

بهبود بخش‌بندی معنایی تصاویر شهری با ارائه یک تابع هزینه جدید مبتنی بر ابرپیکسل

زهره غنائی

گروه مهندسی کامپیوتر، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران
پست الکترونیکی: zahra.ghanaei@mail.um.ac.ir

مجتبی روحانی*

گروه مهندسی کامپیوتر، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران
پست الکترونیکی: rouhani@um.ac.ir

چکیده

بخش‌بندی معنایی^۱ تصاویر یکی از مؤلفه‌های کلیدی در تحلیل تصاویر شهری است که عملکرد مناسب آن تأثیر مستقیمی بر کارایی و سرعت سیستم‌های بینایی ماشین در این حوزه دارد. بر همین اساس، در این پژوهش یک تابع هزینه جدید برای بهبود بخش‌بندی معنایی در تصاویر شهری ارائه شده است. یکی از چالش‌های اصلی در این زمینه، عدم توازن^۲ داده‌ها است که با توجه به تفاوت نرخ عدم توازن در هر کلاس در مجموعه داده‌های چند کلاسه، انتخاب تابع هزینه مناسب را نسبت به مسائل دو کلاسه دشوارتر می‌سازد. علاوه بر این، مشکلاتی مانند شناسایی دقیق مرز اجسام و حفظ پیوستگی نواحی نیز باعث کاهش دقت روش‌های موجود می‌شوند. بنابراین، راهکار پیشنهادی با شناسایی نواحی پیچیده و پرتراکم در تصویر، که می‌تواند شامل کلاس‌هایی با شباهت ظاهری زیاد، مرز بین چند

کلاس، یا کلاس‌ها و اجسام کوچک باشد، و تمرکز بیشتر بر آنها در زمان آموزش شبکه به بهبود عملکرد مدل کمک می‌کند. در این روش، ابتدا هر تصویر توسط الگوریتم SLIC^۳ به ابرپیکسل‌های^۴ مجزا تقسیم شده و تأثیر هر ناحیه با تحلیل ابرپیکسل‌ها و اعمال وزن‌دهی هدفمند در تابع هزینه پیشنهادی تعیین می‌شود. در ادامه برای ارزیابی اثربخشی این روش، مجموعه‌ای از آزمایش‌ها با استفاده از مدل U-Net و دو مجموعه داده از تصاویر شهری انجام شده است. نتایج این آزمایش‌ها نشان می‌دهد که استفاده از تابع هزینه پیشنهادی در مقایسه با سایر روش‌ها به طور میانگین موجب بیش از ۱ درصد بهبود در معیارهای IoU و امتیاز F1 شده و تمرکز بر نواحی پیچیده تصاویر می‌تواند گامی موثر در رفع محدودیت‌های موجود در بخش‌بندی معنایی باشد. واژه‌های کلیدی: تصاویر شهری، بخش‌بندی معنایی، تابع هزینه، ابرپیکسل، شبکه‌های عصبی عمیق.

3-Simple Linear Iterative Clustering
4-Superpixels

1-Semantic Segmentation
2-Imbalance

* نویسنده مسئول

۱- مقدمه

درک ساختاری محتوای تصاویر دیجیتال، از جمله مهم‌ترین چالش‌ها در حوزه تحلیل تصاویر شهری به شمار می‌رود و همواره مورد توجه پژوهشگران بوده است. یکی از راهکارهای کلیدی برای این منظور، بخش‌بندی معنایی تصاویر است؛ فرایندی که در آن، هر تصویر به نواحی دارای معنا افزای می‌شود تا سیستم‌های بینایی ماشین بتوانند درک عمیق‌تری از صحنه‌های شهری به‌دست آورند. در این زمینه، مدل‌های بخش‌بندی برای شناسایی دقیق اجزای مختلف صحنه‌های شهری از جمله پیاده‌روها، مسیرهای دوچرخه، خودروها، علائم راهنمایی و رانندگی، ساختمان‌ها و فضای سبز به کار گرفته می‌شوند. دقت بالای این فرایند برای کاربردهایی همچون سامانه‌های هوشمند کمک‌راننده، خودروهای خودران، مدیریت هوشمند ترافیک و پایش زیست‌محیطی در مناطق شهری حیاتی است. هرگونه ضعف در کیفیت بخش‌بندی می‌تواند منجر به افزایش خطا در تصمیم‌گیری این سامانه‌ها و در نهایت کاهش ایمنی و کارایی آن‌ها شود. این فناوری علاوه بر تحلیل تصاویر شهری، در کاربردهای متعدد دیگری از جمله تحلیل تصاویر پزشکی، پردازش تصاویر ماهواره‌ای برای شناسایی و پایش عوارض طبیعی و انسانی، تحلیل داده‌های سنجش از دور و سیستم‌های نظارت تصویری نقش‌آفرینی می‌کند [۱]. با این وجود، تمرکز اصلی این پژوهش بر بهبود عملکرد مدل‌ها در تحلیل دقیق‌تر تصاویر شهری و ارتقاء سیستم‌های مرتبط با آن است.

با توجه به کاربرد فراوان این مسئله، روش‌های زیادی به صورت بدون راهنما^۵ و با راهنما^۶ برای حل آن ارائه شده است. از جمله این روش‌ها می‌توان به آستانه‌گذاری^۷ [۲]، خوشه‌بندی^۸ و بخش‌بندی بر مبنای

هیستوگرام^۹ [۳]، رشد ناحیه^{۱۰} [۴]، روش آب‌پخش‌شان^{۱۱} [۵] و همچنین روش‌های پیشرفته‌تر مانند برش گراف^{۱۲} [۶]، مدل کانتورهای فعال^{۱۳} [۷]، و میدان تصادفی مارکوف^{۱۴} و شرطی^{۱۵} [۸] اشاره کرد. با ظهور مدل‌های یادگیری عمیق، نسل جدیدی از روش‌های بخش‌بندی تصاویر بر مبنای شبکه‌های عصبی هم‌آمیختگی^{۱۶} شکل گرفت که بهبود قابل توجهی در عملکرد بخش‌بندی تصاویر ایجاد کرد. برخی از این مدل‌ها که به طور ویژه برای بخش‌بندی تصاویر طراحی شده‌اند، با استفاده از ساختار رمزگذار-رمزگشا^{۱۷}، به صورت خودکار و سلسله‌مراتبی، ویژگی‌های معنادار را مستقیماً از تصویر استخراج می‌کنند. این مدل‌ها با وجود عملکرد مناسب در بسیاری از کاربردها، در مجموعه داده‌هایی با توزیع نامتوازن در شناسایی کلاس‌های اقلیت با چالش مواجه می‌شوند. یکی از رویکردهای مطرح برای مقابله با این مشکل، در کنار تحقیقاتی که تا کنون برای طراحی مدل‌هایی با ساختار کارآمدتر انجام شده است، ارائه یک تابع هزینه مناسب است. استفاده از یک تابع هزینه بهینه، بر خلاف توسعه شبکه‌های پیچیده‌تر، مانند شبکه‌های دارای عمق بیشتر، مازول‌های توجه، یا معماری‌های چندشاخه که معمولاً به منابع محاسباتی بالا و تنظیمات پیچیده نیاز دارند، می‌تواند بدون تغییرات عمده در معماری‌های موجود، کارایی آن‌ها را بهبود بخشد. زیرا تابع هزینه نقش کلیدی در پسانتشار خطا و هدایت فرایند یادگیری دارد و تأثیر قابل توجهی بر نتایج به‌دست آمده خواهد داشت. از این رو، طراحی یک تابع هزینه که بتواند فرایند یادگیری را در مواجهه با داده‌های نامتوازن مدیریت کند، اهمیت ویژه‌ای دارد.

این پژوهش با هدف ارائه یک تابع هزینه نوآورانه

9- Histogram-based Bundling
10-Region Growing
11-Watershed Method
12-Graph Cuts
13-Active Contours Model
14-Markov Random Field
15-Conditional Random Fields
16-Convolutional Neural Network
17-Encoder-Decoder

5- Unsupervised Methods
6- Supervised Methods
7- Thresholding
8- Clustering

تابع آنتروپی متقاطع^{۱۸}، که از معیار واگرایی کولبک-لیبلر^{۱۹} الهام گرفته شده است، رایج‌ترین تابع هزینه مبتنی بر توزیع است. این تابع میانگین خطای آنتروپی متقاطع را برای تمام پیکسل‌های تصویر و مقادیر پیش‌بینی شده متناظر محاسبه می‌کند. در نتیجه تمام پیکسل‌ها به طور یکسان در محاسبه میزان خطا نقش دارند. این روش در داده‌های متوازن عملکرد خوبی دارد ولی در مجموعه داده‌های نامتوازن، که در آن کلاس اکثریت بر کلاس اقلیت غالب است، دچار مشکل می‌شود. برای حل این چالش، نسخه اصلاح شده‌ای از این تابع، به نام تابع آنتروپی متقاطع وزن‌دار^{۲۰} [۱۰] ارائه شده است که با وزن‌دهی به کلاس‌ها تأثیر نمونه‌های دشوار را افزایش می‌دهد. علاوه بر این، نوع دیگری از تابع آنتروپی متقاطع به نام TopK در مقاله [۱۱] ارائه شده است که در هر گام آموزش شبکه، تنها از $k\%$ از نمونه‌هایی که بیشترین خطا را دارند، برای محاسبه تابع هزینه استفاده می‌کند. همچنین، تابع هزینه کانونی^{۲۱} [۱۲] نیز بر مبنای تابع آنتروپی متقاطع برای حل مسائلی با عدم توازن بالا طراحی شده است. این تابع با افزودن یک پارامتر جدید، مقدار هزینه را برای نمونه‌هایی که به درستی طبقه‌بندی شده‌اند کاهش داده و توجه بیشتری به نمونه‌های دشوار می‌دهد.

بررسی مقالات منتشر شده بین سال‌های ۲۰۲۰ تا ۲۰۲۵ نشان می‌دهد که تابع آنتروپی متقاطع همچنان پرکاربردترین گزینه در حوزه تحلیل تصاویر شهری است و در بسیاری از پژوهش‌ها، با وجود توسعه ساختارهای جدید، این تابع به دلیل سادگی، پایداری و کارایی مناسب در مسائل چندکلاسه، همچنان به‌عنوان یک گزینه رایج و قابل اعتماد مورد استفاده قرار می‌گیرد [۱۳ و ۱۵]. با این حال، در کنار استفاده گسترده از تابع آنتروپی متقاطع استاندارد، برخی از پژوهش‌ها به سمت استفاده از ترکیب‌های پیچیده‌تر و هدفمندتری از توابع هزینه

برای بهبود عملکرد شبکه‌های عصبی عمیق در مسائل بخش‌بندی تصاویر شهری انجام شده است. تابع هزینه پیشنهادی با شناسایی نواحی پیچیده در تصویر و افزایش وزن آنها در فرایند محاسبه خطا، تلاش می‌کند دقت مدل را در شناسایی اجسام کوچک و مرزهای دشوار افزایش دهد. تمرکز بر بخش‌های دشوار تصویر که احتمالاً شامل کلاس‌های اقلیت هستند، به مدل کمک می‌کند تا دقت بهتری در بخش‌بندی معنایی به‌دست آورد و عملکرد بهتری در تشخیص کلاس‌های کم تکرار داشته باشد.

ساختار مقاله در ادامه به شرح زیر است: در بخش ۲ مروری بر تحقیقات اخیر در زمینه ارائه توابع هزینه برای بخش‌بندی تصاویر انجام شده است. در بخش‌های ۳ و ۴ به ترتیب شرح جزئیات روش پیشنهادی و ارزیابی آن ارائه شده و در نهایت بخش ۵ به جمع‌بندی و نتیجه‌گیری این پژوهش اختصاص دارد.

۲- کارهای مرتبط

مسئله عدم توازن کلاس‌ها در سال‌های اخیر، به دلیل تأثیر قابل‌توجه آن بر دقت و قابلیت اطمینان شبکه‌های عصبی عمیق، مورد توجه ویژه قرار گرفته است. در نتیجه، رویکردهای نوآورانه متعددی برای کاهش اثرات منفی این مشکل و بهبود عملکرد مدل‌های عمیق ارائه شده‌اند. در این راستا، بسیاری از پژوهشگران بر توسعه توابع هزینه مؤثر برای شبکه‌های عصبی عمیق، به‌ویژه با هدف رفع چالش‌های ناشی از عدم توازن کلاس‌ها در بخش‌بندی تصاویر تمرکز کرده‌اند. در این بخش، انواع توابع هزینه که در بخش‌بندی تصاویر با روش‌های یادگیری عمیق استفاده می‌شوند، معرفی شده‌اند. این توابع را می‌توان به چهار دسته اصلی تقسیم کرد که هر یک عدم تطابق بین داده‌های ورودی و پیش‌بینی شده را بر اساس توزیع داده‌ها، میزان همپوشانی نواحی، مرزها و ترکیبی از این موارد کمینه می‌کنند [۹].

18- Cross Entropy (CE)

19- Kullback-Leibler Divergence

20- Weighted Cross Entropy (WCE)

21- Focal Loss

شده‌اند. در ادامه، تابع دایس تعمیم یافته^{۲۹} برای مسائل چندکلاسه معرفی شده است که با وزن‌دهی معکوس نسبت به فراوانی کلاس‌ها، سهم کلاس‌های نادر را در فرایند یادگیری افزایش می‌دهد [۲۳]. تابع تورسکی^{۳۰} نیز با افزودن دو ضریب قابل تنظیم به معادله دایس، امکان کنترل تعادل میان دقت^{۳۱} و یادآوری^{۳۲} را فراهم می‌سازد [۲۴]. این تابع در واقع تعمیمی از توابع دایس و ژاکارد به شمار می‌رود، به گونه‌ای که با انتخاب ضرایب مناسب، می‌توان هر یک از این دو تابع را بازسازی کرد. همچنین، مقاله [۲۵] با ترکیب مفهوم تابع هزینه قانونی با تابع تورسکی، تابعی با نام تورسکی قانونی^{۳۳} ارائه نموده که در آن یادگیری بر روی داده‌های دشوار متمرکز شده است.

به طور مشابه توابع هزینه مبتنی بر مرز، با هدف کاهش فاصله بین مرزهای بخش‌بندی پیش‌بینی شده و مرزهای صحیح معرفی شده‌اند. نخستین نمونه از این نوع توابع، تابع هزینه مرزی^{۳۴} [۲۶] است که به جای نواحی، از انتگرال‌گیری روی مرزها بهره می‌برد تا روشی مشتق‌پذیر برای اندازه‌گیری فاصله بین دو مرز فراهم کند. رویکردی دیگر نیز با استفاده از معیار فاصله هاوسدورف^{۳۵}، که شاخصی شناخته‌شده برای ارزیابی دقت مرزی در بخش‌بندی تصویر به شمار می‌رود، توسعه یافته است [۲۷]. شایان ذکر است که هر دو تابع یاد شده در زمان آموزش شبکه باید با یک تابع هزینه مبتنی بر ناحیه مانند دایس همراه شوند تا پایداری آموزش شبکه حفظ شود [۹].

توابع هزینه ترکیبی نیز، همان طور که از نام آنها پیداست، ترکیبی از توابع ارائه شده در بخش‌های قبل می‌باشند. در مقاله [۲۸]، با اعمال توابع تبدیل نمایی و لگاریتمی بر روی توابع آنتروپی متقاطع و دایس، تابع

حرکت کرده‌اند. به عنوان مثال، در مقاله [۱۶]، به منظور بهبود عملکرد روی پیکسل‌های دشوار یا کلاس‌های نادر، نسخه‌ای از آنتروپی متقاطع همراه با روش استخراج برخط نمونه‌های سخت^{۳۲} به کار گرفته شده است. در این روش توجه مدل بیشتر بر نمونه‌هایی که پیش‌بینی آنها دشوار است و احتمال خطای بالایی دارند متمرکز می‌شود. علاوه بر این، استفاده از خروجی‌های کمکی^{۳۳} برای بهینه‌سازی بهتر لایه‌های میانی شبکه نیز به کار گرفته شده است. همچنین، در مقاله [۱۷] تابع هزینه به صورت ترکیبی از چند تابع آنتروپی متقاطع مجزا در سطوح مقیاسی مختلف تعریف شده است که با وزن‌دهی متفاوت به مسیرهای شبکه، امکان یادگیری چندمقیاسی و افزایش دقت نهایی مدل را فراهم می‌آورد. در مدل‌های پیشرفته‌تری مانند TV3S [۱۸] که برای بخش‌بندی ویدئو طراحی شده است، تابع هزینه آنتروپی متقاطع نه تنها دقت مکانی، بلکه انسجام زمانی بین قاب‌های ویدیویی^{۳۴} را نیز بهینه می‌کند. در این رویکرد دوگانه، فرایند بهینه‌سازی تنها به خروجی نهایی محدود نمی‌شود، بلکه تلاش شده است نمایش‌های میانی ویژگی‌ها در طول پردازش زمانی نیز دقیق و منسجم باقی بمانند. از سوی دیگر، مقاله [۱۹] با معرفی تابع هزینه Joint Edge Seg Loss ترکیبی از مؤلفه‌های مرتبط با اشیاء و نواحی مرزی آنها را لحاظ کرده که موجب بهبود قابل توجهی در شناسایی دقیق مرزها و اشیای کوچک در مناظر شهری می‌شود.

توابع هزینه دایس^{۳۵} [۲۰، ۲۱] و ژاکارد^{۳۶} [۲۲] نمونه‌هایی رایج از توابع هزینه مبتنی بر ناحیه هستند و نقش مؤثری در مسائل دو و چندکلاسه ایفا می‌کنند. این توابع با هدف بیشینه‌سازی میزان هم‌پوشانی میان نواحی صحیح و پیش‌بینی شده، و به ترتیب با الهام از معیارهای شباهت دایس^{۳۷} و نسبت اشتراک به اجتماع^{۳۸} تعریف

29- Generalized Dice

30-Tversky

31-Precision

32-Recall

33-Focal Tversky

34-Boundary (BD)

35-Hausdorff Distance

22- Online Hard Example Mining (OHEM)

23- Auxiliary Head

24- Video Frames

25- Dice

26- Jaccard

27- Dice Similarity Coefficient (DSC)

28- Intersection over Union (IoU)

هم‌آمیختگی شبکه به کار گرفته می‌شود، توابع هزینه نیز ظرفیت بهره‌گیری از این اطلاعات و نیز اطلاعات سطح بالاتر، مانند تعداد نواحی، ابعاد و ساختار آن‌ها را دارند. با بهره‌گیری از این ویژگی‌ها و شناسایی نواحی کلیدی و راهبردی در تصویر و تمرکز بیشتر بر آن‌ها، می‌توان نتایج دقیق‌تری در بخش‌بندی تصاویر به‌دست آورد. چالشی که در این بخش وجود دارد این است که این اطلاعات معمولاً گسسته هستند و استفاده مستقیم از آنها در تابع هزینه می‌تواند به ناپایداری در فرایند آموزش شبکه منجر شود. بنابراین لازم است این اطلاعات در قالب یک معیار پیوسته و مشتق‌پذیر در محاسبه هزینه در نظر گرفته شوند تا امکان به‌کارگیری آن‌ها در فرایند پس‌انتشار فراهم شود [۳۴].

بنا بر آنچه گفته شد، راهکار پیشنهادی بر استخراج و تحلیل اطلاعات سطح بالا و شناسایی نواحی پیچیده و پراهمیت در تصویر تمرکز دارد. سپس با استفاده از این اطلاعات و با طراحی یک الگوی وزن‌دهی برای نواحی مختلف تصویر، میزان تأثیر هر ناحیه در محاسبه جریمه نهایی کنترل می‌شود. در این چارچوب، می‌توان از تمام توابع هزینه متعارف مانند تابع آنتروپی متقاطع که مشتق‌پذیر و پایدار هستند، استفاده کرد. برای تعیین میزان اهمیت هر ناحیه معیارهای مختلفی تعریف می‌شود. به‌عنوان نمونه، با بررسی و تحلیل اسکلتهای موجود در تصویر، یا با استفاده از ارتباط معنایی و قوانین همسایگی میان کلاس‌ها، می‌توان بخش‌هایی از تصویر را که پیچیده‌تر بوده و احتمال خطا در آن بیشتر است، شناسایی کرده و وزن آنها را در تابع هزینه، و در نتیجه اهمیت آن نواحی را افزایش داد. با این حال، این گونه معیارها اغلب به طور خاص برای یک مسئله تعریف شده و قابلیت تعمیم به مجموعه داده‌های دیگر را ندارند. از این رو، با توجه به هدف این پژوهش که ارائه یک تابع هزینه برای مجموعه دادگان متنوع در حوزه تصاویر شهری است، معیار دیگری بر مبنای میزان تراکم کلاس‌ها در

هزینه جدیدی تحت عنوان تابع لگاریتمی نمایی^{۳۶} معرفی شده است. علاوه بر این، در سال‌های اخیر توابع ترکیبی دیگری نیز ارائه شده‌اند که به صورت مجموع تابع هزینه دایس و یکی از توابع مبتنی بر توزیع تعریف می‌شوند، از جمله تابع آنتروپی متقاطع وزن‌دار [۲۹ و ۳۰]، تابع هزینه کانونی [۳۱]، تابع TopK [۳۲] و تابع چندجمله‌ای [۳۳] که در کاربردهای مختلف مورد استفاده قرار گرفته‌اند.

در مجموع، روند طراحی توابع هزینه در تحلیل تصاویر شهری نشان می‌دهد که استفاده از ساختارهای ترکیبی، تمرکز بر نمونه‌های دشوار، توجه به انسجام زمانی و مدل‌سازی دقیق مرزها به محورهای اصلی بسیاری از پژوهش‌های اخیر تبدیل شده‌اند. این روند، نشان‌دهنده درک عمیق‌تر نیازها و چالش‌های موجود در تحلیل مناظر شهری و طراحی دقیق‌تر توابع هزینه برای مواجهه با این پیچیدگی‌هاست.

۳- روش پیشنهادی

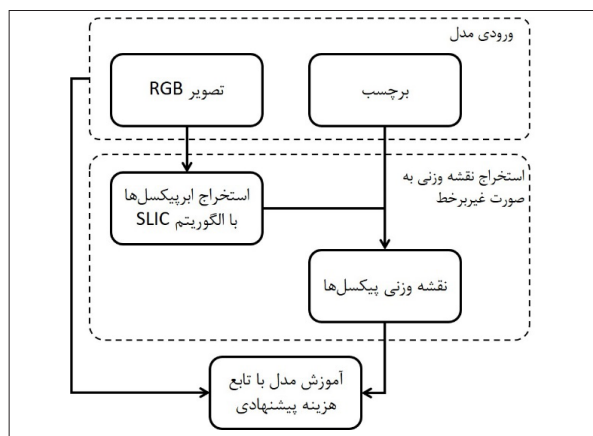
در مسئله بخش‌بندی تصاویر، هر نقطه از تصویر دارای اهمیتی ویژه است و میزان ارزش و اطلاعات هر پیکسل از یک شیء بر اساس ویژگی‌های کلاسی که به آن تعلق دارد، ابعاد آن شیء، و محیط پیرامون آن بستگی دارد. توابع هزینه‌ای که تا کنون ارائه شده‌اند با پیکسل‌های تصویر به صورت جداگانه برخورد کرده و نه تنها اهمیت نسبی آنها، بلکه ارتباط میان پیکسل‌ها را نیز نادیده می‌گیرند. به‌عنوان مثال در تابع هزینه آنتروپی متقاطع، خطای هر پیکسل به طور مستقل محاسبه شده و میانگین این خطاها به‌عنوان هزینه نهایی در نظر گرفته می‌شود. به عبارت دیگر در این تابع، تمامی پیکسل‌های تصویر اهمیت یکسانی دارند. در ادامه تابع هزینه کانونی با هدف افزایش توجه به داده‌هایی که به درستی طبقه‌بندی نشده‌اند، ارائه شده است. با این وجود این تابع نیز از اطلاعات ساختاری یا ارتباطی بین پیکسل‌ها صرف نظر می‌کند. اگرچه اطلاعات مکانی تصویر در لایه‌های

36-Exponential Logarithmic Loss (ELL)

کلاسی و درون کلاسی است. در این روش، بخش‌هایی از تصویر که شامل نمونه‌هایی از کلاس‌های اقلیت مانند ماشین یا انسان هستند، یا شامل اشیاء کوچکتر از کلاس‌های اکثریت مانند درخت‌های تکی یا فضای سبز محدود می‌باشند، وزن بیشتری دریافت می‌کنند. این امر می‌تواند به بهبود عملکرد مدل در مجموعه داده‌هایی با عدم توازن بین کلاسی و درون کلاسی کمک کند. همچنین ابرپیکسل‌هایی که شامل مرز بین دو یا چند کلاس هستند نیز وزن بیشتری خواهند گرفت. این نواحی می‌توانند تنها بخش کوچکی از یک شیء بزرگ (مانند لبه ساختمان در مجاورت پس‌زمینه) را در برگیرند. در نتیجه، نیازی به اختصاص وزن بالا به کل آن شیء یا کلاس نخواهد بود و تنها بخشی از آن که در ناحیه پیچیده قرار دارد مورد توجه قرار خواهد گرفت. علاوه بر این، برخی نواحی از تصویر به علت شباهت ظاهری اشیاء موجود در آن با پس‌زمینه یا سایر اجسام مجاور، فارغ از ابعاد یا نوع کلاس، در دسته نواحی پیچیده قرار می‌گیرند. این موضوع به علت دید محلی روش پیشنهادی و در نظر گرفتن شرایط آن ناحیه و موقعیت مکانی اشیاء رخ می‌دهد. نمونه‌های دیگری از مواردی که الگوی وزندهی پیشنهادی عملکرد مناسبی در آن از خود نشان می‌دهد، شامل موقعیت‌هایی است که یک نمونه از کلاس اقلیت در پس‌زمینه‌ای ساده قرار گرفته باشد، یا بالعکس. همچنین در شرایطی که نمونه‌هایی از یک کلاس، بسته به موقعیت و زمینه پیرامونی خود، به طور هم‌زمان شامل موارد سخت و آسان باشند، استفاده از وزندهی تطبیقی می‌تواند بسیار مؤثر واقع شود.

نمادهای استفاده شده در پیاده‌سازی روش پیشنهادی به شرح زیر است: فرض کنید I یک تصویر رنگی دو بعدی و G و P به ترتیب نقشه برچسب واقعی و نقشه پیش‌بینی شده مربوط برای آن باشند. بنابراین برای پیکسل i -ام از این تصویر، $p_i^f \in [0, 1]$ احتمال پیش‌بینی شده تعلق آن پیکسل به کلاس c و $g_i^f \in \{0, 1\}$ یک نشانگر دودویی

هر ناحیه به عنوان ایده اصلی پیشنهاد شده است. بر این اساس، می‌توان بدون در نظر گرفتن ماهیت تصاویر و نوع اشیاء موجود در آن، تخمینی از میزان پیچیدگی نواحی مختلف به دست آورده و از آن در طراحی تابع هزینه استفاده نمود. نمای کلی روش پیشنهادی در شکل ۱ نمایش داده شده است.



شکل ۱: نمای کلی روش پیشنهادی. گام اول: آماده‌سازی ورودی‌ها. گام دوم: استخراج ابرپیکسل‌ها از تصاویر RGB و محاسبه نقشه وزنی با شمارش تعداد کلاس‌های موجود در هر ابرپیکسل. گام سوم: آموزش مدل با استفاده از تابع هزینه پیشنهادی مبتنی بر نقشه وزنی پیکسل‌ها.

برای این منظور، ابتدا هر تصویر با استفاده از الگوریتم SLIC [۳۵]، به عنوان یکی از روش‌های کلاسیک و بدون راهنما برای بخش‌بندی تصاویر، به نواحی کوچکتری به نام ابرپیکسل تقسیم می‌شود. از آنجا که این نواحی دارای ویژگی‌های ظاهری مشابهی هستند، در حالت ایده‌آل انتظار می‌رود متعلق به کلاس یکسانی باشند. بنابراین، حضور بیش از یک کلاس در یک ابرپیکسل می‌تواند نشانه‌ای از پیچیدگی و دشواری بیشتر آن ناحیه برای شبکه باشد. بر این اساس، وزن پیکسل‌های آن ناحیه در تابع هزینه متناسب با تعداد کلاس‌های موجود در آن افزایش می‌یابد. علت این پیچیدگی ممکن است مجاورت کلاس‌هایی با ویژگی‌های ظاهری مشابه در تصویر، یا حضور کلاس‌های نادر و اشیاء کوچک در آن ناحیه باشد که به ترتیب ناشی از عدم توازن بین

علاوه بر این، برخلاف فرمول‌بندی کلاسیک آنتروپی متقاطع چندکلاسه که تمرکز آن عمدتاً بر نمونه‌های مثبت است، در این پژوهش مسئله چندکلاسه به صورت چندین مسئله دودویی در نظر گرفته شده است. به این ترتیب، پیکسل‌های منفی، جایی که $g_i^c = 0$ است، نیز در صورت پیش‌بینی نادرست سهمی در مقدار تابع هزینه خواهند داشت. این ویژگی باعث می‌شود مدل حساسیت بیشتری نسبت به اشتباهات در کلاس‌های غیرهدف پیدا کرده و کارایی کلی آن ارتقا یابد.

۴- روش ارزیابی

در این فصل، چارچوب ارزیابی روش پیشنهادی ارائه می‌شود. ابتدا مجموعه داده‌های مورد استفاده و مدل پایه انتخاب شده برای انجام آزمایش‌ها معرفی می‌گردد. سپس جزئیات مربوط به تنظیمات پیاده‌سازی و نحوه آموزش مدل شرح داده می‌شود. در ادامه نیز معیارهای مورد استفاده برای ارزیابی عملکرد مدل‌ها معرفی خواهند شد.

۴-۱- مجموعه داده‌ها و مدل

برای ارزیابی عملکرد روش پیشنهادی در زمینه تحلیل تصاویر شهری، از دو مجموعه داده استاندارد و پرکاربرد UAVid و CityScapes استفاده شده است. این دو مجموعه با دارا بودن ویژگی‌های خاص و چالش‌برانگیز همچون عدم توازن بین کلاسی و درون‌کلاسی، تنوع در مقیاس اشیاء، و وجود نواحی پیچیده بسیار، بستر مناسبی را برای بررسی جامع عملکرد مدل در سناریوهای واقعی فراهم می‌سازند.

مجموعه داده UAVid که در سال ۲۰۲۰ ارائه شده است [۱۷]، شامل ۳۰ قطعه فیلم در قالب استاندارد RGB بوده که در مکان‌های مختلف، تحت شرایط آب و هوایی خوب و نور کافی ثبت شده‌اند. برای تهیه این تصاویر، دوربین روی یک پهپاد نصب شده و با سرعت حداکثر ۱۰ متر بر ثانیه حرکت کرده است تا از تار شدن تصاویر جلوگیری

برای تعلق واقعی آن پیکسل به کلاس c است. علاوه بر این تعداد کل پیکسل‌ها در تصویر I با نماد N و تعداد کلاس‌های موجود در مسئله با نماد C نشان داده می‌شود. بر این اساس، و با توجه به راهکار توصیف شده در بخش قبل، تابع هزینه پیشنهادی برای کلاس c و بر مبنای تابع هزینه آنتروپی متقاطع طبق رابطه (۱) تعریف می‌شود.

(۱)

$$L_W^c = -\frac{1}{N} \sum_{i=1}^N d_i (g_i^c \log p_i^c + (1 - g_i^c) \log (1 - p_i^c))$$

در این رابطه d_i بیانگر میزان پیچیدگی پیکسل i -ام است و با استفاده از رابطه (۲) محاسبه می‌شود.

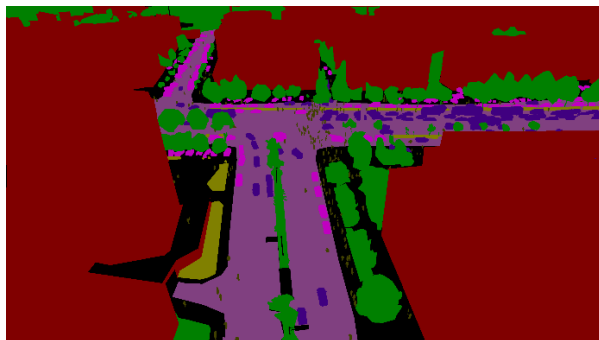
$$d_i = |\{SP_j\}| \quad (2)$$

در این رابطه SP_j نمایانگر ابرپیکسلی است که پیکسل i به آن تعلق دارد، $\{.\}$ نماد مجموعه کلاس‌های یکتای موجود در آن ابرپیکسل و $|\cdot|$ اندازه آن مجموعه است. بنابراین، این رابطه در واقع تعداد کلاس‌های موجود در ابرپیکسلی که پیکسل i در آن قرار دارد را محاسبه می‌کند. در نهایت مقدار هزینه نهایی به صورت میانگین هزینه تمام کلاس‌ها توسط رابطه (۳) به دست می‌آید.

(۳)

$$L = \frac{1}{C} \sum_{c=1}^C L_W^c$$

هر چند فرم کلی رابطه مشابه آنتروپی متقاطع است، اما افزودن d_i که پیچیدگی هر ابرپیکسل را توصیف می‌کند، باعث می‌شود وزن‌دهی به نواحی پیچیده‌تر بیشتر شود. در نتیجه، مدل توجه ویژه‌ای به نواحی با جزئیات بیشتر، مرزها و کلاس‌های اقلیت داشته و ملزم به یادگیری دقیق‌تر این بخش‌ها می‌شود. بنابراین، تفاوت اصلی نه صرفاً در اضافه شدن یک جمله جدید، بلکه در تغییر توزیع اهمیت پیکسل‌ها متناسب با پیچیدگی ابرپیکسل‌ها است که منجر به بهبود نتایج شده است.



(ب)



(آ)

شکل ۲: نمونه‌ای از تصاویر مجموعه داده UAVid. (آ) تصویر اصلی. (ب) برچسب متناظر با آن.

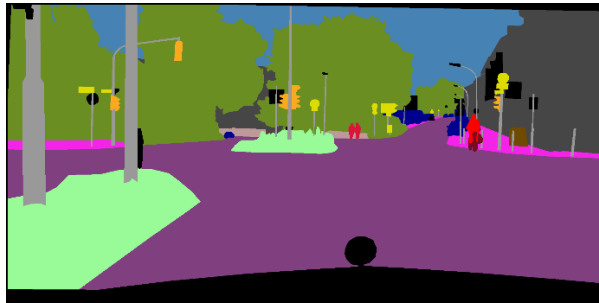
مجموعه داده‌های موجود مزیت‌های قابل توجهی دارد. این مجموعه داده در سال‌های اخیر در پژوهش‌های حوزه بخش‌بندی صحنه‌های شهری مبتنی بر پهنای مورد استفاده قرار گرفته است.

مجموعه داده Cityscapes یکی از مجموعه‌های معتبر در زمینه درک صحنه‌های شهری است که در سال ۲۰۱۶ معرفی شده است [۳۶]. این مجموعه شامل تصاویر RGB با کیفیت بالا از مناظر شهری ۵۰ شهر مختلف در آلمان است که در شرایط نوری مناسب و از دید یک خودروی در حال حرکت ثبت شده‌اند. در این مجموعه، دوربین روی خودرو نصب شده و تصاویر با زاویه دید افقی (نمای روبه‌رو) ثبت گردیده‌اند. این ویژگی آن را برای کاربردهایی همچون رانندگی خودکار و تحلیل محیط شهری، به گزینه‌ای بسیار مناسب تبدیل کرده است. این تصاویر دارای وضوح 256×1024 پیکسل بوده و در قالب PNG ذخیره شده‌اند. نمونه‌ای از تصاویر این مجموعه به همراه برچسب متناظر با آن در شکل ۳ نمایش داده شده است.

در نسخه اصلی این مجموعه، بیش از ۵۰۰۰ تصویر به صورت دستی و با دقت بالا برچسب‌گذاری شده‌اند که شامل ۲۹۷۵ تصویر برای آموزش، ۵۰۰ تصویر برای اعتبارسنجی و ۱۵۲۵ تصویر برای آزمون هستند. برچسب‌ها شامل ۱۹ کلاس معنایی در سطح پیکسل است که عوارض رایج شهری از جمله جاده، پیاده‌رو، ساختمان، درخت، وسایل نقلیه، انسان و دوچرخه

شود. ارتفاع پرواز حدود ۵۰ متر و زاویه دید دوربین بر خلاف تصاویر ماهواره‌ای، مایل (۴۵ درجه) است. این زاویه امکان تصویربرداری با میدان دید وسیع‌تر و جزئیات دقیق‌تر را فراهم می‌سازد. علاوه بر این، عوارض و اشیاء با مقیاس‌های مختلف در تصاویر ظاهر می‌شوند. تصاویر با وضوح 4k و در ابعاد 2160×4096 یا 2160×3840 ثبت و در قالب PNG ذخیره شده‌اند. با توجه به دشواری و هزینه بالای برچسب‌گذاری دستی، از هر فیلم تنها ۱۰ قاب با فاصله زمانی پنج ثانیه انتخاب و برچسب زده شده‌اند. در نتیجه، این مجموعه شامل ۳۰۰ تصویر با مشخصات گفته شده می‌باشد.

برچسب‌های این مجموعه شامل هشت کلاس مختلف با عناوین ساختمان، جاده، درخت، فضای سبز، ماشین ساکن، ماشین در حال حرکت، انسان و کلاس متفرقه است. تصاویر برچسب نیز همانند تصاویر اصلی در قالب PNG به همراه نقشه رنگی مشخص برای هر کلاس ارائه شده‌اند. نمونه‌ای از تصاویر این مجموعه به همراه برچسب آن در شکل ۲ مشاهده می‌شود. از آنجایی که طبقه‌بندی درست کلاس‌هایی مانند انسان و وسیله نقلیه در بسیاری از کاربردهای مرتبط با تحلیل تصاویر شهری اهمیت ویژه‌ای دارد، این مجموعه داده می‌تواند گزینه ارزشمندی برای ارزیابی عملکرد مدل‌ها در این حوزه باشد. افزون بر این، این مجموعه داده از نظر تعداد تصاویر، ابعاد آنها، تعداد پیکسل‌های برچسب خورده، تنوع اشیاء، و پیچیدگی صحنه‌ها نسبت به بسیاری از



(ب)



(آ)

شکل ۳: نمونه‌ای از تصاویر مجموعه داده CityScapes. (آ) تصویر اصلی. (ب) برچسب متناظر با آن.

۴-۲- جزئیات پیاده‌سازی و آموزش

فرایند آموزش مدل با استخراج تکه‌های تصادفی از تصاویر موجود در مجموعه آموزش آغاز می‌شود. به منظور افزایش تنوع داده‌ها و کاهش بیش‌برازش، چرخش افقی تصادفی^{۴۱} به عنوان یک روش داده‌افزایی^{۴۲} روی تمامی تصاویر آموزش اعمال شده است. در مرحله آزمون، تکه‌های تصویر با استفاده از یک پنجره لغزان^{۴۳} و با هم‌پوشانی تقریباً ۵۰ درصد، بدون اعمال هیچ‌گونه روش داده‌افزایی، از تصاویر آزمون استخراج می‌شوند. لازم به ذکر است که نقشه‌های وزنی مورد استفاده در تابع هزینه، برای هر تصویر قبل از شروع فرایند آموزش به صورت غیربرخط^{۴۴} محاسبه و ذخیره می‌شود. در زمان آموزش مدل، هر تصویر به همراه برچسب و نقشه وزنی متناظر آن بارگذاری شده و روش‌های داده‌افزایی یکسانی بر آنها اعمال می‌شود. به این ترتیب، استفاده از تابع هزینه پیشنهادی باعث افزایش پیچیدگی زمانی حین آموزش نمی‌شود.

در تمامی آزمایش‌ها، مدل‌ها با استفاده از روش مقداردهی اولیه گلوروت^{۴۵} [۳۷] تنظیم شده و با بهره‌گیری از بهینه‌ساز آدام^{۴۶} [۳۸] با نرخ یادگیری^{۴۷} ثابت برابر با $1e-4$ آموزش دیده‌اند. در طول فرایند آموزش، عملکرد مدل به طور پیوسته بر روی مجموعه اعتبارسنجی

را پوشش می‌دهند. تنوع مکانی و زمانی، وجود اشیاء متحرک، و چالش‌هایی همچون اجسام کوچک یا نواحی با مرزهای نامشخص، این مجموعه را به گزینه‌ای مناسب برای ارزیابی عملکرد مدل‌ها در حوزه بینایی ماشین تبدیل کرده است. علاوه بر این، این مجموعه داده به طور گسترده در ارزیابی مدل‌های بخش‌بندی معنایی شهری مورد استفاده قرار گرفته و مرجع متداول بسیاری از پژوهش‌های مرتبط است.

برای انجام آزمایش‌ها، از مدل UNet [۱۰] به عنوان یک معماری پایه برای بخش‌بندی تصویر استفاده شده است. این مدل به دلیل سادگی، قابلیت تعمیم مناسب و کارایی بالا در استخراج و بازسازی ویژگی‌های مکانی، انتخاب شده است. UNet به عنوان یک شبکه مبتنی بر ساختار رمزگذار-رمزگشا، در ابتدا برای مسئله بخش‌بندی تصاویر پزشکی طراحی شده بود. در این معماری، برای بهره‌گیری از اطلاعات مکانی دقیق، اتصالات میان‌بر^{۳۷} بین سطوح متناظر رمزگذار و رمزگشا اضافه شده و آن را به گزینه‌ای مناسب برای بخش‌بندی تصویر تبدیل کرده است. در این پژوهش، نسخه پایه مطابق با پیکربندی استاندارد و بدون تغییر پیاده‌سازی شده است. سپس، برای جلوگیری از بیش‌برازش^{۳۸}، یک لایه هم‌آمیخت^{۳۹} اضافی به همراه یک لایه Dropout با نرخ ۰/۱ پیش از لایه خروجی نهایی به مدل افزوده شده است.

40- Patch
41-Random Horizontal Flip
42-Data Augmentation
43-Sliding Window
44-Offline
45-Glorot Uniform Initializer
46-Adam Optimizer
47-Learning Rate

37-Skip Connections
38-Overfitting
39-Convolutional Layer

جدول ۱: پارامترهای آموزش مدل

مجموعه داده	آموزش	اعتبارسنجی	آزمون	کلاس‌ها	تکرار	اندازه دسته	ابعاد تکه‌های تصویر
UAVid	۱۷۰	۳۰	۷۰	۸	۵۰۰	۸	۵۱۲×۱۰۲۴
CityScapes	۲۵۲۸	۴۴۷	۵۰۰	۲۰	۱۰۰	۴	۵۱۲×۱۰۲۴

ارزیابی شده و بر همین اساس، بهترین وزن‌های مدل ذخیره شده و در تحلیل‌های بعدی مورد استفاده قرار گرفته‌اند. آزمایش‌ها با استفاده از یک پردازنده گرافیکی NVIDIA RTX 3090 اجرا شده‌اند.

در هر دو مجموعه داده، تعداد ابرپیکسل‌ها و میزان فشردگی در الگوریتم SLIC به ترتیب برابر با ۲۰۰ و ۱ در نظر گرفته شده است. اندازه جزئیات پارامترهای آموزشی، از جمله تعداد تصاویر در هر بخش از داده (آموزش، اعتبارسنجی و آزمون)، تعداد کلاس‌ها، تعداد تکرارها^{۴۸}، اندازه دسته‌ها^{۴۹} و ابعاد تکه‌های تصاویر^{۵۰} برای هر مجموعه داده، در جدول ۱ ارائه شده است.

۳-۴- معیارهای ارزیابی

برای ارزیابی کمی عملکرد مدل در مسئله بخش‌بندی تصویر، از دو معیار رایج و معتبر شامل $IoU^{۵۱}$ و امتیاز $F1^{۵۲}$ استفاده شده است. معیار IoU یکی از پرکاربردترین معیارها در ارزیابی بخش‌بندی معنایی است که نسبت اشتراک بین ناحیه پیش‌بینی شده و ناحیه واقعی به اجتماع آنها را محاسبه می‌کند. امتیاز $F1$ نیز که به عنوان میانگین هماهنگ دقت^{۵۳} و یادآوری^{۵۴} شناخته می‌شود، معیاری مناسب برای بررسی تعادل بین تعداد نمونه‌های مثبت درست پیش‌بینی شده و نمونه‌های از قلم افتاده یا نادرست است. این دو معیار به ویژه در مواجهه با داده‌هایی با نواحی پیچیده یا توزیع نامتوازن کلاس‌ها، اطلاعات دقیقی درباره کیفیت تفکیک کلاس‌ها فراهم می‌سازند. به منظور کاهش تأثیر توزیع نابرابر بین

کلاس‌ها، هر دو معیار فوق به طور جداگانه برای هر کلاس محاسبه شده و میانگین آنها روی کلاس‌ها گزارش شده تا عملکرد مدل در تشخیص کلاس‌های کم تعداد نیز به درستی سنجیده شود. این دو معیار از روابط زیر به دست می‌آیند.

$$IoU = \frac{TP}{TP + FN + FP} \quad (۴)$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (۵)$$

$$Precision = \frac{TP}{TP + FP} \quad (۶)$$

$$Recall = \frac{TP}{TP + FN} \quad (۷)$$

در این روابط $TP^{۵۵}$ بیانگر تعداد داده‌هایی از کلاس مثبت که به درستی پیش‌بینی شده‌اند، و $FP^{۵۶}$ و $FN^{۵۷}$ به ترتیب نشان دهنده تعداد داده‌های نادرست از کلاس‌های مثبت و منفی هستند.

۵- نتایج تجربی و تحلیل

برای ارزیابی عملکرد مدل‌های مختلف، دو مجموعه داده UAVid و Cityscapes مورد استفاده قرار گرفته‌اند. در ادامه عملکرد مدل‌ها تحت پنج تابع هزینه متفاوت شامل تابع آنتروپی متقاطع، دایس، کانونی، تورسکی و تابع هزینه پیشنهادی، با استفاده از معیارهای IoU و امتیاز $F1$ بررسی می‌شود.

55-True Positive
56-False Positive
57-False Negative

48-Epochs
49-Batch Size
50-Patch Size
51-Intersection over Union
52-F1-Score
53-Precision
54-Recall

۵-۱- بررسی نتایج مجموعه داده UAVid

بررسی نتایج حاصل از آزمایش‌ها بر روی مجموعه داده UAVid نشان می‌دهد که روش پیشنهادی در مقایسه با سایر توابع هزینه، عملکرد بهتری در هر دو معیار IoU و امتیاز F1 داشته است. همان طور که در جدول ۲ مشاهده می‌شود، این روش موفق شده است به مقدار IoU برابر با ۵۹/۴۶ درصد و امتیاز F1 معادل ۷۰/۵۷ درصد دست یابد که در هر دو شاخص بالاترین مقدار در بین روش‌های مورد مقایسه است. در این مجموعه، سه تابع آنتروپی متقاطع، دایس و تورسکی تقریباً عملکردی نزدیک به هم داشته‌اند، اما به طور محسوسی از روش پیشنهادی ضعیف‌تر بوده‌اند. به طور خاص، روش پیشنهادی نسبت به این توابع در معیار IoU حدود ۱/۵ تا ۲ درصد و در امتیاز F1 حدود ۱/۳ تا ۱/۷ درصد بهبود نشان داده است. تابع هزینه کانونی نیز با اختلاف اندکی پس از روش پیشنهادی در جایگاه دوم قرار گرفته است.

با توجه به ماهیت مجموعه داده UAVid، که شامل تصاویر پیچیده با کلاس‌هایی با اندازه‌های مختلف و توزیع نامتوازن است، این نتایج تأیید می‌کند که وزن‌دهی هدفمند در تابع هزینه می‌تواند نقش مهمی در افزایش تمرکز مدل بر نواحی چالش‌برانگیز و کلاس‌های کم‌تعداد ایفا کند. افزایش قابل ملاحظه در امتیاز F1 نشان می‌دهد که روش پیشنهادی، علاوه بر افزایش میزان همپوشانی بین اشیاء پیش‌بینی شده و نواحی صحیح، در کاهش خطاهای مربوط به تشخیص نمونه‌های مثبت و منفی مؤثر بوده و تعادل بین دقت و یادآوری را نیز بهبود بخشیده است.

۵-۲- بررسی نتایج مجموعه داده CityScapes

نتایج حاصل از انجام آزمایش‌های مختلف بر روی مجموعه داده CityScapes نیز نشان‌دهنده برتری روش پیشنهادی نسبت به سایر توابع هزینه است. با توجه

جدول ۲: نتایج به‌دست آمده بر روی مجموعه داده UAVid

تابع هزینه	IoU	امتیاز F1
آنتروپی متقاطع	۵۷/۹۵	۶۹/۲۲
دایس	۵۷/۶۹	۶۹/۲۱
کانونی	۵۹/۳۵	۷۰/۴۳
تورسکی	۵۷/۳۴	۶۸/۸۰
روش پیشنهادی	۵۹/۴۶	۷۰/۵۷

جدول ۳: نتایج به‌دست آمده بر روی مجموعه داده CityScapes

تابع هزینه	IoU	امتیاز F1
آنتروپی متقاطع	۵۳/۹۷	۶۰/۱۶
دایس	۵۰/۵۹	۵۷/۰۸
کانونی	۵۳/۳۵	۵۹/۶۲
تورسکی	۵۴/۲۳	۶۰/۹۶
روش پیشنهادی	۵۴/۳۱	۶۰/۵۳

به جدول ۳، این روش در معیار IoU به مقدار ۵۴/۳۱ درصد و در امتیاز F1 به مقدار ۶۰/۵۳ درصد دست یافته است، که در هر دو مورد بالاتر از توابع هزینه آنتروپی متقاطع، دایس و کانونی می‌باشد. در این مجموعه داده ضعیف‌ترین عملکرد مربوط به تابع هزینه دایس بوده و توابع آنتروپی متقاطع و کانونی با فاصله کمتری پس از تابع هزینه پیشنهادی قرار گرفته‌اند. به طور دقیق، روش پیشنهادی نسبت به تابع هزینه دایس در معیار IoU حدود ۳/۷ درصد و در امتیاز F1 مقدار ۳/۵ درصد بهبود داشته است. همچنین تابع هزینه پیشنهادی در مقایسه با تابع کانونی حدود ۰/۹ درصد و در مقایسه با آنتروپی متقاطع حدود ۰/۳ درصد برای هر دو معیار بهبود نشان می‌دهد. تابع هزینه تورسکی نیز با وجود پیشی گرفتن از تابع هزینه پیشنهادی در امتیاز F1، با اختلاف اندکی در معیار IoU پس از آن در رده دوم قرار گرفته است. این نتایج نشان می‌دهد که الگوی وزن‌دهی نقاط در روش پیشنهادی، در این مجموعه داده نیز توانسته است در فرایند یادگیری مدل، به‌ویژه در نواحی دشوارتر، تأثیر مثبتی بگذارد و دقت تفکیک کلاس‌ها را ارتقا دهد.

۶- جمع‌بندی

شهری بوده و زمینه‌ساز توسعه مدل‌هایی با عملکرد قابل اطمینان‌تر در کاربردهای واقعی باشند.

۷- مراجع

- [1] Minaee, S., Boykov, Y. Y., Porikli, F., Plaza, A. J., Kehtarnavaz, N. & Terzopoulos, D. (2021). "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*
- [2] Otsu, N. (1979). "A threshold selection method from gray-level histograms," *IEEE Trans. Syst. Man. Cybern.*, vol. 9, no. 1, pp. 62–66.
- [3] Dhanachandra, N., Mangle, K. & Chanu, Y. J. (2015). "Image segmentation using K-means clustering algorithm and subtractive clustering algorithm," *Procedia Comput. Sci.*, vol. 54, pp. 764–771, 2015.
- [4] Nock, R. & Nielsen, F. (2004). "Statistical region merging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1452–1458.
- [5] Najman, L., & Schmitt, M. (1994). "Watershed of a continuous function," *Signal Processing*, vol. 38, no. 1, pp. 99–112.
- [6] Boykov, Y., Veksler, O. & Zabih, R. (2001). "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239.
- [7] Kass, M., Witkin, A., & Terzopoulos, D. (1998). "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331.
- [8] Plath, N., Toussaint, M. & Nakajima, S. (2009). "Multi-class image segmentation using conditional random fields and global classification," in *Proceedings of the 26th annual international conference on machine learning*, 2009, pp. 817–824.
- [9] J. Ma et al., "Loss odyssey in medical image segmentation," *Med. Image Anal.*, p. 102035, 2021.
- [10] Ronneberger, O., Fischer, P., & Brox, T. (2015). "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241.
- [11] Wu, Z., Shen, C. & Van den Hengel, A. (2016). "Bridging category-level and instance-level semantic image segmentation," *arXiv Prepr. arXiv1605.06885*.
- [12] Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, pp. 2980–2988.
- [13] Arulananth, T. S. et al., (2024). "Semantic segmentation of urban environments: Leveraging U-Net deep learning model for cityscape image analysis," *PLoS One*, vol. 19, no. 4, p. e0300767.
- [14] Sahragard, E., Farsi, H. & Mohamadzadeh, S. (2025). "Advancing semantic segmentation: Enhanced UNet algorithm with attention mechanism and deformable convolution," *PLoS One*, vol. 20, no. 1, p. e0305561.
- [15] Zhang, Z. & Li, G. (2025). "UAV Imagery Real-Time Semantic Segmentation with Global--Local Information Atten-

در این پژوهش، یک تابع هزینه جدید با هدف بهبود عملکرد مدل‌های بخش‌بندی معنایی تصاویر شهری، به ویژه در شرایط وجود عدم توازن در کلاس‌ها، ارائه شده است. ایده اصلی روش پیشنهادی، تمرکز بیشتر بر نواحی پیچیده تصویر مانند مرز بین کلاس‌ها، اشیاء کوچک یا نادر، یا کلاس‌هایی با شباهت ظاهری بالا است. بدین منظور، ابتدا تصاویر به کمک الگوریتم SLIC به ابرپیکسل‌هایی با ویژگی‌های ظاهری مشابه تقسیم شده و سپس با تحلیل تراکم کلاس‌ها در هر ابرپیکسل، وزن‌دهی هدفمند به پیکسل‌ها در تابع هزینه اعمال می‌شود.

مزیت اصلی این رویکرد آن است که بدون ایجاد تغییرات گسترده در معماری مدل، می‌تواند دقت مدل را در شناسایی کلاس‌های اقلیت و نواحی دشوار بهبود بخشد. نتایج تجربی حاصل از پیاده‌سازی این روش بر روی دو مجموعه داده شهری با درجات مختلف عدم توازن، و با استفاده از مدل UNet، نشان می‌دهد که تابع هزینه پیشنهادی به طور معناداری معیارهای ارزیابی عملکرد از جمله IoU و امتیاز F1 را نسبت به توابع هزینه متداول بهبود می‌بخشد.

در نهایت، یافته‌های این پژوهش بر اهمیت طراحی توابع هزینه آگاه از ساختار تصویر و نواحی بحرانی تأکید دارد. از محدودیت‌های روش پیشنهادی می‌توان به وابستگی آن به کیفیت تقسیم‌بندی ابرپیکسل‌ها اشاره کرد، چرا که عملکرد نهایی تا حدی به نتایج حاصل از الگوریتم SLIC وابسته است. به عنوان مسیرهای آتی پژوهش، می‌توان بررسی تأثیر عواملی همچون اندازه ابرپیکسل، چگالی لبه‌ها و ویژگی‌های بافت را نیز در تعریف شاخص پیچیدگی مدنظر قرار داد، تا وزن‌دهی دقیق‌تری به نواحی تصویر حاصل شود. چنین رویکردهایی می‌توانند گامی مؤثر در جهت حل چالش‌های ذاتی در مسائل بخش‌بندی معنایی تصاویر

- put and output imbalance in multi-organ segmentation,” *Comput. Med. Imaging Graph.*, vol. 75, pp. 24–33.
- [30] Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J. & Maier-Hein, K. H. (2021). “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation,” *Nat. Methods*, vol. 18, no. 2, pp. 203–211.
- [31] Zhu, W. et al., (2019). “AnatomyNet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy,” *Med. Phys.*, vol. 46, no. 2, pp. 576–589.
- [32] Brugnara, G. et al., (2020). “Automated volumetric assessment with artificial neural networks might enable a more accurate assessment of disease burden in patients with multiple sclerosis,” *Eur. Radiol.*, vol. 30, no. 4, pp. 2356–2364.
- [33] Arega, T. W., Bricq, S., & Meriaudeau, F. (2022). “Using polynomial loss and uncertainty information for robust left atrial and scar quantification and segmentation,” in *Challenge on Left Atrial and Scar Quantification and Segmentation*, Springer, 2022, pp. 133–144.
- [34] El Jurdi, R., Petitjean, C., Honeine, P., Cheplygina, V. & Abdallah, F. (2021). “High-level prior-based loss functions for medical image segmentation: A survey,” *Comput. Vis. Image Underst.*, vol. 210, p. 103248.
- [35] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P. & Süsstrunk, S. (2012). “SLIC superpixels compared to state-of-the-art superpixel methods,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282.
- [36] Cordts et al., M. (2016). “The cityscapes dataset for semantic urban scene understanding,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3213–3223.
- [37] Glorot, X. & Bengio, Y. (2010). “Understanding the difficulty of training deep feedforward neural networks,” in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249–256.
- [38] Kingma, D. P. & Ba, J. (2014). “Adam: A method for stochastic optimization,” *arXiv Prepr. arXiv1412.6980*.
- tion,” *Sensors*, vol. 25, no. 6, p. 1786.
- [16] Yang, G., Wang, Y., Shi, D. & Wang, Y. (2025). “Golden Cudgel Network for Real-Time Semantic Segmentation,” in *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 25367–25376.
- [17] Lyu, Y., Vosselman, G., Xia, G.-S., Yilmaz, A. & Yang, M. Y. (2020). “UAVid: A semantic segmentation dataset for UAV imagery,” *ISPRS J. Photogramm. Remote Sens.*, vol. 165, pp. 108–119.
- [18] Hesham et al., S. A. S. (2025). “Exploiting temporal state space sharing for video semantic segmentation,” in *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 24211–24221.
- [19] Li, X. et al., (2020). “Improving semantic segmentation via decoupled body and edge supervision,” in *European Conference on Computer Vision*, pp. 435–452.
- [20] Milletari, F., Navab, N., & Ahmadi, S. A. (2016). “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” in *2016 fourth international conference on 3D vision (3DV)*, 2016, pp. 565–571.
- [21] Drozdal, M., Vorontsov, E., Chartrand, G., Kadoury, S. & Pal, C. (2016). “The importance of skip connections in biomedical image segmentation,” in *Deep learning and data labeling for medical applications*, Springer, 2016, pp. 179–187.
- [22] Rahman M. A., & Wang, Y. (2016). “Optimizing intersection-over-union in deep neural networks for image segmentation,” in *International symposium on visual computing*, 2016, pp. 234–244.
- [23] Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S. & Cardoso, M. J. (2017). “Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations,” in *Deep learning in medical image analysis and multimodal learning for clinical decision support*, Springer, 2017, pp. 240–248.
- [24] Salehi, S. S. M., Erdogmus, D., & Gholipour, A. (2017). “Tversky loss function for image segmentation using 3D fully convolutional deep networks,” in *International workshop on machine learning in medical imaging*, pp. 379–387.
- [25] Abraham, N., & Khan, N. M. (2019). “A novel focal tversky loss function with improved attention u-net for lesion segmentation,” in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 683–687.
- [26] Kervadec, H., Bouchtiba, J., Desrosiers, C., Granger, E., Dolz, J., & Ben Ayed, I. (2021). “Boundary loss for highly unbalanced segmentation,” *Med. Image Anal.*, vol. 67, p. 101851, 2021.
- [27] Karimi, D. & Salcudean, S. E. (2019). “Reducing the hausdorff distance in medical image segmentation with convolutional neural networks,” *IEEE Trans. Med. Imaging*, vol. 39, no. 2, pp. 499–513, 2019.
- [28] Wong, K. C. L., Moradi, M., Tang, H. & Syeda-Mahmood, T. (2018). “3D segmentation with exponential logarithmic loss for highly unbalanced object sizes,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 612–619.
- [29] Taghanaki, S. A. et al., (2019). “Combo loss: Handling in-