25TH INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION MILAN, ITALY 10 | 15 January 2021

Attention Stereo Matching Network

Doudou Zhang[1,*], Jing Cai[1,*], Yanbing Xue[1,†], Zan Gao[2] and Hua Zhang[3]
[1]School of Computer Science and Engineering
Tianjin University of Technology, Tianjin 300384, China
[2]Institute of AI, Shandong Computer Science Center
National Supercomputer Center in Jinan
Qilu University of Technology, Shandong 250101, China
[3]Tianjin Sino-German University of Applied Sciences, Tianjin 300350, China

Submission number: 1371

Challenge and Related work

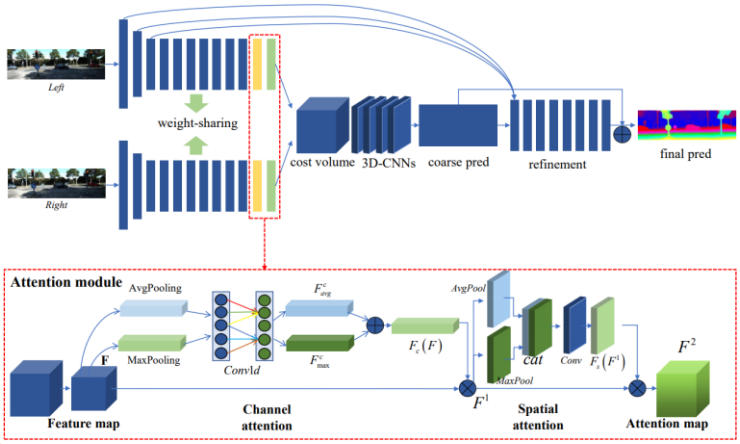
➢ **Challenge**

a. How to improve the matching accuracy in slender structures and textureless regions.
b. How to balance accuracy and speed.

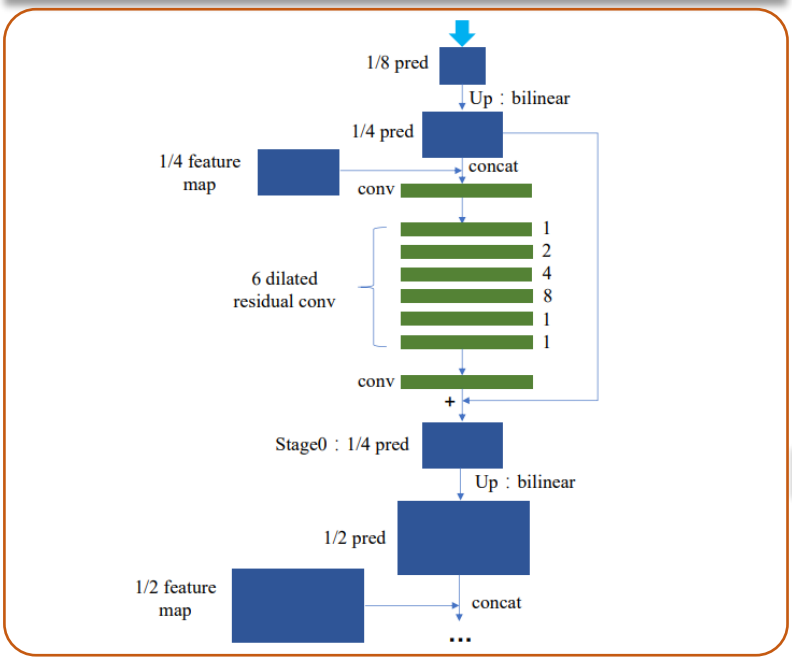
➢ **Related work**

1. The accuracy of the algorithm based on traditional stereo matching can not meet the requirements.
2. In recent years, the algorithm based on end-to-end using multi-scale 3D convolution has a large amount of calculation and slow calculation speed.
3. Some stereo matching networks based on multi-stage end-to-end deep learning can not balance the relationship between precision and speed.
4. There are few application models that can achieve the balance of speed and precision.

Our Proposed Method



Network Architecture of the Refinement Module



Experiments

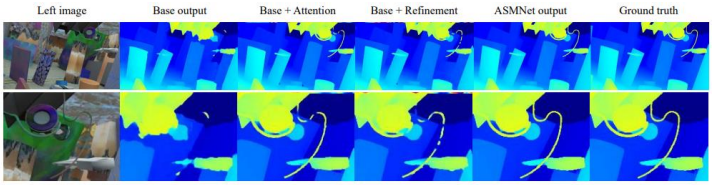


Fig. 4. Visualization of ablation experiments in slender structures.

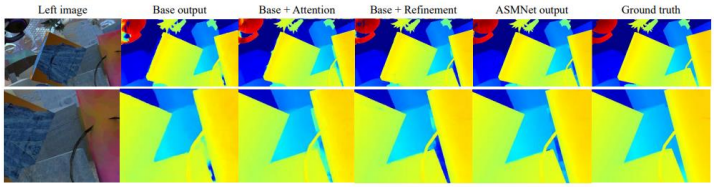


Fig. 5. Visualization of ablation experiments in textureless regions.

Model	EPE(px)	Runtime(s)	Device
Base	2.69	0.0097	GTX1080
Base+Attention	2.20	0.0102	GTX1080
Base+Refinement	1.38	0.0131	GTX1080
Base+Attention+Refinement	1.25	0.0142	GTX1080

TABLE I
EVALUATION OF ASMNET WITH DIFFERENT SETTINGS. WE COMPUTED END-POINT-ERROR ON THE SCENE FLOW TEST SET FOR COMPARING THE EFFECTS OF SEVERAL MODELS.

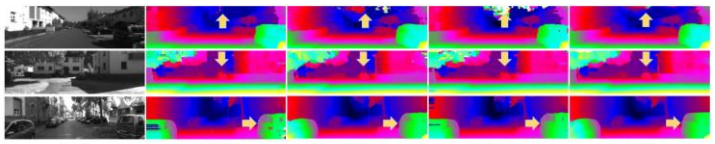


Fig. 7. Visualization results and comparisons on the KITTI2012 dataset. The left panel shows the left input image of stereo image pair. For each input image, the disparity maps obtained by MC-CNN [4], GC-Net [7], SGM-Net [36], and ASMNet are illustrated together.

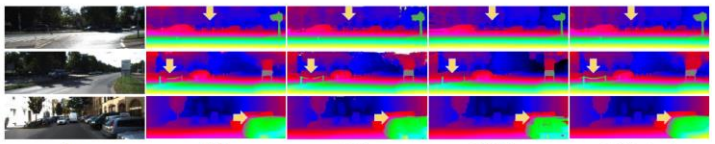


Fig. 8. Visualization results and comparisons on the KITTI2015 dataset. The left panel shows the left input image of stereo image pair. For each input image, the disparity maps obtained by CRL [21], GC-Net [7], SGM-Net [36], and ASMNet are illustrated together.

Methods	D1-bg(%)	D1-fg(%)	D1-All(%)	Runtime(s)
CRL [21]	2.48	3.59	2.67	0.5
GC-Net [7]	2.21	6.16	2.87	0.9
MC-CNN [4]	2.89	8.88	3.89	67
SGM-Net [36]	2.66	8.64	3.66	67
StereoNet [35]	4.30	7.45	4.83	0.015
ASMNet(ours)	3.18	5.98	3.64	0.014

TABLE III
KITTI2015 OFFICIAL ASSESSMENT WAS COMPARED TO A LIST OF OTHER METHODS.

Methods	Out-Noc(%)	Out-All(%)	Avg-Noc(px)	Avg-All(px)	Runtime(s)
GC-Net [7]	2.71	3.46	0.6	0.7	0.9
MC-CNN [4]	3.90	5.45	0.7	0.9	67
SGM-Net [36]	3.60	5.15	0.7	0.9	67
StereoNet [35]	4.91	6.02	0.8	0.9	0.015
ASMNet(ours)	4.30	4.96	0.7	0.7	0.014

TABLE IV
ASMNET RESULTS ARE SUBMITTED TO THE KITTI2012 EVALUATION SYSTEM AND COMPARED WITH OTHER METHODS.

Conclusion and Future work

In this work, we propose ASMNet, a novel end-to-end real time CNN architecture for stereo matching which consists of two main modules to exploit correlation information and disparity refinement: the attention module and the disparity refinement module based on low-level feature maps. Experiments show the effectiveness of the proposed method.

If you have any questions please contact: 247838227@qq.com