Single-Image Depth Estimation Based on Fourier Domain Analysis

By: Shiv Narayan (210978), IIT Kanpur

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

Perception & Intelligence Lab

Introduction

Single-Image depth estimation is crucial In computer vision enabling applications such as 3D reconstruction, autonomous navigation, and augmented reality. But estimating depth accurately from a single image is challenging due to the lack of depth information and motion data. Other approaches often struggle to capture both global depth distribution and local details.

Challenges

- **Depth Ambiguity**: Without stereoscopic data , distinguishing distances in a scene from a single image is ambiguous in itself.
- **Balancing Global and Local Depth**: Models fail to simultaneously capture both the broad depth structure and fine-grain details within the same image.

Objective

This work proposes a novel method that Combines CNN-based depth estimation with Fourier domain frequency analysis generating accurate depth predictions across different scales within an image. This addresses the inherent challenges of single-image depth estimation. We utilize the modified ResNet architecture and introduce the depth-balanced loss function to ensure consistent accuracy across varying depth ranges.

Methodology

1. CNN Architecture Based on ResNet-152

Our model is based on ResNet-152, modified to include Intermediate feature extraction paths, enabling it to capture multi-scale depth features. This improves generalization across both near and far depths in the scene.

2. Depth-Balanced Euclidean (DBE) Loss

- **Problem**: Standard Euclidean loss often biases toward distant objects, causing depth errors to be more significant for close objects.
- **Solution**: We introduce a **Depth-Balanced Euclidean (DBE) Loss** to balance the error across distances. DBE Loss weights the errors, ensuring reliable depth estimation for objects at varying distances within the same scene.

$$L_{\mathrm{DBE}} = \frac{1}{2N} \sum_{\mathbf{x}} \left(g(\hat{d}_{\mathbf{x}}) - g(d_{\mathbf{x}}) \right)^2$$

3. Generating Multiple Depth Map Candidates

- **Approach**: We create multiple depth map candidates by cropping the input image at different scales. Small crops capture local details, while large crops capture the overall depth distribution.

4. Fourier Domain Combination

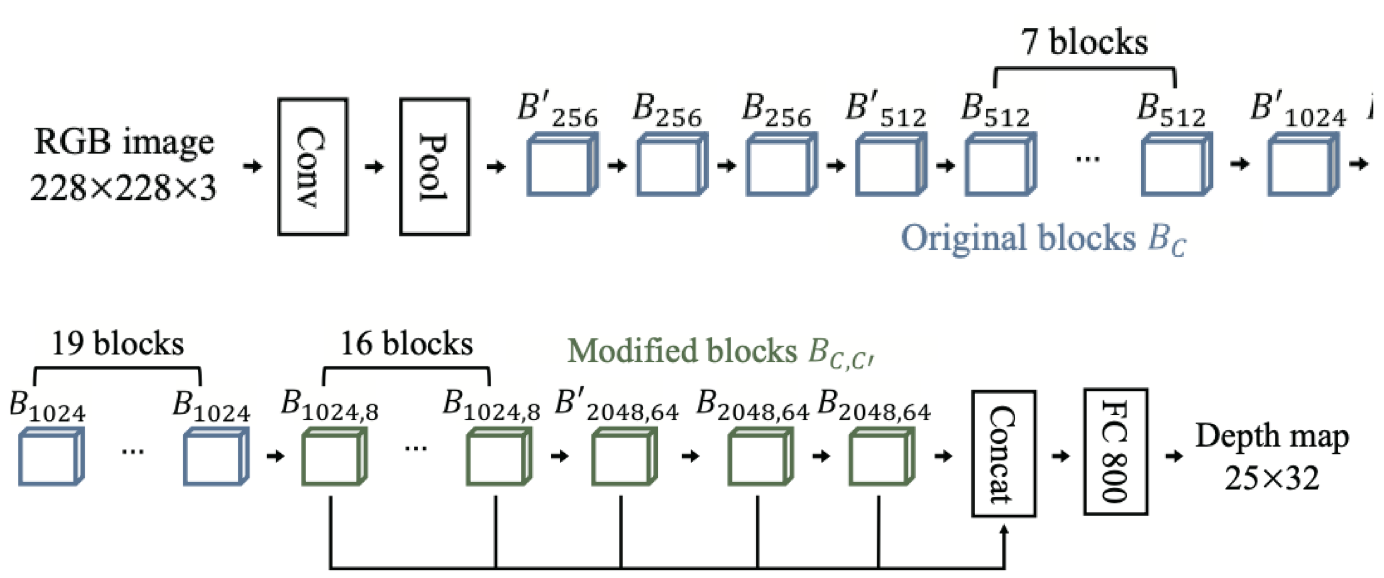
- **Technique**: We combine the depth map candidates in the Fourier domain. Low-frequency components capture the broad depth structure, while high frequencies emphasize fine details, resulting in a more comprehensive final depth map.

$$\hat{f}_k = \sum_{m=1}^{M} w_k^m (\hat{f}_k^m - b_k^m)$$

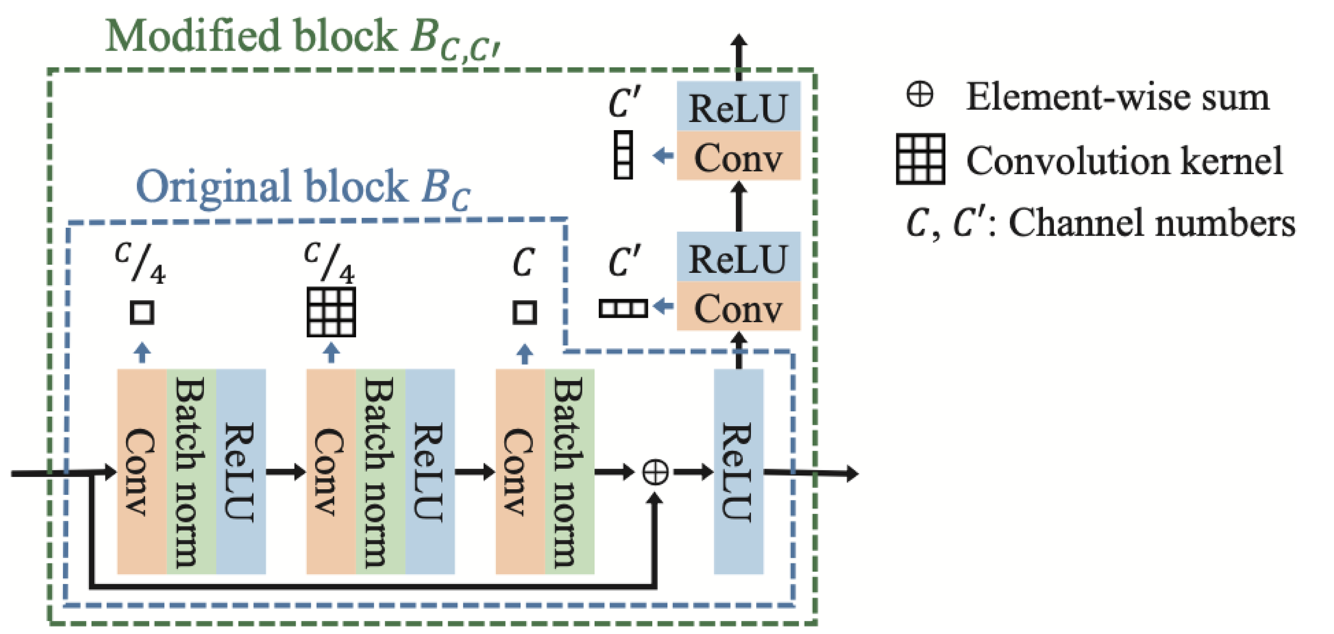
5. Adaptive DBE loss

We propose Adaptive DBE Loss dynamically adjusts a_1 and a_2 based on the mean and standard deviation of ground-truth depths within each mini-batch. This adaptation enhances model sensitivity to varied depth distributions and so improving prediction accuracy and generalization across diverse scenes.

Model Architecture



The Architecture



The detailed structures of original and modified blocks.

Conclusion

Our model achieves superior depth estimation by combining CNN predictions with Fourier domain analysis, effectively capturing both broad structure and fine details. The DBE Loss enhances accuracy across varied depths, and the Adaptive DBE Loss further improves robustness, achieving validation errors of **2.3812 MSE** and **2.3522 MSE**, respectively.