Integrating Convolutional and Graph Neural Networks for Enhanced Dense Crowd Counting

Author: TREFAULT Romain
Supervisor: Pr. BOUTRY Nicolas

École pour l'informatique et les techniques avancées (EPITA)

July 1, 2024

Outline

Introduction

- Background and Motivation
  - Importance of accurate crowd counting
  - Applications: event management, public safety, urban planning, surveillance, etc.

Figure: Image of a Crowd Scene. [1]

- Challenges in Crowd Counting
  - Occlusion, scale variation, density variation, perspective distortion

Figure: Occlusion, Scale and Density Variations, Perspective Distortion Crowd Scenes. [2]

Detection-based Methods

- **Monolithic Detection**
  - Trains classifier using full-body appearance.
  - Uses features like Histogram of Oriented Gradients (HOG).
  - Employs Support Vector Machines (SVMs) and Random Forests with a sliding window approach.
  - Effective for sparse crowds, limited for dense crowds.
- **Part-based Detection**
  - Focuses on parts like the head and shoulders.
  - Combines head and shoulders for more reliable detection.
  - More effective for dense crowds.

Regression-based Methods

- Avoid segmentation or individual tracking.
- **Approach**
  - Extract low-level features: edges, foreground pixels.
  - Apply regression modeling to map features to count.
  - Uses total area and texture of foreground pixels to provide direct estimation of crowd size.

Density-based Methods

- **Approach**
  - Estimate density, incorporating spatial information.
  - Map local features to density maps, tracking groups of individuals.
  - Creates separate forests for crowded and less crowded areas.
- **Advantages**
  - Addresses occlusion and clutter.
  - Provides spatial distribution of individuals.

Deep Learning Approaches

- **CNN-based Models**
  - **Basic CNNs**
    - Initial methodologies with fundamental convolutional layers.
  - **Scale-aware Models**
    - Multi-column designs.
  - **Context-aware Models**
    - Integrate local and global contextual information.
- **Examples of CNN-based Approaches**
  - Wang et al. [3]: CNN regression model using AlexNet.
  - Zhang et al. [4]: Multi-Column CNN for adaptive learning.
  - Boominathan et al. [5]: Hybrid deep and shallow networks.

Limitations of Existing Methods

- **Handling Varying Head Scales**
  - Difficulty in accurately counting individuals with different head sizes due to perspective effects.
- **Scene Variations**
  - Inconsistencies in different environments and lighting conditions.

CNN-GNN Combination

- **Combined Model Architecture**
  - CNN backbone for feature extraction
  - Graph construction and GNN processing

Proposed Approach

- **CNN for Spatial Feature Extraction**
  - U-Net architecture for segmentation

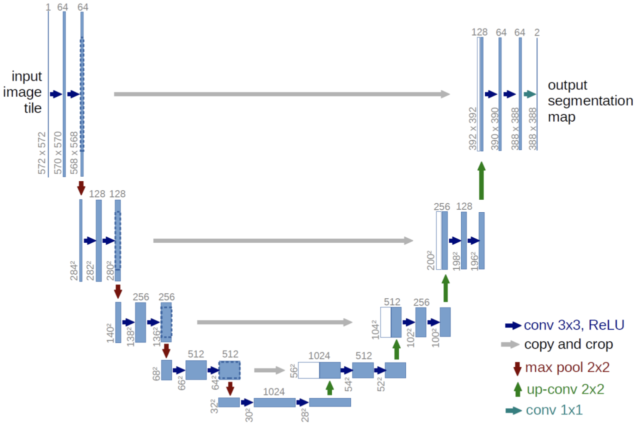

Figure: UNet Architecture. Source: [8]

- **GNN for Structural Information Processing**
  - Graph construction from CNN embeddings
  - Graph convolutional layers for message passing

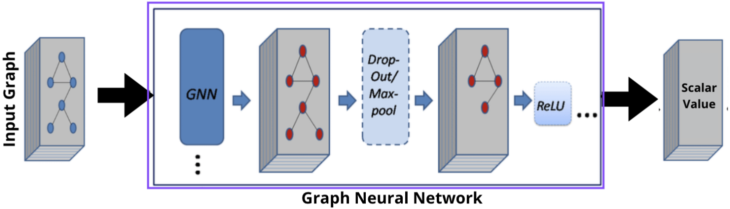

Figure: GNN Architecture.

Experiments

- Dataset: JHU-CROWD++
  - Diverse crowd densities and environmental conditions

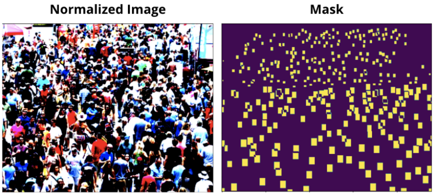

Figure: Sample of the CNN Dataset.

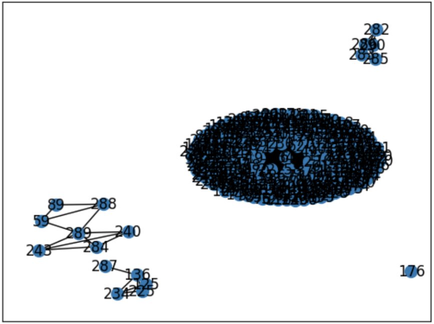

Figure: Sample of the GNN Dataset.

Training Procedures and Hyperparameters

Category	**Hyperparameter**	**Value**
CNN	Optimizer	Adam (lr=1e-4)
	Scheduler	ReduceLROnPlateau (patience=10)
	Loss Functions	Binary Cross-Entropy (BCE), Dice Loss
	Epochs	1000
	Evaluation Metric	Dice Score
GNN	Optimizer	Adam (lr=0.001)
	Loss Function	Mean Squared Error (MSE)
	Epochs	200
	Evaluation Metric	Mean Absolute Error (MAE)

Table: Training Procedures and Hyperparameters

CNN Component

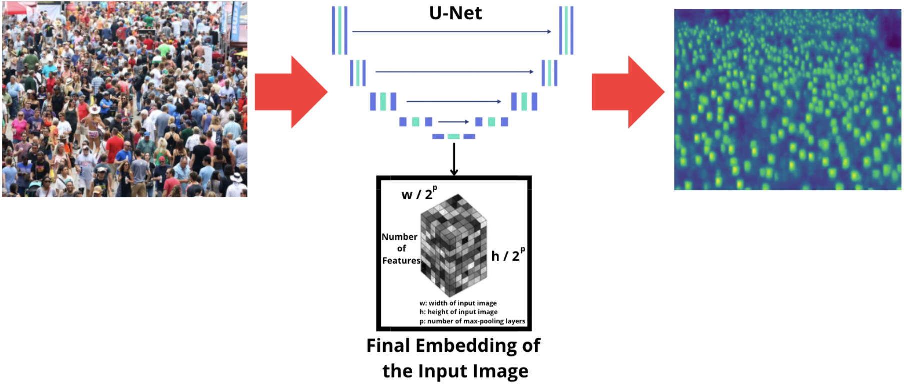

Figure: CNN Component of the Framework.

GNN Component

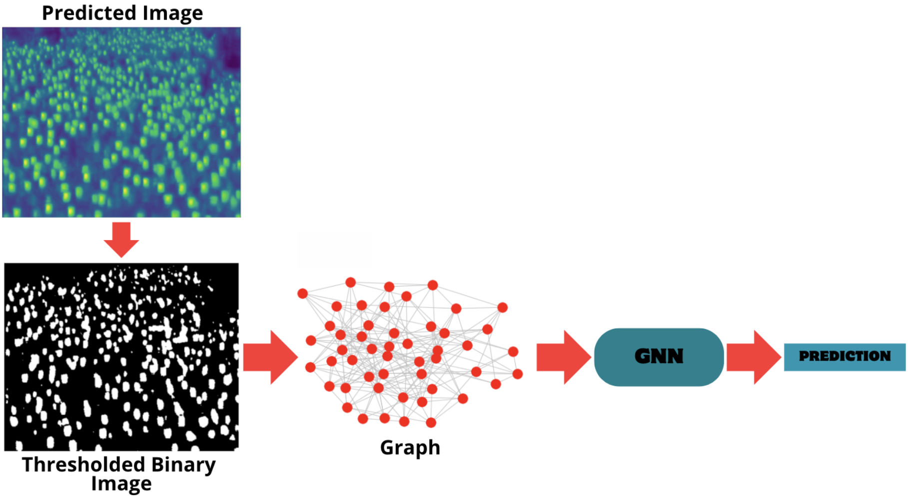

Figure: GNN Component of the Framework.

Results

```
Number of connected components w/o erosion (CNN):  tensor([18, 11, 36, 18, 35, 21,  8, 64])
Number of connected components w erosion (CNN):  tensor([16, 10, 35, 18, 35, 17,  7, 66])
Actual output:  tensor([ 48.,  17.,  86.,  25.,  36.,  54.,  23., 127.])
Predicted output:  tensor([  68.3502,   34.2540,   72.8633,  364.8552,   66.2278,   32.4979,
        3120.0261,  139.6208])
```

Figure: Quantitative Results and CNN Result Image.

Discussion

- **Limitations and Future Work**
  - **Post-processing Improvements**
    - Addressing issues with erosion that may remove small components.
  - **Shape Handling**
    - Better representation of true oval shapes of faces.
    - Mitigating shape deformation from image resizing.
  - **Hyperparameters**
    - Examining hyperparameters and kernel sizes.

Conclusion

- **Key Points**
  - Proposed approach: combining CNNs and GNNs for crowd counting
  - Findings: local spatial features and global structural patterns
- **Contributions and Implications**
  - Impact on crowd management and surveillance
  - Future research directions:
    - Post-processing improvements
    - Handling ovular shapes and mitigating image resizing effects
    - Hyperparameter sensitivity analysis

References I

- Chan, A. B., & Vasconcelos, N. (2008). Counting People With Low-Level Features and Bayesian Regression. *IEEE Transactions on Image Processing*.
- Pham, V.-Q., Kozakaya, T., Yamaguchi, O., & Okada, R. (2015). COUNT Forest: CO-Voting Uncertain Number of Targets Using Random Forest for Crowd Density Estimation.
- Wang, C., Zhang, H., Yang, L., Liu, S., & Cao, X. (2018). Deep People Counting in Extremely Dense Crowds. *State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences*.
- Zhang, Y., Zhou, D., & Ma, Y. (2016). Single-Image Crowd Counting via Multi-Column Convolutional Neural Network. Published in *Computer Vision and Pattern Recognition*.

References II

- Boominathan, L., Kruthiventi, S. S. S., & Babu, R. V. (2016). CrowdNet: A Deep Convolutional Network for Dense Crowd Counting. *Video Analytics Lab, Indian Institute of Science, Bangalore, India*.
- Oñoro-Rubio, D., & López-Sastre, R. J. (2016). Towards Perspective-Free Object Counting with Deep Learning.
- Han, K., Wang, Y., Guo, J., Tang, Y., Wu, E. (2022). Vision GNN: An Image is Worth Graph of Nodes.
- Lu, Y., Chen, Y., Zhao, D., Liu, B., Lai, Z., & Chen, J. (2020). CNN-G: Convolutional Neural Network Combined with Graph for Image Segmentation with Theoretical Analysis.
- Luo, A., Yang, F., Li, X., Nie, D., Jiao, Z., Zhou, S., Cheng, H. (2020). Hybrid Graph Neural Networks for Crowd Counting.
- Peng, F., Lu, W., Tan, W., Qi, K., Zhang, X., & Zhu, Q. (2022). Multi-Output Network Combining GNN and CNN for Remote Sensing Scene Classification.

References III

- Advances in Convolution Neural Networks Based Crowd Counting and Density Estimation by Rafik Gouiaa, Moulay A. Akhloufi, and Mozhdeh Shahbazi (2022).
- Patwal, A., Diwakar, M., Tripathi, V., & Singh, P. (2023). Crowd counting analysis using deep learning: a critical review. *Department of CSE, Graphic Era Deemed to be University, Dehradun, Uttarakhand, India*.
- Sindagi, V. A., & Patel, V. M. (2017). A Survey of Recent Advances in CNN-based Single Image Crowd Counting and Density Estimation. *Dept. of Electrical and Computer Engineering, 94 Brett Road, Piscataway, NJ 08854, USA*.
- Loy, C. C., Chen, K., Gong, S., & Xiang, T. (2013). Crowd Counting and Profiling: Methodology and Evaluation.

Appendix

1. Sunarso, E. (2016, December 7). Crowd Management. Seasoned security professional with extensive experience in corporate & physical security operations & management across APAC & ME.
2. Nguyen, V., & Ngo Duc, T. (2020). Single-image crowd counting: a comparative survey on deep learning-based approaches. International Journal of Multimedia Information Retrieval, 9(11), 1-18. https://doi.org/10.1007/s13735-019-00181-y.