اولین همایش ملی جویشگر بومی
۲۷ و ۲۸ بهمن ماه سال ۱۳۹٤
پژوهشگاه ارتباطات و فناوری اطلاعات
(مرکز تحقیقات مخابرات ایران)
وزارت ارتباطات و فناوری اطلاعات
ITRC

مروری بر ضرورت‌ها و کاربردهای شبکه‌ی تصویری

فرزانه رحمانی[۱]، علی حافظی[۲]، فرزاد زرگری[۳]

[۱] دانشجوی دکتری، پژوهشگاه ارتباطات و فناوری اطلاعات، تهران
rahmani@itrc.ac.ir

[۲] دانشجوی کارشناسی ارشد، دانشگاه آزاد اسلامی واحد علوم و تحقیقات، تهران
a.hafezi@itrc.ac.ir

[۳] عضو هیئت علمی، پژوهشگاه ارتباطات و فناوری اطلاعات، تهران
zargari@itrc.ac.ir

چکیده

دستاوردهای جدید پژوهشی در زمینه‌ی پردازش تصویر و بینایی ماشین نشان داده است که یکی از نیازمندی‌های اصلی برای بهبود عملکرد الگوریتم‌های کاربردی بینایی ماشین، وجود پایگاه داده‌های چندرسانه‌ای است که محتوای آموزشی مناسبی را برای این الگوریتم‌ها فراهم سازد تا حدی که بتوان قدرت بینایی و پردازش عامل‌های نرم‌افزاری هوشمند تولید شده را به قدرت تشخیص و درک انسان از محیط پیرامون نزدیک گرداند. نمونه‌ی با ارزش این‌گونه پایگاه داده‌های چندرسانه‌ای شبکه‌ی تصویری می‌باشد که علاوه بر فراهم آوردن حجم بسیار زیادی از تصاویر، آن‌ها را در یک ساختار سلسله‌مراتبی مفهومی‌طبقه‌بندی می‌کند. در این مقاله نخست شبکه‌ی تصویری **ImageNet** معرفی شده و مروری بر ضرورت و کاربردهای ایجاد این شبکه‌ی تصویری انجام گرفته است. سپس ضرورت‌های بومی‌سازی شبکه‌ی تصویری و ارتباط آن با جویشگر بومی‌شرح داده شده است.

کلمات کلیدی

شبکه‌ی تصویری، ImageNet، بینایی ماشین(Machine Vision)، هوش بصری(Smart Visioning)، زیرساخت داده‌ای تصویری، تشخیص اشیاء، توصیف تصویر، کلان داده ها، یادگیری ماشین.

۱- مقدمه

دیدن و درک صحیح توسط ماشین‌ها همواره یکی از چالش‌برانگیزترین مسایل انسان در صد سال اخیر بوده است. از انقلاب صنعتی تاکنون انسان‌ها به‌دنبال ساخت دنیایی بوده‌اند که بتوانند وظایف انسان‌ها را به ماشین‌ها بسپارند. شاید بتوان گفت، موفقیت‌های بسیاری نیز در راستای این هدف بدست آورده‌اند. از ساخت موتور بخار تا عامل‌های نرم‌افزاری هوشمند، گام‌های به سزایی در این زمینه برداشته شده است. در حال حاضر هوش مصنوعی با سرعت بسیار بالایی در حال رشد است و حتی در مواردی عامل‌های نرم‌افزاری از انسان پیشی گرفته‌اند. طیف وسیعی از عامل‌های نرم‌افزاری تولید شده از بازی‌ها رایانه‌ای نظیر شطرنج تا جویشگرها همگی اثباتی بر این مدعا می‌باشند. ولی با وجود امکانات بالای ضبط تصاویر توسط دوربین‌های چندین مگاپیکسلی و دستگاه‌های ذخیره‌سازی با ظرفیت‌های چندین ترابایتی هنوز یک مشکل اساسی در بینایی ماشین[۱] وجود دارد و آن درست دیدن و درک صحیح تصاویر توسط ماشین‌های تولید شده به دست انسان است.

با وجود پیشرفته‌ترین نرم‌افزارها و ماشین‌های ثبت، ذخیره‌سازی و تشخیص تصاویر، هیچ ساخته‌ی بشر توانایی تشخیص و درک حقایق تصاویر پیرامون را در حد قدرت بینایی چشم و تشخیص مغز انسان دارا نیست. نرم‌افزارهای پردازش و تصاویر دوربین‌های با وضوح مگاپیکسلی به قدر کافی در کمک به نابینایان موفق نبوده‌اند، پردازش تصاویر در دوربین‌های ترافیکی قادر به تشخیص و درک ماشین‌های فرسوده و آلاینده نمی‌باشند، دوربین‌های امنیتی بین سارق و یک فرد عادی تفاوتی قایل نمی‌شوند. اما با توجه به رشد روزافزون حجم تصاویر و ویدئوها، استفاده از قدرت بینایی و تشخیص انسان برای پردازش این تعداد وسیع از تصاویر و ویدئوها ممکن نیست و باید این کار به ماشین‌های هوشمند سپرده شود. هرچند پیشرفته‌ترین و هوشمندترین نرم‌افزارهای امروزی نیز در مدیریت و فهم این حجم عظیم از تصاویر مشکل دارند.

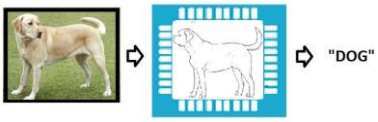
کلید حل این مشکل می‌تواند دنبال کردن رفتار طبیعت و شبیه‌سازی و الگوبرداری از طبیعت باشد. طبیعت حدود ۵۴۰ میلیون سال صرف مسئله تکامل کرده است. بیشتر این تلاش صرف نظر از خود چشم به ابزار پردازش دید در مغز جانوران اختصاص داده شده است. دیدن با چشم آغاز می‌شود ولی در مغزشکل می‌گیرد. اگر چشم‌های یک انسان را مانند یک جفت دوربین بیولوژیکی در نظر بگیریم،از طریق آنها به طور متوسط هر ۲۰۰ میلی ثانیه یک عکس در مغز ذخیره می‌شود. بنابراین یک کودک تا سه سالگی صدها میلیون تصویر از دنیای واقعی در مغز خود ذخیره کرده است. این تعداد بسیار زیاد تصویر، برای یادگیری مفاهیم بصری توسط مغز کودک استفاده می‌شود. این حقایق در مورد بینایی انسان و درک او از محیط پیرامونش ما را با این حقیقت آشنا می‌کند که انسانها این درک و شناخت را تنها با مشاهده و کسب تجربه یاد می‌گیرند. در واقع پیوند یادگیری ماشین به بینایی رایانه‌ای می‌تواند کلیدی برای حل این مشکل باشد.

ترتیب ارائه باقی مطالب در این مقاله بدین شرح است. در ادامه در بخش دو مروری بر بینایی ماشین و ضرورت شبکه ی تصویری در مباحث بینایی ماشین و پردازش تصویر مطرح می گردد. سپس در بخش سوم شبکه ی تصویری ImageNet معرفی شده و یافته‌ها و کاربردهای آن به عنوان یک شبکه ی تصویری موفق بیان می‌گردد. بخش چهارم نیاز به بومی‌سازی یک شبکه ی تصویری را به عنوان زیرساخت داده‌ای برای جویشگر بومی مورد بحث قرار می دهد و در انتها مقاله با ارائه نتیجه گیری جمع بندی می‌شود.

۲- بینایی ماشین و نیاز به شبکه‌ی تصویری

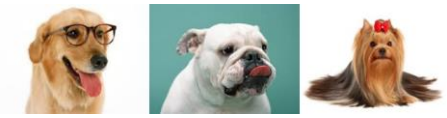
بینایی ماشین به دنبال تبیین روشی برای دیدن و درک تصویر بوسیله ماشین می‌باشد. آنچه محققان بینایی ماشین به دنبال آن هستند، این است که در نهایت ماشین‌ها بتوانند مانند انسان ها، بر روی اشیا نام بگذارند، افراد را تشخیص بدهند، محیط سه بعدی را به همراه عمق درک کنند، ارتباطات، احساسات، اعمال و نیت‌ها را بفهمند.

نخستین گام برای رسیدن ماشین‌ها به درکی صحیح از تصاویر این است که ماشین بتواند در یک تصویر اشیای مختلف را ببیند. برای رسیدن به چنین درکی نیاز به الگوریتمی‌است که به ازای تصاویر ورودی دریافتی، خروجی مناسب بصورت متن بدهد. برای اینکه ماشین بتواند دید درستی از یک شی داشته باشد، به توصیفی نیاز است که آن شی را به صورت ریاضی مدل کند. اگر بخواهیم دیدن و تشخیص یک شی یا موجود خاص مانند سگ را به رایانه آموزش دهیم، یک مدل از سگ‌ها طراحی می‌کنیم. میتوانیم به رایانه بگوییم سگ‌ها دارای گوش بلند و پهن هستند و بدنی کشیده اما جثه‌ای کوچک به همراه دم دارند. به این ترتیب هرگاه رایانه تصویری از سگ‌ها را مشاهده کند که با الگو مطابقت داشته باشد به درستی آن را تشخیص می‌دهد (شکل (۱)).



شکل (۱) : تطبیق الگو

در نتیجه رایانه‌ها بر اساس الگوها و تطابق الگوها با تصاویر، اشیا مورد نظر را شناسایی و پیدا می‌کنند. نکته قابل توجه این است که همانطور که گفتیم با تطابق دادن الگو با عکس‌ها می‌توان شی را تشخیص داد. اما اگر رایانه نتواند شی را با مدلش انطباق بدهد چه می‌شود؟ با توجه به شکل (۲) و مدلی که برای تصویر سگ‌ها طراحی کرده ایم، رایانه نمی‌تواند تصاویر سگ‌های شکل (۲) را تشخیص دهد. چرا که این تصاویر مطابق با الگوی طراحی شده برای رایانه نیستند.



شکل (۲) : تصاویر سگ که با الگو مطابقت ندارند.

اولین راه حلی که به ذهن اغلب افراد می‌رسد این است که به ازای هر تصویر سگی که طبق الگو رفتار نمی‌کند، یک الگوی جدید طراحی کنیم. به گونه‌ای که اگر الگوی شماره یک با تصویر تطابق نداشت، الگوی شماره دو سپس الگوی شماره سه و به همین صورت ادامه یابد تا شی درون عکس با الگوی مورد نظر تطابق داشته باشد. به نظر درست می‌آید. اما طراحی الگو برای یک شی خاص باید تا کجا ادامه پیدا کند و اینکه آیا انسان می‌تواند تمامی‌الگوهای موجود را طراحی کند؟ جواب بسیار بدیهی است. حتی یک حیوان خانگی ساده مانند سگ، بی نهایت الگو دارد. که طراحی تمامی‌آنها کاری تقریبا غیر ممکن است. این درحالی است که سگ‌ها تنها یک مورد هستند. اشیا گوناگون و متفاوت بسیار زیاد دیگری هستند که طراحی تمام آن‌ها کاری غیر عقلانی و محال است. پژوهش‌های نوین نشان می‌دهند که شبکه‌ی تصویری راهکار بسیار ارزشمندی در آموزش مدل‌های گوناگون مفاهیم متفاوت در جهت رسیدن ماشین‌ها به درک و تشخیص تصاویر است. شبکه‌ی تصویری بر اساس یک سلسله مراتب مفهومی‌ساخته می‌شود و به هر مفهوم در این سلسله مراتب تعداد بسیار زیادی تصویر نسبت داده می‌شود.

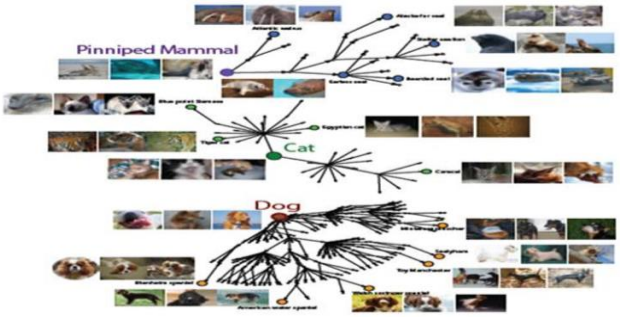
۳- شبکه‌ی تصویری ImageNet

در این بخش به معرفی یک شبکه‌ی تصویری موفق به نام ImageNet پرداخته می‌شود [1]. که تهیه‌ی آن حدود ۴ سال در دانشگاه استنفورد به طول انجامید و اکنون به عنوان موفق ترین شبکه‌ی تصویری شناخته می‌شود. پایگاه داده‌ی سلسله‌مراتبی ImageNet به طور گسترده در تحقیقات پردازش تصویر مانند آموزش شبکه‌های عصبی، یادگیری عمیق[۲]، برچسب زنی خودکار تصاویر و همچنین تحقیقات گوگل مورد استفاده قرار گرفته است به طوری که بیش از۱۹۶۰و ۱۶۰۰ مورد مقاله ارجاع داده شده به دو مقاله اصلی شبکه تصویری وجود دارد و بیش از ۳۲۰ مورد مقاله و پروژه در وبسایت این پایگاه داده ذکر شده است.

یکی از بنیان گذاران اصلی پروژه‌ی ImageNet خانم پروفسور فی فی لی[۳] می‌باشد که هدایت آزمایشگاه بینایی ماشین دانشگاه استنفورد را برعهده دارد. او به همراه تیمی قوی از اساتید و دانشجویان به مدت ۱۵ سال در حوزه‌ی بینایی ماشین تلاش کرد. ایده‌ی اصلی ImageNet از آن جا ناشی شد که خانم فی فی به این نتیجه رسید که یک کودک در سه سالگی بدون هیچ آموزش و تخصصی می‌تواند تصاویر را به درستی ببیند و تشخیص دهد [2]. نکته حایز اهمیت این است که در سنین ابتدایی این مسائل به کودکان آموزش داده نمی‌شود و تشخیص تنها با تکرار دیدن تصاویر صورت می‌گیرد. این مشاهده به ایده ساخت شبکه‌ای تصویری برای اشیا منجر شد. به‌طوری‌که به ازای هر شی تعداد قابل توجهی تصویر با کیفیت بالا و ابعاد مناسب جمع‌آوری شود. بنابراین تنها کافی است که این تصاویر را به عنوان ورودی آموزشی به عامل نرم‌افزاری داد. اگر به تعداد کافی عکس از هر شی وجود داشته باشد، آن شی با دقت بسیار بالایی در هر موقعیتی در تصویر قابل شناسایی است.

پروژه ImageNet در سال ۲۰۰۷ در دانشگاه استنفورد توسط پروفسورفی فی لی با همراهی پروفسور کای لی[۴] از دانشگاه پرینستون راه اندازی شد. هدف این پروژه جمع‌آوری تعداد زیادی تصویر در ابعاد و کیفیت بالا به صورت یک ساختار شبکه‌ای منسجم بود. با رجوع به اینترنت نزدیک به یک میلیارد تصویر جمع‌آوری شد. تلاش برای برچسب زنی این تصاویر به کمک تکنولوژی جمع‌سپاری[۵] صورت گرفت و برای اجرای جمع‌سپاری از پلتفرم جمع سپاری آمازون[۶] استفاده شد و ImageNet در زمان اوج کاری از بزرگترین کارفرماهای پلتفرم جمع سپاری آمازون بود. در مجموع ۴۸۹۴۰ نفر از ۱۶۷ کشور در سراسر دنیا در اصلاح، مرتب سازی و برچسب گذاری یک میلیارد عکس همکاری کردند [3]. در نهایت عکس‌هایی که کار پالایش و برچسب گذاری آن‌ها تمام شده بود به یک ساختار سلسله‌مراتبی از واژگان متصل شد. برای این ساختار سلسله‌مراتبی واژگان از ساختار وردنت[۷][4] استفاده شد.پس از گذشت دو سال ImageNet تبدیل به یک پایگاه داده تصویری بزرگ با ساختار سلسله‌مراتبی شد. حدود ۱۵ میلیون تصویر در وسعت ۲۲ هزار کلاس از اشیا، که این اشیا با هم رابطه سلسله‌مراتبی پدر فرزندی تشکیل می‌دهند، در این پایگاه داده وجود دارد. از لحاظ کیفیت و کمیت این مقیاس در پایگاه داده‌های تصویری بی سابقه است. به عنوان مثال در مورد گربه‌ها حدود ۶۲ هزار تصویر از گربه‌ها در انواع شکل‌ها و فرم بدن و گونه‌های وحشی و اهلی وجود دارد. شکل (۳) بخش کوچکی از شبکه‌ی تصویری ImageNet را به همراه تصاویر نمونه‌ی آن نمایش می‌دهد. پایگاه داده این پروژه در حال حاضر به صورت کاملا رایگان در اختیار پژوهشگران قراردارد[5].

با وجود تصاویر متفاوت و زیاد از اشیاء مختلف توانایی ساخت الگوریتم‌های کارا مبتنی بر یادگیری ماشین در تشخیص و پردازش و فهم تصاویر فراهم می‌گردد. ImageNet برای بهبود الگوریتم‌های یادگیری ماشین از کلاس خاصی از یادگیری ماشین به نام شبکه‌های عصبی در هم تنیده استفاده کرد. شبکه ی عصبی، پردازشگری با ساختار توزیع شده و قابلیت بالای موازی سازی است که از واحدهای پردازشگر سادهای تشکیل شده است که از مغز انسان الگوبرداری شده‌اند. مانند مغز انسان که از میلیاردها نورون بهم پیوسته تشکیل شده است. نورون یک واحد عملیاتی در یک شبکه عصبی است. شبکه عصبی قابلیت ذخیره کردن تجربیات و به کارگیری آن برای استفاده های آتی را دارا می باشد. یادگیری، یکی از مهمترین بخشهای هوش مصنوعی است. برای هوشمند بودن، یک سیستم که در محیطی با شرایط متغیر قرار دارد، باید توانایی آموختن داشته باشد.



شکل (۳) : نمایی بسیار کوچک از ساختار سلسله‌مراتبی و تصاویر ImageNet[1]

در پروژه ImageNet شبکه‌ی عصبی مانند مغز انسان طراحی شده است. این شبکه عصبی که براساس ساختار سلسله‌مراتبی وردنت چیده شده است دارای ۲۴ میلیون نود،۱۴۰ میلیون پارامتر

و ۱۵ میلیارد اتصال است. با استفاده از نیروی عظیم داده‌های ImageNet و CPUها و GPUهای قدرتمند، شبکه عصبی در هم تنیده به شکلی که هیچ کس انتظارش را نداشت شکوفا گشت. معماری برتر این شبکه عصبی برای تولید نتایج تازه و بدیع در تشخیص اشیا برای اولین بار به کار گرفته شد. در ادامه به یافته‌ها و کاربردهای پروژه ی ImageNet پرداخته می شود.

۳-۱- یافته‌های پروژه‌ی ImageNet

تحقیقات مستمر در پروژه‌ی ImageNet و استفاده از سلسله مراتب تصویری و مفهومی‌آن به یافته‌های ارزشمندی انجامیده است[۶،۷] که راه¬گشای نزدیک شدن به قدرت بینایی و تشخیص انسان در ماشین‌ها است. برخی از دستاوردهای حاصل از به کارگیری ImageNet به شرح زیر است:

- تشخیص اشیا و پیدا کردن موقعیت آن ها

در تشخیص اشیا مرحله نخست پیدا کردن موقعیت دقیق آن‌ها در یک عکس، ابتدایی ترین و مهمترین گام است. عامل نرم‌افزاری باید بتواند تک تک اشیای موجود در تصاویر را به درستی تشخیص داده و با برچسب گذاری صحیح بر روی تصاویر محل آن‌ها را نیز نشان دهد. در شکل (۴) برخی از تصاویری که توسط برنامه‌ی ImageNet برچسب گذاری شده، نمایش داده شده است. همانطور که مشخص شده است، سیستم در تمامی‌تصاویر توانسته به درستی اشیا را تشخیص داده و برچسب را به طور صحیح در موقعیت مناسب قرار دهد. حتی در مواردی که در مورد تشخیص شی مردد بوده است، (مانند شکل سمت راست پایین) در سلسله مراتب مفاهیم به یک لایه بالاتر رجوع کرده و اسم گره پدر آن را در نظر گرفته است.

- توصیف تصاویر

با وجود چنین پیشرفتی بزرگی، در تشخیص اشیا تنها یک گام به جلو برداشته شده است و هنوز فاصله‌ی زیادی با رسیدن به بینش یک کودک سه ساله از محیط اطراف وجود دارد. هر کدام از تصاویر شکل (۵) وقتی به یک کودک سه ساله نشان داده شود، او می‌تواند به خوبی درباره آنها قضاوت کند و توصیفات درستی را به کار ببرد. بنابراین در پروژه ImageNet سعی بر این است تا به درکی معادل درک سه سالگی یک انسان برسند. برای یاد دادن دیدن و توصیف درست از تصویر به رایانه باید پیوندی بین کلان داده ها[۸] و یادگیری ماشین[۹] انجام شود. بنابراین رایانه باید هم از تصاویر و هم از جملات زبان طبیعی که توسط انسانها به کار می‌رود یاد بگیرد. درست مانند مغز که بینایی و زبان را بهم می‌آمیزد. در ImageNet مدلی ایجاد شده است که بخش‌های اجسام بصری در بخش‌های مختلف تصاویر را به عبارات و کلمات در جملات پیوند می‌زند. در سال ۲۰۱۵ در پروژه ImageNet این پیوند زده شد و اولین مدل از دید رایانه‌ای را بوجود آوردند که وقتی تصویری را برای بار اول می‌بینند برای آن جملاتی مانند انسان‌ها تولید می‌کند. همانطور که در شکل (۵) نشان داده شده است، رایانه برای هر یک از تصاویر جمله‌ای به زبان انگلیسی بیان کرده است.

این جمله بیان و توصیف هر کدام از تصاویر است. اگرچه در توصیف سه تصویر به خوبی عمل کرده است اما همانطور که مشاهده می‌شود عکس سمت راست پایین دارای توصیف نادرست است. به هر حال با وجود نقایص، این اولین بار است که انسان توانسته است تا سه سالگی خودش را شبیه سازی کند. اکنون میتوان با اطمینان کامل گفت که انسان توانسته است عامل‌های هوشمندی با هوش یک انسان سه ساله را طراحی و پیاده سازی کند.

۳-۲- کاربردهای ImageNet

ImageNet می‌تواند به عنوان زیرساخت داده‌ای تصویری برای تمام الگوریتم‌های ایجاد هوش بصری[۱۰] بکار گرفته شود. بدون داشتن این مقدار عظیم و ساختار یافته از تصاویر، انسان نمیتواند الگوریتم مناسبی با قدرت تشخیص بشری را طراحی کند. حوزه‌ی تحقیقاتی یادگیری عمیق در هوش مصنوعی که به دنبال پردازش داده‌ها مشابه روشی است که مغز انسان انجام می‌دهد، یکی از استفاده کنندگان اصلی این پایگاه داده‌ی تصویری خواهد بود. در این زمینه مثال‌های بسیار زیادی وجود دارد [8-14]. برای درک بهتر از ضرورت‌های ImageNet برخی از این مثال‌ها را که بسیاری از آنها با استفاده از ImageNet امکان محقق شدن پیدا کردند، در ادامه معرفی می‌شوند.

- رانندگی خودکار : یک عامل نرم‌افزاری باید بتواند فرق کاغذی که مچاله شده و در خیابان افتاده را با یک قطعه سنگ به همان اندازه تشخیص دهد. این امر در تشخیص اینکه میتواند از روی آن رد شود یا نه کمک می‌کند. همچنین باید تفاوت انسان را با اشیا دیگر مانند درخت درک کند. چون در زمان تصادف بین برخورد با انسان و یا یک درخت نیاز به تصمیم بلادرنگ و دقیق دارد.
- هواپیماهای بدون سرنشین: بدون در نظر گرفتن کاربرد هواپیماهای بدون سرنشین (مانندکاربردهای نظامی، رهگیری تغییرات جنگلهای بارانی، ساخت و ساز شهری و...) این هواپیماها باید بتوانند از اشیاء مختلفی که بر فراز آنها پرواز می‌کنند دید درستی داشته باشد.
- دوربین‌های امنیتی: بررسی رفتار موجودیت‌هاودرک نیت و اعمال آنها به درستی. به عنوان مثال یک دوربین امنیتی در یک استخر باید فرق یک کودک که در آب شنا میکند با دیگری که در حال غرق شدن است را تشخیص دهد. یک دوربین امنیتی باید یک سارق را در حال سرقت کردن تشخیص دهد.
- دوربین‌های ترافیکی و کنترل شهر: این دوربین‌ها باید دید وسیع و سطح بالایی از اتفاق‌ها و واقعیت‌های موجود داشته باشند. باید بتوانند رفتارهای اجتماعی را در سطح شهر مشاهده و درک کنند. به عنوان مثال میزان خودروهای فرسوده در شهر را شناسایی کنند.
- ربات‌های امدادگر: در زمان جنگ، بلایای طبیعی، آتش سوزی و سایر حوادث زیان بار جانی به ربات‌های امداد گر نیاز است. ربات‌های امداد گر باید در تشخیص اشیاء، توصیف شرایط و شدت وخامت مجروحان، یافتن اقدام مناسب در جهت بهبود مصدوم و انتخاب بهترین رویکرد در کمک به مجروحان همواره به بهترین شکل ممکن عمل کنند. طراحی صحیح هوش بصری در این زمینه در کاهش تلفات کمک به سزایی می‌کند.
- کمک به پزشکان و پرستاران: در نگهداری بیماران، شناسایی بیماری و کنترل وضعیت بیمار، همواره چشم‌های بیدار و دقیق بهترین کمک را می‌کند. پرستاران نیاز به کنترل مداوم مریض ندارند. تنها در زمان نیاز به بیمار رسیدگی می‌کنند. پزشکان در تشخیص بیماری‌ها و درمان بیمار از ماشین‌ها استفاده می‌برند.
- کمک به نابینایان: نابینایان همواره در انجام امور عادی خود در زندگی دچار مشکلات عدیده شده‌اند. طراحی ماشینی که بتواند ببیند و محیط را برای نابینایان توصیف کند یکی از با ارزش ترین خدمت‌های انسانی در این حوزه است.
- مدیریت و سازماندهی مفاهیم چند رسانه ای بر اساس برچسب: ImageNet ایده‌ی سازمان دهی فایل‌های چندرسانه‌ای را به صورت سلسله‌مراتبی و بر اساس مفهوم مطرح نموده است که مدیریت آن‌ها را نیز ساده می‌سازد.
- جویشگرها: در حال حاضر جویشگرها با روند رو به رشد در حال توسعه هستند. به طوری که امروزه علاوه بر جستجوی متنی جستجوهای دیگری نظیر جستجوی تصویری، ویدیویی و صوتی نیز امکان پذیر است. در صورتی که یک جویشگر تصویری قادر به درک صحیح از تصاویر باشد می‌تواند به قابلیتهای استثنائی دست یابد. برخی کاربردهای هوش بصری برای جویشگر به شرح زیر می‌باشد:

۱. تشخیص اشیا و استنباط منطقی و توصیف حالات از تصویر مورد پرسش[۱۱]: به عنوان مثال با اخذ یک پرس وجوی تصویری، جویشگر باید متن‌ها و سایت‌های مرتبط به آن تصویر را به کاربر برگرداند. در این حالت جویشگر باید بتواند پس از دریافت عکس از سمت کاربر، اشیا موجود در آن عکس را به همراه اتفاقی که در آن عکس در حال رخ دادن است و همچنین سایر استنباط‌های منطقی را به درستی انجام دهد.

۲. تشخیص اشیا و استنباط منطقی و توصیف حالات از تصویر در زمان خزش و نمایه گذاری: در بخش نمایه گذاری جویشگر، برچسب زنی خودکار تصاویر خزش شده بر اساس مفهوم و پیشنهاد برچسب برای این تصاویر توسط سلسله مراتب ImageNet و همچنین با استفاده از هوش بصری امکان پذیر است[15]. به عنوان مثال علاوه بر اینکه با جستجوی واژه "مرغابی" عکس تمامی‌مرغابی‌ها را از سایت‌های مختلف استخراج می‌کند باید بتواند با جستجوی "مرغابی‌های نشسته بر آب" عکس تمامی‌مرغابی‌های‌ی که روی آب نشسته‌اند را برگرداند. برای اینکار جویشگر باید بتواند در زمان نمایه گذاری تصاویر اشیا را به درستی تشخیص دهد.

۳. کاربرد در گسترش پرس و جوی تصویری: پرس و جوهای تصویری می‌توانند مفاهیم زیادی را در بر داشته باشند. استفاده از شبکه‌ی تصویری در گسترش پرس و جو با کمک تصاویر مشابه می‌تواند ایده‌ی مناسبی در رسیدن کاربر به هدف جست و جو باشد.

۴. نگاشت مفهوم به پرسش در جستجوی تصویر: با نگاشت تصویر پرس و جو به مفهوم آن و دنبال کردن پرس و جو به صورت متنی می‌توان نتایج کامل تری به کاربر ارائه نمود. بدین ترتیب ImageNet فاصله‌ی معنایی مفهوم و پرسش را نیز کم می‌کند.

شکل (۴) : تشخیص خودکار اشیا و پیدا کردن موقعیت آن ها

شکل (۵) : توصیف خودکار تصاویر به کمک شبکه ی تصویری ImageNet

بسیاری از مثال‌های گوناگون دیگر وجود دارد که نشان می‌دهند تشخیص صحیح و بلادرنگ از شرایط می‌تواند گره گشای بسیاری از مسائل امروز باشد. دنیایی که ربات‌ها همه کارهای لازم برای آسایش و رفاه شهروندان انجام می‌دهند.

۴- ضرورت‌های بومی‌سازی پروژه ImageNet

بومی‌سازی تکنولوژی‌های موجود در دنیا همواره یکی از مهمترین دغدغه‌های جوامع در حال توسعه می‌باشد. بومی‌سازی این پروژه‌ها و تکنولوژی‌ها می‌تواند باعث کاستن وابستگی به جهان و خودکفایی بیشتر گردد. با شرحی که از پروژه ImageNet مطرح شد برای بومی‌سازی این پروژه می‌توان دلایل و مزایای دیگری نیز ذکر نمود. همانطور که مطرح شد یکی از جنبه‌های استفاده از شبکه‌ی تصویری کاربرد آن در جویشگرهای تصویری به منظور درک و تشخیص بهتر مفاهیم مورد پرس و جو می‌باشد. از جهت دیگر یکی از اهداف تولید جویشگر بومی‌توسعه‌ی محتوای بومی‌و اشاعه‌ی فرهنگ ایرانی اسلامی‌می‌باشد. با توجه به اینکه شبکه‌ی تصویری به عنوان یک زیرساخت داده‌ای مدنظر قرار می‌گیرد و ImageNet بر اساس شبکه‌ی کلمات انگلیسی وردنت ساخته شده است، برای کاربرد در جویشگر بومی نیاز به استفاده از خط و زبان فارسی و داده‌های مرتبط با فرهنگ ایرانی اسلامی حس می‌شود. در واقع علاوه بر مستقل شدن از پروژه اصلی نیازمند پیاده سازی این بستر داده‌ای برای فرهنگ ایرانی اسلامی و زبان فارسی هستیم. در فرهنگ ایرانی اسلامی‌بسیاری از مفاهیم و اشیا وجود دارند که به هیچ وجه در پروژه اصلی ImageNet به کاربرده نشده است. به عنوان مثال مفاهیمی مانند کرسی، سجاده، هفت سین، سمنو و هزاران مورد دیگر هستند که ضرورت دارند در این پروژه گنجانده شوند.

شاید این سوال مطرح شود که چه نیازی به گنجاندن مفاهیم و واژگان فارسی است. اگرچه برخی عامل‌های نرم‌افزاری مثل دوربین‌هایی که سارق را تشخیص می‌دهند نیازی به بومی‌سازی ندارند. چون سارق در همه فرهنگ‌ها یک مفهوم است، اما مثال‌های دیگری هستند که نیاز به بومی‌سازی دارند. به عنوان مثال به مورد کمک به نابینایان رجوع می‌کنیم. یک نابینا در ایران دنیا را طوری می‌خواهد ببیند که یک ایرانی میبیند. باید عامل نرم‌افزاری بتواند نماز خواندن را درک کند و آن را برای نابینا توصیف کند. باید بتواند مراسم از آتش پریدن چهارشنبه سوری را با واقعه آتش سوزی تفکیک کند. باید بتواند سفره هفت سین را نسبت به سفره شام تشخیص دهد. و هزاران مورد دیگر که مختص فرهنگ ایرانی اسلامی‌می‌باشد. علاوه بر این جویشگرهای بومی‌نیاز به یک زیر ساخت بومی‌مناسب داده‌ای برای اجرا دارند. نمی‌توانیم ادعای تولید جویشگر بومی را داشته باشیم در حالی که زیر ساخت‌های آن از جمله شبکه تصویری آن غیر بومی باشد. مثال‌ها در این حوزه بسیار هستند. بنابراین نیاز به طرحی عامل‌های نرم‌افزاری داریم که به درستی با فرهنگ و باورهای ما مطابقت داشته باشد.

۵- نتیجه گیری

به منظور درک صحیح و تشخیص درست محیط پیرامون توسط ماشین‌ها علم بینایی ماشین مجاب به استفاده از یک پایگاه داده‌ی سلسله‌مراتبی مفهومی‌به نام شبکه‌ی تصویری می‌شود. در این مقاله به معرفی شبکه‌ی تصویری ImageNet پرداخته شده است و کاربردهای آن مورد بررسی قرار گرفته است. ایده‌ی استفاده از شبکه‌ی تصویری به عنوان زیرساخت داده‌ای در جویشگر تصویری و دیگر کاربردهای بومی‌در حوزه بینایی ماشین نیازمند بومی‌سازی این شبکه‌ی تصویر مبتنی بر مفاهیم و فرهنگ ایرانی اسلامی می‌باشد.

مراجع

[1] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li and L. Fei-Fei,"*ImageNet: A Large-Scale Hierarchical Image Database*", CVPR, 2009.

[2] L. Fei-Fei, O. Russakovsky, "*Analysis of Large-Scale Visual Recognition*", Bay Area Vision Meeting, October, 2013.

[3] L. Fei-Fei, "*ImageNet: crowdsourcing, benchmarking & other cool things*", CMU VASC Seminar, March, 2010,

[4] C. Fellbaum, *WordNet: An Electronic Lexical Database*, MIT Press, 1998.

[5] http://image-net.org/download

[6] O. Russakovsky, J. Deng, H. Su, J. Krause, S .Satheesh, S. Ma, Z. Huang, A.Karpathy, A. Khosla, M. Bernstein, A. C. Berg, L.Fei-Fei, "*ImageNet Large Scale Visual Recognition Challenge*", arXiv:1409.0575, 2014.

[7] J. Deng, O. Russakovsky, J. Krause, M. Bernstein, A. Berg, L. Fei-Fei, "*Scalable multi-label annotation*". ACM conference on human factors in computing (CHI), 2014.

[8] L. V. Ahn, L. Dabbish,"*Labeling images with a computer game*", CHI04, pp. 319-326, 2004.

[9] B. Russell, A. Torralba, K. Murphy, W. Freeman, "*Labelme: A database and web-based tool for image annotation*", IJCV, pp.157-173, May 2008.

[10] B. Yao, X. Yang, S. Zhu, "*Introduction to a large-scale general purpose ground truth database: Methodology, annotation tool and benchmarks*", EMMCVPR07, pp. 169-183, 2007.

[11] A. Torralba, R. Fergus, W. Freeman,"*80 million tiny images: A large data set for nonparametric object and scene recognition*", PAMI, Vol. 30, No. 11, pp. 1958-1970, November 2008.

[12] X. Glorot, Y. Bengio, "*Understanding the difficulty of training deep feedforward neural networks*", Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS), 2010.

[13] J. Deng, A. C. Berg, K. Li, L. Fei-Fei, "*What Does Classifying More Than 10,000 Image Categories Tell Us?*", ECCV 2010, Part V, LNCS 6315, pp. 71–84, 2010.

[14] T. Deselaers, V. Ferrari, "*Visual and Semantic Similarity in ImageNet*", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1777-1784, 2011.

[15] J. MARKOFF, "*Seeking a better way to find web images*", The New York Times, Nov, 19, 2012.

زیرنویس‌ها

۱ Machine Vision

۲ Deep Learning

۳ Fei-Fei Li

۴ Kai Li

۵ Crowd Sourcing

۶ Amazon Mechanical Turk

۷ Wordnet

۸ Big Data

۹ Machine Learning

۱۰ Smart Visioning

۱۱ Query