

## ارائه یک الگوریتم تشخیص شیء عمیق موازی مبتنی بر شبکه‌های عصبی تک لایه رو به جلو ترکیبی و تطبیقی

حامد خرم فر

دانشجوی کارشناسی ارشد هوش مصنوعی دانشگاه آزاد اسلامی واحد فردوس، گروه مهندسی کامپیوتر

حمید غفاری

عضو هیات علمی دانشگاه آزاد اسلامی واحد فردوس

### چکیده

پیشرفت‌های اخیر در حوزه‌های علم رایانه و سیستم‌های هوشمند از جمله توان محاسباتی سیستم‌ها می‌باشد. در سیستم‌های هوشمند یادگیری عمیق از نظر کاربردی برای حوزه‌های مختلف یادگیری ماشین، بینایی ماشین، پردازش تصویر، پردازش زبان‌های طبیعی و داده کاوی مطرح می‌شود. در بین متدهای مطرح شده در یادگیری عمیق، شبکه‌های کانولوشن به طور گسترده برای کاربردهای مختلفی استفاده می‌شوند. یکی از مسائل مهمی که به عنوان چالش در این حوزه مطرح است بحث شناسایی و دسته‌بندی تصاویر در چند دامنه‌ی متفاوت است. دسته‌بندی و تشخیص اشیاء در شرایط متفاوت از نظر ویژگی‌های تصور مانند رنگ، نور، زاویه و تصویر پس زمینه دارای عملکرد یکسان با نرخ تشخیص بالا نیست. در این مقاله یک روش مبتنی بر مدل‌های یادگیری عمیق از پیش آموزش دیده شده در مرحله استخراج ویژگی ارائه شده است. علاوه بر مدل‌های عمیق سعی در به کار گیری داده‌ها از چند دامنه متفاوت می‌باشد به نحوی که سیستم از داده‌های متفاوتی در مرحله آموزش و تست استفاده کند. همچنین جهت افزایش مقاومت سیستم از روش شبکه عصبی تک لایه برای مرحله دسته‌بندی داده‌های بدون برچسب در کنار داده‌های برچسب دار نیز استفاده می‌شود. در این شبکه عصبی علاوه بر عبارت خطای مرسوم از عبارات خطای دیگری نیز استفاده می‌شود که موجب بهبود عملکرد این شبکه در مرحله دسته‌بندی می‌شود. هدف اصلی این مقاله ارائه یک الگوریتم تشخیص شیء عمیق موازی مبتنی بر شبکه‌های عصبی تک لایه رو به جلو ترکیبی و تطبیقی است. نتایج حاصل از این روش با سه مدل عمیق متفاوت با یکدیگر ترکیب شدند تا عملکرد کلی سیستم بهبود پیدا کند. نتایج بدست آمده بر روی مجموعه داده‌های استفاده شده در این مقاله نشان دهنده بهبود بیش از یک درصد در عملکرد دسته‌بندی شیء عمیق دارد.

**واژگان کلیدی:** شبکه عصبی تک لایه، تشخیص شیء، یادگیری عمیق، تشخیص شیء چند دامنه ای، شبکه عصبی کانولوشن

## ۱. مقدمه

تشخیص شیء فرایندی است که طی آن بر روی تصویر ورودی یک سیستم برچسب یا برچسب‌هایی نسبت داده می‌شود که در واقع نشان دهنده میزان احتمال حضور شیء موجود در یک ناحیه از تصویر در یک دسته خاص است. تشخیص اشیاء<sup>۱</sup> را می‌توان به عنوان یکی از مسائل موجود در مجموعه دسته‌بندی تصاویر در نظر گرفت. فرایند تشخیص شیء می‌تواند با تشخیص جزئیات در سطوح متفاوتی انجام شود. یکی از روش‌های تشخیص حضور شیء در تشخیص مکان حضور آن استفاده از پنجره بر روی تصویر است، به این ترتیب که پنجره بر روی تصویر حرکت می‌کند و در صورت بروز اشتراک پنجره با هدف آن ناحیه از تصویر را به عنوان مکان مورد نظر معرفی می‌کند (Guo et al, 2015).

به کارگیری یادگیری عمیق در زمینه‌های مرتبط با تصاویر طی سال‌های اخیر به شدت افزایش پیدا کرده است، به گونه‌ای که در برخی موارد در برنامه‌های کاربردی بر روی سیستم عامل‌ها و چارچوب‌های کاری متفاوتی همچون تلفن‌های همراه نیز به کار گرفته شده است. یکی از روش‌های به کارگیری مدل‌های عمیق از پیش آموزش دیده شده این است که از مدل عمیق به عنوان یک خبره برای استخراج ویژگی‌ها استفاده می‌شود و تنها بخش پایانی سیستم هوشمند که وظیفه تشخیص را بر عهده دارد بر روی داده‌های جدید آموزش می‌بیند. از جمله کاربردهای مهم یادگیری عمیق در استخراج ویژگی‌ها در داده‌های آنالوگ است که به خوبی برتری خود را نسبت به سایر روش‌ها نشان داده است (Glorot et al, 2010).

یادگیری عمیق به دلیل تسلط آن در آموزش مجموعه داده‌های بزرگ در هر زمینه تحصیلی رواج قابل توجهی پیدا می‌کند (Kocher and Kuma, 2021) و کارایی خود را در استخراج ویژگی از اشیاء، تصاویر، ویدئوها و متن ثابت کرده است و از این طریق در کارهای مختلف تشخیص، تقسیم بندی و طبقه بندی اشیاء موفق بوده است (Rostami et al, 2022).

پیش از این نیز روش‌هایی برای حل مسائل در دامنه‌های مختلف نیز بیان شده است، روش‌هایی مانند AELM<sup>۲</sup> که گسترش یافته AELM معمولی است و یک توسعه از مسائل تک دامنه‌ای به مسائل چند دامنه‌ای می‌باشد و تنها از یک مدل عمیق از پیش آموزش دیده و یک مجموعه داده بدون برچسب از دامنه هدف در فرآیند آموزش استفاده می‌شود که به دلایل متعددی مناسب نیستند. در مجموع به دلیل انتخاب تعداد نورون‌های لایه مخفی به صورت از پیش تعریف شده یا تصادفی و یا استفاده از آزمون و خطا برای تعیین تعداد نورون‌ها، این شبکه‌ها قادر به دستیابی به ساختار بهینه و فشرده نمی‌باشند (Wu et al, 2015).

در این مقاله هدف اصلی ارائه یک الگوریتم تشخیص شیء عمیق موازی مبتنی بر شبکه‌های عصبی تک لایه رو به جلو ترکیبی و تطبیقی است که مسائل دسته‌بندی چند دامنه‌ای را بهبود می‌دهد. بطور کلی سعی بر این است که دسته‌بندی تصاویر و تشخیص اشیاء را در چند دامنه<sup>۳</sup> انجام بدهیم. تشخیص شیء روی چند دامنه به معنای داشتن دو دامنه متفاوت در منبع و دامنه هدف است و هر یک دارای توزیع متفاوتی هستند (Zhang et al, 2017). علت این امر نیز چنین است که کارایی روش‌های سنتی برای دسته‌بندی تصاویر و تشخیص اشیاء، از جمله ماشین بردار پشتیبان<sup>۴</sup>، نزدیکترین همسایه<sup>۵</sup>، ماشین یادگیری افراطی<sup>۶</sup> پایین است و عمدتاً پاسخ مناسبی را در مسائل دارای چند دامنه نشان نمی‌دهند درحالی که در مسائل با یک دامنه کارایی بهتری دارند.

ادامه این مقاله به صورت زیر تنظیم شده است: بخش ۲ کار مرتبط را توضیح می‌دهد. در بخش ۳، الگوریتم پیشنهادی را توضیح می‌دهیم. بخش ۴ شامل ارزیابی نتایج آزمایش شبیه سازی است. در نهایت، نتیجه گیری را در بخش ۵ ارائه می‌دهیم.

<sup>۱</sup> Object detection

<sup>۲</sup> Adaptive ELM(AELM)

<sup>۳</sup> Cross domain

<sup>۴</sup> Support vector machine(SVM)

<sup>۵</sup> k-nearest neighbors(KNN)

<sup>۶</sup> Extreme learning machine(ELM)

## ۲. کارهای مرتبط

در این بخش ابتدا مفاهیم اولیه بیان میشود و در ادامه برخی از تحقیقات مرتبط مطالعه و بررسی میشوند.

اتوانکدر<sup>۱</sup>: نوع خاصی از شبکه عصبی مصنوعی است که برای رمزگذاری<sup>۲</sup> بهینه یادگیری مورد استفاده قرار میگیرد. یک اتوانکدر آموزش میبیند تا بجای آموزش شبکه و پیش بینی مقدار هدف  $Y$  در ازای ورودی  $X$ ، ورودی  $X$  خود را بازسازی کند. بنابراین بردارهای خروجی همان ابعاد بردار ورودی را خواهند داشت. در حین این فرآیند اتوانکدر با کمینه سازی خطای نوسازی<sup>۳</sup> بهینه می شود. کد، متناظر همان ویژگی فراگرفته شده است (Guo et al, 2015 و Hanna et al, 2015).

شبکه های ماشین بولتزمان محدود RBM<sup>۴</sup>: یک شبکه عصبی زایشی-تصادفی است که می تواند توزیع احتمال داده های ورودی خود را فرا گیرد. این شبکه ها توسط هینتون<sup>۵</sup> معرفی شدند و می توانند در کاربردهایی نظیر کاهش بعد، دسته بندی، رگرسیون، سیستم های توصیه کننده، مدل سازی و ... به کار روند (Fischer et al, 2012).

شبکه باور عمیق: همانند RBM یک شبکه عصبی زایشی است که در چند مرحله با استفاده از رویکردی مشابه RBM سعی می کند توزیع احتمالی داده های ورودی را فرا گیرد (Bengio et al, 2007).

شبکه های RNN: این شبکه ها در واقع برای پردازش سیگنال های دنباله دار به وجود آمدند. در یک شبکه عصبی معمولی تمام ورودی ها و خروجی ها مستقل از یکدیگر هستند، اما در بسیاری از موارد این ایده مناسب نیست. به عنوان نمونه زمانی که در یک جمله هدف پیش بینی کلمه بعدی باشد، در صورتی که شبکه بدون فراگیری روابط بین کلمات نمی تواند کلمه بعدی را به درستی پیش بینی کند (Sak et al, 2014).

شبکه های عصبی کانولوشن CNN<sup>۶</sup>: مهمترین و مطرح ترین روش در بین روش های یادگیری عمیق هستند و به طور معمول چندین لایه با ساختارهای مشخص و روشی قدرتمند آموزش می بینند. ساختار این شبکه ها و قابلیت آنها به گونه است که در بیشتر موارد این شبکه ها را به صورت آماده و با هدف استخراج ویژگی به کار می گیرند به این ترتیب که پیش از این با حجم داده بسیار زیاد و توسط سیستم های بسیار قدرتمند در زمان طولانی تحت آموزش قرار گرفته اند تا برای کاربردهای خاص تطبیق داده شوند (Zhang et al, 2017).

لایه کانولوشن: در شبکه های عصبی کانولوشن اولین لایه ورودی را دریافت و با پردازش آن طی عملیات مشخص فیلترها را به آن اعمال می کند. این عملیات تحت عنوان لایه کانولوشن شناخته می شود (Bruna et al, 2013).

شبکه های عصبی تک لایه رو به جلو SLFN: به طور گسترده ای در مسائل دسته بندی و رگرسیون مورد استفاده قرار می گیرند و شامل لایه ورودی، نورون های لایه مخفی و یک لایه خروجی می باشد.

در ادامه مروری بر کارهای گذشته خواهیم داشت:

در پژوهش (Cho Kang, 2021) روش تشخیص شیء توجه آگاه از نگاه انسان برای هوش محیطی به نام Ami<sup>۷</sup> ارائه شده است. در این تحقیق، نویسندگان یک چارچوب جدید تشخیص شیء توجه آگاه از نگاه انسان را به عنوان یک فناوری اساسی برای Ami پیشنهاد می کنند. چارچوب پیشنهادی اشیاء توجه کاربران را شناسایی می کند و عملکرد دقیق تر و قوی تری را در برابر تغییرات مقیاس شی نشان می دهد.

در پژوهش (Liu et al, 2021) یک روش تشخیص تغییر با استفاده از رویکرد یادگیری عمیق با تجزیه و تحلیل تصویر مبتنی بر

<sup>1</sup> Auto encoder

<sup>2</sup> Encode

<sup>3</sup> ReconAuto Encoder

<sup>4</sup> Restricted Boltzmann Machine

<sup>5</sup> Hinton

<sup>6</sup> Convolutional Neural Networks

<sup>7</sup> ambient intelligence

شی ارائه شده است. این مقاله یک روش جدید تشخیص تغییر را پیشنهاد می‌کند که استخراج ویژگی تغییر را با استفاده از شبکه‌های عصبی کانولوشن تحت یک چارچوب OBIA<sup>۱</sup> اجرا می‌کند.

در پژوهش (Jammoussi et al, 2020) تشخیص حالات صورت از طریق شبکه عصبی کانولوشن و ماشین یادگیری شدید ارائه شده است. این مقاله رویکرد تشخیص حالات چهره را بررسی می‌کند. فرآیند طبقه‌بندی تصویر براساس تکنیک‌های یادگیری عمیق برای تشخیص عبارات احساسی است.

در پژوهش (Luo et al, 2020) یک سیستم تشخیص هوشمند مبتنی بر یادگیری عمیق تعبیه‌شده ELM برای ابزارهای ماشین کنترل عددی کامپیوتری ارائه شده است. در این مقاله، یک سیستم تشخیص ابزار هوشمند با یک چارچوب ترکیبی جدید از شبکه یادگیری عمیق چند کاناله با شبکه عصبی پیش‌خور غیر تکراری و سریع برای برآوردن نیازهای کارایی و دقت بالا در تولید هوشمند پیشنهاد شده است.

در پژوهش (Haripriya and Porkodi, 2020) شبکه عصبی کانولوشنال عمیق موازی برای بازیابی تصاویر پزشکی مبتنی بر محتوا ارائه شده است. تصاویر DICOM<sup>۲</sup> که به تشخیص و پیش‌آگهی کمک می‌کند جزء حیاتی در سیستم‌های مراقبت بهداشتی است. بازیابی سریع تصاویر DICOM براساس تصویر پرس و جو داده شده به یک نیاز حیاتی برای آزمایشگاه‌ها تبدیل شده است و پزشکان برای استنتاج سریع و قیاس دقیق شرایط بیمار استفاده می‌کنند.

در پژوهش (Khan and Alarabi, 2021) چارچوب یادگیری عمیق یکپارچه آشکارسازهای چند مقیاسی برای تشخیص شی ژئو فضایی در تصاویر ماهواره‌ای با وضوح بالا ارائه شده است. تشخیص اشیاء ژئوفضایی در تصاویر ماهواره‌ای با وضوح بالا کاربردهای زیادی در برنامه ریزی شهری، کاربردهای نظامی، نظارت دریایی، کنترل و مدیریت محیط دارد. مرحله اول از شبکه هرم ویژگی برای به دست آوردن نقشه‌های ویژگی چند مقیاسی استفاده می‌کند و سپس هر سطح از هرم را با الحاق شبکه‌های پیشنهادی منطقه‌ای RPN<sup>۳</sup> به یک تولیدکننده پیشنهاد مستقل در مقیاس چندگانه تبدیل می‌کند.

در پژوهش (Gudžius et al, 2021) تشخیص شیء مبتنی بر یادگیری عمیق در تصاویر ماهواره‌ای چندطیفی برای کاربردهای بلادرنگ ارائه شده است. تصاویر ماهواره‌ای نحوه درک و پیش‌بینی فعالیت‌های اقتصادی در جهان را تغییر می‌دهد. پیشرفت‌ها در ساخت‌افزار ماهواره‌ای و پرتاب‌های موشکی کم‌هزینه، تصاویری تقریباً واقعی و با وضوح بالا را امکان‌پذیر کرده است که کل زمین را پوشش می‌دهد. یک معماری شبکه عصبی کاملاً کانولوشنال را پیشنهاد می‌شود که برای تشخیص دقیق و شتاب‌دار شی در تصاویر ماهواره‌ای چندطیفی بهینه شده است. این روش تشخیص برای کاربردهای حساس به تاخیر مناسب است.

در پژوهش (Sun et al, 2021) آموزش عمیق گروه موازی برای تشخیص چند هدف حسگر ویدئویی از راه دور در زمان واقعی ارائه شده است. این مقاله یک چارچوب یادگیری عمیق مجموعه موازی را برای تشخیص چند هدف ویدیویی وسایل نقلیه هوایی بدون سرنشین پیشنهاد می‌کند که یک استراتژی تشخیص مشترک جهانی و محلی است. این الگوریتم تشخیص هدف یادگیری عمیق را با تطبیق الگو ترکیب می‌کند تا از اطلاعات تصویر به طور کامل استفاده کند. همچنین مکانیزم‌های چند فرآیندی و چند رشته‌ای را برای سرعت بخشیدن به پردازش یکپارچه می‌کند. این سیستم از دقت تشخیص بالایی برخوردار است و در نمایش بلادرنگ و پایدار نتایج تشخیص تحقق می‌یابد.

در پژوهش (Tao et al, 2022) تشخیص خودکار نقص در اشیاء غیر مسطح با استفاده از الگوریتم‌های یادگیری عمیق ارائه شده است. این مقاله درباره استفاده از شبکه عصبی بازگشتی و شبکه عصبی پیش‌خور مصنوعی در ترموگرافی پالسی در طول تشخیص خودکار نمونه‌های پلاستیکی تقویت‌شده با فیبر کربن غیرمستقیم را گزارش می‌کند. سری‌های زمانی، شامل سری‌های دما-زمان خام و سیگنال‌های توالی به‌دست‌آمده از اولین مشتق پس از بازسازی سیگنال ترموگرافی به ترتیب برای آموزش و آزمایش مدل‌ها استفاده شده است. این مدل پیشنهاد شده در مدیریت اطلاعات وابسته به زمان دقیق‌تر است.

<sup>1</sup> object-based image analysis

<sup>2</sup> digital imaging and communications in medicine

<sup>3</sup> region proposal network

در پژوهش (Ketab et al, 2023) معماری یادگیری عمیق موازی با فیلترهای سفارشی و قابل یادگیری برای تشخیص چهره با وضوح پایین ارائه شده است. این مقاله وظیفه تشخیص چهره با وضوح پایین را با ترکیب تکنیک‌های نمونه برای استخراج ویژگی‌های متمایز حل می‌کند و از ویژگی‌هایی استفاده می‌کند که توسط فیلترهای سفارشی و قابل یادگیری آموخته شده و در فرآیند آموزش تزریق می‌شود تا آنها را بهتر با عملکرد مغز انسان تطبیق دهد. هدف تبدیل گابور این است که یک تصویر چهره را با استفاده از طیف وسیعی از ضرایب فیلتر گابور در مقیاس‌ها و جهت‌های مختلف در هم پیچد که منجر به ویژگی‌های تغییر مقیاس و چرخش می‌شود. معماری متناسب با جریان باقیمانده با هدف افزایش نمایش عملکردی و جلوگیری از تأثیر شیب موتور پیش‌بینی بر نقشه عملکردی شبکه ستون فقرات است.

در پژوهش (Zhong et al, 2023) شبکه موازی خود محدود برای تشخیص شی سبک وزن مبتنی بر نقطه کلید به نام SCPNet<sup>۱</sup> ارائه شده است که استنتاج را سرعت می‌بخشد، پارامترها را کاهش می‌دهد، زمینه‌های دریافتی را گسترده می‌کند و پیش‌بینی را دقیق می‌کند. به طور خاص، ماژول همجوشی چند مقیاسی موازی PMFM<sup>۲</sup> با بلوک‌های مخلوط شده موازی ساختار موازی را برای به دست آوردن ویژگی‌های قابل اعتماد و کاهش عمق اتخاذ می‌کند، از تطبیق چند مقیاسی مکرر برای جلوگیری از شاخه‌های موازی زیاد استفاده می‌کند. ماژول تشخیص خود محدود SCDM<sup>۳</sup> دارای ساختار دو شاخه‌ای است که یک شاخه گوشه‌ها را پیش‌بینی می‌کند و از آن استفاده می‌کند. برای مطابقت با جفت‌های گوشه‌ای با کیفیت بالا، و شاخه‌های دیگر نقاط کلیدی مرکز را پیش‌بینی می‌کنند. فواصل بین مراکز هندسی گوشه‌های زوج و نقاط کلیدی مرکزی برای تشخیص خود محدود استفاده می‌شود.

در پژوهش (Yin et al, 2023) تشخیص افسردگی در گفتار با استفاده از مبدل و شبکه‌های عصبی کانولوشنال موازی انجام شده است. بطور کلی یک مدل یادگیری عمیق مبتنی بر یک شبکه عصبی کانولوشنال موازی و یک مبدل به منظور استخراج اطلاعات موثر با پیچیدگی قابل قبول پیشنهاد می‌شود. رویکرد پیشنهادی شامل یک ماژول شبکه عصبی کانولوشنال موازی است که برای تمرکز بر دانش محلی استفاده می‌شود، در حالی که یک ماژول مبدل به عنوان جریان موازی دیگر برای درک اطلاعات متوالی زمانی با استفاده از مکانیسم‌های توجه خطی با توابع هسته استفاده می‌شود.

در پژوهش (Wang et al, 2023) برنامه ریزی نمای با کارایی بالا برای بازرسی سطح براساس یادگیری تقویت عمیق موازی ارائه شده است. ابتدا، روش تخمین دید توسعه داده شده است تا مناطق قابل مشاهده را بتوان به سرعت برای یک دیدگاه مشخص شناسایی کرد. سپس، یک تابع پاداش جدید طراحی می‌شود و از روش مزیت همزمان بازیگر-نقد برای حل مشکل برنامه‌ریزی دیدگاه استفاده می‌شود. روش پیشنهادی می‌تواند به طور بالقوه برای سایر وظایف مشابه مبتنی بر دید نیز اعمال شود.

### ۳. روش تحقیق

با توجه به ماهیت مسائل دسته‌بندی تصاویر در چند دامنه که در آن فرآیند آموزش و تست بر روی دو دامنه متفاوت انجام می‌گیرد، الگوریتم‌های سنتی دسته‌بندی نمی‌توانند در آن نتایج بهینه‌ای را بدست آورند. در این مقاله به دنبال افزایش قابلیت تعمیم و صحت دسته‌بندی، استفاده از مدل‌های عمیق از پیش آموزش دیده برای استخراج ویژگی و همچنین استفاده از مجموعه داده بدون برچسب متفاوت از دامنه هدف را در فرآیند آموزش پیشنهاد می‌کنیم و برای انجام دسته‌بندی از SLFN ترکیبی استفاده می‌کنیم. دسته‌بندی پیشنهادی دارای عبارت جدید در تابع هزینه است که برای مسائل چند دامنه‌ای مناسب می‌باشد. عبارتی که مربوط به خطای داده‌های برچسب دار از دامنه هدف می‌باشد.

<sup>1</sup> self-constrained parallelism keypoint-based lightweight object detection network

<sup>2</sup> parallel multi-scale fusion module

<sup>3</sup> self-constrained detection module



یکی از مهمترین نوآوری‌های مطرح شده در SLFN ترکیبی و تطبیقی پیشنهاد شده است که روش مطرح شده در این تحقیق، گسترش یافته‌ی روش ساده SLFN می‌باشد به نحوی که از حالت یک دامنه‌ای به حالت یادگیری چند دامنه‌ای توسعه پیدا می‌کند.

نوآوری دیگری که در این تحقیق معرفی یک عبارت خطای جدید برای تابع ارزیابی دسته‌بندی است. این عبارت خطا دارای دو بخش است، یک بخش مربوط به خطای حاصل از داده‌های برچسب دار دامنه هدف است و یک بخش هم مربوط به خطای حاصل از داده‌های بدون برچسب دامنه هدف می‌باشد.

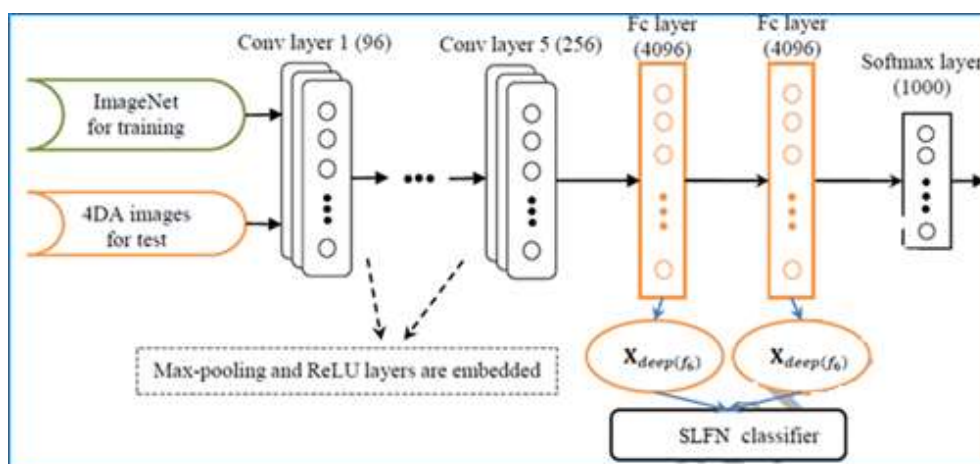
برای دسته‌بندی داده‌های بدون برچسب از دامنه هدف لازم است به نوعی یادگیری بدون نظارت انجام شود. برای رسیدن به این هدف گراف لاپلاسی مبتنی بر یک عبارت تنظیم چندانکه نیز در تابع هدف به کار گرفته می‌شود. این گراف با تحقق هدف استفاده از یادگیری نیمه نظارت شده در مساله دسته‌بندی چند دامنه‌ای به کار می‌رود.

نوآوری دیگر این تحقیق استفاده از داده‌های بدون برچسب از دامنه‌های متفاوت در مرحله آموزش است. به این ترتیب که داده‌ها از یک دامنه برای مرحله آموزش استفاده می‌شوند و در یک دامنه دیگر برای تست استفاده می‌شوند. از سه مجموعه داده بدون برچسب برای سه مدل مجزا استفاده می‌شود که هر یک از یک مدل عمیق برای استخراج ویژگی استفاده می‌کنند.

همچنین نوآوری دیگر تعریف یک مدل موازی از مدل‌های عمیق در مرحله استخراج ویژگی است. از سه مدل عمیق از پیش آموزش دیده مبتنی بر شبکه‌های عصبی کانولوشن برای استخراج ویژگی‌های عمیق و همچنین از سه مجموعه داده بدون برچسب متفاوت از دامنه هدف در فرآیند آموزش استفاده شده است. در واقع ویژگی‌های استخراج شده نه تنها مربوط به داده‌های برچسب دار هستند بلکه برای تعدادی از داده‌ها که بدون برچسب نیز هستند به وسیله مدل‌های عمیق ویژگی‌ها استخراج می‌شود و به عنوان ورودی به دسته‌بندی SLFN وارد می‌شوند و دسته‌بندی روی آنها انجام می‌گیرد. در انتها برای ترکیب نتایج این مدل‌های عمیق و دسته‌بندی از روش میانگین گیری ساده استفاده می‌کنیم.

### ۳-۱ معماری مدل پیشنهادی

طرح به کارگیری مدل‌های عمیق از پیش آموزش داده برای مرحله استخراج ویژگی و استفاده از دسته‌بندی SLFN ترکیبی بر اساس ویژگی‌های استخراج شده در شکل ۱ نشان داده شده است. این شکل برای یک مدل ترسیم شده است که برای تمام مدل‌ها تعمیم داده شده است.

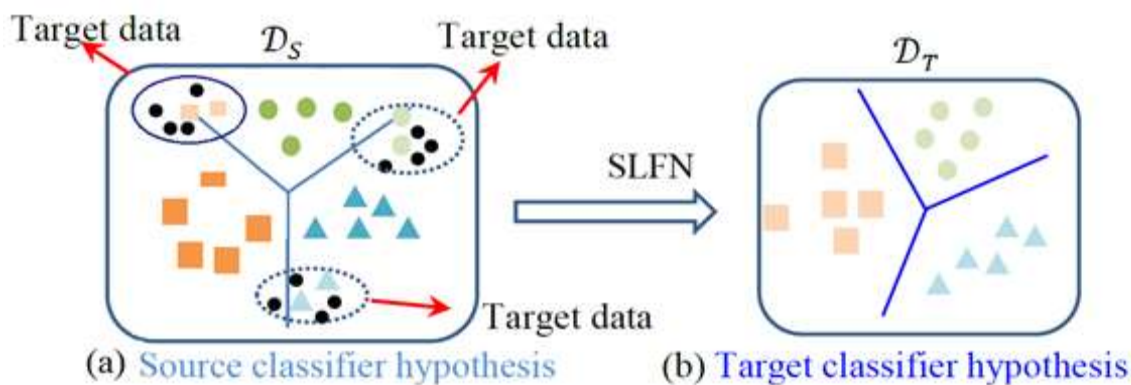


شکل ۱: به کارگیری مدل‌های عمیق برای استخراج ویژگی

همانطور که در شکل ۱ نشان داده شده است ویژگی‌ها توسط هر یک از مدل‌های عمیق از پیش آموزش دیده AlexNet، VGG16 و VGG19 استخراج می‌شود و به واسطه هر یک از لایه‌های fully-connected می‌تواند دریافت شده و در اختیار دسته‌بندی مبتنی بر شبکه عصبی تک لایه رو به جلو ترکیبی ارائه شده قرار گیرد.

### ۳-۱-۱-۳ دسته‌بندی به کمک SLFN ترکیبی

طرح کلی شبکه SLFN ترکیبی ارائه شده در این مقاله، در شکل ۲ نمایش داده شده است، لازم به ذکر است که این طرح نسبت به طرح‌های ارائه شده در کارهای پیشین تفاوت‌هایی دارد که مهمترین آنها به کارگیری SLFN ارائه شده به جای ELM است:



شکل ۲: SLFN ترکیبی

در قسمت a شکل ۲ دایره‌های مشکی رنگ داده‌های بدون برچسب از دامنه هدف را نشان می‌دهند و اشکال با رنگ روشن داده‌های برچسب دار از دامنه هدف هستند. داده‌هایی که با رنگ‌های مختلف تیره در شکل مشخص شده اند بقیه داده‌های آموزشی هستند که برچسب دارند و از دامنه منبع انتخاب شده اند. در قسمت b از شکل ۲ نیز دسته‌بندی تطبیق یافته با مساله چند دامنه‌ای را نشان می‌دهد که می‌تواند داده‌های برچسب دار موجود در چند دامنه هدف را به درستی دسته‌بندی کند.

منظور از دامنه منبع، مجموعه داده اصلی که است که در واقع فرآیند آموزش بر روی آن داده‌ها انجام می‌شود و منظور از دامنه هدف نیز مجموعه‌ای از داده‌هاست که فرایند ارزیابی و تست بر روی آن انجام می‌شود. با به کارگیری چندین نمونه برچسب دار و بدون برچسب در فرآیند آموزش و اضافه کردن عبارت خطای جدید به تابع هدف دسته بند، دسته‌بندی شبکه عصبی تک لایه رو به جلو ترکیبی و تطبیق یافته با مساله چند دامنه‌ای ارائه می‌گردد.

در SLFN ترکیبی ارائه شده در این مقاله مفاهیمی داریم که جهت درک بهتر عملکرد آن در ادامه به آنها اشاره خواهد شد:

پارامترهای S و T به دامنه‌های منبع و هدف اشاره دارند. پارامترهای X شامل  $X_S = [x_s^1, \dots, x_s^{N_s}] \in R^{d \times N_s}$  و  $X_T = [x_t^1, \dots, x_t^{N_T}] \in R^{d \times N_T}$  هستند به ترتیب به مجموعه داده‌های آموزشی منبع و مجموعه داده‌های هدف برچسب دار اشاره می‌کنند. پارامترهای با عنوان Y که شامل دو مورد  $Y_S = [y_s^1, \dots, y_s^{N_s}] \in R^{C \times N_s}$  و  $Y_T = [y_t^1, \dots, y_t^{N_T}] \in R^{C \times N_T}$  هستند نیز اشاره به ماتریس برچسب دو مجموعه فوق دارند. همچنین پارامترهای N شامل  $N_S$  و  $N_T$  هستند به ترتیب تعداد داده‌های مجموعه آموزشی منبع و

تعداد داده‌های برچسب دار از مجموعه داده‌های هدف هستند و پارامترهای  $C$  و  $d$  نیز تعداد کلاس‌ها و ابعاد آن می‌باشند. همچنین پارامتر  $X_{II}$  را به عنوان ماتریس داده بدون برچسب از مجموعه داده‌های انتخاب شده از دامنه هدف تعریف می‌کنیم.

لازم به ذکر است که به صورت پیش فرض تابع هزینه در شبکه عصبی تک لایه رو به جلو معمولی برای الگوریتم‌های یادگیری یک دامنه‌ای می‌باشد. در این مقاله برای تطبیق این شبکه با مسائل چند دامنه‌ای، تابع هزینه جدیدی تعریف بشود که خطای مربوط به داده‌های برچسب دار و بدون برچسب دامنه هدف را پوشش بدهد. بدین منظور تابع هزینه دارای عبارات خطای جدیدی می‌باشد که این عبارات به شکل زیر تعریف می‌شوند:

$$C = \gamma.E_s(\theta) + \alpha.E_T(\theta) + \Omega(\theta) \quad (1)$$

در تابع هزینه فوق، اولین عبارت به خطای حاصل از داده‌های آموزشی دامنه منبع اشاره می‌کند. دومین عبارت نیز نشان دهنده خطای داده‌های برچسب دار انتخاب شده از دامنه هدف می‌باشد. سومین عبارت نیز که با نماد  $\Omega(\theta)$  نشان داده شده است، میزان خطای حاصل از داده‌های بدون برچسب انتخاب شده از دامنه هدف می‌باشد. پارامترهای  $\gamma$  و  $\alpha$  نیز ضرایب وزن دهی بین پارامترهای ذکر شده هستند که به صورت تجربی مقادیر ثابتی برای آنها در نظر گرفته می‌شود.

برای محاسبه بخش اول عبارت خطا یعنی عبارت  $E_s(\theta)$  لازم است که خطای حاصل از داده‌های آموزشی یک دامنه‌ای محاسبه شود. این دامنه می‌تواند هر یک از دامنه‌های مربوط به منبع یا دامنه هدف باشد. بنابراین به ترتیب زیر محاسبه خواهد شد:

$$E_s(\theta) = \frac{1}{2} \sum_{i=1}^{N_s} \|e_s^i\|_2^2, (e_s^i)^T = h(x_s^i)\theta - (y_s^i)^T, i=1, \dots, N_s \quad (2)$$

در رابطه فوق عبارت  $e_s$  نشان دهنده بردار خطای داده‌های دامنه منبع است، پارامتر  $\theta$  نیز نشان دهنده وزن‌های بین لایه مخفی و خروجی است. همچنین پارامتر  $y_s$  نیز خروجی مطلوب نمونه‌های آموزشی از دامنه منبع می‌باشد.

با فرض وجود تعداد کمی داده برچسب دار از دامنه هدف، عبارت خطای  $E_T(\theta)$  به صورت زیر تعریف می‌کنیم:

$$E_T(\theta) = \frac{1}{2} \sum_{j=1}^{N_T} \|e_T^j\|_2^2, (e_T^j)^T = h(x_T^j)\theta - (y_T^j)^T, j=1, \dots, N_T \quad (3)$$

در رابطه فوق عبارت  $e_T$  نشان دهنده بردار خطای داده‌های دامنه هدف است، پارامتر  $\theta$  نیز بیان کننده وزن‌های بین لایه مخفی و لایه خروجی است. همچنین پارامتر  $y_T$  نشان دهنده خروجی مطلوب نمونه‌های آموزشی از دامنه هدف می‌باشد.

جهت افزایش قدرت دسته‌بندی، علاوه بر داده‌های برچسب دار از دامنه هدف، وجود تعدادی داده بدون برچسب از دامنه هدف نیز در تشخیص چند دامنه‌ای می‌تواند مفید باشد. بنابراین بر اساس یادگیری چند گانه گراف بدون نظارت، عبارت  $\Omega(\theta)$  را بر اساس گراف لاپلاسی به گونه‌ای استفاده می‌کنیم که اطلاعات تعدادی از داده‌های بدون برچسب از دامنه جدید می‌تواند به خوبی مورد استفاده قرار گیرد (Zeiler et al, 2014). لازم به ذکر است که منظور از دامنه جدید می‌تواند داده موجود در دامنه هدف باشد. در این روش اگر دو نقطه داده به یکدیگر نزدیک باشند، برچسب‌های پیش بینی شده آنها نیز باید به یکدیگر نزدیک باشند. این عبارت در واقع بیان کننده ماهیت خوشه بندی داده‌ها و یادگیری بدون نظارت است. در حالت کلی عبارت  $\Omega(\theta)$  مبتنی بر گراف لاپلاسی به صورت زیر تعریف می‌شود:



$$\Omega(\theta) = F^T L F \quad (4)$$

$$F = h(X_U) \theta \quad (5)$$

در روابط فوق پارامتر  $F$  نشان دهنده ماتریس برچسب پیش بینی شده داده‌های بدون برچسب از دامنه هدف است. پارامتر  $L$  نیز ماتریس لاپلاسی است که به صورت  $L=D-W$  محاسبه می‌شود که در این محاسبه  $D$  ماتریس قطری بوده که دارای ورودی  $D_{ii} = \sum_j W_{i,j}$  است. ورودی ماتریس  $W$  نیز به صورت زیر محاسبه می‌شود:

$$W_{i,j} = \begin{cases} 1, & \text{if } x_i \in N_k(x_j) \text{ or } x_j \in N_k(x_i) \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

که در آن  $N_k(x_i)$ ، به  $k$  نزدیکترین همسایه  $x_i$  اشاره دارد. بنابراین داریم:

$$\Omega(\theta) = \frac{1}{2} \theta^T h(X_u)^T L h(X_u) \theta \quad (7)$$

بنابراین تابع هزینه در روش SLFN ترکیبی ارائه شده به صورت زیر می‌باشد:

$$C = \lambda \cdot \frac{1}{2} \sum_{i=1}^{N_s} \|e_s^i\|_2^2 + \mu \cdot \frac{1}{2} \sum_{j=1}^{N_T} \|e_T^j\|_2^2 + \frac{1}{2} \theta^T h(X_u)^T L h(X_u) \theta \quad (8)$$

فرآیند کلی SLFN ترکیبی در شکل ۳ نمایش داده شده است:



شکل ۳: روند کلی SLFN ترکیبی

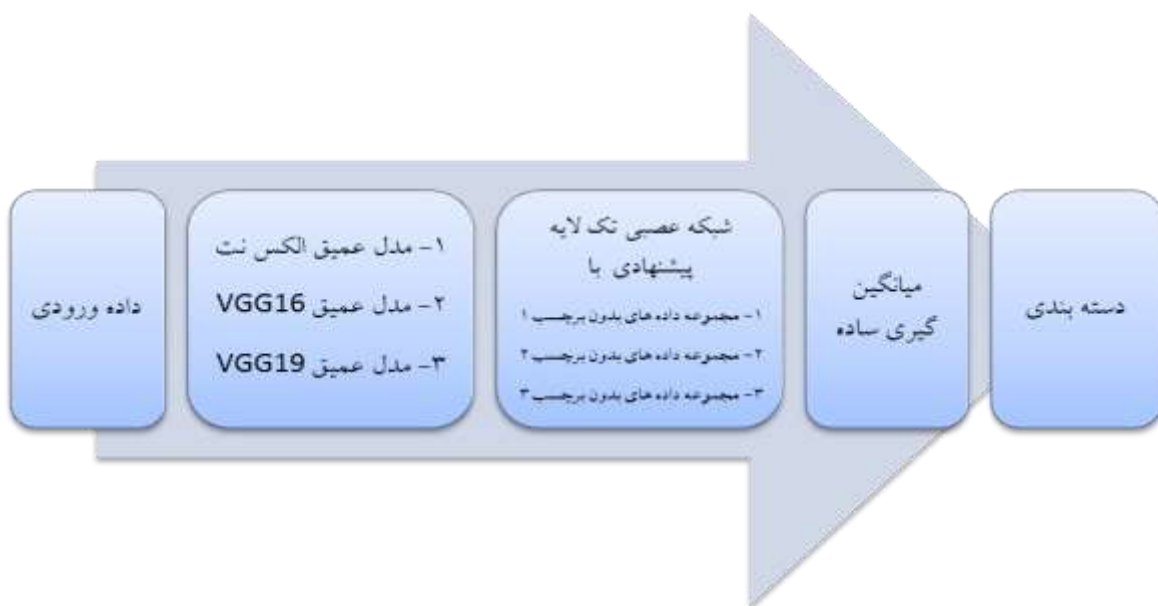
### ۳-۱-۲ روش میانگین گیری ساده

روش های متعددی برای ترکیب نتایج حاصل از دسته بندی بر پایه مدل های عمیق وجود دارد. یکی از مرسوم ترین روش ها برای ترکیب نتایج چند مدل استفاده از متد میانگین گیری ساده<sup>۱</sup> است. برای این روش لازم است هر داده ورودی به هر یک از مدل های عمیق داده شود و خروجی حاصل از آنها که در واقع ویژگی های استخراج شده هستند و در نهایت نتایج حاصل ترکیب می شود و به دسته بندی مبتنی بر شبکه عصبی تک لایه رو به جلو داده شود. این عملیات به گونه ای است که کلاس پیش بینی شده حاصل از ترکیب نتایج با استفاده از رابطه زیر بدست می آید (Zhang et al, 2016):

$$m = \arg \max_{1 \leq i \leq M} \left\{ \frac{1}{N} \sum_{j=1}^N p(y=i | c_j) \right\} \quad (9)$$

در رابطه ذکر شده منظور از  $m$  یک کلاس پیش بینی شده است. در این رابطه پارامتر  $M$  نشان دهنده تعداد کلاس ها و پارامتر  $N$  نشان دهنده تعداد دسته بندها می باشد. همچنین عبارت احتمال شرطی  $p(y=i | c_j)$  نشان دهنده مقدار احتمال پیش بینی شده کلاس شماره  $i$  توسط دسته بندی شماره  $j$  می باشد.

نمودار کلی دسته بندی تصویر ورودی به وسیله چند مدل عمیق کانولوشن از پیش آموزش دیده و همچنین دسته بندی SLFN ترکیبی ارائه شده، در شکل ۴ نمایش داده شده است:



شکل ۴: دسته بندی تصویر در طرح ارائه شده

تفسیر نمودار فوق به این ترتیب است که پس از دریافت داده ورودی توسط مدل عمیق ویژگی های داده استخراج شده و به شبکه عصبی تک لایه آموزش دیده شده با داده های بدون برچسب متفاوت داده می شود و پس از آن خروجی مدل ها میانگین گیری ساده شده و نتیجه دسته بندی اعلام می شود.

به کارگیری داده های بدون برچسب متفاوت و متنوع از دامنه هدف برای هر یک از مدل های عمیق و دسته بندها در طی فرآیند آموزش باعث بالا رفتن قابلیت تعمیم مدل در مساله دسته بندی چند دامنه ای می گردد.

<sup>1</sup> Simple Average

#### ۴. یافته‌ها

در این مقاله از مدل‌های عمیق از پیش آموزش دیده با هدف استخراج ویژگی‌های عمیق از تصاویر استفاده شده تا از ویژگی‌های استخراج شده برای دسته‌بندی تصاویر به کمک شبکه عصبی تک لایه رو به جلو استفاده می‌کنیم. همانطور که ذکر شده از سه مدل عمیق الکسنت، وی جی جی ۱۶ و ۱۹ برای استخراج ویژگی استفاده می‌گردد. خروجی مدل‌های عمیق از لایه fully-connected به عنوان ورودی به دسته‌بندی SLFN معرفی شده وارد می‌شود و فرایند دسته‌بندی به کمک آن انجام می‌گیرد. ابعاد ویژگی‌های استخراج شده از تصاویر توسط هر یک از مدل‌های عمیق ۴۰۹۶ می‌باشد.

همچنین توانایی دسته‌بندی با استفاده از ویژگی‌های عمیق نیز بررسی می‌شود. ویژگی‌های بدست آمده برای آموزش دسته‌بندی شبکه عصبی تک لایه رو به جلو به کار گرفته شده و نتایج حاصل نیز از نظر صحت دسته‌بندی بر روی چند دامنه بررسی و مقایسه می‌شود.

#### ۴-۱ تنظیمات شبیه سازی

مجموعه داده‌های 4DA<sup>۱</sup> معمولاً برای ارزیابی کاربردهای تطبیق دامنه‌ها و انتقال دانش به کار گرفته می‌شود. به همین دلیل در این مقاله برای ارزیابی نتایج طرح پیشنهادی از این مجموعه داده استفاده شده است. این مجموعه داده شامل چهار دامنه از جمله: کلتک ۲۵۶<sup>۲</sup>، آمازون<sup>۳</sup>، وب کم<sup>۴</sup> و دی اس ال آر<sup>۵</sup> است. لازم به ذکر است که نمونه‌های این مجموعه داده‌ها از منابع مختلف که شامل ده کلاس شیء می‌باشند انتخاب شده است (Zeiler et al, 2014).

جزئیات مجموعه داده 4DA با ویژگی‌های عمیق در جدول ۱ نمایش داده شده است.

جدول ۱: جزئیات مجموعه داده 4DA CNN

مجموعه داده	تعداد کلاس	تعداد ابعاد	تعداد نمونه‌ها	$n_s/c$	$n_t/c$
Amazon	۱۰	۴۰۹۶	۹۵۸	۲۰	۳
Dslr	۱۰	۴۰۹۶	۱۵۷	۸	۳
Webcam	۱۰	۴۰۹۶	۲۹۵	۸	۳
Caltech	۱۰	۴۰۹۶	۱۱۲۳	۸	۳

در جدول ۱ پارامتر  $(n_s/c)$  بیان کننده تعداد نمونه‌ها به ازاء هر کلاس از دامنه منبع است و پارامتر  $(n_t/c)$  نشان دهنده تعداد نمونه‌ها به ازاء هر کلاس از دامنه هدف می‌باشد. نمونه‌های ذکر شده نمونه‌هایی هستند که در فرآیند آموزش به کار گرفته می‌شوند.

<sup>۱</sup> Domain Adaptation

<sup>۲</sup> Caltech 256(c)

<sup>۳</sup> Amazon (A)

<sup>۴</sup> Webcam (W)

<sup>۵</sup> Dslr (D)

## ۴-۲ معیارهای عملکرد

برای بررسی عملکرد سیستم لازم است پیکربندی‌های متفاوتی ایجاد و بر روی هر یک ارزیابی انجام شود. این پیکربندی‌ها شامل تشخیص یک دامنه ای، تشخیص چند دامنه‌ای تنها بر روی دامنه منبع و تشخیص چند دامنه‌ای بر روی هر دو دامنه منبع و هدف هستند. جزئیات هر پیکربندی در ادامه ذکر می‌شود.

در تشخیص تک دامنه‌ای فرآیند آموزش و تست بر روی یک دامنه انجام می‌گیرد. به عنوان مثال آموزش مدل بر روی داده‌های آمازون مطابق تعداد مشخص شده نمونه‌ها از هر کلاس از دامنه منبع انجام می‌شود و فرآیند تست نیز بر روی داده‌های باقیمانده از همین دامنه انجام می‌گیرد. نتایج میانگین صحت دسته‌بندی حاصل از ۲۰ بار اجرای آموزش و تست بر روی هر دامنه گزارش شده است.

در تشخیص چند دامنه‌ای که تنها از دامنه منبع استفاده می‌کند، فرآیند آموزش بر روی یک دامنه و تست بر روی دامنه‌ای دیگر انجام می‌گیرد. به عنوان مثال آموزش بر روی آمازون و تست بر روی دی اس ال آر انجام می‌گیرد. با توجه به تعداد دامنه‌ها، تشخیص چند دامنه‌ای در ۱۲ حالت مختلف می‌تواند انجام شود. نتایج میانگین صحت دسته‌بندی حاصل از ۲۰ بار اجرای آموزش و تست برای هر یک از حالات گزارش می‌شود. نکته‌ای که در این حالت باید به آن توجه شود این است که آموزش فقط بر روی دامنه منبع انجام می‌شود و از دامنه هدف برای فرآیند آموزش استفاده‌ای نمی‌شود.

در تشخیص چند دامنه‌ای که هر دو دامنه منبع و هدف مشارکت داشته باشند مشابه به حالت قبلی، فرآیند آموزش بر روی یک دامنه و تست بر روی دامنه‌ای دیگر انجام می‌گیرد. به عنوان مثال آموزش بر روی آمازون است و تست بر روی دی اس ال آر انجام می‌گیرد. با توجه به تعداد دامنه‌ها، در این حالت نیز تشخیص چند دامنه‌ای در ۱۲ حالت مختلف می‌تواند انجام گیرد. تفاوت این حالت با حالت قبل در به کارگیری ۳ نمونه برجسب دار به ازاء هر کلاس از دامنه هدف در فرآیند آموزش می‌باشد. همچنین ۳ نمونه بدون برجسب به ازاء هر کلاس از دامنه هدف نیز در روش SLFN ترکیبی به کار برده می‌شود. نتایج میانگین صحت دسته‌بندی حاصل از ۲۰ بار اجرای آموزش و تست در هر یک از حالات گزارش می‌شود.

بر اساس پیکربندی حالت اول، میانگین صحت و انحراف معیار (عدد اول میانگین صحت و عدد دوم انحراف معیار می‌باشد) حاصل از اجرای (آموزش و تست) ۲۰ قطعه داده که به صورت تصادفی تولید شده اند برای چهار روش شامل: NN، SVM، ELM، و روش پیشنهاد شده در این تحقیق در جدول ۲ نشان داده شده است. همانطور که مشاهده می‌شود عملکرد تشخیص بر اساس ویژگی‌های عمیق به دست آمده از لایه‌های Fully Connected به روشنی با یکدیگر متفاوت است.

همانطور که مشاهده می‌شود روش SLFN عملکرد بهتری نسبت به سایر روش‌ها دارد. همچنین روش ELM نسبت به NN و SVM در تمام دامنه‌ها عملکرد بهتری را از خود نشان می‌دهد. به خصوص نتایج نشان می‌دهند که SLFN نسبت به SVM بر اساس ویژگی‌های عمیق f6 در دامنه‌های Webcam، Dslr، Amazon و Caltech به ترتیب ۰.۳٪، ۰.۱٪ و ۰.۵٪ نتایج را بهبود داده است. همچنین بر اساس ویژگی‌های عمیق f7 نیز نتایج به ترتیب ۰.۳٪، ۰.۲٪ و ۰.۶٪ بهبود پیدا کرده اند.

جدول ۲: صحت تشخیص برای همه روش‌ها در همه دامنه‌ها برای پیکربندی اول

Method	CNN layer	Amazon	DSLR	Webcam	Caltech	CNN layer	Amazon	DSLR	Webcam	Caltech
NN	f6	91.0±0.3	97.3±0.6	95.0±0.4	75.0±0.4	f7	92.4±0.2	96.8±0.5	95.3±0.5	76.2±0.5
SVM	f6	92.9±0.1	97.6±0.6	96.7±0.3	83.9±0.4	f7	93.2±0.1	96.9±0.5	96.5±0.4	83.2±0.5
ELM	f6	92.9±0.1	98.0±0.3	97.7±0.2	84.8±0.3	f7	93.6±0.1	97.2±0.4	97.4±0.3	85.0±0.3
SLFN	f6	95.3±0.3	98.8±0.2	98.7±0.2	88.3±0.4	f7	96.5±0.1	98.9±0.2	98.7±0.3	89.4±0.2

جدول ۳ و ۴ نشان دهنده نتایج به دست آمده بر اساس پیکربندی حالت دوم می‌باشد. این نتایج نشان دهنده میانگین صحت و انحراف معیار حاصل از اجرای (آموزش و تست) ۲۰ قطعه داده که به صورت تصادفی تولید شده اند، می‌باشد. در مجموع ۱۲ حالت مختلف برای تشخیص چند دامنه‌ای نشان داده شده است.

این نتایج نشان می‌دهد که صحت تشخیص با ویژگی‌های عمیق f7 نسبت به ویژگی‌های بدست آمده از لایه f6 عملکرد بهتری دارد که نشان دهنده موثر بودن ویژگی‌های عمیق تر می‌باشد، بنابراین سعی بر این است که از ویژگی‌های عمیق تر استفاده شود.

همچنین این نتایج نشان می‌دهد که عملکرد ELM به روشنی بهتر از SVM می‌باشد. میانگین ۱۲ حالت مختلف تشخیص چند دامنه‌ای نشان دهنده عملکرد بهتر روش ELM نسبت به روش SVM می‌باشد، بنابراین واضح است که انتخاب روشی مشابه ELM آن هم با ویژگی‌های بهتر نسبت به آن (مانند قابلیت افزایش نورون) حرکت درستی بوده است.

علاوه بر این موارد، نتایج نشان می‌دهد که در مسائل مشکل تر تشخیص چند دامنه‌ای روش SLFN پیشنهادی عملکرد بهتر و استحکام بیشتری دارد و دارای برتری کاملاً مشخص نسبت به سایر متدها در خروجی می‌باشد.

جدول ۳: صحت تشخیص برای پیکربندی دوم برای همه روش‌ها

Method	CNN_layer	A→D	C→D	W→D	A→C	W→C	D→C
NN	f6	71.9±0.9	72.0±1.7	92.7±0.5	76.8±0.3	56.6±0.9	64.4±0.4
	f7	78.7±0.5	75.6±1.3	96.9±0.4	77.2±0.4	66.2±0.5	70.7±0.4
SVM	f6	79.6±0.7	75.1±1.8	96.7±0.4	79.5±0.4	59.5±0.9	67.3±1.2
	f7	80.6±0.8	76.4±1.4	96.7±0.4	79.6±0.4	68.1±0.6	74.3±0.6
ELM	f6	80.6±0.6	79.5±1.2	96.7±0.2	80.4±0.3	67.2±0.5	75.6±0.5
	f7	82.3±0.5	81.2±0.7	97.0±0.4	81.8±0.3	74.0±0.3	79.5±0.2
SLFN	f6	84.7±0.2	83.7±0.4	96.2±0.4	84.4±0.2	70.3±0.6	77.9±0.2
	f7	86.8±0.3	86.1±0.5	97.5±0.3	86.6±0.4	78.0±0.3	81.9±0.2



در نتایج ذکر شده مرحله آموزش تنها بر روی داده‌های منبع انجام می‌گیرد.

جدول ۴ نشان دهنده نتایج به دست آمده بر اساس پیکربندی حالت سوم می‌باشد. این نتایج نشان دهنده میانگین صحت و انحراف معیار حاصل از اجرای (آموزش و تست) ۲۰ قطعه داده که به صورت تصادفی تولید شده اند، می‌باشد. در مجموع ۱۲ حالت مختلف برای تشخیص چند دامنه‌ای و ارزیابی آن قابل انجام است.

به کارگیری چندین نمونه برچسب دار از دامنه هدف در فرآیند آموزش باعث بهبود نتایج نسبت به حالت دوم شده است و همانطور که مشاهده می‌شود تفاوت بین روش‌های ELM و SVM کاهش یافته است. همچنین روش SLFN ارائه شده به دلیل استفاده از چندین نمونه بدون برچسب از دامنه هدف و استفاده از یادگیری نیمه نظارت شده در فرآیند آموزش و همچنین بهره‌گیری از سه مدل عمیق به صورت موازی برای استخراج ویژگی‌های عمیق و استفاده از روش میانگین‌گیری ساده، دارای برتری مشهودی نسبت به سایر روش‌ها در این حالت می‌باشد. روش SLFN به طور میانگین در تمام حالات نسبت به ELM به وضوح دارای عملکرد بهتر در دسته‌بندی می‌باشد.

جدول ۴: صحت تشخیص برای همه روش‌ها در پیکربندی حالت سوم

Method	CNN_layer	A→D	C→D	W→D	A→C	W→C	D→C
NN	f6	89.4±0.7	90.1±0.8	97.0±0.4	78.1±0.4	69.0±0.9	72.8±0.8
	f7	93.0±0.5	90.9±0.9	98.6±0.2	78.9±0.4	73.6±0.6	75.6±0.4
SVM	f6	94.5±0.4	92.9±0.8	99.1±0.2	84.0±0.3	81.7±0.5	83.0±0.3
	f7	94.0±0.6	92.7±0.8	98.9±0.2	83.4±0.4	81.2±0.4	82.7±0.4
ELM	f6	94.6±0.5	93.7±0.6	99.2±0.2	83.4±0.3	81.2±0.3	83.5±0.3
	f7	94.9±0.4	93.0±0.6	99.0±0.2	84.1±0.2	82.2±0.4	84.1±0.2
SLFN	f6	98.3±0.3	95.1±0.4	98.4±0.1	87.1±0.3	85.1±0.3	83.7±0.1
	f7	98.5±0.5	95.3±0.2	98.8±0.2	88.4±0.4	85.6±0.4	84.3±0.2

در نتایج فوق داده‌های آموزشی از هر دو دامنه منبع و هدف می‌باشند.

## ۵. بحث و نتیجه‌گیری

در این مقاله به مساله تشخیص شی در چند دامنه به کمک مدل‌های یادگیری عمیق پرداخته شده است. منظور از مسائل چند دامنه‌ای مسائلی است که فرآیند آموزش و تست بر روی دامنه‌های مختلف انجام می‌شود. دامنه منبع و دامنه هدف از نظر معنا و مفهوم مشابه یکدیگرند اما دارای توزیع‌های متفاوتی هستند. به این معنی که داده‌ها از نظر مفهومی یکی هستند اما از نظر ویژگی‌های خاص با هم متفاوتند. روش‌های کلاسیک برای دسته‌بندی تصاویر در یک دامنه جواب‌های قابل قبولی را ارائه می‌دهند اما در مسائل چند دامنه‌ای کارایی مناسبی ندارند که این به دلیل متفاوت بودن تصاویر از نظر زاویه دوربین، رزولوشن تصویر، شدت رنگ، تصویر پس زمینه و...، در مجموعه‌های آموزش و تست در مسائل چند دامنه‌ای، می‌باشد و نیاز به راه حلی برای رفع این مشکل دارد. از این رو در این مقاله برای افزایش قابلیت تعمیم و صحت دسته‌بندی در مسائل چند دامنه‌ای، دسته‌بندی شبکه تک لایه رو به جلو ترکیبی را که تطبیق یافته و مناسب برای مسائل چند دامنه‌ای است پیشنهاد شده است. این دسته‌بندی دارای یک الگوریتم ترکیبی برای آموزش همه

پارامترها و یک روش افزایشی برای دستیابی به ساختار فشرده می‌باشد که موجب می‌شود ابعاد شبکه در طول مرحله آموزش بتواند افزایش داشته باشد.

به دلیل کاربرد مفید و موثر یادگیری عمیق در استخراج ویژگی و دسته‌بندی، به طور خلاصه به بررسی روش‌های مختلف یادگیری عمیق پرداخته شد و برای بهبود عملکرد دسته‌بندی از سه مدل عمیق از پیش آموزش دیده نیز در مرحله استخراج ویژگی‌های ورودی استفاده شد. برای بررسی عملکرد طرح پیشنهادی از مجموعه داده 4DA که شامل چهار دامنه متفاوت برای مسائل تشخیص چند دامنه‌ای است استفاده شد و نتایج در سه حالت متفاوت با سایر روش‌های مرسوم در این حوزه مقایسه گردید. نتایج تجربی نشان دهنده سودمندی به کارگیری مدل‌های عمیق برای استخراج ویژگی در مسائل دسته‌بندی تصاویر به خصوص تشخیص شی چند دامنه‌ای می‌باشد. با بررسی نتایج تجربی مشخص می‌شود که با به کارگیری روش پیشنهادی در حالات مختلف ارائه شده به خصوص در مسائل دسته‌بندی چند دامنه‌ای، بهبود قابل توجهی در عملکرد دسته‌بندی نسبت به سایر روش‌های مرسوم در این حوزه به دست آمده است.

استفاده از روش‌های یادگیری ترکیبی باعث بهبود صحت در کاربردهای مختلف یادگیری ماشین می‌شود. در تحقیق حاضر نیز استفاده از این روش در دسته‌بندی چند دامنه‌ای، بهبود صحت دسته‌بندی را به دنبال داشت. اما هزینه اجرای بسیار بالای این تکنیک (به خصوص با افزایش تعداد دسته‌بندی‌ها) از نکات منفی آن است که به کارگیری آن را حتی در سیستم‌هایی با پردازنده‌های قدرتمند امروزی با مشکل رو به رو می‌کند. از این رو به کارگیری روش‌های جدید یادگیری که هزینه اجرای پایین تر و عملکردی مشابه با روش‌های معمول دارند به خصوص در موضوع دسته‌بندی چند دامنه‌ای می‌تواند مفید باشد. همچنین به عنوان ادامه این مقاله می‌توان در کنار استفاده از چندین مجموعه داده بدون برچسب متفاوت از چندین دسته‌بندی مختلف نیز در فرآیند آموزش استفاده شود تا در نتیجه دسته‌بندی که برای یک مجموعه بهتر عمل می‌کند برای آن داده استفاده شود نه اینکه یک دسته‌بندی برای تمام مجموعه‌ها به کار گرفته شود. با این فرایند می‌توان در نهایت صحت دسته‌بندی را در مسائل چند دامنه‌ای بهبود بخشید.

## منابع

- Y. Guo, et al, (2015) Deep learning for visual understanding: A review, *Neurocomputing*, vol.187, pp. 27-48.
- Glorot, Xavier, et al, (2010) "Understanding the difficulty of training deep feedforward neural networks", *Aistats*, Vol. 9, pp. 249-256.
- Geeta Kocher, Gulshan Kuma (2021) *Machine learning and deep learning methods for intrusion detection systems: recent developments and challenges* © The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature
- Mohsen Rostami, Am.irhamzeh Farajollahi & Hashem Parvin (2022) Deep learning-based face detection and recognition on drones *J Ambient Intell Human Comput* <https://doi.org/10.1007/s12652-022-03897-8>
- Xing Wu, Paweł Rózycki, and Bogdan M. Wilamowski, (2015) A Hybrid Constructive Algorithm for Single-Layer Feedforward Networks Learning, *IEEE Transaction On Neural Networks And Learning Systems*, Vol. 26, pp. 1659-1668.
- Lei Zhang, Zhenwei He, Yan Liu, (2017) Deep Object Recognition Across Domains based on Adaptive Extreme Learning Machine, *Neurocomputing*, vol.239, pp. 194-203.
- Kamyshanska, Hanna, et al, (2015) "The potential energy of an autoencoder", *IEEE transactions on pattern analysis and machine intelligence*, Vol. 37, pp. 1261-1273.
- Fischer, Asja, and Christian Igel, (2012) "An introduction to restricted Boltzmann machines." *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pp. 14-46.
- Bengio, Yoshua, et al, (2007) "Greedy layer-wise training of deep networks", *Advances in neural information processing systems*, Vol. 19, p. 153.
- Sak, Haşim, Andrew Senior, and Françoise Beaufays, (2014) "Long short-term memory recurrent neural network architectures for large scale acoustic modeling." *Fifteenth Annual Conference of the International Speech Communication Association*.
- Bruna, Joan, et al, (2013) "Invariant scattering convolution networks", *IEEE transactions on pattern analysis and machine intelligence*, Vol. 35, pp. 1872-1886.

- Dae-Yong Cho, Min-Koo Kang(2021)Human gaze-aware attentive object detection for ambient intelligence 0952-1976/The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license
- Tao Liu<sup>1</sup>, Lexie Yang, Dalton Lunga(2021) Change Detection Using Deep Learning Approach with Object-based 1 Image Analysis published by Elsevier <https://doi.org/10.1016/j.rse.2021.112308>
- Imen Jammoussi, Mounir Ben Nasr and Mohamed Chtourou(2020) Facial Expressions Recognition through Convolutional Neural Network and Extreme Learning Machine 978-1-7281-1080-6/20 ©IEEE
- LUQING LUO , ZHI-XIN YANG , LULU TANG, AND KUN ZHANG(2020) An ELM-Embedded Deep Learning Based Intelligent Recognition System for Computer Numeric Control Machine Tools Digital Object Identifier 10.1109/ACCESS.2965284
- P. Haripriya, R. Porkodi(2020) Parallel deep convolutional neural network for content based medical image retrieval © Springer-Verlag GmbH Germany, part of Springer Nature
- Sultan Daud Khan, Louai Alarabi(2021) A Unified Deep Learning Framework of Multi-scale Detectors for Geo-spatial Object Detection in High-Resolution Satellite Images in Arabian Journal for Science and Engineering
- Povilas Gudžius, Olga Kurasova, Vytenis Darulis, Ernestas Filatovas(2021) Deep learning based object recognition in multispectral satellite imagery for real time applications © The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature
- Long Sun, Jie Chen, Dazheng Feng and Mengdao Xing(2021) Parallel Ensemble Deep Learning for Real-Time Remote Sensing Video Multi-Target Detection Remote Sen 13, 4377. <https://doi.org/10.3390/rs13214377>
- Yuntao Tao, Caiqi Hu, Hai Zhang, Ahmad Osman, Clemente Ibarra-Castaneda, Qiang Fang, Stefano Sfarra, Xiaobiao Dai, Xavier Maldague & Yuxia Duan(2022)Automated Defect Detection in Non-planar Objects Using Deep Learning Algorithms J Nondestruct Eval 41, 14. <https://doi.org/10.1007/s10921-022-00845-6>
- Faris Ketab, Newlin Shebiah Russel, Arivazhagan Selvaraj & Seyed Mohamed Buhari(2023) Parallel deep learning architecture with customized and learnable filters for low-resolution face recognition Vis Comput. <https://doi.org/10.1007/s00371-022-02757-y>
- Xian Zhong, Mengdie Wang, Wenxuan Liu, Jingling Yuan, Wenxin Huang(2023)SCPNet: Self-constrained parallelism network for keypoint-based lightweight object detection, [doi.org/10.1016/j.jvcir.103719](https://doi.org/10.1016/j.jvcir.103719)
- Faming Yin, Jing Du, Xinzhou Xu and Li Zhao(2023) Depression Detection in Speech Using Transformer and Parallel Convolutional Neural Networks by Electronics, 12(2), 328; <https://doi.org/10.3390/electronics12020328>
- Yuanbin Wang, Tao Peng, Wenhui Wang, Ming Luo(2023)High-efficient view planning for surface inspection based on parallel deep reinforcement learning <https://doi.org/10.1016/j.aei.101849>
- Lei Zhang, Member and David Zhang, (2016) Visual Understanding via Multi-Feature Shared Learning with Global Consistency, IEEE Transactions on Multimedia vol. 18, pp. 247-259.
- Zeiler, Matthew D., and Rob Fergus. (2014) "Visualizing and understanding convolutional networks." European conference on computer vision. Springer, Cham, pp. 818-833.

## Presenting a Deep Parallel Object Recognition Algorithm based on Hybrid and Adaptive Single-Layer Forward Neural Networks

Hamed Khorram Far

Master student of Science in Artificial Intelligence Islamic Azad University Firdous Department of Computer Engineering

Hamid Ghaffary

Member of the academic staff of Isalami Azad University, Firdous branch

### Abstract

Recent developments in the fields of computer science and intelligent systems include the computing power of systems. In intelligent systems, deep learning is applied in different areas of machine learning, machine vision, image processing, natural language processing and data mining. Among the proposed methods in deep learning, convolutional networks are widely used for various applications. One of the important issues that is raised as a challenge in this field is the issue of identifying and classifying images in several different domains. Classification and recognition of objects in different conditions in terms of perception features such as color, light, angle and background image does not have the same performance with high recognition rate. In this article, a method based on pre-trained deep learning models is presented in the feature extraction stage. In addition to deep models, there is an attempt to use data from several different domains so that the system uses different data in the training and testing phase. Also, in order to increase the resistance of the system, the single-layer neural network method is used for the classification stage of unlabeled data along with labeled data. In this neural network, in addition to the conventional error expression, other error expressions are also used, which improves the performance of this network in the classification stage. The main goal of this paper is to present a deep parallel object recognition algorithm based on hybrid and adaptive single-layer forward neural networks. The results of this method were combined with three different deep models to improve the overall performance of the system. The results obtained on the data set used in this article show an improvement of more than one percent in the performance of deep object classification.

**Keywords:** Single layer feed foeward neural network, object recognition, deep learning, multi domain object detection, convolutional neural network.s