



دانشگاه شهید بهشتی
دانشکده علوم ریاضی
گروه ریاضی کاربردی
گرایش علوم داده‌ها

پایان‌نامه کارشناسی ارشد

عنوان

سیستم استخراج نقشه‌ی خیابان با استفاده از پردازش تصویر بر روی نمای ماهواره‌ای

نگارش
محمد جواد عباس‌پور

استاد راهنما
دکتر حسین حاجی ابوالحسن

استاد مشاور
دکتر سعید رضا خردپیشه

بنامه شهزاده خان و

کلیه حقوق اعم از چاپ و تکثیر، نسخه برداری، ترجمه، اقتباس و ... از این پایان‌نامه برای دانشگاه شهید بهشتی محفوظ است.
نقل مطالب با ذکر مأخذ آزاد است.

تعهد نامه اصالت اثر

باسمہ تعالیٰ

اینجانب، محمدجواد عباسپور به شماره دانشجویی رشته ریاضی کاربردی، ۴۰۰۴۲۲۲۵ دانشجویی گرایش علوم داده‌ها مقطع تحصیلی کارشناسی ارشد متعهد می‌شوم که، مطالب مندرج در این پایان‌نامه حاصل کار پژوهشی اینجانب، تحت نظارت و راهنمایی استاد دانشگاه شهید بهشتی بوده و به دستاوردهای دیگران که در این پژوهش از آن‌ها استفاده شده است، مطابق مقررات و روال متعارف ارجاع و در فهرست منابع و مأخذ ذکر گردیده است.

این پایان‌نامه قبلًا برای احراز هیچ مدرک هم‌سطح یا بالاتر ارائه نگردیده است.

در صورت اثبات تخلف در هر زمان، مدرک تحصیلی صادر شده توسط دانشگاه از درجه اعتبار ساقط بوده و دانشگاه حق پیگیری قانونی خواهد داشت.

کلیه نتایج و حقوق حاصل از این پایان‌نامه، متعلق به دانشگاه شهید بهشتی می‌باشد. هرگونه استفاده از نتایج علمی و عملی، واگذاری اطلاعات به دیگران یا چاپ و تکثیر، نسخه‌برداری، ترجمه و اقتباس از این پایان‌نامه، بدون موافقت کتبی دانشگاه شهید بهشتی ممنوع است.

نقل مطالب با ذکر مأخذ بلامانع است.

محمدجواد عباسپور

.....
امضا:

مجوز بهره‌برداری از پایان‌نامه

بهره‌برداری از این پایان‌نامه در چارچوب مقررات کتابخانه و با توجه به محدودیتی که توسط استاد راهنمای شرح زیر تعیین می‌شود، بلامانع است:

- بجهه‌برداری از این پایان‌نامه برای همگان بلامانع است.
- بجهه‌برداری از این پایان‌نامه با اخذ مجوز از استاد راهنمای، بلامانع است.
- بجهه‌برداری از این پایان‌نامه تا تاریخ ممنوع است.

تاریخ:

استاد راهنمای آقای دکتر حاجی ابوالحسن

امضا:

تقدیم به:

مادر عزیز

و

پدر مهربانم

سپاس‌گزاری...

با سپاس فراوان از استادان گرانقدر، دکتر سعیدرضا خردپیشه، دکتر حاجی ابوالحسن، که با راهنمایی‌ها و حمایت‌های بی‌دریغ خود، مرا در مسیر نگارش این پایان‌نامه همراهی کردند.

بدون تردید، زحمات و دانش ارزشمند ایشان نقش بسزایی در پیشبرد این پژوهش داشت و موقفیت‌های به‌دست‌آمده مرهون تلاش‌های بی‌وقفه و حمایت‌های همیشگی آنان است.

از صمیم قلب از تمامی وقتی که برای هدایت و راهنمایی من اختصاص دادند، سپاس‌گزارم و همواره
قدرتان محبت‌ها و آموزه‌های ارزشمند ایشان خواهم بود.

همچنین از پدر و مادر عزیزم که با صبر و حمایت بی‌پایان خود، همیشه پشتیبان من بوده‌اند، عمیقاً
سپاس‌گزارم.

آنان با صبوری، تشویق و محبت‌های بی‌دریغ خود، نقش مهمی در رسیدن من به این مرحله از زندگی
داشته‌اند. بدون همراهی و حمایت بی‌نظیر خانواده‌ام، طی این مسیر ممکن نبود.

محمدجواد عباس‌پور

۱۴۰۳

نام: محمدجواد

نام خانوادگی دانشجو: عباسپور

عنوان: سیستم استخراج نقشه‌ی خیابان با استفاده از پردازش تصویر بر روی نمای ماهواره‌ای

استاد راهنما: دکتر حسین حاجی‌ابوالحسن

استاد مشاور: دکتر سعیدرضا خردپیشه

گرایش: علوم داده‌ها

رشته: ریاضی کاربردی

مقطع تحصیلی: کارشناسی ارشد

علوم ریاضی

تعداد صفحات: ۱۱۳

دانشگاه: شهید بهشتی

تاریخ فارغ‌التحصیلی: ۱۴۰۳

واژگان کلیدی: شبکه عصبی پیچشی، بخش‌بندی تصویر، تصویر سنجش از دور، پردازش تصویر، استخراج خیابان، ترانسفورمر چرخشی

چکیده

این پایان‌نامه به بررسی روش‌های مختلف برای استخراج نقشه‌های خیابانی از تصاویر ماهواره‌ای با استفاده از شبکه‌های عصبی کانولوشنی و سایر روش‌های یادگیری عمیق می‌پردازد.

مدل‌های عمیق مانند شبکه‌های عصبی کانولوشنی و ترانسفورمرهای چرخشی برای استخراج دقیق خیابان‌ها در تصاویر با وضوح و کنترast پایین مورد استفاده قرار گرفتند.

از جمله مشکلات اصلی در این حوزه، دقت پایین در تصاویر با کنترast پایین است که روش‌های سنتی ریاضی مانند مورفولوژی قادر به حل آن نیستند. به همین دلیل، یادگیری عمیق و مدل‌هایی مانند ترانسفورمرهای چرخشی، که توانایی حفظ جزئیات تصاویر را دارند، پیشنهاد می‌شود. ترکیب شبکه‌های عصبی کانولوشنی با ترانسفورمرهای چرخشی نتایج بسیار خوبی را در بهبود دقت و شناسایی ویژگی‌های خیابان‌ها نشان داد.

این روش‌ها می‌توانند در کاربردهایی مانند نقشه‌برداری شهری، مسیریابی خودکار، پایش تغییرات زیست‌محیطی، سیستم‌های رانندگی خودکار، و همچنین توسط ارگان‌های امداد و نجات مورد استفاده قرار گیرند.

فهرست مطالب

۱	مقدمه
۱	۱ مقدمه‌ای بر پردازش تصویر
۱	۱.۱ مقدمه
۱	۲.۱ تاریخچه‌ی پردازش تصویر
۲	۳.۱ تکنیک‌های پیش‌پردازش
۴	۴.۱ مبانی پردازش تصویر
۴	۴.۱.۱ تعریف تصویر دیجیتال
۴	۲.۴.۱ مفاهیم پایه‌ای پردازش تصویر
۶	۳.۴.۱ انواع نویز در تصاویر و روش‌های حذف آنها
۷	۵.۱ یادگیری ماشین
۷	۱.۵.۱ تاریخچه و تعریف یادگیری ماشین
۸	۲.۵.۱ مبانی یادگیری ماشین
۸	۳.۵.۱ انواع الگوریتم‌های یادگیری ماشین
۸	۴.۵.۱ یادگیری ناظارت شده
۱۱	۵.۵.۱ یادگیری بدون ناظارت
۱۲	۶.۵.۱ یادگیری تقویتی
۱۳	۷.۵.۱ نتیجه‌گیری و کاربردها در استخراج نقشه خیابان از تصاویر ماهواره‌ای
۱۴	۸.۵.۱ نحوه آموزش مدل‌های یادگیری ماشین
۱۵	۶.۱ یادگیری عمیق
۱۵	۱.۶.۱ تفاوت یادگیری ماشین و یادگیری عمیق

۱۶	شبکه‌های عصبی پیچشی	۲.۶.۱
۱۸	شبکه‌های عصبی بازگشتی و حافظه طولانی کوتاه‌مدت	۳.۶.۱
۱۹	آموزش شبکه‌های عصبی عمیق	۴.۶.۱
۲۰	مدل‌های ترانسفورمر در یادگیری عمیق	۷.۱
۲۰	ساختار کلی مدل‌های ترانسفورمر	۱.۷.۱
۲۱	مدل ترانسفور چرخشی	۲.۷.۱
۲۲	کاربرد ترانسفورمرها در بینایی کامپیوتوئی	۳.۷.۱
۲۳	استخراج ویژگی‌ها و کاربرد آن در پردازش تصویر	۸.۱
۲۳	استخراج ویژگی‌های کلاسیک	۱.۸.۱
۲۴	ویژگی‌های سطح بالا با استفاده از شبکه‌های عصبی	۲.۸.۱
۲۴	روش‌های کاهش ابعاد	۳.۸.۱
۲۶	پیش‌پردازش تصاویر ماهواره‌ای	۹.۱
۲۶	مقدمه‌ای بر تصاویر ماهواره‌ای	۱.۹.۱
۲۶	مراحل پیش‌پردازش تصاویر ماهواره‌ای	۲.۹.۱
۲۷	روش‌های افزایش وضوح تصاویر ماهواره‌ای	۳.۹.۱
۲۸	بیش‌پردازش	۱۰.۱
۲۸	تعریف و مفهوم بیش‌پردازش	۱.۱۰.۱
۲۹	راهکارهای جلوگیری از بیش‌پردازش	۲.۱۰.۱
۳۰	تأثیر بیش‌پردازش در پروژه‌های پردازش تصویر	۳.۱۰.۱
۳۱	بهینه‌سازی فرآپارامترها	۱۱.۱
۳۱	روش‌های بهینه‌سازی هایپرپارامترها	۱.۱۱.۱
۳۲	کاربرد در استخراج جاده از تصاویر ماهواره‌ای	۲.۱۱.۱
۳۳	نتیجه‌گیری	۳.۱۱.۱
۳۳	خلاصه و جمع‌بندی	۱۲.۱
۳۳	مروری بر مبانی پردازش تصویر	۱.۱۲.۱
۳۳	یادگیری ماشین و الگوریتم‌های آن	۲.۱۲.۱
۳۳	یادگیری عمیق و شبکه‌های عصبی	۳.۱۲.۱

۳۴	۴.۱۲.۱	مدل‌های ترانسفورمر در یادگیری عمیق
۳۴	۵.۱۲.۱	استخراج ویژگی‌ها و کاهش ابعاد
۳۴	۶.۱۲.۱	پیش‌پردازش تصاویر ماهواره‌ای
۳۴	۷.۱۲.۱	بیش‌برازش و راهکارهای مقابله با آن
۳۴	۸.۱۲.۱	نتیجه‌گیری کلی
۳۵	۲	پیش‌نیازها و مفاهیم تخصصی
۳۵	۱.۲	شبکه‌های عصبی پیچشی
۳۵	۱.۱.۲	ساختار لایه‌های کانولوشن
۳۸	۲.۱.۲	لایه‌های تجمعی
۴۰	۳.۱.۲	لایه‌های تماماً متصل
۴۰	۴.۱.۲	کاربرد شبکه‌های عصبی پیچشی در استخراج جاده از تصاویر ماهواره‌ای
۴۱	۵.۱.۲	شبکه‌های عصبی پیچشی برای استخراج مسیر از تصاویر ماهواره‌ای
۴۴	۲.۲	الگوریتم شبکه‌های باقی‌مانده
۴۴	۱.۲.۲	چالش شبکه‌های عمیق
۴۴	۲.۲.۲	ایده اصلی
۴۸	۳.۲.۲	ارتباط شبکه‌های باقی‌مانده با استخراج جاده از تصاویر ماهواره‌ای
۴۹	۴.۲.۲	نتیجه‌گیری
۴۹	۳.۲	روش‌های استخراج خیابان از نقشه‌های ماهواره‌ای
۴۹	۱.۳.۲	روش‌های سنتی استخراج خیابان
۵۰	۲.۳.۲	روش‌های مبتنی بر تجزیه و تحلیل ساختاری
۵۰	۳.۳.۲	روش‌های مبتنی بر یادگیری ماشین
۵۰	۴.۳.۲	روش‌های مبتنی بر یادگیری عمیق
۵۰	۵.۳.۲	استفاده از مکانیزم توجه
۵۱	۶.۳.۲	ترکیب روش‌های مختلف
۵۲	۴.۲	ترانسفور چرخشی
۵۲	۱.۴.۲	معماری و ساختار ترانسفور چرخشی
۵۳	۲.۴.۲	آموزش مدل ترانسسورماتور چرخشی
۵۳	۳.۴.۲	مزیت‌های ترانسفور چرخشی

۵۴	۳ پیشینه‌ی پژوهش
۵۴	۱.۳ روش‌های سنتی
۵۴	۲.۳ روش‌های مبتنی بر شبکه‌های عصبی پیچشی
۵۶	۳.۳ روش‌های مبتنی بر ترانسفور
۵۷	۴.۳ ترکیب شبکه‌های عصبی پیچشی و ترانسفور
۵۸	۴ طراحی و ارزیابی مدل جاده‌ساز
۵۹	۱.۴ طراحی کلی مدل جاده‌ساز
۶۰	۲.۴ رمزگذار: پردازش و استخراج ویژگی‌ها
۶۲	۳.۴ گلوگاه: پالایش ویژگی‌ها
۶۳	۴.۴ رمزگشا و تابع هزینه
۶۵	۵.۴ آزمایش‌ها و ارزیابی عملکرد
۶۶	۱.۵.۴ مجموعه‌داده‌ها
۶۷	۲.۵.۴ پیاده‌سازی آزمایش
۶۸	۳.۵.۴ معیارهای ارزیابی
۶۸	۶.۴ آزمایش‌های حذفی
۷۱	۷.۴ آزمایش‌های مقایسه‌ای
۷۱	۱.۷.۴ آزمایش‌ها بر روی مجموعه داده گلوب عمیق
۷۳	۲.۷.۴ آزمایش‌ها بر روی مجموعه داده ماساچوست
۷۶	۸.۴ نتیجه‌گیری
۷۶	۹.۴ جمع‌بندی
۷۸	۵ نتیجه‌گیری
۷۸	۱.۵ دستاوردهای اصلی
۷۸	۲.۵ مدل‌نهایی
۷۹	۱.۲.۵ ساختار کلی
۸۰	۲.۲.۵ ساختار اتصال تقویت‌شده با ویژگی
۸۱	۳.۲.۵ ساختار اتصال تقویت‌شده با شکل

۸۲	تابع هزینه	۳.۵
۸۳	متريک دقت کلي	۱.۳.۵
۸۴	بررسی مجموعه‌داده‌ی ماساچوست	۴.۵
۸۴	نتایج مربوط به ماساچوست	۱.۴.۵
۸۵	جمع‌بندی	۵.۵
۸۵	محدودیت‌ها	۶.۵
۸۶	پیشنهادات برای کارهای آینده	۷.۵
۸۶	استفاده از الگوریتم‌های گسترش‌یافته برای پیش‌بینی بهتر جاده‌ها	۱.۷.۵
۸۷	استفاده از متريک‌های گرافی برای بهبود دقت مدل	۲.۷.۵
۸۸	استفاده از الگوریتم اسکلت‌سازی برای ترسیم گراف جاده‌ها	۳.۷.۵
۸۹	چالش‌های استخراج جاده از تصاویر هوایی با استفاده از شبکه‌های عصبی گرافی	۴.۷.۵
۹۲	جمع‌بندی و پیشنهادات نهايى	۵.۷.۵
۹۲	نتيجه‌گيرى نهايى	۸.۵
۹۳	واژه‌نامه انگلیسي به فارسي	
۹۹	واژه‌نامه فارسي به انگلیسي	
۱۰۵	مراجع	

فهرست تصاویر

۱.۱	پردازش تصویر در حوزه‌ی پزشکی	۲
۲.۱	ساخت یک هیستوگرام رنگی برای هر بخش از تصویر مشاهده می‌شود.	۳
۳.۱	مثالی از اهمیت آشکارسازی لبه‌ها در پردازش تصویر	۵
۴.۱	مثالی از نویز و فیلترهای حذف نویز.	۶
۵.۱	نمایی از الگوریتم پرسپترون	۷
۶.۱	نمایی از الگوریتم ماشین بردار پشتیبان	۹
۷.۱	روش تحلیل مولفه‌های اصلی	۱۱
۸.۱	مثالی از خوشبندی	۱۲
۹.۱	مثالی از یادگیری تقویتی	۱۳
۱۰.۱	شباهت و تفاوت یادگیری ماشین و یادگیری عمیق	۱۵
۱۱.۱	شبکه‌ی عصبی پیچشی	۱۷
۱۲.۱	معماری شبکه‌ی عصبی حافظه طولانی کوتاه‌مدت	۱۹
۱۳.۱	ساختار مدل ترانسفورمر	۲۱
۱۴.۱	نمای کلی از معماری مدل ترانسفورم چرخشی.	۲۲
۱۵.۱	کاهش ویژگی‌ها سبب سرعت و آموزش بهتر مدل می‌شود.	۲۴
۱۶.۱	بهبود وضوح تصویر با استفاده از تکنیک Pan-sharpening	۲۷
۱۷.۱	نمودار نشان‌دهنده عملکرد مدل در حالت بیش‌برازش و حالت بهینه.	۳۰
۱۸.۱	با استفاده از قطع کردن تابع $(x)^c$ با تابع هدف محدوده‌ی نقاط کمینه به دست می‌آید.	۳۲
۱.۲	نمونه‌ای از فیلترهای مختلف و تاثیر آن‌ها	۳۶
۲.۲	نمونه‌ای از فیلترهای مختلف و تاثیر آن‌ها	۳۷

۳۸	مثال ساختار شبکه‌های عصبی پیچشی	۳.۲
۳۹		۴.۲
۴۲	ساختار معماری مدل U-Net	۵.۲
۴۳	ساختار معماری مدل شبکه عصبی تماماً متصل	۶.۲
۴۳	ساختار معماری مدل DeepLab	۷.۲
۴۷	ساختار معماری مدل شبکه‌های عصبی باقی‌مانده	۸.۲
۴۹		۹.۲
۵۱	mekanizm توجه در برابر باقی روش‌ها	۱۰.۲
۵۱	معماری مکانیزم توجه	۱۱.۲
۵۵	پردازش تصویر با استفاده از ماشین بردار پشتیبان	۱.۳
۵۶	معماری مدل شبکه عصبی تماماً متصل برای تشخیص مسیر	۲.۳
۶۰	معماری جاده‌ساز از یک رمزگذار، گلوگاه و رمزگشا تشکیل شده است. نمایش ویژگی چند مقیاسی می‌تواند توسط رمزگذار تولید شود. نقشه‌های ویژگی با ابعاد بالا را می‌توان با کانولوشن قابل جداسازی و بلوک متسع در گلوگاه به دست آورد. نتایج نهایی توسط رمزگشا ارائه می‌شود.	۱.۴
۶۱	معماری بلوک‌های ترنسفور چرخشی	۲.۴
۶۲	ساختار کانولوشن قابل جداسازی موازی. کانولوشن قابل تفکیک، ویژگی‌ها را به ویژگی‌های کanal و ویژگی‌های فضایی تقسیم می‌کند.	۳.۴
۶۳	ساختار بلوک متسع. میدان‌های گیرنده چند مقیاسی از طریق لایه‌های کانولوشنی متسع ساخته می‌شوند، که باعث می‌شود شبکه بتواند ویژگی‌هایی با مقیاس‌های مختلف را استخراج کند.	۴.۴
۶۴	معماری بلوک رمزگشا	۵.۴
۶۷	چند نمونه از مجموعه داده‌های گلوب عمیق در دو سطر اول و ماساچوست در دو سطر آخر آورده شده است.	۶.۴
۷۰	نتایج تصویری از پیکربندی‌های مختلف گلوگاه. در تصویر اندازه 1024×1024 ، جعبه‌های قرمز با اندازه 320×250 مناطقی را نشان می‌دهند که نتایج استخراج جاده‌ی مدل جاده‌ساز بهتر هستند.	۷.۴

۸.۴	مقایسه تصویری مدل جاده‌ساز و سایر روش‌های روز برای مجموعه داده گلوب عمیق. کادرهای قرمز، مناطقی را که روش ما بهتر از روش‌های دیگر عمل می‌کند، برجسته می‌کند.	۷۲
۹.۴	مقایسه بصری روش مدل جاده‌ساز و سایر روش‌های روز برای مجموعه داده جاده‌ای ماساچوست. جعبه‌های قرمز رنگ مناطقی را نشان می‌دهند که روش ما نسبت به روش‌های دیگر بهتر عمل کرده است.	۷۴
۱.۵	معماری کلی شبکه‌ی عصبی پیچشی چرخشی	۷۹
۲.۵	ساختار مدل شبکه عصبی پیچشی چرخشی	۸۰
۳.۵	ساختار اتصال تقویت‌شده با ویژگی	۸۱
۴.۵	ساختار اتصال تقویت‌شده با شکل	۸۲
۵.۵	نمونه‌ای از نتایج مدل‌های مختلف بر روی دیتاست ماساچوست.	۸۵
۶.۵	کانولوشن معمولی در برابر کانولوشن گسترش‌یافته	۸۶
۷.۵	متريک APLS تفاوت طول مسیر را بين ماسک واقعی و ماسک پيشنهادي مقایسه می کند. سمت چپ: کوتاه ترین مسیر بين گره منبع (سبز) و هدف (قرمز) در نمودار حقیقت زمین به رنگ زرد با طول مسیر ≈ ۹۴۸ متر نشان داده شده است. راست: کوتاه ترین مسیر بين گره منبع و هدف در نمودار پيشنهادي با ۳۰ یال حذف شده، با طول مسیر ≈ ۱۰۲۷ متر. اين تفاوت در طول، اساس متريک را تشکيل می دهد. . .	۸۷
۸.۵	اسکلت پس از کاهش به خط مرکزی	۸۸
۹.۵	چالشی مهم برای ساخت گراف از داده‌های تصویری	۹۰
۱۰.۵	نمونه‌هایی از بخش‌های طولانی جاده (مانند بزرگراه‌ها، پل‌ها) که معمولاً در اين روش از دست می‌روند.	۹۱
۱۱.۵	ذخیره سازی و سرعت بخشیدن به محاسبات الگوريتم‌های انجام شده بر روی اين مسیرها، مانند الگوريتم‌های کوتاه‌ترین مسیر	۹۱

فهرست جداول

۱.۴	مقایسه کمی پشتیبان‌های مختلف برای مدل جاده‌ساز با استفاده از مجموعه داده گلوب عمیق	۶۹
۲.۴	مقایسه کمی پیکربندی‌های مختلف گلوگاه	۶۹
۳.۴	نتایج عملکرد کمی مجموعه داده گلوب عمیق.	۷۳
۴.۴	نتایج عملکرد کمی مجموعه داده جاده‌ای ماساچوست.	۷۵

مقدمه

با پیشرفت تکنولوژی، بینایی ماشین تبدیل به ابزاری برای کمک به انسان‌ها شده است. اکنون می‌توان از این تکنولوژی برای ساخت نقشه‌ی خیابان‌ها با استفاده از نمای ماهواره‌ای پرداخت. برای این کار در ابتدا با استفاده از شبکه عصبی عمیق عمل بخش بندی تصویر انجام می‌شود. بعد از پردازش تصویر و استخراج اولیه خیابان‌ها باید توجه داشت که تمامی خطوط به درستی تشخیص داده شده باشند.

استخراج جاده از تصاویر ماهواره‌ای یکی از موضوعات تحقیقی داغی بوده که به دلیل نقش اساسی آن در برنامه‌هایی از جمله رانندگی خودکار، مسیریابی خودروها و نظارت بر جاده‌ها مورد توجه قرار گرفته است. در دهه‌های گذشته، پژوهشگران با استفاده از روش‌های سنتی مرتبط با مورفولوژی ریاضی و تجزیه و تحلیل ساختاری، نتایج خوبی با تصاویر با کنترast بالا به دست آورده‌اند. با این حال، این روش‌ها معمولاً توسط پارامترهای ثابت محدود شده‌اند و ثابت شده است که زمانی که در تصاویر با کنترast پایین اعمال می‌شوند، عملکرد کمتری دارند.

از منظر یادگیری ماشین، کار استخراج جاده می‌تواند به عنوان یک وظیفه طبقه‌بندی با دو دسته (جاده و پس‌زمینه) در نظر گرفته شود که معادل یک وظیفه تقسیم بندی دودویی است. با توجه به عملکرد عالی یادگیری عمیق در سال‌های اخیر برای وظایف دید کامپیوتری، پژوهشگران اکنون ترجیح می‌دهند از روش‌های یادگیری عمیق برای استخراج جاده استفاده کنند. برخی از اثار اخیر به تکنیک‌های استخراج جاده مبتنی بر شبکه عصبی پیچشی^۱ یا به طور اختصار CNN پرداخته‌اند که با غلبه بر مشکلات مذکور، از روش‌های سنتی عملکرد بهتری داشته‌اند. اما این کارها فقط استخراج جاده را به یک مسئله تقسیم بندی معنایی، ساده می‌کنند و ساختار ذاتی جاده را نادیده می‌گیرند.

استخراج جاده‌ها یک مسئله تقسیم بندی دقیق نیست؛ به دلیل: اولاً، وضوح تصاویر ماهواره‌ای اغلب کمتر از تصاویر در حالت عمومی است، به این معنا که شبکه‌های تقسیم بندی جاده باید یک زمینه دریافت کننده بزرگ داشته باشند. ثانياً، از آنجا که مسیر جاده‌ها در تصاویر ماهواره‌ای اغلب باریک و پیچیده هستند، انتظار می‌رود که شبکه، ویژگی‌های دقیق تصویر را حفظ کند.

¹Convolutional Neural Network

مدل‌های مبتنی بر شبکه‌های عصبی پیچشی برای حل مشکل به اندازه کافی موثر نیستند؛ زیرا زمینه پذیرنده معمولاً^۱ توسط اندازه هسته^۲ کانولوشن تعیین می‌شود. مدل‌های فعلی اکثراً از اندازه هسته‌ی ۳ × ۳ استفاده می‌کنند که نمی‌توانند نیازهای ما را برطرف کنند. در صورت افزایش اندازه هسته‌ی کانولوشن، هزینه محاسباتی افزایش خواهد یافت ولی بهبود کمی رخ می‌دهد.

علاوه بر این، در هنگام نمونه‌برداری تصویر، لایه‌های پولینگ ممکن است جزئیات تصویر را از دست بدهند. بنابراین، نیاز به یک ساختار جدید برای حل وظایف استخراج جاده دارد. خوشبختانه ترنسفورمر بینایی^۳ نشان می‌دهد که معماری ترنسفورمر عملکرد عالی برای حل مشکل بالا دارد.

mekanizm توجه، ترنسفورمر را تشویق می‌کند تا وابستگی‌های دوربرد را بهتر ایجاد کند به طوری که اطلاعات همگی بتواند در هر دو لایه عمیق و کم عمق مورد استفاده قرار گیرد. مبدل پنجره‌های جابجا شونده^۴ در مقایسه با مدل‌های شبکه عصبی پیچشی، ارتباط معنایی زمینه‌ای قوی‌تر و زمینه‌پذیرنده گسترده‌تری دارد. در ادامه با توجه به توضیحات بالا به ساخت مدل مربوط به سیستم پرداخته می‌شود. بخش‌های باقی‌مانده این مقاله به این شرح می‌باشد: نمای کلی از کارهای استخراج جاده که قبل از آن شده است و تفاوت‌های بین روش‌های مرتبط مورد تجزیه و تحلیل قرار می‌گیرد. معماری و طراحی مدل به صورت کامل توضیح داده می‌شود. مقایسه‌ی مختلف آزمایش‌ها ارائه می‌شوند و با مقایسه‌ی آزمایش‌های قبلی تحلیل می‌شود.

غالب مطالب این پایان‌نامه برگرفته از مراجع زیر می‌باشد:

Liu X, Wang Z, Wan J, et al. "RoadFormer: Road Extraction Using a Swin Transformer Combined with a Spatial and Channel Separable Convolution" (2023). [26]

Tao Chen, Yiran Liu, Haoyu Jiang, Ruirui Li. "Swin Transformer coupling CNNs Makes Strong Contextual Encoders for VHR Image Road Extraction" arXiv:2201.03178 (2023). [41]

^۱kernel size

^۲Vision Transformer(ViT)

^۳Swin Transformer

۱ فصل

مقدمه‌ای بر پردازش تصویر

۱.۱ مقدمه

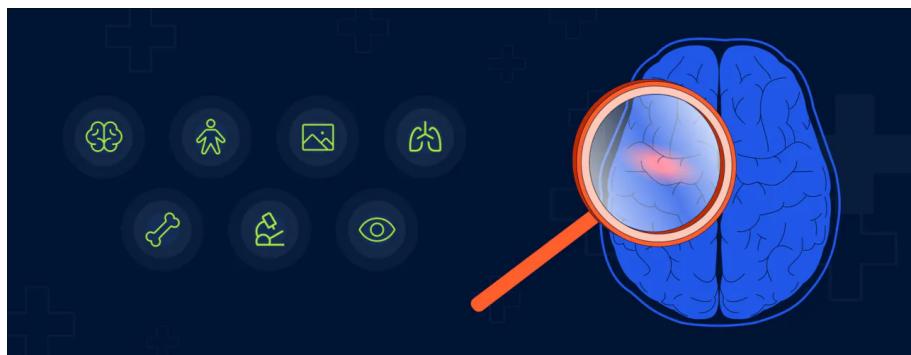
پردازش تصویر^۱ مجموعه روش‌هایی است که هدف‌شان دستکاری یا بهبود تصاویر است. پردازش تصویر به صورت پیکسل به پیکسل است.

در دنیای فناورانه امروز، تصاویر بخش جدایی ناپذیر از ارتباطات، سرگرمی و درک ما از دنیای اطراف هستند. از ثبت خاطرات گرفته تا تجزیه و تحلیل داده‌های علمی و تصاویر نقش مهمی در جنبه‌های مختلف زندگی ما دارند. پردازش تصویر، شاخه‌ای از علوم کامپیوتر است که به دستکاری و تجزیه و تحلیل تصاویر دیجیتال، به منظور استخراج اطلاعات معنادار و بهبود کیفیت آنها می‌پردازد. از کاربردهای این زمینه می‌توان به تشخیص چهره، اشیاء، حرکت، الگو، رنگ، شکل، نوشتار، اثر انگشت، صدا و امضا، پزشکی، بینایی ماشین، امنیت و نظارت، و همچنین در علوم فضایی اشاره کرد.

۲.۱ تاریخچه پردازش تصویر

تاریخچه پردازش تصویر به اوایل قرن بیستم باز می‌گردد، زمانی که دانشمندان شروع به مطالعه خواص تصاویر و چگونگی پردازش آن‌ها توسط ذهن انسان کردند. اولین کاربردهای عملی پردازش تصویر در دهه ۱۹۴۰ و ۱۹۵۰ در زمینه‌های پزشکی و نظامی بود. به عنوان مثال، در پزشکی از پردازش تصویر برای تشخیص سرطان و سایر بیماری‌ها استفاده می‌شد و در نظامی از پردازش تصویر برای شناسایی هواپیماهای دشمن استفاده می‌شد.

^۱Image processing



شکل ۱.۱: پردازش تصویر در حوزه‌ی پزشکی

در دهه ۱۹۶۰، پیشرفت‌های مهمی در زمینه پردازش تصویر صورت گرفت، از جمله توسعه الگوریتم‌های جدید برای تصحیح تصویر، تشخیص اشیا و تجزیه و تحلیل تصویر. در دهه ۱۹۹۰ و ۲۰۰۰، پردازش تصویر به دلیل پیشرفت‌های سریع در فناوری رایانه‌ای، به یک زمینه نوظهور تبدیل شد. در این دوره، الگوریتم‌های پردازش تصویر پیچیده‌تر و کارآمدتر شدند و کاربردهای پردازش تصویر به طور مداوم گسترش یافت. در حال حاضر نیز از آن برای توسعه خودروهای خودران، ربات‌ها و سایر فناوری‌های نوظهور استفاده می‌شود.

۳.۱ تکنیک‌های پیش‌پردازش

برای پیش‌پردازش داده‌های تصویری می‌توان از روش‌های ذیل استفاده کرد:

۱. اصلاح هندسی^۲: اصلاح هندسی یکی از مراحل مهم در پردازش تصاویر ماهواره‌ای و دیگر تصاویر دیجیتال است. هدف از این فرآیند تطبیق تصاویر به نحوی است که خطاهای هندسی حذف شده و تصاویر به مختصات جغرافیایی صحیح منتقل شوند. این اصلاحات برای اطمینان از اینکه داده‌ها به درستی تفسیر و تحلیل می‌شوند، حیاتی است. مراحل و فرآیند اصلاح هندسی عبارتند از:

- شناسایی نقاط کنترل زمین^۳:

نقاط کنترل زمین، نقاطی با مختصات جغرافیایی دقیق هستند که در هر دو تصویر ورودی و نقشه‌های مرجع شناخته شده‌اند. این نقاط به عنوان مرجع برای اصلاح خطاهای هندسی استفاده می‌شوند.

- مدل‌سازی خطاهای هندسی:

خطاهای هندسی می‌توانند ناشی از چندین منبع باشند، از جمله جابجایی ماهواره، انحراف زاویه دوربین، انحنای زمین، و دیگر عوامل محیطی. مدل‌های ریاضی برای شناسایی و تصحیح این خطاهای کار گرفته می‌شوند.

^۲Geometric Correction

^۳GCPs

- تطبیق تصویر با مدل‌های ریاضی:

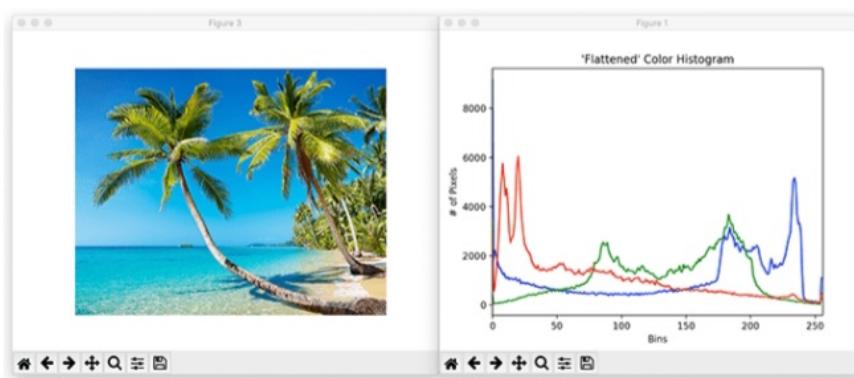
پس از شناسایی خطاهای هندسی، از مدل‌های ریاضی برای تطبیق تصویر با مختصات جغرافیایی دقیق استفاده می‌شود. این مدل‌ها می‌توانند خطی یا غیرخطی باشند و بر اساس نوع خطاهای و دقت مورد نیاز انتخاب می‌شوند.

- اعمال تصحیحات به تصویر:

پس از محاسبه تصحیحات، این تغییرات به تصویر اعمال می‌شوند تا جابجایی‌ها و انحرافات برطرف شوند. این مرحله ممکن است شامل بازنمونه‌برداری تصویر نیز باشد تا تطابق دقیق‌تری حاصل شود[۳۶].

۲. اصلاح رنگ و روشنایی^۴:

تصحیح هیستوگرام، توازن رنگ سفید.



شکل ۲.۱: ساخت یک هیستوگرام رنگی برای هر بخش از تصویر مشاهده می‌شود.

۳. کاهش نویز^۵:

فیلترهای میانگین^۶، فیلتر میانه^۷.

^۴Color and Brightness Adjustment

^۵Noise Reduction

^۶Mean Filter

^۷Median Filter

۴.۱ مبانی پردازش تصویر

۱.۴.۱ تعریف تصویر دیجیتال

تصویر دیجیتال^۸ به مجموعه‌ای از نقاط داده‌ای (پیکسل‌ها^۹) گفته می‌شود که به طور معمول به صورت ماتریسی ذخیره می‌شود. هر پیکسل دارای مقادیر مشخصی برای رنگ^{۱۰} و شدت نور^{۱۱} است که این مقادیر می‌توانند به صورت عددی ذخیره شوند.

یک تصویر دیجیتال معمولاً به صورت یک ماتریس دو بعدی ($I(x, y)$ مدل می‌شود که در آن x و y مختصات پیکسل‌ها در محورهای افقی و عمودی هستند و مقدار ($I(x, y)$) شدت روشنایی آن پیکسل است.

$$I(x, y) = \begin{bmatrix} i_{11} & i_{12} & \dots & i_{1n} \\ i_{21} & i_{22} & \dots & i_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ i_{m1} & i_{m2} & \dots & i_{mn} \end{bmatrix}$$

که در آن i_{ij} مقدار شدت روشنایی پیکسل در مختصات (j, i) است. این ماتریس می‌تواند برای پردازش و تحلیل توسط الگوریتم‌های پردازش تصویر استفاده شود.

۲.۴.۱ مفاهیم پایه‌ای پردازش تصویر

پردازش تصویر شامل مجموعه‌ای از تکنیک‌ها است که برای تحلیل، بهبود، و استخراج اطلاعات از تصاویر دیجیتال به کار می‌رود. این تکنیک‌ها می‌توانند شامل فیلتر کردن^{۱۲}، آشکارسازی لبه‌ها^{۱۳}، و تبدیل‌های هندسی^{۱۴} باشند که هر کدام در مراحل مختلف پردازش تصویر مورد استفاده قرار می‌گیرند.^[۱۵]

فیلتر کردن

فیلتر کردن یکی از تکنیک‌های اصلی پردازش تصویر است که برای تغییر یا بهبود کیفیت تصویر استفاده می‌شود. به عنوان مثال، فیلتر میانگین^{۱۵} که یک فیلتر پایین‌گذر است، می‌تواند برای کاهش نویز استفاده شود. فرمول فیلتر میانگین به صورت زیر است:

^۸Digital Image

^۹Pixels

^{۱۰}Color

^{۱۱}Intensity

^{۱۲}Filtering

^{۱۳}Edge Detection

^{۱۴}Geometric Transformations

^{۱۵}Mean Filter

$$I_{\text{filtered}}(x, y) = \frac{1}{N} \sum_{i=-k}^k \sum_{j=-k}^j I(x+i, y+j)$$

که در آن N تعداد پیکسل‌های درون فیلتر است و k اندازه فیلتر را تعیین می‌کند.

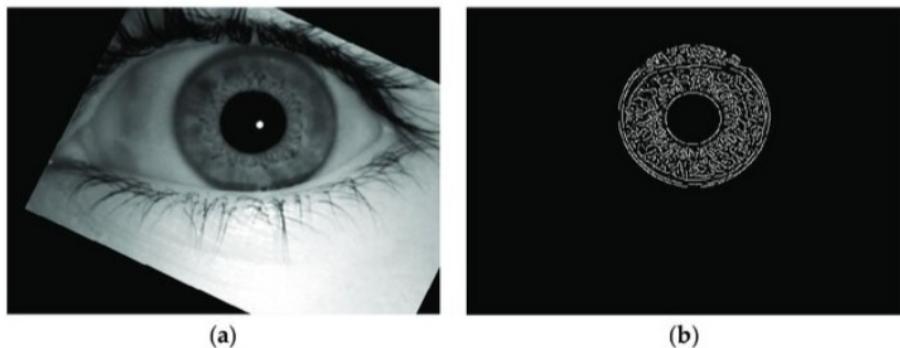
آشکارسازی لبه‌ها

آشکارسازی لبه‌ها یک روش مهم برای استخراج ویژگی‌های اصلی در تصویر است. یکی از روش‌های متداول برای آشکارسازی لبه‌ها، استفاده از فیلتر سوبل^{۱۶} است. فیلتر سوبل برای محاسبه گرادیان شدت روشنایی تصویر در دو جهت افقی و عمودی به کار می‌رود:

$$G_x = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix}, \quad G_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix}$$

که در آن G_x و G_y به ترتیب گرادیان‌های افقی و عمودی هستند. مجموع مربعات این گرادیان‌ها می‌تواند شدت لبه‌ها را در تصویر مشخص کند:

$$G = \sqrt{G_x^2 + G_y^2}$$



شکل ۳.۱: مثالی از اهمیت آشکارسازی لبه‌ها در پردازش تصویر

^{۱۶}Sobel Filter

۳.۴.۱ انواع نویز در تصاویر و روش‌های حذف آنها

نویز در تصاویر می‌تواند به صورت تصادفی^{۱۷} یا ساختاری^{۱۸} باشد و از منابع مختلفی مانند حسگرهای^{۱۹} یا انتقال داده‌ها ناشی شود.

یکی از نویزهای رایج، نویز گوسی^{۲۰} است که با استفاده از فیلتر گوسی^{۲۱} می‌توان آن را حذف کرد. فیلتر گوسی یک فیلتر پایین‌گذر است که تابع گوسی برای فیلتر کردن استفاده می‌کند:

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$$

که در آن σ انحراف معیار تابع گوسی است و تعیین‌کننده میزان صاف‌شدن تصویر است.

همچنین، نویز نمک و فلفل^{۲۲} یکی دیگر از نویزهای رایج است که می‌تواند با استفاده از فیلتر میانه^{۲۳} حذف شود. فیلتر میانه با جایگزینی هر پیکسل با مقدار میانه پیکسل‌های مجاور، نویزهای اسپاکی را حذف می‌کند:

$$I_{\text{median}}(x, y) = \text{median}\{I(x+i, y+j) : -k \leq i, j \leq k\}$$

که در آن میانه مقادیر پیکسل‌های درون پنجره فیلتر محاسبه می‌شود.



شکل ۴.۱: مثالی از نویز و فیلترهای حذف نویز.

^{۱۷}Random Noise

^{۱۸}Structural Noise

^{۱۹}Sensors

^{۲۰}Gaussian Noise

^{۲۱}Gaussian Filter

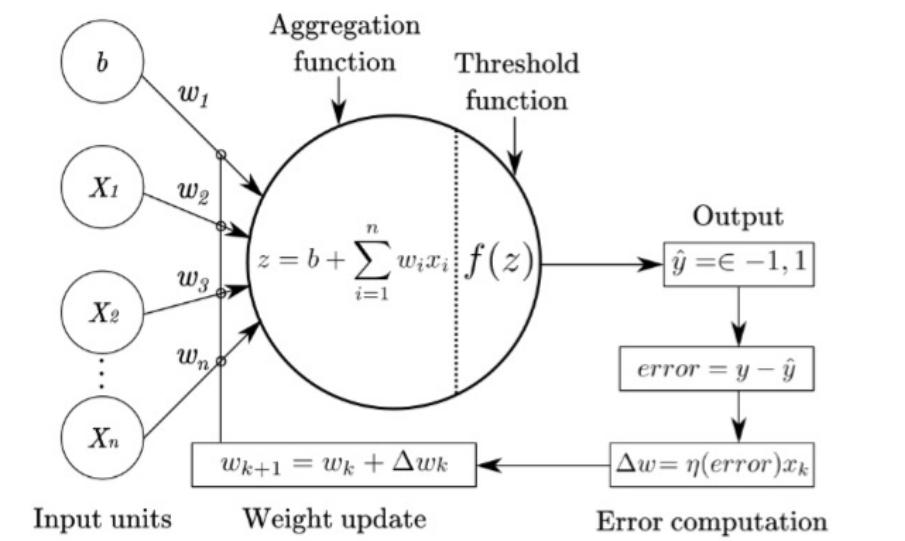
^{۲۲}Salt-and-Pepper Noise

^{۲۳}Median Filter

۱.۵.۱ یادگیری ماشین

۱.۵.۱.۱ تاریخچه و تعریف یادگیری ماشین

یادگیری ماشین^{۲۴} به عنوان یک زیرشاخه از هوش مصنوعی^{۲۵} شناخته می‌شود که هدف آن توسعه الگوریتم‌ها و مدل‌هایی است که بتوانند بدون نیاز به برنامه‌ریزی صریح، از داده‌ها یاد بگیرند و عملکرد خود را بهبود بخشنند. این حوزه در دهه ۱۹۵۰ با معرفی پرسپترون^{۲۶} توسط فرانک روزنبلات به عنوان یکی از نخستین مدل‌های یادگیری ماشین آغاز شد.^[۱۴]



شکل ۱.۵.۱: نمایی از الگوریتم پرسپترون

در دهه‌های بعد، یادگیری ماشین با توسعه الگوریتم‌هایی مانند ماشین‌های بردار پشتیبان^{۲۷} و جنگل‌های تصادفی^{۲۸} به سرعت پیشرفت کرد. در دهه ۲۰۱۰، با پیشرفت سخت‌افزارهای محاسباتی و افزایش حجم داده‌ها، یادگیری عمیق^{۲۹} که زیرمجموعه‌ای از یادگیری ماشین است، به یکی از مهم‌ترین و مؤثرترین روش‌ها در پردازش تصویر^{۳۰}، پردازش زبان طبیعی^{۳۱} و بسیاری از حوزه‌های دیگر تبدیل شد.^[۱۹].

^{۲۴}Machine Learning

^{۲۵}Artificial Intelligence

^{۲۶}Perceptron

^{۲۷}Support Vector Machines - SVM

^{۲۸}Random Forests

^{۲۹}Deep Learning

^{۳۰}Image Processing

^{۳۱}Natural Language Processing - NLP

۲.۵.۱ مبانی یادگیری ماشین

یادگیری ماشین به روش‌های خودآموزی اشاره دارد که در آن الگوریتم‌ها با استفاده از داده‌ها به یادگیری و پیش‌بینی می‌پردازند. این روش‌ها به طور عمده شامل دسته‌بندی^{۳۲}، رگرسیون^{۳۳} و خوش‌بندی^{۳۴} است که در مسائل مختلف از جمله پردازش تصویر، تشخیص الگوها و پیش‌بینی استفاده می‌شوند[۲۹].

۳.۵.۱ انواع الگوریتم‌های یادگیری ماشین

الگوریتم‌های یادگیری ماشین به سه دسته اصلی تقسیم می‌شوند: نظارت شده^{۳۵}، بدون نظارت^{۳۶} و تقویتی^{۳۷}. هر کدام از این الگوریتم‌ها کاربردهای خاص خود را دارند.

۴.۵.۱ یادگیری نظارت شده

پرسپترون یکی از قدیمی‌ترین و ساده‌ترین مدل‌های شبکه عصبی^{۳۸} است که در یادگیری نظارت شده به کار می‌رود. در این مدل، هر نرون یک تصمیم‌گیرنده خطی است که تلاش می‌کند داده‌ها را بر اساس یکتابع خطی به دو دسته مجزا تقسیم کند. معادله تصمیم‌گیری پرسپترون به صورت زیر است:

$$y = f(\mathbf{w} \cdot \mathbf{x} + b)$$

که در آن:

• \mathbf{w} بردار وزن‌ها^{۳۹}

• \mathbf{x} بردار ویژگی‌های ورودی^{۴۰}

• b بایاس^{۴۱}

• f تابع فعال‌سازی^{۴۲} (مانند تابع پله^{۴۳})

^{۳۲}Classification

^{۳۳}Regression

^{۳۴}Clustering

^{۳۵}Supervised Learning

^{۳۶}Unsupervised Learning

^{۳۷}Reinforcement Learning

^{۳۸}Neural Network

^{۳۹}Weights Vector

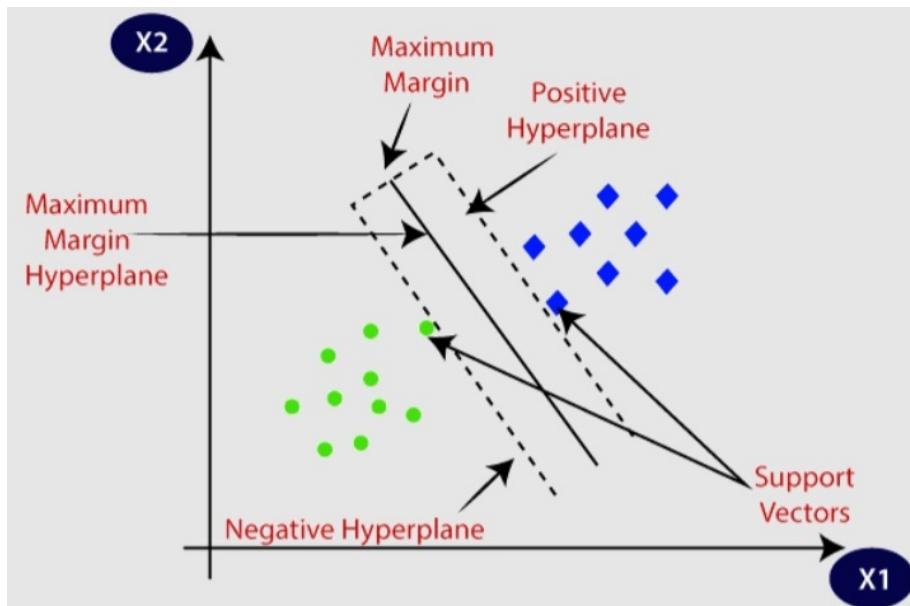
^{۴۰}Input Features Vector

^{۴۱}Bias

^{۴۲}Activation Function

^{۴۳}Step Function

ماشین‌های بردار پشتیبان یکی از قدرتمندترین الگوریتم‌های دسته‌بندی است که هدف آن یافتن یک ابرصفحه^{۴۴} است که دسته‌های مختلف داده را از هم جدا کند.



شکل ۱: نمایی از الگوریتم ماشین بردار پشتیبان

معادله ابرصفحه به صورت زیر است:

$$\mathbf{w} \cdot \mathbf{x} + b = 0$$

که در آن:

^{۴۵} بردار وزنها ●

^{۴۶} بردار ویژگیها ●

^{۴۷} b مقدار بایاس ●

^{۴۴} Hyperplane

^{۴۵} Weights Vector

^{۴۶} Feature Vector

^{۴۷} Bias

کاربرد در پردازش تصویر: ماشین بردار پشتیبان برای دسته‌بندی دقیق بخش‌های مختلف تصویر به خیابان و غیر خیابان استفاده می‌شود. مثلاً در استخراج نقشه خیابان‌ها از تصاویر ماهواره‌ای، ماشین بردار پشتیبان می‌تواند به شناسایی مرزهای خیابان و تفکیک آن از سایر نواحی کمک کند.

تحلیل مولفه‌های اصلی PCA^{۴۸} یک تکنیک کاهش ابعاد^{۴۹} است که هدف آن کاهش پیچیدگی داده‌ها با حفظ بیشترین واریانس ممکن است. این تکنیک با تجزیه ماتریس کوواریانس داده‌ها به مولفه‌های اصلی انجام می‌شود:

$$\mathbf{C} = \frac{1}{m} \sum_{i=1}^m (\mathbf{x}^{(i)} - \mu)(\mathbf{x}^{(i)} - \mu)^T$$

که در آن:

- \mathbf{C} ماتریس کوواریانس^{۵۰}

- $\mathbf{x}^{(i)}$ نمونه داده^{۵۱}

- μ میانگین داده‌ها^{۵۲}

^{۴۸}Principal Component Analysis

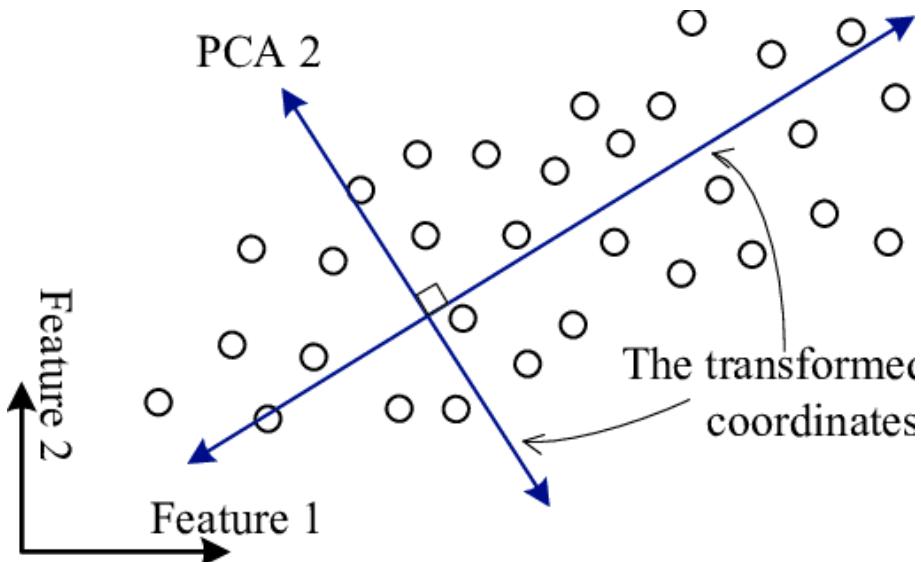
^{۴۹}Dimensionality Reduction

^{۵۰}Covariance Matrix

^{۵۱}Data Sample

^{۵۲}Mean

کاربرد در پردازش تصویر: PCA می‌تواند برای کاهش ابعاد تصاویر ماهواره‌ای و استخراج ویژگی‌های اصلی از تصاویر به کار رود. این ویژگی‌ها سپس به عنوان ورودی به الگوریتم‌های دیگر ماشین بردار پشتیبان داده می‌شوند تا دقیق استخراج خیابان‌ها افزایش یابد.



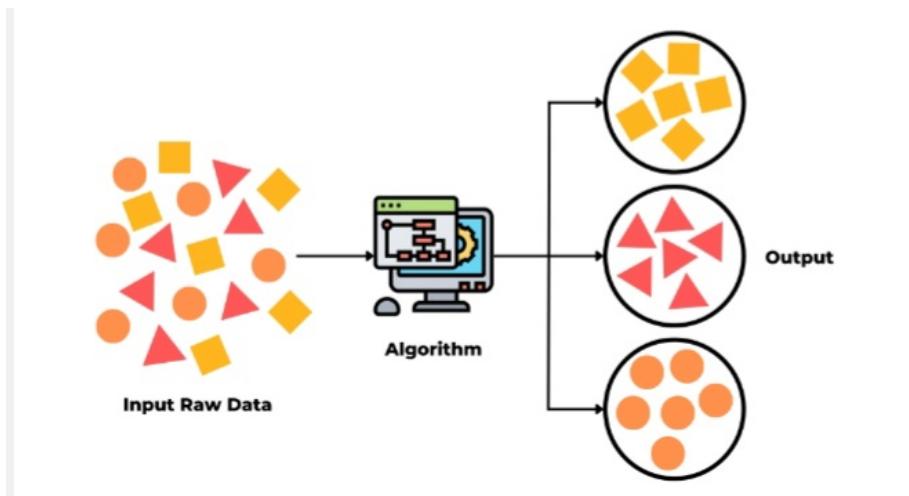
شکل ۷.۱: روش تحلیل مولفه‌های اصلی

۵.۵.۱ یادگیری بدون نظارت

خوشه‌بندی K-Means یکی از معروف‌ترین الگوریتم‌های یادگیری بدون نظارت است که هدف آن تقسیم داده‌ها به k خوشه مختلف است. این الگوریتم بر اساس فاصله اقلیدسی (Euclidean distance) عمل می‌کند و داده‌ها را به نزدیک‌ترین مرکز خوشه اختصاص می‌دهد. فرمول فاصله اقلیدسی به صورت زیر است:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

کاربرد در پردازش تصویر: K-Means برای تقسیم‌بندی تصاویر ماهواره‌ای به نواحی مختلف استفاده می‌شود. این الگوریتم می‌تواند به عنوان یک مرحله پیش‌پردازش برای شناسایی بخش‌های مربوط به خیابان‌ها از سایر نواحی استفاده شود.



شکل ۱.۸.۱: مثالی از خوشه‌بندی

۶.۵.۱ یادگیری تقویتی

Q-Learning یکی از معروف‌ترین الگوریتم‌های یادگیری تقویتی است. در این روش، یک عامل (Agent) با محیط تعامل می‌کند و تلاش می‌کند تا سیاست بهینه‌ای برای حداکثر کردن پاداش‌های دریافتی در طول زمان پیدا کند. فرمول بهروزرسانی مقدار Q به صورت زیر است:

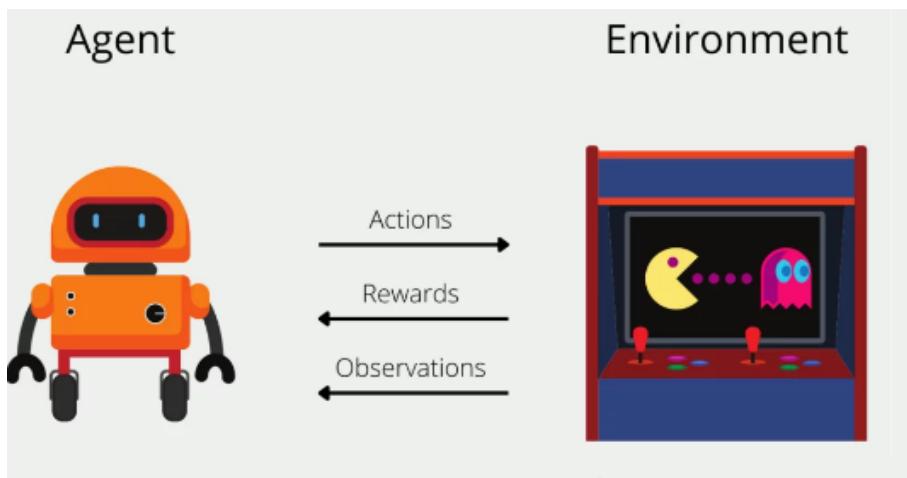
$$Q(s, a) \leftarrow Q(s, a) + \alpha \left[r + \gamma \max_{a'} Q(s', a') - Q(s, a) \right]$$

که در آن:

- α نرخ یادگیری
- r پاداش فوری
- γ ضریب تخفیف
- s حالت فعلی
- a عمل انجام شده

کاربرد در پردازش تصویر: Q-Learning می‌تواند در مسائل رده‌بی اشیاء در ویدئوها و یا تعیین مسیر بهینه برای ربات‌ها در محیط‌های پیچیده استفاده شود. برای مثال، در پروژه‌های پردازش تصویر

ماهواره‌ای، می‌توان از Q-Learning برای بهینه‌سازی مسیر جستجو و ردیابی تغییرات در تصاویر استفاده کرد.



شکل ۹.۱: مثالی از یادگیری تقویتی

۷.۵.۱ نتیجه‌گیری و کاربردها در استخراج نقشه خیابان از تصاویر ماهواره‌ای

هر یک از الگوریتم‌های ذکر شده می‌توانند نقش مهمی در استخراج نقشه خیابان‌ها از تصاویر ماهواره‌ای ایفا کنند. به طور خاص:

- **شبکه‌های عصبی پیچشی**: این شبکه‌ها به طور ویژه برای پردازش داده‌های تصویری طراحی شده‌اند و می‌توانند برای شناسایی و استخراج خیابان‌ها از تصاویر ماهواره‌ای استفاده شوند. با استفاده از لایه‌های کانولوشن، این شبکه‌ها قادرند ویژگی‌های پیچیده‌ای مانند الگوهای خیابانی، تقاطع‌ها و ساختارهای جاده‌ای را شناسایی کنند. در این فرآیند، ابتدا تصاویر ماهواره‌ای به شبکه‌های عصبی پیچشی داده می‌شوند تا بخش‌های مرتبط با خیابان‌ها استخراج شوند.

- **الگوریتم‌های خوشه‌بندی**: این الگوریتم‌ها می‌توانند برای تقسیم‌بندی تصاویر ماهواره‌ای به نواحی مختلف از جمله خیابان‌ها و نواحی غیر خیابانی استفاده شوند. K-Means با گروه‌بندی پیکسل‌های مشابه بر اساس ویژگی‌های رنگ، شدت نور، و بافت، قادر است بخش‌هایی از تصویر که به خیابان‌ها مربوط هستند را از سایر نواحی متمایز کند. این تقسیم‌بندی اولیه می‌تواند به عنوان یک پیش‌پردازش برای الگوریتم‌های پیچیده‌تر مورد استفاده قرار گیرد.

- **تحلیل مؤلفه‌های اصلی (PCA)**: این تکنیک برای کاهش ابعاد داده‌های تصویری و تمرکز بر ویژگی‌های اصلی که در شناسایی خیابان‌ها مؤثر هستند، مورد استفاده قرار می‌گیرد. با استفاده از PCA، می‌توان نویزهای غیرمرتبط را حذف کرد و تنها اطلاعاتی را که برای تشخیص الگوهای خیابانی مهم هستند، حفظ نمود.

– ماشین‌های بردار پشتیبان: این الگوریتم می‌تواند برای دسته‌بندی بخش‌های مختلف تصویر به خیابان و غیر خیابان استفاده شود. ماشین‌های بردار پشتیبان با ایجاد یک مرز تصمیم‌گیری بهینه بین پیکسل‌هایی که به خیابان‌ها تعلق دارند و پیکسل‌های دیگر، می‌تواند دقیق استخراج نقشه خیابان را افزایش دهد.

– یادگیری تقویتی (Q-Learning): در پروژه‌هایی که نیاز به بهروزرسانی و بهینه‌سازی مدام ن نقشه‌ها بر اساس تصاویر جدید دارند، یادگیری تقویتی می‌تواند به بهینه‌سازی فرآیند استخراج خیابان‌ها کمک کند. برای مثال، می‌توان از این الگوریتم برای یادگیری یک سیاست بهینه جهت ردیابی تغییرات در شبکه خیابان‌ها و بهروزرسانی نقشه‌ها استفاده کرد.

۸.۵.۱ نحوه آموزش مدل‌های یادگیری ماشین

مدل‌های یادگیری ماشین با استفاده از داده‌های آموزشی^{۵۳} آموزش داده می‌شوند و سپس برای پیش‌بینی یا تحلیل داده‌های جدید استفاده می‌شوند. فرآیند آموزش شامل تنظیم پارامترهای مدل برای کاهش خطای است. برای مثال، از روش گرادیان کاوشی^{۵۴} برای بهینه‌سازی پارامترها استفاده می‌شود:

$$\theta = \theta - \alpha \frac{\partial J(\theta)}{\partial \theta}$$

که در آن:

● θ پارامترهای مدل^{۵۵}

● α نرخ یادگیری^{۵۶}

● $J(\theta)$ تابع هزینه^{۵۷}

^{۵۳}Training Data

^{۵۴}Gradient Descent

^{۵۵}Model Parameters

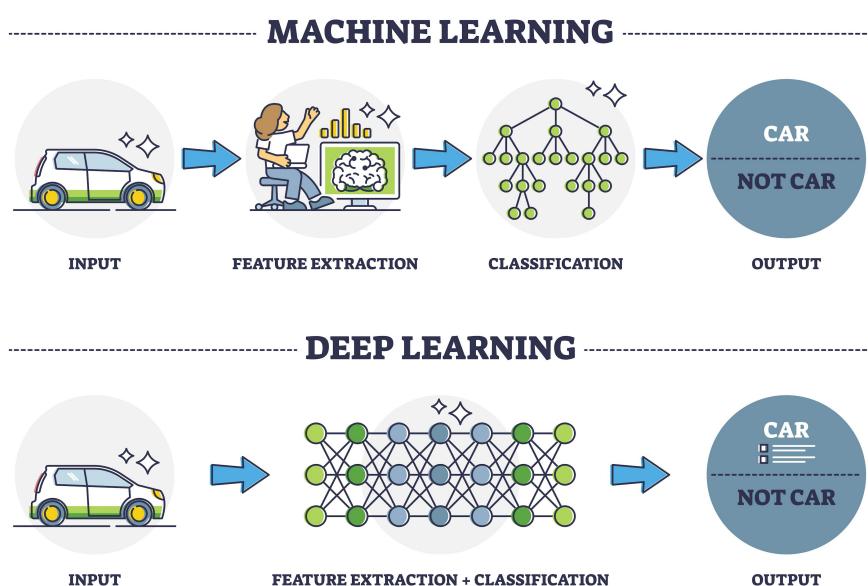
^{۵۶}Learning Rate

^{۵۷}Cost Function

۶.۱ یادگیری عمیق

۱.۶.۱ تفاوت یادگیری ماشین و یادگیری عمیق

یادگیری عمیق^{۵۸} زیرشاخه‌ای از یادگیری ماشین^{۵۹} است که از شبکه‌های عصبی عمیق^{۶۰} با لایه‌های متعدد استفاده می‌کند. تفاوت اصلی بین یادگیری ماشین و یادگیری عمیق در این است که در یادگیری عمیق، مدل‌ها به صورت خودکار ویژگی‌های داده‌ها را استخراج می‌کنند و نیازی به تعریف ویژگی‌ها به صورت دستی نیست[۱۵].



شکل ۱۰.۱: شبهه و تفاوت یادگیری ماشین و یادگیری عمیق

در یادگیری ماشین سنتی، مدل‌ها نیاز به ورودی‌هایی دارند که به صورت دستی مهندسی شده‌اند، در حالی که در یادگیری عمیق، شبکه‌های عصبی چندلایه می‌توانند ویژگی‌های پیچیده‌ای را از داده‌ها استخراج کنند که برای مسائل پیچیده‌تری مانند تشخیص تصویر، پردازش زبان طبیعی و بازی‌های ویدئویی به کار می‌روند.

^{۵۸}Deep Learning

^{۵۹}Machine Learning

^{۶۰}Deep Neural Networks - DNN

شبکه‌های عصبی مصنوعی^{۶۱} شامل مجموعه‌ای از نرون‌ها^{۶۲} هستند که به صورت لایه‌ای سازماندهی شده‌اند. هر نرون یک تابع ریاضی است که ورودی‌ها را دریافت کرده و پس از اعمال تابع فعال‌سازی^{۶۳}، خروجی تولید می‌کند.

$$y = f \left(\sum_{i=1}^n w_i x_i + b \right)$$

که در آن:

- y خروجی نرون
- f تابع فعال‌سازی (مانند سیگموید^{۶۴} یا ReLU)
- w_i وزن ورودی i
- x_i ورودی i
- b بایاس

این شبکه‌ها قابلیت یادگیری الگوهای پیچیده از داده‌ها را دارند و برای مسائل مختلفی از جمله تشخیص تصویر و شناسایی الگوها به کار می‌روند.

۲.۶.۱ شبکه‌های عصبی پیچشی

شبکه‌های عصبی پیچشی^{۶۵} به طور خاص برای پردازش داده‌های تصویری طراحی شده‌اند. این شبکه‌ها از لایه‌های کانولوشن^{۶۶} استفاده می‌کنند که قادر به استخراج ویژگی‌های محلی از تصاویر هستند.

$$f(x, y) = \sum_{i=1}^m \sum_{j=1}^n w_{ij} \cdot I(x+i, y+j)$$

که در آن:

- $f(x, y)$ خروجی فیلتر در موقعیت (x, y) است.
- w_{ij} وزن‌های فیلتر هستند.

^{۶۱} Artificial Neural Networks - ANN

^{۶۲} Neurons

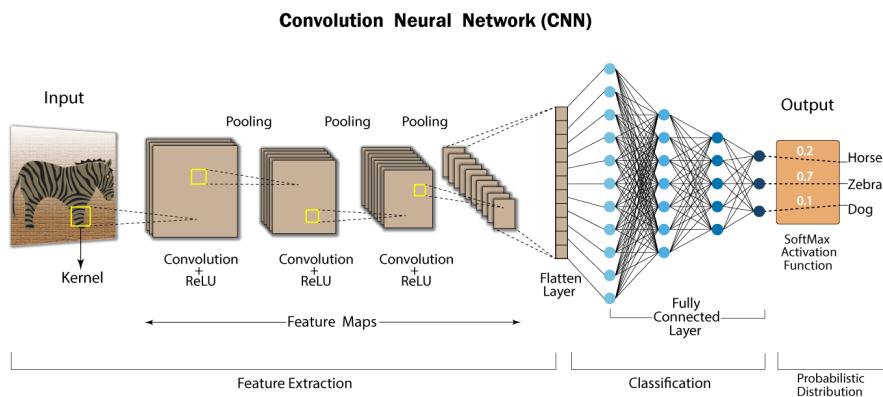
^{۶۳} Activation Function

^{۶۴} Sigmoid

^{۶۵} Convolutional Neural Networks - CNN

^{۶۶} Convolutional Layers

مقدار پیکسل در موقعیت $(x + i, y + j)$ تصویر است. ●



شکل ۱۱.۱: شبکه‌ی عصبی پیچشی

این شبکه‌ها قادر به شناسایی ویژگی‌های پیچیده مانند لبه‌ها، گوش‌ها و بافت‌ها هستند و برای مسائل مرتبط با تصاویر مانند شناسایی اشیاء، تشخیص چهره و استخراج خیابان‌ها از تصاویر ماهواره‌ای کاربرد دارند.

۳.۶.۱ شبکه‌های عصبی بازگشتی و حافظه طولانی کوتاه‌مدت

شبکه‌های عصبی بازگشتی^{۶۷} برای پردازش داده‌های توالی دار مانند متن، صدا و سری‌های زمانی طراحی شده‌اند. این شبکه‌ها قادرند از اطلاعات گذشته برای پیش‌بینی آینده استفاده کنند.

LSTM^{۶۸} نوعی از شبکه‌های عصبی بازگشتی است که توانایی حفظ اطلاعات در طول زمان را دارد و به دلیل استفاده از واحدهای حافظه، قادر به یادگیری روابط بلندمدت در داده‌ها است.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

که در آن:

- خروجی دروازه فراموشی^{۶۹} است. f_t

- وزن‌های دروازه فراموشی هستند. W_f

- وضعیت قبلی^{۷۰} است. h_{t-1}

- ورودی فعلی^{۷۱} است. x_t

- بایاس دروازه فراموشی است. b_f

- σ تابع سیگموید است.

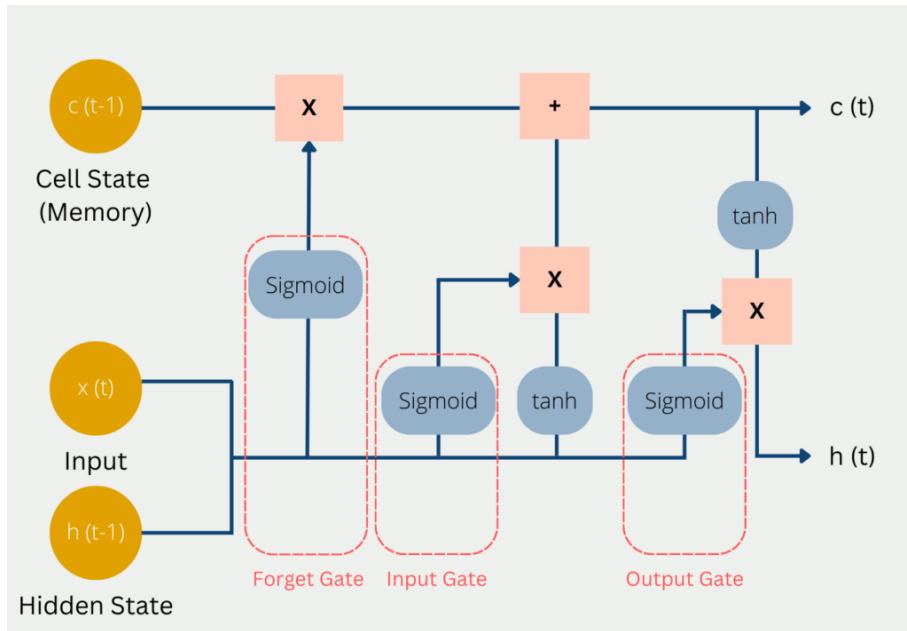
^{۶۷} Recurrent Neural Networks - RNN

^{۶۸} Long Short-Term Memory

^{۶۹} Forget Gate

^{۷۰} Previous State

^{۷۱} Current Input



شکل ۱۲.۱: معماری شبکه‌ی عصبی حافظه طولانی کوتاه‌مدت

این شبکه‌ها برای مسائل مرتبط با داده‌های ترتیبی مانند ترجمه ماشینی، تشخیص گفتار و تحلیل سری‌های زمانی به کار می‌روند.

۴.۶.۱ آموزش شبکه‌های عصبی عمیق

آموزش شبکه‌های عصبی عمیق شامل فرآیند بهروزرسانی وزن‌های شبکه است که معمولاً^{۷۱} با استفاده از الگوریتم انتشار معکوس^{۷۲} و گرادیان کاوهشی^{۷۳} انجام می‌شود. هدف از این فرآیند، کاهش تابع هزینه^{۷۴} و بهبود عملکرد مدل است.

$$\theta = \theta - \alpha \frac{\partial J(\theta)}{\partial \theta}$$

که در آن:

- θ وزن‌ها و بایاس‌های شبکه هستند.
- α نرخ یادگیری است.
- $J(\theta)$ تابع هزینه است.

^{۷۱} Backpropagation

^{۷۲} Gradient Descent

^{۷۳} Cost Function

برای بهینه‌سازی عملکرد شبکه‌های عصبی عمیق، تنظیم مناسب هایپرپارامترها^{۷۵} مانند نرخ یادگیری، تعداد لایه‌ها و تعداد نرون‌ها اهمیت زیادی دارد.

۷.۱ مدل‌های ترانسفورمر در یادگیری عمیق

۱.۷.۱ ساختار کلی مدل‌های ترانسفورمر

مدل‌های ترانسفورمر به طور خاص برای پردازش داده‌های توالی دار^{۷۶} طراحی شده‌اند و از مکانیزم‌های توجه^{۷۷} برای بررسی اهمیت اجزای مختلف داده‌ها استفاده می‌کنند^[۴۷].

ساختار کلی یک مدل ترانسفورمر شامل چندین لایه توجه^{۷۸} و لایه‌های کامل‌کننده^{۷۹} است. در این مدل‌ها، توجه چندسر^{۸۰} به مدل امکان می‌دهد که به بخش‌های مختلف توالی با دقت متفاوت توجه کند.

$$\text{Attention}(Q, K, V) = \text{softmax} \left(\frac{QK^T}{\sqrt{d_k}} \right) V$$

که در آن:

- Q ماتریس پرسش‌ها^{۸۱} است.
- K ماتریس کلیدها^{۸۲} است.
- V ماتریس ارزش‌ها^{۸۳} است.
- ابعاد کلیدها است. d_k

مدل‌های ترانسفورمر به دلیل توانایی‌شان در پردازش موازی و یادگیری روابط پیچیده بین اجزای داده‌ها، در پردازش زبان طبیعی و بینایی کامپیوتری بسیار موفق بوده‌اند.

^{۷۵}Hyperparameters

^{۷۶}Sequential Data

^{۷۷}Attention Mechanism

^{۷۸}Attention Layers

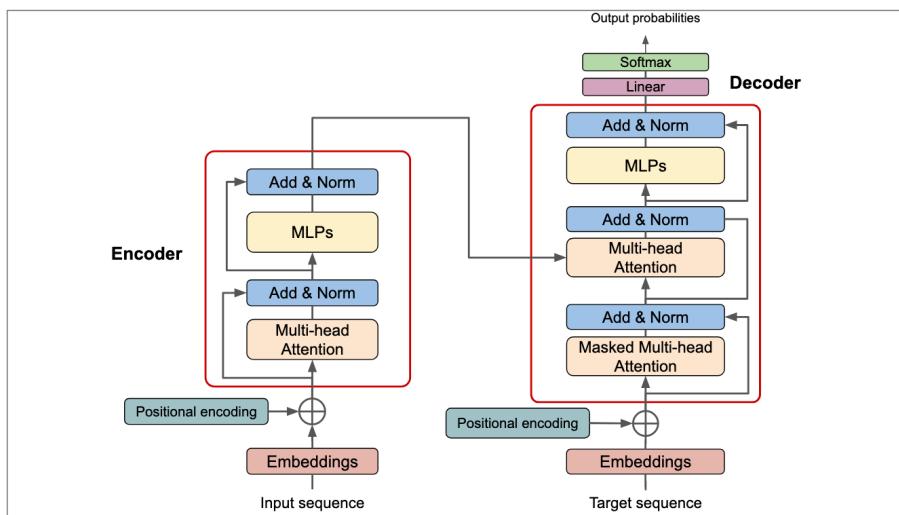
^{۷۹}Feed-Forward Layers

^{۸۰}Multi-Head Attention

^{۸۱}Queries

^{۸۲}Keys

^{۸۳}Values



شکل ۱۳.۱: ساختار مدل ترنسفورمر

۲.۷.۱ مدل ترنسفور چرخشی

مدل ترنسفور چرخشی^{۸۴} یکی از نسخه‌های پیشرفته مدل‌های ترنسفورمر است که به طور خاص برای پردازش داده‌های تصویری طراحی شده است. این مدل از پنجره‌های توجه^{۸۵} با مقیاس‌های مختلف برای پردازش تصویر در سطوح مختلف استفاده می‌کند [۲۷].

در ترنسفور چرخشی، تصویر به چندین بخش کوچک تقسیم می‌شود و هر بخش به طور جداگانه مورد پردازش قرار می‌گیرد. سپس این بخش‌ها به تدریج با مقیاس‌های بزرگتر ترکیب می‌شوند تا یک نمایش کلی از تصویر ایجاد شود.

$$\text{SwinAttention} = \text{Attention}(Q_{\text{win}}, K_{\text{win}}, V_{\text{win}})$$

که در آن:

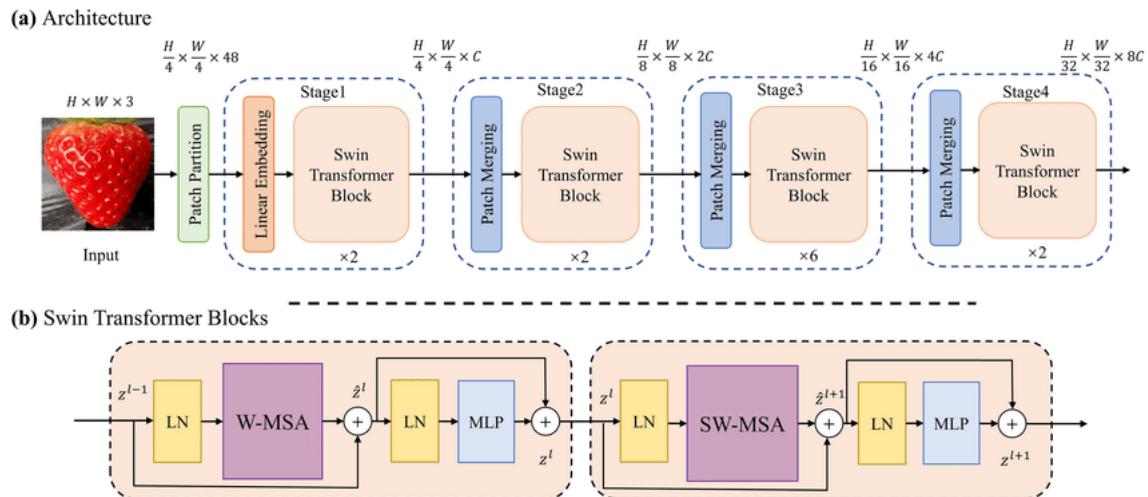
Q_{win} ماتریس پرسش‌های پنجره‌ای است. •

K_{win} ماتریس کلیدهای پنجره‌ای است. •

V_{win} ماتریس ارزش‌های پنجره‌ای است. •

^{۸۴}Swin Transformer

^{۸۵}Attention Windows



شکل ۱۴.۱: نمای کلی از معماری مدل ترانسفور چرخشی.

مدل ترانسفور چرخشی به دلیل توانایی اش در پردازش داده‌های تصویری و بهره‌گیری از ساختار سلسله‌مراتبی، در بسیاری از مسائل بینایی کامپیوتری، از جمله تشخیص اشیاء و تقسیم‌بندی تصویر، عملکرد بسیار خوبی دارد.

۳.۷.۱ کاربرد ترانسفورها در بینایی کامپیوتری

مدل‌های ترانسفورمر، به ویژه مدل‌هایی مانند ترانسفور چرخشی، در پردازش تصویر و بینایی کامپیوتری برای وظایف مختلفی مانند شناسایی اشیاء، دسته‌بندی تصویر و تقسیم‌بندی تصویر استفاده می‌شوند [۱۲].

در مسائل استخراج نقشه خیابان از تصاویر ماهواره‌ای، مدل‌های ترانسفورمر می‌توانند به شناسایی دقیق‌تر مرزها و ویژگی‌های خیابان کمک کنند. این مدل‌ها به دلیل توانایی‌شان در یادگیری روابط پیچیده بین اجزای مختلف تصویر، می‌توانند ویژگی‌هایی مانند تقاطع‌ها، پیچ‌ها و انشعابات خیابان‌ها را با دقت بیشتری شناسایی کنند.

۱۸.۱ استخراج ویژگی‌ها و کاربرد آن در پردازش تصویر

۱.۸.۱ استخراج ویژگی‌های کلاسیک

استخراج ویژگی‌های کلاسیک شامل استفاده از تکنیک‌هایی است که ویژگی‌هایی مانند لبه‌ها، گوش‌ها و بافت‌های تصویر را شناسایی و استخراج می‌کنند. این ویژگی‌ها به مدل‌ها کمک می‌کنند تا الگوها و ساختارهای موجود در تصویر را به طور مؤثری تحلیل کنند [۱۳].

یکی از روش‌های متداول برای استخراج ویژگی‌های کلاسیک، استفاده از فیلترهای گابور^{۸۶} است که می‌توانند ویژگی‌های مکانی-فرکانسی تصویر را شناسایی کنند. همچنین، تبدیل هفت^{۸۷} یک روش دیگر برای شناسایی اشکال هندسی مانند خطوط و دایره‌ها در تصویر است.

$$G(x, y, \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda} + \psi\right)$$

که در آن:

$$x' = x \cos \theta + y \sin \theta \quad \bullet$$

$$y' = -x \sin \theta + y \cos \theta \quad \bullet$$

λ طول موج فیلتر

θ زاویه جهت‌گیری فیلتر

ψ فاز فیلتر

σ انحراف معیار گوسی

γ نسبت ابعاد

این تکنیک‌ها در مسائلی مانند شناسایی لبه‌های خیابان‌ها و مرزهای اشیاء در تصاویر ماهواره‌ای بسیار مؤثر هستند.

^{۸۶}Gabor Filters

^{۸۷}Hough Transform

۲.۸.۱ ویژگی‌های سطح بالا با استفاده از شبکه‌های عصبی

شبکه‌های عصبی، به ویژه شبکه‌های عصبی پیچشی، قادر به استخراج ویژگی‌های پیچیده‌تر و سطح بالاتری از تصاویر هستند. این ویژگی‌ها می‌توانند شامل الگوهای پیچیده‌ای مانند تقاطع‌ها، ساختارهای خیابانی و حتی تشخیص نوع مواد سطحی مانند آسفالت یا بتن باشند [۲۲].

در این شبکه‌ها، لایه‌های کانولوشن وظیفه استخراج ویژگی‌های سطح پایین (مانند لبه‌ها) و لایه‌های بالاتر وظیفه استخراج ویژگی‌های سطح بالاتر (مانند اشکال پیچیده) را بر عهده دارند.

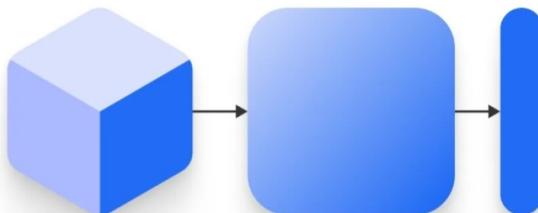
$$f(x, y) = \sum_{i=1}^m \sum_{j=1}^n w_{ij} \cdot I(x + i, y + j)$$

این روش به ویژه در مسائلی که نیاز به شناسایی الگوهای پیچیده در تصاویر دارند، مانند استخراج نقشه خیابان از تصاویر ماهواره‌ای، بسیار کاربرد دارد.

۳.۸.۱ روش‌های کاهش ابعاد

در بسیاری از مسائل پردازش تصویر، تعداد ویژگی‌ها بسیار زیاد است و این می‌تواند منجر به مشکلاتی مانند پیچیدگی محاسباتی و بیش‌بازش شود. برای مقابله با این مشکلات، از روش‌های کاهش ابعاد ^{۸۸} مانند تحلیل مولفه‌های اصلی ^{۸۹} و ^{۹۰} t-SNE استفاده می‌شود [۴۵].

Dimensionality Reduction



شکل ۱۵.۱: کاهش ویژگی‌ها سبب سرعت و آموزش بهتر مدل می‌شود.

^{۸۸}Dimensionality Reduction

^{۸۹}Principal Component Analysis - PCA

^{۹۰}t-Distributed Stochastic Neighbor Embedding

تحلیل مولفه‌های اصلی یکی از روش‌های متداول کاهش ابعاد است که با تجزیه ماتریس کوواریانس داده‌ها، ویژگی‌های اصلی را استخراج می‌کند. این ویژگی‌ها بیشترین واریانس داده‌ها را حفظ می‌کنند و به مدل‌ها اجازه می‌دهند که با تعداد کمتری از ویژگی‌ها، داده‌ها را به خوبی توصیف کنند.

$$\mathbf{C} = \frac{1}{m} \sum_{i=1}^m (\mathbf{x}^{(i)} - \mu)(\mathbf{x}^{(i)} - \mu)^T$$

که در آن:

- \mathbf{C} ماتریس کوواریانس

- $\mathbf{x}^{(i)}$ نمونه داده‌ام

- μ میانگین داده‌ها

روش‌های کاهش ابعاد به ویژه در مسائلی که با حجم زیاد داده‌های تصویری سر و کار دارند، مانند پردازش تصاویر ماهواره‌ای، بسیار مفید هستند. این روش‌ها می‌توانند با کاهش تعداد ویژگی‌ها، کارایی محاسباتی را افزایش داده و خطر بیش‌بازش را کاهش دهند.

۹.۱ پیش‌پردازش تصاویر ماهواره‌ای

۱.۹.۱ مقدمه‌ای بر تصاویر ماهواره‌ای

تصاویر ماهواره‌ای به داده‌هایی اطلاق می‌شود که از طریق حسگرهای مستقر در ماهواره‌ها جمع‌آوری می‌شوند. این تصاویر شامل اطلاعات گوناگونی از سطح زمین و محیط اطراف آن هستند و می‌توانند در کاربردهای مختلفی مانند نقشه‌برداری، کشاورزی، محیط‌زیست و نظامی مورد استفاده قرار گیرند^[۳۹].

تصاویر ماهواره‌ای به دلیل مقیاس وسیع و پوشش جهانی، یکی از منابع اصلی داده‌ها در بسیاری از تحقیقات علمی و صنعتی محسوب می‌شوند. این تصاویر اغلب نیاز به پردازش و تحلیل دارند تا اطلاعات مورد نظر استخراج شود.

۲.۹.۱ مراحل پیش‌پردازش تصاویر ماهواره‌ای

پیش‌پردازش تصاویر ماهواره‌ای به مجموعه‌ای از مراحل اشاره دارد که به منظور آماده‌سازی تصاویر برای تحلیل‌های بیشتر انجام می‌شود. این مراحل شامل تصحیح رادیومتری^{۹۱}، تصحیح هندسی^{۹۲} و فیلتر کردن نویز^{۹۳} است.

تصحیح رادیومتری

تصحیح رادیومتری به فرآیندی اطلاق می‌شود که در آن اختلافات ناشی از شرایط حسگر و تغییرات جوی در تصویر اصلاح می‌شود. این مرحله بهبود کیفیت تصویر و دقت داده‌های استخراجی را تضمین می‌کند^[۲۰].

تصحیح هندسی

تصحیح هندسی شامل اصلاح اعوجاج‌های ناشی از تغییر زاویه دید حسگر یا حرکت ماهواره است. این مرحله باعث می‌شود که تصاویر ماهواره‌ای به طور دقیق با نقشه‌های زمین مرجع هماهنگ شوند^[۳۸].

^{۹۱} Radiometric Correction

^{۹۲} Geometric Correction

^{۹۳} Noise Filtering

فیلتر کردن نویز

نویز در تصاویر ماهواره‌ای ممکن است ناشی از منابع مختلفی مانند تداخل‌های الکترونیکی یا نویزهای محیطی باشد. فیلترهای مختلفی مانند فیلتر میانگین^{۹۴}، فیلتر میانه^{۹۵} و فیلتر گوسین^{۹۶} برای کاهش نویز و بهبود کیفیت تصویر به کار می‌روند.

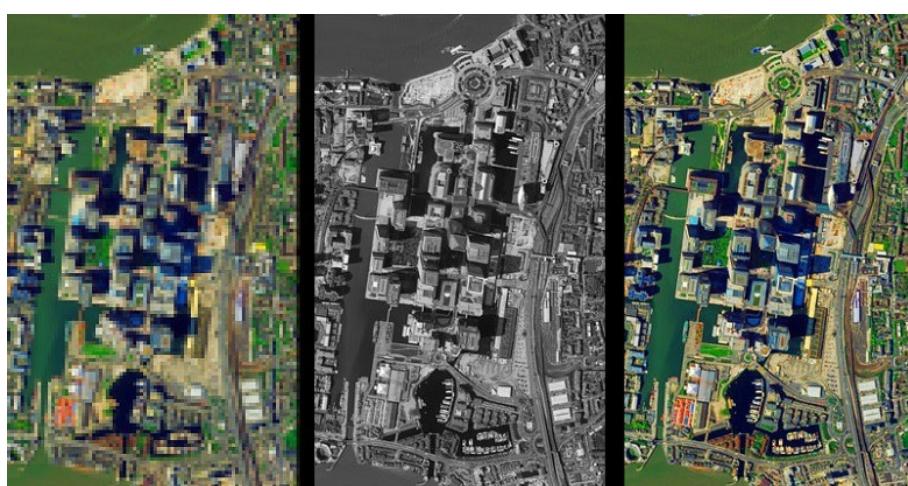
$$I_{\text{filtered}}(x, y) = \sum_{i=-k}^k \sum_{j=-k}^j G(i, j) \cdot I(x + i, y + j)$$

که در آن:

- $I_{\text{filtered}}(x, y)$ تصویر فیلتر شده در موقعیت (x, y) است.
- $G(i, j)$ ماتریس فیلتر گوسین است.
- $I(x + i, y + j)$ مقدار پیکسل در موقعیت $(x + i, y + j)$ تصویر اصلی است.

۳.۹.۱ روش‌های افزایش وضوح تصاویر ماهواره‌ای

وضوح تصاویر ماهواره‌ای به معنای دقت جزئیات تصویری است که حسگر ماهواره قادر به ثبت آن است. برای بهبود وضوح این تصاویر، از تکنیک‌های مختلفی استفاده می‌شود. یکی از این روش‌ها، تکنیک Pan-sharpening است که به ترکیب تصاویر با وضوح بالا و پایین می‌پردازد[۲۴].



شکل ۱۶.۱: بهبود وضوح تصویر با استفاده از تکنیک Pan-sharpening

^{۹۴}Mean Filter

^{۹۵}Median Filter

^{۹۶}Gaussian Filter

تکنیک Pan-sharpening با ترکیب داده‌های پانسکرونیک^{۹۷} و داده‌های چندطیفی^{۹۸}، تصویری با وضوح بالا ایجاد می‌کند که جزئیات بیشتری از سطح زمین را نشان می‌دهد.

$$I_{\text{sharpened}} = I_{\text{pan}} + \lambda(I_{\text{multi}} - I_{\text{blurred multi}})$$

که در آن:

- $I_{\text{sharpened}}$ تصویر نهایی با وضوح بالا است.
- I_{pan} تصویر پانسکرونیک است.
- I_{multi} تصویر چندطیفی است.
- $I_{\text{blurred multi}}$ تصویر چندطیفی با فیلتر تاری است.
- λ ضریبی برای تنظیم میزان ترکیب است.

این روش‌ها به ویژه در کاربردهایی که نیاز به تحلیل دقیق جزئیات زمین دارند، مانند نقشه‌برداری و نظارت بر تغییرات محیطی، بسیار موثر هستند.

۱۰.۱ بیش‌برازش

۱۰.۱.۱ تعریف و مفهوم بیش‌برازش

بیش‌برازش^{۹۹} زمانی اتفاق می‌افتد که مدل یادگیری ماشین بیش از حد با داده‌های آموزشی انطباق پیدا می‌کند و به جای یادگیری الگوهای عمومی، جزئیات و نویزهای موجود در داده‌های آموزشی را یاد می‌گیرد[۱۷]. این وضعیت معمولاً منجر به عملکرد ضعیف مدل در مواجهه با داده‌های جدید و نادیده گرفته شده می‌شود.

در فرآیند آموزش مدل، هدف اصلی یافتن مدلی است که قادر باشد به خوبی تعمیم‌دهی کند؛ یعنی در داده‌های جدید نیز دقت بالایی داشته باشد. اما در حالت بیش‌برازش، مدل به دلیل یادگیری بیش از حد جزئیات داده‌های آموزشی، نمی‌تواند به خوبی تعمیم دهد.

^{۹۷}Panchromatic

^{۹۸}Multispectral

^{۹۹}Overfitting

۲.۱۰.۱ راهکارهای جلوگیری از بیشبرازش

برای جلوگیری از بیشبرازش، چندین روش وجود دارد که می‌توانند به بهبود توانایی تعمیم‌دهی مدل کمک کنند:

استفاده از داده‌های بیشتر

یکی از راهکارهای ساده برای کاهش خطر بیشبرازش، استفاده از داده‌های آموزشی بیشتر است. با افزایش حجم داده‌ها، مدل به جای یادگیری جزئیات خاص، الگوهای عمومی‌تر را یاد می‌گیرد و احتمال بیشبرازش کاهش می‌یابد.

اعمال منظم‌کننده‌ها

منظم‌کننده‌ها^{۱۰۰} به مدل کمک می‌کنند که از یادگیری وزن‌های بیش از حد بزرگ خودداری کند و به جای آن، مدل ساده‌تری را انتخاب کند که توانایی تعمیم‌دهی بهتری داشته باشد. منظم‌کننده‌های L1 و L2 دو نوع رایج از منظم‌کننده‌ها هستند که به صورت زیر تعریف می‌شوند:

$$L1 : \lambda \sum_{i=1}^n |w_i|$$

$$L2 : \lambda \sum_{i=1}^n w_i^2$$

که در آن λ ضریب منظم‌کننده است.

استفاده از تکنیک حذف کردن

تکنیک حذف^{۱۰۱} یکی دیگر از روش‌های جلوگیری از بیشبرازش است که به صورت تصادفی بخشی از نرون‌ها را در هر مرحله از آموزش مدل غیرفعال می‌کند. این کار باعث می‌شود که مدل نتواند به وابستگی زیاد به یک نرون یا مجموعه‌ای از نرون‌ها در طول آموزش داشته باشد و به این ترتیب تعمیم‌دهی بهتری پیدا کند.

$$\text{DropoutRate} = p$$

که در آن p درصد نرون‌هایی است که به صورت تصادفی غیرفعال می‌شوند.

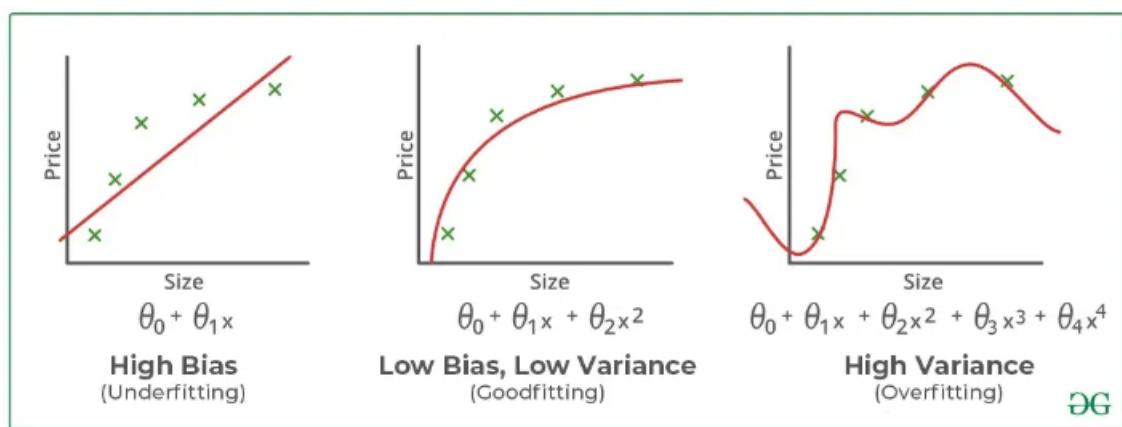
^{۱۰۰}Regularization

^{۱۰۱}Dropout

۳.۱۰.۱ تأثیر بیش‌بازش در پروژه‌های پردازش تصویر

در پروژه‌های پردازش تصویر، بیش‌بازش می‌تواند منجر به عملکرد ضعیف مدل‌ها در شناسایی یا دسته‌بندی تصاویر جدید شود. برای مثال، اگر مدلی برای استخراج نقشه خیابان‌ها از تصاویر ماهواره‌ای بیش‌بازش کند، ممکن است در تشخیص خیابان‌های جدید که قبلاً در داده‌های آموزشی نبوده‌اند، دچار خطأ شود.

برای مقابله با این مشکل، استفاده از تکنیک‌هایی مانند داده‌افزایی^{۱۰۲}، تنظیم مناسب هایپرپارامترها^{۱۰۳}، و استفاده از مجموعه داده‌های مختلف برای آموزش مدل توصیه می‌شود.



شکل ۱۷.۱: نمودار نشان‌دهنده عملکرد مدل در حالت بیش‌بازش و حالت بهینه.

^{۱۰۲}Data Augmentation

^{۱۰۳}Hyperparameter Tuning

۱۱.۱ بهینه‌سازی فرآپارامترها

فرآپارامترها^{۱۰۴} پارامترهایی هستند که در طول فرآیند آموزش مدل تنظیم نمی‌شوند، بلکه باید پیش از آموزش تنظیم شوند. این پارامترها شامل تنظیمات شبکه عصبی، بهینه‌سازی و ساختار مدل هستند. برخی از مهم‌ترین های پارامترهایی که باید بهینه شوند عبارت‌اند از:

- نرخ یادگیری^{۱۰۵}: نرخ یادگیری تعیین می‌کند که مدل با چه سرعتی وزن‌ها را در طول آموزش به روزرسانی کند. نرخ یادگیری خیلی بالا ممکن است مدل را به نوسان بیندازد و از همگرایی جلوگیری کند، در حالی که نرخ یادگیری خیلی پایین باعث می‌شود که آموزش بسیار کند شود.
- ابعاد دسته‌ها^{۱۰۶}: اندازه دسته تعیین می‌کند که در هر مرحله از آموزش، چه تعداد نمونه به مدل وارد شود. این مقدار بر حافظه، زمان محاسباتی و دقت مدل تأثیر می‌گذارد.
- نرخ رهاسازی^{۱۰۷}: نرخ رهاسازی برای جلوگیری از بیش‌برازش^{۱۰۸} استفاده می‌شود و به مدل کمک می‌کند تا تعمیم‌پذیری بهتری داشته باشد.
- تعداد دوره‌های آموزشی^{۱۰۹}: تعداد کل تکرارهای آموزش بر روی داده‌های آموزشی است. انتخاب تعداد مناسب epoch‌ها برای جلوگیری از بیش‌برازش و افزایش دقت مدل ضروری است.

۱.۱۱.۱ روش‌های بهینه‌سازی هایپرپارامترها

جستجوی شبکه‌ای

در روش جستجوی شبکه‌ای^{۱۱۰}، مجموعه‌ای از مقادیر ممکن برای هر هایپرپارامتر تعریف می‌شود و سپس مدل با تمامی ترکیبات ممکن از این مقادیر اجرا می‌شود. اگرچه این روش ساده و پرکاربرد است، اما به دلیل نیاز به ارزیابی همه ترکیبات ممکن، ممکن است زمان بر باشد.

جستجوی تصادفی

در روش جستجوی تصادفی^{۱۱۱} به طور تصادفی مقادیر هایپرپارامترها انتخاب می‌شود. این روش معمولاً سریع‌تر از جستجوی شبکه‌ای است و به نتیجه‌های مشابهی در زمان کمتری دست می‌یابد.

^{۱۰۴}Hyper parameter

^{۱۰۵}Learning Rate

^{۱۰۶}Batch Size

^{۱۰۷}Dropout Rate

^{۱۰۸}Overfitting

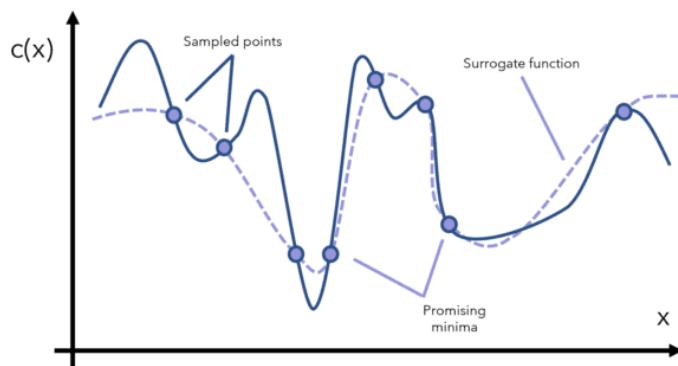
^{۱۰۹}Epoch

^{۱۱۰}Grid Search

^{۱۱۱}Random Search

بهینه‌سازی بیزی

بهینه‌سازی بیزی^{۱۱۲} از مدل‌های احتمالی برای پیش‌بینی عملکرد مدل استفاده می‌کند و با هر بار آموزش مدل، اطلاعات بیشتری در مورد هایپرپارامترهای مناسب به دست می‌آورد. این روش در مواردی که هزینه اجرای مدل بالا است، بسیار کارآمد است.



شکل ۱۸.۱: با استفاده از قطع کردن تابع $C(x)$ با تابع هدف محدوده نقاط کمینه به دست می‌آید.

تنظیم گرادیان چند مرحله‌ای

روش تنظیم گرادیان چند مرحله‌ای^{۱۱۳} یکی از روش‌های مدرن بهینه‌سازی است که ترکیبی از روش‌های جستجوی تصادفی و تنظیم زمان آموزش است. این روش به طور کارآمد هایپرپارامترهای مناسب را انتخاب می‌کند و از زمان کمتری برای آموزش استفاده می‌کند.

۲.۱۱.۱ کاربرد در استخراج جاده از تصاویر ماهواره‌ای

در پژوهشی استخراج جاده از تصاویر ماهواره‌ای، بهینه‌سازی هایپرپارامترها می‌تواند منجر به بهبود عملکرد مدل در بخش‌بندی جاده‌ها شود. برخی از هایپرپارامترهای کلیدی که می‌توانند برای پژوهه بهینه شوند، شامل موارد زیر هستند:

- نرخ یادگیری: با استفاده از جستجوی شبکه‌ای، نرخ یادگیری مناسب را انتخاب کنید. مقادیر پیشنهادی برای شروع: $0.0001, 0.001, 0.01$ و 0.1 .
- اندازه دسته‌ها: اندازه دسته‌های $16, 32$ و 64 را برای ارزیابی زمان و دقت مدل استفاده کنید.
- نرخ رهاسازی: از نرخ رهاسازی‌های $0.003, 0.005$ و 0.007 برای جلوگیری از بیش‌برازش استفاده کنید.
- تعداد epoch‌ها: تعداد epoch‌های $50, 100$ و 150 را برای ارزیابی همگرای مدل انتخاب کنید.

^{۱۱۲}Bayesian Optimization

^{۱۱۳}Hyperband

۳.۱۱.۱ نتیجه‌گیری

بهینه‌سازی های پارامترها نقش مهمی در بهبود دقت و کارایی مدل در پروژه استخراج جاده از تصاویر ماهواره‌ای دارد. با استفاده از روش‌های مختلف بهینه‌سازی، مانند جستجوی شبکه‌ای، تصادفی و بیزی، می‌توان پارامترهایی مانند نرخ یادگیری، اندازه دسته‌ها و نرخ رهاسازی را برای بهبود عملکرد پروژه تنظیم کرد. این بهینه‌سازی‌ها بهویژه در پروژه‌هایی که با داده‌های تصویری حجمی و پیچیده سر و کار دارند، مانند استخراج جاده از تصاویر ماهواره‌ای، می‌توانند تأثیر قابل توجهی بر دقت نهایی داشته باشند.

۱۲.۱ خلاصه و جمع‌بندی

در این فصل، مفاهیم و تکنیک‌های مختلف مرتبط با پردازش تصویر، یادگیری ماشین، یادگیری عمیق، و کاربرد آن‌ها در استخراج نقشه خیابان از تصاویر ماهواره‌ای بررسی شد. این فصل به طور کلی نتایج مهم هر یک از فصل‌ها را مرور و جمع‌بندی می‌کند.

۱.۱۲.۱ مروری بر مبانی پردازش تصویر

در قسمت اول، مبانی پردازش تصویر مورد بررسی قرار گرفت. تعریف تصویر دیجیتال و تکنیک‌های پایه‌ای مانند فیلتر کردن، آشکارسازی لبه‌ها، و حذف نویز از جمله مباحث مطرح شده بودند. این مفاهیم اساسی برای درک بهتر فرآیندهای پیچیده‌تر پردازش تصویر که در فصل‌های بعدی مورد بحث قرار گرفتند، ضروری بودند.

۲.۱۲.۱ یادگیری ماشین و الگوریتم‌های آن

قسمت دوم به یادگیری ماشین و انواع الگوریتم‌های آن اختصاص داشت. مبانی یادگیری ماشین و تاریخچه آن، همچنین الگوریتم‌های پرسپترون، ماشین‌های بردار پشتیبان، (SVM) و تحلیل مولفه‌های اصلی (PCA) به تفصیل توضیح داده شدند. این الگوریتم‌ها به عنوان پایه‌های اولیه برای درک روش‌های پیچیده‌تر یادگیری عمیق در فصل‌های بعدی مورد استفاده قرار گرفتند.

۳.۱۲.۱ یادگیری عمیق و شبکه‌های عصبی

بخش سوم به بررسی یادگیری عمیق و شبکه‌های عصبی اختصاص یافت. تفاوت یادگیری ماشین و یادگیری عمیق، و نیز شبکه‌های عصبی مصنوعی، شبکه‌های عصبی پیچشی، و شبکه‌های عصبی بازگشته به همراه واحدهای حافظه طولانی مدت به تفصیل مورد بحث قرار گرفتند. این مباحث اهمیت یادگیری عمیق در پردازش تصویر را روشن کردند.

۴.۱۲.۱ مدل‌های ترانسفورمر در یادگیری عمیق

در بخش چهارم، مدل‌های ترانسفورمر و به طور خاص مدل ترانسفورمر چرخشی که برای پردازش تصاویر طراحی شده‌اند، بررسی شدند. این مدل‌ها به دلیل قابلیت‌های ایشان در یادگیری روابط پیچیده بین اجزای تصویر و پردازش موازی داده‌ها، به عنوان یکی از مهم‌ترین ابزارها در بینایی کامپیوترا مطرح شده‌اند.

۵.۱۲.۱ استخراج ویژگی‌ها و کاهش ابعاد

قسمت پنجم به روش‌های مختلف استخراج ویژگی‌ها و کاهش ابعاد در پردازش تصویر پرداخت. روش‌های کلاسیک مانند فیلترهای گابور و تبدیل هافت، به همراه تکنیک‌های پیشرفته‌تر که از شبکه‌های عصبی برای استخراج ویژگی‌های سطح بالا استفاده می‌کنند، مورد بررسی قرار گرفتند. همچنین روش‌های کاهش ابعاد مانند PCA برای مقابله با پیچیدگی داده‌ها و جلوگیری از بیش‌برازش معرفی شدند.

۶.۱۲.۱ پیش‌پردازش تصاویر ماهواره‌ای

در بخش ششم، مراحل پیش‌پردازش تصاویر ماهواره‌ای شامل تصحیح رادیومتری، تصحیح هندسی، و فیلتر کردن نویز بررسی شدند. این مراحل به عنوان گام‌های اولیه و ضروری برای آماده‌سازی تصاویر ماهواره‌ای جهت تحلیل‌های دقیق‌تر مطرح شدند. همچنین، روش‌های افزایش وضوح تصاویر مانند تکنیک Pan-sharpening معرفی شدند.

۷.۱۲.۱ بیش‌برازش و راهکارهای مقابله با آن

قسمت هفتم به بررسی پدیده بیش‌برازش و راهکارهای مقابله با آن اختصاص داشت. روش‌های مختلفی برای جلوگیری از بیش‌برازش معرفی شدند که شامل استفاده از داده‌های بیشتر، منظم‌کننده‌ها، و تکنیک حذف کردن بود. این راهکارها به بهبود توانایی تعمیم‌دهی مدل‌ها و جلوگیری از افت عملکرد آن‌ها در مواجهه با داده‌های جدید کمک می‌کنند.

۸.۱۲.۱ نتیجه‌گیری کلی

این فصل با ارائه‌ی یک دید جامع از تکنیک‌ها و مدل‌های پیشرفته‌پردازش تصویر و یادگیری ماشین، به ویژه در زمینه استخراج نقشه خیابان از تصاویر ماهواره‌ای، تلاش کرده است تا راهکارهای مناسبی برای حل مسائل پیچیده در این حوزه ارائه دهد. هر یک از بخش‌ها به طور خاص به جنبه‌های مختلفی از این مسئله پرداخته‌اند و در نهایت، مجموعه‌ای از روش‌ها و تکنیک‌ها معرفی شده‌اند که می‌توانند به بهبود دقت و کارایی در استخراج نقشه خیابان‌ها کمک کنند.

۲ فصل

پیش‌نیازها و مفاهیم تخصصی

در این فصل سه منطق شناختی پویا را با این رویکرد مطرح می‌کنیم که برای هریک اصولی موسوم به اصول موضوعی تحويل^۱ معرفی کرده و با اثبات صحت آنها گامی به سوی تمامیت بر می‌داریم. در انتهای فصل نیز تمامیت را در یک قضیه برای هر سه منطق اثبات خواهیم کرد.

۱.۲ شبکه‌های عصبی پیچشی

شبکه‌های عصبی پیچشی نوعی از شبکه‌های عصبی مصنوعی هستند که به‌ویژه برای پردازش داده‌های تصویری طراحی شده‌اند. این شبکه‌ها از لایه‌های کانولوشن برای استخراج ویژگی‌های فضایی و مکانی از تصاویر استفاده می‌کنند و معمولاً در کاربردهای بینایی ماشین و پردازش تصویر بسیار موفق هستند.

۱.۱.۲ ساختار لایه‌های کانولوشن

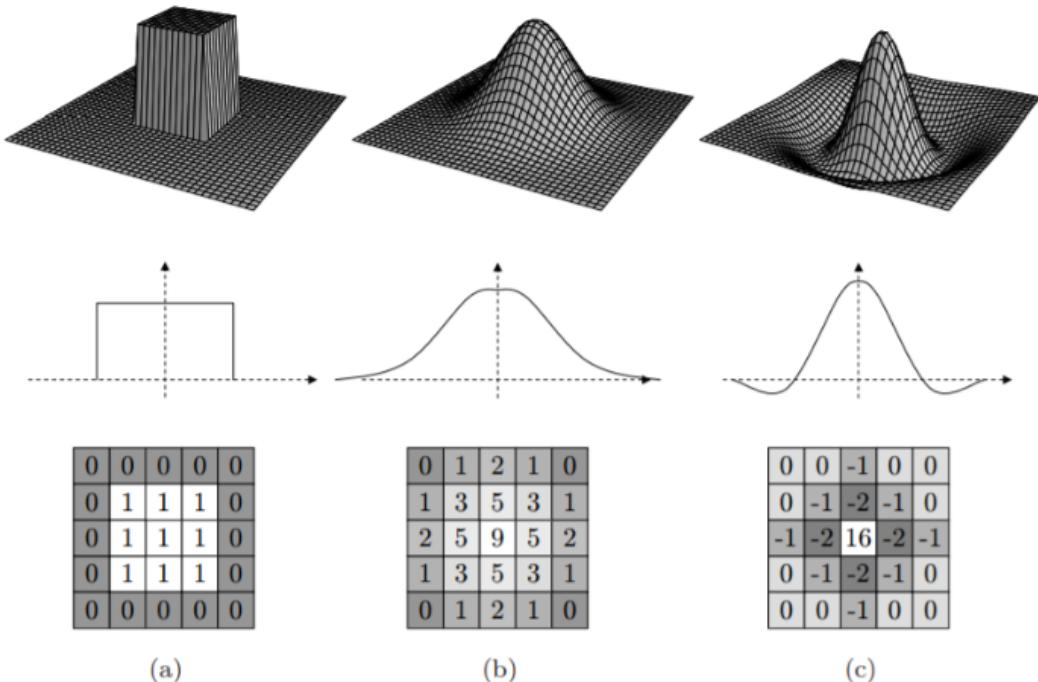
لایه‌های پیچشی اصلی‌ترین بخش شبکه‌های عصبی پیچشی هستند که با اعمال فیلترهای مختلف (هسته‌های کانولوشن) بر روی تصویر ورودی، ویژگی‌های مختلفی از تصویر را استخراج می‌کنند.

۱. هسته یا فیلتر:

هسته یک ماتریس از اعداد است که در طول عملیات کانولوشن بر روی تصویر (یا ورودی چند بعدی دیگر) استفاده می‌شود. این هسته به طور معمول به صورت مربع یا مستطیلی است و ابعاد آن معمولاً از تعداد ابعاد ورودی تعیین می‌شود. به عنوان مثال، برای تصاویر هسته‌ها معمولاً دو بعدی هستند.^[۵۰]

۲. عملگر کانولوشن:

^۱ reduction axioms



شکل ۱.۲: نمونه‌ای از فیلترهای مختلف و تاثیر آن‌ها

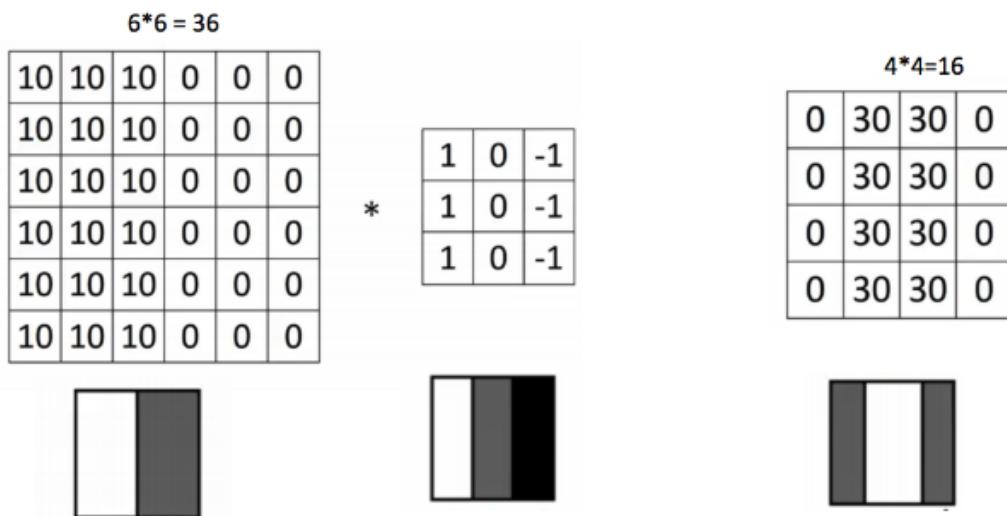
این عملگر بر روی تصویر ورودی با استفاده از کرنل اعمال می‌شود. در هر مرحله، کرنل روی یک قسمت از تصویر (یا ورودی) اعمال می‌شود و یک مقدار خروجی تولید می‌کند که به عنوان ویژگی مورد نظر استفاده می‌شود. این عملگر عمدها شامل ضرب نقطه‌ای بین عناصر کرنل و مقادیر متناظر در تصویر ورودی است.^[۸]

۳. یارامترهای لایه:

پارامترهای اصلی شامل ابعاد هسته (مانند ارتفاع و عرض)، تعداد هسته‌ها (برای لایه‌های چند کاناله) و نوع عملیات (مانند مقداردهی اولیه و توابع فعال‌سازی) هستند. این پارامترها به طور مستقیم تعیین می‌کنند که لایه پیچشی چگونه عمل می‌کند و ویژگی‌های چه نوعی از داده‌های ورودی استخراج می‌کند [۵۰].

۴. خروجی لایه پیچشی:

خروجی لایه پیچشی یک نوع تصویر (یا داده چند بعدی) است که به عنوان ورودی به لایه‌های بعدی در شبکه عصبی پیچشی می‌رود. این خروجی شامل ویژگی‌های استخراج شده است که برای وظایف مختلفی مانند تشخیص الگو، تصویربرداری، تفسیر ورودی و غیره مفید هستند.^[۸]

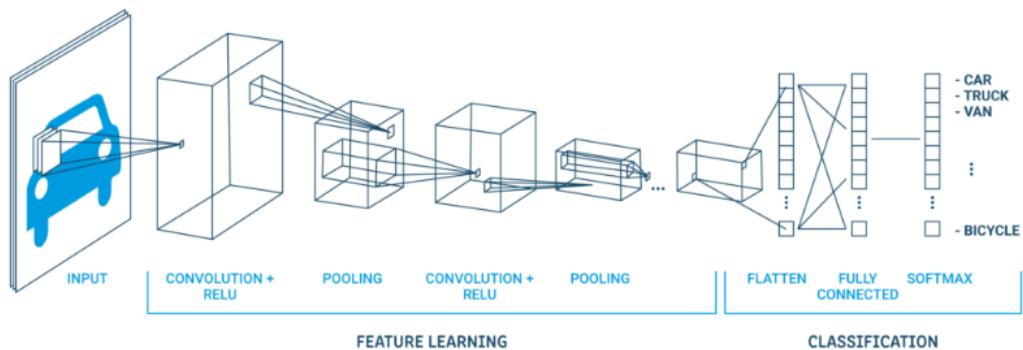
۵. نقشه‌ی ویژگی^۲ یا نقشه‌ی فعال‌سازی^۳:

شکل ۲.۲: نمونه‌ای از فیلترهای مختلف و تاثیر آن‌ها

کانولوشن به‌شکل ساده به معنای ضرب نقطه‌ای دوتابع برای تولید تابع سوم است. در اینجا ۲.۲ یک تابع ماتریس، پیکسل‌های تصویر ما و دیگری فیلتر ما است. فیلتر را روی تصویر می‌کشیم و حاصل ضرب نقطه‌ای دو ماتریس را می‌گیریم. ماتریس حاصل را "نقشه‌ی فعال‌سازی" یا "نقشه‌ی ویژگی" می‌نامند.

^۱Feature Map

^۲Activation Map



شکل ۳.۲: مثال ساختار شبکه‌های عصبی پیچشی

۲.۱.۲ لایه‌های تجمعی

لایه‌های تجمعی^۴ در شبکه‌های عصبی پیچشی برای کاهش ابعاد فضایی (ارتفاع و عرض) خروجی لایه‌های پیچشی استفاده می‌شوند، در حالی که ویژگی‌های مهم را حفظ می‌کنند. دو نوع رایج آن‌ها تجمعی بیشینه^۵ و تجمعی میانگین^۶ هستند[۷]. این کار باعث کاهش تعداد پارامترها و محاسبات در شبکه می‌شود و همچنین کمک می‌کند که مدل به ویژگی‌های مکانی محلی حساسیت کمتری داشته باشد. انواع لایه‌های تجمعی به شرح زیر است:

۱. تجمعی بیشینه:

در این روش، بیشترین مقدار از هر ناحیه‌ای که توسط فیلتر تجمعی پوشش داده می‌شود انتخاب می‌شود[۵۰].

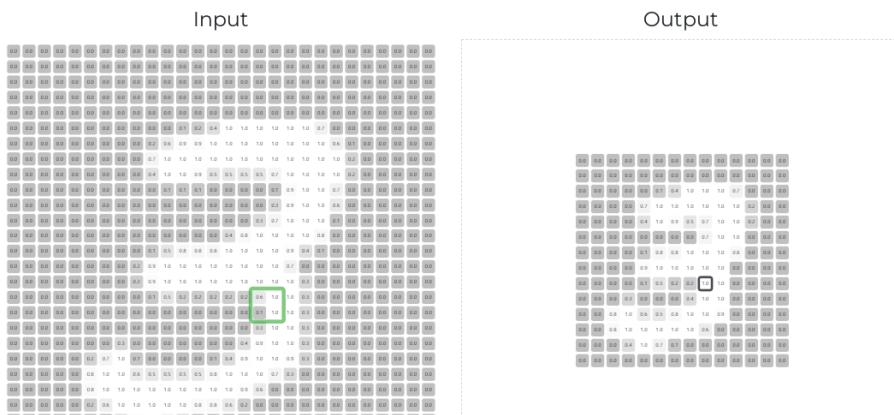
۲. تجمعی میانگین:

در این روش، میانگین مقادیر در هر ناحیه‌ای که توسط فیلتر تجمعی پوشش داده می‌شود محاسبه می‌شود[۸].

^۴Pooling Layers

^۵Max Pooling

^۶Average Pooling



شکل ۴.۲:

در تصویر ۴.۲، عملکرد لایه تجمعی بیشینه نمایش داده شده است. فیلتر 2×2 بر روی ورودی حرکت می‌کند و بیشترین مقدار در هر ناحیه را انتخاب می‌کند تا خروجی تجمعی شده را تولید کند.

پارامترهای لایه‌های تجمعی عبارتنداز:

- ابعاد فیلتر^۷:

ابعاد فیلتر تعیین می‌کند که چند پیکسل در هر بعد برای هر عملیات تجمعی انتخاب شود^[۵۰].

- پوشش^۸:

فاصله‌ای که فیلتر در هر مرحله حرکت می‌کند^[۸].

- لبه‌گذاری^۹:

مقدار پدینگ اضافه شده به مرزهای ورودی برای حفظ اندازه ورودی^[۵۰].

^۷Filter Size

^۸Stride

^۹Padding

۳.۱.۲ لایه‌های تماماً متصل

لایه‌های تماماً متصل^{۱۰}، که به آنها لایه‌های متراکم^{۱۱} نیز گفته می‌شود، یکی از اجزای اساسی شبکه‌های عصبی مصنوعی^{۱۲} هستند. این لایه‌ها در انتهای شبکه‌های عصبی پیچشی و بسیاری از معماری‌های دیگر استفاده می‌شوند تا خروجی نهایی را تولید کنند. ساختار لایه‌های تماماً متصل:

۱. هر راس (نورون) در یک لایه تماماً متصل به تمام نودهای لایه قبلی متصل است.^[۵۰]

۲. وزن‌ها و سوگیری^{۱۳}‌ها:

هر اتصال بین راس‌ها دارای یک وزن است که در طول آموزش تنظیم می‌شود. هر راس همچنین دارای یک سوگیری است که به مجموع ورودی‌ها اضافه می‌شود.^[۸]

۳. تابع فعال‌سازی^{۱۴}: خروجی هر راس پس از اعمال تابع فعال‌سازی (مانند رلو^{۱۵} یا سیگموید^{۱۶}) محاسبه می‌شود.^[۵۰]

عملکرد لایه‌های تماماً متصل:

۱. تابع فعال‌سازی:

جمع وزن‌دار ورودی‌ها از طریق تابع فعال‌سازی عبور می‌کند تا خروجی نهایی راس محاسبه شود.

۲. جمع وزن‌دار ورودی‌ها:

هر راس خروجی از طریق جمع وزن‌دار ورودی‌ها به دست می‌آید.

۴.۱.۲ کاربرد شبکه‌های عصبی پیچشی در استخراج جاده از تصاویر ماهواره‌ای

شبکه‌های عصبی پیچشی در استخراج جاده‌ها از تصاویر ماهواره‌ای به‌طور گسترده‌ای استفاده می‌شوند. فرآیند کلی شامل چند مرحله اساسی است که در ادامه توضیح داده شده‌اند.

۱. پیش‌پردازش تصویر:

تصاویر ماهواره‌ای قبل از ورود به شبکه عصبی پیچشی باید پیش‌پردازش شوند. این شامل اصلاح رنگ و روشنایی، کاهش نویز و تنظیم مقیاس تصویر است.^[۳۶]

^{۱۰}Fully Connected Layers

^{۱۱}Dense Layers

^{۱۲}Artificial Neural Networks

^{۱۳}Bias

^{۱۴}Activation function

^{۱۵}ReLU

^{۱۶}Sigmoid

۲. شبکه عصبی پیچشی برای بخش‌بندی تصاویر:

از این شبکه‌ها برای بخش‌بندی تصاویر ماهواره‌ای به دو دسته (جاده و پس‌زمینه) استفاده می‌شود. این کار با استفاده از لایه‌های پیچشی و تجمعی انجام می‌شود تا ویژگی‌های مکانی جاده‌ها استخراج شود [۵۰].

۳. آموزش مدل:

مدل شبکه عصبی پیچشی با استفاده از مجموعه‌ای از تصاویر آموزشی که شامل برچسب‌های صحیح (جاده و پس‌زمینه) هستند، آموزش داده می‌شود [۴۳].

۴. پیش‌بینی و استخراج جاده:

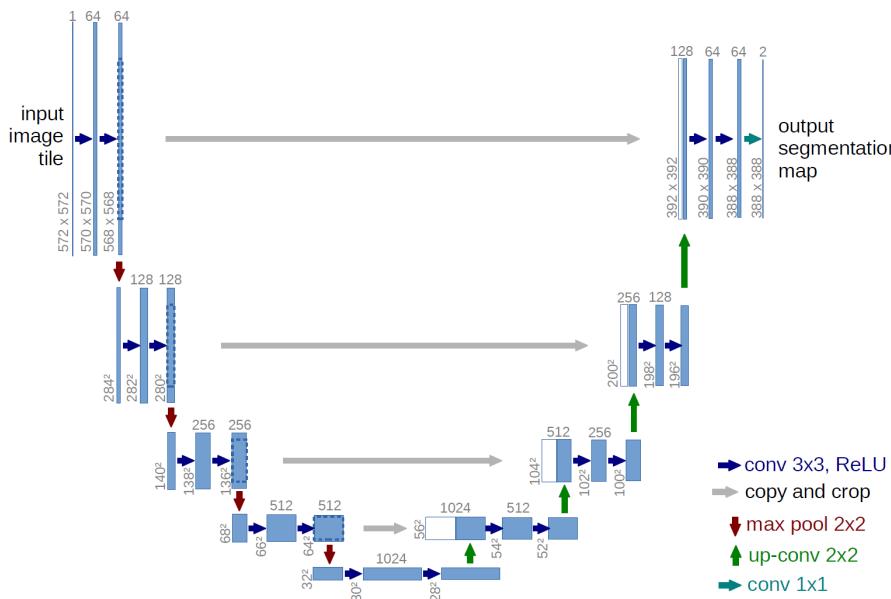
پس از آموزش، مدل قادر است جاده‌ها را در تصاویر ماهواره‌ای جدید شناسایی کند. خروجی مدل به صورت نقشه‌ای از پیکسل‌های جاده و پس‌زمینه است [۳۷].

۵. شبکه‌های عصبی پیچشی برای استخراج مسیر از تصاویر ماهواره‌ای

شبکه‌های عصبی پیچشی به دلیل توانایی بالای خود در شناسایی الگوهای پیچیده در تصاویر، به طور گسترده‌ای برای استخراج مسیرها از تصاویر ماهواره‌ای استفاده می‌شوند. در ادامه چندین مثال کاربردی استفاده از شبکه‌های عصبی پیچشی در این زمینه ارائه شده است.

۱. استفاده از U-Net برای بخش‌بندی جاده‌ها:

این مدل، که یک نوع معماری شبکه‌ی عصبی پیچشی است، به طور موافقی آمیزی برای بخش‌بندی جاده‌ها از تصاویر ماهواره‌ای استفاده می‌کند. این مدل با داشتن ساختار رمزگذار-رمزگشایانه^{۱۷}، قادر به یادگیری ویژگی‌های دقیق و پیچیده تصاویر می‌باشد.^[۳۴]



شکل ۵.۲: ساختار معماری مدل U-Net

۲. استفاده از شبکه SegNet برای استخراج جاده‌ها:

شبکه SegNet که برای بخش‌بندی معنایی طراحی شده است، برای استخراج جاده‌ها از تصاویر ماهواره‌ای استفاده می‌شود. این شبکه با استفاده از لایه‌های پیچشی و دیکدرهای مربوطه، ویژگی‌های جاده‌ها را به خوبی شناسایی می‌کند.^[۴۸]

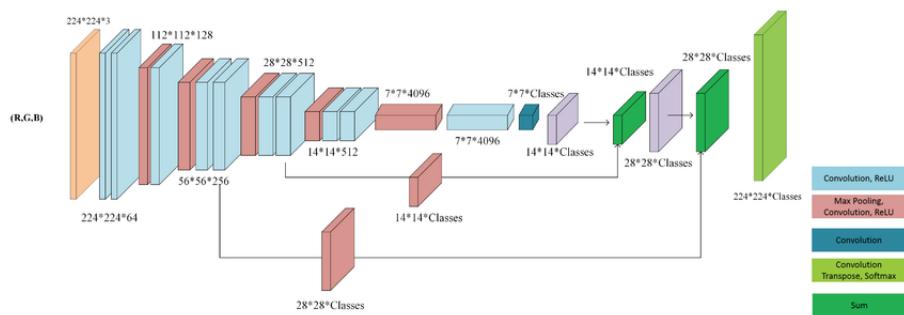
۳. استفاده از شبکه شبکه عصبی تماماً متصل^{۱۸}:

شبکه عصبی تماماً متصل^{۱۹} یکی دیگر از معماری‌های موثر برای بخش‌بندی تصاویر است که می‌تواند برای استخراج جاده‌ها از تصاویر ماهواره‌ای به کار رود. این شبکه به جای استفاده از لایه‌های تماماً متصل، از لایه‌های پیچشی به صورت کامل استفاده می‌کند تا اطلاعات مکانی به خوبی حفظ شود.^[۲۱]

^{۱۷}encoder-decoder

^{۱۸}Fully Convolutional Networks

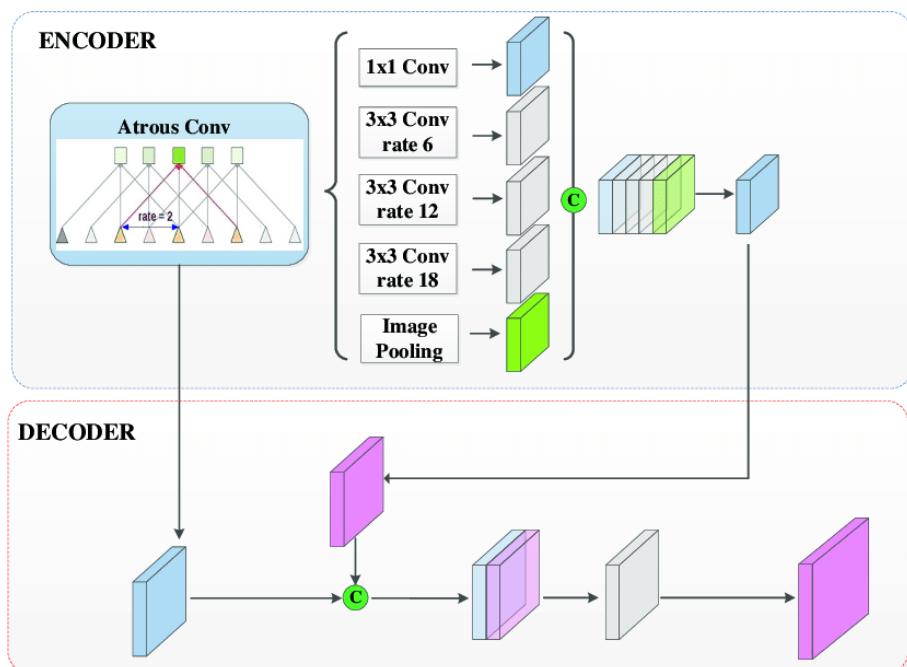
^{۱۹}FCN



شکل ۶.۲: ساختار معماری مدل شبکه عصبی تماماً متصل

۴. استفاده از DeepLab برای استخراج جاده‌ها:

این مدل با استفاده از لایه‌های کانولوشن با نرخ گسترش^{۲۰} و CRF^{۲۱} برای بهبود دقیق بخش‌بندی، در استخراج جاده‌ها از تصاویر ماهواره‌ای بسیار موثر بوده است [۲۵].



شکل ۷.۲: ساختار معماری مدل DeepLab

۵. ترکیب شبکه‌ی عصبی پیچشی با LSTM^{۲۲} برای استخراج جاده‌ها:

ترکیب شبکه‌های عصبی پیچشی با شبکه‌های عصبی حافظه‌ی کوتاه مدت بلند^{۲۳} برای استخراج جاده‌ها از تصاویر ماهواره‌ای استفاده شده است. این ترکیب به دلیل توانایی این شبکه‌ها در حفظ

^{۲۰}Atrous

^{۲۱}Conditional Random Fields

^{۲۲}Long short-term memory

^{۲۳}LSTM

اطلاعات طولانی‌مدت، به بهبود دقت بخش‌بندی جاده‌ها کمک می‌کند.^{۳۱}

۶. استفاده از شبکه عصبی باقی‌مانده برای استخراج جاده‌ها:

شبکه عصبی باقی‌مانده^{۲۴} که برای بخش‌بندی معنایی طراحی شده است، برای استخراج جاده‌ها از تصاویر ماهواره‌ای استفاده می‌شود. این شبکه با استفاده از لایه‌های پیچشی و کدگشایی مربوطه، ویژگی‌های جاده‌ها را به خوبی شناسایی می‌کند.^{۴۸}

۲.۲ الگوریتم شبکه‌های باقی‌مانده

^{۲۵} (شبکه‌های باقی‌مانده) یکی از معروف‌ترین و پراستفاده‌ترین معماری‌های شبکه عصبی عمیق است که برای بهبود دقت و کاهش مشکلات آموزش شبکه‌های عمیق معرفی شد. این معماری شبکه‌های عصبی باقی‌مانده توانست جایگاه اول در رقابت‌های COCO و ImageNet را به دست آورد.

۱.۲.۲ چالش شبکه‌های عمیق

یکی از مشکلات اساسی شبکه‌های عصبی عمیق این است که با افزایش عمق شبکه، ممکن است دقت شبکه کاهش پیدا کند یا به اصطلاح شبکه بیش‌برازش شود و به خوبی نتواند اطلاعات ورودی را تعمیم دهد. همچنین، شبکه‌های عمیق دچار مشکلاتی مانند نابودی گرادیان^{۲۶} می‌شوند، به این معنی که گرادیان در لایه‌های ابتدایی شبکه به صفر نزدیک می‌شود و شبکه نمی‌تواند به خوبی یادگیری کند.

۲.۲.۲ ایده اصلی

شبکه‌های عصبی باقی‌مانده به‌طور کلی این مشکل را با معرفی یک مفهوم به نام اتصال باقی‌مانده^{۲۷} حل می‌کنند. ایده اصلی این است که به جای یادگیری تابع اصلی، شبکه یک تابع باقی‌مانده^{۲۸} یاد می‌گیرد. در این معماری، هر بلوک شبکه‌ی عصبی باقی‌مانده، تابع هویت^{۲۹} را به ورودی اضافه می‌کند. این بدان معنی است که ورودی لایه‌های ابتدایی شبکه، به لایه‌های بعدی شبکه منتقل می‌شود و از این طریق شبکه می‌تواند به‌طور مؤثری لایه‌های عمیق‌تر را آموزش دهد.

^{۲۴} ResNet

^{۲۵} Residual Network (ResNet)

^{۲۶} Vanishing Gradient

^{۲۷} Residual Connection

^{۲۸} Residual Function

^{۲۹} Identity Mapping

ساختار کلی شبکه‌های عصبی باقی‌مانده

معماری شبکه‌های عصبی باقی‌مانده یکی از معماری‌های برجسته در یادگیری عمیق است که به دلیل توانایی آن در حل مشکلات مرتبط با افزایش عمق شبکه، به طور گسترده‌ای در پروژه‌های مختلف استفاده می‌شود. ساختار کلی شبکه‌های عصبی باقی‌مانده از بلوک‌های باقی‌مانده^{۳۰} تشکیل شده است که هدف آن‌ها یادگیری یکتابع باقی‌مانده به جای تابع اصلی است.

در شبکه‌های عمیق سنتی، با افزایش عمق شبکه، مشکلاتی نظیر نابودی گرادیان^{۳۱} یا انفجار گرادیان^{۳۲} رخ می‌دهد. این مشکلات باعث می‌شود که شبکه‌های خیلی عمیق نتوانند به طور مؤثر آموزش داده شوند. شبکه‌های عصبی باقی‌مانده با استفاده از اتصالات باقی‌مانده^{۳۳} این مشکلات را حل می‌کند.

بلوک باقی‌مانده: هر بلوک باقی‌مانده^{۳۴} در شبکه‌های عصبی باقی‌مانده شامل دو یا سه لایه‌ی پیچشی است که با استفاده از یک مسیر میانبر^{۳۵} به ورودی متصل می‌شوند. عملیات اصلی در این بلوک به صورت زیر انجام می‌شود:

$$y = f(x, \{W_i\}) + x$$

که در آن:

- x ورودی به بلوک است.
- $f(x, \{W_i\})$ تابع باقی‌مانده است که توسط لایه‌های پیچشی با وزن‌های $\{W_i\}$ محاسبه می‌شود. این تابع می‌تواند شامل عملیات‌های غیرخطی مانند ReLU^{۳۶} نیز باشد.
- y خروجی نهایی بلوک باقی‌مانده است که از جمع تابع باقی‌مانده و ورودی اولیه x به دست می‌آید.
- عملیات جمع ($x + f(x)$) باعث می‌شود که حتی در صورتی که لایه‌های پیچشی نتوانند به طور کامل یاد بگیرند، حداقل ورودی x به لایه‌های بعدی منتقل شود. به این ترتیب، اطلاعات اصلی ورودی در طول شبکه حفظ می‌شود.

^{۳۰}Residual Blocks

^{۳۱}Vanishing Gradient

^{۳۲}Exploding Gradient

^{۳۳}Residual Connections

^{۳۴}Residual Block

^{۳۵}Shortcut Connection

^{۳۶}Rectified Linear Unit

اتصالات میانبر: اتصالات میانبر^{۳۷} مهم‌ترین بخش در ساختار شبکه‌های عصبی باقی‌مانده هستند. این اتصالات به دو نوع اصلی تقسیم می‌شوند:

- **اتصال هویتی**^{۳۸}: در این نوع اتصال، ورودی مستقیماً به خروجی اضافه می‌شود. این مورد زمانی استفاده می‌شود که ابعاد ورودی و خروجی یکسان باشند.
- **اتصال فرافکن**^{۳۹}: در این نوع اتصال، از یک لایه پیچشی 1×1 برای تغییر ابعاد ورودی استفاده می‌شود تا بتوان ورودی را به خروجی با ابعاد متفاوت اضافه کرد. این نوع اتصال زمانی استفاده می‌شود که ابعاد ورودی و خروجی متفاوت باشند.

این اتصالات به شبکه این امکان را می‌دهند که شبکه‌های بسیار عمیق‌تری را بدون از دست دادن اطلاعات ورودی و بدون مشکلات یادگیری آموزش دهد.

ساختار داخلی هر بلوک: هر بلوک باقی‌مانده شامل دو یا سه لایه پیچشی است که به‌طور معمول به صورت زیر ترتیب داده می‌شوند:

$$y = \text{ReLU}(W_2 \cdot \text{ReLU}(W_1 \cdot x))$$

که در آن:

- W_2 و W_1 وزن‌های لایه‌های پیچشی اول و دوم هستند.
- ReLU تابع غیرخطی است که بعد از هر لایه پیچشی اعمال می‌شود.

پس از اعمال این عملیات‌ها، ورودی x از طریق اتصال میانبر به خروجی اضافه می‌شود:

$$y = f(x, \{W_i\}) + x$$

در صورت تفاوت ابعاد ورودی و خروجی، از یک لایه پیچشی 1×1 به عنوان اتصال فرافکن استفاده می‌شود تا این اختلاف ابعاد برطرف شود.

^{۳۷}Shortcut Connections

^{۳۸}Identity Shortcut

^{۳۹}Projection Shortcut

معماری‌های مختلف شبکه‌های عصبی باقی‌مانده

شبکه‌های عصبی باقی‌مانده در اندازه‌ها و عمق‌های مختلفی طراحی شده است:

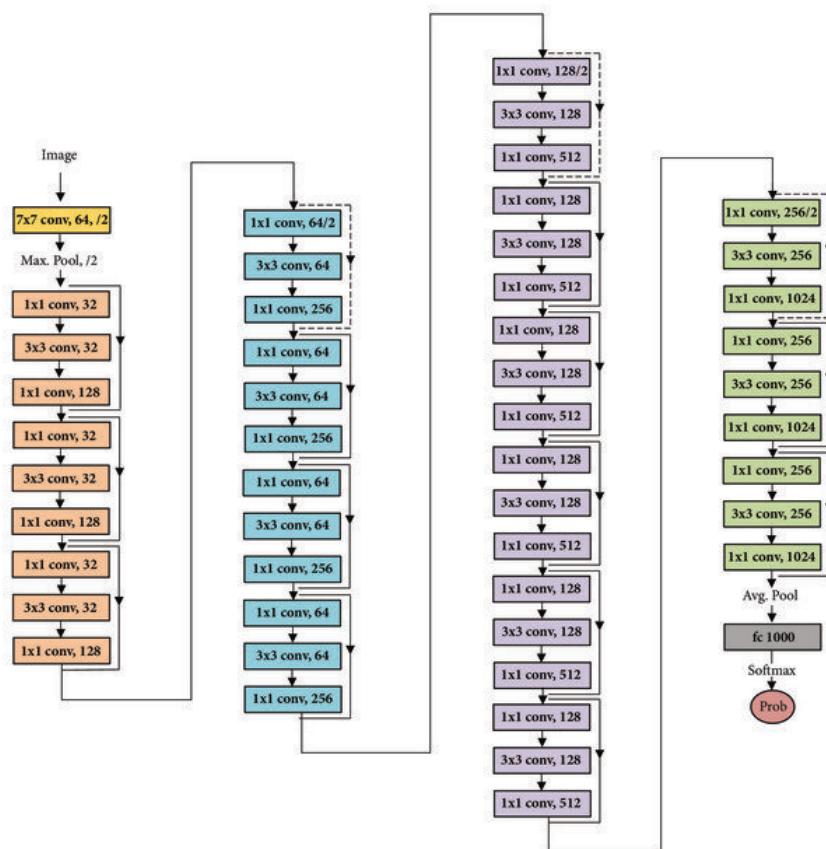
شامل ۱۸ لایه: ResNet-18 ●

شامل ۳۴ لایه: ResNet-34 ●

شامل ۵۰ لایه که یکی از محبوب‌ترین نسخه‌های شبکه‌های عصبی باقی‌مانده است: ResNet-50 ●

به ترتیب شامل ۱۰۱ و ۱۵۲ لایه: ResNet-101 و ResNet-152 ●

این معماری‌ها با افزایش عمق، همچنان قابلیت یادگیری و تعمیم‌دهی مناسبی دارند، به دلیل استفاده از اتصالات باقی‌مانده که مشکل یادگیری عمیق را حل می‌کنند.



شکل ۸.۲: ساختار معماری مدل شبکه‌های عصبی باقی‌مانده

مزایای شبکه‌های عصبی باقی‌مانده

- آموزش شبکه‌های عمیق‌تر: شبکه‌های عصبی باقی‌مانده امکان آموزش شبکه‌های با تعداد لایه‌های بسیار بیشتر را فراهم می‌کند بدون اینکه دچار مشکلات معمول شبکه‌های عمیق شود.
- حل مشکل نابودی گرادیان: با استفاده از مسیرهای میانبر، مشکل نابودی گرادیان در شبکه‌های عمیق به‌طور قابل توجهی کاهش پیدا می‌کند.
- بهبود دقต: معماری شبکه‌های عصبی باقی‌مانده توانست به طور مداوم دقت بالاتری را نسبت به سایر شبکه‌ها ارائه دهد و از شبکه‌های قدیمی‌تر مانند AlexNet و VGG پیشی بگیرد.

۳.۲.۲ ارتباط شبکه‌های باقی‌مانده با استخراج جاده از تصاویر ماهواره‌ای

در پژوهه استخراج نقشه‌های جاده از تصاویر ماهواره‌ای، شبکه‌های باقی‌مانده^{۴۰} می‌تواند به عنوان یک معماری مؤثر در بخش کدگذار^{۴۱} استفاده شود. بخش کدگذار در این گونه پروژه‌ها معمولاً وظیفه استخراج ویژگی‌های اولیه از تصویر را بر عهده دارد. استفاده از شبکه‌های عصبی باقی‌مانده به دلیل قابلیت‌های آن در یادگیری عمیق و توانایی آن در پردازش داده‌های تصویری با جزئیات پیچیده، می‌تواند دقت استخراج ویژگی‌ها را افزایش دهد. همچنین، شبکه‌های عصبی باقی‌مانده به دلیل استفاده از اتصالات باقی‌مانده می‌تواند مشکلات ناشی از افزایش عمق شبکه را کاهش دهد و از افت دقت جلوگیری کند.

در پژوهه‌های مشابه، معماری‌های مبتنی بر شبکه‌های عصبی باقی‌مانده به‌طور گستردگی در بخش استخراج ویژگی‌ها برای کارهای مرتبط با بخش‌بندی تصویر^{۴۲} استفاده می‌شوند. به عنوان مثال، در مدل‌هایی مانند UNet و DeepLab که از شبکه‌های عصبی باقی‌مانده به عنوان کدگذار استفاده می‌کنند، دقت در استخراج مرزهای جاده و تشخیص جاده‌های پیچیده به‌طور قابل توجهی بهبود یافته است.

مزیت اصلی شبکه‌های عصبی باقی‌مانده

- تشخیص دقیق‌تر جاده‌ها: شبکه‌های عصبی باقی‌مانده با توانایی در استخراج ویژگی‌های پیچیده و محلی، به ویژه در محیط‌های پیچیده که جاده‌ها با پس‌زمینه‌ی متنوعی مانند مناطق شهری، روستایی و جنگلی تلفیق شده‌اند، عملکرد بهتری خواهد داشت.
- عمق شبکه بدون افت دقت: استفاده از ResNet-50 یا ResNet-101 به شما اجازه می‌دهد که از شبکه‌های عمیق استفاده کنید و ویژگی‌های جزئی‌تر از جاده‌ها و مرزهای آنها را استخراج کنید، بدون آنکه دچار مشکلات معمول شبکه‌های عمیق شوید.

^{۴۰} ResNet

^{۴۱} Encoder

^{۴۲} Image Segmentation

- قابلیت یکپارچگی با معماری‌های جدید: شبکه‌های عصبی باقی‌مانده می‌تواند به خوبی با معماری‌های مبتنی بر ترانسفورمر ترکیب شود و قابلیت‌های هر دو روش را به طور همزمان بهبود بخشد.

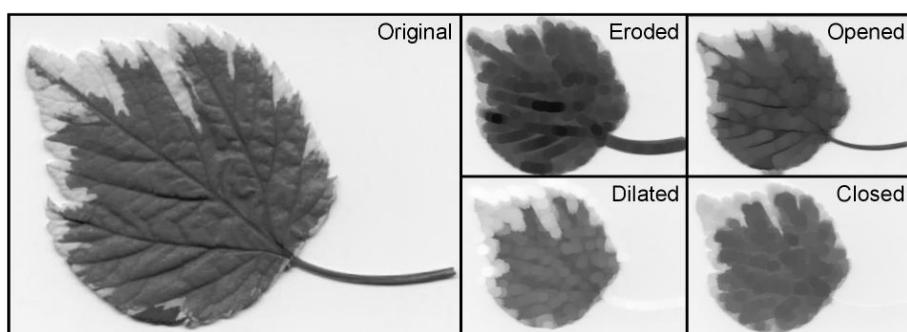
۴۰.۲.۲ نتیجه‌گیری

شبکه‌های عصبی باقی‌مانده یکی از مؤثرترین معماری‌های موجود در حوزه یادگیری عمیق است و استفاده از آن در پروژه شما، به ویژه برای استخراج جاده‌ها از تصاویر ماهواره‌ای، می‌تواند منجر به بهبود دقیق و کارایی مدل شود. با استفاده از ویژگی‌های قدرتمند شبکه‌های عصبی باقی‌مانده در بخش استخراج ویژگی‌ها، می‌توانید از مزایای آن برای بهبود عملکرد پروژه خود بهره ببرید..

۳۰.۲ روش‌های استخراج خیابان از نقشه‌های ماهواره‌ای

۱.۳۰.۲ روش‌های سنتی استخراج خیابان

روش‌های سنتی اغلب شامل تکنیک‌های پردازش تصویر کلاسیک هستند که بر مبنای ویژگی‌های لبه و شکل جاده عمل می‌کنند. از مهم‌ترین این روش‌ها می‌توان به استفاده از فیلترهای لبه‌باب نظری Sobel و Canny اشاره کرد که به شناسایی مرزهای جاده کمک می‌کنند. این روش‌ها به دلیل سادگی و سرعت پردازش بالا، هنوز در برخی کاربردها مورد استفاده قرار می‌گیرند، اما در تصاویر با نویز و تغییرات نوری دچار مشکل می‌شوند. از دیگر روش‌های سنتی، می‌توان به روش‌های مبتنی بر مورفولوژی ریاضی اشاره کرد؛ مورفولوژی ریاضی یکی از تکنیک‌های پردازش تصویر است که برای تغییر شکل و ساختار اشیاء تصویری مورد استفاده قرار می‌گیرد. این تکنیک‌ها معمولاً بر روی تصاویر دودویی (تصاویر سیاه و سفید) با دو مقدار پیکسل ۰ و ۱ اعمال می‌شوند. از این روش‌ها برای تشخیص و استخراج ساختارهای خطی مانند جاده‌ها استفاده می‌کنند و برای تصاویر با کنترast بالا عملکرد خوبی دارند، اما در تصاویر با کنترast پایین و شرایط پیچیده‌تر محدودیت دارند.



شکل ۹.۲:

۲.۳.۲ روش‌های مبتنی بر تجزیه و تحلیل ساختاری

این روش‌ها از ویژگی‌های هندسی و ساختاری برای شناسایی جاده‌ها استفاده می‌کنند. برای مثال، الگوریتم‌های تبدیل هوف^{۴۳} و تبدیل رادون^{۴۴} برای تشخیص خطوط مستقیم و منحنی‌ها کاربرد دارند. این روش‌ها به پارامترهای ثابت وابسته هستند و ممکن است در شرایط متغیر عملکرد ضعیفی داشته باشند.

۳.۳.۲ روش‌های مبتنی بر یادگیری ماشین

این روش‌ها از الگوریتم‌های یادگیری نظارت‌شده و نظارت‌نشده برای استخراج جاده‌ها استفاده می‌کنند. از جمله الگوریتم‌های معروف در این حوزه می‌توان به جنگل تصادفی^{۴۵} و ماشین بردار پشتیبانی^{۴۶} اشاره کرد. این الگوریتم‌ها از ویژگی‌های تصویری مختلف برای مدل‌سازی و شناسایی جاده‌ها استفاده می‌کنند و نسبت به روش‌های سنتی دقت و انعطاف‌پذیری بیشتری دارند ولی این روش‌ها به داده‌های آموزشی با کیفیت نیاز دارند و ممکن است به خوبی روش‌های یادگیری عمیق عمل نکنند.

۴.۳.۲ روش‌های مبتنی بر یادگیری عمیق

مدل‌های یادگیری عمیق، به ویژه شبکه‌های عصبی پیچشی و شبکه‌های عصبی بازگشتی^{۴۷}، به طور گسترده برای استخراج جاده‌ها از تصاویر ماهواره‌ای استفاده می‌شوند. مدل‌هایی نظیر U-Net و SegNet به دلیل دقت بالای خود در تشخیص و تقسیم‌بندی جاده‌ها معروف هستند. این مدل‌ها با استفاده از معماری‌های پیچیده و لایه‌های متعدد، عملکرد بسیار خوبی در شرایط مختلف از خود نشان می‌دهند. این مدل‌ها می‌توانند ویژگی‌های پیچیده تصاویر را به خوبی یاد بگیرند و عملکرد قابل قبولی در استخراج جاده‌ها نشان دهند.

۵.۳.۲ استفاده از مکانیزم توجه

مکانیزم توجه^{۴۸} به عنوان یک ابزار قدرتمند برای بهبود دقت مدل‌های استخراج جاده معرفی شده‌است. این مکانیزم به مدل اجازه می‌دهد تا وابستگی‌های بلندمدت را بهتر مدل‌سازی کرده و جزئیات دقیق‌تری از جاده‌ها استخراج کند. ترنسفور چرخشی^{۴۹} یکی از مدل‌هایی است که از این مکانیزم بهره می‌برد و در تشخیص دقیق جاده‌ها در تصاویر ماهواره‌ای با وضوح کم و شرایط پیچیده عملکرد بهتری دارند.

^{۴۳} Hough Transform

^{۴۴} Radon Transform

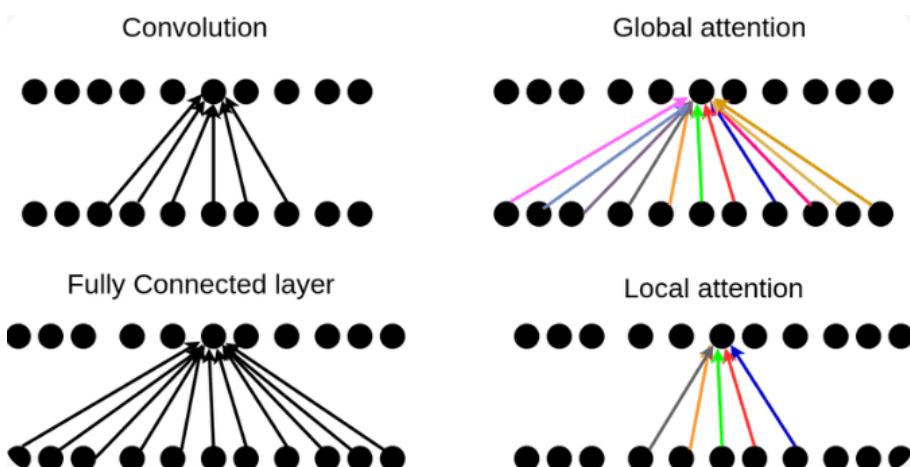
^{۴۵} Random Forest

^{۴۶} Support Vector Machines

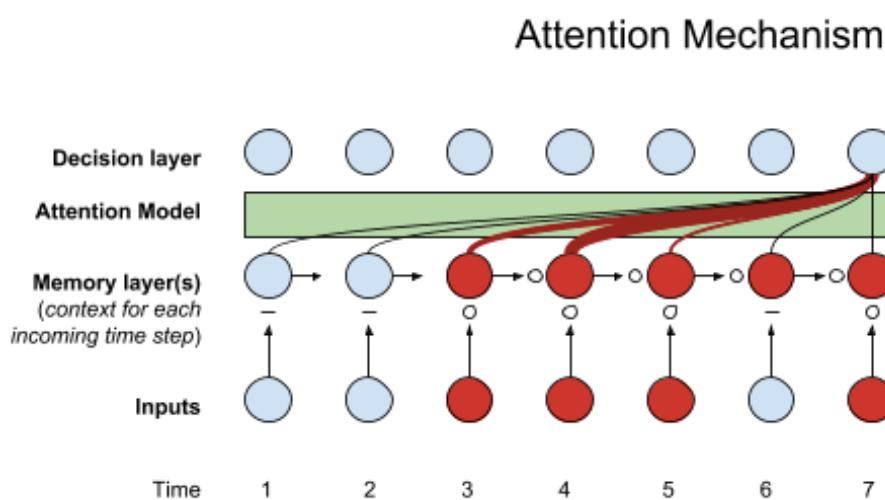
^{۴۷} Recurrent Neural Networks

^{۴۸} Attention Mechanism

^{۴۹} Swin Transformer



شکل ۱۰.۲: مکانیزم توجه در برابر باقی روش‌ها



شکل ۱۱.۲: معماری مکانیزم توجه

۶.۳.۲ ترکیب روش‌های مختلف

بسیاری از تحقیقات جدید به ترکیب روش‌های مختلف پرداخته‌اند تا از مزایای هر یک بهره‌مند شوند. برای مثال، ترکیب شبکه‌های عصبی پیچشی و شبکه‌های عصبی بازگشتی با مکانیزم توجه می‌تواند دقیق و کارایی مدل را به طور قابل توجهی افزایش دهد. این ترکیب‌ها باعث شده‌اند تا مدل‌های جدید بتوانند در شرایط مختلف و با تصاویر با کیفیت‌های متفاوت به خوبی عمل کنند.

٤.٢ ترنسفور چرخشی

ترنسفور چرخشی یک مدل مبتنی بر ترنسفورمر است که برای وظایف دید کامپیوترا^{۵۰} طراحی شده و به خصوص در پردازش تصاویر ماهواره‌ای کاربردهای گسترده‌ای دارد؛ به عنوان روشی نوین برای استخراج خیابان از تصاویر ماهواره‌ای استفاده می‌شود. ترنسفور چرخشی نوعی شبکه عصبی عمیق است که از معماری ترنسفورمر و مکانیزم توجه، برای یادگیری وابستگی‌های فضایی بین پیکسل‌های تصویر استفاده می‌کند. این روش به دلیل دقت بالا، عملکرد قابل اعتماد در مقایسه با روش‌های سنتی و قابلیت‌های پیشرفته‌اش در مدل‌سازی وابستگی‌های بلندمدت و محلی، به عنوان یکی از ابزارهای موثر در استخراج خیابان از تصاویر ماهواره‌ای مطرح شده است.

۱.۴.۲ معماری و ساختار ترنسفور چرخشی

ترنسفور چرخشی یک مدل ترنسفور چشم انداز سلسله‌مراتبی^{۵۱} است که از پنجره‌های جابه‌جا شده^{۵۲} برای پردازش تصاویر استفاده می‌کند. این مدل با تقسیم تصویر به پنجره‌های کوچک‌تر و جابجایی این پنجره‌ها در هر لایه، وابستگی‌های محلی و جهانی را به صورت همزمان مدل‌سازی می‌کند. این روش باعث می‌شود که ترنسفور چرخشی در تشخیص جزئیات دقیق تصاویر و حفظ وابستگی‌های بلندمدت عملکرد بهتری نسبت به مدل‌های سنتی داشته باشد.

معماری ترنسفور چرخشی شامل چندین بلوک تکرارشونده است که هر کدام از سه بخش اصلی تشکیل شده‌اند:

- لایه‌ی نرم‌سازی^{۵۳}: این لایه داده‌های ورودی را به یک مقیاس استاندارد تبدیل می‌کند تا از ناپایداری در حین آموزش شبکه جلوگیری شود.
- لایه‌ی توجه محلی^{۵۴}: این لایه به ترنسفور چرخشی اجازه می‌دهد تا بر روی بخش‌های کوچکی از تصویر تمرکز کند و وابستگی‌های فضایی بین پیکسل‌های مجاور را یاد بگیرد.
- لایه‌ی پیشخور^{۵۵}: این لایه یک تبدیل غیرخطی به داده‌ها اعمال می‌کند و به شبکه اجازه می‌دهد تا روابط پیچیده‌تر بین پیکسل‌ها را یاد بگیرد.

بلوک‌های تکرارشونده به طور متوالی به یکدیگر متصل می‌شوند و در طول فرآیند آموزش، ترنسفور چرخشی یاد می‌گیرد تا ویژگی‌های مرتبط با خیابان‌ها را از تصاویر ماهواره‌ای استخراج کند.

^{۵۰}Computer Vision

^{۵۱}Hierarchical Vision Transformer

^{۵۲}Shifted Windows

^{۵۳}Normalization Layer

^{۵۴}Local Attention Layer

^{۵۵}Feedforward Layer

۴.۴.۲ آموزش مدل ترنسورماتور چرخشی

برای آموزش ترنسفور چرخشی، به یک مجموعه داده از تصاویر ماهواره‌ای همراه با برچسب‌های مربوط به خیابان‌ها نیاز داریم. برچسب‌ها می‌توانند به صورت دودویی باشند، به طوری که هر پیکسل به عنوان خیابان یا غیرخیابان طبقه‌بندی شود.

سپس، ترنسفور چرخشی با استفاده از الگوریتم‌های بهینه‌سازی مانند Adam یا SGD آموزش داده می‌شود تا تابع ضرر را به حداقل برساند. تابع ضرر معمولاً^{۵۶} بر اساس معیارهایی مانند دقت^{۵۶}، فراخوانی^{۵۷} و $F1 - score$ محاسبه می‌شود.

۴.۴.۲ مزیت‌های ترنسفور چرخشی

ترنسفور چرخشی در مقایسه با روش‌های سنتی استخراج خیابان از تصاویر ماهواره‌ای، مزایای متعددی دارد، از جمله:

- دقت بالا: به دلیل استفاده از مکانیزم توجه، ترنسفور چرخشی قادر است وابستگی‌های بلندمدت را به خوبی مدل‌سازی کرده و دقت بالایی در تشخیص و تقسیم‌بندی خیابان‌ها از تصاویر ماهواره‌ای ارائه دهد.
- کارایی محاسباتی: معماری پنجره‌ای این مدل امکان پردازش موازی و بهینه‌سازی‌های سخت‌افزاری را فراهم می‌کند، که منجر به افزایش کارایی محاسباتی می‌شود.
- انعطاف‌پذیری: ترنسفور چرخشی به دلیل ساختار مدولار خود، قابل تنظیم برای کاربردهای مختلف است و می‌تواند در شرایط و داده‌های متفاوت به خوبی عمل کند.
- عملکرد قابل اعتماد: ترنسفور چرخشی در شرایط مختلف، مانند تصاویر باوضوح کم یا تصاویر با نویز زیاد، عملکرد قابل اعتمادی را نشان می‌دهد.

ترنسفور چرخشی نسبت به مدل‌های شبکه عصبی پیچشی و ترنسفورمر بینایی مزایای متعددی دارد. اولاً^{۵۸}، به دلیل استفاده از پنجره‌های جایه‌جا شده، نیاز به محاسبات کمتری دارد و بنابراین سریع‌تر و کارآمدتر است. ثانیاً، معماری سلسله‌مراتبی آن به مدل اجازه می‌دهد تا ویژگی‌های مختلف تصویر را در سطوح مختلف استخراج کند که به بهبود دقت مدل کمک می‌کند. بنابراین این مدل با توانایی در حفظ وابستگی‌های بلندمدت و تشخیص جزئیات دقیق، عملکرد بهتری نسبت به مدل‌های سنتی و حتی برخی مدل‌های یادگیری عمیق دارد.

^{۵۶} Accuracy

^{۵۷} Recall

۳ فصل

پیشینه‌ی پژوهش

۱.۳ روش‌های سنتی

مطالعات بسیاری تلاش کرده‌اند تا اطلاعات جاده‌ای را از تصاویر سنجش از دور استخراج کنند. در [۱۸]، الگوریتمی برای استخراج جاده بر اساس مدل رنگ طراحی شد که اطلاعات مرزی را در تصاویر خاکستری با نتایج استخراج ناحیه‌ی جاده در تصاویر رنگی ترکیب می‌کند. در [۵۱]، نواحی جاده با تحلیل ویژگی‌های بافت جاده و استخراج آن‌ها بر اساس ویژگی‌های مختلف بافت استخراج شدند. در [۴۹]، از مدل میدانی تصادفی شرطی مرتبه بالا برای استخراج جاده استفاده شد. در [۴۰]، مدل دو مرحله‌ای دیگری پیشنهاد شد. در ابتدا، پیکسل‌ها با استفاده از الگوریتم ماشین بردار پشتیبانی بر اساس ویژگی‌های جاده به دو گروه جاده و غیر جاده تقسیم می‌شوند. سپس، گروه جاده با استفاده از الگوریتم‌های بخش‌بندی پالایش شده و نواحی نهایی جاده تولید می‌شوند. مزیت این روش‌های سنتی سادگی و سهولت در پیاده‌سازی است. با این حال، این روش‌ها اغلب کارایی پایینی دارند و ممکن است در سناریوهای پیچیده‌ی جاده عملکرد خوبی نداشته باشند، همچنان در مواجهه با ساختارهای ناقص، تغییرات نور و کنتراست، قادر استحکام هستند.

۲.۳ روش‌های مبتنی بر شبکه‌های عصبی پیچشی

برای حل مشکلات موجود در روش‌های سنتی، رویکردهای مبتنی بر یادگیری عمیق برای استخراج جاده به کار گرفته شدند. به عنوان یک نماینده از روش‌های مبتنی بر شبکه‌های عصبی پیچشی، یک مدل مبتنی بر پچ^۱ برای استخراج جاده از داده‌های سنجش از دور با وضوح بالا پیشنهاد شد [۱]. سپس، RoadNet برای استخراج سطح جاده، خطوط مرکزی و لبه‌ها در چندین وظیفه ارائه شد.

^۱Patch-Based



شکل ۱.۳: پردازش تصویر با استفاده از ماشین بردار پشتیبان

به منظور حفظ اطلاعات جزئیات مکانی بیشتر و افزایش یکپارچگی جاده، یک شبکه‌ی گراف پیچشی و بخش‌بندی سوپرسل اخیراً توسعه یافته است [۹]. روش‌های مبتنی بر شبکه‌های عصبی پیچشی که در بالا ذکر شدند می‌توانند دقت بالایی را به دست آورند، اما سرعت پردازش آن‌ها نیاز به افزایش دارد.

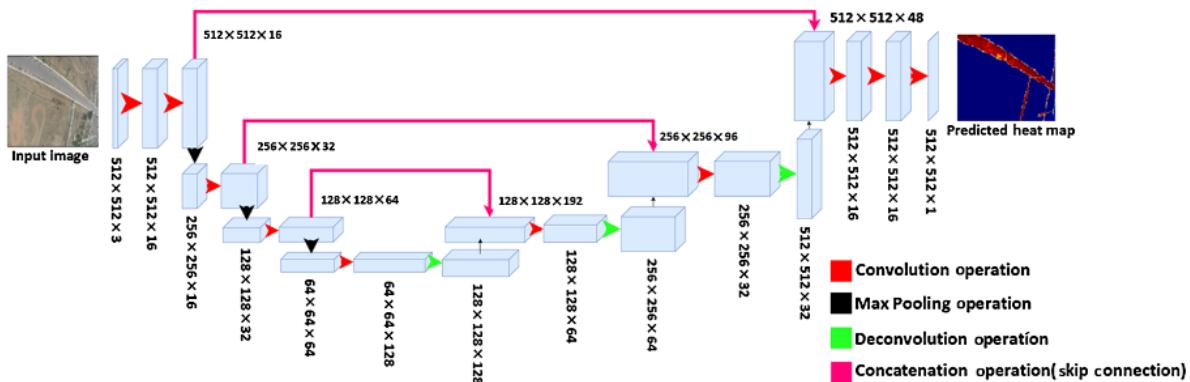
برای رفع کاستی‌های شبکه‌های عصبی پیچشی، شبکه‌ی کاملاً پیچشی^۲ به جای لایه‌ی کاملاً متصل از لایه‌ی دکانولوشن استفاده می‌کند که طبقه‌بندی پیکسلی انتها به انتها را انجام می‌دهد. در کارهای اولیه، ثابت شد که رویکرد FCN در حفظ پیوستگی و یکپارچگی جاده‌ها در وظایف استخراج جاده موفق عمل کرده است [۲۲]. بعدتر، پیشنهاد شد که از UFCN برای استخراج جاده‌ها از تصاویر هوایی گرفته شده توسط پهپاد استفاده شود. پس از آن، 32-FCN برای استخراج جاده از تصاویر با وضوح بالا به کار گرفته شد [۴۶].

برای استفاده‌ی جامع از اطلاعات چندمقیاسی موجود در تصاویر، سری U-Net با مازول‌های اتصال میان‌گذر توسعه یافت [۵۲]. مدل SegNet از ساختار کدگزار-رمزنگشا استفاده کرد، جایی که موقعیت لبه در رمزنگشا با استفاده از مقدار شاخص ذخیره شده در کدگزار بازیابی می‌شود. اخیراً، برای به دست آوردن نتایج بخش‌بندی بهتر، سری روش‌های DeepLab از کانولوشن گسترش یافته برای استخراج اطلاعات دوربرد استفاده کرده و یک لایه‌ی هرمی‌شکل برای حفظ ساختار مکانی توسعه دادند.

با اینکه مدل‌های شبکه عصبی تماماً متصل کارایی استخراج جاده را بهبود می‌بخشند، اما اغلب در سناریوهای بسیار پیچیده دچار خطای دسته‌بندی جاده و پس زمینه می‌شوند. در همین حال، مدل‌های مبتنی بر شبکه عصبی تماماً متصل به دلیل وجود لایه‌های نمونه‌برداری پایین، اطلاعات موقعیت لبه را از دست می‌دهند. علاوه بر این، از دست رفتن اطلاعات دوربرد، دقت بخش‌بندی در مدل‌های U-Net و

^۲FCN

SegNet را محدود می‌کند. همچنین، کانولوشن گسترش یافته باعث می‌شود که DeepLab در استخراج اهداف بزرگ عملکرد خوبی داشته باشد، اما در اهداف کوچک ضعیف عمل کند. برای حل مشکلات فوق، ما ساختار ترانسفور را در وظیفه‌ی بخش‌بندی جاده معرفی می‌کنیم [۳۳].



شکل ۲.۳: معماری مدل شبکه عصبی تماما متصل برای تشخیص مسیر

۳.۳ روش‌های مبتنی بر ترانسفور

اخیراً، معماری ترانسفور به دلیل مکانیزم توجه ویژه‌اش در زمینه‌ی بینایی رایانه‌ای بسیار مورد توجه قرار گرفته است [۴۷]. مکانیزم توجه ترانسفور می‌تواند آن را قادر سازد تا ویژگی‌های دوربرد را یاد بگیرد و اطلاعات جهانی را مدل‌سازی کند، برخلاف تأکید شبکه‌های عصبی پیچشی بر ویژگی‌های محلی. ترانسفور تصویری [۱۱] در دسته‌بندی تصاویر نتایج رضایت‌بخشی کسب کرد و پتانسیل بالایی در بینایی رایانه‌ای نشان داد، جایی که قطعات تصویر به عنوان توکن‌های مازول ترانسفور در نظر گرفته می‌شوند. با این وجود، طراحی این روش هنوز دارای معایب زیادی است. بار محاسباتی درجه دوم تحمیل شده توسط ترانسفورها، هزینه‌ی قابل توجهی را به همراه دارد که در وظایف بخش‌بندی برای تصاویر با اندازه بزرگ غیرقابل تحمل است. علاوه بر این، اگرچه ترانسفور قادر به استخراج اطلاعات دوربرد و متن جهانی است، اما در شناسایی اطلاعات سطح پایین مورد نیاز در بخش‌بندی دشواری دارد [۴].

برای کاهش نیازهای حافظه‌ای ترانسفورها، Liu و همکاران [۳۵] ترانسفور چرخشی را ابداع کردند که از استراتژی ترکیب قطعات مجاور برای ساختار نمایش سلسله‌مراتبی استفاده می‌کند. با این نمایش‌های سلسله‌مراتبی، مدل می‌تواند به راحتی پیش‌بینی‌های متراکم را با استفاده از یک شبکه‌ی هرمی ویژگی انجام دهد. در همین حال، ترانسفور چرخشی توجه به خود را در پنجره‌های غیرهمپوشان با پیچیدگی محاسباتی خطی محاسبه می‌کند. این مزایا آن را برای استفاده به عنوان کدگذار در شبکه‌های بخش‌بندی مناسب می‌سازد.

۴.۳ ترکیب شبکه‌های عصبی پیچشی و ترنسفور

اگرچه توانایی مدل‌های ترنسفور در شناسایی وابستگی‌های دوربرد منجر به بهبود عملکرد بخش‌بندی در وظایف بخش‌بندی تصویر شده است، شبکه‌های عصبی پیچشی همچنان در استخراج اطلاعات ویژگی‌های محلی مزیت ذاتی دارند. اخیراً در زمینه‌ی پژوهش‌های استخراج جاده، تلاش‌هایی برای ترکیب ترنسفورها و شبکه‌های عصبی پیچشی صورت گرفته است. این روش‌ها هم توانایی ترنسفورها را در شناسایی وابستگی‌های دوربرد و هم توانایی شبکه‌های عصبی پیچشی را در استخراج اطلاعات ویژگی‌های محلی دارا هستند.

برای ترکیب مزایای هر دو معماری شبکه‌های عصبی پیچشی و ترنسفور، در [۱۶] مدلی به نام هم‌ساز^۳ پیشنهاد شد که ساختار دو مسیرهای طراحی می‌کند که به آن‌ها اجازه می‌دهد به صورت تعاملی ویژگی‌ها را بیاموزند. آزمایش‌ها نشان دادند که عملکرد کدگذار ترکیبی شبکه‌های عصبی پیچشی – ترنسفور نسبت به استفاده از ترنسفور به‌نهایی به عنوان کدگذار بهتر است. در [۶]، مدل TransUNet پیشنهاد شد که ویژگی‌های ترنسفور و UNet را ترکیب می‌کند. این مدل ابتدا از شبکه‌های عصبی پیچشی برای استخراج ویژگی‌ها از تصاویر استفاده می‌کند تا نقشه‌های ویژگی به دست آید، سپس این نقشه‌ها به ترنسفور کدگذار منتقل می‌شوند. در نهایت، از یک مازول رمزگشا مشابه با UNet برای انجام نمونه‌برداری بالا و اتصالات میان‌گذر لایه‌ای به منظور تولید نقشه‌های بخش‌بندی استفاده می‌شود. این مدل نتایج خوبی در چندین وظیفه‌ی بخش‌بندی تصویر به دست آورده است. در [۲۸]، ترنسفور چرخشی گسترش داده شد و ساختارهای رمزگشا به آن اضافه شدند تا عملکرد بخش‌بندی جاده در مدل بهبود یابد. در [۴۲]، مدل Seg-Road پیشنهاد شد تا اتصال جاده‌ها را بهبود بخشد. این مدل از ترنسفور برای برقراری وابستگی‌های بلندبرد و اطلاعات متنی جهانی استفاده می‌کند تا مشکلات تکه‌تکه شدن بخش‌بندی جاده‌ها را بهبود دهد. علاوه بر این، از ساختار شبکه‌های عصبی پیچشی برای استخراج اطلاعات زمینه‌ای محلی استفاده می‌کند تا بخش‌بندی جزئیات جاده را بهبود بخشد.

^۳Conformer

۴ فصل

طراحی و ارزیابی مدل جاده‌ساز

در این فصل، به بررسی و توضیح کامل روش کلی پژوهش پرداخته می‌شود. هدف اصلی این فصل، ارائه یک دید جامع از معماری مدل پیشنهادی به نام جاده ساز^۱ است که برای استخراج نقشه‌های جاده از تصاویر ماهواره‌ای طراحی شده است. این معماری شامل سه بخش اصلی رمزگذار، گلوگاه و رمزگشا می‌باشد. در هر بخش از این فصل، جزئیات معماری و نحوه عملکرد هر یک از این اجزاء به تفصیل توضیح داده می‌شود.

ابتدا، در بخش اول، به طراحی کلی جاده‌ساز پرداخته و ساختار کلی آن معرفی می‌شود. سپس، در بخش‌های بعدی، جریان کاری رمزگذار و گلوگاه مورد بررسی قرار می‌گیرد و نحوه پردازش و پالایش ویژگی‌های جاده تشریح می‌شود. سپس به بررسی عملکرد رمزگشا و نقش آن در بازسازی جزئیات نقشه‌های جاده و همچنین تابع هزینه مورد استفاده در این مدل می‌پردازیم. در ادامه نتایج آزمایش‌ها و تحلیل‌های انجام شده بر روی مدل جاده‌ساز، پرداخته می‌شود. این فصل به گونه‌ای طراحی شده است که تمامی مراحل و اجزای مدل به طور کامل و دقیق مورد بررسی قرار گیرد تا درک بهتری از فرآیند استخراج نقشه‌های جاده از تصاویر ماهواره‌ای به دست آید.

¹RoadFormer

این بخش به توصیف دقیق معماری مدل پیشنهادی می‌پردازد. روش کلی شامل سه جزء اصلی می‌باشد:

۱. رمزگذار^۲

۲. گلوگاه^۳

۳. رمزگشا^۴

در بخش ۳،۱، طراحی کلی جادهساز معرفی می‌شود. سپس، جریان کاری رمزگذار در بخش ۳،۲ شرح داده می‌شود و طراحی گلوگاه برای پالایش ویژگی‌های جاده در بخش ۳،۳ ارائه می‌شود. در نهایت، بخش ۳،۴ به رمزگشا وتابع هزینه می‌پردازد.

۱.۴ طراحی کلی مدل جادهساز

ما یک مدل استخراج جاده به نام جادهساز ارائه می‌دهیم تا محدودیت‌های میدان گیرنده را برطرف کرده و اطلاعات دقیق در تصاویر سنجش از دور را به دست آوریم. معماری جادهساز به سه بخش تقسیم می‌شود که در شکل ۱.۴ نمایش داده شده است:

۱. رمزگذار مبتنی بر ترنسفور چرخشی: رمزگذار، تصویر ورودی RGB^۵ را به نقشه‌های ویژگی^۶ با ابعاد بالا و چندمقیاسی، که ورودی‌های لازم برای رمزگشا و گلوگاه هستند، کاهش می‌دهد و رمزگذاری می‌کند.

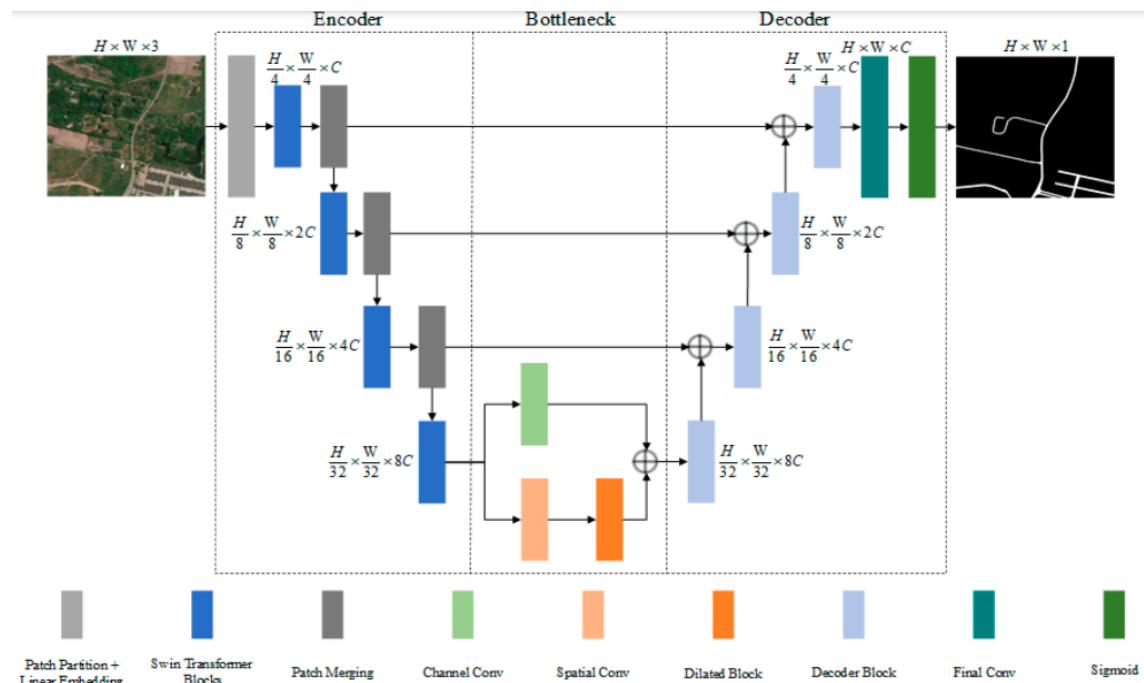
^۱Encoder

^۲Bottleneck

^۳Decoder

^۴Red, Green and Blue

^۵Feature Map



شکل ۱۰.۴: معماری جاده‌ساز از یک رمزگذار، گلوگاه و رمزگشا تشکیل شده است. نمایش ویژگی چند مقیاسی می‌تواند توسط رمزگذار تولید شود. نقشه‌های ویژگی با ابعاد بالا را می‌توان با کانولوشن قابل جداسازی و بلوک متسع در گلوگاه به دست آورد. نتایج نهایی توسط رمزگشا ارائه می‌شود.

۲. گلوگاه جداسازی ویژگی: گلوگاه، نقشه‌های ویژگی با ابعاد بالا را به ویژگی‌های کانالی و مکانی جدا می‌کند. در همین حال، یک بلوک متسع^۷ متشكل از چهار لایه پیچشی متسع به ویژگی فضایی اعمال می‌شود تا میدان گیرنده را گسترش دهد.

۳. رمزگشای سبک: نقشه‌های ویژگی ایجاد شده توسط گلوگاه به طور متناسب نمونه‌برداری می‌شوند و با نقشه‌های ویژگی تولید شده توسط رمزگذار در بلوک رمزگشای بالایی ادغام می‌شوند. سپس، با استفاده از کانولوشن ترانهاده^۸ و Sigmoid نتیجه تقسیم‌بندی از رمزگشای بالایی به دست می‌آید.

۲.۴ رمزگذار: پردازش و استخراج ویژگی‌ها

بدون از دست دادن کلیت، توزیع جاده‌ها باید پیوسته و در سراسر تصویر باشد و مدل باید توانایی جمع‌آوری اطلاعات بلندمدت را داشته باشد. ما به دلیل توانایی ترنسفور چرخشی در مدل‌سازی روابط اطلاعات بلندمدت از آن استفاده کردیم.

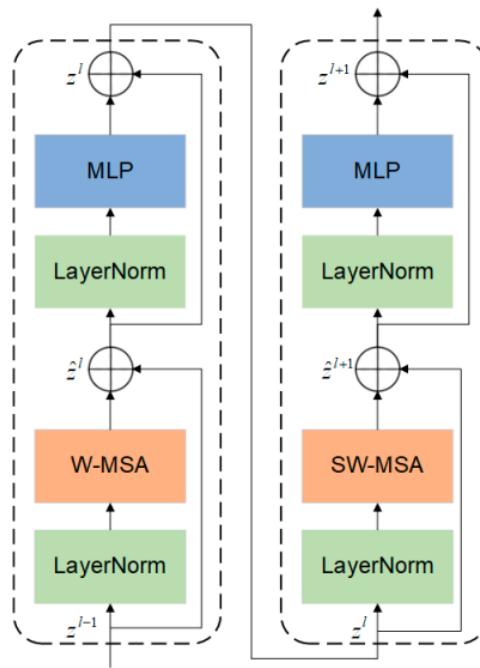
برخلاف ترنسفورمر معمولی، ترنسفور چرخشی مازول MSA^۹ را با بلوکی که می‌تواند شامل

^۷Dilated Block

^۸Transposed Convolution

^۹Multi-Head Self-Attention Module

مبتنی بر پنجره‌ی جابه‌جاشده، 10 MLP، 11 LayerNorm، 12 Residual Connection و اتصال باقیمانده 13 W-MSA، 14 SW-MSA به طور متوالی و متناوب، MSA، با پیکربندی‌های پنجره‌ای منظم و انتقال‌یافته در یک بلوک اعمال می‌شوند. ساختار بلوک‌های ترانسفور چرخشی در شکل ${}^{20.4}$ ارائه شده است.



شکل ${}^{20.4}$: معماری بلوک‌های ترانسفور چرخشی

رمزگذار مدل پیشنهادی از چهار ماثول ترانسفور چرخشی که به صورت پشته 15 آمده، تشکیل شده است. تصویر اصلی ($H \times W \times 3$) به بخش وصله 16 در لایه اول منتقل می‌شود و به وصله‌هایی به ابعاد $(\frac{H}{4} \times \frac{W}{4} \times 48)$ تقسیم می‌شود. سپس، این وصله‌ها با استفاده از نگاشت خطی به توکن‌ها تبدیل می‌شوند. بعد از آن، توکن‌ها به صورت متناوب در بلوک‌های ترانسفور چرخشی و لایه‌های ادغام وصله‌ها 17 تغذیه می‌شوند تا نمایه سلسله مراتبی ایجاد شود. به طور خاص، بلوک‌های ترانسفور چرخشی نقشه‌های ویژگی را در مقیاس لایه فعلی تولید می‌کنند. در حالی که لایه‌های ادغام وصله‌ها، تعدادی از این ویژگی‌ها را کاهش می‌دهند. قابل توجه است که خروجی لایه‌های ادغام وصله‌ها، به صورت همزمان، با استفاده از اتصال میانبر به لایه‌ی مربوط به رمزگشای تأمین می‌شود و به عنوان ورودی بلوک ترانسفور چرخشی بعدی استفاده می‌شود.

¹⁰Multi-Layer Perceptron

¹¹Layer Normalization

¹²Residual Connection

¹³Window-Based MSA

¹⁴Shifted Window MSA

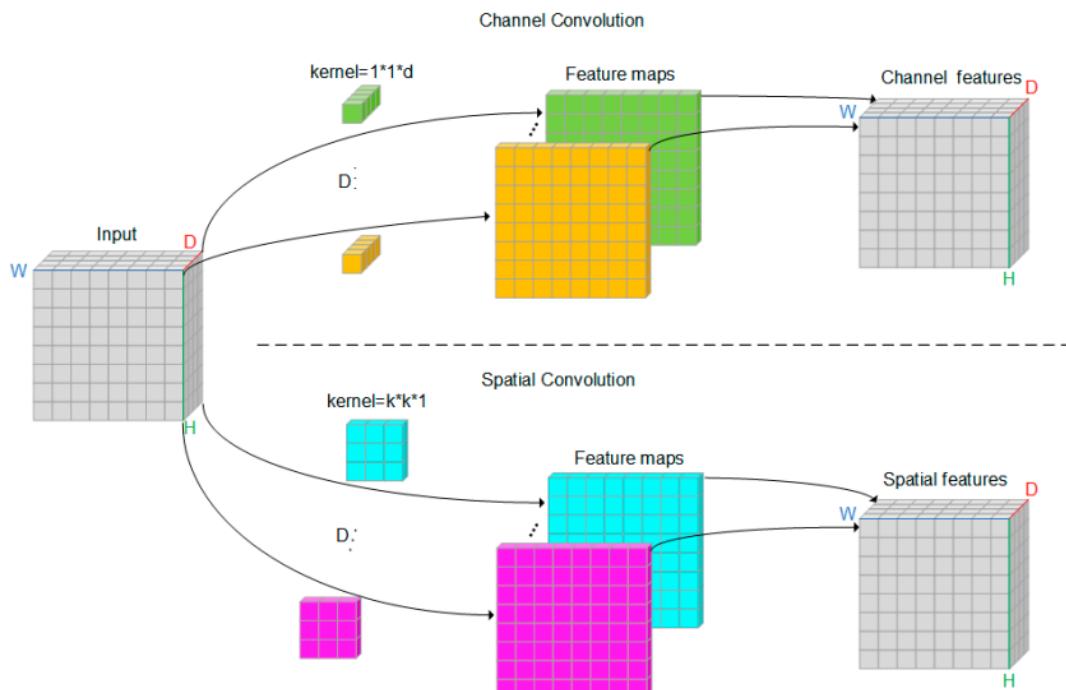
¹⁵Stack

¹⁶Patch Partition

¹⁷Patch Merging Layers

۳.۴ گلوگاه: پالایش ویژگی‌ها

برای به دست آوردن ویژگی‌های مکانی و کanalی به طور مؤثر، یک ساختار موازی که با کانولوشن ^{18}DW و ^{19}PW ترکیب شده است، در جاده‌ساز توسعه یافته است. فرآیند مارژول کانولوشن قابل جداسازی 20 در شکل ۳.۴ نشان داده شده است. به طور خاص، اتصالات موازی بین کانولوشن مکانی 21 و کانولوشن کanalی 22 پس از رمزگذار برقرار می‌شوند. در بخش کانولوشن کanalی، یک هسته کانولوشن یک بعدی برای کانولوشن نقشه‌ی ویژگی در جهت کanal استفاده می‌شود. در بخش کانولوشن مکانی، هر نقشه ویژگی توسط یک هسته $(k \times 1 \times 1)$ کانولوشن شده و به عنوان نقشه‌های ویژگی مکانی ادغام می‌شود. نقشه‌های ویژگی پالایش شده توسط کانولوشن مکانی و کanalی اندازه‌ای برابر با $(H \times W \times N)$ دارند که با ورودی سازگار است.



شکل ۳.۴: ساختار کانولوشن قابل جداسازی موازی. کانولوشن قابل تفکیک، ویژگی‌ها را به ویژگی‌های کanal و ویژگی‌های فضایی تقسیم می‌کند.

کارهای قبلی ثابت کردند که کانولوشن سنتی تمایل به داشتن میدان گیرنده 23 محدودی دارد که در وظایف بخش‌بندی عملکرد خوبی ندارد. خوشبختانه، کانولوشن متسع می‌تواند میدان‌های گیرنده را به

18 Depthwise Convolution

19 Pointwise Convolution

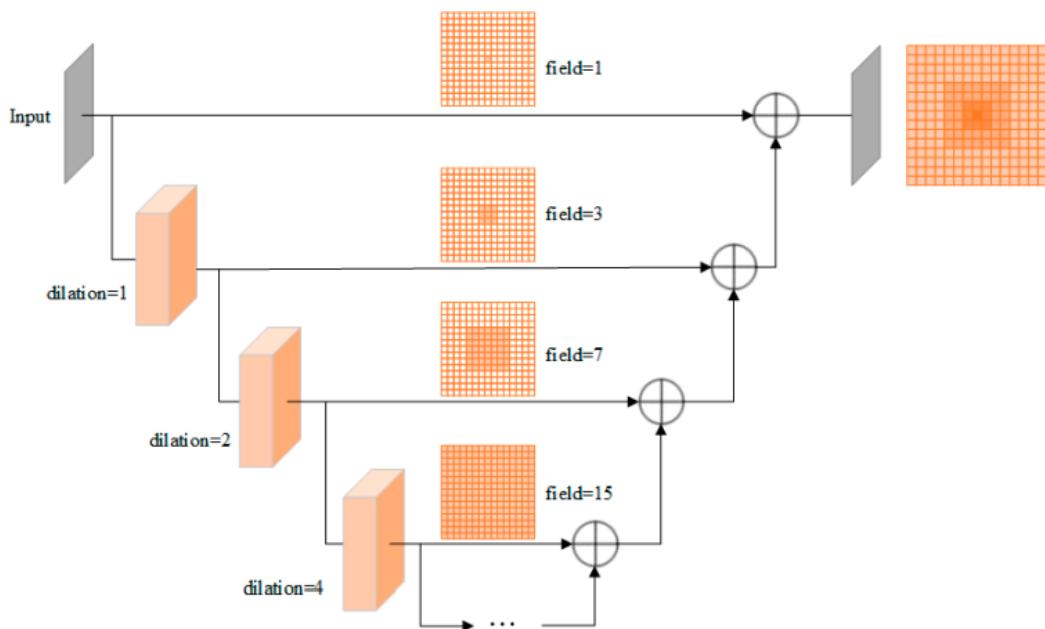
20 Separable Convolution Module

21 Spatial Convolution

22 Channel Convolution

23 Receptive Field

طور مؤثر گسترش دهد؛ در حالی که وضوح ویژگی‌های نقشه را حفظ می‌کند. با اشاره به D -LinkNet ما پس از مارژول کانولوشن مکانی، ساختاری موازی و متواലی از کانولوشن متسع قرار دادیم. همان طور که در شکل ۴.۴ نشان داده شده است؛ اگر نرخ‌های متسع هر لایه به ترتیب ۱، ۲، ۴ و ۸ تنظیم شوند، میدان‌های گیرنده هر لایه ۳، ۷، ۱۵ و ۳۱ خواهد بود. رمزگذار ترانسفور چرخشی ورودی اصلی را با یک کاهش $\frac{1}{4}$ و سه کاهش $\frac{1}{3}$ نمونه‌برداری می‌کند. برای تصویری با اندازه 1024×1024 ، اندازه‌ی نقشه ویژگی خروجی رمزگذار 32×32 خواهد بود. در این حالت، میدان گیرنده بلوک متسع می‌تواند تقریباً تمام محدوده‌ی نقشه‌ی ویژگی را پوشش دهد. از طریق این طراحی، این معماری به طور قابل توجهی ظرفیت مدل ما را برای جمع‌آوری اطلاعات بلندمدت افزایش می‌دهد.



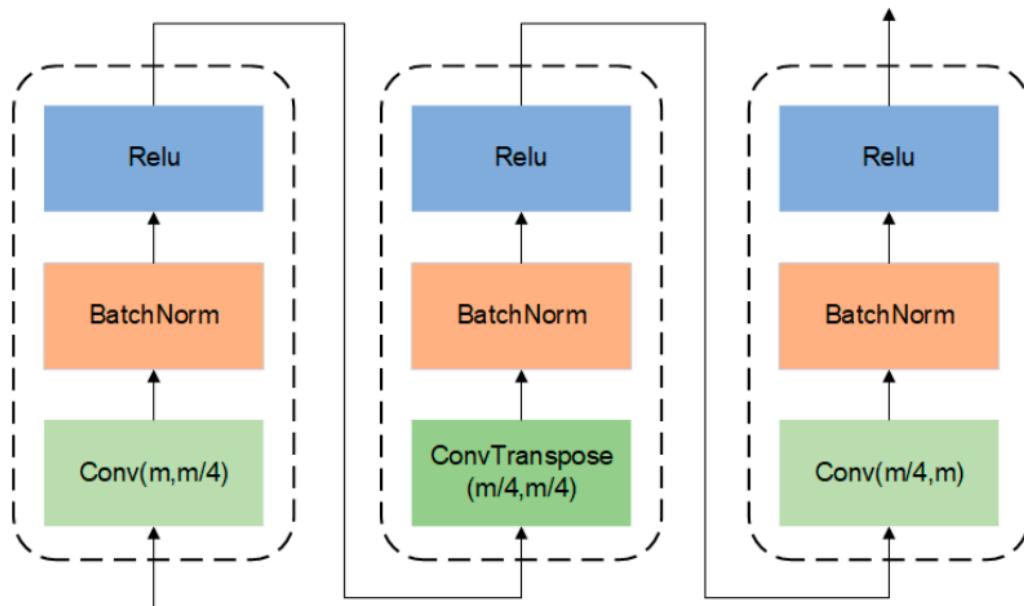
شکل ۴.۴: ساختار بلوک متسع. میدان‌های گیرنده چند مقیاسی از طریق لایه‌های کانولوشنی متسع ساخته می‌شوند، که باعث می‌شود شبکه بتواند ویژگی‌هایی با مقیاس‌های مختلف را استخراج کند.

۴.۴ رمزگشا و تابع هزینه

برای بازیابی جزئیات بخش‌بندی، از یک رمزگشا در جاده‌ساز استفاده می‌شود. به صورت متقارن با رمزگذار، چهار بلوک رمزگشا و یک لایه کانولوشن نهایی برای افزودن نمونه‌برداری^{۲۴} نقشه‌های ویژگی استفاده می‌شوند. شکل ۵.۴ ساختار بلوک رمزگشا را نشان می‌دهد. به طور خاص، در هر بلوک رمزگشا، ویژگی‌ها ابتدا توسط یک لایه کانولوشن 3×3 فیلتر شده و سپس توسط یک لایه کانولوشن ترانهاده، نمونه‌برداری می‌شوند. سپس، ویژگی‌ها دوباره توسط یک لایه کانولوشن 3×3 فیلتر می‌شوند. پس از کانولوشن، ویژگی‌های نمونه‌برداری شده با نتایج رمزگذار در مقیاس مربوطه اضافه می‌شوند.

^{۲۴}Upsampling

پس از عبور از چهار بلوک رمزگشا، یک لایه کانولوشن ترانهاده و دو لایه کانولوشن با هسته‌های 3×3 ، نقشه‌های ویژگی را به اندازه تصویر منبع پردازش می‌کنند. در نهایت، یک طبقه‌بند سیگموئیدی^{۲۵} برای استخراج مناطق جاده با نگاشت خروجی به دامنه $[0, 1]$ اعمال می‌شود، جایی که آستانه 0.5 ، برای طبقه‌بندی مناطق جاده و پس زمینه تنظیم شده است.



شکل ۵.۴: معماری بلوک رمزگشا

تابع هزینه^{۲۶} جاده‌ساز از BCE^{۲۷} و ضریب تاس^{۲۸} تشکیل شده است. تابع هزینه BCE، که اغلب به عنوان بخش‌بندی باینری استفاده می‌شود، به صورت زیر تعریف می‌شود:

$$L_{BCE} = -\frac{1}{N} \sum_i (t_i \cdot \log(o_i) + (1 - t_i) \cdot \log(1 - o_i)) \quad (1.4)$$

که در آن o نتایج پیش‌بینی شده پس از سیگموید، t برچسب واقعی و N اندازه دسته^{۲۹} را نشان می‌دهد.

^{۲۵}Sigmoid Classifier

^{۲۶}Loss Function

^{۲۷}Binary Cross Entropy

^{۲۸}Dice Coefficient Loss

^{۲۹}Batch Size

تقسیم‌بندی جاده یک سنتاریوی خاص است که در آن پیش‌زمینه و پس‌زمینه به شدت نامتوازن هستند. بنابراین، تابع هزینه باید برای توزیع نامتوازن داده‌ها قابل تطبیق باشد. تابع هزینه تاس بیشتر بر استخراج مناطق پیش‌زمینه در حین آموزش مرکز دارد، که سهم نظراتی آن در شبکه، با اندازه تصویر تغییر نمی‌کند. بنابراین، برای حل وضعیتی که پس‌زمینه نسبتاً کم است، مناسب است. فرمول‌بندی تابع هزینه تاس به شرح زیر است:

$$L_{Dice} = 1 - \frac{2 \sum_i o_i \cdot t_i}{\sum_i o_i + \sum_i t_i} \quad (2.4)$$

برای جلوگیری از صفر شدن در مخرج، ما یک پارامتر هموارسازی s اضافه کردیم. L_{Dice} بهینه‌شده به شرح زیر است:

$$L_{Dice} = 1 - \frac{2 \sum_i o_i \cdot t_i + s}{\sum_i o_i + \sum_i t_i + s} \quad (3.4)$$

پارامتر هموارسازی از مشکل تقسیم بر صفر و از بیش‌بازش مدل جلوگیری می‌کند. زیان کلی می‌تواند به صورت زیر محاسبه شود:

$$L_{total} = \alpha L_{BCE} + \beta L_{Dice} \quad (4.4)$$

که در آن α و β وزن‌هایی هستند که می‌توانند دو تابع زیان را متعادل کنند. تابع هزینه طراحی شده‌ی فوق باعث می‌شود که مدل، جاده‌ها را به دقت استخراج کرده و اتصال جاده را حفظ کند. از طریق طراحی تابع هزینه، اطلاعات ویژگی که به طور مؤثر به بخش‌بندی منجر می‌شود، می‌تواند دقت استخراج جاده و حفظ اتصال جاده را به طور همزمان تضمین کند.

۵.۴ آزمایش‌ها و ارزیابی عملکرد

در این قسمت، به بررسی نتایج آزمایش‌ها و تحلیل‌های انجام شده بر روی مدل جاده‌ساز پرداخته می‌شود. ابتدا، مجموعه‌داده‌های مورد استفاده و نحوه آموزش مدل معرفی می‌شوند. سپس معیارهای ارزیابی که معمولاً در وظایف استخراج جاده استفاده می‌شوند، تشریح می‌گردند. به دنبال آن، نتایج آزمایش‌های حذفی به منظور بررسی تأثیر اجزاء مختلف مدل و اعتبارسنجی طراحی آن تحلیل می‌شود. در نهایت، نتایج تصویری و کمی از عملکرد مدل جاده‌ساز در مقایسه با سایر روش‌های روز ۳۰ ارائه و ارزیابی می‌شوند.

^{۳۰}State-of-the-Art

این بخش به گونه‌ای طراحی شده است که خواننده را با جزئیات مربوط به نحوه عملکرد مدل پیشنهادی و مقایسه آن با سایر روش‌ها آشنا کند. همچنین، نقاط قوت و ضعف مدل به دقت بررسی خواهد شد تا درک بهتری از قابلیت‌های مدل جاده‌ساز در استخراج نقشه‌های جاده‌ای از تصاویر ماهواره‌ای به دست آید.

۱.۰.۴ مجموعه‌داده‌ها

در این پژوهش، از مجموعه‌داده‌های گلوب عمیق^{۳۱} و ماساچوست^{۳۲} برای آزمایش استفاده شده است. توضیحات دقیق این دو مجموعه‌داده به شرح زیر است:

۱. مجموعه‌داده گلوب عمیق: این مجموعه‌داده برای چالش استخراج جاده گلوب عمیق ۲۰۱۸ آماده شده است. این مجموعه شامل ۶۲۲۶ تصویر با اندازه 1024×1024 پیکسل می‌باشد.

این تصاویر RGB در فرمت JPG مناطقی از تایلند، هند و اندونزی را پوشش می‌دهند و شامل جاده‌های سیمانی، آسفالتی و کوهستانی هستند. برچسب آن یک تصویر بازیز سه کاناله در فرمت PNG است که به ترتیب از (۲۵۵، ۲۵۵، ۰) و (۰، ۰، ۰) برای نمایش جاده‌ها و پس‌زمینه‌ها استفاده می‌کند. در آزمایش مدل ما، مجموعه‌داده به مجموعه آموزشی (۴۹۸۷ تصویر) و مجموعه تست (۱۲۴۶ تصویر) تقسیم شد.^[۱۰]

۲. مجموعه‌داده ماساچوست: مجموعه‌داده جاده‌های ماساچوست شامل ۱۱۰۸ تصویر برای آموزش، ۱۴ تصویر برای اعتبارسنجی و ۴۹ تصویر برای تست است که همه آن‌ها دارای اندازه 1500×1500 می‌باشند. تصویر منبع در فرمت TIFF یک تصویر رنگی سه کاناله و برچسب آن در فرمت TIFF یک تصویر بازیز است که از رنگ‌های سفید و سیاه برای تمایز جاده‌ها و پس‌زمینه‌ها استفاده می‌کند. جاده‌های سیمانی و آسفالتی نوع اصلی جاده‌ها در این مجموعه‌داده هستند.^[۳۰]

^{۳۱}Deepglobe

^{۳۲}Massachusetts



شکل ۴.۶: چند نمونه از مجموعه داده‌های گلوب عمیق در دو سطر اول و ماساچوست در دو سطر آخر آورده شده است.

۲.۵.۴ پیاده‌سازی آزمایش

در تمامی مدل‌ها، بهینه‌ساز گرادیان نزولی تصادفی^{۳۳} برای دسته‌ای به اندازه‌ی چهار استفاده شده است. SGD یک الگوریتم بهینه‌سازی است که به‌طور گستردۀ در یادگیری ماشین استفاده می‌شود و عمدتاً در فرآیند آموزش مدل به کار می‌رود. این الگوریتم یک نوع از الگوریتم گرادیان نزولی است و بهویژه برای پردازش مجموعه‌داده‌های بزرگ مناسب است. در مقایسه با روش گرادیان نزولی سنتی، در هر تکرار به جای محاسبه گرادیان‌های تمامی نمونه‌ها، تنها گرادیان یک نمونه را به‌طور تصادفی انتخاب می‌کند تا پارامترهای مدل را به‌روزرسانی کند. ضریب کاهش مومنتوم و ضریب کاهش وزن به ترتیب به $0,9 \times 10^{-4}$ و $0,01$ تنظیم شده‌اند. نرخ یادگیری اولیه به $0,01$ تنظیم شده و پس از هر تکرار آموزشی، نرخ یادگیری به مقدار $0,98$ برابر مقدار قبلی کاهش می‌یابد. روش ما با استفاده از چارچوب یادگیری عمیق PyTorch پیاده‌سازی شده و آزمایش‌ها بر روی یک GPU با ۲۴ گیگابایت حافظه انجام شده‌اند. در طول فرآیند آموزش، از چرخش افقی تصادفی، چرخش عمودی تصادفی و مقیاس‌بندی تصادفی به عنوان افزایش داده‌ها استفاده می‌کنیم تا توانایی تعمیم‌دهی مدل را بهبود بخشیم.

^{۳۳}Stochastic Gradient Descent (SGD)

۳.۵.۴ معیارهای ارزیابی

برای ارزیابی عملکرد مدل در وظیفه استخراج جاده، از معیارهای ارزیابی دقت^{۳۴}، بازیابی^{۳۵}، ضریب F1 و شاخص IoU^{۳۶} استفاده شد. این معیارها به صورت زیر تعریف می‌شوند:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5.4)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6.4)$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \quad (7.4)$$

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8.4)$$

۶.۴ آزمایش‌های حذفی

در این بخش، آزمایش حذفی برای ارزیابی عملکرد ماثوله‌های رمزگذار با پشتیبانهای مختلف انجام شده است. ما از ResNet-50 و سری ترنسفور چرخشی به عنوان رمزگذار شبکه استفاده کردیم. بر اساس پیکربندی‌های مختلف، ترنسفور چرخشی می‌تواند به صورت T-Swin-S (اندازه کوچک)، T-Swin-B (اندازه کوچک) و B-Swin (اندازه پایه) تشکیل شود. همانطور که در جدول ۱ نشان داده شده است، T-Swin-B تحت شرایطی که تعداد پارامترهای ResNet-50 قابل مقایسه است، نتایج بهتری کسب کرد. در میان آنها، B-Swin با چهار برابر تعداد پارامترهای ResNet-50 بهترین عملکرد را به دست آورد. برای تجارت عملکرد و هزینه مدل، S-Swin برای آزمایش‌های حذفی بعدی انتخاب شد. در آزمایش‌های مقایسه‌ای، ما عمدتاً از نتایج B-Swin برای مقایسه به دلیل عملکرد بهتر آن استفاده کردیم. در کاربردهایی با نیازهای کارآیی محاسباتی، T-Swin به دلیل اندازه کوچک و سرعت بالای آن یک انتخاب خوب است.

^{۳۴}Precision

^{۳۵}Recall

^{۳۶}Intersection over Union

جدول ۱.۴: مقایسه کمی پشتیبان‌های مختلف برای مدل جاده‌ساز با استفاده از مجموعه‌داده گلوب عمیق

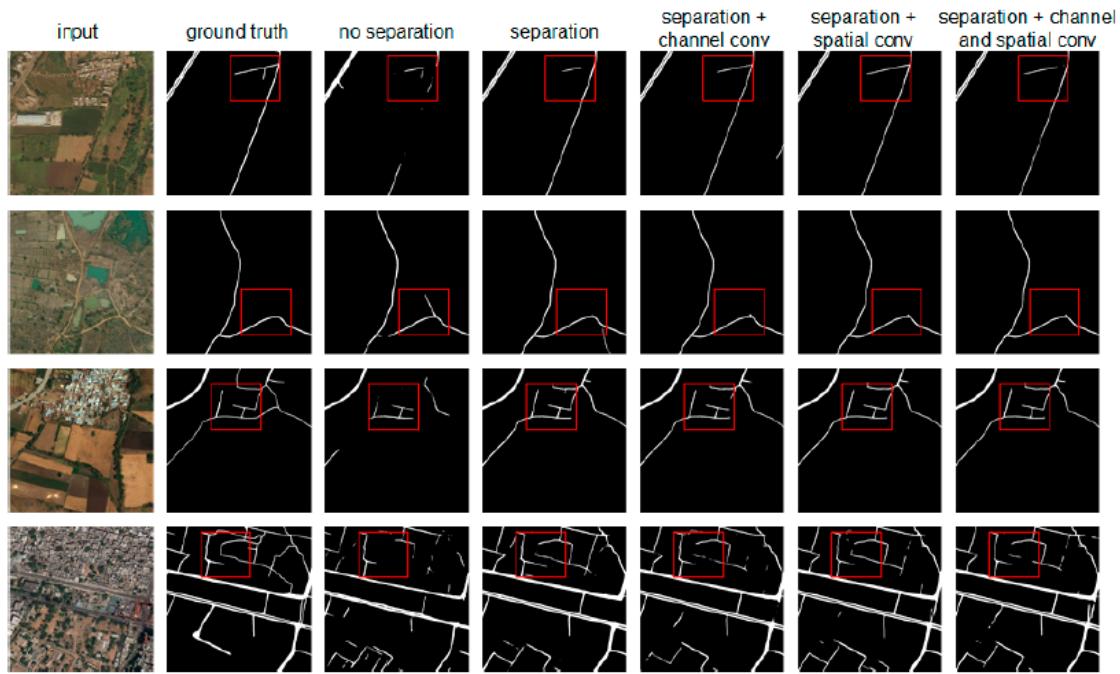
پشتیبان	پارامترها (M)	دقت	بازیابی	IoU	ضریب F1
ResNet-50	۲۱,۶۶	۸۴,۹۱	۷۸,۶۱	۶۸,۱۴	۸۱,۶۴
Swin-T	۲۸,۳۰	۸۴,۰۵	۸۱,۳۴	۶۹,۹۱	۸۲,۶۷
Swin-S	۴۹,۵۹	۸۵,۲۹	۸۲,۵۱	۷۲,۱۸	۸۳,۸۸
Swin-B	۸۸,۰۷	۸۵,۷۶	۸۳,۱۷	۷۳,۱۱	۸۴,۵۰

آزمایش دیگری نیز انجام دادیم تا نشان دهیم بخش گلوبگاه معتبر است. نتایج کمی پیکربندی‌های مختلف ماژول‌ها در جدول ۲.۴ نشان داده شده است. همانطور که از نتایج مشاهده می‌شود، همبستگی فضایی و کanalی عملکرد کلی مدل را به طور قابل توجهی افزایش می‌دهد. بلوک‌های متسع پس از همبستگی فضایی و کanalی قرار داده شدند. به وضوح، جدا کردن ویژگی‌های جهانی و جزئی منطقی است. مدل با جداسازی ویژگی‌ها به طور قابل توجهی بهتر از مدل بدون چنین ساختاری عمل می‌کند. در همین حال، بلوک متسع پس از همبستگی فضایی IoU و ضریب F1 را بهبود می‌بخشد. در مقابل، اگرچه اضافه کردن بلوک متسع پس از همبستگی کanalی دقต را افزایش می‌دهد، سایر معیارهای عملکرد کاهش می‌یابند. این به این دلیل است که اثر بلوک متسع گسترش میدان‌های پذیرش است که با ویژگی‌های فضایی جدا شده سازگار هستند. با این حال، ویژگی‌های کanalی جدا شده توسط همبستگی 1×1 بر اطلاعات پیکسل خود تمرکز دارند و گسترش میدان‌های پذیرش آن‌ها بی‌معنی است. نتایج بالا تأیید می‌کند که همبستگی فضایی به دنبال همبستگی متسع عملکرد را بهبود می‌بخشد، در حالی که این امر پس از همبستگی کanalی بدتر می‌شود.

جدول ۲.۴: مقایسه کمی پیکربندی‌های مختلف گلوبگاه

روش‌ها	جدا سازی ویژگی	بلوک متسع	دقت	بازیابی	IoU	ضریب F1
مدل جاده‌ساز		x			۶۹,۶۶	۸۲,۱۸
مدل جاده‌ساز	✓		x		۸۳,۶۸	۸۳,۲۶
مدل جاده‌ساز	✓			کanal	۷۹,۴۵	۸۲,۹۵
مدل جاده‌ساز	✓			فضایی	۸۲,۵۱	۸۳,۸۸
مدل جاده‌ساز	✓			کanal + فضایی	۸۶,۴۷	۸۳,۵۴
					۸۰,۸۰	۷۱,۷۷
					۸۰,۷۰	۷۱,۲۸

نتایج آزمایش حذف اجزا با پیکربندی‌های مختلف در شکل ۷.۴ ارائه شده است. در بین مدل‌های مختلف پیکربندی شده، نتایج استخراج جاده با جداسازی ویژگی و بلوک‌های متسع، تداوم و جزئیات بهتری دارند. به طور خلاصه، مارژول جداسازی ویژگی عملکرد جامع مدل را به طور مؤثر بهبود می‌بخشد.



شکل ۷.۴: نتایج تصویری از پیکربندی‌های مختلف گلوبگاه. در تصویر اندازه 1024×1024 ، جعبه‌های قرمز با اندازه 320×350 مناطقی را نشان می‌دهند که نتایج استخراج جاده‌ی مدل جاده‌ساز بهتر هستند.

۷.۴ آزمایش‌های مقایسه‌ای

ما آزمایش‌ها را از طریق مقایسه با روش‌های روز ^{۳۷} بر روی مجموعه داده‌های گلوب عمیق و مجموعه داده ماساچوست از نظر دقت، بازیابی، IoU و امتیاز F1 انجام دادیم تا به طور کامل اثربخشی روش‌های پیشنهادی را ارزیابی کنیم. نتایج تصویری مدل پیشنهادی با پنج مدل نمایشی ارائه شده و تحلیل‌های کمی و نتایج در این بخش ارائه شده‌اند.

۱.۷.۴ آزمایش‌ها بر روی مجموعه داده گلوب عمیق

بر روی مجموعه داده گلوب عمیق، مدل جاده‌ساز با سایر مدل‌ها مقایسه شد. در بین روش‌های بالا، FCN و U-Net نماینده مدل‌های کلاسیک بخش‌بندی هستند. PSPNet از یک ساختار هرمی برای جمع‌آوری اطلاعات زمینه استفاده کرد. DeepLabV3 از همپوشانی توسعه‌یافته برای بزرگ‌کردن اندازه ناحیه پذیرش استفاده کرد و ویژگی‌های چندمقیاسی را با استفاده از مازول ASSP تجمعی کرد. SwinUnet یک مدل مبتنی بر ترانسفور است که معمولاً برای بخش‌بندی تصویر پزشکی استفاده می‌شود. ما نتایج تصویری بدست‌آمده توسط مدل جاده‌ساز و این پنج روش نمایشی را در بالا نشان می‌دهیم. برای روش‌های دیگر، نتایج را از منبع اصلی آن‌ها نقل کردیم، بنابراین نتایج تصویری در دسترس نبودند.

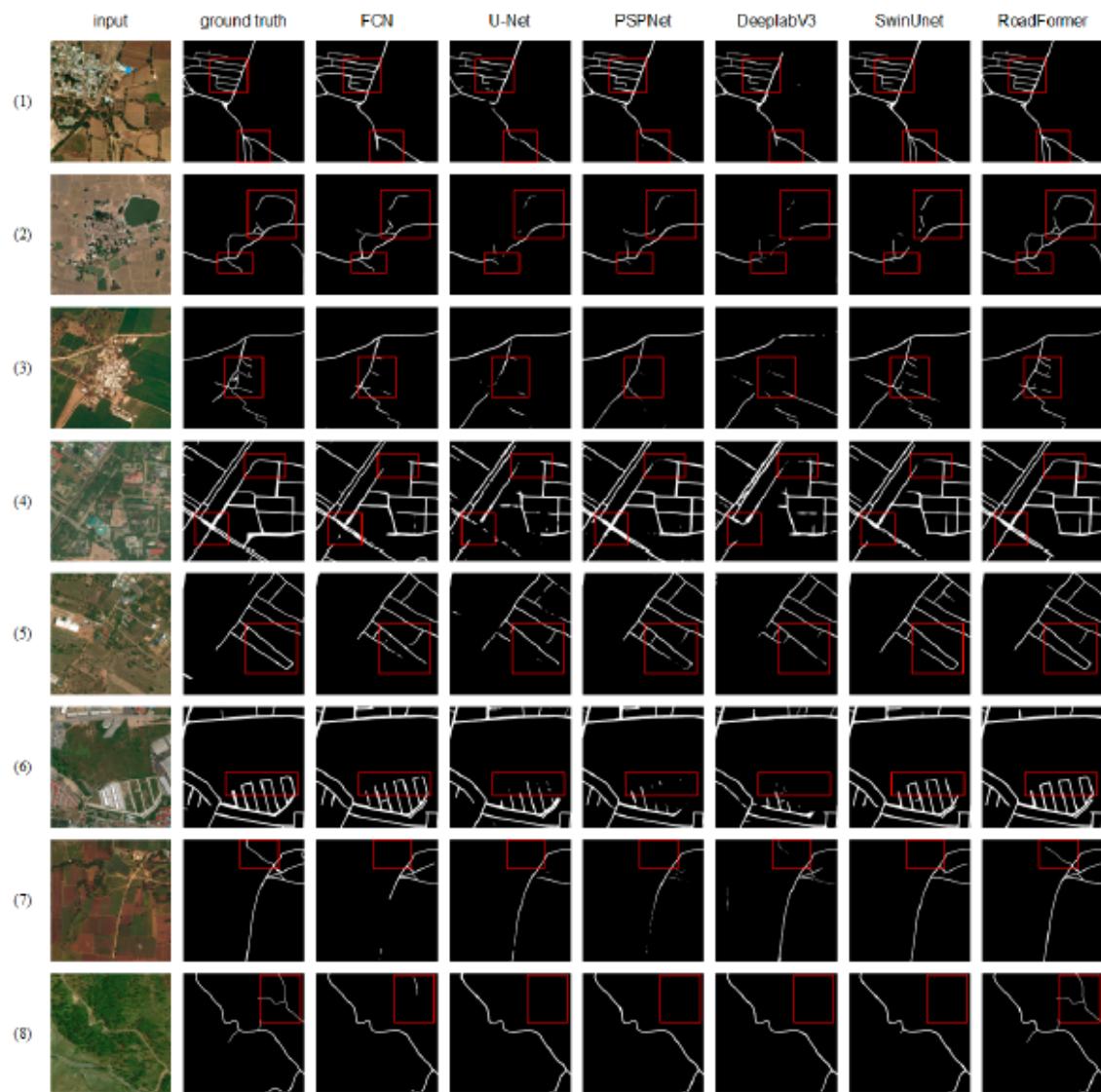
برای ارزیابی شهودی عملکرد استخراج جاده، تصاویر نمایشی از صحنه‌های مختلف از مجموعه تست انتخاب شدند. شکل ۸.۴ نتایج استخراج جاده این تصاویر را با استفاده از شش روش مختلف نشان می‌دهد. تصاویر ورودی، تصاویر واقعیت زمینی، و نتایج مدل‌های نامبرده و مدل جاده‌ساز به ترتیب از چپ به راست نمایش داده شده‌اند.

برای سه تصویر اول از ردیف بالا، نتایج بدست‌آمده توسط U-Net و DeepLabV3 اطلاعات جاده‌ای زیادی را از دست می‌دهند، در حالی که سایر روش‌ها به خوبی عمل می‌کنند.

در صحنه‌های مبهم (چهارم تا ششم ردیف‌ها)، ساختارهای جاده‌ای گستته در نتایج سایر روش‌ها ظاهر شدند و نتایج استخراج مدل جاده‌ساز همچنان کامل باقی می‌مانند.

برای صحنه‌های کم‌کنترast (هفتم و هشتم ردیف‌ها)، هیچ‌یک از پنج روش نمی‌تواند جاده را به طور کامل استخراج کنند، در حالی که مدل جاده‌ساز قادر به استخراج جاده به طور دقیق است. قابل ذکر است که جاده‌های استخراج شده توسط SwinUnet عملکرد بهتری نسبت به سایر مدل‌های مبتنی بر شبکه عصبی پیچشی دارند از نظر تداوم، که به دلیل توانایی طولانی مدت اطلاعات‌گیری که توسط ترانسفور ایجاد شده است. با این حال، SwinUnet برخی از جاده‌های باریک را از دست می‌دهد، در حالی که مدل جاده‌ساز همچنان در این حالت به دلیل طراحی گلوگاه به خوبی عمل می‌کند.

نتایج تصویری نشان می‌دهد که عملکرد مدل جادهساز از سایر روش‌ها برتر است. به دلیل استراتژی جداسازی ویژگی و قابلیت اطلاعات‌گیری طولانی‌مدت مدل جادهساز انسجام و تداوم جاده به خوبی حفظ می‌شود.



شکل ۸.۴: مقایسه تصویری مدل جادهساز و سایر روش‌های روز برای مجموعه داده گلوب عمیق. کادرهای قرمز، مناطقی را که روش ما بهتر از روش‌های دیگر عمل می‌کند، برجسته می‌کند.

جدول ۳.۴: نتایج عملکرد کمی مجموعه داده گلوب عمیق.

FLOPs	پارامترها	F1-Score	IoU	بازیابی	دقت	روش‌ها
۱۹۷,۷G	۴۷,۱M	۷۹,۱	۶۴,۸	۷۵,۵	۸۳,۱	FCN
۶۱,۶G	۳۱,۰M	۷۲,۱	۵۰,۵	۶۴,۶	۸۲,۶	U-Net
۱۷۸,۴G	۴۹,۰M	۷۵,۸	۶۰,۱	۷۰,۱	۸۴,۷	PSPNet
۲۷۰,۰G	–	۷۰,۹	۵۰,۴	۵۸,۳	۷۸,۹	DeepLabV3
–	–	۷۱,۲	۵۳,۵	۷۳,۰	۶۹,۵	SegNet
–	–	۷۲,۸	۵۶,۱	۷۳,۸	۷۱,۸	LinkNet
–	–	۷۶,۶	۶۸,۱	۷۸,۶	۸۴,۹	D-LinkNet
–	–	۷۸,۹	۶۵,۳	۸۰,۱	۷۹,۴	HourGlass
–	–	۷۹,۶	۷۲,۴	۸۴,۱	۸۳,۴	Batra et al.
۲۵۴,۸G	۲۷,۱M	۷۷,۷	۶۲,۹	۷۳,۳	۸۲,۱	SwinUnet
۱۷۶,۵G	۲۸,۳M	۸۲,۷	۶۹,۹	۸۱,۳	۸۴,۱	مدل جاده‌ساز (Swin-T)
۲۶۹,۴G	۴۹,۶M	۸۴,۰	۷۲,۲	۸۲,۵	۸۵,۳	مدل جاده‌ساز (Swin-S)
۴۴۷,۷G	۸۹,۰M	۸۴,۵	۷۳,۱	۸۳,۲	۸۵,۸	مدل جاده‌ساز (Swin-B)

۲.۷.۴ آزمایش‌ها بر روی مجموعه داده ماساچوست

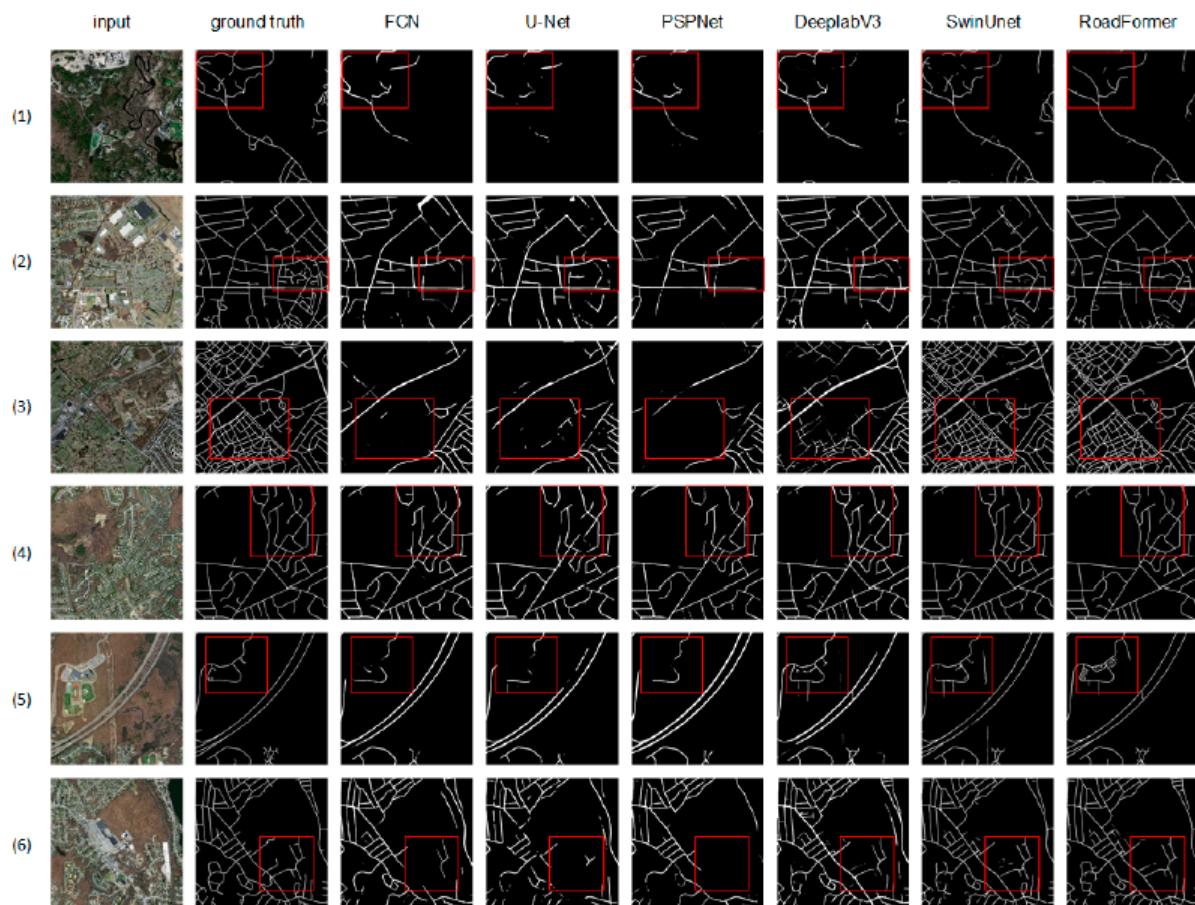
در مجموعه داده ماساچوست، مدل جاده‌ساز با مدل‌های روز که در شکل ۹.۴ آمده‌اند، مقایسه شد. در مقایسه با مجموعه داده گلوب عمیق، مجموعه داده ماساچوست چالش‌برانگیزتر است؛ به دلیل آن که:

۱. وضوح کمتر تصاویر مجموعه داده ماساچوست در مقایسه با گلوب عمیق که به این معنی است که عرض جاده در تصویر باریک‌تر است

۲. تصاویر در مجموعه آموزشی بسیار بزرگ‌تر هستند، که مشکلاتی برای همگرایی مدل ایجاد می‌کند. علاوه بر این، برای اعتبارسنجی کامل توانایی مدل جاده‌ساز در پردازش تصاویر با اندازه بزرگ، ما تصاویر برش‌نخورده را به شبکه ورودی دادیم.

شش تصویر مختلف از مجموعه آزمایشی برای مقایسه تجسم انتخاب شدند. شکل ۹.۴ نتایج تشخیص جاده توسط مدل جاده‌ساز و مدل‌های ذکر شده را در شش ردیف نشان می‌دهد.

در میان نتایج استخراج شده، مدل جاده‌ساز تطبیق‌پذیری قوی‌تری برای صحنه‌های پیچیده نشان داد. برای مثال، جاده‌ها در اولین ردیف توسط محیط اطراف در منطقه کوهستانی مسدود شده‌اند. در این مورد، نتایج استخراج روش‌های مبتنی بر شبکه عصبی کانولوشنی بیشتر اطلاعات جاده را از دست می‌دهند، در حالی که SwinUnet و مدل جاده‌ساز می‌توانند ساختارهای جاده را معتبر را استخراج کنند.



شکل ۹.۴: مقایسه بصری روش مدل جاده‌ساز و سایر روش‌های روز برای مجموعه داده جاده‌ای ماساچوست. جعبه‌های قرمز رنگ مناطقی را نشان می‌دهند که روش ما نسبت به روش‌های دیگر بهتر عمل کرده است.

در ردیف دوم تا چهارم، پنج روش دیگر نمی‌توانند ویژگی‌های جاده‌ای را به درستی استخراج کنند، در حالی که مدل ما همچنان می‌تواند ساختارهای جاده‌ای واضح را به دلیل استفاده از ترانسفور چرخشی استخراج کند. برای جاده‌های باریک (ردیف هفتم و هشتم)، هیچ‌یک از روش‌های دیگر نمی‌توانند ساختارهای جاده‌ای را به طور کامل استخراج کنند، در حالی که مدل جاده‌ساز قادر است این جاده‌های باریک را با دقت استخراج کند. نتایج تجسم در شکل ۹.۴ تطبیق‌پذیری بهتر مدل جاده‌ساز در صحنه‌های پیچیده در مقایسه با روش‌های دیگر را نشان می‌دهد.

جدول ۴.۴: نتایج عملکرد کمی مجموعه داده جاده‌ای ماساچوست.

FLOPs	پارامترها	F1	شاخص IoU	بازخوانی	دقت	روش‌ها
۱۹۷,۷G	۴۷,۱M	۷۴,۷	۵۹,۷	۶۸,۱	۸۲,۸	FCN
۲۰۲,۵G	۲۹,۱M	۷۵,۹	۶۱,۱	۷۰,۳۷	۸۲,۳	U-Net
-	-	۷۶,۴	۶۱,۸	۷۲,۴	۸۰,۹	U-Net++
۱۷۸,۴G	۴۹,۰M	۷۷,۱	۶۲,۷	۷۶,۳	۷۷,۹	PSPNet
۲۷۰,۰G	۶۵,۷M	۷۷,۶	۶۱,۴	۷۶,۰	۷۸,۳	DeepLabV3
-	-	۷۶,۹	۶۲,۵	۷۲,۱	۸۲,۵	SegNet
-	-	۷۷,۹	۶۱,۷	۷۶,۵	۷۹,۵	CADUNet
-	-	۷۵,۱	۶۰,۱	۶۹,۰	۸۱,۹	Batra et al.
-	-	۷۹,۰	۶۵,۳	۷۳,۹	۸۴,۸	SGCN
۲۵۴,۸G	۲۷,۱M	۷۷,۱	۶۲,۸	۷۵,۸	۷۸,۵	SwinUnet
۴۴۷,۷G	۸۹,۰M	۷۹,۲	۶۵,۵	۷۷,۶	۸۰,۷	مدل جاده‌ساز (Swin-B)

نتایج مقایسه بصری و کمی بالا تأیید می‌کنند که روش پیشنهادی ما توانایی بیشتری برای استخراج جاده‌ها دارد. نتایج بصری به وضوح نشان می‌دهند که مدل جاده‌ساز تطبیق‌پذیری بیشتری با جاده‌های پیچیده و سناریوهای جاده‌های دقیق دارد. در جداول ۳.۴ و ۴.۴، می‌توان مشاهده کرد که مدل جاده‌ساز با اندازه کوچک‌تر پیچیدگی محاسباتی کمتر و تعداد پارامترهای کمتری دارد، اما همچنان عملکرد خوبی ارائه می‌دهد. با اینکه مدل جاده‌ساز با اندازه پایه پیچیدگی بیشتری دارد، بهترین عملکرد را نشان می‌دهد. سه عامل زیر عمدهاً دلیل برتری مدل جاده‌ساز هستند:

۱. مازول‌های ترانسفور چرخشی به مدل کمک می‌کنند تا وابستگی بلندمدت را برقرار کند.

۲. با جداسازی ویژگی، یک نقشه ویژگی پالایش شده به مدل اجازه می‌دهد تا ویژگی‌های مختلف را به طور جداگانه درک کند.

۳. میدان دریافتی مدل با بلوک متسع که از کانولوشن فضایی ناشی می‌شود، بیشتر گسترش می‌یابد.

علاوه بر این، بر اساس نتایج کمی، مدل ما بالاترین IoU و شاخص F1 را در مجموعه داده‌های گلوب عمیق و ماساچوست کسب کرده است که برتری روش ما را بیشتر نشان می‌دهد.

۸.۴ نتیجه‌گیری

برای وظایف استخراج جاده، ما یک مدل جدید به نام مدل جاده‌ساز را ارائه دادیم که از ترنسفور چرخشی به عنوان ستون فقرات استفاده می‌کند. کانولوشن جدادشدنی فضایی و کanal در طراحی مدل جاده‌ساز ترکیب شده است تا نمایش ویژگی مدل را بهبود بخشد. علاوه بر این، یک بلوک متسع پس از کانولوشن فضایی به کار گرفته شد که به طور موثر به مدل کمک می‌کند تا اطلاعات متنی جهانی بهتر را بگیرد و میدان‌های دریافتی بزرگتری را به دست آورد. آزمایش‌های حذفی اعتبار طراحی مدل ما را نشان دادند. مجموعه‌داده‌های گلوب عمیق و ماساچوست استفاده شده در آزمایشات ما به دقت ارزیابی شده‌اند. نتایج نشان می‌دهند که روش پیشنهادی ما از روش‌های روز قبلی برتر است، همان‌طور که با مقایسه بصری و نتایج کمی نشان داده شده است، که از برتری و اثربخشی مدل پیشنهادی ما حمایت می‌کنند.

مدل پیشنهادی بر روی مجموعه‌داده‌های تصویربرداری RGB از راه دور گلوب عمیق و ماساچوست آموخته شد. در عمل، دقت استخراج جاده می‌تواند با ادغام داده‌های چندحالتی بیشتر بهبود یابد. به طور خاص، اطلاعات DEM و پس‌زمینه جغرافیایی بسیار مهم هستند که باعث می‌شود ویژگی‌های جاده در مناطق خاص راحت‌تر استخراج شوند. علاوه بر این، کanal‌های چندگانه اطلاعات در تصویربرداری نوری و راداری می‌توانند اطلاعات متفاوتی در مورد جاده ارائه دهند. علاوه بر این، معماری بلوک ترنسفور چرخشی‌می‌تواند بیشتر بهینه‌سازی و تنظیم شود و توابع مختلف از دست دادن می‌توانند برای بهبود عملکرد مدل بررسی شوند. در آینده، ما مجموعه‌داده‌های چندحالتی تصویربرداری از راه دور ذکر شده در بالا را جمع‌آوری خواهیم کرد و عملکرد مدل را با بهینه‌سازی معماری و تابع از دست دادن بیشتر بهبود خواهیم بخشید.

۹.۴ جمع‌بندی

در این فصل، معماری مدل جاده‌ساز به طور جامع بررسی و تحلیل شد. ابتدا، طراحی کلی مدل و اجزاء مختلف آن شامل رمزگذار، گلوگاه و رمزگشا توضیح داده شد. رمزگذار به عنوان بخشی که تصویر ورودی را به نقشه‌های ویژگی با مقیاس‌های مختلف کاهش می‌دهد، معرفی شد. در این مرحله، از ترنسفور چرخشی برای مدل سازی روابط بلندمدت استفاده شد که باعث افزایش دقت و کارایی مدل در استخراج جاده‌ها از تصاویر ماهواره‌ای شد.

سپس، در بخش گلوگاه، فرآیند جداسازی ویژگی‌ها به ویژگی‌های کanalی و مکانی مورد بحث قرار گرفت. استفاده از کانولوشن متسع و مازول‌های قابل جداسازی در این بخش باعث شد تا میدان گیرنده به طور مؤثر گسترش یابد و اطلاعات مکانی و کanalی به خوبی پالایش شوند.

در نهایت، رمزگشا و تابع هزینه مدل مورد بررسی قرار گرفت. رمزگشا با استفاده از بلوک‌های متعدد، توانست جزئیات بخش‌بندی را بازسازی کرده و نقشه‌های جاده را با دقت بالا ارائه دهد.

این فصل نشان داد که مدل جاده‌ساز با استفاده از تکنیک‌های پیشرفته پردازش تصویر و یادگیری ماشین، توانایی بالایی در استخراج نقشه‌های جاده با دقت و کارایی بالا دارد. از طریق طراحی دقیق و استفاده از روش‌های بهینه‌سازی مناسب، این مدل توانست به طور مؤثری به چالش‌های مربوط به استخراج نقشه‌های جاده از تصاویر ماهواره‌ای پاسخ دهد.

سپس مجموعه‌داده‌های آموزشی، شامل مجموعه‌داده‌های گلوب عمیق و ماساچوست معرفی شدند و نحوه آموزش مدل تشریح گردید. بعد از آن، معیارهای ارزیابی عملکرد مدل شامل دقت، بازیابی، ضریب F1 و شاخص IoU توضیح داده شد و تابع هزینه طراحی شده، با ترکیب BCE و ضریب تاس، به مدل کمک کرد تا مناطق جاده را به خوبی از پس‌زمینه تمیز کند.

نتایج آزمایش‌ها نشان داد که مازول‌های مختلف رمزگذار و گلوگاه در مدل جاده‌ساز تأثیر قابل توجهی بر عملکرد مدل دارند. به ویژه، استفاده از ترانسفور چرخشی در رمزگذار و استفاده از جداسازی ویژگی‌ها و بلوک‌های متسع در گلوگاه به بهبود قابل توجهی در دقت و کارایی مدل منجر شد.

در ادامه، نتایج مقایسه‌ای مدل جاده‌ساز با سایر روش‌های برتر بر روی مجموعه‌داده‌های گلوب عمیق و ماساچوست نشان داده شد. این نتایج نشان دادند که مدل جاده‌ساز از نظر دقت، بازیابی، ضریب F1 و شاخص IoU عملکرد بهتری نسبت به سایر روش‌های موجود دارد. به ویژه، این مدل توانست در شرایط پیچیده‌تر و صحنه‌های با وضوح پایین، همچنان نتایج مطلوبی ارائه دهد.

در مجموع، این فصل نشان داد که مدل جاده‌ساز با استفاده از تکنیک‌های پیشرفته پردازش تصویر و یادگیری عمیق، توانایی بالایی در استخراج دقیق و کارآمد نقشه‌های جاده‌ای از تصاویر ماهواره‌ای دارد. با بهکارگیری معماری و تکنیک‌های مناسب، این مدل موفق به ارائه نتایج برتری نسبت به سایر روش‌ها شده و قابلیت‌های بالقوه آن برای استفاده در کاربردهای واقعی اثبات شده است.

۵ فصل

نتیجه‌گیری

این پایان‌نامه به بررسی استخراج نقشه‌های خیابان از تصاویر ماهواره‌ای با استفاده از روش‌های یادگیری عمیق پرداخته است. هدف اصلی این پژوهش ارائه روشی بهینه برای شناسایی و استخراج خیابان‌ها از تصاویر ماهواره‌ای بوده است. در این بخش، نتایج حاصل از آزمایشات انجام‌شده، مورد بررسی قرار گرفته و پیشنهاداتی برای کارهای آینده ارائه می‌شود.

۱.۵ دستاوردهای اصلی

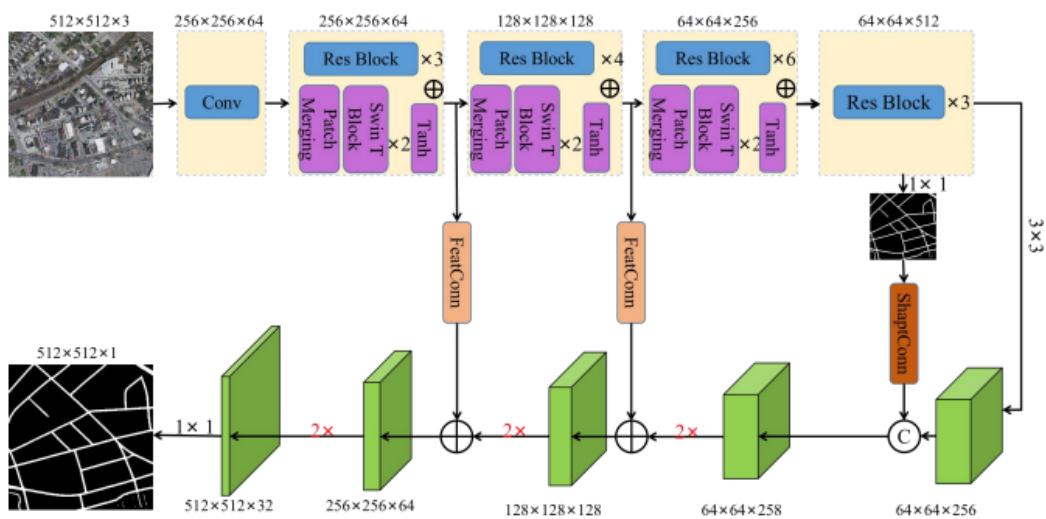
در طول این پژوهش، مدل‌های مختلف یادگیری عمیق برای استخراج خیابان‌ها از تصاویر ماهواره‌ای مورد بررسی و آزمایش قرار گرفتند. نتایج نشان دادند که این مدل‌ها توانایی بالایی در شناسایی ویژگی‌های پیچیده و ساختارهای محلی در تصاویر دارند. استفاده از مدل‌های ترکیبی مانند ترانسفورمر چرخشی به همراه شبکه‌های عصبی کانولوشنی، که در مقاله معرفی شده [۴۱]، توانسته است قابلیت استخراج اطلاعات محلی (ویژگی‌های محلی) و اطلاعات جهانی (زمینه‌یابی جهانی) را ترکیب کند و در نتیجه به بهبود دقیقت در استخراج جاده‌ها از تصاویر ماهواره‌ای کمک کند. این ترکیب باعث شد تا مدل‌ها بتوانند با دقیقت بیشتری نقاط قطع شده جاده‌ها را تشخیص دهند و نواحی با موضع طبیعی و غیرطبیعی (مانند درختان و ساختمان‌ها) را بهبود بخشنند.

۲.۵ مدل‌نهایی

در این قسمت به بررسی مدل‌نهایی که نتیجه‌ی بهتری را نسبت به مدل‌های دیگر به ثبت رساند؛ پرداخته می‌شود.

۱.۲.۵ ساختار کلی

شبکه پیشنهادی شامل یک معماری ساعت شنبی است که از مازول CNNSwin برای کدگذاری و بازسازی اطلاعات استفاده می‌کند. این مازول مزایای شبکه عصبی پیچشی برای استخراج ویژگی‌های محلی و ترانسفورمر چرخشی برای پردازش اطلاعات جهانی را ترکیب می‌کند. مسیر کدگذاری شامل سه مازول متصل است که به منظور به دست آوردن ویژگی‌های محلی و زمینه کلی از تصاویر سنجش از دور طراحی شده‌اند.



شکل ۱.۵: معماری کلی شبکه‌ی عصبی پیچشی چرخشی

مدل شبکه عصبی پیچشی چرخشی یک مازول کدگذاری جدید است که برای استخراج اطلاعات جاده از تصاویر با وضوح بالا طراحی شده است. این مازول از دو مسیر موازی تشکیل شده است: یکی شامل بلوک‌های شبکه عصبی باقی‌مانده و دیگری شامل دو بلوک متوالی ترانسفورمر چرخشی با استفاده از طراحی سلسه مراتبی و مکانیزم پنجره‌ی جابجاشده، اطلاعات جهانی را پردازش می‌کند، در حالی که شبکه عصبی باقی‌مانده ویژگی‌های محلی را استخراج می‌کند.

در حین فرآیند یادگیری، اختلافات آماری بین ویژگی‌های شبکه عصبی باقی‌مانده و ترانسفورمر چرخشی وجود دارد. برای هماهنگ‌سازی این ویژگی‌ها، از تابع تائزانت هیپربولیک^۱ برای کاهش تفاوت‌ها و بهبود تلفیق ویژگی‌ها استفاده می‌شود. فرمول استفاده شده برای این هدف به صورت زیر است:

$$z_j = x_j + \tanh(y_j)$$

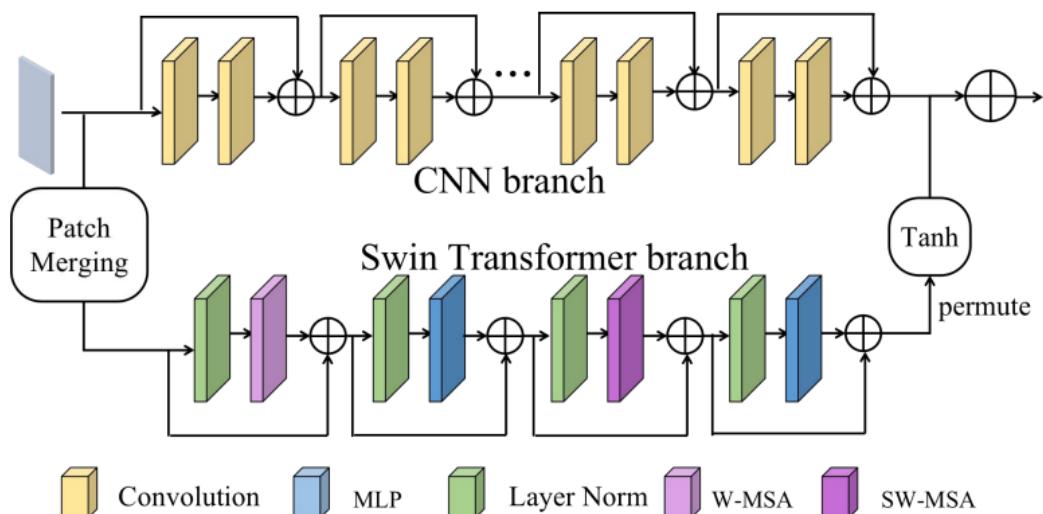
^۱tanh

که در آن:

- z ویژگی نهایی است،

- x_j ویژگی استخراج شده توسط ResNet

- y_j ویژگی استخراج شده توسط ترانسفورمر چرخشی است.

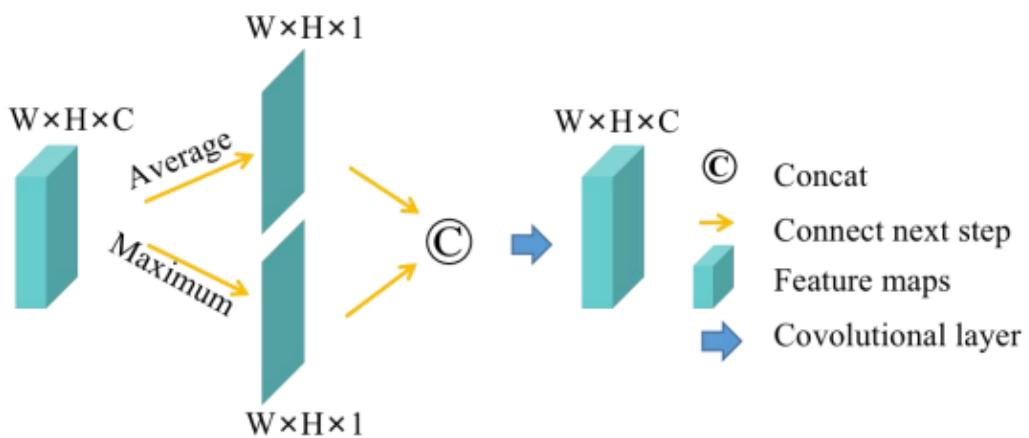


شکل ۲.۵: ساختار مدل شبکه عصبی پیچشی چرخشی

۲.۲.۵ ساختار اتصال تقویت شده با ویژگی

ساختار اتصال تقویت شده^۲ با ویژگی یک ساختار اتصال است که به منظور انتقال اطلاعات جزئی از کدگذار به کدگذار در سطح یکسان طراحی شده است. این ساختار، ویژگی های محلی را از طریق محاسبه میانگین و بیشینه ویژگی های نقشه های ویژگی منتقل می کند.

^۲FeatConn



شکل ۳.۵: ساختار اتصال تقویت شده با ویژگی

ورودی این ساختار به دو بخش تقسیم می‌شود: میانگین و بیشینه. سپس این دو مقدار به یکدیگر متصل می‌شوند و از یک لایه کانولوشنی برای پردازش آنها استفاده می‌شود. فرمول ریاضی برای این فرایند به صورت زیر است:

$$Y = \text{Conv}(\text{Concat}(\text{Avg}(X), \text{Max}(X)))$$

که در آن:

- X ورودی به FeatConn است،

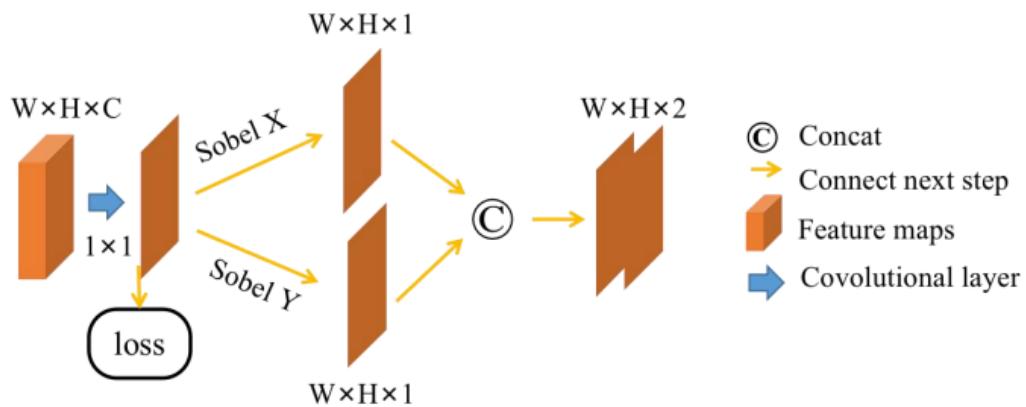
- $\text{Avg}(X)$ میانگین نقشه ویژگی‌ها و $\text{Max}(X)$ بیشینه نقشه ویژگی‌ها است،

- و Concat عملیات الحاق و Conv کانولوشن است.

۳.۲.۵ ساختار اتصال تقویت شده با شکل

ساختار اتصال تقویت شده با شکل ۳ ساختاری است که برای استخراج و ترکیب اطلاعات جهت جاده طراحی شده است. در این ساختار، از یک نقشه بخش‌بندی اولیه جاده استفاده می‌شود تا مؤلفه‌های در جهت‌های x و y محاسبه شوند. این مؤلفه‌ها به عنوان ویژگی‌های جهت‌دار جاده به نقشه ویژگی‌ها اضافه می‌شوند.

^rshapConn



شکل ۴.۵: ساختار اتصال تقویت شده با شکل

فرض کنید ورودی X' نقشه‌ی بخش‌بندی جاده است و مؤلفه‌های در جهت‌های x و y به ترتیب با S_x و S_y نمایش داده می‌شوند. فرمول محاسبه خروجی به شکل زیر است:

$$Y = \text{Concat}(S_x * X', S_y * X')$$

که در آن:

- S_y و S_x فیلترها در جهت‌های x و y هستند،
- Concat عملیات الحق ویژگی‌ها و $*$ نشان‌دهندهٔ عملیات کانولوشن است.

۳.۵ تابع هزینه

در این بخش، تابع هزینه‌ای معرفی می‌شود که برای آموزش شبکه استخراج جاده استفاده شده است. از آنجا که پوشش جاده‌ها به طور معمول بسیار کمتر از پس‌زمینه است، داده‌ها به صورت نامتوازن هستند و این می‌تواند منجر به استخراج ناکافی جاده‌ها شود. برای کاهش این مشکل، مدل بخش‌بندی جاده با استفاده از تابع هزینه‌ی آنتروپی متقاطع وزنی^۴ آموزش داده می‌شود. فرمول تابع هزینه به شکل زیر است:

$$L_1 = \frac{1}{N} \sum_{i=1}^N ((1 + \alpha \cdot w_i) y_i \log(y'_i) + (1 - y_i) \log(1 - y'_i)) + L_{smooth}$$

که در آن:

- y_i مقدار پیش‌بینی شده،
- y'_i مقدار واقعی،

⁴Weighted Cross-Entropy Loss

- N تعداد کل پیکسل‌ها،

- w_i وزن تعریف شده بر اساس نسبت پیکسل‌های جاده به کل پیکسل‌ها است، - و α یک پارامتر ثابت است.

علاوه بر این، از یک تابع منظم‌سازی^۵ برای پایدارسازی فرآیند آموزش استفاده می‌شود. در ساختار اتصالات میانی شبکه، از تابع هزینه‌ی آتروپی باینری^۶ برای دستیابی به یک نتیجه‌ی اولیه‌ی بخش‌بندی جاده استفاده شده است.

به طور کلی، تابع هزینه نهایی شامل دو مؤلفه است:

۱. L_1^{**} که نمایانگر هزینه نهایی نتیجه‌ی بخش‌بندی است.

۲. L_2^{**} که نمایانگر هزینه‌ی نتیجه‌ی میانی بخش‌بندی است.

برای بهینه‌سازی وزن‌دهی این دو تابع هزینه و جلوگیری از تنظیم دستی در طول آموزش، از روش تقریب بیشینه‌ی درست‌نمایی گاوی^۷ استفاده شده است:

$$L_T \approx \frac{1}{\sigma_1^2} L_1 + \frac{1}{\sigma_2^2} L_2 + \log \sigma_1 \sigma_2$$

که در آن:

- σ_1 و σ_2 انحراف استاندارد هستند و با استفاده از تبدیل نمایی پارامترهای قابل یادگیری تخمین زده می‌شوند.

۱.۳.۵ متريک دقت کلي

OA^8 نشان‌دهنده‌ی درصد نمونه‌هایی است که به درستی پیش‌بینی شده‌اند. فرمول محاسبه OA به صورت زیر است:^۹

$$OA = \frac{TP + TN}{TP + FP + FN + TN}$$

که در آن:

- TP نشان‌دهنده‌ی نمونه‌های مثبت به درستی تشخیص داده شده،

- TN نشان‌دهنده‌ی نمونه‌های منفی به درستی تشخیص داده شده،

- FP نشان‌دهنده‌ی نمونه‌های مثبت اشتباه تشخیص داده شده،

- و FN نشان‌دهنده‌ی نمونه‌های منفی اشتباه تشخیص داده شده است.

⁵Smooth-L1

⁶BCE

⁷Gaussian Likelihood Maximization

⁸Overall Accuracy

⁹Overall Accuracy

۴.۵ بررسی مجموعه‌داده‌ی ماساچوست

مجموعه‌داده‌ی ماساچوست شامل تصاویری از مناطق شهری و روستایی است که از تصاویر هوایی به دست آمده‌اند. این مجموعه‌داده شامل ۱۱۰۸ تصویر برای آموزش، ۱۴ تصویر برای اعتبارسنجی و ۴۹ تصویر برای آزمون است. هر تصویر دارای ابعاد 1500×1500 پیکسل است که در آزمایش‌ها به 512×512 پیکسل تغییر اندازه داده شده‌اند.

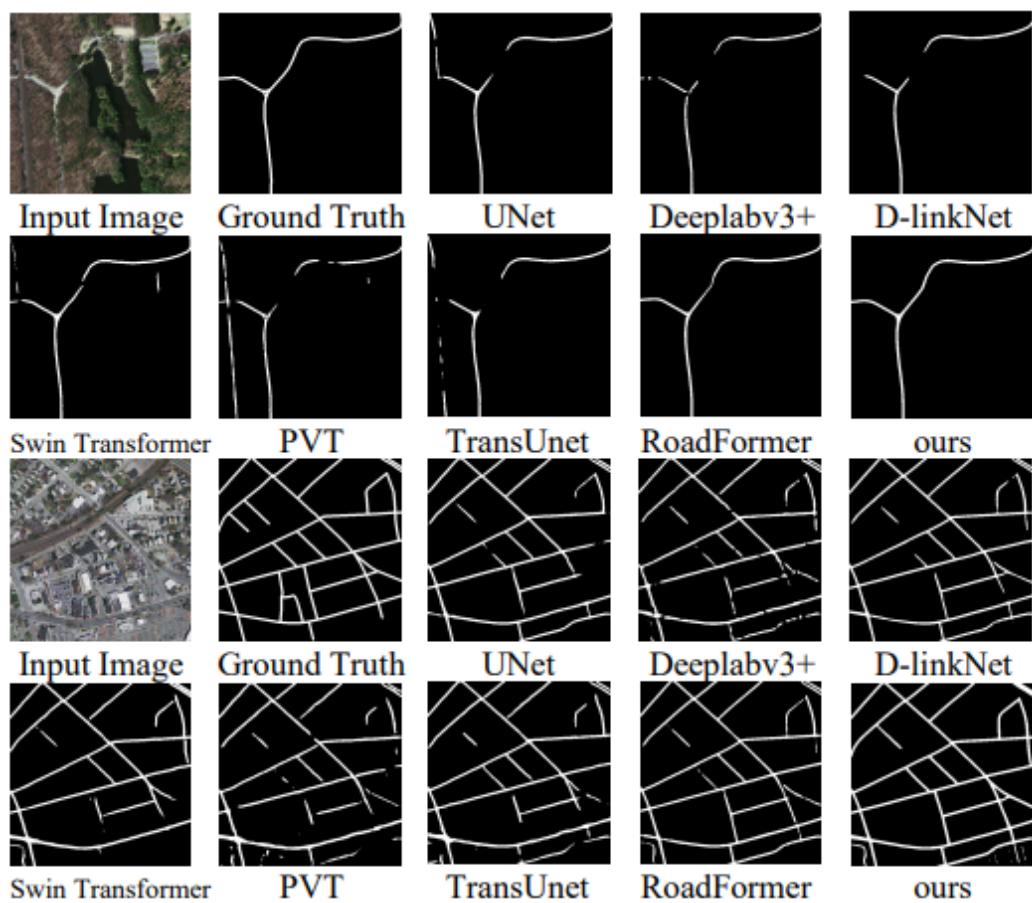
۱.۴.۵ نتایج مربوط به ماساچوست

در آزمایش‌های انجام شده، مدل پیشنهادی ما در مقایسه با روش‌های پیشرفته دیگر عملکرد بهتری را نشان داده است. نتایج مربوط به مجموعه‌داده ماساچوست در جدول زیر به صورت درصد ارائه شده است:

Model	F1	IOU	OA
U-Net	۷۷,۴۵	۶۳,۸۲	۹۷,۹۹
Deeplabv۳+	۷۷,۸۲	۶۴,۱۹	۹۸,۰۳
D-LinkNet	۷۸,۱۱	۶۴,۴۵	۹۸,۰۵
Transformer Swin	۷۸,۹۰	۶۵,۶۵	۹۸,۱۳
PVT	۷۹,۰۶	۶۵,۷۴	۹۸,۱۳
TransUNet	۷۹,۵۱	۶۶,۰۱	۹۸,۱۴
RoadFormer	۷۹,۶۰	۶۶,۱۲	۹۸,۱۵
مدل پیشنهادی	۸۰,۱۳	۶۶,۸۴	۹۸,۱۵

همان‌طور که مشاهده می‌شود، مدل پیشنهادی توانسته است با دقت ۹۸,۱۵٪ در معیار OA بهترین عملکرد را در دیتاست ماساچوست نشان دهد. این مدل همچنین در معیارهای F1 و IOU نیز عملکرد بهتری نسبت به سایر مدل‌ها دارد.

در این تصاویر، مدل پیشنهادی توانسته است با حفظ پیوستگی جاده‌ها و استخراج دقیق‌تر جاده‌ها در مناطق با پوشش گیاهی یا انسداد، عملکرد بهتری نسبت به سایر مدل‌ها داشته باشد.



شکل ۵.۵: نمونه‌ای از نتایج مدل‌های مختلف بر روی دیتابست ماساچوست.

۵.۵ جمع‌بندی

در این مقاله، یک روش جدید برای استخراج جاده از تصاویر با وضوح بسیار بالا معرفی شد. آزمایش‌ها نشان دادند که این مدل توانایی بالایی در استخراج دقیق جاده‌ها داشته و می‌تواند برای مسیریابی در نواحی صعب‌العبور و عملیات‌های امداد و نجات بسیار مفید باشد.

۶.۵ محدودیت‌ها

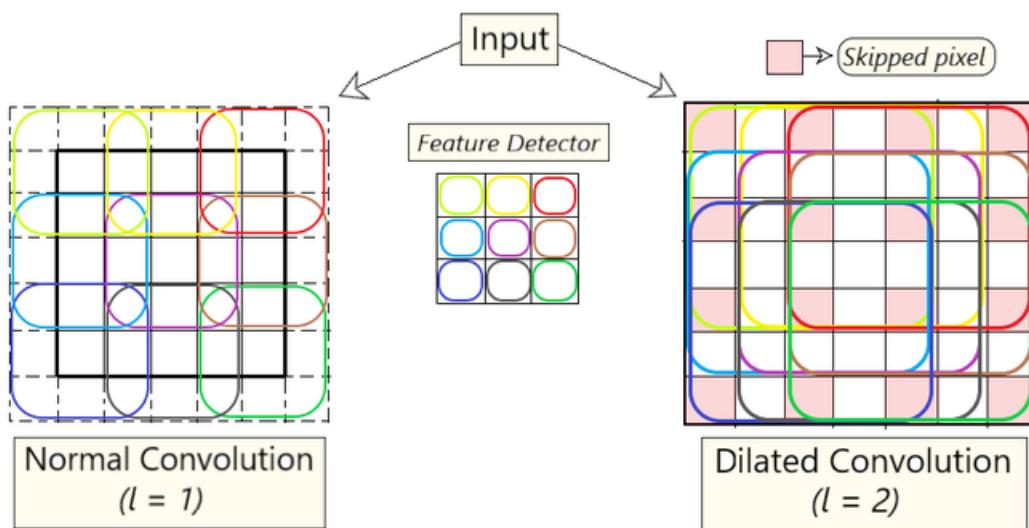
یکی از چالش‌های اصلی در این پژوهش، عدم دقیقت مدل‌ها در تصاویر با کنترast پایین بود. در این نوع تصاویر، مدل‌های یادگیری عمیق معمولاً دچار کاهش دقیقت در شناسایی ویژگی‌های خیابان‌ها می‌شوند. همچنین، محاسبات سنگین و زمان‌بر بودن آموزش مدل‌ها از دیگر محدودیت‌های این پژوهش بود.

۷.۵ پیشنهادات برای کارهای آینده

با توجه به محدودیت‌هایی که مدل مورد استفاده در این پژوهه داشت، به خصوص در مواجهه با موانع طبیعی و غیرطبیعی که منجر به عدم نمایش صحیح جاده‌ها در تصاویر ماهواره‌ای می‌شود، در این بخش به بررسی روش‌ها و ایده‌هایی برای بهبود این مدل پرداخته می‌شود.

۱.۷.۵ استفاده از الگوریتم‌های گسترش‌یافته برای پیش‌بینی بهتر جاده‌ها

یکی از مشکلات عمده این پژوهه، وجود موانع طبیعی مانند درختان و ساختمان‌ها بود که باعث قطع شدن جاده‌ها در تصاویر ماهواره‌ای می‌شد. برای حل این مشکل، می‌توان از الگوریتم گسترش‌یافته^{۱۰} استفاده کرد.



شکل ۶.۵: کانولوشن معمولی در برابر کانولوشن گسترش‌یافته

تعريف الگوریتم گسترش‌یافته‌ی کانولوشنی: این الگوریتم نوعی از کانولوشن است که در آن به جای استفاده از فیلترهای پیوسته در تصویر، فاصله‌هایی بین مقادیر مختلف تصویر قرار می‌گیرد. این فاصله‌ها باعث می‌شوند که شبکه عصبی بتواند ویژگی‌های گسترده‌تر و اطلاعات از نقاط دورتر تصویر را بدون از دست دادن رزولوشن تشخیص دهد. این روش به مدل اجازه می‌دهد تا جاده‌های پوشیده شده توسط موانع طبیعی مانند درختان را بهتر شناسایی کند و نقاط قطع شده را به درستی پر کند.

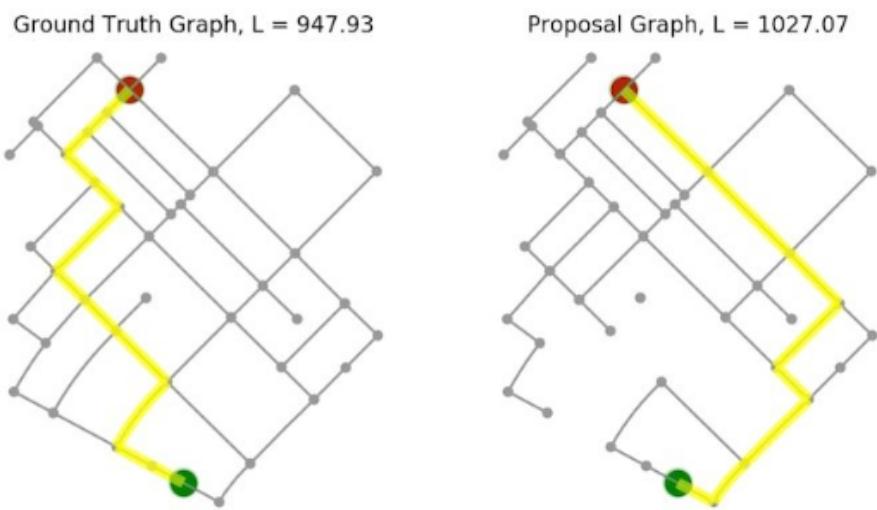
^{۱۰}Dilated

۲.۷.۵ استفاده از متریک‌های گرافی برای بهبود دقت مدل

متریک‌های گرافی ابزار مناسبی برای مدل‌سازی و نمایش شبکه‌های جاده‌ای هستند و می‌توانند به بهبود دقت مدل و آموزش بهتر آن کمک کنند.

معرفی متریک‌های گرافی:

- گره^{۱۱}: گره‌ها نقاط تقاطع یا پایانی در یک شبکه جاده‌ای را نشان می‌دهند.
- یال^{۱۲}: یال‌ها جاده‌ها را به صورت خطوطی که گره‌ها را به هم متصل می‌کنند، نشان می‌دهند.
- درجه گره^{۱۳}: درجه یک گره تعداد یال‌هایی است که به آن گره متصل هستند. این متریک می‌تواند برای تشخیص مناطق پرترافیک یا مهم استفاده شود.
- ضریب خوشبندی^{۱۴}: این متریک میزان تراکم یا اتصال گره‌ها در یک بخش خاص از شبکه را نشان می‌دهد.
- کوتاه‌ترین مسیر^{۱۵}: این متریک نشان‌دهنده حداقل فاصله بین دو گره در شبکه است و می‌تواند برای مسیریابی بهینه استفاده شود.



شکل ۷.۵: متریک APLS تفاوت طول مسیر را بین ماسک واقعی و ماسک پیشنهادی مقایسه می‌کند. سمت چپ: کوتاه‌ترین مسیر بین گره منبع (سبز) و هدف (قرمز) در نمودار حقیقت زمین به رنگ زرد با طول مسیر ≈ 948 متر نشان داده شده است. راست: کوتاه‌ترین مسیر بین گره منبع و هدف در نمودار پیشنهادی با 30 یال حذف شده، با طول مسیر ≈ 1027 متر. این تفاوت در طول، اساس متریک را تشکیل می‌دهد.

^{۱۱}Node

^{۱۲}Edge

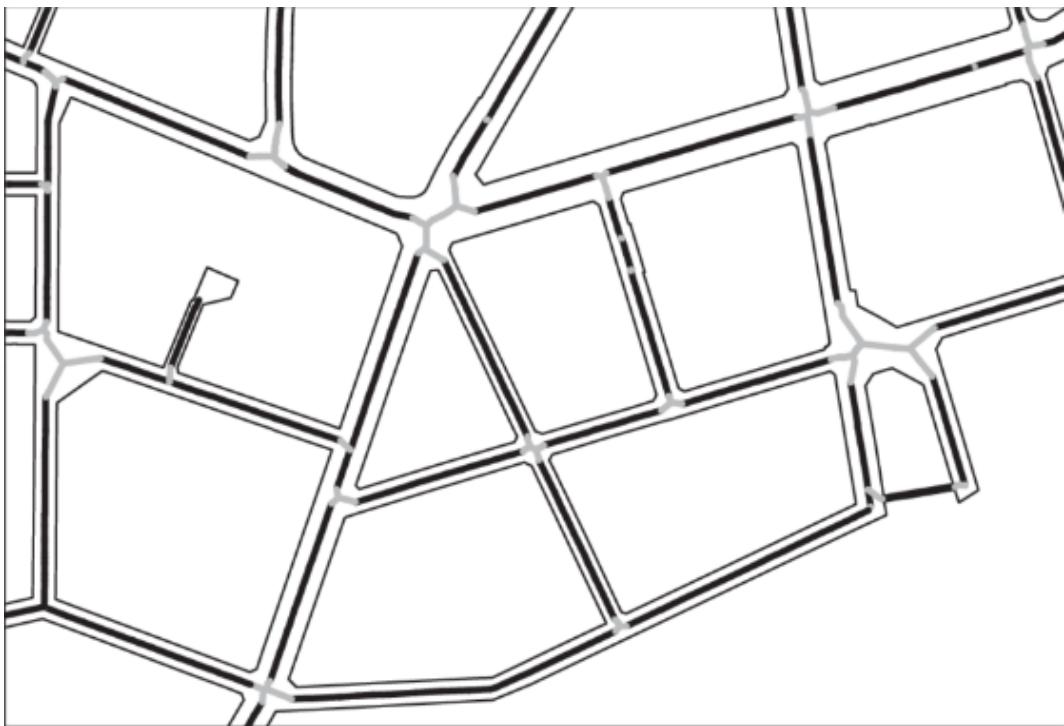
^{۱۳}Degree

^{۱۴}Clustering Coefficient

^{۱۵}Shortest Path

۳.۷.۵ استفاده از الگوریتم اسکلت‌سازی برای ترسیم گراف جاده‌ها

برای نمایش بهتر جاده‌ها به صورت شبکه گرافی، الگوریتم اسکلت‌سازی^{۱۶} می‌تواند مفید باشد.



شکل ۸.۵: اسکلت پس از کاهش به خط مرکزی

تعریف Skeltonize: این الگوریتم وظیفه دارد که از تصاویر دوبعدی جاده‌ها، یک "اسکلت" باریک و ساده از مسیر جاده‌ها استخراج کند. این اسکلت نمایی از ساختار اصلی جاده‌ها است که شامل نقاط و خطوط کلیدی (گره‌ها و یال‌ها) می‌شود.

چرا Skeltonize برای ترسیم گراف مفید است؟

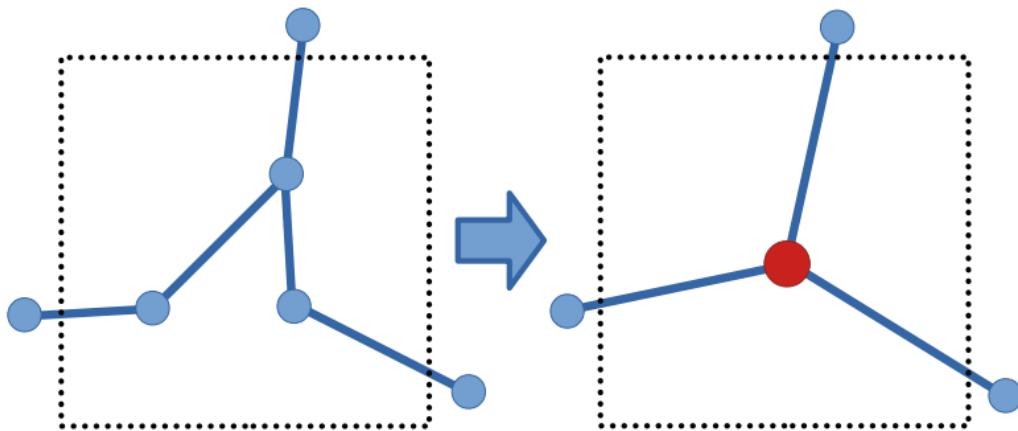
با استفاده از این الگوریتم، می‌توان جاده‌ها را به ساده‌ترین شکل ممکن ترسیم کرد و این اسکلت را به عنوان ورودی برای ساخت گرافی از جاده‌ها استفاده کرد. به کمک این فرآیند، گره‌ها و یال‌ها با دقت بیشتری شناسایی می‌شوند و می‌توان برای حل مسائلی مانند مسیریابی و تخمین مسیرهای قطع شده از آن استفاده کرد.

^{۱۶}Skeltonize

۴.۷.۵ چالش‌های استخراج جاده از تصاویر هوایی با استفاده از شبکه‌های عصبی گرافی

پس از مطالعه مقاله‌ی [۴۴] که در سال ۲۰۲۱ منتشر شده است، چندین چالش اصلی در زمینه استخراج نقشه‌های جاده‌ای با استفاده از شبکه‌های عصبی دیده می‌شود که در ادامه آن‌ها بیان شده است:

- دقت پایین در تشخیص مرزهای جاده‌ها:
یکی از چالش‌های اصلی در استخراج جاده‌ها از تصاویر ماهواره‌ای، دقت پایین در تشخیص مرزهای جاده‌ها است. برخی از مدل‌های پیشین قادر به تشخیص صحیح مرزهای جاده نیستند، به ویژه در مکان‌هایی که جاده‌ها با سایر عناصر، مانند پارکینگ‌ها و مناطق سبز، هم‌پوشانی دارند.
- تراکم بالای تصاویر و تداخل با سایر اشیاء:
در مناطق شهری که تراکم ساختمان‌ها و اشیاء زیاد است، جداسازی جاده‌ها از دیگر اشیاء سخت می‌شود. این تراکم بالا و هم‌پوشانی اشیاء باعث می‌شود که مدل‌های سنتی قادر به شناسایی دقیق جاده‌ها نباشند.
- پیچیدگی اتصالات جاده‌ها:
پیش‌بینی و تشخیص اتصالات بین جاده‌ها، مانند تقاطع‌ها و انشعابات، چالشی مهم است. در بسیاری از موارد، شبکه‌های عصبی گرافی باید بتوانند اتصالات درست بین نقاط مختلف جاده‌ها را با دقت بالا تشخیص دهند، در غیر این صورت مدل نمی‌تواند نقشه کامل و درستی از جاده‌ها ارائه دهد.
- مشکلات مرتبط با وضوح پایین تصاویر:
تصاویر ماهواره‌ای معمولاً دارای وضوح پایین هستند که می‌تواند باعث از دست رفتن اطلاعات جزئی و دقیق جاده‌ها شود. مدل‌های معمولی ممکن است در پردازش و استخراج اطلاعات از چنین تصاویری دچار مشکل شوند.
- شبکه‌های عصبی بازگشتی و مشکلات حافظه:
مدل‌های پیشین برای پردازش تصاویر ترتیبی، به شبکه‌های عصبی بازگشتی وابسته بودند که گاهی مشکلاتی مانند ناپایداری و ناپدید شدن گرادیان‌ها را به همراه داشت. استفاده از شبکه‌های عصبی گرافی این مشکل را حل می‌کند، اما چالش‌های جدیدی مانند مدیریت مقیاس‌پذیری گراف‌ها را مطرح می‌کند.
- تشخیص دقیق نقاط کلیدی:
تشخیص نقاط کلیدی جاده‌ها (مانند تقاطع‌ها) است. در مناطقی با تراکم بالا، ممکن است چندین نقطه کلیدی در یک ناحیه کوچک قرار گیرند که این موضوع باعث کاهش دقت در پیش‌بینی محل تقاطع‌ها می‌شود. در این مقاله، با استفاده از روشی نقاط نزدیک به هم با میانگین‌گیری تجمعی می‌شوند، اما این روش ممکن است دقت کلی تشخیص را کاهش دهد.



شکل ۹.۵: چالشی مهم برای ساخت گراف از داده‌های تصویری

برای آنکه بیشتر با فرآیند تصویر بالا آشنا شوید، در ابتدا چند اصطلاح درباره آنها بیان شده و سپس به توضیح آن پرداخته می‌شود:

- میانگین‌گیری گره:

این فرآیند به این معناست که زمانی که چندین نقطه از داده‌ی تصویری در یک سلول خروجی^{۱۷} قرار می‌گیرند، به جای پردازش تک تک آنها، موقعیت مرکز آنها محاسبه می‌شود.

- شاخه جابجایی^{۱۸}:

شاخه‌ای از شبکه است که مسئول یادگیری موقعیت دقیق این نقاط است و در این حالت برای پیش‌بینی موقعیت میانگین مرکز نقاط به کار می‌رود.

حال همان‌طور که در شکل ۹.۵ مشاهده می‌شود؛ یکی از چالش‌های کار با شبکه‌های عصبی گرافی این است که زمانی که چندین نقطه از داده‌های اصلی در یک سلول خروجی قرار بگیرند، فرآیندی به نام میانگین‌گیری گره^{۱۹} رخ می‌دهد. در این حالت، شاخه جابجایی آموزش می‌بیند تا موقعیت مرکز این نقاط را به جای تک تک آنها پیش‌بینی کند.

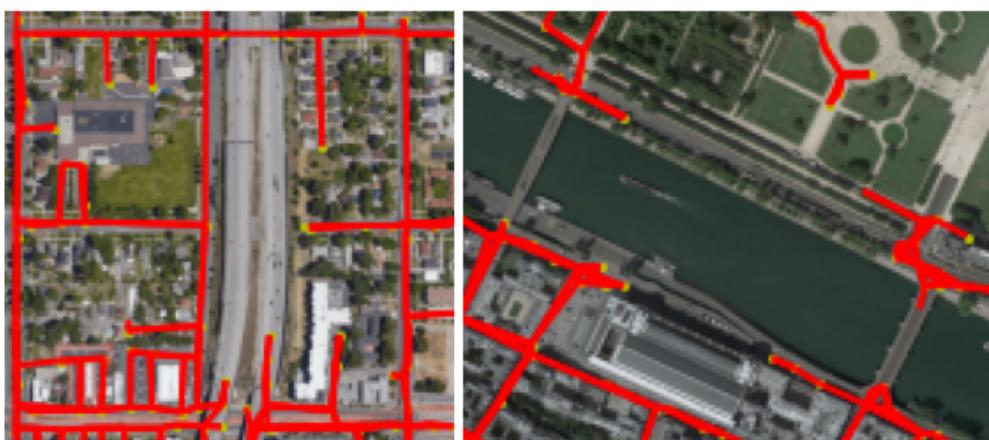
این کار به مدل کمک می‌کند که به جای پردازش تک تک نقاط، تنها مرکز ثقل آنها را در نظر بگیرد و در نتیجه این روش به کاهش پیچیدگی و بهینه‌سازی مدل برای داده‌های پرtraکم کمک می‌کند.

باتوجه به شکل بالا می‌توان گفت که این روش در برابر تشخیص پل یا بزرگراه‌ها ضعف دارد. این در حالی است که روش معروفی شده در ابتدای این فصل مشکلاتی از این قبیل را نداشت. در واقع این دو روش می‌توانند مکمل یک‌دیگر باشند.

^{۱۷}output cell

^{۱۸}offset branch

^{۱۹}Node averaging



شکل ۱۰.۵: نمونه هایی از بخش های طولانی جاده (مانند بزرگراهها، پل ها) که معمولا در این روش از دست می روند.

بنابراین یکی از ایده ها برای کار آینده می تواند ترکیب کردن الگوریتم های شبکه عصبی پیچشی چرخشی با شبکه عصبی گرافی باشد.



شکل ۱۱.۵: ذخیره سازی و سرعت بخشیدن به محاسبات الگوریتم های انجام شده بر روی این مسیرها، مانند الگوریتم های کوتاه ترین مسیر

با توجه به این شکل ۱۱.۵ در روش های گرافی تعداد راس ها کمتر است بنابراین مدل پیچیدگی کمتری خواهد داشت و همچنین برای محاسبه متریک هایی از قبیل کوتاه ترین مسیر، عملکرد بهتری در برابر نمونه های سمت راستی خواهد داشت.

۵.۷.۵ جمع‌بندی و پیشنهادات نهایی

با توجه به ایده‌های مطرح شده، استفاده از الگوریتم Dilated Convolution برای پر کردن نقاط قطع شده جاده‌ها و استفاده از الگوریتم‌های گرافی برای نمایش و بهبود دقت مدل می‌تواند به طور قابل توجهی عملکرد پروژه را بهبود بخشد. علاوه بر این، الگوریتم Skeltonize می‌تواند در استخراج ساختار اصلی جاده‌ها و تبدیل آن‌ها به گراف کمک کند. این ایده‌ها مسیرهای مناسبی برای ادامه کار و تحقیقات آینده در زمینه استخراج جاده‌ها از تصاویر ماهواره‌ای هستند و می‌توانند دقت و کارایی مدل را در شرایط پیچیده و چالش‌برانگیز افزایش دهند.

۸.۵ نتیجه‌گیری نهایی

این پژوهش نشان داد که مدل‌های یادگیری عمیق ابزار قدرتمندی برای استخراج نقشه‌های خیابانی از تصاویر ماهواره‌ای هستند و توانایی بالایی در تشخیص ویژگی‌های پیچیده و محلی تصاویر دارند. همچنین در نهایت با مقدار خوبی از دقت بهوسیله‌ی شبکه عصبی پیچشی چرخشی، استخراج نقشه از تصاویر ماهواره‌ای انجام شد. در مجموع این پژوهش توانسته است گامی موثر در جهت بهبود روش‌های استخراج جاده از تصاویر ماهواره‌ای بردارد و امکانات متنوعی را برای تحقیقات آینده فراهم کند. در حقیقت با استفاده از متريک‌های گرافی می‌توان کارهای با دقت بالایی انجام داد. سرانجام با استفاده از آنچه که در پیشنهادات برای کارهای آینده بيان شد و تجاری که کسب شد می‌توان ایده‌ی استخراج جاده از تصاویر ماهواره‌ای را با شبکه‌های گرافی تلفیق کرد و نتایج جدید به دست آورد.

واژه‌نامه انگلیسی به فارسی

A

Activation Function

تابع فعال‌سازی

Adam

آدم (الگوریتم بهینه‌سازی)

APLS

ارزیابی مسیر پیشنهادی و مسیریابی

Artificial Neural Networks

شبکه‌های عصبی مصنوعی

Average Pooling

تجمیع میانگین

B

BCE (Binary Cross Entropy)

آنتروپی متقابل باینری

Batch Size

اندازه دسته

Bias

بایاس

Bottleneck

گلوگاه

C

CNN (Convolutional Neural Networks)	شبکه‌های عصبی کانولوشنی
CNNSwin	شبکه عصبی پیچشی چرخشی
Channel Convolution	کانولوشن کانالی
Clustering Coefficient	ضریب خوشبندی
Computer Vision	دید رایانه‌ای
Conditional Random Fields (CRF)	میدانی تصادفی شرطی
Conv	کانولوشن
Conformer	هم‌ساز

D

Dense Layers	لایه‌های متراکم
Dilated Block	بلوک متسع
Dilated Convolution	کانولوشن گسترش‌یافته
DeepGlobe	گلوب عمیق

E

Edge	یال
Encoder	رمزگذار

F**F1**

ضریب اف ۱

FeatConn

اتصال تقویت شده با ویژگی

Feature Map

نقشه ویژگی

Filter Size

ابعاد فیلتر

FLOPs

عملیات ممیز شناور

Fully Convolutional Networks (FCN)

شبکه‌های کاملاً کانولوشنی

Fully Connected Layers

لایه‌های تماماً متصل

G**Gaussian Likelihood Maximization**

حداکثرسازی درست‌نمایی گاووسی

GCPs (Ground Control Points)

نقاط کنترل زمین

I**Identity Mapping**

تابع هویت

IoU (Intersection over Union)

همپوشانی بر همگرایی

Intensity

شدت نور

M

Max Pooling

تجمعیع بیشینه

Massachusetts

ماساچوست

MLP (Multi-Layer Perceptron)

پرسپترون چندلایه

MSA (Multi-Head Self-Attention Module)

ماژول خودتوجه چندسری

N

Node

گره

P

Padding

لبه‌گذاری

Patch Partition

بخش وصله

Patch-Based

مبتنی بر پچ

Precision

دقت

Pooling Layers

لایه‌های تجمعیع

PW (Pointwise Convolution)

کانولوشن نقطه‌ای

R

ReLU

واحد خطی اصلاح شده

RGB (Red, Green and Blue)	قرمز، سبز و آبی
RNN (Recurrent Neural Networks)	شبکه‌های عصبی بازگشته
RoadFormer	جاده‌ساز
Residual Connection	اتصال باقی‌مانده
Residual Network (ResNet)	شبکه‌های باقی‌مانده
Recall	بازیابی
Reduction Axioms	اصول موضوعه کاهش
S	
SGD (Stochastic Gradient Descent)	گرادیان نزولی تصادفی
Separable Convolution Module	ماژول کانولوشن قابل جداسازی
Shifted Windows	پنجره‌های جابجا شده
Shortest Path	کوتاه‌ترین مسیر
Sigmoid	سیگموید
Skeltonize	اسکلت‌سازی
Smooth-L1	زیان هموار
SOTA (State-of-the-Art)	روش‌های برتر روز
Spatial Convolution	کانولوشن فضایی
Stack	پشته

Stride

پوشش

Swin Transformer

ترنسفورمر چرخشی

T

tanh

تائزانت هیپربولیک

Transposed Convolution

کانولوشن ترانهاده

V

Vanishing Gradient

نابودی گرادیان

Vision Transformer (ViT)

ترنسفورمر تصویری

W

W-MSA (Window-Based MSA)

خودتوجه مبتنی بر پنجره

Weighted Cross-Entropy Loss

زیان آنتروپی متقاطع وزنی

Weights Vector

بردار وزن‌ها

واژه‌نامه فارسی به انگلیسی

الف

Adam آدم (الگوریتم بهینه‌سازی)

shapConn اتصال تقویت‌شده با شکل

FeatConn اتصال تقویت‌شده با ویژگی

Batch Size اندازه دسته

kernel size اندازه هسته

BCE (Binary Cross Entropy) آتروپی متقابل باینری

reduction axioms اصول موضوعه کاهش

reduction axioms اصول موضوعه کاهش

ب

Bias بایاس

Patch Partition بخش وصله

پ

MLP (Multi-Layer Perceptron)

پرسپترون چندلایه

Stride

پوشش

ت

Activation Function

تابع فعال‌سازی

Identity Mapping

تابع هویت

Dice Coefficient Loss

تابع زیان ضریب دایس

Max Pooling

تجمیع بیشینه

Average Pooling

تجمیع میانگین

Swin Transformer

ترنسفورمر چرخشی

Vision Transformer (ViT)

ترنسفورمر تصویری

Hierarchical Vision Transformer

ترنسفورمر چشم‌انداز سلسله‌مراتبی

ج

RoadFormer

جاده‌ساز

خ

W-MSA (Window-Based MSA)

خودتوجه مبتنی بر پنجره

SW-MSA (Shifted Window MSA)

خودتوجه مبتنی بر پنجره جابجا شده

د

Degree

درجه

Precision

دقت

OA (Overall Accuracy)

دقت کلی

ر

ReLU

واحد خطی اصلاح شده

رمزگذار

Decoder

رمزگشای

ز

Weighted Cross-Entropy Loss

زیان آنتروپی متقاطع وزنی

Smooth-L1

زیان هموار

س

Sigmoid

سیگموید

Skeltonize

اسکلت‌سازی

ش

CNN (Convolutional Neural Networks)

شبکه‌های عصبی پیچشی

Residual Network (ResNet)

شبکه‌های باقی‌مانده

Fully Convolutional Networks (FCN)

شبکه‌های کاملاً کانولوشنی

ص

Shifted Windows

صفحه‌های جابه‌جا شده

ض

Clustering Coefficient

ضریب خوشبندی

F1

ضریب اف یک

ع

FLOPs

عملیات ممیز شناور

ف

Recall

فراخوانی

Channel Convolution

فیلتر کanal

ک

DW (Depthwise Convolution) کانولوشن عمقی

PW (Pointwise Convolution) کانولوشن نقطه‌ای

Dilated Convolution کانولوشن گسترش یافته

Conv کانولوشن

گ

Bottleneck گلوگاه

DeepGlobe گلوب عمیق

Node گره

SGD (Stochastic Gradient Descent) گرادیان نزولی تصادفی

م

Separable Convolution Module مازول کانولوشن قابل جداسازی

MSA (Multi-Head Self-Attention Module) مازول خودتوجه چندسری

Patch-Based مبتنی بر پچ

Receptive Field میدان گیرنده

ن

Feature Map

نقشه ویژگی

Activation Map

نقشه فعال‌سازی

LayerNorm (Layer Normalization)

نرمال‌سازی لایه

و

Weights

وزن‌ها

مراجع

- [1] Alshehhi, R.; Marpu, P.R.; Woon, W.L.; Mura, M.D. Simultaneous Extraction of Roads and Buildings in Remote Sensing Imagery with Convolutional Neural Networks. *ISPRS J. Photogramm. Remote Sens.* 2017, 130, 139–149.
- [2] Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 2017, 39, 2481–2495. [CrossRef] [PubMed]
- [3] Batra, A.; Singh, S.; Pang, G.; Basu, S.; Jawahar, C.; Paluri, M. Improved Road Connectivity by Joint Learning of Orientation and Segmentation. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019.
- [4] Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, Q.; Wang, M. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. *arXiv* 2021, arXiv:2105.05537.
- [5] Chaurasia, A.; Culurciello, E. LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation. In Proceedings of the 2017 IEEE Visual Communications and Image Processing (VCIP), St. Petersburg, FL, USA, 10–13 December 2017; pp. 1–4.
- [6] Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Adeli, E.; Wang, Y.; Lu, L.; Yuille, A.L.; Zhou, Y. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv* 2021, arXiv:2102.04306.
- [7] Christopher M. Bishop, "Pattern Recognition and Machine Learning", 2006.
- [8] "Convolutional Neural Networks for Visual Recognition", (<https://cs231n.github.io/convolutional-networks/>), 2024.
- [9] Cui, F.; Feng, R.; Wang, L.; Wei, L. Joint Superpixel Segmentation and Graph Convolutional Network Road Extraction for High-Resolution Remote Sensing Imagery. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 2178–2181.
- [10] Demir, I.; Koperski, K.; Lindenbaum, D.; Pang, G.; Huang, J.; Basu, S.; Hughes, F.; Tuia, D.; Raskar, R. DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern

Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 172–17209.

- [11] Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image Is Worth Words: Transformers for Image Recognition at Scale. arXiv 2020, arXiv:2010.11929.
- [12] Dosovitskiy, Alexey, et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." *International Conference on Learning Representations*, 2021.
- [13] Duda, Richard O., and Peter E. Hart. *Pattern Classification and Scene Analysis*. Wiley, 2001.
- [14] Gonzalez, Rafael C., and Richard E. Woods. *Digital Image Processing*. Prentice Hall, 2008.
- [15] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [16] Gulati, A.; Qin, J.; Chiu, C.; Parmar, N.; Zhang, Y.; Yu, J.; Han, W.; Wang, S.; Zhang, Z.; Wu, Y.; et al. Conformer: Convolutionaugmented Transformer for Speech Recognition. In Proceedings of the Interspeech 2020, 21st Annual Conference of the International Speech Communication Association, Shanghai, China, 25–29 October 2020; pp. 5036–5040.
- [17] Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer, 2009.
- [18] He, Y.; Wang, H.; Zhang, B. Color based road detection in urban traffic scenes. In Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems, Shanghai, China, 12–15 October 2003; Volume 1, pp. 730–735.
- [19] Jain, Anil K. *Fundamentals of Digital Image Processing*. Prentice Hall, 1989.
- [20] Jensen, John R. *Introductory Digital Image Processing: A Remote Sensing Perspective*. Pearson, 2015.
- [21] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully Convolutional Networks for Semantic Segmentation", 2015 CVPR (Conference on Computer Vision and Pattern Recognition).
- [22] Kestur, R.; Farooq, S.; Abdal, R.; Qadri, E.; Narasipura, O.; Mudigere, M. UFCN: A Fully Convolutional Neural Network for Road Extraction in RGB Imagery Acquired by Remote Sensing from an Unmanned Aerial Vehicle. *J. Appl. Remote Sens.* 2018, 12, 016020.
- [23] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Communications of the ACM*, 2012.
- [24] Laben, Craig A., and Bernard V. Brower. "Process for Enhancing the Spatial Resolution of Multispectral Imagery Using Pan-Sharpening." U.S. Patent No. 6,011,875. 2000.

- [25] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, Alan L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs", 2018 IEEE Transactions on Pattern Analysis and Machine Intelligence.
- [26] Liu X, Wang Z, Wan J, Zhang, J, et al. "RoadFormer: Road Extraction Using a Swin Transformer Combined with a Spatial and Channel Separable Convolution". *Remote Sens*, 2023.
- [27] Liu X, Ze, et al. "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows." *International Conference on Computer Vision*, 2021.
- [28] Luo, L.; Wang, J.X.; Chen, S.B.; Tang, J.; Luo, B. BDTNet: Road Extraction by Bi-Direction Transformer From Remote Sensing Images. *IEEE Geosci. Remote Sens. Lett.* 2022, 19, 2505605.
- [29] Mitchell, Tom M. *Machine Learning*. McGraw-Hill, 1997.
- [30] Mnih, V.; Hinton, G.E. Learning to Label Aerial Images from Noisy Data. In Proceedings of the the 29th International Conference on Machine Learning, ICML 2012, Edinburgh, UK, 26 June–1 July 2012.
- [31] Mohammadkhani Maryam, Jampour Mahyar, Zandkarimi Shahin, Samavi Shadrokh, "Recurrent Convolutional Neural Networks for Road Extraction from Aerial Imagery", 2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS).
- [32] Newell, A.; Yang, K.; Deng, J. Stacked Hourglass Networks for Human Pose Estimation. In Proceedings of the Computer Vision—ECCV 2016, Amsterdam, The Netherlands, 11–14 October 2016; Springer International Publishing: Cham, Switzerland, 2016; pp. 483–499.
- [33] Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention U-Net: Learning Where to Look for the Pancreas. arXiv 2018, arXiv:1804.03999.
- [34] Olaf Ronneberger, Philipp Fischer, Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", 2015, Medical Image Computing and Computer-Assisted Intervention (MICCAI) conference.
- [35] Park, N.; Kim, S. How Do Vision Transformers Work? arXiv 2022, arXiv:2202.06709.
- [36] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing", 2nd ed, kolxo3, 2002.
- [37] Rafael Pires de Lima, kurt Marfurt, "Convolutional Neural Networks for Remote Sensing Image Classification", 2020 IEEE Xplore Remote Sensing.
- [38] Richards, John A. *Remote Sensing Digital Image Analysis*. Springer, 2013.
- [39] Schowengerdt, Robert A. *Remote Sensing: Models and Methods for Image Processing*. Academic Press, 2006.

- [40] Song, M.; Civco, D.L. Road Extraction Using SVM and Image Segmentation. Photogramm. Eng. Remote Sens. 2004, 70, 1365–1371.
- [41] Tao Chen, Yiran Liu, Haoyu Jiang, Ruirui Li. "Swin Transformer coupling CNNs Makes Strong Contextual Encoders for VHR Image Road Extraction" 2023.
- [42] Tao, J.; Chen, Z.; Sun, Z.; Guo, H.; Leng, B.; Yu, Z.; Wang, Y.; He, Z.; Lei, X.; Yang, J. Seg-Road: A Segmentation Network for Road Extraction Based on Transformer and CNN with Connectivity Structures. Remote Sens. 2023, 15, 1602.
- [43] Thomas Lillesand, Ralph W. Kiefer, Jonathan Chipman, "Remote Sensing and Image Interpretation", 7th ed, 2014.
- [44] Gaetan Bahl, Mehdi Bahri, Florent Lafarge: "Road Extraction from Overhead Images with Graph Neural Networks", arXiv:2112.05215 (2021).
- [45] van der Maaten, Laurens, and Geoffrey Hinton. "Visualizing data using t-SNE." *Journal of Machine Learning Research*, 2008.
- [46] Varia, N.; Dokania, A.; Senthilnath, J. DeepExt: A Convolution Neural Network for Road Extraction Using RGB Images Captured by UAV. In Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence (SSCI), Bengaluru, India, 18–21 November 2018; pp. 1890–1895.
- [47] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. Adv. Neural Inf. Process. Syst. 2017, 30, 5998–6008.
- [48] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation", 2017 IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI).
- [49] Wegner, J.D.; Montoya-Zegarra, J.A.; Schindler, K. A Higher-Order CRF Model for Road Network Extraction. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; IEEE Computer Society: Washington, DC, USA, 2013; pp. 1698–1705.
- [50] Yoshua Bengio, Aaro ncourville, Ian Goodfellow, "Deep Learning Book", (<http://www.deeplearningbook.org/contents/convnets.html>), 2016.
- [51] Zhang, Q.; Couloigner, I. Benefit of the angular texture signature for the separation of parking lots and roads on high resolution multi-spectral imagery. Pattern Recognit. Lett. 2006, 27, 937–946.
- [52] Zhang, Z.; Liu, Q.; Wang, Y. Road Extraction by Deep Residual U-Net. IEEE Geosci. Remote Sens. Lett. 2018, 15, 749–753.
- [53] Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239.

- [54] Zhou, L.; Zhang, C.; Wu, M. D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 192–1924.

Surname: Abbas pour

Name: Mohammad javad

Title: Street map extraction system using image processing on satellite view

Supervisor: Dr. Hossein Hajiabolhassan Advisor: Dr. Saeid Reza Kheradpishe

Degree: Master of Science

Subject: Data Science Field

Field: Data Science Field

Shahid Beheshti University

Faculty of Mathematical Sciences

Date: 2024

Number of pages: 113

Keywords: Convolutional neural network, Image segmentation, Remote sensing image, Image processing, Road extraction, Swin transformer

Abstract

This thesis explores various methods for extracting road maps from satellite images using convolutional neural networks (CNNs) and other deep learning techniques.

Deep models, such as convolutional neural networks and rotational transformers, were employed to accurately extract roads from images with low resolution and contrast.

One of the main challenges in this field is the low accuracy in images with low contrast, which traditional mathematical methods, like morphology, cannot resolve. Therefore, deep learning models, such as rotational transformers, which can preserve image details, are proposed as a solution.

The combination of convolutional neural networks and rotational transformers has shown significant improvements in accuracy and road feature detection.

These methods can be applied in various areas, such as urban mapping, automatic routing, environmental change monitoring, autonomous driving systems, and by rescue organizations.



Shahid Beheshti University
Faculty of Mathematical Sciences
Data Science Field

M.Sc. Thesis

Street map extraction system using image processing on satellite view

by
Mohammad javad Abbas pour

Supervisor
Dr. Hossein Hajiabolhassan
Advisor
Dr. Saeid Reza Kheradpishe

2024