

Crowd Counting 연구 동향

21110691 우소연

Crowd Counting 연구 개요

Crowd Counting란 이미지에서 군중들의 대략적인 숫자를 자동으로 추정하는 것을 목표로 합니다. Crowd Counting에 적용되고 있는 이미지의 개체 수를 계산하는 방법은 컴퓨터 비전의 기본적인 목표인 Object Counting입니다. Object Counting은 단일 이미지 또는 비디오에서 개체의 수를 계산하는 방법입니다. Object Counting의 활용범위는 CCTV를 이용한 감시, 미생물학, 교통혼잡도 추정 등 응용 프로그램으로 사용되고 있습니다. Object Counting의 유형은 크게 2가지로 Detection-based Object Counting, Regression-based Object Counting으로 나뉩니다. Detection-based Object Counting은 감지기를 사용하여 이미지에서 대상 객체를 식별하고 몇 개가 존재하는지 계산합니다. 이 방법은 얼굴을 감지하는데 좋은 성능을 보이지만 물체가 밀집된 이미지에서는 정확하게 식별하기 힘듭니다. Regression-based Object Counting은 CNN을 사용하여 end-to-end regression을 구축합니다. 또한 Regression기반의 방법론은 잘린 image patch에서 추출된 특징을 기반으로 예측된 개체 수를 실제 값에 매핑합니다. Crowd Counting 연구에서 주로 나타나는 문제점은 심각한 occlusions(군중 밀도가 증가함에 따라 군중들이 부분적으로 서로를 가리는 것처럼 보이는 것), scale 변환, 원근 왜곡, 날씨변화 등이 존재하였습니다. 앞선 문제점들을 해결하기 위해 어떤 연구가 이루어졌는지 동향을 통해 설명하겠습니다.

Crowd Counting 연구 동향

이미지 속 Crowd Counting은 이미지 처리의 일반적인 응용 분야 중 하나입니다. 빠르게 변화하는 사회환경의 배경 속에서 공적인 데이터 수집을 목적으로 Crowd Counting이 진행되고 있습니다. 초기의 Crowd Counting 연구는 Detection-based, clustering-based, regression-based의 방법론이 나타났습니다. 이러한 방법론은 주로 디지털 이미지 처리 이론에서 파생되어 특징을 추출하는 방향으로 구축합니다. 초기에는 연구를 진행하기 위한 알맞은 Data set이 거의 확립되지 않았습니다. 2010년 V. Lempitsky and A. Zisserman은 이미지 특징을 추출하기 위해 Crowd Counting에 밀도 추정을 도입하여 Counting의 정확도를 크게 향상시켰습니다. 밀도 추정은 군중 이미지를 매핑하고 이미지에 표시된 픽셀당 사람을 계산하는 데 초점을 맞추었습니다. 그 이후, Crowd Counting과 신경망을 이용한 밀도 추정을 위한 새로운 딥 러닝 알고리즘이 활발하게 연구되었습니다.

최근 Estimation-based counting의 발전 동향

밀도 추정 기반의 방법론은 픽셀 값을 사용하여 주어진 지점의 평균 인구 수를 나타낼 수 있을 뿐만 아니라 명시적으로 개인을 감지하지 않고 군중 분포의 위치 관계와 클러스터링 추세를 표시할 수 있어 분석에 유리합니다. 밀도 맵 기반의 방법은 객체의 심각한 occlusion으로 인한 어려움을 우회하여 더 나은 결과를 달성할 수 있으며 개인의 정보보호가 가능하고, 새로운 공간 정보를 보존하여 객체 계산을 해결하는데 효과적으로 나타났습니다.

1) Density Estimation-based Counting (밀도 추정 기반)

밀도 추정 기반 방법의 성능은 주로 사용되는 특징 유형에 따라 달라지며, 많은 방법이 수작업으로 만든 특징 또는 개별적으로 훈련된 무작위 포레스트 특징을 사용합니다. 최근에는 컴퓨터 비전 분야에 대한 CNN의 성과로 일부 문헌이 CNN 배경에서 KNN(K-근접 이웃), CNN(Convolutional Neural Network) 및 심층 CNNs 등과 같은 신경망(NN) 모델을 설정하여 작업을 계산하기 위한 밀도 추정에 딥 러닝 방법을 도입했습니다. 형상 추출 프로세스의 관점에서 밀도 추정 기반 알고리즘은 부분 공간 학습을 통한 추정과 전역 카운트를 통한 추정으로 분류할 수 있습니다.

2) Advances in CNN-based Estimation (CNN기반 추정의 발전)

회귀 기반 방법 중 CNN 네트워크 구조를 사용한 밀도 맵 추정 기반 Counting 접근 방식은 특히 유망합니다. 이러한 방법은 계산 성능을 꾸준히 개선하기 위해 심층 CNN의 힘에 의존하며 최근 작업에서는 CNN 기반 추정이 대부분을 차지합니다. Inception 및 Shortcuts와 같은 구조를 사용하는 모델은 MCNN, CrowdNet, Hydra CNN, Switch-CNN, CSRnet 및 TEDnet, DM-Count등 이 존재합니다.

Model Name	Brief description	Reference
MCNN	세 개의 작은 네트워크로 구성되어 있음. 세 개의 작은 네트워크는 서로 다른 수용 영역을 갖고 있으며 해당 구조에서는 각각의 네트워크에서 다양한 크기의 시맨틱 특징을 추출할 수 있도록 네트워크를 구성함.	[8]
CrowdNet	CrowdNet의 깊은 네트워크에서는 고차원의 시맨틱 특징을 추출하고자 하였고, 얇은 네트워크에서는 저차원의 시맨틱 특징을 추출하고자 함. CrowdNet에서는 UCF_CC_50 데이터에 대해 그 성능을 평가함.	[13]
Hydra CNN	Hydra CNN은 입력 패치의 피라미드를 사용한다. 피라미드의 각 층은 각기 다른 스케일을 나타내며, 히드라의 특정한 머리를 먹여 살림. 모든 헤드 출력물은 연결되어 히드라의 몸을 형성하는 완전히 연결된 층으로 전달됨.	[9]
Switch-CNN	Switch-CNN은 이미지 내에서 군중 밀도의 변동을 활용하여 예측된 군중 수의 품질과 지역화를 개선함. Independent CNN crowd density regressors는 주어진 군중 장면의 그리드에서 샘플링된 패치에 대해 훈련되며, 수용적 영역과 시야가 서로 다르도록 선택된다. 이를 통해 각 CNN crowd density regressors가 학습한 기능이 특정 척도에 맞게 조정됨.	[10]
CSRnet	컴퓨터 비전 분야에서 이미지 분류, 이미지 탐지 등에 특징 추출기로 널리 사용되고 있는 VGGNet에 기반한 밀도 지도 추정 방법을 제안하고 있음. Shanghaitech, UCF_CC_50, UCSD, WorldExpo'10, TRANSCO 데이터를 이용해 성능을 평가함.	[11]
TEDnet	고품질 밀도 추정 맵을 생성하는 데 초점을 맞춘 Crowd Counting을 위한 trellis encoder-decoder network (TEDnet).	[12]
ANF	Crowd Counting을 위한 the attentional neural field (ANF) 제안함. encoder-decoder 네트워크 내에서, 다중 스케일 기능을 집계하기 위해 조건부 랜덤 필드(CRF)를 도입하여 보다 유용한 묘사를 만들 수 있음.	[14]
DM-Count	DM-Count에서는 OT(Optimal Transport)를 사용하여 정규화된 예측 밀도 맵과 정규화된 지상 진실 밀도 맵 사이의 유사성을 측정함. OT 계산을 안정화하기 위해 모델에 총 변동 손실을 포함. Data set은 UCF-QNRF, NWPU, ShanghaiTech, UCF-CC50사용하여 성능을 평가함.	[2]

Estimation-based counting 기반 모델의 성능 비교

Model Name	Year	ShanghaiTech A		UCF CC 50	
		MAE	RMSE	MAE	RMSE
MCNN	2016	110.2	173.2	377.6	509.1
CrowdNet	2016	181.8	277.7	467.0	498.5
Hydra CNN	2016	-	-	333.7	425.3
Switch-CNN	2017	90.4	135.0	318.1	439.2
CSRnet	2018	68.2	115.0	266.1	397.5
TEDnet	2019	64.2	109.1	249.4	354.5
ANF	2019	63.9	99.4	250.2	340.0
DM-Count	2020	59.7	95.7	211.0	291.5

Crowd Counting 사용 Data 분류

1) UCSD

해당 데이터는 캠퍼스의 감시카메라에서 얻어진 2,000 장의 프레임으로 이루어져 있습니다. 군중 집계 초기 데이터로 군중의 밀도가 높지 않은 특징을 갖고 있습니다. Peds1는 카메라를 향해 걸어가거나 카메라에서 멀어지는 그룹의 사람들의 클립과 약간의 원근 왜곡, 34개의 교육 비디오 샘플과 36개의 테스트 비디오 샘플이 포함되어 있습니다. Peds2는 카메라 평면에 평행한 보행자 움직임이 있는 장면, 16개의 교육 비디오 샘플과 12개의 테스트 비디오 샘플이 포함되어 있습니다.

Fig. 1. UCSD 데이터 예시 사진



2) The Mall Dataset

Mall 보행자 데이터 세트는 Chen에 의해 소개되었으며 320 x 240 크기의 주석이 달린 2000개의 프레임을 포함하고 있습니다. 쇼핑몰에 설치된 감시카메라를 이용해 다른 조건의 채광과 군중밀도를 채집했습니다. 이 데이터 세트는 쇼핑 센터의 보행자 교통 및 클러스터링 프로세스를 기록하며, 하나의 이미지에 적은 수의 개체를 포함하고 있습니다.

Fig. 2. The Mall 데이터 예시 사진



3) Shanghaitech

ShanghaiTech 데이터 세트는 군중의 밀도에 따라 A, B 두 타입으로 분류되는 데이터로 1198개의 이미지와 330,165개의 주식 헤드를 포함하는 대규모 인구 통계 데이터 세트입니다. 이 데이터는 파트 A와 파트 B의 두 부분으로 이루어져 있습니다. A는 인터넷에서 무작위로 뽑은 482개의 이미지로 구성되어 있고, B는 상하이 메트로폴리탄 거리의 이미지입니다.

Fig. 3. ShanghaiTech A 데이터 예시 사진



4) UCF_CC_50

UCF_CC_50 데이터 세트에는 서로 다른 해상도의 총 50개 이미지가 포함되어 있으며, 이미지당 평균 1280명이 있습니다. 훈련과 테스트를 위한 cross-validation protocol을 정의했으며, 여기서 데이터 세트는 10개의 그룹으로 나뉘고 5개의 교차 검증을 수행해야 합니다. 이러한 기능 때문에 데이터 세트를 처리하기가 매우 어려움이 있습니다.

Fig. 4. UCF_CC_50 데이터 예시 사진



5) UCF-QNRF

UCF-QNRF는 가장 많은 수의 군중 이미지를 가진 새로운 대규모 데이터 세트입니다. 이 데이터 세트의 이미지에 있는 군중도 밀도가 높지만 자세한 감독 정보와 주석으로 인해 이 데이터 세트는 모델을 교육할 때, 좋은 역할을 합니다.

Fig. 5. UCF-QNRF 데이터 예시 사진



6) WorldExpo10

WorldExpo10 데이터 세트는 2010년 상하이 세계 엑스포에서 108개의 감시 카메라에 캡처된 1,132개의 주석이 달린 비디오 시퀀스가 포함되어 있습니다. 훈련 세트에는 103개 장면의 1분 비디오 시퀀스 1127개와 5개의 다른 장면의 1시간 비디오 시퀀스 5개가 포함된 테스트 세트가 포함됩니다. 각 테스트 장면은 120개의 라벨이 붙은 프레임으로 구성되며, 군중 번호는 1에서 220까지 6605입니다. 다른 Data set과 달리, 이 Data set은 공개 다운로드 접근 방식을 제공하지 않으며, 연구자는 학술적 및 비상업적 다운로드 목적으로 데이터 세트 작성자에게 문의해야 합니다.

Fig. 6. WorldExpo10 데이터 예시 사진



7) NWPU

NWPU 5,109개의 이미지와 총 2,133,238개의 주석이 달린 인스턴스가 포함되어 있습니다. 타 Data set과 비교하여 데이터 양 이외에도 negative samples, 공정한 평가, 고해상도 및 큰 모양 변화 등 장점을 가지고 있습니다.

Fig. 7. NWPU 데이터 예시 사진



참고 문헌

- [1] Luo, Ying, Jinhua Lu, and Baochang Zhang. "Crowd counting for static images: a survey of methodology." *2020 39th Chinese control conference (CCC)*. IEEE, 2020.
- [2] Wang, Boyu, et al. "Distribution matching for crowd counting." arXiv preprint arXiv:2009.13077 (2020).
- [3] 이장호, 윤성로. (2018). 딥러닝을 이용한 군중 집계 동향 분석. 대한전자공학회 학술대회, 735-737.
- [4] Boominathan, et al. "Crowdnet: A deep convolutional network for dense crowd counting." *ACMMM*, 2016.
- [5] Sindagi, et al. "Generating high-quality crowd density maps using contextual pyramid cnns." *ICCV*, 2017
- [6] Li, et al. "CSRNet: Dilated convolutional neural networks for understanding the highly congested scenes." *CVPR*, 2018.
- [7] D. Ryan, S. Denman, S. Sridharan, and C. Fookes, "An evaluation of crowd counting methods, features and regression models," *Comput. Vis. Image Underst.*, vol. 130, pp. 1–17, 2015.
- [8] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, "Single-image crowd counting via multi-column convolutional neural network," in *CVPR*, 2016, pp. 589–597
- [9] D. Onoro-Rubio and R. J. Lopez-Sastre, "Towards perspective-free object counting with deep learning," in *ECCV*. Springer, 2016, pp. 615–629.
- [10] D. B. Sam, S. Surya, and R. V. Babu, "Switching convolutional neural network for crowd counting," in *CVPR*. IEEE, 2017, pp. 4031–4039.
- [11] Y. Li, X. Zhang, and D. Chen, "Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes," in *CVPR*, 2018, pp. 1091–1100.
- [12] X. Jiang, Z. Xiao, B. Zhang, X. Zhen, X. Cao, D. Doermann, and L. Shao, "Crowd counting and density estimation by trellis encoderdecoder network," *CVPR*, 2019.
- [13] Boominathan, et al. "Crowdnet: A deep convolutional network for dense crowd counting." *ACMMM*, 2016.
- [14] A. Zhang, L. Yue, J. Shen, F. Zhu, X. Zhen, X. Cao, and L. Shao, "Attentional neural fields for crowd counting," in *ICCV*, 2019, pp. 5714–5713.