بسم اللہ الرحمن الرحیم

وقل رب زدنی علماً

باسمه تعالی

دانشگاه صنعتی مالک اشتر

مجتمع دانشگاهی برق و کامپیوتر

سمینار کارشناسی‌ارشد رشته مهندسی کامپیوتر- هوش مصنوعی

عنوان

تطبیق دامنه در شبکه های عصبی عمیق برای کاربرد تشخیص اشیا در هنگام استفاده از داده های آموزشی مصنوعی

توسط

هلیا محمدی

استاد راهنما

دکتر محمدعلی کیوان‌راد

شهریور ماه ۱۴۰۲

فهرست مطالب

عنوان **صفحه**

فهرست مطالب ..... ج

فهرست جدول‌ها ..... ه

فهرست شکل‌ها ۶

فصل ۱: مقدمه و کلیات ۱۲

۱-۱. مقدمه ۱۲

۱-۲. ضرورت انجام پژوهش ۱۳

۱-۳. انگیزه و هدف ۱۳

۱-۴. ساختار سمینار ۱۴

فصل ۲: مبانی نظری تحقیق ۱۵

۲-۱. مقدمه ۱۵

۲-۲. تشخیص اشیا ۱۵

۲-۲-۱. مدل‌های تشخیص دهنده یک مرحله‌ای ۱۷

۲-۲-۲. مدل‌های تشخیص دهنده دو مرحله‌ای ۲۱

۲-۳. تطبیق دامنه در یادگیری عمیق ۲۴

۲-۳-۱. روش‌های مبتنی بر اختلاف ۲۵

۲-۳-۲. روشهای خصمانه ۲۷

۲-۳-۳. تطابق چند دامنه ای ۳۲

۲-۳-۴. روشهای ترکیبی ۳۴

۲-۳-۵. تطبیق دامنه چند حالته ۳۴

۲-۳-۶. روشهای نوظهور ۳۷

فصل ۳: مبانی تجربی تحقیق ۳۹

۳-۱. مقدمه ۳۹

۳-۲. تطبیق دامنه برای تشخیص اشیا ۳۹

۳-۲-۱. مقالات ۳۹

3-2-2. پایاننامهها ۵۴

۳-۳. معیارهای ارزیابی ۶۰

۱-۳-۳. معیارهای Precision و recall ۶۱
3-3-2. معیار IoU ۶۲
۳-۳-۳. معیار mAP ۶۲
3-4. مجموعه دادگان ۶۳
۱-۴-۳. مجموعه داده Cityscapes ۶۳
۲-۴-۳. مجموعه داده Foggy Cityscapes ۶۴
۳-۴-۳. مجموعه داده COCO ۶۵
۴-۴-۳. مجموعه داده Caltech ۶۶
۵-۴-۳. مجموعه داده SYNTHIA ۶۷
3-4-6. مجموعه داده KITTI ۶۸
3-4-7. مجموعه داده Pascal VOC ۶۹
۸-۴-۳. مجموعه داده SIM10K ۷۱
۹-۴-۳. مجموعه داده BDD100K ۷۱
۱۰-۴-۳. مجموعه داده YTBB ۷۲
۵-۳. نتایج ۷۴
فصل 4 نتیجه‌گیری و پیشنهادها ۷۶
۱-۴. نتیجه‌گیری ۷۶
۲-۴. پیشنهاد کارهای آینده ۷۷
مراجع ۷۹

فهرست جدول‌ها

شماره جدول	عنوان جدول	صفحه

جدول ۱ نتایج عملی تطبیق دامنه برای تشخیص اشیا در مقالات متناظر ۷۵

فهرست شکل‌ها

شماره شکل	عنوان شکل	صفحه

شکل ۱ نمونه ای از کار تشخیص اشیا ۱۶

شکل ۲ مثالی از ساختار یک بلوک residual learning ۱۷

شکل ۳ معماری شبکه YOLO[5] ۱۸

شکل ۴ ساختار هرمی معماری SSD[6] ۲۰

شکل ۵ معماری و ساختار FPN[7] ۲۱

شکل ۶ ساختار و نحوه عملکرد شبکه R-CNN[8] ۲۲

شکل ۷ ساختار و نحوه عملکرد شبکه Fast R-CNN[9] ۲۳

شکل ۸ ساختار و نحوه عملکرد شبکه Faster R-CNN[10] ۲۳

شکل ۹ ساختار/معماری شبکه معمولی شبکه‌های عصبی عمیق که روش‌های مبتنی بر اختلاف را برای تطبیق دامنه پیاده‌سازی می‌کنند. در این ساختار، لایه‌های استخراج ویژگی به اشتراک گذاشته می‌شوند منظم می‌شوند. لایه‌های بعدی، "لایه‌های تطبیق" شناخته می‌شوند که شامل معیارهای اختلاف یا نمایشی از شبکه به همراه متریک اختلاف است. لایه‌های خاص وظیفه نقش زیادی در تطبیق ویژگی‌های منبع و هدف ندارند[14]. ۲۶

شکل ۱۰ ساختار/معماری شبکه‌های عصبی عمیق که روش‌های مبتنی بر خصومت را برای تطبیق دامنه پیاده‌سازی می‌کنند. مولد(ها) اختیاری هستند و در صورت نیاز فقط برای ایجاد داده‌های مصنوعی استفاده می‌شوند. متمایزکننده دامنه، به دنبال تمایز (طبقه بندی) دامنه های منبع و هدف است، در حالی که وظیفه خاص (مثلاً طبقه بندی یا تشخیص شی) توسط شبکه خاص وظیفه انجام می‌شود[14]. ۲۹

شکل ۱۱ شبکه عصبی متخاصم دامنه‌ای (DANN) دو شبکه را همزمان آموزش می‌دهد. DANN یک استخراج‌کننده ویژگی (شبکه سبز) و پیش‌بینی‌کننده کلاس/برچسب (شبکه آبی) را روی داده منبع آموزش می‌دهد. DANN همچنین یک استخراج‌کننده ویژگی (شبکه سبز) و طبقه‌بندی‌کننده دامنه (شبکه صورتی) را بر روی داده منبع و هدف آموزش می‌دهد. لایه تغییر گرادیان (GRL) به شبکه پیش‌روی اجازه می‌دهد تا به عنوان معمول پیش برود؛ با این حال، در زمان پس‌انتشار، گرادیان از تمییز دامنه تغییر می‌کند (برعکس می‌شود اما با یک مقدار منفی ضرب می‌شود) که منجر به درک ویژگی‌های ثابت دامنه (سردرگمی دامنه) توسط استخراج‌کننده ویژگی (شبکه سبز) می‌شود. پارامتر λ به یادگیری ویژگی‌های دسته‌بندی کمک می‌کند و به تدریج ویژگی‌های دامنه را یاد می‌گیرد[14]. ۳۰

شکل ۱۲ ساختار/معماری شبکه معمولی برای شبکه‌های عصبی عمیق که اجرای تطبیق دامنه چند منظوره همگن را انجام می‌دهند. در بلوک درون دامنه ای، تمیزدهنده‌ها ویژگی بردارها را مجبور می‌کنند تا ویژگی‌های بی‌تغییر در دامنه را یاد بگیرند. اما دو بلوک بین حوزه ای و بین دامنه ای اختیاری هستند و در هر تنظیم تطبیق دامنه چند منظوره ممکن است مشاهده نشوند[14].....۳۶

شکل ۱۳ ساختار روش Faster-RCNN Robust. این شبکه از مدل آموزش دیده در منبع برای به دست آوردن برچسب‌های نویزی در مجموعه داده دامنه مقصد استفاده می‌کند. یک ماژول دسته‌بندی برای به‌روزرسانی برچسب‌های نویزی استفاده می‌شود و کیفیت آنها را بهبود می‌بخشد. در نهایت، مدل تشخیص بر روی داده‌های دامنه مقصد با کمک یک تابع ضرر قوی آموزش می‌بیند که عدم قطعیت را جهت در نظر گرفتن احتمال اشتباهات ممکن در برچسب‌های نویزی شبه حاصل از مدل آموزش دیده در منبع می‌گیرد[41]..... ۴۰

شکل ۱۴ مقایسه کیفی روش فوق با Faster R-CNN در آزمایش Cityscapes → KITTI[41] ۴۰

شکل ۱۵ مروری بر معماری "Mean Teacher with Object Relations (MTOR)" برای تشخیص اشیاء در دامنه‌های مختلف، با مدل‌های معلم و دانش‌آموز تحت یک backbone مشترک از Faster R-CNN[42]..... ۴۱

شکل ۱۶ معماری شبکه پیشنهادی. این روش با استفاده از یک شبکه تشخیص دهنده دامنه محلی و یک شبکه تشخیص دهنده دامنه جهانی تطبیق قوی-محلی انجام می‌دهد و تطبیق ضعیف-جهانی را به انجام می‌رساند. بردار متناظر از طریق تشخیص دهنده دامنه استخراج می‌شود و در لایه قبل از لایه کاملاً متصل نهایی ادغام می‌شود[43]..... ۴۲

شکل ۱۷ ساختار شبکه MAF . این شبکه الهام گرفته از Faster-RCNN مبتنی بر VGG16 است، MAF ماژول‌های تطبیق ویژگی را هم بر روی ویژگی‌های دامنه و هم بر روی ویژگی‌های پیشنهاد اعمال می‌کند. برای ماژول تطبیق ویژگی دامنه سلسله مراتبی، چندین زیرماژول تطبیق کننده دامنه متخاصم روی بلوک ۳، ۴، ۵ VGG16 پیاده‌سازی شده است. لایه‌های GRL برای استراتژی یادگیری تخاصمی استفاده می‌شوند و اندازه نقشه‌های ویژگی توسط ماژول‌های کاهش مقیاس SRM کاهش می‌یابد. در ماژول تطبیق ویژگی، امتیازهای دسته‌بندی و نتایج بازیابی جعبه مرتبط با ویژگی‌ها در دامنه دسته‌بند ترکیب شده است و این در حالی است که WGRL برای استراتژی یادگیری تخاصمی معرفی می‌شود. SRM از دو بخش تشکیل شده است، بخش اول یک لایه پیکسل‌های کانوولوشن ۱×۱ است که برای کاهش اندازه کانال اعمال می‌شود. بعد از آن، بخش کاهش مقیاس برای ترکیب ویژگی‌های مجاور s×s استفاده می‌شود تا اندازه نقشه‌های ویژگی کاهش یابد[44]..... ۴۲

شکل ۱۸ یک نمای کلی از مدل Domain Adaptive Faster R-CNN. در این روش با دو سطح، سطح تصویر و سطح نمونه، تغییر دامنه مورد بررسی قرار داده شده است. در هر سطح، یک طبقه‌بند دامنه بر پایه ساخته شده است که به شیوه آموزش تخاصمی آموزش داده می‌شود. یک تنظیم‌کننده

تناسب همراه با این دو طبقه‌بند در داخل این دو طبقه‌بند به منظور یادگیری یک RPN ثابت دامنه برای مدل Faster R-CNN وارد شده است[45]. ۴۳

شکل ۱۹ UDA با (MIC). در UDA، یک شبکه معمولاً با یک Loss نظارت شده در دامنه منبع (آبی) و یک تابع ضرر تطبیقی بدون نظارت در دامنه هدف (سبز) آموزش داده می شود. MIC سازگاری بین پیش‌بینی‌های تصاویر هدف پوشانده شده (بنفش) و برچسب‌های ساختگی که بر اساس تصویر کامل توسط (EMA Teacher) تولید می‌شوند، اعمال می‌کند. برای به حداقل رساندن Loss مربوط به MIC، شبکه باید بیاموزد که پیش بینی های مناطق پوشانده شده را از بافت آن ها استخراج کند[46]. ۴۵

شکل ۲۰ نمای کلی از روش ConfMix پیشنهادی. ما نمونه‌های منبع $\boldsymbol{xs}$ و نمونه‌های هدف $\boldsymbol{xt}$ را به مدل تشخیص دهنده $\boldsymbol{F(\theta)}$ منتقل می‌کنیم و پیش‌بینی‌های $\boldsymbol{ys}$ و $\boldsymbol{yt}$ را به ترتیب دریافت می‌کنیم. ما ناحیه هدف با بالاترین اعتماد منطقی را برای تشکیل نمونه مخلوط $\boldsymbol{xM}$ با تصویر منبع انتخاب می‌کنیم، سپس این نمونه به مدل تشخیص دهنده $\boldsymbol{F(\theta)}$ منتقل می‌شود و پیش‌بینی‌های $\boldsymbol{yM}$ را تولید می‌کند. ما مدل را با استفاده از از دست دادن تشخیص نظارت شده با استفاده از اشارات منبع $\boldsymbol{ys}$ و از دست دادن تناسب خود-نظارتی با مقایسه $\boldsymbol{yM}$ با پیش‌بینی‌های ترکیبی منبع و هدف $\boldsymbol{ys, t}$ آموزش می‌دهیم[47]. ۴۶

شکل ۲۱ مروری بر چارچوب CycleGAN در این برنامه. G_Nتصاویر را از دامنه روز به دامنه شب نگاشت می‌کند، در حالی که G_Dنگاشت را به صورت معکوس انجام می‌دهد. تشخیص‌دهندگان D_N و D_D ارزیابی می‌کنند که آیا این تصویر واقعی است یا به ترتیب تصویر تقلبی در دامنه شب و روز است. محدودیت تطابق چرخه با استفاده از تابع ضرر L1 برای اطمینان از توانایی بازسازی مورد استفاده قرار می‌گیرد[51]. ۴۸

شکل ۲۲ بررسی کلی چارچوب رویکرد پیشنهادی TRKP. پیکان‌های سفید به انتقال به جلو اشاره دارند و پیکان‌های خط‌خورده نظارت را نمایان می‌کنند. تشخیص دهنده معلم بر روی تصاویر منبع دارای برچسب آموزش داده می‌شود و برچسب‌های سیاه برای تصاویر بدون برچسب در دامنه هدف تولید می‌کند که تشخیص کننده دانش آموز را راهنمایی می‌کند. TRKP از ماژول تبادل چند-منبعی تخاصمی (AMSD) برای حفظ دانش ویژگی‌های مخصوص دامنه منبع و از طرح معدن دانش کلی مرتبط با هدف (HTRM) برای تقویت تمرکز دانش مرتبط با هدف استفاده می‌کند، که به طور قابل توجهی به تطبیق دانش چند-منبع با دامنه هدف کمک می‌کند[52]. ۴۹

شکل ۲۳ شکل ۷ چارچوب پیشنهادی روش تطبیق دامنه پیشرو را نشان می‌دهد. الگوریتم شامل دو مرحله تطابق است، که در a) و b) نشان داده شده است. در a)، ابتدا تصاویر منبع را تبدیل کرده و تصاویر مصنوعی را با استفاده از مولد G که از طریق CycleGAN یادگرفته شده است، تولید می‌شود. سپس از دامنه منبع با برچسب استفاده کرده و تطابق مرحله اول را با دامنه مصنوعی انجام

می‌دهیم. در b)، مدل یک تطابق مرحله دوم انجام می‌دهد که دامنه مصنوعی با برچسب‌های از منبع به ارث برده شده را می‌گیرد و ویژگی‌های دامنه مصنوعی را با توزیع هدف هماهنگ می‌کند. علاوه بر این، وزن w از تشخیص‌دهنده Dcycle در CycleGAN به منظور تعادل کیفیت تصاویر مصنوعی در از دست دادن تشخیص به دست می‌آید. ساختار کلی شبکه تطابق ما در c) نشان داده شده است. تصاویر با برچسب و بدون برچسب هر دو از طریق شبکه رمزگذار E منتقل می‌شوند تا ویژگی‌های CNN featL و featU استخراج شود. سپس از آنها برای: ۱) یادگیری تشخیص اشیاء تحت نظر با شبکه تشخیص از featL استفاده می‌شود، و ۲) هر دو ویژگی را به GRL و یک تشخیص‌دهنده دامنه ارسال می‌کنیم تا ویژگی‌های بی‌دامنه را به صورت مقابله‌ای و با تحدید نوعی دامنه یادگرفته شود[53]. ۴۹

شکل ۲۴ نمونه‌هایی از نتایج تشخیص از سه وظیفه تطبیق روش تطبیق دامنه پیشرو. دو ردیف اول به ترتیب نتایج آزمایش KITTI به Cityscapes و Cityscapes به Foggy Cityscapes هستند، در حالی که دو ردیف آخر نتایج Cityscapes به BDD100k هستند. نتایج تشخیص روی دامنه مقصد قبل و بعد از اعمال روش تطبیق دامنه به همراه برچسب‌های واقعی نمایش داده شده است[53]. ۵۰

شکل ۲۵ معماری چارچوب تطبیق دامنه روش فوق برای تشخیص اشیاء. این چارچوب بر پایه شبکه تشخیص اشیاء ساخته شده است[54]. ۵۱

شکل ۲۶ توضیح مختصری از AsyFOD. پس از استخراج نقشه ویژگی تصاویر ورودی از طریق بخش اصلی، نمونه‌های ویژگی مرتبط با نمونه به وسیله ماژول پروژه‌ای و حقایق ارائه شده استخراج می‌شوند. سپس، AsyFOD نمونه‌های منبع را به دو مجموعه تشخیص هدف مشابه (آبی تیره) و مجموعه تشخیص هدف نامشابه (آبی روشن) تقسیم می‌کند. نمونه‌های منبع مشابه هدف برای افزایش نمونه‌های هدف استفاده می‌شوند. با این تقسیم‌بندی، تطبیق ناهمگام بین نمونه‌های منبع نامشابه هدف و نمونه‌های هدف افزایش یافته انجام می‌شود که باعث کاهش یادگیری زودهنگام ناشی از توزیع داده‌های بسیار نامتعادل می‌شود. همچنین، AsyFOD آموزش نظارت شده متمرکز بر وظیفه را برای وظایف طبقه‌بندی و مکان‌یابی به طور جداگانه انجام می‌دهد. بهترین نمایش در حالت رنگی مشاهده می‌شود[61]. ۵۳

شکل ۲۷ ساختار DannFixbi برای رویکرد اول[63] ۵۵

شکل ۲۸ ساختار DannFixbi برای رویکرد دوم[63] ۵۵

شکل ۲۹ فلوچارت روشهای پیشنهادی تطبیق دامنههای افزایشی[66] ۵۸

شکل ۳۰ مدل ارائه شده برای انجام تطبیق وجه[67] ۶۰

شکل ۳۱ چند نمونه از تصاویر مختلف مجموعه داده Cityscapes ۶۴

شکل ۳۲ مجموعه داده Cityscapes (سمت چپ) در مقابل Foggy Cityscapes (سمت راست) ۶۵
شکل ۳۳ چند نمونه از تصاویر مختلف مجموعه داده COCO ۶۶
شکل ۳۴ چند نمونه از تصاویر مختلف مجموعه داده Caltech ۶۷
شکل ۳۵ چند نمونه از تصاویر مختلف مجموعه داده SYNTHIA ۶۸
شکل ۳۶ چند نمونه از تصاویر محموعه داده KITTI ۶۹
شکل ۳۷ چند نمونه از تصاویر مختلف مجموعه داده Pascal VOC 2007 ۷۰
شکل ۳۸ چند نمونه از تصاویر مختلف مجموعه داده SIM10K ۷۱
شکل ۳۹ چند نمونه از تصاویر مختلف مجموعه داده BDD100K ۷۲
شکل ۴۰ چند نمونه از تصاویر مختلف مجموعه داده YTBB ۷۳

چکیده

شبکه های عصبی عمیق[1] (DNNs) با سرعت زیادی در سال‌های گذشته توسعه یافته اند. این مدل‌ها نتایج امیدوارکننده‌ای در وظایف مختلف بینایی ماشینی از جمله شناسائی اشیاء نشان داده اند، اما معمولاً به دلیل تغییرات دامنه از دامنه منبع به دامنه هدف، با کاهش دقت مواجه می‌شوند. برای مقابله با این اتفاق مخصوصا در شرایطی که برچسب‌های دامنه هدف در دسترس نیستند، رویکردهای تطبیق دامنه بدون نظارت (UDA) پیشنهاد می‌شوند. روش‌های موجود در شناسایی اشیاء به صورت بدون نظارت، ویژگی‌ها را از سطح تصویر استخراج می‌کنند و همانطور که در این حالت برای وظایف طبقه‌بندی عمل می‌کنند، ویژگی‌های کامل را به‌صورت مستقیم تطبیق می‌دهند. با این حال، تطبیق ویژگی‌ها به صورت کامل بر سطح تصویر به عنوان یک کل برای وظیفه شناسایی اشیاء مطلوب نیست.

کلید واژه‌ها: یادگیری عمیق، داده‌های برچسب‌گذاری شده، توزیع داده‌ها، دامنه منبع، دامنه هدف، تطبیق توزیع، شناسایی اشیا

[1] Deep Neural Networks

فصل ۱: مقدمه و کلیات

۱-۱. مقدمه

شبکه‌های عصبی عمیق (DNNs) با موفقیت در بسیاری از وظایف بینایی ماشینی، مانند شناسایی اشیاء، به دقت بالایی و تقریبا معادل دقت انسانی دست‌یافته‌اند. با اینکه تحقیقات با استفاده از مدل‌های DNNs به نتایج بسیار خوبی رسیده اند، اما هنوز این مدل‌ها قابل مقایسه با قابلیت بینایی انسانی در ارتباط با شناسایی اشیاء نیستند. عملکرد بینایی انسانی کمتر تحت تأثیر محیط اطراف، تغییرات محیطی و غیره قرار می‌گیرد[1]. افزودن نویز تصادفی یا ایجاد تغییرات مختلف در پس‌زمینه باعث کاهش اعتماد به پیش‌بینی مدل های DNNs خواهد شد و این قضیه به صورت تجربی در بسیاری از مقالات و تحقیقات مختلف به صورت تجربی ثابت شده است. ارائه پیش‌بینی قابل اعتماد در محیط‌های مختلف بسیار حیاتی و چالش‌برانگیز است.

کمبود قابلیت سازگاری جلوگیری شدیدی از کاربردهای مدل‌های DNNs می‌کند. در تحقیقات، کمبود قابلیت سازگاری معمولاً با استفاده از دو یا بیشتر مجموعه‌داده‌ها اندازه‌گیری می‌شود، به این معنا که وقتی یک شبکه بر روی یک یا چند مجموعه‌داده آموزش داده می‌شود و بر روی یک مجموعه‌داده دیگر برای حل همان وظیفه استفاده می‌شود، مدل احتمالاً عملکرد خوبی از خود نشان نمی‌دهد. این پدیده به دلیل تفاوت توزیع داده‌ها بین دو مجموعه‌داده ایجاد می‌شود، که به عنوان "فاصله دامنه" نیز ارجاع می‌شود.

۱-۲. ضرورت انجام پژوهش

تطبیق دامنه[1] یک مفهوم مهم در حوزه یادگیری ماشین و شبکه‌های عصبی است که به تطابق یک مدل یادگیری با داده‌های آموزشی و داده‌های آزمایشی از دامنه‌های مختلف اشاره دارد. در مواردی که داده‌ها از دامنه‌های مختلف باشند، ممکن است مدل یادگیری عمیق عملکرد نامطلوبی داشته باشد، زیرا اختلاف‌های بین دامنه‌ها می‌تواند به تخریب دقت مدل منجر شود. به عبارت دیگر، مدلی که بر روی یک دامنه آموزش دیده شده است، ممکن است بر روی دامنه‌های دیگر کارایی خود را از دست دهد.

در زمینه تشخیص اشیا، این مسئله حائز اهمیت است. چرا که داده‌های آموزشی مصنوعی که برای آموزش مدل‌های تشخیص اشیا استفاده می‌شوند، معمولاً از دامنه‌های مصنوعی و محدود به شرایط خاصی تولید می‌شوند. اما در محیط‌های واقعی، اشیا در شرایط مختلف از زوایا مختلف، با شرایط تابش تور متفاوت و پس‌زمینه‌های متنوع قرار می‌گیرند. بنابراین، انتقال دانش از دامنه‌های مصنوعی به دامنه‌های واقعی و تطبیق مدل‌های یادگیری به شرایط واقعی اهمیت زیادی دارد.

از جمله دلایل دیگری که میتوان در وصف اهمیت تطبیق دامنه اشاره نمود، محدودیت منابع داده‌های برچسب‌دار است. عموماً جمع‌آوری داده‌های برچسب‌دار در دامنه مقصد هزینه‌بر و زمان‌بر می‌باشد. بنابراین، انتقال دامنه به ما امکان می‌دهد تا با بهره‌گیری از دانش ویژگی‌ها و الگوهای یادگرفته شده در دامنه منبع، عملکرد مدل‌ها در دامنه مقصد را بهبود بخشیم.

همچنین استفاده از دانش عمومی مدل‌ها نیز دارای اهمیت است. مدل‌های یادگیری عمیق عموماً قادرند الگوها و ویژگی‌های کلان و عمومی را از داده‌ها استخراج کنند که در چندین دامنه قابل استفاده هستند. بدین ترتیب، اگر یک مدل در یک دامنه خاص خوب عمل کند، قابل انتظار است که بتواند بخشی از این دانش را به دامنه‌های مشابه منتقل نماید.

این پژوهش‌ها می‌توانند به تعریف روش‌ها و الگوریتم‌هایی کمک کنند که مدل‌های عصبی را قادر به تطبیق با تغییرات دامنه‌ها کنند و در شرایط واقعی به عملکرد بهتری دست یابند.

۱-۳. انگیزه و هدف

یکی از اهداف اصلی تحقیق در این حوزه، بهبود دقت و کارایی شبکه‌های عصبی در تشخیص اشیا در تصاویر است. این بهبودات می‌توانند در بسیاری از کاربردهای عملیاتی مانند خودروهای بدون راننده، پزشکی تصویری، امنیت و نظارت ترافیکی مفید باشند. همچنین بسیاری از شبکه‌های عصبی برای آموزش و تشخیص اشیا نیاز به داده‌های آموزشی بزرگ و متنوع دارند. اما تهیه و برچسب‌گذاری این داده‌ها ممکن است پرهزینه و زمان‌بر باشد. از این رو، هدف می‌تواند تلاش برای بهبود کارایی شبکه‌ها

[1] Domain Adaptation

با استفاده از داده‌های آموزشی مصنوعی و ایجاد تطبیق بین داده‌های مصنوعی و واقعی باشد. از طرفی تطبیق دامنه به شبکه‌های عصبی امکان می‌دهد تا در مواجهه با تغییرات و تنوع‌های محیطی، عملکرد خود را بهتر حفظ کنند و از سیستم‌های مقاوم‌تری برخوردار باشند.

سوالات این پژوهش عبارت‌اند از:

۱. چگونه روش‌های انتقال یادگیری می‌توانند عملکرد مدل را در شرایط تطبیق دامنه بهبود بخشند؟

۲. چگونه می‌توان دامنه‌های مشابه با داده‌های آموزشی مصنوعی را انتخاب کرد تا تطبیق دامنه بهبود یابد؟

۳. آیا استفاده از تکنیک‌های ترکیب داده‌های مصنوعی با داده‌های واقعی می‌تواند به بهبود تطبیق دامنه کمک کند؟

۴. چگونه می‌توان از ترکیب روش‌های مختلف مانند انتقال یادگیری، یادگیری نیمه نظارتی و یادگیری تقویتی بهره برد تا تطبیق دامنه بهبود یابد؟

۵. دامنه میانی چه تاثیری بر بهبود عملکرد تطبیق دامنه دارد؟

۱-۴. ساختار سمینار

در فصل اول، به مقدمات و بررسی دلایل نیاز به تطبیق دامنه و استفاده از آن برای کاربرد تشخیص اشیا پرداخته شده است. در فصل دوم با نگاهی مبسوط، به الگوریتم‌های پایه و اساسی تشخیص اشیا و تطبیق دامنه به صورت کلی پرداخته می‌شود و یک دسته بندی کلی از روش های تطبیق دامنه ارائه می‌شود. در فصل سوم نتایج تجربی حاصل شده از مقالات و پایان نامه های مرتبط و توضیحات مربوط به معیارهای ارزیابی و مجموعه دادگان مختلف مربوط به کار با این موضوع پرداخته می‌شود. و در نهایت در فصل چهارم نتیجه‌گیری‌ها بیان و پیشنهاداتی جهت کارهای آتی مطرح می‌شود.

فصل ۲: مبانی نظری تحقیق

۲-۱. مقدمه

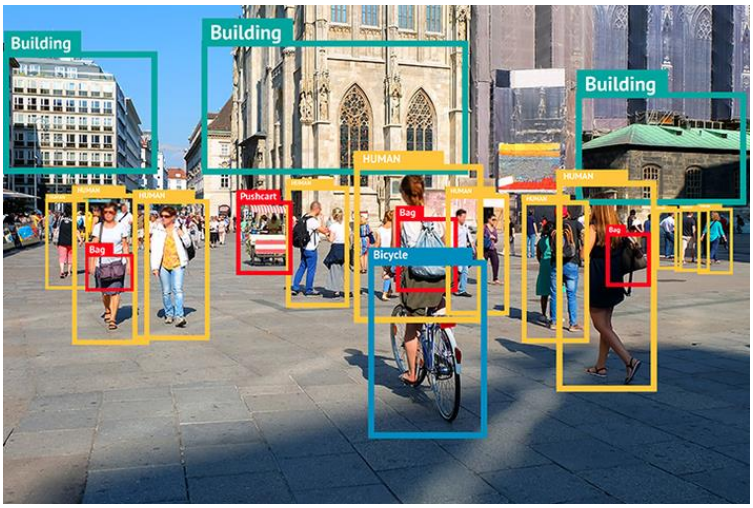
در این فصل، در ابتدا به الگوریتم‌های پایه و اساسی تشخیص اشیا و تطبیق دامنه به صورت کلی پرداخته می‌شود و سپس یک دسته بندی کلی از روش های تطبیق دامنه ارائه می‌شود.

۲-۲. تشخیص اشیا

تشخیص اشیا[1] یک وظیفه مهم در زمینه هوش مصنوعی و بینایی کامپیوتری است که به کامپیوتر اجازه می‌دهد تا اشیا و اجزای مختلف یک تصویر یا ویدئو را تشخیص دهد و آن‌ها را به صورت خودکار مشخص کند. این وظیفه برای بسیاری از کاربردها از جمله خودران‌های اتومبیل، تحت نظر گرفتن اشیا در تصاویر پزشکی، شناسایی اشیا در تصاویر مداربسته، تشخیص چهره در عکس‌ها و ویدئوها و بسیاری از سیستم‌های هوش مصنوعی دیگر مورد استفاده قرار می‌گیرد.

[1] Object detection

همچنین تشخیص اشیا به وسیله تکنولوژی‌های مختلفی انجام می‌شود، از جمله شبکه‌های عصبی عمیق مانند YOLO[1]، SSD[2] و Faster R-CNN[3]. این شبکه‌ها از ترکیب الگوریتم‌های یادگیری ماشین و شبکه‌های عصبی عمیق برای تشخیص اشیا با دقت و سرعت بالا استفاده می‌کنند.



شکل ۱ نمونه ای از کار تشخیص اشیا

وظیفه تشخیص اشیا می‌تواند به عنوان ترکیبی از دو مرحله اصلی دیده شود: یافتن مناطق تصویری که ممکن است شامل اشیا باشند، و سپس دسته‌بندی اشیا در این مناطق به صورت مستقل. قبل از ظهور یادگیری عمیق، این کار با استفاده از رویکرد sliding window انجام می‌شد. در این رویکرد، یک دسته‌بند تصویر به مناطق مختلف تصویر اعمال می‌شد و تنها پیش‌بینی‌هایی با احتمال بالاتر حفظ می‌شدند. اما در حال حاضر، دو خانواده اصلی از آشکارسازهای اشیا شناخته می‌شوند.

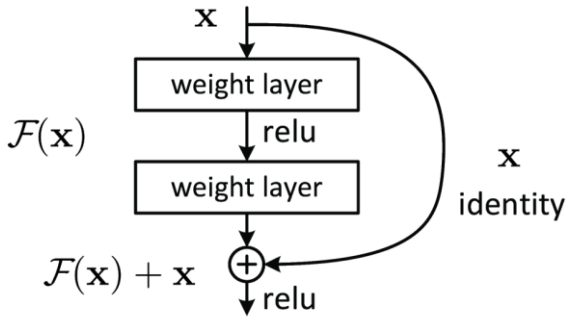
چندین معماری شبکه به عنوان شبکه‌های پایه در آشکارسازهای اشیا استفاده شده‌اند تا ویژگی‌های تصاویر ورودی را برای وظیفه تشخیص اشیا استخراج کنند. به دلیل عملکرد بالا، یکی از محبوب‌ترین این معماری‌ها، معماری residual neural network (ResNet) است[2].

شبکه‌های ResNet توسط استفاده از اتصال‌های کوتاه یا اتصالات پرش مشخص می‌شوند. از طریق این اتصالات پرش، خروجی یک بلوک شبکه به سیگنال ورودی همان بلوک اضافه می‌شود. این امر منجر به انتقال گرادیان از لایه‌های سطح بالا به لایه‌های سطح پایین در مرحله پس‌انتشار می‌شود و از ویژگی‌های اولیه استخراج‌شده در لایه‌های ابتدایی شبکه در فرآیند استنباط بهره‌بری می‌کند. این به

[1] You Only Look Once
[2] Single Shot MultiBox Detector
[3] Region-based Convolutional Neural Network

ایجاد شبکه‌های عمیق بسیاری که مشکلات معماری‌های قبلی مربوط به آموزش و دقت را حل کرده‌اند، امکان پذیر شده است. ساختار یک بلوک باقیمانده در شکل 2 نشان داده شده است.

تعداد لایه‌های یک شبکه ResNet معمولاً به عنوان یک پسوند مشخص می‌شود. مقادیر معمول شامل ۱۸، ۳۴، ۵۰ (ResNet-50)، ۱۵۲ و حتی بیش از ۱۰۰۰ لایه[3] هستند.



شکل ۲ مثالی از ساختار یک بلوک residual learning

مدل های تشخیص دهنده را میتوان به دو دسته یک مرحله و دو مرحله ای تقسیم کرد که در ادامه به توضیح هرکدام خواهیم پرداخت.

۲-۲-۱. مدل‌های تشخیص دهنده یک مرحله‌ای[1]

مدل‌های تشخیص دهنده یک مرحله‌ای با تفاوت اساسی نسبت به مدل‌های دو مرحله‌ای عمل می‌کنند. این مدل‌ها فاز پیشنهاد منطقه را نادیده می‌گیرند و به جای آن تمرکز خود را بر روی پیش‌بینی مناطق و کلاس‌های اشیاء در تصاویر می‌گذارند. این باعث می‌شود که پیش‌بینی‌ها نسبت به مدل‌های دو مرحله‌ای سریع‌تر انجام شود و برای برنامه‌های واقعی زمان واقعی اما با هزینه کیفیت پیش‌بینی کمتری مناسب باشند. کار پیشرو در زمینه تشخیص اشیاء در یک مرحله کار توسط Pierre Sermanet و همکارانش[4] انجام شده است. در این روش، لایه‌های طبقه‌بندی آخر شبکه عصبی کانولوشنی استاندارد با یک شبکه رگرسیون برای هر کلاس جایگزین می‌شوند تا مختصات جعبه‌های محدودکننده اشیاء را پیش‌بینی کنند.

۲-۲-۱-۱. شبکه YOLO

شبکه YOLO[2] یک مدل عمیق یادگیری عمیق برای شناسایی اشیاء و تشخیص اشیاء در تصاویر و ویدئوها است. این مدل در سال ۲۰۱۶ توسط جوزف ردمان و مجتبی کارویو[5] ارائه شد و به عنوان یک روش موثر و سریع برای شناسایی اشیاء در زمان واقعی شناخته می‌شود.

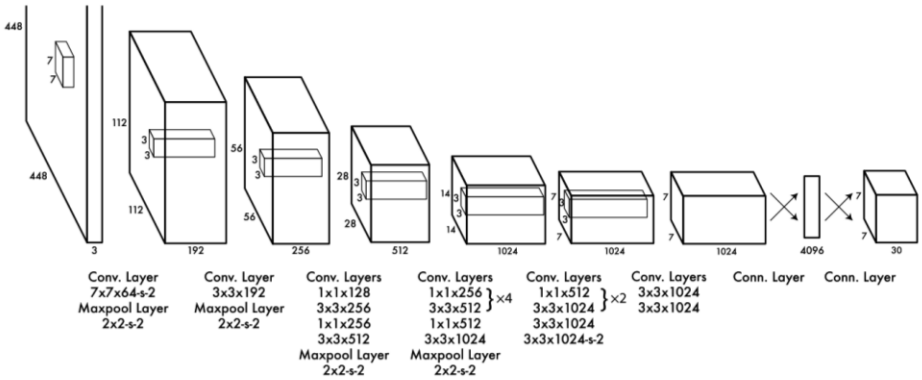
[1] One stage object detectors
[2] You Only Look Once

این مدل، بحث شناسایی اشیاء را به عنوان یک مسئله‌ی رگرسیون معرفی می‌کند و به گونه‌ای عمل می‌کند که تصویر را به سلول‌های کوچکتر تقسیم می‌کند. اگر مرکز یک شیء درون یک سلول قرار بگیرد، آن سلول مسئولیت تشخیص و پیش‌بینی مکان محدوده مستطیلی[1]، امتیاز اعتماد[2] و احتمال کلاس برای اشیاء تشخیص داده شده را دارد. برای یادگیری بهینه پیش‌بینی مکان محدوده مستطیلی و کلاس‌های اشیاء به صورت همزمان، یک تابع خسارت خاص استفاده می‌شود.

مهمترین ویژگی YOLO این است که به عنوان یک مدل تک مرحله‌ای عمل می‌کند، به این معنا که به جای اجرای چندین مرحله از پیش‌پردازش و اجرای شبکه‌های پیچیده، تصویر ورودی را یک‌بار پردازش می‌کند و اشیاء موجود در تصویر را در یک گام تشخیص می‌دهد. این ویژگی باعث می‌شود که YOLO بسیار سریع باشد و به راحتی در زمان واقعی برای کاربردهایی مانند تشخیص اشیاء در ویدئوها یا مانیتورینگ ترافیک استفاده شود.

معماری شبکه YOLO الهام گرفته از مدل GoogLeNet برای دسته‌بندی تصاویر است. با این حال، به جای استفاده از ماژول‌های inception که در GoogLeNet استفاده می‌شوند، YOLO از لایه‌های کاهنده[3] ۱×۱ به آن‌ها پیوسته به لایه‌های کانولوشن ۳ × ۳ استفاده می‌کند. پیش‌بینی نهایی یک تنسور ۷ × ۷ × ۳۰ است.

علاوه بر این، یک نسخه سریع از YOLO وجود دارد که از یک شبکه عصبی با تعداد کمتری لایه کانولوشن (۹ به جای ۲۴) و تعداد کمتری فیلتر در این لایه‌ها استفاده می‌کند.



شکل ۳ معماری شبکه YOLO[5]

شبکه تشخیص از ۲۴ لایه کانولوشنال[4] و پس از آن دو لایه تماماً متصل[5] تشکیل شده است. لایه‌های کانولوشنال برای استخراج ویژگی‌های تصویر به کار می‌روند و لایه‌های تماماً متصل برای

[1] bounding box
[2] confidence score
[3] Reduction Layers
[4] Convolutional Layers
[5] Fully Connected Layers

دسته‌بندی و تصمیم‌گیری نهایی استفاده می‌شوند. در معماری این شبکه، از لایه‌های کانولوشنال 1×1 به صورت متناوب استفاده می‌شود. این لایه‌ها به کاهش ابعاد ویژگی‌های تصویر از لایه‌های قبلی می‌پردازند. از این لایه‌ها برای کاهش تعداد کانال‌ها و بهبود محاسبات در شبکه استفاده می‌شود.

در مرحله‌ای از آموزش، لایه‌های کانولوشنال ابتدا بر روی مسئله دسته‌بندی تصاویر ImageNet آموزش داده می‌شوند. این آموزش با استفاده از تصاویری با وضوح نصف (224×224 پیکسل) انجام می‌شود. سپس، در مرحله تشخیص، وضوح تصاویر به دو برابر افزایش پیدا می‌کند تا برای شناسایی بهتر اشیاء استفاده شود.

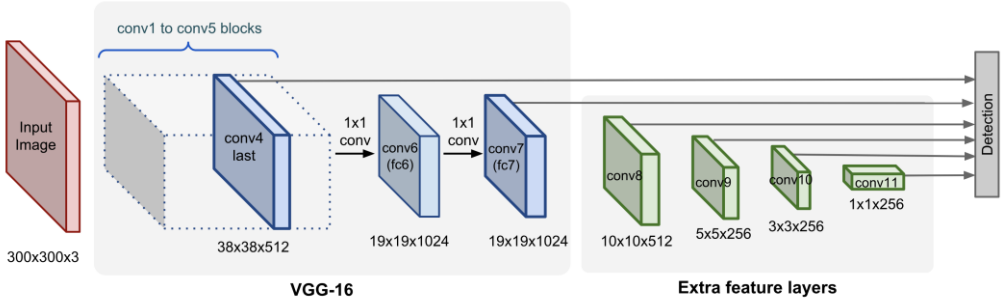
نویسندگان مقاله عملکرد YOLO را با استفاده از مجموعه داده PASCAL VOC بررسی کرده و نشان داده‌اند که این روش بیش از دو برابر دقت متوسط میانگین دیگر سیستم‌های زمان واقعی را دارد و در حین پردازش ویدئوهای جاری، دارای کمتر از ۲۵ میلی‌ثانیه تاخیر است. مقاله همچنین مزایای YOLO نسبت به روش‌های سنتی شناسایی اشیاء را بررسی می‌کند و یک پیوند به یک نمونه از سیستم کار در زمان واقعی با دوربین وب ارائه می‌دهد.

YOLO به دلیل سرعت بالا و عملکرد برتر در شناسایی اشیاء در زمان واقعی در بسیاری از کاربردهای بینایی ماشین و پردازش تصویر مورد استفاده قرار گرفته است. این مدل در نسخه‌های مختلفی مانند YOLOv1 تا YOLOv5 و YOLOX عرضه شده و بهبودهای متعددی در زمینه دقت و سرعت اجرا داشته است.

۲-۱-۲-۲. شبکه SSD

SSD[1] [6] روشی برای تشخیص اشیاء در تصاویر با استفاده از یک شبکه عصبی عمیق تکین است. اصلی‌ترین اجزای SSD شامل پیش‌بینی امتیازهای دسته‌بندی و انتقال‌های جعبه (مربع) برای یک مجموعه ثابت از جعبه‌های مربعی پیش‌فرض با استفاده از فیلترهای کانولوشنی کوچکی است که به نقشه‌های ویژگی اعمال می‌شوند. SSD از یک شبکه عصبی کانولوشنی به نام "VGG-16" به عنوان پایه استفاده می‌کند و به روی آن چندین لایه کانولوشنی با اندازه‌های کمتر اضافه می‌کند. این کار به منظور ساخت یک نمایش پیرامونی از تصاویر و تشخیص کارهای مختلف با اندازه‌های مختلف در هر لایه پیرامونی انجام می‌شود. به جای تقسیم تصاویر به سلول‌ها، از جعبه‌های مرجع[2] برای تشخیص سریع‌تر استفاده می‌شود.

[1] Single Shot MultiBox Detector
[2] Anchor boxes



شکل ۴ ساختار هرمی معماری SSD[6]

در اصل، SSD یک روش برای تشخیص اشیاء در تصاویر است که با استفاده از یک شبکه عصبی کانولوشنی به تصویر ورودی نگاه می‌کند و اشیاء مختلفی که در تصویر وجود دارند را به صورت همزمان و با دقت تشخیص می‌دهد. این به این شکل کار می‌کند که تصاویر را در لایه‌های مختلف پیرامونی تجزیه و تحلیل می‌کند تا بتواند اشیاء با اندازه‌های مختلف را تشخیص دهد. همچنین، از جعبه‌های مرجع به عنوان نقاط مرجع برای تشخیص اشیاء استفاده می‌شود تا فرآیند تشخیص سریع‌تر و کارآمدتری داشته باشد.

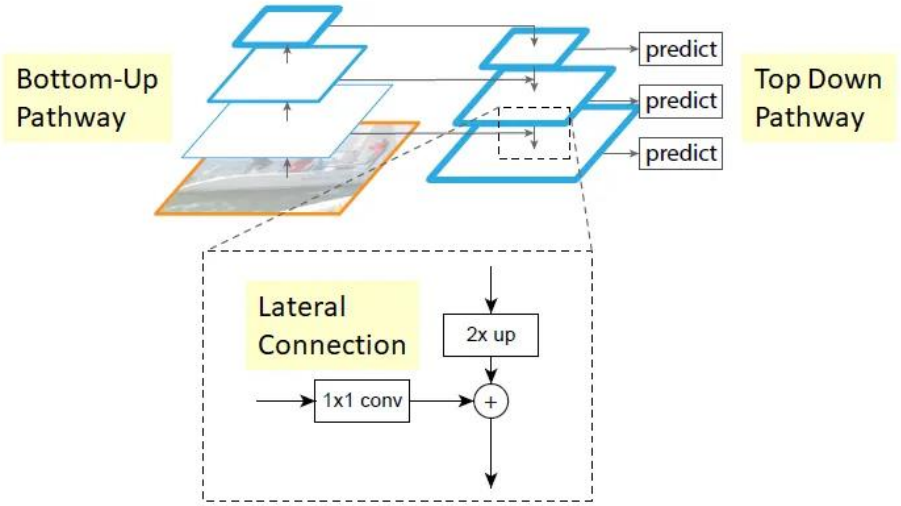
برای دستیابی به دقت بالای تشخیص، SSD پیش‌بینی‌های مقیاس‌های مختلف را از نقشه‌های ویژگی با مقیاس‌های مختلف تولید می‌کند و به صراحت پیش‌بینی‌ها را بر اساس نسبت ابعاد متمایز می‌کند. این ویژگی‌های طراحی باعث می‌شوند که آموزش از ابتدا تا انتها ساده باشد و دقت بالایی را به ارمغان بیاورد، حتی بر روی تصاویر ورودی با رزولوشن پایین، که تجارت سرعت در مقایسه با دقت را بهبود می‌بخشد. SSD نسبت به تکنیک‌های قبلی تک مرحله ای (مانند YOLO) سریع‌تر است و با اختلاف قابل توجهی دقت بالاتری دارد، به حدی که به دقت تکنیک‌های کندتری که اقتراح‌های منطقه‌ای و پولینگ صریح را انجام می‌دهند (شامل Faster R-CNN) را داراست. آزمایشات شامل تحلیل زمانی و دقت بر روی مدل‌های با اندازه ورودی متغیری است که بر روی مجموعه داده‌های PASCAL VOC، COCO، و ILSVRC ارزیابی شده و با مجموعه‌ای از رویکردهای معاصر وضعیت آخرین مقالات مقایسه می‌شوند.

۳-۱-۲-۲. شبکه FPN[1]

شبکه FPN[7] از تعدادی سطح هرمی تشکیل شده است که با مراحل مختلف شبکه متناظر هستند. هر مرحله از چندین لایه کانولوشنی با اندازه یکسان تشکیل شده است و اندازه این لایه‌ها به ازای مراحل مختلف دو برابر می‌شود. به علاوه، مراحل مختلف با یک مسیر بالا به پایین در جهت معکوس از یکدیگر جفت می‌شوند و اتصالات جانبی دارند. این اتصالات و مسیر بالا به پایین امکان می‌دهند تا شبکه یک هرم ویژگی چندسطحی و غنی را برای هر تصویر ورودی ایجاد کند.

[1] Feature Pyramid Network

پس از معرفی این روش در دسته‌ی مدل‌های تشخیص دهنده تک مرحله ای، FPNها همچنین به عنوان بخشی از اسکلت مدل‌های تشخیص دهنده دو مرحله ای مورد استفاده قرار گرفته‌اند تا به بهبود عملکرد این مدل‌های تشخیص دهنده کمک کنند.



شکل ۵ معماری و ساختار FPN[7]

۲-۲-۲. مدل‌های تشخیص دهنده دو مرحله‌ای[1]

همانطور که اشاره شد، مدل‌های بسیار متداول برای تشخیص اشیاء بر مبنای یک رویکرد دو مرحله‌ای هستند و به خانواده شبکه‌های عصبی کانولوشنی مبتنی بر مناطق[2] (R-CNNs) تعلق دارند.

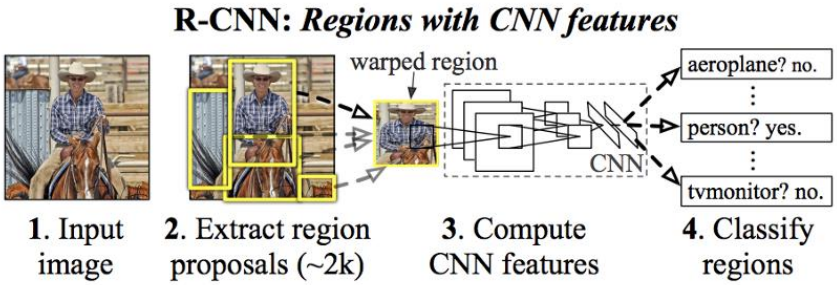
۱-۲-۲-۲. شبکه R-CNN

در R-CNN ابتدایی[8]، تعداد قابل مدیریتی از مناطق ممکن جهت تمرکز[3] (RoIs) با استفاده از الگوریتم جستجوی انتخابی[4]، از تصویر استخراج می‌شوند. جستجوی انتخابی به هدف محلی‌سازی اشیاء در تصاویر ورودی انجام می‌شود و در مقابل جستجوی جامع استفاده شده در گذشته قرار می‌گیرد، که در آن از رویکرد پنجره‌های لغزان برای شناسایی نواحی مهم استفاده می‌شود. در جستجوی انتخابی، تصویر ورودی ابتدا به نواحی کوچک تقسیم می‌شود، سپس به تدریج بر اساس شباهت آنها به یکدیگر ادغام می‌شوند. این عملیات تا زمانی ادامه می‌یابد که مناطق مستقل بزرگتری ایجاد شوند. این نواحی به عنوان مناطق ممکن برای تمرکز نامزد بازمی‌گردانده می‌شوند.

سپس یک CNN به صورت مستقل بر روی هر منطقه تمرکز پیشنهادی ارزیابی می‌شود تا ویژگی‌ها استخراج شوند که به یک ماشین بردار پشتیبانی (SVM) داده می‌شوند تا تشخیص دهد آیا یک شیء

[1] Two stage object detectors
[2] Region-Based Convolutional Neural Networks
[3] Regions of Interest
[4] Selective Search

در این جعبه محدودکننده وجود دارد (وظیفه دسته‌بندی) و همچنین مکان آن را با مکان شیء متناظر تطابق دهد (وظیفه رگرسیون). از آنجایی که SVM یک الگوریتم دسته‌بندی دودویی است، به همین دلیل تعدادی SVM برای وظیفه دسته‌بندی برای هر کلاس مورد علاقه استفاده می‌شوند. خلاصه‌ای از روش در شکل ۵ گزارش شده است.



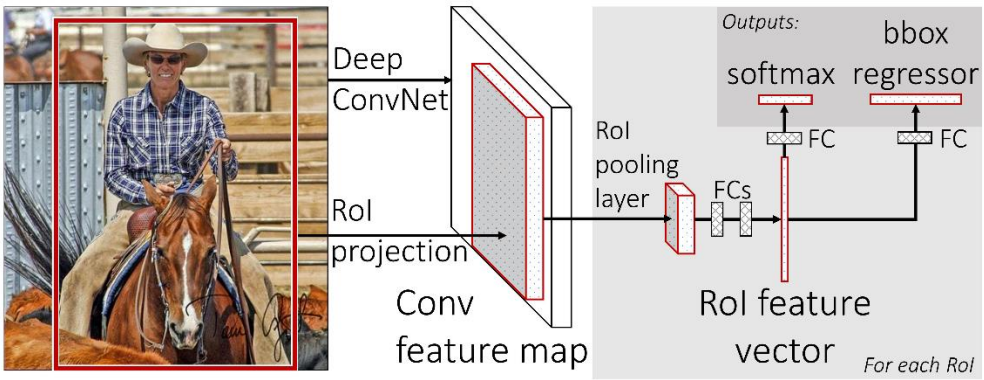
شکل ۶ ساختار و نحوه عملکرد شبکه R-CNN[8]

مدل‌های Fast R-CNN[9] و Faster R-CNN[10] معماری و عملکرد R-CNN را بهبود دادند و زمان استنتاج را کاهش دادند.

۲-۲-۲-۲. شبکه Fast R-CNN

شبکه Fast R-CNN[9] بهبودهایی در عملکرد نسبت به روش R-CNN ارائه می‌دهد که از طریق اضافه کردن لایه RoI pooling و یک روش آموزش تک‌مرحله‌ای با یک تابع هدف چند وظیفه‌ای بهینه‌سازی شده به دست آمده است. لایه RoI pooling به شبکه اجازه می‌دهد تا پردازش را به صورت موازی انجام دهد و این موضوع باعث مشود تا یکی از علت‌های اصلی کندی معماری R-CNN حل شود. به عبارت دیگر، در R-CNN هر پیشنهاد RoI به صورت مستقل به یک شبکه عصبی کانولوشنی (CNN) برای استخراج ویژگی‌ها ارسال می‌شود. با استفاده از RoI pooling، تصویر ورودی کامل فقط یک بار به استخراج کننده ویژگی‌ها (CNN) ارسال می‌شود. سپس نقشه‌های ویژگی مستطیلی که متناظر با هر پیشنهاد RoI است از ویژگی‌های تصویر استخراج می‌شوند.

برای افزایش سرعت تست، Fast R-CNN از تکنیک "بازگردانی جعبه محدود کننده" استفاده می‌کند که به شبکه امکان بهبود دقت مکانی پیشنهادهای اشیاء را می‌دهد. این کار منجر به کاهش تعداد پیشنهادهایی می‌شود که باید پردازش شوند و در نتیجه به سرعت تست کاهش می‌دهد.

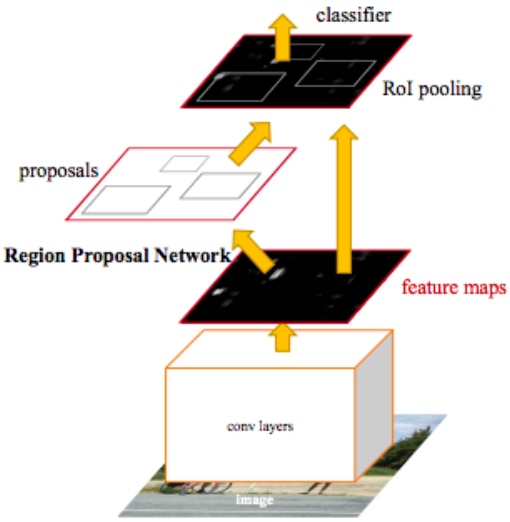


شکل ۷ ساختار و نحوه عملکرد شبکه Fast R-CNN[9]

۳-۲-۲-۲. شبکه Faster R-CNN

شبکه Faster R-CNN[10]، بهبودهایی نسبت به شبکه قبلی (Fast R-CNN) دارد. یکی از اصلی‌ترین بهبودهای این روش نسبت به Fast R-CNN، معرفی یک شبکه به نام Region Proposal Network (RPN) است. این شبکه به گونه‌ای طراحی شده است که بخشی از لایه‌های شبکه عصبی کانولوشنی برای مرحله پیشنهاد مناطق تصویر را جایگزین الگوریتم جستجوی انتخابی که گران و زمان‌بر است، می‌کند. این تغییر به بهبود عملکرد و کاهش زمان مورد نیاز برای تشخیص اشیاء در تصاویر کمک می‌کند.

در شکل ۷ نحوه عملکرد Faster R-CNN به اختصار توصیف شده است.



شکل ۸ ساختار و نحوه عملکرد شبکه Faster R-CNN[10]

هدف اصلی RPN این است که مجموعه‌ای از RoIs از یک تصویر ورودی استخراج کند. این CNN با به اشتراک گذاشتن یک مجموعه مشترک از لایه‌های کانولوشنی با یک شناسایی‌کننده شیء Fast R-CNN همراه است و مناطق تصویری را که باید بر روی آن‌ها تمرکز کند، مشخص می‌کند.

برای تولید پیشنهادهای RoI، RPN ویژگی‌های استخراج شده توسط آخرین لایه کانولوشنی مشترک را به عنوان ورودی دریافت می‌کند و از یک روش پنجره‌ای لغزان برای استخراج پیشنهادهای چندگانه استفاده می‌کند. هر RoI به یک نقشه ویژگی با ابعاد کمتر نگاشت می‌شود که به دو لایه مشابه (یک لایه تصمیم‌گیری دودویی برای تخمین احتمال این که آیا RoI ورودی واقعاً یک شیء را شامل می‌شود یا خیر و یک لایه بهبود مستطیل‌های محیطی برای بهبود پیشنهادهای مستطیلی) تغذیه می‌شود. نواحی که در هر مکان پنجره لغزان تولید می‌شوند، ابعاد و نسبت‌های جانبی مختلفی دارند و به عنوان Anchor Boxes شناخته می‌شوند. این مفهوم مهم در تشخیص اشیاء مدرن است.

RPN می‌تواند با هدف کمینه کردن یک تابع از دست دادن چند وظیفه‌ای آموزش داده شود. اما چون RPN و شناسایی‌کننده شیء Fast R-CNN از یک مجموعه از لایه‌های کانولوشنی به اشتراک گذاشته می‌کنند، آموزش مستقل از این دو مؤلفه ممکن است وزن‌ها را به روش‌های مختلفی تغییر دهد و نتایج ناسازگار باشند. برای این کار، سه تکنیک معرفی شده است:

- آموزش متناوب: در این روش، RPN به عنوان یک شبکه عصبی کانولوشنی از پیش آموزش داده شده مقدماتی شده و به طور کامل بر روی وظیفه پیشنهاد RoI به طور انتها به انتها آموزش داده می‌شود، و پیشنهادهای تولید شده برای آموزش یک شناسایی‌کننده شیء Fast R-CNN استفاده می‌شود.

- آموزش مشترک تقریبی: در این روش، مؤلفه‌های RPN و Fast R-CNN به یک شبکه ترکیب می‌شوند، اما این آموزش مشترک به نتایج تقریبی منجر می‌شود.

- آموزش مشترک غیرتقریبی: در این روش، مشکل تقریبی آموزش مشترک با توجه به مختصات مستطیل‌های پیشنهادی را حل کرد. این نیاز به توسعه یک لایه تجمع RoI که با توجه به مختصات مستطیل‌ها قابل تفکیک باشد، دارد.

به دلیل عملکرد خوب و سرعت مناسب آن، Faster R-CNN به عنوان یک مدل مرجع برای ارزیابی و به عنوان پایه برای ایجاد مدل‌های مشتق شده در حوزه تشخیص اشیاء در تصاویر گسترش یافته است.

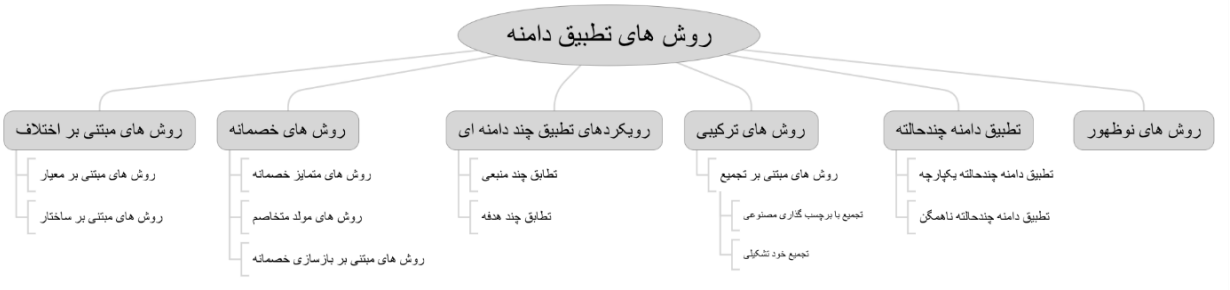
۳-۲. تطبیق دامنه در یادگیری عمیق

از آنجایی که شبکه‌های عصبی عمیق با دقت بالا (یا هر معیار مورد نیاز دیگری) مرتبط هستند و می‌توانند مطابق آخرین پیشرفت ها[1] نتایج را ارائه دهند، استفاده از آن‌ها در بسیاری از برنامه‌ها و وظایف هوش مصنوعی و یادگیری ماشین افزایش یافته است. با این حال، این شبکه‌ها با مشکلات تغییر دامنه مواجه هستند و قادر به تطبیق خود با توزیع‌های مختلف داده‌ها (نسبت به دامنه منبع) نمی‌باشند

[1] State-of-the-art

و نتایج بهترین‌های فعلی را ارائه نمی‌دهند. علاوه بر این، از آنجایی که شبکه‌های عصبی عمیق به حجم زیادی از داده‌های برچسب‌خورده برای آموزش نیاز دارند و دسترسی همیشگی به داده‌های برچسب‌خورده مشکل است (زمان‌بر، دشوار یا گاهی غیرممکن می‌باشد)، استفاده از یادگیری انتقالی برای شبکه‌های عصبی عمیق بسیار ضروری است. برخلاف تطبیق دامنه در یادگیری سطحی، تمرکز و هدف تطبیق دامنه شامل گنجاندن آن در فرآیند یادگیری عمیق و کانال ارتباطی[1] است به طوری که بازنمایی های قابل انتقال آموخته شوند[11].

نمودار ۱ روش‌ها و رویکردهای یادگیری عمیق تطبیق دامنه را نشان می‌دهد و طبقه‌بندی روش‌های تطبیق دامنه که توسط وانگ و دنگ[12] اشاره شده است را بسط می‌دهد. با این حال، این مقاله تنها بر تکنیک‌های تطبیق دامنه عمیق در حوزه‌ی تصویری تمرکز دارد. در مقابل، هدف ما این است که رویکردهای تطبیق دامنه عمیقی که برای کاربرد تشخیص اشیا بکار رفته اند را بررسی و طبقه بندی کنیم.



نمودار ۱ درخت‌واره روش های تطبیق دامنه

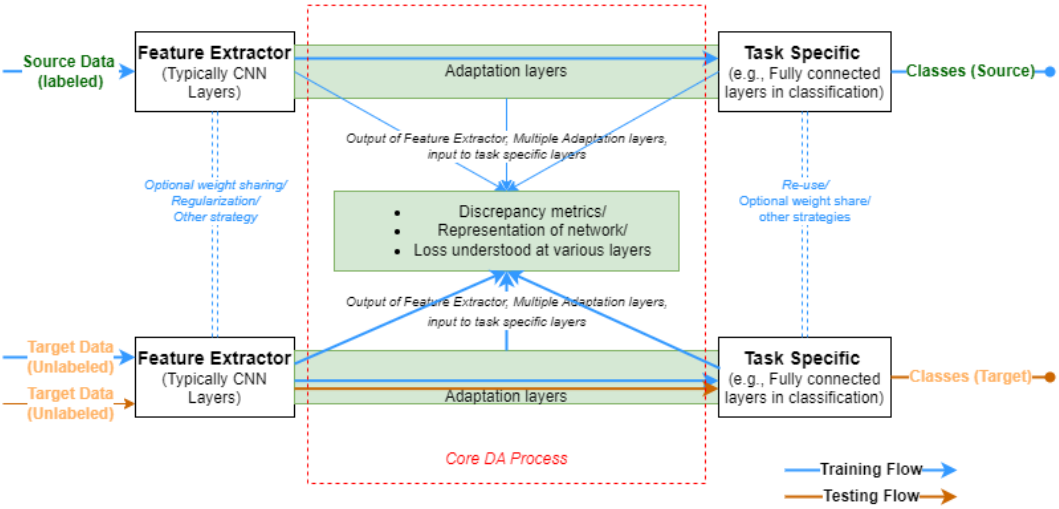
۲-۳-۱. روش‌های مبتنی بر اختلاف

Mathelin و همکارانش[13] یک رویکرد نوآورانه برای طراحی استراتژی‌های یادگیری فعال ارائه می‌دهند که منجر به تطبیق دامنه می‌شود. نویسندگان مرزهای خطاهای تعمیم را استخراج نموده و یک الگوریتم عملی را پیشنهاد می‌کنند که قادر به مدیریت مجموعه‌های داده بزرگ می‌باشد.

این رویکرد پیشنهادی با کارهای پیشین در زمینه یادگیری فعال تفاوت دارد زیرا به جای تطابق کشفی یا توزیع، بر اساس اندازه‌گیری اختلاف بین توزیع‌های دارای برچسب و بدون برچسب می‌باشد. نویسندگان نشان می‌دهند که این اندازه‌گیری اختلاف می‌تواند برای استخراج مرزهای خطاهای تعمیم، که به ریسک هدف وابسته هستند، به کار رود. آنها یک الگوریتم عملی پیشنهاد می‌کنند که این اندازه‌گیری اختلاف را به حداقل می‌رساند تا نقاطی که برای برچسب‌گذاری اطلاعات بیشتری ارائه می‌کنند، انتخاب شوند.

[1] Pipeline

الگوریتم پیشنهادی بر اساس یک شرط منظم است که توابع ضرر باید آن را رعایت کنند تا مرزهای خطاهای تعمیم حفظ شوند. نویسندگان توضیح می‌دهند که این شرط توسط بسیاری از توابع ضررهای معمولی که به طور معمول استفاده می‌شوند، از جمله squared loss و logistic loss، رعایت می‌شود.



شکل ۹ ساختار/معماری شبکه معمولی شبکه‌های عصبی عمیق که روش‌های مبتنی بر اختلاف را برای تطبیق دامنه پیاده‌سازی می‌کنند. در این ساختار، لایه‌های استخراج ویژگی به اشتراک گذاشته می‌شوند منظم می‌شوند. لایه‌های بعدی، "لایه‌های تطبیق" شناخته می‌شوند که شامل معیارهای اختلاف یا نمایشی از شبکه به همراه متریک اختلاف است. لایه‌های خاص وظیفه نقش زیادی در تطبیق ویژگی‌های منبع و هدف ندارند[14].

۱-۱-۳-۲. روش‌های اختلاف: مبتنی بر معیارها

سردرگمی دامنه عمیق[1] (DDC)[15] اولین ایده کلیدی بود که به طور مشترک وظایف (طبقه بندی) و سردرگمی دامنه را بهینه کرد. تلفیق دامنه برای تطبیق دو لایه متصل کننده با ایده استفاده شده است: ویژگی‌هایی که شبکه یاد می‌گیرد باید بی‌طرف باشند، یعنی باید در فضای ویژگی قرار بگیرند که در این شرایط اطلاعات دامنه از بین می‌رود در حالی که اطلاعات کلاس ها باقی می‌ماند. روش MMD[16] به عنوان یک معیار اختلاف برای ضرر دامنه استفاده می شود. DDC این تابع را به عنوان ضرر مشترک برای تطبیق دامنه معرفی می‌کند:

$$L_{\text{Domain_Adaptation}} = L_{\text{classification}} + \lambda\, MMD^2(P(X^s), P(X^t))$$

معادله ۱

معادله ۱ همچنین به ما کمک می‌کند تا درک کنیم که معیار اختلاف (مانند MMD یا مشابه آن) چگونه به عنوان یک تنظیم‌کننده برای کل شبکه عمل می‌کند. در ادامه، بسیاری از کارها بر ایده کلیدی DDC با استفاده از معیارهای اختلاف متفاوت یا مشابه ساخته‌اند.

[1] Deep Domain Confusion

Kashyap و همکارانش[17] معیار اختلاف مبتنی بر معیار را به سه دسته تقسیم کرده‌اند: دسته اول هندسی است که به فاصله بین بردارها اشاره دارد، دسته دوم اطلاعاتی-نظریه‌ای است که به فاصله بین توزیع‌های احتمال اشاره دارد و دسته سوم اندازه‌گیری‌های مرتبه‌بالا است که در آن از فاصله بین لحظه‌های بالاتر توزیع‌ها، یا فاصله بین پروژکشن‌ها یا حتی فاصله بین نمایش‌ها گفته می‌شود.

۲-۱-۳-۲. روش های اختلاف: مبتنی بر معماری (ساختار)

در این روش ها، تمرکز بیشتری روی یادگیری ویژگی ها و معماری قابل انتقال نسبت به معیارهاست. این روش‌ها به تطابق ویژگی‌ها و ساختارها در معماری مدل‌ها توجه می‌کنند. اصل اساسی در روش‌های تطابق بر مبنای ساختار این است که اطلاعات درباره تغییر دامنه (از دامنه مبدأ به دامنه مقصد) تنها به یک تبدیل تساوی مختصات (تبدیل آفین) بستگی دارد. به عبارت دیگر، وجود یک تبدیل کوچک بر روی وزن‌ها می‌تواند در فرآیند تبدیل ویژگی‌های دامنه مبدأ به دامنه مقصد کمک کند. این تبدیل کوچک می‌تواند تبدیل آفین یا شبکه‌های چند لایه‌ای (MLP/شبکه‌های عمیق) خود باشد.

یکی از مثال‌های این روش‌ها، شبکه انطباق عمیق (DAN) است که از اصلی استفاده می‌کند که در شبکه‌های عمیق پیچشی[1]، لایه‌های ابتدایی ویژگی‌های عمومی را درک می‌کنند در حالی که لایه‌های بعدی ویژگی‌های وظیفه‌محور را درک می‌کنند. در این روش، لایه‌های ابتدایی منجمد می‌شوند، لایه‌های میانی بهبود یافته و از روش‌های مبتنی بر اختلاف مانند MK-MMD (تغییرهای هسته‌ای چندگانه - MMD) برای تطبیق لایه‌های بعدی استفاده می‌شود. به طور معمول، روش‌های مبتنی بر اختلاف به تراز کردن توزیع‌های حاشیه‌ای داده‌های دامنه مبدأ و مقصد نگاه می‌کنند، اما رویکردهای مختلفی نیز وجود دارند،

یک مفهوم دیگر در این متن مربوط به شبکه‌های پروتوتایپی انتقال‌پذیر (TPN)[18] است که بر روی تفاوت‌ها (فواصل) برای هر کلاس در یک فضای تعبیه‌شده از سه مجموعه داده تمرکز دارد: فقط دامنه مبدأ، فقط دامنه مقصد و ترکیبی از دامنه مبدأ و مقصد. همچنین به نمونه‌های بدون برچسب دامنه مقصد، شبه برچسب‌ها را اختصاص می‌دهد. تطبیق طوری انجام می‌شود که پروتوتایپ هر کلاس در فضای تعبیه شده نزدیک یکدیگر باشند.

۲-۳-۲. روشهای خصمانه

ایده پشت روش های خصمانه این است که به همراه افزایش ابهام دامنه، همزمان به طور قوی آموزش داده می‌شوند تا تجزیه و تفکیک دامنه را درک کند. این ایده ارتباط نزدیکی با شبکه متخاصم

[1] Deep convolutional networks

مولد یا GANها[19] دارد که شامل دو بخش تولید کننده[1] و متمایز کننده[2] در یک محیط خصمانه است. هدف مولد تولید خروجی (معمولاً تصاویر) برای فریب دادن متمایزکننده است، در حالی که تمایزکننده، از سوی دیگر، سعی می‌کند تصاویر واقعی و مصنوعی را ازهم دیگر تفکیک کند.

در تطبیق دامنه (DA)، ایده‌ای که اقتباس می‌شود، این است که تمیز دهنده باید قادر باشد تا توزیع دامنه‌های منبع و هدف را با استفاده از ویژگی‌های نامتغیر دامنه از یکدیگر تفکیک کند. تطبیق دامنه متمایز کننده متخاصمانه[3] (ADDA)[20] یک چارچوب عمومی برای تطبیق دامنه با استفاده از مدل‌های متخاصم معرفی کرد.

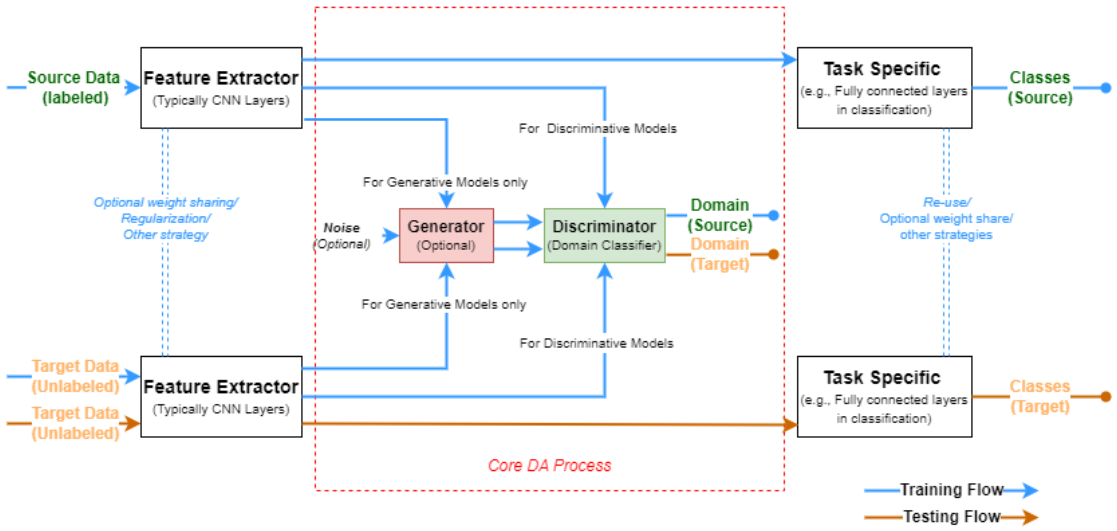
ADDA یک چارچوب عمومی برای تطابق داده‌ها از دو دامنه مختلف با استفاده از مدل‌های تخاصمی ارائه می‌دهد. شکل۵، این چارچوب را نشان می‌دهد.

معماری معمول برای تطابق داده‌های تخاصمی متمایزکننده، از یک معماری سیامی با دو شاخه داده استفاده می‌کند: یک شاخه از داده‌های منبع و یک شاخه از داده‌های هدف. این معماری با استفاده از تابع ضرر (معمولاً دسته‌بندی) و یا یک تابع ضرر تخاصمی یا مبتنی بر اختلاف، آموزش می‌بینند.

معماری تخاصمی تولیدی (در ساده‌ترین شکل خود) شامل یک مولد است که نگاشتی از یک دامنه به دامنه دیگر تولید می‌کند؛ سپس نگاشت تولیدی و سایر نگاشت‌ها به معماری متمایزکننده تخاصمی متصل می‌شوند و پس از بهبود یافتن با یکدیگر، نگاشت تولید شده و سایر نگاشت ها از معماری تبعیض آمیز تبعیت می کنند.

در کل، ADDA یک چارچوب عمومی برای تطابق داده‌ها از دو دامنه مختلف با استفاده از مدل‌های تخاصمی است، در این فرآیند داده‌ها از دامنه منبع به دامنه هدف منتقل می‌شوند و مدعی است که این چارچوب بهبودهای قابل توجهی در کارایی مدل در دامنه هدف ایجاد می‌کند.

[1] Generator
[2] Discriminator
[3] Adversarial Discriminative Domain Adaptation



شکل ۱۰ ساختار/معماری شبکه‌های عصبی عمیق که روش‌های مبتنی بر خصومت را برای تطبیق دامنه پیاده‌سازی می‌کنند. مولد(ها) اختیاری هستند و در صورت نیاز فقط برای ایجاد داده‌های مصنوعی استفاده می‌شوند. متمایزکننده دامنه[1]، به دنبال تمایز (طبقه بندی) دامنه های منبع و هدف است، در حالی که وظیفه خاص (مثلاً طبقه بندی یا تشخیص شی) توسط شبکه خاص وظیفه انجام می‌شود[14].

۱-۲-۳-۲. مدل های تمایزگر متخاصم

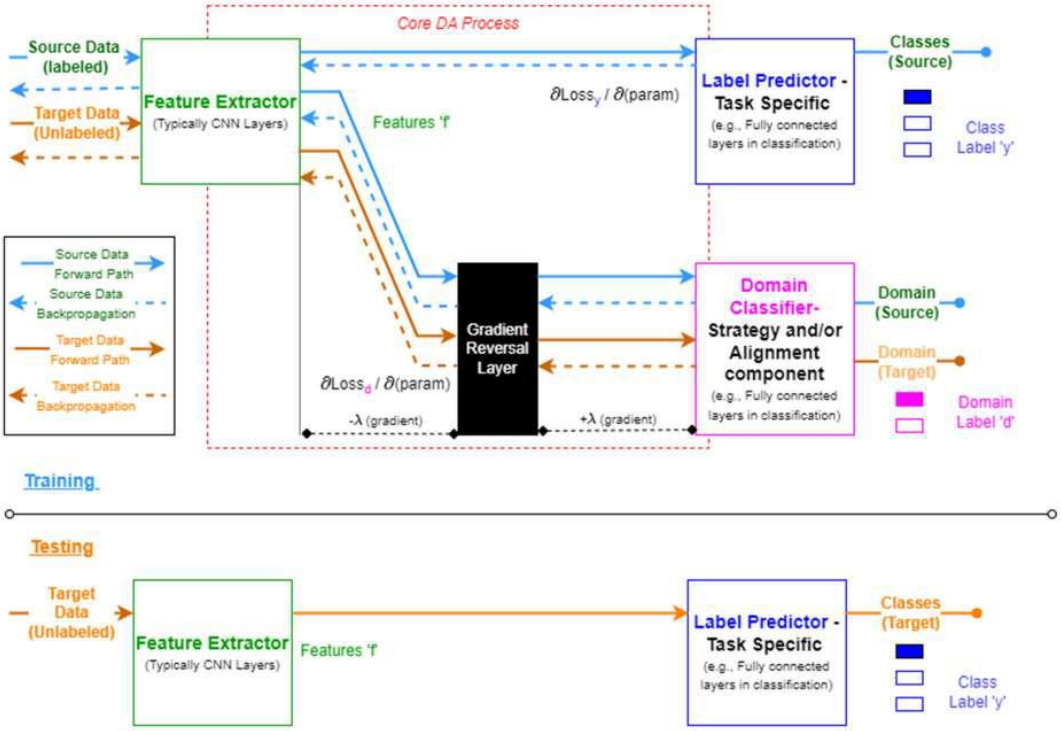
یکی از کارهای اساسی در حوزه تطبیق دامنه در یادگیری عمیق، Domain-Adversarial Neural Network (DANN) است. یادگیری دامنه باید تمایزدهنده باشد اما همزمان آن را فریب دهد. این مدل نشان داد که هر مدل feedforward می‌تواند اگر با یک لایه گرادیان معکوس[2] (GRL) نوآوری شده تقویت شود، تطبیق دامنه را پشتیبانی کند.

DANN گسترده‌ترین رویکرد تطبیق دامنه را در تمام حوزه‌های داده دارد. در ابتدا این روش در حوزه بینایی ماشین ابتدا برای تشخیص اعداد و دسته‌بندی تصاویر استفاده شد و سپس برای وظایف پیچیده‌تری مانند تقسیم‌بندی معنایی و تشخیص اشیاء نیز استفاده شد. در مورد تقسیم‌بندی معنایی، یک شبکه سیامی (که از دو مسیر موازی تشکیل شده است) استفاده می‌شود، که یک مسیر نمونه‌های منبع و مسیر دیگر نمونه‌های هدف را پردازش می‌کند. به دلیل پیچیدگی ذاتی وظایف، تطبیق دامنه (طبقه‌بند دامنه - شبکه صورتی در شکل ۶) در لایه‌ها/مراحل مختلف وجود دارد، و لایه‌های کانولوشن (ورودی به استخراج ویژگی) و لایه‌های دکانولوشن (استخراج ویژگی به نقشه معنایی[3]) هم‌راستا هستند (اشتراک دارند، نگاشت مشترک دارند یا معیار آماری استفاده می‌شود). Hoffman و همکارانش[21] به جز تابع ضرر اصلی، از دو تابع ضرر دیگر هم استفاده کردند؛ یک تابع ضرر برای سازگاری با پارامترهای

[1] Domain Discriminator
[2] Gradient Reversal Layer
[3] Feature map

مختص دسته، یعنی سازگاری خاص دسته و تابع ضرر دیگری برای کاهش فاصله توزیع جهانی، یعنی تطبیق جهانی دامنه. Huang و همکارانش[22] نیز به تطابق ویژگی‌ها در هر لایه شبکه نگاه کردند.

ADDA نیز از فلسفه ای مشابه با DANN استفاده می‌کند، اما با این تفاوت که استخراج ویژگی‌ها بین منبع و هدف به اشتراک گذاشته نمی‌شود، و تابع هزینه استفاده‌شده در ADDA تابع هزینه GAN است در حالی که DANN از تابع هزینه min-max استفاده می‌کند و آموزش چند مرحله‌ای است. شبکه‌های تخاصمی دامنه مشروط [23](CDAN) از یک تمییزدهنده شرطی استفاده می‌کنند که ورودی را از هردو استخراج کننده ویژگی و طبقه‌بندی‌کننده می‌گیرد. Shen و همکارانش[24] به جای استفاده از یک طبقه‌بند خالص در تمییزدهنده، از تابع ضرر به عنوان فاصله Wasserstein (مشابه Wasserstein GAN توسط Arjovsky و همکارانش[25]) در طول آموزش بین نمونه‌های منبع و هدف استفاده کردند.



شکل ۱۱ شبکه عصبی متخاصم دامنه‌ای (DANN) دو شبکه را همزمان آموزش می‌دهد. DANN یک استخراج‌کننده ویژگی (شبکه سبز) و پیش‌بینی‌کننده کلاس/برچسب (شبکه آبی) را روی داده منبع آموزش می‌دهد. DANN همچنین یک استخراج‌کننده ویژگی (شبکه سبز) و طبقه‌بندی‌کننده دامنه (شبکه صورتی) را بر روی داده منبع و هدف آموزش می‌دهد. لایه تغییر گرادیان (GRL) به شبکه پیش‌روی اجازه می‌دهد تا به عنوان معمول پیش برود؛ با این حال، در زمان پس‌انتشار، گرادیان از تمییز دامنه تغییر می‌کند (برعکس می‌شود اما با یک مقدار منفی ضرب

می‌شود) که منجر به درک ویژگی‌های ثابت دامنه (سردرگمی دامنه) توسط استخراج‌کننده ویژگی (شبکه سبز) می‌شود. پارامتر λ به یادگیری ویژگی‌های دسته‌بندی کمک می‌کند و به تدریج ویژگی‌های دامنه را یاد می‌گیرد[14].

۲-۲-۳-۲. مدل های مولد متخاصم

مدل‌های تولیدی متخاصم شامل یک بخش مولد (معمولاً یک مولد در یک GAN) به همراه بخش تمایزدهنده‌ای از مدل‌های تمایزدهنده هستند. این مولد به طور معمول داده‌های هدف مصنوعی را از داده‌های منبع با برچسب تولید می‌کند. این داده‌های هدف مصنوعی باعث کاهش نیاز به نمونه‌های دارای برچسب در دامنه‌های هدف می‌شوند. سپس شبکه آموزش داده می‌شود که فرض کند که تغییرات دامنه کم یا هیچ‌گونه تأثیری روی داده‌های مصنوعی ندارد. جزء نگاشت منبع، مولدی است که دامنه منبع را به دامنه هدف نگاشت می‌کند. بنابراین به طور عام، این مولدها به نام دامنه‌نگارها هم شناخته می‌شوند. یکی از کارهای اولیه در زمینه مدل‌های مولد متخاصم شبکه مولد متخاصم مشترک (CoGAN) می‌باشد[26]. همانطور که از نامش پیداست، دو GAN به طور موازی اجرا می‌شوند و در لایه‌های ابتدایی برای مولدها و لایه‌های نهایی برای تمایزدهنده‌ها به اشتراک گذاشته می‌شود. این لایه‌ها ویژگی‌های سطح بالا در تمایزدهنده‌ها و مفاهیم سطح بالا در مولدها را ثبت می‌کنند. این کار کمک می‌کند تا GAN توزیع مشترک دامنه‌ها را درک کند. در CoGAN، دامنه هدف به دامنه منبع تبدیل شده و سپس دسته‌بندی انجام می‌شود.

به طور معمول، تنها برای یک کار، تطبیق دامنه اختصاصی است(در دو دامنه به اشتراک گذاشته می‌شود). با این حال، PixelDA[27] یک تنظیم تطبیق دامنه تولیدی متخاصم را استفاده کرد تا چارچوبی را ارائه دهد که از جزئیات مربوط به کار جدا شده باشد. به طور معمول، تصاویر منبع به تصاویر شبیه به هدف تبدیل می‌شوند؛ با این حال، Generate-to-adapt[28] از GAN برای تطبیق دامنه با ایجاد تصاویر شبیه به منبع برای موارد دامنه هدف استفاده می‌کند. این از تعبیرهای (در طول آموزش) تصاویر به عنوان فضای نهفته به عنوان ورودی فرعی به GAN استفاده می‌کند تا تصاویر شبیه به منبع را از مولد و تمایزدهنده ایجاد کند که دامنه واقعی را از مصنوعی را تمایز می‌دهد و برچسب‌های کلاس را ارائه می‌دهد.

۲-۲-۳-۳. روش‌های مبتنی بر بازسازی تخاصمی

نوع دیگری از روش‌های تولیدی تخاصمی، روش‌های مبتنی بر بازسازی هستند (به همان خطوط راهبرد تطبیق دامنه با تطبیق ویژگی‌های سطح کم عمق): روش‌های بازسازی معمولاً از شبکه‌های GAN یا شبکه‌های رمزگذار خودکار[1] (AE) استفاده می‌کنند تا محتوای یک دامنه را با سبک دامنه دیگر بازسازی کنند. در ادامه به بررسی ایده‌های کلیدی پشت روش‌های بازسازی تخاصمی می‌پردازیم.

[1] AutoEncoder

شبکه جداساز دامنه[1] (DSN)[29] یک چارچوب مبتنی بر AE است. AE یک مدل یادگیری عمیق است که برای تبدیل ویژگی‌های ورودی به یک نمایش مخفی و سپس بازیابی ویژگی‌های ورودی از این نمایش استفاده می‌کند. این چارچوب سعی دارد تا ویژگی‌ها و نمایش‌هایی که ویژگی‌های خصوصی یک دامنه خاص هستند را درک کند، همچنین ویژگی‌ها و نمایش‌هایی که در چندین دامنه به اشتراک گذاشته می‌شوند. نمایش‌های اشتراکی و نمایش‌های خصوصی در این چارچوب به گونه‌ای طراحی شده‌اند که مستقل از یکدیگر هستند و با یکدیگر تداخل ندارند. همچنین قادر است اطلاعات مختص به هر دامنه را از اطلاعاتی که در چندین دامنه به اشتراک گذاشته می‌شوند جدا کند.

این چارچوب از یک معیار به نام "reconstruction loss" (تلفیق با سایر معیارها) استفاده می‌کند که شبکه رمزگشا را مجبور به بازسازی مثال ورودی اصلی با استفاده از ویژگی‌های موجود در نمایش‌های خصوصی و اشتراکی دامنه مربوطه می‌کند.

CycleGAN[15] یک نوع مدل مبتنی بر GAN است که به منظور ترجمه تصاویر از یک دامنه مبدا به دامنه مقصد (ترجمه تصویر به تصویر) استفاده می‌شود. هدف اصلی این مدل این است که تصویر مبدا را به دامنه مقصد ترجمه کرده و سپس تلاش کند تا تصویر مبدا را بازسازی کند، به طوری که تفاوت‌ها و تغییرات ناشی از ترجمه را تا جای ممکن کاهش دهد.

برای انجام این کار، CycleGAN از مفهوم "Cycle consistency" استفاده می‌کند. یعنی اگر یک تصویر از دامنه مبدا به دامنه مقصد ترجمه شود و سپس باز به دامنه مبدا بازگردانده شود، تصویر بازگردانده شده باید به تصویر اصلی مبدا شباهت داشته باشد؛ یعنی ترجمه و بازگردانی باید به اندازه کافی دقیق باشد تا تصاویر نهایی به یکدیگر متناظر باشند و به نوعی اطمینان می‌یابیم که اطلاعات معنایی در طول ترجمه از دست نرفته است.

یکی از نکات مهم در CycleGAN این است که برخلاف بسیاری از روش‌های ترجمه تصویر به تصویر دیگر، نیازی به داشتن تصاویر جفت (تصاویر متناظر در دو دامنه) نیست. این به معنای این است که برای آموزش این مدل، مجموعه‌ای از تصاویر از دامنه مبدا و دامنه مقصد بدون نیاز به تصاویر متناظر در دو دامنه مختلف کافی است. این ویژگی به عنوان یکی از مزایای این روش مطرح می‌شود.

۲-۳-۳. تطابق چند دامنه ای

تطابق چند دامنه ای[2] در واقع تفاوت‌هایی با مدل‌های تطبیق دامنه معمولی دارد. در تطبیق دامنه معمولی، معمولاً تنها یک دامنه منبع و یک دامنه مقصد وجود دارد که مدل باید از حوزه منبع به حوزه مقصد تطابق پیدا کند. اما در تطابق چند حوزه‌ای، دو حالت ممکن است رخ دهد:

[1] Domain Separation Networks
[2] Multi Domain Adaptation

تطابق چند منبعی[1]: در این حالت، تعداد حوزه‌های منبع بیشتر از یکی است. به عبارت دیگر، مدل باید از چندین حوزه منبع به حوزه مقصد تطابق پیدا کند. این تطابق می‌تواند در مواقعی کاربرد داشته باشد که داده‌های مختلفی از منابع مختلف برای تطابق در یک حوزه مقصد مشترک استفاده می‌شوند.

تطابق چند هدفه[2]: در این حالت، تعداد حوزه‌های مقصد بیشتر از یکی است. این بدان معناست که مدل باید از یک حوزه منبع به چند حوزه مقصد متفاوت تطابق پیدا کند. این مفهوم زمانی به کار می‌رود که یک مدل یکپارچه با چندین حوزه مقصد متفاوت نیاز دارد و باید توانایی تطابق با همه این حوزه‌ها را داشته باشد.

۲-۳-۳-۱. تطابق چند منبعی

به منظور ایجاد مدل‌های تطبیقی دامنه‌ای قوی‌تر، منطقی است که مدل‌ها را بر روی چند منبع آموزش دهیم. در مطالعات اولیه (قبل از یادگیری عمیق)، Sun، Shi و Wu[30] اشاره کرده‌اند که می‌توان یک طبقه‌بند فردی را بر روی دامنه‌های منبع فردی و یک دامنه هدف آموزش داد و سپس طبقه‌بندهای پایه را یا تمام منابع را به عنوان یک منبع یکپارچه در نظر گرفت و آموزش داد. همچنین برای یادگیری عمیق، Zhao و همکاران[31] به استفاده از استراتژی‌های اختلاف، تخاصمی و تطابق ویژگی اشاره کرده‌اند.

یکی از استراتژی‌ها و یا حوزه‌هایی که برای تطبیق استفاده شده است، تولید دامنه میانی برای تطبیق است. در این حالت، یک دامنه با استفاده از مولفه‌های دامنه (معمولاً بر اساس GAN) ایجاد می‌شود.

۲-۳-۳-۲. تطابق چند هدفه

به طور معمول، تطبیق دامنه از روش pairwise پیروی می‌کند، یعنی دامنه منبع به دامنه هدف متصل می‌شود و تلاش می‌شود تا مدلی که در دامنه منبع آموزش دیده است، به دامنه هدف منتقل شود.

با الهام از روش ذکر شده در مقاله[32]، Gholami و همکاران[33] به دنبال یافتن اطلاعات مشترک بین دامنه‌ها هستند. آن‌ها رویکردی به نام "روش اطلاعاتی تطبیق چند هدفه" یا MTDA-ITA ارائه می‌دهند. این روش از فضاهای خصوصی و مشترک بین ترکیب‌های دامنه‌های منبع و هدف استفاده می‌کند، مشابه روش DSN[29]. همچنین Isobe و همکاران[34] از تطابق چند هدفه برای وظایف تقسیم بندی معنایی با استفاده از مدل‌های مبدأ-مقصد و پل‌های مجزایی که بین جفت‌های مدل مبدأ-مقصد ایجاد شده‌اند برای همکاری استفاده کردند. یک مدل دانش‌آموز بر اساس تمام جفت‌های مدل

[1] Multi-Source Adaptation
[2] Multi-Target Adaptation

مبدأ-مقصد فردی با استفاده از تنظیمات متغیر در هر جفت مدل مبدأ-مقصد فردی یادگرفته می‌شود. دانش مشابهی نیز در یادگیری توجه چند معلمه چند مقصدی (MT-MTDA) درک شده‌است[35].

۲-۳-٤. روش‌های ترکیبی

در این بخش به ادغام تکنیک‌های متعددی که قبلاً برای تطبیق دامنه مورد بحث قرار گرفته است، می‌پردازیم.

۲-۳-۴-۱. روش‌های مبتنی بر تجمیع

روش‌های مبتنی بر تجمیع[1] شامل چندین مدل مختلف هستند، که خروجی چندین مدل با هم ترکیب می‌شود. به طور معمول در وظایف رگرسیون از میانگین‌گیری، و در وظایف طبقه‌بندی از رای‌گیری استفاده می‌شود. تنوع مدل ها باعث می شود که انحراف از میزان درستی آن زیاد نباشد. یکی از مشکلات مهم این روش‌ها این است که از نظر محاسباتی پرهزینه هستند.

روش‌های تجمیع برای تطبیق دامنه به دو زیر روش تقسیم می‌شوند:

- **تجمیع با برچسب‌گذاری مصنوعی**

در این روش، برچسب‌های دامنه هدف بر اساس دیدگاه ترکیب مدل‌های مختلف تعیین می‌شوند. اگر اکثر مدل‌ها به یک نتیجه مشابه برسند و اطمینان زیادی در مورد کلاس برچسب برای یک مثال خاص در دامنه هدف وجود داشته باشد، شبه برچسب‌ها برای آن مثال تعیین می‌شود. در واقعیت، یک مثال از دامنه هدف (نه منبع) برای آموزش طبقه‌بند هدف استفاده می‌شود.

- **تجمیع خودتشکیلی**

در این روش، ترکیب خروجی‌ها بر روی خروجی‌های یک مدل تکراری در طول زمان انجام می‌شود. ترکیب خروجی‌ها در طول زمان همچنین به عنوان تجمیع زمانی[2] نیز شناخته می‌شود. به عبارت دیگر، اطلاعات زمانی نیز به ترکیب خروجی‌ها اضافه می‌شود.

۲-۳-۵. تطبیق دامنه چند حالته

تطبیق دامنه چند حالته[3]، به معنای تطبیق مدل‌های یادگیری ماشینی در مقابل دامنه‌های مختلفی است که داده‌های آنها دارای حالت‌های متفاوتی هستند. این داده‌ها معمولاً از چندین منبع ورودی به شکل‌های مختلفی تشکیل شده‌اند، مثلاً تصاویر و متون ممکن است با یکدیگر ترکیب شوند.

[1] Ensemble-Based Methods
[2] temporal ensembling
[3] Multi-Modal Domain Adaptation

بزرگترین چالش در تطبیق دامنه چند حالته این است که فرآیند تطبیق دامنه باید ساختارهای مختلف حالت و تغییرات مختلف دامنه برای هر حالت را در نظر بگیرد. به عبارت دیگر، هنگامی که با داده‌های چند حالته سر و کار داریم، لازم است که مدل‌ها و الگوریتم‌های تطبیق دامنه به اختلاف‌ها و تفاوت‌هایی که در ساختار حالت‌ها و تغییرات دامنه وجود دارد، توجه داشته باشند.

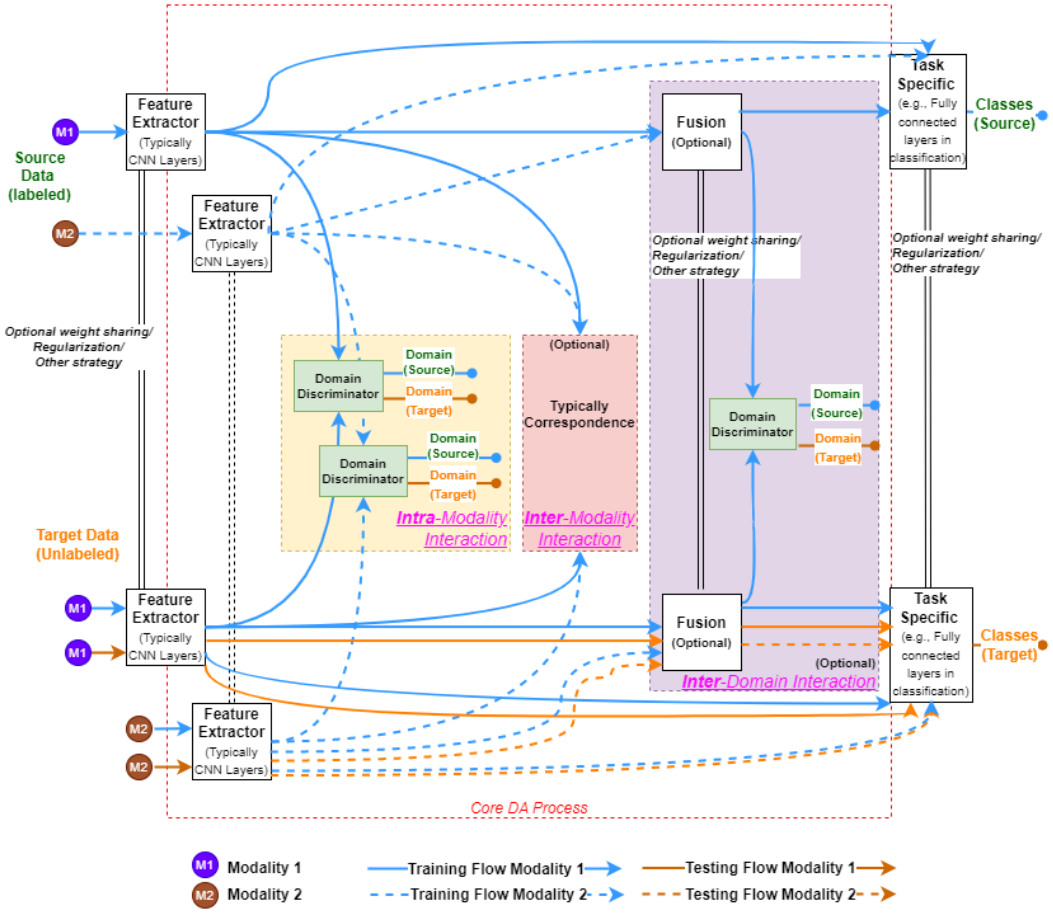
در مواجهه با تطبیق دامنه چند حالته، فرآیند تطبیق دامنه باید همچنین به فضاهای ویژگی مختلف، نمایش‌های مختلف و ابعاد مختلف فضای ویژگی داده‌ها نیز توجه کند. به عبارت دیگر، ویژگی‌های مختلف (مانند ویژگی‌های تصویری و ویژگی‌های متنی) باید با چالش تطبیق دامنه مرتبط شوند و مدل‌ها باید بتوانند از هر دو حالت بهره ببرند تا در محیط‌های مختلف به خوبی عمل کنند.

۱-۵-۳-۲. تطبیق دامنه چندحالته یکپارچه

بیشتر کارهای مرتبط با تطبیق دامنه، داده‌های یکپارچه را پشتیبانی می‌کنند، و این بدین معناست که فضای ویژگی ثابت می‌ماند ($X^s = X^t$)، اما تغییر به دلیل توزیع‌های متفاوت داده‌ها اتفاق می‌افتد، به عبارت دیگر $P(X^s) \neq P(X^t)$. وقتی که هر دو دامنه منبع و مقصد حداقل دو حالت داده داشته باشند، به عبارت دیگر چندگانه باشند، اما همچنان فضای ویژگی (ویژگی‌های ورودی برای مسئله) یکسان باشد، حالت تطبیق دامنه چندگانه یکپارچه رخ می‌دهد. این شبکه‌ها توانایی ارتباط داخل حالتی را بطور اجباری اجرا می‌کنند، اما مشاهده شده است که پیاده‌سازی ارتباطات میان حالتی و میان دامنه اختیاری است.

Xu,Yang و همکارانش[36] یک شبکه تطبیق دامنه چندگانه را با ماژول‌های توجه و ادغام به همراه محدودیت‌های دامنه ترکیبی ایجاد کردند تا ویژگی‌های دامنه‌ای ثابت را یاد بگیرند. ماژول توجه به درک ارتباط بین حالت‌ها کمک می‌کند و مدل دوخطی[1] برای ادغام استفاده می‌شود.

[1] bilinear model



شکل ۱۲ ساختار/معماری شبکه معمولی برای شبکه‌های عصبی عمیق که اجرای تطبیق دامنه چند منظوره همگن را انجام می‌دهند. در بلوک درون دامنه ای[1]، تمیزدهنده‌ها ویژگی بردارها را مجبور می‌کنند تا ویژگی‌های بی‌تغییر در دامنه را یاد بگیرند. اما دو بلوک بین حوزه ای[2] و بین دامنه ای[3] اختیاری هستند و در هر تنظیم تطبیق دامنه چند منظوره ممکن است مشاهده نشوند[14].

۲-۵-۳-۲. تطبیق دامنه چند حالته ناهمگن[4]

در دنیای واقعی، داده‌های چند حالته و ناهمگن بسیار شایع هستند. با توجه به استفاده روزافزون از شبکه‌های عمیق در پردازش این داده‌های چند حالته و ناهمگن، یادگیری تطبیق دامنه در محیط‌های چند حالته و ناهمگن بسیار مهم است.

در تطبیق دامنه ناهمگن، معمولاً از شبکه‌های جداگانه برای استخراج ویژگی‌های دو دامنه مختلف استفاده می‌شود. سپس، اطلاعات ویژگی‌های مشترک بین دو دامنه به سطح وظیفه منتقل می‌شود، و این انتقال می‌تواند به دو صورت صورت بگیرد: در حالت اول وزن‌ها به صورت قوی(کامل) بین دو دامنه

[1] Intra Modality
[2] Inter Modality
[3] Inter Domain
[4] Heterogeneous Multi-Modal Domain Adaptation

به اشتراک گذاشته می‌شوند[1] و در حالت دوم وزن‌ها به صورت ضعیف به اشتراک گذاشته می‌شوند[2]، که در کار Shu و همکارانش[37] ذکر شده است.

اهمیت تطبیق دامنه چند حالته ناهمگن زمانی مشخص می‌شود که یکی از حالت‌ها (مثلاً یکی از داده‌ها) در دامنه هدف وجود نداشته باشد. به عنوان مثال، اگر دامنه منبع دارای دو حالت m1 و m2 باشد ولی دامنه هدف تنها شامل m3 باشد و m4 را نداشته باشد، مسئله‌ی فقدان حالت[3] پیش می‌آید. این مسئله توسط Ding و همکارانش[38] مورد بررسی قرار گرفته و روشی با عنوان یادگیری انتقال فقدان حالت از طریق محدودیت نرم‌افزاری نهان [4](M2TL) ارائه شده است. انتقال یادگیری در اینجا دوجانبه است: انتقال از یک پایگاه داده به پایگاه داده دیگر (انتقال میان پایگاه داده) و انتقال از حالت منبع به حالت هدف (انتقال میان حالت). در این روش از محدودیت ماتریس کم‌رتبه[5] برای یادگیری زیرفضای داده در چارچوبی مختلف برای حالت‌های مختلف و با استفاده از MMD متغیرهای ویژگی بین پایگاه داده‌ها در دامنه منبع (حالت‌های مشخص شده) را به اشتراک می‌گذارند.

از جمله روش‌های مختلفی که در تطبیق دامنه چند حالته ناهمگن مورد استفاده قرار می‌گیرند میتوان به استفاده از شبکه‌های متخاصم دامنه مشروط[6] (CDAN)[23] که توانایی تطبیق به توزیع‌های داده چند حالته دارند، اشاره کرد. این شبکه‌ها توسط مدل‌های مقابله‌ای ساخته می‌شوند و از اطلاعات تمایزدهنده برای تطابق با دامنه‌های مختلف استفاده می‌کنند. در اینجا، متمایزکننده بر اساس ماتریس ترکیبی داده‌های ویژگی خاص به هر دامنه و پیش‌بینی‌های طبقه‌بندی استفاده می‌شود. این رویکردها معمولاً برای مسائلی با فرآیندهای مختلف و داده‌های مختلف (مثلاً تصویر و صدا) مورد استفاده قرار می‌گیرند.

۲-۳-۶. روش‌های نوظهور

در این بخش، به تکنیک‌ها و مدل‌هایی اشاره می‌کنیم که در ادبیات موجود در زمینه تطبیق دامنه جای نمی‌گیرند. این تکنیک‌ها و مدل‌ها، توجه زیادی به خود جلب کرده‌اند و به نوعی نسبت به تنظیمات عملی بیشتری و/یا چالش‌های واقعی جهان واقعی تطبیق یافته‌اند.

[1] strong parameter sharing weights
[2] weakly parameter-shared weights
[3] Missing Modality Problem
[4] Missing Modality Transfer Learning
[5] low-rank matrix
[6] conditional domain adversarial networks

۲-۳-۶-۱. تطبیق دامنه با داده‌های کم

چالش اصلی در تطبیق دامنه با داده‌های کم[1] این است که تعداد داده‌های هدف کافی نیست تا به طور قاطع به نیازهای همزمان تطبیق دامنه پاسخ دهد. این نیازها شامل تبدیل دامنه و تطبیق نمایش بین دو دامنه می‌شوند.

۲-۳-۶-۲. تطبیق دامنه بدون داده

تطبیق دامنه بدون داده[2] به وضعیتی اشاره دارد که در آن داده‌های واقعی دامنه هدف در زمان آموزش حاضر نیستند، تنها برخی اطلاعات مرتبط با آن (معمولاً ابرداده‌های هدف) در دسترس است. تطبیق دامنه بدون داده از تطبیق دامنه معمولی متمایز می‌شود، زیرا در تطبیق دامنه معمولی هیچ اطلاعاتی در مورد داده‌های هدف موجود نیست، حتی ابرداده‌ها.

یادگیری بدون داده با استفاده از داده‌های مرتبط با وظیفه: Peng و همکارانش[39] در وظیفه‌های بینایی کامپیوتر از اطلاعات موجود در داده‌های مرتبط با وظیفه (جفت‌های دامنه) برای کمک به درک اطلاعات شبکه در مورد دامنه هدف که در زمان آموزش در دسترس نیست، استفاده می‌کنند.

یادگیری بدون داده با برچسب‌های جدید در دامنه هدف: هدف در اینجا این است که برچسب‌های کلاس‌های "متفاوت" در دامنه هدف را بر اساس برچسب‌های موجود در منبع یاد بگیریم. این موضوع در واقعیت یک سناریوی تطبیق دامنه نیست، زیرا دامنه برچسب در هر دو منبع یک مثال ذکر شده در مقاله کودیروف و همکارانش[40] برچسب "خرس قطبی" را می‌توان با استفاده از بردارهای تعبیه "پوست دارد"، "سفید است" و "ماهی می‌خورد" نمایان کرد. هر تعبیه معنایی که به این بردارهای تعبیه نزدیک باشد، به کمک در برچسب‌گذاری مؤثر خواهد بود.

[1] Few-Shot Domain Adaptation
[2] Zero-Shot Domain Adaptation

فصل ۳: مبانی تجربی تحقیق

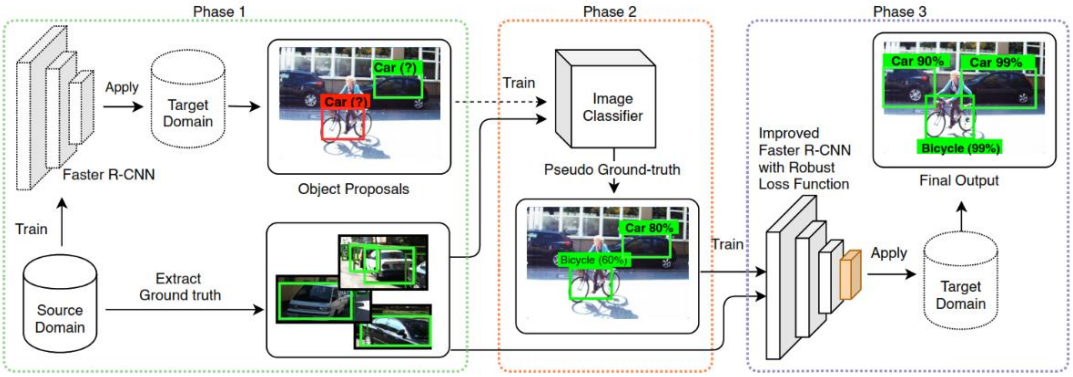
۳-۱. مقدمه

در این فصل در ابتدا به بررسی مقاله‌ها و پایان نامه‌های انجام شده در زمینه تطبیق دامنه برای تشخیص اشیا که الگوریتم های پایه آن در بخش مبانی نظری بررسی شده اند، می‌پردازیم. سپس مجموعه دادگان استفاده شده در این زمینه را بررسی کرده و در نهایت، نتایج مقاله ها و پایان نامه های ذکر شده را مشاهده می‌کنیم.

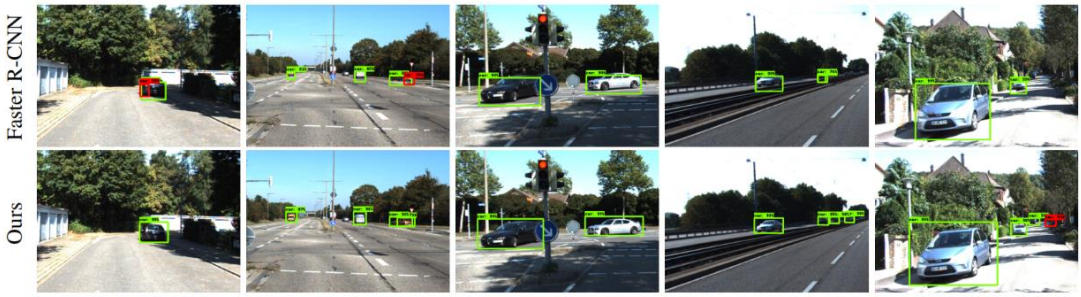
۳-۲. تطبیق دامنه برای تشخیص اشیا

۳-۲-۱. مقالات

Khodabandeh و همکارانش[41] یک رویکرد یادگیری قوی برای تشخیص اشیاء با تطبیق دامنه ارائه داده‌اند. نویسندگان مسئله را به عنوان آموزش با برچسب‌های نویزی فرموله کرده‌اند. بر اساس مجموعه‌ای از جعبه‌های محدود نویزی که توسط یک مدل تشخیص تنها در دامنه مبدأ آموزش داده شده‌اند، مدل نهایی تشخیص آموزش داده می‌شود.

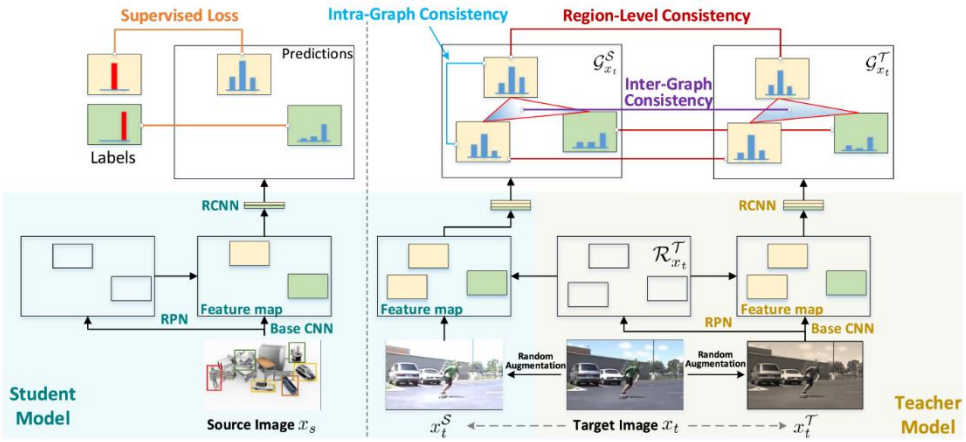


شکل ۱۳ ساختار روش Faster-RCNN Robust. این شبکه از مدل آموزش دیده در منبع برای به دست آوردن برچسب‌های نویزی در مجموعه داده دامنه مقصد استفاده می‌کند. یک ماژول دسته‌بندی برای به‌روزرسانی برچسب‌های نویزی استفاده می‌شود و کیفیت آنها را بهبود می‌بخشد. در نهایت، مدل تشخیص بر روی داده‌های دامنه مقصد با کمک یک تابع ضرر قوی آموزش می‌بیند که عدم قطعیت را جهت در نظر گرفتن احتمال اشتباهات ممکن در برچسب‌های نویزی شبه حاصل از مدل آموزش دیده در منبع می‌گیرد[41].



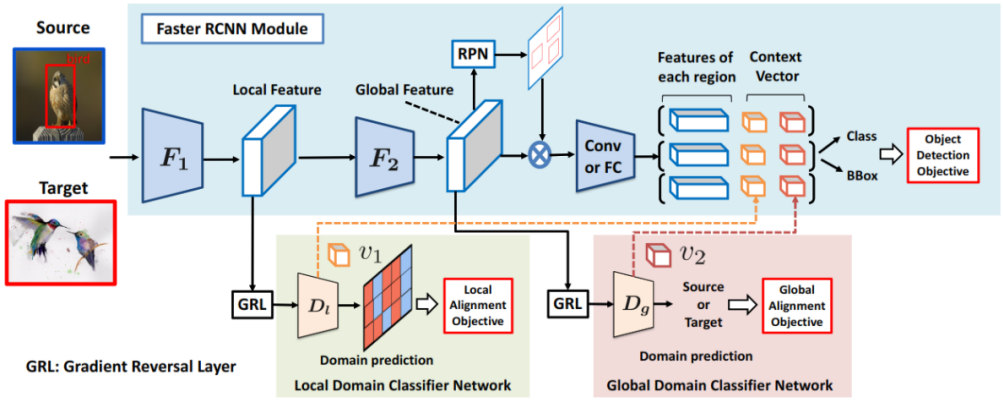
شکل ۱۴ مقایسه کیفی روش فوق با Faster R-CNN در آزمایش Cityscapes → KITTI[41]

Cai و همکارانش[42] رویکرد Mean Teacher را برای تشخیص اشیاء در دامنه‌های متفاوت بهبود داده‌اند و روش Mean Teacher با روابط اشیاء (MTOR) را ارائه داده‌اند. این روش به طور نوآورانه Mean Teacher را تحت تأسیس Faster R-CNN تغییر می‌دهد و روابط اشیاء را با محاسبه هزینه سازگاری ادغام می‌کند.



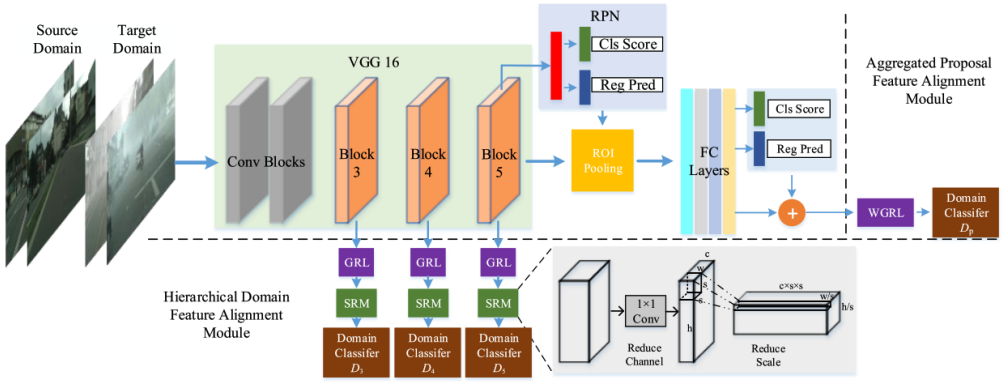
شکل ۱۵ مروری بر معماری "Mean Teacher with Object Relations (MTOR)" برای تشخیص اشیاء در دامنه‌های مختلف، با مدل‌های معلم و دانش‌آموز تحت یک backbone مشترک از Faster R-CNN[42].

Kuniaki Saito و همکارانش[43] یک رویکرد جدید برای تطبیق بدون نظارت تشخیص‌دهنده‌های اشیاء از دامنه‌های دارای برچسب به دامنه‌های کم‌برچسب پیشنهاد می‌دهد که می‌تواند هزینه‌های برچسب‌گذاری مرتبط با تشخیص را به طرز قابل توجهی کاهش دهد. روش پیشنهادی بر تطبیق محلی قوی و تطبیق جهانی ضعیف تمرکز دارد و در چندین مجموعه داده با میزان بسیار بالایی از روش‌های موجود ارتقاء می‌یابد. مدل تطبیق ضعیف برای تمرکز بر تلفیق نقضی روی تصاویری طراحی شده است که به طور جهانی مشابه هستند و کمتر تأکیدی بر تطبیق تصاویری دارند که به طور جهانی متفاوت هستند. مدل تطبیق دامنه قوی برای تنها نگاه کردن به میدان‌های محلی از نقشه ویژگی طراحی شده است. روش پیشنهادی در صورتی که دو دامنه متفاوت و مشابه باشند، تأثیربخش در مواقع تطبیق اثبات شده است.



شکل ۱۶ معماری شبکه پیشنهادی. این روش با استفاده از یک شبکه تشخیص دهنده دامنه محلی و یک شبکه تشخیص دهنده دامنه جهانی تطبیق قوی-محلی انجام می‌دهد و تطبیق ضعیف-جهانی را به انجام می‌رساند. بردار متناظر از طریق تشخیص دهنده دامنه استخراج می‌شود و در لایه قبل از لایه کاملاً متصل نهایی ادغام می‌شود[43].

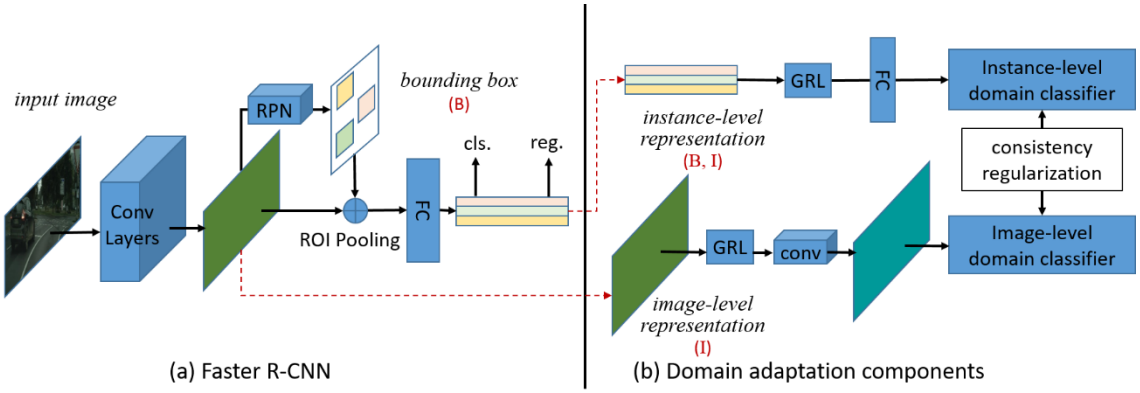
Zhenwei He و همکارانش[44] یک روش جدید برای تشخیص اشیاء در محیط‌های نا محدود با بهره‌گیری از دانش دامنه آموزش داده شده، از یک دامنه منبع فرعی پیشنهاد شده است. چارچوب multi-adversarial Faster-RCNN (MAF) پیشنهادی را برای به حداقل رساندن اختلاف دامنه برای تطبیق دامنه در نمایش ویژگی بیان می‌کند و شامل یک ماژول تطبیق ویژگی دامنه سلسله مراتبی، یک ماژول کاهش مقیاس اطلاعات ثابت (SRM) و یک ماژول تطبیق ویژگی پیشنهاد جمع شده می‌شود. تشخیص دهنده MAF می‌تواند به صورت تماماً انتها به انتها با بهینه‌سازی تابع از دست دادن تطبیق دامنه و تابع از دست دادن تشخیص Faster-RCNN آموزش داده شود. آزمایشات انجام شده بر روی چندین مجموعه داده با دامنه‌های مختلف نشان می‌دهد که رویکرد پیشنهادی نتایج برتری دارد و اثربخشی مدل را تایید می‌کند. به طور کلی، این مقاله یک راه‌حل موثر برای مسئله چالش‌برانگیز تشخیص اشیاء در محیط‌های نامحدود ارائه می‌دهد.



شکل ۱۷ ساختار شبکه MAF . این شبکه الهام گرفته از Faster-RCNN مبتنی بر VGG16 است، MAF ماژول‌های تطبیق ویژگی را هم بر روی ویژگی‌های دامنه و هم بر روی ویژگی‌های پیشنهاد اعمال می‌کند. برای ماژول تطبیق ویژگی دامنه سلسله مراتبی، چندین زیرماژول تطبیق کننده دامنه متخاصم روی بلوک ۳، ۴، ۵ VGG16

پیاده‌سازی شده است. لایه‌های GRL برای استراتژی یادگیری تخاصمی استفاده می‌شوند و اندازه نقشه‌های ویژگی توسط ماژول‌های کاهش مقیاس SRM کاهش می‌یابد. در ماژول تطبیق ویژگی، امتیازهای دسته‌بندی و نتایج بازیابی جعبه مرتبط با ویژگی‌ها در دامنه دسته‌بند ترکیب شده است و این در حالی است که WGRL برای استراتژی یادگیری تخاصمی معرفی می‌شود. SRM از دو بخش تشکیل شده است، بخش اول یک لایه پیکسل‌های کانوولوشن ۱×۱ است که برای کاهش اندازه کانال اعمال می‌شود. بعد از آن، بخش کاهش مقیاس برای ترکیب ویژگی‌های مجاور s×s استفاده می‌شود تا اندازه نقشه‌های ویژگی کاهش یابد[44].

Yuhua Chen و همکارانش[45] یک رویکرد جدید برای بهبود انطباق دامنه‌ای و کاهش آسیب‌پذیری در تشخیص اشیاء ارائه می‌دهند، به نام Faster R-CNN تطبیق دامنه‌ای. چالش اصلی در تشخیص اشیاء هنگامی که داده‌های آموزش و آزمون از توزیع‌های مختلفی گرفته شوند، تغییر دامنه است. برای حل این مشکل، رویکرد پیشنهادی دو مؤلفه تطبیق دامنه را در سطح تصویر و سطح نمونه به مدل Faster R-CNN اضافه می‌کند تا کمینه کردن همگرایی H بین دو دامنه را انجام دهد. در هر مؤلفه، یک طبقه‌بند دامنه آموزش داده می‌شود و استراتژی آموزش دشمن‌گرایی برای یادگیری ویژگی‌های مقاوم در برابر دامنه‌های مختلف استفاده می‌شود که دامنه-ثابت باشند. یک تنظیم مرتب‌سازی همچنین بین طبقه‌بند دامنه‌ها در سطوح مختلف جهت یادگیری یک شبکه پیشنهاد منطقه (RPN) ثابت دامنه در مدل Faster R-CNN اضافه می‌شود. رویکرد پیشنهادی بر روی چندین مجموعه داده ارزیابی می‌شود و کارآیی خود را در تشخیص قوی اشیاء اثبات می‌کند. مشارکت‌های این کار شامل تحلیل تئوری مشکل تغییر دامنه، طراحی دو مؤلفه تطبیق دامنه برای کاهش اختلاف دامنه در سطح تصویر و سطح نمونه، تنظیم مرتب‌سازی جهت تشویق RPN به تبدیل‌شدن به ثابت دامنه و ادغام مؤلفه‌های پیشنهادی در مدل Faster R-CNN است که می‌تواند به صورت کامل آموزش داده شود.



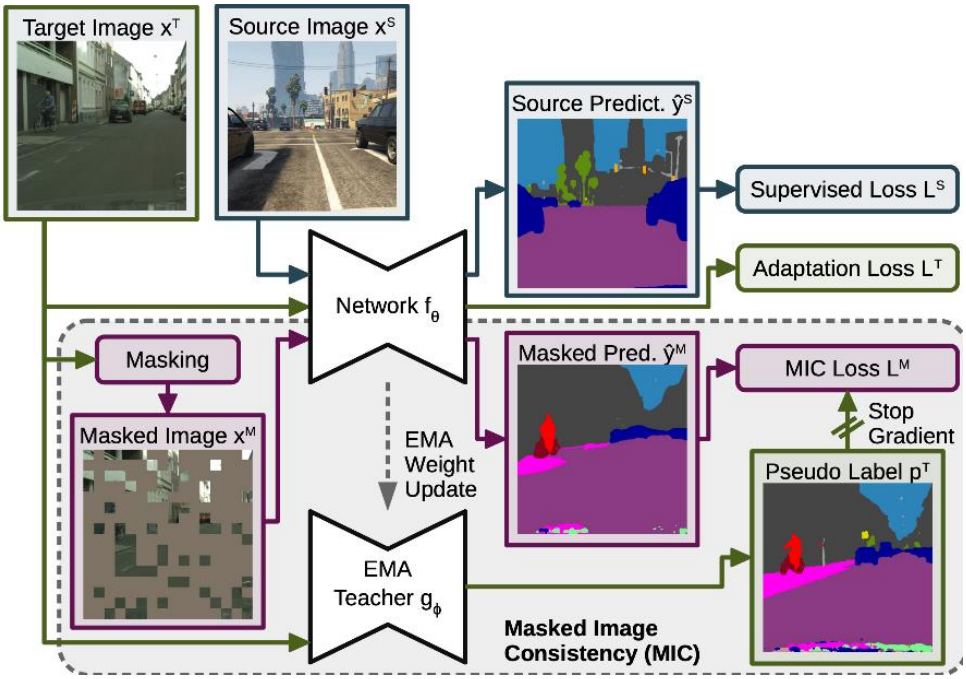
شکل ۱۸ یک نمای کلی از مدل Domain Adaptive Faster R-CNN. در این روش با دو سطح، سطح تصویر و سطح نمونه، تغییر دامنه مورد بررسی قرار داده شده است. در هر سطح، یک طبقه‌بند دامنه بر پایه ساخته شده است که به شیوه آموزش تخاصمی آموزش داده می‌شود. یک تنظیم‌کننده تناسب همراه با این دو طبقه‌بند در داخل این دو طبقه‌بند به منظور یادگیری یک RPN ثابت دامنه برای مدل Faster R-CNN وارد شده است[45].

Lukas Hoyer و همکارانش[46] رویکرد نوآورانه ای به نام MIC را ارائه دادند. در این روش ابتدا شبکه f_θ روی دامنه منبع آموزش می‌بیند و تابع ضرر بسته به اینکه کدام کار بینایی کامپیوتر انجام می‌شود انتخاب می‌گردد که در اینجا $\mathcal{L}^s$ مربوط به قطعه بندی معنایی است و نماد تابع ضرر مربوط به دامنه منبع است. سپس متناسب با روش UDA انتخابی و الگوریتم آن $\mathcal{L}^T$ که نماد تابع ضرر مربوط به دامنه هدف است انتخاب می‌شود. شبکه f_θ یکبار روی دامنه منبع آموزش می‌بیند و وزن های آپدیت می‌شوند. بعد از ان روی دامنه هدف اجرا می‌شود و یک وزنی می‌گیرد. سپس ماژول MIC شروع به کار می‌کند. تصاویر دامنه هدف به صورت تصادفی پوشیده می‌شوند و به شبکه f_θ وارد می‌شوند. شبکه برای هر کدام از تصاویر پوشیده شده باید پیش بینی انجام دهد و حتی برای قسمت های پوشیده شده نیز باید پیش بینی کند که متعلق به چه کلاسی هستند و باید این کار را با استقاده از اطلاعات باقی مانده از تصویر انجام دهد و از اطلاعات زمینه ای استفاده کند.

از آن جایی که در دامنه هدف برچسبی وجود ندارد برای ارزیابی پیش بینی های شبکه از یک شبکه EMA Teacher استفاده می‌شود. این شبکه با گرفتن تصاویر پوشیده نشده دامنه هدف برچسب های ساختگی تولید می‌کند. شبکه f_θ به عنوان شبکه Student فقط آموزش می‌بیند و سپس در هر مرحله از آموزش با ضریبی وزن های خود را به مدل EMA Teacher تخصیص می‌دهد. شبکه EMA Teacher با یک ضریبی وزن های قبلی خود را نگه می‌دارد. پس در واقع مدلEMA Teacher به یک بازه زمانی از وزن های Student دسترسی دارد. اولین وزنی که از student به EMA Teacher تخصیص می‌یابد وزنی است که حاصل از آموزش f_θ روی دامنه هدف است.

مدل EMA Teacher علاوه بر اینکه به وزن های f_θ دسترسی دارد به تصاویر دامنه هدف که پوشیده شده نیستند نیز دسترسی دارد. پس در واقع این مدل هم به اطلاعات محلی و هم به اطلاعات زمینه ای دسترسی دارد. اطلاعات محلی از تصاویر پوشیده نشده و اطلاعات زمینه ای از تخصیص وزن

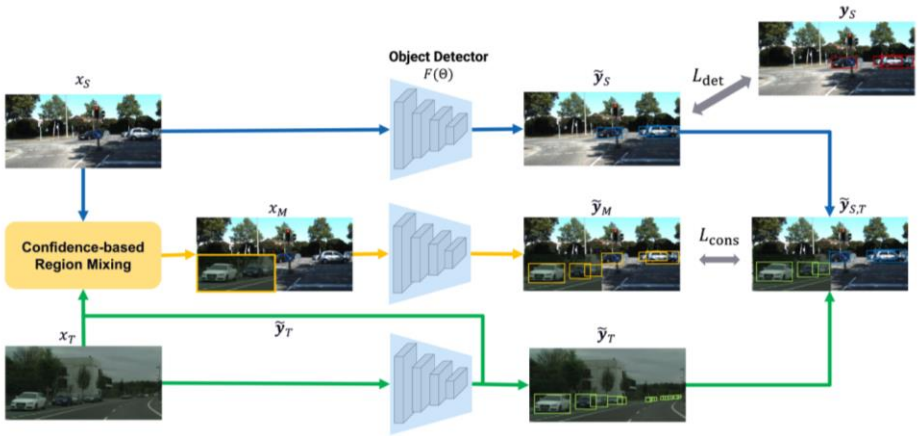
های f_θ به مدل EMA Teacher حاصل می‌شود. این مدل با داشتن این اطلاعات می‌تواند در نهایت برچسب های ساختگی قوی تولید بکند.



شکل ۱۹ UDA با (MIC). در UDA، یک شبکه معمولاً با یک Loss نظارت شده در دامنه منبع (آبی) و یک تابع ضرر تطبیقی بدون نظارت در دامنه هدف (سبز) آموزش داده می شود. MIC سازگاری بین پیش‌بینی‌های تصاویر هدف پوشانده شده (بنفش) و برچسب‌های ساختگی که بر اساس تصویر کامل توسط (EMA Teacher) تولید می‌شوند، اعمال می‌کند. برای به حداقل رساندن Loss مربوط به MIC، شبکه باید بیاموزد که پیش بینی های مناطق پوشانده شده را از بافت آن ها استخراج کند[46].

Giulio Mattolin و همکارانش[47] یک روش نوآورانه برای تطبیق دامنه بدون نظارت در تشخیص اشیاء ارائه دادند که یک استراتژی مخلوط‌کردن نمونه‌ها بر اساس اعتماد تشخیص در سطح ناحیه را معرفی می‌کند. رویکرد پیشنهادی هدف، آموزش یک تشخیص‌دهنده اشیاء پیش‌آموزش دیده به یک دامنه هدف جدید است که برای آن توصیف‌ها در دسترس نیست. ConfMix از دو مرحله اصلی تشکیل شده است: (۱) مرحله آموزش فقط در دامنه منبع، در این مرحله تشخیص‌دهنده بر روی دامنه منبع آموزش داده می‌شود؛ و (۲) مرحله تطبیق دامنه، در این مرحله تشخیص‌دهنده بر روی مخلوطی از نمونه‌های منبع و هدف تنظیم می‌شود. استراتژی مخلوط‌کردن بر اساس امتیازهای اعتماد تشخیص دهنده در دامنه هدف مبتنی است و به تدریج نمونه‌های با اعتماد کم را فیلتر می‌کند تا از انتقال منفی جلوگیری شود. ConfMix در چندین مجموعه داده، از جمله Sim10k، KITTI و Cityscapes، کارآیی بهتر از روش‌های تطبیق دامنه بدون نظارت موجود را داشته و عملکرد بهتری را از خود نشان می‌دهد.

همچنین نشان داده شده است که رویکرد پیشنهادی در حالات با تغییرات آب و هوا و دوربین نیز موثر است و پتانسیل کاربردهای واقعی را نشان می‌دهد.



شکل ۲۰ نمای کلی از روش ConfMix پیشنهادی. ما نمونه‌های منبع x_s و نمونه‌های هدف x_t را به مدل تشخیص دهنده $F(\theta)$ منتقل می‌کنیم و پیش‌بینی‌های $\tilde{y}_s$ و $\tilde{y}_t$ را به ترتیب دریافت می‌کنیم. ما ناحیه هدف با بالاترین اعتماد منطقی را برای تشکیل نمونه مخلوط x_M با تصویر منبع انتخاب می‌کنیم، سپس این نمونه به مدل تشخیص دهنده $F(\theta)$ منتقل می‌شود و پیش‌بینی‌های $\tilde{y}_M$ را تولید می‌کند. ما مدل را با استفاده از از دست دادن تشخیص نظارت شده با استفاده از اشارات منبع $\tilde{y}_s$ و از دست دادن تناسب خود-نظارتی با مقایسه $\tilde{y}_M$ با پیش‌بینی‌های ترکیبی منبع و هدف $\tilde{y}_{s,t}$ آموزش می‌دهیم[47].

Cycle-Consistent Adversarial Domain Adaptation (CyCADA)[48] الهام گرفته از روش CycleGAN است و از توابع cycle-consistent loss به عنوان یکی از مؤلفه‌های آن استفاده می‌کند. در واقع، CyCADA از یک ترکیب از تفاوت‌های تصویری، اطلاعات ویژگی و هدف برای انجام یک وظیفه خاص به عنوان یک مجموعه از توابع ضرر برای آموزش مدل استفاده می‌کند.

اولین تابع ضرری که در CyCADA استفاده می‌شود، تابع ضرر cycle-consistent است. این تابع مطمئن می‌شود که تصاویر بین دو دامنه (دامنه مبدا و دامنه مقصد) به درستی ترجمه شده و بازگردانده شده و تغییرات ناشی از ترجمه در حداقل کمیت‌هایی از این تابع هدف نمایان می‌شوند.

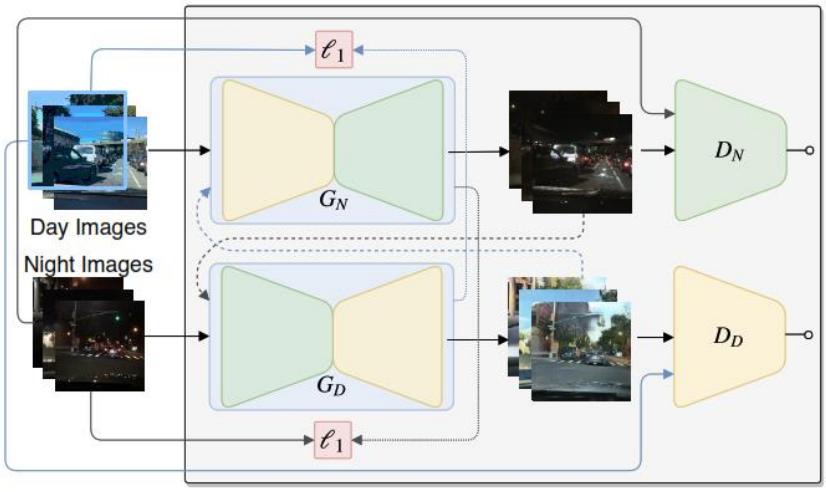
علاوه بر آن سایر توابع ضرر در سه سطح ارائه می‌شوند: image-level، feature-level و task. در سطح تصویری معمولاً به مقایسه تصاویر میان دو دامنه و محاسبه تفاوت‌های تصویری بین آن‌ها می‌پردازد. سطح ویژگی به معنای مقایسه ویژگی‌های تصاویر و محاسبه تفاوت‌های ویژگی بین دو دامنه است. این تابع‌های هدف به کمک آموزش مدل به تفاوت‌ها و تطابق‌های بین دو دامنه کمک می‌کنند. تابع task نیز برای حل یک وظیفه خاص (مثلاً تشخیص شیء در یک تصویر) در دامنه مقصد مورد استفاده قرار می‌گیرد. در نهایت، تمام این تابع‌های هدف با هم ترکیب می‌شوند و سعی در به حداقل رساندن مقدار آن‌هاست.

شبکه‌های بازگردنده در تصویر[1] (DRCN)[49] یک روش در حوزه بینایی ماشین و پردازش تصویر hsj که برای مسائل بازگردنده و پیچیده در تصویربرداری و پردازش تصاویر مورد استفاده قرار می‌گیرد. این روش تلاش می‌کند تا با استفاده از اتصالات بازگردنده میان اطلاعات مختلف در تصاویر، اطلاعات ساختاری و معنایی عمیق‌تری از تصاویر استخراج کند. هدف اصلی DRCN بهبود کیفیت تصاویر است. این بهبود کیفیت می‌تواند شامل افزایش وضوح، کاهش نویز، افزایش تبیین و ویژگی‌های دیگری باشد که به تصاویر اضافه می‌شود. ایده اصلی در DRCN این است که از شبکه‌های بازگردنده برای استخراج ویژگی‌های تصویر استفاده شود و سپس از این ویژگی‌ها برای بهبود تصویر اصلی استفاده کنیم.

در ادبیات، روش‌های دیگری وجود دارند که به طور کامل با تعریف بازسازی تخاصمی همخوانی ندارند، اما به شدت به روش‌کار آن نزدیک هستند و یک مثال آن روش DiscoGAN است. DiscoGAN[50] همانند روش CycleGAN عمل می‌کند، با این تفاوت که از cycle-consistent loss استفاده نمی‌کند.

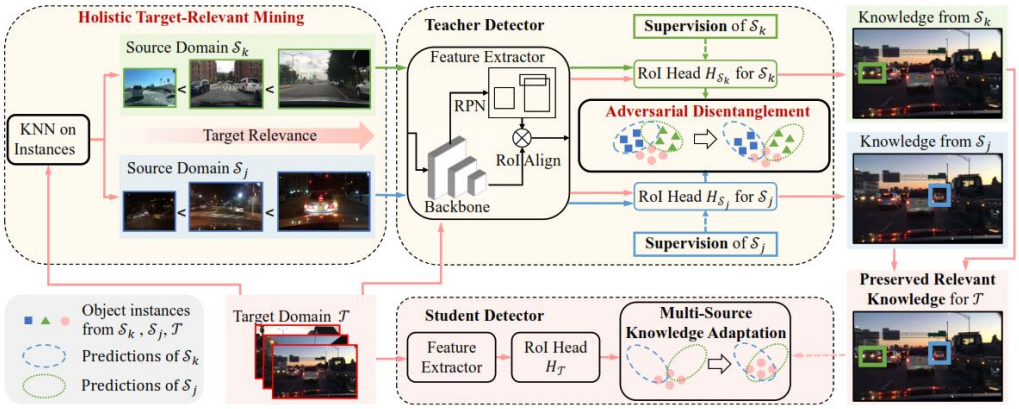
Arruda و همکارانش[51] یک روش تشخیص اتومبیل در دامنه‌های مختلف با استفاده از ترجمه تصویر به تصویر بدون ناظر ارائه داده‌اند. آن‌ها از CycleGAN به منظور ترجمه تصاویر از دامنه روز به دامنه شب به منظور ایجاد مجموعه داده مصنوعی استفاده کردند. در این مرحله، تصاویر از دامنه روز به دامنه شب ترجمه شده و مجموعه داده مصنوعی ایجاد می‌شود. در نهایت، مدل تشخیص (مدلی که اتومبیل‌ها را تشخیص می‌دهد) بر روی مجموعه داده مصنوعی آموزش داده می‌شود و اطلاعات برچسب‌گذاری شده از دامنه منبع به مجموعه داده مصنوعی منتقل می‌شود و مدل تشخیص از تصاویر ترجمه شده از دامنه روز به دامنه شب به طور خودکار یاد می‌گیرد. این امر به ایجاد یک مدل تشخیص قادر به تشخیص اتومبیل‌ها در دامنه‌های مختلف در شرایط مختلف نورپردازی کمک می‌کند.

[1] Deep Reconstruction Classification Networks



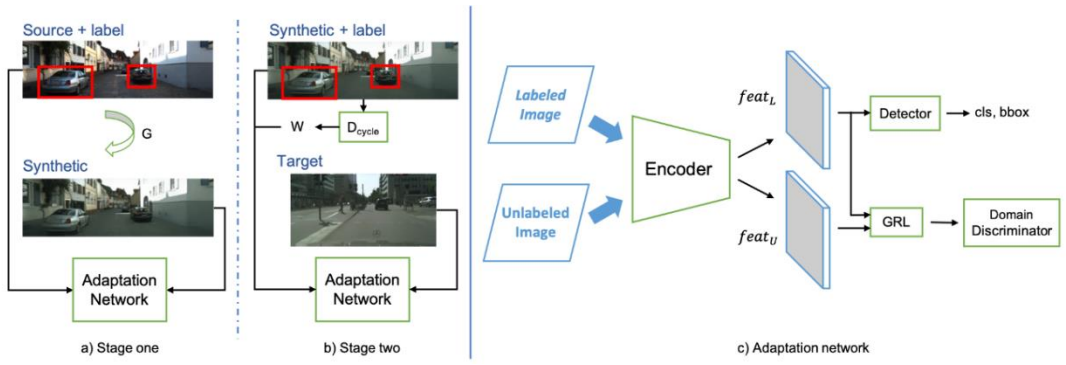
شکل ۲۱ مروری بر چارچوب CycleGAN در این برنامه. G_Nتصاویر را از دامنه روز به دامنه شب نگاشت می‌کند، در حالی که G_Dنگاشت را به صورت معکوس انجام می‌دهد. تشخیص‌دهندگان D_N و D_D ارزیابی می‌کنند که آیا این تصویر واقعی است یا به ترتیب تصویر تقلبی در دامنه شب و روز است. محدودیت تطابق چرخه با استفاده از تابع ضرر L1 برای اطمینان از توانایی بازسازی مورد استفاده قرار می‌گیرد[51].

Jiaxi Wu و همکارانش[52]، یک رویکرد جدید برای تشخیص چند منبعی بدون نظارت و تطبیق دامنه (DAOD) ارائه شده است که به مسئله کاهش دانش در طی ترکیب آن‌ها پرداخته است. روش پیشنهادی، به نام حفظ دانش مرتبط با هدف (TRKP)، از دو مؤلفه اصلی تشکیل شده است: ماژول تبادل چند منبعی تخاصمی (AMSD) و یک طرح معدن دانش کلی مرتبط با هدف (HTRM). ماژول AMSD ویژگی‌های مختص منبع و ویژگی‌های مرتبط با هدف را از یکدیگر جدا می‌کند، در حالی که طرح HTRM دانش مرتبط با هدف را از ویژگی‌های چند منبع استخراج می‌کند. رویکرد پیشنهادی بر روی چندین مجموعه داده مرجع ارزیابی شده و با روش‌های برتر مقایسه شده است. نتایج نشان می‌دهند که TRKP در اصطلاح عملکرد و کارآیی نسبت به روش‌های دیگر از جمله عملکرد برتری دارد.



شکل ۲۲ بررسی کلی چارچوب رویکرد پیشنهادی TRKP. پیکان‌های سفید به انتقال به جلو اشاره دارند و پیکان‌های خط‌خورده نظارت را نمایان می‌کنند. تشخیص دهنده معلم بر روی تصاویر منبع دارای برچسب آموزش داده می‌شود و برچسب‌های سیاه برای تصاویر بدون برچسب در دامنه هدف تولید می‌کند که تشخیص کننده دانش آموز را راهنمایی می‌کند. TRKP از ماژول تبادل چند-منبعی تخاصمی (AMSD) برای حفظ دانش ویژگی‌های مخصوص دامنه منبع و از طرح معدن دانش کلی مرتبط با هدف (HTRM) برای تقویت تمرکز دانش مرتبط با هدف استفاده می‌کند، که به طور قابل توجهی به تطبیق دانش چند-منبع با دامنه هدف کمک می‌کند[52].

Han-Kai Hsu و همکارانش[53] یک رویکرد نوآورانه به نام تطبیق دامنه پیشرو[1] برای شناسایی اشیا ارائه می‌دهد تا شکاف دامنه را پوشش دهد و عملکرد مدل‌های شناسایی اشیا را بهبود ببخشد. این روش از یک دامنه میانی و یادگیری متخاصم برای تطابق توزیع‌ها در سطح ویژگی استفاده می‌کند. همچنین از یک از تابع ضرر وزن‌دار ویژگی‌های تصویر نامتعادل استفاده می‌کند. این روش به تدریج به دامنه مقصد با استفاده از فضای ویژگی میانی سازگار می‌شود و نسبت به روش پایه در معیار mAP بهبود می‌یابد. روش پیشنهادی در موارد مختلف مانند موارد دوربین متقاطع، شرایط آب و هوا و تطبیق با مجموعه داده‌های بزرگ عملکرد بهتری داشته است.



شکل ۲۳ شکل ۷ چارچوب پیشنهادی روش تطبیق دامنه پیشرو را نشان می‌دهد. الگوریتم شامل دو مرحله تطابق است، که در a) و b) نشان داده شده است. در a)، ابتدا تصاویر منبع را تبدیل کرده و تصاویر مصنوعی را با استفاده

[1] progressive domain adaptation

از مولد G که از طریق CycleGAN یادگرفته شده است، تولید می‌شود. سپس از دامنه منبع با برچسب استفاده کرده و تطابق مرحله اول را با دامنه مصنوعی انجام می‌دهیم. در b)، مدل یک تطابق مرحله دوم انجام می‌دهد که دامنه مصنوعی با برچسب‌های از منبع به ارث برده شده را می‌گیرد و ویژگی‌های دامنه مصنوعی را با توزیع هدف هماهنگ می‌کند. علاوه بر این، وزن w از تشخیص‌دهنده Dcycle در CycleGAN به منظور تعادل کیفیت تصاویر مصنوعی در از دست دادن تشخیص به دست می‌آید. ساختار کلی شبکه تطابق ما در c) نشان داده شده است. تصاویر با برچسب و بدون برچسب هر دو از طریق شبکه رمزگذار E منتقل می‌شوند تا ویژگی‌های CNN featL و featU استخراج شود. سپس از آنها برای: ۱) یادگیری تشخیص اشیاء تحت نظر با شبکه تشخیص از featL استفاده می‌شود، و ۲) هر دو ویژگی را به GRL و یک تشخیص‌دهنده دامنه ارسال می‌کنیم تا ویژگی‌های بی‌دامنه را به صورت مقابله‌ای و با تحدید نوعی دامنه یادگرفته شود[53].

شکل ۲۴ نمونه‌هایی از نتایج تشخیص از سه وظیفه تطبیق روش تطبیق دامنه پیشرو. دو ردیف اول به ترتیب نتایج آزمایش KITTI به Cityscapes و Cityscapes به Foggy Cityscapes هستند، در حالی که دو ردیف آخر نتایج Cityscapes به BDD100k هستند. نتایج تشخیص روی دامنه مقصد قبل و بعد از اعمال روش تطبیق دامنه به همراه برچسب‌های واقعی نمایش داده شده است[53].

T. Kim و همکارانش[54] یک چارچوب تطبیق دامنه جهانی برای تشخیص اشیاء پیشنهاد می‌شود که لایه‌های تشخیص اشیاء بدون تغییر دامنه را با داده‌های برچسب‌گذاری شده و تنوع دار آموزش می‌دهد، در عین حال دامنه‌های منتشر را به سوی یک فضای ویژگی مشترک تشویق می‌کند. این چارچوب به مسئله تمایل به تمایز بر مبنای منبع و مسئله ترجمه ناقص پرداخته است. رویکرد شامل دو مرحله به نام تنوع دامنه (DD) و یادگیری نمایندگی چند دامنه‌ای (MRL) است و به نسبت زیادی روش‌های جاری را در انواع مختلف مجموعه داده‌ها پیش‌می‌گیرد.

مرحله اول، تنوع دامنه، روشی است که با تولید عمدی تغییرات متفاوت در دامنه از طریق متغیرهای دامنه، دامنه منبع را متنوع‌تر می‌کند. توزیع متنوع داده‌های برچسب‌گذاری شده، مدل را تشویق به استنباط میان داده‌های با تفاوت داخلی کلاس بزرگ و تفاوت بیرون کلاس می‌کند. این امکان را به مدل می‌دهد تا ویژگی‌های معنایی بدون تعصب از دامنه هدف استخراج کند که متمایزتر از ویژگی‌های تعصبی به منبع هستند.

مرحله دوم، یادگیری نمایندگی چند دامنه‌ای، یک روش است که لایه‌های تشخیص اشیاء بدون تغییر دامنه را با داده‌های برچسب‌گذاری شده و تنوع دار آموزش می‌دهد، در عین حال دامنه‌های منتشر را به سوی یک فضای ویژگی مشترک تشویق می‌کند. این مرحله به مسئله تمایل به تمایز بر مبنای منبع و مسئله ترجمه ناقص پرداخته است. رویکرد نشان داده است که عملکرد برتری را در مجموعه‌های داده مختلف از جمله مجموعه‌های داده جهان واقعی، مجموعه‌های داده رسانه‌هنری و مجموعه‌های داده صحنه‌های شهری به دست آورده است.

شکل ۲۵ معماری چارچوب تطبیق دامنه روش فوق برای تشخیص اشیاء. این چارچوب بر پایه شبکه تشخیص اشیاء ساخته شده است[54].

یک مثال از تجمیع خودتشکیلی در تطبیق دامنه، استفاده از معماری "Teacher-Student" است که Valpola و همکارانش[55] ارائه داده‌اند و Fisher و همکارانش[56] از آن در مقاله خود استفاده کرده اند. در این روش، یک شبکه معلم به عنوان مدل منبع روی وظیفه آموزش می‌شود و خروجی‌هایی که احتمالاتی هستند (برخلاف برچسب‌های صحیح) را تولید می‌کند. سپس شبکه دانش‌آموز از معلم یاد می‌گیرد و می‌تواند اطلاعات بهتری را فهمیده و به عنوان نتیجه بهتری داشته باشد. ترکیب خروجی‌ها برای آموزش شبکه دانش‌آموز با استفاده از کاهش گرادیان[1] انجام می‌شود، و وزن معلم به عنوان میانگین متحرک از شبکه دانش‌آموز استفاده می‌شود.

(M3DA[2])[57] مجموعه داده چند دامنه‌ای به نام DomainNet با ۶ دامنه ایجاد کرد. علاوه بر این، آن‌ها لحظات توزیع ویژگی‌ها را بین دامنه‌های منبع دارای برچسب چندگانه و دامنه هدف به صورت پویا تطابق دادند.

[1] gradient descent
[2] Moment Matching for Multi-Source Domain Adaptation

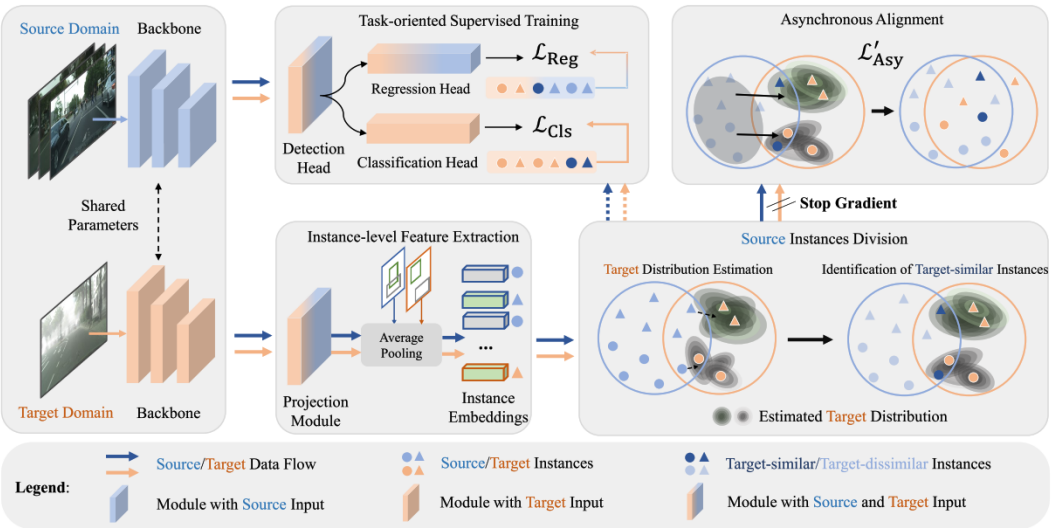
Zhao و همکارانش[34] شبکه تجمع دامنه تخاصمی چند منبعی (MADAN) را معرفی کردند که اساساً از CycleGAN (تفکیک کننده زیر دامنه برای دامنه های منبع و متمایز کننده چرخه متقابل دامنه برای دامنه های منبع- هدف) استفاده می‌کند و یک دامنه تطبیقی پنهان برای تمام داده‌های منبع و داده‌های هدف ایجاد می‌کند.

روش Few-Shot Adversarial Domain Adaptation (FADA) توسط Motiian و همکارانش[58] معرفی شده است و از یادگیری تضادی با تمرکز بر سرعت تطبیق استفاده می‌کند. آنها با ترکیب نمونه‌های منبع و هدف به چهار دسته بر اساس دامنه و برچسب‌های کلاسی تقسیم کردند و سپس طبقه‌بند بر روی این چهار دسته به جای دو دسته معمولی کار می‌کند. آنها شبکه را (استخراج کننده و طبقه‌بند برچسب) فقط با داده‌های منبع مقداردهی اولیه کردند، سپس متمایز کننده دسته دامنه (استخراج کننده ویژگی را فریز کردند) را به‌روز کردند. در نهایت، متمایزکننده دسته دامنه را فریز کردند و استخراج کننده و طبقه‌بند برچسب را به‌روز کردند.

روش Domain-Adaptive Few-Shot Learning (DA-FSL)[59] به حل یک مسئله پیچیده‌تر مرتبط با یادگیری Few-Shot پرداخته می‌شود، یعنی داده‌های هدف ممکن است کلاس‌هایی داشته باشند که متعلق به دامنه‌های مختلفی باشند. تمرکز شبکه تضادی دامنه‌ای (DAPN) در DA-FSL بر دستیابی به تطابق در توزیع دامنه جهانی است، و این کار با معرفی توابع ضرر جدید صورت می‌گیرد در حالی که تمایز کلاسی را حفظ می‌کند. این توابع ضرر از طریق یک مکانیزم تغییر وزن تطبیقی وزن‌دهی می‌شوند. جنبه نوآورانه دیگر در استفاده از مکانیزم توجه قبل از تعبیه منبع است.

روش Prototypical Cross-domain Self-Supervised Learning (PCS) توسط Yue و همکارانش[60] معرفی شده و شامل یادگیری خودکار (که به عنوان چارچوب یادگیری خودی متقابل میان دامنه‌ای تقلبی (PCS) نامیده می‌شود) و بدون نظارت است. ایده اصلی انتقال دانش از منبع به هدف این است که شباهت‌هایی بین نمونه و نماینده (نمونه نماینده) پیدا کند تا انتقال به شکل مستحکم‌تری انجام شود.

تشخیص اشیاء در دامنه‌های تطبیقی کم داده (FSDAOD) یک وظیفه چالش‌برانگیز است که تنها تعداد محدودی تصاویر هدف برای آموزش در کنار تعداد کافی تصاویر منبع برچسب‌گذاری شده در دسترس است. کمبود داده در دامنه هدف منجر به عدم تعادل شدید داده بین دامنه‌های منبع و هدف می‌شود، که احتمالاً در تطبیق سنتی ویژگی باعث انجام یادگیری بیش از حد می‌شود. برای حل این مشکل، نویسندگان مقاله[61] یک پارادایم تطبیق نامتقارن به نام AsyFOD ارائه می‌دهند که از نمونه‌های منبع و هدف از دیدگاه‌های مختلف بهره می‌برد.



شکل ۲۶ توضیح مختصری از AsyFOD. پس از استخراج نقشه ویژگی تصاویر ورودی از طریق بخش اصلی، نمونه‌های ویژگی مرتبط با نمونه به وسیله ماژول پروژه‌ای و حقایق ارائه شده استخراج می‌شوند. سپس، AsyFOD نمونه‌های منبع را به دو مجموعه تشخیص هدف مشابه (آبی تیره) و مجموعه تشخیص هدف نامشابه (آبی روشن) تقسیم می‌کند. نمونه‌های منبع مشابه هدف برای افزایش نمونه‌های هدف استفاده می‌شوند. با این تقسیم‌بندی، تطبیق ناهمگام بین نمونه‌های منبع نامشابه هدف و نمونه‌های هدف افزایش یافته انجام می‌شود که باعث کاهش یادگیری زودهنگام ناشی از توزیع داده‌های بسیار نامتعادل می‌شود. همچنین، AsyFOD آموزش نظارت شده متمرکز بر وظیفه را برای وظایف طبقه‌بندی و مکان‌یابی به طور جداگانه انجام می‌دهد. بهترین نمایش در حالت رنگی مشاهده می‌شود[61].

AsyFOD ابتدا نمونه‌های مشابه به هدف را با استفاده از تخمین توزیع هدف تشخیص می‌دهد، که به افزایش تعداد محدودی نمونه هدف کمک می‌کند. سپس، AsyFOD تطبیق ناهمگام بین مجموعه نمونه‌های منبع نامشابه هدف و مجموعه محدود نمونه هدف انجام می‌دهد تا از یادگیری بیش از حد جلوگیری کند. نویسندگان همچنین یک عملیات توقف-گرادیان را ارائه داده‌اند تا از جریان گرادیان به دامنه منبع در طول فرآیند تطبیق جلوگیری کنند.

Jaemin Na و همکارانش[62] رویکرد جدیدی برای تطبیق دامنه بدون نظارت[1] (UDA) برای یادگیری نمایش‌های مقاوم در برابر تغییرات دامنه به نام FixBi ارائه دادند. بیشتر مطالعات بر اساس تطبیق مستقیم از دامنه منبع به دامنه مقصد انجام شده‌اند و از تفاوت‌های بزرگ دامنه رنج می‌برند. در این مقاله، یک روش UDA ارائه شده است که به طور موثر با تفاوت‌های بزرگ دامنه‌ای مقابله می‌کند.

در این روش، از مخلوط مبتنی بر نسبت ثابت برای افزایش دامنه‌های میانی چندگانه بین دامنه منبع و دامنه مقصد استفاده می‌شود. از این دامنه‌های افزوده، مدل مغرور از دامنه منبع و مدل مغرور از دامنه مقصدی را آموزش می‌دهیم که ویژگی‌های تکمیلی دارند. با استفاده از روش‌های یادگیری مبتنی بر اعتماد، مانند مطابقت دوطرفه با پیش‌بینی‌های با اعتماد بالا و خود تنبیه با استفاده از

[1] Unsupervised Domain Adaptation

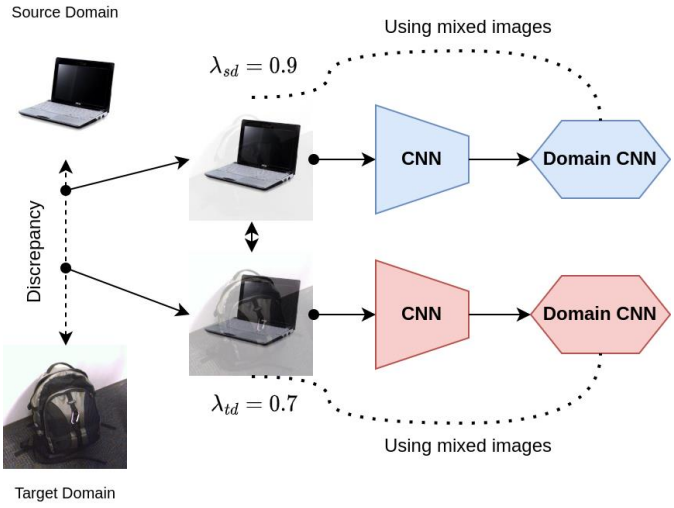
پیش‌بینی‌های با اعتماد پایین، مدل‌ها می‌توانند از یادگیری از یکدیگر یا از نتایج خود یاد بگیرند. از طریق روش‌های پیشنهادی نویسندگان، مدل‌ها به تدریج دانش دامنه را از دامنه منبع به دامنه مقصد منتقل می‌کنند. آزمایش‌های گسترده نشان می‌دهند که روش پیشنهادی فوق برتری خود را در سه بنچمارک عمومی، Office-31، Office-Home و VisDA-2017، اثبات کرده است.

۳-۲-۲. پایان‌نامه‌ها

Artem Bitiutskii در پایان نامه خود[63]، در مورد روش جدید توسعه‌یافته ای به نام "DannFixbi" بحث می‌کند که رویکرد Fixbi[62] و رویکرد بازگشت عصبی را ترکیب می‌کند. تصمیم برای ترکیب این دو رویکرد بر اساس نقاط قوت متقابل و قابلیت تعامل متقابل میان آن‌ها اتخاذ شده است. با استفاده از تکنیک Fixbi که اختلاف حوزه را مدیریت می‌کند و بهره‌برداری از مزایای بازگشت عصبی، DannFixbi به بهبود عملکرد و استحکام تطبیق حوزه در زمینه تصاویر هدف می‌گیرد. توسعه DannFixbi نمایانگر یک مشارکت اصیل در این زمینه است.

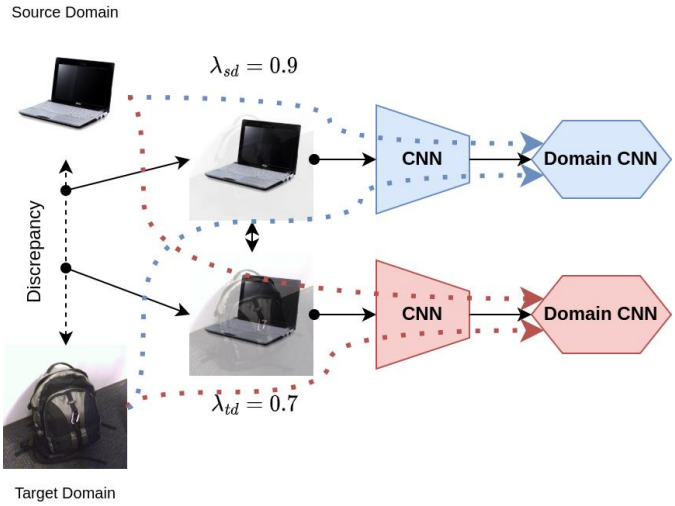
این روش جدید از دو شبکه عصبی تشکیل شده است که با استفاده از یک نسخه اصلاح شده از رویکرد Fixbi آموزش داده می‌شوند. برای بهبود عملکرد این روش، دو طبقه‌بند حوزه به هر دو این دو شبکه اضافه شده است. دو رویکرد مختلف برای یادگیری این طبقه‌بندهای حوزه مورد بررسی قرار گرفته است.

رویکرد اول شامل افزودن یک طبقه‌بند حوزه به هر شبکه عصبی است(اشاره به شکل۲۷). در طول مرحله آموزش، تصاویر حاصل از میکس کردن داده‌های حوزه منبع و حوزه مقصد با نسبت‌های mixup از پیش تعیین شده به این طبقه‌بندها وارد می‌شوند.



شکل ۲۷ ساختار DannFixbi برای رویکرد اول[63]

در رویکرد دوم، از یک طبقه‌بند حوزه برای هر یک از شبکه‌ها با تصاویر از حوزه منبع و حوزه مقصد بدون هیچ مخلوطی استفاده می‌شود، مشابه روش بازگشت عصبی(اشاره به شکل ۲۸).



شکل ۲۸ ساختار DannFixbi برای رویکرد دوم[63]

Gabriele Degola در پایان‌نامه خود[64] در مورد انتقال دانش بین حوزه‌های مختلف در یک سناریوی تشخیص اشیا مطالعه داشته است. این وظیفه با چالش‌های زیادی همراه است و کاربردهای متعددی در حوزه‌های تجاری مختلف دارد.

یکی از چالش‌های اصلی در این پروژه، کمبود داده‌های توصیف شده (annotated data) برای مورد مطالعه مورد نظر است. این مشکل اغلب یکی از مهم‌ترین موانع برای پروژه‌های یادگیری ماشین است. مهندسان باید ساعت‌ها صرف جمع‌آوری و امکاناً توصیف داده‌های مرتبط کنند، به جای اینکه بر روی الگوریتم‌ها و بهینه‌سازی تمرکز کنند. از سوی دیگر، داده‌های مصنوعی به راحتی و به تعداد زیادی قابل

تولید هستند. علاوه بر این، وظایف در دست معمولاً مشابه هستند، اما داده‌ها تحت شرایط مختلفی از نور، طیف نور، زاویه دوربین و موارد مشابه جمع‌آوری شده‌اند.

مدل‌های آموزش داده شده در گذشته قابل استفاده مجدد نیستند، و به همین دلیل فرآیند آموزش طولانی نیاز دارد. به همین منظور، تحقیق انجام شده در این پایان‌نامه، رویکردها ساده‌ای برای تشخیص اشیا با تطبیق دامنه را ارائه و نشان می‌دهد که مستقل از معماری‌های شبکه مورد استفاده و بدون معرفی تغییرات پیچیده در معماری می‌توانند نتایج قابل قبولی را به دست آورند و با روش‌های پیچیده‌تری که پیش‌تر در ادبیات معرفی شده‌اند، رقابت کنند.

در بخش نتایج آزمایشی پایان‌نامه خود، نویسنده عملکرد شبکه‌های RetinaNet و Faster R-CNN را روی مجموعه داده CityScapes ارزیابی کرد. او از رویکردهای یادگیری ویژگی مقاوم در برابر دامنه و تطبیق دامنه برای بهبود عملکرد این شبکه‌ها استفاده کرد. نتایج نشان داد که تغییرات در روشنایی، تضاد، اشباع و رنگ‌های تصاویر باعث افزایش نتایج تشخیص در مجموعه داده CityScapes می‌شود. تغییرات اضافی مانند Solarization (تغییر دیگری در نور و رنگ تصویر)، Equalization (متعادل کردن تصویر) یا تقسیم‌کردن تصویر به اندازه و شکل‌های مختلف تأثیر مثبت زیادی بر عملکرد نهایی ندارند یا تأثیر اندکی دارند.

همچنین نویسنده اعلام کرده که رویکردهای اجرا شده برای تشخیص اشیا با تطبیق دامنه از منبع دامنه تولید ماشین به دامنه هدف واقعی دنیا، موثر هستند. تغییرات بسیار ساده در ظاهر تصاویر آموزشی تأثیر زیادی بر عملکرد تشخیص دارند. بهترین نتایج برای هر دو از شبکه‌های RetinaNet و Faster R-CNN از طریق تطبیق دامنه بر اساس تطبیق آماره‌های تصاویر منبع و هدف به دست می‌آید. این روش بدون هیچ تغییر معماری‌ای انجام می‌شود و نیازی به هیچ توصیف برای دامنه هدف ندارد. قبل از این، آزمایش‌های انجام شده به وضوح نشان می‌دهند که تنها تغییر دامنه، که از طریق عملگرهای اصلاحی ابتدایی داده، می‌تواند نتایج قابل قبولی را به دست آورد و تفاوت بین دامنه منبع و هدف را به طور قابل توجهی کاهش دهد. این روش بسیار ساده برای تنظیم و آزمایش است و اگر با تکنیک‌های دیگر ترکیب شود، می‌تواند به نتایج بهتری منجر شود.

Yuchen Yang در پایان نامه خود[65] در ابتدا چالش تغییر دامنه و به طبع آن کاهش دقت مدل‌ها را بیان می‌کند برای کم کردن این کاهش دقت بدون نیاز به برچسب از دامنه هدف، روش‌های تطبیق دامنه بدون نظارت (UDA) پیشنهاد شده‌اند.

در مقایسه با بیشتر مطالعات UDA که به تصویربرداری از کلاس‌ها و تصویربرداری سطح پیکسل (تقسیم تصویر به قطعات کوچک) می‌پردازند، UDA برای تشخیص اشیا یک حوزه نسبتاً جدید است. یک روند پردازش معروف برای اعمال آموزش مقاوم در برابر دامنه شامل آموزش معکوس با دامنه تمییزدهنده است. دامنه تمییزدهنده توزیع ویژگی‌های دامنه منبع و دامنه هدف را همگام می‌کند.

روش‌های موجود در تشخیص اشیا با تطبیق دامنه ویژگی‌ها را از سطح تصویر استخراج می‌کنند و به طور مستقیم آن‌ها را به عنوان ویژگی‌های کامل در UDA برای وظایف دسته‌بندی تطبیق می‌دهند. با این حال، همگام‌سازی ویژگی‌های کامل تصویر به عنوان یک کلیت برای وظیفه تشخیص اشیا مناسب نیست، زیرا وجود پس‌زمینه‌های متنوع ممکن است با نتایج تطبیق تداخل داشته باشد. برای جلوگیری از همگام‌سازی ویژگی‌های کامل، این پایان‌نامه یک چارچوب تطبیق دامنه متمرکز بر روی جلوه‌های اصلی (FFDA) را پیشنهاد می‌کند. این چارچوب FFDA از توابع ضرر تمیزدهنده‌های دامنه به منظور تمرکز همگام‌سازی روی جلوه‌های اصلی در زمان بازگشت به عقب استفاده می‌کند. FFDA پیش‌بینی‌های هدف و برچسب‌های تصویر منبع را جمع‌آوری می‌کند و از آنها برای تولید ماسک‌های جستجو استفاده می‌کند که مناطق جلوه را مشخص می‌کنند. سپس این ماسک‌ها را به تمیزدهنده‌های دامنه سطح تصویر و نمونه اعمال می‌کند تا فقط در مناطق جستجویی به عقب برگردد. به علاوه، با تقویت این تطبیق متمرکز بر روی جلوه در طول چند لایه مختلف در مدل تشخیص اشیا، FFDA باعث افزایش قابل توجهی در پیش‌بینی دامنه هدف می‌شود.

روش FFDA، نسبت به روش‌های قبلی به بهترین دقت جدید در تطبیق از Cityscape به مجموعه داده Foggy Cityscape می‌رسد. FFDA نیز نتایج رقابتی را در مجموعه‌های داده دیگری که شامل سناریوهای مختلفی برای برنامه‌های رانندگی خودکار هستند، نشان می‌دهد.

علاوه بر مشکل تشخیص اشیا، این پایان‌نامه به بررسی کاربرد UDA برای دسته‌بندی تصویر کلی اسلاید WSI[1] نیز می‌پردازد. دسته‌بندی تصویر برای WSI در مقایسه با دسته‌بندی تصویر عمومی به دلیل دقت بالا و اطلاعات کلیدی پراکنده‌ای که دارد، یک وظیفه چالش‌برانگیز است. کارهای قبلی یک لوله‌کدینگ جدید برای دسته‌بندی WSI ارائه داده‌اند. با این حال، این لوله‌کدینگ همچنان هنگام استفاده از مجموعه داده‌های WSI متفاوت از یک مؤسسه مختلف برای انجام همان وظیفه از همان پدیده کاهش دقت رنج می‌برد. این مسئله محدودیت‌های استفاده عملی از این لوله‌کدینگ را، به خصوص وقتی تشخیص‌های WSI سخت به دست آمده‌اند، ایجاد می‌کند.

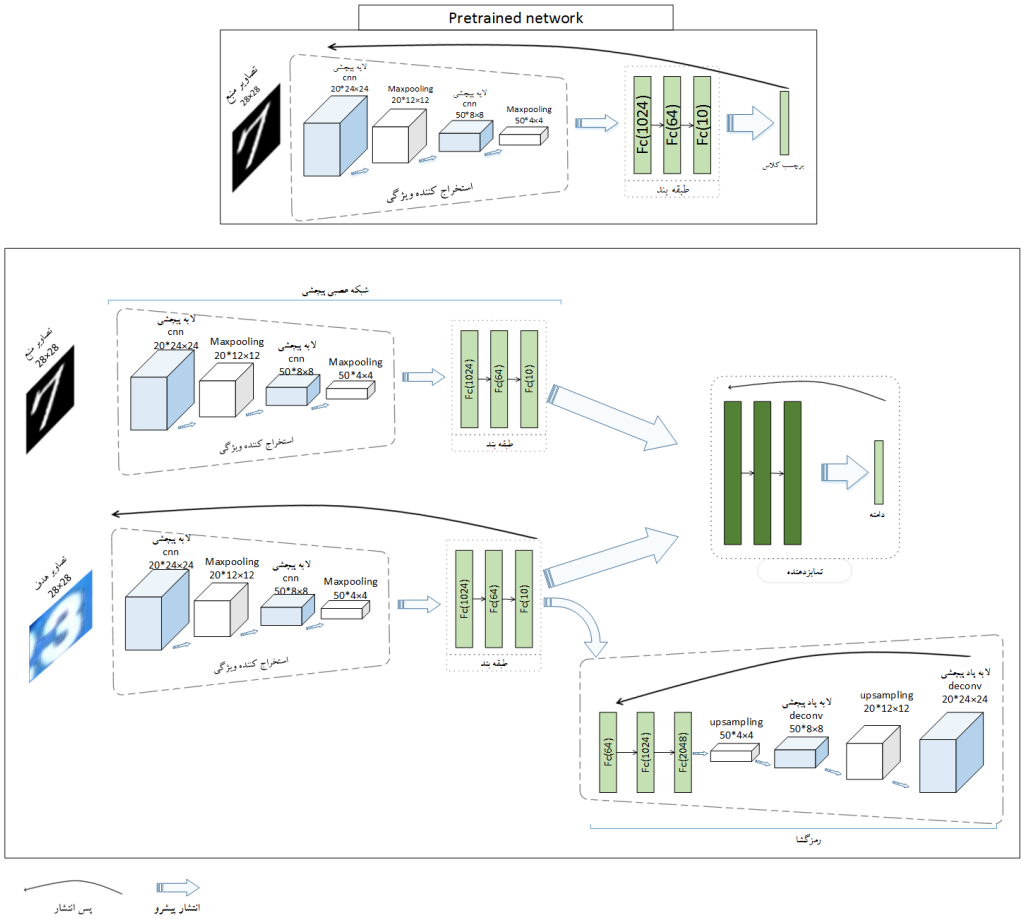
از سوی دیگر، کارهای قبلی که از UDA در تصویربرداری پزشکی استفاده کرده‌اند، معمولاً بر روی نمونه‌های کوچک تصویرهای میکروسکوپی یا قطعات تصویر استخراج شده از WSI تمرکز دارند. UDA برای کاربرد در دسته‌بندی تمام WSI به دلیل محدودیت تعداد لوله‌کدینگ‌ها و مجموعه‌داده‌هایی که پشتیبانی از دسته‌بندی WSI را دارند، هنوز بررسی نشده است.

هدف این پایان‌نامه ارائه یک راه‌حل UDA برای افزایش مقاومت لوله‌کدینگ‌های قبلی با کاهش دقت ناشی از مجموعه داده‌های مختلف WSI است. این راه‌حل دسته‌بند دامنه را در مراحل مختلف میانه‌ای در لوله‌کدینگ قبلی درج می‌کند تا ویژگی‌ها را در طول آموزش همگام سازد. نتایج نشان

[1] Whole Slide Image

می‌دهند که با قرار دادن دسته‌بند دامنه در مراحل مختلف، لوله‌کدینگ نمایش دقت بالاتری را بر روی داده‌های WSI هدف نشان می‌دهد.

روش پیشنهادی در پایان نامه حمیده خادم‌پیر[66]، توسعه ای از روش ADDA[20] می‌باشد. در این پایان نامه برای اینکه دقت مدل افزایش پیدا کند و کاستی های روش ADDA جبران شود از روش تطبیق دامنه افزایشی استفاده می‌شود که در آن به تدریج تعدادی از مطمئن ترین داده های دامنه هدف شناسایی شده و به مجموعه داده های دامنه منبع اضافه میشوند. برای تشخیص مطمئن ترین دادههای دامنه هدف از یک شبکه خودرمزنگار استفاده می‌شود. این شبکه بخش رمزگذار خود را از قسم مولدِ شبکه تمییزدهنده که در مرحله قبل آموزش دیده است می‌گیرد[66]. در شکل دیاگرام چارچوب کلی مسئله رسم شده است[66].



شکل ۲۹ فلوچارت روشهای پیشنهادی تطبیق دامنه‌های افزایشی[66]

همانطور که مشاهده میشود دیاگرام از سه قسم اصلی تشکیل شده است. قسمت اول شبکه عصبی پیچشی اس که خود از دو قسم استخراج کننده ویژگی و طبقه بند تشکیل شده است. استخراج کننده ویژگی شامل لایه‌های پیچشی و max_pooling میباشد که تصاویر را به یک بردار ویژگی نگاشت می‌دهد. طبقه بند متشکل از لایه تماما متصل است که ویژگیهای استخراج شده را به کلاس هر دامنه

نگاشت میدهد. قسمت دوم که رمزگشا نام دارد یک شبکه پادپیچشی است که ترانهاده شبکه پیچشی می‌باشد. قسم سوم تمایزدهنده می‌باشد که به وسیله لایه‌های تماما متصل وظیفه تشخیص دامنه ورودی را دارد[66].

روش پیشنهادی در پایان نامه نیلوفر شعبانی[67]، متشکل از دو بخش می باشد. در بخش اول روشی که به منظور انجام تطبیق وجه برای داده های چند وجهی طراحی شده است، توضیح داده می شود. سپس در بخش دوم، روش پیشنهادی این پایان نامه به منظور انجام تطبیق دامنه برای داده های چند وجهی معرفی خواهد شد[67].

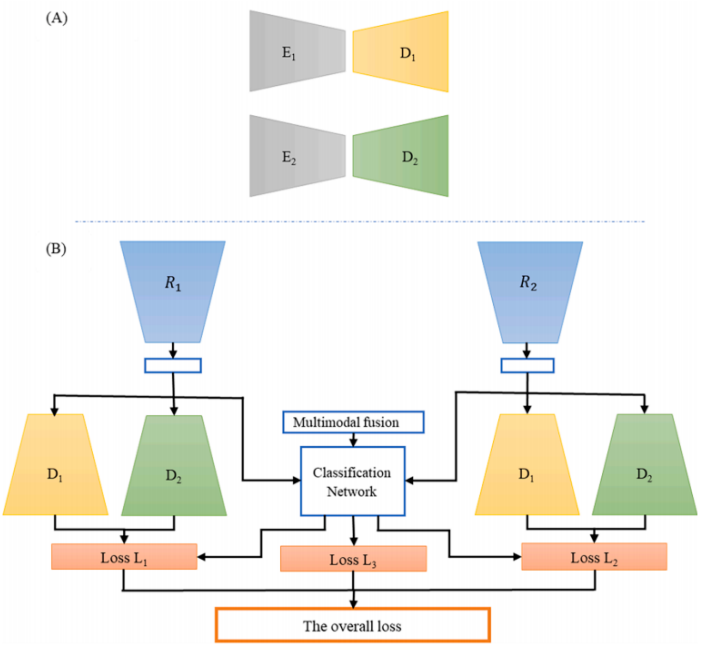
مدل پیشنهادی نویسنده در بخش اول دارای سه ویژگی عمده می باشد که موجب تمایز آن با سایر روش ها می شود، این سه ویژگی عبارتند از:

۱. روش پیشنهاد شده در این پایان نامه برای تطبیق وجه علاوه بر استخراج اطلاعات معمول وجه‌های مختلف داده، اطلاعاتی را که در وجه های مختلف داده مکمل یکدیگر هستند را نیز استخراج می کند[67].

۲. در این روش اطلاعات موجود در فضای برچسب نیز در نظر گرفته می شود[67].

۳. روش پیشنهادی در این بخش یک معماری یکپارچه می باشد بدین معنا که تمامی پارامترهای شبکه در یک گام بروزرسانی می شوند. این در حالی است که در سایر روش های انجام تطبیق وجه که براساس یادگیری رقابتی عمل می کنند، در هر گام بخشی از پارامتر های شبکه ثابت نگه داشته می شوند و بقیه پارامترها بروزرسانی می شوند و این فرآیند تا زمانی که مدل همگرا شود ادامه می یابد[67].

مدل پیشنهادی در این بخش تلاش دارد تا با استفاده از ساختار خودکدگذارها، اطلاعات مشترک بین وجه های مختلف داده را استخراج نماید.



شکل ۳۰ مدل ارائه شده برای انجام تطبیق وجه[67]

روش پیشنهادی در بخش اول دارای دو گام مختلف می باشد. در گام اول برای هر وجه داده یک خودکدگذار آموزش داده می شود. در گام دوم برای هر وجه داده یک کدگذار جدید در نظر گرفته می شود. بدین ترتیب شبکه در این گام به ازای هر وجه داده یک زیرشبکه خواهد داشت[67].

بخش دوم شامل سه مرحله می باشد که عبارتند از: مرحله پیش آموزش که خود شامل دو گام مختلف است، مرحله اصلی و مرحله آزمایش. در روش پیشنهادی این بخش از ایده به دست آوردن اطلاعات مشترک بین دو وجه در حین به دست آوردن اطلاعات خاص هر وجه که در بخش اول توضیح داده شد بهره برده شده است[67].

برای سنجش مدل پیشنهادی بخش اول این پایان نامه از مجموعه داده های چند نمایی COIL100 و ALOI100 به منظور انجام طبقه بندی اشیا استفاده شده است[67].

در بخش دوم این پایان نامه نیز از چهار مجموعه داده مختلف به منظور انجام طبقه بندی استفاده شده است که عبارتند از: MNIST, USPS, NYU Depth Dataset V2 و RGB+D DATABASE[67].

۳-۳. معیارهای ارزیابی

در این بخش به مفاهیم اساسی و معیارهای رایج در اندازه گیری عملکرد الگوریتم های تشخیص اشیاء می‌پردازیم.

ارزیابی الگوریتم یک بخش اساسی از پروژه‌های یادگیری ماشین است و می‌تواند با توجه به وظیفه و هدف، به روش‌های مختلفی انجام شود. تشخیص اشیاء یک وظیفه پیچیده‌تر از دسته‌بندی تصاویر

سنتی است زیرا تعداد غیرثابتی از اشیاء ممکن است در یک تصویر وجود داشته باشد و مدل تشخیص باید به دقت کلاس آن‌ها را همراه با موقعیت آن‌ها پیش‌بینی کند. به همین دلیل، باید معیارهای اختصاصی معرفی و تعریف شوند.

۳-۳-۱. معیارهای Precision و recall

برای توضیح این معیار ابتدا باید مفاهیم زیر در زمینه تشخیص اشیاء تعریف شوند:

• مقدار مثبت واقعی (TP): یک bounding box به درستی به یک شیء اختصاص داده می‌شود که در واقع در تصویر ظاهر شده است.

• مقدار مثبت غلط (FP): تشخیص نادرست، یعنی شیءی تشخیص داده می‌شود که در واقع در bounding box حاضر نیست یا فقط جزئی از آن است.

• مقدار منفی غلط (FN): شیء حاضر در تصویر توسط مدل تشخیص اشیاء تشخیص داده نمی‌شود.

مفهوم منفی واقعی (TN) در زمینه تشخیص اشیاء معنی ندارد، زیرا عدم تشخیص یک شیء غیر حاضر در تصویر مرتبط نیست. در واقع، این به معنای این است که مدل به درستی پس‌زمینه را به عنوان متعلق به هیچ کلاسی تشخیص داده است.

Precision: این معیار دقت، اندازه‌گیری می‌کند که چقدر پیش‌بینی‌های مدل دقیق هستند، بنابراین درصد یا کسری از پیش‌بینی‌های صحیح را بازمی‌گرداند. به عبارت دیگر، این معیار دقت نشان می‌دهد که از تمام پیش‌بینی‌هایی که مدل انجام داده است، چه تعداد به درستی اشتباه نیافته‌اند.

این معیار طبق رابطه‌ی زیر محاسبه می‌شود:

$$Precision = \frac{TP}{TP + FP}$$

معادله ۲

Recall: این معیار توانایی تشخیص مدل را اندازه‌گیری می‌کند، بنابراین درصد یا کسری از اشیاءی که به درستی تشخیص داده شده‌اند را بازمی‌گرداند. به عبارت دیگر، این معیار نشان می‌دهد که از تمام اشیاء موجود در تصویر، چه تعداد به درستی تشخیص داده شده‌اند.

این معیار طبق رابطه‌ی زیر محاسبه می‌شود:

$$Recall = \frac{TP}{TP + FN}$$

معادله ۳

۳-۳-۲. معیار IoU[1]

این معیار برای هر bounding box پیش‌بینی شده، اشتراک آن را با bounding box واقعی[۲] محاسبه می‌کند تا یک پیش‌بینی به عنوان مثبت واقعی یا مثبت غلط ارائه دهد. برای یک جفت bounding box، معیار IoU به صورت زیر محاسبه می‌شود:

$$IoU = \frac{area\ of\ overlap}{area\ of\ union}$$

معادله ۴

در اینجا، area of overlap بخشی از تصویر است که توسط هر دو bounding box پوشش داده می‌شود، و area of union کل مساحت هر دو bounding box را اشغال می‌کند. اگر IoU بالا باشد، به معنای این است که میزان تطابق و دقت بیشتری در تشخیص اشیاء وجود دارد.

۳-۳-۳. معیار mAP

mAP[3] یک معیار ارزیابی است که برای ارزیابی عملکرد سیستم‌های بازیابی اطلاعات[۴] استفاده می‌شود. این معیار به طور خاص در مواردی که در جستجوی اطلاعات بازیابی چندگانه (مثلاً در موتورهای جستجو) مورد استفاده قرار می‌گیرد.

mAP معیاری است که هم دقت و هم بازیابی و امتیازهای اعتماد مرتبط را در نظر می‌گیرد. برای محاسبه MAP، ابتدا برای هر جستجو، باید لیستی از مجموعه‌ای از نتایج بازیابی را داشته باشیم، که هرکدام دارای یک مقدار ارزش (مثلاً ثبت شده به عنوان یک برچسب) هستند. سپس با استفاده از این لیست نتایج، Precision@K (دقت در مقدار K) محاسبه می‌شود، که نسبت تعداد نتایج صحیح به تعداد کل نتایج بازیابی شده در K نتیجه را نشان می‌دهد.

با محاسبه Precision@K برای تمامی جستجوها، می‌توان معیار Average Precision (AP) را محاسبه کرد. سپس با محاسبه میانگین AP برای تمام جستجوها، متوسط عملکرد دقیق (MAP) به دست می‌آید.

[1] Intersection over Union
[2] ground truth
[3] Mean Average Precision
[4] Information Retrieval

MAP معیاری از کلیت و کیفیت عملکرد یک سیستم بازیابی اطلاعات را نشان می‌دهد، زیرا سیستم‌هایی که در جستجوها به خوبی عمل می‌کنند و نتایج با ارزش بالا را بازیابی می‌کنند، در این معیار امتیاز بالاتری خواهند داشت.

۳-۴. مجموعه دادگان

۳-٤-۱. مجموعه داده Cityscapes

مجموعه داده Cityscapes مجموعه ای برای بررسی و پژوهش در زمینه تشخیص و تفسیر تصاویر شهری و خودروها در محیط‌های شهری است. این دیتاست ابتدا توسط دانشگاه Tübingen در آلمان ایجاد شد و با همکاری شرکت‌ها، دانشگاه‌ها و موسسات تحقیقاتی دیگر توسعه یافته است. هدف اصلی از ایجاد این دیتاست، تشویق تحقیقات در زمینه تشخیص اشیاء، ترافیک، خیابان‌ها و معماری شهری با استفاده از تصاویر واقعی از محیط‌های شهری بوده است.

این مجموعه داده دارای تصاویر معنایی و اصلی (۵۰۰۰ برچسب با کیفیت بالا، ۲۰۰۰۰ برچسب بزرگ) از ۵۰ شهر مختلف هستند. هر تصویر در مجموعه داده Cityscapes دارای برچسب‌های دقیق برای اشیاء مختلف موجود در تصویر می‌باشد. این برچسب‌ها از جمله دسته‌بندی اشیاء، تشخیص مرزهای اشیاء (Segmentation)، تشخیص نقاط مختلف روی اشیاء (Keypoints) و سایر اطلاعات مرتبط با تصویر می‌شوند. این به عنوان یکی از مجموعه داده های معیار در تقسیم بندی معنایی جاده/مناظر شهری عمل می کند و شامل تصاویر با وضوح بالا (p۱۰۸۰) از محیط‌های شهری مختلف است که از نقاط مختلف شهرهای مختلف جمع‌آوری شده‌اند. علاوه بر تصاویر محیط شهری، این مجموعه داده شامل تصاویر از خودروها با زوایای مختلف و در شرایط مختلف ترافیکی نیز می‌شود و دارای تنوع محیطی بالایی است.

از دیگر کاربردهای این مجموعه داده می‌توان به توسعه الگوریتم‌ها و مدل‌های هوش مصنوعی برای خودکارسازی خودروها، ارزیابی ایمنی خودروها، پژوهش‌های مرتبط با شهرهوشمند و بهینه‌سازی ترافیک اشاره کرد. به دلیل حجم بالای داده‌ها و پیچیدگی تسک‌های مرتبط، این دیتاست به عنوان یک چالش مهم برای پژوهش‌های مختلف در زمینه بینایی ماشین و یادگیری عمیق مطرح است.

شکل ۳۱ چند نمونه از تصاویر مختلف مجموعه داده Cityscapes

۳-٤-۲. مجموعه داده Foggy Cityscapes

مجموعه داده "Foggy Cityscapes" یک مجموعه داده تصویری است که برای آموزش و ارزیابی مدل‌های یادگیری عمیق مخصوصاً در زمینه بینایی ماشین، تقسیم بندی معنایی و تشخیص اشیاء در شرایط آب و هوای مه آلود و غبارآلود توسعه داده شده است. ویژگی اصلی این مجموعه داده، حضور مه و غبار در تصاویر است که می‌تواند برای مدل‌های یادگیری عمیق به چالشی تبدیل شود. مه و غبار معمولاً باعث کاهش وضوح و کیفیت تصاویر می‌شوند و تشخیص اشیاء و ویژگی‌های تصویری به طور

معمول دشوارتر می‌شود. این مجموعه داده شامل تصاویری از صحنه‌های شهری در شرایط آب و هوای مه و غبارآلود است.

از این مجموعه داده می‌توان برای آموزش مدل‌های یادگیری عمیق مانند شبکه‌های عصبی کانولوشنی برای تقسیم بندی معنایی، تشخیص اشیاء، تشخیص ترافیک، تشخیص افراد و سایر وظایف بینایی ماشین استفاده کرد. این مجموعه داده معمولاً به همراه برچسب‌هایی مشخص برای اشیاء مختلف ارائه می‌شود که به مدل‌ها کمک می‌کند تا اشیاء را در تصاویر شناسایی کنند.

شکل ۳۲ مجموعه داده Cityscapes (سمت چپ) در مقابل Foggy Cityscapes (سمت راست)

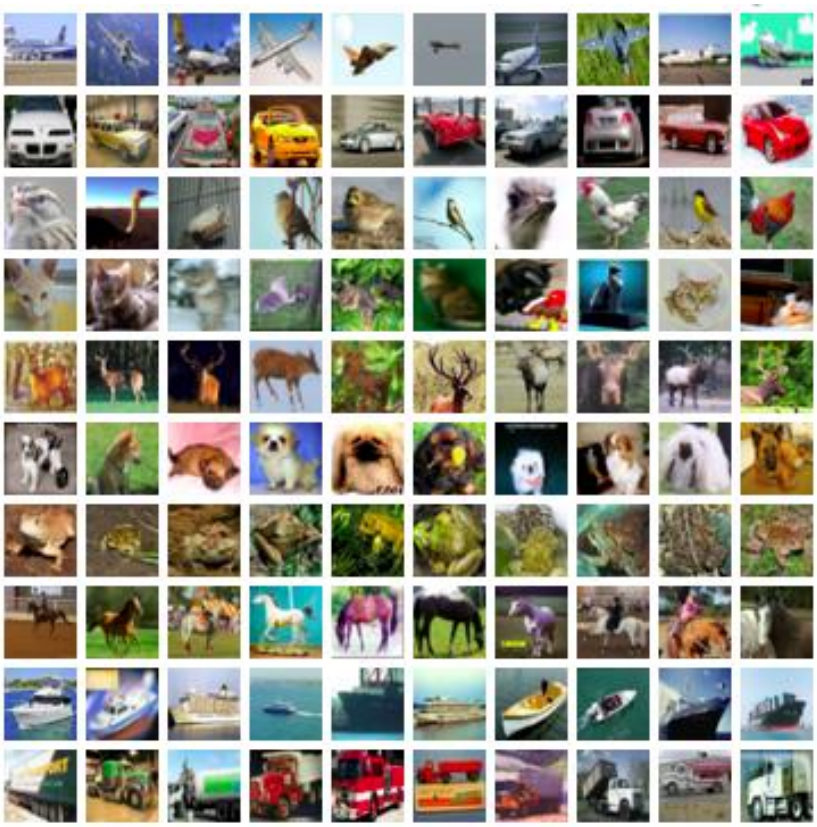
۳-٤-۳. مجموعه داده COCO

مجموعه داده‌ی COCO یکی از مجموعه‌های داده معروف و گسترده در زمینه تشخیص اشیاء و تفسیر صحنه در تصاویر است. نام "COCO" مخفف "Common Objects in Context" می‌باشد، به این معنا که در این مجموعه داده، اشیاء مختلف در متناوب و با توجه به مفهوم کلی تصویر قرار دارند. این مجموعه داده در سال ۲۰۱۴ توسط محققانی از دانشگاه استنفورد ایجاد شد و به منظور ترویج تحقیقات در زمینه تشخیص اشیاء، دسته‌بندی، تفسیر صحنه و وظایف مشابه دیگر در حوزه بینایی ماشین و یادگیری عمیق ساخته شد.

مجموعه COCO شامل بیش از ۳۳۰ هزار تصویر با بیش از ۱.۵ میلیون جفت تصویر و برچسب می‌باشد. این تصاویر به اشیاء مختلف در محیط‌های مختلف اشاره دارند و دارای تنوعی از شرایط

نورپردازی، پس‌زمینه، مقیاس و انواع اشیاء هستند. همچنین، این مجموعه داده دارای ۸۰ دسته اشیاء مختلف است که از جمله آنها می‌توان به اشیاء روزمره مانند خودروها، حیوانات، میوه‌ها، ابزارها، مبلمان و غیره اشاره کرد.

از این مجموعه داده به عنوان یک چالش معتبر برای ارزیابی الگوریتم‌ها و مدل‌های مختلف در حوزه تشخیص اشیاء و تفسیر صحنه استفاده می‌شود. این مجموعه داده باعث توسعه و پیشرفت در زمینه‌هایی مانند شناخت تصویری، یادگیری عمیق، شبکه‌های عصبی کانولوشنی (CNN) و ترکیب تشخیص اشیاء با درک صحنه شده است.



شکل ۳۳ چند نمونه از تصاویر مختلف مجموعه داده COCO

۳-٤-٤. مجموعه داده Caltech

مجموعه داده Caltech 101 یک مجموعه داده معروف در زمینه تشخیص اشیاء در تصاویر است. این مجموعه داده به منظور توسعه و ارزیابی الگوریتم‌ها و مدل‌های تشخیص اشیاء در حوزه بینایی ماشین و یادگیری عمیق ایجاد شده است. نام "Caltech 101" اشاره دارد به دانشگاه کالیفرنیای جنوبی (Caltech) که این مجموعه داده از آنجا نامگذاری شده است و "۱۰۱" به تعداد دسته‌های مختلف اشیاء در این مجموعه داده اشاره دارد.

مجموعه داده Caltech 101 دارای ۱۰۱ دسته اشیاء مختلف می‌باشد. این دسته‌ها شامل اشیاء مختلفی مانند حیوانات، وسایل نقلیه، اشیاء روزمره و غیره هستند. همچنین شامل تقریباً ۹،۰۰۰ تصویر از اشیاء مختلف در تنوعی از شرایط نورپردازی، زوایا و پس‌زمینه‌ها است.

مجموعه داده Caltech 101 به عنوان یک مجموعه داده معتبر و استاندارد در زمینه تشخیص اشیاء و یادگیری عمیق مورد استفاده قرار می‌گیرد. این مجموعه داده از ابزارهای معتبر برای ارزیابی عملکرد مدل‌ها و الگوریتم‌های جدید در زمینه تشخیص اشیاء بهره می‌برد.

شکل ۳۴ چند نمونه از تصاویر مختلف مجموعه داده Caltech

۳-٤-٥. مجموعه داده SYNTHIA

مجموعه داده SYNTHIA یک دیتاست مصنوعی است که برای تحقیقات در زمینه بینایی ماشین، یادگیری عمیق و واقع‌گرایی مصنوعی ایجاد شده است. این دیتاست به منظور ارائه یک محیط تستی و شبیه‌سازی واقعی برای تجزیه و تحلیل الگوریتم‌ها و مدل‌های یادگیری ماشینی در شرایط مختلف تصاویر شهری و خودروها تولید شده است. این محموعه دارای ۹۴۰۰ تصویر شهر مجازی در ۱۳ کلاس و شامل تصاویری با نور مصنوعی در شرایط روز و شب است. همچنین دارای تنوع محیطی خوب و میزان شباهت نسبتا بالا به تصاویر واقعی است. این ویژگی امکان آزمایش الگوریتم‌ها و مدل‌ها در شرایط مختلف نوردهی را به ارمغان می‌آورد.

دیتاست SYNTHIA به عنوان یک ابزار تستی برای تحقیقات در زمینه بینایی ماشینی، تشخیص اشیاء و خودروها، ترافیک شهری و سایر حوزه‌های مرتبط با تصاویر شهری و محیط‌های خارجی بسیار مناسب است.

شکل ۳۵ چند نمونه از تصاویر مختلف مجموعه داده SYNTHIA

۳-٤-٦. مجموعه داده KITTI

مجموعه داده KITTI یکی از مجموعه‌های داده معروف در زمینه بینایی ماشین و پردازش تصویر برای وظایف مختلفی از جمله شناسایی اشیاء، تخمین عمق، رهگیری اشیاء و غیره است. این مجموعه داده به ویژه برای تحقیقات مربوط به خودروهای هوش مصنوعی و خودروهای خودران مناسب است. KITTI[1] یک پروژه تحقیقاتی است که در دانشگاه کارلسروهه و دانشگاه مونیخ در آلمان انجام شده است.

این پلتفرم برای تعیین و ارزیابی الگوریتم‌ها و مدل‌های مرتبط با بینایی ماشین در شرایط واقعی و چالش‌آور استفاده می‌شود. وظایف مورد علاقه در این پلتفرم شامل تخمین عمق از تصاویر استریو، جریان نوری[2]، تعیین حرکت دوربین[3]، تشخیص اشیاء سه‌بعدی و پیگیری اشیاء سه‌بعدی می‌باشند.

این مجموعه داده شامل ساعت‌ها فیلم از سناریوهای ترافیکی است که با استفاده از حسگرهای با کیفیت بالا متنوعی مانند حسگرهای RGB، تصویر سیاه و سفید و حسگرهای عمق ضبط شده‌اند. مجموعه داده تشخیص اشیاء KITTI شامل ۷,۴۸۱ تصویر آموزشی و ۷,۵۱۸ تصویر آزمون می‌شود که به مجموع ۸۰,۲۵۶ شیء با برچسب تعلق دارند. این اشیاء به ۸ دسته مختلف تقسیم شده‌اند که عبارتند از: خودرو، ون، کامیون، عابر پیاده، شخص نشسته، دوچرخه‌سوار، تراموا و سایر.

مجموعه داده KITTI شامل چندین زیرمجموعه داده مختلف است، از جمله:

Stereo 2012: برای تخمین عمق از تصاویر استریو.

[1] KITTI Vision Benchmark Suite
[2] optical flow
[3] visual odometry

Vision Odometry: برای تخمین حرکت دوربین.

Object Detection: برای تشخیص و پیگیری اشیاء مانند خودروها و افراد.

Semantic Segmentation: برای تشخیص اجزاء مختلف محیط از جمله جاده، خودروها و درخت‌ها.

Road Detection: برای تشخیص و تعیین مسیر جاده.

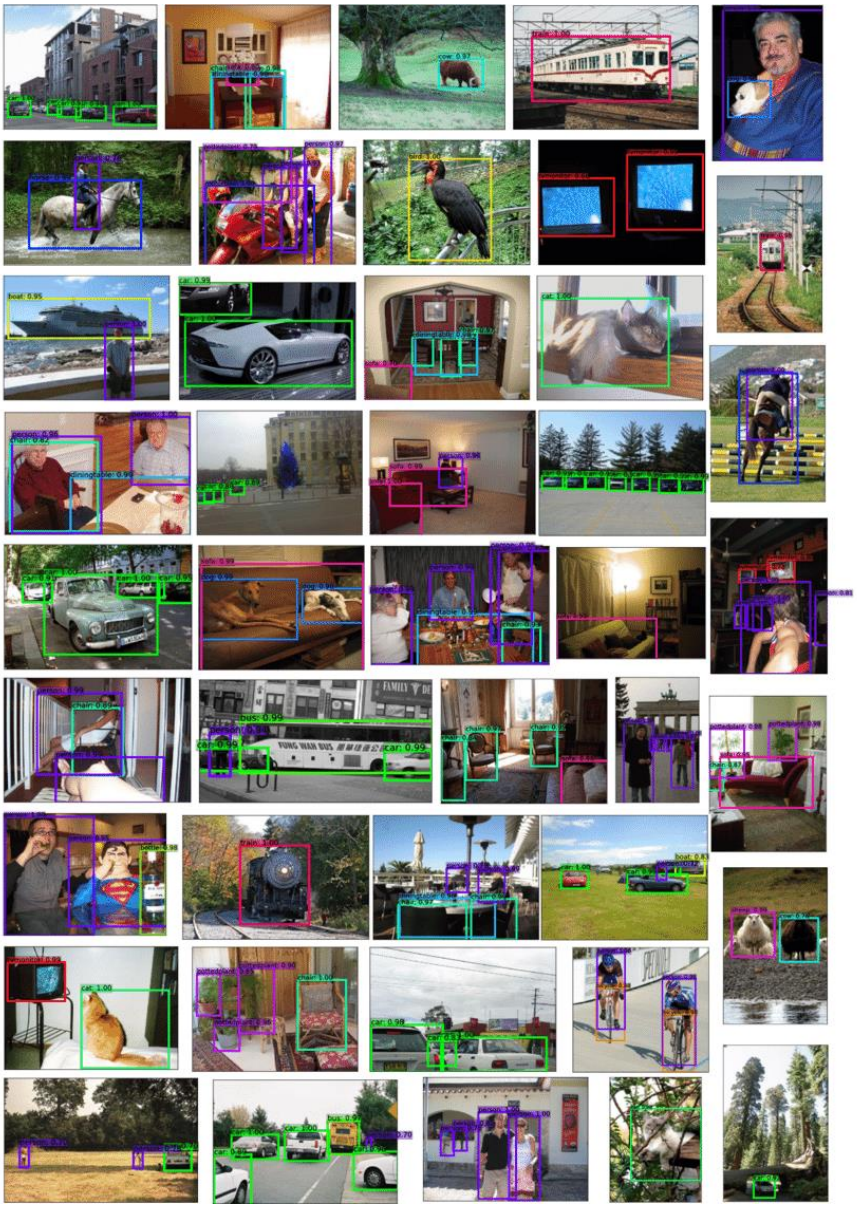
شکل ۳۶ چند نمونه از تصاویر مجموعه داده KITTI

۳-٤-۷. مجموعه داده Pascal VOC

مجموعه داده Pascal VOC یک مجموعه داده واقعی است که شامل اشیاء عمومی است. این مجموعه داده برای وظایف متنوعی مانند دسته‌بندی به مقیاس بزرگ، تشخیص اشیاء و تقسیم‌بندی معرفی شده است. مجموعه تصاویر Pascal VOC در بازه زمانی بین سال ۲۰۰۵ تا ۲۰۱۲ چندین بار به روزرسانی شده است.

در این مجموعه داده، نسخه‌های محبوب‌تر ۲۰۰۷ و ۲۰۱۲ تصاویر با تقسیم‌بندی آموزش و اعتبارسنجی وجود دارند که در مجموع شامل ۱۵,۰۰۰ تصویر هستند. تصاویر موجود در این مجموعه داده بیش از ۲۰ دسته از اشیاء عمومی را شامل می‌شوند که عبارتند از: هواپیما، دوچرخه، پرنده، قایق، بطری، اتوبوس، خودرو، گربه، صندلی، گاو، میز، سگ، اسب، دوچرخه‌سوار، شخص، گیاه، گوسفند، مبل، قطار و تلویزیون.

این مجموعه داده از اهمیت بسیاری در زمینه تحقیقات بینایی ماشین برخوردار بوده و برای ارزیابی و توسعه الگوریتم‌ها و مدل‌های مختلف در وظایف مختلف مانند تشخیص اشیاء و تقسیم‌بندی تصاویر به کار می‌رود.

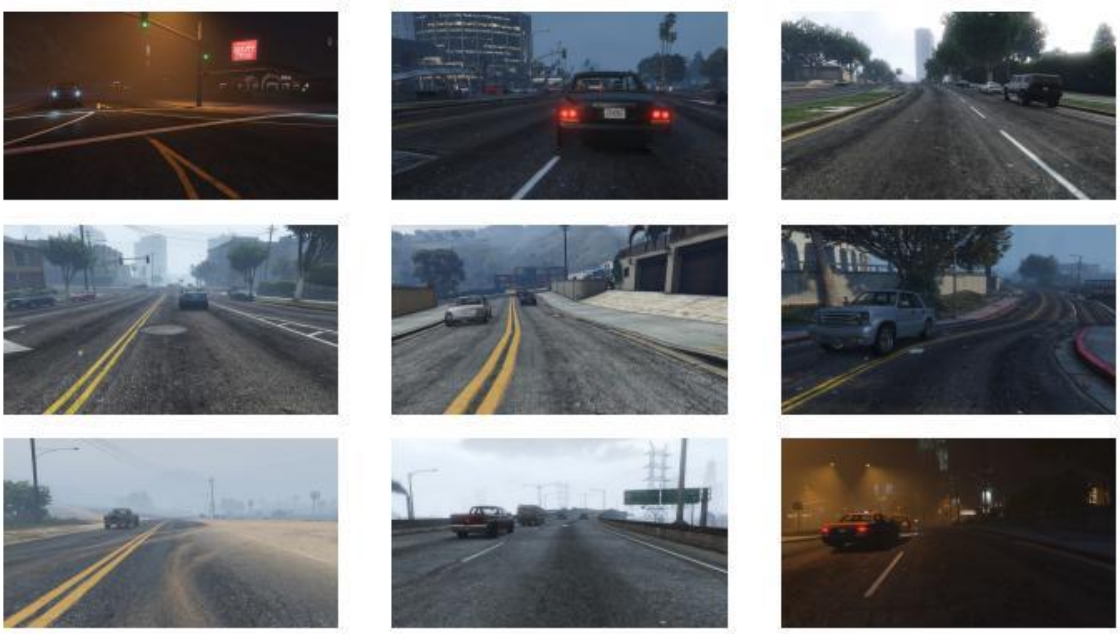


شکل ۳۷ چند نمونه از تصاویر مختلف مجموعه داده Pascal VOC 2007

۳-٤-۸. مجموعه داده SIM10K

مجموعه داده SIM10K از یک موتور شبیه‌سازی(GTA[1]) برای تولید مجموعه داده SIM10K استفاده کردند، که شامل ۱۰۰۰۰ تصویر واقع‌گرایانه از ماشین‌ها در محیط‌های شهری مختلف است. ایجاد شده است.

این مجموعه داده شامل حاشیه‌نویسی‌ها برای وظایف تشخیص اشیاء و تقسیم‌بندی معنایی است، که آن را مناسب برای آموزش شبکه‌های یادگیری عمیق می‌سازد. نویسندگان[68] عملکرد شبکه‌های آموزش دیده بر روی مجموعه داده SIM10K را با شبکه‌های آموزش دیده بر روی مجموعه داده‌های واقعی مقایسه کردند، مانند Cityscapes، و یافتند که داده‌های شبیه‌سازی شده در تمام سطوح پیچیدگی از داده‌های واقعی بهتر عمل کردند. نویسندگان پیشنهاد می‌دهند که این رویکرد می‌تواند برای به سرعت افزایش اندازه مجموعه داده‌های آموزشی مفید از طریق شبیه‌سازی و تأیید آن که مجموعه داده‌های بزرگ کافی است برای یادگیری مدل‌های عمومی مورد استفاده قرار گیرد.



شکل ۳۸ چند نمونه از تصاویر مختلف مجموعه داده SIM10K

۳-٤-۹. مجموعه داده BDD100K

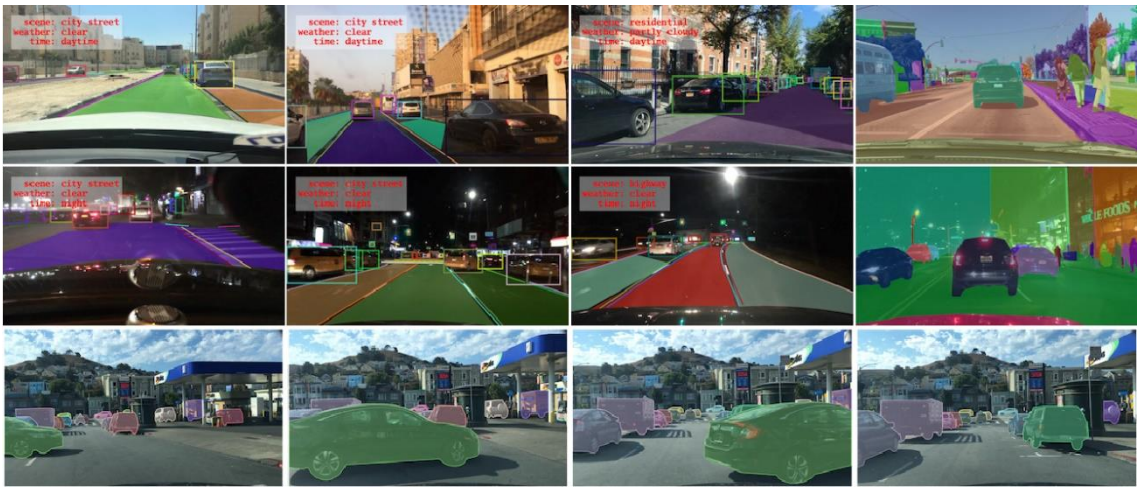
این مجموعه داده در حال حاضر بزرگترین مجموعه برای هوش مصنوعی خودران[2] به شمار می‌آید. مجموعه مذکور حاوی بیش از 100000 ویدئوی 1100 ساعت تجربه رانندگی در زمان‌های گوناگون روز

[1] Grand Theft Auto
[2] self-driving AI

و شرایط‌های مختلف آب‌وهوایی است. تصاویر این مجموعه داده دارای توضیحات و متعلق به نواحی نیویورک و سان فرانسیسکو هستند.

مجموعه داده BDD100K برای تسهیل مطالعات الگوریتمی بر روی داده‌های تصویری بزرگ و متنوع و چندین وظیفه طراحی شده است. این مجموعه داده شامل وظایف متنوعی مانند تشخیص اشیاء، تقسیم‌بندی نمونه، تقسیم‌بندی معنایی، تشخیص خطوط، تقسیم‌بندی مناطق رانندگی و موارد دیگر می‌شود. این مجموعه داده برای آموزش مدل‌هایی که می‌توانند ترکیبی از وظایف ادراکی با پیچیدگی‌های مختلف انجام دهند، مناسب است، به جای انجام تنها وظایف یکسان با ساختار پیش‌بینی یکسان.

این مجموعه داده بزرگ‌تر و متنوع‌تر از مجموعه‌های داده موجود برای رانندگی خودکار است و دارای حاشیه‌نویسی‌های غنی و توزیع جغرافیایی متنوعی است. همچنین، این مجموعه داده یک مجموعه داده MOTS را ارائه می‌دهد که بزرگ‌تر از مجموعه داده‌های KITTI و MOTS Challenge است و تعداد حاشیه‌نویسی‌های آن با مجموعه داده YouTube VOS بزرگ مقایسه می‌شود. به طور کلی، مجموعه داده BDD100K درهای تحقیقات آینده در این حوزه مهم را باز می‌کند و منبع ارزشمندی برای پژوهشگران و توسعه‌دهندگانی است که در برنامه‌های رانندگی خودکار کار می‌کنند.



شکل ۳۹ چند نمونه از تصاویر مختلف مجموعه داده BDD100K

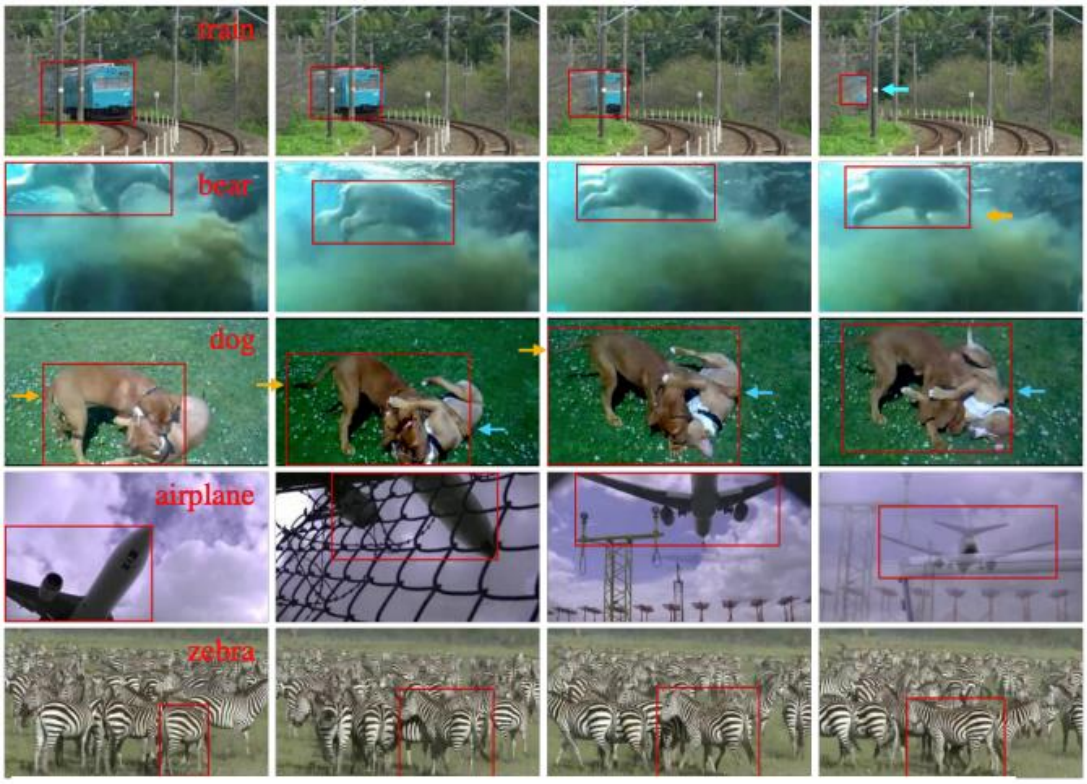
۳-٤-۱۰. مجموعه داده YTBB

YTBB[1] مجموعه‌ای از آدرس‌های ویدئو با برچسب‌های جعبه‌های محدودکننده برای اشیاء با فراوانی داده‌های بزرگ می‌باشد. این مجموعه داده حاوی تقریباً ۳۸۰،۰۰۰ قطعه ویدئویی به طول حدود ۱۹ ثانیه است که به صورت خودکار انتخاب شده‌اند تا اشیاء را در محیط‌های طبیعی بدون ویرایش یا پردازش پس از تولید نمایش دهند. کیفیت ضبط این ویدئوها اغلب شبیه به کیفیت دوربین موبایلی

[1] YouTube BoundingBoxes

دستی است. اشیاء در این ویدئوها یک زیرمجموعه از مجموعه برچسب‌هایMS COCO را نمایش می‌دهند. تمام قطعات ویدئو با برچسب‌های دقیق با دقت بالا و جعبه‌های محدودکننده در هر فریم در ثانیه واحد توسط انسان توسط توسعه‌دهندگان دیتا آناتومی شده‌اند.که برای مسائل تشخیص اشیاء در ویدئوهای YouTube توسعه داده شده است. این مجموعه داده به منظور تعلیم ماشین و توسعه الگوریتم‌های هوش مصنوعی کاربرد دارد.

مجموعه داده YTBB یکی از مجموعه‌های داده مهم برای تحقیقات در حوزه بینایی ماشین و پردازش تصویر است و محققان و توسعه‌دهندگان از آن برای آموزش و ارزیابی مدل‌های مختلف استفاده می‌کنند. این مجموعه داده به توسعه اپلیکیشن‌های تشخیص در ویدئوها و افزایش امکانات تحلیل ویدئویی نیز کمک می‌کند.



شکل ۴۰ چند نمونه از تصاویر مختلف مجموعه داده YTBB

۵-۳. نتایج

در بخش ۲.۴ به موارد و روش های بیشتری در زمینه های مختلف تطبیق دامنه پرداخته شده است. در جدول ۱ نتایج تجربی مقاله هایی که در بخش ۱.۲.۴ بر روی روش های تطبیق دامنه برای کاربرد تشخیص اشیا کار کرده اند، آورده شده است.

شماره	روش	Labeled target	Basic detector	مجموعه داده	دقت (mAP)
1	Khodabandeh و همکارانش[41]	Unsupervised	Faster RCNN	SIM10k→ Cityscapes Cityscapes→ Foggy[1] Cityscapes → KITTI KITTI → Cityscapes	car AP: 42.6, oracle: 68.1 mAP: 36.5, oracle: 43.5 car AP: 77.6, oracle: 90.1 car AP: 43.0, oracle: 68.1
2	Cai و همکارانش[42]	Unsupervised	Faster RCNN	SIM10k → Cityscapes Cityscapes → Foggy Synthetic → COCO Synthetic → YTBB	car AP: 46.6 mAP: 35.1 mAP: 20.7 mAP: 22.8
3	K. Saito و همکارانش[43]	Unsupervised	Faster RCNN	SIM 10k → Cityscapes Cityscapes → Foggy PASCAL → Clipart PASCAL → Watercolor	car AP:47.7 34.3 38.1 53.3
4	Zhenwei He و همکارانش [44]	Unsupervised	Faster RCNN	SIM 10k → Cityscapes Cityscapes → Foggy KITTI → Cityscapes Cityscapes → KITTI	car AP: 41.1 34.0 car AP :41.0 car AP:72.1
5	Yuhua Chen و همکارانش [45]	Unsupervised	Faster RCNN	Cityscapes → Foggy Cityscapes → KITTI KITTI → Cityscapes	27.6 64.1 38.5
6	Lukas Hoyer و همکارانش [46]	Unsupervised	Faster RCNN	Cityscapes → Foggy	47.6
7	Giulio Mattolin و همکارانش [47]	Unsupervised	YOLOv5	Sim10K → Cityscapes KITT → Cityscapes Cityscapes → Foggy	56.3 52.2 40.8
8	Arruda و همکارانش [51]	Unsupervised	Faster RCNN	BDD100k (day) → night	41.9/67.0

[1] Foggy Cityscapes

9	Jiaxi Wu و همکارانش [52]	Unsupervised	Faster RCNN	Cityscapes + KITTI BDD100K day + night Cityscapes + MS COCO + Synscapes	58.4 39.8 37.1
10	Han-Kai Hsu و همکارانش [53]	Unsupervised	Faster RCNN	KITTI → Cityscapes Cityscapes → Foggy Cityscapes →BDD100k	car AP: 43.9,oracle: 55.8 36.9,oracle: 39.2 24.3,oracle: 43.3
11	T. Kim و همکارانش[54]	Unsupervised	Faster RCNN	PASCAL → Clipart1k PASCAL → Watercolor2k PASCAL → Comic2k Cityscapes → Foggy	41.8 52.0 34.5 34.6
12	Yue و همکارانش [60]	Unsupervised	Faster RCNN Yolov5 X	Cityscapes → Foggy(Yolov5 X) Sim10K → Cityscapes KITTI → Cityscapes ViPeD → COCO (Yolov5 X)	44.3 64.5 64.1 47.4

جدول ۱ نتایج عملی تطبیق دامنه برای تشخیص اشیا در مقالات متناظر

فصل ۴: نتیجه‌گیری و پیشنهادها

۴-۱. نتیجه‌گیری

تحقیق حاضر به بررسی و مقایسه روش‌های مختلف تطبیق دامنه در شبکه‌های عصبی عمیق برای کاربرد تشخیص اشیا در هنگام استفاده از داده‌های آموزشی مصنوعی پرداخت. در این تحقیق، ابتدا دسته‌بندی کلی از روش‌های تطبیق دامنه ارائه شد و دسته‌های مختلفی مانند روش‌های مبتنی بر اختلاف، روش‌های خصمانه، تطابق چند دامنه، روش‌های ترکیبی، تطبیق دامنه چند حالتی و روش‌های نوظهور معرفی شدند. سپس، هر یک از دسته‌های مذکور با جزئیات بیشتر مورد بررسی قرار گرفت و مقالاتی که برای تطبیق دامنه در کاربرد تشخیص اشیا مورد استفاده قرار گرفته بودند، در دسته‌های متناظر با موضوعاتشان قرار گرفتند و نقدهایی درباره نقاط قوت و ضعف هر روش ارائه شد.

در ادامه، مجموعه داده‌هایی که در این مقالات برای آموزش و ارزیابی استفاده شدند، توضیح داده شدند. این مجموعه‌های داده از اهمیت بسیاری برخوردار هستند زیرا کیفیت و تنوع داده‌ها می‌تواند تأثیر مستقیمی بر عملکرد تطبیق دامنه داشته باشد.

در نهایت، با تجمیع نتایج مقالات مورد بررسی و مقایسه درصدهای دقت به دست آمده در هر یک از روش‌ها، مشخص شد که کدام روش‌ها در کاربرد تشخیص اشیا با استفاده از داده‌های آموزشی مصنوعی بهترین عملکرد را داشته‌اند. این نتایج می‌توانند به پژوهشگران و متخصصان در حوزه شبکه‌های عصبی عمیق کمک کنند تا روش‌های بهتری برای تطبیق دامنه در این کاربرد توسعه دهند و عملکرد سیستم‌های تشخیص اشیا را بهبود ببخشند.

در مجموع، تحقیق حاضر نشان می‌دهد که تطبیق دامنه در شبکه‌های عصبی عمیق با استفاده از روش‌های مناسب می‌تواند کارایی و دقت سیستم‌های تشخیص اشیا را افزایش دهد و بهبود بخشد. این تحقیق به عنوان یک منبع مفید برای پژوهش‌های آتی در این حوزه مطالعه و مورد استفاده قرار گیرد.

۲-۴. پیشنهاد کارهای آینده

با توجه به مطالعات انجام شده و روش های بررسی شده در راستای آن ها، قصد داریم تا در آینده حداقل با یکی از چند رویکرد زیر پیش برویم :

۱- تبدیل و آزمایش بر روی کارهایی که در موضوع تطبیق دامنه انجام شده اند اما برای کاربرد تشخیص اشیا نبوده اند؛ مخصوصا در بخش کارهای نوین که در این تحقیق تنها به برخی از آن ها اشاره شد میتوان آزمایش های متعددی را بر روی این روش ها انجام داد تا به نتایج بهتری دست یافت.

یکی از روش‌هایی که میتوان به آن اشاره کرد، استفاده از Batch Normalization برای تطبیق دامنه است. این روش برای کار طبقه بندی در مقاله[69] استفاده شده است، و قصد داریم تا با آزمایش-های متخلف روی تشخیص دهنده های مختلفی مثل Faster RCNN، YOLO و غیره، از این روش برای کار تشخیص اشیا استفاده کنیم.

۲- بهبود نتایج برخی از کارهای انجام شده در زمینه تطبیق دامنه برای تشخیص اشیا

۳- بررسی تاثیر دامنه میانی در عملکرد مدل و نحوه چگونگی تولید آن

در کارهای تطبیق دامنه اصولا دو بخش اساسی یک مرحله ای و چند مرحله ای وجود دارد: روش یک مرحله‌ای: در این روش، تطبیق دامنه به صورت مستقیم و در یک مرحله صورت می‌گیرد. به عبارت دیگر، برای انتقال دانش از یک دامنه مبدأ به یک دامنه هدف، از یک مدل یادگیری ماشینی به صورت مستقیم استفاده می‌شود. بدین ترتیب، داده‌های دامنه مبدأ و دامنه هدف در یک مستقل‌سازی تصادفی آورده می‌شوند و مدل عملکرد خود را بهبود می‌بخشد.

روش چند مرحله‌ای: در این روش، تطبیق دامنه از طریق مراحل متعدد صورت می‌گیرد. در اینجا، تلاش برای کاهش تفاوت بین دامنه مبدأ و دامنه هدف در بین مراحل اصلی روش وجود دارد تا به تدریج دانش به دامنه هدف منتقل شود. یک روش رایج در این رویکرد، استفاده از یادگیری نیمه‌نظارتی است که علاوه بر داده‌های برچسب‌دار از دامنه مبدأ، از داده‌های بدون برچسب از دامنه هدف نیز استفاده می‌کند. در مراحل متعدد، مدل با استفاده از داده‌های دامنه هدف خود را بهبود می‌بخشد و تطبیق دامنه پیشرفت می‌کند.

در روش‌های چند مرحله‌ای برای تطبیق دامنه، دامنه میانی[1] نقش بسیار مهم دارد. این دامنه میانی در مراحل مختلف تطبیق دامنه به عنوان یک مرحله واسط عمل می‌کند تا اطلاعات از دامنه مبدا به دامنه هدف منتقل شوند.

در مرحله اول، یادگیری از داده‌های دامنه مبدا صورت می‌گیرد و یک مدل مبتنی بر داده‌های دامنه مبدا آموزش داده می‌شود. این مدل تلاش می‌کند ویژگی‌های مهم و تطبیق‌پذیری از دامنه مبدا را یاد بگیرد. سپس در مرحله بعد، از مدلی که در مرحله اول آموزش داده شده استفاده می‌شود تا ویژگی‌های مهمی که از دامنه مبدا یاد گرفته است را به دامنه میانی انتقال دهد. این انتقال ویژگی‌ها می‌تواند از روش‌های مختلفی مانند تبدیل‌های مخصوص یا شبکه‌های تبدیل دامنه استفاده کند. و در مرحله آخر، ویژگی‌های تطبیق‌داده شده از دامنه میانی به دامنه هدف انتقال داده می‌شوند تا مدل بتواند بر روی داده‌های دامنه هدف عملکرد خوبی داشته باشد. این مرحله نیز ممکن است با استفاده از تبدیل‌های یا شبکه‌های تبدیل دامنه انجام شود.

[1] Intermediate Domain

مراجع

[1] ‘Unsupervised domain adaptation for object detection and whole slide image classification.pdf’.

[2] K. He, X. Zhang, S. Ren, and J. Sun, ‘Deep Residual Learning for Image Recognition’. arXiv, Dec. 10, 2015. Accessed: Sep. 16, 2023. [Online]. Available: http://arxiv.org/abs/1512.03385

[3] K. He, X. Zhang, S. Ren, and J. Sun, ‘Identity Mappings in Deep Residual Networks’. arXiv, Jul. 25, 2016. Accessed: Sep. 16, 2023. [Online]. Available: http://arxiv.org/abs/1603.05027

[4] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, ‘OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks’. arXiv, Feb. 23, 2014. Accessed: Sep. 16, 2023. [Online]. Available: http://arxiv.org/abs/1312.6229

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, ‘You Only Look Once: Unified, Real-Time Object Detection’, in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.

[6] W. Liu *et al.*, ‘SSD: Single Shot MultiBox Detector’, 2016, pp. 21–37. doi: 10.1007/978-3-319-46448-0_2.

[7] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, ‘Feature Pyramid Networks for Object Detection’. arXiv, Apr. 19, 2017. Accessed: Sep. 17, 2023. [Online]. Available: http://arxiv.org/abs/1612.03144

[8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, ‘Rich feature hierarchies for accurate object detection and semantic segmentation’. arXiv, Oct. 22, 2014. Accessed: Sep. 17, 2023. [Online]. Available: http://arxiv.org/abs/1311.2524

[9] R. Girshick, ‘Fast R-CNN’. arXiv, Sep. 27, 2015. Accessed: Sep. 17, 2023. [Online]. Available: http://arxiv.org/abs/1504.08083

[10] S. Ren, K. He, R. Girshick, and J. Sun, ‘Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks’. arXiv, Jan. 06, 2016. Accessed: Sep. 17, 2023. [Online]. Available: http://arxiv.org/abs/1506.01497

[11] ‘Domain_Adaptation_Challenges_Methods_Datasets_and_Applications.pdf’.

[12] M. Wang and W. Deng, ‘Deep Visual Domain Adaptation: A Survey’, *Neurocomputing*, vol. 312, pp. 135–153, Oct. 2018, doi: 10.1016/j.neucom.2018.05.083.

[13] A. de Mathelin, F. Deheeger, M. Mougeot, and N. Vayatis, ‘Discrepancy-Based Active Learning for Domain Adaptation’. arXiv, Sep. 14, 2022. Accessed: Aug. 30, 2023. [Online]. Available: http://arxiv.org/abs/2103.03757

[14] P. Singhal, R. Walambe, S. Ramanna, and K. Kotecha, ‘Domain Adaptation: Challenges, Methods, Datasets, and Applications’, *IEEE Access*, vol. 11, pp. 6973–7020, 2023, doi: 10.1109/ACCESS.2023.3237025.

[15] ‘Deep Domain Confusion و Maximizing for Domain Invariance.pdf’.

[16] K. M. Borgwardt, A. Gretton, M. J. Rasch, H.-P. Kriegel, B. Schölkopf, and A. J. Smola, ‘Integrating structured biological data by Kernel Maximum Mean Discrepancy’, *Bioinformatics*, vol. 22, no. 14, pp. e49–e57, Jul. 2006, doi: 10.1093/bioinformatics/btl242.

[17] A. R. Kashyap, D. Hazarika, M.-Y. Kan, and R. Zimmermann, ‘Domain Divergences: a Survey and Empirical Analysis’. arXiv, Apr. 19, 2021. Accessed: Aug. 30, 2023. [Online]. Available: http://arxiv.org/abs/2010.12198

[18] Y. Pan, T. Yao, Y. Li, Y. Wang, C.-W. Ngo, and T. Mei, ‘Transferrable Prototypical Networks for Unsupervised Domain Adaptation’. arXiv, Apr. 25, 2019. Accessed: Aug. 29, 2023. [Online]. Available: http://arxiv.org/abs/1904.11227

[19] I. J. Goodfellow *et al.*, ‘Generative Adversarial Networks’. arXiv, Jun. 10, 2014. Accessed: Aug. 31, 2023. [Online]. Available: http://arxiv.org/abs/1406.2661

[20] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, ‘Adversarial Discriminative Domain Adaptation’. arXiv, Feb. 17, 2017. Accessed: Aug. 31, 2023. [Online]. Available: http://arxiv.org/abs/1702.05464

[21] J. Hoffman, D. Wang, F. Yu, and T. Darrell, ‘FCNs in the Wild: Pixel-level Adversarial and Constraint-based Adaptation’. arXiv, Dec. 08, 2016. Accessed: Sep. 04, 2023. [Online]. Available: http://arxiv.org/abs/1612.02649

[22] H. Huang, Q. Huang, and P. Krähenbühl, ‘Domain Transfer Through Deep Activation Matching’, in *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., in Lecture Notes in Computer Science, vol. 11220. Cham: Springer International Publishing, 2018, pp. 611–626. doi: 10.1007/978-3-030-01270-0_36.

[23] M. Long, Z. Cao, J. Wang, and M. I. Jordan, ‘Conditional Adversarial Domain Adaptation’. arXiv, Dec. 29, 2018. Accessed: Sep. 03, 2023. [Online]. Available: http://arxiv.org/abs/1705.10667

[24] J. Shen, Y. Qu, W. Zhang, and Y. Yu, ‘Wasserstein Distance Guided Representation Learning for Domain Adaptation’. arXiv, Mar. 09, 2018. Accessed: Sep. 04, 2023. [Online]. Available: http://arxiv.org/abs/1707.01217

[25] M. Arjovsky, S. Chintala, and L. Bottou, ‘Wasserstein GAN’. arXiv, Dec. 06, 2017. Accessed: Sep. 04, 2023. [Online]. Available: http://arxiv.org/abs/1701.07875

[26] Y. Du, Z. Tan, Q. Chen, X. Zhang, Y. Yao, and C. Wang, ‘Dual Adversarial Domain Adaptation’. arXiv, Jan. 01, 2020. Accessed: Sep. 03, 2023. [Online]. Available: http://arxiv.org/abs/2001.00153

[27] H. Rangwani, S. K. Aithal, M. Mishra, A. Jain, and R. V. Babu, ‘A Closer Look at Smoothness in Domain Adversarial Training’. arXiv, Jun. 16, 2022. Accessed: Sep. 01, 2023. [Online]. Available: http://arxiv.org/abs/2206.08213

[28] G. Shi *et al.*, ‘Genre Separation Network with Adversarial Training for Cross-genre Relation Extraction’, in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium: Association for Computational Linguistics, 2018, pp. 1018–1023. doi: 10.18653/v1/D18-1125.

[29] K. Bousmalis, G. Trigeorgis, N. Silberman, D. Krishnan, and D. Erhan, 'Domain Separation Networks'. arXiv, Aug. 21, 2016. Accessed: Sep. 01, 2023. [Online]. Available: http://arxiv.org/abs/1608.06019

[30] S. Sun, H. Shi, and Y. Wu, 'A survey of multi-source domain adaptation', *Inf. Fusion*, vol. 24, pp. 84–92, Jul. 2015, doi: 10.1016/j.inffus.2014.12.003.

[31] S. Zhao, B. Li, C. Reed, P. Xu, and K. Keutzer, 'Multi-source Domain Adaptation in the Deep Learning Era: A Systematic Survey'. arXiv, Feb. 26, 2020. Accessed: Sep. 01, 2023. [Online]. Available: http://arxiv.org/abs/2002.12169

[32] S. Zhao *et al.*, 'Multi-source Domain Adaptation for Semantic Segmentation'. arXiv, Oct. 27, 2019. Accessed: Sep. 01, 2023. [Online]. Available: http://arxiv.org/abs/1910.12181

[33] B. Gholami, P. Sahu, O. Rudovic, K. Bousmalis, and V. Pavlovic, 'Unsupervised Multi-Target Domain Adaptation: An Information Theoretic Approach'. arXiv, Oct. 26, 2018. Accessed: Sep. 02, 2023. [Online]. Available: http://arxiv.org/abs/1810.11547

[34] T. Isobe *et al.*, 'Multi-Target Domain Adaptation with Collaborative Consistency Learning'. arXiv, Jun. 07, 2021. Accessed: Sep. 02, 2023. [Online]. Available: http://arxiv.org/abs/2106.03418

[35] L. T. Nguyen-Meidine, A. Belal, M. Kiran, J. Dolz, L.-A. Blais-Morin, and E. Granger, 'Unsupervised Multi-Target Domain Adaptation Through Knowledge Distillation'. arXiv, Nov. 19, 2020. Accessed: Sep. 02, 2023. [Online]. Available: http://arxiv.org/abs/2007.07077

[36] F. Qi, X. Yang, and C. Xu, 'A Unified Framework for Multimodal Domain Adaptation', *Proc. 26th ACM Int. Conf. Multimed.*, 2018, [Online]. Available: https://api.semanticscholar.org/CorpusID:53034461

[37] X. Shu, G.-J. Qi, J. Tang, and J. Wang, 'Weakly-Shared Deep Transfer Networks for Heterogeneous-Domain Knowledge Propagation', Oct. 2015, pp. 35–44. doi: 10.1145/2733373.2806216.

[38] Z. Ding, M. Shao, and Y. Fu, 'Missing Modality Transfer Learning via Latent Low-Rank Constraint', *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4322–4334, 2015, doi: 10.1109/TIP.2015.2462023.

[39] K.-C. Peng, Z. Wu, and J. Ernst, 'Zero-Shot Deep Domain Adaptation'. arXiv, Jul. 24, 2018. Accessed: Sep. 04, 2023. [Online]. Available: http://arxiv.org/abs/1707.01922

[40] E. Kodirov, T. Xiang, Z. Fu, and S. Gong, 'Unsupervised Domain Adaptation for Zero-Shot Learning', in *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile: IEEE, Dec. 2015, pp. 2452–2460. doi: 10.1109/ICCV.2015.282.

[41] 'A Robust Learning Approach to Domain Adaptive Object Detection.pdf'.

[42] Q. Cai, Y. Pan, C.-W. Ngo, X. Tian, L. Duan, and T. Yao, 'Exploring Object Relation in Mean Teacher for Cross-Domain Detection'. arXiv, Dec. 25, 2019. Accessed: Aug. 29, 2023. [Online]. Available: http://arxiv.org/abs/1904.11245

[43] K. Saito, Y. Ushiku, T. Harada, and K. Saenko, 'Strong-Weak Distribution Alignment for Adaptive Object Detection'. arXiv, Apr. 05, 2019. Accessed: Sep. 05, 2023. [Online]. Available: http://arxiv.org/abs/1812.04798

[44] Z. He and L. Zhang, 'Multi-adversarial Faster-RCNN for Unrestricted Object Detection'. arXiv, Sep. 06, 2019. Accessed: Sep. 05, 2023. [Online]. Available: http://arxiv.org/abs/1907.10343

[45] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. Van Gool, 'Domain Adaptive Faster R-CNN for Object Detection in the Wild'. arXiv, Mar. 08, 2018. Accessed: Sep. 05, 2023. [Online]. Available: http://arxiv.org/abs/1803.03243

[46] L. Hoyer, D. Dai, H. Wang, and L. Van Gool, 'MIC: Masked Image Consistency for Context-Enhanced Domain Adaptation'. arXiv, Mar. 24, 2023. Accessed: Sep. 06, 2023. [Online]. Available: http://arxiv.org/abs/2212.01322

[47] G. Mattolin, L. Zanella, E. Ricci, and Y. Wang, 'ConfMix: Unsupervised Domain Adaptation for Object Detection via Confidence-based Mixing'. arXiv, Oct. 20, 2022. Accessed: Sep. 06, 2023. [Online]. Available: http://arxiv.org/abs/2210.11539

[48] J. Hoffman *et al.*, 'CyCADA: Cycle-Consistent Adversarial Domain Adaptation'. arXiv, Dec. 29, 2017. Accessed: Sep. 01, 2023. [Online]. Available: http://arxiv.org/abs/1711.03213

[49] M. Ghifary, W. B. Kleijn, M. Zhang, D. Balduzzi, and W. Li, 'Deep Reconstruction-Classification Networks for Unsupervised Domain Adaptation'. arXiv, Aug. 01, 2016. Accessed: Sep. 01, 2023. [Online]. Available: http://arxiv.org/abs/1607.03516

[50] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, 'Learning to Discover Cross-Domain Relations with Generative Adversarial Networks'. arXiv, May 15, 2017. Accessed: Sep. 01, 2023. [Online]. Available: http://arxiv.org/abs/1703.05192

[51] V. F. Arruda *et al.*, 'Cross-Domain Car Detection Using Unsupervised Image-to-Image Translation: From Day to Night', in *2019 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2019, pp. 1–8. doi: 10.1109/IJCNN.2019.8852008.

[52] J. Wu *et al.*, 'Target-Relevant Knowledge Preservation for Multi-Source Domain Adaptive Object Detection'. arXiv, Apr. 17, 2022. Accessed: Sep. 06, 2023. [Online]. Available: http://arxiv.org/abs/2204.07964

[53] H.-K. Hsu *et al.*, 'Progressive Domain Adaptation for Object Detection'. arXiv, Oct. 24, 2019. Accessed: Sep. 05, 2023. [Online]. Available: http://arxiv.org/abs/1910.11319

[54] T. Kim, M. Jeong, S. Kim, S. Choi, and C. Kim, 'Diversify and Match: A Domain Adaptive Representation Learning Paradigm for Object Detection'. arXiv, May 14, 2019. Accessed: Sep. 05, 2023. [Online]. Available: http://arxiv.org/abs/1905.05396

[55] A. Tarvainen and H. Valpola, 'Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results'. arXiv, Apr. 16, 2018. Accessed: Sep. 02, 2023. [Online]. Available: http://arxiv.org/abs/1703.01780

[56] G. French, M. Mackiewicz, and M. Fisher, ‘Self-ensembling for visual domain adaptation’. arXiv, Sep. 23, 2018. Accessed: Sep. 02, 2023. [Online]. Available: http://arxiv.org/abs/1706.05208

[57] X. Peng, Q. Bai, X. Xia, Z. Huang, K. Saenko, and B. Wang, ‘Moment Matching for Multi-Source Domain Adaptation’. arXiv, Aug. 27, 2019. Accessed: Sep. 01, 2023. [Online]. Available: http://arxiv.org/abs/1812.01754

[58] S. Motiian, Q. Jones, S. M. Iranmanesh, and G. Doretto, ‘Few-Shot Adversarial Domain Adaptation’. arXiv, Nov. 05, 2017. Accessed: Sep. 04, 2023. [Online]. Available: http://arxiv.org/abs/1711.02536

[59] A. Zhao *et al.*, ‘Domain-Adaptive Few-Shot Learning’. arXiv, Mar. 19, 2020. Accessed: Sep. 04, 2023. [Online]. Available: http://arxiv.org/abs/2003.08626

[60] X. Yue *et al.*, ‘Prototypical Cross-domain Self-supervised Learning for Few-shot Unsupervised Domain Adaptation’. arXiv, Mar. 30, 2021. Accessed: Sep. 04, 2023. [Online]. Available: http://arxiv.org/abs/2103.16765

[61] Y. Gao, K.-Y. Lin, J. Yan, Y. Wang, and W.-S. Zheng, ‘AsyFOD: An Asymmetric Adaptation Paradigm for Few-Shot Domain Adaptive Object Detection’.

[62] J. Na, H. Jung, H. J. Chang, and W. Hwang, ‘FixBi: Bridging Domain Spaces for Unsupervised Domain Adaptation’. arXiv, Mar. 25, 2021. Accessed: Sep. 18, 2023. [Online]. Available: http://arxiv.org/abs/2011.09230

[63] A. Bituitskii, ‘Unsupervised Domain Adaptation with Deep Neural-Network’. arXiv, Jul. 10, 2023. Accessed: Sep. 18, 2023. [Online]. Available: http://arxiv.org/abs/2307.05601

[64] G. Degola, ‘Approaches for Domain Adaptive Object Detection in Production Environments’.

[65] Y. Yang, ‘Unsupervised domain adaptation for object detection and whole slide image classification’.

[66] ‘توسعه ی شبکه های تقابلی عمیق برای تطبیق دامنه.pdf’.

[67] ‘ارائه یک چهارچوب جامع برای تطبیق دامنه و وجه در داده های چند وجهی.pdf’.

[68] M. Johnson-Roberson, C. Barto, R. Mehta, S. N. Sridhar, K. Rosaen, and R. Vasudevan, ‘Driving in the Matrix: Can Virtual Worlds Replace Human-Generated Annotations for Real World Tasks?’ arXiv, Feb. 25, 2017. Accessed: Sep. 05, 2023. [Online]. Available: http://arxiv.org/abs/1610.01983

[69] W.-G. Chang, T. You, S. Seo, S. Kwak, and B. Han, ‘Domain-Specific Batch Normalization for Unsupervised Domain Adaptation’. arXiv, May 27, 2019. Accessed: Sep. 16, 2023. [Online]. Available: http://arxiv.org/abs/1906.03950

In the name of God

Malek-ashtar University of Technology

Faculty of Electrical and Computer

M.Sc. Seminar

Title

Domain adaptation in deep neural networks for object detection applications when using synthetic training data

By

Helia Mohamadi

Supervised by

Dr. Mohammad Ali Keyvanrad

September. 2023