

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

شیمای جوانمردی^۱، دانشجوی دکتری؛ علی محمد 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. از این‌رو 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 کانولوشن 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 کانولوشن 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. Javanmardi¹, Ph.D Student; A. M. Latif², Associate Professor; V. Derhami³, 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: علی محمد 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 برچسب‌گذاری شده 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 بازیابی شده حاوی تمام درخواست‌های مورد جستجو باشد. در روش ارائه شده 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 استفاده شده 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 ایجاد شده 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 و موفق‌ترین 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 معرفی شده 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 عنوان شده 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 استفاده شده 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 ارائه شده در [۸]، روش‌های تکامل و پالایش برچسب‌های تصاویر را می‌توان به دو دسته روش‌های مبتنی بر استقرا و روش‌های مبتنی بر استنتاج تقسیم کرد. این دسته‌بندی براساس ایجاد و

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 در نظر 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 کاهش 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 از دست‌رفته 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

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 جداسازی 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 هر کدام 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 مشتعل 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 شده است. هدف پژوهش آن‌ها بهینه‌سازی اندازه‌گیری چهار جنبه (۱) همبستگی معنایی میان تگ‌های کم‌رتبه

می‌شود. با این عمل نگاهت فعال‌ساز^{۲۹} دوبعدی برای آن فیلتر ایجاد می‌شود.

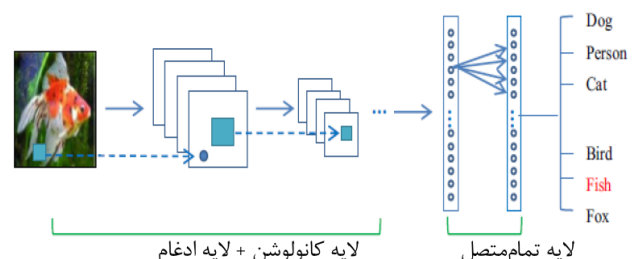
ویژگی‌های استخراج شده از لایه‌های کانولوشن خاصیت سلسله‌مراتبی دارند، به این معنی که به‌طور مثال در لایه‌های ابتدایی گوشه‌ها و خط‌ها و لبه‌ها یادگرفته می‌شوند و در لایه‌های عمیق‌تر به‌ترتیب ویژگی‌هایی با سطوح بالاتر، همانند اشیای درون تصویر، استخراج می‌شوند. با این‌روش شبکه‌های عصبی کانولوشنی تصویر اصلی را طی لایه‌های مختلف از مقادیر اصلی 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 و درنهایت 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 گرادبان 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 تا به حال 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

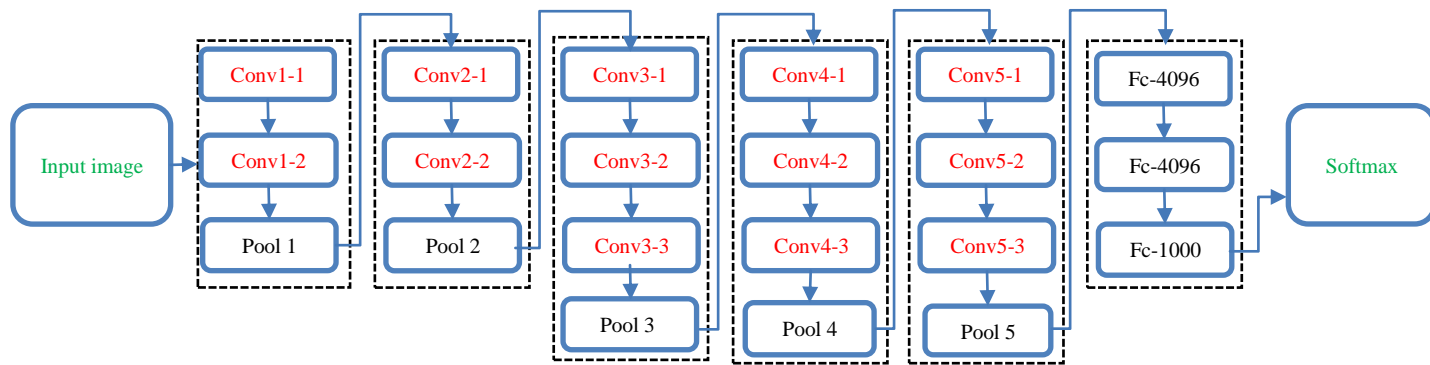
ویژگی‌های سطح پایین تصویر و مفاهیم سطح بالای آن، می‌باشد. در سال‌های اخیر از یادگیری عمیق به‌عنوان رویکردی مؤثر در کاهش این شکاف معنایی استفاده شده است [۱۶]. این رویکرد، تکنیکی مبتنی بر شبکه‌های عصبی است و به‌عنوان زیر مجموعه‌ای از روش‌های یادگیری ماشین به‌شمار می‌رود. معماری‌های یادگیری عمیق متنوعی وجود دارد که از جمله آن‌ها می‌توان به شبکه عصبی کانولوشن (CNN)^{۳۵} [۱۷]، شبکه باور عمیق^{۳۱} [۱۸] و شبکه عصبی بازگشت‌کننده^{۳۲} [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 چهار دسته (۱) CNN، (۲) ماشین بولترمن محدود (RBMS)^{۳۳}، (۳) رمزنگارخودکار و (۴) کدنویسی پراکنده^{۳۴} 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)^{۳۵}، لایه ادغام^{۳۶} و لایه کاملاً متصل (FC)^{۳۷}. 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 در حالی 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 از لحاظ 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 نگاهت و ویژگی^{۳۸} متفاوت از تصاویر در هر لایه کانولوشن، این فیلترها در امتداد پهنا و ارتفاع بر سطح تصویر غلطانده



شکل ۲. نمایی از معماری شبکه VGGNet

داخلی می‌باشد که در این مقاله از نرم اقلیدسی استفاده شده است. الگوریتم FCM به صورت الگوریتم ۱ می‌باشد.

۱ الگوریتم FCM

ورودی: X, c, m, n_s

خروجی: V, U

While $\max_{1 \leq k \leq c} \{ \|V_{k,new} - V_{k,old}\|^2 \} > \epsilon$ 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 $\epsilon = 10^{-2}$ در نظر گرفته شده است. در صورت در نظر 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 به دست می‌آید:

$$Jm(U, V) = \sum_{i=1}^c \sum_{j=1}^n W_j u_{ij} \|X_j - V_i\|_A^2 \quad (2)$$

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

۲ الگوریتم SPFCM

ورودی: X, c, m, n_s

خروجی: V

انتخاب زیرمجموعه‌ای از X با طول n_s

$X = \{X_1, X_2, \dots, X_{n_s}\}$

1. $W = 1n_s$

2. $U, V = \text{WFCM}(X, c, m, W)$

3. For $L=2$ to s do

$$w'_i = \sum_{j=1}^{n_s} (u_{ij}) w_j, i = 1, \dots, c$$

4. $W = \{W' \cup 1n_s\}$

5. $U, V = \text{WFCM}(\{V \cup X_L\}, c, m, W, V)$

پایان

حوزه آنالیز تصاویر، معماری VGGNet^{۳۵} است. شبکه عصبی کانولوشن عمیقی است که توسط کارن سیمونیان^{۳۶} و اندرو زیسرمن^{۳۷} توسعه داده شده است. این معماری به عنوان 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 ارائه شده در آن رقابت، انتخاب شده است [۲۳]. نسخه از پیش‌آموزش دیده شده از این شبکه، توسط ۲/۱ میلیون تصویر موجود در دادگان ImageNet [۲۴] با رزولوشن بالا و سایز 227×227 در ۱۰۰۰ طبقه یادگیری شده است. نسخه نهایی بهترین شبکه از این معماری شامل شانزده لایه Conv/FC بوده و یک معماری به شدت همگن می‌باشد که در این مقاله مورد استفاده قرار می‌گیرد. شکل ۲ نمایی از معماری ارائه شده 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 به منظور دسته‌بندی تصاویر مشابه دیداری، در بخش بعد به معرفی کامل الگوریتم FCM و نسخه SPFCM ارائه شده از آن، می‌پردازیم.

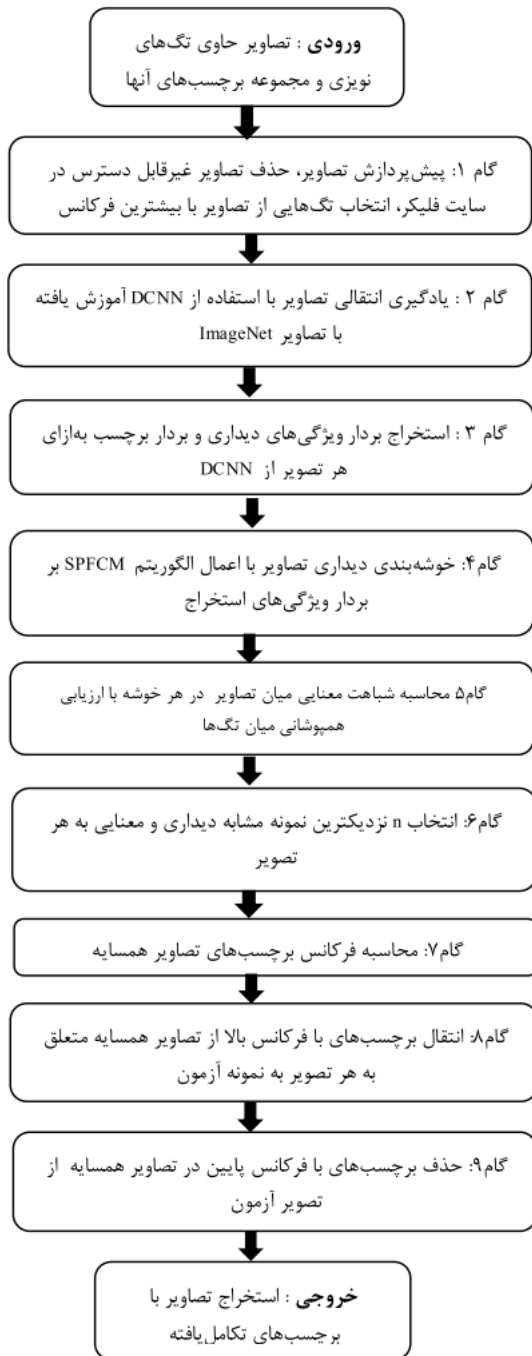
۴ الگوریتم FCM

این الگوریتم نخستین بار در سال ۱۹۷۳ توسط دان^{۴۱} [۲۵] ارائه شد و از جمله الگوریتم‌های خوشه‌بندی فازی به شمار می‌آید. در این الگوریتم درجه عضویت n نمونه ورودی به c خوشه مشخص، محاسبه می‌شود. تعداد خوشه‌ها توسط کاربر مشخص می‌شود. تابع هدف محاسبه شده توسط این الگوریتم به صورت رابطه (۱) می‌باشد.

(۱)

در این رابطه X بیانگر ویژگی‌های متعلق به نمونه‌های ورودی، U نشان‌دهنده ماتریس بخش‌بندی با ابعاد $c \times n$ 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 بر روی دادگان انتخابی، مشابه [۷]، $m = 2/1$ در نظر گرفته شده است. $\| \cdot \|_A$ نرم ضرب

می‌پردازیم. ابتدا به منظور 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 ارائه شده در این پژوهش

۴- الگوریتم SPFCM

الگوریتم SPFCM نسخه‌ای توسعه یافته 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 محاسبه شده از مرحله قبل (V) می‌باشد. بنابراین در هر تکرار این الگوریتم ($c+ns$) نمونه خوشه‌بندی شده 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 به منظور 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 $\{V \cup X_L\}$ 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 به منظور 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، به منظور 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 رویکرد نخست کلیه تصاویر مورد پیش‌پردازش قرار می‌گیرند. در این راستا ابتدا تصاویری که در وبسایت فلیکر^{۴۳} با لینک‌های معتبر قابل دسترس هستند دریافت می‌شوند. عملیات انجام شده پیش از فرآیند تکامل تگ‌های تصاویر عبارت‌اند از: ۱- تنک کردن تگ‌ها، ۲- استخراج ویژگی‌های تصاویر با به کارگیری CNN، ۳-

در این بخش رویکرد ارائه شده در این مقاله به منظور 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 به معرفی مدل پیشنهادی

هر تصویر در حافظه، از الگوریتم خوشه‌بندی فازی SPFCM استفاده شده است. بهره‌مندی از این خوشه‌بند به مقابله با چالش‌های مطرح شده می‌پردازد.

پیش از ارائه روابط استفاده شده در این مقاله ابتدا به معرفی متغیرهای موجود می‌پردازیم. مجموعه $X = \{x_1, x_2, \dots, x_N\}$ بیانگر مجموعه تصاویر اولیه می‌باشد که تگ‌های تخصیص داده شده به هر تصویر با $T = \{w_1, w_2, \dots, w_M\}$ 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 به منظور بیان تعلق و عدم تعلق 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. $y_i = (y_{i1}, y_{i2}, \dots, y_{im})$ بیانگر بردار درجه اطمینان تخصیص تگ‌ها به i مین تصویر و بردار $v_i = (v_{i1}, v_{i2}, \dots, v_{iz})$ نشان‌دهنده بردار ویژگی‌های دیداری تصاویر می‌باشد.

همان‌طور که 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 هر کدام 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 برجسب، و از دومین لایه FC شبکه، بردار v_i حاوی ۴۰۹۶ 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.

همان‌طور که 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 در نظر 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، از این‌رو 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 در نظر گرفتن یک حد آستانه، بردار y_i که

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 درجه احتمال تعلق هر نمونه به هر کلاس، کلیه تصاویر به‌عنوان ورودی به DCNN داده می‌شود. در این راستا ابتدا هر کدام از نمونه‌های x_i را به‌عنوان ورودی به شبکه کانولوشن VGGNet داده می‌شود و به‌ازای هر نمونه، بردار y_i با $M=1000$ ضریب برجسب از آخرین لایه FC و بردار v_i با $Z=4096$ ویژگی از دومین لایه FC شبکه استخراج می‌شود.

در فاز سوم، یک بردار ویژگی به ابعاد 1×4096 و یک بردار برجسب 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 از جمله چالش‌های مطرح در این حوزه می‌باشد. لذا در مرحله چهارم به‌منظور مقیاس‌پذیری فرآیند تکامل برجسب‌های 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 برجسب‌های هر تصویر، تشابه معنایی میان تصاویر ارزیابی شده و تصاویر همسایه متعلق به هر نمونه محاسبه می‌شود. در نهایت 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 کانولوشن عمیق برای استخراج ویژگی‌های مؤثر از تصاویر، مؤثر بودن این رویکرد را در فرآیند تکامل تگ‌های تصاویر نشان می‌دهد.

رویکرد ارائه شده علاوه بر 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 یادگیری شده 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 محاسبه شده برای

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 مورد استفاده در ارزیابی رویکرد CNN-TC و معیارهای ارزیابی محاسبه‌شده، نتایج ارزیابی روش پیشنهادی بر مجموعه دادگان معرفی شده ارائه می‌گردد. به منظور ارزیابی کیفیت رویکرد CNN-TC از مجموعه تصاویر با مقیاس بزرگ NUS-WIDE-270K استفاده شده است [۳۰]. این مجموعه شامل ۲۶۹۶۴۸ تصویر جمع‌آوری شده از سایت فلیکر به همراه ۵۰۱۸ تگ ایجاد شده توسط کاربران می‌باشد که با حذف آدرس‌های نامعتبر و تصاویر غیرقابل دسترس 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 به عنوان NUS-WIDE-220K یاد می‌شود.

با توجه به آنکه تگ‌های ابتدایی ایجاد شده توسط کاربران به شدت ذهنی و یا غیر مرتبط با محتوای تصویر هستند، بنابراین با هدف بهبود کیفیت فرآیند تکامل برچسب‌های تصویر به حذف آن‌ها می‌پردازیم. از این رو 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 LAB^{۴۸} با ۶۴ بعد، ۲- ممان رنگ^{۴۹} مبتنی بر بلاک با ۲۲۵ بعد، ۳- کورلوگرام رنگ HSV^{۵۰} 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. به منظور بررسی میزان کارایی رویکرد CNN-TC پیشنهاد شده بر فرآیند تکامل شرح‌گذاری تصاویر، نتایج حاصل از آزمایش‌های انجام شده با به کارگیری شبکه کانولوشن VGGNet و بدون استفاده از آن مورد مقایسه قرار می‌گیرد. در ارزیابی نتایج حاصل از رویکرد CNN-TC، مشابه روش ارائه شده در [۲۱-۱۹] نسخه ۸۱ تگی از مجموعه NUS-WIDE-220k که توسط افراد خبره شرح‌گذاری شده است، به عنوان ماتریس مبنا در نظر گرفته می‌شود. برای ارزیابی نتایج تکامل تگ‌های تصاویر سه معیار ارزیابی شناخته شده 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، به منظور 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

حاوی ضرایب تعلق تصاویر به ۱۰۰۰ طبقه از تصاویر ImageNet می‌باشد، به یک بردار باینری تبدیل می‌شود. به این صورت که ضرایب بیشتر از یک حد آستانه فرضی همچون β را به ۱ و سایر ضرایب به صفر که نشان از عدم تعلق 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 $d(W_i, W_j)$ فاصله میان تگ‌های W_i و W_j 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 $q(W_i)$ و $q(W_j)$ تعداد تصاویر حاوی تگ W_i و W_j می‌باشد. $q(W_i, W_j)$ تعداد تصاویری می‌باشد که حاوی هر دو تگ W_i و W_j است. در نهایت با استفاده از تشابه دیداری و معنایی به کار گرفته شده تصاویر همسایه متعلق به هر تصویر محاسبه می‌شود.

(۴)

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

(۵)

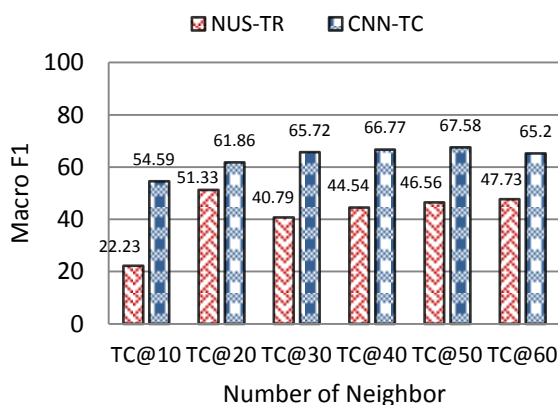
در راستای تکامل تگ‌های تصاویر انتخابی، پیش‌بینی لیستی از تگ‌ها بر مبنای تگ‌های تصاویر همسایه مشابه، می‌تواند مؤثر باشد. از این رو در چارچوب پالایش تگ ارائه شده مشابه [۲۳-۲۱]، از روش رأی‌گیری همسایه^{۴۷} استفاده شده است. در طول انجام فرآیند تکامل تگ تصاویر، تگ‌های هر تصویر با تگ‌های تصاویر همسایه مورد 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 به کار گرفته شده در شبکه کانولوشن VGGNet متفاوت می‌باشد و فرض بر این است که هیچ‌گونه ارتباطی میان آن‌ها وجود ندارد.

به منظور 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، به منظور محاسبه میزان کارایی رویکرد پیشنهادی با استفاده از

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

همان‌طور که 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 مشاهده شده در مقادیر به دست آمده حاکی از مؤثر بودن استخراج ویژگی‌های تصاویر با به کارگیری شبکه کانولوشن 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 به کارگیری ویژگی‌های حاصل از DCNN و ویژگی‌های اصلی ارائه‌شده توسط NUS-WIDE

محاسبه شده است. از این‌رو 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 کانولوشن 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 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 n-CNN-TC بر دادگان NUS-WIDE-220k به‌ازای n همسایه مختلف و به کارگیری خوشه‌بند SPFCM

روش	دقت میکرو (%)	فراخوان میکرو (%)	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	۵۲/۶۸	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 عدم بهره‌مندی 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 n-NUS-TR بر دادگان NUS-WIDE-220k به‌ازای n همسایه مختلف

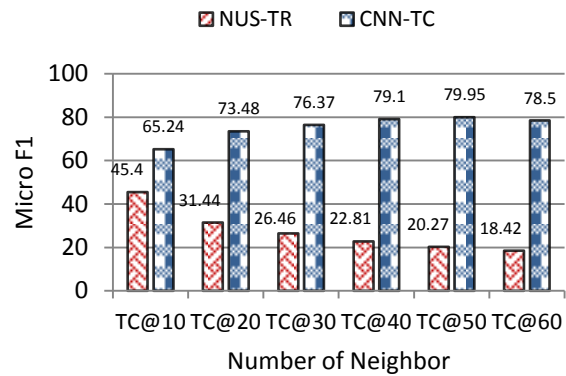
تصویر هستند که در صورت پالایش و حذف آن‌ها از مجموعه تگ‌های تصویر، بر روی آن‌ها خط کشیده شده است. همان‌طور که مشاهده می‌شود بسیاری از برچسب‌های غیرمرتبط با محتوای تصویر بعد از انجام فعالیت پالایش تگ، حذف شده‌اند و برچسب‌های مرتبط بعد از اعمال این رویکرد، به تصاویر افزوده شده‌اند. کلیه آزمایش‌ها در محیط 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 بر روی 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 به منظور استخراج ویژگی‌های سطح بالا از تصاویر استفاده نشده است، لذا مقایسه رویکرد پیشنهادی این مقاله با آن‌ها منطقی نمی‌باشد. به منظور 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 ارائه شده نشان می‌دهد که دقت روش CNN-TC در معیار ماکرو F1 نسبت به NUS-TR و همچنین نسبت به سایر روش‌های تکامل تگ اعمال شده بر مجموعه تصاویر NUS-WIDE در دیگر پژوهش‌ها، بیشترین مقدار را به خود اختصاص داده است. مقایسه میان این رویکردها در شکل 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 پیش‌پردازش شده توسط ویکی‌پدیا ارائه داده‌اند. بهترین معیار ماکرو F1 محاسبه شده در پژوهش آن‌ها درازای پیاده‌سازی رویکرد TC-CC-ES-LR^{۵۶} مقدار $F1 = \frac{3}{3.5}$ را به خود اختصاص می‌دهد.





به کارگیری رویکرد رأی‌دهی نزدیک‌ترین همسایه به‌عنوان روش



شکل ۵. مقایسه معیار میکرو F1 حاصل از فرآیند تکامل تگ تصاویر با به کارگیری ویژگی‌های حاصل از DCNN و ویژگی‌های اصلی ارائه شده توسط NUS-WIDE.

همان‌طور که در شکل ۴ نشان داده شده است، نتایج حاصل از فرآیند تکامل تگ در معیار ماکرو F1 با استفاده از ویژگی‌های استخراج شده توسط DCNN و دسته‌بندی تصاویر مشابه با خوشه‌بندی SPFCM نسبت به تکرار این آزمایش با استفاده از ویژگی‌های اصلی ارائه شده 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 به کارگیری ویژگی‌های حاصل از DCNN و خوشه‌بندی تصاویر مشابه با روش SPFCM همسایگی $n = 50$ بهترین مقدار را به خود اختصاص می‌دهد. نتایج حاصل بیانگر مؤثر بودن بهره‌مندی از ویژگی‌های استخراج شده از تصاویر توسط DCNN و خوشه‌بندی فازی آن‌ها به منظور 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 افزوده شده به تصویر به رنگ آبی است و در زیر آن‌ها خط کشیده شده است. در صورت عدم پیش‌بینی تگ‌های مرتبط اولیه طی فرآیند پالایش، تگ‌ها به رنگ آبی و بر روی آن‌ها خط کشیده شده است. تگ‌های قرمز رنگ، برچسب‌های غیر مرتبط با محتوای

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

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

خبره‌ها شرح‌گذاری شده‌اند صورت گرفته است. همان‌طور که مشخص است معیار ماکرو FI محاسبه شده با استفاده از رویکرد تکامل تگ CNN-TC در این پژوهش، با مقدار $0.58/67\%$ ، بهتر از دیگر رویکردها می‌باشد. این درحالی است که در صورت عدم بهره‌مندی از ویژگی‌های استخراج شده از شبکه کانولوشن VGGNet و عدم استفاده از خوشه‌بند فازی به‌منظور محاسبه تصاویر مشابه دیداری معیار ماکرو FI، $0.73/47\%$ 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 4096 ویژگی و 1000 برچسب به‌ازای هر تصویر استخراج شد. به‌منظور مقیاس‌پذیری سیستم تکامل تگ‌های تصاویر، خوشه‌بند فازی SPFCM بر بردار ویژگی‌های استخراج شده 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

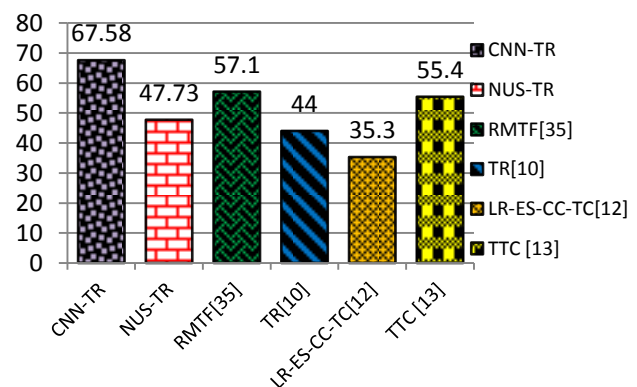
- [1] R. Datta, D. Joshi, J. Li and J. Z. Wang, "Image retrieval: ideas, influences and trends of the new age," ACM Computing Surveys, vol. 40, no. 2, 2008.
- [۲] مریم تقی‌زاده و عبدالله چاله‌چاله، «مدلی به‌منظور بازیابی تصاویر مبتنی بر چند درخواست»، مجله مهندسی برق دانشگاه تبریز، دوره ۴۷، شماره ۳، صفحه 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.

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 238251 تصویر با 684 تگ تطبیق داده شده با پایگاه داده وردنت می‌باشد. همچنین در پژوهش آن‌ها ضریب $\alpha = 5$ در نظر گرفته شده است. این درحالی است که در این پژوهش تعداد نمونه‌های مورد بررسی 221817 تصویر با 1000 تگ با بیشترین فرکانس بوده و ضریب $\alpha = 10$ در نظر گرفته شده است. بهترین میزان معیار ماکرو FI ارائه‌شده در پژوهش آن‌ها حاصل از رویکرد $57TR$ [۲۹] برابر با 0.44% 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 124099 تصویر از میان 247849 تصویر بارگذاری شده توسط 50120 کاربر از دادگان NUS-WIDE به رفع نویز از تگ‌های تصاویر می‌پردازند. معیار FI محاسبه شده توسط این رویکرد میزان 0.157% را به خود اختصاص می‌دهد و ضریب α نیز مشابه این پژوهش، 10 در نظر گرفته شده است.

همان‌طور که 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 ارائه شده در این پژوهش بر روی 247849 تصویر بارگذاری شده توسط 49528 کاربر از دادگان NUS-WIDE اعمال شده و به‌ازای هر تصویر 1000 برچسب با بیشترین فرکانس در نظر گرفته شده است. بهترین میزان معیار FI محاسبه شده توسط رویکرد TTC مقدار $0.55/4\%$ 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 معرفی شده مقایسه نتایج تکامل تگ‌های تصاویر با نسخه ۸۱ تگی از دادگان NUS-WIDE-270k که توسط

- [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.
- [7] C. Blake and C. J. Merz, *UCI Repository of Machine Learning Databases*, <http://mllearn.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.

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

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

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