مجله ماشین‌بینایی و پردازش‌تصویر
سال هفتم، شماره اول، تابستان ۱۳۹۹

استفاده از سیگنال‌های بالابه‌پایین مبتنی بر محتوا برای بهبود بازشناسی شیء

الهه سادات سلطاندوست ناری[۱]، رضا ابراهیم پور[۲] و کریم رجایی[۳]

چکیده

بازشناسی شیء در صحنه‌های پیچیده‌ی ازجمله توانایی‌های شگرف سامانه بینایی انسان است که تاکنون مدل‌های محاسباتی بینایی در پیاده‌سازی آن چندان موفق نبوده‌اند. در این راستا محققان سعی دارند با شناسایی سازوکار مغز و الهام از آن این مدل را بهبود بخشند. یکی از موفق‌ترین مدل‌های ارائه‌شده در بازشناسی شیء شبکه‌های عصبی کانولوشنی (CNN's) هستند. این مدل‌ها تنها قادر به شبیه‌سازی مسیر پیش‌روی بینایی انسان می‌باشند. با این حال شواهد مطالعات علوم اعصاب نشان می‌دهند سامانه بینایی انسان سیگنال‌های بالابه‌پایین انتظار را در راستای افزایش دقت و سرعت بازشناسی شیء در زمینه‌های پیچیده به کار می‌بندد. در این مقاله با بهره‌مندی از سیگنال‌های بالابه‌پایین انتظار، سعی بر شبیه‌سازی مسیر بازخوردی سیستم بینایی انسان شده است. به این منظور مدل کانولوشنی AlexNet به‌عنوان مسیر پیش‌رو سیستم بینایی استفاده شد. برای بازشناسی شیء از مدل آموزش یافته با مجموعه داده‌ی ImageNet و برای بازشناسی صحنه از مدل آموزش یافته با مجموعه تصاویر صحنه Places استفاده شد. شبکه آموزش دیده بر روی تصاویر صحنه (Place_CNN) برای تولید بردار بازخورد مبتنی بر اطلاعات حاصل از صحنه در نظر گرفته شد. سیگنال‌های بازخوردی شامل اطلاعاتی از فراوانی تکرار شیء موردنظر در صحنه‌ی جاری هستند. این سیگنال‌ها با قاعده‌ی پس‌انتشار در قالب سیگنال‌های بالابه‌پایین با اطلاعات مسیر پیش‌رو تلفیق و در شبکه‌ی تشخیص شیء بازخورد می‌شوند. به‌منظور سنجش مدل پیشنهادی آزمایش‌هایی با استفاده از چند مجموعه داده صورت گرفت. نتایج نشان داد که ترکیب اطلاعات بازخوردی با مسیر پیش‌رو باعث بهبود معنی دار عملکرد مدل پیشنهادی نسبت به مدل پایه‌ی AlexNet می شود. استفاده از اطلاعات محتوایی تصاویر باعث بهبود عملکرد بازشناسی شیء می‌شود به‌خصوص هنگامی که شیء هدف در شرایط چالشی قرار گرفته است.

کلیدواژه‌ها

شبکه عصبی کانولوشنی، بازشناسی شیء، محتوا، شبکه‌ی Place_CNN، شبکه‌ی AlexNet

۱ مقدمه

بازشناسی شیء در دنیای واقعی یکی از پیچیده‌ترین کارهایی است که سرعت و دقت بالایی توسط دستگاه بینایی انسان و سایر پستانداران، با آزمایش‌ها نشان می‌دهند که بینایی انسان در دو مسیر کلی .انجام می‌شود جریان پیدا می‌یابد: مسیر قدامی[1] (چگونگی و مکان شیء را پردازش می‌کند) که از قشر بینایی تا لوب آهیانه[2] گسترده است و مسیر شکمی[3] (چیستی و ماهیت شیء را پردازش می‌کند) که از قشر بینایی تا قشر

این مقاله در خرداد ماه سال ۹۶ دریافت، در آبان ماه سال ۹۶ بازنگری و در تیر ماه سال ۹۸ پذیرفته شد.

[۱] کارشناس ارشد کامپیوتر، دانشکده مهندسی کامپیوتر، دانشگاه تربیت دبیر شهید رجایی.
رایانامه: elahe.soltandoostn@srttu.edu
[۲] گروه هوش مصنوعی، دانشکده مهندسی کامپیوتر، دانشگاه تربیت دبیر شهید رجایی.
رایانامه: rebrahimpour@sru.ac.ir
[۳] دانشجوی دکتری علوم اعصاب شناختی، پژوهشکده علوم‌شناختی، پژوهشگاه دانش‌های بنیادی.
رایانامه: rajaei.k@ipm.ir
نویسنده مسئول: رضا ابراهیم‌پور

[1] Dorsal Pathway
[2] Parietal Lobe
[3] Ventral Pathway

) گسترش‌یافته است. در مسیر بازشناسی شیء ITگیجگاهی تحتانی[1] (طی سلسله‌مراتبی از پردازش‌ها، ویژگی‌های تصویر ورودی به ترتیب از بازشناسی ITساده به پیچیده استخراج می‌شوند و درنهایت در ناحیه‌ی اتفاق می‌افتد.

برای مدت‌ها محققان دقت و سرعت بازشناسی شیء را تنها حاصل سامانه‌ی سلسله‌مراتبی پیش‌رو[2] ی قشر بینایی می‌دانستند، اما تحقیقات تأثیر سیگنال‌های بالابه‌پایین[3] را در پردازش بینایی نشان می‌دهد [۱] تا [۸]. یافته‌ها حاکی از آن است که درک صحنه های بینایی پیچیده مستلزم حجم بالایی از محاسبات است و اجرای این محاسبات، تنها با تکیه بر سازوکار پیش‌رو میسر نیست. مسیرهای بالابه‌پایین حامل اطلاعات غنی و متفاوتی درباره ی مفاهیم رفتاری هستند که درک محتوای بینایی را تسهیل کرده و اجازه می‌دهند جزئیات بیشتری از ورودی استخراج شود. سیگنال های بالابه‌پایین بازخوردی با توجه به سازوکاری که پیش می‌گیرند و انعکاسی که در رفتار ایجاد می‌کنند، متفاوت هستند. اگرچه سازوکار این سیگنال‌ها همواره ناشناخته است، اثرات آن‌ها در سازوکارهایی از مغز مانند توجه[4]، انتظار[5]، وظیفه‌ی ادراکی[6]، موتور کنترل[7] و حافظه کاری[8] مشاهده شده است [۲]. به‌طورکلی اتصالات بالابه‌پایین را می‌توان در دو نوع بازخوردهای محلی[9] و بالابه‌پایین[10] بیان کرد. این سیگنال‌ها از لحاظ زمان فعال‌شدن و منشأ شکل‌گیری با یکدیگر متفاوت هستند. همچنین به هر یک از آن‌ها وظیفه‌ای نسبت داده می‌شود. تأثیر اتصالات بازخوردی محلی در حدود ۱۰۰ میلی‌ثانیه در مسیر شکمی دیده شده است [۹]. این سیگنال‌ها به‌صورت غیرارادی تولید می‌شوند [۱۰] و [۱۱]. اما سیگنال‌های بالابه‌پایین با تأخیر بیشتری رخ می‌دهند (حدود ۱۵۰ تا ۲۰۰ میلی‌ثانیه و بیشتر [۱۲] و [۱۳]). این سیگنال‌ها برخلاف سیگنال‌های محلی با اراده فرد تولید می‌شوند [۱۴] و [۱۵]. منشأ این سیگنال‌ها اغلب لوب پیشانی[11] و لوب آهیانه پیشنهاد شده است. اثرگذاری سیگنال‌های بالابه‌پایین در حدی است که انتخاب‌پذیری نورون‌ها تحت تأثیر آن‌ها، گاه پیچیده‌تر و گاه ساده‌تر می‌شوند. به‌طوری که پاسخ‌دهی نورون‌های V_1 به‌جای خطوط صاف و جهت‌دار، تحت تاثیر سیگنال‌های بالابه‌پایین به اشکال هندسی ساده تغییر می‌کند (برای جزئیات بیشتر مراجعه شود به [۱۶]). همچنین مطالعات نشان می‌دهند نیاز به سیگنال‌های بالابه‌پایین با بالا رفتن میزان پیچیدگی وظیفه‌ی بصری[12] افزایش می‌یابد [۱] تا [۳].

توجه و انتظار دو سازوکاری هستند که در سال‌های اخیر نسبت به سایر سیگنال‌های بالابه‌پایین مورد توجه بیشتری قرارگرفته‌اند. سیگنال انتظار، احتمالی است از آن چیزی که در آینده‌ی نزدیک در مجاورت زمانی و مکانی ما پدیدار خواهد شد [۱۷]. به‌عنوان مثال، مکان شیء در حال حرکت، زاویه‌ی چرخش، شیء که ممکن است در صحنه‌ی پس از صحنه‌ی جاری پدیدار شوند (مکان و هویت آن‌ها) و غیره برای انسان به راحتی قابل پیشگویی هستند. اما این پیشگویی‌ها بر چه مبنایی ایجاد می شوند؟ با توجه به اینکه شیء معمولاً دارای وابستگی‌های انجمنی[13] بوده و در زمان و مکان خاصی حاضر می‌شوند، مغز انسان با ذخیره‌ی فراوانی حضور شیء با یکدیگر، تعداد دفعات حضور آن‌ها در مکان‌های متفاوت و اطلاعات آماری مشابه، همچنین با اعتماد به ثابت بودن جهان اطراف، به پیشگویی می پردازد [۱۸]. از دیگر آثار سیگنال انتظار، می‌توان افزایش اعتماد به نفس در تصمیم‌گیری و افزایش تأثیر فراشناختی در تصمیمات ادراکی را نام برد [۱۹].

نقش محتوا در بازشناسی شیء را نیز می‌توان اثر سیگنال‌های انتظار در مغز دانست. محتوا درواقع ترکیبی از شرایطی است که به تعریف، بازنمایی و در نهایت، معنا یافتن محیط (صحنه) کمک می‌کنند. این شرایط همان وابستگی‌های انجمنی هستند که در بلندمدت و به دلیل تکرار در رخداد، ذخیره شده و ارتباط محتوایی را ایجاد می‌کنند. ارتباط محتوایی بیان‌گر ارتباط میان شیء مرتبط به یک محتوا (صحنه)، مانند اتاق و صندلی راحتی، یا ارتباط مربوط به رابطه‌ی فضایی (صفحه کلید در پایین مانیتور قرار می گیرد) است. ازآنجایی که شیء در دنیای اطراف ما منفرد نیستند و همواره در محیط و با سایر اشیاء ظاهر می‌شوند، این اطلاعات در بازشناسی شیء نقش مؤثری دارند [۲۰] و [۲۱]. به‌عنوان مثال، یک ساحل به‌طور معمول شامل شیء مانند چتر ساحلی و یا صندلی ساحلی می‌باشد و حضور شیء مانند تلویزیون در این صحنه، احتمالی نزدیک به صفر خواهد داشت و می‌توانند از رقابت بازشناسی شیء در این صحنه حذف شود. همچنین وجود یک گربه در این صحنه احتمال کمی خواهد گرفت؛ پس با وزن کمتری در رقابت شرکت داده خواهد شد. بنابراین شکستن فضای جستجو از تمام شیء ذخیره شده در مغز به فضایی محدود به شیء محتمل در یک صحنه، با کاستن از کلاس‌های هدف موجود در لایه‌ی IT، دقت و سرعت بازشناسی شیء را بهبود می‌بخشد [۱۸]. بر این اساس، بسیاری از تحقیقات کلاسیک نیز نشان داده‌اند که شیء در صحنه‌ی مرتبط (به‌عنوان مثال یک ببر در باغ‌وحش) با اعتماد و دقت بیشتری نسبت به شیء در صحنه‌ی

[1] Inferior Temporal Cortex
[2] Feed-forward
[3] Top-down
[4] Attention
[5] Expectation
[6] Perceptual Task
[7] Motor Control
[8] Working Memory
[9] Local Feedback
[10] Top-Down Feedback
[11] Frontal Lobe
[12] Visual Task
[13] Dependence Aassociative

غیر مرتبط (به‌عنوان مثال یک ببر در اتاق‌خواب) تشخیص داده می شوند [۲۲] و [۲۳]. این شواهد اهمیت صحنه‌ی جاری در بازشناسی شیء را نشان می‌دهد.

در سوی دیگر، درحالی‌که بازشناسی شیء توسط دستگاه بینایی انسان و سایر پستانداران، بدون هیچ زحمتی و با کارایی بالا انجام می‌شود، مدل‌های بینایی ارائه شده در این حوزه هنوز با کارایی مغز فاصله دارند. یکی از مدل‌های موفقی که در حوزه‌ی بینایی ماشین مطرح شده است شبکه‌های عصبی کانولوشنی[1] می‌باشند. این مدل‌ها عملکرد قابل قبولی در طبقه‌بندی تعداد زیادی شیء (به‌عنوان مثال ۱۰۰۰ دسته شیء [۲۴]) دارند. همچنین شواهدی به نفع تطابق آن‌ها با عملکرد مغز وجود دارد [۲۵] تا [۲۶] تاکنون روش‌های متفاوتی جهت افزایش تطابق این شبکه‌ها با عملکرد انسان [۲۷] و [۲۷] و همچنین رفع ناتوانی آن‌ها در بازشناسی شیء پیچیده ارائه شده است. اما همواره مسائل زیادی در این حوزه مبهم هستند. مشکل عمده‌ی شبکه‌های عصبی کانولوشنی محاسبات حجیم [۲۴] و ناتوانی‌شان در بازشناسی شیء در صحنه‌ی بینایی پیچیده [۳] است. پیش‌بینی می‌شود افزودن اطلاعاتی در قالب سیگنال‌های بازخوردی بالابه‌پایین علاوه بر افزایش تطابق‌پذیری این مدل‌ها با عملکرد مغز، در بازشناسی شیء حاضر در صحنه‌های بینایی پیچیده نیز به آن‌ها کمک کند.

همان‌طور که پیش‌تر بیان شد، یکی از سیگنال‌های بازخوردی مؤثر در بازشناسی شیء سیگنال‌هایی است که از صحنه نشأت گرفته و بر مسیر پیش‌روی بینایی اعمال می‌شوند. تاکنون مطالعه‌های بسیاری به صورت مدل‌سازی [۲۸] تا [۳۰]، رفتاری، تصویربرداری عصبی [۳۱]تا [۳۴] و غیره نقش صحنه را در بازشناسی شیء بررسی نموده‌اند. گروهی از این مطالعات نیز بر روی CNN ها صورت گرفته است [۳۵] و [۳۶]؛ که از جمله‌ی آن‌ها می‌توان به شبکه‌های عصبی کانولوشنی مبتنی بر ناحیه یا R-CNN ها اشاره کرد [۳۷]. در این شبکه‌ها قسمت‌هایی از تصویر ورودی به‌عنوان نواحی پیشنهادی انتخاب می‌شوند. ویژگی‌های نواحی پیشنهادی پس از عبور از یک شبکه‌ی کانولوشنی به دست می‌آید و در نهایت تابع بیشینه‌گیر نرم (Softmax) معین می‌کند که کدام یک از این نواحی مرتبط به پس‌زمینه و کدام یک مرتبط به شیء مورد نظر هستند. مدل ارائه شده در [۲۸] با نگاهی دیگر از اطلاعات صحنه جهت بازشناسی شیء استفاده می‌نماید. در این دیدگاه آن دسته از نواحی که احتمال حضور شیء در آن‌ها وجود دارد با استفاده از یک شبکه‌ی کانولوشنی بازشناسی صحنه مشخص می‌شوند. این نواحی ورودی‌های شبکه‌ی بازشناسی شیء را می‌سازند.

اگرچه این مدل‌ها نتایج قابل قبولی از خود ارائه داده‌اند، به نظر می‌رسد هنوز از روندی که در شبکه‌ی محتوایی مغز اتفاق می افتد فاصله دارند. چالش اصلی اغلب این مدل‌ها چگونگی ترکیب اطلاعات بازخوردی محتوایی با یکدیگر و با اطلاعات پیش‌رو است [۳۰]. در این مطالعه با اعمال مستقیم اطلاعات محتوایی به شبکه‌ی شناخته شده‌ی AlexNet تأثیر این اطلاعات بر افزایش کارایی این شبکه نشان داده شده است. در مقاله‌ی حاضر از شبکه‌ای کانولوشنی که برای تشخیص صحنه آموزش دیده برای به دست آوردن اطلاعات محتوایی استفاده شد. بنابر خروجی این شبکه بردار بازخوردی خاصی فعال شده و اطلاعات محتوایی را با نتایج به دست آمده از مسیر پیش‌رو ترکیب می‌کند. در اینجا از یک شبکه‌ی کانولوشنی مشابه که برای بازشناسی شیء آموزش دیده نیز به‌عنوان مسیر پیش‌روی تشخیص شیء بینایی استفاده شده است. در بخش دوم پیکربندی مدل کانولوشنی پایه‌ی به کار برده شده در این مقاله تشریح و در بخش سوم جزئیات پیاده‌سازی و بازخورد اعمال شده، بیان می‌شود؛ در بخش چهارم، مجموعه داده‌ها، آزمایش‌ها و نتایج حاصل آورده و در بخش نهایی به نتیجه‌گیری پرداخته شده است.

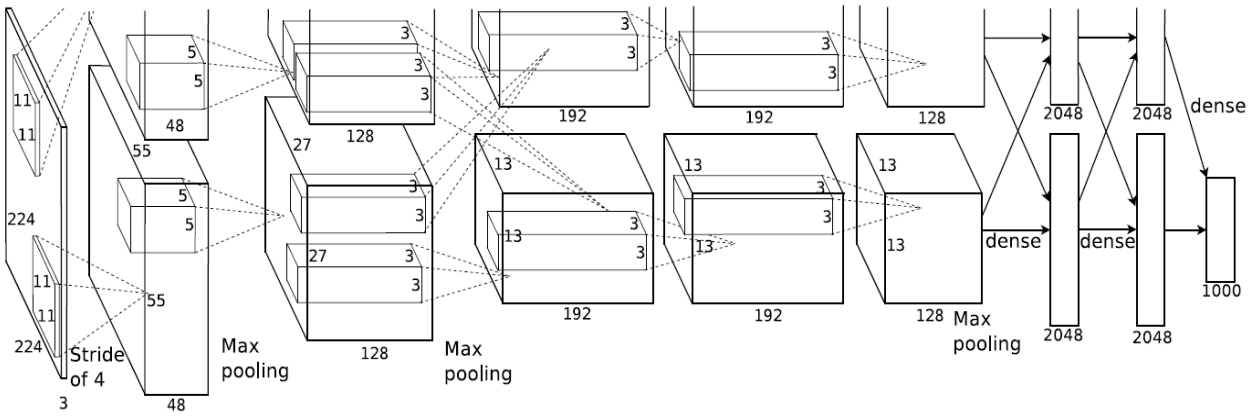
۲ شبکه عصبی کانولوشنی

شبکه عصبی کانولوشنی اولین بار در سال ۱۹۹۸ توسط لیکان ارائه شد. شبکه عصبی کانولوشنی بسیار مشابه شبکه عصبی مصنوعی می‌باشد و الهام گرفته از مسیر پیش‌روی شکمی بینایی انسان است [۳۸]. این شبکه‌ها برای پردازش ورودی‌های چند بعدی؛ دوبعدی (مانند آرایه‌ی دو بعدی برای تصویر یا طیف نگاره‌های صدا) و سه بعدی (برای ویدئو یا تصاویر حجمی) طراحی شده‌اند [۳۹]. در سال ۲۰۱۲ نسخه جدیدی از این مدل توسط گروه هینتون[2] ارائه شد و به نام AlexNet شهرت یافت. این این شبکه با حضور درخشان خود در مسابقه‌های ILSVRC-۲۰۱۲ توانست به دقت ۶۲,۵ درصد بر روی مجموعه داده ۱۰۰۰ کلاسه‌ی ImageNet دست یابد [۲۴]. مطالعات نشان می‌دهد در صورتی که داده‌های بزرگ[3] برای آموزش پارامترهای نسبتا زیاد این شبکه‌ها در دست باشد نه تنها در طبقه‌بندی شیء، بلکه برای وظایفی مانند بازشناسی محتوا و صحنه نیز قدرتمند ظاهر می‌شوند [۴۰]. مدل پیشنهادی برای بازشناسی شیء از مدل آموزش یافته با مجموعه داده‌ی ImageNet (Object_CNN) و برای بازشناسی صحنه از مدل آموزش یافته با مجموعه تصاویر صحنه Places (Place_CNN) استفاده می‌کند. ساختار هر دو شبکه مورد استفاده در مدل پیشنهادی یکسان است با این تفاوت که یکی برای بازشناسی شیء و دیگری برای بازشناسی صحنه آموزش داده شده‌اند. در بخش بعد ساختار این شبکه تشریح می‌شود.

[1] Convolutional Neural Networks

[2] Hinton

[3] Big Data



شکل ۱ ساختمان شبکه‌ی AlexNet. ساختار این شبکه از پنج لایه‌ی کانولوشنی به‌عنوان استخراج‌کننده‌ی ویژگی و سه لایه‌ی کاملاً متصل انتهایی به‌عنوان طبقه‌بند تشکیل شده است. بعد از برخی لایه‌های کانولوشنی لایه‌های ادغام و نرمال‌سازی اعمال می‌شوند. به دلیل حجم بالای محاسبات، آموزش این شبکه بر روی چند واحد گرافیکی مجزا صورت می‌گیرد [۲۴].

۳ معماری AlexNet

ساختار اصلی شبکه‌ی AlexNet را سه نوع لایه‌ی ادغام[1]، کانولوشن[2] و لایه‌های کاملاً متصل[3] تشکیل می‌دهد. این شبکه از پنج لایه‌ی کانولوشنی به‌عنوان لایه‌های استخراج ویژگی و از سه لایه‌ی کاملاً متصل برای طبقه‌بندی استفاده می‌کند. روی خروجی برخی لایه‌های کانولوشنی عملگر ادغام اعمال می‌شود. لایه‌های ادغام علاوه‌بر کاهش بعد فضایی، در شبکه نوعی مقاومت نسبت به تغییرات مختصر ورودی ایجاد می‌کنند (شکل ۱). این شبکه دارای حدود ۶۵۰,۰۰۰ واحد محاسباتی و ۶۰ میلیون پارامتر است. برای جلوگیری از بیش‌برازش[4] در این مدل علاوه‌بر لایه‌های ادغام، از عملگر حذف تصادفی[5] استفاده می‌شود. همچنین تابع یکسوساز واحد خطی تصحیح (RELU)[6] به‌عنوان تابع فعالیت به‌کارگرفته شده است [۲۴].

به دلیل هزینه‌ی بالای محاسباتی آموزش این شبکه‌ها معمولاً از آموزش آن‌ها خودداری شده و از وزن‌های از پیش آموزش داده شده که به صورت رایگان در اینترنت قابل دسترس هستند، استفاده می‌شود. به دلیل سازگار ساختن شبکه با فضای مسئله‌ی مورد نظر تنها برخی لایه‌های این شبکه‌ها با مجموعه داده‌های مورد آزمایش، آموزش می‌بینند. به این روش سازگار کردن[7] می‌گویند. در این مقاله مقاله نیز برای هر آزمایش لایه‌ی انتهایی شبکه‌ها روی مجموعه داده‌های مورد آزمایش سازگار شدند.

۴ مدل پیشنهادی

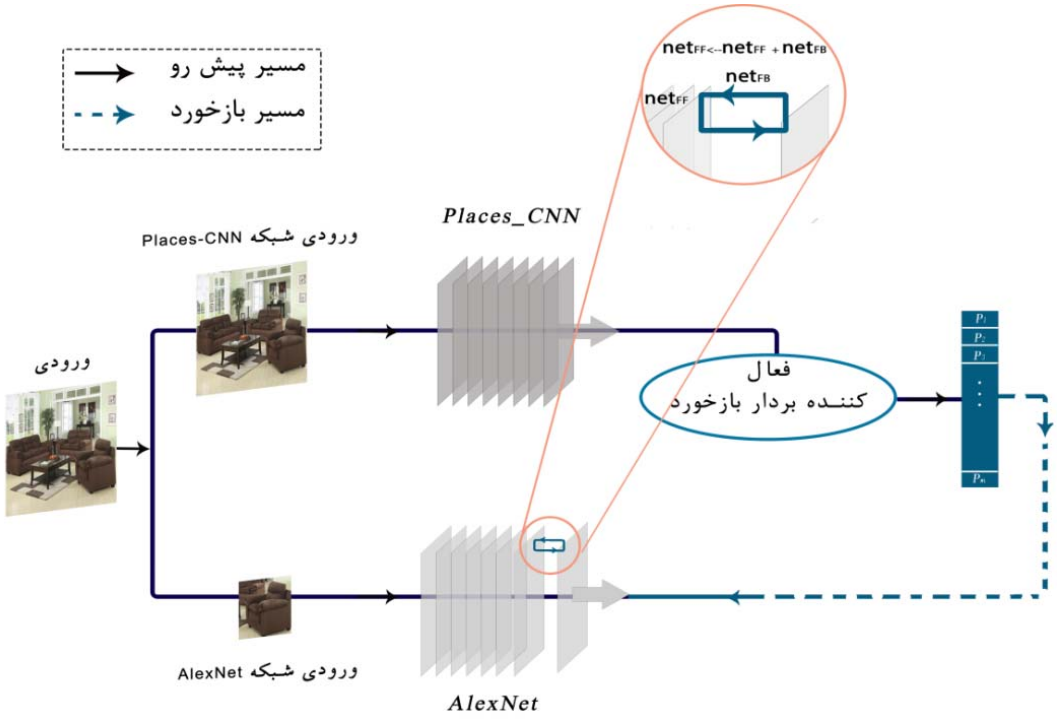
مدل پیشنهادی با الهام از مسیر بینایی و سیگنال‌های بازخوردی از نواحی بالایی مغز شکل گرفته است. در این مدل با ایجاد بازخورد در شبکه‌ی بازشناسی شیء پایه، کلاس‌های محتمل‌تر تقویت و کلاس‌های با احتمال کم تضعیف شده و از دور رقابت حذف می‌شوند. بنابراین با دخالت سیگنال‌های بازخوردی دقت مدل افزایش خواهد یافت. به نظر می‌رسد روند مشابهی در مغز وجود دارد. سیگنال‌های بالابه‌پایینی که از نواحی بالایی مغز دریافت می‌شوند دقت و سرعت بازشناسی را افزایش می‌دهند [۱] و [۲]. اگرچه هنوز سازوکار دقیق این سیگنال‌ها مشخص نیست، اما مطالعات نشان می‌دهند که تأثیرهایی مانند کاهش تعداد کلاس‌های شرکت‌کننده در رقابت موجود در سطح IT و تأکید بر کلاس‌های محتمل و تأثیر آن در لایه‌های پایین‌تر از جمله‌ی این پردازش‌هاست [۱۸].

در مدل پیشنهادی با اعمال بردار بازخورد سعی بر شبیه‌سازی این روند شده است. در بردار بازخورد اعمال شده در مدل پیشنهادی، وزن بیشتری به کلاس‌های محتمل و وزن کمی به کلاس‌های غیرمحتمل انتساب داده می‌شود و با تکرار اعمال بازخورد پردازش‌های بیشتری بر روی ورودی صورت می‌گیرد. این عمل باعث هدایت مدل به سمت یافتن شیء هدف و اجتناب از تشخیص شیء غیر هدف به‌عنوان کلاس برنده می‌شود. در این مقاله مشخص شدن شیء غیر هدف توسط شبکه‌ای که عمل بازشناسی محتوا یا صحنه را انجام می‌دهد (Place_CNN) اتفاق می‌افتد. این شبکه با دریافت تصویر ورودی، صحنه‌ی مربوط به آن را شناسایی می‌کند. با توجه به نتایج بازشناسی صحنه، بردار بازخوردی شامل فراوانی حضور شیء محتمل فعال می‌شود. از طرفی شبکه‌ای که وظیفه‌ی بازشناسی شیء را بر عهده دارد، تصویر ورودی را دریافت و روی ماهیت آن قضاوت می‌کند. سپس بردار بازخوردی حاصل از شبکه‌ی Place_CNN در شبکه‌ی

[1] Pooling
[2] Convolution
[3] Fully connected
[4] Over-fitting
[5] Dropout
[6] Rectified Linear Units
[7] Fine-tuning

بازشناسی شیء بازخورد شده و بر قضاوت این شبکه تأثیر می‌گذارد (شکل ۲).

در بردار بازخوردی فعال شده، شیء نامرتبط با کلاس صحنه‌ی برنده، غیر هدف معرفی می‌شوند. کلاس‌های غیرمحتمل و محتمل توسط وزن‌های اختصاص داده شده به هر کلاس در بردار بازخورد مشخص می‌شوند. در این بردار هرچه وزن کمتری به کلاس‌ها نسبت داده شود، در واقع احتمال حضور آن‌ها در صحنه‌ی مورد نظر کمتر است. با این حال با روش‌های متفاوتی می‌توان وزن این کلاس‌ها را مطابق با فراوانی حضورشان در صحنه‌ی مورد نظر، به شبکه‌ی بازشناسی شیء انتقال داد (در بخش «انتخاب بردار بازخوردی» چند روش متفاوت از مقداردهی به بردارهای بازخوردی مطابق با احتمال حضور شیء در صحنه‌ی مورد نظر، مطرح شده است).



شکل ۲ مدل پیشنهادی. مسیر پیش‌رو با خط صاف و مسیر بازخوردی با خطوط مقطع مشخص شده است. بعد از مشخص شدن صحنه‌ی تصویر ورودی توسط شبکه‌ی Place_CNN، بردار مربوطه فعال شده و به لایه‌ی آخر شبکه‌ی بازشناسی شیء اعمال و پنج مرتبه تکرار می‌شود. بعد بردار بازخوردی برابر است با تعداد کلاس‌های شیء که شبکه‌ی بازشناسی شیء بر روی آن‌ها قضاوت می‌کند. این بردار شامل تناوب حضور شیء مورد نظر در صحنه‌ی منتخب توسط شبکه‌ی بازشناسی صحنه است.

پس از ایجاد بردار بازخوردی متناسب، با بهره ۵۰ از قاعده‌ی پس‌انتشار[1] این بردار در لایه‌ی انتهایی شبکه‌ی بازشناسی شیء چندین بار بازخورد می‌شود. هر بار اطلاعات بالابه‌پایین با داده‌های مسیر پیش‌رو تلفیق شده و مجددا از شبکه عبور می‌کنند. در بخش بعد جزئیات بیشتری از چگونگی پیاده‌سازی بازخورد و ترکیب آن با اطلاعات پیش‌رو تشریح شده است.

۴-۱ پیاده‌سازی بازخورد

ساختار بینایی انسان سلسله‌مراتبی است. ارتباطات در این سلسله-مراتب به‌صورت دوطرفه هستند. به گونه‌ای که سیگنال‌های بازخوردی بر سیگنال‌های حاصل از مسیر پیش‌رو تأثیر گذاشته و آن‌ها را تقویت و یا تضعیف می‌کند. در فرمول بازخورد استفاده شده در این مقاله (۱)، این عملیات با عملگر ضرب شبیه‌سازی شده است [۴۱]. پس از عبور ورودی از شبکه، عمل بازخورد چندین بار تکرار می‌شود. در هر تکرار اطلاعات محتوایی با استفاده از قاعده‌ی پس‌انتشار تا لایه‌ی مورد نظر بازخورد و با اطلاعات پیش‌رو در آن لایه مطابق فرمول (۱) ترکیب می‌شوند. درنهایت ورودی جدید تولید شده و در مسیر پیش‌روی شبکه تا لایه‌ی انتهایی پیش می‌رود. سپس مجددا با اطلاعات محتوایی ترکیب شده و تا لایه‌ی مورد نظر بازخورد می‌شود. این عملیات چندین مرتبه تکرار می‌شود.

$$\mathbf{net^{FF}[n+1] = net^{FF}[n] \times (1 + \eta . net^{FB}[n])} \quad (۱)$$

در این رابطه $\mathrm{net}^{FF}[n+1]$ ورودی جدیدی است که در هر تکرار تولید می‌شود. ضریب اعشاری η جهت کنترل میزان تأثیر اطلاعات بازخوردی است و مقدار آن طی آزمون و خطا برابر ۰٫۴ قرار داده شد. $net^{FB}[n]$ پاسخ حاصل از بردار بازخوردی از نواحی بالاتر و $net^{FF}[n]$ پاسخ حاصل از مسیر پیش‌رو در مرحله‌ی قبل است. با این روش اطلاعات حاصل از مفاهیم

[1] Back-propagation

صحنه‌ی بصری، پردازشی که در مسیر پیش‌روی شبکه‌ی بازشناسی شیء انجام می‌شود را تحت تأثیر قرار می‌دهند. انتظار می‌رود این ادغام عملکرد مدل را بهبود بخشد.

فرآیند کلی مدل پیشنهادی به این ترتیب می‌باشد. پس از یک مرتبه عبور تصویر ورودی از لایه‌های مدل، بردار بازخوردی که به صورت موازی از مدل بازشناسی صحنه ایجاد شده است، از طریق قاعده‌ی پس‌انتشار تا لایه‌ی موردنظر بازگردانده می‌شود. همان‌طور که پیش‌تر بیان شد این نوع بازخورد در مغز، در لایه‌ی IT اعمال شده و از تعداد کلاس‌های هدف می‌کاهد. به این ترتیب عملیات بازشناسی شیء را تسهیل می‌بخشد [۱۸]. در مدل Object_CNN، لایه‌ی نهایی که وظیفه‌ی طبقه‌بندی را برعهده دارد و می‌توان آن را به‌عنوان لایه‌ی IT در نظر گرفت برای اعمال بازخورد انتخاب شد. در نتیجه اطلاعات بازخوردی تنها یک لایه برگردانده می‌شود. پس از اعمال بردار بازخورد مطابق با فرمول (۱)، ورودی لایه‌ی نهایی شبکه تحت تأثیر قرار می‌گیرد. حاصل این ترکیب به‌عنوان ورودی جدید لایه‌ی نهایی، مسیر پیش‌روی این لایه را طی می‌کند. این روند در مدل پیشنهادی چند بار تکرار می‌شود.

بردار بازخورد با توجه به محتوای تشخیص داده شده توسط شبکه‌ی Place_CNN ایجاد می‌شود. به‌عنوان مثال، در طبقه‌بندی کلاس شیء ابتدا با استفاده از تصاویر حاشیه‌نویسی شده‌ی مجموعه داده‌ی SUN [۴۲] فراوانی حضور اشیاء در پنج صحنه‌ی مورد نظر جمع آوری می‌شود. در نهایت، پنج بردار (به تعداد صحنه‌های این آزمایش) به ابعاد ۱×۵۰ به‌عنوان اطلاعات محتوایی ذخیره شده در حافظه ایجاد می‌شوند. پس از مشخص شدن صحنه‌ی تصویر ورودی، بردار بازخورد مربوط به صحنه‌ی برنده فعال می‌گردد. سپس بردار بازخورد از لایه‌ی انتهایی شبکه‌ی Object_CNN با قاعده‌ی پس‌انتشار یک لایه به عقب برگردانده می‌شود. این عمل چند بار تکرار می‌شود و هر بار با تأکید بر کلاس‌های محتمل در محتوای تشخیص داده شده، شبکه را به سوی کلاس‌های محتمل‌تر هدایت می‌کند.

۵ آزمایش‌ها و نتایج

برای ارزیابی مدل پیشنهادی، کارایی آن روی مجموعه داده‌های مجزا با مدل پایه به رقابت گذاشته شد. در بخش‌های بعد به تشریح و بررسی نتایج این آزمایش‌ها پرداخته شده است.

۵-۱ مقایسه‌ی مدل پیشنهادی با مدل پایه برای طبقه بندی مجموعه داده‌ی ۵۰ کلاسه

به‌عنوان اولین تلاش، شبکه‌ی Place_CNN با پنج کلاس صحنه، که از مجموعه داده‌ی Palces انتخاب شده بود، آموزش داده شد. این کلاس‌ها شامل صحنه‌های فضای باز[1] و محیط داخل ساختمان[2] بودند. همچنین ۵۰ کلاس جهت طبقه‌بندی در شبکه‌ی Object_CNN به‌گونه‌ای انتخاب شدند که هر کلاس شیء حداقل در یکی از کلاس‌های صحنه‌ی برگزیده شده موجود باشد. در نهایت، مدل پیشنهادی و شبکه‌ی Object_CNN به‌عنوان مدل پایه در بازشناسی این ۵۰ کلاس شیء به رقابت گذاشته شدند. نتایج حاصل، بهبود دقت و جدایی‌پذیری کلاس‌ها در مدل پیشنهادی را نشان داد.

- آزمایش

برای آموزش شبکه‌ی Place_CNN، از مجموعه داده‌ی Places استفاده شد. این مجموعه داده از تصاویر طبیعی ایجاد شده و با ۴۷۶ کلاس و حداقل ۱۶۰۰ نمونه در هر کلاس، از بزرگ‌ترین مجموعه داده‌های صحنه‌ی موجود می‌باشد [۴۰]. پنج کلاس صحنه (آشپزخانه، حیاط، مهدکودک، اتاق نشیمن و آسیاب بادی) از مجموعه داده‌ی Places انتخاب و برای هر کلاس ۸۰۰ نمونه آموزشی و ۲۰۰ نمونه اعتبارسنجی به‌صورت تصادفی اختیار شد. برای آموزش شبکه‌ی Object_CNN نیز ۵۰ کلاس با روش بیان شده از مجموعه داده ImageNet انتخاب شد. مجموعه داده ImageNet ازجمله بزرگ‌ترین مجموعه داده‌های شیء است و هر ساله با استفاده از تصاویر موجود در اینترنت به‌روزرسانی می‌شود [۲۴] و [۴۳]. مجموعه داده‌های Places و ImageNet که در این آزمایش استفاده شده‌اند به ترتیب توسط دانشگاه‌های ام‌آی‌تی[3] و استانفورد[4] به‌صورت رایگان در دسترس هستند.

- مجموعه داده

انتخاب شیء هر صحنه با استفاده از مجموعه داده SUN صورت گرفته است که شامل تصاویر صحنه‌ها، مکان‌های حاشیه گذاری شده[5] و اشیاء موجود در آن‌ها است. مجموعه داده‌ی SUN شامل ۹۰۸ کلاس صحنه و ۳۸۱۹ کلاس شیء می‌باشد [۴۲]. در این مجموعه داده تناوب حضور اشیاء مشاهده شده در هر صحنه مشخص شده است. از میان اشیاء متناوب در صحنه‌های اختیار شده، کلاس‌های شیء انتخاب گردید. این ۵۰ کلاس شیء به گونه‌ای انتخاب شده است که هر یک از آن‌ها حداقل در یکی از پنج کلاس صحنه موجود باشند. چگونگی توزیع ۵۰ کلاس شیء میان پنج کلاس صحنه در شکل ۴ آمده است. در این شکل تعلق هر شیء به هر صحنه با یک قاب مستطیل شکل مشخص شده است و میزان شدت رنگ این قاب میزان وزن تعلق یافته به آن شیء خاص را در آن صحنه نشان می‌دهد. همان‌طور که مشاهده می‌شود تعداد کمی از اشیاء میان صحنه‌های مختلف به صورت مشترک وجود دارند و اغلب آن‌ها منحصر به یک صحنه هستند. در مجموعه داده‌ی آزمایش از داده‌های Places به‌طور مستقیم به‌عنوان ورودی شبکه‌ی Place_CNN استفاده شد.

[1] Outdoor

[2] Indoor

[3] http://places.csail.mit.edu/

[4] http://imagenet.stanford.edu /

[5] Annotated Image

برای داده‌های آزمایش شبکه‌ی Object_CNN، با توجه به محدودیت موجود در نحوه‌ی هماهنگی میان ورودی شبکه‌ی بازشناسی شیء و شبکه‌ی بازشناسی صحنه، اشیاء مورد نظر از تصاویر صحنه به‌طور دستی بریده شدند. حداقل ابعاد این تصاویر جدا شده ۱۵۰ پیکسل در نظر گرفته شد (شکل ۳). تصاویر شیء بریده شده به‌عنوان مجموعه داده‌ی آزمون برای شبکه‌ی Object_CNN و صحنه‌های متناظر ورودی شبکه‌ی Place_CNN در نظر گرفته شدند. از این مجموعه داده علاوه بر مقایسه‌ی دقت دو مدل پایه و پیشنهادی، برای انتخاب بهترین روش جهت انتقال اطلاعات مربوط به فراوانی حضور شیء در صحنه‌ی جاری به شبکه‌ی بازشناسی شیء استفاده شد. در بخش بعدی نتایج حاصل از برخی روش‌های ممکن جهت مقداردهی به بردار بازخوردی، مطابق با اطلاعات آماری از فراوانی حضور شیء در صحنه آمده است.

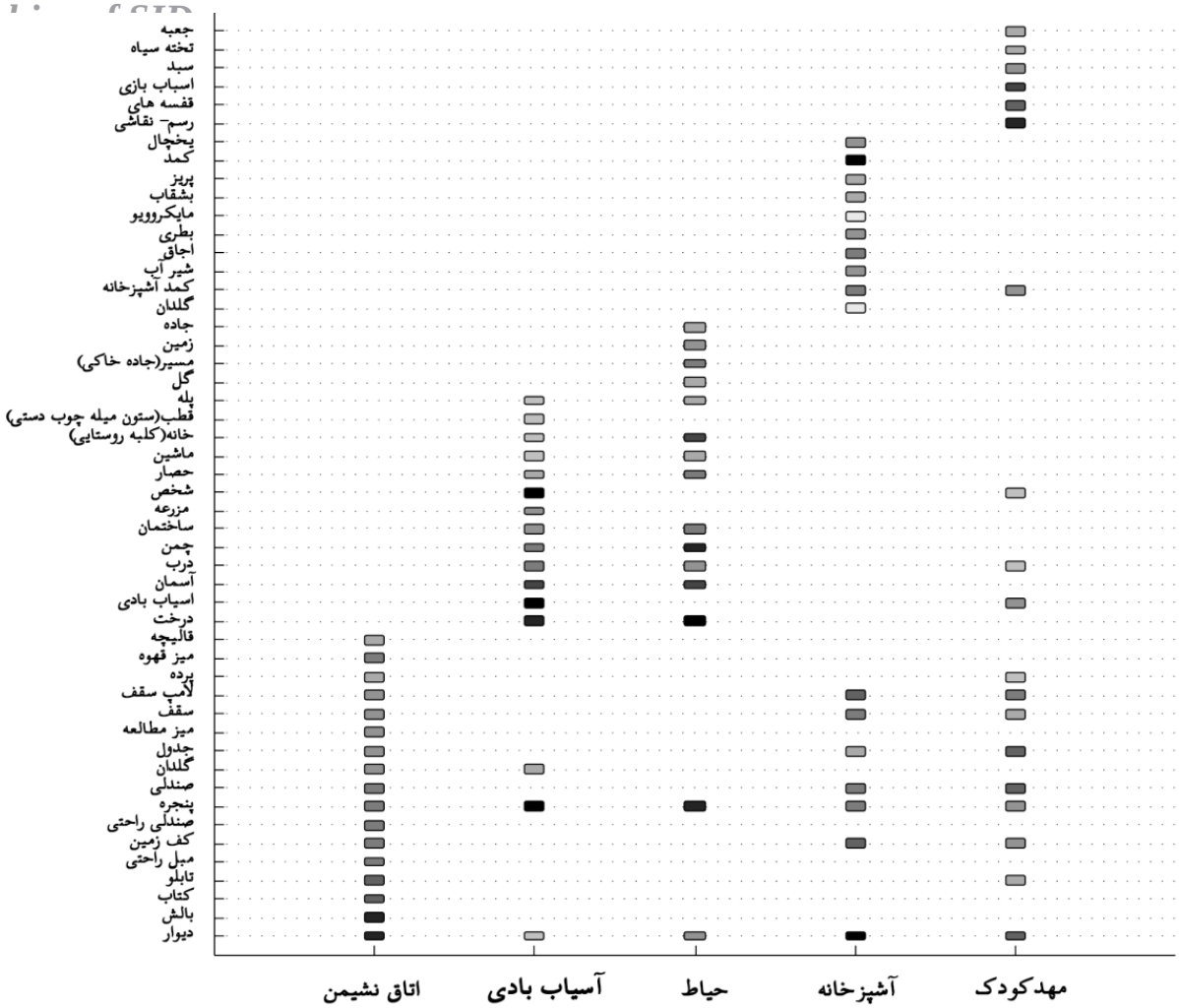
۵-۲ انتخاب بردار بازخورد

همان‌طور که در بخش‌های قبل اشاره شد، پس از تعیین صحنه‌ی ورودی بردار بازخوردی متناسب ایجاد شده و پس از بازخورد در شبکه‌ی بازشناسی شیء، با داده‌های مسیر پیش‌رو تلفیق و مجددا مسیر پیش‌روی شبکه را طی می‌کند. در این آزمایش، با در نظر گرفتن روش‌های متفاوت در مقداردهی به بردار بازخورد، عملکرد مدل مورد بررسی قرار گرفت و در طی چندین آزمایش بردارهای بازخوردی متفاوت امتحان شد. روش‌های متفاوت انتخاب بردار بازخورد به این ترتیب می‌باشد:

- مقادیر غیر صفر بردار، با مقادیر بیشینه و مقادیر صفر با مقدار کمینه‌ی لایه‌ی امتیازها جایگزین شد (بازخورد ثابت).

در اولین تلاش از بازخورد ثابت برای اعمال اطلاعات محتوایی استفاده شد. در این حالت با رویکردی قاطعانه تمام کلاس‌های محتمل با حداکثر احتمال در شبکه تقویت شده و تمام کلاس‌های غیرمحتمل قاطعانه حذف می‌شوند. به این ترتیب کلاس‌های با احتمال کم، شانس برابری در مقابل کلاس‌های با احتمال بالا را دارند. به‌علاوه کلاس‌های غیرمحتمل از دور رقابت حذف می‌شوند.

شکل ۳ مجموعه داده‌ی ایجاد شده برای بررسی مدل. تصاویر شیء به‌صورت دستی از تصاویر صحنه جدا شده و مجموعه داده‌ی آزمایش برای Object_CNN تولید می‌شود. (الف) دو تصویر صحنه و شیء موجود در صحنه‌ها و (ب) و (ج) به ترتیب مجموعه داده‌های تولید شده برای شیء و صحنه، از دو تصویر حاضر در (الف) را نشان می دهند.



شکل ۴ نمودار نمایش توزیع کلاس‌ها میان پنج صحنه. در این شکل مستطیل‌ها نماینده‌ی شیء حاضر در یک کلاس صحنه هستند. میزان رنگ هر مستطیل نشان‌دهنده‌ی میزان فراوانی آن کلاس شیء در آن صحنه‌ی خاص است. همان‌طور که مشاهده می‌شود تعداد کمی از شیء میان صحنه‌های مختلف به صورت مشترک وجود دارند و اغلب اشیاء منحصر به یک صحنه هستند.

با ترکیب این اطلاعات با داده‌های حاصل از ورودی جاری، رقابتی کاملاً عادلانه میان کلاس‌های محتمل برقرار خواهد بود. بنابراین انتظار می‌رود با کاهش تعداد کلاس‌های هدف از تمام کلاس‌های موجود به تعداد کلاس‌های محتمل، دقت مدل افزایش یابد. همچنین تأثیر این نوع بازخورد با شیب تندی ایجاد و با همان سرعت کاهش می‌یابد. اگر رقابت را تنها در میان کلاس‌های محتمل در نظر بگیرید، تمام کلاس‌های محتمل با حداکثر توان هر بار بازخورد می‌شوند و این باعث از بین رفتن اطلاعات ورودی در طی تکرار می‌شود. با توجه به اینکه نتایج زیستی نظریه‌ی مغز بیزی[1] (در نظریه مغز بیز، مغز به عنوان یک سیستم آماری در نظر گرفته می‌شود که با استفاده از اطلاعات پیشین حالت های مختلف ورودی حسی را پیش بینی می‌کند) را در ایجاد سیگنال‌های بازخوردی مطرح می‌کنند [۴۴] در ادامه، اعمال بردار بازخورد در حالت احتمالاتی نیز مورد آزمایش قرار گرفت. در اینجا چند روش از مقداردهی بیزی به بردار بازخورد آورده شده‌است.

- ✓ مقادیر بردار بین مقدار بیشینه و کمینه‌ی لایه‌ی امتیازها نرمال شد (بازخورد ترکیبی).
- ✓ مقادیر بردار بین صفر و یک نرمال شده و مقادیر صفر با مقدار کمینه‌ی بردار لایه‌ی امتیازها جایگزین شد (بازخورد مستقیم).
- ✓ مقادیر بردار پس از نرمال‌سازی به تابع سیگموید داده شده و کمینه‌ی بردار با کمینه‌ی بردار امتیازها جایگزین شد (بازخورد هلالی).

توجه به این نکته ضروری است که تعداد کلاس‌های با احتمال متوسط و همچنین فراوانی حضور آن‌ها در صحنه، اختلاف زیادی با کلاس‌های با احتمال بالا دارند. بنابراین اگرچه حذف کلاس‌های غیرمحتمل و کاهش کلاس‌های هدف باعث بهبود عملکرد مدل در حالت‌های ترکیبی و مستقیم شده است، شانس کلاس‌های با احتمال کم در مقابل کلاس‌های با احتمال بالا بسیار ضعیف است. به همین جهت عملکرد مدل قابل‌توجه نیست. در حالت ترکیبی به

[1] Bayesian Brain Theory

دلیل استفاده از محدوده اعداد بزرگ در بردار، بازخورد باعث کاهش سریع دقت شده است. در آزمایش بعدی با استفاده از تابع سیگموید تلاش شد تا فاصله‌ی میان کلاس‌های محتمل تعدیل شود. انتظار می‌رود که استفاده از این تابع شانس رقابت کلاس‌های با احتمال کم با کلاس‌های با احتمال بالا را افزایش داده و نتایج بهتر و عادلانه‌تری تولید شود؛ زیرا فرض ما این است که ممکن است کلاس برنده در میان کلاس‌های با احتمال کم باشد. عملکرد مدل برای یک تا ۱۰ بار تکرار عمل بازخورد، برای هر چهار بردار بررسی گردید. نتایج حاصل در شکل ۶ آورده شده است. در رقابتی نزدیک، بهترین نتیجه متعلق به بردار بازخورد هلالی در پنج بار تکرار می‌باشد. این بردار در آزمایش‌ها مورد استفاده قرار گرفته است.

همان‌گونه که مشاهده شد، در تمام نمودارها بعد از تعداد دفعاتی تکرار بازخورد دقت مدل افت می‌کند. در واقع با افزایش تکرار بازخورد نقش اطلاعات محتوایی افزایش می‌یابد و گویی بازشناسی با تکیه بر اطلاعات پیشین صورت می‌گیرد و جزئیات حاصل از پردازش‌های مسیر پیش‌رو در نظر گرفته می‌شود. بنابراین دقت مدل شروع به کاهش یافتن می‌کند. مطالعات نشان می‌دهد عدم وجود ارتباط صحیح میان ورودی و نواحی بالایی مغز که در ایجاد سیگنال‌های بالابه‌پایین مبتنی بر محتوا نقش بسزایی ایفا می‌کنند، باعث بروز مشکلات بینایی می‌شود. حتی در برخی دیدگاه‌ها تاکید بیش از حد بر اطلاعات بالابه‌پایین را در ایجاد توهمات بینایی در بیماران اسکیزوفرنی[1] دخیل می‌دانند [۴۵].

بنابراین اگرچه سیگنال‌های بالابه‌پایین باعث بهبود عملکرد مدل می‌شوند اما تأثیر ورودی را نباید از یاد برد.

- نتایج

با استفاده از بهترین حالت دفعات تکرار بازخورد در مدل پیشنهادی، دقت آن در مقایسه با مدل پایه به‌طور معناداری از ۵۵٫۳% به ۶۰% افزایش یافته است (معناداری با استفاده از آزمون رتبه ویلکاکسون[2] و p-value=۰٫۰۰۰۱). نمودار میله‌ای مربوطه در شکل ۵ آمده است.

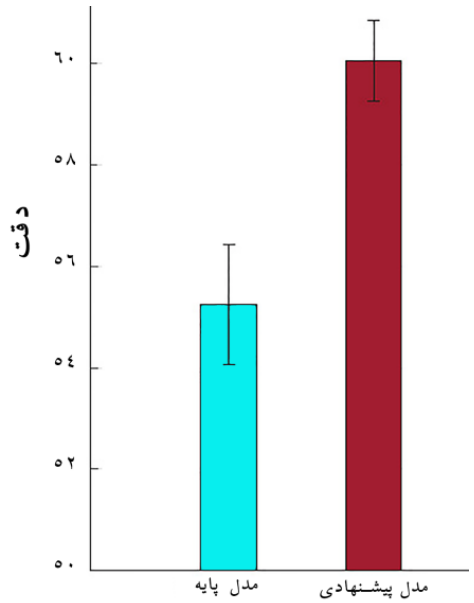
برای مقایسه خروجی دو مدل، ماتریس‌های عدم شباهت[3] (RDM) محاسبه شد (شکل ۷). همان‌طور که مشخص است، این ماتریس‌ها در خروجی مدل پیشنهادی نسبت به مدل پایه تفکیک‌پذیرتر شده‌اند؛ که نقش بردار بازخورد در بهبود بازشناسی و تمییز کلاس‌ها را نشان می‌دهد. به‌علاوه معناداری این بهبود توسط معیار آماری کندال تاو[4] (p-value =۰٫۰۰۰۵۸۲) تایید شد.

۳-۵ بررسی عملکرد مدل پیشنهادی در تصاویر پیچیده

یکی از مسائل مهم در حوزه‌ی بازشناسی، پیچیدگی شیء است. این پیچیدگی در عناوین متفاوتی ازجمله تفاوت در روشنایی، اندازه، چرخش در عمق، پنهان شدن قسمتی از شیء در پشت شیء دیگر، چرخش در صفحه و غیره مطرح می‌شود. این آزمایش سعی دارد تا در تصاویر پیچیده مدل پیشنهادی را در مقایسه با مدل پایه به چالش بکشد. تصاویر در یکی از حالت‌های پیچیدگی قرار می‌گیرد و عملکرد مدل در این شرایط سنجیده می‌شود.

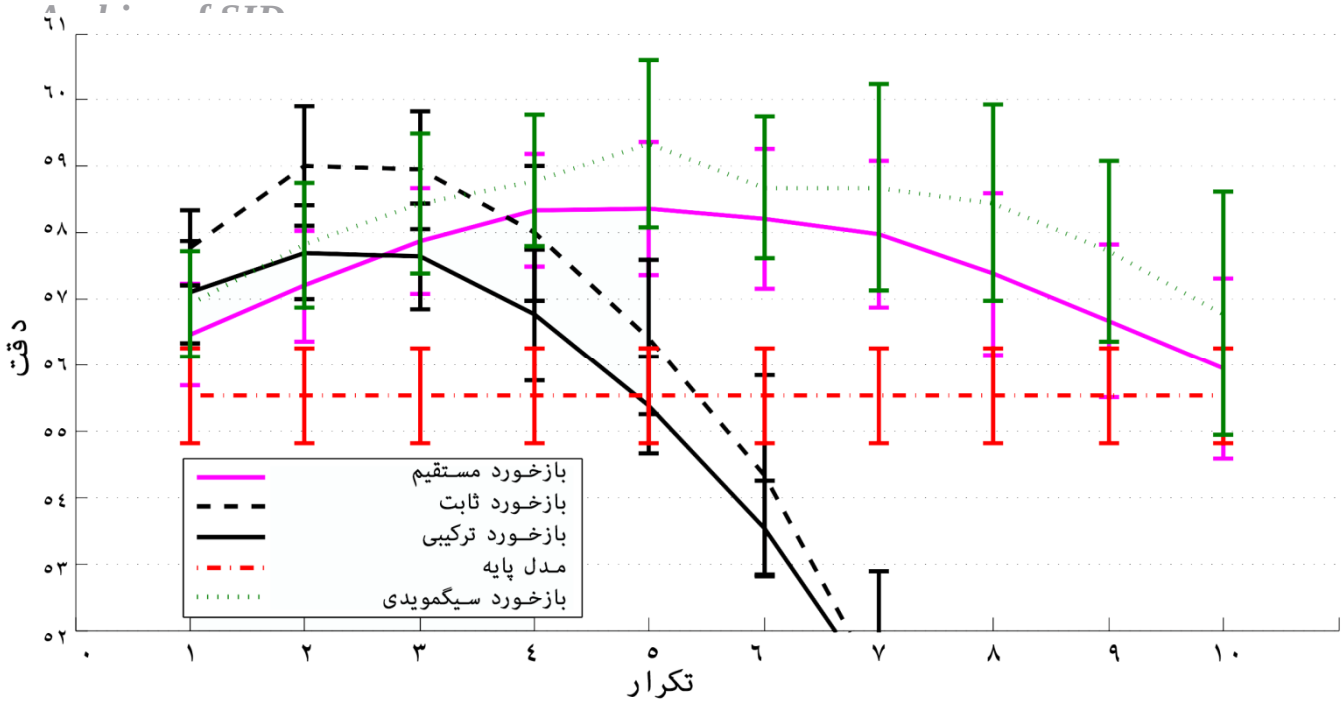
- آزمایش

در این آزمایش مدل پیشنهادی در دو شبکه‌ی Object_CNN و Place_CNN، همانند بخش قبل، به‌ترتیب برای طبقه‌بندی روی ۵۰ کلاس شیء و پنج کلاس صحنه آموزش دیده شده است.



شکل ۵ کارایی مدل پایه و مدل پیشنهادی در بازشناسی ۵۰ کلاس شیء. مدل پیشنهادی به صورت معنی‌داری از مدل پایه بهتر عمل می‌کند (معناداری با استفاده از آزمون رتبه ویلکاکسون و p-value=۰٫۰۰۰۱).

[1] Schizophrenia
[2] Wilcoxon Singed-rank Test
[3] Representational Dissimilarity Matrix
[4] Kendall Tau



شکل ۶ بررسی انواع مقداردهی اولیه‌ی بردار بازخورد. هرچند انواع دیگر بردارهای بازخورد آزمایشی توانسته‌اند نسبت به مدل پایه بهبود معناداری ایجاد نمایند، اما بهترین عملکرد در پنج بار تکرار برای بردار بازخورد هلالی دیده می‌شود.

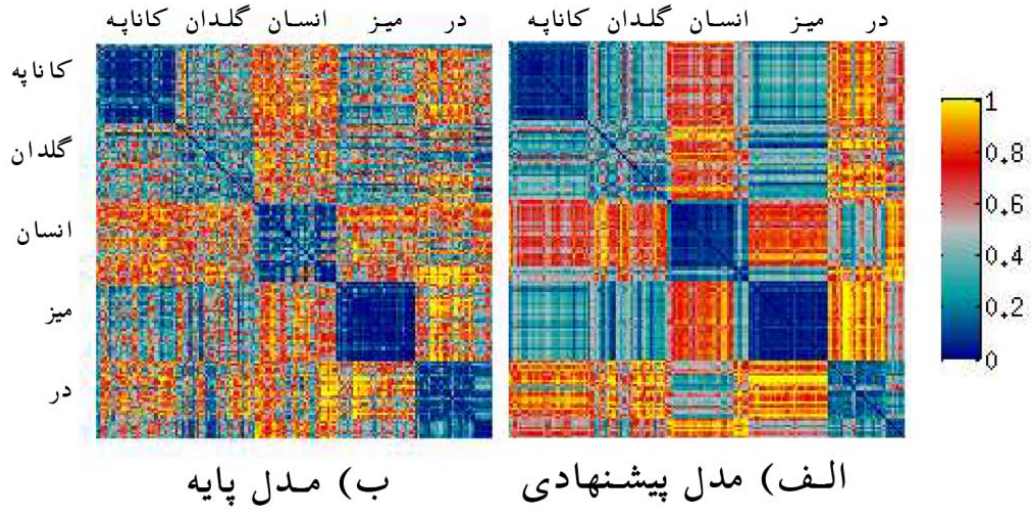
- مجموعه داده

در این آزمایش نیز مجموعه داده‌ی ۵۰ کلاسه‌ی مذکور در بخش قبل مورد استفاده قرار گرفت. برای ایجاد پیچیدگی در شی، تصاویری با ابعاد بزرگ‌تر از ۱۵۰ پیکسل و شامل شیء متمرکز و کامل اختیار شد. این تصاویر در طی آزمایش‌های متفاوت از ۱۲۸×۱۲۸ تا ۸×۸ تغییر اندازه یافته و به‌عنوان ورودی به مدل بازشناسی شیء داده شدند. ابعاد تصاویر صحنه‌ی مربوطه نیز متناسب با تغییر ابعاد تصویر شیء تغییر یافت.

- نتایج

عملکرد مدل در رویارویی با تصاویر تغییر اندازه یافته در شکل ۸ آمده است. همان‌طور که مشاهده می‌شود با افزایش میزان پیچیدگی تصاویر، اثر بازخورد محسوس‌تر می‌شود. زمانی که تصاویر نزدیک به اندازه طبیعی خود هستند و وضوح نسبتا متناسبی دارند، بازشناسی شیء تقریبا با مسیر پیش‌رو قابل حل است. اما زمانی که تصویر به ابعاد کوچک‌تر در می‌آید، مدل پیش‌روی استاندارد به تنهایی قادر به بازشناسی شیء نیست و نیاز آن به اطلاعات بازخوردی افزایش می‌یابد. تا جایی که در ابعاد ۳۲×۳۲ مدل پیشنهادی نسبت به سایر سطوح به‌طور معناداری از مدل پایه پیشی می‌گیرد. با کاهش بیش‌تر وضوح، تصویر به اندازه‌ای کوچک می‌شود که بازخورد نیز نمی‌تواند به بهبود دقت مدل کمک کند. تاجایی که در دو سطح ۸×۸ و ۱۶×۱۶ پس از اعمال بازخورد دقت مدل پایه به‌طور معناداری بهبود نمی‌یابد (معناداری با استفاده از آزمون رتبه ویلکاکسون و p-value برای سه سطح اول تقریبا ۰٫۰۰۰۱ و برای دو سطح چهارم و پنجم به ترتیب ۰٫۰۳۶۹ و ۰٫۰۲۰۲ محاسبه شد). در نهایت مدل پیشنهادی در وضوح ۸×۸ با مدل پایه هم‌روند می‌شود. تصاویر در این اندازه برای انسان نیز به سختی قابل تشخیص است.

این آزمایش تأثیر محتوا را در بازشناسی شیء نشان می‌دهد. اگرچه نقش محتوا در بازشناسی شیء با وضوح بالا پررنگ



شکل ۷ نمودار RDM برای لایه انتهایی هر دو مدل پایه و مدل پیشنهادی برای پنج کلاس از آزمایش طبقه‌بندی ۵۰ کلاس. همان‌طور که به صورت بصری نیز مشخص است خروجی بعد از بازخورد نسبت به مدل پایه به‌طور معناداری بهبود یافته است (بررسی معناداری توسط آزمون کندال تآو و p-value ۰,۰۰۰۵۸۲ =).

نیست، اما در تصاویر با وضوح کم میزان اهمیت آن افزایش می‌یابد. در واقع اطلاعات محتوایی در مدل پیشنهادی، در وضوح پایین بهبود بیش‌تری در مدل پایه ایجاد می‌کند. مطابق با این نتیجه، نتایج آزمایش‌های رفتاری و تصویربرداری عصبی نشان می‌دهند انسان برای بازشناسی شیء در وضوح پایین نیاز بیش‌تری به اطلاعات محتوا دارد [۴۶] و [۴۷].

۵-۴ بررسی مدل پیشنهادی در بازشناسی تصاویر متجانس و نامتجانس

شواهد نشان می‌دهد که میزان تأثیر محتوا در بازشناسی شیء بسیار حائز اهمیت است [۲۳] و [۳۲] و [۴۸]. در این بخش تلاش می‌شود تناسب و تعادل میان صحنه و شیء در مدل پیشنهادی بررسی و با نتایج رفتاری مقایسه گردد. برای «بررسی نقش محتوا در بازشناسی شیء» در اغلب مطالعات در دو دسته تصاویر صحنه‌ی طبیعی[1] در مقابل صحنه‌ی بشری[2] برای ساخت تصاویر متجانس و نامتجانس در دسته‌بندی حیوان و غیرحیوان به کار گرفته می‌شود [۴۹] و [۵۰]. این انتخاب به دو دلیل عمده صورت می‌گیرد. اول توجه به این واقعیت که حیوان‌ها بیش‌تر در صحنه‌های طبیعی و شیء بشری در صحنه‌های بشری ظاهر می‌شوند. دوم اینکه پردازش‌های محتوایی در طبقه‌بندی جزئی‌تر نیاز به زمان بیشتری برای پردازش دارند [۵۱]. بنابراین این نوع تصاویر برای بررسی تأثیر محتوا مناسب به نظر می‌رسند.

- آزمایش

به جهت آزمودن مدل پیشنهادی مجموعه داده‌هایی مورد استفاده قرار گرفتند که بتوانند میزان تأثیر محتوا در بازشناسی شیء را به چالش بکشند. بنابراین مدل Place_CNN برای دو محتوای متفاوت از یکدیگر و متناسب برای کلاس‌های حیوان- غیرحیوان (یعنی طبیعی و بشری) آموزش دید. به دلیل حفظ تنوع تصاویر، برای هر کلاس اعم از کلاس‌های شیء و یا صحنه، از چند طبقه مرتبط استفاده شد. در شکل ۹ تنوع طبقه‌های متفاوت در هر کلاس را مشاهده می‌کنید.

- مجموعه داده

برای ایجاد داده‌های آزمون شبکه، حدود ۳۰ داده از هریک از دو دسته (حیوان- غیرحیوان، طبیعی- بشری) به‌صورت تصادفی انتخاب شد. این داده‌ها از تصاویری ساخته می‌شود که شیء هدف در مرکز قرار گرفته و تقریبا شیء مربوط دیگری در کل تصویر وجود ندارد. تصاویر از مجموعه داده‌ی Places انتخاب شده و هر کدام یک صحنه به حساب می‌آیند. سپس به‌صورت دستی تصاویر شیء این بار بدون در نظر گرفتن حاشیه‌ای خارج از محیط شیء، اختیار شده و به‌طور تصادفی بر روی صحنه‌های کامل متجانس و غیر متجانس قرار گرفتند. نمونه‌ای از این تصاویر در شکل ۹ آمده است. در نهایت، از هر کلاس ۳۰ نمونه ایجاد و هر شیء در مرکز صحنه‌های متجانس و نامتجانس قرار گرفت. نسبت تعداد پیکسل‌های شیء به تعداد پیکسل‌های صحنه برای تمام تصاویر یکسان است. تصویر صحنه به شبکه‌ی Place_CNN و تصویر صحنه‌ی حاوی شیء به Object_CNN داده شد.

- نتایج

طی دو آزمایش نتایج مدل برای تصاویر شیء و صحنه متجانس و تصاویر شیء و صحنه نامتجانس بررسی شد.

✓ بررسی از لحاظ کارایی و دقت

شکل ۱۰ عملکرد مدل پیشنهادی و مدل پایه را در حالات متفاوت ورودی و صحنه به نمایش می‌گذارد. در حالت متجانس

[1] Natural
[2] Man Made

افزودن اطلاعات محتوایی به مدل پایه باعث بهبود دقت و در تصاویر نامتجانس، سبب افت دقت می‌شود. تأثیر در هر دو دسته‌ی داده‌ها معنادار است (معناداری با استفاده از آزمون رتبه ویلکاکسون و p-value برای حالت متجانس ۰,۰۰۰۰۴۱۱ و برای حالت نامتجانس ۰,۰۱۶۷). این رفتار مدل با نتایج مطالعات رفتاری [۲۲] و [۲۳] و [۴۹] و [۵۰] و [۵۲] که افت دقت انسان درصحنه‌های نامتجانس را نسبت به حالت متجانس نشان می‌دهند، مطابقت دارد. به نظر می‌رسد مدل پایه که در اینجا می‌توان آن را نماینده‌ای از مدل‌های کانولوشنی در نظر گرفت، به پس‌زمینه‌ی شیء ورودی بسیار حساس بوده به طوری که با مشاهده‌ی شیء در صحنه‌های نامتجانس افت دقتی بیش از ۲۰ درصد از خود نشان می‌دهد.

این افت شدید بیان می‌کند که مدل پایه برای بازشناسی شیء، تکیه‌ی زیادی بر ویژگی‌های صحنه‌ی آن دارد. هرچند شیء در مجموعه داده‌ی آموزشی در مرکز تصویر واقع شده بودند و تقریبا مقدار قابل توجه‌ای از تصویر را تشکیل می‌دادند، مدل پایه از ویژگی‌های صحنه بسیار بهره برده است. با این حال مشاهده می‌شود، رفتار مدل در مواجه با تصاویر نامتجانس مشابه نتایج گزارش شده‌ای است که رفتار انسان را در آزمایش‌های مرتبط اعلام کردند [۴۹] و [۵۰] و [۵۲] .به‌علاوه در تصاویر متجانس با وجود دقت بالای مدل پایه، اطلاعات محتوایی توانسته‌اند دقت آن را به طور معناداری افزایش دهند.

جدول ۱ کلاس‌های انتخاب شده برای آموزش شبکه‌های Object_CNN و Place_CNN در آزمایش طبقه‌بندی دو کلاس در حالت‌های متجانس و نامتجانس. جهت ایجاد تنوع در داده‌های آموزشی هر شبکه از چند کلاس متفاوت را در هر دسته مشاهده می‌کند.

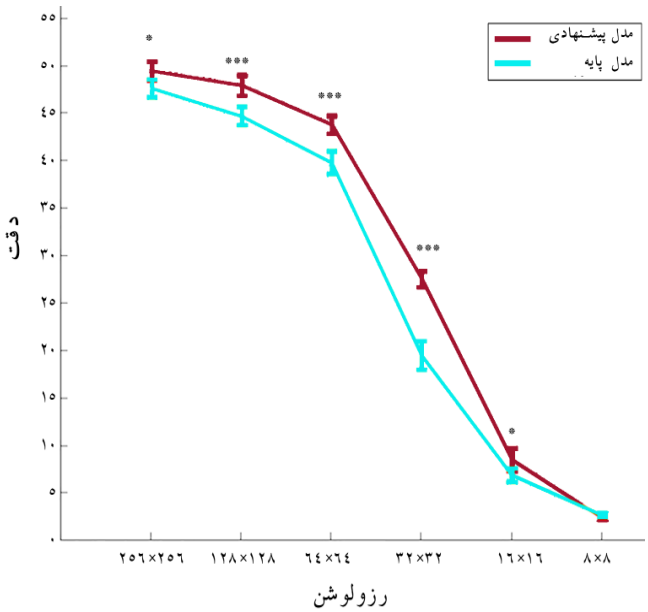
غیرحیوان	حیوان	بشری	طبیعی
نردبان	گوزن	راهرو	گودال آب
بطری	فیل	آشپزخانه	چشمه
ماکرویو	حیوانات مزرعه	رختکن	آبشار
مبل راحتی	زرافه	مهمان‌سرا	-

✓ بررسی از لحاظ میزان جدایی‌پذیری

سرعت بازشناسی تصاویر متجانس در آزمایش‌های رفتاری کمتر از حالت نامتجانس گزارش شده است [۴۹] و [۵۰] و [۵۲]. با توجه به نتایج تحقیقات به‌نظر می‌رسد تفکیک‌پذیر بودن نمونه‌ها نشان‌دهنده‌ی ساده‌تر بودن مسئله است که باعث افزایش سرعت بازشناسی خواهد بود. بنابراین پیش‌بینی می‌شود افزودن اطلاعات محتوایی به مدل پایه باعث بهبود تفکیک‌پذیری در تصاویر متجانس و تخریب آن در تصاویر نامتجانس شود. جهت بررسی این ویژگی در مدل با استفاده از جعبه‌ابزار Drtoolbox در نرم‌افزار متلب خروجی دو مدل پایه و مدل پیشنهادی در ابعاد کوچک بازنمایی شد.

با استفاده از الگوریتم تحلیل مؤلفه‌های اصلی[1] (PCA) ابعاد بردار خروجی هر دو مدل کاهش داده شد. نمایش دو بعدی خروجی این الگوریتم در شکل ۱۱ آورده شده است. نقاط قرمز مربوط به نمونه‌های کلاس غیرحیوان و نقاط آبی مربوط به نمونه‌های کلاس حیوان می‌باشند. هر نقطه نمایانگر یک نمونه‌ی آزمون است.

مشاهده می‌شود که خروجی هر دو مدل در تصاویر متجانس نسبت به تصاویر نامتجانس قابل‌تمییزتر هستند (شکل ۱۱ الف و ب). به‌علاوه همان‌طور که پیش‌بینی می‌شد اطلاعات محتوایی در تصاویر نامتجانس باعث کاهش تفکیک‌پذیری داده‌ها شده است (شکل ۱۱ ج و د). اما در حالت متجانس این اطلاعات بهبود چشم‌گیری در موقعیت نمونه‌ها در فضای بازنمایی نداشته‌اند. این عدم بهبود می‌تواند به دلیل سادگی فضای مسئله باشد که نمونه‌های دو دسته قبل از اعمال بازخورد تفکیک پذیری قابل قبولی دارند.



شکل ۸ به چالش کشیدن مدل پیشنهادی و مدل پایه در تصاویر با وضوح‌های متفاوت. در این آزمایش مدل به‌طور جداگانه روی داده‌های آماده شده مورد آزمایش قرار می‌گیرد. همان‌طور که در شکل مشخص است با پیچیده شدن و کاهش بعد تصاویر میزان تأثیر بازخورد نمایان‌تر می‌شود. حداکثر تأثیرگذاری بردارهای بازخوردی در تصاویر ۳۲×۳۲ دیده می‌شود.

[1] Principal Component Analysis

۶ نتیجه‌گیری

در این مقاله بر نقش اطلاعات انجمنی و محتوایی در بازشناسی شیء اشاره شد. سپس مدلی با الهام از نقش سیگنال‌های بالابه‌پایین محتوا، جهت بازشناسی شیء ارائه شد. در این مدل با استفاده از پاسخ شبکه‌ای مجزا (Place_CNN) که درباره‌ی صحنه‌ی شیء مورد نظر قضاوت می‌کرد، بردارهای بازخوردی فعال شدند. این بردارها حاوی اطلاعات انجمنی مرتبط با حضور شیء در صحنه‌ها بودند که تحقیقات تأثیر به سزای این داده‌های محتوایی را در بازشناسی شیء نشان دادند. افزودن این اطلاعات به شبکه پایه‌ی Object_CNN باعث بهبود معنی‌دار تقریبا شش درصدی کارایی آن در مسئله‌ی طبقه‌بندی ۵۰ کلاسه گردید.

شواهد نشان داده‌اند که سیگنال‌های بازخوردی بالابه‌پایین باعث بهبود بازشناسی شیء، به خصوص در تصاویر پیچیده می‌شوند [۱۳] و [۴۸]. نتایج آزمایش‌ها در این مقاله نیز با تاکید بر این واقعیت افزایش اثر اطلاعات محتوایی را در افزایش کارایی بازشناسی شیء به ویژه در تصاویر پیچیده نشان داد.

به‌علاوه مطابق با نتایج حاصل از آزمایش‌های رفتاری، انسان شیء قرار گرفته در صحنه‌ی متجانس را با سرعت و دقت بیشتری نسبت به شیء قرار گرفته در صحنه‌ی نامتجانس بازشناسی می‌کند [۲۰]. در آزمایش نهایی نیز مشاهده شد که هم‌سو با آزمایش‌های رفتاری، اطلاعات محتوایی ناصحیحی که در مدل پایه بازخورد شدند باعث افت دقت، و اطلاعات محتوایی صحیح بازخوردی باعث بهبود عملکرد مدل پایه می‌شود.

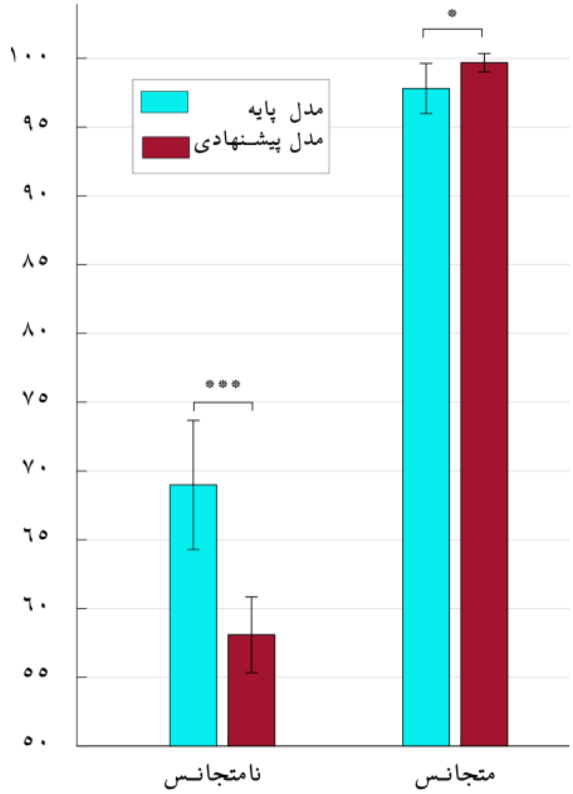
(الف) متجانس (ب)نامتجانس

شکل ۹ چهار نوع داده‌ی مورد استفاده در بازشناسی تصاویر متجانس و نامتجانس. (الف) شیء حیوان روی صحنه‌ی طبیعی و شیء غیر حیوان روی صحنه‌ی بشری به‌عنوان تصاویر متجانس. (ب) شیء حیوان روی صحنه‌ی بشری، شیء غیرحیوان روی صحنه طبیعی به‌عنوان تصاویر نامتجانس.

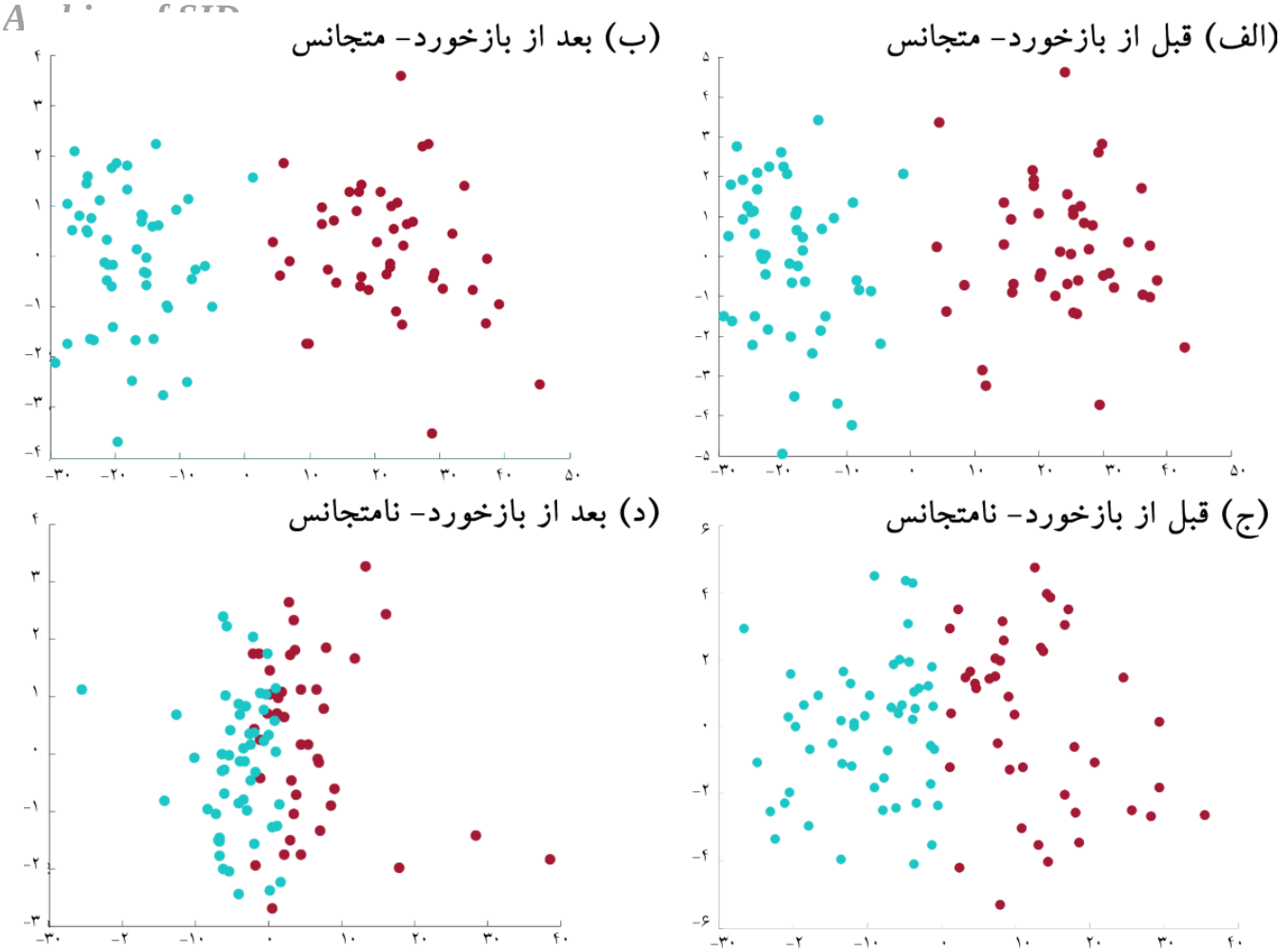
همچنین تفکیک‌پذیری کلاس‌ها در خروجی دو مدل در تصاویر نامتجانس بعد از بازخورد کاهش یافته است که می‌توان آن را معادل سخت‌تر شدن مسئله و در نهایت کاهش سرعت بازشناسی دانست. انتظار می‌رود که اطلاعات محتوایی باعث افزایش تفکیک‌پذیری داده‌ها در تصاویر متجانس شود. اما در حالت متجانس اطلاعات بازخوردی بهبود چشم‌گیری در میزان تفکیک‌پذیری نمونه‌ها در فضای بازنمایی نداشته‌اند که می‌تواند به دلیل سادگی فضای مسئله باشد که نمونه‌های دو دسته قبل از اعمال بازخورد تفکیک پذیری قابل قبولی دارند.

۷ تقدیر و تشکر

در این بخش از دانشگاه تربیت دبیر شهید رجایی به جهت امکاناتی که در راستای انجام این پروژه برای ما فراهم آوردند نهایت تشکر و قدردانی را می‌نماییم.



شکل ۱۰ بررسی عملکرد مدل در تصاویر متجانس و نامتجانس. نمودارهای آبی مربوط به عملکرد مدل پایه (قبل از اعمال بازخورد) و نمودارهای قرمز مربوط به عملکرد مدل پیشنهادی (بعد از اعمال بازخورد) می‌باشند. مشاهده می‌شود که داده‌های بازخوردی زمانی که شیء در صحنه‌ی متجانس قرار دارد باعث بهبود معنادار عملکرد مدل شده و در حالتی که شیء در صحنه‌ی نامتجانس است، سیگنال‌های بازخوردی که مخالف انتظارات و اطلاعات ذخیره شده در مغز هستند، سبب افت دقت مدل پایه می‌شوند.



شکل ۱۱ تفکیک‌پذیری نمونه‌ی کلاس‌ها در تصاویر متجانس و نامتجانس. نقاط قرمز مربوط به نمونه‌های کلاس غیرحیوان و نقاط آبی مربوط به نمونه‌های کلاس حیوان می‌باشند. هر نقطه نمایانگر یک نمونه‌ی آزمون است. (الف) و (ب) بازنمایی خروجی نهایی برای تصاویر متجانس را به ترتیب در دو مدل قبل از بازخورد و بعد از آن را نشان می‌دهند. (ج) و (د) بازنمایی خروجی نهایی برای تصاویر نامتجانس را به ترتیب در دو مدل قبل از بازخورد و بعد از آن نشان می‌دهند. مشاهده می‌شود که در حالت متجانس محتوا به تفکیک‌پذیری هرچه بیشتر نمونه‌ها کمک نموده است. همان‌طور که انتظار می‌رفت افزودن داده‌های محتوایی ناصحیح در تصاویر نامتجانس باعث دشوار شدن مسئله و در نتیجه افت دقت و تلفیق بیشتر داده‌ها شده است.

۸ منابع

[1] D. Wyatte, D. J. Jilk, and R. C. O'Reilly, "Early recurrent feedback facilitates visual object recognition under challenging conditions," 2014.

[2] C. D. Gilbert and W. Li, "Top-down influences on visual processing," *Nature Reviews Neuroscience,* vol. 14, no. 5, pp. 350-363, 2013.

[3] M. Ghodrati, A. Farzmahdi, K. Rajaei, R. Ebrahimpour, and S.-M. Khaligh-Razavi, "Feedforward object-vision models only tolerate small image variations compared to human," *Frontiers in computational neuroscience,* vol. 8, p. 74, 2014.

[4] H. Karimi-Rouzbahani, N. Bagheri, and R. Ebrahimpour, "Hard-wired feed-forward visual mechanisms of the brain compensate for affine variations in object recognition," *Neuroscience,* vol. 349, pp. 48-63, 2017.

[5] K. Rajaei, Y. Mohsenzadeh, R. Ebrahimpour, and S.-M. Khaligh-Razavi, "Beyond Core Object Recognition: Recurrent processes account for object recognition under occlusion," *bioRxiv,* p. 302034, 2018.

[6] P. C. Klink, B. Dagnino, M.-A. Gariel-Mathis, and P. R. Roelfsema, "Distinct Feedforward and Feedback Effects of Microstimulation in Visual Cortex Reveal Neural Mechanisms of Texture Segregation," *Neuron,* vol. 95, no. 1, pp. 209-220. e3, 2017.

[7] V. A. Lamme and P. R. Roelfsema, "The distinct modes of vision offered by feedforward and recurrent processing," *Trends in neurosciences,* vol. 23, no. 11, pp. 571-579, 2000.

[8] H. Kafaligonul, B. G. Breitmeyer, and H. Öğmen, "Feedforward and feedback processes in vision," *Frontiers in psychology,* vol. 6, p. 279, 2015.

[9] C. Boehler, M. Schoenfeld, H.-J. Heinze, and J.-M. Hopf, "Rapid recurrent processing gates awareness in primary visual cortex," *Proceedings of the National Academy of Sciences,* vol. 105, no. 25, pp. 8742-8747, 2008.

[10] P. E. Roland *et al.*, "Cortical feedback depolarization waves: a mechanism of top-down influence on early visual areas," *Proceedings of the National Academy of Sciences,* vol. 103, no. 33, pp. 12586-12591, 2006.

[11] P. E. Roland, "Six principles of visual cortical dynamics," *Frontiers in systems neuroscience,* vol. 4, p. 28, 2010.

[12] J. J. Fahrenfort, H. S. Scholte, and V. A. Lamme, "Masking disrupts reentrant processing in human visual cortex," *Journal of cognitive neuroscience,* vol. 19, no. 9, pp. 1488-1497, 2007.

[13] J. Fahrenfort, H. Scholte, and V. Lamme, "The spatiotemporal profile of cortical processing leading up to visual perception," *Journal of Vision,* vol. 8, no. 1, pp. 12-12, 2008.

[14] K. G. Thompson, K. L. Biscoe, and T. R. Sato, "Neuronal basis of covert spatial attention in the frontal eye field," *Journal of Neuroscience,* vol. 25, no. 41, pp. 9479-9487, 2005.

[15] S. L. Bressler, W. Tang, C. M. Sylvester, G. L. Shulman, and M. Corbetta, "Top-down control of human visual cortex by frontal and parietal cortex in anticipatory visual spatial attention," *Journal of Neuroscience,* vol. 28, no. 40, pp. 10056-10061, 2008.

[16] J. N. McManus, W. Li, and C. D. Gilbert, "Adaptive shape processing in primary visual cortex," *Proceedings of the National Academy of Sciences,* vol. 108, no. 24, pp. 9739-9746, 2011.

[17] C. Summerfield and T. Egner, "Expectation (and attention) in visual cognition," *Trends in cognitive sciences,* vol. 13, no. 9, pp. 403-409, 2009.

[18] S. Trapp and M. Bar, "Prediction, context, and competition in visual recognition," *Annals of the New York Academy of Sciences,* vol. 1339, no. 1, pp. 190-198, 2015.

[19] M. Sherman, A. Seth, A. Barrett, and R. Kanai, "Prior expectations facilitate metacognition for perceptual decision," *Consciousness and cognition,* vol. 35, pp. 53-65, 2015.

[20] E. M. Aminoff, K. Kveraga, and M. Bar, "The role of the parahippocampal cortex in cognition," *Trends in cognitive sciences,* vol. 17, no. 8, pp. 379-390, 2013.

[21] A. Oliva and A. Torralba, "The role of context in object recognition," *Trends in cognitive sciences,* vol. 11, no. 12, pp. 520-527, 2007.

[22] M. Bar, "Visual objects in context," *Nature Reviews Neuroscience,* vol. 5, no. 8, pp. 617-629, 2004.

[23] A.-C. Collet, D. Fize, and R. VanRullen, "Contextual Congruency Effect in Natural Scene Categorization: Different Strategies in Humans and Monkeys (Macaca mulatta)," *PloS one,* vol. 10, no. 7, p. e0133721, 2015.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[25] R. M. Cichy, A. Khosla, D. Pantazis, and A. Oliva, "Dynamics of scene representations in the human brain revealed by magnetoencephalography and deep neural networks," *NeuroImage,* vol. 153, pp. 346-358, 2017.

[26] W. Yu, K. Yang, Y. Bai, H. Yao, and Y. Rui, "Visualizing and Comparing Convolutional Neural Networks," *arXiv preprint arXiv:1412.6631,* 2014.

[۲۷] عباسی، فریبا و ابراهیم‌پور، رضا و رجایی، کریم، "مدل‌سازی محاسباتی جداسازی شیء هدف از پس‌زمینه در بازشناسی اشیاء با الهام سیستم بینایی انسان"، مجله ماشین بینایی و پردازش تصویر، دوره ٤، شماره ١، ص ١٦-١، تابستان ٩٦.

[41] S. Tscheehne and H. Neumann, "Hierarchical representation of shapes in visual cortex—from localized features to figural shape segregation," *Frontiers in computational neuroscience,* vol. 8, 2014.

[42] J. Xiao, J. Hays, K. A. Ehinger, A. Oliva, and A. Torralba, "Sun database: Large-scale scene recognition from abbey to zoo," in *Computer vision and pattern recognition (CVPR), 2010 IEEE conference on*, 2010, pp. 3485-3492: IEEE.

[43] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 248-255: IEEE.

[44] R. Kanai, Y. Komura, S. Shipp, and K. Friston, "Cerebral hierarchies: predictive processing, precision and the pulvinar," *Phil. Trans. R. Soc. B,* vol. 370, no. 1668, p. 20140169, 2015.

[45] H. Tsukada, H. Fujii, K. Aihara, and I. Tsuda, "Computational model of visual hallucination in dementia with Lewy bodies," *Neural Networks,* vol. 62, pp. 73-82, 2015.

[46] D. Parikh, C. L. Zitnick, and T. Chen, "Exploring tiny images: The roles of appearance and contextual information for machine and human object recognition," *IEEE transactions on pattern analysis and machine intelligence,* vol. 34, no. 10, pp. 1978-1991, 2012.

[47] T. Brandman and M. V. Peelen, "Interaction between scene and object processing revealed by human fMRI and MEG decoding," *Journal of Neuroscience,* vol. 37, no. 32, pp. 7700-7710, 2017.

[48] L. J. Jenkins, Y.-J. Yang, J. Goh, Y.-Y. Hong, and D. C. Park, "Cultural differences in the lateral occipital complex while viewing incongruent scenes," *Social Cognitive and Affective Neuroscience,* vol. 5, no. 2-3, pp. 236-241, 2010.

[49] J. J. Summerfield, J. Lepsien, D. R. Gitelman, M. M. Mesulam, and A. C. Nobre, "Orienting attention based on long-term memory experience," *Neuron,* vol. 49, no. 6, pp. 905-916, 2006.

[50] A. F. Rossi, R. Desimone, and L. G. Ungerleider, "Contextual modulation in primary visual cortex of macaques," *the Journal of Neuroscience,* vol. 21, no. 5, pp. 1698-1709, 2001.

[51] M. Poncet and M. Fabre - Thorpe, "Stimulus duration and diversity do not reverse the advantage for superordinate - level representations: the animal is seen before the bird," *European Journal of Neuroscience,* vol. 39, no. 9, pp. 1508-1516, 2014.

[52] J. L. Davenport and M. C. Potter, "Scene consistency in object and background perception," *Psychological Science,* vol. 15, no. 8, pp. 559-564, 2004.

[28] J. H. Bappy and A. K. Roy-Chowdhury, "Inter-dependent CNNs for joint scene and object recognition," in *Pattern Recognition (ICPR), 2016 23rd International Conference on*, 2016, pp. 3386-3391: IEEE.

[29] A. Torralba, K. P. Murphy, W. T. Freeman, and M. A. Rubin, "Context-based vision system for place and object recognition," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, 2003, pp. 273-280: IEEE.

[30] P. Duho, "CONTEXTUAL OBJECT CATEGORIZATION WITH ENERGY-BASED MODEL," *The Journal of Engineering,* vol. 1, no. 1, 2017.

[31] T. H. Cornelissen and M. L.-H. Võ, "Stuck on semantics: Processing of irrelevant object-scene inconsistencies modulates ongoing gaze behavior," *Attention, Perception, & Psychophysics,* vol. 79, no. 1, pp. 154-168, 2017.

[32] S. Vanmarcke, F. Calders, and J. Wagemans, "The Time-Course of Ultrarapid Categorization: The Influence of Scene Congruency and Top-Down Processing," *i-Perception,* vol. 7, no. 5, p. 2041669516673384, 2016.

[33] D. Crafa, C. Hawco, and M. B. Brodeur, "Heightened Responses of the Parahippocampal and Retrosplenial Cortices during Contextualized Recognition of Congruent Objects," *Frontiers in behavioral neuroscience,* vol. 11, 2017.

[34] O. S. Cheung and M. Bar, "Visual prediction and perceptual expertise," *International Journal of Psychophysiology,* vol. 83, no. 2, pp. 156-163, 2012.

[35] Y. Zhu, R. Urtasun, R. Salakhutdinov, and S. Fidler, "segdeepm: Exploiting segmentation and context in deep neural networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4703-4711.

[36] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," *arXiv preprint arXiv:1312.6229,* 2013.

[37] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580-587.

[38] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE,* vol. 86, no. 11, pp. 2278-2324, 1998.

[39] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature,* vol. 521, no. 7553, pp. 436-444, 2015.

[40] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in *Advances in neural information processing systems*, 2014, pp. 487-495.

الهه‌سادات سلطاندوست مدرک کارشناسی خود را به عنوان دانشجوی ممتاز در رشته مهندسی کامپیوتر گرایش نرم‌افزار در سال ۱۳۹۳ از دانشگاه شهید رجایی دریافت نمود. سپس ایشان مدرک کارشناسی ارشد خود را در رشته مهندسی کامپیوتر گرایش نرم‌افزار در سال ۱۳۹۶ به عنوان دانشجوی ممتاز از دانشگاه شهید رجاییکسب نمود. علاقه‌مندی‌های علمی ایشان شامل علوم اعصاب بینایی، بینایی‌ماشین، مدل‌سازی شناختی است.

رضا ابراهیم پور استاد تمام گروه هوش مصنوعی، دانشکده مهندسی کامپیوتر دانشگاه تربیت دبیر شهید رجایی می باشند. 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 به عنوان 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 به عنوان پژوهشگر ارشد با پژوهشگاه دانش‌های بنیادی همکاری پژوهشی دارند. آقای دکتر ابراهیم پور بیش از ۱۰۰ مقاله علمی در مجلات و کنفرانس های علمی ارائه نموده اند و همچنین در کمیته علمی و داوری متجاوز از ۲۰ مجله و کنفرانس علمی فعالیت داشته اند. ایشان سرگروه داوری گروه مکاترونیک جشنواره جوان خوارزمی می باشند و بعلاوه از منتخبین سرآمدان علمی کشور توسط فدراسیون سرآمدان علمی کشور در سال ۱۳۹۴ می باشند. زمینه های تخصصی ایشان عبارتند از: علوم اعصاب شناختی، مدلسازی شناختی، بینایی انسان و ماشین.

کریم رجایی مدرک کارشناسی خود را در رشته علوم کامپیوتر در سال ۱۳۸۸ از دانشگاه قم دریافت نمود. مدرک کارشناسی ارشد خود را در رشته علوم کامپیوتر گرایش سیستم های هوشمند در سال ۱۳۹۰ از دانشگاه امیرکبیر دریافت نمود. ایشان از سال ۱۳۹۱ تا ۱۳۹۳ به عنوان محقق در پژوهشگاه دانش های بنیادی (IPM) مشغول به فعالیت پژوهشی بوده اند و از سال ۱۳۹۳ تا اکنون دانشجوی مقطع دکتری تخصصی در رشته علوم اعصاب شناختی، پژوهشگاه دانش های بنیادی هستند. علاقه‌مندی‌های علمی ایشان شامل علوم اعصاب، مدلسازی محاسباتی، یادگیری ماشین و فلسفه ذهن است.