

مدل‌سازی محاسباتی جداسازی شیء هدف از پس‌زمینه در بازشناسی اشیاء با الهام سیستم بینایی انسان

فریبا عباسی^۱، رضا ابراهیم‌پور^۲ و کریم رجایی^۳

چکیده

قرار گرفتن شیء در پس‌زمینه باعث پیچیده‌شدن مسئله بازشناسی اشیاء و در نتیجه افت عملکرد مدل‌های محاسباتی بینایی می‌شود. در حالی که انسان‌ها علی‌رغم این پیچیدگی، شیء هدف را با دقت و سرعت زیادی که متأثر از ارتباطات جانبی و بازخورد از نواحی بالاتر بینایی است بازشناسی می‌کنند.

یکی از مدل‌های بینایی که اخیراً به عملکرد چشمگیری در بازشناسی اشیاء دست یافته است، شبکه عصبی کانولوشنی است که مسیر پیش‌خور بینایی را شبیه‌سازی می‌کند. در این مقاله مدلی بازگشتی بر پایه‌ی این مدل و با الهام از یافته‌های بیولوژیک ارائه شده است که شامل اتصال‌های بازخوردی از نواحی بالاتر و همچنین اتصال‌های جانبی در همان لایه است. برای ارزیابی مدل از مجموعه داده‌ی پنج دسته‌ای، شامل تصاویر دارای پس‌زمینه و بدون پس‌زمینه، استفاده شد. با بصری‌سازی بازنمایی‌هایی ایجاد شده در لایه‌های مدل مشاهده شد که با پیش‌روی در لایه‌های مدل، پس‌زمینه‌ی بیشتری از تصویر ورودی حذف می‌شود. سپس با انجام آزمایش‌هایی نشان داده شد که مدل بازگشتی با سازوکارهای پیشنهادی بازخوردی از نواحی بالاتر و سرکوب پیرامون باعث بهبود معنی‌دار عملکرد مدل، در حذف پس‌زمینه‌ی شیء هدف و در نتیجه بازشناسی اشیاء می‌شود. با توجه به نتایج، در حالی که هر دو سازوکار پیشنهادی همزمان به مدل افزوده شدند، این افزایش عملکرد بیشتر بود که این یافته با شواهد بیولوژیک نیز تطابق دارد.

کلید واژه‌ها

جداسازی شیء هدف از پس‌زمینه، شبکه‌های عصبی کانولوشنی، مدل‌های محاسباتی بینایی، بازخورد، ارتباطات جانبی

۱ مقدمه

با شناسایی سازوکار مغز و شبیه‌سازی آن می‌توان ماشین‌هایی تولید کرد که از لحاظ دقت و سرعت مشابه مغز انسان باشند. بازشناسی اشیاء به توانمندی برچسب زدن به اشیای خاص، از برچسب‌های جزئی (شناسایی^۱) تا برچسب‌های کلی (دسته‌بندی^۲) گفته می‌شود. انسان‌ها با دقت و سرعت زیادی می‌توانند اشیای پیرامونشان را، علی‌رغم پیچیدگی و درهم‌ریختگی تصویر، بازشناسی کنند [۱]. صحنه‌های بینایی که توسط سیستم بینایی پردازش می‌شوند، ترکیبی از تعداد زیادی عناصر تصویری (پیکسل) هستند که در روشنایی، رنگ، شکل و حرکت مقادیر

کامپیوترهای مدرن بسیاری از محاسبات پیچیده را خیلی سریع، حتی سریع‌تر و کارآمدتر از انسان انجام می‌دهند. اما در برخی از مسائل مانند بازشناسی اشیاء هنوز با انسان فاصله زیادی دارند.

این مقاله در دی‌ماه سال ۱۳۹۴ دریافت، در مردادماه ۱۳۹۵ بازنگری و در آبان‌ماه همان سال پذیرفته شد.

^۱ کارشناس ارشد برق، دانشکده مهندسی برق، دانشگاه تربیت دبیر شهید رجایی، رایانامه: f.abbasi@srttu.edu

^۲ دانشیار گروه کامپیوتر، دانشکده مهندسی کامپیوتر، دانشگاه تربیت دبیر شهید رجایی، رایانامه: ebrahimpour@ipm.ir

^۳ دانشجوی دکتری علوم اعصاب شناختی، پژوهشکده علوم شناختی، پژوهشگاه دانش‌های بنیادی، رایانامه: rajaei.k@ipm.ir

¹ Recognition

² Categorization

چندین بار به اشتراک گذاشته می‌شود که این کار در واقع مشابه با عمیق‌تر کردن شبکه است.

در سال ۲۰۱۵ لیانگ و هو^۵ یک شبکه‌ی کانولوشنی ارائه کردند که در آن از لایه‌های کانولوشنی بازگشتی (درون لایه‌ای) استفاده کردند [۱۷]. لایه‌های کانولوشنی بازگشتی سبب ایجاد بازخوردهای درون لایه‌ای شده و عملکرد شبکه در بهترین حالت برابر با شبکه‌ی عمیق‌تر (شبکه‌ای که از لحاظ تعداد لایه برابر با شبکه‌ی بازگشتی باز شده^۶ باشد) خواهد بود؛ زیرا در این مدل از اطلاعات سطح بالا در لایه‌های پایین‌تر استفاده نشده است.

در [۱۸] از بازخورد از لایه‌های بالاتر مدل استفاده شده است. لایه‌های بازخورد در این شبکه با استفاده از اطلاعات حاصل از نواحی بالاتر تنها به نورون‌هایی که مرتبط به بخشی از شیء هدف هستند اجازه فعالیت می‌دهند. ساختار این شبکه طوری طراحی شده است که از نقشه‌ی برجستگی مبتنی بر شیء^۷ برای مکان‌یابی استفاده می‌کند.

در این مقاله مدل بازگشتی بر پایه‌ی شبکه‌های عصبی کانولوشنی ارائه می‌شود که از یافته‌های بیولوژیکی الهام گرفته است. شبکه‌های کانولوشنی پیشین تصویر ورودی را در مسیر پیش‌خور و با پردازش‌های ایستا، پردازش می‌کردند اما در مدل بازگشتی پیشنهاد شده، دو سازوکار بازخورد از نواحی بالاتر بینایی و سرکوب پیرامون امکان پردازش پویا در ورودی را فراهم می‌کنند. این پردازش‌های پویا در پردازش دقیق‌تر تصاویر پیچیده، مانند حالتی که شیء هدف روی پس‌زمینه قرار گرفته است، نقش مؤثری دارند.

در بخش دوم این مقاله ساختار شبکه‌های عصبی کانولوشنی و تشریح یکی از مطرح‌ترین این مدل‌ها به نام AlexNet که اساس و پایه‌ی مدل پیشنهادی است ارائه می‌شود. در بخش سوم به مدل پیشنهادی پرداخته و جزئیات افزودن بازخورد و سرکوب پیرامون توضیح داده خواهد شد. سپس در بخش چهارم مجموعه داده‌ی استفاده‌شده و آزمایش‌های طراحی شده برای ارزیابی مدل و همچنین نتایج حاصل از این آزمایش‌ها ارائه می‌شود. در بخش پنجم به بحث و نتیجه‌گیری پرداخته می‌شود.

۲ شبکه عصبی کانولوشنی

شبکه‌های عصبی کانولوشنی بسیار شبیه به شبکه‌های عصبی اولیه هستند. لایه‌های اصلی مورد استفاده در ساختار شبکه‌های عصبی کانولوشنی عبارت‌اند از: لایه کانولوشنی، لایه ادغام^۸ و لایه کاملاً متصل (دقیقاً مشابه همان چیزی که در شبکه‌های عصبی رایج وجود دارد). با پشت‌هم قرار دادن این لایه‌ها ساختار یک شبکه‌ی

متنوعی دارند. سیستم بینایی این ویژگی‌ها را به صورت موازی ثبت کرده و آن‌ها را با الگوی فعالیت توزیع‌شده در نورون‌های زیادی در نواحی قشر بینایی بازنمایی می‌کند. اشیاء از به هم پیوستن ویژگی‌های متعلق به یک شیء ایجاد می‌شوند و از پس‌زمینه جدا می‌شوند [۲]. بنابراین، بازشناسی شیء نیازمند جداسازی شیء هدف از میان اشیای دیگر است.

شواهد زیستی نشان می‌دهند که مسیر بینایی انسان یک مسیر صرفاً پیش‌خور^۱ نیست [۳]. سیستم بینایی سازوکار سرکوب^۲ پیرامون را که توسط اتصالات جانبی حاصل می‌شود در مسیر پیش‌خور برای جداسازی شیء هدف از پس‌زمینه به کار می‌برد. همچنین، اخیراً نقش سیگنال‌های بازخورد^۳ نیز در مسئله‌ی جداسازی شیء هدف از پس‌زمینه مشخص شده است. پیشنهاد شده است که بعد از جاروب پیش‌خور آغازین، قشر بینایی اولیه مانند یک تخته‌سیاه عمل می‌کند [۴] و برای مقایسه و یکپارچه‌سازی محاسبات انجام‌گرفته در نواحی بالاتر در دسترس است. متناسب با این دیدگاه محققان ادعا کرده‌اند که بازخورد از نواحی بالاتر برای پردازش اطلاعات جزئی ضروری هستند. در سال‌های اخیر اهمیت این تعاملات در سلسله‌مراتب قشری توسط مدل‌هایی از سیستم بینایی نیز توضیح داده شده است [۵] تا [۷].

با وجود تحقیقات و یافته‌های بسیار در مورد سیستم بینایی انسان و همچنین سازوکار جداسازی شیء هدف از پس‌زمینه، این سازوکار هنوز تا حد زیادی ناشناخته باقی‌مانده است. مدل‌های محاسباتی بسیاری با الهام از سیستم بینایی انسان ارائه شده‌اند [۸] تا [۱۰]؛ با کشف ویژگی‌های بیشتری از سیستم بینایی انسان، این مدل‌ها همواره در حال توسعه و بررسی‌های بیشتری برای فهم کمبودهای مدل‌ها هستند [۱۱] تا [۱۳]. یکی از مدل‌های قدرتمندی که در سال‌های اخیر در حوزه یادگیری ماشین مطرح شده است، شبکه‌های عصبی کانولوشنی است [۱۴]. این مدل‌ها در ابتدا توسط لیکان^۴ برای بازشناسی حروف دست‌نویس استفاده شدند [۱۵]. این مدل‌ها در سال‌های اخیر به بهترین عملکرد در حوزه بازشناسی اشیاء رسیده‌اند اما یکی از کمبودهای شبکه‌های عصبی کانولوشنی عدم وجود بازخورد در آن‌هاست. اگرچه اخیراً بازخوردهایی به این مدل‌ها اضافه‌شده‌اند [۱۶] تا [۱۸]، اما هنوز سازوکاری که با سیستم بینایی انسان نیز مطابقت داشته باشد ارائه نشده است.

در [۱۶] لایه کاملاً متصل حذف شده و از شبکه‌ی تماماً کانولوشنی که دارای خروجی دو بعدی است برای برچسب‌دهی پیکسل‌های تصویر استفاده شده است. در این شبکه از خروجی شبکه مجدداً به عنوان ورودی استفاده شده است و وزن‌های شبکه

⁵Liang and Hu

⁶Unfolded

⁷ Object based saliency map

⁸ Pooling layer

¹Feedforward

²Suppression

³Feedback Signals

⁴LeCun

اولین لایه کانولوشنی تصویر ورودی با اندازه $3 \times 224 \times 224$ را با کرنل با اندازه‌های $3 \times 11 \times 11$ با گام چهار پیکسل فیلتر می‌کند. دومین لایه کانولوشنی خروجی لایه اول کانولوشنی (پاسخ نرمال شده و ادغام شده) را به‌عنوان ورودی دریافت کرده و آن را با 256 کرنل با اندازه $5 \times 5 \times 48$ فیلتر می‌کند. سومین، چهارمین لایه‌های کانولوشنی بدون لایه ادغام و نرمال‌سازی به یکدیگر متصل هستند. سومین لایه 384 کرنل با اندازه $3 \times 3 \times 256$ دارد که به خروجی‌های دومین لایه کانولوشنی (ادغام و نرمالیزه شده) متصل است. چهارمین لایه کانولوشنی 384 کرنل با اندازه $3 \times 3 \times 192$ و پنجمین لایه کانولوشنی 256 کرنل با اندازه $3 \times 3 \times 192$ دارد. هر کدام از لایه‌های کاملاً متصل 4096 نورون دارند.

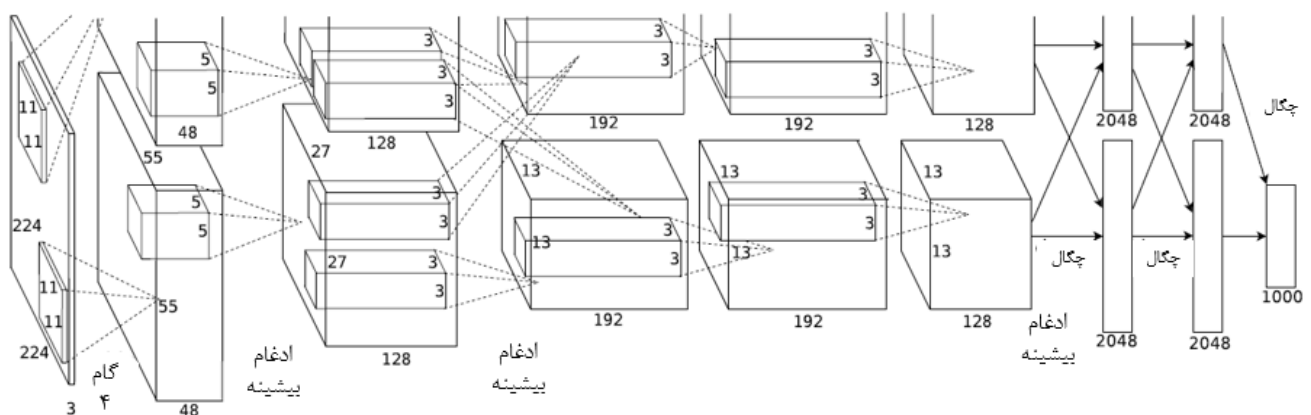
۳ مدل پیشنهادی

در این قسمت مدل بازگشتی ارائه می‌شود که مبتنی بر شبکه عصبی AlexNet بوده و الهام گرفته از سیستم بینایی انسان است. دو ویژگی اصلی مدل پیشنهادی، بازخورد از نواحی بالاتر بینایی و ویژگی سرکوب پیرامون در همان لایه هستند که در پردازش تصاویر پیچیده نقش بسزایی دارند. بازخورد از نواحی بالاتر بینایی با تغییر ورودی اولیه، در واقع باعث پردازش پویا روی تصویر ورودی می‌شود. با این پردازش پویا در هر مرحله، ورودی اصلی در معرض پردازش‌های بیشتری قرار گرفته و باعث تقویت ویژگی‌های مهم‌تر در ورودی می‌شود. علاوه بر این، افزودن سرکوب پیرامون نیز با رقابتی که بین نورون‌های همسایه ایجاد می‌کند، به این هدف کمک کرده و باعث تقویت پاسخ‌های قوی‌تر و حذف پاسخ‌های ضعیف‌تر می‌شود. در طی تکرار این پردازش‌ها، به تدریج بخش‌های اصلی تصویر ورودی که در واقع شیء هدف است، باقی مانده و بخش‌های غیر اصلی که مربوط به پس‌زمینه است حذف می‌شود. در این بخش به تشریح دو سازوکار پیشنهادی و نحوه مدل‌سازی آن‌ها می‌پردازیم.

عصبی کانولوشنی ایجاد می‌شود که تصویر اصلی را لایه به لایه از مقادیر اصلی پیکسل‌ها به امتیاز کلاس نهایی تبدیل می‌کنند [۱۹]. لایه‌های ادغام سبب کاهش پارامترها و در نتیجه آموزش ساده‌تر این‌گونه مدل‌ها نسبت به مدل‌های کاملاً متصل با تعداد لایه‌های پنهان برابر، می‌شوند. به علاوه، این لایه‌ها سبب افزایش مقاومت مدل نسبت به تغییرات مختصر در ورودی می‌شوند. این مدل‌ها در سال‌های اخیر به موفقیت‌های چشمگیری در حوزه‌های مختلف از جمله بازشناسی اشیاء رسیده‌اند. از جمله مدل‌های مطرح در این زمینه، شبکه عصبی کانولوشنی AlexNet است که در مسابقات ILSVRC-2012 به عملکرد $62/5$ درصد روی مجموعه داده‌ی ImageNet رسید و در این مقاله از آن به‌عنوان اساس مدل پیشنهادی استفاده شده است [۲۰]. ساختار مدل AlexNet در شکل ۱ نشان داده شده است. در بخش بعدی معماری این مدل تشریح می‌شود.

۱-۲ معماری AlexNet

همان‌طور که در شکل ۱ مشاهده می‌شود، مدل شامل هشت لایه‌ی وزن‌دار است؛ پنج لایه اول کانولوشنی و سه لایه باقی‌مانده لایه‌های کاملاً متصل هستند. خروجی آخرین لایه کاملاً متصل به یک بیشینه‌گیر نرم (softmax) هزارتایی داده می‌شود که توزیعی روی برچسب‌های 1000 کلاس ایجاد می‌کند. این مدل برای سرعت بخشیدن به آموزش، روی دو GPU آموزش داده شده است به طوری که نیمی از وزن‌ها در یک GPU و نیم دیگر روی یک GPU دیگر قرار داده شدند و تنها در برخی از لایه‌ها اتصال کامل لایه به لایه قبلی برقرار است. لایه‌های ادغام در این مدل برخلاف بیشتر مدل‌های پیشین، همراه با هم‌پوشانی هستند. اندازه پنجره‌های ادغام سه و گام حرکت دو پیکسل است. تابع غیرخطی غیرقابل اشباع $f(x) = \max(0, x)$ به خروجی همه لایه‌های کانولوشنی و لایه‌های کاملاً متصل اعمال شده است.



شکل ۱- نمایش ساختار مدل، با نمایش آشکار بین دو GPU. یک GPU قسمت‌هایی از لایه را که در بالای شکل است اجرا می‌کند در حالی که دیگری بخش‌های از لایه را که در قسمت پایینی شکل است را اجرا می‌کند. GPUها تنها در لایه‌های مشخصی ارتباط دارند [۱۸].

۳-۱ پیاده سازی بازخورد

همان طور که در مقدمه گفته شد سیستم بینایی انسان یک مسیر تماماً پیش خور نیست و لایه های بالاتر مانند IT بازخوردهایی به لایه های پایین تر مانند لایه های V_1 و V_2 ارسال می کنند.

سازوکار بازخورد استفاده شده در مدل پیشنهادی از بازخوردهای موجود در سیستم بینایی انسان الهام گرفته شده است. پس از ایجاد بردار خروجی احتمالاتی توسط مسیر پیش خور، برداری هم اندازه با بردار خروجی ایجاد شده و مقادیر متناظر با کلاس های قوی تر با یک و بقیه مقادیر با صفر جایگزین می شوند. با این کار، در واقع به کلاس های محتمل تر وزن یک و به کلاس های غیر محتمل وزن صفر نسبت داده می شود. با استفاده از قاعده پس انتشار خطا این بردار به لایه مورد نظر (در این مقاله بازخورد به لایه دوم کانولوشنی اعمال شده است) بازخورد شده، $net^{FB}[n]$ و ورودی اولیه (در اینجا خروجی لایه دوم کانولوشنی)، net^{FF} ، را مطابق فرمول ۱ تحت تأثیر قرار می دهد. به این ترتیب به کرنل هایی که در پیش بینی کلاس محتمل نقش داشتند وزن بیشتر و به کرنل های مربوط به کلاس غیر محتمل وزن کمتری نسبت داده می شود. ورودی تغییر یافته، $net^{FF}[n + 1]$ دوباره وارد مسیر پیش خور شده و خروجی جدید حاصل شده و این عمل چندین بار تکرار می شود. در طی این تکرارها در هر مرحله پاسخ های قوی که مربوط به شیء هدف هستند، قوی تر شده و پاسخ های ضعیف که احتمالاً مربوط به پس زمینه هستند، ضعیف تر می شوند و در نتیجه شیء هدف باقی مانده و پس زمینه تصویر حذف می شود.

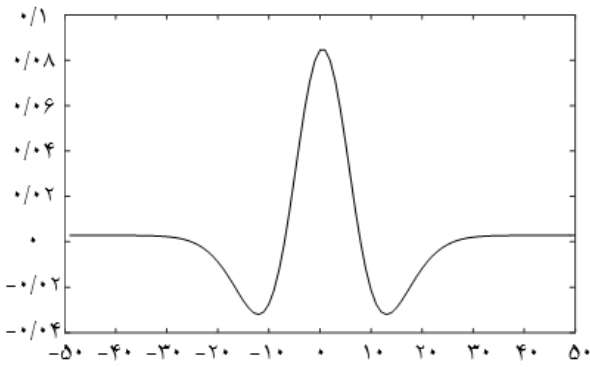
$$net^{FF}[n + 1] \leftarrow net^{FF}[n](1 + \eta \cdot net^{FB}[n]) \quad (1)$$

η ضریب بازخورد، net^{FB} سیگنال بازخورد از نواحی بالاتر، $net^{FF}[n]$ ورودی در مرحله قبل و $net^{FF}[n + 1]$ ورودی جدید است که توسط بازخورد مدوله شده است [۲۱].

۳-۲ مدل سازی سرکوب پیرامون

همان طور که در مقدمه گفته شد، یکی از ویژگی های سیستم بینایی انسان سرکوب پیرامون است. به همین منظور در مدل پیشنهادی از روشی که در مقاله ایتی و کخ^۱ (۲۰۰۰) ارائه شده برای اعمال این ویژگی استفاده شده است [۲۲]. در این روش رقابت بین ویژگی های یک لایه، با الهام از مطالعات انجام گرفته در ناحیه ی اولیه بینایی پیاده سازی شده است. این اتصالات سبب مدولاسیون پاسخ نورون ها توسط نورون های همسایه می شود اما چون فهم انسان از این تعاملات هنوز اندک است [۲۳] به یک مدل سازی ساده از این تعاملات اکتفا می شود. این مدل سه ویژگی مهم این تعاملات جانبی را ایجاد می کند: اول، تعاملات بین مکان مرکزی و پیرامون آن تحت سلطه ی یک جزء مهارتی از پیرامون به مرکز است

[۲۴] و این اثر به کنتراست نسبی بین مرکز و پیرامون بستگی دارد [۲۵]. دوم، مهار ناشی از مکان های پیرامون زمانی که نورون ها به ویژگی مشابه با نورون مرکزی حساس هستند، قوی تر است. سوم، مهار در فاصله های خاصی از مرکز، قوی تر ظاهر می شود و در فاصله های دورتر و یا نزدیک تر ضعیف تر است. این سه ویژگی به عنوان ساختار تعاملات غیر کلاسیک می تواند توسط گوسی های تفاضلی (DoG) دوبعدی مدل شود که در شکل ۲ یک نمونه از فیلتر DoG قابل مشاهده است.



شکل ۲- فیلتر DoG.

پیاده سازی خاص این تعاملات در این مدل به صورت زیر است: ورودی در هر تکرار با یک فیلتر DoG کانوالو شده، تصویر اصلی به حاصل اضافه می شود و نتایج منفی با صفر جایگزین می شوند.

معادله فیلتر DoG استفاده شده به صورت معادله ۲ است:

$$DoG(x, y) = \frac{C_{ex}^2}{2\pi\sigma_{ex}^2} e^{-(x^2+y^2)/2\sigma_{ex}^2} - \frac{C_{inh}^2}{2\pi\sigma_{inh}^2} e^{-(x^2+y^2)/2\sigma_{inh}^2} \quad (2)$$

ورودی M وارد معادله ۳ می شود:

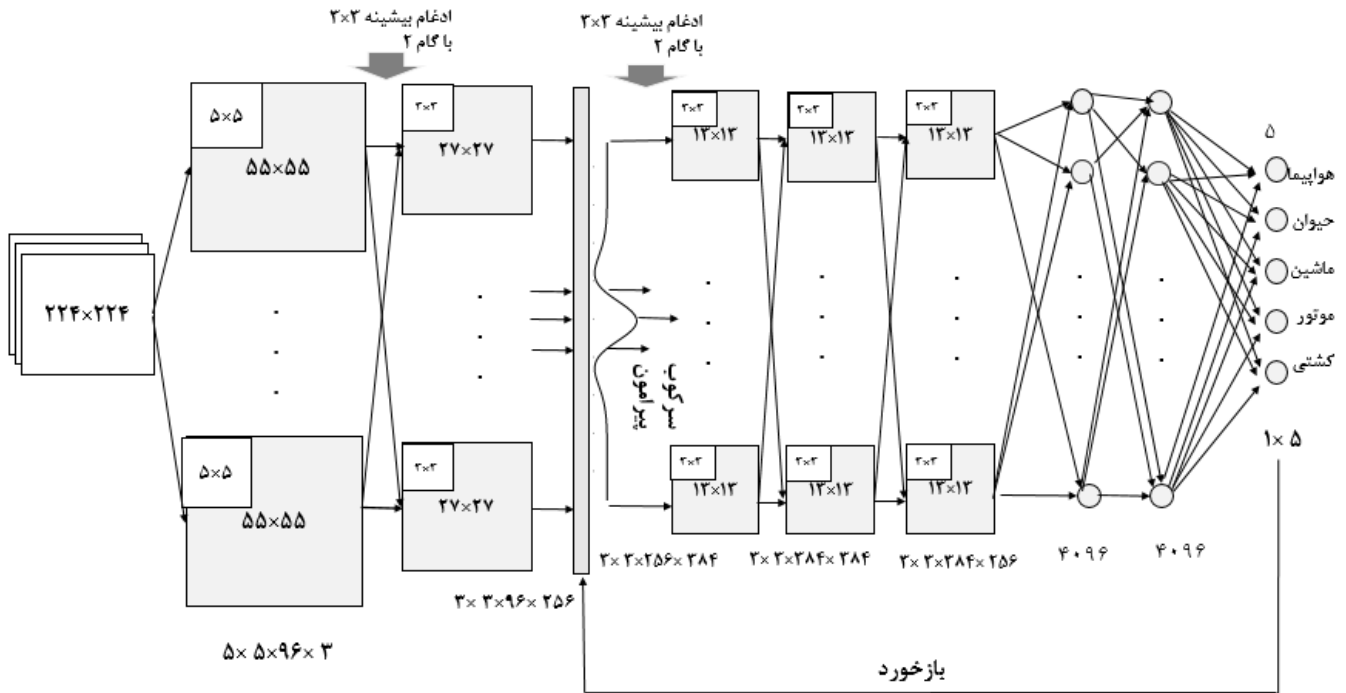
$$M \leftarrow |M + M * DoG - C_{inh}| \quad (3)$$

C_{inh} یک عبارت مهارتی ثابت است که مناطقی را که در آن ها سیگنال های مهارتی و تحریکی برابری را سرکوب می کند تا به عنوان بخش برجسته انتخاب نشوند.

۳-۳ مدل پیشنهادی: شبکه عصبی کانولوشنی بازگشتی

اخیراً آزمایش هایی تأثیر متقابل بازخورد و سرکوب پیرامون را بررسی کرده و نشان داده اند که این دو سازوکار با همکاری همدیگر منجر به بهبود بازشناسی اشیاء می شوند [۲۶]. شکل ۳ مدل بازگشتی حاصل از افزودن بازخورد و سرکوب پیرامون به مدل اصلی را نشان می دهد.

¹ Itti and Koch



شکل ۳- مدل بازگشتی. بازخورد از آخرین لایه کاملاً متصل به خروجی لایه دوم کانولوشنی اعمال می‌شود. پس از اعمال بازخورد، سرکوب پیرامون روی نتیجه حاصل شده اعمال می‌شود. سپس این پردازش بازگشتی در ده چرخه تکرار می‌شود.

بدون پس‌زمینه است، به مقایسه شبکه‌های عصبی کانولوشنی با انسان پرداخته و همچنین اثر عمق را در این مدل‌ها بررسی خواهیم کرد. سپس مدل پیشنهادی ارزیابی شده و نتایج تحلیل خواهند شد.

۴-۱ مجموعه داده

داده و نتایج آزمایش روان‌فیزیک این تحقیق از مرجع (قدرتی و همکاران) استفاده شده است [۲۷]. این مجموعه داده شامل پنج دسته شیء (ماشین، موتور، حیوان، کشتی و هواپیما) بود که به‌طور میانگین ۱۶ تصویر سه‌بعدی از هر کدام وجود داشت. این تصاویر روی پس‌زمینه‌هایی که به‌صورت تصادفی از بین ۴۰۰۰ تصویر پس‌زمینه‌ی موجود انتخاب می‌شدند قرار می‌گرفتند. پس‌زمینه‌ها شامل تصاویری از طبیعت و همچنین مناطق انسانی بودند. چهار تغییر مختلف به تصاویر اعمال شد: تغییر در مکان، تغییر در مقیاس، چرخش در عمق و چرخش در صفحه (شکل ۴). برای ایجاد میزان سختی‌های مختلف، هفت سطح مختلف تغییرات به تصاویر اعمال شد.

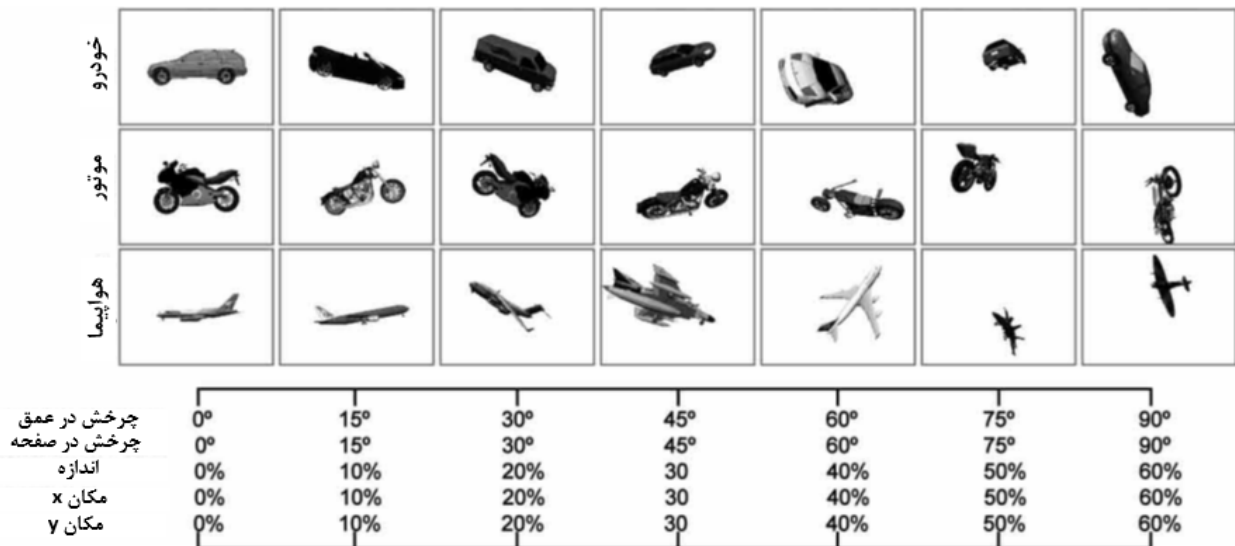
برای هر دسته در هر سطح ۳۰۰ تصویر به‌صورت تصادفی انتخاب شد. ۱۵۰ تصویر برای آموزش و ۱۵۰ تصویر دیگر برای مرحله‌ی آزمون به کار رفت. تصاویر آزمون شامل دو نوع تصاویر بدون پس‌زمینه و دارای پس‌زمینه بودند. تصاویر آموزشی برای تنظیم دقیق^۱ مدل AlexNet برای ایجاد مدل بازگشتی به کار رفتند.

همان‌طور که در شکل ۳ مشاهده می‌شود تصویر ورودی که دارای ابعاد $224 \times 224 \times 3$ است وارد مدل شده و خروجی که یک بردار 1×5 در فضای احتمالاتی است ایجاد می‌شود (مسئله مورد آزمایش پنج دسته‌ای است). سپس با توجه به خروجی حاصل، یک بردار با همین اندازه ایجاد شده به‌طوری‌که در این بردار به کلاس‌های قوی‌تر وزن یک و به کلاس‌های ضعیف‌تر وزن صفر داده می‌شود. این بردار با قاعده پس‌انتشارخطا تا لایه دوم کانولوشنی بازگردانده شده و سپس طبق فرمول شماره یک به خروجی لایه دوم اعمال می‌شود. سپس سازوکاری که برای سرکوب پیرامون پیش‌بینی شده است روی خروجی لایه دوم که اکنون با بازخورد از نواحی بالاتر مدوله شده است، اعمال می‌شود. سپس مسیر پیش‌خور ادامه پیدا کرده و این کار به همین صورت ده مرتبه تکرار می‌شود.

۴ آزمایش‌ها و نتایج

همان‌طور که در مقدمه گفته شد مدل‌های محاسباتی بینایی در حل مسائل پیچیده، مانند بازشناسی شیء هدفی که روی پس‌زمینه قرار گرفته، در مقایسه با انسان عملکرد ضعیف‌تری دارند. مطالعاتی که روی سیستم بینایی انسان انجام گرفته است نشان داده‌اند که بازخورد از نواحی بالاتر بینایی و سرکوب پیرامون نقش تعیین‌کننده‌ای در جداسازی شیء هدف از پس‌زمینه دارند. به همین دلیل در بخش سوم با الهام از یافته‌های بیولوژیکی، مدل بازگشتی بر اساس شبکه‌های عصبی کانولوشنی پیشنهاد شد. در این قسمت برای بررسی مسئله‌ی جداسازی شیء هدف از پس‌زمینه، با مجموعه داده‌ای که شامل دو مجموعه تصاویر دارای پس‌زمینه و

^۱Finetuen



شکل ۴- نمونه تصاویر در سطوح مختلف در پس زمینه ساده. شروع از حالت بدون تغییر در شکل سمت چپ تا تغییرات میانی و پیچیده در سمت راست.

آزموده شدند. این کار ده بار تکرار شد و از میانگین ده بار آزمایش به همراه نتایج آزمایش روان فیزیک، شکل ۶ حاصل شد.

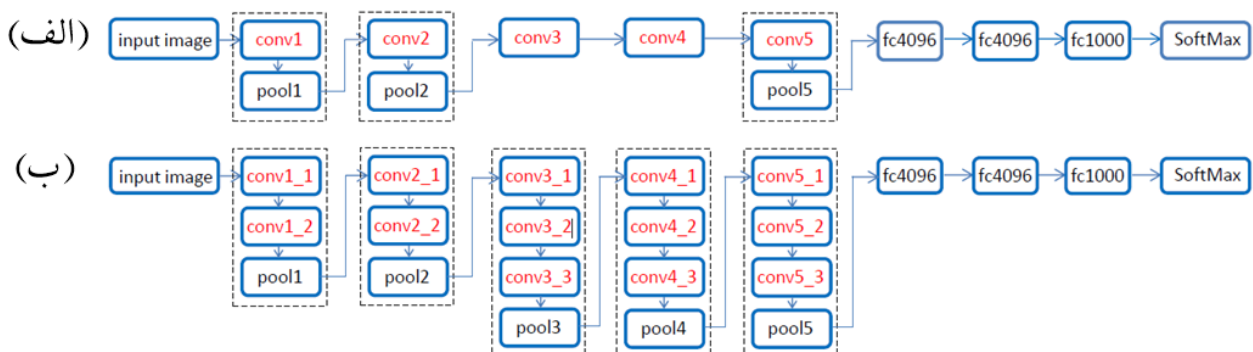
با توجه به شکل ۶-الف، در حالت تصاویر بدون پس زمینه هر دو مدل در سطوح پایین تغییرات عملکردی بهتر از عملکرد انسان دارند. اما با بیشتر شدن تغییرات مدل‌ها کاهش عملکرد بیشتری نسبت به انسان دارند. این مشاهده نشان‌دهنده عدم وجود استقلال نسبت به تغییرات است. در این میان مدل عمیق‌تر نسبت به AlexNet که تعداد لایه‌های کمتری دارد در سطوح بالاتر عملکرد بهتری دارد زیرا به دلیل برخورداری از پارامترهای بیشتر قادر به یادگیری بهتر پیچیدگی‌های مسئله است.

همانطور که در شکل ۶-ب مشخص است با افزودن پس زمینه به تصاویر آزمون، عملکرد هر سه (دو مدل و همچنین انسان) کاهش قابل ملاحظه‌ای داشته است. اما همانطور که مشاهده می‌شود کاهش عملکرد انسان نسبت به مدل‌ها کمتر است. می‌توان این‌گونه استنباط کرد که با بالا رفتن سطح تغییرات و همچنین افزودن پس زمینه به تصاویر، مسئله بسیار پیچیده شده و مدل پیش‌خور قادر به حل مسئله نیست. همچنان VGGNet در سطوح بالاتر نسبت به AlexNet عملکرد بهتری دارد که دلیل آن یادگیری بهتر فضای مسئله است.

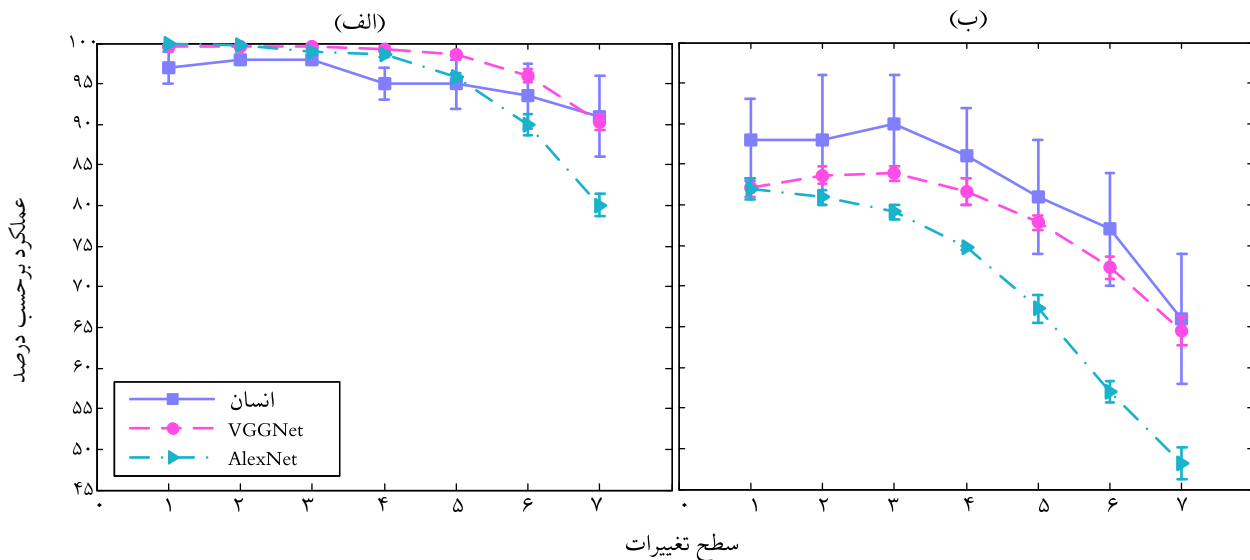
۲-۴ مقایسه عملکرد شبکه‌های عصبی کانولوشنی و عملکرد رفتاری انسان

علاوه بر AlexNet یک مدل عمیق‌تر با نام VGGNet نیز ارزیابی و بررسی شد [۲۸]. این مدل‌ها توسط چارچوب Caffe پیاده‌سازی و اجرا شدند [۲۹]. این مدل نیز در سال ۲۰۱۴ به بهترین عملکرد در مسابقات ILSVRC-2014 رسید. VGGNet دارای ۱۳ لایه کانولوشنی و سه لایه کاملاً متصل، در مجموع ۱۶ لایه است. ساختار این مدل‌ها در شکل ۵ نشان داده شده است. مدل عمیق‌تر دارای بار محاسباتی بسیار بالایی است. دلیل استفاده نکردن از این ساختار برای مدل پیشنهادی، زمان‌بر بودن اجرای مدل و هزینه‌ی محاسباتی بالای آن است.

به دلیل آموزش زمان‌بر این شبکه‌ها و همچنین کمبود داده برای آموزش، معمولاً از شبکه‌های آموزش دیده‌ی قبلی که وزن‌های آن‌ها در دسترس عموم قرار می‌گیرند استفاده می‌شود. برای این قسمت از شبکه‌های آموزش دیده‌ی AlexNet و VGGNet استفاده شد. تنها آخرین لایه کاملاً متصل با ۷۵۰ تصویر (۱۵۰ تصویر از هر دسته) و در هفت سطح مختلف تغییرات، آموزش داده شد. سپس این شبکه‌ها با ۷۵۰ تصویر دیگر در هفت سطح مختلف تغییرات



شکل ۵- مدل‌های مورد آزمایش. (الف) ساختار مدل AlexNet (ب) ساختار مدل VGGNet



شکل ۶- مقایسه‌ی عملکرد انسان با AlexNet و VGGNet در مسئله‌ی پنج دسته‌ای. الف) عملکرد انسان و مدل‌ها در تصاویر بدون پس‌زمینه. ب) عملکرد انسان و مدل‌ها در تصاویر دارای پس‌زمینه.

در شکل ۷-ج مشاهده می‌شود که VGGNet با تعداد لایه‌های بیشتری که نسبت به AlexNet دارد در سطوح بالای تغییرات در تصاویر بدون پس‌زمینه، بهتر عمل کرده است. اما در شکل ۷-د که مربوط به تصاویر دارای پس‌زمینه است مشاهده می‌شود که در سطوح پایین تغییرات، VGGNet بدتر از AlexNet عمل کرده است. این مشاهده را این‌طور می‌توان تفسیر کرد که مدلی که لایه‌های بیشتری دارد، ویژگی‌های بیشتری را یاد گرفته و این ویژگی‌های بیشتر، باعث پیدا شدن اشیای بیشتری در تصویر می‌شود (مدل به اشتباه بخش‌هایی از تصویر را به‌عنوان شیء در نظر می‌گیرد).

۴-۳ نتایج حاصل از مدل پیشنهادی

در این بخش به بررسی مدل پیشنهادی پرداخته خواهد شد. همان‌طور که گفته شد دو سازوکار سرکوب پیرامون و بازخورد به خروجی لایه دوم کانولوشنی افزوده شد. ابتدا به بررسی اثر تک تک این سازوکارها پرداخته می‌شود.

برای یافتن تعداد تکرارها برای افزودن بازخورد، عملکرد مدل در طی ده تکرار محاسبه شد. با توجه به شکل ۸ مشخص است که تعداد ده تکرار برای مدل بازگشتی کافی است زیرا تقریباً بعد از آن، عملکرد مدل بازگشتی بر اثر بازخورد از ناحیه بالاتر تغییر چندانی نمی‌کند. برای ارزیابی مدل پیشنهادی سه آزمایش مختلف طراحی گردید. در اولین آزمایش تنها سازوکار سرکوب پیرامون، در دومین آزمایش نیز تنها سازوکار بازخورد و در آخرین آزمایش هر دو سازوکار به‌صورت همزمان به خروجی لایه دوم کانولوشنی اعمال شدند. میانگین نتایج برای ۱۰ بار اجرا در شکل ۹ قابل مشاهده است.

برای بررسی اثر عمق در مدل‌های کانولوشنی بر روی جداسازی شیء هدف از پس‌زمینه، به مقایسه دو مدل با ماتریس عدم شباهت بازنمایی^۱ (RDM) پرداخته خواهد شد. ماتریس عدم شباهت یک روش مؤثر برای نشان دادن عدم شباهت بین الگوهای پاسخ حاصل از دو تصویر است. برای هر دو مدل خروجی‌های سه لایه آخر کانولوشنی و همچنین لایه کاملاً متصل دوم استخراج شدند و ماتریس عدم شباهت آن‌ها رسم شد که برای حالت‌های مختلف در شکل‌های ۱۶، ۱۷، ۱۸ و ۱۹ قابل مشاهده هستند (برای دیدن شکل‌ها به پیوست مراجعه شود). برای رسم ماتریس عدم شباهتاز جعبه‌ابزار تحلیل شباهت بازنمایی^۲ (RSA) در MATLAB استفاده شد [۳۰] (برای اطلاعات بیشتر به پیوست مراجعه شود).

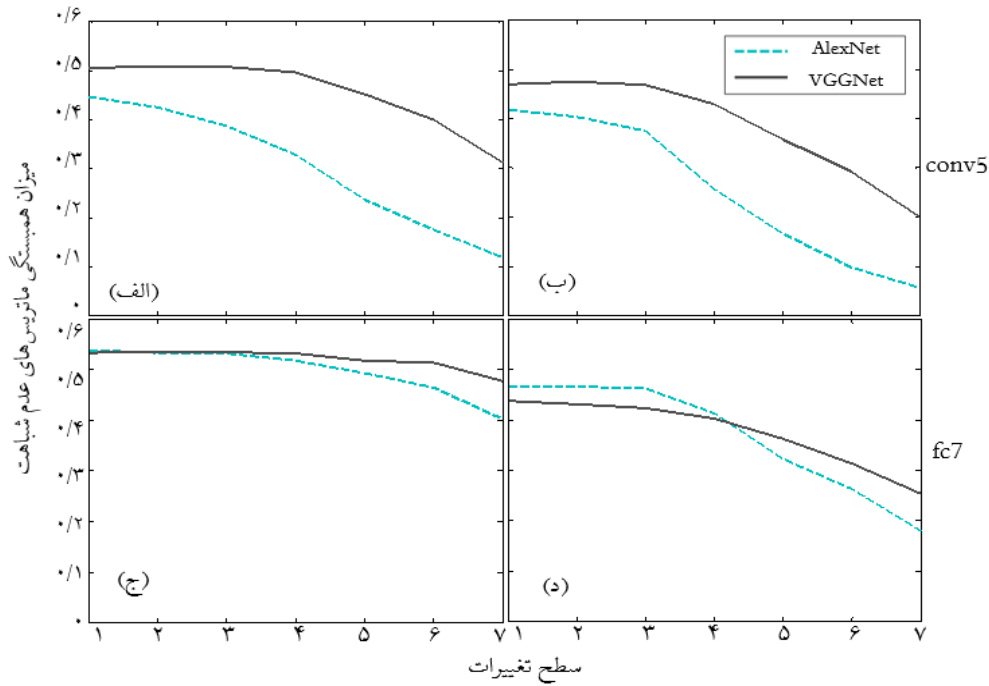
برای بررسی و مقایسه کمی بین ماتریس عدم شباهت دو مدل از معیار کندانال تائو^۳ استفاده می‌کنیم که همبستگی بین ماتریس‌های عدم شباهت حاصل از دو مدل را با ماتریس عدم شباهت ایده‌آل محاسبه می‌کند (ماتریس عدم شباهت ایده‌آل در پیوست قابل مشاهده است).

با توجه به شکل ۷-الف و ب مشاهده می‌شود که VGGNet در هر دو مجموعه تصاویر بدون پس‌زمینه و دارای پس‌زمینه، به اندازه ثابتی از AlexNet بهتر است. در تصاویر بدون پس‌زمینه، این اختلاف را می‌توان به بازشناسی بهتر اشیاء در VGGNet تفسیر کرد. اما در تصاویر دارای پس‌زمینه نیز این میزان اختلاف تغییری نکرده است. احتمالاً این ثابت بودن اختلاف در دو حالت تصاویر دارای پس‌زمینه و بدون پس‌زمینه به این دلیل است که VGGNet با وجود لایه‌های بیشتر، در حذف پس‌زمینه بهتر عمل نکرده است و این افزایش عملکرد تنها حاصل از بازشناسی بهتر اشیاء است.

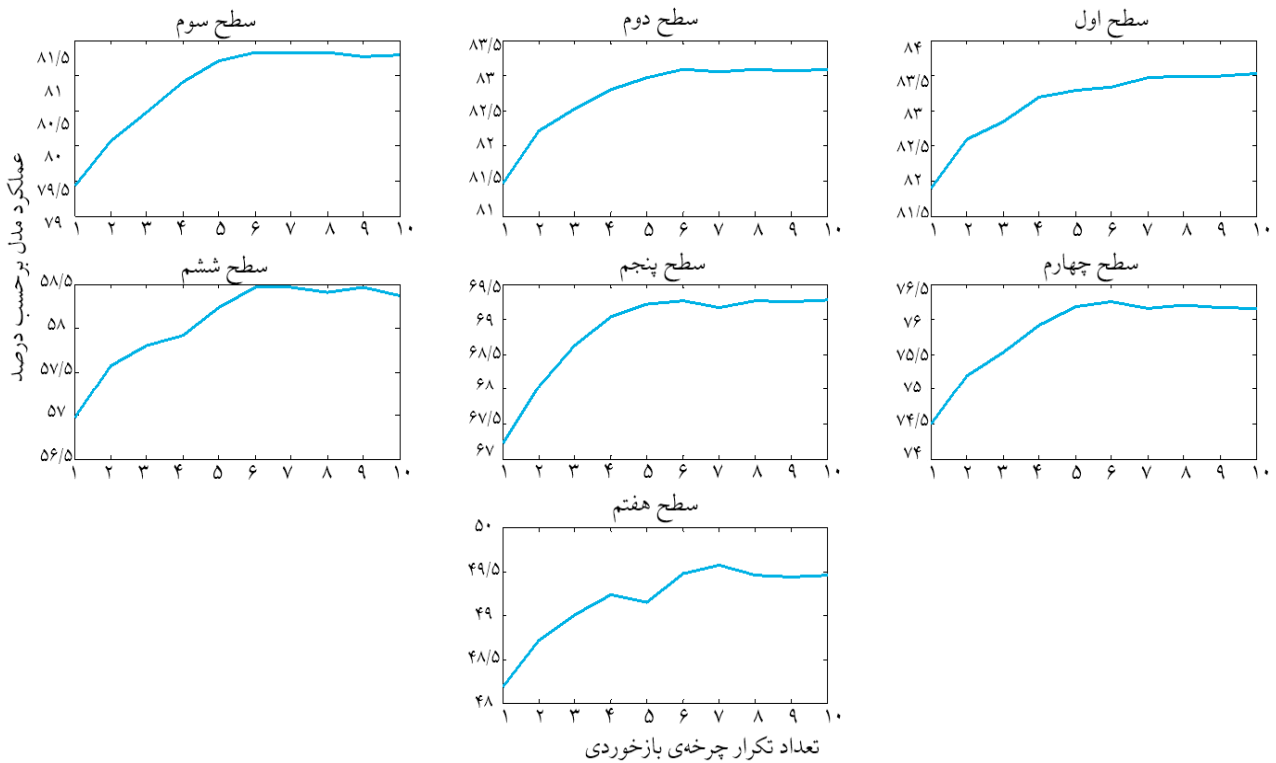
¹Representational Dissimilarity Matrix

²Representational Similarity Analysis

³Kendall tau



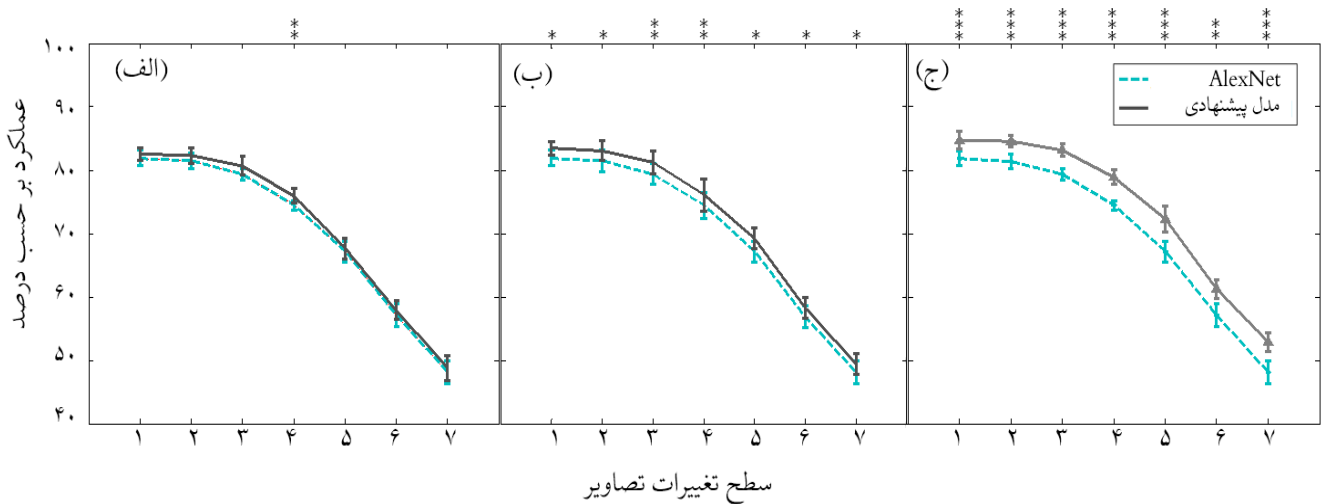
شکل ۷- میزان همبستگی ماتریس‌های عدم‌شباهت با ماتریس عدم‌شباهت ایده‌آل. (الف و ب) میزان همبستگی ماتریس عدم‌شباهت حاصل از آخرین لایه کانولوشنی با ماتریس عدم‌شباهت ایده‌آل به ترتیب برای تصاویر بدون پس‌زمینه و دارای پس‌زمینه. (ج و د) میزان همبستگی ماتریس عدم‌شباهت حاصل از دومین لایه کاملاً متصل با ماتریس عدم‌شباهت ایده‌آل برای تصاویر بدون پس‌زمینه و دارای پس‌زمینه.



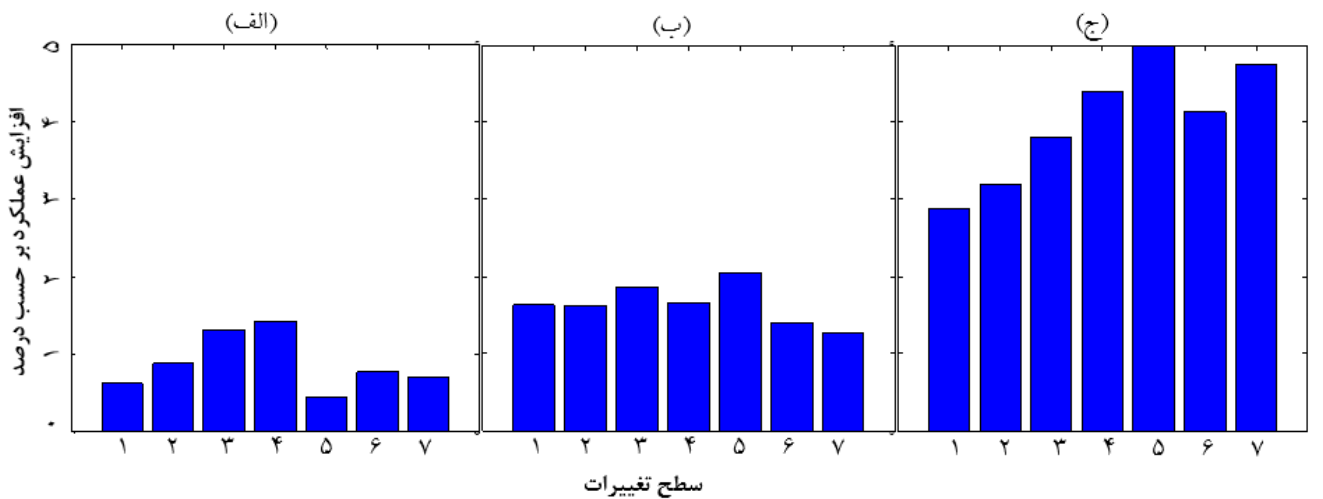
شکل ۸- میزان تغییر عملکرد مدل بر اثر افزودن بازخورد در طی ده چرخه‌ی بازخوردی

مشاهده می‌شود که افزودن بازخورد به‌تنهایی عملکرد مدل را به‌طور معنی‌داری افزایش می‌دهد درحالی‌که افزودن سازوکار سرکوب پیرامون به‌تنهایی تأثیر چشمگیری در عملکرد مدل ندارد. اما نتیجه‌ی جالب افزایش چشمگیر عملکرد مدل در حالت ترکیب دو سازوکار است. با توجه به شکل ۹ نتیجه گرفته می‌شود زمانی

وجود ستاره بالای هر سطح از تغییرات در شکل ۹ نشان‌دهنده‌ی معنی‌دار بودن اختلاف بین AlexNet و مدل پیشنهادی است، به‌طوری‌که افزایش تعداد ستاره‌ها، معنی‌دارتر بودن اختلاف را نشان می‌دهد. معنی‌داری توسط ارزش احتمال یا همان P-value محاسبه شده است (برای اطلاعات بیشتر به پیوست مراجعه شود).



شکل ۹- مقایسه عملکرد مدل پیشنهادی و AlexNet. الف) عملکرد مدل پیشنهادی با افزودن سرکوب پیرامون ب) عملکرد مدل پیشنهادی با افزودن بازخورد از آخرین لایه کاملاً متصل ج) عملکرد مدل پیشنهادی با افزودن بازخورد از آخرین لایه کاملاً متصل و سرکوب پیرامون به طور همزمان. وجود ستاره نشان دهنده معنی دار بودن اختلاف بین دو عملکرد است و بیشتر شدن این ستاره‌ها، معنی دارتر بودن اختلاف را نشان می‌دهد. آزمون معنی داری توسط روش آزمون p-value سنجیده شده است. * یعنی $p < 0/05$, ** یعنی $p < 0/01$ و *** یعنی $p < 0/001$



شکل ۱۰- مقدار افزایش عملکرد مدل پیشنهادی بر حسب درصد. الف) افزایش ناشی از افزودن سرکوب پیرامون ب) افزایش ناشی از افزودن بازخورد ج) افزایش ناشی از افزودن بازخورد و سرکوب پیرامون به طور همزمان.

۴-۴ مقایسه‌ی عملکرد مدل پیشنهادی با مدل‌های دیگر و انسان

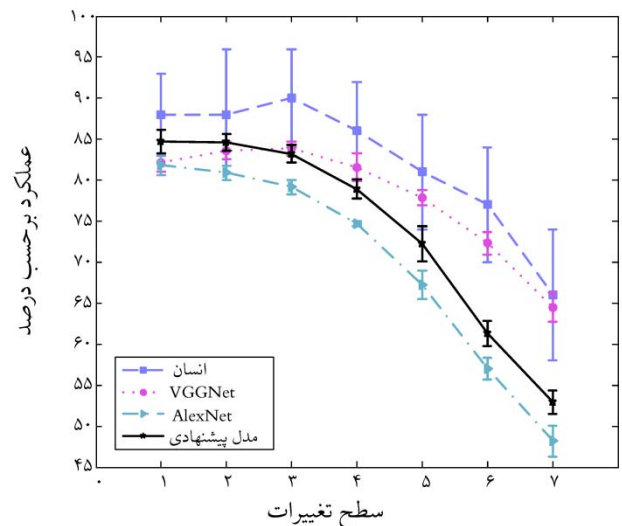
همانطور که مشاهده شد افزودن سرکوب پیرامون و بازخورد سبب بالاتر رفتن عملکرد مدل پیشنهادی شد. در این بخش به مقایسه‌ی نتایج حاصل از مدل پیشنهادی با انسان و مدل‌های AlexNet و VGGNet خواهیم پرداخت (شکل ۱۱). همانطور که در شکل ۱۱ مشخص است با وجود بهتر شدن عملکرد مدل پیشنهادی، هنوز این مدل به عملکرد انسان نرسیده است. مدل پیشنهادی در همه‌ی سطوح از مدل AlexNet عملکرد بهتری دارد.

که هر دو سازوکار به صورت همزمان به مدل افزوده می‌شوند بهترین عملکرد حاصل خواهد شد. از منظر زیستی هم شواهدی بر این موضوع وجود دارد که نشان می‌دهند برای بهبود عملکرد بینایی انسان، به همراهی هر دو سازوکار نیازمند است [۲۶]. میزان افزایش عملکرد بر حسب درصد، برای هر سه آزمایش مطرح شده در شکل ۱۰ آمده است. با توجه به شکل ۱۰-ج، یکی از نتایج قابل توجه دیگر سیر صعودی افزایش عملکرد در طی سطوح مختلف تغییرات تصاویر است. این نتیجه بیانگر این موضوع است که با بیشتر شدن پیچیدگی تصاویر و در نتیجه سخت‌تر شدن فضای مسئله، نیاز به بازخورد از نواحی بالاتر بیشتر حس شده و بازخورد باعث بهبود بیشتری در عملکرد می‌شود.

برای مشاهده بصری تأثیر افزودن بازخورد و سرکوب پیرامون، خروجی لایه پنجم کانولوشنی را برای AlexNet و همچنین مدل پیشنهادی، بصری سازی شد که در شکل ۱۳ آورده شده است. همان طور که مشاهده می شود، پس زمینه تصوی در مدل پیشنهادی بهتر از مدل بدون بازخورد حذف شده است. همچنین قابل مشاهده است که حذف پس زمینه زمانی بهتر انجام می شود که هر دو سازوکار بازخورد از نواحی بالاتر و سرکوب پیرامون توأمان به مدل اضافه شده اند.

۵ نتیجه گیری

در این مقاله قابلیت مدل های AlexNet و VGGNet در جداسازی شیء هدف از پس زمینه بررسی و با بصری سازی نشان داده شد که با پیش روی در طی لایه های مدل، بازشناسی شیء بهبود می یابد. زیرا بیشتر بودن فیلترها باعث یادگیری ویژگی های پیچیده تری در مدل شده و در نتیجه ویژگی های پیچیده تر و مفیدتری از شیء استخراج می شود؛ همچنین نشان داده شد که افزایش عمق نسبت به تغییرات استقلال بیشتری ایجاد می کند. اما مدل های خیلی عمیق با وجود عملکرد بهتر مقرون به صرفه نیستند در حالی که هنوز هم با عملکرد انسان فاصله دارند. در نتیجه سازوکارهای بازخورد و سرکوب پیرامون به مدل اضافه شدند. بازخورد و سرکوب پیرامون که در مغز به سیستم بینایی اولیه اعمال می شود در اینجا نیز به یکی از پایین ترین لایه های مدل یعنی لایه دوم کانولوشنی اعمال شد. بازخورد از خروجی لایه آخر که امتیازهای کلاس ها است حاصل شد و از طریق قاعده پس انتشار خطا به لایه های پایین تر آورده شد. مشاهده شد که همراهی این سازوکارها باعث افزایش عملکرد مدل در جداسازی شیء هدف از پس زمینه در حدود ۴ درصد خواهد شد. ولی در حالتی که این سازوکارها به تنهایی به مدل افزوده شدند افزایش عملکرد کمتر بود؛ البته بازخورد به تنهایی هم قادر به افزایش عملکرد است اما نکته مهم این است که زمانی که بازخورد با سرکوب پیرامون همراه می شود افزایش عملکرد بسیار چشمگیرتر و معنی دارتر است.



شکل ۱۱- مقایسه عملکرد مدل پیشنهادی و مدل های AlexNet و VGGNet و انسان.

اما مشاهده می شود که مدل پیشنهادی تنها در سطح اول از VGGNet بهتر عمل کرده است که دلیل این موضوع همانطور که در بخش ۲-۴ عنوان شد یادگیری بهتر فضای مسئله به دلیل پارامترهای بیشتر است که باعث می شود مدل VGGNet در سطوح بالای تغییرات عملکرد بهتری داشته باشد. افزودن بازخورد و سرکوب پیرامون سبب افزایش پیچیدگی محاسباتی در حدود ۲۰ برابر می شود که این یکی از نقص های مدل پیشنهادی است.

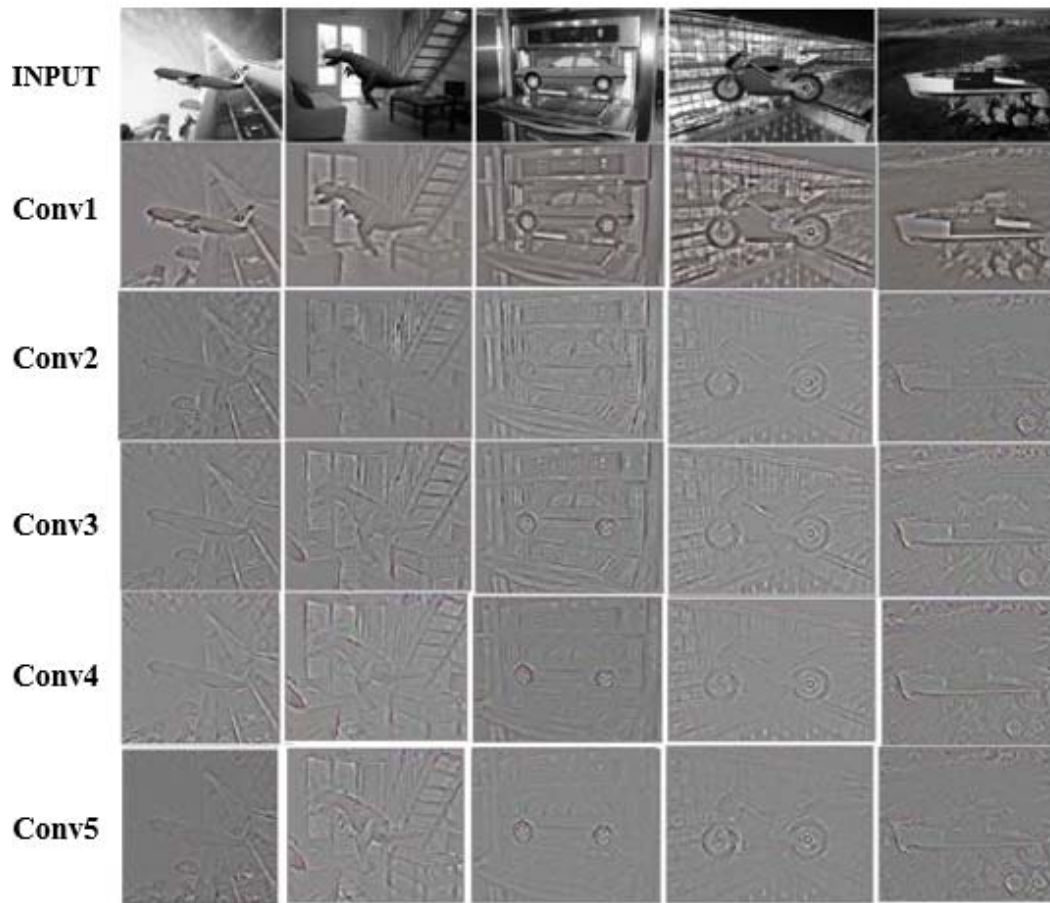
۴-۵ بصری سازی بازنمایی های لایه های کانولوشنی

یکی از راه های شناخت بیشتر شبکه های عصبی کانولوشنی بصری کردن لایه های درونی آن ها و فهم هر چه بیشتر آن ها است. از روش معکوس کانولوشن^۱ برای بصری سازی بازنمایی های لایه های درونی کانولوشنی استفاده می شود [۳۱]. در این روش خروجی هر لایه کانولوشنی در ترانهاد^۲ فیلترها ضرب شده و ورودی لایه حاصل می شود.

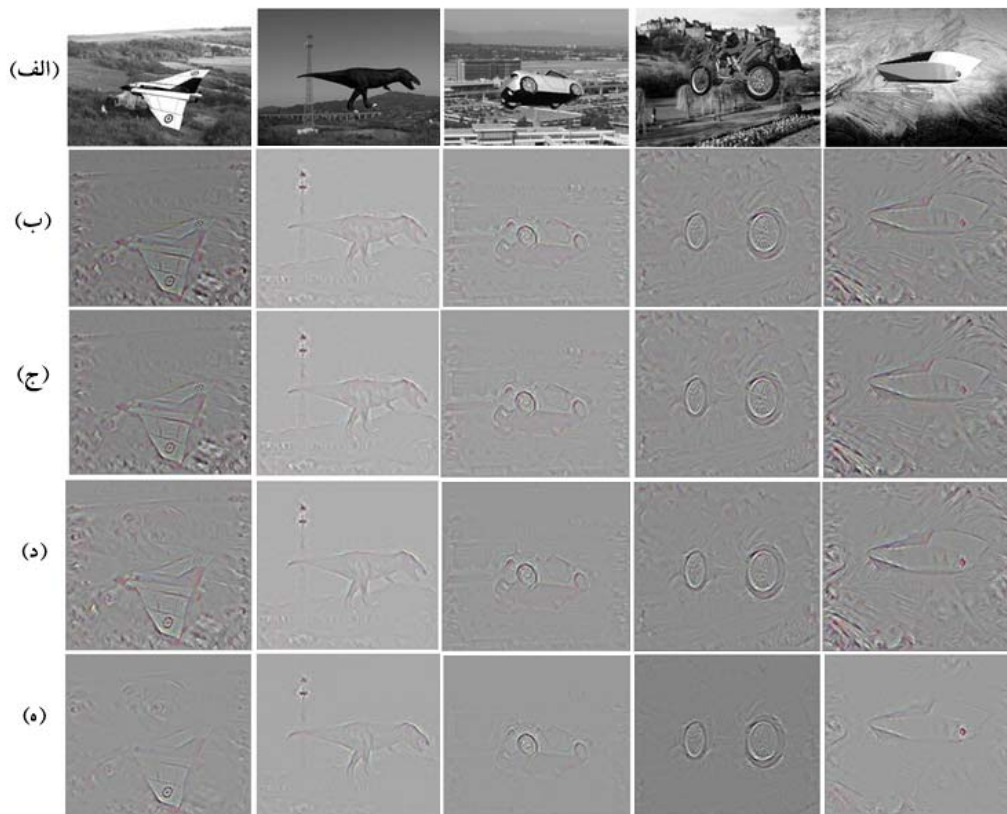
برای معکوس لایه های ادغام نیز، مکان هایی که فعالیت در آن ها بیشینه بوده در حافظه ذخیره می شوند و در راه برگشت از آن ها استفاده می شود. نتایج حاصل از بصری سازی در شکل ۱۲ نشان داده شده است. با توجه به شکل مشاهده می شود که با پیش روی در لایه های مدل، پس زمینه بیشتری از تصویر حذف می شود. همان طور که در [۳۱] نشان داده شده است، با بالاتر رفتن در سلسله مراتب شبکه های کانولوشنی، ویژگی های پیچیده تر و مهم تری توسط مدل یاد گرفته می شوند و در نتیجه ویژگی های غیر مهم که عموماً جزء پس زمینه تصوی هستند در طی سلسله مراتب حذف می شوند.

¹Deconvolution

²Transpose



شکل ۱۲ - بصری سازی لایه های درونی CNN. مشاهده می شود که با پیش رو در لایه ها پس زمینه ی بیشتری از تصویر حذف می شود.

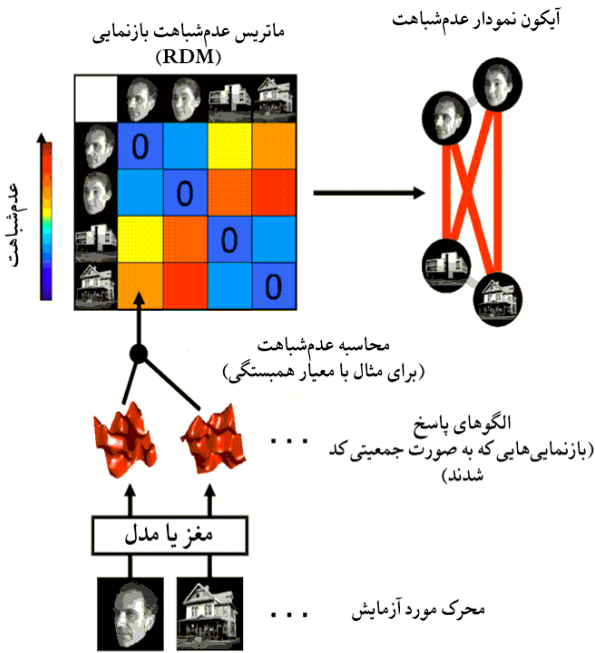


شکل ۱۳ - بصری سازی لایه های کانولوشنی بعد از افزودن بازخورد و سرکوب پیرامون. الف) تصویر اصلی. بصری سازی خروجی لایه ی پنجم کانولوشنی برای ب) مدل AlexNet ج) مدل پیشنهادی با افزودن بازخورد د) مدل پیشنهادی با افزودن سرکوب پیرامون ه) مدل پیشنهادی با افزودن بازخورد و سرکوب پیرامون.

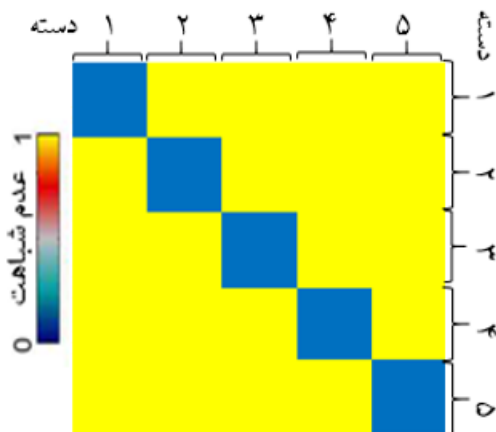
۶ مراجع

- [16] Pinheiro, P. H., & Collobert, R. (2013). Recurrent convolutional neural networks for scene parsing. arXiv preprint arXiv:1306.2795.
- [17] Liang, M., & Hu, X. (2015). Recurrent convolutional neural network for object recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3367–3375).
- [18] Cao, C., Liu, X., Yang, Y., Yu, Y., Wang, J., Wang, Z., ... & Ramanan, D. (2015). Look and think twice: Capturing top-down visual attention with feedback convolutional neural networks. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2956–2964).
- [19] Bengio, Y., Goodfellow, I. J., & Courville, A. (2015). Deep Learning.
- [20] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).
- [21] Brosch, Tobias, and Heiko Neumann. "Computing with a canonical neural circuits model with pool normalization and modulating feedback." *Neural computation* (2014).
- [22] Itti, Laurent, and Christof Koch. "A saliency-based search mechanism for overt and covert shifts of visual attention." *Vision research* 40.10 (2000): 1489–1506.
- [23] Zenger, Barbara, and Dov Sagi. "Isolating excitatory and inhibitory nonlinear spatial interactions involved in contrast detection." *Vision Research* 36.16 (1996): 2497–2513.
- [24] Cannon, Mark W., and Steven C. Fullenkamp. "Spatial interactions in apparent contrast: inhibitory effects among grating patterns of different spatial frequencies, spatial positions and orientations." *Vision research* 31.11 (1991): 1985–1998.
- [25] Levitt, Jonathan B., and Jennifer S. Lund. "Contrast dependence of contextual effects in primate visual cortex." *Nature* 387.6628 (1997): 73–76.
- [26] Nassi, Jonathan J., Stephen G. Lomber, and Richard T. Born. "Corticocortical feedback contributes to surround suppression in V1 of the alert primate." *The Journal of Neuroscience* 33.19 (2013): 8504–8517.
- [27] Ghodrati, M., Farzmaadi, A., Rajaei, K., Ebrahimpour, R., & Khaligh-Razavi, S. M. (2014). Feedforward object-vision models only tolerate small image variations compared to human. *Frontiers in computational neuroscience*, 8.
- [28] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).
- [29] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., ... & Darrell, T. (2014, November). Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the ACM International Conference on Multimedia* (pp. 675–678). ACM.
- [1] DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, 73(3), 415–434.
- [2] Roelfsema, P. R., Tolboom, M., & Khayat, P. S. (2007). Different processing phases for features, figures, and selective attention in the primary visual cortex. *Neuron*, 56(5), 785–792.
- [3] Arall, M., Romeo, A., & Supèr, H. (2012). Role of feedforward and feedback projections in figure-ground responses. S. Molotchnikoff (Ed.). INTECH Open Access Publisher.
- [4] Bullier, J. (2001). Feedback connections and conscious vision. *Trends in cognitive sciences*, 5(9), 369–370.
- [5] Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience*, 2(1), 79–87.
- [6] Thielscher, A., & Neumann, H. (2003). Neural mechanisms of cortico-cortical interaction in texture boundary detection: a modeling approach. *Neuroscience*, 122(4), 921–939.
- [7] O'Reilly, R. C., Wyatte, D., Herd, S., Mingus, B., & Jilk, D. J. (2013). Recurrent processing during object recognition. *Frontiers in psychology*, 4, 1–14.
- [8] Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4), 193–202.
- [9] Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148(3), 574–591.
- [10] Rolls, Edmund T. "Invariant visual object and face recognition: neural and computational bases, and a model, VisNet." *Frontiers in Computational Neuroscience* 6 (2012).
- [11] Ghodrati, Masoud, Karim Rajaei, and Reza Ebrahimpour. "The importance of visual features in generic vs. specialized object recognition: a computational study." *Frontiers in computational neuroscience* 8 (2014).
- [12] Rajaei, Karim, et al. "A stable biologically motivated learning mechanism for visual feature extraction to handle facial categorization." (2012): e38478.
- [13] Ghodrati, Masoud, et al. "How can selection of biologically inspired features improve the performance of a robust object recognition model." *PloS one* 7.2 (2012): e32357.
- [14] Kriegeskorte, Nikolaus. "Deep neural networks: a new framework for modelling biological vision and brain information processing." *bioRxiv* (2015): 029876.
- [15] LeCun, Yann, et al. "Gradient-based learning applied to document recognition." *Proceedings of the IEEE* 86.11 (1998): 2278–2324.

محرك با خودش بیشترین مقدار و شباهت هر محرك با محرك دیگر کمترین مقدار ممکن است.



شکل ۱۴- محاسبه‌ی ماتریس عدم شباهت بازنمایی. برای هر جفت محرك مورد آزمایش، الگوهای پاسخ مستخرج از یک ناحیه مغز یا بازنمایی مدل، برای مشخص کردن عدم شباهت بازنمایی محرك، مقایسه شدند. عدم شباهت بین دو الگو می‌تواند از تفاضل یک و میزان همبستگی محاسبه شود. (شباهت کامل دارای مقدار صفر و عدم شباهت دارای مقدار یک و تضاد کامل دارای مقدار ۲ خواهد شد). این عدم شباهت‌ها برای تمام جفت محرك‌ها در ماتریس عدم شباهت بازنمایی (RDM) گردآوری شده‌اند. هر سلول ماتریس الگوهای پاسخ مستخرج شده برای دو تصویر را مقایسه می‌کند و ماتریس متقارن بوده و قطر آن صفر است. برای بصری‌سازی بازنمایی، می‌توان محرك را برحسب عدم شباهت الگوی پاسخ مرتب کرد به طوری که محرك‌هایی که الگوهای پاسخ یکسانی دارند نزدیک یکدیگر و محرك‌هایی که الگوهای پاسخ غیرمشابهی دارند دور از هم قرار بگیرند.



شکل ۱۵- ماتریس عدم شباهت ایده‌آل. مسئله‌ی مورد آزمایش پنج دسته‌ای بوده و ماتریس عدم شباهت بیانگر این موضوع است که شباهت بین بازنمایی‌های هر دسته دارای شباهت کامل بوده و رنگ آن‌ها آبی است در حالی که بین دسته‌های دیگر عدم شباهت کامل برقرار بوده و رنگ آن‌ها زرد است.

[30] Nili, Hamed, et al. "A toolbox for representational similarity analysis." *PLoSComput. Biol* 10.4 (2014): e1003553.
 [31] Yu, Wei, et al. "Visualizing and Comparing Convolutional Neural Networks." *arXiv preprint arXiv:1412.6631* (2014).

۷ پیوست

۷-۱ آزمون آماری

در هر مطالعه، پژوهشگر برای اینکه تفاوتی را نشان دهد اول با این فرض شروع می‌کند که «در واقعیت تفاوتی بین گروه‌ها وجود ندارد»؛ سپس با فرض برقراری فرضیه H_0 (عدم تفاوت)، احتمال مشاهده نتایج مختلف را تخمین می‌زند. به H_0 فرضیه صفر نیز گفته می‌شود. یکی از مهم‌ترین روش‌ها برای نشان دادن تفاوت معنی‌دار بین دو گروه، محاسبه P-value است.

روش محاسبه P-value به این صورت است که ابتدا فرض می‌شود فرض صفر برقرار است و تفاوتی بین گروه‌ها وجود ندارد. سپس محاسبه می‌کنیم که چقدر احتمال دارد که به صورت تصادفی بین دو گروه تفاوتی را مشاهده کنیم. اگر این احتمال کم باشد، یعنی احتمال مشاهده‌ی تفاوت بین دو گروه به صورت تصادفی کم است. برعکس اگر مقدار p-value زیاد باشد یعنی احتمال اینکه بین دو گروه اختلافی به صورت تصادفی مشاهده شود زیاد است و این اختلاف بین دو گروه معنی‌دار نیست و فرض صفر برقرار است. سطح معنی‌داری معمولاً در علوم انسانی ۵٪ و در علوم پزشکی ۱٪ در نظر گرفته می‌شود، اما این سطح با توجه به حساسیت مسئله می‌تواند برای مسائل گوناگون تغییر کند. سطح معنی‌داری ۵٪ یعنی در ۱۰۰ نمونه از گروه در کمتر از ۵ نمونه، اختلاف بین نمونه‌ها تصادفی بوده و در ۹۵ نمونه دیگر اختلاف مشاهده شده واقعی است و بدین ترتیب این اختلاف بین دو گروه را پذیرفته و فرضیه صفر رد می‌شود.

۷-۲ تحلیل شباهت بازنمایی (RSA)

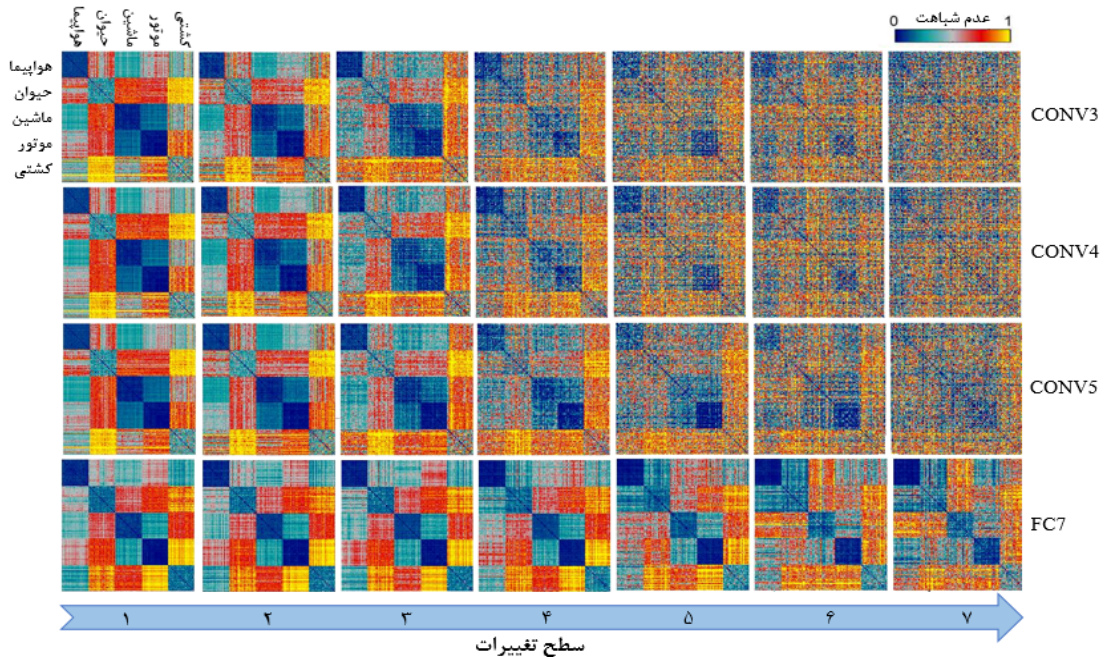
تحلیل شباهت بازنمایی یک ابزار است که ما را قادر می‌سازد تا آزمایش‌هایی را بدون اینکه دسته‌ی محرك از پیش تعریف شده باشد را به کار ببریم تا مدل‌های محاسباتی و ادراکی را آزمایش کرده و بازنمایی‌های بین انسان و میمون و یا مدل را مرتبط سازیم. در تحلیل شباهت بازنمایی، بازنمایی هر منطقه از مغز را ماتریس عدم شباهت بازنمایی توصیف می‌کند. (شکل ۱۴) یک ماتریس عدم شباهت بازنمایی یک ماتریس متقارن مربعی است که هر سلول به عدم شباهت بین الگوهای فعالیت مرتبط با دو محرك اشاره می‌کند. در شکل ۱۵ یک ماتریس عدم شباهت ایده‌آل نمایش داده شده است که نشان می‌دهد که شباهت هر

کلاس‌های مختلف). همان‌طور که در ماتریس‌های عدم‌شباهت مشاهده می‌شود، بازنمایی‌های دسته‌های هواپیما و کشتی نسبت به دسته‌های دیگر شباهت بیشتری دارند. با توجه به شباهت ظاهری این دو دسته (بدنه، بادبان، باله و غیره)، نتیجه گرفته می‌شود که مدل‌های مورد مقایسه، مانند انسان، اطلاعاتی از شکل اجسام را برای دسته‌بندی استفاده می‌کنند. علاوه بر این با توجه به ماتریس‌های عدم‌شباهت در سطوح بالای تغییرات، کاملاً مشخص است که VGGNet عملکرد بهتری دارد.

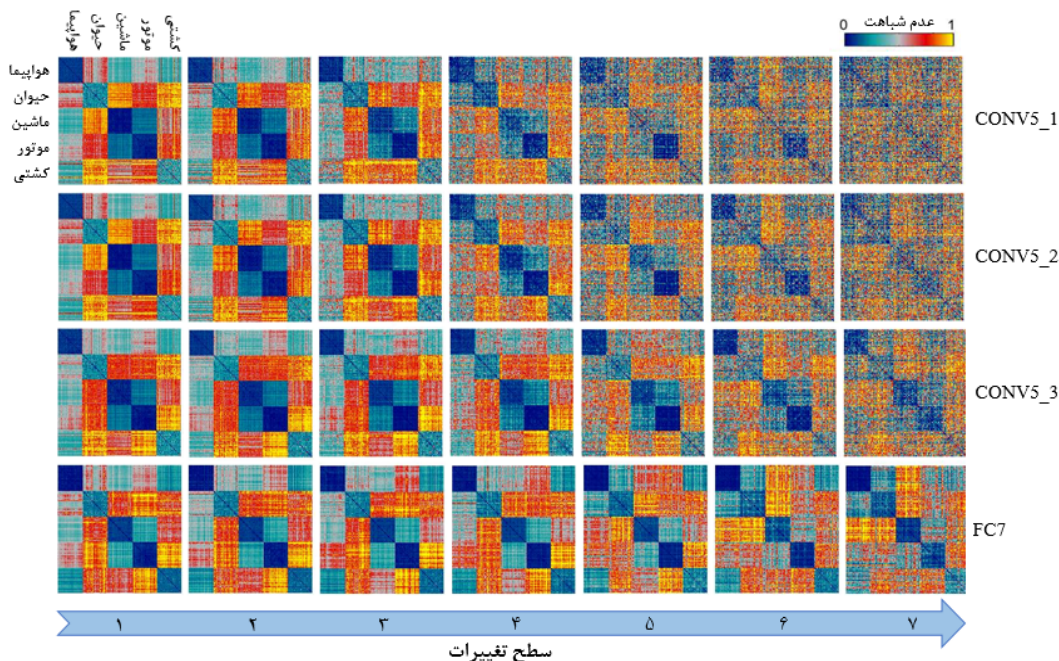
۳-۷ ماتریس‌های عدم‌شباهت بازنمایی

همان‌طور که در متن گفته شد ماتریس‌های عدم‌شباهت برای خروجی سه لایه کانولوشنی آخر و همچنین دومین لایه کاملاً متصل رسم شد.

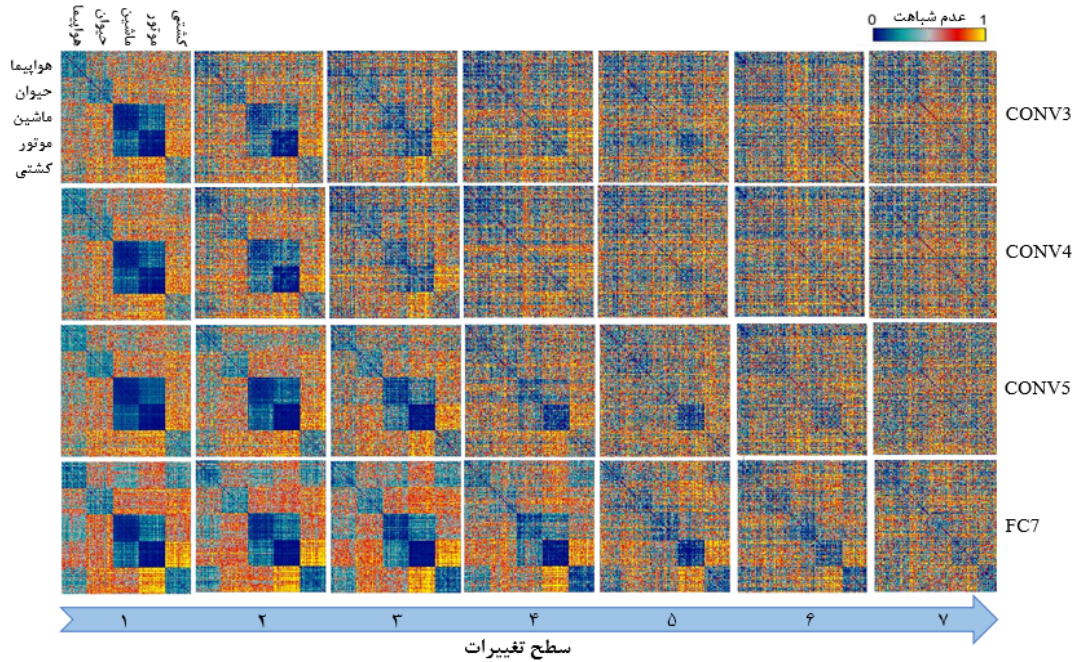
در این ماتریس‌ها رنگ آبی نشان‌دهنده شباهت زیاد و رنگ زرد نشان‌دهنده عدم‌شباهت است بنابراین انتظار داریم که قطر اصلی دارای رنگ آبی بوده و بقیه نواحی زرد رنگ باشد (شباهت زیاد بین اعضای یک کلاس و تفاوت زیاد بین



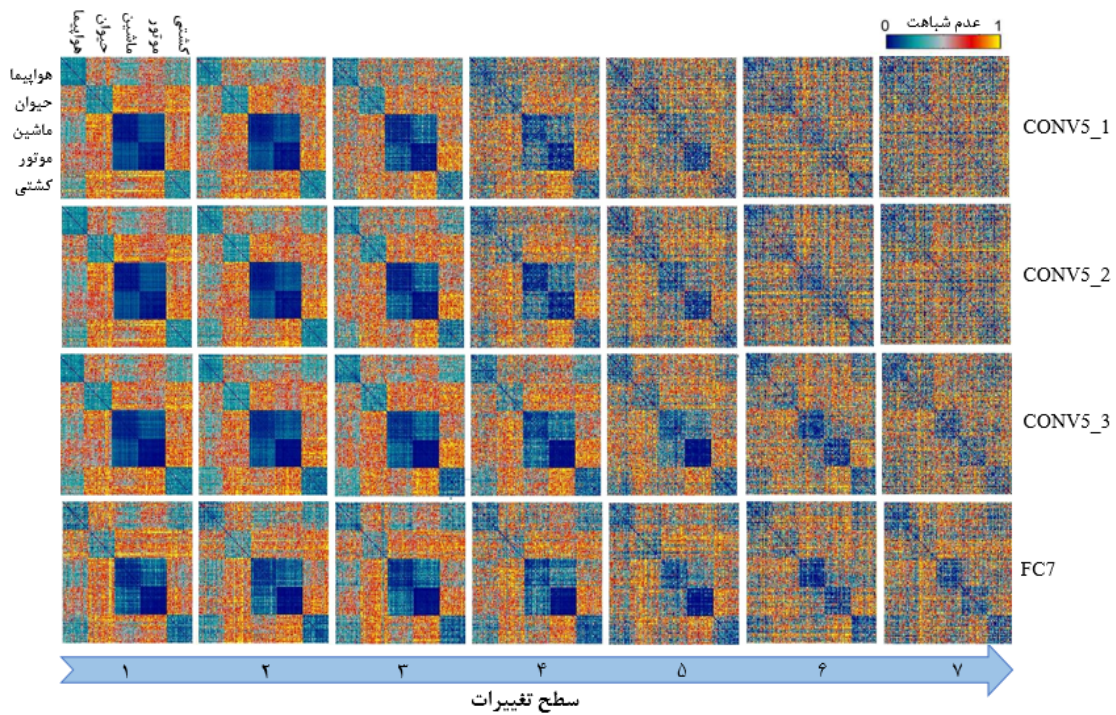
شکل ۱۶- رسم ماتریس‌های عدم‌شباهت برای خروجی‌های حاصل از لایه کاملاً متصل دوم و سه لایه کانولوشنی آخر AlexNet برای تصاویر بدون پس‌زمینه. مسئله دارای پنج دسته بوده و هر دسته شامل ۱۵۰ تصویر است. هر عنصر در ماتریس عدم‌شباهت بین بازنمایی حاصل از مدل برای یک جفت تصویر را نشان می‌دهد. هر ستون نشان‌دهنده سطح تغییرات تصاویر ورودی است. میله‌ی رنگی بالای شکل، میزان عدم‌شباهت را نشان می‌دهد (۱-correlation).



شکل ۱۷- رسم ماتریس‌های عدم‌شباهت برای خروجی‌های حاصل از لایه کاملاً متصل دوم و سه لایه کانولوشنی آخر مدل VGGNet برای تصاویر بدون پس‌زمینه.



شکل ۱۸- رسم ماتریس‌های عدم‌شباهت برای خروجی‌های حاصل از لایه کاملاً متصل دوم و سه لایه کانولوشنی آخر مدل AlexNet برای تصاویر دارای پس‌زمینه.



شکل ۱۹- رسم ماتریس‌های عدم‌شباهت برای خروجی‌های حاصل از لایه کاملاً متصل دوم و سه لایه کانولوشنی آخر مدل VGGNet برای تصاویر دارای پس‌زمینه.



فریبا عباسی مدرک کارشناسی خود را در رشته مهندسی برق گرایش الکترونیک در سال ۱۳۹۲ از دانشگاه شهید رجایی دریافت نمود. سپس ایشان مدرک کارشناسی ارشد خود را در رشته مهندسی برق گرایش کنترل در سال ۱۳۹۴ از دانشگاه شهید رجایی کسب نمود. ایشان از سال ۱۳۹۴ تاکنون به عنوان پژوهشگر در پژوهشکده علوم شناختی، پژوهشگاه دانش‌های بنیادی (IPM) فعالیت پژوهشی دارند. علاقه‌مندی‌های علمی ایشان شامل علوم اعصاب بینایی، بینایی ماشین، مدلسازی شناختی است.



رضا ابراهیم پور 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-بیوالکترونیک 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 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 بعلاوه از منتخبین سرآمدان علمی کشور توسط فدراسیون سرآمدان علمی کشور در سال ۱۳۹۴ می‌باشند. زمینه‌های 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 سیستم‌های هوشمند در سال ۱۳۹۰ از دانشگاه امیرکبیر دریافت نمود. ایشان از سال ۱۳۹۱ تا ۱۳۹۳ به عنوان محقق در پژوهشگاه دانش‌های بنیادی (IPM) مشغول به فعالیت پژوهشی بوده‌اند و از سال ۱۳۹۳ تاکنون دانشجوی مقطع دکتری تخصصی در رشته علوم اعصاب شناختی، پژوهشگاه دانش‌های بنیادی هستند. علاقه‌مندی‌های علمی ایشان شامل علوم اعصاب، مدلسازی محاسباتی، یادگیری ماشین و فلسفه ذهن است.