف به بررسی روش بهینه برای نمایه سازی و جستجو از میان مجموعه داده‌های انبوه در پایگاه داده بازیابی تصویر است. نمایه سازی ویژگی^۲ را می‌توان به معنی ساختاردهی پایگاه داده برای بهبود بخشیدن به سرعت جستجو دانست. (Sadeghi-Tehran et al., 2019) امروزه به دلیل استفاده از روش‌های مبتنی بر یادگیری عمیق برای بالا بردن دقت در استخراج ویژگی، بردارهای به دست آمده برای نمایش تصویر دارای ابعاد بالایی هستند که این امر مقایسه با بردار تصویر پرس و جو را بسیار سخت می‌کند. درحالی که مقایسه بردار ویژگی پرس و جو با کل مجموعه داده تصویر ممکن است برای مجموعه داده‌های کوچک امکان پذیر باشد، اما این هنوز یک عملیات خطی از مرتبه $O(N)$ است. بنابراین، برای مجموعه داده‌های مقیاس بزرگ و زمانی که تعداد این بردارهای تصویر از میلیاردها بردار عبور می‌کند، این عمل از نظر محاسباتی کارآمد نیست. (Sadeghi-Tehran et al., 2019) در اینجا یک ساختار شاخص گذاری پایگاه داده مناسب می‌تواند تا حدود زیادی از مدت زمان لازم برای جستجو بکاهد و نیز از مقایسه‌های غیر ضروری با بردارهای غیر مشابه جلوگیری کند. در واقع نمایه سازی یا ایندکس یک مرحله میانی بین طیف مراحل جستجو بصری است که در برخی منابع آن را در مازول نمایش تصویر و در برخی منابع آن را در مازول جستجوی شباهت قرار می‌دهند اما امروزه به دلیل اهمیت آن در کاهش زمان پاسخ به خصوص در زمانی که حجم داده‌های قابل بررسی خیلی زیاد است، بسیاری از تحقیقات به طور خاص متمرکز بر ایندکس پایگاه داده هستند و حتی مازول جداگانه‌ای برای این کار در نظر می‌گیرند.

پس از مرحله نمایه سازی پایگاه داده نوبت به مرحله جستجو می‌رسد که کاربر باید پرس و جو خود را به عنوان ورودی به سیستم بدهد و تلاش بر آن است تا با حداقل مقایسه بتوان شبیه‌ترین تصاویر به درخواست کاربر را از میان انبوهی از داده‌های موجود در پایگاه داده، پیدا نمود. در انتها نتایج بازیابی شده بر اساس امتیاز مشخصی رتبه بندی می‌شوند و به عنوان خروجی سیستم برای کاربر ارسال می‌شوند. این نتایج را می‌توان بر اساس معیارهای مختلف دقت، سرعت و ... ارزیابی نمود.

¹ features folder

² feature indexing

۲-۳. نمایش تصویر

بازنمایی تصویر یکی از وظایف کلیدی سیستم جستجوی بصری است که اجرای صحیح آن، هدف اساسی سیستم یعنی مقایسه تصاویر را محقق می‌کند و در کیفیت نتایج حاصل از آن نقش مستقیم دارد. همانطور که در قسمت قبل هم ذکر شد، بخشی از این عملیات می‌تواند به صورت آفلاین و بخشی از آن به صورت آنلاین صورت پذیرد.

هنگامی که اطلاعات ورودی به یک الگوریتم برای عملیات بازیابی، بسیار بزرگ باشد، می‌توان آن را به مجموعه کاهش یافته‌ای از ویژگی‌های مهم به نام "بردار ویژگی" تبدیل کرد. به این عمل استخراج ویژگی گفته می‌شود. در برخی از سیستم‌ها پیش از عمل استخراج ویژگی، رویکردی موسوم به "پیش پردازش" وجود دارد که هدف آن بهبود ظاهری و کیفیت تصاویر برای پردازش محاسباتی دقیق‌تر است. (Ghosh et al., 2018)

برای انجام این بخش روش‌های گوناگون و بسیاری وجود دارد که می‌توان آن‌ها را در ۲ دسته‌بندی کلی قرار داد:

۱) روش‌های مرسوم مبتنی بر ویژگی^۱

این دسته اشاره به روش‌هایی دارند که در آن‌ها از هیچ رویکرد یادگیری عمیق استفاده نشده است و بیشتر با روش‌های اکتشافی^۲ طراحی شده‌اند. روش‌های مرسوم مبتنی بر ویژگی می‌توانند در دو دسته ویژگی‌های محلی^۳ و سراسری^۴ قرار گیرند.

روش‌های سراسری بیشتر شامل روش‌هایی هستند که از رنگ، بافت، شکل و سایر ویژگی‌های ساختاری خود تصویر برای استخراج استفاده می‌کنند که تاحدودی باعث بهبود عملکرد حاشیه نویسی تصویر می‌شوند و یا در برخی مسائل برای تشخیص نمونه‌های تکراری می‌توانند مفید باشند اما از سمت دیگر ممکن است برخی از درهم ریختگی‌های پس زمینه را شامل شوند.

روش‌های محلی دسته دیگری از روش‌های مبتنی بر ویژگی هستند که از ویژگی‌های هدفمندتری استفاده می‌کنند. یکی از معروفترین الگوریتم‌های موجود در این دسته، الگوریتم SIFT است که شامل ۲ مرحله اساسی تشخیص نقاط موردعلاقه^۵ و توصیف منطقه محلی^۶ است. در سال‌های اخیر بسیاری از روش‌ها را می‌توان الهام گرفته از الگوریتم SIFT دانست. مقایسه بین روش‌های محلی

^۱ Conventional feature-based methods

^۲ heuristically designed

^۳ local

^۴ global

^۵ interest point detection

^۶ local region description

را می توان از سه جنبه انجام داد: سرعت، فشردگی و تمایز. از آن جایی که توصیف کننده های محلی بسیاری با انواع گوناگون از برتری در سال های اخیر توسعه یافته اند، انتخاب روش بهینه برای نمایش و عملیات بازیابی تصویر می تواند یکی از چالش ها در این زمینه باشد.

اگرچه مطالعات پیرامون رویکردهای جدیدتر برای استخراج ویژگی به روش های مرسوم مبتنی بر ویژگی همچنان ادامه دارد اما امروزه اکثر این روش ها به خاطر محدودیت قدرت نمایشی تصویر، کمتر مورد استفاده قرار می گیرند. (X. Li et al., 2021)

۲) روش های مبتنی بر شبکه های CNN و شبکه های عمیق

در پژوهش ها نشان داده شده که ویژگی هایی که از طریق شبکه های عمیق حاصل می شوند، به طور کلی بهتر از سایر روش های استخراج ویژگی عمل می کنند و تجربیات جستجو را بهبود می دهند. (Camli, 2020) برای به دست آوردن یک بازنمایی از تصویر، می توان از انواع گوناگونی از مدل های CNN استفاده نمود: همچون مدل هایی که از پیش آموزش دیده اند (pre-trained) و یا مدل هایی که برای داده های بزرگتر و جدید تنظیم و بهینه سازی شدند (fine-tuned) و یا مدل هایی که از ابتدا برای اهداف بازیابی ساخته و آموزش داده می شوند (built/trained from scratch)

مدل های pre-trained

مدل های از پیش آموزش دیده در بسیاری از پژوهش های در زمینه بازیابی تصویر مبتنی بر محتوا با هدف برقراری تعادل بین پیچیدگی و کارایی استفاده می شوند. یک مدل آموزش دیده از قبل، پس از پردازش تعداد زیادی تصویر در یک مجموعه داده (معمولاً مجموعه داده ImageNet) بدون نیاز به مرحله پرهزینه محاسباتی آموزش می تواند به توصیفگرهای بصری دست یابد. (Sadeghi-Tehran et al., 2019) امروزه مدل های از پیش آموزش دیده بسیاری برای اهداف مختلف به کار گرفته می شوند ولی شناخت صحیح کاربردهای هر مدل و جایگاه مناسب استفاده از آن از اهمیت بالاتری برخوردار است. بنابراین بخشی از پژوهش ها بر ارائه تحلیل و مقایسه های کاربردی از مدل های از پیش آموزش دیده (Canziani et al., 2016) و نیز نقش استفاده ترکیبی از چندین مدل و نتایج آن در جستجوی بصری (Camli, 2020)، تمرکز دارند.

مدل های fine-tuned

کلمه fine-tuned به معنی بهینه سازی دقیق یا تنظیم دقیق است. این اصطلاح در حوزه هوش مصنوعی به کار می رود تا به تنظیم و بهبود دقت و عملکرد الگوریتم های ماشین یادگیری اشاره کند. به طور کلی مدل fine-tuned به مدلی گفته می شود که با استفاده از داده های بیشتر بهبود یافته و بهینه سازی شده است.

مدل‌های built/trained from scratch

در حالی که بسیاری از محققان مشغول مطالعه بر روی ویژگی‌های استخراج شده از شبکه‌های عمیق از پیش آموزش داده شده برای کار طبقه‌بندی بودند، بعداً نشان داده شد که شبکه‌های عمیق را می‌توان مستقیماً برای کار بازیابی نمونه به روشی انتها به انتها آموزش داد. در این روش‌ها معمولاً از یادگیری متریک عمیق^۱ استفاده می‌شود که هدف آن یادگیری فضایی ست که در آن بردارهای مشابه تشویق می‌شوند تا بهم نزدیک‌تر باشند و بردارهای غیر مشابه از هم دور می‌شوند. نکته کلیدی در بسیاری از این مدل‌ها، تعریف یک تابع زیان است که رتبه‌بندی را به جای طبقه‌بندی بهینه می‌کند و دارای یک تکنیک استخراج سخت منفی است که کیفیت فضای برداری آموخته شده را بهبود می‌بخشد. (X. Li et al., 2021)

در سال‌های اخیر انواع دیگری از شبکه‌های عصبی کانولوشن تحت عنوان مدل‌های Siamese Network و مدل‌های Triplet Network محبوبیت گسترده‌ای یافته‌اند که هدف آنها یافتن تصاویر مشابه به وسیله بهینه سازی فضای ویژگی بین جفت‌های تصویری است. نگاشت تصاویر مشابه نزدیک به یکدیگر و همچنین نگاشت تصاویر غیرمشابه به دور از یکدیگر، وظیفه اصلی این مدل‌هاست. مهم‌ترین تفاوت این دو مدل در نوع ورودی به شبکه است که در شبکه‌های Siamese یک دوتایی تصویری برای مشابهت یابی به شبکه وارد می‌شوند که این دو می‌توانند مثبت (مشابه) یا منفی (غیرمشابه) باشند؛ در حالی که در شبکه‌های Triplet یک سه تایی تصویر برای یادگیری به شبکه وارد می‌شوند که عبارتند از یک تصویر پرس و جو به نام anchor، یک نمونه تصویر مشابه به نام مثبت و یک نمونه تصویر غیر مشابه به نام منفی و در پایان مرحله آموزشی هر شاخه CNN یک جاسازی^۲ برای نمایش هر تصویر خروجی می‌دهد. (Camli, 2020)

همچنین اخیراً روش‌هایی نیز برای استخراج مستقیم کدهای باینری با استفاده تلفیق روش‌های ایجاد هش باینری و شبکه‌های عصبی عمیق پیشنهاد شده‌اند که می‌توان آن را همان مرز مشترک بین مراحل استخراج ویژگی توسط شبکه‌ها عمیق و نمایه سازی بردارهای ویژگی تصویر در نظر گرفت. (Y. Li et al., 2019; Liu et al., 2016; W. Zhang et al., 2021)

^۱ Deep metric learning

^۲ embedding

۴-۲. نمایه سازی

همانطور که پیش تر نیز بیان شد، محاسبه شباهت بین تصاویر می تواند به عنوان اصلی ترین وظیفه سیستم بازیابی تصویر مبتنی بر محتوا در نظر گرفته شود. بنابراین باید روشی برای محاسبه شباهت در نظر گرفته شود که به قدر کافی متمایزکننده، قوی و کارآمد باشد. (Dubey, 2021)

یکی از چالش هایی که در این زمینه وجود دارد این است که معمولاً بردارهای استخراج شده از روش های استخراج ویژگی، دارای ابعاد بسیار بالایی هستند که این باعث می شود جستجو و یافتن مطابقت برای تصویر پرس و جو در بین این داده ها تبدیل به امری دشوار شود. این مساله بخصوص زمانی اهمیت بیشتری پیدا می کند که تعداد تصاویر موجود در پایگاه داده یا به عبارت دیگر حجم داده ها به مرور افزایش یابد که سبب بالا رفتن حجم محاسبات و نیاز به فضای ذخیره سازی بیشتر می شود. یک از راه حل هایی که برای بهبود این مساله پیشنهاد شده است، نمایه سازی یا ایندکس است که بیشتر به ساختار بندی و نوع ذخیره سازی در پایگاه داده یعنی ساختاردهی بردارهای ویژگی استخراج شده از مرحله نمایش تصویر، ارتباط پیدا می کند. (Y. Li et al., 2021) در برخی منابع این مرحله را نیز در ماژول بخش اول یعنی نمایش ویژگی و بلافاصله پس از استخراج بردار ویژگی در نظر می گیرند (Hussain & Surendran, 2021; Mezzoudj et al., 2021) و بعضی دیگر از منابع پژوهشی نیز فرآیندهای پردازشی این مرحله را به مرحله جستجو شباهت و بازیابی نسبت می دهند. (X. Li et al., 2021) اما برخی منابع به دلیل اهمیت آن، یک بخش جداگانه را در سیستم به عملیات ایندکس اختصاص می دهند. (Camli, 2020; Y. Li et al., 2021) به دلیل این که عملیات ایندکس باعث افزایش سرعت جستجو می شود و همچنین به دلیل اهمیت زمان پاسخ، سبب می شود که این موضوع جایگاه ویژه ای در سیستم های جستجوی بصری به خصوص در یک پایگاه داده تصویری در مقیاس بسیار بزرگ پیدا کند. (Y. Li et al., 2021) البته بدیهی است که بین سرعت و دقت رابطه معکوس وجود دارد. از آن جایی که بردارهای ویژگی دارای ابعاد بالایی هستند، یکی از روش های موجود استفاده از الگوریتم ها و تکنیک های کاهش ابعاد در یادگیری ماشین می باشد. (Camli, 2020)

روش های ایندکس پایگاه داده را که از رویکردهای حریصانه استفاده می کنند می توان در ۳ دسته

قرار داد: (Y. Li et al., 2021)

۱) روش های مبتنی بر درخت

این روش ها به صورت بازگشتی فضای ویژگی با ابعاد بالا را به زیرفضاهای با ابعاد کمتر تقسیم می کنند تا در نهایت یک ساختار درختی را از داده های موجود در پایگاه داده شکل بدهند. در فاز بازیابی در این روش ها معمولاً از تکنیک های تحدید و انشعاب استفاده می شود و جستجو معمولاً به

صورت جستجو تقریبی نزدیکترین همسایگان است. این گروه از روش‌ها معمولاً دچار مشکل محدودیت‌های حافظه هستند و در کاربردهایی که پایگاه داده با ابعاد بالا وجود دارد، چندان کارآمد نیستند. البته استفاده از روش‌های کاهش ابعاد می‌تواند تا حدودی در این زمینه راه‌گشا باشد (Halavataya, 2020; Y. Li et al., 2021)

۲) روش‌های مبتنی بر خوشه بندی

روش‌های مبتنی بر خوشه‌بندی، بردارهای ویژگی در قالب خوشه‌ها (یعنی کلمات بصری) جمع‌آوری می‌کند که در آن همه کلمات بصری یک کتاب رمز بصری را تشکیل می‌دهند. در این دسته می‌توان از تکنیک‌های شاخص معکوس فایل^۱ برای ذخیره‌سازی داده‌ها استفاده نمود. مزیت این روش‌ها اینست که سرعت جستجو را تا حد خوبی بهبود می‌بخشند اما چون خوشه‌ها ثابت هستند، اضافه نمودن تصاویر جدید ممکن است منجر به ایجاد مشکلاتی در این روش‌ها شود. همچنین در فضاهای با ابعاد بالا، داده‌ها به صورت پراکنده هستند و این موضوع سبب می‌شود که دقت معیارهای ارزیابی فاصله کاهش یابد. (Y. Li et al., 2021)

۳) روش‌های مبتنی بر درهم‌سازی یا هش

یکی از روش‌های جستجوی پایگاه داده که به طور گسترده مورد استفاده قرار می‌گیرد، ایندکس مبتنی بر هش است که یک تصویر را به مجموعه‌ای از کدهای فشرده هش یا کدهای باینری تبدیل می‌کند، به طوری که جستجو را می‌توان به صورت مقایسه فواصل همینگ در بین کدهای هش در نظر گرفت. از مزایای این روش می‌توان به کاهش پیچیدگی محاسباتی و هزینه ذخیره سازی اشاره نمود. (X. Li et al., 2021) استفاده از روش‌های مبتنی بر هش به همراه روش‌های جستجوی تقریبی می‌تواند به طور ویژه‌ای به بالا رفتن سرعت جستجو و کاهش چشمگیر فضای ذخیره سازی کمک کند. روش‌های هش در ابتدا توابع هش را ایجاد می‌کنند تا برای هر تصویر در بایگانی اعمال شود تا کد هش باینری تصویر به دست بیاید. سپس، یک جدول هش تولید می‌شود، که در آن تصاویر مشابه دارای کد هش یکسان یا مشابه در یک سطل هش قرار می‌گیرند. (Y. Li et al., 2021) متدهای مبتنی بر هش را می‌توان به دو دسته مستقل از داده و وابسته به داده تقسیم کرد که دو زیردسته روش‌های هش با ناظر و هش بدون ناظر نیز در گروه روش‌های هش وابسته به داده قرار داده می‌شوند. (X. Li et al., 2021)

امروزه استفاده از روش‌های مبتنی بر کدهای هش و باینری با هدف بالا بردن سرعت و کاهش مصرف حافظه (Kulis & Grauman, 2009; Mu et al., 2018) به همراه شبکه‌های یادگیری

¹ Inverted File Index

عمیق مورد توجه بسیاری از محققان قرار گرفته است و بخش مهمی از پژوهش های این حوزه را شامل می شود. (Cheng et al., 2019; Y. Li et al., 2019; W. Zhang et al., 2021)

در سال های اخیر همچنین بخشی از پژوهش ها معطوف به روش های ایجاد کدهای فشرده به ازای بردارهای ویژگی با ابعاد بالا از تصاویر در مرحله نمایه سازی، قرار گرفته اند. این روش ها ابتدا یک بردار ورودی را به یک کد کوتاه تبدیل می کنند، جستجو با استفاده از کدهای کوتاه حاصل انجام می شود. این کدهای فشرده در فرآیند بازیابی جایگزین بردارهای ویژگی می شوند و گاهی حتی معیارهای شباهت به خصوصی را برای وظیفه جستجوی شباهت و ارزیابی معرفی می کنند. به صورت کلی این روش های مبتنی بر کد کوتاه فشرده^۱ به روش های مبتنی بر باینری و مبتنی بر PQ^۲ دسته بندی می شوند که این دو روش مکمل هم هستند. روش های مبتنی بر باینری محاسبات فاصله سریع تری را با استفاده از اندازه گیری فاصله همینگ ارائه می دهند، در حالی که روش های مبتنی بر PQ به نتایج دقیق تری با توجه محدودیت حافظه داده شده، دست می یابند. به طور خلاصه روش های مبتنی بر کدهای PQ ابتدا بردارهای ویژگی با ابعاد بالا را به چندین قسمت با ابعاد کمتر تقسیم می کنند و با کمک یکی از روش های خوشه بندی و یک تابع رمزگذاری (که معمولاً قابلیت رمزگشایی یا بازسازی بردار اصلی از روی کد PQ را دارد)، به هر قسمت یک sub-codeword اختصاص می دهند و در نهایت با کنارهم قرار دادن sub-codeword قسمت های مختلف به یک نمایش کد کوتاه PQ به ازای هر بردار ورودی دست پیدا می کنند. همچنین معمولاً در روش های مبتنی بر Product Quantization برای انجام جستجوی میلیونی یا حتی میلیاردی، از یک سیستم جستجو با نمایه سازی معکوس استفاده می شود. (Matsui et al., 2018)

گروهی دیگر از پژوهشگران نیز برای استفاده بهینه از منابع و بالابردن سرعت در سیستم بازیابی تصویر مبتنی بر محتوا، از برخی رویکردهایی استفاده می کنند که به بررسی تبدیل ویژگی های شبکه عصبی با ابعاد بالا به فرم های متنی مناسب برای نمایه سازی توسط یک موتور بازیابی کامل متن استاندارد مانند Elasticsearch می پردازند. (Amato et al., 2018, 2020; Mu et al., 2018)

۲-۵. جستجوی شباهت

در بخش جستجوی شباهت و بازیابی، تمرکز بر پیدا کردن نزدیک ترین تصویر به یک تصویر خاص که به عنوان تصویر پرس و جو یا کوئری به سیستم داده می شود، است. با نمایش مؤثر تصویر در قالب بردار ویژگی و استفاده از روش مناسب ایندکس در پایگاه داده می توان شاخص های تصویر را

^۱ short-codes

^۲ Product Quantization

از طریق اندازه‌گیری شباهت با بردار تصویر پرس و جو جستجو کنیم. این فرآیند به عنوان جستجوی پایگاه داده نامیده می‌شود. (X. Li et al., 2021) به طور خلاصه در این مرحله بردار ویژگی یک تصویر پرس و جو محاسبه شده و با بردارهای ویژگی تصاویر پایگاه داده مقایسه می‌شود. در نتیجه، سیستم CBIR نزدیک‌ترین تصاویر را به کاربر برمی‌گرداند. (Mezzoudj et al., 2021) برای یک بردار ویژگی پرس و جو، جستجوی جامع ویژگی در N بردار ویژگی موجود در پایگاه داده شامل محاسبات $O(N)$ می‌شود. روش‌های مرسوم از یک متریک شباهت برای مقایسه بردار ویژگی تصویر پرس و جو با هر بردار ویژگی در پایگاه داده استفاده می‌کنند. به عبارت دیگر در حالی که مقایسه بردار ویژگی پرس و جو با کل مجموعه داده تصویر ممکن است برای یک مجموعه داده کوچک امکان پذیر باشد، اما برای مجموعه داده‌های مقیاس بزرگ با میلیاردها تصویر، جستجوی ویژگی‌های جامع غیرعملی می‌شود. در اینجا استفاده از تکنیک‌های جستجوی تقریبی اهمیت بیشتری پیدا می‌کند (Y. Li et al., 2021) و انتخاب روش مناسب در این باب همچنان چالش برانگیز است.

برخی از پژوهش‌ها بر روی کدهای فشرده کوتاه شده، ایده‌هایی را در مرحله جستجو ارائه می‌دهند که بتوانند هزینه جستجو را تا حد امکان پایین بیاورند. برای مثال در پژوهشی تلاش دارند تا با فیلتر کردن تعدادی از کدهای باینری که با در نظر گرفتن ویژگی‌های ذاتی این کدها در اندازه‌گیری فاصله همینگ به انجام می‌رسد، از جستجوی کامل و با مصرف حافظه کمتر در مقابل روش‌های دیگر استفاده کنند. (Mu, Yang, et al., 2019; Mu, Zhao, et al., 2019)

همچنین بخشی از پژوهش‌ها در مورد کدهای فشرده و کوتاه شده PQ، از محاسبه فاصله نامتقارن^۱ (ADC) برای اندازه‌گیری و معیار شباهت بین تصاویر استفاده می‌کنند که معمولاً در آنها از نمایه سازی معکوس برای کاهش فضای جستجو کمک گرفته می‌شود. (Matsui et al., 2018)

۲-۶. مرور کارهای مرتبط

در این قسمت به صورت خلاصه بخشی از آخرین ایده‌های موجود برای کل فرآیند جستجوی بصری در قالب جدول ۱-۲: خلاصه روش‌های پیشین جمع آوری و تدوین شده‌اند.

¹ Asymmetric Distance Computation (ADC)

جدول ۲-۱: خلاصه روش های پیشین

مقاله	ایده کلی	نمایش تصویر	ابعاد	نمایه سازی	جستجو	معیار شباهت
(Mu, Zhao, et al., 2019)	فیلتر کردن کدهای باینری غیر ضروری	SSDH model	128 and 256 (binary)	1. Bit Operation 2. Sub-Code Filtering 3. Data Preprocessing with Permutation	کامل	فاصله همینگ
(Camli, 2020)	ترکیب ویژگی ها جهت ایجاد بهترین نمایش برای تصویر	different (low level features + CNN-based fatures)	different (12, 300, 512, 1024, 2048, ...)	کاهش ابعاد (PCA, UMAP, Random Projection, LDA and NCA)	کامل	فاصله اقلیدسی
(Mu et al., 2018)	تبدیل هر بردار به نمایش متنی	Inception-ResNet-V2	1536	1) Element-wise Rounding 2) Subvector-wise Clustering	کامل	Elasticsearch
(Zhao et al., 2019)	استفاده از گراف توزیع شده برای نمایه سازی و جستجو روی کدهای باینری + Map-Reduce	GoogLeNet V1 Inception V1 (binary)	512	1- Clustering using k-means 2- K-nn Graph	جستجو در میان زیرگراف ها و سپس ادغام نتایج	فاصله همینگ

مقاله	ایده کلی	نمایش تصویر	ابعاد	نمایه سازی	جستجو	معیار شباهت
(Gennaro et al., 2010)	تبدیل ویژگی‌ها به نمایش متنی	low level features (such as colors and textures)	50	1-Assume references 2-Ranking references according to distance 3-Converting to text 4-Using Lucene	تقریبی	فاصله Spearman Rho
(Amato et al., 2016)	تولید شناسه برای نمایه سازی مبتنی بر جایگشت	HybridNet	4096	generate sequence of identifiers	کامل	فاصله Spearman Rho
(Amato et al., 2018)	تبدیل ویژگی‌ها به نمایش متنی	ResNet-101 (R-MAC feature)	4096	Permutation-base indexing + transform into surrogate text	تقریبی	Elasticsearch
(Dawn et al., 2020)	استفاده از موجک Haar برای تجزیه تصویر	low level image features (such as color histogram)	1440	using Haar Wavelet for image decomposition	کامل	فاصله chi-squared
(Matsui et al., 2018)	نمایش ویژگی‌ها به PQ عنوان کد	input base High-dimensional vector	-	encoding to PQ code	تقریبی	محاسبه فاصله نامتقارن (ADC)

مقاله	ایده کلی	نمایش تصویر	ابعاد	نمایه سازی	جستجو	معیار شباهت
(Wan et al., 2017)	با استفاده از کدهای باینری و PQ	input base High-dimensional vector	-	SH(spectral hashing) with Multi-index Hashing & PQ-codes using inverted file system (IVF)	غیر کامل (تقریبی)	فاصله همینگ + محاسبه فاصله نامتقارن (ADC)
(Sagharichian & Mirmarouf, 2021)	تبدیل ویژگی ها به زیر کدهای فشرده	VGG16	4096	- convert to Binary code (using LSH) - create subcode - index in Elasticsearch	تقریبی	Elasticsearch
(Kalyankar et al., 2022)	ترکیب ویژگی های سطح پایین و تشخیص خودکار ویژگی های انسانی	1) Image preprocessing 2) Attribute detection 3) Patch-level LBP feature computation 4) Attribute-enhanced sparse coding 5) Regression technique	1600	کاهش ابعاد (1600 => 9)	تقریبی	فاصله همینگ
(Moghadam Charkari & Shakibian, 2021)	استفاده از شبکه های پیچیده چندلایه برای یادگیری ویژگی	low level image features (such as color histogram and dominant color)	different	ساخت شبکه پیچیده پویا چندلایه	تقریبی	پیدا کردن meta-paths

۷-۲. جمع بندی

در این فصل با ساختار کلی یک سیستم جستجوی بصری به صورت کامل آشنا شدیم و همچنین به مرور آخرین روش‌های ارائه شده در زمینه بازیابی تصویر مبتنی بر محتوا پرداختیم. در این زمینه تحقیقات زیادی انجام شده است که تعدادی از این روش‌ها بیشتر متمرکز بر قسمت نمایش تصویر هستند اما بیشتر پژوهش‌ها تلاش بر ارائه یک سیستم end-to-end دارند که بتوانند تمام مراحل نمایش، ذخیره، جستجو و بازیابی را تا حد امکان بهبود ببخشند و پیشرفت در این حوزه را رقم بزنند.

فصل سوم

روش پیشنهادی

۳. روش پیشنهادی

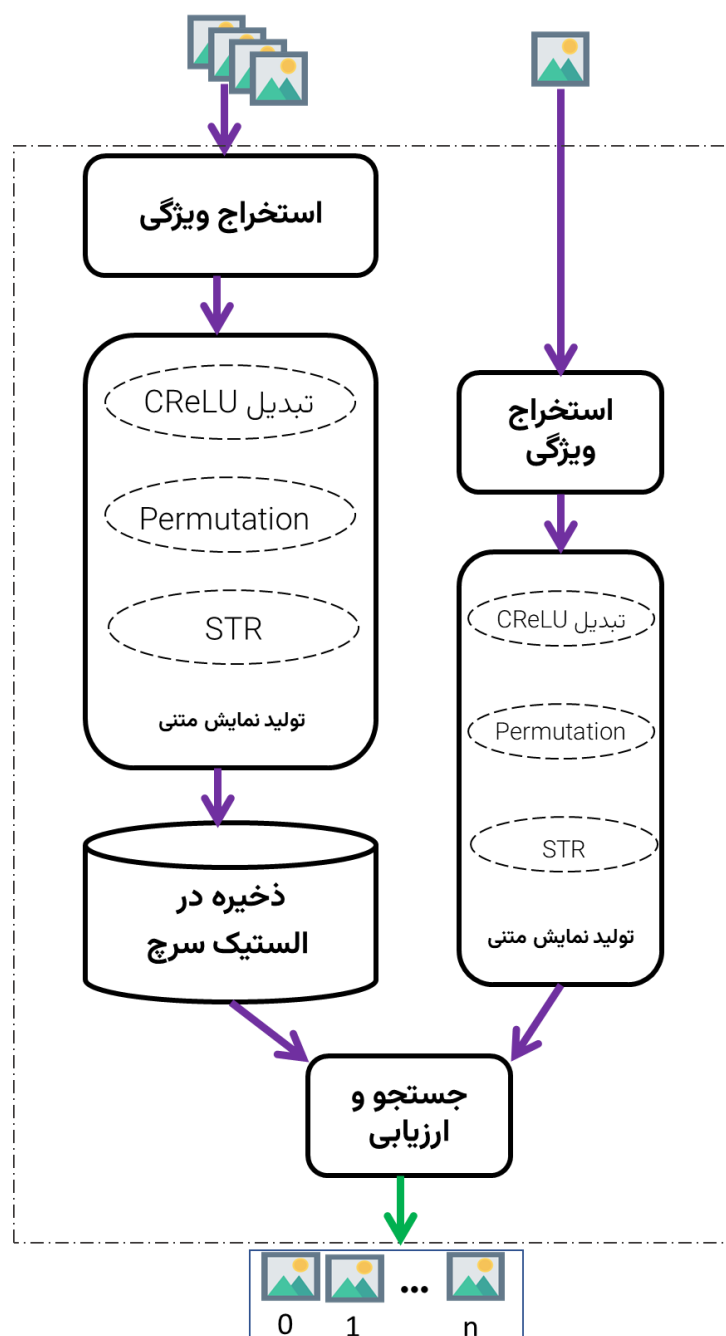
در فصل‌های قبلی با ساختار کلی روش‌های بازیابی مبتنی بر محتوا آشنا شدیم و آخرین تکنیک‌های موجود را در این زمینه مرور کردیم. اگرچه هرکدام از این روش‌های شامل نقاط قوت و ضعف به خصوصی بودند اما تلاش برای یافتن رویکرد بهینه در محیط‌های آموزشی و صنعتی همچنان ادامه دارد. هدف اصلی در این فصل ارائه رویکردی بهینه با ایجاد تعادلی قابل قبول بین دقت و سرعت با بهره‌گیری از نقاط قوت در الگوریتم‌های پیشین می‌باشد.

۳-۱. مقدمه

همانطور که پیش‌تر نیز بیان شد، در سال‌های اخیر با پیشرفت ابزارهای فناوری اطلاعات حجم قابل توجهی از تصاویر تولید و برای بهره‌برداری در دسترس قرار گرفته است. از این رو یافتن اطلاعات تصویری دلخواه از میان حجم زیادی از داده‌های موجود، به چالشی دشوار در میان فناوری‌های مرتبط با علوم کامپیوتر بدل شده است. برای رفع این چالش، روش‌های مختلفی ارائه شده‌اند که یکی از روش‌های پرکاربرد در این زمینه، بازیابی تصویر مبتنی بر محتوا است. بازیابی تصویر مبتنی بر محتوا، با استفاده از ویژگی‌های مختلفی که با استفاده از شبکه‌های عصبی از تصویر استخراج می‌شود، اقدام به یافتن تصاویر مشابه و یا دقیقاً همانند تصویر اولیه در پایگاه داده‌هایی با ابعاد بزرگ می‌کند. در این روش ابتدا تصاویر با بهره‌گیری از فرآیند استخراج ویژگی توسط شبکه‌های عمیق به بردارهای ویژگی (عموماً با ابعاد بالا) تبدیل می‌شوند که این عمل نگهداری و پردازش آن در کامپیوتر میسر می‌نماید. پس از آن، مجموعه از پردازش‌ها در فرآیند ایندکس یا نمایه‌سازی، بر روی بردارهای ویژگی اعمال می‌شوند تا به واسطه ساختار بندی صحیح در حافظه و بهره‌مندی مناسب از امکانات موجود، بتوان سرعت جستجو را تا حد قابل قبولی بهبود بخشید. در نهایت عمل جستجو با استفاده از موتور جستجوی مبتنی بر متن الستیک سرچ انجام می‌شود تا بتوان شبیه‌ترین تصاویر را به تصویر درخواستی کاربر پیدا نمود.

۲-۳. پیش زمینه

در این پژوهش، یکی از روش‌های نوین در زمینه بازیابی تصویر مبتنی بر محتوا را مورد تحلیل و بررسی قرار می‌دهیم و با پیشنهاد تغییراتی در الگوریتم بخش نمایه سازی تلاش می‌کنیم تا این روش را به رویکردی بهینه در این زمینه تبدیل نماییم. در شکل ۱-۳ می‌توانید بلوک دیاگرام کلی روش پیشنهادی را که برگرفته از مقاله (Amato et al., 2018) می‌باشد، مشاهده نمایید.



شکل ۱-۳: چارچوب کلی روش پیشنهادی

در ادامه ابتدا به معرفی برخی از ابزارها و فناوری‌های مورد استفاده در پیاده سازی این روش می‌پردازیم و پس از آن جزییات روش پیشنهادی را مورد بررسی قرار می‌دهیم.

معرفی الاستیک سرچ

الستیک سرچ (Elasticsearch) یک سامانه متن باز و توزیع شده برای جستجو در محتوای متنی و غیرمتنی است که قابلیت انجام عملیات جستجوی پیشرفته را بر روی داده‌های ساختار یافته و نیمه ساختار یافته دارد. این موتور جستجو و تحلیل و ذخیره سازی داده در اصل یک سیستم مبتنی بر Apache Lucene بوده که با زبان برنامه نویسی جاوا پیاده سازی شده است و قابلیت‌هایی همچون جستجوی متنی پیشرفته، تحلیلگرهای نحوی و ... را فراهم می‌کند. با وجود اینکه الاستیک سرچ با زبان برنامه نویسی جاوا نوشته شده است اما حاوی API های گوناگونی است که امکان استفاده آسانتر از آن را برای کاربران مختلف به زبان‌های دیگر فراهم می‌آورد.

الستیک سرچ به عنوان یک سیستم جستجو و تحلیل داده‌های پیچیده، در بسیاری از شرکت‌های مشهور مورد استفاده قرار می‌گیرد. برای مثال، شرکت‌های بزرگی مانند Walmart و Netflix از الاستیک سرچ برای جستجوی داده‌های مورد نیاز خود استفاده می‌کنند.

این فناوری به کاربران اجازه می‌دهد تا داده‌های مختلف را ذخیره کرده و به سرعت و با دقت به آن‌ها دسترسی پیدا کنند و در میان آن‌ها به جستجو بپردازند. برای مثال، می‌توان داده‌های لاگ‌های سرور، داده‌های جستجو در سایت‌های وب و حتی داده‌های حسابداری را در الاستیک سرچ ذخیره کرد و برای جستجو و تحلیل آن‌ها از آن استفاده نمود.

از مزایای دیگر آن می‌توان به سرعت و قابلیت پذیرش بار بالا، راحتی در استفاده، پروتکل‌های متعدد جستجو، ابزارهای تحلیل و داشبوردهای قدرتمند، پشتیبانی از معماری توزیع شده و ... اشاره کرد. الاستیک سرچ در حال حاضر یکی از ابزارهای بسیار محبوب در مجموعه ابزارهای متن باز برای جستجو و تحلیل داده‌های پیچیده است.

برای درک شیوه کارکرد این ابزار، در ادامه به معرفی برخی از مفاهیم اصلی مهم در این چارچوب می‌پردازیم.

ایندکس یا شاخص: به صورت مفهومی ایندکس می‌تواند مشابه با جدول^۱ در پایگاه داده رابطه‌ای در نظر گرفته شود که اسناد در آن نگهداری می‌شوند. به عبارت دیگر منظور از ایندکس در ساختار الاستیک سرچ یک گروه از اسناد است که دارای شاخصه‌های مشابهی با یکدیگر هستند.

¹ Table

اسناد: اصلی‌ترین واحدهای اطلاعاتی اسناد^۱ یا داکيومنت‌ها هستند که در الاستیک سرچ ایندکس گذاری می‌شوند. در حقیقت می‌توان مفهوم اسناد در ایندکس را به شکل مشابه با ردیف‌ها (رکوردها) در یک جدول در پایگاه داده رابطه‌ای تصور نمود. همچنین یک داکيومنت را می‌توان به صورت مجموعه‌ای از فیلدها^۲ به روشی خاص در نظر گرفت که در قالب JSON تعریف شده و دارای یک شناسه منحصر به فرد در ایندکس است.

فیلدها: در واقع فیلدها را می‌توان مشابه با ستون‌های یک جدول در پایگاه داده رابطه‌ای فرض نمود با این تفاوت که در الاستیک سرچ فیلدها attribute های یک سند هستند که می‌توانند شامل اطلاعات متنی، عددی، تاریخی و یا دیگر انواع داده‌ها باشند. هر فیلد می‌تواند ویژگی‌های خاصی داشته باشد که برای جستجو، فیلتر کردن و تحلیل داده‌ها استفاده می‌شود. یکی از نکات مهمی که در این بخش وجود دارد اینست که Elasticsearch از ذخیره انواع داده‌های متعدد برای یک فیلد در یک ایندکس پشتیبانی نمی‌کند. در Elasticsearch، هر فیلد با یک نوع داده خاص مانند متن، کلمه کلیدی، عددی، تاریخ و... مرتبط است. هنگامی که یک نوع داده به یک فیلد اختصاص داده شد، تمام اسناد موجود در آن ایندکس باید به آن نوع داده پایبند باشند و اگر سندی با نوع داده‌ای متفاوتی برای یک فیلد فرستاده شود، Elasticsearch یا سند را رد می‌کند یا بر اساس قوانین نگاشت پویا خود، تبدیل نوع خودکار را انجام می‌دهد که این می‌تواند منجر به رفتار غیرمنتظره و تناقضات احتمالی داده‌ها شود. البته Elasticsearch همچنین از قابلیت MultiFields پشتیبانی می‌کند که به ما انعطاف پذیری می‌دهد تا از فیلد یک ایندکس به روش‌های مختلف برای اهداف مختلف استفاده کنیم.

نگاشت^۳: این واژه به طرح کلی ذخیره اسناد در ایندکس اشاره دارد و مشابه مفهوم طرح‌واره^۴ در پایگاه داده رابطه‌ای می‌باشد. در mapping، فیلدها، انواع داده‌ای برای هر فیلد و نحوه کنترل هر فیلد توسط الاستیک سرچ تعریف می‌شود.

گره^۵: این مفهوم به یک نمونه در حال اجرا از Elasticsearch اشاره دارد. به طور کلی یک سرور فیزیکی و مجازی بسته به قابلیت‌های منابع فیزیکی خود مانند رم، حافظه ذخیره سازی و قدرت پردازش، می‌تواند چندین گره را در خود جای دهد.

¹ Documents

² Fields

³ Mapping

⁴ schema

⁵ Node

خوشه^۱ یا کلاستر: خوشه مجموعه‌ای از یک یا چند گره است که در کنار یکدیگر قرار می‌گیرند. کلاستر قابلیت نمایه سازی و جستجوی جمعی را در تمام گره‌ها برای کل داده‌ها فراهم می‌کند. تکه شکسته^۲ یا شارد: ایندکس‌ها به صورت افقی به شاردها تقسیم می‌شوند. هر شارد یک نمونه مستقل از لوسین^۳ است و یک واحد کاری است که به طور خودکار توسط Elasticsearch مدیریت می‌شود. به عنوان کاربر، ما فقط باید تعداد شاردهای اولیه و replica را برای یک ایندکس مشخص کنیم و هرگز نباید به صورت جداگانه با شاردها برخورد کنیم، بلکه فقط در سطح ایندکس کار می‌کنیم.

ماکت^۴ یا رپلیکا: این قابلیت به کاربر این امکان را می‌دهد تا برای شاردها و ایندکس‌های خود همانندسازی کند. این عمل نه تنها به افزایش در دسترس بودن داده‌ها در صورت خرابی کمک می‌کند، بلکه با انجام عملیات جستجوی موازی در این کپی‌ها، عملکرد جستجو را نیز بهبود می‌بخشد. در این پروژه از روش‌های گوناگون برای ایجاد ایندکس‌ها و همچنین جستجوی داده‌های مورد نظر در الستیک سرچ استفاده نموده و دقت و سرعت عملکرد هر ایده را به طور جداگانه گزارش می‌کنیم.

معرفی شبکه عصبی مصنوعی عمیق ResNet101

شبکه مصنوعی ResNet101 یکی از مدل‌های پرکاربرد در زمینه تشخیص تصویر است. این شبکه مبتنی بر شبکه‌های عصبی عمیق و با استفاده از روش‌های جدید در آموزش شبکه‌های عصبی، به صورت خودکار و با دقت بالا قابل آموزش است.

ResNet101 از معماری شبکه‌های عصبی پیچشی (CNN) استفاده می‌کند که با استفاده از بلوک‌های کوچک سعی در بهبود عملکرد شبکه دارد. این شبکه با داشتن ۱۰۱ لایه، در تشخیص تصاویر با دقت بسیار بالایی عمل می‌کند.

در این شبکه، از روش Residual learning برای بهبود عملکرد و جلوگیری از مشکل ناپایداری در آموزش شبکه استفاده شده است. همچنین، این شبکه با استفاده از ترکیب لایه‌های کانولوشن، لایه‌های ادغامی و لایه‌های کاملاً متصل، قابلیت تشخیص اشیاء و ویژگی‌های مختلف در تصاویر را دارا می‌باشد.

¹ Cluster

² Shard

³ Lucene

⁴ Replica

به طور خلاصه، شبکه مصنوعی ResNet101 یکی از بهترین و پرکاربردترین شبکه‌های عصبی عمیق در زمینه تشخیص تصویر است که با استفاده از روش‌های جدید و معماری‌های پیچیده، قابل آموزش و دارای دقت بالایی در تشخیص تصاویر مختلف است.

استخراج ویژگی یا feature extraction با استفاده از شبکه مصنوعی ResNet101 به معنای استفاده از لایه‌های پیچشی و ادغامی این شبکه برای استخراج ویژگی‌های مختلف از تصاویر است. در این روش، تصویر ورودی به شبکه داده می‌شود و سپس با استفاده از لایه‌های پیچشی، ویژگی‌های مختلفی از تصویر استخراج می‌شود. سپس با استفاده از لایه‌های ادغامی، این ویژگی‌ها به صورت خلاصه شده و کاهش داده می‌شوند. در نهایت، با استفاده از لایه‌های کاملاً متصل، این ویژگی‌ها به صورت بردار ورودی به یک الگوریتم دسته‌بندی داده می‌شوند. استخراج ویژگی با استفاده از شبکه مصنوعی ResNet101 به دلیل دقت بالای این شبکه در تشخیص تصاویر، در بسیاری از برنامه‌های کاربردی مانند تشخیص چهره، تشخیص اشیاء و... استفاده می‌شود.

۳-۳. سیستم پیشنهادی

همان طور که در شکل ۱-۳ چارچوب کلی این روش معرفی شد، این پروژه از ۴ بخش کلی ایجاد شده است که یک سیستم کامل بازیابی تصویر مبتنی بر محتوا را تشکیل می‌دهند. به طور خلاصه ایده اصلی در تکنیک نمایه سازی پیشنهادی برای جستجوی بصری، محاسبه بردار جایگشت به ازای بردار ویژگی استخراج شده و سپس تبدیل آن بردار به نمایش متنی و در نهایت ذخیره و جستجو با استفاده از موتور جستجوی مبتنی بر متن الستیک سرچ می‌باشد.

در ادامه به شرح بخش‌های مختلف معرفی شده می‌پردازیم:

بخش اول - استخراج ویژگی

نخستین مرحله در فرآیند جستجوی بصری، استخراج بردارهای ویژگی به ازای هر یک از تصاویر موجود در مجموعه داده است. همانطور که پیش‌تر نیز بیان شد، امروزه در اکثر سیستم‌های بازیابی تصویر مبتنی بر محتوا از شبکه‌های عمیق برای انجام این کار استفاده می‌شود. همچنین طراحی و معرفی شبکه‌های عصبی عمیق با هدف بهبود عمل بازیابی، بسیاری از پژوهش‌های این حوزه را شامل می‌شود. از این رو، ما در این پروژه از شبکه عصبی مصنوعی عمیق ResNet101 برای عمل استخراج بردارهای ویژگی استفاده نمودیم. نتیجه بهره گیری از این شبکه CNN یک توصیفگر یا بردار ویژگی استخراج شده با ابعاد ۲۰۴۸ و با مقادیر حقیقی^۱ است.

^۱ real-valued

بخش دوم - پردازش و تولید نمایش متنی

از آن جایی که الستیک سرچ یک موتو جستجوی مبتنی بر متن است، از این رو در ادامه تلاش می‌کنیم تا به یک نمایش متن جایگزین^۱ مناسب از بردارهای استخراج شده دست یابیم که مهم‌ترین ویژگی‌های هر تصویر را حفظ کرده و علاوه بر آن به ما امکان می‌دهد تا از یک موتور بازیابی متن موجود برای ایجاد یک سیستم بازیابی تصویر مبتنی بر محتوا استفاده کنیم.

پس از اتمام مرحله نمایش تصویر توسط بردارهای ویژگی، باید با استفاده از رویکردهای معرفی شده در (Amato et al., 2016; Gennaro et al., 2010) فرم متنی بردارهای به دست آمده را محاسبه کنیم. به طور خلاصه می‌توان گفت که ابتدا باید بردارهای جایگشت متناسب با هر بردار ویژگی استخراج شده را به دست آوریم و سپس به ازای هر تصویر به کمک الگوی خاص مبتنی بر پارامتر k جهت کوتاه سازی، بردار را به فرم متنی متناظر با آن تبدیل کنیم.

نخستین مرحله در تبدیل بردار ویژگی به نمایش متنی، اعمال تبدیل واحد خطی اصلاح شده پیوسته^۲ یا CReLU بر بردارهای استخراج شده است.

در بردار استخراج شده از تصویر، هر یک از ابعاد بیانگر یک ویژگی پایه در تصویر ورودی هستند. هر چه مقدار این ویژگی پایه در بردار بیشتر باشد تاثیری که در مقایسه شباهت دو تصویر نیز دارد، بیشتر است زیرا معمولاً روش‌های اندازه گیری شباهت از یک معیار خاص برای تعیین فاصله بین ابعاد استفاده می‌کنند و هر چه مقدار این فاصله کمتر باشد میزان شباهت بالاتر است. بنابراین به عبارت دیگر، دو عکس که در مقادیر ویژگی‌های پایه بیشتر به هم شباهت دارند، در نتیجه شباهت کلی آن‌ها هم بیشتر خواهد بود. بردارهای به دست آمده از شبکه‌های عمیق CNN شامل تعداد زیادی از اعداد مثبت و منفی هستند ولی هنگام استفاده از برخی از رویکردها، عناصر منفی عملاً نادیده گرفته می‌شوند در حالی که همان طور که گفته شد، مقادیر عناصر صرف نظر از علامتی که دارند، در تعیین میزان شباهت مهم و تاثیرگذار خواهند بود. در این رویکرد به منظور جلوگیری از این عدم تعادل بین مقادیر ابعاد مثبت و ابعاد منفی، از تبدیل واحد خطی اصلاح شده پیوسته (CReLU) استفاده می‌کنیم.

در همین راستا، ابتدا یک کپی یکسان از عناصر بردار ایجاد می‌کنیم و سپس علامت مقادیر بردار کپی شده را برعکس می‌کنیم یا به طور ساده بردار دوم را منفی می‌کنیم و در ادامه بردار کپی شده و

¹ Surrogate Text Representation (STR)

² Concatenated Rectified Linear Unit transformation

منفی شده را به انتهای بردار اولیه می‌چسبانیم. در پایان تابع ReLU را روی بردار جدید اعمال می‌کنیم یعنی به ازای هر یک از عناصر در بردار جدید، اگر علامت مثبت بود مقدار عنصر و اگر علامت منفی بود صفر در نظر می‌گیریم. بدین ترتیب بردار حاصل، برداری با ابعاد ۴۰۹۶ خواهد بود که اگرچه این ابعاد ۲ برابر ابعاد بردار اولیه است اما مزیت آن در این است که شامل تمامی ویژگی مهم تصویر است و در ادامه خواهیم دید که با کمک روش کوتاه سازی با استفاده از پارامتر K برای تبدیل به نمایش متنی، می‌توان عناصر برداری فراتر از K را نادیده گرفت و این به ما امکانی می‌دهد که طول بردارها را تعدیل کنیم و اندازه ایندکس را کاهش دهیم.

دومین مرحله از فرآیند تولید نمایش متنی، ایجاد جایگشت‌ها^۱ به ازای هر یک از بردارهای ویژگی خواهد بود.

در این مرحله هدف به دست آوردن نمایش خاصی از شناسه (شاخص^۲) های بردار است که مربوط به ابعاد دارای ویژگی های پایه هستند. به عبارت دیگر جایگشت‌ها همان شاخص‌های عناصر در بردارهای ویژگی عمیق هستند و ما در اینجا می‌خواهیم چیدمانی از جایگشت‌ها را به دست آوریم که شامل مهم ترین ویژگی های استخراج شده از تصویر باشند که مقادیر آن‌ها از سایر عناصر در ابعاد دیگر بالاتر است زیرا همان طور که پیش‌تر بیان شد، این ویژگی‌ها تاثیر بیشتری در مقایسه شباهت بین تصاویر دارند.

بدین منظور، پس از اعمال تبدیل CReLU و به دست آوردن بردار با ابعاد جدید، برای به دست آوردن جایگشت متناظر، باید شاخص‌های عناصر بردار را به ترتیب نزولی و با توجه به مقادیر عناصر مربوطه مرتب سازی کنیم. برای مثال بردار زیر را در نظر بگیرید:

$$fv = [0.1, 0.3, 0.4, 0, 0.2]$$

(البته توجه به این نکته اهمیت دارد که این تنها مثالی برای درک بهتر شیوه اجرا شده در این پروژه می‌باشد زیرا که در واقعیت ابعاد این بردار ۲۰۴۸ یا بعد از اعمال CReLU ۴۰۹۶ خواهد بود). نمایش (اولیه) مبتنی بر جایگشت^۳ برای این بردار با مرتب سازی ابعاد این بردار بر اساس مقادیر آن‌ها به ترتیب نزولی به دست خواهد آمد که نتیجه آن مطابق زیر خواهد بود:

$$\Pi_{fv} = (3, 2, 5, 1, 4)$$

¹ Permutant

² index

³ Permutation-Based Representation

این بدان معناست که اولین ویژگی پایه با بیشترین مقدار در موقعیت شاخص ۳، دومین ویژگی پایه در موقعیت شاخص ۲، سومین ویژگی پایه در موقعیت شاخص ۵ و به همین ترتیب بقیه جایگشت ها قرار می گیرند.

در ادامه باید نمایش معکوس این بردارها را به دست بیاوریم. توجه داشته باشید که نمایش معکوس جایگشتها Π^{-1} ، برداری است که ما از آن به عنوان بردار جایگشت یاد می کنیم و به ما اجازه می دهد تا به راحتی بیشتر توابع فاصله بین جایگشتها را تعریف کنیم. جایگشتها به طور کلی با استفاده از فاصله Spearman rho، Kendall Tau یا Spearman Footrule مقایسه می شوند. در Π مقدار در هر موقعیت از دنباله، شناسه محور در آن موقعیت است ولی در نمایش معکوس جایگشتها Π^{-1} ، هر موقعیت مربوط به یک محور و مقدار در هر موقعیت با رتبه محور مربوطه مطابقت دارد. (Amato et al., 2016, 2018)

محاسبه نمایش معکوس توابع به این صورت است که اگر تابع $G(x)=y$ باشد آنگاه $G^{-1}(y)=x$ است. پس برای مثال قبلی نمایش معکوس جایگشتها یا همان بردار جایگشت به این صورت است:

$$\Pi_{fv}^{-1} = (4, 2, 1, 5, 3)$$

مرحله سوم و پایانی در این بخش، تولید نمایش متنی بر اساس بردار جایگشت به دست آمده به ازای هر یک از تصاویر موجود در مجموعه داده می باشد.

ما در نهایت، به منظور ذخیره سازی بردارهای جایگشت با یک موتور بازیابی متن و به عنوان مثال Elasticsearch، از نمایش متن جایگزین معرفی شده در (Gennaro et al., 2010) استفاده می کنیم. در این رویکرد پس از به دست آوردن بردار جایگشت (نمایش معکوس جایگشت) ابتدا عملیات کوتاه سازی^۱ بر اساس پارامتر k (که توسط کاربر تعیین می شود) را اعمال می کنیم. هدف از این کار در نظر گرفتن عناصر با اهمیت بیشتر یا به عبارتی top-K در بردار جایگشت است که باعث می شود از عناصر با اهمیت کم تر چشم پوشی شود و طول فرم متنی کاهش پیدا کند.

روش کار به این صورت است که در هر یک از موقعیت های عناصر در بردار جایگشت، اگر مقدار موجود بزرگتر از مقدار k بود، مقدار بردار در آن موقعیت را برابر $k + 1$ قرار می دهیم و اگر مقدار موجود از k کوچک تر و یا با آن مساوی باشد، بدون تغییر باقی خواهد ماند.

در ادامه، شناسه هر عنصر $ri \in \Pi^{-1}$ را با یک کلمه کلیدی الفبایی عددی منحصر به فرد مانند ti مرتبط می کنیم. سپس مقدار تکرار مربوط به هر شناسه در بردار را به این صورت به دست می آوریم

¹ truncate

که اگر p در موقعیت ri در بردار جایگشت (کوتاه شده) ظاهر شود، آنگاه اصطلاحی مانند τ_i به تعداد $p - (k + 1)$ بار در متن جایگزین تکرار می‌شود. در نهایت از الحاق عبارتهای به دست آمده بر اساس میزان تکرار آنها به ترتیب نزولی، متن جایگزین نهایی به دست می‌آید.

برای درک بیشتر روش کار در این بخش به طور کلی، فرض کنید بردار ویژگی استخراج شده f_v را به صورت زیر تعریف کنیم (در واقعیت ابعاد این بردار ۲۰۴۸ است):

$$f_v = [0.1, -0.3, -0.4, 0, 0.2]$$

پس از اعمال تبدیل CReLU بر روی آن، نتیجه زیر حاصل می‌شود:

$$f_v^+ = [0.1, 0, 0, 0, 0.2, 0, 0.3, 0.4, 0, 0]$$

برای به دست آوردن بردار جایگشت ابتدا باید نمایش اولیه مبتنی بر جایگشت برای این مثال را محاسبه کنیم که همان مرتب کردن شناسه‌های عناصر بردار بر اساس مقادیر آنها به ترتیب نزولی است:

$$\Pi_{f_v} = (8, 7, 5, 1, 2, 3, 4, 6, 9, 10)$$

و پس از آن با معکوس کردن بردار فوق، بردار جایگشت بدین ترتیب به دست می‌آید:

$$\Pi^{-1}_{f_v} = (4, 5, 6, 7, 3, 8, 2, 1, 9, 10)$$

حالا نوبت به تبدیل به نمایش متنی می‌رسد که برای این منظور، ابتدا کوتاه سازی بر اساس پارامتر k (فرضاً $k=4$) را اعمال می‌کنیم که بردار جایگشت جدید به این شکل در می‌آید:

$$\Pi^{-1}_{f_v} = (4, 5, 5, 5, 3, 5, 2, 1, 5, 5)$$

پس از آن مقدار تکرار در هر یک از موقعیت‌ها را بر اساس قاعده $p - (k + 1)$ به این ترتیب محاسبه می‌کنیم:

$$\tau_1 = (4+1) - 4 = 1$$

$$\tau_{2,3,4,6,9,10} = (4+1) - 5 = 0$$

$$\tau_5 = (4+1) - 3 = 2$$

$$\tau_7 = (4+1) - 2 = 3$$

$$\tau_8 = (4+1) - 1 = 4$$

و در نهایت نمایش متن جایگزین برای بردار f_v بدین شکل خواهد بود:

$$\tau_8 \tau_8 \tau_8 \tau_8 \tau_7 \tau_7 \tau_7 \tau_5 \tau_5 \tau_1$$

بخش سوم - ذخیره در الستیک سرچ

در بخش بعد از به دست آوردن نمایش متنی از بردار ویژگی، نوبت به ذخیره داده‌ها در الستیک سرچ به منظور انجام عمل جستجو در میان آن‌ها می‌رسد. از این مرحله همچنین با عنوان نمایه سازی

یا indexing یاد می‌شود چرا که هدف آن، طراحی ساختار بهینه برای ذخیره سازی داده‌ها در ایندکس‌های الستیک سرچ خواهد بود.

برای انجام این بخش باید مراحل زیر را دنبال نمود:

(۱) ایجاد یک index: همانطور که پیش‌تر نیز اشاره شد، ایندکس را می‌توان مشابه با جدول در پایگاه داده رابطه‌ای در نظر گرفت. بنابراین قبل از ارسال داده‌ها در قالب اسناد به الستیک سرچ باید ایندکسی را که قرار است داده‌ها در آن ذخیره شوند، با مشخصات مناسب ایجاد کنیم.

(۲) تعریف mapping برای index: الستیک سرچ، این امکان را می‌دهد تا ساختار اسناد خود را با ایجاد یک mapping تعریف کنید. mapping می‌تواند انواع داده‌ها و تنظیمات هر فیلد را در یک سند مشخص کند که این به الستیک سرچ کمک می‌کند تا داده‌ها را درک کند و به طور خودکار فرآیند نمایه سازی را بهینه کند. یکی از مهم‌ترین چالش‌ها در این بخش تعیین یک ساختار مناسب mapping برای هر ایندکس به منظور بهینه سازی فرآیند جستجو در آینده خواهد بود.

(۳) آماده سازی داده‌ها: در مرحله بعدی باید داده‌ها به فرمت مناسب (JSON) تبدیل کنیم که این فرمت باید با ساختار mapping تعریف شده در مرحله پیشین، مطابقت داشته باشد. بدین ترتیب سندی را که قرار است در الستیک سرچ ذخیره شود، آماده می‌کنیم.

(۴) ایندکس کردن اسناد: در این مرحله باید با سندی که تا اینجا آماده شده است، یک درخواست نمایه سازی به الستیک سرچ ارسال کنیم. الستیک سرچ سند را تجزیه می‌کند، آن را در برابر mapping تعریف شده، اعتبارسنجی می‌کند و آن را در index مشخص شده ذخیره می‌کند.

ما در این پروژه روش‌های گوناگونی را برای طراحی mapping و نوع ذخیره سازی داده‌ها در ایندکس‌های الستیک سرچ امتحان کردیم که در بخش ۳-۴ به طور مفصل‌تری به جزئیات آنها خواهیم پرداخت.

بخش چهارم - جستجو و ارزیابی

پس از اتمام بخش indexing، باید جستجو برای پرس و جوهای از پیش تعریف شده انجام شود. در این بخش ابتدا باید تصاویری را که قرار است به عنوان کوئری به سیستم جستجوی بصری فرستاده شوند، تعیین کنیم. این تصاویر می‌توانند به صورت دلخواه و یا تصادفی از خود پایگاه داده مورد آزمایش انتخاب گردند که در این پروژه ما برای مجموعه داده (INRIA Holidays) تصاویری را که در شرح خود مجموعه داده برای آزمایش و ارزیابی روش بازیابی معرفی شده‌اند، انتخاب کردیم و در ادامه برای مجموعه داده فلیکر، تعدادی تصویر را به صورت تصادفی انتخاب کردیم زیرا در آزمون‌های مربوط به این مجموعه داده، ارزیابی سرعت سیستم پیشنهادی و مقیاس پذیری آن در ابعاد بالا به

عنوان هدف در نظر گرفته شده است و اهمیت هر تصویر به تنهایی مورد توجه قرار نمی گیرد بلکه تعداد بالای آن ها محور آزمایش ها قرار دارد.

پس از این مرحله، باید تصویر (تصاویر) پرس و جو به روشی کاملاً مشابه با روش تولید نمایش متنی اجزای پایگاه داده (یعنی همان بخش اول و دوم سیستم) استخراج ویژگی و کد گذاری شود. سپس باید این نمایش متنی ایجاد شده را به شکل مناسب برای جستجو در الاستیک سرچ تبدیل نمود. برای جستجو در الاستیک سرچ، می توان از RESTful API یا رابط کاربری رسمی الاستیک سرچ موجود برای زبان های برنامه نویسی مختلف استفاده نمود. ولی به هر حال هر کدام از روش ها که انتخاب شوند، باید ابتدا طراحی کوئری به صورت صحیح با استفاده از Query DSL انجام گردد.

Query DSL یک زبان دامنه خاص است که برای ساخت پرس و جو در Elasticsearch استفاده می شود. با Query DSL، کاربران می توانند پرس و جوهای جستجوی پیچیده را فرموله کنند و اسناد مربوطه را بر اساس شرایط و فیلترهای مختلف بازیابی کنند و بدین شکل امکان تعریف معیارهای جستجو با استفاده از نحوی شبیه JSON فراهم می آید.

Query DSL طیف وسیعی از انواع پرس و جو را ارائه می دهد، از جمله پرس و جوهای اصطلاحی (term query)، پرس و جوهای مطابقت (match query)، پرس و جوهای محدوده ای (range query)، پرس و جوهای ترکیبی (bool query) و موارد دیگر. هر نوع پرس و جو دارای پارامترها و عملکرد خاصی است که می تواند برای دستیابی به نتایج جستجوی مورد نظر سفارشی شود. علاوه بر این، Query DSL امکان ترکیب چند پرس و جو را با استفاده از عملگرهای منطقی مانند AND، OR، و NOT می دهد.

در مرحله بعد از آماده کردن پرس و جو به شکل مناسب و همچنین سازگار با mapping تعریف شده برای ایندکس، باید یک درخواست جستجو با کوئری طراحی شده برای الاستیک سرچ ارسال شود. الاستیک سرچ کوئری را دریافت کرده و آن را تجزیه و تحلیل کرده و به ترتیب مشابه ترین اسناد موجود در ایندکس را به عنوان پاسخ برمی گرداند.

در آخرین مرحله باید نتایج بازگردانده شده را دریافت کرده و آن ها را به فرمت مناسب برای مازول ارزیابی درآوریم و پس از بررسی، نتایج نهایی سیستم را گزارش کنیم.

در این پروژه روش های مختلفی برای طراحی کوئری جهت حصول نتایج بهینه برای سیستم بازیابی تصویر مبتنی بر محتوا، مورد آزمایش قرار گرفت که در بخش ۳-۴ به جزئیات بیشتری در این خصوص می پردازیم.

۳-۴. جزییات بخش ذخیره سازی و جستجو

در قسمت های پیشین توضیحاتی درباره چارچوب کلی سیستم پیشنهادی و شیوه پیاده سازی بخش های مختلف آن برای استخراج بردار ویژگی از تصویر و سپس تبدیل آن به نمایش متنی برای ذخیره سازی و جستجو در موتور جستجوی مبتنی بر متن الاستیک سرچ گفته شد. اما همچنان این سوال وجود دارد که چگونه می توان با ارائه نحوه ذخیره سازی و جستجو خاص و سازگار با قابلیت های الاستیک سرچ و نیز ایجاد تغییر در ساختاردهی داده ها، عملکرد کلی سیستم بازیابی تصویر را بهبود بخشید؟

این سوال پژوهشی مرکز توجه پایان نامه حاضر است و امیدواریم با پاسخ به آن، سهم خویش را از نوآوری در این عرصه، به جامعه علمی و صنعتی ارائه دهیم. در این قسمت ایده هایی برای پیاده سازی در بخش indexing و جستجو در الاستیک سرچ بیان می شود و جزییات بیشتری در رابطه با نحوه طراحی ایندکس ها و نحوه ذخیره سازی داده ها در هر ایندکس و همین طور شیوه نگارش کوئری ها با استفاده از Query DSL توضیح داده می شود.

برای درک دقیق تر چگونگی پیاده سازی و اجرای پیشنهادات ارائه شده، مثالی بسیار ساده از نمایش های متنی که باید در الاستیک سرچ، ایندکس و جستجو شوند بیان می شوند. البته بدیهی ست که در واقعیت تعداد رشته های متنی بسیار بالاتر و طول ها آن ها بسیار بیشتر است و این ها صرفاً مثال هایی برای درک چگونگی کارکرد ایده ها در سیستم پیشنهادی خواهند بود.

فرض کنید رشته های متنی که قرار هست ذخیره سازی (index) کنیم این ها باشند:

1. T8T8T8T8T2T2T2T3T3T9 (image1.jpg)
2. T5T5T5T5T1T1T1T6T6T2 (image2.jpg)

و رشته ای که قرار هست کوئری شود به این صورت باشد:

T5T5T5T5T2T2T2T3T3T6

در ادامه به بررسی جزییات پیشنهادی جهت پیاده سازی ایده های مربوطه می پردازیم.

ایده نخست – جستجوی فازی

جستجوی فازی در الستیک سرچ یک روش جستجو است که به کاربر امکان می‌دهد نتایج جستجو را براساس شباهت‌های مفهومی به جای دقت کامل بررسی کند. این روش معمولاً در مواردی استفاده می‌شود که کاربر نتایج جستجوی دقیق را نمی‌داند یا نمی‌تواند به طور کامل و دقیق عبارات مورد نظر خود را تایپ کند.

با استفاده از جستجوی فازی، الستیک سرچ قادر است نتایجی را پیدا کند که شامل کلمات مشابه، مترادف یا مرتبط با عبارات جستجو شده هستند. برای این منظور، الگوریتم‌های پیچیده‌ای مانند تحلیلگر لغوی (Fuzzy Analyzer) و تحلیلگر معنایی (Semantic Analyzer) استفاده می‌شود. این الگوریتم‌ها قادرند به صورت هوشمندانه با عبارات جستجو شده مطابقت دهند و نتایجی را با استفاده از معیارهای شباهت و ارزیابی مفهومی به کاربر نمایش دهند.

در جستجوی فازی مطابقت عین عبارت در اسناد بررسی نمی‌شود و می‌توان تعیین کرد که اگر اختلافی در حد یک یا چند کاراکتر متنی نیز وجود داشت، سند مورد نظر به عنوان پاسخ بازگردانده شود که این قابلیت برای جستجو و تعیین میزان شباهت (و نه تطبیق کامل) مفید خواهد بود.

در این قسمت برای طراحی ساختار ایندکس، از ۲ فیلد که یکی برای نگهداری عنوان تصویر (از نوع keyword) و دیگری برای نگهداری رشته متنی جایگزین برای بردار ویژگی تصویر (از نوع text) هستند، استفاده می‌کنیم.

به طور مثال، برای نمونه‌های بیان شده، ساختار ایندکس در این روش به صورت زیر خواهد بود:

title	text_code
image1.jpg	T8 T8 T8 T8 T2 T2 T2 T3 T3 T9
image2.jpg	T5 T5 T5 T5 T1 T1 T1 T6 T6 T2

و در قسمت جستجو از کوئری (API REST) به شکلی که در ادامه می‌آید، استفاده می‌کنیم:

```

1. {
2.   "query": {
3.     "bool": {
4.       "should": [
5.         {
6.           "match": {
7.             "text_code": {
8.               "query": "T5 T5 T5 T5 ",
9.               "fuzziness": "AUTO"
10.            }
11.          }
12.        },
13.        {
14.          "match": {
15.            "text_code": {
16.              "query": "T2 T2 T2 T2 ",
17.              "fuzziness": "AUTO"
18.            }
19.          }
20.        },
21.        {
22.          "match": {
23.            "text_code": {
24.              "query": "T3 T3 T3 T3 ",
25.              "fuzziness": "AUTO"
26.            }
27.          }
28.        },
29.        {
30.          "match": {
31.            "text_code": {
32.              "query": "T6 T6 T6 T6 ",
33.              "fuzziness": "AUTO"
34.            }
35.          }
36.        }
37.      ]
38.    }
39.  }
40.}

```

پرس و جو ۳-۱: جستجو به روش فازی

در پرس و جو ۳-۱: جستجو به روش فازی مهمترین متغیر fuzziness است که مقدار آن برابر AUTO قرار داده شده است. وقتی این متغیر برای جستجو فیلدی با نوع داده ای text و یا keyword به کار می رود، به عنوان فاصله ویرایشی لوناشتاین^۱ تفسیر می شود و به طور خلاصه به معنی کمترین تعداد عملیات مورد نیاز برای تبدیل یک رشته متنی به رشته متنی دیگر است و به نوعی می توان آن را تعمیم فاصله همینگ دانست. به عنوان مثال فاصله لوناشتاین بین "kitten" و "sitting" برابر ۳ است زیرا حداقل سه ویرایش برای تبدیل یکی به دیگری وجود دارد و کمتر از آن ممکن نیست.

متغیر fuzziness در الستیک سرچ می تواند مقادیر ۰، ۱، ۲ را بپذیرد که ۰ عملاً به معنای تطبیق کامل دو رشته متنی پرس و جو و رشته متنی ایندکس شده است و ۱ به معنای مجاز بودن حداکثر یک تغییر برای ایجاد تطبیق و ۲ به معنای مجاز بودن حداکثر دو تغییر برای ایجاد تطبیق بین دو رشته می باشد.

¹ Levenshtein Edit Distance

اگر مقدار متغیر fuzziness برابر AUTO قرار داده شود، فاصله ویرایش بر اساس طول عبارت متنی تعیین می شود. معمولاً در این حالت مقدار متغیر می تواند به شکل $AUTO:[low],[high]$ به کار رود که مقادیر آرگومان های low و high به ترتیب بیانگر کم ترین و بیشترین طول عبارت متنی هستند. اگر طول عبارت از مقدار low کمتر باشد فاصله ویرایش برابر ۰ و اگر طول عبارت اندازه ای بین low و high داشته باشد مقدار فاصله ویرایش برابر ۱ و در صورتی که طول عبارت از high بیشتر باشد فاصله ویرایش برابر ۲ تعیین می شود.

اگر هیچ مقداری برای آرگومان های low و high در نظر گرفته نشود به طور پیش فرض مقادیر آن ها به صورت $AUTO:3,6$ خواهد بود که در این صورت اگر طول عبارت متنی کوتاه تر از ۳ باشد، باید تطبیق کامل بین دو عبارت مورد جستجو صورت بگیرد (فاصله برابر ۰) و اگر طول عبارت بین ۳ تا ۶ باشد این فاصله ویرایشی حداکثر برابر ۱ و اگر طول عبارت متنی بیشتر از ۶ باشد، فاصله ویرایش حداکثر برابر ۲ در نظر گرفته می شود.

ایده دوم - جستجو عین عبارت

یکی از روش های جستجو در Elasticsearch جستجوی match برای بررسی شباهت های میان یک رشته متنی و رشته متنی موجود در یک فیلد از یک سند ذخیره شده است. معمولاً در این روش از الگوریتم Term Frequency-Inverse Document Frequency (TF-IDF) برای بررسی شباهت استفاده می شود.

الگوریتم term frequency-inverse document frequency (TF-IDF) یک روش شمارشی است که برای تخمین اهمیت یک واژه در یک سند و نیز برای ساماندهی و جستجو در اسناد، به خصوص در روش های بازیابی اطلاعات، استفاده می شود. در این الگوریتم، ابتدا شمارش تعداد تکرار واژه ها (term frequency) در یک سند صورت می گیرد و هر چه این مقدار بالاتر باشد، میزان مرتبط بودن آن واژه با پاسخ پرس و جو را نشان می دهد. سپس برای محاسبه اهمیت واژه ها، تعداد سندهایی که واژه مورد نظر در آنها وجود دارد (inverse document frequency) محاسبه می شود و هر چه این تعداد نسبت به کل سندهای موجود، بیشتر باشد یعنی اهمیت آن واژه در جستجو پایین تر است. در نهایت، مقدار tf و idf ضرب شده و میزان اهمیت کلی واژه مشخص می شود. اگر عبارت مورد جستجو از بیشتر از یک واژه تشکیل شده باشد، به طور جداگانه برای هر یک TF-IDF محاسبه می شود و نتیجه نهایی از مجموع آن ها به دست می آید.

در این قسمت نیز مشابه قسمت قبلی، برای طراحی ساختار ایندکس، از ۲ فیلد که یکی برای نگهداری عنوان تصویر (از نوع keyword) و دیگری برای نگهداری رشته متنی جایگزین برای بردار ویژگی تصویر (از نوع text) هستند، استفاده می‌کنیم.

به طور مثال، برای نمونه‌های بیان شده قبلی، ساختار ایندکس در این روش به صورت زیر خواهد بود:

title	text_code
image1.jpg	T8 T8 T8 T8 T2 T2 T2 T3 T3 T9
image2.jpg	T5 T5 T5 T5 T1 T1 T1 T6 T6 T2

و در قسمت جستجو از کوئری (API REST) به شکل زیر استفاده می‌کنیم:

```

1. {
2.   "query": {
3.     "match": {
4.       "text_code": "T5 T5 T5 T5 T2 T2 T2 T3 T3 T6"
5.     }
6.   }
7. }
```

پرس و جو ۳-۲: جستجوی عین عبارت

ایده سوم - جستجو با حذف تکرار

در این قسمت ابتدا کلمات تکراری را در هر عبارت متن جایگزین حذف می‌کنیم و فقط ترتیب کلمات کلیدی الفبایی عددی منحصر به فرد را در عبارت حفظ می‌کنیم. دلیل پشت این کار، اینست که احتمال می‌رود با کوتاه شدن نمایش متنی جایگزین، سرعت عمل جستجو تا حدودی افزایش یابد. سپس برای طراحی ساختار ایندکس، از ۲ فیلد که یکی برای نگهداری عنوان تصویر (از نوع keyword) و دیگری برای نگهداری رشته متنی جایگزین با حذف تکرار عبارت‌ها (از نوع text) هستند، استفاده می‌کنیم.

به طور مثال، برای نمونه‌های بیان شده، ساختار ایندکس در این روش به صورت زیر خواهد بود:

title	text_code
image1.jpg	T8 T2 T3 T9
image2.jpg	T5 T1 T6 T2

و در قسمت جستجو از کوئری (API REST) به شکل زیر استفاده می‌کنیم:

```

1. {
2.   "query": {
3.     "match": {
4.       "text_code": "T5 T2 T3 T6"
5.     }
6.   }
7. }

```

پرس و جو ۳-۳: جستجو با حذف تکرار

ایده چهارم - جستجو با در نظر گرفتن پیشوندها

برای پیاده سازی در این قسمت ابتدا مشابه روش قبلی، کلمات تکراری را در عبارت متن جایگزین حذف می‌کنیم و فقط ترتیب در کلمات کلیدی الفبایی عددی منحصر به فرد را حفظ می‌کنیم. سپس برای طراحی ساختار ایندکس از تعداد $k + 1$ فیلد استفاده می‌کنیم (برای اطلاعات بیشتر در مورد پارامتر k به فصل قبل مراجعه نمایید). که یک فیلد برای نگهداری عنوان تصویر (از نوع keyword) و بقیه برای نگهداری نمایش متنی به صورت پیشوند در نظر گرفته می‌شوند که از هر دو نوع داده keyword و text در جستجو پشتیبانی می‌کنند. (با بهره گیری از قابلیت MultiFields)

نحوه قرارگیری نمایش‌های متنی در ایندکس به این صورت خواهد بود که در هر فیلد یک پیشوند یعنی نخستین کلمه (یا کلمات) کلیدی از عبارت نمایش متنی جای گذاری می‌شود و در فیلد بعدی کلمه کلیدی بعدی به پیشوند اضافه می‌شود و به همین ترتیب تا آخرین فیلد ادامه می‌یابد. در واقع ارتباط این دنباله‌های پیشوندی در الگوی آن‌ها برای اضافه کردن یک یا چند کلمه کلیدی جدید به انتهای دنباله قبلی است. هر دنباله بر روی دنباله قبلی ساخته می‌شود و پیشرفتی از اصطلاحات را ایجاد می‌کند.

جستجو در این ایندکس می‌تواند بر اساس مقدار داده‌ای در نوع text یا همان مقدار در نوع keyword و یا هر دو حالت باشد.

به طور مثال، برای نمونه‌های بیان شده، ساختار ایندکس در این روش به صورت زیر خواهد بود:

title	prefix1	prefix2	prefix3	prefix4
image1.jpg	T8	T8 T2	T8 T2 T3	T8 T2 T3 T9
image2.jpg	T5	T5 T1	T5 T1 T6	T5 T1 T6 T2

و در قسمت جستجو می‌توان از کوئری (API REST) به شکل زیر استفاده نمود:

```

1. {
2.   "query": {
3.     "bool": {
4.       "should": [
5.         {
6.           "match": {
7.             "prefix1.disjoint": "T5 "
8.           }
9.         },
10.        {
11.          "match": {
12.            "prefix2.disjoint": "T5 T2 "
13.          }
14.        },
15.        {
16.          "match": {
17.            "prefix3.disjoint": "T5 T2 T3 "
18.          }
19.        },
20.        {
21.          "match": {
22.            "prefix4.disjoint": "T5 T2 T3 T6 "
23.          }
24.        }
25.      ]
26.    }
27.  }
28.}

```

پرس و جو ۳-۴-۱: جستجو با در نظر گرفتن پیشنوندها برای text

این کوئری را می‌توان هم برای جستجو به صورت keyword (بدون استفاده از عبارت disjoint که زیر آن خط کشیده شده است) و هم برای جستجو به صورت text (با استفاده از عبارت disjoint) به کار برد و ساختار کلی آن به همین صورت خواهد بود. لازم به ذکر است نام گذاری disjoint به صورت دلخواه انجام شده است. برای جستجوی ترکیبی هم به صورت text و هم به صورت keyword نیز از کوئری به شکلی که در ادامه می‌آید، استفاده خواهد شد که در آن هر دو حالت (با و بدون استفاده از عبارت disjoint) دیده می‌شود.

```

1. {
2.   "query": {
3.     "bool": {
4.       "should": [
5.         {
6.           "multi_match": {
7.             "query": "T5 ",
8.             "fields": [
9.               "prefix1",
10.              "prefix1.disjoint"
11.            ]
12.          }
13.        },
14.        {
15.          "multi_match": {
16.            "query": "T5 T2 ",
17.            "fields": [
18.              "prefix2",
19.              "prefix2.disjoint"
20.            ]
21.          }
22.        },
23.        {
24.          "multi_match": {
25.            "query": " T5 T2 T3 ",
26.            "fields": [
27.              "prefix3",
28.              "prefix3.disjoint"
29.            ]
30.          }
31.        },
32.        {
33.          "multi_match": {
34.            "query": "T5 T2 T3 T6 ",
35.            "fields": [
36.              "prefix4",
37.              "prefix4.disjoint"
38.            ]
39.          }
40.        }
41.      ]
42.    }
43.  }
44.}

```

پرس و جو ۳-۴-۲: جستجو با در نظر گرفتن پیشنودها برای text و keyword

ایده پنجم - جستجو بر اساس پارتیشن بندی

پیاده سازی در این قسمت اندکی با روش های گفته شده پیشین متفاوت خواهد بود زیرا که انجام تغییرات در نمایش جایگزین برای بردار، پیش از بخش تولید نمایش متنی آغاز می شود. در این قسمت پس از استخراج و تعیین بردار ویژگی، هر بردار (مربوط به یک تصویر) به تعداد مشخص n قطعه تقسیم می شود (طول این قطعه ها تقریباً برابر است) و برای هر قطعه به طور جداگانه الگوریتم بخش تولید نمایش متنی انجام شده و در نهایت به جای یک نمایش متنی به ازای هر بردار، حالا تعداد n نمایش متنی مجزا خواهیم داشت.

برای طراحی ساختار ایندکس، از $n + 1$ فیلد استفاده خواهیم کرد که یک فیلد برای نگهداری عنوان تصویر (از نوع keyword) و بقیه برای نگهداری n نمایش متنی مجزا به ازای هر تصویر (از نوع text) در نظر گرفته می‌شوند.

برای بخش جستجو نیز، مشابه با بخش ایندکس، ابتدا بردار ویژگی استخراج شده برای تصویر پرس و جو را به n قطعه تقسیم می‌کنیم و برای هر قطعه به صورت جداگانه الگوریتم بخش تولید نمایش متنی جایگزین را اجرا می‌کنیم که در آخر تعداد n نمایش متنی مجزا به ازای هر تصویر پرس و جو خواهیم داشت.

برای مثال نمونه‌ای از ساختار ایندکس در نظر گرفته شده برای این قسمت ($n=4$) را در ادامه

می‌بینید:

title	part1	part2	part3	part4

و نیز نمونه‌ای از کوئری (API REST) برای این قسمت به صورتی که ادامه می‌آید، خواهد بود:

```

1. {
2.   "bool": {
3.     "should": [
4.       {
5.         "match": {
6.           "part1": "T62 T143 T39 T181 "
7.         }
8.       },
9.       {
10.        "match": {
11.          "part2": "T196 T90 T329 T224 "
12.        }
13.      },
14.      {
15.        "match": {
16.          "part3": "T257 T111 T42 T66 "
17.        }
18.      },
19.      {
20.        "match": {
21.          "part4": "T326 T11 T65 T349 "
22.        }
23.      }
24.    ]
25.  }
26.}
```

پرس و جو ۳-۵: جستجو بر اساس پارتیشن بندی

۳-۵. جمع بندی

در این فصل ابتدا بخشی از چارچوب ها و مواد استفاده شده در این پروژه را بیان کردیم و به توضیح روش پیشنهادی به کار رفته در این پژوهش پرداختیم و چگونگی عملکرد بخش های مختلف سیستم را به طور جداگانه مورد بحث قرار دادیم. همچنین ایده هایی را برای ایجاد بهبود در قسمت نمایه سازی و جستجو با استفاده از الستیک سرچ ارائه نمودیم.

در فصل بعدی به پیاده سازی ایده های پیشنهادی و بررسی کارایی هر یک به طور جداگانه و نیز در مقایسه با یکدیگر، خواهیم پرداخت.

فصل چهارم

آزمایشات و ارائه نتایج

۴. پیاده سازی و ارائه نتایج

در فصل‌های گذشته با جستجوی بصری در سیستم بازیابی تصویر مبتنی بر محتوا آشنا شدیم و اهداف و کاربردهای این سیستم‌ها در حوزه‌های پژوهشی و صنعتی و حتی زندگی روزمره را مرور نمودیم و همچنین آخرین پژوهش‌ها در این زمینه را بیان کردیم. در ادامه به پیشنهاد روشی نوین برای انجام جستجوی بصری پرداختیم و مراحل مختلف آن را گام به گام معرفی نمودیم.

به طور خلاصه در سیستم پیشنهادی ابتدا با استفاده از شبکه عصبی مصنوعی ResNet101 بردارهای ویژگی مربوط به تصاویر در مجموعه‌های داده INRIA Holidays و Flickr1M استخراج می‌شود. سپس با استفاده از الگوریتم بیان شده در فصل قبل، نخست برای نادیده گرفته نشدن عناصر منفی، تبدیل CReLU بر بردار ویژگی اعمال می‌شود. سپس نمایش اولیه مبتنی بر جایگشت به ازای هر یک از بردارهای ویژگی ایجاد شده و معکوس آن به عنوان بردار جایگشت در نظر گرفته می‌شود. پس از آن، نمایش متن جایگزین مربوط به هر بردار محاسبه و برای ایجاد و تکمیل ایندکس، به الاستیک سرچ فرستاده می‌شود. در بخش انتهایی، جستجو برای یافتن مشابه‌ترین نمایش‌ها به نمایش متنی تصویر (تصاویر) پرس و جو انجام شده و پاسخ‌های رتبه بندی شده در قالب نتایج نهایی، بازگردانده و ارزیابی می‌شوند. مهم‌ترین هدف این سیستم بهینه سازی و اجرای طراحی ساختاری با کارایی مناسب به خصوص در بخش ایندکس و همچنین بخش جستجو خواهد بود تا بتوان عکس‌هایی با بیشترین شباهت به پرس و جو را از میان انبوهی از تصاویر بازیابی نمود.

در فصل گذشته به مشخصات محیط پیاده سازی و روش مربوطه و همچنین مجموعه داده‌های مورد استفاده پرداختیم. در این فصل، بیشتر به جزئیات اجرا و توصیف روش‌های آزمایش شده مربوط به بخش نمایه سازی (indexing) و جستجو (search) در رویکرد پیشنهادی معرفی شده می‌پردازیم و نتایج حاصل از این آزمایش‌ها در ادامه گزارش می‌شوند. لازم به ذکر است که تمامی پیاده سازی‌ها به زبان پایتون انجام شده است.

هدف نهایی از ارائه این فصل بررسی نتایج حاصل از اجرای آزمایشات در این پروژه با استفاده از الاستیک سرچ به صورت توزیع شده بر مجموعه داده مقیاس بزرگی همچون Flickr1M خواهد بود.

۴-۱. مشخصات سیستم مورد آزمایش

پیاده سازی روش پیشنهادی در محیط کامپیوتری با مشخصات زیر صورت پذیرفته است:

نوع سیستم عامل: ویندوز ۱۰ (64-bit operating system)

مشخصات پردازنده: Intel(R) Core(TM) i3-2350M CPU @ 2.30GHz 2.30 GHz

میزان حافظه RAM: 8.00 GB

نسخه زبان پایتون: Python 3.10.8

نسخه الاستیک سرچ: Elasticsearch 8.8.0

۴-۲. مجموعه داده

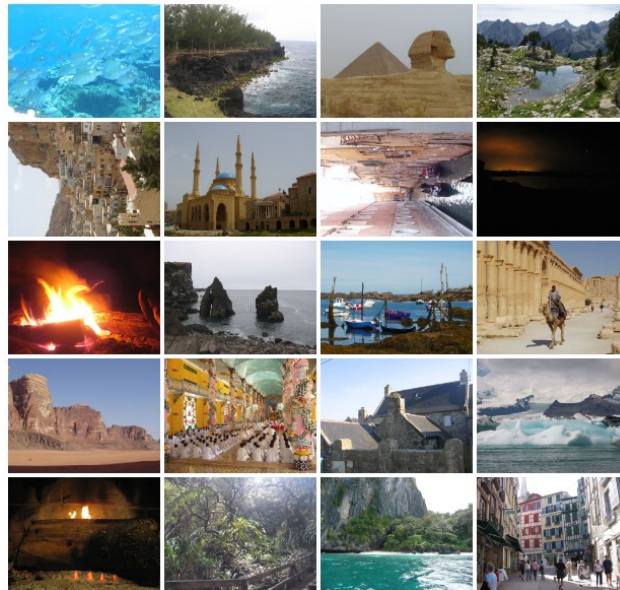
در این پروژه عملکرد روش پیشنهادی را با استفاده از ۲ مجموعه داده ارزیابی می‌کنیم. نخستین مجموعه داده، به نام تعطیلات برای ارزیابی دقت روش پیاده سازی شده و دومین مجموعه با عنوان Flickr1M برای ارزیابی سرعت روش پیشنهادی در نظر گرفته شده است. در ادامه به شرح خواص و مشخصات هر یک از این مجموعه‌های داده می‌پردازیم.

مجموعه داده تعطیلات (INRIA Holidays dataset)

این مجموعه، یکی از مجموعه داده‌های استاندارد برای بررسی کارایی سیستم‌های جستجوی بصری می‌باشد که شامل طیف وسیعی از انواع صحنه‌ها (طبیعی، مصنوعی، آب و آتش و غیره) است و نیز مواردی که عمدتاً به جهت آزمایش استحکام در برابر حملات مختلف همچون چرخش، تغییر زاویه دید و روشنایی، تاری، و موارد مشابه دیگر گردآوری شده‌اند.

تعداد تصاویر موجود در این مجموعه ۱۴۹۱ تصویر است و این مجموعه داده ۵۰۰ گروه تصویری را در بر می‌گیرد که هر کدام نمایش دهنده یک صحنه یا شی مجزا هستند و تعداد تصاویر در گروه‌های مختلف می‌تواند باهم متفاوت باشند. (Amato et al., 2018; Jegou et al., 2008)

نمونه‌ای از تصاویر موجود در این مجموعه را می‌توانید در شکل ۴-۱ مشاهده نمایید.



شکل ۴-۱: نمونه تصاویر مجموعه داده تعطیلات

یکی دیگر از انگیزه‌های انتخاب این مجموعه داده اینست که در آن از یک پروتکل ارزیابی بخصوص برای سنجش دقت در سیستم پیاده سازی شده استفاده می‌شود که نشان دهنده طراحی ویژه آن برای عمل جستجوی بصری است. این شیوه به گونه‌ای در نظر گرفته شده است که در هر گروه تصویری (یکی از ۵۰۰ گروه تصویری طراحی شده) تصویر نخست به عنوان پرس و جو قرار داده می‌شود و سایر تصاویری که در گروه وجود دارند را به عنوان پاسخ به آن پرس و جو و بر اساس میزان شباهت رتبه بندی می‌کند. در انتها میزان دقت سیستم جستجوی پیشنهادی به وسیله مقایسه نتایج با پاسخ‌های معیار ارزیابی می‌گردد. این بسته ارزیابی که از نرم افزار آکسفورد الهام گرفته شده است، عملکرد سیستم معین را به وسیله معیار mAP یا میانگین متوسط دقت اندازه گیری می‌کند.

مجموعه داده فلیکر ۱ میلیون (Flickr1M)

این مجموعه داده یکی از مجموعه داده‌های در مقیاس عظیم در حوزه بازیابی تصویر مبتنی بر محتوا به شمار می‌رود که شامل ۱ میلیون داده به صورت تصویر و ابرداده‌های مرتبط است که از پلتفرم محبوب اشتراک گذاری عکس، فلیکر جمع آوری شده‌اند. هدف از ایجاد این مجموعه پشتیبانی از تحقیق و توسعه در زمینه بینایی کامپیوتر، به ویژه برای کارهایی مانند طبقه بندی تصویر، تشخیص اشیاء، و بازیابی تصویر است که منبعی غنی از اطلاعات برای تحقیق و تحلیل را فراهم آورده است.

مجموعه داده Flickr1M طیف گسترده‌ای از موضوعات را پوشش می‌دهد و حاوی تصاویری از ژانرهای مختلف مانند طبیعت، حیوانات، اشیاء، افراد و مکان‌های دیدنی است. تصاویر از نظر محتوا، سبک و کیفیت متنوع هستند و طیف گسترده‌ای از صحنه‌ها و مفاهیم بصری را به تصویر می‌کشند. تصاویر با ابر داده مانند توضیحات متنی، برچسب‌ها و حاشیه نویسی‌های تولید شده توسط کاربر همراه هستند. این ابر داده را می‌توان برای کارهایی مانند حاشیه نویسی خودکار، درک معنایی و سیستم‌های توصیه استفاده کرد.

نمونه‌ای از تصاویر موجود در این مجموعه را می‌توانید در شکل ۴-۲ مشاهده نمایید.



by [Silke Gerstenkorn](#)



by [Dave Wild](#)



by [Hugo A.B. Olivas](#)



by [Martin P. Szymczak](#)



by [Mani Babbar](#)



by [Lee Otis](#)

شکل ۴-۲: نمونه تصاویر مجموعه داده فلیکر ۱ میلیون

از آن جایی که تصاویر موجود در مجموعه داده Flickr1M مستقیماً از پلتفرم Flickr جمع آوری شده‌اند، ممکن است برخی از تصاویر دارای مجوزها و حقوق استفاده مختلف باشند و شایان ذکر است که محققان هنگام استفاده از این مجموعه داده برای کار خود باید به این محدودیت‌های مجوز توجه داشته باشند.

نکته دیگری که در رابطه با مجموعه داده Flickr وجود دارد اینست که باید توجه داشته باشیم که مجموعه داده‌های Flickr دیگری نیز در دسترس هستند، مانند مجموعه داده Flickr30k و Flickr25k که معمولاً برای وظایف توصیف تصویر مبتنی بر جمله استفاده می‌شود. همچنین باید توجه داشت که در مجموعه داده Flickr1M تعداد ۲۵۰۰۰ تصویر اول این مجموعه مربوط به مجموعه MIRFLICKR-25k است. (Huiskes & Lew, 2008)

۳-۴. معیارهای ارزیابی

به منظور تحلیل و بررسی کارکرد روش پیشنهادی برای بازیابی تصویر مبتنی بر محتوا، بهترین راه بهره گیری از معیارهایی رایجی است که به طور گسترده برای اندازه گیری دقت و سرعت در تحقیقات گوناگون استفاده می شود. همان طور که هدف اصلی پژوهش پیش رو، برقراری تعادل قابل قبول در زمینه مفهوم سرعت و دقت معرفی شده بود و به منظور بررسی دقیق تر روش بیان شده، از معیار mAP یا میانگین متوسط دقت^۱ برای اندازه گیری درستی نتایج استفاده شده است و جهت مقایسه سرعت از اندازه گیری مجموع و میانگین زمان اجرای ماژول های مختلف سیستم بر حسب ثانیه برای پرس و جوها استفاده می شود.

توجه به این نکته از اهمیت ویژه ای برخوردار است که همانطور که در نرم افزار ارزیابی آکسفورد تعریف شده است و این بسته ارزیابی مجموعه داده نیز الهام گرفته از همان نرم افزار است، هنگام محاسبه mAP، تصاویر پرس و جو به عنوان مثبت صحیح^۲ محاسبه نمی شوند (و البته همچنین نه به عنوان مثبت کاذب^۳) بلکه تصور می شود که آنها تصاویر "ناخواسته"^۴ هستند. برای معرفی معیار میانگین متوسط دقت ابتدا نیاز به تعریف مفاهیم اولیه دیگری است که در ادامه به مرور آن ها می پردازیم:

معیار مثبت صحیح (TP, True Positive): در این جا به معنای تعداد تصاویر بازیابی شده و مرتبط است.

معیار مثبت کاذب (FP, False Positive): در این جا به معنای تعداد تصاویر بازیابی شده و نامرتب است.

معیار منفی صحیح (TN, True Negative): در این جا به معنای تعداد تصاویر بازیابی نشده و نامرتب است.

معیار منفی کاذب (FN, False Negative): در این جا به معنای تعداد تصاویر بازیابی نشده و مرتبط است.

معیار دقت (Precision): این معیار نشان می دهد به چه میزان می توان به خروجی اعتماد کرد یعنی چه میزان از نتایج بازیابی شده واقعاً مرتبط (نسبت اسناد بازیابی شده مرتبط با درخواست کاربر به کل اسناد بازیابی شده) هستند. روش محاسب این معیار به این صورت است:

^۱ mean Average Precision

^۲ True Positive

^۳ False Positive

^۴ junk

$$Precision = \frac{TP}{TP+FP} \quad (۱-۴)$$

معیار فراخوانی (Recall): این معیار دقت بازیابی را نسبت به کل مجموعه داده می‌سنجد و به عبارت دیگر نشان می‌دهد چه مقدار از تصاویر مرتبط به درستی بازیابی شده‌اند. (یعنی نسبت تصاویر مرتبط بازیابی شده به کل تصاویر مرتبط)

$$Recall = \frac{TP}{TP+FN} \quad (۲-۴)$$

به‌طور پیش‌فرض دقت، تمام اسناد بازیابی‌شده را در نظر می‌گیرد، اما با این حال، می‌توان آن را در تعداد معینی از اسناد بازیابی شده نیز ارزیابی کرد، که معمولاً به عنوان رتبه‌بندی برش^۱ شناخته می‌شود، که در آن ارزیابی تنها با در نظر گرفتن بالاترین پاسخ‌ها انجام می‌شود. این رویکرد اندازه‌گیری دقت در k یا $P@K$ نامیده می‌شود.

بازیابی اطلاعات عموماً به این معنی است که کاربر یک پرس و جو را به یک پایگاه داده ارائه کند و اطلاعاتی بسیار شبیه به پرس و جو را بازیابی کند. برای اندازه‌گیری $p@k$ در هر مرحله K فقط تعداد K تصویر بازیابی شده و رتبه‌بندی شده را برای یک پرس و جو در نظر می‌گیریم و میزان دقت را برای این مجموعه محاسبه می‌کنیم. به عبارت دیگر دقت $p@k$ برای یک جستجوی خاص، میزان نتایج مثبت صحیح را در بین k نتیجه برگردانده شده توسط الگوریتم جستجو محاسبه می‌کند. به طور ساده، این معیار نسبت تعداد نتایج صحیح (مثبت صحیح) به تمامی نتایج برگردانده شده تا رتبه k را به ما نشان می‌دهد.

برای مثال مجموعه بازیابی و رتبه‌بندی شده زیر را در نظر بگیرید (برای $k \geq 6$ پاسخ مثبت صحیح دیگری وجود ندارد).

✓	✗	✗	✓	✓	✗	...	✗
0	1	2	3	4	5		n

معیار $p@k$ به ازای k های مختلف به این صورت محاسبه می‌شود:

$$P@1 = 1/1 = 1$$

$$P@2 = 1/2 = 0.5$$

$$P@3 = 1/3 = 0.33$$

$$P@4 = 2/4 = 0.5$$

$$P@5 = 3/5 = 0.6$$

$$P@n = 3/n$$

¹ cut-off rank

همچنین $rel@k$ یک تابع ارتباط و به معنای فراخوانی در رتبه k است. تابع ارتباط یک تابع نشانگر است که اگر پاسخ بازیابی شده در رتبه k مرتبط باشد مقدار تابع برابر با ۱ و در غیر این صورت برابر با ۰ است.

متوسط دقت (AP): اکنون می‌توانیم با معرفی میانگین دقت^۱ به محاسبات ادامه دهیم. برای اندازه‌گیری این معیار از فرمول زیر استفاده کنیم:

$$AP@n = \frac{1}{GTP} \sum_k^n P@k \times rel@k \quad (۳-۴)$$

که در آن GTP به تعداد کل موارد مثبت حقیقت پایه^۲، n به تعداد کل اسناد بازیابی و رتبه بندی شده، $P@k$ به دقت و $rel@k$ به فراخوانی در رتبه k اشاره دارد. موارد مثبت حقیقت پایه داده‌هایی هستند که به عنوان مثبت برچسب گذاری شده‌اند و به عبارت دیگر اسناد مربوطه را تعریف می‌کنند. متوسط دقت به ازای هر پرس و جو می‌تواند اندازه‌گیری شود.

تعریف میانگین متوسط دقت (mAP): در انجام فرآیند بازیابی، برای هر یک از پرس و جو ها می‌توانیم AP مربوطه را محاسبه کنیم. همچنین یک کاربر می‌تواند به اندازه دلخواه خود پرس و جو داشته باشد. mAP صرفاً میانگین AP تمام پرس و جو هایی است که برای ارزیابی عملکرد سیستم مربوطه انجام شده است و فرمول آن مطابق زیر است:

$$mAP = \frac{1}{N} \sum_{i=0}^N AP_i \quad (۴-۴)$$

۴-۴. آزمایش ها در مجموعه داده تعطیلات

در ابتدا برای حصول اطمینان از صحت پیاده سازی و آزمایش کارایی روش‌ها مجموعه داده تعطیلات را انتخاب نمودیم زیرا این مجموعه داده از نظر تعداد تصاویر در مقیاس پایین‌تری قرار دارد و همچنین دارای سازوکار های مورد نیاز و استاندارد برای اندازه گیری میزان دقت mAP روش پیشنهادی است. در این مجموعه داده، ایده‌هایی را برای اجرا در بخش نمایه سازی توسط الستیک سرچ، پیاده سازی و مورد امتحان قرار گرفت و نتایج آن‌ها در ادامه بیان می‌شوند. گفتنی است مقدار دقت mAP در همه آزمایش های پیش رو با اندازه گیری میانگین AP مربوط به اجرای جستجوی ۵۰۰ کوئری تصویری در مجموعه داده INRIA Holidays dataset به دست آمده است.

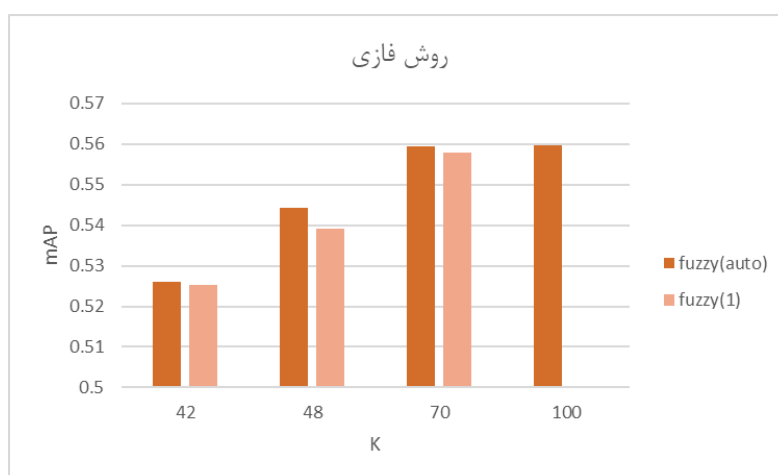
^۱ Average Precision

^۲ Ground Truth Positives

آزمایش اول:

در ابتدا به پیاده سازی و بررسی روش فازی در مجموعه داده INRIA Holidays dataset پرداختیم.

در این روش یعنی در پرس و جو ۱-۳: جستجو به روش فازی، مقدار fuzziness می تواند به صورت دلخواه تغییر کند. (برای توضیحات بیشتر به بخش ۳-۴ قسمت جستجوی فازی مراجعه نمایید) ما مقدار auto و مقدار 1 را مورد بررسی قرار دادیم که نتایج دقت (mAP) آن بر حسب K در نمودار ۱-۴ مشخص شده است:



نمودار ۱-۴: مقایسه تأثیر متغیر fuzziness در جستجوی فازی

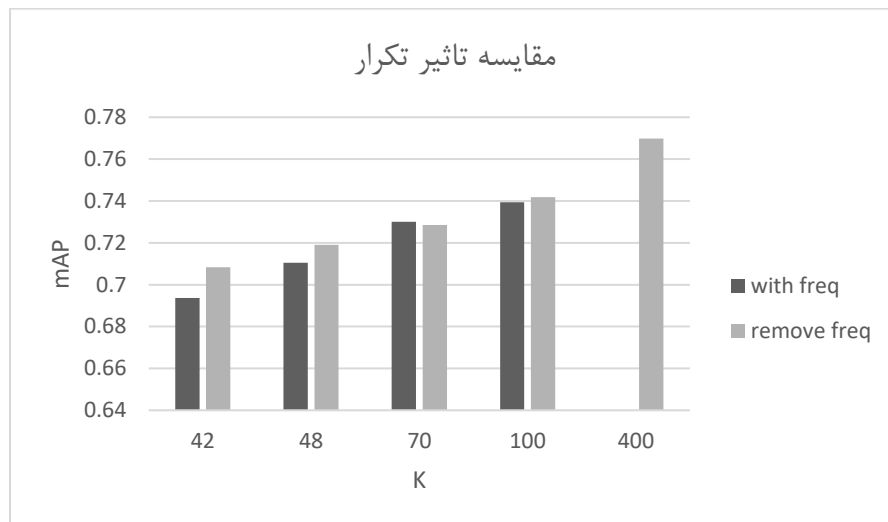
مقدار fuzziness(0) به آن دلیل در نظر گرفته نشده است که همان طور که پیش تر در توضیحات بخش ۳-۴ بیان شد، این مقدار برابر جستجو برای یافتن مطابقت کامل بین دو عبارت متنی در نظر گرفته می شود که در اینجا هدف ما نبوده و هدف اصلی یافتن مشابهت بین عبارات متنی جایگزین تصاویر است.

لازم به ذکر است دلیل آن که نمودار ۱-۴ در قسمت K=100 برای مقدار fuzzy(1) فاقد داده است اینست که محیط سخت افزاری پروژه فاقد نیازمندی های لازم برای محاسبه کامل نتیجه با این فرضیات بود و پاسخ با خطای محاسباتی Max clause count همراه شد و از آنجایی که برای برطرف نمودن این خطا، نیاز به تغییر محیط محاسباتی وجود داشت، تلاش ها در این راستا بی ثمر ماند. برطرف نمودن این موضوع و محاسبه مجدد نتایج به طور کامل، بخشی از اهداف آینده برای توسعه این پروژه خواهد بود.

با توجه به نمودار ۱-۴ نتیجه می گیریم که مقدار "auto" برای متغیر fuzziness در این بخش، با نتایج بهتری همراه بوده است و برای مقایسه های بعدی نیز این مقادیر مورد توجه قرار می گیرند.

آزمایش دوم:

با توجه به این که روش‌های ایده جستجوی عین عبارت و جستجو با حذف تکرار، ساختار ایندکس و کوئری مشابهی دارند و تنها در شیوه ذخیره داده متنی با یکدیگر متفاوت هستند (برای توضیحات بیشتر به بخش ۳-۴ مراجعه نمایید)، برای مقایسه دقت بین آن‌ها، به پیاده سازی و ارزیابی نتایج حاصل پرداختیم که در نمودار ۲-۴ نشان داده شده است:



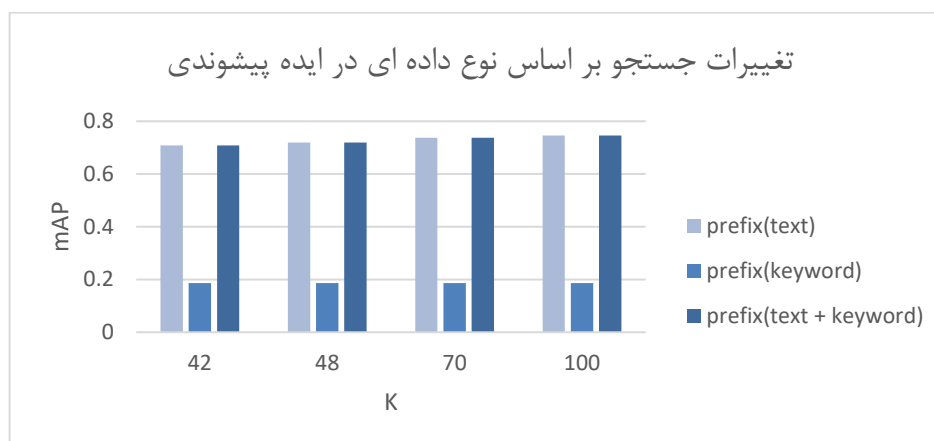
نمودار ۲-۴: مقایسه تأثیر تکرار در جستجوی ساده

علت کمبود داده در $K=400$ این نمودار نیز مشابه بخش فازی، خطای محاسباتی Max clause count بوده است که برطرف نمودن این موضوع و محاسبه مجدد نتایج به طور کامل، بخشی از اهداف آینده برای توسعه این پروژه خواهد بود.

مطابق نمودار ۲-۴ نتایج حاصل از ایده جستجو با حذف تکرار در $K=42, 48, 100, 400$ با دقت بالاتری همراه بوده است و پاسخ به دست آمده از ایده جستجوی عین عبارت (با تکرار) تنها در $K=70$ دقت بیشتری را کسب نموده است.

آزمایش سوم:

برای بررسی بیشتر درباره جستجو با در نظر گرفتن پیشنوندها، ابتدا تفاوت نتایج به دست آمده از ۲ نوع مختلف داده‌ای و کوئری های بیان شده متعلق به هرکدام را بررسی می‌کنیم؛ به عبارت دیگر در این آزمایش به بررسی تاثیر ذخیره سازی به صورت Multifields و تفاوت پرس و جو های ۱-۴-۳ و ۲-۴-۳ خواهیم پرداخت که یافته های به دست آمده در نمودار ۳-۴ نشان داده شده است:



نمودار ۲-۳-۴: مقایسه تفاوت نوع داده‌ای در جستجو با در نظر گرفتن پیشنوندها

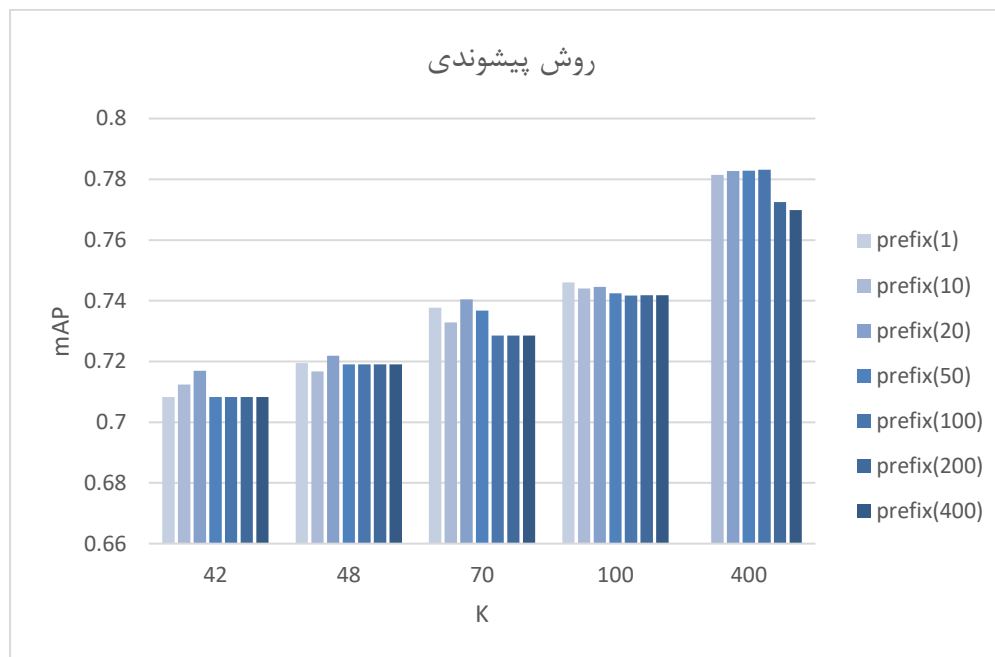
مطابق نمودار ۳-۴ با فرض تعریف نوع داده‌ای text نتایج به دست آمده از نظر دقت، از نوع داده‌ای keyword پیشی می‌گیرند و دلیل آن هم اینست که در حالت keyword سختگیرانه ترین نوع جستجو در نظر گرفته می‌شود و فقط رشته‌های متنی که نسبت به رشته متنی مورد جستجو، ترتیب کاملاً یکسان برای کلمات کلیدی منحصر به فرد دارند و بین آن‌ها مطابقت کامل وجود دارد، به عنوان جواب بازگردانده می‌شوند در حالی که ما به دنبال مشابه‌ترین پاسخ‌ها خواهیم بود و نه صرفاً پاسخ‌های دقیقاً یکسان.

همین طور برای حالت داده‌ای text + keyword نیز از آنجایی که مجموعه امتیازهای هر دو حالت در نظر گرفته می‌شود و مقدار امتیاز کسب شده توسط پاسخ‌های حالت text با اختلاف از حالت keyword بیشتر است، پس نتایج نهایی این بخش هم به نتایج حالت text نزدیک‌تر است.

لازم به ذکر است که با توجه به آنچه در رابطه با نمودار ۳-۴ بیان شد، در آزمایش‌های بعدی فقط از نوع داده‌ای text برای جستجو استفاده می‌نماییم.

آزمایش چهارم:

در سناریو آزمایشی بعدی، می‌خواهیم بررسی کنیم که در ایده جستجو با نظر گرفتن پیشوندها، اگر هر بار به جای اضافه کردن یک کلمه کلیدی، تعداد بیشتری از کلمات کلیدی را به مقادیر قبلی اضافه کنیم آیا نتایج مربوط به دقت تغییری خواهند کرد یا خیر؟ پس به جای افزایش هر بار یک کلمه کلیدی به دنباله قبلی، این تعداد را تغییر می‌دهیم و میزان افزایش را مقادیر ۱۰، ۲۰، ۵۰، ۱۰۰، ۲۰۰، ۴۰۰ قرار داده و دقت پاسخ‌های بازگردانده شده را اندازه‌گیری می‌کنیم که نتایج آن در نمودار ۴-۴ نشان داده شده است:



نمودار ۴-۴-۲: مقایسه تأثیر میزان افزایش پیشوندی

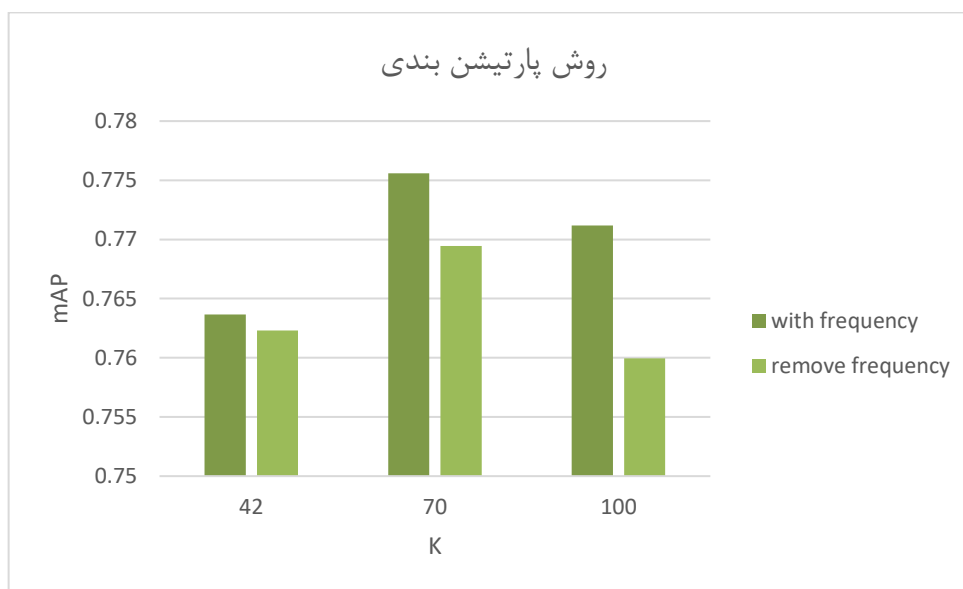
همان طور که در نمودار ۴-۴ نیز دیده می‌شود، برای $K=42, 48, 70$ میزان افزایش پیشوندی ۲۰ یعنی $\text{prefix}(20)$ مقدار mAP بالاتری را نسبت به بقیه کسب نموده است و برای $K=100$ بهترین مقدار دقت متعلق به میزان افزایش ۱ یعنی همان حالت اولیه $\text{prefix}(1)$ بوده است. همچنین بالاترین میزان دقت mAP در $K=400$ و مربوط به افزایش پیشوندی ۱۰۰ بوده است. به طور معمول با افزایش مقدار K دقت نتایج نیز (در اکثر روش‌ها) افزایش می‌یابد و دلیل آن هم مربوط به مقادیر بالاتر ویژگی‌های پایه است که در top-K در نظر گرفته می‌شود.

آزمایش پنجم:

در این سناریو به پیاده سازی و آزمایش ایده جستجو به روش پارتیشن بندی خواهیم پرداخت و ۲ شیوه برای ذخیره سازی و جستجو عبارت های متن جایگزین را از نظر دقت مورد مقایسه قرار خواهیم داد. گفتنی است تمامی آزمایش های این بخش را با $n=10$ انجام خواهیم داد (برای توضیحات بیشتر به بخش ۳-۴ قسمت جستجوی بر اساس پارتیشن بندی مراجعه نمایید)

ابتدا پس از استخراج بردار ویژگی و تقسیم آن به تعداد ۱۰ پارتیشن (قطعه)، برای هر قطعه به طور جداگانه الگوریتم تبدیل به نمایش متن جایگزین را اجرا می کنیم و مشابه ایده جستجو عین عبارت، نمایش متنی مربوط به هر قطعه را بدون ایجاد هیچ تغییری به تفکیک در فیلد های یک سند (مربوط به یک تصویر) ایندکس می کنیم و با استفاده از پرس و جو ۳-۵ جستجو را انجام داده و دقت را اندازه گیری می نماییم.

سپس عملیات گفته شده را بار دیگر و مشابه ایده جستجو با حذف تکرار، مجددا اجرا می کنیم و نمایش متنی مربوط به هر قطعه را با حذف کلمات کلیدی تکراری و تنها با حفظ ترتیب کلمات، ایندکس کرده و جستجو را نیز با حذف تکرار انجام می دهیم و دقت را محاسبه می کنیم. نتایج حاصل از این آزمایش در نمودار ۴-۵ ارائه شده است:

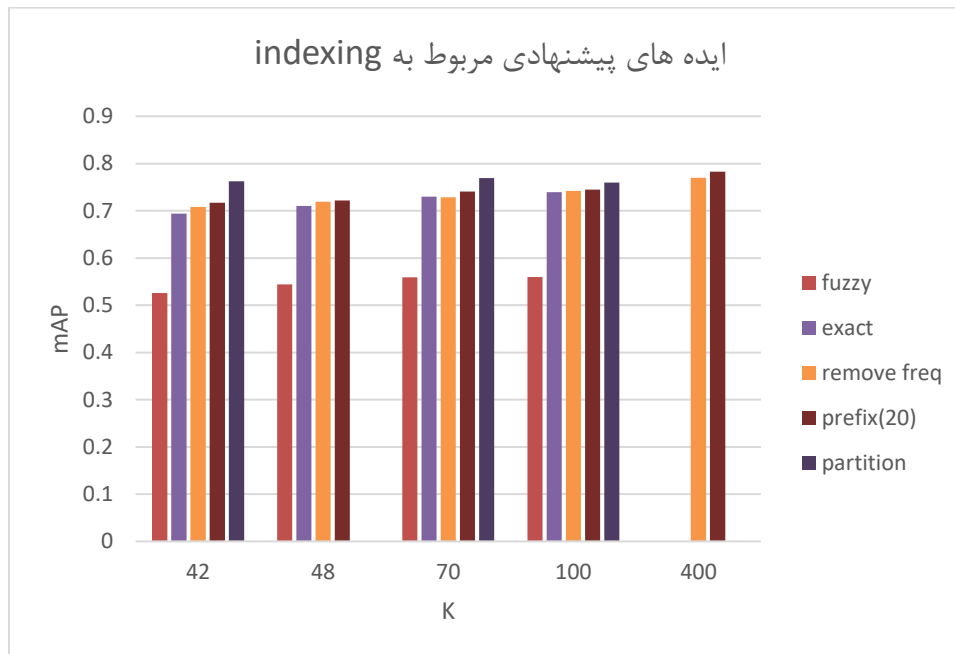


نمودار ۲-۵-۴: مقایسه تأثیر تکرار در روش پارتیشن بندی

مطابق نمودار ۴-۵ پاسخ های حاصل از جستجوی عین عبارت های هر قطعه به لحاظ دقت در جایگاه بالاتری قرار می گیرند. بنابراین این شیوه ذخیره سازی در آزمایش های بعدی مورد توجه قرار می گیرد.

آزمایش ششم:

در پایان دقت همه ایده‌های ارائه شده را بر اساس داده‌های موجود با یکدیگر مقایسه نمودیم و در نمودار ۴-۶ نتایج آن را ارائه نمودیم.



نمودار ۲-۶-۴: مقایسه کلی روش‌ها

در نمودار ۴-۶ به طور کلی بالاترین میزان mAP مربوط به روش پارتیشن بندی و کمترین میزان مربوط به روش جستجوی فازی مشاهده شده است. همچنین به طور خاص بالاترین دقت مربوط به روش پیشوندی با در نظر گرفتن ۲۰ پیشوند در هر گام و با $K=400$ ، مقدار ۰.۷۸۲۶۷ است و کمترین میزان دقت در روش فازی با $K=42$ و با مقدار ۰.۵۲۵۹۵ گزارش شده است. بدیهی است که با افزایش مقدار پارامتر K ، تعداد ویژگی‌های پایه در نظر گرفته شده در هر روش، افزایش می‌یابد و به همین ترتیب دقت نیز بالاتر می‌رود.

علت این کاهش دقت در جستجوی فازی را می‌توان مربوط به ماهیت جستجوی فازی در تشخیص شباهت‌های عبارات بدون در نظر گرفتن تطابق کامل دانست؛ برای مثال دو کلمه کلیدی T10 و T26 با تعیین میزان fuzziness=1 مشابه تلقی شده و به عنوان جواب برای جستجو در نظر گرفته می‌شوند در حالی که این پاسخ مدنظر ما نیست و همین سبب پایین آمدن میزان دقت در نتایج نهایی می‌شود.

دلیل افزایش دقت در ایده جستجو بر اساس پارتیشن بندی نیز آنست که در این روش به جای صرفاً در نظر گرفتن ویژگی‌های پایه در کل بردار، (ویژگی‌هایی که مقدار بیشتری دارند و تأثیر آن‌ها در محاسبه شباهت بیشتر است) بالاترین ویژگی‌های پایه در هر قطعه به صورت جداگانه در نظر

گرفته می‌شود و با توجه به این موضوع، می‌توان گفت که حتی برخی از ویژگی‌های پایه که مقدار کمتری نسبت به بقیه داشتند و ارزش آن‌ها در محاسبه شباهت تقریباً نادیده گرفته می‌شد، در روش پارتیشن بندی مورد توجه قرار می‌گیرند و همین امر موجب بالا رفتن دقت در این روش بوده است.

مقایسه با روش های دیگر

در نمودار ۴-۶ بالاترین میزان دقت mAP گزارش شده مربوط به روش پارتیشن بندی در $K=70$ با مقدار 0.76946 (یعنی حدودا 76.9%) و همچنین روش در نظر گرفتن پیشوندها (برای حالت 20 پیشوند) در $K=400$ با مقدار 0.78267 (یعنی حدودا 78.3%) است.

در جدول ۴-۱: مقایسه با روش های دیگر، تعدادی از رویکرد هایی که از مجموعه داده تعطیلات برای ارزیابی عملکرد سیستم بازیابی خود استفاده کرده اند، جمع آوری شده و میزان دقت روش های پیشنهادی در این پژوهش، با آن ها مقایسه شده است.

جدول ۴-۱: مقایسه با روش های دیگر

نام روش	mAP (%)
(Perronnin et al., 2010)	70
(Douze et al., 2011)	69.9
(Shen et al., 2012)	76.2
(Torii et al., 2013)	74.95
(Do et al., 2015)	75.8
(T. Zhang et al., 2015)	64.4
روش پیشنهادی (پارتیشن بندی)	76.9
روش پیشنهادی (پیشوندی)	78.3

۴-۵. آزمایش در مجموعه داده فلیکر

در این بخش به شرح سناریوهای آزمایشی در مجموعه داده فلیکر می پردازیم تا مقیاس پذیری و توزیع شدگی سیستم پیشنهادی را مورد بررسی قرار دهیم. از آن جایی که با توجه به آزمایش های قبلی، ایده جستجو بر اساس پارتیشن بندی، نتایج بهتری را ثبت نمود، در این قسمت نیز این روش را مورد آزمایش قرار می دهیم و نتایج آن را بیان خواهیم نمود.

اگر بار دیگر به نمودار ۴-۶ توجه کنیم، مشاهده می کنیم که دو روش پارتیشن بندی در $K=42$ و روش حذف تکرار در $K=400$ دقت تقریباً یکسانی را ثبت نموده اند. در روش پارتیشن بندی، بردار ویژگی با ابعاد ۴۰۹۶ به ۱۰ پارتیشن تقسیم می شود که هر کدام حدوداً ابعاد ۴۱۰ دارند. در هر پارتیشن تقریباً ۴۰ ویژگی (ابعاد) پایه برای تبدیل شدن به نمایش متنی برگزیده می شوند که با توجه به داشتن ۱۰ پارتیشن به صورت کلی، می توان آن را معادل با در نظر گرفتن ۴۰۰ ویژگی پایه برای هر بردار در نظر گرفت و بدین صورت این یکسانی دقت در نمودار توجیه می شود. در ادامه می خواهیم بررسی کنیم که این دو روش به لحاظ سرعت در پاسخ دهی به پرس و جوها چه نسبتی با یکدیگر دارند؟

در ابتدا تعداد ۱۰۰ هزار داده تصویر ابتدایی در مجموعه داده فلیکر ۱ میلیون را جدا می کنیم و مطابق ایده روش پارتیشن بندی با $K=42$ و تعداد ۱۰ پارتیشن، نمایش های متنی مربوطه را به دست می آوریم و در ایندکس معرفی شده در الاستیک سرچ، ذخیره می کنیم. سپس به صورت کاملاً تصادفی ۱۰۰۰ داده از این ۱۰۰ هزار سند را انتخاب می کنیم و به عنوان کوئری به سیستم می دهیم و مدت زمان مربوط به هر کوئری را اندازه گیری نموده و در پایان میانگین همه آن ها را محاسبه می کنیم.

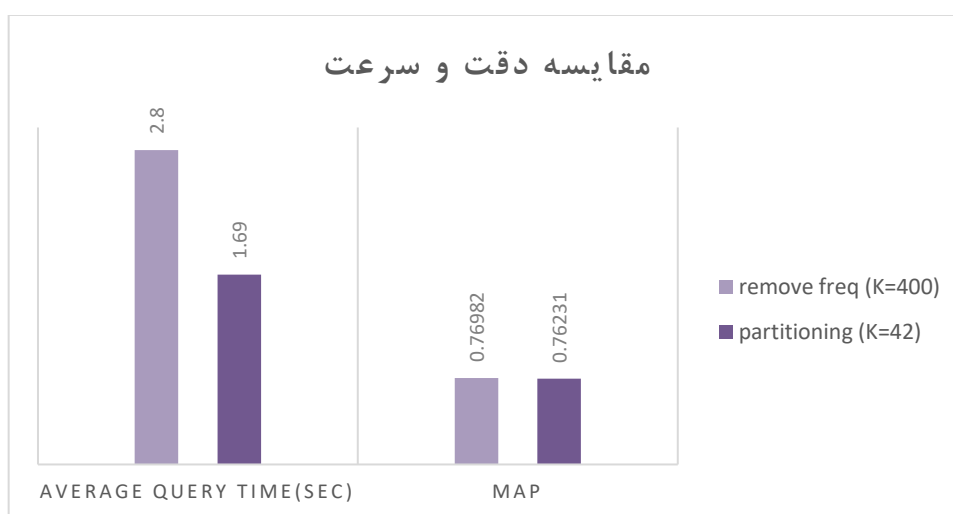
در ادامه نوبت به روش نمایش و جستجو ساده با حذف تکرار می رسد. به طور مشابه برای ۱۰۰ هزار تصویر ابتدایی مجموعه داده فلیکر بردار های ویژگی را استخراج کرده و سپس با $K=400$ نمایش متنی مربوط به هر بردار را به دست می آوریم و در ایندکس جداگانه ای در الاستیک سرچ ذخیره می کنیم. سپس مجدداً ۱۰۰۰ داده را به صورت کاملاً تصادفی انتخاب کرده و جستجو را با این داده ها به انجام می رسانیم و مدت زمان پاسخگویی به هر کوئری را در نظر گرفته و در پایان میانگین همه آن ها را به دست می آوریم.

جدول ۴-۱ مدت زمان گذشته برای کل فرآیند ذخیره سازی، کل فرآیند جستجو و نیز میانگین زمان جستجو برای هر کوئری را نشان می دهد.

جدول ۲-۴: مقایسه زمان برای دو روش پارتیشن بندی و حذف تکرار

	Indexing time(sec)	whole search time(sec)	Average Query Time(sec)
remove frequency (K=400)	3191.23	2805.18	2.8
partitioning (K=42)	2287.82	1706.94	1.69

همان طور که در جدول ۲-۴ مشخص شده است روش پارتیشن بندی چه در بخش indexing و چه در بخش searching مدت زمان کوتاه تری را ثبت نموده است که این می تواند به دلیل برخی از عملکردهای خودکار موازی سازی در الاستیک سرچ جهت ایجاد قابلیت مقیاس پذیری و توزیع شدگی در سیستم موردنظر باشد و با استفاده درست از آن می توان نتایج بهتری را ثبت نمود. بنابراین خلاصه ای از مقایسه دقت و سرعت روش پارتیشن بندی با $K=42$ و روش جستجوی ساده با حذف تکرار با $K=400$ را در نمودار ۷-۴ می بینید.



نمودار ۷-۴: مقایسه دقت و سرعت روش پارتیشن بندی و روش حذف تکرار

از مشاهده نمودار ۷-۴ در می یابیم که روش پارتیشن بندی با وجود ثبت دقت تقریباً مشابه با روش جستجو ساده با حذف تکرار، متوسط زمان پاسخ به کوئری بسیار پایین تری دارد و توانسته است این زمان را تا حدود ۳۹/۶۴ درصد کاهش دهد. بنابراین می توان اظهار نمود که سیستم پیشنهادی با رویکرد پارتیشن بندی برای ذخیره سازی و جستجو، قادر است با حفظ دقت فرآیند بازیابی تصویر بر اساس محتوا، تا حد امکان سرعت جستجو را نیز افزایش دهد و مهم ترین هدف این پژوهش را برآورده سازد.

۴-۶. جمع بندی

در این فصل ابتدا به معرفی مجموعه داده های مورد استفاده برای آزمایش ایده های پیشنهادی پرداختیم و به طور کامل معیارهای ارزیابی مورد استفاده در آزمایش ها را شرح دادیم. سپس هر یک از ایده های پیشنهادی را پیاده سازی کردیم و برای هر کدام از آن ها به طور جداگانه آزمون هایی برای ارزیابی عملکرد بهینه دقت و سرعت در نظر گرفتیم و نتایج آن ها را ارائه کردیم. همچنین به مقایسه همه روش ها با یکدیگر پرداختیم و بهترین ایده را با توجه به نتایج به دست آمده، معرفی نمودیم.

در فصل بعدی و نهایی این پایان نامه، به جمع بندی کلی و نتیجه گیری این پروژه خواهیم پرداخت.

فصل پنجم

جمع بندی

۵. جمع بندی

جستجوی بصری در سیستم بازیابی تصویر مبتنی بر محتوا، یک راه حل کارآمد برای پیدا کردن تصاویر دلخواه از میان داده‌های تصویری در مقیاس بسیار بزرگ است. امروزه با گسترش روش‌های ثبت و تولید عکس‌ها، نگهداری و جستجو در میان تصاویر به چالشی جدی در فضای صنعتی و دانشگاهی تبدیل شده است. پژوهش‌های اخیر، روش‌های گوناگونی برای رویارویی با چالش‌های به وجود آمده حاصل از افزایش تعداد و ابعاد داده‌ها پیشنهاد داده‌اند اما تحقیقات در این حوزه همچنان ادامه دارد. استفاده از فناوری‌های موجود همچون موتورهای جستجو مبتنی بر متن مانند Elasticsearch برای سرعت بخشیدن به فرآیند جستجو با حفظ تعادل میان سرعت و دقت، یکی از راهکارهای ارائه شده در این زمینه است که ما در این پژوهش از آن استفاده کردیم.

در این پروژه، به طور خلاصه، ابتدا بردار ویژگی را از تصاویر در مجموعه داده‌های معرفی شده، استخراج کردیم و با بهره‌گیری از رویکردهای نمایه سازی مبتنی بر جایگشت و در نهایت ایجاد نمایش متنی برای تصویر و ذخیره آن در Elasticsearch توانستیم عملیات جستجوی بصری برای یافتن تصویر دلخواه از مجموعه داده‌ای بزرگ را اجرا کنیم و با ارائه پیشنهادهایی این فرآیند را بهبود بخشیم.

در این فصل به جمع بندی پژوهش انجام شده خواهیم پرداخت و نتیجه گیری کلی بر اساس روش پیشنهادی و نتایج به دست آمده را بیان خواهیم کرد. همچنین در ادامه مسیرهای پژوهشی آینده را مرور می‌کنیم و پیشنهادهای خود برای ایجاد بهبودهای ممکن را ارائه می‌نماییم.

۵-۱. نتیجه گیری

در این پژوهش رویکردهای مختلف موجود برای بازیابی تصویر مبتنی بر محتوا مورد بررسی قرار گرفت. سپس روش پیشنهادی در این حوزه تعریف شده و جزییات آن به طور کامل معرفی شدند. در ادامه ایده‌های برای بهبود نمایه سازی و جستجو در سیستم ارائه شده بیان گردید و آزمایش‌هایی برای سنجش میزان دقت و سرعت و نیز قابلیت مقیاس پذیری آن‌ها انجام شد که نتایج آن‌ها در فصل گذشته ارائه داده شد. با توجه به آخرین یافته‌ها، روش پارتیشن بندی برای ایجاد ایندکس و جستجو در الستیک سرچ از باقی روش‌های ارائه شده نتایج بهتری را ثبت نمود.

۵-۲. پژوهش‌های آینده

در اینجا به عنوان بخش پایانی این پژوهش مهم‌ترین مسیرهای تحقیقاتی آینده را تعیین می‌کنیم که بدین ترتیب هستند:

- تغییر در بخش استخراج ویژگی و استفاده از شبکه‌های عصبی CNN با ساختارهای متفاوت و جدیدتر که برای عمل جستجوی بصری تنظیم و مناسب سازی شده‌اند.
- استفاده از الگوریتم‌های دیگر برای تبدیل بردار ویژگی به نمایش متنی
- استفاده از نمایش‌های فشرده دیگر همچون کدهای باینری و کدهای PQ و تلفیق آن‌ها با روش پیشنهاد شده
- به کارگیری ایده‌های متفاوت در بخش نمایه سازی و جستجو با استفاده از الستیک سرچ
- برطرف نمودن خطاهای محاسباتی مربوط به کمبودهای سخت افزاری
- به کارگیری موتورهای جستجوی مبتنی بر متن دیگر مانند Solr و مقایسه آن‌ها با نتایج به دست آمده توسط الستیک سرچ
- اجرای آزمایش‌های دقت و سرعت بر روی تلفیق دو مجموعه داده Flickr1M + INRIA Holidays
- اجرای آزمایش‌ها بر روی مجموعه داده‌های متفاوت و در مقیاس بالاتر برای بررسی مقیاس پذیری و کارایی روش پیشنهاد شده

فهرست منابع و مآخذ

- ❖ A picture is worth a thousand words. (2023). In *Wikipedia*.
https://en.wikipedia.org/w/index.php?title=A_picture_is_worth_a_thousand_words&oldid=1144992814
- ❖ Amato, G., Bolettieri, P., Carrara, F., Falchi, F., & Gennaro, C. (2018). Large-scale image retrieval with elasticsearch. *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 925–928.
- ❖ Amato, G., Carrara, F., Falchi, F., Gennaro, C., & Vadicamo, L. (2020). Large-scale instance-level image retrieval. *Information Processing & Management*, 57(6), 102100.
- ❖ Amato, G., Falchi, F., Gennaro, C., & Vadicamo, L. (2016). Deep permutations: Deep convolutional neural networks and permutation-based indexing. *Similarity Search and Applications: 9th International Conference, SISAP 2016, Tokyo, Japan, October 24-26, 2016, Proceedings 9*, 93–106.
- ❖ Camli, G. (2020). *Visual Search: Finding similar images* [Master thesis, Universitat Politècnica de Catalunya]. <http://hdl.handle.net/2117/185623>
- ❖ Canziani, A., Paszke, A., & Culurciello, E. (2016). An analysis of deep neural network models for practical applications. *arXiv Preprint arXiv:1605.07678*.
- ❖ Chang, X., Ren, P., Xu, P., Li, Z., Chen, X., & Hauptmann, A. (2021). *A Comprehensive Survey of Scene Graphs: Generation and Application*.
<https://doi.org/10.48550/ARXIV.2104.01111>
- ❖ Cheng, S., Wang, L., & Du, A. (2019). An adaptive and asymmetric residual hash for fast image retrieval. *IEEE Access*, 7, 78942–78953.
- ❖ Curse of dimensionality. (2023). In *Wikipedia*.
https://en.wikipedia.org/w/index.php?title=Curse_of_dimensionality&oldid=1145413427
- ❖ Dawn, S., Tulsyan, S., Bhattarai, S., Gopal, S., & Saxena, V. (2020). An efficient approach to image indexing and retrieval using Haar cascade and perceptual similarity index. *2020 6th International Conference on Signal Processing and Communication (ICSC)*, 108–113.
- ❖ Degenhardt, J., Kallumadi, S., Porwal, U., & Trotman, A. (2021). Report on the SIGIR 2019 Workshop on eCommerce (ECOM19). *ACM SIGIR Forum*, 53(2), 11–19.
- ❖ Do, T.-T., Tran, Q. D., & Cheung, N.-M. (2015). FAemb: A function approximation-based embedding method for image retrieval. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3556–3564.
- ❖ Douze, M., Ramisa, A., & Schmid, C. (2011). Combining attributes and fisher vectors for efficient image retrieval. *CVPR 2011*, 745–752.

- ❖ Dubey, S. R. (2021). A decade survey of content based image retrieval using deep learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(5), 2687–2704.
- ❖ Gennaro, C., Amato, G., Bolettieri, P., & Savino, P. (2010). An approach to content-based image retrieval based on the lucene search engine library. *Research and Advanced Technology for Digital Libraries: 14th European Conference, ECDL 2010, Glasgow, UK, September 6-10, 2010. Proceedings 14*, 55–66.
- ❖ Ghosh, N., Agrawal, S., & Motwani, M. (2018). A survey of feature extraction for content-based image retrieval system. *Proceedings of International Conference on Recent Advancement on Computer and Communication: ICRAC 2017*, 305–313.
- ❖ Halavataya, K. (2020). Local feature descriptor indexing for image matching and object detection in real-time applications. *Pattern Recognition and Image Analysis*, 30, 16–21.
- ❖ Huiskes, M. J., & Lew, M. S. (2008). The MIR Flickr Retrieval Evaluation. *MIR '08: Proceedings of the 2008 ACM International Conference on Multimedia Information Retrieval*.
- ❖ Hussain, D. M., & Surendran, D. (2021). The efficient fast-response content-based image retrieval using spark and MapReduce model framework. *Journal of Ambient Intelligence and Humanized Computing*, 12, 4049–4056.
- ❖ Jegou, H., Douze, M., & Schmid, C. (2008). *Hamming embedding and weak geometry consistency for large scale image search-extended version*.
- ❖ Johnson, J., Douze, M., & Jégou, H. (2019). Billion-scale similarity search with gpus. *IEEE Transactions on Big Data*, 7(3), 535–547.
- ❖ Kalyankar, P. A., Mulani, A. O., Thigale, S. P., Chavhan, P. G., & Jadhav, M. M. (2022). Scalable face image retrieval using AESC technique. *Journal Of Algebraic Statistics*, 13(3), 173–176.
- ❖ Khawandi, S., Abdallah, F., & Ismail, A. (2019). A survey on image indexing and retrieval based on content based image. *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 222–225.
- ❖ Kulis, B., & Grauman, K. (2009). Kernelized locality-sensitive hashing for scalable image search. *2009 IEEE 12th International Conference on Computer Vision*, 2130–2137.
- ❖ Li, X., Yang, J., & Ma, J. (2021). Recent developments of content-based image retrieval (CBIR). *Neurocomputing*, 452, 675–689.
- ❖ Li, Y., Ma, J., & Zhang, Y. (2021). Image retrieval from remote sensing big data: A survey. *Information Fusion*, 67, 94–115.
- ❖ Li, Y., Wan, L., Fu, T., & Hu, W. (2019). Piecewise supervised deep hashing for image retrieval. *Multimedia Tools and Applications*, 78, 24431–24451.
- ❖ Liu, H., Wang, R., Shan, S., & Chen, X. (2016). Deep supervised hashing for fast image retrieval. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2064–2072.

- ❖ Matsui, Y., Uchida, Y., Jegou, H., & Satoh, S. (2018). *Paper A Survey of Product Quantization*. 6(1). <https://doi.org/10.3169/mta.6.2>
- ❖ Mezzoudj, S., Behloul, A., Seghir, R., & Saadna, Y. (2021). A parallel content-based image retrieval system using spark and tachyon frameworks. *Journal of King Saud University-Computer and Information Sciences*, 33(2), 141–149.
- ❖ Moghadam Charkari, N., & Shakibian, H. (2021). A Multilayered Complex Network Model for Image Retrieval. *International Journal of Information and Communication Technology Research*, 13(4), 36–42.
- ❖ Mu, C., Yang, B., & Yan, Z. (2019). An empirical comparison of FAISS and FENSHSES for nearest neighbor search in hamming space. *arXiv Preprint arXiv:1906.10095*.
- ❖ Mu, C., Zhao, J., Yang, G., Yang, B., & Yan, Z. (2019). *Fast and Exact Nearest Neighbor Search in Hamming Space on Full-Text Search Engines* (arXiv:1902.08498). arXiv. <http://arxiv.org/abs/1902.08498>
- ❖ Mu, C., Zhao, J., Yang, G., Zhang, J., & Yan, Z. (2018). Towards practical visual search engine within elasticsearch. *arXiv Preprint arXiv:1806.08896*.
- ❖ Nguyen, M.-D., Nguyen, B. T., & Gurrin, C. (2021). Graph-based indexing and retrieval of lifelog data. *MultiMedia Modeling: 27th International Conference, MMM 2021, Prague, Czech Republic, June 22–24, 2021, Proceedings, Part II 27*, 256–267.
- ❖ Perronnin, F., Liu, Y., Sánchez, J., & Poirier, H. (2010). Large-scale image retrieval with compressed fisher vectors. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 3384–3391.
- ❖ Sadeghi-Tehran, P., Angelov, P., Virlet, N., & Hawkesford, M. J. (2019). Scalable database indexing and fast image retrieval based on deep learning and hierarchically nested structure applied to remote sensing and plant biology. *Journal of Imaging*, 5(3), 33.
- ❖ Sagharichian, M., & Mirmarouf, M. (2021). *A new perspective of content-based image retrieval based on Elasticsearch*. 9.
- ❖ Shen, X., Lin, Z., Brandt, J., Avidan, S., & Wu, Y. (2012). Object retrieval and localization with spatially-constrained similarity measure and k-nn re-ranking. *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 3013–3020.
- ❖ Stanley, T., Vanjara, N., Pan, Y., Pirogova, E., Chakraborty, S., & Chaudhuri, A. (2020). SIR: Similar Image Retrieval for Product Search in E-Commerce. *Similarity Search and Applications: 13th International Conference, SISAP 2020, Copenhagen, Denmark, September 30–October 2, 2020, Proceedings 13*, 338–351.
- ❖ Torii, A., Sivic, J., Pajdla, T., & Okutomi, M. (2013). Visual place recognition with repetitive structures. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 883–890.
- ❖ Wan, J., Tang, S., Zhang, Y., Li, J., Wu, P., & Hoi, S. C. (2017). Hdidx: High-dimensional indexing for efficient approximate nearest neighbor search. *Neurocomputing*, 237, 401–404.

- ❖ Zhang, T., Qi, G.-J., Tang, J., & Wang, J. (2015). Sparse composite quantization. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4548–4556.
- ❖ Zhang, W., Wu, D., Zhou, Y., Li, B., Wang, W., & Meng, D. (2021). Binary neural network hashing for image retrieval. *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1318–1327.
- ❖ Zhang, Y., Pan, P., Zheng, Y., Zhao, K., Zhang, Y., Ren, X., & Jin, R. (2018). Visual search at alibaba. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 993–1001.
- ❖ Zhao, K., Pan, P., Zheng, Y., Zhang, Y., Wang, C., Zhang, Y., Xu, Y., & Jin, R. (2019). Large-Scale Visual Search with Binary Distributed Graph at Alibaba. *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2567–2575. <https://doi.org/10.1145/3357384.3357834>
- ❖ Zhou, W., Li, H., & Tian, Q. (2017). Recent advance in content-based image retrieval: A literature survey. *arXiv Preprint arXiv:1706.06064*.

❖ ساغریچیان، مسعود و میرمعروف زبینه، میررضا، ۱۴۰۰، ارائه دیدگاه جدید مشابهت یابی تصویر مبتنی بر محتوا بر پایه الستیک سرچ، چهارمین همایش بین المللی مهندسی فناوری اطلاعات، کامپیوتر و مخابرات ایران، تهران، ۱۲۵۹۸۱۸، <https://civilica.com/doc/1259818>

پیوست‌ها

توضیح: پیوست‌ها از نظر عملی مواردی هستند که نسبت به اصل پایان‌نامه/رساله حاشیه‌ای به نظر می‌رسند و نبود آن‌ها خللی در روند آن نداشته ولی وجود آن‌ها در انتهای پایان‌نامه/رساله برای خوانندگان تخصصی و علاقه‌مند می‌تواند مفید باشد. به عبارتی، پیوست شامل مطالبی است که تکمیل‌کننده برخی از اطلاعات نظری و عملی در متن اصلی پایان‌نامه/رساله می‌باشد که ترجیحاً به علت جلوگیری از طولانی شدن متن اصلی پایان‌نامه/رساله، باید خارج از متن اصلی یعنی پس از فهرست منابع در پایان‌نامه/رساله ارائه گردد.

هدف از ارائه پیوست، درک بهتر پایان‌نامه/رساله و حفظ یکپارچگی متن آن است. محتوای پیوست را نامه‌ها، نمونه‌بزارها، بررسی‌ها، نوشته‌های بیشتر، آمار، ارقام و مانند آن‌ها در بر دارد که به‌گونه‌ای در انجام پژوهش به‌کار رفته‌اند؛ ولی آوردن آن‌ها در متن اصلی نیاز نیست.

پیوست ۱. پیوست الف

در این پیوست جدول زمان‌های پاسخ به پرس و جو در آزمایش ششم در مجموعه داده تعطیلات آورده می‌شود:

	42	48	70	100	400
fuzzy(auto)	0.4653	1.0074	2.1640	7.1887	
exact	0.0433	0.0282	0.0323	0.0367	
remove freq	0.0212	0.0072	0.0068	0.0111	0.042
prefix(20)	0.0457	0.0379	0.0259	0.042	1.2819
partition	0.1013		0.2396	0.2754	

پیوست ۲. پیوست ب

در این پیوست نمونه ای از پاسخ های ارائه شده توسط سیستم پیشنهادی در این پژوهش برای تعدادی از پرس و جوها در مجموعه تعطیلات آورده می شود. روش به کار رفته برای فرآیند نمایه سازی و جستجو در الستیک سرچ، روش در نظر گرفتن پیشوند ها (۲۰ پیشوند در هر گام) و با پارامتر $K=400$ است. تعداد اصلی پرس و جو های مشخص شده ۵۰۰ کوئری است و برای هر کوئری در هر درخواست جستجو، ۱۰ جواب که شامل عنوان تصاویر و نمایش متنی جایگزین برای آن تصاویر بازگردانده می شود. نمونه های این پیوست به صورت تصویری از پاسخ به پرس و جو های تعیین شده برای درک بهتر و شهودی دقت عملکرد سیستم آورده شده است.



واژه‌نامه توصیفی

تعریف اصطلاحات فنی و واژگان کلیدی

در این قسمت بخشی از واژگانی که در این پژوهش نقش اساسی دارند و بیشتر مورد تکرار قرار می‌گیرند، به صورت خلاصه تعریف می‌گردند.

جستجوی بصری و بازیابی تصویر بر اساس محتوا

جستجوی بصری (تصویری) به معنای دادن یک پرس و جو به یک سیستم جستجو (کامل) و یافتن مشابه‌ترین موارد با پرس و جو از یک پایگاه داده بسیار بزرگ که معمولاً دارای میلیون‌ها آیتم (تصویر، بردار ویژگی، کد هش و یا ...) است و برگرداندن یک یا تعداد بیشتری نتیجه به ترتیب مناسبی که موارد مشابه در بالاترین رتبه قرار می‌گیرند. به سیستمی که چنین کاری را برای کاربران انجام می‌دهد سیستم بازیابی تصویر مبتنی بر محتوا گفته می‌شود.

بردار ویژگی

بردار ویژگی یک نمایش عددی از یک تصویر است که برای کامپیوتر قابل فهم باشد. در یادگیری ماشین و داده کاوی، ویژگی یک خصیصه یا مشخصه قابل اندازه گیری از یک پدیده در حال مشاهده است. بردار ویژگی یک بردار با ابعاد n از ویژگی‌های عددی است که می‌تواند نمایانگر برخی از اشیا باشد. به طور خلاصه در واقع این بردار نمایش تصویر در قالبی است که برای کامپیوتر قابل ذخیره و پردازش باشد.

استخراج ویژگی - انتخاب ویژگی

نخستین مرحله از عملیات بازیابی محتوای محور تصاویر، استخراج ویژگی از تصاویر است. استخراج ویژگی فرآیندی است که طی آن یک تصویر به عنوان ورودی سیستم دریافت می‌شود و به یک بردار عددی که به بردار ویژگی شناخته شده، تبدیل می‌شود. گاهی این بردار ویژگی به دست آمده، ممکن است به دلیل ابعاد بالا یا ویژگی‌های نامناسب، چندان بهینه نباشد بنابراین یک سری پردازش‌هایی پیش از مرحله ذخیره سازی می‌توان بر آن اعمال کرد تا برای محاسبات آتی مفیدتر باشد. به این پردازش‌ها انتخاب ویژگی گفته می‌شود.

جاسازی یا Embeddings

بسیاری از پیشرفته‌ترین سیستم‌های بازیابی تصویر بر ویژگی‌های ابعادی بسیار بالایی تکیه می‌کنند که به عنوان جاسازی‌هایی شناخته می‌شوند که یا از یک شبکه از پیش آموزش دیده یا با تنظیم دقیق یک شبکه عصبی عمیق استخراج می‌شوند.

پرس و جو یا کوئری

به طور کلی Query در علوم کامپیوتر، به معنای درخواست یک یا تعدادی داده معین از میان مجموعه ای از داده هاست که معمولاً در پایگاه داده ذخیره شده اند. در سیستم باز یابی تصویر مبتنی بر محتوا کوئری به درخواست مورد نظر کاربر اشاره دارد و عموماً منظور همان تصویر مورد جستجو است اما در سیستم های مختلف می تواند به اشکال گوناگون از جمله تصویر، متن، یک طراحی دستی و ... باشد.

ایندکسینگ یا نمایه سازی

نمایه سازی یا ایندکسینگ بیشتر به فرآیند ساختار بندی و نوع ذخیره سازی در پایگاه داده یعنی ساختاردهی بردارهای ویژگی استخراج شده از مرحله نمایش تصویر، ارتباط پیدا می کند. برخی منابع به دلیل اهمیت آن، یک بخش جداگانه را در سیستم به عملیات ایندکس اختصاص می دهند. به دلیل این که عملیات ایندکس باعث افزایش سرعت جستجو می شود و همچنین به دلیل اهمیت زمان پاسخ، این موضوع امروزه جایگاه ویژه ای در سیستم های جستجوی بصری به خصوص در یک پایگاه داده تصویر در مقیاس بزرگ پیدا کرده است.

ایندکس (در الستیک سرچ)

منظور از ایندکس در ساختار الستیک سرچ یک گروه از اسناد است که دارای شاخصه های مشابهی با یکدیگر هستند. به صورت مفهومی ایندکس را می توان مشابه با جدول در پایگاه داده رابطه ای در نظر گرفت که اسناد در آن نگهداری می شوند.

قالب json

قالب JSON (JavaScript Object Notation) یک قالب ساختاری است که برای نمایش و تبادل داده ها استفاده می شود. این قالب بر پایه زبان برنامه نویسی جاوا اسکریپت استوار است و به راحتی قابل فهم و خواندن برای انسان ها و قابل پردازش برای کامپیوترها است. در قالب JSON، داده ها به صورت اشیاء (Objects) و آرایه ها (Arrays) سازماندهی می شوند. هر شیء در قالب JSON از یک مجموعه از جفت های کلید و مقدار تشکیل شده است که با کاما از یکدیگر جدا می شوند. قالب JSON بسیار پر کاربرد در برنامه نویسی و تبادل داده ها در وب است و به عنوان یک استاندارد مشترک بین زبان ها و سرویس های مختلف استفاده می شود.

واژه‌نامه انگلیسی به فارسی

Asymmetric Distance Computation (ADC)	محاسبه فاصله نامتقارن
Average Precision	دقت متوسط
Building retrieval	بازیابی ساختمان
classification	طبقه بندی
Cluster	خوشه
clustering	خوشه بندی
computer vision	بینایی ماشین
Concatenated Rectified Linear Unit transformation (CReLU)	تبدیل واحد خطی
	اصلاح شده الحاقی
Content-Based Image Retrieval	بازیابی تصویر مبتنی بر محتوا
Conventional feature-based methods	روش های متداول مبتنی بر
	ویژگی
cut-off rank	رتبه برش
Deep metric learning	یادگیری متریک عمیق
Description Based Image Retrieval (DBIR)	بازیابی تصویر مبتنی بر
	توضیحات
disjoint	جدا از هم - از هم گسسته
Documents	اسناد
Elasticsearch	الستیک سرچ
embedding	تعبیه، جاسازی
False Negative	منفی اشتباه
False Positive	مثبت کاذب
fashion image retrieval	بازیابی تصویر مد
feature indexing	نمایه سازی ویژگی
feature selection	انتخاب ویژگی
Fields	فیلدها
global	کلی
Graphics Processing Unit (GPU)	واحد پردازش گرافیکی
Ground Truth Positives	پاسخ صحیح واقعی
heuristically designed	طراحی به صورت اکتشافی
index	ایندکس

intention gap	شکاف مقصود
interest point detection	تشخیص نقطه مورد علاقه
Inverted File Index	شاخص فایل معکوس
Levenshtein Edit Distance	ویرایش فاصله لونشتاین
local	محلی
local region description	توضیحات منطقه محلی
Lucene	لوسین
Mapping	نقشه برداری، ساختار ایندکس
mean Average Precision	میانگین دقت متوسط
metadata	فراداده
Node	گره، دستگاه
object recognition	تشخیص شی
Permutant	بردار جایگشت
Permutation-Based Representation	نمایش مبتنی بر جایگشت
person re-identification	شناسایی مجدد شخص
Product Quantization.....	کمی سازی حاصل
ranking	رتبه بندی
real-valued.....	با ارزش واقعی (اعداد حقیقی)
Re-ID	شناسه مجدد
remote sensing.....	سنجش از دور
Replica	همانند، کپی دقیق
scene graph	گراف صحنه
schema.....	طرحواره
semantic gap.....	شکاف معنایی
Shard	تکه شکسته
short-codes	کدهای کوتاه شده
sorting	مرتب سازی
Surrogate Text Representation (STR).....	نمایش متن جایگزین
Table	جدول
Text-Based Image Retrieval	بازیابی تصویر مبتنی بر متن
True Negative	منفی واقعی
True Positive	مثبت واقعی
truncate	کوتاه سازی

واژه‌نامه فارسی به انگلیسی

Documents	اسناد
Elasticsearch	الستیک سرچ
feature selection	انتخاب ویژگی
index	ایندکس
real-valued	با ارزش واقعی (اعداد حقیقی)
Description Based Image Retrieval (DBIR)	بازیابی تصویر مبتنی بر توضیحات
Text-Based Image Retrieval	بازیابی تصویر مبتنی بر متن
Content-Based Image Retrieval	بازیابی تصویر مبتنی بر محتوا
fashion image retrieval	بازیابی تصویر مد
Building retrieval	بازیابی ساختمان
Permutant	بردار جایگشت
computer vision	بینایی ماشین
Ground Truth Positives	پاسخ صحیح واقعی
features folder	پوشه ویژگی ها
Concatenated Rectified Linear Unit transformation (CReLU)	تبدیل واحد خطی
	اصلاح شده الحاقی
object recognition	تشخیص شی
interest point detection	تشخیص نقطه مورد علاقه
embedding	تعبیه، جاسازی، بردار ویژگی
Shard	تکه شکسته
local region description	توضیحات منطقه محلی
disjoint	جدا از هم - از هم گسسته
Table	جدول
Junk	جنس اوراق و شکسته
Cluster	خوشه
Clustering	خوشه بندی
Average Precision	دقت متوسط
cut-off rank	رتبه برش
ranking	رتبه بندی

Conventional feature-based methods	روش های متداول مبتنی بر
ویژگی	
remote sensing	سنجش از دور
Inverted File Index	شاخص فایل معکوس
semantic gap	شکاف معنایی
intention gap	شکاف مقصود
person re-identification	شناسایی مجدد شخص
Re-ID	شناسه مجدد
Classification	طبقه بندی
heuristically designed	طراحی به صورت اکتشافی
schema	طرحواره
Metadata	فراداده
Fields	فیلدها
short-codes	کدهای کوتاه شده
global	کلی
Product Quantization.....	کمی سازی حاصل
Truncate	کوتاه سازی
scene graph	گراف صحنه
Node	گره، دستگاه
Lucene	لوسین
False Positive	مثبت کاذب
True Positive	مثبت واقعی
Asymmetric Distance Computation (ADC)	محاسبه فاصله نامتقارن
Local	محلی
sorting	مرتب سازی
False Negative	منفی اشتباه
True Negative	منفی واقعی
Mean Average Precision	میانگین دقت متوسط
Mapping	ساختار ایندکس
Permutation-Based Representation	نمایش مبتنی بر جایگشت
Surrogate Text Representation (STR).....	نمایش متن جایگزین
feature indexing	نمایه سازی ویژگی
Replica	همانند، کپی دقیق

Graphics Processing Unit (GPU) واحد پردازش گرافیکی
Levenshtein Edit Distance ویرایش فاصله لונشتاین
Deep metric learning یادگیری متریک عمیق

Abstract

Today, content-based image retrieval has become an active and widely used research area in many fields of computer science and computer vision. Content-based image retrieval and visual search systems allow users to find related images based on content similarity with the desired image, to obtain the best results in image retrieval by using textual or image inputs. Researches in this field are mainly focused on providing solutions to increase speed and accuracy, optimal use of resources, as well as improving the experience of searching and retrieving images, and finding the right method in this field is still a challenging and attractive thing in industrial and academic environments

Today, with the evolution of image recording and production tools, as well as the development of the Internet and cloud space, we have witnessed a significant growth in the number and dimensions of available image data and an increase in the demand for browsing and searching among them, and this has led to the creation of the concept of visual search. The challenges of visual search on a very high scale and dimensions have recently become the most important issue in this field, and efforts to propose solutions to deal with them have been especially noticed by researchers

Therefore, in this research, we decided to investigate the previous methods in the field of image recovery and research that provide optimal solutions to face the challenges in this field, and develop our proposed solution. In short, in this solution, we use Elastic Search for storage and search, and with the help of ResNet101 deep neural network and algorithm for creating alternative text for images, we perform the recovery process

We tested our approach on the INRIA Holidays and Flickr1M datasets and were able to propose a method that maximizes search speed while maintaining retrieval precision.

Keywords: Content-based Image Retrieval, Visual Search, High-dimensional data, Image indexing, Elasticsearch



Alzahra University
Faculty of Engineering
Department of Computer Engineering

**A Thesis Submitted in Partial Fulfillment of the Requirements for
the Degree of M.Sc. in Software Engineering**

Title

Visual Search in Content-Based Image Retrieval System

By
Sarvin Saravi

Supervisor
Dr. Masoud Sagharichian

October 2023



Alzahra University
Faculty of Engineering
Department of Computer Engineering

**A Thesis Submitted in Partial Fulfillment of the Requirements for
the Degree of M.Sc. in Software Engineering**

Title
**Visual Search in Content-Based Image Retrieval
System**

By
Sarvin Saravi

Supervisor
Dr. Masoud Sagharichian

October 2023