بازیابی تصاویر پوشاک با استفاده از شبکه های کپسولی چهارگانه

مهدیه سادات خاتمی[۱]، دکتر محمد جواد فدائی اسلام[۲]

دانشجوی کارشناسی ارشد هوش مصنوعی دانشگاه سمنان

mahdiye_khatami@semnan.ac.ir

چکیده

هدف از بازیابی تصاویر پوشاک یافتن تصاویر مشابه با تصویر مورد جستجو در خرید های اینترنتی می باشد. یکی از چالش های موجود، بازیابی تصاویر بدون حساسیت به زاویه دید و بدون استفاده از اطلاعات جانبی تصاویر میباشد. در مطالعات پیشین معماری RCCapsNet در راستای جایگزین نمودن شبکه های کپسولی با شبکه های عصبی کانولوشنی و حل مشکلات ذکر شده ارائه شده است که از تابع زیان سه گانه جهت رسیدن به این امر بهره برده است. این پژوهش با هدف ارتقاء معماری RCCapsNet انجام شده است. جهت رسیدن به این امر نسخه چهار گانه معماری RCCapsNet با بهره گیری از تابع زیان چهار گانه تحت عنوان QRCCapsNet پیشنهاد شده است. روش پیشنهادی روی مجموعه داده DeepFashion-Inshop آموزش داده شده است و عملکرد بر اساس معیار ارزیابی Recall@K ارزیابی شده است. نتایج پیاده سازی حاکی از موفقیت روش پیشنهادی می باشد. سورس کد در گیت هاب منتشر شده است.[۳]

کلمات کلیدی: بازیابی تصاویر پوشاک، بازیابی پوشاک درون فروشگاهی، شبکه های کپسولی، تابع زیان چهار گانه

[۱] دانشجوی کارشناسی ارشد، رشته مهندسی کامپیوتر، گرایش هوش مصنوعی و رباتیکز، دانشگاه سمنان

[۲] استادیار گروه آموزشی مهندسی نرم افزار کامپیوتر، دانشکده مهندسی برق و کامپیوتر، دانشگاه سمنان

[3] https://github.com/MahdiyeKhatami/QRCCapsNet

۱- مقدمه

بازیابی تصاویر پوشاک یکی از چالش برانگیزترین امور در حوزه صنعت مد و پوشاک می باشد. هدف از این امر یافتن تصاویر مشابه با تصویر مورد جستجو در خرید های اینترنتی می باشد. این امر نقش مهمی در زمینه نیاز رو به رشد تجارت الکترونیک و توصیه‌های مبتنی بر وب دارد. یکی از چالش های موجود، تشخصیص تصویر با توجه به تغییرات زاویه دید می باشد. مطالعات متعددی جهت به کارگیری شبکه های عصبی کانولوشنی برای حل این مسئله صورت گرفته است. با این وجود، شبکه های کانولوشنی دارای محدودیت هایی هستند مانند از دست رفتن اطلاعات فضایی اشیاء و عدم مقاومت نسبت به تبدیلات آفین. این محدودیت ها سبب میشود تا عملیات بازیابی تصاویر وابسته به مجموعه داده غنی به همراه اطلاعات جانبی تصاویر باشد. جهت رفع این مشکلات در سال ۲۰۱۹، طرحی سه گانه بر پایه شبکه های کپسولی تحت عنوان RCCapsNet[1] ارائه شده است. در این معماری از تابع زیان سه گانه استفاده شده است. این پژوهش با هدف ارتقاء معماری RCCapsNet انجام شده است. جهت رسیدن به این امر نسخه چهار گانه معماری RCCapsNet با بهره گیری از تابع زیان چهار گانه تحت عنوان QRCCapsNet پیشنهاد شده است.

۱-۱- مرور کارهای گذشته

در سال ۲۰۱۷ جهت رفع مشکلات موجود در شبکه های عصبی کانولوشنی همچون از دست رفتن اطلاعات فضایی اشیاء و عدم مقاومت نسبت به تبدیلات آفین، طرحی جدید به نام شبکه های کپسولی توسط صبور، هینتون و همکاران[۴] [2] پیشنهاد شده است. در این طرح، با کمک الگوریتم مسیریابی پویا، می توان به اطلاعات توصیفی بیشتری در مورد اشیاء، بدون از دست رفتن ارتباط فضایی بین اشیاء و اجزای آن دست یافت. به این ترتیب، شبکه های کپسولی از این امکان بهره مند هستند که بدون حساسیت به زاویه دید، تصاویر را تشخیص دهند چرا که در ذات خود توانایی فراگیری پیکربندی بالاتری از تصاویر را دارا هستند. لالوند و همکاران[۵] [3] در سال ۲۰۱۸، نشان می‌دهند که شبکه‌های کپسولی با تغییر معماری و الگوریتم مسیریابی پویا می‌توانند با پارامترهای کمتری نسبت به معماری اولیه، بر روی تصاویر بزرگ کار کند. ژانگ و همکاران[۶] [4] در سال ۲۰۱۸، دو متد مسیریابی سریع پیشنهاد میکنند که به کمک این متدها شبکه های کپسولی میتوانند با بهره زمانی بهتر و با تصاویر بزرگ‌تری کار نمایند. جایسال و همکاران[۷] [5] در سال ۲۰۱۸، چارچوبی ارائه می‌کند که از معماری شبکه کپسولی در شبکه‌های مولد تخاصمی (GAN)[۸] استفاده می‌کند. کوشیرک و همکاران[۹] [6] در سال ۲۰۱۹، نسخه بدون نظارت شبکه‌های کپسولی را ارائه می‌کنند که در آن یک رمزگذار، تصویر را به بخش‌های تشکیل‌دهنده تقسیم می‌کند و یک رمزگشا پیکربندی شی را پیش بینی می‌کند. گو و همکاران[۱۰] [7] در سال ۲۰۲۰، معماری Aff-CapsNets را جهت بهبود مقاومت شبکه های کپسولی نسبت تبدیلات آفین ارائه نموده اند. وانگ و همکاران[۱۱] [8] در سال ۲۰۲۰، سعی دارند تا از شبکه های کپسولی برای اولین بار درتوسعه مدل‌های طبقه‌بندی مربوط به دارو استفاده نمایند. الدیفراوی و همکاران[۱۲] [9] در سال ۲۰۲۱، معماری جدیدی با عنوان Deep-FECapsNet را معرفی مینمایند که از مسیریابی پویا مبتنی بر کانولوشن یک بعدی با فرآیند ضرب سریع

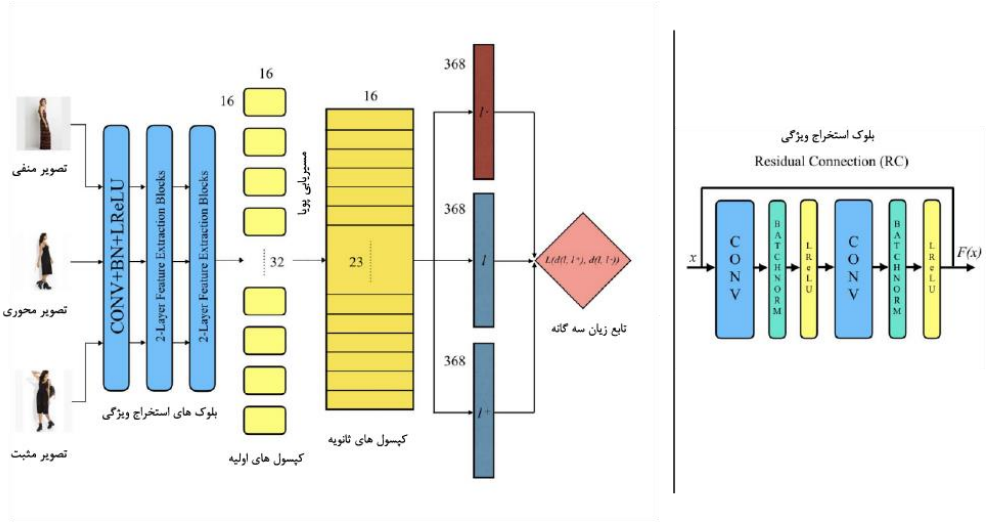
[4] Sabour and Hinton et al.
[5] LaLonde et al.
[6] Zhang et al.
[7] Jaiswal et al.
[8] Generative Adversarial Networks
[9] Kosiorek et al.
[10] Gu et al.
[11] Wang et al.
[12] Eldifrawi et al.

درایه ای استفاده می کند. گو و همکاران[13] [10] در سال ۲۰۲۱، شبکه کانولوشنی پیشرفته تحت عنوان ConvNet-Avg را پیشنهاد میکنند و سعی میکنند مولفه هایی که موجب پیشرفت شبکه های کپسولی شده است را به شبکه های عصبی کانولوشنی اضافه نمایند. ما و همکاران[14] [11] در سال ۲۰۲۱، معماری کپسولی تحت عنوان CapsuleRRT را جهت انجام رگرسیون با در نظر گرفتن روابط بین بخش های تشکیل دهنده اشیاء پیشنهاد میدهند.

در حوزه بازیابی تصاویر پوشاک، کیاپور و همکاران[15] [12] در سال ۲۰۱۵، پژوهشی بسیار چالش‌برانگیز با عنوان "Exact Street to Shop" را معرفی می‌کنند. هدف این مطالعه تطابق دقیق تصاویر گرفته شده توسط مشتریان با تصاویر متناظر فروشگاهی می باشد. هانگ و همکاران[16] [13] در سال ۲۰۱۵، شبکه ویژگی محور دوگانه (DARN) را برای حل مشکل تطبیق تصاویر فروشگاهی با تصاویر مشتریان پیشنهاد میدهند. لیو و همکاران[17] [14] در سال ۲۰۱۶، مجموعه داده DeepFashion را معرفی می کنند که حجم وسیعی از تصاویر پوشاک در ابعاد بزرگ را در خود جای داده است و شامل ویژگی های متعدد و اطلاعات برجسته ای از تصاویر پوشاک می باشد. همچنین معماری FashionNet جهت بازیابی تصاویر پوشاک پیشنهاد شده‌است. شاف و همکاران[18] [15] در سال ۲۰۱۵، شبکه سه گانه را معرفی می نمایند که طرحی برای یادگیری تشابه میان جفت‌ها میباشد به طوری که شبکه بتواند نزدیک ترین و دورترین نمونه نسبت به نمونه ورودی را بیاموزد. در نهایت معماری FaceNet با بهره گیری از شبکه سه گانه در حوزه تشخیص چهره پیشنهاد شده است. هرمانس و همکاران[19] [16] در سال ۲۰۱۷، به منظور ارتقاء شبکه سه گانه دو تکنیک نمونه گیری سخت تحت عنوان های Batch hard وBatch all را پیشنهاد می کنند. تا در نهایت با اعمال نمونه های سخت به شبکه سه گانه، عملکرد نهایی افزایش یابد. چن و همکاران[20] [17] در سال ۲۰۱۷، شبکه چهار گانه را با هدف ارتقاء شبکه سه گانه معرفی میکنند. همچنین در این پژوهش تکنیک MargOHNM جهت انجام نمونه برداری سخت معرفی شده است. میائو و همکاران[21] [18] در سال ۲۰۲۰، معماری ClothingNet را جهت بازیابی تصاویر پوشاک چند دامنه ای پیشنهاد کرده اند. همچنین در این پژوهش تعریف جدیدی از تابع زیان چهار گانه ارائه شده است که متناسب با حوزه بازیابی پوشاک می باشد.

اساس کار ما بر اساس معماری RCCapsNet [1] می باشد که در سال ۲۰۱۹ توسط کینلی و همکاران ارائه شده است. در این مقاله نویسندگان سعی دارند تا با طراحی ساختاری سه گانه بر پایه شبکه های کپسولی، عملیات بازیابی پوشاک بدون حساسیت به زاویه دید و بدون استفاده از اطلاعات جانبی تصاویر انجام گردد. شکل ۱ ساختار معماری RCCapsNet را نشان می دهد. ورودیِ این طرح شامل سه تصویر می باشد که به صورت تصویر محوری، تصویر مثبت و تصویر منفی تعریف شده‌اند. در این معماری ساختارهای کپسولی یکسان هستند و بلوک های استخراج ویژگی به کپسول های اولیه متصل میشوند. در نهایت از تابع زیان سه گانه جهت ارزیابی ویژگی های استخراج شده، استفاده شده است.

[13] Gu et al.
[14] Ma et al.
[15] Kiapour et al.
[16] Huang et al.
[17] Liu et al.
[18] Schroff et al.
[19] Hermans et al.
[20] Chen et al.
[21] Miao et al.

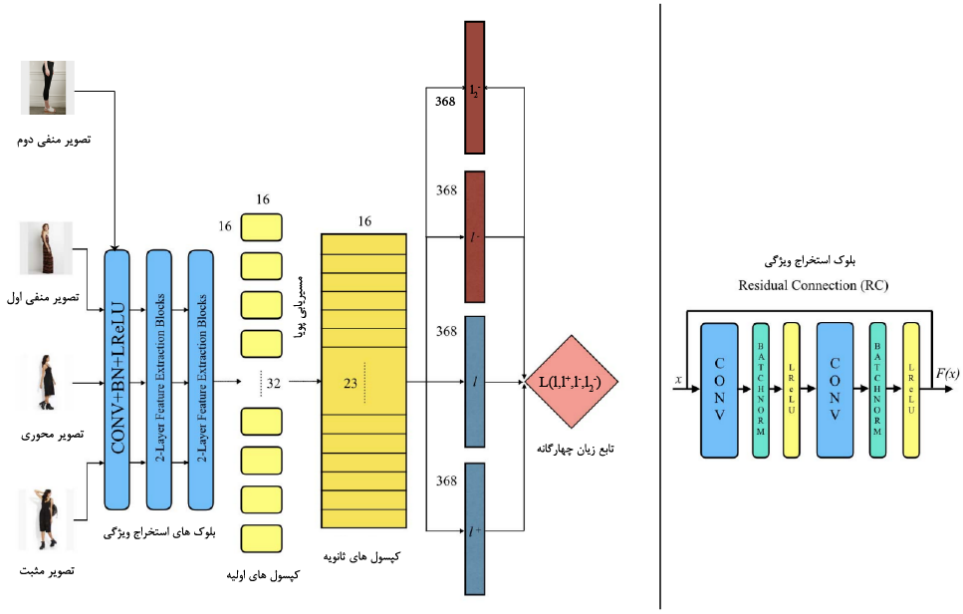


شکل ۱- معماری RCCapsNet [۱]

۲- روش پیشنهادی

این پژوهش با هدف ارتقاء معماری RCCapsNet انجام شده است. جهت رسیدن به این امر نسخه چهار گانه معماری RCCapsNet با بهره گیری از تابع زیان چهار گانه [18,17] توسعه یافته است. بهره گیری از تابع زیان چهار گانه به جای تابع زیان سه گانه موجب میشود تا شبکه درک بهتری نسبت به تفاوت های درون کلاسی و بین کلاسی با توجه به ساختار سلسله مراتبی تصاویر پوشاک داشته باشد. از آنجایی که تابع زیان سه گانه تنها از یک برچسب پشتیبانی می نماید، تنها قادر به درک تفاوت های درون کلاسی می باشد در حالی که تابع زیان چهار گانه با پشتیبانی از دو برچسب قادر می باشد تا علاوه بر درک تفاوت های درون کلاسی، تفاوت های بین کلاسی را نیز تشخیص دهد.

به عنوان مثال تصاویر پوشاک در کلاس هایی همچون پیراهن، شلوار، تیشرت و غیره طبقه بندی میشوند. علاوه بر این در کلاس پیراهن، محصولات با شناسه هایی همچون id_0001 و id_0002 و در کلاس شلوار محصولات با شناسه هایی همچون id_0010 و id_0011 دسته بندی می شوند. همچنین در هر شناسه مانند id_0001 ، تصاویر مرتبط با آن محصول از زوایای گوناگون مانند جلو، پشت و غیره قرار داده میشود. با توجه به شرایط ذکر شده، تابع زیان سه گانه که از یک نمونه منفی پشتیبانی می نماید، تنها قادر به درک تفاوت در سطح شناسه می باشد. حال آنکه تابع زیان چهار گانه با پشتیبانی از دو نمونه منفی قادر به درک تفاوت در هر دو سطح شناسه و کلاس می باشد و به همین دلیل می تواند جایگزین مناسبی برای تابع زیان سه گانه با توجه به ساختار سلسله مراتبی تصاویر پوشاک باشد وکارایی بهتری را حاصل نماید. شکل ۲ ساختار معماری چهارگانه پیشنهادی را نشان میدهد که تحت عنوان QRCCapsNet نامگذاری شده است.



شکل ۲- معماری پیشنهادی QRCCapsNet

همانطور که در تصویر ۲ نشان داده شده است ورودی این معماری یک چهار گانه میباشد که از تصویر محوری، تصویر مثبت، تصویر منفی اول و تصویر منفی دوم تشکیل شده است. تصویر مثبت دارای شناسه یکسان نسبت به تصویر محوری می باشد. تصویر منفی اول متعلق به کلاسی یکسان نسبت به تصویر محوری می باشد و تصویر منفی دوم متعلق به کلاسی متفاوت نسبت به تصویر محوری می باشد. نمونه ای از یک چهار گانه در شکل ۳ نشان داده شده است.

شکل ۳- نمونه ای از یک چهار گانه

در معماری فوق تابع زیان چهارگانه مطابق رابطه (۱) تعریف شده است.

$$L = \lambda \,.\, L_{id} + \mu.\, L_{class} = \lambda \,.\, Max(AP - AN1 + \alpha_1, 0) + \mu \,.\, Max(AN1 - AN2 + \alpha_2, 0) \quad (۱)$$

$$\alpha_1 = 0.1\,, \alpha_2 = 1$$

$$\lambda = 0.1\,, \mu = 0.9$$

مطابق رابطه (۱) ، تابع زیان چهارگانه از دو بخش L_{id} و L_{class} تشکیل شده است. بخش L_{id} مسئول این است که در یک کلاس یکسان مانند پیراهن، فاصله بین شناسه های مختلف بزرگتر از فاصله بین تصاویر متعلق به یک شناسه خاص باشد. بخش L_{class} مسئول این است که فاصله بین کلاس های مختلف بزرگتر از فاصله درون کلاسی میان شناسه های متعلق به یک کلاس خاص باشد. در ادامه AP بیانگر فاصله اقلیدسی میان بردار ویژگی های تصویر محوری و تصویر مثبت می باشد $AN1$.بیانگر فاصله بین تصویر محوری و تصویر منفی اول و $AN2$ بیانگر فاصله بین تصویر محوری و تصویر منفی دوم می باشد .همچنین α_1 بیانگر حداقل فاصله میان شناسه های مختلف درون یک کلاس خاص و α_2 بیانگر حداقل فاصله میان کلاس های گوناگون می باشد. از آنجاییکه معمولا تفاوت های درون کلاسی کمتر از تفاوت های بین کلاسی می باشد، مقدار α_1 کوچکتر از α_2 در نظر گرفته شده است. علاوه براین با رویکردی مشابه، مقدار ضریب λ کوچکتر از ضریب μ در نظر گرفته شده است.

۲-۱- مجموعه داده

این پژوهش با بهره گیری از مجموعه داده DeepFashion-Inshop[14] انجام شده است که یکی از برجسته ترین مجموعه های داده در زمینه تصاویر پوشاک درون فروشگاهی میباشد و شامل بیش از ۵۲ هزار تصویر در ۲۳ دسته متفاوت می باشد. علاوه بر این، اطلاعات جانبی مربوط به تصاویر مانند بافت،جنس، شکل، سبک و ... نیز ارائه شده است اما از این اطلاعات استفاده نشده است زیرا شبکه های کپسولی در ذات خود قادر به یادگیری این موارد هستند. نحوه توزیع تصاویر به تفکیک گروه در جدول ۱ ارائه شده است.

جدول ۱- نحوه توزیع تصاویر به تفکیک گروه در مجموعه داده DeepFashion- in-shop

عنوان گروه	تعداد دسته	تعداد تصاویر
بانوان	۱۴	۴۴.۸ هزار
آقایان	۹	۷.۸ هزار

۲-۲- محیط پیاده سازی

پژوهش فعلی در پلتفرم Google Colab-GPU انجام شده است. همچنین با توجه به حجم بالای تصاویر آموزشی و طولانی بودن زمان آموزش در شرایط سخت افزاری موجود، جهت مقایسه روش پیشنهادی با معماری مبنا در شرایط یکسان، پارامتر epochs که تعیین کننده تعداد دفعات تکرار فرایند آموزش می باشد برابر با ۱ دور در نظرگرفته شده است. سورس کد در گیت هاب منتشر شده است.[۲۲]

[22] https://github.com/MahdiyeKhatami/QRCCapsNet

۲-۳- معیار ارزیابی

جهت ارزیابی راهکار های پیشنهادی از معیار Recall@K استفاده شده است تا میزان کارایی با توجه به k نمونه اول مورد ارزیابی قرار گیرد. در فرایند ارزیابی، به ازای هر تصویر از مجموعه پرس و جو، k تصویر از مجموعه گالری بازیابی میشود. حال در هر تصویر پرس و جو در صورتی که حداقل یکی از نتایج به درستی بازیابی شده باشد، یک امتیاز برای آن در نظر گرفته میشود. این فرایند به ازای تمامی تصاویر موجود در مجموعه پرس و جو تکرار شده و در نهایت میانگین امتیازات به عنوان میزان کارایی اعلام میگردد.

۳- نتایج

در ادامه نتایج حاصل از روش پیشنهادی با روش مبنا مقایسه شده است. نتایج مقایسه بر اساس معیار ارزیابی Recal@K در دو مجموعه پوشاک بانوان و پوشاک آقایان به ترتیب در جدول ۲ و ۳ ارائه شده است.

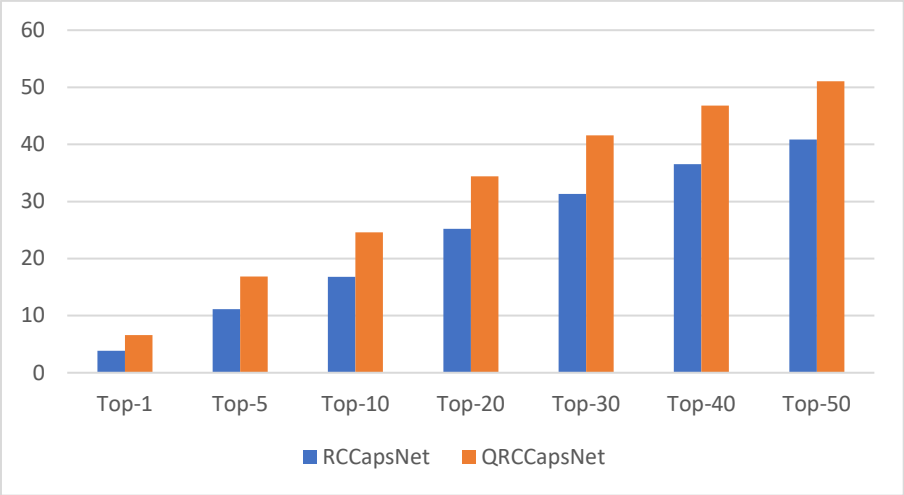
جدول ۲- مقایسه روش پیشنهادی با مبنا بر اساس معیار ارزیابی Recal@K در مجموعه پوشاک بانوان

	Top-1	Top-5	Top-10	Top-20	Top-30	Top-40	Top-50	زمان آموزش	زمان آزمایش
RCCapsNet (مبنا)	۳.۸۳	۱۱.۱۲	۱۶.۸۱	۲۵.۱۹	۳۱.۳۴	۳۶.۵۳	۴۰.۸۵	۰:۳۵:۳۶	۰:۲۹:۵۶
QRCCapsNet (پیشنهادی)	۶.۵۹	۱۶.۸۷	۲۴.۵۹	۳۴.۴	۴۱.۵۶	۴۶.۷۸	۵۱.۰۸	۰:۴۵:۲۰	۰:۳۱:۴۰

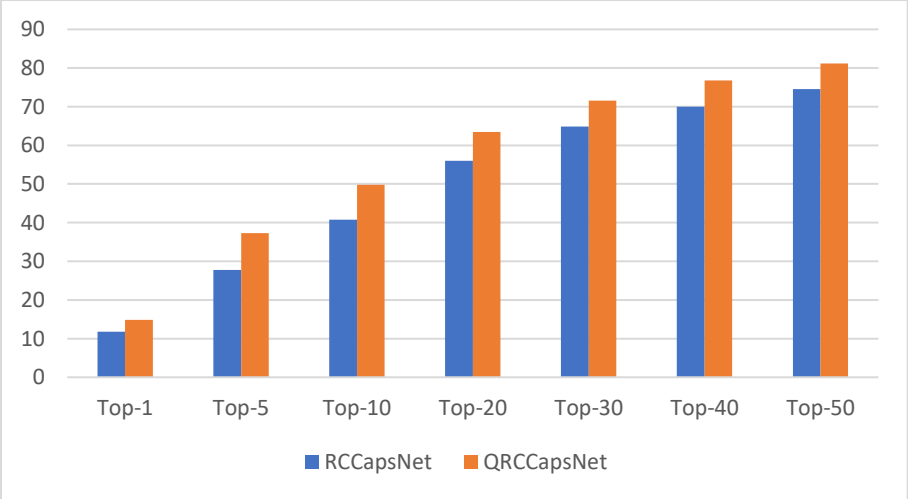
جدول ۳- مقایسه روش پیشنهادی با مبنا بر اساس معیار ارزیابی Recal@K در مجموعه پوشاک آقایان

	Top-1	Top-5	Top-10	Top-20	Top-30	Top-40	Top-50	زمان آموزش	زمان آزمایش
RCCapsNet (مبنا)	۱۱.۷۷	۲۷.۷۸	۴۰.۷۵	۵۶	۶۴.۸۷	۷۰	۷۴.۵۹	۰:۰۴:۵۸	۰:۰۱:۱۵
QRCCapsNet (پیشنهادی)	۱۴.۸۵	۳۷.۳۲	۴۹.۷۵	۶۳.۴۹	۷۱.۶	۷۶.۷۷	۸۱.۱۹	۰:۰۶:۲۸	۰:۰۱:۱۷

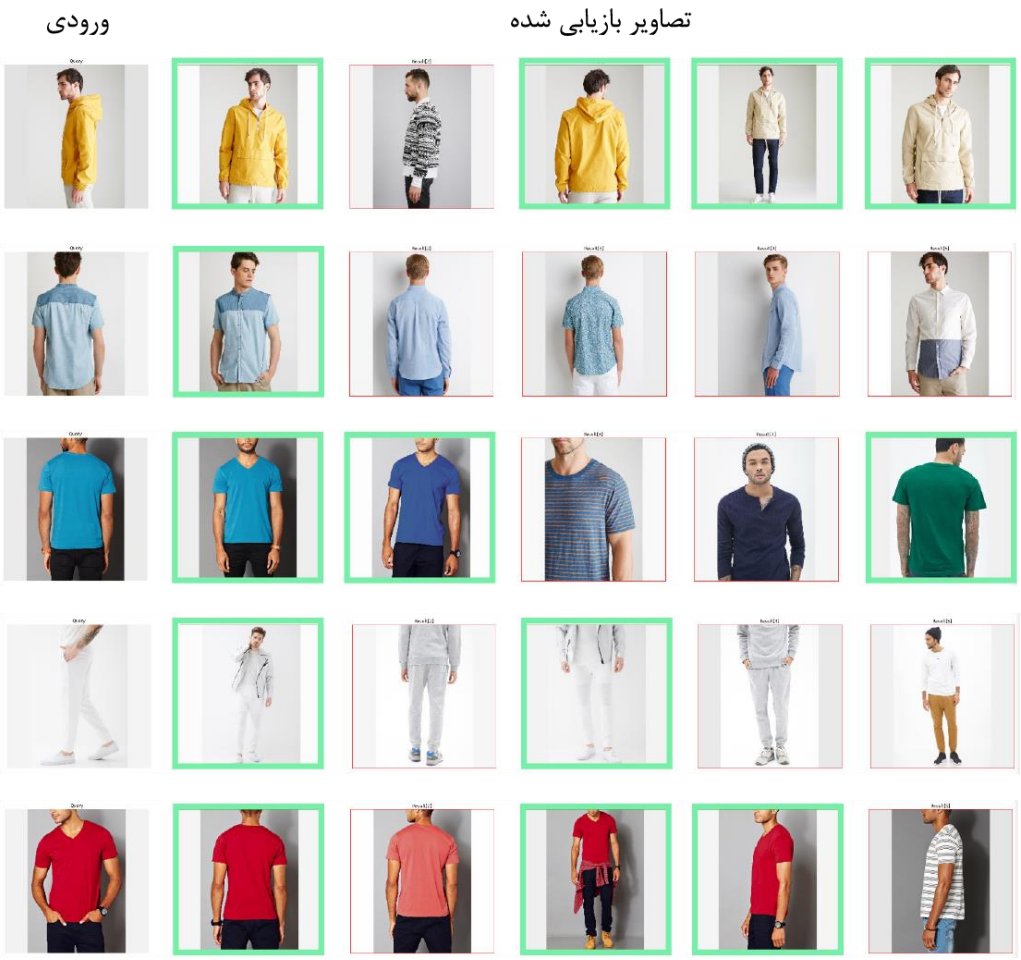
نتایج حاصل شده نشان میدهد روش پیشنهادی موجب بهبود در نتایج شده است. میزان بهبود حاصل شده بر اساس معیار های Top-10، Top-20 و Top-30 در مجموعه پوشاک بانوان به ترتیب برابر با ۷.۷۸ درصد،۹.۲۱ درصد و ۱۰.۲۲ درصد و در مجموعه پوشاک آقایان به ترتیب برابر با ۹ درصد، ۷.۴۹ درصد و۶.۷۳ درصد می باشد. با این وجود زمان آموزش در مجموعه پوشاک بانوان و آقایان به ترتیب ۲۷.۳۴ درصد و ۳۰.۲۰ درصد افزایش یافته است. در ادامه نمودار مقایسه روش پیشنهادی با مبنا در مجموعه های پوشاک بانوان و پوشاک آقایان به ترتیب در نمودار های ۱ و ۲ ارائه شده است. همچنین نمونه ای از تصاویر بازیابی شده در شکل ۴ ارائه شده است.



نمودار ۱- مقایسه روش پیشنهادی (QRCCapsNet) با مبنا (RCCapsNet) در مجموعه پوشاک بانوان



نمودار ۲- مقایسه روش پیشنهادی (QRCCapsNet) با مبنا (RCCapsNet) در مجموعه پوشاک آقایان



شکل ۴- نمونه ای از تصاویر بازیابی شده

۴- بحث و نتیجه‌گیری

این پژوهش با هدف ارتقاء معماری RCCapsNet انجام شده است. جهت رسیدن به این امر نسخه چهار گانه معماری RCCapsNet با بهره گیری از تابع زیان چهار گانه تحت عنوان QRCCapsNet ارائه شده است. بهره گیری از تابع زیان چهار گانه به جای تابع زیان سه گانه موجب شده است تا شبکه درک بهتری نسبت به تفاوت های درون کلاسی و بین کلاسی با توجه به ساختار سلسله مراتبی تصاویر پوشاک داشته باشد و در نهایت افزایش کارایی شبکه را به همراه داشته است. میزان بهبود کارایی بر اساس معیار های Top-10، Top-20 و Top-30در مجموعه پوشاک بانوان به ترتیب برابر با ۷.۷۸ درصد،۹.۲۱ درصد و ۱۰.۲۲درصد و در مجموعه پوشاک آقایان به ترتیب برابر با ۹ درصد، ۷.۴۹ درصد و۶.۷۳ درصد می باشد. با این وجود زمان آموزش در مجموعه پوشاک بانوان و آقایان به ترتیب ۲۷.۳۴ درصد و ۳۰.۲۰ درصد افزایش یافته است.

منابع

1- F. Kinli, B. Ozcan, F. Kirac: Fashion Image Retrieval with Capsule Networks.Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019 .

2- S. Sabour, N. Frosst, and G. E. Hinton. Dynamic routing between capsules. In Advances in Neural Information Processing Systems 30, pages 3856–3866. 2017.

3- R. LaLonde and U. Bagci, “Capsules for object segmentation,” in arXiv preprintarXiv:1804.04241, 04 2018.

4- S. Zhang, Q. Zhou, and X. Wu, “Fast Dynamic Routing Based on Weighted Kernel Density Estimation,” in Cognitive Internet of Things, 2018.

5- A. Jaiswal, W. AbdAlmageed, Y. Wu, and P. Natarajan, “CapsuleGAN: Generative Adversarial Capsule Network,” in The European Conference on Computer Vision (ECCV) Workshops, September 2018.

6- A. R. Kosiorek, S. Sabour, Y. W. Teh, and G. E. Hinton, “Stacked Capsule Autoencoders,” in arXiv preprint arXiv:1906.06818, 06 2019.

7- J. Gu and V. Tresp, "Improving the Robustness of Capsule Networks to Image Affine Transformations," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 7283-7291, doi: 10.1109/CVPR42600.2020.00731.

8-Y. Wang, L. Huang, S. Jiang, Y. Wang, J. Zou, H. Fu, S. Yang, "Capsule Networks Showed Excellent Performance in the Classification of hERG Blockers/Nonblockers," Front Pharmacol. 2020 Jan 28;10:1631. doi: 10.3389/fphar.2019.01631. PMID: 32063849; PMCID: PMC6997788.

9- I. Eldifrawi, M. Abo-Zahhad, A. H. Abd El-Malek and M. Abdelwahab, "Deep Fast Embedded CapsNet: Going Faster with Deep-Caps," 2021 IEEE International Midwest Symposium on Circuits and Systems (MWSCAS), 2021, pp. 187-191, doi: 10.1109/MWSCAS47672.2021.9531794.

10- J. Gu, V. Tresp and H. Hu, "Capsule Network is Not More Robust than Convolutional Network," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 14304-14312, doi: 10.1109/CVPR46437.2021.01408.

11- D. Ma and X. Wu, "CapsuleRRT: Relationships-aware Regression Tracking via Capsules," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 10943-10952, doi: 10.1109/CVPR46437.2021.01080.

12- M. Hadi Kiapour, X. Han, S. Lazebnik, A. C. Berg, and T. L. Berg, “Where to Buy It: Matching street clothing photos in online shops,” pp. 3343–3351, 12 2015.

13- J. Huang, R. Feris, Q. Chen, and S. Yan, “Cross-Domain Image Retrieval with a Dual Attribute-aware Ranking Network,” in Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1062–1070, 2015.

14- Z. Liu, P. Luo, S. Qiu, X. Wang, and X. Tang, “DeepFashion: Powering robust clothes recognition and retrieval with rich annotations,” in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016.

15- F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A unified embedding for face recognition and clustering,” CoRR, vol. abs/1503.03832, 2015.

16- A.Hermans, L. Beyer, and B. Leibe, "In defense of the triplet loss for person re-identification." arXiv preprint arXiv:1703.07737 (2017).

17- W. Chen, X. Chen, J. Zhang and K. Huang, "Beyond Triplet Loss: A Deep Quadruplet Network for Person Re-identification," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1320-1329, doi: 10.1109/CVPR.2017.145.

18- Y. Miao, G. Li, C. Bao, J. Zhang and J. Wang, "ClothingNet: Cross-Domain Clothing Retrieval With Feature Fusion and Quadruplet Loss," in IEEE Access, vol. 8, pp. 142669-142679, 2020, doi: 10.1109/ACCESS.2020.3013631.

Fashion Image Retrieval With Quadruplet Capsule Networks

Mahdiye Sadat Khatami, Dr Mohammad Javad Fadaeieslam

mahdiye_khatami@semnan.ac.ir

Abstract

The purpose of fashion image retrieval is to find images similar to the image searched in online shopping. one of the challenges is to retrieval images without sensitivity to the viewing angle and without using additional information. In previous studies, the RCCapsNet architecture has been proposed with the aim of replacing capsule networks with convolutional neural networks to solve the mentioned problems, which has used the triplet loss function to achieve this. This research is done with the aim of improving the RCCapsNet. To achieve this, the quadruplet version of the RCCapsNet is proposed using the quadruplet loss function which is named as QRCCapsNet. The proposed method is trained on the DeepFashion-Inshop dataset and the performance is evaluated by Recall@K metric. The results indicate the success of our proposed method

Keywords: fashion image retrieval, in-shop clothing retrieval, capsule networks, quadruplet loss function