

ارائه یک روش مؤثر بازشناسی شیء مبتنی بر هرس کردن احتمالی مشخصه‌های دیداری تصویر در مدل HMAX

محمداسماعیل اکبرپور^۱، دانشجوی دکتری؛ ناصر مهرشاد^۲، دانشیار؛ سیدمحمد رضوی^۳، دانشیار

۱- دانشکده مهندسی برق و کامپیوتر- دانشگاه بیرجند- بیرجند- ایران - esmaeil.akbarpour@birjand.ac.ir

۲- دانشکده مهندسی برق و کامپیوتر- دانشگاه بیرجند- بیرجند- ایران - nmehrshad@birjand.ac.ir

۳- دانشکده مهندسی برق و کامپیوتر- دانشگاه بیرجند- بیرجند- ایران - smrazavi@birjand.ac.ir

چکیده: سیستم بینایی انسان قادر به بازشناسی شیء در صحنه‌های شلوغ با سرعت و دقت بالاست. مدل‌های زیادی برای بازشناسی شیء با الهام از سیستم بینایی انسان از جمله مدل HMAX معرفی شده‌اند. در این تحقیق، روشی مؤثر با عنوان انتخاب احتمالی HMAX (PSHMAX) برای بازشناسی شیء با حفظ ساختار مدل HMAX ارائه شده است. مشکل مدل HMAX انتخاب تصادفی تکه‌های تصویر است که سبب استخراج دو دسته از تکه‌های نامطلوب می‌شود. دسته اول تکه‌هایی با اطلاعات کم که بدون تولید خروجی مفید، بار محاسباتی سیستم را افزایش می‌دهند و دسته دوم تکه‌هایی با اطلاعات غیرمفید از پس‌زمینه که باعث تولید خروجی اشتباه می‌شوند. در مدل پیشنهادی، تکه‌هایی شامل اطلاعات حداکثری مفید با رویکرد تصادفی در دو مرحله استخراج می‌شوند: مرحله اول ایجاد استخری از تکه‌های شامل بیشترین اطلاعات و مرحله دوم استخراج تکه‌های شامل اطلاعات مفیدتر و بهینه از استخری برای ارزیابی، نرخ بازشناسی روش پیشنهادی با مدل HMAX و سایر روش‌های جدید توسعه‌یافته آن روی پایگاه‌های داده Caltech5 و Caltech101 مقایسه شده است. نتایج حاصل نشانگر برتری روش پیشنهادی است.

واژه‌های کلیدی: بازشناسی شیء، تکه‌های تصویر، سیستم بینایی انسان، لبه، مدل HMAX

New Effective Method for Object Recognition based on Probabilistic Pruning of Visual Characteristics in HMAX

Mohammadesmaeil Akbarpour¹, Ph.D. student; Nasser Mehrshad², Associate Professor; Seyyed-Mohammad Razavi³, Associate Professor

1- Faculty of Electrical and Computer Engineering, University of Birjand, Birjand, Iran, Email: esmaeil.akbarpour@birjand.ac.ir

2- Faculty of Electrical and Computer Engineering, University of Birjand, Birjand, Iran, Email: nmehrshad@birjand.ac.ir

3- Faculty of Electrical and Computer Engineering, University of Birjand, Birjand, Iran, Email: smrazavi@birjand.ac.ir

Abstract: The human visual system (HVS) recognizes object in the crowded scenes with high speed and accuracy. So far, many object recognition models based on HVS, like HMAX, have been developed. In this paper, the new effective method based on HMAX is proposed called Probabilistic Selection HMAX (PSHMAX). HMAX main problem is random patch extraction which extracts two useless patches. First, patches involving low information that cause more computational complexity with no useful result. Second, patches with wrong information from background that produce wrong output. In the proposed method, the optimum patches involving maximum useful information are extracted in the random way which has two steps: first is producing poll of patches involving maximum information, second is patches extracting with useful information from poll. To evaluate the proposed method, we apply it to object categorization and conduct experiment on the Caltech5 and Caltech101 databases. Results demonstrate that the proposed method has a higher performance than the HMAX and existing architectures having a similar framework.

Keywords: Object Recognition, Image Patches, Human Visual System, Edge, HMAX Model

تاریخ ارسال مقاله: ۱۳۹۶/۰۲/۲۲

تاریخ اصلاح مقاله: ۱۳۹۶/۰۶/۱۶

تاریخ پذیرش مقاله: ۱۳۹۶/۰۸/۲۷

نام نویسنده مسئول: ناصر مهرشاد

نشانی نویسنده مسئول: ایران - بیرجند - دانشگاه تبریز - دانشکده مهندسی برق و کامپیوتر.

۱- مقدمه

یکی از موضوعات مطرح در زمینه تحلیل تصویر، بازشناسی سریع و درعین حال دقیق شیء می‌باشد [۱] که سیستم بینایی انسان به راحتی و در یک زمان بسیار کوتاه قادر به انجام آن است [۲]. این در حالی است که عملکرد ماشین‌های محاسباتی علیرغم بالا بودن سرعت و دقت عملیات پایه‌ای فاصله زیادی با سیستم بینایی انسان دارند. به دلیل شناخت ناکافی از ساختار و عملکرد سیستم بینایی انسان از یک طرف و فقدان الگوریتم‌های مناسب از طرف دیگر، امکان مدل‌سازی مؤثر این توانایی خارق‌العاده انسان هنوز فراهم نشده است.

عملکرد ایده‌آل سیستم بینایی انسان در بسیاری از زمینه‌های پردازش تصویر باعث شده است که ارزیابی بسیاری از الگوریتم‌ها به صورت کیفی و تنها با نگاه کردن به تصویر خروجی انجام شود. در بسیاری از زمینه‌های تحلیل تصویر نیز ارزیابی‌ها با مقایسه خواص استخراج شده از تصویر با خواص مشابه درک شده توسط سیستم بینایی انسان و به صورت کیفی انجام می‌شود. به عبارت دیگر می‌توان گفت که ارزیابی نهایی الگوریتم‌های پردازش تصویر سیستم بینایی انسان می‌باشد [۲]. این در حالی است که تلاش محققان برای شناخت آن تاکنون بی‌نتیجه مانده و بیشتر کارهای انجام شده در این زمینه هنوز در مراحل اولیه خود قرار دارد.

سیستم بینایی انسان به طور عمده توسط هابل و ویسل^۲ بر پایه مطالعات روی قشر بینایی گربه [۳] و میمون [۴] مدل و منجر به معرفی سلول‌های ساده و پیچیده شد. این مدل‌ها که به صورت پایه‌ای در بسیاری از الگوریتم‌های پیچیده پردازش و تحلیل تصویر جایگاه خود را پیدا کرده‌اند بیشتر بر اساس حساسیت سیستم بینایی انسان به خطوط و لبه‌ها توسعه یافته‌اند. مدل‌سازی سیستم بینایی انسان از اولین ناحیه پردازشی در قشر بینایی تا سطوح بالاتر، مربوط به میلی‌ثانیه‌های ابتدایی پردازش بینایی را شامل می‌شود [۵]. با بررسی نورون‌های بینایی، یک نظریه قوی در بینایی انسان به نام نظریه مقیاس-مکان ارائه شده است. طبق نظریه مقیاس-مکان، از ابتدا تا انتهای قشر بینایی با یک دسته از عملگرهای گوسی محلی و مدرج^۳ مدل می‌شود. این مدل‌سازی را می‌توان به کمک شبکه‌های عصبی چندلایه و بر اساس آرایش متناوب سلول‌های ساده و پیچیده قشر بینایی اجرا کرد. برای سلول‌های ساده از طرح کانولوشن و برای سلول‌های پیچیده از طرح تجمیع استفاده می‌شود. رایزن‌ها^۴ و پوگیو^۶ با گسترش مطالعات هابل و ویسل یک مدل سلسله‌مراتبی برای بازشناسی اشیاء معرفی کردند. آن‌ها عملگر مشتق گوسی (DOG^۵) را برای مدل کردن سلول‌های ساده و دو عملگر مجموع و بیشینه را برای مدل‌سازی سلول‌های پیچیده قشر بینایی معرفی کردند.

نتیجه تحقیق سر^۷ و همکاران [۷، ۸] روی گسترش مدل رایزن‌ها^۴، معرفی مدل HMAX بود که به دسته مدل‌های سلسله‌مراتبی تعلق دارد. این مدل، از مدل‌های سلسله‌مراتبی بدون پس‌خوردی است که ساختار و پارامترهای آن بر اساس خصوصیات

بیولوژیکی قشر بینایی انتخاب شده است. مهم‌ترین پارامتر در بازشناسی شیء مقاوم بودن آن به جابجایی و تغییر اندازه است که با جابجا و یا بزرگ و کوچک شدن شیء، سیستم دچار تشخیص اشتباه نشود. عملکرد فیلترهای مدل HMAX از عملکرد فیلترهای مقاوم در برابر جابجایی و تغییر اندازه (SIFT^۹) بهتر است [۹، ۱۰].

اساس کار در مدل HMAX ابتدا مقاوم‌سازی همه تصاویر به جابجایی و تغییر اندازه، سپس انتخاب تکه در مرحله آموزش و استفاده از این تکه‌ها در مرحله آزمون است. مشکل اصلی مدل HMAX انتخاب کاملاً تصادفی این تکه‌هاست به طوری که هیچ پالایشی روی آن‌ها صورت نمی‌گیرد. یک نمونه از تکه‌های غیرمفید که در انتخاب تصادفی ممکن است استخراج شوند تکه‌های با اطلاعات کم است. چون شناخت ما از ساختار و عملکرد سیستم بینایی کامل نیست در انتخاب تکه‌های تصویر می‌توانیم فرضیه‌هایی را مطرح و مدل کنیم، خود این امر می‌تواند به شناخت بهتر ساختار و عملکرد سیستم بینایی انسان کمک کند. به عنوان فرضیه اول، سیستم بینایی انسان در انتخاب تکه‌های تصویر تصادفی عمل نکرده و تکه‌هایی انتخاب می‌کند که شامل بیشترین اطلاعات باشند. نمونه دیگر تکه‌های غیرمفید، تکه‌های تصادفی از پس‌زمینه است، بنابراین نمی‌توان تضمین کرد که تکه‌ها به شیء اختصاص داشته باشند. این اتفاق نه تنها بار محاسباتی بیشتری به سیستم تحمیل می‌کند، حتی باعث تولید خروجی اشتباه نیز می‌شود. به عنوان فرضیه دوم، سیستم بینایی انسان در انتخاب تکه‌ها بین تصاویر یک گروه، تکه‌هایی را به عنوان شاخص انتخاب می‌کند که بیشترین شباهت را با تکه‌های دیگر هم‌گروه داشته باشند.

اگر تصویر خود دارای پس‌زمینه از شیء دیگر باشد قابل حذف کردن نیست، ولی بعضی از تصاویر طبیعی پس‌زمینه‌های یکرنگ (مانند آسمان یا دیوار) دارند که با الگوریتم‌های حذف پس‌زمینه [۱۱] قابل حذف می‌باشند؛ بنابراین در ابتدا برای کل تصاویر از الگوریتم حذف پس‌زمینه استفاده می‌کنیم تا برای تصاویر دارای پس‌زمینه قابل حذف، پس‌زمینه حذف شود. یک سیستم طبقه‌بندی خوب باید همواره به اشیاء داخل یک گروه پاسخ مشابه (خاصیت تغییرناپذیری) و به اشیاء بین گروه‌ها پاسخ متفاوت (خاصیت انتخاب‌پذیری) بدهد [۱۲]. برای نزدیک شدن به این خاصیت، باید تکه‌هایی انتخاب شوند که شامل اطلاعات بیشتر و درعین حال مفیدتر باشند. تکه‌های شامل اطلاعات بیشتر تکه‌هایی با سطح رنگ بالاتر و لبه بیشتر می‌باشند. برای تشخیص تکه‌های مفیدتر نیز کافی است شباهت اطلاعات تکه با تکه‌های داخل یک گروه سنجیده شود. از آنجایی که مشکل جابجایی و تغییر اندازه قبلاً حل شد تکه‌های یک گروه باید حداکثر شباهت با یکدیگر را داشته باشند.

برای استخراج تکه‌هایی با خصوصیات فوق، تکه‌ها را در دو مرحله جداگانه پالایش می‌کنیم. در مرحله اول تکه‌های شامل اطلاعات بیشتر را انتخاب می‌کنیم، به این منظور تکه‌هایی که شامل اطلاعات کمی می‌باشند حذف می‌شوند و از بین تکه‌های باقی‌مانده، متناسب با میزان

را دارند و سبب رسیدن به حداکثر دقت توأم مکان و فرکانس می‌شوند [۱۴]. فرمول فیلترهای گابور در مختصات دکارتی x و y که در حقیقت ضرب تابع گوسین و سینوسی است در رابطه (۱) بیان شده است [۷]:

$$h_{\sigma,\lambda,\theta} = \exp\left(-\frac{x_0^2 + y_0^2}{2\sigma^2}\right) \cos\left(\frac{2\pi}{\lambda} x_0\right) \quad (1)$$

$$x_0 = x \cos(\theta) + y \sin(\theta)$$

$$y_0 = y \cos(\theta) - x \sin(\theta)$$

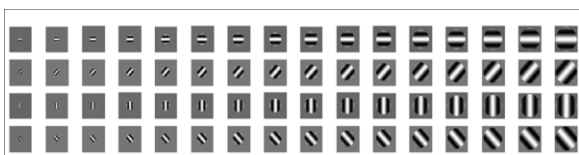
که در آن σ طول تابع گوسین، θ جهت گوسین، λ نرخ نمو و طول موج است. در شکل ۱ فیلترهای گابور استفاده شده در لایه S_1 رسم شده است. اگر تصویر ورودی را I بنامیم، خروجی لایه اول حاصل کانولوشن تصویر ورودی و فیلتر گابور می‌باشد در رابطه (۲) بیان شده است:

$$S_{I,\theta}^1 = h_{\sigma,\lambda,\theta} * I \quad (2)$$

که در آن I مقیاس می‌باشد و از یک تا شانزده تغییر می‌کند. تصویر ورودی با فیلترهای گابور در ابعاد مختلف (از 7×7 تا 37×37 با گام دو پیکسل) کانولوشن می‌شود بنابراین خروجی لایه اول، شانزده حالت مختلف دارد که در این تحقیق هر حالت یک مقیاس نامیده می‌شود. همه فیلترها دارای هر چهار جهت زاویه $(0^\circ, 45^\circ, 90^\circ, 135^\circ)$ می‌باشند، در نتیجه تعداد کل خروجی‌های لایه اول برابر ۶۴ (چهار جهت و شانزده مقیاس) می‌باشد. شانزده تصویر خروجی لایه اول در هر جهت، به هشت باند تقسیم می‌شوند که پارامترهای مربوط به چهار باند اول در جدول ۱ آمده است (برای اطلاعات هر هشت باند به مرجع [۷] مراجعه کنید). خروجی‌های این لایه برای تصویر قایق از پایگاه داده Caltech101 که در شکل ۲-الف نمایش داده شده است، در چهار مقیاس ۱، ۲، ۷ و ۸ در زوایای $\theta = 0^\circ, 90^\circ$ در ۰-ب آمده است.

۲-۴ لایه C_1

خروجی‌های لایه اول به این لایه فرستاده می‌شوند، این لایه فعالیت سلول‌های پیچیده را شبیه‌سازی می‌کند. این سلول‌ها نسبت به تغییر موقعیت و اندازه تصاویر مقاوم می‌باشند که توسط عملگر بیشینه در این لایه شبیه‌سازی می‌شود. عملگر بیشینه روی دو مقیاس مجاور با جهت‌های مشابه (مانند مقیاس‌های (۱ و ۲) با پنجره‌هایی به ابعاد $N_s \times N_s$ و همپوشانی Δ_s (اطلاعات چهار باند اول در جدول ۱ نشان داده شده است) به منظور ایجاد ویژگی تغییرناپذیری به اندازه و موقعیت اعمال می‌شود، به این ترتیب ۳۲ تصویر به دست می‌آید (۴



شکل ۱: فیلترهای گابور لایه اول در چهار زاویه و شانزده اندازه [۱۵]

اطلاعات موجود در تکه‌ها، تکه‌های شامل اطلاعات بیشتر انتخاب می‌شوند. به این معنی که هرچه اطلاعات موجود در یک تکه بیشتر باشد احتمال انتخاب آن نیز بیشتر است، در این مرحله استخری از تکه‌ها ایجاد می‌شود. در مرحله دوم، از بین تکه‌های استخر تکه‌هایی نهایی می‌شوند که درون گروهی دارای حداکثر شباهت باشند.

از بین تکه‌های نهایی شده، اگر شباهت یک تکه به تکه‌های هم‌گروه زیاد باشد یعنی تکه بسیار خوبی برای آن گروه آموزشی است. لذا با اطمینان از بهینه بودن تکه در گروه موردنظر، تکه مربوطه متمم شده $(1-p)$ و به گروه‌های دیگر اضافه می‌شود. فرض کنید تکه بهینه متعلق به گروه A بوده و متمم آن به گروه‌های دیگر اضافه شود، در این صورت اگر تصویر آزمون متعلق به گروه A باشد احتمال تشخیص آن به گروه‌های دیگر بسیار کاهش می‌یابد. چون این تصویر با تکه متمم دارای بیشترین فاصله و کمترین شباهت است. به این ترتیب احتمال تشخیص اشتباه تصویر آزمون متعلق به گروه A در گروه‌های دیگر به حداقل می‌رسد.

نتایج شبیه‌سازی نشان می‌دهد که مدل معرفی شده نسبت به مدل HMAX و مدل‌های توسعه یافته مربوطه، کارایی بهتری دارد. برای ارزیابی مدل معرفی شده، از کمیت‌های نرخ بازشناسی، دقت و بازخوانی بر روی پایگاه‌های داده استاندارد Caltech5، Caltech101 و GRAZ-01 استفاده شده است که در مقایسه با روش‌های موجود نتایج بهتری می‌دهد. با توجه به این‌که تغییر تنها در مرحله آموزش انجام می‌شود، سرعت اجرای مدل در مرحله آزمون تغییری نکرده است.

۲-معرفی مدل HMAX

مدل عمومی HMAX، بر اساس آرایش سلول‌های ساده و پیچیده موجود در قشر بینایی که به طور متناوب قرار گرفتند، پیشنهاد شد [۶، ۷]. این مدل که بر اساس نظریه سلسله‌مراتبی پردازش بینایی مدل هابل و ویسل استنتاج و بنا شده، به دسته مدل‌های سلسله‌مراتبی تعلق دارد و از طرح متناوب کانولوشن/تجمیع [۱۳] استفاده می‌کند. مدل HMAX، از چهار لایه تشکیل شده است که انتخاب‌پذیری و تغییرناپذیری به صورت سلسله‌مراتبی در آن افزایش می‌یابد. واحدهای ساده برای افزایش انتخاب‌پذیری، ورودی‌هایشان را با یک سری توابع خاص مثل گابور و مشتق گوسین کانولوشن می‌کنند که در هر مرحله از کانولوشن مجموعه‌ای از نقشه‌های ویژگی فراهم می‌شود. واحدهای پیچیده نیز برای افزایش تغییرناپذیری، خروجی‌های واحدهای ساده را به عملگرهای غیرخطی مثل بیشینه و مجموع اعمال می‌کنند. در ادامه لایه‌های مختلف معرفی می‌شوند:

۲-۴ لایه S_1

این لایه فعالیت سلول‌های ساده را شبیه‌سازی می‌کند. در این لایه تصویر ورودی با مجموعه‌ای از فیلترهای گابور $(h_{\sigma,\lambda,\theta})$ کانولوشن می‌شود. فیلترهای گابور، موجک‌های گوسی مدوله شده‌ای هستند که توانایی استخراج ویژگی‌های تصویر در جهت‌ها و فرکانس‌های متفاوت

جدول ۱: پارامترهای لایه‌های اول و دوم برای هر جهت [۷]

I	لایه S ₁			لایه C ₁			
	فیلتر گالور			خروجی S ₁ و ورودی C ₁	عملگر بیشینه		خروجی
	سایز				N _s × N _s	Δ _s	
تصویر ورودی	۷×۷	۲/۸	۳/۵	l=1 l=2	۸×۸	۴	باند ۱
	۹×۹	۳/۶	۴/۶				
	۱۱×۱۱	۴/۵	۵/۶	l=3 l=4	۱۰×۱۰	۵	باند ۲
	۱۳×۱۳	۵/۴	۶/۸				
	۱۵×۱۵	۶/۳	۷/۹	l=5 l=6	۱۲×۱۲	۶	باند ۳
	۱۷×۱۷	۷/۳	۹/۱				
	۱۹×۱۹	۸/۲	۱۰/۳	l=7 l=8	۱۴×۱۴	۷	باند ۴
	۲۱×۲۱	۹/۲	۱۱/۵				

۲ لایه C₂

خروجی نهایی شبکه که پاسخی مقاوم در مقابل تغییر مقیاس و جابجایی می‌باشد، با یافتن بیشینه عمومی روی همه مقیاس‌ها و موقعیت‌ها برای هر نمونه لایه S₂ محاسبه می‌شود. پس از تطبیق تصویر ورودی با تکه‌های ذخیره‌شده در همه موقعیت‌ها و مقیاس‌ها،

جهت و ۸ باند). خروجی این لایه (C_{1b,θ}), بیشینه بین مقیاس‌های l و l+1 است که در رابطه (۳) بیان شده است:

$$C_{b,\theta}^l(i, j) = \text{Max}\{S_{l,\theta}^1(i, j), S_{l+1,\theta}^1(i, j)\} \quad (3)$$

$$1 \leq i, j \leq N_s$$

که در آن ابعاد پنجره‌های عملگر بیشینه، i و j شماره پیکسل‌ها، b شماره باند، θ جهت گوسین و l نشان‌دهنده مقیاس‌هاست (l=1,3,5,7,9,11,13,15). خروجی‌های میانی C_{1,0}, C_{1,90}, C_{1,45} و C_{1,135} در ۰-ج نمایش داده شده است.

فرآیند استخراج تکه در مرحله آموزش و پس از عبور از دو لایه اول انجام می‌گیرد. تعداد زیادی تکه‌های تصادفی از تصاویر آموزشی با ابعاد مختلف (n×n, n=4, 8, 12, 16) در سطح لایه C₁ در هر چهار جهت (۰°, ۴۵°, ۹۰°, ۱۳۵°) استخراج و ذخیره می‌شوند. اندازه تکه‌های استخراجی در مدل HMAX صرف‌نظر از اندازه لایه C₁ می‌باشد و در همه باندها ثابت است. از این تکه‌های ذخیره‌شده در مرحله آزمون استفاده می‌شود.

۲ لایه S₂

این لایه به‌عنوان تابع پایه شعاعی^۱ (RBF) رفتار می‌کند. خروجی لایه S₂ فاصله اقلیدسی گوسی‌شکل بین ورودی جدید و یک تکه نمونه ذخیره‌شده (P_i) می‌باشد. فرض کنید تصویر لایه C₁ به ابعاد K×L باشد و پنجره‌های به ابعاد W×W از آن که از مختصات (i, j) آغاز می‌شود با X نشان داده شود. خروجی این لایه در رابطه (۴) بیان شده است:

$$S_2(i, j) = \exp(-\beta \|X - P_i\|) \quad (4)$$

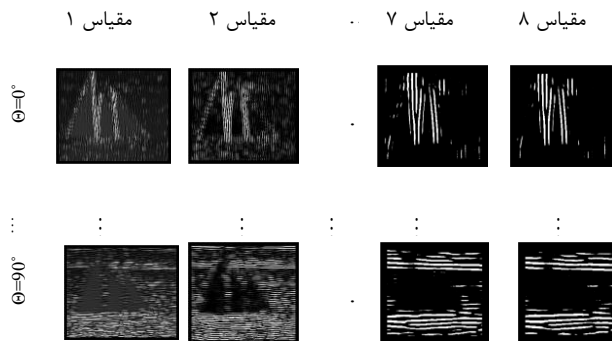
$$1 \leq i \leq K - W + 1$$

$$1 \leq j \leq L - W + 1$$

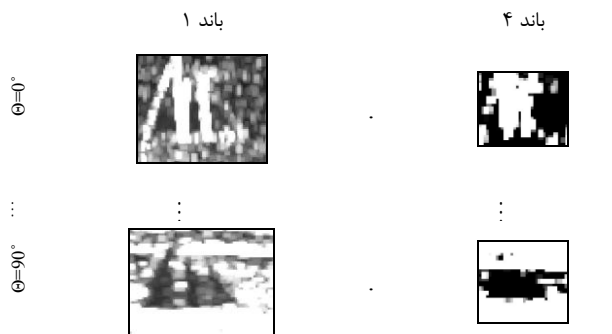
که در آن β میزان وضوح^۱ و P_i همان تکه تصویر است که به‌عنوان مرکز واحد RBF در نظر گرفته می‌شود. تعداد کل مراکز RBF متناسب با تعداد نمونه‌های اولیه استخراج‌شده در مرحله آموزش است.



(الف)



(ب)



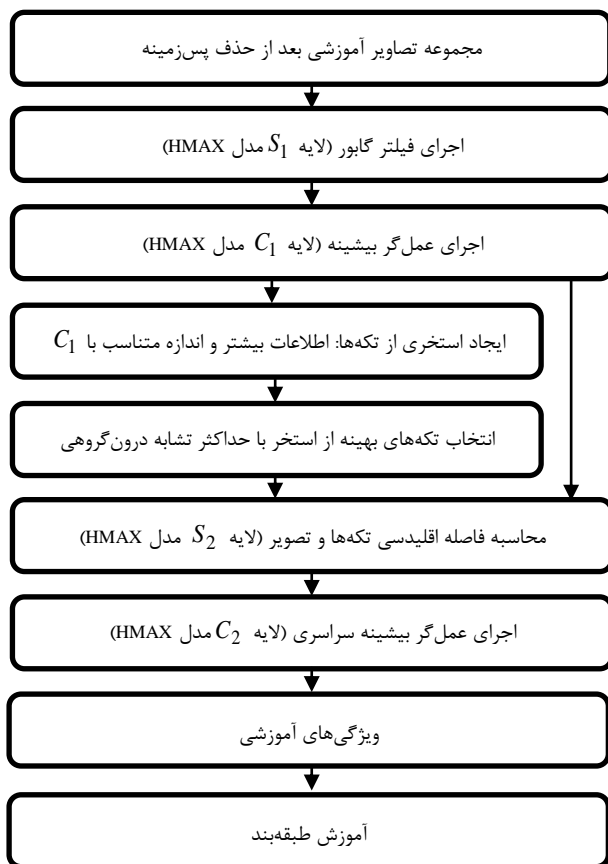
(ج)

شکل ۲: (الف) شکل اولیه از پایگاه داده، (ب) خروجی میانی لایه اول، (ج) خروجی لایه دوم برای لایه اول داده شده.

مدل GMAX پرداختند. آن‌ها تصاویر را به سه بخش آموزشی، ارزیابی و آزمون تقسیم کردند و نرخ بازشناسی روی تصاویر ارزیابی را به‌عنوان تابع شایستگی انتخاب کردند. مدل معرفی شده، به دلیل کاهش تعداد تکه‌ها و جمع‌آوری تکه‌های مفید، علاوه بر افزایش کارایی سبب کاهش زمان شبیه‌سازی نیز شد. در این مدل روی تکه‌ها تمرکز نشد بلکه فقط تکه‌هایی که تابع شایستگی را بهینه کنند به‌عنوان تکه‌های نهایی انتخاب شدند. کروپا^{۱۹} و همکارانش [۲۶] با تعریف یک فرآیند و الگوریتم یادگیری، تکه‌های بهتر را انتخاب کردند. هوانگ و همکارانش [۲۷] بجای استخراج تصادفی تکه، از نواحی برجسته تصویر (نواحی با مقادیر بیشتر پیکسل‌ها) تکه استخراج کردند.

۴- مدل پیشنهادی

یکی از مشکلات اساسی مدل HMAX این است که تکه‌های ذخیره‌شده در مرحله آموزش کاملاً تصادفی هستند، لذا بسیاری از تکه‌های تصویر یا شامل اطلاعات کافی نبوده یا شامل اطلاعات اشتباه می‌باشند (همانند تکه‌هایی با اعداد تقریباً مشابه و یا تکه‌هایی با اطلاعات پس‌زمینه). این تکه‌ها فقط بار محاسباتی سیستم را افزایش می‌دهند و حتی گاهی سبب تولید خروجی اشتباه نیز می‌شوند، بنابراین پالایش تکه‌های تصویر ضروری به نظر می‌رسد. از آنجایی که مدل پیشنهادی یک مدل طبقه‌بند تصاویر می‌باشد و دارای دو مرحله



شکل ۳: مرحله آموزش مدل پیشنهادی

بازرزش‌ترین‌ها به‌عنوان ویژگی در نظر گرفته می‌شوند. به این ترتیب خروجی مورد تأیید نهایی که خاصیت تغییرناپذیری کلی را نیز دارد، به دست می‌آید.

۳- کارهای انجام شده

با توجه به این که مدل عمومی HMAX بر اساس نظریه سلسله‌مراتبی پردازش بینایی بنا شده است، با مطالعات بیشتر روی سیستم بینایی می‌توان آن را بهبود بخشید. لی^{۱۰} و همکاران [۱۶] با الگوبرداری از سازوکار توجه مربوط به لایه V_1 ، در لایه S_1 از مدل برجستگی استفاده کردند. آن‌ها همچنین با الگوبرداری از سازوکارهای یادگیری، خوشه‌بندی و همچنین حافظه‌ها در لایه V_2 و IT ، از روش خوشه‌بندی تکراری برای استخراج تکه استفاده کردند. کیاو^{۱۱} و همکاران [۱۷] با الگوبرداری از شبکه‌های عصبی، یک تشخیص اولیه برای انواع مختلف ورودی ارائه نمودند و ویژگی‌های استخراج‌شده را خوشه‌بندی کردند. هوانگ^{۱۲} و همکاران [۱۸] یک مدل بهبودیافته بیولوژیکی از روی مدل HMAX ارائه نمودند که با دو تغییر اساسی در مدل HMAX معرفی شد: تغییر اول کاهش ویژگی‌ها با استفاده از گرادینان افقی و عمودی و تغییر دوم استفاده از پس‌خورد بر اساس یک الگوریتم خاص بود. سریالت^{۱۳} [۱۹] با معرفی ECHMAX بجای استفاده از بیشینه سراسری روی لایه S_2 ، به‌ازای هر تکه، شش بیشینه در نظر گرفت. بیشینه‌ها در ناحیه‌هایی به مرکز تکه و شعاع‌هایی با اندازه ۵، ۱۰، ۳۰، ۵۰، ۷۰ و ۱۰۰ درصد تصویر محاسبه می‌شوند، در این مدل روی انتخاب تکه‌های مفید تمرکز نشد.

از آنجایی که مدل HMAX یک مدل سلسله‌مراتبی چهارلایه‌ای می‌باشد می‌توان هر یک از لایه‌های آن را به‌صورت جداگانه تغییر و بهبود بخشید. در مدل معرفی شده HMAX-s [۲۰] توسط سریالت و همکاران، در لایه S_2 از ضرب کسینوسی به‌جای فاصله اقلیدوسی استفاده شده است. نتایج نشان داد نه تنها سرعت بلکه نرخ بازشناسی نیز افزایش یافت. در این روش نیز روی انتخاب تکه بهینه تمرکز نشد. ایستاشیو^{۱۴} و همکاران [۲۱] زمان اجرا را، با استفاده از تکنیک خوشه‌بندی در لایه C_2 کاهش دادند. ژانگ^{۱۵} و همکاران [۲۲] با اضافه کردن یک لایه به نام H به لایه‌های قبلی به معرفی یک مدل جدید پرداختند.

بسیاری از کارهای انجام شده در راستای انتخاب تکه‌های مفید بجای تکه‌های تصادفی هستند، لی^{۱۶} و همکاران [۲۳] با ترکیب الگوریتم FAST و مدل HMAX به معرفی مدل جدید به نام DOHMAX^{۱۷} پرداختند که نهایتاً منجر به استخراج پیش‌نمونه بهتر گردید. اساس کار انتخاب تکه‌ها از گوشه‌های تصویر بود که به‌تنهایی نمی‌تواند برای همه تصاویر مفید باشد. موج و لاو^{۱۸} [۲۴] با اعمال محدودیت‌هایی در ویژگی‌های ورودی مانند محاسبه جهت غالب توانستند تعداد ویژگی‌ها را کاهش دهند. قدرتی و همکارانش در [۲۵] با استفاده از الگوریتم‌های تکاملی برای انتخاب تکه بهینه به معرفی

جدول ۲: اندازه تکه‌های هر باند متناسب با اندازه لایه C_1

شماره باند	اندازه لایه C_1	اندازه تکه‌ها
باند ۱	30×30	$12 \times 12 \times 16$
باند ۲	22×22	$10 \times 10 \times 14$
باند ۳	17×17	$8 \times 8 \times 12$
باند ۴	13×13	$6 \times 6 \times 10$
باند ۵	10×10	$5 \times 5 \times 8$
باند ۶	8×8	$4 \times 4 \times 6$
باند ۷	6×6	$3 \times 3 \times 4$
باند ۸	5×5	$2 \times 2 \times 3$

شد. برای حل مشکلات مذکور، در مدل پیشنهادی برای انتخاب تکه از یک الگوریتم استفاده می‌شود.

دو تفاوت عمده بین تکه‌های مدل HMAX و مدل پیشنهادی وجود دارد: تفاوت اول در اندازه تکه‌هاست، در مدل HMAX در هر هشت باند اندازه تکه‌ها یکسان است ولی در مدل پیشنهادی با توجه به تغییر اندازه لایه C_1 در هر باند، اندازه تکه‌ها نیز برای هر باند متفاوت است. اندازه تکه‌ها در باندهای مختلف، با فرض اندازه تصویر ورودی ۱۴۰ در ۱۴۰، در جدول ۲ بیان شده است. تفاوت دوم، اطلاعات موجود در تکه‌هاست، معیار اصلی انتخاب تکه‌های نهایی این است که شامل اطلاعات بیشتر و درعین حال مفیدتر باشند.

گزینش اطلاعات بیشتر و مفیدتر در دو مرحله انجام می‌شود. مرحله اول انتخاب تکه‌ها با اطلاعات بیشتر است. همان‌گونه که در مقدمه اشاره شد، تکه‌هایی که دارای سطح رنگ بالاتر و لبه بیشتر باشند، شامل اطلاعات بیشتری هستند. برای شبیه‌سازی سطح رنگ از میانگین و لبه از واریانس استفاده می‌شود. با توجه به این شبیه‌سازی، برای استخراج تکه‌های شامل اطلاعات بیشتر از تکه‌ها واریانس و میانگین گرفته می‌شود که معرف میزان اطلاعات موجود در تکه‌ها هستند. محاسبه و مقایسه این دو معیار در هر باند و برای هر جهت به‌طور جداگانه انجام می‌شود. برای هر کدام از این کمیت‌ها مقدار آستانه در نظر می‌گیریم (آستانه می‌تواند ضریبی از مقدار متوسط این کمیت‌ها باشد) و پس از تعریف آستانه مناسب، به‌عنوان گام نخست تکه‌هایی که هر دو پارامتر آن‌ها از مقدار آستانه بیشتر باشند را انتخاب می‌کنیم. به‌عنوان گام دوم، از بین آن‌ها و با حفظ رویکرد تصادفی تکه‌های شامل اطلاعات بیشتر را انتخاب می‌کنیم. برای حفظ رویکرد تصادفی از یک طرف و اطمینان از انتخاب تکه‌های شامل اطلاعات بیشتر از طرف دیگر، احتمال انتخاب تکه‌ها را متناسب با درصد اطلاعات موجود در یک تکه در مقایسه با اطلاعات موجود در کل تکه‌ها می‌گیریم (PS_f). این درصد برای هر تکه با توجه به اندازه، باند و جهت آن، جداگانه انجام می‌شود که برای تحقق آن از چرخه رولت

آموزش و آزمون است، معماری مدل پیشنهادی در دو مرحله توضیح داده می‌شود.

۴ مرحله آموزش مدل پیشنهادی

معماری مرحله آموزش مدل پیشنهادی در ۰ نشان داده شده است. برخلاف بعضی کاربردها همچون ردیابی اهداف متحرک [۲۸] که تصاویر پس‌زمینه می‌تواند مفید باشد، در بازشناسی شیء تصاویر پس‌زمینه می‌توانند سبب تولید خروجی اشتباه شوند. همان‌گونه که در شکل مشخص است در ابتدا از الگوریتم حذف پس‌زمینه [۱۱] استفاده می‌کنیم. هرچند استفاده از این الگوریتم نمی‌تواند در همه تصاویر مفید باشد ولی در تصاویری که قابلیت حذف پس‌زمینه وجود دارد بسیار مفید می‌باشد و سبب می‌شود تصویر به‌صورت حداکثری شامل شیء باشد. نمونه مفید حذف پس‌زمینه در شکل ۴ نشان داده شده است، در این شکل در تصویری مانند هواپیما که پس‌زمینه آن آسمان است الگوریتم حذف پس‌زمینه بسیار مفید است و از تحمیل بار محاسباتی اضافه به سیستم جلوگیری می‌کند. از آنجایی که این الگوریتم بر پایه آستانه‌گذاری خودکار می‌باشد، فقط برای تصاویری که دارای پس‌زمینه یکسان هستند قابل استفاده می‌باشد. لذا در صورت وجود پس‌زمینه غیریکسان و یا وجود اشیاء دیگر در پس‌زمینه، این الگوریتم مفید نخواهد بود. در بسیاری از تصاویر پایگاه‌های داده همانند هواپیما و موتور این الگوریتم کمک شایانی خواهد کرد.

پس از حذف پس‌زمینه، تصویر باید از دو لایه عبور کند که همانند دو لایه اول مدل HMAX هستند. اولین لایه همانند لایه I_1 در مدل HMAX می‌باشد، لایه‌ای که با اعمال شانزده فیلتر گابور با اندازه‌های مختلف و در چهار جهت ($0^\circ, 45^\circ, 90^\circ, 135^\circ$) به‌دست می‌آید که در مجموع برای یک تصویر ورودی ۶۴ خروجی تولید خواهد کرد. لایه بعداز آن همانند لایه C_1 در مدل HMAX می‌باشد، لایه‌ای که با محاسبه بیشینه بین هر دو مقیاس متفاوت در هر چهار جهت به‌دست می‌آید که در مجموع ۳۲ خروجی تولید خواهد کرد.

از آنجایی که تصویر پس از عبور از لایه C_1 نسبت به جابجایی و تغییر اندازه مقاوم می‌شود، لذا تکه‌ها در این مرحله انتخاب می‌شوند. تفاوت اصلی مدل پیشنهادی با مدل HMAX در این مرحله است، در مدل HMAX تکه‌ها به‌صورت کاملاً تصادفی انتخاب می‌شوند. این نوع انتخاب مشکلات زیادی برای سیستم ایجاد می‌کند که قبلاً به آن اشاره



شکل ۴: استفاده از الگوریتم حذف پس‌زمینه در تصویر هواپیما از پایگاه داده Caltech5

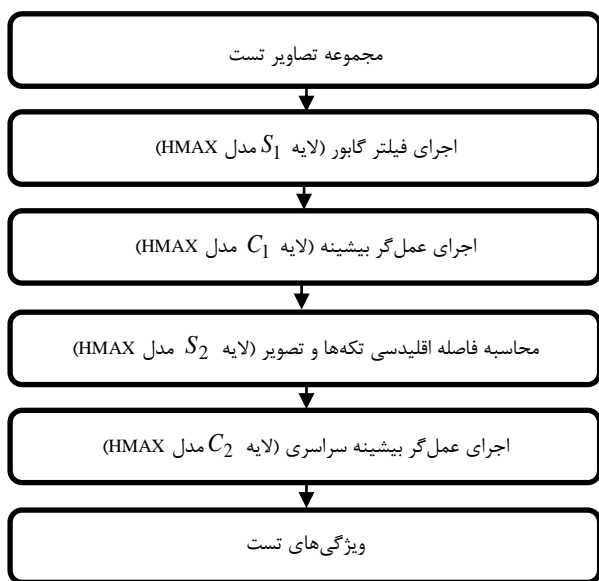
پس از نهایی شدن انتخاب، تکه‌هایی که با تکه‌های هم‌گروه شباهت بالایی دارند را انتخاب کرده و متمم آن $(1-P)$ را به گروه‌های دیگر اضافه می‌کنیم. با اضافه شدن متمم به گروه‌های دیگر احتمال تشخیص تصویر مربوط به این گروه به گروه‌های دیگر به حداقل می‌رسد. چون تصویر مربوط به این گروه با تکه متمم حداقل شباهت را دارد. خلاصه مراحل انتخاب N تکه بهینه در الگوریتم ۱ آورده شده است.

الگوریتم ۱: انتخاب N تکه بهینه از بین تکه‌های استخراج شده تصادفی

- (۱) ابتدا میانگین و واریانس تکه‌ها را به دست می‌آوریم.
- (۲) برای هر کدام از این کمیت‌ها مقدار آستانه در نظر می‌گیریم.
- (۳) تکه‌های کمتر از مقدار آستانه را به عنوان تکه‌هایی که اطلاعات مطلوبی ندارند حذف می‌کنیم.
- (۴) مقدار PS_k را طبق فرمول گفته شده محاسبه می‌کنیم.
- (۵) با استفاده از چرخ رولت با حفظ رویکرد تصادفی و متناسب با PS_k ، استخری از تکه‌ها به تعداد MN تکه حاوی بیشترین اطلاعات برای هر باند، اندازه و جهت ایجاد می‌کنیم.
- (۶) با استفاده از فرمول شباهت (SI)، دسته‌های M تایی با حداکثر شباهت را برحسب میزان شباهت به ترتیب می‌نویسیم.
- (۷) از هر دسته فقط یک تکه که بیشترین شباهت با بقیه دارد را به عنوان انتخاب نهایی در نظر می‌گیریم.
- (۸) با تعریف یک آستانه برای شباهت، متمم تکه‌های با شباهت بیشتر را به گروه‌های دیگر اضافه می‌کنیم.

۴-۴ مرحله آزمون

معماری مرحله آزمون مدل پیشنهادی در شکل ۵ نشان داده شده است. همان‌گونه که در شکل مشخص است همه قسمت‌های این مرحله مشابه مدل HMAX است. در حقیقت در مدل پیشنهادی فقط مرحله آموزش مدل HMAX که شامل استخراج تکه است تغییر کرده و این



شکل ۵: مرحله آزمون مدل پیشنهادی

استفاده می‌شود. درصد اطلاعات موجود در تکه شماره k در مقایسه با اطلاعات موجود در کل تکه‌ها در رابطه (۵) بیان شده است:

$$PS_k = \frac{1}{2} \times \left(\frac{\sigma_{p_k}}{\sum_{i=1}^{Num} \sigma_{p_i}} + \frac{\bar{p}_k}{\sum_{j=1}^{Num} \bar{p}_j} \right) \quad (5)$$

که در آن Num تعداد کل تکه‌های انتخاب شده در گام نخست در اندازه، جهت و باند مورد نظر، \bar{p}_k میانگین پیکسل‌های تکه p_k و σ_{p_k} واریانس پیکسل‌های تکه p_k است. اگر تعداد تکه‌های مورد نیاز برای هر باند، اندازه و جهت در نهایت برابر N تکه باشد، در این مرحله $M \times N$ تکه انتخاب می‌شود و استخری از تکه‌ها ایجاد می‌شود. مرحله دوم استخراج تکه‌های مفیدتر و بهینه از استخر تکه‌هاست. برای استخراج تکه‌های شامل اطلاعات مفیدتر، تکه‌هایی که در گروه خود دارای تشابه بیشتر هستند را جستجو می‌کنیم. برای این کار تکه‌ها به گروه‌های M تایی تقسیم می‌شوند. از آنجایی که در این مرحله تعداد کل تکه‌های باقیمانده در اندازه، باند و جهت مربوطه برابر MN تکه است، تعداد کل حالت‌ها برابر $\binom{MN}{M}$ خواهد بود (انتخاب M تکه از MN تکه). اگر واریانس تکه‌ای صفر باشد و با داشتن میانگین بالا از گام نخست عبور کرده باشد در اینجا حذف می‌شود (واریانس صفر یعنی همه پیکسل‌ها هم‌اندازه هستند و تکه اطلاعات مفیدی ندارد). برای محاسبه میزان شباهت دو تکه هم‌اندازه p و q ، می‌توانیم دو تکه را دو بردار عمود بر هم فرض کنیم که با ضرب داخلی بردارهای نظیر به نظیر، میزان شباهت آن‌ها محاسبه می‌شود. لذا از ضرب داخلی نرمالیزه شده دو تکه استفاده می‌کنیم [۲۰] که در رابطه (۶) بیان شده است:

$$SI = \frac{\sum_{k=1}^K \sum_{l=1}^L (p_{k,l} - \bar{p})(q_{k,l} - \bar{q})}{\sqrt{\sum_{k=1}^K \sum_{l=1}^L (p_{k,l} - \bar{p})^2} \times \sqrt{\sum_{k=1}^K \sum_{l=1}^L (q_{k,l} - \bar{q})^2}} \quad (6)$$

که در آن K و L ابعاد تکه‌ها می‌باشند، $p_{k,l}$ و $q_{k,l}$ به ترتیب معرف مقدار پیکسل در ردیف k و ستون l در تکه‌های p و q می‌باشند. در این مرحله تکه‌ها به دسته‌های M تایی که بیشترین شباهت به هم دارند تقسیم‌بندی می‌شوند (یک تکه می‌تواند در چند گروه قرار داشته باشد). دسته‌ها به ترتیب برحسب شباهت از بیشترین به کمترین مرتب می‌شوند. از هر دسته که شامل M تکه است باید یک تکه انتخاب شود چون اطلاعات آن‌ها تقریباً مشابه است و فقط بار محاسباتی سیستم را افزایش می‌دهد؛ بنابراین از بین M تکه مشابه اول یک تکه، از M تکه مشابه دوم یک تکه تا نهایتاً از M تکه مشابه N ام یک تکه انتخاب می‌شود (از آنجایی که تکه‌های انتهایی شامل اطلاعات مفید کمتری می‌باشند، این روند می‌تواند به دلخواه و تا کمتر از N ادامه یابد).

انتخاب شده‌اند. دو تصویر نمونه از هر گروه نام‌برده در شکل ۸ نمایش داده شده است.

مرحله (مرحله آزمون) تغییری نکرده است.

۵- پیاده‌سازی و ارزیابی نتایج تجربی

برای شبیه‌سازی، تصاویر به دو مجموعه آموزش و آزمون تقسیم می‌شوند، تعداد تصاویر آموزش برابر ۵، ۱۰، ۲۰، ۳۰ و ۴۰ و تعداد تصاویر آزمون به ترتیب برابر ۳، ۵، ۱۰، ۱۵ و ۲۰ با انتخاب کاملاً تصادفی در نظر گرفته می‌شوند. همه تصاویر به تصاویر خاکستری با اندازه 140×140 تغییر داده می‌شوند.

تکه‌های بهینه از تصاویر آموزشی و پس از عبور از لایه‌های D_1 و C_1 استخراج می‌شوند. منظور از تکه‌های بهینه تکه‌هایی با حداکثر اطلاعات از یک طرف و حداکثر تشابه با تکه‌های هم‌گروه از طرف دیگر می‌باشند. با عبور لایه C_1 و تکه‌ها از دو لایه D_2 و C_2 ویژگی‌های آموزش برای طبقه‌بند تولید می‌شوند. در مرحله آزمون نیز همین فرآیند به‌طور مشابه تکرار می‌شود ولی از تکه‌های ذخیره‌شده مرحله آموزش استفاده می‌شود. در نهایت ویژگی‌های به‌دست‌آمده به طبقه‌بند داده می‌شود تا گروه مربوط به تصویر به‌دست آید. تصاویر مورد آزمایش از پایگاه‌های داده استاندارد [۲۹] Caltech5، [۳۰] Caltech101 و [۳۱] GRAZ-01 می‌باشند.

۵-۴ پایگاه‌های داده

پایگاه‌های داده‌ای که در این مقاله برای ارزیابی روش پیشنهادی استفاده شده است پایگاه‌های داده Caltech5، Caltech101 و GRAZ-01 می‌باشند. از هر پایگاه داده دو گروه از تصاویر به‌عنوان تصاویر مثبت و گروه تصاویر پس‌زمینه به‌عنوان تصاویر منفی استفاده شده‌اند.

Caltech5: این پایگاه داده شامل ۳۱۲۲ تصویر طبیعی در پنج گروه مختلف و یک گروه تصاویر پس‌زمینه می‌باشد: چهره از روبرو، موتور، خودرو از پشت، هواپیما، برگ و پس‌زمینه. اندازه هر تصویر حدود 300×200 می‌باشد. تعداد تصاویر در هر گروه از ۱۲۷ تصویر در اتومبیل تا ۱۰۷۶ تصویر در هواپیما متفاوت است. از این پایگاه داده گروه‌های چهره و موتور به‌عنوان گروه تصاویر مثبت و گروه پس‌زمینه به‌عنوان گروه تصاویر منفی انتخاب شده‌اند. دو تصویر نمونه از هر گروه نام‌برده در شکل ۶ نشان داده شده است.

Caltech101: این پایگاه داده شامل ۹۱۴۴ تصویر طبیعی در ۱۰۱ گروه و یک گروه تصاویر پس‌زمینه می‌باشد که تصاویر هر گروه از ۴۰ تا ۸۰۰ تغییر می‌کند. از این پایگاه داده گروه‌های هواپیما و خودرو به‌عنوان تصاویر مثبت و گروه پس‌زمینه به‌عنوان تصاویر منفی انتخاب شده‌اند. دو تصویر نمونه از هر گروه نام‌برده در شکل ۷ نمایش داده شده است.

GRAZ-01: این پایگاه داده شامل ۱۱۰۳ تصویر طبیعی می‌باشد. از این پایگاه داده گروه‌های دوچرخه با ۳۷۳ تصویر و اشخاص با ۴۶۰ تصویر به‌عنوان تصاویر مثبت و بقیه تصاویر به‌عنوان تصاویر منفی

۵-۴ نتایج تجربی

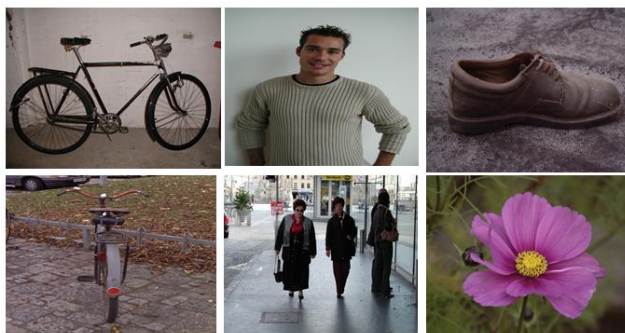
در این تحقیق، مدل HMAX به‌عنوان مدلی موفق برای بازشناسی شیء با الهام از سیستم بینایی انسان بهبودیافته و نرخ بازشناسی آن ارتقاء داده شده است. برای ارزیابی و مقایسه، نرخ بازشناسی مدل پیشنهادی (PSHMAX) با مدل‌های SIFT (۱۹۹۹)، HMAX (۲۰۰۷)، GMAX (۲۰۱۲) و DOHMAX (۲۰۱۶) مقایسه شده است. نتایج بعد از پنج مرتبه تکرار و میانگین‌گیری آورده شده است. پیاده‌سازی و شبیه‌سازی با استفاده از نرم‌افزار متلب انجام شده است. تصاویر نهایی به چهار دسته مثبت صحیح، مثبت کاذب، منفی صحیح و منفی کاذب تقسیم‌بندی می‌شوند. معیار کمی استفاده شده برای مقایسه نرخ بازشناسی، دقت و بازخوانی می‌باشند. نرخ بازشناسی برابر با نسبت مجموع منفی صحیح و مثبت صحیح به کل تصاویر، دقت برابر با نسبت مثبت صحیح به مجموع مثبت صحیح و مثبت کاذب و نهایتاً



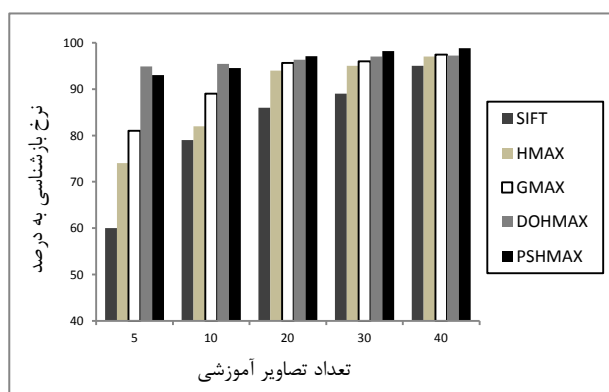
شکل ۶: موتور، چهره و پس‌زمینه از پایگاه داده Caltech5



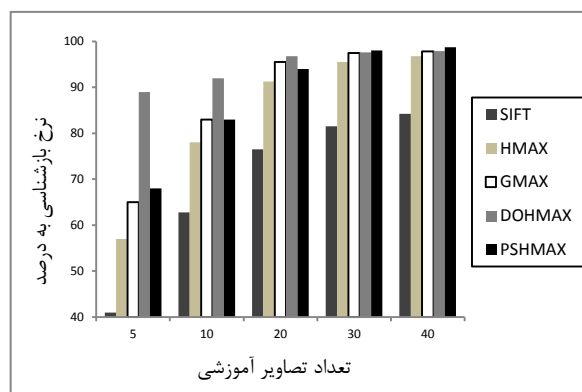
شکل ۷: هواپیما، خودرو و پس‌زمینه از پایگاه داده Caltech101



شکل ۸: دوچرخه، اشخاص و پس‌زمینه از پایگاه داده GRAZ-01



(ب)



(الف)

شکل ۹: مقایسه درصد نرخ بازشناسی روش پیشنهادی با روش‌های دیگر، (الف) تصاویر چهره و (ب) تصاویر موتور از پایگاه داده Caltech5

نتایج نرخ بازشناسی برای گروه‌های دوچرخه و اشخاص از پایگاه داده GRAZ-01 در شکل ۱۱ نشان داده شده است. گروه‌های دوچرخه و اشخاص به‌عنوان تصاویر مثبت و تصاویر پس‌زمینه به‌عنوان تصاویر منفی در نظر گرفته می‌شوند. همان‌طور که در شکل ۱۱-الف دیده می‌شود برای گروه دوچرخه روش پیشنهادی برتری مطلق نسبت به روش‌های دیگر دارد. در شکل ۱۱-ب که مربوط به گروه اشخاص است نیز همین روند حاکم است فقط در تصاویر آموزشی کم نرخ بازشناسی آن کمتر از GMAX است که آن هم در تصاویر آموزشی بالاتر پیشی می‌گیرد.

در جدول ۳ نتایج نرخ بازشناسی برای ۴۰ تصویر آموزشی و ۲۰ تصویر آزمون از هر گروه آموزشی برای کل گروه‌های پایگاه داده Caltech5 آورده شده است. با توجه به بالا بودن تعداد تصاویر آموزشی، طبق انتظار، نتایج حاصل نشان از برتری روش پیشنهادی نسبت به روش‌های دیگر دارد. تنها در گروه خودرو نرخ بازشناسی روش پیشنهادی بیشترین نیست، در این گروه نیز فقط از روش DOHMAX کمتر است و برتری آن نسبت به سایر روش‌ها مشهود است.

جدول ۳: مقایسه درصد نرخ بازشناسی روش پیشنهادی با روش‌های دیگر برای تعداد تصاویر آموزشی ۴۰

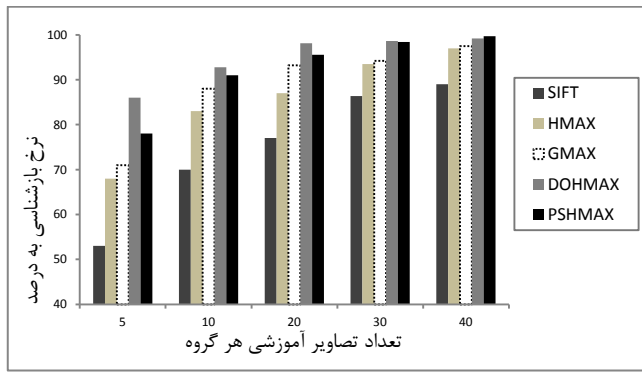
موتور	خودرو، پشت	برگ	هوایما	چهره، روبرو
۹۵	۸۹	۸۵	۸۳	۸۴/۲
۹۷	۹۷	۹۵	۹۳/۵	۹۶/۸
۹۷/۴	۹۷/۵	۹۵/۵	۹۵/۳	۹۷/۸
۹۷/۲	۹۹/۸	۹۶/۱	۹۷/۱	۹۷/۹
۹۸/۸	۹۹/۵	۹۷/۶	۹۸/۸	۹۸/۷

در دو گروه موتور و هوایما، به‌دلیل پیچیدگی تصاویر، یافتن تکه‌ها در روش پیشنهادی به سمت تکه‌های بهینه میل می‌کند و تکه‌هایی انتخاب می‌شوند که علاوه بر پیچیدگی زیاد، شباهت درون‌گروهی نیز دارند. لذا در مرحله آزمون احتمال تشخیص اشتباه بسیار کم خواهد بود. گروه دیگری که در آن روش پیشنهادی اختلاف زیادی با روش‌های دیگر دارد گروه تصاویر برگ است. تصاویر برگ شاید دارای پیچیدگی زیادی نباشند ولی به‌علت داشتن لبه زیاد و شکل خاص، یافتن تکه‌های بهینه در این گروه را ممکن کرده است. در

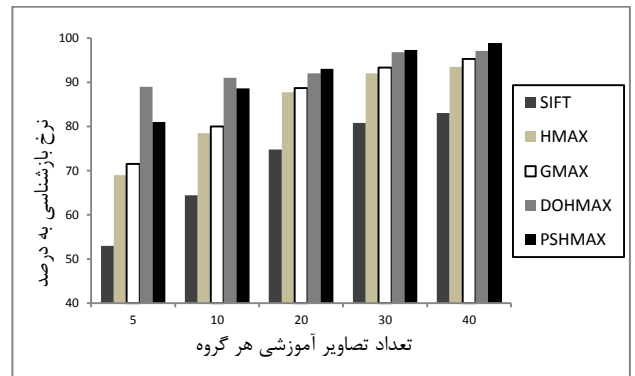
بازخوانی برابر با نسبت مثبت صحیح به مجموع مثبت صحیح و منفی کاذب است. در نهایت سرعت شبیه‌سازی نیز مقایسه می‌شود.

در شکل ۹، نتایج نرخ بازشناسی برای گروه‌های چهره و موتور از پایگاه داده Caltech5 نشان داده شده است. تصاویر چهره و موتور به‌عنوان تصاویر مثبت و تصاویر پس‌زمینه به‌عنوان تصاویر منفی در نظر گرفته شده‌اند. همان‌طور که در شکل ۹-الف دیده می‌شود برای گروه چهره روش پیشنهادی برتری مطلق نسبت به روش‌های دیگر دارد و فقط در مقایسه با روش DOHMAX آن هم در تعداد تصاویر آموزشی کم، نرخ بازشناسی آن کمتر است که در تصاویر آموزشی بالاتر نرخ بازشناسی آن پیشی می‌گیرد. در شکل ۹-ب که مربوط به گروه موتور است نیز همین روند حاکم است. چون در تعداد تصاویر آموزشی بالا، یافتن تکه‌های مشابه هم‌گروه ولی با اطلاعات بالا به واقعیت نزدیک‌تر است. به‌بیان‌دیگر اگر تعداد تصاویر آموزشی کم باشد ممکن است تکه‌های نامطلوبی که انتخاب می‌شوند، تصادفی با تکه‌های هم‌گروه مشابه باشند. لذا مدل پیشنهادی آن‌ها را به‌عنوان تکه‌های مطلوب می‌گیرد، بنابراین در این حالت مدل پیشنهادی کارایی چندانی ندارد. با بالا رفتن تعداد تصاویر آموزشی، احتمال تشابه تصادفی تکه‌های نامطلوب به حداقل می‌رسد و فقط تکه‌هایی که مربوط به شیء و از قسمت‌های مشترک آن‌ها هستند انتخاب می‌شوند، بنابراین تکه‌ها در تصاویر آموزشی بالا به سمت بهترین تکه‌ها میل می‌کنند.

نتایج نرخ بازشناسی برای گروه‌های هوایما و خودرو از پایگاه داده Caltech101 در شکل ۱۰ نشان داده شده است. گروه تصاویر هوایما و خودرو به‌عنوان تصاویر مثبت و تصاویر پس‌زمینه به‌عنوان تصاویر منفی در نظر گرفته می‌شوند. همان‌طور که در شکل ۱۰-الف دیده می‌شود برای گروه هوایما روش پیشنهادی برتری مطلق نسبت به روش‌های دیگر دارد فقط در مقایسه با روش DOHMAX آن هم در تعداد تصاویر آموزشی کم، نرخ بازشناسی آن کمتر از DOHMAX است که در تصاویر آموزشی بالاتر نرخ بازشناسی روش پیشنهادی پیشی می‌گیرد. در شکل ۱۰-ب که مربوط به گروه خودرو است نیز همین روند حاکم است. با استدلالی مشابه قبل، با بالا رفتن تعداد تصاویر آموزشی روش پیشنهادی برتری مطلق نسبت به روش‌های دیگر دارد.

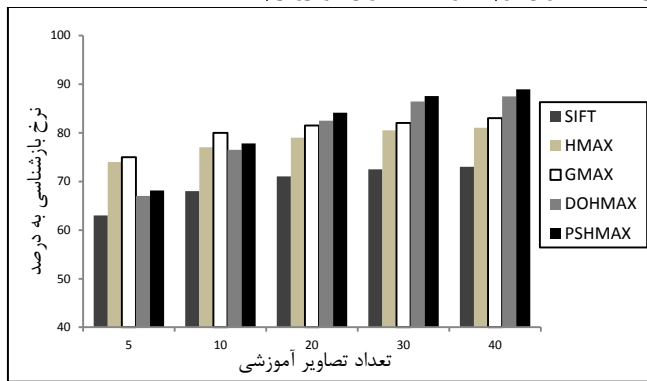


(ب)

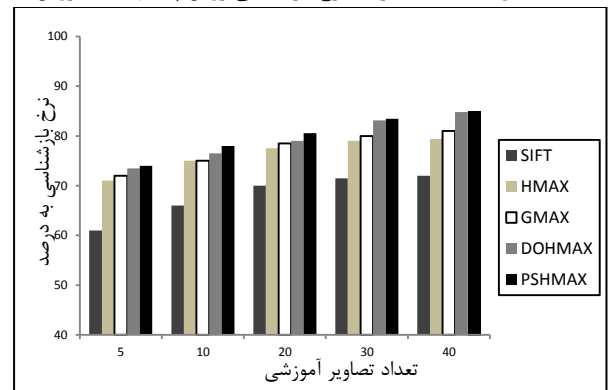


(الف)

شکل ۱۰: مقایسه درصد نرخ بازشناسی روش پیشنهادی با روش‌های دیگر: (الف) تصاویر هوایما و (ب) تصاویر خودرو از پایگاه داده Caltech101



(ب)



(الف)

شکل ۱۱: مقایسه درصد نرخ بازشناسی روش پیشنهادی با روش‌های دیگر: (الف) تصاویر دوچرخه و (ب) تصاویر اشخاص از پایگاه داده GRAZ-01

۶ - نتیجه‌گیری و کارهای آینده

در روش پیشنهادی دو دسته اصلی از تکه‌های غیرمفید حذف شدند، دسته اول تکه‌هایی با اطلاعات کم که فقط بار محاسباتی سیستم را افزایش می‌دهند و دسته دوم تکه‌هایی با اطلاعات غیرمفید از پس‌زمینه که باعث تولید خروجی اشتباه می‌شوند. لذا در مقایسه با روش‌های مشابه موجود هم سرعت و هم دقت افزایش یافته است. نتایج که حاصل ارزیابی سه پارامتر نرخ بازشناسی، دقت و بازخوانی روی سه پایگاه داده مختلف می‌باشد نشان‌دهنده برتری روش پیشنهادی است.

در این تحقیق روی انتخاب تکه بهینه به جای تکه تصادفی مطالعه شده و نتایج مطلوبی حاصل شده است ولی به نظر می‌رسد هنوز بهترین تکه‌ها انتخاب نشده‌اند و تحقیقات بعدی ممکن است منجر به

گروه تصاویر چهره نیز برتری مطلق از روش پیشنهادی است ولی اختلاف زیادی ایجاد نشده است.

در جدول ۴ نتایج مقایسه سرعت روش پیشنهادی با سایر روش‌ها آورده شده است. از سیستم با حافظه رم ۴ گیگابایت و پردازنده مرکزی Core i7 استفاده شده است. از آنجایی که در روش پیشنهادی تکه‌های حاوی اطلاعات کم حذف شده است، سرعت شبیه‌سازی پیشرفت خوبی داشته است. از طرفی با حذف تکه‌های حاوی اطلاعات پس‌زمینه، خروجی دچار اشتباه نمی‌شود بنابراین نرخ شبیه‌سازی بالا می‌رود. نتایج، مربوط به زمان متوسط لازم برای بازشناسی یک تصویر از پایگاه‌های داده مختلف می‌باشد که در جدول ۴ ذکر شده است.

در جدول ۵ نتایج پارامتر دقت برای تعداد متفاوتی از تصاویر آموزشی آورده شده است. در این جدول چهره از Caltech5، هوایما از Caltech101 و دوچرخه از GRAZ-01 آورده شده است. غیر از دوچرخه که روش پیشنهادی فقط از DOHMAX با اختلاف ناچیزی کمتر است، در بقیه موارد برتری مطلق با روش پیشنهادی است.

در جدول ۶ نتایج پارامتر بازخوانی برای تعداد متفاوتی از تصاویر آموزشی آورده شده است. در این جدول موتور از Caltech5، خودرو از Caltech101 و اشخاص از GRAZ-01 آورده شده است. در همه موارد، در تصاویر آموزشی بالا برتری مطلق از روش پیشنهادی است.

جدول ۴: مقایسه زمان بازشناسی روش پیشنهادی با روش‌های دیگر

	زمان متوسط بازشناسی هر تصویر برحسب ثانیه		
	دوچرخه و اشخاص از GRAZ-01	هوایما و خودرو از Caltech101	چهره و موتور از Caltech5
SIFT	۹/۴۳۲	۹/۷۲۱	۱۰/۰۴۵
HMAX	۶/۱۸۹	۶/۳۷۵	۶/۵۶۳
GMAX	۶/۳۴۱	۶/۵۷۸	۶/۹۰۱
DOHMAX	۶/۱۳	۶/۴۰۸	۶/۵۴۲
PSHMAX	۵/۱۴۸	۵/۳۰۲	۵/۴۶۲

جدول ۵: مقایسه پارامتر دقت روش پیشنهادی با روش‌های دیگر برای گروه‌های چهره، هواپیما و دوچرخه

	چهره از Caltech5					هواپیما از Caltech101					دوچرخه از GRAZ-01				
	تعداد تصاویر آموزشی					تعداد تصاویر آموزشی					تعداد تصاویر آموزشی				
	۵	۱۰	۲۰	۳۰	۴۰	۵	۱۰	۲۰	۳۰	۴۰	۵	۱۰	۲۰	۳۰	۴۰
SIFT	۴۱/۳	۶۱/۸	۷۴/۳	۷۸/۸	۸۱/۷	۵۲/۶	۶۳/۳	۷۱/۷	۷۹/۶	۸۱/۷	۶۰/۵	۶۵/۶	۶۸/۵	۷۱/۲	۷۱/۷
HMAX	۵۶/۳	۷۸/۳	۸۰/۹	۹۳/۱	۹۴/۹	۶۵/۸	۷۵/۳	۸۵/۴	۹۱/۱	۸۴/۳	۷۰/۶	۷۳/۵	۷۶/۹	۷۸/۴	۷۸/۸
GMAX	۵۸/۸	۸۰/۸	۹۳/۳	۹۵/۹	۹۵/۹	۶۶/۴	۷۵/۸	۸۶/۳	۹۱/۹	۹۳/۹	۷۰/۷	۷۵	۷۷/۹	۷۹/۴	۸۰/۶
DOHMAX	۸۵/۵	۸۱/۱	۹۵	۹۶	۹۶/۳	۸۷/۶	۸۹/۷	۹۰/۳	۹۵/۶	۹۶/۳	۷۱/۹	۷۵/۷	۷۹	۸۲/۸	۸۴/۷
PSHMAX	۶۵/۵	۸۱/۸	۹۳/۱	۹۷/۸	۹۸/۵	۷۸/۵	۸۶/۲	۹۱/۳	۹۶/۵	۹۸/۴	۷۱/۸	۷۶/۳	۹۷/۸	۸۲/۴	۸۴

جدول ۶: مقایسه پارامتر بازخوانی روش پیشنهادی با روش‌های دیگر برای گروه‌های چهره، هواپیما و دوچرخه

	موتور از Caltech5					خودرو از پهلو از Caltech101					اشخاص از GRAZ-01				
	تعداد تصاویر آموزشی					تعداد تصاویر آموزشی					تعداد تصاویر آموزشی				
	۵	۱۰	۲۰	۳۰	۴۰	۵	۱۰	۲۰	۳۰	۴۰	۵	۱۰	۲۰	۳۰	۴۰
SIFT	۶۲	۸۲/۵	۹۰	۹۳	۹۶/۵	۶۵	۷۶/۸	۸۱/۳	۸۹/۲	۹۱/۵	۶۴	۷۱	۷۲	۷۳	۷۲/۵
HMAX	۷۸	۸۶/۳	۹۸/۵	۹۸/۸	۹۹/۱	۸۰/۵	۸۶/۳	۹۲/۱	۹۵/۱	۹۷/۸	۷۶	۷۸	۸۱	۸۰/۸	۸۱/۴
GMAX	۹۲/۴	۹۲/۳	۹۸/۱	۹۸/۷	۹۸/۹	۸۰/۵	۹۱/۴	۹۳/۸	۹۵/۶	۹۸/۱	۷۶	۸۰/۷	۸۲	۸۲/۷	۸۳/۵
DOHMAX	۹۷/۲	۹۸/۳	۹۸/۳	۹۸/۵	۹۹	۸۹/۳	۹۴/۶	۹۸/۵	۹۹/۱	۹۹/۴	۷۰	۷۸	۸۳	۸۷/۸	۸۹
PSHMAX	۹۷	۹۸/۱	۹۸/۸	۹۹/۲	۹۹/۸	۸۶/۳	۹۳/۸	۹۷/۱	۹۸/۹	۹۹/۸	۷۱	۷۹/۶	۸۵	۸۸/۳	۹۵/۱

گرفته شد. به نظر می‌رسد وزن میانگین و واریانس در سنجش میزان اطلاعات تکه متفاوت است، لذا می‌توان برای آن‌ها وزن تعریف کرد و با توجه به وزن آن‌ها نتایج شبیه‌سازی را ارزیابی کرد. فرمول لایه ۲ در مدل HMAX خود معرف میزان شباهت تکه با لایه C_1 است لذا برای محاسبه میزان شباهت می‌توان از همان فرمول لایه ۲ در صورت مستقیم و یا روابط دیگر شباهت تصویر استفاده کرد.

انتخاب تکه‌های بهتر شود. در این تحقیق فقط تکه‌هایی که در گروه خود دارای بیشترین شباهت باشند به‌عنوان بهینه در نظر گرفته شدند و به ارتباط تکه با گروه‌های دیگر پرداخته نشده است. ارتباط یک تکه با گروه‌های دیگر و تأثیر آن روی انتخاب یا عدم انتخاب تکه می‌تواند مورد مطالعه قرار گیرد. در این تحقیق، میانگین و واریانس به‌عنوان معیارهای سنجش میزان اطلاعات موجود در تکه در نظر گرفته شدند که در رابطه (۵) مقدار متوسط آن‌ها به‌عنوان سنجش نهایی در نظر

مراجع

- [6] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," *Nature neuroscience*, vol. 2, pp. 1019-1025, 1999.
- [7] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust object recognition with cortex-like mechanisms," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, 2007.
- [8] T. Serre, L. Wolf, and T. Poggio, "Object recognition with features inspired by visual cortex," *Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 994-1000, 2005.
- [9] D. G. Lowe, "Object recognition from local scale-invariant features," *The proceedings of the seventh IEEE international conference on Computer vision*, pp. 1150-1157, 1999.
- [10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, pp. 91-110, 2004.
- [11] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, pp. 23-27, 1975.
- [12] Y. LeCun, F. J. Huang, and L. Bottou, "Learning methods for generic object recognition with invariance to pose and
- [1] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. II-II, 2003.
- [2] R. Peng and P. K. Varshney, "A human visual system-driven image segmentation algorithm," *Journal of Visual Communication and Image Representation*, vol. 26, pp. 66-79, 2015.
- [3] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurones in the cat's striate cortex," *The Journal of physiology*, vol. 148, p. 574, 1959.
- [4] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *The Journal of physiology*, vol. 195, pp. 215-243, 1968.
- [5] M. Mishkin and L. G. Ungerleider, "Contribution of striate inputs to the visuospatial functions of parieto-occipital cortex in monkeys," *Behavioural brain research*, vol. 6, pp. 57-77, 1982.

- recognition," *Signal Processing and Signal Processing Education Workshop (SP/SPE), 2015 IEEE*, 2015, pp. 403-408.
- [22] Y. Zhang, L. Zhang, and P. Li, "A novel biologically inspired ELM-based network for image recognition," *Neurocomputing*, vol. 174, pp. 286-298, 2016.
- [23] Y.-F. Lu, H.-Z. Zhang, T.-K. Kang, and M.-T. Lim, "Dominant orientation patch matching for HMAX," *Neurocomputing*, vol. 193, pp. 155-166, 2016.
- [24] J. Mutch and D. G. Lowe, "Multiclass object recognition with sparse, localized features," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 11-18, 2006.
- [25] M. Ghodrati, S.-M. Khaligh-Razavi, R. Ebrahimpour, K. Rajaei, and M. Pooyan, "How can selection of biologically inspired features improve the performance of a robust object recognition model?," *PloS one*, vol. 7, p. e32357, 2012.
- [26] E. Krupka, A. Navot, and N. Tishby, "Learning to select features using their properties," *Journal of Machine Learning Research*, vol. 9, pp. 2349-2376, 2008.
- [27] K. Huang, D. Tao, Y. Yuan, X. Li, and T. Tan, "Biologically inspired features for scene classification in video surveillance," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, pp. 307-313, 2011.
- [۲۸] وحید آزادزاده، علی محمد لطیف، «دسته بندی ویژگی‌های استخراج شده از پیش زمینه و پس زمینه تصویر برای ردیابی اهداف متحرک»، *مجله مهندسی برق دانشگاه تبریز*، جلد ۴۶، شماره ۳، پاییز ۱۳۹۵.
- [29] <http://vision.caltech.edu/archive.html>.
- [30] http://www.vision.caltech.edu/Image_Datasets/Caltech101/.
- [31] http://www.emt.tugraz.at/~pinz/data/GRAZ_01/.
- lighting," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 83-104, 2004.
- [13] S. M. Stringer, G. Perry, E. T. Rolls, and J. Proske, "Learning invariant object recognition in the visual system with continuous transformations," *Biological cybernetics*, vol. 94, pp. 128-142, 2006.
- [۱۴] سلیمه بامری، سعید سریزدی و حسین نظام آبادی پور، «فیلترهای چندجمله‌ای مدوله شده با دوره محدود و کاربرد آن‌ها در طبقه بندی تصویر»، *مجله مهندسی برق دانشگاه تبریز*، جلد ۴۰، شماره ۱.
- [15] A. W. Bitar, M. M. Mansour, and A. Chehab, "Efficient Implementation of a Recognition System using the Cortex Ventral Stream Model," *VISAPP*, pp. 138-147, 2015.
- [16] Y. Li, W. Wu, B. Zhang, and F. Li, "Enhanced HMAX model with feedforward feature learning for multiclass categorization," *Frontiers in computational neuroscience*, vol. 9, 2015.
- [17] H. Qiao, X. Xi, Y. Li, W. Wu, and F. Li, "Biologically inspired visual model with preliminary cognition and active attention adjustment," *IEEE transactions on cybernetics*, vol. 45, pp. 2612-2624, 2015.
- [18] Y. Huang, K. Huang, D. Tao, T. Tan, and X. Li, "Enhanced biologically inspired model for object recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, pp. 1668-1680, 2011.
- [19] C. Theriault, N. Thome, and M. Cord, "Extended coding and pooling in the HMAX model," *IEEE Transactions on Image Processing*, vol. 22, pp. 764-777, 2013.
- [20] C. Theriault, N. Thome, and M. Cord, "HMAX-S: deep scale representation for biologically inspired image categorization," *18th IEEE International Conference on Image Processing (ICIP)*, pp. 1261-1264, 2011.
- [21] O.-R. H. Eustacio, C.-M. M. Ignacio, and R.-Q. J. Alberto, "A neural bio-inspired scheme for head pose

زیرنویس‌ها

- ¹ Probabilistic Selection HMAX
- ² Hubel & Wiesel
- ³ A Family of Local and Scaled Gaussian Operators
- ⁴ Riesenhuber
- ⁵ Derivative of Gaussian
- ⁶ Serre
- ⁷ Scale Invariant Feature Transform
- ⁸ Radial Basis Filter
- ⁹ Sharpness of the Tuning
- ¹⁰ Li
- ¹¹ Qiao
- ¹² Huang
- ¹³ Theriault
- ¹⁴ Eustacio
- ¹⁵ Zhang
- ¹⁶ Yan-Feng
- ¹⁷ Dominant Orientation Patch matching for HMAX
- ¹⁸ Much and Lowe
- ¹⁹ Krupka