بازشناسی اشیاء

علوم اعصاب بینایی، مدل های محاسباتی– شناختی، یادگیری عمیق

تألیف و تصنیف:

دکتر رضا ابراهیم‌پور

عضو هیأت علمی دانشگاه تربیت دبیر شهید رجائی

سرشناسه	:	ابراهیم‌پور، رضا، ۱۳۵۴ -
عنوان و نام پدیدآور	:	بازشناسی اشیاء: علوم اعصاب بینایی، مدل‌های محاسباتی- شناختی، یادگیری عمیق/ تالیف و تصنیف رضا ابراهیم‌پور.
مشخصات نشر	:	تهران: دانشگاه تربیت دبیر شهید رجائی، ۱۳۹۵.
مشخصات ظاهری	:	د، ۳۹۰ ص.: مصور، جدول، نمودار.
شابک	:	۹۷۸-۶۰۰-۶۵۹۴-۸۱-۱
وضعیت فهرست نویسی	:	فیپا
یادداشت	:	واژه‌نامه.
عنوان دیگر	:	علوم اعصاب بینایی، مدل‌های محاسباتی- شناختی، یادگیری عمیق.
موضوع	:	عصب پایه‌شناسی شناختی
موضوع	:	Cognitive neuroscience
موضوع	:	الگوهای عصب شناختی
موضوع	:	* Models, Neurological
موضوع	:	شبکه‌های عصبی (کامپیوتر)
موضوع	:	Neural networks (Computer science)
موضوع	:	عصب پایه‌شناسی کامپیوتری
موضوع	:	Computational neuroscience
موضوع	:	الگوشناسی -- سیستم‌ها
موضوع	:	Pattern recognition systems
موضوع	:	بینایی ماشین
موضوع	:	Computer vision
موضوع	:	ادراک بصری
موضوع	:	Visual perception
شناسه افزوده	:	دانشگاه تربیت دبیر شهید رجایی
شناسه افزوده	:	Shahid Rajaee Teacher Training University
رده بندی کنگره	:	QP۳۶۰/۵
رده بندی دیویی	:	۶۱۲/۸۲
شماره کتابشناسی ملی	:	۴۵۵۲۵۲۷

دانشگاه تربیت دبیر شهید رجائی

عنوان	:	بازشناسی اشیاء علوم اعصاب بینایی، مدل‌های محاسباتی- شناختی، یادگیری عمیق
تألیف و تصنیف	:	دکتر رضا ابراهیم‌پور، عضو هیأت علمی دانشگاه تربیت دبیر شهید رجائی
ویراستار علمی	:	دکتر رضا ابراهیم‌پور
ویراستار ادبی	:	دکتر یدالله بهمنی
نوبت چاپ	:	اول- بهار ۱۳۹۶
انتشارات	:	دانشگاه تربیت دبیر شهید رجائی
لیتوگرافی	:	رجاء نقشینه
چاپ	:	شریف
طراح جلد	:	محمد مشتاقی
ناظر چاپ	:	محمد معتمدی‌نژاد
کارشناس چاپ و صفحه‌آرا	:	نیره فیروزی
کارشناسان	:	طاهره کیاء/ علی رضایی اهوانوئی
شمارگان	:	۱۰۰۰ جلد
قیمت	:	۲۵۰,۰۰۰ ریال
شابک	:	۹۷۸-۶۰۰-۶۵۹۴-۸۱-۱ ISBN: 978-600-6594-81-1

کلیهٔ حقوق این اثر برای مؤلفین و مترجمین و دانشگاه تربیت دبیر شهیدرجائی محفوظ است.

نشانی: تهران، لویزان – کد پستی ۱۶۷۸۸-۱۵۸۱۱ – صندوق پستی ۱۶۳ – ۱۶۷۸۵ – تلفن: (۲۶۳۲) ۹ – ۲۲۹۷۰۰۶۰،
۲۲۹۷۰۰۷۰، تلفکس: ۲۲۹۷۰۰۴۲، پست الکترونیکی: Publish@srttu.edu، وب سایت: http://Publish.srttu.edu

فهرست مطالب

پیشگفتار	ذ
فصل ۱: مقدمه	**۱**
۱-۱- مقدمه	۱
۱-۲- بازشناسی اشیا	۳
۱-۳- سامانه بینایی انسان	۷
۱-۴- مدل‌های محاسباتی بینایی مبتنی‌بر سیستم بینایی انسان	۱۰
۱-۵- یادگیری عمیق	۱۲
۱-۶- ساختار کتاب	۱۵
۱-۷- جمع‌بندی	۱۵
۱-۸- فهرست مراجع	۱۷
فصل ۲: مروی بر ساختار مغز، چشم و سامانه‌ی بینایی انسان	**۱۹**
۲-۱- مقدمه	۱۹
۲-۲- مروری بر ساختار مغز انسان	۲۰
۲-۲-۱- ساختمان نورون	۲۰
۲-۲-۲- نحوه‌ی انتقال پیام در شبکه‌ی عصبی	۲۱
۲-۲-۲-۱- پتانسیل غشایی	۲۳
۲-۲-۲-۲- تکانه‌ها	۲۴
۲-۲-۲-۳- انتقالات سیناپسی	۲۵
۲-۲-۳- انواع نورون‌ها	۲۸

۲-۲-۳-۱- تقسیم‌بندی بر اساس تعداد آکسون‌ها و دندریت‌ها ۲۹
۲-۲-۳-۲- تقسیم‌بندی بر مبنای دندریت‌ها ۳۰
۲-۲-۳-۳- تقسیم‌بندی بر اساس اتصالات ۳۰
۲-۲-۳-۴- تقسیم‌بندی بر اساس طول آکسون ۳۱
۲-۲-۳-۵- تقسیم‌بندی بر اساس نوع انتقال‌دهندهی عصبی ۳۲
۲-۲-۴- بخش‌های مختلف مغز و سامانه عصبی بدن ۳۲
۲-۲-۵- تقسیم‌بندی مغز براساس نواحی برادمن ۳۵
۲-۳- مروری بر ساختار چشم و سامانه‌ی بینایی انسان ۳۶
۲-۳-۱- نور ۳۶
۲-۳-۲- چشم ۴۰
۲-۳-۲-۱- شبکیه و گیرنده‌های نوری ۴۳
۲-۳-۲-۲- انواع سلول‌های گانگلیون ۴۶
۲-۳-۳- مسیر بینایی مرکزی ۴۹
۲-۳-۳-۱- تالاموس ۵۰
۲-۳-۳-۲- هسته‌ی خمیده‌ی جانبی ۵۳
۲-۳-۳-۳- ناحیه‌ی V_1 ۵۵
۲-۳-۳-۴- ناحیه‌ی V_2 ۵۹
۲-۳-۳-۵- ناحیه‌ی V_4 ۵۹
۲-۳-۳-۶- ناحیه‌ی IT ۶۰
۲-۳-۳-۷- ناحیه‌ی قشر حرکتی ۶۰
۲-۳-۴- مسیر توجه بینایی ۶۳
۲-۴- جمع‌بندی ۶۵
۲-۵- فهرست مراجع ۶۷

فصل ۳: مدل‌های محاسباتی– شناختی بینایی ۶۹

۳-۱- مقدمه ۶۹
۳-۲- مدل Neocognitron ۷۰
۳-۲-۱- مقدمه ۷۰
۳-۲-۲- معرفی مدل Neocognitron ۷۱
۳-۲-۳- عملکرد لایه‌های Neocognitron ۷۴

۳-۲-۳-۱- لایه‌ی ورودی (U0) ۷۴
۳-۲-۳-۲- لایه‌ی شبکیه (UG) (یا لایه‌ی استخراج تباین) ۷۴
۳-۲-۳-۳- لایه‌ی سلول‌های ساده (US) ۷۵
۳-۲-۳-۴- لایه‌ی سلول‌های پیچیده (UC) ۷۹
۳-۲-۳-۵- لایه‌ی بازشناسی ۸۰
۳-۲-۴- مروری بر قواعد آموزش ۸۱
۳-۳- مدل HMAX ۸۳
۳-۳-۱- مقدمه ۸۳
۳-۳-۲- سلول‌ها یا واحدهای ساده و پیچیده ۸۴
۳-۳-۲-۱- واحدهای S_1 ۸۴
۳-۳-۲-۲- واحدهای C_1 ۸۶
۳-۳-۳- مرحله‌ی یادگیری ۸۸
۳-۳-۴- گام طبقه‌بندی ۸۹
۳-۳-۵- عملکرد مدل در مقایسه با انسان ۹۰
۳-۳-۵-۱- آزمایش اول ۹۰
۳-۳-۵-۲- آزمایش دوم ۹۲
۳-۳-۵-۳- آزمایش سوم ۹۶
۳-۴- مدل VisNet ۹۷
۳-۴-۱- مقدمه ۹۷
۳-۴-۲- معرفی مدل VisNet ۹۸
۳-۴-۳- معیارهای اندازه‌گیری کارایی مدل VisNet ۱۰۲
۳-۴-۴- قانون اثر ۱۰۵
۳-۴-۵- آزمایش‌های انجام شده بر روی VisNet با استفاده از قانون اثر ۱۰۶
۳-۴-۶- T، L و + به عنوان محرک: آموختن عدم تغییر نسبت به جابه‌جایی ۱۰۶
۳-۴-۷- اندازه‌گیری ظرفیت VisNet ۱۰۹
۳-۴-۷-۱- مجموعه داده استفاده شده ۱۱۰
۳-۴-۷-۲- مدل HMAX مورد استفاده برای مقایسه ۱۱۱
۳-۴-۷-۳- کارایی مدل بر روی مجموعه داده Caltech-256 ۱۱۱
۳-۴-۸- کارایی مدل بر روی مجموعه داده‌ی ALOI ۱۱۲
۳-۵- مدل HMAX مبتنی‌بر نورون ضربه‌ای ۱۱۴

۳-۵-۱- مقدمه ۱۱۴
۳-۵-۲- مدل‌های ضربه‌ای نورون ۱۱۵
۳-۵-۲-۱- مدل ادغام و آتش ۱۱۶
۳-۵-۲-۲- مدل هاجکین-هاکسلی ۱۱۷
۳-۵-۳- شبکه‌های نورونی و قاعده یادگیری ۱۱۹
۳-۵-۳-۱- قانون STDP ۱۲۰
۳-۵-۴- مدل سلسله‌مراتبی بازشناسی اشیا مبتنی بر ضربه ۱۲۱
۳-۵-۴-۱- ساختار لایه‌ها ۱۲۳
۳-۶- ارائه‌ی مدل محاسباتی مبتنی‌بر ویژگی‌های کارکردی هسته‌های تالامیک در فرآیند بازشناسی اشیا (LGN-model) ۱۳۵
۳-۶-۱- رمزگذاری بهینه ۱۳۶
۳-۶-۲- بازخوردها به LGN و اهمیت آن‌ها ۱۳۸
۳-۶-۳- ارزیابی مدل ۱۳۸
۳-۶-۴- سازوکار LGN-model و شواهد زیستی آن ۱۳۹
۳-۶-۵- تحلیل عملکرد و خروجی مدل ۱۴۱
۳-۶-۶- تابع خودهمبستگی و طیف توان ۱۴۲
۳-۶-۷- انتقال اطلاعات ۱۴۳
۳-۶-۸- تُنُکی ۱۴۴
۳-۶-۹- آزمایش بازشناسی اشیا ۱۴۵
۳-۶-۹-۱- آزمایش روان- فیزیک ۱۴۵
۳-۶-۹-۲- مدل محاسباتی بازشناسی اشیا ۱۴۶
۳-۶-۹-۳- محرک‌ها ۱۴۷
۳-۶-۹-۴- نتایج ۱۵۰
۳-۶-۱۰- ارزیابی نابسته سازی زمانی با استفاده از طیف توان و تابع خودهمبستگی ۱۵۰
۳-۶-۱۱- آزمایش بازشناسی اشیا ۱۵۴
۳-۷- ویژگی‌های بینایی سطح میانی برای بازشناسی اشیا ۱۵۶
۳-۷-۱- تعیین ویژگی‌های بینایی سطح میانی براساس الگوریتم ژنتیک ۱۵۷
۳-۷-۱-۱- ساختار کروموزوم ۱۵۸

۳-۷-۱-۲- تابع شایستگی ۱۵۹
۳-۷-۱-۳- معماری سامانه ۱۵۹
۳-۷-۱-۴- معرفی پایگاه داده ۱۶۰
۳-۷-۱-۵- طراحی آزمایش ۱۶۲
۳-۷-۱-۶- نتایج ۱۶۴
۳-۷-۲- تعیین ویژگی‌های بینایی سطح میانی بر اساس نظریه‌ی تشدید تطبیقی ۱۷۳
۳-۷-۲-۱- مجموعه‌ی داده تصاویر ۱۷۸
۳-۷-۲-۲- طبقه‌بندی به وسیله‌ی مدل Stable ۱۷۹
۳-۷-۲-۳- نتایج ۱۸۰
۳-۷-۲-۴- مقایسه با مدل محتمل زیستی دیگر ۱۸۰
۳-۷-۲-۵- مقایسه با انسان ۱۸۱
۳-۷-۲-۶- پایداری مدل Stable ۱۸۴
۳-۷-۲-۷- بررسی مدل Stable در بازشناسی چهره‌ی مستقل از زاویه دید ۱۸۸
۳-۸- ارائه مدل محاسباتی پردازش چهره مبتنی‌بر سیستم بینایی میمون با قابلیت شبیه‌سازی خروجی‌های رفتاری در انسان ۱۹۲
۳-۸-۱- بازشناسی چهره مقاوم در برابر تغییرات و اهداف مدلسازی مبتنی‌بر سامانه بینایی ۱۹۳
۳-۸-۲- Face-Net ۱۹۵
۳-۸-۲-۱- لایه S_1 ۱۹۵
۳-۸-۲-۲- لایه C_1 ۱۹۷
۳-۸-۲-۳- لایه S_2 ۱۹۷
۳-۸-۲-۴- لایه C_2 ۱۹۷
۳-۸-۲-۵- لایه VSL ۱۹۷
۳-۸-۲-۶- لایه ISL ۱۹۸
۳-۸-۳- ارزیابی Face-Net ۲۰۰
۳-۸-۳-۱- پایگاه داده ۲۰۰
۳-۸-۳-۲- ارزیابی توسط محاسبه‌ی ماتریس مشابهت ۲۰۱
۳-۸-۳-۳- ارزیابی مدل با شاخص‌های حساسیت به زاویه‌ی دید و هویت افراد ۲۰۲
۳-۸-۳-۴- اثر معکوس کردن تصاویر چهره ۲۰۴

۳-۸-۳-۵- اثر هم‌ترازی تصاویر چهره ۲۰۵
۳-۸-۳-۶- اثر نژادهای مختلف ۲۰۵
۳-۹- جمع‌بندی ۲۰۷
۳-۱۰- فهرست مراجع ۲۱۰

فصل ۴: شبکه‌های عصبی کانولوشنی ۲۱۹

۴-۱- مقدمه ۲۱۹
۴-۲- ساختار شبکه‌های عصبی کانولوشنی ۲۲۲
۴-۳- لایه‌های استفاده شده در شبکه‌های عصبی کانولوشنی ۲۲۳
۴-۳-۱- لایه‌ی کانولوشنی ۲۲۳
۴-۳-۱-۱- عملگر کانولوشن ۲۲۳
۴-۳-۱-۲- ویژگی‌های کانولوشن ۲۲۶
۴-۳-۲- لایه‌ی ادغام ۲۳۱
۴-۳-۳- لایه‌ی اتصال کامل ۲۳۴
۴-۳-۴- لایه‌ی یک‌سوساز خطی ۲۳۵
۴-۴- انتخاب ابرپارامترها ۲۳۵
۴-۵- روش‌های تنظیم ۲۳۶
۴-۵-۱- روش‌های تجربی ۲۳۶
۴-۵-۲- روش‌های صریح ۲۳۸
۴-۶- آموزش شبکه‌های کانولوشنی ۲۳۹
۴-۶-۱- آموزش با سرپرست ۲۴۰
۴-۶-۱-۱- محاسبه حساسیت خطا ۲۴۰
۴-۶-۱-۲- محاسبه گرادیان خطا ۲۴۲
۴-۶-۱-۳- قانون یادگیری دلتا ۲۴۴
۴-۶-۲- آموزش بدون سرپرست ۲۴۴
۴-۶-۲-۱- تجزیه‌ی اجزای مؤلفه اصلی (PCA) در آموزش ۲۴۵
۴-۶-۲-۲- یادگیری هِب ۲۴۵
۴-۶-۲-۳- خود رمزگذارها ۲۴۶
۴-۷- مصورسازی و معکوس کانولوشن ۲۴۹
۴-۸- اساس‌های زیستی در CNN ۲۵۲

۴-۸-۱- انطباق مدل‌های محاسباتی با مغز از نظر بازنمایی ۲۵۵
۴-۹- شبکه‌های ارائه شده و کاربردها ۲۵۸
۴-۹-۱- LeNet ۲۵۹
۴-۹-۲- AlexNet ۲۶۳
۴-۹-۳- VGGNet ۲۶۹
۴-۹-۴- GoogLeNet ۲۷۳
۴-۱۰- شبکه‌های عصبی بازخوردی ۲۷۵
۴-۱۰-۱- شبکه‌های عصبی کانولوشنی بازخوردی ۲۷۶
۴-۱۰-۱-۱- لایه‌ی کانولوشنی بازخوردی ۲۷۷
۴-۱۰-۲- استفاده از شبکه‌های عصبی بازگشتی برای فهم بهتر شبکه‌های عمیق ۲۷۸
۴-۱۰-۳- شبکه‌های عصبی حافظه بلند-کوتاه مدت (LSTMs) ۲۸۰
۴-۱۰-۴- کاربردهای شبکه‌های عصبی بازخوردی ۲۸۲
۴-۱۱- توانایی مسیر پیشرو قشر بینایی در مقابله با تغییرات در بازشناسی اشیا: آزمایش‌های انسانی و مدل محاسباتی سازگار با قشر بینایی ۲۸۵
۴-۱۱-۱- مواد و روش‌ها ۲۸۸
۴-۱۱-۱-۱- آزمایش‌های بازشناسی اشیا در انسان ۲۹۰
۴-۱۱-۱-۲- مدل محاسباتی قشر بینایی ۲۹۱
۴-۱۱-۱-۳- ارزیابی دقت بازشناسی در انسان و مدل محاسباتی و شاخص‌های خوشه‌بندی ۲۹۳
۴-۱۱-۲- نتایج ۲۹۶
۴-۱۱-۳- نتیجه‌گیری ۳۰۳
۴-۱۲- مدل‌سازی محاسباتی جداسازی شیء هدف از پس‌زمینه در بازشناسی اشیا با الهام از سیستم بینایی انسان ۳۰۶
۴-۱۲-۱- مواد و روش‌ها ۳۰۸
۴-۱۲-۱-۱- شبکه‌ی عصبی کانولوشنی ۳۰۸
۴-۱۲-۱-۲- مدل Recurrent ۳۱۰
۴-۱۲-۱-۳- ساختار مدل Recurrent ۳۱۲
۴-۱۲-۲- نتایج ۳۱۳
۴-۱۲-۲-۱- مجموعه‌ی داده ۳۱۴

۴-۱۲-۲-۲- مقایسه عملکرد شبکه‌های عصبی کانولوشنی و عملکرد رفتاری انسان ۳۱۵
۴-۱۲-۲-۳- ماتریس‌های عدم شباهت بازنمایی ۳۱۷
۴-۱۲-۲-۴- نتایج حاصل از مدل Recurrent ۳۲۲
۴-۱۲-۲-۵- مقایسه‌ی عملکرد مدل Recurrent با مدل‌های دیگر و انسان ۳۲۴
۴-۱۲-۲-۶- بصری‌سازی بازنمایی‌های لایه‌های کانولوشنی ۳۲۵
۴-۱۲-۳- نتیجه‌گیری ۳۲۷
۴-۱۳- جمع‌بندی ۳۲۷
۴-۱۴- فهرست مراجع ۳۲۹

واژه‌نامه‌ی فارسی به انگلیسی ۳۳۷

واژه‌نامه‌ی انگلیسی به فارسی ۳۵۳

پیشـــگفتار

در آزمایش‌هایی که هابل و ویزل در سال ۱۹۵۹ میلادی بر روی مغز گربه انجام دادند، دریافتند که در بخش‌های ابتدایی قشر بینایی، سلول‌هایی وجود دارند که به خطوط یا لبه‌های اشیا، در زوایا، اندازه‌ها و نواحی خاص پاسخ می‌دهند. همچنین، سلول‌های دیگری نیز یافته شدند که به زوایایی دیگر از محرک‌های مشابه پاسخ می‌دادند؛ آنها این سلول‌ها را سلول‌های ساده نامیدند. نیز، در لایه‌های بالاتر سلول‌هایی به دست آوردند که مانند سلول‌های ساده، به زاویه‌ای مشخص از یک خط، اما با حساسیت کمتری نسبت به موقعیت و اندازه‌ی آن پاسخ می‌دادند؛ این سلول‌ها نیز سلول‌های پیچیده نامیده شدند. آنها پیشنهاد دادند که توالی این سلول‌ها در لایه‌های مختلف می‌تواند منجر به کاهش حساسیت به موقعیت، اندازه و زاویه‌ی محرک شده و همچنین بازشناسی الگوهای پیچیده‌تر در لایه‌های بالاتر را میسر می‌نماید.

یافته‌ی اولیه‌ی هابل و ویزل، و ارائه‌ی مدلی مفهومی از قشر بینایی براساس سلول‌های ساده و پیچیده، قدم بزرگی در مدل‌سازی قشر بینایی بود. بسیاری از مدل‌ها که بعد از کار هابل و ویزل ارائه شدند، به نحوی از این یافته الهام گرفته و سعی در توسعه و تکمیل کار آنها داشتند. یکی از این مدل‌ها، مدل محاسباتی Neocognitron بود، که یک سامانه‌ی سلسله‌مراتبی متشکل از لایه-های متوالی سلول‌های ساده و پیچیده است. مدل VisNet نیز در سال ۱۹۹۷ میلادی، توسط والیز و رولز معرفی گردید. همان‌گونه که می‌دانیم، سامانه‌ی بینایی انسان، علی‌رغم تغییرات اشیا نظیر چرخش، موقعیت و اندازه، قادر به بازشناسی آنها است که این قابلیت، یکی از اساسی‌ترین اصول در ساختار VisNet می‌باشد. از دیگر مدل‌های مبتنی بر یافته‌های فیزیولوژی و آناتومی رشته‌ی قدامی قشر بینایی، مدل HMAX است. این مدل در سال ۱۹۹۹ میلادی توسط ریزنهوبر و پوجیو در دانشگاه تکنولوژی ماساچوست (MIT) معرفی گردید. عامل مهمی که موجب توجه فراوان به این مدل شد، استفاده از عملگر بیشینه‌گیر برای مدل‌سازی عملکرد ادغام در سلول‌های پیچیده در ناحیه‌ی V_1 (و لایه‌های بالاتر) بوده است. آنها نشان دادند که عملگر بیشینه‌گیر می‌تواند رفتار سلول‌های پیچیده را به خوبی توصیف کرده و با ثبت‌های نورونی انطباق بالایی دارد.

در سال‌های اخیر، دانش پزشکی و به صورت ویژه علوم اعصاب بینایی، به کمک پیشرفت‌های روزافزونی که ابزارهای مختلف تصویربرداری از مغز فراهم آورده‌اند، توانسته است بسیاری از

سؤالات مطرح و ابهامات زیستی در حوزه‌ی بازشناسی اشیا را پاسخ دهند. به علاوه، به صورت همزمان، مهندسان و متخصصان مدل‌سازی در این حوزه پیشرفت‌های شگرفی در استفاده و بکارگیری سامانه‌های پردازشی قدرتمند و موازی در این عرصه نموده‌اند. از این رو، نویسنده معتقد است در سالیان آینده، شاهد معرفی ماشین‌هایی خواهیم بود که با استفاده از مدل‌های محاسباتی‌ـشناختیِ مبتنی بر سازوکار بینایی مغز، نقش‌های بسیار مهمی را در زندگی روزمره‌ی بشر در حوزه‌های صنعتی، پزشکی، نظامی و غیره ایفا خواهند کرد.

نویسنده طی سالیان متمادی تحصیل، پژوهش و تدریس خود به مطالعه‌ی مدل‌های محاسباتی-ـشناختی بازشناسی اشیا، مبتنی بر سامانه‌ی بینایی پرداخته است. همچنین، یک آزمایشگاه پژوهشی در این حوزه دایر نموده (http://ccvlab.ir)، در این زمینه به تربیت پژوهشگران و دانشجویان برجسته پرداخته و موفق به انتشار مقالات متعددی در مجلات معتبر بین المللی علوم شناختی شده است. با توجه به اشتیاق روزافزون پژوهشگران و دانشجویان به حوزه‌ی بازشناسی اشیا و همچنین نبود مرجعی که در آن به صورت جامع به معرفی و بررسی مدل‌های محاسباتی مذکور پرداخته باشد، بر آن شدم که نتیجه‌ی سال‌ها پژوهش و تجربه‌ی خود را در قالب کتابی به رشته‌ی تحریر درآورم.

ساختار کتاب در چهار فصل طراحی شده است. پس از فصل اول که در آن به معرفی اجمالی مفاهیم و موضوعات پایه‌ی مطرح در حوزه‌ی بازشناسی اشیا پرداخته شده، در فصل دوم، ساختار مغز، بخش‌های مختلف آن و سامانه‌ی بینایی مرور گشته است. در فصل سوم مدل‌های محاسباتی-ـشناختی بازشناسی اشیا معرفی شده و ساختار و عملکرد آنها مورد بررسی قرار گرفته است. در نهایت، کتاب در فصل چهارم با معرفی یکی از مدل‌های موفق بازشناسی اشیا ــ که از خانواده‌ی شبکه‌های عصبی کانولوشنی است و از رویکرد یادگیری عمیق بهره می‌برد ــ به پایان می‌رسد.

اگرچه در نگارش کتاب تمام سعی نویسنده بر این بوده که دقت کافی را در نگارش مطالب در نظر بگیرد، بی‌شک این اثر نیز مانند دیگر آثار انسانی خالی از اشکال نخواهد بود؛ بنابراین، باعث خوشحالی نویسنده خواهد بود که تمامی خوانندگان و پژوهشگران در این زمینه نظرات و پیشنهادات اصلاحی خود را جهت بهبود این اثر در چاپ‌های بعدی از طریق آدرس rebrahimpour@srttu.edu در اختیار نویسنده قرار دهند.

بر خود لازم می‌دانم از زحمات اساتید و همکاران بزرگوارم جناب آقایان دکتر حسین استکی، دکتر احسان‌اله کبیر، دکتر بابک نجار اعرابی، دکتر سیدمهدی خلیق رضوی، دکتر محمد گنج تابش و دکتر نصور باقری و همچنین پژوهشگران و دانشجویان بزرگوارم خانم‌ها معصومه مکاری

محلاتی، فریبا عباسی، الهه سلطان‌دوست، محبوبه حبیبی و فاطمه فلاح و آقایان کریم رجایی، امین میرزایی، سجاد ذباح، مسعود قدرتی، امیرحسین فرض‌مهدی، حمید کریمی روزبهانی، رضا اقدام و حمید گل‌محمدی در راستای نگارش این کتاب کمال تشکر و قدردانی را نمایم.
در خاتمه مراتب سپاس و قدردانی خود را به همسر عزیزم که با صبوری و شکیبایی، همواره مشوق اینجانب در تمامی این سال‌ها بوده‌اند، اعلام می‌دارم. همچنین، این اثر را با تمام وجود، به دختران دلبندم هستی و دلسا تقدیم می‌کنم.

رضا ابراهیم‌پور
عضو هیأت علمی دانشگاه تربیت دبیر شهید رجائی
زمستان ۱۳۹۵

فصل ۱

مقـــدمه

۱-۱- مقدمه

از دهه‌ی ۶۰ میلادی و کار پیشگام هاجکین هاکسلی[1] برای مدل‌سازی فرآیند تولید ضربه[2][۱]، توسعه‌ی مدل‌های محاسباتی برای علوم اعصاب[3] شاهد جهش بزرگی بوده است. برای مثال تاکنون تنها برای ناحیه‌ی بینایی اولیه[4] مدل‌های فراوانی ارائه شده است که در واقع تنها برخی از آن‌ها در شبیه‌سازی پردازش ناحیه‌ی بینایی اولیه از هسته‌ی خمیده‌ی جانبی[5](LGN) تا V_1 کاملاً موفق بوده‌اند[۲].

تحقیقات انجام‌شده برروی نواحی قشری مخطط[6] و فرامخطط[7] در دهه‌های گذشته، اطلاعات قابل‌توجهی را تولید کرده‌اند. آگاهی از سازوکار[8] بازشناسی اشیا[9] نیازمند ایجاد پلی میان سطوح مختلف یافته‌ها از پردازش اطلاعات یا سطح محاسباتی گرفته تا سطح مدارها و سازوکارهای سلولی و بیوفیزیکی است. چگونگی بازشناسی اشیا در حقیقت برای هر مدل مغزی بسیار پیچیده

---

[1] Hodgkin And Huxley
[2] Spike
[3] Neuroscience
[4] Primary Visual Cortex (V_1)
[5] Lateral Geniculate Nucleus
[6] Striate
[7] Extrastriate
[8] Mechanism
[9] Object Recognition

است. از این رو، به نظریه‌های محاسباتی کمّی نیاز است که بتوانند: ۱) اطلاعات موجود را خلاصه و سازمان‌دهی کنند. ۲) به برنامه‌ریزی، هماهنگی و تفسیر آزمایش‌های جدید کمک کنند. با ظهور ابررایانه‌ها و معماری مختص شبیه‌سازی مدل‌های دقیق پردازش نورونی[۱]، مانند مغز آبی[۲][۳]، به‌زودی شبیه‌سازی مدارهای دقیق تشکیل‌دهنده مغز در طی نواحی بینایی ممکن خواهد بود.

در این میان، تلاش‌ها برای شناخت هر چه بیشتر کارکرد مغز منجر به کشف اطلاعاتی در مورد نحوه‌ی انجام پردازش‌های مرتبط با بازشناسی اشیا شده است. یافته‌ها حاکی از آن است که انسان‌ها حتی در صحنه‌های به‌هم‌ریخته با سرعت و دقّت قابل‌ملاحظه‌ای قادر به بازشناسی اشیا هستند که این موضوع بیانگر پیچیدگی شگفت‌آور سامانه‌ی بینایی[۳] انسان است. این پیچیدگی از زمان مطالعات هابل و ویزل[۴] در مورد قشر بینایی[۵] گربه و میمون در اواخر دهه‌ی ۵۰ و دهه‌ی ۶۰ میلادی آشکار شد. تاکنون نواحی بینایی دیگری نیز کشف شده‌اند که بررسی‌های اخیر بالغ بر ۳۰ ناحیه را که به طور پیچیده و با یک الگوی مبهم به یکدیگر متصل شده‌اند نشان می‌دهند[۴] و [۵]. این طرح ارتباطی پیچیده می‌تواند به دو مسیر اصلی تقسیم‌بندی شود: مسیر "چیستی[۶]" (جریان شکمی[۷]) که از ناحیه‌ی بینایی اولیه، یعنی V_1 شروع می‌شود و از طریق V_2 و V_4 به قشر فروگیجگاهی[۸](IT) می‌رسد و مسیر "کجایی[۹]" (جریان پشتی[۱۰]) که از V_1 به V_2، V_3، MT، MST و دیگر نواحی قشری می‌رود. در این چهارچوب، مسیر "چیستی" برای بازشناسی اشیا شناخته شده در حالی که مسیر "کجایی" مختص چشم‌انداز فضایی است. با یک نگاه دقیق‌تر به جریان شکمی، شاهد افزایش پیچیدگی ویژگی‌های محرک[۱۱] سلول‌ها و همچنین اندازه‌ی میدان دریافت[۱۲] آن‌ها هستیم[۶]. درحالی‌که نورون‌های ناحیه‌ی V_1 به محرک‌های میله‌ای جهت‌دار پاسخ می‌دهند و میدان دریافت کوچکی دارند، سلول‌های ناحیه‌ی IT به

---

[1] Neuron
[2] Blue Brain
[3] Visual System
[4] Hubel And Wiesel
[5] Visual Cortex
[6] What
[7] Ventral Stream
[8] Inferior Temporal
[9] Where
[10] Dorsal Stream
[11] Stimulus
[12] Receptive Field

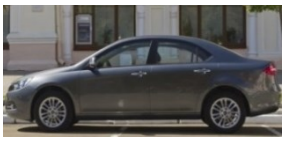
محرک‌های پیچیده‌تری نظیر چهره در مکان‌ها و مقیاس‌های متفاوتی پاسخ می‌دهند که به نقش حیاتی قشر IT در بازشناسی شیء اشاره دارد.
بسیاری از مدل‌های محاسباتی بینایی برای شبیه‌سازی جریان شکمی بینایی، به عنوان مسیر بازشناسی اشیا، ارائه شده‌اند که در این کتاب به بررسی برخی از آن‌ها خواهیم پرداخت.

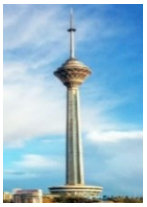
۱-۲- بازشناسی اشیا

یکی از اساسی‌ترین فعالیت‌های شناختی انسان‌ها و جانوران بازشناسی اشیا است. بسیاری از امور روزمره اساساً به توانایی بازشناسی اشیا، صحنه و طبقه‌بندی[1]ها متکی هستند.
بازشناسی اشیا به توانمندی برچسب زدن اشیا از برچسب[2]های دقیق[3] (شناسایی) تا برچسب‌های کلی[4] (طبقه‌بندی) اطلاق می‌شود (شکل ۱-۱). در مورد اول هدف شناسایی نمونه‌هایی خاص از یک شیء، مکان و یا فرد است. برای مثال، چهره‌ی دکتر حسابی، برج میلاد و یا جلد یک کتاب خاص از آن جمله‌اند. در مقابل، در طبقه‌بندی به دنبال بازشناسی نمونه‌های مختلفی از یک دسته‌ی کلی هستیم که از لحاظ مفهومی به یک کلاسِ[5] یکسان مانند ساختمان‌ها، فنجان‌های قهوه یا ماشین‌ها تعلق دارند. در هر دو مورد می‌توان انتظار داشت که در ظاهرِ نمونه‌های مختلف یک کلاس و یا یک شیء تغییراتی وجود داشته باشد.
اولین سؤالی که در حل مسأله‌ی بازشناسی مطرح می‌شود این است که چه چیزهایی می‌توانند بر پایه‌ی اطلاعات بینایی شناخته شوند؟ برای پاسخ به این سؤال بهتر است که چگونگی سازمان‌دهی دانش در انسان‌ها را مرور کنیم. این سؤال در روانشناسی شناختی[6] بسیار مورد توجه است[۷].
به‌عنوان مثال براون[7] می‌گوید که یک سگ می‌تواند به‌عنوان یک "سگ"، یک "باکسر[8]"، یک "چهارپا" و یا یک "جاندار" تصور شود[۷]. اما هنوز هم سگ عبارتی است که ساده‌تر به ذهن خطور می‌کند و به هیچ عنوان تصادفی نیست. آزمایش‌ها نشان می‌دهند که یک سطح پایه در طبقه‌بندی در انسان وجود دارد که بیشتر دانش در آن سازمان‌دهی می‌شود[۸].

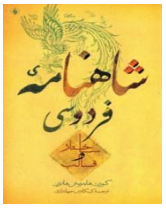
---

[1] Category
[2] Label
[3] Specific
[4] Generic
[5] Class
[6] Cognitive Psychology
[7] Brown
[8] Boxer


ماشین همسایه‌ی من


برج میلاد

دکتر حسابی



شاهنامه‌ی فردوسی

بازشناسی

ساختمان

ساختمان

ماشین

ماشین

دسته‌بندی

شکل ۱-۱- مسأله‌ی بازشناسی شامل شناسایی نمونه‌هایی از یک شیء خاص یا صحنه (بالا) و همچنین تعمیم دادن آن برای دسته‌بندی نمونه‌ها در سطح پایه (پایین).

بر طبق[۸] و [۹] سطح پایه به صورت زیر است:

- بالاترین سطحی که در آن اعضای یک دسته، شکل ادراکی مشابه تولید می‌کنند.
- بالاترین سطحی که در آن یک تک تصویر ذهنی می‌تواند کل دسته را منعکس کند.
- بالاترین سطحی که در آن یک فرد حرکت‌های مشابهی را برای تعامل با اعضای دسته به کار می‌برد.
- بالاترین سطحی که در آن معمولاً سریع‌ترین شناسایی اعضای دسته توسط انسان اتفاق می‌افتد.
- اولین سطحی که توسط کودکان فهمیده و نام‌گذاری می‌شود.

دسته‌بندی عام‌تر از سطح پایه دسته‌های مافوق[۱] هستند و دسته‌های خاص‌تر دسته‌های مادون[۲] نامیده می شوند. بنابراین، دسته‌بندی اشیا در انسان‌ها شامل سه سطح است: مافوق (حیوان، وسیله‌ی نقلیه، مبلمان)، سطح پایه (سگ، ماشین، صندلی) و مادون (سگ گله اسکاتلندی[۳]، پورشه، صندلی من)[۸]. اگرچه تمام این دسته‌بندی‌ها مهم هستند اما راش[۴] و همکارانش نشان دادند که یکی از آن‌ها بر بقیه برتری دارد. زمانی که از افراد خواسته می شود که تصاویر اشیای رایج را نام‌گذاری کنند، معمولاً از نام‌های سطح پایه استفاده می‌کنند ([۱۰] را ببینید). همچنین وقتی از افراد خواسته می شود که تعلق یک تصویر به یک دسته‌ی خاص را بررسی کنند، تصمیمات افراد برای سطح پایه سریع‌تر است[۸]. این یافته‌ها نشان می‌دهند که اولین ارتباط بین درک شیء و اطلاعات معناییِ[۵] آن در سطح پایه اتفاق می‌افتد که در بازشناسی شیء با نام دسترسی اولیه[۶] و یا نقطه‌ی ورود[۷] نیز شناخته می‌شود.

تصور می‌شود که پردازش‌های بیشتری در دسته‌بندی مافوق نیاز است و این دسته‌بندی با دسته‌بندی‌های مادون و پایه متفاوت است درحالی‌که دسته‌بندی‌های پایه و مادون شباهت بیشتری دارند. راش و همکارانش نشان دادند که نام‌های پایه و مادون شناسایی را تسریع می‌کنند درحالی‌که نام‌های مافوق چنین قابلیتی ندارند.

به دلایلی که گفته شد مقدار قابل توجهی از تحقیقات بازشناسی اشیا بر دسته‌بندی پایه تمرکز کرده‌اند[۱۱].

- **چالش‌ها**

مسأله‌ی بازشناسی اشیا دارای چالش‌های بسیاری است[۱۲] و [۱۳]. نمونه‌های یک دسته از اشیا می‌توانند با توجه به عوامل مؤثری مانند روشنایی[۸]، جهت‌گیری عمقی (وضعیت)[۹] شیء، زاویه‌ی

[1] Superordinate
[2] Subordinate
[3] Collie
[4] Rush
[5] Semantics
[6] Primal Access
[7] Entry Point
[8] Illumination
[9] Pose

دید (نقطه‌ی دید[1])، انسدادهای جزئی[2] و پس‌زمینه[3] "در هم ریختگی[4]"، تصاویر بسیار متفاوتی را بر روی شبکیه‌ی[5] چشم و یا حسگر نوری در دوربین ایجاد کند (شکل ۱-۲). نمونه‌های مختلف اشیا از یک طبقه نیز می‌توانند تغییرات چشمگیری در ظاهر شیء ایجاد کنند. علاوه بر این، در بسیاری از موارد زمانی که شیء به‌تنهایی بررسی می‌شود، مبهم است و برای بازشناسی شیء به ارتباط شیء با صحنه‌ی پس‌زمینه و آمادگی اولیه برای مواجهه با شیء نیاز است.

شکل ۱-۲- تصاویر محتوی نمونه‌های یک دسته‌ی مشابه می‌توانند متفاوت ظاهر شوند. مدل‌های بازشناسی اشیا باید نسبت به این تغییرات مقاوم باشند.

[1] View point
[2] Partial Occlusions
[3] Background
[4] Clutter
[5] Retina

علاوه بر مسائل مربوط به مقاومت[1]، الگوریتم‌های بازشناسی امروزه با چالش عمده‌ای در پیچیدگی محاسباتی و مقیاس‌پذیری مواجه هستند. این حقیقت که در حدود نیمی از قشر مغز[2] در پستانداران درگیر پردازش اطلاعات بینایی می‌شود، نشانه‌ی بار محاسباتی زیاد در بازشناسی اشیا است[۴]. از این رو، نیاز است که الگوریتم‌های بسیار بهینه برای جستجوی پایگاه‌های داده بزرگ تصویر و یا برای تعمیم بازشناسی به هزاران نوع دسته، با بازنمایی‌های[3] غنی از تصویر با ابعاد بالا توسعه داده شوند[۱۴].

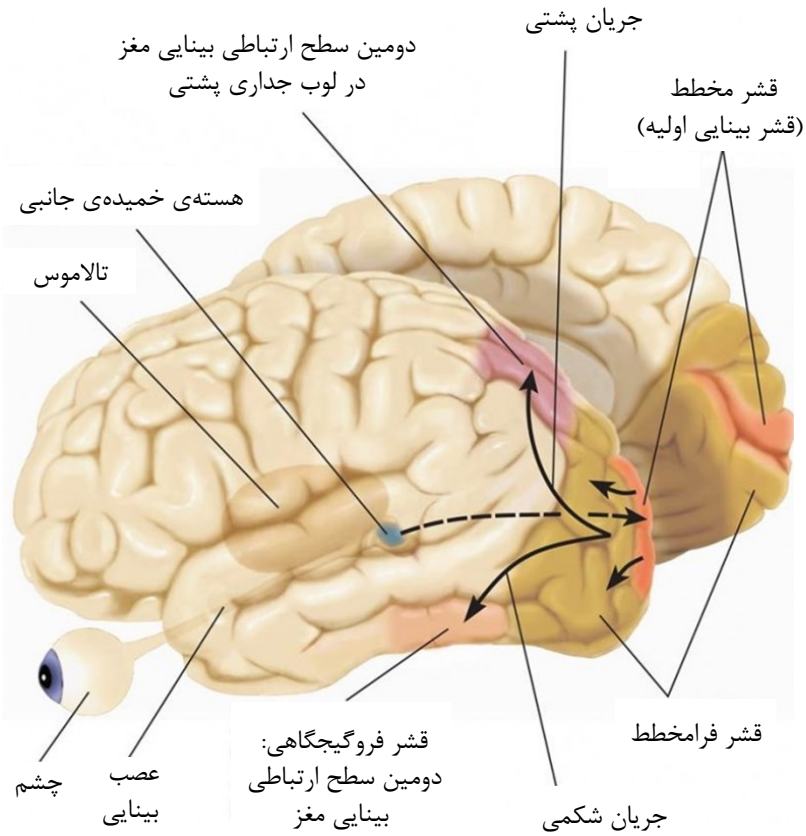
۱-۳- سامانه‌ی بینایی انسان

سامانه‌ی بینایی انسان از شبکیه‌ی چشم شروع می‌شود. سیگنال‌های بینایی در شبکیه‌ی چشم تولید و از مسیر عصب بینایی، پس از انتقال به نیمکره‌ی[4] مقابل، به ناحیه‌ی هسته‌ی خمیده‌ی جانبی در تالاموس[5] می‌رسند. در آن نقطه، پس از پردازش‌های اولیه به‌سمت قشر اولیه‌ی بینایی در ناحیه‌ی V_1 حرکت می‌کنند. اطلاعات بینایی سپس به ناحیه‌ی V_2 رفته و در آن ناحیه به دو مسیر مجزا می‌روند: مسیرهای شکمی و پشتی (شکل ۱-۳). در پژوهش‌های پیشین نشان داده شده است که عمده‌ی اطلاعات مربوط به ماهیت شیء در نواحی شکمی پردازش می‌شوند و ناحیه‌ی پشتی مسئول پردازش اطلاعات رنگ و موقعیت اشیا است. از این رو، نواحی شکمی و پشتی به‌ترتیب به‌عنوان نواحی "چیستی" و "کجایی" نیز شناخته می‌شوند. البته پژوهش‌های اخیر مسأله‌ی استقلال کامل مسیرهای شکمی و پشتی را به چالش کشیده‌اند[۱۵] و [۱۶].

با توجه به موضوع کتاب (بازشناسی اشیا)، در ادامه عمدتاً به ویژگی‌های عملکردی مسیر شکمی پرداخته خواهد شد. در امتداد مسیر بینایی شکمی، اطلاعات بینایی از ناحیه‌ی V_2 به V_4 رسیده و از آن‌جا به IT انتقال می‌یابند. اطلاعات بینایی که در شبکیه‌ی چشم تولید و در طول مسیر شکمی قشر بینایی پیش می‌روند در هر مرحله دچار تغییرات زیادی می‌شوند. از آن جمله می‌توان به غیرهمبسته‌سازی[6] زمانی و مکانی و افزایش تباین[7] اشاره کرد. به‌طور کل در هر لایه اطلاعات

[1] Robustness
[2] Cortex
[3] Representation
[4] Hemisphere
[5] Thalamus
[6] De-correlation
[7] Contrast

ورودی فیلتر شده و ویژگی‌های[1] خاصی از آن استخراج می‌شود. بر اساس یافته‌های موجود از مسیر شکمی عموم دانشمندان بر یافته‌های ذیل در خصوص این مسیر اتفاق نظر دارند:



شکل ۱-۳- مسیرهای انتقال اطلاعات بینایی از شبکیه‌ی چشم به قشر اولیه‌ی بینایی و مسیرهای پشتی و شکمی.

۱. سامانه‌ی بینایی انسان دارای ساختاری سلسله‌مراتبی[2] است و بدین ترتیب در سلسله‌مراتب با تغییراتی همچون اندازه[3]، موقعیت[4]، زاویه‌ی دید و . . . مقابله می‌کند[۱۷].

[1] Feature
[2] Hierarchical
[3] Size
[4] Position

۲. در طول این سلسله‌مراتب، میدان دریافت نورون‌ها بزرگ‌تر و محرک‌های مطلوب آن‌ها پیچیده‌تر می‌شوند.
۳. پردازش اولیه در سامانه‌ی بینایی به‌صورت پیش‌رو[1] انجام می‌شود (در مسیر پیش‌رو بازشناسی صورت می‌گیرد و در صورت نیاز، سازوکارهایی همچون توجه درگیر می‌شوند تا به بازشناسی اشیا کمک نمایند).
۴. یادگیری با احتمال بالا در تمامی لایه‌های سامانه بینایی و به‌طور قطع در لایه‌های نهایی آن نظیر قشر فروگیجگاهی و جلویی پیشانی[2] صورت می‌گیرد.
بر اساس این ویژگی‌های سامانه‌ی بینایی، اطلاعات بینایی به‌صورت یک بازنمایی در مغز انسان شکل می‌گیرند و تمام فعالیت‌های شناختی بعدی مرتبط با بینایی بر روی بازنمایی‌های شکل گرفته از اشیا در مغز صورت می‌پذیرد. یکی از مهم‌ترین این فعالیت‌های شناختی بازشناسی اشیا است که پیش‌تر تشریح شد.

- **ضرورت آگاهی از نحوه‌ی بازشناسی اشیا در انسان**

آگاهی از سازوکارهای انجام عمل بازشناسی اشیا در انسان می‌تواند راه‌گشای حل بسیاری از مسائل بینایی ماشین[3] باشد. در راستای پاسخ به چگونگی ایجاد قابلیت بازشناسی اشیا، متخصصان علوم شناختی[4]، علوم اعصاب بینایی[5] و همچنین مهندسان مدل‌سازی عصبی هر کدام از جنبه‌ای متفاوت سعی در آشکار ساختن حقایق این مسأله نموده‌اند.
با بهره‌گیری از یافته‌های متخصصان علوم اعصاب بینایی، مهندسان مدل‌سازی عصبی وارد عرصه شده و مدل‌هایی شناختی-محاسباتی[6] از سامانه بینایی ارائه نموده‌اند[۱۸]–[۲۲]. این مدل‌ها در سطوح مختلف از جمله سطح سلولی[۲۴]، سطح سامانه‌ای[7][۲۲] و سطح شناختی[۲۵] توسعه داده شده و قابلیت پیش‌بینی رفتار سامانه بینایی را تحت شرایط خاص فراهم نموده‌اند. همچنین، این مدل‌ها که از عملکردهای محاسباتی و شناختی مغز استفاده می‌کنند، توانسته‌اند تا حدود زیادی سازوکارهای به‌خدمت گرفته شده توسط مغز برای حل مسأله بازشناسی اشیا را الگوبرداری نموده و مسائل کنونی در حوزه‌ی بینایی ماشین را حل نمایند.

[1] Feed-Forward
[2] Prefrontal
[3] Machine Vision
[4] Cognitive Science
[5] Visual Neuroscience
[6] Cognitive-Computational
[7] System-Level

۱-۴- مدل‌های محاسباتی بینایی مبتنی‌بر سامانه‌ی بینایی انسان

یکی از رایج‌ترین روش‌های آگاهی از سازوکارهای بازشناسی اشیا در انسان مدل‌سازی محاسباتی این سامانه است. مهم‌ترین کاربردی که مدل‌های محاسباتی قشر بینایی دارند فراهم آوردن قابلیت پیش‌بینی رفتار این سامانه تحت شرایط مختلف، بدون نیاز به انجام آزمایش‌های انسانی و حیوانی است. همچنین، این مدل‌ها می‌توانند در توضیح فعالیت بازشناسی اشیا در انسان مفید واقع شوند. بر همین اساس، تا به امروز مدل‌های متنوعی با الهام از مسیر شکمی قشر بینایی توسعه یافته‌اند. هابل و ویزل بر اساس نتایج مطالعات الکتروفیزیولوژیکی[1] روی قشر بینایی گربه، اولین مدل مفهومی را ارائه دادند که متشکل از سلسله‌مراتبی از لایه‌ها با واحدهای ساده و پیچیده بود. فوکوشیما[2] در سال ۱۹۸۰ بر اساس مدل مفهومی، شبکه[3] Neocognitron را معرفی کرد[۱۸]. در توسعه‌ی این مدل تلاش شد که نه تنها ساختار سلسله‌مراتبی قشر بینایی انسان مدنظر قرار گیرد بلکه از ویژگی‌های سلول‌های ساده[4] و پیچیده[5] که در ساختار سامانه‌ی بینایی شناخته شده‌اند، بهره برده شود. بر این اساس سلول‌های ساده که از موقعیت‌های مختلفی از فضای ورودی نمونه‌برداری می‌کنند، ورودی لازم برای تحریک سلول‌های پیچیده را فراهم می‌آورند. لایه‌هایی متوالی از ترکیب این سلول‌ها در نهایت ساختاری سلسله‌مراتبی را شکل می‌دهند که در طی روند پایین به بالا از حساسیت عناصر محاسباتی به تغییرات موقعیت مکانی محرک ورودی کاسته و به پیچیدگی محرک‌های تحریک‌کننده افزوده می‌شود.

در ادامه‌ی توسعه‌ی مدل‌های محاسباتی قشر بینایی، مدل VisNet در سال ۱۹۹۷ توسط رالز[6] و همکارانش در دانشگاه آکسفورد معرفی شد[۲۶]. این مدل که از ساختاری سلسله‌مراتبی بهره می‌برد یک تفاوت اساسی با مدل Neocognitron دارد و آن در نحوه‌ی آموزش مدل نهفته است. مدل VisNet تنها بر اساس الگوریتم بدون سرپرستِ[7] اثر[8] آموزش می‌بینند. این الگوریتم بر این اصل اساسی استوار است که دنیای پیرامون ما عموماً به آهستگی تغییر می‌کند و این تغییرِ آهسته، نورون‌های سامانه‌ی بینایی را آموزش می‌دهد. بر اساس این واقیعت و با استفاده از قانون اثر، تصاویر مختلفی که از تغییرات یک شیء حاصل می‌شود همگی به یک نورون و یا یک دسته

---

[1] Electrophysiology
[2] Fukushima
[3] Network
[4] Simple Cell
[5] Complex Cell
[6] Rolls
[7] Unsupervised
[8] Trace Rule

از نورون‌ها آموزش داده می‌شود و آن نورون تغییرات مذکور را که همگی متعلق به یک شیء هستند یاد می‌گیرد. به همین دلیل بیشتر از این شبکه به‌عنوان مدلی برای بررسی ویژگی پیوستگی زمانی-مکانی مغز استفاده می‌شود. این تغییرات شامل تغییر موقعیت، اندازه و زاویه‌ی دید می‌باشند. یکی از مسائلی که استفاده از این مدل را در کاربردهای بازشناسی اشیا محدود نموده است همین مسأله است، زیرا به‌جز تعداد محدودی مجموعه‌داده که تصاویری از اشیا را تحت تأثیر تغییرات مختلف ارائه می‌دهند، مجموعه‌داده‌های رایج قابلیت اِعمال به این شبکه را ندارند، زیرا در آن صورت کارایی شبکه به شدت افت می‌کند.

مدل HMAX که در سال ۱۹۹۹ توسط گروه پوجیو[1] ارائه شد، افقی جدید در مدل‌سازی محاسباتی سامانه‌ی بینایی انسان گشود[۲۳]. مدل ارائه شده توسط این گروه که با ارائه‌ی شواهد زیستی منطبق بر آن همراه بود شواهدی را در دفاع از استفاده سامانه‌ی بینایی انسان از عملگر بیشینه‌گیری[2] مطرح نمود. این مدل که ساختاری سلسله‌مراتبی داشت براساس یافته‌های هابل و ویزل در خصوص سلول‌های ساده و پیچیده توسعه داده شد. در این مدل طی سلسله مراتب، پیچیدگی محرک‌ها افزایش می‌یابد و بازنمایی‌های ورودی نسبت به تغییر اندازه و موقعیت پایا[3] می‌گردند. پایایی نسبت به جابجایی با نمونه‌برداری از تمام فضای لایه پیشین و پایایی نسبت به تغییر اندازه با کمک عملگر بیشینه‌گیری صورت می‌پذیرد. نسخه‌ی اولیه‌ی این مدل دارای پیچیدگی کمتری بود و تنها برای بازشناسی محرک‌های بینایی ساده مانند گیره کاغذ مورد استفاده قرار گرفت، درحالی‌که در سالیان بعد نسخه‌های پیشرفته آن برای بازشناسی تصاویر پیچیده و در کاربردهای عملی مورد استفاده قرار گرفت. از جمله مواردی که محبوبیت این مدل را افزود می‌توان به توانایی بالای مدل در حل مسائل بینایی ماشین و در عین حال انطباق با مسیر شکمی قشر بینایی از جمله موارد ذیل اشاره نمود: استفاده از عملگر بیشینه‌گیری، ساختار سلسله‌مراتبی با افزایش پیچیدگی محرک‌ها و استقلال از تغییرات محرک، آموزش به روش بدون سرپرست و با نمونه‌برداری تصادفی از تصاویر آموزشی، استفاده از فیلتر گابور[4].

البته علاوه بر مدل‌های محاسباتی مذکور، مدل‌هایی دیگر نیز در راستای توضیح عملکرد سامانه‌ی بینایی انسان توسعه یافته‌اند که عموماً نسخه‌هایی نه‌چندان متفاوت از مدل‌های بالا ارائه می‌کنند که در این بخش از تشریح آن‌ها خودداری می‌کنیم.

---

[1] Poggio
[2] Maximum
[3] invariant
[4] Gabor Filter