

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشکده مهندسی کامپیوتر

## پایان نامه کارشناسی ارشد مهندسی هوش مصنوعی

ارائه یک الگوریتم بهینه برای هم‌بخش‌بندی تصویر

نگارنده : محمد صالح ملایی پاشایی

استاد راهنما :

دکتر علی اکبر پویان

استاد مشاور :

دکتر منصور فاتح

شهریور ۱۳۹۶

## پاس گذاری

خدای را به یساکرم که از روی کرم پدر و مادری فداکار نصیبم ساخته تا در سایه  
درخت پر بار وجودشان بیایم و از ریشه آنها شاخ و برگ که می‌رم و از سایه وجودشان  
در راه کرب علام و دانش تلاش نمایم.

که بولاندشیلین تاج افتخاری است بر سرم و نامشان دایم است بر بودم چرا که  
این دو وجود پس از پروردگاریه همیشه ام بوده اند و تم را گرفتند و راه رفتن  
را در این وادی زندگی پر از فراز و نشیب آموختند.

آموزگارانی که برایم زندگی؛ بودن و انسان بودن را معنا کردند  
اینجا بگم سبزی است تخم درویش تقدیم وجود با ارزش آنان...

## تسکرو قدردانی

اسی را کہ پہلین فضلان، در ستودن او بمانند و شمارندگان، شمردن نعمت های او ندانند و کوشندگان، حق او را گزاردن نتوانند. و سلام و دورد بر چه دو خندان پاک او، طاهران موصوم، ہم آفتاب و کمان و امدار و جودشان است و نغمه ترین پیوسته بردشیران ایشان تاروزر تا خمیر...

بدون شک جایگاه و منزلت معلم، اجل از آن است کہ در مقام قدردانی از زحمات بنی شائبی او، بازبان قاصد مرو درست ناتوان، چیزی بکاریم. اما از آنجایی کہ تبدیل از معلم، ریاس از انسانی است گفت و غایت آفرینش را تأمین می کند و سلامت نمایانست کہ به دستش رسیده اند، تقدیر بنی خود و واجب میدانم ابتدا با بوسه ای بردستان پدر و مادر عزیزم. این دو معلم بنی که او را هر که تو با بنی و در شتی من، قام عذو کشیده و کریمانه از کنار غفلت هایم گذشته انباز گذارشان باشم. و از اتاد با کمالات و شایسته، جناب آقای دکتر عالی که مر و میان که در کمال سعادت، با حسن مذاق و فروتنی، از بیچ امکان در این عرصه بر من دریغ نند و دندوز حرمت را بنه مای این رساله را بر عهده گرفتند و پندین از اتاد صبر و رباته تو، جناب آقای دکتر که در فلاح ترحم مشاوره این رساله را در حای ممتقل شدند که بدون مساعدت ایشان، این پروژه به نتیجه مطلوب نمی رسید کمال تسکرو قدردانی را بلاشکلم این خردترین، بنه شتی از زحمات آنان را ریاس گوید.

## تعهد نامه

اینجانب محمد صالح ملایی پاشایی دانشجوی دوره کارشناسی ارشد رشته مهندسی کامپیوتر- گرایش هوش مصنوعی دانشکده مهندسی کامپیوتر دانشگاه صنعتی شاهرود نویسنده پایان نامه ارائه یک الگوریتم بهینه به منظور هم‌بخش‌بندی تصویر تحت راهنمایی آقای دکتر علی‌اکبر پویان متعهد می‌شوم:

- تحقیقات در این پایان‌نامه توسط اینجانب انجام شده است و از صحت و اصالت برخوردار است.
- در استفاده از نتایج پژوهش‌های محققان دیگر به مرجع مورد استفاده استناد شده است.
- مطالب مندرج در پایان‌نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی در هیچ جا ارائه نشده است.
- کلیه حقوق معنوی این اثر متعلق به دانشگاه صنعتی شاهرود می‌باشد و مقالات مستخرج با نام «دانشگاه صنعتی شاهرود» و یا «Shahrood University of Technology» به چاپ خواهد رسید.
- حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی پایان‌نامه تأثیرگذار بوده‌اند در مقالات مستخرج از پایان‌نامه رعایت می‌گردد.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که از موجود زنده (یا بافتهای آنها) استفاده شده‌است، ضوابط و اصول اخلاقی رعایت شده‌است.
- در کلیه مراحل انجام این پایان‌نامه، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده‌است اصل رازداری، ضوابط و اصول اخلاق انسانی رعایت شده‌است.

تاریخ

امضای دانشجو

### مالکیت نتایج و حق نشر

- کلیه حقوق معنوی این اثر و محصولات آن (مقالات مستخرج، کتاب، برنامه‌های رایانه‌ای، نرم افزارها و تجهیزات ساخته شده است) متعلق به دانشگاه صنعتی شاهرود می‌باشد.
- این مطلب باید به نحو مقتضی در تولیدات علمی مربوطه ذکر شود.
- استفاده از اطلاعات و نتایج موجود در پایان‌نامه بدون ذکر مرجع مجاز نمی‌باشد.

## چکیده :

بخش‌بندی تصویر و پیدا کردن اشیاء در تصاویر همواره یک کار مهم در بینایی ماشین بوده است. هم-بخش‌بندی به معنای بخش‌بندی قسمت‌های مشترک (شی‌های مشترک) در دو تصویر یا یک مجموعه تصاویر به صورت همزمان است این قسمت‌های مشترک می‌تواند یک شی صلب، غیر صلب، صحنه و شی‌های مشابه از یک کلاس باشند. همچنین، اشیا می‌توانند از دیدگاه‌های مختلف نیز دیده شوند. در این پژوهش به یک مسئله‌ای در زمینه هم‌بخش‌بندی تصویر می‌پردازیم که آن را هم‌بخش‌بندی چند پیش‌زمینه‌ای می‌نامند به این صورت که تعدادی پیش‌زمینه در کل مجموعه تصاویر خود داریم و در هر تصویر، زیرمجموعه‌ای از کل پیش‌زمینه‌ها وجود دارد. این قضیه در تضاد با مشکل هم‌بخش‌بندی کلاسیک که توسط بیشتر الگوریتم‌های موجود ارائه شده است می‌باشد. چون که فرضشان ساده‌تر می‌باشد. این پژوهش یک روش بهینه برای هم‌بخش‌بندی با چند پیش‌زمینه ارائه می‌کند که هیچ فرضی روی پس‌زمینه نمی‌سازد و محدودیت‌های موجود در روش‌های قبل را که باعث تحمل خطا می‌شدند را ندارد.

در این تحقیق از یک طرح تکراری استفاده می‌شود، که متناوباً یک روال مدل کردن پیش‌زمینه و مدل تخصیص ناحیه را اجرا می‌کند. هر دو روشی که در دو مرحله فوق مورد استفاده قرار می‌دهیم دارای بهینگی مناسبی می‌باشند. بطور کلی، الگوریتم ارائه شده، به اندازه کافی برای تشکیل کلاس‌بندی پیشرفته هر پیش‌زمینه انعطاف‌پذیر است و ناحیه مشخص شده در یک روال انتخابی ترکیبی مورد ارزیابی قرار می‌گیرد. این الگوریتم دارای چندین مشخصه خوب قابل لمس مانند پیچیدگی خطی و دقت در تشخیص است.

کلمات کلیدی: هم‌بخش‌بندی، چند پیش‌زمینه، ناحیه کاندیدا، هرم تطبیق مکانی، گراف حالت

## فهرست مطالب

### ۱. مقدمه

- ۱-۱- تعریف هم‌بخش‌بندی تصویر ..... ۲
- ۲-۱- تاریخچه هم‌بخش‌بندی تصویر ..... ۳
- ۳-۱- کاربردهای هم‌بخش‌بندی تصویر ..... ۴
- ۱-۳-۱- دسته بندی و برچسب گذاری تصاویر ..... ۴
- ۱-۳-۲- بازیابی تصاویر مربوط به یک یا چند شی خاص از یک پایگاه داده ..... ۵
- ۱-۴- تعریف مسئله ..... ۵
- ۱-۵- هدف و نوآوری پایان نامه ..... ۷
- ۱-۶- ساختار پایان نامه ..... ۷

### ۲. کارهای گذشته

- ۱-۲- مقدمه ..... ۱۰
- ۲-۲- مدل کردن و تشخیص شی (پیش‌زمینه) ..... ۱۱
- ۱-۲-۲- روش‌های مبتنی بر گروه‌بندی ..... ۱۲
- ۲-۲-۲- روش‌های امتیازدادن به پنجره کاندید ..... ۲۰
- ۲-۲-۳- روش‌های پایه یا مرجع در پیشنهاد کاندید ..... ۲۱
- ۲-۲-۴- کنترل تعداد کاندیدهای پیشنهادی ..... ۲۲
- ۲-۲-۵- مقایسه بین روش‌های نواحی کاندید ..... ۲۴
- ۲-۳- روش‌های تطبیق پیش‌زمینه‌های کاندید با پیش‌زمینه‌های مورد نظر ..... ۲۵

- ۲-۳-۱- بسته کلمات ..... ۲۵
- ۲-۳-۲- روش هرم تطابق مکانی ..... ۲۹
- ۲-۳-۱- هسته تطبیق هرمی ..... ۲۹
- ۲-۳-۲- هرم تطابق مکانی ..... ۳۱
- ۲-۳-۳- کدگذاری تنک ..... ۳۱
- ۲-۳-۴- کدگذاری با پایه‌های محلی ..... ۳۲
- ۲-۳-۵- کدگذاری محلی و خطی ..... ۳۳
- ۲-۳-۶- کدگذاری نرم با برش ..... ۳۴
- ۲-۳-۷- هسته بهینه برای تشخیص شباهت ویژگی‌ها ..... ۳۵
- ۲-۳-۷-۱- یادگیری ویژگی با بعد محدود ..... ۳۶
- ۲-۳-۹- نتیجه‌گیری ..... ۳۸

### ۳. روش پیشنهادی

- ۳-۱- مقدمه ..... ۴۰
- ۳-۲- فرموله کردن مسئله ..... ۴۱
- ۳-۳- مدل کردن پیش‌زمینه ..... ۴۲
- ۳-۳- تخصیص ناحیه ..... ۴۴
- ۳-۳- روش پیشنهادی به منظور بخش‌بندی همزمان چند پیش‌زمینه ..... ۴۵
- ۳-۳-۱- درخت محدود کننده تخصیص ناحیه ..... ۴۵
- ۳-۳-۲- تشکیل مجموعه کاندیدها ..... ۴۸
- ۳-۳-۳- انتساب نواحی قابل تشخیص ..... ۵۰
- ۳-۴- شبه کد الگوریتم پیشنهادی ..... ۵۲
- ۳-۵- جمع بندی ..... ۵۴



## نتایج و آمارها

- ۴-۱- مقدمه ..... ۵۶
- ۴-۲- معرفی پایگاه داده ..... ۵۶
- ۴-۳- نتایج روش پیشنهادی ..... ۵۸
- ۴-۴- مقایسه الگوریتم پیشنهادی با روش‌های موجود ..... ۶۴
- ۴-۵- جمع‌بندی ..... ۶۵

## ۵. نتیجه‌گیری و کارهای آینده

- ۵-۱- بحث و نتیجه‌گیری ..... ۶۸
- ۵-۲- کارهای آینده ..... ۶۸
- منابع ..... ۷۰

## فهرست اشکال

- شکل ۱-۱ نمونه ای از تصویر پایگاه داده با برچسب کودک و سبد سیب ..... ۶
- شکل ۱-۲ نمونه ای از ناحیه کاندید ..... ۱۱
- شکل ۲-۳ الگوریتم گروه بندی سلسله مراتبی در جستجو انتخابی [۱۴] ..... ۱۳
- شکل ۲-۴ نمونه ای از خروجی الگوریتم [۱۴]. محدوده قرمز ناحیه کاندید روش جستجو انتخابی و محدوده سبز ناحیه هدف است ..... ۱۳
- شکل ۲-۵ مراحل کلی الگوریتم چانگ [۱۹] شامل نمونه گیری تصافی جدید و انتصاب مقدار برجستگی به هر پنجره، تخمین شی بودن می باشد ..... ۱۵
- شکل ۲-۶ روال پیشنهاد شده در [۲۸] ..... ۱۶
- شکل ۲-۷ مراحل مختلف از روش [۲۱، الف] تصویر ورودی، ب) قطعه بندی سلسله مراتبی، ج) ناحیه های کاندید، د) رتبه بندی ناحیه ها ..... ۱۷
- شکل ۲-۸ مراحل کلی از روش Geodesic مطرح شده در مرجع [۲۲] (الف) تصویر ورودی و پیش قطعه بندی آن به همراه نگاشت احتمال مرز، ب) مکان Seeds، پ) ماسک پیش-زمینه و پس زمینه تولید شده برای دو Seeds، ج) استفاده از تبدیل فاصله Geodesic (SGDB) برای این ماسک ها، د) کاندیدهای پیشنهادی با توجه به محاسبه مجموعه های سطوح بحرانی در هر SGDB ..... ۱۸
- شکل ۲-۹ اشیاء کاندید (نواحی قرمز) توسط الگوریتم [۲۲] ..... ۱۹
- شکل ۲-۱۰ مقایسه بین روش های پیشنهاد کاندید بر روی مجموعه داده PASCAL VOC 2007 با روش R-CNN سریع [۱۳] ..... ۲۴
- شکل ۲-۱۱ مراحل یادگیری پایه ها در الگوریتم [۲۸] ..... ۲۶
- شکل ۲-۱۲ ایجاد بردار ویژگی در الگوریتم [۲۸] ..... ۲۷
- شکل ۲-۱۳ مراحل کلی استخراج ویژگی در روش هرم تطابق مکانی ..... ۳۰

- شکل ۳-۱ شمای کلی الگوریتم هم‌بخش‌بندی ..... ۴۲
- شکل ۳-۱ یک مثال برای مدل پیش‌زمینه نوزاد ..... ۴۳
- شکل ۳-۲ پیش تقسیم تصویر I به  $S_i$  ..... ۴۶
- شکل ۳-۳ نمایشی از امتیاز دادن به کاندیدهای پیشنهادی هر پیش‌زمینه ..... ۴۷
- شکل ۳-۴ خروجی الگوریتم CABOB برای تصویر با برچسب کودک و سبد سیب از پایگاه داده flicker ..... ۵۲
- شکل ۴-۱ نمونه ای از تصاویر پایگاه داده Flickr با برچسب کودک و سبد سیب ..... ۵۷
- شکل ۴-۲ نمونه ای از تصاویر پایگاه داده Flickr با برچسب گاو و چراگاه ..... ۵۷
- شکل ۴-۳ بخش‌بندی تصویر I با برچسب سبد سیب و نوزاد ..... ۵۸
- شکل ۴-۴ نحوه تشکیل گراف حالت برای تصویر I با برچسب سبد سیب و نوزاد ..... ۵۹
- شکل ۴-۵ خروجی الگوریتم جستجو CABOB ..... ۵۹
- شکل ۴-۲ خروجی الگوریتم sift برای یک تصویر مجموعه گاو و چراگاه ..... ۶۰
- شکل ۴-۶ خروجی الگوریتم پیشنهادی برای هم‌بخش‌بندی ..... ۶۰
- شکل ۴-۷ خروجی الگوریتم هم‌بخش‌بندی برای تصویر با برچسب گاو و چمنزار ..... ۶۱
- شکل ۴-۸ خروجی الگوریتم از مجموعه تصاویر با برچسب سنگ ..... ۶۲
- شکل ۴-۹ خروجی الگوریتم پیشنهادی برای مجموعه تصاویر با برچسب پاندا ..... ۶۳

## فهرست جداول

- جدول ۱-۲ مقایسه‌ای بین روش‌های معرفی شده. علامت +/- به معنی، دارا بودن یا نبودن خاصیت مشخص است و علامت \* در کنترل تعداد کاندید، نشان دهنده‌ی کنترل غیر مستقیم است. .... ۲۳
- ۱-۳-۲- بسته کلمات ..... ۲۵
- جدول ۱-۳ شبه کد مدل الگوریتم جستوجوی BEAM برای تولید ناحیه‌های کاندید ..... ۴۹
- جدول ۲-۳ نمایش شبه کد الگوریتم پیشنهادی ..... ۵۳
- جدول ۱-۴ مقایسه نتایج الگوریتم پیشنهادی با سایر الگوریتم‌ها ( به درصد) ..... ۶۴
- جدول ۱-۴ میانگین تشخیص درست پیش‌زمینه‌ها در پایگاه داده **fliker** ..... ۶۵
- جدول ۳-۴ زمان اجرای الگوریتم پیشنهادی بر روی دسته‌های مختلف با تعداد ۱۰ تصویر ..... ۶۵





فصل اول:

# مقدمه

## ۱-۱- تعریف هم‌بخش‌بندی تصویر

یکی از مسایل پایه‌ای در بینایی ماشین مسئله بخش‌بندی<sup>۱</sup> است که نقش مهمی در حل دیگر مسایل سطح بالای این حوزه نظیر کشف شی<sup>۲</sup>، تشخیص شی<sup>۳</sup>، ردیابی شی<sup>۴</sup> و درک صحنه<sup>۵</sup> دارد. اگرچه اقدامات بسیار زیادی بر روی این مسئله انجام شده، اما چالش‌های آن مانند نمایش ویژگی تا طراحی مدل و بهینه‌سازی به طور کامل حل نشده است. بدین معنا که روش‌های ارائه شده با خطا همراه است و هیچ الگوریتم معتبر کامل پایین به بالایی<sup>۶</sup> برای بخش بندی یک تصویر معرفی نشده است [۱-۲]. برای تصحیح بخشی از این خطاها، می‌توان از تعدادی تصاویر مرتبط استفاده کرد و همه آنها را به صورت همزمان بخش‌بندی نمود، این مفهوم «هم‌بخش‌بندی»<sup>۷</sup> نامگذاری شده است [۳] که در سال‌های اخیر بسیار مورد توجه پژوهشگران قرار گرفته است.

هم‌بخش‌بندی به معنای بخش‌بندی قسمت‌های مشترک (شی‌های مشترک) در دو تصویر [۳-۴] یا یک مجموعه تصاویر [۲-۴-۵] به صورت همزمان است این قسمت‌های مشترک می‌تواند یک شی صلب، غیر صلب، صحنه و شی‌های مشابه از یک کلاس باشند. همچنین، اشیا می‌توانند از دیدگاه‌های مختلف نیز دیده شوند [۳].

مدلی که در هم‌بخش‌بندی استفاده می‌شود از قبل آموزش داده نشده است. یعنی هیچ اطلاعات قبلی راجع به تصاویر (مانند کلاس شی) در دسترس نیست. بنابراین، مجموعه تصاویر از پیش تعریف شده‌ای<sup>۸</sup> آموزش داده نمی‌شود، بلکه مدل روی تصاویر داده‌شده جدید ایجاد می‌گردد.

---

<sup>1</sup> Segmentation  
<sup>2</sup> Object Detection  
<sup>3</sup> Object Recognition  
<sup>4</sup> Object Tracking  
<sup>5</sup> Scene Understanding  
<sup>6</sup> Bottom-up  
<sup>7</sup> cosegmentation  
<sup>8</sup> Predefined Set Of Images



این موضوع باعث می‌شود که استفاده از هم‌بخش‌بندی مناسب باشد زیرا نسبت به روش‌های معمول تنظیمات انعطاف‌پذیرتری دارد و نیز دغدغه تنظیم پارامترهای پایگاه داده وجود ندارد. همچنین می‌تواند برای بخش‌بندی مجموعه داده‌های بزرگ و بسیاری از کاربردهای عملی مورد استفاده قرار گیرد [۶].

به طور کلی، الگوریتم‌های ارائه شده برای هم‌بخش‌بندی از نظر میزان نظارت<sup>۹</sup> به دسته‌های هم-بخش‌بندی بدون سرپرست<sup>۱۰</sup> و هم‌بخش‌بندی تعاملی<sup>۱۱</sup> یا با سرپرست<sup>۱۲</sup> قابل تقسیم هستند [۷].

## ۱-۲- تاریخچه هم‌بخش‌بندی تصویر

در ابتدا، اکثر الگوریتم‌های ارائه شده برای هم‌بخش‌بندی به صورت بدون نظارت و برای یک جفت تصویر دارای یک شی دقیقاً مشترک بودند [۳-۶-۸]. ایده مشترک در این دسته الگوریتم‌ها، فرمول‌بندی هم‌بخش‌بندی به منظور کمینه کردن یک تابع انرژی، مبتنی بر یک مسئله ی برچسب-گذاری دودویی<sup>۱۳</sup> است. در این روش‌ها، یک تابع انرژی واحد بر روی همه پیکسل‌ها در همه تصاویر مجموعه تعریف می‌شد. این تابع انرژی معمولاً متشکل از دو عبارت است. یک عبارت مشخص‌کننده یکنواختی<sup>۱۴</sup> بخش‌بندی در هر تصویر و دیگری نشانگر میزان شباهت نواحی مشترک در همه تصاویر است [۷]. به عنوان نمونه، در [۳]، تابع انرژی شامل میدان تصادفی مارکوف معمولی است (که بخش‌بندی را در هر تصویر یکنواخت‌تر و ملایم‌تر می‌کند) همچنین دارای یک عبارت تطبیق هیستوگرام<sup>۱۵</sup> است که به ازای غیرمشابه بودن<sup>۱۶</sup> هیستوگرام‌های پیش‌مینه، جریمه‌ای<sup>۱۷</sup> در نظر می‌گیرد.

<sup>9</sup> Supervision Degree

<sup>10</sup> Unsupervised Co-segmentation

<sup>11</sup> Interactive Co-segmentation

<sup>12</sup> Supervised

<sup>13</sup> Binary Annotation

<sup>14</sup> Smoothness

<sup>15</sup> histogram Matching

### ۱-۳-کاربردهای هم‌بخش‌بندی تصویر

بخش‌بندی شی، همواره در بینایی ماشین یک امر بسیار مهم و قابل ملاحظه بوده است. چون بسیاری از مسائل سطح بالای این زمینه مانند ردیابی شی [۹]، شناسایی شی [۱۰] و درک صحنه [۱۱] را به راحتی می‌توان با این الگوریتم‌ها حل نمود. در ادامه به چند نمونه از کاربردهای آن اشاره می‌کنیم.

#### ۱-۳-۱-دسته‌بندی و برچسب‌گذاری تصاویر

دسته‌بندی<sup>۱۸</sup> و برچسب‌زنی<sup>۱۹</sup> تصویر دو وظیفه<sup>۲۰</sup> مهم در بینایی ماشین می‌باشند. هدف از دسته‌بندی تصویر توصیف تصویر به صورت سراسری<sup>۲۱</sup> با یک برچسب<sup>۲۲</sup> توصیف‌کننده می‌باشد (برای نمونه، ساحل<sup>۲۳</sup>، فضای باز<sup>۲۴</sup>، درون شهر<sup>۲۵</sup> و غیره) در حالی که برچسب‌زنی تصویر بر برچسب-گذاری<sup>۲۶</sup> بر محتوا<sup>۲۷</sup> محلی<sup>۲۸</sup> یک تصویر تاکید دارد (برای نمونه آیا یک تصویر شامل، آسمان، ماشین، درخت و غیر می‌باشد). از آنجایی که این دو مسئله به هم مرتبط می‌باشند تلاش برای حل مشترک این مسائل طبیعی می‌باشد. برای نمونه، یک تصویر که به عنوان خیابان برچسب‌گذاری شده، احتمال توصیف تصویر با حضور "ماشین"، "پیاده" یا "ساختمان" در آن بیشتر از "ساحل" یا "آب دریا" می‌باشد.

---

<sup>16</sup> Dissimilarity

<sup>17</sup> Penalty

<sup>18</sup> Classification

<sup>19</sup> Annotation

<sup>20</sup> Tasks

<sup>21</sup> Globally

<sup>22</sup> Single

<sup>23</sup> Coast

<sup>24</sup> Outdoor

<sup>25</sup> Inside city

<sup>26</sup> Tagging

<sup>27</sup> Content

<sup>28</sup> Local

### ۱-۳-۲- بازیابی تصاویر مربوط به یک یا چند شی خاص از یک پایگاه داده

از آنجایی که هم‌بخش‌بندی نیاز به آموزش ندارد یک کاربر به راحتی می‌تواند با مشخص کردن شی مورد نظر خود در یک تصویر، در کمترین زمان اشیاء مورد نظر خود را در سایر تصاویر پایگاه داده پیدا کند و آنها را مورد استفاده قرار دهد.

### ۱-۴- تعریف مسئله

در پژوهش انجام شده هدف ما پیدا کردن اشیای مشترک در یک مجموعه تصاویر از یک پایگاه داده می‌باشد که در اصطلاح ما به آن هم‌بخش‌بندی چند پیش‌زمینه‌ای<sup>۲۹</sup> می‌گوییم. هم‌بخش‌بندی چند پیش‌زمینه‌ای، به وظیفه‌ای از بخش‌بندی اشتراکی  $K$  پیش‌زمینه مشترک  $F = \{f^1, \dots, f^k\}$  (مجموعه کلیه پیش‌زمینه‌های مورد نظر ما می‌باشد) از  $m$  تصویر ورودی، اشاره دارد که هر کدام از آنها شامل یک زیر مجموعه ناشناخته مختلف از  $K$  پیش‌زمینه است. هر کاربر ممکن است پیش‌زمینه‌های مورد نظر خود را با توجه به علاقه و نیاز خود، در یک جفت تصویر ورودی، از روی محدوده‌ها و یا تفسیرهای وابسته به پیکسل مشخص کند.

اکثر روش‌های هم‌بخش‌بندی تصویر، برای استفاده روی مجموعه تصاویر کاربران عمومی دچار مشکل هستند. یکی از بزرگترین مشکلات این روش، نیاز به وجود داشتن تمام پیش‌زمینه‌ها در تصویر ورودی است (پیش‌زمینه مورد نظر ما باید در تمام مجموعه تصاویر ورودی وجود داشته باشد) تا این الگوریتم‌ها بتوانند پاسخ مطلوبی داشته باشند.



شکل ۱-۱ نمونه ای از تصویر پایگاه داده با برجسب کودک و سبد سیب

شکل ۱-۱ یک مثال را نشان می‌دهد که این مشکل را به تصویر کشیده است. تعداد اشیا موجود در جریان یک عکس محدود است اما این اشیا در هر عکس وجود ندارند. برای مثال در شکل ۱، دو دختر و یک کودک و یک سبد سیب، در جریان عکس‌ها موجود تکرار شده‌اند اما هر عکس فقط شامل یک زیرمجموعه ناشناخته از این پیش‌زمینه‌ها است. این مشکل برای اکثر الگوریتم‌های هم‌بخش‌بندی که در حال حاضر وجود دارند قابل قبول نیست. چون آنها بر این باورند که اشیا ی مورد نظر کاربر در تمام تصاویر وجود دارند. برای استفاده از الگوریتم‌هایی که مجموعه تصاویری با این خصوصیات را استفاده می‌کنند اگر نمونه‌هایی از پیش‌زمینه‌ها را که به طور نامنظم در سراسر تصویر رخ می‌دهند در نظر بگیریم حتماً باید یک کاربر حضور داشته باشد تا تصاویری که دارای پیش‌زمینه‌های مشترک هستند را شناسایی و دسته‌بندی کند این مشکل واقعاً وقت‌گیر و دشوار است بخصوص در زمانی که ما یک پایگاه داده با تعداد زیادی تصویر در اختیار داشته باشیم.

در این تحقیق از روش بهینه سازی ترکیبی برای هم‌بخش‌بندی تصویر استفاده می‌شده است که با محدودیت ذکر شده در بالا، دچار مشکل نمی‌شود. این الگوریتم اجازه می‌دهد که چند پیش‌زمینه با چینش و جزئیات مختلف در هر عکس ظاهر شوند و محدودیتی برای آن ایجاد نمی‌کند و مستقیماً آنها را هم‌بخش‌بندی می‌کند.

## ۱-۵- هدف و نوآوری پایان نامه

در این پژوهش از روشی مبتنی بر گراف استفاده کرده ایم تا آسان تر و سریع تر بتوان تخصیص ناحیه‌های مربوط به یک پیش‌زمینه را انجام داد. همچنین با استفاده از این روش سعی داریم تعداد حالات به وجود آمده برای هر پیش‌زمینه را کاهش دهیم. همچنین الگوریتم ارائه شده اجازه می‌دهد که هر تعداد پیش‌زمینه که کاربر انتظار دارد استخراج شوند. این موضوع در تضاد با مشکل هم‌بخش‌بندی کلاسیک که توسط بیشتر الگوریتم‌های موجود ارائه شده است می‌باشد چون فرضشان ساده‌تر است. در هم‌بخش‌بندی کلاسیک تمامی پیش‌زمینه‌های مورد نظر ما، باید در هر تصویر وجود داشته باشند تا الگوریتم دچار خطا نشود. پژوهش انجام شده یک روش بهینه برای هم-بخش‌بندی با چند پیش‌زمینه ارائه می‌دهد که هیچ فرضی روی پیش‌زمینه نمی‌سازد و محدودیت‌های موجود در روش‌های قبل را که باعث تحمل خطا می‌شد را ندارد.

## ۱-۶- ساختار پایان‌نامه

این پایان‌نامه در ۵ فصل تنظیم شده است. در فصل دوم مروری بر کارهای انجام شده در حوزه هم‌بخش‌بندی تصویر و تشخیص اشیاء خواهیم داشت. در این فصل تکنیک‌های شناسایی شیء، انتخاب ناحیه‌های کاندید و تشکیل ناحیه‌های کاندید مورد بررسی قرار خواهد گرفت و چند روش کار انجام شده در این زمینه نیز مورد بررسی قرار می‌گیرد. در فصل سوم روش پیشنهادی خود را ارائه خواهیم کرد. و در فصل ۴ به ارزیابی روش پیشنهادی و نتایج آن می‌پردازیم. و نهایتاً در فصل ۵ نتیجه‌گیری و کارهای پیشنهادی که در آینده انجام خواهد شد را مورد بررسی قرار می‌دهیم.



فصل دوم:

# مروری بر کارهای انجام شده

امروزه حجم بالای تصاویر اینترنتی وجود دارند که اغلب این تصاویر دارای وجه مشترکی از نظر جزئیات و محتوا می‌باشند. از آنجایی که اشیا باهم در یک مجموعه تصاویر مرتبط رخ می‌دهند بهتر و قابل قبول‌تر است که بخشبندی تصاویر دارای اشیا و پیش‌زمینه‌های یکسان را بصورت اشتراکی انجام دهیم تا اینکه این بخشبندی به صورت مجزا انجام گیرد چون بخشبندی به صورت اشتراکی باعث افزایش قدرت تشخیص پس‌زمینه شده و آسان‌تر می‌توان اشیا پیش‌زمینه را از پس‌زمینه تمیز داد. این راه کار جدید به طور جدی در زمینه بینایی ماشین مورد مطالعه قرار گرفته است [۳-۴-۶].

اکثر الگوریتم‌هایی که در زمینه هم‌بخشبندی تصویر ارائه شده‌اند همواره در حال بهینه کردن دو زیر مسئله می‌باشند. مسئله اول پیدا کردن ناحیه‌هایی از تصویر است که قابلیت شی بودن را دارا باشند این عمل را در اصطلاح مدل کردن پیش‌زمینه می‌گوییم. در سال‌های اخیر الگوریتم‌های زیادی در این راستا ارائه شده است که الگوریتم‌های ارائه دهنده ناحیه‌های کاندید<sup>۳۰</sup> نامیده می‌شوند. مسئله بعدی تخصیص هر یک از ناحیه‌های پیشنهادی به یکی از پیش‌زمینه‌های مورد نظر می‌باشد ما این زیر مسئله را با نام تخصیص ناحیه<sup>۳۱</sup> معرفی می‌کنیم.

الگوریتم ارائه شده توسط این پژوهش نیز شامل این دو زیرمسئله می‌باشد و مکرراً از این دو زیر مسئله استفاده کرده تا به جواب بهینه برسد. در ادامه روش‌های مورد استفاده در الگوریتم‌های گوناگون را در هر یک از مباحث پیشنهاد ناحیه‌های کاندید و تخصیص ناحیه مورد بررسی قرار می‌دهیم.

---

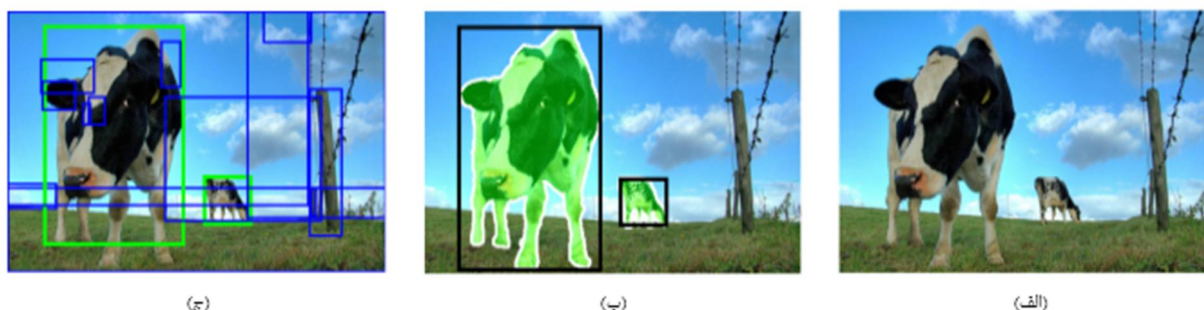
<sup>30</sup> Region Proposal

<sup>31</sup> Region Assignment



## ۲-۲- مدل کردن و تشخیص شی (پیش‌زمینه)

سیستم‌های تشخیص اشیاء که در سال‌های اخیر پیشنهاد شده‌اند، به الگوریتم‌های ناحیه‌های کاندید<sup>۳۲</sup> برای تخمین مکان اشیاء بسیار وابسته‌اند. از مهمترین شبکه‌های تشخیص می‌توان به SPPnet [۱۲] و R-CNN<sup>۳۳</sup> سریع [۱۳] اشاره نمود که در سال‌های اخیر، مورد توجه بسیاری از محققان و کاربرهای صنعتی قرار گرفته است. در این روش‌ها نمی‌توان سراسر تصویر را به دلیل محدودیت‌های زمانی و حالت‌های مختلف از اشیاء جستجو کرد. در همین راستا، ابتدا باید ناحیه‌های کاندید وجود شی استخراج شوند. پیشرفت‌های انجام شده سبب بهبود زمان اجرای شبکه‌های تشخیص شده است و همچنین نشان داده شده تعیین نواحی کاندید مهمترین تنگنای<sup>۳۴</sup> محاسباتی این گونه روش‌ها می‌باشد. نمونه‌ای از ناحیه‌های کاندید جهت بررسی وجود شی مورد نظر در شکل ۲-۱ نشان داده شده است. همانطور که در شکل نشان داده شده، سیستم تشخیص شی، انتخاب نهایی خود را از ناحیه‌های کاندید انجام می‌دهد.



شکل ۲-۱ نمونه‌ای از ناحیه کاندید. (الف) تصویر اصلی، (ب) اشیاء موجود در تصویر، (ج) ناحیه‌های کاندید [۱۷]

روش‌های نواحی کاندید به طور معمول بر ویژگی‌های ارزان و سریع تکیه می‌کنند. جستجوی انتخابی [۱۴] یکی از معروف‌ترین روش‌های پیشنهاد ناحیه است که به صورت حریصانه ابرپیکسل-هایی<sup>۳۵</sup> را که بر اساس ویژگی‌های سطح پایین (به مانند میزان شدت روشنایی و رنگ در فضاها)

32 Region proposal

33 Region-convolutional neural network

34 bottleneck

35 Superpixels به نواحی اتمیک به هم پیوسته و دارای مفهوم واحد در تصویر گفته می‌شود

رنگی متفاوت) طراحی شده ادغام می‌کند. اما این روش در مقایسه با شبکه‌های شناسایی کارآمد به مانند R-CNN سریع‌تر می‌باشد.

همانطور که اشاره شد، بهبود عملکرد سیستم‌های تشخیص و شناسایی اشیاء وابستگی زیادی به روش‌های ناحیه‌کاندید دارند. در همین راستا، در اینجا به کارهای انجام شده در این زمینه می‌پردازیم و روش‌های موجود را معرفی و مقایسه می‌کنیم.

## ۲-۲-۱- روش‌های مبتنی بر گروه‌بندی

روش‌های مبتنی بر گروه‌بندی برای استخراج ناحیه‌های کاندید در سه گروه و بر اساس چگونگی تولید کاندید، دسته‌بندی می‌شوند. به طور کلی روش‌های تولید کاندید بر پایه گروه‌بندی، به گروه‌بندی ابرپیکسل (SP) [۱۵]، حل کردن مسائل برش گراف چندگانه<sup>۳۶</sup> (GC) با seeds<sup>۳۷</sup>های متنوع [۱۶] یا به طور مستقیم از کانتورهای لبه<sup>۳۸</sup> (EC) [۲۳-۲۴] تقسیم می‌شوند. از مهمترین روش‌های موجود در این حوزه می‌توان به موارد زیر اشاره نمود:

- جستجو انتخابی [۱۴]: در این روش از ادغام ابرپیکسل‌ها، برای تولید کاندیدها استفاده می‌شود. این روش، پارامترهای یادگیری ندارد و از ویژگی‌ها و تابع‌های شباهت برای ادغام ابرپیکسل‌ها استفاده می‌کند. در سال‌های اخیر این روش، به صورت گسترده به عنوان روش کاندید برای بسیاری از آشکارسازهای شی مورد استفاده قرار گرفته شده است.

در مقاله [۱۴] جستجوی انتخابی با ترکیب استراتژی جستجو جامع<sup>۳۹</sup> و قطعه‌بندی ارائه شده است و به مانند قطعه‌بندی، برای نمونه‌گیری از ساختار تصویر استفاده شده است. همچنین از جستجو جامع برای محدود کردن تمامی موقعیت‌های ممکن از اشیاء استفاده شده است. نتایج بدست آمده این روش، بر روی مجموعه کوچکی، که وابسته به داده و مستقل از کلاس بوده، بیشترین دقت در مکان-

---

36 Multiple Graph Cut

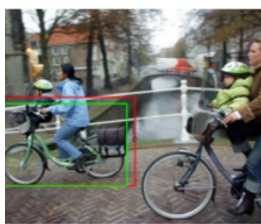
37 Seeds: Superpixels Sxtracted Via Energy-Driven Sampling

38 Edge contours

39 Exhaustive

یابی را نشان داده است. آزمایش های انجام شده دقتی برابر با ۹۹٪ و متوسط میانگین بهترین هم-پوشانی بین نواحی پیشنهادی و نواحی صحیح ۰.۸۷۹ در ۱۰۰۹۷ موقعیت را از خود نشان داده است. در این روش از یک الگوریتم گروه بندی سلسله مراتبی استفاده شده است. ناحیه در مقایسه با پیکسل اطلاعات بیشتری دارد، به همین دلیل، در این مقاله از ویژگی ها بر پایه ناحیه بیشتر استفاده شده است. برای انتخاب مجموعه ای از ناحیه های هم پوشان با اشیاء، از روش سریع [۱۵] استفاده شده است. الگوریتم کلی در شکل ۲-۱ آورده شده است. همانطور که اشاره شد، از روش ارائه شده در مرجع [۱۵] برای تولید ناحیه های اولیه استفاده شده

است. سپس از یک الگوریتم حرصانه برای گروه بندی ناحیه ها با هم استفاده می شود. در این الگوریتم، ابتدا شباهت بین همه ی ناحیه های همسایه محاسبه می شود و در گام بعدی، ناحیه هایی با بیشترین شباهت، باهم ترکیب می شوند. این کار تا زمانی ادامه می یابد که همه ی تصویر به یک ناحیه واحد تبدیل شوند. در این مقاله اشاره شد که ناحیه های تولید شده می توانند در صحنه ها و شرایط روشنایی متفاوت باشند در همین راستا الگوریتم گروه بندی در فضاهای رنگی متفاوت با ویژگی های مستقل و متفاوت اجرا شده است. این ویژگی ها شامل، رنگ در فضاهای رنگی RGB،



Bike: 0.863



Person: 0.882

شکل ۲-۴ نمونه ای از خروجی الگوریتم [۱۴]. محدوده

قرمز ناحیه کاندید روش جستجو انتخابی و محدوده سبز

ناحیه هدف است

#### Algorithm 1: Hierarchical Grouping Algorithm

Input: (colour) image

Output: Set of object location hypotheses  $L$

Obtain initial regions  $R = \{r_1, \dots, r_n\}$  using [9]

Initialise similarity set  $S = \emptyset$

While Neighbouring region pair  $(r_i, r_j)$  do

    Calculate similarity  $s(r_i, r_j)$

$S = S \cup s(r_i, r_j)$

While  $S \neq \emptyset$  do

    Get highest similarity  $s(r_i, r_j) = \max(S)$

    Merge corresponding regions  $r_t = r_i \cup r_j$

    Remove similarities regarding  $r_i: S = S \setminus s(r_i, r_*)$

    Remove similarities regarding  $r_j: S = S \setminus s(r_*, r_j)$

    Calculate similarity set  $S_t$  between  $r_t$  and its neighbours

$S = S \cup S_t$

$R = R \cup r_t$

Extract object location boxes  $L$  from all regions in  $R$

شکل ۲-۳ الگوریتم گروه بندی سلسله مراتبی در جستجو

انتخابی [۱۴]

شدت روشنایی<sup>۴۰</sup> سطح خاکستری (I) در تصاویر، کانال رنگی rg به همراه I, HSV, RGB نرمال شده، در H و HSV و C می‌باشند. معیار شباهت استفاده شده وابسته به چهار پارامتر رنگ، بافت، اندازه و همپوشانی ناحیه‌ها با هم می‌باشد.

شکل ۲-۳ نمونه‌ای از نتایج بدست آمده از روش جستجو انتخابی را نشان می‌دهد. در هر تصویر محدوده قرمز ناحیه کاندید و محدوده سبز مربوط به ناحیه هدف است. مقداری که زیر هر عکس ذکر شده، میزان همپوشانی این دو محدوده را نشان می‌دهد.

• پریم تصادفی<sup>۴۱</sup>: در مرجع [۱۷] از شباهت ویژگی‌ها به مانند روش جستجوی انتخابی، استفاده شده است. در این روش یک فرایند ادغام جدید به صورت تصادفی برای ابرپیکسل‌ها معرفی شده است. در این فرایند ادغام همه‌ی احتمالات ممکن آموزش داده شده است. با این روش سرعت به صورت قابل ملاحظه‌ای افزایش پیدا کرده است.

• جستجو محلی و سراسری: در مرجع [۱۸] یک استراتژی مشابه جستجو انتخابی، برای ادغام ابرپیکسل‌ها پیشنهاد شده است. ولی از ویژگی‌های مختلف و بیشتری استفاده کرده است. در مرحله بعدی، قطعات<sup>۴۲</sup> تولید شده به عنوان Seeds بکار گرفته شده و برای حل کردن برش گراف از CPMC<sup>۴۳</sup> برای تولید کاندیدهای بیشتر استفاده شده است.

• چانگ [۱۹] ترکیبی از برجستگی<sup>۴۴</sup> و شی‌بودن با مدل گرافیکی، برای ادغام ابرپیکسل در قطعه-بندی تصویر/پس‌زمینه معرفی کرده است. مراحل کلی از مدل گرافی پیشنهادی در شکل ۲-۴ نشان داده شده است. این روش با نمونه‌گیری تصادفی از پنجره‌های زیاد شروع می‌شود. به هر پنجره مقدار شی‌بودن و برای هر پیکسل یا ابرپیکسل مقدار برجستگی اختصاص داده می‌شود. برای ارتباط شی-

---

40 intensity

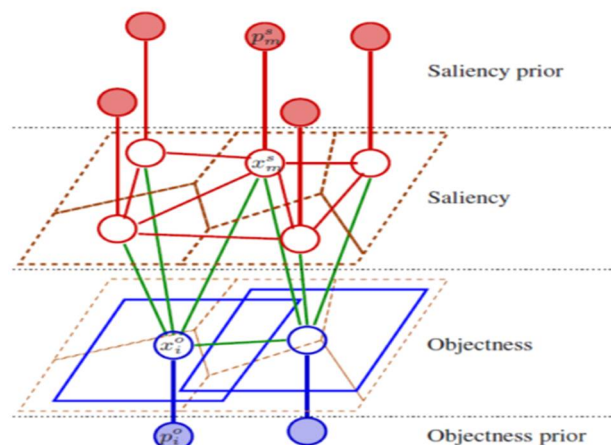
41 RandomizedPrim

42 segments

43 Constrained Parametric Min-Cuts

44 Saliency

بودن و برجستگی، مقداری به عنوان سطح برجستگی شی<sup>۴۵</sup> برای هر پنجره معرفی شده است. این مقدار برای نمایش برجستگی جسم زمینه در هر پنجره استفاده می‌شود. از شی بودن برای تخمین برجستگی استفاده شده است. همچنین از برجستگی برای تخمین شی بودن استفاده شده است. مقدار شی بودن هر پنجره در صورتی بالا خواهد بود که



شکل ۲-۵ مراحل کلی الگوریتم چانگ [۱۹] شامل نمونه‌گیری تصافی جدید و انتصاب مقدار برجستگی به هر پنجره، تخمین شی بودن می‌باشد

برجستگی سطح شی به خوبی بتواند بسیاری از مقادیر برجستگی پیکسل‌های را توضیح دهد. در این مقاله، برجستگی سطح شی توسط اطلاعات به دست آمده از ابرپیکسل‌ها با یک معیار اندازه‌گیری جدید محاسبه شده است.

• قطعه‌بندی خودکار با برش‌های کمینه<sup>۴۶</sup> (CPMC) [۲۰] از قطعه‌بندی اولیه اجتناب می‌کند و برش‌های گراف را با Seedهای مختلف مستقیماً بر روی پیکسل محاسبه می‌کند. نتایج قطعه‌بندی با حجم بزرگی از ویژگی‌ها رتبه‌بندی می‌شوند. در مقاله [۲۰] یک روش جدید برای تولید و رتبه‌بندی اشیاء قابل قبول در تصویر با فرآیند پایین به بالا<sup>۴۷</sup> و نشانه‌های<sup>۴۸</sup> سطح متوسط ارائه شده است. منظور

45 Object-level saliency

46 Cpmc: Automatic object segmentation using constrained parametric min-cuts

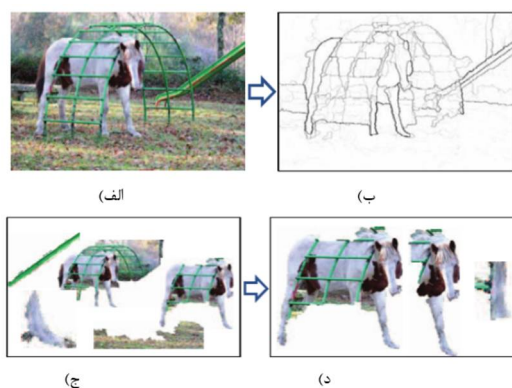
47 Down-Top

48 cues



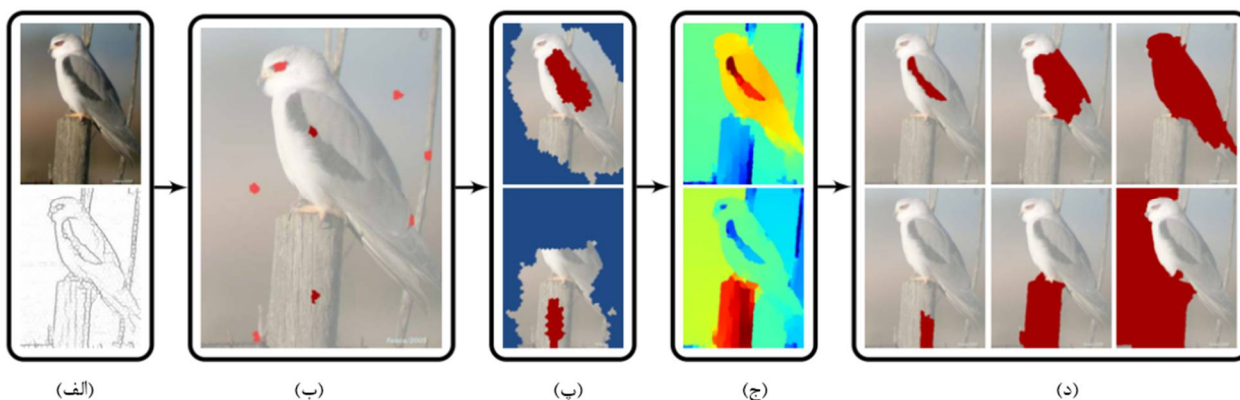
امتداد مرز تصویر، جز Seed منفی می‌شوند. از الگوریتم‌ها برپایه گراف به منظور قطعه‌بندی این Seedها استفاده شده است. همچنین میزان شباهت بین پیکسل‌های همسایه به عنوان وزن گراف در نظر گرفته شده است. در ادامه یک تابع انرژی تعریف می‌شود. با این تعریف، برچسب‌های اختصاص داده شده به پیکسل‌ها کاهش داده می‌شوند و اشیاء مورد نظر استخراج می‌شوند. شکل ۲-۵ روال کلی این روش را نشان می‌دهد.

• در مراجع [۲۱] یک قطعه‌بندی سلسله‌مراتبی از مرزهای انسداد<sup>۵۰</sup> و حل برش گراف با Seedها و پارامترهای متفاوت، برای تولید قطعات استفاده شده است. کاندیدها بر اساس طیف گسترده‌ای از نشانه‌ها رتبه‌بندی شده است. این عمل در واقع سبب تشویق تنوع انتخابات می‌شود.



شکل ۲-۷ مراحل مختلف از روش [۲۱]، (الف) تصویر ورودی، (ب) قطعه‌بندی سلسله‌مراتبی، (ج) ناحیه‌های کاندید، (د)

رتبه‌بندی ناحیه‌ها



شکل ۲-۸ مراحل کلی از روش **Geodesic** مطرح شده در مرجع [۲۲] (الف) تصویر ورودی و بیش‌قطعه‌بندی آن به همراه نگاشت احتمال مرز، (ب) مکان **Seeds**، (پ) ماسک پیش‌زمینه و پس‌زمینه تولید شده برای دو **Seeds**، (ج) استفاده از تبدیل فاصله **Geodesic (SGDB)** برای این ماسک‌ها، (د) کاندیدهای پیشنهادی با توجه به محاسبه

مجموعه‌های سطوح بحرانی در هر **SGDB**

مراحل کلی پیشنهاد شده در مقاله [۲۱] در شکل ۲-۶ نشان داده شده است. این قطعه تولید شده، ناحیه‌های **Seed** تصویر را به وجود می‌آوردند. در گام بعدی، با استفاده از این **Seed**ها و ویژگی‌های مانند مرز، رنگ و بافت، مجموعه متنوعی از نواحی تولید می‌شود. این نواحی، می‌توانند به عنوان قطعات شی معرفی شوند. در نهایت با روش یادگیری ساخت‌یافته، ناحیه‌های بدست آمده رتبه‌بندی می‌شوند که به احتمال زیاد مناطق با امتیاز بالاتر مربوط به اشیاء مختلف هستند.

• روش  $Geodesic^{51}$  [۲۲] توسط یک روش قطع‌بندی تصویر پیشنهادی شروع می‌شود. طبقه‌بندها از مکان **Seeds** برای تبدیل فاصله  $Geodesic^{52}$  استفاده می‌کنند. در این مقاله، **Seeds** توسط طبقه‌بند بهینه برای کشف اشیاء، آموزش داده شده است.

ایده اصلی در این مقاله، معرفی مجموعه‌های سطوح بحرانی  $^{53}$  در تبدیل فاصله **Geodesic** است. تبدیل فاصله **Geodesic** برای مکان‌های **Seeds** در تصویر محاسبه شده است. تبدیل فاصله، در

منظور از فاصله **Geodesic**، اندازه کوتاه‌ترین مسیر بین نودها می‌باشد 51

، معیاری برای اندازه‌گیری کوتاه‌ترین مسیر بین مجموعه‌ای از نودها تا هر نود منظور از تبدیل فاصله 52

53 Critical level sets



زمان نزدیک برخط<sup>۵۴</sup> محاسبه می‌شود. همچنین محاسبه هر تبدیل برای تولید کاندیدها در مقیاس-های متفاوت استفاده شده است. از این رو، استفاده از خطلوله<sup>۵۵</sup> در محاسبات، می‌تواند بسیار موثر باشد.

همانطور که در شکل ۲-۷ نشان داده شده، در این روش ابرپیکسل‌های تصاویر ورودی با قطعه‌بندی محاسبه می‌شوند و نگاشت احتمال مرز<sup>۵۶</sup> به عنوان وزن گراف در نظر گرفته می‌شود. در واقع در این گراف هر نود به یک ابرپیکسل مربوط بوده و هر یال اتصال‌های بین ابرپیکسل‌های مجاور است. وزن یال‌ها میزان درست‌نمایی<sup>۵۷</sup> مرز اشیاء با توجه به لبه‌های تصویر است.

فاصله Geodestic بین دو نود، اندازه کوتاه‌ترین مسیر بین نودهای موجود است. تبدیل فاصله Geodestic معیاری برای اندازه‌گیری کوتاه‌ترین مسیر بین مجموعه‌ای از نودها تا هر نود است. این معیار را می‌توان توسط الگوریتم دایجسترا در زمان  $O(n \log n)$  محاسبه کرد. در این الگوریتم  $n$  تعداد ابرپیکسل‌ها است.



شکل ۲-۹ اشیاء کاندید (نواحی قرمز) توسط الگوریتم [۲۲]

---

54 online  
 55 Pipeline  
 56 Boundary probability map  
 57 likelihood

## ۲-۲-۲- روش‌های امتیازدادن به پنجره کاندید

یک روش جایگزین برای تولید کاندیدها، امتیازدهی هر پنجره کاندید است. در واقع هر تصویر پنجره-گذاری شده و در گام بعدی به هر پنجره امتیازدهی بر اساس احتمال شی بودن پنجره کاندید صورت می‌گیرد. در مقایسه با روش‌های گروه‌بندی، این روش‌ها معمولا فقط مرز محدوده‌ها را برگشت می‌دهند و سریع‌تر هستند. در مواردی که نمونه‌گیری پنجره با تعداد بالا باشد، این روش‌ها معمولا پیشنهادهای با دقت محلی‌سازی پایین تولید می‌کنند. در ادامه به معرفی بعضی از این روش‌ها می‌پردازیم.

• شی‌بودن [۲۳] یکی از اولین و بهترین روش‌های شناخته شده در بحث روش‌های کاندید است. یک مجموعه اولیه از کاندیدها از محل‌های برجسته در تصویر انتخاب می‌شود. این کاندیدها بر اساس چندین نشانه به مانند رنگ، لبه‌ها، محل، اندازه و استحکام "Supapixel stradding(SS)" امتیازدهی می‌شوند. برای بررسی استحکام SS معیاری معرفی شده که میزان اشتراک بین ابرپیکسل و پنجره پیشنهادی را محاسبه می‌کند.

• Rahtu [۲۴] با مخزن بزرگی از ناحیه‌های کاندید تولید شده توسط ابرپیکسل منحصر به فرد، جفت و سه‌تایی و چندین محدوده انتخاب شده، به صورت تصادفی شروع می‌شود. استراتژی امتیازدهی بازبینی شده و بهبودهای در آن ارائه شده است.

• Bing [۲۵] از یک طبقه‌بند خطی ساده با ویژگی‌های لبه استفاده می‌کند. در این روش، از پنجره لغزان استفاده شده است. سرعت تخمین در این روش بسیار بالا است. در این روش، طبقه‌بند تاثیر اندکی دارد و می‌تواند عملکردی مشابه با زمانی که به تصویر نگاه نمی‌شود را داشته باشد. این روش‌های مستقل از تصویر، با نام CrackingBing معروف هستند.

• EdgeBoxes [۲۶] همچنین از یک الگو پنجره لغزان بزرگ<sup>۵۸</sup> شروع می‌شود. اما بر اساس تخمین مرز اشیاء، که توسط ساختار جنگل تصمیم [۱۷] بدست آمده، ساخته شده است و یک گام پالایش<sup>۵۹</sup> بعدی را برای افزایش محلی سازی اضافه می‌کند. در این روش، پارامتری یاد گرفته نمی‌شود.

## ۲-۲-۳- روش‌های پایه یا مرجع<sup>۶۰</sup> در پیشنهاد کاندید

بعضی از روش‌ها به عنوان روش پایه و به عنوان نقاط مرجع معرفی می‌شوند. تمامی این روش‌ها در قسمت‌های قبل مورد ارزیابی قرار گرفته شده‌اند. از روش‌های زیر به عنوان روش پایه استفاده می‌شود.

• روش گاوسی<sup>۶۱</sup>: به همین ترتیب، یک توزیع گوسین چند متغیره برای موقعیت مرکز محدوده مرزبندی، ناحیه ریشه مربع و لگاریتم نسبت ابعاد تخمین زده می‌شود. در گام بعدی مقدار میانگین و واریانس در مجموعه آموزشی محاسبه می‌شود. در نهایت کاندیدها از این توزیع‌ها نمونه برداری شده‌اند.

• روش پنجره لغزان: در این روش پنجره در یک شبکه (گرید) منظم مکان‌یابی می‌شود که برای آشکارسازی شی در پنجره‌ی لغزان رایج است. تعداد درخواست کاندیدها بر اساس اندازه پنجره توزیع شده است. برای هر اندازه پنجره به صورت یکنواخت پنجره‌ها مکان‌یابی شده‌اند. این روش از پیاده‌سازی مراجع [۲۵] الهام گرفته شده است.

• روش ابرپیکسل: همانطور که در قسمت‌های قبل بیان شد، ابرپیکسل تاثیر مهمی بر روی رفتار روش‌های پیشنهادی دارند. با توجه به تعریف مرجع [۲۷]، ابرپیکسل به نواحی اتمیک<sup>۶۲</sup> متراکم یا به هم پیوسته<sup>۶۳</sup> و ادراکی<sup>۶۴</sup> معنادار<sup>۶۵</sup> در تصویر گفته می‌شود.

---

58 coarse  
59 refinement  
60 Baseline  
61 Gaussian  
62 atomic  
63 compact  
64 perceptually

## ۲-۲-۴- کنترل تعداد کاندیدهای پیشنهادی

برای مقایسه جامع بین روش‌های مطرح شده باید تعداد کاندیدهای ارائه شده در هر تصویر توسط روش‌هایی کنترل شوند، زیرا برای مقایسه صحیح بین روش‌های موجود باید تعداد کاندیدهای ارائه شده یکسان باشد. در بسیاری از موارد، روش‌ها با ارائه‌ی تعداد کاندیدها بین ۱۰۰ تا ۱۰۰۰۰۰ مورد ارزیابی قرار می‌گیرند. برخی از روش‌ها به کاندیدهای ارائه شده امتیاز می‌دهند و در مقابل بعضی دیگر این توانایی را ندارند. البته نمی‌توان در همه‌ی روش‌ها به صورت صریح و روشن تعداد کاندیدها را کنترل نمود. در روش‌هایی که امتیاز کاندید را ارائه می‌دهند،  $k$  برترین ناحیه کاندید انتخاب می‌شوند. ولی در برخی از روش‌ها، نمی‌توان کنترل مستقیم بر روی تعداد کاندیدها داشت و همچنین امتیاز کاندیدها نیز مشخص نمی‌شوند. در این روش‌ها می‌توان با کنترل غیر مستقیم به  $k$  کاندید دسترسی داشت. در این راستا نیاز است که پارامترهای دیگر را تغییر داد تا به این کنترل، دسترسی پیدا نمود. به همین جهت تعداد کاندیدهای تولید شده با پارامترهای متفاوت در زیر مجموعه‌ای از تصویر ذخیره می‌شود. در نهایت با استفاده از یک درونیابی خطی بین تنظیم‌های پارامترها،  $k$  کاندید انتخاب می‌شوند. در بعضی از روش‌ها که هیچ کنترلی بر روی تعداد کاندیدها امکان‌پذیر نمی‌باشد  $k$  نمونه به صورت تصادفی انتخاب می‌شود.

بعضی از روش‌ها کاندیدهای تکراری تولید می‌کنند که آن‌ها را باید از مجموعه انتخابی در هر روش پاک نمود. مقایسه بین روش‌های ارائه شده از لحاظ کنترل پذیر بودن تعداد کاندیدها در جدول ۲-۱ نشان داده شده است. همچنین در این جدول مقایسه‌ای بین روش‌ها از لحاظ نوع روش و خروجی حاصل صورت گرفته است.

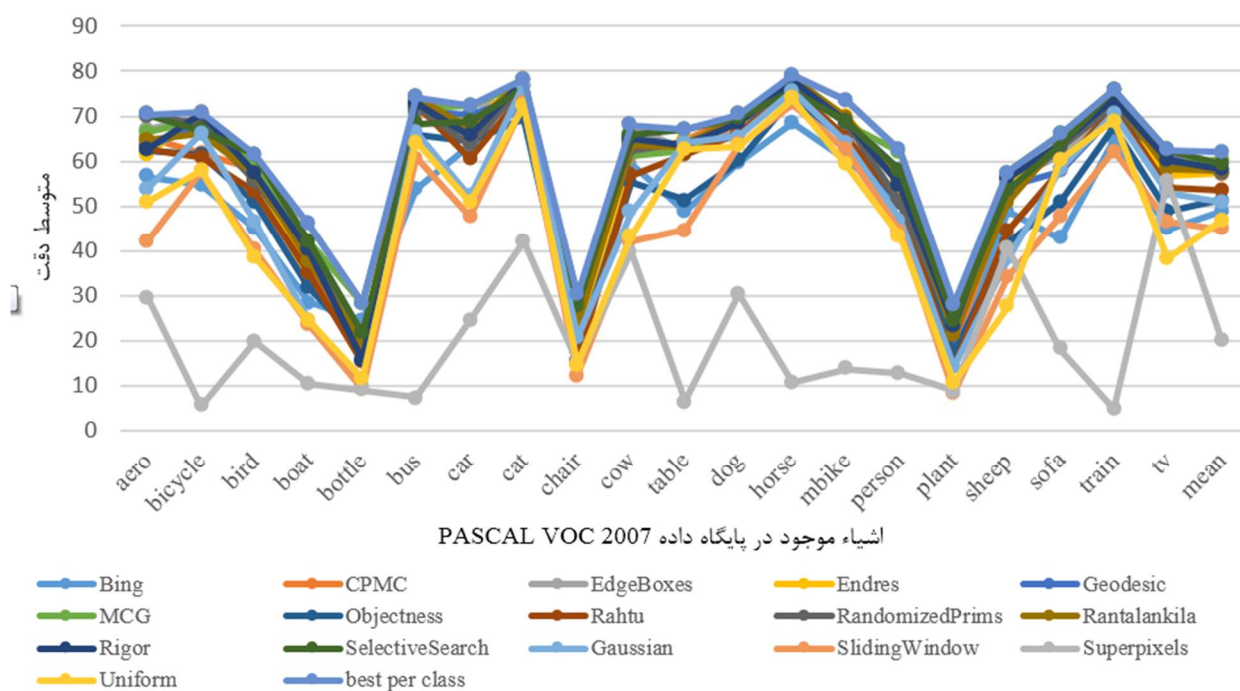
جدول ۱-۲ مقایسه‌ای بین روش‌های معرفی شده. علامت +/- به معنی، دارا بودن یا نبودن خاصیت مشخص است و

علامت \* در کنترل تعداد کاندید، نشان دهنده‌ی کنترل غیر مستقیم است.

کنترل	خروجی امتیاز	خروجی قطعات	نوع روش	روش
+	+	-	پنجره لغزان	<b>Bing</b>
*	+	+	گروه‌بندی	<b>CPMC</b>
+	+	-	پنجره لغزان	<b>EdgeBox</b>
+	+	+	گروه‌بندی	<b>Endres</b>
*	-	+	گروه‌بندی	<b>Geodesic</b>
*	+	+	گروه‌بندی	<b>MCG</b>
+	+	-	پنجره لغزان	<b>Objectness</b>
+	+	-	پنجره لغزان	<b>Rahtu</b>
+	-	+	گروه‌بندی	<b>RandPrim</b>
*	-	+	گروه‌بندی	<b>Rantalankila</b>
*	-	+	گروه‌بندی	<b>Rigor</b>
*	+	+	گروه‌بندی	<b>SelectiveSearch</b>
+	-	-	-	<b>Gussian</b>
+	-	-	-	<b>SlidingWindow</b>
-	-	+	-	<b>Superpixels</b>
+	-	-	-	<b>Uniform</b>

## ۲-۲-۵- مقایسه بین روش‌های نواحی کاندید

همانطور که در قسمت مقدمه بیان شد، یکی از چالش‌های اساسی و تاثیر گذار در بحث شناسایی اشیاء، پیشنهاد ناحیه‌های کاندید بوده است. از طرفی یکی از روش‌های موفق و مشهور مطرح در زمینه شناسایی، روش R-CNN سریع بوده که توسط دانشگاه MIT پیاده سازی شده است. در همین راستا به بررسی تاثیر روش پیشنهاد کاندید بر روی این سیستم شناسایی پرداخته شده است. در آزمایش اول از مجموعه داده PASCAL VOC 2007 برای ارزیابی کلاس‌های موجود در این مجموعه داده به صورت



شکل ۲-۱۰ مقایسه بین روش‌های پیشنهاد کاندید بر روی مجموعه داده PASCAL VOC 2007 با روش R-CNN

سریع [۱۳]

جداگانه استفاده شده است. شکل ۲-۹ نتایج بدست آمده از روش R-CNN سریع با روش‌های کاندید متفاوت در هر کلاس را نشان می‌دهد. با توجه به نتایج بدست آمده روش‌های EdgeBox و MCG در این مجموعه داده بر روی اکثر کلاس‌ها دارای بهترین نتایج بوده و عملکرد مناسبی را از خود نشان

داده‌اند. تعداد نواحی کاندید استفاده شده در این آزمایش ۲۰۰۰ کاندید بوده است. با توجه به نتایج بدست آمده روش R-CNN سریع با نواحی کاندید در اشکالی به مانند باطری و گیاه نتایج خوبی از خود نشان نداده است ولی در مواردی به مانند قطار، اسب، گربه و ماشین این روش نتایج قابل قبولی را از خود نشان داده است.

## ۲-۳- روش‌های تطبیق پیش‌زمینه‌های کاندید با پیش‌زمینه‌های مورد نظر

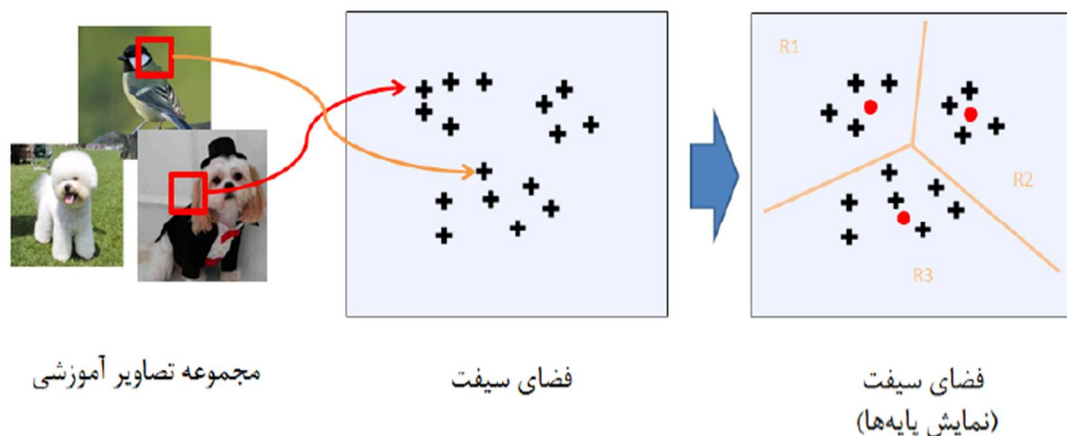
یک مرحله مهم در هم‌بخش‌بندی تصویر تطبیق ناحیه‌های کاندید با پیش‌زمینه‌های مورد نظر است. الگوریتم‌های زیادی در این زمینه وجود دارد و پایه‌ی تمامی این الگوریتم‌ها روش‌های استخراج ویژگی می‌باشد. به این صورت که در ابتدا از ناحیه‌های کاندید پیشنهاد شده، ویژگی‌هایی مناسب و قابل قبول استخراج شده و سپس وارد فاز مقایسه می‌شوند. در ادامه به بررسی برخی از این روش‌ها می‌پردازیم.

### ۲-۳-۱- بسته کلمات

در این روش [۲۸] در مرحله اول از هر تصویر تعدادی نقطه به همراه همسایگی مشخصی از آن انتخاب می‌گردد که به آنها نقاط محلی می‌گویند. در مرحله بعد تمامی این نواحی محلی با روش یکسان‌سازی هیستوگرام از نظر شدت نور و اندازه، یکسان‌سازی می‌گردند. بدین ترتیب نواحی محلی نسبت به تغییرات نور و اندازه نواحی تا حد خوبی مستقل می‌گردند. سپس از این نقاط ویژگی سیفت استخراج می‌گردد. ویژگی سیفت هیستوگرامی از اندازه و جهت بردارهای گرادیان تصویر در نقاط تصویر است. این ویژگی نسبت به دوران تصویر مستقل است. خروجی این الگوریتم یک بردار ۱۲۸ بعدی ویژگی محلی است که نسبت به تغییرات شدت نور، اندازه و دوران تا حد خوبی مستقل است. در اینجا یک بردار ویژگی سیفت را به صورت  $S_i$  نشان می‌دهیم که به معنی ویژگی سیفت استخراج شده از  $I$ -امین نقطه محلی تصاویر است.

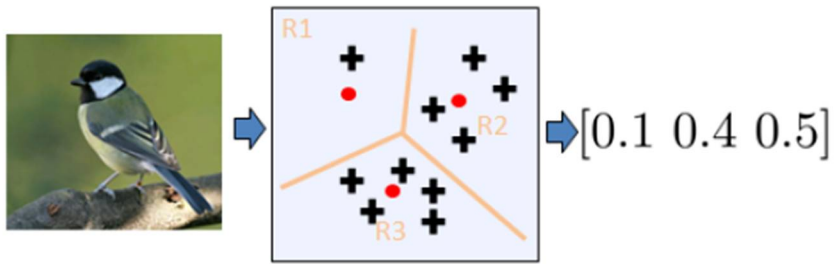
در مرحله دوم بر روی بردارهای سیفت بدست آمده از تمامی تصاویر الگوریتم خوشه‌بندی انجام می‌شود و مرکز هر خوشه به عنوان ویژگی سرآمد انتخاب می‌گردد. در واقع مرکز هر خوشه یک ویژگی سرآمد است که به بهترین شکل می‌تواند داده‌های موجود در یک خوشه را توصیف نماید. منظور از توصیف در روش مختلف، متفاوت است و می‌تواند به معنی کم‌ترین خطای بازسازی و یا نزدیکی اقلیدسی به داده‌ها باشد.  $l$  امین مرکز خوشه را به صورت  $d_l$  نشان می‌دهیم. به هر مرکز خوشه به اصطلاح پایه و به مجموعه پایه‌ها واژه‌نامه می‌گوییم. در شکل ۲-۱۰ مراحل استخراج ویژگی و خوشه‌بندی نمایش داده شده است.

در مرحله سوم که به کدگذاری معروف است، به هر ویژگی سیفت  $S_i$  یک کد  $a_i$  اختصاص می‌دهیم. ابعاد بردار کد  $a_i$  برابر با تعداد مراکز خوشه است و همه عناصر آن به جز یک عنصر برابر با ۰ است. اگر فرض کنیم داده  $S_i$  به  $l$  امین خوشه تعلیق داشته باشد عنصر  $l$  ام کد  $a_i$  برابر با یک و باقی عناصر برابر با صفر قرار داده می‌شوند. با کمی تامل مشخص می‌شود که در واقع بردار کد، یک نمایش سطح بالا از



شکل ۲-۱۱ مراحل یادگیری پایه‌ها در الگوریتم [۲۸]





شکل ۲-۱۲ ایجاد بردار ویژگی در الگوریتم [۲۸]

ویژگی‌های سطح پایین سیفت است زیرا به جای استفاده از ویژگی سطح پایین، از ارتباط آن با ویژگی‌های شاخص دیگر برای نمایش آن استفاده می‌شود.

در مرحله چهارم کدهای بدست آمده از هر تصویر را تجمیع می‌کنیم تا به یک ویژگی کلی که معرف تصویر اولیه است برسیم. در اینجا تجمیع با عمل هیستوگرام انجام می‌شود به این ترتیب بردار ویژگی نهایی بصورت زیر بدست می‌آید:

$$v_i = \frac{1}{m} \sum_{j \in \text{imag}(i)} a_j \quad 1-2$$

که  $v_i$  بردار ویژگی نهایی تصویر  $h$ ام است و  $m$  تعداد نواحی محلی استخراج شده از این تصویر است. لازم است توجه شود که عمل هیستوگرام‌گیری روی المان‌های بردار  $a_j$  انجام می‌شود و بنابراین ابعاد  $v_i$  با ابعاد بردار  $a_j$  برابر است. در شکل ۲-۱۱ شیوه کدگذاری یک تصویر به شکل نمادین نشان داده شده است.

گرچه پس از این بر روی این روش تغییرات عمده‌ای صورت پذیرفته است، اما می‌توان چارچوبی ارائه داد و این روش و روش‌های آتی را در قالب آن بررسی کرد. اغلب روش‌های استخراج ویژگی محلی شامل این مراحل هستند:

۱- انتخاب نواحی محلی بر روی تصویر

۲- استخراج ویژگی از نواحی محلی

۳- ساخت پایه‌ها که همان ویژگی‌های سرآمد هستند (یادگیری واژه‌نامه<sup>۶۶</sup>)

۴- نمایش هر ویژگی بر اساس پایه‌ها (کدگذاری<sup>۶۷</sup>)

۵- ترکیب کد نواحی محلی برای بدست آوردن ویژگی نهایی (نظرسنجی<sup>۶۸</sup>)

نحوه انتخاب نواحی محلی یکی از مسائل باز است. اگر نواحی انتخاب شده شامل مفاهیم اصلی موجود در تصویر نباشند، میزان اطلاعاتی که در این مرحله از دست خواهد رفت کار طبقه‌بندی را بسیار مشکل خواهد کرد. این نواحی معمولاً به طور تصادفی، بر روی لبه‌ها و یا بطور یکنواخت از تصویر انتخاب و استخراج می‌گردند. آزمایش‌های بسیاری برای مقایسه این روش‌ها با یکدیگر انجام شده است [۲۹]. نتایج نشان می‌دهد هیچ کدام از این روش‌ها بطور کلی بهتر از دیگری نیست.

در مرحله استخراج ویژگی روشی مطلوب است که بردار ویژگی آن تا حد ممکن مستقل از دوران باشد. از طرفی با توجه به این لبه‌های تصویر حاوی اطلاعات بیشتری نسبت به نواحی هموار تصویر هستند استفاده از روش‌هایی که مبتنی بر هیستوگرام بردار گرادیان<sup>۶۹</sup> هستند مانند سیفت و هیستوگرام گرادیان، سودمند است و در استخراج ویژگی محلی از آنها استفاده شده است.

در مرحله کدگذاری ایرادی که به این الگوریتم وارد است خطای بازسازی<sup>۷۰</sup> بالای آن است. به طور دقیق‌تر، بازسازی ویژگی اولیه از ویژگی که داده با آن توصیف می‌گردد خطای بسیاری دارد. برای حل این مشکل ضمن ارائه چارچوب تئوری، پیشنهاد شده است که هر داده را با چندین پایه نزدیک به آن توصیف نماییم که این روش‌ها در بخش بعدی مورد بررسی قرار خواهند گرفت. ایراد مهم دیگر این روش این است که محل ویژگی‌ها تاثیری در نتیجه نهایی ندارد یعنی با تغییر مکان نقاط محلی نسبت به هم، بردار ویژگی نهایی تغییر نخواهد کرد. این ایراد را می‌توان در مرحله نظرسنجی، به نحوی که مکان رویداد ویژگی‌ها در تصویر را نیز لحاظ کند، برطرف کرد. چالش دیگر مرحله نظرسنجی،

---

<sup>66</sup> dictionary learning

<sup>67</sup> coding

<sup>68</sup> pooling

<sup>69</sup> histogram of gradients (HoG)

<sup>70</sup> reconstruction error

عملگری است که از آن در تجمیع کدها استفاده می‌کنیم. روش معمول در این بخش همانطور که گفته شد استفاده از هیستوگرام است، اما به تازگی به طور شهودی مشاهده شده است که استفاده از عملگر بیشینه‌گیری نتایج بهتری خواهد داشت اما هنوز تئوری مطمئنی برای آن وجود ندارد.

### ۲-۳-۲- روش هرم تطابق مکانی

همان‌گونه که اشاره شد یکی از ایرادها در روش بسته کلمات این است که ویژگی نهایی مستقل از مکان قرارگیری نقاط محلی در تصویر است. روش هرم تطابق مکانی<sup>۷۱</sup> [۳۰] برای حل این مشکل پیشنهاد شده است. ایده اصلی این روش از روش هسته تطبیق هرمی [۳۱] الهام گرفته شده است. بنابراین قبل از آن به بررسی این روش می‌پردازیم.

### ۲-۳-۲-۱- هسته تطبیق هرمی

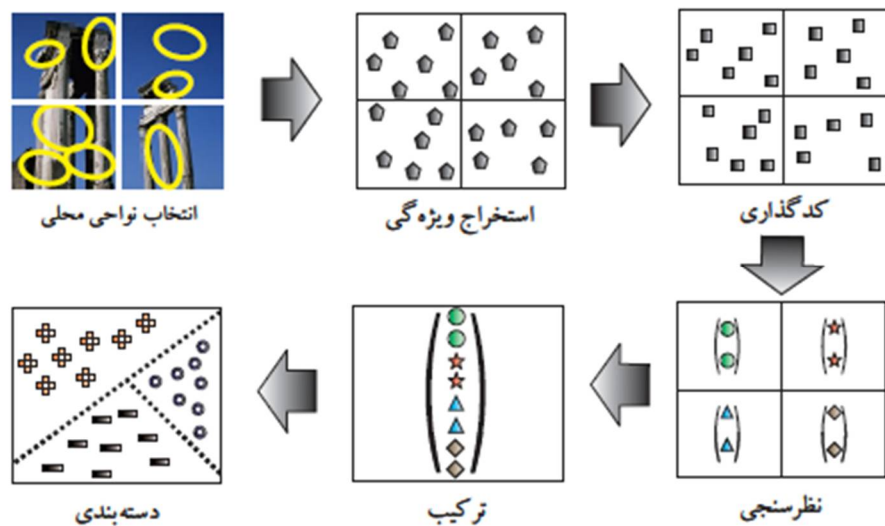
این روش برای تشخیص شباهت میان دو مجموعه به کار می‌رود. فرض کنید  $X$  و  $Y$  مجموعه‌ای از بردارهای با بعد  $d$  باشند. هسته تطبیق هرمی، ابتدا یک شبکه منظم را در فضای برداری در نظر می‌گیرد و با شمارش تعداد داده‌هایی از مجموعه  $X$  و  $Y$  که در هر سلول شبکه قرار می‌گیرند مقدار هسته در آن شبکه را مشخص می‌کند. به طور دقیق‌تر اگر شبکه‌هایی در سطوح  $0, \dots, L$  داشته باشیم، در شبکه  $l$ ام هر بعد به  $2^l$  سلول یک اندازه تقسیم می‌شود و در نهایت  $D=2^{dl}$  سلول خواهیم داشت. فرض کنید  $H_X^l(i)$  و  $H_Y^l(i)$  به ترتیب تعداد داده‌هایی از مجموعه  $X$  و  $Y$  است که در شبکه  $l$ ام در سلول  $l$ ام قرار گرفته‌اند. حال تعداد تطبیق‌های دو مجموعه در سطح  $l$ ام به این صورت نمایش داده می‌شود:

---

71spatial pyramid matching (SPM)

$$I_l = \sum_i^D f(H_Y^l(i), H_X^l(i)) \quad ۲-۲$$

که  $f$  تابع شمارش است و  $I_l$  تعداد تطبیق‌ها در سطح  $l$ ام است. از آنجایی که تعداد تطبیق‌ها در این سطح، تطبیق‌ها در شبکه ریزتر سطح  $l+1$  را شامل می‌شود. تعداد تطبیق‌های جدیدی که در این سطح پیدا شده است برابر  $I_l - I_{l+1}$  است. در نهایت تطابق در همه سطوح با یکدیگر به صورت وزن‌دار جمع می‌شوند تا هسته نهایی شکل پیدا کند. وزن هر سطح به صورت  $\frac{1}{2^{L-l}}$  در نظر گرفته می‌شود زیرا عدم تطابق در سطوح ریزتر حائز اهمیت بیشتری است. بنابراین شباهت نهایی میان دو مجموعه  $X$  و  $Y$  به فرم زیر تعریف می‌گردد:



شکل ۲-۱۳ مراحل کلی استخراج ویژگی در روش هرم تطابق مکانی

$$K^L(X, Y) = I^L + \sum_{l=0}^{L-1} \frac{1}{2^{L-l}} (I_l - I_{l+1}) \quad ۳-۲$$

همچنین اثبات شده است که این هسته شرط مرسر<sup>۷۲</sup> را دارد [۳۱].

<sup>72</sup> Mercer's condition

### ۲-۳-۲-هرم تطابق مکانی

برای در نظر گرفتن ترتیب مکانی کدها در نظرسنجی از روش هسته تطبیق هرمی استفاده شده است. اما در این روش شبکه‌بندی‌ها بر روی تصاویر انجام می‌شود حال آنکه ویژگی‌هایی که با استفاده از تابع  $f$  تطبیق داده می‌شوند همان کدهای تصویر هستند. به طور دقیق‌تر فرض کنید  $X$  مجموعه ویژگی‌های یک تصویر است. ویژگی این تصویر در سطح  $\lambda_m$  با عمل هیستوگرام کدها در هر سلول بدست می‌آید، بنابراین به ازای هر سلول یک بردار ویژگی خواهیم داشت. در نهایت با پشت سرهم قرار دادن ویژگی‌های همه سلول‌ها در همه سطوح ویژگی نهایی بدست می‌آید. شکل ۲-۱۲ کلیه مراحل در استخراج ویژگی را با فرض  $L=1$  نشان می‌دهد. همانطور که در شکل نیز مشخص این روش تنها در مرحله نظرسنجی با روش بسته کلمات تفاوت دارد و بردار ویژگی را به نحوی تولید کرده است که مکان ویژگی‌ها نیز در بردار ویژگی نهایی تاثیر داشته باشند. در پایان با استفاده از ماشین بردار پشتیبان و با هسته غیر خطی مساله چند دسته‌ای حل می‌گردد و برچسب تصاویر مشخص می‌شود.

### ۲-۳-۳-کدگذاری تنک

روش کدگذاری تنک [۲۹] با تغییر مکانیزم کدگذاری اولیه باعث کاهش خطای بازسازی، که در بخش‌های قبل به عنوان مشکل اصلی در کدگذاری روش بسته کلمات به آن اشاره شد، می‌شود. کدگذاری به روش بسته کلمات را که هر ویژگی به نزدیک‌ترین پایه در واژه‌نامه تعلق می‌گیرد، می‌توان به فرم بهیته‌سازی زیر نوشت:

$$\min_{D,A} \sum_m \|s_m - Da_m\|^2 \quad ۴-۲$$

به طوریکه: تنها یک عنصر بردار  $a_m$  برابر با ۱ و باقی صفر می‌باشند. که در آن  $s_m$  و  $a_m$  به ترتیب ویژگی و کد متناظر با ناحیه محلی  $m$  است و  $D$  و  $A$  ماتریس‌هایی هستند که ستون‌های آنها به ترتیب پایه‌های واژه‌نامه و کدها هستند. منظور از عملگر  $\|\cdot\|$  نرم دو است. در واقع برای هر  $m$  المانی از  $a_m$  برابر ۱ می‌شود که پایه متناظر آن در واژه‌نامه  $D$  به  $s_m$  نزدیکتر باشد. شرط "تنها یک پایه انتخاب

شود" شرط محکمی است که می‌توان آن را با شرط "تعداد پایه‌هایی که انتخاب می‌شوند محدود باشند" متعادل نمود. با انجام این کار یک ویژگی با ترکیب خطی تعداد محدودی پایه ساخته خواهد شد که خطای بازسازی را تا حد زیادی کاهش می‌دهد. این راه حل را می‌توان به فرم بهینه‌سازی زیر نوشت:

$$\min_{D,A} \sum_m \|s_m - Da_m\|^2 + \lambda |a_m| \quad \forall_i \|d_i\| \leq 1 \quad 5-2$$

در اینجا منظور از عملگر  $\|\cdot\|$  نرم یک است و تضمین می‌کند عناصر کمی از بردار  $a_m$  غیر صفر باشند. شرط در نظر گرفته شده بر روی اندازه پایه‌ها باعث حذف جواب‌های بدیهی می‌گردد. نشان داده شده است که با بدست آوردن کد داده‌ها بر اساس کدگذاری تنک دقت دسته‌بندی در حد چشم‌گیری افزایش می‌یابد و همچنین دسته‌بندی با یک ماشین‌بردار پشتیبان با هسته خطی نتایج مانند هسته غیرخطی دارد. این مساله بسیار مهمی است زیرا آموزش یک ماشین‌بردار پشتیبان با هسته غیرخطی هزینه زمانی از مرتبه  $O(n^3)$  و آزمون آن هزینه زمانی از مرتبه  $O(n)$  دارد در حالی که با هسته خطی هزینه زمانی آموزش به  $O(n)$  و آزمون به  $O(1)$  کاهش می‌یابد. ادعا می‌شود که این ویژگی‌ها تا حد زیادی معنایی هستند و به همین دلیل جداسازی آنها با خط امکان‌پذیر است.

کار دیگری که در این روش انجام شده است، استفاده از عملگر بیشینه‌گیری به جای هیستوگرام در محله نظرسنجی است. پشتوانه این کار مشاهداتی است که نشان می‌دهد عملکرد نرون‌ها در مغز شبیه عمل بیشینه‌گیر است. نتایج عملی نیز پیشرفت محسوس در نتایج دسته‌بندی را نشان می‌دهد.

## ۲-۳-۴- کدگذاری با پایه‌های محلی

طبق مشاهدات عملی که در [۳۲] گزارش شده، هنگامی که از پایه‌های محلی برای ساخت یک ویژگی استفاده می‌گردد نتایج دسته‌بندی تصویر نسبت به روش کدگذاری تنک بهبود می‌یابد. از طرف دیگر روش کدگذاری تنک با وجود دقت دسته‌بندی بالا، از پشتوانه تئوری قوی در یادگیر ماشین

برخودار نیست [۴۵]. در این روش به طور تئوری نشان داده می‌شود که با انتخاب محلی پایه‌ها برای کدگذاری یک داده و استفاده از یک روش یادگیر خطی می‌توان هر تابعی را بروی منیفولدی که داده بر روی آن قرار گرفته‌اند یاد گرفت. همچنین ثابت می‌شود پیچیدگی یادگیر مرتبط با بعد منیفولد و نه بعد کل فضا است. این نتیجه بسیار مهم است زیرا نشان می‌دهد این روش در زمره روش‌های نیمه-نظارتی قرار می‌گیرد. تابع بهینه‌سازی این روش به صورت زیر است:

$$\min_{D,A} \sum_m \|s_m - Da_m\|^2 + \lambda \sum_i |a_{mi}| + \|s_m - d_i\|^2 \quad ۶-۲$$

که منظور از  $a_{mi}$  درایه  $i$ ام بردار  $a_m$  است. بنابراین اگر داده  $s_m$  از پایه  $d_i$  دور باشد، مقدار عبارت  $\|s_m - d_i\|$  بزرگ خواهد بود و در بهینه‌سازی با کوچک انتخاب کردن المان متناظر آن یعنی  $a_{mi}$  مقدار استفاده از این پایه در کدگذاری کاهش می‌یابد. بدین ترتیب در کدگذاری هر داده تنها از پایه‌های محلی آن استفاده می‌گردد.

### ۲-۳-۵- کدگذاری محلی و خطی

در رابطه ۶-۲ فرم بهینه‌سازی روش کدگذاری با پایه‌های محلی شامل دو جمله است. جمله اول کد را به نحوی انتخاب می‌کند که خطای بازسازی داده تا حد امکان کوچک باشد در حالی که جمله دوم پایه‌های محلی را برای انتخاب در اولویت قرار می‌دهد. با استفاده از این ایده روش کدگذاری محلی و خطی [۳۳] یک راه‌حل دو مرحله‌ای را پیشنهاد می‌دهد که به شکل زیر است:

۱- برای ویژگی  $s_m$  تعداد  $k$  نزدیک‌ترین پایه را انتخاب کن و در ستون‌های ماتریس  $D_{km}$  قرار ده.

۲- عبارت  $\min_{\alpha_m} \sum_m \|s_m - D_{km} \alpha_m\|^2$  را با فرض واژه‌نامه ثابت برای بدست آوردن ضرایب متناظر

کدنهایی حل کن.

در گام اول تضمین می‌شود که پایه‌ها محلی انتخاب می‌شوند و در گام دوم ضرایب کد به گونه‌ای انتخاب می‌شوند که خطای بازسازی داده کمینه گردد. در واقع فرم بهینه‌سازی پایه‌های محلی که در یک مرحله حل می‌شد به دو بخش تقسیم شده است. با این کار نه تنها دقت دسته‌بندی کاهش نیافته است، بلکه زمان اجرا به مقدار قابل توجهی کاهش می‌یابد. علت این امر است که عبارت بهینه‌سازی در مرحله دوم تنها بر روی تعداد محدودی از پایه‌ها اجرا می‌گردد. برای یادگیری واژه‌نامه نیز یک روش برخط مبتنی بر عبارت بهینه‌سازی اولیه پیشنهاد شده است اما به طور عملی نشان داده شده است که استفاده از یک روش خوشه‌بندی ساده نیز نتایج مشابهی را تولید می‌کند.

همانطور که گفته شده با توجه به سرعت بالای این روش نسبت به روش کدگذاری با پایه‌های محلی استفاده از آن در دسته‌بندی مجموعه تصاویر بزرگ ممکن است اما ایراد واضح این روش این است که تعداد پایه‌هایی که در کدگذاری شرکت می‌کنند همواره ثابت و برابر با  $k$  است اما در روش کدگذاری با پایه‌های محلی، که از نرم اول برای تضمین تنگ بودن استفاده می‌کند، تعداد پایه‌های شرکت‌کننده در ساخت یک داده می‌تواند داده به داده تغییر کند و با ویژگی آن داده تنظیم شود.

## ۲-۳-۶- کدگذاری نرم با برش

در روش کدگذاری نرم [۳۴] درایه متناظر هر پایه در بردار کد برابر مقدار هسته گاوسی در نظر گرفته می‌شود:

$$a_{ij} = \exp\left(-\frac{\|s_i - d_j\|^2}{2\sigma}\right) \quad ۷-۲$$

این روش برخلاف روش کدگذاری استفاده شده در بسته کلمات تعلق به هر پایه را به صورت نرم و با استفاده از هسته گاوسی اندازه‌گیری می‌کند. با وجود پیشرفت قابل ملاحظه، این روش یک ایراد واضح دارد: از آنجایی که فاصله اقلیدیس فقط در فاصله‌های کوتاه معیار مناسبی برای فاصله منیفولدی است مقدار رابطه ۷-۲ تنها برای مقادیر کوچک  $\|s_i - d_j\|^2$  مقادیر درستی است. در [۱۶] مقدار رابطه ۷-۲



تنها برای  $k$  نزدیکترین پایه به ویژگی سنجیده می‌شود و باقی درایه‌ها برابر صفر گذاشته می‌شود. با کمی دقت مشاهده می‌شود که در این روش درایه‌های یکسانی با روش کدگذاری محلی و خطی غیر صفر است و تنها تفاوت در دو روش انتخاب مقادیر درایه‌هاست که در این روش از معیار شباهت و در روش کدگذاری محلی و خطی از ضرایب بازسازی استفاده شده است.

### ۲-۳-۷- هسته بهینه برای تشخیص شباهت ویژگی‌ها

در این روش [۳۵] ابتدا به مساله کدگذاری از نگاه طراحی هسته نگریسته می‌شود و بعد تلاش می‌شود یک هسته بهینه برای حل مساله ارائه گردد. با این کار یک چارچوب ارائه می‌گردد که می‌توان یک هسته بهینه طراحی نمود در حالی که روش‌هایی که معرفی شد تنها یک راه‌حل ارائه می‌دهند و از این نظر قابل مقایسه و بررسی نیستند. این روش که همزمان با روش کدگذاری محلی و خطی معرفی شده است، با ارائه هسته بهینه سعی در کاهش خطای بازسازی در روش کدگذاری بسته کلمات دارد. گرچه نگاه این دو روش به مشکل متفاوت است و راه‌حل‌های نسبتاً متفاوتی نیز ارائه می‌نمایند اما نتایج عملی مشابهی دارند.

همانطور که در روش بسته کلمات گفته شد در روش هیستوگرام‌گیری ویژگی نهایی از روی بردار کدها به شکل زیر محاسبه می‌گردد:

$$v_i = \frac{1}{m} \sum_c a_c \quad ۸-۲$$

با فرض استفاده از هسته خطی در دسته‌بندی می‌توان نوشت:

$$K(v_k, v_l) = v_k^T v_l = \left( \frac{1}{m} \sum_c a_c \right)^T \times \frac{1}{n} \sum_f a_f = \frac{1}{mn} \times \sum_c \sum_f a_c^T a_f = \frac{1}{mn} \times \sum_c \sum_f K_s(s_c, s_f) \quad ۹-۲$$

که در آن  $n$  و  $m$  تعداد ویژگی‌های محلی در دو تصویر است. نتیجه این محاسبات این است که در حالت هیستوگرام‌گیری می‌توان به جای هسته  $a_c^T a_f$  از یک هسته دلخواه بر روی دو ویژگی اولیه استفاده کرد. همانطور که گفته شد مشکلی که در استفاده از هسته وجود دارد زمانی بالای فاز

یادگیری است، به همین منظور راه‌حلی پیشنهادی می‌گردد که بتوان به‌طور مستقیم فضای نگاشت هسته را پیدا کرد.

هسته  $K_s(s_c, s_f)$  هسته‌ای با بعد محدود خوانده می‌شود اگر نگاشت متناظر آن  $\Phi(\cdot)$  بعد محدود داشته باشد که در این صورت می‌توان نوشت:

$$K_s(s_c, s_f) = \Phi(s_c)^T \Phi(s_f) \quad 10-2$$

در این صورت با توجه به رابطه ۱۰-۲ می‌توان هسته اصلی بر روی دو ویژگی را به صورت زیر تعریف کرد:

$$K(v_k, v_l) = \frac{1}{mn} \times \sum_c \sum_f K_s(s_c, s_f) = \frac{1}{m} \sum_c \Phi(s_c) \times \frac{1}{n} \sum_f \Phi(s_f) = \Phi(v_k)^T \Phi(v_l) \quad 11-2$$

اما در عمل فضای نگاشت بسیاری از هسته‌ها بعد محدود ندارند. برای مثال در این روش از هسته گاوسی به صورت

$$K_s(s_c, s_f) = \exp\left(-\frac{\|s_c - s_f\|^2}{\sigma}\right) \quad 12-2$$

استفاده می‌گردد که فضای نگاشت با بعد محدود ندارد. راه‌حلی که پیشنهاد می‌شود یادگیری نگاشتی با بعد محدود است که هسته آن نزدیک به هسته اصلی باشد. در ادامه به بررسی این روش می‌پردازیم.

### ۲-۳-۷-۱- یادگیری ویژگی با بعد محدود

روش کار به این صورت است که ویژگی با بعد زیاد را بر روی یک فضای  $D$  بعدی تصویر می‌کنیم. فرض کنیم پایه‌های این فضا به صورت  $\{\Phi(z_i)\}_{i=1}^D$  داده شده است. اگر فرض کنیم ستون‌های ماتریس  $H$  این پایه‌ها هستند ضرایب تصویر بر روی این پایه‌ها به شکل زیر نوشته محاسبه می‌گردد.

$$\bar{p}_x = \underset{p_x}{\operatorname{argmin}} \|\Phi(x) - Hp_x\|^2 \quad 13-2$$

که  $\bar{p}_x$  ضرایب تصویر بردار  $\Phi(x)$  در فضای است که ستون‌های  $H$  ایجاد می‌کنند. این عبارت بهینه-سازی محدب و پاسخی به فرم بسته دارد:

$$\bar{p}_x = (H^T H)^{-1} H^T \Phi(x) \quad 14-2$$

بنابراین هسته تعریف شده در فضای تصویر شده به صورت زیر تعریف می‌گردد:

$$K(x, y) = (H\bar{p}_x)^T (H\bar{p}_y) = \bar{p}_x^T H^T H \bar{p}_y = \bar{p}_x^T L \bar{p}_y \quad 15-2$$

که  $L = H^T H$  یک ماتریس  $D \times D$  بعدی است که المان  $(i, j)$  آن  $K(z_i, z_j)$  در فضای نگاشت اولیه است. بنابراین با توجه به قضیه مرسر فشای نگاشت در فضای تصویر بصورت زیر نوشته می‌شود:

$$\Phi(x) = G\bar{p}_x \quad 16-2$$

که در آن  $G = V^{0.5}$  ماتریس  $D \times D$  بعدی است که  $V$  و  $S$  به ترتیب ماتریس ضرایب و مقادیر ویژه در تجزیه طیفی ماتریس  $L$  است. سوال دیگر این است که پایه‌های مناسب برای ایجاد فضا را چگونه پیدا کنیم. برای این کار استفاده از روش آنالیز مولفه‌های اصلی هسته<sup>۷۳</sup> [۳۶] پیشنهاد می‌گردد. تا پایه‌ها به نحوی انتخاب شوند که در جهت بیشترین واریانس داده‌ها در فضای نگاشت باشند.

برتری تئوری این روش نسبت به روش‌های دیگر این است که در چارچوب هسته بهینه، کدگذاری را بررسی می‌کند، در واقع این چارچوب به اندازه‌ای کلی است که روش‌های دیگر کدگذاری را نیز می‌توان ذیل آن بررسی کرد و با محاسبه هسته آنها قدرت آنها را با هم مقایسه نمود. اما در حیطه عمل با وجود نتایج مشابه روش کدگذاری محلی و خطی، به زمان اجرای بالاتری برای محاسبه ماتریس  $G$ ، به علت انجام عمل ماتریس معکوس، نیاز دارد.

<sup>73</sup> Kernel Principal Component Analyses (KPCA)

## ۲-۳-۹- نتیجه‌گیری

یکی از تنگناهای محاسباتی در سیستم‌های تشخیص و شناسایی اشیاء در تصویر، بدست آوردن ناحیه‌های کاندید هستند. در واقع عملکرد این نوع سیستم‌ها به روش انتخاب نواحی کاندید وابسته است. در همین راستا، به معرفی و مقایسه‌ی روش‌های ناحیه‌های کاندید پرداخته شده و بررسی شد که این روش‌ها در یک مجموعه داده یکسان چه عملکردی را از خود نشان می‌دهند. به منظور مقایسه این روش‌ها از یک سیستم شناسایی مشهور و محبوب R-CNN سریع که مورد استقبال بسیاری از محققان و پژوهشگران در این زمینه بوده، استفاده شده است.

اساس همه روش‌های یادگیری ویژگی، که در این فصل بررسی نموده‌ایم، روش بسته کلمات است. روش‌هایی که در ادامه مطرح شدند، اغلب برای حل مشکلاتی است که در طراحی اولیه روش بسته کلمات وجود داشته‌اند. یکی از این مشکلات مستقل بودن ویژگی نخایی نسبت به تغییرات مکانی نقاط محلی است. برای حل این مشکل روش هرم تطابق مکانی ارائه شد، این روش هنوز هم به همراه روش‌های کدگذاری، که بعدها ارائه شدند، مورد استفاده قرار می‌گیرد. روش‌هایی که برای کدگذاری ارائه شده همگی حول یک نقطه اشتراک دارند و آن کاهش خطای بازسازی داده در عمل کدگذاری است. در این میان روش‌های کدگذاری با پایه‌های محلی و روش هسته بهینه برای تشخیص شباهت ویژگی‌ها، از نظر تئوری و روش کدگذاری تنک و کدگذاری محلی و خطی، به لحاظ کارکرد عملی مورد توجه قرار گرفته‌اند.

فصل روم:

# روش شناسی مادی

### ۳-۱- مقدمه

در تحقیقات انجام شده در سال‌های اخیر، بدست آوردن ناحیه‌های کاندید به عنوان یک مرحله اساسی و مهم در سیستم‌های تشخیص و شناسایی اشیای موجود در تصویر معرفی شده‌است. همچنین در این تحقیقات بیان شده که بدست آوردن این ناحیه‌ها به مانند یک تنگنا بوده و بیشترین بار محاسبات در این نوع از سیستم‌ها وابسته به آن می‌باشد. در همین راستا انتخاب روش مناسب و سریع می‌تواند در بهبود عملکرد سیستم‌های تشخیص بسیار حائز اهمیت باشد. الگوریتم ارائه شده از این امر مستثنا نبوده و بیشترین بار هزینه زمانی را در این قسمت دارا می‌باشد. بنابراین ارائه روشی مناسب و کارآمد به منظور بدست آوردن ناحیه کاندید برای هم‌بخش‌بندی تصویر می‌تواند موجب ارائه یک الگوریتم بهینه و کارآمد گردد. به همین منظور ما در روش پیشنهادی خود بیشترین تمرکز خود را در مرحله مدل کردن ناحیه‌های کاندید گذاشته‌ایم.

قبل از طراحی و ارائه یک الگوریتم ما چالش‌ها و مشکلات پیش‌روی را که در دقت الگوریتم تاثیر گذار است را معرفی می‌کنیم، در تصاویر طبیعی معمولاً سه ویژگی اساسی وجود دارد که کار دسته‌بندی را مشکل می‌کند:

۱- پس زمینه‌ها اغلب پیچیده هستند و دسته‌های مختلف ممکن است پس‌زمینه مشترک داشته باشند.

۲- اشیا ممکن است در هر نقطه‌ای از تصویر، با زوایای مختلف و نورپردازی‌های متفاوت وجود داشته باشند.

۳- اشیا ممکن است توسط یک شی دیگر دچار پوشیده‌گی جزئی شده باشند.

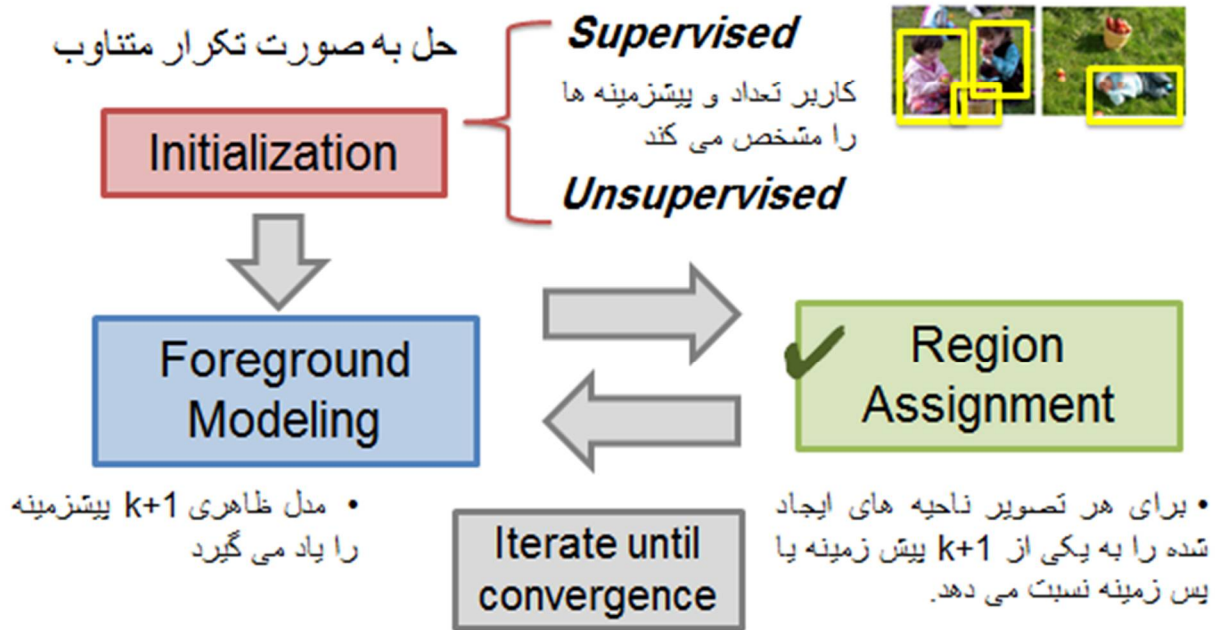
روش‌های استفاده شده در این نوع از سیستم‌ها باید با توجه به این چالش‌ها و پیچیده‌گی سنجیده شوند و بتواند تا حد ممکن بر آنها فائق آیند.

### ۳-۲- فرموله کردن مسئله

ما تصاویر ورودی را با  $I = \{I_1, \dots, I_M\}$  نشان می‌دهیم. هدف ما بیرون کشیدن  $k$  پیش-زمینه مشترک  $F = \{F^1, \dots, F^K\}$  از میان تصاویر موجود در مجموعه  $I$  است، هر تصویر شامل یک زیرمجموعه ناشناخته از  $F$  می‌باشد. الگوریتم ما با دو سناریو مختلف سروکار دارد. در طراحی بدون-ناظر، یک کاربر تنها به ورودی  $k$  مقدار می‌دهد و الگوریتم پیشنهادی به صورت خودکار  $k$  پیش‌زمینه مشخص، که بیشترین مطابقت را در مجموعه  $I$  دارد استنتاج می‌کند. در طراحی باناظر، کاربر می‌تواند با مشخص کردن محدوده شی یا براساس پیکسل، مرزها را برای  $k$  پیش‌زمینه از تصاویر انتخاب شده مشخص کند.

ما مسئله تعریف شده در بالا را به دو زیر مسئله تقسیم می‌کنیم به صورتی که ما همواره مدل کردن پیش‌زمینه و تخصیص ناحیه را حل می‌کنیم. مدل کردن پیش‌زمینه، مدل ظاهری از  $K$  پیش-زمینه یا پس‌زمینه را یاد می‌گیرد و تخصیص ناحیه، ناحیه‌هایی از هر تصویر را به یکی از  $K$  پیش‌زمینه یا پس‌زمینه نظیر می‌کند. روش ما یک راه‌حل برای هر یک از دو زیر مسئله می‌دهد که قابل حل است. این روش می‌تواند از یک ناحیه تخصیص داده شده اولیه،  $k+1$  مدل پیش‌زمینه را یاد بگیرد و به نوبه خود در هر تصویر، ناحیه تخصیصی را بهبود ببخشد. این دو پردازش به صورت نتاوبی تا زمانی که به راه حل همگرا شوند ادامه پیدا می‌کنند. نمایی کلی از روش کار ما در زمینه بخشبندی تصویر در شکل ۳-۱ آورده شده است. البته کلیت کار اکثر الگوریتم‌هایی که در این زمینه کار می‌کنند به صورت نمایش داده شده در شکل ۳-۱ می‌باشد.

## شمای کلی از الگوریتم



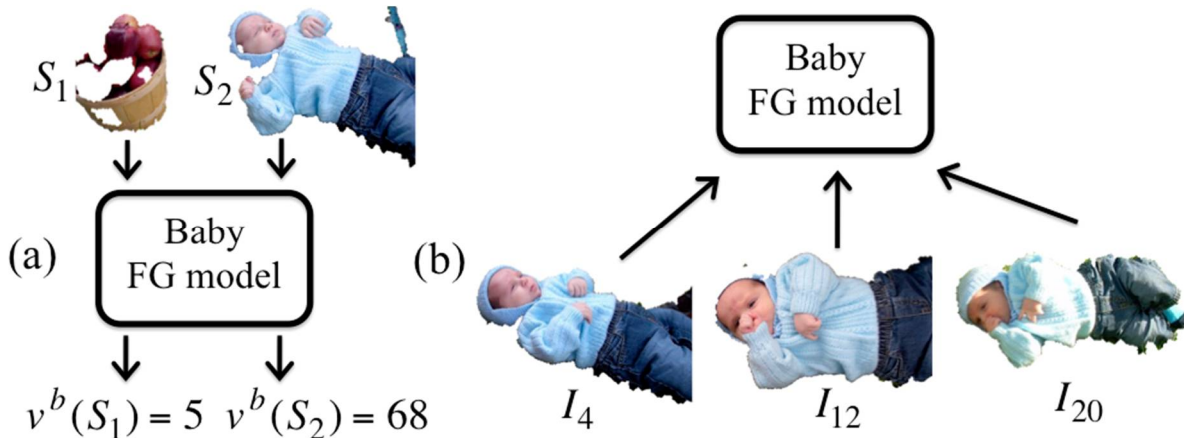
شکل ۱-۳ شمای کلی الگوریتم هم‌بخش‌بندی

### ۳-۳- مدل کردن پیش‌زمینه

بدون این که تغییری در کلیت گفته شده در بخش قبل ایجاد شود، ما  $K$  آمین پیش‌زمینه را مدل می‌کنیم، این کار را با یک تابع پارامتریک انجام می‌دهیم. به صورت  $V^k: S \rightarrow R$  که هر ناحیه  $S \in S$  در یک تصویر را به یک مقدار درست برای  $K$  آمین پیش‌زمینه نظیر می‌کند. اگر یک تصویر مانند  $I_i$  به صورت  $S_i$  بخش‌بندی شود آنگاه  $V^k: 2^{|S_i|} \rightarrow R$  هر زیرمجموعه از  $S \subset S_i$  را به عنوان ورودی می‌گیرد و مقدار بازگشتی آن را به  $K$  آمین پیش‌زمینه در آن تصویر می‌دهد. در مدت تخصیص ناحیه، هر مدل پیش‌زمینه ارزیابی می‌شود که چگونه یک ناحیه برای پیش‌زمینه بودن مناسب است، همان‌گونه که در شکل ۱-۳ (a) نشان داده شده است. در زمانه مدل کردن پیش‌زمینه،



هر مدل پیش‌زمینه با یادگرفتن از بخش‌های تعیین شده برای پیش‌زمینه، بروز رسانی می‌شود. مانند شکل ۱-۳(b).



شکل ۱-۳ یک مثال برای مدل پیش‌زمینه نوزاد. (a) یک مدل پیش‌زمینه یک تابع پارامتریک می‌باشد که هر ناحیه را با یک مقدار به یک پیش‌زمینه نسبت می‌دهد. (b) قبل از تخصیص ناحیه، مدل پیش‌زمینه با ناحیه‌های اختصاص داده شده به پیش‌زمینه آموزش داده می‌شود.

یک قسمت مهم از روش ما این است که هر الگوریتم طبقه‌بندی یا رتبه‌بندی می‌تواند به عنوان مدل پیش‌زمینه استفاده شود البته این قابلیت تا زمانی قابل پذیرش است که این روش‌های مدل کردن پیش‌زمینه بتوانند یک ناحیه را ارزیابی کنند و بتوانند به وسیله یادگیری از ناحیه‌های تخصیص داده شده بروز رسانی شوند. ما از دو روش مدل ترکیب گوسین (GGM) و دیگری تطبیق هرم فضایی (SPM) یا SVM خطی برای مدل کردن پیش‌زمینه استفاده می‌کنیم.

$$v^k(S) = \alpha \cdot v_{GSM}^k(s) + (1 - \alpha) \cdot v_{SPM}^k(S) \quad 1-3$$

محاسبه می‌شود که  $\alpha$  بین صفر و یک تغییر می‌کند.

### ۳-۳- تخصیص ناحیه

با توجه به مدل پیش زمینه، تخصیص ناحیه برای هر تصویر به صورت جداگانه انجام می‌شود. هدف از این مرحله تقسیم کردن  $S_i$  (مجموعه قسمت های هر تصویر  $I_i$ ) به صورت زیرمجموعه‌هایی مجزا از پیش‌زمینه‌ها ( $F_i^k$  ( $k = \{1, \dots, K\}$ ) و پس‌زمینه ( برای سادگی در نوشتار ما از  $F_i^{K+1}$  برای نمایش پس‌زمینه استفاده می‌کنیم) است. با توجه به این که تمام پیش‌زمینه‌ها در یک تصویر ظاهر نمی‌شود بعضی از پیش‌زمینه‌ها ( $F_i^k$ ) مجموعه‌ای تهی هستند. برای سادگی کار، ممکن است ما هر قسمت  $S \in S_i$  را به یکی از  $F_i^k$  که بیشترین مقدار  $v^k(S)$  برای آن وجود دارد تعمیم دهیم. به هر حال، در بخش‌بندی تصویر، ارزش بدست آمده برای یک بخش می‌تواند بیشتر یا کمتر از مجموع ارزش‌های بخش‌های جدا باشد. برای مثال، یک ناحیه سیاه رنگ بیشترین مقدار ارزش را برای پیش‌زمینه گاو ماده دارد. اما اگر یک ناحیه سیاه رنگ با یک مجموعه از ناحیه‌های رنگی ترکیب شود، مقدار ارزشی که برای آن به دست می‌آید ممکن است برای پیش‌زمینه یک شخص با ارزش تر باشد تا پیش‌زمینه یک گاو ماده.

نهایتاً، بخش‌بندی ناحیه به یافتن یک قسمت مجزا  $S_i = \bigcup_{k=1}^{K+1} f_i^k$  با  $k \neq l$  اگر  $f_i^k \cap f_i^l = \emptyset$  با بیشترین  $\sum_{k=1}^{K+1} v_k(f_i^k)$  خلاصه می‌شود.

در نهایت به رابطه ۲-۳ خواهیم رسید.

$$\begin{aligned} \max \quad & \sum_{k=1}^{K+1} \sum_{S \subseteq S_i} v^k(S) x^k(S) \\ \text{s.t.} \quad & \sum_{k=1}^{K+1} \sum_{s \in S, S \subseteq S_i} x^k(S) \leq 1, \quad \forall s \in S_i, \\ & x^k(S) \in \{0, 1\} \end{aligned} \quad 2-3$$

در اینجا متغیر  $x^K(S)$  تخصیصی از  $S$  به  $k$ امین پیش‌زمینه  $f_i^k$  را شرح می‌دهد برای مثال  $x^K(S) = 1$  اگر و تنها اگر  $k$ امین پیش‌زمینه دسته  $S$  را شامل شود. اولین محدودیتی که چک می‌شود این است که هر بخش  $S \in S_i$  نمی‌تواند بیش از دوبار اختصاص داده شود.

تخصیص ناحیه در معادله یک نیازمند به بررسی تمامی احتمال‌های ممکن زیرمجموعه  $S \subseteq S_i$  می‌باشد. متأسفانه  $2^{|S_i|}$  احتمال ممکن از زیرمجموعه‌ها وجود دارد. بنابراین انجام آن نشدنی است.

### ۳-۳- روش پیشنهادی به منظور بخش‌بندی همزمان چند پیش‌زمینه

ما برای حل مشکل بالا از یک روش هم‌بخش‌بندی چند پیش‌زمینه‌ای ساده‌تر که به طور عملی دو مشکل ایجاد شده در قسمت قبل را حل می‌کند استفاده می‌کنیم. در قسمت قبل دیدیم که مدل کردن پیش‌زمینه ساده و روان است اما تخصیص ناحیه غیر قابل اجرا می‌باشد. ما در این قسمت به توسعه الگوریتم زمان چندجمله‌می پردازیم که برای حل مشکل تخصیص ناحیه از ویژگی‌های ساختاری که به طور عادی در تصویر قابل مشاهده است استفاده می‌کند.

### ۳-۳-۱- درخت محدود کننده تخصیص ناحیه

با توجه به  $k+1$  مدل پیش‌زمینه، مدل تخصیص ناحیه جاری به شرح زیر است: اول، هر تصویر  $I_i$  پیش تقسیم می‌شود مانند  $S_i$  که در شکل ۳-۲ نمایش داده شده است. از هر الگوریتم بخش‌بندی می‌تواند استفاده کرد، و ما زیر مدل‌های بخش‌بندی تصویر  $sift$  را برای هر تصویر استفاده می‌کنیم.

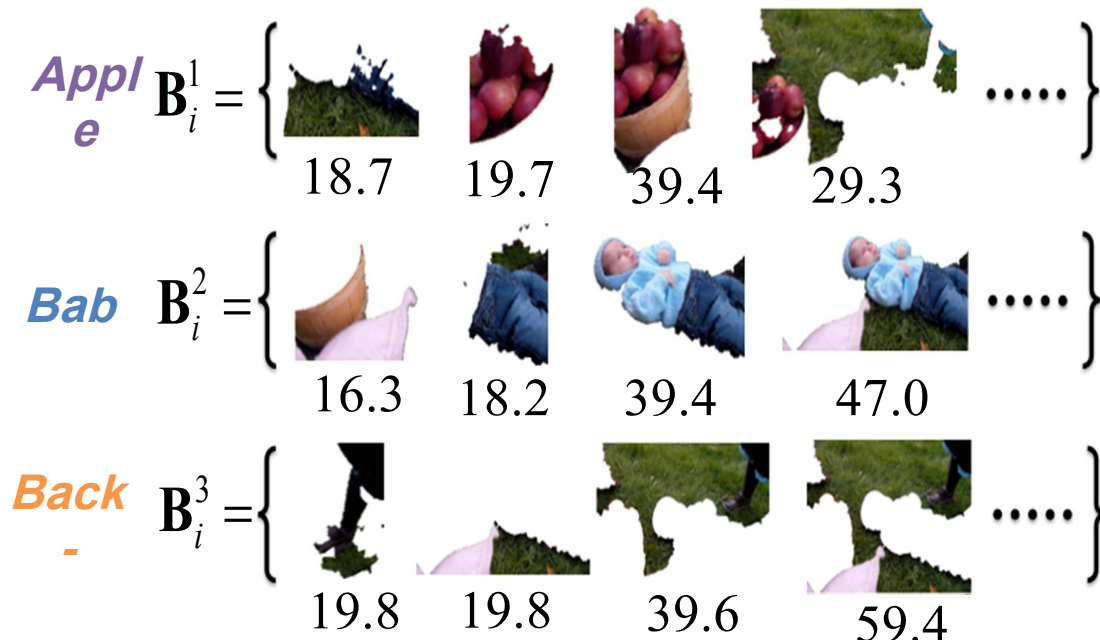


$S_i$

شکل ۲-۳ پیش تقسیم تصویر  $I_i$  به  $S_i$

باتوجه به مجموعه قسمت‌های ایجاد شده  $S_i$  از تصویر  $I_i$ ، هر پیش‌زمینه در  $F$  یک مجموعه از پیش‌زمینه‌های کاندید  $B_j^k = \{B_1^k, \dots, B_n^k\}$  را ایجاد می‌کند که هر کاندید یک چندتایی  $B_j^k = \langle K_j, C_j, W_j \rangle$  است که  $K_j$  اندیس پیش‌زمینه‌ای است که کاندید  $j$  پیشنهاد می‌دهد،  $C_j \subseteq S_i$  دسته‌ای از بخش‌ها مشخص شده در تصویر  $I_i$  است و  $W_j$  برابر با مقدار  $W_j = v^k C_j$  است. یک مثال از آن را در شکل ۳-۳ مشاهده می‌کنید. در این مرحله ما اجازه می‌دهیم هر پیش-

زمینه، تعداد نامحدودی از کاندیدها را که مایل به انتخاب آنها است را پیشنهاد کند. نهایتاً بخش‌بندی تصویر در معادله ۲-۳، با انتخاب بعضی از کاندیدهای پیش‌زمینه قابل قبول‌تر، از میان همه پیشنهادهای  $B_i = \{B_i^1, \dots, B_i^{k+1}\}$ ، به منظور به حداکثر رساندن ارزش کلی آن، حل می‌شود.



شکل ۳-۳ نمایشی از امتیاز دادن به کاندیدهای پیشنهادی هر پیش‌زمینه

دو روش احتمالی برای حل مسئله تخصیص ناحیه در رابطه ۲-۳ وجود دارد :

قابلیت قطعه شدن (tractable) : اعمال یک محدودیت روی مقدار تابع  $v^k$  یا یک محدودیت روی ایجاد کردن پیش‌زمینه‌های کاندید  $B_i$ . ما همانند روش قبلی اعمال محدودیت روی  $B_i$  را ادامه می‌دهیم چون که یکی از هدف‌های طراحی شده ما قابلیت انعطاف پذیری در انتخاب مدل‌های پیش‌زمینه است. (به عنوان مثال این مشکل وجود دارد که روی امتیازهای خروجی که SPM به یک دسته از قسمت‌های مجموعه  $S_i$  می‌دهد یک محدودیت قانونمند اعمال کرد).

فرض : ما فرض می‌کنیم که یک نمونه پیش‌زمینه در یک تصویر به نمایندگی از یک مجموعه بخش-های همجوار باشد. دویخش را همجوار می‌گوییم اگر کمترین فضای فاصله بین آنها در یک تصویر کمتر از یا برابر با  $\rho$  باشد. این فرض عقلانی است زیرا بیشتر پیش‌زمینه‌ها، علاقه به اشغال مناطق متصل در یک تصویر را دارند. روش ما اجازه فضای چندگانه را می‌دهد (برای مثال . چند سطل سیب در یک تصویر) که به عنوان ناحیه‌های متصل چندگانه در نظر گرفته می‌شوند.

فرض کنید که ما یک گراف مجاورت  $\mathcal{G}_i = (S_i, \mathcal{E}_i)$  ایجاد کنیم که هر بخش ( هر زیر مجموعه از  $S_i$  ) یک راس و  $(s_l, s_m) \in \mathcal{E}_i$  اگر  $\min(s_l, s_m) \leq \rho$  ( بعنوان مثال  $\rho = 5$  پیکسل ) برای همه  $s_l, s_m \in S_i$  (نمونه آن در شکل ۳ (c) آمده است) . آن وقت ، ناحیه متصل در یک تصویر می‌تواند به نمایندگی از زیر درخت های آن گراف باشند و به این صورت انتساب ناحیه‌های نهایی ممکن است جنگلی از زیر درخت ها باشد . به دنبال آن ، بدون از دست دادن کلیت کار، ما هر پیش-زمینه کاندید شده  $\beta_i \in B_i$  را به یک زیر درخت از گراف  $\mathcal{G}_i$  محدود می‌کنیم و هدف ما از انتساب ناحیه ، انتخاب بعضی از  $B_i$  است که به تنهایی قابل قبول نیستند ولی همچنان بیشترین امتیاز را از معادله یک دارند.

### ۳-۳-۲- تشکیل مجموعه کاندید ها

در این قسمت ما در مورد چگونگی ایجاد یک مجموعه پیش‌زمینه کاندید برای یک پیش-زمینه  $B_i^k$  بحث می‌کنیم که هر یک از آنها یک زیر درخت از گراف  $\mathcal{G}_i$  هستند .

فرض کنید در گراف مجاورت  $\mathcal{G}_i$  ، هر نمونه پیش‌زمینه، یک زیردرخت بسیار با ارزش مانند کاندید  $B_i^k$  باشد که با استفاده از جست‌وجوی *beam* با  $v^k$  به عنوان یک تابع هیروستیک و یک  $D$  به عنوان پهنای شعاع بدست آمده باشد (در تست ما  $D=10$ ). ما با تمام قسمت های مجزا شروع می‌کنیم تا  $\forall s \in S_i$  به  $B_i^k$  اضافه شود. در هر تکرار، ما تمام زیر درخت‌هایی که می‌توانند با اضافه

کردن یک یال به کاندیدهای قبل به دست آیند را نیز به شمار می آوریم. عرض پهنای  $D$  ( برای جستجوی *beam*) بیشترین مقدار زیر درخت که در هر دور باید حفظ شود را مشخص می کند.

ما فقط بیشترین  $D$  زیر درخت های با ارزش را به عنوان  $B_i^k$  نگه می داریم بدون این که روی آنهایی که ارزش کمی دارند زمان را تلف کنیم. در مرحله ۳ از الگوریتم یک مشاهده می کنید). در عمل، این جستجوی *beam* کاندیدای خوب و مناسب را انتخاب می کند زیرا هر پیش زمینه معمولا یک قسمت از تصویر را اشغال می کند. محاسبه زمان در این مرحله توسط پیش زمینه ای است که بیشترین  $O(D|S_i|^2)$  را دارد، و تعدادی از کاندیدهای پیش زمینه که  $|B_i|$  آنها بیشتر از  $(D|S_i|)$  باشد.

جدول ۱-۳ شبه کد مدل الگوریتم جستجوی BEAM برای تولید ناحیه های کاندید

Algorithm 1: Build candidates $B_i^k$ from $\mathcal{G}_i$ by beam search.
Input: (1) Adjacency graph $G_i = (S_i; \mathcal{E}_i)$ . (2) Value function $v_k$ of the $k$ -th foreground model. (3) $D$ : Beam width.
Output: $k$ -th foreground candidates $B_i^k$ .
1: Set the initial open set to be $\mathcal{O} \leftarrow \forall s \in S_i$ . $B_i \leftarrow \forall s \in S_i$ .
for $i = 1$ to $ S_i  - 1$ do
foreach $o \in \mathcal{O}$ do
2: Enumerate all subgraphs $\mathcal{O}_o$ that can be obtained by adding an edge to $o$ . $\mathcal{O} \leftarrow \forall \mathcal{O}_o$ and $\mathcal{O} \leftarrow \mathcal{O} / \mathcal{O}_o$ .
3: Compute values $v_o \leftarrow v^k(o)$ for all $o \in \mathcal{O}$ and remove $o$ from $\mathcal{O}$ if it is not top $D$ highly valued $B_i \leftarrow \forall \mathcal{O}$ .

### ۳-۳-۳- انتساب نواحی قابل تشخیص

فرض کنیم  $B_i$  را داریم، در این صورت ما می توانیم معادله ۱ را با انتخاب بعضی از کاندیدای قابل قبول تر از میان  $B_i$  ها حل کنیم. برای ایجاد یک راه حل قابل فهم ، ما اول یک قضیه در [۳۷] را مطرح می کنیم که فرمول بندی دوباره برای منطبق شدن با کار ما را شرح می دهد.

قضیه ۱ [۳۷]. برنامه نویسی پویا می تواند معادله اول را در بدترین حالت با  $O(|B_i||S_i|)$  حل کند اگر هر کاندیدا در  $B_i$  را بتوان با یک زیر گراف متصل از یک درخت  $T_i^*$  نشان داد.

قضیه ۱ ، یک الگوریتم دارای زمان خطی را برای انتساب ناحیه پیشنهاد می دهد، اگر  $B_i$  را بتوان مانند یک درخت سازماندهی کرد. در مجموعه کاندیدهای پیش زمینه  $B_i$  ، هر  $\beta_i \in B_i$  یک زیر درخت است اما تجمعی از  $\beta_i$  ها ممکن نیست. بنابراین ما بعضی از  $\beta_i$  ها را که باعث ایجاد حلقه می شوند ولی ارزش بالایی ندارد را رد می کنیم ، چون که راه حل نهایی یک جنگل از زیر درختان کاندید است. حرص  $B_i$  با  $B_i^*$  نشان داده می شود. در ادامه ما در مورد چگونگی به دست آوردن  $T_i^*$  و  $B_i^*$  از  $B_i$  بحث می کنیم.

استنتاج کردن درخت از مجموعه کاندید : با توجه به مجموعه کاندید ها  $B_i$  (یک مجموعه از زیردرختها)، هدف ما در اینجا بدست آوردن محتمل ترین درخت  $T_i^*$  است. می توان آن را مانند بیشترین برآورد احتمال (MLE) (maximum likelihood estimation) در یک روش مشابه به ساختار درخت یادگیری فرمول بندی کرد. [7]:

$$T_i^* = \operatorname{argmax} P(B_i|T) \quad , \quad T \in \quad \quad \quad ۳-۳$$

$$T(g_i)$$

اینجا  $P(B_i|T)$  داده احتمالی و  $T(g_i)$  مجموعه ای از همه درخت های پوشای ممکن روی گراف  $g_i$  می باشند. در تکمیل موارد بالا ، خلاصه کار ما یک الگوریتم ساده *chow-liu* است که بیشترین



احتمال درخت  $T_i^*$  از  $B_i$  را در زمان  $O(|B_i||S_i|^2)$  می دهد. این نیز ثابت شده که راه حل این الگوریتم ارزش  $\beta_l$  های مردود در  $B_i$  را تحت محدودیت ساختار درختی به صورت زیر کمینه می کند :

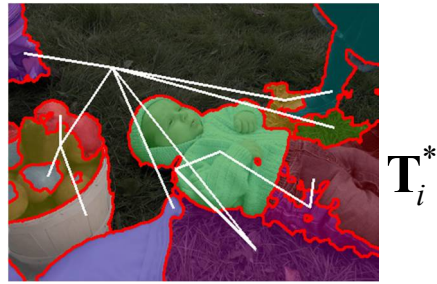
۴-۳

$$T_i^* = \underset{T \in T(g_i)}{\operatorname{argmin}} \sum_{B_l \in B_i, B_l \notin T_i} v(B_l)$$

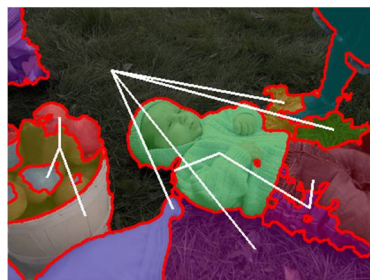
ما قبلاً یکبار  $T_i^*$  را بدست آورده بودیم و فقط کاندید های  $B_i^* \subset B_i$  را که زیر گرافی از  $T_i^*$  هستند نگه می داریم.

الگوریتم جستجو : همانطور که در قضیه ۱ مشخص شد ، راه حل بهینه در معادله ۱، با توجه به  $B_i^*$  می تواند موثر و کارا باشد. ما از برنامه نویسی پویا بر پایه الگوریتم جستجو برای اصلاح و تعدیل کردن الگوریتم CABOB<sup>۷۴</sup> استفاده می کنیم [۳۸]. الگوریتم CABOB یک الگوریتم جستجوی درختی بر پایه جستجوی اول سطح می باشد. این الگوریتم زیر درخت های از درخت ورودی استخراج می کند. نمونه ای از خروجی الگوریتم CABOB را در شکل ۴-۳ مشاهده می کنید.

<sup>74</sup> Combinatorial Auction Branch On Bids



CABOB algorithm



شکل ۳-۴ خروجی الگوریتم CABOB برای تصویر با برجسب کودک و سبد سیب از پایگاه داده **fliker**

### ۳-۴- شبه کد الگوریتم پیشنهادی

کل الگوریتم در الگوریتم ۲ خلاصه شده است. ما مدل کردن پیشزمینه و تخصیص ناحیه را تا زمانی انجام می‌دهیم که ارزش هدف نهایی برای تخصیص یک ناحیه جدید در معادله ۱ افزایش پیدا نکند و تا زمانی که این اتفاق نیافتد این تکرار متوقف نمی‌شود و تا آن زمان ما مدل کردن پیشزمینه را مانند یک جعبه سیاه مورد بررسی قرار می‌دهیم. به هر حال اگر ما فقط از مدل GMM به عنوان مدل پیش‌زمینه خودمان استفاده کنیم، الگوریتم تضمین می‌کند که حداقل در یک مینیمم محلی همگرا شود [۳۹].

Algorithm 2: Multiple foreground cosegmentation
<p>Input: (1) Input image set <math>I</math>. (2) Number of foregrounds (FGs) <math>K</math>.  (3) (In supervised case) annotations <math>A = \{A^1, \dots, A^K\}</math></p> <p>Output: Foregrounds <math>F = \{f^1, \dots, f^k\}</math> for all <math>I_i \in I</math></p> <p style="text-align: center;">Initialization</p> <p>foreach <math>I_i \in I</math> do</p> <p style="padding-left: 2em;">1: Oversegment <math>I_i</math> to <math>S_i</math> and build adjacency graph <math>\mathcal{G}_i = (S_i, \mathcal{E}_i)</math>  where <math>(s_l, s_m) \in \mathcal{E}_i</math> if <math>\min d(s_l, s_m) \leq \rho</math></p> <p>if unsupervised then</p> <p style="padding-left: 2em;">2: Apply diversity ranking of [13] to the similarity graph of  <math>S_i = \cup_{l=1}^M S_i</math> to find <math>K</math> regions <math>A = \{A^1, \dots, A^K\}</math> that are highly  repeated in <math>S</math> and diverse with respect to one another.</p> <p>3: Set <math>F \leftarrow A</math>.</p> <p style="text-align: center;">Iterative Optimization</p> <p>/* Stopping condition. */</p> <p>We stop the iteration if a new region assignment does not increase  the objective value (i.e. <math>\sum_{i=1}^M \sum_{k=1}^{k+1} v^k (f_i^k)</math> from Eq.(3-2)).</p> <p>/* Foreground Modeling (Any methods can be used). */</p> <p>foreach <math>k = 1: K</math> do</p> <p style="padding-left: 2em;">1: Learn GMM and SPM FG models from <math>F^k</math> (See Table 2).</p> <p>/* Region assignment */</p> <p>foreach <math>I_i \in I</math> do</p> <p style="padding-left: 2em;">foreach <math>k = 1: K</math> do</p> <p style="padding-left: 4em;">2: Generate FG candidates <math>B_j^k</math> by Alg. 1 as a set of <math>B_j^k = \langle K_j, C_j, W_j \rangle</math>,  where <math>K_j</math> is the foreground index, <math>C_j \subseteq S_i</math> is a subtree of <math>\mathcal{G}_i</math> and <math>W_j = v^k C_j</math>.</p> <p style="padding-left: 4em;">3: Compute the most probable candidate tree <math>T_i^*</math> and pruned <math>B_i^*</math> by Eq.(2) from  <math>B_i = \cup_{k=1}^{k+1} B_i</math></p> <p style="padding-left: 4em;">4: Obtain <math>F_i</math> to solve region assignment in Eq.(1) by using dynamic programming  on <math>B_i^*</math> (in the supplementary material).</p>

### ۳-۵- جمع بندی

همانطور که در طول فصل توضیح داده شده است در روش پیشنهادی از دو روش  $spm$  و  $Gmm$  به صورت ترکیبی برای استخراج ویژگی و یادگیری مدل شی پیش‌زمینه و پس‌زمینه استفاده کرده‌ایم و از روش  $sift$  برای تقسیم هر تصویر به زیر مجموعه‌های آن استفاده نمودیم و توانستیم به وسیله برنامه نویسی پویا مدل تخصیص ناحیه را بهینه کنیم. و در این راه از روش‌های مبتنی بر گراف حالت کمک گرفته‌ایم. تمام تلاش خود را کرده‌ایم تا عوامل طبیعی از جمله نویزهای موجود در تصویر و دیگر عوامل موجود سبب تاثیر منفی در خروجی الگوریتم نگردد. اما این‌ها جز جدا نشدی روش‌های پردازش تصویر می‌باشند. البته برای کارایی بهتر این الگوریتم‌ها می‌توان ابتدا یک پیش پردازش اولیه بروی تصاویر انجام داد. در فصل بعد نتایج حاصل از روش پیشنهادی و دیگر کارهای انجام شده را مورد ارزیابی قرار می‌دهیم.

فصل چهارم:

# نتایج آزمون‌ها

#### ۴-۱- مقدمه

پس از بیان روش پیشنهادی در فصل قبل، در این فصل به بررسی نتایج حاصل از روش خود می‌پردازیم. در ادامه ابتدا پایگاه داده‌ای که روش خود را بر روی آن مورد آزمایش قرار داده‌ایم معرفی می‌کنیم. و همچنین نتایج را به صورت مجزا در دو قسمت مدل کردن شی و تخصیص ناحیه مورد ارزیابی قرار می‌دهیم و نتایج را در قالب نمودار و تصاویر با سایر الگوریتم‌ها مورد مقایسه قرار می‌دهیم.

#### ۴-۲- معرفی پایگاه داده

ما از یک پایگاه داده استفاده می‌کنیم که Flickr نام دارد. این مجموعه داده به صورت کاملاً دستی در ۱۴ گروه دسته بندی شده است و هر گروه شامل ۱۰ تا ۲۰ تصویر می‌باشد. هر گروه شامل عکس‌هایی است که به صورت لحظه‌ای در جریان یک رویداد گرفته شده است همچنین یک یا چند شی در این تصاویر حضور دارند که مکرراً در بعضی از تصاویر حضور دارند در شکل ۴-۱ و شکل ۴-۲ نمونه‌ای از این تصاویر را مشاهده می‌کنید.



شکل ۱-۴ نمونه ای از تصاویر پایگاه داده Flickr با برچسب کودک و سبد سیب



شکل ۲-۴ نمونه ای از تصاویر پایگاه داده Flickr با برچسب گاو و چراگاه



## ۴-۳- نتایج روش پیشنهادی

همانگونه که در فصل سوم بیان شد اولین قدم در بحث هم‌بخش‌بندی تصویر، قطعه‌بندی و تقسیم هر تصویر به مجموعه‌ای از نواحی می‌باشد. قطعه‌بندی تصویر را می‌توان به وسیله الگوریتم‌های مختلفی انجام داد و ما از الگوریتم Sift استفاده می‌کنیم و مجموعه بخش‌های هر تصویر تشکیل می‌دهیم در ادامه و در شکل ۴-۳ نمونه‌ای از این تصاویر را نشان می‌دهیم.

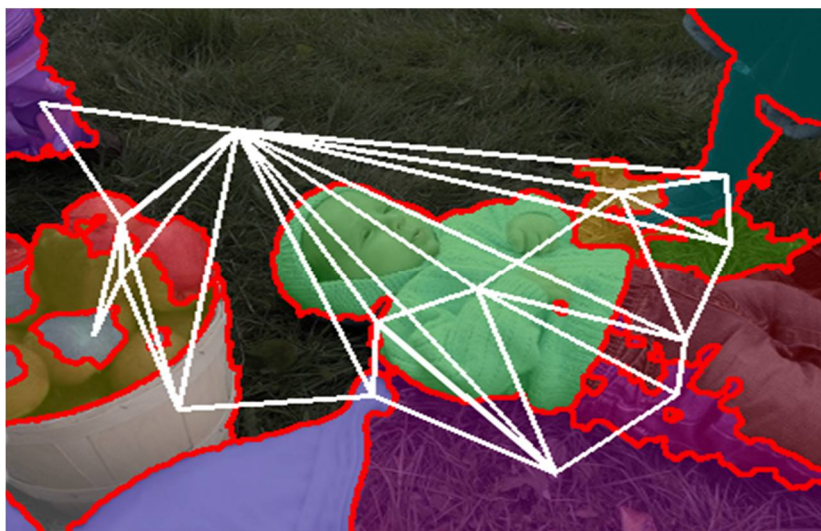


شکل ۴-۳ بخش‌بندی تصویر I با برچسب سبب سیب و نوزاد

در شکل ۴-۳ مشاهده می‌کنید که یک تصویر ابتدا به مجموعه‌ای از سگمنت‌ها افزای می‌شود و یک شی پیش‌زمینه از چند سگمنت تشکیل می‌شود.

در شکل ۴-۴ نحوه تولید گراف حالت یک تصویر را مشاهده می‌کنید. هر سگمنت به عنوان یک نود در نظر گرفته می‌شود. طبق همسایگی تعریف شده برای هر سگمنت اگر فاصله سگمنت‌های دیگر کمتر از یک تعداد پیکسل مشخص باشد بین آنها یک یال خواهد بود. بعضی از سگمنت‌ها هیچ همسایگی ندارند در نتیجه هیچ یالی از آنها خارج نمی‌شود.

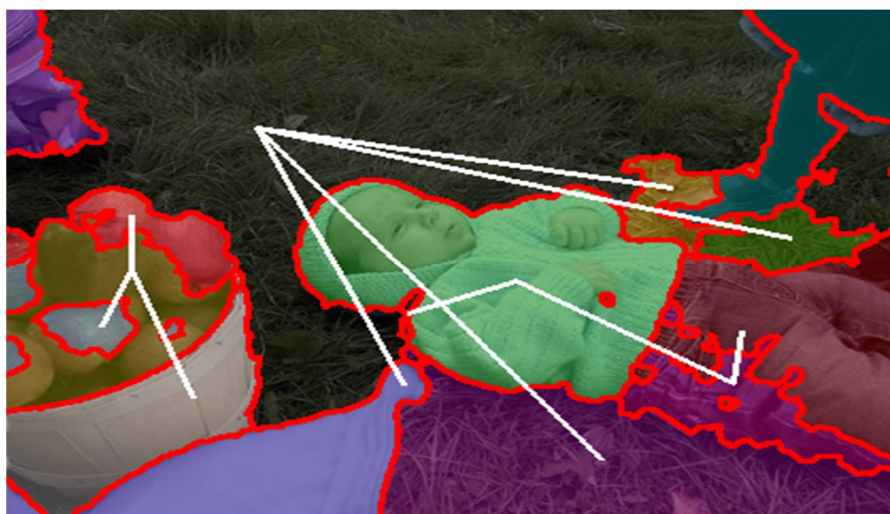




$$\mathbf{G}_i = (\mathbf{S}_i, \mathbf{E}_i)$$

شکل ۴-۴ نحوه تشکیل گراف حالت برای تصویر I با برچسب سبید سیب و نوزاد

با استفاده از الگوریتم جستجوی beam تمامی درختان پوشای گراف حالت تشکیل شده در شکل ۴-۴ را استخراج می‌کنیم این درخت‌های پوشای استخراج شده ورودی الگوریتم CABOB می‌باشند. نهایتاً با استفاده از این الگوریتم جنگلی از زیر درختان ایجاد می‌شود که در شکل ۴-۵ نشان داده شده است.



شکل ۴-۵ خروجی الگوریتم جستجو CABOB

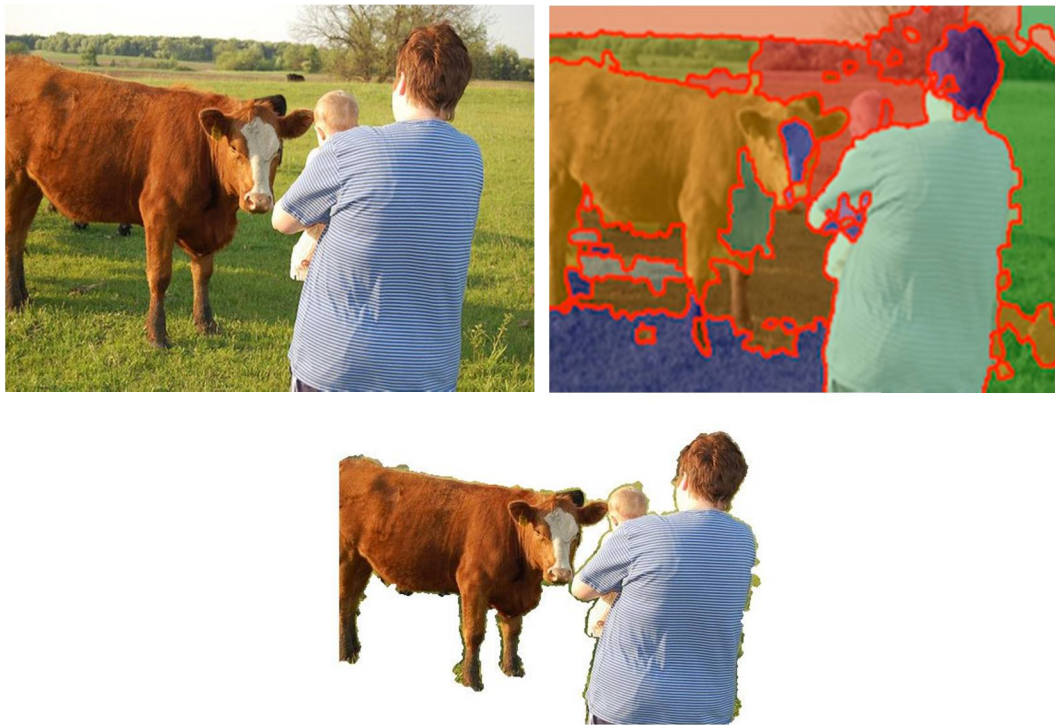
در شکل ۴-۶ خروجی نهایی را بعد از بهینه شدن یال‌های گراف حالت مشاهده می‌کنید.



شکل ۴-۶ خروجی الگوریتم پیشنهادی برای همبختبندی

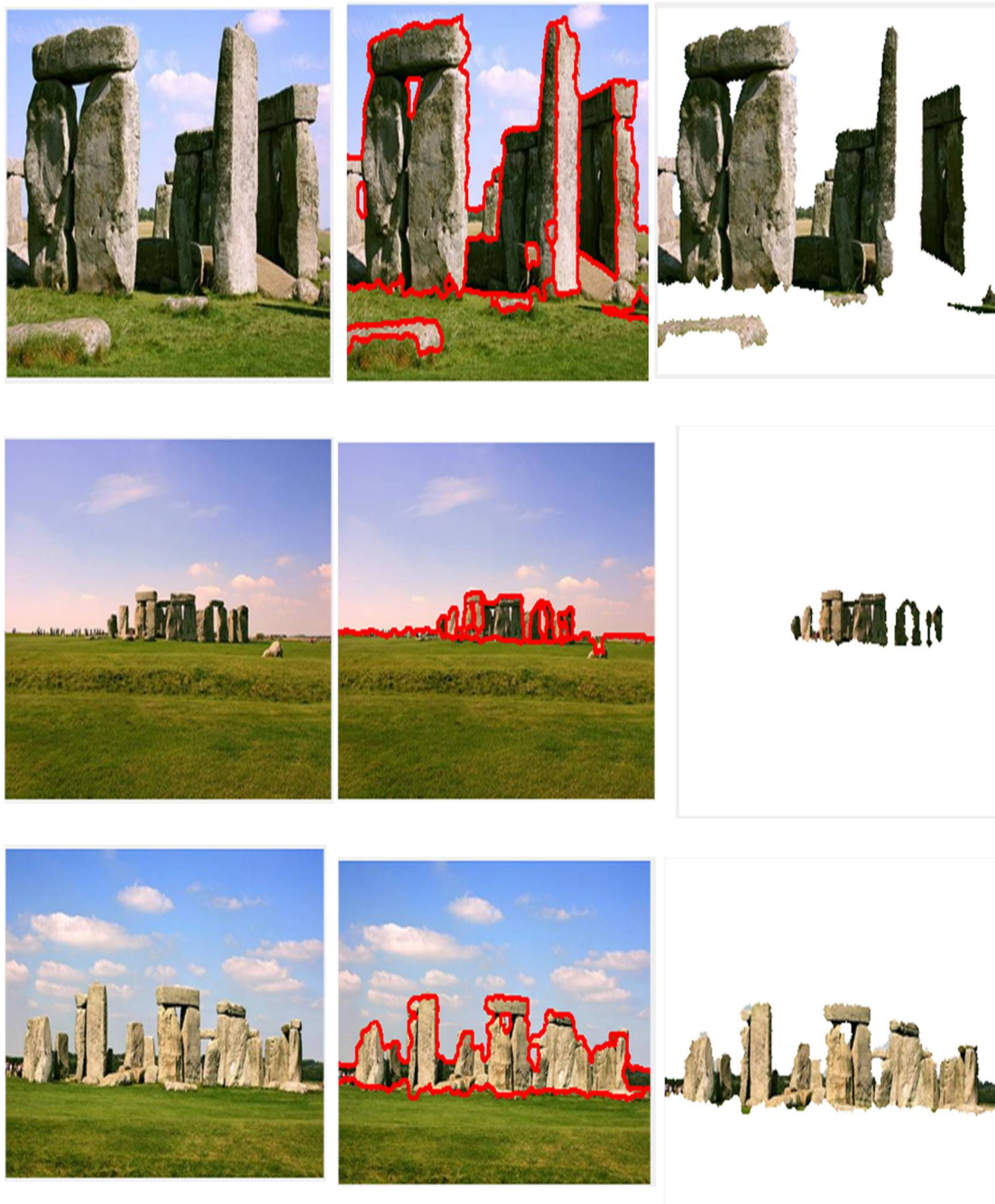
همان‌گونه که در شکل مشخص است مقداری از چمنزار به عنوان پیش زمینه سبد سیب و نوزاد شناخته شده است این خطای انجام شده مربوط به الگوریتم سگمنت کننده اولیه تصویر می باشد البته تشخیص لبه در این قسمت ها کمی دشوار است چون شدت پیکسل ها در این نقاط به آرامی تغییر می کند. همچنین کیفیت پایین تصاویر نیز می‌تواند دلیل دیگری بر این ضعف باشد.

در شکل ۴-۷ یک نمونه دیگر از سگمنت کردن تصویر را مشاهده می‌کنید. یک تصویر به چندین جز کوچکتر تقسیم می‌شود در این تصویر می‌توان دلیل استفاده از گراف حالت و الگوریتم‌های جستجو را توجیه نمود. همان‌گونه که در شکل ۴-۷ مشخص شده است این تصویر شامل ۴۸ بخش کوچک و بزرگ می باشد اگر بخواهیم تمام حالت‌های ممکن را برای معادله رابطه ۳-۱ در نظر بگیریم باید  $2^{48}$  را بررسی کنیم اگر این کار را برای چندین تصویر بخواهیم انجام دهیم مرتبه زمانی الگوریتم بسیار زیاد خواهد بود.



شکل ۴-۷ خروجی الگوریتم بخشبندی برای تصویر با برچسب گاو و چمنزار

در ادامه تعدادی از تصاویر یک مجموعه از تصاویر را در چند شکل نمایش می‌دهیم که حال خروجی الگوریتم ما می‌باشد. در این شکل‌ها ابتدا تصاویری را که می‌خواهیم همبخش‌بندی کنیم می‌آوریم و سپس خروجی الگوریتم سیفت نمایش داده می‌شود و در انتها نیز پیش‌زمینه‌های مورد نظر نیز نشان داده می‌شوند.



شکل ۴-۸ خروجی الگوریتم از مجموعه تصاویر با برچسب سنگ





شکل ۴-۹ خروجی الگوریتم پیشنهادی براری مجموعه تصاویر با برچسب پاندا

#### ۴-۴- مقایسه الگوریتم پیشنهادی با روش‌های موجود

ما الگوریتم خود را از نظر میزان تشخیص پیش‌زمینه‌ها یا همان اشیای موجود در تصویر با سایر روش‌ها مقایسه می‌کنیم. این کار را با مقایسه بروی هر دسته از تصاویر پایگاه داده، که در ابتدای فصل معرفی کرده‌ایم انجام می‌دهیم. این نتایج در جدول ۴-۱ مشخص شده است. نتایج موجود در جدول ۴-۱ بر اساس درصد درستی تشخیص پیش‌زمینه‌های یک مجموعه تصاویر توسط هر الگوریتم می‌باشد.

جدول ۴-۱ مقایسه نتایج الگوریتم پیشنهادی با سایر الگوریتم‌ها ( به درصد)

gorilla	stonehenge	fishing	dolphin	dog	cow	baseball	apple	
%۱۵	%۲۳	%۲۰	%۲۷	%۲۸	%۲۸	%۱۸	%۲۳	LDA
%۴۰	%۳۹	%۳۰	%۴۶	%۳۸	%۴۴	%۳۲	%۳۲	COS
%۲۸	%۳۲	%۲۰	%۳۵	%۳۶	%۳۹	%۳۴	%۴۱	DC
%۲۵	%۳۳	%۳۳	%۴۵	%۴۳	%۴۱	%۴۰	%۶۲	الگوریتم پیشنهادی

جدول ۴-۱ میانگین تشخیص درست پیش‌زمینه‌ها در پایگاه داده fliker

الگوریتم پیشنهادی	DC	COS	LDA	
				میانگین تشخیص درست
	٪۴۱	٪۳۹	٪۲۴	
الگوریتم پیشنهادی				٪۴۳

جدول ۴-۱ نمایشی از میزان درست تشخیص دادن الگوریتم پیشنهادی در مقایسه با سایر الگوریتم‌ها می‌باشد این نتایج به صورت درصد آورده شده است.

از آنجایی که عوامل زیادی در زمان اجرای یک الگوریتم تاثیر گذار است بنابراین مقایسه دو الگوریتم از نظر زمانی نیازمند ایجاد شرایط یکسان می‌باشد این شرایط یکسان شامل سخت افزار و سیستم عامل یکسان می‌باشد. همچنین با در اختیار نداشتن زمان اجرای دیگر الگوریتم‌ها، تنها زمان اجرای الگوریتم خود را برای هر مجموعه عکس می‌گذاریم

جدول ۴-۳ زمان اجرای الگوریتم پیشنهادی بر روی دسته‌های مختلف با تعداد ۱۰ تصویر بر حسب ثانیه

gorilla	stonehenge	fishing	dolphin	dog	cow	baseball	apple	
۱۲۱	۱۰۱	۱۰۴	۱۰۸	۱۱۸	۱۰۲	۱۰۵	۱۱۲	الگوریتم پیشنهادی

#### ۴-۵- جمع‌بندی

در این فصل سعی بر آن داشتیم تا با نمایش خروجی‌های الگوریتم پیشنهادی این پژوهش، دقت این الگوریتم را نشان دهیم. این کار را با به تصویر کشیدن خروجی‌های دیگر الگوریتم‌ها و الگوریتم روش پیشنهادی انجام داده ایم. همچنین به صورت آماری یکسری از مقایسه‌ها را بین روش این پژوهش و دیگر روش‌ها انجام داده‌ایم. اکثر الگوریتم‌های موجود از پیچیدگی و روی هم افتادگی

پیش‌زمینه‌ها و پس‌زمینه‌ها رنج می‌برند و این یک چالش برای آنها می‌باشد. هرچقدر اشیا موجود در تصویر قابل لمس‌تر باشند هم‌بخش‌بندی دقیق‌تر خواهد بود.



فصل پنجم:

# مشاوره و کارهای آینده

## ۵-۱- بحث و نتیجه‌گیری

در این پایان نامه الگوریتمی برای هم‌بخش‌بندی تصویر ارائه شده است تا علاوه بر اینکه زمان اجرای قابل قبولی داشته باشد بتواند نسبت به اسکیل شدن تصویر و تعداد تصاویر پایگاه داده مقاوم باشد و خروجی‌های قابل قبولی را ارائه دهد. اکثر روش‌های موجود از زاویه‌هایی که عکس از شی مورد نظر گرفته شده است رنج می‌برند و نمی‌توانند به خوبی عمل بخش‌بندی و جدا کردن شی از پس-زمینه را انجام دهند ما این کار این مشکل را به خوبی در الگوریتم خود حل کرده‌ایم و توانسته‌ایم نتایج مناسبی بگیریم.

همانگونه که در فصل‌های قبل توضیح داده شده است الگوریتم ارائه شده توسط این پایان نامه و اکثر الگوریتم‌های ارائه شده در زمینه کاری این پژوهش، از چند الگوریتم به صورت مرحله‌ای و تکراری برای گرفتن نتیجه مناسب استفاده می‌کنند. اگر بتوان الگوریتم مناسبی را با یکی از این الگوریتم‌ها جایگزین نمود مطمئناً می‌توان در کار این الگوریتم‌ها بهبود ایجاد کرد.

## ۵-۲- کارهای آینده

هر چند که الگوریتم ارائه شده توانسته نتایج خوبی را ارائه دهد ولی همچنان شکاف‌هایی در این الگوریتم و دیگر الگوریتم‌های موجود در این زمینه وجود دارد و این زمینه را فراهم می‌کند تا علاقه‌مندان در زمینه پردازش تصویر با ارائه الگوریتم‌های بهینه و مناسب این شکاف‌ها را پوشش دهند. در این راستا می‌توان از ما چند شکاف و تکنیک را که می‌توان از آن در این زمینه کمک گرفت را بیان می‌کنیم.

- استفاده از روش‌های پردازش سطح بالای تصویر به منظور شناسایی اشیای موجود در تصویر.
- استفاده از ویژگی‌های بافتی و معنایی به منظور شناسایی موقیت شی‌های پیش‌زمینه.

- استفاده از روش‌های یادگیری عمیق برای شناسایی پیش‌زمینه‌ها.

ما در ادامه کار خود به دنبال استفاده از روش‌های معنایی و ویژگی‌های سطح بالای تصویر می‌باشیم تا بتوان پیش‌زمینه‌هایی که از یک خانواده هستند را در یک مجموعه تصاویر پیدا و هم‌بخش‌بندی نمود. البته این کار مستلزم تحقیق و تلاش بیشتر می‌باشد.

- [1] A. Joulin, "Convex Optimization for Cosegmentation," Ph. D., École normale supérieure de Cachan-ENS Cachan, 2012.
- [2] J. C. Rubio, J. Serrat, A. Lopez, and N. Paragios, "Unsupervised Co-segmentation through Region Matching," in Computer Vision and Pattern Recognition (CVPR) IEEE Conference on, 2012, pp. 749-756.
- [3] C. Rother, T. Minka, A. Blake, and V. Kolmogorov, "Cosegmentation of Image Pairs by Histogram Matching - Incorporating a Global Constraint into MRFs," in Computer Vision and Pattern Recognition IEEE Computer Society Conference on, 2006, pp. 993-1000.
- [4] D. Batra, A. Kowdle, D. Parikh, J. Luo, and T. Chen, "Interactively Co-segmenting Topically Related Images with Intelligent Scribble guidance," International journal of computer vision, vol. 93, pp. 273-292, 2011.
- [5] L. Mukherjee, V. Singh, and J. Peng, "Scale Invariant Cosegmentation for Image Groups," in Computer Vision and Pattern Recognition (CVPR) IEEE Conference on, 2011, pp. 1881-1888.
- [6] S. Vicente, C. Rother, and V. Kolmogorov, "Object Cosegmentation," in Computer Vision and Pattern Recognition (CVPR) IEEE Conference on, 2011, pp. 2217-2224.
- [7] A. Torralba and A. A. Efros, "Unbiased Look at Dataset Bias," in Computer Vision and Pattern Recognition (CVPR) IEEE Conference on, 2011, pp. 1521-1528.
- [8] M. Fanman, L. Hongliang, L. Guanghui, and N. King Ngi, "Image Cosegmentation by Incorporating Color Reward Strategy and Active Contour Model," Cybernetics, IEEE Transactions on, vol. 43, pp. 725-737, 2013.
- [9] Chai, Y., Rahu, E., Lempitsky, V., Gool, L.V., Zisserman, A.: Tricos: A tri-level class-discriminative co-segmentation method for image classification. In: European Conference on Computer Vision. (2012)
- [10] Rother, C., Kolmogorov, V., Minka, T., Blake, A.: Cosegmentation of image pairs by histogram matching-incorporating a global constraint into mrfs. In: IEEE Conference on Computer Vision and Pattern Recognition. (2006) 993{1000

- [11] Zhu, H., Lu, J., Cai, J., Zheng, J., Thalmann, N.M.: Multiple foreground recognition and cosegmentation: an object-oriented crf model with robust higher-order potentials. In: IEEE Winter Conference on Applications of Computer Vision. (2014)
- [12] He, K., X. Zhang, S. Ren, and J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2015.
- [13] Ren, S., K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. in *Advances in Neural Information Processing Systems* .2015.
- [14] Uijlings, J.R., K.E. van de Sande, T. Gevers, and A.W. Smeulders, Selective search for object recognition. *International journal of computer vision*, 2015.
- [15] Felzenszwalb, P.F. and D.P. Huttenlocher, Efficient graph-based image segmentation. *International Journal of Computer Vision*, 2014.
- [16] Van den Bergh, M., X. Boix, G. Roig, and L. Van Gool, Seeds: Superpixels extracted via energy-driven sampling. *International Journal of Computer Vision*, 2011.
- [17] Manen, S., M. Guillaumin, and L. Gool. Prime object proposals with randomized prim's algorithm. in *Proceedings of the IEEE International Conference on Computer Vision*. 2013.
- [18] Rantalankila, P., J. Kannala, and E. Rahtu. Generating object segmentation proposals using global and local search. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014.
- [19] Carreira, J. and C. Sminchisescu, Cpmc: Automatic object segmentation using constrained parametric min-cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2012.
- [20] Harada, T., Y. Ushiku, Y. Yamashita, and Y. Kuniyoshi. Discriminative spatial pyramid. in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. 2011. IEEE.
- [21] Endres, I. and D. Hoiem, Category-independent object proposals with diverse ranking. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2014.

- [22] Krähenbühl, P. and V. Koltun, Geodesic object proposals, in Computer Vision–ECCV 2014. 2014, Springer.
- [23] Alexe, B., T. Deselaers, and V. Ferrari, Measuring the objectness of image windows. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2012.
- [24] Rahtu, E., J. Kannala, and M. Blaschko. Learning a category independent object detection cascade. in *Computer Vision (ICCV), 2012 IEEE International Conference on*. 2012. IEEE.
- [25] Cheng, M.-M., Z. Zhang, W.-Y. Lin, and P. Torr. BING: Binarized normed gradients for objectness estimation at 300fps. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2014.
- [26] Zitnick, C.L. and P. Dollár, Edge boxes: Locating object proposals from edges, in *Computer Vision–ECCV 2014*. 2014, Springer.
- [27] Yan, J., Y. Yu, X. Zhu, Z. Lei, and S.Z. Li. Object detection by labeling superpixels .in *Proceedings of image features*. in *Tenth IEEE International Conference on Computer Vision*. 2014.
- [28] Csurka, G., C. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. in *Workshop on statistical learning in computer vision, ECCV*. 2004. Prague.
- [29] Nowak, E., F. Jurie, and B. Triggs. Sampling strategies for bag-of-features image classification. in *European conference on computer vision*. 2006. Springer.
- [30] Grauman, K. and T. Darrell. The pyramid match kernel: Discriminative classification with sets of image features. in *Tenth IEEE International Conference on Computer Vision (ICCV Volume 1)*. 2005. IEEE.
- [31] Lazebnik, S., C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. 2006. IEEE.
- [32] Yu, K., T. Zhang, and Y. Gong. Nonlinear learning using local coordinate coding. in *Advances in neural information processing systems*. 2009.

- [33] Wang, J., J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. in Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. 2010. IEEE.
- [34] Van Gemert, J.C., C.J. Veenman, A.W. Smeulders, and J.-M. Geusebroek, Visual word ambiguity. IEEE transactions on pattern analysis and machine intelligence, 2010.
- [35] Bo, L. and C. Sminchisescu. Efficient match kernel between sets of features for visual recognition. in Advances in neural information processing systems. 2009.
- [36] Schölkopf, B., A. Smola, and K.-R. Müller, Nonlinear component analysis as a kernel eigenvalue problem. Neural computation, 1998.
- [37] T. Sandholm and S. Suri. BOB: Improved Winner Determination in Combinatorial Auctions and Generalizations Artificial Intelligence 2003 .
- [38] T. Sandholm, S. Suri, A. Gilpin, and D. Levine. CABOB: A Fast Optimal Algorithm for Winner Determination in Combinatorial Auctions. Manage. Sci., 51:374–390, 2005. 5
- [39] C. Rother, V. Kolmogorov, and A. Blake. GrabCut – Interactive Foreground Extraction using Iterated Graph Cuts. In SIGGRAPH ,2004 .

## **Abstract**

Image segmentation and detection of objects in pictures have always been important tasks in artificial vision. The segmentation of similar sections -or similar objects- between two pictures or a collection of pictures is called cosegmentation. The similar sections shall contain a rigid object, a non-rigid object, a scene, or similar objects of the same category. Objects can also be segmented from different viewpoints. Current paper studies background cosegmentation in which there are a number of similar backgrounds in picture collections though every picture contains a subset of the backgrounds. Because of the simpler assumption, this issue is in contrast to the classic cosegmentation problem in most available algorithms. Current study offers an optimal procedure for cosegmentation of multiple backgrounds that assigns no hypothesis on the backgrounds and surpasses limitations of previous procedures caused error. Current study used the same pattern which alternately runs a process of background modeling and a process of areal allocation modeling. These two processes both have a good optimality. The proposed algorithm, in general, is considered flexible enough for advanced categorization of any background or specified area through a complex selection procedure. The proposed algorithm offers several astonishing features as linear complexity and accuracy in detection.

Keyword:

Cosegmentation , multiple foreground , region proposal , spatial pyramid matching, state graph





**Faculty of Computer Engineering**  
**M.Sc. Thesis in Artificial Intelligence Engineering**

**An efficient algorithm for image co-segmentation**

**By:**

**Mohammad Saleh Mollae Pashae**

**Supervisor:**

**Dr Ali Akbar Pouyan**

**Advisor:**

**Dr Mansour Fateh**

**August 2017**