

بررسی روش‌های شمارش جمعیت در تصاویر

سیدمحمدحسین جعفری^{۱*}، حسین فقیه علی آبادی^۲

* hjafari303@yahoo.com

hosseinfaghieh1995@gmail.com

چکیده

در سال‌های اخیر، نیازهای فوری برای شمارش جمعیت و وسایل نقلیه، تحقیقات در مورد شمارش جمعیت و تخمین تراکم را به شدت ارتقا داده است. تخمین دقیق تعداد اشیا در یک تصویر یک کار چالش برانگیز و در عین حال معنادار است و در بسیاری از کاربردها مانند برنامه ریزی شهری و ایمنی عمومی استفاده شده است. خوشبختانه، توسعه تکنیک‌های جمعیت‌شمار را می‌توان به سایر زمینه‌های مرتبط مانند شمارش وسایل نقلیه و بررسی محیطی، بدون در نظر گرفتن ویژگی‌های آن‌ها، تعمیم داد. با بهره‌مندی از توسعه سریع یادگیری عمیق، عملکرد شمارش تا حد زیادی بهبود یافته است، و سناریوهای کاربردی بیشتر گسترش یافته‌اند. ما در این مقاله رویکردهای موجود را در چهار دسته مبتنی بر تشخیص، مبتنی بر رگرسیون، مبتنی بر شبکه عصبی کانولوشن و مبتنی بر ویدئو خلاصه می‌کنیم. ما بیشتر در مورد مجموعه داده‌ها و معیارها برای جامعه شمارش جمعیت توضیح می‌دهیم و در مورد کار حل مشکل شمارش بر اساس نمونه‌های کوچک، روش‌های حاشیه نویسی مجموعه داده‌ها و غیره بحث می‌کنیم. در نهایت، چالش‌های مختلف پیش روی جمعیت‌شمار و راه‌حل‌های مربوط به آن‌ها را خلاصه می‌کنیم و مجموعه‌ای از روندهای توسعه را در آینده پیشنهاد می‌کنیم.

واژه‌های کلیدی: شمارش جمعیت- تصاویر- یادگیری عمیق- شبکه عصبی کانولوشن

۱- مقدمه

شمارش جمعیت و تخمین تراکم چندین سال است که بحث چالش برانگیزی در تجزیه و تحلیل تصویر و ویدئو بوده است. شمارش دقیق جمعیت برای تجزیه و تحلیل عابر پیاده و تخمین تراکم جمعیت مفید است و دارای طیف گسترده‌ای از کاربردها مانند ایمنی عمومی، حمل و نقل هوشمند و نظارت تصویری است. جمعیت‌شماري هنوز با چالش‌های زیادی مانند انسداد شدید، تغییر صحنه، نویز پیچیده، مقیاس‌های مختلف، دیدگاه‌های مختلف و توزیع غیریکنواخت افراد مواجه است.

۱- کارشناسی ارشد مهندسی نرم افزار

۲- کارشناسی ارشد شبکه‌های کامپیوتری

شمارش اولیه جمعیت و رویکردهای برآورد تراکم عمدتاً بر اساس تشخیص عابر پیاده است [۱]. در صحنه‌های شلوغ، به دلیل عواملی مانند انسداد و مقیاس‌های مختلف، عملکرد این روش‌ها برای دستیابی به نتایج رضایت‌بخش با مشکل مواجه می‌شود و استفاده از آن در کاربردهای عملی را دشوار می‌سازد.

یک روش جایگزین، تخمین تراکم جمعیت در تصویر و در نهایت ارائه یک سطح تراکم جمعیت است، ولی کیفیت این روش طبقه‌بندی تراکم جمعیت نسبتاً بد است که کاربرد آن را در بسیاری از سناریوها محدود می‌کند [۲]. در سال‌های اخیر، با توسعه سریع یادگیری عمیق، عملکرد جمعیت شماری پیشرفت زیادی داشته است و دقت و سرعت شمارش در شرایط شلوغ به طور قابل توجهی بهبود یافته است [۳]. این مقاله به منظور مرتب سازی روش‌های تحقیق و سیر تحول رویکردهای جمعیت شماری، ایده‌ها و روش‌های اصلی جمعیت شماری را مرور می‌کند. همچنین ما یک بررسی دقیق از آخرین روش‌های مبتنی بر یادگیری عمیق انجام می‌دهیم. در این مقاله توسعه شمارش جمعیت و تخمین تراکم را به چهار شاخه روش‌های مبتنی بر تشخیص، روش‌های مبتنی بر رگرسیون، روش‌های مبتنی بر CNN و روش‌های مبتنی بر ویدئو تقسیم می‌کنیم. روش‌های مبتنی بر تشخیص، تعداد اشیاء را از طریق یک آشکارساز شی که بر روی ویژگی‌های تصویر استخراج شده آموزش دیده است، شمارش می‌کنند. روش‌ها در سناریوهای کم تراکم به خوبی کار می‌کنند، اما با افزایش تراکم جمعیت، عملکرد چنین روش‌هایی متناسب با آن کاهش می‌یابد. در مورد ازدحام متراکم، نویسندگان در گذشته به این نتیجه رسیدند که یادگیری نگاشت بین ویژگی‌های تصویر به تعداد افراد مفید است و عملکرد روش‌های مبتنی بر این نگاشت‌ها از روش‌های مبتنی بر تشخیص بهتر عمل می‌کند [۴]. این روش‌ها معمولاً یک مدل رگرسیون را از نگاشت‌های آموخته شده آموزش می‌دهند و به عنوان روش‌های مبتنی بر رگرسیون نامیده می‌شوند. با این حال، روش‌ها به شدت به ویژگی‌های دست ساز و عدم استحکام در سناریوها با تغییرات بزرگ در نور، پرسپکتیو، توزیع جمعیت، تراکم جمعیت و غیره متکی هستند. با توجه به قابلیت‌های قدرتمند استخراج ویژگی‌های CNN در یادگیری عمیق، محققان سعی کردند از الگوریتم برای استخراج خودکار ویژگی‌ها استفاده کنند. روش‌های یادگیری عمیق می‌توانند با تغییرات در عوامل مختلف سازگار شوند، تعداد افراد را با دقت بیشتری پیش‌بینی کنند و در بسیاری از معیارهای ارزیابی به وضعیت خوبی دست یافته‌اند [۵]. به منظور ارائه مجموعه داده‌ها و معیارهایی که تا حد امکان به صحنه واقعی نزدیک باشد، محققان مجموعه داده‌های بسیاری را از شمارش جمعیت ساخته‌اند. این مجموعه داده‌ها به میزان زیادی توسعه جمعیت شماری را ترویج کردند. روش‌های زیادی در حال مطالعه برای استفاده از داده‌های دارای برچسب کمتر برای شمارش دقیق تعداد افراد هستند، که از جمله روش‌های معمولی می‌توان به [6] L2R، [7] SL2R، اشاره کرد، که در ادامه بیشتر توضیح خواهیم داد. پس از مقدمه‌ای که در این بخش بیام کردیم، بخش دوم به تحلیل آثار اصلی چهار شاخه در شمارش جمعیت می‌پردازیم. در بخش سوم برخی از مجموعه داده‌های محبوب و برخی از معیارهای ارزیابی را بیان می‌کنیم. در بخش چهارم چالش‌ها و کاربردها را معرفی می‌کنیم و در بخش پنجم به نتیجه بحث، کاربرد و کار آینده شمارش جمعیت را مورد بحث قرار می‌دهیم.

۲- روش‌های انجام شده

روش‌های شمارش جمعیت را می‌توان به چهار دسته روش‌های مبتنی بر تشخیص، مبتنی بر رگرسیون، مبتنی بر CNN و مبتنی بر ویدئو تقسیم کرد. روش‌های مبتنی بر رگرسیون را می‌توان به روش‌های مبتنی بر نقشه فردی و مبتنی بر نقشه چگالی نیز تقسیم کرد.

۲-۱ روش های مبتنی بر تشخیص

اکثر کارهای اولیه جمعیت شماری بر اساس تشخیص است. در این روش‌ها از ویژگی‌های استخراج شده شمارش اهداف استفاده می‌کنند. روش‌های مبتنی بر تشخیص به شدت به ویژگی‌های اهداف متکی هستند. روش‌های استخراج ویژگی را می‌توان به دو دسته مبتنی بر انتگرال و مبتنی بر قطعات تقسیم کرد. روش‌های تشخیص مبتنی بر انتگرال ابتدا ویژگی‌های کل تصویر را مانند لبه‌ها، شکل‌ها، بافت‌ها، استخراج می‌کنند، سپس از ماشین بردار پشتیبان [۸]، جنگل تصادفی [۹]، خوشه بندی [۱۰] یا الگوریتم‌های دیگر برای تشخیص یا طبقه بندی اشیاء برای شمارش جمعیت استفاده می‌کنند. اکثر این روش‌ها زمانی که اجسام کم هستند عملکرد خوبی دارند، اما در مواجهه با جمعیت تراکم، اثر شمارش به طور قابل توجهی کاهش می‌یابد. بنابراین، پژوهشگران شروع به کشف روش‌های شمارش مؤثر در سناریوهای تراکم جمعیت کردند. در اکثر سناریوهای تراکم جمعیت، استفاده از ویژگی‌های محلی می‌تواند عملکرد شمارش را در مقایسه با ویژگی‌های جهانی تا حد زیادی بهبود بخشد. بسیاری از آثار [۱۱] بر اساس ویژگی‌های محلی کار می‌کنند.

۲-۲ روش‌های تشخیص مبتنی بر رگرسیون

روش‌های تشخیص مبتنی بر رگرسیون اغلب عملکرد بالاتری دارند و توجهات رو به رشدی را در شمارش جمعیت جلب می‌کنند. با توجه به اهداف رگرسیون مختلف، روش‌ها را می‌توان به رگرسیون مبتنی بر فردی و رگرسیون مبتنی بر نقشه چگالی تقسیم کرد که روش‌های مبتنی بر فردی عملکرد شمارش را بیشتر بهبود بخشیدند [۱۲]. به عنوان مثال، Zhou D و همکاران [۱۳] ابتدا پیش زمینه تصویر را عادی کردند و سپس از ویژگی‌های پیش زمینه، لبه و بافت محلی استخراج شده با استفاده از رگرسیون چندگانه برای بدست آوردن تعداد افراد در تصویر استفاده کردند. در مقایسه با مدل‌های رگرسیون قبلی، این روش با یادگیری یک ویژگی کم‌بعدی و یک تابع خروجی چند ساختاری، استحکام مدل را افزایش داده و آن را برای سناریوهای عملی‌تر کاربردی‌تر کرده است. با تحقیقات بیشتر، مفهوم نقشه چگالی ارائه شده توسط Sindagi [۱۴] توجه گسترده محققان را به خود جلب کرده است. با یادگیری نگاشت تصاویر به نقشه‌های چگالی، از وابستگی به آشکارساز جلوگیری می‌کند. رودریگز و همکاران [۱۵] تایید کرد که شمارش با استفاده از نقشه چگالی می‌تواند عملکرد شمارش را به شدت بهبود بخشد. با توجه به اینکه نقشه چگالی نه تنها اطلاعات توزیع فضایی جمعیت را منعکس می‌کند، بلکه دقت شمارش را نیز افزایش می‌دهد، رگرسیون مبتنی بر نقشه چگالی به تدریج به یک دسته بندی محبوب تبدیل می‌شود.

مشابه روش‌های مبتنی بر تشخیص، روش‌های رگرسیون را می‌توان به دو دسته مبتنی بر انتگرال و مبتنی بر patch [۱۶] تقسیم کرد. روش‌های رگرسیون مبتنی بر انتگرال همیشه در برخورد با تغییرات مقیاس بزرگ و چگالی مشکلاتی دارند در حالی که روش‌های رگرسیون مبتنی بر patch حاوی اطلاعات محلی بیشتری از تصویر هستند و کمتر تحت تأثیر تغییرات مقیاس و چگالی قرار می‌گیرند. بنابراین، عملکرد روش‌های رگرسیون مبتنی بر patch اغلب بهتر از روش‌های مبتنی بر انتگرال است. فام و همکاران [۱۷] یک تصویر را به چند تکه تقسیم کردند و از جنگل تصادفی برای طبقه‌بندی ویژگی‌ها استفاده کردند، به طوری که گره‌های برگ هر درخت فقط دارای ویژگی‌های مشابه بودند. اگرچه روش‌های مبتنی بر رگرسیون وابستگی به آشکارساز را کاهش می‌دهند، اما همچنان به شدت به ویژگی‌های دست ساز متکی هستند. در نتیجه، الگوریتم استخراج ویژگی به یک گلوگاه مهم برای روش‌های مبتنی بر رگرسیون تبدیل شد. با توسعه سریع یادگیری عمیق، قابلیت‌های

قدرتمند استخراج ویژگی شبکه‌های عصبی کانولوشنال (CNN) محققان را مجذوب خود کرده است و روش‌های جمعیت‌شمار مبتنی بر CNN و روش‌های تخمین تراکم جمعیت به سرعت در حال توسعه هستند.

۲-۲ روش‌های مبتنی بر CNN

در سال‌های اخیر، یادگیری عمیق به طور فزاینده‌ای توجه محققان را به خود جلب کرده است. CNN قابلیت‌های یادگیری قوی در پردازش تصویر نشان داده‌اند که الهام‌بخش بسیاری از کارهای شمارش جمعیت مبتنی بر CNN هستند. مین و همکاران [۱۸] اولین رویکردی را از CNN برای شمارش جمعیت استفاده کردند. با این حال، فقط سطح تراکم جمعیت را تخمین زد. از آن زمان، کار شمارش بر اساس CNN به سرعت پیشرفت کرده است. روش‌های مبتنی بر CNN عملکرد بهتری در سناریوهایی مانند گستره بزرگ مقیاس سرانسان، توزیع چگالی غیریکنواخت و تغییرات بزرگ در پرسپکتیو و صحنه دارند، که باعث می‌شود رویکردهای مبتنی بر CNN بر تحقیقات کنونی شمارش جمعیت تسلط داشته باشند. برای چالش‌های جمعیت‌شمار کنونی، محققان روش‌های مختلفی را برای مقابله با آن‌ها اتخاذ کرده‌اند. به منظور توانمندسازی محققان برای درک جامع مشکلات فعلی و استراتژی‌های پردازش متناظر آن‌ها، ما آثار موجود را به چند دسته زیر تقسیم می‌کنیم که هر کدام نشان‌دهنده یک استراتژی شمارش جریان اصلی است. در مقایسه با آثار فوق، ما بر مشکلات فعلی و استراتژی‌های پردازش متناظر آن‌ها تأکید بیشتری داریم.

ترکیب چند چگالی با در نظر گرفتن مشکل شرایط مختلف تصویر ورودی، یکی از راه‌حل‌ها ترکیب نقشه‌های چگالی در مقیاس‌های چندگانه برای شمارش جمعیت است. به منظور غلبه بر مشکلات تغییرات پرسپکتیو دوربین و انسداد موانع در شمارش جمعیت، Zhang [۱۹] مدل شبکه‌ای را پیشنهاد کرد که به طور تطبیقی پاسخ مسیر را با کشف می‌کند. چارچوب اول از یک شبکه عصبی کانولوشن برای استخراج ویژگی‌ها استفاده می‌کند و سپس از سه شاخه برای تولید یک نقشه چگالی استفاده می‌کند. شن و همکاران [۲۰] چارچوبی را پیشنهاد کرد که تشخیص و رگرسیون را ترکیب می‌کند که شامل سه بخش شبکه رگرسیون RegNet، شبکه تشخیص DetNet و QualityNet است. دو قسمت اول به ترتیب دو نقشه چگالی تولید کردند و دومی نقشه‌های چگالی تولید شده را وزن کرده و فیوز می‌کند تا یک نقشه نهایی را تشکیل دهد. مدل او نسبت به اندازه‌های مختلف جمعیت قوی‌تر است و برای طیف وسیع تری از سناریوها مناسب‌تر است. لیو و همکاران [۲۱] یک شبکه آگاه از دیدگاه برای شمارش پیشنهاد کردند که اطلاعات پرسپکتیو را به عنوان اطلاعات کمکی برای تغییرات مقیاس جمعیت گرفته و نقشه‌های چگالی چند سطحی را وزن کرده و آن‌ها را ترکیب کرده است. Shi M [۲۲] کل تصویر را به عنوان ورودی گرفت، از سه شاخه برای پیش‌بینی تعداد افراد در تصویر استفاده کرد و سپس از یک شاخه برای وزن کردن نتایج پیش‌بینی قبلی و ترکیب کردن همه آن‌ها استفاده کرد است.

مبتنی بر GAN، شبکه‌های متخاصم مولد (GAN) یک مدل یادگیری عمیق است که در سال‌های اخیر، یکی از امیدوارکننده‌ترین روش‌ها برای یادگیری بدون نظارت توزیع‌های پیچیده بوده است. GAN شامل دو ماژول، یعنی یک مدل تولیدی و یک مدل متمایز است که برای درک توزیع داده‌های واقعی تا حد امکان با یکدیگر رقابت می‌کنند. در کارهای مرتبط در مورد شمارش جمعیت، برخی از محققان از یک ژنراتور برای به دست آوردن نقشه چگالی استفاده کردند و سپس از یک

تمایز برای تشخیص نقشه چگالی از حقیقت زمینی استفاده کردند. این رقابت با یکدیگر در نهایت نقشه های چگالی حاصل را دقیق تر می کند [۲۳]. شبیه به GAN، مدل MS-GAN پیشنهاد شده توسط یانگ و همکاران [۲۴] شامل یک مولد و یک ممیز است. ژنراتور یک شبکه کانولوشن کامل چند مقیاسی است که ویژگی های لایه های کانولوشنی مختلف را برای ایجاد یک نقشه چگالی ترکیب می کند، نقشه چگالی تولید شده توسط ژنراتور به عنوان نمونه های منفی استفاده می شود و در ترکیب با حقیقت زمین توسط تفکیک کننده آموزش داده می شود. به این ترتیب عملکرد ژنراتور در شبکه متخاصم به طور مکرر بهبود می یابد و نقشه چگالی بهتری به دست می آید. وانگ و همکاران [۲۵] یک روش یادگیری تطبیقی بدون نظارت را پیشنهاد کردند که برای بهبود عملکرد مدل در یک صحنه نادیده طراحی شده است. نویسندگان از تکه های هر می چند مقیاسی در حوزه های منبع و هدف برای آموزش خصوصت آمیز برای مدیریت مقیاس های مختلف جمعیت و توزیع تراکم استفاده کردند. شن و همکاران [۲۶] یک چارچوب شمارش جمعیت بر اساس مدل شبکه مولد متخاصم طراحی کرد: یک شبکه U شکل به عنوان شبکه مولد مدل استفاده شد و یک نقشه چگالی با وضوح بالا با استفاده از یک تمایز غربال شد. اولمچنک و همکاران [۲۷] به مطالعه نحوه آموزش یک مدل شبکه شمارش جمعیت با استفاده از مقدار کمی داده اختصاص داشت.

مبنتی بر زمینه، برخی از کارهای شمارش جمعیت از معناسناسی متنی تصاویر برای هدایت روند شمارش استفاده می کنند. این روش عمدتاً از زمینه و اطلاعات معنایی صحنه جمعیت برای محدود کردن نقشه چگالی برای دستیابی به عملکرد بهتر استفاده می کند. CP-CNN [۲۸] یک شبکه هر می زمینه را برای استفاده کامل از اطلاعات زمینه ای برای تولید نقشه های چگالی با دقت بالا پیشنهاد کرد. شبکه از چهار بخش GCE (برآورنده زمینه جهانی)، LCE (برآورنده زمینه محلی)، DME (برآورد نقشه چگالی) و F-CNN تشکیل شده است. GCE اطلاعات جهانی را رمزگذاری می کند و ویژگی های معنایی سطح بالا را استخراج می کند، در حالی که کل تصویر ورودی را در سطوح مختلف چگالی طبقه بندی می کند. LCE اطلاعات محلی را رمزگذاری می کند و ویژگی های محلی را استخراج می کند، در حالی که هر path را به سطوح تراکم مختلف طبقه بندی می کند. DME برای تولید مستقیم نقشه چگالی استفاده می شود. در نهایت، خروجی های سه بخش توسط F-CNN برای به دست آوردن یک نقشه چگالی با کیفیت بالا ترکیب می شوند. با توجه به اینکه استفاده از تلفات اقلیدسی تنها باعث تار شدن نقشه چگالی می شود، ترکیب وزنی از تلفات اقلیدسی در سطح پیکسل و تلفات متخاصم به عنوان تابع ضرر استفاده می شود. شانگ و همکاران [۲۹] به طور مستقیم تعداد جمعیت را بر اساس کل تصویر شمارش نکردند، اما تعداد نهایی افراد را با استفاده از محاسبات مشترک بر روی مناطق همپوشانی محاسبه کرد. لیو و همکاران [۳۰] ویژگی های چند اندازه فلد گیرنده و هر مکان تصویر را ترکیب کردند و سپس آن ها را با استفاده از یک شبکه آموزش پذیر سرتاسر آموزش دادند. در نهایت، شبکه یک نقشه چگالی با کیفیت بالا تولید می کنند.

coarse-to-fine، بسیاری از کارهای coarse-to-fine ابتدا یک نقشه چگالی درشت به دست می آورند و سپس آن را برای به دست آوردن نقشه چگالی دانه بندی شده نهایی بهینه یا تنظیم می کنند. به منظور حل مشکلات چرخش ها، مقیاس ها و پرسپکتیوهای مختلف ناشی از تغییر نماهای دوربین ها، لیو و همکاران یک شبکه آگاه از فضای عمیق را پیشنهاد کردند. این شبکه از مدل Global Feature Embedding بر اساس VGG-16 به عنوان قسمت جلویی برای تولید نقشه چگالی اولیه استفاده می کند و سپس از مدل Recurrent SpatialAware Refinement برای بهینه سازی نقشه چگالی تولید شده استفاده می کند [۳۱]. کارهای قبلی مانند S-DCNet [32] نقشه های ویژگی لایه های پیچیدگی مختلف را ادغام کردند و اطلاعات چند

مقیاسی را از طریق ساختار شبکه ای هرم ویژگی به دست آوردند. در مقابل، [۳۳] تنها بهبود اطلاعات چند مقیاسی را بر روی یک نقشه ویژگی تک لایه پیاده سازی می کند و این عملیات را روی لایه های پیچیدگی مختلف تکرار می کند تا اطلاعات غنی را به ماژول رگرسیون بعدی بیاورد. [34] ic-CNN یک مدل جمعیت شماری دو مرحله ای را پیشنهاد کرد، شاخه LR یک نقشه چگالی با وضوح پایین ایجاد می کند، و شاخه HR نقشه ویژگی و پیش بینی وضوح پایین را برای تولید یک نقشه چگالی با وضوح بالا ترکیب می کند. این مدل همچنین می تواند به یک مدل چند مرحله ای گسترش یابد، یعنی از ترکیب تکراری برای بهبود عملکرد مدل استفاده می شود. بدین ترتیب یک نقشه چگالی با کیفیت بالا به دست می آید.

۴-۲ شمارش جمعیت مبتنی بر ویدئو

با توجه به اینکه توالی ویدئو حاوی اطلاعات زمانی است که برای شمارش مفید است، برخی از محققان در حال حاضر روی شمارش تعداد افراد در ویدئو کار می کنند. برخی از این وظایف معمولی را به صورت مختصر شرح خواهیم داد.

ConvLSTM اکثر روش های شمارش مبتنی بر ویدئو فقط تک فریم ویدئو را جداگانه در نظر می گیرند و همبستگی زمانی بین فریم های ویدئو را نادیده می گیرند، که برای زمینه شمارش آموزنده و مفید است. [35] ConvLSTM به طور موثر از همبستگی زمانی برای کمک به کار شمارش استفاده کرد. مدل ConvLSTM توسعه FC-LSTM است. این مدل ساختار کاملاً متصل با یک ساختار پیچشی برای انجام تبدیل ویژگی جایگزین می شود و یک تانسور سه بعدی ترکیب شده با اطلاعات مکانی زمانی به عنوان نمایش اطلاعات برای انتقال اطلاعات و دروازه کنترل استفاده می شود. برخلاف روش های مبتنی بر CNN که فقط اطلاعات مکانی را در نظر می گیرند، مدل ConvLSTM توجه بیشتری را به همبستگی زمانی بین فریم های مجاور ویدئو معطوف می کند، بنابراین می تواند به طور موثر از اطلاعات حوزه زمانی استفاده کند. این روش ارتباط بین فضا و زمان را در ویدئو به میزان کافی نشان می دهد و در نتیجه دقت شمارش در صحنه های پیچیده را بهبود می بخشد.

LSTM برخلاف روش مدل سازی ضمنی مبتنی بر LSTM [36]، LSTM از یک ماژول ترانسفورماتور فضایی با محدودیت محلی برای ثبت وابستگی های مکانی زمانی در ویدئو استفاده کرد. این مدل عمدهً از دو ماژول ترانسفورماتور رگرسیون نقشه چگالی و ماژول ترانسفورماتور فضایی مبتنی بر محدودیت موقعیت (LST) تشکیل شده است. ماژول رگرسیون نقشه چگالی مستقیماً نقشه چگالی یک فریم را تخمین میزند و سپس از ماژول LST برای مرتبط کردن نقشه چگالی فریم های مجاور برای خروجی نقشه چگالی دقیق تر استفاده می کند. وو و همکاران [۳۷] از تصاویر انبوه و نقشه های چگالی پیش بینی کردند تا اطلاعات زمانی را به صراحت مدل سازی کنند. آن ها از مجموعه ای از بلوک های باقیمانده گشاد شده برای مدل سازی رابطه بین ویژگی های فریم های مجاور ویدئو استفاده کردند. در هر مرحله، مجموعه گسترده ای از پیش ها در طول زمان برای تولید نقشه چگالی اولیه استفاده می شود که برای بهینه سازی نقشه چگالی بعدی به طور مکرر در مرحله بعد استفاده می شود.

E3D با توجه به عملکرد برتر کانولوشن سه بعدی در تشخیص حرکت، [38] Zou et al. سعی کرد از کانولوشن سه بعدی برای رمزگذاری ویژگی های مکانی زمانی در ویدئو استفاده کند. این اطلاعات زمینه جهانی را در وزن های مدولاسیون رمزگذاری می کند، در حالی که پاسخ مشخصه هر کانال را تغییر می دهد، ویژگی های مفید را به طور تطبیقی برجسته می کند و از اتصالات کوتاه پرش برای ساده سازی آموزش مدل استفاده می کند. معماری جدید موقت کانال آگاه (TCA) ساخته شده در

این مقاله نه تنها می‌تواند وابستگی زمانی دنباله‌های ویدئویی را به طور موثر ضبط کند، بلکه اطلاعات مکانی و زمانی محلی و جهانی را نیز ترکیب می‌کند.

شمارش عابر پیاده روی خط متقاطع، ژنگ و همکاران [۳۹] یک روش جمعیت‌شمار مقیاس‌پذیر را پیشنهاد کردند، که برای شمارش عابران پیاده در حال عبور از خطوط مجازی زمانی که جمعیت بسیار پویا و متراکم است، طراحی شده است. این روش شامل دو بخش برآورد تراکم جمعیت محلی و شمارش عابر پیاده از خط متقاطع است. به منظور برآورد دقیق تراکم جمعیت محلی، آن‌ها محله را در خط مجازی به چندین بلوک تقسیم کردند و سازگاری فضایی بین شمارش محلی و تعداد منطقه بسته را افزایش دادند تا از ثبات برآورد تراکم جمعیت محلی اطمینان حاصل کنند.

تقسیم منطقه پویا، با توجه به اینکه روش شمارش عابر پیاده دو ناحیه مستقیم ممکن است یک سر را به دو قسمت تقسیم کند و در نتیجه خطاهای شمارش را معرفی کند، او و همکاران [۴۰] یک الگوریتم پارتیشن‌بندی منطقه پویا را برای اطمینان از یکپارچگی شمارش پیشنهاد کردند. با فرض حفظ یکپارچگی سر، آنها از جعبه مرزی و خطوط تقسیم‌بندی صحنه اشیاء به دست آمده توسط YOLOV3 [41] برای تقسیم‌بندی نواحی دیستال و پروگزیمال استفاده کردند.

۳- مجموعه داده‌ها و معیارها

مجموعه داده‌ها برای آموزش و ارزیابی مدل‌های شمارش جمعیت اهمیت زیادی دارند. برای آموزش مدل‌های تعمیم‌یافته‌تر، محققان مجموعه‌های داده‌ای متنوعی ساخته‌اند. مجموعه داده‌های اولیه عمدتاً تصاویر یا فریم‌های ویدیویی با تراکم جمعیت کم و سناریوهای مشابه هستند. اکثر مجموعه داده‌های بعدی اغلب دارای اندازه نمونه بزرگ و برچسب‌های دقیق تری هستند. از نظر اندازه نمونه، این مجموعه داده‌ها عمدتاً طیف گسترده‌ای از عوامل را پوشش می‌دهند: سناریوهای متنوع، تراکم‌های مختلف جمعیت، گستره مقیاس بزرگ سر و غیره، که بسیار به توزیع داده‌ها در برنامه‌های واقعی نزدیک‌تر هستند. از منظر دقت برچسب‌گذاری، نقشه‌های چگالی تولید شده توسط این مجموعه داده‌ها دقیق‌تر و معقول‌تر هستند. ما این مجموعه داده‌ها را شرح داده و تجزیه و تحلیل خواهیم کرد و مقایسه عملکرد روش‌های محبوب شمارش جمعیت را در مجموعه داده‌ها انجام خواهیم داد.

۳-۱ معرفی مجموعه دادگان

ShanghaiTech Campus dataset (Anomaly Detection)، مدل تشخیص ناهنجاری آموزش دیده می‌تواند به طور مستقیم در صحنه‌های متعدد با زاویه دید چندگانه اعمال شود. با این حال، تقریباً تمام مجموعه داده‌های موجود فقط حاوی ویدیوهایی هستند که با یک دوربین زاویه ثابت گرفته شده‌اند و فاقد تنوع صحنه‌ها و زوایای دید هستند. برای افزایش تنوع صحنه، یک مجموعه داده تشخیص ناهنجاری جدید ایجاد می‌شود. علاوه بر این، ناهنجاری‌های ناشی از حرکت ناگهانی در این مجموعه داده را معرفی می‌شود. این ویژگی‌ها مجموعه داده را در سناریوهای واقعی مناسب‌تر می‌کند. برای درک بهتر تفاوت‌های بین

این مجموعه داده و مجموعه داده‌های تشخیص ناهنجاری موجود، به طور خلاصه همه مجموعه داده‌های تشخیص ناهنجاری را به شرح زیر خلاصه می‌کنیم:

مجموعه داده CUHK Avenue شامل ۱۶ فیلم آموزشی و ۲۱ ویدیوی آزمایشی با مجموع ۴۷ رویداد غیرعادی، از جمله پرتاب اشیا، پرسه زدن و دویدن است. اندازه افراد ممکن است به دلیل موقعیت و زاویه دوربین تغییر کند.

مجموعه داده عابر پیاده (Ped1) شامل ۳۴ فیلم آموزشی و ۳۶ ویدیوی آزمایشی با ۴۰ رویداد نامنظم است. همه این موارد غیرعادی مربوط به وسایل نقلیه ای مانند دوچرخه و ماشین است.

مجموعه داده عابر پیاده (Ped2) شامل ۱۶ فیلم آموزشی و ۱۲ ویدیوی آزمایشی با ۱۲ رویداد غیرعادی است. تعریف ناهنجاری برای Ped2 با Ped1 یکسان است.

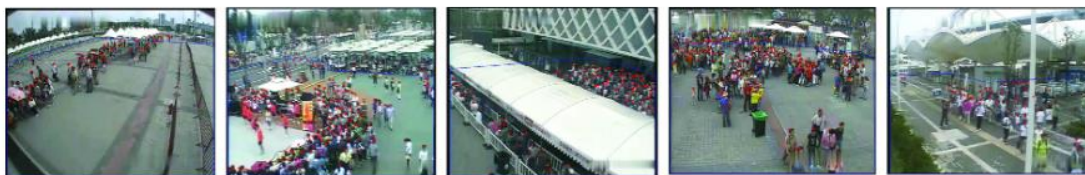
رویدادهای غیرمعمول شامل راه رفتن در مسیرهای اشتباه و پرسه زدن است. مهم‌تر از آن، این مجموعه داده در محیط داخلی ثبت می‌شود در حالی که موارد فوق در محیط بیرون ثبت می‌شود. مجموعه داده شانگهای دارای ۱۳ صحنه با شرایط نوری پیچیده و زوایای دوربین است. این شامل ۱۳۰ رویداد غیر عادی و بیش از ۲۷۰۰۰۰ فریم آموزشی است. علاوه بر این، حقیقت سطح پیکسل رویدادهای غیرعادی نیز در مجموعه داده مشروح شده است [۴۲].

مجموعه داده worldexpo'10، یک مجموعه داده شمارش جمعیت در مقیاس بزرگ را معرفی می‌کنیم. تا جایی که ما می‌دانیم، این بزرگترین مجموعه داده است که بر شمارش صحنه‌های متقاطع تمرکز دارد. این شامل ۱۱۳۲ دنباله ویدیویی مشروح شده است که توسط ۱۰۸ دوربین نظارتی از نمایشگاه جهانی شانگهای ۲۰۱۰ گرفته شده است. از آنجایی که بیشتر دوربین‌ها دارای نماهای پرنده‌ای متفاوت هستند، طیف وسیعی از صحنه‌ها را پوشش می‌دهند. مجموعه داده ما به دو بخش تقسیم می‌شود. ۱۱۲۷ سکانس ویدیویی طولانی یک دقیقه ای از ۱۰۳ صحنه به عنوان مجموعه آموزشی و اعتبار سنجی در نظر گرفته می‌شود. در هر فیلم آموزشی ۳ فریم برچسب دار وجود دارد و فاصله بین دو فریم برچسب دار ۱۵ ثانیه است. برخی از نمونه‌ها به صورت زیر در شکل ۱ نشان داده شده است.



شکل ۱: نمونه‌ای از مجموعه داده worldexpo'10 dataset

مجموعه آزمایشی دارای ۵ سکانس ویدیویی یک ساعته از ۵ صحنه مختلف است. در هر صحنه تست ۱۲۰ فریم برچسب دار وجود دارد و فاصله بین دو فریم برچسب دار ۳۰ ثانیه است. برخی از نمونه‌ها در شکل ۲ نشان داده شده است:



شکل ۲: نمونه‌ای از مجموعه داده worldexpo'10 dataset

مجموعه آموزشی شامل هیچ صحنه‌ای در مجموعه تست نمی‌شود. ناحیه درون چند ضلعی‌های آبی، مناطق مورد نظر (ROI) هستند و موقعیت سرهای عابر پیاده با نقاط قرمز برچسب گذاری شده است [۴۳].

مجموعه داده SmartCity، در مجموع ۵۰ تصویر از ده صحنه شهر شامل ورودی اداری، پیاده رو، دهلیز، مرکز خرید و غیره جمع آوری شده است. برخی از نمونه‌ها در شکل ۳ نشان داده شده است.



شکل ۳: مجموعه داده SmartCity

همه آن‌ها با زاویه بالا برای نظارت تصویری گرفته شده‌اند. مجموعه داده‌های شمارش جمعیت موجود شامل تصاویر صدها یا حتی هزاران عابر پیاده است و تقریباً همه تصاویر در فضای باز گرفته شده‌اند. بنابراین ما به طور خاص مجموعه داده‌ای را ایجاد می‌کنیم که عابران پیاده کمی در تصاویر داشته باشد و از صحنه‌های بیرونی و داخلی تشکیل شده است. میانگین تعداد عابران پیاده در هر تصویر تنها ۷,۴ است که حداقل ۱ و حداکثر ۱۴ است. از این مجموعه داده برای آزمایش توانایی تعمیم چارچوب پیشنهادی در صحنه‌های بسیار کم جمعیت استفاده می‌شود [۴۴].

مجموعه داده UCF-QNRF، بزرگترین مجموعه داده تا به امروز (از نظر تعداد حاشیه نویسی) را برای آموزش و ارزیابی روش‌های شمارش جمعیت و بومی سازی معرفی می‌کنیم. این شامل ۱۵۳۵ تصویر است که به ترتیب به مجموعه های آزمایش و تست ۱۲۰۱ و ۳۳۴ تصویر تقسیم می‌شوند. مجموعه داده برای آموزش شبکه‌های عصبی کانولوشنال بسیار عمیق (CNN) مناسب‌تر است، زیرا در مقایسه با سایر مجموعه داده‌های شمارش ازدحام موجود، انسان‌های حاشیه‌نویسی بیشتری در صحنه‌های انبوه جمعیت دارد. در حالی که شکل ۴ شش تصویر را نشان می‌دهد که به طور تصادفی از مجموعه داده انتخاب شده‌اند.



شکل ۴: مجموعه داده UCF-QNRF

مجموعه داده‌های UCF-QNRF دارای بیشترین تعداد تصاویر و حاشیه‌نویسی‌های پرتعداد، و طیف گسترده‌تری از صحنه‌ها شامل متنوع‌ترین مجموعه‌ای از دیدگاه‌ها، تراکم و تغییرات نور است. وضوح در مقایسه با WorldExpo10 و ShanghaiTech بزرگ است. تراکم متوسط، یعنی تعداد افراد در هر پیکسل در تمام تصاویر نیز کمترین میزان است، که نشان دهنده تصاویر بزرگ با کیفیت بالا است. تراکم کمتر در هر پیکسل تا حدی به دلیل گنجاندن مناطق پس زمینه است، که در آن مناطق با چگالی بالا و همچنین مناطق با چگالی صفر وجود دارد. بخش A از مجموعه داده شانگهای نیز دارای تصاویر جمعیتی زیاد است، با این حال، آنها به شدت برش داده شده‌اند تا فقط ازدحام داشته باشند. از سوی دیگر، مجموعه داده‌های جدید UCF-QNRF شامل ساختمان‌ها، پوشش گیاهی، آسمان و جاده‌ها می‌شود که در سناریوهای واقعی که در طبیعت ثبت شده‌اند، حضور دارند. این امر این مجموعه داده را واقعی‌تر و همچنین دشوار می‌کند [۴۵].

مجموعه داده UCF CC 50، این مجموعه داده حاوی تصاویری از جمعیت بسیار متراکم است. تصاویر عمدتاً از جمع آوری شده‌اند. مجموعه داده UCF CC 50 شامل ۵۰ تصویر در مقیاس خاکستری با مجموع ۶۳۹۷۴ شخص حاشیه نویسی شده است. تعداد افراد از ۹۴ تا ۴۵۴۳ با میانگین ۱۲۸۰ نفر در هر تصویر متغیر است [۴۶].

مجموعه داده Venice، این مجموعه داده شامل اندازه گیری های ساعتی سطح آب در ونزی از سال ۱۹۸۳ تا ۲۰۱۵ است. مجموعه داده حاصل شامل ۴ دنباله مختلف و در مجموع ۱۶۷ فریم حاشیه نویسی با وضوح ۱۲۸۰×۷۲۰ ثابت

است. ۸۰ تصویر از یک دنباله طولانی به عنوان داده‌های آموزشی گرفته می‌شود و از تصاویر ۳ دنباله باقی مانده برای اهداف آزمایشی استفاده می‌شود [۴۷].

مجموعه داده JHU-CROWD، با ۴۳۷۲ تصویر با وضوح متوسط (1430x910) است که در شرایط مختلف و موقعیت‌های جغرافیایی مختلف جمع آوری شده است و ۱,۵۱ میلیون حاشیه نویسی نقطه با میانگین ۳۴۶ نقطه در هر تصویر و حداکثر ۲۵ هزار نقطه است. در مقایسه با مجموعه داده‌های موجود، مجموعه داده پیشنهادی تحت انواع سناریوها و شرایط محیطی متنوع جمع آوری می‌شود. علاوه بر این، مجموعه داده‌ها نسبتاً غنی‌تری از حاشیه‌نویسی‌ها مانند نقاط، جعبه‌های مرزی تقریبی، سطوح تاری و غیره را ارائه می‌کند. برجسب‌های سطح سر (نقاط، تقریباً کادر محدود، سطح تاری، و غیره) و برجسب‌های سطح تصویر (نوع صحنه و وضعیت آب و هوا) را ارائه می‌کند. تراکم‌های مختلف، تغییرات روشنایی، شرایط نامساعد جوی مانند مه، باران و برف، نکات برجسته حاشیه نویسی مجموعه ای غنی از حاشیه نویسی: نقاط، کادرهای مرزی تقریبی، سطوح تاری و غیره را ارائه می‌کند [۴۸].

۲,۳ روش‌های حاشیه نویسی مجموعه داده‌ها

مجموعه داده‌های مختلف در تولید نقشه چگالی کاملاً متفاوت هستند. روش‌های اصلی تولید نقشه چگالی به شرح زیر است.

الف. پیچیدگی نقطه‌ای سر انسان

لمپیتسکی و زیسرمن [۴۹] ابتدا مفهوم نقشه چگالی را در شمارش جمعیت معرفی کردند که این موضوع را وارد مرحله جدیدی کرد و تأثیر زیادی بر کارهای بعدی داشت. برای جلوگیری از مشکلات تشخیص و مکان‌یابی اشیا، مسئله شمارش به عنوان یک مشکل نگاشت بین ورودی و نقشه چگالی در نظر گرفته می‌شود. در نهایت، نقشه چگالی با روش رگرسیون برای به دست آوردن تعداد کل افراد استفاده می‌شود. به طور خاص، آن‌ها سر انسان را در تصویر به عنوان یک نقطه برجسب گذاری کردند و سپس پیچیدگی گاوسی دوبعدی را روی هر نقطه انجام دادند تا حقیقت زمینی مربوطه را به دست آورند.

ب. نقشه چگالی آگاه از محتوا

اوغاز و همکاران [۵۰] روش‌های تولید نقشه چگالی قبلی را به دو دسته روش گاوسی دو بعدی استاتیک و روش گاوسی دو بعدی پویا تقسیم می‌کند. در روش استاتیکی تغییرات اندازه سر انسان را در نظر نمی‌گیرد که از دقت بیشتر نقشه چگالی جلوگیری می‌کند. روش گاوسی دو بعدی پویا سعی در تکمیل این نقص دارد. با این حال، اطلاعات محتوای جمعیت در تصویر را در نظر نمی‌گیرد، که ممکن است نویز زیادی ایجاد کند و بر دقت شمارش تأثیر منفی بگذارد. نویسنده استراتژی تقسیم‌بندی Chan-Vese، هسته کانولوشن گاوسی دو بعدی و جستجوی نزدیکترین نقطه را با نیروی brute-force ترکیب کرد تا عملکرد را بهبود بخشد، و از فناوری آگاه از محتوا برای جبران کمبود دقت روش‌های قبلی استفاده کرد. آزمایش‌ها نشان می‌دهند که مدل‌هایی که از نقشه‌های چگالی تولید شده توسط این روش استفاده می‌کنند، می‌توانند به دقت بسیار بالاتری دست یابند.

ج. نقشه IKNN

اولمچنک و همکاران [۵۱] متوجه شدند که روش‌های تولید نقشه چگالی قبلی هنوز دو جنبه دارند که می‌توان آنها را بهبود بخشید. اولاً، یک مورد شدید را در نظر بگیرید، هر عابر پیاده کاملاً روی یک پیکسل در نقشه چگالی ساکن است، شبکه ای که تراکم ۱ پیکسل دور از برجسب گذاری صحیح را پیش بینی می‌کند، به همان اندازه نادرست در نظر گرفته می‌شود که ۱۰ پیکسل دور از برجسب گذاری صحیح است. بدیهی است که مطلوب نیست زیرا یک گرادیان آموزشی ناپیوسته ایجاد می‌شود. جنبه دیگر این است که برخی از محلات دارای توزیع گاوسی بسیار بزرگی هستند که همچنین منجر به اطلاعات مکانی نادرست مکان های تراکم می‌شود. آنها به طور تجربی نشان دادند که استفاده از نقشه IKNN برای آموزش مدل‌های شمارش جمعیت موجود می‌تواند عملکرد را به طور قابل توجهی بهبود بخشد.

۳-۳ معیارهای ارزیابی

ما به طور خلاصه معیارهای ارزیابی موجود مدل شمارش جمعیت را از دو جنبه کمیت اشیاء و نقشه چگالی تجزیه و تحلیل و ارزیابی می‌کنیم. در زیر پرکاربردترین معیارهای ارزیابی آورده شده است.

برای ارزیابی کمیت اشیاء، MAE (میانگین خطای مطلق) یک معیار ارزیابی رایج در مدل‌های رگرسیونی است و مجموع مقادیر مطلق تفاوت‌های بین مقادیر پیش‌بینی شده و حقیقت پایه را نشان می‌دهد که در معادله (۱) به صورت زیر تعریف می‌شود:

(۱)

$$MAE = \frac{1}{M} \sum_{I=1}^M |Y_I - \hat{Y}_I|$$

برای ارزیابی نقشه چگالی، نسبت سیگنال به نویز (PSNR) یک معیار رایج برای تشابه دو تصویر است که در معادله (۲) نمایش داده شده است. بر اساس خطای بین پیکسل‌های مربوطه است. این یک استاندارد ارزیابی عینی است که ویژگی‌های بصری چشم انسان را در نظر نمی‌گیرد. بنابراین، در برخی موارد، نتایج ارزیابی با ادراک ذهنی انسان ناسازگار است.

(۲)

$$PSNR = 10 \lg \frac{(2^n - 1)^2}{MSE}$$

برای L2R و SL2R از چارچوب یادگیری به رتبه برای مرتب کردن تصاویر بدون برجسب و تسهیل آموزش مدل شمارش استفاده کرد. آنها تصویر را به چند تکه کوچک برش دادند و سپس آنها را با توجه به این که تعداد افراد در تصویر فرعی کمتر یا مساوی تعداد افراد در تصویر اصلی است، رتبه بندی کردند. علاوه بر این، آنها دو مدل شبکه تخمین چگالی با وصله‌های ورودی چند مقیاسی و شبکه‌ای که داده‌های بدون برجسب را رتبه‌بندی می‌کند، پیشنهاد کردند. آنها سه روش آموزشی را با استفاده از داده‌های برجسب‌گذاری شده و داده‌های مرتب‌سازی شده تأیید کردند. نتایج نشان می‌دهد که عملکرد سه مدل آموزشی بهتر از مدلی است که فقط توسط داده‌های برجسب‌گذاری شده آموزش داده شده است. در میان آنها، عملکرد مدل آموزشی چند وظیفه ای معمولاً بهترین است [۵۲].

۴- کاربرد و بحث

۴-۱ کاربرد

در حال حاضر، شمارش جمعیت عمدتاً در سناریوهایی مانند امنیت جمعیت، نظارت تصویری و تجزیه و تحلیل ترافیک استفاده می شود. نظارت بر تعداد افراد در فعالیت‌های تجمع مانند رویدادهای ورزشی، تظاهرات عمومی، گردهمایی‌های سیاسی، کنسرت‌ها و غیره بخش مهمی از امنیت جمعیت است. اطلاعات مربوط به تعداد افراد می‌تواند نه تنها برای کمک به نیروهای امنیتی بلکه برای تخلیه افراد به موقع و موثرتر مورد استفاده قرار گیرد و در نتیجه احتمال بروز حوادث کاهش یابد. ثانیاً، شمارش جمعیت همچنین می‌تواند برای نظارت بر اطلاعات ترافیک استفاده شود، که نه تنها به ساخت جاده کمک می‌کند، بلکه برنامه‌های زمان‌بندی وسایل نقلیه را معقول‌تر می‌کند.

۴-۲ چالش‌ها و راه حل‌ها

تنوع مقیاس یک چالش سخت در شمارش جمعیت و تخمین تراکم است. تعداد افراد در صحنه‌های مختلف بسیار متفاوت است، زمانی که جمعیت متراکم باشد، تعداد افراد می‌تواند به هزاران نفر برسد، و در صورت پراکندگی، تنها ده‌ها نفر وجود دارند. به دلیل زاویه دوربین، اندازه سر انسان در تصویر به ناچار بسیار متفاوت است. چنین تفاوت کمیت بسیار چالش برانگیزی برای مدل است. همانطور که در بخش قبلی توضیح داده شد در حال حاضر دو راه حل اصلی ادغام ویژگی‌های چند مقیاسی و ادغام نقشه‌های چگالی چند مقیاسی وجود دارد. با ادغام ویژگی‌ها یا نقشه‌های چگالی سطوح مختلف، می‌توان تغییرات مقیاس را تا حدودی کاهش داد.

انسداد، تقریباً همه تصاویر دارای مشکل انسداد هستند و با متراکم شدن جمعیت این مشکل شدیدتر می‌شود. هنگامی که جمعیت متراکم است، انسداد شدید می‌شود که شمارش را بسیار دشوار می‌کند. اکثر آثار فعلی از توانایی استخراج ویژگی قدرتمند و توانایی یادگیری شبکه عصبی کانولوشن برای کاهش این مشکل استفاده می‌کنند.

تنوع دیدگاه‌ها، تغییر در موقعیت دوربین و زاویه دید مستقیماً منجر به تغییر مقیاس در تصویر، انسداد و توزیع ناهموار می‌شود.

۵ نتیجه گیری

این مقاله به طور خلاصه روش‌های سنتی شمارش جمعیت و روند توسعه آنها را خلاصه می‌کند. ما بر روش‌های جمعیت‌شمار مبتنی بر CNN تمرکز کردیم و آنها را با توجه به ایدئولوژی راهنمایان به چند دسته تقسیم کردیم و سپس به طور جداگانه توضیح دادیم. همچنین ما برخی از مجموعه داده‌ها را شرح دادیم و روش‌های مختلف حاشیه نویسی موجود را به تفصیل اشاره کردیم. سپس روش‌های ارزیابی به منظور مقایسه دقیق‌تر عملکرد مدل را بیان کردیم. علاوه بر این، کار شمارش جمعیت مبتنی بر ویدیو را نیز بطور خلاصه بیان کردیم. در نهایت، مشکلاتی را که در شمارش جمعیت و راه‌حل‌ها و نتیجه‌گیری‌های مربوط به آنها، روندهای آتی با آن مواجه می‌شوند، مورد بحث قرار دادیم. امیدواریم که این بررسی به درک کلی از شمارش جمعیت و تخمین تراکم بدهد.

مراجع

- 1- Subburaman VB, Descamps A, Carincotte C (2012) Counting people in the crowd using a generic head detector. In: 2012 IEEE ninth international conference on advanced video and signal-based surveillance.
- 2- Zhu L, Li C, Yang Z, Yuan K, Wang S. Crowd density estimation based on classification activation map and patch density level. *Neural Computing and Applications*. 2020 May;32(9):5105-16.
- 3- Bhuiyan MR, Abdullah J, Hashim N, Al Farid F, Uddin J, Abdullah N, Samsudin MA. Crowd density estimation using deep learning for hajj pilgrimage video analytics. *F1000Research*. 2021;10.
- 4- Al Farid F, Hashim N, Abdullah J, Bhuiyan MR, Shahida Mohd Isa WN, Uddin J, Haque MA, Husen MN. A Structured and Methodological Review on Vision-Based Hand Gesture Recognition System. *Journal of Imaging*. 2022 Jun;8(6):153.
- 5- Dong L, Zhang H, Ji Y, Ding Y. Crowd counting by using multi-level density-based spatial information: A Multi-scale CNN framework. *Information Sciences*. 2020 Aug 1;528:79-91.
- 6- Liu X, Weijer JVD, Bagdanov AD (2018) Leveraging unlabeled data for crowd counting by learning to rank. In: 2018 IEEE/CVF conference on computer vision and pattern recognition (CVPR)
- 7- Liu X, van de Weijer J, Bagdanov AD (2019) Exploiting unlabeled data in cnns by self-supervised learning to rank. *IEEE Trans Pattern Anal Mach Intell* 41:1862–1878
- 8- Jing ZO, Liang HE, Binke HU. An incremental learning algorithm based on support vector machine for multi-scenario crowd counting. *微电子学与计算机*. 2022;39(2):75-83.
- 9- De Sanctis M, Di Domenico S, Fioravanti D, Abellán EB, Rossi T, Cianca E. RF-Based Device-Free Counting of People Waiting in Line: A Modular Approach. *IEEE Transactions on Vehicular Technology*. 2022 Jun 13;71(10):10471-84.
- 10- Wang H, Zhang K, Su Z, Lu J, Xiong Z. Graph clustering-based crowd counting with very limited labelled samples. *Electronics Letters*. 2020 Jul;56(14):709-12.
- 11- Xu J, Bo C, Wang D. A novel multi-target multi-camera tracking approach based on feature grouping. *Computers & Electrical Engineering*. 2021 Jun 1;92:107153.
- 12- Liu X, Yang J, Ding W, Wang T, Wang Z, Xiong J. Adaptive mixture regression network with local counting map for crowd counting. In *European Conference on Computer Vision 2020 Aug 23* (pp. 241-257). Springer, Cham.
- 13- Zhou D, He Q. Cascaded multi-task learning of head segmentation and density regression for rgbd crowd counting. *IEEE Access*. 2020 May 29;8:101616-27.
- 14- Sindagi VA, Yasarla R, Babu DS, Babu RV, Patel VM. Learning to count in the crowd from limited labeled data. In *European Conference on Computer Vision 2020 Aug 23* (pp. 212-229). Springer, Cham.
- 15- Rodriguez M, Laptev I, Sivic J et al (2011) Density-aware person detection and tracking in crowds. In: *IEEE international conference on computer vision, ICCV 2011, Barcelona, Spain, November 6–13, 2011*
- 16- Li J, Huang L, Liu C (2011) Robust people counting in video surveillance: dataset and system. In: *2011 8th IEEE international conference on advanced video and signal based surveillance (AVSS)*. pp 54–59

- 17- Pham VQ, Kozakaya T, Yamaguchi O et al (2015) Count forest: co-voting uncertain number of targets using random forest for crowd density estimation. In: International conference on computer vision (ICCV 2015)
- 18- Min F, Pei X, Li X et al (2015) Fast crowd density estimation with convolutional neural networks. *Eng Appl Artif Intell* 43:81–88
- 19- Zhang L, Shi M, Chen Q. Crowd counting via scale-adaptive convolutional neural network. In 2018 IEEE Winter Conference on Applications of Computer Vision (WACV) 2018 Mar 12 (pp. 1113-1121). IEEE.
- 20- Shen Z, Xu Y, Ni B et al (2018) Crowd counting via adversarial cross-scale consistency pursuit. *IEEE CVF Conf Comput Vis Pattern Recognit* 2018:5245–5254
- 21- Liu J, Gao C, Meng D et al (2017) Decidenet: counting varying density crowds through attention guided detection and density estimation. *IEEE CVF Conf Comput Vis Pattern Recognit* 2018:5197–5206
- 22- Shi M, Yang Z, Xu C et al (2018) Revisiting perspective information for efficient crowd counting. *IEEE CVF Conf Comput Vis Pattern Recognit CVPR* 2019:7271–7280
- 23- Huang S, Zhou H, Liu Y, Chen R. High-resolution crowd density maps generation with multi-scale fusion conditional GAN. *IEEE Access*. 2020 Jun 8;8:108072-87.
- 24- Yang J, Zhou Y, Kung S-Y (2018) Multi-scale generative adversarial networks for crowd counting. In: 2018 24th international conference on pattern recognition (ICPR).
- 25- Wang L, Li Y, Xue X (2019) Coda: counting objects via scaleaware adversarial density adaption. *IEEE Int Conf Multimed Expo (ICME)* 2019:193–198
- 26- Shen Z, Xu Y, Ni B et al (2018) Crowd counting via adversarial cross-scale consistency pursuit. *IEEE CVF Conf Comput Vis Pattern Recognit* 2018:5245–5254
- 27- Olmschenk G, Hao T, Zhu Z (2018) Crowd counting with minimal data using generative adversarial networks for multiple target regression. In: 2018 IEEE winter conference on applications of computer vision (WACV)
- 28- Sindagi VA, Patel VM. Generating high-quality crowd density maps using contextual pyramid cnns. In Proceedings of the IEEE international conference on computer vision 2017 (pp. 1861-1870).
- 29- Shang C, Ai H, Bai B. End-to-end crowd counting via joint learning local and global count. In 2016 IEEE International Conference on Image Processing (ICIP) 2016 Sep 25 (pp. 1215-1219). IEEE.
- 30- Liu W, Salzmann M, Fua P (2018) Context-aware crowd counting. *IEEE CVF Conf Comput Vis Pattern Recognit CVPR* 2019:5094–5103
- 31- Liu L, Wang H, Li G, Ouyang W, Lin L. Crowd counting using deep recurrent spatial-aware network. *arXiv preprint arXiv:1807.00601*. 2018 Jul 2.
- 32- Xiong H, Lu H, Liu C et al (2019) From open set to closed set: counting objects by spatial divide-and-conquer. *IEEE CVF Int Conf Comput Vis (ICCV)* 2019:8361–8370
- 33- Liu X, Yang J, Ding W (2020) Adaptive mixture regression network with local counting map for crowd counting. *arXiv :2005.05776*
- 34- Ranjan V, Le H, Hoai M. Iterative crowd counting. In Proceedings of the European Conference on Computer Vision (ECCV) 2018 (pp. 270-285).
- 35- Xiong F, Shi X, Yeung D-Y (2017) Spatiotemporal modeling for crowd counting in videos. *IEEE Int Conf Comput Vis ICCV* 2017:5161–5169

- 36- ang Y, Zhan B, Cai W et al (2019) Locality-constrained spatial transformer network for video crowd counting. IEEE Int Conf Multimed Expo ICME 2019:814–819
- 37- Wu X, Xu B, Zheng Y et al (2019) Video crowd counting via dynamic temporal modeling. arXiv:1907.02198
- 38- Zou Z, Shao H, Qu X et al (2019) Enhanced 3d convolutional networks for crowd counting. arXiv:1908.04121
- 39- Zheng H, Lin Z, Cen J et al (2019) Cross-line pedestrian counting based on spatially-consistent two-stage local crowd density estimation and accumulation. IEEE Trans Circuits Syst Video Technology 29(3):787–799
- 40- He G, Ma Z, Huang B et al (2019) Dynamic region division for adaptive learning pedestrian counting. IEEE Int Conf Multimed Expo ICME 2019:1120–1125
- 41- Redmon J, Farhadi A (2018) Yolov3: an incremental improvement. arXiv:1804.02767
- 42- https://svip-lab.github.io/dataset/campus_dataset.html
- 43- <http://www.ee.cuhk.edu.hk/~xgwang/expo.html>
- 44- <http://iot.ee.surrey.ac.uk:8080/datasets.html>
- 45- <https://www.crcv.ucf.edu/data/ucf-qnrf/>
- 46- <https://www.crcv.ucf.edu/data/ucf-cc-50/>
- 47- <https://grail.cs.washington.edu/projects/bal/venice.html>
- 48- <http://www.crowd-counting.com/>
- 49- Lempitsky V, Zisserman A. Learning to count objects in images. Advances in neural information processing systems. 2010;23.
- 50- Oghaz MM, Khadka AR, Argyriou V et al (2019) Content-aware density map for crowd counting and density estimation. arXiv :1906.07258
- 51- Olmschenk G, Tang H, Zhu Z (2019) Improving dense crowd counting convolutional neural networks using inverse k-nearest neighbor maps and multiscale upsampling. arXiv:1902.05379
- 52- Liu X, Weijer JVD, Bagdanov AD (2018) Leveraging unlabeled data for crowd counting by learning to rank. In: 2018 IEEE/CVF conference on computer vision and pattern recognition (CVPR)