

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِيْمِ



دانشکده

مهندسی برق و کامپیوتر

بسم الله الرحمن الرحيم

تاییدیه اعضا هیات داوران

آقای علیرضا دری به شماره دانشجویی ۴۰۰۶۱۲۶۱۰۰۵ رشته مهندسی برق، مخابرات - مخابرات سیستم پایان نامه خود را با عنوان: طراحی شبکه CNN مبتنی بر فیلتر گابور با الهام از سامانه بینایی انسان برای طبقه‌بندی تصاویر سنجش از دور با نمونه‌های آموزشی محدود

Title: Design Gabor based CNN inspired by human vision system for classification of remote sensing images with limited training samples

در تاریخ ۱۴۰۲/۰۵/۳۰ ارائه گردند.

اعضا هیات داوران نسخه نهایی پایان نامه را از نظر فرم و محتوا بررسی نموده و پذیرش آنرا برای تکمیل درجه کارشناسی ارشد تایید می نمایند.

اعضاء	رتبه علمی	نام و نام خانوادگی	اعضا هیات داوران
	استاد	دکتر محمدحسن قاسمیان یزدی	استاد راهنمای
	دانشیار	دکتر مریم ایمانی آرانی	استاد مشاور
	دانشیار	دکتر نادر مکاری یامچی	استاد داور داخلی
	دانشیار	دکتر حمید دققانی	استاد داور خارجی
	دانشیار	دکتر نادر مکاری یامچی	نماینده شورای تحصیلات تکمیلی

دستورالعمل حق مالکیت مادی و معنوی در مورد نتایج پژوهش‌های علمی دانشگاه تربیت مدرس

مقدمه: با عنایت به سیاست‌های پژوهشی دانشگاه در راستای تحقق عدالت و کرامت انسانها که لازمه شکوفایی علمی و فنی است و رعایت حقوق مادی و معنوی دانشگاه و پژوهشگران، لازم است اعضای هیات علمی، دانشجویان، دانش آموختگان و دیگر همکاران طرح، در مورد نتایج پژوهش‌های علمی که تحت عنوانین پایان‌نامه، رساله و طرحهای تحقیقاتی که با هماهنگی دانشگاه انجام شده است، موارد ذیل را رعایت نمایند:

ماده ۱ - حقوق مادی و معنوی پایان نامه‌ها / رساله‌های مصوب دانشگاه متعلق به دانشگاه است و هرگونه بهره‌برداری از آن باید با ذکر نام دانشگاه و رعایت آئین نامه‌ها و دستورالعمل‌های مصوب دانشگاه باشد.

ماده ۲ - انتشار مقاله یا مقالات مستخرج از پایان نامه / رساله به صورت چاپ در نشریات علمی و یا ارائه در مجتمع علمی باید به نام دانشگاه بوده و استاد راهنما مسئول مکاتبات مقاله باشد. تبصره: در مقالاتی که پس از دانش آموختگی بصورت ترکیبی از اطلاعات جدید و نتایج حاصل از پایان نامه / رساله نیز منتشر می‌شود نیز باید نام دانشگاه درج شود.

ماده ۳ - انتشار کتاب حاصل از نتایج پایان نامه / رساله و تمامی طرحهای تحقیقاتی دانشگاه باید با مجوز کتبی صادره از طریق حوزه پژوهشی دانشگاه و بر اساس آئین نامه‌های مصوب انجام می‌شود.

ماده ۴ - ثبت اختراع و تدوین دانش فنی و یا ارائه در جشنواره‌های ملی، منطقه‌ای و بین‌المللی که حاصل نتایج مستخرج از پایان نامه / رساله و تمامی طرحهای تحقیقاتی دانشگاه باید با هماهنگی استاد راهنما یا مجری طرح از طریق حوزه پژوهشی دانشگاه انجام گیرد.

ماده ۵ - این دستورالعمل در ۵ ماده و یک تبصره در تاریخ ۱۳۸۷/۴/۱ در شورای پژوهشی دانشگاه به تصویب رسیده و از تاریخ تصویب لازم الاجرا است و هرگونه تخلف از مفاد این دستورالعمل، از طریق مراجع قانونی قابل پیگیری می‌شود.

«اینجانب علیرضا دری دانشجوی رشته مهندسی برق - مخابرات ورودی سال تحصیلی ۱۴۰۰ مقطع کارشناسی ارشد دانشکده مهندسی برق و کامپیوتر متعدد می‌شوم کلیه نکات مندرج در آئین نامه حق مالکیت مادی و معنوی در مورد نتایج پژوهش‌های علمی دانشگاه تربیت مدرس را در انتشار یافته‌های علمی مستخرج از پایان نامه / رساله تحصیلی خود رعایت نمایم. در صورت تخلف از مفاد آئین نامه فوق به دانشگاه و کالت و نمایندگی می‌دهم که از طرف اینجانب نسبت به لغو امتیاز اختراع بنام بنده و یا هرگونه امتیاز دیگر و تغییر آن به نام دانشگاه اقدام نماید. ضمناً نسبت به جبران فوری ضرر و زیان حاصله براساس برآورد دانشگاه اقدام خواهم نمود و بدینوسیله حق هرگونه اعتراض را از خود سلب نمودم.»

علیرضا دری امضا



آیین نامه چاپ پایان نامه (رساله) های دانشجویان دانشگاه تربیت مدرس

نظر به اینکه چاپ و انتشار پایان نامه (رساله) های تحصیلی دانشجویان دانشگاه تربیت مدرس، مبین بخشی از فعالیتهای علمی - پژوهشی دانشگاه است بنابراین به منظور آگاهی و رعایت حقوق دانشگاه، دانش آموختگان این دانشگاه نسبت به رعایت موارد ذیل متعهد می شوند:

ماده ۱: در صورت اقدام به چاپ پایان نامه (رساله) خود، مراتب را قبلاً به طور کتبی به «دفتر نشر آثار علمی» دانشگاه اطلاع دهد.

ماده ۲: در صفحه سوم کتاب (پس از برگ شناسنامه) عبارت ذیل را چاپ کند: «کتاب حاضر، حاصل پایان نامه کارشناسی ارشد/رساله دکتری نگارنده در رشته مهندسی برق است که در سال ۱۴۰۲ در دانشکده مهندسی برق و کامپیوتر دانشگاه تربیت مدرس به راهنمایی جناب آقای دکتر محمد حسن قاسمیان یزدی، مشاوره سرکار خانم دکتر مریم ایمانی از آن دفاع شده است.»

ماده ۳: به منظور جبران بخشی از هزینه های انتشارات دانشگاه، تعداد یک درصد شمارگان کتاب (در هر نوبت چاپ) را به «دفتر نشر آثار علمی» دانشگاه می تواند مازاد نیاز خود را به نفع مرکز نشر درعرض فروش قرار دهد.

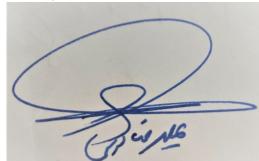
ماده ۴: در صورت عدم رعایت ماده ۳، ۵۰٪ بهای شمارگان چاپ شده را به عنوان خسارت به دانشگاه تربیت مدرس، تأديه کند.

ماده ۵: دانشجو تعهد و قبول می کند در صورت خودداری از پرداخت بهای خسارت، دانشگاه می تواند خسارت مذکور را از طریق مراجع قضایی مطالبه و وصول کند؛ به علاوه به دانشگاه حق می دهد به منظور استیفای حقوق خود، از طریق دادگاه، معادل وجه مذکور در ماده ۴ را از محل توقيف کتابهای عرضه شده نگارنده برای فروش، تامین نماید.

ماده ۶: اینجانب علیرضا دری دانشجوی رشته مهندسی برق، مخابرات سیستم مقطع کارشناسی ارشد تعهد فوق وضمانت اجرایی آن را قبول کرده، به آن ملتزم می شوم.

علیرضا دری

امضا



۱۴۰۲
مرداد



پایان نامه کارشناسی ارشد
گروه مخابرات (گرایش سیستم)

طراحی شبکه CNN مبتنی بر فیلتربانک گابور با الهام از سامانه بینایی
انسان برای طبقه‌بندی تصاویر سنجش از دور با نمونه‌های آموزشی
محدود

نگارش
علیرضا دری
استاد راهنما
دکتر محمد حسن قاسمیان یزدی
استاد مشاور
دکتر مریم ایمانی آرانی

سکنر و قدردانی

از زحمات و تلاش های بی دین استاد محترم، جناب آقا^ی دکتر قاسمیان نژدی برای بهمراه ایده ها، میش و مشاوره، برای به سر انجام رسیدن این پایان نامه و همچنین درس ها و نهاده ارزشمند برای ادامه مسیر زندگی صمیمانه سپاس گذاری و قدردانی می نایم.

همچنین از سرکار خانم دکتر رایانی برای تمام تلاش ها و راهنمایی ارزشمند شان، سپاس گذاری و قدردانی می نایم.
از خانواده عزیزم که همواره پشتیبان بند بوده اند سپاسگزاری می نایم و همچنین از سایر بھکاران و دوستان به خصوص عزیزان حاضر در آزمایشگاه پردازش تصویر و آنالیز اطلاعات که به هر نحوی در تهیه این مجموعه با این جانب بھکاری داشته اند سکنر نموده و موفقیت بهم آن ها را آرزو مندم.

علی‌رضادی

۱۴۰۲

لعدیم به تمام کسانی که با تقدیم نفس هایشان، جانی دوباره به ایران بخواهند.

علیرضا دری
مرداد ۱۴۰۲

چکیده

طبقه‌بندی تصاویر سنجش از دور مسئله‌ای مهم و کاربردی است که اهمیت آن در سال اخیر روندی افزایشی داشته است. سنجنده‌های جدید سنجش از دور تصاویر سطح زمین را در باندهای مختلف طیفی الکترومغناطیسی با دقت مکانی بالا اخذ می‌کنند. این تصاویر حاوی اطلاعات طیفی و مکانی از صحنه مورد نظر می‌باشند. اطلاعات مکانی شامل شکل، موقعیت مکانی، نوع بافت و ساختار هندسی و چینش اشیاء در صحنه می‌باشد و اطلاعات طیفی کیفیت و نوع مواد تشکیل‌دهنده اشیاء موجود در صحنه را بیان می‌کند. ماهواره‌های اخیر سنجش از دور با ایجاد تصاویر چندطیفی و با دقت مکانی بالا، لزوم استفاده توأم از اطلاعات طیفی و مکانی در طبقه‌بندی تصاویر سنجش از دور را ایجاب می‌کند. برای مثال ماهواره WorldView3 دارای ۸ باند طیفی و با دقت مکانی حدود ۱.۲ متر و تصویر PAN با دقت ۳۰ سانتیمتر است. مسئله اصلی این تحقیق نیاز به ارائه یک روش کم هزینه، که بتواند در سامانه‌های ارزان قیمت سنجش از دور با صحت و اعتبار بالا فرایند تشخیص کاربری‌های سطح زمین را انجام دهد، می‌باشد. مسئله دوم: نداشتن نقشه مرجع و محدودیت نمونه‌های آموزشی با برچسب صحیح و قابل قبول و بالا بودن هزینه تهیه این گونه نمونه‌ها می‌باشد، که فرایند طبقه‌بندی را با محدودیت نمونه‌های آموزشی مواجه می‌کند. هدف این تحقیق طراحی تلفیقی از مجموعه فیلتری گابور و طبقه‌بندی با الهام از سامانه بینایی انسان برای استخراج ویژگی و طبقه‌بندی پوشش‌های زمینی با قید محدودیت نمونه‌های آموزشی می‌باشد. شبکه‌های CNN به دلیل شباهت زیادشان به سیستم بینایی انسان شهرت زیادی یافته‌اند، این شبکه‌ها با وجود تئوری قوی به عنوان پشتیبان، اخیراً مورد توجه بیشتری قرار گرفته است که از جمله دلایل آن پیشرفت سخت افزارهای کامپیوترا مربوط به انجام پردازش‌های شبکه CNN است. این شبکه در سال‌های اخیر نشان داده‌اند که قدرت زیادی در طبقه‌بندی تصاویر سنجش از دور دارند اما متاسفانه یکی از مشکلات این شبکه‌ها نیاز به دیتای آموزشی زیاد است، که از دلایل آن شروع شدن روند آموزش از نقاط تصادفی برای وزن‌های داخل شبکه است. یک روش مرسوم برای حل این مشکل، استفاده از روش‌های استخراج ویژگی قبل از شبکه و سپس ورود ویژگی‌ها به شبکه است. در این پایان نامه برای حل مشکل نیاز به داده آموزشی زیاد، سعی شده تعیین وزن‌ها تا جای ممکن از نقاط تصادفی شروع نشود و برای این کار از مدل بینایی انسان الهام گرفته شده است تا به این طریق علاوه بر افزایش سرعت آموزش و حفظ دقت و اعتبار طبقه‌بندی، تعداد نمونه‌های آموزشی مورد نیاز شبکه نیز کاهش یابد.

واژه‌های کلیدی:

سنجش از دور، تصاویر چندطیفی، شبکه کانولوشنی، ساختار بینایی، مدل چشم، فیلتربانک گابور، نقشه مرجع.

فهرست مطالب

عنوان

صفحة

۱ مقدمه	۲
۱-۱ سنجش از دور	۳
۲-۱ تصاویر چندطیفی و اهمیت طبقه‌بندی آن‌ها	۴
۳-۱ مشخصات و ویژگی‌های تصاویر سنجش از دور	۶
۴-۱ سنجنده‌های تصاویر چندطیفی	۶
۵-۱ بیان مسئله و چالش‌های طبقه‌بندی تصاویر سنجش از دور	۹
۲ اصول اولیه و ابزارهای مورد نیاز	۱۱
۱-۲ بررسی عملکرد و ساختمان چشم انسان	۱۲
۱-۱-۱ ساختمان کلی چشم و شبکیه انسان	۱۲
۱-۱-۲ ویژگی‌های تصاویر تولید شده در شبکیه و ساختمان جزئی‌تر چشم	۱۵
۲-۱-۲ شبکه‌های عصبی	۱۸
۱-۲-۲ نورون‌های مصنوعی	۱۸
۲-۲-۲ شبکه‌های پرسپترون چندلایه	۲۰
۳-۲-۲ الگوریتم انتشار به عقب	۲۰
۳-۲ شبکه‌های کانوالوشنی	۲۴
۱-۳-۲ تاریخچه شبکه‌های کانوالوشنی	۲۴
۲-۳-۲ شبکه کانوالوشنی نماینده‌ای از سیستم بینایی انسان	۲۶
۳-۳-۲ عملکرد مغز انسان و شبکه CNN	۲۷
۴-۳-۲ شبکه‌های کانوالوشنی	۲۹
۱-۴-۳-۲ لایه کانوالوشن دو بعدی	۳۱
۲-۴-۳-۲ لایه‌های ادغام	۳۳
۳-۴-۳-۲ نحوه به روزرسانی وزن‌ها در شبکه کانوالوشنی	۳۴
۴-۴-۳-۲ بررسی یک شبکه کانوالوشنی بزرگ	۳۵
۴-۲ فیلتربانک گابور	۳۶
۵-۲ معیارهای ارزیابی	۳۸
۳ عملکرد طبقه‌بندها و ویژگی‌های متفاوت بر روی دیتاست‌های مختلف	۴۳
۱-۳ معرفی روش‌های استخراج ویژگی‌های مکانی با فیلتربانک گابور	۴۴
۱-۱-۳ روش اول، استفاده از فیلتربانک گابور کامل	۴۴
۱-۲-۳ روش دوم، استفاده از فیلتربانک گابور برش خورده	۴۵
۲-۳ فیلتربانک کامل و برش خورده و مقایسه عملکرد آن‌ها در خوشبندی	۴۶
۳-۳ فیلتربانک کامل و برش خورده و مقایسه عملکرد آن‌ها در طبقه‌بندی	۴۸
۱-۳-۳ آزمایش طبقه‌بندی با دیتاست اول	۴۹

۵۱	آزمایش طبقه‌بندی با دیتاست دوم	۲-۳-۳
۵۶	معرفی داده مناسب با نقشه مرجعی مناسب برای اشیاء	۴-۳
۶۳	انجام طبقه‌بندی بر روی دیتاست جدید	۵-۳
۶۳	نتایج طبقه‌بندی با ML	۱-۵-۳
۶۴	نتایج طبقه‌بندی با 1D-CNN	۲-۵-۳
۶۶	نتایج طبقه‌بندی با Tree	۳-۵-۳
۶۷	نتایج طبقه‌بندی با 2D-CNN	۴-۵-۳
۶۹	طرح اولیه برای روش پیشنهادی	۶-۳
۷۰	۴ روش پیشنهادی	
۷۱	۱-۴ مقدمه	۱-۴
۷۱	۲-۴ ایده اولیه روش پیشنهادی	۲-۴
۷۶	۳-۴ مرحله دوم در مسیر دست یابی به روش پیشنهادی	۳-۴
۸۰	۴-۴ مرحله نهایی و روش پیشنهادی کلی	۴-۴
۸۶	۵-۴ نگرشی دیگر در طبقه‌بندی تصاویر چندطیفی	۵-۴
۹۳	۵ جمع‌بندی و پیشنهادات	
۹۴	۱-۵ جمع‌بندی	۱-۵
۹۵	۲-۵ پیشنهاد	۲-۵
۹۶	منابع و مراجع	
۱۰۰	واژه‌نامه‌ی انگلیسی به فارسی	۰۰

فهرست اشکال

شکل

صفحة

۱-۱	روند اخذ و ثبت تصاویر سنجش از دور	۴
۲-۱	روند رشد تعداد مقالات با موضوع طبقه‌بندی تصاویر چندطیفی سنجش از دور	۴
۳-۱	چند نمونه از کاربردهای سنجش از دور	۵
۴-۱	تفاوت تصاویر چندطیفی و ابرطیفی در دقت طیفی آنها	۷
۱-۲	ساختمان کلی چشم و لایه‌های شبکیه	۱۳
۲-۲	نمایی از شبکیه و سلول‌های موجود در هر لایه	۱۴
۳-۲	حفره مرکزی	۱۵
۴-۲	تفاوت حفره مرکزی و سایر شبکیه	۱۶
۵-۲	نحوه اتصال سلول‌های AII و استفاده از سلول‌های خاموش و روشن	۱۷
۶-۲	تصاویری برای درک بهتر شبکه عصبی	۱۸
۷-۲	یک نورون مصنوعی	۱۹
۸-۲	انواع مختلف توابع فعال‌ساز	۲۰
۹-۲	نحوه نام‌گذاری لایه‌ها در یک شبکه پرسپترون چندلایه با یک لایه مخفی	۲۰
۱۰-۲	نحوه تغییر وزن لایه میانی	۲۳
۱۱-۲	کشف سلول‌های ساده و پیچیده توسط هابل و ویزل	۲۵
۱۲-۲	شبکه نوادرانکی و مدل سازی سلول‌های S و C به صفحات S و C	۲۵
۱۳-۲	تشابه شبکه HMAX و لایه maxpooling به مدل سلول‌های S و C	۲۶
۱۴-۲	پیش‌بینی مکان احتمالی فعالیت مغزی با کمک از فعالیت شبکه CNN	۲۷
۱۵-۲	بخش‌های مختلف مرتبط با بینایی در مغز انسان	۲۸
۱۶-۲	نحوه عملکرد و وظیفه شبکه CNN	۲۹
۱۷-۲	بخش‌های مختلف عملکرد مغز در سیستم بینایی	۳۰
۱۸-۲	نمایش اتصالات محدود شده در شبکه کانوالوشنی	۳۰
۱۹-۲	نمایش اتصالات در لایه‌های بالاتر شبکه کانوالوشنی	۳۱
۲۰-۲	نحوه اعمال فیلتر بر تصویر در حوزه مکان	۳۱
۲۱-۲	نحوه عملکرد لایه کانوالوشنی	۳۲
۲۲-۲	رفتار کانوالوشن در دو بعد اول نسبت به پارامتر S	۳۲
۲۳-۲	رفتار کانوالوشن در دو بعد اول نسبت به پارامتر	۳۳
۲۴-۲	عملکرد لایه ادغام	۳۴
۲۵-۲	نحوه عملیات اصلاح وزن‌ها در شبکه CNN	۳۴
۲۶-۲	برخی از فیلترهای لایه اول شبکه AlexNet	۳۵
۲۷-۲	نمایش خروجی ۳ کanal از هر لایه که حساسیت بالایی به ورودی دایره شکل دارند.	۳۶

۲۸-۲	نمایش فیلتربانک گابور	۳۷
۲۹-۲	نمایش فیلتر دو نوع فیلتر گابور	۳۹
۳۰-۲	جایگاه پارامترهای متفاوت در ماتریس درهم ریختگی	۴۰
۳۱-۲	ROC curve نمودار	۴۱
۳۲-۲	نمایش دو حالت متفاوت از ROC	۴۲
۱-۳	فرایند بررسی و مقایسه عملکرد دو نوع فیلتربانک گابور	۴۴
۲-۳	نمونه برش یک فیلتر گابور برای انجام آزمایشات بر روی آنها	۴۵
۳-۳	اعمال فیلتر برش خورده	۴۶
۴-۳	نتایج خوشبندی	۴۷
۵-۳	اعمال فیلتر برش خورده	۴۸
۶-۳	داده اول برای آزمایش طبقه‌بندی	۴۹
۷-۳	نتایج طبقه‌بندی با طبقه‌بند ML برای داده اول	۵۱
۸-۳	اصلاح و پیش‌پردازش تصویر ابرطیفی دانشگاه پاویا	۵۲
۹-۳	روندنمای انجام آزمایش دوم	۵۳
۱۰-۳	نتایج طبقه‌بندی با طبقه‌بند ML برای داده اول	۵۴
۱۱-۳	مثالی از توانایی شبکه‌های عصبی	۵۶
۱۲-۳	نمای کلی شبکه عصبی استفاده شده در آزمایش دوم طبقه‌بندی	۵۶
۱۳-۳	نتایج خوشبندی فرضی	۵۸
۱۴-۳	بخش اول نحوه ساخت نقشه مرجع با الگوریتم مورد نظر	۵۹
۱۵-۳	بخش دوم نحوه ساخت نقشه مرجع با الگوریتم مورد نظر	۶۰
۱۶-۳	نقشه نهایی به دست آمده از الگوریتم مورد نظر به همراه تصاویر اصلی	۶۱
۱۷-۳	نمای کلی شبکه عصبی استفاده شده در تولید نقشه مرجع	۶۲
۱۸-۳	نتایج طبقه‌بند ML برای داده جدید	۶۴
۱۹-۳	اطلاعات کلی روش پیشنهادی پایان نامه یاد شده	۶۵
۲۰-۳	نتایج طبقه‌بند 1D-CNN برای داده جدید	۶۶
۲۱-۳	نتایج طبقه‌بند Tree برای داده جدید	۶۷
۲۲-۳	نتایج طبقه‌بند 2D-CNN برای داده جدید	۶۸
۱-۴	اطلاعات کلی روش پیشنهادی اولیه	۷۲
۲-۴	جزئیات شبکه مورد استفاده در روش پیشنهادی اولیه	۷۳
۳-۴	نتایج شبکه مورد استفاده در روش پیشنهادی اولیه	۷۵
۴-۴	بلوک گابور برای استفاده در مراحل بعدی شبکه پیشنهادی (بخش داخل مستطیل)	۷۶
۵-۴	جزئیات شبکه مورد استفاده در مرحله دوم طراحی شبکه پیشنهادی	۷۷
۶-۴	نتایج شبکه مورد استفاده در مرحله دوم طراحی شبکه روش پیشنهادی	۷۹

۷-۴	بلوک مکانی برای استفاده در مراحل بعدی شبکه پیشنهادی	۸۰
۸-۴	جزئیات شبکه پیشنهادی	۸۱
۹-۴	مزیت پنجره با طول فرد برای پیکسل مرکزی	۸۱
۱۰-۴	نتایج تصویری روش پیشنهادی برای داده همۀ داده ها	۸۴
۱۱-۴	تصویر یک خودروی متحرک در چند باند مختلف از تصویر چندطیفی	۸۵
۱۲-۴	بررسی پیکسل های هدف: در تصویر تک باند در مقابل هر پیکسل چندطیفی، ۱۶ پیکسل به عنوان هدف در نظر گرفته می شوند.	۸۷
۱۳-۴	نحوه حرکت پنجره بر روی تصویر. در اینجا ۳ مربع ۱۶ پیکسلی با همسایگی ۵ از هر طرف به عنوان داده متناظر با یک پیکسل از تصویر چندطیفی انتخاب شده‌اند که هدف برچسب زدن به ۱۶ پیکسل مرکزی است.	۸۸
۱۴-۴	نتایج تصویری روش جدید	۹۲

صفحة	فهرست جداول	جدول
		1-۱ جدول اطلاعات تصاویر ماهواره WV3 (earth.esa.int) ۸
		2-۱ جدول اطلاعات تصاویر ماهواره QB2 (earth.esa.int) ۸
		3-۱ اطلاعات فیلتربانک‌های مورد آزمایش برای مقایسه عملکرد آن‌ها در خوشبندی ۴۷
		2-۳ اطلاعات فیلتربانک‌های مورد آزمایش برای مقایسه آن‌ها در طبقه‌بندی دیتاست اول ۴۹
		3-۳ نتایج طبقه‌بند ML با فیلتربانک برش‌خورده ۵۰
		4-۳ نتایج طبقه‌بند ML با فیلتربانک کامل ۵۰
		5-۳ اطلاعات فیلتربانک‌های مورد آزمایش برای مقایسه آن‌ها در طبقه‌بندی دیتاست دوم ۵۲
		6-۳ نتایج طبقه‌بند CNN با فیلتربانک کامل ۵۳
		7-۳ نتایج طبقه‌بند CNN با فیلتربانک برش‌خورده ۵۳
		8-۳ نمونه‌ای از ایرادات طبقه‌بندی با فیلترهای کامل ۵۵
		9-۳ اطلاعات شبکه کانوالوشنی استفاده شده ۵۷
		10-۳ تعداد نقاط هر خوشة در هر کدام از نتایج ۵۷
		11-۳ مشخصات نقشه مرجع تولید شده ۶۱
		12-۳ اطلاعات شبکه MLP استفاده شده در تولید نقشه مرجع ۶۲
		13-۳ نتایج طبقه‌بند ML برای داده جدید ۶۴
		14-۳ نتایج طبقه‌بند 1D-CNN برای داده جدید ۶۶
		15-۳ نتایج طبقه‌بند Tree برای داده جدید ۶۷
		16-۳ نتایج طبقه‌بندی 2D-CNN برای داده جدید ۶۸
		1-۴ اطلاعات فیلتربانک انتخاب شده برای قرار گرفتن در شبکه CNN پیشنهادی ۷۲
		2-۴ اطلاعات کامل شبکه اولیه برای روش پیشنهادی ۷۴
		3-۴ اطلاعات شبکه مربوط به آموزش شبکه پیشنهادی اولیه ۷۴
		4-۴ نتایج شبکه مورد استفاده در روش پیشنهادی اولیه برای داده‌های آزمایشی ۷۴
		5-۴ اطلاعات کامل شبکه در مرحله دوم برای روش پیشنهادی ۷۷
		6-۴ نتایج شبکه مورد استفاده در مرحله دوم طراحی شبکه پیشنهادی ۷۸
		7-۴ اطلاعات کامل شبکه برای روش پیشنهادی ۸۲
		8-۴ نتایج طبقه‌بندی روش پیشنهادی (فقط برای داده‌های تست) ۸۳
		9-۴ مشخصات نقشه مرجع تولید پس از هم سایز کردن به تصویر چندطیفی ۸۸
		10-۴ اطلاعات کامل بلوک گابور برای طرح جدید ۸۹
		11-۴ اطلاعات کامل بلوک مکانی برای طرح جدید ۸۹
		12-۴ اطلاعات کامل شبکه برای روش جدید ۹۰
		13-۴ اطلاعات مربوط به آموزش شبکه جدید ۹۰

۹۱ ۱۴-۴ نتایج طبقه‌بندی با روش جدید ..

فهرست نمادها

نماد	
κ	Kappa coefficient
P_n	Noise Power
P_s	Signal Power
x_0	Gaussian peak at x-axis
y_0	Gaussian peak at y-axis
σ_x	x-axis deviation of Gaussian envelope for Gabor filter
σ_y	y-axis deviation of Gaussian envelope for Gabor filter
ΔS	Spatial Resolution
Δt	Temporal Resolution
$\Delta \lambda$	Spectral Resolution
E	Energy
P	Phase Offset & padding
S	Stride
x	Longitude
y	Latitude

فهرست اختصارات

1-D CNN	1-Dimensional Convolutional Neural Network
2-D CNN	2-Dimensional Convolutional Neural Network
3-D CNN	3-Dimensional Convolutional Neural Network
AA	Average Accuracy
AROC	Approximate Rank-Order Clustering
AV	Average Validity
CNN	Convolutional Neural Network
CNN-MRF	Convolutional Neural Network-Markov Random Fields
CNNE	Convolutional Neural Network Ensemble
DNN	Deep Neural Network
DP	Depthwise & Pointwise
GLCM	gray-level co-occurrence matrices
GSHCN	Gabor & Shapelet Convolutional Network
GTM	Ground Truth Map
LSTM	Long Short-Term Memory
M3D-CNN	Multi-Scale Densely 3D-CNN
ML	Maximum Likelihood
MLP	Multi-Layer Perceptron
OA	Overall Accuracy
P	Padding
PAN	panchromatic
PCA	Principal Component Analysis
QB2	Quick Bird 2
RF	Random Forest
RGB	Red Green Blue
RNN	Recurrent Neural Network
S	Stride
SGDM	stochastic gradient descent with momentum
SNR	Signal to Noise Ratio
SV	Single-view
WV3	World view 3

فصل اول

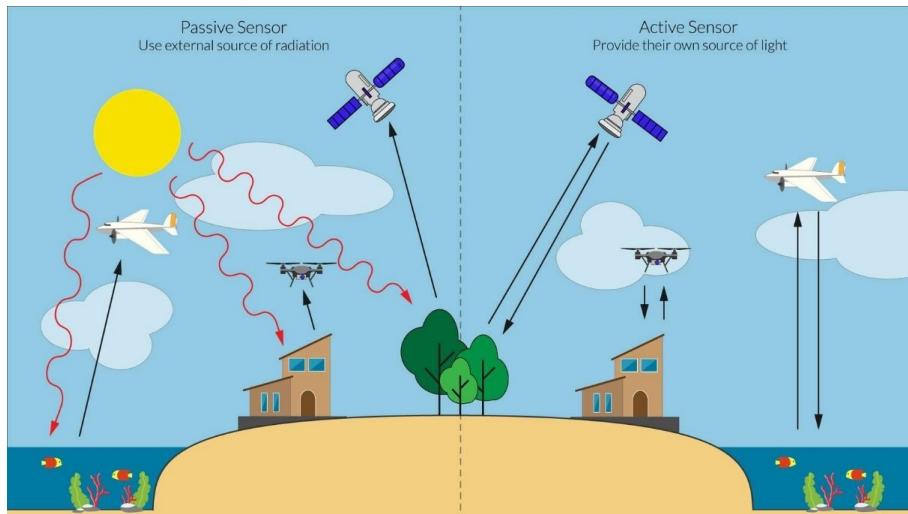
مقدمه

۱-۱ سنجش از دور

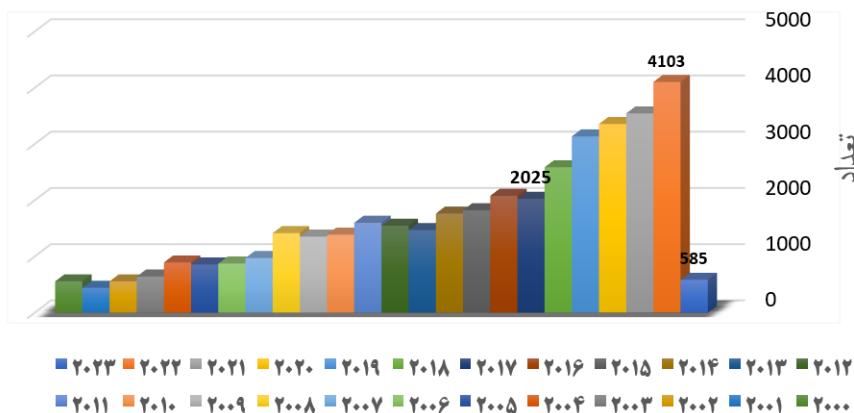
سنجش از دور، فرآیند تشخیص و نظارت به مشخصه‌های پوشش زمین (شیء یا عارضه) بدون تماس فیزیکی است که با اندازه‌گیری بازتاب نور ساطع شده از ناحیه تصویر برداری شده انجام می‌گیرد.^[۱] تصاویر طیفی که در سنجش از دور مورد استفاده قرار می‌گیرند شامل تصاویر تکباند، چندطیفی و ابرطیفی می‌شوند. بازتاب اشیاء مختلف مانند درخت، ساختمان، جاده و... یکتا است و به طور کلی هر شیء همچنان که یک شکل و بافت خاص دارد، یک امضا طیفی مربوط به خودش را دارد که از این امضاهای طیفی برای طبقه‌بندی سطح زمین استفاده می‌شود. تعداد باندهای تصاویر چندطیفی بین ۳ تا ۱۵ باند است و تصاویر ابرطیفی دارای چندصد باند مختلف هستند. طبقه‌بندی تصاویر ابرطیفی به جهت تعداد زیاد باندها دقت بالایی به همراه دارد در مقابل، در تصاویر چندطیفی دستیابی به دقت بالای طبقه‌بندی به جهت کم بودن تعداد باندها بدون استفاده از ویژگی‌های مکانی ممکن نیست. نویز و عوامل محیطی مانند ابر، سایه اشیاء، رطوبت و... فرایند طبقه‌بندی را مشکل می‌کنند از این رو هرچه تعداد باندها بیشتر باشد دقت طبقه‌بندی بیشتر خواهد بود.

أخذ تصاویر سنجش از دور از سطح زمین بسته به نوع کاربرد تصاویر، از طریق نصب سنجنده بر روی پهپاد، هواپیما یا ماهواره انجام می‌شود.^[۲] در گذشته این تصویر برداری با نصب سنجنده بر روی بالن صورت می‌گرفت است، در سال ۱۸۵۸ یکی از اولین تصاویر سنجش از دور از شهر لندن و از روی بالن ثبت شده است.^[۳] سنجنده‌ها به دو دسته کلی فعال و غیرفعال تقسیم می‌شوند. در عکس برداری با سنجنده‌های فعال، سنجنده با ارسال موج در باندهای مختلف و سپس دریافت و ثبت شدت موج بازتابی از اشیاء در باندهای ارسال شده، تصویر محل مورد نظر را تشکیل می‌دهند. در عکس برداری با سنجنده غیرفعال، سنجنده شدت بازتاب امواج در طیف‌های مختلف را دریافت و ثبت می‌کند عموماً منبع اصلی این امواج نور خورشید است و خود سنجنده در تولید این امواج نقشی ندارد.^[۴] روند تولید تصاویر سنجش از دور و تفاوت سنجنده‌های مختلف در شکل ۱-۱ نمایش داده شده است.

در نمودار شکل ۱-۲ نشان داده شده است که در سال‌های اخیر طبقه‌بندی تصاویر چندطیفی مورد توجه قرار گرفته است.



شکل ۱-۱: روند اخذ و ثبت تصاویر سنجش از دور.[۵]



شکل ۱-۲: روند رشد تعداد مقالات با موضوع طبقه‌بندی تصاویر چندطیفی سنجش از دور.(scopus.com)

۱-۲ تصاویر چندطیفی و اهمیت طبقه‌بندی آن‌ها

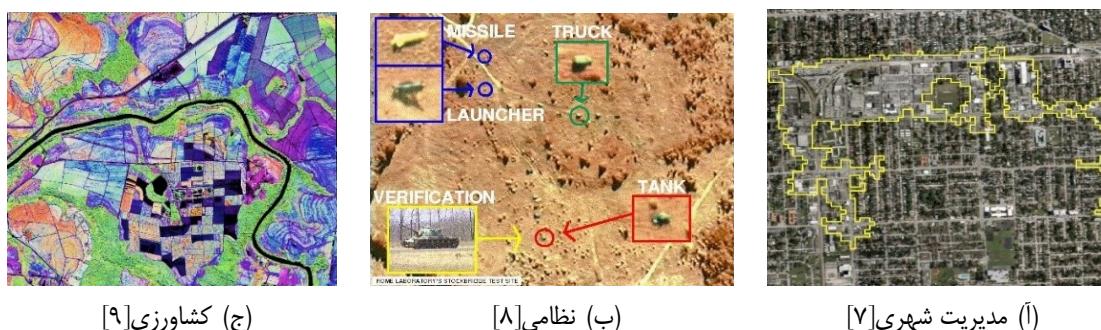
تصاویر چندطیفی متشکل از چند طول موج مشخص هستند که معمولاً طول موج نور مرئی و طول موج‌های بزرگتر را شامل می‌شوند. این تصاویر دقت مکانی بیشتری نسبت به تصاویر ابرطیفی در اختیار ما قرار می‌دهند ولی اطلاعات طیفی آن‌ها به اندازه تصاویر ابرطیفی نیست. در تصویر برداری چندطیفی شدت نور در تعداد محدودی باند (بین ۳ تا ۱۵ باند مختلف) اندازه‌گیری و ثبت می‌شود.[۶]

طبقه‌بندی تصاویر سنجش از دور در مدیریت شهری، کشاورزی و کاربردهای نظامی اهمیت زیادی دارد. برتری تصاویر چندطیفی نسبت به ابرطیفی دقت مکانی بیشتر با حفظ دقت طیفی کافی است. در تصاویر ابرطیفی به دلیل وجود اطلاعات طیفی بالا استفاده از اطلاعات مکانی اهمیت چندانی در نتایج نخواهد داشت.

اما استفاده از اطلاعات مکانی در تصاویر چندطیفی بسیار مفید و ارزشمند خواهد بود چرا که در این تصاویر با کاهش اطلاعات طیفی، اطلاعات مکانی افزایش یافته است.

کاربردهای متنوعی را برای تصاویر سنجش از دور و به خصوص تصاویر چندطیفی می‌توان برشمرد که در هر کدام از کاربردها، نوع اطلاعات مورد نیاز می‌تواند متفاوت باشد. برای مثال در شکل ۱-۳(آ) در زمینه مدیریت شهری، برای احداث و یا عدم احداث یک پل می‌توان با استفاده از اطلاعات طیفی تصویر، جنس و سختی محل احداث را تقریب زد که آیا از نظر سختی مناسب احداث پل هست یا خیر. یا با استفاده از ویژگی‌های مکانی تاثیر پل را بر بافت شهری ارزیابی کرد. با استفاده از اطلاعات طیفی تصاویر سنجش از دور از منطقه نیز می‌توان تاثیر ساخت پل بر آلودگی هوا را ارزیابی کرد. در مثال فوق نقش اطلاعات مختلف یک تصویر از یک محل خاص به خوبی نشان داده شده است.

در کاربردهای نظامی نقش این دو دسته اطلاعات در کنار یکدیگر بیشتر خود را نشان می‌دهد برای مثال در شکل ۱-۳(ب) با استفاده از اطلاعات طیفی جنس مواد مختلف موجود در صحنه تا حد زیادی قابل شناسایی است. فرض کنید اطلاعات طیفی وجود مقدار زیادی آهن را در یک نقطه نشان دهد، بدون اطلاعات مکانی نمی‌توانیم متوجه شویم که در آن نقطه یک تانک قرار دارد یا یک هدف فرضی، پس اهمیت این دو دسته اطلاعات در کنار یکدیگر بودن آنها است. در بحث کشاورزی در شکل ۱-۳(ج) نیز می‌توان با استفاده از همین اطلاعات گیاهان مناسب برای کشت در مناطق مختلف را شناسایی کنیم.



شکل ۱-۳: چند نمونه از کاربردهای سنجش از دور

۳-۱ مشخصات و ویژگی‌های تصاویر سنجش از دور

داده‌های ابرطیفی یا چندطیفی تابعی از انرژی دریافتی توسط سنجنده و بر حسب دقت‌های مکانی، طیفی، پرتوسنجی و زمانی است که در رابطه‌های ۱-۱ تا ۴-۱ مشخص شده‌اند.

$$f(x, y, \lambda, t) = \int_{\Delta t} \int_{\Delta \lambda} \int_{\Delta x} \int_{\Delta y} E_r(x, y, \lambda, t) dx dy d\lambda dt \quad (1-1)$$

$$\text{Radiometric Resolution} \geq \log_2^{SNR} \quad (2-1)$$

$$\Delta E = \Delta S \Delta \lambda \Delta t \quad , \quad \Delta S = \Delta x \Delta y \quad (3-1)$$

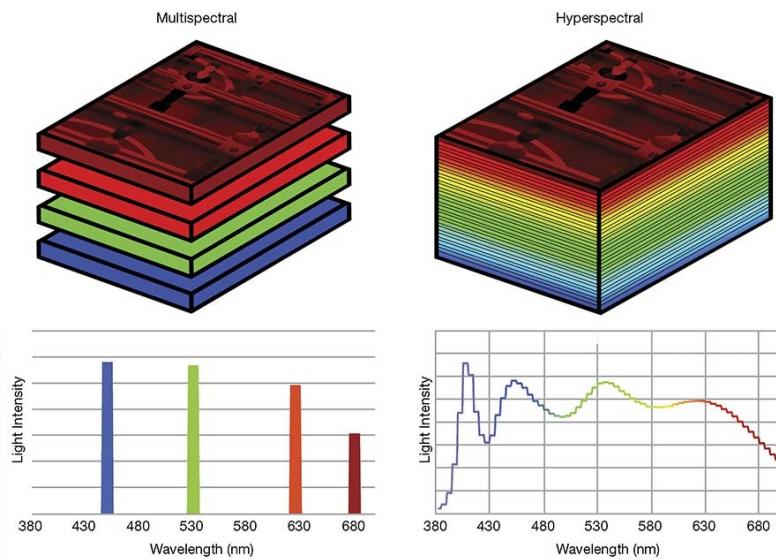
$$SNR = \frac{P_s}{P_n} = \frac{\Delta S \Delta \lambda}{\sigma_2^n} \Rightarrow \Delta \lambda \propto \frac{1}{\Delta S} \quad (4-1)$$

دقت مکانی ΔS بیانگر ابعاد و فاصله فیزیکی است که هر پیکسل از تصویر دربر دارد. دقت طیفی $\Delta \lambda$ فاصله طول موج‌های تصویر برداری شده را بیان می‌کند که در تصویر برداری ابرطیفی این فواصل برابر هستند. هر چه این فاصله کوچک‌تر باشد، تعداد باندهای بیشتری تصویر برداری می‌شود ولی باعث کاهش انرژی ورودی به سنجنده می‌شود از این رو به ناجار باید ΔS یا Δt افزایش یابد تا SNR بالا باقی بماند. دقت پرتوسنجی تعداد سطوح روشنایی برای پیکسل‌های تصویر را نشان می‌دهد که معمولاً برای تصاویر چندطیفی ۸ تا ۱۶ بیتی است. دقت زمانی Δt طول مدت زمان تصویر برداری یا به عبارتی مدت زمان باز بودن ورودی سنجنده برای SNR دریافت انرژی تابشی از سطح زمین است. طبق رابطه ۴-۱ ضرب دقت طیفی و معکوس دقت مکانی با SNR متناسب است. سنجنده تصویر برداری برای ارسال دیتای قابل تشخیص نیاز به میزان حداقلی و ثابت SNR دارد همچنین با در نظر گرفتن میزان نویز گوسی با میانگین و واریانس یکسان در سیستم، ملاحظه می‌شود دقت طیفی و دقت مکانی با یکدیگر رابطه عکس دارند و چون دقت طیفی در تصاویر ابرطیفی زیاد است یا به عبارت دیگر $\Delta \lambda$ از نظر عددی کوچک است، دقت مکانی دچار تخریب می‌شود.

۱-۴ سنجنده‌های تصاویر چندطیفی

پیش‌تر گفتیم که تعداد باندها در تصاویر چندطیفی کمتر از تصاویر ابرطیفی است اکنون با دقت به شکل ۴-۱ اثر کاهش تعداد باندها را بر بردار اطلاعات طیفی هر پیکسل مشاهده می‌کنیم. در تصاویر چندطیفی به دلیل بزرگ بودن بازه‌های طول موج برای هر باند تصویر نهایی برای هر پیکسل میانگینی از چند باند مختلف از

تصویر ابرطیفی است. از طرف دیگر باید دقت داشته باشیم که در تصویر ابرطیفی اطلاعات مکانی پیکسل‌ها میانگینی از چند پیکسل از تصویر چندطیفی است.



شکل ۱-۴: تفاوت تصاویر چندطیفی و ابرطیفی در دقت اندازه‌گیری طول موج‌های مختلف از مکان. راست: تصویر یک مکعب ابرطیفی. چپ: تصویر یک مکعب چندطیفی با ۴ باند متفاوت [۱۰].

از جمله ماهواره‌هایی که برای تصویر برداری چندطیفی موجود هستند می‌توان به ماهواره‌های QuickBird2 و WorldView3 اشاره کرد. این ماهواره‌ها تصاویری در باندهای مختلف از زمین اخذ می‌کنند به طور کلی ماهواره WV3 ۴ نوع تصویر متفاوت اخذ می‌کند که یک نوع از آن‌ها تصویر تک‌باند است نوع دوم تصویر چندطیفی با ۸ باند در محدوده دید انسان است که مشخصات این دو تصویر در جدول ۱-۱ نمایش داده شده است. نوع سوم تصویر در طول موج‌های حرارتی، شامل ۸ باند مختلف و نوع چهارم در طول موج‌هایی بزرگتر، شامل ۱۲ باند که مناسب بررسی‌های هواشناسی است تولید می‌شوند. دو نوع تصویر آخر مناسب بررسی و طبقه‌بندی نیستند از این رو ما فقط از نوع ۱ و نوع ۲ برای انجام آزمایش‌های خود استفاده خواهیم کرد.

ماهواره QB2 نیز به طور کلی دو نوع تصویر اخذ می‌کند که اطلاعات این تصاویر نیز در جدول ۲-۱ قابل مشاهده است. با مقایسه این دو ماهواره برتری ماهواره WV3 QB2 نسبت به چه از نظر اطلاعات طیفی و چه از نظر دقت مکانی کاملاً مشهود است.

دقت مکانی بالا در تصاویر چند طیفی باعث می‌شود که مشکل تداخل اشیاء در تصاویر ابرطیفی با دقت مکانی پایین حل شود. این مسئله در تصاویر ابرطیفی با دقت مکانی پایین به این گونه عمل می‌کند که در یک پیکسل چند ماده مختلف وجود دارد و طیف بازتابی این مواد مختلف با هم جمع می‌شوند و عملاً آن پیکسل

جدول ۱-۱: جدول اطلاعات تصاویر ماهواره WV3 (earth.esa.int)

Spectral Range	Band Name	Spectral Band	GSD (Ground Sample Distance)
Panchromatic Band		450 - 800 nm	Nadir: 0.31 m, 20° off-nadir: 0.34 m
MS (Multispectral) Bands (8) in VNIR (Visible Near Infrared)	Coastal Blue	400 – 450 nm	Nadir: 1.24 m 20° off-nadir: 1.38 m 13 bit/sample
	Blue	450 – 510 nm	
	Green	510 – 580 nm	
	Yellow	585 – 625 nm	
	Red	630 – 690 nm	
	Red edge	705 – 745 nm	
	Near-IR1	770 – 895 nm	
	Near-IR2	860 – 1040 nm	

جدول ۱-۲: جدول اطلاعات تصاویر ماهواره QB2 (earth.esa.int)

Spectral Range	Band Name	Spectral Band	GSD (Ground Sample Distance)
Panchromatic Band		450 - 900 nm	Nadir: 0.61 m, 20° off-nadir: 0.72 m
MS(Multispectral) Bands (4) in VNIR (Visible Near Infrared)	Blue	450 – 520 nm	Nadir: 2.4 m 20° off-nadir: 2.6 m 11 bit/sample
	Green	520 – 600 nm	
	Red	630 – 690 nm	
	NIR	760 – 900 nm	

امضای طیفی خاصی از یک ماده واحد را نشان نمی‌دهد و آنچه نمایش داده می‌شود میانگینی از طیف تمام مواد موجود در آن پیکسل است.^[۱۱] اما در تصاویر چندطیفی به جهت دقت مکانی بیشتر احتمال وجود چند ماده در یک پیکسل کمتر است پس می‌توان چنین گفت که تصاویر چندطیفی دقت طیفی پایینی دارند ولی در همان دقت پایین خطای تداخل طیفی کمتر است.

ما علاقه‌مندیم تصاویر چندطیفی را به گونه‌ای طبقه‌بندی کنیم که بتوانیم اشیاء مختلف را در صحنه از هم تفکیک کنیم و همچنین از تصاویر چندطیفی ماهواره‌ها استفاده کنیم دلیل انتخاب ماهواره WorldView3 اول از همه در دسترس داشتن یک داده از این ماهواره است (عدم دسترسی داشتن به تصویری از ماهواره‌های جدیدتر). در تصاویر چندطیفی، معمولاً این گونه طبقه‌بندی (بر اساس اشیاء) بر روی تصاویر چندطیفی با دقت مکانی ۵ سانتی‌متر که با نصب سنجنده بر روی پهپاد اخذ می‌شود، (UAV Multispectral Imagery) انجام می‌گیرد.^{[۱۲][۱۳]} در اکثر موارد طبقه‌بندی بر روی تصاویر چندطیفی ماهواره‌ای به گونه‌ای انجام شده است که یک ناحیه خاص مثلاً ناحیه کشاورزی یا شهری از سایر ناحیه‌های موجود در صحنه جدا شود.^[۱۴]

۱-۵ بیان مسئله و چالش‌های طبقه‌بندی تصاویر سنجش از دور

تصاویر چندطیفی، تصویری از سطح زمین هستند که برای هر پیکسل یک بردار محدود از ویژگی‌های طیفی را دارا هستند ولی خوبی‌خانه به دلیل دقت مکانی زیاد مشکل تداخل طیفی در این تصاویر نسبت به تصاویر ابرطیفی کمتر است. از آنجایی که فرایند طبقه‌بندی برای هر پیکسل انجام می‌شود فرایند طبقه‌بندی کند صورت می‌گیرد. به خصوص در تصاویر چندطیفی که باید از اطلاعات مکانی نیز در کنار اطلاعات طیفی استفاده شود و با توجه به اینکه اطلاعات مکانی نسبت به حجم ذخیره‌سازی و پردازشی آن، اطلاعات مفید کمتری در اختیار ما قرار می‌دهد، فرایند طبقه‌بندی کنتر و دشوارتر صورت می‌گیرد. برای افزایش سرعت طبقه‌بندی راه کارهای پردازش موازی و استفاده از شبکه‌های عصبی بعد از آموزش ارائه شده است.

از جمله مشکلات طبقه‌بندی تصاویر چندطیفی وجود اغتشاشات اتمسفری مانند وجود ابر در هوا و یا آلودگی هوا است که باعث ایجاد اختلال در طیف دریافتی و مشکل تفسیر پذیری اطلاعات مکانی می‌شود برای رفع مشکلات طیفی ناشی از جو، استفاده از اطلاعات مکانی بسیار مفید خواهد بود. علاوه بر موارد ذکر شده تهییه نقشه مرجع به صورت میدانی بسیار دشوار بوده و هزینه زیادی را متحمل می‌شود که در اکثر موارد، محدودیت در تعداد نمونه‌های آموزشی وجود دارد به همین دلیل بهره‌گیری از روشی که بتواند با تعداد نمونه آموزشی محدود به صحتی بالا در زمانی محدود و معقول در طبقه‌بندی برسد و همچنین از نظر ارزیابی بصری کیفیت خوبی را داشته باشد، مد نظر است که در این پایان نامه سعی بر دست‌یابی به این اهداف شده است.

در فصل دوم تئوری‌های مورد نیاز برای انجام روش پیشنهادی بیان می‌شود. در بخش ۱-۲ ساختار چشم و تصاویر تولید شده توسط چشم بررسی می‌شود. در ۲-۲ در مورد شبکه‌های عصبی صحبت‌هایی ارائه می‌گردد و در ۲-۳ ارتباط میان شبکه‌های عصبی و ساختار بینایی انسان و پارامترهای شبکه‌های کانوالوشنی اطلاعاتی در اختیار قرار می‌گیرد. همچنین در ۴-۲ به فیلتربانک گابور و شباهت آن به پردازش‌های تصویر داخل چشم اشاره می‌گردد. در نهایت در بخش ۵-۲ به بررسی معیارهای ارزیابی خواهیم پرداخت.

در فصل سوم آزمایش‌های انجام شده برای بررسی عملکردهای متفاوت فیلتربانک گابور ارائه شده است. در بخش ۱-۳ به معرفی دو نوع پیاده‌سازی فیلتربانک گابور خواهیم پرداخت و در بخش ۲-۳ و ۳-۳ به بررسی توانایی این دو فیلتربانک در زمینه خوشبندی و طبقه‌بندی پرداخته خواهد شد. سپس در ۴-۳ مراحل تولید یک نقشه مرجع برای تصویر چندطیفی بیان می‌شود. در ۵-۳ بر روی داده تولید شده آزمایش‌های گوناگونی

انجام می‌شود که از انجام هرکدام هدف خاصی را دنبال خواهیم کرد. در نهایت در ۲-۵ اطلاعات اولیه روش پیشنهادی ارائه خواهد شد.

در فصل چهارم بعد از بیان یک مقدمه از بخش ۴-۲ تا بخش ۴-۴ به توضیح، پیاده‌سازی و بررسی نتایج روش پیشنهادی خواهیم پرداخت. همچنین در ۴-۵ یک نگرشی دیگر در طبقه‌بندی تصاویر چندطیفی بررسی می‌شود که می‌تواند یک پیشنهاد خوب برای ادامه کار در نظر گرفته شود. در نهایت فصل پنجم، شامل خلاصه‌ای از نتیجه‌گیری، جمع‌بندی و پیشنهادات خواهد بود.

فصل دوم

اصول اولیه و ابزارهای مورد نیاز

۱-۲ بررسی عملکرد و ساختمان چشم انسان

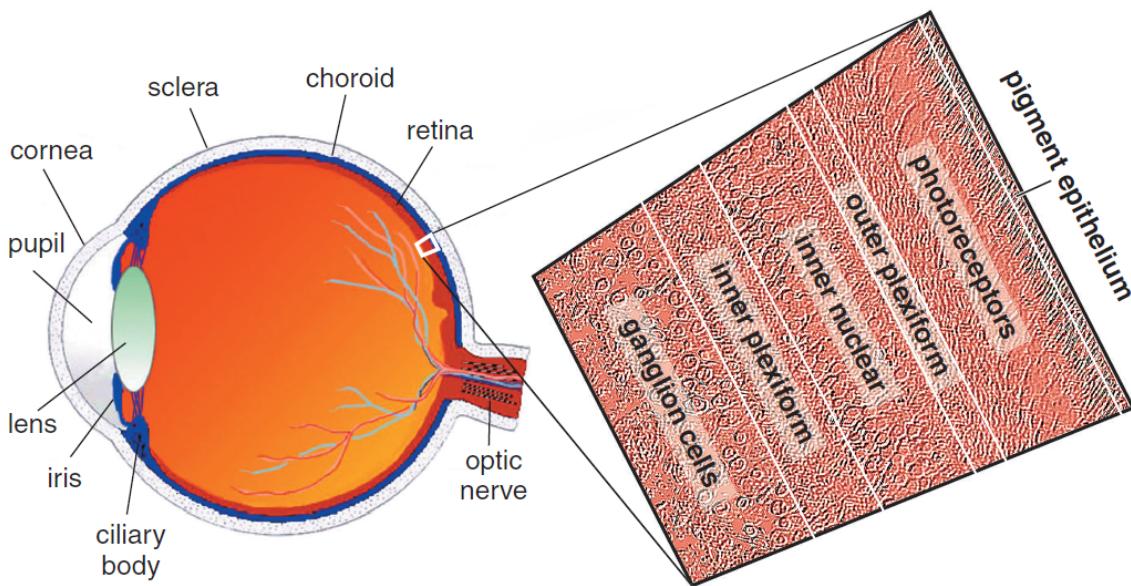
با توجه به شباهت شبکه CNN به ساختار بینایی مغز انسان و موجودات زنده، بررسی ساختمان چشم و تصاویر تولید شده توسط چشم از این نظر برای ما اهمیت دارد تا بدانیم چشم چگونه تصاویری برای این شبکه تولید می‌کند و با پیش‌پردازش‌های لازم بر روی تصاویر، قبل از ورود به شبکه عصبی بینایی مغز آشنا شویم. سپس تا حد ممکن تلاش کنیم تا پیش‌پردازش‌هایی را نظیر پیش‌پردازش‌های چشم، برای تصاویر چندطیفی انجام دهیم. در ادامه، در دو قسمت به طور کامل ساختمان و عملکرد چشم را توضیح خواهیم داد.

۱-۱ ساختمان کلی چشم و شبکیه انسان

مهمن ترین اجزای چشم انسان عبارت‌اند از قرنیه، مردمک، عدسی، عنبه و شبکیه که در اینجا قصد داریم در مورد شبکیه بیشتر صحبت کنیم.

شبکیه دارای قطری کمتر از نیم میلی‌متر است که به صورت یک لایه، داخل تخم چشم قرار گرفته است. نکته جالب این است که شبکیه قسمتی از مغز است. یکی از مهم‌ترین بخش‌های چشم شبکیه است که دارای سه لایه اصلی می‌باشد و هر لایه در فرایند اخذ تصویر وظیفه مهمی بر عهده دارد. پردازش‌های اولیه سیستم بینایی انسان از شبکیه شروع می‌شود و سپس اطلاعات پردازش شده برای پردازش‌های پیشرفته‌تر و درک بینایی توسط نورون‌ها به شبکه عصبی مغز منتقل می‌شوند.^[۱۵]

در ظاهر ممکن است فکر کنیم که سطح شبکیه (منظور از سطح شبکیه همان لایه داخلی شبکیه است که در تماس با مایع داخل چشم است) شامل سلول‌های گیرنده است اما در واقع این سلول‌ها در پشت شبکیه قرار گرفته‌اند و پرتوهای نور برای رسیدن به این سلول‌ها باید از عرض شبکیه عبور کنند. اگر این لایه را لایه‌ی اول بنامیم، در لایه دوم از سه لایه اصلی شبکیه یعنی لایه هسته داخلی یک تا چهار نوع سلول افقی، یازده نوع سلول دوقطبی و همچنین ۳۰ تا ۲۲ نوع سلول آماکرین (که تعداد آن‌ها بسته به فضای موجود متغیر است) وجود دارد و در نهایت در لایه سطحی شبکیه سلول‌های غده‌ای قرار گرفته‌اند که شامل ۲۰ نوع مختلف می‌باشند. همچنین دو زیر لایه دیگر در بین این سه لایه قرار دارد. یکی شبه لایه خارجی، که شامل سیناپس‌هایی است که سلول‌های گیرنده را به سلول‌های افقی، و همچنین به یک سر سلول‌های دو قطبی متصل می‌کند. شبه لایه بعدی، شبه لایه داخلی است. ناحیه که در آن سلول‌های دو قطبی و آماکرین‌ها به سلول‌های غده‌ای متصل می‌شوند. در شکل ۱-۲ ساختمان چشم و شبکیه نمایش داده شده است.^[۱۶]

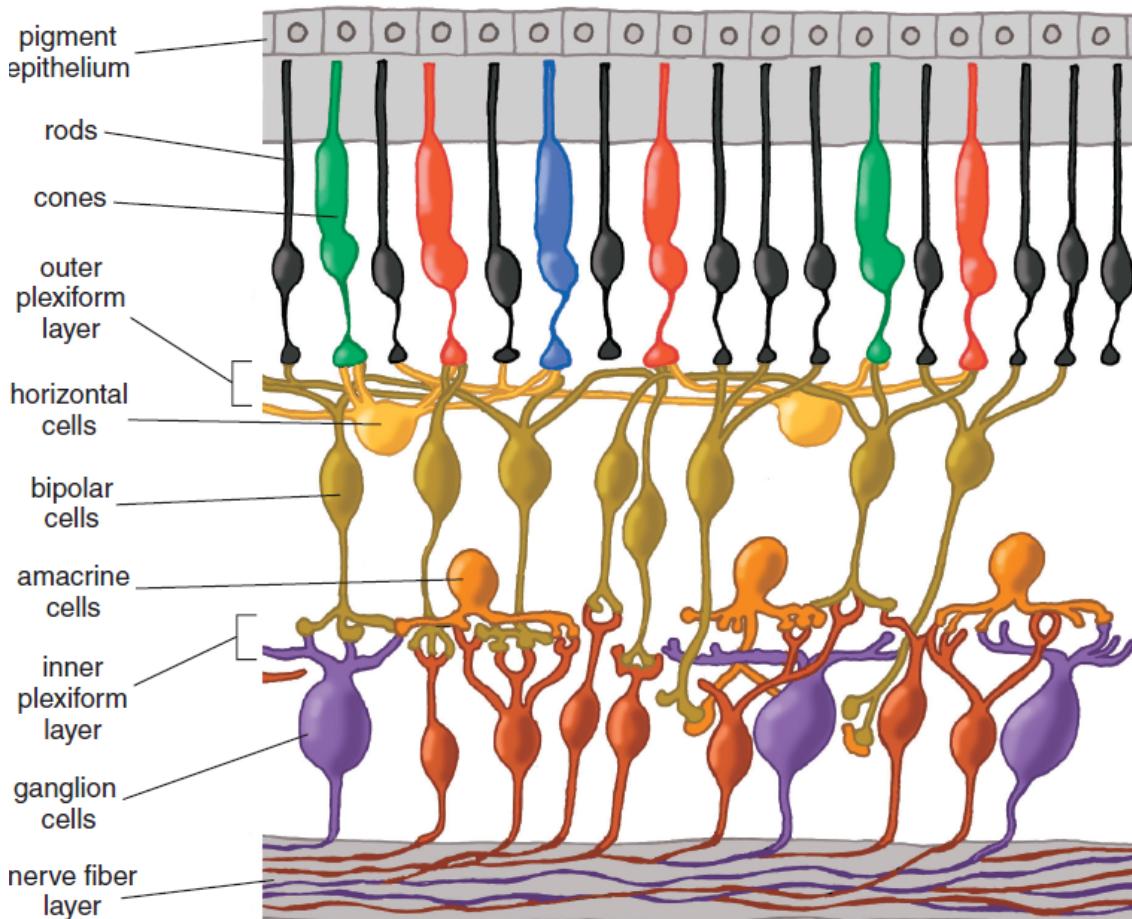


شکل ۱-۲: تصویر ساختمان کلی چشم و لایه‌های شبکیه [۱۶]

سلول‌های گیرنده نور به دو نوع مخروطی و میله‌ای تقسیم می‌شود. نوع میله‌ای برای دیدن در شرایط نوری پایین کاربرد دارد. در حالی که نوع مخروطی برای دیدن در شرایط نوری بالا بهتر عمل می‌کند و از همین رو برای درک تغییرات سریع در محیط عمل کرد خوبی دارد. در حالی که نوع میله‌ای با توجه به کارکرد در شرایط نوری پایین، تغییرات آهسته را بهتر درک می‌کند. تعداد سلول‌های میله‌ای بسیار بیشتر از سلول‌های مخروطی است. سلول‌های میله‌ای حساس به روشنایی هستند و طول موج نور در عملکرد آن‌ها تاثیر ندارد این درحالی است که سلول‌های مخروطی، نسبت به طول موج‌های مختلف حساسیت دارند از این رو ما می‌توانیم اشیاء را رنگی بینیم. نکته مشترک در هر دو نوع سلول‌ها (به طور کلی در سلول‌های گیرنده) این است که فرایند تجزیه تصویر به قسمت‌های جدا از هم از همین بخش شروع می‌شود. یک تصویر در اولین سیناپس‌های سیستم بینایی به کوچک شدن و تجزیه شدن ادامه می‌دهد، یعنی مسیر بین سلول‌ها دوقطبی مختلف و گیرنده‌های مختلف. [۱۷]

قابلیت پردازش موازی شبکه عصبی مغز، باعث شده شبکیه چشم به این گونه عمل کند که یک تصویر را به چند بخش کوچک تبدیل کند از طرفی ویژگی پردازش موازی باعث شده است چشم تصاویر ورودی را به دو صورت دریافت کند که عبارت‌اند از روشن (ON) و خاموش (OFF). تصاویر روشن در واقع شناسایی نواحی روشن در پس‌زمینه تاریک را ممکن می‌سازد و تصاویر خاموش شناسایی نواحی تاریک در پس‌زمینه روشن را ممکن می‌سازد. [۱۸] در این باره در بخش بعد به طور مفصل همراه با سایر ویژگی‌های تولید و استخراج ویژگی

در شبکیه صحبت می‌کنیم. در شکل ۲-۲ نمونه بزرگنمایی شده تصویر شبکیه را مشاهده می‌کنیم و جزئیات دقیق‌تری را پس از مشاهده آن بیان خواهیم کرد.

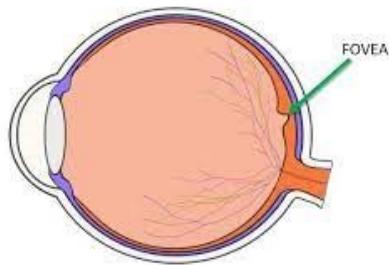


شکل ۲-۲: نمایی از شبکیه و سلول‌های موجود در هر لایه.[۱۶]

تا به اینجا در مورد برخی از خواص و ویژگی‌های سلول‌های شبکیه صحبت کردیم. سلول‌های افقی مانند شکل ۲-۲ به تعداد زیادی میله و مخروط متصل است از این رو یک پهنه‌ای دید وسیع برای ما فراهم می‌کند. از طرفی سلول‌های دو قطبی ناحیه دید محدودی دارند که دلیل آن تعداد کم میله‌ها و مخروط‌های متصل به هر کدام است. آمیکرین‌ها کنار سلول‌های غده‌ای قرار دارند. سلول آمیکرین نقش مهمی در انتقال اطلاعات از میله‌ها، به سلول‌های غده‌ای دارد. این سلول‌ها، پیام‌ها را از تعداد زیادی میله‌های متصل به دو قطبی جمع‌آوری می‌کنند که به ما این امکان را می‌دهد که در شرایط نوری ضعیف قدرت تشخیص داشته باشیم.[۱۸]

برخلاف بقیه شبکه چشم، حفره چشم (یا گودی چشم) شامل سلول‌های غده‌ای کوچک است که شاخه‌های دندانی کوچکی دارند که به نسبت یک به یک به سلول‌های دو قطبی کوچک متصل هستند. حفره در چشم تمام موجودات قرار ندارد. در گونه‌های دارای حفره تصاویر قبل از ارسال به مغز پردازش‌های پیشرفته‌تری انجام

می‌شود به طوری که شبکیه چشم آن‌ها می‌تواند بلافاصله اطلاعات مربوط به حرکت در تصاویر را به مغز منتقل کند.^[۱۶] محل قرار گرفتن حفره چشم انسان در شکل ۳-۲ مشخص شده است. در بخش بعد با تصاویر تولید شده در شبکیه و حفره آشنا می‌شویم.



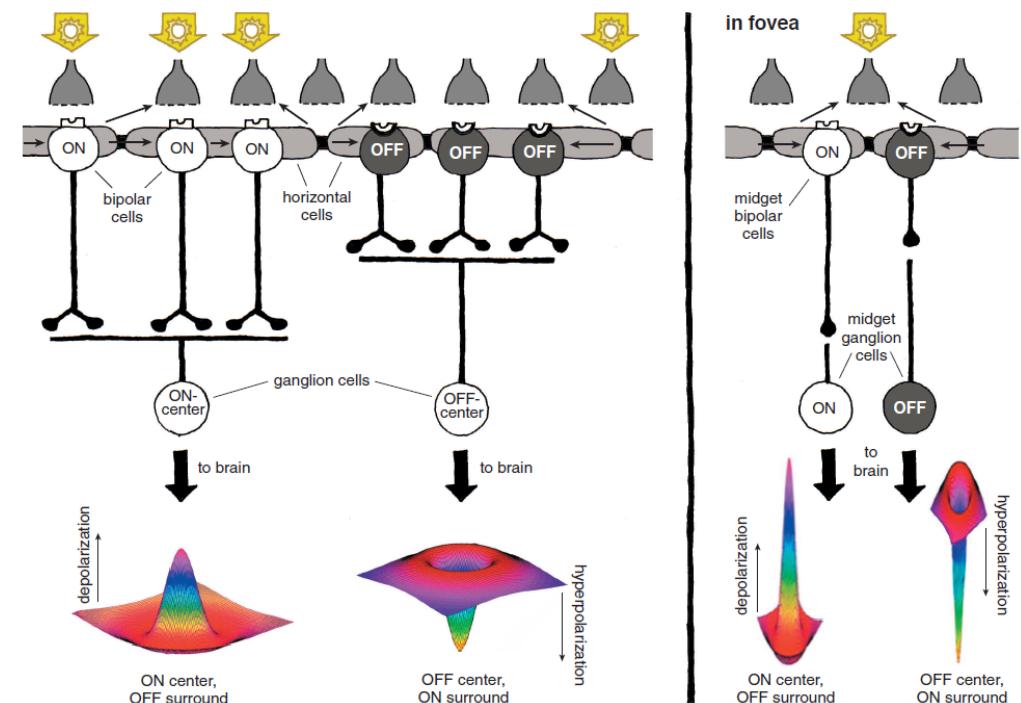
شکل ۳-۲: حفره مرکزی^[۱۶]

۲-۱-۲ ویژگی‌های تصاویر تولید شده در شبکیه و ساختمان جزئی‌تر چشم

همانطور که گفتیم شبکیه تصاویر را در شرایط نوری مختلف در دو نوع روشن و خاموش از هم جدا می‌کند. اگر شبکیه تصاویر با کنتراست مخالف را با هم برای پردازش به مغز منتقل کند، دید حاصل احتمالاً درشت دانه و تار خواهد بود. همچنین در شبکیه پردازش‌هایی نیز برای جداسازی لبه انجام می‌شود و باعث می‌شود بتوانیم روی جزئیات دقیق تمرکز کنیم. صاف کردن تصویر نیز از اولین سطح سیناپسی روی شبکیه آغاز می‌شود. سلول‌های افقی به مخروط‌های متعددی متصل‌اند از این رو میدان دید وسیعی برای ما فراهم می‌کنند. در نتیجه سلول‌های افقی به نور در یک منطقه بزرگ پاسخ می‌دهند. در حالی که سلول‌های دو قطبی تعداد ورودی انگشت شماری از مخروط‌ها را دریافت می‌کنند و بنابراین دارای یک میدان گیرنده با اندازه متوسط هستند. سلول‌های غده‌ای دارای میدان ورودی وسیعی هستند که به صورت دایره‌های متحدم‌المرکز در شبکیه قرار گرفته‌اند که مدار سلول‌های آماکرین در لایه پلکسی، اطلاعات زیادی را به سلول‌های غده‌ای منتقل می‌کنند. آزمایشات اخیر نشان می‌دهد که احتمالاً سلول‌های غده‌ای که با آرایش یاد شده در شبکیه قرار دارند، در تیز کردن لبه‌ها مخصوصاً، لبه‌های مرکز به اطراف نقش بیشتری از سلول‌های افقی دارند. همچنین وظیفه تولید تصاویر مرکز روشن و مرکز تاریک نیز بر عهده سلول‌های غده‌ای است. به این ترتیب که سلول‌های مرکز روشن زمانی فعال می‌شوند که نقطه‌ای از نور در مرکز میدان پذیرنده آن‌ها بتابد، و زمانی که نور به اطراف میدان می‌تابد غیرفعال می‌شوند. نوع فعالیت سلول‌های خارج از مرکز، عکس رفتار فوق را نشان می‌دهد به این صورت که فعالیت آن‌ها

زمانی زیاد می‌شود که محیط میدان گیرنده‌شان روشن باشد و اگر نور به مرکز بتاید فعالیت‌شان کم می‌شود. در واقع نحوه مدل‌سازی میدان‌های پذیرنده سلول‌های غدهای به عنوان تفاوت بین توزیع‌های گوسی (از نظر شدت نور و نه فرکانس) می‌باشد که به آن‌ها شکل کلاه مکریکی می‌گویند.^{[۱۵][۱۶][۱۸][۱۹]}

در حفره به دلیل ساختار یاد شده، کanal سلول‌های دو قطبی کوچک، اطلاعات را از یک مخروط منفرد تا سلول غدهای کوچک حمل می‌کنند. نتیجه آن انتقال یک تصویر نقطه به نقطه از حفره به مغز است. هر مخروط قرمز یا سبز به دو سلول غدهای کوچک متصل می‌شود که یکی برای دید روشن (ON) و یکی برای دید خاموش (OFF) کاربرد دارد. بنابراین پیام منتقل شده به مغز، هم اطلاعات طیفی و هم اطلاعات مکانی بسیار بالایی دارد. بینایی حاصل از حفره برای انجام کارهای دقیق مثل خواندن، نوشتن و حتی رانندگی کاربرد دارد.^[۱۶] برای دیدن خوب باید تصویر را روی حفره مرکزی متتمرکز کنید، دلیل این که شما هنگام خواندن این متن ناخودآگاه سر یا چشمان خود را روی تکان می‌دهید همین است که بتوانید کلمات را روی حفره چشم متتمرکز کنید. در واقع عملکرد حفره مانند سایر شبکیه است ولی با دقت بیشتر و محدوده بینایی باریک‌تر. در شکل ۲-۴ (سمت راست) مدل حفره و سایر چشم (سمت چپ) و سلول‌های روشن و خاموش نشان داده شده است.

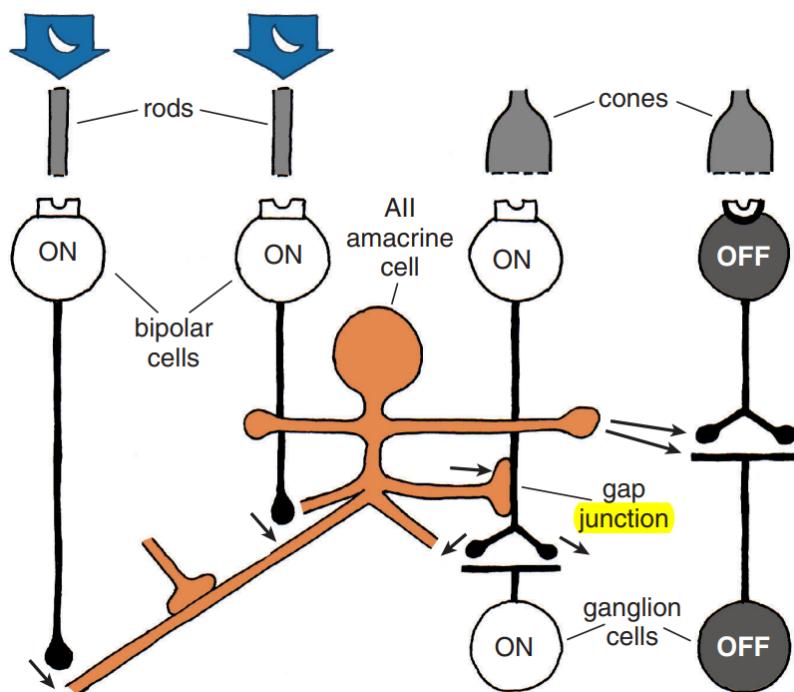


شکل ۲-۴: تفاوت حفره مرکزی و سایر شبکیه در تولید تصاویر روشن و خاموش.^[۱۶]

در بحث قبلی در مورد سلول‌های ON و OFF که از مخروط‌ها سرچشمه می‌گیرند و به سلول‌های غدهای

می‌رسند از صحبت در مورد میله‌ها غافل شدیم. در حالی که سلول‌های مخروطی به طور مستقیم به دوقطبی و سپس به غده‌ای متصل شده‌اند، سلول‌های دوقطبی که ورودی خود را از میله‌ها دریافت می‌کنند، مستقیماً با سلول‌های غده‌ای در ارتباط نیستند.^[16]

سلول‌های دوقطبی متصل به میله‌ها همه از یک نوع هستند و فقط سیگنال ON را ارسال می‌کنند. و از سلول‌های آماکرین "AII" و "A17" به عنوان واسطه برای دریافت سیگنال به سلول‌های غده‌ای استفاده می‌کنند. سلول "AII" اطلاعات را از میدان کوچکی (حدود ۳۰ سلول دوقطبی متصل به میله) جمع‌آوری می‌کند، و پیام جداکنندگی را هم به سلول‌های مخروطی دوقطبی ON (تشخیص نور) و سلول‌های غده‌ای ON و هم به سلول‌های دوقطبی OFF و سلول‌های غده‌ای OFF ارسال می‌کنند. در واقع سلول‌های "AII" در بخش‌های تحت تسلط میله‌ای شبکیه چشم، به عنوان یک درک کننده ثانویه عمل می‌کنند و این نتیجه از معماری سلول‌های مخروطی به غده‌ای حاصل می‌شود. در شکل ۲-۵ نحوه این ارتباط‌ها نشان داده شده است.^[16]

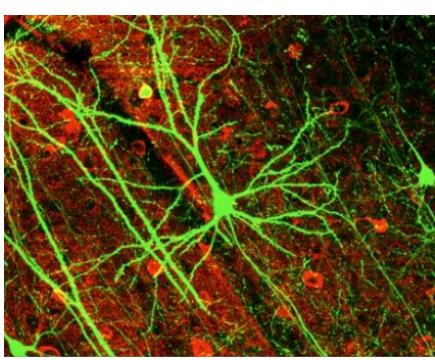


شکل ۲-۵: نحوه اتصال سلول‌های AII و استفاده از سلول‌های خاموش و روشن.^[16]

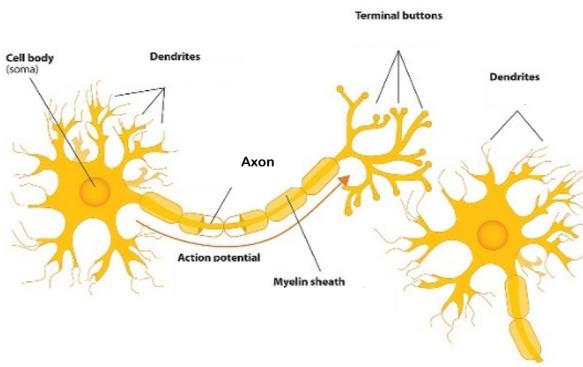
تصاویر تولید شده در شبکیه چشم به کمک جداسازی ON از OFF می‌تواند لبه‌های تیزتری برای ما فراهم کند سپس این تصاویر به مغز منتقل شده و در آنجا همانطور که در بخش قبل توضیح دادیم مورد پردازش‌های پیشرفته‌تری قرار می‌گیرد.

۲-۲ شبکه‌های عصبی

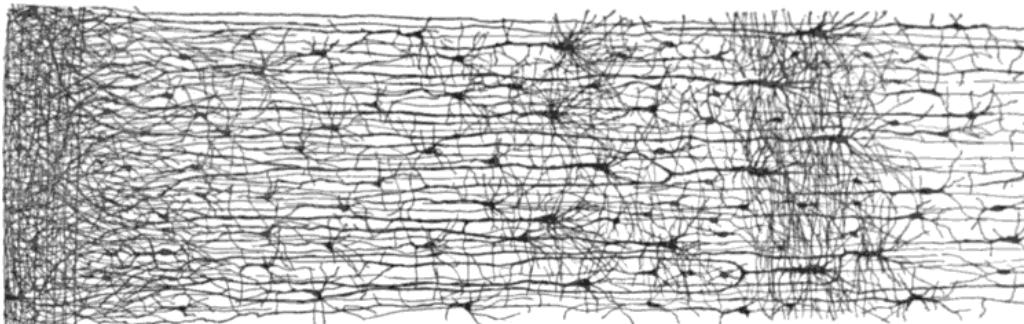
شبکه‌های عصبی یک مدل شبیه‌سازی شده از شبکه‌های عصبی زیستی هستند. عناصر تشکیل‌دهنده یک شبکه زیستی نورون می‌باشد. در شکل ۲-۶(آ) اجزای یک نورون نشان داده شده است. نورون‌ها یک سینگال الکتریکی را از سمت دندریت‌ها دریافت می‌کنند و از طریق اکسون به خروجی منتقل می‌کنند در شکل ۲-۶(ب) نمونه واقعی این نوع سلول‌ها نیز نمایش داده شده است. با ارتباط این نورون‌ها به یکدیگر شبکه‌های عصبی زیستی به وجود می‌آیند. که در شکل ۲-۶(ج) نمونه‌ای از آن‌ها مشاهده می‌شود. در این شبکه‌ها پیغام‌های الکتریکی درون شبکه با وزن‌های موجود در هر سلول درون شبکه انتقال پیدا می‌کند.



(ب) تصویری از نمونه واقعی یک نورون زیستی [۲۱]



(آ) اجزای مختلف یک نورون [۲۰]



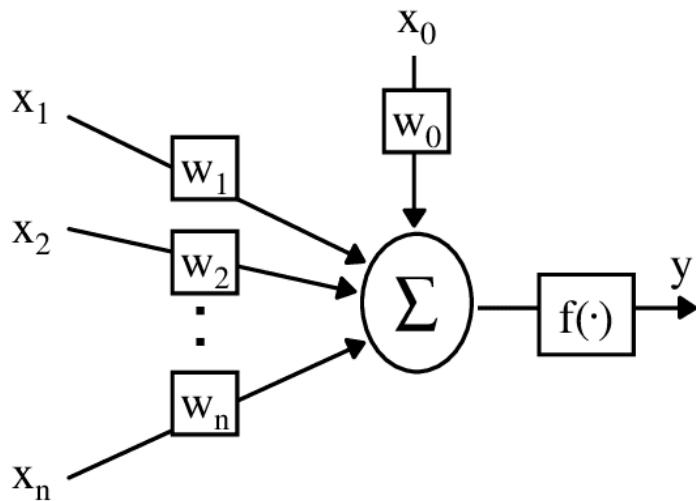
(ج) یک شبکه زیستی متشكل از نورون‌ها [۲۲]

شکل ۲-۶: تصاویری برای درک بهتر شبکه‌های عصبی

۱-۲-۲ نورون‌های مصنوعی

در یک نورون زیستی معمولی، سینگال‌ها پس از دریافت و اعمال یک وزن بخصوص توسط سلول اصلی در هر یک، به سمت دیگر هدایت می‌شود. یک نورون مصنوعی نیز کاری مشابه با نورون زیستی انجام می‌دهد. در

واقع نورون مصنوعی متشكل از یک ضرب شونده و سپس یک جمع شونده است به این صورت که سیگنال‌های ورودی در یک وزن خاص (که نحوه تعیین آن در بخش‌های بعد توضیح داده می‌شود) ضرب می‌شود و سپس با یک مقدار ثابت جمع شده از یکتابع به نام تابع فعال‌ساز عبور می‌کند و به خروجی منتقل می‌شود با کنار هم قرار گرفتن چند نورون یک شبکه عصبی مصنوعی تولید می‌شود. در شکل ۷-۲ خروجی و ورودی از رابطه

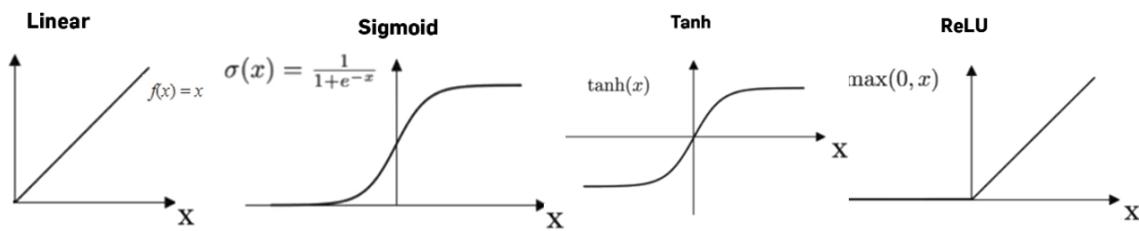


شکل ۷-۲: یک نورون مصنوعی و نحوه اعمال وزن‌ها و مقدار ثابت برای یک مجموعه سیگنال ورودی

۱-۲ پیروی می‌کند:

$$y = f((x_1 w_1 + x_2 w_2 + \dots + x_n w_n) + w_0) \quad (1-2)$$

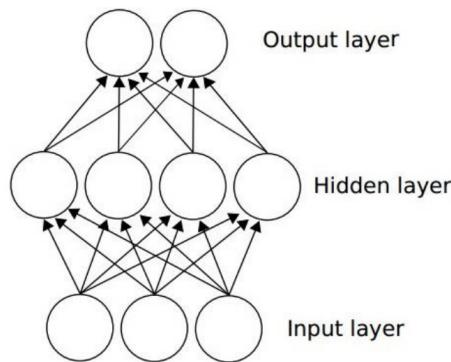
تابع فعال‌ساز (f) وظیفه دارد تا به طور کلی تعیین کند چه مقداری به خروجی منتقل شود. با توجه به این که هیچ محدودیتی برای وزن‌ها و مقدار ثابت وجود ندارد مقدار خروجی می‌تواند به اندازه‌های بسیار بزرگ ($\pm\infty$) برسد. برای اینکه مقادیر خروجی هر نورون کنترل شده باشد، از توابع فعال‌ساز استفاده می‌شود که عموماً خروجی را بین 1 ± 1 محدود می‌کند البته لزومی به حفظ این محدودیت نیست و توابعی نیز وجود دارد که این محدودیت را رعایت نمی‌کنند. توابع فعال‌ساز انواع مختلفی دارد که هر کدام در شرایط و دیتاها مختلف، عملکرد متفاوتی دارند. در شکل ۸-۲ انواع مختلفی از این توابع نشان داده شده است.



شکل ۲-۸: انواع مختلف توابع فعال‌ساز

۲-۲-۲ شبکه‌های پرسپترون چندلایه

پرسپترون چندلایه یک شبکه عصبی انتشار روبه جلو است که متسلسل از چندین لایه متصل به یکدیگر است. در هر لایه تعدادی نورون قرار دارد که به تمام نورون‌های لایه‌ی قبل، متصل است. هرچقدر شبکه‌ای لایه‌های بیشتری داشته باشد، برای انجام محاسبات پیچیده بهتر عمل می‌کند. عموماً توابع Tanh و sigmoid مناسبی برای شبکه‌های پرسپترون چندلایه می‌باشند ولی استفاده از تابع ReLU در شبکه‌های پرسپترون آنچنان رایج نیست. نوع تابع فعال‌ساز یکی از ویژگی‌های مهم یک شبکه است از این رو توابع غیرخطی برای دیتاهای پیچیده مناسب‌تر هستند زیرا باعث می‌شوند ما یک خروجی غیرخطی نسبت به ورودی‌ها داشته باشیم. به هر حال انتخاب درست شبکه، تابع فعال‌ساز، و سایر پارامترهای قابل تغییر در یک شبکه ارتباط نزدیکی با نوع دیتا دارد که تغییر هر کدام می‌تواند بر نتایج نهایی اثرگذار باشد.



شکل ۲-۹: نحوه نام‌گذاری لایه‌ها در یک شبکه پرسپترون چندلایه با یک لایه مخفی [۲۳]

۳-۲-۲ الگوریتم انتشار به عقب

هدف الگوریتم انتشار به عقب تنظیم وزن‌های بین نورون‌ها است، به طوریکه شبکه عصبی مصنوعی خروجی دلخواه را تولید کند. این الگوریتم فرآیند آموزش را توصیف می‌کند. نتیجه این الگوریتم یک شبکه عصبی

پیکربندی شده برای به حداقل رساندن خطا در حل مسئله داده شده است. آموزش باید بر روی داده‌های برچسب‌گذاری شده انجام شود و بنابراین جزء الگوریتم‌های با نظارت قرار می‌گیرد. قبل از اینکه الگوریتم شروع شود، وزن‌ها باید مقداردهی اولیه شوند. مقداردهی اولیه بخشنی از مشخصات الگوریتم نیست، زیرا ممکن است رویکردهای مختلفی وجود داشته باشد. رایج‌ترین مدل مقداردهی اولیه استفاده از توزیع‌های مختلف آماری به صورت اعداد تصادفی برای وزن‌ها می‌باشد. سپس الگوریتم آموزش شروع می‌شود.

هر پترون ورودی را می‌توان به صورت (x_i, d_i) در نظر گرفت که x_i بردار ورودی و d_i برچسب آن است. در اینجا هدف پیش‌بینی برچسب ورودی است. برای این منظور باید وزن‌های شبکه به گونه‌ای تنظیم شوند که بتوان به این مطلوب رسید. این یادگیری در چند تکرار انجام می‌شود که تعداد این تکرارها به پارامترهای مختلفی مانند سایز شبکه و سایز مجموعه آموزش بستگی دارد.

توانایی یادگیری، مفهوم کلیدی شبکه‌های عصبی است. هدف از این فرآیند یافتن پارامترهای مطلوب شبکه برای حل مسئله داده شده است. قبل از شروع آموزش، پارامترهای شبکه باید تنظیم شود. مقادیر اولیه وزن‌ها اغلب به طور تصادفی انتخاب می‌شوند، با این حال با استفاده از برخی از اکتشافات ممکن است تنظیم پارامترها سریعتر انجام شود. یادگیری پس از انجام آموزش در مجموعه داده‌ها توسط تغذیه داده‌های آموزشی از طریق شبکه انجام می‌شود. این یک فرآیند تکراری است که در آن خروجی‌های تولید شده از هر مجموعه ورودی آموزش، تجزیه و تحلیل می‌شود و شبکه به صورت تکراری تنظیم می‌شود تا نتایج بهتر تولید کند. معیارهای مختلفی برای ارزیابی عملکرد شبکه وجود دارد.^[۲۴]

برای بیان نحوه یادگیری شبکه با جزئیات، x^p را به عنوان بردار ورودی و d^p را به عنوان خروجی مطلوب فرض می‌کنیم. که در آن p یک عدد صحیح بین ۱ تا P است که P نشان‌دهنده تعداد کل نمونه‌های آموزشی است. اگر خروجی شبکه (خروجی غیر مطلوب که باید طی فرآیند یادگیری به خروجی مطلوب تغییر کند) را y^p بنامیم این خطاب را نویس j از طریق رابطه‌ی ۲-۲ محاسبه می‌شود. انرژی خطا نیز طبق رابطه ۳-۲ محاسبه می‌شود.

$$e_j^p = y_j^p - d_j^p \quad (۲-۲)$$

$$E_p = \frac{1}{m_k} \sum_{j=1}^{m_k} (y_j^p - d_j^p)^2 \quad (۳-۲)$$

که m_k تعداد نورون‌های خروجی است. اگر این انرژی صفر شود سیگنال خروجی و سیگنال مطلوب باهم برابر می‌شوند که در رابطه ۴-۲ این امر نشان داده شده است. همچنین انرژی میانگین خطأ نیز از رابطه ۵-۲ قابل محاسبه است.

$$\forall_j : E_p = 0 \implies e_j^p = 0 \implies y_j^p = d_j^p \quad (4-2)$$

$$E_{ave} = \frac{1}{P} \sum_{p=1}^P E_p \quad (5-2)$$

در مرحله یادگیری، وزن اتصال به نورون i از لایه $1 - L$ به نورون j از لایه L داده می‌شود. به منظور حداقل کردن انرژی باید این وزن‌ها به گونه‌ای تنظیم شوند که E_{ave} به حداقل مقدار ممکن برسد. مشتق انرژی میانگین نسبت به وزن‌های اتصالی طبق رابطه ۶-۲ محاسبه می‌شود. همچنین اگر از انرژی استفاده کنیم به رابطه ۷-۲ خواهیم رسید. در رابطه ۸-۲ و ۹-۲ دو بخش سمت راست معادله مشتق، با استفاده از انرژی (رابطه ۷-۲) بازنویسی شده‌اند. و در نهایت، رابطه ۱۰-۲ با ترکیب بازنویسی شده معادله مشتق انرژی حاصل می‌شود.

$$\frac{\partial E_{ave}}{\partial \omega_{i,j}} = \frac{1}{P} \frac{\partial}{\partial \omega_{i,j}} \sum_{p=1}^P E_p = \frac{1}{P} \sum_{p=1}^P \frac{\partial E_p}{\partial \omega_{i,j}} \quad (6-2)$$

$$\frac{\partial E}{\partial \omega_{i,j}} = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial \omega_{i,j}} \quad (7-2)$$

$$\frac{\partial E}{\partial y_j} = (y_j - d_j) \quad (8-2)$$

$$\frac{\partial y_j}{\partial \omega_{i,j}} = \frac{\partial y_j}{\partial \zeta_j} \frac{\partial \zeta_j}{\partial \omega_{i,j}} = f'(\zeta_j) \frac{\partial}{\partial \omega_{i,j}} \sum_k \omega_{k,j} y_k = f'(\zeta_j) y_i \quad (9-2)$$

$$\frac{\partial E}{\partial \omega_{i,j}} = (y_j - d_j) f'(\zeta_j) y_i \quad (10-2)$$

اکنون با تعریف گردایان محلی مطابق رابطه ۱۱-۲ می‌توانیم رابطه ۱۰-۲ را مطابق رابطه ۱۲-۲ بازنویسی

کنیم.

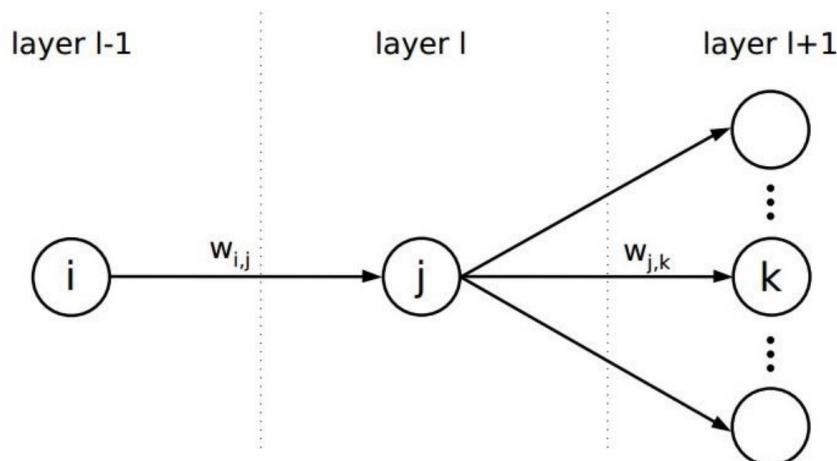
$$\delta_j = \frac{\partial E}{\partial \zeta_j} = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial \zeta_j} = (y_j - d_j) f'(\zeta_j) \quad (11-2)$$

$$\frac{\partial E}{\partial \omega_{i,j}} = \delta_j y_i \quad (12-2)$$

اکنون با استفاده از رابطه ۱۲-۲ که به دست آورده‌ایم، و پارامتر یادگیری $1 < \eta < 0$ تنظیم وزن‌ها مطابق رابطه ۱۳-۲ صورت می‌گیرد.

$$\Delta \omega_{i,j} = -\eta \frac{\partial E}{\partial \omega_{i,j}} = -\eta \delta_j y_i \quad (13-2)$$

تنظیم وزن $\Delta \omega_{i,j}$ که در رابطه ۱۳-۲ به آن رسیدیم، فقط بر روی نورون‌های لایه آخر صورت می‌گیرد. اگر پرسپترون چند لایه شکل ۱۰-۲ را در نظر بگیریم، در لایه‌های میانی روابط ۱۴-۲ تا ۱۸-۲ برای بهروزرسانی وزن‌ها حاصل می‌شود.



شکل ۱۰-۲: نمایش تغییر وزن $\omega_{i,j}$ در لایه مخفی $l - 1$ که بر روی $\omega_{i,j}$ در لایه l اثر می‌گذارد.[۲۵]

$$\delta_j = \frac{\partial E}{\partial \zeta_j} = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial \zeta_j} = \frac{\partial E}{\partial y_j} f'(\zeta_j) \quad (14-2)$$

$$\frac{\partial E}{\partial y_j} = \sum_{k=1}^{m_l+1} \frac{\partial E}{\partial \zeta_k} \frac{\partial \zeta_k}{\partial y_j} = \sum_{k=1}^{m_l+1} \frac{\partial E}{\partial \zeta_k} \omega_{j,k} = \sum_{k=1}^{m_l+1} \delta_k \omega_{j,k} \quad (15-2)$$

از ترکیب دو رابطه ۱۴-۲ و ۱۵-۲ به رابطه ۱۶-۲ خواهیم رسید. در نهایت تنظیم وزن‌ها به صورت رابطه ۱۷-۲ خواهد بود و اگر بخواهیم نحوه به روزرسانی تمام لایه‌ها را نشان دهیم می‌توانیم به سادگی از رابطه شرطی ۱۸-۲ استفاده کنیم.

$$\delta_j = \left(\sum_{k=1}^{m_l+1} \delta_k \omega_{j,k} \right) f'(\zeta_j) \quad (16-2)$$

$$\Delta \omega_{i,j} = -\eta \delta_j y_j \quad (17-2)$$

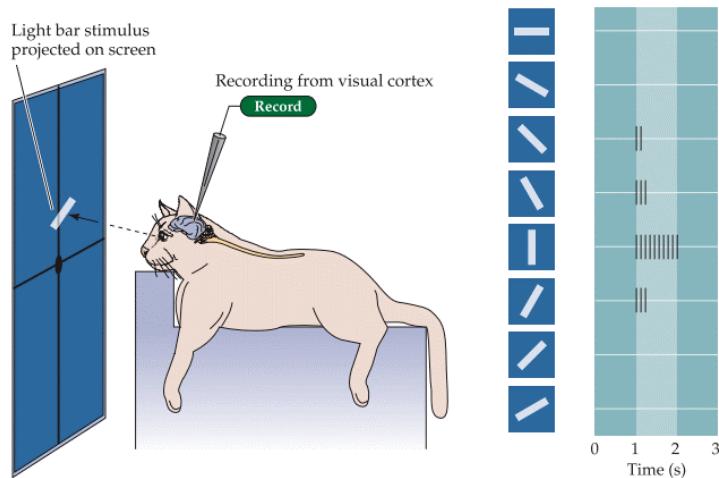
$$\delta_j = \begin{cases} \left(\sum_{k=1}^{m_l+1} \delta_k \omega_{j,k} \right) f'(\zeta_j) & \forall j \quad \text{in} \quad l < L \\ (y_j - d_j) f'(\zeta_j) & \forall j \quad \text{in} \quad L \end{cases} \quad (18-2)$$

۳-۲ شبکه‌های کانوالوشنی

۱-۳-۲ تاریخچه شبکه‌های کانوالوشنی

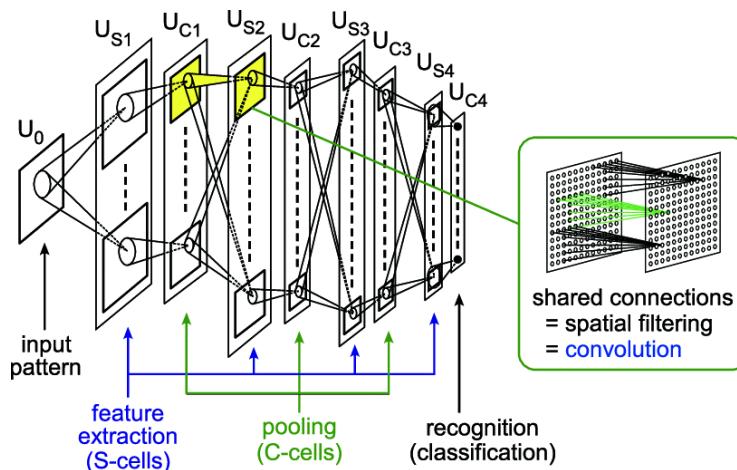
در میانه قرن ۲۰ هابل و ویزل، دو نوع سلول اصلی را در ابتدای سیستم بینایی گربه کشف کردند. نوع اول یعنی سلول‌های ساده به نوارهای روشن یا تاریکی که در مکان خاصی قرار گرفته باشند پاسخ می‌دهند. هر سلولی دارای یک جهت گیری خاص برای نوارها است که در آن جهت پاسخ بیشتری به ورودی چشم خواهد داد و با تغییر زاویه نوار از این جهت‌گیری خاص پاسخ آن سلول کاهش می‌یابد. نوع دوم، سلول‌های پیچیده هستند که روش پاسخ دادن‌شان سختگیری کمتری دارد.^[۲۶] این سلول‌ها نیز جهت‌گیری ترجیحی دارند ولی می‌توانند به یک شدت یکسان، به یک نوار با جهت مشخص، در چند مکان نزدیک به هم پاسخ دهند. هابل و ویزل نتیجه گرفتند که احتمالاً این سلول‌های پیچیده ورودی خود را از چند سلول ساده می‌گیرند، که همه آن‌ها جهت‌گیری ترجیحی یکسانی دارند ولی مکان‌های ترجیحی آن‌ها با هم کمی تفاوت دارند.

در سال ۱۹۸۰ فوکوشیما یافته‌های هابل و ویزل را به یک مدل سیستم بینایی کارآمد تبدیل کرد. این مدل نوادرانکی نام داشت که پیشرو شبکه‌های کانوالوشنی مدرن امروزی است. این مدل شامل دو نوع سلول اصلی است. سلول‌های S از ساده نام‌گذاری شده‌اند و همان وظیفه‌های ساده را پیاده می‌کنند. به طور خاص، یک صفحه دو بعدی از وزن در هر مکان تصویر ورودی اعمال می‌شود تا پاسخ سلول ساده S را تولید کند. به



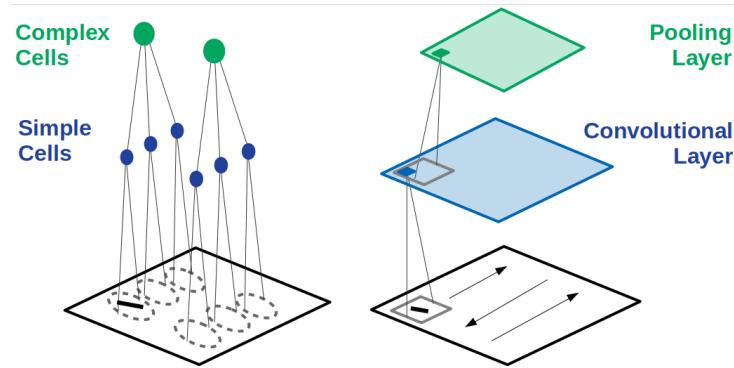
شکل ۱۱-۲: کشف سلول های ساده و پیچیده توسط هابل و ویزل [۲۷]

این ترتیب یک صفحه S کاری که سلولهای ساده انجام می‌دهند را برای لایه بعد فراهم می‌کند. سلول‌های C نوع دوم سلول‌ها هستند که نامشان از پیچیده گرفته شده است. درست مانند سلول‌های پیچیده است و ورودی آن‌ها ازیک صفحه S است ولی با مکان‌های مختلف از همان صفحه.



شکل ۱۲-۲: شبکه نوادرانی و مدل سازی سلول‌های S و C به صفحات S و C [۲۸].

مدل‌های مشابه بسیاری تا سال ۱۹۹۰ ارائه شده است که یکی از برجسته‌ترین آن‌ها HMAX است. که در واقع نمونه به روز شده مدل قبلی است. HMAX از یک تابع ماکسیمم معمولی برای انتقال از صفحه S به صفحه C استفاده می‌کند. در واقع عملکرد آن مانند MAXpooling در شبکه‌های کانولوشنی است. در شکل ۱۳-۲ نمونه‌ای از عملکرد این مدل نشان داده شده است.

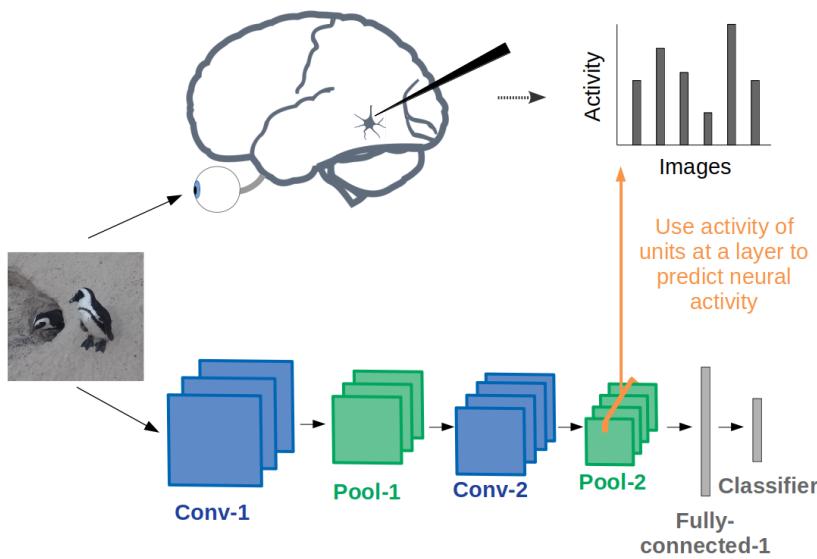


شکل ۲-۳-۲: تشابه شبکه HMAX و لایه maxpooling به مدل سلول های S و C [۲۹]

۲-۳-۲ شبکه کانوالوشنی نماینده‌ای از سیستم بینایی انسان

معماری CNN شباهت‌های زیادی به معماری سیستم بینایی انسان دارد. تصویر وارد شده به این شبکه‌ها معمولاً ابتدا تحت بهنجارش قرار می‌گیرند و معمولاً به سه کanal رنگی مختلف (RGB) تقسیم می‌شوند که محاسبات مشخصی که توسط شبکه انجام می‌شود را ضبط می‌کنند (در واقع پیش پردازش در شبکه CNN همان کاریست که شبکه انجام می‌دهد). در شبکه هر کدام از فیلترهای غیرخطی کانوالوشنی تشکیل یک تصویر جدید می‌دهند که این تصاویر پشت سر هم چیده می‌شوند و به لایه‌ی بعد منتقل می‌شوند. هرچقدر تعداد فیلترهای لایه‌ها بیشتر باشد تصویر منتقل شده به لایه بعد، عمق بیشتر و در عین حال ویژگی‌های بیشتر و در نتیجه پیچیدگی بیشتری خواهد داشت.

یکی از دلایل اصلی که باعث شده شبکه CNN به عنوان شبکه‌ای مشابه سیستم بینایی معرفی شود و در بین دانشمندان نورون‌شناسی نیز جایگاه ویژه‌ای پیدا کند این است که دانشمندان نورون‌شناسی می‌توانند جریان اطلاعات را بهتر دنبال کنند. به طور خاص، زمانی که یک تصویر مشخص به ورودی شبکه CNN و یک موجود زنده نشان داده شود، می‌توان با توجه به فعالیت واحدهای مصنوعی شبکه عصبی فعالیت نورون‌های واقعی موجودات زنده را پیش‌بینی کرد.[۲۹] همچنین در مطالعات اخیر نشان داده شده است که شبکه‌ای که عملکرد بهتری در شناسایی اشیاء دارد، در پیش‌بینی بخش فعالیت موجودات زنده نیز عملکرد بهتری دارد. که می‌توان نتیجه گرفت هرچه مدل شبکه CNN به مدل موجودات زنده نزدیک‌تر باشد، عملکرد بهتری خواهد داشت.



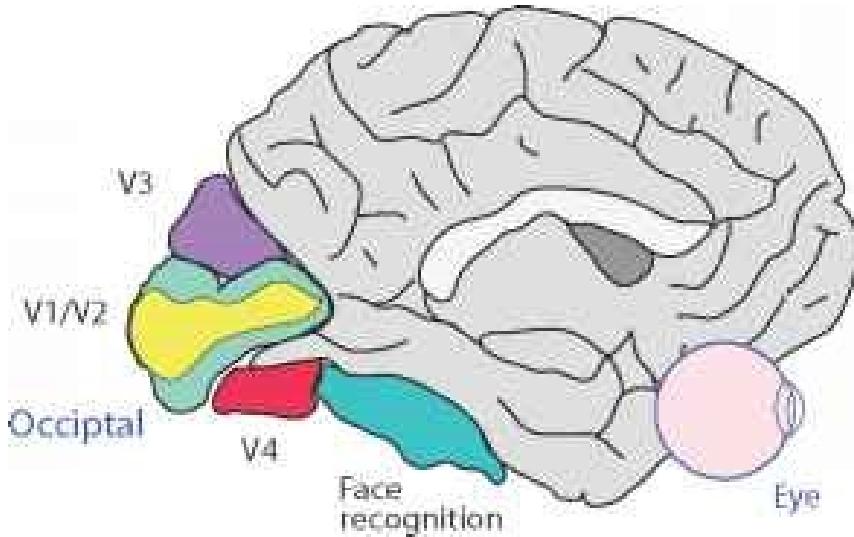
شکل ۱۴-۲: پیش‌بینی مکان احتمالی فعالیت مغزی با کمک از فعالیت شبکه CNN [۲۹]

۳-۳-۲ عملکرد مغز انسان و شبکه CNN

از نظر عملکردی مغز انسان برای بینایی از چند بخش استفاده می‌کند که به ترتیب در اینجا بیان می‌کنیم. اولاً بینایی رنگی به واسطه گیرنده‌های مخروطی ۳ گانه در شبکیه است که به طول موج‌های مختلف پاسخ می‌دهند. در واقع این بخش شروع یک فرایند فوق العاده است که با آن آشنا خواهیم شد. تشخیص دادن رنگ به هر کدام از این طول موج‌ها فرایندی است که توسط سیستم بینایی انجام می‌شود و ما می‌توانیم جهانی رنگارنگ بسته به میزان تشعشع هر طول موج را درک کنیم. پس از دریافت تصویر در شبکیه آکسون‌ها اطلاعات دریافت شده را به واحد LGN منتقل می‌کنند در اینجا تصاویر دو چشم با هم ادغام و هم مرکز می‌شوند و پس از آن به اولین بخش سیستم بینایی V1 منتقل می‌شوند.

نزدیک به ۲۰ الی ۳۰ بخش مختلف برای بینایی انسان در مغز معرفی شده است که هر بخش فرایند و وظیفه خاصی را به همراه دارد. مثلاً لایه V1 برای شناسایی رنگ‌ها عمل می‌کند هرچند با توجه به پیچیده بودن تشخیص رنگ احتمالاً بخش‌های دیگری نیز در آن دخیل است. نقشه‌های بصری در اطراف بخش‌های V1، V2 و V3 ساخته می‌شوند. سپس اطلاعات از دو مسیر، از V1 حرکت می‌کنند و وارد بخش زمانی MT می‌شوند. این بخش وظیفه بررسی حرکت در لحظه‌های مختلف را دارد. به این بخش "مسیر کجاست" گفته می‌شود. بررسی حرکت اشیاء و حتی وظیفه حرکت غیر ارادی چشم برای دنبال کردن اهداف متحرک از وظایف این بخش است. بخش V8 با حافظه بلند مدت خود وظیفه بررسی و شناسایی افراد و اشیاء و به طور

کلی طبقه‌بندی را بر عهده دارد.

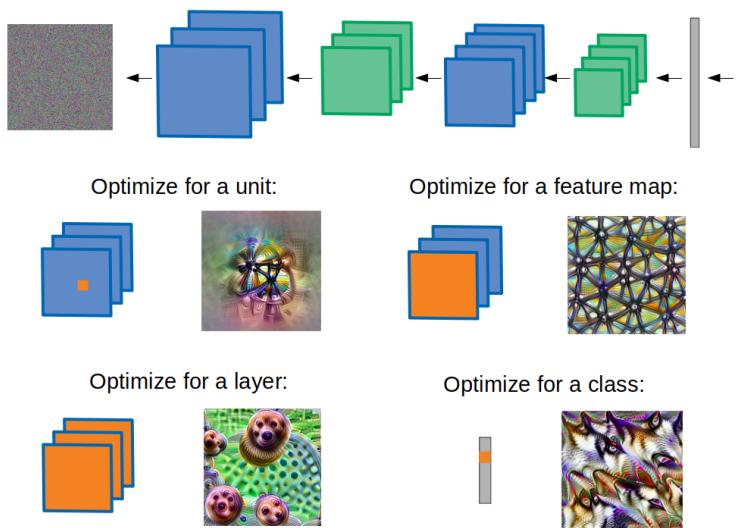


شکل ۲: بخش‌های مختلف مرتبط با بینایی در مغز انسان [۳۰]

در مطالعات زیادی شبکه CNN را با سیستم بینایی موجودات زنده مقایسه کرده‌اند تا به این وسیله بتوانند عملکرد بخش‌های جلویی مغز را توصیف کنند چرا که نحوه پاسخ دادن این بخش‌ها به صحنه‌های مختلف برعکس لایه‌های اولیه بینایی بسیار پیچیده است. در قسمت‌های جلوتر سیستم بینایی فرایندهای پیچیده‌ای مثل تشخیص بافت و شکل‌هایی خاص صورت می‌گیرد که درست شبیه به شبکه‌های CNN است. به طور کلی کرنل‌های ایجاد شده در لایه اول شبکه CNN شبیه به فیلترهایی خواهند بود که هر کدام یک بازه فرکانسی خاص را پوشش خواهند داد که به معنی این است که می‌توانند لبه‌هایی را در جهت‌های مختلف پیدا کنند و یا بافت خاصی را پیدا کنند. [۳۱]

سپس تصویر خروجی هر فیلتر پشت سر هم چیده می‌شود و با اعمال یک فیلتر سه بعدی بر روی این تصویر جدید بررسی می‌شود که در یک بخش خاص از تصویر فیلترهایی که توانایی تشخیص یک شکل خاص را دارند پاسخ درخوری به تصویر ورودی داده‌اند یا خیر. برای مثال اگر فیلتر شماره یک وظیفه یافتن خطوط افقی و فیلتر شماره دو وظیفه یافتن خط عمودی را داشته باشد پس از چیدن تصویر خروجی هر فیلتر روی هم، یک فیلتر از لایه بعدی بررسی می‌کند که اگر در تصویر اول خط عمودی و در تصویر دوم خط افقی در یک مکان خاص باشد، نشان دهنده یک "+" در آن مکان خاص از تصویر است در واقع لایه دوم شبکه وظیفه تشخیص یک واحد مشخص را دارد. و به همین ترتیب لایه سوم از قرار گرفتن این واحدهای تشخیص داده شده کنار هم در می‌یابد که آیا یکسری شکل خاص از واحدها در تصویر وجود دارد یا خیر. در نهایت تصویر خروجی در مسیر حرکت خود ابتدا استخراج ویژگی می‌شود که همان لایه اول است. سپس در لایه بعدی واحدهای

مشخص به وجود می‌آید. که در نتیجه آن هر تصویر خروجی از این لایه نشان دهنده یک سری واحد مشخص در تمام بخش‌های تصویر است (که به آن Feature MAP می‌گویند). از اینجا به بعد تصاویر خروجی یک ویژگی خاص از تصویر را بررسی می‌کنند و در مجموع مشخص می‌کنند که کلاس داده مورد نظر با توجه به این ویژگی‌ها کدام می‌تواند باشد.^{[۳۲][۳۳]}



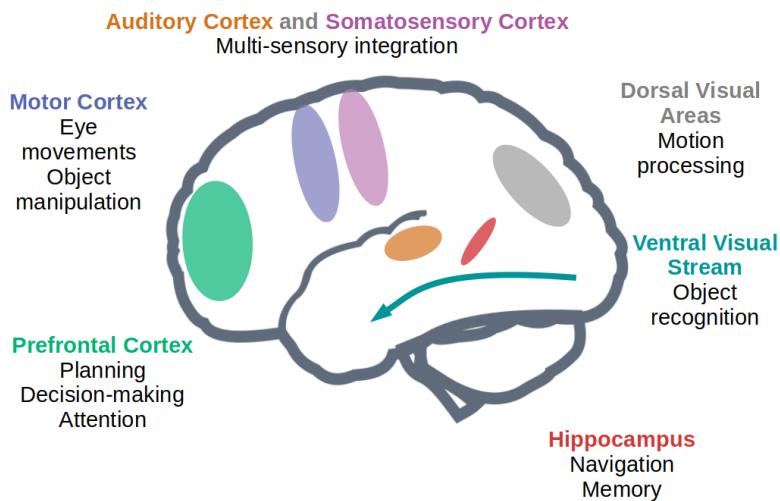
شکل ۲: نحوه عملکرد و وظیفه شبکه CNN^[۲۹]

این خاصیت عملکرد لایه‌ای در شبکه CNN، در مغز انسان به عنوان Ventral Visual Stream شناخته می‌شود. در شکل ۲ بخش‌های مختلف دیگری از مغز انسان و سیستم بینایی مشاهده می‌شود.

۳-۴ شبکه‌های کانولوشنی

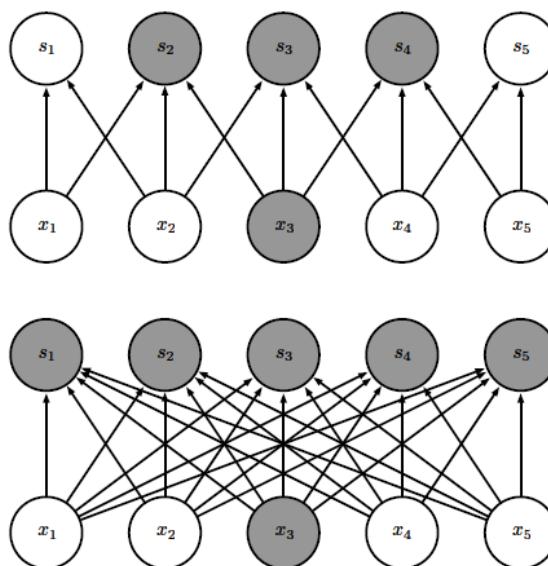
شبکه‌های کانولوشن ساختار خاصی از شبکه‌های عمیق هستند که از مدل بینایی انسان الهام گرفته شده‌اند. شبکه‌های کانولوشن سه ویژگی اساسی را دارا هستند. یکی ارتباطات تنک، اشتراک‌گذاری پارامتر و نمایش معادل. در شبکه‌های عصبی معمولی، هر نورون به کلیه نورون‌های لایه قبل متصل است اما در شبکه‌های کانولوشن هر نورون به تعدادی از نورون‌های لایه قبل متصل است که به آن ناحیه پذیرش گفته می‌شود. در عمل این کار با کوچک‌تر انتخاب شدن هسته کانولوشن انتخاب می‌شود.

نتیجه این ویژگی آن است که تعداد پارامترها برای آموزش در شبکه کمتر می‌شود و همچنین میزان حافظه موردنیاز برای ذخیره آن نیز بسیار کمتر می‌شود. شکل ۲-۱۸ نحوه ارتباط تنک را نشان می‌دهد. در شکل



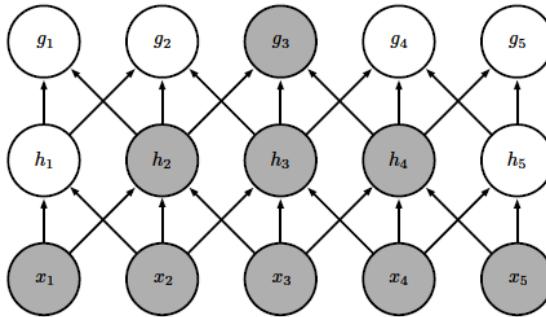
شکل ۱۷-۲: بخش‌های مختلف عملکرد مغز در سیستم بینایی [۲۹]

۱۹-۲ نیز ارتباط نورون‌های لایه‌های بالاتر را نشان می‌دهد که در لایه‌های بالاتر این ارتباط گسترش می‌باید.



شکل ۲: نمایش اتصالات محدود شده در شبکه کانوالوشنی [۳۴]

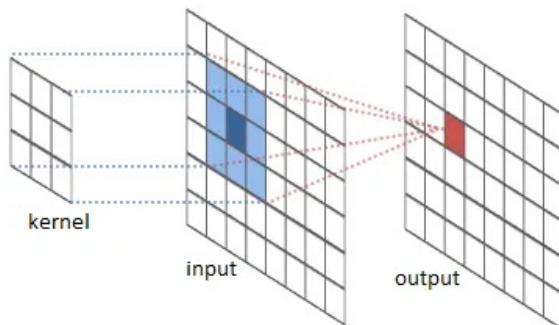
در شبکه‌های کانوالوشنی لایه‌های متفاوت با عملکردهای متفاوتی وجود دارد در ادامه با این لایه‌ها و عملکرد هر کدام به صورت خلاصه آشنا می‌شویم. از آنجایی که ما با شبکه‌های کانوالوشنی دو بعدی سرکار داریم این لایه‌ها را فقط برای حالت دو بعدی بیان خواهیم کرد.



شکل ۲-۱۹: نمایش اتصالات در لایه‌های بالاتر شبکه کانوالوشنی [۳۴]

۱-۴-۳-۲ لایه کانوالوشن دوبعدی

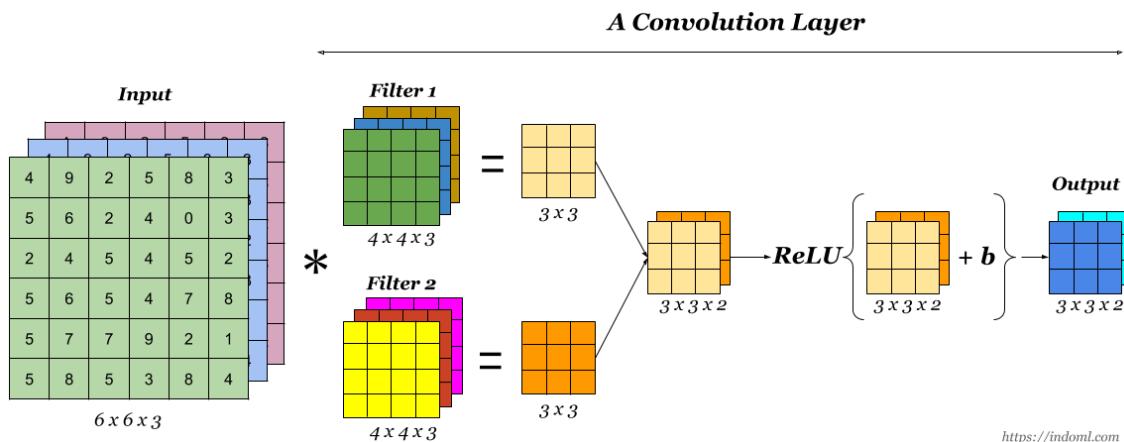
مهم‌ترین لایه‌ی یک شبکه کانوالوشنی همین لایه است. در این لایه اتفاقی نظیر فرایند فیلتر کردن تصویر در حوزه مکان اتفاق می‌افتد. همانطور که می‌دانیم در حوزه مکان، اعمال فیلتر بر تصویر برابر با کانوالو فیلتر بر تصویر است. برای یادآوری کانوالو می‌توانید به شکل ۲-۲۰ دقت کنید. یکی از ویژگی‌های کانوالو این است که اگر تصویر سایزی برابر $B \times p \times q$ داشته باشد و فیلتری با سایز $B \times n \times m$ با این تصویر کانوالو شود خروجی با اندازه‌ای برابر $(m - n + 1) \times (p - q + 1)$ خواهد داشت که نشان‌دهنده کاهش ابعاد تصویر می‌شود. به یک فیلتر دوبعدی اصطلاحاً کرنل نیز می‌گویند.



شکل ۲-۲۰: نحوه اعمال فیلتر بر تصویر در حوزه مکان [۳۵]

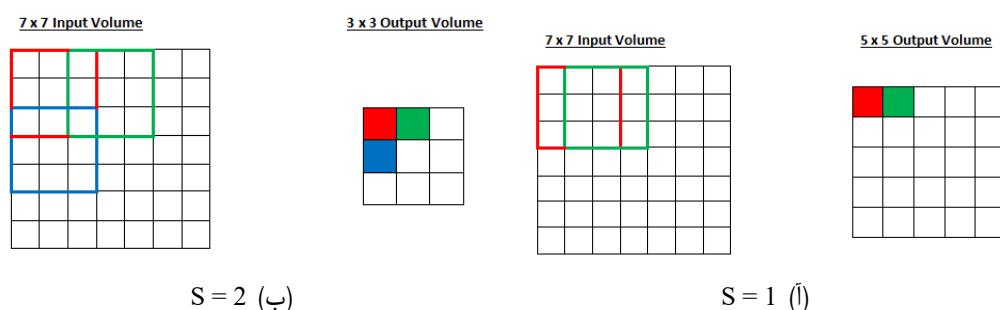
حال باید بدانیم لایه کانوالوشنی چه کاری در شبکه انجام می‌دهد. این لایه یک تصویر با ابعاد $m \times p \times B$ دریافت می‌کند و این تصویر را با K تا فیلتر را با $n \times q \times B$ کانوالو می‌کند تصویر خروجی تمام این K تا فیلتر، یک تصویر جدید تولید می‌کند که عمقی برابر K خواهد داشت. در حالت استاندارد تصویر خروجی باید به فرم $(m - n + 1) \times (p - q + 1) \times K$ در بیاید ولی با تنظیم دو پارامتر به نام‌های padding و stride که از این به بعد آن‌ها را P و S می‌نامیم می‌توان اندازه دو بعد اول را کنترل کرد. برای درک این موضوع در

ابتدا فرض کنید که پارامترهای یاد شده وجود نداشته باشند در یک مثال عددی فرایнд کانوالوشن در این لایه‌ها در شکل ۲۱-۲ نشان داده شده است. مشاهده می‌شود بعد از عملیات کانوالو، خروجی با یک مقدار بایاس جمع شده و از تابع فعال ساز عبور می‌کند.



شکل ۲۱-۲: نحوه عملکرد لایه کانوالوشنی با صرف نظر از وجود پارامترهای S و P

پارامتر S در واقع تنظیم کننده این است که حرکت فیلترها بر روی تصویر با چه گامی صورت گیرد. همانطور که در شکل ۲۰-۲ مشاهده می‌شود در یک کانوالو استاندارد ما فیلتر را به گونه‌ای جابجا می‌کنیم که پس از هر بار اعمال بر پیکسل‌های زیر فیلتر، یک پیکسل به سمت جلو حرکت می‌کنیم و سپس خروجی در مکان جدید قرار می‌گیرد(حتی در حرکت به سمت پایین). به عبارت دیگر فیلتر بر تمام پیکسل‌های تصویر اعمال می‌شود. در این حالت گوییم S برابر با یک است. اکنون اگر S برابر دو باشد حرکت فیلتر بر روی تصویر با گام دو پیکسل انجام می‌شود. در شکل ۲۲-۲ تفاوت دو مقدار برای این پارامتر به صورت عددی نشان داده شده است.



شکل ۲۲-۲: رفتار کانوالوشن در دو بعد اول نسبت به پارامتر S

پارامتر دیگر P به معنای اضافه کردن صفر به اطراف تصویر است. به این شکل که به اندازه این پارامتر در

اطراف تصویر صفر اضافه می‌شود و این تصویر بزرگتر با فیلترها کانوالو می‌شود. برخی ویژگی‌های دیگر این پارامتر این است که می‌توان در آن رشتہ^{'same'} قرار داد که به گونه‌ای در اطراف تصویر صفر اضافه می‌کند که اندازه تصویر خروجی با تصویر ورودی برابر شود. در شکل ۲۳-۲ اثر این پارامتر را بر تصویر مشاهده می‌کنیم.

$$I = \begin{matrix} 3 & 5 & 9 & 1 & 10 \\ 13 & 2 & 4 & 6 & 11 \\ 16 & 24 & 9 & 13 & 1 \\ 7 & 1 & 6 & 8 & 3 \\ 8 & 4 & 9 & 1 & 9 \end{matrix}$$

$$P = 2 \quad (\text{ب})$$

$$J = 'same' \quad (\text{ج})$$

شکل ۲۳-۲: رفتار کانوالوشن در دو بعد اول نسبت به پارامتر $[35]P$

در نهایت با در نظر گرفتن پارامترهای توضیح داده شده اگر تصویر ورودی اندازه‌ای برابر $B \times p \times q \times S$ باشد وارد یک لایه کانوالوشن با K تا فیلتر به اندازه $n \times m \times p \times q$ شود، با در نظر گرفتن پارامترهای P خروجی در دو بعد اول از رابطه ۱۹-۲ پیروی می‌کند.

$$S_{out} = \left\lfloor \frac{S_{in} - F_s + 2P}{S} + 1 \right\rfloor \quad (19-2)$$

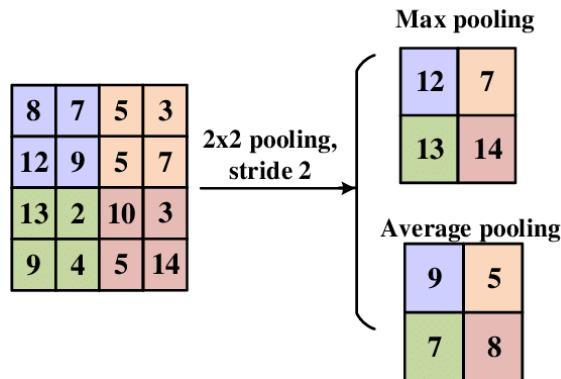
این رابطه برای طول و عرض تصویر صادق است و برای استفاده از آن باید به جای S_{in} ، m یا p قرار گیرد و به جای F_s ، n یا q قرار گیرد تا اندازه خروجی به ترتیب برای طول و عرض تصویر محاسبه شود. دقت شود در نهایت اگر عدد اعشاری شود اندازه خروجی به سمت پایین گرد می‌شود. اکنون خروجی کلی به در نظر گرفتن ۱۹-۲ به فرم رابطه ۲۰-۲ خواهد شد. که A ، B و C به ترتیب اندازه طول، عرض و عمق خروجی می‌باشد.

$$A \times B \times C = \left\lfloor \frac{m - n + 2P}{S} + 1 \right\rfloor \times \left\lfloor \frac{p - q + 2P}{S} + 1 \right\rfloor \times K \quad (20-2)$$

۲-۴-۳-۲ لایه‌های ادغام

این لایه انواع مختلفی دارد. در تمام انواع آن یک صفحه به عمق تصویر ورودی به این لایه بر روی تصویر حرکت می‌کند ولی به جای اعمال فیلتر، بر ناحیه زیرین خود یک مقدار برای آن ناحیه و برای هر عمق تصویر به خروجی منتقل می‌کند. این مقدار می‌تواند، مقدار بیشینه هر یک از صفحات در ناحیه تحت اعمال باشد که

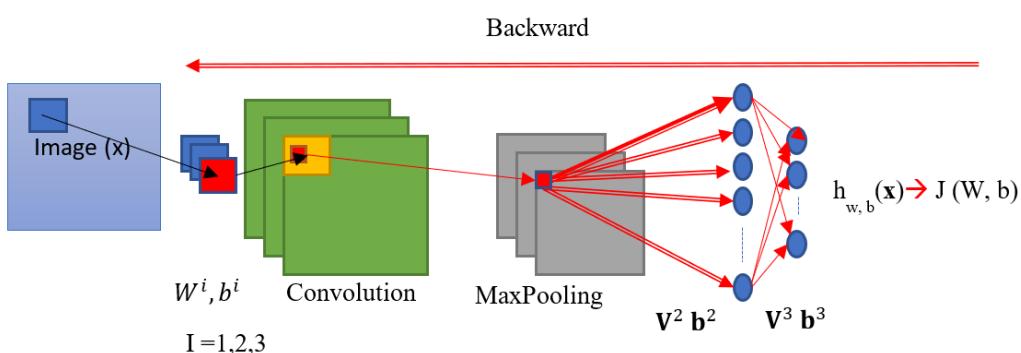
به آن ادغام بیشینه گفته می‌شود یا می‌تواند مقدار میانگین هر یک از صفحات تحت اعمال باشد که به آن ادغام میانگین گفته می‌شود. دو پارامتر S و P در اینجا نیز کاربرد دارد و رابطه خروجی و وردی طبق رابطه ۱۹-۲ می‌باشد. در شکل ۲۴-۲ عملکرد این لایه نشان داده شده است.



شکل ۲۴-۲: عملکرد لایه ادغام [۳۶]

۳-۴-۳-۲ نحوه به روزرسانی وزن‌ها در شبکه کانوالوشنی

اکنون توضیح خلاصه‌ای می‌دهیم که چرا پارامترهای کانوالوشنی زمان بیشتری برای آموزش نیاز دارند البته وارد ریاضیات نمی‌شویم، باید بدانیم که در یک شبکه کانوالوشنی اصلاح وزن‌ها به صورت (EBP) مشابه اصلاح وزن در شبکه‌های معمولی چند لایه است ولی دو تغییر بزرگ دارد به شکل ۲۵-۲ دقت کنید.



شکل ۲۵-۲: نحوه عملیات اصلاح وزن‌ها در شبکه CNN

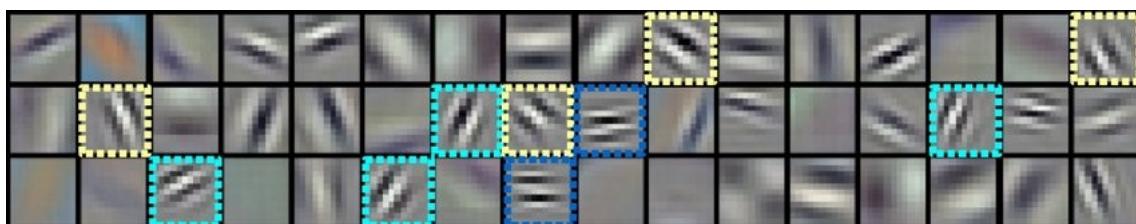
برای این مثال فقط یک لایه ادغام بیشینه و یک لایه کانوالوشنی قرار دارد. ابتدا درمورد لایه ادغام بیشینه صحبت کنیم. می‌دانیم این لایه پارامتری ندارد و قرار نیست در EBP تغییری در این پارامترها حاصل شود چون پارامتری وجود ندارد. در واقع این لایه در مسیر EBP مشخص می‌کند که کدام واحد Feature map باید

انتخاب شود. تمام واحدهایی که از Feature map انتخاب نمی‌شوند، در مسیر تعیین loss قرار نمی‌گیرند. (البته در همان مرحله از آموزش) به طور خلاصه در هر مرحله از آموزش قرار است برخی از واحدهای انتخاب شوند و برخی انتخاب نشوند. بنابراین لایه ادغام اعلام می‌کند که کدام یک از واحدهای کانوالوشنی مسیر انتشار رو به جلو برای تعیین تابع هزینه از نمونه‌های ورودی دخیل هستند و در نمونه‌های ورودی انتخاب‌های لایه ادغام متفاوت است. اهمیت در این است که با این کار برای محاسبه زنجیرهای مشتق، سراغ مسیرهایی خواهیم رفت که در تابع هزینه تاثیر گذار هستند.

حالا باید بدانیم تمام مسیرهایی که در خروجی نقش داشته‌اند از یک بردار پارامتر Feature map استفاده می‌کنند. در واقع مسیر مشتق برگشتی به تمام آن مسیرها(مسیرهایی که در خروجی نقش دارند)، نسبت به این بردار پارامتر مشترک محاسبه می‌شود، تمام این مشتق‌ها باید با هم جمع شوند و برای به روزرسانی آن فیلتر خاص استفاده شوند. به طور خلاصه برای اصلاح وزن‌ها باید بدانیم، وزن‌ها در تابع هزینه چه تاثیری گذاشته‌اند یا به عبارتی از چه مسیرهایی تاثیر خود را بر روی خروجی گذاشته‌اند و سپس مشتق زنجیرهای را برای تمام مسیرها انجام دهیم و مقدار این گرادیان‌ها را با هم جمع کنیم و سپس با توجه به آن وزن‌ها اصلاح شود. به همین دلیل حجم محاسبات شبکه‌های کانوالوشنی بسیار بیشتر از حجم محاسبات شبکه‌های پرسپترون است.

۴-۳-۴-۲ بررسی یک شبکه کانوالوشنی بزرگ

پس از به وجود آمدن ساختار شبکه کانوالوشنی، شبکه‌های بزرگی ساخته و با داده‌های سنگینی آموزش داده شده‌اند که معروف‌ترین آن‌ها عبارت‌اند؛ VGG، inception، ResNet، AlexNet و AlexNet نشان داده شده است. که با کمی دقیق به آن مشاهده می‌کنیم بسیاری از این فیلترها شبیه به فیلتر گابور هستند.



شکل ۲۶-۲: برخی از فیلترهای لایه اول شبکه AlexNet

در سایت www.microscope.openai.com اکثر مدل‌های کانوالوشنی بزرگ آورده شده است و خروجی هر لایه از این شبکه‌ها، که برای یک دسته خاص از داده‌ها، پاسخ‌های قابل ملاحظه‌ای داده‌اند به همراه چند

کanal از خروجی آن شبکه به تصاویر آن دسته از داده، نمایش داده می شود. در شکل ۲۷-۲ تصویر خروجی های چند لایه از شبکه AlexNet و ۳ کanal از این خروجی ها که بیشترین پاسخ را به شکل های دایره ای دارند، نسبت به تصویر ورودی یک ساعت مشاهده می کنیم.



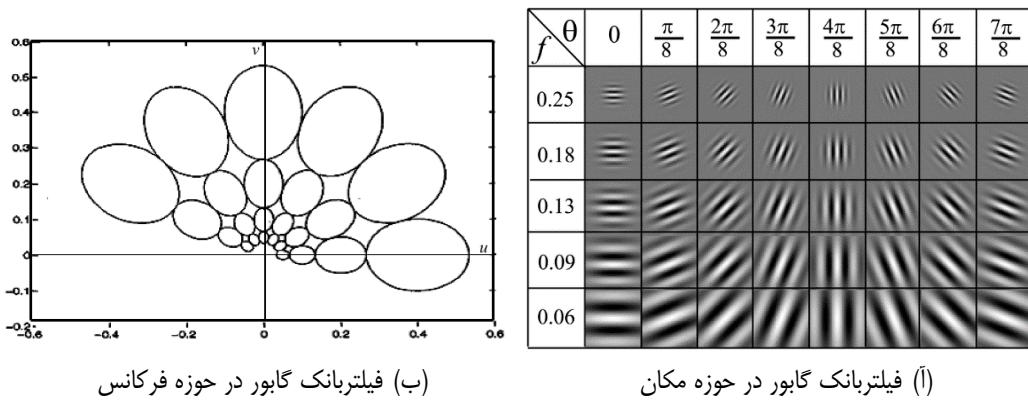
شکل ۲۷-۲: نمایش خروجی ۳ کanal از هر لایه که حساسیت بالایی به ورودی دایره شکل دارد.

در شکل ۲۷-۲ مشاهده می کنیم در لایه های بالاتر شکل خروجی فاصله بیشتری از تصویر ورودی دارد به گونه ای که کم کم این تصویر غیر قابل تشخیص برای چشم می شود در واقع در لایه های بالاتر ویژگی های تصویر است که موجب پاسخ قابل توجه در خروجی می شود.

۴-۴ فیلتربانک گابور

فیلتربانک گابور در بین بسیاری از دانشمندان بینایی، یکی از مشابه ترین ابزارها به سیستم بینایی انسان می باشد.[۳۷] این فیلتربانک در عین کارایی بسیار بالا در استخراج ویژگی های بافت قابلیت پیاده سازی بسیار ساده ای دارد. یک فیلتربانک گابور متشکل از چند فیلتر گابور با پارامترهای مختلف است. به طور کلی یک فیلتر

گابور دو پارامتر اصلی دارد که یکی طول موج فیلتر (فرکانس فیلتر) است و دیگری جهت فیلتر می‌باشد. در ادامه در مورد این پارامترها توضیحاتی ارائه خواهد شد. پیش از شروع بیان ریاضیات برای درک بهتر روابط به شکل‌های ۲۸-۲(آ) و ۲۸-۲(ب) دقت کنید، این دو شکل به ترتیب نشان دهنده فیلتربانک گابور در حوزه مکان و حوزه فرکانس می‌باشند.



شکل ۲۸-۲: نمایش فیلتربانک گابور [۳۸]

با کمی دقت به شکل فیلتربانک گابور کاملاً مفهوم جهت فیلتر نشان داده می‌شود. از طرف دیگر مفهوم فرکانس فیلتر گابور نیز مشخص است به طوری که هرچه فرکانس کمتر می‌شود طول موج‌های مشخص شده در شکل ۲۸-۲(آ) بزرگتر می‌شود. با توجه به شکل ۲۸-۲(ب) نیز می‌توانیم مشاهده کنیم که فیلتربانک گابور یکسری اندازه مشخص در حوزه فرکانس دارد که در جهت‌های مختلف چرخیده است. در واقع چرخش طیف حول مرکز نشان دهنده چرخش تصویر است که در حوزه مکان همان جهت فیلتر نامیده می‌شود.

لازم به تذکر است که بیضی‌های مشخص شده در طیف فرکانسی در واقع توابع گوسی با واریانس‌های مختلف است. برای چرخش تصویر (همان جهت‌های فیلتر گابور) باید سیگنال در حوزه مکان در توابع سینوسی و کسینوسی بر حسب بردار X و Y ضرب شود. با توجه به این نکته که توابع سینوس و کسینوس باعث شیفت فرکانسی در حوزه فرکانسی می‌شود و تابع گوسی در حوزه فرکانس یک گوسی می‌شود انتظار خواهیم داشت که فیلتر گابور از حاصل ضرب یک تابع گوسی در یک تابع سینوسی و کسینوسی بر حسب دو بعد مکانی باشد. اکنون به روابط فیلترهای گابور دقت می‌کنیم.

$$g(x, y) = s(x, y)w(x, y) \quad (21-2)$$

$$s(x, y) = \exp(j(2\pi(ux + vy) + P)) \quad (22-2)$$

$$w(x, y) = K \times \exp\{-\pi(a^2(x - x_0)_r^2 + b^2(y - y_0)_r^2)\}, \quad (23-2)$$

طبق رابطه ۲۱-۲ فیلتر گابور از ضرب دو بخش پوش گوسی و سینوسی مختلط تشکیل می‌شود. سینوسی مختلط در رابطه ۲۲-۲ آورده شده است که در این رابطه u و v فرکانس‌های مکانی و P آفست فاز است که معمولاً مقدار صفر برای آن در نظر گرفته می‌شود. بخش پوش گوسی در رابطه ۲۳-۲ آورده شده است که در آن K میزان مقیاس پوش گوسی، (a, b) مقدار مقیاس محورها و (x_0, y_0) مختصات پیک گوسی هستند. همچنین رابطه ۲۳-۲ به صورت قطبی نوشته شده است که طبق روابط ۲۴-۲ و ۲۵-۲ به دست می‌آید.

$$(x - x_0) = (x - x_0) \cos \theta + (y - y_0) \sin \theta \quad (24-2)$$

$$(y - y_0) = -(x - x_0) \sin \theta + (y - y_0) \cos \theta \quad (25-2)$$

فیلتر گابور دو بعدی به دلیل همبستگی زیاد پیکسل‌های تصویر در راستای x و y عملکردی مؤثر، در استخراج ویژگی‌های مکانی بافت دارد اما در حالت کلی‌تر به خصوص در داده‌های ابرطیفی باندهای مختلف این گونه داده نیز با یکدیگر میزانی از همبستگی را دارند که با استفاده از فیلتر گابور سه بعدی که بعد λ نیز در آن وجود دارد می‌توان این همبستگی‌ها را بررسی کرد. در [۳۹] از این روش برای طبقه‌بندی تصاویر ابرطیفی استفاده شده است. یک بیان دیگر برای فیلتریانک دو بعدی گابور در رابطه ۲۶-۲ آمده است.

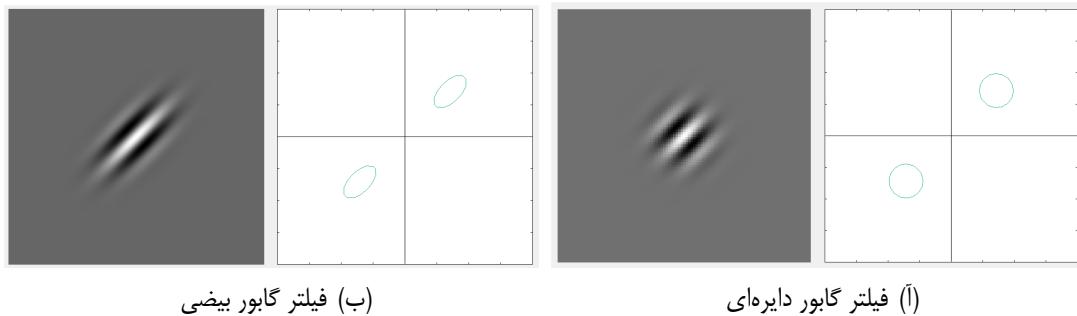
$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left[\left(\frac{x}{\sigma_x}\right)^2 + \left(\frac{y}{\sigma_y}\right)^2\right]} \cos(2\pi(F_x x + F_y y)) \quad (26-2)$$

که در آن F_x و F_y فرکانس قسمت سینوسی در جهت x و y می‌باشد و σ_x و σ_y انحراف معیار پوش گوسی در جهت x و y می‌باشند. نکته قابل توجه این است که اگر σ_x و σ_y برابر باشند فیلترها در حوزه فرکانس به شکل دایره خواهند بود (متقارن) و اگر نابرابر باشند به شکل بیضی (نامتقارن) در خواهند آمد. در شکل ۲۹-۲ می‌توان این پدیده را مشاهده کرد. به طور استاندارد اندازه واریانس‌های گوسی به گونه‌ای انتخاب می‌شود که دو طول موج کسینوسی در داخل تابع گوسی قرار بگیرد و منظور آن است که ۹۷ درصد از انرژی کل گوسی متناظر با اندازه‌ای باشد که دو طول موج در آن قرار بگیرد.

۲-۵ معیارهای ارزیابی

از جمله معیارهای اندازه گیری می‌توان به موارد زیر اشاره کرد:

دقت طبقه‌بندی کلاس: به نسبت تعداد نمونه‌هایی که طبقه‌بند در یک کلاس درست پیش‌بینی می‌کند



[شکل ۲-۲۹: نمایش دو نوع فیلتر گابور] [۳۹]

به تعداد کل نمونه‌ها در همان کلاس دقت طبقه‌بندی کلاس می‌گویند. اگر N_{pi} تعداد نمونه‌های کلاس i باشند که درست طبقه‌بندی شده‌اند و N_i تعداد کل نمونه‌های کلاس i م باشد، آنگاه دقت طبقه‌بندی کلاس i طبق رابطه ۲۷-۲ محاسبه می‌شود.

$$Acc_i = \frac{N_{pi}}{N_i} \quad (27-2)$$

متوسط دقت: به میانگین دقت کلاس‌ها متوسط دقت گفته می‌شود که طبق ۲۸-۲ محاسبه می‌گردد.

$$AA = \frac{\sum Acc_i}{C} \quad , \quad C = number of classes \quad (28-2)$$

دقت طبقه‌بندی کل: نسبت تعداد نمونه‌هایی که توسط طبقه‌بند درست پیش‌بینی شده‌اند به تعداد کل نمونه‌ها. در رابطه ۲۹-۲، N_{ii} تعداد نمونه‌های کلاس i هستند که درست طبقه‌بندی شده‌اند و N_{ij} در صورت ناکام بودن i و j برابر تعداد نمونه‌های کلاس i است که به اشتباه، کلاس j تشخیص داده شده‌اند. و اگر برابر باشد برابر نقاط طبقه‌بندی شده صحیح هستند.

$$OA = \frac{\sum N_{ii}}{\sum \sum N_{ij}} classes \quad (29-2)$$

اعتبار: به نسبت تعداد نمونه‌های درست تشخیص داده شده در یک کلاس، به کل نمونه‌هایی که به آن

کلاس تعلق گرفته‌اند اعتبار آن کلاس گفته می‌شود. که طبق رابطه ۳۰-۲ قابل محاسبه است.

$$V_i = \frac{N_{pi}}{\sum N_{i,j}} \quad (30-2)$$

ماتریس درهم‌ریختگی: اطلاعات مربوط به یک طبقه‌بند را بصورت یک جدول ارائه می‌کند. تمامی پارامترهای گفته شده در بالا را می‌توان توسط این ماتریس محاسبه کرد.

ROC (Receiver Operating Characteristic): معیار ROC یک معیار اندازه‌گیری برای طبقه‌بندی‌های دو کلاسه است یعنی این معیار اندازه‌گیری، زمانی کاربرد دارد که طبقه‌بند ما قرار است بین وجود یا عدم وجود یک هدف، یا صحیح و غلط بودن یک پدیده تصمیم‌گیری کند. برای مثال در یک رادار، یک هدف می‌تواند وجود داشته باشد و رادار وجود هدف را تایید یا رد کند. به طور کلی این معیار دو پارامتر مهم دارد. TPR(True Positive Rate) و FPR(False Positive Rate) این دو معیار در رابطه‌های ۳۱-۲ و ۳۲-۲ مشخص شده است.

$$TPR = \frac{TP}{TP + FN} \quad (31-2)$$

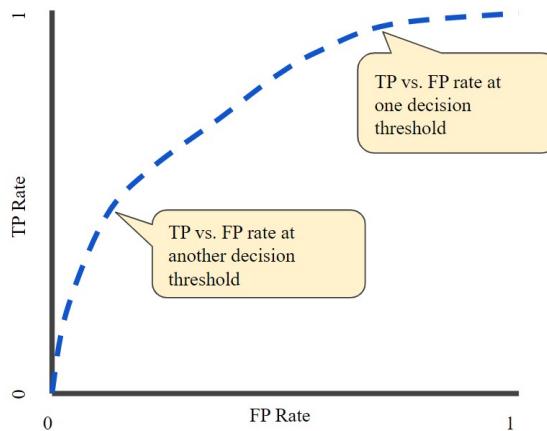
$$FPR = \frac{FP}{TN + FP} \quad (32-2)$$

متغیرهای موجود در رابطه‌های بالا طبق ماتریس درهم ریختگی دو کلاسه مانند شکل ۳۰-۲ نام گذاری شده‌اند.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

شکل ۳۰-۲: جایگاه پارامترهای متفاوت در ماتریس درهم ریختگی [۴۰]

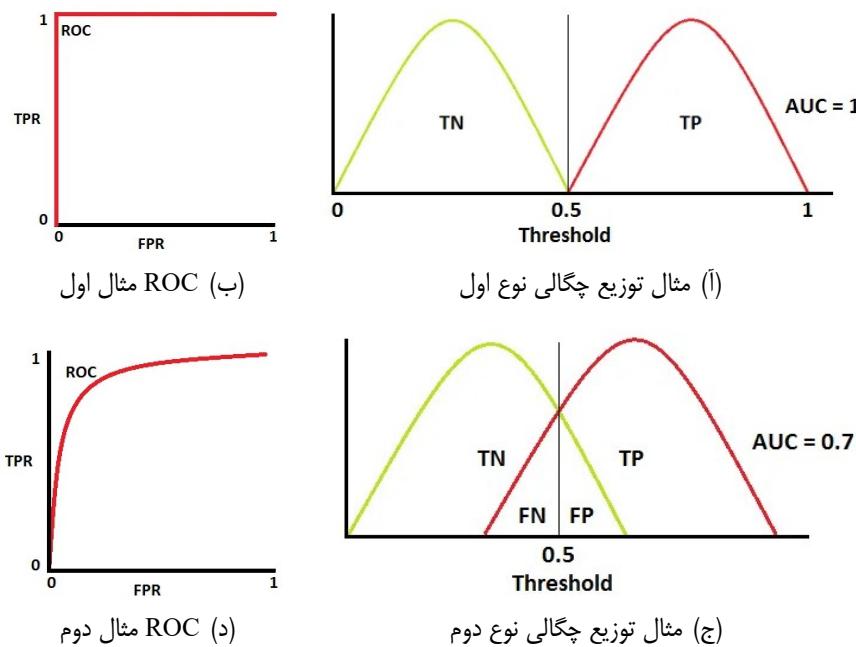
اکنون در یک طبقه‌بندی کننده دو کلاسه مانند یک رادار یا تشخیص یک هدف (Target Detection) با تغییر آستانه تصمیم‌گیری (Threshold)، می‌توان به اعداد مختلف برای دو پارامتر بالا دست یافت. با رسم نمودار TPR، FPR به نمودار منحنی ROC (ROC curve) دست می‌یابیم که در مقابل آستانه‌های متفاوت، نقاط متفاوتی در صفحه FPR/TPR در اختیار ما قرار می‌دهد. نمونه‌ای از این نمودار در شکل ۳۱-۲ آورده شده است.



[۴۱] نمودار ROC

ROC یک تابع توزیع احتمال است یکی از ویژگی‌های این منحنی این است که در نقطه $(0, 0)$ مقداری برابر صفر دارد و در نقطه $(1, 1)$ مقداری برابر یک دارد. یکی از معیارهای ارزیابی با منحنی ROC مقدار AUC(Area Under The Curve) است که برابر با مساحت زیر منحنی می‌باشد. با توجه به این نکته اگر ما آستانه را به گونه‌های متفاوت تغییر دهیم (یا حتی در شرایطی ثابت در نظر بگیریم) که منحنی ROC شرط صفر و یک در ابتدا و انتهای منحنی را رعایت کند، در این صورت بیشینه تفکیک پذیری مدل را نشان داده‌ایم که برابر با AUC خواهد بود و مقداری بین ۰ و صفر خواهد داشت که هرچه به یک نزدیک‌تر باشد تفکیک کنندگی طبقه‌بند بیشتر است. به عنوان مثال به شکل ۳۲-۲ دقت کنید. در شکل ۳۲-۲(آ) و ۳۲-۲(ج) نمودار چگالی احتمال دو حالت متفاوت از تشخیص دو کلاسه آورده شده است. در شکل ۳۲-۲ تفکیک پذیری مثال اول بیشتر از مثال دوم است. [۴۰]

اکنون برای حالت چند کلاسه معیار ROC به این گونه است که برای هر کلاس در مقابل سایر کلاس‌ها منحنی ROC رسم شود. نکته‌ای که وجود دارد این است که یک طبقه‌بندی کننده قابلیت این را دارد که این معیار را به ازای آستانه تصمیم‌گیری‌های متفاوت اندازه‌گیری کند ولی در یک شبکه عصبی ما نمی‌توانیم آستانه تصمیم‌گیری را برای کلاس‌های متفاوت تغییر دهیم. به عبارت دیگر آستانه تصمیم‌گیری در شبکه عصبی ثابت



شکل ۲-۳۲: نمایش دو حالت متفاوت از ROC [۴۰]

و بهینه‌ترین نقطه انتخاب می‌شود. حالا دو پارامتر TPR و FPR برای هر کلاس ثابت خواهد بود.

پارامتر TPR برای هر کلاس طبق رابطه ۳۱-۲ برابر با اعتبار هر کلاس می‌باشد اما به طور کلی چون پارامترهای TPR و FPR برای محاسبه منحنی ROC استفاده می‌شوند کاربرد و اطلاعات مفیدی در برای نتایج طبقه‌بندی ما وجود ندارد و پارامتر TPR نیز همان اعتبار هر کلاس است که در نتایج آورده شده است.

(یکی از راه‌های تغییر آستانه تصمیم‌گیری در شبکه‌های عصبی، تغییر تعداد داده‌های آموزشی و یا تغییر نمونه‌های انتخاب شده برای آموزش است در هر کدام از این حالت‌ها ما نخواهیم توانست به منحنی ROC با شرایط ذکر شده بررسیم. زیرا با تغییر تعداد نمونه‌های آموزشی به مقادیر بسیار کم مثلاً یک نمونه، با هر بار اجرای برنامه، به جهت نقطه شروع تصادفی وزن‌ها در شبکه، منحنی ROC به صفر نخواهد رسید و با افزایش تعداد نمونه‌های آموزشی نیز این منحنی به یک نمی‌رسد. به طور کلی این منحنی در شرایطی که آستانه تصمیم‌گیری به صورت دقیق قابل جابجایی باشد معیاری قابل اعتماد خواهد بود در غیر این صورت اعتبار هر کلاس نماینده‌ای از این معیار خواهد بود.)

فصل سوم

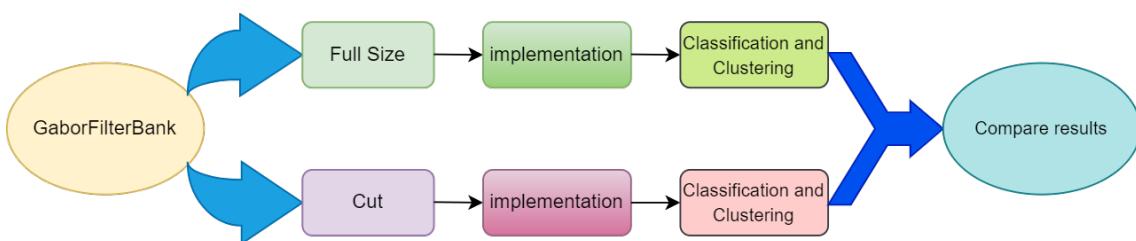
عملکرد طبقه‌بندها و ویژگی‌های متفاوت بر روی دیتاست‌های مختلف

چنان که در فصل ۱ گفتیم قرار است در این فصل نتایج شبیه‌سازی‌های متفاوت را با روش‌های متفاوت استخراج ویژگی و طبقه‌بندهای متفاوت نشان دهیم تا اثر هر کدام از روش‌های استخراج ویژگی را نشان دهیم.

۳-۱ معرفی روش‌های استخراج ویژگی‌های مکانی با فیلتربانک گابور

با توجه به این که فیلتربانک گابور به صورت کامل ممکن است اندازه‌های بزرگی داشته باشد، سعی می‌کنیم نتایج طبقه‌بندی با فیلترهای کامل و فیلترهایی که اندازه آن‌ها برای کوچکتر شدن، برش خورده‌اند را نشان دهیم و مزایا و معایب هر کدام را برشماریم (به طور کلی دو روش استخراج ویژگی).

ابتدا باید بدانیم از جمله علت‌هایی که علاقه داریم نتیجه استخراج ویژگی با فیلترهای برش خورده را بررسی کنیم، قابلیت قرارگیری فیلتر برش خورده در شبکه‌های CNN می‌باشد (که در فصل ۴ به این موضوع می‌پردازیم). یکی دیگر از دلایل ما این است که در اندازه‌های خیلی بزرگ همبستگی پیکسل مرکزی با پیکسل‌های دور کمتر می‌شود در نتیجه اطلاعات استخراج شده از پیکسل‌های دور از مرکز، برای پیکسل مرکزی می‌توانند با سایر کلاس‌ها تداخل داشته باشند در نتیجه ویژگی کلاس‌های مختلف، همبستگی زیادی با یکدیگر خواهند داشت. در شکل ۳-۱ بلوک دیاگرام فرایند بررسی عملکرد این دو فیلتر نشان داده شده است.



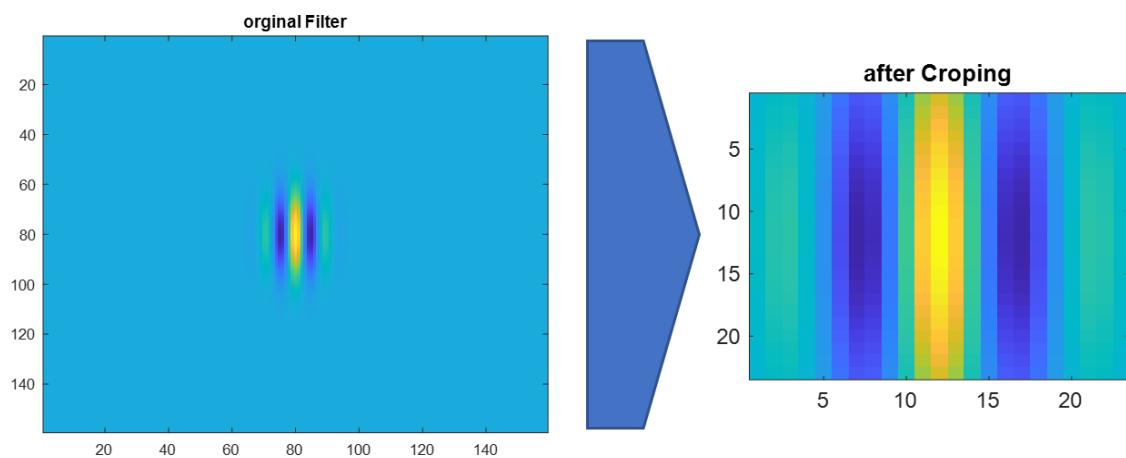
شکل ۳-۱: فرایند بررسی و مقایسه عملکرد دو نوع فیلتربانک گابور که در ادامه نحوه کار و نتایج بررسی خواهد شد.

۳-۱-۱ روش اول، استفاده از فیلتربانک گابور کامل

در این روش ما فیلترهای گابور را با همان اندازه کامل خود که در فصل قبل نمایش دادیم استفاده می‌کنیم. نتایج این روش در کنار سایر نتایج در بخش بعد نمایش داده و مقایسه خواهد شد.

۲-۱-۳ روش دوم، استفاده از فیلتربانک گابور برش خورده

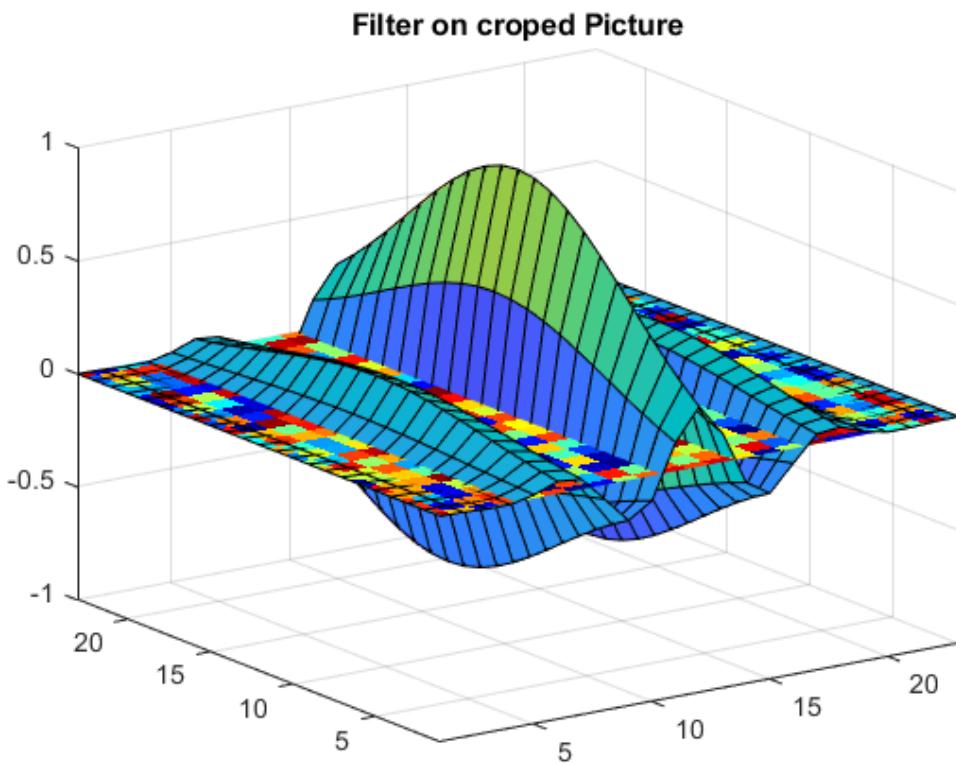
در این روش ما برای فیلتربانک یک اندازه معین انتخاب می‌کنیم، سپس هر فیلتری که اندازه بزرگتری از حد تعیین شده توسط ما داشته باشد برش خورده و کوچک می‌شود. اندازه تعیینی توسط ما بایست عدد فردی باشد تا فیلتر به صورت قرینه در حوزه مکان باقی بماند. نکته دیگر این است که با توجه به اندازه تعیین شده، اندازه فیلترهای اولیه نیز زیاد بزرگ نباشد تا اطلاعات زیادی از فیلترها از بین نروید. در شکل ۲-۳ نحوه برش فیلترها نمایش داده شده است. برای مثال در این شکل تمام فیلترها را با سایز ۲۳ برش می‌دهیم. با دقت به شکل



شکل ۲-۳: نمونه برش یک فیلتر گابور برای انجام آزمایشات بر روی آن‌ها

در می‌یابیم این کار مثل ضرب فیلتر در یکتابع مربع دو بعدی است که طبیعتاً نتایجی مثل ایجاد لوب فرعی در طیف فرکانسی خواهد داشت، در حال حاضر با چشم پوشی از این اثر صرفاً به دنبال نتایج حاصل هستیم تا بینیم برش فیلترها برای قرارگیری در شبکه CNN نتایج مطلوب را خواهد داشت؟ و آیا این کار ویژگی‌های کلاس‌های متفاوت را با کاهش همبستگی همراه خواهد کرد یا نه. مجدداً تاکید می‌شود در این روش نباید اندازه فیلترهای اولیه زیاد بزرگ باشد چرا که برش با یک اندازه مشخص، در این فیلترها اطلاعات زیادی را از بین می‌برد و در نهایت نسبت توان سایدلوب‌ها به توان سیگنال اصلی فیلتر افزایش می‌یابد.

در استخراج ویژگی‌های مکانی هر فیلتر گابور اطلاعات بافت به خصوصی را استخراج می‌کند. این ویژگی‌ها که از پیکسل مرکزی تا چند پیکسل مجاور آن اندازه‌گیری و استخراج می‌شود، همگی به پیکسل مرکزی نسبت داده می‌شود. برای مثال به شکل ۳-۳ دقت کنید. در شکل ۳-۳ مشاهده می‌کنیم فیلتر برش خورده بر روی پیکسل مرکزی تا همسایگی به اندازه ۱۱ پیکسل از پیکسل مرکزی اعمال شده است. در واقع ما برای کاهش همبستگی ویژگی‌های استخراج شده استفاده از فیلترهای کوچک را توصیه می‌کنیم البته درست است که برش



شکل ۳-۳: اعمال فیلتر برش خورده شکل ۲-۳ جهت استخراج ویژگی برای پیکسل مرکزی تصویر فرضی در صفحه

فیلتر نیز همبستگی ویژگی‌های مختلف را کاهش می‌هد ولی دلیل اصلی بررسی ما بر روی این فیلترها الزام اندازه یکسان فیلترها، در شبکه‌های CNN است.

۲-۳ فیلتربانک کامل و برش خورده و مقایسه عملکرد آن‌ها در خوشبندی

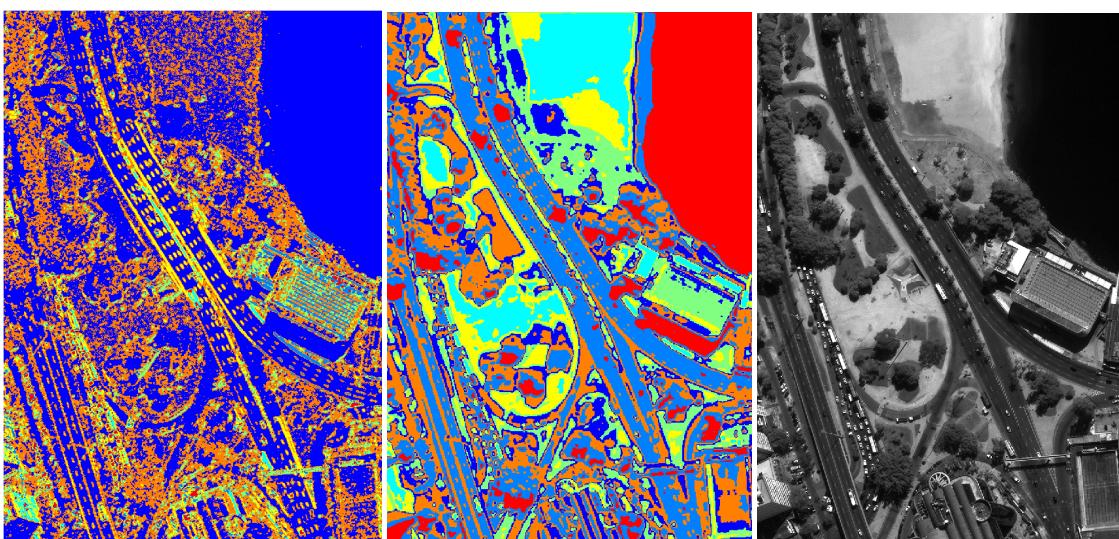
یکی از نکاتی که در مورد فیلتربانک‌ها وجود دارد این است که بتوانند به نحوی عمل کنند که هر فیلتر آن، برای هر کلاس متفاوت، ویژگی‌های متفاوت استخراج کند. به عبارتی فاصله بین کلاسی زیاد باشد. هر چقدر فاصله بین کلاسی بیشتر باشد الگوریتم‌های خوشبندی بهتر عمل خواهند کرد. در مورد آزمایشات باید به این نکات توجه کنیم که در هر دو حالت فیلترها بر روی یک تصویر PAN اعمال شده‌اند و ویژگی‌های ما همگی ویژگی‌های مکانی هستند. در ابتدا اطلاعات فیلتربانک‌ها را در جدول ۱-۳ مشاهده می‌کنیم. (این آزمایشات به

جدول ۱-۳: اطلاعات فیلتربانک‌های مورد آزمایش برای مقایسه عملکرد آن‌ها در خوشبندی

	فیلتربانک گابور با اندازه کامل	فیلتربانک گابور برش خورده
طول موج بر حسب پیکسل	$\{3, 5, 9, 15\}$ pixel	3 pixel
اندازه برش	9 pixel	unCut
زاویه‌ها	$\{0, 30, 60, 120, 150\}^\circ$	$\{0, 30, 60, 120, 150\}^\circ$

دفعات زیاد انجام شده است و در اینجا ما نتایج بهترین حالت مورد آزمایش را قرار داده‌ایم،) باید دقت داشته باشیم چنانچه در فصل قبل گفتیم در یک فیلتر گابور معمولاً تابع گوسی به شکلی انتخاب می‌شود که تقریباً ۲ طول موج سینوسی در آن جای بگیرد.

برای فیلتربانک با اندازه کامل بهترین نتیجه زمانی حاصل می‌شود که فقط از یک فیلتر کوچک استفاده کنیم. در شکل ۳-۴(آ) تصویر اصلی مورد آزمایش نمایش داده شده است در شکل ۳-۴(ب) نتیجه خوشبندی با فیلتربانک برش خورده و در شکل ۳-۴(ج) نتایج خوشبندی با فیلتربانک کامل مشاهده می‌شود.

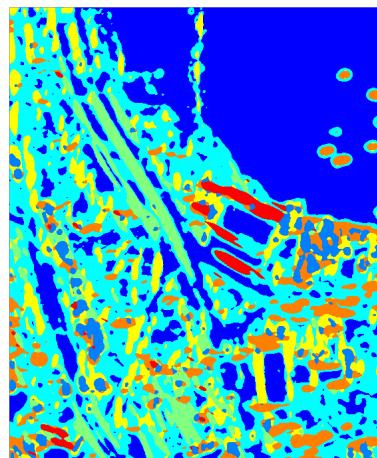


(آ) تصویر اصلی (ب) نتایج فیلترهای برش خورده (ج) نتایج فیلترهای کامل

شکل ۳-۴: نتایج خوشبندی

الگوریتم خوشبندی مورد استفاده در این آزمایش‌ها، الگوریتم K-means است. در هر دو آزمایش تعداد کل کلاس‌های خروجی برابر ۷ کلاس است. همانطور که مشاهده می‌شود شکل ۳-۴(ب) خروجی مناسب‌تری نسبت به شکل ۳-۴(ج) به ما ارائه داده است. می‌توان گفت برش فیلترهای گابور با اینکه فرایندی است که به اجرار دچار آن هستیم تا بتوانیم آن‌ها را در شبکه CNN قرار دهیم، باعث شده است ویژگی‌های اشیاء مختلف موجود در صحنه فاصله بیشتری نسبت به هم داشته باشند. در شکل ۳-۵ نیز نتیجه همین خوشبندی با ۴

فیلتر با اندازه ۲، ۴، ۸، ۱۶ که برش نخورده‌اند و با اندازه اصلی مورد استفاده قرار گرفته‌اند نشان داده شده است. این فیلترها همان فیلترهای آزمایش قبل در حالت برش نخورده‌اند که نتیجه شکل ۳-۴(ب) را به همراه داشته‌اند اما این بار نتیجه را بدون برش مشاهده می‌کنیم.



شکل ۳-۵: خوشبندی تصویر بدون برش فیلترها و استفاده از فیلتر سایزهایی که در حالت برش نخورده نتیجه شکل ۳-۴(ب) را به همراه داشته‌اند.

یکی از دلایلی که نتیجه خوشبندی با فیلترهای کامل اما کوچک (شکل ۳-۴(ج)) نتیجه بهتری نسبت به فیلترهای بزرگ (شکل ۳-۵) دارد این است که اطلاعات موجود در همسایگی‌های دور، بر روی پیکسل مرکزی کشیده نمی‌شود. دقیقاً به همین دلیل است که نتیجه خوشبندی با فیلترهای برش نخورده بهترین نتیجه را به همراه داشته است. یکی از مشکلات خوشبندی مرزهای اشیاء هستند که راه حل‌های برای برخورد با آن‌ها موجود است ولی چون از حوصله تحقیق ما خارج است برای بهبود این خوشبندی علاقه چندانی نداریم و هدف ما بررسی نتیجه فیلترهای برش نخورده گابور است.

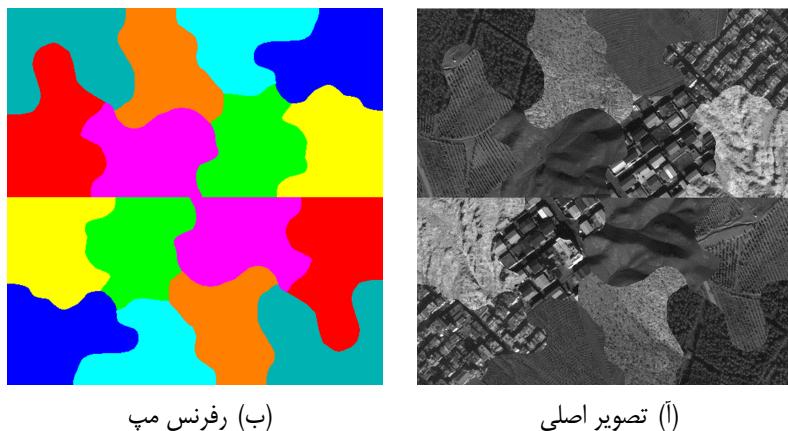
۳-۳ فیلتربانک کامل و برش نخورده و مقایسه عملکرد آن‌ها در طبقه‌بندی

در بخش قبلی در مورد نتایج خوشبندی مشاهداتی انجام دادیم. این بار در مورد طبقه‌بندی آزمایش‌هایی را بر روی چند دیتاست که دارای نقشه مرجع هستند بررسی‌هایی انجام می‌دهیم. در اینجا نیز هدف ما بررسی نحوه عملکرد دو نوع فیلتربانک است از این رو فقط از دو طبقه‌بند ML و طبقه‌بند CNN استفاده می‌کنیم و هدف ما نیز بهبود طبقه‌بندی نخواهد بود بلکه فقط بررسی نتایج در شرایط یکسان است. همچنین این آزمایش را بر

روی ۲ دیتاست مختلف انجام می‌دهیم تا بتوانیم به طور صحیح عملکرد هر دو نوع فیلتربانک را بررسی کنیم.

۱-۳-۳ آزمایش طبقه‌بندی با دیتاست اول

دیتاست اول یک تصویر مصنوعی از مناطق مختلف ایران است که کنار یکدیگر چسبانده شده‌اند. این تصویر دارای ۶ کلاس است که از هر کلاس دو تکه متفاوت در تصویر گنجانده شده است. نقشه‌مرجع این تصویر نیز به صورت ناحیه‌ای تولید شده است یعنی برای ناحیه شهری یک کلاس، برای ناحیه کشاورزی یک کلاس و ... در شکل ۳-۶ تصویر این دیتاست به همراه نقشه‌مرجع آن نشان داده شده است. این دیتاست بنابراین بررسی ویژگی‌های بافت تصویر تولید شده است و از این رو یک تصویر تک باند است (بدون ویژگی‌های طیفی). در هر دو آزمایش نمونه‌های آموزشی، تصادفی و یکسان و برابر ۲ درصد کل نمونه‌های برچسب خورده می‌باشد. در جدول ۳-۲ اطلاعات فیلتربانک‌ها قرار دارد. همچنین نتایج طبقه‌بند ML در جدول ۳-۳ و جدول ۴-۳ نمایش داده شده است.



شکل ۳-۶: داده اول برای آزمایش طبقه‌بندی

جدول ۳-۳: اطلاعات فیلتربانک‌های مورد آزمایش برای مقایسه آن‌ها در طبقه‌بندی دیتاست اول

	فیلتربانک گابور برش خورده	فیلتربانک گابور با اندازه کامل
طول موج بر حسب پیکسل	$\{3, 5, 9, 15, 31, 63\}$ pixel	$\{3, 5, 9, 15, 31, 63, 129\}$ pixel
اندازه برش	19 pixel	unCut
زاویه‌ها	$\{0, 30, 60, 120, 150\}^\circ$	$\{0, 30, 60, 120, 150\}^\circ$
تعداد کل ویژگی‌ها	$6 \times 6 = 36$	$6 \times 7 = 42$

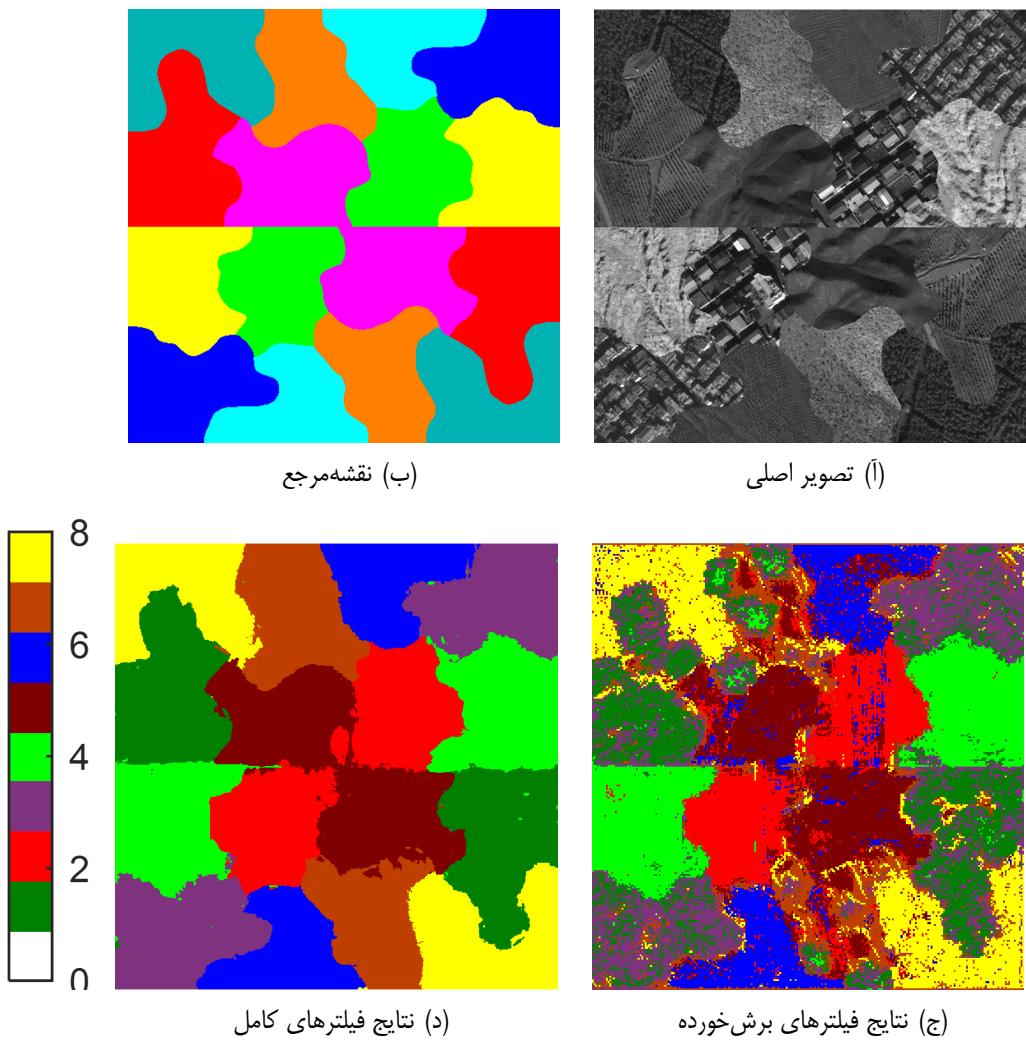
جدول ۳-۳: نتایج طبقه‌بند ML با فیلتربانک برش‌خورده

Overall Accuracy=68 ,Average Accuracy=68, Average Validity=68, $\kappa=63$									
Class	1	2	3	4	5	6	7	8	Acc
1	24276	16	9349	651	575	13	2041	1433	63
2	412	23588	567	1706	2227	1167	1463	357	75
3	9359	0	20967	460	30	97	1904	641	63
4	1118	96	534	31216	0	0	66	0	95
5	1033	2093	1017	171	24679	1785	1979	868	73
6	405	2380	1327	29	3864	16709	1783	811	61
7	3637	1526	3016	1379	7808	614	10248	2314	34
8	1891	538	1371	9	756	791	2624	26360	77
Val	58	78	55	88	62	79	46	80	

جدول ۳-۴: نتایج طبقه‌بند ML با فیلتربانک کامل

Overall Accuracy=94 , Average Accuracy=94 , Average Validity=94 , $\kappa=93$									
Class	1	2	3	4	5	6	7	8	Acc
1	36934	0	1	679	158	0	0	582	96
2	62	27926	483	1276	1473	56	211	0	89
3	0	0	32434	962	0	62	0	0	97
4	24	46	526	32434	0	0	0	0	98
5	903	765	8	346	30611	0	746	246	91
6	0	126	1584	35	0	24564	999	0	90
7	20	149	6	4	438	170	29303	452	96
8	1380	0	0	34	16	0	1009	31901	93
Val	94	96	93	91	94	99	91	96	

با دقت به جداول ۳-۳ و ۳-۴ مشاهده می‌کنیم که دقت طبقه‌بندی با فیلتربانک کامل بسیار بیشتر از حالت برش‌خورده است. اما نباید فراموش کنیم که این داده برای تحلیل و طبقه‌بندی ویژگی‌های بافت تولید شده است به همین دلیل با فیلترهای بزرگ ناحیه وسیع‌تری را می‌توان تحلیل کرد و از همین رو دقت با این روش بیشتر است. ولی اگر قرار باشد طبقه‌بندی بر روی اشیاء مختلف صورت گیرد این فیلترها روش مناسبی نخواهند بود. برای مثال اگر قرار باشد در یک بافت شهری خیابان، ساختمان، فضای سبز و ... را طبقه‌بندی کنیم فیلترهای گابور کامل روش ایده‌الی برای آن نیستند اما در جایی که قرار است طبقه‌بندی به شکل ناحیه‌ای صورت گیرد مثلاً ناحیه شهری (شامل اشیایی مانند ساختمان، خیابان و ...) و ناحیه کشاورزی (شامل گیاهان کشت شده متفاوت) و ... فیلتربانک گابور با آنالیز دقیق بافت بهترین نتیجه را خواهد داشت. در شکل ۷-۳ نتیجه تصویری آزمایشات بالا آورده شده است که گفته‌های بالا در مورد آن‌ها صدق می‌کند.



شکل ۷-۳: نتایج طبقه‌بندی با طبقه‌بند ML برای داده اول

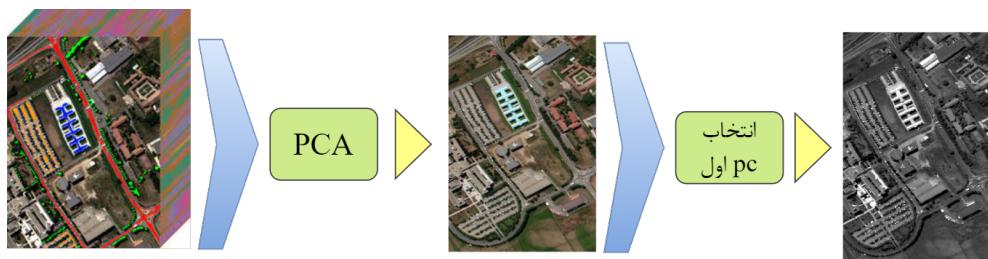
۲-۳-۳ آزمایش طبقه‌بندی با دیتاست دوم

به طور کلی اگر در یک تصویر قرار باشد طبقه‌بندی بر روی اشیاء صورت گیرد، فیلترهای برش‌خورده نتیجه قابل اطمینان‌تری نسبت به فیلتر کامل به ما ارائه می‌دهند، زیرا اطلاعات اشیاء دورتر، کمتر بر پیکسل مرکزی اعمال می‌شود. در این آزمایش می‌خواهیم این گفته را نشان دهیم برای این که نتایج و توضیحات ما قابل اطمینان باشد می‌بایست از داده‌ای استفاده کنیم که نقشه‌مرجع آن بر اساس اشیاء موجود در تصویر باشد و نه یک ناحیه بزرگ و خاص (برعکس داده اول که در آزمایش قبل از آن استفاده کردیم). از همین رو سراغ داده‌های ابرطیفی می‌رویم این داده‌ها عموماً با تکیه بر بردار طیفی دارای نقشه‌مرجعی قابل اطمینان برای جداسازی اشیاء مختلف هستند. یکی از داده‌های ابرطیفی معروف که با وجود دقت طیفی بالا از دقت مکانی کافی نیز برخوردار است داده دانشگاه پاویا است. چون این داده ابرطیفی است و قصد داریم فقط از یک تصویر PAN استفاده کنیم، از

جدول ۳-۵: اطلاعات فیلتربانک‌های مورد آزمایش برای مقایسه آن‌ها در طبقه‌بندی دیتاست دوم

	فیلتربانک گابور با اندازه کامل	
طول موج بحسب پیکسل	{3, 5, 9, 15, 31} pixel	{3, 5, 9, 15, 31, 63, 129 }pixel
اندازه برش	11 pixel	unCut
زاویه ها	{0, 30, 60, 120, 150}°	{0, 30, 60, 120, 150}°
تعداد کل ویژگی‌ها	$6 \times 5 = 30$	$6 \times 7 = 42$

تصویر ابرطیفی دانشگاه پاویا تبدیل PCA اخذ می‌کیم و سپس از PC1 به عنوان تصویر تک باند خود استفاده می‌کنیم.

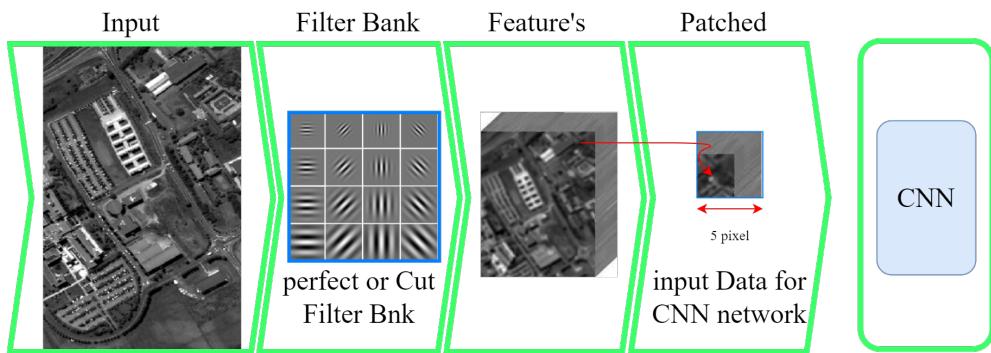


شکل ۳-۸: اصلاح و پیش‌پردازش تصویر ابرطیفی دانشگاه پاویا برای انجام آزمایش‌ها. [۴۲]

پس از انجام اصلاحات لازم بر روی این تصویر قصد داریم آزمایش قبلی را بر روی این تصویر پیاده کنیم. ولی قبل از آن باید با طبقه‌بند مورد آزمایش در این قسمت آشنا شویم.

طبقه‌بند مورد استفاده در این آزمایش یک شبکه CNN دو بعدی است. روند انجام آزمایش به این شکل است که داده اصلاح شده را با یکی از روش‌های مورد نظر خود، مورد استخراج ویژگی قرار می‌دهیم. سپس به یک تصویر سه بعدی می‌رسیم که در بعد سوم ویژگی‌های استخراج شده از تصویر قرار دارد و اندازه بعد سوم برابر تعداد فیلترهای موجود در فیلتر بانک است. تا به اینجا روند کار مشابه قبل است. شبکه مورد استفاده چنان که گفته شد یک شبکه CNN دو بعدی است، پس تصویر ورودی به این شبکه باید هم دارای ارتفاع و هم عرض و هم عمق باشد. تصویر ورودی برای این شبکه تصویر فیلتر شده است که عمق آن برابر تعداد فیلترهای موجود در فیلتر بانک است. نکته‌ای که وجود دارد این است که ما تصویر را به صورت پنجره‌هایی به ارتفاع و عرض ۵ پیکسل تبدیل می‌کنیم. هدف ما تعیین کلاس پیکسل مرکزی هر کدام از پنجره‌ها است. به عبارتی دیگر برای هر پیکسل یک پنجره 5×5 تولید می‌کنیم و آن پنجره را متعلق به کلاسی می‌دانیم که مربوط به پیکسل مرکزی آن است. سپس با ۱۰ درصد از دیتا شبکه عصبی را تعلیم می‌دهیم و با سایر داده‌ها نتایج را بررسی خواهیم کرد. روند کار مطابق شکل ۳-۹ می‌باشد و اطلاعات فیلتر بانک‌ها در جدول ۳-۵ موجود است.

فصل سوم: عملکرد طبقه‌بندها و ویژگی‌های متفاوت بر روی دیتاست‌های مختلف



شکل ۳-۹: روند نمای انجام آزمایش دوم

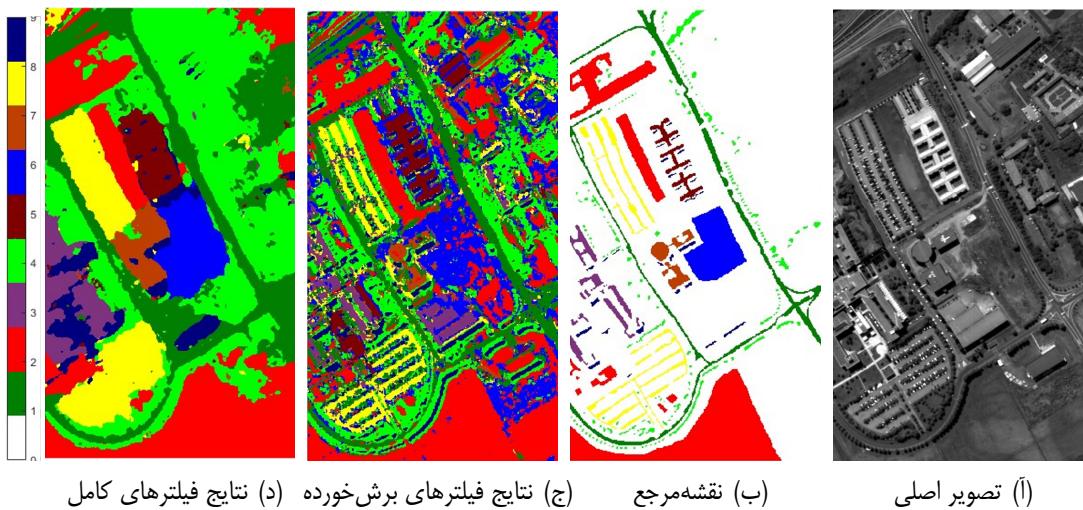
جدول ۳-۶: نتایج طبقه‌بند CNN با فیلتربانک کامل

Overall Accuracy=99 , Average Accuracy=98 , Average Validity=98 , $\kappa=98$										Acc
Class	1	2	3	4	5	6	7	8	9	Acc
1	6619	0	0	12	0	0	0	0	0	100
2	0	18649	0	0	0	0	0	0	0	100
3	0	0	2084	0	0	0	0	0	15	99
4	14	26	3	2947	0	0	6	37	31	96
5	0	0	0	0	1338	0	0	0	7	99
6	0	0	0	0	0	5029	0	0	0	100
7	0	0	0	0	0	0	1330	0	0	100
8	0	0	0	27	0	0	0	3655	0	100
9	0	0	4	2	99	0	5	10	827	87
Val	100	100	100	99	93	100	99	99	94	

جدول ۳-۷: نتایج طبقه‌بند CNN با فیلتربانک برش خورده

Overall Accuracy=95 , Average Accuracy=93 , Average Validity=94 , $\kappa=93$										Acc
Class	1	2	3	4	5	6	7	8	9	Acc
1	6270	96	17	82	0	96	39	31	0	95
2	146	18195	7	48	0	248	5	0	0	98
3	37	22	1863	2	0	46	40	89	0	89
4	22	50	0	2921	0	58	2	8	3	95
5	2	1	0	0	1337	5	0	0	0	99
6	49	420	46	18	29	4362	61	42	2	87
7	88	89	8	12	0	13	1103	17	0	83
8	49	1	149	5	0	12	11	3455	0	94
9	0	0	0	0	0	0	0	0	947	100
Val	94	96	89	95	98	90	87	95	99	

اکنون به جدول‌های ۳-۶ و ۳-۷ دقت کنید و نتایج طبقه‌بندی را مشاهده کنید. با توجه به جدول‌ها، فیلتربانک کامل نتیجه بهتری به همراه داشته است. اکنون به شکل نقشه‌های تولید شده دقت می‌کنیم. در شکل ۳-۱۰ نتایج تصویر برچسب خورده را مشاهده می‌کنیم.



شکل ۱۰-۳: نتایج طبقه‌بندی با طبقه‌بندی باز ML برای داده اول

با کمی دقیق به جداول ۳-۶ و ۳-۷ مشاهده می‌کنیم که دقیق طبقه‌بندی در حالت استفاده از فیلترهای کامل بیشتر است. ولی با دقیق به شکل ۱۰-۳(ج) مشاهده می‌کنیم طبقه‌بندی به شکلی انجام شده است که فقط دقیق طبقه‌بندی افزایش یابد، بدون این که طبقه‌بندی واقع گرایانه‌ای نسبت به صحنه داشته باشیم در جدول ۳-۸-۳ چند نمونه از این موارد را نشان داده‌ایم. به طور کلی چون این یک طبقه‌بندی طیفی نیست اما دقیق طبقه‌بندی فیلترهای برش‌خورده قابل رقابت با برخی روش‌های مبتنی بر طیف است. باز در اینجا متذکر می‌شویم هدف ما بهبود دقیق طبقه‌بندی نیست و صرفاً هدف ما بررسی عملکرد دو نوع فیلتربانک یاد شده است.

همانطور که گفتیم برای طبقه‌بندی داده دانشگاه پاویا از شبکه CNN استفاده کردہ‌ایم. به طور کلی یکی از مزیت‌های شبکه‌های عصبی طبقه‌بندی داده‌های است که برای یک کلاس خاص، بخش‌های مختلفی از فضای ویژگی را اشغال کرده باشند، یا توزیع خاصی در فضای ویژگی نداشته باشند. در این صورت یک شبکه عصبی با انتخاب مناسب پارامترهای قابل تنظیم شبکه، می‌تواند این داده‌ها را طبقه‌بندی کند ولی برای انجام این کار نیاز به داده آموزشی زیادی دارد تا بتواند به طور کامل مرز بین کلاس‌ها را استخراج کند. از این رو ما برای این آزمایش از شبکه عصبی استفاده کردیم و ۱۰ درصد از دیتای برچسب دار را برای آموزش و سایر داده‌ها را برای تست انتخاب کردیم.

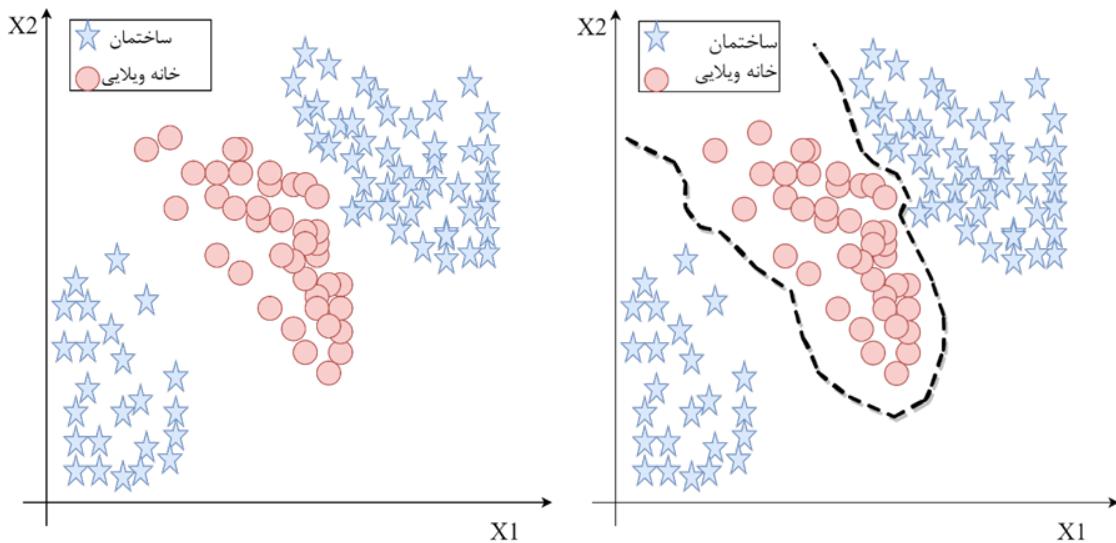
در واقع وقتی می‌خواهیم با استفاده از ویژگی‌های مکانی یک تصویر را طبقه‌بندی کنیم و نوع طبقه‌بندی هم مبتنی بر شئ باشد، داده‌های یک کلاس خاص فاصله زیادی از هم خواهند داشت. مثلاً یک ساختمان ممکن است یک بافت متفاوت از ساختمان دیگر داشته باشد، در نتیجه ویژگی‌های استخراج شده برای این دو ساختمان فاصله زیادی از هم خواهند داشت و ممکن است در بین این فاصله یک کلاس دیگر نیز قرار بگیرد

جدول ۸-۳: نمونه‌ای از ایرادات طبقه‌بندی با فیلترهای کامل

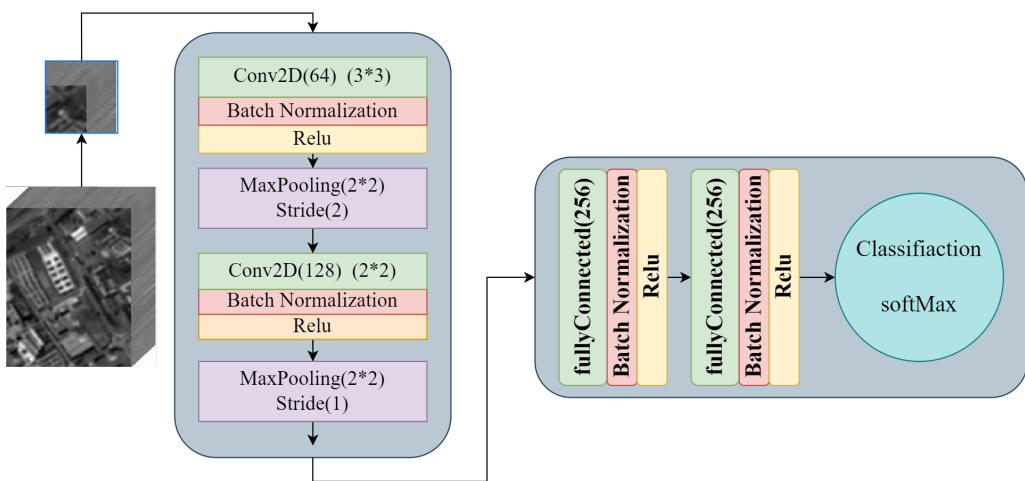
توضیحات	خروجی طبقه‌بندی با فیلتربانک برش خورده (۲)	خروجی طبقه‌بندی با فیلتربانک کامل (۱)	تصویر pc1 (خاکستری) و تصویر [pc1,pc2,pc3] به صورت رنگ مصنوعی	نمره
سطح پوشیده از آجر (زرد): در (۱) حتی درختان بین آجرها زردنگ شده‌اند ولی در (۲) سطح آجری مستقل از سایر اشیاء برچسب خورده است.				۱
فلز رنگ شده (قمه‌هایی): این شیء به عنوان سقف بعضی از ساختمان‌ها به کار رفته است. ساختمان‌های سمت چپ در (۱) به طور کامل با اطرافشان برچسب قمه‌هایی خورده‌اند. ولی در (۲) فقط سقف برچسب خورده است همچنین ساختمان سمت راست بالا فقط در (۲) شناسایی شده و در (۱) هیچ نشانی از رنگ قمه‌هایی در این ناحیه نیست.				۲
سایه‌ها (سرمه‌ای یا آبی خیلی پررنگ): در (۱) فقط جایی که نقشه‌مرجع سایه را نشان داده برچسب سایه خورده است آن هم بسیار پهن‌تر از سایه واقعی. اما در (۲) تقریباً تمام سایه‌ها به درستی برچسب خورده‌اند و مخصوصاً اندازه آن‌ها زیاد پهن‌تر است.				۳

که مشکل طبقه‌بندی را افزایش می‌دهد. به مثال شکل ۱۱-۳ که برای طبقه‌بندی یک داده دو بعدی (داده‌ای با دو ویژگی) آورده شده است دقت کنید.

در شکل ۱۱-۳ متوجه می‌شویم که طبقه‌بند ML نمی‌تواند این داده را به خوبی طبقه‌بندی کند. سایر روش‌های کلاسیک یادگیری ماشین در این دیتا دقت کافی نخواهد داشت اما شبکه عصبی با مصرف زیاد دیتای آموزشی توانایی دست یابی به طبقه‌بندی با دقت بالایی برای این داده را خواهد داشت. در شکل ۱۲-۳ و جدول ۹-۳ نمای کلی شبکه استفاده شده و اطلاعات آموزش این شبکه را نمایش داده‌ایم.



شکل ۱۱-۳: مثالی از توانایی شبکه‌های عصبی با تعداد لایه و نورون کافی در جداسازی کلاس‌هایی که پراکندگی زیادی دارند.



شکل ۱۲-۳: نمای کلی شبکه عصبی استفاده شده در آزمایش دوم طبقه‌بندی

۳-۴ معرفی داده مناسب با نقشه مرجعی مناسب برای اشیاء

در مقدمه مشخصات تصویرهای چندطیفی ماهواره QB2 و WV3 را نشان دادیم و در مورد دقت مکانی بالای این تصاویر صحبت‌هایی را ارائه نمودیم. اکنون هدف ما استفاده از این تصاویر برای طبقه‌بندی با اطلاعات مکانی است. متأسفانه هیچ نقشه مرجع قابل اطمینانی از این تصاویر برای طبقه‌بندی بر اساس اشیاء (مانند تصاویر ابرطیفی) در دسترس نیست. از همین رو یکی از فعالیت‌هایی که در طول انجام روند پایان‌نامه مد نظر ما بوده است تولید یک نقشه مرجع مناسب برای یک تصویر از این ماهواره بوده است. اهمیت آزمایش اول

جدول ۹-۳: اطلاعات شبکه کانوالوشنی استفاده شده

TrainPercent	Validation	Batch size	epochs	Learning rate	Training options
8%	2%	64	35	$0.005 = > 10^{-6}$	"SGDM"

که در مورد خوشبندی انجام شده در این جا مشخص می‌شود. در واقع ما یک نقشه مرجع با دقت کافی را با استفاده از خوشبندی و اصلاح آن به صورت دستی تولید نموده‌ایم که در ادامه روند تولید این نقشه مرجع را توضیح و نمایش داده‌ایم.

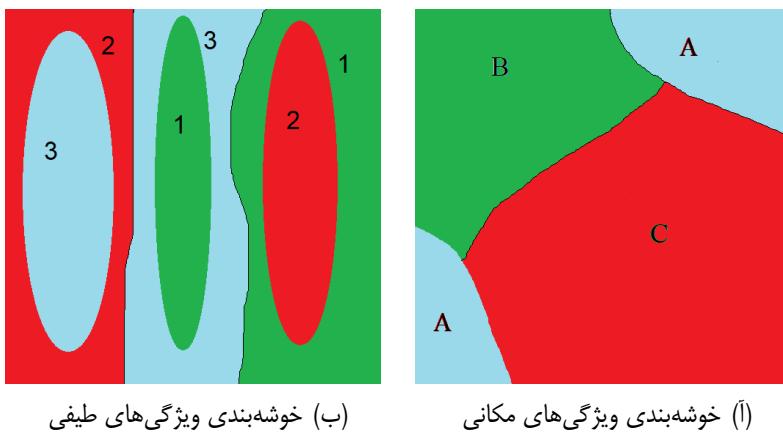
ابتدا الگوریتم مورد استفاده خود را توضیح می‌دهیم. نقطه شروع این الگوریتم خوشبندی یک تصویر چندطیفی است. به این صورت که ما یکبار ویژگی‌های مکانی را با تکیه بر نتایج بخش ۲-۳ خوشبندی می‌کنیم و یکبار ویژگی‌های طیفی را نیز به طور مستقل مورد خوشبندی قرار می‌دهیم (در هر دو حالت از الگوریتم K-means استفاده شده است). نکته قابل توجه این است که ما تصویر چندطیفی را به اندازه تصویر تک‌باند بزرگ کرده‌ایم. تا به اینجا گام اول انجام شده است و ما دو نوع نتیجه خوشبندی متفاوت داریم که یکی با ویژگی‌های طیفی و یکی با ویژگی‌های مکانی به دست آمده است. اکنون در بخش دوم می‌بایست، گروه‌های متفاوت خوشبندی شده را هم‌برچسب کنیم. یعنی اگر برای یک ناحیه در تصویر اول، شماره ۲ قرار داده شده است باید در تصویر دوم نیز در آن ناحیه شماره ۲ قرار بگیرد. نکته‌ای که وجود دارد این است که این فرایند بسیار پیچیده است.

به طور خلاصه ما هم‌برچسب کردن را به گونه‌ای انجام داده‌ایم که بیشترین تعداد پیکسل هم‌برچسب بشوند. یک شرایط فرضی را تصور می‌کیم: داده‌ها را به ۳ خوشبندی تقسیم کرده‌ایم، برچسب خروجی ویژگی‌های طیفی ۱، ۲ و ۳ است و برچسب خوشبندی‌های مکانی A، B و C است و نقشه‌های خروجی مطابق شکل ۱۳-۳ می‌باشد و تعداد کل نقاط برابر ۲۵۰۰ است (یک تصویر با ابعاد 50×50) و همچنین تعداد نمونه‌های قرار گرفته در هر خوشبندی مطابق جدول ۱۰-۳ می‌باشد.

جدول ۱۰-۳: تعداد نقاط هر خوشبندی در هر کدام از نتایج

تعداد نمونه‌های خوشبندی ویژگی‌های مکانی	تعداد نمونه‌های خوشبندی ویژگی‌های طیفی
کلاس ۱: ۹۰۰	۵۰۰: کلاس A
کلاس ۲: ۸۰۰	۱۱۰۰: کلاس B
کلاس ۳: ۸۰۰	۹۰۰: کلاس C

اکنون می‌خواهیم برچسب‌های A، B و C را به گونه‌ای تغییر دهیم که پس از تغییر این برچسب‌ها به برچسب‌های ۱، ۲ و ۳ بیشترین تعداد نقاط ممکن، در هر دو نتیجه هم‌برچسب شوند. فرض کنید اگر برچسب B به برچسب



شکل ۱۳-۳: نتایج خوشبندی فرضی

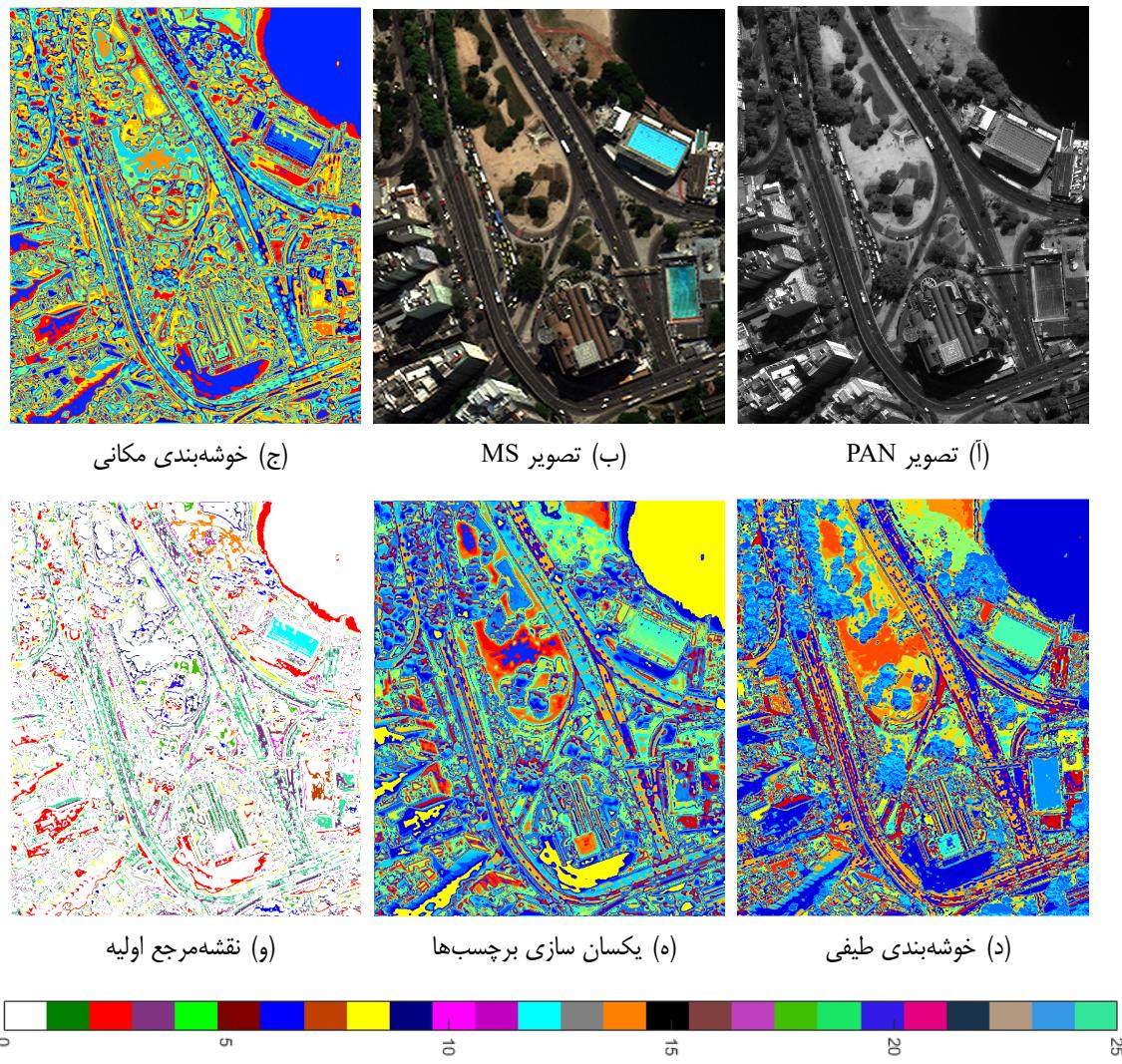
۱ تغییر کند ۵۰ نمونه هم در تصویر یک و هم در تصویر دو برچسب ۱ خواهند داشت و اگر B به ۳ تغییر کند ۲۰۰ نقطه هم در تصویر یک و هم در تصویر دو برچسب ۳ خواهند داشت.

به همین ترتیب اگر هر کدام از برچسب‌های A، B و C به یکی از برچسب‌های ۱، ۲ و ۳ تغییر کند، تعداد نقاطی که در هر دو تصویر، هم‌برچسب می‌شوند متفاوت خواهد. الگوریتم پیاده شده برای هم‌برچسب کردن دو نقشه به گونه‌ای طراحی شده است که در پایان فرایند، بیشترین تعداد نقاط ممکن در هر دو تصویر هم‌برچسب باشند. یک پیششرط نیز وجود دارد.

تصویر کنید در شکل ۱۳-۳(ب) ۹۰ درصد از داده‌ها برچسب ۱ داشته باشند. اکنون اگر هر سه کلاس A، B و C به برچسب ۱ تغییر کنند ۹۰ درصد از هر دو تصویر هم‌برچسب می‌شوند، ولی همه نقاط هم‌چسب، برچسب ۱ دارند. متوجه دیگری که وجود دارد این است که ما هم‌چسب کردن را به گونه‌ای انجام دهیم که در ازای هر برچسب از ۳(ب) یک برچسب از ۳(آ) را به آن تعلق دهیم و این کار به گونه‌ای انجام شود که در نهایت علاوه بر حفظ این شرط بیشترین نقاط هم‌برچسب نیز در هر دو تصویر موجود باشد.

بعد از هم‌برچسب کردن تصویر خوشبندی مکانی به تصویر خوشبندی طیفی، باید نقاطی که در هر دو تصویر یک برچسب دارند را به عنوان نقشه مرجع اولیه در نظر بگیریم. دیتای مورد استفاده همان دیتای استفاده شده در ۲-۳ است و در شکل ۱۴-۳ نتایج به دست آمده تا به این قسمت از الگوریتم نمایش داده شده است. ضمناً لازم است بدانیم الگوریتم خوشبندی مورد استفاده ما الگوریتم K-means و تعداد کلاس‌های خروجی به طور کلی در خروجی الگوریتم خوشبند برابر ۲۵ است.

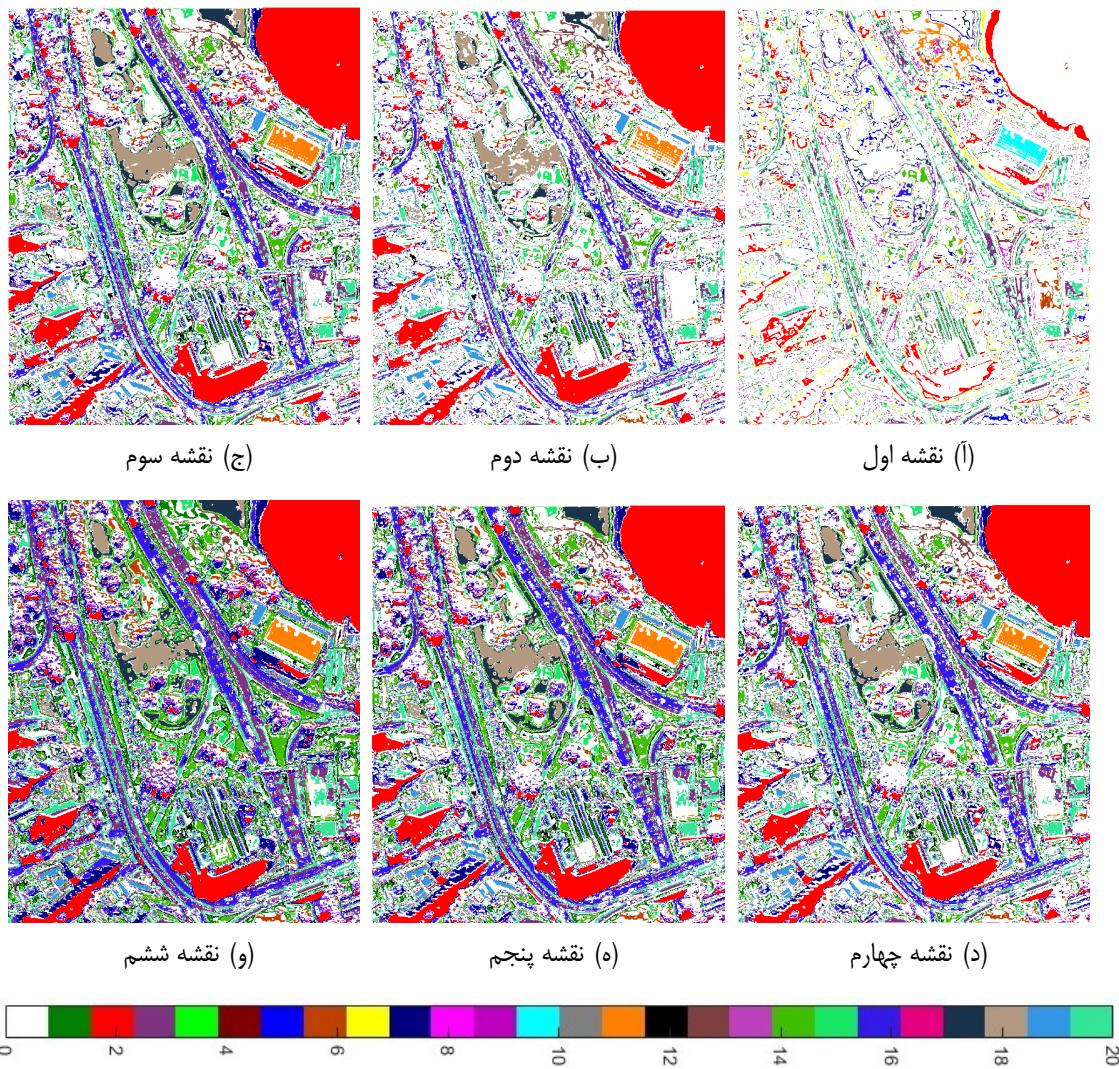
در تصویر ۱۴-۳(۵) به این نکته دقت داشته باشید که ما روشی را برای هم‌برچسب کردن نقشه‌ها انتخاب



شکل ۱۴-۳: بخش اول نحوه ساخت نقشه مرجع با الگوریتم مورد نظر

کردیم که بیشترین کلاس متفاوت به دست آید (همانطور که در بالاتر توضیح دادیم). سپس در تصویر ۱۴-۳(و) نقاطی که هم در تصویر ۱۴-۳(د) و هم در تصویر ۱۴-۳(ه) دارای برچسب یکسان هستند نشان داده شده‌اند و این تصویر به عنوان نقشه مرجع اولیه ما انتخاب می‌شود. دقت کنید در این مرحله برخی از کلاس‌ها ممکن است حذف شده باشد یعنی ممکن است در دو تصویر قبل برخی کلاس‌ها هیچ همپوشانی نداشته باشند و ما شماره‌گذاری را مجدداً از ۰ تا تعداد کلاس‌های باقیمانده انجام داده‌ایم. همچنین ۰ به معنی آن است که آن نقطه برچسب ندارد و با رنگ سفید مشخص شده است. از این رو ممکن است رنگ بعضی از نقاط تصویر ۱۴-۳(و) با دو تصویر دیگر متفاوت باشد. پس در طی این فرایند ممکن است برخی از ۲۵ کلاس به دلیل عدم همپوشانی حذف شده باشند.

بعد از انجام بخش اول الگوریتم، اکنون با همان نقشه مرجع به دست آمده تصویر را طبقه‌بندی می‌کنیم.

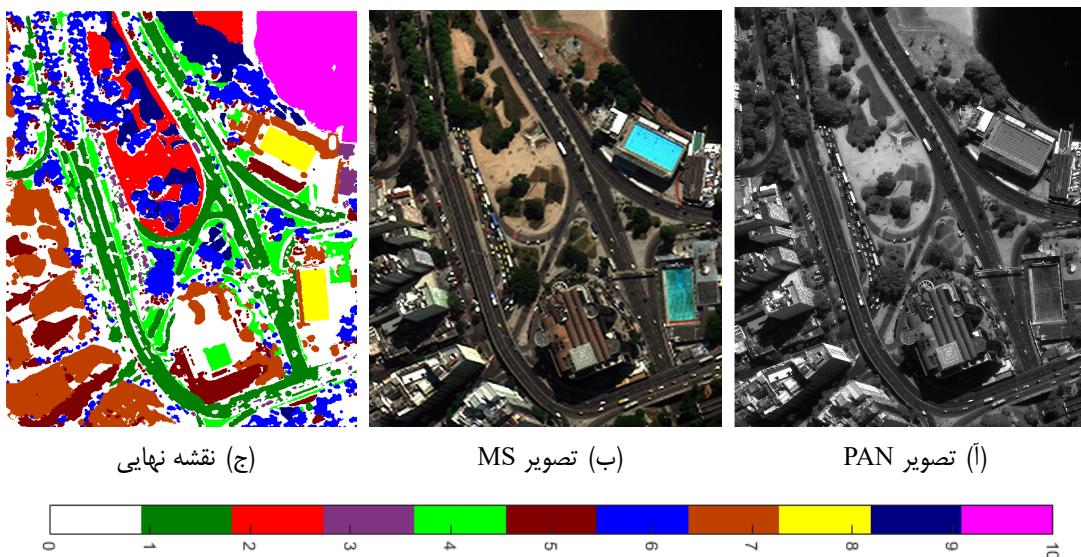


شکل ۱۵-۳: بخش دوم نحوه ساخت نقشه مرجع با الگوریتم مورد نظر

ابزارهای طبقه‌بندی بسیار زیادند و ما پس از بررسی چندین مورد از این ابزارها یک شبکه عصبی MLP ساده با سه لایه مخفی را برای این کار انتخاب نمودیم. حالا یکبار با استفاده از ویژگی‌های طیفی و یکبار با استفاده از ویژگی‌های مکانی تصویر را با نقشه مرجع مرحله قبل طبقه‌بندی می‌کنیم. بعد از این دو خروجی را در اختیار داریم و سپس نقاط هم برچسب را نگاه می‌داریم و نقاطی که برچسب یکسان ندارند صفر قرار می‌دهیم. اکنون نقشه مرجعی داریم که می‌تواند به عنوان نقشه مرجع جدید ما انتخاب شود. حالا همین فرایند را تا چند بار تکرار می‌کنیم.

در شکل ۱۵-۳ این فرایند را از مرحله ۱ تا مرحله ۶ نشان داده‌ایم. ممکن است در هر مرحله چند کلاس از بین بروند و تعداد کلاس‌ها کمتر شود که هر بار این اتفاق صورت گیرد برچسب گذاری مجدداً از صفر تا تعداد کلاس باقی‌مانده انجام می‌شود.

در نقشه ۶ از ۲۵ کلاس اولیه ۲۰ کلاس باقی‌مانده است اکنون با دخالت و به طور دستی برخی کلاس هارا ادغام و برخی را جدا می‌کنیم. در طی این فرایند از نرم‌افزار فتوشاپ استفاده نموده‌ایم سپس با استفاده از برنامه Image Segmenter در نرم‌افزار متلب، تصاویر هر کدام از کلاس‌ها را با استفاده از فیلترهای ریخت شناسی و عملیات‌های باز کردن و بستن و سایر ابزارهای موجود در این برنامه همزمان با مقایسه با تصویر اصلی به طور چشمی بهینه می‌نماییم. در نهایت ما برای این تصویر به نقشه مرجع شکل ۱۶-۳(ج) می‌رسیم.



شکل ۱۶-۳: نقشه نهایی به دست آمده از الگوریتم مورد نظر به همراه تصاویر اصلی

چنان که در شکل ۱۶-۳(ج) مشخص است ما در نهایت به ده کلاس دست یافته‌ایم که دقت برچسب‌های زده شده نیز به اندازه کافی بالا است. هدف اصلی ما در فصل ۴ این است که طبقه‌بندی را معرفی کنیم که این تصویر را با دقیقی بالا طبقه‌بندی کند. در ادامه این فصل با چند روش مختلف طبقه‌بندی طیفی و مکانی سعی می‌کنیم این تصویر را طبقه‌بندی کنیم و نتایج هر کدام را نشان دهیم. اما قبل از شروع در جدول ۱۱-۳ مشخصات این داده و نقشه مرجع آن را مشاهده می‌کنیم.

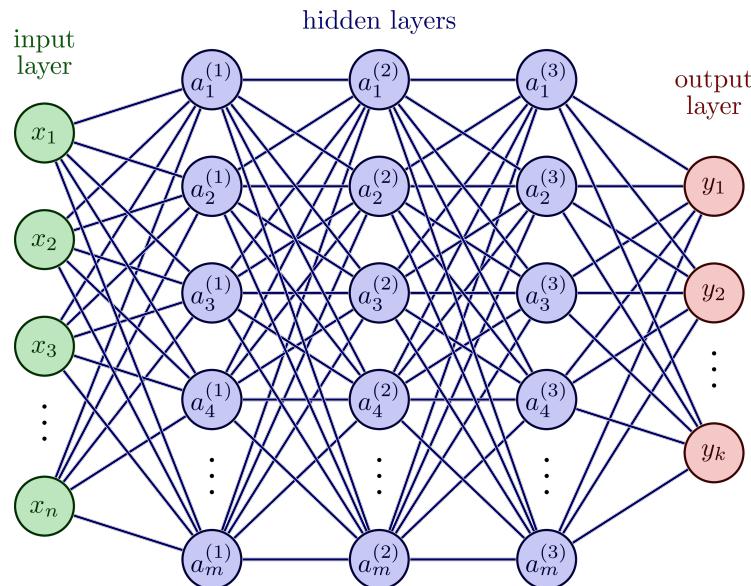
جدول ۱۱-۳: مشخصات نقشه مرجع تولید شده

شماره	نام کلاس	شماره	نام کلاس	تعداد نمونه	تعداد نمونه	شماره
۱	خیابان	۶	درخت	۱۴۹۳۶۱	۱۸۸۴۸۶	
۲	خاک و ماسه	۷	ساختمان	۱۷۰۹۴۳	۷۳۵۹۷	
۳	ماشین غیر مشکی	۸	استخر	۲۵۶۸۵	۱۶۹۶۱	
۴	سنگ فرش	۹	چمن	۶۸۰۴۶	۸۵۳۸۰	
۵	سایه	۱۰	دریا	۱۰۰۳۴۱	۹۷۶۸۸	
اندازه تصویر			۱۱۰۰ × ۱۳۰۱			

در ادامه و در جدول ۱۲-۳ مشخصات شبکه MLP استفاده شده در تولید این نقشه مرجع را نشان داده‌ایم. دقت شود چون هدف ما تولید نقشه مرجع بوده است درصد داده آموزشی بالا است و شبکه مورد استفاده نیز یک شبکه ساده است تا فرایند سریع‌تر انجام شود. همچنین گفتیم برای خوشبندی اولیه و طبقه‌بندی‌های پشت سر هم، یکبار از ویژگی‌های طیفی و یکبار از ویژگی‌های مکانی استفاده کردی‌ایم. ویژگی‌های طیفی همان بردار طیفی هر پیکسل است که در این تصویر برابر ۸ ویژگی است که در جدول ۱-۱ مشخصات این باندها درج شده است. برای استخراج ویژگی‌های مکانی نیز از تصویر تک‌باند استفاده کردی‌ایم که فیلتربانک مورد استفاده ما یک فیلتربانک گابور برش‌خورده بوده است که در جدول ۵-۳ (و در ستون مربوط به فیلتربانک برش‌خورده) اطلاعات فیلتربانک نمایش داده شده است. در شکل ۱۷-۳ نیز شمای کلی شبکه عصبی مشخص است.

جدول ۱۲-۳: اطلاعات شبکه MLP استفاده شده در تولید نقشه مرجع

TrainPercent	Validation	Batch size	epochs	Learning rate	Training options
50%	0%	64	10	$0.01 \Rightarrow 10^{-6}$	"SGDM"



شکل ۱۷-۳: نمای کلی شبکه عصبی استفاده شده در تولید نقشه مرجع

در شکل ۱۷-۳ تعداد نورون‌های لایه مخفی برابر $a_m^{(1)} = 64$, $a_m^{(2)} = 128$, $a_m^{(3)} = 64$ است. لایه خروجی مقدار مشخصی در تمام مراحل ندارد زیرا که گفتیم در هر مرحله بخشی از کلاس‌ها از بین می‌رود ولی میدانیم تعداد این نورون‌ها بین ۲۰ تا ۲۵ خواهد بود. در لایه اول نیز n دو مقدار دارد یکی برای ویژگی‌های طیفی که برابر ۸ است و یکی برای ویژگی‌های مکانی که برابر ۳۰ است.

۳-۵ انجام طبقه‌بندی بر روی دیتاست جدید

در این بخش سعی می‌کنیم تصویر حاصل از بخش ۴-۳ را با چند روش مختلف طبقه‌بندی کنیم تا با رفتار این دیتا بیشتر آشنا شویم. در واقع چون هر طبقه‌بند برای یکسری داده خاص طراحی شده است (منظور نوع توزیع داده‌ها در فضای ویژگی است) تلاش می‌کنیم با طبقه‌بندی کردن آن دید کامل تری نسبت به داده به دست آوریم. قبل از بررسی نتایج ذکر این نکته لازم است که در تمام آزمایش‌های صورت گرفته دیتای آموزشی یکسان است و تعداد آن‌ها برابر ۳ درصد از کل داده برچسب خورده است.

فرم داده‌ها برای طبقه‌بندی به این صورت است که ابتدا ویژگی‌های مکانی را از تصویر تکباند با استفاده از فیلترانک گابور برش خورده که در جدول ۵-۳ (و در ستون مربوط به فیلترانک برش خورده) اطلاعات آن نمایش داده شده است استخراج می‌کنیم. پس تا به اینجا برای هر پیکسل ۳۰ ویژگی مکانی استخراج کردہ‌ایم. همچنین با تغییر اندازه تصویر چندطیفی به اندازه تصویر تکباند، برای هر پیکسل ۸ ویژگی طیفی نیز خواهیم داشت ناگفته نماند که اندازه تصویر تکباند برابر 1100×1301 می‌باشد. در این صورت داده نهایی در ابعاد $38 \times 1100 \times 1301$ خواهد بود. با دانستن این اطلاعات نتایج را در چند بخش مشاهده خواهیم کرد. (در بعضی از بخش‌ها ممکن است فرم داده‌ها عوض شده باشد که در همان بخش به توضیحات لازم پرداخته شده است).

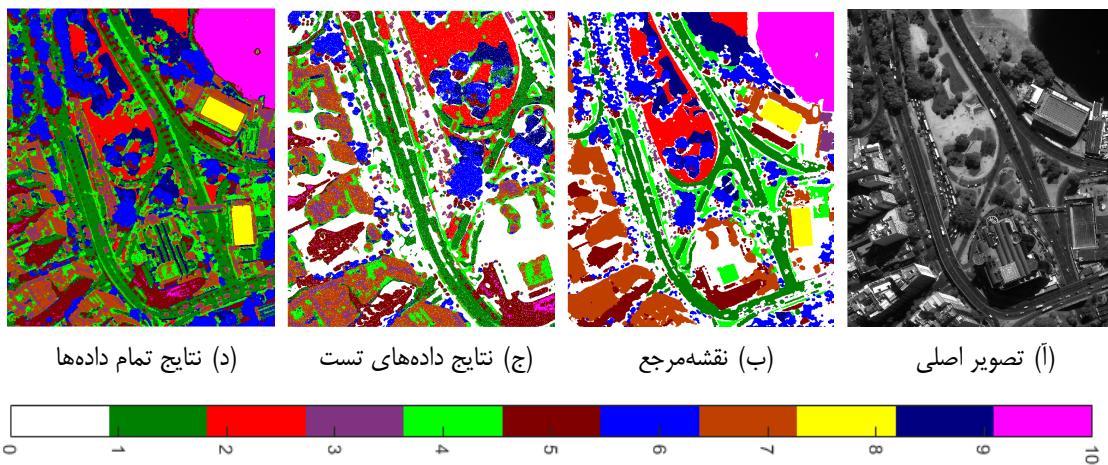
۳-۵-۱ نتایج طبقه‌بندی با ML

با توجه به اینکه طبقه‌بند ML در فضای ویژگی یکتابع گوسی با بعد فضای ویژگی تخمین می‌زند، برای داده‌های مناسب است که در فضای ویژگی یکتوزيع گوسی با حداقل همپوشانی را داشته باشند. به عبارتی هر کلاس به قدری ویژگی‌های یکسان داشته باشد که در بین آن ویژگی‌ها، ویژگی‌های کلاس دیگری قرار نگیرد و تا جای ممکن به یکتوزيع گوسی با واریانس کم نزدیک باشند. برای یادآوری، این موضوع به شکل ۱۱-۳ مجددأً دقت کنید. با درک نحوه کارکرد این طبقه‌بند می‌توان اطلاعات مهمی در باره داده خود به دست آوریم. اکنون به جدول ۱۳-۳ و شکل ۱۸-۳ دقت کنید.

با توجه به جدول ۱۳-۳ و با دانستن این نکته که از طبقه‌بند ML استفاده نموده‌ایم متوجه می‌شویم دو کلاس سنگ‌فرش و ماسه بسیار به هم نزدیک هستند همچنین تعداد زیادی از کلاس ماشین‌ها به ساختمان نسبت داده شده است که می‌توان با توجه به شکل ۱۸-۳(ج) دریافت که ماشین‌های سفید رنگ بیشتر به

جدول ۱۳-۳: نتایج طبقه‌بند ML برای داده جدید

Class	Overall Accuracy=76 , Average Accuracy=79 , Average Validity=74 , $\kappa=73$										
	1	2	3	4	5	6	7	8	9	10	Ac
1	138711	121	521	8988	11624	9130	5667	9	2140	13	78
2	7	60419	52	6290	0	16	1509	0	2361	0	85
3	0	0	12516	196	0	23	3326	0	0	0	78
4	2181	20391	2658	45990	0	95	9887	0	6514	0	52
5	5279	0	0	138	74347	4304	762	63	14	5038	83
6	3794	11	96	1177	3552	117603	3088	0	9332	1	85
7	7566	802	29596	30244	2640	3899	98675	259	2130	10	56
8	0	0	147	0	4	0	264	24235	8	0	98
9	479	2915	52	4984	7	4928	1670	0	51155	0	77
10	117	0	0	0	2192	243	95	29	0	93652	97
Val	87	71	27	46	79	84	79	98	70	95	

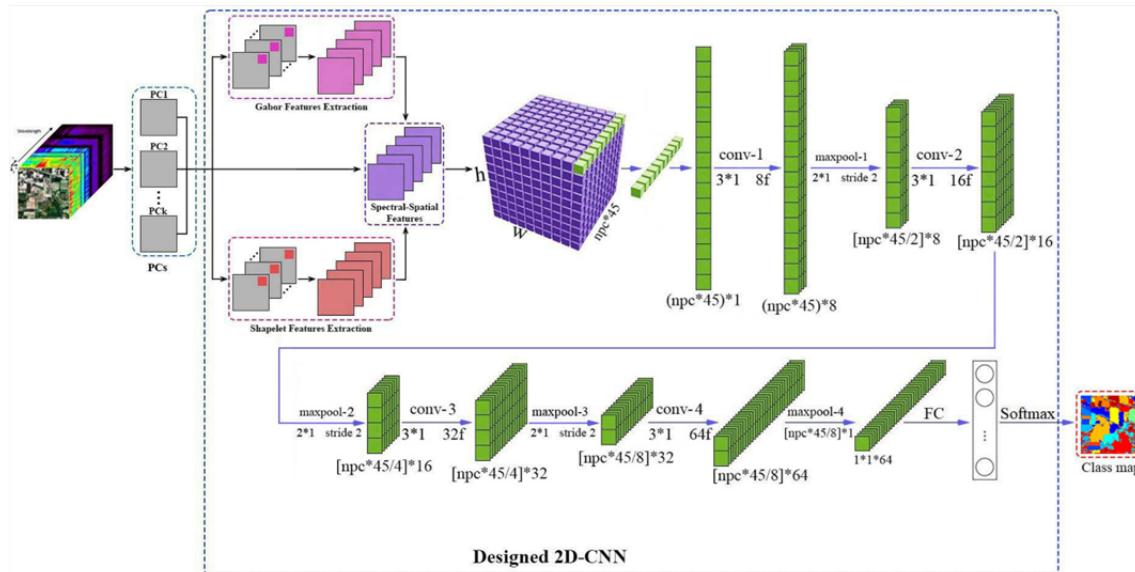


ساختمان نسبت داده شده‌اند که به خاطر رنگ مشابه آن‌ها است. در واقع کلاس ماشین پراکندگی زیادی دارد و بخشی از نمونه‌ها در لایه لای کلاس ساختمان قرار گرفته است. چمن، درخت و سنگفرش نیز نزدیکی قابل توجهی دارند که با توجه به نزدیکی سنگفرش و ماسه می‌توان گفت این ۴ کلاس برای طبقه‌بندی چالش برانگیز خواهد بود. در تصویر ۱۸-۳(ج) یکسری نقاط سفید رنگ وجود دارد که در واقع نقاط آموزشی انتخاب شده است.

۲-۵-۳ نتایج طبقه‌بندی با 1D-CNN

با استفاده از شبکه یک بعدی ارائه شده در [۴۳] می‌توان این داده را طبقه‌بندی کرد ما این روش را اندکی تغییر داده‌ایم به این صورت که داده‌ای که با روش پیشنهادی این پایان نامه مورد استخراج ویژگی قرار می‌گیرد چند

PC اول تصویر ابرطیفی است به طوری که شامل ۹۸ درصد انرژی کل باشد که معمولاً بین ۳ تا ۵ PC باقی می‌ماند. ما نیز از تصویر چندطیفی مانند این پایان‌نامه تبدیل PCA می‌گیریم که طبق روش ارائه شده ۴ PC برای ما باقی می‌ماند یعنی تصویر ما دارایی عمقی برابر ۴ است. اکنون تصویر تکباند را نیز به این داده اضافه می‌کنیم که در نتیجه عمق تصویر برابر ۵ خواهد شد. حال از این داده طبق روش ارائه شده در پایان نامه [۴۳] استفاده می‌کنیم که ابتدا با فیلتربانک گابور(به حالت کامل و نه برش خورده) و شکلینه از تصویر استخراج ویژگی صورت می‌گیرد(برای مطالعه اندازه و تعداد فیلتر بانک می‌توان به این پایان نامه رجوع کرد) و سپس با شبکه پیشنهادی طبقه‌بندی صورت می‌گیرد. در شکل ۱۹-۳ روش پیشنهادی این پایان نامه آورده شده است که همانطور که گفتیم بخش پیش‌پردازش طبق گفته‌های بالا اندکی متفاوت است. در جدول ۱۴-۳ و شکل ۲۰-۳ می‌توان نتایج استفاده از این روش را مشاهده نمود.

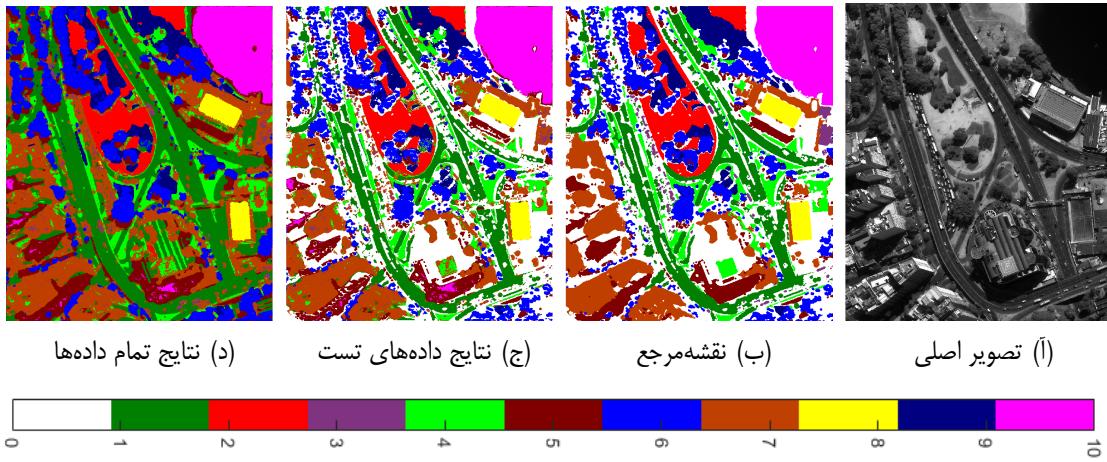


شکل ۱۹-۳: اطلاعات کلی روش پیشنهادی پایان نامه یاد شده [۴۳]

با دقت به جدول ۱۴-۳ مشاهده می‌کنیم چالش تفکیک درخت از چمن تا حد زیادی حل شده است اما مشکل ماسه و سنگفرش هنوز موجود است. ولی چون چالش تفکیک درخت از چمن حل شده است می‌توان گفت با یک روش صحیح استخراج ویژگی و طبقه‌بندی این مشکل هم رفع خواهد شد. مشکل بزرگ دیگر کلاس ساختمان و ماشین است که با بررسی نتایج بخش ۳-۵-۳ برای حل آن راهکاری ارائه خواهیم کرد. در پایان ذکر این نکته لازم است که در این پایان نامه محاسبه جدول صحت و اعتبار با تمام نقاط دارای برچسب انجام شده است (یعنی نقاط آموزشی هم در نتایج جدول و تصویر ۲۰-۳(ج) موجود است).

جدول ۳-۱۴: نتایج طبقه‌بند 1D-CNN برای داده جدید

Class	Overall Accuracy=87 , Average Accuracy=83 , Average Validity=87 , $\kappa=85$										
	1	2	3	4	5	6	7	8	9	10	Ac
1	165277	79	112	2096	2336	6428	7850	16	125	71	90
2	14	6679	0	2930	0	7	2326	0	1517	10	91
3	54	16	5064	69437	5	183	16541	0	1551	0	76
4	1640	1846	167	69437	5	183	16541	0	1551	0	76
5	292	0	17	7	82305	3179	2327	4	34	5527	88
6	2557	7	71	110	3393	130033	2568	4	5687	1	90
7	4215	802	2373	6842	3114	2123	163113	209	353	2	89
8	66	0	26	0	15	21	428	25192	0	0	98
9	537	1332	14	4538	30	6285	2809	5	53397	0	77
10	33	9	2	20	3518	83	122	0	7	96547	96
Val	95	94	65	81	87	88	78	99	85	94	



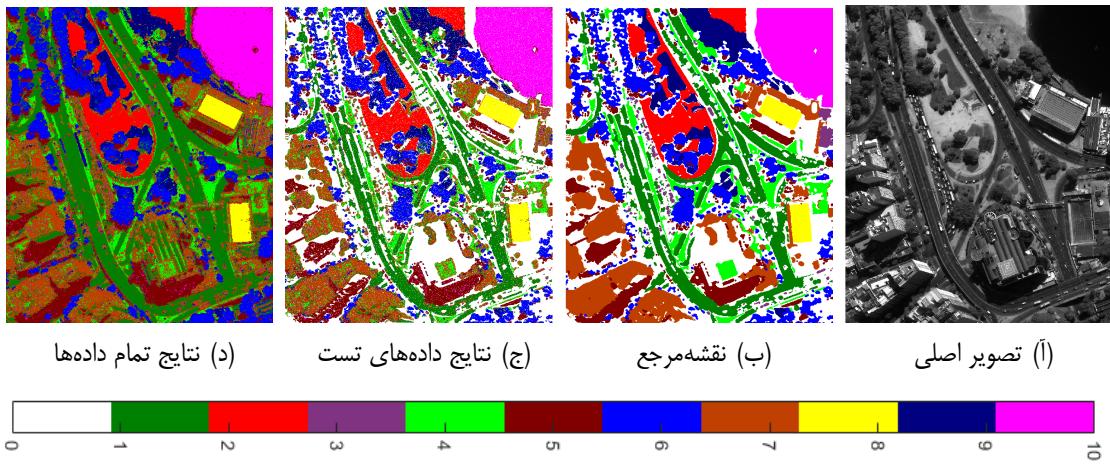
شکل ۳-۲۰: نتایج طبقه‌بند 1D-CNN برای داده جدید

۳-۵-۳ نتایج طبقه‌بندی با tree

درخت طبقه‌بندی یک طبقه‌بند ساده است و قدرت بالایی در تفکیک کلاس‌های نزدیک به هم ندارد. در اینجا ما صرفاً نتایج این طبقه‌بندی را در جدول ۱۵-۳ و شکل ۲۱-۳ مشاهده می‌کنیم. در این طبقه‌بند دو کلاس ساختمان و ماشین بیشتر با هم اشتباه شده‌اند که بررسی‌های قسمت ۱-۵-۳ را در این دو کلاس بهتر نشان داده است. با بررسی بیشتر این دو طبقه‌بند (tree و ML) می‌توانیم حدس بزنیم که در فضای ویژگی بخش زیادی از کلاس ماشین دقیقاً در لابه لایه کلاس ساختمان قرار دارد. به این ترتیب لزوم استفاده بیشتر از اطلاعات مکانی، بیش از پیش نمایان می‌شود. منظور از اطلاعات مکانی بیشتر این است که، برای استخراج ویژگی مکانی همسایگی بزرگتری را در نظر بگیریم.

جدول ۱۵-۳: نتایج طبقه‌بند Tree برای داده جدید

Class	Overall Accuracy=80 , Average Accuracy=77 , Average Validity=77 , $\kappa=77$										
	1	2	3	4	5	6	7	8	9	10	Ac
1	153886	63	70	1805	1866	5561	12951	22	756	35	87
2	26	61058	106	4142	0	8	3177	0	2137	0	86
3	174	260	4457	1434	3	217	9231	15	270	0	28
4	1843	3996	990	5641	0	324	20063	1	3558	0	65
5	1666	0	0	0	79239	3562	3242	0	1	2235	88
6	6640	14	16	283	3719	115167	4748	1	8007	59	83
7	13265	3992	8406	19285	3241	4047	120211	265	3013	96	68
8	53	6	51	50	18	2	508	23962	4	4	97
9	1171	1966	61	4301	33	8872	3451	1	46334	0	70
10	20	0	0	0	2578	40	123	0	0	93567	97
Val	86	86	31	94	87	84	68	99	72	98	



شکل ۲۱-۳: نتایج طبقه‌بندی Tree برای داده جدید

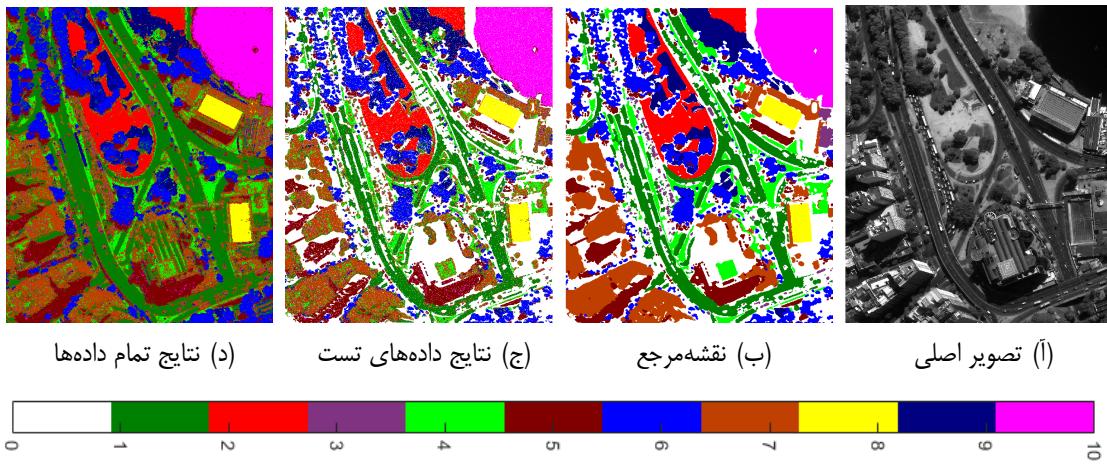
۲-۵-۴ نتایج طبقه‌بندی با 2D-CNN

در بخش ۲-۵-۲ نتایج طبقه‌بندی را با یک شبکه 1D-CNN که روش استخراج ویژگی پیشرفته‌تری نسبت به سایر روش‌ها داشت مشاهده کردیم. اکنون با استفاده از یک شبکه 2D-CNN ساده با همان روش استخراج ویژگی که در ابتدای قسمت ۳-۵ توضیح دادیم فرایند طبقه‌بندی را مشاهده می‌کنیم. همچنین این روش از مقاله [۴۴] الهام گرفته شده است. ولی نحوه استخراج ویژگی ما همان روشی است که در ابتدای قسمت ۳-۵ توضیح دادیم و شبکه انتخاب شده همان شبکه بخش ۲-۳-۳ که در شکل ۱۲-۳ اطلاعات این شبکه آمده است. همچنین در جدول ۱۶-۳ و شکل ۲۲-۳ نتایج این بخش نشان داده شده است.

با بررسی نتایج این بخش در می‌باییم که استفاده از همسایگی‌ها در شبکه کانوالوشنی حتی بعد از استخراج ویژگی اثر به سزاپی در صحت و اعتبار نتایج دارد. همچنین یکی از چالش برانگیزترین کلاس‌ها، کلاس

جدول ۳-۱۶: نتایج طبقه‌بندی 2D-CNN برای داده جدید

Class	Overall Accuracy=91 , Average Accuracy=88 , Average Validity=90 , $\kappa=90$										
	1	2	3	4	5	6	7	8	9	10	Ac
1	165034	80	30	1141	1713	4591	4233	0	193	0	93
2	82	67349	4	1612	0	10	758	0	839	0	95
3	43	5	8252	262	0	72	7422	0	5	0	51
4	1581	883	238	74178	0	247	9232	0	1357	0	85
5	596	0	0	0	86261	1872	1106	15	0	95	96
6	2159	9	18	191	2572	127994	1284	0	4424	3	92
7	5814	413	3176	8769	32	2009	151791	117	435	11	86
8	0	0	0	3	1	0	112	24542	0	0	99
9	391	937	11	3236	41	4105	1342	0	56125	2	85
10	0	0	0	0	511	0	48	0	0	95769	99
Val	94	94	70	83	91	91	86	99	89	100	



شکل ۳-۲۲: نتایج طبقه‌بندی 2D-CNN برای داده جدید

ساختمان (کلاس ۷) است که نمونه‌هایی از آن به چندین کلاس مختلف اختصاص یافته است دلیل آن این است که این کلاس به دلیل وجود شکل‌ها و بافت‌های متفاوت ساختمان‌ها بازه گسترده‌ای از فضای ویژگی‌های مکانی را اشغال کرده است، و هم به دلیل زاویه تابش متفاوت نور برای هر طرف از ساختمان‌ها بازه گسترده‌ای از فضای ویژگی‌های طیفی را نیز اشغال کرده است.

پس تا به اینجا می‌توانیم بگوییم که بزرگترین مشکل ما کلاس ساختمان است. کلاس ۳ یعنی کلاس ماشین نیز با ساختمان تداخل زیادی پیدا می‌کند که از دلایل مهم آن کم بودن تعداد نمونه‌های این کلاس است. باید سعی کنیم با روش‌هایی برای این کلاس داده مصنوعی تولید کنیم.

۳-۶ طرح اولیه برای روش پیشنهادی

از آزمایش‌های که تا به حال انجام دادیم، و نتایجی که گرفتیم مشخص است که فیلتربانک گابور مخصوصاً به صورتی که برش خورده باشد توانایی بالایی در استخراج متنوع ویژگی‌های مکانی دارد. از طرفی داده‌ای که تولید کردیم در عین دقت بالای نقشه مرجع ممکن است دارای اشکالاتی باشد. از این رو باید مدلی که پیشنهاد می‌دهیم به نوعی باشد تا با تصویر ابرطیفی نیز سازگار باشد تا به بتوانیم به طور منصفانه نتایج را با سایر روش‌ها بررسی کنیم و میدانیم که بیشتر روش‌ها برای تصاویر ابرطیفی استفاده می‌شوند. روش ارائه ما در فصل بعد ارائه یک مدل اولیه است و این که چگونه با ایجاد تغییراتی این مدل را، مرحله به مرحله بهبود بخشیم تا نتیجه خوبی برای تصویر چندطیفی ما به همراه داشته باشد. در ابتدا و در بخش مرور روش‌های پیشین به مقالاتی اشاره کردیم که در آن‌ها با استفاده از فیلتربانک گابور و شبکه‌های CNN چه به صورت یک‌بعدی و چه به صورت دو‌بعدی تصاویر ابرطیفی را طبقه‌بندی نموده‌اند به خصوص در [۴۴] که در بخش ۳-۵-۴ بررسی‌هایی را بر روی آن انجام دادیم. یک روش جالب دیگر در پایان نامه [۴۵] عنوان شده است که در آن فیلتربانک گابور در لایه‌های کانوالوشنی قرار داده شده است که طی فرایند یادگیری ممکن است این فیلترها اصلاح شده و به شکل ایده‌آلی برای تصاویر مورد نظر تغییر کنند. به نظر می‌رسد این روش برای استفاده از فیلتربانک گابور مناسب‌تر باشد.

روش کار این پایان نامه [۴۵] به طور خلاصه این است که از تصویر ابرطیفی (پس به طور کلی این پایان نامه برای طبقه‌بندی تصاویر ابرطیفی طراحی شده است). تبدیل PCA اخذ شده است و سپس چند pc اول به عنوان ورودی به شبکه 2D-CNN اعمال می‌شود به طوری که مقدار اولیه فیلترهای لایه اول آن شبکه به جای شروع از نقاط تصادفی، از فیلتربانک گابور شروع می‌شوند. در طی فرایند یادگیری فیلترهای لایه اول که از گابور شروع شده‌اند با توجه به نوع داده می‌توانند تغییراتی داشته باشند تا با داده مورد نظر تطابق بیشتری پیدا کنند. به نظر می‌رسد این فرایند برای استفاده از فیلترهای گابور مناسب‌تر باشد و ما نیز با توجه به توضیحاتی که در فصل ۲ دادیم در نظر داریم تا این کار را بر روی شبکه مورد نظر خود انجام دهیم.

فصل چهارم

روش پیشنهادی

۴-۱ مقدمه

پس از مطالعه فصل ۲ و ۳ اکنون با کنار هم گذاشتن تمام مطالب، قصد داریم روشی ارائه دهیم که تطابق نسبی با ساختار چشم انسان داشته باشد. طبق آنچه در بخش ۴-۲ در مورد تطابق فیلتربانک گابور با ساختار چشم گفتیم می‌بایست از تصویر PAN برای استخراج ویژگی‌های مکانی استفاده کنیم. چرا که در مرکز بینایی رزولوشن تصویر بیشتر است و مانیز همین کار را انجام می‌دهیم در مورد ادغام ویژگی‌های طیفی با توجه به اندازه کوچک‌تر (رزولوشن کمتر) تصویر چندطیفی نسبت به تصویر تکباند دو متده را برای تطابق این داده‌ها ارائه خواهیم کرد.

لازم به تذکر دوباره است که چون تصویر چندطیفی رایج و برچسب خورده‌ای برای داده‌های چندطیفی وجود ندارد تصویری برای بررسی روش پیشنهادی خود تولید نمودیم و از آنجایی که قصد داریم روش خود را با سایر روش‌ها نیز مقایسه کنیم باید این روش به اندازه کافی قابلیت انعطاف داشته باشد تا با تصاویر ابرطیفی نیز سازگار باشد تا به این صورت بتوانیم این روش را با سایر روش‌ها مقایسه کنیم.

ما روش پیشنهادی خود را در ۳ مرحله بیان می‌کنیم هر کدام از این مراحل، در طول مسیر انجام پایان نامه و آزمایش‌های انجام شده بر روی چند تصویر و داده متفاوت، شکل گرفته است. ما با بررسی هر مرحله، سعی می‌کنیم دلیل نیاز به شکل گرفتن مرحله بعد را بیان کرده و سپس راه حل خود را بیان و توجیه کنیم. و با بررسی نتایج در هر مرحله از بهبود نتایج اطمینان حاصل می‌کنیم.

۴-۲ ایده اولیه روش پیشنهادی

با توجه به حجم زیاد اطلاعات مکانی، در تصاویر تکباند ماهواره‌های جدید، ما ابتدا متناسب با آنچه خواندیم سعی در طراحی یک شبکه عصبی می‌کنیم که بهترین نتیجه را برای استفاده از اطلاعات مکانی به ما ارائه دهد. که با توجه به مطالب گفته شده در قدم اول باید فیلترهای گابور را در اولین لایه از این شبکه قرار دهیم. در ابتدا مشخصات فیلتربانک مورد نظر خود را در جدول ۱-۴ مشاهده می‌کنیم و سپس به توضیح شبکه و اطلاعات آن خواهیم پرداخت.

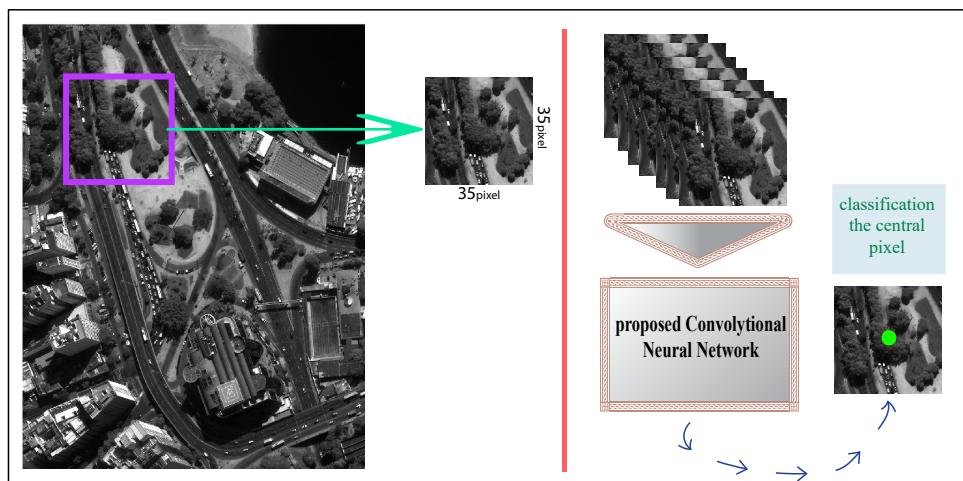
دقیق کنید که در گام اول قصد داریم فقط با استفاده از تصویر PAN تصویر را طبقه‌بندی کنیم که قطعاً نتایج

جدول ۱-۴: اطلاعات فیلتربانک انتخاب شده برای قرار گرفتن در شبکه CNN پیشنهادی

فیلتربانک گابور برش خورده	
طول موج بر حسب پیکسل	{3, 5, 9, 15 }pixel
اندازه برش	23 pixel
زاویه ها	{0,30,60,90,120,150}°
تعداد کل ویژگی ها	$4 \times 6 = 24$

ضعیفی خواهیم گرفت ولی هدف ما این بوده که یک شبکه ایجاد کنیم که نهایت استفاده را از اطلاعات مکانی داشته باشیم. با توجه به اندازه بزرگ فیلتربانک مورد نظر، ارائه‌ی شبکه‌ای که هم از نظر سرعت و تعداد پارامترهای قابل تعلیم بھینه باشد و هم از دقت طبقه‌بندی قابل قبولی برخوردار باشد یک مسئله مهم است ما اینجا یک بدء-بستان بین این دو معیار انجام داده‌ایم.

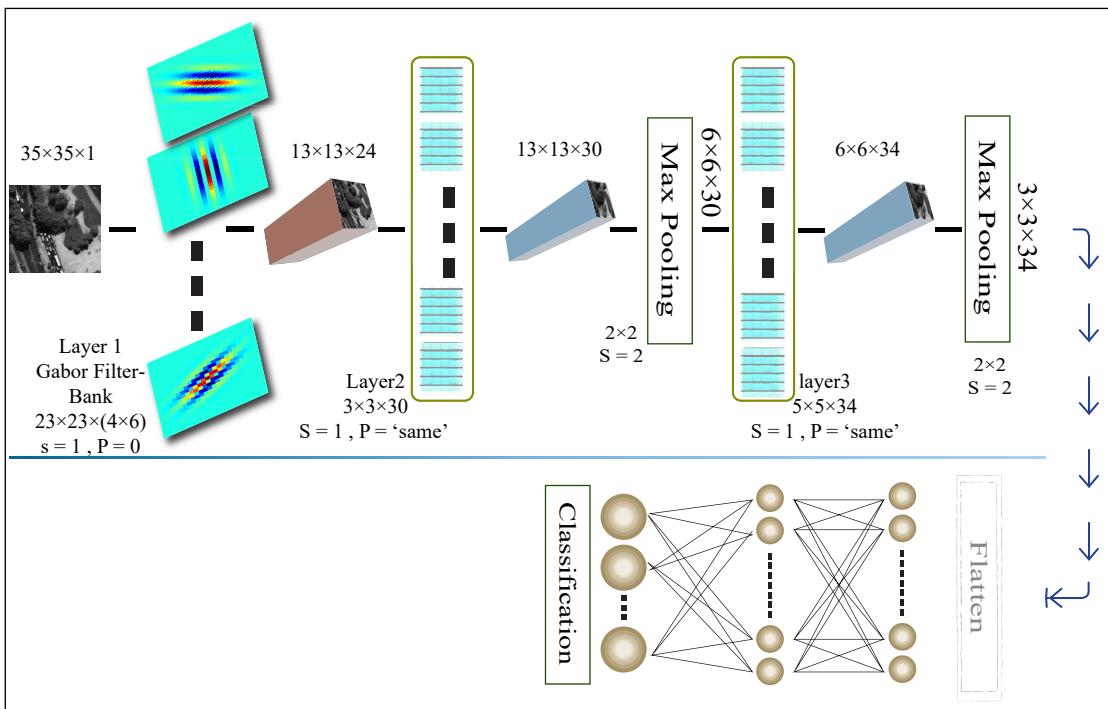
روند کلی استفاده از ویژگی‌های مکانی با شبکه CNN و فیلترهای گابور به این شکل است که ما یک پنجره 35×35 برای هر پیکسل از تصویر تک‌باند ایجاد می‌کنیم. از آنجایی که تصویر ما دارای سایز 1301×1100 است یعنی ما به همین تعداد پیکسل، پنجره‌های 35×35 داریم. سپس برای هر پنجره یک برچسب داریم که برابر برچسب پیکسل مرکزی است. پس به طور کلی شکل داده ورودی به شبکه یک پنجره به سایز $1-4$ تعریف شده است و خروجی یک کلاس است که مربوط به پیکسل مرکزی این تصویر می‌باشد. در شکل ۱-۴ روند کلی مرحله اولیه نشان داده شده است.



شکل ۱-۴: اطلاعات کلی روش پیشنهادی اولیه

شبکه مورد نظر به گونه‌ای است که در لایه اول آن فیلترهای گابور که در جدول ۱-۴ اطلاعات آن نمایش داده شده است قرار گرفته است. در شکل ۲-۴ و جدول ۲-۴ اطلاعات این شبکه موجود است. همچنین

اطلاعات مربوط به بخش Flatten و پس از آن فقط در جدول قرار دارد. در جدول ۳-۴ نیز اطلاعات مربوط به آموزش این شبکه موجود است.



شکل ۳-۴: جزئیات شبکه مورد استفاده در روش پیشنهادی اولیه

در فصل قبل مشکلات طبقه‌بندی داده‌ای که تولید کردیم را بررسی نمودیم و دیدیم مشکلات زیادی در طبقه‌بندی این داده حتی با استفاده از اطلاعات مکانی و طیفی موجود است. در نتیجه فقط استفاده از اطلاعات مکانی قطعاً مشکلات بیشتری به همراه دارد. اما می‌بایست برای پیدا کردن بهینه‌ترین شبکه برای استخراج حداقلی ویژگی‌های مکانی، این بررسی صورت گیرد. یک نکته مهم دیگر که باید توضیح داده شود این است که، همانطور که در بخش ۳-۵-۴ اشاره کردیم می‌بایست برای کلاس ماشین (کلاس ۳) که دارای نمونه برچسب خورده کمتری است نمونه آموزشی مصنوعی تولید کنیم.

برای تولید داده مصنوعی، آن ۲۰۴ درصد از دیتا‌یی که از این کلاس انتخاب می‌شود را با زاویه‌های ۹۰، ۱۸۰ و ۲۷۰ درجه می‌چرخانیم. با همین کار ساده نتایج با تصویر تک‌باند تا حدود ۱۵ درصد برای این کلاس بهبود می‌یابد. در جدول ۳-۴ نتایج این شبکه با داده مورد نظر آمده است و همچنین نتایج تصویری آن در شکل ۳-۴ قرار گرفته است. در این جدول فقط داده‌های آزمایشی شمرده شده‌اند ولی در شکل ۳-۴(د) تمام نقاط برچسب دار نشان داده شده است. دلیل استفاده از فیلتربانک گابور، در اولویت اول این است که فرایند آموزش سریع‌تر اتفاق بیفتند چرا که در بخش ۳-۴-۳-۲ با فیلترهایی که در لایه اول شبکه AlexNet ساخته

جدول ۴-۲: اطلاعات کامل شبکه اولیه برای روش پیشنهادی

ردیف	نوع لایه	اطلاعات لایه	شکل خروجی
۱	input Image	IM:35×35	-
۲	2D-Conv gabor init	23×23×24 , S=1 , P=0	13×13×24
۳	Batch norm	-	13×13×24
۴	relu	-	13×13×24
۵	2D-Conv	3×3×30 , S=1 , P='same'	13×13×30
۶	Batch norm	-	13×13×30
۷	relu	-	13×13×30
۸	MaxPool	2×2 , S=2 , P=0	6×6×30
۹	2D-Conv	5×5×34 , S=1 , P='same'	6×6×34
۱۰	Batch norm	-	6×6×34
۱۱	relu	-	6×6×34
۱۲	MaxPool	2×2 , S=2 , P=0	3×3×34
۱۳	flatten	-	306 neuron
۱۴	Fully connected	256	256 neuron
۱۵	Batch norm	-	256 neuron
۱۶	relu	-	256 neuron
۱۷	Fully connected	تعداد کلاس‌ها	تعداد کلاس‌ها
۱۸	SoftMax	-	برچسب

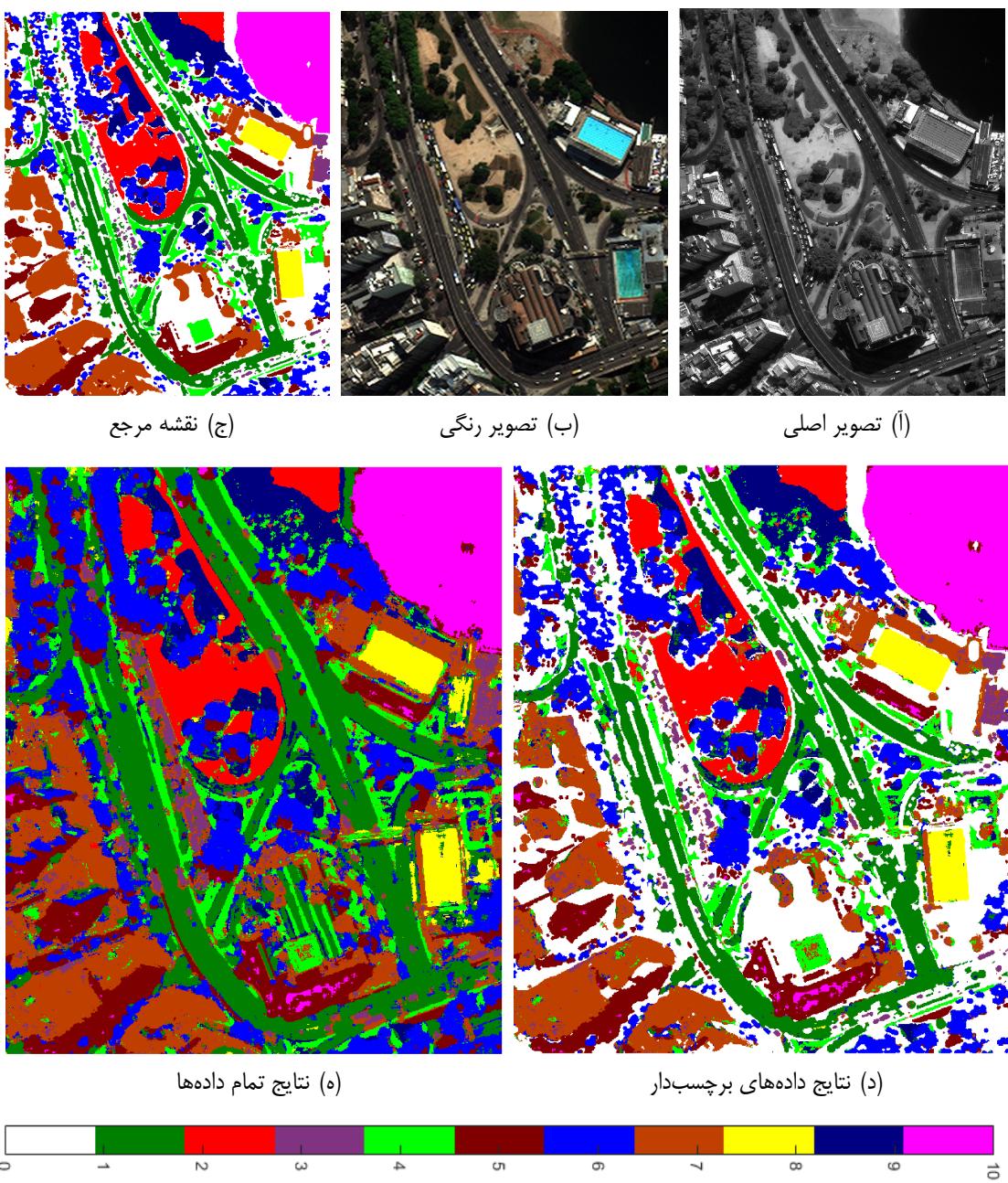
جدول ۴-۳: اطلاعات شبکه مربوط به آموزش شبکه پیشنهادی اولیه

TrainPercent	Validation	Batch size	epochs	Learning rate	Training options
2.4%	0.6%	64	25	$0.01 \Rightarrow 10^{-6}$	"SGDM"

جدول ۴-۴: نتایج شبکه مورد استفاده در روش پیشنهادی اولیه برای داده‌های آزمایشی

Class	Overall Accuracy=89 , Average Accuracy=89 , Average Validity=88 , $\kappa=87$										
	1	2	3	4	5	6	7	8	9	10	Ac
1	167937	170	394	854	1516	4903	2437	808	3754	57	92
2	52	66058	17	3555	1	10	869	89	720	0	93
3	36	2	14042	178	1	28	2165	1	0	0	85
4	612	3486	399	66066	0	1289	5066	337	5563	0	80
5	964	0	14	1	84293	3454	1367	42	19	4604	89
6	2966	47	119	1069	4442	129298	3463	166	3296	15	89
7	2454	633	7575	5916	2434	4783	140320	794	891	15	85
8	268	0	0	43	4	13	79	24466	42	0	98
9	1893	88	1	5326	31	4499	1316	217	51838	2	78
10	1	0	3	0	1928	2	15	0	0	95382	98
Val	95	93	62	80	89	87	89	91	78	95	

شدند آشنا شدیم و دیدیم که همانند فیلترهای گابور می‌باشند. در اولویت دوم عملکرد بهتر شبکه است. با این حال اگر همین شبکه را بدون استفاده از فیلربانک گابور در لایه اول استفاده کنیم، در طی epoch ۲۵ که برای این شبکه در نظر گرفته‌ایم دقیق شبکه در تمام پارامترها حدود ۵ درصد کمتر از جدول ۴-۴ است. و با افزایش



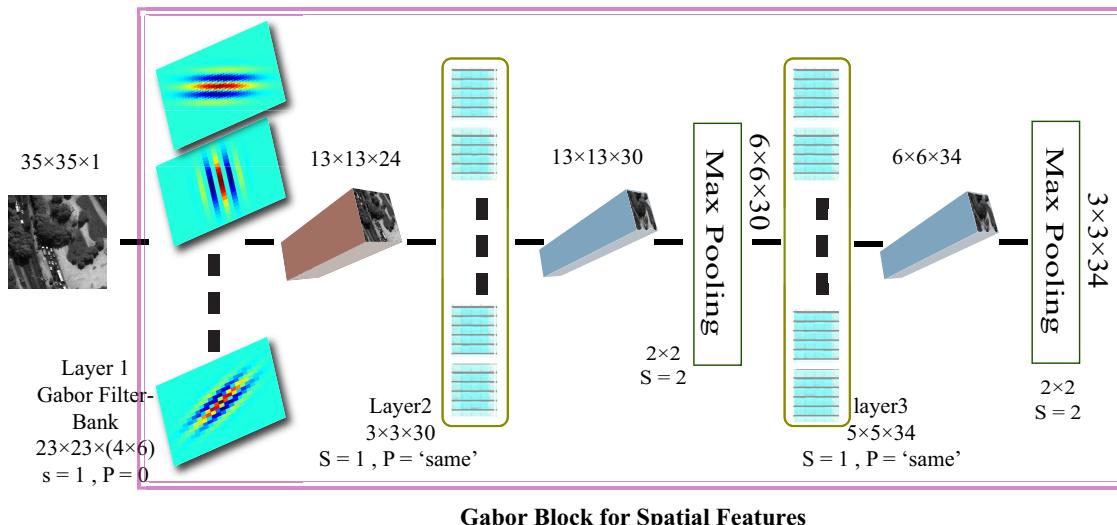
شکل ۴-۳: نتایج شبکه مورد استفاده در روش پیشنهادی اولیه

تعداد epoch ها به عدد ۵۰ باز هم دقت حدود یک درصد پایین تر است.

تنها حالتی که دقت شبکه بدون فیلتربانک گابور با ۲۵ epoch به حدود اعداد جدول ۴-۴ برسد این بود که علاوه بر نقاط آموزشی در نظر گرفته شده برای شبکه، ۱ درصد دیتای آموزشی جدید به داده‌های آموزشی اضافه کنیم. پس به طور کلی با استفاده از فیلتربانک گابور به هدف بیشتر کردن سرعت و کم کردن داده آموزشی دست یافتیم، ولی دقت کلی چنان تغییری نداشت و با افزایش تعداد epoch و یا افزایش داده آموزشی می‌توان

بدون استفاده از فیلتربانک گابور نیز به همان دقت دست یافت.

پیش از شروع مرحله دوم روش پیشنهادی ابتدا از شکل ۴-۲ یا به عبارتی از مرحله اول روش پیشنهادی، لایه‌های کانوالوشنی اولیه را برای مرحله دوم استفاده می‌کنیم. و نام این بخش را بلوک گابور (Gabor Block) می‌نامیم و در مرحله دوم روش پیشنهادی از این بخش استفاده می‌کنیم. در شکل ۴-۴ این بخش نشان داده شده است که در واقع بخش خاصی از شبکه اولیه است.



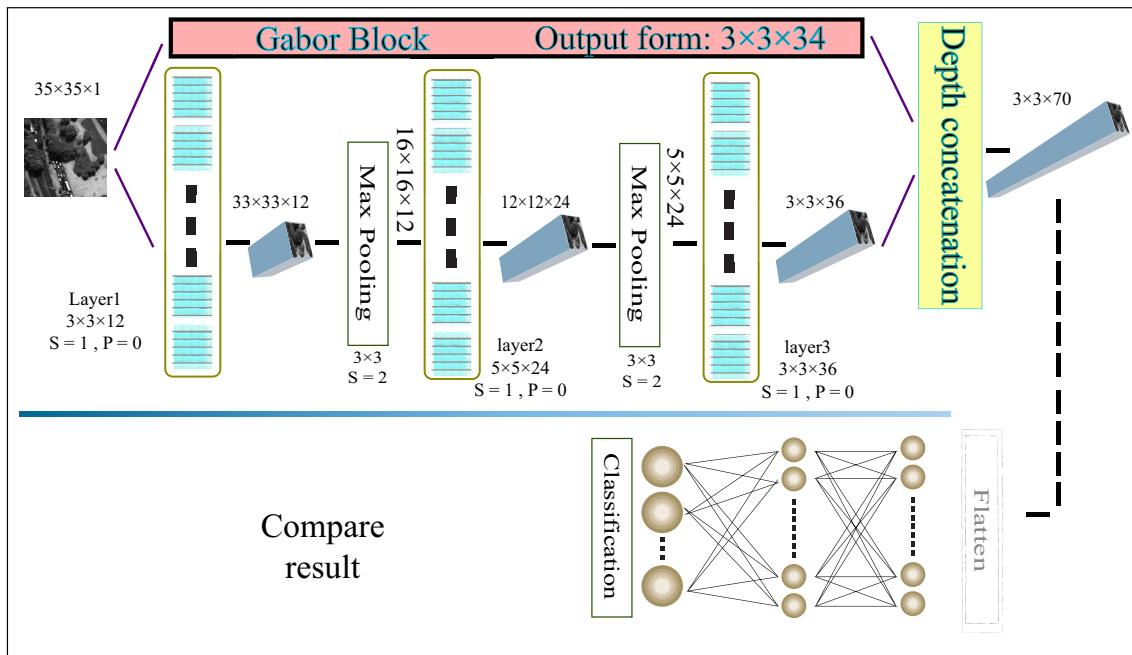
شکل ۴-۴: بلوک گابور برای استفاده در مراحل بعدی شبکه پیشنهادی(بخش داخل مستطیل)

۴-۳ مرحله دوم در مسیر دست یابی به روش پیشنهادی

تا به اینجا توانستیم با استفاده از فیلتربانک گابور و شبکه CNN تا حد زیادی اطلاعات مکانی را استخراج کنیم. اکنون نوبت آن است تا مرحله دوم را برای بهبود استخراج ویژگی‌های مکانی معرفی کنیم. اما یک اصل مهم در شبکه CNN که در اکثر روش‌های استفاده از آن رعایت شده است این است که فیلترهای لایه اول اندازه کوچکی داشته باشد تا بتواند تغییرات جزئی تصویر در زاویه‌های مختلف را پیدا کند و سپس اندازه فیلترها بزرگ و بزرگتر شود.

اکنون با توجه به این نکته، یک بلاک دیگر را موازی با بلاک گابور در شبکه قرار می‌دهیم که طبق این استاندارد عمل کند. با این کار روند آموزش شبکه 30×30 درصد کندرتر می‌شود ولی دقت نزدیک به 5 درصد افزایش می‌یابد که با در نظر گرفتن این که هنوز از ویژگی‌های طیفی استفاده نکرده‌ایم رقم قابل توجهی است. در شکل ۴-۵ جزئیات شبکه در مرحله دوم را مشاهده می‌کنیم. همچنین در جدول ۴-۵ اطلاعات جزئی شبکه مشاهده

می شود. این نکته نیز قابل توجه است روند کلی همانند شکل ۱-۴ است فقط شبکه مورد استفاده تغییر کرده است.



شکل ۴-۵: جزئیات شبکه مورد استفاده در مرحله دوم طراحی شبکه پیشنهادی

جدول ۴-۵: اطلاعات کامل شبکه در مرحله دوم برای روش پیشنهادی

ردیف	نوع لایه	اطلاعات لایه	شکل خروجی
۱	input Image	IM: 35×35	-
۲	2D-Conv	$23 \times 23 \times 24, S=1, P=0$	$33 \times 33 \times 12$
۳	Batch norm	-	$33 \times 33 \times 12$
۴	relu	-	$33 \times 33 \times 12$
۵	MaxPool	$3 \times 3, S=2, P=0$	$16 \times 16 \times 12$
۶	2D-Conv	$5 \times 5 \times 24, S=1, P=0$	$12 \times 12 \times 24$
۷	Batch norm	-	$12 \times 12 \times 24$
۸	relu	-	$12 \times 12 \times 24$
۹	MaxPool	$3 \times 3, S=2, P=0$	$5 \times 5 \times 24$
۱۰	2D-Conv	$3 \times 3 \times 36, S=1, P=0$	$3 \times 3 \times 36$
۱۱	Batch norm	-	$3 \times 3 \times 36$
۱۲	relu	-	$3 \times 3 \times 36$
۱۳	concat 2 Block	$3 \times 3 \times 70$	Gabor Block shape: $3 \times 3 \times 34$
۱۴	flatten	-	630 neuron
۱۵	Fully connected	256	256 neuron
۱۶	Batch norm	-	256 neuron
۱۷	relu	-	تعداد کلاس‌ها
۱۸	Fully connected	تعداد کلاس‌ها	برچسب
۱۹	SoftMax	-	

حالا که با شبکه جدید آشنایی می‌توانیم نتایج را بررسی کنیم اما قبل از آن اطلاعات آموزش این شبکه نیز مانند مدل اولیه است و در همان جدول ۴-۳ موجود است. همچنین ذکر این نکته لازم است که با حذف فیلتر گابور از لایه اول بلوک مربوطه دقت طبقه‌بندی با epoch ۱۵ بیشتر با اختلاف جزئی به نتایج حالتی که در آن فیلترهای گابور موجود باشد تزدیک می‌شود. پس با اضافه کردن بلوک جدید وجود فیلترهای گابور فقط به افزایش سرعت شبکه محدود شده است.

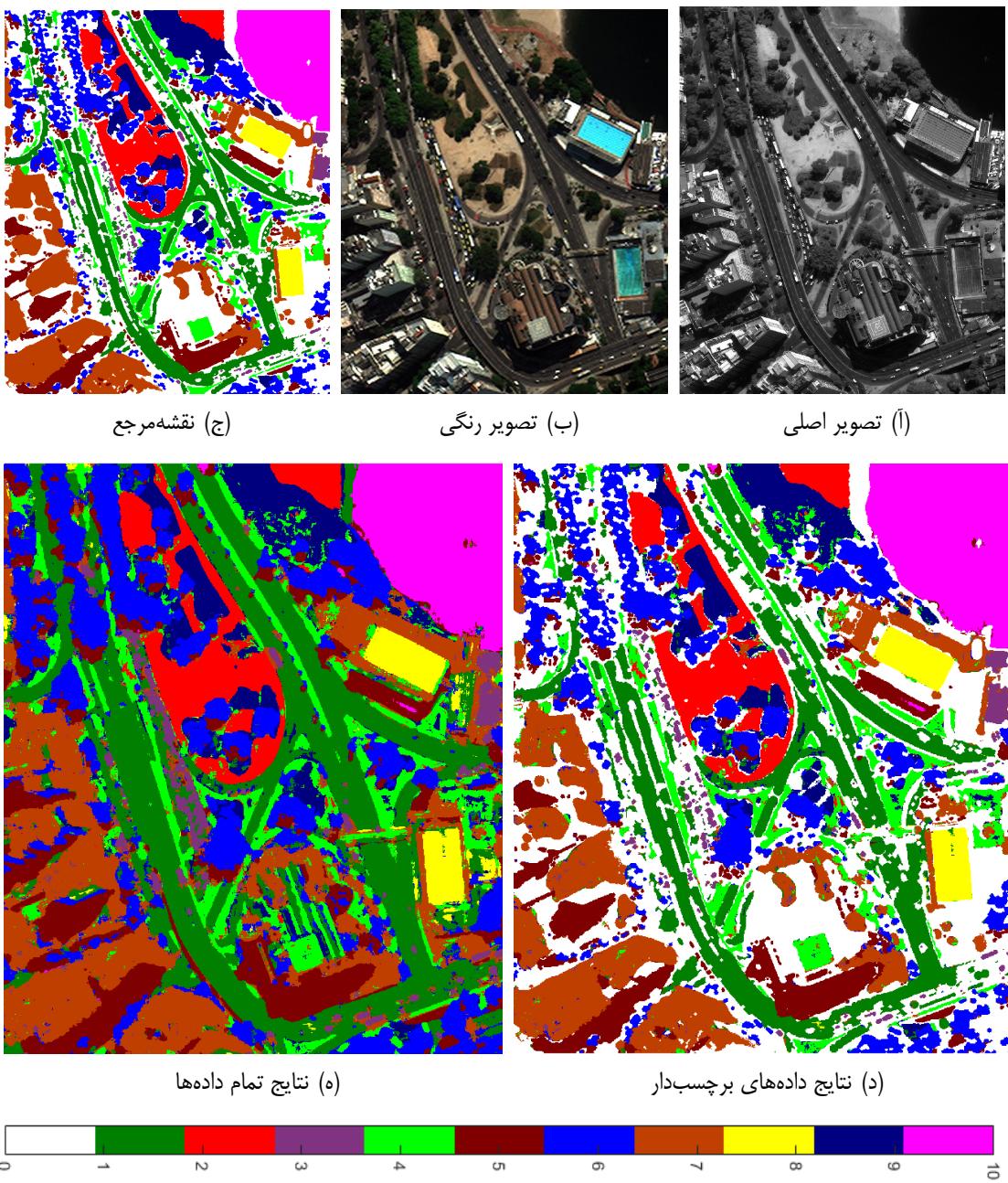
در جدول ۴-۶ نتایج این شبکه به صورت عددی موجود است و همچنین در شکل ۴-۶ نیز نتایج به صورت تصویری نشان داده شده است. در این جدول فقط داده‌های آزمایشی شمرده شده‌اند ولی در شکل ۴-۶(د) تمام نقاط برچسب دار نشان داده شده است.

جدول ۴-۶: نتایج شبکه مورد استفاده در مرحله دوم طراحی شبکه پیشنهادی

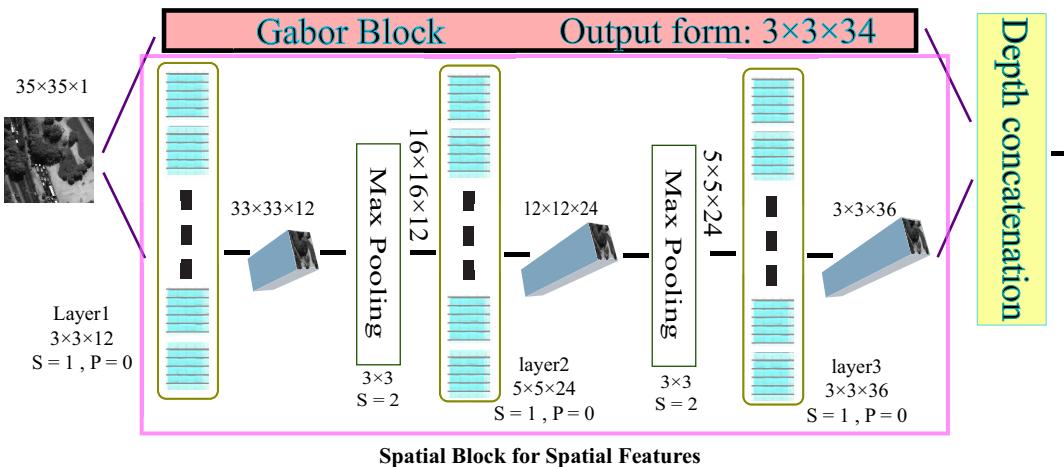
Class	Overall Accuracy=93 , Average Accuracy=93 , Average Validity=92 , $\kappa=92$										
	1	2	3	4	5	6	7	8	9	10	Ac
1	174730	139	427	908	1199	1999	1457	344	1626	1	96
2	36	67842	15	2174	0	31	484	27	762	0	95
3	44	2	151138	132	2	28	1102	1	3	1	92
4	665	1914	323	89967	4	723	3123	177	4075	0	87
5	594	0	7	1	89697	2376	1320	5	13	745	95
6	2090	78	118	827	3609	132865	2274	105	2913	2	92
7	1461	124	2036	3262	2058	1971	154179	333	22		93
8	118	0	0	45	0	11	57	24681	3	0	99
9	1308	653	0	4024	36	3090	892	51	55951	0	85
10	3	0	0	0	828	2	4	0	0	96494	99
Val	96	96	84	86	92	93	93	96	85	99	

با توجه به جدول‌های ۴-۴ و ۴-۶ می‌توانیم شاهد رشد همه معیارهای باشیم همچنین کلاس‌های ۳ و ۷ فاصله بیشتری از هم گرفتند. پس در حل این مشکل یک گام به سمت جلو حرکت کردیم. سایر کلاس‌ها نیز همگی در اعتبار و صحبت رشد داشته‌اند. با دقت به دو شکل ۴-۳(ه) و ۴-۶(ه) نیز مشاهده می‌شود دقت طبقه‌بندی برای اشیاء بدون برچسب بیشتر شده است.

تا به اینجا توانستیم از اطلاعات مکانی حداقل استفاده را بکنیم و به دقتی بررسیم، که با روش‌هایی که در فصل ۳ بررسی نمودیم قابل قیاس باشد. اگر یادمان باشد آن روش‌ها با استفاده از اطلاعات مکانی و طیفی به چنین دقتی دست یافتند که ما فقط با تصویر تک‌باند به این دقت رسیده‌ایم. پیش از شروع بخش مرحله سوم از روش پیشنهادی باید بلوک جدیدی را که ساخته‌ایم نیز نام گذاری کنیم. این بلوک کانولوشنی جدید را بلوک مکانی (Spatial Block) می‌نامیم. در شکل ۷-۴ این بخش از شبکه داخل مستطیل قرار گرفته است.



شکل ۴-۶ نتایج شبکه مورد استفاده در مرحله دوم طراحی شبکه روش پیشنهادی



شکل ۴-۷: بلوک مکانی برای استفاده در مراحل بعدی شبکه پیشنهادی

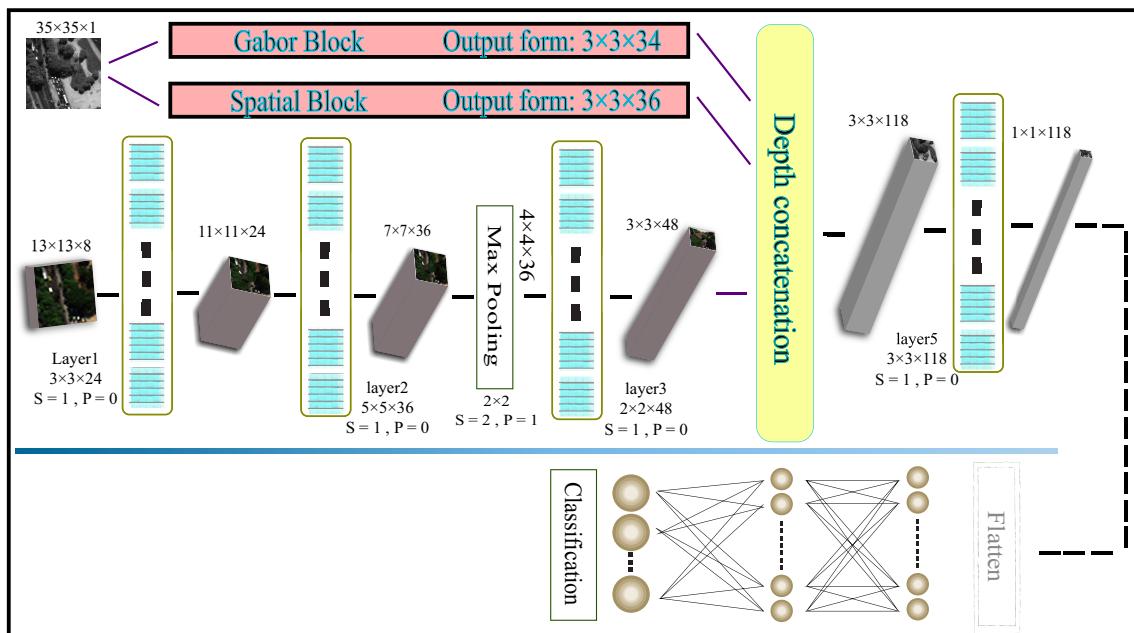
۴-۴ مرحله نهایی و روش پیشنهادی کلی

اکنون که توانستیم از اطلاعات مکانی به دقت بالایی دست یابیم زمان استفاده از ویژگی‌های طیفی است. برای استفاده از ویژگی‌های طیفی در کنار ویژگی‌های مکانی راههای زیادی موجود است. به عبارت دیگر ویژگی‌های طیفی می‌توانند به روش‌های مختلف در کنار ویژگی‌های مکانی قرار گرفته و فرایند طبقه‌بندی صورت پذیرد. ما قصد داریم این کار را با یک شبکه انجام دهیم و تمام ۸ باند تصویر خود را به این شبکه اضافه کنیم و در یک لایه در انتهای شبکه تمام ویژگی‌ها را پشت سر هم قرار دهیم تا شبکه طبقه‌بندی را انجام دهد.

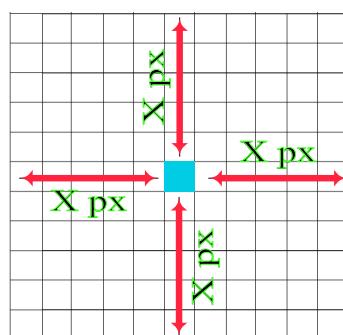
ساده‌ترین شبکه می‌تواند شبکه MLP باشد. اما اگر خاطرمان باشد ما برای هر پیکسل از تصویر چند طیفی ۱۶ پیکسل (یک مربع 4×4) در تصویر تک‌باند داریم، در نتیجه طی هماندازه کردن تصویر چند طیفی به تک باند، اطلاعات طیفی در پیکسل مورد نظر نمی‌تواند قابل اعتماد باشد از این رو به ناچار برای ویژگی‌های طیفی نیز از شبکه CNN استفاده می‌کنیم. استفاده از ۳ بلوک CNN برای طبقه‌بندی البته که نتیجه بسیار خوبی خواهد داشت اما زمان اجرای برنامه نسبت به مرحله ۲ حدود ۵۰ درصد و نسبت به مرحله ۱ حدود ۱۰۰ درصد افزایش می‌یابد. در بخش بعدی یک راه حل برای افزایش سرعت ارائه خواهیم کرد که انتظار خواهیم داشت نتایج دقت خود را از دست دهنند.

با توجه به بلوک مکانی و بلوک گابور، دیاگرام شبکه به صورت کلی شکل ۴-۸ خواهد بود. همچنین

اطلاعات جزئی تر شبکه در جدول ۷-۴ درج شده است. مجدداً تاکید می‌شود روند کلی همچنان مانند شکل ۱-۴ است و فقط شبکه تغییر کرده است. با توجه به اطلاعات شبکه می‌بینیم که تصویر چندطیفی نیز به صورت یک پنجره $13 \times 13 \times 8$ به صورت موازی در کنار تصویر تک‌باند به شبکه وارد شده است.



شکل ۸-۴: جزئیات شبکه پیشنهادی



شکل ۹-۴: مزیت پنجره با طول فرد برای پیکسل مرکزی

همچنان که گفتیم با توجه به هماندازه کردن دو تصویر، ممکن است پیکسل‌هایی که روی هم قرار گرفته‌اند اطلاعات طیفی دقیق محل را به ما ارائه ندهند لذا از شبکه CNN استفاده می‌کنیم تا این مشکل را حل کنیم و از اطلاعات طیفی همسایگی‌ها نیز استفاده شود. نکته دیگر دلیل فرد بودن اندازه پنجره‌ها است. با دقت به شکل ۹-۴ توجیه می‌کنیم که انتخاب پنجره با طول فرد باعث می‌شود پیکسل هدف دقیقاً در مرکز پنجره قرار گیرد. این باعث می‌شود ما به طور منصفانه‌ای همسایگی را در اطراف پیکسل هدف تولید کنیم. به همین دلیل

جدول ۴-۷: اطلاعات کامل شبکه برای روش پیشنهادی

ردیف	نوع لایه	اطلاعات لایه	شكل خروجی
۱	input PAN Image	IM:35×35	-
۲	Gabor Block	Connect to depth concat	3×3×34
۳	Spatial Block	Connect to depth concat	3×3×36
۴	input MUL Image	IM:13×13×8	-
۵	2D-Conv	3×3×24 , S=1 , P=0	11×11×24
۶	Batch norm	-	11×11×24
۷	relu	-	11×11×24
۸	2D-Conv	5×5×36 , S=1 , P=0	7×7×36
۹	Batch norm	-	7×7×36
۱۰	relu	-	7×7×36
۱۱	MaxPool	2×2 , S=2 , P=1	4×4×36
۱۲	2D-Conv	2×2×48 , S=1 , P=0	3×3×48
۱۳	Batch norm	-	3×3×48
۱۴	relu	Connect to depth concat	3×3×48
۱۵	Depth concatenation	concat 3 Block	3×3×118
۱۶	2D-Conv	3×3×118 , S=1 , P=0	1×1×118
۱۷	Batch norm	-	1×1×118
۱۸	relu	-	1×1×118
۱۹	flatten	-	118 neuron
۲۰	Fully connected	256	256 neuron
۲۱	Batch norm	-	256 neuron
۲۲	relu	-	256 neuron
۲۳	Fully connected	تعداد کلاس ها	تعداد کلاس ها
۱۹	SoftMax	-	برچسب

هم در تصویر چندطیفی و هم در تصویر تکباند از پنجره هایی با طول فرد استفاده کرده ایم. با توجه به این که تمایل داریم همسایگی ها از همه طرف برابر باشند، طول پنجره به همراه پیکسل مرکزی عدد فردی می شود. یعنی $1 + 2X = \text{odd}$ که در واقع عدد ۱ همان پیکسل مرکزی است و X طول همسایگی از هر طرف است.

برای درک بهتر جدول ۴-۸ توضیحاتی لازم است. در ردیف اول تصویر تکباند وارد دو بلوک مکانی و گابور می شود و هر کدام یک خروجی با فرمت مشخص شده تولید می کنند. در ردیف ۴ تا ۱۴ بلوکی برای تصویر چندطیفی وجود دارد که اطلاعات آن در هر ردیف درج شده است. سپس خروجی هر سه بلوک چون در دو بعد اول هم اندازه هستند، در بعد سوم که عمق آن ها است توسط لایه Depth concatenition پشت سر هم قرار می گیرند.

سپس توسط لایه کانوالوشنی دیگری که اطلاعات آن در جدول مذکور در ردیف ۱۶ قرار دارد اندازه خروجی را به تصویری با یک پیکسل در صفحه و عمق ۱۱۸ تبدیل می کنیم. در واقع این لایه وظیفه دارد ۱۱۸ عدد

ویژگی، از ترکیب‌های متفاوت $1082 = 118 \times 3 \times 3$ استخراج کند. دلیل انتخاب عدد ۱۱۸ مربوط به عمق تصویر قبلی نیست و صرفاً چون در لایه fully connected بعدی ۲۵۶ نورون قرار داده‌ایم خواستیم تعداد ویژگی‌های این لایه تقریباً نصف تعداد نورون‌های لایهی بعدی باشد.

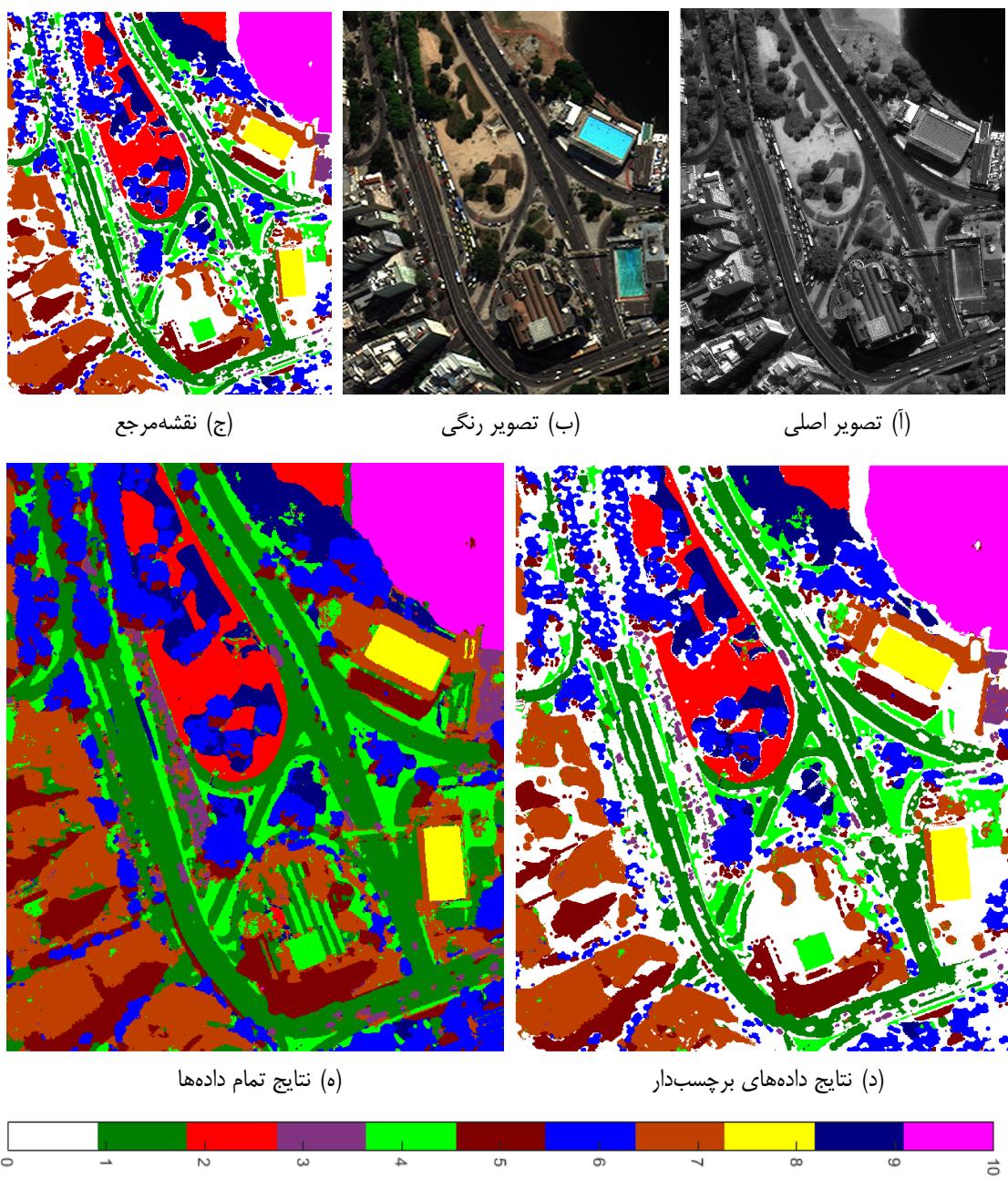
اطلاعات آموزش شبکه نیز در همان جدول ۳-۴ موجود است و هیچ تفاوتی نکرده است. اکنون که با ساختار کلی روش پیشنهادی آشنا شدیم زمان آن رسیده است که نتایج را مشاهده و مقایسه کنیم. در جدول ۴-۸ نتایج این شبکه به صورت عددی موجود است و همچنین در شکل ۱۰-۴ نیز نتایج به صورت تصویری نشان داده شده است. در این جدول فقط داده‌های آزمایشی شمرده شده‌اند ولی در شکل ۱۰-۴(د) تمام نقاط برچسب دار نشان داده شده است.

جدول ۴-۸: نتایج طبقه‌بندی روش پیشنهادی (فقط برای داده‌های تست)

Class	Overall Accuracy=96 , Average Accuracy=96 , Average Validity=96 , $\kappa=95$										Ac
	1	2	3	4	5	6	7	8	9	10	
1	178177	133	361	670	1144	1328	930	0	87	0	98
2	39	69677	12	872	0	26	168	0	577	0	98
3	76	1	14891	153	2	27	1301	0	0	2	91
4	851	142	240	77727	0	307	2320	0	1262	0	94
5	586	0	1	0	90351	2386	1384	29	17	4	95
6	507	21	48	315	2003	138646	787	0	2554	0	96
7	959	21	955	288	1757	1104	157717	222	89	9	95
8	0	0	0	2	0	0	57	24856	0	0	100
9	87	732	0	1786	23	2647	335	0	60395	0	92
10	0	0	0	0	176	0	2	0	0	97153	100
Val	98	98	90	92	95	95	96	99	93	100	

همچنان که در جدول ۴-۸ مشخص است دقت طبقه‌بندی نسبت به روش‌های دیگر پیشرفت زیادی داشته است. همچنین در کلاس‌های چالش برانگیزی که در مورد آن‌ها صحبت‌هایی را مطرح نمودیم مشکلات طبقه‌بندی تا حد زیادی برطرف شده است و صحت و اعتبار تمام کلاس‌ها ۹۰ درصد یا بیشتر است.

در شکل ۱۰-۴(ه) به یک نقشه یکنواخت و نرم دست یافته‌ایم. برخی نقاط که در این تصویر و در شکل ۱۰-۴(د) اشتباه برچسب خورده‌اند نقاط شک برانگیزی حتی برای تصمیم گیری توسط انسان هستند. از طرفی برچسب داده‌ها نیز دقت بالایی دارد ولی در نهایت توسط چشم انسان (نویسنده این متن) بررسی، اصلاح و بازبینی شده است. در نتیجه دستیابی به دقت ۱۰۰ درصد یا نزدیک به آن تقریباً غیر ممکن است. حتی اگر به دقت ۱۰۰ درصد بررسیم به دلیل مشکل عدم اعتماد به نقشه مرجع و وجود حتمی چند پیکسل اشتباه در آن صحت کار انجام شده زیر سوال خواهد رفت.



شکل ۴-۱۰: نتایج تصویری روش پیشنهادی برای داده همه داده ها

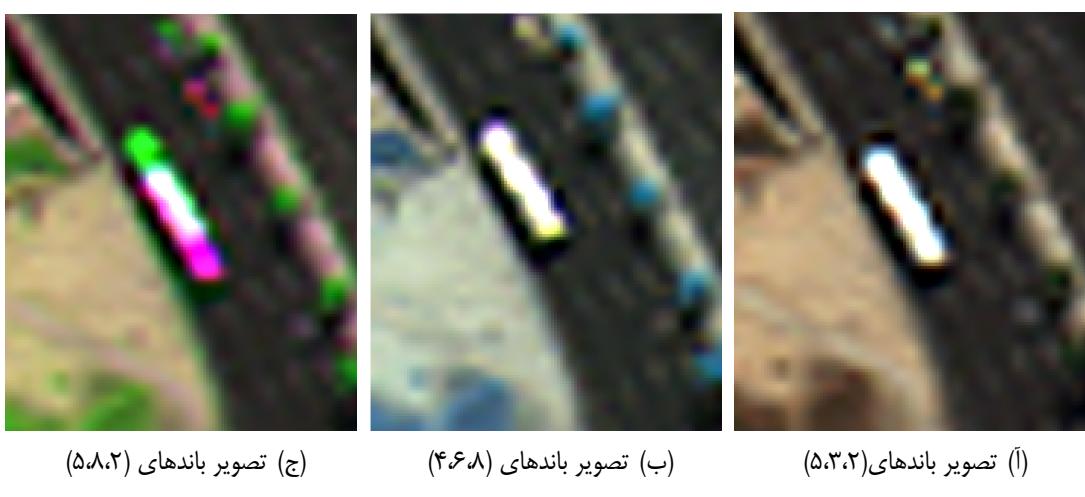
در هر صورت این نقشه مرجع با استفاده از ویژگی های مکانی، ویژگی های طیفی و بررسی و گسترش توسط چشم انسان ساخته شده است ولی وجود خطا در آن به هیچ عنوان رد نمی شود. از طرف دیگر خطاهای آن به دلیل نظارت انسان بسیار کم است. ما در این عکس نزدیک به ۱ میلیون پیکسل را برچسب زده ایم که با اعتماد کامل بالای ۹۷ درصد آن ها برچسب صحیح دریافت کردند.

همچنین در تولید نقشه مرجع هیچ گونه اهمال کاری یا سرسی گذشتن از موضوع صورت نگرفته اما

وجود خطا رد نمی‌شود. از طرفی فرایند اصلی تولید نقشه مرجع بر عهده الگوریتم‌های خوشه‌بندی و طبقه‌بندی بوده است و وظیفه نویسنده فقط ناظارت و در پایان ترکیب و گسترش کلاس بوده است. می‌توان گفت روش تولید نقشه مرجع مورد نظر یک روش نیمه ناظارت شده است به طوری که ابتدا ماشین برچسب‌هایی را برای ما تولید می‌کند(یادگیری بدون ناظارت) و سپس ما نمونه‌ها را تایید یا رد، ترکیب و گسترش می‌دهیم(یادگیری با ناظارت). در واقع فرض بر این است که آنچه انسان انجام داده است صد درصد درست است و به عنوان داده برچسب دار در نظر گرفته می‌شود.

شبکه مورد استفاده در این بخش شبکه‌ای است که از اطلاعات مکانی با دقت بالایی استفاده می‌کند. و زمانی که علاوه بر تصویر تک‌باند تصویر چندطیفی نیز به آن اضافه می‌شود کلاس‌ها تا حد زیادی از هم فاصله گرفته و فرایند طبقه‌بندی با دقت زیادی انجام می‌شود.

در نهایت کلاس خودرو همچنان از مشکلات حل نشده می‌باشد. در مرحله ۱ و ۲ از روش پیشنهادی با تکیه بر تصویر تک‌باند خودروها با ساختمان‌ها تداخل زیادی دارند و وقتی از تصویر چندطیفی کمک می‌گیریم این خطا تا حد زیادی برطرف می‌شود ولی همچنان نسبت به برخی کلاس‌ها دقت کمتری دارد. با کمی دقت به تصاویر شکل ۱۱-۴ مشاهده می‌کنیم برخی خودروها هنوز مشکلات مرحله ۱ و ۲ را در طبقه‌بندی دارند. سوال اینجا است که چرا ویژگی‌های طیفی توانستند برخی مشکلات در این کلاس را حل کنند ولی بقیه این کلاس همچنان مشکل دارند! برای پاسخ به این سوال به شکل‌های (آ) تا (ج) دقت کنید.



شکل ۱۱-۴: تصویر یک خودروی متحرک در چند باند مختلف از تصویر چندطیفی

با دقت به شکل ۱۱-۴(آ) و ۱۱-۴(ب) که دقیقاً از یک ناحیه‌ی تصویر چندطیفی، برش خورده‌اند، مشاهده می‌شود که تصویر اتوبوس موجود در این کادر جایجا شده است. در واقع در شکل ۱۱-۴(ب) نسبت شکل

۴-۱۱(آ) کمی به سمت جلو حرکت کرده است. همچنین در تصویر ۴-۱۱(j) که ترکیبی از باندهای دو تصویر قبل است این گفته تایید می‌شود. در واقع در تصویر آخر باند ۸ به رنگ سبز در آمده و اندکی جلوتر از دو باند قرمز و آبی قرار دارد.

به عبارتی ماهواره QB3 تصاویر چندطیفی خود، در طول موج‌های متفاوت را، در دو زمان متفاوت اخذ می‌کند به همین جهت اشیاء متحرک در یک دسته از باندها در یک مکان قرار دارند و در دسته دیگر کمی در جهت حرکت شیء، در مکانی دیگر قرار می‌گیرد. از همین رو تصویر چندطیفی نمی‌تواند کمک چندانی به کلاس خودرو بنماید چرا که در این تصویر بعضی خودروها ثابت هستند و برخی متحرک، در نتیجه برای یک کلاس خاص، به نام خودرو، دو دسته ویژگی متفاوت وجود دارد. که یکی از این دسته‌ها نیز به کلاس ساختمان نزدیک است.

شاید در نگاه اول، این که تصویر چند طیفی در دو زمان مختلف اخذ شود یک ایراد یا کم توانی در ساخت ماهواره تصور شود ولی با کمی دقیق توجه می‌شویم این کار باعث می‌شود خیلی راحت اشیاء متحرک را در یک تصویر شناسایی کنیم، به شرطی که اولاً، در هر دو دسته از باندها بازتابی تقریباً یکسان داشته باشند (چون باندها فاصله زیادی ندارند و ترتیب آن‌ها نیز به گونه‌ای است که در لابه لای هم هستند، این امر بسیار محتمل است) و دوماً، بخشی از شیء در هر دو دسته قرار گیرد یا یک شیء دیگر از همان جنس و رنگ به صورت ثابت نیز در تصویر وجود داشته باشد. در نهایت با توجه به جدول ۱-۲ ذکر این نکته مهم است که باندهای ۲، ۳، ۴ و ۷ زودتر از باندهای ۱، ۴، ۶ و ۸ اخذ می‌شوند.

۴-۵ نگرشی دیگر در طبقه‌بندی تصاویر چندطیفی

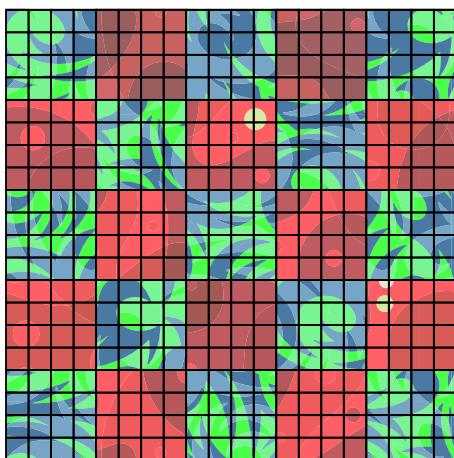
چنین که در بخش‌های قبلی توضیح دادیم تصویر تک‌باند ماهواره WV3 دارای رزولوشن ۴ برابر نسبت به تصویر چندطیفی آن است. ما در این کار طبقه‌بندی را بر اساس پیکسل‌های تصویر تک‌باند انجام دادیم اما راهکار دیگر آن است که پیکسل‌های تصویر چندطیفی را مورد طبقه‌بندی قرار دهیم.

در بخش قبل با تغییر اندازه تصویر چندطیفی به اندازه تصویر تک‌باند، و استفاده از یک بلوک CNN برای ویژگی‌های طیفی توانستیم دقیق تر طبقه‌بندی پیکسل به پیکسل تصویر تک‌باند را افزایش دهیم. اکنون هدف اصلی ما این است که تصویر چندطیفی را طبقه‌بندی نماییم. نکته مهم این است که اگر تصویر تک‌باند را به اندازه تصویر چندطیفی کوچک کنیم اطلاعات مکانی از بین خواهد رفت.

در اینجا روشی ارائه می‌دهیم که با طبقه‌بندی تصویر چندطیفی، نهایت استفاده از تصویر تکباند را نیز داشته باشیم. در قسمت قبل ما برای هر پیکسل از تصویر تکباند یک پنجره 35×35 انتخاب کردیم و به طور متناظر یک پنجره 13×13 برای تصویر چندطیفی که به اندازه تصویر تکباند بزرگ شده بود.

برای طبقه‌بندی تصویر چندطیفی ابتدا نیاز به نقشه مرجع داریم. نقشه مرجعی که در ابعاد تصویر تکباند ساختیم با ضریب 0.25 کوچک می‌کنیم تا به اندازه تصویر چندطیفی شود. همچنین این کار نیاز به کمی تصحیح وهم مرکز کردن برای تصویر چندطیفی دارد. پس از در دسترس قرار گرفتن نقشه مرجع، از همان شبکه قبلی برای طبقه‌بندی استفاده می‌کنیم. با این تفاوت که اطلاعات ورودی کمی متفاوت‌تر از قبل است.

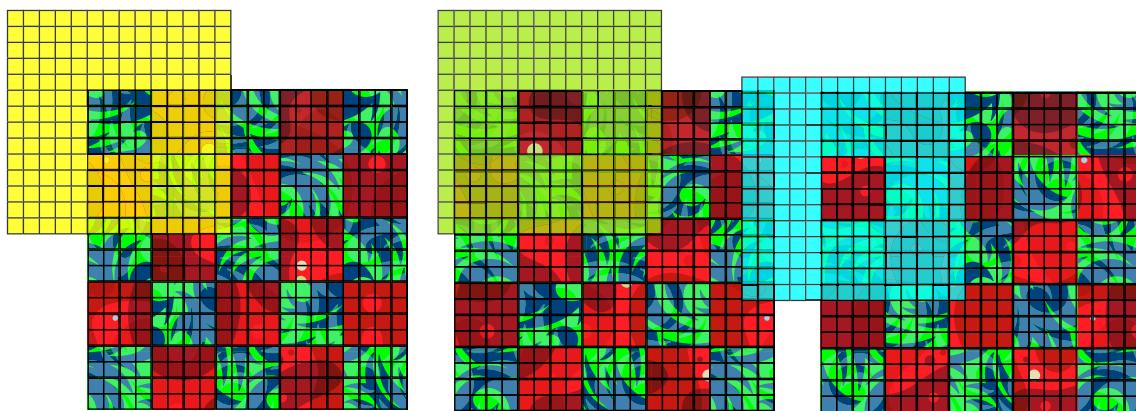
در اینجا برای تصویر چند طیفی (با اندازه واقعی) یک پنجره 3×3 تولید می‌کنیم. و هدف برچسب زدن به پیکسل مرکزی است. از طرف دیگر و برای استفاده از ویژگی‌های مکانی برای هر قسمت متناظر با یک پیکسل تصویر چندطیفی، یک پنجره با اندازه 50×50 تولید می‌کنیم که می‌کنیم که همسایگی برابر 23 پیکسل از هر طرف برای ناحیه مورد نظر دارد. به عبارت دیگر ما برای تعیین کلاس یک پنجره 4×4 یک پنجره با همسایگی 23 تولید می‌کنیم که قصد داریم 16 پیکسل مرکز آن را طبقه‌بندی کنیم. با توجه به شکل ۱۲-۴ ابتدا پیکسل‌های هدف را در هر دو تصویر مشاهده می‌کنیم.



شکل ۱۲-۴: بررسی پیکسل‌های هدف: در تصویر تکباند در مقابل هر پیکسل چندطیفی، 16 پیکسل به عنوان هدف در نظر گرفته می‌شوند.

اکنون باید برای هر مربع 4×4 یک همسایگی انتخاب کنیم، یکی از بهترین نقاط از نظر سرعت و نتایج عدد 23 پیکسل از هر طرف از این مربع 16 پیکسلی است. که با طول خود این مربع، اندازه هر پنجره برابر 50×50 می‌شود. دقت شود این بار حرکت بر روی تصویر تکباند به صورت پیکسل به پیکسل نیست و از هر طرف 4 پیکسل این اتفاق می‌افتد مجدداً در شکل ۱۳-۴ سه تا از این پنجره‌های کنار هم را مشاهده

می‌کنیم. البته در این شکل مثال ما برای همسایگی به اندازه ۵ پیکسل از هر طرف است که با اندازه خود هدف ۱۶ پیکسلی، برابر یک پنجره با طول و عرض ۱۴ می‌شود.



شکل ۱۳-۴: نحوه حرکت پنجره بر روی تصویر. در اینجا ۳ مربع ۱۶ پیکسلی با همسایگی ۵ از هر طرف به عنوان داده متناظر با یک پیکسل از تصویر چندطیفی انتخاب شده‌اند که هدف برچسب زدن به ۱۶ پیکسل مرکزی است.

پیش از نمایش نتایج باید به یکسری اطلاعات در مورد داده توجه کنیم. چنان که گفته شد ما نقشه مرجع تولید شده در فصل ۳ را به اندازه تصویر چندطیفی کوچک می‌کنیم. تعداد داده‌های برچسبدار پس از این کار کاهش می‌یابد که تعداد جدید این داده‌ها در جدول ۹-۴ قابل مشاهده است.

جدول ۹-۴: مشخصات نقشه مرجع تولید پس از هم سایز کردن به تصویر چندطیفی

شماره	نام کلاس	شماره	نام کلاس	تعداد نمونه	نام کلاس	تعداد نمونه	تعداد کلاس	شماره نمونه
۱	خیابان	۶	درخت	۱۱۵۳۸	۶	درخت	۹۱۰۰	
۲	خاک و ماسه	۷	ساختمان	۴۵۸۶	۷	ساختمان	۱۱۵۳۵	
۳	ماشین غیر مشکی	۸	استخر	۱۰۵۷	۸	استخر	۱۶۰۷	
۴	سنگ فرش	۹	چمن	۵۷۲۲۴	۹	چمن	۴۳۰۰	
۵	سایه	۱۰	دریا	۵۸۷۸	۱۰	دریا	۶۲۴۵	
اندازه تصویر				۳۲۷ × ۲۷۵				

پس از توضیح فرم ورودی دیتا باید شکل شبکه را بیان کنیم. شبکه مورد نظر ما دقیقاً همان شبکه مورد استفاده در روش پیشنهادی اصلی است. برای شبکه می‌توان مجدداً به شکل ۸-۴ رجوع کرد. مشاهده جزئیات این شبکه و تغییرات آن در هر بلوک به جداول مندرج در ادامه دقت کنید. در جدول ۱۰-۴ مشخصات بلوک گابور، در جدول ۱۱-۴ مشخصات بلوک مکانی و در جدول ۱۲-۴ مشخصات بلوک طیفی و بقیه شبکه موجود است.

با توجه و دقت به جداول‌های مندرج اطلاعات متفاوت بلوک‌ها را که نسبت به قبل تغییر کرده‌اند مشاهده

جدول ۱۰-۴: اطلاعات کامل بلوک گابور برای طرح جدید

ردیف	نوع لایه	اطلاعات لایه	شکل خروجی
۱	input Image	IM:50×50	-
۲	2D-Conv gabor init	23×23×24 , S=1 , P=0	28×28×24
۳	Batch norm	-	28×28×24
۴	relu	-	28×28×24
۵	2D-Conv	3×3×30 , S=1 , P='same'	28×28×30
۶	Batch norm	-	28×28×30
۷	relu	-	28×28×30
۸	MaxPool	3×3 , S=3 , P=0	9×9×30
۹	2D-Conv	5×5×34 , S=1 , P=0	5×5×34
۱۰	Batch norm	-	5×5×34
۱۱	relu	-	5×5×34
۱۲	MaxPool	2×2 , S=2 , P=1	3×3×34

جدول ۱۱-۴: اطلاعات کامل بلوک مکانی برای طرح جدید

ردیف	نوع لایه	اطلاعات لایه	شکل خروجی
۱	input Image	center crop IM from :50×50	center crop IM to:26×26
۲	2D-Conv	3×3×24 , S=1 , P=0	24×24×24
۳	Batch norm	-	2×24×24
۴	relu	-	24×24×24
۵	MaxPool	2×2 , S=2 , P=0	12×12×24
۶	2D-Conv	5×5×36 , S=1 , P=0	8×8×36
۷	Batch norm	-	8×8×36
۸	relu	-	8×8×36
۹	2D-Conv	6×6×36 , S=1 , P=0	3×3×36
۱۰	Batch norm	-	3×3×36
۱۱	relu	-	3×3×36

می‌کنیم. اکنون که دیتا جدید و شبکه پیشنهادی اصلاح شده با این دیتا را مشاهده کردیم، زمان بررسی نتایج است. پیش از بررسی نتایج باید خاطر نشان کرد این روش بیشتر جنبه پیشنهادی دارد و یک فرایند موثر در افزایش سرعت طبقه‌بندی می‌باشد. و برای بهبود نتایج کارهای زیادی می‌توان انجام داد. در اینجا ما صرفاً این ایده را پیاده سازی نموده‌ایم ولی کار بسیاری برای بهینه کردن نتایج انجام نداده‌ایم.

همچنین مجدداً مثل روش پیشنهادی، برای دو کلاسی که تعداد داده آموزشی کمتری داشتند به وسیله دوران، در ۳ جهت ۹۰، ۱۸۰ و ۳۷۰ درجه داده مصنوعی تولید نموده‌ایم تا تعداد داده‌های آموزشی در هر کلاس برای شبکه تقریباً یکسان باشد. زیرا اگر تعداد داده‌ها اختلاف زیادی داشته باشد شبکه به سمت کلاسی که تعداد داده بیشتری داشته باشد بایاس می‌شود. در جدول ۱۳-۴ مشخصات آموزشی شبکه نیز موجود است.

جدول ۱۲-۴: اطلاعات کامل شبکه برای روش جدید

ردیف	نوع لایه	اطلاعات لایه	شکل خروجی
۱	input PAN Image	IM:50×50	-
۲	Gabor Block	Connect to depth concat	3×3×34
۳	Spatial Block	Connect to depth concat	3×3×36
۴	input MUL Image	IM:7×7×8	-
۵	2D-Conv	3×3×24 , S=1 , P='same'	7×7×24
۶	Batch norm	-	7×7×24
۷	relu	-	7×7×24
۸	2D-Conv	5×5×36 , S=1 , P='same'	7×7×36
۹	Batch norm	-	7×7×36
۱۰	relu	-	7×7×36
۱۱	MaxPool	2×2 , S=2 , P=0	3×3×36
۱۲	Depth concatention	concat 3 Block	3×3×106
۱۳	2D-Conv	3×3×118 , S=1 , P=0	1×1×118
۱۴	Batch norm	-	1×1×118
۱۵	relu	-	1×1×118
۱۶	flatten	-	118 neuron
۱۷	Fully connected	256	256 neuron
۱۸	Batch norm	-	256 neuron
۱۹	relu	-	256 neuron
۲۰	Fully connected	تعداد کلاس‌ها	تعداد کلاس‌ها
۲۱	SoftMax	-	برچسب

جدول ۱۳-۴: اطلاعات مربوط به آموزش شبکه جدید

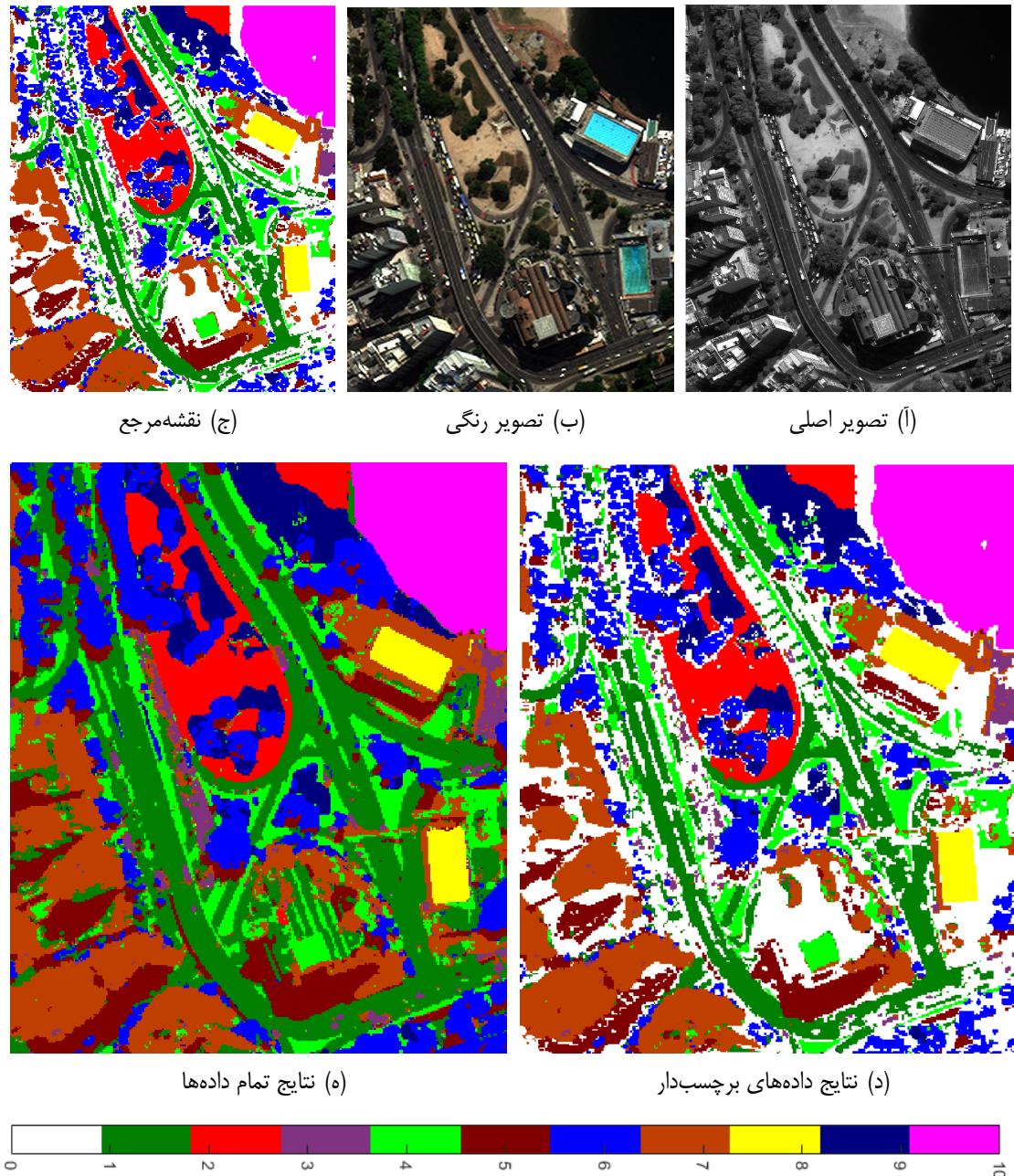
TrainPercent	Validation	Batch size	epochs	Learning rate	Training options
8%	2%	64	25	$0.01 \Rightarrow 10^{-6}$	"SGDM"

اکنون زمان مشاهده نتایج است ضمن بیان این نکته که فرایند آموزش در این حالت ۹۰٪ نسبت به روش پیشنهادی اصلی کاهش می‌یابد. زمان خروجی گرفتن از تمام داده نیز ۹۵٪ درصد کاهش می‌یابد. پس این روش در تسريع فرایند طبقه‌بندی کاملاً موثر بوده است. با دقت به شکل ۱۴-۴ و جدول ۱۴-۴ بررسی می‌کنیم که این افزایش سرعت چه تاثیری در نتایج ما داشته است.

با توجه به نتایج این روش درست است که به یک سرعت بسیار عالی رسیدیم ولی دقت تا ۴ درصد کاهش یافته است. از طرفی در این روش برای دستیابی به چنین دقتی تقریباً ۳ برابر دیتا آموزشی بیشتر استفاده شده است. همچنین با حذف فیلتربانک گابور از ابتدای بلوک گابور، صحت کل به عددی زیر ۹۰ درصد می‌رسد. به هر حال این روش شامل روش پیشنهادی ما نبوده است و در فصل بعد به عنوان پیشنهادی برای علاقه‌مندان به این حوزه بیان می‌شود که ما در اینجا شبیه‌سازی‌های اولیه این طرح و ایده انجام آن را بیان نموده‌ایم.

جدول ۱۴-۴: نتایج طبقه‌بندی با روش جدید

Class	Overall Accuracy=92 , Average Accuracy=92 , Average Validity=92 , $\kappa=91$										
	1	2	3	4	5	6	7	8	9	10	Ac
1	9752	11	33	89	96	282	120	0	2	0	94
2	7	4023	0	29	0	6	6	0	57	0	97
3	7	1	793	31	0	2	118	0	0	0	83
4	112	45	24	4476	0	44	341	1	109	0	87
5	34	0	2	0	4872	263	146	0	1	0	92
6	132	3	6	28	183	7492	105	0	241	0	92
7	235	3	181	317	168	89	9363	11	14	1	90
8	0	0	0	0	2	0	12	1433	0	0	99
9	7	53	3	80	1	324	21	0	3381	0	87
10	0	0	0	0	12	1	4	0	0	5604	100
Val	95	97	76	87	91	88	92	99	89	100	



شکل ۱۴-۴: نتایج تصویری روش جدید

فصل پنجم

جمع‌بندی و پیشنهادات

۱-۵ جمع‌بندی

با گذشت زمان ماهواره‌های سنجش از دور دچار تغییر و تحول شده‌اند. یکی از این تغییرات، افزایش دقت مکانی تصاویر ماهواره‌ها است که در نتیجه آن، اطلاعات مکانی نیز افزایش یافته است و تغییر دیگر، محدود شدن تعداد باندها از چند صد باند (تصاویر ابرطیفی) به چند باند محدود (تصاویر چندطیفی) می‌باشد.

با توجه به این نکته باید طبقه‌بندی این داده‌ها به گونه‌ای صورت پذیرد که از حداقل اطلاعات مکانی در فرایند طبقه‌بندی استفاده شود. مهم‌ترین منبع برای اطلاعات مکانی، تصویر تک باند است و مهم‌ترین منبع برای اطلاعات طیفی، تصویر چندطیفی از ماهواره مورد نظر می‌باشد. فرایند طبقه‌بندی باید به گونه‌ای باشد که هم از حداقل ویژگی‌های مکانی استفاده شود و هم از حداقل ویژگی‌های طیفی، بنابراین در طبقه‌بندی این تصاویر باید از هر دو نوع تصویر استفاده شود.

یکی از مشکلات طبقه‌بندی تصاویر چندطیفی عدم وجود یک داده با نقشه مرجع قابل اطمینان است. این مشکل باعث شده است تحقیق در زمینه طبقه‌بندی تصاویر چندطیفی دشوار شود. می‌توان برای حل مشکل نبود نقشه مرجع، از روش‌های بدون نظارت یا نیمه نظارت شده استفاده کرد. یادگیری تقویتی یکی از روش‌هایی است که می‌توان برای تولید نقشه مرجع به کار برد شود.

طبقه‌بندی بر اساس پیکسل‌های تصویر تک باند فرایند زمان گیری است و الزامی در چنین دقت بالایی برای طبقه‌بندی نیست ولی دقت طبقه‌بندی در صورت طبقه‌بندی پیکسل‌های تک باند بیشتر است. با طبقه‌بندی تصویر چندطیفی می‌توان به سرعت بیشتر و مقیاس طبقه‌بندی قابل قبولی دست یافت.

تصاویر چندطیفی ماهواره‌ها ارزان قیمت هستند و برای یک محل خاص، تعداد زیادی از این تصاویر در دوره‌های زمانی متفاوت ثبت می‌شود. همچنین ناحیه تحت پوشش ماهواره‌ها وسیع‌تر از پهپاد است و به همین دلیل می‌توان منطقه وسیع‌تری را مورد طبقه‌بندی قرار داد.

چون دقت مکانی تصاویر چندطیفی ماهواره‌ای به اندازه تصاویر چندطیفی اخذ شده با پهپاد نیست و از طرفی دقت طیفی این تصاویر هم به اندازه تصاویر ابرطیفی نیست، طبقه‌بندی و کار کردن با این تصاویر بسیار دشوار است. از این رو رفتار طبقه‌بندی با این تصاویر متفاوت از دو تصویر دیگر بوده است و به صورت ناحیه‌ای طبقه‌بندی صورت می‌گیرد. با روشی که در پیش گرفتیم توانستیم با به کارگیری هر دو اطلاعات طیفی و مکانی روش طبقه‌بندی مرسوم در تصاویر ابرطیفی را بر روی این تصاویر انجام دهیم.

۲-۵ پیشنهاد

استفاده از تصاویر ماهواره‌ایی که تصویر با نقشه مرجع معتبر دارن فرایند کار را تسریع می‌بخشد و روش‌های پیشنهادی سریع‌تر شکل می‌گیرد. همچنین قابلیت مقایسه عادلانه روش پیشنهادی با سایر روش‌ها به وجود می‌آید. یکی از پیشنهادهای خوب برای طبقه‌بندی تصاویر چندطیفی، طبقه‌بندی تصویر چندطیفی ماهواره به جای تصویر تک‌باند است. برای این کار شبیه سازی اولیه در فصل ۴ ام شرح داده شده است. و نیاز بالا بردن دقیق به وسیله استفاده بهینه‌تر از ویژگی‌های موجود در دو تصویر است.

یک راه دیگر برای طبقه‌بندی این تصاویر استخراج توام ویژگی‌های مکانی می‌باشد که با توجه به روش‌های گوناگون طبقه‌بندی می‌تواند روش مناسبی برای کار با این تصاویر باشد.

مراجع و مراجع

- [1] R. Schowengerdt. *Remote Sensing: Models and Methods for Image Processing*. Academic Press, 2007.
- [2] J. Schott. *Remote Sensing: The Image Chain Approach*. Oxford University Press, USA, 2007.
- [3] Maksel and Rebecca, “Flight of the giant,” *Air and Space Magazine*, Aug 2021.
- [4] H. Guo, Q. Huang, X. Li, Z. Sun, and Y. Zhang, “Spatiotemporal analysis of urban environment based on the vegetation–impervious surface–soil model,” *Journal of Applied Remote Sensing*, vol.8, no.1, p.084597, 2013.
- [5] R. Brumbaugh, “Principles of remote sensing,” may 2017. publisher: Reef Resilience Network.
- [6] N. Hagen and M. W. Kudenov, “Review of snapshot spectral imaging technologies,” *Optical Engineering*, vol.52, no.9, pp.090901–090901, 2013.
- [7] P. Lynch, L. Blesius, and E. Hines, “Classification of urban area using multispectral indices for urban planning,” *Remote Sensing*, vol.12, no.15, p.2503, 2020.
- [8] J. Pike, “Intelligence - hyperspectral imaging,” Jul 2011, publisher: Global Security.
- [9] I. Igor, “Hyperspectral imaging in agriculture: Opportunities, benefits and future perspectives,” Aug 2020, publisher: Gamaya blog.
- [10] J. BRODERSEN, “Multispectral lighting: A practical option for difficult industrial imaging situations,” may 2017. publisher: photonics.com.
- [11] S. Zhou, L. Sun, W. Xing, G. Feng, Y. Ji, J. Yang, and S. Liu, “Hyperspectral imaging of beet seed germination prediction,” *Infrared Physics & Technology*, vol.108, p.103363, 2020.
- [12] D. C. Santana, M. C. M. Teixeira Filho, M. R. da Silva, P. H. M. d. Chagas, J. L. G. de Oliveira, F. H. R. Baio, C. N. S. Campos, L. P. R. Teodoro, C. A. da Silva Junior, P. E. Teodoro, *et al.*, “Machine learning in the classification of soybean genotypes for primary macronutrients’ content using uav–multispectral sensor,” *Remote Sensing*, vol.15, no.5, p.1457, 2023.

-
- [13] H. Xu, W. He, L. Zhang, and H. Zhang, “Unsupervised spectral–spatial semantic feature learning for hyperspectral image classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol.60, pp.1–14, 2022.
 - [14] F. Melgani, B. A. Al Hashemy, and S. M. Taha, “An explicit fuzzy supervised classification method for multispectral remote sensing images,” *IEEE Transactions on Geoscience and Remote Sensing*, vol.38, no.1, pp.287–295, 2000.
 - [15] S. N. Blythe and J. H. Krantz, “A mathematical model of retinal receptive fields capable of form & color analysis,” *Impulse: The Premier Journal for Undergraduate Publications in the Neurosciences*, vol.1, no.1, pp.1–58, 2004.
 - [16] H. Kolb, “How the retina works: Much of the construction of an image takes place in the retina itself through the use of specialized neural circuits,” *American scientist*, vol.91, no.1, pp.28–35, 2003.
 - [17] F. Gaillard, M. A. Arbib, F. J. Corbacho, and H. B. Lee, “Modeling the physiological responses of anuran r3 ganglion cells,” *Vision research*, vol.38, no.17, pp.2551–2568, 1998.
 - [18] R. M. S. B. Jose Antonio Henriquez Germano, “Artificial retina:development of a bio-inspired model with configurable implementation,” 2003.
 - [19] M. Puoèiauskas, “Artificial neural network modeling of visual system,” 2003.
 - [20] J. Walinga. *Introduction to Psychology: 1st Canadian Edition*. Online access: Center for Open Education Open Textbook Library, BCcampus, 2010.
 - [21] W.-C. A. Lee, H. Huang, G. Feng, J. R. Sanes, E. N. Brown, P. T. So, and E. Nedivi, “Dynamic remodeling of dendritic arbors in gabaergic interneurons of adult visual cortex,” *PLoS biology*, vol.4, no.2, p.e29, 2006.
 - [22] S. Cajal. *Comparative Study of the Sensory Areas of the Human Cortex*. Clark University, 1899.
 - [23] B. Jia, W. Feng, and M. Zhu, “Obstacle detection in single images with deep neural networks,” *Signal, Image and Video Processing*, vol.10, pp.1033–1040, 2016.
 - [24] S. Herculano-Houzel and R. Lent, “Isotropic fractionator: a simple, rapid method for the quantification of total cell and neuron numbers in the brain,” *Journal of Neuroscience*, vol.25, no.10, pp.2518–2521, 2005.
 - [25] P. Ghamisi, J. A. Benediktsson, and J. R. Sveinsson, “Automatic spectral–spatial classification framework based on attribute profiles and supervised feature extraction,” *IEEE Transactions on Geoscience and Remote Sensing*, vol.52, no.9, pp.5771–5782, 2013.
 - [26] D. H. Hubel and T. N. Wiesel, “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex,” *The Journal of physiology*, vol.160, no.1, p.106, 1962.

- [27] D. Berga Garreta. *Understanding eye movements: Psychophysics and a model of primary visual cortex*. 2019.
- [28] K. Fukushima, “Recent advances in the deep cnn neocognitron,” *Nonlinear Theory and Its Applications, IEICE*, vol.10, no.4, pp.304–321, 2019.
- [29] G. W. Lindsay, “Convolutional neural networks as a model of the visual system: Past, present, and future,” *Journal of cognitive neuroscience*, vol.33, no.10, pp.2017–2031, 2021.
- [30] L. Thulin Nilsson, “The role of primary visual cortex in visual awareness,” 2015.
- [31] J. J. DiCarlo and D. D. Cox, “Untangling invariant object recognition,” *Trends in cognitive sciences*, vol.11, no.8, pp.333–341, 2007.
- [32] S. Chung, D. D. Lee, and H. Sompolinsky, “Classification and geometry of general perceptual manifolds,” *Physical Review X*, vol.8, no.3, p.031003, 2018.
- [33] U. Cohen, S. Chung, D. D. Lee, and H. Sompolinsky, “Separability and geometry of object manifolds in deep neural networks,” *Nature communications*, vol.11, no.1, p.746, 2020.
- [34] I. Goodfellow, Y. Bengio, and A. Courville. *Deep learning*. MIT press, 2016.
- [35] S. Kim, “A beginner’s guide to convolutional neural networks (cnns),” *towards data science*, 2019.
- [36] H. Yingge, I. Ali, and K.-Y. Lee, “Deep neural networks on chip-a survey,” in *2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp.589–592, IEEE, 2020.
- [37] B. A. Olshausen and D. J. Field, “Emergence of simple-cell receptive field properties by learning a sparse code for natural images,” *Nature*, vol.381, no.6583, pp.607–609, 1996.
- [38] R. E. Soans, D. C. Lim, B. T. Keenan, A. I. Pack, and J. A. Shackleford, “Automated protein localization of blood brain barrier vasculature in brightfield ihc images,” *PloS one*, vol.11, no.2, p.e0148411, 2016.
- [٣٩] احمد، طاهری، ”استفاده از ویژگی های طیفی-مکانی استخراج شده از فیلترهای گابور سه بعدی به منظور طبقه بندی تصاویر ابر طیفی،“ پایان نامه کارشناسی ارشد، دانشکده برق و کامپیوتر، دانشگاه دانشگاه تربیت مدرس، ۱۳۹۷.
- [40] S. Narkhede, “Understanding auc-roc curve,” *Towards Data Science*, vol.26, no.1, pp.220–227, 2018.
- [41] Z. H. Hoo, J. Candlish, and D. Teare, “What is an roc curve?,” 2017.
- [42] U. B. Gewali, S. T. Monteiro, and E. Saber, “Machine learning based hyperspectral image analysis: a survey,” *arXiv preprint arXiv:1802.08701*, 2018.

- [۴۳] محمد ابراهیم عقیلی، ”طراحی شبکه‌های عمیق کانوالوشنی به منظور ادغام ویژگی‌های طیفی مکانی و طبقه‌بندی تصاویر سنجش از دور ایران“، پایان‌نامه کارشناسی ارشد، دانشکده برق و کامپیوتر، دانشگاه دانشگاه تربیت‌مدرس، ۱۴۰۱.
- [44] M. Ghassemi, H. Ghassemian, and M. Imani, “Hyperspectral image classification by optimizing convolutional neural networks based on information theory and 3d-gabor filters,” *International Journal of Remote Sensing*, vol.42, no.11, pp.4380–4410, 2021.
- [۴۵] شیریان حسین‌آبادی حمید، ”ادغام ویژگی‌های طیفی - مکانی تصاویر ابرطیفی با استفاده از فیلتر گابور در شبکه CNN“، پایان‌نامه کارشناسی ارشد، دانشکده برق و کامپیوتر، دانشگاه دانشگاه تربیت‌مدرس، ۱۳۹۸.

واژه‌نامه‌ی انگلیسی به فارسی

Adobe Photoshop	فتوشاپ	سنسور فعال
Artificial Neural Network	شبکه عصبی مصنوعی	آماکرین
Axon	اکسون	ادغام کننده میانگین
Bias	ثابت	انتشار به عقب
Cell Body	بخش اصلی سلول	شبکه عصبی زیستس
Confusion Matrix	ماتریس درهم‌ریختگی	پیچیده
Cornea	قرنیه	شبکه عصبی کانوالوشنی
Dendrites	دندریت‌ها	دندریتی
Feed-Forward	تعذیب از جلو	خطای انتشار به عقب
Fova	حفره چشم	انتشار به جلو
Ganglion Cell	سلول غده‌ای	فعال ساز
Horizontal Cells	سلول افقی	صف کردن
Inner Nuclear Layer	لایه هسته داخلی	ابرهیفی
Iris	عنیبه	لایه پلکسی داخلی
Learning Rate	نرخ یادگیری	کرنل
Loss Function	تابع هزینه	عدسی
Mexican Hat	کلاه مکزینکی	ادغام بیشینه
Multispectral	چندطیفی	پرسپکترون چندلایه
Neural Network	عصبی	نوادراتی
Normalize	بهنجارش	نورون
Panchromatic	تکباند	لایه پلکسی خارجی
Perceptrons	نورون مصنوعی	سنسور غیرفعال
Pupil	مردمک	سلول گیرنده
Rectangular	مربع	پرتوسنجی
Reinfircement Learning	یادگیری تقویتی	تابع خطی اصلاح شده
Retina	شبکیه	سنجهش از دور
Smooth	نرم	ساده
Spectral	طیفی	مکانی
Temporal	زمانی	سیناپسی
		بده-بستان

Abstract

Modern remote sensing sensors take images of the earth's surface in different electromagnetic spectral bands with high spatial accuracy. These images contain spectral and spatial information of the scene. Spatial information includes the shape, location and type of texture and geometric structure of objects in the scene. Spectral information represents the quality and type of materials of the objects in the scene. Recent remote sensing satellites in creating multi-spectral images with high spatial accuracy require the simultaneous use of spectral and spatial information in the classification of remote sensing images. For example The WorldView3 satellite has 8 spectral bands with a spatial accuracy of about 1.2 meters and a PAN image with an accuracy of 30 cm. The main issue of this research is the need to provide a low-cost method that can detect land use in low-cost remote sensing systems with high process accuracy and reliability. The second problem is the lack of a reference map and the limitation of training samples with correct and acceptable stickers and the high cost of preparing such samples, which causes the classification process to face the limitation of training samples. The purpose of this research is to use the combined design of the filter bank and classifier inspired by the human vision system for feature extraction and classification of land covers with the limitation of training samples. CNN networks have gained a lot of fame due to their similarity to the human visual system, these networks have recently received more attention despite the strong theory as a support, which is one of the reasons for the advancement of computer hardware related to performing CNN network processing. In recent years, these networks have shown that they have great power in classifying remote sensing images, but unfortunately, one of the problems of these networks is there need for high training data, one of the reason for that, is the start of the training process from random points for the weights inside the network. A conventional method to solve this problem is to use feature extraction methods before the network and then enter the features into the network. In this thesis, to solve the need of high training data problem, tried to determine the weights as far as possible not to start from random points, and for this purpose, it is inspired by the human vision model, in addition to increasing the speed of training and maintaining the accuracy and validity of the classification. the number of training examples required by the network will be reduced.

Key Words: Remote sensing, multispectral images, convolutional network, vision structure, eye model, Gabor filterbank, reference map.



Tarbiat Modares University

**Submitted in Partial Fulfillment of the Requirements for the Master of Science
(M.S.) Degree in Communications Systems Engineering
Faculty of Electrical and Computer Engineering Tarbiat Modares University**

**Design Gabor based CNN inspired by human vision system for
classification of remote sensing images with limited training samples**

By

Alireza Dori

Supervisor

Dr. Mohammad Hassan Ghassemlan

Advisor

Dr. Maryam Imani

August 2023