Crowd Counting Deep-NCL

Goural (210393) , Electrical Engineering, IIT Kanpur

Professor: Dr. Tushar Sandhan Course: Intelligent Pattern Recognition (EE798R)

Abstract

This project presents an enhanced approach to crowd counting using a multi-scale attention mechanism within the Deep Negative Correlation Learning (D-ConvNet) framework. Crowd counting is a critical task with applications in surveillance, public safety, but it faces significant challenges, especially in densely populated scenes. Standard models often struggle with occlusions, scale variations, and density fluctuations within a single image, leading to inaccuracies in high-density areas.

We initially implemented the D-ConvNet model, which leverages decorrelated regressors through Negative Correlation Learning (NCL) to improve generalization. While this model showed promising results, it lacked adaptability to scale variations and density variability across different image regions. To address these limitations, we introduced a multi-scale attention mechanism that enables the model to focus on crowd regions adaptively, prioritizing areas with high density and adjusting for scale changes.

Multi-Scale Attention Mechanism

The multi-scale attention mechanism enhances the D-ConvNet-1 model by allowing it to adaptively focus on regions of varying crowd densities and scales within an image. This mechanism operates by applying attention at multiple scales, enabling the model to selectively prioritize features from different resolutions. Specifically:

•**Multi-Scale Attention Layers**: Attention modules are integrated at various levels of the network, allowing the model to adjust its focus across low, medium, and high-density regions. Each scale captures unique details that contribute to accurate density estimation.

•**Adaptive Weighting**: The mechanism dynamically assigns weights to features, enhancing focus on high-density areas while maintaining accuracy in sparser regions.

•**Improved Feature Representation**: By combining information from multiple scales, the model achieves a richer feature representation, making it more resilient to changes in crowd size and distribution across the image.

Improvements

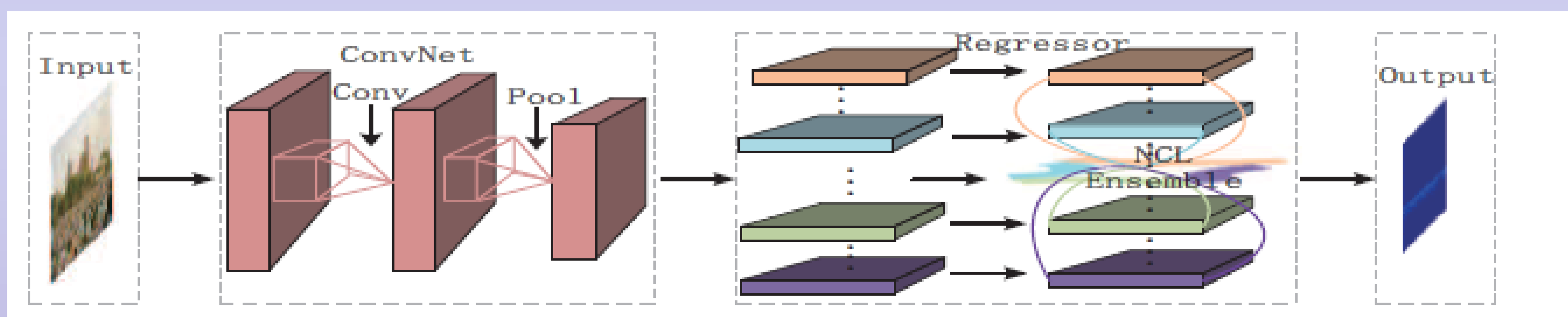
The integration of the multi-scale attention mechanism into the D-ConvNet-1 model led to significant improvements in crowd counting accuracy and adaptability:

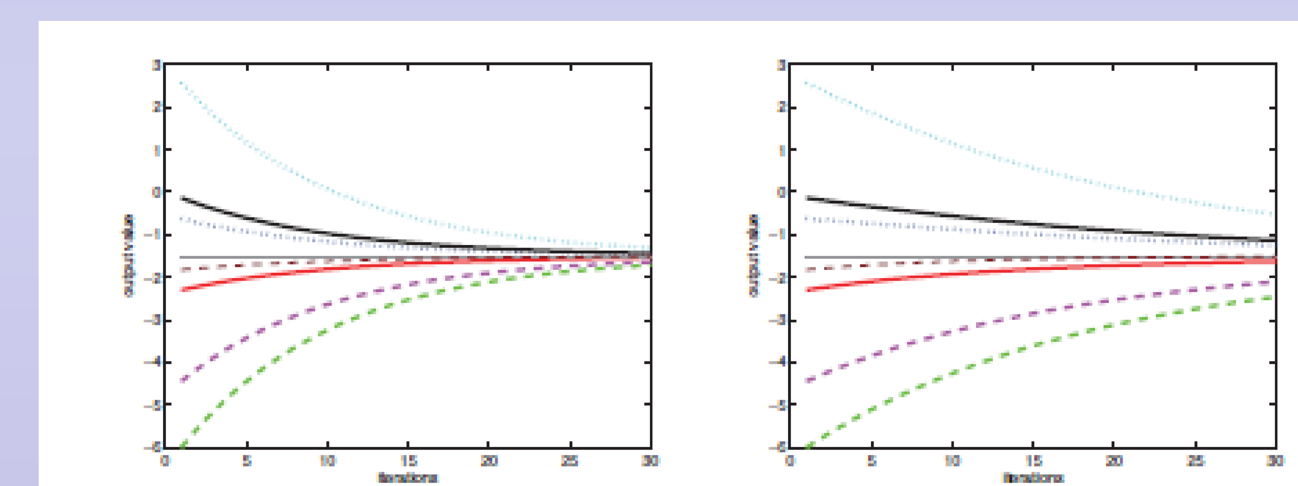
1.Enhanced Density Estimation: By applying attention across multiple scales, the model became more capable of identifying and accurately estimating high-density regions within an image.

2.Improved Scale-Invariance: Multi-scale attention enabled the model to adjust to varying scales of people within an image. This was particularly beneficial for images with people at different distances from the camera.

3.Reduced Overfitting: The multi-scale attention mechanism helped the model focus on relevant features, reducing noise from irrelevant regions.

4.Overall, these improvements were validated through lower Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) on benchmark datasets,





(a) Conventional ensemble learning (b) NCL

D-ConvNet-1

The D-ConvNet-1 model employs Deep Negative Correlation Learning (NCL) to enhance crowd counting by promoting diversity among regressors. Built on a modified VGG-16 backbone, D-ConvNet-1 extracts features up to the conv4_3 layer, followed by dilated convolutions to increase the receptive field without losing spatial details.

Architecture:

Feature Extractor: Uses VGG-16 layers up to conv4_3 with dilated convolutions, capturing detailed spatial information crucial for density estimation.

Regressors with Group Convolution: A set of group convolutions produces density maps for different image parts, ensuring spatial diversity.

Negative Correlation Learning: NCL penalizes correlated outputs among regressors, fostering unique predictions that improve model generalization.

.

Why Multi-Scale Attention Mechanism

The original D-ConvNet-1 model, while effective, had limitations in handling crowd density and scale variations across different regions of an image:

1.Density Variability: D-ConvNet-1 struggled to adjust its focus based on crowd density, often underperforming in highly dense areas. The multi-scale attention mechanism enables the model to give higher priority to crowded regions, improving accuracy where it's most needed.

2.Scale Adaptability: In images with people at varying distances, D-ConvNet-1 lacked the ability to handle scale differences effectively. Multi-scale attention addresses this by allowing the model to process features at different resolutions, capturing both close and distant individuals more accurately.

3.Enhanced Robustness: The adaptive nature of multi-scale attention allows the model to ignore irrelevant details and focus on essential crowd features, leading to better generalization in diverse crowd scenarios.

Conclusion

This project successfully improved the D-ConvNet-1 model for crowd counting by introducing a multi-scale attention mechanism, addressing key limitations in handling density variability and scale adaptation. Crowd counting in densely populated scenes remains a challenging problem due to overlapping individuals, perspective distortions, and fluctuating densities within a single image. The original D-ConvNet-1 model, though effective in promoting diversity among regressors, struggled with these challenges, particularly when crowd densities and scales varied significantly.

The multi-scale attention mechanism introduced here proved highly effective in overcoming these challenges. By allowing the model to selectively focus on relevant features at different scales, the attention mechanism enabled the model to accurately capture details across varying densities and distances within an image. This adaptability resulted in more precise density map predictions, as the model could dynamically prioritize high-density regions while also maintaining accuracy in sparser areas.