

توانایی مسیر پیش‌رو قشر بینایی در مقابله با تغییرات در بازشناسی اشیاء: آزمایش‌های انسانی و مدل محاسباتی سازگار با قشر بینایی

حمید کریمی روزبهانی^۱، رضا ابراهیم‌پور^۲ و منصور باقری^۳

چکیده

در این مقاله بازشناسی اشیاء در انسان در مواجهه با تغییرات مختلف بصورت کمی مورد ارزیابی قرار می‌گیرد. در سالیان اخیر سامانه بینایی انسان به جهت سرعت و دقت بسیار بالا در بازشناسی اشیاء به عنوان الگویی برای توسعه بسیاری از الگوریتم‌های محاسباتی در حوزه بینایی ماشین بوده است. از این رو، شناخت هرچه بهتر این سامانه و کمی‌سازی رفتار آن در شرایط مختلف می‌تواند تا حد زیادی به مدل‌سازی بهتر آن کمک نماید. در این پژوهش، توانایی بازشناسی انسان زمانی که اشیاء دچار تغییر زاویه تابش نور، جهت‌گیری عمقی، اندازه و موقعیت می‌شوند مورد ارزیابی قرار گرفته است. برای این منظور دسته‌ای از تصاویر با کنترل هر یک از تغییرات مذکور ساخته و تحت وضعیت‌های خاص به سوژه‌ها انسانی و مدل محاسباتی قشر بینایی ارائه داده می‌شود. مدل محاسباتی به منظور کمی‌سازی تاثیر هر یک از تغییرات بر روی بازشناسی استفاده می‌شود. نتایج آزمایش‌های انسانی حاکی از پیچیدگی بالای جهت‌گیری اشیاء در بازشناسی است. مشخص می‌شود که افت کارایی انسان به دلیل عدم توانایی در کاهش درهم‌تنیدگی بازنمایی‌های اشیاء در جهت‌گیری‌های شدید است. به صورت کمی بازنمایی‌های حاصل از تغییر زاویه تابش نور، جهت‌گیری عمقی، اندازه و موقعیت با رفتن از فضای پیکسل به قشر بینایی به ترتیب ۰/۵۷، ۰/۳۳، ۰/۵۵ و ۰/۷۳ بهبود می‌یابند.

کلیدواژه‌ها

بازشناسی اشیاء، قشر بینایی انسان، شبکه عصبی کانولوشنی، بازنمایی مستقل از تغییرات

قابلیت با سرعت و دقت زیاد توام است [۳] تا [۵]. به نظر می‌رسد که پی بردن به سازوکارهای سامانه بینایی انسان در بازشناسی اشیاء راه حلی برای بسیاری از مسائل بینایی ماشین فراهم می‌کند [۶]. بر این اساس است که بسیاری از قدرتمندترین الگوریتم‌های محاسباتی بازشناسی اشیاء که تا به امروز توسعه یافته‌اند از سامانه بینایی انسان الهام گرفته‌اند [۷] تا [۱۰].

اگرچه چگونگی فراهم آمدن توانایی بازشناسی مستقل از تغییرات اشیاء همچنان مورد سوال است، عموم پژوهشگران علوم بینایی بر یافته‌های ذیل در خصوص سازوکار و ساختار سامانه بازشناسی اشیاء در انسان توافق نظر دارند [۱۱]:

۱ مقدمه

انسان و بسیاری از پستانداران در بازشناسی اشیاء در حضور تغییرات زیاد دارای قابلیت بسیار بالایی هستند [۱] و [۲]. این

این مقاله در دی‌ماه سال ۱۳۹۴ دریافت، در خردادماه سال ۹۵ بازنگری و در شهریورماه همان سال پذیرفته شد.

^۱ دانشگاه تربیت دبیر شهید رجایی، دانشکده مهندسی برق.

رایانامه: hkarimi@srutu.edu

^۲ دانشگاه تربیت دبیر شهید رجایی، دانشکده مهندسی کامپیوتر.

رایانامه: rebrahimpour@srutu.edu

^۳ دانشگاه تربیت دبیر شهید رجایی، دانشکده مهندسی برق.

رایانامه: na.bagheri@gmail.com

نویسنده مسئول: رضا ابراهیم‌پور

[۱۴]. این در حالی است که، در آزمایش‌هایی دیگر دیده شده که دقت بازشناسی اشیاء از سیگنال‌های مغزی تا حد زیادی نامتأثر از تغییراتی چون اندازه و موقعیت است [۵]. از طرف دیگر، در آزمایش‌های بازشناسی اشیاء که در سطح رفتاری انجام شده است نیز یافته‌های ضد و نقیضی در خصوص توانایی انسان در بازشناسی مستقل از تغییرات اشیاء ارائه شده است. نتایج برخی آزمایش‌ها نشان می‌دهد که کاهش اندازه شیء و دور شدن آن از مرکز میدان بینایی باعث کاهش دقت در بازشناسی می‌شود [۱۵]. همچنین نشان داده شده است که جهت‌گیری عمقی و زاویه تابش نور اثرات زیادی بر دقت بازشناسی اشیاء دارند [۱۶] و [۱۷]. با این حال، آزمایش‌هایی دیگر نشان می‌دهند که وارونگی و چرخش تصویر شیء باعث افت نرخ بازشناسی اشیاء در افراد نمی‌شود [۱۸].

بنظر می‌رسد فارغ از روش‌های متفاوت به خدمت گرفته شده در آزمایش‌های اشاره شده، منشا این نتیجه‌گیری‌های ضد و نقیض نبود یک چارچوب محاسباتی برای محاسبه میزان تأثیرگذاری تغییرات مختلف بر روی بازشناسی اشیاء در انسان باشد. برای مثال، در یک آزمایش ثبت سیگنال مغزی (به روش MEG⁷) که در آن به سوژه‌های انسانی تصاویری از اشیاء تحت تأثیر تغییرات اندازه و موقعیت نمایش داده شد، مشخص شد که در صورتی که سیگنال‌های مغزی افراد که حاصل از نمایش اشیاء است مورد بازشناسی قرار گیرد، دقت بازشناسی در مواجهه با تغییر موقعیت مکانی شیء بیش از زمانی که با تغییر اندازه شیء مواجه می‌شود، افت خواهد کرد [۱۹]. علیرغم نتیجه بدست آمده در [۱۹]، نمی‌توان در مورد تأثیر بیشتر موقعیت مکانی بر نرخ بازشناسی در قیاس با اندازه شیء اظهار نظر نمود چراکه میزان پیچیدگی هر کدام از آن تغییرات در دسته تصاویر ساخته شده از ابتدا مورد مقایسه قرار نگرفته است. به بیان دیگر، ممکن است که میزان تغییر اندازه اعمال شده بر روی اشیاء از ابتدا و در ساخت تصاویر، کمتر از میزان تغییر موقعیت بوده است. بنابراین، وجود یک چارچوب محاسباتی که بر اساس آن بتوان تغییرات مختلف را مقایسه نمود بنظر لازم می‌آید.

در این پژوهش با ساخت دسته تصاویری^۸ از اشیاء با تغییرات کنترل شده، انجام آزمایش‌های بازشناسی اشیاء با شرکت سوژه‌های انسانی و استفاده از مدل محاسباتی نوین الهام یافته از قشر بینایی انسان روشی محاسباتی برای بررسی تأثیر هر یک از تغییرات بر روی نرخ بازشناسی پیشنهاد می‌شود. ساخت دسته تصاویر اشیاء به منظور کنترل هر یک از تغییرات اشیاء و محیط به صورت پارامتریک و به صورت مستقل از هم انجام می‌شود. آزمایش‌های بازشناسی اشیاء که در سطح رفتاری طراحی شدند به جهت مقایسه کیفی تأثیر هر یک از تغییرات بر روی صحت بازشناسی صورت گرفتند. استفاده از مدل محاسباتی الهام یافته از

یکی از اصلی‌ترین عملکردهای مسیر شکمی قشر بینایی^۱ فراهم آوردن مصالحه‌ای بسیار ظریف بین حساسیت و عدم حساسیت به تغییرات محرک‌های بینایی بخصوص در سطح سلول‌های فروگیجگاهی^۲ است که حاصل آن توانایی بالا در انجام بسیاری از عملکردهای بازشناسی است.

سامانه بینایی انسان دارای ساختاری سلسله‌مراتبی است و بدین ترتیب در سلسله مراتب با تغییراتی همچون اندازه^۳، موقعیت^۴، جهت‌گیری عمقی^۵ و زاویه تابش نور^۶ مقابله می‌کند.

پردازش اولیه در سامانه بینایی به صورت پیش‌رو انجام می‌شود (در مسیر پیش‌رو بازشناسی صورت می‌گیرد و در صورت نیاز، سازوکارهایی همچون توجه درگیر می‌شوند تا به بازشناسی کمک نمایند).

در آزمایش‌ها مشخص شده است که انسان‌ها قادر هستند بازشناسی اشیاء را در تصویری که برای چند ده میلی‌ثانیه نمایش داده شده است، در کسری از ثانیه انجام دهند [۱]. دشواری مسئله بازشناسی اشیاء زمانی بهتر آشکار می‌شود که می‌دانیم تصاویر ایجاد شده از یک شیء بر روی شبکه تقریباً هیچ‌گاه مشابه با یکدیگر نیستند. این مسئله، که در اثر تغییر موقعیت بیننده در فضا، حرکت چشم‌ها و یا تغییر وضعیت شیء و یا محیط رخ می‌دهد. از جمله این تغییرات می‌توان به تغییر اندازه، موقعیت مکانی، چرخش در صفحه، جهت‌گیری عمقی، زاویه تابش نور و تغییر شکل شیء اشاره نمود.

با توجه به اینکه، این توانایی بالا بسیاری از متخصصین پردازش و بازشناسی تصویر را در توسعه الگوریتم‌های محاسباتی به الگوبرداری از سازوکارهای قشر بینایی انسان ترغیب نموده است، سوالی بدین شکل قابل طرح است: آیا انسان در بازشناسی اشیاء، از جنبه‌های مختلف، الگویی بهینه برای الگوریتم‌های محاسباتی محسوب می‌شود؟ به طور مشخص، انسان در مواجهه با هر یک از تغییرات اشیاء (به عنوان یکی از مهم‌ترین مسائل حل نشده در توسعه الگوریتم‌های هوشمند) تا چه حد دچار مشکل می‌شود؟ چرا؟ در این پژوهش، با ارائه یک روش محاسباتی، پاسخی کمی برای این سوال ارائه می‌شود. به این منظور، در ابتدا به بررسی توانایی انسان در بازشناسی اشیاء توسط انسان تحت تأثیر تغییرات مختلف پرداخته می‌شود.

در سطح سیگنال‌های عصبی، پیش‌تر نشان داده شده است که پاسخ نرون‌های قشر فروگیجگاهی تا حدود زیادی تحت تأثیر تغییرات موقعیت، جهت‌گیری و اندازه قرار می‌گیرند [۱۲] تا

¹ Ventral visual stream

² Inferotemporal

³ Size

⁴ Position

⁵ Pose

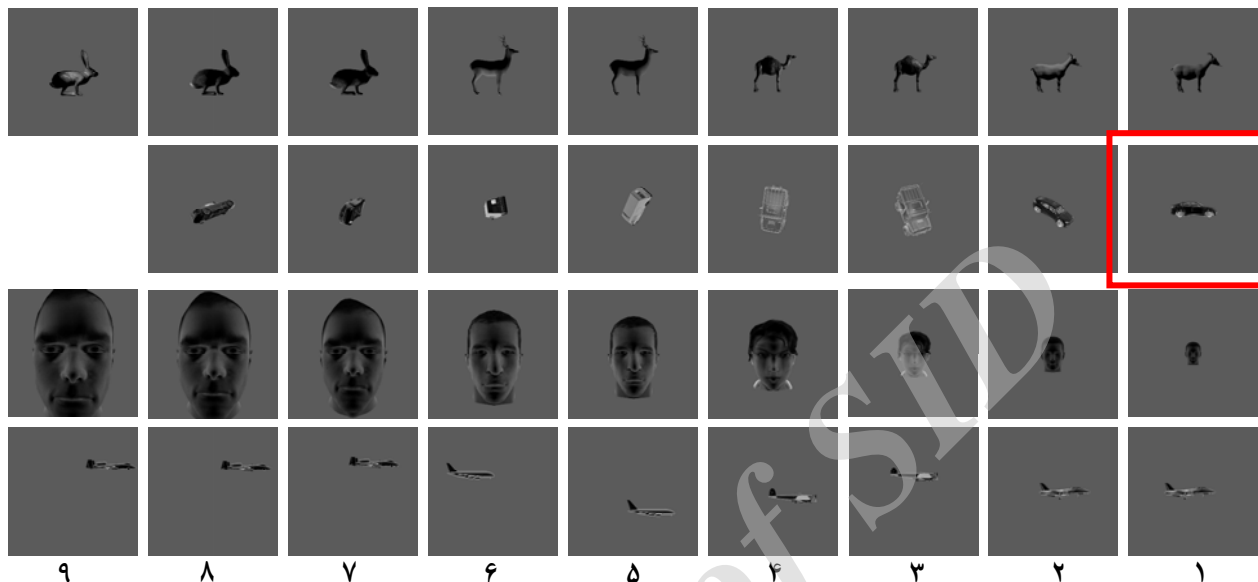
⁶ Lighting

⁷ Magneto-encephalography

⁸ Image Dataset

به‌طور مشخص، دسته تصاویر باید شامل تغییرات مورد مطالعه در این پژوهش باشد، هر تغییر باید بدون تاثیرگذاری بر روی تغییرات دیگر بر اشیاء اعمال شوند و اشیاء در تصاویر ساخته شده باید به طور پارامتریک در معرض تغییرات قرار گیرند. هیچ‌یک از دسته‌های تصاویر مذکور دارای چنین ویژگی‌هایی نیستند [۲۰] تا [۲۲]. علاوه بر این، در عموم آزمایش‌هایی که در خصوص بررسی بازشناسی اشیاء انجام شده است، محرک‌های بصری مورد استفاده

مغز نیز به عنوان جایگزینی برای سامانه بینایی انسان و در راستای توضیح نرخ بازشناسی رفتاری با استفاده از بازنمایی نرونی انجام شده است. روش پیشنهادی با کمی‌سازی تاثیر هر یک از تغییرات بر روی بازشناسی نه تنها بستری را برای مقایسه آن‌ها فراهم می‌آورد، بلکه می‌تواند به عنوان روشی برای توضیح پدیده‌های دیده شده در بازشناسی اشیاء در انسان با کمک بازنمایی اشیاء در سطح جمعیت نرونی باشد.



شکل ۱ نمونه‌هایی از تصاویر دوبعدی از اشیاء سه‌بعدی که تغییرات مختلف را متحمل شده‌اند. سطرها به ترتیب از بالا به پایین تغییر زاویه تابش نور، جهت‌گیری عمقی، اندازه و موقعیت مکانی را نشان می‌دهند. ستون‌ها (و اعداد نوشته شده در سطر آخر) به ترتیب از راست به چپ وضعیت‌های مختلف در هر تغییر را نشان می‌دهند. برای مثال سطر چهارم از بالا و ستون پنجم از راست نشان دهنده نمونه هواپیمای سوم تحت وضعیت پنجم تغییر (که از مرکز تصویر با شعاع بیشتری نسبت به وضعیت چهارم و کمتر از وضعیت ششم قرار گرفته‌اند) در موقعیت مکانی را نشان می‌دهد. همچنین، ستون هشتم از سطر دوم نشان دهنده نوع چهارم اتومبیل است که حول محورهای x ، y و z به میزان 315° درجه دوران داده شده است. تصویر داخل مربع نشان‌دهنده وضعیت خنثی است که در تغییرات اندازه و موقعیت نیز تکرار شده است (در این شکل نمایش داده نشده است).

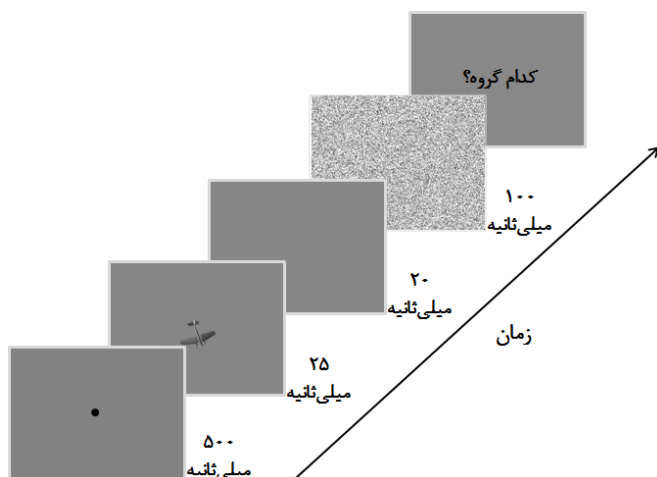
از دسته تصاویر طبیعی که در آن شیء مورد نظر حضور دارد استفاده نموده‌اند [۲۳] و [۲۴] و این مسئله بررسی کمی تاثیر هر یک از تغییرات را دچار خلل می‌کند. این مشکل به این دلیل رخ می‌دهد که بطور مثال نمی‌توان بیان کرد که تصویر شماره یک از شیء در قیاس با تصویر شماره دو از همان شیء به صورت کمی دچار چه میزان تغییر اندازه، جهت‌گیری عمقی، موقعیت و... شده است، چراکه تصاویر به طور کنترل شده گردآوری نشده‌اند. بنابراین، نیاز به طراحی تصاویری دیده می‌شود که در آنها از محرک‌های بصری با تغییرات کنترل شده بهره گرفته می‌شود [۲۴]. بر این اساس، در این پژوهش دسته‌ای از تصاویر شامل محرک‌های بصری با ویژگی‌های زیر ساخته شد. تصاویر از گروه‌هایی از مدل‌های سه‌بعدی از اشیاء شامل اتومبیل، هواپیما، حیوان و همچنین تصاویر چهره انسان (در ادبیات بازشناسی اشیاء تصاویر انسان، حیوان و چهره نیز در زمره اشیاء تلقی می‌شوند) تشکیل شده است (شکل ۱). در هر گروه چهار نمونه از اشیاء آن گروه قرار گرفتند. برای مثال گروه حیوان، شامل گوزن، خرگوش، شتر و بز و گروه اتومبیل شامل وانت، ون، سواری و اتومبیل

این مقاله بدین صورت مرتب شده است که در بخش دو روش بکار گرفته شده در راستای ساخت دسته تصاویر، انجام آزمایش‌های انسانی، شبیه‌سازی مدل قشر بینایی و معیارهای ارزیابی مدل بیان می‌شوند. در بخش سه نتایج آزمایش‌های انسانی و شبیه‌سازی‌های مدل محاسباتی ارائه می‌شوند و در بخش چهارم به جمع‌بندی نتایج پرداخته خواهد شد.

۲ روش

گام نخست در این پژوهش بررسی سطح کارایی افراد در بازشناسی در مواجهه با تغییرات مختلف اشیاء است. برای این منظور نیاز است تا دسته‌ای از تصاویر که اشیاء هدف را در خود دارند به صورت مصنوعی ساخته شوند. اگرچه دسته‌های تصاویر مختلفی وجود دارند که حالت‌های مختلف اشیاء سه بعدی را تحت تاثیر تغییرات ایجاد می‌کنند [۲۰] تا [۲۲]، اما با توجه به اینکه هر یک از این دسته تصاویر در راستای پژوهشی مشخص طراحی شده‌اند، فاقد ویژگی‌های مدنظر پژوهش جاری هستند.

روی شبکه افراد می‌نمود. در آزمایش بازشناسی طراحی شده، در هر آزمون بازشناسی، ابتدا برای مدت ۵۰۰ میلی‌ثانیه یک نقطه تمرکز به فرد نشان داده می‌شد که فرد موظف بود به آن توجه نماید. سپس یکی از تصاویر اشیاء که به صورت تصادفی از دسته تصاویر انتخاب شده بود برای بازشناسی توسط فرد برای مدت ۲۵ میلی‌ثانیه به نمایش درآمد و بعد از ۲۰ میلی‌ثانیه از اتمام نمایش آن (که در این فاصله صفحه نمایشگر مجدداً خاکستری می‌شد)، نقابی حاوی یک تصویر نویزی برای ۱۰۰ میلی‌ثانیه نمایش داده شد (شکل ۲).



شکل ۲ رویه انجام آزمایش‌های بازشناسی اشیاء در انسان.

پیش‌تر نشان داده شده است که با این رویه آزمایش تا حدود زیادی از خنثی شدن نقش مسیرهای بازخوردی اطمینان حاصل می‌شود [۲۶]. از فرد خواسته شده بود که پس از نمایش نقاب نویزی، با زدن یکی از چهار کلید از پیش تعریف‌شده بر روی صفحه‌کلید گروهی که شیء به آن تعلق داشت (یکی از چهار گروه حیوان، اتومبیل، هواپیما و چهره) را مشخص نماید. از افراد شرکت‌کننده خواسته شده بود که تا حد امکان سریع و دقیق پاسخ دهند. علاوه بر گروه شیء که فرد با زدن کلید مشخص می‌کرد، سرعت پاسخ افراد نیز ثبت می‌شد.

۱۲ شرکت‌کننده (۶ نفر مرد، با میانگین سنی ۲۳ سال) در آزمایش‌های بازشناسی اشیاء شرکت کردند. همگی شرکت‌کنندگان پیش از شروع آزمایش اصلی یک دور آموزش با تصاویر متفاوت جهت آشنایی با آزمایش را با حداقل نرخ بازشناسی ۸۰ درصد گذراندند. افراد شرکت‌کننده از نظر بینایی سالم و یا اصلاح شده (با استفاده از عینک) بودند. همگی شرکت‌کنندگان هر ۵۶۰ تصویر غیرتکراری موجود در دسته تصاویر را در ۴ بسته شامل ۱۴۰ تصویر مورد بازشناسی قرار دادند. بین هر کدام از بسته‌های تصویر ۵ دقیقه زمان استراحت پیش‌بینی شده بود که زمان انجام آزمایش را به طور میانگین به حدود ۳۰ دقیقه می‌رساند. آزمایش بازشناسی اشیاء با استفاده از جعبه‌ابزار Psychotool نرم‌افزار Matlab اجرا و نتایج با کمک آن ثبت شد.

اسپرت بود. بدین ترتیب ۴ گروه ۴ نمونه‌ای انتخاب و مدل‌های سه‌بعدی از آنها از اینترنت دانلود شد.

در گام بعد، توسط نرم‌افزار گرافیکی Blender تغییرات شیء، محیط و پس‌زمینه بر اشیاء اعمال شد. هرکدام از مدل‌های سه‌بعدی متحمل تغییرات در چهار بُعد شدند: تغییر در اندازه، موقعیت، جهت‌گیری عمقی و زاویه تابش نور. این تغییرات در هر یک از چهار بُعد در وضعیت‌های مختلف صورت پذیرفتند. در تغییر موقعیت، اشیاء از مرکز تصویر با شعاع‌هایی پلکانی-افزایشی در ۹ گام به سمت حاشیه‌های تصویر منتقل و در هر شعاع نمونه‌هایی از تصاویر شیء ساخته شدند. بدین ترتیب از یک وضعیت به وضعیت بعد شیء از مرکز تصویر دور شده و بر روی دویاری از پیش تعیین‌شده قرار می‌گرفت. در تغییر اندازه نیز اندازه شیء در ۹ اندازه از کوچک به بزرگ تغییر داده شد و تصاویر ثبت شد. ۸ جهت‌گیری عمقی مشتمل بر زوایای افزایشی از ۰ الی ۳۶۰ درجه در گام‌های ۴۵ درجه‌ای با چرخش شیء حول سه محور X، Y و Z در فضای سه‌بعدی انجام پذیرفت. تغییرات جهت تابش نور با تغییر موقعیت منبع نور محیط سه‌بعدی صورت گرفت. این موقعیت‌ها شامل چهار جهت اصلی بالا، پایین، راست و چپ و چهار جهت فرعی بالا راست، بالا چپ، پایین راست و پایین چپ و یک جهت مقابل است.

در مجموع، ۹ حالت مختلف برای جهت تابش نور، ۸ حالت جهت‌گیری عمقی، ۱۰ حالت موقعیت و اندازه در دسته تصاویر ساخته شده در نظر گرفته شدند (این تعداد با احتساب وضعیت خنثی برای تغییر اندازه و موقعیت است که در شکل ۱ نشان داده شده است). جزئیات بیشتر در خصوص دسته تصاویر در شکل ۱ ارائه شده است. در نهایت ۵۶۰ تصویر سطح خاکستری یکتا بر روی پس‌زمینه خاکستری از تغییرات و وضعیت‌های تغییرات مختلف ساخته شد که در آزمایش‌های بازشناسی اشیاء در انسان و مدل محاسباتی بکار رفت.

۲-۱ آزمایش‌های بازشناسی اشیاء در انسان

آزمایش‌های بازشناسی اشیاء در انسان که در این پژوهش مدنظر هستند توانایی بازشناسی اشیاء به صورت سریع را در انسان مورد ارزیابی قرار می‌دهند. با توجه به فرضیات اولیه حوزه بازشناسی اشیاء مبنی بر حل شدن بخش اصلی مسئله بازشناسی در مسیرهای پیش‌رو قشر بینایی، آزمایش‌های انسانی طوری طراحی شده‌اند که تا جای امکان توانایی مسیرهای پیش‌رو قشر بینایی انسان را مورد ارزیابی قرار دهند و نقش مسیرهای بازخوردی کمک‌کننده به بازشناسی اشیاء را خنثی نمایند [۲۶].

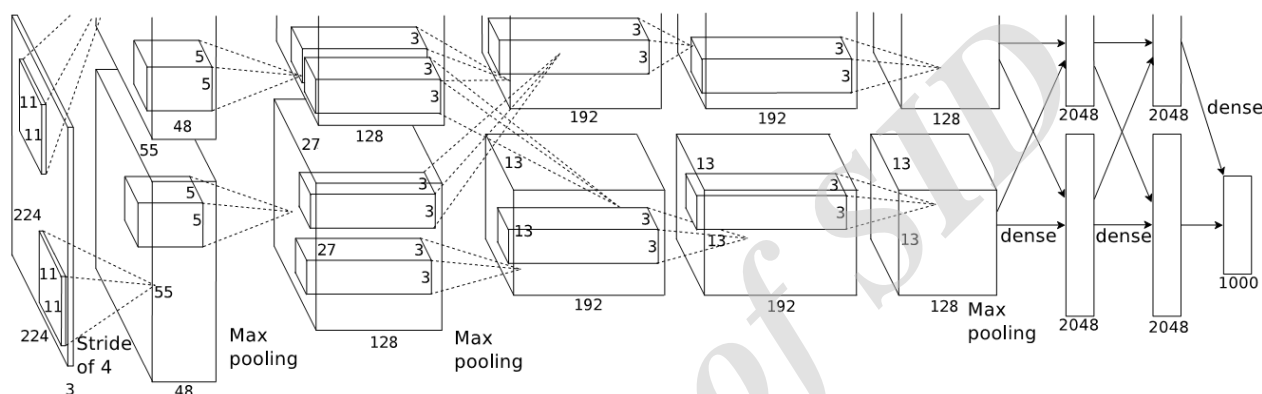
آزمایش بازشناسی اشیاء در اتافی نسبتاً تاریک و با نمایش تصاویر اشیاء بر روی نمایشگر با فرکانس ۸۰ هرتز انجام شد. اندازه تصاویر به نمایش درآمده ۲۵۶×۲۵۶ پیکسل بود که با احتساب فاصله شرکت‌کنندگان از نمایشگر (که ۶۰ سانتی‌متر انتخاب شد) تولید تصویری به اندازه حدود ۱۰ درجه بینایی بر

(مدل نرونی) و حدود ۶۰ میلیون پارامتر (وزن سیناپسی)، پنج لایه کانولوشنی و سه لایه با اتصالات کامل تشکیل شده است. هر کدام از عملیات مذکور بخشی از رفتارهای نرونی سامانه بینایی را شبیه‌سازی می‌کنند. برای مثال عمل فیلترینگ با اعمال فیلترهای گبور^۱ رفتار استخراج خط و لبه در سلول‌های ساده سامانه بینایی را شبیه‌سازی می‌کند. عمل آستانه‌گذاری و اشباع فعالیت بازنمایی کننده فعالیت نرونی، عملگر جمع‌آوری نماینده رفتار سلول‌های پیچیده و عملگر نرمال‌سازی بیانگر مهار جانبی بین نرون‌های لایه‌های مختلف سامانه بینایی است.

این مدل به روش باسرپرست کاهش شیب تصادفی^۲ آموزش داده می‌شود. در این پژوهش، نسخه پیاده‌سازی شده تحت نرم‌افزار

۲-۲ مدل محاسباتی قشر بینایی

نظر به اینکه ثبت فعالیت‌های مغزی در انسان عموماً با روش‌های غیرتهاجمی (EEG/MEG/FMRI) و از خارج مجموعه انجام می‌شوند، وجود محدودیت‌های زیادی از جمله عدم دقت مکانی مناسب (به معنی برهم‌افتادگی سیگنال‌های عصبی) و همچنین وجود نویز در سیگنال‌های ثبت‌شده اجتناب‌ناپذیر است. از این رو با توجه به انطباق بالای شبکه عصبی کانولوشنی با مسیر پیش‌رو قشر بینایی و همچنین در راستای حصول بازنمایی در سطح لایه‌های مسیر بینایی به بررسی بازنمایی‌های تولیدشده توسط این شبکه‌ها به‌عنوان مدلی محاسباتی از قشر بینایی پرداخته خواهد شد.



شکل ۳ ساختار شبکه عصبی کانولوشنی با عنوان مدلی از قشر بینایی انسان [۷].

Matlab و وزن‌هایی که پیش‌تر برای حل مسئله بازشناسی اشیاء ۱۰۰۰ دسته‌ای ILSVRC توسط مدل آموخته شده بودند دانلود و مورد استفاده قرار گرفتند [۲۸]. دسته تصاویر ILSVRC که زیرمجموعه‌ای از مجموعه بزرگ تصاویر ImageNet است شامل حدود ۱۰۰۰ تصویر برای هر یک از ۱۰۰۰ دسته موجود در آن مجموعه است که در مجموع تعداد تصاویر دسته تصاویر را به حدود یک میلیون تصویر می‌رساند [۲۹]. با توجه به تعداد کم نمونه‌های ساخته‌شده در دسته تصاویر پژوهش جاری و تعداد بسیار بالای پارامترهای آزاد مدل، در صورت آموزش مدل با دسته تصاویر جاری احتمال بیش‌برازش^۳ مدل بسیار بالا پیش‌بینی می‌شد. بعلاوه، با توجه به اینکه دسته‌های اشیاء موجود در دسته تصاویر این مقاله در دسته تصاویر ILSVRC نیز موجود بودند نیاز به آموزش مجدد به چشم نمی‌خورد. همچنین، مدل مذکور از سامانه بینایی که با تعداد گروه‌های بالا آموزش دیده است می‌تواند بهتر بازنمایی کننده رفتار سامانه بینایی انسان در دنیای نامحدود واقعی باشد. در این مدل، خروجی هر لایه پاسخی از جمعیت نرون‌هاست که حتی در بسیاری از روش‌های ثبت موجود مشابه آن در دسترس نیست. در این پژوهش با بررسی دقیق ۷ لایه ابتدایی

شبکه عصبی کانولوشنی که برای نخستین مرتبه در سال ۱۹۹۸ توسط یان لیکان معرفی شد یک شبکه عصبی مصنوعی مبتنی بر سامانه بینایی انسان است که ساختار سلسله‌مراتبی پیش‌رو خود را از مسیر شکمی بینایی انسان الهام گرفته است [۲۷]. انتخاب نام این شبکه (که در این مقاله به آن «مدل» خطاب می‌شود) بر اساس عملگر کانولوشن که در لایه‌های مدل برای ایجاد بازنمایی مستقل از تغییر موقعیت صورت می‌گیرد. ویژگی‌های سطح پایین تصاویر (از جمله خطوط جهت‌دار و لبه‌ها) در لایه‌های ابتدایی مدل با اعمال فیلترهایی که عملکرد سلول‌های نواحی پایین قشر بینایی را شبیه‌سازی می‌کنند، استخراج می‌شوند. این ویژگی‌ها با ترکیب در لایه‌های بعد منجر به ساخت ویژگی‌های پیچیده‌تر و مقاوم‌تر به تغییرات می‌شوند.

نسخه نوین این شبکه‌ها تحت عنوان ConvNet و یا AlexNet که در سال ۲۰۱۲، توسط گروه هینتون معرفی شد قابلیت‌های منحصر بفردی را نه تنها در دست‌یابی به کارایی بازشناسی قابل قیاس با سامانه بینایی انسان بلکه در خصوص پیش‌بینی نواحی بالای سامانه بینایی مانند ناحیه فروگیجگاهی از خود نشان داد [۷]. این مدل از ۸ لایه تشکیل شده است که بر روی هر تصویر ورودی به هر لایه آن به ترتیب عملیات فیلترینگ، آستانه‌گذاری و اشباع فعالیت، جمع‌آوری و نرمال‌سازی انجام می‌شود (شکل ۳). این مدل از حدود ۶۵۰۰۰۰ واحد محاسباتی

^۱ Gabor

^۲ Stochastic gradient descent

^۳ Overfitting

بینایی از یکدیگر فاصله گرفته و بنابراین جدایی‌پذیری گروه‌ها طی مسیر افزایش می‌یابد. با توجه به اینکه نقش اصلی مسیر پیش‌رو قشر بینایی جداسازی خوشه‌های بازنمایی اشیاء است، هر تغییر در میزان در هم تنیدگی خوشه‌های متناظر با هر گروه از اشیاء که در اثر تغییرات شیء اتفاق می‌افتد میزان پیچیدگی تحمیل شده به بازشناسی اشیاء را تعیین می‌کند. در این پژوهش از شاخص ارزیابی خوشه‌بندی دان^۳ جهت تعیین این درهم تنیدگی استفاده شده است. این شاخص به صورت رابطه (۲) تعریف و استفاده می‌شود [۳۲]:

$$D = \frac{\min_{1 \leq i \leq j \leq n} d(i, j)}{\max_{1 \leq k \leq n} d'(k)} \quad (2)$$

که در آن $d(i, j)$ فاصله خوشه متعلق به گروه i از خوشه متعلق به گروه j است و $d'(k)$ فاصله درون-گروهی گروه k را نشان می‌دهد. در اینجا در محاسبه شاخص دان، هر کدام از ۴ گروه اشیاء یکی از گروه‌ها در فرمول در نظر گرفته شده‌اند. فاصله بین-گروهی $d(i, j)$ می‌تواند با انواع روش‌های اندازه‌گیری فاصله محاسبه شود. در این پژوهش این فاصله، فاصله بین مراکز خوشه‌های گروه‌های i و j است. روش‌های متنوعی نیز برای محاسبه $d'(k)$ وجود دارد که در اینجا بیشترین فاصله بین تمام جفت بردارهای بازنمایی ممکن در هر گروه در نظر گرفته شده است.

از آنجا که افزایش و کاهش شاخص دان رابطه مستقیمی با افزایش و کاهش درهم‌تنیدگی خوشه‌های گروه‌ها دارد، و همچنین می‌توان آن را در هر ابعاد فضایی از جمله فضای پیکسل یا نرونی (ثبت حقیقی و یا مدل محاسباتی) اعمال نمود، این شاخص می‌تواند معیار مناسبی برای محاسبه اثر هر یک از تغییرات روی نرخ بازشناسی اشیاء تلقی شود. شاخص‌هایی دیگری از جمله شاخص دیویس-بولدین^۴ و ضریب سایه^۵ نیز در ارزیابی درهم‌تنیدگی خوشه‌ها وجود دارند، اما در آزمایش‌ها تفاوت معنی‌داری را با شاخص دان نشان ندادند.

در محاسبه شاخص دان، محاسبات در فضایی به ابعاد تعداد نرون‌های موجود در هر لایه مدل و یا پیکسل‌های تصویر انجام می‌شود. برای مثال، در لایه هفتم مدل که ۴۰۹۶ نرون وجود دارد مقدار شاخص در فضای ۴۰۹۶ بُعدی و در فضای ورودی (پیکسل) در ابعاد ۶۵۵۳۶ (۲۵۶×۲۵۶) محاسبه می‌شوند. گرچه در نهایت این مقادیر با توجه به یکسان بودن واحد قابل مقایسه هستند. بر همین اساس، برای محاسبه میزان تغییر شاخص دان از ورودی (فضای پیکسل) به خروجی لایه‌های مدل از شاخص پردازش رابطه (۳) استفاده می‌شود:

$$P_index = \frac{D_2 - D_1}{D_2 + D_1} \quad (3)$$

مدل (با توجه به شباهت بازنمایی این لایه‌ها با لایه‌های نواحی قشر بینایی و نزدیکی هر چه بیشتر آنها به رفتار انسان) توضیحی مستدل و کمی برای رفتارهای دیده‌شده در آزمایش‌ها حاصل می‌شود.

۲-۳ ارزیابی دقت بازشناسی در انسان و مدل محاسباتی و شاخص‌های خوشه‌بندی

به منظور محاسبه کارایی بازشناسی اشیاء در مدل محاسباتی، بازنمایی‌های متناظر با هر تصویر به یک طبقه‌بند پیشینه همبستگی^۱ اعمال و نرخ بازشناسی صحیح محاسبه شد. این طبقه‌بند به صورت رابطه (۱) تعریف می‌شود:

$$i^* = \arg \max_i [corr(x^*, \bar{x}_i)] \quad (1)$$

که x^* به ترتیب به بردار بازنمایی شیء آزمایشی و \bar{x}_i به میانگین بازنمایی بردارهای دسته آموزشی i اشاره می‌کنند. گروه بردار آزمایشی x^* ، i خواهد بود اگر x^* بیشترین شباهت را به بردار میانگین بازنمایی‌های دسته آموزشی i یعنی \bar{x}_i داشته باشد. $corr$ تابع همبستگی است که در این مقاله از همبستگی خطی پیرسن^۲ استفاده شده است [۳۰].

برای جلوگیری از اشباع و در نتیجه بی‌معنی شدن کارایی مدل در اندازه‌گیری نرخ طبقه‌بندی، تنها ۵۰٪ از نمونه‌های هر گروه به عنوان نمونه آموزشی و مابقی ۵۰٪ به عنوان نمونه آزمایشی به طبقه‌بند اعمال و نرخ طبقه‌بندی مورد محاسبه قرار گرفته است. انتخاب نمونه‌های آموزشی به صورت تصادفی از کل مجموعه داده هر گروه انجام شده و این رویه ۲۰۰ مرتبه تکرار شد تا نویز داده‌ها از میان رود. اگرچه نسبت ۵۰٪ برای نمونه‌های آموزشی-آزمایشی تمام توان مدل را به خدمت نمی‌گیرد، اما، با توجه به اینکه در این پژوهش هدف مقایسه کارایی مدل با انسان نیست خللی به تحلیل‌ها وارد نخواهد شد. در این پژوهش هدف استفاده از مدل محاسباتی در راستای توضیح رفتار بازشناسی انسان و به عنوان ابزاری برای دسترسی به لایه‌های مختلف پردازشی در سامانه بینایی است.

از دیدگاه نظری، اعمال هر یک از تغییرات بر روی شیء باعث تغییر فعالیت مجموعه نرون‌های تحریک‌پذیر از آن محرک در نواحی مختلف قشر بینایی می‌شود [۱۱]. در صورتی که میانگین فعالیت یک دسته از نرون‌ها در پاسخ به یک شیء در فضایی با ابعاد تعداد نرون‌های آن مجموعه با یک نقطه مشخص شود، تغییر وضعیت شیء در تصویر ورودی باعث بوجود آمدن نقاط متعدد و در نهایت شکل‌گیری یک خوشه برای بازنمایی‌های هر شیء و به همین شکل برای هر گروه از اشیاء در فضای جمعیت نرونی می‌شود [۳۱]. خوشه‌های حاصل از بازنمایی‌های اشیاء هر گروه با حرکت سیگنال‌های بصری طی مسیر پیش‌رو قشر

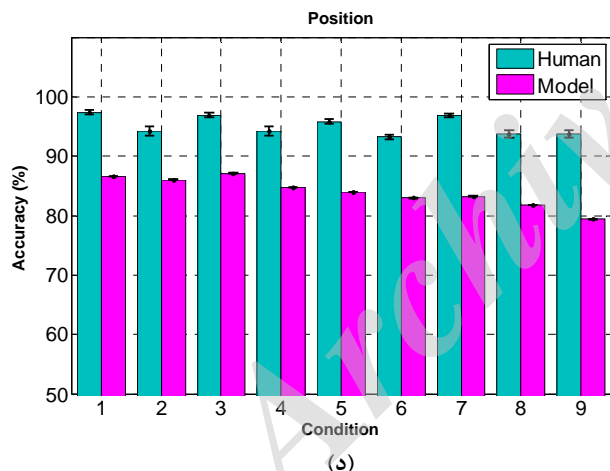
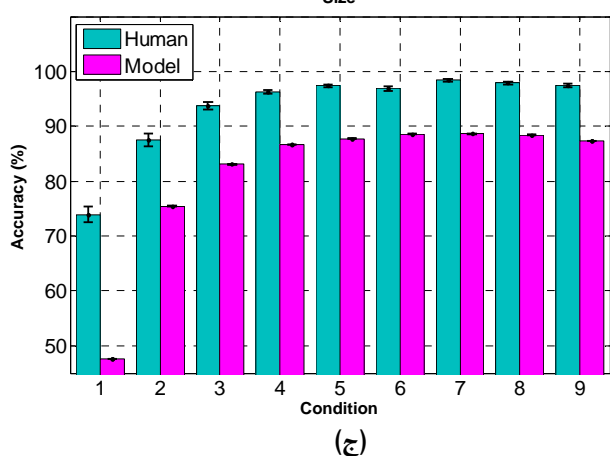
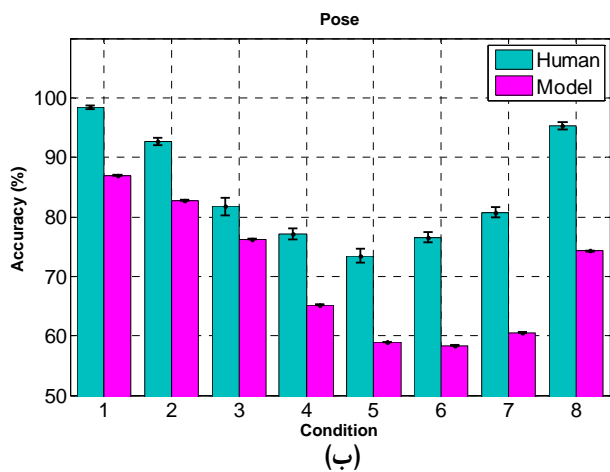
³ Dunn clustering index

⁴ Davies-Bouldin index

⁵ Silhouette coefficient

¹ Maximum correlation

² Pearson linear correlation



شکل ۴ میانگین کارایی بدست آمده در آزمایش‌های بازشناسی اشیاء در انسان و مدل محاسباتی تحت تاثیر تغییرات زاویه تابش نور (الف)، جهت‌گیری عمقی (ب)، اندازه (ج) و موقعیت (د). محور افقی وضعیت‌های متناظر با هر تغییر را بر اساس چیدمان شکل ۱ نمایش می‌دهد. میله‌های خطا، میزان خطای استاندارد را نشان می‌دهند.

در مدل محاسباتی، نرخ بازشناسی صحیح که روی ۲۰۰ اجرای مدل بدست آمده برابر ۸۰/۳۲٪ و برای تغییرات جهت تابش نور، جهت‌گیری عمقی، اندازه و موقعیت به ترتیب برابر ۸۴/۲۶٪، ۷۰/۵۶٪، ۸۱/۵۱٪ و ۸۳/۸۷٪ بدست آمد. در شکل ۴ می‌توان میزان کارایی مدل را در هر یک از تغییرات دید.

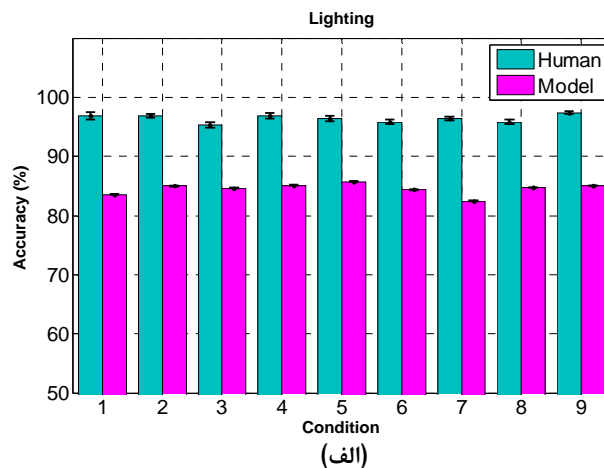
که D_1 و D_2 مقادیر شاخص دان به ترتیب در فضای اول (مثلاً پیکسل) و دوم (مثلاً مدل) و P_index شاخص پردازش را نشان می‌دهند. مقدار مثبت برای شاخص پردازش بیانگر بهبود بازنمایی از فضای اول به فضای دوم است.

۳ نتایج

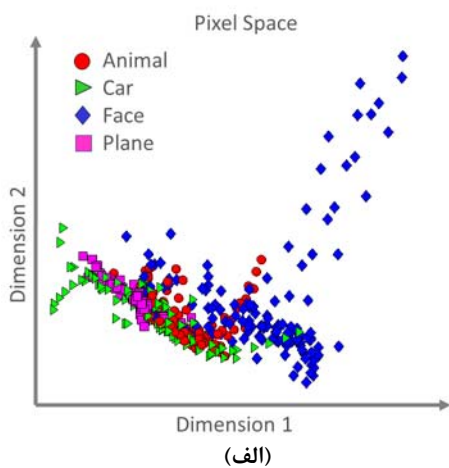
پاسخ و زمان پاسخ افراد در هر آزمون از آزمایش‌های بازشناسی اشیاء ثبت و نتایجی که در ادامه می‌آیند از میانگین‌گیری پاسخ‌ها روی تمام شرکت‌کنندگان تحت هر یک از وضعیت‌ها حاصل شده است. به‌طور کل شرکت‌کنندگان با کارایی بالایی آزمایش را انجام دادند. میانگین نرخ بازشناسی صحیح ۹۲/۵۵٪ (با انحراف از معیار ۷/۴۶٪) محاسبه شد. میانگین زمان پاسخ‌دهی افراد ۸۷۶ میلی‌ثانیه (با انحراف از معیار ۱۳۵ میلی‌ثانیه) بدست آمد. نرخ بازشناسی صحیح در مواجهه با تغییر جهت تابش نور، جهت‌گیری عمقی، اندازه و موقعیت به ترتیب برابر ۹۶/۴۱٪، ۸۴/۵٪، ۹۳/۲۸٪ و ۹۵/۱۴٪ و زمان‌های پاسخ‌دهی به ترتیب برابر با ۸۱۵، ۹۵۱، ۹۲۰ و ۸۲۹ میلی‌ثانیه بدست آمد.

در شکل ۴ نتایج بازشناسی اشیاء در مواجهه با هر یک از تغییرات با جزئیات بیشتر نمایش داده شده است. چنانکه مشاهده می‌شود تغییرات زاویه تابش نور در هیچ‌یک از وضعیت‌های این تغییر باعث تفاوتی معنی‌دار با وضعیت دیگر نشده است (بر اساس آزمون رتبه ویلکاکسون^۱، $P > 0/1$). این در حالی است که وضعیت‌های ۳ الی ۷ در تغییر جهت‌گیری عمقی، وضعیت‌های ۱ و ۲ در تغییر اندازه و وضعیت ۶ در تغییر موقعیت تفاوتی معنی‌دار با دیگر وضعیت‌های تغییر نشان می‌دهد (آزمون رتبه ویلکاکسون، $P < 0/05$).

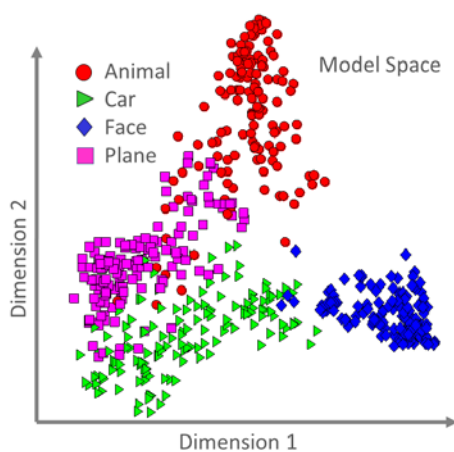
در مجموع، نتایج آزمایش‌های بازشناسی اشیاء در انسان در سطح رفتار نشان می‌دهد که تغییر زاویه تابش نور تاثیر زیادی در کارایی بازشناسی را باعث نمی‌شود. این در حالی است که، چرخش شیء در عمق (تغییر جهت‌گیری عمقی)، کوچک شدن اندازه شیء و دور شدن آن از مرکز تصویر باعث کاهش کارایی بازشناسی اشیاء در انسان می‌شود.



¹ Wilcoxon signed-rank test



(الف)

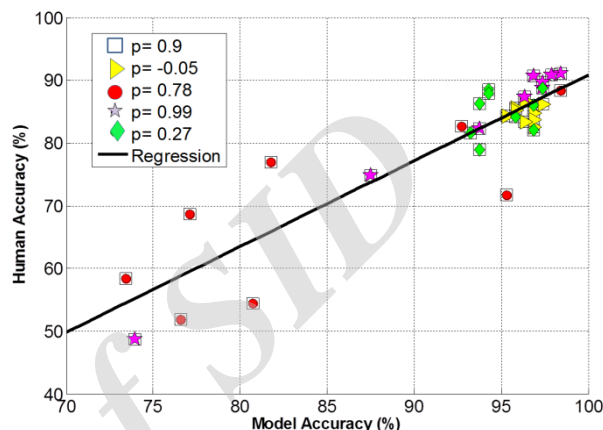


(ب)

شکل ۶ دیاگرام MDS تصاویر اشیاء در فضای دو بُعدی پیکسل (الف)، و در فضای خروجی مدل (ب) به عنوان مدلی از نواحی انتهایی سامانه بینایی انسان. در دیاگرام‌های MDS فضای اصلی به فضای دو بُعدی نگاشت می‌شوند تا یک شمای شهودی از نحوه پراکندگی نمونه‌ها حاصل نمایند. محورهای عمودی و افقی ابعاد اول و دوم فضای MDS را نشان می‌دهند. با توجه به اینکه مقادیر این محورها اهمیتی ندارند از نمایش آنها صرف‌نظر شده است. شکل‌های دایره، مثلث، لوزی و مربع به ترتیب نشان‌دهنده گروه‌های حیوان، اتومبیل، چهره و هواپیما هستند.

همانطور که از شکل ۶ مشخص است، پراکندگی نمونه‌های گروه‌های مختلف اشیاء دچار درهم‌تنیدگی زیادی است که با گذر از لایه‌های مدل، در لایه آخر به صورت خوشه‌هایی درمی‌آیند که تا حدود زیادی به صورت خطی از یکدیگر جداپذیر هستند. شاخص دان برای پراکندگی در فضای ورودی و خروجی مدل به ترتیب برابر با ۰/۰۶۲ و ۰/۳۳۳ هستند. به همین شکل MDS هر یک از تغییرات به تفکیک در شکل ۷ گردآوری شده‌اند.

چنانکه مشاهده می‌شود بازشناسی اشیاء در انسان و مدل محاسباتی همبستگی بالایی را نشان می‌دهد (همبستگی پیرسن برابر ۰/۹۱، با ضریب اهمیت^۱ $p < ۰/۰۵$). همبستگی انسان و مدل محاسباتی در هر یک از تغییرات در شکل ۵ نشان داده شده است. بنابراین، مدل محاسباتی انتخاب شده، همان‌طور که پیش‌تر هم انتظار آن می‌رفت، تا حد بالایی رفتار مسیر پیش‌رو قشر بینایی را شبیه‌سازی می‌کند. در ادامه برای توضیح رفتار دیده شده در آزمایش‌های بازشناسی اشیاء در انسان به بررسی کمی بازنمایی‌های لایه‌های مدل محاسباتی به عنوان مدلی از مسیر پیش‌رو قشر بینایی با کمک شاخص دان پرداخته می‌شود.

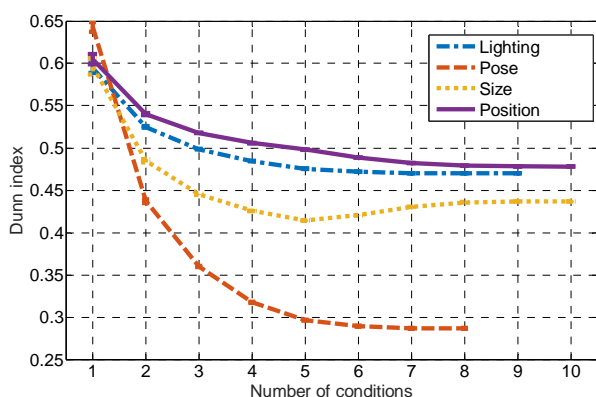


شکل ۵ همبستگی بالای بین بازشناسی اشیاء در مدل محاسباتی و انسان. محور افقی دقت بازشناسی اشیاء در مدل محاسباتی و محور عمودی در انسان را نشان می‌دهد. هر کدام از نقاط بیانگر یکی از وضعیت‌های دسته تصاویر است. به ترتیب شکل‌های مثلث، دایره، ستاره و لوزی بیانگر تغییرات اشیاء تحت تاثیر تغییر زاویه تابش نور، جهت‌گیری عمقی، اندازه و موقعیت هستند. مقادیر همبستگی برای هر تغییر در راهنمای شکل آمده است. مقدار همبستگی انسان و مدل در مجموع نیز با مربع توخالی برابر با ۰/۹ نشان داده شده است. خط تیره نشان‌دهنده روند همبستگی مجموع است که از برازش نمودن یک معادله خط با آن بدست آمده است.

در راستای بوجود آوردن حسی شهودی از بازنمایی‌های حاصل از مدل محاسباتی از روش Multi-Dimensional Scaling (MDS) استفاده شده است. روش MDS که بر اساس یک نگاشت تقریبی توسعه می‌یابد، با کاهش یک تابع هزینه، تلاش می‌کند که فاصله نسبی نقاط در فضای با ابعاد جدید حدوداً برابر با فضای اصلی باشد. در این پژوهش، نگاشتی از فضای بازنمایی با تعداد نرون‌های لایه مورد نظر از مدل محاسباتی به فضای با ابعاد دو بُعدی فراهم آمده است. در شکل ۶ دیاگرام‌های MDS بازنمایی در فضای ورودی (پیکسل) و خروجی (لایه هفتم مدل) ارائه شده است.

¹ Significance

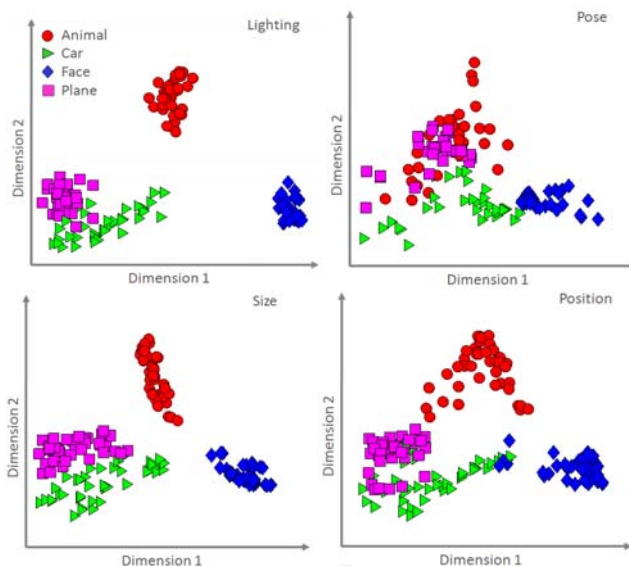
یکی از نکاتی که ممکن است نتایج گردآوری شده در جدول ۱ را تحت تاثیر قرار داده باشد نابرابری تعداد وضعیت‌ها در تغییرات مختلف است (در زاویه تابش نور ۹، در جهت‌گیری عمقی ۸ و در دو تغییر دیگر ۱۰ وضعیت). همچنین، برای اینکه بتوان حدس زد که رفتار شاخص دان (و یا به‌طور کل پراکندگی نمونه‌ها در فضا) تا چه حد تحت تاثیر تعداد نمونه‌های موجود در هر گروه قرار می‌گیرد، در شکل ۸ تاثیر تعداد نمونه‌ها با تغییر در تعداد وضعیت‌ها در مقدار شاخص دان نشان داده شده است. برای دستیابی به این شکل شاخص دان در هر یک از تغییرات برای تعداد وضعیت‌های مختلف محاسبه شد. برای مثال برای محاسبه مقدار شاخص دان در تغییر جهت‌گیری عمقی در تعداد وضعیت‌های ۴، شاخص دان برای تمامی ترکیب‌های ۴ از ۸ وضعیت برای تغییر مذکور محاسبه شده است.



شکل ۸ تاثیر تعداد وضعیت‌های در هر تغییر بر شاخص دان. محور افقی تعداد وضعیت‌های هر تغییر از کل وضعیت‌های موجود در آن تغییر (به ترتیب ۹، ۸، ۱۰ و ۱۰ وضعیت برای زاویه تابش نور، جهت‌گیری عمقی، اندازه و موقعیت) را نشان می‌دهد. میله‌های خطا، میزان خطا استاندارد را نشان می‌دهند.

همانطور که شکل ۸ نشان می‌دهد زمانی که تعداد وضعیت‌های هر تغییر ۱ است شاخص دان برای تغییر جهت‌گیری عمقی و زاویه تابش نور بیشتر از دیگر تغییرات است که این برتری با افزایش تعداد وضعیت‌ها با روندی نزولی کاهش پیدا می‌کند تا در نهایت در تعداد ۴ به حالتی پایدار می‌رسد. بنابراین، با توجه به تعداد وضعیت‌های بیش از ۸ در این دسته تصاویر، این فرضیه که تفاوت تعداد وضعیت‌ها در نتایج دخیل بوده است رد می‌شود. همچنین با توجه به عدم تغییر شیب نمودارها از تعداد ۵ وضعیت به بعد تا حدود زیادی می‌توان از کفایت تعداد وضعیت‌های هر تغییر در دسته تصاویر اطمینان حاصل نمود.

با مشخص شدن میزان پیچیدگی هر یکی از تغییرات اشیاء، یکی از سوال‌هایی که پیش می‌آید ترتیب و اولویت جبران‌سازی تغییرات مختلف اشیاء در مسیر پیش‌رو قشر بینایی است. برای پاسخ به این سوال مقدار شاخص دان برای خروجی هر یک از لایه‌های مدل و هر یک از تغییرات محاسبه شد. شکل ۹ روند



شکل ۷ دیاگرام MDS تصاویر اشیاء در فضای دوبعدی خروجی مدل به تفکیک هر یک از تغییرات. شکل‌های بالا چپ، بالا راست، پایین چپ و پایین راست به ترتیب پراکندگی بازنمایی اشیاء را در فضا تحت تاثیر تغییر در زاویه تابش نور، جهت‌گیری عمقی، اندازه و موقعیت نشان می‌دهند.

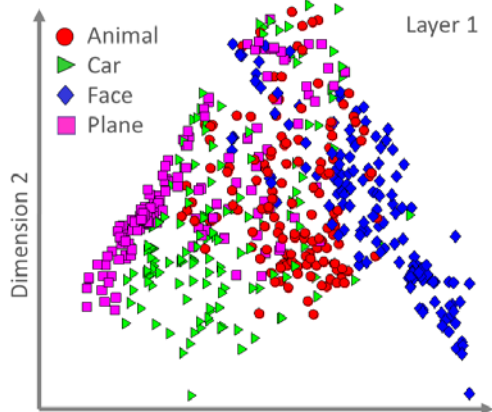
شکل ۷ نشان می‌دهد که تغییر در زاویه تابش نور باعث درهم تنیدگی اندکی در گروه‌های اشیاء شده است. این در حالی است که تغییر جهت‌گیری عمقی و اندازه باعث درهم فرو رفتن بازنمایی‌های اشیاء شده است. جدول ۱ مقادیر شاخص دان را در ورودی، خروجی مدل و میزان شاخص پردازش بین ورودی و خروجی را نشان می‌دهد.

جدول ۱ مقادیر شاخص دان در ورودی، خروجی مدل و شاخص پردازش از فضای ورودی به خروجی مدل.

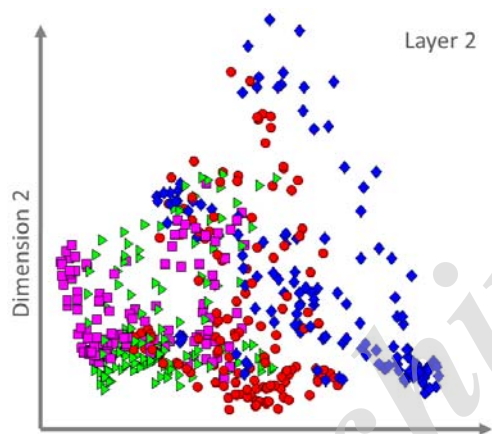
تغییر مورد بررسی	شاخص دان در فضای ورودی	شاخص دان در فضای خروجی	شاخص پردازش
زاویه تابش نور	۰/۱۲۳	۰/۴۵۸	۰/۵۷۶
جهت‌گیری عمقی	۰/۱۲۴	۰/۲۴۶	۰/۳۳۱
اندازه	۰/۱۱۳	۰/۳۹۶	۰/۵۵۷
موقعیت	۰/۰۷۴	۰/۴۷۷	۰/۷۳۱
همه تغییرات	۰/۰۶۲	۰/۳۳۳	۰/۶۸۷

با توجه به جدول ۱ مشخص است که تغییر در جهت‌گیری عمقی با شاخص پردازش ۰/۳۳۱ مشکل‌ترین تغییری است که مسیر پیش‌رو قشر بینایی با آن مواجه می‌شود تغییر در اندازه، زاویه تابش نور و موقعیت به ترتیب با شاخص پردازش ۰/۵۵۷، ۰/۵۷۶ و ۰/۷۳۱ مشکلات کم و کمتری را برای آن سامانه ایجاد نموده‌اند.

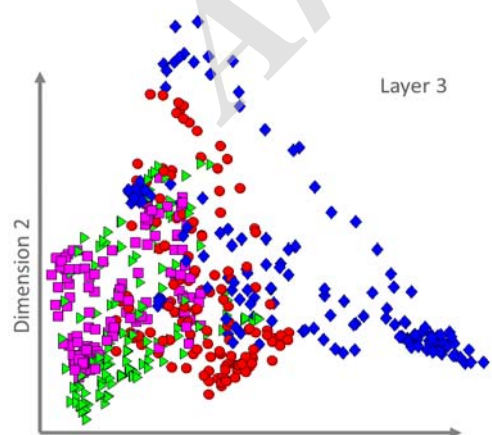
وجود قسمت منحصربه‌فرد پردازش چهره در ساختار سامانه بینایی انسان باشد [۳۱]. همچنین نزدیکی خوشه‌های اتومبیل و هواپیما از لایه نخست تا انتهای سلسله‌مراتب می‌تواند شاهدهی به نفع وجود سازوکار استفاده از ویژگی‌های سطح پایین در سامانه بینایی انسان باشد [۳۲].



Dimension 2
Dimension 1 (الف)

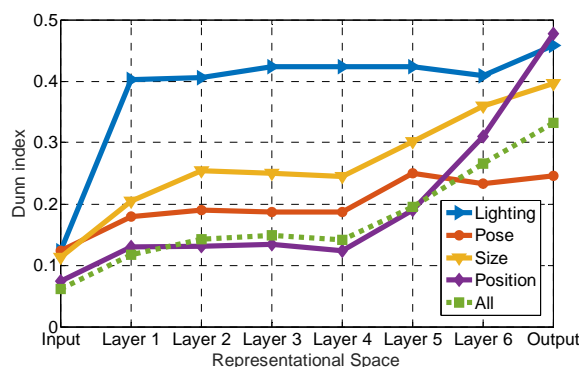


Dimension 2
Dimension 1 (ب)



Dimension 2
Dimension 1 (ج)

کاهش درهم‌تنیدگی خوشه‌های بازنمایی اشیاء را با عبور از لایه‌های مختلف مدل نشان می‌دهد.



شکل ۹ روند افزایش شاخص دان با عبور بازنمایی‌های اشیاء از لایه‌های مدل. محور افقی فضاهایی را که شاخص دان در آنها محاسبه شده است نشان می‌دهد. این فضاها از چپ به راست به ترتیب فضای پیکسل و خروجی هر یک از لایه‌های مدل را نشان می‌دهند.

چنان‌که مشخص است شاخص دان طی مسیر پیش‌رو برای بیشتر تغییرات روندی افزایشی دارد. نکته جالب این است که جبران تغییر زاویه تابش نور تا حد زیادی توسط لایه اول مدل و دیگر تغییرات عموماً در لایه‌های انتهایی مدل (۵ تا ۷) جبران می‌شوند. این درحالی است که لایه‌های میانی ۱ الی ۴ در قیاس با دیگر لایه‌ها نقش چندانی در کاهش درهم‌تنیدگی بازنمایی‌ها ندارند. جدول ۲ مقدار پردازش انجام شده در هر لایه را نسبت به لایه پیشین آن را بر اساس شاخص پردازش نشان می‌دهد. در این میان بیشترین شاخص پردازش مربوط به خروجی لایه اول سپس به لایه‌های پنجم و ششم و کمترین پردازش مربوط به لایه چهارم است.

جدول ۲ مقادیر شاخص پردازش در هر یک از لایه‌های مدل که نسبت به لایه پیشین محاسبه شده است.

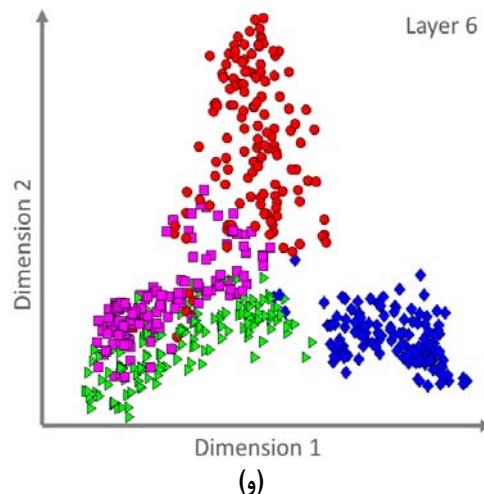
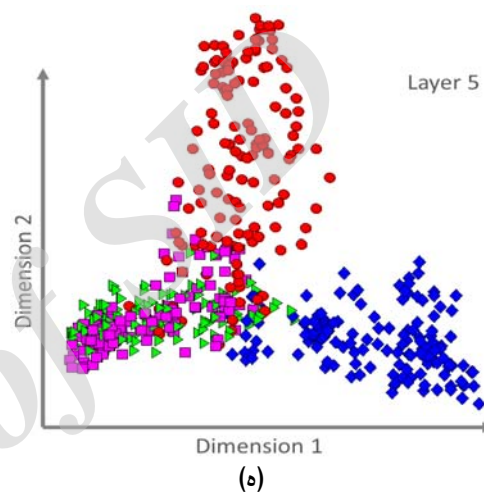
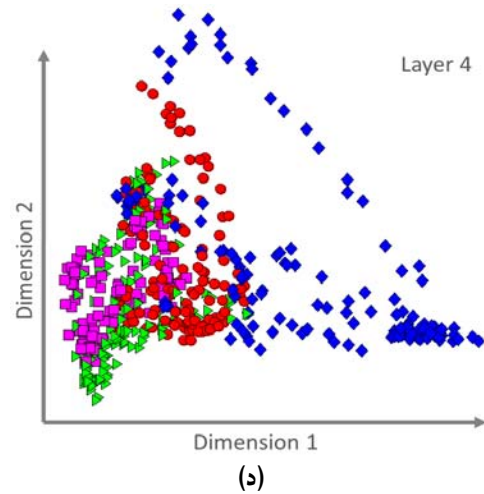
تغییر فضا	زاویه تابش	جهت‌گیری عمقی	اندازه	موقعیت	همه تغییرات
لایه ۱	۰/۵۳۱	۰/۱۸۴	۰/۲۹۰	۰/۲۷۲	۰/۳۱۳
لایه ۲	۰/۰۰۵	۰/۰۲۷	۰/۱۰۷	۰/۰۰۷	۰/۰۹۷
لایه ۳	۰/۰۲۱	-۰/۰۰۷	-۰/۰۰۸	۰/۰۱۱	۰/۰۲۲
لایه ۴	~۰	~۰	-۰/۰۱۰	-۰/۰۴۲	-۰/۰۲۵
لایه ۵	~۰	۰/۱۴۴	۰/۱۰۳	۰/۲۱۰	۰/۱۵۷
لایه ۶	-۰/۰۱۸	-۰/۰۳۶	۰/۰۸۸	۰/۲۴۱	۰/۱۵۳
لایه ۷	۰/۰۵۷	۰/۰۲۸	۰/۰۴۹	۰/۱۵۳	۰/۱۱۲

روند جداسازی خوشه‌های اشیاء با گذشت از لایه‌های مدل در شکل ۱۰ نمایش داده شده است. مشخصاً در ابتدا خوشه‌ها بسیار درهم‌تنیده و پس از گذر از لایه‌های مدل بسیار جداپذیرتر می‌گردند. از نکات جالب در پراکندگی نمونه‌ها می‌توان به تراکم بالای تصاویر چهره و نزدیکی خوشه‌های مربوط به اتومبیل و هواپیما اشاره نمود. نزدیکی نمونه‌های چهره می‌تواند توجیه‌کننده

چهار تغییر مدنظر قرار گرفتند: تغییر در زاویه تابش نور به اشیاء، جهت‌گیری عمقی، اندازه و موقعیت اشیاء. برای بررسی تاثیر این تغییرات دسته‌ای از تصاویر شامل گروه‌های منتخب از اشیاء ساخته شدند. به‌منظور مشاهده رفتار انسان در بازشناسی اشیاء، آزمایش‌هایی انسانی ترتیب داده شد و توانایی سوژه‌های انسانی در مواجهه با هر یک از تغییرات مورد ارزیابی قرار گرفت. در این آزمایش‌ها نشان داده شد که کارایی بازشناسی اشیاء زمانی که اشیاء دچار چرخش در عمق می‌شوند دچار افتی معنی‌دار می‌شود. همچنین، مشخص شد که دور شدن شیء از مرکز تصویر و کوچک نمودن اندازه آن کاهش نرخ بازشناسی را به دنبال خواهند داشت. این در حالی بود که تغییر در زاویه تابش نور تاثیر چندانی در دقت بازشناسی انسان نشان نداد.

به‌منظور توضیح رفتارهای دیده‌شده در آزمایش‌ها، از مدل محاسباتی قشر بینایی در شبیه‌سازی بهره برده شد. ابتدا نشان داده شده که این مدل تا حد بسیار بالایی رفتار انسان را تقلید می‌کند. سپس، از بازنمایی‌های تصاویر که توسط مدل تولید شدند استفاده شد تا به‌صورت کمی مشخص شود که کدام یک از تغییرات مشکل بیشتری را و به چه ترتیب برای سامانه بینایی انسان ایجاد می‌کند. نشان داده شد که، از میان تغییرات مطالعه شده، چرخش در عمق، با درهم فروبردن بازنمایی گروه‌های اشیاء مشکل بیشتری را برای بازشناسی اشیاء ایجاد می‌کند.

در پژوهش جاری از شبکه عصبی کانولوشنی به‌عنوان مدلی از سامانه بینایی انسان در راستای دستیابی به بازنمایی‌های مغزی اشیاء استفاده شده است. یکی از اشکالات اساسی عموم مدل‌های سلسله‌مراتبی سامانه بینایی، ازجمله مدل حاضر، عدم لحاظ نمودن پارامتر زمان در ساختار آن است. این در حالی است که زمان به‌عنوان یکی از تاثیرگذارترین مولفه‌ها در بازشناسی اشیاء توسط انسان بشمار می‌رود [۱]. عموم مدل‌های بینایی از جمله مدل حاضر به فعالیت‌های عصبی قشر بینایی مغز به شکل میانگین فعالیت (یا به عبارت دیگر میانگین نرخ آتش نرون‌ها) نگاه می‌کنند و رمزگذاری اطلاعات را در مغز بر اساس میانگین فعالیت عصبی در نظر می‌گیرند که این خود موضوعی چالشی است. البته برای دخیل نمودن مولفه زمان، مدل‌هایی نیز بر اساس شبکه‌های عصبی ضربه‌ای [۳۳] و پایداری زمانی [۳۴] ارائه شده‌اند که توانسته‌اند در سطحی محدود سازوکارهای زمانی مغز را در بازشناسی اشیاء توضیح دهند، اما به علت پیچیدگی بسیار بالا در طراحی و محدودیت‌های پیاده‌سازی هنوز توفیق چندانی در کاربردهای عملی نداشته‌اند. بر اساس دلایل مذکور توسعه این مدل‌ها هنوز چندان مورد توجه پژوهش‌گران حوزه علوم بینایی و بینایی ماشین قرار نگرفته است. از طرف دیگر، توفیق روزافزون مدل‌های یادگیری عمیق در حل مسائل هوش مصنوعی [۳۵] و همچنین در شبیه‌سازی سامانه بینایی انسان باعث محبوبیت بالای مدل‌های بر اساس میانگین فعالیت عصبی از جمله شبکه‌های عصبی کانولوشنی شده است [۶]. علاوه بر موارد مذکور، با توجه به اینکه



شکل ۱۰ بازنمایی دو بُعدی پراکندگی اشیاء در هر یک از لایه‌های مدل (بازنمایی در فضای ورودی و لایه‌های آخر مدل در شکل ۶ نشان داده شده است).

۴ جمع‌بندی و نتیجه‌گیری

با توجه به اهمیت آگاهی از سازوکارهای سامانه بینایی انسان در بازشناسی اشیاء در راستای توسعه الگوریتم‌های محاسباتی با الهام از این سامانه، در این پژوهش نحوه برخورد این سامانه با تغییرات رایج در بازشناسی اشیاء مورد مطالعه قرار گرفت. در این راستا

بازشناسی اشیاء در انسان همچون مسیرهای بازخوردی و پردازش بازگشتی باشد.

مراجع

- [1] Thorpe, S., Fize, D., Marlot, C., "Speed of processing in the human visual system," *Nature*, Vol. 381, pp. 520–522, 1995.
- [2] Fabre-Thorpe, M., Richard G., Thorpe S., "Rapid categorization of natural images by rhesus monkeys," *Neuroreport*, Vol. 9, pp. 303–308, 1998.
- [3] Keysers, C., Xiao, D., Földiák, P., Perrett, D., "The speed of sight," *J. Cognitive Neuroscience*, Vol. 13, pp. 90–101, 2001.
- [4] Potter, M. C., Wyble, B., Haggmann, C. E., McCourt, E. S., "Detecting meaning in RSVP at 13 ms per picture," *Attention, Perception, & Psychophysics*, Vol. 76, pp. 270–279, 2013.
- [5] Hung, C. P., Kreiman, G., Poggio, T., DiCarlo, J. J., "Fast readout of object identity from macaque inferior temporal cortex," *Science*, Vol. 310, pp. 863–866, 2005.
- [6] Cadieu, C. F., Hong, H., Yamins, D. L. K., Pinto, N., Ardila, D., et al., "Deep neural networks rival the representation of primate IT cortex for core visual object recognition," *PloS Comp. Biol.*, Vol. 10, no. 12, 2014.
- [7] Zeiler, M. D., Fergus, R. "Visualizing and understanding convolutional networks," *Computer Vision—ECCV 2014*, Vol. 8689, pp. 818–833, 2014.
- [8] Krizhevsky, A., Sutskever, I., Hinton, G., "ImageNet: classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, Vol. 25, pp. 1106–1114, 2012.
- [9] Jarrett, K., Kavukcuoglu, K., Ranzato, M. A., LeCun, Y., "What is the best multi-stage architecture for object recognition?" *Computer Vision, IEEE 12th International Conference on*, pp. 2146–2153, 2009.
- [10] Lazebnik, S., Schmid, C., Ponce, J., "Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*. Vol. 2, pp. 2169–2178, 2006.
- [11] DiCarlo, J. J., Zoccolan, D., Rust, N. C., "how does the brain solve visual object recognition?" *Neuron*, Vol. 73, pp. 415–434, 2012.
- [12] DiCarlo J. J., Maunsell, J. H. R., "Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position," *J. Neurophysiology*, Vol. 89, pp. 3264–3278, 2003.
- [13] Desimone, R., Albright, T. D., Gross, C. G., Bruce, C., "Stimulus-selective properties of inferior temporal neurons in the macaque," *J. Neuroscience*, Vol. 4, pp. 2051–2062, 1984.

نوع رمزگذاری اطلاعات در مغز همچنان محل سوال است و عموماً در تحلیل سیگنال‌های عصبی از میانگین فعالیت استفاده می‌شود، استفاده از مدل‌های میانگین فعالیت در بازسازی رفتارهای مغزی چندان غیرمعقول به نظر نمی‌آید. بر همین اساس در بسیاری از مطالعات امروزی از این مدل‌ها در راستای توضیح رفتار انسان استفاده می‌شود [۶]، [۱۹] و [۲۶].

در این پژوهش برای نخستین مرتبه با ارائه روشی محاسباتی به مقایسه تغییرات مختلف پرداخته شد. روشی که در این مقاله ارائه شد با فراهم آوردن یک استاندارد عددی، فضایی مشترک میان فضاها متفاوت فراهم می‌آورد که می‌توان در آن بازنمایی خوشه‌ها را از نظر جدایی‌پذیری مورد مقایسه قرار داد. در این پژوهش با استفاده روش مذکور، فضای پیکسل و هر کدام از لایه‌های مدل محاسباتی با استفاده از شاخص دان مورد مقایسه قرار گرفتند و با معرفی شاخص پردازش مشخص شد که هر یک از تغییرات به‌صورت کمی با رفتن از فضای پیکسل به فضای خروجی مدل در راستای انجام بازشناسی تا چه حد بهبود یافته‌اند. روش ارائه شده را می‌توان برای بررسی بازنمایی‌های حاصل از ثبت سیگنال‌های عصبی (مانند الکتروفیزیولوژی، EEG، MEG و fMRI) نیز مورد استفاده قرار داد.

از دیگر دستاوردهای مهم این پژوهش می‌توان به این یافته اشاره نمود که: چرخش در عمق از دیگر تغییرات بررسی شده در این آزمایش مشکل‌تر دیده شد. از طرف دیگر، تغییر جهت تابش نور تاثیر چندان را در بازشناسی اشیاء در انسان نشان نداد. این مسئله پیش از این به صورت کمی مورد مطالعه قرار نگرفته بود. این یافته مشخص می‌کند که، علیرغم توانایی‌های بسیار بالا در بازشناسی اشیاء، مسیر پیش‌رو قشر بینایی انسان به تنهایی نمی‌تواند برای توسعه الگوریتم‌های محاسباتی بازشناسی اشیاء تحت چرخش در عمق کفایت کند و احتمالاً برای حل این‌گونه مسائل باید از راهبردهای دیگر بهره برد.

در راستای مقابله با جهت‌گیری عمقی یکی از راهکارهایی که بنظر می‌رسد راهگشا باشد استفاده از سازوکارهای بازخوردی و همچنین پردازش‌های بازگشتی (Recurrent) است که در سامانه بینایی انسان نیز مورد استفاده هستند [۳۶]. در آزمایش انسانی انجام شده در این پژوهش با توجه به اینکه این مسیرها تا حد زیادی از مدار خارج شده‌اند، تاثیر خود را در بازشناسی نشان ندادند. این در حالی است که پیش‌تر نشان داده شده است، سازوکارهای مذکور نقش کمک‌کننده‌ای در حل مسائل بازشناسی پیچیده ایفا می‌کنند [۲۶].

گرچه تا به امروز عموم مدل‌های محاسباتی بازشناسی اشیاء که از مسیر پیش‌رو قشر بینایی الهام گرفته‌اند نقش سازوکارهای بازخوردی را نادیده گرفته‌اند، اما نتایج این پژوهش نشان می‌دهد که شاید برای حصول بازشناسی مستقل از تغییرات اشیاء یکی از راهکارها استفاده از سازوکارهای کمکی درگیر در حل مسئله

- [29] Deng, J., Berg, A., Satheesh, S., Su, H., Khosla, A., Fei-Fei, L., ILSVRC-2012, 2012. <http://www.image-net.org/challenges/LSVRC/2012/>.
- [30] Pearson K., "Notes on regression and inheritance in the case of two parents," Proceedings of the Royal Society of London, Vol. 58, pp. 240-242, 1895.
- [31] DiCarlo, J. J., Cox, D. D., "Untangling invariant object recognition," Trends Cogn. Sci., Vol. 11, pp. 333-341, 2007.
- [32] Dunn, J. C., "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," J. Cybernetics, Vol. 3, pp. 32-57, 1973.
- [33] Masquelier, T., Thorpe, S.J., "Unsupervised learning of visual features through spike timing dependent plasticity," PLoS Comp. Biol., Vol. 3, no. 2, 2007.
- [34] Wyss, R., Konig, P., Verschure, M.J., "A model of the ventral visual system based on temporal stability and local memory," Plos Biol., Vol. 4, no. 5, 2006.
- [35] Lecun, Y., Bengio, Y., Hinton, G., "Deep learning," Nature, Vol. 521, no. 7553, pp. 436-444, 2015.
- [36] Freiwald, W. A., Tsao, D. Y., "Functional compartmentalization and viewpoint generalization within the macaque face-processing system," Science, Vol. 330, no. 6005, pp. 845-851, 2010.
- [14] Ashbridge, E., Perret, D. I., Oram, W. M., Jellema, T., "Effect of image orientation and size on object recognition: responses of single units in the macaque monkey temporal cortex," Cognitive Neuropsychology, Vol. 17, no. 1/2/3, pp. 13-34, 2000.
- [15] Kravitz, D. J., Vinson, L. D., Baker, C. I., "How position dependent is visual object recognition," Neuron, Vol. 2, no. 3, pp. 114-122, 2008.
- [16] Tjan, B. S., Legge, G. E., "The viewpoint complexity of an object recognition task," Vision Research, Vol. 38, no. 15, pp. 2335-2350, 1998.
- [17] Braje, W. L., Kersten, D., Tarr, M., Troje, N. F., "Illumination effects in face recognition," Psychobiology, Vol. 26, no. 4, pp. 371-380, 1998.
- [18] Guyonneau, R., Kirchner, H., Thorpe S. J., "Animals roll around the clock: The rotation invariance of ultrarapid visual processing," J Vision, Vol. 6, pp. 1008-1017, 2006.
- [19] Isik, L., Meyers, E. M., Leibo, J. Z., Poggio, T., "The dynamics of invariant object recognition in the human visual system," J Neurophysiology, Vol. 111, pp. 91-102, 2014.
- [20] LeCun, Y., Huang, F. J., Bottou, L., "learning methods for generic object recognition with invariance to pose and lighting," Proceedings of CVPR, 2004.
- [21] Geusebroek, J. M., Burghouts, G. J., Smeulders, A. W. M., "The Amsterdam library of object images," Int. J. Comput. Vision, vol. 61, no. 1, pp. 103-112, 2005.
- [22] Peters, G., Zitova, B., and von der Malsburg, C., "How to measure the pose robustness of object views," Image and Vision Computing, vol. 20, no. 4, pp. 249-256, 2002.
- [23] Felsen, G., Dan, Y., "A natural approach to studying vision," Nat. Neurosci., Vol. 8, pp. 1643-1646, 2005.
- [24] Gallant, J. L., Connor, C. E., Van Essen, D. C., "Neural activity in areas V1, V2, and V4 during free viewing of natural scenes compared to control images," Neuroreport, Vol. 9, pp. 85-89, 1998.
- [25] Pinto, N., Cox, D. D., DiCarlo, J. J., "Why is real-world visual object recognition hard," PLoS Comp. Bio., Vol. 4, pp. 1, 2008.
- [26] Serre, T., Oliva, A., Poggio, T., "A feedforward architecture accounts for rapid categorization," Proc. Natl. Acad. Sci., Vol. 104, no. 15, pp. 6424-6429, 2007.
- [27] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., "Gradient-based learning applied to document recognition," Proc. IEEE, Vol. 86, pp. 2278-2324, 1998.
- [28] Vedaldi, A., Lenc, K., "MatConvNet-Convolutional Neural Networks for MATLAB," Proc. of the ACM Int. Conf. on Multimedia, 2015. <http://www.vlfeat.org/matconvnet/>



حمید کریمی روزبهانی مدرک کارشناسی خود را در رشته مهندسی برق گرایش الکترونیک در سال ۱۳۸۸ از دانشگاه شهید رجایی دریافت نمود. سپس ایشان مدرک کارشناسی ارشد خود را در رشته مهندسی پزشکی گرایش بیوالکتریک در سال ۱۳۹۱ از دانشگاه علم و صنعت ایران کسب نمود.

ایشان هم‌اکنون دانشجوی مقطع دکتری تخصصی در رشته مهندسی برق، گرایش الکترونیک دانشکده مهندسی برق دانشگاه شهید رجایی است. ایشان بین سال‌های ۱۳۹۲ الی ۱۳۹۴ به‌عنوان پژوهشگر در پژوهشکده علوم شناختی، پژوهشگاه دانش‌های بنیادی (IPM) فعالیت پژوهشی داشتند. علاقه‌مندی‌های علمی ایشان شامل علوم اعصاب بینایی، بینایی ماشین، پردازش و بازشناسی سیگنال و تصویر است.



رضا ابراهیم‌پور دانشیار دانشکده مهندسی کامپیوتر دانشگاه تربیت دبیر شهید رجایی می‌باشند. ایشان مدرک کارشناسی مهندسی برق-الکترونیک را در سال ۱۳۷۸ از دانشگاه مازندران و مدرک کارشناسی ارشد مهندسی پزشکی-بیوالکتریک را در سال ۱۳۸۰ از دانشگاه تربیت مدرس دریافت نمودند. در فروردین ۱۳۸۱ به‌عنوان

دانشجوی اولین دوره دکتری علوم اعصاب شناختی در پژوهشکده علوم شناختی، پژوهشگاه دانش‌های بنیادی (IPM) شروع به تحصیل نمودند

و در سال ۱۳۸۶ موفق به اخذ مدرک دکتری تخصصی گردیدند. ایشان به‌عنوان پژوهشگر ارشد با پژوهشگاه دانش‌های بنیادی همکاری پژوهشی دارند. آقای دکتر ابراهیم پور بیش از ۱۰۰ مقاله علمی در مجلات و کنفرانس‌های علمی ارائه نموده‌اند و همچنین در کمیته علمی و داوری متجاوز از ۲۰ مجله و کنفرانس علمی فعالیت داشته‌اند. ایشان سرگروه داوری گروه مکترونیک جشنواره جوان خوارزمی می‌باشند و به‌علاوه از منتخبین سرآمدان علمی کشور توسط فدراسیون سرآمدان علمی کشور در سال ۱۳۹۴ می‌باشند. زمینه‌های تخصصی ایشان عبارت‌اند از: علوم اعصاب شناختی، مدلسازی شناختی، بینایی انسان و ماشین.



نصویر باقری استادیار گروه مهندسی مخابرات دانشکده مهندسی برق دانشگاه تربیت دبیر شهید رجایی می‌باشد. ایشان مدرک دکتری خود را در سال ۱۳۸۹ از دانشگاه علم و صنعت ایران دریافت کرد. زمینه تخصصی ایشان امنیت و رمزنگاری می‌باشد. علاوه بر این، ایشان در زمینه

مدلسازی شناختی سیستم بینایی انسان نیز فعالیت تحقیقاتی دارند. از ایشان تاکنون بیش از ۵۰ مقاله در مجلات و همایش‌های بین‌المللی منتشر شده است.