2021
Computer Vision
Final Report

군중 집계 동향분석

엄태선 (umts2358@naver.com) **세종대학교 인공지능대학원 석사과정**

군중 집계는 효율적인 자원 배분, 응급 상황의 효과적인 관리 등 많은 분야에 적용된다. 본 보고서에서는 다양한 군중 계수 방법을 조사하고 비교한다. 또한 기존 접근 방식의 한계를 식별하고 식별된 개방형 연구 과제를 해결하기 위한 향후 잡업의 의제를 제안한다.

* 본 동향분석은 2021년 세종대학교 대학원 컴퓨터비전 수업 기말고사 대체과제를 위한 제출 레포트임을 알린다.

I. 서론

기계 학습에 대한 연구가 시작된 이후로 군중 집계는 항상 중요한 연구 중 하나로 다뤄져왔다. 특히 2020년 코로나 바이러스 감염증-19(이하 코로나 19)의 장기화로 인해 이미지나 영상을 기반으로 사람들의 밀집도를 계산하는 것은 국가 단위의 중대사한 문제가 되었다. 그 외에도 군중 집계는 실제 이미지나 영상 등을 기반으로 통행량 조사, 인구 밀도 조사 등 다양한 부분에서 사용될 수 있다. 이는 그림 1에서 볼 수 있듯이 CCTV가 꾸준히 증가했다는 것과 관련이 있기도 하다. CCTV의 수가 증가한다는 것은 무작위 이미지와 영상에서 객체 인식 및 군중 분석을 할 원시 데이터의 수가 늘어난다고 할 수 있기 때문이다[1].

이러한 비전 기반 딥 러닝 접근 방식을 사용해 인간을 개별적으로 인식하거나 장면의 밀도 혹은 열 지도를 만든 다음 Convolutional Neural Network(CNN)를 사용한 뒤, 분석하여 생성된 밀도 지도의 픽셀 값을 합산하여 군중 수를 추정할 수 있다. 또한 이 과정에서 캡쳐된 이미지는 추후에 군중 집계 딥러닝 학습을 위한 네트워크에 공급된다.

이러한 군중 집계 및 분석은 세 가지 범주로 분류가 가능하다.

- *인구 수 계산 또는 밀도 추정*은 혼잡도를 밀도 또는 희소성 기준으로 정의하고 군중에 존재하는 참가자 수를 추정한다. 이는 일반적으로 정적 영상과 비디오 시퀀스를 분석하여 수행된다. 이 때 사용된 집계 방법에는 검출, 군집화, 회귀 등이 포함된다[1].
- *사람 추적*은 붐비는 장면의 연속적인 프레임, 즉 영상에서 개인의 위치를 파악하고 움직임을 추적한다.
- *군집 행동 분석*은 움직임, 속도, 흐름 방향 및 싸움이나 달리기와 같은 비정상적 사건 감지 측면에서 군중 행동을 모니터링한다.

본 보고서는 그림 2에 제시된 분류법에 따른 군중 계수 및 군중 행동 분석 기법을 조사한다.

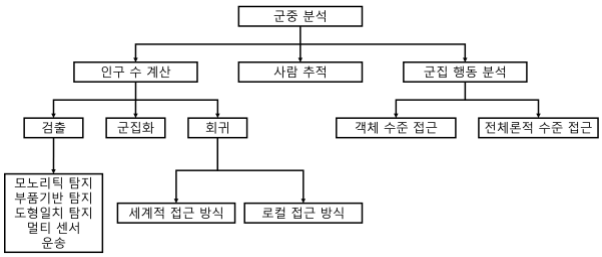

그림 1 군중 분석 및 모니터링 분류법

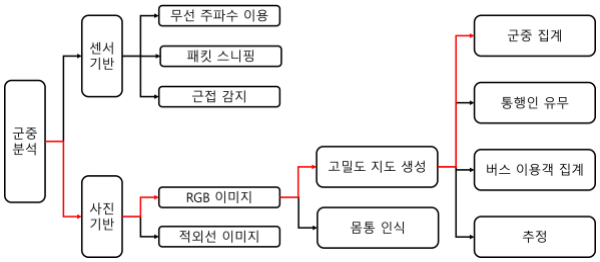

그림 2 군중 분석 기술의 분류법

본 보고서의 나머지 부분은 다음과 같이 구성되어 있다. 2장에서는 전파 신호 처리 접근 방식을 논의한다. 3장은 비전 기반 딥러닝 접근 방식을 제시한다. 4장은 군중 집계를 위한 과제와 해결책을 논의하고 5장에서 남은 연구 질문을 제시하고 이를 해결하기 위한 연구 의제를 스케치한다.

II. 센서 기반 접근법

센서 기반 군중 집계 접근 방식에는 패킷 스니핑, 무선 주파수 이용 및 근접 감지가 포함된다.

A. 패킷 스니핑

그림 3에서 볼 수 있듯이, 군중 분야에서 와이파이 또는 블루투스 액세스 포인트를 배치하면 액세스 포인트 영역에서 와이파이 패킷[2]과 블루투스 비콘을 감지할 수 있다. 송신기 및 수신기 MAC 주소는 교환된 패킷에서 추출하여 해당 영역 내의 장치 개수를 셀 수 있다. 이것은 가장 간단한 해결책이지만 많은 단점이 있다. 어떤 사람들은 하나 이상의 기기를 가지고 있을 수 있으며, 네트워크에 장치가 연결되지 않았다면 집계를 실패할 수도 있다.

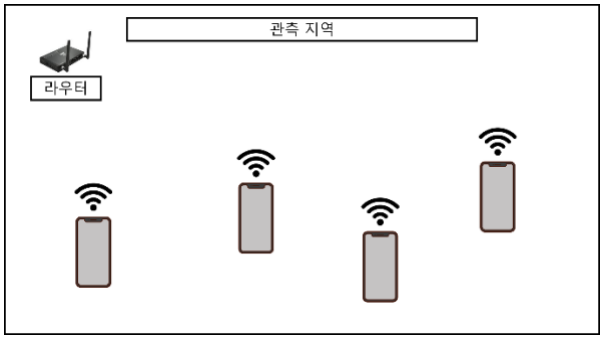

그림 3 패킷 또는 신호 스니핑을 사용하여 액세스 포인트 범위 내에서 인구 집계 실행

B. 무선 주파수 이용

대부분의 사람들은 유심칩이 있는 전자 기기를 적어도 한 개 이상 가지고 있다. 이 장치들은 송신탑에 정기적인 전파 신호를 보내 가용성을 나타내거나 대략적인 위치를 업데이트한다.

그림 4와 같이 간단한 송수신기를 이용하여 휴대폰이 동작하는 주파수 범위에서 신호의 힘을 측정할 수 있으며, 신호의 강도를 바탕으로 얼마나 많은 장치가 사용 가능한지를 예측하는 모델을 구축할 수 있다. 단점은 전화기의 무선 주파수가 통화 중에 집계 과정을 방해할 수 있는 경우에 도 높은 에너지를 가진다[3].

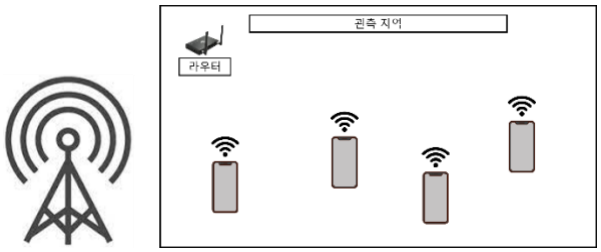

그림 4 휴대폰 작동 범위 내의 주파수를 감지

C. 근접 감지

이전의 군중 계수 방법은 운동 센서와 같은 근접 센서를 문 위에 배치하여 통과 여부를 판단한 다음 모니터링 대상 내부 또는 외부의 신체 흐름에 따라 특정 장소에서 사람들의 수를 증가시키거나 감소시키는 무선, 소리, 적외선 신호 처리를 사용했다[4].

III. 비전 기반 접근법

군중 집계 문제에서 가장 많이 사용되는 해결책은 사람 키보다 높은 곳에 장착된 카메라를

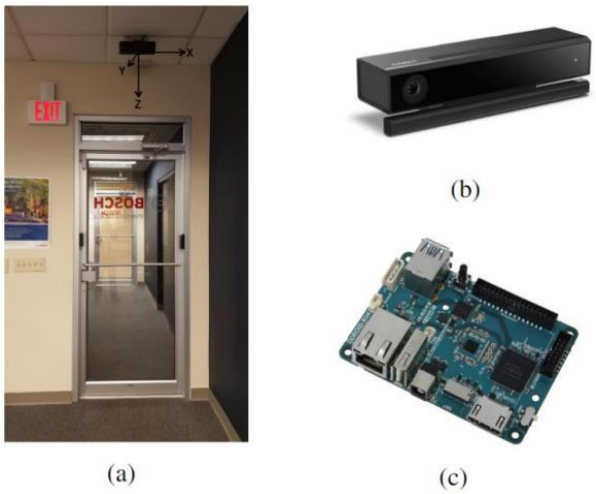

그림 5 (a) 천장에 Kinect 센서 배치. (b) XBOX One용 Kinect 센서. (c) 임베디드 컴퓨터 Odroid-XU4.

사용하는 것이다. 카메라를 높게 둘수록 프레임이 넓어진다. 카메라는 대부분 기둥, 가로등, 천장에 설치되어 있다. 어떤 사람들은 공중 사진으로 군중 집계를 시도하기도 했다[5]. 또 다른 접근 방식은 열화상 카메라를 사용하여 프레임에서 가장 가열된 영역을 기준으로 사람을 감지하는 것이다[6]. 열화상 이미지는 그림 6과 같이 파란색과 빨간색 사이의 색상으로만 구성되어 있기 때문에 RGB 이미지를 회색 스케일로 변환할 필요가 없기 때문에 처리가 용이하다.

다양한 이미지 처리와 컴퓨터 비전 기술은 군중 집계를 위한 다양한 방법을 제시했다. 해당 방법들은 그림 7에 묘사된 것과 같이 인식 기반 접근법과 특징 기반 접근법으로 분류된다.

그림 6 CNN 예측 모델의 입력으로 사용되는 적외선 이미지.

그림 7 군중 밀도 추정 및 집계 시스템을 위한 분류법

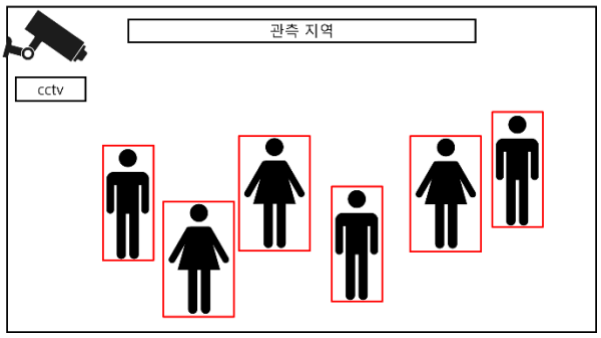

그림 8 프레임 내에 있는 사람들 개별적으로 탐지

A. 인식 기반 접근법

인식 기반 접근법은 Convolutional Neural Network(CNN)을 사용하여 이미지에서 각 사람을 감지하여 물체를 인간으로 분류하는 분류기를 훈련시킨다. 이를 통해 그림 8과 같이 프레임의 각 사람을 개별적으로 감지하고 계산할 수 있다[8]. 그러나 이 방법은 군중 밀도가 증가하면 성능이 저하될 수 있다.

B. 특징 기반 접근법

특징 기반 접근법은 그림 9와 같이 전체 장면의 밀도 지도 또는 열 지도를 생성하는 것을 주로 한다. 밀도 지도는 군중의 공간 정보를 제공한다. 밀도가 높은 맵의 각 픽셀은 지역 군중 밀도를 포함하며, 집계는 밀도가 높은 맵의 픽셀을 합산하여 추정한다[7]. 이것은 사람들을 개별적으로 감지하는 것보다 더 효율적이다. 또한 대부분의 프로세스 동안 개인적으로 식별할 수 있는 기능이 필요하지 않으므로 사람들의 프라이버시를 보호할 수 있다.

그 외에도 군중 집계를 위해 비지도 학습, 지도 학습이 제안되었다.

그림 9 입력 이미지 및 해당 암호화 밀도 맵

C. 비지도 학습

해당 이미지 해상도의 약 1/3을 여전히 볼 수 있고 식별 가능한 상대적으로 더 높은 품질의 이미지를 사용한다. 그 후, 다음과 같은 방식으로 해당 이미지에 대한 알고리즘을 훈련한다. 이미지가 완전히 채워진 s1이면 논리적으로 해당 이미지의 잘라낸 구역 s2는 주 이미지인 s1보다 적거나 같은 수의 사람을 포함하게 된다. 이 때, count는 혼잡한 이미지를 취하고 그 안에 있는 대략적인 수의 사람을 계산하는 함수이다.

$$\boldsymbol{Count(S2) \geq Count(S1)} \qquad \textbf{(1)}$$

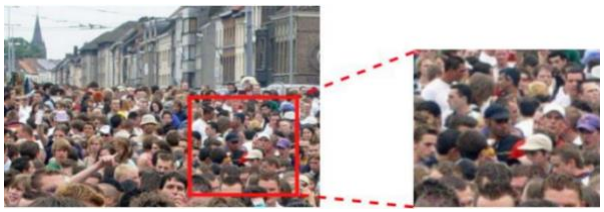

그림 10 각각 이미지 S1 및 S2이다. 이미지가 잘라낸 부분보다 더 많거나 동등한 사람을 가진 학습 방법론이다.

그러한 논리에 기초하여 [9]에서 논의한 것과 같이 자른 이미지에 비해 적은 수의 사람을 예측할 수 있도록 모델을 훈련시킬 수 있다.

D. 지도 학습

CNN은 이전에 캡처한 이미지와 전체 주석이 달린 밀도 맵을 기반으로 군중 밀도를 예측하는 데 사용할 수 있다. 사용된 데이터 셋의 이미지는 최대 너비가 1024px이고 높이가 768px다. 모든 이미지가 정확한 크기를 가진 것은 아니지만 밀도맵 생성 모델에는 컨볼루션과 최대 풀링

레이어만 포함되어 있어 별다른 입력 크기가 필요하지 않았다. 훈련 시간을 줄이기 위해 입력 이미지가 그레이 스케일로 변환되는데, 이는 장면의 색상이 매우 다양하기 때문에 일반화에도 도움이 된다. 또한, 훈련 세트의 일부 이미지는 회색조로만 구성된다.

표1은 [9]에 따라 여러 방법이 동일한 "ShanghaiTech_A" 데이터셋에서 Mean Square Error(MSE), Mean Approximate Error(MAE)를 계산한 결과이다. 현재 실험중인 모델은 CMTL이다.

그림 11은 지도 학습과 비지도 학습을 모두 사용하여 제안된 하이브리드 접근 방식을 요약하고 있다. 두 경우 모두 모델에 대한 입력은 사용된 모델 요구를 준수할 뿐만 아니라 일부는 기본적으로 흑백으로 촬영되었기 때문에 데이터셋의 모든 영상에 대처하기 위해 회색조여야 한다. 제안된 CNN 모델은 [14]에

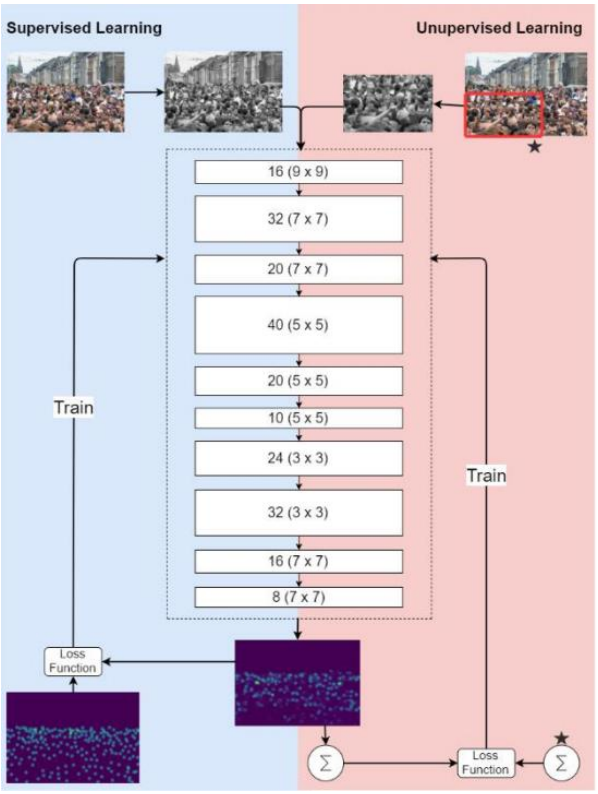

그림 11 동일한 모델을 가르치는 것에 대한 지도 학습과 비지도 학습의 차이점

제시된 접근 방식에 기초한다. 지도 학습의 경우 출력 밀도 맵을 원본 맵과 비교하고 그 중에서 모델에 피드백하는 MSE, MAE를 계산한다. 반면에 우리는 카운트에 의존하지만 합리적인 출력 밀도 맵을 기대하지 않는다. 앞에서 설명한 대로, 자른 이미지에는 더 적은 수의 이미지가 포함되어 있어야하며, 그렇지 않으면 상위 이미지에 대해 동일한 사용자가 카운트된다. 따라서 모델이 상위 추정치 카운트보다 더 높은 카운트를 추측한 경우 두 카운트 간의 손실을 기반으로 다시 학습한다. 주석이 달린 데이터가 부족한 경우 비지도 학습을 사용하여 훈련할 수 있는 군중 이미지를 무작위로 수집할 수 있다.

IV. 군중 집계 문제

이번 장에서는 주요 공개 군중 집계 과제를 요약한다. 주요 사항은 프라이버시 문제와 광역 인원 집계와 관련이 있다.

제일 먼저 나온 군중 집계 방식은 얼굴 또는 지배적인 신체 특징을 인식한 다음 군중 카운터를 증가시킨 것이다[8]. 이는 프레임에 있는 여러 사람에게도 작용하지만, 모든 머리나 사람의 크기가 같지 않기 때문에 CNN에서 매우 정확한 분할 알고리즘이 필요하다. 이 접근법은 사람들의 얼굴을 인식하여 저장해야 하기 때문에 사생활에 대한 우려를 야기한다. 포착된 장면에서 직접 밀도 지도를 생성해 밀도 계산을 하는 것이 사생활 보장을 더 잘 할 수 있다. 사람의 존재가 예상되는 픽셀 색상에서 큰 차이를 만들 수 있는 히트맵을 만들면 된다. 이제 이 히트맵은 CNN을 통해 별도로 분석하거나 픽셀 값을 합산하여 군중 수를 구할 수 있다.

표 1 모델 성능 비교

모델	MAE	MSE
CMTL [10]	101.3	152.4
Switching-CNN [9]	90.4	135.0
TopDownFeedback [11]	97.5	145.1
CP-CNN [12]	73.6	106.4
IG-CNN [13]	72.5	118.2
Crowd-CNN [16]	181.8	277.7
MCNN [17]	110.2	173.2
Cascaded-MTL [18]	101.3	152.4
Switch-CNN [19]	90.4	135.0
ACSCP [20]	75.7	102.7
CP-CNN [21]	73.6	106.4
D-ConvNet [22]	73.5	112.3
IG-CNN [23]	72.5	118.2
Ic-CNN [24]	69.8	116.2
CSRNet [25]	68.2	115.0
SANet [26]	67.0	104.5
LSC-CNN [27]	66.4	117.0
CAN [28]	62.3	100.0
S-DCNet [29]	58.3	95.0
SGANet+CL [30]	57.6	101.1

V. 참고문헌

[1] S. Lamba and N. Nain, “Crowd Monitoring and Classification : A Survey Crowd Monitoring and Classification : A Survey,” no. October, pp.20–31, 2017.

[2] T. Oransirikul, I. Piumarta, and H. Takada, “Classifying passenger and non-passenger signals in public transportation by analysing mobile device Wi-Fi activity,” J. Inf. Process., vol. 27, pp. 25–32, 2019.

[3] C. Kowalczuk et al., “Absence of nonlinear responses in cells and tissues exposed to RF energy at mobile phone frequencies using a doubly resonant cavity,” Bioelectromagnetics, vol. 31, no. 7, pp. 556–565, 2010

[4] S. Munir et al., “Real-time fine grained occupancy estimation using depth sensors on ARM embedded platforms,” Proc. IEEE Real-Time Embed. Technol. Appl. Symp. RTAS, vol. 1, pp. 295–306, 2017.

[5] V. Kurama, “Dense and Sparse Crowd Counting Methods and Techniques: A Review,” 2019. [Online]. Available: https://nanonets.com/blog/crowd-counting-review/.

[6] I. J. Amin, A. J. Taylor, F. Junejo, A. Al-Habaibeh, and R. M. Parkin, “Automated people-counting by using low-resolution infrared and visual cameras,” Meas. J. Int. Meas. Confed., vol. 41, no. 6, pp. 589–599, 2008.

[7] E. Walach and L. Wolf, “Learning to count with CNN boosting,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect.Notes Bioinformatics), vol. 9906 LNCS, pp. 660–676, 2016.

[8] T. Parthornratt, N. Burapanonte, and W. Gunjarueg, “People identification and counting system using raspberry Pi (AU-PiCC: Raspberry Pi customer counter),” Int. Conf. Electron. Information, Commun. ICEIC 2016, 2016.

[9] V. A. Sindagi and V. M. Patel, “Inverse Attention Guided Deep Crowd Counting Network,” 2019.

[10] V. A. Sindagi and V. M. Patel, “CNN-Based cascaded multi-task learning of high-level prior and density estimation for crowd counting,” 2017 14th IEEE Int. Conf. Adv. Video Signal Based Surveillance, AVSS 2017, 2017

[11] D. Babu Sam and R. Venkatesh Babu, “Top-down feedback for crowd counting convolutional neural network,” 32nd AAAI Conf. Artif. Intell. AAAI 2018, pp. 7323–7330, 2018.

[12] V. A. Sindagi and V. M. Patel, “Generating High-Quality Crowd Density Maps Using Contextual Pyramid CNNs,” Proc. IEEE Int. Conf. Comput. Vis., vol. 2017-Octob, pp. 1879–1888, 2017.

[13] D. B. Sam, N. N. Sajjan, R. V. Babu, and M. Srinivasan, “Divide and Grow: Capturing Huge Diversity in Crowd Images with Incrementally Growing CNN,” Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., pp. 3618–3626, 2018.

[14] F. Tong, Z. Zhang, H. Wang, and Y. Wang, “Concise Convolutional Neural Network for Crowd Counting,” 2018 10th Int. Conf. Adv. Infocomm Technol. ICAIT 2018, pp. 174–178, 2019.

[15] B. Yang, J. Cao, X. Liu, N. Wang, and J. Lv, “Edge computing-based real-time passenger counting using a compact convolutional neural network,” Neural Comput. Appl., no. November 2018, 2018

[16] Zhang, C., Li, H., Wang, X., & Yang, X. (2015). Cross-scene crowd counting via deep convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 833-841).

[17] Zhang, Y., Zhou, D., Chen, S., Gao, S., & Ma, Y. (2016). 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).

[18] Sindagi, V. A., & Patel, V. M. (2017, August). Cnn-based cascaded multi-task learning of high-level prior and density estimation for crowd counting. In *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)* (pp. 1-6). IEEE.

[19] Babu Sam, D., Surya, S., & Venkatesh Babu, R. (2017). Switching convolutional neural network for crowd counting. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5744-5752).

[20] Shen, Z., Xu, Y., Ni, B., Wang, M., Hu, J., & Yang, X. (2018). Crowd counting via adversarial cross-scale consistency pursuit. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5245-5254).

[21] Sindagi, V. A., & Patel, V. M. (2017). Generating high-quality crowd density maps using contextual pyramid cnns. In *Proceedings of the IEEE international conference on computer vision* (pp. 1861-1870).

[22] Shi, Z., Zhang, L., Liu, Y., Cao, X., Ye, Y., Cheng, M. M., & Zheng, G. (2018). Crowd counting with deep negative correlation learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5382-5390).

[23] Sam, D. B., Sajjan, N. N., Babu, R. V., & Srinivasan, M. (2018). Divide and grow: Capturing huge diversity in crowd images with incrementally growing cnn. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3618-3626).

[24] Ranjan, V., Le, H., & Hoai, M. (2018). Iterative crowd counting. In *Proceedings of the European*

Conference on Computer Vision (ECCV) (pp. 270-285).
[25] Li, Y., Zhang, X., & Chen, D. (2018). Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1091-1100).
[26] Cao, X., Wang, Z., Zhao, Y., & Su, F. (2018). Scale aggregation network for accurate and efficient crowd counting. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 734-750).
[27] Sam, D. B., Peri, S. V., Sundararaman, M. N., Kamath, A., & Radhakrishnan, V. B. (2020). Locate, size and count: Accurately resolving people in dense crowds via detection. *IEEE transactions on pattern analysis and machine intelligence*.
[28] Liu, W., Salzmann, M., & Fua, P. (2019). Context-aware crowd counting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5099-5108).
[29] Xiong, H., Lu, H., Liu, C., Liu, L., Cao, Z., & Shen, C. (2019). From open set to closed set: Counting objects by spatial divide-and-conquer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 8362-8371).
[30] Wang, Q., & Breckon, T. P. (2019). Crowd Counting via Segmentation Guided Attention Networks and Curriculum Loss. *arXiv preprint arXiv:1911.07990*.