**بازیابی ریزدانه‌ای تصویر مبتنی بر محتوا به کمک شبکه خودتوجهی مکانی**

سید نیما سید آقا یزدی1، نام و نام خانوادگی نویسنده دوم2، نام و نام خانوادگی نویسنده سوم3

1 رتبه علمی نویسنده در صورت تمایل، گروه آموزشی یا واحد سازمانی مربوطه، نام سازمان ، شهر،

آدرس پست الکترونیکی

2 رتبه علمی نویسنده در صورت تمایل، گروه آموزشی یا واحد سازمانی مربوطه، نام سازمان ، شهر

آدرس پست الکترونیکی

3 رتبه علمی نویسنده در صورت تمایل، گروه آموزشی یا واحد سازمانی مربوطه، نام سازمان ، شهر

آدرس پست الکترونیکی

چكیده

در این مقاله، شیوه نگارش یك مقاله برای کنفرانس پردازش سیگنال و سیستم‌های هوشمند تشریح می‌شود. روش قالب‌بندی مقاله، بخش‌های مختلف آن، انواع قلم‌ها و اندازه آن‌ها، به طور كامل مشخص شده است. كلیه سبك **(Style)** های مورد نیاز برای بخش‌های مختلف مقاله، از جمله عنوان‌ها، نویسندگان، چكیده، متن، و ... از پیش تعریف شده‌اند و تنها كافی است سبك مورد نظر را برای بخشی از مقاله انتخاب كنید. نویسندگان محترم مقاله‌ها باید توجه داشته باشند، كنفرانس از پذیرش مقاله‌هایی كه خارج از این چارچوب تهیه شده باشند، معذور است. چكیده مقاله باید در یك بند (پاراگراف) تهیه شود و حداكثر شامل 200 كلمه باشد. چكیده باید بطور صریح و شفاف موضوع پژوهش و نتایج آن را مطرح كند؛ یعنی بیان كند چه كاری، چگونه، و برای چه هدفی انجام و چه نتایجی حاصل شده است. در چكیده از ذكر جزییات كار، شكل‌ها، جدول‌ها، فرمول‌ها، و مراجع‌ پرهیز كنید.

كلمات كلیدی

بازیابی تصویر – بازیابی ریزدانه‌ای تصویر – بازیابی تصویر مبتنی بر محتوا - عنوان

# مقدمه

یکی از مهم‌ترین شاخه‌های پردازش تصویر، بازیابی تصاویر می‌باشد. 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]](#endnote-1) معرفی گردید. پس از آن رویکردی متفاوت، با عنوان مبتنی بر محتوا[[2]](#endnote-2) معرفی شد که بر اساس ویژگی‌های استخراج شده از تصاویر، کار می‌کرد. این رویکرد به‌سرعت جایگزین رویکرد پیشین شد و در حوزه‌های پزشکی، احراز هویت، پیشگیری از وقوع جرم، امنیت محیط و... مورداستفاده قرار گرفت. در این میان چالش‌های بسیاری به هنگام استفاده از روش‌های مبتنی بر این رویکرد، پیش می‌آمد. از جمله آنکه 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]](#endnote-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 دسته‌ عمده‌ بیان نمود: بازیابی تصویر مبتنی بر طرح، بازیابی تصویر مبتنی بر محتوا و بازیابی تصویر مبتنی بر ریزدانه. این دسته بندی‌ها هر کدام دارای زیرروش های مختلفی هستند که می‌توان آن‌ها را با توجه به انواع استخراج ویژگی، پردازش ویژگی و طبقه‌بندی تصاویر، دسته‌بندی کرد. در ادامه به بررسی نمونه‌هایی از این دسته‌ها پرداخته‌شده‌است.

در دسته روش‌های مبتنی بر طرح، در [10] یک مدل بازیابی تصویر مبتنی بر طرح با نام صفرشات را بررسی می‌کند که در آن دسته‌های آزمایشی در مرحله آموزش ظاهر نمی‌شوند. سعی می‌شود از طریق گسستگی نامتقارن[[4]](#endnote-4) 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 بین حوزه طرح و حوزه تصویر انجام می‌شود. در [11] بازیابی تصویر ریزدانه‌ای مبتنی بر طرح به‌عنوان یک فرایند درشت به ریز فرموله شده است و یک مدل رتبه‌بندی متقابل با نام آبشاری عمیق[[5]](#endnote-5) پیشنهاد می‌شود که می‌تواند از تمام اطلاعات چندوجهی مفید در طرح‌ها و تصاویر حاشیه‌نویسی بهره‌برداری کند و کارایی بازیابی را بهبود بخشد. هدف ساختن بازنمایی‌های عمیق برای طرح‌ها، تصاویر، توضیحات و یادگیری همبستگی‌های عمیق بهینه شده در چنین حوزه‌های مختلف، متمرکز است؛ بنابراین برای یک طرح ورودی داده شده، تصاویر مربوط به آن با شباهت‌های ریز در سطح نمونه، در یک دسته خاص می‌توانند برگردانده شوند.

در دسته روش‌های مبتنی بر محتوا، در [1] استفاده از جنگل‌های مسیر بهینه (بدون نظارت و با نظارت) و رویکردهای یادگیری فعال را برای بازخورد مرتبط در سیستم‌های بازیابی تصویر پزشکی مبتنی بر محتوا بررسی می‌کند. آموزنده‌ترین 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 ارائه می‌دهند. در [2] هدف ایجاد یک روش بازیابی تصویر برای مشخص‌کردن دسته‌بندی یک محصول است و یک مدل شبکه کانولوشنی سیامی[[6]](#endnote-6) پیشنهاد می‌شود که شامل برچسب‌های دسته و مورد در آموزش برای تولید ویژگی آگاه از دسته است. این مدل با اصلاح رویه آموزشی همراه می‌باشد که به طور هم‌زمان دسته و برچسب مورد را یاد می‌گیرد.

در [5] یک روش بازیابی متقابل رسانه‌ای مبتنی بر ترکیب چند ویژگی[[7]](#endnote-7) پیشنهاد می‌شود. این روش قادر به ادغام چندین ویژگی برای ارتقای درک معنایی و اتخاذ یادگیری متخاصم برای بهبود بیشتر دقت بازنمایی زیرفضای عمومی است. سپس از شباهت در همان فضا برای مرتب‌سازی نتایج بازیابی استفاده می‌شود. در [6] یک روش بازیابی تصویر مبتنی بر محتوا پیشنهاد می‌شود. در مرحله توصیف تصویر، این روش ابتدا توصیفگر ریزساختار سنتی را اصلاح می‌کند تا رابطه مستقیم بین ویژگی‌های شکل و بافت و بین ویژگی‌های رنگ و بافت را به تصویر بکشد. سپس هیستوگرام الگوهای باینری محلی یکنواخت[[8]](#endnote-8) تصویر را استخراج می‌کند تا اطلاعات تفاوت رنگ را به تصویر بکشد.

در [7] یک چارچوب چند وظیفه‌ای مبتنی بر جداسازی و بازسازی ویژگی [[9]](#endnote-9) برای بازیابی متقابل وجهی بر اساس روش‌های رایج یادگیری مکانی پیشنهاد می‌شود که ماژول جداسازی ویژگی را برای مقابله با عدم تقارن اطلاعات بین روش‌های مختلف معرفی می‌کند. در [8] بازیابی تصویر مبتنی بر محتوا با یک شبکه عصبی سیامی کانولوشنی پیشنهاد می‌شود. ابتدا، تکه‌های ضایعه برای ایجاد دو مجموعه‌های داده، برش داده می‌شوند و جفت‌های دوتکه دلخواه یک مجموعه‌داده پچ-جفت را تشکیل می‌دهند. دوم، این مجموعه‌داده پچ-جفت 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. در [9] روشی پیشنهاد می‌شود که از قدرت شبکه‌های عصبی کانولوشن برای پیش‌بینی دسته‌بندی تصویر ورودی برای همه کلاس‌های خروجی و بازیابی تصاویر با استفاده از تابع فاصله تغییریافته در فضای ویژگی موجک، استفاده می‌کند.

در [13] بافت کانال و اطلاعات توالی مکانی برای بازیابی مبتنی بر محتوا مورد تمرکز قرار می‌گیرند. ابتدا یک مدل عمیق پیشنهاد می‌شود که هدف آن استنباط نقشه‌های توجه، در امتداد ابعاد کانال و مکان است. با بهبود ماژول‌های توجه کانال و توجه مکانی و کاوش ترانسفورماتور، توانایی ساخت و درک مدل افزایش می‌یابد. در [14] با اشاره به روش‌هایی که با خطای ویژگی‌های عمومی به استخراج ویژگی‌های متمایزتر کمک می‌کنند، یک تابع محاسبه خطای جدید به نام خطای متمرکز سخت ارائه می‌شود. این تابع در استخراج ویژگی برای تمایز در تقسیم مشابه‌ترین دسته‌ها کمک می‌کند. در [15] یک ژنراتور لنگر استخراج ویژگی محلی[[10]](#endnote-10) برای شبیه‌سازی اشکال ویژگی‌های نامنظم پیشنهاد می‌شود؛ بنابراین، ویژگی‌های متمایز را می‌توان به طور کامل در ویژگی‌های استخراج شده گنجاند. علاوه بر این، یک ماژول استخراج ویژگی محلی متقارن مؤثر[[11]](#endnote-11) بر اساس مکانیزم توجه پیشنهاد شده است تا به طور کامل از رابطه مکانی بین ویژگی‌های محلی استخراج‌شده، استفاده و ویژگی‌های متمایز را برجسته کند.

در [16] به طبقه‌بندی تصاویر با استفاده از یادگیری انتقالی و شبکه عصبی کانولوشن دوخطی پرداخته می‌شود. از شبکه متقارن VGG16 برای استخراج ویژگی بهره می‌گیرد و پس از آموزش به یک چارچوب پیشنهادی منتقل می‌کند. سپس ویژگی‌های عمومی را از دو شبکه گرفته و مورد بررسی قرار می‌دهد. در [17] روش یادگیری هش با دو مشکل بررسی می‌شود: 1- ویژگی‌های 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 به یک شبکه پاک‌کننده توجه مبتنی بر ثبات ویژگی[[12]](#endnote-12) پرداخته‌می‌شود. برای مشکل نخست، از یک ماژول پاک‌کردن ناحیه انتخاب‌شده[[13]](#endnote-13) استفاده می‌کند که با پوشش تطبیقی برخی از مناطق تصاویر خام، شبکه را در برابر تفاوت‌های ظریف ریزدانه‌ای مقاوم می‌کند. بدین ترتیب کدهای هش متمایزتری در پایگاه‌داده هش ذخیره می‌شوند. سپس برای پایدارترکردن رابطه بین کد هش تصویر ورودی و کد هش پایگاه‌داده، از ماژول افزایش خطای رابطه مکانی[[14]](#endnote-14) استفاده می‌کند.

در [18] به طور هم‌زمان، نشانه‌های سطحی شیء[[15]](#endnote-15) و نشانه‌های سطحی بخشی[[16]](#endnote-16) از تصاویر را ضبط می‌کند. نشانه‌های سطح شیء تصاویر محصول توسط، یک شبکه جدید طبقه‌بندی بازسازی[[17]](#endnote-17) تولید می‌شود. برای مدل‌سازی بدون حاشیه‌نویسی نشانه‌های سطح جزئی، قسمت‌های تبعیض‌آمیز، تصاویر محصول در اطراف نقاط کلیدی شناسایی می‌شوند. این بخش‌ها به‌صورت توالی‌های مرتب‌شده توسط یک حافظه کوتاه‌مدت-بلندمدت کانولوشنی، کدگذاری می‌شوند و محصولات را به طور منحصربه‌فرد توصیف می‌کنند. در [19] یک شبکه ترکیبی مبتنی بر خودتوجهی[[18]](#endnote-18) برای یادگیری بازنمایی‌های رایج داده‌های رسانه‌های مختلف[[19]](#endnote-19) پیشنهاد می‌شود. به طور خاص، ابتدا از یک لایه خودتوجهی محلی برای یادگیری فضای توجه مشترک بین داده‌های رسانه‌های مختلف استفاده می‌شود. سپس یک روش الحاق شباهت برای درک رابطه محتوایی بین ویژگی‌ها پیشنهاد می‌شود. برای بهبود بیشتر استحکام مدل، یک کدگذاری موقعیت محلی را یاد می‌گیرد تا روابط مکانی بین ویژگی‌ها را ثبت کند؛ بنابراین، این روش می‌تواند به طور مؤثر، شکاف بین توزیع‌های ویژگی‌های مختلف در وظایف بازیابی بین رسانه‌ای را کاهش دهد.

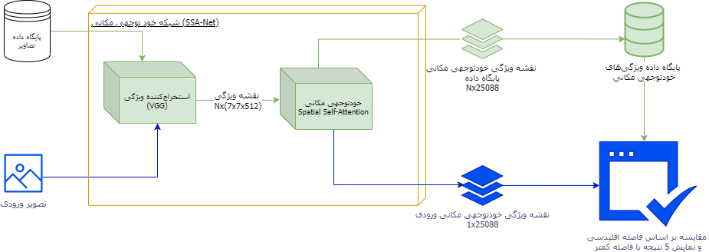
در[20] یک چارچوب سبک‌تر برای نمونه‌برداری تدریجی از قطعات متمایز، جهت یادگیری جزئیات ارائه می‌شود. در این روش ابتدا شیء از تصویر اصلی تقویت‌شده و سپس یک نمونه‌برداری خودتطبیقی برای شناسایی بیشتر منطقه تقویت‌شده انجام می‌گردد. پس این چارچوب می‌تواند از کل به شیء و از شیء به جزئیات برسد. در این میان ویژگی‌های سلسله‌مراتبی نیز سنجیده می‌شوند که هزینه‌های محاسباتی را کاهش می‌دهد. در [12] که روی طبقه‌بندی تصاویر تمرکز دارد، یک سیستم یادگیری نیمه نظارت تهیه‌شده است. 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.

در دسته روش‌های مبتنی بر ریزدانه، در [3] انتخاب توصیف‌گرهای عمیق مفید به خوبی به تشخیص تصویر با دانه‌ریز کمک می‌کند. به طور خاص، یک مدل جدید شبکه عصبی کانولوشنی ماسک دار[[20]](#endnote-20)، بدون لایه‌های کاملاً متصل پیشنهاد شده است. بر اساس حاشیه‌نویسی‌های بخش، مدل پیشنهادی 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. پس از آن، یک مدل سه‌جریانی برای تجمیع توصیفگرهای انتخاب شده در سطح جسم و قطعه به طور هم‌زمان ساخته می‌شود. با کنار گذاشتن پارامتر لایه‌های کاملاً متصل اضافی، این شبکه دارای ابعاد کوچک و سرعت استنتاج کارآمد در مقایسه با سایر روش‌های ریزدانه است. در [4] یک روش تخمین ریزدانه، برای تخمین نمره زیبایی‌شناسی پیشنهاد می‌شود و مکانیسم‌های توجه موقعیت و کانال را برای افزایش ترکیب ویژگی‌های زیبایی‌شناسی ترکیب می‌کند. با آموزش شبکه رگرسیون، جدا از شبکه طبقه‌بندی، طبقه‌بندی وظیفه رگرسیون را تکمیل می‌کند. محققان به استفاده از میانگین مربع خطا[[21]](#endnote-21) به‌عنوان معیار ارزیابی اصلی عادت کرده‌اند، که در اندازه‌گیری خطای هر بازه ناکافی است. به‌منظور درنظرگرفتن کامل تصاویر، بخش‌های مختلف امتیاز زیبایی‌شناسی، به‌جای تمرکز بر بخش‌های این نمره‌ها متوسط ​​به دلیل عدم تعادل مجموعه‌داده‌ها ، یک معیار ارزیابی جدید به نام خطاهای میانگین مربع تقسیم شده[[22]](#endnote-22) برای اثبات مزایا پیشنهاد می‌شود.

در ادامه در بخش 2 به مرور مقاله پایه پرداخته می‌شود. در بخش 3 راهکار پیشنهادی مبتنی بر روش برجستگی ارائه می‌گردد. در بخش 4 به ارزیابی دقت و پایداری نتایج آزمایشگاهی و در نهایت در بخش 5 به جمع بندی این مقاله پرداخته شده است.

# بازیابی تصویر ریزدانه‌ای با استفاده از شبکه خودتوجهی مکانی و مکانیسم برجسته‌سازی

تغییر کند



در [21] شبکه خودتوجهی مکانی[[23]](#endnote-23) پیشنهاد شده‌است که شامل دو جزء اصلی می‌باشد. ابتدا یک شبکه عصبی کانولوشنی به‌عنوان استخراج‌کننده ویژگی[[24]](#endnote-24) پیاده‌سازی می‌شود که ویژگی‌های اولیه را از تصاویر ورودی از طریق چندین لایه کانولوشن استخراج می‌کند. سپس ماژول خودتوجهی مکانی با استفاده از مکانیسم توجه، ویژگی‌های جدید را ذخیره می‌کند.

## 1-2-استخراج‌کننده ویژگی

اخیراً، برای وظایف پردازش تصویر، یک رویکرد مرسوم برای استخراج ویژگی‌های اولیه، استفاده از یک شبکه عصبی کانولوشنی از قبل آموزش‌دیده، به‌منظور بهره‌مندی از مقدار اولیه وزن معنادار است. چنین شبکه‌های عصبی کانولوشنی می‌توانند ویژگی‌های سطح بالا را از تصاویر استخراج کنند. برای مقایسه منصفانه با سایر روش‌های پیشرفته، از VGG-16 از پیش آموزش‌دیده بر روی مجموعه‌داده ImageNet استفاده می‌شود.

برای استخراج اولیه، سه‌لایه آخر که کاملاً متصل هستند حذف می‌شوند. نقشه‌های ویژگی هر تصویر از مجموعه‌داده به عنوان X از خروجی لایه کانولوشنی نهایی گرفته‌می‌شود. این فرآیند به صورت زیر نشان داده شده است

|  |  |
| --- | --- |
| (1) | 𝐹 = 𝑉𝐺𝐺(𝑋) |

به طور خاص، استخراج‌کننده ویژگی، یک تصویر ورودی 𝑋 را به یک نقشه ویژگی با ابعاد 𝐹 ∈ نگاشت می‌کند، که در آن 𝐻، W و 𝐾 نشان دهنده ارتفاع مکانی، عرض مکانی، تعداد کانال ها/کرنل حاوی کانال هستند.

## 2-2-ماژول خودتوجهی مکانی

ماژول خودتوجهی مکانی از مکانیسم خودتوجهی پیشنهاد شده استفاده می‌کند که توجه محلی را از طریق یک تابع فعال‌ساز softmax جمع می‌کند. این ایده گسترش می‌یابد تا به موقعیت‌های پیکسل مکانی ویژگی‌های اصلی توجه و از تجمیع ویژگی‌ها برای به‌دست‌آوردن نقشه‌های ویژگی خودتوجهی مکانی استفاده شود. باتوجه‌به نقشه‌های ویژگی اولیه 𝐹 ∈ به دست آمده از استخراج‌کننده ویژگی، ابتدا سه نقشه ویژگی جدید 𝐴، 𝐵 و 𝐶 با استفاده از کانولوشن 1×1 تولید می‌شود. {A,B,C} ∈ همان ابعاد فضای F را داراست. سپس A و B و C به تغییر شکل می‌یابد، که در آن 𝑁 = 𝐻x𝑊 تعداد پیکسل‌ها است. سپس، ضرب عناصر بین 𝐴 و ترانهاده 𝐵 محاسبه می‌شود. تابع فعال‌ساز softmax از نظر مکانی برای محاسبه نقشه خودتوجهی مکانی اعمال می‌شود 𝑆 ∈ که:

|  |  |
| --- | --- |
| (2) |  |

در فرمول (2) ⊗ نشان‌دهنده ضرب عنصر به عنصر است. نشان می‌دهد که چگونه شبکه تاثیر 𝑖مین موقعیت مکانی را بر موقعیت مکانی 𝑗مین اندازه گیری می‌کند. از این رو، بازنمایی ویژگی‌های مرتبط‌تر بین 𝐴 و 𝐵 منجر به همبستگی معنی‌دار و غنی‌تر بین آنها می‌شود و بالعکس. برای تقویت موقعیت‌های مکانی، ضرب عناصر بین 𝑆 ∈ و 𝐶 ∈ انجام می‌شود و نتایج به تغییر شکل داده‌می‌شود.

در نهایت، یک مکانیسم تجمیع ویژگی برای بررسی تأثیر مناطق خودتوجهی مکانی در همه موقعیت‌ها، در نقشه ویژگی اصلی از طریق فرمول (3) پیاده‌سازی می‌شود:

|  |  |
| --- | --- |
| (3) |  |

می‌توان از فرمول (3) استنباط کرد که ویژگی‌های به‌دست‌آمده توسط نشان‌دهنده یک تجمع کلی از نمای زمینه‌ای بر اساس نقشه‌های خودتوجهی مکانی است. مجموعه این ویژگی‌ها به عنوان یک پایگاه داده ذخیره می‌شوند.

## 3-2-بازیابی تصویر

در این بخش یک تصویر به عنوان ورودی به شبکه داده می‌شود و طبق معادلات (1)، (2) و (3) نقشه ویژگی‌های خودتوجهی مکانی آن به دست می‌آید. سپس با نقشه‌های ویژگی ذخیره شده در قسمت 2-2 و با استفاده از معادله زیر مقایسه می‌شوند:

|  |  |
| --- | --- |
| (4) |  |

در فرمول (4) X نقشه‌ ویژگی خودتوجهی مکانی تصویر ورودی و Y نقشه ویژگی خودتوجهی مکانی هر تصویر از پایگاه داده است. سپس فاصله‌های به‌دست‌آمده، که هرکدام نگاشتی به تصویری از پایگاه داده دارند، به صورت نزولی مرتب شده و پنج نتیجه برتر بازیابی می‌شود. خروجی سیستم بر اساس کلاسی که بیشترین احتمال را در بین این پنج نتیجه دارد، تعیین می‌گردد.

# 4-2-مکانیسم برجسته‌سازی

یکی از مشکلات روش استفاده از شبکه خودتوجهی مکانی آن است که تصویر ورودی، با ویژگی‌های با اهمیت کم‌تر بررسی می‌شود و ممکن است بخش‌های حاشیه‌ای در نتیجه نهایی عملکرد شبکه، تأثیرگذار باشند.

در این مقاله، روش XRAI برجسته‌سازی[[25]](#endnote-25) پیشنهاد شده‌است. این روش از گراف Felzenswalb برای تقسیم‌بندی بهره می‌گیرد. روش‌های تقسیم‌بندی، معمولاً دارای چندین مجموعه از پارامترها هستند که تعداد و شکل بخش‌ها را تغییر می‌دهند. از آنجا که امکان‌پذیر نیست نتایج انتساب، به مجموعه خاصی از پارامترهای فوق یا کیفیت روش تقسیم‌بندی، بستگی داشته باشد، تصویر چندین بار با استفاده از مجموعه پارامترهای مختلف قطعه‌بندی می‌شود. به طور خاص، از یک پارامتر مقیاس در مجموعه [50، 100، 150،250، 500، 1200] استفاده شده و بخش‌های کوچکتر از بیست پیکسل نادیده گرفته‌می‌شود (پارامتر مقیاس عمدتاً بر اندازه بخش‌ها تأثیر می‌گذارد). یک پارامتر واحد، اجتماع بخش‌ها در کل تصویر را محاسبه می‌کند. بنابراین، اجتماع همه بخش ها مساحتی برابر با شش برابر مساحت تصویر ایجاد می‌کند و در نتیجه بخش‌های جداگانه به طور قابل توجهی همپوشانی دارند. مرزهای بخش معمولاً با لبه‌های تصویر همسو می‌شوند. برای استخراج نقشه‌های برجسته، مطلوب است که بخش‌ها شامل لبه‌ها باشند، زیرا تصاویر در دو طرف یک لبه نازک اغلب به یکدیگر مرتبط هستند. برای این منظور، ماسک های بخش‌ها پنج پیکسل گسترش می‌یابند تا مجموعه نهایی قطعات به دست آید.

الگوریتم

# ارزیابی دقت و نتایج آزمایشگاهی

در تکمیل بهبود استفاده از شبکه خودتوجهی مکانی، از روش شفاف‌سازی XRAI استفاده شده‌است. مجموعه داده Stanford Dogs [22] شامل 20580 تصویر از 120 نژاد سگ از سراسر جهان است. این مجموعه‌داده با استفاده از تصاویر و حاشیه نویسی از ImageNet، برای طبقه بندی تصاویر ریز دانه ساخته شده است. این مجموعه‌داده که در ابتدا برای دسته‌بندی تصاویر ریزدانه جمع آوری شد، یک مشکل چالش برانگیز داشت، زیرا برخی از نژادهای سگ ویژگی های تقریباً یکسانی دارند یا از نظر رنگ و سن متفاوت هستند.

در این بخش یک تصویر به عنوان ورودی به شبکه داده می‌شود و طبق معادلات (1)، (2) و (3) نقشه ویژگی‌های خودتوجهی مکانی آن به دست می‌آید. سپس با نقشه‌های ویژگی ذخیره شده در قسمت 2 و با استفاده از معادله زیر مقایسه می‌شوند.

|  |  |
| --- | --- |
| (5) |  |

فرمول محاسبه امتیاز

در فرمول (5)، X نقشه‌ ویژگی خودتوجهی مکانی تصویر ورودی و Y نقشه ویژگی خودتوجهی مکانی هر تصویر از پایگاه داده است. سپس فاصله‌های به‌دست‌آمده، که هرکدام نگاشتی به تصویری از پایگاه داده دارند، به صورت نزولی مرتب شده و 5 نتیجه برتر بازیابی می‌شود. خروجی سیستم بر اساس کلاسی که بیشترین احتمال را در بین این 5 نتیجه دارد، تعیین می‌گردد.

در آزمایش پیش‌ رو، این پایگاه‌داده ازآن‌جهت که ساختار سلسله‌مراتبی داشته و تا حد بالایی رفتارهای ریزدانه‌ای ارائه می‌دهد، انتخاب شده است. در این آزمایش از یک کامپیوتر شخصی مدل Macbook Air با پردازنده M1 و حافظه اصلی 8 گیگابایت استفاده شده است. محیط توسعه و زبان مورد استفاده python می‌باشد.



شکل 1. نمونه‌ای از تصاویر مجموعه‌داده [23]

جدول 1 بیان‌کننده مقایسه نتایج عملکرد روش‌های مختلف در بازیابی تصویر ریزدانه‌ای بر روی مجموعه‌داده Stanford Dogs می‌باشد. در روش EfficeintNet-B0 از کانولوشن‌های قابل تفکیک عمیق استفاده شده و در میان سایر روش‌ها از عملکرد ضعیف‌تری برخوردار است. در روش PC از آموزش با تقسیم‌بندی‌های پیچیده، در جهت ریزدانه‌ای کردن بازیابی تصویر استفاده‌شده‌است که نسبت به روش پیشین 0.7 درصد بهبود داشته‌است.

جدول1. نتایج آزمایشگاهی

|  |  |
| --- | --- |
| Score | Method |
| 84.5% | FCAN [24] |
| 71.9% | PDFR [25] |
| 83.6% | PC-DenseNet-161[26] |
| 79.6% | HDWE [27] |
| 61.2% | EfficientNet-B0[28] |
| 61.9% | PC [29] |
| 86% | SSA [30] |
| 88% | روش پیشنهادی |

مراجع

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). Elsevier BV. https://doi.org/10.1016/j.jksuci.2022.03.005

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). Elsevier BV. 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 Cognitive Robotics. Elsevier BV. https://doi.org/10.1016/j.cogr.2022.07.003

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). Elsevier BV. 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

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

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

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

Yang, M., Xu, Y., Wu, Z., & Wei, Z. (2022). Symmetrical irregular local features for fine-grained visual classification. In Neurocomputing (Vol. 505, pp. 304–314). Elsevier BV. https://doi.org/10.1016/j.neucom.2022.07.056

Yuan, P. , Qian, S. , Zhai, Z. , FernánMartínez, J. , & Xu, H. (2022). Study of chrysanthemum image phenotype on-line classification based on transfer learning and bilinear convolutional neural network. Computers and Electronics in Agriculture, 194, 106679. https: //doi. org/10. 1016/j. compag. 2021. 106679

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

Santra, B. , Shaw, A. K. , & Mukherjee, D. P. (2022). Part-based annotation-free fine-grained classification of images of retail products. Pattern Recognition, 121, 108257. https://doi.org/10.1016/j.patcog.2021.108257

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.2021.1076

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

Khosla, A., Jayadevaprakash, N., Yao, B., & Fei-Fei, L. (2012). Novel Dataset for Fine-Grained Image Categorization : Stanford Dogs.

Khosla, A., Jayadevaprakash, N., Yao, B., & Fei-Fei, L. (2012). Novel Dataset for Fine-Grained Image Categorization : Stanford Dogs.

Liu, X., Xia, T., Wang, J., & Lin, Y. (2016). Fully convolutional attention localization networks: Efficient attention localization for fine-grained recognition. arXiv preprint arXiv:1603.06765, 1(2), 4.

X. Zhang, H. Xiong, W. Zhou, W. Lin and Q. Tian, "Picking Deep Filter Responses for Fine-Grained Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 1134-1142, doi: 10.1109/CVPR.2016.128.

Dubey, A., Gupta, O., Guo, P., Raskar, R., Farrell, R., Naik, N. (2018). Pairwise Confusion for Fine-Grained Visual Classification. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science(), vol 11216. Springer, Cham. https://doi.org/10.1007/978-3-030-01258-8\_5

Yu, J., Huang, Y., Gbur, G., Wang, F., & Cai, Y. (2019). Enhanced backscatter of vortex beams in double-pass optical links with atmospheric turbulence. In Journal of Quantitative Spectroscopy and Radiative Transfer (Vol. 228, pp. 1–10). Elsevier BV. https://doi.org/10.1016/j.jqsrt.2019.02.021

D. Haase and M. Amthor, "Rethinking Depthwise Separable Convolutions: How Intra-Kernel Correlations Lead to Improved MobileNets," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 14588-14597, doi: 10.1109/CVPR42600.2020.01461.

Dubey, A., Gupta, O., Guo, P., Raskar, R., Farrell, R., & Naik, N. (2017). Training with confusion for fine-grained visual classification. CoRR

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

زیرنویس‌ها

1. Text Based Image Retrieval [↑](#endnote-ref-1)
2. Content Based Image Retrieval [↑](#endnote-ref-2)
3. Fine-Grained Content Based Image Retrieval [↑](#endnote-ref-3)
4. Asymmetric Disentanglement [↑](#endnote-ref-4)
5. Deep Cascaded Cross-modal Ranking Model [↑](#endnote-ref-5)
6. Siamese Convolutional Network [↑](#endnote-ref-6)
7. Multi-feature Fusion based Cross-Media Retrieval [↑](#endnote-ref-7)
8. Uniform local binary patterns [↑](#endnote-ref-8)
9. multi-task framework based on feature separation and reconstruction [↑](#endnote-ref-9)
10. Local Feature Extraction Anchor Generator [↑](#endnote-ref-10)
11. Symmetrized Local Feature Extraction Module [↑](#endnote-ref-11)
12. Feature Consistency Driven Attention Erasing Network: FCAENet [↑](#endnote-ref-12)
13. Selective Region Erasing Module: SREM [↑](#endnote-ref-13)
14. Enhancing Space Relation Loss: ESRL [↑](#endnote-ref-14)
15. Object-level [↑](#endnote-ref-15)
16. Part-level [↑](#endnote-ref-16)
17. Reconstruction-Classification Network: RC-Net [↑](#endnote-ref-17)
18. Self-Attention Network [↑](#endnote-ref-18)
19. Cross-Media [↑](#endnote-ref-19)
20. Mask-Convolutional Neural Network [↑](#endnote-ref-20)
21. Mean Square Errors [↑](#endnote-ref-21)
22. Segmented Mean Square Errors [↑](#endnote-ref-22)
23. Spatial Self-Attention Network (SSA.Net) [↑](#endnote-ref-23)
24. Feature Extractor: FE [↑](#endnote-ref-24)
25. Saliency [↑](#endnote-ref-25)