



توسعه روش های یادگیری فعال عمیق برای طبقه بندی تصاویر سنجش از دوری

سجاد روشن میمند



استاد راهنما اول : دکتر سعید نیازمردی
استاد راهنمای دوم : دکتر علی اسماعیلی



Development of active deep learning methods for remotely sensed data classification

Sajjad Roshan Meymandi

Highlights

- Accurate remotely sensed data classification with less training data by active learning for when we do not have sufficient training data
- Recent methods can develop and make class prediction more efficient
- Active-learning can use for CNNs in the image classification task

First supervisor: Dr. Saeid Niazmardi
Second supervisor: Dr. Ali Esmaeeli

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



وزارت علوم، تحقیقات و فناوری
دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته
دانشکده عمران
گروه عمران-نقشه برداری

توسعه روش‌های یادگیری فعال عمیق برای طبقه‌بندی تصاویر سنجش از دوری

پایان‌نامه کارشناسی ارشد نقشه برداری - سنجش از دور

سجاد روشن میمندی

اساتید راهنما:

دکتر سعید نیازمردی

دکتر علی اسماعیلی

شهریور ۱۴۰۰



وزارت علوم تحقیقات و فناوری
دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته
دانشکده مهندسی عمران و نقشه برداری
گروه مهندسی نقشه برداری

آقای سجاد روشن میمندی به شماره دانشجویی 97100186، پایان نامه خود را با عنوان ((توسعه روش های یادگیری فعال عمیق برای طبقه بندی تصاویر سنجش از دوری)) در تاریخ 1400/06/20، به منظور یکی از شرایط احراز درجه کارشناسی ارشد ارائه و اعضا هیات داوران آن را با درجه علمی و نموده تایید نموده و پذیرش آن را برای تکمیل درجه کارشناسی ارشد پیشنهاد می نمایند.

اعضاء هیات داوران	نام و نام خانوادگی	رتبه علمی	امضاء
1- استاد راهنمای اول	دکتر سعید نیازمردی	استادیار	
2- استاد راهنمای دوم	دکتر علی اسماعیلی	استادیار	
3- استاد مشاور اول			
4- استاد مشاور دوم			
5- استاد داور	دکتر رضا حسن زاده	استادیار	
6- استاد داور	دکتر عصمت راشدی	دانشیار	
8- نماینده	دکتر فرشاد همائی	استادیار	

دانشکده/پژوهشکده

حق چاپ محفوظ و متعلق به دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته می باشد.

تعهدنامه اصالت اثر و رعایت حقوق مادی و معنوی

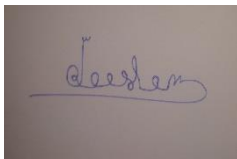
اینجانب سجاد روشن میمندی تعهد می‌نمایم که:

- ۱- مطالب مندرج در این پایان‌نامه، حاصل کار پژوهشی اینجانب است و به دستاوردهای پژوهشی دیگران که در این نوشته از آن استفاده شده است، ارجاع گردیده است.
- ۲- این پایان‌نامه از نظر محتوایی همانندجویی شده و دارای حداقل در صد همانندی مورد تایید دانشگاه می‌باشد و پیش از این برای احراز هیچ مدرکی ارائه نشده است.
- ۳- انتشار مقاله یا مقالات مستخرج از پایان‌نامه به صورت چاپ در نشریات علمی و یا ارائه در مجامع علمی به نام دانشگاه بوده و استاد راهنما مسئول مکاتبات مقاله باشد.
- ۴- انتشار کتاب حاصل از نتایج پایان‌نامه با مجوز کتبی صادره از طریق حوزه پژوهشی و مطابق با ضوابط دانشگاه انجام شود.
- ۵- ثبت اختراع و تدوین دانش فنی حاصل از نتایج پایان‌نامه با هماهنگی استاد راهنما، از طریق حوزه پژوهشی و مطابق با ضوابط دانشگاه انجام گیرد.

نام و نام خانوادگی دانشجو:

سجاد روشن میمندی

امضاء:



کلیه حقوق مادی و معنوی این اثر متعلق به دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته می‌باشد.

چکیده

امروزه علم رایانه و پیشرفت‌های روز افزون آن بر سایر علم‌های مرتبط از جمله سنجش از دور اثر بسزایی دارد. طبقه‌بندی تصاویر با استفاده از روش‌های یادگیری عمیق که از جمله مهمترین و پر توجه‌ترین قسمت‌های سنجش از دور است نیز با پیشرفت علم رایانه به سرعت پیشرفت می‌کند. یکی از مشکلات طبقه‌بندی تصاویر با روش‌های یادگیری عمیق کمبود داده‌های آموزشی است که یکی از راه‌حل‌های پیشنهادی برای حل مشکل کمبود داده استفاده از یادگیری فعال است. در این تحقیق سعی بر این شده تا با استفاده از یادگیری فعال با سایر روش‌های یادگیری عمیق مشکل کمبود داده این روش‌ها را جهت طبقه‌بندی تصاویر ماهواره‌ای برطرف کنیم. در واقع با توسعه دادن یادگیری فعال برای روش یادگیری عمیق Segnet که تا به حال انجام نشده است، به اثر یادگیری فعال بر این شبکه و صحت نتایج می‌پردازیم. داده‌های مورد استفاده در این تحقیق از ماهواره World view3 تهیه شده‌اند. هم چنین Ground truth این تصاویر نیز به صورت دستی در چهار کلاس آماده شده‌اند. پس از اجرای یادگیری فعال برای سه روش ANN، MLC، و DBN نتایج بهتری را ارائه کردند، همچنین روش پیشنهادی یادگیری فعال ارائه شده در این تحقیق که برای Segnet توسعه داده شد، نیز به بهترین نحو عمل کرده و نتایج یادگیری فعال دقت ۹۱٪ که از صحت Random sampling ۴٪ بالاتر است را ارائه کرد.

کلید واژه: یادگیری فعال، یادگیری عمیق، سنجش از دور، طبقه بندی تصویر، شبکه عصبی مصنوعی

فهرست مطالب

عنوان	صفحه
فصل اول: مقدمه	۱
۱-۱-مقدمه.....	۲
۲-۱-طرح مسأله و هدف.....	۳
۳-۱-اهداف و سؤالات تحقیق.....	۶
۴-۱-ساختار پایان نامه.....	۷
فصل دوم: پیشینه تحقیق	۸
۱-۲-طبقه بندی تصاویر در سنجش از دور.....	۹
۱-۱-۲- طبقه بندی تصاویر ویژگی‌های پیکسل.....	۹
۲-۱-۲- طبقه بندی بر اساس اجزاء تشکیل دهنده پیکسل.....	۹
۳-۱-۲- طبقه‌بندی‌های شیء گرا.....	۱۰
۲-۲- CNN ها.....	۱۰
۳-۲-Semantic segmentation.....	۱۲
۴-۲- یادگیری فعال.....	۱۲
۵-۲- یادگیری فعال در سنجش از دور.....	۱۵
فصل سوم: مواد و روش‌ها	۱۷
۱-۳-مقدمه.....	۱۸
۱-۱-۳-تعریف تصویر و پردازش‌های مربوط به صورت برداری.....	۱۸
۲-۱-۳-مواد و داده‌ها.....	۱۹

۲۱	۳-۱-۳-برنامه نویسی پایتون.....
۲۲	۳-۲-روش MLC.....
۲۲	۳-۲-۱-کلیات روش MLC.....
۲۴	۳-۲-۲-روش MLC با استفاده از یادگیری فعال.....
۲۵	۳-۳-روش ANN.....
۲۵	۳-۳-۱-کلیت شبکه‌های ANN.....
۲۹	۳-۳-۲-شبکه‌های ANN و یادگیری فعال.....
۳۰	۳-۴-روش DBN.....
۳۰	۳-۴-۱-کلیات روش DBN.....
۳۲	۳-۴-۲-روش DBN با یادگیری فعال.....
۳۶	۳-۵-روش CNN.....
۳۶	۳-۵-۱-روش segnet.....
۳۹	۳-۵-۲-یادگیری فعال و segnet.....
۴۴	فصل چهارم: نتایج و ارزیابی
۴۵	۴-۱-مقدمه.....
۴۵	۴-۲-روش MLC.....
۴۸	۴-۳-روش ANN.....
۵۲	۴-۴-روش DBN.....
۵۶	۴-۵-مقایسه روش‌های پیکسل مبنا.....
۶۰	۴-۶-روش Segnet.....
۶۸	فصل پنجم: نتیجه‌گیری

۶۹	۱-۵-مقدمه.....
۶۹	۲-۵-نتیجه‌گیری در رابطه با یادگیری فعال با روش MLC.....
۶۹	۳-۵-نتیجه‌گیری در رابطه با یادگیری فعال با روش ANN.....
۷۰	۴-۵-نتیجه‌گیری در رابطه با یادگیری فعال با روش DBN.....
۷۰	۵-۵-نتیجه‌گیری در رابطه با یادگیری فعال با روش Segnet.....
۷۱	۶-۵-پیشنهادهات.....
۷۳	منابع و مآخذ.....

فهرست شکل‌ها

صفحه	عنوان
۵	شکل ۱-۱- فضای احتمالات و نتیجه طبقه بندی توسط یک الگوریتم.....
۱۳	شکل ۱-۲- الگوریتم کلی یادگیری فعال.....
۲۰	شکل ۱-۳- تصویر منطقه مورد مطالعه.....
۲۱	شکل ۲-۳- راهنمای patch های اولیه برای train.....
	شکل ۳-۳- فلوجارت نحوه بدست آوردن بردار بیشترین قطعیت از بین کلاسها، فلش با رنگ
۲۴	آبی کمرنگ بیانگر حلقه تکرار است.....
۲۷	شکل ۴-۳- شبکه عصبی با یک لایه مخفی.....
۳۲	شکل ۵-۳- ساختار DBN.....
۳۷	شکل ۶-۳- معماری شبکه segnet و لایههای رمزگذاری و رمز گشایی.....
۳۸	شکل ۷-۳- تابع maxpooling (سمت چپ)، تابع unpooling (سمت راست).....
	شکل ۸-۳- خروجی لایه softmax، طبق شکل کلاس ۳ کلاسی است که پیکسل به آن تعلق
۴۱	میگیرد.....
	شکل ۹-۳- فلوجارت تحقیق - پیکان از دادههای pool به دادههای train به علت انتقال
۴۳	دادهها در هر تکرار به رنگ متفاوت کشیده شده است.....
	شکل ۱-۴- نتایج MLC، الف: Active Learning، ب: Ground truth، پ: Random sampling
۴۶
	شکل ۲-۴- نمودار دقت-تعداد دادههای آموزشی MLC، نمودار آبی مربوط به یادگیری فعال
۴۶	و نمودار قرمز مربوط به Random sampling با روش طبقه‌بندی MLC هستند.....
	شکل ۳-۴- نمودار دقت-زمان MLC، محور افقی زمان طی شده از شروع تا پایان یادگیری
۴۷	فعال و محور عمودی میزان دقت را نمایش میدهد.....
	شکل شماره ۴-۴- نتایج ANN، الف: Active Learning، ب: Ground truth، پ: Random
۴۹	sampling.....
	شکل ۵-۴- نمودار دقت-تعداد دادههای آموزشی ANN، نمودار نیلی مربوط به یادگیری فعال
۵۰	و نمودار نارنجی مربوط به Random sampling بار روش طبقه بندی ANN هستند.....

شکل ۴-۶- نمودار دقت-زمان ANN، محور افقی زمان طی شده از شروع تا پایان یادگیری فعال و محور عمودی میزان دقت را برای روش ANN نمایش میدهد.....	۵۱
شکل شماره ۴-۷- نتایج DBN، الف: Active Learning، ب: Ground truth، پ: Random sampling.....	۵۳
شکل ۴-۸- نمودار دقت-تعداد داده‌های آموزشی DBN، نمودار سبز مربوط به یادگیری فعال و نمودار دیگر مربوط به Random sampling بار روش طبقه بندی DBN هستند.....	۵۴
شکل ۴-۹- نمودار دقت-زمان DBN، محور افقی زمان طی شده از شروع تا پایان یادگیری فعال و محور عمودی میزان دقت را برای روش DBN نمایش میدهد.....	۵۵
شکل ۴-۱۰- نتایج کلی روشهای pixel-base، الف: DBN یادگیری فعال، ب: ANN یادگیری فعال، پ: MLC یادگیری فعال، ت: Image، ث: DBN Random sampling، ج: ANN Random sampling، چ: MLC Random sampling، ح: Ground truth.....	۵۷
شکل ۴-۱۱- نمودار دقت بدست آمده بر حسب تعداد داده‌های آموزشی با روشهای MLC، ANN و DBN در حالت‌های یادگیری فعال و Random sampling.....	۵۸
شکل ۴-۱۲- نمودار دقت بدست آمده بر حسب زمان صرف شده با روشهای MLC، ANN و DBN در حالت‌های یادگیری فعال و Random sampling.....	۵۹
شکل ۴-۱۳- قطعه‌تصویرهای آموزشی، در ابتدا این قطعه‌تصویرها به عنوان داده آموزشی استفاده شده و سپس به تعداد آنها در هر دور افزوده شده‌است، بالا (تصویری اصلی)، پایین (Ground truth).....	۶۰
شکل ۴-۱۴- نتایج Segnet، الف: Ground truth، ب: Image، پ: یادگیری فعال، ت: Random sampling.....	۶۲
شکل ۴-۱۵- نمودار دقت بر اساس تعداد قطعه تصویر آموزشی با شبکه Segnet.....	۶۳
شکل ۴-۱۶- دقت بدست آمده در دو حالت یادگیری فعال و Random sampling بر اساس زمان.....	۶۴
شکل ۴-۱۷- فلوجارت نتایج تحقیق.....	۶۷

فهرست جدول‌ها

صفحه	عنوان
۴۵	جدول ۴-۱- توزیع داده‌ها جهت ارزیابی و آموزش
۴۸	جدول ۴-۲- جدول نتایج MLC
۵۲	جدول ۴-۳- جدول نتایج ANN
۵۶	جدول ۴-۴- جدول نتایج DBN
۵۹	جدول ۴-۵- مقایسه نتایج روشهای Pixel base
۶۰	جدول ۴-۶- داده‌های مورد استفاده برای Segnet
۶۵	جدول ۴-۷- بالا : ماتریس confusion، بالا: Randm sampling، پایین: یادگیری فعال روش Segnet
۶۶	جدول ۴-۸- امتیازهای F1 بدست آمده برای هر کلاس در هر دو حالت Random sampling و یادگیری فعال
۶۶	جدول ۴-۹- جدول نتایج Segnet

فصل اول:

مقدمه

۱-۱- مقدمه

از زمان‌های بسیار دور بشر به دنبال راه‌های مختلفی برای نمایش و ثبت وقایع پیرامون خود بوده است. این وقایع می‌توانستند شامل خاطرات و اتفاقات خاص (تاریخ و سپس خط) و یا به منظورهای دیگر، از جمله ایجاد راهنمایی برای اشخاص دیگر در جهت استفاده بهینه از عوارض موجود در زمین باشند، این راهنمایی که جهت سهولت استفاده از عوارض به کار می‌رفت، رفته رفته به نقشه‌های امروزی تبدیل شد.

به وضوح می‌توان دید که بهترین دید برای تولید یک نقشه، دید از بالا به عوارضی است که می‌خواهیم از آن‌ها نقشه تهیه کنیم. با پیشرفت علم و تکنولوژی ابتدا بالون‌ها، سپس هواپیماها و در آخر ماهواره‌ها برای تصویربرداری جهت تولید نقشه‌ها به کار گرفته شدند، با توجه به این تفاسیر رابطه تنگاتنگی بین تصویربرداری از راه دور و نقشه‌ها وجود دارد.

تهیه نقشه‌ها به وسیله تصاویر هوایی یا ماهواره‌ای به دو طریق دستی و طبقه‌بندی انجام می‌شود، از مزیت‌های تهیه نقشه به روش دستی می‌توان به دقت بالای نقشه تهیه شده از روش دستی اشاره کرد. اما تهیه نقشه با استفاده از این روش بسیار وقت‌گیر است. این مشکل با روش دوم تهیه نقشه (طبقه‌بندی تصویر) حل می‌شود؛ اما معمولاً دقت روش اولیه را ندارد، طبقه‌بندی تصاویر با علمی به نام پردازش تصویر و غالباً به وسیله بینایی کامپیوتر^۱ انجام می‌شود، استخراج عوارض به صورت خودکار یکی از زیر شاخه‌های طبقه‌بندی و یکی از چالش‌های مهم پیش روی بینایی کامپیوتر است که در این پایان نامه به آن‌ها پرداخته می‌شود.

طبقه‌بندی به دو روش نظارت شده^۲ و بدون نظارت^۳ انجام می‌شود، تفاوت این دو روش در این است که روش‌هایی که در دسته اول قرار می‌گیرند به نمونه‌های آموزشی نیاز دارند؛ اما روش‌های دسته دوم، که به روش‌های خوشه‌بندی نیز شناخته می‌شوند، نیازی به نمونه‌های آموزشی ندارند [۱].

^۱ Computer vision

^۲ supervised

^۳ unsupervised

روش‌های بسیاری برای طبقه‌بندی نظارت شده تصاویر ارائه شده‌اند که می‌توان به روش‌های متداول‌تر مانند ^۱MLE و روش‌های جدید مانند ^۲ANN های عمیق و ^۳CNN ها اشاره کرد که از زیر مجموعه‌های یادگیری عمیق هستند [۲]، در کلیه این روش‌ها، نیاز به داده‌های آموزشی یکی از مشکلات پیش روی کاربران است.

هر چقدر که روش‌های طبقه‌بندی نظارت شده به روزتر می‌شدند، دقت و حساسیت آن‌ها به تفاوت کلاس‌ها نیز بالاتر می‌رفت، به علت حساسیت بالای روش‌های نوین به خصوص CNN ها داده‌های آموزشی نقش بسزایی در صحت و کارایی آن‌ها دارند [۳]، تهیه نمونه‌های آموزشی برای روش‌های طبقه‌بندی بسیار زمانبر و پرهزینه است، پس در اکثر اوقات برای آموزش با کمبود داده‌های آموزشی روبرو هستیم، علاوه بر این، اگر داده‌های زیادی نیز موجود باشند، با مشکل زمان زیاد آموزش در روش‌های یادگیری عمیق مواجه خواهیم بود [۴].

۱-۲- طرح مسأله و هدف

پس از بیان ایرادات مطرح شده در رابطه با طبقه‌بندی نظارت شده با روش‌های یادگیری عمیق، باید به دنبال این بود که نمونه‌های آموزشی را به گونه‌ای به شبکه معرفی کرد که با داده‌هایی کمتر و بهینه به دقت مورد نظر برای طبقه‌بندی دست یافت. همچنین باید در نظر داشته باشیم تا حدی که دقت طبقه‌بندی پایین نیاید، زمان طبقه بندی را نیز کاهش دهیم. برای پاسخ به مشکلات مطرح شده می‌توان این حالت موضوع را در نظر گرفت اگر ما سیستم طبقه بندی را مانند مغز یک انسان در نظر بگیریم و فرض بر این باشد که این انسان تا به حال هواپیما را از نزدیک ندیده و با ماهیت آن آشنایی ندارد، اگر انواع هواپیما را ۱۰۰ دسته در نظر بگیریم و در حالت اول ۱۰۰۰ تصویر از هواپیمای دسته اول به فرد نشان دهیم، وی با دیدن هواپیمایی از ۹۹ دسته دیگر ممکن است در تشخیص، دچار خطا شود. اما در حالت دوم از هر دسته هواپیما یک تصویر به وی نشان می‌دهیم که وی با دیدن هر تصویر جدید تشخیص می‌دهد که تصویر

^۱ Maximum likelihood estimation

^۲ Artificial neural network

^۳ Convolutional neural network

متعلق به هواپیما است پس با ۱۰۰ تصویر به جای ۱۰۰۰ تصویر آموزش انجام شده و خطا نیز کمتر شده است. این مثال نشان داد که با داده‌های آموزشی کمتر به شرط بهینه بودن آن‌ها می‌توان آموزش را به نحوه بهتری انجام داد.

اکنون که ماهیت مسأله با مثال مشخص شد، به سراغ این مهم خواهیم رفت که چگونه این مسأله را با یادگیری ماشین^۱ حل کنیم، در واقع یادگیری ماشین و سایر الگوریتم‌های یادگیری نیز از نحوه یادگیری موجودات زنده الگوبرداری می‌کنند، در یادگیری ماشین ما باید مجموعه پیکسل‌هایی که به صورت درجات خاکستری هستند را به ماشین نشان داده و به ماشین بیاموزیم که هر مجموعه پیکسل به چه دسته‌ای تعلق دارد. بر اساس انواع الگوریتم یادگیری ماشین، ما نتایجی را بدست می‌آوریم که این نتیجه‌ها گاهی از نظر دسته‌ای که به آن تعلق گرفته‌اند، قطعیت^۲ داشته و گاهی نیز با عدم قطعیت^۳ مواجه هستند، ما در اینجا به دنبال راهی هستیم تا داده‌هایی که با عدم قطعیت بیشتری همراه هستند را به عنوان داده‌های آموزشی جدید به الگوریتم معرفی کنیم.

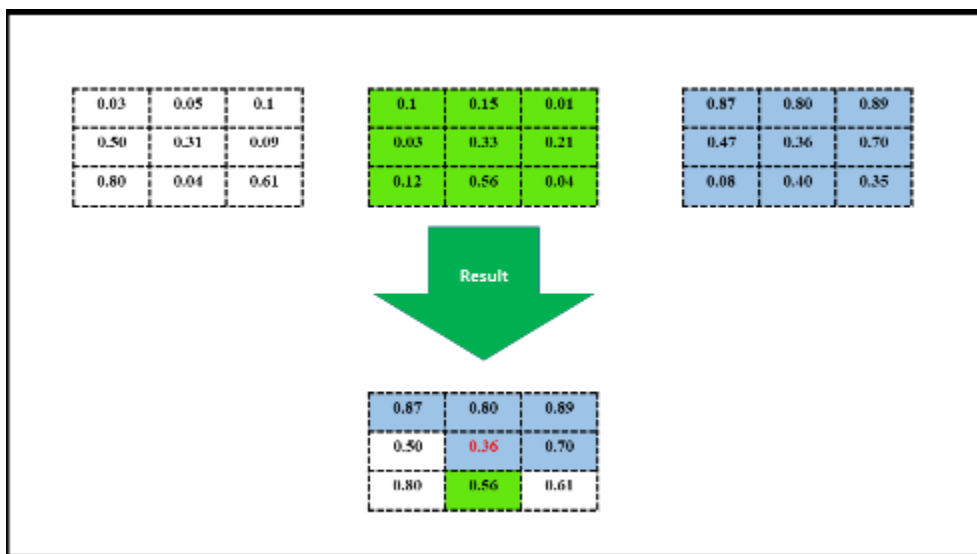
الگوریتم‌های زیادی برای طبقه‌بندی وجود دارند، اکثر الگوریتم‌هایی که برای طبقه‌بندی تصویر استفاده می‌شوند؛ زیر مجموعه یادگیری ماشین هستند، در ادامه برای روشن شدن ساز و کار این الگوریتم‌ها توضیحاتی در رابطه با یک الگوریتم فرضی داده خواهد شد.

شکل ۱-۱ را در نظر می‌گیریم، در این جدول طبقه‌بندی پیکسل‌های تصویر توسط یک الگوریتم فرضی دیده می‌شود، این شکل یک patch از یک تصویر فرضی را به ما نشان می‌دهد، کلاس‌های مورد نظر ما در طبقه بندی این تصویر سه کلاس هستند، پوشش گیاهی با رنگ سبز، ساختمان به رنگ آبی و زمین بایر به رنگ سفید طبقه بندی خواهند شد، برای هر سه کلاس الگوریتم یک patch احتمال به اندازه تصویر می‌سازد، اعداد هر patch نشانگر این هستند که آن پیکسل با چه احتمالی به آن کلاس تعلق دارد، نتیجه نهایی نیز بصورت یک patch در پایین شکل ۱-۱ آورده شده است.

^۱ Machine learning

^۲ certainty

^۳ uncertainty



شکل ۱-۱- فضای احتمالات و نتیجه طبقه بندی توسط یک الگوریتم

در شکل ۱-۱ مشاهده می‌شود، پیکسلی به کلاسی تعلق می‌گیرد که بیشترین احتمال حضور در آن را دارد، اما در مورد پیکسل‌هایی مانند پیکسل (۲،۲)، که با رنگ قرمز مشخص شده، احتمال تعلق به کلاس ساختمان برابر با ۰/۳۶ و اختلاف آن با تعلق به کلاس پوشش گیاهی که پس از این کلاس بالاترین احتمال را کسب کرده‌است، برابر با ۰/۰۳ است که اختلاف بسیار کمی‌است، در اصطلاح به این پیکسل، پیکسلی با عدم قطعیت بیشتر گفته می‌شود. اگر پیکسل-هایی که ویژگی خاصی مانند عدم قطعیت دارند را یافته و به داده‌های آموزشی الگوریتم اضافه کنیم، باید از یکی از توابع QUERY استفاده کنیم، اگر انتخاب داده‌ها بر اساس تابع عدم قطعیت باشد، می‌توان گفت که آموزش الگوریتم با استفاده از یکی از روش‌های QUERY در یادگیری فعال به نام نمونه برداری بر اساس عدم قطعیت^۱ انجام شده‌است.

یادگیری فعال در واقع یک الگوریتم به منظور بهبود آموزش به وسیله یک روش یادگیری ماشین است که با انتخاب داده‌های بهینه باعث بهبود آموزش می‌شود. یادگیری فعال می‌تواند

^۱ Uncertainty sampling

برای تمام الگوریتم‌های طبقه بندی نظارت شده، استفاده شود، با اعمال این الگوریتم می‌توان تعداد داده‌های آموزشی را به صورت بهینه انتخاب کرده و عملکرد الگوریتم اصلی را نیز بهبود بخشید، در مواقعی نیز می‌توان زمان آموزش را تا رسیدن به دقت مورد نظر کاهش داد. استفاده از یادگیری فعال در همه روش‌های طبقه‌بندی جایگاه خاص خود را دارد، زیرا تهیه داده‌های آموزشی برای اکثر روش‌ها بسیار زمانبر و پرهزینه است، در CNN ها نیز این مشکل به علت patch بودن داده‌ها بسیار بیشتر است. حتی این ایراد در انتخاب داده‌ها با یادگیری فعال نیز مشکلاتی ایجاد می‌کند که در این تحقیق ما به دنبال راهی جهت حل مشکلات استفاده از یادگیری فعال برای patch‌های مورد استفاده در CNN ها هستیم.

دانش سنجش از دور پیوسته با طبقه‌بندی تصاویر در ارتباط است. یکی از بهترین الگوریتم‌های مورد استفاده در طبقه‌بندی تصویر، الگوریتم CNN است، این الگوریتم کارآمد نتایج بسیار دقیقی را ارائه می‌کند. اما با مشکلاتی از جمله کمبود داده‌های آموزشی رو برو است، داده‌های این الگوریتم در بیشتر مواقع از نوع قطعه تصویر بوده و همین باعث سخت‌تر شدن تهیه داده‌ای آموزشی برای CNN ها شده‌است [۵].

یکی دیگر از تفاوت‌های CNN با سایر الگوریتم‌ها، داده‌های آن است که بصورت قطعات تصویر هستند، پس عدم قطعیت گفته شده را نباید فقط درباره یک پیکسل در نظر گرفته بلکه باید عدم قطعیت برای کل قطعه (تمام پیکسل‌های موجود) تصویر محاسبه شود، همین مورد کار را برای ما کمی مشکل‌تر از به کار بردن یادگیری فعال برای الگوریتم‌های دیگر کرده‌است.

استفاده از یادگیری فعال بر روی یک CNN در تحقیقات سنجش از دوری تا به حال انجام نشده و همین نکته نیز تحقیق ما از سایر تحقیقات انجام شده در سنجش از دور متمایز می‌سازد.

۳-۱- اهداف و سؤالات تحقیق

هدف اصلی تحقیق پیش رو توسعه روش‌های یادگیری فعال عمیق به منظور طبقه‌بندی تصاویر سنجش از دور است.

سؤالاتی که در این پایان نامه پاسخ داده می‌شوند، به شرح زیر هستند:

۱- آیا یادگیری فعال می‌تواند راه حل مناسبی جهت انتخاب داده‌های آموزشی بهینه در

روش‌های طبقه بندی تصویر با یادگیری عمیق در سنجش از دور باشد؟

۲- آیا یادگیری فعال می‌تواند زمان طبقه بندی را کاهش دهد؟

هدف کلی تحقیق پاسخ به دو سؤال بالا برای روش‌های مختلف طبقه بندی تصاویر سنجش از دوری است، هدف دیگر که هدف خاص این پایان است، توسعه روش یادگیری فعال برای روش CNN از نوع Segnet است.

۴-۱- ساختار پایان نامه

این پایان نامه شامل شش فصل است که در فصل اول آن در مورد ضرورت انجام تحقیق، طرح مسأله و مشخص کردن هدف بحث شد. در فصل دوم درباره تحقیقات پیشین که به گونه‌ای با موضوع پایان‌نامه مرتبط هستند، بحث می‌شود. در فصل سوم مبانی نظری، شامل مباحثی است که در فصل‌های بعدی مورد استفاده قرار گرفته‌اند. در فصل چهارم درباره روش‌های یادگیری ماشین (اغلب یادگیری عمیق) می‌شود. فصل پنجم شامل نتایج و ارزیابی آن‌ها و فصل ششم نتیجه گیری از تحقیق پیش رو است.

فصل دوم:

پیشینه تحقیق

۱-۲- طبقه بندی تصاویر در سنجش از دور

طبقه‌بندی تصاویر را می‌توان بر اساس روش‌های مربوط به ساختار پیکسل‌ها، با سه روش کلی انجام داد که کلیه روش‌های طبقه‌بندی تصویر زیر مجموعه‌ای از این سه روش هستند، این سه روش، شامل طبقه‌بندی بر اساس ویژگی‌های پیکسل^۱، بر اساس اجزاء تشکیل دهنده یک پیکسل^۲ و یا براساس شیء گرایی^۳ هستند که در ادامه به توضیح و تاریخچه مختصری از این روش‌ها پرداخته می‌شود [۶].

۱-۱-۲- طبقه‌بندی تصاویر ویژگی‌های پیکسل

طبقه‌بندی بر اساس این روش در دو نوع Land-Cover و Land-Use ممکن است که در نهایت به هر پیکسل برجستگی بر اساس شبیه بودن به یک کلاس اختصاص می‌دهد [۷].
MLC^۴ یکی از پرکاربردترین روش‌های زیر مجموعه طبقه‌بندی بر اساس ویژگی‌های پیکسلی است که برای اولین بار در سال ۱۹۸۷ میلادی برای طبقه‌بندی داده‌های سنجش از دور به کار گرفته شد، این روش از روش‌های نظارت شده است و بر اساس قوانین احتمال طبقه‌بندی را انجام می‌دهد [۸]. یکی دیگر از روش‌های مورد استفاده در سنجش از دور روش KNN^۵ است که طبقه‌بندی تصاویر را از ویژگی‌های پیکسل‌های همسایه انجام می‌دهد [۹].

۲-۱-۲- طبقه بندی بر اساس اجزاء تشکیل دهنده پیکسل

در این روش هر پیکسل بصورت خالص در نظر گرفته نمی‌شود، به عنوان مثال هر پیکسل فقط شامل آب یا خاک نیست بلکه یک ترکیب^۶ از این دو کلاس است، این روش بیشتر برای داده‌های Hyperspectral کاربرد دارد و هر پیکسل را به بخش‌های کوچکتر تقسیم می‌کند [۱۰].

^۱Pixel-based techniques

^۲ Sub-pixel-based techniques

^۳ Object-based techniques

^۴ Maximum Likelihood Classifier

^۵ K-Nearest Neighbors Classifier

^۶ Mixed

۲-۱-۳- طبقه‌بندی‌های شیء گرا

این روش‌ها به جای قلمداد کردن تصویر به عنوان مجموعه‌ای از پیکسل‌ها، آن را به عنوان یک مجموعه از اشیاء^۱ در نظر گرفته و به قسمت‌های مختلف تقسیم می‌کنند^۲. این روش‌ها از همه ویژگی‌های تصویر اعم از بافتی^۳، مکانی^۴، نوعی^۵ و طیفی^۶ استفاده می‌کنند و طبقه‌بندی را اغلب با الگویی بر اساس مجموعه‌ای از پیکسل‌ها انجام می‌دهند [۱۱]. اکثر CNN^۷ ها نیز از زیر مجموعه‌های این روش طبقه بندی به حساب می‌آیند، زیرا طبقه بندی پیکسل‌ها را معمولاً با توجه به پیکسل‌های اطراف انجام می‌دهند.

۲-۲- CNN ها

این روش از روش‌های بسیار کارآمد و موفق در علم بینایی کامپیوتر^۸ است [۱۲] که از زیر مجموعه‌های شبکه‌های عصبی مصنوعی است. در سال ۱۹۵۹ Hubel و Wiesel یک مدل به منظور شناسایی الگو پیشنهاد کردند که با تکیه بر ساختارهایی مانند گوشه و ساختار اشیاء عمل می‌کرد، این مدل ساختاری مانند ساختار چشم موجودات زنده داشت [۱۳]، همین دو محقق در سال ۱۹۶۲ میلادی دریافتند که با اعمال فیلترهایی بر یک تصویر و جمع نتیجه حاصل از آن‌ها در قسمت‌هایی که الگویی مشخصی دارند، نتایج تقریباً یکسان است [۱۴]. در سال ۱۹۸۰ میلادی یک دانشمند که تحت تاثیر یافته‌های Hubel و Wiesel قرار گرفته بود، یک مدل یادگیری جدید با convolution به نام NC^۹ پیشنهاد کرد، اولین مدل جدید از CNN از مدل این دانشمند

^۱ Object

^۲ Image segmentation

^۳ textural

^۴ spatial

^۵ contextual

^۶ spectral

^۷ convolutional neural network

^۸ Computer vision

^۹ Neocognitron

نشأت گرفته شده است، این مدل توسط Yann LeCun و همکارانش ارائه شد، مدل آن‌ها از شیوه یادگیری Gradient-Base برای تشخیص دست خط (مجموعه داده ^۱ MNIST) استفاده کرده بود [۱۵]. سرانجام Alex Krizhevsky و همکاران توانستند شبکه‌ای^۲ به نام خود سازنده Alex-Net و با یک معماری پیچیده جهت برچسب گذاری تصاویر ارائه کنند که این شبکه فقط مشخص می‌کرد که آیا یک شیء در تصویر وجود دارد یا خیر؟ شبکه‌های دیگری نیز مانند VGG16^۳ [۱۶] و Google-net [۱۷] نیز با ساختاری مشابه و با تفاوت‌های جزئی نیز از دیگر CNN‌های مورد استفاده هستند.

با وجود اینکه در گذشته از یادگیری عمیق^۴ به منظورهای تلفیق تصاویر^۵، هم مرجع سازی^۶ و یافتن اشیاء^۷ استفاده می‌شد، امروزه بیشتر کاربرد یادگیری عمیق در سنجش از دور به طبقه بندی‌های Land-cover و Land-use متمرکز شده است [۱۷]. در سال ۲۰۱۵ در یک تحقیق با استفاده از DBN^۸ به طبقه بندی تصاویر SAR^۹ پرداخته شد که نتایج خوبی (دقت بالای ۸۵٪) نیز حاصل شد [۱۸]، در سال ۲۰۱۷، Sharma و همکاران یک CNN با ساختار patch-base پیشنهاد کردند که برای طبقه بندی تصاویری با اندازه پیکسل متوسط مناسب بود [۱۹]، این دو کار از کارهایی هستند که نتایج خوبی را برای طبقه بندی ارائه کرده‌اند.

از نمونه کارهای دیگر انجام شده در سنجش از دور به وسیله CNN می‌توان به تحقیقی در سال ۲۰۱۷ اشاره کرد که به استخراج همزمان راه‌ها و ساختمان‌ها می‌پرداخت و دقت‌هایی بالاتر از ۹۰٪ ارائه می‌کند [۲۰]. یا تحقیق دیگری در سال ۲۰۱۸ که به طبقه بندی تصاویر Hyperspectral پرداخته است نیز دقت‌های طبقه بندی همگی بالای ۹۰٪ بود [۲۱]. در تحقیقی

^۱ Modified National Institute of Standards and Technology

^۲ Alex-Net

^۳ Visual Geometry Group

^۴ Deep learning

^۵ image Fusion

^۶ Image Registration

^۷ Object Detection and Scene classification

^۸ Deep Belief Network

^۹ Synthetic-aperture radar

دیگر در همان سال ۲۰۱۸ نویسندگان به یافتن خودکار اشیا مختلف از جمله هواپیما در تصاویر اپتیکی و کشتی در تصاویر راداری پرداختند [۲]. نمونه‌هایی نیز موجود است که change detection در سنجش از دور با شبکه عصبی CNN انجام شده‌است و با تخمین‌های دقیق دقت-های بالای ۸۵٪ درصد را ارائه می‌کند [۲۲]، در تحقیقی نیز به استخراج عوارض از طریق semantic segmentation در پنج کلاس مختلف در تصاویر با قدرت تفکیک بالا پرداخته شده است که معمولاً با آموزش خوب، دقتی بالای ۸۵ درصد را ارائه می‌کند [۲۳].

۳-۲- Semantic segmentation

قطعه بندی تصاویر به معنای تقسیم یک تصویر به قطعه‌هایی با پیکسل‌هایی است که ویژگی یکسانی دارند، این نوع طبقه بندی به هر پیکسل یک بر اساس ساختار پیکسل‌های اطرافش یک بر حسب اطلاق می‌کند [۲۴]، این کار معمولاً با FCN^۱ ها که خود زیر مجموعه‌ای از CNNها هستند انجام می‌شود [۲۵]، یکی از مهم ترین مشکلات این شبکه تار شدن لبه‌های اشیا موجود در تصاویر است که چندان مربوط به بحث ما نیست، اما از جمله کارهایی که برای بهبود آن انجام شده‌است می‌توان به یک تحقیق مهم در ۲۰۱۶ اشاره کرد، که با استفاده از CNN عمیق به نتایج خوبی از نظر دقت طبقه‌بندی دست یافت [۲۶]. یکی از کارهای انجام شده در گذشته که شبیه به تحقیق پیش رو است، می‌توان به تحقیق Kemker در سال ۲۰۱۸ اشاره کرد.، در این تحقیق به جای تغییر در معماری و ساختار شبکه، فرآیند آموزش را بصورت دقیق‌تر و آموزش دوباره انجام داد [۲۷].

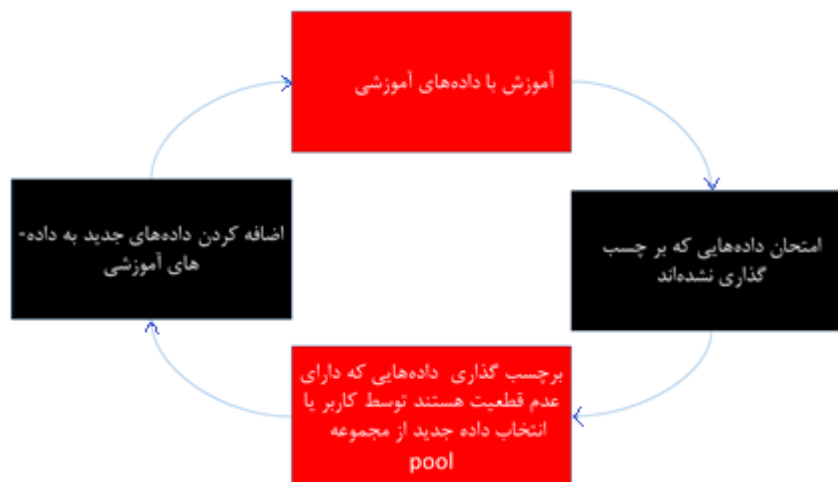
۴-۲- یادگیری فعال

یکی از مشکلات مهم شبکه‌های عصبی کمبود داده‌های آموزشی است، CNN ها نیز از این مشکل مستثنی نیستند، علاوه بر کمبود داده‌های آموزشی که در CNNها محسوس‌تر است [۴]، زمان لازم برای آموزش این شبکه‌ها نیز با افزایش داده‌های آموزشی افزایش می‌یابد، یادگیری

^۱ Fully Convolutional Network

فعال می‌تواند راه مناسبی برای تهیه داده‌های آموزشی بهینه باشد.

یادگیری فعال شاخه‌ای خاص از یادگیری ماشین است که از یک کاربر یا یک منبع اطلاعاتی داده‌های جدید آموزشی را دریافت کرده و فرآیند آموزش را تا رسیدن به معیار مورد نظر دوباره با داده‌های جدید انجام می‌دهد، در علم آمار گاهی به این عمل فرآیند طراحی بهینه تجربی^۱ گفته می‌شود [۲۸]. شکل ۱-۲ الگوریتم کلی یادگیری فعال را نشان می‌دهد.



شکل ۱-۲- الگوریتم کلی یادگیری فعال

در ادامه به اختصار درباره بعضی از زیر مجموعه‌های الگوریتم یادگیری فعال که با استفاده از انواع queryها (توابع مورد استفاده در یادگیری فعال) برای اضافه نمودن داده‌های جدید به داده‌های آموزشی خواهیم پرداخت.

Balance exploration and exploitation: این الگوریتم که در سال ۲۰۱۴ ارائه شده است، یک فرآیند تکرار به نام ATS^۲ در نظر می‌گیرد که بر اساس چگالی (تعداد یک نوع داده نسبت

^۱ - Optimal experimental design

^۲ Active Thompson Sampling

به کل داده‌ها) داده‌ها، داده‌های جدید را از مجموعه داده pool انتخاب کرده و به داده‌های آموزشی اضافه می‌کند [۲۹].

Expected model change: این الگوریتم داده‌هایی را انتخاب می‌کند که بیشترین تغییر را در عملکرد مدل فعلی ایجاد می‌کنند.

Expected error reduction: با این الگوریتم داده‌هایی انتخاب می‌شوند که خطای تعمیم^۱ مدل را به کمترین حد ممکن برسانند.

Exponentiated Gradient Exploration for Active Learning: این الگوریتم راهکاری به نام EG-active ارائه می‌کند که می‌تواند هر الگوریتم یادگیری فعال را بهینه سازد.

Uncertainty sampling: داده‌هایی که کمترین قطعیت را دارند، جهت بر چسب‌گذاری به داده‌های آموزشی قبلی اضافه می‌کند. در تحقیق پیش رو نیز از همین روش Query استفاده شده‌است.

Query by committee: در این الگوریتم چند مدل آموزش داده شده و داده‌ای انتخاب می‌شود که بیشترین اختلاف بر چسب‌گذاری را در بین مدل‌ها متفاوت باشد.

Querying from diverse subspaces or partitions: این حالت داده‌هایی را از یک مجموعه داده برای یک جنگل تصادفی انتخاب می‌کند که کمترین هم پوشانی را ما بین نتایج بر چسب‌گذاری کل درخت‌های تشکیل دهنده جنگل تصادفی داشته باشند.

Variance reduction: داده‌هایی را علامت‌گذاری می‌کند که واریانس بین خروجی‌ها را به حداقل برساند.

Mismatch-first farthest-traversal: با استفاده از الگوریتم طبقه‌بندی نزدیکترین همسایه^۲ داده‌هایی که بیشترین فاصله با همسایه‌ها دارند را به عنوان داده‌هایی که باید بر چسب‌گذاری شوند، انتخاب می‌کند [۳۰].

^۱ Generalization error

^۲ -Nearest neighbour

۲-۵- یادگیری فعال در سنجش از دور

از جمله کارهایی که با یادگیری فعال و در زمینه سنجش از دور انجام شده می‌توان به تحقیقی در زمینه قطعه بندی تصاویر چند طیفی اشاره کرد که به وسیله SVM^۱ و یادگیری فعال، تصاویر ماهواره IRS-1A در چهار باند طیفی، با تعداد پیکسل‌های کمتر اما در مدت زمان بیشتری با دقت‌های بالای ۸۷٪ درصد برای تصاویر مختلف انجام شد [۳۱]. از نمونه‌های دیگر که با یادگیری فعال در زمینه سنجش از دور انجام شده است، می‌توان به مقاله روش‌های یادگیری فعال برای طبقه بندی تصاویر ماهواره‌ای اشاره کرد که با آموزش SVM با روش یادگیری فعال که در هر دور یادگیری فعال یک داده (پیکسل) به داده‌های آموزشی قبل اضافه می‌کند به دقت میانگین بیشتری برای دقت طبقه‌بندی تصویر دست یافت [۳۲].

یادگیری فعال batch-mode از جمله روش‌های متداول یادگیری فعال در سنجش از دور است، که به انتخاب تعدادی داده به جای یک داده در هر تکرار برای SVM می‌پردازد، استفاده از این روش موجب شده تا زمان کمتری صرف یادگیری فعال شود، همچنین دقت بدست آمده نیز ۳٪ بالاتر یعنی از ۸۷٪ (انتخاب یک داده در هر دور) به بالای ۹۰٪ (در حالت انتخاب چند داده در هر دور) رسیده است [۳۳]. در سال ۲۰۱۳ در مقاله‌ای در رابطه با تصاویر Hyperspectral، به حل مشکل افزونگی داده و استفاده از روش‌های کاهش ابعاد مانند PCA پرداخته و در طبقه‌بندی تصاویر Hyperspectral و برای طبقه‌بندی تصاویر از یادگیری فعال استفاده شده‌است، در نهایت دقت‌ها نیز بالای ۸۷٪ بدست آمده‌اند [۳۴].

تحقیق دیگری نیز از ترکیب انتخاب ویژگی تصاویر و کاندیدا شدن برای افزوده شدن به داده‌های آموزشی در هر تکرار یادگیری فعال استفاده کرده است [۳۵]، یکی دیگر از کارهایی که با استفاده از یادگیری فعال و ANN در زمینه سنجش از دور استفاده شده‌است و نتایج آن نیز با استفاده از داده‌های کمتر، دقت‌هایی حدود ۹۰٪ را ارائه کرده است، می‌توان به تحقیق یادگیری فعال برای طبقه بندی داده‌های Hyperspectral اشاره کرد که یادگیری فعال را برای یک شبکه

^۱ Support vector machine

^۱DBN را با استفاده از توابعی برای بدست آوردن احتمال در هرلایه و اندازه‌گیری شباهت هر پیکسل با پیکسل‌های دیگر به کار برده‌است و به نتایج قابل قبولی (دقت بالای ۹۲٪) دست یافته است [۳۶].

در فصل آخر نیز به یک نتیجه‌گیری کلی در رابطه با استفاده از یادگیری فعال در سنجش از دور خواهیم رسید.

^۱ Deep belief network

فصل سوم:

مواد و روش‌ها

۳-۱-مقدمه

در این فصل در ابتدا به توضیحی مختصر درباره روش‌های مورد استفاده برای طبقه‌بندی تصویر خواهیم پرداخت، سپس نحوه استفاده از یادگیری فعال برای هر روش جداگانه توضیح داده خواهد شد، ساختار روش‌ها در این فصل به این گونه است که روش‌های قدیمی‌تر در ابتدای فصل و با جلوتر رفتن فصل، رفته رفته روش‌ها جدیدتر می‌شوند. در انتهای فصل به معرفی روش پیشنهادی می‌پردازیم که برای استفاده از یادگیری فعال در CNN ها به کار گرفته شده است.

۳-۱-۱-تعریف تصویر و پردازش‌های مربوط به صورت برداری

اگر مجموعه $S = \{1, 2, 3, 4, 5 \dots N\}$ را مجموعه اندیس‌های پیکسل‌های تصویری با N پیکسل با m باند و مجموعه $L = \{[1, 2, 3, 4, 5 \dots, C]\}$ را مجموعه تعداد کلاس‌هایی بدانیم که پیکسل‌های تصویر متعلق به این کلاس‌ها هستند، در نظر بگیریم، تصویر را بصورت مجموعه بردار $X = \{x_1, x_2, \dots, x_i, \dots, x_N\}$ که N تعداد بردارهای ویژگی^۱ و بردار متعلق به هر پیکسل به صورت X_i $\{x_i^1, x_i^2, x_i^3 \dots, x_i^m\}$ ، که در آن m تعداد باندهای تصویر و X_i بردار ویژگی پیکسل^۲ است، در نظر می‌گیریم. مجموعه $Y = \{y_1, y_2, \dots, y_i, \dots, y_N\}$ بیانگر برچسب‌های^۳ N بردار موجود در تصویر است، اگر تعداد کلاس‌ها را با C کلاس نشان دهیم، مجموعه احتمال هر پیکسل را با Y_i $\{y_i^1, y_i^2, y_i^3, \dots, y_i^C\}$ نشان می‌دهیم که مجموعه‌ای برداری به ابعاد C است. به هر پیکسل موجود در تصویر یک بردار Y تعلق گرفته و احتمال حضور پیکسل^۴ i ام در هر کلاس را نشان می‌دهد، برچسب نهایی پیکسل مربوط به کلاسی است که بیشترین احتمال^۴ حضور در آن کلاس را کسب کند، رابطه زیر اندیس^۵ بیشترین احتمال را به عنوان کلاس مورد نظر تعیین می‌کند [۳۶].

^۱ Feature vectors

^۲ Band

^۳ Label

^۴ Probability

^۵ index

$$L_i = \arg.\max(\{y_i^1, y_i^2, y_i^3, \dots, y_i^C\}) \quad 1-3$$

۳-۱-۲-مواد و داده‌ها

داده‌های مورد استفاده در این پایان نامه شامل یک تصویر از ماهواره world-view3 از شهر کرمان است (شکل ۱-۳) که در سال ۲۰۱۹ اخذ شده‌است. کرمان یکی از کلانشهری در جنوب شرقی ایران و مرکز استان کرمان است. جمعیت این شهر طبق سرشماری بر اساس آمار سال ۹۵ معادل ۵۳۷،۰۸۱۸ نفر بوده‌است. این منطقه از نظر تنوع ساخت و ساز بسیار مورد توجه است و می‌توان آن را به علت وجود بافت‌های مختلف شهری چالشی جهت طبقه‌بندی تصاویر سنجش از دوری دانست، از همین رو تصویر ماهواره‌ای این شهر جهت طبقه‌بندی و ارزیابی روش‌های موجود و روش پیشنهادی ما انتخاب شده‌است تا عملکرد شبکه‌ها به بهترین نحو مورد بررسی و ارزیابی قرار گیرد.

در ادامه مشخصات این ماهواره در جدول ۱-۳ آورده شده است، قدرت تفکیک برای روش segnet برابر با ۱/۲۴ متر است.

جدول ۱-۳- مشخصات ماهواره world-view3

تاریخ پرتاب	panchromatic resolution	multispectral resolution	revisit time
۱۳ آگوست ۲۰۱۴	۰/۳۱ متر	۱/۲۴ متر	کمتر از ۱ روز



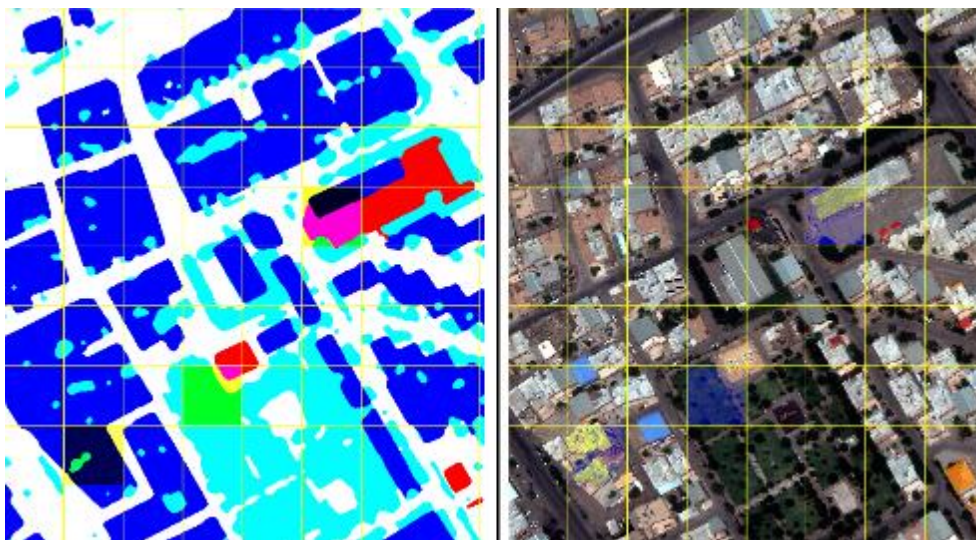
شکل ۳-۱- تصویر منطقه مورد مطالعه

داده‌ها به منظور استفاده برای یادگیری فعال در روش‌های pixel-base طبق جدول ۳-۲ در ۴ کلاس آماده شده‌اند.

جدول ۳-۲- داده‌های مورد استفاده در روش‌های pixel-base

داده/کلاس	راه	ساختمان	پوشش گیاهی	تداخلی
train	۱۶۵	۹۲	۳۷	۶
test	۴۰۶۱	۲۲۸۴	۸۷۴	۱۴۱
poll	۳۰۲۶	۱۷۲۷	۶۸۵	۱۱۴

داده‌های مورد استفاده در روش segnet بصورت قطعه تصویر بوده و بصورتی انتخاب شده‌اند که شامل همه کلاس‌های موجود باشند، بنابراین این patch ها با رنگ‌های مختلف بر روی تصویر اصلی مشخص در شکل ۳-۲ مشخص شده‌اند:



شکل ۳-۲- راهنمای patch‌های اولیه برای train

۳-۱-۳- برنامه نویسی پایتون

پایتون^۱ یک زبان برنامه نویسی شیء گرا و سطح بالا است، کدهایی که با این زبان نوشته شوند قابل حمل بوده و بر روی سیستم عامل‌های مختلف قابل اجرا هستند [۳۹]، در این پایان نامه طبقه‌بندی به روش‌های غیر از CNN از افزونه sklearn استفاده شده‌است، این افزونه برای یادگیری فعال از افزونه دیگری به نام modAL استفاده می‌کند [۴۰].

برای طبقه بندی با CNN از افزونه ای به نام Pytorch استفاده می‌شود که سرعت بسیار بالایی برای اجرا دارد، یادگیری فعال برای segnet نیز توسط خود ما برای این افزونه طراحی شده‌است.

توجه : قسمت برنامه نویسی برای هر روش در هر بخش نیز به صورت مختص آن بخش

^۱ python

توضیح داده شده است.

کلیه برنامه‌ها در Google-COLAB اجرا شده‌اند، که یک سرویس شرکت گوگل برای برنامه نویسان پایتون است، سخت افزار اجرا کننده کدها به شرح زیر است.

CPU: Intel(R) Xeon(R) CPU @ 2.20GHz

Ram: 13GB

GPU: Tesla k80-PCIE-12GB

۳-۲-روش MLC

۳-۲-۱-کلیات روش MLC

این روش ابتدا برای دسته‌بندی مشابهت‌های آماری استفاده می‌شد که بعداً با نیاز به طبقه‌بندی تصاویر، به سنجش از دور راه پیدا کرد. می‌توان گفت که در سنجش از دور در ابتدا دو روش بیشترین کاربرد را در طبقه‌بندی تصاویر داشتند، اولی روش^۱ MLC که بصورت نظارت شده عمل می‌کرد و دومی روش‌های نظارت نشده که همان روش‌های خوشه‌بندی بودند [۴۱]. با توجه به تحقیقات انجام شده با این روش می‌توان دریافت که استفاده از این روش در طبقه‌بندی تصاویر اهمیت بسیاری دارد، در ادامه به توضیح این روش خواهیم پرداخت. این روش از روش‌های نظارت شده است و به داده آموزشی نیاز دارد، برای عملکرد بهینه این روش بهتر است که توزیع داده‌ها به صورت نرمال^۲ باشد. پس از معرفی داده‌ها به این روش، برای هر کلاس یک تابع چگالی شرطی^۳ جداگانه ساخته می‌شود، اگر فرض شود که C کلاس داریم و تعداد باندهای تصویری برابر m است، احتمال تعلق به هر کلاس موجود با رابطه زیر تعریف می‌شود [۴۲]:

$$P(G_c|X) = \frac{P(X|G_c)P(G_c)}{P(X)} \quad (۲-۳)$$

^۱ Maximum likelihood classifier

^۲ Normal distribution

^۳ Probability density function

X : بردار ویژگی پیکسل

$P(G_c)$: احتمال هر کلاس است که با توجه به چگالی کلاس بدست می‌آید

$P(X|G_c)$: احتمال شرطی X به شرط اینکه متعلق به کلاس G_c باشد

$P(X)$: احتمال رخداد هر پیکسل که این مقدار برای تمام پیکسل‌ها برابر در نظر گرفته

می‌شود.

اگر ما از قبل درباره $P(G_c)$ اطلاعاتی در دست نداشته باشیم، می‌توان احتمال همه کلاس‌ها را یکسان در نظر گرفت، پس از یکسان شدن همه $P(G_c)$ ‌ها می‌توان این پارامتر را نادیده گرفت، بنابراین تابع احتمال توسط $P(X|G_c)$ تعیین می‌شود که به آن تابع شباهت^۱ X به G_c گفته می‌شود، بنابراین رابطه می‌تواند به صورت زیر بازنویسی شود:

$$M_c(X) = \ln(P(G_c)) + \ln\left(\frac{S_c^{-1}}{(2\pi)^{\frac{m}{2}}}\right) - \frac{1}{2}(X - \mu_c)^T S_c^{-1}(X - \mu_c) \quad (3-3)$$

$M_c(X)$: تابع شباهت بردار X

μ_c : بردار میانگین کلاس c ام

S_c : ماتریس کواریانس کلاس c ام

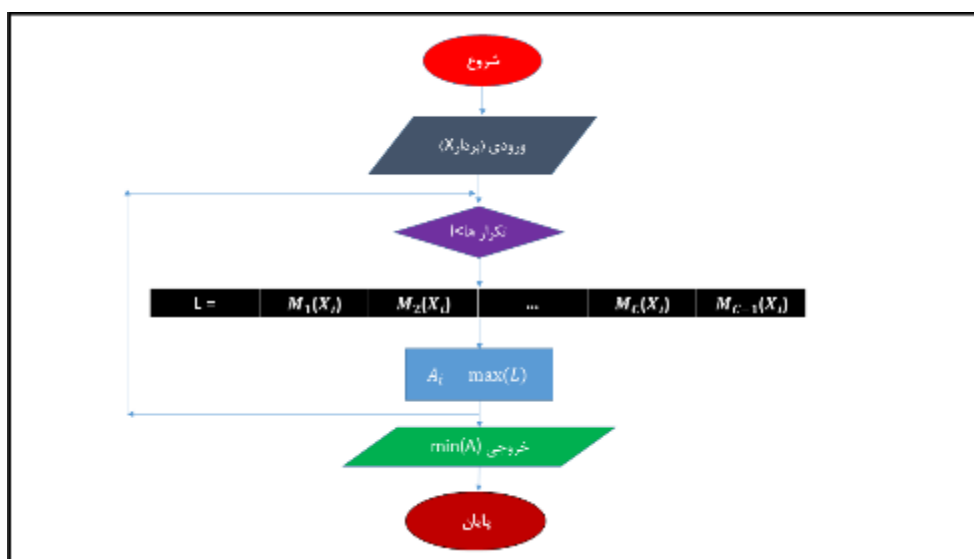
این روش برای هر پیکسل یک بردار X در نظر گرفته و طبقه بندی را بر اساس تابع شباهت برای هر کلاس یک شباهت بدست می‌آورد [۴۳]، سپس با استفاده از رابطه ۳-۱ کلاس مورد نظر را تعیین می‌کند.

کدهای مربوط به این روش فقط با استفاده از کتابخانه numpy، به صورت کلاس و با استفاده از شی‌گرایی در ۹۷ خط نوشته شده تا در قسمت یادگیری فعال به دفعات مورد استفاده قرار گیرند.

^۱ likelihood function

۳-۲-۲- روش MLC با استفاده از یادگیری فعال

اگر یک تصویر به ابعاد w و h با m باند داشته باشیم، تعداد پیکسل‌های تصویر برابر با $w \times h$ خواهد شد، همانطور که در فصل سوم اشاره شد برای هر پیکسل یک بردار با ابعاد $X_{[m,1]}$ خواهیم داشت، طبق رابطه ۳-۳ برای هر پیکسل به ازاء هر کلاس یک احتمال تخمین زده می‌شود که کلاس مورد نظر بیشترین احتمال را کسب می‌کند؛ برای درک بهتر به شکل ۳-۳ توجه کنید:



شکل ۳-۳- فلوجارت نحوه بدست آوردن بردار بیشترین قطعیت از بین کلاس‌ها، فلش با رنگ آبی کمرنگ بیانگر حلقه تکرار است

با توجه به شکل ۳-۳ می‌بینیم که برداری که به ترتیب شامل کوچکترین قطعیت‌های نتایج MLC است، بدست آمده که A نامگذاری شده‌است، ابعاد A به اندازه پیکسل‌هایی است که مورد طبقه بندی قرار گرفته‌اند، کوچکترین مقدار بدست آمده در بردار A مربوط به پیکسلی است که کمترین قطعیت را دارا است، پیکسل‌های جدید نیز به عنوان داده آموزشی با انتخاب از داده‌های pool بر اساس کوچکترین مقدار در A انتخاب و به عنوان داده آموزشی به داده‌های قبلی اضافه می‌شوند، البته این نکته را باید ذکر کرد که اختیار اینکه در هر بار یادگیری فعال چند پیکسل به داده‌های آموزشی اضافه کند بر عهده کاربر است [۳۲].

۳-۳-۳- روش ANN

۳-۳-۱- کلیت شبکه‌های ANN

شبکه‌های عصبی از الگوریتم‌های محاسباتی هستند که برای پیش‌بینی در بسیاری از علوم استفاده می‌شوند. الگوریتم کلی شبکه‌های عصبی الهام گرفته شده از سیستم عصبی موجودات زنده است. شبکه‌های عصبی با الگوریتمی به نام back-propagation آموزش می‌بیند، این الگوریتم، از دسته‌ای از داده‌ها شامل مقادیر ورودی و مقادیر خروجی به عنوان داده‌های آموزشی برای آموزش شبکه استفاده می‌کند. فرآیند یادگیری در شبکه‌های عصبی از لایه ورودی^۱ شروع شده، لایه‌های مخفی^۲ پس از لایه ورودی قرار گرفته و فرآیند محاسبات خروجی را انجام می‌دهند، وزن دهی مناسب نیز بر اساس مقدار خروجی داده‌های آموزشی انجام می‌شود. این وزن دهی بر اساس خطی بودن یا نبودن رابطه میان مقادیر ورودی و خروجی متفاوت است. شبکه‌های عصبی می‌توانند بیش از یک لایه مخفی داشته باشند که در این صورت به این شبکه‌های شبکه چند لایه^۳ گفته می‌شود [۴۴].

در شبکه‌های چند لایه، هر لایه به عنوان ورودی لایه بعدی است، به همین ترتیب هر ورودی در هر لایه، وزن مخصوص به خودش را دارد. در ادامه به طور خلاصه فرآیند یادگیری شبکه عصبی با یک لایه مخفی توضیح داده می‌شود که برای شبکه‌هایی با لایه‌های بیشتر نیز قابل تعمیم است.

اولین قسمت از شبکه‌های عصبی مربوط به لایه ورودی است، لایه ورودی را که با حرف I نمایش داده می‌شود و شامل نورهایی است که مجموعه پیکسل‌های تصویر X با m باند را به عنوان ورودی به شبکه وارد می‌کند، معمولاً اگر از روش‌های پیش پردازش برای کاهش ابعاد تصویر استفاده نکنیم، به تعداد باندهای تصویر برای شبکه عصبی pixel-base، ورودی وجود دارد،

^۱ Input Layer

^۲ Hidden Layers

^۳ Multi-Layer

دومین قسمت مورد بررسی لایه‌های مخفی است، لایه‌های مخفی شامل مجموعه نرون^۱هایی است که دارای تابع فعالیت^۲ هستند، نرون‌ها ورودی‌های موجود در هر لایه هستند که خود شامل توابع ریاضی فعالیت هستند، معمولاً برای طبقه‌بندی تصویر از نرون‌های غیر خطی استفاده می‌شود، این نرون‌ها شامل توابعی ریاضی مانند sigmoid، Hyperbolic tangent و غیره هستند، ورودی این نرون‌ها مجموع وزن دار لایه I می‌باشد، این ورودی را با z نشان می‌دهیم که رابطه محاسبه آن به صورت زیر است:

$$Z = \sum_{i=1}^m x_i \times w_i \quad (4-3)$$

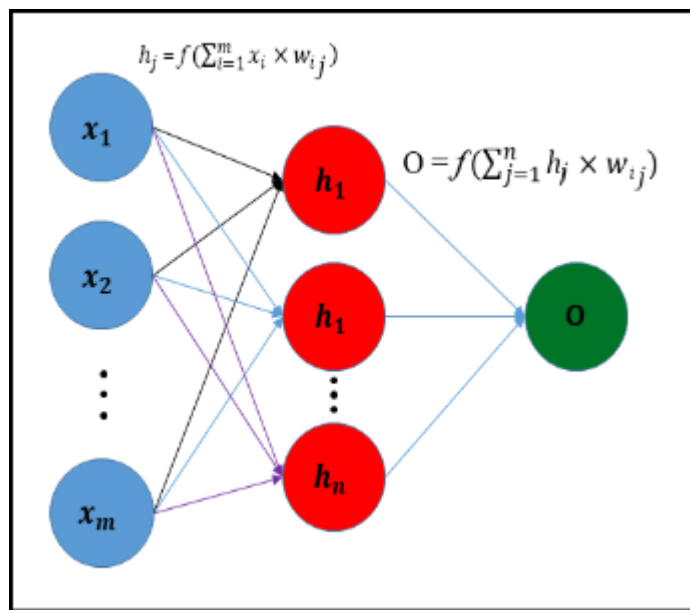
وزن‌ها بیشتر اوقات اعداد کوچکی هستند که در هر بار یادگیری شبکه‌های عصبی به روز می‌شوند. اگر تابع عملگر بر روی مقدار ورودی را f بنامیم، خروجی لایه مخفی با رابطه زیر محاسبه می‌شود:

$$h = f(Z) \quad (5-3)$$

وابسته به تعداد نرون‌هایی که برای لایه مخفی در نظر می‌گیریم، برای لایه خروجی ورودی وجود دارد که با رابطه ۳-۴ قابل محاسبه است، لایه خروجی خود شامل تابعی است که مانند رابطه ۳-۵ محاسبه می‌شود در شکل ۳-۴ می‌توان ساختمان شبکه عصبی با یک لایه مخفی را مشاهده کرد.

^۱ Neuron

^۲ Activation function



شکل ۳-۴- شبکه عصبی با یک لایه مخفی

نحوه یادگیری شبکه بدین صورت است که شبکه ابتدا با وزن‌های ابتدایی^۱ به ازاء هر ورودی یک خروجی محاسبه می‌کند، با هر خروجی مقداری به عنوان خطا^۲ بدست خواهد آمد که از رابطه زیر قابل محاسبه است:

$$E = \sum_{i=1}^k \frac{1}{2} (t_k - o_k)^2 \quad (۳-۶)$$

E: خطا

K: تعداد نمونه‌ها

O_k : خروجی شبکه

t_k : مقدار هدفی که خروجی شبکه باید به آن مقدار نزدیک شود

^۱ Initial weights

^۲ Error

الگوریتم شبکه‌های عصبی با تکرار آموزش می‌یابند، بدین صورت که بر اساس خطای بدست آمده هر بار وزن‌ها تغییر کرده و شبکه یک خروجی جدید ارائه می‌کند که این خروجی نیز خطایی جدید ارائه می‌کند، هر بار تکرار را یک epoch یادگیری می‌گویند، همین منوال ادامه می‌یابد تا جایی که شبکه به یکی از شروطی که ما برای آن تعیین کرده‌ایم رسیده و کار آموزش شبکه تمام شود، این شروط می‌تواند رسیدن شبکه به دقتی خاص یا رسیدن epochها به تعدادی مشخص یا شروطی دیگر مانند کوچکتر شدن مشتق وزن‌ها و ... باشد. رابطه محاسبه وزن جدید در هر epoch به شرح زیر است:

$$w_{ij}(n+1) = \eta \left(x_i \times \frac{\partial E}{\partial w_{ij}} \right) + \alpha \Delta w_{ij} \quad (۷-۳)$$

: نرخ یادگیری شبکه است که معمولاً عددی کوچک و بین ۰ و ۱ است

$\frac{\partial E}{\partial w_{ij}}$: مشتق خطا نسبت به وزن مورد نظر است

α : momentum خطا است و بین ۰ و ۱ انتخاب می‌شود

Δw_{ij} : تغییرات وزن در epoch قبلی است

محاسبه مشتق در هر دور برای همه وزن‌ها انجام می‌شود و نرخ تغییر هر وزن به مقدار این مشتق بستگی دارد برای محاسبه این مشتق از قانون مشتق گیری زنجیری طبق رابطه زیر استفاده می‌شود [۴۵]:

$$\frac{\partial E}{\partial w_{ij}} = \frac{1}{2} \sum_{t=1}^k \frac{\partial net^t}{\partial w_{ij}} \frac{\partial E^t}{\partial net^t} \quad (۸-۳)$$

$\frac{\partial net^t}{\partial w_{ij}} \frac{\partial E^t}{\partial net^t}$: مشتق شبکه ضرب شده در وزن ij ام است

در قسمت کد نویسی برای این بخش از کتابخانه sk-learn استفاده شده است، این کتابخانه یک شیء از نوع شبکه عصبی import می کند که برای سایر قسمت های یادگیری فعال نیز قابل استفاده است، به علت استفاده از توابع کتابخانه ای کدنویسی این بخش در ۱۰ تا ۲۰ خط با توجه به روش یادگیری فعال قابل انجام است. با استفاده از این کتابخانه به راحتی می توان تعداد لایه ها، نرون ها و دیگر ابر پارامترهای مربوط به شبکه عصبی را به راحتی تغییر داد. توجه : کد نویسی مربوط به این بخش در ۱۰ خط و به صورت پیش فرض با ۵ لایه و نرون خطی انجام شده است.

۳-۲-۳- شبکه های ANN و یادگیری فعال

در شبکه های عصبی خروجی ها به صورت امتیازهایی هستند که توسط لایه آخر به هر کلاس تعلق می گیرد، این امتیازها به عنوان معیاری برای انتخاب داده های جدید در یادگیری فعال استفاده می شوند. شبکه های عصبی از راه دیگری نیز به ما کلاس ها را تحویل می دهند به این صورت که با استفاده از داده های آموزشی بر اساس شماره کلاس معمولاً یک تابع غیر خطی بر روی داده های آموزشی برازش می دهند. تفاضل مقدار برازش داده شده تا کلاس مورد نظر یک خطا برای ما ایجاد می کند، در حالت Batch-mode می توان خطای کمترین مربعات پیکسل های Batch یا همان MSE^1 برای تعدادی از پیکسل ها را معیاری برای انتخاب داده ها توسط یادگیری در نظر گرفت، در واقع معمولاً شبکه های عصبی معمولی از روش های تخمین وزن و بعضی از برازش تابع برای رسیدن به کلاس مورد نظر استفاده می کنند از همین رو می توان MSE و یا خروجی لایه آخر را معیاری برای QUERY یادگیری فعال دانست [۴۶]. معیار MSE مانند خروجی آخرین لایه، نوعی امتیاز دهی را انجام می دهد، بنابراین می تواند به عنوان معیاری در انتخاب داده ها در یادگیری فعال مورد استفاده قرار می گیرد.

¹ Mean square error

۳-۴-روش DBN

۳-۴-۱-کلیات روش DBN

DBN یک معماری خاص از شبکه‌های عصبی عمیق است که در هر لایه آن از یک RBM^1 استفاده می‌شود. هر RBM بصورت جداگانه یک شبکه عصبی مصنوعی است که می‌تواند توزیع احتمال داده‌های ورودی را یادگیری کند. فرض کنیم که لایه h_l بردار خروجی خود را از لایه h_{l-1} دریافت می‌کند (مقادیر این خروجی vv نامیده می‌شود) و یک بردار خروجی (مقادیر این خروجی hv نامیده می‌شوند) برای لایه h_{l+1} فراهم می‌سازد، در یک شبکه RBM هیچ ارتباط مستقیمی^۲ بین لایه‌های مخفی نیست و در واقع چند RBM پشت سر هم به عنوان لایه قرار دارند، شبکه برای هر دو لایه که بتوانند مقادیر vv و hv به ما ارائه کنند با رابطه زیر یک توزیع احتمال بدست می‌آورد:

$$\rho(vv, hv) = \frac{1}{\sum_h e^{-E(vv, hv)}} e^{-E(vv, hv)} \quad (۹-۳)$$

از رابطه ۹-۳ می‌توان دریافت که مقداری که به ما ارائه می‌شود بین ۰ و ۱ قرار دارد، همانطور که گفته شد در RBM ها لایه‌های مخفی به طور مستقیم به یکدیگر متصل نیستند و می‌توان گفت که مستقل هستند؛ از همین رو می‌توان رابطه احتمال شرطی زیر را نوشت:

$$\begin{aligned} \rho(vv|hv) &= \prod_{ii} \rho(hv_{ii}.vv) \\ \rho(hv|vv) &= \prod_{jj} \rho(vv_{jj}.hv) \end{aligned} \quad (۱۰-۳)$$

^۱ Restricted Boltzmann machine

^۲ Direct connection

که در توابع بالا، توابع احتمال شرطی به صورت زیر نوشته می‌شوند که تابع f تابع سیگموئید^۱ است :

$$\begin{aligned} p(hv_{ii} = 1/vv) &= f(W_{ii}vv + b_i) \\ p(vv_{jj} = 1/hv) &= f(W_{jj}hv + c_{jj}) \end{aligned} \quad (۱۱-۳)$$

W : ماتریس وزن

c و b : بردارهای offset

نحوه رمزگذاری^۲ و رمز گشایی^۳ هر لایه با تابع انرژی که با E^4 نشان داده می‌شود به مقادیر b و c و W مربوط می‌شود که به صورت زیر تعریف می‌شوند:

$$E(vv, hv) = -b'vv - c'hv - h'Wvv \quad (۱۲-۳)$$

در قسمت‌های قبل مشخص شد که برای آموزش شبکه نیاز به مشتق گیری داریم، مشتق تابع احتمال در RBM از رابطه زیر بدست می‌آید:

$$\frac{\partial \ln \rho(vv)}{\partial \theta} = \sum_h \rho(hv|vv) \frac{\partial E(vv, hv)}{\partial \theta} - \sum_h \rho(vv|hv) \frac{\partial E(vv, hv)}{\partial \theta} \quad (۱۳-۳)$$

θ : مشتق را توصیف می‌کند

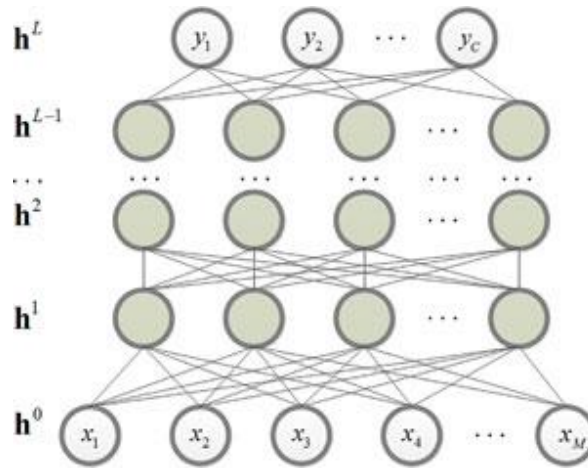
θ : بیانگر پارامترهای W و b یا W و c است [۶].

^۱ Sigmoid function

^۲ Encoding

^۳ Decoding

^۴ Energy function



شکل ۳-۵- ساختار DBN

۳-۴-۲- روش DBN با یادگیری فعال

در یادگیری فعال داده‌ها به این نیاز دارند که بر اساس معیاری خاص که اهمیت آن‌ها برای بهتر شدن کارایی روش طبقه بندی مرتب و سپس انتخاب شوند، معیاری که برای انتخاب داده‌ها در این تحقیق استفاده شده است، معیار وزن دهی بر اساس چگالی است که در تحقیقی انجام شده در سال ۲۰۰۸ به چاپ رسیده است [۴۷]، این معیار نیز قطعی‌ترین احتمال را از برچسب داده‌ها استخراج کرده و بصورت رابطه ۳-۱۴ تعریف می‌شود:

$$x^* = \max_x \left\{ \Phi(x) \frac{1}{N} \sum_{n=1}^N S(x, x_n) \right\} \quad (۳-۱۴)$$

x : نمونه انتخاب شده از یک دسته پیکسل مشابه است

x_n : نمونه کاندید شده برای مقایسه با پیکسل x است

$S(x, x_n)$: این تابع نیز شباهت میان نمونه‌های x و x_n اندازه گیری می‌کند

$\Phi(x)$: تابعی است که عدم قطعیت x را اندازه گیری می‌کند

در یادگیری فعال تابع مشابهتی (S) استفاده می‌شود که بتواند نمونه‌هایی که به عنوان نماینده دسته در نظر گرفته می‌شوند را به خوبی بیابد، به عبارت دیگر این تابع باید هر دو مورد ساختار و توزیع در داده‌ها به طور مؤثری ارزیابی کند. شکل ۳-۵ را برای یک ساختار DBN در نظر می‌گیرد.

مجموعه بردار بایاس‌ها^۱ که آن را دیکشنری در نظر می‌گیریم، شامل مجموعه‌ای است که اعضای آن اتم نامیده می‌شوند و می‌تواند کل ویژگی تعدادی داده را توصیف کند که به عنوان نماینده داده‌ها در یادگیری فعال در نظر گرفته می‌شود.

برای بدست آوردن عدم قطعیت در این روش از دیکشنری‌ها نیز استفاده می‌شود. با توجه به شکل ۳-۵، خروجی لایه l ام را که بصورت $H^l = \{h_1^l \dots h_n^l \dots h_N^l\}$ نشان می‌دهیم، می‌تواند به عنوان یک خروجی از تعداد N بردار ورودی $X = \{x_1 \dots x_n \dots x_N\}$ در نظر گرفت. اگر coding هایی که میان لایه‌ها در RBMها اتفاق می‌افتد به عنوان یادگیری نظارت نشده و کاهش ابعاد ویژگی^۲ در یادگیری عمیق قلمداد می‌شود، به این علت که نظارت مستقیمی بر RBMها وجود ندارد و همه این فرآیند به کمک اطلاعاتی بدست می‌آید که از داده‌های استخراج شده توسط DBN محاسبه می‌شود.

اتم‌ها می‌توانند از آخرین لایه مخفی (لایه L ام طبق شکل ۳-۵) استخراج شود $H_l^n \in R^C$ می‌تواند به صورت $H_l^n = D\alpha$ تعریف شود که $D, D \in R^{C \times p}$ یک دیکشنری با p اتم است و α یک بردار ضریب است که بر اساس یکسان بودن پیکسل انتخابی و پیکسل‌های مشابه در داده‌های آموزشی تعیین می‌شود، یادگیری دیکشنری را می‌توان به صورت زیر رابطه بیان کرد:

$$\min_{D, \alpha} ||h_n - D\alpha||_2^2 \quad ||\alpha||_0 \leq k \quad (۱۵-۳)$$

subjected to

^۱ Bias

^۲ feature learning and dimension reduction

منظور از $\|\cdot\|_0$. نرم L_0 و منظور از $\|\cdot\|_2$. نرم L_2 است. وقتی که اندازه داده‌های ما $H^l = \{h_1^l \dots h_n^l \dots h_N^l\}$ برابر N باشد بنابراین به تعداد N نیز ضریب α خواهیم داشت، پس رابطه ۱۵-۳ به این صورت بازنویسی می‌شود:

$$\min_{D, \alpha_1, \dots, \alpha_N} \sum_{i=1}^N \|h_i - D\alpha_i\|_2^2 + \lambda \sum_{i=1}^N \|\alpha_i\|_0 \quad (۱۶-۳)$$

λ : پارمتری برای منظم سازی^۱ رابطه است
با استفاده از روشی به نام بی نظمی اطلاعات^۲ که از آن برای عدم قطعیت استفاده خواهیم کرد، تابع بی نظمی بصورت زیر تعریف می‌شود:

$$\Phi(x) = - \sum_{j=1}^c p(y_j|x) \log(p(y_j|x)) \quad (۱۷-۳)$$

$p(y_j|x)$: احتمال این است که نمونه x متعلق به کلاس j ام باشد
حال می‌خواهیم با استفاده از تابع بی‌نظمی اطلاعات برای اندازه‌گیری عدم قطعیت و sparse representation برای اندازه‌گیری مشابهت به سراغ یادگیری فعال برویم و با ترکیبی از این دو به آموزش یک DBN بپردازیم.

اگر $X = \{x_1 \dots x_N\}$ را مجموعه داده‌ای برچسب گذاری نشده و خروجی شبکه را به صورت $H = \{h_1 \dots h_N\}$ در نظر بگیریم، برای برچسب گذاری مجموعه X باید پردازش‌هایی با یک الگوریتم جست و جو انجام شود. مقادیر ابتدایی خروجی شبکه را به صورت دیکشنری $D = \{d_1 \dots d_N\} \subset H$ در یادگیری فعال دیکشنری D در هر بار یادگیری با دیکشنری E جایگزین

^۱ regularization

^۲ Information Entropy

می‌شود، دیکشنری E را می‌توان به صورت $E = \{d_{n+1} \dots d_{N+m}\} \subset H$ نشان داد، اکنون انتخاب داده‌ها^۱ براساس مشابهت را می‌توان به صورت رابطه ۱۸-۳ تعریف کرد:

$$j(E, \beta_1 \dots \beta_N) = \sum_{i=1}^N \|h_i - [DE] \begin{bmatrix} \alpha_i \\ \beta_i \end{bmatrix}\|_2^2 + \lambda \sum_1^N \|\beta_i\|_0 \quad (18-3)$$

$$\alpha = \begin{bmatrix} \alpha_{i,1} \\ \vdots \\ \alpha_{i,n} \end{bmatrix} \quad \beta = \begin{bmatrix} \alpha_{i,n+1} \\ \vdots \\ \alpha_{i,n+m} \end{bmatrix}$$

ضریب β ، ضریب E و ضریب α ضریب D است، E مجموعه اتم‌هایی است که بعد از D بوجود آمده است. رابطه ۱۹-۳ را با جایگذاری $r_i = h_i - D\alpha_i$ می‌توان به صورت زیر نیز باز نویسی کرد:

$$j(E, \beta_1 \dots \beta_N) = \sum_{i=1}^N \|r_i - E\beta_i\|_2^2 + \lambda \sum_1^N \|\beta_i\|_0 \quad (19-3)$$

با توجه به اینکه ما با هر دو معیار معیار شباهت و عدم قطعیت سر و کار داریم، لذا تابع بدست آوردن عدم قطعیت باید در رابطه ۱۹-۳ اثر داده شود، بدین منظور رابطه ۱۹-۳ بصورت زیر بازنویسی می‌شود:

$$j(E, \beta_1 \dots \beta_N) = \sum_{i=1}^N \|r_i - E\beta_i\|_2^2 + \lambda \sum_1^N \|\Gamma\beta_i\|_0 \quad (20-3)$$

تابع Γ بصورت زیر تعریف می‌شود:

^۱ Data selection

$$\Gamma = \text{diag.} [\Phi(d_{n+1}), \dots, \Phi(d_{n+m})] \quad (21-3)$$

هر چقدر که $\Phi(d_i)$ بزرگتر باشد نشان دهنده عدم قطعیت بیشتر اتم λ_i است، پس مسأله بصورت مسأله کمینه سازی رابطه ۳-۲۰ با جایگزینی دیکشنری E است، پس رابطه نهایی برای یافتن داده با بیشترین عدم قطعیت بصورت زیر نوشته می شود [۳۶]:

$$d = \underset{\xi=n+1 \dots n+m}{\operatorname{argmax}} \{ \phi(d_{\xi}) | r_i \cdot d_{\xi} | \} \quad (22-3)$$

کد نویسی مربوط به این بخش نیز با کمک توابع کتابخانه sklearn انجام شده است، کدها در ۵۷ خط آماده شده و برای بخش یادگیری فعال نیز از کتابخانه numpy کمک گرفته شده است.

۳-۵-روش CNN

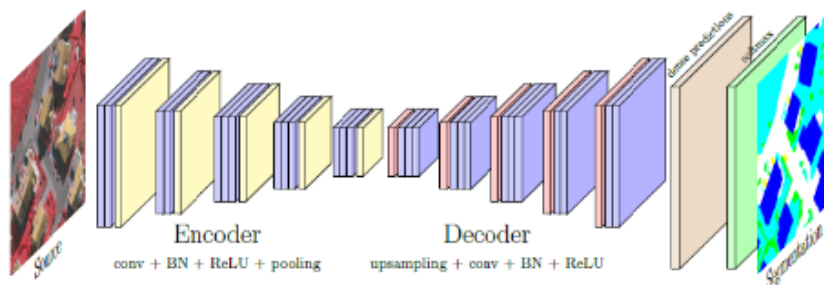
۳-۵-۱-روش segnet

برچسب گذاری معنایی^۱ با روش Semantic segmentation توسط طبقه بندی pixel-wise با در نظر گرفتن تراکم پیکسل^۲ های موجود در همسایگی پیکسل مورد طبقه بندی انجام می شود، برای انجام این نوع طبقه بندی از معماری شبکه های عصبی FCN^۳ استفاده می شود. یکی از شبکه های کارآمد که ما در این پایان نامه مورد استفاده قرار می دهیم segnet نام دارد که زیر مجموعه FCN ها است [۴۸]. این شبکه بر اساس معماری رمزگذاری-رمزگشایی است که یک خروجی به تعداد پیکسل های تصویر ورودی اش تولید می کند، معماری این شبکه در شکل ۳-۶ آورده شده است.

^۱ Semantic labeling

^۲ dense pixel-wise classification

^۳ Fully Convolutional Network



شکل ۳-۶- معماری شبکه segnet و لایه‌های رمزگذاری و رمز گشایی

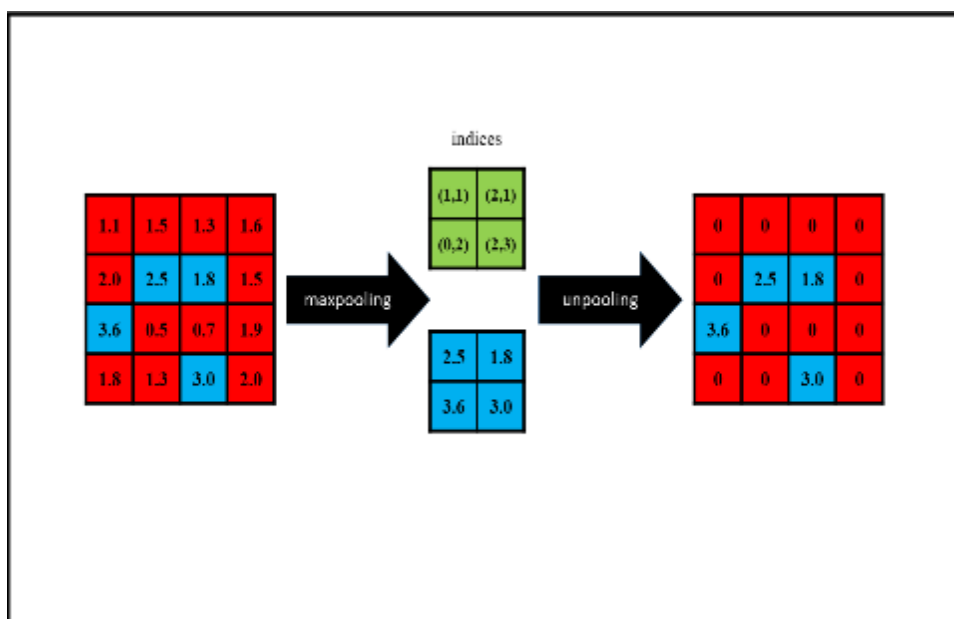
در ادامه برای توضیح ساز و کار segnet به توضیحات مختصری بسنده کرده و بیشتر توضیحات را به بخش یادگیری فعال معطوف می‌کنیم، ورودی این شبکه‌ها بر خلاف شبکه‌هایی مانند DBN پیکسل نیست، بلکه مجموعه قطعه‌هایی از تصویر^۱ است، اندازه این قطعه تصویرها توسط کاربر تعیین می‌شود، مثلاً اگر اندازه لایه ورودی را 32×32 در نظر بگیریم، تصویر به قسمت‌هایی به همین اندازه تقسیم خواهد شد، درخصوص لایه‌های مخفی در این شبکه با توجه به شکل ۳-۶ می‌بینیم که لایه‌های رمز گذار به صورت $\text{conv}^2 + \text{BN}^3 + \text{ReLU} + \text{pooling}$ هستند، لایه‌های conv شامل فیلترهای پردازش تصویر هستند، به این فیلترها وزن نیز گفته می‌شود. مجموعه فیلترهای مورد استفاده در segnet فیلترهای VGG-16 نام دارند. BN نیز خروجی لایه‌های قبل را با یک تابع مشخص به ReLU می‌فرستد، سپس تابع ReLU نیز با یک رابطه ریاضی اگر ورودی بزرگتر از صفر باشد خود ورودی را به عنوان خروجی و اگر کوچکتر از صفر باشد، صفر را به عنوان خروجی باز می‌گرداند.

توابع pooling اگر پنجره آن‌ها را 2×2 در نظر بگیریم، مانند صورت شکل ۳-۷ عمل می‌کنند. سمت چپ شکل تابع maxpooling مربوط به مرحله رمزگزاری و سمت راست شکل unpooling مربوط به قسمت رمزگشایی آورده شده‌اند:

^۱ Image patches

^۲ Convolution

^۳ Batch normalization



شکل ۳-۷- تابع maxpooling (سمت چپ)، تابع unpooling (سمت راست)

با توجه به شکل ۳-۶ می‌توان دریافت که فرآیند آموزش شبکه به این گونه است که قطعه تصویر ورودی به عملیات رمزگذاری به یک قطعه تصویر کوچکتر از لایه ورودی تبدیل خواهد شد. سپس این قطعه تصویر رمزگذاری شده با اعمال دوباره فیلترها منتهی به صورت بر عکس به اندازه تصویر ورودی باز خواهد گشت، بنابراین ساینز ورودی و خروجی این شبکه‌ها برابر است [۴۹].

معمولاً همه شبکه‌های عصبی مصنوعی از یک معماری مخصوص به نوع خود تبعیت کرده و فرق آن‌ها در نوع لایه‌ها است، در معماری آن‌ها پس از آخرین لایه یک لایه معیار قرار دارد، هدف کلی شبکه کمینه سازی مقدار عددی این معیار است. در شبکه segnet معیار تابع loss است، که در واقع مجموعه احتمال‌های کلاس‌هایی است که به عنوان کلاس مورد نظر طبقه‌بندی انتخاب نشده‌اند و به صورت زیر تعریف می‌شود:

$$loss = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^k y_j^i \log \left(\frac{\exp(z_j^i)}{\sum_{l=1}^k \exp(z_l^i)} \right) \quad (۳-۲۳)$$

در رابطه بالا N تعداد پیکسل‌ها، k تعداد کلاس‌ها، y^i بیانگر برچسب و $(z_1^i \dots z_k^i)$ بردار پیش‌بینی برچسب‌ها است.

۳-۵-۲- یادگیری فعال و segnet

این بخش تا به حال در سنجش از دور تحقیق مرتبطی نداشته است، در علوم مرتبط با پردازش تصویر نیز تحقیقات زیادی در این زمینه انجام نگرفته است، بنابراین اکثر معیارهای استخراجی بر اساس نوآوری در این پایان نامه و با استفاده کمتری از تحقیقات دیگر نسبت به بخش‌های قبل خواهد بود.

یکی از پرکاربردترین معیارهایی که برای یادگیری فعال به کار برده می‌شود، عدم قطعیت است. عدم قطعیت در CNNها با احتمال سر و کار داشته و برای بدست آوردن آن باید احتمال را نیز به دست آورد.

در شبکه‌های عصبی عمیق که برای طبقه‌بندی استفاده می‌شوند، معمولاً آخرین لایه یک لایه جهت طبقه‌بندی هر عنصر ورودی است، پس از این لایه یک معیار برای آموزش شبکه قرار می‌گیرد، با کمینه شدن این معیار آموزش شبکه نیز انجام می‌شود. در شبکه‌های CNN و به تبع segnet نیز از این لایه که softmax نام دارد، استفاده می‌شود.

Softmax تابعی است که برداری از مقادیر را دریافت کرده و برداری جدید با مجموع ۱ به محاسبه می‌کند که مقادیر موجود در این بردار بین ۰ تا ۱ هستند، این مقادیر را به عنوان احتمال نیز می‌توان در نظر گرفت. عملکرد این تابع برای دو حالت دو متغیره که ساده ترین حالت است به صورت زیر نوشته می‌شود:

$$softmax(z_1, z_2) = \left(\frac{e^{z_1}}{e^{z_1} + e^{z_2}}, \frac{e^{z_2}}{e^{z_1} + e^{z_2}} \right) \quad (۲۴-۳)$$

می توان گفت که ساز و کار این تابع حاصل تقسیم تابع نمایی یک عنصر به حاصل مجموع توابع نمایی تمام عناصرها است، که به صورت نهایی زیر نوشته می شود:

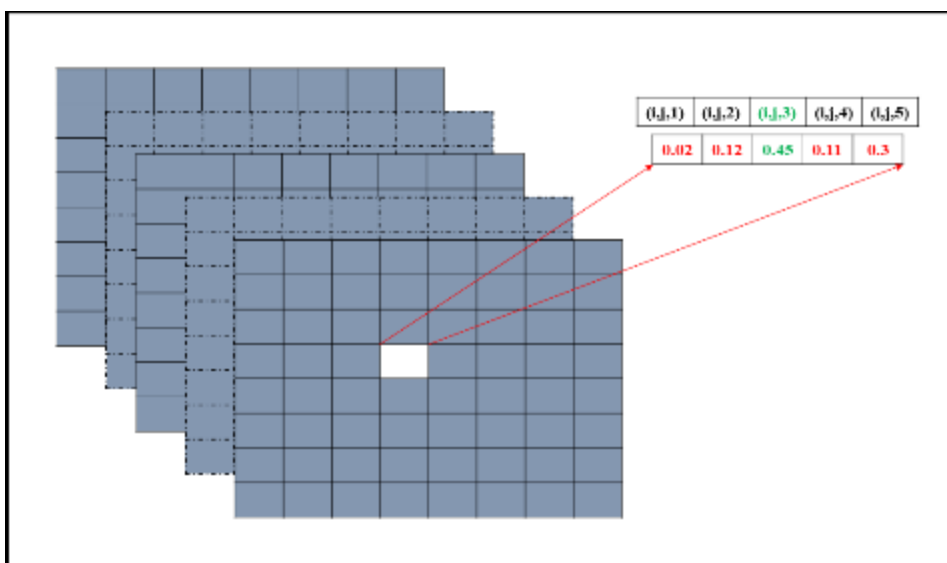
$$\text{softmax}(z_1 \dots z_n) = \left(\frac{e^{z_1}}{e^{z_1} + e^{z_2}} \dots \frac{e^{z_n}}{e^{z_1} + \dots + e^{z_n}} \right) \quad (25-3)$$

تابع softmax در یک CNN به ازای هر پیکسل یک بردار خروجی به ابعاد [1,c] محاسبه می کند که c تعداد کلاس ها است [50]. بر اساس رابطه 3-23 هدف شبکه segnet نیز کمینه کردن معیار loss است.

اگر فرض کنیم پیکسل (i,j) ام توسط segnet طبقه بندی شده، c برابر با 5 بوده و خروجی لایه softmax مانند شکل 4-6 باشد، کلاسی که پیکسل به آن تعلق می گیرد با رنگ سبز مشخص شده و با رابطه زیر کلاس مشخص می شود:

$$\text{classnumber} = \text{argmax} \left(\frac{e^{z_1}}{e^{z_1} + e^{z_2}} \dots \frac{e^{z_n}}{e^{z_1} + \dots + e^{z_n}} \right) \quad (26-3)$$

با رابطه 3-26 شماره کلاسی که پیکسل به آن تعلق می گیرد، مشخص خواهد شد. در شکل 3-8 نیز خروجی لایه softmax نشان داده شده است:



شکل ۳-۸- خروجی لایه softmax، طبق شکل کلاس ۳ کلاسی است که پیکسل به آن تعلق می‌گیرد

طبق شکل ۳-۸ عدم قطعیت پیکسل برابر با جمع مقادیر قرمز است و قطعیت پیکسل عددی که با رنگ سبز نشان داده شده است، پس می‌توان گفت که هر چقدر که عدد سبز کوچکتر باشد عدم قطعیت نیز بیشتر است، تابع مقدار قطعیت را با رابطه زیر تعریف می‌کنیم:

$$\varphi(x) = \max \left(\frac{e^{z_1}}{e^{z_1} + e^{z_2} + \dots + e^{z_n}} \dots \frac{e^{z_n}}{e^{z_1} + e^{z_2} + \dots + e^{z_n}} \right) \quad (27-3)$$

$$= \max(\text{softmax}(z_1 \dots z_n))$$

طبق رابطه ۳-۲۷ عدم قطعیت به دست نخواهد آمد بلکه قطعیت پیکسل طبقه بندی شده بدست می‌آید، رابطه بالا مربوط به یک قطعیت تنها یک پیکسل است اما مشکلی که وجود دارد در CNN ها ورودی باید به صورت قطعه تصویر باشد، اگر قطعه تصویر ورودی را با X و با N پیکسل در نظر بگیریم قطعیت طبقه بندی هر قطعه تصویر با رابطه زیر بدست می‌آید:

(۲۸-۳)

هر قطعه تصویر آموزشی $\emptyset(X)$ مربوط به خود را دارد، اگر مجموعه $T = \{\emptyset_1(X) \dots \emptyset_t(X)\}$ را در نظر بگیریم، قطعه تصویری که برای افزوده شدن به داده‌های آموزشی با یادگیری فعال در نظر گرفته می‌شود، $Q = \min(T)$ ^۱ است، زیرا کمترین قطعیت و در نتیجه بیشترین عدم قطعیت را در بین قطعه تصاویری که در داده‌های pool قرار دارند را داراست.

نوآوری اصلی ما در این تحقیق مربوط به بخش segnet است که ما برای شبکه‌ای از این خانواده آموزش با یادگیری فعال را به کار برده‌ایم، تا به حال در هیچ تحقیق سنجش از دوری یادگیری فعال برای patchهای تصویر به کار گرفته نشده و به طور کلی یادگیری فعال با استفاده از patch ها در سایر علوم مربوط به بینایی ماشین نیز نوع جدیدی است.

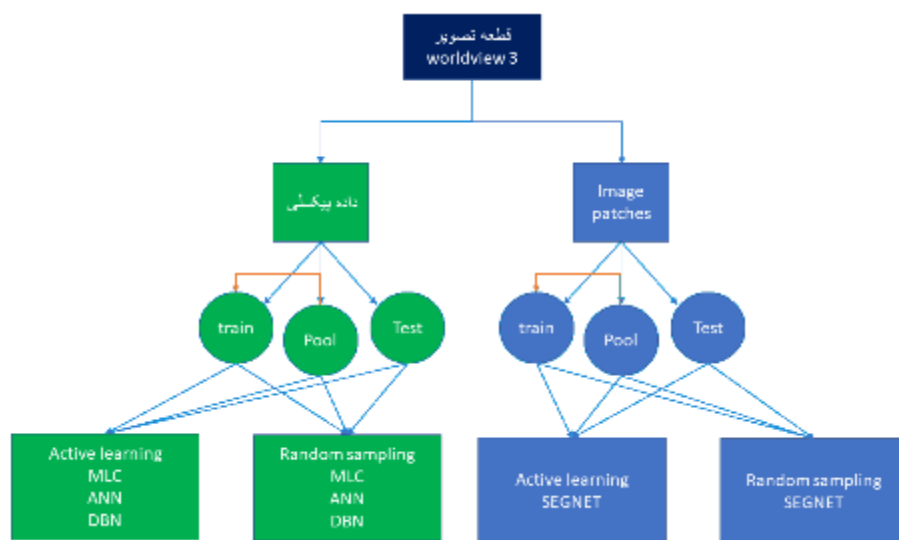
کد نویسی این بخش در بیش از ۱۰۰۰ خط و استفاده از انواع کتابخانه‌های پایتون از قبیل sklearn، skimage و pytorch نوشته شده‌است، بخش یادگیری فعال با استفاده از دو کتابخانه numpy و pytorch انجام شده که به تنهایی ۱۵۰ خط کد را شامل می‌شود.

۳-۶- خلاصه فصل

به طور کلی روش‌های استفاده شده در این تحقیق را می‌توان به دو بخش تقسیم کرد: ۱- روش‌هایی که با استفاده از ویژگی‌های یک پیکسل به طبقه‌بندی تصویر می‌پردازند، ۲- روش SegNET که با استفاده از قطعه تصاویر (مجموعه‌ای از پیکسل‌ها) آموزش دیده و تصاویر را طبقه‌بندی می‌کند.

در روش‌های بخش اول یادگیری فعال قبلاً به کار برده شده‌بود، اما در روش SegNET ما قطعه کدی برای یادگیری فعال طراحی کرده و در فصل بعد به بررسی نتایج حاصل از آن خواهیم پرداخت، فلوجارت تحقیق را نیز می‌توانیم به صورت شکل ۳-۹ نمایش دهیم :

^۱ Query function



شکل ۹-۳ فلوجارت تحقیق - پیکان از داده‌های pool به داده‌های train به علت انتقال داده‌ها در هر تکرار به رنگ متفاوت کشیده شده است

فصل چهارم:

نتایج و ارزیابی

۴-۱-مقدمه

در این فصل به پیاده سازی روش‌های که در فصل ۴ توضیح داده شد پرداخته می‌شود، پس از پیاده سازی روش‌های ذکر شده با استفاده از معیارهای مختلف مورد ارزیابی و مقایسه قرار خواهند گرفت. داده‌های مورد استفاده برای این پایان نامه یک قطعه تصویر ماهواره view world 3 از شهر کرمان است، در روش‌های pixel-base به علت عدم کارایی این روش‌ها (صرف زمان بسیار بالا برای آموزش) این تصویر به ۰/۱ اندازه واقعی تغییر اندازه^۱ داده شده است. روش Random-sampling در این پایان نامه به عنوان روشی برای مقایسه با یادگیری فعال انتخاب شده است، این روش به جای انتخاب همان تعداد داده که با یادگیری فعال به داده‌های آموزشی اضافه می‌شوند، به صورت کاملاً تصادفی همان تعداد داده‌ها به داده‌های آموزشی اضافه می‌کند.

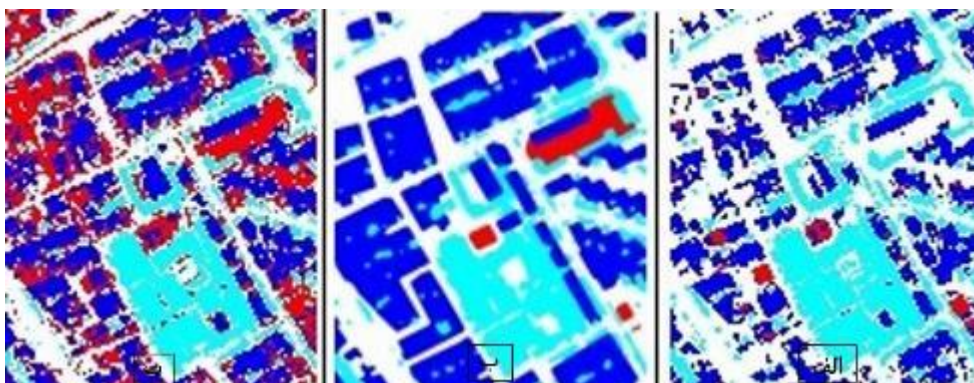
۴-۲-روش MLC

جدول ۴-۱ نشان‌دهنده تعداد داده‌ها در هر کلاس برای آموزش ابتدایی برای آموزش روش MLC است، که میزان داده‌ها آموزش و test و pool را نشان می‌دهد. جدول ۴-۱- توزیع داده‌ها جهت ارزیابی و آموزش

sum	poll	test	Train	
۱۳۲۰۸	۵۵۲۲	۷۳۶۰	۳۰۰	تعداد
%۱۰۰	%۴۲	%۵۶	%۲	نسبت به کل داده‌ها (درصد)

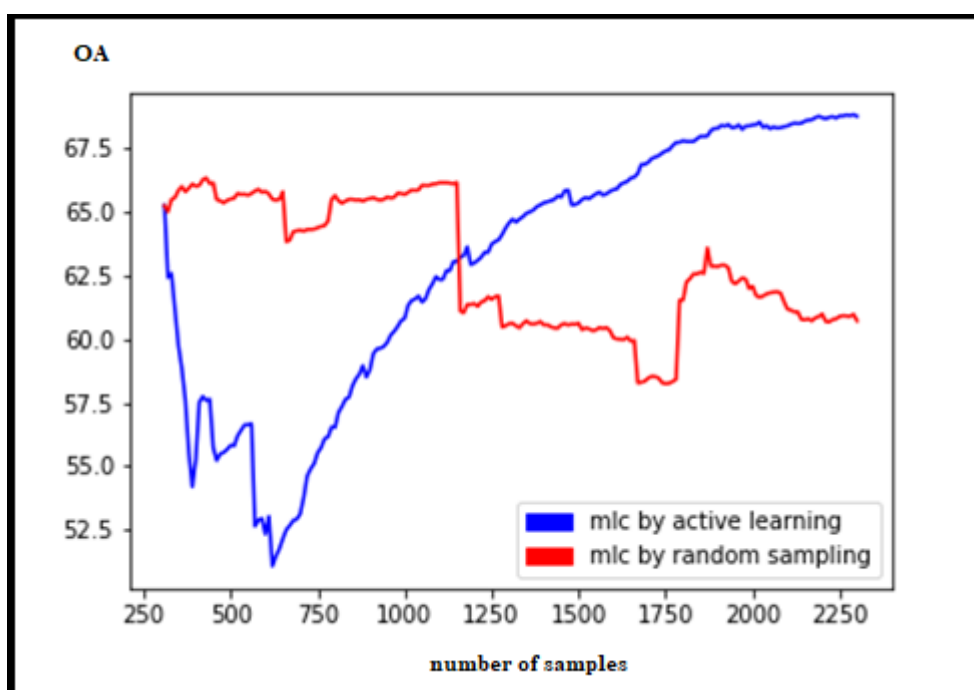
به منظور اجرای یادگیری فعال با MLC تعداد تکرارها برابر ۲۰۰ در نظر گرفته شده است و با هر تکرار ۱۰ پیکسل بر اساس بیشترین عدم قطعیت به داده‌های آموزشی اضافه می‌شوند. به منظور مقایسه یادگیری فعال در ۲۰۰ تکرار و در هر تکرار ۱۰ پیکسل به صورت کاملاً اتفاقی به داده‌های آموزشی اضافه شده‌اند، نتایج در شکل ۵-۱ نمایش داده شده‌اند.

^۱ Resize



شکل ۴-۱. نتایج MLC، الف: Active Learning، ب: Ground truth، پ: Random sampling

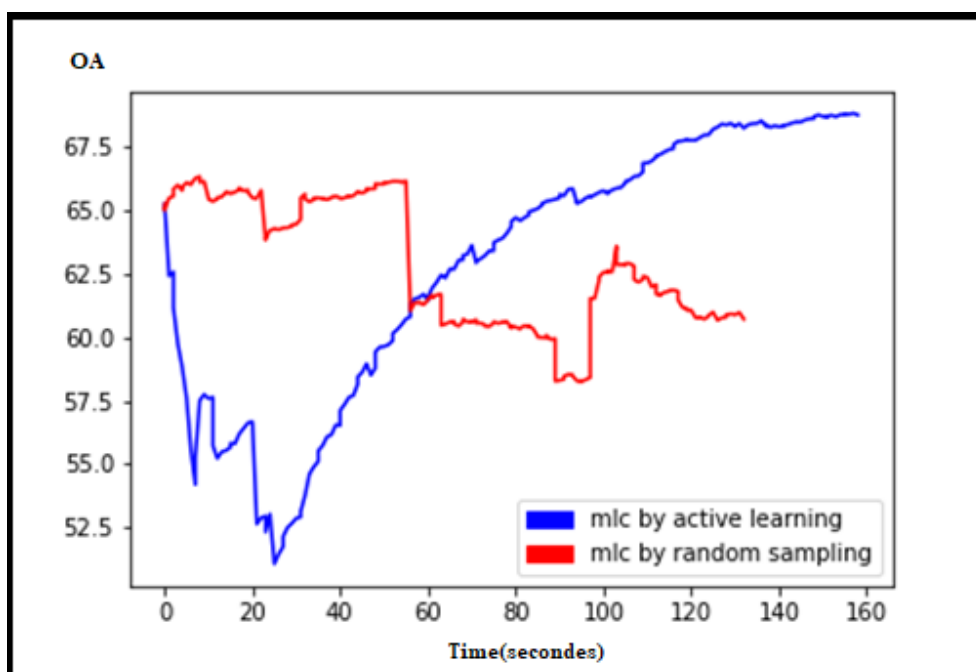
شکل ۴-۲ نمودار Acuracy (صحت) به درصد نسبت به تعداد داده‌های آموزشی را در تکرار-های یادگیری فعال و Random sampling به صورت جداگانه نمایش می‌دهد.



شکل ۴-۲- نمودار دقت-تعداد داده‌های آموزشی MLC، نمودار آبی مربوط به یادگیری فعال و نمودار قرمز مربوط به Random sampling با روش طبقه‌بندی MLC هستند
با توجه به شکل نمودار یادگیری فعال در اولین دوره‌های یادگیری دقت کاهش یافته است و

با رسیدن داده‌های آموزشی به ۳۳۰ عدد، روند عوض شده و دقت حداکثر تا ۶۸/۸۰ درصد افزایش یافته است. این در حالی است که دقت حاصل از روش Random sampling حداکثر به ۶۶/۳۱ درصد رسیده است، همچنین اگر کل داده‌های poll + train را برای آموزش به کار ببریم دقت ۶۳/۵ درصد خواهد شد که نتیجه می‌دهد یادگیری فعال در این روش باعث بهبود دقت نیز شده است.

شکل ۳-۴ نمودار دقت بر اساس مدت زمان بر حسب ثانیه را نمایش می‌دهد، در این بازه زمانی یادگیری فعال و Random sampling در ۲۰۰ تکرار انجام شده‌اند، این بازه زمانی برای یادگیری فعال برابر با ۱۵۸ ثانیه و برای Random sampling برابر با ۱۳۲ ثانیه بوده است.



شکل ۳-۴- نمودار دقت-زمان MLC، محور افقی زمان طی شده از شروع تا پایان یادگیری فعال و محور عمودی میزان دقت را نمایش می‌دهد

برای آموزش روش MLC با همه داده‌های train + poll حدود ۰/۵ ثانیه زمان لازم است و دقت به ۶۳/۵ می‌رسد، با استفاده از یادگیری فعال از ۸۳ ثانیه به بعد نمودار دقت بالا ۶۳/۵ را نشان می‌دهد، بیشترین دقت بدست آمده با یادگیری فعال ۶۸/۸۰ و در حالت Random

sampling برابر با ۶۶/۳۱ بوده است که یادگیری فعال ۵/۳ درصد دقت بهتر و حالت Random sampling ۲/۸ درصد دقت بهتری را نتیجه داده‌اند.

جدول ۲-۴ ماتریس‌های دقت هر کلاس حاصل از یادگیری فعال و Random sampling را نمایش می‌دهد، کلاس‌ها به ترتیب سطح^۱، ساختمان، پوشش گیاهی و کلاس تداخلی (به عوارض خاص در تصویر اختصاص داده می‌شود، مانند زمین‌های ورزشی) هستند.

جدول ۲-۴. جدول نتایج MLC

کلاس / حالت	یادگیری فعال	Random sampling
سطح	۶۴/۴۰	۴۶/۴۹
ساختمان	۶۷/۷۸	۶۹/۹۰
پوشش گیاهی	۹۶/۸۷	۹۳/۰۴
تداخلی	۱۳/۷۹	۸۳/۵۲
OA	۶۸/۸۰	۶۶/۳۲
Kappa	۰/۵۰	۰/۴۵

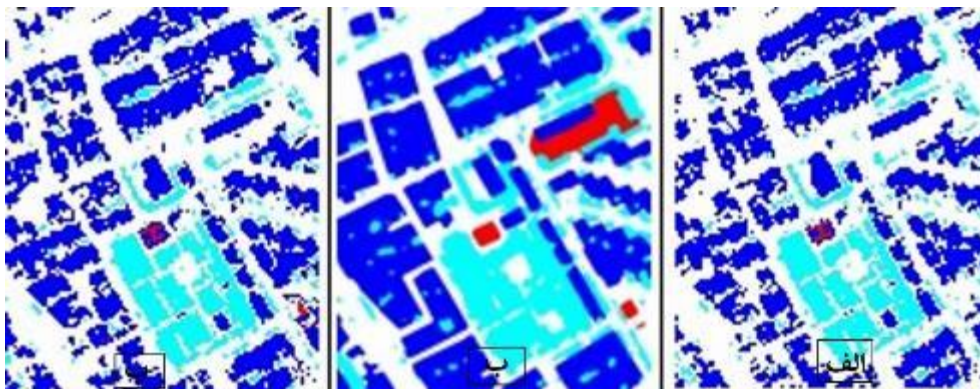
با توجه به جدول ۲-۵ می‌توان دریافت که با یادگیری فعال تعداد پیکسل‌هایی که به اشتباه در کلاس دیگری طبقه‌بندی شده شده‌بودند، کاهش یافته است به طور مثال در کلاس تداخلی با روش Random sampling ۱۸۸۳ پیکسل به اشتباه به عنوان سطح طبقه‌بندی شده‌بودند که در یادگیری فعال به ۱۴۲ پیکسل کاهش یافته است، این امر در سایر کلاس‌ها نیز مشهود است، با توجه به شکل ۱-۵ می‌توان به صحت این ادعا پی برد.

۴-۳- روش ANN

در این روش تعداد Epoch ها برابر با ۱۵۰۰ و نرخ یادگیری برابر با ۰/۰۱ در نظر گرفته

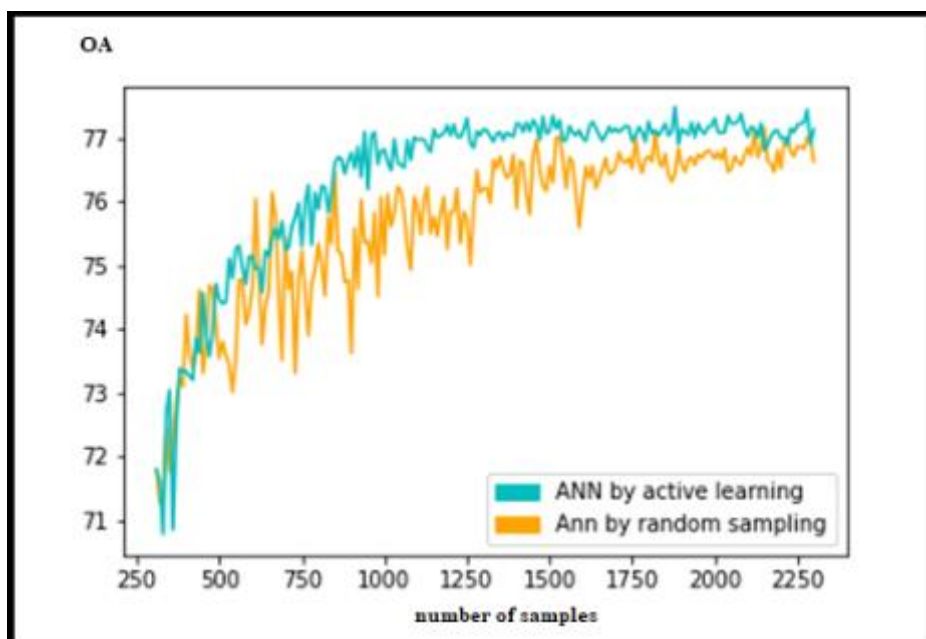
^۱ impervious surface

شده است. در هر تکرار ۱۰ داده به داده‌های آموزشی دور قبل اضافه شده و آموزش دوباره انجام می‌شود، نتایج نیز به منظور مقایسه به صورت دقت (OA) ثبت می‌شوند. داده‌های مورد استفاده برای روش ANN نیز مانند جدول ۴-۱ هستند، شکل ۴-۴ نتایج بدست آمده از یادگیری فعال و Random sampling را نمایش می‌دهد. بر خلاف روش MLC نتایج بدست آمده از این روش تفاوت محسوسی در دو حالت یادگیری فعال و Random sampling ندارند، تقریباً پیکسل‌ها به صورت یکسان در هر دو روش طبقه بندی شده‌اند، اما باز با یادگیری فعال نتایج کمی بهتر هستند.



شکل شماره ۴-۴- نتایج ANN، الف: Active Learning، ب: Ground truth، پ: Random sampling

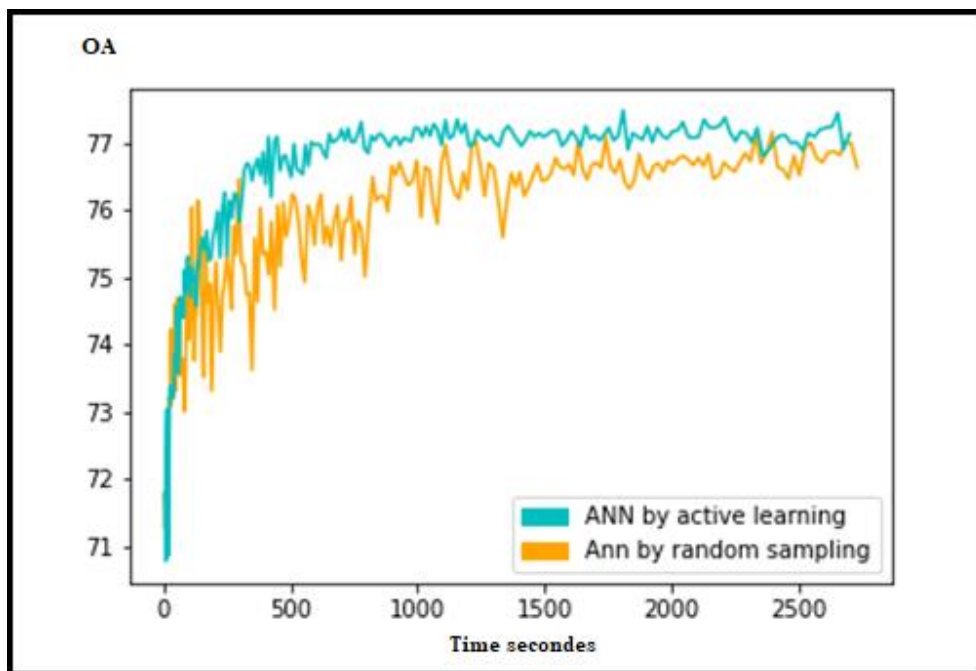
شکل ۴-۵ نمودار دقت در تکرارهای یادگیری فعال و Random sampling با روش طبقه‌بندی ANN را نمایش می‌دهد. در این شکل نیز نمودار یادگیری فعال نیز کمی بالاتر از نمودار Random sampling قرار گرفته است.



شکل ۴-۵- نمودار دقت-تعداد داده‌های آموزشی ANN، نمودار نیلی مربوط به یادگیری فعال و نمودار نارنجی

مربوط به Random sampling بار روش طبقه بندی ANN هستند

نمودار مربوط به روش ANN بر خلاف روش MLC که در یادگیری فعال که در ابتدا در حالت یادگیری فعال روندی نزولی داشت و سپس روند صعودی به خود گرفت دارای یک روند صعودی یکنواخت همراه با نوسانات کوچک است، نمودار Random sampling نیز روندی صعودی دارد اما دقت بدست آمده در این حالت اغلب کمتر از روش یادگیری فعال بوده‌است.



شکل ۴-۶- نمودار دقت-زمان ANN، محور افقی زمان طی شده از شروع تا پایان یادگیری فعال و محور عمودی

میزان دقت را برای روش ANN نمایش می‌دهد

شکل ۴-۳ که زمان سپری شده در یادگیری فعال و Random sampling را نشان می‌دهد، بیانگر این است که زمان‌های لازم برای هر دو روش تقریباً برابر است و می‌توان مشاهده کرد که زمان مورد نیاز برای انجام یادگیری فعال با ۲۰۰ تکرار برابر با ۲۷۰۴ ثانیه است، در حالی که همین مقدار برای حالت Random sampling برابر با ۲۷۳۴ ثانیه بوده است.

اگر آموزش ANN با همه داده‌های train + pool انجام شود زمان آموزش برابر با ۵۳/۲۷ ثانیه است و بیشترین دقتی که بدست آمده برابر ۷۷/۲۴ درصد است، در حالت Random sampling به این دقت دست یافته نشده است، اما یادگیری در زمان ۱۶۵۰ ثانیه، با داده‌های آموزشی ۱۷۹۰ به دقت ۷۷/۳۱ رسیده است، بیشترین دقت بدست آمده با یادگیری فعال نیز برابر با ۷۷/۴۹ درصد است که نسبت به حالتی که از تمام داده‌های train + pool برای آموزش استفاده کردیم ۰/۱۸ بیشتر است.

جدول ۴-۳ نتایج بدست آمده از دو حالت Random sampling و یادگیری فعال با روش

ANN را نشان داده‌اند.

جدول ۳-۴. جدول نتایج ANN

Random sampling	یادگیری فعال	کلاس / حالت
۷۷/۴۹	۷۹/۶۵	سطح
۷۷/۵۵	۷۷/۶۷	ساختمان
۷۹/۳۱	۷۸/۷۵	پوشش گیاهی
۸/۴۲	۶/۹۰	تداخلی
۷۷/۱۶	۷۷/۴۸	OA
۰/۶۰	۰/۶۱	Kappa

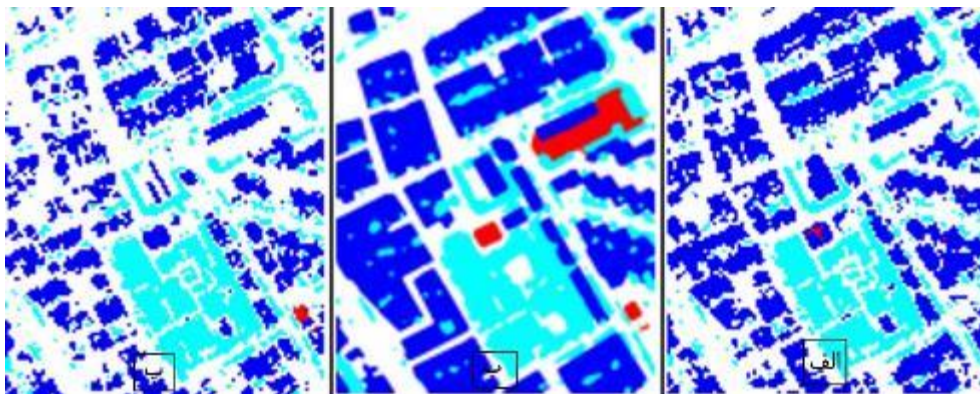
با توجه به جدول ۳-۴ می‌توان دریافت که دقت کلی برای هر کلاس به غیر از کلاس تداخلی با یادگیری فعال نسبت به حالت Random sampling افزایش داشته‌است، به طور مثال برای کلاس پوشش گیاهی در حالت Random sampling، ۳۳۵ پیکسل به اشتباه به عنوان سطح طبقه‌بندی شده بودند اما در حالت یادگیری فعال به ۲۰۳ پیکسل کاهش یافته است، همچنین تعداد پیکسل‌هایی که به درستی طبقه‌بندی شده‌اند نیز در کلاس سطح که بیشترین پیکسل‌های تصویر را دارا است افزایش یافته است.

۴-۴-روش DBN

شبکه DBN مورد استفاده در این پایان نامه دارای سه RBM است که تعداد عناصر استخراج شده هر کدام به ترتیب ۱۵ و ۱۰ و ۵ عنصر است، در پایان نیز به کمک یک Regressor خطی نتایج محاسبه می‌شوند، تعداد تکرار در هر RBM برابر با ۱۵۰ در نظر گرفته شده است. داده‌های مورد استفاده در این روش نیز طبق جدول ۴-۱ بوده و در هر تکرار ۱۰ داده به داده‌های آموزشی دور قبل اضافه شده‌اند.

شکل ۴-۷ نتایج بدست آمده از این روش را در دو حالت یادگیری فعال و Random sampling

نمایش می‌دهد، در این روش نیز مقداری دقت بدست آمده در یادگیری فعال نسبت به Random sampling بالاتر است.

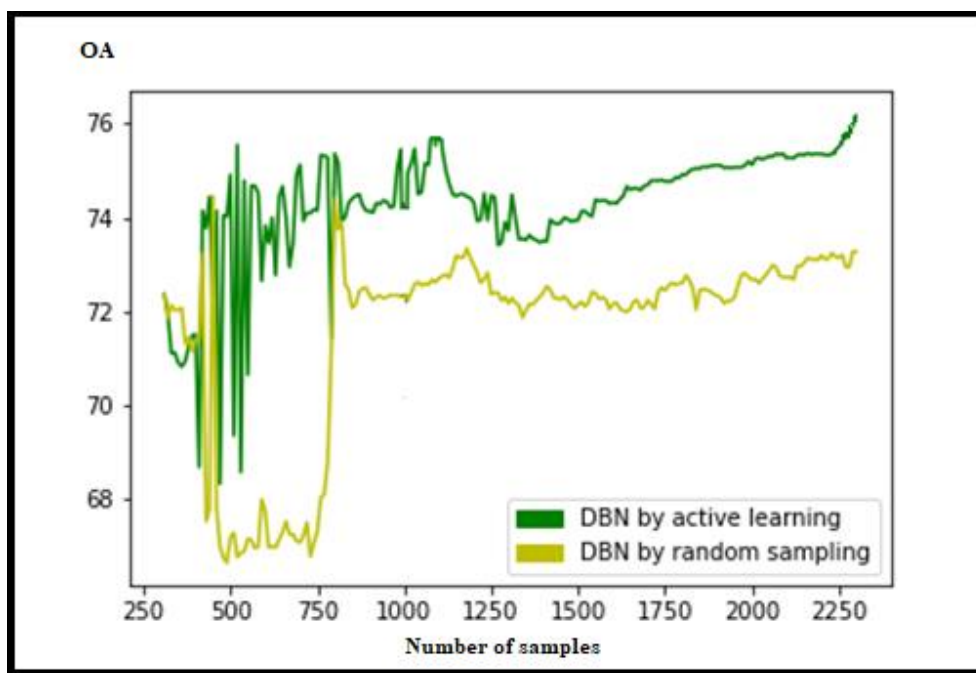


شکل شماره ۴-۷- نتایج DBN، الف: Active Learning، ب: Ground truth، پ: Random sampling

از مقایسه نتایج می‌توان به این پی برد که DBN در تشخیص کلاس تداخلی بسیار ضعیف عمل کرده‌است، ولی در کلاس پوشش گیاهی عملکرد نسبتاً خوبی داشته است.

شکل ۴-۸ نمودار دقت بدست آمده با DBN در دو حالت Random sampling و یادگیری فعال را بر اساس تعداد داده‌های آموزشی نشان می‌دهد، نمودارها در دو حالت تا زمانی که تعداد داده‌ها ۵۰۰ است عملکرد مشابهی دارند اما پس از آن نمودار Random sampling به کلی پایین‌تر از نمودار یادگیری فعال قرار می‌گیرد.

هر دو نمودار حین افزایش داده‌ها نوسان‌های ناگهانی به سمت پایین دارند که به علت معرفی داده‌های ناشناخته به روش DBN است، اما یادگیری فعال با اضافه کردن پیکسل‌های مناسب به داده‌های آموزشی این نوسان را جبران کرده و به تبع دقت را نیز بالاتر می‌برد. در صورتی که حالت Random sampling با اضافه شدن داده‌ها فقط حالتی نوسانی پیدا کرده‌است. بالاترین دقت بدست آمده در حالت یادگیری فعال ۷۶/۲۴ درصد و در حالت Random sampling برابر با ۷۴/۴۶ درصد بوده‌است، در صورتی که دقت بدست آمده در حالتی که همه داده‌های train و pool را برای آموزش به کار ببریم برابر با ۷۳/۴۰ درصد است. حالت یادگیری فعال ۲/۲۴ درصد و حالت Random sampling ۱/۰۶ درصد دقت بالاتری را نتیجه داده‌اند.



شکل ۴-۸- نمودار دقت-تعداد داده‌های آموزشی DBN، نمودار سبز مربوط به یادگیری فعال و نمودار دیگر

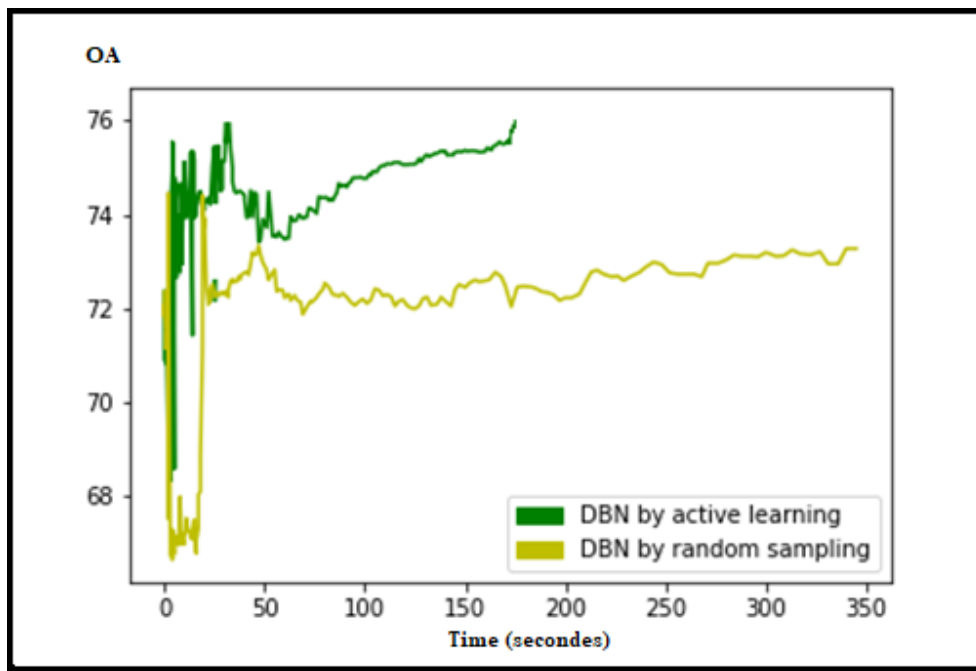
مربوط به Random sampling بار روش طبقه بندی DBN هستند

شکل ۴-۹ نیز نمودار دقت بر حسب زمان را نمایش می‌دهد، در این روش نیز مانند ANN یادگیری فعال سریعتر به پایان رسیده‌است که به علت استفاده از مازول برنامه نویسی سریعتر برای یادگیری فعال است. یادگیری فعال ۱۷۱ ثانیه و حالت Random sampling ۳۷۵ ثانیه برای ۲۰۰ تکرار زمان برده‌اند.

هم چنین یادگیری فعال برای رسیدن به دقت $73/40$ درصد ۱۷ ثانیه زمان برده‌است و Random sampling نیز در ثانیه ۳۳ به این دقت رسیده‌است، این در حالی است که در صورت استفاده از همه داده‌های train + poll آموزش $38/5$ ثانیه طول کشیده‌است، این موضوع می‌تواند بیانگر این باشد که برخی از داده‌ها اگر برای آموزش استفاده شوند ممکن است دقت را پایینتر آورند.

کمت‌ر شدن زمان نیز می‌تواند به علت این باشد که احتمالات به صورت آنی در روش DBN توسط مازولی که از GPU برای محاسبات استفاده می‌کند و DBN با این مازول اجرا می‌شود،

محاسبه می‌شوند و عدم قطعیت آن مشخص می‌شود، حال آن که Random sampling باید با یک ماژول جداگانه و با سرعتی کمتر از ماژول DBN عملیاتی را انجام دهد و تعدادی داده انتخاب کند که خود زمان بیشتری لازم دارد.



شکل ۴-۹- نمودار دقت-زمان DBN، محور افقی زمان طی شده از شروع تا پایان یادگیری فعال و محور عمودی میزان دقت را برای روش DBN نمایش می‌دهد

جدول ۴-۴ ماتریس‌های confusion در دو حالت یادگیری فعال و Random sampling را نشان می‌دهد، با توجه به این ماتریس‌ها کلاس ساختمان و پوشش گیاهی در حالت یادگیری فعال نسبت به Random sampling دقت بالاتری داشته‌اند، اما کلاس سطح با کمی کاهش دقت همراه بوده‌است. کلاس تداخلی در هر دو حالت بد ظاهر شده و دقت خوبی نشان نمی‌دهد اما در حالت Random smpling کمی وضعیت بهتری دارد.

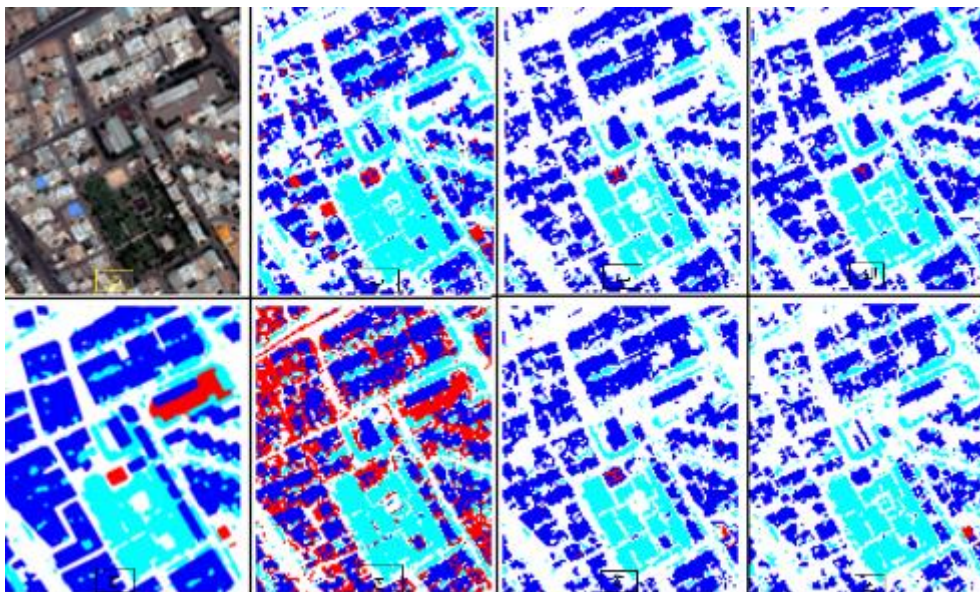
جدول ۴-۴- جدول نتایج DBN

کلاس / حالت	یادگیری فعال	Random sampling
سطح	۷۳/۵۷	۷۸/۳۴
ساختمان	۷۸/۷۷	۶۱/۵۴
پوشش گیاهی	۸۶/۴۶	۸۹/۹۷
تداخلی	۳/۴۵	۵/۳۶
OA	۷۶/۲۴	۷۴/۴۵
Kappa	۰/۵۹	۰/۵۴

۴-۵- مقایسه روش‌های پیکسل مبنا

شکل ۴-۱۰ مقایسه نتایج بدست آمده از سه روش ذکر شده را نمایش می‌دهد که هر سه از روش‌هایی بودند که بر اساس ویژگی‌های یک پیکسل به آن برجستگی اختصاص می‌دادند، از بین این روش‌ها ANN در حالت یادگیری فعال بالاترین دقت و پایینترین دقت مربوط به روش MLC در حالت Random sampling است.

از شکل ۴-۱۰ مشخص است که همه روش‌ها در تشخیص کلاس تداخلی مشکل داشته و به علت شباهت این کلاس با کلاس‌های دیگر این کلاس را به غلط طبقه‌بندی کرده‌اند، با مقایسه نتایج بدست آمده از یادگیری فعال و Random sampling نیز می‌توان متوجه شد که همه روش‌ها در حالت یادگیری فعال نتایج بهتری را بدست آورده‌اند.



شکل ۴-۱۰ نتایج کلی روش‌های pixel-base، الف: DBN یادگیری فعال، ب: ANN یادگیری فعال، پ: MLC

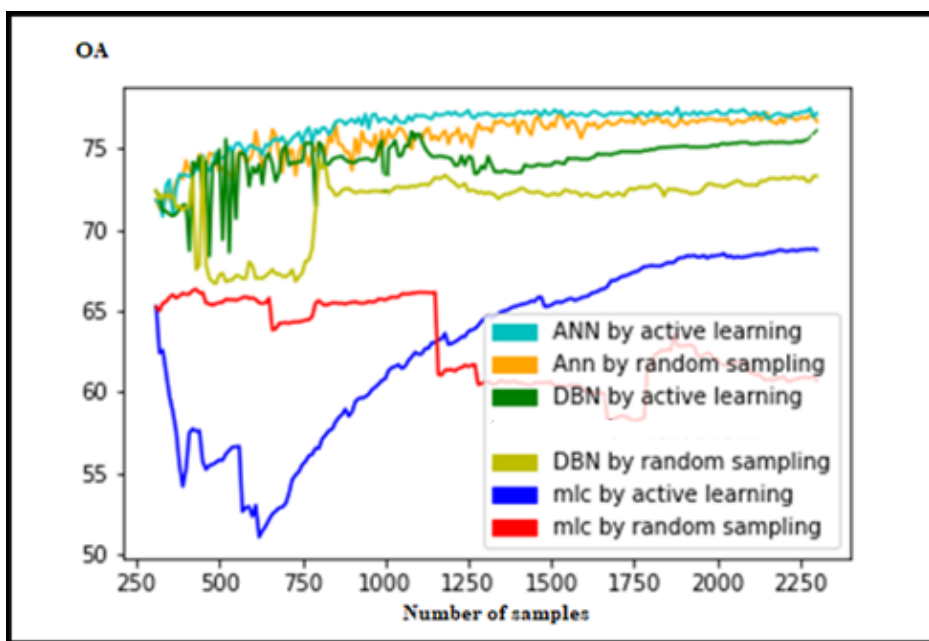
یادگیری فعال، ت: Image، ث: DBN Random sampling، ج: ANN Random sampling، چ: MLC Random

sampling، ح: Ground truth

شکل ۴-۱۱ نمودارهای دقت بر حسب تعداد داده‌های آموزشی را نشان می‌دهد، در همه-روش‌ها در حالت یادگیری فعال با بالاتر رفتن تعداد داده‌ها دقت نیز افزایش می‌یابد، اما در حالت Random sampling فقط در روش ANN روند صعودی است که باز هم نمودار پایین‌تر از نمودار یادگیری فعال قرار می‌گیرد.

بالاترین دقت بدست آمده با روش ANN و حالت یادگیری فعال بوده که دقت برابر با ۷۷/۴۹ درصد بوده‌است و کمترین دقت با روش MLC در حالت Random sampling بدست آمده که برابر با ۶۶/۳۱ درصد است.

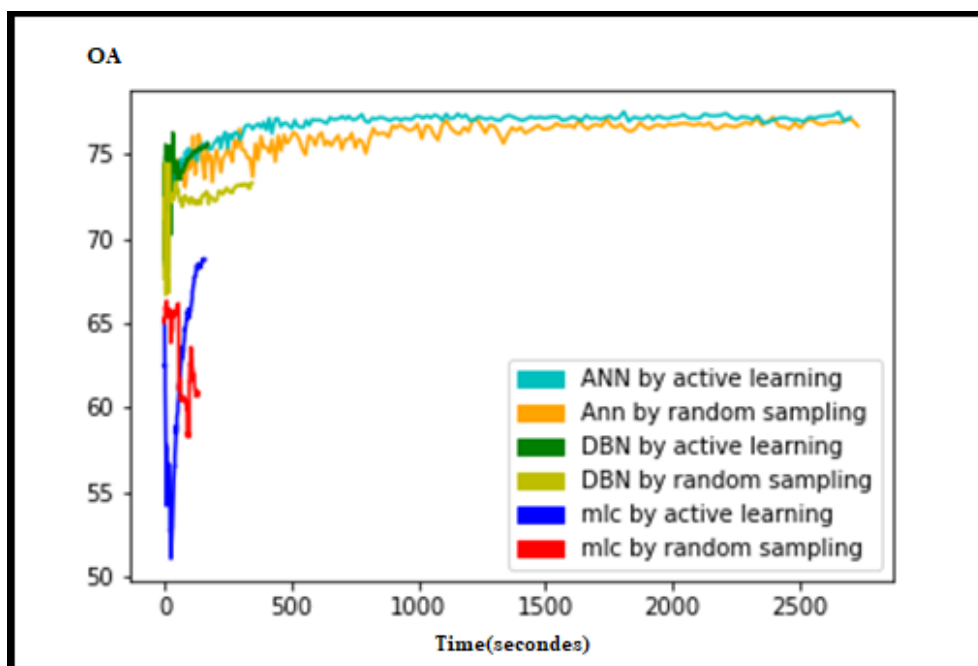
بیشترین افزایش دقت با یادگیری فعال در روش ANN مشاهده می‌شود که نمودار آن در ابتدای آموزش دقت ۶۷/۵ درصد را نشان می‌دهد، این در حالی است که بیشترین دقت بدست آمده با ANN برابر با ۷۷/۴۹ درصد است که نشان از افزایش ۱۰ درصدی دقت است.



شکل ۴-۱۱- نمودار دقت بدست آمده بر حسب تعداد داده‌های آموزشی با روش‌های ANN، DBN و MLC در

حالت‌های یادگیری فعال و Random sampling

شکل ۴-۱۲ دقت بدست آمده بر حسب زمان طی شده در همه حالات در روش‌های ذکر شده را نشان می‌دهد، کمترین زمان صرف شده برای روش MLC در Random smpling و بیشترین زمان مربوط به روش ANN در حالت Random sampling است.



شکل ۴-۱۲- نمودار دقت بدست آمده بر حسب زمان صرف شده با روش‌های ANN، DBN و MLC در حالت-

های یادگیری فعال و Random sampling

با توجه به جدول ۵-۵ می‌توان مشاهده کرد که روش ANN در بین همه روش‌ها بالاترین ضریب Kappa و دقت را نتیجه داده‌است، حالت Random sampling با MLC به علت مشکل در طبقه‌بندی کلاس تداخلی پایین‌ترین مقدار Kappa را از میان سایر روش‌ها نتیجه داده‌است.

جدول ۵-۴- مقایسه نتایج روش‌های Pixel base

DBN RANDOM	DBN ACTIVE	ANN RANDOM	ANN ACTIVE	MLC RANDOM	MLC ACTIVE	کلاس / حالت
۷۸/۳۴	۷۳/۵۷	۷۷/۴۹	۷۹/۶۵	۴۶/۴۹	۶۴/۴۰	سطح
۶۱/۵۴	۷۸/۷۷	۷۷/۵۵	۷۷/۶۷	۶۹/۹۰	۶۷/۷۸	ساختمان
۸۹/۹۷	۸۶/۴۶	۷۹/۳۱	۷۸/۷۵	۹۳/۰۴	۹۶/۸۷	پوشش گیاهی
۵/۳۶	۳/۴۵	۸/۴۲	۶/۹۰	۸۳/۵۲	۱۳/۷۹	تداخلی
۷۴/۴۵	۷۶/۲۴	۷۷/۱۶	۷۷/۴۸	۶۶/۳۲	۶۸/۸۰	OA
۰/۵۴	۰/۵۹	۰/۶۰	۰/۶۱	۰/۴۵	۰/۵۰	Kappa

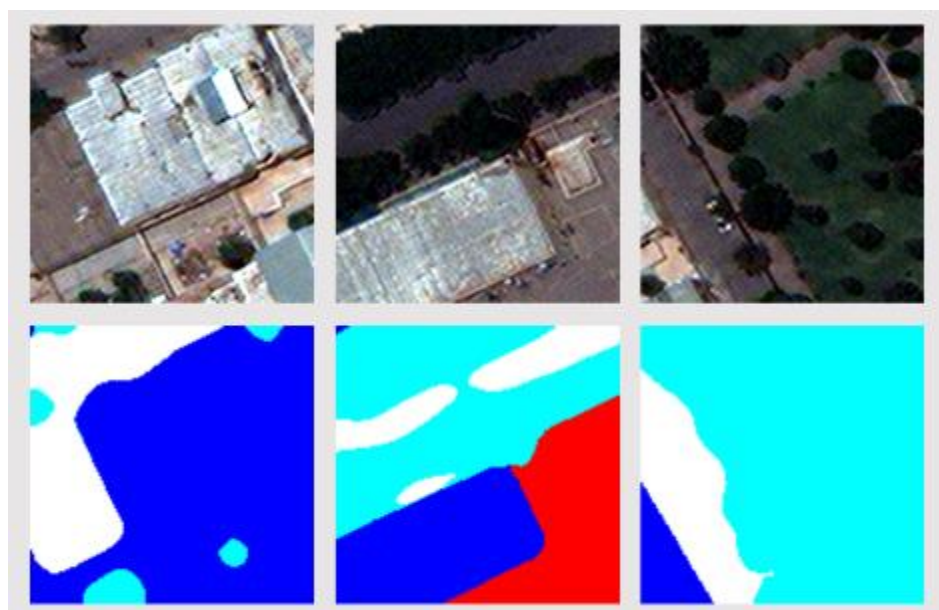
۴-۶- Segnet روش

داده‌های روش Segnet بر خلاف روش‌هایی که تا حالا در این پایان نامه مورد بررسی قرار گرفتند به صورت قطعه تصویر هستند، و هر پیکسل بر اساس ماهیت خود پیکسل و پیکسل‌های اطراف برچسبی را به خود اختصاص می‌دهد.

داده‌های مورد استفاده در این شبکه قطعه تصویرهایی با اندازه 128×128 پیکسل هستند، که سه نمونه اولیه را می‌توان در شکل ۴-۱۳ مشاهده کرد. یک قطعه تصویر بزرگتر 1037×1265 پیکسل برای test در نظر گرفته شده‌است، داده‌های مورد استفاده در این بخش مطابق جدول ۵-۶ هستند.

جدول ۴-۶- داده‌های مورد استفاده برای Segnet

sum	poll	test	train	
۳۸	۳۵ قطعه تصویر 128×128	یک تصویر بزرگ (۷۲ قطعه تصویر)	۳ قطعه تصویر 128×128	تعداد



شکل ۴-۱۳- قطعه‌تصویرهای آموزشی، در ابتدا این قطعه‌تصویرها به عنوان داده آموزشی استفاده شده و سپس

به تعداد آن‌ها در هر دور افزوده شده‌است، بالا (تصویری اصلی)، پایین (Ground truth)

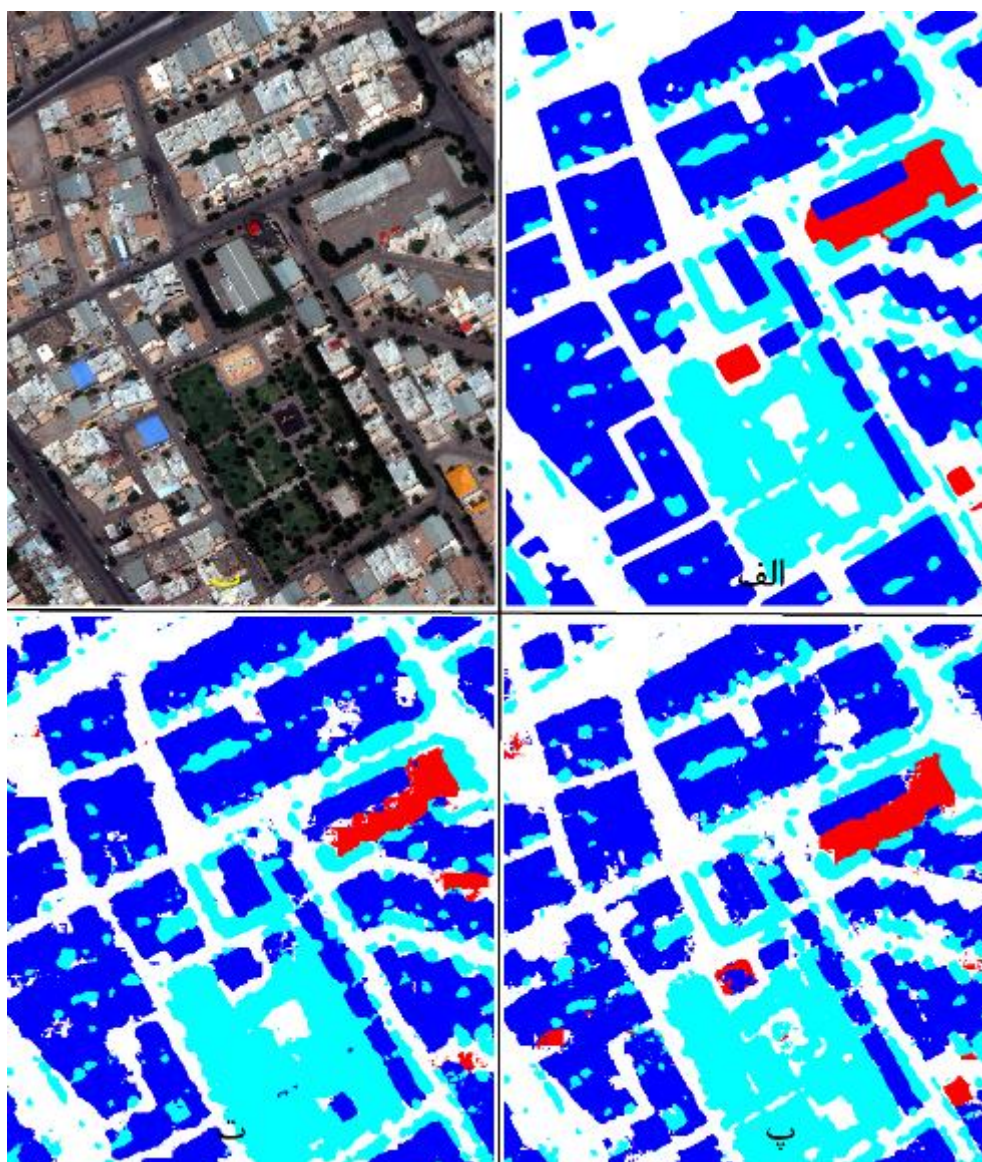
آموزش شبکه با ۳ قطعه تصویر شروع شده و در هر تکرار یادگیری فعال و یا Random sampling در هر تکرار ۲ قطعه تصویر به داده‌های آموزشی اضافه شده است تا تعداد داده‌های آموزشی به ۲۳ عدد رسیده است.

قطعه تصویرها به گونه‌ای انتخاب شده‌اند که همه‌ی کلاس‌های موجود در تصویر بزرگتر را شامل شوند، قطعه تصاویر مورد استفاده در این بخش از پایان نامه از ماهواره 3 World view با دقت مکانی^۱ ۱/۲۴ تهیه شده‌اند، تصویر test از همان منطقه تصویر مورد استفاده در سه روش pixel-base است، با این تفاوت که قدرت تفکیک در این تصویر و قطعه تصویرهای آموزشی ۱۰ برابر بالاتر است.

نتایج بدست آمده در هر دو حالت در شکل ۴-۱۴ مشاهده می‌شوند. با توجه به این شکل نیز بهتر بودن نتیجه حاصل از یادگیری فعال مشخص است، نکته قابل توجه این است که کلاس تداخلی در این حالت یادگیری فعال بسیار خوب شناسایی شده‌است.

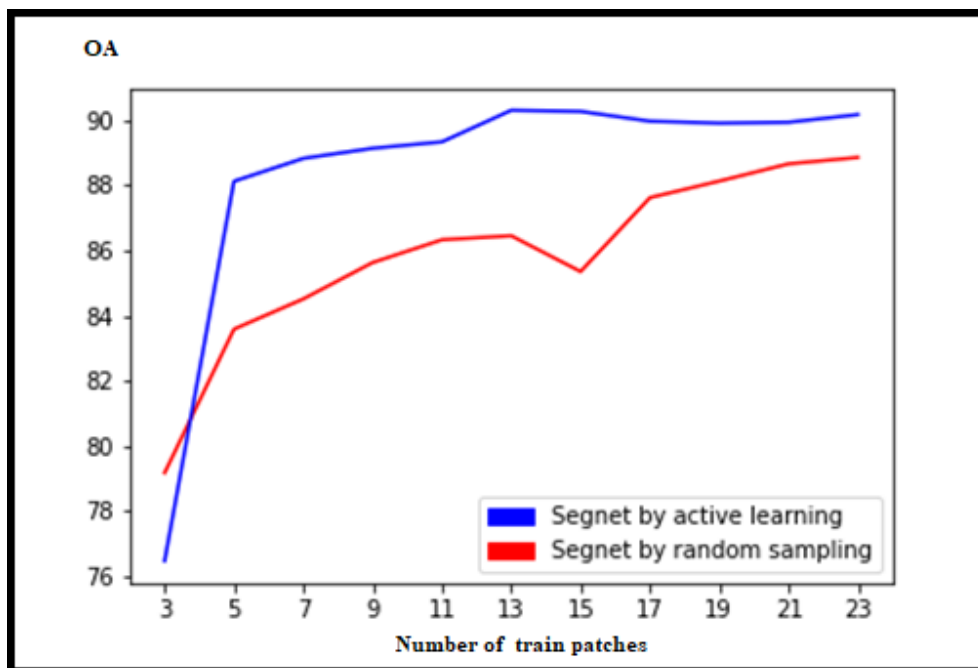
کلاس‌های دیگر نیز در حالت یادگیری فعال نسبت به روش Random sampling بهتر شناسایی شده‌اند، زمانی که از همه داده‌های pool و train برای آموزش استفاده کنیم، دقت بدست آمده برابر با ۸۵/۵۵ است، بالاترین دقت بدست آمده در روش Random sampling برابر با ۸۸/۸۶ درصد است که نسبت به حالت استفاده از همه داده‌ها ۳/۳۱ درصد بهتر است و زمانی که از یادگیری فعال استفاده می‌شود بالاترین دقت بدست آمده برابر ۹۰/۱۷ درصد است که نسبت به حالت Random sampling ۱/۳۱ درصد و نسبت به حالت استفاده از همه داده‌ها ۴/۶۲ درصد دقت بالاتری را نتیجه داده است.

^۱ Spatial Resolution



شکل ۴-۱۴- نتایج Segnet، الف: Ground truth، ب: Image، پ: یادگیری فعال، ت: Random sampling

شکل ۴-۱۵ نمودار دقت بر اساس اعداد قطعه تصویرهایی که برای آموزش استفاده شده‌اند را نشان می‌دهد، به وضوح با مقایسه این نمودارها برتری یادگیری فعال به Random sampling مشخص است.



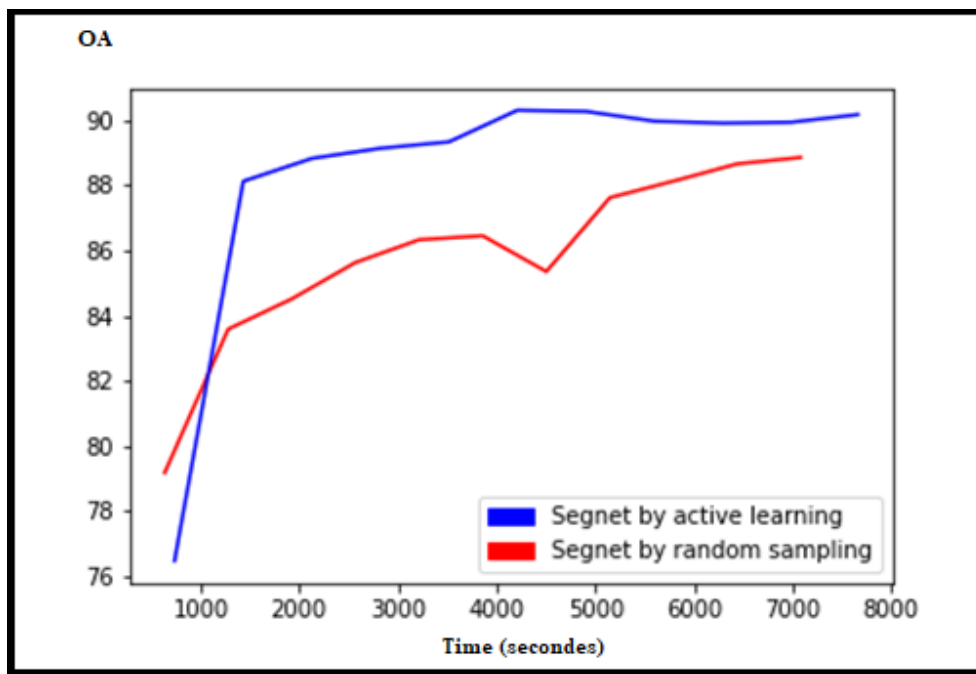
شکل ۴-۱۵- نمودار دقت بر اساس تعداد قطعه تصویر آموزشی با شبکه Segnet

دقت بدست آمده زمانی که از همه داده‌های pool و train استفاده می‌شود برابر با ۸۵/۵۵ است، در اولین تکرار یادگیری فعال با ۳ قطعه تصویر آموزشی دقت ۷۶/۱۳ درصد بدست آمده- است، این در حالی است که در تکرار دوم و زمانی که تعداد قطعه تصویرها به ۵ عدد می‌رسد دقت برابر با ۸۶/۱۳ درصد است که از دقت بدست آمده در حالت آموزش با تمام داده‌ها ۰/۵۸ درصد بالاتر است.

زمانی که در حالت Random sampling تعداد قطعه تصویرهای آموزشی به ۱۱ عدد می‌رسد نیز دقت به ۸۵/۶۳ درصد رسیده است که از دقت در حالت آموزش با تمام داده‌ها بالاتر است، علت افزایش دقت در هر دو حالت استفاده شبکه Segnet از Checkpoint ها برای ذخیره وزن- هایی است که نتایج درست را محاسبه کرده‌اند و در واقع در هر تکرار یادگیری فعال و Random smpling وزن‌های ذخیره شده در Checkpoint های قبلی بروز می‌شوند و این موضوع مربوط به ذات Segnet است و ارتباطی با یادگیری فعال ندارد.

شکل ۴-۱۶ نمودار دقت بدست آمده نسبت به زمان طی شده از شروع اولین تکرار را نشان

می‌دهد، با کمی توجه می‌توان متوجه شد که زمان صرف شده برای حالت یادگیری فعال از زمان صرف شده در حالت Random sampling بیشتر است، اما دقت بالاتری را نیز نتیجه داده‌است.



شکل ۴-۱۶- دقت بدست آمده در دو حالت یادگیری فعال و Random sampling بر اساس زمان

در زمان ۱۴۳۳ ثانیه دقت یادگیری فعال به بالاتر از دقت بدست آمده از حالتی که از تمام داده‌ها برای آموزش استفاده کرده‌ایم، رسیده است. این در حالی است که در حالت Random sampling در زمان ۲۵۶۸ ثانیه دقت به کمی بالاتر از این حد رسیده است، گفتنی است که با استفاده از همه داده‌های poll + train برای آموزش در زمان ۶۲۶ ثانیه به دقت ۸۵/۵۵ درصد رسیده‌است.

جدول ۴-۷ ماتریس‌های Confusion بدست آمده در دو حالت Random sampling و یادگیری فعال برای شبکه Segnet را نشان می‌دهد، در ادامه به مقایسه این دو ماتریس خواهیم پرداخت.

جدول ۷-۴. بالا : ماتریس confusion، بالا: Randm sampling ، پایین: یادگیری فعال روش Segnet

	سطح	ساختمان	پوشش گیاهی	تداخلی
سطح	۳۳۲۹۵۳	۵۲۶۴۴	۲۱۰۰۱	۵۴۲۴
ساختمان	۲۶۰۴۹	۴۸۶۱۲۰	۷۱۰۸	۹۴
پوشش گیاهی	۳۵۸۸	۴۸۶۸	۲۴۳۷۱۲	۱۰۲
تداخلی	۴۷۷۵	۹۱۱۰	۱۰۱۹	۱۹۹۹۹

	سطح	ساختمان	پوشش گیاهی	تداخلی
سطح	۳۶۶۶۰۴	۲۹۰۷۱	۱۳۸۹۵	۲۴۵۲
ساختمان	۴۵۴۷۰	۴۶۴۵۸۶	۷۵۰۳	۱۸۱۲
پوشش گیاهی	۶۶۴۶	۴۷۶۸	۲۴۰۷۶۳	۹۳
تداخلی	۴۳۷۸	۳۱۳۱	۵۵۵	۲۶۸۳۹

$$K_r = 0.66$$

$$K_a = 0.81$$

با توجه به ماتریس‌های بدست آمده در حالت یادگیری فعال دقت بدست آمده در حالتی که از یادگیری فعال استفاده می‌شود بالاتر از روش Random sampling است، با بررسی ضریب کاپا که این ضریب برای حالت یادگیری فعال با K_a و برای حالت Random sampling با K_r نشان داده شده‌اند، متوجه می‌شویم که اعتماد پذیری نتایج نیز در صورت استفاده از یادگیری فعال بالاتر رفته است.

امتیاز F1 برای هر کلاس در جدول ۷-۴ آورده شده‌اند، این امتیاز روش مناسبی جهت مقایسه صحت و دقت هر کلاس است، برای بررسی توانایی شناسایی کلاس‌ها در هر دو حالت از این ضریب استفاده می‌کنیم.

جدول ۴-۸- امتیازهای F1 بدست آمده برای هر کلاس در هر دو حالت Random sampling و یادگیری

فعال

کلاس/حالت	Random sampling	یادگیری فعال
سطح	۰/۸۵	۰/۸۸
ساختمان	۰/۹۱	۰/۹۱
پوشش گیاهی	۰/۹۳	۰/۹۴
کلاس تداخلی	۰/۶۶	۰/۸۱

با توجه به جدول ۴-۸ در همه کلاس‌ها یادگیری فعال ضریب F1 بالاتری را بدست آورده است، در کلاس تداخلی اختلاف محسوس‌تری بین ضرایب F1 یادگیری فعال و Random sampling وجود دارد که خود دلیلی بر برتری یادگیری فعال است.

جدول ۴-۹- جدول نتایج Segnet

کلاس / حالت	یادگیری فعال	Random sampling
سطح	۸۸/۹۸	۸۰/۸۱
ساختمان	۹۱/۴۵	۹۳/۶۰
پوشش گیاهی	۹۵/۴۴	۹۶/۶۰
تداخلی	۷۶/۹۰	۵۷/۳۰
OA	۹۰/۱۷	۸۸/۸۶
Kappa	۰/۸۱	۰/۶۶

با توجه به جدول ۵-۹ می‌توان بیان کرد که اکثر کلاس‌ها و کلاس تداخلی در حالت یادگیری فعال دقت‌های بالاتری را بدست آورده‌اند، با توجه با ماهیت کلاس تداخلی با بالاتر رفتن دقت این کلاس معمولاً از دقت بقیه کلاس‌ها کاسته می‌شود، در جدول بالا نیز این موضوع مشاهده

می‌شود، در روش‌های Pixel-base روش MLC در حالت Random بالاترین دقت را برای این کلاس بدست آورده بود، اما در روش MLC با کاهش شدید دقت همراه بوده‌است، اما با روش Segnet در حالت یادگیری فعال دقت‌ها به میزان بسیار کمی کاسته شده است و دقت کلاس تداخلی از ۵۷/۳۰ درصد به ۷۶/۹۰ درصد (۱۹/۶ درصد) افزایش یافته‌است.



شکل ۴-۱۷-فلوچارت نتایج تحقیق

فصل پنجم:

نتیجه گیری

۵-۱-مقدمه

بر اساس نتایج بدست آمده در فصل پنجم یادگیری فعال در روش‌های pixel base مانند ANN، MLC و DBN به خوبی عمل می‌کند، اکنون در این فصل به مقایسه نتایج بدست آمده در این تحقیق، با نتایج بدست آمده در تحقیقات دیگر می‌پردازیم. سپس اثر بخشی یادگیری فعال در شبکه Segnet را مورد بررسی قرار می‌دهیم و در آخر نتیجه‌گیری کلی برای استفاده از یادگیری فعال را بیان می‌کنیم.

۵-۲-نتیجه‌گیری در رابطه با یادگیری فعال با روش MLC

بر اساس نتایج بدست آمده (شکل‌های ۵-۱ تا ۵-۳ و جدول ۵-۲) یادگیری فعال برای روش MLC نتیجه مثبتی داشته است، براساس تحقیقات دیگر و مقایسه نتایج بدست آمده در این تحقیق با تحقیقات دیگر به این نتیجه می‌رسیم که یادگیری فعال ابتدا دقت طبقه‌بندی را به صورت مقطعی کاهش می‌دهد، اما پس از چند تکرار نتیجه را بهبود داده و حتی دقتی بالاتر از دقت اولیه را نتیجه می‌دهد، این نتیجه با نتایج تحقیقات^۱ دیگر (آورده شده در منبع (۵۱)) نیز هم‌خوانی دارد [۵۱]. با توجه به زمان بسیاری کمی که یادگیری فعال با MLC نیاز دارد، یادگیری فعال را می‌توان بدون صرف وقتی زیاد برای این روش به منظور رسیدن به دقت مطلوب مورد استفاده قرار داد.

۵-۳-نتیجه‌گیری در رابطه با یادگیری فعال با روش ANN

اکثر تحقیقات انجام شده با این روش مربوط به مقالات جدید است، در تنها تحقیقی که با این روش به منظور یافتن اشیاء^۲ در تصاویر سنجش از دور انجام شده است نتایج مطلوبی گزارش

^۱ An active learning approach to hyperspectral data classification

^۲ Object detection

نشده است زیرا شناسایی اشیا با شبکه های عصبی مستلزم تعداد داده های آموزشی زیاد است [۵۲]، اما در تحقیقی دیگر که از یادگیری فعال بر روی این روش به منظور شناسایی تغییرات^۱ استفاده شده است نتایج خوب هستند [۵۳]، در تحقیقی که ما انجام دادیم نیز این روش بالاترین دقت را نتیجه داد اما در شناسایی کلاس تداخلی به علت ذات کلاس تقریباً از همه روش های دیگر ضعیفتر عمل کرد، زیرا معمولاً هر چقدر دقت طبقه بندی کلاس های دیگر افزایش پیدا می کند، دقت این کلاس کاهش میابد.

یادگیری فعال با صرف وقتی بسیار زیاد نسبت به سایر روش های pixel-base، در این روش به دقت بالاتری رسیده است، بنابراین می توان نتیجه گرفت که اگر زمان آموزش برای کاربر اهمیت زیادی نداشته باشد این روش، روش مناسبی برای بدست آوردن نتایج مطلوب است.

۵-۴- نتیجه گیری در رابطه با یادگیری فعال با روش DBN

تنها یک تحقیق در با این روش در زمینه سنجش از دور انجام شده است که نتایج بدست آمده از آن تحقیق نیز باعث بالاتر رفتن دقت شده است، در تحقیقی که ما انجام دادیم نیز دقت بالاتری نسبت به سایر حالات غیر از یادگیری فعال بدست آمد، با توجه به بالا بودن دقت و زمان نسبتاً کم استفاده از یادگیری فعال در این روش نیز به صرفه خواهد بود [۳۶].

۵-۵- نتیجه گیری در رابطه با یادگیری فعال با روش Segnet

تنها تحقیقی که بوسیله یادگیری فعال با CNN انجام شده است در زمینه یافتن اشیا است که نتایج بسیار خوبی را نسبت به Random sampling نتیجه داده است [۵۴]، در تحقیق ما بحث طبقه بندی تصویر مد نظر بوده که یادگیری فعال با روش Segnet هم از نظر دقت و هم از نظر زمان صرف شده نسبت به حالتی که از همه داده های pool + test برای آموزش استفاده شده عملکرد بهتری داشته است، برای اولین بار در این پایان نامه از یادگیری فعال با شبکه Segnet

^۱ Change Detection

استفاده شده است که نتایج مثبتی نیز گزارش کرده‌است، این روش نسبت به سایر روش‌ها در شناسایی کلاس‌های مختلف بسیار توانمند است، با توجه به اینکه کلاس تداخلی کلاسی است که در ظاهر به کلاس‌های دیگر شباهت دارد اما کاربردی متفاوت دارد و یادگیری فعال با Segnet توانست نسبت به سایر روش‌ها به خوبی این کلاس را تشخیص دهد، نتیجه می‌شود که یادگیری فعال با Segnet در هر دو حالت طبقه‌بندی Land-use و Land-cover نتایج خوبی را ارائه کند، یادگیری فعال برای Segnet به دلیل ارائه دقت بسار بالاتر از حالت‌های دیگر بسیار کارآمد است، به گونه‌ای که می‌توان به نتایج بسیار خوب با دقت‌های بالاتر نیز با ترکیب Segnet و یادگیری فعال دست پیدا کرد.

۵-۶-پیشنهادهات

- از آن جایی که یادگیری فعال برای همه روش‌های طبقه بندی مؤثر بوده است، بنابراین استفاده از آن در سایر تحقیقات و پروژه‌هایی که نیازمند طبقه‌بندی تصویر هستند، توصیه می‌شود.
- با توجه به اینکه CNNها برای عملکرد بهتر به داده‌های زیادی نیاز دارند و همیشه داده کافی در دسترس نیست، توصیه می‌شود که از یادگیری فعال برای این دسته از شبکه‌های عصبی استفاده شود.
- در این تحقیق تنها از یک CNN (SegNet) برای ارزیابی یادگیری فعال استفاده شده- است، پیشنهاد می‌شود که برای ارزیابی بهتر از CNN های دیگر نیز استفاده شود.
- عدم قطعیت فقط یکی از روش‌های Query یادگیری فعال است، با توجه به اینکه ما تعداد زیادی Query برای انجام یادگیری فعال داریم، می‌توان هر کدام از این Queryها را برای سایر روش‌های طبقه بندی تعمیم داد.
- با توجه به قابلیت شبکه Segnet در شناسایی کلاس تداخلی و بهبود آن با یادگیری فعال می‌توان از این روش برای برچسب‌گذاری Land cover استفاده کرد.
- در این تحقیق به علت کم بودن تعداد کلاس‌ها، شباهت کلاس‌ها به یکدیگر و وجود

کلاس تداخلی، داده‌های نویزی زیادی وجود داشت. گاهی نمودار یادگیری فعال-تکرار به جای روند صعودی، به صورت موقت دارای یک روند نزولی می‌شد که این به علت وجود نویز در داده‌های اضافه شده در آن تکرار یادگیری فعال بود، به همین جهت هنگام استفاده از یادگیری فعال پیشنهاد می‌شود که حداقل مکان از کلاس‌های زیادی استفاده شود و داده‌ها نیز عاری از نویز باشند تا یادگیری فعال بهتر عمل کند.

منابع و مأخذ

- [1] Donalek, C. *Supervised and unsupervised learning*. in *Astronomy Colloquia. USA*. 2011.
- [2] Deng, Z., et al., *Multi-scale object detection in remote sensing imagery with convolutional neural networks*. *ISPRS journal of photogrammetry and remote sensing*, 201 : ١٤٥ . ٨p. 3-22.
- [3] Ciregan, D., U. Meier, and J. Schmidhuber. *Multi-column deep neural networks for image classification*. in *2012 IEEE conference on computer vision and pattern recognition*. 2012. IEEE.
- [4] Luo, Y., et al. *Hsi-cnn: A novel convolution neural network for hyperspectral image*. in *2018 International Conference on Audio, Language and Image Processing (ICALIP)*. 2018. IEEE.
- [5] Albawi, S., T.A. Mohammed, and S. Al-Zawi. *Understanding of a convolutional neural network*. in *2017 International Conference on Engineering and Technology (ICET)*. 2017. IEEE.
- [6] Li, M., et al., *A review of remote sensing image classification techniques: The role of spatio-contextual information*. *European Journal of Remote Sensing*, 2014. **47**(1): p. 389-411.
- [7] Lu, D. and Q. Weng, *A survey of image classification methods and techniques for improving classification performance*. *International journal of Remote sensing*, 2007. **28**(5): p. 823-870.
- [8] Settle, J. and S. Briggs, *Fast maximum likelihood classification of remotely-sensed imagery*. *International Journal of Remote Sensing*, 1987. **8**(5): p. 723-734.
- [9] Zhu, H. and O. Basir, *An adaptive fuzzy evidential nearest neighbor formulation for classifying remote sensing images*. *IEEE Transactions on Geoscience and Remote Sensing*, 2005. : (٨) ٤٣ p. 1874-1889.
- [10] Mannan, B. and A. Ray, *Crisp and fuzzy competitive learning networks for supervised classification of multispectral IRS scenes*. *International Journal of Remote Sensing*, 2003. **24**(17): p. 3491-3502.
- [11] PAL, N.R. and D. BHANDARI, *On object background classification*. *International journal of systems science*, 1992. **23**(11): p. 1903-1920.
- [12] Al-Saffar, A.A.M., H. Tao, and M.A. Talab. *Review of deep convolution neural network in image classification*. in *2017 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET)*. 2017. IEEE.
- [13] Hubel, D.H. and T.N. Wiesel, *Receptive fields of single neurones in the cat's striate cortex*. *The Journal of physiology*, 1959. **148**(3): p. 574.
- [14] Hubel, D.H. and T.N. Wiesel, *Receptive fields, binocular interaction and functional architecture in the cat's visual cortex*. *The Journal of physiology*, 1962. **160**(1): p. 106.
- [15] LeCun, Y., et al., *Gradient-based learning applied to document recognition*.

- Proceedings of the IEEE, 1998. **86**(11): p. 2278-2324.
- [16] Krizhevsky, A., I. Sutskever, and G.E. Hinton, *Imagenet classification with deep convolutional neural networks*. Communications of the ACM, 2017. **60**(6): p. 84-90.
 - [17] Ma, L., et al., *Deep learning in remote sensing applications: A meta-analysis and review*. ISPRS journal of photogrammetry and remote sensing, 2019. **152**: p. 166-177.
 - [18] Lv, Q., et al., *Urban land use and land cover classification using remotely sensed SAR data through deep belief networks*. Journal of Sensors, 2015. **2015**.
 - [19] Sharma, A., et al., *A patch-based convolutional neural network for remote sensing image classification*. Neural Networks, 2017. **95**: p. 19-28.
 - [20] Alshehhi, R., et al., *Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks*. ISPRS Journal of Photogrammetry and Remote Sensing, 2017. **130**: p. 139-149.
 - [21] Han, M., et al., *Joint spatial-spectral hyperspectral image classification based on convolutional neural network*. Pattern Recognition Letters, 2020. **130**: p. 38-45.
 - [22] Wang, Q., et al., *Change detection based on Faster R-CNN for high-resolution remote sensing images*. Remote sensing letters, 2018. **9**(10): p. 923-932.
 - [23] Papadomanolaki, M., et al. *Stacked Encoder-Decoders for Accurate Semantic Segmentation of Very High Resolution Satellite Datasets*. in *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*. 2018. IEEE.
 - [24] Xing, H., et al., *Exploring geo-tagged photos for land cover validation with deep learning*. ISPRS journal of photogrammetry and remote sensing, 2018. **141**: p. 237-251.
 - [25] Long, J., E. Shelhamer, and T. Darrell. *Fully convolutional networks for semantic segmentation*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
 - [26] Sherrah, J., *Fully convolutional networks for dense semantic labelling of high-resolution aerial imagery*. arXiv preprint arXiv:1606.02585, 2016.
 - [27] Kemker, R., C. Salvaggio, and C. Kanan, *Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning*. ISPRS journal of photogrammetry and remote sensing, 2018. **145**: p. 60-77.
 - [28] Settles, B., *Active learning literature survey*. 2009, University of Wisconsin-Madison Department of Computer Sciences.
 - [29] Bouneffouf, D., et al. *Contextual bandit for active learning: Active thompson sampling*. in *International Conference on Neural Information Processing*. 2014. Springer.
 - [30] Bouneffouf, D., *Exponentiated gradient exploration for active learning*. Computers, 2016. **5**(1): p. 1.
 - [31] Mitra, P., B.U. Shankar, and S.K. Pal, *Segmentation of multispectral remote*

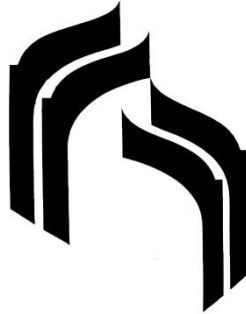
- sensing images using active support vector machines*. Pattern recognition letters, 2004. **25**(9): p. 1067-1074.
- [32] Tuia, D., et al., *Active learning methods for remote sensing image classification*. IEEE Transactions on Geoscience and Remote Sensing, 2009. **47**(7): p. 2218-2232.
- [33] Demir, B., C. Persello, and L. Bruzzone, *Batch-mode active-learning methods for the interactive classification of remote sensing images*. IEEE Transactions on Geoscience and Remote Sensing, 2010. **49**(3): p. 1014-1031.
- [34] Camps-Valls, G., et al., *Advances in hyperspectral image classification: Earth monitoring with statistical learning methods*. IEEE signal processing magazine, 2013. **31**(1): p. 45-54.
- [35] Zou, Q., et al., *Deep learning based feature selection for remote sensing scene classification*. IEEE Geoscience and Remote Sensing Letters, 2015. **12**(11): p. 2321-2325.
- [36] Liu, P., H. Zhang, and K.B. Eom, *Active deep learning for classification of hyperspectral images*. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2016. **10**(2): p. 712-724.
- [37] Deng, X., et al., *An improved method to construct basic probability assignment based on the confusion matrix for classification problem*. Information Sciences, 2016. **340**: p. 250-261.
- [38] Ma, Z. and R.L. Redmond, *Tau coefficients for accuracy assessment of classification of remote sensing data*. Photogrammetric Engineering and Remote Sensing, 1995. **61**(4): p. 435-439.
- [39] Kuhlman, D., *A Python Book: Beginning Python*. Advanced Python, and Python Exercises, 2012.
- [40] Danko, T. and P. Horvath, *modAL: A modular active learning framework for Python*. arXiv preprint arXiv:1805.00979, 2018.
- [41] Strahler, A.H., *The use of prior probabilities in maximum likelihood classification of remotely sensed data*. Remote sensing of Environment, 1980. **10**(2): p. 135-163.
- [42] Hagner, O. and H. Reese, *A method for calibrated maximum likelihood classification of forest types*. Remote sensing of environment, 2007. **107**(3): p. 438-444.
- [43] Sun, J., et al., *Automatic remotely sensed image classification in a grid environment based on the maximum likelihood method*. Mathematical and Computer Modelling, 2013. **58**(3-4): p. 573-581.
- [44] Biswajeet, P. and L. Saro, *Utilization of optical remote sensing data and GIS tools for regional landslide hazard analysis using an artificial neural network model*. Earth Science Frontiers, 2007. **14**(6): p. 143-151.
- [45] Abraham, A., *Artificial neural networks*. Handbook of measuring system design, 2005.
- [46] Seliya, N. and T.M. Khoshgoftaar. *Active learning with neural networks for*

- intrusion detection*. in *2010 IEEE International Conference on Information Reuse & Integration*. 2010. IEEE.
- [47] Settles, B. and M. Craven. *An analysis of active learning strategies for sequence labeling tasks*. in *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*. 2008.
 - [48] Badrinarayanan, V., A. Kendall, and R. Cipolla, *Segnet: A deep convolutional encoder-decoder architecture for image segmentation*. IEEE transactions on pattern analysis and machine intelligence, 2017. **39**(12): p. 2481-2495.
 - [49] Audebert, N., B. Le Saux, and S. Lefèvre, *Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks*. ISPRS Journal of Photogrammetry and Remote Sensing, 2018. **140**: p. 20-32.
 - [50] Stevens, E., L. Antiga, and T. Viehmann, *Deep Learning with PyTorch*. 2020, Manning.
 - [51] Rajan, S., J. Ghosh, and M.M. Crawford, *An active learning approach to hyperspectral data classification*. IEEE Transactions on Geoscience and Remote Sensing, 2008. **46**(4): p. 1231-1242.
 - [52] Qu, Z., et al., *Deep Active Learning for Remote Sensing Object Detection*. arXiv preprint arXiv:2003.08793, 2020.
 - [53] Ruzicka, V., et al. *Deep active learning in remote sensing for data efficient change detection*. in *Proceedings of MACLEAN: MACHine Learning for EArth ObservatioN Workshop co-located with the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD 2020)*. 2020. RWTH Aachen University.
 - [54] Xu, G., X. Zhu, and N. Tapper, *Using convolutional neural networks incorporating hierarchical active learning for target-searching in large-scale remote sensing images*. International Journal of Remote Sensing, 2019. **40**: p. 4057-4079.

Abstract:

Today, computer science and its increasing advances have a great impact on other related sciences, including remote sensing. Image classification using deep learning methods, which is one of the most important and noteworthy parts of remote sensing, is also advancing rapidly with the advancement of computer science. One of the problems of image classification with deep learning methods is the lack of educational data, which is one of the proposed solutions to solve the problem of data lack is the use of active learning. In this research, we have tried to solve the problem of data shortage by using active learning with other deep learning methods to classify satellite images. In fact, by developing active learning for the Segnet deep learning method, which has not been done before, we will address the effect of active learning on this network and the accuracy of the results. The data used in this study were obtained from the World view3 satellite. Ground truth These images are also manually prepared in four classes. After implementing active learning for three methods, MLC, ANN and DBN provided better results. Also, the proposed active learning method presented in this research, which was developed for Segnet, worked best and the active learning results were 91% accurate. Random sampling accuracy is 4% higher.

Keywords: Active learning, Deep learning, Remote sensing, Image classification, Artificial neural network



**Ministry of Science
Research and Technology**

**Graduate University of Advanced Technology
Faculty of Civil and Surveying Engineering
Department of Surveying Engineering**

**Development of active deep learning methods for remotely sensed data
classification**

A Thesis for Degree of Master of Science (M.Sc.) in Remote Sensing

**By:
Sajjad roshan meymandi**

**Supervisors
Saeid Niazmardi
Ali Esmacili**

February 2021