**دسته بندی تصاویر عمومی با استفاده از یادگیری مبتنی بر درخت: یک مطالعه تجربی گسترده**

**چکیده**

در این مقاله بدون هیچ دانش قبلی از گروه های تصویر[[1]](#footnote-1)، مساله عمومی دسته بندی تصاویر بررسی می شود. ما انواع مختلف روش های مبتنی بر یادگیری نظارتی را بررسی می کنیم که از گام های مشابهی برای استخراج زیرپنجره های تصادفی استفاده می کنند. این زیرپنجره ها توسط پیکسل های سطری با تراکم بالا تشریح می شوند و با ترکیب درختان بی نهایت تصادفی، مستقیما می توان تصادیر را دسته بندی کرد یا ویژگی های تصویر را آموزش داد. تاثیر پارامترها و نسخه های مختلف این روش به طور کامل مورد بررسی قرار می گیرد و بنابراین، خط مبنا و راهنمایی برای بررسی های آتی ارائه می شود. نتایج دقیق برای 80 مجموعه داده در دسترس عمومی ارائه شده است که انواع مختلف تصاویر را نشان می دهد (شامل بیش از 3800 گروه تصویر و بیش از 1.5 میلیون تصویر).

1. **مقدمه**

هدف از دسته بندی نظارتی تصاویر، طراحی خودکار مدل های کامپیوتری است تا بعد از آموزش با مجموعه ای از تصاویر برچسب دار، به صورت دقیق بتوان گروه تصاویر جدید را (از بین گروه های از پیش تعریف شده ی قبلی) پیش بینی کرد. برای حل این مساله ی عمومی در دنیای واقعی، عملیات مشخصی صورت می گیرد، همچون تشخیص خودکار تصاویر حروف دستخط، تشخیص چهره، سلول ها و علائم خیابان و موارد دیگر.

در آغاز فعالیت های تصویرسازی کامپیوتری، هنگامی که یک محقق یک روش جدید برای دسته بندی تصاویر ارائه میداد، اغلب یک الگوریتم اختصاصی طراحی کرده بود که با استفاده از دانش پیشین افراد، دنباله ای از عملیات مشخص را پیاده سازی می کرد که روش دستی نیز نامیده می شود. این روش ها اغلب شامل طراحی و محاسبه ی فیلترهای مناسب و نیز ثبت ویژگی های موردانتظار از مشخصه های مختلف تصاویر هستند. در زمینه ی برنامه های کاربردی مورد نظر ما، تصویربرداری علوم زیستی، هرچند مقالات خاص و کارآمدی ارائه شده است، اما گزینه های طراحی به ندرت واضح هستند و بنابراین، این راهکارها برای هر مساله ی خاص، به تحقیق و توسعه ی زیادی نیاز دارند. در صورتی که پارامترهای مساله متغیر باشند (مثلا پروتکل های آماده سازی نمونه، تعیین مدالیته تصاویر، تشخیص فنوتیپ ها و ...)، تنظیمات بیشتری مورد نیاز است. به عبارت دیگر، این روش های مهندسی را نمی توان برای مقیاسی برابر صدها هزار موجودیت زیستی مورد استفاده قرار داد. زیرا برای نمایش آنها از تکنیک های مختلف نمونه برداری و مدالیتی تصاویر استفاده می شود. بنابراین بررسی های علمی معمولا محدودیت مقیاس (اندازه) دارند و یا همچنان به صورت دستی و جزئی عمل می کنند (مثلا در پروژه مبتنی بر وب GalaxyZoo، 50 میلیون تصویر به صورت دستی توسط 150 هزار نفر در مدت 1 سال برچسب گذاری شده و در گروه های ریخت شناسی (مورفولوژیکی) قرار داده شدند). روش های دیگر به زیربنای محاسباتی بسیار عظیمی نیاز دارند، زیرا مبتنی بر محاسبات ویژگی های متراکم هستند (مثلا کامپیوترهای مورد استفاده در پروژه ی Help Conquer Cancer Project ، که بیش از 100 میلیون CPU برای دسته بندی خودکار 10 میلیون تصویر آزمایش تبلور پروتئین به کار گرفته شدند که به طور متوسط در هر روز، 55 CPU مورد استفاده قرار گرفته است (کاتسروبا و همکاران[[2]](#footnote-2) (2012))).

**1-1-این مقاله**

در این مقاله براساس تحقیقات قبلی (ماری[[3]](#footnote-3) و همکاران، 2003 و 2004 و 2005 و 2007) و با توسعه ی آنها، بدون هیچ پیش فرضی از دسته ی تصویر[[4]](#footnote-4) یا موارد دیگر، مساله ی عمومی دسته بندی تصاویر به شیوه ی نظارتی[[5]](#footnote-5) را بررسی می کنیم که شامل تشخیص انواع مختلف تصاویر، تحت شرایط متفاوت تصویربرداری است. در واقع با طراحی ذهنی یک دسته بند همه منظوره ی تصاویر که به سادگی قابل اجرا باشد، ابتدا یک روش یادگیری فرم گرا[[6]](#footnote-6) ارائه شده است که به صورت تصادفی و با تراکم بالا، تعدادی زیرپنجره[[7]](#footnote-7) را از تصویر استخراج می کند و سپس با استفاده از مقادیر پیکسل های سطری، آنها را تشریح می کند و برای دسته بندی این زیرپنجره ها (و بنابراین دسته بندی تصویر) از مجموعه درختان بی نهایت تصادفی استفاده می کند. این راهکار با وجود سادگی مفهومی و پیچیدگی نسبتا اندک در زمان اجرا، نتایج جالبی را برای چند مجموعه داده به ارمغان آورده است. سپس در مقالات (موسمن[[8]](#footnote-8) و همکاران (2008)، ماری و همکاران (2009)، دومونت[[9]](#footnote-9) و همکاران (2009) و استرن[[10]](#footnote-10) و همکاران (2011)) برای طبقه بندی، تقسیم بندی تصویر، تشخیص نقاط مورد نظر و بازیابی تصاویر براساس محتوا، روش های متنوعی پیشنهاد شده است.

در این مقاله، یک چهارچوب[[11]](#footnote-11) کلی برای دسته بندی تصاویر ارائه داده و آن را به صورت کامل ارزیابی می کنیم. فعالیت های مهم ما عبارتند از:

* استخراج زیرپنجره ها و استفاده از درختان شدیدا تصادفی، مهمترین بخش چهارچوب پیشنهادی ما است که در تحقیقات قبلی نیز مطرح شده بود و الگوریتم های متنوعی ارائه شده که نیازمند تست و بررسی هستند. به ویژه ما در این مقاله، با بسط مقاله ی (موسمن و همکاران (2008)) چندین روش جدید برای یادگیری ویژگی ها مطرح می کنیم و در نتیجه راه های مختلفی برای استخراج ویژگی ها از درختان ارائه می شود. همچنین محدوده ی برخی پارامترهای نامشخص (همچون فاصله و اندازه ی زیرپنجره ها) را بررسی می کنیم که برای مجموعه داده های مختلف می تواند سودمند باشد. در نتیجه این الگوریتم های جدید و متنوع در چهارچوب پیشنهادی ما می توانند عملیات دسته بندی را بر روی طیف وسیعی از تصاویر انجام دهند و جامعیت طرح ما را بهبود بخشند.
* برای ارزیابی چهارچوب پیشنهادی، کارایی 80 مجموعه داده ی در دسترس عمومی (شامل 25 مجموعه داده ی زیستی[[12]](#footnote-12)) را به صورت منظم و گسترده بررسی می کنیم. با انجام این تحقیق بزرگ می توان کارایی این روش ها و نسخه های جدید آنها را ارزیابی نمود تا تاثیر طرح دسته بندی و پارامترهای آن مشخص شود و سپس، قدرتمندترین طرح انتخاب می شود و دستورالعمل های عمومی برای استفاده های بعدی طراحی می شود. بنابراین، از این طرح با سرعت بالا برای حل مسائل بعدی استفاده می شود.
* تا جایی که می دانیم، تا کنون هیچ یک از روش های دسته بندی تصویر به این وسعت مورد ارزیابی قرار نگرفته است. عمیقا معتقدیم که برای ارزیابی جامع و منصفانه ی روش های عمومی لازم است که عملیات مختلفی توسط آنها اجرا شود و با خلاصه سازی پایگاه داده های عمومیِ در دسترس و با ارائه ی نتایج مثبت و منفی، تاثیر پارامترها مورد بررسی قرار گیرد و بنابراین امیدواریم که با ارائه ی این پژوهش، محققان دیگر نیز تشویق شوند که روش های خود را در طیف وسیعی از تصاویر مورد ارزیابی و مقایسه قرار دهند.

**2- تنظیمات آزمایش**

ما با مجموعه داده های بزرگ و متنوعی در محدوده کاربردهای مختلف کار می کنیم. فرض ما این است که با بررسی کلی مساله ی دسته بندی تصاویر، می توان راهکارهای عمومی ارزشمندی ارائه کرد، که در زمینه های مختلف قابل استفاده هستند. به عنوان مثال، مشاهدات تجربی در رابطه با تشخیص علائم ترافیک (ثبت شده بر روی دوربین های رومیزی) یا تصاویر کهکشان ها (حاصل از بررسی گسترده ی تصاویر فضایی)، می تواند برای تشخیص سلول ها (ی ثبت شده توسط میکروسکوپ) مفید باشد. زیرا این مجموعه داده ها، برخی ویژگی های ضروری را به صورت مشترک دارا هستند (و شامل گروه های مختلفی از شکل ها هستند و به دلیل متفاوت بودن نوع روش کسب اطلاعات، تغییرات نویز و شدت نور متفاوتی دارند). به طور مشابه، مشاهدات حاصل از مجموعه داده های دسته بندی مواد نیز برای تشخیص بافت های زیستی جالب توجه است (زیرا تصاویر آنها الگوهای بافتی دارند).

**2-1- معیار ارزیابی و مجموعه داده ها**

تنظیمات آزمایش ما شامل 80 مجموعه داده ی تصویر است که قبلا منتشر شده اند و به صورت عمومی و رایگان در دسترس هستند و مجموعا شامل حدود 1.5 میلیون تصویر هستند که 3850 گروه متمایز را به نمایش می گذارند. انتخاب مجموعه داده ها به صورت پیش فرض صورت گرفته است و ارتباطی با نتایج حاصل از روش پیشنهادی ما ندارد. جزئیات بیشتر این مجموعه داده ها در بخش تکمیلی ارائه شده است. به ویژه خلاصه ای از مشخصه های آنها در جدول تکمیلی 1 ارائه شده و مروری بر گروه های تصویر برای تمام مجموعه داده ها در شکل های تکمیلی 1 و 2 و 3 و 4 مطرح شده است. این تصاویر با استفاده از تجهیزات تخصصی و با تنظیمات آزمایش، در شرایط کنترلی و غیرکنترلی از سراسر جهان به دست آمده اند و از یک دوربین دیجیتال در دنیای واقعی و چندین ابزار تصویرسازی طبی-پزشکی (میکروسکوپ فلوئورسانس یا میدان روشن، فیلم رادیولوژی ساده و غیره)، تلسکوپ رباتیک، رادار دهانه ترکیبی و غیره استفاده شده است. برای هر مجموعه داده، گروه های تصویر می توانند تغییرات ظاهری کوچک یا بزرگ را نمایش دهند. زیرا منابع مختلف و سطوح متنوعی وجود دارد، که می توانند تغییرات بالقوه ای در مکان، شدت نور، مقیاس، و دیدگاه و/یا وجود پس زمینه ی شلوغ، ازدحام و نویز ایجاد کنند. به علاوه، تغییرات زیاد بین کلاس ها یا شباهت زیاد بین کلاس های متمایز را می توان مطرح کرد. چندین مدل از این مجموعه داده ها مصنوعی هستند و بنابراین تغییرات آنها تحت کنترل است و به خوبی توصیف می شوند، در حالی که برخی مجموعه داده های دیگر حاوی تصاویر واقعی هستند و بنابراین تغییرات آنها ترکیب می شوند. توجه شود که ما در آزمایشات خود فقط از دو مجموعه داده ی تشخیص چهره ی پرکاربرد (از بین 10 مجموعه داده) استفاده می کنیم که این مجموعه داده ها اخیرا در مقالات (هوانگ و همکاران (2007)، شمیر (2008) و پینتو و همکاران (2008)) به صورت خلاصه مطرح شده و مورد ارزیابی کامل قرار گرفته اند. همچنین ما از مجموعه داده چالش پاسکال VOC (اورینگهام و همکاران (2010)) استفاده نمی کنیم. زیرا معیار ارزیابی آن، دقت/ منحنی فراخوانی برای هر گروه از اشیاء است و در چهارچوب ارزیابی ما قرار نمی گیرد.

پروتکل های ارزیابی ما به صورت خلاصه در جدول تکمیلی 1 ارائه شده اند. معیار ارزیابی ما، نرخ خطای دسته بندی نادرست برای تصاویر تست جداگانه است. در صورتی که در مقالات، یک پروتکل دقیق برای مجموعه داده ها تعریف شده و مورد پذیرش قرار گرفته باشد، ما نیز از آن استفاده می کنیم. هرچند برای اغلب مجموعه داده ها (مثل مجموعه داده هایی که پروتکل آنها به دقت تشریح نشده است، یا مجموعه داده هایی که در مقالات مختلف با هم تفاوت دارند، یا آنهایی که تعداد تصاویر تست آنها نسبتا اندک است)، 10 آزمایش را اجرا می کنیم که هر اجرا از تعداد خاصی تصاویر تصادفی برای مرحله ی یادگیری استفاده می کند (مثل 80 درصد از تعداد کل تصاویر) و تصاویر باقی مانده برای مرحله ی تست مورد استفاده هستند (مثلا 20 درصد تصاویر). سپس میانگین نرخ خطای دسته بندی نادرست برای تمام مجموعه تست ها محاسبه می شود که در نتیجه به صورت قابل اعتماد می توان تاثیر پارامترها در این روش را بررسی کرد.

**3- روش ها**

ما دو مولفه ی مهم از متغیرهای دسته بندی تصویر خود را ارائه می دهیم. این روش شامل استخراج زیرپنجره های تصادفی است که توسط مقادیر پیکسل های سطری تشریح می شوند و از ابزارهای مختلف برای ترکیب درخت های بی نهایت تصادفی استفاده می شود.

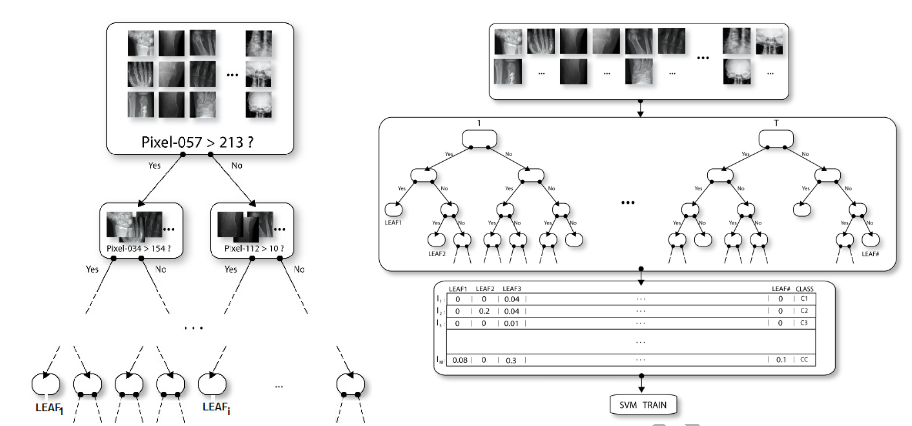
**3-1- زیرپنجره های تصادفی**

ما قبلا طرح های مختلف نمونه برداری زیرپنجره های تصادفی را تعریف کرده ایم. زیرپنجره های تصادفی قطعات مربعی با اندازه ی تصادفی هستند که از مکان های تصادفی تصویر استخراج می شوند. سپس به قطعه ای با اندازه ی ثابت تبدیل می شوند و از پیکسل ها به عنوان ورودی الگوریتم یادگیری ماشین استفاده می شود (زیربخش بعد را ببینید). با تغییر اندازه ی قطعات، قدرت آنها برای تغییر مقیاس افزایش می یابد و می توان از آنها در روش های عمومی یادگیری ماشین استفاده کرد که از بردارهای ویژگی با اندازه ثابت استفاده می کنند. همچنین در این روش، برای نمونه برداری از مجموعه آموزشی زیرپنجره ها با پیکسل های اندک و متغیرهای شدت استفاده می شود. این پروسه می تواند به یادگیری الگوریتم کمک کند، به گونه ای که در مقابل تغییرات طبیعی در تصاویر تست دیده نشده مفید باشد. همچنین این تفاوت ها شامل فعالسازی زاویه چرخش راست و مستقیم و معکوس کردن زیرپنجره ها است و بنابراین این مدل می تواند به گونه ای آموزش ببیند که در مقابل چرخش، مقاوم باشد.

در این مقاله ابتدا به صورت دقیق، تاثیر فاصله زیراندازه ها و روش های کدگذاری تصادفی آنها در تمامی 80 مجموعه داده را بررسی می کنیم. مجموعه های پیش فرض با استفاده از یک میلیون زیرپنجره ی آموزشی (از بین 100 هزار زیرپنجره) ساخته می شوند و تست های گسترده ای بر روی بیش از 50 میلیون زیرپنجره انجام می شود. برای هر مجموعه داده، همین تعداد زیرپنجره به صورت تصادفی از تصویر استخراج می شود که برابر است با که در این جا تعداد تصاویر آموزشی است. در این جا زیرپنجره ها به صورت متن پیکسلی نشان داده می شوند و مناطق پشتیبان یا فیلدهای پذیرنده ی مقیاس ها/ اندازه های مختلف، به صورت منظم تست می شوند: ما یک پیکسل 1\*1 را به عنوان مبنا در نظر می گیریم و 13 پیکربندی مختلف از زیرپنجره های مربع را در محدوده ی مناطق کوچک تصویر (0 تا 10 درصد) تا مناطق بزرگ (90 تا 100 درصد) مدنظر قرار می دهیم که اندازه های نامحدود پیش فرض (0 تا 100 درصد) در مقاله (ماری و همکاران در 2005) مورد استفاده قرار گرفته است. محدود کردن اندازه ها مثلا به میزان (25 تا 50 درصد) بدین معناست که اندازه ی هر زیرپنجره به صورت تصادفی، بین 25 تا 50 درصد مینیممِ (عرض و ارتفاع) تصویر انتخاب می شود و سپس مکان آن نیز به صورت تصادفی انتخاب می شود تا تضمین شود که زیرپنجره های مربعی همواره در محدوده ی تصاویر قرار دارند. توجه شود که در پیکربندی با مینیمم صفر ، مینیمم اندازه دقیقا برابر است با 1\*1. برای تمامی پیکربندی ها (به جز پیکربندی مبنای 1\*1 که سایزبندی مجدد انجام نمی شود)، هر زیرپنجره با الحاق دوسویه به یک قطعه با اندازه ثابت (8\*8 و 16\*16 (پیش فرض) و 32\*32) تبدیل می شود و مقادیر پیکسل های آن به سطوح خاکستری یا HSV کدگذاری می شوند و به عنوان توصیفگر زیرپنجره مورد استفاده قرار می گیرند. در حالی که نیازمند ارائه ی طرح های نمونه برداری خاص یا دقیق تری هستیم تا نتایج برای مجموعه داده های خاصی بهبود یابد (مثلا نمونه برداری محلی برای مجموعه داده هایی که در الگوی موردنظر مشخص هستند، زیرپنجره های مستطیلی برای اشیای باریک بلند، نمونه برداری تطبیقی (موسمن و همکاران در 2008) و ...)، ما در اینجا میخواهیم بررسی کنیم که نمونه برداری تصادفی عمومی، منظم و اولیه، از نظر دقت بر روی مجموعه داده های مختلف به چه صورت عمل می کند تا بتوان قبل از توسعه ی طرح های نمونه برداری پیچیده تر، یک راهکار مبنا ارائه نمود.

**3-2- درخت های بی نهایت تصادفی برای دسته بندی مستقیم تصاویر یا آموزش ویژگی ها**

کاربرد درختان تصادفی در یادگیری ماشین و تصویرسازی کامپیوتری رو به افزایش است. الگوریتم Extra-Trees در مقاله (گرتس و همکاران در 2006) پیشنهاد شده است که خواننده می تواند توصیف دقیق الگوریتم را در آنجا ببیند. در این مقاله، ما با استفاده از دو ابزار مختلف یعنی دسته بندهای مستقیم (ET-DIC) و آموزش ویژگی (ET-FL) کاربرد درختان بی نهایت تصادفی را ارزیابی می کنیم. به این دلیل که هنگام استفاده از آموزش ویژگی ها، کارایی کلی بهتری به دست آمد، ما فقط در اینجا این نسخه ی خاص را بررسی می کنیم و نسخه های دیگر در بخش تکمیلی ارائه می شوند. در طرح دسته بندی ET-FL، به جای نگهداری مقادیر تخمین احتمال در گره های پایانی و استفاده از درخت ها برای دسته بندی زیرپنجره ها (و بنابراین دسته بندی تصاویر)، هر گره پایانی (برگ) درخت به عنوان یک "کدبوک"[[13]](#footnote-13) یا "کلمه ی بصری"[[14]](#footnote-14) در نظر گرفته می شود. این روش با الهام از مقالات قبلی، از کدبوک های بصری استفاده می کند. در این تنظیمات، بعد از انتشار زیرپنجره ها به پایین درخت، هر تصویر با یک بردار ویژگی عمومی تشریح می شود که از نظر ابعادی، معادل با تعداد گره های پایانی در کل درخت ها است و ویژگی ها دارای مقادیر کمّی هستند (و متناظر با تعداد زیرپنجره های تصویر هستند که به گره پایانی رسیده اند/ تقسیم بر تعداد کل زیرپنجره های استخراج شده از تصویر و در نتیجه یک مقدار دودویی در محدوده [0, 1] دارند و مجموع تمامی گره های پایانی برای هر درخت در هر تصویر برابر 1 است). این "گروه ویژگی ها" وارد دسته بند می شود تا مدل نهایی دسته بندی تصاویر ساخته شود. در بررسی های ما، از یک دسته بند خطی ماشین بردار پشتیبان استفاده شده است، همان طور که در شکل 1 می بینیم. برای پیش بینی گروه یک تصویر جدید، زیرپنجره های تصادفی آن به کل درخت ها انتشار می یابند تا بردار ویژگی عمومی آن ساخته شود و سپس گروه بندی با استفاده از دسته بند SVM انجام می شود.



**شکل 1: سمت چپ: درخت حاصل از مجموعه آموزشی زیرپنجره های تصادفی که برای طرح ET-FL از تست گره ها با مقادیر آستانه ی پیکسل استفاده میکند. راست: ترکیبی از درخت T، نمایش عمومی استنتاجی با اندازه های کمّی برای آموزش تصاویر و آموزش یک دسته بند نهایی خطی SVM در مد ET-FL**

در هر دو حالت، ما به صورت دقیق 80 مجموعه داده را بررسی کرده و با انتخاب تعداد اندکی از مقادیر احتمالی (از 1 تا 100 در ET-DIC و از 1 تا 50 هزار در ET-FL)، تعداد تست های تصادفی k (از 1 تا ماکسیسمم تعداد متغیرهای ورودی) و تعداد درخت ها T (از 1 تا 20 در ET-DIC و از 1 تا 40 در ET-FL، هرچند که در تست های گسترده، مقدار T بیش از 1000 درخت است)، تاثیر مینیمم اندازه ی نمونه گره را مورد تحقیق قرار می دهیم. همچنین در ET-FL به صورت دقیق، تاثیر کدگذاری بردار ویژگی عمومی را بررسی می کنیم: ما نمایش عددی و کدهای دودویی را ارزیابی می کنیم (مقدار آنها برابر 1 است اگر حداقل یکی از زیرپنجره ها به گره پایانی انتشار یافته باشد، و در غیراینصورت مقدار آن صفر است) که آیا در گره های پایانی درخت قرار دارند و یا در تمامی گره های درخت (گره های داخلی و گره های پایانی) هستند. ما در ET-FL از یک دسته بند خطی SVM برای انجام دسته بندی نهایی استفاده می کنیم که پارامترهای آن به مقادیر پیش فرض تنظیم شده اند (برای مشاهده ی جزئیات دقیق پیاده سازی به بخش تکمیلی مراجعه نمایید).

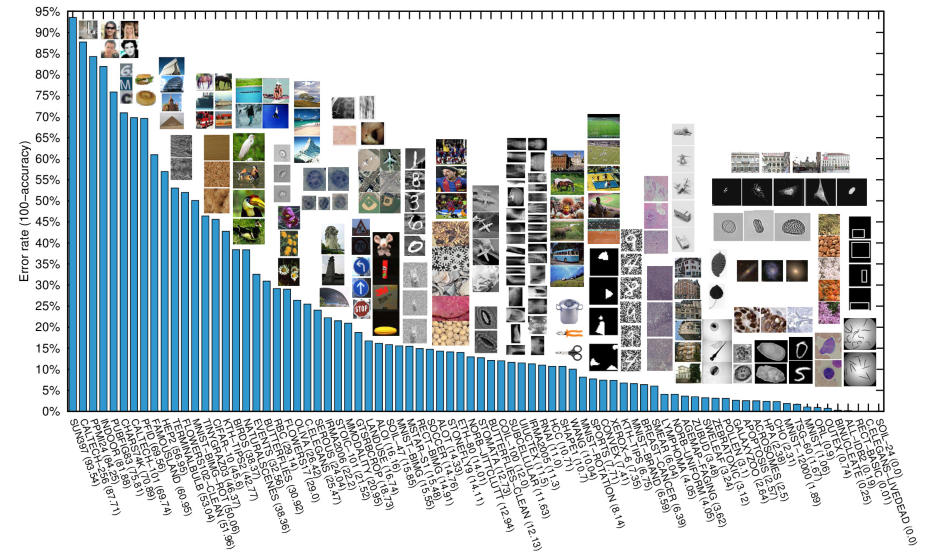
**4- نتایج**

**4-1- نتایج کلی**

در مورد کارایی کلی، ما برای 52 مجموعه داده (از بین 80 مجموعه)، نرخ تشخیص بیش از 80 درصد را ارائه دادیم و برای 30 مجموعه داده، میزان آن بیش از 90 درصد است (شکل 2 را ببینید). هرچند نتایج ما برای برخی مجموعه داده های دیگر یا مجموعه داده هایی که اخیرا منتشر شده اند، ضعیف تر هستند. زیرا تغییرات زیادی در آنها رخ داده است و به ویژه ما برای 13 مجموعه داده، به نرخ تشخیص کمتر از 50 درصد رسیدیم که اغلب آنها حاوی تصاویری از وب بودند و گروه های دانه درشت را نمایش می دهند (حس ذاتی یا گروه های اشیاء با بک گراند پیچیده و شدت نور قوی و تغییرات نور). در کل، میانگین بهترین خطا برای تمامی 80 مجموعه داده برابر 22.22 درصد است. جالب است بدانید برای زیرمجموعه ی 25 مجموعه داده ی تصاویر زیستی، میانگین بهترین نرخ خطا برابر 12.03 درصد است (در شکل (5-2) مروری بر این مجموعه داده صورت گرفته است). برای بررسی اطلاعات حیاتی در این تصاویر با نمایش بصری، مشابه زیرمجموعه های پیشین لازم است که ابتدا از یک الگوریتم دسته بندی با مقادیر پارامتر مشابه استفاده شود تا بهترین نرخ تشخیص برای این مجموعه داده حاصل شود (برای بررسی نتایج دقیق برای هر مجموعه داده به بخش تکمیلی مراجعه نمایید).

**4-1-1- مقایسه EL-DIC و ET-FL**

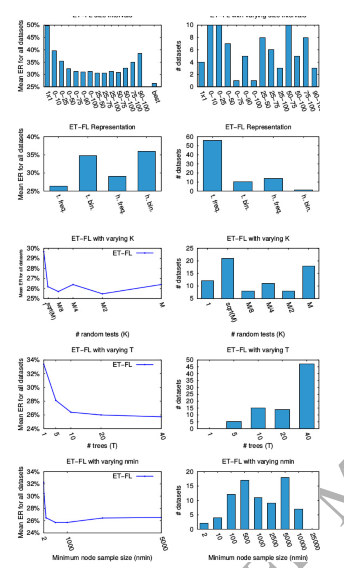
ET-DIC برای یک چهارم از مجموعه داده ها (شامل مجموعه داده های خاص شناسایی اشیاء در شرایط کنترلی) بهتر عمل می کند، و ET-FL برای مابقی مجموعه داده ها عملکرد بهتری دارد (60 مجموعه داده از بین 80 مجموعه). این نتایج نشان می دهد که برای اغلب مجموعه داده ها، ساخت یک نمایش عمومی از تصویر، براساس مقادیر گره های پایانی درخت، و سپس گروه بندی با استفاده از یک دسته بند خطی (ET-FL)، نتایج بهتری را در قیاس با دسته بندی مستقیم زیرپنجره ها (ET-DIC) ارائه می دهد. هرچند که این زیرپنجره ها با توجه به تصویر ورودی (در هنگامی که ET-DIC عملکرد بهتری دارد) می توانند پیشگویی قوی تری داشته باشند. با استفاده از ET-FL می توان تصاویر را با نمایش سطح بالاتری (در قیاس با پیکسل های سطری) ارائه داد. ویژگی های تصویر (با استفاده از الگوهای کوچک یا بزرگ) آموزش داده می شوند و هر برگ درخت، حاوی زیرپنجره هایی است که مجموعه ای از تست ها را بر روی مقادیر پیکسل ها (از کوچک به بزرگ) در زیرپنجره ها اجرا می کنند. مدل دسته بندی نهایی که این ویژگی های "پاسخ" را ترکیب می کند، قدرت تفکیک بالاتری نسبت به حالتی دارد که پیشگویی های جداگانه برای هر زیرپنجره با هم ترکیب می شوند (ET-DIC).



**شکل 2: خلاصه ای از بهترین نرخ خطای حاصل برای هر مجموعه داده، بدون بهینه سازی، که از جداول تکمیلی 2 تا 7 به دست آمده است. نمونه تصاویر تقریبا از 50 مجموعه داده به دست آمده اند و براساس کارایی تشخیص مجموعه داده، محل تقریبی آنها مشخص می شود. توجه شود که با بهینه سازی بیشتر، برخی نتایج (مثلا CIFAR-10 و GTSRB و IRMA-2005) بهبود قابل توجهی داشته اند. بخش (4-3) و جدول تکمیلی 8 را ببینید.**

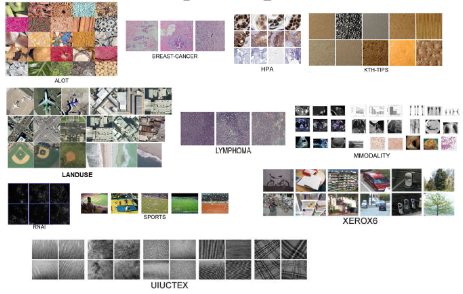
**4-2- بررسی تاثیر پارامتر**

تاثیر مقادیر تمامی پارامترها برای هر دو نسخه به صورت دقیق ارزیابی شده است. نتایج اصلی ما برای تاثیر پارامترها در بهترین روش (ET-FL) به صورت خلاصه در شکل 3 ارائه شده است و در ادامه بیان می شود (نتایج دقیق در بخش تکمیلی در جداول 2 تا 7 ارائه شده است).



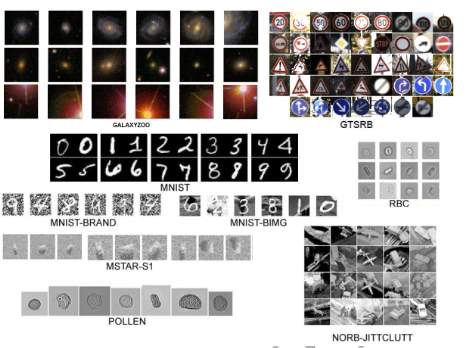
**شکل 3: میانگین نتایج برای تمامی 80 مجموعه داده برای نسخه ی ET-FL: ستون اول: میانگین نرخ خطا برای تمامی مجموعه داده ها متناسب با اندازه ی زیرپنجره ها (سطر اول)، نمایش تصویر (سطر دوم)، تعداد تست های تصادفی (سطر سوم)، تعداد درخت ها (سطر چهارم)، مینیمم اندازه ی نمونه های گره (سطر پنجم). ستون دوم: تعداد مجموعه داده ها به گونه ای که مقادیر پارامترها بتوانند بهترین نرخ خطا را داشته باشند. برای بررسی نتایج دقیق تر به جداول تکمیلی 2 تا 7 مراجعه کنید.**

با توجه به طرح استخراج تصادفی زیرپنجره ها، تاثیرگذارترین پارامتر، فاصله اندازه ی زیرپنجره ها است که سبب می شود این روش با انواع مختلف مسائل سازگار باشد. اندازه ی بهینه می تواند بسیار کوچک یا بسیار بزرگ باشد، و به اندازه ی تصویر بستگی دارد. مشاهده می شود که زیرپنجره های کوچک می توانند جزئیات دقیق را ثبت کنند و عموما برای تصاویری با قابلیت تکرار زیاد (مثلا تصویر بافت ها مثل بافت های هیستولوژیکی، بافت های سازنده ی بدن یا سنجش جمعیت سلول ها، شکل 4 را ببینید)، بهترین هستند. در حالی که زیرپنجره های بزرگتر، نتایج بهتری را برای مجموعه داده های شکل گرا ارائه می دهند (مثل سلول های قرمز خون، برگ ها، مشخصه های دستخط و علائم ترافیکی، شکل 5 را ببینید). برای این نوع دوم از مجموعه داده ها، استخراج زیرپنجره های بزرگ سبب تقویت مجموعه آموزشی (با مقیاس کوچک) و تبدیل نسخه ها می شود و در نتیجه مدل ها می توانند مستقیما الگوهای عمومی را ثبت کنند. در مورد تعداد زیرپنجره های استخراج شده، مشاهده شد که در مجموع 1 میلیون زیرپنجره ی آموزشی مورد استفاده قرار گرفته است. اما با نمونه برداری متراکم تر نیز می توان نتایج این مجموعه داده ها را بهبود بخشید.



**شکل 4: چندین مجموعه داده برای اینکه زیرپنجره های کوچکتر، کمترین نرخ خطا را داشته باشند.**

در نسخه ET-FL، با افزایش تعداد درخت ها تا 40 عدد (و بنابراین افزایش تعداد ویژگی ها برای دسته بند خطی نهایی)، پیشرفت هایی مشاهده می شود. هرچند که میزان این پیشرفت در قیاس با هنگامی که فقط از 10 درخت استفاده می شود، چندان زیاد نیست. درخت ها باید هرس شوند و در نتیجه مقدار باید تقریبا برابر یک هزارم تعداد کل زیرپنجره های مجموعه آموزشی باشد (تا ویژگی هایی ساخته شود که چندان خاص نیستند). این مساله فقط در برخی مسائل خاص صورت نمی گیرد که در آنها، شناسایی اشیاء در محیط کنترلی انجام می شود (و برای ویژگی های خاص بهتر عمل می کنند). مقدار عددی گره پایانی، نتایجی بهتر از کدگذاری دودویی یا سلسله مراتبی ارائه می دهد. به طور متوسط، مقدار پیش فرض پارامتر فیلترینگ، نتایج بهتری نسبت به ساخت ویژگی های غیرنظارتی ارائه می دهد. اما به نظر نمی رسد که افزایش این پارامتر به مقادیر بالاتر، اهمیتی داشته باشد. هرچند که این مورد برای مسائل مختلف (تصاویر نویزی و دارای شکل خاص) همچنان مفید است.



**شکل 5:** **چندین مجموعه داده برای اینکه زیرپنجره های بزرگتر، کمترین نرخ خطا را داشته باشند.**

**4-3- بهینه سازی بیشتر**

در عمل در صورتی که در این روش، نتایج رضایت بخشی برای یک مساله ی خاص حاصل نشود، می توان پارامترهای آن را بیشتر بهینه سازی نمود و نسخه های مختلف الگوریتم را پیاده سازی کرد تا نتایج بهتری به دست آید. هرچند که با این عمل، جامعیت راهکار به خطر می افتد، اما معتقدیم که بهینه سازی بهتر از طراحی کامل یک روش جدید و خاص است. هرچند که در ادامه ی مقاله نیاز است ارزیابی شود که آیا می توان این نسخه ها را تعمیم داد و به صورت موفقیت آمیز در مجموعه داده های زیادی مورد استفاده قرار داد یا خیر. در بخش تکمیلی، نتایج نویدبخشی از بهینه سازی های ساده برای 12 مجموعه داده ارائه شده است. این بهینه سازی ها شامل موارد زیر هستند: بسط محدوده پارامترها (افزایش تعداد درخت ها)، استفاده از داده های مصنوعی (تقویت داده ها با افزودن زیرپنجره های معکوس و چرخشی با زوایای مشخص)، نرمال سازی توصیفگرهای زیرپنجره های تصادفی (مثلا با کاهش میانگین و سپس تقسیم بر انحراف استاندارد، برای هر کانال زیرپنجره)، ارزیابی تست های گره های مختلف در Extra-tree (تست های گره که آستانه ی پیکسل آنها با 8 گره مجاور متفاوت است)، استفاده از فیلترها بر روی تصاویر اصلی (مثلا استفاده از فیلترهای خطی و عملیات جمع آوری فضایی)، اضافه کردن ویژگی های آماری به توصیفگرهای زیرپنجره ها (استفاده از ویژگی ها در مقاله (اورلو و همکاران در 2008)). این بهینه سازی ها و ارزیابی آنها در بخش تکمیلی مورد بحث قرار گرفته است.

**5- مقایسه با مدل های دیگر**

بدون ذخیره ی مرکزی این نتایج، نتایج نمونه های مختلف برای تمامی مجموعه داده های ارائه شده در این مقاله، از مقالات مختلف تصویرسازی کامپیوتری جمع آوری می شود تا نتایج ارائه شده کاملا به روز باشند. تا جایی که می دانیم، هیچ روش دیگری برای ارزیابی راهکار دسته بندی تصاویر، از این تعداد مجموعه داده استفاده نکرده است. بنابراین، این تنها راهکار عمومیِ ارائه شده است. مقایسه ی دقیق زیرمجموعه ای از مجموعه داده ها در بخش تکمیلی ارائه شده است.

ابتدا با استفاده از درخت های بی نهایت تصادفی، روش خود را برای مجموعه داده های مختلف با راهکارهای دیگر مقایسه می کنیم. ابتدا بر روی تعداد مجموعه داده های اندک با اندازه تصویر ثابت، روش خود را با کاربردهای مستقیم Extra-Tree مقایسه می کنیم. در این جا استخراج زیرپنجره انجام نمی شود و در نتیجه هر تصویر با یک بردار ورودی نمایش داده می شود که تمامی مقادیر پیکسل های آن را کدگذاری می کند. نتایج نشان می دهد هنگامی که در روش ما از استخراج زیرپنجره ها استفاده می شود، نتایج بهتری تولید می شود. به ویژه برای مجموعه داده هایی که زیرپنجره های کوچکی دارند، نتایج بهتر است. اما روش ما برای زیرپنجره های بزرگ نیز بهترین نتایج را ارائه می دهد. در قیاس با (ماری و همکاران، 2005) که از زیرپنجره هایی با اندازه نامحدود و ET-DIC برای برخی مجموعه ها استفاده می کند، مشاهده می کنیم که تنظیم پارامترها (همچون اندازه زیرپنجره ها، تعداد آنها، تعداد تست های تصادفی، و طرح دسته بندی) می تواند به نتایج بسیار دقیق تری منجر شود. در قیاس با (موسمن و همکاران در 2008) که از ET-FL با کدگذاری دودویی در گره های پایانی و تعداد ثابتی از ویژگی ها بر روی برخی کلاس های اشیاء استفاده می کند، مشاهده کردیم که کدگذاری عددی و تعداد مسائل وابسته به ویژگی های آموزشی، تاثیر قابل توجهی بر نتایج دارند.

دوم آنکه مشاهدات نشان می دهد روش پیشنهادی، اغلب بهتر از راهکارهای ارائه شده در مقالات قبلی برای مجموعه داده های مختلف عمل می کند. به ویژه این مورد برای روش های عمومی صحیح است، مثلا هنگام استفاده از دسته بندها (دسته بند نزدیکترین همسایه با فاصله اقلیدسی، رگرسیون منطقی یا SVM) در تصاویر نمونه برداری معکوس. همچنین گاهی برای مجموعه داده های جدید، بهتر از روش های خاص عمل می کند، مثلا برای ساخت مجموعه داده ی تشخیص، مجموعه داده ی گروه بندی ورزشی، عملیات تشخیص برگ، یک مجموعه داده در مورد زمین که از تصاویر از نمای بالا استفاده می کند، و چندین مجموعه داده ی تصاویر زیستی (بخش تکمیلی را ببینید). همچنین روش پیشنهادی ما برای مجموعه داده های مختلف، همتراز یا بهتر از روش هایی است که از ویژگی های خاص کاربردی استفاده می کنند و از اغلب روش های دیگر، بهتر عمل می کند، در حالی که برای هر مساله ای نمی تواند به کارایی روش های نوین دست یابد (مثلا برای سلول های فلوروسانس ایمنی).

نتایج ما برای مسائل دیگر (به ویژه مجموعه داده هایی که تصاویر آنها از وب گرفته شده است و حیوانات وحشی، چهره ی سلبریتی ها یا حس و فعالیت های طبیعی را نشان می دهد)، با استفاده از مقادیر پیکسل های سطری از تصاویر اصلی، چندان رضایت بخش نیست. در اغلب این مجموعه داده ها، روش ما بدون انجام بهینه سازی، بدترین نتایج را نسبت به GIST (اولویا و ترالب در 2001) دارد و همچنین ضعیف تر از روش های ارائه شده است، همچون روش هایی که از ترکیب تعداد زیادی توصیفگر تصویر ساخته شده اند (گلر و نوزین در 2009) یا معماری چندمرحله ای (عمیق) دارند که گام های مختلف نرمال سازی، فیلترینگ و جمع آوری فضایی را ترکیب می کنند (پینتو و همکاران در 2009 و سیرسن و همکاران در 2012 و کاتونی و تورالبا در 2009 و ژااو و همکاران در 2010). در مجموعه داده های تشخیص اشیاء در مقیاس وب، که بهینه سازی را با استفاده از فیلتر تصاویر انجام دادیم (بخش 4-3 را ببینید)، راهکار ما بهتر از GIST است و همچنین اندکی بهتر از روش های چندمرحله ای دیگر (همچون شبکه های عصبی کانوالوشنی) و ماشین های بولتزمن قطعه ای است. اما همچنان به طرز قابل توجهی، ضعیف تر از بهترین روش های ارائه شده برای این مجموعه داده است. به علاوه، مشاهده کردیم که برای مسائل دیگر (همچون تشخیص علائم ترافیک، و تصاویر مصنوعی گروه های اشیاء)، برای امکان رقابت با روش های متنوع چندمرحله ای، نیازی به فیلتر کردن تصویر نیست. این نتایج مختلف نشان می دهند که هرچند یادگیری عمیق، اغلب به صورت یک چهارچوب یکپارچه نمایش داده میشود، اما روش ها و معماری "یادگیری عمیق" در هنگام ارزیابی برای مجموعه داده های مختلف، کارایی تشخیص بسیار متفاوتی دارند. در بخش تکمیلی، مقایسه های بیشتری ارائه شده است، اما به نظر میرسد مقایسه با نسخه های مختلف یادگیری عمیق، فراتر از حوزه ی این مقاله باشد.

6- راهنماها

با وجود خوب بودن کارایی کلی الگوریتم پیشنهادی، معتقدیم که یک روش بسیار خوب برای دسته بندی تصاویر قفسه ای است. واضح است که روش پیشنهادی، بدون انجام تنظیمات زیاد، بهترین کارایی را برای حل هر مساله ای ندارد، و برای حل مسائل جدید به شاخص های مناسبی نیاز دارد.

با خلاصه سازی آنالیز گسترده ی ارائه شده در این مقاله، پیشنهاد می شود که در هنگام استفاده از این روش برای دسته بندی تصویر جدید، از رویه ی زیر استفاده شود:

* بدون هیچ دانش قبلی از مساله ی در دست، پیشنهاد می شود که از تنظیمات پیش فرض زیر برای پارامترهای این روش استفاده شود: 1 میلیون زیرپنجره ی آموزشی در فضای رنگی HSV با قطعات 16\*16 کدگذاری شوند و 1000 زیرپنجره برای هر تصویر تست در نظر گرفته شود و مد ET-FL با T=10 و k=28 و و کدکردن تعداد گره های پایانی. با در نظر گرفتن اندازه ی زیرپنجره ها، ابتدا تنظیمات زیر را انجام می دهیم: زیرپنجره های کوچک (از 0 تا 10 درصد)، زیرپنجره های متوسط (25 تا 50 درصد) و زیرپنجره های بزرگ (75 تا 100 درصد) و سپس بر نتایج اولیه، اندازه ها را اصلاح می کنیم. با استخراج این اندازه ها از شبیه ترین مجموعه داده در مخزن 80 مجموعه داده ی مورد استفاده در مقاله نیز می توان به نتایج بهتری دست یافت.
* در صورتی که نتایج حاصل از تنظیمات پیش فرض، رضایت بخش نباشند، پیشنهاد می شود که برخی پارامترها تنظیم شوند. همان طور که قبلا بحث شد، تعداد درخت ها و تعداد زیرپنجره ها فقط براساس منابع محاسباتی در دسترس مشخص می شود (هر چقدر این منابع بیشتر باشند، نتیجه بهتر است). برای غنی سازی مجموعه آموزشی، در صورتی که کلاس ها وابسته به جهت نباشند، پیشنهاد می شود که داده افزایی صورت گیرد (با استفاده از چرخش و عملیات معکوس). همان طور که در بخش 4 دیدیم، بعد از اندازه ی زیرپنجره که نقش مهمی دارد، مساله ی مهمتر، پارامترهای خاص هستند که عبارتند از پارامتر فیلتر k و تنظیمات اولیه و باید بر روی این پارامترها تمرکز شود. تنظیم و سوییچ به ET-DIC نیز می تواند در نهایت انجام شود، اما در آزمایشات ما، بهبود قابل توجهی مورد انتظار نیست.
* در نهایت در صورتی که نتایج همچنان به قدر کافی خوب نیستند، پیشنهاد می شود که از فیلترسازی تصاویر (با فیلترهای خطی و عملیات جمع آوری فضایی)، یا استخراج ویژگی های واضح جدید استفاده شود تا توصیفگر ویژگی زیرپنجره ها تقویت شود. این کار می تواند مبتنی بر استخراج ویژگی تصاویر عمومی یا مسائل خاص تر باشد، در صورتی که بتوان این ویژگی ها را براساس دانش قبلی استخراج کرد.

این که آیا این سه مرحله مورد نیاز هستند یا خیر، وابسته به نوع کاربرد است.

**7- نتیجه گیری**

در این مقاله، مساله ی عمومی "دسته بندی تصاویر" به شیوه ی نظارتی، بدون هیچ پیش فرضی از گروه های تصویر، مطرح شد. یک بررسی تجربی گسترده انجام شد تا با استخراج زیرپنجره های تصادفی، با کمک توصیفگر شدت پیکسل سطری، و در نهایت استفاده از درختان بی نهایت تصادفی برای دسته بندی مستقیم تصاویر یا آموزش ویژگی ها، کارایی کلی نسخه های روش های ساده و مرکب ارزیابی شود.

نتایج ارائه شده توسط روش پیشنهادی ما، ضعیف تر از نتایج نوین برای حل هر مساله ای هستند، اما در عوض، راهکار پیشنهادی به سادگی می تواند کارایی را ارزیابی کند و کارایی خوبی برای مجموعه تصاویر متنوع دارد، مثلا تصاویر دنیای واقعی که تغییرات زیادی را شامل می شوند. بنابراین پیشنهاد می شود که از روش پیشنهادی برای دسته بندی تصاویر جدید استفاده شود و راهنماهایی نیز برای آن ارائه شده است. ما قبلا با موفقیت از این راهنماها و نسخه های مختلف راهکار پیشنهادی برای کاربردهای تجربی طبی- زیستی استفاده کرده ایم که شامل مقالات (دلگا و همکاران در 2014 و جینری و همکاران در 2015) است.

در نهایت، در آینده ی نزدیک، یک پیاده سازی پایتون از الگوریتم خود تحت مجوز منبع باز ارائه می دهیم و از CYTOMINE برای توزیع آن استفاده می کنیم که یک برنامه ی قوی اینترنتی برای آنالیز مشترک تصاویر طبی- زیستی چند گیگا پیکسلی است.

1. Image classes [↑](#footnote-ref-1)
2. Kotseruba et al [↑](#footnote-ref-2)
3. Marée [↑](#footnote-ref-3)
4. Image class [↑](#footnote-ref-4)
5. supervised [↑](#footnote-ref-5)
6. appearance-based [↑](#footnote-ref-6)
7. subwindow [↑](#footnote-ref-7)
8. Moosmann [↑](#footnote-ref-8)
9. Dumont [↑](#footnote-ref-9)
10. Stern [↑](#footnote-ref-10)
11. framework [↑](#footnote-ref-11)
12. bioimaging datasets [↑](#footnote-ref-12)
13. codebook [↑](#footnote-ref-13)
14. Visual word [↑](#footnote-ref-14)