

Saliency Detection in Eye Gaze Prediction by Using Deep Learning

Reza Nasiripour¹, Hassan Farsi² and Sajad Mohamadzadeh^{3*}

1- Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran.

2- Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran.

3*- Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran.

¹reza.nasiripour@birjand.ac.ir, ²hfarsi@birjand.ac.ir, and ^{3*}s.mohamadzadeh@birjand.ac.ir

Corresponding author address: Sajad Mohamadzadeh, Department of Electrical and Computer Engineering, University of Birjand, Birjand, Iran. Postal Code: 9717434765.

Abstract- Salient object detection attracted the attention of researchers in various fields, so that it is used in many applications of the visual machine, such as object detection and tracking. Most of the work in this area is based on bottom-top saliency models and use low-level features to extract the final saliency map that these works do not have a significant accuracy in salient object detection. On the other hand, top-down visual models are used for the specific applications. In this paper, a different method has been proposed to extract the salient object map that uses low-up and top-down attributes for extracting the salient and it is based on the learning process. The simultaneous selection of these features makes the proposed algorithm for various applications and increases the accuracy of the salient object detection. The learning process is performed by using the and Convolutional Neural Network (CNN) structure. After the decomposition of image to its super pixels, different features of image are extracted. Then, the extracted features are normalized to have zero mean and unit variance, and CNN is used to train the features. The accuracy of the proposed method is improved by using of the 8 salient and CNN. The performance of the proposed method has been compared to twenty method by applying four new databases including MSRA-100, ECSSD, MSRA-10K and Paskal-S. The proposed method provides better results compared to the other methods in respect to prediction of salient object.

Keywords- Salient, low-up and top-down feature, super pixel, CNN

شناسایی شی برجسته در پیش بینی تمرکز نگاه با استفاده از یادگیری عمیق

رضا نصیری پور^۱، حسن فرسی^۲، سجاد محمدزاده^{۳*}

۱- دانشکده مهندسی برق و کامپیوتر، دانشگاه بیرجند، بیرجند، ایران.

۲- دانشکده مهندسی برق و کامپیوتر، دانشگاه بیرجند، بیرجند، ایران.

۳- دانشکده مهندسی برق و کامپیوتر، دانشگاه بیرجند، بیرجند، ایران.

¹reza.nasiripour@birjand.ac.ir, ²hfarsi@birjand.ac.ir, and ³*s.mohamadzadeh@birjand.ac.ir

* نشانی نویسنده مسئول: سجاد محمدزاده، خراسان جنوبی، بیرجند، دانشگاه بیرجند، دانشکده مهندسی برق و کامپیوتر، کد پستی: ۹۷۱۷۴۳۴۷۶۵

چکیده- شناسایی شی برجسته، توجه محققین در حوزه‌های مختلف را به خود جلب کرده‌است بگونه‌ای که در بسیاری از کاربردهای ماشین بینایی از جمله شناسایی و رهگیری شیء بکار گرفته می‌شود. بیشتر کارهای انجام شده در این حوزه مبتنی بر مدل‌های بصری پائین به بالا می‌باشند و از ویژگی‌های سطح پائین برای استخراج تصویر برجستگی نهایی استفاده می‌کنند که از دقت قابل توجهی در تشخیص شیء چشمگیر برخوردار نمی‌باشند. از طرفی مدل‌های بصری بالا به پائین برای کاربردهای خاصی مورد استفاده قرار می‌گیرند. در این مقاله روش متفاوتی برای استخراج تصویر برجستگی شیء ارائه شده است که از ویژگی‌های سطح پائین به بالا و بالا به پائین برای استخراج تصویر برجستگی استفاده می‌کند و مبتنی بر فرآیند یادگیری می‌باشد. انتخاب همزمان این ویژگی‌ها باعث بهبود الگوریتم پیشنهادی در کاربردهای مختلف می‌شود و باعث افزایش دقت شناسایی شیء چشمگیر می‌شود. فرآیند یادگیری با استفاده از شبکه عصبی کانولوشن انجام می‌گیرد. پس از این که تصویر به سوپرپیکسل‌های آن تجزیه شد، ویژگی‌های متفاوتی از آن استخراج می‌گردد. سپس ویژگی‌های استخراج شده به میانگین صفر و واریانس واحد نرمالیزه شده و از الگوریتم شبکه عصبی کانولوشن به منظور آموزش دادن ویژگی‌ها استفاده می‌شود. دقت روش پیشنهادی با استفاده از ۸ تصویر برجستگی و یادگیری عمیق بهبود داده شده است. روش پیشنهادی با بیش از ۲۰ روش دیگر و در ۴ پایگاه داده شناخته شده ECSSD, MSRA-1000, MSRA-10k و PASCAL-S مورد ارزیابی و مقایسه قرار گرفته شده است. نتایج حاکی از کارایی روش پیشنهادی نسبت به سایر روش‌ها در زمینه پیش‌بینی شیء چشمگیر در تعیین تمرکز نگاه می‌باشد.

واژه‌های کلیدی: تصویر برجستگی، ویژگی‌های سطح پائین به بالا و بالا به پائین، سوپرپیکسل، CNN.

۱- مقدمه

یکی از مهم‌ترین چالش‌ها در حوزه پردازش ویدئو، تصویر و بینایی ماشین می‌باشد که از جمله کاربردهای گوناگون آن می‌توان به مواردی مانند فشرده‌سازی تصویر و ویدئو، جداکردن خودکار نواحی از تصویر^۱، طراحی و سیستم‌های مبتنی بر تعامل کاربر با کامپیوتر اشاره کرد [۱].

پیش‌بینی توجه بصری را می‌توان به دو دسته کلی برجستگی پائین به بالا^۲ و برجستگی بالا به پائین^۴ دسته‌بندی کرد [۲]. در روش‌های مبتنی بر توجه پائین به بالا، ساختارهای سیستم بینایی انسان که از تحرک بالایی برخوردار می‌باشند، مورد توجه قرار

بررسی ناحیه‌ای از تصویر که توسط مغز انسان انتخاب می‌شود به سیستم بینایی انسان تبدیل می‌شود و به این فرآیند توجه بصری^۱ گفته می‌شود. ساختار سیستم بینایی انسان به گونه‌ای می‌باشد که نواحی خاصی از تصویر، توجه بیشتری نسبت به نواحی دیگر تصویر از خود نشان می‌دهد. توجه بصری فرآیندی است که سیستم بیولوژیکی را قادر می‌سازد که نواحی از صحنه را انتخاب کند که نگاه انسان به آن خیره شده است. پیش‌بینی توجه بصری

داده‌اند. سپس سطح هر هرم از نصف کردن ابعاد تصویر در سطح قبلی بدست آورده می‌شود و در انتها برای هر سطح، سه تصویر ویژگی شدت روشنایی، رنگ و جهت استخراج می‌شود.

در روش‌های مبتنی بر طیف فرکانسی^۶، طیف گابور یا طیف موجک یا طیف فوریه از تصویر محاسبه می‌شود. روش ارائه شده در [۵]، مبتنی بر طیف فرکانسی می‌باشد که از تبدیل موجک برای استخراج تصویر برجستگی استفاده کرده است. نویسندگان مقاله [۸] پیشنهاد استفاده از دو طیف فوریه و گابور را داده‌اند. براساس معیار وزن‌دهی، تصویر برجستگی برای هر طیف فرکانسی محاسبه و با ترکیب این دو طیف، تصویر برجستگی نهایی محاسبه می‌شود. در روش‌هایی که از قانون توان استفاده می‌کنند، تعداد مختلفی الگو برای تصویر تعریف می‌شود. در تضاد با دسته قبلی، این روش‌ها بر روی توان موجود میان پیکسل‌ها تمرکز دارند. این الگوها را می‌توان برای توان موجود در میان پیکسل‌ها و یا برای الگوریتم بیشترین شباهت تعریف کرد [۹]. عیب اصلی مدل ارائه شده در [۹]، استفاده از پارامترهای گوناگون به منظور بهینه‌سازی می‌باشد که باعث افزایش زمان و پیچیدگی محاسباتی می‌گردد.

مکانیزم کار در روش‌های مبتنی بر کنتراست، محاسبه میانگین شدت روشنایی تمامی پیکسل‌های تصویر و کنتراست هر پیکسل می‌باشد. در روشی که در مرجع [۱۰] پیشنهاد شده است، پس از اینکه تبدیل فضای رنگ از RGB به LAB انجام گرفت، برای هر کانال LAB کنتراست محاسبه می‌شود و از میانگین آن کم می‌شود. برای هر کانال رنگ، تصویر برجستگی محاسبه می‌شود. تصویر برجستگی نهایی ترکیب این سه کانال می‌باشد.

روش‌هایی که مبتنی بر آنتروپی می‌باشند، از الگوهای تعریف شده‌ای استفاده می‌شود تا مقادیر آنتروپی را بهینه‌سازی کنند. مبنای این روش‌ها استفاده از کنتراست محلی و بکارگیری ICA^۷ در تولید تصویر برجستگی شیء می‌باشد. مدلی که در مرجع [۱۱] پیشنهاد شده است از دو مکانیزم مختلف استفاده می‌کند. در ابتدا، با استفاده از مدل پائین به بالا، تصویر ویژگی‌های رنگ و DOG چهار مقیاسی حاصل از شدت روشنایی استخراج می‌شود. در فاز دوم، از ICA برای بدست آوردن تصویر برجستگی استفاده شده است. به عبارت دیگر، ۱۲ تصویر از DOG و ۳۶۲ ویژگی ICA در وجود آمدن تصویر برجستگی نهایی شرکت می‌کنند.

در روش‌های مبتنی بر مرکز-اطراف، تصویر برجستگی هر شیء براساس ویژگی‌های پنجره‌ای می‌باشد که در اطراف هر شیء در نظر گرفته می‌شود. به عبارت دیگر، کادری پنجره‌ای شکل در اطراف شیء به عنوان مرکز و کادری بزرگتر از مرکز به عنوان اطراف ترسیم می‌شوند و از ویژگی‌های مرکز-اطراف برای بدست آوردن تصویر برجستگی استفاده می‌شود. مرجع [۱۲] برای

می‌گیرد درحالی‌که در روش‌های مبتنی بر توجه بالا به پائین، حافظه و جستجو مبتنی بر حافظه مورد توجه قرار گرفته است. به عبارت دیگر روش‌های توجه بالا به پائین یک روش هدف‌گرا می‌باشد که نیاز به اطلاعات اولیه‌ای از تصویر از جمله شناسایی منظره و پس‌زمینه آن را دارد. یکی از اصلی‌ترین معایب رویکرد توجه بالا به پائین، زمان محاسباتی بالای آن در استخراج تصویر برجستگی است.

در مدل‌های پائین به بالا از ویژگی‌های سطح پائین تصویر برای استخراج تصویر برجستگی استفاده می‌شود [۳،۴]. این ویژگی‌ها عبارتند از: رنگ، بافت، سطوح روشنایی و غیره. Itti و همکارانش در مرجع [۴] روشی پیشنهاد کرده‌اند که مبتنی بر مکانیزم پائین به بالا می‌باشد. نویسندگان این مقاله با بهره‌گیری از ساختار سلسله مراتبی، از سه تصویر ویژگی رنگ، شدت و جهت برای استخراج تصویر نهایی برجستگی استفاده کرده‌اند. مدلی که در مرجع [۵] و به منظور تشکیل تصویر برجستگی نهایی پیشنهاد شده است، مبتنی بر ضرائب تبدیل موجک بالاگذر است. در مرحله اول تصویر ورودی از فضای رنگ RGB به فضای رنگ LAB انتقال داده می‌شود. سپس از تبدیل موجک برای استخراج کردن نویز موجود استفاده شده است. مرحله بعدی را اعمال IWT^۸ برای استخراج کردن دو تصویر برجستگی دربرمی‌گیرد. در مرحله آخر و به منظور تشکیل تصویر برجستگی نهایی، دو تصویر برجستگی با یکدیگر ترکیب می‌شوند.

مدل‌های بالا به پائین براساس مکانیزمی عمل می‌کنند که از قبل برای آن تعریف شده است [۶]. به عبارت دیگر، این مدل‌ها برای کاربردهایی از پیش تعیین شده، مورد استفاده قرار می‌گیرند. Sang و همکارانش در مرجع [۶] برای هر تصویر یازده تصویر ویژگی تعریف و استخراج کرده‌اند. این یازده تصویر ویژگی عبارتند از: سه ویژگی رنگ، دو ویژگی شدت روشنایی، چهار ویژگی جهت و دو ویژگی بافت. تصویر برجستگی نهایی با ترکیب یازده تصویر ویژگی را بوجود می‌آورند.

بطور کلی می‌توان روش‌های استخراج تصویر برجستگی شیء را به هفت گروه تقسیم کرد [۷] که عبارتند از: روش‌های سلسله مراتبی، روش‌های طیف فرکانسی، روش‌های مبتنی بر توان، روش‌های مبتنی بر کنتراست، آنتروپی، روش‌های مرکز-اطراف و روش‌های ترکیبی.

در روش‌های مبتنی بر ساختار سلسله مراتبی، چندین مقیاس از تصویر (تصویر برجستگی) ساخته می‌شود. از تفاضل میان این چندین مقیاس برای بدست آوردن تصویر برجستگی نهایی استفاده می‌شود. در مرجع [۴] از ساختار سلسله مراتبی استفاده شده است. نویسندگان برای هر تصویر یک ساختار هرمی تشکیل

اشاره کرد. در این روش‌ها از مجموعه آموزشی استفاده می‌کنند که بصورت نقاطی از تصویر با کلاس مثبت (نقاط برجسته) و نقاطی با کلاس منفی (نقاط غیر برجسته) می‌باشد. سپس با استفاده از طبقه‌بند RBF SVM به یادگیری مدل پرداخته می‌شود. روش‌های استفاده شده در [۱۷-۱۶-۱۵] ایده‌ای را مطرح کردند که مبتنی بر ادغام ویژگی‌های سطح پائین، متوسط و بالا با یکدیگر و به‌منظور استخراج وزن‌های بهینه^۹ در تشکیل تصویر برجستگی نهایی می‌باشد. در مرجع [۱۷] با استفاده از رگرسیون خطی حداقل مربعات^{۱۰} به آموزش وزن‌های مطلوب پرداخته شده است. در مرجع [۱۵] از طبقه‌بند غیر خطی AdaBoost به‌منظور استخراج وزن‌ها استفاده شده است. بکارگیری اپراتور غیرخطی باعث بهبود نتایج در پیش‌بینی برجستگی نسبت به رگرسیون خطی می‌شود. یادگیری با استفاده از ماشین بردار پشتیبان خطی دودویی در مرجع [۱۶] پیشنهاد شده است. نتایج این روش نیز نسبت به طبقه‌بند غیر خطی از دقت کمتری برخوردار می‌باشد.

نویسندگان مقاله [۱۸] زنجیره مارکوف را برای استخراج تصویر برجستگی پیشنهاد داده‌اند. به‌عبارت دیگر از زنجیره مارکوف برای تبدیل تصویر به سوپریکسل‌های تشکیل دهنده آن استفاده کرده‌اند. آنان دیورژانس نما و توزیع مکانی را به‌عنوان اشیاء برجسته و پس‌زمینه تصویر تعریف کرده‌اند. در این روش، از پیکسل‌های مرزی به‌عنوان گره‌های برگشت‌پذیر در زنجیره مارکوف استفاده شده و زمان برگشتی مورد محاسبه قرار گرفته است. با توجه به اینکه زمان طی شده از گره گذرا به برگشت پذیر وابسته به وزن‌هایی می‌باشد که برای هر مسیر در نظر گرفته می‌شود، آنان برای هر مسیر احتمالی را در نظر گرفتند که بیانگر نواحی است که در مرکز هر تصویر قرار می‌گیرد و حضور اشیاء برجسته محتمل می‌باشد. عیب این روش در ماهیت زنجیره مارکوف می‌باشد. مارکوف فرآیند تصادفی و بدون حافظه می‌باشد که توزیع احتمال شرطی حالت بعدی به حالت فعلی آن وابسته است و به رخداد‌های قبلی خود وابسته نمی‌باشد. این ماهیت باعث کاهش دقت در تشکیل سوپریکسل‌ها می‌شود. روشی که در مرجع [۱۹] ارائه گردیده است از تحلیل‌های چند مقیاسی^{۱۴} بر روی سوپریکسل‌ها بهره می‌برد. در ابتدا از ناحیه‌بندی چندمقیاسی برای استخراج سوپریکسل‌ها استفاده می‌شود. برای دستیابی به دقت بیشتر، نویسندگان استفاده از پارامترهای مختلف گوسین را پیشنهاد داده‌اند. در هر مقیاس سه تصویر ویژگی کنتراست محلی، یکپارچگی و بایاس مرکزی استخراج می‌شود. نقشه‌های برجستگی با در نظر گرفتن وزن‌های مجموع این نقشه‌ها و نرمالیزه‌سازی بدست می‌آیند. در انتها این نقشه‌ها با یکدیگر ادغام می‌شوند و تصویر برجستگی نهایی را بوجود می‌آورند. با توجه به تحلیل‌های

استخراج ویژگی‌های مرکز-اطراف، استفاده از کنتراست را پیشنهاد داده است. در این روش، تصویر به نواحی تشکیل دهنده آن تجزیه و تصویر کنتراست هر ناحیه تشکیل می‌شود. براساس معیار تعریف شده‌ای، تصویرها با یکدیگر ترکیب و تصویر برجستگی نهایی تشکیل می‌شود.

روش‌های ترکیبی از ترکیب کردن چندین روش برای بدست آوردن تصویر برجستگی نهایی استفاده می‌کنند. طبقه بندی کننده‌های مختلف برای این گروه تعریف می‌شود [۱۳]. روش پیشنهاد شده در مرجع [۱۳] از حرکت چشم و ماشین بردار پشتیبان برای استخراج تصویر برجستگی استفاده می‌کند. این روش مبتنی بر آموزش دادن می‌باشد. در روشی که در مرجع [۱۴] ارائه شده است، در ابتدا مولفه‌های روشنایی و حالت رنگ تصویر از یکدیگر تجزیه می‌شوند و بصورت جداگانه در استخراج تصویر برجستگی نهایی استفاده می‌شوند. به عبارت دیگر تصویر ورودی به دو فضای رنگی جدید CIELAB و CIELUV نگاشت داده می‌شوند. بر روی هر یک از این فضاها رنگ، تبدیل فوریه Hypercomplex اعمال می‌شود. سپس سه کانال A, L و B از فضای رنگی CIELAB و سه کانال U, L و V از فضای رنگی CIELUV، به‌عنوان ماتریس کوواریانس تعریف می‌شوند. در ادامه برای ترکیب کانال‌ها از ماتریس همگن استفاده می‌شود.

ساختار این مقاله بدینگونه می‌باشد که کارهای مرتبط در بخش ۲ آورده شده است. توضیحات مربوط به روش پیشنهادی در بخش ۳ ارائه شده است. این بخش شامل توضیحاتی در رابطه با استخراج نقشه‌های برجستگی، وزن‌دهی، نرمالیزه‌سازی و تصویر برجستگی نهایی می‌باشد. معرفی پایگاه‌های داده، نتایج شبیه‌سازی و مقایسه آن با روش‌های دیگر در بخش ۴ ارائه شده است. در بخش ۵، نیز نتیجه‌گیری مقاله ارائه شده است.

۲- کارهای مرتبط

در یک دهه اخیر، مدل‌های محاسباتی متنوعی برای پیش‌بینی خیرگی انسان ارائه شده است. در تضاد با روش‌های قبلی که از ویژگی‌های همچون رنگ، بافت و غیره برای استخراج تصویر برجستگی استفاده کرده‌اند، روش‌های جدید مبتنی بر یادگیری می‌باشند. به‌عبارت دیگر، یک مدل آموزش می‌بیند تا محل تمرکز نگاه^۸ انسان را پیش‌بینی کند. این فرآیند با استفاده از اطلاعات ضبط شده از خیرگی چشم و آموزش دادن طبقه بند انجام می‌گیرد [۱۷-۱۶-۱۵]. اطلاعات ذخیره شده از خیرگی چشم انسان بصورت ground-truth برای هر تصویر ارائه شده است.

از اولین کارهایی که در زمینه‌ی پیش‌بینی خیرگی مبتنی بر یادگیری پیشنهاد شده است می‌توان به روش‌های مرجع [۱۶]

بهره می برد که منجر به استخراج وزن های کارآمد در معماری شبکه عصبی کانولوشن و به تبع آن نفشه برجستگی نهایی می شود.

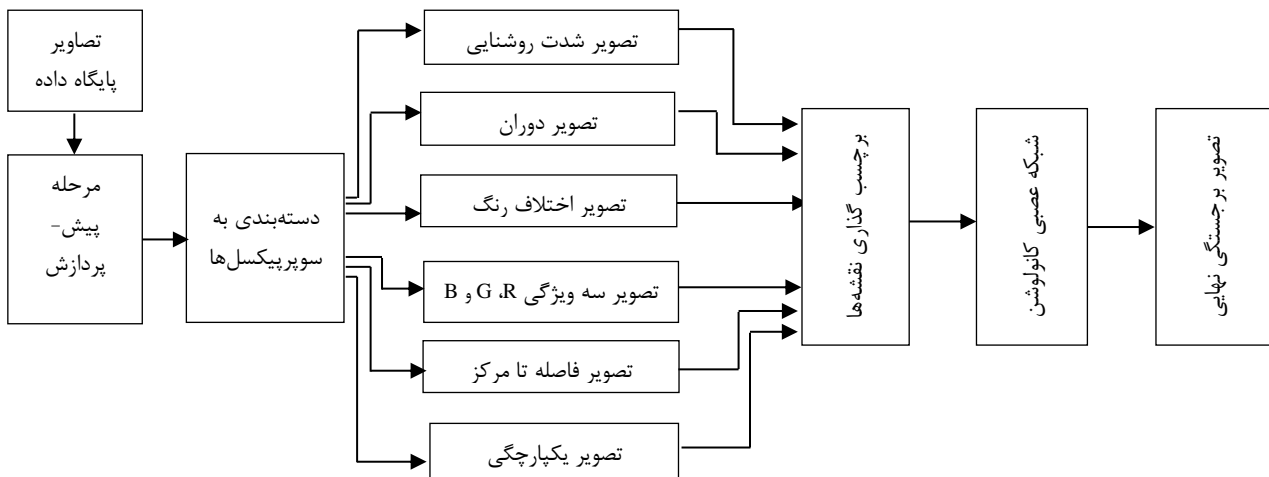
با توجه به اینکه اکثر کارهای انجام شده از ویژگی های سطح بالا به پائین در سوپرپیکسل ها استفاده می کنند که این امر باعث می شود کارآیی این روش ها محدود به کاربردهای خاص شود و همچنین روش هایی که از تحلیل های چند مقیاسی استفاده می کنند باعث افزایش زمان محاسباتی می شوند، در این مقاله روش متمایزی ارائه شده است که همزمان از ویژگی های پائین به بالا و بالا به پائین استفاده می شود که مشکل استفاده در کارآیی های خاص را برطرف می کند و همچنین روش پیشنهادی مبتنی بر یادگیری می باشد که در حالت آفلاین انجام می شود و باعث کاهش زمان محاسباتی می گردد. بنابراین روش پیشنهادی را می توان برای کاربردهای عمومی و خاص از جمله ردیابی اشیاء در ویدئو استفاده کرد. در روش پیشنهادی با استفاده از شبکه عصبی کانولوشن به آموزش دادن تصویر ویژگی های استخراج شده در تعیین تمرکز نگاه می پردازد. مراحل کار روش پیشنهادی برای مشخص نمودن شیء چشمگیر بدین گونه می باشد که در ابتدا تصویر به سوپرپیکسل های تشکیل دهنده آن تجزیه می شود. برای هر سوپرپیکسل ۸ تصویر ویژگی استخراج می شود. این ۸ تصویر ویژگی عبارتند از: سه ویژگی فضای رنگ R, G, B، یکپارچگی، بایاس مرکزی یا فاصله تا مرکز، ویژگی دوران، شدت روشنایی و تفاضل رنگ. در مرحله بعد، این ۸ تصویر ویژگی به میانگین صفر واریانس واحد نرمالیزه می شوند. برای آموزش دادن از شبکه عصبی کانولوشن استفاده شده است. توضیحات روش پیشنهادی در بخش ۳ آورده شده است. برای ارزیابی روش پیشنهادی با سایر روش های موجود، از ۴ پایگاه داده شناخته شده MSRA-1000, ECSSD, MSRA-10k و PASCAL-S استفاده شده است.

۳- روش پیشنهادی

در این بخش به توضیحات روش پیشنهادی پرداخته می شود که بلوک دیاگرام آن در شکل ۱ نمایش داده شده است. اولین بخش از روش پیشنهادی به این صورت است که پیش پردازشی بر روی تصاویر پایگاه داده انجام می گیرد و ابعاد تصاویر به 200×200 تغییر اندازه داده می شوند. این پیش پردازش بدلیل متفاوت بودن اندازه تصاویر پایگاه داده ها انجام می شود. از این تصاویر برای پردازش در مراحل بعدی استفاده می شوند. توضیحات روش پیشنهادی پیش بینی تمرکز نگاه در زیربخش های بعد آورده شده است. توضیحات مربوط به تبدیل تصویر به سوپرپیکسل ها در بخش ۳-۱ و مراحل استخراج ویژگی از سوپرپیکسل ها در بخش

چند مقیاسی که در این مقاله برای هر سوپرپیکسل در نظر گرفته شده است، زمان محاسباتی افزایش می یابد که مناسب برای کارهای برخط نمی باشد. روش دیگری که برای این منظور استفاده شده است، استفاده از ویژگی های سطح بالا به پائین برای سوپرپیکسل ها می باشد [۲۰]. در ابتدا و با استفاده از اطلاعات پیکسل هایی که بر روی مرز تصویر قرار گرفته شده است، سوپرپیکسل هایی برای پس زمینه تعریف می شود. در مرحله بعد با بهره گیری از حد آستانه وفقی و ناحیه بندی، سوپرپیکسل هایی برای پیش زمینه تعریف می شود. در انتها از روش ترکیبی ابتکاری برای ادغام نقشه های ویژگی حاصل از سوپرپیکسل ها استفاده می شود. عیب این روش این است که تنها برای کاربردهای خاص می توان از آن استفاده کرد که آن نیز بخاطر استفاده از ویژگی های بالا به پائین می باشد. در روش مرجع [۲۱] پس از اینکه تصویر به سوپرپیکسل های تشکیل دهنده آن تجزیه گردید، تصویر ویژگی های کنتراست محلی، رنگ و پس زمینه استخراج می شود و در ادامه از روش Quantum Cuts برای استخراج تصویر پیش زمینه استفاده می شود. خروجی هر دو مرحله با یکدیگر ترکیب می شوند و تصویر برجستگی نهایی را تشکیل می دهند. استفاده از ویژگی های سطح بالا به پائین که مبتنی بر آموزش دیکشنری بصری و یک حالت تصادفی است در مرجع [۲۲] پیشنهاد شده است. از این ویژگی ها بصورت وفقی استفاده می شود و به عنوان حالت تصادفی تعریف می شود. در ادامه با بهره گیری از یک ساختار باناظر مرحله آموزش انجام می گیرد. عیب این روش در بهره گیری از ویژگی های سطح بالا به پائین می باشد که برای کاربردهای خاص می توان از آن استفاده کرد.

در روشی که در مرجع [۲۳] پیشنهاد شده است، به منظور تعیین خیرگی از توصیف گر ویژگی SURF استفاده شده است. توصیف گر SURF منجر به تولید بردار ویژگی با ابعاد ۶۴ می شود. به عبارت دیگر، برای هر تصویر بک بردار ۶۴ بعدی استخراج می شود. به منظور آموزش دادن از اتوماتای یادگیر استفاده شده است. با استفاده از این طبقه بند، مرز میان پیکسل های برجسته و غیر برجسته شناسایی می شود. روش مرجع [۲۳] تنها از یک تصویر ویژگی برای تولید تصویر برجستگی نهایی استفاده کرده است و از طرفی برای مرحله آموزش دادن از بردار ویژگی با ابعاد ۶۴ بهره می گیرد. در روش پیشنهادی در این مقاله از هشت تصویر ویژگی در استخراج تصویر برجستگی نهایی استفاده شده است. روش پیشنهادی از تبدیل تصویر به سوپرپیکسل ها استفاده می کند که باعث کاهش ابعاد بردار ویژگی در مرحله آموزش می شود. به عبارت دیگر، تنها از ۲۰ ویژگی در مرحله آموزش استفاده می کند. همچنین روش پیشنهادی از شبکه عصبی کانولوشن در آموزش



شکل ۱: بلوک دیاگرام روش پیشنهادی

تصویر (r, c) می‌باشد. به منظور تعیین مراکز خوشه‌ها در سطح تصویر، شبکه منظمی از نقاط با فاصله برابر d روی سطح تصویر تشکیل می‌شود که از رابطه زیر محاسبه می‌شود:

$$d = \sqrt{P/K} \quad (1)$$

در این رابطه P بیانگر پیکسل‌های موجود در تصویر است. هر مرکز در یک پمجره همسایگی با ابعاد ۳ به طرف پیکس دارای کمترین مقدار گرادیان حرکت می‌کند. در مرحله بعد و به منظور نسبت دادن هر پیکسل به یکی از K دسته، پیکسل i با نزدیکترین مرکز دسته که همپوشانی با ناحیه جستجو دارد، نسبت داده می‌شود.

یکی از مزیت‌های روش SLIC نسبت به روش دسته‌بندی K -means این است که بجای جستجو در تمامی مراکز دسته‌ها، تنها با مرکز دسته‌ای جستجو انجام می‌گیرد که همپوشانی با آن دسته را داراست. نسبت دادن برای تمامی پیکسل‌ها انجام می‌گیرد. در الگوریتم K -means فاصله بین هر مرکز خوشه با تمام پیکسل‌ها در تصویر محاسبه می‌شود. در حالیکه در الگوریتم SLIC و به منظور کاهش حجم محاسبات، فاصله مراکز خوشه‌ها با پیکسل‌ها در یک همسایگی پیرامون مرکز سوپرپیکسل محاسبه می‌شود. در ساده ترین حالت این همسایگی بصورت مربعی با طول ضلع دو برابر فاصله شبکه سوپرپیکسل‌ها تعریف می‌شود. شکل ۲ بیانگر این موضوع می‌باشد. در مرحله بعد مراکز دسته‌های جدید مبتنی بر مرکز بردار پیکسل‌ها انجام می‌گیرد و دوباره عمل نسبت دادن برای پیکسل‌ها محاسبه می‌گردد.

این امر تا زمانیکه خطای همگرایی از یک مقدار آستانه از پیش تعریف شده کمتر گردد، تکرار می‌شود. در انتها نیز عمل پس-پردازش بر روی تصویر انجام می‌گیرد.

چنانچه پیکسلی به صورت جداافتاده‌ای از دسته قرار گرفته باشد، در این مرحله پیکسل به نزدیکترین دسته اختصاص داده می‌شود.

۳-۲ توضیح داده شده است. سپس در بخش ۳-۳ به آموزش دادن الگوریتم پیشنهادی بوسیله شبکه عصبی کانولوشن و تصویر برجستگی نهایی پرداخته شده است.

۳-۱- دسته‌بندی به سوپرپیکسل

در این مرحله تصاویر ورودی به سوپرپیکسل‌های تشکیل دهنده آن دسته‌بندی می‌شوند. این عمل با استفاده از روش ناحیه‌بندی مبتنی بر گراف انجام می‌گیرد. در این روش، اطلاعات مکانی و فضای ویژگی‌ها بطور همزمان و به نحو مناسبی در گراف ساخته شده، لحاظ می‌شوند. هر سوپرپیکسل یا ناحیه، متشکل از پیکسل‌هایی می‌باشد که از نظر خواصی مانند رنگ، بافت و یا کنتراست به یکدیگر شباهت دارند و هر یک از سوپرپیکسل‌ها، اطلاعات ناشی از یک شیء چشمگیر را در خود نگه می‌دارد. برای این منظور، تصویر ورودی بصورت گراف‌های تشکیل دهنده آن، $G=(V,E)$ نمایش داده می‌شود. که در این نمایش، V بیانگر تعداد یا مجموعه گره‌ها می‌باشد و E بیانگر مجموعه‌ای از یال‌های وزندار می‌باشد. برای محاسبه وزن یال‌ها، الگوریتم‌های مختلفی ارائه شده است [۲۴]. در این مقاله از روش SLIC برای تشکیل سوپرپیکسل‌ها استفاده شده است [۲۵]. روش SLIC از روش ساده‌ای در تشکیل سوپرپیکسل‌ها برخوردار می‌باشد. در این روش K بیانگر تعداد سوپرپیکسل‌ها می‌باشد. در این مقاله $K=200$ در نظر گرفته شده است. روش دسته‌بندی سوپرپیکسل SLIC بدین گونه می‌باشد که در گام اول K مرکز دسته اولیه با مقدار دهی اولیه انتخاب می‌شود. به عبارت دیگر فرآیند خوشه‌بندی با مقدار دهی اولیه مراکز خوشه‌ها بصورت $[l_i \ a_i \ b_i \ r_i \ c_i]^T$ در سیستم رنگی CIELAB شروع می‌شود. یک فضای ویژگی پنج بُعدی متشکل از مقادیر رنگی در سیستم رنگی CIELAB شامل l, a و b به همراه مقادیر حاوی اطلاعات مکانی مرکز خوشه در سطح

تصویر برجستگی دوران

با توجه به اینکه سیستم بینایی انسان به جهت تصویر واکنش نشان می‌دهد، از جهت به عنوان ابزاری در استخراج تصویر برجستگی استفاده شده است. برای بدست آوردن تصویر برجستگی جهت، از فیلتر گابور در چهار جهت ۰، ۴۵، ۹۰ و ۱۳۵ درجه و ترکیب آن‌ها با یکدیگر استفاده شده است. در روش ارائه شده در مرجع [۴]، از فیلتر گابور در سه مقیاس و در چهار جهت ۰، ۴۵، ۹۰ و ۱۳۵ درجه استفاده شده است که منجر به بدست آوردن ۱۲ تصویر ویژگی برای جهت می‌شود و افزایش زمان محاسباتی را در پی دارد. در روش پیشنهادی در این مقاله تنها از یک مقیاس استفاده شده است که منجر به تولید یک تصویر ویژگی می‌شود.

تصویر برجستگی اختلاف رنگ و سه ویژگی R، G و B

بر اساس اختلاف رنگی که در میان کانال‌های رنگ وجود دارد، تصویر ویژگی رنگ تعریف می‌شود [۴]. از اختلاف رنگ قرمز-سبز و آبی-زرد در استخراج کردن تصویر برجستگی استفاده می‌شود. روابط زیر بیانگر کانال‌های رنگ قرمز، سبز، آبی و زرد در فضای رنگ RGB می‌باشد:

$$R = r - (g + b) / 2 \quad (2)$$

$$G = g - (r + b) / 2 \quad (3)$$

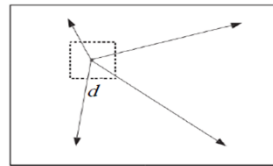
$$B = b - (g + r) / 2 \quad (4)$$

در روش پیشنهادی، از سه رابطه (۲)، (۳) و (۴) به عنوان سه تصویر ویژگی R، G و B استفاده شده است. علاوه بر این از اختلاف رنگ موجود میان سه رنگ R، G و B، اختلاف رنگ تعریف شده است که رابطه آن بصورت زیر است:

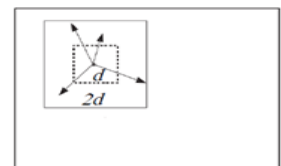
$$Y = (r + g) / 2 - (|r - g| / 2) - b \quad (5)$$

تصویر برجستگی فاصله تا مرکز

آزمایش‌ها نشان داده است که سیستم بینایی انسان به مرکز تصویر توجه بیشتری در مقایسه با سایر قسمت‌های تصویر از خود نشان می‌دهد [۲۶]. در روش پیشنهادی، مبتنی بر این خاصیت، تصویر برجستگی تعریف شده است. پس از اینکه مرکز تصویر مشخص گردید، فاصله هر سوپر پیکسل تا این مرکز مورد محاسبه قرار می‌گیرد. سوپر پیکسل‌هایی که به مرکز تصویر نزدیک‌تر می‌باشند، از وزن بیشتری برخوردار می‌باشند.



محدوده جست‌وجو در الگوریتم K-means



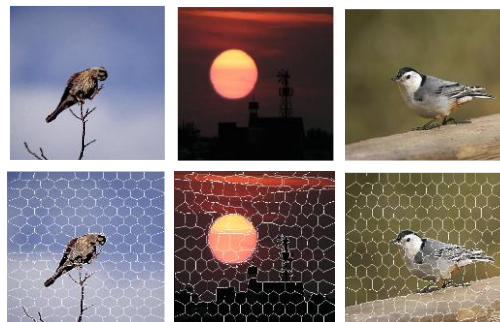
محدوده جست‌وجو در الگوریتم SLIC

شکل ۲: تفاوت جست‌وجو در دو الگوریتم SLIC و K-means

خروجی مرحله سوپرپیکسل‌ها که با استفاده از ناحیه‌بندی مبتنی بر گراف و با استفاده از روش SLIC انجام شده است، در شکل ۳ نمایش داده شده است.

۲-۳- استخراج ویژگی

پس از اینکه تصاویر به سوپرپیکسل‌های تشکیل دهنده آن تجزیه شدند، ۸ تصویر برجستگی ویژگی از سوپرپیکسل‌ها استخراج می‌شوند. این ویژگی‌ها عبارتند از: سه ویژگی فضای رنگ R، G، B، یکپارچگی، بایاس مرکزی، دوران، شدت روشنایی و اختلاف رنگ. در مرحله بعد، ویژگی‌های استخراج شده به میانگین صفر و واریانس واحد نرمالیزه می‌شوند. توضیحات هر یک از این ویژگی‌ها در ادامه توضیح داده شده‌اند.



شکل ۳: تجزیه تصویر به سوپرپیکسل‌های تشکیل دهنده آن

تصویر برجستگی شدت روشنایی

بررسی‌ها نشان می‌دهد که ناحیه‌ای برجسته از تصویر بوسیله مغز انسان انتخاب شده و وظیفه بررسی جزئیات آن ناحیه را به سیستم بینایی محول می‌کند. فرآیندی که بیانگر این انتخاب برجسته باشد، بطور کامل هنوز شناخته شده نیست [۱۶]. یکی از مهم‌ترین ابزارهای موجود در تصویر که واکنش سریع چشم را به همراه دارد، شدت روشنایی می‌باشد. با استفاده از این خاصیت، تصویر برجستگی شدت روشنایی استخراج می‌شود.

تصویر برجستگی یکپارچگی

تصویر برجستگی یکپارچگی^{۱۵}، از مهم‌ترین ویژگی در برجسته کردن شیء چشمگیر می‌باشد [۱۹، ۲۷]. رابطه یکپارچگی برای سوپرپیکسل بصورت زیر تعریف می‌شود:

$$q(u) = \begin{cases} \exp\left(-\frac{\mu}{\lambda \times E}\right) & \frac{\mu}{E} \leq \eta \\ 0 & O.W \end{cases} \quad (۶)$$

در این رابطه، μ تعداد پیکسل‌هایی که بر روی مرزهای یک سوپرپیکسل قرار گرفته اند را نشان می‌دهد، E بیانگر مجموعه پیکسل‌های مرزی تصویر ورودی است، η و λ پارامترهای حد آستانه می‌باشند که در این شبیه‌سازی به ترتیب دارای مقادیر ۰/۰۳۵ و ۰/۰۵ می‌باشند.

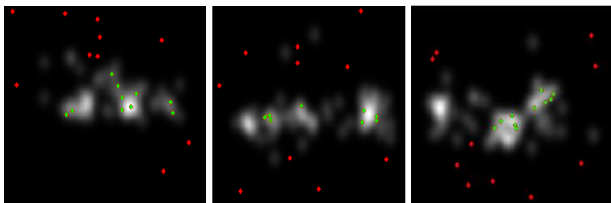
تمامی این هشت تصویر برجستگی ویژگی به میانگین صفر و واریانس واحد نرمالیزه می‌شوند. برای هر سوپر پیکسل، ماتریس ویژگی با ابعاد ۸ بُعدی تولید می‌شود. بنابراین یک ماتریس ویژگی برای هر تصویر تعریف می‌شود. این ماتریس ویژگی دارای ابعاد ۸ در تعداد سوپرپیکسل‌های تصویر می‌باشد. در شبیه‌سازی انجام گرفته شده، تعداد سوپرپیکسل‌ها برای هر تصویر ۲۰۰ در نظر گرفته شده است. مرحله بعدی آموزش دادن این ماتریس ویژگی می‌باشد. توضیحات آن در بخش بعدی شرح داده شده است.

۳-۳- آموزش دادن

در این مقاله به منظور آموزش دادن بردارهای ویژگی، از شبکه‌های عصبی کانولوشن استفاده می‌شود. شبکه‌های عصبی کانولوشن یکی از زیرشاخه‌های یادگیری عمیق است. همانگونه که در مرحله قبل اشاره شد، برای هر تصویر یک بردار یا ماتریس ویژگی ۸ بُعدی تعریف می‌شود. ابعاد این ماتریس ۲۰۰×۸ می‌باشد. عدد ۲۰۰ بیانگر تعداد سوپرپیکسل‌های تعریف شده برای هر تصویر می‌باشد. مقادیر این ماتریس به میانگین صفر و واریانس یک نرمالیزه شده است.

به منظور آموزش دادن، می‌بایست برجسته گذاری برای هر بردار از این ماتریس ویژگی در نظر گرفته شود. بنابراین هر بردار با برجسته ۱+ و ۱- تعریف می‌شود که با استفاده از تصویر برجستگی ground truth تصویر، این برجسته گذاری انجام می‌گیرد. تصویر برجستگی ground truth دارای مقادیر ۰ یا ۱ می‌باشند. به مقدار یک تصویر برجستگی ground truth، برجسته و به مقدار صفر، غیر برجسته گفته می‌شود. اگر تعداد پیکسل‌های برجسته تصویر برجستگی ground truth که درون سوپر پیکسل نام قرار گرفته است، از ۷۵ درصد تعداد پیکسل‌های آن سوپر پیکسل بیشتر باشد، برجسته

۱+، در غیر اینصورت برجسته ۱- به سوپر پیکسل نام اختصاص داده می‌شود. در این مقاله، ۱۰ سوپرپیکسل برجسته و ۱۰ سوپرپیکسل غیربرجسته در نظر گرفته شده است. شکل ۴ بیانگر تصاویر نمونه‌ای از سوپر پیکسل‌های برجسته و غیربرجسته می‌باشد که نقاط سبز بیانگر سوپرپیکسل‌های برجسته و نقاط قرمز بیانگر سوپرپیکسل‌های غیربرجسته می‌باشد.



شکل ۴: نمایش سوپرپیکسل‌های برجسته و غیر برجسته

پس از مرحله برجسته گذاری، مرحله آموزش دادن با استفاده از شبکه عصبی کانولوشن می‌باشد. یکی از مهم‌ترین و کاربردی‌ترین روش‌های یادگیری عمیق، شبکه عصبی کانولوشن می‌باشد. این شبکه از ساختارهای سلسله مراتبه‌ای تشکیل شده است که باعث می‌شود عمل آموزش به شیوه جدیدی انجام گیرد. ورودی شبکه عصبی کانولوشن، نواحی مختلف تصویر می‌باشد که با اعمال فیلتر بر روی ورودی، عمل کانوالو انجام می‌گیرد. به‌طور کلی شبکه عصبی کانولوشن از سه بخش کلی لایه کانولوشن، لایه ادغام و لایه تماماً اتصال تقسیم شده است [۲۸].

در لایه اول در شبکه عصبی کانولوشن، چندین لایه کانولوشن وجود دارد. وظیفه این لایه کانوالو کردن تصویر با فیلتر با ابعاد $n \times n$ می‌باشد. یکی از مزیت‌های این فیلتر این است که ضرائب آن قابل آموزش است. فیلترها بر روی تصویر ورودی حرکت داده می‌شوند. عمق فیلتر کانوالو با عمق تصویر ورودی یکسان است. وزن‌های این فیلترها در حین آموزش دادن به‌روز می‌شوند. نتایج حاصل از کانوالو با عدد بایاس جمع می‌شود و ذخیره می‌شود.

روش استفاده شده در این مقاله برگرفته از معماری AlexNet می‌باشد [۲۸]. این معماری از ۵ لایه کانوالو به همراه سه لایه تماماً اتصال تشکیل شده است. دلیل انتخاب معماری AlexNet این می‌باشد که از تعداد پارامترهای کمتری در ساختار آموزش تشکیل شده است. بلوک دیگرام لایه‌های تشکیل دهنده معماری شبکه عصبی کانولوشن در شکل ۵ نمایش داده شده است. در ادامه به توضیحات هر یک از این لایه‌ها پرداخته شده است.

لایه کانولوشن: تعداد ۵ لایه در لایه کانولوشن قرار گرفته شده است. در لایه اول کانولوشن، فیلتر با ابعاد 11×11 قرار گرفته است. این فیلتر در ابتدا بصورت تصادفی مقداردهی می‌شود و در ادامه ضرائب آن به‌روز می‌شوند. عمق این فیلتر دو می‌باشد که با عمق تصویر ورودی برابر می‌باشد. فیلتر بر روی نواحی تصویر عمل

می‌کند. بنابراین نیاز به دسته‌بندی داده‌ها با کمترین خطا می‌باشد. در این مقاله از روش پس انتشار خطا و گرادیان نزولی استفاده شده است. در این روش وزن‌ها در هر تکرار از رابطه (۷) محاسبه می‌شوند.

$$x = x - (Learning\ Rate) \times dx \quad (7)$$

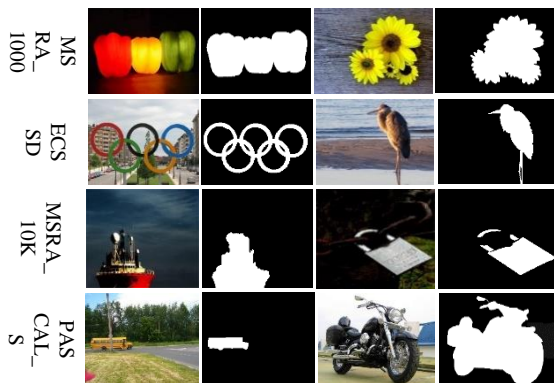
در این رابطه x بیانگر پارامترها و وزن‌های فیلترهای لایه کانولوشن و ادغام می‌باشد و Learning Rate نیز نرخ آموزش می‌باشد که در ابتدای فرآیند آموزش مقداردهی می‌شود.

پس از اینکه مرحله آموزش تمام شد، الگوریتم فرآیند استخراج پیکسل‌های برجسته و غیر برجسته را از تصویر ورودی آموزش دیده است که این امر منجر به شناسایی شیء چشمگیر می‌شود.

۴- نتایج شبیه‌سازی

۴-۱- پایگاه‌های داده

برای ارزیابی روش پیشنهادی، از پایگاه‌های داده‌ی MSRA-1000 [۲۹]، ECSSD [۳۰]، MSRA-10k [۳۱] و PASCAL-S [۳۲] استفاده شده است. در شکل ۶، چند نمونه از تصاویر این پایگاه‌های داده همراه با تصویر ground truth آن‌ها نمایش داده شده است.



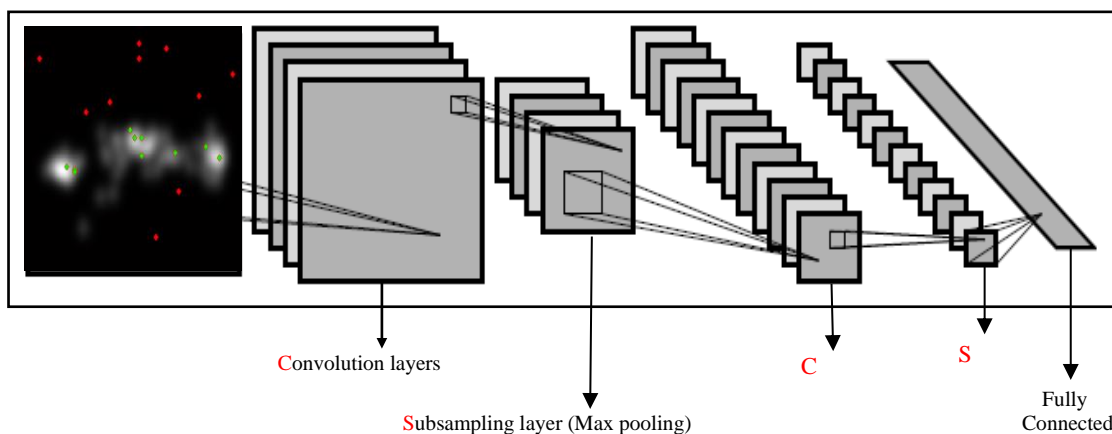
شکل ۶: تصاویر پایگاه‌های داده‌ی MSRA-1000، ECSSD، MSRA-10k و PASCAL-S به همراه ground truth.

کانوالو را انجام می‌دهد. به عبارت دیگر، وزن‌های فیلتر 11×11 با مقدار شدت روشنایی پیکسل‌ها کانوالو می‌شود و حاصل آن در یک پیکسل ذخیره می‌شود. سپس فیلتر بر روی ناحیه دیگر از تصویر اعمال می‌شود و عمل کانوالو را برای ناحیه جدید انجام می‌دهد و حاصل در پیکسل جدید ذخیره می‌شود. لایه دوم کانولوشن از ابعاد 5×5 برخوردار است و لایه‌های سوم تا پنجم این لایه از فیلتر یکسان با ابعاد 3×3 تشکیل شده است.

لایه ادغام: خروجی لایه کانولوشن، نقشه‌ای با تعداد پارامترهای زیاد می‌باشد. به منظور کاهش ابعاد و افزایش سرعت محاسبات، لایه ادغام تعریف می‌شود. هدف از لایه ادغام، نمونه‌برداری می‌باشد. روش‌های مختلفی برای نمونه‌برداری تعریف شده است: ماکزیمم، میانگین، مینیمم و غیره. در این مقاله از لایه ادغام ماکزیمم استفاده است. در این لایه پنجره‌ای با ابعاد 3×3 در نظر گرفته می‌شود. این پنجره بر روی خروجی حاصل از لایه کانولوشن اعمال می‌شود. این پنجره از میان ۹ پیکسل موجود، ماکزیمم مقدار را انتخاب می‌کند و به لایه بعدی انتقال می‌دهد. این عمل برای تمامی خروجی‌های حاصل از لایه کانولوشن انجام می‌گیرد. تعداد ۳ لایه در این بخش در نظر گرفته شده است. این سه لایه از ابعاد یکسان 3×3 تشکیل شده است.

لایه تماما اتصال: پس از ۵ لایه کانولوشن و ۳ لایه ادغام، لایه تماما اتصال قرار دارد. هدف تبدیل خروجی حاصل از لایه ادغام به یک بردار یک بعدی است. در این مقاله از سه لایه تماما اتصال استفاده شده است.

با توجه به برجسب‌های مشخص شده برای هر تصویر، در انتها خروجی لایه ادغام به طبقه‌بند Softmax داده می‌شود که عمل دسته‌بندی داده‌ها را انجام دهد. در فرآیند مرحله آموزش تصویر برجستگی، هدف پیدا کردن پارامترهای مجهول شبکه عصبی کانولوشن از جمله وزن‌های فیلترها و ضرائب لایه تماما متصل می‌باشد که نقش اساسی در فرآیند یادگیری تصویر برجستگی ایفا



شکل ۵: لایه‌های تشکیل دهنده در معماری شبکه عصبی کانولوشن [۲۸].

۴-۲- معیارهای ارزیابی

در این رابطه S بیانگر تصویر برجستگی روش پیشنهادی و G تصویر برجستگی $ground\ truth$ می‌باشد. به عبارت دیگر، MAE ، تفاضل میانگین مطلق میان تصویر برجستگی S و G را نشان می‌دهد. معیار OR نیز بصورت زیر تعریف می‌شود [۳۵]:

$$OR = |S' \cap G| / |S \cup G| \quad (12)$$

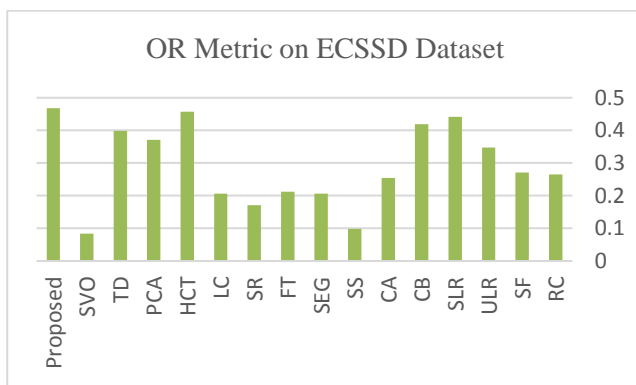
در این رابطه S' بیانگر باینری کردن تصویر برجستگی با استفاده از یک حد آستانه وقتی می‌باشد.

۴-۳- نتایج

در این قسمت به مقایسه روش پیشنهادی با سایر روش‌ها در پایگاه‌های داده‌ی MSRA-1000، ECSSD، MSRA-10k و PASCAL-S پرداخته می‌شود.

پایگاه داده ECSSD

روش پیشنهادی با روش‌های RC [۳۳]، SF [۳۴]، ULR [۳۶]، SLR [۳۷]، CB [۳۸]، CA [۳۹]، SS [۴۰]، SEG [۴۱]، FT [۴۲]، SR [۴۳]، LC [۴۴]، HCT [۴۵]، PCA [۴۶]، TD [۴۷] و SVO [۴۸] مقایسه و مورد ارزیابی قرار گرفته است. در شکل‌های ۷ تا ۱۰، ارزیابی روش پیشنهادی با سایر روش‌ها با استفاده از معیارهای OR ، AUC ، معیار F و MAE به ترتیب نشان داده شده است. معیارهای OR ، AUC ، FC و MAE برای روش پیشنهادی به ترتیب برابر است با: ۰/۴۶۸، ۰/۷۹۱، ۰/۴۵۱ و ۰/۲۳۱. برای معیار MAE ، عدد کوچکتر بیانگر عملکرد خوب روش می‌باشد. منحنی‌های $Precision-Recall$ و ROC برای پایگاه داده ECSSD در شکل‌های ۱۱ و ۱۲ نمایش داده شده است. نتایج حاکی از عملکرد بهتر روش پیشنهادی در این منحنی‌ها می‌باشد.



شکل ۷: معیار OR انواع روش‌ها در پایگاه داده ECSSD

به منظور مقایسه روش پیشنهادی با سایر روش‌های موجود، از چندین معیار ارزیابی استفاده شده است که عبارتند از: MAE ، AUC ، OR ، معیار F ، منحنی ROC و منحنی PR . برای مقایسه تصویر برجستگی روش پیشنهادی با $ground\ truth$ ، تصویر برجستگی باینری تعریف می‌شود.

تعدادی حد آستانه در بازه $[0, 1]$ در نظر گرفته شده است و مبتنی بر آن تصویر برجستگی باینری ساخته می‌شود. برای هر تصویر پایگاه داده، ۱۰ تصویر برجستگی باینری در نظر گرفته می‌شود و با تصویر $ground\ truth$ مورد ارزیابی قرار می‌گیرد. برای این ارزیابی مقادیر $precision$ و $recall$ بصورت زیر تعریف می‌شوند:

$$precision = TP / (TP + FP) \quad (8)$$

$$recall = TP / (TP + FN) \quad (9)$$

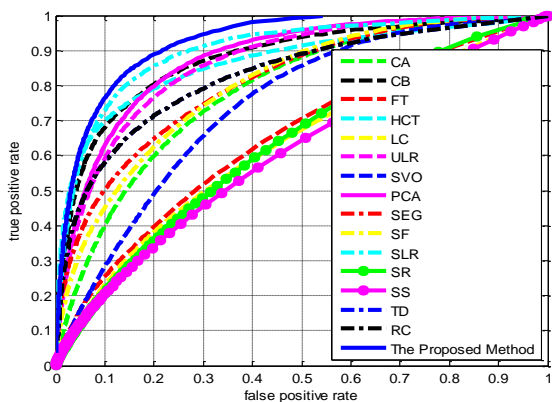
در این روابط، TP بیانگر پیکسل‌های برجسته در هر دو تصویر برجستگی باینری و $ground\ truth$ ، FP بیانگر تعداد پیکسل‌های برجسته در تصویر برجستگی باینری و غیر برجسته در تصویر $ground\ truth$ و FN بیانگر تعداد پیکسل‌های غیر برجسته در تصویر برجستگی باینری و برجسته در $ground\ truth$ می‌باشد. برای هر حد آستانه مقادیر $precision$ و $recall$ محاسبه می‌شود و میانگین آن منحنی $precision-recall$ را تشکیل می‌دهد. در این مقاله از معیار دیگری به عنوان F نیز استفاده شده است. رابطه F بصورت زیر بیان می‌شود:

$$F = \frac{(1 + \beta^2) \cdot precision \cdot recall}{\beta^2 \cdot precision + recall} \quad (10)$$

مقدار β در این رابطه ۰/۳ در نظر گرفته شده است [۳۳]. معیار بعدی، معیار ارزیابی منحنی PR می‌باشد. این منحنی با در نظر گرفتن ۱۰ تصویر برجستگی باینری ترسیم می‌شود. منحنی PR بیانگر این نکته می‌باشد که پیکسل متعلق به شیء چشمگیر می‌باشد یا خیر. منحنی ROC نیز با استفاده از منحنی PR بوجود می‌آید. این منحنی با استفاده از پارامترهای TP و FP ترسیم می‌شود.

معیار AUC بصورت اسکالر بیان می‌شود. این معیار ارزیابی، ناحیه زیر منحنی ROC را محاسبه کرده و بصورت عدد نمایش می‌دهد. معیار ارزیابی دیگر، معیار MAE می‌باشد که بصورت زیر تعریف می‌شود [۳۴]:

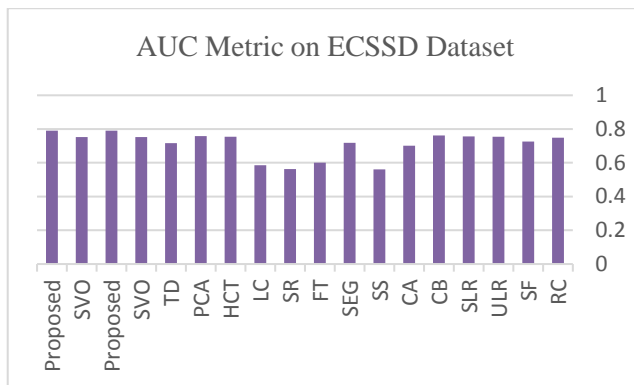
$$MAE = mean(|S - G|) \quad (11)$$



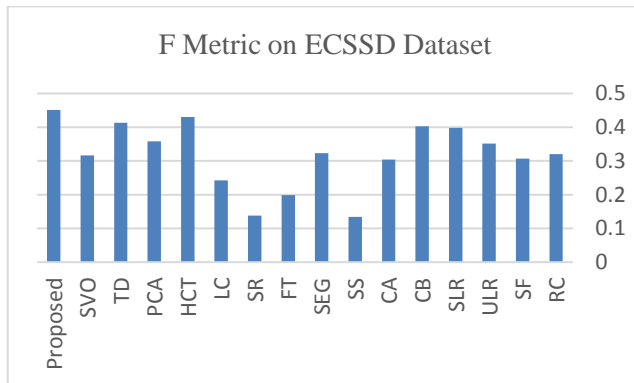
شکل ۱۲: منحنی ROC انواع روش‌ها در پایگاه داده ECSSD

پایگاه داده MSRA-10k

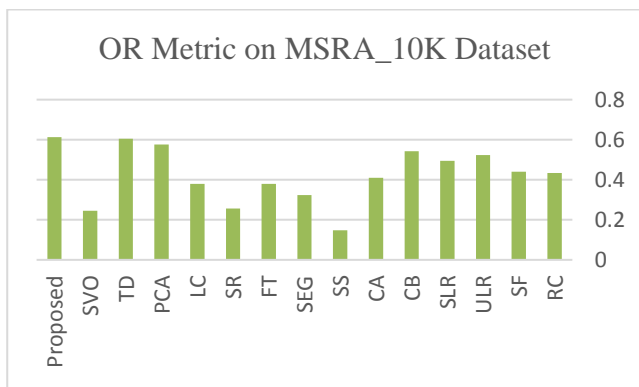
در این پایگاه داده، روش پیشنهادی با روش‌های RC [۳۳]، SF [۳۴]، ULR [۳۶]، SLR [۳۷]، CB [۳۸]، CA [۳۹]، SS [۴۰]، SEG [۴۱]، FT [۴۲]، SR [۴۳]، LC [۴۴]، PCA [۴۶]، TD [۴۷] و SVO [۴۸] مورد ارزیابی قرار گرفته است. نتایج این ارزیابی در شکل‌های ۱۳ تا ۱۶ نمایش داده شده است.



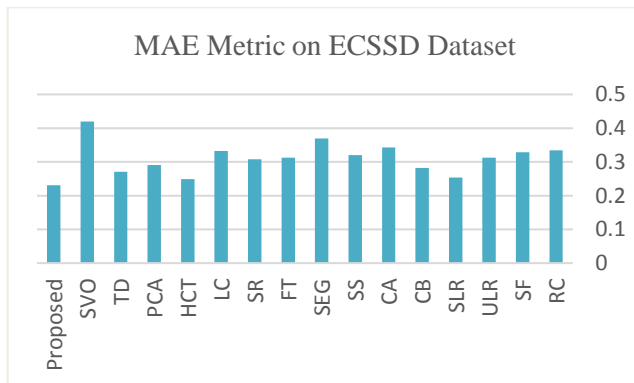
شکل ۸: معیار AUC انواع روش‌ها در پایگاه داده ECSSD



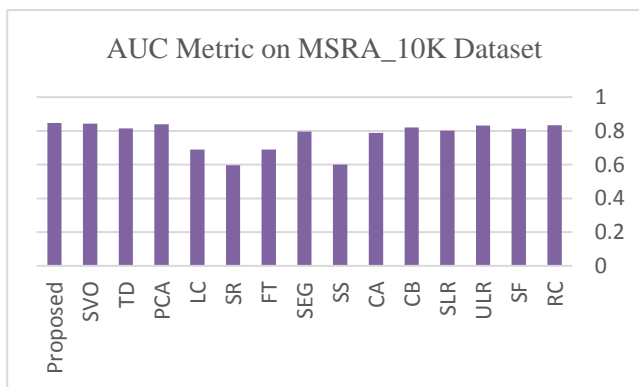
شکل ۹: معیار F انواع روش‌ها در پایگاه داده ECSSD



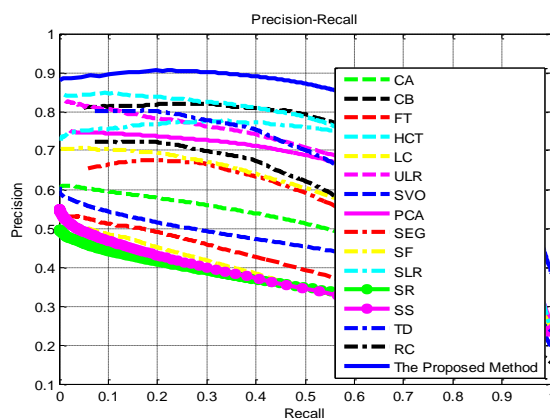
شکل ۱۳: معیار OR انواع روش‌ها در پایگاه داده MSRA_10K



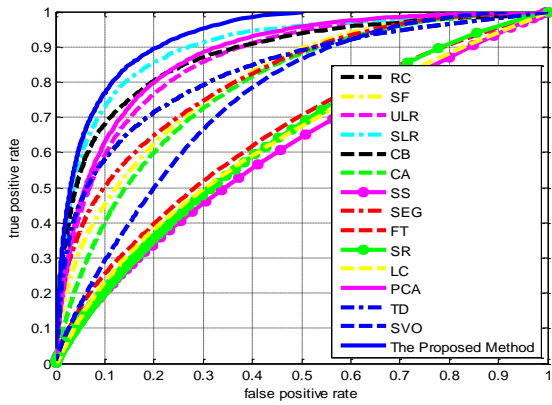
شکل ۱۰: معیار MAE انواع روش‌ها در پایگاه داده ECSSD



شکل ۱۴: معیار AUC انواع روش‌ها در پایگاه داده MSRA_10K



شکل ۱۱: منحنی PR انواع روش‌ها در پایگاه داده ECSSD



شکل ۱۸: منحنی ROC انواع روش‌ها در پایگاه داده MSRA-10k

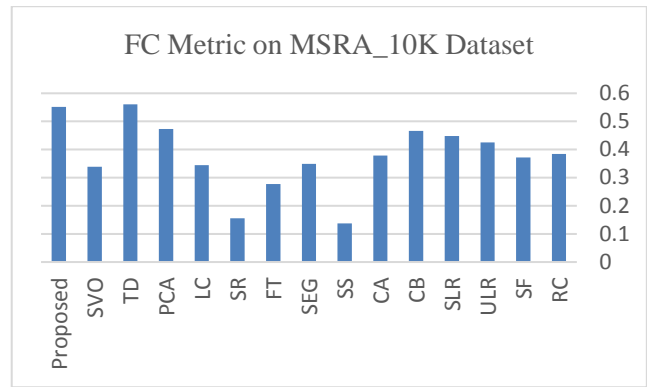
در جدول ۱، زمان اجرایی روش‌های مختلف در استخراج تصویر برجستگی در پایگاه داده MSRA-10K نمایش داده شده است. در این جدول، C بیانگر زبان برنامه نویسی C و M بیانگر متلب می‌باشد.

جدول ۱: مقایسه زمان اجرایی روش پیشنهادی با سایر روش‌ها در پایگاه داده MSRA-10K

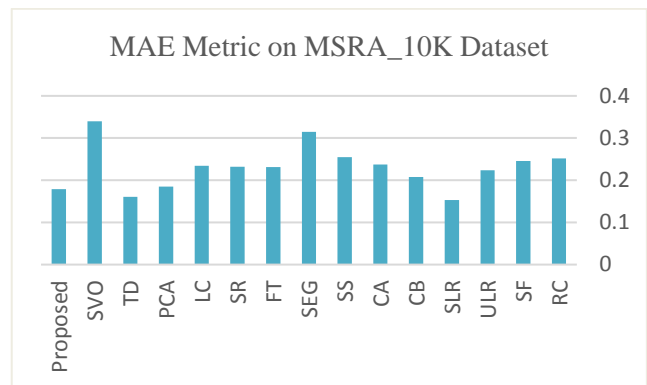
Time (s)	Code	روش
۰/۱۲۶	C	RC
۰/۲۰۲	C	SF
NA	NA	ULR
NA	NA	SLR
۲/۲۴	M+C	CB
۴۰/۹	M+C	CA
۷۹/۱	M+C	SS
۱۰/۹	M+C	SEG
۰/۷۲	C	FT
NA	NA	SR
NA	NA	LC
۴/۳۴	M+C	PCA
NA	NA	TD
۵۶/۵	M+C	SVO
۲/۱	M+C	روش پیشنهادی

پایگاه داده MSRA-1000

روش پیشنهادی با روش‌های IT [۴]، ZH [۱۱]، AC [۴۹]، AIM [۵۰]، GB [۵۱]، MZ [۵۲] و SR [۵۳] مورد ارزیابی قرار گرفته است. همانگونه که در شکل‌های ۱۹ تا ۲۲ قابل مشاهده است، روش پیشنهادی از عملکرد بهتری نسبت به سایر روش‌ها در پایگاه داده MSRA-1000 برخوردار می‌باشد. منحنی‌های Precision-Recall و ROC روش پیشنهادی و سایر روش‌ها در شکل‌های ۲۳ و ۲۴ به ترتیب نمایش داده شده است.



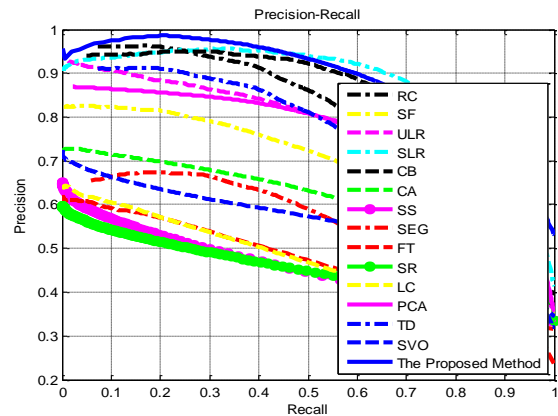
شکل ۱۵: معیار انواع روش‌ها در پایگاه داده MSRA_10K



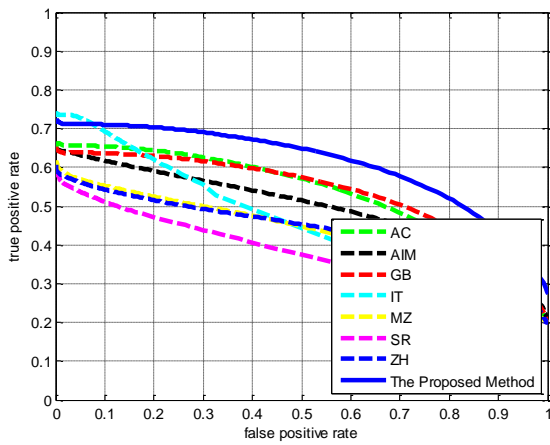
شکل ۱۶: معیار MAE انواع روش‌ها در پایگاه داده MSRA_10K

همانگونه که از شکل‌های ۱۳ تا ۱۶ قابل مشاهده است، معیارهای OR، AUC، FC و MAE برای روش پیشنهادی به ترتیب برابر است با: ۰/۶۱۳، ۰/۸۴۷، ۰/۵۵۳ و ۰/۱۷۹. روش پیشنهادی از عملکرد بهتری نسبت به سایر روش‌ها در پایگاه داده MSRA-10k برخوردار می‌باشد.

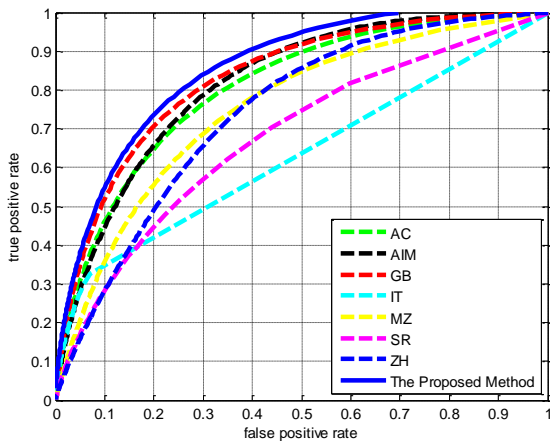
همانگونه که در شکل‌های ۱۷ و ۱۸ قابل مشاهده است، روش پیشنهادی دارای عملکرد بهتری در منحنی‌های Precision-Recall و ROC نسبت به روش‌های موجود در پایگاه داده MSRA-10k می‌باشد.



شکل ۱۷: منحنی PR انواع روش‌ها در پایگاه داده MSRA-10k



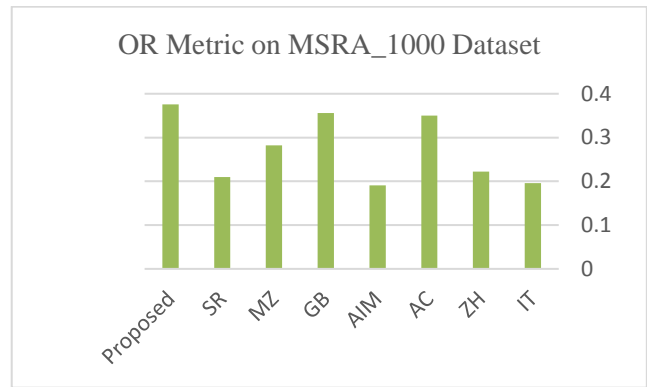
شکل ۲۳: منحنی PR انواع روش‌ها در پایگاه داده MSRA-1000



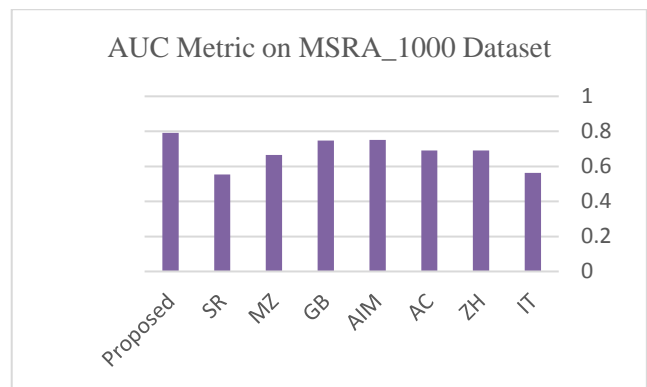
شکل ۲۴: منحنی ROC انواع روش‌ها در پایگاه داده MSRA-1000

پایگاه داده PASCAL-S

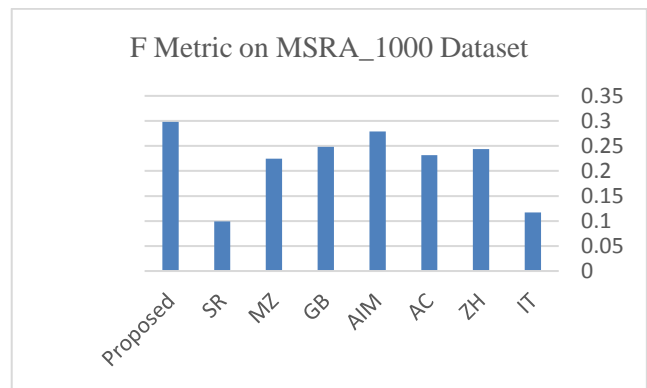
در این پایگاه داده نیز روش پیشنهادی با ۷ روش موجود مورد ارزیابی و مقایسه قرار گرفته شده است. این ۷ روش عبارتند از: IT [۴]، SUN [۱۱]، SS [۴۰]، AIM [۵۰]، GB [۵۱]، AWS [۵۴] و DVA [۵۵]. نتایج این ارزیابی در شکل‌های ۲۵ تا ۲۸ آورده شده است. معیارهای MAE، AUC، OR و FC برای روش پیشنهادی به ترتیب برابر است با: ۰/۲۷۴، ۰/۷۵۱، ۰/۳۴۱ و ۰/۴۲۷. روش پیشنهادی از عملکرد بهتری نسبت به سایر روش‌ها در پایگاه داده PASCAL-S برخوردار می‌باشد.



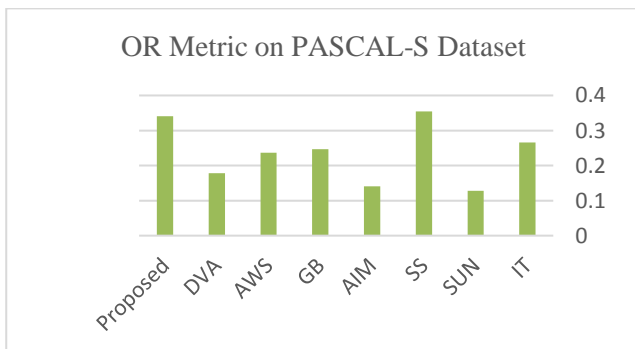
شکل ۱۹: معیار OR انواع روش‌ها در پایگاه داده MSRA_1000



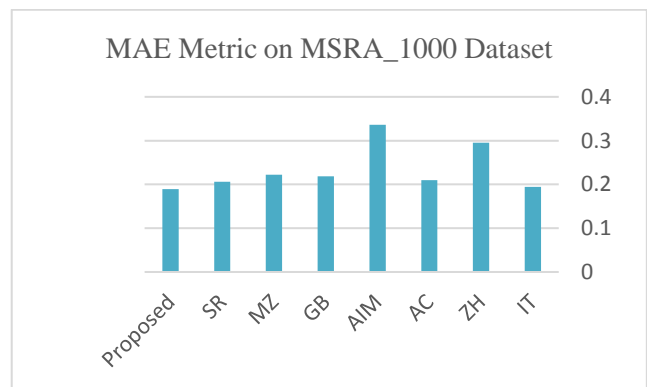
شکل ۲۰: معیار AUC انواع روش‌ها در پایگاه داده MSRA_1000



شکل ۲۱: معیار F انواع روش‌ها در پایگاه داده MSRA_1000

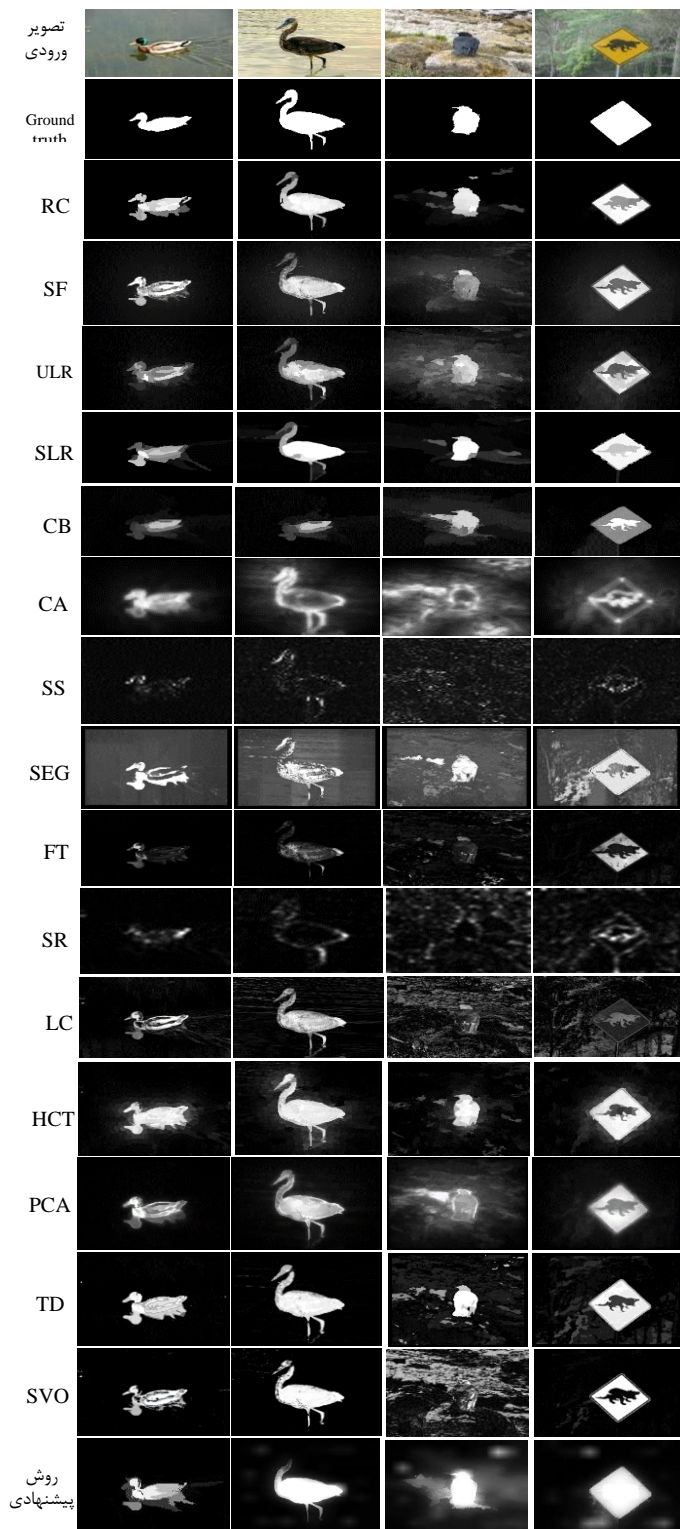


شکل ۲۵: معیار OR انواع روش‌ها در پایگاه داده PASCAL-S

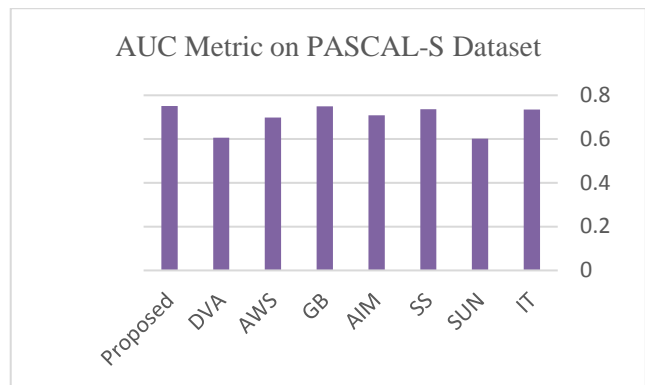


شکل ۲۲: معیار MAE انواع روش‌ها در پایگاه داده MSRA_1000

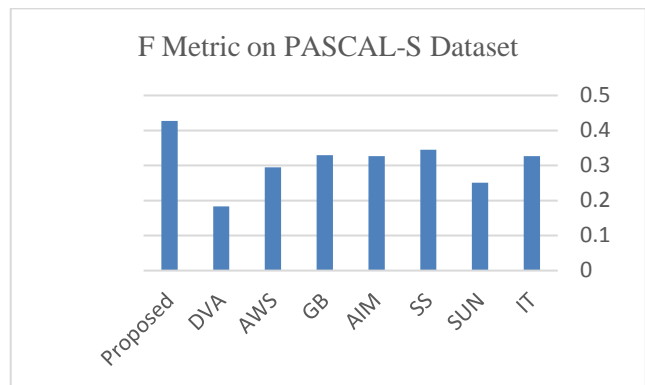
از شکل ۲۹ تا ۳۲ تصویر برجستگی روش پیشنهادی با سایر روش‌ها به ترتیب در ۴ پایگاه داده ECSSD، MSRA-10k، MSRA-1000 و PASCAL-S به همراه ground truth آن‌ها نمایش داده شده است.



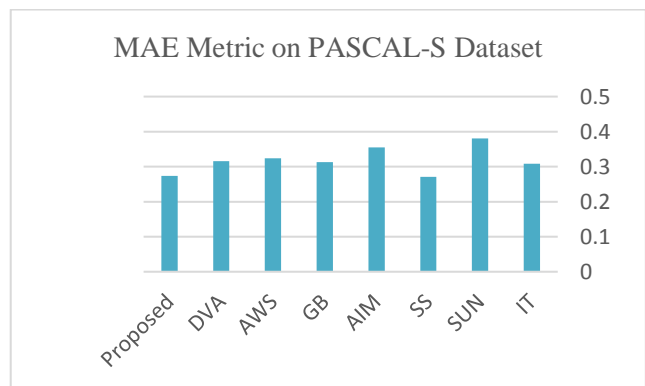
شکل ۲۹: تصویر برجستگی روش پیشنهادی و سایر روش‌ها در پایگاه داده ECSSD



شکل ۲۶: معیار AUC انواع روش‌ها در پایگاه داده PASCAL-S



شکل ۲۷: معیار F انواع روش‌ها در پایگاه داده PASCAL-S

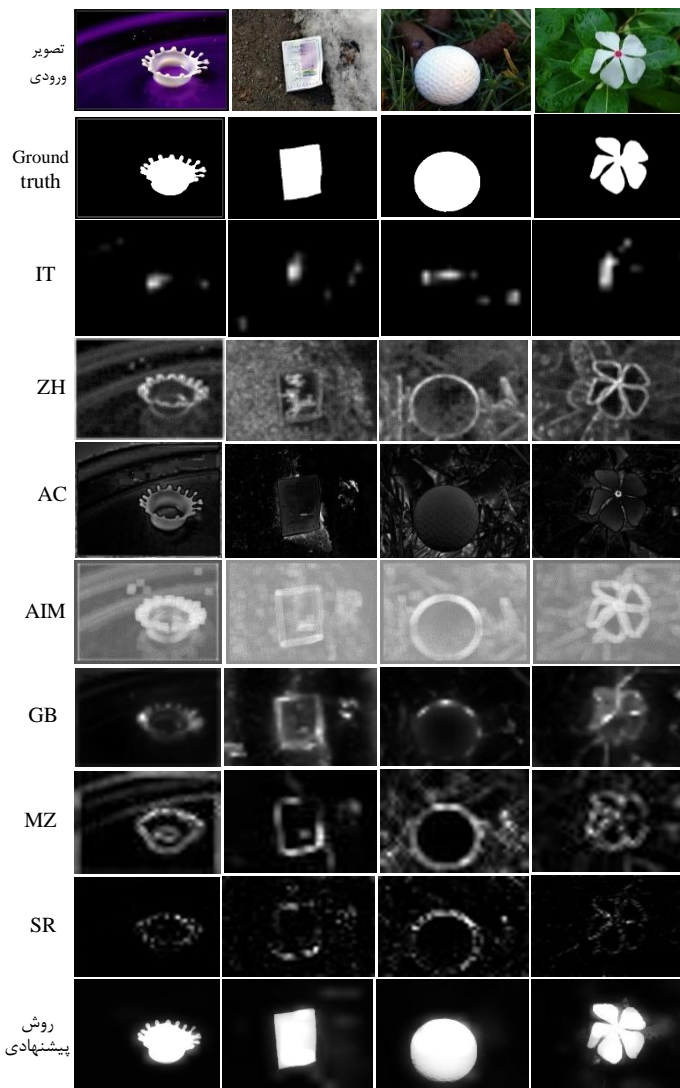


شکل ۲۸: معیار MAE انواع روش‌ها در پایگاه داده PASCAL-S

در جدول ۲ نیز به بررسی روش پیشنهادی با سایر روش‌ها در پایگاه داده MIT [۵۶] برای معیار AUC مورد ارزیابی قرار گرفته است.

جدول ۲: مقایسه معیار AUC روش پیشنهادی با سایر روش‌ها در پایگاه داده MIT

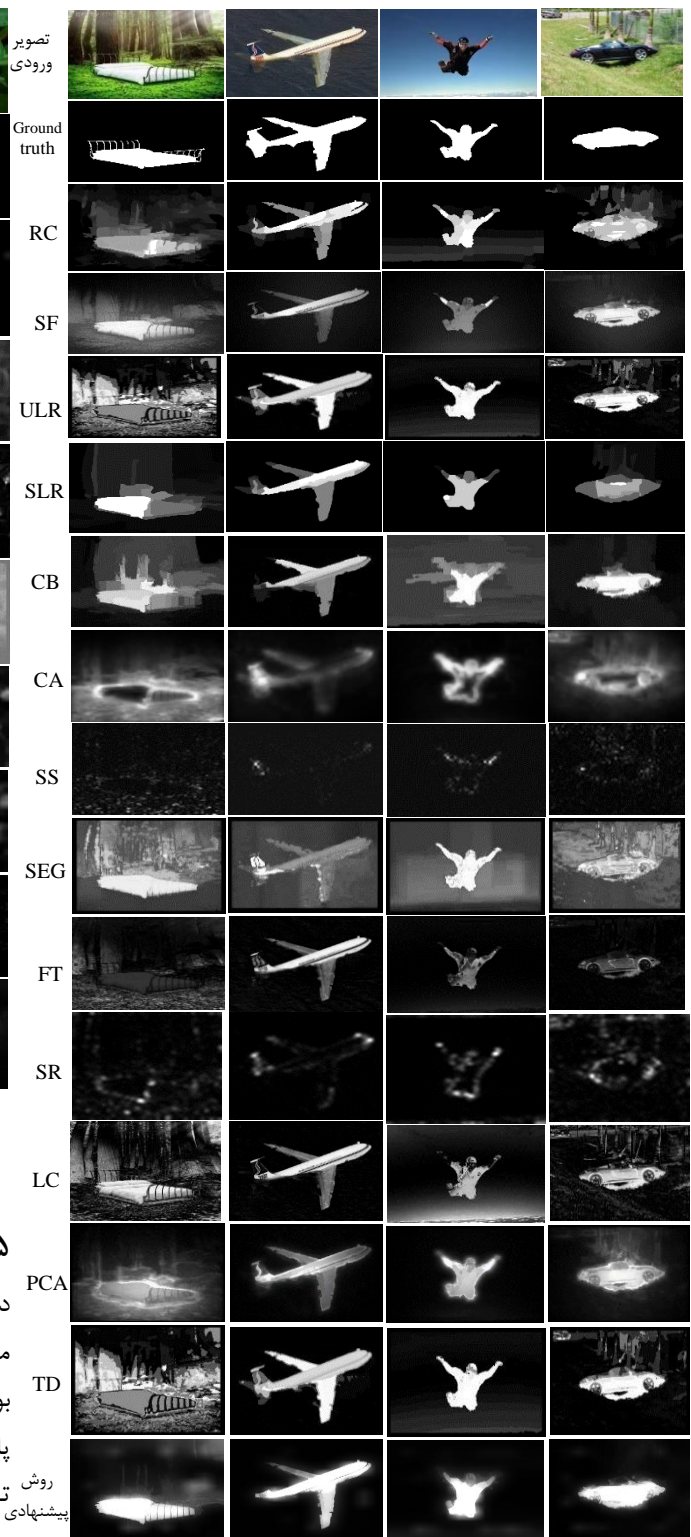
AUC	روش
۰/۱۶۸	HFT [57]
۰/۱۶۷۱	LG [58]
۰/۱۶۸۷	FES [59]
۰/۱۵۶۰	LDS [60]
۰/۱۶۹۱	Deep [61]
۰/۱۶۹۷	روش پیشنهادی



شکل ۳۱: تصویر برجستگی روش پیشنهادی و سایر روش‌ها در پایگاه داده MSRA-1000

۵- نتیجه گیری

در این مقاله روش متفاوتی برای تمرکز نگاه ارائه داده شد که مبتنی بر الگوریتم آموزش دادن و با استفاده از یادگیری عمیق بود. در این مقاله پیشنهاد استفاده از ویژگی‌های سطح بالا و سطح پائین داده شد که علاوه بر کاربردهای عام در کاربردهای خاص که تنها از ویژگی‌های سطح بالا استفاده می‌کرد، نیز قابل استفاده بود. روش پیشنهادی باعث کاهش زمان محاسباتی نیز شد. روش کار بدین صورت بود که پس از اینکه تصویر به سوپر پیکسل‌های آن تجزیه شد، برای تشکیل تصویر برجستگی، ۸ تصویر ویژگی از آن استخراج می‌شود که عبارتند از: سه ویژگی فضای رنگ B, G, R ، یکپارچگی، بایاس مرکزی، ویژگی دوران، شدت روشنایی و تفاضل رنگ. در این مقاله از معماری شبکه عصبی کانولوشن به منظور آموزش دادن استفاده شد که نقش بسزایی در استخراج تصویر برجستگی داشت. روش پیشنهادی با بسیاری از روش‌های موجود



شکل ۳۰: تصویر برجستگی روش پیشنهادی و سایر روش‌ها در پایگاه داده MSRA-10k

[5] Chenlei Guo and Liming Zhang, "A Novel Multiresolution Spatiotemporal Saliency Detection Model and Its Applications in Image and Video Compression," IEEE Transactions On Image Processing, Vol. 19, No. 1, pp. 10-25, 2010.

[6] N. Sang, L. Wei, and Y. Wang, "A Biologically-inspired Top-down Learning Model Based on Visual Attention," International conference on pattern recognition, 2010.

[7] Tadmeri Narayan Vikram, Marko Tscherepanow, Britta Wrede, "A saliency map based on sampling an image into random rectangular," Pattern Recognition, 2012.

[8] M. Wang, J. Li, T. Huang, Y. Tian, L. Duan, G. Jia, Saliency detection based on 2d log-gabor wavelets and center bias, in: Proceedings of the International Conference on Multimedia, 2010, pp. 979-982.

[9] V. Yanulevska, J.M. Geusebroek, Significance of the Weibull distribution and its sub-models in natural image statistics, in: Proceedings of the International Conference on Computer Vision Theory and Applications, 2009, pp. 355-362.

[10] R. Achanta and S. Susstrunk, "Saliency Detection for Contentaware Image Resizing," in IEEE International Conference on Image Processing, 2009.

[11] L. Zhang, M. H. Tong, T. K. Marks, H. Shan, and G. W. Cottrell, "SUN: A Bayesian framework for saliency using natural statistics," Journal of Vision, vol. 8, no. 7, pp. 1-20, 2008.

[12] T. Avraham, M. Lindenbaum, "Esaliency (extended saliency): meaningful attention using stochastic image modeling, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 4, pp. 693-708, 2010.

[13] W. Kienzle, F.A. Wichmann, B. Sch"olkopf, M.O. Franz, "A nonparametric approach to bottom-up visual saliency, Advances in Neural Information Processing Systems, 2007, pp. 689-696.

[14] K. Huang, C. Zhu, and G. Li, "Saliency Detection by Adaptive Channel Fusion," IEEE Signal Processing Letters, vol. 25, no. 7, pp. 1059-1063, 2018.

[15] Q. Zhao and C. Koch, Learning visual saliency by combining feature maps in a nonlinear manner using adaboost", Journal of Vision, vol. 12, no. 6, pp. 1-15, 2012.

[16] T. Judd, K. Ehinger, F. Durand, and A. Torralba, Learning to predict where humans look", in Proceedings of International Conference on Computer Vision, 2009.

[17] Zhao and C. Koch, Learning a saliency map using _xated locations in natural scenes", Journal of Vision, vol. 11, no. n3, pp. 1-15, 2011.

[18] B. Jiang, L. Zhang, H. Lu, C. Yang, and M.-H. Yang, "Saliency Detection via Absorbing Markov Chain," 2013 IEEE International Conference on Computer Vision, Dec. 2013.

[19] N. Tong, H. Lu, L. Zhang, X. Ruan, "Saliency Detection with Multi-Scale Superpixels", IEEE Signal Processing Letters, Vol. 21, No. 9, pp. 12-19, 2014.

[20] Wang, H. Lu, X. Li, N. Tong, and W. Liu, "Saliency detection via background and foreground seed selection," Neurocomputing, vol. 152, pp. 359-368, Mar. 2015.

[21] C. Aytekin, H. Possegger, T. Mauthner, S. Kiranyaz, H. Bischof, and M. Gabbouj, "Spatiotemporal Saliency Estimation by Spectral Foreground Detection," IEEE Transactions on Multimedia, vol. 20, no. 1, pp. 82-95, Jan. 2018.

[22] J. Yang and M.-H. Yang, "Top-Down Visual Saliency via Joint CRF and Dictionary Learning," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 3, pp. 576-588, Mar. 2017.

[23] H. Farsi, R. Nasiripour, S. Mohammadzadeh, "Eye gaze detection based on learning automata by using SURF descriptor," Information Systems & Telecommunication, vol. 6, no. 1, pp. 41-49, 2018.

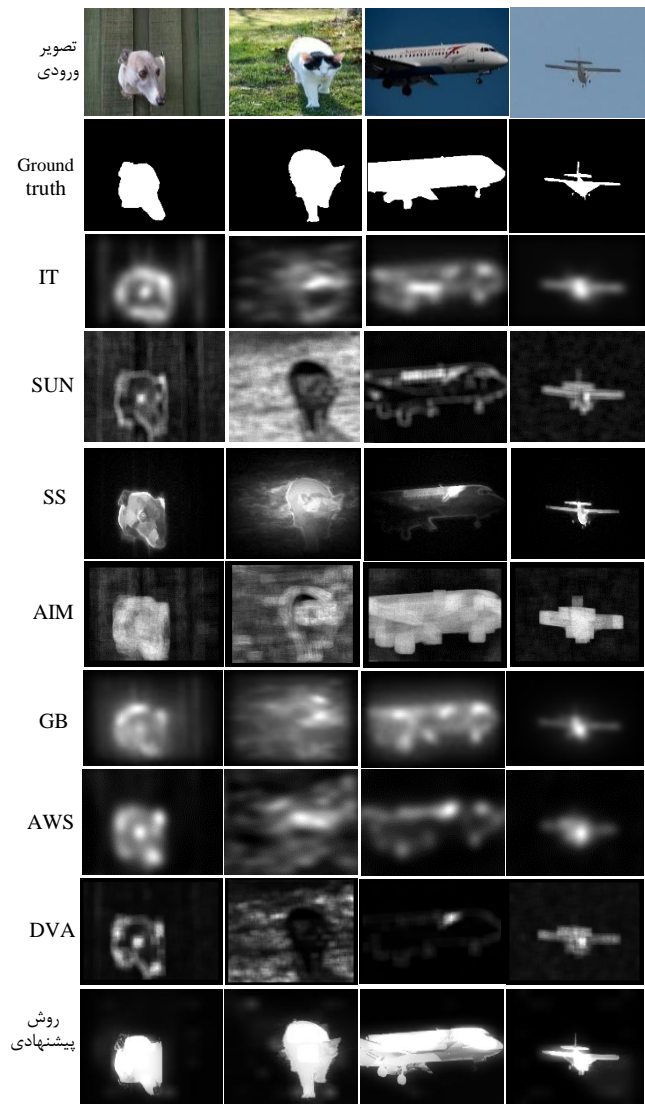
[24] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," IJCV, vol. 59, no. 2, pp. 167-181, 2004.

[25] O., Csillik, "Fast segmentation and classification of very high resolution remote sensing data using SLIC superpixels," Remote Sensing, vol. 9, no. 3, pp. 243-250, 2017.

[26] N. Tong, Lu H, R. Xiang, M. Hsuan. "Salient object detection via bootstrap learning", IEEE Conference on Computer Vision and Pattern Recognition, USA, pp. 1884-1892, 2015.

[27] G., Li, Y., Yu, "Deep contrast learning for salient object detection," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 478-487, 2016.

و در ۴ پایگاه داده شناخته شده ECSSD, MSRA-1000, MSRA-10k و PASCAL-S مورد ارزیابی و مقایسه قرار گرفت. نتایج بدست آمده بیانگر کارایی روش پیشنهادی نسبت به سایر روش‌ها در زمینه پیشگویی شیء چشمگیر و در تعیین خیرگی نگاه بود.



شکل ۲۲: تصویر برجستگی روش پیشنهادی و سایر روش‌ها در پایگاه داده PASCAL-S

مراجع

[1] A. Borji, D. Sihite, and L. Itti, "Quantitative analysis of humanmodel agreement in visual saliency modeling: A comparative study," IEEE TIP, vol. 22, no. 1, pp. 55-69, 2013.

[2] A. Borji. Boosting bottom-up and top-down visual features for saliency estimation. In CVPR, 2012.

[3] Q. Zhou, "Object-based attention: saliency detection using contrast via background prototypes," electronics letters, vol. 50, No. 14, pp. 997-999, 2014.

[4] L. Itti, C. Koch, and E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 11, pp. 1254-1259, 1998.

- [52] Y. F. Ma and H. J. Zhang, Contrast-based image attention analysis by using fuzzy growing. In ACM International Conference on Multimedia, 2003.
- [53] X. Hou and L. Zhang. Saliency detection: A spectral residual approach. IEEE Conference on Computer Vision and Pattern Recognition, 2007.
- [54] A. Garcia-Diaz, V. Lebor'an, X. R. Fdez-Vidal, and X. M. Pardo. On the relationship between optical variability, visual saliency, and eye fixations: A computational approach. Journal of Vision, Vol. 12, No. 6, 2012.
- [55] X. Hou and L. Zhang. Dynamic visual attention: Searching for coding length increments. In NIPS, pages 681–688, 2008.
- [56] B., Jiang, L., Zhang, H., Lu, Saliency detection via absorbing Markov chain. IEEE Int. Conf. on Computer Vision, Sydney, Australia, 2013.
- [57] X., Hu, L., Zhu, J., Qin, C.W., Fu, P.A. Heng. Recurrently aggregating deep features for salient object detection. In Thirty-Second AAAI Conference on Artificial Intelligence, 2018 .
- [58] X., Zhang, T., Wang, J., Qi, H., Lu, G. Wang. Progressive attention guided recurrent network for salient object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 714-722 .
- [59] W., Wang, J., Shen, X., Dong, A. Borji. Salient object detection driven by fixation prediction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 1711-1720 .
- [60] S., Fang, J., Li, Y., Tian, T., Huang, X., Chen, Learning discriminative subspaces on random contrasts for image saliency analysis. IEEE transactions on neural networks and learning systems, vol. 28, no. 5, pp.1095-1108, 2018.
- [61] B., Ghariba, M.S., Shehata, P. McGuire. Visual Saliency Prediction Based on Deep Learning. Information, vol. 10, no. 8, 2019.
- [28] E. A. Smirnov, D. M. Timoshenko, and S. N. Andrianov, "Comparison of Regularization Methods for ImageNet Classification with Deep Convolutional Neural Networks," AASRI Procedia, vol. 6, 2014, pp. 89–94.
- [29] R. Achanta, S. S. Hemami, F. J. Estrada, and S. Susstrunk, "Frequency-tuned salient region detection," in Proc. IEEE Conf. Computer. Vis. Pattern Recognition, 2009, pp. 1597–160.
- [30] Yan Q, Xu L, Shi J, Jiaya Jia, "Hierarchical saliency detection", IEEE Conference, USA, 2013, pp. 1155-1162.
- [31] <http://mmcheng.net/gsal/>.
- [32] Y. Li, X. Hou, C. Koch, J. Rehg, and A. Yuille, "The secrets of salient object segmentation," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2014, pp. 280–287.
- [33] Cheng M, Mitra NJ, Huang X, Torr PH, Hu S. Global contrast based salient region detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2015; 37(3):569–82.
- [34] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2012, pp. 733–740.
- [35] X. Li, Y. Li, C. Shen, A. R. Dick, and A. van den Hengel, "Contextual hypergraph modeling for salient object detection," in Proc. IEEE Int. Conf. Comput. Vis., 2013, pp. 3328–3335.
- [36] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2012, pp. 2296–2303.
- [37] C. Lang, G. Liu, J. Yu, and S. Yan, "Saliency detection by multitask sparsity pursuit," IEEE Trans. Image Process., vol. 21, no. 3, pp. 1327–1338, 2012.
- [38] H. Jiang, J. Wang, Z. Yuan, T. Liu, N. Zheng, and S. Li, "Automatic salient object segmentation based on context and shape prior," in Proc. Brit. Mach. Vis. Conf., 2011, pp. 1–12.
- [39] S. Goferman, L. Z. Manor, and A. Tal, "Context-aware saliency detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2010, pp. 1915–1926.
- [40] X. Hou, J. Harel, and C. Koch, "Image signature: Highlighting sparse salient regions," IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 1, pp. 194–201, 2012.
- [41] E. Rahtu, J. Kannala, M. Salo, and J. Heikkila, "Segmenting salient objects from images and videos," in Proc. Eur. Conf. Comput. Vis., 2010, pp. 366–379.
- [42] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, "Frequencytuned salient region detection," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2009, pp. 1597–1604.
- [43] X. Hou and L. Zhang, "Saliency detection: A spectral residual approach," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2007, pp. 1–8.
- [44] Y. Zhai and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," in Proc. 14th ACM Int. Conf. Multimedia, 2006, pp. 815–824.
- [45] J. Kim, D. Han, Y. Tai, and J. Kim, "Salient region detection via high-dimensional color transform," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2014, pp. 883–890.
- [46] R. Margolin, A. Tal, and L. Zelnik-Manor, "What makes a patch distinct?" in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2013, pp. 1139–1146.
- [47] C. Scharfenberger, A. Wong, K. Fergani, J. S. Zelek, and D. A. Clausi, "Statistical textural distinctiveness for salient region detection in natural images," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., 2013, pp. 979–986.
- [48] K.-Y. Chang, T.-L. Liu, H.-T. Chen, and S.-H. Lai, "Fusing generic objectness and visual saliency for salient object detection," in Proc. IEEE Int. Conf. Comput. Vis., 2011, pp. 914–921.
- [49] R. Achanta, F. J. Estrada, P. Wils, and S. Susstrunk, "Salient region detection and segmentation," in Proc. 6th Int. Conf. Comput. Vis. Syst., 2008, pp. 66–75.
- [50] N.D.B. Bruce, J.K. Tsotsos, Saliency based on information maximization, Advances in Neural Information Processing Systems, 2005, pp. 155–162.
- [51] J. Harel, C. Koch, P. Perona, Graph-based visual saliency, Advances in Neural Information Processing Systems, 2007, pp. 545–552.

زیر نویس ها

- 1 Visual Attention
- 2 Automatic image cropping
- 3 Bottom-Top Saliency
- 4 Top-Down Saliency
- 5 Inter Wavelet Transform
- 6 Spectral domain
- 7 Independent Component Analysis
- 8 Eye Gaze Information
- 9 Optimal Weights
- 10 Linear, Least Square Regressions
- 11 Binary Linear SVM
- 12 Ensembles of Deep Networks
- 13 Convolutional Neural Networks
- 14 Multi-resolutions
- 15 Integrity
- 16 Mean Absolute Errors
- 17 Area Under ROC Curve
- 18 Overlap Ratio
- 19 F-measure score
- 20 The receiver operating characteristic
- 21 Precision-recall
- 22 True Positive
- 23 False Positive
- 24 False Negative
- 25 Convolutional Neural Network
- 26 Deep Learning
- 27 Pooling
- 28 Fully-connected
- 29 Simple Linear Iterative Clustering