



تشخیص برجستگی‌های تصاویر ۳۶۰ درجه برای کاربردهای شناسایی و تشخیص اشیا

سید حمید خاتمی

دانشجوی دکتری مهندسی برق، گروه الکترونیک، دانشکده برق و کامپیوتر، دانشگاه بیرجند

h.khatami@birjand.ac.ir

ارسال: مرداد ماه ۱۴۰۲ پذیرش: مهر ماه ۱۴۰۲

چکیده

تشخیص اشیا، ردیابی اشیا و پیش‌بینی سری‌های زمانی، از چالش‌های اساسی در بینایی ماشین است. یادگیری عمیق، گام‌های بسیار بزرگی در حل این چالش‌ها برداشته است؛ اما برای بسیاری از مشکلات، راه‌حل‌های رضایت‌بخشی که در واقعیت، کاربردهای مفیدی داشته باشد و بتوان از آن استفاده کرد، هنوز پیدا نشده است. در این مقاله، بر روی مسئله آشکارسازی برجستگی‌ها در تصاویر ۳۶۰ درجه تمرکز شد. در بحث آشکارسازی برجستگی‌ها از تصاویر حرارتی استفاده شد و برتری‌های تصاویر ۳۶۰ درجه با سرعتی بهتر از قبل، تشخیص داده شد. در سال‌های اخیر، استفاده از شبکه‌های عصبی کانولوشن برای پیش‌بینی برجستگی‌ها یا بینایی بهتر، مورد توجه روز افزون قرار گرفته است. در این مقاله، این موضوع را کمی گسترش دادیم. در این کار، فراتر از تشخیص برجستگی برای تصاویر دو بعدی معمولی پیش‌رفتیم و نقشه‌های برجسته‌سازی برای تصاویر ۳۶۰ درجه همه جانبه، پیش‌بینی شد. یک تصویر همه جانبه کروی، به ۶ وجه مکعب تقسیم شد و بطور جداگانه، پردازش انجام گرفت. یک مدل تشخیص برجستگی که در ابتدا برای تصاویر غیرهمه جانبه آموزش دیده است، به عنوان شبکه اصلی برای پیش‌بینی برجستگی در هر یک از این ۶ وجه مکعب، تنظیم شده است. نقشه‌های برجستگی پیش‌بینی شده، برای تمام وجه‌های مکعب، به همدیگر متصل شدند و با استفاده از یک رمزگذار (Encoder) خودکار کانولوشن، حذف نویز هموار شد. همچنین یک ماسک جدید، برای از بین بردن آثار محل اتصال، پیشنهاد شده است.

کلمات کلیدی: تشخیص برجستگی در تصاویر ۳۶۰ درجه، تصاویر ۳۶۰ درجه، شناسایی اشیا در تصاویر ۳۶۰ درجه.

۱- مقدمه

تصاویر و ویدئوهای ۳۶۰ درجه در سال‌های اخیر، با توسعه و پیشرفت در فناوری‌های واقعیت مجازی (VR) و واقعیت افزوده (AR) محبوبیت فزاینده‌ای یافته‌اند [۱]. مطالعات علوم شناخت بصری نشان می‌دهد که انسان‌ها، صحنه‌ها را با شدت تمرکز یکسان مشاهده نمی‌کنند؛ بلکه نگاه انسان به قسمت‌های برجسته‌تر و مرتبط‌تر در صحنه جذب می‌شود. تشخیص برجستگی‌ها یا همان قسمت‌های Bold، دارای طیف وسیعی از برنامه‌ها مانند محلی‌سازی شی، ردیابی شی، فشرده‌سازی ویدئو و زیرنویس تصاویر است. پیش‌بینی نقشه برجستگی‌ها در تصاویر و فیلم‌های ۳۶۰ درجه نیز، برای بسیاری از برنامه‌های VR و AR با تکیه بر این اطلاعات اساسی درک صحنه، بسیار مهم است. همچنین، پیش‌بینی برجستگی‌ها، برای پخش ویدئوی ۳۶۰ درجه نیز مفید خواهد بود [۲]. از آنجا که بخش ویدئویی که مشاهده می‌شود، دائماً توسط تغییرات میدان دید (FoV) کاربر هدایت می‌شود، دانستن منطقه برجستگی در صحنه کنونی، به پیش‌بینی FoV کمک می‌کند [۳]. کدگذاری ویدئویی مبتنی بر شفافیت یا تفاوت کیفیت نیز از استفاده محدود از منابع کدگذاری و پهنای باند سود خواهد برد [۴]. تشخیص نقاط برجسته تصویر، باعث کاهش محاسبات و

تسهیل در پردازش اطلاعات در زمینه شناسایی و ردیابی اشیاء می‌شود. در نتیجه، شاهد کاهش اشغال پهنای باند و صرفه‌جویی در پهنای باند می‌شود [۵]. شکل ۱، یک نمونه تصویر ۳۶۰ درجه را نشان می‌دهد.

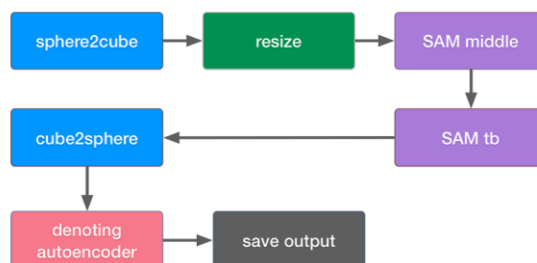


شکل ۱- نمونه تصویر ۳۶۰ درجه [۶]

روش‌های سنتی تشخیص برجستگی‌ها، بر اساس ویژگی‌های ساده‌ای مانند رنگ اشیاء یا اطلاعات تشخیص اشیاء سطح بالا مانند چهره است. در سال‌های اخیر، شبکه‌های عصبی پیچیده عمیق، با اختلاف زیادی از روش‌های سنتی بهتر عمل کرده‌است. معماری‌های جدید و مجموعه داده‌های بزرگ برای مشارکت در این زمینه عمومی می‌شوند [۷]. رایج‌ترین روش‌های پیش‌بینی برجستگی برای تصاویر ۳۶۰ درجه از روش‌هایی که برای تصاویر دوبعدی سنتی طراحی شده‌اند، گسترش یافته‌است. این روش‌ها به این صورت است که یک تصویر ۳۶۰ درجه را به چند تصویر دوبعدی تقسیم می‌کند و سپس عملیات پردازش تصاویر را انجام می‌دهد. این روش‌ها، زمان زیادی را هدر می‌دهد و پهنای باند زیادی مصرف می‌کند [۸].

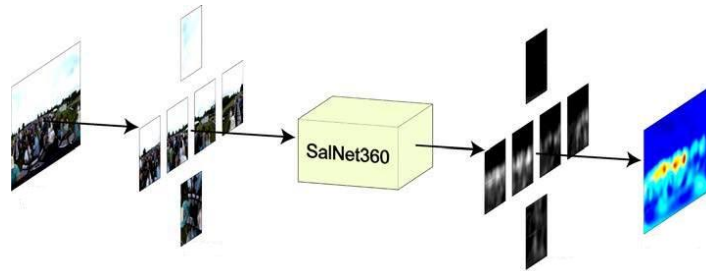
۲- روش پیشنهادی

در روش پیشنهادی این مقاله، ابتدا نقشه‌های مستطیلی به ۶ وجه مکعب تقسیم می‌شود؛ سپس سائز آن تغییر می‌کند. در مرحله بعدی که مرحله یادگیری و آموزش است، از دو مدل SAM استفاده می‌شود. در این مدل SAM، وجه‌ها به ترتیب به مجموعه SAM middle و SAM tb نشان تقسیم می‌شوند. به عبارت دیگر، در این مرحله با آگاهی از این که اثر اعوجاج و منحرف شدن توجه به بالا و پایین وجه‌ها، از چهار تصویر باقی مانده در وسط (چپ، جلو، راست و عقب) متفاوت است؛ دو مجموعه جداگانه از همان مدل SAM که از پیش آماده شده است، استفاده می‌شود. سپس نقشه‌های تبدیل شده به ۶ وجه مکعب، به تصویر مستطیلی تبدیل می‌شود. به عبارت دیگر، یک تبدیل معکوس از وجه مکعب به کره انجام می‌شود. سپس یک رمزگذار خودکار کانولوشنال ارائه شد تا بتوان نویز را حذف کرده و خروجی مطلوب را دریافت کرد. در شکل ۲ و شکل ۳، روش پیشنهادی این مقاله نشان داده شده‌است. سپس با رویکرد "تقسیم‌بندی کنید و سپس مغلوب کنید" ابتدا نقشه برجستگی‌ها را در هر وجه مکعب با تنظیم دقیق SAM که عملکردهای خوبی را در تصاویر معمولی دوبعدی نشان می‌دهد، تشخیص داده شد. سپس نقشه‌های برجسته پیش‌بینی شده برای تمام وجه‌های مکعب، دوخته شده و متصل شده و سپس از مکعب به نمای مستطیل شکل ترسیم می‌شوند.



شکل ۲- روش پیشنهادی سیستم تشخیص برجستگی‌ها

برای از بین بردن ناپیوستگی و آرتیفکت اضافه شده در محل اتصال و دوخت، یک رمزگذار خودکار کاهش‌دهنده نویز کانولوشنال طراحی شد تا نقشه برجستگی در هر وجه مکعب را صاف و دوباره نمایان سازد. به عبارت دیگر، از یک انکدر خودکار برای اتصال بهتر و تمیز ۶ وجه مکعب استفاده شد تا تصویر کره یا همان ۳۶۰ درجه بدست آید.



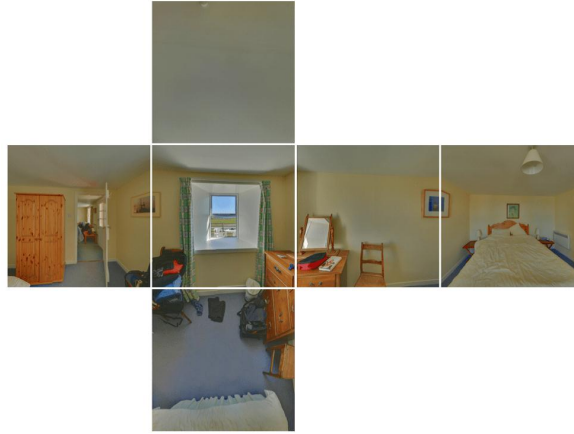
شکل ۳- تشخیص برجستگی‌ها با روش پیشنهادی به زبان ساده

در این مقاله از مدل سطح بالاتری برای تشخیص برجستگی در تصاویر دوبعدی منظم، یعنی همان مدل ویژگی‌های برجستگی (SAM-ResNet) استفاده شد، که از پیش در مجموعه داده SALICON آموزش داده شده بود [۹]. همچنین از ResNet برای استخراج ویژگی‌ها استفاده شد. در مرحله پیش‌پردازش داده‌ها، نمایش‌های مستطیلی را به نقشه‌های وجه مکعب برای تصاویر محرک و نقشه‌های برجستگی تبدیل کردیم. این عملیات را می‌توان در شکل ۴ و شکل ۵ مشاهده کرد.



شکل ۴- نمونه تصویر مستطیلی از مجموعه داده آموزش

بر این اساس، کره به شش نقشه برجستگی وجه‌های مکعب تبدیل می‌شود. در حین آموزش، یک تصویر خاص از مکعب و نقشه برجستگی مکعب به عنوان یک جفت در نظر گرفته می‌شود. در مرحله آموزش، با آگاهی از این که اثر اعوجاج و منحرف شدن توجه به بالا و پایین وجه‌ها، از چهار تصویر باقی مانده در وسط (چپ، جلو، راست و عقب) متفاوت است؛ دو مدل جداگانه از همان مدل SAM که از پیش آماده شده است، استفاده می‌شود. در این دو مجموعه، وجه‌ها به ترتیب به عنوان SAM middle و SAM tb نشان داده شده‌اند. همچنین در طول آموزش، مشاهده شد که آموزش با نسبت ابعاد مختلف، به بهبود نتایج کمک می‌کند. از این رو، دو مجموعه از وجه‌های مکعب و نقشه‌های برجستگی، پیش‌پردازش می‌شوند؛ یکی با رزولوشن 192×256 پیش‌پردازش می‌شود و دیگری با رزولوشن 256×256 . ابتدا مدل را روی وجه‌های مکعبی 192×256 تنظیم شد و سپس راه-اندازی مجدد مدل روی جفت 256×256 ادامه یافت. پس از آموزش، عملکرد شبکه حاصل در ۲ مقیاس متفاوت ارزیابی شد. واضح است که پس از تنظیم دقیق و بررسی، عملکرد بهتری را نسبت به مدل SAM از پیش تنظیم شده در تصاویر معمولی دو بعدی شاهد هستیم.



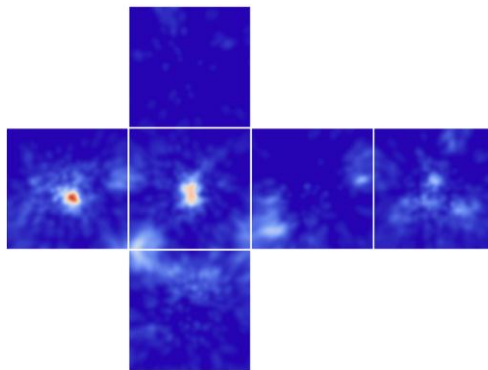
شکل ۵- تصویر حاصل از تصویر مستطیلی که تبدیل به شش وجه مکعب شده است

پس از بدست آوردن همه پیش‌بینی‌ها از هر وجه مکعب، تبدیل معکوس از وجه مکعب به کره انجام می‌شود تا نقشه برجستگی تصویر اصلی "مستطیلی شکل" ایجاد شود. از آنجا که نقشه برجستگی برای هر وجه مکعب به طور مستقل پیش‌بینی شده است؛ نقشه برجستگی‌ها فقط از نظر موضعی و در مکانی خاص، صاف و هموار است. برای کاهش این موضوع، از تکنیک بلور کردن گوسی استفاده شد. ولی هنوز نتایج رضایت بخشی نمی‌دهد. در این مقاله، نگاشت از "نقشه برجستگی دوخته شده" به نقشه برجستگی ground truth را به عنوان یک مشکل کاهش تصویر در نظر گرفتیم. این محصول مصنوعی و تصویری ارائه شده، در واقع متمایز و وابسته به مکان است. یک رمزگذار خودکار کانولوشنال را ارائه دادیم تا بتوان نویز را حذف کرده و پیش‌بینی را در هر وجه مکعب به طور همزمان حفظ کرد.

۳- نتایج

برای تصاویر ۳۶۰ درجه، بایاس بصری مشاهده شده این است که توجه در امتداد نوار افقی میانی متمرکز شده است. در این مقاله از مدل SAM در زمان پیش‌بینی نقشه برجستگی برای هر وجه مکعب استفاده شد. مدل SAM مجموعه‌ای از نقشه‌های قبلی گوسی را یاد می‌گیرد. سپس جهت‌گیری مرکز برای هر قطعه در نظر گرفته شد. به منظور کنترل بایاس نوار افقی، با الهام از کارهای قبلی، نقشه برجستگی میانگین را با پیش‌بینی‌های برجستگی جدا شده به عنوان ورودی برای رمزگذار خودکار نویز زدایی ترکیب شد. نقشه برجستگی میانگین، به عنوان راهنما برای شبکه عمل می‌کند و به شبکه کمک می‌کند تا نتیجه نهایی را براساس نقشه برجستگی تجربی، به عنوان مثال، برای ارتقا یا مهار پاسخ‌های برجسته محاسبه شده، مرتب کند. در شکل ۶، نقشه برجستگی تصویر را مشاهده می‌کنید.

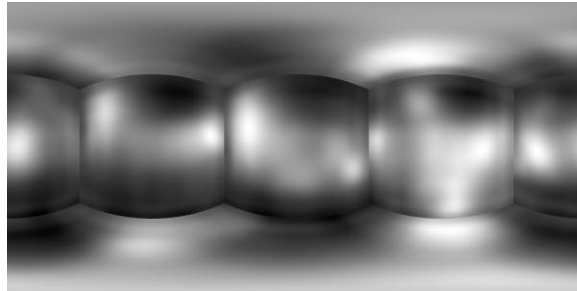
همچنین از لحاظ آماری مشاهده شده است که سهم قابل توجهی از تثبیت چشم انسان در ناحیه مرکزی یک تصویر قرار می‌گیرد. این تمایل انسان‌ها به خیره شدن به مرکز تصویر، در حالی که دید و بینایی آزاد به عنوان بایاس مرکزی در ثابت نگه داشتن چشم‌ها نامیده می‌شود، به طور گسترده در زمینه‌های علوم اعصاب و روان‌شناسی مورد مطالعه قرار گرفته است. برای مدیریت این بایاس مرکزی، در گذشته از یک بایاس مرکزی صریح و آشکار در مدل‌های پیش‌بینی برجستگی استفاده شده است.



شکل ۶- نقشه برجستگی تصویر

همچنین یک لایه کانولوشن با فضای دوبعدی (LBC)^۱ ارائه شده است. ۱۶ حباب یا لکه گوسی با میانگین صفر و واریانس‌های افقی و عمودی مختلف به عنوان ویژگی‌های خاص مکان، مورد استفاده قرار گرفته و به نقشه‌های ویژگی موجود متصل شده است. این ویژگی‌های خاص مکان در تمام مراحل آموزش ثابت هستند. در حالی که وزن فیلتر که بر روی ویژگی‌های الحاق شده عمل می‌کند در طول آموزش، یاد گرفته می‌شود.

با یادگیری مجموعه‌ای از پارامترهای گوسی، پیشرفت‌های ثابت قبلی را بهبود بخشید و آن‌ها را به صورت مداوم آموزش داد. هنگام تهیه نقشه برجستگی برای یک محل دوخت و اتصال، با ارائه اطلاعات اضافی در مورد اینکه در حوزه اصلی در کجا قرار دارد، مرحله اصلاح را پیشنهاد کرد و نقشه‌های مختصات کروی به نقشه‌های موجود متصل شد. شکل ۷، نقشه برجستگی متصل شده و دوخته شده را نشان می‌دهد.



شکل ۷- نقشه برجستگی متصل شده

با الهام از U-Net، یک شبکه عمیق U-Net مانند Encoder-Decoder، برای پیش‌بینی کل نقشه برجستگی براساس پیش‌بینی‌های نویزدار استفاده شد. در مقایسه با رمزگذار خودکار کانولوشن پیشنهادی قبلی، AE U-net، لایه‌های بیشتری دارد و همچنین مجهز به اتصالات پرشی (skip) است، که می‌تواند به حفظ اطلاعات مکان و ویژگی‌های دقیق‌تر کمک کند. در آزمایش‌ها، دریافتیم که استفاده از کانولوشن متسع (منبسط شده) برای ۶ لایه کانولوشن اول، در مقایسه با هسته‌های کانولوشن نرمال، عملکرد بهتری دارد. با دور شدن از هسته‌های کانولوشن معمول، هسته‌های کانولوشن متسع حفره‌هایی را در داخل هسته‌ها ایجاد می‌کنند و بنابراین بدون اضافه کردن پارامترهای اضافی، اندازه میدان‌های دریافتی را افزایش می‌دهند. نویز، پس از لایه ورودی برای اعمال فیلترهای متحرک برای یادگیری ویژگی‌های قوی‌تر معرفی شد. با توجه به داده‌های محدودی که داریم، نظم-دهی قوی‌تری را به شبکه اضافه می‌کنیم (نرخ افت ۰/۷). اندازه هسته شش لایه کانولوشن اول و هفت لایه آخر (جدا از آخرین لایه‌ها) ۷ است. یک لایه باریک (اندازه هسته ۱×۱) برای آخرین لایه کانولوشن استفاده می‌شود.

جدول ۱- مدل پیشنهادی آموزش داده شده با تصاویر ground truth

	فایل‌های باینری ground truth		تصاویر ground truth	
	KL	CC	KL	CC
1	0.6811	0.6901	0.0861	0.6623
2	0.8475	0.6631	0.0818	0.6823
3	0.4217	0.6953	0.0583	0.7468
4	0.5312	0.6032	0.0738	0.6449
5	0.4303	0.6945	0.0663	0.7204
6	0.7892	0.6159	0.0716	0.6262
7	0.8203	0.6628	0.0789	0.7036
8	0.543	0.5059	0.0766	0.5369
9	0.8609	0.6781	0.106	0.6899
10	1.1445	0.6703	0.1093	0.7001
11	0.6476	0.3847	0.0718	0.3866

¹ Location Biased Convolutional

12	0.3985	0.4421	0.0722	0.4713
13	0.4876	0.6994	0.0669	0.7491
14	0.7167	0.6169	0.067	0.6493
15	0.8752	0.686	0.084	0.7225
16	1.2011	0.6948	0.0883	0.7331
17	0.7677	0.696	0.0856	0.7276
18	0.7621	0.6351	0.0842	0.6595
19	0.6461	0.5547	0.0735	0.611
20	0.9346	0.5152	0.0936	0.5767
21	0.7467	0.7133	0.0797	0.7509
22	2.4683	0.6691	0.1161	0.7174
23	0.8144	0.5625	0.0809	0.5947
24	0.6101	0.4795	0.0567	0.4839
Average	0.7977	0.6178	0.0803	0.6477

نقشه‌های برجسته ground truth در دو قالب ارائه می‌شود؛ تصاویر برجستگی ground truth (فایل‌های با فرمت jpg) و فایل‌های باینری با دقت بسیار بالاتر برای هر پیکسل. برای سادگی آموزش، ابتدا با استفاده از تصاویر برجستگی، آزمایش انجام شد. نتایج نشان داده شده در جدول ۱ و جدول ۲ در برابر تصاویر ارزیابی می‌شوند. با این حال، از فایل‌های باینری با دقت بالا برای ارزیابی عملکردها استفاده شد. به طور خاص، ضرایب همبستگی ارزیابی شده مشابه هستند. در حالی که مدل این مقاله، تمایل به واگرایی KL بزرگتری با فایل باینری دارد که این انتظار می‌رود که با آموزش مدل بر روی نسخه کوانتیزه شده توزیع داده‌ها (یعنی استفاده از تصاویر با مقادیر پیکسل عدد صحیح) به جای توزیع داده‌های اصلی، به این امر دست پیدا کند.

جدول ۲- عملکرد مدل SAM تنظیم شده در وجه‌های مکعب. آموزش و یادگیری بر روی تصاویر

	KL متقارن	KL	CC
از قبل آموزش دیده‌اند و تست در:			
192×256 middle	0.9889	1.1867	0.5475
256×256 middle	1.3485	1.6690	0.5321
192×256 tb	1.3421	1.4994	0.3511
256×256 tb	1.5582	1.9204	0.3365
Model 1 middle (آموزش و تست در 192×256)	0.5472	0.2658	0.6615
Model 1* middle (ادامه تنظیم دقیق و 256×256 و تست در 192×256)	0.5929	0.2761	0.6618
Model 1 middle (آموزش در 192×256 و تست در 256×256)	0.0778	0.0851	0.7150
Model 1* middle (ادامه تنظیم دقیق و 256×256 و تست در 256×256)	0.0421	0.0422	0.7310
Model 1 tb (آموزش و تست در 192×256)	0.7588	0.3618	0.2436
Model 1* tb (ادامه تنظیم دقیق و 256×256 و تست در 192×256)	0.6899	0.3185	0.3329
Model 1 tb (آموزش در 192×256 و تست در 256×256)	0.0876	0.0905	0.4041
Model 1* tb (ادامه تنظیم دقیق و 256×256 و تست در 256×256)	0.0517	0.0525	0.4112

هنگامی که مقادیر برجستگی بسیار زیاد باشد، تصاویر توانست اشباع رنگ زیادی داشته باشد؛ بنابراین توزیع داده‌های اصلی را قطع کرد [۱۰]. انتظار می‌رفت که در صورت استفاده از آموزش فایل‌های باینری برای تنظیم مجدد و همچنین برای کاهش نویز، عملکرد بهتری داشته باشیم.

در جدول ۳، عملکرد رمزگذار خودکار کانولوشن را مشاهده می‌کنید. در این جدول، آموزش و ارزیابی بر روی تصاویر ground truth انجام شده است. جدول ۴، نتایج را در زیرمجموعه آزمایش مجموعه داده نشان می‌دهد. طبق این جدول، اگر از فایل‌های باینری برای آموزش مدل پیشنهادی استفاده شود، عملکرد بسیار امیدوارکننده‌ای را شاهد خواهیم بود.

جدول ۳- عملکرد رمزگذار خودکار کانولوشن. آموزش و ارزیابی بر روی تصاویر ground truth

	KL متقارن	KL	CC
کانولوشن AE بدون ماسک Loss	0.0667	0.0749	0.4172
کانولوشن AE با ماسک Loss	0.0666	0.0748	0.5359
U-net AE	0.0668	0.0803	0.6477

جدول ۴- عملکرد U-Net AE (آموزش داده شده و ارزیابی شده بر روی فایل‌های gt باینری) در زیرمجموعه آزمایشی یک مجموعه آموزشی [۱۱]

پیش‌بینی وزن‌ها	0.3620	0.7556
پیش‌بینی برجستگی‌ها	0.1641	0.7922
Adversarial Loss	0.1603	0.7921

۴- نتیجه‌گیری

در این مقاله، خط لوله‌ای برای تشخیص برجستگی‌های تصویر ۳۶۰ درجه پیشنهاد شد که می‌تواند برای به پایان رساندن هر مدل تشخیص برجستگی که در تصاویر معمولی دویعدی از پیش تنظیم شده‌بود، استفاده شود. یک رمزگذار خودکار کانولوشن با یک ماسک جدید برای از بین بردن نقشه‌های برجسته دوخته‌شده و متصل شده در سراسر شکل و حذف آرتیفکت‌های دوخت و اتصال پیشنهاد شد. برای افزایش عملکرد و کاهش نویز، پیشنهاد شد از یک رمزگذار خودکار U-net استفاده شود و با تنظیم متفاوت آموزش داده شود. به طور کلی، مدل پیشنهادی U-Net AE با حذف نویز (آموزش دیده با تصاویر ground truth)، عملکرد بهتری از نظر ضریب همبستگی بین پیش‌بینی‌ها از خود نشان داد.

۵- مراجع

- Schoop, E., J. Smith, and B. Hartmann. Hindsight: enhancing spatial awareness by sonifying detected objects in real-time 360-degree video. in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018.
- Havangi, R. and S.H. Khatami, Simultaneous identification and tracking of objects using deep learning. Journal of Machine Vision and Image Processing, 2023. 10(3): p. 79-92
- Zhang, Z., et al. Saliency detection in 360 videos. in Proceedings of the European conference on computer vision (ECCV). 2018.
- Zhu, Y., et al., The prediction of saliency map for head and eye movements in 360 degree images. IEEE Transactions on Multimedia, 2019. 22(9): p. 2331-2344.
- Li, C., et al. Very long term field of view prediction for 360-degree video streaming. in 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). 2019. IEEE..
- Monroy, R., et al., Salnet360: Saliency maps for omni-directional images with cnn. Signal Processing: Image Communication, 2018. 69: p. 26-34.
- Jetley, S., N. Murray, and E. Vig, End-to-end saliency mapping via probability distribution prediction. 2017, Google Patents.
- Goferman, S., L. Zelnik-Manor, and A. Tal, Context-aware saliency detection. IEEE transactions on pattern analysis and machine intelligence, 2011. 34(10): p. 1915-926.
- Jiang, M., et al. Salicon: Saliency in context. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.
- Zhu, Y., G. Zhai, and X. Min, The prediction of head and eye movement for 360 degree images. Signal Processing: Image Communication, 2018. 69: p. 15-25.

11. Chao, F.-Y., et al. Salgan360: Visual saliency prediction on 360 degree images with generative adversarial networks. in 2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW). 2018. IEEE.