Archive of SID.ir

مدل چندوظیفه برای تشخیص برجستگی و لبه با استفاده از تابع هزینه ترکیبی

سجاد دهقان[۱]، محمدجواد فدائی‌اسلام[۲*]

اطلاعات مقاله

نوع مقاله: پژوهشی
دریافت مقاله: ۱۴۰۱/۰۲/۲۳
بازنگری مقاله: ۱۴۰۱/۰۴/۲۳
پذیرش مقاله: ۱۴۰۱/۰۵/۳۰

واژگان کلیدی:
تشخیص شئ برجسته،
تشخیص لبه،
تابع هزینه ترکیبی،
شبکه کاملاً کانولوشنی،
یادگیری عمیق،
پردازش تصویر.

چکیده

تشخیص شئ برجسته با هدف شناسایی و بخش‌بندی برجسته‌ترین و متمایزترین اشیاء یا نواحی در یک تصویر انجام می‌شود. شبکه‌های کاملاً کانولوشنی (FCN)، مزایای خود را در مسأله تشخیص شئ برجسته نشان داده‌اند، با این حال، بسیاری از کارهای قبلی بر دقت ناحیه برجسته تمرکز کرده‌اند اما به کیفیت مرز توجّهی ندارند. در این پژوهش، ما بر مکمل بودن بین اطلاعات لبه و اطلاعات شئ برجسته تمرکز می‌کنیم و یک ماژول تشخیص لبه را برای مدل‌سازی صریح اطلاعات لبه برای حفظ مرزهای شیء برجسته به شبکه پیشنهادی اضافه می‌کنیم. شبکه پیشنهادی ما سعی دارد این دو وظیفه مکمل را با کمک متقابل هم بهبود دهد. از طرف دیگر حضور اشیاء چند مقیاسی در مجموعه داده‌های تشخیص شئ برجسته نیاز به مدل‌سازی دقیق در سطح تابع هزینه برای مقابله با مشکل عدم تعادل بین پیش‌زمینه و پس‌زمینه در تصاویر دارد. از این رو، ما از تابع هزینه ترکیبی در مرحله آموزش استفاده می‌کنیم که به مقیاس اشیاء حساس نیست، و می‌تواند مسأله انسجام فضایی را بهتر مدیریت کند و به طور یکنواخت مناطق برجسته را بدون پارامترهای اضافی برجسته کند. مقایسه نتایج کمّی، کیفی به دست آمده توسط روش پیشنهادی با سایر روش‌های پیشرفته در شش مجموعه داده پرکابرد تشخیص برجستگی، نشان می‌دهد، روش پیشنهادی از عمل‌کرد خوبی برخوردار است و به سرعت می‌تواند مناطق برجسته را شناسایی کند. به طور خاص، روش ما بهترین عملکرد را در سه مجموعه‌داده آزمایشی پرکابرد از نظر معیارهای F-measure و MAE دریافت می‌کند که کارایی روش پیشنهادی را نشان می‌دهد.

۱-مقدمه

هدف تشخیص شئ برجسته[۲]، شناسایی متمایزترین اشیاء از نظر بصری در یک تصویر ورودی است و یک مرحله پیش-پردازش موثر در بسیاری از مسائل پردازش تصویر و بینایی رایانه‌ای به شمار می‌رود. تشخیص برجستگی، نیازمند درک کل تصویر و شناسایی دقیق جزئیات مناطق برجسته است. از این رو، تشخیص برجستگی یک مسأله تحقیقاتی چالش برانگیز است. ناحیه برجسته‌ای که توجّه انسان را به خود جلب می‌کند، اکثر بخش‌های اطلاعاتی یک صحنه را در خود دارد. در مقایسه با پردازش کل تصویر، تمرکز بر ناحیه برجسته می‌تواند هزینه محاسباتی را کاهش داده و برای بسیاری از مسائل بینایی رایانه‌ای عملکرد مطلوبی داشته باشد. تشخیص شئ برجسته با کمک روش‌های یادگیری عمیق، به سرعت در حال رشد است و در بسیاری از حوزه‌های بینایی رایانه‌ای، گرافیک، و رباتیک استفاده شده است[۱].

* پست الکترونیک نویسنده مسئول: fadaei@semnan.ac.ir
۱. فارغ‌التحصیل کارشناسی ارشد مهندسی کامپیوتر، دانشکده مهندسی برق و کامپیوتر، دانشگاه سمنان
۲. استادیار، دانشکده مهندسی برق و کامپیوتر، دانشگاه سمنان

[2] Salient object detection

DOI: 10.22075/jme.2022.27156.2272

Archive of SID.ir

Archive of SID.ir

مدل‌های سنّتی تشخیص شئ برجسته، متعلق به پیشرفت های مختلفی هستند که قبل از ظهور تکنیک‌های یادگیری عمیق در تشخیص برجستگی تصویر انجام شده است. در تمام مدل‌های سنّتی از روش‌های اکتشافی برای شناسایی اشیاء برجسته استفاده می‌شود. با وجود اینکه، ویژگی‌های استخراج شده به صورت دستی تشخیص درلحظه را امکان‌پذیر می‌کنند، اما دارای کمبودهای متعددی هستند که توانایی آن‌ها برای به دست آوردن اشیاء برجسته در سناریوهای چالش برانگیز را محدود می‌سازد. مدل‌های عمیق اولیه تشخیص برجستگی از پرسپترون‌های چند لایه (MLP) استفاده می‌کنند[۲]. در این مدل‌ها، تصویر ورودی معمولاً به نواحیِ کوچک تک‌مقیاس یا چند مقیاس تقسیم می‌شود. سپس، از شبکه عصبی کانولوشنی برای استخراج ویژگی‌های سطح بالا استفاده می‌شود. در نهایت، ویژگی‌های استخراج شده برای پیش بینیِ امتیاز برجستگی هر ناحیه کوچک، به یک طبقه‌بندی‌کننده MLP وارد می‌شوند. اگرچه روش‌های تشخیص برجستگی تصویر مبتنی بر MLP بسیار بهتر از همتایان غیرعمیق عمل می‌کنند، اما آن‌ها نمی‌توانند به طور کامل از اطلاعات فضایی ضروری استفاده کنند و بسیار زمان‌بر هستند، چون باید تمام زیرواحدهای بصری را یک به یک پردازش کنند. اخیراً، اغلب مدل‌های عمیق تشخیص برجستگی دارای چارچوب کدگذار-کدگشا مبتنی بر شبکهٔ کاملاً کانولوشنی (FCN) [۳] هستند که با استفاده از نقشه‌های برجستگی حاشیه‌نویسی شده پیکسلی، به شیوه ای سر به سر[۱] آموزش داده شده‌اند. به طور خاص، اکثر این نوع مدل‌ها تحت ساختاری شبیه به شبکهٔ FCN کار می‌کنند، که در ابتدا برای حل سایر وظایف یادگیری تصویر به تصویر از جمله: تقسیم‌بندی معنایی و تشخیص لبه پیشنهاد شده است. تکنیک‌های خاصی، مانند: اتصال کنارگذر[۲]، کانولوشن - آتروسی[۳]، شبکه کاملاً تحت نظارت با خروجی جانبی کلی نگر[۴] و ماژول نمونه‌برداری هرمی[۵] و غیره، را می‌توان به طور ضمنی یا صریح برای توسعه مدل های جدید شبکهٔ عصبی کانولوشنی عمیق برای تشخیص برجستگی شئ، معرفی-کرد[۴].

اگرچه در حال حاضر پیشرفت‌های زیادی حاصل شده است، اما هنوز باید به دو موضوع توجه شود؛ (۱) بسیاری از روش‌های تشخیص شئ برجسته مبتنی بر یادگیری عمیق موجود، از مرزهای شئ برجسته نامناسب رنج می‌برند. اخیراً، برای به دست آوردن جزئیات لبه مناسب، برخی از کارهای تشخیص شئ برجسته مانند مراجع [۸/۵] از یک روش دوطرفه یا بازگشتی برای اصلاح ویژگی‌های سطح بالا با اطلاعات محلی استفاده کردند، اما در این روش‌ها، مرزهای اشیاء برجسته هنوز به صراحت مدل نشده‌اند و مکمل بودن اطلاعات لبه و اطلاعات شئ برجسته مورد توجه قرار نگرفته است. همچنین، روش‌هایی نیز از تکنیک های پیش‌پردازش یا پس‌پردازش برای حفظ مرزهای شئ برجسته استفاده می‌کنند که مشکل اصلی این روش‌ها سرعت استنتاج پایین آن‌ها می‌باشد[۹]. یک نتیجه خوب تشخیص لبه، می‌تواند به کار تشخیص شئ برجسته در هر دو بخش تقسیم‌بندی و محلی‌سازی کمک کند. براساس این ایده، ما یک ماژول تشخیص لبه را برای مدل‌سازی و ترکیب اطلاعات تکمیلی لبه با اطلاعات شئ برجسته به شبکه پیشنهادی اضافه می‌کنیم. (۲) بیشتر روش‌های تشخیص شئ برجسته از تابع آنتروپی متقاطع[۶] به عنوان تابع هزینه در فرآیند آموزش استفاده می‌کنند. این تابع هزینه، رابطه بین پیکسلی را نادیده می‌گیرد در حالی که هزینه هر پیکسل را برای یک اندازه دسته ثابت جمع می‌کند[۱۰]. مدل‌هایی که با آنتروپی متقاطع آموزش داده می‌شوند، معمولاً نسبت به پیکسل‌های مرزی، قابلیت اطمینان کمتری دارند که این امر منجر به تار شدن مرزها می‌شود. همچنین، مشکل عدم توازن کلاسی[۷] ناشی از تغییر مقیاس، اثر تابع هزینه آنتروپی متقاطع را تضعیف می کند و منجر به ناسازگاری فضایی پیش‌بینی‌ها می‌شود. علاوه بر این، مسأله چند مقیاسی باعث ناهماهنگی بین مناطق پیش‌زمینه و پس‌زمینه در مجموعه داده‌ها می شود[۱۱]، به عبارتی، حضور آشکار اشیاء چند مقیاسی در مجموعه داده‌های تشخیص شئ برجسته نیاز به مدل‌سازی دقیق در سطح تابع هزینه برای مقابله با مشکل عدم تعادل بین پیش‌زمینه و پس‌زمینه در تصاویر دارد. برای پرداختن به این مشکل، ما از تابع هزینه ترکیبی استفاده می‌کنیم تا مدل خود را برای برجسته کردن یکنواخت کل منطقه

[1] End-to-End
[2] Skip connection
[3] Atrous convolution
[4] Holistically-nested side-way output
[5] Pyramid pooling module
[6] Cross Entropy (CE)
[7] Class imbalance

Archive of SID.ir

Archive of SID.ir

برجسته و مدیریت بهتر مشکل عدم تعادل پیکسلی بین مناطق پیش‌زمینه و پس‌زمینه ناشی از مقیاس‌های مختلف اشیاء، بدون هیچگونه پس‌پردازش یا پارامتر اضافی آموزش دهیم.

به طور خلاصه، ما یک شبکه چند وظیفه را برای مدل‌سازی و ترکیب اطلاعات تکمیلی لبه و اطلاعات شئ برجسته در یک شبکه واحد به روشی سر به سر پیشنهاد می‌کنیم. همچنین، تابع هزینه ترکیبی را بین پیش‌بینی حاصل و پاسخ صحیح مبنا، بهینه‌سازی می‌کنیم.

۲-پیشینه پژوهش

۲-۱-تشخیص شی برجسته

مدل‌های سنّتی تشخیص شئ برجسته، مبتنی بر ویژگی های سطح پایین و دانش قبلی اکتشافی هستند. مدل‌های سنّتی را می‌توان به دو دسته اصلی طبقه‌بندی کرد: روش-های پایین به بالا و روش‌های بالا به پایین. به طور خاص، روش‌های پایین به بالا با تمرکز بر ویژگی سطح پایین به عنوان عوامل بیرونی که مکانیسم توجه بصری را فعال می-کند، راهکاری برای تعریف برجستگی نشان می‌دهند و روش‌های بالا به پایین برجستگی را از دیدگاه درونی تعریف می‌کنند، به این معنی که توجه بصری با تجربه، حافظه و احساسات افراد، ارتباط نزدیکی دارد[۴]. این مدل‌های سنّتی عمدتاً به ویژگی‌های سطح پایین و شیوه‌های اکتشافی خاص تمرکز می‌کنند (برای مطالعه بیشتر رجوع کنید به[۱۲]). واضح است که ویژگی‌های سطح پایین و اکتشافی برای دستیابی به معانی سطح بالا کافی نیستند، بنابراین روش‌های مبتنی بر آن‌ها اغلب برای ورودی‌های بی‌اهمیت با شکست مواجه می‌شوند. بنابراین، لازم است که اطلاعات تصویر سطح بالا و بافت را برای پیش‌بینی برجستگی در نظر بگیریم.

روش‌های یادگیری عمیق تمام ابعاد بینایی کامپیوتر را تحت تاثیر قرار دادند[۱۵/۱۳]. همین‌طور روش‌های متعدد تشخیص شئ برجسته با روش عمیق ایجاد شد. روش‌های عمیق اولیّه، از پرسپترون‌های چند لایه[1] (MLP) برای پیش‌بینی امتیاز برجستگی هر پیکسل در یک تصویر استفاده می‌کردند. این روش‌ها از زیربخش‌های تصویر (به عنوان مثال؛ ابرپیکسل یا پچ‌ها و پیشنهادهای اشیاء عمومی[2] تصویر) به عنوان واحدهای پردازشی، استفاده می‌کنند[۲]. اگرچه مدل‌های مبتنی بر MLP از روش‌های تشخیص شئ برجسته سنّتی بهتر عمل می‌کنند، اما به دلیل استفاده از لایه‌های کاملاً متصل ثابت، از کارایی پایین و آسیب ساختارهای ویژگی ناشی از مسطح‌سازی[3] رنج می برند و قادر به گرفتن اطلاعات مکانی به طور مؤثر نیستند. روش‌های عمیق بعدی با استفاده از شبکه‌های کاملاً کانولوشنی[4] [۳]با این مسأله مقابله کردند و موفقیت خود را بر اساس تقسیم‌بندی معنایی نشان دادند. بسیاری از مدل‌های تشخیص شئ برجسته مبتنی بر یادگیری عمیق اخیر بر اساس شبکه‌های کاملاً کانولوشنی با استراتژی‌های مختلف برای ترکیب اطلاعات زمینه‌ای چند مقیاسی ایجاد شده‌اند. آن‌ها عمدتاً ماژول‌هایی را برای استخراج و تجمیع ویژگی‌های بافت از لایه‌های مختلف شبکه تعبیه کرده‌اند و به طور معمول، از خروجی‌های جانبی لایه‌های مختلف در کدگذار استفاده می‌کنند و این خروجی‌های جانبی را با لایه‌های موجود در کدگشا تجمیع می‌کنند. به طور کلی، روش‌های مبتنی بر شبکه‌های کاملاً کانولوشنی، در بین روش‌های یادگیری عمیق تشخیص شئ برجسته، غالب و برتر هستند.

۲-۲- مدل‌های چندوظیفه

در یادگیری ماشین، روش استاندارد یادگیری، انجام یک کار در یک زمان است. بیشتر روش‌های تشخیص برجستگی شئ عمیق به این حوزه از الگوی یادگیری تعلق دارند، یعنی برای آموزش، از نظارت یک حوزه‌ی دانش منفرد استفاده می‌کنند. با الهام از فرآیند یادگیری انسان که در آن دانش آموخته شده از وظایف مرتبط می‌تواند به یادگیری یک کار جدید کمک‌کند، یادگیری چند وظیفه[5] (MTL) قصد دارد تا عملکرد چندین وظیفه مرتبط را با یادگیری همزمان آن ها بهبود بخشد[۲]. وظایف یادگیری که به طور همزمان انجام می‌شوند، مرتبط با یکدیگر فرض می‌شوند. علاوه بر این، یک رابطه صریح بین وظایف ممکن است در معماری مدل تعریف و اجرا شود. همان‌طور که تشخیص شئ برجسته از دانش موجود در سایر وظایف استفاده می‌کند و بالعکس، توانایی تعمیم شبکه به سناریوهای دیده نشده بهتر می‌شود. علاوه بر این، یادگیری چند وظیفه‌ای برای

[1] Multi-layer perception
[2] Object proposals
[3] Flattening
[4] Fully convolutional networks
[5] Multi-task learning

Archive of SID.ir

Archive of SID.ir

رسیدگی به مشکل کمبود داده که در آن داده‌های مخصوص یک وظیفه برای آموزش یک مدل عمیق/ کم‌عمق بسیار محدود است، نیز مفید است[۱۰]. با بهره‌گیری از دانش اضافی در زمینه وظایف مرتبط، مدل‌ها می‌توانند قابلیت تعمیم بهتری پیدا کنند. در مرجع[۱۶]، از شمارش اشیا برجسته[1] به عنوان یک وظیفه کمکی استفاده کردند. زیرشبکه شمارش اشیا برجسته که از قبل آموزش دیده است، توسط یک لایه وزن تطبیقی به زیرشبکه تشخیص شئ برجسته متصل می‌شود. در مرجع[۱۷]، ویژگی‌های چند سطحی تشخیص شئ برجسته و تشخیص لبه به طور همزمان از طریق تعامل دو طرفه بین این دو وظیفه اصلاح می‌شوند. ژانگ و همکاران[۱۸] از عنوان تصویر[2] به عنوان یک وظیفه کمکی برای کدگذاری دانش معنایی اشیاء برجسته استفاده کردند. دو زیرشبکه تشکیل دهنده، یک شبکه پایه مشترک دارند و نقشه‌های حاصل حاوی نشانه‌های برجستگی مکمل هستند که برای برجستگی نهایی ترکیب می‌شوند. وانگ و همکاران [۱۹]، مکانیسم توجه سلسله مراتبی که براساس ویژگی‌های چند سطحی شبکه پایه عمل می‌کند را ایجاد کردند. این ویژگی‌های چند مقیاسی پیشرفته با اطلاعات لبه‌های منطبق بر مقیاس برای ترازکردن مرز شئ برجسته تلفیق می‌شوند. در مرجع[۲۰]، یادگیری مشترک تقسیم‌بندی معنایی با نظارت ضعیف و تشخیص شئ برجسته را انجام دادند، دو زیرشبکه معماری بر روی عمیق‌ترین ویژگی‌ها از یک شبکه پایه مشترک اجرا می‌شوند و در دو مرحله آموزش می‌بینند. ژائو و همکاران [۹]، ویژگی‌های لبه را با ویژگی‌های برجستگی چند لایه در یک ساختار دو مرحله‌ای ترکیب کردند. در مرجع [۲۱]، از نظارت‌های چندگانه برای آموزش چارچوب خود برای تشخیص شئ برجسته استفاده کردند. به طور خاص، تشخیص حد فاصل پیش‌زمینه[3] (FC)، تشخیص لبه (ED) و تشخیص شئ برجسته بانظارت استفاده می‌شود. در مرجع [۲۲]، مشکل عدم تعادل در توزیع پیکسل لبه را با تجزیه ماسک‌های برجسته برای نظارت‌های بهبود یافته حل کردند.

۳-۲-توابع هزینه

تابع هزینه آنتروپی متقاطع باینری، معیار هزینه‌ای است که به طور گسترده در مدل‌های تشخیص شئ برجسته عمیق به کار رفته است. توابع هزینه دیگر مانند؛ تابع هزینه اجتماع بر اشتراک[4]، تابع هزینه F-measure و تابع هزینه امتیاز dice برای مجموعه‌های آموزشی بایاس نیز پیشنهاد شده‌اند، اما آن‌ها به طور خاص برای گرفتن ساختارهای ظریف طراحی نشده‌اند[۲۳]. یک مشکل عدم تعادل مشابه، بین کلاس‌های مثبت/ منفی برای یال‌ها در مرجع [۲۴] با استفاده از یک مکانیزم وزن‌دهی در هر پیکسل مورد بررسی قرار گرفته است. تابع هزینه به کار رفته در مرجع [۲۵] نیز ترکیبی از تابع هزینه آنتروپی متقاطع وزنی با معیارهای ارزیابی برای رسیدگی به عدم تعادل پیش‌زمینه/پس‌زمینه است. توابع هزینه در مراجع [۲۶, ۲۷] از اصطلاحات هزینه اضافی مانند اجتماع بر اشتراک (IoU) برای بهبود مرزهای شئ برجسته استفاده کرده‌اند. اخیراً نیز در مرجع [۲۳]، از یک تابع هزینه ترکیبی استفاده کردند که این تابع، شبکه را راهنمایی می‌کند تا تبدیل بین تصویر ورودی و پاسخ صحیح مبنا را در یک سلسله مراتب سه سطحی شامل؛ سطح پیکسل، سطح پچ و سطح نقشه ویژگی با ادغام توابع هزینه آنتروپی متقاطع باینری (BCE)، تشابه ساختاری[5] (SSIM) و اجتماع بر اشتراک[6] (IoU) بیاموزد. این تلاش‌ها برای طراحی تابع هزینه توجیه می‌کنند که استفاده از یک تابع هزینه با هدف طراحی مدلی کلی و یا خاص، همراه با تحلیل گرادیان می‌تواند در بهبود عملکرد مدل تعیین‌کننده باشد.

۳-روش پیشنهادی

در کار قبلی خود [۲۸]، متفاوت از روش‌های عمیق ارائه شده برای مسأله تشخیص شئ برجسته، به بررسی چگونگی حل مسأله تشخیص شئ برجسته با گسترش نقش شبکه های پایه در معماری شبکه‌های کاملاً کانولوشنی عمیق پرداختیم. ما یک بلوک ساختاری[7] جدید ارائه شده [۲۹] را برای استخراج ویژگی‌های عمیق چند سطحی و چند مقیاسی قوی‌تر در بخش شبکه پایه[8] (کدگذار[9])، به معماری مرجع[۳۰] اضافه کردیم و توانستیم سطح عملکرد را بهبود بخشیم. شکل (۱)، معماری کلی روش پیشنهادی

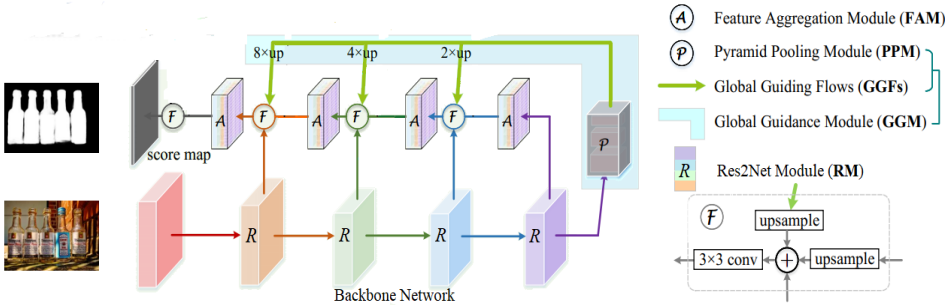
[1] Salient object subitization
[2] Image captioning
[3] Foreground contour detection
[4] Intersection over union (IoU) loss
[5] Structural Similarity
[6] Intersection over Union
[7] Building block
[8] Backbone network
[9] Encoder

Archive of SID.ir

Archive of SID.ir

قبلی ما را نشان می‌دهد که ساختار پایه آن براساس شبکه‌های هرمی ویژگی[1] (FPN) [۳۱] می‌باشد.

در این مقاله، برای بهبود بیشتر کیفیت نقشه‌های برجستگی تولید شده توسط رویکرد پیشنهادی قبلی، ما سعی می‌کنیم وظیفه تشخیص لبه را با وظیفه تشخیص شئ برجسته ترکیب کنیم. ما یک شبکه چند وظیفه را برای مدل‌سازی این ویژگی‌های مکمل درون شبکه پیشنهاد می‌کنیم که در آن مرزهای شئ برجسته و محلی‌سازی با کمک ویژگی‌های لبه بهبود می‌یابد. در شاخه تشخیص لبه، شبکه پیشنهادی از ویژگی‌های سطوح مختلف مسیر بالا به پایین برای انجام پیش‌بینی پیکسلی به شکل تصویر به تصویر استفاده می کند. در نتیجه، قادر به دستیابی به نمایش دقیق اشیاء یا بخش‌هایی از شئ در مقیاس‌های مختلف است. علاوه بر این، ما از تابع هزینه ترکیبی برای کاهش مشکل عدم تعادل پیش‌زمینه و پس‌زمینه در شبکه پیشنهادی خود استفاده می‌کنیم، که می‌تواند به برجسته کردن یکنواخت مناطق شی برجسته کمک کند. به طور خلاصه، ما یک شبکه کاملاً کانولوشنی را برای ترکیب موثر ویژگی‌های لبه و اشیاء برجسته پیشنهاد می‌کنیم. ساختار کلی شبکه پیشنهادی را در شکل (۲) مشاهده می‌کنید که در ادامه به توضیح آن خواهیم پرداخت.



شکل ۱: معماری کلی روش پایه

۳-۱- ماژول تشخیص لبه

تشخیص لبه، که با هدف استخراج لبه‌های برجسته بصری و مرزهای اشیاء از تصاویر طبیعی انجام می‌شود، معمولاً به عنوان یک تکنیک سطح پایین در نظر گرفته می‌شود و انواع وظایف سطح بالا تا حد زیادی از توسعه تشخیص لبه بهره برده‌اند. ما چهار بلوک باقیمانده با طراحی مشابه مرجع [۳۲]را در چهار سطح ویژگی در مسیر بالا به پایین شبکه پیشنهادی اضافه می‌کنیم، که برای تبدیل اطلاعات استفاده می‌شود. این بلوک‌های باقی‌مانده دارای شماره کانال {۱۲۸، ۲۵۶، ۲۵۶، ۵۱۲} در سطوح مختلف هستند. هر لایه بلوک باقیمانده توسط یک لایه کانولوشنی ۳×۳ با ۱۶ کانال برای فشرده‌سازی ویژگی به اضافه یک لایه کانولوشنی ۱×۱ با ۱ کانال برای پیش‌بینی لبه دنبال می‌شود. همچنین، ما این چهار لایه کانولوشنی ۳×۳ با ۱۶ کانال را به هم متصل می‌کنیم و آن‌ها را به سه لایه کانولوشنی ۳×۳ متوالی با ۱۲۸ کانال وارد می‌کنیم تا اطلاعات لبه گرفته شده را برای افزایش جزئیات به شاخه تشخیص اشیاء برجسته منتقل کنیم. در تشخیص لبه تک مقیاس، یک تصویر اصلی را به شاخه تشخیص لبه آموزش دیده خود وارد می‌کنیم و خروجی یک نقشه احتمال لبه است. به این صورت که پیش‌بینی‌های نقشه لبه را هم از لایه‌های خروجی جانبی و هم از لایه ترکیب وزنی مطابق رابطه (۱) به دست می‌آوریم.

$$(\acute{Y}_{fuse}, \acute{Y}^{1}_{side}, \dots, \acute{Y}^{M}_{side}) = CNN((W, w, h)^{*}) \qquad (۱)$$

در رابطه بالا CNN(.) نقشه‌های لبه تولید شده توسط شاخه تشخیص لبه را نشان می‌دهد. خروجی یکپارچه نهایی را می‌توان با تجمیع این نقشه‌های لبه تولید شده مطابق رابطه (۲) به دست آورد.

$$\acute{Y}_{edge\ branch} = Average(\acute{Y}_{fuse}, \acute{Y}^{(1)}_{side}, \dots, \acute{Y}^{(M)}_{side}) \qquad (۲)$$

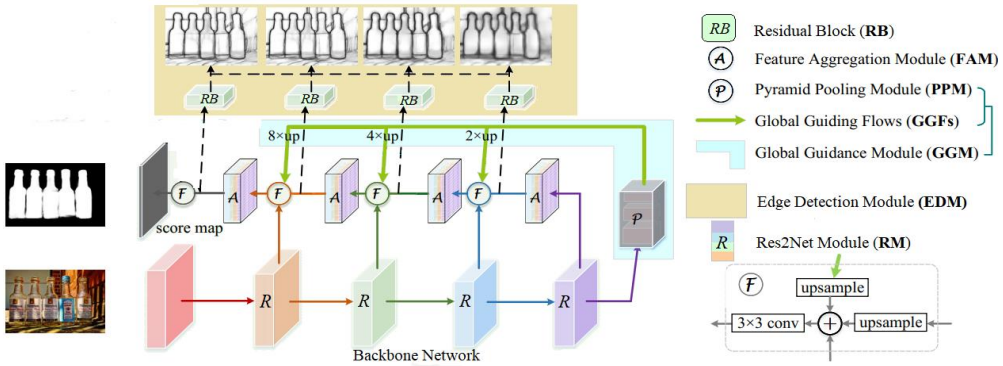
لایه‌های خروجی جانبی بعد از لایه‌های کانولوشن قرار داده می‌شوند. نظارت عمیق در هر لایه خروجی جانبی اعمال می‌شود و خروجی‌های جانبی را به سمت پیش‌بینی‌های لبه با ویژگی‌های مورد نظر ما هدایت می‌کند. خروجی‌های شاخه تشخیص لبه، چند مقیاسی و چند سطحی هستند.

[1] Feature pyramid networks

Archive of SID.ir

Archive of SID.ir

یک لایه ترکیب وزنی به شبکه اضافه می‌شود تا به طور خودکار نحوه ترکیب خروجی‌های چند مقیاسی را یاد بگیرد. برای بهبود بیشتر کیفیت لبه‌ها، از هرم‌های تصویری در هنگام فرآیند آزمایش استفاده می‌کنیم. به طور خاص، ما اندازه تصویر ورودی را برای ایجاد هرم تصاویر تغییر می‌دهیم و هر یک از این تصاویر به طور جداگانه به لبه‌یاب تک مقیاس ما وارد می‌شوند. ما در این پژوهش از چهار مقیاس ۰/۵، ۰/۱ ، ۱/۵ و ۲ استفاده می‌کنیم. سپس، تمام نقشه‌های احتمال لبه به دست آمده با استفاده از درون‌یابی دوخطی به اندازه تصویر اصلی تبدیل می‌شوند و برای به دست آوردن یک نقشه پیش‌بینی نهایی، میانگین‌گیری (میانگین ساده) می‌شوند. در آخر، نقشه‌ویژگی پیش بینی‌شده نهایی شاخه تشخیص لبه با نقشه ویژگی پیش بینی‌شده نهایی شاخه تشخیص شئ برجسته به هم پیوسته می‌شوند تا خروجی نهایی شبکه به دست آید. در فرآیند آموزش مشترک با تشخیص لبه، سعی می‌کنیم از ویژگی های لبه استخراج شده از شاخه تشخیص لبه برای کمک به اصلاح جزئیات نواحی لبه اشیاء برجسته استفاده کنیم.



شکل ۲: معماری شبکه پیشنهادی

۳-۲- تابع هزینه ترکیبی

در مسأله تشخیص شئ برجسته، تابع آنتروپی متقاطع باینری به طور گسترده مورد استفاده قرار می‌گیرد، اما این تابع، هزینه هر پیکسل را در کل دسته جمع می‌کند و روابط بین پیکسلی را در نظر نمی‌گیرد. در نتیجه، نمی‌تواند به صراحت مدل را وادار کند که ناحیه پیش‌زمینه را تا حد امکان بدون مشکل برجسته کند و به خوبی با مسأله عدم تعادل نمونه مقابله کند. برای این منظور، ما از تابع هزینه سازگاری افزایش یافته (CEL) [۱۱] در کنار تابع هزینه آنتروپی متقاطع باینری برای شاخه تشخیص شئ برجسته استفاده می‌کنیم. تابع هزینه آنتروپی متقاطع باینری طبق رابطه (۳) نوشته می‌شود که در آن log() یک عملیات عنصری است و $G \in \{0,1\}^{N\times H\times W\times 1}$ نشان‌دهنده پاسخ صحیح مرجع است P نشان دهنده مقدار پیش‌بینی شده است. تابع هزینه CEL طبق رابطه (۴) نوشته می‌شود[۱۱].

$$L_{BCEL} = \sum_{p\in P, g\in G} -[glogp + (1-g)\log(1-p)] \qquad (۳)$$

$$L_{CEL} = \frac{|FP+FN|}{|FP+2TP+FN|} = \frac{\sum(p-pg)+\sup(g-pg)}{\sum p\sum g} \qquad (۴)$$

که در آن TP، FP و FN به ترتیب نشان‌دهنده مثبت درست، مثبت کاذب و منفی کاذب هستند و | ۰ | ناحیه را محاسبه می‌کند. FP + FN نشان‌دهنده مجموعه تفاوت بین اجتماع و اشتراک ناحیه پیش‌زمینه پیش‌بینی شده و پاسخ صحیح مرجع است. در حالی که FP + 2TP + FN نشان‌دهنده مجموع این مجموعه اجتماع و اشتراک است. sup () کوچکترین کران بالای مجموعه است.

گرادیان L_{BCEL} تنها به پیش‌بینی موقعیت فردی متکی است. در حالی که گرادیان L_{CEL} به همه پیکسل‌ها؛ هم در پیش‌بینی شبکه و هم پاسخ صحیح مرجع، مرتبط است. بنابراین، تابع هزینه CEL برای اعمال محدودیت سراسری بر روی نتایج پیش‌بینی در نظر گرفته می‌شود، که می‌تواند انتشار گرادیان موثرتری ایجاد کند[۱۱]. در نهایت، تابع هزینه کل شاخه تشخیص شئ برجسته را می‌توان به صورت زیر نوشت.

$$L_{saliency} = L_{BCEL}(P,G) + L_{CEL}(P,G) \qquad (۵)$$

Archive of SID.ir

Archive of SID.ir

در شاخه تشخیص لبه برای سادگی، مجموعه تمام پارامترهای شاخه تشخیص لبه را به صورت W نشان می‌دهیم. فرض کنید در شبکه، M لایه خروجی جانبی داریم. هر لایه خروجی جانبی نیز با یک طبقه‌بندی کننده مرتبط است که در آن وزن‌های مربوطه با $W = (w^1, w^2, \dots, w^M)$ نشان داده می‌شوند. تابع هدف را طبق رابطه (۶) در نظر می‌گیریم.

$$L_{side}(W, w) = \sum_{m=1}^{M} \alpha_m l_{side}^{(m)}(W, w^{(m)}) \qquad (۶)$$

در رابطه بالا L_{side} تابع هزینه سطح تصویر برای خروجی‌های جانبی است.

در فرآیند آموزش، تابع هزینه روی تمام پیکسل‌های تصویر آموزشی $X = (x_j, j = 1, \dots, |X|)$ و پاسخ صحیح مبنا $Y = (y_j, j = 1, \dots, |X|), y_j \in \{0,1\}$ محاسبه می‌شود. برای یک تصویر طبیعی معمولی، توزیع پیکسل‌های لبه و غیر لبه به شدت بایاس است؛ تقریباً ۹۰ درصد از پاسخ صحیح مبنا غیر لبه است. ما از یک استراتژی ساده برای متعادل کردن هزینه بین کلاس‌های مثبت و منفی استفاده می‌کنیم. ما وزن متعادل کننده کلاسی β را بر مبنای مقدار هر پیکسل معرفی می‌کنیم. سپس، از این وزن متعادل کننده کلاسی، به عنوان یک روش ساده برای جبران عدم تعادل بین لبه و غیر لبه استفاده می‌کنیم. تابع هزینه متقاطع آنتروپی متعادل با کلاس طبق رابطه (۷) تعریف می‌شود[۲۴].

$$l_{slide}^{(m)}(W, w) = -\beta \sum_{j \in Y_+} logPr(y_i = 1|X; W, w^{(m)}) - (1-\beta) \sum_{j \in Y_-} logPr(y_i = 0|X; W, w^{(m)}) \qquad (۷)$$

در رابطه بالا $1 - \beta = |Y_+|/|Y| و \beta = |Y_-|/|Y|$ است، و Y_- و Y_+ به ترتیب مجموعه برچسب صحیح مبنای لبه و غیرلبه را نشان می‌دهد. مقدار $Pr(y_i = 1|X; W, w^{(m)}) = \sigma(a_j^{(m)}) \in [0,1]$ با استفاده از تابع سیگموئید $\sigma(.)$ روی مقدار فعال‌سازی در پیکسل j محاسبه می‌شود. در هر لایه خروجی جانبی، پیش‌بینی نقشه لبه $\acute{Y}_{side}^{(m)} = \sigma(\acute{A}_{slide}^{(m)})$ را به دست می آوریم، به طوریکه $\acute{A}_{slide}^{(m)} \equiv \{a_j^{(m)}, j = 1, \dots, |Y|\}$ فعال‌سازی خروجی جانبی لایه m است. برای استفاده مستقیم از پیش‌بینی‌های خروجی جانبی، یک لایه ترکیب وزن را به شبکه اضافه می‌کنیم که به طور همزمان وزن ترکیب را در طول فرآیند آموزش یاد می‌گیرد. تابع هزینه در لایه ترکیب طبق رابطه (۸) تعریف می‌شود.

$$L_{f_{use}}(W, w, h) = Dist(Y, \acute{Y}_{f_{use}}) \qquad (۸)$$

در رابطه بالا $\acute{Y}_{f_{use}} \equiv \sigma(\sum_{m=1}^{M} h_m)$ که در آن h وزن ترکیب است. $Dist(.,.)$ فاصله بین پیش‌بینی‌های ترکیب شده و پاسخ صحیح مبنا است که ما تابع هزینه آنتروپی متقاطع را به کار می‌بریم. با کنار هم قراردادن تمام موارد گفته شده، تابع هزینه رابطه (۹) را برای شاخه تشخیص لبه کمینه می‌کنیم[۲۴].

$$(W, w, h)^* = argmin(L_{side}(W, w), L_{f_{use}}(W, w, h)) \qquad (۹)$$

در نهایت، تابع هزینه کل روش پیشنهادی را می‌توان به صورت زیر نوشت.

$$L = L_{saliency} + L_{edge} \qquad (۱۰)$$

۳-۳- جزئیات پیاده سازی

چارچوب پیشنهادی براساس مخزن پایتورچ[1] پیاده‌سازی شده است. پارامترهای شبکه پایه مدل پیشنهادی (Res2Net-50)[۲۹]با مدل‌های متناظر از قبل آموزش دیده شده بر روی مجموعه‌داده ImageNet راه‌اندازی می شوند و بقیه پارامترها به طور تصادفی مقداردهی اولیه می شوند. برای اطمینان از همگرایی مدل، شبکه ما ۱۲ دوره[2] بر روی یک پردازنده گرافیکی NVIDIA GTX ۱۰۸۰ Ti آموزش دیده است. تمامی آزمایش‌ها با استفاده از بهینه ساز آدام[3] با کاهش وزن[4] $5e-4$ و نرخ یادگیری[5] اولیه $5e-5$که پس از ۸ دوره تقسیم بر ۱۰ می‌شود، صورت می‌گیرد. با پیروی از اکثر روش‌های پیشرفته موجود، ما از مجموعه‌داده DUTS[۳۳] به عنوان مجموعه‌داده آموزشی برای تشخیص شئ برجسته استفاده می‌کنیم. مجموعه-داده DUTS جدیدترین مجموعه‌داده در مقیاس بزرگ منتشر شده برای تشخیص برجستگی است که به ترتیب

[1] Pytorch repository
[2] Epoch
[3] Adam optimizer
[4] Weight decay
[5] Learning rate

Archive of SID.ir

Archive of SID.ir

شامل ۱۰۵۵۳ و ۵۰۱۹ تصویر در مجموعه‌های آموزشی و آزمایشی است. هر دو مجموعه آموزشی و آزمایشی شامل صحنه‌های پیچیده برای تشخیص شئ برجسته هستند. برای تشخیص لبه، ما با استفاده از مجموعه‌داده‌های تشخیص لبه BSDS500[۳۴] و PASCAL VOC Context[۳۵]، آموزش مشترک تشخیص شئ برجسته را با تشخیص لبه انجام می‌دهیم. در طول آموزش مدل، تصاویر از مجموعه‌داده تشخیص شئ برجسته و مجموعه-داده تشخیص لبه به طور متناوب وارد می‌شوند.

4- ارزیابی

در این بخش، ما شبکه پیشنهادی خود را با روش‌های پیشرفته قبلی؛ MSR [۳۶]، Amulet [۸]، NLDF [۲۶]، DSS [۳۷]، SRM [۳۸]، BMPM [۷]، PAGR [۳۹]، PICANet [۶]، DGRL [۴۰]، MLMS [۲۱]، CPD [۴۱]، ICNet [۴۲]، BASNet [۲۳]، AFNet [۲۷]، PoolNet [۳۰]، BANet [۴۳]، EGNet [۹]، R3NET [۴۴] و MINet [۱۱] مقایسه می‌کنیم. ما عمل‌کرد رویکرد خود را با استفاده از سه معیار پرکاربرد؛ منحنی‌های دقت و حساسیت (PR)، امتیاز بیشینه F-measure و میانگین قدر مطلق خطا (MAE)، با توجه به محبوبیت، مزایا و کامل بودن، بر روی شش مجموعه داده رایج SOD[۴۵]، DUT-OMRON[۴۶]، ECSSD[۴۷]، PASCAL-S[۴۸]، HKU-IS[۴۹]، DUTS-TE[۳۳] در مسأله تشخیص شئ برجسته ارزیابی می‌کنیم. شرح معیارها در [۲۸] آمده است. در جدول ۱، نتایج کمّی با توجه به معیار MAE و معیار بیشینه F-measure بر روی شش مجموعه داده پرکاربرد ارائه شده است (رنگ قرمز، آبی و سبز به ترتیب: رتبه اول، رتبه دوم و رتبه سوم از نظر عملکرد). در مجموعه داده‌های HKU–IS، SOD و DUTS–TE، مدل پیشنهادی ما بهترین عملکرد را با توجه به معیارهای MAE و بیشینه F-measure به دست آورد. در مجموعه داده DUT–OMRON با توجه به معیارهای MAE و بیشینه F-measure به ترتیب مراجع [۳۰] و [۲۸] بهترین عملکرد را دارند و مدل پیشنهادی ما، دومین بهترین عملکرد را دارد. در مجموعه داده PASCAL براساس هر دو معیار F-measure، مرجع [۱۱] بهترین عملکرد را دارد و مدل پیشنهادی با توجه به معیار MAE رتبه دوم را به خود اختصاص می‌دهد. در مجموعه داده ECSSD، مدل پیشنهادی ما با توجه به معیار بیشینه F-measure رتبه سوم از نظر عملکرد را دارد.

شکل (۳)، نتایج نمودارهای دقت-حساسیت (PR) را نشان می‌دهد. همانطور که مشاهده می‌شود، منحنی‌های PR روش پیشنهادی ما (رنگ آبی) در مقایسه با سایر روش‌های قبلی، برجسته می‌باشد. با نزدیک شدن امتیاز حساسیت به ۱، امتیاز دقت ما بالاتر از روش‌های دیگر است و نشان می دهد که نقاط مثبت کاذب در نقشه برجسته ما کم است. به خاطر اطلاعات لبه مکمل، نتایج به دست آمده حاکی از اطلاعات لبه تیز و مکان‌یابی دقیق است، که منجر به منحنی دقت-حساسیت بهتری می‌شود. شکل (۴) تصویر منحنی‌های معیار F-measure روش پیشنهادی و سایر روش‌های پیشرفته را نشان می‌دهد. به طور خلاصه، با توجه به شکل (۳) (منحنی‌های PR) و شکل (۴) (منحنی‌های F-measure)، همانطور که مشاهده می‌شود، خط آبی یکپارچه که نشان دهنده روش پیشنهادی است در اکثر آستانه‌ها به طور کلی بالاتر از روش‌های دیگر است و این پدیده نشان می‌دهد که روش ما از سایر رویکردها در مجموعه داده‌های چالش برانگیز عملکرد بهتری دارد.

برای توضیح بیشتر مزایای روش پیشنهادی، ما برخی از نتایج کیفی را در شکل (۵) نشان داده‌ایم. برای مقایسه منصفانه، نقشه‌های برجستگی روش‌های پیشرفته که به‌طور مستقیم توسط نویسندگان آن‌ها ارائه شده است را آورده‌ایم. سطر اول، عملکرد را بر روی تصاویر حاوی شئ شفاف نشان می‌دهد. سطر دوم، تصاویر با اجسام کوچک را نشان می دهد، برای اجسام کوچک در صحنه‌ها، مسائلی مانند تشخیص مناسب در سطوح عمیق شبکه و استراتژی‌های تجمع ویژگی‌ها به منظور جلوگیری از حواس‌پرتی در طول آموزش بسیار مهم هستند. سطر سوم، عملکرد را بر روی تصاویر با یک جسم بزرگ نشان می‌دهد، تشخیص یک شئ بزرگ در یک صحنه به اطلاعات زمینه‌ای چند مقیاسی اضافی نیاز دارد تا گستره کامل این اشیاء برجسته را پوشش دهد. یکی دیگر از موارد پیچیده، وجود چندین اشیا در یک صحنه است که در سطر چهارم ارائه شده است، وجود چندین شئ با معنایی مشابه یک موضوع چالش برانگیز است، زیرا ممکن است تعداد اشیا، شکل، اندازه، مکان و روشنایی آن‌ها در صحنه متفاوت باشد. سطر پنجم، عملکرد را بر روی تصاویر با یک صحنه پیچیده که دقیقاً حاوی یک

Archive of SID.ir

Archive of SID.ir

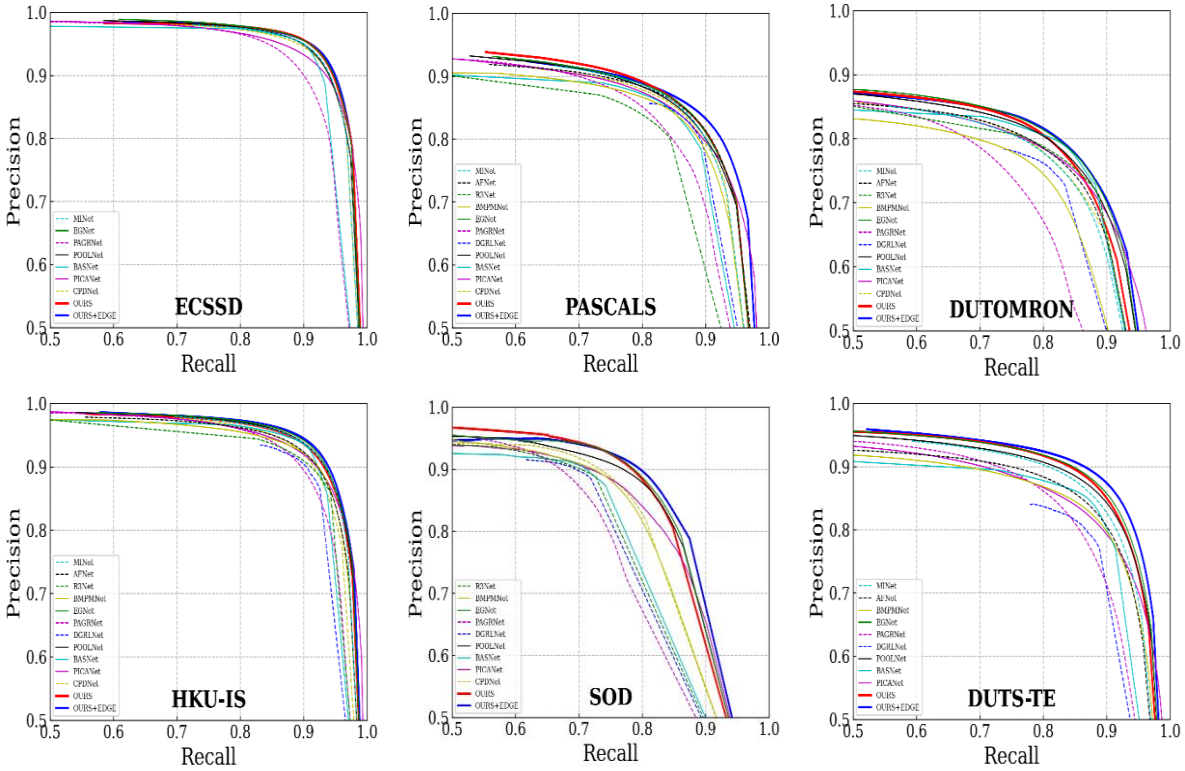
شئ برجسته در پس‌زمینه‌ای به هم ریخته است را نشان می‌دهد. سطر آخر، شامل تصویری با کنتراست پایین بین پیش‌زمینه و پس‌زمینه است. همانطور که از نتایج مشخص است، رویکرد پیشنهادی ما عملکرد بسیار خوبی را در مقایسه با سایر روش‌های پیشرفته از خود نشان داده است.

جدول ۱: نتایج کمّی تشخیص شئ برجسته براساس معیار MAE و معیار بیشینه F-measure

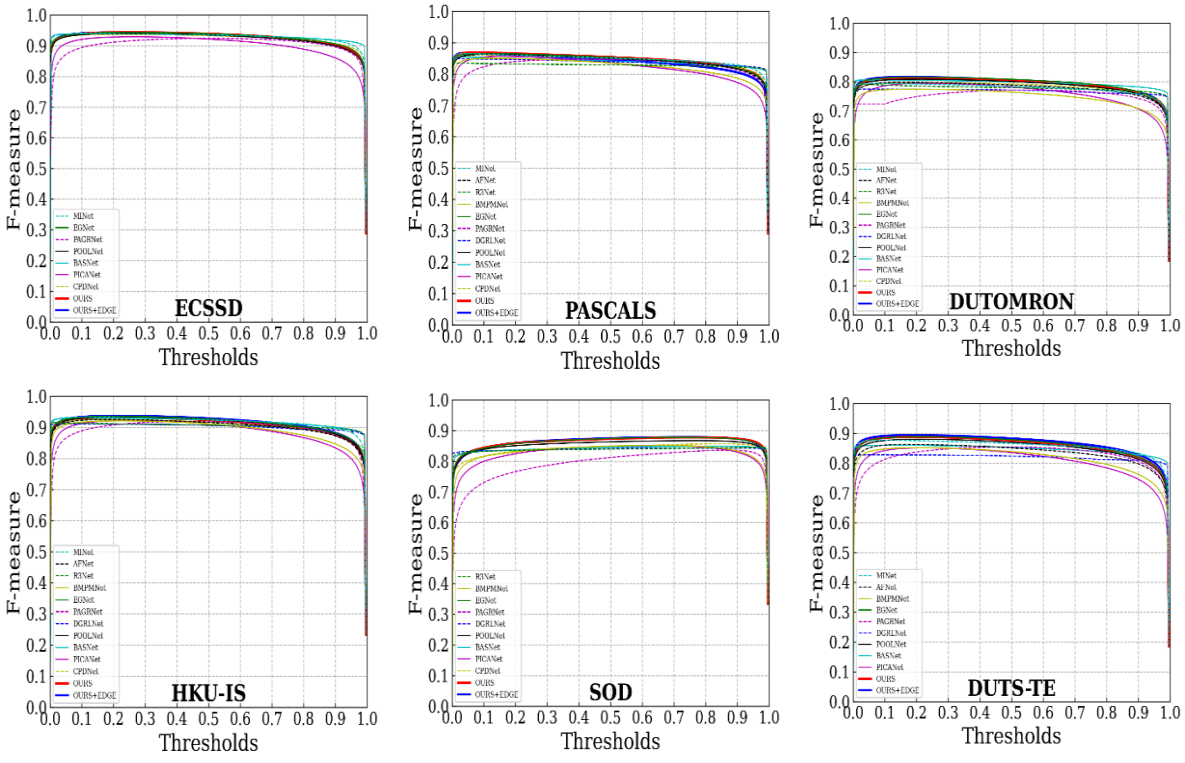
Models	Public	ECSSD		PASCAL		DUT-O		HKU-IS		SOD		DUTS – TE	
		MaxF	MAE	MaxF	MAE	MaxF	MAE	MaxF	MAE	MaxF	MAE	MaxF	MAE
VGG Backbone													
MSR[۳۶]	CVPR 2017	۰/۹۰۳	۰/۰۵۹	۰/۸۳۹	۰/۰۸۳	۰/۷۹۰	۰/۰۷۳	۰/۹۰۷	۰/۰۴۳	۰/۸۴۱	۰/۱۱۱	۰/۸۲۴	۰/۰۶۲
Amulet[۸]	ICCV 2017	۰/۹۱۱	۰/۰۶۲	۰/۸۲۶	۰/۰۹۲	۰/۷۳۷	۰/۰۸۳	۰/۰۸۳	۰/۰۵۲	۰/۷۹۹	۰/۱۴۶	۰/۷۷۳	۰/۰۷۵
NLDF[۲۶]	CVPR 2017	۰/۹۰۳	۰/۰۶۵	۰/۸۲۲	۰/۰۹۸	۰/۷۵۳	۰/۰۷۹	۰/۹۰۲	۰/۰۴۸	۰/۸۳۷	۰/۱۲۳	۰/۸۱۶	۰/۰۶۵
DSS[۳۷]	CVPR 2017	۰/۹۰۶	۰/۰۶۴	۰/۸۲۱	۰/۱۰۱	۰/۷۶۰	۰/۰۷۴	۰/۹۰۰	۰/۰۵۰	۰/۸۳۴	۰/۱۲۵	۰/۸۱۳	۰/۰۶۵
BMPM[۷]	CVPR 2018	۰/۹۲۸	۰/۰۴۴	۰/۸۶۲	۰/۰۷۴	--	--	۰/۹۲۰	۰/۰۳۸	۰/۸۵۱	۰/۱۰۶	۰/۸۵۰	۰/۰۴۹
PAGR[۳۹]	CVPR 2018	۰/۹۲۴	۰/۰۶۴	۰/۸۴۷	۰/۰۸۹	۰/۷۷۱	۰/۰۷۱	۰/۹۱۹	۰/۰۴۷	--	--	۰/۸۵۴	۰/۰۵۵
MLMS[۲۱]	CVPR 2019	۰/۹۲۸	۰/۰۴۵	۰/۸۶۴	۰/۰۷۵	۰/۷۷۴	۰/۰۶۴	۰/۹۲۰	۰/۰۳۹	۰/۸۱۱	۰/۱۰۶	۰/۸۵۲	۰/۰۴۹
CPD[۴۱]	CVPR 2019	۰/۹۳۶	۰/۰۴۰	۰/۸۷۳	۰/۰۷۴	۰/۷۹۴	۰/۰۵۷	۰/۹۲۴	۰/۰۳۳	۰/۸۵۶	۰/۱۰۹	۰/۸۶۴	۰/۰۴۳
MINet[۱۱]	CVPR 2020	۰/۹۴۳	۰/۰۳۶	۰/۸۸۲	۰/۰۶۵	۰/۷۹۴	۰/۰۵۷	۰/۹۳۲	۰/۰۳۰	--	--	۰/۸۸۷	۰/۰۳۹
ResNet Backbone													
SRM[۳۸]	ICCV 2017	۰/۹۱۶	۰/۰۵۶	۰/۸۳۸	۰/۰۸۴	۰/۷۶۹	۰/۰۶۹	۰/۷۶۹	۰/۰۴۶	۰/۸۴۰	۰/۱۲۶	۰/۸۲۶	۰/۰۵۸
PICA[۶]	CVPR 2018	۰/۹۳۵	۰/۰۴۶	۰/۸۵۷	۰/۰۶۷	۰/۸۰۳	۰۶۵/۰	۰/۰۶۵	۰/۰۴۳	۰/۸۵۶	۰/۱۰۴	۰/۸۶۰	۰/۰۵۰
DGRL[۴۰]	CVPR 2018	۰/۹۲۱	۰/۰۴۳	۰/۷۴۴	۰/۰۷۲	۰/۷۷۴	۰/۰۶۲	۰/۹۱۰	۰/۰۳۶	۰/۷۴۳	۰/۱۰۳	۰/۸۲۸	۰/۰۴۹
ICNet[۴۲]	CVPR 2019	۰/۹۳۸	۰/۰۴۱	۰/۸۶۶	۰/۰۷۱	۰/۸۱۳	۰/۰۶۱	۰/۹۲۵	۰/۰۳۷	--	--	۰/۸۵۵	۰/۰۴۸
BASNet[۲۳]	CVPR 2019	۰/۹۴۲	۰/۰۳۷	۰/۸۵۴	۰/۰۷۶	۰/۸۰۵	۰/۰۵۶	۰/۹۲۸	۰/۰۳۲	۰/۸۵۱	۰/۱۱۴	۰/۸۶۰	۰/۰۴۷
AFNet[۲۷]	CVPR 2019	۰/۹۳۵	۰/۰۴۲	۰/۸۷۱	۰/۰۷۱	۰/۷۹۷	۰/۰۵۷	۰/۹۲۵	۰/۰۳۶	--	--	۰/۸۶۳	۰/۰۴۶
PoolNet[۳۰]	CVPR 2019	۰/۹۴۰	۰/۰۴۲	۰/۸۶۳	۰/۰۷۵	۰/۸۳۰	۰/۰۵۵	۰/۹۳۴	۰/۰۳۲	۰/۸۶۷	۰/۱۰۰	۰/۸۸۶	۰/۰۴۰
BANet[۴۳]	ICCV 2019	۰/۹۵۴	۰/۰۳۵	۰/۸۷۹	۰/۰۷۰	۰/۸۰۳	۰/۰۵۹	۰/۹۳۰	۰/۰۳۲	--	--	۰/۸۷۲	۰/۰۴۰
EGNet[۹]	ICCV 2019	۰/۹۴۷	۰/۰۳۷	۰/۸۷۸	۰/۰۷۵	۰/۸۱۵	۰/۰۵۳	۰/۹۳۵	/۰۳۱۰	۰/۸۸۷	۰/۰۹۷	۰/۸۸۹	۰/۰۳۹
Res2Net Backbone													
مدل مرجع [۲۸]	--	۰/۹۴۳	۰/۰۳۸	۰/۸۶۹	۰/۰۷۰	۰/۸۱۲	۰/۰۵۱	۰/۹۳۱	۰/۰۳۱	۰/۸۷۷	۰/۱۰۰	۰/۸۸۷	۰/۰۳۶
روش پیشنهادی	--	۰/۹۴۴	۰/۰۳۸	۰/۸۶۹	۰/۰۶۷	۰/۸۱۵	۰/۰۵۳	۰/۹۳۶	۰/۰۲۹	۰/۸۷۸	۰/۰۹۷	۰/۹۱۱	۰/۰۳۱

Archive of SID.ir

Archive of SID.ir



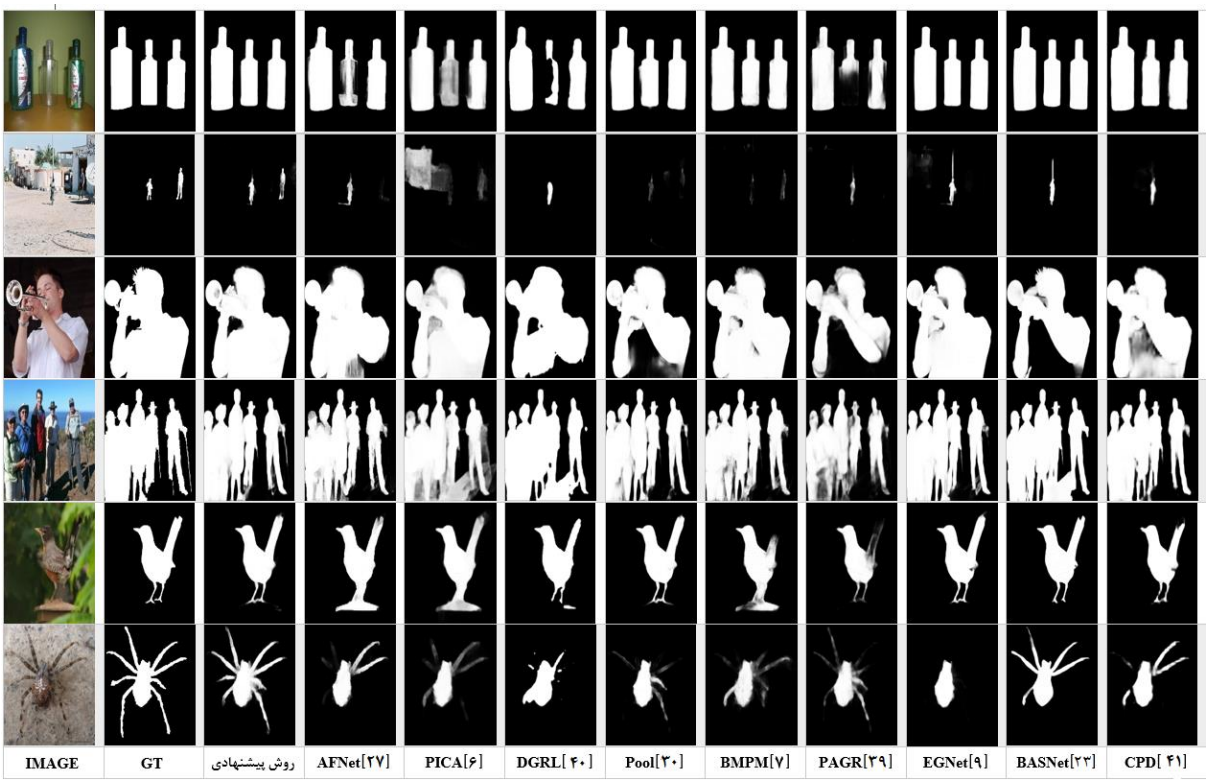
شکل ۳: تصویر منحنی‌های دقت-حساسیت (PR) روش پیشنهادی و سایر روش‌های پیشرفته



شکل ۴: تصویر منحنی های معیار F-measure روش پیشنهادی و سایر روش های پیشرفته

Archive of SID.ir

Archive of SID.ir



شکل ۵: مقایسه کیفی رویکرد پیشنهادی با سایر روش‌های پیشرفته

۵- نتیجه گیری

در این مقاله، روشی چند وظیفه برای مسأله تشخیص شئ برجسته ارائه شد. در شبکه پیشنهادی، از یک ماژول تشخیص لبه که بر اهمیت اطلاعات لبه برجسته تاکید دارد استفاده شد، زیرا اطلاعات مفیدی برای تقسیم‌بندی بهتر اشیاء برجسته و اصلاح مرزهای اشیاء ارائه می‌دهد. این ماژول تشخیص لبه، تخمین مرز برجسته دقیق را یاد می گیرد. بنابراین، تقسیم‌بندی شئ برجسته با حفظ لبه را به دقت مورد توجه قرار می‌دهد که در بهبود نتایج تاثیرگذار است. همچنین، ما از یک تابع هزینه ترکیبی استفاده کردیم تا مدل خود را برای برجسته کردن یکنواخت کل منطقه برجسته و مدیریت بهتر مشکل عدم تعادل پیکسلی بین مناطق پیش‌زمینه و پس‌زمینه ناشی از مقیاس‌های مختلف اشیاء، بدون هیچگونه پس‌پردازش یا پارامتر اضافی آموزش دهیم. نتایج آزمایش‌های انجام شده بر روی شش مجموعه داده پرکاربرد در مسأله تشخیص شئ برجسته، نشان می دهد که این روش از کارایی و دقت خوبی در مقایسه با سایر روش‌های پیشرفته موجود برخوردار است.

مراجع

[1] A. Borji et al., "Salient Object Detection: A Survey", Computational Visual Media, Vol. 5, 2019, pp. 117–150.

[2] W. Wang et al., "Salient Object Detection in the Deep Learning Era: An In-Depth Survey", arXiv: 1904.09146, 2019.

[3] J. Long, E. Shelhamer, and T. Darrell. “Fully Convolutional Networks for Semantic Segmentation”, IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431-3440.

[4] Y. Ji, H. Zhang, Z., Zhang and M. Liu, "CNN-Based Encoder-Decoder Networks for Salient Object Detection: A Comprehensive Review and Recent Advances, "Information Sciences", Vol. 546, 2021, pp. 835-857.

[5] N. Liu and J. Han. “DHSNet: Deep Hierarchical Saliency Network for Salient Object Detection”, IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 678-686.

[6] N. Liu, J. Han, and M.-H. Yang., “Picanet: Learning Pixel-wise Contextual Attention for Saliency Detection” , IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3089-3098.

Archive of SID.ir

Archive of SID.ir

[7] L. Zhang et al., “A Bi-directional Message Passing Model for Salient Object Detection”, IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1741-1750.

[8] P. Zhang et al., “Amulet: Aggregating Multi-level Convolutional Features for Salient Object Detection”, IEEE International Conference on Computer Vision, 2017, pp. 202-211.

[9] J.-X. Zhao et al., “EGNet: Edge Guidance Network for Salient Object Detection”, IEEE International Conference on Computer Vision, 2019, pp. 8779-8788.

[10] A.K. Gupta et al., "Salient Object Detection Techniques in Computer Vision—A Survey", Entropy, Vol. 22, NO. 10, 2020, pp. 1174.

[11] Y. Pang et al., "Multi-scale Interactive Network for Salient Object Detection", IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 9413-9422.

[12] A. Borji et al., "Salient Object Detection: A Benchmark", IEEE Transactions on Image Processing, Vol. 24, NO. 12, pp. 5706-5722.

[۱۳] محمود معلم، علی‌اکبر پویان، "کشف ناهنجاری با استفاده از کد کننده خودکار مبتنی بر بلوک‌های LSTM"، مدل سازی در مهندسی، دوره ۱۷، شماره ۵۶، بهار ۱۳۹۸، صفحه ۱۹۱-۲۱۱.

[۱۴] الهام پارسایی مهر، مهدی فرتاش و جواد اکبری ترکستانی، "بهبود استخراج ویژگی با استفاده از یک مدل یادگیری عمیق گروهی برای تشخیص موجودیت"، مدل سازی در مهندسی، دوره ۲۰، شماره ۶۹، تیر ۱۴۰۱، صفحه ۱۰۳-۱۱۲.

[۱۵] راضیه راستگو و کوروش کیانی، "شناسایی چهره با استفاده از تنظیم دقیق شبکه های کانولوشنی عمیق و رویکرد یادگیری انتقالی"، مدل‌سازی در مهندسی، دوره ۱۷، شماره ۵۸ ، پائیز ۱۳۹۸، صفحه ۱۰۳-۱۱۱.

[16] S. He et al., “Delving into Salient Object Subitizing and Detection”, IEEE International Conference on Computer Vision, 2017, pp. 1059-1067.

[17] Z. Wu, L. Su, and Q. Huang, “Stacked Cross Refinement Network for Edge-aware Salient Object Detection”, IEEE/CVF International Conference on Computer Vision, 2019, pp. 7264-7273.

[18] L. Zhang et al., “Capsal: Leveraging Captioning to Boost Semantics for Salient Object Detection”, IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 6024-6033.

[19] W. Wang et al., “Salient Object Detection with Pyramid Attention and Salient Edges”, IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 1448-1457.

[20] Y. Zeng et al., “Joint Learning of Saliency Detection and Weakly Supervised Semantic Segmentation”, IEEE/CVF International Conference on Computer Vision, 2019, pp. 7223-7233.

[21] R. Wu et al., “A Mutual Learning Method for Salient Object Detection with Intertwined Multi-superVision”, IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 8150-8159.

[22] J. Wei et al., “Label Decoupling Framework for Salient Object Detection”, IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 13025-13034.

[23] Qin, X., et al. “Basnet: Boundary-aware Salient Object Detection”, IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 7479-7489.

[24] S. Xie and Z. Tu. “Holistically-nested Edge Detection”, IEEE international Conference on Computer Vision, 2015, pp. 1395-1403.

[25] Wang, L., et al., "Salient Object Detection with Recurrent Fully Convolutional Networks", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 41, NO. 7, 2018, pp. 1734-1746.

[26] Luo, Z., et al. “Non-local Deep Features for Salient Object Detection”, IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 6609-6617.

[27] M. Feng, H. Lu, and E. Ding, “Attentive Feedback Network for Boundary-aware Salient Object Detection”, IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 1623-1632.

[۲۸] سجاد دهقان و محمدجواد فدائی اسلام، "بهبود تشخیص شئ برجسته با استفاده از ویژگی های چند مقیاسی در شبکه های عمیق"، چهارمین کنفرانس بین المللی محاسبات نرم، ۱۴۰۰.

Archive of SID.ir

Archive of SID.ir

[29] S. Gao et al., "Res2net: A New Multi-scale Backbone Architecture", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 43, NO. 2, 2019, pp. 652-662.

[30] J.-J. Liu et al., “A Simple Pooling-based Design for Real-time Salient Object Detection”, IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 3917-3926.

[31] T.-Y. Lin et al., “Feature Pyramid Networks for Object Detection”, IEEE Conference on Computer Vision and Pattern Recognition. 2017, pp. 2117-2125.

[32] K. He et al., “Deep Residual Learning for Image Recognition”, IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770-778.

[33] L. Wang et al., “Learning to Detect Salient Objects with Image-level SuperVision”, IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 136-145.

[34] P. Arbelaez et al., "Contour Detection and Hierarchical Image Segmentation", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, NO. 5, 2010, pp. 898-916.

[35] R. Mottaghi et al., “The Role of Context for Object Detection and Semantic Segmentation in the Wild”, IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 891-898.

[36] G. Li et al., “Instance-level Salient Object Segmentation”, IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2386-2395.

[37] Q. Hou et al., “Deeply Supervised Salient Object Detection with Short Connections”, IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3210-3212.

[38] T. Wang et al., “A Stagewise Refinement Model for Detecting Salient Objects in Images”, IEEE International Conference on Computer Vision, 2017, pp. 4019-4028.

[39] X. Zhan et al., “Progressive Attention Guided Recurrent Network for Salient Object Detection”, IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 714-722.

[40] T. Wang et al., “Detect Globally, Refine Locally: A Novel Approach to Saliency Detection”, IEEE Conference on Computer Vision and Pattern Recognition. 2018, pp. 3127-3135.

[41] Z. Wu, L. Su, and Q. Huang, “Cascaded Partial Decoder for Fast and Accurate Salient Object Detection”, IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3907-3916.

[42] W. Wang et al., “An Iterative and Cooperative Top-down and Bottom-up Inference Network for Salient Object Detection”, IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 5968-5977.

[43] J. Su et al., “Selectivity or Invariance: Boundary-aware Salient Object Detection”, IEEE International Conference on Computer Vision, 2019, pp. 3799-3808.

[44] Z. Deng et al., “R3net: Recurrent Residual Refinement Network for Saliency Detection”, International Joint Conference on Artificial Intelligence, 2018, pp. 684-690.

[45] V. Movahedi and J.H. Elder, “Design and Perceptual Validation of Performance Measures for Salient Object Segmentation”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition-workshops, 2010, pp. 49-56.

[46] C. Yang et al., “Saliency Detection via Graph-based Manifold Ranking”, IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3166-3173.

[47] Q. Yan et al., “Hierarchical Saliency Detection”, IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 1155-1162.

[48] Y. Li et al., “The Secrets of Salient Object Segmentation”, IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 280-287.

[49] G. Li and Y. Yu. “Visual Saliency Based on Multiscale Deep Features”, IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 5455-5463.

Archive of SID.ir