

# 딥러닝 기반 군중 계수 모델 개발 및 연구

장성욱, 박주현, 송근재, 므드 파이살 아흐메드 부이안, 라크모노브 아크로존, 김정홍  
경북대학교

jeff4556@knu.ac.kr, juheona@gmail.com, skj2491@naver.com,  
faysalahmedtushan@gmail.com, r.akhror@knu.ac.kr, jhk@knu.ac.kr

## A Study on the Development and Research of a Deep Learning-based Crowd Counting Model

Jang SeongUk, Park Juheon, Song Geunjae, Md Faysal Ahmed Bhuiyan,  
Rakhmonov Akhrorjon, Kim JeongHong  
Kyungpook Nat'l Univ.

### 요약

최근 발생한 다중인파밀집 사고들로 공공장소의 안전 보장과 대규모 행사의 효율적 관리를 위한 군중 관리 기술의 중요성이 대두되었다. 기존의 군중 계수 기술은 밀집된 군중 이미지에서 발생하는 중첩과 원근에 의한 변화의 구분이 어려웠다. 본 연구에서는 명시적인 감지 절차 없이 이미지로부터 인구 밀도 지도를 추정할 수 있는 딥러닝 모델을 제안하였다. 이 모델은 공간 피라미드 풀링(Spatial Pyramid Pooling, SPP)을 적용하여 다양한 스케일의 이미지 특징을 효과적으로 추출하고 통합한다. 이렇게 추출된 특징을 활용하여, 모델은 주어진 훈련 이미지와 해당하는 ground-truth 밀도 지도 간의 비선형 매핑을 학습하여 밀도 지도를 예측한다. 제안한 모델의 성능 평가를 위해 평균 절대 오차(MAE)와 평균 제곱근 오차(RMSE) 지표를 사용하였다. 제안한 모델은 기존 모델에 비해 두 오차 지표에서 각각 35%, 29%의 감소를 보였다. 이 실험 결과는 모델의 성능이 우수함을 나타내며, 군중 계수 모델의 개선 방향을 제시하였다..

### I. 서론

최근 서울과 런던에서 발생한 대형 사고들이 군중 관리의 중요성을 조명하였다. 이러한 다중인파밀집 사고들은 공공장소에서의 안전과 대규모 행사의 효율적 관리를 위해서 보다 정확한 군중 계수에 대한 연구가 필요하다. 기존의 군중 계수 기술은 밀집된 환경에서 개인 간의 중첩과 가려짐, 원근에 의한 인물의 크기 변화 등으로 인해 정확성에 한계를 보였다[1].

본 논문에서는 기존 기술의 한계를 극복하고자 딥러닝을 활용한 새로운 군중 계수 모델의 개발을 다룬다. 이 모델은 명시적인 감지 없이 군중의 밀도를 추정하고, 다중 스케일 문맥 정보를 통합하여 이미지에서 인구 밀도 지도를 생성한다. 본 연구에서는 N 개의 훈련 이미지와 그에 해당하는 ground-truth 밀도 지도를 사용하여 입력 이미지를 예측된 밀도 지도로 매핑하는 비선형 매핑을 학습한다[2].

이 연구가 제안하는 고도화된 군중 분석 기술은 군중 계수의 정확도를 향상시키고, 대규모 행사 및 공공장소에서의 안전 관리를 개선한다. 또한 잠재적인 위험 상황을 예방하고 보행자 동선 및 도시 교통 계획에 대한 새로운 통찰을 제공함으로써 사회적 및 경제적 효과를 가져올 것으로 기대된다.

### II. 본론

[2]에서는 고도로 맥락화된 군중 계수를 목표로 하며, 복잡한 환경에서의 군중 밀도 추정을 위한 다양한

기술적 접근법을 제시했다. 본 논문에서는 다음과 같이 이러한 접근법에서의 모델을 참고하여, 이를 기반으로 성능 개선을 위한 실험을 수행하였다.

Figure 1 은 본 연구에서 사용된 모델의 아키텍처를 보여준다.

첫 번째 단계는 이미지로부터 기본적인 시각적 특징을 추출하는 것이다. 이를 위해 사전 훈련된 VGG-16 네트워크의 첫 10 개 계층을 사용하여, 이미지의 기본적인 특징을 추출한다. 이후, 공간 피라미드 풀링(Spatial Pyramid Pooling, SPP) 기법을 적용하여 다양한 스케일에서의 정보를 효과적으로 포착하고, 모델이 이미지의 해상도를 가리지 않고 작동할 수 있도록 하였다.

두 번째로, 추출된 특징은 이중선형 보간(bilinear interpolation)을 사용하여 원래의 크기로 업샘플링된다. 이 특징 맵에서 기존에 추출한 특징과의 차이를 계산하여 대비 특징을 생성하고, 이 대비 특징을 다음 단계의 네트워크에 입력하여 각 스케일에 대한 가중치 맵을 학습한다.

세 번째로, 앞서 구한 대비 특징과 SPP로부터의 스케일 가중치를 조합하여 최종적으로 가중치가 적용된 특징 맵을 생성한다. 이 가중치 맵은 원본 이미지 특징 맵과 연결되어 백엔드 디코더의 입력으로 사용된다.

마지막으로, 백엔드 디코더는 여러 단계의 확장 컨볼루션(dilated convolution)을 포함하며, 이를 통해 최종적인 밀도 지도를 생성한다. 이 밀도 지도는 실제 이미지 내의 개별 인원 수를 나타내며, 정확도를 높이기 위해 여러 크기와 스케일의 특징을 효과적으로 통합한다.

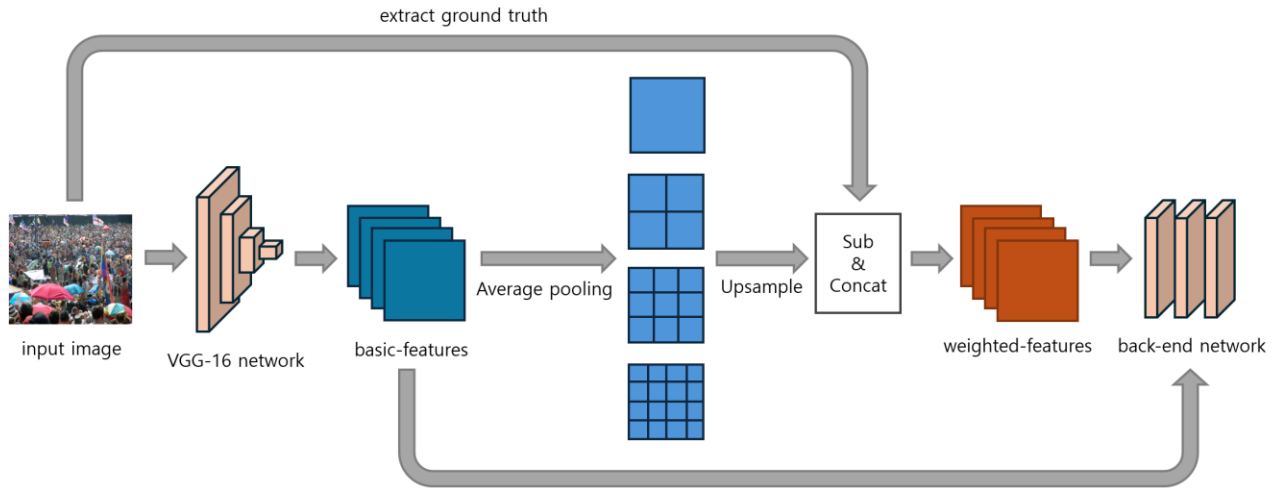


Figure 1. 제안된 모델의 아키텍처 모식도

마지막으로, 백엔드 디코더는 여러 단계의 확장 컨볼루션(dilated convolution)을 포함하며, 이를 통해 최종적인 밀도 지도를 생성한다. 이 밀도 지도는 실제 이미지 내의 개별 인원 수를 나타내며, 정확도를 높이기 위해 여러 크기와 스케일의 특징을 효과적으로 통합한다.

본 연구에서는 상기한 모델을 기반으로 다양한 하이퍼파라미터와 활성화 함수의 조합을 실험하였다. 모델의 학습률(learning rate), 배치 크기(batch size), 가중치 감소율(decay), 세대(epochs), 그리고 활성화 함수를 변경하면서 각 조합이 모델의 성능에 미치는 영향을 평가하였다.

본 연구에서는 [3]에서 제시한 모델과, 시험 모델 중 LeakyReLU 활성화 함수를 사용한 모델 간의 성능을 비교하였다. 각 설정에서 모델의 성능 평가 척도로 Mean Absolute Error(MAE)와 Root Mean Squared Error(RMSE)를 사용하였다.

본 연구의 실험에서는 [2]에서 제공한 데이터셋을 기반으로, 위에서 제시한 모델을 다양한 설정 하에서 훈련시켰으며, 훈련 데이터셋과 평가 데이터셋은 서로 다른 이미지와 truth map 으로 구성되어 모델의 일반화 능력을 검증할 수 있도록 하였다. 이를 기반으로 훈련된 모델과, 기존에 제시된 모델 사이의 성능 차이를 비교하였으며, 실험 결과는 Table 1 으로 나타내었다.

Table 1. 기존 모델과 시험 개발 모델 간 성능 비교

Model	Performance	
	MAE	RMSE
Zhang et al. [3]	181.8	277.7
<b>Ours</b>	<b>118.7</b>	<b>198.2</b>

실험 결과, 실험 모델의 MAE, RMSE 양 지표가 기존 모델의 지표에 비해 상당한 감소를 보였으며, 이는 실험 모델이 기존 모델에 비해 더 강력한 인구 밀도 추정 성능을 보였으며, 본 연구의 접근법이 군중 밀도 추정 문제에 있어서 어떠한 개선점을 제공할 수 있는지를 시사한다.

### III. 결론

본 연구는 최근 발생한 다중인파밀집사고를 계기로 공공 장소에서의 군중 관리의 중요성을 재조명하였다. 이를 바탕으로, 기존 군중 계수 기술의 한계를 극복하기 위해 딥러닝 기반의 새로운 군중 계수 모델을 개발하였다. 이 모델은 다중 스케일 문맥 정보를 통합하여 이미지에서 보다 정확한 인구 밀도 지도를 생성할 수 있으며, 다양한 하이퍼파라미터의 조정을 통해 성능을 향상시킬 수 있었다. 제안한 모델의 성능 평가를 위해 평균 절대 오차(MAE)와 평균 제곱근 오차(RMSE) 지표를 사용하였으며, 기존 모델에 비해 두 오차 지표에서 우수한 성능을 확인하였다.

제안한 모델은 군중 계수의 정확도를 높이고, 대규모 행사 및 공공 장소에서의 안전 관리를 개선하는 데 기여할 것으로 기대된다.

### ACKNOWLEDGMENT

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업의 연구결과로 수행되었음"(2021-0-01082)

### 참 고 문 헌

- [1] Xu, B., & Qiu, G. "Crowd density estimation based on rich features and random projection forest." IEEE Winter Conference on Applications of Computer Vision. 2016.
- [2] Liu, Weizhe, et al. "Context-Aware Crowd Counting." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.
- [3] Cong Zhang, Hongsheng Li, Xiaogang Wang, and Xiaokang Yang. Cross-Scene Crowd Counting via Deep Convolutional Neural Networks. In Conference on Computer Vision and Pattern Recognition, pages 833-841, 2015.