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 با استفاده از داده آموزشی مصنوعی با محدودیت Source-Free

توسط

امیررضا کاظم لو

استاد راهنما

دکتر محمدعلی کیوان‌راد

فهرست مطالب

فهرست مطالب 2

چکیده 6 1

فصل 1: مقدمه 7 1

مقدمه 7 1.1

ضرورت انجام پژوهش 8 1.2

انگیزه و هدف 9 1.3

هدف و سوالات پژوهش (Goal and Research Questions) 9 1.3.1

ساختار سمینار 10 1.4

فصل 2: مروری بر تطبیق دامنه و روش های آن 11 2

مقدمه 11 2.1

تطبیق دامنه در یادگیری عمیق 13 2.2

روش‌های مبتنی بر اختلاف (Discrepancy-based Methods) 14 2.2.1

روشهای خصمانه 18 2.2.2

تطابق چند دامنه ای 26 2.2.3

روش های ترکیبی 27 2.2.4

تطبیق دامنه چند حالته. 29 2.2.5

روش های نوظهور 32 2.2.6

فصل3: مجموعه دادگان و نتایج تجربی: 34 3

مقدمه 34 3.1

3.2 مجموعه دادگان 34
3.2.1 مجموعه داده Cityscapes 34
3.2.2 مجموعه داده Foggy Cityscapes 35
3.2.3 مجموعه داده COCO 36
3.2.4 مجموعه داده Caltech 38
3.2.5 مجموعه داده SYNTHIA 39
3.2.6 مجموعه داده KITTI 40
3.2.7 مجموعه داده Pascal VOC 41
3.2.8 مجموعه داده SIM10K 43
3.2.9 مجموعه داده BDD100K 44
3.2.10 مجموعه داده YTBB 45
3.3 نتایج 46

فهرست جداول

جدول 1 - نتایج عملی تطبیق دامنه برای تشخیص اشیا 47

جدول 2 ..... **Error! Bookmark not defined.**

فهرست اشکال

شکل 1 - درخت واره روش های تطبیق دامنه 14

شکل 2 ساختار Robust Faster-RCNN-این شکل، ساختار را نشان می‌دهد که تطبیق دامنه را از طریق یادگیری با برچسب‌های نویزی انجام می‌دهد.در این روش، ابتدا مدلِ آموزش‌دیده روی دامنه منبع، شبه-برچسب‌های نویزی برای داده‌های هدف تولید می‌کن سپس یک ماژول پالایش، کیفیت این برچسب‌ها را مدل‌سازی کرده و از یادگیری خطاها جلوگیری می‌کند 17

شکل 3 - چند نمونه از تصاویر مختلف مجموعه داده Cityscapes 35

شکل 4 - مجموعه داده Cityscapes شکل 20 (سمت چپ) در مقابل Foggy Cityscapes(سمت راست) 36

شکل 5 - چند نمونه از تصاویر مختلف مجموعه داده COCO مجموعه داده 37

شکل 6 - چند نمونه از تصاویر مختلف مجموعه داده COCO 38

شکل 7 - چند نمونه از تصاویر مختلف مجموعه داده Caltech 39

شکل 8 - چند نمونه از تصاویر مختلف مجموعه داده SYNTHIA 39

شکل 9 - چند نمونه از تصاویر محموعه داده KITTI 41

شکل 10 - چند نمونه از تصاویر مختلف مجموعه داده Pascal VOC 2007 42

شکل 11 - چند نمونه از تصاویر مختلف مجموعه داده SIM10K 43

شکل 12 - چند نمونه از تصاویر مختلف مجموعه داده BDD100K 44

شکل 13 - چند نمونه از تصاویر مختلف مجموعه داده YTBB 45

1 چکیده

شبکه‌های عصبی عمیق (DNNs)[1] به موفقیت‌های چشمگیری در وظیفه تشخیص اشیاء دست یافته‌اند. با این حال، عملکرد این مدل‌ها هنگامی که از دامنه منبع (Source Domain) که در آن آموزش دیده‌اند، مانند داده‌های مصنوعی(Synthetic Data)، به یک دامنه هدف (Target Domain) جدید، مانند تصاویر دنیای واقعی، منتقل می‌شوند، به دلیل پدیده "شکاف دامنه (Domain Shift) " با افت شدید مواجه می‌شود.

برای حل این چالش، رویکردهای تطبیق دامنه بدون نظارت (UDA) توسعه یافته‌اند که تلاش می‌کنند مدل را با داده‌های بدون برچسب از دامنه هدف سازگار کنند. بسیاری از این روش‌های اولیه، با الهام از وظایف طبقه‌بندی، به تطبیق سراسری ویژگی‌ها در سطح کل تصویر می‌پردازند. اما این رویکرد برای تشخیص اشیاء، که یک وظیفه محلی‌سازی شده (localized) است، ایده‌آل نیست، زیرا تغییرات دامنه ممکن است بر اشیاء و پس‌زمینه به صورت متفاوتی تأثیر بگذارد.

اخیراً، یک محدودیت عملی‌تر و چالش‌برانگیزتر تحت عنوان "تطبیق دامنه بدون منبع - (Source-Free Domain Adaptation SFDA) "مطرح شده است. در این سناریو، فرآیند تطبیق تنها با دسترسی به مدل از پیش آموزش‌دیده و داده‌های دامنه هدف انجام می‌شود و هیچ دسترسی به داده‌های اصلی دامنه منبع (داده‌های مصنوعی) وجود ندارد. این محدودیت، به دلیل مسائل مربوط به حریم خصوصی یا حجم بالای داده منبع، در کاربردهای واقعی بسیار رایج است.

این سمینار به بررسی رویکردها و چالش‌های موجود در تطبیق دامنه برای تشخیص اشیاء، با تمرکز ویژه بر سناریوی Source-Free و استفاده از داده‌های مصنوعی به عنوان منبع اولیه آموزش، می‌پردازد.

کلمات کلیدی : تشخیص اشیاء، داده مصنوعی، تطبیق دامنه، روش‌های مبتنی بر ویژگی تطبیق دامنه.

[1] Deep Neural Networks

1 فصل 1: مقدمه

1.1 مقدمه

شبکه‌های عصبی عمیق (DNNs) در حوزه‌های مختلف بینایی ماشین، به‌ویژه در تشخیص اشیاء، به پیشرفت‌های چشمگیری دست یافته و دقتی نزدیک به سطح انسان را به نمایش گذاشته‌اند. با وجود این موفقیت‌ها، این مدل‌ها هنوز از نظر توانایی با سیستم بینایی انسان فاصله دارند. بینایی انسان به مراتب استواری بیشتری در مقابل تغییرات محیطی، نویزهای بصری و شرایط متغیر پیرامون نشان می‌دهد. تحقیقات متعددی به صورت تجربی ثابت کرده‌اند که افزودن نویز یا تغییرات جزئی در پس‌زمینه تصاویر، می‌تواند به شدت بر اطمینان و دقت پیش‌بینی‌های مدل‌های DNN تأثیر منفی بگذارد. به همین دلیل، دستیابی به پیش‌بینی‌های قابل اعتماد در محیط‌های پویا و متغیر، یکی از چالش‌های اساسی و حیاتی در این حوزه محسوب می‌شود.

یکی از موانع اصلی در گسترش کاربرد مدل‌های DNN،[1], [2], [3] ضعف آن‌ها در سازگاری با شرایط جدید است. این نقص که به آن "عدم قابلیت انطباق" گفته می‌شود، معمولاً زمانی آشکار می‌گردد که مدلی که بر روی یک یا چند مجموعه داده آموزش

دیده، برای حل همان مسئله روی یک مجموعه داده جدید و متفاوت به کار گرفته می‌شود. در چنین شرایطی، عملکرد مدل به دلیل تفاوت در توزیع آماری داده‌ها بین مجموعه داده‌های آموزشی و آزمایشی، دچار افت قابل‌توجهی می‌شود. این پدیده که به آن "شکاف دامنه" (Domain Gap) نیز می‌گویند، یکی از موانع اصلی در تعمیم‌پذیری مدل‌های یادگیری عمیق است.

1.2 ضرورت انجام پژوهش

یادگیری عمیق، به‌ویژه در زمینه تشخیص اشیاء(Object Detection) ، به پیشرفت‌های چشمگیری دست یافته است. با این حال، یکی از بزرگترین موانع در به‌کارگیری عملی این مدل‌ها، وابستگی شدید آن‌ها به داده‌های آموزشی است. مدلی که بر روی داده‌های یک دامنه منبع (Source Domain) آموزش دیده است، هنگام مواجهه با داده‌های یک دامنه هدف (Target Domain) متفاوت، با افت قابل توجهی در دقت و کارایی روبرو می‌شود. این پدیده که به "شکاف دامنه (Domain Gap) "معروف است، ضرورت انجام پژوهش در زمینه تطبیق دامنه (Domain Adaptation) را برجسته می‌سازد.

اهمیت این مسئله در سناریوی انتقال از داده‌های مصنوعی به دنیای واقعی (Sim-to-Real) دوچندان می‌شود. تولید داده‌های آموزشی مصنوعی (Synthetic) مزایای فراوانی دارد: این داده‌ها ارزان، مقیاس‌پذیر و دارای برچسب‌های دقیق و کامل هستند و می‌توانند شرایط خاص یا خطرناکی (مانند تصادفات رانندگی) را که نمونه‌برداری از آن‌ها در دنیای واقعی دشوار است، شبیه‌سازی کنند [4]. با این حال، این داده‌ها هرگز نمی‌توانند پیچیدگی‌های دنیای واقعی مانند تنوع بی‌پایان در شرایط نوری، بافت‌ها، زوایا و پس‌زمینه‌ها را به طور کامل بازسازی کنند. در نتیجه، مدلی که تنها با داده‌های مصنوعی آموزش دیده، در محیط واقعی عملکرد مطلوبی نخواهد داشت.

از سوی دیگر، جمع‌آوری و برچسب‌زنی حجم عظیمی از داده‌ها در دامنه هدف (دنیای واقعی) برای آموزش مجدد مدل، فرآیندی بسیار پرهزینه و زمان‌بر است [5]. اینجاست که تطبیق دامنه بدون نظارت (UDA) به عنوان یک راه‌حل کلیدی مطرح می‌شود. این رویکردها به مدل امکان می‌دهند تا با بهره‌گیری از دانش آموخته‌شده از دامنه منبع، خود را با داده‌های بدون برچسب دامنه هدف وفق دهد.

اخیراً، یک گام فراتر و عملی‌تر، پژوهش‌ها را به سمت سناریوی "تطبیق دامنه بدون منبع - Source-Free Domain Adaptation) " (SFDAسوق داده است[6]. در این سناریوی واقع‌گرایانه، پس از اتمام آموزش اولیه، دیگر هیچ دسترسی به داده‌های حجیم دامنه منبع (داده‌های مصنوعی) وجود ندارد. این محدودیت به دلایل عملی مانند حفظ حریم خصوصی، ملاحظات مربوط به مالکیت معنوی داده‌ها، یا صرفه‌جویی در هزینه‌های ذخیره‌سازی، اهمیت پیدا می‌کند. بنابراین، چالش اصلی، تطبیق مدلی است که تنها "دانش" را در وزن‌های خود ذخیره کرده است.

این پژوهش با تمرکز بر این چالش‌های مدرن، به دنبال یافتن روش‌هایی است که بتوانند دانش عمومی و الگوهای استخراج‌شده توسط مدل از داده‌های مصنوعی را به طور مؤثری به دامنه واقعی منتقل کنند، آن هم در شرایطی که داده‌های منبع دیگر در

دسترس نیستند. موفقیت در این زمینه می‌تواند راه را برای استفاده گسترده‌تر و اقتصادی‌تر از مدل‌های تشخیص اشیاء در کاربردهای دنیای واقعی هموار سازد.

1.3 انگیزه و هدف

انگیزه اصلی این پژوهش، غلبه بر یکی از بزرگترین موانع در پیاده‌سازی عملی سیستم‌های تشخیص اشیاء در دنیای واقعی است که شکاف دامنه (Domain Gap) بین داده‌های آموزشی و داده‌های عملیاتی. استفاده از داده‌های مصنوعی (Synthetic Data) [7]برای آموزش شبکه‌های عصبی، به دلیل هزینه‌ی پایین، مقیاس‌پذیری بالا و دسترسی به برچسب‌های دقیق، یک راهکار بسیار جذاب است. این داده‌ها به ما اجازه می‌دهند تا مدل‌ها را برای کاربردهای حیاتی مانند خودروهای خودران، رباتیک و نظارت هوشمند، بدون نیاز به جمع‌آوری و برچسب‌زنی دستی هزاران ساعت داده واقعی، آموزش دهیم.

با این حال، مدلی که تنها بر روی داده‌های مصنوعی آموزش دیده، در مواجهه با پیچیدگی‌های دنیای واقعی (مانند شرایط نوری غیرمنتظره، بافت‌های متنوع و انسداد اشیاء) عملکرد ضعیفی از خود نشان می‌دهد. هدف اولیه تطبیق دامنه (Domain Adaptation)، پر کردن این شکاف و بهبود کارایی مدل در دامنه هدف (دنیای واقعی) بدون نیاز به برچسب‌گذاری داده‌های آن است .

این پژوهش یک گام فراتر می‌رود و به سناریوی چالش‌برانگیز و واقع‌گرایانه تطبیق دامنه بدون منبع (Source-Free Domain Adaptation - SFDA) می‌پردازد. انگیزه این رویکرد از محدودیت‌های عملی نشأت می‌گیرد:

1. حریم خصوصی و مالکیت داده : شرکت‌ها و مؤسسات ممکن است به دلیل قوانین حفظ حریم خصوصی مانند GDPR یا ملاحظات تجاری، تمایلی به اشتراک‌گذاری داده‌های آموزشی اصلی خود نداشته باشند.
2. بهره‌وری و هزینه : ذخیره‌سازی و انتقال مجموعه داده‌های عظیم مصنوعی (که ممکن است صدها گیگابایت حجم داشته باشند) برای هر فرآیند تطبیق، ناکارآمد و پرهزینه است.

بنابراین، انگیزه اصلی، توسعه سیستم‌های تشخیص اشیاء مقاوم، کارآمد و عملی است که بتوانند تنها با استفاده از یک مدل از پیش آموزش‌دیده، خود را با محیط‌های جدید و دیده نشده تطبیق دهند[8].

1.3.1 هدف و سوالات پژوهش(Goal and Research Questions)

هدف این سمینار، بررسی عمیق روش‌ها، چالش‌ها و پیشرفت‌های اخیر در حوزه تطبیق دامنه بدون منبع برای تشخیص اشیاء است. این پژوهش به دنبال پاسخ به سوالات اساسی زیر است:

1. چگونه می‌توان دانش نهفته در یک مدل (که روی داده‌های مصنوعی آموزش دیده) را استخراج و از آن برای تطبیق با دامنه هدف (داده‌های واقعی) استفاده کرد، بدون آنکه به داده‌های مصنوعی اصلی دسترسی داشته باشیم؟

2. با توجه به عدم دسترسی به داده‌های منبع، چگونه می‌توان شبه-برچسب‌های (Pseudo-Labels) قابل اعتماد برای اشیاء در دامنه هدف تولید کرد و اثرات نویز ناشی از برچسب‌های نادرست را در طول فرآیند خود-آموزی (Self-Training) کاهش داد؟

3. چه مکانیزم‌هایی (مانند تولید داده از خود مدل یا تقطیر دانش) می‌توانند جایگزین داده‌های منبع شوند تا توزیع ویژگی‌های مدل را به دامنه هدف نزدیک‌تر کنند؟

4. چالش‌های خاص تطبیق دامنه بدون منبع در تشخیص اشیاء (که نیازمند محلی‌سازی و رگرسیون جعبه مرزی است) در مقایسه با وظایف ساده‌تر مانند طبقه‌بندی تصویر چیست و چگونه می‌توان بر آن‌ها غلبه کرد؟

1.4 ساختار سمینار

ساختار این سمینار در چهار فصل به شرح زیر سازماندهی شده است:

فصل اول به معرفی مبانی نظری و ضرورت پژوهش در حوزه تطبیق دامنه اختصاص دارد. در این فصل، چالش «شکاف دامنه» به ویژه در کاربرد تشخیص اشیاء مورد بررسی قرار گرفته و اهمیت استفاده از این تکنیک‌ها، به خصوص در سناریوی انتقال از داده‌های مصنوعی به دنیای واقعی، تشریح می‌شود.

فصل دوم مروری جامع بر ادبیات موضوع و روش‌شناسی‌های موجود ارائه می‌دهد. در این بخش، سناریوهای مختلف تطبیق دامنه (مانند بدون نظارت، نیمه نظارت و بدون منبع (Source-Free))تعریف شده و رویکردهای اصلی مورد استفاده در هر یک، از جمله روش‌های مبتنی بر یادگیری تخاصمی، بازسازی تصویر و خود-آموزی(Self-Training) ، به تفصیل شرح داده می‌شوند.

فصل سوم به تحلیل نتایج تجربی و بررسی عملکرد روش‌های کلیدی در این حوزه می‌پردازد. در این راستا، مجموعه داده‌های استاندارد مورد استفاده برای ارزیابی) مانند ,Cityscapes, Foggy Cityscapes و (BDD100K معرفی شده و نتایج گزارش‌شده در مقالات برجسته، مورد بحث و مقایسه قرار می‌گیرند.

فصل چهارم به جمع‌بندی و نتیجه‌گیری نهایی اختصاص یافته است. در این فصل، یافته‌های اصلی مرور شده و در نهایت، پیشنهاداتی برای مسیرهای تحقیقاتی آینده در زمینه تطبیق دامنه برای تشخیص اشیاء ارائه می‌گردد.

2 فصل 2: مروری بر تطبیق دامنه و روش های آن

2.1 مقدمه

این فصل به تشریح عمیق مبانی نظری و روش‌شناسی‌های تطبیق دامنه (Domain Adaptation) اختصاص دارد و به این پرسش پاسخ می‌دهد که این تکنیک‌ها چگونه به شبکه‌های عصبی عمیق کمک می‌کنند تا بر چالش "شکاف دامنه[2]"غلبه کرده و کارایی خود را در محیط‌های جدید حفظ نمایند.

در ابتدا، مفاهیم بنیادین تعریف می‌شوند:

- دامنه منبع :(Source Domain) دامنه‌ای که مدل در آن با استفاده از داده‌های برچسب‌دار آموزش دیده است (در این پژوهش، داده‌های مصنوعی).
- دامنه هدف :(Target Domain) دامنه‌ای جدید که مدل باید در آن به خوبی عمل کند، اما داده‌های آن فاقد برچسب هستند (در این پژوهش، تصاویر دنیای واقعی).

سپس، سناریوهای مختلف تطبیق دامنه بر اساس میزان دسترسی به داده‌ها و برچسب‌ها بررسی می‌شوند:

1. تطبیق دامنه نظارت‌شده :(Supervised DA) سناریویی ایده‌آل که در آن به مجموعه داده‌ای کاملاً برچسب‌دار در دامنه هدف دسترسی داریم. این حالت بیشتر به عنوان یک حد بالای تئوریک (Upper Bound) برای عملکرد در نظر گرفته می‌شود.

[2] Domain Gap

2. تطبیق دامنه نیمه-نظارت‌شده :(Semi-Supervised DA) در این حالت، تعداد کمی داده برچسب‌دار به همراه حجم زیادی داده بدون برچسب از دامنه هدف در دسترس است.

3. تطبیق دامنه بدون نظارت :(Unsupervised DA - UDA) رایج‌ترین سناریو در پژوهش‌ها، که در آن به داده‌های برچسب‌دار منبع و داده‌های بدون برچسب هدف به طور همزمان دسترسی وجود دارد.

4. تطبیق دامنه بدون منبع :(Source-Free DA - SFDA) سناریوی محوری این سمینار، که در آن فرآیند تطبیق تنها با دسترسی به مدل آموزش‌دیده و داده‌های بدون برچسب هدف انجام می‌شود و داده‌های اصلی منبع در دسترس نیستند.

در ادامه، رویکردهای اصلی مورد استفاده در این حوزه، به ویژه برای تشخیص اشیاء، به سه دسته اصلی تقسیم و تحلیل می‌شوند:

الف) روش‌های مبتنی بر هم‌ترازی توزیع(Discrepancy-Based Methods): این دسته از روش‌ها با هدف کاهش مستقیم شکاف دامنه از طریق هم‌تراز کردن توزیع آماری ویژگی‌های استخراج‌شده از دو دامنه عمل می‌کنند. تکنیک‌های کلیدی در این زمینه شامل استفاده از شبکه‌های تخاصمی (Adversarial Networks) است که در آن یک تفکیک‌کننده دامنه (Domain Discriminator)تلاش می‌کند ویژگی‌های دو دامنه را از هم تشخیص دهد و استخراج‌کننده ویژگی (Feature Extractor) آموزش می‌بیند تا ویژگی‌هایی تولید کند که برای تفکیک‌کننده غیرقابل تشخیص باشند. (Domain-Invariant Features) در تشخیص اشیاء، این هم‌ترازی هم در سطح کل تصویر[3] و هم در سطح نمونه‌های محلی (Instance-Level) اهمیت دارد.

ب) روش‌های مبتنی بر بازسازی تصویر (Image Reconstruction Methods): این رویکردها تلاش می‌کنند شکاف دامنه را در سطح پیکسل‌ها کاهش دهند. با استفاده از مدل‌های تولیدگر تخاصمی[4]، مانند CycleGAN، تصاویر از یک دامنه به سبک دامنه دیگر ترجمه می‌شوند. برای مثال، تصاویر مصنوعی به گونه‌ای تغییر می‌یابند که گویی در دنیای واقعی گرفته شده‌اند. این کار به مدل کمک می‌کند تا با داده‌هایی که به دامنه هدف نزدیک‌تر هستند، آموزش ببیند. لازم به ذکر است که این روش‌ها معمولاً در سناریوی UDA کاربرد دارند، زیرا نیازمند دسترسی به تصاویر هر دو دامنه هستند.

ج) روش‌های مبتنی بر خود-آموزی (Self-Training Methods): این دسته از روش‌ها، که برای سناریوی Source-Free حیاتی هستند، بر تولید شبه-برچسب[5] برای داده‌های بدون برچسب دامنه هدف متمرکز هستند. فرآیند کلی به این صورت است:

1. مدل آموزش‌دیده روی دامنه منبع، برای پیش‌بینی برچسب روی داده‌های هدف استفاده می‌شود.

2. پیش‌بینی‌هایی که دارای اطمینان[6] بالایی هستند، به عنوان برچسب‌های موقت در نظر گرفته می‌شوند.

---

[3] Image-Level
[4] GANs
[5] Pseudo-Labels
[6] Confidence

3. مدل مجدداً بر روی این داده‌های شبه-برچسب‌دار آموزش می‌بیند تا خود را با توزیع دامنه هدف تطبیق دهد. چالش اصلی در این روش، خطا و نویز در شبه-برچسب‌هاست که می‌تواند منجر به انباشت خطا[7] شود. تکنیک‌هایی مانند فیلتر کردن بر اساس آستانه اطمینان و منظم‌سازی سازگاری[8] برای مقابله با این مشکل توسعه یافته‌اند.

در نهایت، این فصل به بررسی چالش‌های منحصر به فرد تطبیق دامنه در تشخیص اشیاء می‌پردازد و نشان می‌دهد که چگونه روش‌های فوق برای این وظیفه که نیازمند طبقه‌بندی و محلی‌سازی[9] همزمان است، سازگار شده‌اند.

2.2 تطبیق دامنه در یادگیری عمیق

عملکرد چشمگیر شبکه‌های عصبی عمیق (DNNs) که در بسیاری از معیارها به نتایج پیشرفته[10] دست یافته‌اند، منجر به استفاده گسترده از آن‌ها در کاربردهای مختلف هوش مصنوعی شده است. با این حال، یک چالش اساسی، آسیب‌پذیری این شبکه‌ها در برابر پدیده «تغییر دامنه [11]»است. به عبارت دیگر، مدلی که بر روی داده‌های دامنه منبع (Source Domain) آموزش دیده، قادر به تعمیم دانش خود به دامنه هدف[12] با توزیع داده متفاوت نیست و با افت شدید عملکرد مواجه می‌شود.

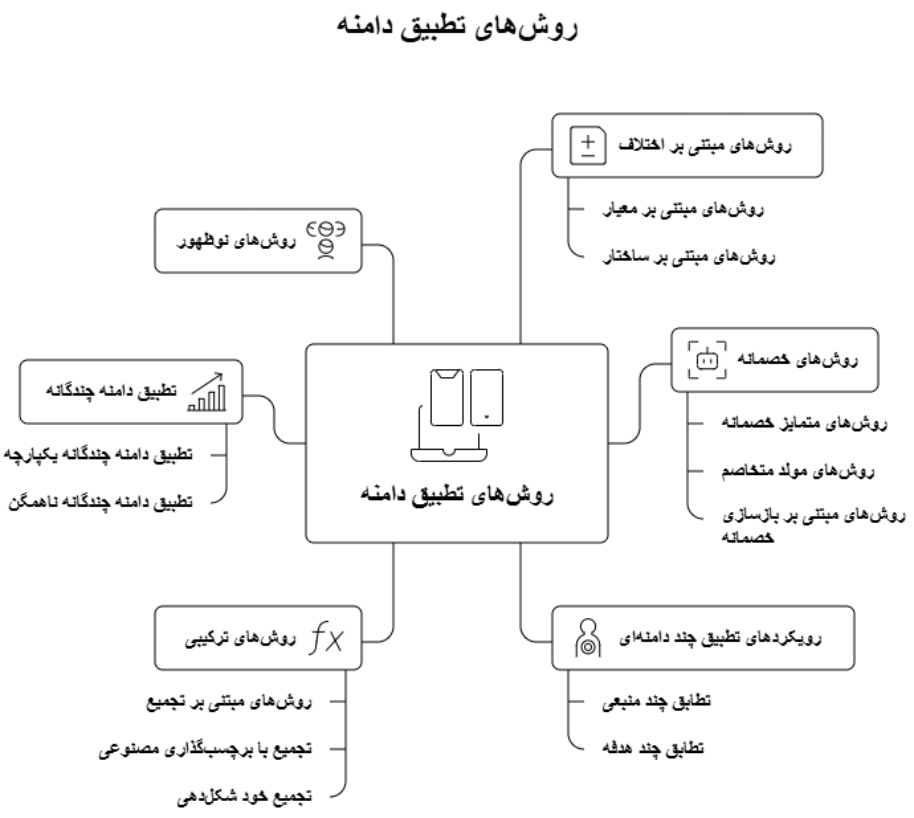
علاوه بر این، ماهیت داده-محور [13] این شبکه‌ها، که نیازمند حجم عظیمی از داده‌های برچسب‌دار برای آموزش هستند، یک مانع بزرگ محسوب می‌شود. فرآیند جمع‌آوری و برچسب‌گذاری داده در مقیاس بزرگ، اغلب دشوار، پرهزینه و گاهی غیرممکن است. این محدودیت‌ها، ضرورت استفاده از رویکردهای یادگیری انتقال [14] و به طور خاص تطبیق دامنه (Domain Adaptation) را بیش از پیش نمایان می‌سازد.

برخلاف رویکردهای کلاسیک، هدف اصلی در تطبیق دامنه عمیق، صرفاً تطبیق سطحی داده‌ها نیست؛ بلکه هدف، گنجاندن مکانیزم‌های تطبیق در خود فرآیند یادگیری عمیق است تا شبکه بتواند بازنمایی‌های ویژگی (Feature Representations) قابل انتقال و نامتغیر از دامنه[15] را بیاموزد.[9]

برای سازماندهی این حوزه وسیع، چندین مقاله مروری، روش‌های تطبیق دامنه عمیق را طبقه‌بندی کرده‌اند. به عنوان مثال، یک طبقه‌بندی شناخته‌شده توسط Wang و Deng ارائه شده است[10]. با این حال، بسیاری از این کارهای مروری، عمدتاً بر روی وظایف حوزه تصویر مانند طبقه‌بندی (Classification) تمرکز دارند. در مقابل، هدف این سمینار، ارائه یک بررسی متمرکز و طبقه‌بندی روش‌هایی است که به طور خاص برای چالش‌های منحصر به فرد تشخیص اشیاء (Object Detection) توسعه یافته‌اند.

---

[7] Error Accumulation
[8] Consistency Regularization
[9] Classification and Localization
[10] State-of-the-Art
[11] Domain Shift
[12] Target Domain
[13] Data-Hungry
[14] Transfer Learning
[15] Domain-Invariant



شکل 1 - درخت واره روش های تطبیق دامنه

2.2.1 روش‌های مبتنی بر اختلاف (Discrepancy-based Methods)

روش‌های مبتنی بر اختلاف، یکی از اصلی‌ترین خانواده‌های الگوریتم‌ها در تطبیق دامنه بدون نظارت (UDA) هستند. ایده اصلی این رویکردها، کاهش مستقیم فاصله یا اختلاف (Discrepancy) بین توزیع ویژگی‌های دامنه منبع و دامنه هدف است. با وادار کردن شبکه به یادگیری بازنمایی‌های نامتغیر از دامنه(Domain-Invariant Representations)، مدل می‌تواند دانش آموخته‌شده از منبع را به طور مؤثرتری به هدف تعمیم دهد. این روش‌ها عمدتاً به دو دسته تقسیم می‌شوند:

1. روش‌های آماری : که از معیارهای آماری صریح برای اندازه‌گیری فاصله بین دو توزیع استفاده می‌کنند. معروف‌ترین معیار در این دسته Maximum Mean Discrepancy (MMD) است که فاصله بین توزیع‌ها را در یک فضای کرنل با ابعاد بالا (RKHS)محاسبه می‌کند.[11]

2. روش‌های تخاصمی : (Adversarial) این رویکرد که محبوبیت بسیار زیادی دارد، از یک بازی دو نفره بین استخراج‌کننده ویژگی (Feature Extractor) و یک تفکیک‌کننده دامنه (Domain Discriminator) استفاده می‌کند. تفکیک‌کننده تلاش

می‌کند تا تشخیص دهد که یک نمونه ویژگی از دامنه منبع آمده است یا هدف، در حالی که استخراج‌کننده ویژگی آموزش می‌بیند تا ویژگی‌هایی تولید کند که تفکیک‌کننده را فریب دهند. این فرآیند منجر به یادگیری ویژگی‌های مشترک و غیرقابل تمایز بین دو دامنه می‌شود.

در کنار این رویکردهای اصلی، ایده‌ی "اندازه‌گیری اختلاف" در سناریوهای دیگری نیز به کار رفته است. به عنوان مثال، در یک پژوهش نوآورانه [4]، از این مفهوم برای هدایت یک استراتژی یادگیری فعال (Active Learning) به منظور تطبیق دامنه استفاده شده است. در این رویکرد، به جای هم‌تراز کردن مستقیم ویژگی‌ها، یک معیار اختلاف بین توزیع داده‌های برچسب‌دار و بدون برچسب تعریف می‌شود. نویسندگان نشان می‌دهند که این معیار به آن‌ها اجازه می‌دهد تا مرزهای تئوریک برای خطای تعمیم (Generalization Error Bounds)را در دامنه هدف استخراج کنند.

سپس، الگوریتمی عملی ارائه می‌شود که با کمینه‌سازی این معیار اختلاف، به صورت هوشمند نقاطی را از دامنه هدف برای برچسب‌زنی انتخاب می‌کند که بیشترین اطلاعات را برای کاهش خطای مدل فراهم می‌کنند. این الگوریتم تحت یک شرط منظم‌سازی بر روی تابع زیان (Loss Function) عمل می‌کند که توسط توابع زیان رایج مانند Squared Loss و Logistic Loss ارضا می‌شود. نتایج تجربی کارایی این رویکرد را در وظایف طبقه‌بندی و رگرسیون نشان داده‌اند.

2.2.1.1 روش‌های اختلاف: مبتنی بر معیارها

یکی از اولین و تأثیرگذارترین رویکردها برای هم‌تراز کردن توزیع ویژگی‌ها در تطبیق دامنه عمیق، استفاده از معیارهای آماری صریح (Explicit Statistical Metrics) است. این روش‌ها با تعریف یک تابع زیان (Loss Function) که مستقیماً فاصله بین توزیع‌های دامنه منبع و هدف را اندازه‌گیری می‌کند، شبکه را به یادگیری بازنمایی‌های نامتغیر از دامنه (Domain-Invariant) وادار می‌کنند.

مقاله [12] (DDC) "Deep Domain Confusion" به عنوان یکی از کارهای پیشگام در این حوزه شناخته می‌شود. ایده اصلی DDC، بهینه‌سازی همزمان دو هدف بود:

1. به حداقل رساندن خطای طبقه‌بندی در دامنه منبع.(Task Loss)
2. به حداقل رساندن اختلاف دامنه بین ویژگی‌های استخراج‌شده از دو دامنه.(Discrepancy Loss)

برای دستیابی به هدف دوم، DDC از معیار Maximum Mean Discrepancy (MMD) به عنوان یک تابع زیان برای اختلاف دامنه استفاده کرد .MMD به طور مؤثری فاصله بین دو توزیع را در یک فضای کرنل با ابعاد بالا (RKHS) اندازه‌گیری می‌کند. تابع زیان کلی در این رویکرد به شکل زیر تعریف می‌شود:

$$L_total = L_task(X_s, Y_s) + \lambda * MMD^2(f(X_s), f(X_t))$$

در این معادله، L_task زیان مربوط به وظیفه اصلی (مانند طبقه‌بندی یا تشخیص اشیاء)، f(X) ویژگی‌های استخراج‌شده توسط شبکه و λ یک ضریب برای تنظیم اهمیت زیان اختلاف دامنه است. در عمل، این عبارت MMD به عنوان یک منظم‌ساز (Regularizer) عمل می‌کند که شبکه را تشویق به یادگیری ویژگی‌هایی می‌کند که اطلاعات مربوط به دامنه را حذف کرده و تنها دانش مرتبط با وظیفه اصلی را حفظ نمایند.

ایده کلیدی DDC سنگ بنای بسیاری از کارهای بعدی شد که با استفاده از معیارهای مشابه یا پیشرفته‌تر به بهبود تطبیق دامنه پرداختند. این معیارها را می‌توان به دسته‌های مختلفی تقسیم کرد. به عنوان مثال، در یک طبقه‌بندی جامع‌تر [10]، این معیارها به سه دسته اصلی تقسیم شده‌اند:

- معیارهای هندسی (Geometrical) : که بر فاصله بین بردارهای ویژگی تمرکز دارند.
- معیارهای مبتنی بر نظریه اطلاعات (Information-Theoretical): که فاصله بین توزیع‌های احتمال را اندازه‌گیری می‌کنند مانند واگرایی KL
- معیارهای آماری مرتبه بالا (Higher-Order Statistics): که از گشتاورهای بالاتر توزیع‌ها (مانند کوواریانس) برای هم‌تراز کردن دقیق‌تر ویژگی‌ها استفاده می‌کنند .[13]

2.2.1.2روش‌های مبتنی بر معماری

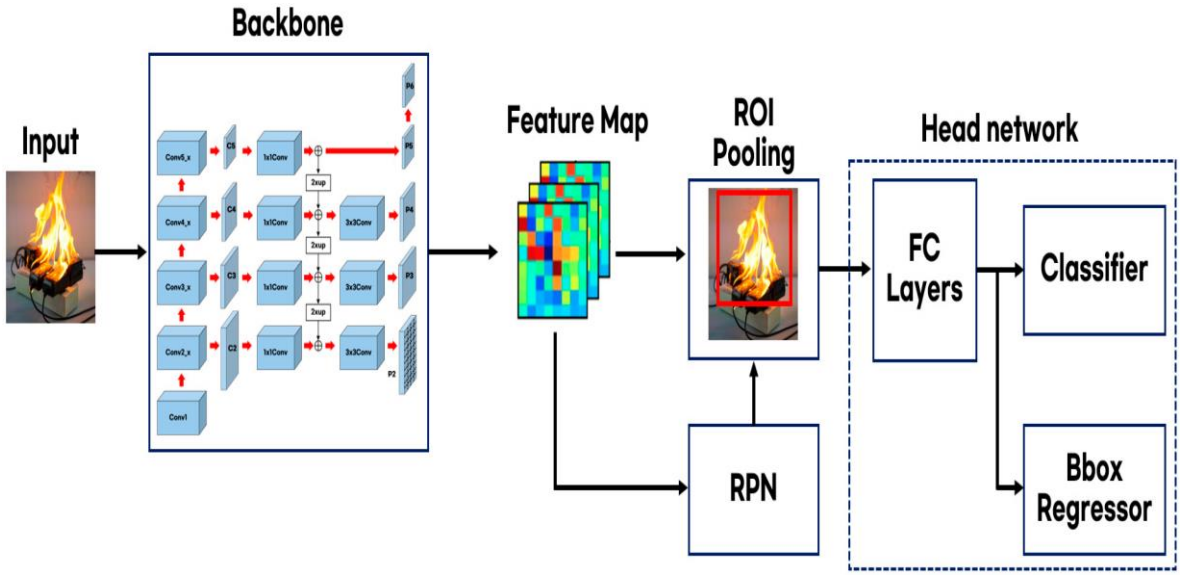
فراتر از روش‌هایی که صرفاً به هم‌تراز کردن کلی توزیع ویژگی‌ها می‌پردازند، دسته‌ای از رویکردها بر ساختار خود معماری شبکه و فضای ویژگی (Feature Space) تمرکز دارند. این روش‌ها به جای تکیه بر یک معیار اختلاف (Discrepancy Metric) واحد، به دنبال یادگیری بازنمایی‌های قابل انتقال از طریق دستکاری ساختار یادگیری یا اعمال محدودیت‌های هوشمندانه‌تر بر روی فضای ویژگی هستند. این خانواده بزرگ را می‌توان به دو زیرشاخه اصلی تقسیم کرد:

2.2.1.2.1 روش‌های مبتنی بر ساختار فضای ویژگی(Feature Space Structuring)

ایده اصلی این رویکردها، تحمیل یک ساختار معنادار و قابل انتقال بر فضای ویژگی است. به جای هم‌تراز کردن کورکورانه توزیع‌ها، این روش‌ها تلاش می‌کنند تا روابط معنایی بین نمونه‌ها را در دامنه‌های مختلف حفظ کنند.

- شبکه‌های تطبیق عمیق (Deep Adaptation Networks - DAN) و جانشینان آن، از اولین تلاش‌ها در این راستا بودند. آن‌ها با اعمال معیار MK-MMD بر روی لایه‌های نهایی شبکه (که ویژگی‌های وظیفه‌محور را یاد می‌گیرند)، تلاش کردند تا توزیع ویژگی‌ها را در سطح کلاس‌ها هم‌تراز کنند و نه فقط به صورت کلی .[14]

- رویکردهای جدیدتر مانند شبکه‌های پروتوتایپی قابل انتقال(Transferable Prototypical Networks - TPN)، این ایده را یک گام فراتر بردند. در این روش‌ها، برای هر کلاس یک پروتوتایپ (Prototype) یا نماینده در فضای ویژگی تعریف می‌شود. سپس فرآیند تطبیق به گونه‌ای انجام می‌شود که فاصله بین پروتوتایپ‌های متناظر از دامنه منبع و هدف به حداقل برسد. این کار تضمین می‌کند که فضای ویژگی به صورت معنایی و کلاس-محور (Class-wise) هم‌تراز شود.[15]



شکل2 -این شکل، ساختار Robust Faster-RCNN را نشان می‌دهد که تطبیق دامنه را از طریق یادگیری با برچسب‌های نویزی انجام می‌دهد.در این روش، ابتدا مدلِ آموزش‌دیده روی دامنه منبع، شبه-برچسب‌های نویزی برای داده‌های هدف تولید می‌کن سپس یک ماژول پالایش، کیفیت این برچسب‌ها را مدل‌سازی کرده و از یادگیری خطاها جلوگیری می‌کند

2.2.1.2.2 روش‌های مبتنی بر خود-آموزی و سازگاری(Self-Training and Consistency-Based)

این دسته از روش‌ها، که برای سناریوی Source-Free بسیار حیاتی هستند، از خود مدل برای هدایت فرآیند تطبیق استفاده می‌کنند. معماری شبکه در یک حلقه یادگیری قرار می‌گیرد تا به تدریج خود را با دامنه هدف وفق دهد.

- یادگیری با برچسب‌های نویزی: یک رویکرد قدرتمند، فرموله کردن تطبیق دامنه به عنوان یک مسئله یادگیری با برچسب‌های نویزی (Learning with Noisy Labels) است. در این سناریو، ابتدا مدل آموزش‌دیده روی دامنه منبع، برای تولید شبه-برچسب (Pseudo-Labels) برای داده‌های هدف استفاده می‌شود. از آنجایی که این برچسب‌ها ناگزیر دارای خطا هستند، تکنیک‌های پیشرفته‌ای برای آموزش یک مدل ثانویه به کار گرفته می‌شود که بتواند در حضور این نویز، به صورت مقاوم آموزش ببیند.[16]

- منظم‌سازی سازگاری :(Consistency Regularization) این روش‌ها، به ویژه مدل Mean Teacher، بر این اصل استوارند که یک مدل باید در برابر اغتشاشات (Perturbations) جزئی در ورودی، پیش‌بینی‌های سازگاری داشته باشد. در زمینه تطبیق دامنه، دو نسخه از مدل (دانش‌آموز و معلم) به کار گرفته می‌شوند. مدل دانش‌آموز بر روی داده‌های هدف آموزش می‌بیند و مدل معلم (که میانگین متحرک وزن‌های مدل دانش‌آموز است)، شبه-برچسب‌های پایدارتری تولید می‌کند. سپس یک زیان سازگاری (Consistency Loss) تعریف می‌شود تا خروجی دانش‌آموز به خروجی معلم نزدیک شود. این ایده برای تشخیص اشیاء با در نظر گرفتن روابط بین اشیاء (Object Relations) نیز توسعه یافته است.[17]

- بهره‌گیری از دانش چندوجهی :(Multi-Modal Knowledge) در کاربردهای خاصی مانند تشخیص عابر پیاده با استفاده از داده‌های حرارتی و بصری، می‌توان از اطلاعات تکمیلی یک مُدالیته برای بهبود برچسب‌زنی در مُدالیته دیگر به صورت تکراری استفاده کرد و سپس یک مدل تشخیص چندوجهی را با این برچسب‌های پالایش‌شده آموزش داد.

- تولید داده در سناریوی :Source-Free جدیدترین رویکردها در تطبیق دامنه بدون منبع، تلاش می‌کنند تا جای خالی داده‌های منبع را با تولید داده‌های جدید پر کنند. این روش‌ها از مدل آموزش‌دیده برای تولید نمونه‌های جدید که شبیه به داده‌های منبع هستند، استفاده می‌کنند و سپس از این داده‌های تولیدشده برای حفظ دانش مدل و جلوگیری از فراموشی فاجعه‌بار (Catastrophic Forgetting) در حین تطبیق با دامنه هدف بهره می‌برند. این یک حوزه بسیار فعال و پیشرفته در سال‌های اخیر است[16].

2.2.2 روشهای خصمانه

روش‌های مبتنی بر یادگیری تخاصمی (Adversarial Learning) با الهام از موفقیت چشمگیر شبکه‌های تخاصمی مولد (GANs) انقلابی در حوزه تطبیق دامنه ایجاد کردند.[18] ایده اصلی این رویکردها، ایجاد یک بازی مینی‌مکس (Minimax Game)بین دو مؤلفه شبکه است تا مدل به صورت خودکار وادار به یادگیری ویژگی‌های نامتغیر از دامنه-(Domain Invariant)شود. این روش‌ها به دو دسته اصلی تقسیم می‌شوند:

1. *روش‌های تخاصمی تمایزی(Discriminative Adversarial Methods)*

این رویکرد، که توسط مقالاتی مانند DANN و ADDA به شهرت رسید[19]، بر روی هم‌تراز کردن ویژگی‌ها در فضای ویژگی (Feature Space)تمرکز دارد. معماری آن شامل دو بخش اصلی است:

استخراج‌کننده ویژگی :(Feature Extractor) که تلاش می‌کند بازنمایی‌هایی از داده‌های هر دو دامنه (منبع و هدف) تولید کند.

تفکیک‌کننده دامنه :(Domain Discriminator) که وظیفه دارد تشخیص دهد که یک بردار ویژگی ورودی، متعلق به دامنه منبع است یا دامنه هدف.

در این بازی تخاصمی، تفکیک‌کننده آموزش می‌بیند تا هرچه بهتر دو دامنه را از هم متمایز کند، در حالی که استخراج‌کننده ویژگی آموزش می‌بیند تا ویژگی‌هایی تولید کند که تفکیک‌کننده را فریب دهند و در نتیجه، برای آن غیرقابل تشخیص باشند. این فرآیند منجر به یادگیری بازنمایی‌هایی می‌شود که اطلاعات مختص دامنه را حذف کرده و تنها دانش مرتبط با وظیفه اصلی (مانند تشخیص اشیاء) را حفظ می‌کنند. معمولاً معماری این روش‌ها به صورت دوشاخه (Siamese) است که یک شاخه داده‌های منبع و شاخه دیگر داده‌های هدف را پردازش می‌کند و تابع زیان کلی ترکیبی از زیان وظیفه و زیان تخاصمی است.

2. *روش‌های تخاصمی تولیدی(Generative Adversarial Methods)*

این دسته از روش‌ها، به جای هم‌تراز کردن ویژگی‌ها، تلاش می‌کنند تا شکاف دامنه را در سطح پیکسل‌ها (Pixel-Level) از بین ببرند. در این رویکرد، یک مولد (Generator) آموزش می‌بیند تا تصاویر یک دامنه را به سبک دامنه دیگر ترجمه (Translate) کند.

برای مثال، مولد می‌تواند یک تصویر مصنوعی (از دامنه منبع) را به گونه‌ای بازسازی کند که گویی یک تصویر واقعی (از دامنه هدف) است.

سپس یک تفکیک‌کننده تلاش می‌کند تا تصاویر واقعی هدف را از تصاویر ترجمه‌شده و مصنوعی تشخیص دهد. این فرآیند به مدل اجازه می‌دهد تا با داده‌هایی که به توزیع هدف نزدیک‌تر هستند، آموزش ببیند.

چالش در سناریوی Source-Free: لازم به ذکر است که هر دو رویکرد تخاصمی فوق، در شکل کلاسیک خود، نیازمند دسترسی همزمان به داده‌های هر دو دامنه منبع و هدف هستند. این پیش‌فرض در سناریوی "تطبیق دامنه بدون منبع" (SFDA)، که در آن داده‌های منبع در دسترس نیستند، برقرار نیست. بنابراین، پژوهشگران روش‌های نوآورانه‌ای را برای شبیه‌سازی این فرآیند تخاصمی بدون دسترسی به داده‌های منبع توسعه داده‌اند. به عنوان مثال، برخی کارهای جدید تلاش می‌کنند تا با استفاده از خود مدل آموزش‌دیده، ویژگی‌های دامنه منبع را بازسازی یا تولید کرده و سپس از آن‌ها در یک حلقه یادگیری تخاصمی استفاده کنند. [21] ,[20]

2.2.2.1 مدل های تمایزگر متخاصم

رویکرد شبکه عصبی تخاصمی دامنه[1](DANN) ، به عنوان یکی از کارهای بنیادی و پرکاربردترین روش در تطبیق دامنه شناخته می‌شود. فلسفه اصلیDANN ، ایجاد یک بازی تخاصمی است که در آن مدل به طور همزمان برای دو هدف متضاد آموزش می‌بیند:

1. یادگیری ویژگی‌های تمایزدهنده :(Discriminative) شبکه باید بتواند وظیفه اصلی (مانند طبقه‌بندی) را به خوبی انجام دهد.
2. ایجاد سردرگمی برای دامنه :(Domain Confusion) شبکه باید ویژگی‌هایی تولید کند که تفکیک آن‌ها از یکدیگر برای یک تفکیک‌کننده دامنه (Domain Discriminator) غیرممکن باشد.

نوآوری کلیدیDANN ، معرفی یک لایه به نام »لایه گرادیان معکوس (Gradient Reversal Layer - GRL) «بود. این لایه در مسیر پیش‌رو (Forward Pass) به عنوان یک لایه همانی عمل می‌کند، اما در مسیر پس‌رو(Backward Pass) ، گرادیان‌های دریافتی از تفکیک‌کننده دامنه را معکوس می‌کند. این مکانیزم هوشمندانه به استخراج‌کننده ویژگی اجازه می‌دهد تا با استفاده از بهینه‌سازی استاندارد، پارامترهای خود را در جهتی به‌روزرسانی کند که زیان تفکیک‌کننده دامنه را بیشینه کند و در نتیجه، ویژگی‌های نامتغیر از دامنه را بیاموزد.

- تکامل و بهبودهای روش‌های تخاصمی

ایده اصلی DANN الهام‌بخش کارهای متعددی شد که تلاش کردند این چارچوب را بهبود بخشند:

- ADDA (Adversarial Discriminative Domain Adaptation): برخلاف DANN که پارامترهای استخراج‌کننده ویژگی را بین دو دامنه به اشتراک می‌گذارد، ADDAاز معماری بدون اشتراک‌گذاری وزن استفاده می‌کند و تابع زیان آن مستقیماً از GAN الهام گرفته شده است.
- CDAN (Conditional Domain Adversarial Networks): این رویکرد با شرطی کردن (Conditioning)تفکیک‌کننده دامنه بر روی خروجی طبقه‌بند، تلاش کرد تا ویژگی‌ها را به صورت چندوجهی (Multimodal)و کلاس-محور (Class-wise) هم‌تراز کند[22].
- استفاده از فاصله Wasserstein : برخی پژوهش‌ها به جای تابع زیان تخاصمی استاندارد، از فاصله Wasserstein که در WGAN معرفی شد (برای اندازه‌گیری اختلاف توزیع‌ها استفاده کردند که منجر به آموزش پایدارتر می‌شود.
- تخصصی‌سازی برای تشخیص اشیاء

تطبیق دامنه برای تشخیص اشیاء (Object Detection) چالش‌های منحصر به فردی دارد، زیرا نیازمند هم‌ترازی در سطوح مختلف است. پژوهشگران روش‌های تخاصمی را برای این وظیفه پیچیده تخصصی‌سازی کرده‌اند:

- هم‌ترازی چندسطحی (Multi-level Alignment) : در تشخیص اشیاء، شکاف دامنه می‌تواند در سطوح مختلفی مانند ظاهر کلی تصویر (Global) و ویژگی‌های خاص هر شیء (Local) وجود داشته باشد. رویکردهایی مانند مقاله [19]، از یک استراتژی دوگانه استفاده می‌کنند :هم‌ترازی ضعیف برای ویژگی‌های سراسری تصویر و هم‌ترازی قوی برای ویژگی‌های محلی مربوط به هر شیء.

- چارچوب‌های چند-تخاصمی :(Multi-Adversarial Frameworks) روش‌هایی مانند MAF (Multi-Adversarial Faster-RCNN) [20]، از چندین تفکیک‌کننده دامنه در مراحل مختلف معماری Faster-RCNN(مانند سطح ویژگی‌های پایه و سطح ویژگی‌های مربوط به هر پیشنهاد) استفاده می‌کنند تا هم‌ترازی دقیق‌تری را در سرتاسر شبکه اعمال کنند.

- چالش و پیشرفت‌ها در سناریوی Source-Free

همانطور که پیش‌تر ذکر شد، روش‌های تخاصمی کلاسیک نیازمند دسترسی به داده‌های منبع هستند. در سناریوی Source-Free، پژوهشگران رویکردهای جدیدی را برای شبیه‌سازی این فرآیند توسعه داده‌اند:

- بازسازی ویژگی منبع (Source Feature Regeneration): برخی از جدیدترین روش‌ها، از یک شبکه مولد برای بازسازی توزیع ویژگی‌های دامنه منبع استفاده می‌کنند. سپس، از این ویژگی‌های تولیدشده به همراه ویژگی‌های دامنه هدف، در یک حلقه یادگیری تخاصمی استاندارد برای هم‌تراز کردن توزیع‌ها استفاده می‌شود. این کار به مدل اجازه می‌دهد تا بدون دسترسی مستقیم به داده‌های منبع، از مزایای یادگیری تخاصمی بهره‌مند شود[24] ,[23].

یکی از رویکردهای بنیادین برای تطبیق دامنه در تشخیص اشیاء، مقاله "Domain-Adaptive Faster R- CNN" است.[25] این پژوهش به طور مستقیم به چالش «تغییر دامنه» در معماری Faster R-CNN می‌پردازد و یک چارچوب تخاصمی چندسطحی برای حل آن پیشنهاد می‌دهد. ایده اصلی این است که شکاف دامنه در تشخیص اشیاء در سطوح مختلفی رخ می‌دهد. بنابراین، نویسندگان دو مؤلفه تطبیق دامنه مجزا را معرفی می‌کنند:

هم‌ترازی در سطح تصویر(Image-Level Alignment) :

یک تفکیک‌کننده دامنه (Domain Discriminator) بر روی ویژگی‌های استخراج‌شده از کل تصویر اعمال می‌شود تا ویژگی‌های سراسری و وابسته به استایل (مانند آب و هوا یا زمان روز) را هم‌تراز کند.

هم‌ترازی در سطح نمونه :(Instance-Level Alignment) تفکیک‌کننده دیگری بر روی ویژگی‌های مربوط به هر شیء پیشنهادی (Region Proposal) اعمال می‌شود تا بازنمایی‌های خاص هر شیء را نامتغیر از دامنه کند.

این دو مؤلفه با استفاده از یک استراتژی یادگیری تخاصمی (Adversarial Learning) آموزش می‌بینند تا شبکه را به یادگیری ویژگی‌های مقاوم در برابر تغییرات دامنه وادار کنند. علاوه بر این، یک عبارت منظم‌سازی سازگاری (Consistency Regularization)بین این دو تفکیک‌کننده اضافه می‌شود تا اطمینان حاصل شود که شبکه پیشنهاد منطقه (RPN) نیز به صورت نامتغیر از دامنه عمل می‌کند. این معماری به صورت انتها به انتها-End-to)

(Endقابل آموزش است و به عنوان یکی از چارچوب‌های پایه برای روش‌های تخاصمی در تطبیق دامنه تشخیص اشیاء شناخته می‌شود.

ارتباط با پژوهش‌های جدید:

ایده هم‌ترازی چندسطحی، الهام‌بخش بسیاری از کارهای جدیدتر بوده است. به عنوان مثال، پژوهش‌های اخیر در سناریوی Source-Free نیز تلاش می‌کنند تا با استفاده از تکنیک‌های تولید شبه-برچسب، هم‌ترازی را هم در سطح کلی تصویر و هم در سطح اشیاء به صورت جداگانه انجام دهند تا از دانش چندوجهی مدل به بهترین شکل استفاده کنند.[26]

در گذار به سناریوی چالش‌برانگیزتر تطبیق دامنه بدون منبع[16]، که در آن داده‌های منبع در دسترس نیستند، روش‌های مبتنی بر خود-آموزی (Self-Training) و منظم‌سازی سازگاری (Consistency Regularization) اهمیت ویژه‌ای پیدا می‌کنند. مقاله که از تکنیکی به نام "سازگاری تصویر پوشش‌دار - Masked Image Consistency) " (MICاستفاده می‌کند، یک نمونه برجسته از این رویکرد است.[27]

این روش بر این فرض استوار است که یک مدل قدرتمند باید بتواند محتوای بخش‌های حذف‌شده یک تصویر را از روی اطلاعات زمینه‌ای (Contextual Information) باقی‌مانده، بازسازی یا پیش‌بینی کند. فرآیند کلی به شرح زیر است:

مدل اولیه: ابتدا یک مدل (که در اینجا Student نامیده می‌شود) بر روی دامنه منبع آموزش دیده و سپس داده‌های منبع کنار گذاشته می‌شوند.

وظیفه خود-نظارتی (Self-Supervised Task): در مرحله تطبیق، به طور تصادفی بخش‌هایی از تصاویر دامنه هدف پوشانده (Mask) شده و به مدل Student داده می‌شود. از مدل خواسته می‌شود که پیش‌بینی کاملی برای کل تصویر (شامل بخش‌های پوشیده‌شده) ارائه دهد. این کار، مدل را وادار می‌کند تا به جای تکیه بر ویژگی‌های سطحی و محلی، الگوهای معنایی و زمینه‌ای عمیق‌تری را یاد بگیرد.

تولید برچسب با مدل معلم (Teacher Model) : از آنجایی که برچسب‌های واقعی برای دامنه هدف وجود ندارد، از یک مدل معلم (EMA Teacher) برای ارزیابی و هدایت مدل Student استفاده می‌شود. مدل Teacher که وزن‌های خود را به صورت میانگین متحرک نمایی (Exponential Moving Average) از وزن‌های مدل

[16] Source-Free Domain Adaptation - SFDA

Studentدریافت می‌کند، نسخه‌ای پایدارتر و قابل اعتمادتر از دانش‌آموز است. این مدل، تصاویر پوشیده‌نشده دامنه هدف را مشاهده کرده و شبه-برچسب‌های باکیفیتی تولید می‌کند.

زیان سازگاری: در نهایت، یک زیان سازگاری (Consistency Loss) بین پیش‌بینی مدل Student (بر روی تصویر پوشیده‌شده) و شبه-برچسب‌های تولیدشده توسط مدل Teacher (بر روی تصویر کامل) تعریف می‌شود. این فرآیند به مدل اجازه می‌دهد تا بدون دسترسی به داده‌های منبع یا برچسب‌های هدف، خود را با دامنه جدید تطبیق دهد.

این چارچوب Student-Teacher و استفاده از یک وظیفه خود-نظارتی هوشمندانه، سنگ بنای بسیاری از روش‌های پیشرفته در حوزه SFDA است[28].

2.2.2.2 مدل های مولد متخاصم

روش‌های تخاصمی تولیدی، به جای هم‌تراز کردن ویژگی‌ها در یک فضای نهفته، به دنبال کاهش مستقیم «شکاف دامنه» در سطح پیکسل‌ها (Pixel-Level) هستند. ایده اصلی این است که با استفاده از یک مولد(Generator) ، تصاویر یک دامنه را به سبک دامنه دیگر ترجمه (Translate) کنیم. در یک سناریوی کلاسیک، مولد آموزش می‌بیند تا تصاویر دامنه منبع (مثلاً تصاویر مصنوعی) را به گونه‌ای بازسازی کند که از دید یک تفکیک‌کننده(Discriminator) ، از تصاویر واقعی دامنه هدف غیرقابل تشخیص باشند. سپس مدل تشخیص اشیاء بر روی این تصاویر ترجمه‌شده و واقع‌گرایانه‌تر آموزش می‌بیند. کارهای اولیه مانند CoGAN [29]با استفاده از شبکه‌های موازی و به اشتراک‌گذاری وزن، تلاش کردند تا یک توزیع مشترک بین دو دامنه را یاد بگیرند. رویکردهای دیگری مانند PixelDA و Generate-to-Adapt نیز این ایده ترجمه تصویر را با استراتژی‌های مختلف پیاده‌سازی کردند.

نقطه ضعف اصلی روش‌های تولیدی کلاسیک، نیاز آن‌ها به دسترسی همزمان به تصاویر هر دو دامنه منبع و هدف است. این پیش‌فرض در سناریوی "تطبیق دامنه بدون منبع(SFDA) "، که در آن داده‌های منبع در دسترس نیستند، نقض می‌شود.

برای غلبه بر این چالش، پژوهش‌های اخیر به جای ترجمه تصاویر، بر روی بازسازی یا تولید ویژگی‌های دامنه منبع تمرکز کرده‌اند. در این روش‌ها، یک شبکه مولد آموزش می‌بیند تا تنها با استفاده از دانش ذخیره‌شده در مدل اولیه، توزیع ویژگی‌های دامنه منبع را بازسازی کند. سپس از این ویژگی‌های تولیدشده به عنوان یک «پروکسی» یا جایگزین برای داده‌های منبع واقعی استفاده می‌شود تا فرآیندهای تطبیق کلاسیک (مانند یادگیری تخاصمی یا هم‌ترازی پروتوتایپ‌ها) امکان‌پذیر شود. این تکنیک به مدل کمک می‌کند تا دانش خود را از دامنه منبع فراموش نکند و در عین حال با دامنه هدف سازگار شود.[23] ,[20]

دسته دیگری از روش‌های قدرتمند، بر پایه تکنیک‌های ترکیب نمونه (Sample Mixing) مانند Mixup بنا شده‌اند. این رویکردها به جای هم‌تراز کردن توزیع‌های کلی، با ایجاد نمونه‌های ترکیبی و مجازی، به مدل کمک می‌کنند تا مرزهای تصمیم‌گیری نرم‌تر و مقاوم‌تری را بیاموزد.

مقاله[30] ConfMix یک نمونه برجسته از این رویکرد برای تطبیق دامنه بدون نظارت در تشخیص اشیاء است. ایده اصلی آن، ترکیب هوشمندانه نمونه‌ها بر اساس میزان اطمینان (Confidence) مدل است. در ابتدا، یک مدل تشخیص اشیاء تنها بر روی دامنه منبع آموزش می‌بیند. در مرحله تطبیق، این مدل بر روی داده‌های هدف اجرا می‌شود تا برای هر شیء پیشنهادی، یک امتیاز اطمینان تولید کند. سپس، نمونه‌هایی از دامنه منبع و هدف با یکدیگر ترکیب (Mix) می‌شوند. این ترکیب به صورت آگاهانه انجام می‌شود؛ نمونه‌های هدف با اطمینان بالا، نقش بیشتری در آموزش ایفا می‌کنند، در حالی که نمونه‌های با اطمینان پایین به تدریج فیلتر می‌شوند تا از انتقال دانش منفی و تأثیر نویز جلوگیری شود.این استراتژی به مدل اجازه می‌دهد تا به آرامی و به صورت کنترل‌شده با ویژگی‌های دامنه هدف آشنا شود.

مشابه روش‌های تولیدی، ConfMix نیز در فرم اصلی خود نیازمند دسترسی به داده‌های منبع برای ترکیب با داده‌های هدف است. برای حل این مشکل در سناریوی SFDA، پژوهش‌های جدیدتر ایده‌های مشابهی را پیاده‌سازی کرده‌اند و در غیاب داده‌های منبع، این روش‌ها به ترکیب داده‌های خودِ دامنه هدف روی می‌آورند. به عنوان مثال، می‌توان نمونه‌های «آسان» (با اطمینان بالا) و نمونه‌های «دشوار» (با اطمینان پایین) از دامنه هدف را با یکدیگر ترکیب کرد. این کار مدل را تشویق می‌کند تا از دانش خود در مورد نمونه‌های آسان برای بهبود پیش‌بینی در نمونه‌های دشوار استفاده کند و در نتیجه، به صورت خود-نظارتی (Self-Supervised) خود را با دامنه هدف تطبیق دهد[31].

2.2.2.3 روش‌های مبتنی بر بازسازی تخاصمی

وش‌های تولیدی تخاصمی مبتنی بر بازسازی، یکی از رویکردهای پیشرفته در تطبیق دامنه هستند که بر اساس همخوانی با استراتژی‌های تطبیق ویژگی‌های سطح کم عمق طراحی می‌شوند. این روش‌ها معمولاً از شبکه‌های تولیدکننده-متقابل (GAN) یا رمزگذارهای خودکار (AE) بهره می‌برند تا محتوای یک دامنه را با سبک دامنه دیگر بازسازی کنند. در ادامه، به بررسی مفاهیم کلیدی پشت این روش‌های بازسازی تخاصمی می‌پردازیم.

شبکه جداساز دامنه (DSN) یک چارچوب مبتنی بر رمزگذار خودکار است. رمزگذار خودکار یک مدل یادگیری عمیق است که ویژگی‌های ورودی را به یک نمایش پنهان تبدیل کرده و سپس آن‌ها را بازسازی می‌کند. این چارچوب تلاش می‌کند تا ویژگی‌ها و نمایش‌هایی که منحصر به فرد برای یک دامنه خاص هستند را شناسایی کند، و همچنین ویژگی‌ها و نمایش‌هایی که بین چندین دامنه مشترک هستند را جدا کند. نمایش‌های مشترک و خصوصی به گونه‌ای طراحی شده‌اند که مستقل از یکدیگر باشند و تداخلی نداشته باشند. علاوه بر این، قادر است اطلاعات ویژه هر دامنه را از اطلاعات مشترک جدا کند[32].

این چارچوب از معیار زیان بازسازی(reconstruction loss)، همراه با سایر معیارها، استفاده می‌کند که شبکه رمزگشا را وادار به بازسازی نمونه ورودی اصلی با استفاده از ویژگی‌های موجود در نمایش‌های خصوصی و مشترک دامنه مربوطه می‌کند.

CycleGAN یک مدل مبتنی بر GAN است که برای ترجمه تصاویر از یک دامنه مبدأ به دامنه مقصد (ترجمه تصویر به تصویر) طراحی شده است. هدف اصلی آن این است که تصویر مبدأ را به دامنه مقصد ترجمه کند و سپس تصویر مبدأ را بازسازی کند، به طوری که اختلافات ناشی از ترجمه را به حداقل برساند[33].

برای این کار، CycleGANاز مفهوم سازگاری چرخشی (cycle consistency) استفاده می‌کند. به عبارت دیگر، اگر تصویری از دامنه مبدأ به مقصد ترجمه شود و سپس به مبدأ بازگردانده شود، تصویر نهایی باید به تصویر اصلی شبیه باشد؛ یعنی ترجمه و بازگردانی باید دقیق باشند تا تصاویر متناظر باشند و اطمینان حاصل شود که اطلاعات معنایی در طول فرایند حفظ شده است.

یکی از ویژگی‌های کلیدی CycleGAN این است که برخلاف بسیاری از روش‌های دیگر ترجمه تصویر به تصویر، نیازی به تصاویر جفت (تصاویر متناظر در دو دامنه) ندارد. این بدان معناست که برای آموزش مدل، مجموعه‌ای از تصاویر از هر دو دامنه بدون نیاز به جفت‌های متناظر کافی است، که این ویژگی به عنوان یکی از مزایای اصلی آن مطرح می‌شود.

CyCADA (Cycle-Consistent Adversarial Domain Adaptation)از ایده CycleGAN الهام گرفته و از توابع زیان سازگار چرخشی به عنوان یکی از اجزای کلیدی خود استفاده می‌کند CyCADA .از ترکیبی از اختلافات تصویری، ویژگی‌ها و هدف برای انجام یک وظیفه خاص به عنوان مجموعه‌ای از توابع زیان برای آموزش مدل بهره می‌برد[34].

اولین تابع زیان در CyCADA، زیان سازگار چرخشی است که اطمینان می‌دهد تصاویر بین دو دامنه به درستی ترجمه و بازگردانده شوند و تغییرات ناشی از ترجمه در حداقل میزان این تابع هدف ظاهر شوند. علاوه بر این، سایر توابع زیان در سه سطح ارائه می‌شوند: سطح تصویری (مقایسه تصاویر بین دو دامنه و محاسبه اختلافات تصویری)، سطح ویژگی (مقایسه ویژگی‌های تصاویر و اختلافات ویژگی بین دو دامنه) و سطح وظیفه (برای حل یک وظیفه خاص مانند تشخیص شیء در تصویر در دامنه مقصد). این توابع کمک می‌کنند تا مدل تفاوت‌ها و تطابق‌ها بین دو دامنه را یاد بگیرد. در نهایت، تمام این توابع زیان با هم ترکیب می‌شوند و هدف به حداقل رساندن آن‌ها است.

شبکه‌های بازگردنده در تصویر (DRCN) یک روش در حوزه بینایی ماشین و پردازش تصویر است که برای مسائل پیچیده بازگردانی تصویر مورد استفاده قرار می‌گیرد. این روش با استفاده از اتصالات بازگردنده بین اطلاعات مختلف در تصاویر، ویژگی‌های ساختاری و معنایی عمیق‌تری استخراج می‌کند. هدف اصلی DRCN بهبود کیفیت تصاویر است، مانند افزایش وضوح، کاهش نویز، بهبود تفکیک و سایر ویژگی‌ها. ایده اصلی این است که از شبکه‌های بازگردنده برای استخراج ویژگی‌های تصویر استفاده شود و سپس از این ویژگی‌ها برای بهبود تصویر اصلی بهره گرفته شود.

در ادبیات علمی، روش‌های دیگری وجود دارند که کاملاً با تعریف بازسازی تخاصمی همخوانی ندارند، اما به آن نزدیک هستند؛ مثالی از این روش DiscoGAN است که مانند CycleGAN عمل می‌کند، اما از زیان سازگار چرخشی استفاده نمی‌کند.

Arruda و همکارانش یک روش تشخیص خودرو در دامنه‌های مختلف با استفاده از ترجمه تصویر به تصویر بدون نظارت ارائه داده‌اند. آن‌ها از CycleGAN برای ترجمه تصاویر از دامنه روز به دامنه شب استفاده کردند تا مجموعه داده مصنوعی ایجاد شود. سپس، مدل تشخیص (که خودروها را شناسایی می‌کند) بر روی این مجموعه داده آموزش داده شد و اطلاعات برچسب‌گذاری شده از دامنه منبع به داده‌های مصنوعی منتقل شد. این کار به مدل کمک می‌کند تا خودروها را در شرایط مختلف نورپردازی تشخیص دهد.[16]

2.2.3 تطابق چند دامنه ای

ک دامنه منبع و یک دامنه مقصد وجود دارد و مدل باید از منبع به مقصد تطابق یابد. اما در رویکرد چند دامنه‌ای، دو سناریو اصلی ممکن است پیش بیاید:

- تطابق چند منبعی [17]:در این حالت، تعداد دامنه‌های منبع بیش از یکی است. به عبارت دیگر، مدل باید از چندین منبع مختلف به یک دامنه مقصد مشترک تطابق پیدا کند. این رویکرد زمانی مفید است که داده‌ها از منابع متعدد برای تطبیق در یک حوزه مشترک استفاده شوند.
- تطابق چند هدفه [18]:در این سناریو، تعداد دامنه‌های مقصد بیش از یکی است. این بدان معناست که مدل باید از یک دامنه منبع واحد به چندین دامنه مقصد متفاوت تطابق یابد. این مفهوم در مواقعی کاربرد دارد که یک مدل یکپارچه نیاز به سازگاری با چندین حوزه هدف متفاوت داشته باشد[35].

2.2.3.1 تطابق چند منبعی

ای ساخت مدل‌های تطبیقی دامنه‌ای قوی‌تر، منطقی است که مدل‌ها را روی چندین منبع آموزش دهیم. در مطالعات اولیه (قبل از ظهور یادگیری عمیق)، سان، شی و وو اشاره کرده‌اند که می‌توان یک طبقه‌بند جداگانه را روی هر دامنه منبع فردی و دامنه هدف آموزش داد، سپس طبقه‌بندهای پایه را با هم ترکیب کرد یا همه منابع را به عنوان یک منبع واحد در نظر گرفت و آموزش داد. در زمینه یادگیری عمیق، ژائو و همکاران نیز به استفاده از استراتژی‌های اختلاف، تخاصمی و تطابق ویژگی اشاره کرده‌اند. یکی از استراتژی‌های کلیدی، تولید یک دامنه میانی برای تطابق است. در این رویکرد، دامنه‌ای جدید با استفاده از مولفه‌های دامنه معمولاً مبتنی بر GAN ایجاد می‌شود.

M3DA یک مجموعه داده چند دامنه‌ای به نام DomainNet با شش دامنه ساخته و لحظات توزیع ویژگی‌ها را بین چندین دامنه منبع دارای برچسب و دامنه هدف به صورت پویا تطابق داده است. ژائو و همکاران شبکه تجمع دامنه تخاصمی چند منبعی (MADAN) [36]را معرفی کردند که اساساً از CycleGAN استفاده می‌کند (با تفکیک‌کننده زیر دامنه برای دامنه‌های منبع و تفکیک‌کننده چرخه متقابل برای دامنه‌های منبع-هدف) و یک دامنه تطبیقی پنهان برای همه داده‌های منبع و هدف ایجاد می‌کند. به طور مشابه، روسو، توماسی و کاپوتو از CoGAN برای تطابق هر دامنه منبع با هدف استفاده کردند[37].

در یک مقاله، رویکرد جدیدی برای تشخیص چند منبعی بدون نظارت و تطابق دامنه (DAOD) ارائه شده که به مسئله کاهش دانش در حین ترکیب آن‌ها می‌پردازد. روش پیشنهادی به نام حفظ دانش مرتبط با هدف (TRKP) از دو بخش اصلی تشکیل شده: ماژول تبادل چند منبعی تخاصمی (AMSD) که ویژگی‌های خاص منبع را از ویژگی‌های مرتبط با هدف جدا می‌کند، و طرح معدن دانش کلی مرتبط با هدف (HTRM) که دانش مرتبط با هدف را از ویژگی‌های چند منبع استخراج می‌کند. این رویکرد روی

[17] Multi-Source Domain Adaptation - MSDA
[18] Multi-Target Domain Adaptation - MTDA

چندین مجموعه داده مرجع ارزیابی شده و با روش‌های برتر مقایسه شده، و نتایج نشان می‌دهد که TRKP در عملکرد و کارآیی برتر است.[38]

2.2.3.2 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[39]، Gholami و همکاران[40] 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[41]. همچنین Isobe و همکاران[42] 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 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.[43]

2.2.4 روش های ترکیبی

»در حالی که هر یک از رویکردهای مبتنی بر اختلاف، تخاصمی و بازسازی، راهکارهای مؤثری برای کاهش «شکاف دامنه» ارائه می‌دهند، محدودیت‌های خاص خود را نیز دارند .روش‌های ترکیبی (Hybrid Methods) با هدف غلبه بر این محدودیت‌ها، از طریق ادغام هوشمندانه چندین تکنیک به صورت همزمان، به دنبال ایجاد چارچوب‌های تطبیق دامنه‌ای قوی‌تر و جامع‌تر هستند. در ادامه، دو نمونه برجسته از این پارادایم را بررسی می‌کنیم.

یکی از چالش‌های تطبیق دامنه، انتقال ناگهانی و مستقیم از دامنه منبع به دامنه هدف است که می‌تواند منجر به آموزش ناپایدار شود. مقاله "تطبیق دامنه پیشرونده (Progressive Domain Adaptation) "[44]با الهام از ایده‌های یادگیری برنامه‌درسی (Curriculum Learning)، این مشکل را از طریق یک فرآیند تطبیق تدریجی و مرحله‌ای حل می‌کند.

این رویکرد ترکیبی، از چندین مؤلفه کلیدی بهره می‌برد:

- ایجاد دامنه میانی :به جای پرش مستقیم، یک یا چند دامنه میانی (Intermediate Domain) بین منبع و هدف ایجاد می‌شود تا یک مسیر هموارتر برای انتقال دانش فراهم گردد.
- یادگیری تخاصمی :از یادگیری تخاصمی برای هم‌تراز کردن توزیع ویژگی‌ها بین دامنه‌های متوالی (مثلاً، منبع به میانی و میانی به هدف) استفاده می‌شود.
- تابع زیان وزن‌دار :یک تابع زیان هوشمندانه برای مدیریت عدم تعادل در ویژگی‌های تصویر به کار گرفته می‌شود.

این استراتژی ترکیبی به مدل اجازه می‌دهد تا به صورت تدریجی خود را با دامنه هدف وفق دهد و نتایج نشان داده‌اند که این روش به ویژه در شرایط چالش‌برانگیز مانند تغییرات آب و هوا یا دوربین‌های مختلف، عملکرد بسیار خوبی از خود نشان می‌دهد.

انتقال دانش منفی، ناشی از تمایل مدل به یادگیری ویژگی‌های مختص دامنه منبع(Source-Bias) ، یکی دیگر از مشکلات اساسی است. مقاله "تنوع‌بخشی و هم‌ترازی (Diversify and Match) " [45] یک چارچوب دو مرحله‌ای نوآورانه برای مقابله با این مشکل ارائه می‌دهد:

1. مرحله تنوع‌بخشی دامنه (Domain Diversification - DD): در این مرحله، به جای استفاده مستقیم از دامنه منبع، با اعمال عمدی اغتشاشات و تغییرات سبکی، چندین نسخه متنوع از دامنه منبع تولید می‌شود. این کار مدل را وادار می‌کند تا بر روی ویژگی‌های معنایی و پایدار تمرکز کند و از یادگیری بیش از حد بر روی ویژگی‌های سطحی و مختص دامنه منبع، پرهیز نماید.

2. مرحله یادگیری بازنمایی چند-دامنه‌ای (Multi-domain Representation Learning - MRL): در این مرحله، مدل بر روی داده‌های تنوع‌یافته آموزش می‌بیند و همزمان، یک مکانیزم هم‌ترازی (مانند یادگیری تخاصمی) به کار گرفته می‌شود تا فضای ویژگی این دامنه‌های متنوع را به همراه دامنه هدف، به یک فضای مشترک نگاشت کند.

این رویکرد ترکیبی با موفقیت به هر دو چالش «تمایل به منبع» و «ترجمه ناقص» پاسخ می‌دهد و به نتایج پیشرفته‌ای در مجموعه داده‌های مختلف دست یافته است .هر دو روش فوق در فرم اصلی خود نیازمند دسترسی به داده‌های دامنه منبع هستند. در سناریوی SFDA، پژوهشگران در حال توسعه روش‌های ترکیبی جدیدی هستند که بتوانند بدون داده‌های منبع عمل کنند. به عنوان مثال، برخی از جدیدترین کارها، ایده‌های خود-آموزی (Self-Training) را با تولید داده مجازی (Data Generation) ترکیب می‌کنند. در این روش‌ها، ابتدا از مدل معلم برای تولید شبه-برچسب استفاده می‌شود و سپس با استفاده از یک مولد، ویژگی‌های مجازی منبع بازسازی شده و با داده‌های هدف ترکیب می‌شوند تا یک فرآیند آموزش مقاوم و ترکیبی شکل گیرد [23], [46].

2.2.4.1 روش‌های مبتنی بر تجمیع

روش‌های مبتنی بر تجمیع (Ensemble Methods) با هدف بهبود استحکام (Robustness) و دقت پیش‌بینی‌ها، از طریق ترکیب خروجی چندین مدل عمل می‌کنند. این رویکردها، که در حالت کلاسیک از مکانیزم‌هایی مانند رأی‌گیری (برای طبقه‌بندی) یا میانگین‌گیری (برای رگرسیون) استفاده می‌کنند، با بهره‌گیری از تنوع مدل‌ها، واریانس پیش‌بینی را کاهش می‌دهند. اگرچه این روش‌ها به طور سنتی از نظر محاسباتی پرهزینه هستند، اما ایده‌های بنیادین آن‌ها الهام‌بخش تکنیک‌های بسیار کارآمدی در تطبیق دامنه شده است.

در حوزه تطبیق دامنه، این روش‌ها را می‌توان به دو دسته اصلی تقسیم کرد:

۱ .تجمیع برای تولید شبه-برچسب(Ensembling for Pseudo-Labeling)

در این رویکرد، چندین مدل مختلف که بر روی دامنه منبع آموزش دیده‌اند، به صورت مستقل بر روی داده‌های بدون برچسب دامنه هدف اجرا می‌شوند. سپس، با استفاده از یک مکانیزم اجماع (مانند رأی اکثریت)، شبه-برچسب‌ها برای آن دسته از نمونه‌های هدف تولید می‌شوند که مدل‌ها بر سر پیش‌بینی آن‌ها توافق بالایی دارند. این شبه-برچسب‌های باکیفیت، سپس برای آموزش یک مدل نهایی بر روی دامنه هدف به کار گرفته می‌شوند.

۲ .تجمیع خود-شکل‌دهی(Self-Ensembling)

این پارادایم نوآورانه، به جای تجمیع چندین مدل مجزا، خروجی‌های یک مدل واحد را در طول زمان (ایترشن‌های مختلف آموزش) تجمیع می‌کند. این فرآیند که به آن تجمیع زمانی (Temporal Ensembling) نیز گفته می‌شود، به مدل اجازه می‌دهد تا از دانش گذشته خود برای پایدارسازی فرآیند یادگیری فعلی استفاده کند.

یک نمونه برجسته و بسیار تأثیرگذار از این رویکرد، معماری "معلم-دانش‌آموز (Mean Teacher) "است که توسط Tarvainenو[47] Valpola معرفی و توسط French و همکاران[48] برای تطبیق دامنه به کار گرفته شد. این چارچوب از دو مؤلفه تشکیل شده است:

مدل دانش‌آموز :(Student Model) این مدل اصلی است که به طور فعال با استفاده از گرادیان کاهشی آموزش می‌بیند.

مدل معلم :(Teacher Model) این مدل آموزش نمی‌بیند، بلکه وزن‌های آن به صورت میانگین متحرک نمایی (Exponential Moving Average - EMA)از وزن‌های مدل دانش‌آموز به‌روزرسانی می‌شود.

این مکانیزم هوشمندانه، مدل معلم را به یک تجمیع زمانی از حالت‌های مختلف مدل دانش‌آموز در طول تاریخچه آموزش تبدیل می‌کند. در نتیجه، پیش‌بینی‌های مدل معلم بسیار پایدارتر و قابل اعتمادتر از پیش‌بینی‌های نوسانی مدل دانش‌آموز است. در فرآیند تطبیق، از مدل معلم برای تولید شبه-برچسب‌های نرم و باکیفیت (Probabilistic Outputs) برای داده‌های هدف استفاده می‌شود و مدل دانش‌آموز وادار می‌شود تا پیش‌بینی‌های خود را با این برچسب‌های پایدار، سازگار کند (Consistency Regularization).

معماری Mean Teacher به دلیل عدم نیاز به دسترسی همزمان به داده‌های منبع در مرحله تطبیق، به یکی از سنگ بناهای اصلی روش‌های تطبیق دامنه بدون منبع (SFDA)، به ویژه برای تشخیص اشیاء، تبدیل شده است. بسیاری از روش‌های پیشرفته در این حوزه، چارچوب Student-Teacher را به عنوان پایه اصلی خود به کار گرفته و آن را با تکنیک‌های دیگری مانند تولید داده مجازی یا هم‌ترازی پروتوتایپ‌ها ترکیب می‌کنند تا به نتایج پیشرفته دست یابند[50] ,[49].

2.2.5 تطبیق دامنه چند حالته.

در حالی که اکثر روش‌های تطبیق دامنه بر روی داده‌های تک‌حالته (Uni-Modal) تمرکز دارند، تطبیق دامنه چندحالته(Multi-Modal Domain Adaptation - MMDA)به سناریوی پیچیده‌تری می‌پردازد که در آن، داده‌ها از چندین نوع یا حالت (Modality)مختلف، مانند تصویر و متن، تشکیل شده‌اند.این حوزه چالش‌های منحصر به فردی را معرفی می‌کند که فراتر از شکاف دامنه استاندارد است:

1. فضاهای ویژگی ناهمگن : (Heterogeneous Feature Spaces) هر حالت داده، ساختار، ابعاد و فضای ویژگی کاملاً متفاوتی دارد. برای مثال، ویژگی‌های استخراج‌شده از یک تصویر توسط یکCNN، با ویژگی‌های استخراج‌شده از یک متن توسط یک مدل زبانی، کاملاً متفاوت هستند.

2. شکاف دامنه دوگانه (Dual Domain Shift) : تغییر دامنه به طور همزمان *درون هر حالت* (e.g., from synthetic images to real images) و *بین حالت‌های مختلف* رخ می‌دهد. مدل باید بتواند هر دو نوع شکاف را به طور همزمان مدیریت کند.

3. هم‌ترازی و ادغام (Alignment and Fusion) : چالش کلیدی، یافتن راهی برای هم‌تراز کردن این فضاهای ناهمگن و سپس ادغام (Fusion) مؤثر اطلاعات از حالت‌های مختلف برای رسیدن به یک پیش‌بینی نهایی و یکپارچه است.

- تطبیق دامنه همگن در مقابل ناهمگن

برای درک بهتر، باید بین دو سناریو تمایز قائل شد:

- تطبیق دامنه همگن : (Homogeneous DA) در این حالت، که اکثر کارهای این حوزه را شامل می‌شود، فضای ویژگی دامنه‌های منبع و هدف یکسان است($X_s = X_t$)، اما توزیع داده‌ها متفاوت است.($P(X_s) \neq P(X_t)$)

- تطبیق دامنه ناهمگن : (Heterogeneous DA - HDA) در این سناریوی چالش‌برانگیزتر، نه تنها توزیع داده‌ها، بلکه خود فضای ویژگی نیز بین دامنه‌ها متفاوت است .($X_s \neq X_t$) تطبیق دامنه چندحالته، یک نمونه بارز و مهم از HDA است.

- رویکردهای حل مسئله

برای حل این چالش‌ها، پژوهشگران به دنبال یادگیری یک فضای مشترک (Common Space) هستند که در آن بتوان بازنمایی‌های حالت‌های مختلف را با یکدیگر مقایسه و هم‌تراز کرد.

- مقاله Qi و همکارانش [51]یک چارچوب تطبیق دامنه چندحالته را با استفاده از ماژول‌های توجه (Attention Modules)و ادغام دوخطی (Bilinear Fusion) ارائه می‌دهد. ماژول توجه به مدل کمک می‌کند تا روابط متقابل و وابستگی‌ها را بین حالت‌های مختلف درک کند و سپس ماژول ادغام، این اطلاعات را به صورت مؤثری برای یادگیری بازنمایی‌های نامتغیر از دامنه ترکیب می‌کند.

- رویکردهای مدرن‌تر، از مدل‌های پایه بزرگ (Foundation Models) مانند CLIPکه از ابتدا بر روی حجم عظیمی از داده‌های تصویر-متن آموزش دیده‌اند، بهره می‌برند. این مدل‌ها به طور ذاتی یک فضای مشترک و هم‌تراز شده بین تصویر و متن را فراهم می‌کنند که می‌توان از آن به عنوان یک نقطه شروع قدرتمند برای تطبیق دامنه چندحالته استفاده کرد[53].[52]

2.2.5.1 *تطبیق دامنه چند حالته ناهمگن*

رواج داده‌های چندحالته و ناهمگن در دنیای واقعی، توسعه روش‌های تطبیق دامنه‌ای را که قادر به مدیریت فضاهای ویژگی متفاوت هستند، به یک ضرورت تبدیل کرده است. تطبیق دامنه ناهمگن (HDA) به این سناریوی چالش‌برانگیز می‌پردازد که در آن، نه تنها توزیع داده‌ها، بلکه خود فضاهای ویژگی و ابعاد آن‌ها نیز بین دامنه‌های منبع و هدف متفاوت است.($X_s \neq X_t$)

رویکرد کلی درHDA ، یادگیری یک نگاشت (Mapping) از هر دامنه به یک فضای مشترک (Common Space) است که در آن، بازنمایی‌ها قابل مقایسه و هم‌تراز کردن باشند. این انتقال دانش می‌تواند به صورت‌های مختلفی انجام شود؛ به عنوان مثال، در کار Shu و همکارانش [54]، این انتقال از طریق اشتراک‌گذاری ضعیف (Weakly-Shared) یا قوی-(Strongly Shared)پارامترها بین شبکه‌های استخراج‌کننده ویژگی هر دامنه، مدل‌سازی می‌شود.

یکی از مسائل اصلی، چالش فقدان حالت(Missing Modality Challenge) است. اهمیت و پیچیدگی HDA زمانی به اوج خود می‌رسد که با مسئله فقدان حالت مواجه می‌شویم. این سناریو زمانی رخ می‌دهد که یک یا چند حالت داده که در دامنه منبع موجود هستند، در دامنه هدف وجود نداشته باشند. به عنوان مثال، یک مدل ممکن است بر روی داده‌های تصویر-متن آموزش دیده باشد، اما در زمان آزمون، تنها به داده‌های تصویری دسترسی داشته باشد.

مقاله Ding و همکارانش [55] این مسئله را به طور خاص مورد بررسی قرار داده و روش M^2TL را برای مدیریت این انتقال دوطرفه (هم بین دامنه‌ها و هم بین حالت‌ها) ارائه می‌دهد. این روش با استفاده از یک محدودیت ماتریس کم-رتبه (Low-Rank Constraint)، یک زیرفضای مشترک و فشرده برای بازنمایی حالت‌های مختلف یاد می‌گیرد و همزمان با استفاده از معیار MMD، ویژگی‌های قابل انتقال را در این زیرفضا هم‌تراز می‌کند.

بسیاری از پارادایم‌های اصلی تطبیق دامنه برای این سناریوی ناهمگن نیز سازگار شده‌اند. یکی از روش هاروش‌های تخاصمی است به عنوان مثال، CDAN [22]با شرطی کردن تفکیک‌کننده دامنه بر روی اطلاعات چندوجهی (مانند ترکیب ویژگی‌ها و پیش‌بینی‌های طبقه‌بند)، قادر است توزیع‌های چندحالته را به صورت مؤثری هم‌تراز کند. این قابلیت، CDAN را به یک ابزار قدرتمند برای تطبیق دامنه بین داده‌هایی با ماهیت کاملاً متفاوت (مانند تصویر و صدا) تبدیل می‌کند. یکی دیگر از روش ها، روش‌های مبتنی بر مدل‌های پایه (Foundation Models) است که جدیدترین رویکردها در این حوزه، از قدرت مدل‌های پایه بزرگ مانند CLIP که به طور ذاتی یک فضای مشترک بین تصویر و متن را یاد گرفته‌اند، بهره می‌برند. این مدل‌ها یک نقطه شروع بسیار قوی برای هم‌تراز کردن دامنه‌های ناهمگن فراهم می‌کنند و می‌توانند به طور مؤثری با چالش فقدان حالت نیز مقابله کنند.[57] ,[56]

2.2.6 روش های 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.

2.2.6.1 تطبیق دامنه با داده‌های کم

چالش اصلی در تطبیق دامنه با داده‌های کم[19] این است که تعداد داده‌های هدف کافی نیست تا همزمان به نیازهای تبدیل دامنه و تطبیق نمایش بین دو دامنه پاسخ دهد.

روش Few-Shot Adversarial Domain Adaptation (FADA) توسط موتیان و همکاران [58]معرفی شده و از یادگیری تضادی با تمرکز بر سرعت تطبیق استفاده می‌کند. آن‌ها نمونه‌های منبع و هدف را بر اساس دامنه و برچسب‌های کلاسی به چهار دسته تقسیم کردند و سپس طبقه‌بند را روی این چهار دسته به جای دو دسته معمولی اعمال کردند. شبکه (استخراج‌کننده ویژگی و طبقه‌بند برچسب) را ابتدا فقط با داده‌های منبع مقداردهی اولیه کردند، سپس متمایزکننده دسته دامنه (با فریز کردن استخراج‌کننده ویژگی) را به‌روزرسانی کردند. در نهایت، متمایزکننده دسته دامنه را فریز کردند و استخراج‌کننده و طبقه‌بند برچسب را به‌روزرسانی کردند.

روش Domain-Adaptive Few-Shot Learning (DA-FSL) به حل مسئله پیچیده‌تر یادگیری few-shot می‌پردازد[59]، یعنی داده‌های هدف ممکن است کلاس‌هایی از دامنه‌های مختلف داشته باشند. تمرکز شبکه تضادی دامنه‌ای (DAPN)در DA-FSL بر دستیابی به تطبیق توزیع دامنه جهانی است، که با معرفی توابع ضرر جدید و حفظ تمایز کلاسی انجام می‌شود. این توابع ضرر از طریق مکانیزم تغییر وزن تطبیقی وزن‌دهی می‌شوند. جنبه نوآورانه دیگر استفاده از مکانیزم توجه قبل از تعبیه منبع است.

روش Prototypical Cross-domain Self-Supervised Learning (PCS) توسط یو و همکاران[60] معرفی شده و شامل یادگیری خودارزشیابی (به عنوان چارچوب یادگیری خودی متقابل میان دامنه‌ای تقلبی) و بدون نظارت است. ایده اصلی انتقال دانش از منبع به هدف، یافتن شباهت بین نمونه و نماینده (نمونه نماینده) برای انتقال مستحکم‌تر است.

تشخیص اشیاء در دامنه‌های تطبیقی کم داده (FSDAOD) یک وظیفه چالش‌برانگیز است که تنها تعداد محدودی تصویر هدف برای آموزش در کنار تصاویر منبع برچسب‌گذاری شده کافی 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 یادگیری بیش از حد می‌شود. برای حل این مشکل، نویسندگان یک پارادایم تطبیق نامتقارن به نام AsyFOD ارائه داده‌اند که از نمونه‌های منبع و هدف از دیدگاه‌های مختلف بهره می‌برد.

[19] Few-Shot Domain Adaptation

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[61] 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 عملیات توقف-گرادیان را ارائه داده‌اند تا از جریان گرادیان به دامنه منبع در طول فرآیند تطبیق جلوگیری کنند.

2.2.6.2 *تطبیق دامنه بدون داده*

تطبیق دامنه بدون داده[20] به وضعیتی اشاره دارد که داده‌های واقعی دامنه هدف در زمان آموزش موجود نیستند، و تنها برخی اطلاعات مرتبط با آن (معمولاً ابرداده‌های هدف) در دسترس است. این رویکرد از تطبیق دامنه معمولی متمایز می‌شود، زیرا در تطبیق معمولی هیچ اطلاعاتی در مورد داده‌های هدف وجود ندارد، حتی ابرداده‌ها.

یادگیری بدون داده با استفاده از داده‌های مرتبط با وظیفه: پنگ و همکاران در وظایف بینایی کامپیوتر از اطلاعات موجود در داده‌های مرتبط با وظیفه (جفت‌های دامنه) برای کمک به درک شبکه در مورد دامنه هدف که در زمان آموزش در دسترس نیست، استفاده می‌کنند.[62]

یادگیری بدون داده با برچسب‌های جدید در دامنه هدف: هدف 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 یکسان است. مثالی که در مقاله کودیروف و همکاران ذکر شده، این است که 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 برچسب‌گذاری مؤثر کمک خواهد کرد.[64] ,[63]

[20] Zero-Shot Domain Adaptation

3 فصل 3: مجموعه دادگان و نتایج تجربی

3.1 مقدمه

در این فصل به بررسی دیتاست های استفاده شده در این موضوع، نتایج مقاله ها و پایان نامه های ذکر شده می پردازیم.

3.2 مجموعه دادگان

3.2.1 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 (5000 برچسب با کیفیت بالا، 20000 برچسب بزرگ) از 50 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 (p1080) 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 - 3 چند نمونه از تصاویر مختلف مجموعه داده Cityscapes

3.2.2 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.

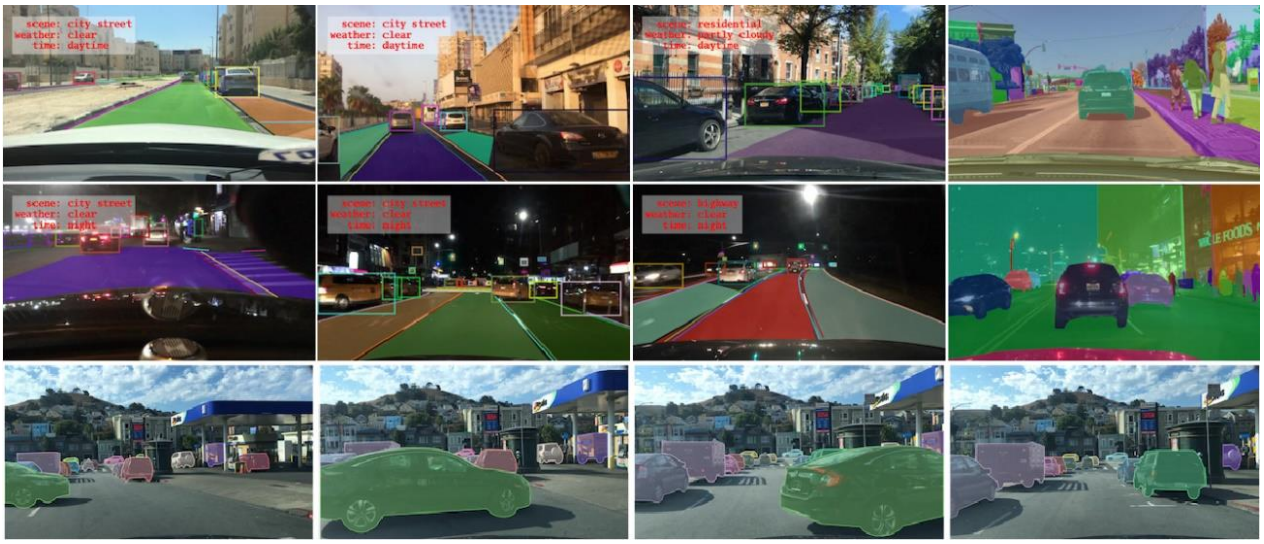
شکل - 4 شکل 20 مجموعه داده Cityscapes (سمت چپ) در مقابل Foggy Cityscapes (سمت راست)

3.2.3 مجموعه دادهCOCO

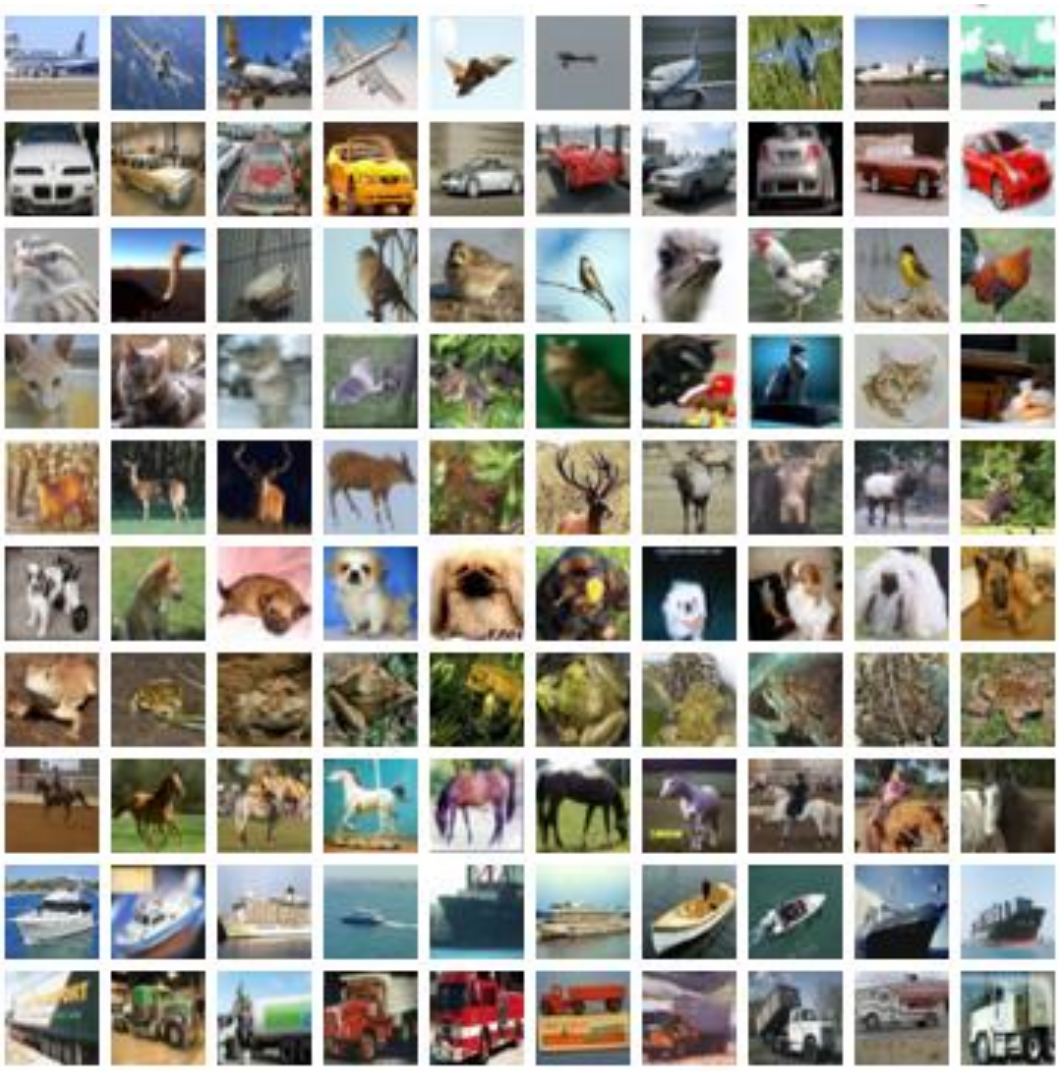
مجموعه داده‌ی COCO یکی از مجموعه‌های داده معروف و گسترده در زمینه تشخیص اشیاء و تفسیر صحنه در تصاویر است. نام "COCO"مخفف "Common Objects in Context" می‌باشد، به این معنا که در این مجموعه داده، اشیاء مختلف در متناوب و با توجه به مفهوم کلی تصویر قرار دارند. این مجموعه داده در سال 2014 توسط محققانی از دانشگاه 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 .

مجموعهCOCO شامل بیش از 330 هزار تصویر با بیش از 1.5 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 80 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 می¬شود. 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 - 5 چند نمونه از تصاویر مختلف مجموعه داده مجموعه داده COCO



شکل - 6 چند نمونه از تصاویر مختلف مجموعه داده COCO

3.2.4 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 "101" به تعداد دسته‌های مختلف اشیاء در این مجموعه داده اشاره دارد.

مجموعه داده Caltech 101 دارای 101 دسته اشیاء مختلف می‌باشد. این دسته‌ها شامل اشیاء مختلفی مانند حیوانات، وسایل نقلیه، اشیاء روزمره و غیره هستند. همچنین شامل تقریباً 9،000 تصویر از اشیاء مختلف در تنوعی از شرایط نورپردازی، زوایا و پس‌زمینه‌ها است.

شکل - 7 چند نمونه از تصاویر مختلف مجموعه دادهCaltech

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.

3.2.5 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 9400 تصویر شهر مجازی در 13 کلاس و شامل تصاویری با نور مصنوعی در شرایط روز و شب است. همچنین دارای تنوع محیطی خوب و میزان

شکل - 8 چند نمونه از تصاویر مختلف مجموعه دادهSYNTHIA

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.

3.2.6 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[21] یک پروژه تحقیقاتی است که در دانشگاه کارلسروهه و دانشگاه مونیخ در آلمان انجام شده است.

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 7,481 تصویر آموزشی و 7,518 تصویر آزمون می‌شود که به مجموع 80,256 شیء با برچسب تعلق دارند. این اشیاء به 8 دسته مختلف تقسیم شده‌اند که عبارتند از: خودرو، ون، کامیون، عابر پیاده، شخص نشسته، دوچرخه‌سوار، تراموا و سایر.

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[22]، تعیین حرکت دوربین[23]، تشخیص اشیاء سه‌بعدی و پیگیری اشیاء سه‌بعدی می‌باشند.

آزمایشات اولیه نشان می‌دهند که روش‌هایی که در معیارهای معتبر مانند Middlebury به نتایج بالایی دست می‌یابند، وقتی به محیط واقعی منتقل می‌شوند، نتایجی متوسط را نشان می‌دهند. هدف اصلی این پژوهش اصلاح این انحراف و تکمیل معیارهای موجود با ارائه معیارهای واقعی با چالش‌های جدید به جامعه تحقیقاتی است.

KITTI شامل چندین زیرمجموعه داده مختلف است، از جمله:

Stereo 2012: برای تخمین عمق از تصاویر استریو.

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.

---

[21] KITTI Vision Benchmark Suite
[22] optical flow
[23] visual odometry

Road Detection: برای تشخیص و تعیین مسیر جاده.

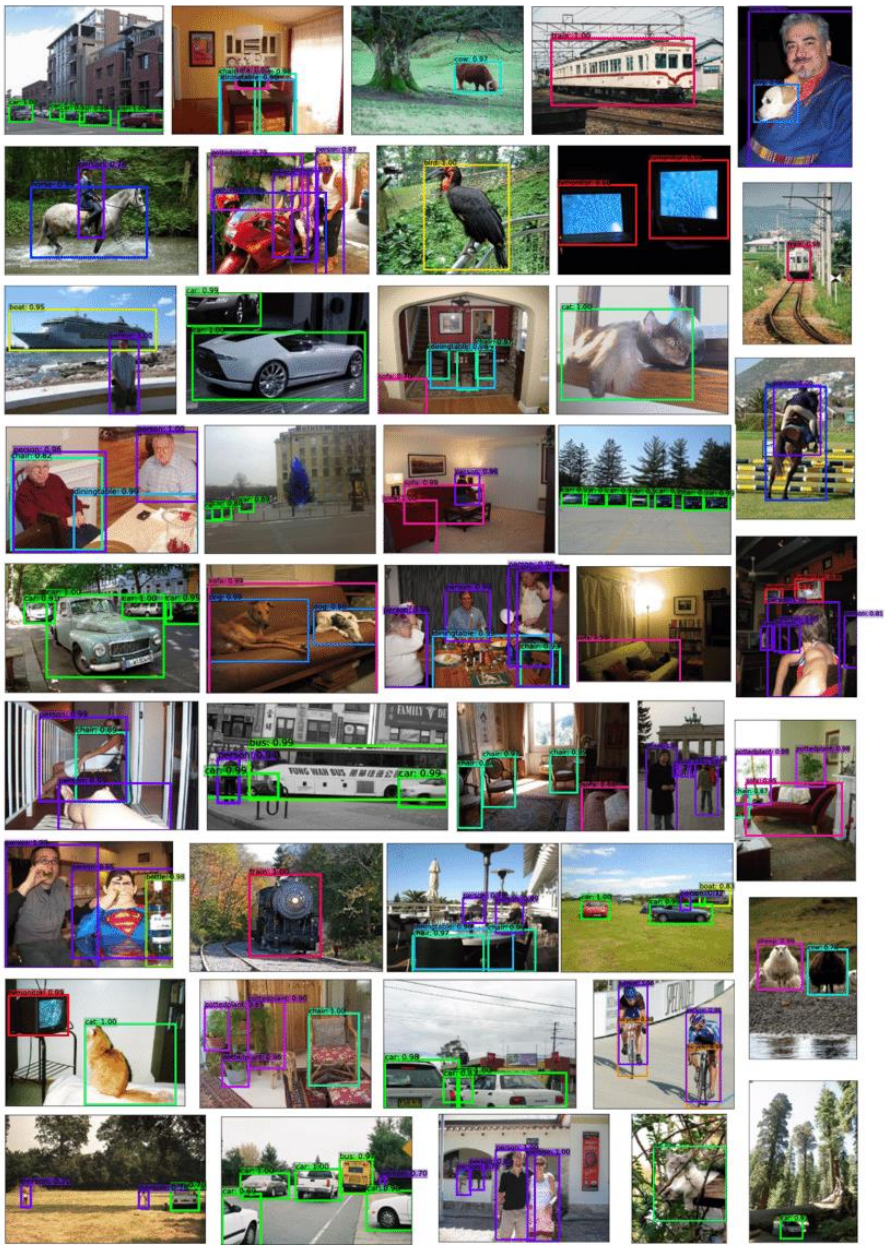
شکل9 - چند نمونه از تصاویر محموعه داده KITTI

3.2.7 مجموعه دادهPascal 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 2005 تا 2012 چندین بار به روزرسانی شده است.

در این مجموعه داده، نسخه‌های محبوب‌تر 2007 و 2012 تصاویر با تقسیم‌بندی آموزش و اعتبارسنجی وجود دارند که در مجموع شامل 15,000 تصویر هستند. تصاویر موجود در این مجموعه داده بیش از 20 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.

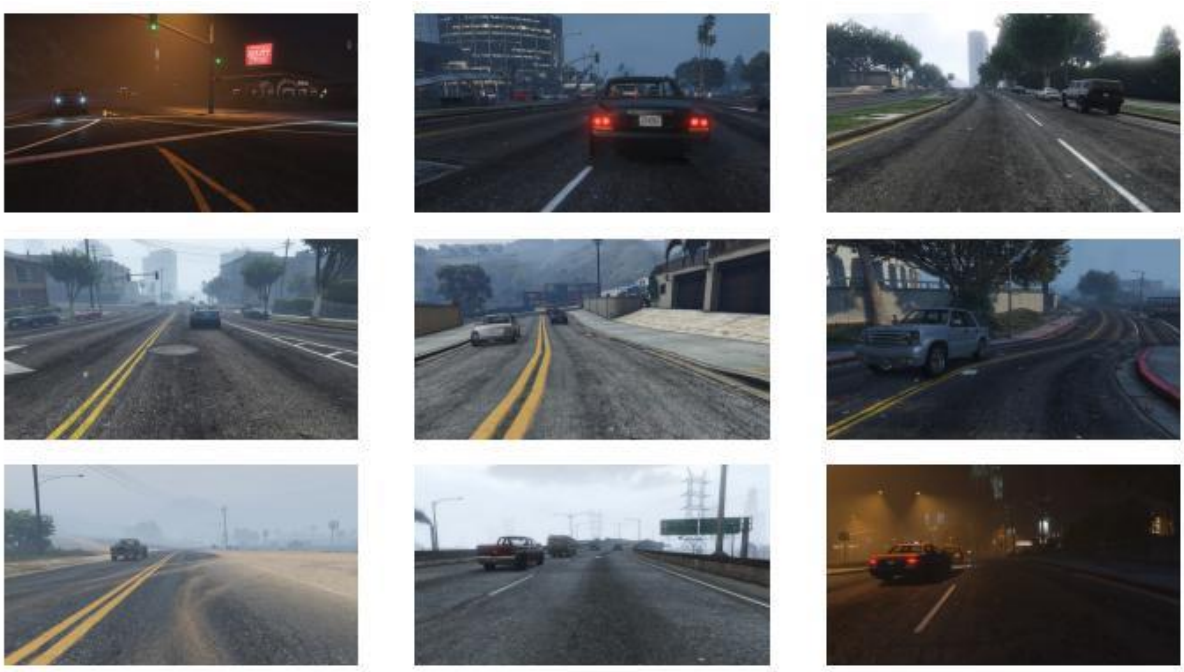


شکل - 10 چند نمونه از تصاویر مختلف مجموعه داده Pascal VOC 2007

3.2.8 مجموعه داده SIM10K

استفاده کردند، که شامل SIM10K 10000 برای تولید مجموعه داده (GTA)از یک موتور شبیه‌سازی SIM10K مجموعه داده .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 را با SIM10K عملکرد شبکه‌های آموزش دیده بر روی مجموعه داده[65] شبکه‌های یادگیری عمیق می‌سازد. نویسندگان ، و یافتند که داده‌های شبیه‌سازی Cityscapes شبکه‌های آموزش دیده بر روی مجموعه داده‌های واقعی مقایسه کردند، مانند 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



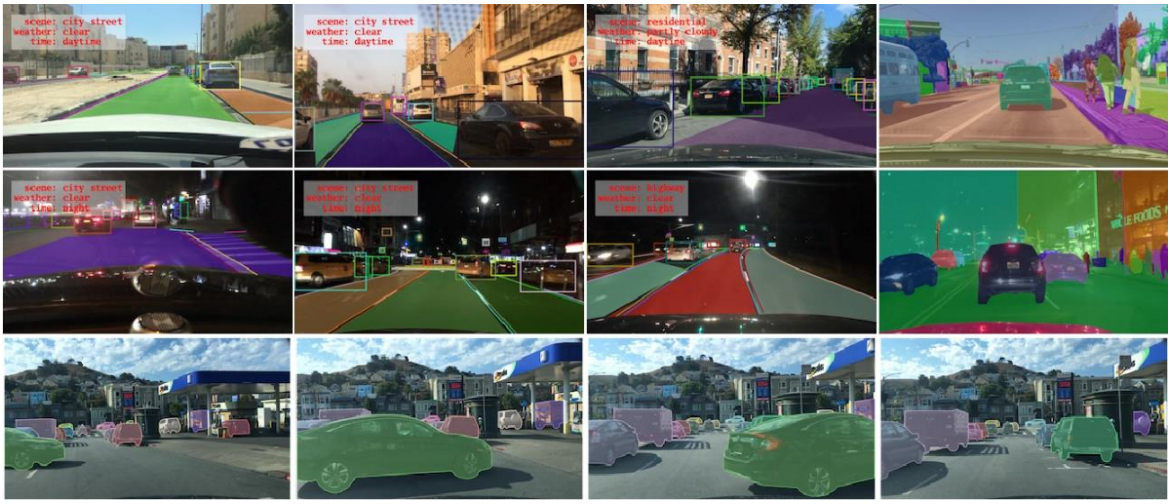
شکل - 11 چند نمونه از تصاویر مختلف مجموعه داده SIM10K

3.2.9 مجموعه داده BDD100K

این مجموعه داده در حال حاضر بزرگترین مجموعه برای هوش مصنوعی خودران[24] به شمار می‌آید. مجموعه مذکور حاوی ساعت تجربه رانندگی در زمان‌های گوناگون روز و شرایط‌های مختلف آب‌وهوایی است. تصاویر 1100 ویدئوی 100000بیش از این مجموعه داده دارای توضیحات و متعلق به نواحی نیویورک و سان فرانسیسکو هستند.

برای تسهیل مطالعات الگوریتمی بر روی داده‌های تصویری بزرگ و متنوع و چندین وظیفه BDD100Kمجموعه داده 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 را ارائه می‌دهد که بزرگ‌تر از مجموعه MOTSو توزیع جغرافیایی متنوعی است. همچنین، این مجموعه داده یک مجموعه داده بزرگ مقایسه YouTube VOS است و تعداد حاشیه‌نویسی‌های آن با مجموعه داده MOTS Challenge و KITTIداده‌های درهای تحقیقات آینده در این حوزه مهم را باز می‌کند و منبع ارزشمندی برای BDD100Kمی‌شود. به طور کلی، مجموعه داده پژوهشگران و توسعه‌دهندگانی است که در برنامه‌های رانندگی خودکار کار می‌کنند.



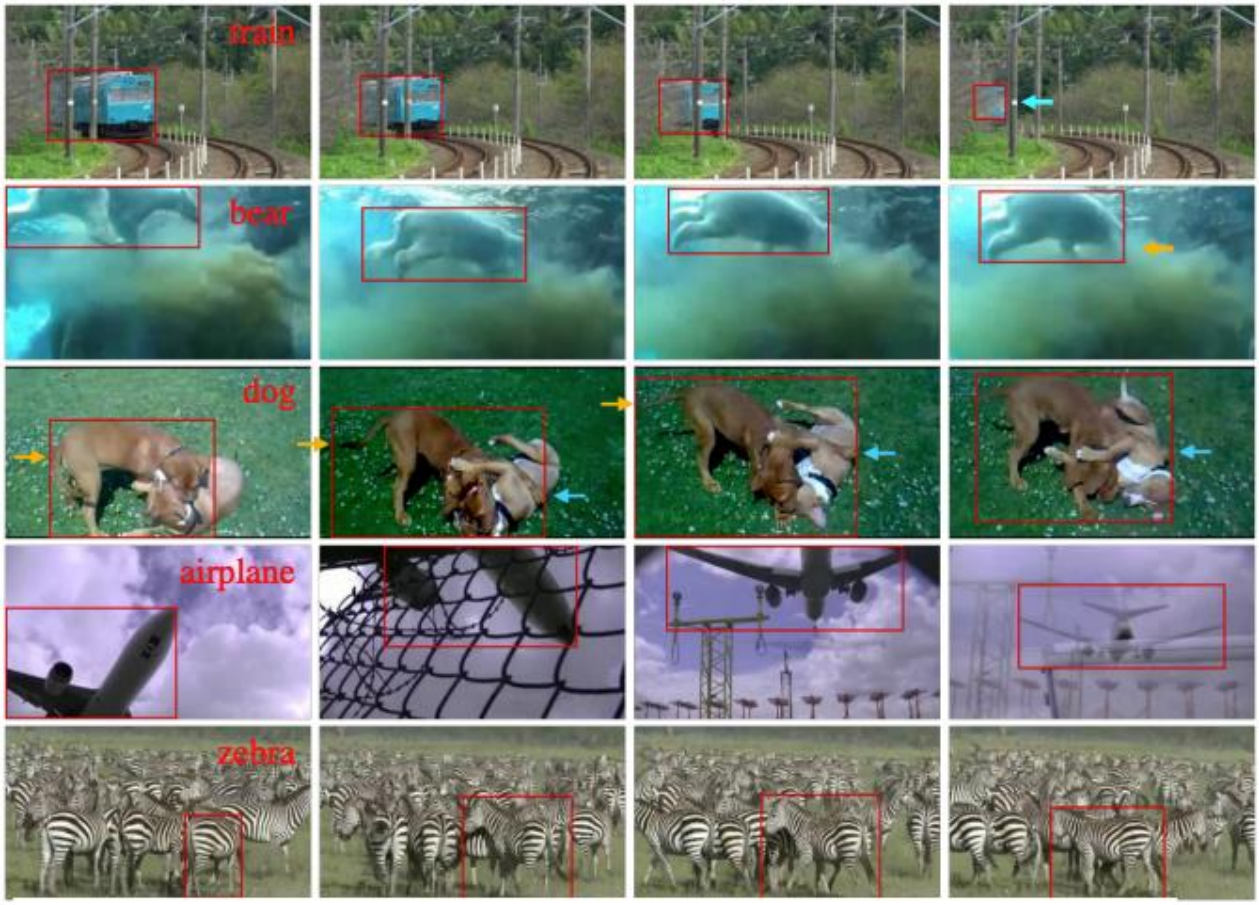
شکل - 12 چند نمونه از تصاویر مختلف مجموعه داده BDD100K

[24] self-driving AI

3.2.10 مجموعه داده YTBB

YTBB[25] مجموعه‌ای از آدرس‌های ویدئو با برچسب‌های جعبه‌های محدودکننده برای اشیاء با فراوانی داده‌های بزرگ می‌باشد. این مجموعه داده حاوی تقریباً 380،000 قطعه ویدئویی به طول حدود 19 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 دستی است. اشیاء در این ویدئوها یک زیرمجموعه از مجموعه برچسب‌های MS 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



شکل - 13 چند نمونه از تصاویر مختلف مجموعه دادهYTBB

[25] YouTube BoundingBoxes

تشخیص در ویدئوها و افزایش امکانات تحلیل ویدئویی نیز کمک می‌کند.

3.3 نتایج

در بخش کارهای پیشین به موارد و روش های بیشتری در زمینه های مختلف تطبیق دامنه ذکر شده است اما در جدول زیر نتایج تجربی روش های تطبیق دامنه برای کاربرد تشخیص اشیا آورده شده است.

شماره	روش	Labeled target	Basic detector	مجموعه داده	دقت(mAP)
1	[10]	Unsupervised	Faster RCNN	SIM10k→ Cityscapes Cityscapes→ Foggy[26] 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	[11]	Unsupervised	Faster RCNN	SIM10kCityscapes → CityscapesFoggy → SyntheticCOCO → SyntheticYTBB→	car AP: 46.6 mAP: 35.1 mAP: 20.7 mAP: 22.8
3	[19]	Unsupervised	Faster RCNN	SIM 10k → Cityscapes Cityscapes → Foggy PASCAL → Clipart PASCAL → Watercolor	car AP:47.7 34.3 38.1 53.3
4	[20]	Unsupervised	Faster RCNN	SIM 10k Cityscapes → CityscapesFoggy → KITTI Cityscapes → Cityscapes KITTI →	car AP: 41.1 34.0 car AP :41.0 car AP:72.1
5	[21]	Unsupervised	Faster RCNN	CityscapesFoggy→ Cityscapes → KITTI	27.6 64.1

[26] Foggy Cityscapes

				KITTI → Cityscapes	38.5
6	[22]	Unsupervised	Faster RCNN	CityscapesFoggy→	47.6
7	[26]	Unsupervised	YOLOv5	Sim10K Cityscapes → KITT Cityscapes → CityscapesFoggy →	56.3 52.2 40.8
8	[31]	Unsupervised	Faster RCNN	BDD100k (day) → night	41.9/67.0
9	[36]	Unsupervised	Faster RCNN	Cityscapes + KITTI BDD100K day + night Cityscapes + MS COCO + Synscapes	58.4 39.8 37.1
10	[40]	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	[41]	Unsupervised	Faster RCNN	PASCAL Clipart1k → PASCAL → Watercolor2k PASCAL → Comic2k Cityscapes Foggy →	41.8 52.0 34.5 34.6
12	[49]	Unsupervised	Faster RCNN Yolov5 X	CityscapesFoggy(Yolov5 → X) Sim10KCityscapes → KITTI Cityscapes → ViPeD COCO (Yolov5 X) →	44.3 64.5 64.1 47.4

- در بخش کارهای پیشین به موارد و روش های بیشتری در زمینه های مختلف تطبیق دامنه ذکر شده است اما در جدول زیر نتایج 1 جدول تجربی روش های تطبیق دامنه برای کاربرد تشخیص اشیا آورده شده است.

4 فصل ۴ نتیجه‌گیری و پیشنهادها

4.1 نتیجه‌گیری

این سمینار، یک بررسی جامع و ساختاریافته از روش‌های تطبیق دامنه (Domain Adaptation) برای وظیفه تشخیص اشیاء، با تمرکز ویژه بر چالش انتقال دانش از داده‌های مصنوعی (Source) به داده‌های واقعی(Target) ، ارائه داد. در این راستا، با مرور نظام‌مند ادبیات موضوع، روش‌های موجود در دسته‌های اصلی شامل رویکردهای مبتنی بر اختلاف، تخاصمی، بازسازی، ترکیبی و چند-دامنه‌ای طبقه‌بندی و تحلیل شدند.

تحلیل مقالات نشان داد که سیر تکاملی این حوزه، یک حرکت مشخص از هم‌ترازی کلی و سراسری (Global Alignment)به سمت رویکردهای ریزدانه، ساختاریافته و خود-محور (Fine-grained, Structured, and Self-driven)بوده است:

- روش‌های اولیه مانند DANN [13] و ADDA [13] بر هم‌تراز کردن کلی توزیع ویژگی‌ها تمرکز داشتند.
- رویکردهای نسل بعد، با درک پیچیدگی تشخیص اشیاء، به هم‌ترازی چندسطحی (مانند تفکیک سطح تصویر و سطح نمونه در[21]) و هم‌ترازی کلاس-محور (مانند استفاده از پروتوتایپ‌ها در[8]) روی آوردند.
- مدرن‌ترین و کارآمدترین روش‌ها، پارادایم را به سمت خود-آموزی (Self-Training) و منظم‌سازی سازگاری (Consistency Regularization)سوق داده‌اند. معماری Teacher-Student [42]و تکنیک‌های نوآورانه مانند MIC [22]و ConfMix [26]، با تولید شبه-برچسب‌های باکیفیت و آموزش مقاوم در برابر نویز، به نتایج پیشرفته‌ای دست یافته‌اند.

مهم‌ترین یافته این بررسی، برجسته شدن اهمیت و ضرورت حرکت به سمت سناریوی واقع‌گرایانه‌تر "تطبیق دامنه بدون منبع (Source-Free Domain Adaptation - SFDA) "است. در این سناریو که به دلیل مسائل حریم خصوصی و حجم داده، دسترسی به داده‌های منبع ممکن نیست، روش‌هایی که قادر به بازسازی دانش منبع یا تولید داده‌های مجازی هستند، آینده این حوزه را شکل می‌دهند. در نهایت، این تحقیق نشان داد که تطبیق دامنه دیگر یک راه‌حل اختیاری نیست، بلکه یک مؤلفه ضروری برای ساخت سیستم‌های تشخیص اشیاء قوی و قابل اعتماد در دنیای واقعی است.

4.2 پیشنهاد کارهای آینده

بر اساس تحلیل انجام‌شده و با نگاهی به آخرین مرزهای دانش در این حوزه، مسیرهای تحقیقاتی زیر برای آینده پیشنهاد می‌گردد:

1. توسعه روش‌های پیشرفته در سناریوی:Source-Free
   - تولید داده و ویژگی :تحقیق بر روی روش‌های مولد کارآمدتر برای بازسازی ویژگی‌های دامنه منبع، بدون نیاز به ذخیره‌سازی اطلاعات اضافی از مدل اولیه.
   - مدیریت فراموشی فاجعه‌بار :طراحی مکانیزم‌های منظم‌سازی جدید برای جلوگیری از فراموشی دانش طبقه‌بندی‌کننده اصلی در حین فرآیند تطبیق با داده‌های هدف.
2. ترکیب با مدل‌های پایه:(Foundation Models)

- استفاده از دانش بصری-زبانی مدل‌های بزرگی مانند CLIP برای هدایت فرآیند تطبیق. این مدل‌ها می‌توانند به تولید شبه-برچسب‌های دقیق‌تر و درک بهتر روابط معنایی بین اشیاء کمک شایانی کنند.

3. تطبیق دامنه چند-منبعی و چند-هدفی بدون منبع:(SF-MSDA & SF-MTDA)

- طراحی الگوریتم‌هایی که قادر به تجمیع دانش از چندین مدل منبع (به جای یکی) و تطبیق آن‌ها با یک یا چند دامنه هدف، بدون دسترسی به داده‌های اولیه باشند. این سناریو بسیار به کاربردهای واقعی نزدیک‌تر است.

4. تطبیق دامنه برای وظایف فراتر از تشخیص اشیاء دوبعدی:

- گسترش روش‌های موجود برای وظایف پیچیده‌تر مانند تشخیص اشیاء سه‌بعدی (3D Object Detection) از داده‌های LiDAR و ردیابی اشیاء (Object Tracking) که در آن‌ها، شکاف دامنه در ابعاد زمانی نیز ظاهر می‌شود.

5. افزایش قابلیت اطمینان و تفسیرپذیری:

- توسعه روش‌هایی برای تخمین عدم قطعیت (Uncertainty Estimation) در شبه-برچسب‌ها و پیش‌بینی‌های نهایی مدل، که برای کاربردهای حساس و حیاتی (مانند خودروهای خودران) ضروری است.
- طراحی مدل‌های تطبیق دامنه‌ای که تفسیرپذیر (Interpretable) باشند و بتوان درک کرد که مدل چگونه و چرا تصمیمات خود را در دامنه هدف اتخاذ می‌کند.

6. تمرکز بر سناریوی واقع‌گرایانه و پیشرفته "تطبیق دامنه بدون منبع:(SFDA) "

از آنجایی که دسترسی به داده‌های منبع (مانند شبیه‌سازها) به دلیل مسائل حریم خصوصی یا حجم بالا همیشه ممکن نیست، این سناریو اهمیت ویژه‌ای دارد. می‌توان روشی نوآورانه برای بازسازی یا تولید ویژگی‌های دامنه منبع از روی مدل از پیش آموزش‌دیده طراحی کرد و از آن برای جلوگیری از فراموشی فاجعه‌بار (Catastrophic Forgetting) در حین تطبیق استفاده نمود.

7. بهره‌گیری از دانش مدل‌های پایه بزرگ:(Foundation Models)

تحقیق بر روی ادغام مدل‌های بصری-زبانی (Vision-Language Models) مانند CLIP در فرآیند تطبیق دامنه برای تشخیص اشیاء. می‌توان از فضای ویژگی مشترک و هم‌تراز شده این مدل‌ها برای تولید شبه-برچسب‌های دقیق‌تر، درک بهتر روابط معنایی بین اشیاء و حتی هدایت فرآیند تطبیق با استفاده از توصیفات متنی (Text Prompts) از دامنه هدف استفاده کرد.

8. طراحی روش‌های ترکیبی پیشرفته برای سناریوی:Source-Free

به جای استفاده از یک تکنیک واحد، می‌توان یک چارچوب ترکیبی (Hybrid Framework) طراحی کرد که به طور هوشمندانه چندین پارادایم را با هم ادغام کند. به عنوان مثال، ترکیب چارچوب Teacher-Student برای تولید شبه-برچسب‌های پایدار، با یک شبکه مولد برای بازسازی ویژگی‌های منبع و یک ماژول توجه برای تمرکز بر روی مناطق مهم‌تر تصویر.

9. توسعه روش‌ها برای سناریوهای پیچیده‌تر چند-دامنه‌ای:

حرکت به سمت سناریوی "تطبیق دامنه چند-منبعی بدون منبع .(SF-MSDA) "در این حالت، به جای یک مدل، چندین مدل از پیش آموزش‌دیده (مثلاً از چندین شبیه‌ساز مختلف) در دسترس است. می‌توان یک الگوریتم تجمیع دانش (Knowledge Aggregation)طراحی کرد که به صورت بهینه، اطلاعات را از این مدل‌های متعدد استخراج و با دامنه هدف تطبیق دهد.

10. افزایش قابلیت اطمینان و تفسیرپذیری مدل تطبیق‌یافته:

طراحی یک ماژول برای تخمین عدم قطعیت (Uncertainty Estimation) در پیش‌بینی‌های مدل. این کار به ما اجازه می‌دهد تا به پیش‌بینی‌هایی که مدل در مورد آن‌ها اطمینان کمی دارد، اعتماد نکنیم و آن‌ها را برای بازبینی بیشتر مشخص کنیم. این قابلیت برای کاربردهای حساس مانند رانندگی خودکار بسیار حیاتی است.

مراجع

[1] Y. Ganin *et al.*, “Domain-Adversarial Training of Neural Networks,” May 26, 2016, *arXiv*: arXiv:1505.07818. doi: 10.48550/arXiv.1505.07818.

[2] R. Wu, S. Yan, Y. Shan, Q. Dang, and G. Sun, “Deep Image: Scaling up Image Recognition,” July 06, 2015, *arXiv*: arXiv:1501.02876. doi: 10.48550/arXiv.1501.02876.

[3] I. J. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and Harnessing Adversarial Examples,” Mar. 20, 2015, *arXiv*: arXiv:1412.6572. doi: 10.48550/arXiv.1412.6572.

[4] Y. Zhu and J. Pei, “Thermal fission rates with temperature dependent fission barriers,” *Phys. Rev. C*, vol. 94, no. 2, p. 024329, Aug. 2016, doi: 10.1103/PhysRevC.94.024329.

[5] T.-Y. Lin *et al.*, “Microsoft COCO: Common Objects in Context,” Feb. 21, 2015, *arXiv*: arXiv:1405.0312. doi: 10.48550/arXiv.1405.0312.

[6] J. N. Kundu, N. Venkat, R. M. V, and R. V. Babu, “Universal Source-Free Domain Adaptation,” Apr. 09, 2020, *arXiv*: arXiv:2004.04393. doi: 10.48550/arXiv.2004.04393.

[7] J. Deng, W. Li, Y. Chen, and L. Duan, “Unbiased Mean Teacher for Cross-domain Object Detection,” June 23, 2021, *arXiv*: arXiv:2003.00707. doi: 10.48550/arXiv.2003.00707.

[8] Y. Liu, W. Zhang, and J. Wang, “Source-Free Domain Adaptation for Semantic Segmentation,” Mar. 30, 2021, *arXiv*: arXiv:2103.16372. doi: 10.48550/arXiv.2103.16372.

[9] Y. Ganin *et al.*, “Domain-Adversarial Training of Neural Networks,” May 26, 2016, *arXiv*: arXiv:1505.07818. doi: 10.48550/arXiv.1505.07818.

[10] 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.

[11] “Domain Adaptation via Transfer Component Analysis | IEEE Journals & Magazine | IEEE Xplore.” Accessed: Oct. 08, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/5640675

[12] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, “Deep Domain Confusion: Maximizing for Domain Invariance,” Dec. 10, 2014, *arXiv*: arXiv:1412.3474. doi: 10.48550/arXiv.1412.3474.

[13] C. Chen, Z. Zheng, X. Ding, Y. Huang, and Q. Dou, “Harmonizing Transferability and Discriminability for Adapting Object Detectors,” Mar. 13, 2020, *arXiv*: arXiv:2003.06297. doi: 10.48550/arXiv.2003.06297.

[14] M. Long, Y. Cao, J. Wang, and M. I. Jordan, “Learning Transferable Features with Deep Adaptation Networks,” May 27, 2015, *arXiv*: arXiv:1502.02791. doi: 10.48550/arXiv.1502.02791.

[15] X. Yue *et al.*, “Prototypical Cross-domain Self-supervised Learning for Few-shot Unsupervised Domain Adaptation,” Mar. 31, 2021, *arXiv*: arXiv:2103.16765. doi: 10.48550/arXiv.2103.16765.

[16] “Source-Free Object Detection by Learning to Overlook Domain Style.” Accessed: Oct. 10, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/9878540

[17] Q. Cai, Y. Pan, C.-W. Ngo, X. Tian, L. Duan, and T. Yao, "Exploring Object Relation in Mean Teacher for Cross-Domain Detection," Dec. 25, 2019, *arXiv*: arXiv:1904.11245. doi: 10.48550/arXiv.1904.11245.

[18] I. J. Goodfellow *et al.*, "Generative Adversarial Networks," June 10, 2014, *arXiv*: arXiv:1406.2661. doi: 10.48550/arXiv.1406.2661.

[19] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial Discriminative Domain Adaptation," Feb. 17, 2017, *arXiv*: arXiv:1702.05464. doi: 10.48550/arXiv.1702.05464.

[20] W. Zhang, L. Shen, and C.-S. Foo, "Source-Free Domain Adaptation Guided by Vision and Vision-Language Pre-training," *Int J Comput Vis.*, vol. 133, no. 2, pp. 844–866, Aug. 2024, doi: 10.1007/s11263-024-02215-3.

[21] C. Eze and C. Crick, "A3: Active Adversarial Alignment for Source-Free Domain Adaptation," Oct. 07, 2024, *arXiv*: arXiv:2409.18418. doi: 10.48550/arXiv.2409.18418.

[22] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional Adversarial Domain Adaptation," Dec. 29, 2018, *arXiv*: arXiv:1705.10667. doi: 10.48550/arXiv.1705.10667.

[23] S. Varailhon, M. Aminbeidokhti, M. Pedersoli, and E. Granger, "Source-Free Domain Adaptation for YOLO Object Detection," Sept. 25, 2024, *arXiv*: arXiv:2409.16538. doi: 10.48550/arXiv.2409.16538.

[24] D. Kothandaraman, S. Shekhar, A. Sancheti, M. Ghuhan, T. Shukla, and D. Manocha, "SALAD: Source-free Active Label-Agnostic Domain Adaptation for Classification, Segmentation and Detection," Oct. 22, 2022, *arXiv*: arXiv:2205.12840. doi: 10.48550/arXiv.2205.12840.

[25] Y. Chen, W. Li, C. Sakaridis, D. Dai, and L. V. Gool, "Domain Adaptive Faster R-CNN for Object Detection in the Wild," Mar. 08, 2018, *arXiv*: arXiv:1803.03243. doi: 10.48550/arXiv.1803.03243.

[26] "SepRep-Net: Multi-source Free Domain Adaptation via Model Separation And Reparameterization." Accessed: Oct. 11, 2025. [Online]. Available: https://arxiv.org/html/2402.08249v2

[27] L. Hoyer, D. Dai, H. Wang, and L. V. Gool, "MIC: Masked Image Consistency for Context-Enhanced Domain Adaptation," Mar. 24, 2023, *arXiv*: arXiv:2212.01322. doi: 10.48550/arXiv.2212.01322.

[28] W. Li, K. Fan, and H. Yang, "Teacher–Student Mutual Learning for efficient source-free unsupervised domain adaptation," *Knowl.-Based Syst.*, vol. 261, p. 110204, Feb. 2023, doi: 10.1016/j.knosys.2022.110204.

[29] M.-Y. Liu and O. Tuzel, "Coupled Generative Adversarial Networks," Sept. 20, 2016, *arXiv*: arXiv:1606.07536. doi: 10.48550/arXiv.1606.07536.

[30] G. Mattolin, L. Zanella, E. Ricci, and Y. Wang, "ConfMix: Unsupervised Domain Adaptation for Object Detection via Confidence-based Mixing," in *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Jan. 2023, pp. 423–433. doi: 10.1109/WACV56688.2023.00050.

[31] S.-Y. Lo, P. Oza, S. Chennupati, A. Galindo, and V. M. Patel, "Spatio-Temporal Pixel-Level Contrastive Learning-based Source-Free Domain Adaptation for Video Semantic Segmentation," Mar. 25, 2023, *arXiv*: arXiv:2303.14361. doi: 10.48550/arXiv.2303.14361.

[32] S. Wang, J. Shi, and H. Li, “Adversarial Reconstruction Feedback for Robust Fine-grained Generalization,” July 29, 2025, *arXiv*: arXiv:2507.21742. doi: 10.48550/arXiv.2507.21742.
[33] “[1703.10593] Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks.” Accessed: Oct. 11, 2025. [Online]. Available: https://arxiv.org/abs/1703.10593
[34] J. Hoffman *et al.*, “CyCADA: Cycle-Consistent Adversarial Domain Adaptation,” Dec. 29, 2017, *arXiv*: arXiv:1711.03213. doi: 10.48550/arXiv.1711.03213.
[35] S. Sun, H. Shi, and Y. Wu, “A survey of multi-source domain adaptation,” *Inf. Fusion*, vol. 24, pp. 84–92, July 2015, doi: 10.1016/j.inffus.2014.12.003.
[36] Z. Pei, Z. Cao, M. Long, and J. Wang, “Multi-Adversarial Domain Adaptation,” Sept. 04, 2018, *arXiv*: arXiv:1809.02176. doi: 10.48550/arXiv.1809.02176.
[37] X. Peng, Q. Bai, X. Xia, Z. Huang, K. Saenko, and B. Wang, “Moment Matching for Multi-Source Domain Adaptation,” Aug. 27, 2019, *arXiv*: arXiv:1812.01754. doi: 10.48550/arXiv.1812.01754.
[38] D. Wright and I. Augenstein, “Transformer Based Multi-Source Domain Adaptation,” Sept. 16, 2020, *arXiv*: arXiv:2009.07806. doi: 10.48550/arXiv.2009.07806.
[39] S. Zhao *et al.*, “Multi-source Domain Adaptation for Semantic Segmentation,” Oct. 27, 2019, *arXiv*: arXiv:1910.12181. doi: 10.48550/arXiv.1910.12181.
[40] B. Gholami, P. Sahu, O. Rudovic, K. Bousmalis, and V. Pavlovic, “Unsupervised Multi-Target Domain Adaptation: An Information Theoretic Approach,” Oct. 26, 2018, *arXiv*: arXiv:1810.11547. doi: 10.48550/arXiv.1810.11547.
[41] K. Bousmalis, G. Trigeorgis, N. Silberman, D. Krishnan, and D. Erhan, “Domain Separation Networks,” Aug. 22, 2016, *arXiv*: arXiv:1608.06019. doi: 10.48550/arXiv.1608.06019.
[42] T. Isobe *et al.*, “Multi-Target Domain Adaptation with Collaborative Consistency Learning,” June 07, 2021, *arXiv*: arXiv:2106.03418. doi: 10.48550/arXiv.2106.03418.
[43] 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,” Nov. 19, 2020, *arXiv*: arXiv:2007.07077. doi: 10.48550/arXiv.2007.07077.
[44] H.-K. Hsu *et al.*, “Progressive Domain Adaptation for Object Detection,” Oct. 24, 2019, *arXiv*: arXiv:1910.11319. doi: 10.48550/arXiv.1910.11319.
[45] T. Kim, M. Jeong, S. Kim, S. Choi, and C. Kim, “Diversify and Match: A Domain Adaptive Representation Learning Paradigm for Object Detection,” May 14, 2019, *arXiv*: arXiv:1905.05396. doi: 10.48550/arXiv.1905.05396.
[46] “Adaptive Pseudo Labeling for Source-Free Domain Adaptation in Medical Image Segmentation | Request PDF.” Accessed: Oct. 12, 2025. [Online]. Available: https://www.researchgate.net/publication/360793126_Adaptive_Pseudo_Labeling_for_Source-Free_Domain_Adaptation_in_Medical_Image_Segmentation
[47] A. Tarvainen and H. Valpola, “Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results,” Apr. 16, 2018, *arXiv*: arXiv:1703.01780. doi: 10.48550/arXiv.1703.01780.
[48] G. French, M. Mackiewicz, and M. Fisher, “Self-ensembling for visual domain adaptation,” Sept. 23, 2018, *arXiv*: arXiv:1706.05208. doi: 10.48550/arXiv.1706.05208.

[49] “Simplifying Source-Free Domain Adaptation for Object Detection: Effective Self-Training Strategies and Performance Insights.” Accessed: Oct. 12, 2025. [Online]. Available: https://arxiv.org/html/2407.07586v1

[50] Q. Liu, L. Lin, Z. Shen, and Z. Yang, “Periodically Exchange Teacher-Student for Source-Free Object Detection,” Nov. 23, 2023, *arXiv*: arXiv:2311.13930. doi: 10.48550/arXiv.2311.13930.

[51] “A Unified Framework for Multimodal Domain Adaptation | Proceedings of the 26th ACM international conference on Multimedia.” Accessed: Oct. 12, 2025. [Online]. Available: https://dl.acm.org/doi/10.1145/3240508.3240633

[52] J. Li *et al.*, “CLIP-Powered Domain Generalization and Domain Adaptation: A Comprehensive Survey,” Apr. 19, 2025, *arXiv*: arXiv:2504.14280. doi: 10.48550/arXiv.2504.14280.

[53] C. Simons, D. S. Raychaudhuri, S. M. Ahmed, S. You, K. Karydis, and A. K. Roy-Chowdhury, “SUMMIT: Source-Free Adaptation of Uni-Modal Models to Multi-Modal Targets,” Aug. 23, 2023, *arXiv*: arXiv:2308.11880. doi: 10.48550/arXiv.2308.11880.

[54] “Weakly-Shared Deep Transfer Networks for Heterogeneous-Domain Knowledge Propagation | Proceedings of the 23rd ACM international conference on Multimedia.” Accessed: Oct. 12, 2025. [Online]. Available: https://dl.acm.org/doi/10.1145/2733373.2806216

[55] “Missing Modality Transfer Learning via Latent Low-Rank Constraint | IEEE Journals & Magazine | IEEE Xplore.” Accessed: Oct. 12, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/7172522

[56] L. Li, Y. Zhou, and G. Yang, “Robust Source-Free Domain Adaptation for Fundus Image Segmentation,” Oct. 25, 2023, *arXiv*: arXiv:2310.16665. doi: 10.48550/arXiv.2310.16665.

[57] “MaMMUT: A Simple Architecture for Joint Learning for MultiModal Tasks | OpenReview.” Accessed: Oct. 12, 2025. [Online]. Available: https://openreview.net/forum?id=FqOG4osY7C

[58] S. Motiian, Q. Jones, S. M. Iranmanesh, and G. Doretto, “Few-Shot Adversarial Domain Adaptation,” Nov. 05, 2017, *arXiv*: arXiv:1711.02536. doi: 10.48550/arXiv.1711.02536.

[59] A. Zhao *et al.*, “Domain-Adaptive Few-Shot Learning,” Mar. 19, 2020, *arXiv*: arXiv:2003.08626. doi: 10.48550/arXiv.2003.08626.

[60] X. Yue *et al.*, “Prototypical Cross-domain Self-supervised Learning for Few-shot Unsupervised Domain Adaptation,” Mar. 31, 2021, *arXiv*: arXiv:2103.16765. doi: 10.48550/arXiv.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,” in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 3261–3271. doi: 10.1109/CVPR52729.2023.00318.

[62] “Zero-Shot Domain Adaptation for SAR Target Recognition Based on Cooperative Learning of Domain Alignment and Task Alignment | IEEE Journals & Magazine | IEEE Xplore.” Accessed: Oct. 12, 2025. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/11037615

[63] H. Azuma, Y. Matsui, and A. Maki, “ZoDi: Zero-Shot Domain Adaptation with Diffusion-Based Image Transfer,” Sept. 25, 2024, *arXiv*: arXiv:2403.13652. doi: 10.48550/arXiv.2403.13652.

[64] “Task guided representation learning using compositional models for zero-shot domain adaptation - ScienceDirect.” Accessed: Oct. 12, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0893608023002769

[65] 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?,” Feb. 25, 2017, *arXiv*: arXiv:1610.01983. doi: 10.48550/arXiv.1610.01983.