مروری بر شمارش جمعیت و تخمین تراکم بر اساس شبکه عصبی پیچشی مبتنی بر رگرسیون

نام و نام خانوادگی نویسنده: نسرین رنجبران*

گروه آموزشی کامپیوتر دانشکده مهندسی دانشگاه بوعلی سینا، همدان

پست الکترونیک: ranjbaran.n32@gmail.com

خلاصه

روش های شمارش جمعیت و تخمین تراکم در حوزه امنیت عمومی از اهمیت بالایی برخوردار است. تخمین تراکم جمعیت و شمارش از روی تصویر یا فریم ویدئو به بخش اساسی سیستم بینایی کامپیوتری در سناریوهای مختلف تبدیل شده است. در نوشتار حاضر مروری بر چند جنبه ی شمارش جمعیت با کمک شبکه عصبی پیچشی مبتنی بر رگرسیون داریم. سپس بررسی ، مقایسه روش ها و نتیجه گیری را ارائه می دهیم.

کلمات کلیدی:
Crowd Counting - Density Estimation - Convolutional Neural Network - Deep Learning

مقدمه

در حالت کلی روش های جمعیت شماری به ترتیب تکامل در چند دسته قرار می گیرند:
الف) روش های مبتنی بر تشخیص: هدف در روش‌های مبتنی بر تشخیص این است که ابتدا قسمت‌هایی از بدن یا کل بدن هر فرد را با استفاده از ویژگی‌های سطح پایین شناسایی کنند و سپس بر اساس آنها تعداد افراد را شمارش کنند.
ب) روشهای مبتنی بر رگرسیون: در این روشها نگاشت بین تصاویر یا فریم های دیتاست به تعداد تعداد کل افراد در تصویر با استفاده از روش‌های یادگیری ماشین انجام می شود.
ج) روش‌های مبتنی بر نقشه چگالی: نقشه چگالی، متفاوت از دادن صرفاً تعداد افراد، برچسبی برای نشان دادن مکان افراد از نظر هندسی ارائه می دهد که می توان در سایر کاربردها استفاده کرد.

تصویر ۱: یک نمونه از عکس از دیتاست ShanghaiTech و نقشه چگالی آن

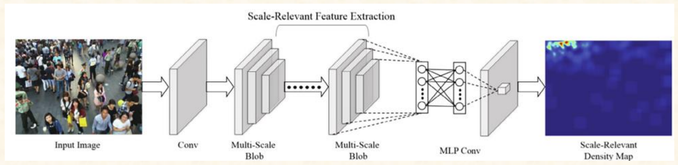
د) شمارش توسط شبکه عصبی پیچشی(CNN): شمارش از طریق CNN از نظر دقت کارآمدتر است، اما پیچیدگی محاسباتی بالاتری دارد. دلیل اصلی موفقیت CNN در کار شمارش جمعیت و تخمین تراکم جمعیت، توانایی آن در یادگیری روابط غیر خطی بین تصاویر و تعداد افراد در تصاویر یا نقشه چگالی متناظر آنها است.

با توجه به دیتاست استفاده شده و خروجی شبکه، روش‌های شمارش جمعیت مبتنی بر CNN را به سه دسته دسته‌بندی می‌کنیم:
(۱) روش CNN مبتنی بر تشخیص (۲) روش CNN مبتنی بر رگرسیون. (۳) ترکیب دو روش قبلی

در این پوستر مروری بر مدل های ارائه شده با روش CNN مبتنی بر رگرسیون را داریم.

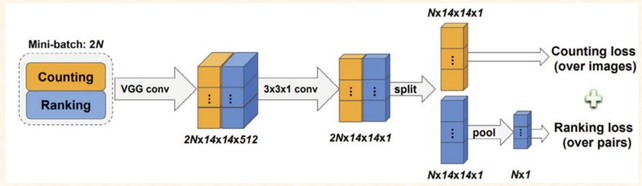
مواد و روشها

نقشه چگالی رگرسیون می تواند شمارش را بدست آورد و تراکم جمعیت را تخمین بزند. در مقایسه با روش‌های CNN مبتنی بر تشخیص، تولید روش‌های مبتنی بر نقشه چگالی می‌تواند نتایج پیش‌بینی بهتری در تراکم بالاتر داشته باشد. انواع مدل های این روش را در زیر بحث خواهیم کرد:
۱- مدل چند مقیاسی: این روش‌ها عمدتاً اطلاعات مقیاس‌های مختلف در تصویر را از طریق ساختار چند ستونی یا شبکه‌های هرمی ویژگی یا مقیاس‌بندی استخراج می‌کنند.



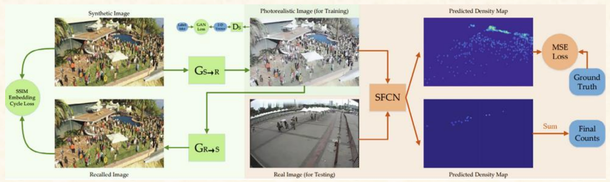
تصویر ۲: شبکه عصبی پیچشی چند مقیاسی برای شمارش جمعیت پیشنهاد شده توسط Zeng و همکاران.

۲- مدل متن آگاه(Context-aware): این روش ها برای بهبود دقت تشخیص، اطلاعات پس زمینه محلی و عمومی را به شبکه اضافه می کنند.
۳- مدل وظیفه کمکی(Auxiliary-task): یک یا چند وظیفه مربوط به شمارش جمعیت به عنوان وظایف کمکی برای آموزش با هم به شبکه اضافه می شود.



تصویر ۳: مدل چند وظیفه ای ترکیبی از شمارش و رتبه بندی پیشنهاد شده توسط لیو و همکاران.

۴- مدل های مقابله ای با کمبود داده برچسب دار: این روش مدل ها را به مدل نیمه نظارت شده، مدل ضعیف نظارت شده و مدل خود نظارتی تقسیم می کند.
۵- مدل تطبیق دامنه (Domain Adaptation): این روش‌ها می‌توانند در تعداد هر شیئی را در دامنه بشمرند.



تصویر ۴: مدل تطبیق دامنه پیشنهاد شده توسط وانگ و همکاران.

۶- مدل نقشه پرسپکتیو (Perspective Map): نقشه پرسپکتیو منعکس کننده اعوجاج پرسپکتیو در هر موقعیت در تصویر است (یعنی چند پیکسل با یک متر قد شخص مطابقت دارد). این روشها نقشه پرسپکتیو را برای کمک به تولید نقشه چگالی به شبکه اضافه می کنند.
۷- مدل مکانیسم توجه (Attentional mechanism): این روش ها مکانیسم توجه را به شبکه اضافه می کنند تا عملکرد شمارش جمعیت را بهبود بخشد.
۸- مدل جستجوی شبکه ای (Network Search): این روش‌ها ساختار شبکه ای را برای کار شمارش جمعیت توسط متد NAS به دست می‌آورند.

نتایج

جدول زیر نتایج اجرای مدل های مختلف هر روش ذکر شده در قسمت قبل را روی دیتاست های مختلف نشان میدهد.

Dataset	UCSD		MAll		UCF_CC_50		WorldExpo'10		SHTech partA		SHTech part B	
Evaluation Criteria	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
CSCC [19]	1.60	3.31	–	–	467.0	498.5	12.9	–	181.8	277.7	32.0	49.8
MCNN [12]	1.07	1.35	–	–	377.6	509.1	11.6	–	110.2	173.2	26.4	41.3
CNN-Boosting [43]	1.10	–	2.01	–	364.4	–	–	–	–	–	–	–
CSRNet [18]	1.16	1.47	–	–	266.1	397.5	8.6	1.47	68.2	115.0	10.6	16.0
CNN-pixel [49]	1.12	2.06	–	–	406.2	404.0	13.4	–	–	–	–	–
CMTL [21]	–	–	–	–	332.8	341.4	–	–	101.3	152.4	20.0	31.1
Switching-CNN [15]	1.62	2.10	–	–	318.1	439.2	9.4	–	90.4	135.0	21.6	33.4
MSCNN [90]	–	–	–	–	363.7	468.4	11.7	–	83.8	127.4	17.7	30.2
NetVLAD [122]	–	–	–	–	311.3	401.8	10.3	–	107.6	169.3	21.4	33.9
SaCNN [132]	–	–	–	–	314.9	424.8	8.5	–	86.8	139.2	16.2	25.8
DA-Net [128]	1.03	1.31	–	–	290.8	326.5	–	–	71.6	104.9	15.0	21.9
FCNN [150]	–	–	–	–	253.1	356.4	–	–	67.6	110.6	10.1	18.8
BSAD [161]	1.00	1.40	–	–	409.5	563.7	10.5	–	–	–	20.2	35.6
IG-CNN [162]	–	–	–	–	291.4	349.4	11.3	–	72.5	118.2	13.6	21.1
IC-CNN [169]	–	–	–	–	260.9	365.5	10.3	–	68.5	116.2	10.7	16.0
SANet [170]	1.02	1.29	–	–	258.4	334.9	8.2	–	67.0	104.5	8.4	13.6
DRSAN [172]	–	–	**1.72**	**2.1**	219.2	250.2	7.76	–	69.3	96.4	11.1	18.2
HA-CCN [174]	–	–	–	–	256.2	348.4	–	–	62.9	94.9	8.1	13.4
PACNN [177]	**0.89**	**1.18**	–	–	241.7	320.7	–	–	62.4	102.2	7.6	11.8
CANet [182]	–	–	–	–	**212.2**	**243.7**	**7.2**	–	62.3	100.0	7.8	12.2
L2SM [194]	–	–	–	–	**188.4**	315.3	–	–	64.2	98.4	7.2	11.1
SAANet [196]	–	–	**1.28**	**1.68**	271.6	391.0	–	–	–	–	16.86	28.4
E3DNet [197]	**0.93**	**1.17**	**1.64**	**2.13**	–	–	8.32	–	–	–	–	–
DSSINet [198]	–	–	–	–	216.9	302.4	**6.67**	–	60.63	**96.04**	**6.85**	**10.34**
SPANet [200]	**1.00**	**1.28**	–	–	232.6	311.7	–	–	**59.4**	**92.5**	**6.5**	**9.9**
PGCNet [188]	–	–	–	–	244.6	361.2	8.1	–	**57.0**	**86.0**	8.8	13.7
AMSNet [71]	–	–	–	–	**208.6**	**296.3**	**6.8**	–	**58.0**	96.2	**7.1**	**10.4**

بحث و نتیجه گیری

تجزیه و تحلیل نتایج متدها نشان می دهد که در مقایسه با شبکه چند ستونی، شبکه تک ستونی با پارامترهای شبکه کمتر عملکرد بهتری دارد. نیز استفاده از مکانیسم توجه باعث می شود شبکه بیشتر بر روی قسمت مورد نظر تمرکز کند و نتیجه بهتری حاصل می شود.روش های نیمه نظارتی عملکرد قابل قبول ندارند. همچنین افزودن اطلاعات پرسپکتیو و استفاده از شبکه ترانسفورماتور فضایی (STN)، به حل مشکل چرخش و تغییر شکل جمعیت کمک می کند و برای صحنه های شلوغ مناسب است.

لیست اختصارات

CNN: Convolutional Neural Network

منابع

1. Zizhu Fan a , Hong Zhang Zheng Zhang , Guangming Lu b , YuCrowd counting Crowd density estimation Convolutional neural network Deep learningdong Zhang, Yaowei Wang (2021). A survey of crowd counting and density estimation based on convolutional neural network, Nurocomputing, 224 -251.
2. Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, “Single-image crowd counting via multi-column convolutional neural network,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 589–597, Jun. 2016 .”.
3. D. B. Sam, S. Surya, and R. V. Babu, “Switching convolutional neural network for crowd counting,” in EEE Conference on Computer Vision and Pattern Recognition, pp. 4031–4039, IEEE, Jul. 2017.”.
4. Ning Liu, Yongchao Long, Changqing Zou, Qun Niu, Li Pan and Hefeng Wu, "Adcrowdnet: An attention-injective deformable convolutional network for crowd understanding", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3225-3 .”