تکامل برچسب‌های تصاویر با اعمال خوشه‌بندی فازی تک‌گذر C-Means بر ویژگی‌های یادگیری‌شده توسط شبکه عصبی کانولوشن عمیق

شیما جوانمردی[۱]، دانشجوی دکتری؛ علی‌محمد لطیف[۲]، دانشیار؛ ولی درهمی[۳]، دانشیار

۱- گروه کامپیوتر - دانشگاه یزد - یزد - ایران - sh.javanmardi@stu.yazd.ac.ir
۲- گروه کامپیوتر - دانشگاه یزد - یزد - ایران - alatif@yazd.ac.ir
۳- گروه کامپیوتر - دانشگاه یزد - یزد - ایران - vderhami@yazd.ac.ir

چکیده: تکامل برچسب‌های تصاویر، فرآیندی است که هم‌زمان به غنی‌سازی تگ‌های تصاویر و رفع نویز از آن‌ها می‌پردازد. بسیاری از تصاویر در وب، توسط تگ‌های مبهم و بی‌ارتباط با محتوای تصویر برچسب‌گذاری شده‌اند. وجود این برچسب‌های غیرمرتبط با تصویر، موجب کاهش دقت بازیابی آن‌ها می‌شود. ازاین‌رو در سال‌های اخیر، به‌منظور رفع نویز و تکمیل برچسب‌های تصاویر، الگوریتم‌هایی با عنوان تکامل تگ مطرح‌شده‌اند که هدف آن‌ها دستیابی به برچسب‌های مرتبط با محتوای تصاویر و حذف برچسب‌های غیرمرتبط می‌باشد. با توجه به کارآمدی فرآیند یادگیری عمیق در بسیاری از حوزه‌های پژوهشی، در این مقاله نیز به‌منظور استخراج ویژگی‌های دیداری و معنایی مناسب از تصاویر، از شبکه‌های عصبی کانولوشنال عمیق استفاده شده است. همچنین با توجه به چالش‌های مطرح در بارگذاری مجموعه تصاویر با مقیاس بزرگ در حافظه، به‌منظور دسته‌بندی تصاویر مشابه دیداری و پالایش برچسب‌های هر تصویر با توجه به نمونه‌های مشابه، از الگوریتم خوشه‌بندی فازی تک‌گذر C-Means استفاده شده است. نتایج آزمایش‌ها بیانگر مؤثر بودن رویکرد ارائه‌شده، در فرآیند تکامل برچسب‌های تصاویر می‌باشد.

واژه‌های کلیدی: تکامل تگ تصویر، شبکه عصبی کانولوشنال عمیق، پالایش تگ، خوشه‌بندی فازی تک‌گذر C-Means، بازیابی تصاویر

Image Tag Completion by Applying SPFCM Clustering on the Features Learned by Deep Convolutional Neural Networks

Sh. Javanmardi[1], Ph.D Student; A. M. Latif[2], Associate Professor; V. Derhami[3], Associate Professor

1- Computer Department, Yazd University, Yazd, Iran, Email: sh.javanmardi@stu.yazd.ac.ir
2- Computer Department, Yazd University, Yazd, Iran, Email: alatif@yazd.ac.ir
3- Computer Department, Yazd University, Yazd, Iran, Email: vderhami@yazd.ac.ir

Abstract: Image tag completion is a process that aims to simultaneously enrich the missing tags and remove noisy tags. many of the images have vague, incomplete and irrelevant tags. These untrusted tags, reduce the accuracy of image retrieval. Hence, in recent years, many tag completion algorithms have been proposed in order to access to the tags associated with the content of images. Due to the effectiveness of deep learning in many research fields, in this paper a deep convolutional neural network has been used to extract suitable visual and semantic features of images. Also, considering the challenges involved in loading a large-scale image databases in memory, a Single Pass Fuzzy C-Means clustering algorithm is used in order to compute visually similar images and refining the image's tags according to similar samples. The results show the effectiveness of proposed approach in images tag completion.

Keywords: Image Tag Completion, Deep Convolutional Neural Network, Tag Refinement, Single Pass Fuzzy C-Means Clustering, Image Retrieval.

تاریخ ارسال مقاله: ۱۳۹۶/۰۵/۰۸
تاریخ اصلاح مقاله: ۱۳۹۶/۰۸/۲۲
تاریخ پذیرش مقاله: ۱۳۹۶/۱۰/۳۰
نام نویسنده مسئول: علی‌محمد لطیف
نشانی نویسنده مسئول: ایران - یزد - بلوار پژوهش - دانشگاه یزد - دانشکده مهندسی برق و کامپیوتر- گروه کامپیوتر.

۱ -مقدمه

موفقیت رسانه‌های اجتماعی منجر به دسترس‌پذیری به حجم بالایی از تصاویر برچسب‌گذاری‌شده توسط کاربران شده است. این رسانه‌ها، قابلیت‌هایی را برای کاربران ایجاد می‌کنند که با برچسب‌گذاری تصاویر و فراداده‌ها، نتایج حاصل از فعالیت‌هایی چون بازیابی تصاویر را بهبود دهند. در این شرایط مدیریت مؤثر رسانه‌های اجتماعی و فراداده‌ها با چالش‌هایی روبروست. از جمله چالش‌های مطرح، وجود برچسب‌های مبهم، ناقص و یا متضاد با محتوای تصویر است که با توجه به تنوع کاربران و حجم بالای تصاویر، وقوع این پدیده در فرآیند شرح‌گذاری تصاویر[۱] امری اجتناب‌ناپذیر است. این امر سبب می‌شود تا قابلیت استفاده از سیستم‌های بازیابی تصاویر در موتورهای جستجو با محدودیت مواجه شود.

بازیابی تصاویر به دو صورت بازیابی مبتنی بر محتوا (CBIR)[۲] و بازیابی مبتنی بر برچسب (TBIR)[۳] صورت می‌گیرد [۱]. در CBIR با استفاده از ویژگی‌های دیداری از تصویر همچون رنگ، بافت و لبه‌های اشیای درون تصویر بازیابی تصاویر مرتبط با پرس‌وجوی کاربر انجام می‌شود. از جمله چالش‌های مطرح در CBIR می‌توان به وجود شکاف معنایی میان ویژگی‌های پایین‌رتبه و مفاهیم سطح بالا تصاویر و وجود ابعاد بالای ویژگی‌های دیداری تصاویر اشاره کرد. به‌منظور حل این چالش در [۲] یک روش جستجوی ترکیبی ارائه شده است که با ترکیب چند درخواست تصویری با محتویات متفاوت به بازیابی تصاویر می‌پردازد به‌گونه‌ای که تصاویر بازیابی‌شده حاوی تمام درخواست‌های مورد جستجو باشد. در روش ارائه‌شده در مقاله مذکور نواحی و اشیاء درون تصویر محاسبه شده، سپس همه درخواست‌ها با استفاده از عملگرهای منطقی مبتنی بر نوع جستجو ترکیب شده تا یک بردار باینری برمبنای درخواست‌ها برای جستجوی تصاویر درخواستی ایجاد شود.

در مقابل، TBIR تنها با توجه به اطلاعات متنی تصاویر به جستجوی آن‌ها می‌پردازد. اطلاعات متنی در مقایسه با اطلاعات دیداری با سادگی بیشتری گویای مفاهیم تصویر هستند. اطلاعات متنی استفاده‌شده در فرآیند TBIR از عناوین تصاویر [۳]، متن اطراف آن‌ها [۴] و شرح‌گذاری‌های صورت‌گرفته توسط کاربران حاصل می‌شود [۵]. کارایی TBIR تا حد زیادی به کیفیت تگ‌های اعمال شده توسط کاربران بر تصاویر بستگی دارد. وجود نویز در اطلاعات متنی، شرح‌گذاری‌های غیرمرتبط با محتوای تصویر، عدم وجود اطلاعات مرتبط با تصاویر و رفع نویز از تگ‌های مجموعه‌های تصاویر با مقیاس بالا از جمله چالش‌های مطرح شده در این حوزه می‌باشد. به‌منظور حل چالش سیستم‌های نامطلوب شرح‌گذاری تصاویر در [۶] یک روش شرح‌گذاری مطرح شده است که در آن ابتدا بافتار ناحیه‌ای تصویر را که نشان‌دهنده ارتباط بین نواحی تصویر هستند محاسبه می‌کند و جایگزین توزیع مستقیم نواحی از توزیع موضوعی میان نواحی در شرح‌گذاری تصاویر استفاده می‌کند.

به‌منظور مقابله با مشکلات مطرح در فرآیند شرح‌گذاری، در سال‌های اخیر، الگوریتم‌هایی تحت عنوان تکامل تگ تصویر (ITC)[۴] مطرح شده‌اند که به رفع ابهام از شرح‌گذاری‌های صورت‌گرفته توسط روش‌های شرح‌گذاری خودکار تصاویر می‌پردازند. در این رویکردها، تگ‌های غیرمرتبط با مفاهیم تصویر از میان تگ‌های اولیه ایجادشده توسط کاربران، حذف می‌شوند و عملیات غنی‌سازی تگ‌ها با دیگر برچسب‌های مرتبط صورت می‌گیرد. این فرآیند تا حد زیادی می‌تواند باعث کاهش شکاف معنایی میان ویژگی‌های سطح پایین و مفاهیم معنایی سطح بالای تصاویر شود، بنابراین رویکردی مؤثر در فرآیند بازیابی تصاویر مبتنی بر برچسب به‌شمار می‌آید.

یکی از چالش‌های مطرح در حوزه تکامل تگ‌های تصاویر، استخراج ویژگی‌های مهم از تصویر می‌باشد. تمام ویژگی‌های کنونی دارای محدودیت در توصیف تصاویر هستند و تحلیل و پردازش ویژگی‌های تصاویر با ابعاد بالا، مسئله پیچیده‌ای است. با توجه به قابلیت بالای فرآیند یادگیری عمیق (DL)[۵] در حوزه استخراج ویژگی‌های سطح بالا و مرتبط با محتوای تصویر، در این مقاله نیز به‌منظور استخراج ویژگی‌های کارآمد در محاسبه تشابه دیداری تصاویر و مجموعه مفاهیم سطح بالای پایگاه‌های تصویری برچسب‌دار، از شبکه عصبی کانولوشنال عمیق (DCNN)[۶] استفاده شده است. DCNN از معروف‌ترین وموفق‌ترین معماری یادگیری عمیق در حوزه آنالیز تصاویر می‌باشد که در بسیاری از حوزه‌های پژوهشی به‌طور وسیع مورد استفاده قرار گرفته است. به‌منظور ایجاد قابلیت مقیاس‌پذیری در فرآیند تکامل تگ‌های تصاویر، کلیه نمونه‌ها با استفاده از الگوریتم خوشه‌بند فازی تک‌گذر C-Means (SPFCM)[۷] معرفی‌شده در [۷]، دسته‌بندی می‌شوند و سپس هر تصویر به کمک نمونه‌های مشابه خود در هر خوشه، مورد پالایش قرار می‌گیرند. در ادامه به معرفی رویکردهای عنوان‌شده خواهیم پرداخت.

ساختار بیان مطالب در این مقاله به این شرح می‌باشد: در بخش دوم به معرفی گزیده‌ای از روش‌های تکامل تگ تصاویر می‌پردازیم. در بخش سوم فرآیند یادگیری عمیق و ساختار شبکه‌های عصبی کانولوشنال عمیق مورد بررسی قرار می‌گیرد. در بخش چهارم خوشه‌بند فازی تک‌گذر SPFCM استفاده‌شده در این پژوهش معرفی می‌شود. در بخش پنجم رویکرد پیشنهادی مقاله معرفی می‌گردد. در بخش ششم مشخصات دادگان مورد استفاده، معرفی معیارهای ارزیابی و نتایج آزمایش‌های صورت‌گرفته ارائه می‌شود. در نهایت در بخش هفتم نتیجه‌گیری کلی از پژوهش و بیان پیشنهاد برای پژوهش‌های آتی صورت می‌گیرد.

۲ -مروری بر روش‌های تکامل برچسب‌های تصاویر

با توجه به تقسیم‌بندی ارائه‌شده در [۸]، روش‌های تکامل و پالایش برچسب‌های تصاویر را می‌توان به دو دسته روش‌های مبتنی‌براستقرا و روش‌های مبتنی‌براستنتاج تقسیم کرد. این دسته‌بندی براساس ایجاد و

یا عدم‌ایجاد تمایز میان دادگان آموزشی و آزمون و نحوه استخراج قوانین بر روی دادگان صورت می‌گیرد.

در روش‌های مبتنی بر استقرا، یادگیرنده سعی می‌کند تا با استفاده از استقرا، تابع تصمیمی را نتیجه بگیرد که دارای نرخ خطای پایینی در تمامی توزیع‌های داده‌های آموزشی و آزمون برای یک یادگیری خاص باشد. از این رو در این روش‌ها بدون توجه به داده‌های آزمون، یکسری قواعد کلی بر اساس یادگیری نمونه‌های آموزشی نتیجه می‌شود و یا به تخمین مدلی عام بر اساس نمونه‌های یادگیری شده می‌پردازد. بر خلاف روش‌های استقرایی، در روش‌های استنتاجی قواعد به دست آمده از داده‌های آموزشی، تنها بر مجموعه‌ای خاص از داده‌های آزمون قابل اعمال می‌باشد. در این روش‌ها می‌توان بدون ایجاد تمایز میان داده‌های آموزشی و آزمون به استنتاج قوانین بر کل مجموعه داده‌های ارائه شده پرداخت.

الگوریتم‌های مبتنی بر استقرا با توجه به اینکه از نوع جداسازهای پارامتری هستند یا از نوع مولدهای غیر پارامتری، خود به دو دسته (۱) مبتنی بر نمونه[۸] و (۲) مبتنی بر مدل[۹] تقسیم می‌شوند. الگوریتم‌های مبتنی بر استنتاج نیز با توجه به مدل یادگیری شامل دو زیر مجموعه (۱) مبتنی بر تجزیه ماتریس و (۲)مبتنی بر گراف می‌باشند. در ادامه هرکدام از این روش‌ها به‌طور خلاصه معرفی می‌شوند.

در الگوریتم‌های مبتنی بر نمونه، هر نمونه آزمون با تمام نمونه‌های آموزشی مقایسه می‌شود. این روش‌ها زیرمجموعه‌ای از الگوریتم‌های غیر پارامتری هستند که در آن‌ها کلیه فرضیه‌ها بر مبنای نمونه‌های آموزشی ساخته می‌شود. از جمله ویژگی‌های این دسته از الگوریتم‌ها این است که با افزایش تعداد نمونه‌های آموزشی، پیچیدگی آن‌ها افزایش می‌یابد. انواع الگوریتم‌های رأی‌دهی همسایه از جمله این الگوریتم‌ها هستند [۹]. یوریچیو[۱۰] و همکارانش در [۱۰] چارچوبی برای پالایش تگ مبتنی بر تکنیک k همسایه نزدیک‌تر ارائه داده‌اند. ایده اصلی این تکنیک انتخاب مجموعه‌ای از تصاویر مشابه دیداری و سپس در نظر گرفتن مجموعه‌ای از تگ‌های مرتبط بر اساس یک روند انتقال تگ می‌باشد. در پژوهش آن‌ها، از معیاری برای ارتباط تگ استفاده می‌شود که میزان توزیع هر تگ در مجموعه تصاویر همسایه تصویر آزمون و در کل مجموعه تصاویر را محاسبه می‌کند.

در [۱۱] چن[۱۱] و همکارانش چارچوبی مشتمل از فاز یادگیری و تخمین ضرایب مرتبط پرس‌وجو ارائه می‌دهند. در پژوهش آن‌ها از روشی تحت عنوان AFSVM[۱۲] به‌منظور تخمین میزان ارتباط تگ‌ها به تصاویر استفاده می‌شود، سپس با بهره‌گیری از فاز پالایشی تحت عنوان LapRLS[۱۳] عملیات پالایش برچسب‌های پیش‌بینی شده برای بهبود بازیابی تصاویر فلیکر[۱۴] مبتنی بر تگ انجام می‌گیرد.

در [۱۲] مسئله پالایش تگ به‌صورت تجزیه ماتریس تگ‌های ایجاد شده توسط کاربر به یک ماتریس تگ تصفیه شده و یک ماتریس پراکندگی خطا، انجام شده‌است. هدف پژوهش آن‌ها بهینه سازی اندازه‌گیری چهار جنبه (۱) همبستگی معنایی میان تگ‌های کم‌رتبه فراهم شده توسط کاربر، (۲) سازگاری محتوای تصویر با تگ‌ها، (۳) همبستگی میان تگ‌ها و (۴) ماتریس پراکندگی خطا می‌باشد.

رویکردهای اخیر پالایش تگ به‌طور عمده تمرکز خود را بر ویژگی‌های دیداری و اطلاعات معنایی استخراج شده از برچسب‌های تصاویر قرار می‌دهند و اطلاعات به‌دست‌آمده از سوی کاربران که تأثیر زیادی در تشخیص برچسب‌های درست و غلط دارند را درنظر نمی‌گیرند. به‌این‌منظور در [۱۳] رویکردی تحت عنوان تکامل تنسور سه خوشه‌ای[۱۵] (TTC) ارائه می‌شود. در این روش روابط درونی میان کاربران، مفاهیم دیداری تصاویر و همبستگی معنایی برچسب‌های تصاویر به‌وسیله یک تنسور مدل می‌شود و روابط بیرونی میان این سه عامل توسط سه تنظیم‌کننده کاوش می‌شود. رویکرد تکامل تگ ارائه شده قادر به تقسیم هر تنسور به تعدادی زیر مجموعه بوده و به‌طور هم‌زمان به خوشه‌بندی کاربران، تصاویر و تگ‌ها می‌پردازد. این عمل موجب بهبود کیفیت تگ‌های تصاویر و کاهش شکاف معنایی میان تصاویر و برچسب‌های آن شده است.

در [۱۴] یک رویکرد نوین تکامل تگ ارائه شده است که در آن به‌جای تکامل تگ‌های ازدست‌رفته هر تصویر با المان‌های آن تصویر، یک بردار رتبه‌دهی برچسب[۱۶] به‌ازای هر تصویر معرفی می‌کند و از آن برای پیش‌بینی برچسب‌های مرتبط با هر تصویر استفاده می‌شود. در پژوهش مذکور به‌منظور ساخت بردار رتبه‌دهی، از یک تابع خطی محلی[۱۷] در همسایگی هر تصویر و بر مبنای بردار ویژگی‌های دیداری تصاویر استفاده شده است.

در [۱۵] روشی تحت عنوان تکامل تگ با استفاده از بازیابی ماتریس نویز[۱۸] (TCMR) ارائه شده است که در آن فرض می‌شود که کلیه تگ‌های متعلق به تصاویر از یک ماتریس تگ ناشناخته نمونه‌برداری شده‌اند. هدف کلی این پژوهش، بازیابی ماتریس تگ بر مبنای برچسب‌های نمونه‌برداری شده می‌باشد. در این مقاله الگوریتم بازیابی ماتریس نویزی، قادر به بازیابی تگ‌های مرتبط از دست‌رفته تصاویر و تشخیص تگ‌های نویزی می‌باشد. روش TCMR به‌منظور بهبود کارایی الگوریتم تکامل تگ پیشنهادی، با استفاده از یک گراف لاپلاسی مبتنی بر جزء[۱۹]، ماتریس تگی را بازیابی می‌کند که به بهره‌برداری مؤثر وابستگی‌های آماری میان ویژگی‌های تصاویر و تگ‌ها می‌پردازد. این روش موجب کاهش تأثیر تگ‌های نویزی می‌شود. به این‌صورت که به برچسب‌های سازگار با ویژگی‌های تصاویر امتیاز بالاتری اختصاص می‌دهد و به تگ‌های غیرمرتبط وزن کمتری می‌دهد. نتایج ارائه شده حاکی از مؤثر بودن روش پیشنهادی در فرآیند تکامل تگ‌های تصاویر می‌باشد.

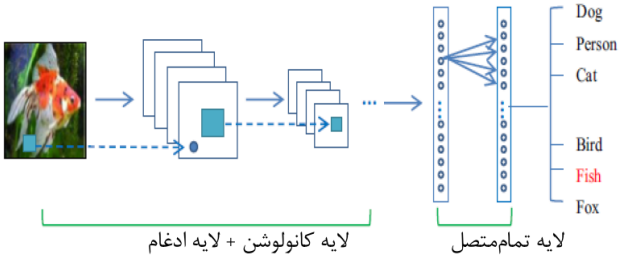
۳ -یادگیری عمیق

طی سال‌های اخیر، یادگیری عمیق به‌صورت گسترده در حوزه بینایی کامپیوتر مورد مطالعه قرار گرفته است. هدف از این پژوهش کاهش شکاف معنایی در بازیابی تصاویر با برقراری ارتباط معنایی مناسب بین

ویژگی‌های سطح پایین تصویر و مفاهیم سطح بالای آن، می‌باشد. در سال‌های اخیر از یادگیری عمیق به‌عنوان رویکردی مؤثر در کاهش این شکاف معنایی استفاده شده است [۱۶]. این رویکرد، تکنیکی مبتنی‌بر شبکه‌های عصبی است و به‌عنوان زیر مجموعه‌ای از روش‌های یادگیری ماشین به‌شمار می‌رود. معماری‌های یادگیری عمیق متنوعی وجود دارد که از جمله آن‌ها می‌توان به شبکه عصبی کانولوشن (CNN)[۲۰] [۱۷]، شبکه باور عمیق[۲۱] [۱۸]و شبکه عصبی بازگشت‌کننده[۲۲] [۱۹] اشاره کرد. مؤثر بودن معماری‌های معرفی‌شده در زمینه‌های متنوعی از قبیل تشخیص اشیاء تصویر [۲۰]، تشخیص خودکار گفتار [۲۱]، شناسایی چهره [۱۷]، پردازش زبان طبیعی [۲۲] به اثبات رسیده است.

به‌طورکلی، روش‌های یادگیری عمیق به چهاردسته (۱) CNN، (۲) ماشین بولتزمن محدود (RBMS)[۲۳]، (۳) رمزنگارخودکار و (۴) کدنویسی پراکنده[۲۴] تقسیم می‌شوند. CNN از مهم‌ترین روش‌های یادگیری عمیق به‌شمار می‌روند که از سه لایه اصلی تشکیل می‌شود. وظیفه اصلی این لایه‌ها استخراج سلسله‌ای از ویژگی‌های غیرخطی از تصاویر می‌باشد. سه‌نوع لایه اصلی مورد استفاده در ساختار CNN عبارت‌اند از: لایه کانولوشن (Conv)[۲۵]، لایه ادغام[۲۶] و لایه کاملاً متصل (FC)[۲۷] با قرار دادن این لایه‌ها به‌صورت متوالی ساختار شبکه کانولوشنی ایجاد می‌شود. در انتهای شبکه لایه‌ای تحت عنوان Softmax قرار می‌گیرد که تصاویر را دسته‌بندی کرده و برچسب تصویر را مشخص می‌کند. شکل ۱ نمایی از معماری CNN را نشان می‌دهد.

وظیفه اصلی این لایه‌ها استخراج سلسله‌ای از ویژگی‌های غیرخطی از تصاویر می‌باشد. توده خروجی شبکه‌های عمیق را می‌توان به‌صورت توده‌ای سه‌بعدی از نورون‌ها تفسیر کرد. این درحالی است که در شبکه‌های عصبی معمولی هر لایه به‌صورت لیستی یک‌بعدی از نورون‌ها می‌باشد. لایه کانولوشن، هسته اصلی تشکیل‌دهنده CNN می‌باشد. پارامترهای لایه کانولوشن شامل مجموعه‌ای از فیلترهای قابل یادگیری هستند. هر فیلتر ازلحاظ مکانی کوچک بوده اما در امتداد عمق توده تصویر ورودی ادامه پیدا می‌کند.



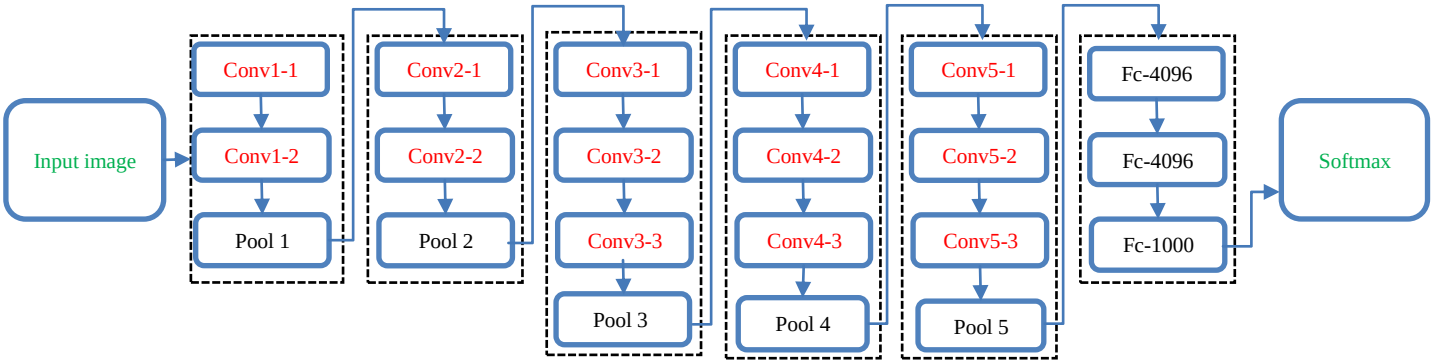
شکل ۱. طرح کلی از معماری شبکه عصبی کانولوشن [۳۳].

به‌منظور استخراج نگاشت ویژگی[۲۸] متفاوت از تصاویر در هر لایه کانولوشن، این فیلترها در امتداد پهنا و ارتفاع بر سطح تصویر غلتانده می‌شود. با این‌عمل نگاشت فعال‌ساز[۲۹] دوبعدی برای آن فیلتر ایجاد می‌شود.

ویژگی‌های استخراج شده از لایه‌های کانوولوشن خاصیت سلسله‌مراتبی دارند، به این‌معنی که به‌طور مثال در لایه‌های ابتدایی گوشه‌ها و خط‌ها و لبه‌ها یادگرفته می‌شوند و در لایه‌های عمیق‌تر به‌ترتیب ویژگی‌هایی با سطوح بالاتر، همانند اشیای درون تصویر، استخراج می‌شوند. با این‌روش شبکه‌های عصبی کانولوشنی تصویر اصلی را طی لایه‌های مختلف از مقادیر اصلیِ پیکسل‌ها به‌احتمال تعلق در هر کلاس تبدیل می‌کنند. کارکرد لایه ادغام کاهش اندازه طول و عرض تصویر ورودی به‌جهت کاهش تعداد پارامترها و محاسبات در داخل شبکه و بنابراین کنترل بیش‌برازش[۳۰] می‌باشد. در لایه FC، نورون‌هایی که در یک لایه قرار دارند، دقیقاً همانند شبکه‌های عصبی معمولی، با تمام نورون‌های موجود در لایه قبلی ارتباط دارند. تنها تفاوت بین لایه تمام‌متصل و لایه کانولوشنی این است که نورون‌ها در هر لایه کانولوشن تنها به ناحیه‌ای محلی از نورون‌های لایه قبل متصل هستند و به اشتراک پارامترها با یکدیگر می‌پردازند. آخرین لایه شبکه لایه Softmax است که دسته‌بندی تصاویر در ۱۰۰۰ طبقه مشخص، که هر طبقه نماینده‌ای از یک یا چند برچسب مشابه می‌باشد را نشان می‌دهد.

در هر شبکه عصبی کانولوشن دو مرحله، (۱) انتشاربه‌جلو[۳۱] و (۲) پس‌انتشار[۳۲] برای آموزش وجود دارد. در مرحله اول تصویر ورودی به شبکه تغذیه می‌شود و این عمل چیزی جز ضرب نقطه‌ای بین ورودی و پارامترهای هر نورون ودرنهایت اعمال عملیات کانولوشن در هر لایه نیست که بعد از این عمل خروجی شبکه محاسبه می‌شود. در این مرحله به‌منظور تنظیم پارامترهای شبکه و یا به‌عبارت‌دیگر همان آموزش شبکه، از نتیجه خروجی به‌منظور محاسبه میزان خطای شبکه استفاده می‌شود. برای این کار خروجی شبکه را با استفاده از یک تابع خطا[۳۳] با پاسخ صحیح مقایسه کرده و به این‌صورت میزان خطا محاسبه می‌شود. در مرحله بعد، بر اساس میزان خطای محاسبه‌شده، مرحله پس‌انتشار آغاز می‌شود. در این مرحله گرادیان هر پارامتر با توجه به قاعده‌زنجیره[۳۴] محاسبه می‌شود و تمامی پارامترها با توجه به تأثیری که بر خطای ایجادشده در شبکه دارند تغییر پیدا می‌کنند. بعد از به‌روزرسانی پارامترها، مرحله انتشاربه‌جلو شروع می‌شود. بعد از تکرار تعداد مناسبی از این مراحل، آموزش شبکه پایان می‌یابد. توده خروجی شبکه‌های عمیق را می‌توان به‌صورت توده‌ای سه‌بعدی از نورون‌ها تفسیر کرد. این در حالی است که در شبکه‌های عصبی معمولی هر لایه به‌صورت لیستی یک‌بعدی از نورون‌ها می‌باشد.

یکی از رویکردهایی که تابه‌حال در حل چالش دقت محاسبات تکامل تگ‌های تصاویر مبتنی بر ویژگی‌های دیداری و معنایی به آن پرداخته نشده است، استفاده از CNN از روش‌های یادگیری عمیق می‌باشد. معماری‌های متفاوتی برای CNN وجود دارد که در تعداد و ساختار لایه‌های میانی متفاوت می‌باشند. از جمله معماری‌های مفید در



شکل ۲. نمایی از معماری شبکه VGGNet

حوزه آنالیز تصاویر، معماری VGGNet[۳۵] است. VGGNet شبکه عصبی کانولوشن عمیقی است که توسط کارن سیمونیان[۳۶] و اندرو زیسرمن[۳۷] توسعه داده شده است. این معماری به‌عنوان دومین شبکه پیشنهادی برنده مسابقات ILSVRC2014[۳۸] در عملیات طبقه‌بندی تصاویر[۳۹] اما با بهترین عملکرد در عملیات تشخیص اشیاء[۴۰] نسبت به دیگر معماری‌های ارائه‌شده در آن رقابت، انتخاب شده است [۲۳]. نسخه از پیش‌آموزش دیده‌شده از این شبکه، توسط ۲/۱ میلیون تصویر موجود در دادگان ImageNet [۲۴] با رزولوشن بالا و سایز ۲۲۷×۲۲۷ در ۱۰۰۰ طبقه یادگیری شده است. نسخه نهایی بهترین شبکه از این معماری شامل شانزده لایه Conv/FC بوده و یک معماری به‌شدت همگن می‌باشد که در این مقاله مورد استفاده قرار می‌گیرد. شکل ۲ نمایی از معماری ارائه‌شده برای نسخه شانزده لایه‌ای از این شبکه، که دارای سیزده لایه کانولوشن و سه لایه تمام‌متصل می‌باشد، را نمایش می‌دهد.

با توجه به استفاده از خوشه‌بند SPFCM به‌منظور دسته‌بندی تصاویر مشابه دیداری، در بخش بعد به معرفی کامل الگوریتم FCM و نسخه SPFCM ارائه‌شده از آن، می‌پردازیم.

۴- الگوریتم FCM

این الگوریتم نخستین بار در سال ۱۹۷۳ توسط دان[۴۱] [۲۵] ارائه شد و ازجمله الگوریتم‌های خوشه‌بندی فازی به‌شمار می‌آید. در این الگوریتم درجه عضویت n نمونه ورودی به c خوشه مشخص، محاسبه می‌شود. تعداد خوشه‌ها توسط کاربر مشخص می‌شود. تابع هدف محاسبه‌شده توسط این الگوریتم به‌صورت رابطه (۱) می‌باشد.

(۱)

در این رابطه X بیانگر ویژگی‌های متعلق به نمونه‌های ورودی، U نشان‌دهنده ماتریس بخش‌بندی با ابعاد c×n و V نشان‌دهنده مراکز خوشه و حاوی c مرکز خوشه می‌باشد. m>1 بیانگر ثابت فازی‌سازی می‌باشد که روش‌های مختلفی برای محاسبه آن ارائه شده است [۱۸،۱۷]. در این پژوهش به دلیل عملکرد مناسب برروی دادگان انتخابی، مشابه [۷]، m=۷/۱ درنظر گرفته شده است. $\|.\|_A$ نرم ضرب داخلی می‌باشد که در این مقاله از نرم اقلیدسی استفاده شده است. الگوریتم FCM به‌صورت الگوریتم ۱ می‌باشد.

الگوریتم ۱: FCM

ورودی : $\mathbf{X}, c, m, n_s$

خروجی : $\mathbf{V}, \mathbf{U}$

While max $1\le k\le c$ $\{||V_{k,new} - V_{k,old}||^2\} > \varepsilon$ do

$$u_{ij} = \left[\sum_{k=1}^{c}\left(\frac{||X_j - V_i||}{||X_j - V_k||}\right)^{\frac{2}{m-1}}\right]^{-1}, \forall\, i,j$$

$$V_i = \frac{\sum_{j=1}^{n}(u_{ij})^m X_j}{\sum_{j=1}^{n}(u_{ij})^m}, \forall\, i$$

پایان

ε ضریب ثابتی است که در کلیه آزمایش‌ها مقدار آن $\varepsilon=10^{-3}$ درنظر گرفته شده است. در صورت درنظر گرفتن ضریب وزن W برای مشخص کردن درجه اهمیت نمونه‌ها، الگوریتم WFCM حاصل می‌شود. در الگوریتم WFCM ماتریس بخش‌بندی مشابه با FCM محاسبه می‌شود، اما مراکز خوشه V با استفاده از رابطه ۲ به‌دست می‌آید:

$$\boldsymbol{Jm}(\mathbf{U},\mathbf{V}) = \sum_{i=1}^{c}\sum_{j=1}^{n}\mathbf{W}_j\boldsymbol{u}_{ij}\left\|\boldsymbol{X}_j - \boldsymbol{V}_i\right\|_{\boldsymbol{A}}^{2} \qquad (۲)$$

از الگوریتم WFCM، به‌منظور بهره‌گیری از آن در الگوریتم خوشه‌بند SPFCM [۷]، استفاده می‌شود که در بخش بعد ارائه خواهد شد .

الگوریتم ۲ : SPFCM

ورودی : $\mathbf{X}, c, m, n_s$

خروجی : $\mathbf{V}$

n_s انتخاب زیرمجموعه‌ای از $\mathbf{X}$ با طول

$\mathbf{X}=\{X_1, X_2, \ldots, X_s\}$

1. $\mathbf{W}=1n_s$
2. $\mathbf{U}$, $\mathbf{V}$= WFCM $(X_1, c, m, \mathbf{W})$
3. For L=2 to s do

$$w_i' = \sum_{j=1}^{ns}(u_{ij})w_j, i = 1, \ldots, c$$

4. $\mathbf{W}=\{W' \cup 1ns\}$
5. $\mathbf{U},\mathbf{V}$ = WFCM($\{\mathbf{V} \cup \mathbf{X}_L\}, c, m, \mathbf{W}, \mathbf{V}$)

پایان

۴ ۱ -الگوریتم SPFCM

الگوریتم SPFCM نسخه‌ای توسعه‌یافته از الگوریتم FCM می‌باشد. این الگوریتم در خوشه‌بندی فازی مجموعه داده‌های با مقیاس بزرگ که مشکل بارگذاری در سیستم دارند نقش مؤثری ایفا می‌کند [۷]. در ادامه، الگوریتم SPFCM ارائه شده است.

در این الگوریتم ابتدا کل مجموعه به s قسمت به طول ns تقسیم می‌شود. در خط اول الگوریتم بردار وزن ورودی WFCM با مقادیر اولیه یک تشکیل می‌شود. در خط دوم با استفاده از الگوریتم WFCM اولین زیرمجموعه از داده، پارتیشن‌بندی می‌شود و c مرکز خوشه از الگوریتم حاصل می‌شود. در هر تکرار، از الگوریتم WFCM برای خوشه‌بندی مجموعه داده $\{V \cup X_L\}$ استفاده می‌شود. این مجموعه، حاوی داده‌های زیرمجموعه مورد بررسی (X_L) و مراکز خوشه محاسبه‌شده از مرحله قبل (V) می‌باشد. بنابراین در هر تکرار این الگوریتم (c+ns) نمونه خوشه‌بندی‌شده و ماتریس بخش‌بندی U که حاوی درجه عضویت نمونه‌ها به هر خوشه و ماتریس V که حاوی مراکز خوشه‌ها می‌باشد، محاسبه می‌شود. در خط سوم و چهارم بردار وزن استفاده شده در الگوریتم WFCM به‌منظور خوشه‌بندی مجموعه $\{V \cup X_L\}$ اجرا می‌شود. در خط سوم وزن‌های متعلق به مرکز خوشه c با استفاده از ماتریس درجه عضویت‌های متعلق به هر نمونه محاسبه می‌شود.

در خط چهارم بردار وزن به طول c+ns ایجاد می‌شود. این بردار حاوی وزن‌های محاسبه شده برای مراکز خوشه که از مرحله قبل به دست آمده است و تعداد ns وزن با مقدار یک برای زیر مجموعه X_L مورد بررسی، می‌باشد.

در خط پنجم، الگوریتم WFCM با پارامترهای ورودی $\{V \cup XL\}$ داده، c خوشه، بردار وزن W و مجموعه V حاوی مراکز خوشه محاسبه شده از مرحله قبل، عملیات خوشه‌بندی را انجام می‌دهد. این روند تکرار می‌شود تا کلیه زیرمجموعه‌های داده خوشه‌بندی شوند. خروجی نهایی الگوریتم ماتریس حاوی مراکز خوشه می‌باشد.

در این پژوهش از الگوریتم SPFCM به‌منظور خوشه‌بندی ماتریس ویژگی‌های تصاویر و محاسبه نمونه‌های مشابه استفاده می‌شود. در هر گام، با استفاده از ماتریس بخش‌بندی حاوی درجه عضویت نمونه‌ها در هر خوشه، بیشترین درجه عضویت هر نمونه به یک خوشه انتخاب می‌شود. در نهایت تصاویر مشابه، دسته‌بندی می‌شوند. تصاویر مشابه دیداری متعلق به هر تصویر، به‌منظور انجام عملیات رفع نویز از برچسب‌های تصویر به فاز بعد انتقال داده می‌شود. در بخش بعد رویکرد پیشنهادی این مقاله مورد بررسی قرار خواهد گرفت.

۵ -روش پیشنهادی

در این بخش رویکرد ارائه‌شده در این مقاله به‌منظور تکامل تگ‌های تصاویر و رفع نویز از آن‌ها معرفی می‌شود. در ادامه از این رویکرد با عنوان CNN-TC[۴۲] یاد می‌شود. در این بخش به‌معرفی مدل پیشنهادی می‌پردازیم. ابتدا به‌منظور ایجاد دیدی از مدل پیشنهادی خلاصه‌ای از آن با استفاده از گام‌های مطرح شده در شکل ۳ ارائه می‌گردد.

شکل ۳. چارچوب رویکرد CNN-TC ارائه‌شده در این پژوهش

در این رویکرد نخست کلیه تصاویر مورد پیش‌پردازش قرار می‌گیرند. در این راستا ابتدا تصاویری که در وب‌سایت فلیکر[۴۳] با لینک‌های معتبر قابل‌دسترس هستند دریافت می‌شوند. عملیات انجام‌شده پیش از فرآیند تکامل تگ‌های تصاویر عبارت‌اند از: ۱- تنک کردن تگ‌ها، ۲- استخراج ویژگی‌های تصاویر با به‌کارگیری CNN، ۳-

خوشه‌بندی تصاویر مشابه با استفاده از خوشه‌بند SPFCM و ۴- محاسبه نزدیک‌ترین تصاویر همسایه با ارزیابی تشابه دیداری و معنایی میان نمونه‌ها. در ادامه نحوه انجام این عملیات ارائه خواهد شد.

با هدف محاسبه ویژگی‌ها و درجه‌احتمال تعلق هر نمونه به هر کلاس، کلیه تصاویر به‌عنوان ورودی به DCNN داده می‌شود. در این راستا ابتدا هرکدام از نمونه‌های xi را به‌عنوان ورودی به شبکه کانولوشن VGGNet داده می‌شود و به‌ازای هر نمونه، بردار yi با M=۱۰۰۰ ضریب برچسب از آخرین لایه FC و بردار vi با Z=۴۰۹۶ ویژگی از دومین لایه FC شبکه استخراج می‌شود.

در فاز سوم، یک بردار ویژگی به ابعاد ۴۰۹۶×۱ و یک بردار برچسب ۱۰۰۰ تایی حاوی ضرایب احتمالی عضویت در هر کلاس، به‌ازای هر تصویر از DCNN استخراج می‌شود. حجم بالای مجموعه تصاویر و مشکلات ناشی از بارگذاری کل مجموعه داده‌های تصاویر در سیستم، پیچیدگی محاسباتی زیاد و پیچیدگی زمانی بالا ازجمله چالش‌های مطرح در این حوزه می‌باشد. لذا در مرحله چهارم به‌منظور مقیاس‌پذیری فرآیند تکامل برچسب‌های مجموعه دادگان با مقیاس بزرگ و مقابله با چالش‌های فوق، با استفاده از بردار ویژگی‌های پایین‌رتبه، تصاویر خوشه‌بندی شده تا تصاویر مشابه دیداری دسته‌بندی شوند. در این مرحله به‌منظور دسته‌بندی تصاویر مشابه از الگوریتم خوشه‌بندی فازی SPFCM استفاده شده است.

در مرحله پنجم در هر خوشه با بررسی همبستگی میان تگ‌ها و ضرایب متعلق به برچسب‌های هر تصویر، تشابه معنایی میان تصاویر ارزیابی شده و تصاویر همسایه متعلق به هر نمونه محاسبه می‌شود. درنهایت در گام‌های بعد، با به‌کارگیری تکنیک رأی‌گیری نزدیک‌ترین همسایه[۴۴] عملیات انتقال تگ‌های مرتبط و با فرکانس بالا از تصاویر همسایه به تصویر آزمون و حذف تگ‌های غیرمرتبط از آن، عملیات تکامل تگ هر نمونه صورت می‌گیرد. بهره‌گیری از فرآیند خوشه‌بندی فازی به‌منظور دسته‌بندی تصاویر مشابه دیداری و استفاده از شبکه‌های عصبی کانولوشنال عمیق برای استخراج ویژگی‌های مؤثر از تصاویر، مؤثر بودن این رویکرد را در فرآیند تکامل تگ‌های تصاویر نشان می‌دهد.

رویکرد ارائه‌شده علاوه‌بر بهره‌گیری از مزایای فرآیند یادگیری انتقالی از جمله کاهش زمان و حجم عملیات مورد نیاز، در آنالیز مجموعه تصاویر با مقیاس بزرگ نیز مفید خواهد بود. لذا در این پژوهش با بهره‌گیری از فرآیند یادگیری انتقالی از شبکه از پیش یادگیری‌شده با تصاویر آموزشی ImageNet استفاده می‌شود. بدیهی است که با توجه به تنوع و همه‌منظوره[۴۵] بودن دادگان آموزشی ImageNet [۲۴]، بارگذاری وزن‌های یادگیری شده برای اهداف مختلف در دادگان دیگر امکان‌پذیر بوده و موجب صرفه‌جویی در زمان یادگیری شبکه و انجام محاسبات خواهد شد. همچنین با توجه به حجم بالای تصاویر و عدم امکان محاسبه تشابه میان کلیه تصاویر به‌صورت یک‌جا و دشواری بارگذاری ماتریس ویژگی محاسبه‌شده برای هر تصویر در حافظه، از الگوریتم خوشه‌بندی فازی SPFCM استفاده شده است. بهره‌مندی از این خوشه‌بند به مقابله با چالش‌های مطرح‌شده می‌پردازد.

پیش از ارائه روابط استفاده‌شده در این مقاله ابتدا به معرفی متغیرهای موجود می‌پردازیم. مجموعه $X=\{x_1, x_2,\ldots,x_N\}$ بیانگر مجموعه تصاویر اولیه می‌باشد که تگ‌های تخصیص‌داده‌شده به هر تصویر با $T=\{w_1,w_2,\ldots,w_M\}$ نشان داده می‌شود. در این مجموعه‌ها N بیانگر تعداد تصاویر و M بیانگر تعداد تگ‌های تصاویر می‌باشد. $y_{N,M}$ یک متغیر باینری است که به‌منظور بیان تعلق و عدم‌تعلق تگ w_M به تصویر x_N مورد استفاده قرار می‌گیرد. نتیجه فرآیند تکامل تگ‌های تصاویر، توسط ماتریس Y معرفی می‌شود که هر عضو آن دارای مقادیر مثبت بوده و نشان‌دهنده درجه تعلق هر تگ w_i به x_j می‌باشد. به‌ازای هر تصویر دو بردار معرفی می‌شود. $y_i=(y_{i1},y_{i2},\ldots,y_{im})$ بیانگر بردار درجه اطمینان تخصیص تگ‌ها به iمین تصویر و بردار $v_i=(v_{i1},v_{i2},\ldots,v_{iz})$ نشان‌دهنده بردار ویژگی‌های دیداری تصاویر می‌باشد.

همان‌طورکه در شکل ۳ نشان داده شده است طی فرآیند تکامل تگ‌های تصاویر، ابتدا با استفاده از شبکه کانولوشنی VGGNet ویژگی‌های دیداری و معنایی از تصاویر، استخراج می‌شوند. در این راستا ابتدا هرکدام از نمونه‌های x_i به‌عنوان ورودی به شبکه کانولوشن VGGNet داده می‌شود. سپس از آخرین لایه شبکه (لایه Softmax)، بردار y_i حاوی ۱۰۰۰ ضریب احتمالاتی به‌عنوان برچسب، و از دومین لایه FC شبکه، بردار v_i حاوی۴۰۹۶ ویژگی دیداری استخراج می‌شود. از ضریب‌های احتمالاتی برای انتخاب تصاویر مشابه معنایی و از ویژگی‌های دیداری، برای انتخاب تصاویر مشابه دیداری استفاده می‌شود. این ویژگی‌ها به‌منظور اعمال عملیات خوشه‌بندی به الگوریتم SPFCM ارسال می‌شوند.

همان‌طورکه عنوان شد، از جمله چالش‌های پیش‌رو در فرآیند تکامل تگ‌های تصاویر، پیچیدگی محاسباتی بالا و عدم مقیاس‌پذیری سیستم می‌باشد. از این‌رو به‌منظور ساده‌سازی محاسبات و مقیاس‌پذیری سیستم، تصاویر خوشه‌بندی می‌شوند. به این صورت که بردارهای ویژگی v_i مربوط به تصاویر، به خوشه‌بند SPFCM وارد می‌شوند و ماتریس بخش‌بندی ایجاد شده و با استفاده از آن تصاویر خوشه‌بندی می‌شوند. این ماتریس حاوی ضرایب تعلق هر تصویر به c خوشه درنظر گرفته شده توسط کاربر می‌باشد. بعد از خوشه‌بندی نمونه‌ها عملیات پالایش در هر خوشه به‌طور مجزا صورت می‌گیرد. از آنجا که ویژگی‌های دیداری تصویر به‌تنهایی قادر به شناسایی دقیق محتوای تصویر نیستند، ازاین‌رو با استفاده از بردار y_i متعلق به هر تصویر، محاسبه تشابه معنایی میان تصاویر مشابه در هر خوشه انجام می‌گیرد. به‌منظور جلوگیری از وقوع پدیده پالایش تگ مشابه برای تصاویر متفاوت، به استفاده ترکیبی از تشابه دیداری و معنایی تصاویر می‌پردازیم. بنابراین ابتدا با درنظرگرفتن یک حد آستانه، بردار y_i که

حاوی ضرایب تعلق تصاویر به ۱۰۰۰ طبقه از تصاویر ImageNet می‌باشد، به یک بردار باینری تبدیل می‌شود. به‌این‌صورت که ضرایب بیشتر از یک حد آستانه فرضی همچون β را به ۱ و سایر ضرایب به صفر که نشان از عدم‌تعلق تگ W_m به تصویر x_N می‌باشد، تبدیل می‌شود. در مرحله بعد با محاسبه تشابه میان تگ‌ها و هم‌پوشانی میان آن‌ها تشابه معنایی میان تصاویر ارزیابی شده است. برای محاسبه تشابه معنایی میان تگ‌های W_i و W_j مشابه [۱۲] از رابطه زیر استفاده می‌شود:

(۳)

در این رابطه $d(Wi, Wj)$ فاصله میان تگ‌های Wi و Wj می‌باشد. برای محاسبه این فاصله مشابه [۱۲] از روش فاصله‌سنجی گوگل[۴۶] استفاده می‌شود که با استفاده از رابطه ۴ به محاسبه هم‌پوشانی میان تگ‌ها می‌پردازد. در این رابطه q(Wi) و $q(W_j)$ تعداد تصاویر حاوی تگ Wi و Wj می‌باشد. q(Wi,Wj) تعداد تصاویری می‌باشد که حاوی هر دو تگ Wi و Wj است. درنهایت با استفاده از تشابه دیداری و معنایی به‌کارگرفته‌شده تصاویر همسایه متعلق به هرتصویر محاسبه می‌شود.

(۴)

N تعداد کل تصاویر موجود در مجموعه دادگان می‌باشد. به‌منظور محاسبه تشابه معنایی میان تصاویر مشابه، از روش ارائه‌شده در [۲۸] استفاده می‌کنیم. به‌این‌صورت که از ضرب نقطه‌ای ماتریس تشابه میان تگ‌ها (S) و بردار درجه اطمینان تعلق هر تگ به تصویر (Y) استفاده می‌کنیم. رابطه ارائه‌شده به‌صورت زیر می‌باشد:

(۵)

در راستای تکامل تگ‌های تصاویر انتخابی، پیش‌بینی لیستی از تگ‌ها بر مبنای تگ‌های تصاویر همسایه مشابه، می‌تواند مؤثر باشد. از این‌رو در چارچوب پالایش تگ ارائه‌شده مشابه [۲۳-۲۱]، از روش رأی‌گیری همسایه[۴۷] استفاده شده است. در طول انجام فرآیند تکامل تگ تصاویر، تگ‌های هر تصویر با تگ‌های تصاویر همسایه مورد مقایسه قرار می‌گیرد. کلیه عملیات مقایسه تصاویر و تکامل برچسب‌های تصاویر، با توجه به برچسب‌های متعلق به دادگان تست صورت می‌گیرد. بدیهی است که با توجه به بهره‌مندی از فرآیند یادگیری انتقالی، دامنه برچسب‌های دادگان تست با دامنه برچسب‌های به‌کارگرفته شده در شبکه کانولوشن VGGNet متفاوت می‌باشد و فرض بر این است که هیچ‌گونه ارتباطی میان آن‌ها وجود ندارد.

به‌منظور انجام فرآیند تکامل تگ تصاویر، ابتدا با در نظر گرفتن حد آستانه α، تمامی تگ‌های با فرکانس وقوع بالاتر از α در تصاویر همسایه، به تصویر آزمون منتقل می‌شوند. سپس تگ‌هایی از تصویر آزمون که در تصاویر همسایه آن، رخ نداده‌اند از آن حذف می‌شود و در صورت وقوع برچسبی در تصاویر همسایه با فرض محتمل شدن صحت آن، حفظ می‌شود. ماتریس برچسب حاصل از فرآیند تکامل تگ، به‌منظور محاسبه میزان کارایی رویکرد پیشنهادی با استفاده از معیارهای ارزیابی مورد بررسی قرار می‌گیرد. نتایج حاصل از پیاده‌سازی رویکرد پیشنهادی مقاله در بخش بعد ارائه می‌گردد.

۶ -نتایج تجربی

در این بخش ضمن معرفی مجموعه دادگان مورداستفاده در ارزیابی رویکرد CNN-TC و معیارهای ارزیابی محاسبه‌شده، نتایج ارزیابی روش پیشنهادی بر مجموعه دادگان معرفی‌شده ارائه می‌گردد. به‌منظور ارزیابی کیفیت رویکرد CNN-TC از مجموعه تصاویر با مقیاس بزرگ NUS-WIDE-270K استفاده شده است [۳۰]. این مجموعه شامل ۲۶۹۶۴۸ تصویر جمع‌آوری‌شده از سایت فلیکر به همراه ۵۰۱۸ تگ ایجادشده توسط کاربران می‌باشد که با حذف آدرس‌های نامعتبر و تصاویر غیرقابل‌دسترس از کل مجموعه تصاویر NUS-WIDE، زیرمجموعه‌ای با تعداد ۲۲۱۸۱۷ تصویر حاصل شد که کلیه آزمایش‌ها بر این زیرمجموعه از تصاویر انجام می‌گیرد. در ادامه از این مجموعه به‌عنوان NUS-WIDE-220K یاد می‌شود.

با توجه به آنکه تگ‌های ابتدایی ایجادشده توسط کاربران به‌شدت ذهنی و یا غیرمرتبط با محتوای تصویر هستند، بنابراین با هدف بهبود کیفیت فرآیند تکامل برچسب‌های تصویر به حذف آن‌ها می‌پردازیم. ازاین‌رو با فرض مؤثرتر بودن برچسب‌های با فرکانس بالا مشابه فرآیند پالایش شرح‌گذاری در [۲۳، ۲۴]، کلیه آزمایش‌ها بر مجموعه تصاویر حاوی ۱۰۰۰ تگ با بیشترین فرکانس انجام گرفته است. همچنین برای این مجموعه از دادگان پنج‌دسته ویژگی سراسری محاسبه شده است که در آزمایش‌ها مورد استفاده قرار می‌گیرد. این ویژگی‌ها عبارت‌اند از: ۱- هیستوگرام رنگ[۴۸] LAB با ۶۴ بعد، ۲- ممان رنگ[۴۹] مبتنی بر بلاک با ۲۲۵ بعد، ۳- کورلوگرام رنگ[۵۰] HSV با ۱۴۴ بعد، ۴- هیستوگرام جهت لبه[۵۱] با ۷۳ بعد و ۵- موجک بافت[۵۲] با ۱۲۸ بعد. اطلاعات بیشتر در رابطه با ویژگی‌های فوق در [۳۴] ارائه شده است. به‌منظور بررسی میزان کارایی رویکرد CNN-TC پیشنهادشده بر فرآیند تکامل شرح‌گذاری تصاویر، نتایج حاصل از آزمایش‌های انجام شده با به‌کارگیری شبکه کانوولوشن VGGNet و بدون استفاده از آن مورد مقایسه قرار می‌گیرد. در ارزیابی نتایج حاصل از رویکرد CNN-TC، مشابه روش ارائه‌شده در [۲۱-۱۹] نسخه ۸۱ تگی از مجموعه NUS-WIDE-220k که توسط افراد خبره شرح‌گذاری شده است، به‌عنوان ماتریس مبنا درنظر گرفته می‌شود. برای ارزیابی نتایج تکامل تگ‌های تصاویر سه معیار ارزیابی شناخته‌شده میانگین دقت[۵۳]، میانگین فراخوان[۵۴] و F1 با دو استراتژی میکرو و ماکرو محاسبه شده است.

با توجه به حجم بالای دادگان مورد استفاده، به‌منظور تنظیم پارامترهای مورد نیاز در رویکرد CNN-TC، نسخه کاهش یافته از دادگان اصلی با ۲۵۰۰۰ تصویر مورد استفاده قرار گرفته است. کلیه پارامترها با اعمال روش اعتبارسنجی متقابل[۵۵] با 2-fold و با در نظر گرفتن ۵۰۰۰ داده تست ۱۵۰۰۰ داده آموزشی و ۵۰۰۰ داده ولیدیشن

محاسبه شده است. ازاین‌رو در این پژوهش مقادیر بهینه محاسبه‌شده برای پارامتر β به‌عنوان حد آستانه انتخابی برای تبدیل احتمال عضویت برچسب‌های استخراج‌شده به ۱۰۰۰ کلمه کلیدی، از شبکه کانوولوشن VGGNet به مقادیر ۰، ۱ و ۵ در نظر گرفته شده است. مقدار پارامتر c به‌عنوان تعداد خوشه‌ها، α به‌عنوان حد آستانه فرضی برای انتخاب تگ‌های تصاویر با فرکانس بالا و s به‌عنوان تعداد زیرمجموعه داده‌ها در الگوریتم خوشه‌بند فازی SPFCM به ترتیب ۱۰۰ و ۱۰ و ۲۰ در نظر گرفته شده است. همچنین حداکثر تعداد همسایگی‌ها به‌ازای هر نمونه آزمون، N=۶۰ فرض شده است و به‌ازای تعداد همسایگان ۴۰،۵۰،۶۰، ۳۰، ۲۰، ۱۰=n نیز نتایج آزمایش‌ها ارائه می‌گردد.

با توجه به اینکه در اکثریت پژوهش‌ها معیار ماکرو F1 مورد محاسبه قرار می‌گیرد، بنابراین تمرکز این پژوهش بر معیار ماکرو F1 می‌باشد. نتایج حاصل از آزمایش‌های صورت‌گرفته با استفاده از رویکرد CNN-TC بر دادگان NUS-WIDE-220k با ۱۰۰۰ تگ با بیشترین فرکانس در جدول ۱ ارائه شده است.

نتایج حاصل از پیاده‌سازی رویکرد CNN-TC با بهره‌مندی از شبکه VGGNet به‌منظور استخراج ویژگی‌های مؤثر از تصاویر و در صورت استفاده از خوشه‌بند فازی SPFCM به‌منظور محاسبه شباهت دیداری میان تصاویر نشان می‌دهد که مقدار پارامتر ماکرو F1 در بهترین حالت برابر با ۵۸/۶۷٪ و بهترین مقدار پارامتر میکرو F1، ۹۵/۷۹٪ به‌دست آمده است.

جدول ۱. نتایج حاصل از پیاده‌سازی رویکرد CNN-TC@n بر دادگان NUS-WIDE-220k به‌ازای n همسایه مختلف و به‌کارگیری خوشه‌بند SPFCM در محاسبه تشابه دیداری میان تصاویر

روش	دقت میکرو (%)	فراخوان‌میکرو (%)	F1 میکرو (%)	دقت ماکرو (%)	فراخوان ماکرو (%)	F1 ماکرو (%)
CNN-TC@10	۰۸/۷۳	۹۲/۵۸	۲۴/۶۵	۹۲/۵۶	۴۳/۶۲	۵۴/۵۹
CNN-TC@20	۲۱/۷۹	۵۳/۶۸	۴۸/۷۳	۹۲/۵۸	۱۳/۶۵	۸۶/۶۱
CNN-TC@30	۲۱/۸۱	۰۹/۷۴	۳۷/۷۶	۶۸/۶۱	۳۴/۷۰	۷۲/۶۵
CNN-TC@40	۵۲/۸۰	۵۶/۷۷	۰۱/۷۹	۱۶/۶۲	۱۲/۷۲	۷۷/۶۶
CNN-TC@50	۷۱/۷۸	۲۳/۸۱	۹۵/۷۹	۹۱/۶۲	۰۲/۷۳	۵۸/۶۷
CNN-TC@60	۷۶/۷۴	۶۵/۸۲	۵۰/۷۸	۹۸/۵۹	۰۲/۷۱	۰۲/۶۵

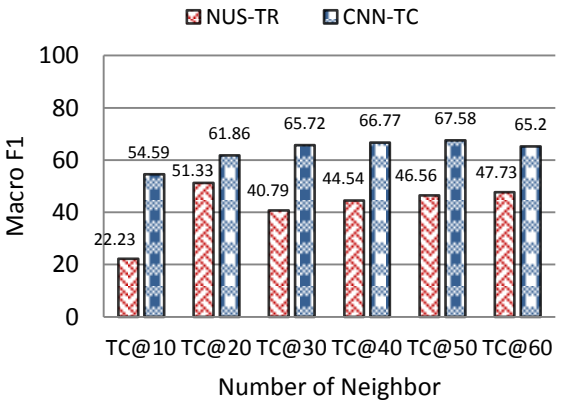
به‌منظور نمایش صلاحیت رویکرد پیشنهادی کلیه آزمایش‌ها به‌صورت مجدد، بدون استفاده از شبکه VGGNet در فرآیند استخراج ویژگی و در صورت عدم‌بهره‌مندی از خوشه‌بند فازی در دسته‌بندی تصاویر مشابه، انجام شده است. نتایج حاصل از فرآیند تکامل تگ با اعمال بر پنج دسته ویژگی معرفی‌شده در بخش قبل در جدول ۲، ارائه شده است.

جدول ۲. نتایج حاصل از پیاده‌سازی رویکرد NUS-TR@n بر دادگان NUS-WIDE-220k به ازای n همسایه مختلف

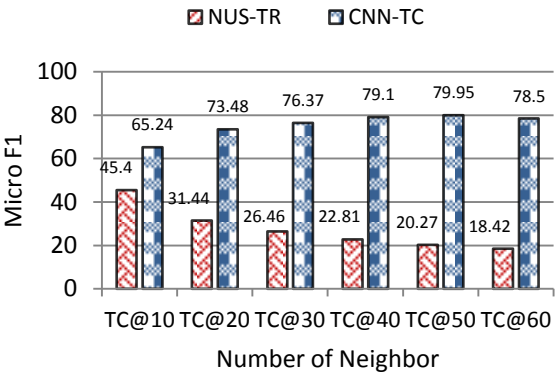
روش	دقت میکرو (%)	فراخوان-میکرو (%)	F1 میکرو (%)	دقت ماکرو (%)	فراخوان ماکرو (%)	F1 ماکرو (%)
NUS-TR@10	۷۹/۱۷	۳۱/۸۳	۴۵/۴۰	۱۲/۶۲	۹۳/۵۹	۲۲/۲۳
NUS-TR@20	۲۳/۷۲	۴۶/۶۰	۳۱/۴۴	۲۱/۳۸	۷۷/۳۷	۳۳/۵۱
NUS-TR@30	۱۷/۳۳	۵۵/۸۹	۲۶/۴۶	۲۸/۷۰	۷۰/۴۸	۴۰/۷۹
NUS-TR@40	۱۳/۹۹	۶۱/۷۳	۲۲/۸۱	۳۳/۹۵	۶۴/۷۴	۴۴/۵۴
NUS-TR@50	۱۱/۹۷	۶۶/۱۱	۲۰/۲۷	۳۸/۲۷	۵۹/۴۴	۴۶/۵۶
NUS-TR@60	۱۰/۶۱	۶۹/۵۲	۱۸/۴۲	۴۵/۵۳	۵۰/۱۵	۴۷/۷۳

همان‌طورکه نتایج نشان می‌دهد مقدار پارامتر ماکرو F1 در بهترین حالت برابر با ۷۳/۴۷٪ و در بهترین حالت میکرو F1 برابر با ۴۰/۴۵٪ می‌باشد. اختلاف مشاهده‌شده در مقادیر به‌دست آمده حاکی از مؤثر بودن استخراج ویژگی‌های تصاویر با به‌کارگیری شبکه کانوولوشن VGGNet و تأثیرگذاری خوشه‌بند فازی SPFCM در محاسبه تشابه دیداری میان تصاویر در فرآیند تکامل برچسب‌های مجموعه تصاویر با مقیاس بزرگ می‌باشد.

مطابق جدول ۲، بهترین نتیجه ماکرو F1 حاصل، به‌ازای همسایگی n=۶۰ حاصل شده است که اختلاف آن با معیار ماکرو F1 نهایی محاسبه‌شده توسط رویکرد CNN-TC نزدیک به ۸۵/۱۹٪ می‌باشد. به‌منظور مقایسه تأثیر ویژگی‌های حاصل از DCNN و ویژگی‌های اصلی ارائه‌شده از تصاویر توسط NUS-WIDE در فرآیند تکامل تگ‌های تصاویر، نتایج حاصل از دو معیار ماکرو و میکرو F1 به دست آمده از آزمایش‌ها در شکل ۴ و شکل ۵ مورد بررسی قرار گرفته است.



شکل ۴. مقایسه معیار ماکرو F1 حاصل از فرآیند تکامل تگ تصاویر با به‌کارگیری ویژگی‌های حاصل از DCNN و ویژگی‌های اصلی ارائه‌شده توسط NUS-WIDE.



شکل ۵. مقایسه معیار میکرو F1 حاصل از فرآیند تکامل تگ تصاویر با به‌کارگیری ویژگی‌های حاصل از DCNN و ویژگی‌های اصلی ارائه‌شده توسط NUS-WIDE.

همان‌طورکه در شکل ۴ نشان داده شده است، نتایج حاصل از فرآیند تکامل تگ در معیار ماکرو F1 با استفاده از ویژگی‌های استخراج‌شده توسط DCNN و دسته‌بندی تصاویر مشابه با خوشه‌بند SPFCM نسبت به تکرار این آزمایش با استفاده از ویژگی‌های اصلی ارائه‌شده توسط NUS-WIDE بهبود داشته است. مقایسه میان این ویژگی‌ها، با استفاده از معیار میکرو F1 نیز در شکل ۵ ارائه شده است. معیار میکرو F1 حاصل از فرآیند تکامل تگ با به‌کارگیری ویژگی‌های حاصل از DCNN و خوشه‌بندی تصاویر مشابه با روش SPFCM در همسایگی ۵۰= n بهترین مقدار را به خود اختصاص می‌دهد. نتایج حاصل بیانگر مؤثر بودن بهره‌مندی از ویژگی‌های استخراج‌شده از تصاویر توسط DCNN و خوشه‌بندی فازی آن‌ها به‌منظور دسته‌بندی تصاویر مشابه، در فرآیند تکامل برچسب‌های تصاویر می‌باشد.

شکل ۶ نمونه‌ای از تصاویر حاصل از اعمال رویکرد CNN-TC بر دادگان NUS-WIDE را نشان می‌دهد. در این تصویر تگ‌های جدید و مرتبط با محتوای افزوده‌شده به تصویر به رنگ آبی است و در زیر آن‌ها خط کشیده شده است. در صورت عدم‌پیش‌بینی تگ‌های مرتبط اولیه طی فرآیند پالایش، تگ‌ها به رنگ آبی و بر روی آن‌ها خط کشیده شده است. تگ‌های قرمزرنگ، برچسب‌های غیر مرتبط با محتوای تصویر هستند که در صورت پالایش و حذف آن‌ها از مجموعه تگ‌های تصویر، بر روی آن‌ها خط کشیده شده است. همان‌طورکه مشاهده می‌شود بسیاری از برچسب‌های غیرمرتبط با محتوای تصویر بعد از انجام فعالیت پالایش تگ، حذف شده‌اند و برچسب‌های مرتبط بعد از اعمال این رویکرد، به تصاویر افزوده شده‌اند. کلیه آزمایش‌ها در محیط MATLAB-2016a انجام گرفته است.

مقایسه نتایج حاصل از رویکرد پیشنهادی این پژوهش با رویکردهای اخیر فرآیند دشواری است. در پژوهش‌های اخیر اغلب عملیات پالایش و تکامل تگ‌های تصاویر برروی زیرمجموعه‌ای از مجموعه تصاویر NUS-WIDE انجام گرفته است و یا اینکه مشخصات واضحی از تصاویر مورد استفاده و ویژگی‌های متعلق به آن‌ها ارائه نشده است. از طرفی در رویکردهای گذشته، از فرآیند یادگیری عمیق به‌منظور استخراج ویژگی‌های سطح بالا از تصاویر استفاده نشده است، لذا مقایسه رویکرد پیشنهادی این مقاله با آن‌ها منطقی نمی‌باشد. به‌منظور مشاهده عملکرد دیگر رویکردها که بیشترین شباهت را به رویکرد ارائه شده در این پژوهش دارند، در ادامه به معرفی تعدادی از آن‌ها می‌پردازیم.

بررسی رویکردهای تکامل تگ ارائه‌شده نشان می‌دهد که دقت روش CNN-TC در معیار ماکرو F1 نسبت به NUS-TR و همچنین نسبت به سایر روش‌های تکامل تگ اعمال‌شده بر مجموعه تصاویر NUS-WIDE در دیگر پژوهش‌ها، بیشترین مقدار را به خود اختصاص داده است. مقایسه میان این رویکردها درشکل ۷ نشان داده شده است. در ادامه رویکردهای مطرح شده در نمودار شکل ۷ معرفی شده و با روش پیشنهادی در این پژوهش مورد مقایسه قرار می‌گیرند.

در [۱۲] ژو و همکارانش نتایج فعالیت‌های پالایش تگ خود را بر مجموعه NUS-WIDE-270k با ۵۲۱ تگ پیش‌پردازش‌شده توسط ویکی‌پدیا ارائه داده‌اند. بهترین معیار ماکرو F1 محاسبه‌شده در پژوهش آن‌ها درازای پیاده‌سازی رویکرد $TC-CC-ES-LR^{56}$ مقدار F1=٪۳/۳۵ را به خود اختصاص می‌دهد.

به‌کارگیری رویکرد رأی‌دهی نزدیک‌ترین همسایه به‌عنوان روش

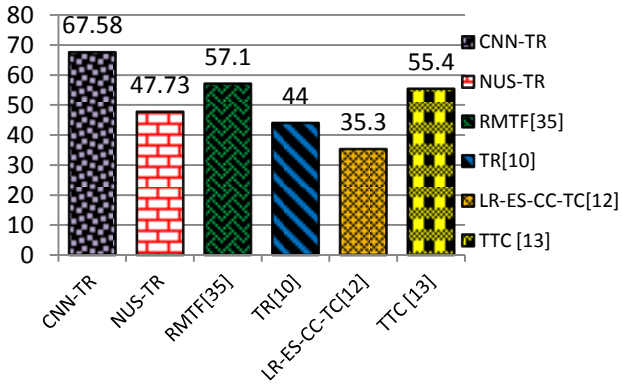
نوع برچسب				
برچسب‌های اولیه	explore, wildlife, animals, Africa, family, elephant, safari	water, tree, birds, reflections, eyes, flight, earth, branch	sunset, beach, trees, reflection, winter, snow, lake, sand, silhouette, ice, path, Michigan	Canada, England, bird, country, Britain, hawk
برچسب‌های تکامل‌یافته با روش NUS-TR	Wildlife, animals, family, nature, sky, blue, trees, water, clouds, explore, art, life, ~~safari~~, ~~elephant~~ , ~~Africa~~	branch, water, reflection, tree, birds, earth, nature, sky, landscape, sea, building, beach, bravo, art, sun, snow, earth, winter, church, ~~eyes~~, ~~flight~~	sunset, trees, winter, nature, sky, water, landscape, tree, autumn, fall, blue, beautiful, photo, color, photograph, brown, colorful, ~~sand~~, ~~path~~, ~~silhouette~~, ~~Michigan~~	hawk, bird, brown, birds, eagle, nature, clouds, sky, ,beautiful, England, Canada, ~~country~~, ~~Britain~~,
برچسب‌های تکامل‌یافته با روش CNN-TC	safari, animals, elephant, wildlife, Africa, nature, elephants, zoo, tusks, cloud, ~~sand~~, ~~family~~, ~~explore~~	tree, reflections, birds, branch, water , view, house, sea, lake, nature, reflection, blue ,flight, art, ~~eyes~~, ~~earth~~	beach, ice, lake, sunset, winter trees, snow, nature, sea, fall, view, reflection, ~~Michigan~~ ~~sand~~, ~~silhouette~~, ~~path~~, ~~snow~~,	hawk, bird, birds, merlin, eagle, nature, ~~Canada~~, ~~country~~ ~~Britain~~, , ~~England~~

شکل ۶. نمونه‌ای از تصاویر با برچسب‌های تکامل یافته توسط رویکرد پیشنهادی در این پژوهش

پالایش تگ ارائه‌شده در این پژوهش، مشابه با پژوهش یوریچیو و همکارانش در [۱۰] می‌باشد، با این تفاوت که تعداد نمونه‌های مورد بررسی در این پژوهش ۲۳۸۲۵۱ تصویر با ۶۸۴ تگ تطبیق داده‌شده با پایگاه داده وردنت می‌باشد. همچنین در پژوهش آن‌ها ضریب α =۵ درنظر گرفته شده است. این درحالی است که در این پژوهش تعداد نمونه‌های مورد بررسی ۲۲۱۸۱۷ تصویر با ۱۰۰۰ تگ با بیشترین فرکانس بوده و ضریب α =۱۰ درنظر گرفته شده است. بهترین میزان معیار ماکرو F1 ارائه‌شده در پژوهش آن‌ها حاصل از رویکرد TR^{57} [۲۹] برابر با ۴۴٪ می‌باشد.

در [۳۵] روشی تحت عنوان RMTF ارائه شده است که در آن براساس یک ساختار یکپارچه سه‌گانه از تصویر، تگ و کاربر به تکامل تگ‌های تصاویر انجام می‌گیرد. سانگ[۵۸] و همکارانش در این پژوهش با انتخاب ۱۲۴۰۹۹ تصویر از میان ۲۴۷۸۴۹ تصویر بارگذاری شده توسط ۵۰۱۲۰ کاربر از دادگان NUS-WIDE به رفع نویز از تگ‌های تصاویر می‌پردازند. معیار F1 محاسبه‌شده توسط این رویکرد میزان ۱/۵۷٪ را به خود اختصاص می‌دهد و ضریب α نیز مشابه این پژوهش، ۱۰ درنظر گرفته شده است.

همان‌طورکه در بخش ۲ عنوان شد، در [۱۳] رویکردی تحت عنوان تکامل تنسور سه‌خوشه‌ای (TTC) ارائه شده است که با توجه به ویژگی‌های پایین‌رتبه تصاویر، اطلاعات ناشی از تگ‌ها و روابط میان کاربران، به تکامل تگ‌های تصاویر می‌پردازد. در این روش یک تنسور سه‌بعدی بر مبنای روابط میانی اطلاعات مذکور ایجاد شده و از آن برای پالایش تگ‌های تصاویر استفاده می‌شود. روش TTC ارائه‌شده در این پژوهش برروی ۲۴۷۸۴۹ تصویر بارگذاری‌شده توسط ۴۹۵۲۸ کاربر از دادگان NUS-WIDE اعمال شده و به‌ازای هر تصویر ۱۰۰۰ برچسب با بیشترین فرکانس درنظر گرفته شده است. بهترین میزان معیار F1 محاسبه‌شده توسط رویکرد TTC مقدار ۵۵/۴٪ را به خود اختصاص می‌دهد.



شکل ۷. نمودار مقایسه روش CNN-TR ارائه شده با سایر روش‌های پالایش تگ در مجموعه تصاویر NUS-WIDE 270k

در کلیه پژوهش‌های معرفی‌شده مقایسه نتایج تکامل تگ‌های تصاویر با نسخه ۸۱ تگی از دادگان NUS-WIDE-270k که توسط خبره‌ها شرح‌گذاری شده‌اند صورت گرفته است. همان‌طورکه مشخص است معیار ماکرو F1 محاسبه‌شده با استفاده از رویکرد تکامل تگ CNN-TC در این پژوهش، با مقدار ۵۸/۶۷٪، بهتر از دیگر رویکردها می‌باشد. این درحالی است که در صورت عدم بهره‌مندی از ویژگی‌های استخراج شده از شبکه کانولوشن VGGNet و عدم‌استفاده از خوشه‌بند فازی به‌منظور محاسبه تصاویر مشابه دیداری معیار ماکرو F1، ۷۳/۴۷٪ محاسبه شده است.

۷ -نتیجه‌گیری

در این پژوهش به‌منظور حل چالش پایین بودن دقت فرآیند بازیابی تصاویر در موتورهای جستجو، رویکردی در دسته الگوریتم‌های مبتنی بر نمونه در تکامل تگ‌های تصاویر ارائه شد. در این رویکرد ابتدا کلیه تصاویر به‌عنوان ورودی به شبکه کانولوشن VGGNet، داده شد و از شبکه، دو بردار با ۴۰۹۶ ویژگی و ۱۰۰۰ برچسب به‌ازای هر تصویر استخراج شد. به‌منظور مقیاس‌پذیری سیستم تکامل تگ‌های تصاویر، خوشه‌بند فازی SPFCM بر بردار ویژگی‌های استخراج‌شده اعمال گردید که در نتیجه تصاویر مشابه دیداری دسته‌بندی شدند. سپس در هر خوشه با استفاده از تشابه معنایی میان تصاویر و محاسبه هم‌پوشانی میان تگ‌ها تصاویر همسایه انتخاب شدند. در نهایت با بهره‌گیری از تکنیک رأی‌گیری از تصاویر همسایه، فرآیند تکامل برچسب‌های تصاویر انجام شد. نتایج حاصل بیانگر مؤثر بودن رویکرد ارائه شده در حوزه تکامل تگ‌های تصاویر و رفع نویز از برچسب‌های هر تصویر می‌باشد. در پژوهش‌های آتی سعی داریم که به بررسی تأثیر رویکرد تکامل تگ ارائه شده بر فرآیند شرح‌گذاری خودکار تصاویر بپردازیم و بهره‌مندی از دیگر روش‌های یادگیری عمیق، در این حوزه را مورد بررسی قرار دهیم.

مراجع

[1] R. Datta, D. Joshi, J. Li and J. Z. Wang, “Image retrieval: ideas, influences and trends of the new age,” ACM Computing Surveys, vol. 40, no. 2, 2008.

[۲] مریم تقی‌زاده و عبداله چاله‌چاله، «مدلی به‌منظور بازیابی تصاویر مبتنی بر چند درخواست»، مجله مهندسی برق دانشگاه تبریز، دوره ۴۷، شماره ۳، صفحه ۹۰۳-۸۹۳، ۱۳۹۶.

[3] X. Li, L. Chen, L. Zhang, F. Lin, and W.-Y. Ma, “Image annotation by large-scale content-based image retrieval,” ACM International Conference on Multimedia, 2006.

[4] X. Rui, M. Li, Z. Li, W.-Y. Ma, and N. Yu, “Bipartite graph reinforcement model for web image annotation,” ACM International Conference on Multimedia, 2007.

[5] M. J. Huiskes and M. S. Lew, “The MIR flickr retrieval evaluation”, ACM International Conference on Multimedia Information retrieval, 2008.

[۶] هنگامه دلجویی و امیرمسعود افتخاری مقدم، «حاشیه‌نویسی خودکار تصویر با استفاده از ارتباط معنایی بین نواحی مبتنی بر تئوری تصمیم چند شرطی»، مجله مهندسی برق دانشگاه تبریز، دوره ۴۲، شماره ۲، صفحه ۳۹-۵۲، ۱۳۹۲.

[7] C. Blake and C. J. Merz, UCI Repository of Machine LearningDatabases,http://mlearn.ics.uci.edu/MLRepository.html, University of California, Irvine, School of Information and Computer Sciences, vol 55. 1998.
[8] T. C. Havens, J. C. Bezdek, C. Leckie, L. O. Hall and M. Palaniswami, "Fuzzy c-means algorithms for very large data," IEEE Transactions on Fuzzy Systems, vol. 20, no. 6, 2012.
[9] X. Li, T. Uricchio, L. Ballan, M. Bertini, C. G. M. Snoek and A. Del Bimbo, "Socializing the semantic gap: a comparative survey on image tag assignment, refinement, and retrieval," ACM Computing Surveys (CSUR), vol. 49, no. 1, 2016.
[10] S. Lee, W. De Neve and Y. M. Ro, "Visually weighted neighbor voting for image tag relevance learning," Multimedia Tools Applications, vol. 72, no. 2, pp. 1363–1386, 2014.
[11] T. Uricchio, L. Ballan, M. Bertini and A. Del Bimbo, "An evaluation of nearest-neighbor methods for tag refinement," International Conference on Multimedia and Expo (ICME), 2013.
[12] L. Chen, D. Xu, I. W. Tsang and J. Luo, "Tag-based image retrieval improved by augmented features and group-based refinement," IEEE Transactions on Multimedia, vol. 14, no. 4, pp. 1057–1067, 2012.
[13] G. Zhu, S. Yan and Y. Ma, "Image tag refinement towards low-rank, content-tag prior and error sparsity," International Conference of Multimedia, pp. 461–470, 2010.
[14] J.Tang, X.Shu, G.J.Qi, Z.Li, M.Wang, S.Yan and R.Jain, "Tri-clustered tensor completion for social-aware image tag refinement," IEEE Transactions on Pattern Analysis and Machine Intelligence., vol. 39, no. 8, pp. 1662–1674, 2017.
[15] X. Yang and F. Yang, "Completing tags by local learning: a novel image tag completion method based on neighborhood tag vector predictor," Neural Computing and Applications , vol. 27, no. 8, pp. 2407–2416, 2016.
[16] Z. Feng, S. Feng, R. Jin and A. K. Jain, "Image tag completion by noisy matrix recovery," European Conference on Computer Vision, pp. 424–438, 2014.
[17] Y. Bengio "Learning deep architectures for AI," Foundations and Trends in Machine Learning, vol. 2, no. 1, 2009.
[18] S. Lawrence, C. L. Giles, A. C. Tsoi and A. D. Back, "Face recognition: a convolutional neural-network approach," IEEE Transactions on Neural Networks , vol. 8, no. 1, 1997.
[19] G. E. Hinton, "Deep belief networks," Scholarpedia, vol. 4, no. 5, 2009.
[20] T. Mikolov, M. Karafiát, L. Burget, J. Cernock and S. Khudanpur, "Recurrent neural network based language model," Interspeech, vol. 2, pp.3, 2010.
[21] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," Computer Vision and Pattern Recognition (CVPR), pp. 580-587, 2014.
[22] G. Hinton, L. Deng, D. Yu, G. Dahl, AR .Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, TN. Sainath and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," IEEE Signal Processing Magazine, vol. 29, no. 6, 2012.
[23] R. Collobert and J. Weston, "A unified architecture for natural language processing: deep neural networks with multitask learning," International Conference on Machine Learning, 2008.
[24] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," International Conference on Learning Representations arXiv preprint arXiv:1409.1556, 2014.
[25] J. Deng, W. Dong, R. Socher, L. Li, K. Li and L. Fei-Fei, "Imagenet: a large-scale hierarchical image database," Computer Vision and Pattern Recognition, 2009.
[26] J. C. Dunn, "A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters," 1973.
[27] V. Schwämmle and O. N. Jensen, "A simple and fast method to determine the parameters for fuzzy c–means cluster analysis," Bioinformatics, vol. 26, no. 22, 2010.
[28] D. Dembélé and P. Kastner, "Fuzzy c-means method for clustering microarray data," Bioinformatics, vol. 19, no. 8, 2003.
[29] D. Liu, X.-S. Hua, M. Wang and H.-J. Zhang, "Image retagging," International Conference on Multimedia, 2010.
[30] X. Li, C. G. M. Snoek and M. Worring, "Learning social tag relevance by neighbor voting," IEEE Transactions on Multimedia, vol. 11, no. 7, 2009.
[31] T.-S. Chua, J. Tang, R. Hong, H. Li, Z. Luo and Y. Zheng, "NUS-WIDE: a real-world web image database from national university of Singapore," ACM International Conference on Image and Video Retrieval, 2009.
[32] Z. Lin, G. Ding, M. Hu, Y. Lin and S. S. Ge, "Image tag completion via dual-view linear sparse reconstructions," Computer Vision Image Understanding, vol. 124, 2014.
[33] S. Zhu, S. Aloufi and A. El Saddik, "Utilizing image social clues for automated image tagging," IEEE International Conference on Multimedia and Expo (ICME), 2015.
[34] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu and M. S. Lew, "Deep learning for visual understanding: A review," Neurocomputing, vol. 187, 2016.
[35] NUS-WIDE Homepage, Lab for Media Search, http://lms.comp.nus.edu.sg/research/NUS-WIDE.html, Accessed 07.07.2017.
[36] J. Sang, C. Xu, and J. Liu, "User-aware image tag refinement via ternary semantic analysis," IEEE Transactions on Multimedia, vol. 14, no. 3, 2012.

زیرنویس‌ها

1 Image Annotation
2 Content Base Image Retrieval
3 Tag Base Image Retrieval
4 Image Tag Completion
5 Deep Learning
6 Deep Convolutional Neural Network
7 Single Pass Fuzzy C-Means
8 Instance Based
9 Model Based

10 Uricchio
11 Chen
12 Augmented Feature Support Vector Machine
13 Laplacian Regularize Least Square
14 Flickr
15 Tri-clustered Tensor Completion
16 Scoring Vector
17 Local Linear Function
18 Tag Completion by Noisy Matrix Recovery
19 Laplacian-based Component
20 Convolution Neural Network
21 Deep Belief Network
22 Recurrent Neural Network
23 Restricted Boltzmann Machines

24 Sparse Coding
25 Convolution
26 Pooling layer
27 Fully connected layer
28 Features Map
29 Activation Map
30 Overfitting
31 Feed Forward
32 Back Propagation
33 Loss Function
34 Chain Rule
35 Visual Geometric Group Network
36 Karen Simonyan
37 Andrew Zisserman
38 ImageNet Large Scale Visual Recognition Challenge 2014
39 Image Classification
40 Object Detection
41 Dunn
42 Deep Convolutional Neural Network Tag Refinement
43 www.flickr.com
44 Nearest Neighbor Voting
26 Multi Objective
46 Google Distance
47 Neighbor Voting
48 Color Histogram
49 Color Moment
50 Color Correlogram
51 Edge Direction Histogram
52 Wavelet Texture
53 Average Precision (AP)
54 Average Recall (AR)
55 Cross Validation
56 Tag Correlation-Content Consistency-Error Sparsity-Low Rank
57 Learning Tag Relevance from Visual Neighbors
58 Sung