VISVESVARAYA TECHNOLOGICAL UNIVERSITY

"JNANA SANGAMA" BELAGAVI-590018, KARNATAKA

Project Synopsis
on

"CAT-CNN: Crowd counting with crowd attention convolutional neural network using Deep Learning"

Submitted in the partial fulfillment of the requirement of the Seventh Semester

Bachelor of Engineering
In
Computer Science & Engineering

CHINTAPARTHI RETISH REDDY	1BO19CS031
CH. MAHESWAR REDDY	1BO19CS027
ALDI ROHITH	1BO19CS004
BIJJULA JAYA SIMHA REDDY	1BO19CS025

Under The Guidance of
Padmavathi H G
Asst. Professor, Dept of CSE

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Brindavan College of Engineering

DWARAKANAGAR, BAGALUR MAIN ROAD, YELAHANKA, BANGALORE-63,

2022-23

AIM & OBJECTIVE

Aim:

The aim of the project is to propose a novel end-to end model called Crowd Attention Convolutional Neural Network (CAT-CNN) that can adaptively judge the position of a human head at each pixel location by automatically encoding a density map and count the number of Humans.

Objective:

Crowd counting is a challenging problem due to the scene complexity and scale variation. Although

Deep learning has achieved great improvement in crowd counting, scene complexity affects the judgment of these methods and they usually regard some objects as people mistakenly; causing potentially enormous errors in the crowd counting result.

1. The Crowd Dataset is collected from Machine Learning Repository
2. CAT-CNN that can adaptively assess the importance of a human head at each pixel location to avoid enormous misjudgments in crowd counting.
3. design a novel classification model that can take input of arbitrary size for training in crowd counting.
4. And we first explicitly map the prior information of the population-level category of images to feature maps to automatically contribute in encoding a highly refined density map.
5. Predicting the Human count based on the density map.

ABSTRACT

Counting objects in images is one of the fundamental tasks in deep learning . Currently, deep learning (DL) methods provide the state-of-the-art performance in digital image processing. However, they require collecting a lot of annotated data, which is usually time consuming and prone to labeling errors. Crowd counting is a challenging problem due to the scene complexity and scale variation. Although deep learning has achieved great improvement in crowd counting, scene complexity affects the judgment of these methods and they usually regard some objects as people mistakenly; causing potentially enormous errors in the crowd counting result. To address the problem, we propose a novel end-to end model called Crowd Attention Convolutional Neural Network (CAT-CNN). Our CAT-CNN can adaptively assess the importance of a human head at each pixel location by automatically encoding a confidence map. With the guidance of the confidence map, the position of human head in estimated density map gets more attention to encode the final density map, which can avoid enormous misjudgments effectively. The crowd count can be obtained by integrating the final density map. To encode a highly refined density map, the total crowd count of each image is classified in a designed classification task and we first explicitly map the prior of the population-level category to feature maps. To verify the efficiency of our proposed method, extensive experiments are conducted on three highly challenging datasets. Results establish the superiority of our method over many state-of-the-art methods.

INTRODUCTION

As the phenomenon of crowd congestion is becoming serious, safety- and security-oriented tasks– such as public safety control and traffic safety monitoring– face huge challenges. Manual analysis of the degree of crowd aggregation not only cannot achieve high accuracy but also will perform low efficiently. In contrast, deep-learning-based methods are more applicable at present since their process not only eliminates manual efforts but also can analyze crowd aggregation accurately and quickly. Among them, crowd estimation at the pixel level through the crowd distribution density maps has achieved tremendous progress. A crowd density map is a kind of image label that can reflect the distribution of crowd heads by processing the head coordinate value through Gaussian convolution.

As the convolutional layer and pooling layer of Convolutional Neural Network (CNN) strengthen the relationship between pattern recognition and the context in the image, the density estimation methods of CNN are with strong learning ability. They have achieved high accuracy in dense scenes. The accuracy of crowd counting mainly depends on the quality of the estimated density map which is limited by the image scale. Since the convolution kernel of CNN owns a static size, heads of dynamic scales will worsen the network's performance, resulting in misjudgments and missing judgments. To solve this problem, the common methods are as follows: (1) introducing a multicolumn structure to estimate the crowd of different scales]; (2) introducing the idea of dilated convolution in the field from image segmentation . This is a special convolution for extracting feature information of different scales, consisting of a 33 convolution kernel and a dilated parameter. By setting the dilated parameter to replace redundant branches of different sizes of convolution kernels, the computational cost of multiscale detection can be reduced; (3) applying different detection methods to regions of different scales in the image. To generate a high-quality density map, spatial continuity should be ensured during the generation process so that the adjacent pixels in the output density graph can transition smoothly.

Crowd counting by computer vision technology plays an important role in safety management, video surveillance, and urban planning. The method of crowd counting can be also extended to other applications, such as cell counting, animal counting, and vehicle counting. However, due to the severe occlusion, scale variation, and high density in the crowd scene, crowd counting is still a challenging task.

To address these problems, a lot of efforts have been done in previous works including detection based methods and regression-based methods. Detection-based methods usually detect the instances of each person with pre-trained detectors. In the sparse crowd scene, they count the crowd accurately, while their accuracies are downgraded in the congested scene. Regression-based methods regress the number of the crowd without detecting people. They implement an implicit mapping between low-level features and crowd counts. However, the location information of the crowd is omitted. So that many CNN-based methods with state-of-the-art results are proposed recently. Most of them map the image to a density map that is more robust than the hand-crafted features. The quantity and location of the crowd at each pixel location are recorded in the density map. The crowd count can be obtained by integrating the density map.

Although CNN-based methods have achieved significant success in crowd counting, we find an important problem that needs to be solved urgently. Due to the complexity of crowd scenes, CNN-based methods usually mistake some objects as the head of people. As shown in Fig. 1, there are no people inside the red box, however, MCNN regards the dense shrubberies as human heads by mistake, which results in enormous errors of crowd counting. To address the above problem, we propose a novel end to-end model called CAT-CNN. An overview of the proposed CAT-CNN, It contains four modules: Multi-information Handling Module, Confidence Module, Density Map Estimation Module, and Fusion Module. The Multi-information Handling Module is utilized to extract robust features for crowd counting. Motivated by [2, 45], we leverage different convolution kernels to encode the input image at the beginning, then we fuse rich hierarchies from different convolutional layers, which is significant for extracting multi-scale features. In addition, the total crowd count of each image is classified in a designed crowd count group classifier. To the best of our knowledge, we first explicitly map the weights of predicted class to feature maps to automatically contribute in encoding a highly refined density map. In the Confidence Module, we classify each pixel to obtain the probability of a human head at each pixel location to encode the confidence map. Unfortunately, the ground-truth confidence map is not provided in present crowd counting datasets. We propose a simple but effective way to obtain the ground-truth confidence map by pasting the ones template on a binary map. The intensive cost of manual labeling is saved. Meanwhile, to address the problem of unbalanced population distribution, we propose the weighted Binary Cross-Entropy Loss (BCELoss) to encode a robust confidence map for population distribution. In the Density Map Estimation Module, the estimated density map is encoded.

EXISTING SYSTEM (PROBLEM IDENTIFICATION)

Problem Statement

In recent years, crowd counting has drawn much attention and various methods have been proposed, especially in deep learning. Next, we will give these methods some introductions.

1. Traditional detection-based algorithms such as Haar wavelets , HOG , and LBP occupy an important position in early works.
2. Regression-based methods learn a mapping between high level features and crowd counts.
3. CNN-based methods
4. Image based Methods

PROPOSED SYSTEM (OBJECTIVE AND METHODOLOGY)

An overview of the proposed CAT-CNN is shown in Fig. 1. Our CAT-CNN is composed of three stages. The first stage contains the first module where the features which can automatically adapt different scales and different crowd count groups are extracted. The second stage consists of two modules in the middle to encode confidence map and estimated density map respectively. The third stage contains the final module. With the guidance of the confidence map, final density map is encoded from the estimated density map in this stage. Next, we will elaborate these modules in each stage.

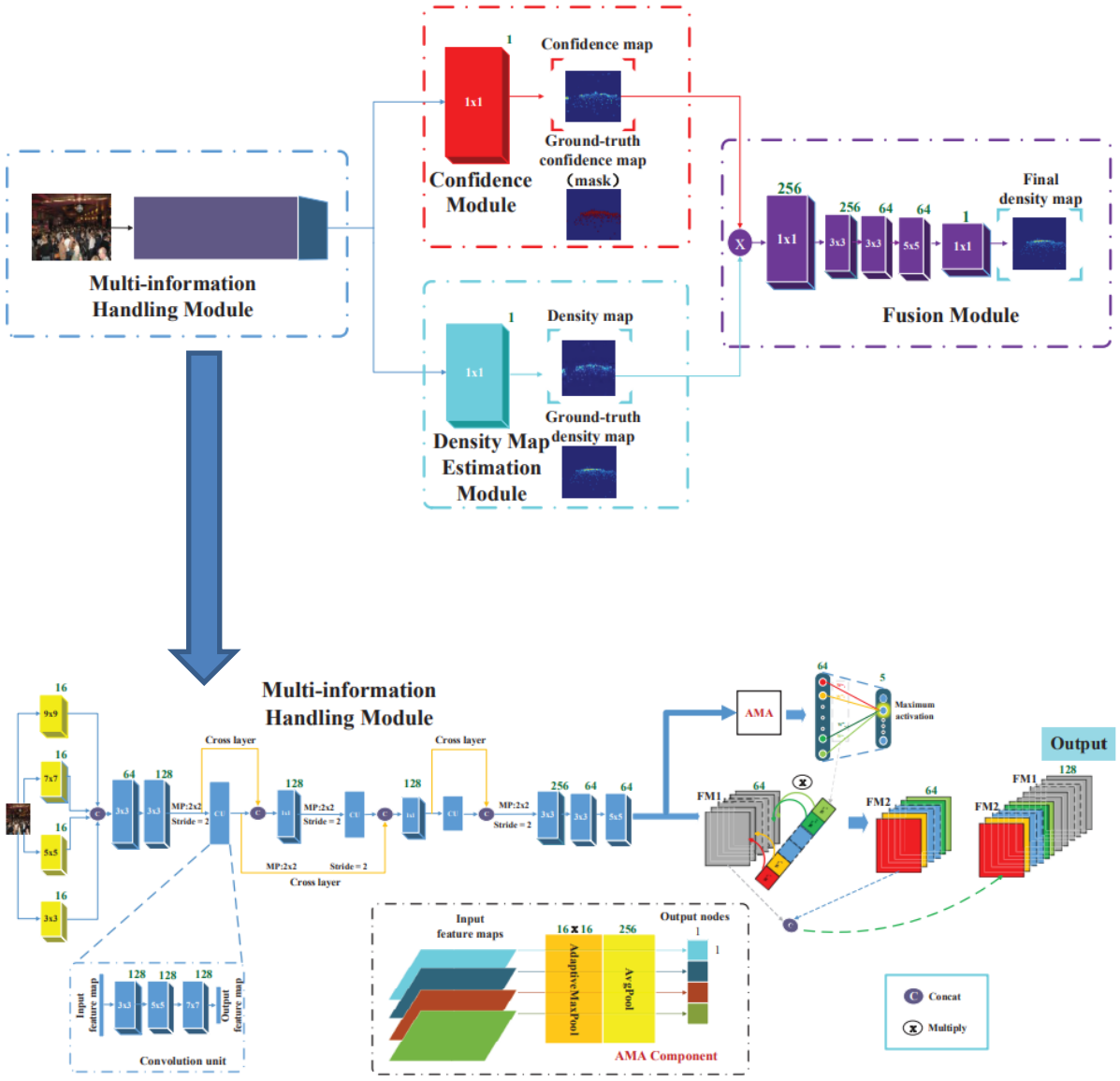

Figure 1: The proposed architecture of our CAT-CNN.

SYSTEM REQUIREMENT & SPECIFICATIONS

Requirements analysis is critical for project development. Requirements must be documented, actionable, measurable, testable and defined to a level of detail sufficient for system design. Requirements can be architectural, structural, behavioural, functional, and functional.

A software requirements specification (SRS) is a comprehensive description of the intended purpose and the environment for software under development.

Software Requirements

Scripting language : Python Programming

Scripting Tool : Anaconda Navigator (Jupyter Notebook) or Google Colab

Operating System : Microsoft Windows 8/ 10 or 11

Dataset : Crowd Dataset

Deep Learning Packages : Numpy, Pandas, Matplotlib , Seaborn Packages etc..

Hardware Requirements

Processor : 3.0 GHz and Above

Output Devices : Monitor (LCD)

Input Devices : Keyboard

Hard Disk : 1 TB

RAM : 16GB or Above

Graphics : 2GB or Higher

REFERENCES

1. Babu Sam, D., Sajjan, N.N., Venkatesh Babu, R., 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.
2. Boominathan, L., Kruthiventi, S.S., Babu, R.V., 2016. Crowdnet: A deep convolutional network for dense crowd counting, in: Proceedings of the 24th ACM international conference on Multimedia, ACM. pp. 640–644.
3. Chan, A.B., Liang, Z.S.J., Vasconcelos, N., 2008. Privacy preserving crowd monitoring: Counting people without people models or tracking, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, IEEE. pp. 1–7.
4. Chan, A.B., Vasconcelos, N., 2009. Bayesian poisson regression for crowd counting, in: Computer Vision, 2009 IEEE 12th International Conference on, IEEE. pp. 545–551.
5. Collobert, R., Weston, J., 2008. A unified architecture for natural language processing: Deep neural networks with multitask learning, in: Proceedings of the 25th international conference on Machine learning, ACM. pp. 160–167.
6. Dalal, N., Triggs, B., 2005. Histograms of oriented gradients for human detection, in: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, IEEE. pp. 886– 893.
7. Glorot, X., Bordes, A., Bengio, Y., 2011. Deep sparse rectifier neural networks, in: Proceedings of the fourteenth international conference on artificial intelligence and statistics, pp. 315–323.
8. He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, in: Proceedings of the IEEE international conference on computer vision, pp. 1026–1034.
9. He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.
10. Hossain, M., Hosseinzadeh, M., Chanda, O., Wang, Y., 2019. Crowd counting using scale-aware attention networks, in: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE. pp. 1280–1288.