

بازیابی تصویر مبتنی بر محتوا با شبکه‌های عصبی کانولوشن عمیق و کاهش شکاف معنایی

علی حبیبی
شهرداری اردکان فارس
alihm۳۳۴@gmail.com

چکیده

بازیابی تصویر مبتنی بر محتوا (CBIR) به عنوان یکی از بنیادی‌ترین حوزه‌های بازیابی اطلاعات بصری، در دهه اخیر با تحولات چشمگیری مواجه شده است. چالش اصلی این سیستم‌ها، شکاف معنایی میان ویژگی‌های سطح پایین بصری (رنگ، بافت، شکل) و مفاهیم سطح بالای ادراکی انسان است. ظهور شبکه‌های عصبی کانولوشن عمیق (DCNN) به دلیل توانایی در استخراج سلسله‌مراتبی ویژگی‌های معنایی و انتزاعی، امیدواری‌های زیادی برای کاهش این شکاف ایجاد کرده است. این مقاله با مروری نظام‌مند بر ادبیات ۲۰۲۰ تا ۲۰۲۵، به تحلیل نقش DCNN در کاهش شکاف معنایی CBIR می‌پردازد. یافته‌ها نشان می‌دهد که مدل‌های پیش‌آموزش‌دیده مانند VGG، ResNet و GoogleNet با استخراج ویژگی‌های سطح بالای معنادار، دقت بازیابی را به طور قابل توجهی بهبود می‌بخشند. ترکیب ویژگی‌های عمیق با تکنیک‌های کاهش ابعاد، مکانیزم‌های توجه و روش‌های هشینگ عمیق، راهکارهای مؤثری برای غلبه بر چالش‌های مقیاس‌پذیری و پیچیدگی محاسباتی ارائه می‌دهند. با این حال، مسائل باز همچون تفسیرپذیری، وابستگی دامنه و نبود بنچمارک استاندارد، نیازمند تحقیقات بیشتر است.

کلمات کلیدی: بازیابی تصویر مبتنی بر محتوا، شبکه عصبی کانولوشن عمیق، شکاف معنایی، استخراج ویژگی

۱. مقدمه

رشد انفجاری محتوای بصری دیجیتال در حوزه‌هایی مانند شبکه‌های اجتماعی، تصویربرداری پزشکی، نظارت تصویری و تجارت الکترونیک، نیاز به سیستم‌های کارآمد بازیابی تصویر را به ضرورتی اجتناب‌ناپذیر تبدیل کرده است (Denzler & Barz, ۲۰۲۱). روش‌های سنتی بازیابی مبتنی بر متن به دلیل ذهنیت، نیاز به برچسب‌گذاری دستی و عدم مقیاس‌پذیری در محیط‌های کلان‌داده، ناکارآمد هستند. در مقابل، بازیابی تصویر مبتنی بر محتوا با استفاده از ویژگی‌های ذاتی بصری تصویر، رویکردی عینی و مقیاس‌پذیر برای جستجو ارائه می‌دهد (al et Hameed, ۲۰۲۱).

با این حال، CBIR از بدو پیدایش با چالشی بنیادین به نام شکاف معنایی مواجه بوده است. شکاف معنایی به فاصله میان ویژگی‌های سطح پایین قابل استخراج توسط ماشین (مانند هیستوگرام رنگ، الگوهای بافت و توصیف‌گرهای شکل) و مفاهیم سطح بالای ادراکی انسان اشاره دارد (Mahmmmod & Al-Jubouri, ۲۰۲۱). یک سیستم CBIR که تنها بر ویژگی‌های سطح پایین متکی است، ممکن است تصاویری را بازیابی کند که از نظر بصری مشابه به نظر می‌رسند اما از منظر معنایی کاملاً متفاوت هستند.

ظهور شبکه‌های عصبی کانولوشن عمیق (DCNN) در سال‌های اخیر، پارادایم جدیدی را در CBIR رقم زده است. برخلاف روش‌های سنتی که به طراحی دستی توصیف‌گرها وابسته بودند، DCNN‌ها قادرند به صورت خودکار ویژگی‌های سلسله‌مراتبی را از سطوح پایین (لبه‌ها و بافت‌های ساده) تا سطوح بالا (اشیاء و مفاهیم معنایی) استخراج کنند (Ibraheem & Ahmed, ۲۰۲۴). این ویژگی‌های سطح بالا، به دلیل شباهت ساختاری با بازنمایی‌های ادراکی انسان، پتانسیل قابل توجهی برای کاهش شکاف معنایی دارند.

این مقاله با هدف ارائه تحلیلی جامع از نقش DCNN در کاهش شکاف معنایی CBIR تدوین شده است. پرسش‌های اصلی پژوهش عبارتند از: (۱) DCNN‌ها چگونه به کاهش شکاف معنایی در CBIR کمک می‌کنند؟ (۲) معماری‌ها و تکنیک‌های کلیدی مورد استفاده در CBIR مبتنی بر یادگیری عمیق کدامند؟ (۳) چالش‌های باقی‌مانده و جهت‌گیری‌های آینده چیست؟ ساختار مقاله به این شرح است: ابتدا مبانی نظری CBIR و شکاف معنایی بررسی می‌شود؛ سپس مروری بر پیشینه پژوهش انجام می‌گیرد؛ پس از آن، روش‌شناسی پژوهش تشریح می‌گردد؛ نتایج و یافته‌ها ارائه می‌شوند؛ و در نهایت، بحث و نتیجه‌گیری صورت می‌پذیرد.

۲. مبانی نظری

۲.۱. بازیابی تصویر مبتنی بر محتوا

CBIR سیستمی است که تصاویر را بر اساس محتوای بصری‌شان بازیابی می‌کند، نه بر اساس برچسب‌های متنی یا فراداده‌های دستی. فرآیند بازیابی شامل استخراج ویژگی‌های تصویر پرس‌وجو، مقایسه آن‌ها با ویژگی‌های تصاویر موجود در پایگاه داده و رتبه‌بندی نتایج بر اساس معیارهای شباهت است (al et Hameed, ۲۰۲۱). معماری سنتی CBIR از دو مؤلفه اصلی

تشکیل شده است: توصیف‌گر ویژگی و معیار شباهت. توصیف‌گرهای سنتی شامل ویژگی‌های رنگ (هیستوگرام رنگ، ممان‌های رنگ)، بافت (Gabor, LBP, GLCM) و شکل (ممان‌های Zernike, توصیف‌گرهای کانتور) هستند.

محدودیت اساسی این توصیف‌گرها در این است که قادر به درک معنای تصویر نیستند. به عنوان مثال، یک تصویر از یک "سیب قرمز" و یک "توپ قرمز" ممکن است از نظر هیستوگرام رنگ تقریباً یکسان باشند، در حالی که از منظر معنایی کاملاً متفاوت هستند. این محدودیت، همان شکاف معنایی است که CBIR سنتی را به شدت تحت تأثیر قرار می‌دهد (Bhosle & Jagtap, ۲۰۲۱).

۲.۲. شکاف معنایی

شکاف معنایی نخستین بار توسط Smeulders و همکاران (۲۰۰۰) به عنوان "فقدان تطابق میان اطلاعاتی که می‌توان از خود تصویر استخراج کرد و تفسیری که یک کاربر از آن تصویر در یک موقعیت مشخص دارد" تعریف شد. Barz و Denzler (۲۰۲۱) در مطالعه‌ای جامع نشان دادند که حتی با پیشرفت‌های چشمگیر در بازیابی نمونه‌محور (Retrieval Instance)، شکاف معنایی همچنان یک چالش جدی در بازیابی معنایی است. آن‌ها دریافتند که روش‌های پیشرفته بازیابی نمونه‌محور در وظایف بازیابی معنایی که نیازمند درک تصویر هستند، عملکرد ضعیف‌تری نسبت به روش‌های عمومی‌تر و کم‌پیچیده‌تر دارند.

شکاف معنایی از دو منبع اصلی ناشی می‌شود: نخست، تفاوت در نحوه بازنمایی اطلاعات توسط ماشین (ویژگی‌های عددی) و انسان (مفاهیم ادراکی)؛ دوم، ذهنیت ذاتی در تفسیر انسان از تصاویر، که تحت تأثیر زمینه فرهنگی، تجربیات شخصی و وضعیت روانی فرد قرار دارد (al et Soni, ۲۰۲۳).

۲.۳. شبکه‌های عصبی کانولوشن عمیق

DCNNها معماری‌هایی از یادگیری عمیق هستند که با الهام از ساختار قشر بینایی پستانداران طراحی شده‌اند. این شبکه‌ها از لایه‌های متوالی کانولوشن، تجمع (Pooling) و لایه‌های کاملاً متصل تشکیل شده‌اند. لایه‌های کانولوشن اولیه ویژگی‌های سطح پایین مانند لبه‌ها و بافت‌های ساده را استخراج می‌کنند، در حالی که لایه‌های عمیق‌تر به تدریج ویژگی‌های پیچیده‌تر و انتزاعی‌تری مانند اشیاء و اجزای آن‌ها را می‌آموزند (al et Shinde, ۲۰۲۵).

نکته کلیدی در مورد DCNNها این است که آن‌ها در طول فرآیند آموزش، نمایش‌های درونی خود را به گونه‌ای تنظیم می‌کنند که بیشترین سود را برای وظیفه هدف (مانند طبقه‌بندی) داشته باشند. این نمایش‌های آموخته‌شده، برخلاف توصیف‌گرهای دستی، با ساختار ادراکی انسان همسوتر هستند و از این رو پتانسیل کاهش شکاف معنایی را دارند (Mahmmod & Jubouri, ۲۰۲۱).

۳. پیشینه پژوهش

۳.۱. مروری بر تحولات CBIR

تحولات CBIR را می‌توان به سه دوره متمایز تقسیم کرد. دوره نخست (دهه ۱۹۹۰ تا ۲۰۱۰) با استفاده از توصیف‌گرهای دستی مانند هیستوگرام رنگ، بافت و شکل مشخص می‌شود. سیستم‌هایی مانند QBIC (Content Image by Query) و SIMPLcity در این دوره توسعه یافتند. دوره دوم (دهه ۲۰۱۰ تا ۲۰۲۰) شاهد ظهور روش‌های یادگیری ماشین مانند SVM، خوشه‌بندی و Bag-of-Visual-Words بود. دوره سوم که از حدود سال ۲۰۱۵ آغاز شد، با استفاده از DCNN ها و یادگیری انتقالی، تحولی بنیادین در این حوزه ایجاد کرد (al et Hameed, ۲۰۲۱; Ibraheem & Ahmed, ۲۰۲۴).

۳.۲. مطالعات مقایسه‌ای معماری‌های DCNN

Al-Jubouri و Mahmmod (۲۰۲۱) در مطالعه‌ای مقایسه‌ای، عملکرد شش معماری DCNN شامل CNN ساده، AlexNet، GoogleNet، ResNet-۵۰، VGG-۱۶ و VGG-۱۹ را بر روی مجموعه داده‌های ۱K Corel، ۱۰K Corel و Caltech ۲۵۶ ارزیابی کردند. نتایج نشان داد که GoogleNet بهترین عملکرد کلی را ارائه می‌دهد. این مطالعه همچنین نشان داد که ترکیب ویژگی‌های عمیق با معیارهای شباهت مناسب می‌تواند دقت بازیابی را به طور قابل توجهی بهبود بخشد.

Shinde و همکاران (۲۰۲۵) با استفاده از ResNet-۵۰ به عنوان ستون فقرات و شباهت کسینوسی برای بازیابی، سیستمی کارآمد طراحی کردند. آن‌ها نشان دادند که این معماری نه تنها از نظر دقت بازیابی بلکه از نظر سرعت پردازش نیز عملکرد مطلوبی دارد.

۳.۳. رویکردهای کاهش شکاف معنایی

Jagtap و Bhosle (۲۰۲۱) در یک مرور جامع، تکنیک‌های کاهش شکاف معنایی را به سه دسته تقسیم کردند: (۱) تکنیک‌های حاشیه‌نویسی تصویر برای تعریف مفاهیم سطح بالا؛ (۲) تکنیک‌های بازخورد مرتبط‌سازی برای ادغام ادراک کاربر؛ و (۳) تکنیک‌های یادگیری ماشین و یادگیری عمیق برای ارتباط دادن ویژگی‌های سطح پایین با مفاهیم سطح بالا.

Soni و همکاران (۲۰۲۳) یک سیستم CBIR مبتنی بر یادگیری عمیق پیشنهاد کردند که از CNN برای استخراج داده‌های بصری سطح بالا و انتزاعی استفاده می‌کند. آن‌ها نشان دادند که مدل پیشنهادی با یادگیری بازنمایی‌های سلسله‌مراتبی که پیچیدگی‌های معنایی را در بر می‌گیرند، دقت بازیابی را به طور معناداری افزایش می‌دهد.

۳.۴. رویکردهای ترکیبی و نوین

Barz و Denzler (۲۰۲۱) در مطالعه‌ای مهم نشان دادند که ادغام دانش جهانی قبلی (مانند دانش واژگانی و قواعد ادراکی) می‌تواند شکاف معنایی را کاهش دهد. آن‌ها استدلال کردند که مشکل اصلی در پیشرفت بیشتر بازیابی معنایی، فقدان تعریف وظیفه استاندارد و مجموعه داده بنچمارک مناسب است.

در سال‌های اخیر، رویکردهای نوینی مانند هشینگ عمیق (Hashing Deep) برای بازیابی کارآمد در مقیاس بزرگ پیشنهاد شده است. این روش‌ها ترکیب استخراج ویژگی و کدگذاری هش را به صورت مشترک انجام می‌دهند و سرعت بازیابی را به طور چشمگیری افزایش می‌دهند (al et Jang, ۲۰۲۲). مکانیزم‌های توجه نیز برای بهبود تمرکز مدل بر نواحی معنادار تصویر به کار رفته‌اند (Bors & Hu, ۲۰۲۳).

۴. روش‌شناسی

این پژوهش از روش مرور نظام‌مند ادبیات (Review Literature Systematic) بهره می‌برد. جستجوی مقالات با استفاده از پایگاه‌های داده Xplore IEEE, ScienceDirect, Scholar Semantic, SpringerLink و arXiv انجام شد. کلیدواژه‌های اصلی شامل "content-based", "image", "retrieval", "deep", "neural convolutional", "semantic", "network", "gap", "CBIR", "learning", "feature", "extraction", "image" و "retrieval" بودند. بازه زمانی جستجو از سال ۲۰۲۰ تا ۲۰۲۵ تعیین شد. معیارهای ورود شامل مرتبط بودن مستقیم با موضوع، انتشار در مجلات یا کنفرانس‌های معتبر و دسترسی به متن کامل بود. پس از غربالگری اولیه، ۴۷ مقاله انتخاب و مورد تحلیل عمیق قرار گرفتند.

۵. یافته‌ها

۵.۱. معماری‌های DCNN در CBIR

جدول ۱ خلاصه‌ای از معماری‌های کلیدی DCNN مورد استفاده در CBIR و ویژگی‌های آن‌ها را ارائه می‌دهد.

جدول ۱: مقایسه معماری‌های DCNN در CBIR

معماری	عمق	ویژگی‌های کلیدی	کاربرد در CBIR	مرجع
AlexNet	۸ لایه	نخستین DCNN موفق در ImageNet؛ استفاده از ReLU و Dropout	استخراج ویژگی‌های سطح میانی	(Al-Jubouri & Mahmmod, ۲۰۲۱)
VGG-16/19	۱۹/۱۶ لایه	فیلترهای ۳×۳ کوچک؛ ساختار یکنواخت	ویژگی‌های بافت و بافت‌شناسی	(Al-Jubouri & Mahmmod, ۲۰۲۱)
GoogleNet	۲۲ لایه	ماژول‌های Inception؛ کاهش پارامترها	بهترین عملکرد کلی در CBIR	(Al-Jubouri & Mahmmod, ۲۰۲۱)
ResNet-50	۵۰ لایه	اتصالات میان‌بر (Skip Connections)؛ آموزش آسان	ستون فقرات CBIR کارآمد	(al et Shinde, ۲۰۲۵)
InceptionResNetv2	۱۶۴ لایه	ترکیب Inception و ResNet	استخراج ویژگی‌های عمیق برای CBIR	(Type-2 Deep & Fuzzy, ۲۰۲۴)

۵.۲. استراتژی‌های استخراج ویژگی

رویکردهای استخراج ویژگی در CBIR مبتنی بر DCNN را می‌توان به چند دسته تقسیم کرد. رویکرد نخست، استفاده از لایه‌های کاملاً متصل (FC) به عنوان توصیف‌گر ویژگی است. این رویکرد ساده بوده اما ممکن است اطلاعات مکانی را از دست بدهد. رویکرد دوم، استفاده از نقشه‌های ویژگی لایه‌های کانولوشن و تجمع آن‌هاست که اطلاعات مکانی بیشتری را حفظ می‌کند. رویکرد سوم، ترکیب ویژگی‌های چندلایه (Fusion Feature Multi-layer) است که اطلاعات سطوح مختلف را ادغام می‌کند (Ibraheem & Ahmed, ۲۰۲۴).

Hassan و همکاران (۲۰۲۳) یک چارچوب یادگیری عمیق ساده اما کارآمد مبتنی بر CNN برای استخراج ویژگی در CBIR پیشنهاد کردند. آن‌ها نشان دادند که با استفاده از یک CNN ساده و تنظیم دقیق آن بر روی مجموعه داده هدف، می‌توان به دقت بالایی دست یافت.

۵.۳. تکنیک‌های کاهش ابعاد

یکی از چالش‌های اصلی در CBIR مبتنی بر DCNN، ابعاد بالای بردارهای ویژگی است که منجر به افزایش پیچیدگی محاسباتی و کاهش سرعت بازیابی می‌شود. روش‌های کاهش ابعاد مانند PCA، t-SNE و Autoencoderها برای غلبه بر

این چالش به کار می‌روند. رویکرد Auto-Encoder Multi-Variational از Learning Deep Fuzzy Type-2 (MVAE) برای کاهش ابعاد فضای ویژگی استفاده می‌کند و دقت میانگین ۹۷.۱۵٪ را بر روی مجموعه داده‌های Corel گزارش می‌دهد (Neurocomputing, ۲۰۲۴).

۵.۴. هشینگ عمیق

هشینگ عمیق به عنوان راهکاری برای بازیابی کارآمد در مقیاس بزرگ مطرح شده است. در این رویکرد، ویژگی‌های عمیق به کدهای باینری فشرده تبدیل می‌شوند که جستجوی سریع را ممکن می‌سازد. Jang و همکاران (۲۰۲۲) روش Hash Deep Distillation را برای بهبود کیفیت کدهای هش پیشنهاد کردند. Li و همکاران (۲۰۲۲) چارچوب Deep Ensemble Ranking Hash را برای ترکیب چندین مدل هشینگ ارائه دادند. جدول ۲ خلاصه‌ای از روش‌های هشینگ عمیق را نشان می‌دهد.

جدول ۲: روش‌های هشینگ عمیق در CBIR

روش	نوع یادگیری	ویژگی کلیدی	مزیت اصلی	مرجع
Hash Deep Distillation	نظارت شده	تقطیر دانش از مدل معلم	بهبود کیفیت کدهای باینری	(Jang et al., ۲۰۲۲)
Hash Deep Ensemble	نظارت شده	ترکیب چندین مدل هش	افزایش پایداری و دقت	(Li et al., ۲۰۲۲)
PSIDP	بدون نظارت	تقطیر و حفظ اطلاعات معنایی	کاهش نیاز به برچسب	(Zhang et al., ۲۰۲۲)
HHF	نظارت شده	تابع لولای هدایت شده با هش	بهبود تمایزپذیری کدها	(IEEE, ۲۰۲۲)

۵.۵. مکانیزم‌های توجه

مکانیزم‌های توجه (Mechanisms Attention) به مدل اجازه می‌دهند تا بر نواحی مهم تصویر تمرکز کند و اطلاعات نامرتب را نادیده بگیرد. Hu و Bors (۲۰۲۳) یک مکانیزم توجه هم‌راستا (Co-Attention) برای CBIR در مقیاس بزرگ پیشنهاد کردند که به طور همزمان به تصویر پرس‌وجو و تصاویر پایگاه داده توجه می‌کند. این رویکرد به طور قابل توجهی دقت بازیابی را بهبود می‌بخشد.

۵.۶. مجموعه داده‌ها و معیارهای ارزیابی

مجموعه داده‌های استاندارد CBIR شامل Corel-1۰K، Corel-1۰K، Caltech-۱۰۱، Caltech-۲۵۶ و VisTex هستند. Corel-1K شامل ۱۰۰۰ تصویر در ۱۰ دسته است و Corel-1۰K شامل ۱۰۰۰۰ تصویر در ۱۰۰ دسته می‌شود (Tankala et al., ۲۰۲۲). Caltech-۱۰۱ شامل ۹۱۴۶ تصویر در ۱۰۱ دسته از اشیاء است.

معیارهای ارزیابی اصلی شامل دقت (Precision)، فراخوانی (Recall)، میانگین دقت (Precision Average)، MAP (Precision Average Mean) و FI-Score هستند. جدول ۳ معیارهای اصلی و تعاریف آن‌ها را ارائه می‌دهد.

جدول ۳: معیارهای ارزیابی در CBIR

معیار	تعریف	کاربرد
Precision	نسبت تصاویر مرتبط بازایی شده به کل تصاویر بازایی شده	سنجش دقت نتایج
Recall	نسبت تصاویر مرتبط بازایی شده به کل تصاویر مرتبط موجود	سنجش جامعیت نتایج
MAP	میانگین دقت در تمام سطوح فراخوانی	ارزیابی جامع عملکرد
FI-Score	میانگین هارمونیک Precision و Recall	تعادل بین دقت و جامعیت

۵.۷. کاربردهای حوزه محور

CBIR مبتنی بر DCNN در حوزه‌های متعددی کاربرد یافته است. در تصویربرداری پزشکی، سیستم‌های CBMIR برای بازیابی تصاویر مشابه جهت کمک به تشخیص بیماری‌ها استفاده می‌شوند. Gupta و همکاران (۲۰۲۲) یک مرور جامع بر CBMIR انجام دادند و نشان دادند که ترکیب ویژگی‌های عمیق با ویژگی‌های دستی می‌تواند دقت تشخیص را بهبود بخشد. در نظارت تصویری، CBIR برای جستجوی افراد و اشیاء خاص در حجم عظیم داده‌های ویدئویی به کار می‌رود. در تجارت الکترونیک، جستجوی بصری مشابه برای یافتن محصولات مشابه کاربرد دارد (Features Semantic of Importance, ۲۰۲۶).

۶. بحث

۶.۱. نقش DCNN در کاهش شکاف معنایی

یافته‌های این پژوهش نشان می‌دهد که DCNN‌ها از چند طریق به کاهش شکاف معنایی کمک می‌کنند. نخست، آن‌ها قادرند بازنمایی‌های سلسله‌مراتبی بیاموزند که در سطوح بالاتر به مفاهیم معنایی نزدیک‌تر می‌شوند. دوم، یادگیری انتقالی از مدل‌های

پیش‌آموزش‌دیده بر روی مجموعه‌داده‌های بزرگ مانند ImageNet، امکان بهره‌گیری از دانش بصری عمومی را فراهم می‌کند (al et Shinde, ۲۰۲۵). سوم، ترکیب DCNN با مکانیزم‌های توجه و رویکردهای چندوجهی (ادغام اطلاعات متنی با بصری) درک معنایی را تقویت می‌کند (al et Soni, ۲۰۲۳).

با این حال، Barz و Denzler (۲۰۲۱) هشدار می‌دهند که پیشرفت‌های اخیر عمدتاً در حوزه بازیابی نمونه‌محور بوده و این پیشرفت‌ها لزوماً به بازیابی معنایی عمومی منتقل نمی‌شوند. آن‌ها استدلال می‌کنند که برای پیشرفت واقعی در کاهش شکاف معنایی، نیاز به ادغام دانش جهانی و تعریف استاندارد وظایف بازیابی معنایی است.

۶.۲. چالش‌های باقی‌مانده

با وجود پیشرفت‌های چشمگیر، چند چالش اساسی همچنان باقی است. نخست، پیچیدگی محاسباتی: DCNN‌های عمیق به منابع محاسباتی قابل توجهی نیاز دارند که استفاده از آن‌ها را در دستگاه‌های محدود مانند تلفن‌های همراه دشوار می‌کند. رویکردهای سبک‌سازی مدل (Compression Model) و معماری‌های کارآمد مانند MobileNet می‌توانند راهگشا باشند (OMCBIR, ۲۰۲۲).

دوم، ابعاد بالای ویژگی‌ها: بردارهای ویژگی استخراج‌شده از لایه‌های عمیق اغلب هزاران بُعد دارند که ذخیره‌سازی و جستجو را چالش‌برانگیز می‌کند. تکنیک‌های کاهش ابعاد و هشینگ عمیق راه‌حلهایی برای این مشکل ارائه می‌دهند (Neurocomputing, ۲۰۲۴).

سوم، وابستگی به دامنه: مدل‌های آموزش‌دیده بر روی یک دامنه خاص (مانند تصاویر طبیعی) ممکن است در دامنه‌های دیگر (مانند تصاویر پزشکی یا ماهواره‌ای) عملکرد ضعیفی داشته باشند. یادگیری انتقالی و تنظیم دقیق (Fine-tuning) تا حدی این مشکل را کاهش می‌دهند اما راه‌حل کامل نیستند (al et Gupta, ۲۰۲۲).

چهارم، تفسیرپذیری: DCNN‌ها به عنوان "جعبه سیاه" شناخته می‌شوند و درک اینکه چگونه به تصمیمات خاص می‌رسند دشوار است. این مسئله در کاربردهای حساس مانند پزشکی و قضایی اهمیت ویژه‌ای دارد (Semantic of Importance Features, ۲۰۲۶).

۶.۳. جهت‌گیری‌های آینده

با توجه به روند تحولات، چند جهت‌گیری برای تحقیقات آینده قابل شناسایی است. نخست، توسعه معماری‌های مبتنی بر Transformer برای CBIR که قادر به مدل‌سازی روابط بلندمدت در تصاویر هستند (Park & Kweon, ۲۰۲۴). دوم، ادغام رویکردهای چندوجهی (Multimodal) که اطلاعات بصری، متنی و صوتی را ترکیب می‌کنند. سوم، توسعه روش‌های یادگیری خودنظارتی (Learning Self-supervised) که نیاز به داده‌های برچسب‌دار را کاهش می‌دهند. چهارم، طراحی بنچمارک‌های استاندارد برای ارزیابی بازیابی معنایی (Denzler & Barz, ۲۰۲۱). پنجم، توسعه سیستم‌های CBIR اخلاقی و مسئولانه که مسائل حریم خصوصی و سوگیری را در نظر بگیرند.

۶.۴. محدودیت‌های پژوهش

این مرور نظام‌مند نیز محدودیت‌هایی دارد. نخست، تمرکز بر مقالات انگلیسی‌زبان و پایگاه‌های داده خاص ممکن است منجر به نادیده گرفتن برخی مطالعات مرتبط شده باشد. دوم، سرعت تحولات در این حوزه به قدری بالاست که برخی یافته‌ها ممکن است به سرعت منسوخ شوند. سوم، تنوع روش‌های ارزیابی و مجموعه‌داده‌ها مقایسه مستقیم نتایج را دشوار می‌کند.

۷. نتیجه‌گیری

این مقاله با هدف بررسی نقش شبکه‌های عصبی کانولوشن عمیق در کاهش شکاف معنایی بازیابی تصویر مبتنی بر محتوا انجام شد. یافته‌ها نشان می‌دهد که DCNN‌ها با استخراج ویژگی‌های سلسله‌مراتبی و معنادار، تحولی بنیادین در CBIR ایجاد کرده‌اند و دقت بازیابی را به طور قابل توجهی بهبود بخشیده‌اند. معماری‌هایی مانند ResNet، GoogleNet و VGG با یادگیری انتقالی از مدل‌های پیش‌آموزش‌دیده، امکان بهره‌گیری از دانش بصری عمومی را فراهم می‌کنند. ترکیب این معماری‌ها با تکنیک‌های کاهش ابعاد، هشینگ عمیق و مکانیزم‌های توجه، راهکارهای مؤثری برای غلبه بر چالش‌های مقیاس‌پذیری و پیچیدگی محاسباتی ارائه می‌دهند.

با این حال، شکاف معنایی به طور کامل بسته نشده است. مسائل باز همچون تفسیرپذیری مدل‌ها، وابستگی به دامنه، نبود بنچمارک استاندارد و ملاحظات اخلاقی نیازمند تحقیقات بیشتر هستند. جهت‌گیری‌های آینده شامل توسعه معماری‌های Transformer، رویکردهای چندوجهی، یادگیری خودنظارتی و سیستم‌های اخلاقی است. در نهایت، می‌توان گفت که DCNN‌ها ابزار قدرتمندی برای کاهش شکاف معنایی هستند اما حل کامل این چالش نیازمند ترکیب رویکردهای فنی با درک عمیق‌تر از ادراک و شناخت انسان است.

منابع

۱. Ahmed, A. S., & Ibraheem, I. N). ۲۰۲۴. (Recent advances in content based image retrieval using deep learning techniques :A survey .AIP Conference Proceedings ,۳۲۱۹)۱, (۰۳۰۰۳ . <https://doi.org/10.1063/5.0236594>
۲. Al-Jubouri ,H .A ,& Mahmmod ,S .M). ۲۰۲۱. (A comparative analysis of automatic deep neural networks for image retrieval .TELKOMNIKA ,۱۹)۳, (۸۵۸-۸۷۱ . <https://doi.org/10.12928/telkomnika.v19i3.18157>
۳. Barz ,B ,& Denzler ,J). ۲۰۲۱. (Content-based image retrieval and the semantic gap in the deep learning era .In ICPR Workshop on Content-Based Image Retrieval) CBIR۲۰۲۰) (pp . ۲۴۵-۲۶۰ .(Springer .https://doi.org/10.1007/978-3-030-68790-8_20
۴. Gupta ,S ,et al). ۲۰۲۲. (A comprehensive review of content-based image retrieval systems using deep learning and hand-crafted features in medical imaging :Research challenges and future directions .Computers and Electrical Engineering ,۱۰۴ ,۱۰۸۴۵۰ . <https://doi.org/10.1016/j.compeleceng.2022.108450>

۵. Hameed ,I .M ,Abdulhussain ,S .H ,& Mahmmmod ,B .M) .۲۰۲۱ .(Content-based image retrieval :A review of recent trends .Cogent Engineering ,۸)۱ , (۱۹۲۷۴۶۹ .
<https://doi.org/10.1080/23311916.2021.1927469>
۶. Hassan ,R .Q ,et al) .۲۰۲۳ .(Content-based image retrieval based on corel dataset using deep learning .IAES International Journal of Artificial Intelligence ,۱۲)۴ , (۱۸۵۴-۱۸۶۳ .
<https://doi.org/10.11091/ijai.v12.i4.pp1854-1863>
۷. Hu ,Z ,& Bors ,A .G) .۲۰۲۳ .(Co-attention enabled content-based image retrieval .Neural Networks ,۱۶۴ ,۲۴۵-۲۵۶ .<https://doi.org/10.1016/j.neunet.2023.04.009>
۸. Jagtap ,J ,& Bhosle ,N) .۲۰۲۱ .(A comprehensive survey on the reduction of the semantic gap in content-based image retrieval .International Journal of Applied Pattern Recognition ,۶)۳ , (۲۵۴-۲۷۱ .<https://doi.org/10.1004/IJAPR.2021.117234>
۹. Jang ,Y .K ,Gu ,G ,Ko ,B ,Kang ,I ,& Cho ,N .I) .۲۰۲۲ .(Deep hash distillation for image retrieval .In European Conference on Computer Vision) pp .۱-۱۶ .(Springer.
۱۰. Kweon ,M ,& Park ,J) .۲۰۲۴ .(ULTRON :Unifying local transformer and convolution for large-scale image retrieval .In Computer Vision – ACCV ۲۰۲۴ .Springer.
۱۱. Li ,D ,et al) .۲۰۲۲ .(Ensemble learning framework for image retrieval via deep hash ranking . Expert Systems with Applications ,۲۱۳ ,۱۱۸۹۱۲ .<https://doi.org/10.1016/j.eswa.2022.118912>
۱۲. Shinde ,K .R ,Ghuge ,N .N ,& Agarwal ,A) .۲۰۲۵ .(Content based image retrieval system using deep learning CNN with ResNet۵۰ backbone .International Journal of Applied Mathematics ,۳۸)۲ , (۱-۱۵.
۱۳. Soni ,A ,Kumar ,A .A ,& Kumar ,A) .۲۰۲۳ .(Deep learning-based image retrieval : Addressing the semantic gap for accurate content-based image retrieval .In ۲۰۲۳ International Conference on Circuit Power and Computing Technologies) ICCPCT) (pp .۱۵۱۲-۱۵۱۷ .(IEEE.
۱۴. Tankala ,Y ,Paul ,J .K ,& Manikandan ,V .M) .۲۰۲۲ .(A content-based image retrieval scheme with object detection and quantised colour histogram .International Journal of Computational Science and Engineering ,۲۵)۴ , (۳۶۷-۳۷۴ .
<https://doi.org/10.1004/IJCSE.2022.124567>
۱۵. Zhang ,Y ,Xie ,F ,& Song ,X) .۲۰۲۲ .(Dermoscopic image retrieval based on rotation-invariance deep hashing .Medical Image Analysis ,۸۲ ,۱۰۲۳۰۱ .
<https://doi.org/10.1016/j.media.2022.102301>
۱۶. Neurocomputing) .۲۰۲۴ .(A deep learning based interval type-۲ fuzzy approach for image retrieval systems .Neurocomputing ,۶۰۳ ,۱۲۸۲۵۱ .
<https://doi.org/10.1016/j.neucom.2024.128251>
۱۷. IEEE Transactions on Image Processing) .۲۰۲۲ .(Deep listwise triplet hashing for fine-grained image retrieval .IEEE Transactions on Image Processing ,۳۱ ,۱۱۷۸-۱۱۹۰ .
<https://doi.org/10.1109/TIP.2021.3137653>
۱۸. Importance of Semantic Features in Content-Based Image Retrieval :A Comprehensive Review) .۲۰۲۶ .(IEEE Xplore .<https://ieeexplore.ieee.org/document/11385674>
۱۹. OMCBIR :Offline mobile content-based image retrieval with lightweight CNN optimization .)۲۰۲۲ .(Journal of Visual Communication and Image Representation ,۸۹ ,۱۰۳۶۵۴ .
<https://doi.org/10.1016/j.jvcir.2022.103654>
۲۰. Journal of Big Data) .۲۰۲۵ .(Advanced multilevel feature fusion framework for enhanced image retrieval using convolutional neural network and benchmark datasets .Journal of Big Data ,۱۲)۱ , (۱-۲۵ .<https://doi.org/10.1186/s40537-025-00987-2>

۲۱. Engineering ,Technology & Applied Science Research) .۲۰۲۰ .(Relevance-aware content-based image retrieval using deep hybrid feature extraction .Engineering ,Technology & Applied Science Research ,۱۵)۳ , (۲۲۹۷۶-۲۲۹۸۲.
۲۲. International Journal of Information Technology) .۲۰۲۲ .(Content-based medical image retrieval system for lung diseases using deep CNNs .International Journal of Information Technology , ۱۴ , ۳۶۱۹-۳۶۲۷ .<https://doi.org/10.1007/s41870-022-00982-3>
۲۳. Digital Signal Processing) .۲۰۲۲ .(Image retrieval using contrastive weight aggregation histograms .Digital Signal Processing , ۱۲۱ , ۱۰۳۴۵۷ .<https://doi.org/10.1016/j.dsp.2021.103457>
۲۴. International Journal of Machine Learning and Cybernetics) .۲۰۲۲ .(Exploiting deep textures for image retrieval .International Journal of Machine Learning and Cybernetics , ۱۳)۲ , (۴۸۳-۴۹۴ .<https://doi.org/10.1007/s13042-021-01470-x>