A Review On Single Image Depth Prediction with Wavelet Decomposition

Arya J Kumar [1], Prof. Parvathi V S [2]

[1]*PG Student, Dept. of Electronics & Communication Engineering, LBSITW, Kerala, India*
[2] *Assistant Professor, Dept. of Electronics & Communication Engineering, LBSITW, Kerala, India*

Abstract - *Wavelet decomposition method can predict accurate depth from single RGB image. Wavelet based method to reduce computational complexity for monocular depth estimation compare this with other methods, this supervise only final depth image reconstruct through inverse discrete transform. In this method can cut down number of necessary operation in the decoder by half while producing a drop in accuracy of less than 2%.More specifically goal is to reconstruct full resolution depth map from low resolution estimate which is progressively upsampled and refined using wavelet coefficient projection at increasing scale. It present new way of reducing computational complexity in image to image which is complementary with efficient seeking method.*

***Key Words*: Monocular Depth Estimation, Wavelet Decomposition, Inverse Discrete Transform , Resolution**

1. INTRODUCTION

Single image depth estimation methods are used in the field of robotics, autonomous driving, augment reality,etc.3D structure environment captures from single image is a studied problem in computer vision. Accurate depth is important for 3D reconstruction .Prediction time is important for robotics, augment reality and autonomous driving.

In this work introduce wavelet monodepth .Wavelet based method to reduce computational complexity for monocular depth estimation.This method can cut down number of necessary operation in the decoder by half while producing a drop in accuracy of less than 2%. Monocular depth estimation method usually train a neural network to predict dense depth maps from single RGB image. Did it so the typically employ a new net encoder decoder architecture.The input image is first processed by encoder which produce feature map at multiple scales. This feature map then fetch to decoder which typically alternate up sampling and convolutional operation in order to be full resolution depth estimation.

Depth map typically comprise of move or flat region together with depth edges .There edges are typically more informative than flat regions as the linear objects in an important geometry in the scenes. This depth edges corresponding region is strong depth region which are usually sparse in lateral scenes. Standard decoder usually runs convolutional kernels at every pixel location for every scales. How ever this can be very expensive especially for high resolution images furthermore ,this is highly in efficient as most of the decoder operations are used for less for even regions. This method implemented with sparse or dense convolution with the baselines. KITTI and NYU are used as the datasets.

2. REVIEW ON RELATED WORK

Qiufuli *et al.* [1] proposed pooling, strided convolution and average pooling in CNNs is replaced by discrete wavelet transform. Discrete wavelet transform and inverse discrete wavelet transform layer are applicable for various wavelets. Wavelet integrated CNNs are designed for image classification. Down sampling, feature maps are decomposed into high frequency and low frequency components. Low frequency components contain the information including basic object structures. High frequency components contain data noise. This low frequency components transmitted to subsequent layer. By using this WaveCNets that give better noise robustness and accuracy than vanilla version.

Chenchi Luo *et al.*[2] introduce WSN architecture for disparity estimation.Today smart phones introduce computational methods to overcome physical lens and sensors limitations. This methods utilize depth map to synthesis narrow DoF. High quality depth maps act as differentiator between computational bokeh and DSLR optical bokh. Wavelet synthesis neural network to produce high disparity map on smartphones.The evaluation matrix quantify the quality of disparity of real image. This may have better accuracy as compared to other CNN based algorithm.

Xiaotong Luo *et al.*[3] proposed a deep wavelet network with domain adaptation mechanism for single image demoireing . Feature mapping is done with wavelet domain .It reduce the information loss and cannot cut down computational complexity. In this Vnet structure up sampling and down sampling is replaced with DWT and IDWT. By this change can reducing information loss and computational complexity. In this method extracting more texture information by using residual – inresidual structures. When the given moire image is self similar by add global context block in the structure for learning the dependence between long distance pixel. Reducing domain shift in the training dataset by fine tuning of pretrained model.

Ji awang Bian *et al.*[4] introduce trained network can predict accurate depth over a video. Depth accuracy is evaluated both in the indoor and outdoor scenes.By using KITTI and NYUV2 dataset obtained high quality depth estimation and predict depth between adjascent views. ORB-SLAM 2 system have monocular trained deep networks which is more robust and accurate tracking .Depth network is unsupervised by trained with monocular video for camera tracking and dense reconstruction.

Clement Godard *et al.*[5] compares both qualitative and quantitative improved maps with self supervised method. To handle occlusion it designed with a minimum reprojection loss. The visual artefact is reduced with full resolution multi scale sampling method. This is a simple and efficient method for depth estimation.

Tobias Koch *et al.*[6] proposed the evaluation of single image depth estimation methods.High quality dataset is used for extended the ground truth of metrics.Here,quality is compared with meaningful properties are depth consistency ,planer region , preservation of edges, planer surface has lack of accuracy and edges gives depth .This is crucial for many applications.

Gabries J Brostow *et al.*[7] proposed unsupervised deep neural network for single image depth estimation .The network is trained with disparity images and image reconstruction loss.Training loss can reduce the disparity between both left and right images.By using supervised baseline is doesnot require expensive ground truth.Currently this method is frame independent .In this model estimate pixel depth and also predict full occupancy of the scence.

G. Huang *et al.*[8] proposed dense convolutional network is connected each layer in feed forward fashion. In this network L(L+1)/2 direct connections. Subsequent layer is used input as the feature map of preceding layers .In this method dense convolutional network as reference.Tuning the hyper parameters and learning rate schedules and obtained further gains in accuracy of DenseNet.DenseNet is good feature extractor and build on convolutional features.DesNets requiring less computation and high performance.

3. CONCLUSION

Depth estimation from image is a major task for scene understanding ,augment reality and autonomous driving . In Monocular depth estimation ,single RGB image given as input and to predict the depth values of each pixels.In wavelet method the depth prediction of single image is done by combining wavelet representation and deep learning.The main goal is to reconstruct full resolution depth map from low resolution, which is progressively upsampled and refined using wavelet coefficient projection at increasing scale. Wavelet method can cut down number of necessary operation in the decoder by half while producing a drop in accuracy of less than 2%. This paper is the summary of depth estimation of single image using wavelet method

ACKNOWLEDGEMENT

We would like to thank the Director of LBSITW and the Principal of the institution for providing the facilities and support for our work.

REFERENCES

[1] Qiufu Li, Linlin Shen, Sheng Guo, and Zhihui Lai. Wavelet integrated cnns for noise-robust image classification. In CVPR, June 2020.

[2] Chenchi Luo, Yingmao Li, Kaimo Lin, George Chen, Seok-Jun Lee, Jihwan Choi, Youngjun Francis Yoo, and Michael O. Polley. Wavelet synthesis net for disparity estimation to synthesize dslr calibre bokeh effect on smartphones. In CVPR, 2020

[3] Xiaotong Luo, Jiangtao Zhang, Ming Hong, Yanyun Qu, Yuan Xie, and Cuihua Li. Deep wavelet network with domain adaptation for single image demoireing. In CVPR Workshops, 2020.

[4] Jiawang Bian, Zhichao Li, Naiyan Wang, Huangying Zhan, Chunhua Shen, Ming-Ming Cheng, and Ian Reid. Unsupervised scale-consistent depth and ego-motion learning from monocular video. In NeurIPS, 2019.

[5] Clement Godard, Oisin Mac Aodha, Michael Firman, and ´ Gabriel J. Brostow. Digging into self-supervised monocular depth estimation. In ICCV, 2019.

[6] Tobias Koch, Lukas Liebel, Friedrich Fraundorfer, and Marco Korner. Evaluation of CNN-Based Single-Image ¨ Depth Estimation Methods. In ECCV, 2018.

[7] Clement Godard, Oisin Mac Aodha, and Gabriel J Bros-tow. Unsupervised monocular depth estimation with left�right consistency. In CVPR, 2017.

[8] G. Huang, Z. Liu, L. Van Der Maaten and K. Weinberger, "Densely Connected Convolutional Networks," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017.