

بازیابی تصویر مبتنی بر محتوا با استفاده از شبکه‌های عصبی کانولوشن

عمیق

امیر سزاوار^۱، دانشجوی کارشناسی ارشد؛ حسن فرسی^۲، استاد؛ سجاد محمدزاده^۳، استادیار

۱- دانشکده مهندسی برق و کامپیوتر - دانشگاه بیرجند - بیرجند - ایران - a.sezavar@birjand.ac.ir

۲- دانشکده مهندسی برق و کامپیوتر - دانشگاه بیرجند - بیرجند - ایران - hfarsi@birjand.ac.ir

۳- دانشکده فنی و مهندسی فردوس - دانشگاه بیرجند - بیرجند - ایران - s.mohamadzadeh@birjand.ac.ir

چکیده: بازیابی تصویر یکی از موضوعات مهم و چالش‌برانگیز در حوزه بینایی ماشین و پردازش تصویر است. تاکنون تحقیقات زیادی در زمینه بازیابی بر روی روش‌های مختلف انجام شده است. از دهه ۷۰ میلادی بازیابی مبتنی بر کلمه معرفی شد و پس از گذشت دو دهه با توجه به افزایش حجم داده ذخیره شده و ناکارآمدی این روش‌ها، بازیابی مبتنی بر محتوا پایه‌گذاری شد. در این روش محققان با توجه به استخراج ویژگی از تصاویر، به نتیجه بهتری دست یافتند. اما وجود فاصله معنایی میان ادراک انسان و ویژگی‌های سطح پایین استخراج شده از تصاویر از یک سو و از سوی دیگر افزایش بی‌رویه حجم داده پیرو پیشرفت فناوری، باعث شد تا روش‌های ابداع شده ناکارآمد شوند و پژوهش‌ها به سمت الگوریتم‌های جدیدتر معطوف شوند. با توجه به موفقیت و رشد چشم‌گیر الگوریتم‌های یادگیری عمیق، بر آن شدیم تا روشی جدید مبتنی بر یادگیری عمیق به منظور بازیابی تصویر پیاده‌سازی کنیم. در این مقاله پس از معرفی شبکه‌های عصبی کانولوشن به عنوان زیرمجموعه‌ای از روش‌های یادگیری عمیق، سیستم بازیابی مبتنی بر شبکه عصبی کانولوشن پیشنهاد می‌دهیم و با اجرای این سیستم بر روی سه پایگاه داده ALOI، Corel و Mpeg7 و محاسبه سه معیار $P(0.5)$ ، $P(1)$ و ANMRR و مقایسه آن‌ها با روش‌های دیگر، نشان می‌دهیم که این مدل از دقت خیلی بالاتری نسبت به سایر الگوریتم‌های مرسوم، برخوردار است.

واژه‌های کلیدی: بازیابی تصویر، روش مبتنی بر کلمه، روش مبتنی بر محتوا، یادگیری عمیق، شبکه‌های عصبی کانولوشن

Content-Based Image Retrieval using Deep Convolutional Neural Networks

A. Sezavar¹, M.Sc Student; H. Farsi², Professor; S. Mohamadzadeh³, Assistant professor

1- Faculty of Electrical and Computer Engineering, University of Birjand, Birjand, Iran, Email: a.sezavar@birjand.ac.ir

2- Faculty of Electrical and Computer Engineering, University of Birjand, Birjand, Iran, Email: hfarsi@birjand.ac.ir

3- Faculty of Technical and Engineering of Ferdows, University of Birjand, Birjand, Iran, Email: s.mohamadzadeh@birjand.ac.ir

Abstract: Image retrieval is an important issue of machine vision and image processing. Many researches have been done in image retrieval. In 70's, Text-Based image retrieval had been created before Content-Based image retrieval have been introduced since 90's cause of large amount of data stored and inefficient previous methods. On this way, researcher reached better conclusion by extracting features from pictures. Semantic gap between these features and human concept, and burst increase in amount of images which were saved, caused researchers to think about new algorithms. Excellent successes on deep learning algorithms encourage us to implant a new method for image retrieval based on deep learning. In this paper, after reviewing deep convolutional neural networks as a kind of deep learning methods, we introduce a new retrieval system based on deep convolutional neural networks and by testing it on three famous databases, ALOI, Corel and MPEG7, computing $P(0.5)$, $P(1)$ and ANMRR and comparing them with other methods which have been used since recent years, we show the superior accuracy of this method in comparison to the other methods.

Keywords: Image retrieval, Text-Based image retrieval, Content-Based image retrieval, Deep learning, convolutional neural networks

تاریخ ارسال مقاله: ۱۳۹۶/۰۱/۲۱

تاریخ اصلاح مقاله: ۱۳۹۶/۰۵/۱۶، ۱۳۹۶/۰۷/۰۲ و ۱۳۹۶/۰۷/۲۴

تاریخ پذیرش مقاله: ۱۳۹۶/۰۹/۰۳

نام نویسنده مسئول: حسن فرسی

نشانی نویسنده مسئول: ایران - بیرجند - انتهای بلوار شهیدآوینی - دانشگاه بیرجند - دانشکده مهندسی برق و کامپیوتر.

۱- مقدمه

الگوریتم‌های یادگیری عمیق^۴ موفقیت چشم‌گیری داشته‌اند. زیرا ویژگی‌های مناسبی را می‌توانند از تصویر استخراج کنند [۴]. بنابراین در این مقاله روش جدیدی مبتنی بر یادگیری عمیق به منظور بازیابی تصویر ارائه می‌دهیم. بدین منظور از شبکه‌های عصبی کانولوشن عمیق در بازیابی استفاده کرده و با استخراج ویژگی‌های عمیق تری از تصویر، بازیابی را بهبود می‌بخشیم. در بخش ۲ به مروری بر کارهای انجام شده و پیشینه تحقیق می‌پردازیم. بخش ۳ را به معرفی یادگیری عمیق و شبکه‌های عصبی کانولوشن به‌عنوان زیرمجموعه‌ای از این الگوریتم‌ها اختصاص می‌دهیم. در بخش ۴ مدل پیشنهادی برای بازیابی بر مبنای شبکه عصبی کانولوشن را ارائه می‌دهیم و در بخش ۵ به مقایسه نتایج حاصل از این روش با روش‌های پیشین می‌پردازیم.

۲- پیشینه تحقیق

CBIR به مدت چندین سال زمینه تحقیقاتی مورد علاقه محققان بوده است. برخی محققان برای بازیابی از استخراج ویژگی رنگ تصاویر استفاده کرده‌اند. این شیوه استخراج برای پایگاه داده‌ای که رنگ تصاویر آن یکسان باشد انتخاب مناسبی نخواهد بود. به‌عنوان مثال در [۵] نویسندگان از مقایسه مستقیم بین هیستوگرام‌های رنگ یا کروماتیسی^۵ استفاده کرده‌اند. همان‌طور که ذکر کردیم، استفاده از ویژگی وابسته به رنگ به‌تنهایی مناسب نیست. در [۶] سیستمی معرفی شده که بر اساس ویژگی رنگ و بافت عمل می‌کند. در این سیستم با روش HMMD-HDWT^۶، ابتدا بردار ویژگی همه تصاویر محاسبه می‌شود. بعد از مقایسه بردار ویژگی هر تصویر ورودی با بردارهای ویژگی پایگاه داده، تمامی تصاویر مشابه بازیابی می‌شوند. در این روش برای استخراج ویژگی تصویر، ماتریس هادامارد^۷ و تبدیل موجک گسسته (DWT)^۸ ترکیب شده‌اند. سپس با استفاده از معیار فاصله (فاصله اقلیدسی)، بردارهای ویژگی تصویر مورد سؤال و تصاویر پایگاه داده با هم مقایسه شده و تصاویر مشابه، نمایش داده می‌شوند. در پژوهشی دیگر در سال ۱۳۹۶، محققان روشی به‌منظور بازیابی با چند درخواست با محتوای متفاوت ارائه دادند [۷]. این روش اگرچه دارای این مزیت بود که کاربر می‌توانست هم‌زمان چند درخواست روی یک نمونه تصویر انجام دهد اما بازهم استفاده از ویژگی‌های سطح پایین، مانع رسیدن به دقت ایده‌آل شد. همچنین در سال ۱۳۸۹ نیز برای افزایش دقت بازیابی، روشی ابداع شد که در آن، حین بازیابی از کاربر راجع به تصاویر مطلوب، پرسش به عمل می‌آید تا خطا کاهش یابد [۸]. اما این روش به دلیل توقف حین بازیابی، در بازیابی بلادرنگ مطلوب نمی‌باشد.

با رشد استفاده از تکنیک‌های یادگیری عمیق، در ماه‌های اخیر شاهد استفاده از این الگوریتم‌ها در بازیابی تصاویر بوده‌ایم. در سال ۲۰۱۶ نویسندگان در مقاله [۹] از ترکیب شبکه عصبی کانولوشن عمیق و توابع هش^۹ برای بازیابی تصاویر استفاده کردند. بدین صورت که در این سیستم، شبکه عمیق و تابع Hash به‌طور هم‌زمان آموزش می‌بینند و بردار ویژگی استخراج شده از هر تصویر با طول زیاد، توسط

امروزه اهمیت بازیابی تصویر^۱ بر کسی پوشیده نیست. در پی پیشرفت تکنولوژی تصویری هم‌روزه شاهد افزایش حجم تصاویر ذخیره شده بر روی اینترنت و یا دستگاه‌های حافظه هستیم. هر کاربر روزانه محتوای زیادی از تصاویر و ویدیوها را در فضای مجازی ذخیره و منتشر می‌کند، و یا یک بیمارستان در طول روز صدها تصویر از بیماران گرفته و ذخیره می‌کند. بنابراین نیاز مبرم به سیستمی کارآمد به‌منظور بازیابی تصویر از میان پایگاه داده‌های عظیم تصویر احساس می‌شود. بازیابی به معنی جست‌وجوی تصاویر دیجیتال از بین تعداد عظیمی از تصاویر و نمایش تصاویر مشابه آن است. یکی از چالش‌های پیش رو در سرعت بازیابی، فهرست‌سازی داده‌ها در ابعاد بزرگ است [۱].

بازیابی تصویر برای اولین بار در سال ۱۹۷۰ با معرفی بازیابی تصویر مبتنی بر کلمه معرفی شد به‌نحوی که تصاویر با عبارت‌هایی حاشیه‌نویسی می‌شدند و جست‌وجو بر اساس عبارت وارد شده انجام می‌شد. دو دهه پس از آن بازیابی مبتنی بر محتوا ابداع شد که در آن جست‌وجو بر اساس ویژگی‌های استخراج شده از تصاویر انجام می‌شود. این نوع بازیابی به‌سرعت جایگزین روش قبلی شد و پس از آن در حوزه‌های پزشکی، کتابخانه‌های دیجیتال، تشخیص اثر انگشت، پیشگیری از وقوع جرم (بازیابی چهره مجرمان و یا اثر انگشت آنان)، تحقیقات تنوع زیستی و غیره استفاده می‌شود. این روش‌ها بر پایه استخراج ویژگی سطح پایین از تصاویر استوار است. اما با توجه به شکاف معنایی موجود میان ادراک انسان و این ویژگی‌های سطح پایین، و از طرفی با توجه به افزایش حجم روزافزون تصاویر، نیازمند روش کارآمدی هستیم تا به فرآیند بازیابی سرعت بخشد و فاصله معنایی میان ادراک انسان و ویژگی‌های استخراج شده را کاهش دهد.

دو روش سنتی و معمول در بازیابی تصویر، بازیابی مبتنی بر کلمه^۲ (TBIR) و بازیابی مبتنی بر محتوا^۳ (CBIR) می‌باشد. در TBIR کاربرد عبارت مورد نظر خود مانند نام تصویر، تاریخ ایجاد تصویر و غیره را وارد کرده و سیستم، تصویر را با توجه به اطلاعات ورودی توسط کاربر بازیابی می‌کند [۲]. استفاده از این روش با مشکلاتی روبه‌رو بود، از قبیل این‌که بیان احساسات در قالب کلمه میسر نیست و یا گاه غلط‌های املائی و یا کلماتی که املائی مشابه اما معنی متفاوت دارند، باعث می‌شد تا دقت بازیابی کاهش یابد. برای حل این مشکل، بازیابی مبتنی بر محتوا ابداع شد، به‌نحوی که سیستم تصویر ورودی توسط کاربر را دریافت می‌کند و تصاویری را که از لحاظ محتوا (مانند رنگ، بافت و شکل) مشابه تصویر تقاضا هستند، بازیابی می‌کند. از مزایای CBIR می‌توان به امکان بازیابی اتوماتیک اشاره کرد و این باعث برتری نسبت به روش TBIR که با حاشیه‌نویسی زمان‌بر همراه است، می‌شود [۳]. در این روش از هر تصویر یک بردار ویژگی استخراج می‌شود و مقایسه تصاویر با مقایسه بردارهای ویژگی آن‌ها صورت می‌گیرد.

بررسی تحقیقات سال‌های اخیر، نشان می‌دهد که روش‌های ماشین‌بینایی در این حوزه به‌خوبی عمل کرده و در میان آن‌ها،

بنابراین توجه پژوهش‌گران به روش‌هایی که این فاصله را کم کند معطوف است. یکی از روش‌هایی که در کاهش این فاصله مؤثر بوده است، یادگیری عمیق است. زیرا این روش به‌جای استفاده از ویژگی‌های استخراج شده دستی، به‌صورت خودکار و طی معماری سلسله‌مراتبی، از تصاویر ویژگی‌های مناسب و دقیق را استخراج می‌کند. بنابراین در این مقاله، استفاده از یادگیری عمیق را نسبت به روش‌های مرسوم برگزیدیم. در روش‌های ذکر شده بازیابی توسط یادگیری عمیق در مراجع [۹-۱۲]، به دلیل ترکیب مدل با الگوریتم‌های دیگر، محاسبات پیچیده‌تر شده است. اما مشاهده می‌کنیم که دقت استفاده از یادگیری عمیق، نسبت به سایر روش‌های پیشین، افزایش یافته است.

۲- یادگیری عمیق

یادگیری عمیق یکی از زیرمجموعه‌های یادگیری ماشین است که هدف آن، یادگرفتن چکیده‌ای سطح بالا از داده‌ها با استفاده از معماری‌های سلسله‌مراتبی است و یک رویکرد در حال ظهور است که به‌طور گسترده در حوزه هوش مصنوعی مانند پردازش زبان طبیعی و بینایی ماشینی استفاده شده است [۴].

ایده‌سازی اولیه یادگیری عمیق به سال‌ها قبل بازمی‌گردد. اما عدم وجود سخت‌افزار قوی و مناسب، مانع پیاده‌سازی آن شده است. در چند سال اخیر توانایی پردازش چپ‌ها افزایش یافته درحالی‌که از قیمت آن‌ها کاسته شده است. هم‌چنین پیشرفت‌های چشم‌گیری در الگوریتم‌های یادگیری عمیق به وجود آمده است. بنابراین روزبه‌روز شاهد رونق بیش‌تر استفاده از این الگوریتم‌ها می‌باشیم. شبکه‌های عمیق می‌توانند ویژگی‌های مناسبی از داده‌های ورودی را طی یک روند سلسله‌مراتبی استخراج کنند. بنابراین در چند سال اخیر در حوزه بینایی ماشین موفق عمل کرده‌اند. عمده‌ترین دلیل ترغیب پژوهش‌گران به الگوریتم‌های عمیق در سال‌های اخیر، موفقیت این الگوریتم‌ها در چالش ILSVRC^{۱۱} بوده است. این چالش را شرکت ImageNet هر ساله برگزار می‌کند و در آن، تعداد بسیار زیادی تصویر در اختیار شرکت‌کنندگان قرار می‌گیرد تا الگوریتمی برای طبقه‌بندی دقیق این حجم زیاد تصاویر معرفی کنند و الگوریتمی که روی پایگاه‌داده مربوط به این چالش کم‌ترین خطا را برای طبقه‌بندی و تشخیص تصاویر داشته باشد، به‌عنوان الگوریتم برتر انتخاب می‌شود. الگوریتم‌های یادگیری عمیق به‌شدت توسط محققان در این چالش مورد استفاده قرار گرفته و به نتایج خوبی دست یافته‌اند [۴].

به‌عنوان چهاردهمین مهم‌ترین الگوریتم‌های عمیق، می‌توان از شبکه‌های عصبی کانولوشن^{۱۱}، ماشین بولتزن^{۱۲}، خودرمزگذارها^{۱۳} و کدینگ پراکنده^{۱۴} نام برد. با توجه به مقایسه انجام شده در [۴] استفاده از روش‌های ماشین بولتزن محصور زمان‌بر و نیازمند محاسبات طولانی است. همچنین این روش‌ها و خودرمزگذارها، در مقابل تغییراتی از قبیل چرخش تصویر، مقاوم نمی‌باشند. از سوی

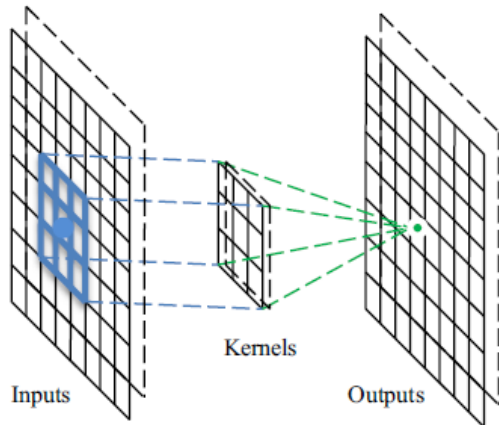
تابع هش، به رشته کد ۰ و ۱ با طول خیلی کم‌تر تبدیل می‌شود. بنابراین مهم‌ترین ویژگی این مقاله، کاهش بار محاسباتی بوده است. بیشترین دقتی که در این مقاله به‌دست آمده، بر روی پایگاه‌داده Oxford با میانگین دقت بازیابی ۰/۷۲ می‌باشد. بنابراین اگرچه بار محاسباتی کاهش و سرعت افزایش یافته، اما به‌دقت مطلوب برای بازیابی دست نیافته است.

در پژوهشی دیگر نیز در سال ۲۰۱۷ [۱۰] از ترکیب شبکه عصبی کانولوشن عمیق و تابع هش استفاده کردند و تفاوت آن با مقاله قبل، استفاده از معماری متفاوت به نام VGGNet با تعداد لایه‌های بیشتر بود. استفاده از این معماری به‌تنهایی به دلیل تعداد لایه‌ها و پارامترهای بیشتر، مستلزم زمان بیشتر برای پردازش است. بنابراین با ترکیب آن با تابع هش، مقداری از این زمان را کاسته و به دلیل عمق بیشتر شبکه، دقت افزایش یافته است. به‌نحوی‌که در بهترین حالت، برای پایگاه‌داده CIFAR-10 به میانگین دقت بازیابی ۰/۸۵۵ دست یافته‌اند.

در همین سال نیز در مقاله [۱۱] نویسندگان از یادگیری عمیق در بازیابی تصاویر پزشکی استفاده کرده‌اند. برای این منظور پایگاه‌داده‌ای مصنوعی از تصاویر پزشکی ایجاد کرده و سپس با استفاده از شبکه عصبی کانولوشن، سیستمی برای بازیابی ارائه دادند، به‌نحوی‌که ابتدا کلاس مربوط به هر تصویر پیش‌بینی شود، سپس بازیابی از داخل تصاویر همان کلاس صورت گیرد. نتیجه اجرای سیستم بر روی پایگاه‌داده مصنوعی، منجر به میانگین دقت بازیابی ۰/۶۹ شده است. مهم‌ترین ایراد این مقاله این است که در صورتی‌که سیستم، در تشخیص کلاس تصویر ورودی دچار خطا شود، همه تصاویر بازیابی شده برای آن تصویر، اشتباه خواهند بود. بنابراین دقت بازیابی در این مقاله مطلوب نیست.

در پژوهشی دیگر در سال جاری، H Liu و همکارانش [۱۲] از ترکیب ویژگی‌های دو مدل از شبکه‌های کانولوشن عمیق، استفاده کردند. در این سیستم مدل AlexNet و مدل بهبودیافته LeNet به‌صورت جداگانه روی پایگاه‌داده، آموزش داده می‌شوند و از ترکیب ویژگی‌های هر دو مدل، برای بازیابی استفاده کرده‌اند. دقت بازیابی به‌دست‌آمده توسط این مدل روی پایگاه‌داده Corel برای تعداد ده تصویر بازیابی شده، برابر ۰/۹۴۸ است. اگرچه این سیستم دقت بالاتری نسبت به روش‌های پیشین به دست آورده است، اما به دلیل نیاز به استفاده از دو شبکه کانولوشن در کنار هم، به زمان بیشتری برای پردازش نیازمند می‌باشد. همچنین این دقت بر روی تنها یک پایگاه‌داده گزارش شده است.

در روش‌هایی که تاکنون مطرح شده، از ویژگی‌های سطح پایین تصاویر استفاده می‌شود. به‌طور دقیق‌تر، در این روش‌ها، ویژگی‌های استفاده شده به‌صورت دستی استخراج می‌شوند. بنابراین بازمی‌آید یکی از معایب مشترک این روش‌ها این است که فاصله معنایی میان ادراک انسان و این ویژگی‌ها مشهود است و دقت بازیابی را پایین می‌آورد.



شکل ۲: نمایش عمل کانولوشن بین تصویر ورودی و فیلتر 3×3 [۴]

بعد از چندین لایه کانولوشن و ادغام، در نهایت ابعاد صفحات ویژگی بسیار کوچک شده اما تعداد آن‌ها زیاد می‌شود. بنابراین در آخر از لایه‌های تماماً متصل استفاده می‌کنیم. این لایه‌ها صفحه ویژگی دو بعدی را به بردار یک بعدی تبدیل می‌کنند.

برای آموزش شبکه کانولوشن دو مرحله پیش رو است: ابتدا مرحله رو به جلو و سپس رو به عقب. در مرحله رو به جلو تصویر از داخل شبکه عبور داده می‌شود و بردار ویژگی آن استخراج می‌شود. با توجه به آگاهی از کلاس مربوط به تصاویر ورودی در مرحله آموزش، یک تابع هزینه تعریف می‌شود به نحوی که میزان مرتبط بودن تصویر به آن کلاس مربوطه را اندازه بگیرد و هر چه این تابع برای یک کلاس کمتر از سایر کلاس‌ها باشد، حاکی از این است که تصویر به آن کلاس مرتبط است. در این شبکه از خطای طبقه‌بند softmax به‌عنوان تابع هزینه استفاده شده و وزن‌های فیلترها را طوری انتخاب می‌کنیم که این تابع کمینه شود به نحوی که طبقه‌بند، کمترین خطا را داشته باشد. به‌منظور انتخاب وزن‌های بهینه، از گرادینان در مرحله بازگشتی استفاده می‌کنیم. گرادینان‌ها با استفاده از قانون بازگشتی در امتداد لایه‌ها محاسبه می‌شود و وزن‌ها و پارامترها با توجه به گرادینان به‌روز می‌شوند. پس از چندین تکرار روی تصاویر ورودی مختلف، مقادیر بهینه فیلترها و پارامترها محاسبه شده و در واقع شبکه برای تشخیص نوع ورودی آموزش دیده است.

۴- روش پیشنهادی

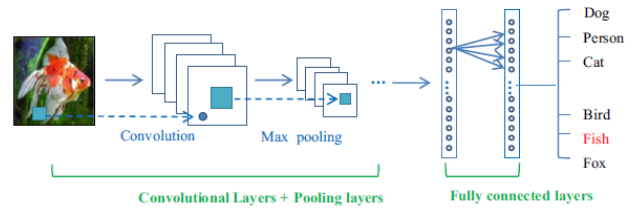
۴-۱- انواع شبکه‌های کانولوشن

اخیراً شبکه‌های قدرتمند کانولوشن به‌منظور طبقه‌بندی تصویر ابداع شده‌اند. در دهه ۹۰ میلادی LeCun و همکارانش [۱۴] برای اولین بار از شبکه کانولوشن برای بازیابی ارقام دست‌نویس استفاده کردند. شبکه آن‌ها LeNet-5 نام داشت. مدل شبکه آن‌ها شامل هفت لایه و تصویر ورودی به شبکه دارای اندازه 32×32 بوده است. در سال ۲۰۱۲ Krizhevsky و همکارانش [۱۵] شبکه کانولوشن عمیقی ایجاد کردند که برنده چالش ImageNet شد. اگر به توازن دقت و سرعت بنگریم،

دیگر، در کدینگ پراکنده، امکان آموزش ویژگی وجود ندارد. از میان این روش‌ها ما شبکه عصبی کانولوشن را برای ادامه کار برگزیدیم. زیرا به‌طور خاص برای داده‌های دو بعدی مانند تصویر کاربرد دارد و نشان داده شده است که در مقابل تغییرات تصویر مقاوم بوده و علاوه بر طبقه‌بندی تصاویر، قابل تعمیم به سایر کاربردها نیز می‌باشد.

۳-۱- شبکه عصبی کانولوشن عمیق^{۱۵}

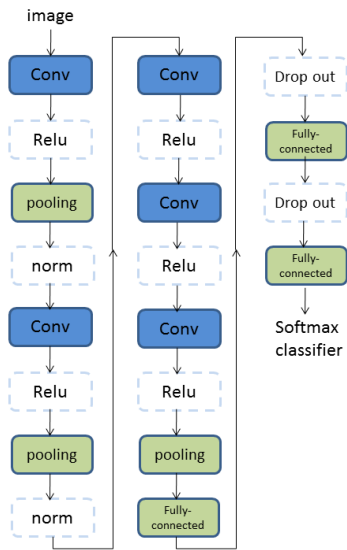
شبکه عصبی کانولوشن (CNN) یکی از مهم‌ترین روش‌های یادگیری عمیق است که لایه‌های متعدد در آن، به شیوه‌ای جدید و مستحکم آموزش می‌بینند. این روش، مؤثر واقع شده و به‌صورت رایج در کاربردهای بینایی ماشین استفاده می‌شود. این شبکه‌ها زیرمجموعه‌ای از شبکه‌های عصبی چندلایه هستند که برای داده‌های دوبعدی مانند تصویر، طراحی شده‌اند. بخش‌های مختلف تصویر به‌عنوان ورودی به لایه‌های سلسله‌مراتبی داده می‌شوند و در هر لایه با اعمال فیلترهایی دیجیتال، ویژگی‌های چشم‌گیری از تصویر استخراج می‌شود [۱۱]. به‌طور کلی CNN از سه نوع لایه اصلی کانولوشن، ادغام^{۱۶} و لایه تماماً متصل تشکیل می‌شود^{۱۷}. در شکل ۱ شمای کلی یک شبکه عصبی کانولوشن را مشاهده می‌کنید.



شکل ۱: شمای کلی معماری یک شبکه عصبی کانولوشن [۴]

در هر معماری CNN چندین لایه کانولوشن وجود دارد. در این لایه‌ها تصویر ورودی با فیلترهایی که ضرایب آن‌ها قابل آموزش است، کانولوشن می‌شود. این فیلترها روی تصویر حرکت داده می‌شوند. عمق فیلتر با عمق تصویر برابر است. به ازای هر فیلتر مجزا، یک صفحه ویژگی ایجاد می‌شود. اگر از n فیلتر استفاده کنیم، n صفحه ویژگی به وجود می‌آید. وزن‌های هر فیلتر که w نامیده می‌شوند، قابل آموزش هستند و در طول آموزش شبکه، مرتباً به‌روز می‌شوند. بعد از عمل کانولوشن، حاصل با یک عدد بایاس جمع شده و در صفحه ویژگی ذخیره می‌شود. به‌عنوان مثال اندازه فیلترها می‌تواند 3×3 باشد. در شکل ۲ لایه کانولوشن را مشاهده می‌کنید.

بعد از این لایه، معمولاً لایه ادغام می‌آید. برای ادغام کردن چندین استراتژی وجود دارد که روش معمول آن، استفاده از ادغام ماکزیمم است. در این روش پنجره‌هایی معمولاً با اندازه 4×4 انتخاب می‌شوند و روی تصویر حرکت داده می‌شوند و از هر چهار پیکسل، مقدار ماکزیمم آن‌ها انتخاب شده و به لایه بعدی انتقال داده می‌شود. با این کار، تعداد صفحات ویژگی ثابت مانده اما اندازه آن کاهش پیدا کرده است.



شکل ۳: لایه‌های اصلی معماری استفاده شده

هر تصویر ابتدا با تعدادی فیلتر با اندازه 11×11 با مقداردهی اولیه گوسی کانولوشن می‌شود به نحوی که فیلتر روی همه قسمت‌های تصویر حرکت داده می‌شود و حاصل کانولوشن، در صفحات ویژگی ذخیره می‌شود. پس از اعمال تابع فعال‌ساز به این صفحات، هر کدام از آن‌ها وارد مرحله ادغام ماکزیمم می‌شوند. به نحوی که پنجره‌هایی با اندازه 4×4 روی تصویر حرکت داده می‌شوند و از میان هر ۴ پیکسل، ماکزیمم آن‌ها انتخاب می‌شود. بنابراین پس از این لایه، اندازه صفحات ویژگی کاهش می‌یابد. در مرحله بعد، دوباره هر کدام از صفحات ویژگی با فیلترها کانولوشن شده و روی این صفحات، ادغام انجام می‌شود. در مجموع پس از پنج لایه کانولوشن و سه لایه ادغام، سه لایه تماماً متصل وجود دارد که این ویژگی‌های دوبعدی را تبدیل به بردار یک بعدی می‌کنند. پس از اینکه کاربر تصویر تقاضا را ارائه کند، تصویر از شبکه عبور داده شده و بردار خروجی لایه آخر، با ماتریس ویژگی مقایسه شده به نحوی که فاصله اقلیدسی این بردار با تک‌تک بردارهای موجود در ماتریس ویژگی محاسبه شده و نزدیک‌ترین تصاویر به تصویر تقاضا، که بردار ویژگی آن‌ها فاصله اقلیدسی کمتری تا بردار ویژگی تصویر تقاضا دارند، به عنوان تصاویر بازیابی شده انتخاب می‌شوند. در شکل ۴ بلوک دیاگرام روش پیشنهادی را مشاهده می‌کنید.

در این شکل دو بخش کلی آموزش آفلاین و بازیابی آنلاین، هم‌زمان نمایش داده شده است. به نحوی که قسمت بالایی شکل، مربوط به مرحله آموزش می‌باشد و قسمت پایینی، مرحله بازیابی را به تصویر می‌کشد. بدیهی است که در مرحله اجرا، لایه‌های dropout و norm که منحصر به مرحله آموزش بوده‌اند، به کار گرفته نخواهند شد.

در این نوع سیستم بازیابی پس از چندلایه کانولوشن، ویژگی‌های عمیق و دقیق‌تری استخراج شده و باعث می‌شود تا این سیستم از دقت بالاتری نسبت به سایر روش‌ها که ویژگی‌های سطح پایین را استخراج می‌کنند، برخوردار شود. در سیستم پیشنهادی، از لایه آخر شبکه، ویژگی‌ها را استخراج کرده و به عنوان بردار ویژگی هر تصویر در نظر

می‌توانیم مدل AlexNet که زیربنای همه الگوریتم‌های دیگر است را به عنوان مبنای کار خود و برای استفاده در سیستم خود انتخاب کنیم که علاوه بر دقت مناسب، از سرعت قابل قبولی برخوردار است.

برای استفاده از شبکه‌های عصبی عمیق، سه راه پیش رو داریم [۱۶]. اول این که می‌توانیم یک شبکه را از اول تعریف کنیم و روی پایگاه داده مورد تحقیق، آموزش دهیم. این روش در صورتی نتیجه مطلوب می‌دهد که پایگاه داده تصاویر آموزشی شامل تعداد عظیمی از تصاویر باشد. بدیهی است به دلیل این که تعداد پارامترهایی که نیاز به آموزش دارند در شبکه‌های عمیق زیاد است، هر چه تعداد تصاویر ورودی بیشتر باشد، آموزش بهتر انجام می‌شود. روش دوم، استفاده از مدل از قبل آموزش داده شده است. این کار در صورتی بهینه است که تفاوت پایگاه داده مورد تحقیق و داده اولیه که مدل تحت آن آموزش دیده است، اندک باشد. و اما روش سوم فاین تیون^{۱۸} کردن نام دارد و به معنی استفاده از وزن‌های از پیش آموزش دیده شده برای تعداد زیادی از لایه‌ها و آموزش تعداد کمتری از لایه‌ها بر روی پایگاه داده مورد نظر است. ما برای ادامه کار، روش سوم را برگزیدیم. زیرا در این صورت نیازمند آموزش تعداد کمتری از لایه‌ها می‌باشیم و آموزش سریع‌تر انجام می‌شود. ضمن این که بعضی از پایگاه‌های داده، برای آموزش مدل از ابتدا، دارای تعداد تصویر کافی نیستند. در شکل ۳ لایه‌های این معماری را مشاهده می‌کنید. در این معماری علاوه بر لایه‌های اصلی، لایه‌های دیگری نیز به چشم می‌خورند. بعد از لایه کانولوشن که عملیات آن خطی است، تابع فعال‌ساز ReLU به منظور غیرخطی کردن سیستم قرار می‌گیرد. بعد از اولین و دومین لایه ادغام و به منظور نرمال کردن داده‌های خروجی از آنان، از لایه norm استفاده می‌شود. این کار باعث صفر شدن میانگین و واحد شدن انحراف از معیار شده و مزیت آن، کاهش وابستگی سیستم به مقداردهی اولیه ضرایب است. از آن جا که تعداد زیادی از پارامترها در سه لایه آخر تماماً متصل قرار دارند، در حین آموزش احتمال عدم همگرایی زیادتر می‌شود. بنابراین، در میان لایه‌های تماماً متصل، از لایه drop out برای کاهش پارامترها استفاده شده است. عملکرد این لایه بدین صورت است که در هر تکرار از آموزش، تعدادی از نرون‌ها را به صورت تصادفی از شبکه حذف می‌کند و در مرحله بعد، با وزن قبلی به شبکه بازگردانده می‌شوند. این کار باعث افزایش سرعت و همگرایی بهتر می‌شود.

۴-۲- نحوه عملکرد الگوریتم پیشنهادی

در این مقاله، هدف ما ارائه مدلی است که بتوانیم از شبکه‌های عصبی کانولوشن عمیق به منظور بازیابی استفاده کنیم. به منظور تحقق بازیابی دقیق، مدل پیشنهادی بدین صورت است که ابتدا مدل را بر روی پایگاه داده مورد تحقیق، تنظیم کنیم. برای این منظور وزن‌های لایه‌ها به غیر از دولایه آخر ثابت نگه داشته می‌شوند و دو لایه آخر آموزش می‌بینند. پس از آموزش، همه تصاویر را از این شبکه عمیق عبور می‌دهیم.

$$K(q) = \min [4 \cdot |NG(q)|, 2 \cdot \max \{|NG(q)|, \forall q\}]$$

▪ $R(k)$ رتبه تصویر مورد سؤال k

▪ $Rank(k)$ به صورت زیر محاسبه می شود:

$$Rank(k) = \begin{cases} R(k) & \text{if } R(k) \leq K(q) \\ 1.25K & \text{otherwise} \end{cases} \quad (1)$$

با استفاده از رابطه بالا میانگین رتبه یعنی $AVR(q)$ برای تصویر مورد سؤال (q) به صورت زیر محاسبه می شود:

$$AVR(q) = \langle Rank(k) \rangle. \quad (2)$$

رتبه بازیابی اصلاح شده به صورت زیر به دست می آید:

$$MRR(q) = AVR(q) - 0.5 [1 + |NG(q)|]. \quad (3)$$

سپس $NMRR$ با استفاده از رابطه زیر محاسبه می شود:

$$NMRR(q) = \frac{AVR(q) - 0.5 [1 + |NG(q)|]}{1.25K(q) - 0.5 [1 + |NG(q)|]}. \quad (4)$$

در این معیار صفر به یک بازیابی کامل و ۱ به یک بازیابی پراشتباه دلالت می کند و $ANMRR$ با میانگین گیری از $NMRR$ های همه تصاویر مورد سؤال موجود در پایگاه های داده به دست می آید.

$$ANMRR = \langle NMRR(q) \rangle. \quad (5)$$

۵-۲- معیارهای دقت و فراخوانی

معیارهای دقت (P) و فراخوانی (R)، از مهم ترین معیارهای ارزیابی سیستم های بازیابی هستند [۱۷]. معیار P نشان گر دقت بازیابی است و معیار R قابلیت بازیابی موردهای مرتبط از پایگاه داده را اندازه گیری می کند. معیارهای دقت و فراخوانی معکوس هم عمل می کنند یعنی به تناسب اینکه P افزایش می یابد، R کاهش می یابد. معیارهای دقت و فراخوانی از روابط زیر به دست می آیند:

$$Precision = \frac{\text{تعداد تصاویر بازیابی شده مرتبط}}{\text{تعداد تصاویر بازیابی شده}} \quad (6)$$

$$Recal = \frac{\text{تعداد تصاویر بازیابی شده مرتبط}}{\text{تعداد کل تصاویر مرتبط پایگاه داده}} \quad (7)$$

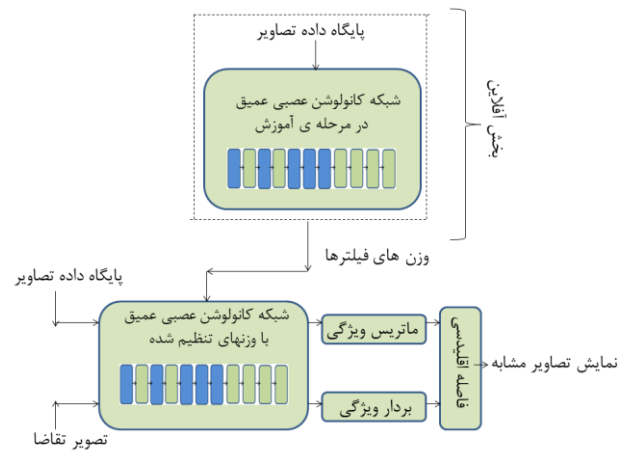
ما با ترکیب معیارهای دقت و فراخوانی، معیارهای بهتری پیشنهاد دادیم تا به جای استفاده جداگانه از معیارهای P و R از هر دو معیار هم زمان استفاده کنیم [۶]. این معیارها عبارت اند از:

▪ $P(0.5)$ دقت در ۵۰٪ فراخوانی (یعنی دقت پس از بازیابی نصف تصاویر مربوطه).

▪ $P(1)$ دقت در ۱۰۰٪ فراخوانی (یعنی دقت پس از بازیابی

کل تصاویر مربوطه که همان نقطه برخورد نمودارهای دقت و فراخوانی است).

می گیریم. انتخاب ویژگی از آخرین لایه (softmax)، دارای دو مزیت است. اول این که به دلیل اینکه خروجی این لایه برابر با تعداد کلاس های هر پایگاه داده است، تعداد ویژگی ها در پایگاه داده های معمول، کم است و نیازمند استفاده از روش های ادغام و هش مانند مقالات [۱۰ و ۱۲]، نمی باشیم. به عنوان دومین مزیت، می توان به این نکته اشاره کرد که در روش پیشنهادی ویژگی ها از بعد از لایه طبقه بند استخراج می شوند، بنابراین اطلاعات مربوط به کلاس هر تصویر، در بردار ویژگی وجود خواهد داشت و این خود باعث کاهش خطا و افزایش دقت می گردد.



شکل ۴: بلوک دیاگرام الگوریتم پیشنهادی

۵- معیارهای ارزیابی

به منظور ارزیابی سیستم، سه معیار $P(0.5)$ و $P(1)$ و $ANMRR$ را که در ادامه توضیح داده شده اند، برای سه پایگاه داده Corel، Aloj و MPEG7 محاسبه و نتایج را با روش های پیشین بازیابی، مقایسه می کنیم و نشان می دهیم که دقت بازیابی در مدل پیشنهادی نسبت به سایر روش ها افزایش می یابد.

برای ارزیابی سیستم های بازیابی، معیار ارزیابی مؤثر و مفیدی لازم است. در این راستا معیارهای ارزیابی مختلفی پیشنهاد شده است. در این بخش به شرح بعضی از معیارهای ارزیابی مشهور که در مقالات برای مقایسه سیستم های بازیابی استفاده شده است، می پردازیم. برای ارزیابی سیستم های بازیابی معیارهای سنجش بسیاری وجود دارند که ما از سه معیار $ANMRR$ ، دقت و فراخوانی استفاده می کنیم.

۵-۱- معیار $ANMRR$

$ANMRR$ (میانگین نرمال شده و اصلاح شده رتبه بازیابی) یک معیار ارزیابی از عملکرد سیستم های بازیابی می باشد که کارایی سیستم را در یک مقدار عددی خلاصه می کند [۱۷]. ابتدا $NG(q)$ ، $K(q)$ و $R(k)$ را به صورت زیر تعریف می کنیم:

▪ $NG(q)$: تعداد تصاویر درست برای تصویر مورد سؤال (q)

۶ - نتایج اجرای برنامه

شبیه‌سازی با استفاده از نرم‌افزارهای Caffé [۱۸] که یک چهارچوب مناسب برای پیاده‌سازی یادگیری عمیق است و Matlab، بر روی سیستم دارای پردازنده Intel Core i5 1.7 GHz و حافظه 8 GB RAM و سیستم‌عامل Ubuntu 14.04 انجام شده است.

در جدول ۱ نتایج حاصل از شبیه‌سازی و محاسبه معیارها را برای مدل پیشنهادی و نتایج استخراج شده از سایر مقالات در مدل‌های HMMMD-HDWT [۶]، CLD، DCD، SCD، pp^۲، HI و [۵] و هم‌چنین روش الگوریتم تنک (sparse-DALM) [۱۹] روی پایگاه‌داده ALOI [۲۲] که شامل ۴۶۰۸ تصویر در ۶۴ کلاس می‌باشد، مشاهده می‌کنید. از میان روش‌های تنک، بیش‌ترین مقدار دقت به‌دست‌آمده در [۱۹] را انتخاب و برای مقایسه در جدول قرار داده‌ایم. در میان سایر روش‌ها، بیش‌ترین مقدار P(0.5) برابر ۰.۸۲/۲۹٪ و مربوط به روش HMMMD-HDWT می‌باشد. در حالی که این مقدار در الگوریتم پیشنهادی به ۰.۹۹/۰۹٪ افزایش یافته است.

منظور از اندازه بردار ویژگی، تعداد اعضای برداری است که به‌عنوان ویژگی‌های هر تصویر در هر کدام از روش‌ها، استخراج می‌شود و برای مقایسه تصاویر، از آن استفاده می‌شود. در جدول ۲ نتایج آزمایش بر روی پایگاه‌داده Corel [۲۱] که شامل هزار تصویر در ۱۰ کلاس است، آورده شده است. در این پایگاه‌داده نیز از میان سایر روش‌ها، بیش‌ترین مقدار P(0.5) مربوط به الگوریتم AL [۱۲] و برابر ۰.۸۷/۶۹٪ بوده که در آن از ترکیب دو الگوریتم عمیق استفاده شده است. اما در روش پیشنهادی این معیار به ۰.۹۶/۳۰٪ بهبود یافته است. همچنین نتایج بر روی پایگاه‌داده Mpeg7 [۲۲] در جدول ۳ قابل مشاهده است. در این جدول نیز مشاهده می‌کنیم که بیش‌ترین مقدار P(0.5) برابر ۰.۸۸/۳۴٪ در روش HI با بردار ویژگی به طول ۱۰۲۴ و بیش‌ترین مقدار P(1) برابر ۰.۵۸/۳۱٪ مربوط به روش تنک می‌باشد. در حالی که در روش پیشنهادی مقدار P(0.5) برابر ۰.۸۶/۰۷٪ و مقدار P(1) برابر ۰.۷۸/۵۷٪ و با طول بردار ویژگی بسیار کم‌تر از روش‌های دیگر شده است.

جدول ۱: پایگاه‌داده ALOI

روش	اندازه بردار ویژگی	P(0.5)	P(1)	ANMRR
Sparse-DALM	۶۴	٪۷۹/۹۹	٪۷۹/۸۹	۰/۲۰۰۵
HMMMD-HDWT	۳×۶۴	٪۸۲/۲۹	٪۶۶/۶۸	۰/۲۵۵۱
CLD	۱۲	٪۱۹/۴۹	٪۱۷/۵۶	۰/۸۵۲۶
DCD	۳۲	٪۵۰/۲۱	٪۴۱/۸۷	۰/۶۳۵۴
SCD	۱۱×۱۲۱	٪۵۲/۲۱	٪۴۳/۵۵	۰/۵۲۱۵
PP	۴۹۶	٪۳۷/۷۹	٪۷/۷۱	۰/۹۰۲۳
HI	۱۰۲۴	٪۳۶/۲۶	٪۶/۷۷	۰/۹۱۸۶
روش پیشنهادی	۶۴	٪۹۹/۰۹	٪۹۸/۳۴	۰/۰۱۰۱

جدول ۲: پایگاه‌داده Corel

روش	اندازه بردار ویژگی	P(0.5)	P(1)	ANMRR
Sparse-DALM	۶۴	٪۶۹/۹۸	٪۶۹/۴۶	۰/۲۰۲۴
HMMMD-HDWT	۳×۶۴	٪۴۱/۴۵	٪۳۵/۴۴	۰/۵۶۱۵
CLD	۱۲	٪۵۷/۷۹	٪۴۱/۸۲	۰/۵۳۹۱
DCD	۳۲	٪۴۸/۶۴	٪۳۶/۳۷	۰/۵۸۲۶
SCD	۱۱×۱۲۱	٪۴۳/۶۳	٪۳۳/۶۵	۰/۶۱۰۶
PP	۴۹۶	٪۳۸/۵۶	٪۲۸/۶۵	۰/۶۵۳۱
HI	۱۰۲۴	٪۴۰/۲۱	٪۲۹/۹۲	۰/۶۳۲۴
AL	۱۰۸۶	٪۸۷/۶۹	---	---
روش پیشنهادی	۱۰	٪۹۶/۳۰	٪۹۳/۰۰	۰/۰۳۸۷

از مقایسه معیارهای P(0.5)، P(1) و ANMRR به‌دست‌آمده در روش پیشنهادی با سایر روش‌ها، به‌دقت بالای این الگوریتم مبتنی بر یادگیری عمیق پی می‌بریم. زیرا در روش‌های دیگر مهم‌ترین عامل کاهش دقت، استخراج ویژگی‌های سطح پایین تصاویر و عدم کاهش فاصله معنایی میان ادراک انسان و این ویژگی‌ها است. در حالی که یادگیری عمیق، با استخراج سلسله‌مراتبی و عمیق ویژگی از تصاویر، ویژگی‌های سطح بالاتری را استخراج کرده و الگوریتم پیشنهادی ما که در آن از شبکه‌های کانولوشن عمیق به‌منظور بازیابی استفاده شده، به‌دقت مطلوبی دست یافته است.

جدول ۳: پایگاه‌داده MPEG7

روش	اندازه بردار ویژگی	P(0.5)	P(1)	ANMRR
Sparse-DALM	۶۴	٪۵۹/۴۱	٪۵۸/۳۱	۰/۴۲۵۰
HDWT	۳×۶۴	٪۷۰/۴۱	٪۵۳/۹۰	۰/۴۲۸۵
PP	۴۹۶	٪۷۷/۷۷	٪۳۷/۲۱	۰/۶۸۲۱
HI	۱۰۲۴	٪۸۸/۳۴	٪۵۷/۵۶	۰/۴۴۳۲
روش پیشنهادی	۷۲	٪۸۶/۰۷	٪۷۸/۵۷	۰/۱۷۳۵

معیار ANMRR که صفر بودن آن به معنی بازیابی کامل است، در این الگوریتم برای پایگاه‌داده ALOI دارای مقدار ۰/۰۱، برای پایگاه‌داده Corel دارای مقدار ۰/۰۳۸۷ و برای تصاویر MPEG7 دارای مقدار ۰/۱۷۳۵ می‌باشد که نشان از یک بازیابی مطلوب و کم‌نقص دارد.

برتری دیگر این مدل نسبت به سایر روش‌ها، استفاده از محاسبات غیر پیچیده است. در سایر الگوریتم‌ها استخراج ویژگی‌های بهتر، مستلزم استفاده از تبدیل‌ها و روش‌های بسیار پیچیده‌تر است که باعث کند شدن روند بازیابی می‌شود. اما عمل کانولوشن که مبنای اصلی الگوریتم پیشنهادی و ترکیبی از عمل ضرب و جمع است، به‌مراتب

[۸] اسما شمسی گوشکی، سعید سریزدی، حسین نظام آبادی پور، محمد شهرام معین، «روشی جدید در بازخورد ربط برای بازیابی تصویر بر اساس محتوا به شیوه چند پرسشی»، مجله مهندسی برق دانشگاه تبریز، دوره ۴۰، شماره ۲، صفحه ۵۱-۶۲، ۱۳۸۹.

- [9] D. Varga and T. Szirányi, "Fast content-based image retrieval using convolutional neural network and hash function," IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 2636-2640, 2016.
- [10] T. q. Peng and F. Li, "Image retrieval based on deep Convolutional Neural Networks and binary hashing learning," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1742-1746, 2017
- [11] A. Qayyum, S.M. Anwar, M. Awais and M. Majid, "Medical image retrieval using deep convolutional neural network," Neurocomputing, vol. 266, pp. 8-20, 2017.
- [12] H. Liu, B. Li, X. Lv and Y. Huang, "Image retrieval using fused deep convolutional features," Procedia Computer Science, vol. 107, pp. 749-754, 2017
- [13] I. Arel, D. C. Rose and T. P. Karnowski, "Deep machine learning - A new frontier in artificial intelligence research [Research Frontier]," IEEE Computational Intelligence Magazine, vol. 5, no. 4, pp. 13-18, 2010.
- [14] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.
- [15] A. Krizhevsky, I. Sutskever and G.E. Hinton, "ImageNet classification with deep convolutional neural networks," Advances in Neural Information Processing Systems 25, 2012.
- [16] H. C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Noguez, et al., "Deep convolutional neural networks for Computer-Aided detection: CNN Architectures, Dataset Characteristics and Transfer Learning," IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1285-1298, 2016.
- [17] Y. D. Chun, N. C. Kim, and I. H. Jang, "Content-Based image retrieval using multiresolution color and texture features," IEEE Transactions on Multimedia, vol. 10, no. 6, pp. 1073-1084, 2008.
- [18] <http://caffe.berkeleyvision.org/>
- [19] S. Mohamadzadeh and H. Farsi, "Content-based image retrieval system via sparse representation," IET Computer Vision, vol. 10, no. 1, pp. 95-102, 2016.
- [20] J. M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders, "The Amsterdam library of object images," International Journal Computer Vision," vol. 61, no. 1, pp. 103-112, 2005.
- [21] Corel Database <http://wang.ist.psu.edu/docs/related/> (last referred on June, 10th, 2009)
- [22] I. O. f. Standardisation., "MPEG-7 overview," Available at: <http://mpeg.chiariglione.org/standards/mpeg-7/mpeg-7.htm>, accessed 15 November 2011.

ساده تر از محاسبات سایر روش‌ها مانند استفاده از تبدیل فوری، تبدیل موجک و غیره می‌باشد. بنابراین سرعت بازیابی در این مدل نیز افزایش می‌یابد که از دیگر مزیت‌های استفاده از این الگوریتم است.

۷- نتیجه

بازیابی تصویر یکی از مهم‌ترین شاخه‌های بینایی ماشین است که از سال‌ها قبل مورد توجه و استفاده محققان بوده است. اما پیش از این الگوریتم‌های موجود در بازیابی نتوانستند شکاف معنایی بین ادراک انسان و ویژگی‌های استخراج شده را کم کنند. بنابراین دقت بازیابی کم بوده و از طرفی با توجه به حجم زیاد تصاویر ذخیره شده، سرعت این الگوریتم‌ها نیز کم بوده است. در این مقاله روشی مبتنی بر یادگیری عمیق با استفاده از شبکه‌های عصبی کانولوشن بنا نهادیم و پس از شبیه‌سازی مدل پیشنهادی و تست آن بر روی سه پایگاه داده ALOI، COREL و MPEG7 و محاسبه سه معیار مهم بازیابی، نشان دادیم که این مدل به دلیل استخراج عمیق و سلسله مراتبی ویژگی‌های تصاویر، نسبت به سایر روش‌ها از دقت بالاتری برخوردار است. همچنین استفاده از عمل کانولوشن به جای استفاده از سایر تبدیل‌های مرسوم در روش‌های دیگر، پیچیدگی محاسبات را کم کرده و به عمل بازیابی سرعت بخشیده است.

مراجع

- [1] A. N. Tikle, C. Vaidya, and P. Dahiwal, "A survey of indexing techniques for large scale Content-Based image retrieval," in 2015 International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO), pp. 1-5, 2015.
- [2] J. A. Silva Júnior, R. E. Marçal and M. A. Batista, "Image retrieval importance and applications," Workshop de Visao Computacional - WVC 2014, 2014.
- [3] R. D. S. Torres and A. X. Falcao, "Content-Based image retrieval theory and applications," RITA, vol. 13, no. 2, pp. 161-185, 2006.
- [4] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu and M. S. Lew, "Deep learning for visual understanding: A review," Neurocomputing, vol. 187, pp. 27-48, 2016.
- [5] R. Montagna and G.D. Finlayson, "Padua point interpolation and Lp-norm minimisation in colour-based image indexing and retrieval," IET Image Processing, vol. 6, no. 2, pp. 139-147, 2012.
- [6] H. Farsi and S. Mohamadzadeh, "Colour and texture feature-based image retrieval by using hadamard matrix in discrete wavelet transform," IET Image Processing, vol. 7, no. 3, pp. 212-218, 2013.
- [۷] مریم تقی‌زاده، عبدالله چاله چاله، «مدلی به منظور بازیابی تصویر مبتنی بر چند درخواست»، مجله مهندسی برق دانشگاه تبریز، مقالات آماده انتشار، پذیرفته شده، انتشار آنلاین از تاریخ ۱۲ فروردین ۱۳۹۶

زیرنویس‌ها

-
- ¹ Image retrieval
 - ² Text-Based Image Retrieval
 - ³ Content-Based Image Retrieval
 - ⁴ Deep learning
 - ⁵ Chromaticity
 - ⁶ Hue Maximum Minimum Difference-Hadamard Discrete Wavelet Transform
 - ⁷ Hadamard matrix
 - ⁸ Discrete Wavelet Transform
 - ⁹ Hash Functions
 - ¹⁰ ImageNet Large Scale Visual Recognition Challenge
 - ¹¹ Convolutional Neural Networks
 - ¹² Restricted Boltzmann Machines
 - ¹³ Autoencoder
 - ¹⁴ Sparse Coding
 - ¹⁵ Deep convolutional neural networks
 - ¹⁶ Pooling
 - ¹⁷ Fully-connected
 - ¹⁸ Fine tune
 - ¹⁹ Average Normalized Modified Retrieval Rank
 - ²⁰ Padua point