# مقدمه

1- پاورقی + مخفف کلمات پاورقی در متن آورده شود.

پاورقی ها پایین هر صفحه درج شود.

هدف روش­های بازیابی تصاویر، دسته‌بندی دقیق تصاویر، با استفاده از شباهت‌ها و تفاوت‌های موجود در بافت، رنگ، فرم و سایر ویژگی‌های تصویر است. این شاخه از علم پردازش تصویر، برای اولین بار در سال ۱۹۷۰ با رویکرد مبتنی بر متن[[1]](#footnote-1)معرفی گردید[] . پس از آن رویکردی متفاوت، با عنوان رویکرد مبتنی بر محتوا[[2]](#footnote-2) مورد توجه قرار گرفت که بر اساس ویژگی‌های استخراج شده از تصاویر، کار می‌کرد. این رویکرد به‌سرعت جایگزین رویکرد پیشین شد. این موضوع به دلیل کاربردهای متنوع در حوزه­های مختلف پزشکی []، احراز هویت[]، پیشگیری از وقوع جرم مورد توجه قرار دارد []. در این میان چالش‌های بسیاری به هنگام استفاده از روش‌های مبتنی بر محتوا، پیش می‌آمد. از جمله، ویژگی‌های استخراج شده با ادراک انسان فاصله معنایی بسیاری داشتند. اما با انتخاب و استخراج درست ویژگی‌های مورد محاسبه این فاصله کمتر به چشم آمده‌است، به‌گونه‌ای که اکنون با نیاز به بررسی دقیق‌تر دسته‌بندهای تصاویر، بازیابی تصاویر ریزدانه‌ای[[3]](#footnote-3) معرفی شده‌است که در پیداکردن ویژگی‌های مشابه، تا حد ادراک انسان رفتار می‌کند. بازیابی تصویر برای پرس‌وجوی مبتنی بر تصویر، شامل رویکردهای متفاوتی است که می‌توان آن‌ها را در سه گروه عمده‌ بیان نمود: 1)بازیابی تصویر مبتنی بر طرح 2)بازیابی تصویر مبتنی بر محتوا 3)بازیابی تصویر مبتنی بر ریزدانه. در ادامه به بررسی نمونه‌هایی از این گروه‌ها پرداخته‌شده‌است.

در رویکرد اول بازیابی تصویر مبتنی بر طرح، یک مدل بازیابی تصویر مبتنی بر طرح با نام صفرشات[[4]](#footnote-4) را بررسی می‌کند که در آن دسته‌های آزمایشی در مرحله آموزش ظاهر نمی‌شوند. سعی می‌شود از طریق گسستگی نامتقارن[[5]](#footnote-5) به بازیابی آگاهانه از ساختار رسید. که در آن ویژگی‌های تصویر به ویژگی‌های ساختار و ویژگی‌های ظاهری تفکیک می‌شوند، درحالی‌که ویژگی‌های طرح تنها به فضای ساختار، پیش‌بینی می‌شوند. از طریق جداسازی ساختار و فضای ظاهری، ترجمه دامنه دو‌جهته[[6]](#footnote-6) بین حوزه طرح و حوزه تصویر انجام می‌شود [1]. بازیابی تصویر مبتنی بر طرح به‌عنوان یک فرایند درشت به ریز فرموله شده است و یک مدل رتبه‌بندی متقابل با نام آبشاری عمیق[[7]](#footnote-7) که می‌تواند از تمام اطلاعات چندوجهی مفید در طرح‌ها و تصاویر حاشیه‌نویسی[[8]](#footnote-8) بهره‌برداری کند و کارایی بازیابی را بهبود بخشد، پیشنهاد شده است [2]. بافت، کانال و اطلاعات توالی مکانی برای بازیابی مبتنی بر محتوا مورد تمرکز قرار می‌گیرند. این کار با استفاده از یک مدل بازیابی تصویر ریزدانه‌ای مبتنی بر طرح عمیق [[9]](#footnote-9) که هدف آن، استنباط نقشه‌های توجه در امتداد ابعاد کانال و مکان است، انجام می­شود [10].

در رویکرد دوم بازیابی تصویر مبتنی بر محتوا، استفاده از جنگل‌های مسیر بهینه[[10]](#footnote-10) (بدون نظارت و با نظارت) و رویکردهای یادگیری فعال[[11]](#footnote-11) را برای بازخورد مرتبط در سیستم‌های بازیابی تصویر مبتنی بر محتوا بررسی می‌کند. پراطلاعات[[12]](#footnote-12)‌ترین تصاویری که با رویکرد یادگیری فعال انتخاب می‌شوند، آن‌هایی هستند که بهترین تعادل را بین شباهت با تصویر پرس‌وجو و درجات خاصی از تنوع و عدم قطعیت را ارائه می‌دهند [3]. بازیابی تصویر برای مشخص‌کردن دسته‌بندی تصاویر و یک مدل شبکه کانولوشنی سیامی[[13]](#footnote-13) که شامل برچسب‌های دسته و شیء در آموزش برای تولید ویژگی آگاه از دسته است، پیشنهاده شده است. این مدل با اصلاح رویه آموزشی همراه می‌باشد که به طور هم‌زمان دسته و برچسب مورد را یاد می‌گیرد [4]. رویکرد دیگر بازیابی متقابل رسانه‌ای مبتنی بر ترکیب چند ویژگی[[14]](#footnote-14) می­باشد. این روش قادر به ادغام چندین ویژگی برای ارتقاء درک معنایی و اتخاذ یادگیری متخاصم[[15]](#footnote-15) برای بهبود بیشتر دقت بازنمایی زیرفضای عمومی است [5]. بازیابی تصویر مبتنی بر محتوا [6] پیشنهاد می‌شود. در مرحله توصیف تصویر، این روش ابتدا توصیفگر ریزساختار سنتی[[16]](#footnote-16) را اصلاح می‌کند تا رابطه مستقیم بین ویژگی‌های شکل، بافت و بین ویژگی‌های رنگ و بافت را نشان دهد. سپس هیستوگرام الگوهای باینری محلی یکنواخت[[17]](#footnote-17) تصویر را استخراج می‌کند تا اطلاعات تفاوت رنگ را به تصویر بکشد.

یک چارچوب چند وظیفه‌ای مبتنی بر جداسازی و بازسازی ویژگی[[18]](#footnote-18) برای بازیابی متقابل وجهی بر اساس روش‌های رایج یادگیری مکانی پیشنهاد شده است، که ماژول جداسازی ویژگی را برای مقابله با عدم تقارن اطلاعات بین روش‌های مختلف معرفی می‌کند [7]. بازیابی تصویر مبتنی بر محتوا با یک شبکه عصبی سیامی کانولوشنی معرفی شده است. ابتدا، تکه‌های ضایعه برای ایجاد دو مجموعه‌های داده، برش داده می‌شوند و جفت‌های دوتکه دلخواه یک مجموعه‌داده جفت مرتب را تشکیل می‌دهند. دوم، این مجموعه‌داده جفت مرتب برای آموزش یک شبکه استفاده می‌شود. سوم، یک پچ آزمایشی به‌عنوان یک پرس‌وجو در نظر گرفته می‌شود. فاصله بین این پرس‌وجو و بیست مورد در هر دو مجموعه‌داده با استفاده از شبکه عصبی کانولوشنی سیامی آموزش‌دیده محاسبه می‌شود. موارد نزدیک به پرس‌وجو برای ارائه پیش‌بینی نهایی با بالاترین امتیاز استفاده می‌شود]8-9[.با اشاره به روش‌هایی که با خطای ویژگی‌های عمومی به استخراج ویژگی‌های متمایزتر کمک می‌کنند، یک تابع محاسبه خطا، به نام خطای متمرکز سخت معرفی می‌شود[11]. یک شبکه ترکیبی مبتنی بر خودتوجهی[[19]](#footnote-19) برای یادگیری بازنمایی‌های رایج داده‌های رسانه‌های مختلف[[20]](#footnote-20) پیشنهاد شده است [12]. در این روش ابتدا از یک لایه خودتوجهی محلی برای یادگیری فضای توجه مشترک[[21]](#footnote-21) بین داده‌های رسانه‌های مختلف استفاده می‌شود، سپس یک روش الحاق شباهت[[22]](#footnote-22) برای درک رابطه محتوایی بین ویژگی‌ها پیشنهاد می‌شود. یک چارچوب سبک‌تر برای نمونه‌برداری تدریجی از قطعات متمایز[[23]](#footnote-23)، جهت یادگیری جزئیات ارائه می‌شود [13]. در این روش ابتدا شیء از تصویر اصلی تقویت‌شده و سپس یک نمونه‌برداری[[24]](#footnote-24) خودتطبیقی برای شناسایی بیشتر منطقه تقویت‌شده انجام می‌گردد.

در رویکرد نهایی مبتنی بر ریزدانه که موضوع پژوهش پیش­رو است ، انتخاب توصیف‌گرهای عمیق[[25]](#footnote-25) مفید به خوبی به تشخیص تصویر با دانه‌ریز کمک می‌کند. به طور خاص، یک مدل جدید شبکه عصبی کانولوشنی ماسک دار[[26]](#footnote-26)، بدون لایه‌های کاملاً متصل پیشنهاد شده است [14]. بر اساس حاشیه‌نویسی‌های بخش، این روش شامل یک شبکه کاملاً کانولوشنی برای مکان‌یابی قسمت‌های متمایز و مهم‌تر از آن تولید ماسک‌های جسم/قطعه وزن‌دار برای انتخاب توصیف‌گرهای کانولوشنی، مفید و معنادار است. یک رویکرد دیگر تخمین ریزدانه برای تخمین نمره زیبایی‌شناسی[[27]](#footnote-27) پیشنهاد شده است .[15] مکانیسم‌های توجه، موقعیت و کانال را برای افزایش ترکیب ویژگی‌های تصویر ترکیب می‌کند. با آموزش شبکه رگرسیون، جدا از شبکه طبقه‌بندی، طبقه‌بندی وظیفه رگرسیون را تکمیل می‌کند. یک سیستم یادگیری نیمه نظارت معرفی ‌شده است [16]. در این روش یک مکانیسم توجه تعاملی ریزدانه‌ای تعبیه شده، که در ابتدا از تصاویر برچسب‌دار استفاده کرده و به تهیه بردارهای احتمالی حاصل از این تصویر، می‌پردازد. سپس داده‌های آموزشی بدون برچسب را با این بردارها مقایسه و طبقه‌بندی می‌کند. روش یادگیری هش[[28]](#footnote-28)با دو مشکل بررسی می‌شود: 1- ویژگی‌ها با ابعاد کم، فرایند بازیابی را تسریع می‌بخشند اما به دلیل ازدست‌رفتن اطلاعات، دقت را کاهش می‌دهند. 2- تصاویر ریزدانه منجر به ایجاد کدهای هش جستجوی یکسان در خوشه‌های مختلف در فضای پنهان پایگاه‌داده می‌شوند. از این‌رو رویکرد دیگر برای حل مشکل یاد شده استفاده از یک شبکه پاک‌کننده توجه مبتنی بر ثبات ویژگی[[29]](#footnote-29)می­باشد [17]. یک شبکه بازیابی و استخراج اطلاعات متمایز به نام شبکه بازیابی و استخراج اطلاعات متمایز [[30]](#footnote-30)(DRE-Net) معرفی شده است [18]. این شبکه از دو شبکه فرعی تشکیل شده است: 1- زیرشبکه بازیابی اطلاعات متمایز ریز [[31]](#footnote-31) 2- زیرشبکه شناسایی با رابطه معنایی خطای تقطیر[[32]](#footnote-32). ماژول اول با استفاده از ویژگی‌ها، به بازیابی جزئیات بافت حیاتی[[33]](#footnote-33) پیکسل‌ها کمک می‌کند. ماژول دوم به روابط صحیح بین هر دو پیکسل در نقشه ویژگی می‌پردازد. در نتیجه ماژول دوم می‌تواند به ماژول اول برای پیداکردن جزئیات دقیق و قابل‌اعتماد کمک کند. روشی برای استفاده از یک مدل توجه چند سطحی[[34]](#footnote-34) پیشنهاد شده است ]19-21[.در این روش ابتدا سه اندازه میدان گیرنده معمولی[[35]](#footnote-35)، برای نقشه‌های توجه چند سطحی انتخاب می‌شوند. سپس یادگیری چندسطحی برای استخراج ویژگی‌های متمایز از این مناطق محلی معرفی می‌گردند. این روش نگرش جدیدی در مورد چگونگی استفاده از فعال‌سازهای شبکه عصبی، برای تولید مناطق چند مقیاسی که برای طبقه‌بندی ریزدانه‌ای مفید هستند، ارائه می‌دهد. یکی از مشکلات بازیابی تصویر ریزدانه‌ای، تنوع کم در بین کلاس‌های مختلف و درعین‌حال تنوع زیاد در هر کلاس می­باشد. با بررسی خطای آنتروپی متقاطع[[36]](#footnote-36) برای ایجاد ویژگی متمایز شبکه عصبی کانولوشنال[[37]](#footnote-37) (CNN) با برخی عملیات اضافی مانند نرمال‌سازی مقیاس، می‌توان بهینه‌تر عمل کرد. سپس برای حل این مشکل یک نوع خطای آنتروپی متقاطع تکه‌ای[[38]](#footnote-38) معرفی شده است [20]. یک رویکرد هرمی دو جهته موثر[[39]](#footnote-39) برای بهبود بازنمایی‌های داخلی ویژگی‌ها پیشنهاد می‌شود [23]. در این رویکرد وظیفه تشخیص تصویر ریزدانه در یادگیری چند شات[[40]](#footnote-40) انجام می­شود. به طور خاص، یک هرم ویژگی چند مقیاسی و یک هرم توجه چند سطحی در شبکه پایه مستقر می‌شود و به تدریج ویژگی‌ها را از فضاهای دانه‌ای مختلف از طریق هر دو آنها جمع‌آوری می‌کند. سپس یک روش پردازش هدایت‌شده توجه[[41]](#footnote-41) را با همکاری یک هرم توجه چند سطحی ارائه می‌دهد تا عدم قطعیت ناشی از پس‌زمینه‌های مشروط‌شده به نمونه‌های محدود[[42]](#footnote-42) را کاهش دهد. در رویکرد دیگر به طبقه‌بندی سلسله‌مراتبی[[43]](#footnote-43) پرداخته شده‌است [24]. ازآنجاکه وظیفه اصلی سیستم به چند وظیفه فرعی تقسیم می‌شود تا ساختار برچسب درختی ارائه گردد، این وظایف فرعی همبستگی بالایی دارند. سپس می‌توانند کاندیدهایی جهت قرارگیری در دسته‌بندی سطح پایین‌تر (ریزدانه)[[44]](#footnote-44) ارائه دهند که خود آن‌ها دارای ویژگی‌هایی هستند که سطح بالاتر (درشت‌دانه)[[45]](#footnote-45) را نیز توصیف می‌کنند؛ بنابراین می‌توان یک شبکه چند وظیفه‌ای عمیق مشترک[[46]](#footnote-46) برای طبقه‌بندی سلسله‌مراتبی تصویر ارائه داد. روش‌های موجود عموماً از داده‌های کمکی[[47]](#footnote-47) برای هدایت شبکه استفاده می‌کنند. این مورد دو مشکل دارد: 1- استفاده از داده‌های کمکی مانند جعبه‌ی محدودکننده[[48]](#footnote-48) که نیازمند دانش تخصصی می­باشد. 2- استفاده از چند شبکه فرعی که موجب می‌شود شبکه پیچیده و آموزش نیز مراحل متعددی داشته باشد. در ادامه برای حل مشکل یاد شده یک شبکه خودتوجهی مکانی سراسری[[49]](#footnote-49) معرفی می­شود که شامل یک ماژول خودتوجهی مکانی[[50]](#footnote-50) و یک تکنیک تقطیر خودتوجهی[[51]](#footnote-51) است. ماژول اول اطلاعات متنی را در ویژگی‌های محلی رمزگذاری کرده و موجب بهبود نمایش درون کلاسی می‌شود. سپس ماژول دوم اطلاعات را از ماژول اول گرفته و به یک نقشه ویژگی اولیه تقطیر می‌کند. این مورد موجب نمایش بین طبقاتی می‌شود. با محاسبه خطای طبقه‌بندی از این دو ماژول، شبکه می‌تواند هر دو ویژگی بین کلاسی و درون کلاسی را آموزش ببیند [25]. ( می تواند بعد از رفرنس 17 باشد.) + یک مشکل از روش قبلی آورده شود. یک شبکه پاک‌سازی پیش‌رونده[[52]](#footnote-52) پیشنهاد می‌شود [26]. در این شبکه، یک مکانیسم پاک‌سازی چند شبکه‌ای، نمونه‌های داده را افزایش می‌دهد و به ثبت ویژگی‌های متمایز محلی کمک می‌کند، جایی که ساختار کلی تصویر به‌طور غیرمستقیم از طریق پاک‌سازی پیکسلی تخریب می‌شود. تجمع ویژگی‌های متقاطع با استخراج ویژگی‌های کلاس برجسته[[53]](#footnote-53) از اهمیت زیادی در بازیابی بصری ریزدانه‌ای برخوردار است [26].

تغییر رفرنس 27 و 28و 15 و11

اضافه کردن روش پیشنهادی (2پاراگراف)

# مراجع

Li, J., Ling, Z., Niu, L., & Zhang, L. (2022). Zero-shot sketch-based image retrieval with structure-aware asymmetric disentanglement. In Computer Vision and Image Understanding (Vol. 218, p. 103412). Elsevier BV. https://doi.org/10.1016/j.cviu.2022.103412

Wang, Y., Huang, F., Zhang, Y., Feng, R., Zhang, T., & Fan, W. (2020). Deep cascaded cross-modal correlation learning for fine-grained sketch-based image retrieval. In Pattern Recognition (Vol. 100, p. 107148). Elsevier BV. https://doi.org/10.1016/j.patcog.2019.107148

Bressan, R. S., Bugatti, P. H., & Saito, P. T. M. (2022). Optimum-path forest and active learning approaches for content-based medical image retrieval. In Optimum-Path Forest (pp. 95–107). Elsevier. https://doi.org/10.1016/b978-0-12-822688-9.00012-8

Rahman, A., Winarko, E., & Mustofa, K. (2022). Product image retrieval using category-aware siamese convolutional neural network feature. In Journal of King Saud University - Computer and Information Sciences (Vol. 34, Issue 6, pp. 2680–2687).ElsevierBV. https://doi.org/10.1016/j.jksuci.2022.03.005

Jiang, Y., Du, J., Xue, Z., & Li, A. (2022). Cross-Media Retrieval of Scientific and Technological Information Based on Multi-Feature Fusion. In Neurocomputing. Elsevier BV. https://doi.org/10.1016/j.neucom.2022.06.061

Niu, D., Zhao, X., Lin, X., & Zhang, C. (2020). A novel image retrieval method based on multi-features fusion. In Signal Processing: Image Communication (Vol. 87, p. 115911).ElsevierBV.https://doi.org/10.1016/j.image.2020.115911

Zhang, L., & Wu, X. (2022). Multi-task framework based on feature separation and reconstruction for cross-modal retrieval. In Pattern Recognition (Vol. 122, p. 108217). Elsevier BV. https://doi.org/10.1016/j.patcog.2021.108217

Zhang, K., Qi, S., Cai, J., Zhao, D., Yu, T., Yue, Y., Yao, Y., & Qian, W. (2022). Content-based image retrieval with a Convolutional Siamese Neural Network: Distinguishing lung cancer and tuberculosis in CT images. In Computers in Biology and Medicine (Vol. 140, p. 105096). Elsevier BV. https://doi.org/10.1016/j.compbiomed.2021.105096

Yelchuri, R., Dash, J. K., Singh, P., Mahapatro, A., & Panigrahi, S. (2022). Exploiting deep and hand-crafted features for texture image retrieval using class membership. In Pattern Recognition Letters (Vol. 160, pp. 163–171). Elsevier BV. https://doi.org/10.1016/j.patrec.2022.06.017

Chen, Y. , Zhang, Z. , Wang, Y. , Zhang, Y. , Feng, R. , Zhang, T. , & Fan, W. (2022). AE-Net: Fine-grained sketch-based image retrieval via attention-enhanced network. Pattern Recognition, 122, 108291. https://doi.org/10.1016/j.patcog.2021.108291

Zeng, X. , Liu, S. , Wang, X. , Zhang, Y. , Chen, K. , & Li, D. (2021). Hard Decorrelated Centralized Loss for fine-grained image retrieval. Neurocomputing, 453, 26–37. https: //doi. org/10. 1016/j. neucom. 2021. 04. 030

Shan, W., Huang, D., Wang, J., Zou, F., & Li, S. (2022). Self-Attention based fine-grained cross-media hybrid network. In Pattern Recognition (Vol. 130, p. 108748). Elsevier BV. https://doi.org/10.1016/j.patcog.2022.108748

Guo, C. , Lin, Y. , Chen, S. , Zeng, Z. , Shao, M. , & Li, S. (2022). From the whole to detail: Progressively sampling discriminative parts for fine-grained recognition. Knowledge-Based Systems, 235, 107651. https://doi.org/10.1016/j.knosys.2021.107651

Wei, X.-S., Xie, C.-W., Wu, J., & Shen, C. (2018). Mask-CNN: Localizing parts and selecting descriptors for fine-grained bird species categorization. In Pattern Recognition (Vol.76,pp.704–714).ElsevierBV. https://doi.org/10.1016/j.patcog.2017.10.002

Jin, X., Deng, Q., Lou, H., Li, X., & Xiao, C. (2022). Fine-grained Regression for Image Aesthetic Scoring. In CognitiveRobotics.ElsevierBV. https://doi.org/10.1016/j.cogr.2022.07.003

Ha, Y. , Du, Z. , & Tian, J. (2022). Fine-grained interactive attention learning for semi-supervised white blood cell classification. Biomedical Signal Processing and Control, 75, 103611. https://doi.org/10.1016/j.bspc.2022.103611

Zhao, Q. , Wang, X. , Lyu, S. , Liu, B. , & Yang, Y. (2022). A feature consistency driven attention erasing network for fine-grained image retrieval. Pattern Recognition, 128, 108618. https: //doi. org/10. 1016/j. patcog. 2022. 108618

Yan, T. , Shi, J. , Li, H. , Luo, Z. , & Wang, Z. (2022). Discriminative information restoration and extraction for weakly supervised low-resolution fine-grained image recognition. Pattern Recognition, 127, 108629. https://doi.org/10.1016/j.patcog.2022.108629

Ji, J. , Guo, Y. , Yang, Z. , Zhang, T. , & Lu, X. (2021). Multi-level dictionary learning for fine-grained images categorization with attention model. Neurocomputing, 453, 403–412. https://doi.org/10.1016/j.neucom.2020.07.147

Zeng, X. , Zhang, Y. , Wang, X. , Chen, K. , Li, D. , & Yang, W. (2019). Fine Grained Image Retrieval via Piecewise Cross Entropy loss. Image and Vision Computing. https://doi.org/10.1016/j. imavis.2019.10. 006

liu, X. , Wang, L. , & Han, X. (2022). Transformer with peak suppression and knowledge guidance for fine-grained image recognition. Neurocomputing, 492, 137–149. https://doi.org/10.1016/j.neucom.2022.04.037

Chen, H. , Sun, C. , Liao, P. , Lai, Y. , Fan, F. , Lin, Y. , Deng, Z. , & Zhang, Y. (2022). A fine-grained network for human identification using panoramic dental images. Patterns, 100485. https: //doi. org/10. 1016/j. patter. 2022. 100485

Tang, H., Yuan, C., Li, Z., & Tang, J. (2022). Learning attention-guided pyramidal features for few-shot fine-grained recognition. In Pattern Recognition (Vol. 130, p. 108792).ElsevierBV. https://doi.org/10.1016/j.patcog.2022.108792

Zhou, Y. , Li, X. , Zhou, Y. , Wang, Y. , Hu, Q. , & Wang, W. (2022). Deep collaborative multi-task network: A human decision process inspired model for hierarchical image classification. Pattern Recognition, 124, 108449. https: //doi. org/10. 1016/j. patcog. 2021. 108449

Baffour, A. A. , Qin, Z. , Wang, Y. , Qin, Z. , & Choo, K. -K. R. (2021). Spatial self-attention network with self-attention distillation for fine-grained image recognition. Journal of Visual Communication and Image Representation, 81, 103368. https: //doi. org/10. 1016/j. jvcir. 2021. 103368

Peng, J., Wang, Y., & Zhou, Z. (2022). Progressive Erasing Network with consistency loss for fine-grained visual classification. In Journal of Visual Communication and Image Representation (Vol. 87, p. 103570). Elsevier BV. https://doi.org/10.1016/j.jvcir.2022.103570

Xu, S., Muselet, D., & Trémeau, A. (2022). Sparse coding and normalization for deep Fisher score representation. In Computer Vision and Image Understanding (Vol. 220, p. 103436).ElsevierBV. https://doi.org/10.1016/j.cviu.2022.103436

Ahmad, J., Muhammad, K., Bakshi, S., & Baik, S. W. (2018). Object-oriented convolutional features for fine-grained image retrieval in large surveillance datasets. In Future Generation Computer Systems (Vol. 81, pp. 314–330).ElsevierBV. https://doi.org/10.1016/j.future.2017.11.002

1. Text Based Image Retrieval [↑](#footnote-ref-1)
2. Content Based Image Retrieval [↑](#footnote-ref-2)
3. Fine-Grained Content Based Image Retrieval [↑](#footnote-ref-3)
4. Zero-Shot Sketch based Image Retrieval (ZS-SBIR) [↑](#footnote-ref-4)
5. Asymmetric Disentanglement [↑](#footnote-ref-5)
6. bi-directional domain translation [↑](#footnote-ref-6)
7. Deep Cascaded Cross-modal Ranking Model [↑](#footnote-ref-7)
8. annotated [↑](#footnote-ref-8)
9. Deep FG-SBIR Model [↑](#footnote-ref-9)
10. # Optimum-path forest

    [↑](#footnote-ref-10)
11. active learning approaches [↑](#footnote-ref-11)
12. Informative [↑](#footnote-ref-12)
13. Siamese Convolutional Network [↑](#footnote-ref-13)
14. Multi-feature Fusion based Cross-Media Retrieval [↑](#footnote-ref-14)
15. Adversarial Learning [↑](#footnote-ref-15)
16. traditional micro-structure descriptor [↑](#footnote-ref-16)
17. Uniform local binary patterns [↑](#footnote-ref-17)
18. multi-task framework based on feature separation and reconstruction [↑](#footnote-ref-18)
19. Self-Attention Network [↑](#footnote-ref-19)
20. Cross-Media [↑](#footnote-ref-20)
21. common attention space [↑](#footnote-ref-21)
22. similarity concatenation [↑](#footnote-ref-22)
23. progressively sampling discriminative parts [↑](#footnote-ref-23)
24. Sampling [↑](#footnote-ref-24)
25. Deep Descriptors [↑](#footnote-ref-25)
26. Mask-Convolutional Neural Network [↑](#footnote-ref-26)
27. Image Aesthetic Scoring [↑](#footnote-ref-27)
28. Hash Learning Method [↑](#footnote-ref-28)
29. Feature Consistency Driven Attention Erasing Network: FCAENet [↑](#footnote-ref-29)
30. discriminative information restoration and extraction network (DRE-Net) [↑](#footnote-ref-30)
31. Fine-Grained discriminative Information Restoration: FDR [↑](#footnote-ref-31)
32. Semantic Relation Distillation Loss: SRD-Loss [↑](#footnote-ref-32)
33. critical texture details [↑](#footnote-ref-33)
34. Multi-level Attention Model [↑](#footnote-ref-34)
35. typical receptive field sizes [↑](#footnote-ref-35)
36. Cross Entropy Loss [↑](#footnote-ref-36)
37. convolutional neural network [↑](#footnote-ref-37)
38. Piecewise Cross Entropy loss [↑](#footnote-ref-38)
39. effective bidirectional pyramid architecture [↑](#footnote-ref-39)
40. Few-Shot [↑](#footnote-ref-40)
41. Attention-Guided Refinement Strategy [↑](#footnote-ref-41)
42. conditioned by limited samples [↑](#footnote-ref-42)
43. Hierarchical classification [↑](#footnote-ref-43)
44. Fine-Graind [↑](#footnote-ref-44)
45. coarse-grained [↑](#footnote-ref-45)
46. deep collaborative multi-task network [↑](#footnote-ref-46)
47. auxiliary data [↑](#footnote-ref-47)
48. Bounding Box [↑](#footnote-ref-48)
49. End-to-end Spatial Self-Attention Network: SSANet [↑](#footnote-ref-49)
50. Spatial Self-Attention: SSA [↑](#footnote-ref-50)
51. Self-attention distillation: Self-AD [↑](#footnote-ref-51)
52. Progressive Erasing Network [↑](#footnote-ref-52)
53. great significance [↑](#footnote-ref-53)