بسمه تعالی

دانشگاه صنعتی مالک اشتر

معاونت آموزش و تحصیلات تکمیلی

پیشنهاد پایان‌نامه مقطع کارشناسی ارشد

کمیته محترم تحصیلات تکمیلی مجتمع دانشگاهی برق و کامپیوتر

احتراماً، اینجانب **فاطمه صالح نیا** به شماره دانشجویی **۹۷۱۴۱۹۰۴۰** دانشجوی کارشناسی ارشد رشته **مهندسی کامپیوتر** گرایش **هوش مصنوعی و رباتیک** شاغل در - که تحصیلات خود را در رشته فوق از نیمسال **اول** سال تحصیلی **۱۳۹۷** شروع نمودم و تاکنون تعداد **۲۶** واحد از دروس اصلی و اختصاصی را با موفقیت و با معدل کل **۱۹,۶۱** گذرانده‌ام در مرحلهٔ شروع پایان‌نامه خود هستم.

پس از انجام مشاوره‌های لازم و با عنایت به زمینهٔ علمی موردنیاز، آقای **دکتر علی مهجور** را بعنوان استاد راهنمای اول و آقای **دکتر محمدعلی کیوان راد** را به عنوان استاد راهنمای دوم پایان‌نامه انتخاب کردم و با استفاده از نظرات و راهنمائیهای ایشان پس از بررسی و مطالعه دقیق موضوع تحقیق، موضوع پایان‌نامه خود را کتباً و در قالب فرم پیوست تقدیم می‌نمایم. بدیهی است پذیرش قطعی موضوع پایان‌نامه‌ام منوط بر تصویب آن درکمیته می‌باشد.

ضمناً با پیشنهاد استاد راهنمای پایان‌نامه آقای را نیز بعنوان استاد مشاور انتخاب نموده‌ام.

امضاء دانشجو **تاریخ :**

عنوان کامل پایان‌نامه :

تشخیص اشیاء در تصاویر هوایی با داده های اندک

نوع پروژه:

بدون طبقه‌بندی	محرمانه	خیلی محرمانه	سری	بکلی سری

مراتب فوق موردتأیید است:

۱ ـ نام و نام خانوادگی استاد راهنما **تاریخ و امضاء**

۲ ـ نام و نام خانوادگی استاد راهنمای دوم **تاریخ و امضاء**

۳ ـ نام و نام خانوادگی استاد مشاور **تاریخ و امضاء**

۴- نام ونام خانوادگی کارشناس ارشد آموزش و تحصیلات تکمیلی مجتمع **تاریخ و امضاء**

۵- نام و نام خانوادگی مدیریت آموزش و تحصیلات تکمیلی مجتمع **تاریخ و امضاء**

*** کلیهٔ صفحات به صورت تایپ شده تحویل شوند.**

بسمه تعالی

دانشگاه صنعتی مالک اشتر

معاونت آموزش و تحصیلات تکمیلی

پرسشنامه تهیه پیشنهاد اولیه (پروپوزال) جهت پایان‌نامه کارشناسی ارشد

نام و نام خانوادگی دانشجو : فاطمه صالح نیا

شماره دانشجویی : ۹۷۱۴۱۹۰۴۰

رشته و گرایش تحصیلی : مهندسی کامپیوتر – هوش مصنوعی و رباتیکز

عنوان پایان نامه :

به فارسی : تشخیص اشیاء در تصاویر هوایی با داده های اندک

به لاتین : Few Shot Object Detection in Aerial Image

نوع تحقیق :

بنیادی ☐ **توسعه‌ای** ☐ **کاربردی** ☐

۱ ـ اطلاعات مربوط به دانشجو :

۱-۱- تعداد واحدهای گذرانده : ۲۶ تعداد واحدهای باقیمانده : ۶ معدل (تاکنون) : ۱۹,۶۱

۱-۲- آدرس : خراسان جنوبی، شهرستان سرایان، شهر آیسک، بلوار امام رضا- خیابان آزادی- پلاک ۳۸

۱-۲-۱- شماره تلفن ثابت : ۰۵۶۳۲۸۷۳۲۹۶ ۱-۲-۲- شماره تلفن همراه: ۰۹۱۵۷۲۱۴۲۱۳

۱-۳ ـ درصورت بورسیه بودن، نام سازمان بورسیه دهنده :

۲ ـ اطلاعات مربوط به اساتید راهنما و مشاور :

سمت	نام و نام خانوادگی	تخصص یا رشته	رتبه علمی	تعداد پایان‌نامه‌های راهنمائی شده		تعداد پایان‌نامه‌های تحت راهنمائی	
				ارشد	دکتری	ارشد	دکتری
استاد راهنمای اول	دکتر علی مهجور	مهندسی نرم افزار	استادیار		2		1
استاد راهنمای دوم	دکتر محمدعلی کیوان راد	مهندسی نرم افزار	استادیار				
استاد مشاور							

* تذکر : درصورتی‌که اساتید فوق عضو هیأت علمی دانشگاه صنعتی مالک اشتر نمی‌باشند؛ آدرس و شماره تلفن آن‌ها ذیلاً قید شود.

آدرس استاد (اساتید) راهنما :

شماره تلفن استاد (اساتید) راهنما :

آدرس استاد مشاور :

شماره تلفن استاد مشاور :

۳ ـ اطلاعات مربوط به پایان‌نامه :

۳-۱ ـ مسئله و اهداف اصلی تحقیق :

الگوریتم‌های پرکاربرد در یادگیری ماشین، برای عملکرد دقیق به داده‌های زیادی نیاز دارند، به‌عبارتی متکی به یادگیری از طریق داده های بزرگ هستند. اکثر تکنیک‌های فعلی در صورتی به درستی عمل می‌کنند که نمونه‌های آموزشی زیادی برای یادگیری در اختیار داشته باشند. این تکنیک‌ها زمانی‌که نمونه‌های آموزشی کم باشند به سرعت قابل تعمیم نیستند. به همین دلیل، امروزه بر ارائه روش هایی که برای عملکرد مناسب متکی به حجم داده ورودی موجود نبوده و نیاز به تهیه مستقیم داده ندارند، تأکید می‌شود. از این رو، تحقیق حاضر، با هدف ارائه و ارتقاء روش‌هایی که بتوانند عملکرد قابل قبولی را در تشخیص اشیاء در داده‌های اندکی که از تصاویر هوایی تهیه می‌شود، انجام می‌گردد.

۳-۲ ـ تشریح و بیان موضوع :

تشخیص اشیاء از دیدگاه فنی، یک چالش مهم و مفید در زمینه بینایی کامپیوتر است. تشخیص اشیاء با شناسایی اشیاء مختلف موجود در تصاویر یا ویدیوها سرو کار دارد. در رابطه با تشخیص و شناسایی اشیاء مختلف در تصویر در یک محیط کنترل شده، موفقیت های بزرگی حاصل شده‌است[۱].

راه حل‌های موجود در تشخیص اشیاء، در دو دسته مبتنی بر یادگیری عمیق و روش‌های سنتی تشخیص دسته‌بندی می‌شوند. هدف از تشخیص اشیاء تعیین این مورد است که آیا مواردی از اشیاء نظیر انسان، هواپیما و ... در تصویر وجود دارد یا خیر. در صورت وجود آیتم مربوطه در تصویر، مکان و اندازه شی مشخص می‌شود[۲].

یکی از موارد به روز در این حوزه، تشخیص اشیاء در تصاویر هوایی یا تصاویر راه دور است. تشخیص اشیاء، به عنوان یک مسئله اساسی اما چالش برانگیز در زمینه تجزیه و تحلیل تصاویر هوایی و ماهواره‌ای، در طیف وسیعی از برنامه‌ها، نقش مهمی را ایفا می کند و در سال‌های اخیر بسیار مورد توجه قرار گرفته است[۳].

تشخیص اشیاء در تصاویر هوایی با وضوح بالا، یک گام اساسی برای طیف وسیعی از برنامه‌ها مانند برنامه‌های نظامی، برنامه‌ریزی شهری و مدیریت محیط زیست است. راهکارهای ارائه شده در تشخیص شی در تصاویر هوایی، در سال‌های اخیر به دلیل گسترش روش‌های مبتنی بر شبکه‌های عصبی به طرز چشمگیری بهبود یافته است[۴]. یادگیری ماشین به طور معمول در مواردی که حاوی داده‌های زیاد هستند، بسیار موفق عمل کرده‌است، اما زمانی‌که مجموعه داده کوچک باشد، اغلب با مشکل روبرو می‌شود. در مقابل، انسان با استفاده از آنچه در گذشته آموخته، می‌تواند در موارد جدید نیز قادر به یادگیری سریع باشد. به عنوان مثال، کودکی که عمل جمع را یادگرفته است، می‌تواند با درک نحوه انجام ضرب، به سرعت ضرب را از طریق جمع زدن یاد بگیرد (۳*۱ = ۱ + ۱+ ۱). یا به عنوان مثالی دیگر، اگر یک یا دو تصویر از یک فرد غریبه به کودک نشان داده شود، به راحتی قادر است از بین تعداد عکس‌های گوناگون، به شناسایی همان شخص بپردازد. این موارد این سوال را ایجاد می‌کند که چگونه می‌توان برنامه‌هایی ساخت که با توجه به تجربیات قبلی، به طور خودکار عملکرد خود را بهبود دهند. به منظور یادگیری از تعداد محدودی از مثال‌های دارای اطلاعات باناظر، روش یادگیری ماشین جدیدی تحت عنوان few-shot learning یا به طور خلاصه FSL ارائه شده‌است. این روش، با استفاده از دانش قبلی، می‌تواند به سرعت روی وظایف جدید که فقط حاوی چند نمونه اطلاعات باناظر هستند، تعمیم یابد[۵].

یادگیری با تصاویر اندک، با هدف نزدیک شدن هوش مصنوعی به انسان و با توجه به نیاز موجود در صنعت برای یادگیری ارزان، اخیراً توجهات زیادی را به خود جلب کرده و موضوع به روزی تلقی می‌شود. رویکردهای یادگیری ماشین زیادی برای انجام این هدف، ارائه شده‌اند که به عنوان مثال می‌توان به یادگیری متا[۱]، یادگیری تعبیه شده[۲]، و ایجاد مدل‌های مولد[۳]، اشاره کرد[۵].

بنابراین استفاده از روش FSL در حوزه تشخیص اشیاء این امکان را می‌دهد که با حداقل داده، بتوانیم نوع شی موجود در تصویر را تشخیص دهیم. اما چالش‌های این حوزه، باعث دقت‌های پایین در مطالعات انجام شده، می شود که برای غلبه بر تمام این چالش‌ها، هنوز مطالعات زیادی باید صورت بگیرد. در بخش بررسی مقالات، به این چالش‌ها پرداخته می‌شود.

---

[1] . meta-learning
[2] . embedding learning
[3] . generative modeling

۳-۳ ـ ضرورت انجام تحقیق :

در سال‌های اخیر، به دلیل ظهور دستگاه‌های محاسباتی قدرتمند (به عنوان مثال، GPU و سیستم عامل‌های توزیع شده)، مجموعه داده‌های بزرگ (به عنوان مثال، ImageNet با ۱۰۰۰ کلاس (۳۰))، مدل‌ها و الگوریتم‌های پیشرفته (به عنوان مثال، شبکه‌های کانولوشن CNN و LSTM)، هوش مصنوعی سرعت رقابت خود با انسان را افزایش داده و در بسیاری از زمینه‌ها نیز قادر به شکست انسان است. به عنوان نمونه، می‌توان به AlphaGo اشاره کرد که قهرمانی انسان را در بازی قدیمی Go با شکست روبرو می‌کند و یا شبکه ResNet، که در دسته‌بندی، روی مجموعه‌داده ImageNet عملکرد بهتری را نسبت به عملکرد انسانی بدست می‌آورد. همچنین هوش مصنوعی در بیشتر جنبه های زندگی انسان از جمله دستیارهای صوتی، موتورهای جستجو، ماشین‌های خودران و ... وارد شده است[۵]. بنابراین درخواست برای بالا بردن توانایی ابزارهای هوش مصنوعی، برای نزدیکی حداکثری عملکرد این ابزارها، در مسائل بینایی ماشین، به عملکرد انسان نیز افزایش یافته است.

یکی از موارد مطرح در بینایی ماشین، که از پرکاربردترین شاخه‌های هوش مصنوعی نیز هست، و شباهت دستگاه‌های هوش مصنوعی به انسان را بیش از پیش تقویت می‌کند، مسئله تشخیص اشیاء است. انسان می‌تواند آنچه را در گذشته آموخته است، با نمونه‌های جدید ترکیب کرده و به سرعت آن روی کارهای جدید تعمیم دهد. اما در مقابل، برنامه‌های هوش مصنوعی، معمولاً به یادگیری از طریق داده‌های بزرگ متکی هستند. وجود این فاصله بین هوش مصنوعی و یادگیری شبه انسانی، اهمیت زیادی دارد و مطالعات پایه در خصوص مدل‌ها، الگوریتم‌ها و نظریه‌ها را تحت الشعاع قرار می‌دهد. به طور خاص، یادگیری ماشین با این پرسش روبروست که چگونه می‌توان برنامه های رایانه‌ای ساخت که با استفاده از تجربیات، بتوانند به طور خودکار عملکرد خود را بهبود دهند. اینجاست که یادگیری با تصاویر اندک، مدنظر قرار می‌گیرد.

یکی از مسائلی که از توانایی طبیعی انسان خارج است، نظارت از راه دور بر تغییرات محیطی است. برای مسائل در مسئله کنترل ترافیک، تغییرات زیست محیطی، عملیات نجات، شناسایی اشیاء پرنده و ... چشم انسان به طور مسلح و غیرمسلح نمی‌تواند دقیق عمل کند. در این رابطه ابزارهایی برای انجام عملیات تصویربرداری ارائه شده‌اند که در حال حاضر نیز با پیشرفت فوق العاده‌ای در تکنولوژی‌های این حوزه، روبرو هستیم. از این موارد می‌توان به توانایی ماهواره‌ها در تصویربرداری با رزولوشن بالا و گستره دید وسیع، قابلیت تصویربرداری در شرایط آب و هوایی نامساعد یا شرایط اضطراری و همچنین استفاده از پهپادها به عنوان ابزاری با قابلیت کنترل بالا و کم هزینه اشاره کرد. در این زمان، که تصویربرداری به روش‌های اعلام شده تصاویر بیشماری را در اختیار ما قرار می‌دهد، نیاز به بررسی و پردازش این تصاویر بیش از پیش احساس می‌شود. اما با توجه به چالش‌هایی که در مجموعه دیتاست های تصاویر هوایی وجود دارد. اطلاعات خالص دریافتی از این دیتاست‌ها ممکن است در موارد بلادرنگ یا تشخیص اشیاء در زوایای خاص، پاسخگو نباشد. به همین دلیل، باید از روش‌هایی استفاده شود که ابزار بتواند با استفاده از داده‌های اندکی که دریافت می کند، برای شرایط جدید نیز تصمیم گیری کند، که این مورد، ویژگی است که الگوریتم‌های FSL را نسبت به سایر روش‌ها برجسته می‌کند.

تشخیص اشیاء در زمینه‌های گوناگونی نظیر شمارش اشیاء، ماشین‌های خودران، تشخیص هویت، استخراج اشیاء از تصویر یا ویدیو، تشخیص رفتار و ... کاربرد دارد. اما در رابطه با کاربرد آن در حوزه تشخیص اشیاء در تصاویر هوایی می‌توان به مدیریت و کنترل حوادث طبیعی، عملیات نجات، شناسایی فعالیت‌های نامتعارف، کنترل ترافیک، حملات هوایی و ... اشاره کرد.

۳-۴ - سابقه تحقیقات و مطالعات انجام گرفته :

با توجه به مطالعات صورت گرفته ابتدا به بررسی انواع دیتاست‌های موجود برای تصاویر هوایی و چالش‌های موجود در تهیه این دیتاست‌ها پرداخته شده است.

۳-۴ - ۱- تشخیص اشیاء و تصاویر هوایی:

۳-۴ -۱-۱- بررسی دیتاست‌های تصاویر هوایی در بینایی ماشین:

در زمینه تهیه تصاویر هوایی، عمدتاً دو روش، استفاده از پهپادها، استفاده از تصاویر ماهواره‌ای وجود دارد. این موارد ابتدا به صورت تصویر یا فیلم تهیه شده و در خصوص دیتاستهایی که داده‌های اولیه آنها به صورت فیلم‌برداری توسط پهپادها، جمع‌آوری شده است، فریم‌ها به عنوان تصاویر در نظر گرفته شده و پس از حاشیه‌نویسی، در دسته‌های مشخص به عنوان مجموعه‌داده مرتبط با

تصاویر هوایی یا تصاویر راه دور، ارائه شده است. در جدول زیر، رایج‌ترین دیتاست‌های حوزه تصاویر هوایی ارائه شده است که در آن، تعداد دسته و نمونه‌های هر دسته نیز مشخص شده است.

شماره	منبع	نام دیتاست	توصیف	ویژگی تصاویر	کاربرد اصلی
۱	Zamir et al., (2019)[6]	iSAID	دیتاست مقیاس بزرگ و حاشیه نویسی شده دارای ۶۵۵,۴۵۱ نمونه شی در ۱۵ دسته است که هر دسته دارای ۲,۸۰۶ تصویر با وضوح بالاست	اشیا در این دیتاست، در محدوده ۱۰ تا ۱۴۴ پیکسل به عنوان کوچک، از ۱۴۴ تا ۱۰۲۴ به عنوان متوسط و از ۱۰۲۴ به بالا، به عنوان بزرگ در نظر گرفته شده اند. درصد تصاویر کوچک، متوسط و بزرگ در iSAID به ترتیب ۵۲,۰، ۳۳,۷ و ۹,۷ است.	Instance Segmentation
۲	Lam et al., (2018)[7]	xView	شامل بیش از ۱ میلیون شی در ۶۰ کلاس با سطح پوشش بیش از ۱۴۰۰ کیلومتر مربع. شامل ۶۰ کلاس با تنوع طبقاتی بالاست شامل، عابرپیاده، زمین، ساختمان ها، ماشین ها و ...	اندازه اشیاء در xview از سه متر (۱۰ پیکسل در ۰,۳ متر (فاصله نمونه از زمین یا GSD) تا بیش از (۱۰هزار پیکسل در فاصله ۰,۳متر از زمین متفاوت است. بالابودن رزولوشن به دلیل تنوع نمونه برداری از زمین. از مزایای آن، بهبود تشخیص ریزدانه ها (زیرکلاس ها) است مثلاً در کلاس والد وسایل نقلیه، کلاس فرزند بیل مکانیکی وجود دارد.	Instance Segmentation, Object Detection
۳	SpaceNet Challenge,(2020)[8]	Spacenet	یک مجموعه عظیم تصاویر ماهواره ای، شامل بیش از ۴۰۰۰۰ کیلومتر مربع از تصاویر است که شامل ۱۱,۰۸۰,۰۰۰ حاشیه نویسی برای ساختمان هاست. ساختمان ها بیشترین جلوه های بصری این دیتاست هستند. داده های تولید شده از چندین ماهواره گرفته شده بنابراین از نظر ویژگی هایی مانند رزولوشن، دقت موقعیت جغرافیایی و ... متفاوت هستند.آخرین مجموعه از این پیکره، با GSD ۴ متر در هر پیکسل ارائه شده است.	اولین مجموعه از این پیکره مجموعه شامل ۲۵۴۴ کیلومتر مربع از تصاویر با وضوح کامل ۵۰ سانتی متر از ریودوژانیرو، است. در این مجموعه ۳۸۲,۵۳۴ ساختمان برچسب گذاری شده و جاده وجود دارد. ساختمان ها در این مجموعه از لحاظ نوع ساختمان در دسته های مختلف نظیر اداری، اقتصادی، فرودگاه، راه آهن، مسکونی، زندان، مراکز پلیس، مراکز بیمارستانی و ... تفکیک و برچسب گذاری شده اند. این پیکره شامل مربع های ۵۰سانتی متر در هر پیکسلس (GSD) با هشت باندطیفی است. داده ها در فرمت GeoJson و ۱۶ بیتی ارائه می شوند که چند چالش دارد: برنامه های کمی برای نمایش تصاویر ۱۶ بیتی	Object Tracking, Object Detection

				طراحی شده که نیاز به تبدیل دارند(راه حل مشاهده داده خام با QGIS). برچسب های وکتور به ندرت با معماری یادگیری ماشین سازگار هستند. همه جاده ها و ساختمان ها مساوی نیستند و ممکن است تعداد خطوط، سطح جاده و عرض آنها بسیار متفاوت باشد.	
۴	T. Nathan Mundhenk (2016)[9]	COWC	اتومبیل ها بهترین قابلیت شناسایی را در این مجموعه دارند. این مجموعه حاوی تصویر ۳۲۷۱۶ ماشین از ۶ مجموعه مختلف از موقعیت های جغرافیایی شامل تورنتو کانادا، سلوین نیوزلند، پوتسدام، وایینگن آلمان، کلمبوس و یوتا در آمریکا است. این مجموعه حاوی ۵۸۲۴۷ نمونه هدف منفی نیز هست، مواردی که ممکن است با ماشین اشتباه شوند از قبیل قایق ها، یدک کش ها، ساختمان ها و....	تصاویر با رزولوشن ۱۵سانتی متر در هر پیکسل و گستره بالای ۴ کیلومتر مربع ارائه شده، که باعث می شود اندازه اتومبیل بین ۲۴ تا ۴۸ پیکسل باشد. دومجموعه از ۶ مجموعه(کلمبوس و وایینگن)، تصاویر سیاه سفید و بقیه RGB هستند. کامیون های خیلی بزرگ حذف شده و ون ها و وانت ها به عنوان ماشین درنظرگرفته شده اند. هر تصویر در اندازه 2048 در 2048 است که اندازه ۳۰۷ در ۳۰۷ متر، در سطح زمین را تشکیل می دهد.	Object Detection
۵	Cheng et. al (2016)[10]	NWPU VHR-10 Dataset	شامل ۸۰۰ تصویر ماهواره ای در ۱۰ کلاس برای تشخیص اشیاء ارائه شده است. روش تصویربرداری سنجش از راه دور نوری است که در آن ۷۱۵ تصویر رنگی از گوگل ارث، با وضوح مکانی ۰,۵ تا ۲ متر و ۸۵ تصویر مادون قرمز با وضوح بالا از داده های وبینگن با رزولوشن ۰,۰۸ متر وجود دارد.	داده ها در دو مجموعه دسته بندی می شوند:۱)مجموعه تصاویر مثبت به تعداد ۶۵۰ تصویر، شامل حداقل یک هدف در هر تصویر. ۲) مجموعه تصاویر منفی، شامل ۱۵۰ تصویر که هیچ هدفی را در خود ندارند. اشیاء موجود در این مجموعه تصاویر مثبت، شامل، ۷۵۷ هواپیما، ۳۰۲ کشتی، ۶۵۵ مخزن ذخیره سازی، ۳۹۰ زمین بیس بال، ۵۲۴ زمین تنیس، ۱۵۹ زمین بسکتبال، ۲۲۴ بندر، ۱۲۴ پل و ۴۷۷ وسیله نقلیه به صورت دستی با حاشیه نویسی به صورت bounding box استفاده می شود.	Object Detection
۶	Ke Li et.al (2020)[11]	DIOR	در مقایسه با دیتاست های قبلی، این است که اشیا نه تنها از نظر تفکیک پذیری فضایی بلکه از نظر تنوع اندازه بین کلاس و درون کلاسی دامنه	این مجموعه شامل ۲۳,۴۶۳ تصویر و ۱۹۲,۴۷۲ نمونه در ۲۰ کلاس شی است. که هر کلاس ۱۲۰۰ تصویر دارد.	Object Detection

			وسیعی از تغییرات دارند. کلاس های اشیاء در dior شامل: هواپیما، فرودگاه، زمین بیسبال، بسکتبال، پل، دودکش، سد،بزرگراه، ایستگاه عوارض، بندرگاه، زمین گلف، پیست زمینی، پل هوایی، کشتی، استادیوم، مخازن ذخیره سازی، زمین تنیس، ایستگاه قطار، وسایل نقلیه و آسیاب بادی است.	ابعاد تصویر در این دیتاست ۸۰۰ پیکسل است.	
۷	Xia et. al (2017)[12]	DOTA	این مجموعه شامل ۲۸۰۶ تصویر هوایی از صحنه ها و پلتفرم های مختلف است و مجموعه ای از اشیاء با طیف گسترده ای از مقیاس ها، جهت ها و شکل ها را به نمایش می گذارد. کلاس های این مجموعه شامل زمین بیسبال، بسکتبال، پل، بندر، دوومیدانی، هلیکوپتر، وسایل نقلیه بزرگ، هواپیما، چهارراه، کشتی، وسایل نقلیه کوچک، زمین فوتبال، مخازن ذخیره سازی، استخر شنا، زمین تنیس است.	شامل ۱۵ کلاس از ۲۸۰۶ تصویر و ۱۸۸,۲۸۲ نمونه است که هر تصویر حدوداً از ۸۰۰ در ۸۰۰ تا ابعاد ۴۰۰۰ در ۴۰۰۰ پیکسل ارائه شده است. تصاویر RGB بوده و GSD این مجموعه متنوع است. به دلیل اینکه اشیاء موجود در تصاویر هوایی محدود است. در مجموع کل نمونه های حاشیه نویسی شده در این مجموعه داده ۱۸۸,۲۸۲ نمونه است که هر نمونه توسط یک چهارضلعی دلخواه برچسب گذاری شده است.	Object Detection
۸	Robicquet et al. (2016)[13]	Stanford Drone Data	این مجموعه داده شامل بیش از ۱۹,۰۰۰ هدف شامل ۱۱,۲۰۰ عابر پیاده، ۶۴۰۰ دوچرخه سوار، ۱۳۰۰ اتومبیل، ۳۰۰ اسکیت سوار، ۲۰۰ ماشین گلف و ۱۰۰ اتوبوس است. این داده ها از فیلم های مربوط به محیط دانشگاه جمع آوری شده اند. تصویربرداری از ۶ منطقه انجام شده و بیش از ۱۰۰ صحنه برچسب زده وجود داردکه شامل: جاده، پیاده رو، دور برگردان، چمن، ساختمان و مسیر دوچرخه است.	داده ها معمولاً در ساعات تعطیلی بین کلاس ها جمع آوری شده تا چگالی اشیاء در تصاویر بالا برود. بنابراین در هر دوره ۲۰ ثانیه ای بین ۲۰ تا ۶۰ هدف مشاهده شده است (تقریباً در ۹۰۰ مترمربع). دقت تصاویر در این دیتاست، ۱۴۰۰ در ۱۹۰۴ است.	Object Tracking Object Detection
۹	Kaggle, (2018)[14]	Airbus Ship	مجموعه داده حاوی بیش از ۱۰۰,۰۰۰ تصویر ماهواره ای در حجم ۳۰ گیگابایت است. این مجموعه داده کاملاً نامتوازن است به طوری که فقط ۱ از ۴ تصویر داده دارای کشتی هستند.	ابعاد تصاویر در این مجموعه ۷۶۸ در ۷۶۸ است. یک فایل csv حاوی شناسه تصاویر و مختصات پیکسل ها وجود دارد که مختصات موقعیت کشتی را نشان می هد. اگر مختصات پیکسلی وجود نداشته باشد، یعنی تصویر حاوی وسیله نقلیه نیست.	Object Detection
۱۰	Du et al(2018)[15]	UAVDT	این مجموعه شامل حدود ۸۰,۰۰۰ فریم از ۱۰ ساعت فیلم برداری هستند.	وضوح تصویر در این مجموعه ۱۰۸۰ در ۵۴۰ پیکسل است و	Object Detection Object tracking

			این فیلم ها از مکان های دیدنی در مناطق شهری، شامل میادین، بزرگراه ها، گذرگاه ها و چهارراه ها تهیه شده است.	فیلم ها با نرخ ۳۰ فریم در ثانیه تهیه شده اند.	
۱۱	Zhu et. al (2020)[16]	VisDrone	تصاویر این مجموعه توسط هواپیماهای بدون سرنشین مختلفی در بیش از ۱۴ شهر در هزاران کیلومتر در چین، و در شرایط آب و هوایی و نورپردازی متفاوت ثبت شده است.	این مجموعه در ۱۰ کلاس، شامل ۲۶۳ توالی ویدیویی است که ۱۷۹,۲۶۴ فریم و ۱۰۲۰۹ تصویر ثابت دارد و شامل اشیاء مختلفی از جمله عابرین پیاده، وسایل نقلیه، دوچرخه و ... با تراکم کم و زیاد در صحنه هاست. فریم ها به صورت دستی با بیش از ۲,۵ میلیون bounding box محدود و با برخی ویژگی ها نظیر کلاس شی و موانع، حاشیه نویسی شده اند.	Object Detection Single-Object Tracking
۱۲	Xia et. Al (2016)[17]	AID	شامل ۱۰,۰۰۰ تصویر در ۳۰ کلاس که همه تصاویر برچسب دار هستند.	اندازه تصاویر در این مجموعه ۶۰۰ در ۶۰۰ پیکسل بوده و به صورت RGB تهیه شده اند. مقدار GSD در این مجموعه داده ۵۰-۸۰۰ سانتی متر در هر پیکسل است.	classification
۱۳	Helber et . al(2019)[18]	EuroSAT	شامل ۲۷,۰۰۰ تصویر در ۱۰ کلاس مختلف به صورت برچسب گذاری شده موجود است.	ابعاد تصاویر در آن ۲۵۶ در ۲۵۶ پیکسل است و در ۱۳ باند طیفی تهیه شده است. مقدار GSD در این مجموعه داده ۳۰ سانتی متر در هر پیکسل است.	classification
۱۴	Basu et.al (2015)[19]	SAT4	شامل ۵۰۰,۰۰۰ تصویر در ۴ کلاس به صورت برچسب دار است که در مساحت حدود ۸۰۰ کیلومتر مربع تهیه شده و شامل زمین های بایر، درختان، چمنزار، و کلاسی شامل سایر کلاس های پوشش زمین که در این سه مجموعه قرار ندارد، است. این مجموعه از یک مجموعه بزرگتر NAIP که شامل کل قاره آمریکاست استخراج شده است.	ابعاد تصاویر در آن ۲۸ در ۲۸ پیکسل بوده و به صورت RGB و NIR تهیه شده است. مقدار GSD در آن ۶۰۰ سانتی متر در پیکسل است.	classification
۱۵	Basu et.al (2015)[19]	SAT 6	شامل ۴۰۵,۰۰۰ تصویر در ۶ کلاس برچسب دار است که در مساحت حدود ۸۰۰ کیلومتر مربع تهیه شده و شامل زمین های بایر، درختان، چمنزار، جاده ها، ساختمان ها و سطوح آبی است. این مجموعه از یک مجموعه بزرگتر NAIP که شامل کل قاره آمریکاست استخراج شده است.	ابعاد تصاویر در آن ۲۸ در ۲۸ پیکسل بوده و به صورت RGB و NIR تهیه شده است. مقدار GSD در آن ۶۰۰ سانتی متر در پیکسل است.	classification

جدول ۱:دیتاست های مربوط به تصاویر هوایی

۳-۴ - ۱- ۲- مقالات تشخیص اشیا در تصاویر هوایی:

در ادامه، به برر سی انواع روش های ت شخیص ا شیاء مورد ا ستفاده در پردازش ت صاویر هوایی پرداخته شده ا ست. در جدول زیر، خلاصه ای از روش های رایج در این حوزه و دقت های بدست آمده در آنها ارائه شده است:

شماره	منبع	تکنیک	دیتاست	معیار سنجش دقت F score / mAP	سخت افزار
۱	Ke Li et.al (2020)[11]	a fully convolutional neural network pipeline (YOLT)	COWC	F score: 0.90±0.09	-
			SpaceNet	F score: 0.69	
۲	Kundid Vasi´c et.al (2020)[20]	RFCCD (RPN + FPN + Classification + Context + Deviation)	HERIDAL	68.89	Intel Xeon E5-2640v4 of 3.40 GHz, 4 * 16 GB DDR4 memory, and multi-GPU ۴* NVIDIA GeForce GTX 1080Ti Turbo with 11 GB memory
۳	Pang et. al (2019)[21]	Remote sensing region-based convolutional neural network(R2-CNN) از یک شبکه Tiny net به عنوان backbone استفاده می کند.	ترکیبی از دو مجموعه تصاویر ماهواره ای Gaofen-1 و Gaofen-2	96.04	-
۴	Ophoff et.al (2020)[22]	D-YOLO : Deconvolutional YOLO مبتنی بر YOLOv2	DOTA (روی کلاس های وسایل نقلیه و کشتی)	وسایل نقلیه 60% کشتی 66%	Nvidia GTX 1080 Ti GPU 4 ms per patch
۵	Acatay et. al (2018)[23]	D-YOLO : Deconvolutional YOLO مبتنی بر YOLOv2	DOTA	69.0	NVIDIA Titan XGPU
۶	Koga et. al (2020)[24]	Single Shot Multibox Detector (SSD) + unsupervised domain adaptation(DA) method	COWC (روی ۴ کلاس رنگی)	78.2	۱۸ ساعت روی Nvidia Tesla T4 (۲۰,۰۰۰ تکرار)
۷	Shen et. al (2020)[25]	ترکیبی از شبکه استخراج ویژگی های سبک وزن با الگوریتم تشخیص شی Faster R-CNN که برای تشخیص اشیا کوچک در صحنه های هوایی مناسب تر است.	Munich dataset و دیتاست نانجینگ، چین	85.5 و 91.2	زمان تشخیص بر روی دو دیتاست با Nvidia Titan XP ۱,۷۸ و ۰,۰۴۸ ثانیه
۸	Alganci et.al (2020)[26]	Faster R-CNN + Augmentation	DOTA(airplane)	F score=0.93	-

-	69.56 86.16	DOTA HRSC2016	RoI Transformer + Light-Head R-CNN with backbone ResNet101	Ding et.al (2019)[27]	۹
NVIDIA Titan X GPUs	95.01 72.45 95.67	NWUH VHR-10 DOTA UCAS-AOD	ماژولهای استفاده شده در روش پیشنهادی:-end-to-end CNN + image cascade net-work (ICN),deformable inception network (DIN), FPN, multi-scale rotational region-proposal network (R-RPN),multi-scale rotational region of interest net-work (R-ROI), and rotational non-maximum suppression (R-NMS)	Azimi et.al (2018)[28]	۱۰
Intel i7-6700K CPU and NVIDIA GTX1080 GPU	84.4 88.3	NWPU VHR-10 SORSI (دو مجموعه کشتی و هواپیما)	Deformable Faster RCNN (ResNet-50) with TCB Deformable Faster RCNN with TCB (+OHEM)	Ren et.al (2018)[29]	۱۱
64-bit Ubuntu 16.04 computer with 8GB memory GeForce GTX1070Ti GPU	79.2	DOTA	Double Multi-scale Feature Pyramid Network (DM-FPN)	Zhang et.al (2019)[30]	۱۲
TitanXp GPU	51.95 79.65 56.24 80.02	VEDAI DOTA DLR-3K هرسه باهم(ABD)	one-stage vehicle detection network (AVDNet : AVDNet Convolutional Network)	Mandal et.al (2019)[31]	۱۳
8 Geforce RTX 2080 Ti GPUs	76.24 89.94	DOTA HRSC2016	FPN(ResNet101) as backbone + Arbitrary-Oriented Region Proposal Network (AO-RPN) + RRoI + Multi-Head Network (MH-Net)	Qin et.al (2021)[32]	۱۴
3 hours for 200 epochs on an NVIDIAGTX 1080 TI GPU	45.40	VEDAI	Generative Data Augmentation Pluralisticand (as generator) + DeepFill: fully convolutional neural net-work (as detector)	Kumdakcı et.al (2020)[33]	۱۵
on 4 Quadro RTX 6000 GPU	75.36	DOTA	U-shaped network: ResNet101 Conv1-5(backbone: pre-train on ImageNet) + box boundary-aware vectors (BBAVectors hr): oriented bounding box (OBB) + rotation bounding boxes (RBB)	Yi et. al (2020)[34]	۱۶
3 Titan Xp GPUs	75.02 88.2	DOTA HRSC2016	fasterR-CNN + RPN + RoIAlign	Xu et.al (2019)[35]	۱۷
RTX 2080 Ti GPU	89.77 76.95 88.3	HRSC2016 DOTA NWPU VHR-10	RetinaNet (baseline) + Dynamic Anchor Learning(DAL) Dynamic Anchor Learning(DAL)+ S2A-Net RetinaNet + DAL	Ming et.al (2020)[36]	۱۸
V100 GPU	79.42	DOTA	Single-Shot Alignment Network (S2A-Net): backbone + FPN (Feature Pyramid Network) + Feature Alignment Module(FAM) + Oriented Detection Module (ODM)	Han et.al (2020)[37]	۱۹

۲۰	Ding et.al (2021)[38]	re-implement of Faster R-CNN OBB (R-50-FPN)+ RoI Transformer (Aug) re-implement of Faster R-CNN HBB (R-50-FPN)+ RoI Transformer (Aug)	DOTA DOTA	79.82 80.7	4 GPUs for training with a total batch size of 8 (2 images per GPU) A single NVIDIA Tesla V100 GPU

جدول ۲: تشخیص اشیاء در تصاویر هوایی

۳-۴ - ۱- ۳ – چالش های موجود برای تشخیص اشیاء در تصاویر هوایی:

پاسخ به این سؤال، که تشخیص اشیاء در تصاویر با چه چالش‌هایی روبرو است، آسان نیست. از آنجا که تشخیص اشیاء در کاربردهای متفاوتی به کار می‌رود، با توجه به تنوع زاویه تصویربرداری، تغییرات نور، تغییرات نمونه‌های یک کلاس، اهداف و محدودیت‌های موجود در هر مسئله، کاملاً متفاوت است. به عنوان برخی از چالش‌های رایج در این حوزه، می‌توان به چرخش اشیاء، تغییر مقیاس، وضوح متفاوت، تراکم، سرعت متفاوت در شناسایی و ... اشاره کرد[۳۹].

در حوزه تشخیص اشیاء در تصاویر راه دور، علاوه بر وجود چالش‌های عمومی تشخیص اشیاء، با چالش‌های دیگری نیز روبرو هستیم. با بررسی صورت گرفته روی دیتاست‌های فوق، می‌توان چالش‌های موجود در زمینه تهیه دیتاست تصاویر هوایی در کاربردهای تشخیص اشیاء را به طور خلاصه بیان کرد.

یکی از مسائل موجود این است، زمانی که از تصاویر ماهواره ای استفاده شود، مقدار زیادی داده تولید خواهد شد. با این‌حال، بخش عظیمی از اطلاعات آموزنده نیستند، چون معمولاً بخش بزرگی از تصاویر فقط شامل سطح اقیانوس است که هیچ چیز مهم و مورد علاقه‌ای ندارد و یا مناطق بالقوه خشکی‌های مجاور دریا را شامل می‌شود. بنابراین تنک بودن داده‌ها می‌تواند بخش بزرگی از داده‌های ما را بلااستفاده کند[۴۰]. مورد بعدی، وجود مقیاس‌های متنوع برای اشیاء است. با توجه به فاصله ابزار تصویربرداری و دقت آن، ممکن است اشیاء مشابه در تصاویر با اندازه‌های متفاوت یا زاویه قرارگیری متفاوت ظاهر شوند که یک سیستم عادی تشخیص، نمی تواند این موارد را به درستی تشخیص دهد. یکی دیگر از چالش‌های موجود، تراکم در تصاویر است. وجود تراکم در تصاویر یا تجمع پیش زیاد اشیاء در یک صحنه، برای مثال، تصویر یک پارکینگ، ممکن است باعث مشکل در تشخیص تعداد اشیاء و تخمین صحیح همه موارد شود. از دیگر چالش های این حوزه، وجود مانع برای مشاهده کامل شی است، برای مثال، ممکن است به دلیل هوای نامساعد یا وجود ابر یا دود، بخشی از شی قابل مشاهده نباشد. این مورد گاهاً حاشیه نویسی تصویر را نیز با مشکل روبرو می‌کند، در این صورت سیستم باید قادر به تشخیص صحیح اشیاء با مشاهده از تصویر بخشی از آن نیز باشد.

یکی دیگر از مسائلی که در تهیه دیتاست با آن روبرو هستیم، محدودیت در تعداد اشیاء قابل مشاهده در تصاویر هوایی است که باعث محدود بودن تعداد کلاس‌ها در دیتاست‌ها می‌شود و کمبود تعداد کلاس دقت روش‌های فعلی را تحت تأثیر قرار می‌دهد.

نیاز به تنوع موقعیت جغرافیایی برای تصویربرداری، اندازه بزرگ تصاویر مربوط به داده‌های هوایی، رزولوشن بالا و سرعت پردازش نیز از سایر چالش‌های موجود برای دیتاست‌های تصاویر هوایی است که در دیتاست‌های ارائه شده سعی شده تا حدودی با پیش‌پردازش و حاشیه‌نویسی دقیق، این موارد جبران شود.

به طور کلی، می‌توان چالش های مرتبط با دیتاست‌های این حوزه را در چند مورد خلاصه کرد:

- تنک بودن داده‌ها در سطح کل مجموعه داده (مانند سطح اقیانوس ها، زمین‌های بایر و ...)
- احتمال تراکم بالا در ابعاد کوچک (مانند پارکینگ‌ها)
- تغییرات نور که باعث کاهش قدرت تشخیص می‌شود
- وجود موانع که باعث عدم مشاهده کامل شی می‌شود
- احتمال وجود شرایط خاص، نظیر تغییرات آب و هوایی، باران، آتش‌سوزی که باعث افت قدرت تشخیص اشیاء می‌شود
- محدود بودن تعداد کلاس‌ها به دلیل عدم امکان تشخیص همه اشیاء در تصاویر ماهواره‌ای
- مقیاس‌های متفاوت برای اشیاء بسته به ارتفاع شی تصویربردار از سطح زمین
- تغییرات زاویه دید اشیاء در تصاویر هوایی

بنابراین، باید برخی از این موارد از طریق انتخاب بهترین دیتاست و برخی از طریق بهبود الگوریتم تشخیص، جبران شود.

۳-۴ - ۲- تشخیص اشیاء با تصاویر اندک:

در دامنه‌های خاص، مانند مراقبت‌های بهداشتی(پیش‌بینی بیمارهای نادر)، حرکت ربات‌ها یا بررسی حوادث طبیعی، به‌دست آوردن مجموعه داده‌های بزرگ تقریباً غیرممکن است یا هزینه زیادی دارد. علاوه بر این، مجموعه داده‌های بزرگتر به زمان آموزش و هزینه های محاسباتی بیشتری نیاز دارند. اما از آنجا که قدرت بازنمایی یادگیری عمیق از نقص بالقوه آموزش در آن بیشتر است، معماری های عمیق معمولاً در بازنمایی دانش مورد پسند قرار می‌گیرند. برای پرداختن به مسئله داده‌های بزرگ آموزشی، به‌طور فزاینده‌ای از یادگیری نیمه نظارتی یا بدون نظارت در معماری‌های عمیق، استفاده می‌شود که در این حالت، مدل‌ها وظیفه دارند با ارائه تنها چند نمونه برچسب‌دار، یک مفهوم جدید را یاد بگیرند. زمانی‌که تنها یک مثال برچسب‌دار وجود دارد یا هیچ نمونه برچسب‌داری نداریم، رویکردهایی مورد استفاده قرار می‌گیرد که به ترتیب one-shot learning و zero-shot گفته می‌شود. اخیراً رویکرد k-shot learning مورد توجه قرار گرفته و هدف از این رویکرد این است که با استفاده از نمونه‌های انگشت شمار و با کمک الگوریتم‌های موجود و بهره‌برداری از ماهیت داده‌ها، بتواند به طور مؤثر یاد بگیرد[۴۱].

یک سناریوی دیگر برای لزوم FSL، زمانی است که به دلیل حفظ حریم خصوصی، امنیت و رعایت اصول اخلاقی، امکان دستیابی به نمونه‌های با ناظر، سخت یا غیرممکن است. به عنوان مثال، می‌توان به کشف دارو اشاره کرد که سعی دارد با کشف خواص مولکول های جدید، مواد مفید را به عنوان داروهای جدید شناسایی کند. به دلیل سمی بودن احتمالی، فعالیت و حلالیت کم، مولکول‌های جدید امکان تست به صورت بالینی را نخواهند داشت. از این‌رو، یادگیری مؤثر از طریق تعداد نمونه‌های کم بسیار اهمیت دارد. با استفاده از روش FSL، یادگیری مدل‌های مناسب برای مواردی که نمونه‌های نادری دارند، امکان پذیر می‌شود.

از مزایای FSL این است که می‌تواند از مشکلات جمع‌آوری داده‌های باناظر در مقیاس بزرگ، خلاص شود. به عنوان مثال، اگرچه ResNet(55)، نسبت به انسان روی مجموعه داده ImageNet عملکرد بهتری دارد، اما لازم است که هر کلاس، دارای تصاویر برچسب دار کافی باشد، که جمع‌آوری این داده‌ها دشوار است. FSL می‌تواند نیاز به جمع‌آوری داده‌ها را برای برنامه‌های کاربردی پرمحتوا[۴]، کاهش دهد. به عنوان نمونه‌هایی از مسائل این حوزه، می‌توان به طبقه‌بندی تصاویر، بازیابی تصویر[۵]، ردیابی اشیاء[۶]، تشخیص حالت[۷]، عنوان سازی تصاویر[۸]، پاسخگویی به سؤالات تصویری[۹]، تشخیص رویدادهای ویدیویی[۱۰]، مدل‌سازی زبان[۱۱] و جستجوی معماری عصبی[۱۲]، اشاره کرد[۵].

FSL، با هدف نزدیک شدن هوش مصنوعی به انسان و با توجه به نیاز موجود در صنعت برای یادگیری ارزان، اخیراً توجهات زیادی را به خود جلب کرده و موضوع بروزی تلقی می‌شود. رویکردهای یادگیری ماشین زیادی برای انجام این هدف، ارائه شده‌اند که به عنوان مثال می‌توان به یادگیری[۱۳]، یادگیری تعبیه‌شده[۱۴] و مدل‌سازی تولیدی[۱۵]، اشاره کرد[۵].

در این بخش به بررسی مقالاتی در حوزه تشخیص اشیاء با داده‌های اندک پرداخته شده و پس از آن چالش‌های این حوزه اشاره شده است.

شماره	منبع	استراتژی	تکنیک	دیتاست	دقت

[4] . data-intensive applications
[5] . image retrieval
[6] . object tracking
[7] . gesture recognition
[8] . image captioning
[9] . visual question answering
[10] . video event detection
[11] . language modeling
[12] . neural architecture search
[13] . meta-learning
[14] . embedding learning
[15] . generative modeling

5-shot:55.7 1-shot: 39.8	PASCAL VOC	two-stage fine-tuning approach (TFA)+ Faster R-CNN + cosine similarity	Transfer Learning	Wang et.al (2020)[42]	۱
1-shot:24.4 5-shot: 25.9	COCO				
1-shot: 14.8 – 21.3 5-shot:33.9-42.8	PASCAL VOC	meta feature extractor + a reweighting module	meta-learning	Kang et.al (2019)[43]	۲
50cat-5shot: 41.3 50cat-5shot: 44.1	COCO FSOD	a weight-shared network+ attention RPN module+ multi-relation detector + (fine-tune by ImageNet)	Transfer Learning	Fan et.al (2020)[44]	۳
20cat-10shot:31.3	FSOD	a weight-shared network+ attention RPN module+ multi-relation detector	Transfer Learning		
10shot: 51.5 5shot: 45.7 1-shot:19.9	PASCAL VOC	Meta R-CNN:1) Faster/ Mask R-CNN; 2) Predictor-head Remodeling Network (PRN)	Transfer Learning	Yan et.al (2019)[45]	۴
1-shot-5way:24.2 2-shot-5way:35.3 3-shot-5way:42.2 5-shot-5way:49.1 10-shot-5way:57.4	PASCAL VOC	Faster R-CNN+ RPN+ proposal-level feature alignment module	meta-learning	Xiao et al (2020)[46]	۶
10shot-20way:27.3 30shot-20way:30.6	COCO				
1shot:34 2shot:51.9 5shot:60.9 10shot:65.5 30shot: 69.7	COCO	LSTD	Transfer Learning	Chen et al(2018)[47]	۷
1shot:33.6 2shot:42.5 5shot:50.9 10shot:54.5 30shot: 58.3	ImageNet2015				

جدول ۳ :تشخیص اشیاء با تصاویر اندک

رویکردهای متنوعی برای یادگیری با نمونه‌های اندک برچسب‌دار، وجود دارد که در بخش‌های زیر متفاوت هستند:

۱) ساخت مدل‌های کم عمق (شبکه‌های بیزی، سیستم‌های مبتنی بر قاعده، مدل‌های مارکوف) و شبکه‌های عمیق

۲) مرحله آموزش (سازگاری دامنه برای انتقال یادگیری، آموزش مبتنی بر حافظه انجمنی)

۳) دامنه‌های وظایف (یادگیری مفاهیم بصری، کارهای مرتبط با کنترل حرکت در رباتیک)

۴) نوع داده (تصاویر نمادین، تصاویر دنیای واقعی، چهره انسان، خواص شیمیایی نمایش داده شده به عنوان گفتار، گراف‌ها و زبان طبیعی، داده‌های حسگر در رباتیک)

۵) ماهیت داده‌ها (پیوسته یا گسسته، متوالی یا تصادفی).

با توجه به چنین تنوعی در ورودی، نمایش داده‌ها، الگوریتم های یادگیری و کار هدف، ادبیات یادگیری با تصاویر اندک را می‌سازند که گاهی فهم این مفاهیم نیز دشوار است[۴۱].

از آنجا که یادگیری با تصاویر اندک، با هدف غلبه بر روش‌هایی که نیاز به داده‌های زیاد دارد، ارائه شده است، از یادگیری بدون ناظر/نیمه نظارتی استفاده می‌کند. با این حال، به طور سنتی روش‌های یادگیری بدون ناظر، برای یافتن ساختار داده‌ها بیشتر بر روی درک داده متمرکز می‌شوند. در روش‌های یادگیری با تصاویر اندک، تک تصویر یا بدون تصویر، از یادگیری باناظر استفاده می‌شود. ممکن است استفاده از یادگیری با ناظر در مواردی که داده‌های آموزشی بسیار کم هستند، برخلاف تصور باشد، اما در این مورد، درست است. در این رویکرد، برای آموزش مؤثر، از داده‌های آموزشی سایر کارهای مرتبط استفاده می‌شود. به طور خاص، شکاف موجود در دانش عنصر یادگیرنده، که به دلیل فقدان نمونه‌های آموزشی کافی وجود دارد، توسط این موارد برطرف می‌گردد:

۱) قرض گرفتن دانش: رویکردهای transfer learning، مبتنی بر استفاده مجدد از معماری مدل و پارامترهایی است که از داده‌های آموزشی کارهای مرتبط(منابع)، آموخته می‌شود. دانش زمینه‌ای یا قبلی، که در کار منبع، آموخته شده است، برای کار هدف می‌تواند مجدداً استفاده شود.

۲) ایجاد دانش از طریق تعمیم داده‌ها: تعمیم رویکردهایی که بازنمایی غنی را به صورت انتزاعی و سریع می‌آموزد.

۳. یادآوری دانش آموخته شده (از قبل)، از طریق حافظه و توجه: دادن مدل‌هایی با توانایی شناخت برای ذخیره دانش خاص هر وظیفه[۱۶]، که آنها را قادر می‌سازد تا دانش مرتبط را به طور مؤثر به یاد آورند.

۴. به‌دست آوردن/تولید داده هایی که به ایجاد دانش جدید کمک می‌کند. رویکردهای تولید[۱۷]، با افزودن داده‌ها یا با یادگیری بازنمایی‌های غنی، داده‌های ترکیبی ایجاد کرده و با استفاده از مدل‌های موجود، به حل مسئله few-shot کمک می‌کنند.

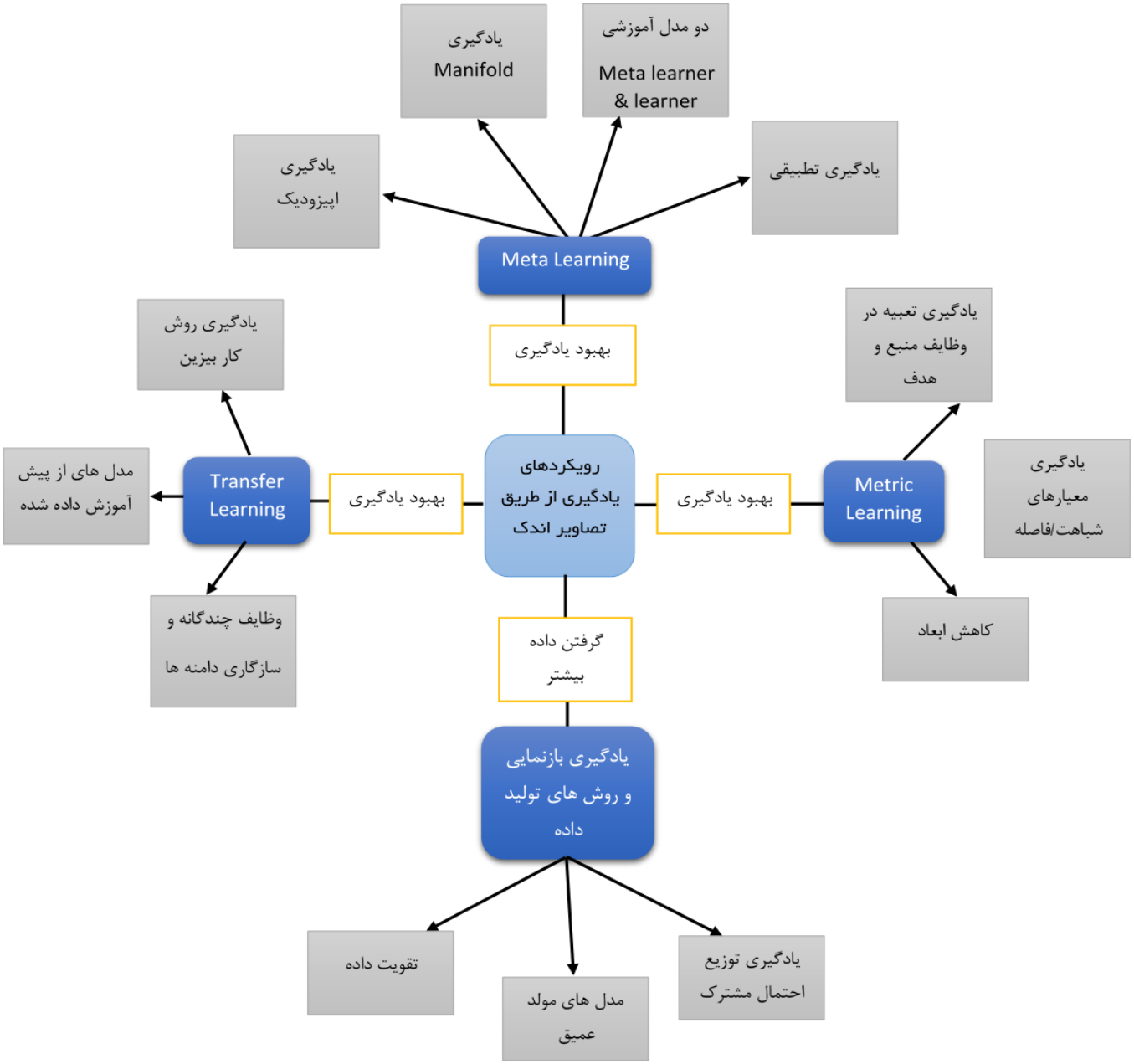
همه انواع یادگیری با تصاویر اندک، یک هدف مشترک دارند که آن کسب دانش هرچه بیشتر برای رسیدن به عملکرد مناسب روی وظایف هدف، است[۴۸].

معمولاً از دو جنبه، می‌توان رویکردهای یادگیری ماشین در رابطه با تصاویر اندک را بررسی کرد. اولین مورد رویکردهای مبتنی بر داده و دومین مورد، تغییر در رویکردهای مبتنی بر یادگیری است. رویکردهای مبتنی بر داده، از طریق اعمال تغییر روی داده، نظیر یافتن بازنمایی‌های غنی تر/ نمونه اولیه داده‌ها، تولید داده‌های برچسب‌دار ترکیبی، از طریق افزایش یا تبدیل داده‌های اولیه، کار می کنند. افزایش یا تقویت داده‌ها، می‌تواند با استفاده از تقویت[۱۸] یا تغییر شکل[۱۹] انجام شود. تقویت داده، می‌تواند با اعمال انواع انتقال یا افزودن نویز تصادفی به تصاویر اندک اولیه، داده‌های بیشتری تولید کنند. این مورد باعث تقویت روش‌های یادگیری عمیق می‌شود، مخصوصاً در مسائل اندک که از ابتدا داده مناسبی وجود ندارد، این کار می‌تواند مفید باشد. اما مطالعات بسیار کمی در خصوص تأثیر تقویت داده در مسائل few-shot وجود دارد.

از طرفی، مدل‌های مولد، نظیر شبکه‌های GAN، یا رمزگذارهای خودکار کانولوشن، می‌توانند بازنمایی‌های خوبی را یاد بگیرند و بنابراین در تولید داده‌های بیشتر مفید باشند. با این‌حال، آموزش شبکه‌های GAN هنوز یک زمینه تحقیقاتی در حال رشد دارد و کاربرد آن در سناریوی تصاویر اندک، هنوز رایج نیست. تصویر زیر، یک طبقه‌بندی از رویکردها و استراتژی‌های موجود برای حل مسائل تصاویر اندک، به منظور بهبود یادگیری در این نوع مسائل را نشان می‌دهد. این رویکردها غالباً انحصاری نیستند و بعضاً به یکدیگر وابسته‌اند. برای نمونه، برخی از رویکردهای meta-learning، از معیارهای فاصله برای یافتن مفاهیم مشابه استفاده می‌کنند. Transfer learning، metric-learning و اخیراً meta-learning، از مؤثرترین رویکردهای یادگیری در سناریوهای few-shot هستند [۴۱]. مطالعات نشان می‌دهد که روش‌های meta-learning در حالت معمول few-shot learning، بهتر از روش‌های transfer learning عمل می‌کنند[۴۸].

---

16 . task-specific
17 . Generative approaches
18 . augmentation
19 . transformation



شکل ۱ : روش های موجود در FSL (Kadam,2020)[41]

۱. **Transfer learning**: از یک مدل آموزش دیده روی وظایف منبع مجدداً برای رسیدن به عملکرد مناسب روی وظیفه هدف، استفاده می‌کند. تکنیک‌های این روش، از توانایی معماری‌های عمیق، برای یادگیری ویژگی‌های مشترک بین وظایف موجود در لایه‌های پایین استفاده می‌کنند. مدل‌های منبع از پیش آموزش‌دیده (آموزش داده شده روی مجموعه داده‌های بزرگ و چالش برانگیز)، که برای هدف مناسب‌تر هستند، فقط با آموزش لایه‌های بالایی برای کارهای هدف، اجرا می‌شوند.

۲. **Metric-learning**: یک یادگیری نیمه نظارتی یا بدون ناظر است که در آن داده‌های بدون برچسب به منظور پیش‌بینی، خوشه‌بندی می‌شوند. اما روش‌های قابل قبول زیادی برای تشکیل خوشه‌ها، به ویژه برای ورودی‌های با ابعاد بالا مثل تصاویر، وجود دارد. معیارهای فاصله برای تشکیل خوشه‌ها حیاتی هستند و تا یافتن خوشه‌های قوی، اصلاح می‌شوند. یادگیری متریک، تغییر پارامترها را به عنوان مسئله بهینه سازی مطرح می‌کند تا معیارهای فاصله از داده‌های آموزش داده‌شده، یاد گرفته شوند. یادگیری متریک، به دلیل اینکه کمک می‌کند از مشابهت یک نمونه جدید بدون برچسب با مثال‌های قبلی برچسب خورده، در پیش‌بینی یا طبقه‌بندی برچسب آن، استفاده شود، در یادگیری با تصاویر اندک بسیار مفید است.

۳. **Meta-learning**: به مسئله هدایت مؤثر الگوریتم‌های بهینه‌سازی مبتنی بر شیب، در فضای پارامترهای بزرگ می‌پردازد. این امر به ویژه در یادگیری با تصاویر اندک چالش برانگیز است، زیرا نمونه‌های آموزشی کمتری برای اصلاح شیب در صورت بروز خطا، در دسترس است. الگوریتم‌های متداول بهینه‌سازی مانند روش گرادیان نزولی (SGD)، تا زمان همگرایی، نیاز به تکرارهای زیادی دارند. این تکرارها در فضای پارامتری با ابعاد بالا و داده‌های آموزشی کوچک به خوبی کار نمی‌کنند، چون موجب overfit می‌شوند. تکنیک learning to learn، حتی زمانی‌که مثال های راهنمای کمتری وجود دارد نیز، بر روی توانایی الگوریتم‌های بهینه سازی برای هدایت هوشمند فضای پارامتری برای همگرایی سریعتر، تمرکز دارند.

رویکردهای transfer learning، از شبکه آموزش دیده شده‌ای استفاده می‌کنند که دانش خود را از وظایف منبع کسب کرده و با ذخیره مدل، از آن برای پیش بینی وظایف هدف استفاده می‌کند. رویکردهای مبتنی بر انتقال معمولا زمانی خوب عمل می‌کنند که روی وظایف منبع مرتبط آموزش داده شده باشند. رویکردهای یادگیری متریک، سعی می‌کنند از ترسیم فضاهای تعبیه‌شده در وظایف منبع و هدف، برای سنجش شباهت و طبقه‌بندی نمونه‌های جدید استفاده کنند. در حال حاضر، رویکردهای متا، امیدوارکننده‌ترین روش‌ها هستند که معمولاً دارای دو مدل، meta-learner و learner model هستند[۴۱]. اما ترکیب دو روش نیز نیاز به بررسی دقیق تری دارد.

۳-۴ - ۲-۱- تشخیص اشیاء با تصاویر اندک در تصاویر هوایی:

شماره	منبع	استراتژی	تکنیک	دیتاست	دقت	سخت افزار
۱	Deng et.al (2020)[49]	Meta and Transfer learning	FSODM	NWPU VHR-10	10shot: 65 5shot: 53 3shot: 32	-
				DIOR	20shot: 36 10shot: 32 5shot: 25	
			YOLOv3	NWPU VHR-10	20shot: 28 10shot: 14	
				DIOR	20shot: 25 10shot: 20 5shot: 14	
			Feature Reweighting	NWPU VHR-10	10shot: 40 5shot: 24 3shot: 12	
				DIOR	20shot: 34 10shot:28 5shot: 22	
۲	Xiao et al (2020)[50]	Transfer learning	Self-Adaptive Attention Network (SAAN) + relation GRU	RSOD (تصاویر پانوراما)	(Airplane) 1-shot-5way:10.8 2-shot-5way:12.34 3-shot-5way:23.20 5-shot-5way:34.38 10-shot-5way:45.09	-
				NWPUVHR-10	(Airplane) 1-shot-5way:9.09 2-shot-5way:15.72 3-shot-5way:25.71 5-shot-5way:34.68 10-shot-5way:38.78	

جدول ۴: تشخیص اشیاء در تصاویر هوایی، با داده های اندک

۳-۴ - ۲- ۲ – چالش‌های موجود برای تشخیص اشیاء در تصاویر اندک:

در دنیایی که با داده‌های نامحدود و منابع محاسباتی، ممکن است به تکنیک دیگری بجز یادگیری عمیق نیاز نداشته باشیم. با این حال، ما در یک دنیای واقعی زندگی می‌کنیم که داده‌ها هرگز بی نهایت نیستند، به‌ویژه در مباحث سنجش از راه دور که داده‌های دریافتی در مقیاس بزرگ بوده و هزینه جمع آوری داده بسیار بالاست[۴۲].

اگرچه مجموعه داده‌های تصاویر هوایی روز به روز در حال افزایش مقیاس هستند، اما بیشتر آنها برای کاربرد در یادگیری عمیق هنوز کوچک هستند. به همین دلیل، در این مسائل، روش‌های FSL به عنوان روشی جایگزین برای پرداختن به مسائلی که تشنه داده هستند، موضع جدیدی از بررسی را ارائه می‌دهند. FSL به جای تمرکز بر مقیاس دیتاست، سعی می‌کند مدلی را بیاموزد که بتواند به سرعت از چند نمونه دارای برچسب، تعمیمی برای کارهای جدید بیاموزد. مسلماً FSL به عملکرد انسانی شباهت بیشتری دارد.

به طور خلاصه می‌توان گفت مزایای استفاده از few-shot را در مقایسه با روش‌های دیگر یادگیری به شرح زیر است:

- کمک به کاهش بار محاسباتی و هزینه جمع‌آوری داده‌ها
- راه حلی برای مسائلی که برای آن داده زیادی وجود ندارد به عنوان مثال، یک بیماری جدید، یا تصاویر هوایی
- کاهش شکاف بین هوش مصنوعی و یادگیری مانند انسان

اما یکی از مشکلات موجود برای این روش‌ها، این است که با توجه به تنوع مسائل و داده‌های موردنیاز هر مسئله، روش‌ها و نتایج برای هر نوع داده می‌تواند کاملا متفاوت با روش دیگر باشد و تنها به تغییر دیتاست نمی‌توان اکتفا کرد. بنابراین برای هر نوع مسئله، نیاز است روش‌های مختلفی بررسی شود تا به الگوریتمی با دقت مناسب برسیم. بنابراین مشکلات و چالش‌های پیش روی FSL در زمینه تشخیص اشیاء در تصاویر هوایی را می‌توان در موارد زیر خلاصه کرد:

- با توجه به کمبود داده در این روش، باید دانش قبلی را تا حد امکان تقویت کرد.
- نیاز به آزمایش روش‌های متفاوت تقویت و دستکاری داده برای رسیدن به نتیجه دلخواه
- باتوجه به اینکه هنگام یادگیری، نمونه‌های آموزشی کمی داریم، به راحتی امکان overfit وجود دارد.
- عدم بررسی روش‌های موجود، روی دیتاست‌های تصاویر پهپادی نظیر visDrone یا Stanford Drone Data
- محدود بودن تعداد اشیاء قابل مشاهده در تصاویر هوایی، باعث محدود بودن تعداد کلاس‌ها در دیتاست هاست.
- باتوجه به گسترش کاربردهای پهپادها، افزایش کارایی، دقت و سطح استقلال آنها در تشخیص اشیاء مدنظر است. اما باتوجه به جدید بودن این موضوع، مطالعات زیادی در این حوزه صورت نگرفته است.
- وجود زاویه دید متفاوت، ارتفاع تصویربرداری متفاوت، ابعاد متنوعی را در داده‌های تصاویر هوایی ایجاد می‌کند. علاوه بر این، مواردی نظیر تغییر شرایط روشنایی، وجود مانع، زمینه‌های پیچیده که باعث می‌شود مرز اشیاء مبهم باشد و ... باعث وجود دقت های پایین در این حوزه شده است.
- در صورت وجود چندین شی در یک تصویر واحد، ممکن است تشخیص اشیاء در کلاس‌های جدید با کلاس‌های دیگر تداخل داشته باشد.
- مطالعات در خصوص دسته‌بندی تصاویر هوایی با داده اندک زیاد است، اما چون تشخیص شی علاوه بر تعیین کلاس، باید محل شی را نیز تعیین کند، امکان استفاده از کارهای انجام شده برای طبقه‌بندی، در تشخیص اشیاء در تصاویر هوایی وجود ندارد.

همه این موارد، نیاز به مطالعات بیشتر در این حوزه را بیش از پیش تقویت می‌کند.

۳-۵ ـ فرضیه‌های تحقیق (سؤالات تحقیق) :

در ابتدا دیتا ست‌های ارائه شده به طور دقیق مورد بررسی قرار می‌گیرد و ویژگی‌هایی که باعث انتخاب این دیتا ست‌ها در تشخیص اشیاء با تصاویر اندک، می‌شود، بررسی خواهد شد تا دیتاست تصاویر هوایی مبتنی بر این اطلاعات ویرایش و آماده گردد.

بنابراین این سوال مطرح می‌شود که چه ویژگی‌هایی باید در دیتاست انتخابی وجود داشته باشد، تا بتواند با داده‌های اندک نیز نتایج مناسب بدهد؟ در حال حاضر باتوجه به مطالعه صورت گرفته، دیتا ست‌های visDrone و UAVDT، DIOR و NWPU VHR-10 نسبت به سایر موارد در اولویت قرار دارند، ساختار داده‌های موردنظر باید چگونه باشد تا به نتایج بهتری دست پیدا کنیم؟

با بررسی دقیق تر روش‌های موجود در تشخیص اشیاء در تصاویر اندک، یکی از روش‌های کارآمدتر انتخاب شده و استفاده می‌شود. روشی مدنظر قرار می‌گیرد که چالش‌های مطرح شده در این حوزه را تا حد امکان به حداقل برساند. با توجه به اینکه در کاربردهای

دیگر تشخیص اشیاء با تصاویر کم، در حال حاضر روش‌های meta-learning کارآمدتر شناخته شده‌اند، آیا استفاده از این روش ها در تشخیص اشیاء در تصاویر هوایی نیز عملکرد مناسبی خواهد داشت؟

برای غلبه بر چالش‌های کمبود داده‌های برچسب‌دار در تصاویر راه دور و هوایی، آیا روش‌های تقویت داده می‌توانند در تشخیص اشیاء در این حالت، مؤثر باشند؟

با توجه به مطالعات قبلی، در تشخیص اشیاء با داده های اندک، روش های meta-learning نتایج بهتری می دهند اما در آخرین مطالعه صورت گرفته، روش ارائه شده مبتنی بر ترکیب دو روش meta-learning و deep transfer learning بوده و اعلام شده ترکیب دو روش، روی یک دیتاست مشخص، نتایج بهتری ارائه داده است، آیا ترکیب این دو روش، در سایر دیتاست های منتخب نیز منجر به نتایج بهتر می شود؟ آیا امکان ترکیب مؤثرتر دو روش، برای رسیدن به نتایج بهتر وجود دارد؟

باتوجه به اینکه برخی از دیتاست‌های موجود به صورت فیلم ارائه شده‌اند، آیا برای غلبه بر مشکل وجود موانع موقت، می‌توان از دانش موجود در فریم‌های قبل و بعد مرتبط با فریم موجود استفاده کرد؟

۳-۶ ـ نتایج مورد انتظار پس از انجام تحقیق :

انتظار می‌رود با انجام این پژوهش، روش جدیدی برای تشخیص اشیاء در صحنه‌های راه دور با تصاویر اندک، ارائه گردد و کارآمدی آن روی داده‌های هدف، قابل قبول باشد.

۳-۷ ـ درصورت کاربردی بودن تحقیق، چه سازمانهایی می‌توانند از نتایج بدست آمده استفاده کنند .

ردیف	نام سازمان	نوع استفاده
۱	آتش نشانی و ستادهای بحران	شناسایی حوادث غیرمترقبه، عملیات نجات
۲	راهنمایی و رانندگی	کنترل ترافیک و حرکات نامتعارف
۳	نیروهای انتظامی و نظامی	بررسی حرکات مشکوک و اشیاء ناشناس
۴	شهرداری	بررسی و مدیریت تراکم ساختمانی

۴ ـ روش انجام تحقیق :

۴-۱ ـ روش گردآوری اطلاعات :

در ابتدا با مطالعه بیشتر روی منابع و مقالات معتبر در حوزه تشخیص اشیاء، تشخیص اشیاء در تصاویر هوایی، روش یادگیری با تصاویر اندک، سعی می شود تا اطلاعات اولیه در رابطه با روش‌های تقویت داده، مدل و الگوریتم برای کاربرد موردنظر بررسی شود. سپس با انتخاب و ترکیب روش‌هایی که در تشخیص اشیاء در حوزه مدنظر توانایی لازم را داشته باشند، روش کار مشخص می‌شود. سپس با توجه به تجربیات بدست آمده در طول کار، ایده‌هایی که پتانسیل بالا و قابلیت پیاده‌سازی در این حوزه را دارند، مطرح شده و در نهایت پس از پیاده‌سازی آن، دقت روش مورد ارزیابی قرار گرفته و نتایج ثبت خواهد شد. در پایان نیز مواردی که در تحقیقات آینده مفید تلقی شود، ارائه خواهد شد.

۴-۲ ـ روش و مراحل تحقیق :

در اولین مرحله، باید دیتاستی که در تشخیص اشیاء مناسب و غنی‌تر باشد، چه از نظر نوع حاشیه‌نویسی و چه از نظر تعداد و نوع دسته‌های در نظر گرفته شده در آن، انتخاب می‌شود. با مطالعات فعلی صورت گرفته، و نتایج ارائه شده در جدول ۱، مجموعه داده های visDrone و UAVDT، DIOR و NWPU VHR-10، از نظر تعداد تصاویر، تنوع کلاس و نوع تصویربرداری برای استفاده در پژوهش این حوزه، مناسب هستند.

سپس باید روش‌های تقویت داده مورد بررسی قرار بگیرد و چنانچه برای افزایش دانش اولیه، نیاز به اعمال مواردی از قبیل چرخش، تغییر اندازه، افزایش نویز و ... وجود داشته باشد، در این فاز صورت خواهد گرفت. در مرحله بعد، برای پیاده‌سازی مواردی باید در نظر گرفته شود که به آن می‌پردازیم. برای حل مسئله تشخیص اشیاء در تصاویر هوایی با داده اندک، معمولاً پیاده‌سازی باید در دو فاز انجام شود. ۱) مرحله آموزش ۲) مرحله تشخیص

در مرحله آموزش، تمرکز بر روی دانش اولیه است و معمولاً آموزش از طریق دیتاست‌های بزرگ، برای تقویت مدل صورت می‌گیرد. به عبارتی، در این مرحله کلاس‌های پایه در آموزش اولیه مدل نقش مهمی ایفا می‌کنند. اگر تمرکز بر روی ترکیب مدل‌ها باشد، معمولاً در این فاز، از روش‌های meta learning، استفاده می‌شود و مدل روی یک مجموعه بزرگ از کلاس‌های پایه pre-train می شود. در این مرحله، مدل با استفاده از تصاویر اولیه، به استخراج ویژگی می‌پردازد.

در زیر به معرفی برخی از روش‌های موجود در meta learning می‌پردازیم:

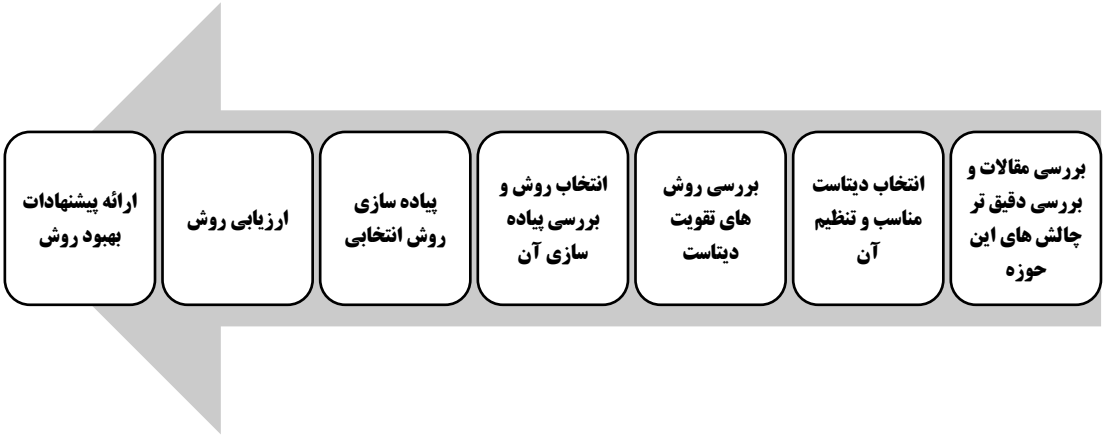
- MAML
- Meta-LSTM
- Meta-RCNN
- MANN
- Relation Network
- MatchingNet
- ProtoNet

در مرحله تشخیص، تعداد اندکی از نمونه‌های مجموعه‌های جدید، که با کلاس‌های پایه همپوشانی ندارند، برای تنظیم مدل استفاده می‌شوند تا مجموعه را با کلاس‌های جدید سازگار کنند. در روش ترکیبی، در این فاز از روش‌های transfer learning استفاده می شود.

در زیر به معرفی برخی از روش‌های موجود در transfer learning می‌پردازیم:

- RCNN
- Faster R-CNN
- YOLO
- SSD
- EfficientDet
- CenterNet
- Mask R-CNN

در آخرین مرحله، میزان دقت مدل مورد ارزیابی قرار می‌گیرد. در یک شمای کلی، مراحل زیر در روند کار بایستی صورت گیرد.



شکل ۲: شمایی از روند کلی پژوهش

۴-۳ ـ قلمرو تحقیق :

این پژوهش با تمرکز بر روی داده‌های مربوط به تصاویر هوایی انجام می‌شود و در این رابطه از الگوریتم‌های ارائه شده در حوزه یادگیری با تصاویر اندک، استفاده می‌شود. الگوریتم‌های متنوعی در حوزه یادگیری با داده‌های اندک ارائه شده‌اند، اما نکته قابل توجه این است که با توجه به نوع مسئله، الگوریتم‌ها می‌توانند کاملاً متفاوت عمل کنند. علاوه بر الگوریتم‌های اصلی استفاده شده، یکی از موارد مهمی که در مطالعات حاضر، مورد بررسی قرار گرفته، وظایفی از قبیل استخراج ویژگی و تعیین محدوده تصاویر در وظایف مرتبط با تشخیص اشیاء است. به طور کلی، تشخیص اشیاء سابقه مطالعاتی طولانی‌تری نسبت به یادگیری با داده‌های اندک دارد. اما

هر دو مفهوم، با توجه به پیشرفت تکنولوژی در تصویربرداری هوایی، و چالش‌های موجود در استخراج داده، اخیراً در این حوزه مورد توجه قرار گرفته‌اند.

قلمرو موضوعی: این پژوهش، در حوزه تشخیص اشیاء در تصاویر هوایی با داده‌های اندک با استفاده از الگوریتم‌های ارائه شده در هوش مصنوعی، ارائه می‌گردد.

قلمرو مکانی: این پژوهش در دانشگاه صنعتی مالک اشتر انجام شده و به بررسی فعالیت‌های انجام شده در حوزه تشخیص اشیاء در تصاویر هوایی با داده اندک، در سطح داخلی و خارجی می‌پردازد.

قلمرو زمانی: مطالعات انجام شده در این زمینه، در بازه زمانی سال‌های ۲۰۱۵ تا ۲۰۲۱، مورد بررسی قرار می‌گیرد.

۴-۴ ـ روش‌های مورد نظر برای تجزیه و تحلیل اطلاعات و آزمون فرضیه‌ها :

عملکرد هر مدل، براساس سه معیار قابل سنجش است که شامل IoU (Intersection over union)، mAP (mean Average Precision) و زمان استنتاج با واحد فریم در ثانیه (fps) ارزیابی می شود. در زمینه تشخیص اشیاء، معیار IoU، میزان ناحیه مشترک بین محدوده پیش‌بینی شده برای شی و منطقه حقیقی آن را در تصویر معین می‌کند. معیار دقت (P)، معیاری است که به طور گسترده در حوزه تشخیص اشیاء در ارزیابی استفاده می‌شود و به صورت نسبت مثبت‌های واقعی در بین تمام موارد شناسایی شده تعریف می‌شود. مقدار میانگین دقت، برای هر دسته از اشیاء محاسبه شده در تمام IoU های احتمالی، تحت عنوان mAP مشخص می‌شود. زمان استنتاج، نرخی است که یک آشکارساز اشیاء، طبق آن قادر به پردازش فریم‌های دوربین ورودی است. این معیار به صورت تعداد فریم در ثانیه محاسبه می‌شود.

در این پژوهش، از معیار mAP برای ارزیابی عملکرد سیستم استفاده می‌شود.

۵- نیازهای تحقیق :

۵-۱ ـ هزینه انجام تحقیق :

باتوجه به وجود دیتا ست‌های تصاویر هوایی، نیاز به بررسی بیشتر این مجموعه داده‌ها از نظر نوع حاشیه نویسی، تعداد کلاس و ... وجود دارد. بنابراین، نیازی به تهیه دادگان مجزا وجود ندارد، اما هزینه‌های زمانی در خصوص مطالعه و در صورت نیاز پیش پردازش داده‌های موجود وجود دارد.

۵-۲ ـ امکانات لازم (سخت‌افزاری ـ نرم‌افزاری و منابع و ...) :

برای پیاده سازی روش‌های مربوط به آموزش اولیه مدل، دریافت داده‌های تصاویر هوایی، با توجه به حجم پردازش‌های اولیه، نیاز به سخت‌افزار قوی می‌باشد.

۶ ـ حداکثر زمان موردنیاز برای انجام تحقیق :

زمان تخمین زده شده برای انجام این پژوهش بین ۱۰ تا ۱۲ ماه تخمین زده شده است که جزئیات در نمودار زیر قابل مشاهده است.

*** کلیهٔ صفحات به صورت تایپ شده تحویل شوند.**

۷ ـ زمان‌بندی تحقیق (نمودار گانت):

زمان موردنیاز مراحل تحقیق	ماه اول	ماه دوم	ماه سوم	ماه چهارم	ماه پنجم	ماه ششم	ماه هفتم	ماه هشتم	ماه نهم	ماه دهم	ماه یازدهم	ماه دوازدهم
بررسی دیتاست های موجود و پیش پردازش آن	*											
ارزیابی روش های موجود	*	*										
پیاده سازی روش های پایه روی دیتاست منتخب			*	*								
طرح و پیاده سازی ایده های جدید					*	*						
مقایسه و ارزیابی ایده های مطرح شده و روش های پایه							*	*				
ثبت نتایج، تکمیل گزارش پایان نامه									*	*		
بازنگری و تکمیل نهایی											*	*

۸ ـ فهرست منابع و مأخذ :

1. Prasad, D. K. (2012). Survey of The Problem of Object Detection In Real Images. International Journal of Image Processing (IJIP). 6. 441.

2.Liu, L. Ouyang, W. Wang, X. et al. Deep Learning for Generic Object Detection: A Survey. Int J Comput. Vis 128, 261–318 (2020). https://doi.org/10.1007/s11263-019-01247-4

3. Cheng, G. Han, J. (2016). A Survey on Object Detection in Optical Remote Sensing Images. ISPRS Journal of Photogrammetry and Remote Sensing. 117. 11-28. 10.1016/j.isprsjprs.2016.03.014.

4. Tayara, H. Chong, K. (2018). Object Detection in Very High-Resolution Aerial Images Using One-Stage Densely Connected Feature Pyramid Network. Sensors. 18. 3341. 10.3390/s18103341.

5. Wang, Y. Yao, Q. Kwok, J. Ni, L. (2020). Generalizing from a Few Examples: A Survey on Few-shot Learning. ACM Computing Surveys. 53. 1-34. 10.1145/3386252.

6. Zamir, S. W. Arora, A. Gupta, A. Khan, S. Sun, G. Khan, F. S. Zhu, F. Shao, L. Xia, G.-S. Bai, X. (2019). iSAID: A Large-scale Dataset for Instance Segmentation in Aerial Images. CVPR'19 Workshops (Detecting Objects in Aerial Images). http://arxiv.org/abs/1905.12886

7. Lam, D. Kuzma, R. McGee, K. Dooley, S. Laielli, M. Klaric, M. Bulatov, Y. McCord, B. (2018). xView: Objects in Context in Overhead Imagery. ArXiv:1802.07856 [Cs]. http://arxiv.org/abs/1802.0785.

8. Spacenet.ai – Accelerating Geospatial Machine Learning. (n.d.). Retrieved March 6, 2021, from https://spacenet.ai/

9. Mundhenk, T. Konjevod, G. Sakla, W. Boakye, K. (2016). A Large Contextual Dataset for Classification, Detection and Counting of Cars with Deep Learning. 9907. 10.1007/978-3-319-46487-9_48.

10. Cheng, G. Zhou, P. Han, J. (2016). Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images. in IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 12, pp. 7405-7415, doi: 10.1109/TGRS.2016.2601622.

11. Li, K & Wan, G. Cheng, G. Meng, L. Han, J. (2020). Object detection in optical remote sensing images: A survey and a new benchmark. ISPRS Journal of Photogrammetry and Remote Sensing. 159. 296-307. 10.1016/j.isprsjprs.2019.11.023.

12. Xia, G. Bai, Xi. Zhang, L. Belongie, S. Luo, J. Datcu, M. Pelilo, M. (2018). DOTA: A Large-Scale Dataset for Object Detection in Aerial Images. 10.1109/CVPR.2018.00418.

13. Robicquet, A. Sadeghian, A. Alahi, A. Savarese, S. (2016). Learning Social Etiquette: Human Trajectory Understanding In Crowded Scenes. 9912. 549-565. 10.1007/978-3-319-46484-8_33.

14. Airbus Ship Detection Challenge | Kaggle. (n.d.). Retrieved March 6, 2021, from https://www.kaggle.com/c/airbus-ship-detection

15. Yu, H. Li, G. Zhang, W. Du, D. Tian, Qi. Sebe, N. (2020). The Unmanned Aerial Vehicle Benchmark: Object Detection, Tracking and Baseline. International Journal of Computer Vision. 128. 10.1007/s11263-019-01266-1.

16. Zhu, P. Wen, L. Du, D. Bian, Xi. Hu, Q. Ling, H. (2020). Vision Meets Drones: Past, Present and Future.

17. Xia, G. Hu, J. Hu, F. Shi, B. Bai, Xi. Zhong, Y. Lu, Xi. Zhang, Li. (2017). AID: A Benchmark Data Set for Performance Evaluation of Aerial Scene Classification. IEEE Transactions on Geoscience and Remote Sensing. 55. 3965 - 3981. 10.1109/TGRS.2017.2685945.

18. Helber, P. Bischke, B. Dengel, A. Borth, D. (2017). EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification. 10.1109/JSTARS.2019.2918242.

19. Basu, Saikat & ganguly, sangram & Mukhopadhyay, Supratik & Dibiano, Robert & Karki, Manohar & Nemani, Ramakrishna. (2015). DeepSat: a learning framework for satellite imagery. 1-10. 10.1145/2820783.2820816.

20. Kundid Vasić, M. Papić, V. (2020). Multimodel Deep Learning for Person Detection in Aerial Images. Electronics Journal. Volume 9.Issue 9. 10.3390/electronics9091459

21. Pang, J., Li, C., Shi, J., Xu, Z., & Feng, H. (2019). R2-CNN : Fast Tiny Object Detection in Large-scale Remote Sensing Images. arXiv Comput. Vis. Pattern Recognit.2019,1902, 06042.

22. Ophoff, T. Puttemans, S. Kalogirou, V. Robin, J. Goedemé, T. (2020). Vehicle and Vessel Detection on Satellite Imagery: A Comparative Study on Single-Shot Detectors. Remote Sensing. 12. 1217. 10.3390/rs12071217.

23. Acatay, O. Sommer, L. Schumann, A. & Beyerer, J. (2018). Comprehensive Evaluation of Deep Learning based Detection Methods for Vehicle Detection in Aerial Imagery. 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). doi:10.1109/avss.2018.8639127

24. Koga, Y. Miyazaki, H. Shibasaki, R. (2020). A Method for Vehicle Detection in High-Resolution Satellite Images that Uses a Region-Based Object Detector and Unsupervised Domain Adaptation. Remote Sensing. 12. 575. 10.3390/rs12030575.

25. Shen, J. Liu, N. Sun, H. Zhou, H. (2019). Vehicle Detection in Aerial Images Based on Lightweight Deep Convolutional Network and Generative Adversarial Network. IEEE Access. PP. 1-1. 10.1109/ACCESS.2019.2947143.

26. Alganci, U. Soydas, M. Sertel, E. (2020). Comparative Research on Deep Learning Approaches for Airplane Detection from Very High-Resolution Satellite Images. Remote Sensing. 12. 458. 10.3390/rs12030458.

27. Ding, J. Xue, N. Long, Y. Xia, G. Lu, Q. (2019). Learning RoI Transformer for Oriented Object Detection in Aerial Images. 2844-2853. 10.1109/CVPR.2019.00296.

28. Azimi, S. M., Vig, E., Bahmanyar, R., Körner, M., & Reinartz, P. (2018). Towards Multi-class Object Detection in Unconstrained Remote Sensing Imagery. Computer Vision - ACCV 2018 - 14th Asian Conference on Computer Vision, Perth, Australia, December 2-6, 2018, Revised Selected Papers, Part III, 150–165. https://doi.org/10.1007/978-3-030-20893-6_10

29. Ren, Y. Zhu, Ch. Xiao, Sh. (2018). Deformable Faster R-CNN with Aggregating Multi-Layer Features for Partially Occluded Object Detection in Optical Remote Sensing Images. Remote Sensing. 10. 1470. 10.3390/rs10091470.

30. Guo, W. Yang, W. Zhang, H. Hua, G. (2018). Geospatial Object Detection in High Resolution Satellite Images Based on Multi-Scale Convolutional Neural Network. Remote Sensing. 10. 131. 10.3390/rs10010131.

31. Mandal, M. Shah, M. Meena, P. Devi, S. Vipparthi, S. (2019). AVDNet: A Small-Sized Vehicle Detection Network for Aerial Visual Data. Project: Vision Intelligence Lab.

32. Qin, R. Liu, Q. Gao, G. Huang, D. Wang, Y. (2020). MRDet: A Multi-Head Network for Accurate Oriented Object Detection in Aerial Images. ArXiv.

33. Kumdakci, H. Ongun, C. Temizel, A. (2021). Generative Data Augmentation for Vehicle Detection in Aerial Images. 10.1007/978-3-030-68793-9_2. Workshop on Analysis of Aerial Motion Imagery (WAAMI 2020) in 25th International Conference on Pattern Recognition (ICPR 2020).

34. Yi, J. Wu, P. Liu, B. Huang, Q. Qu, H. Metaxas, D. (2020). Oriented Object Detection in Aerial Images with Box Boundary-Aware Vectors. Accepted to WACV2021.

35. Xu, Y. Fu, M. Wang, Q. Wang, Y. Chen, K. Xia, G. Bai, X. (2019). Gliding vertex on the horizontal bounding box for multi-oriented object detection. Accepted by TPAMI 2020.

36. Ming, Q. Zhou, Zh. Miao, L. Zhang, H. Li, L. (2020). Dynamic Anchor Learning for Arbitrary-Oriented Object Detection (AAAI2021).

37. Han, J. Ding, J. Li, J. Xia, G.-S. (2021). Align Deep Features for Oriented Object Detection. ArXiv:2008.09397 [Cs]. http://arxiv.org/abs/2008.09397

38. Ding, J. Xue, N. Xia, G.-S. Bai, X. Yang, W. Yang, M. Y. Belongie, S. Luo, J. Datcu, M. Pelillo, M. Zhang, L. (2021). Object Detection in Aerial Images: A Large-Scale Benchmark and Challenges. ArXiv:2102.12219 [Cs]. http://arxiv.org/abs/2102.12219

39. Zou, Z. Shi, Z. Guo, Y. Ye, J. (2019). Object Detection in 20 Years: A Survey. ArXiv,abs/1905.05055.

40. Kolouri, S. Eaton, E. Kim, K. (2019). Deep Transfer Learning for Few-shot SAR Image Classification. 10.20944/preprints201905.0030.v1.

41. Kadam, S. Vaidya, V. (2019). Review and Analysis of Zero, One and Few Shot Learning Approaches. 10.1007/978-3-030-16657-1_10.

42. Wang, X. Huang, T. E. Darrell, T. Gonzalez, J. E. Yu, F. (2020). Frustratingly Simple Few-Shot Object Detection. ArXiv:2003.06957 [Cs]. http://arxiv.org/abs/2003.06957.

43. Kang, B. Liu, Zh. Wang, X. Yu, F. Feng, J. Darrell, T. (2019). Few-Shot Object Detection via Feature Reweighting. ICCV 2019. 8419-8428. 10.1109/ICCV.2019.00851.

44. Fan, Q. Zhuo, W. Tai, Y-W. (2019). Few-Shot Object Detection with Attention-RPN and Multi-Relation Detector. CVPR2020 Camera Ready. ArXiv:1908.01998 [Cs]. http://arxiv.org/abs/1908.01998.

45. Yan, X. Chen, Z. Xu, A. Wang, X. Liang, X. Lin, L. (2019). Meta R-CNN: Towards General Solver for Instance-Level Low-Shot Learning. 9576-9585. 10.1109/ICCV.2019.00967.

46. Xiao, Y. Marlet, R. (2020). Few-Shot Object Detection and Viewpoint Estimation for Objects in the Wild. Accepted as Poster at ECCV 2020.

47. Chen, H. Wang, Y. Wang, G. Qiao, Y. (2018). LSTD: A Low-Shot Transfer Detector for Object Detection. Accepted by AAAI2018.

48. Guo, Y. Codella, N. Karlinsky, L. Codella, J. Saenko, K. Rosing, T. Feris, R. (2020). A Broader Study of Cross-Domain Few-Shot Learning. 10.1007/978-3-030-58583-9_8.

49. Deng, J. Li, X. Fang, Y. (2020). Few-shot Object Detection on Remote Sensing Images. *ArXiv:2006.07826 [Cs]*. http://arxiv.org/abs/2006.07826.

50. Xiao, Z. Xue, W. Zhong, P. (2020). Few-shot Object Detection with Self-adaptive Attention Network for Remote Sensing Images. ArXiv:2009.12596 [Cs]. http://arxiv.org/abs/2009.12596.

* کلیهٔ صفحات به صورت تایپ شده تحویل شوند.

۹ـ نظریه شورای تحصیلات تکمیلی مجتمع دانشگاهی برق و کامپیوتر

(دقت شود که این صفحه به طور مجزا تهیه و تکمیل شود)

پایان نامه دانشجو خانم/ آقای فاطمه صالح نیا

با موضوع پایان‌نامه تشخیص اشیاء در تصاویر هوایی با داده های اندک به راهنمایی آقای دکتر علی مهجور و آقای دکتر کیوان راد و استاد مشاوری در شورای گروه علمی..... مورخ ..... مورد بررسی قرار و به تصویب رسید.

ردیف	نام و نام خانوادگی اعضاء گروه علمی	محل امضاء
۱		
۲		
۳		
۴		
۵		
۶		
۷		
۸		
۹		
۱۰		
۱۱		
۱۲		
۱۳		