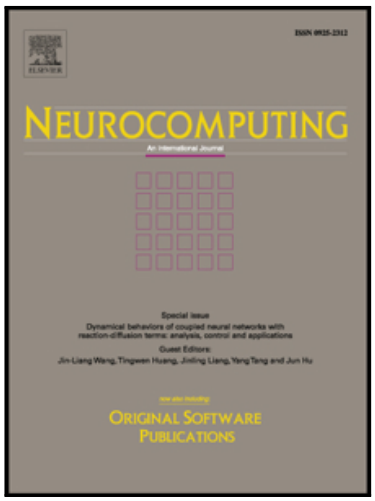

Monocular Depth Estimation with Guidance of Surface Normal Map

Han Yan, Shunli Zhang, Yu Zhang, Li Zhang

PII: S0925-2312(17)31771-X
DOI: 10.1016/j.neucom.2017.08.074
Reference: NEUCOM 19077

To appear in: *Neurocomputing*

Received date: 24 January 2017
Revised date: 5 July 2017
Accepted date: 16 August 2017

Please cite this article as: Han Yan, Shunli Zhang, Yu Zhang, Li Zhang, Monocular Depth Estimation with Guidance of Surface Normal Map, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.08.074

Monocular Depth Estimation with Guidance of Surface Normal Map

Han Yan[a,*], Shunli Zhang[b], Yu Zhang[a], Li Zhang[a]

[a]*Department of Electronic Engineering, Tsinghua University, Beijing 100084, P.R.China*
[b]*School of Software Engineering, Beijing Jiaotong University, Beijing 100044, PR China*

Abstract

In this work, we aim to tackle the task of monocular depth estimation, i.e., estimating depth map from only one single image. Without the references to determine the scale of the scene, the monocular depth estimation suffers from an inherent problem: depth ambiguity, which means the objects with similar appearances in image might have different depths. The depth ambiguity not only makes the depth estimation model hard to train, but also reduces the depth estimation accuracy. In this work, we present a new method to alleviate this problem. We observe the surface normal map is invariant with respect to the scale of the scene, thus we use the surface normal as reference to assist the depth prediction. Firstly, we present a multitask CNN to simultaneously produce the superpixel-wise depth and surface normal predictions. Then we introduce a CRF with an autoencoder based pairwise potential to refine the superpixel-wise predictions of CNN. At last, we propose a novel joint optimization algorithm which not only can enhance the depth prediction in accordance with the surface normal prediction, but also can transform the superpixel-wise depth map into a fine-grained pixel-wise depth estimation result. The proposed model is evaluated on NYU-D2, SUN RGB-D and Make3D datasets. Experimental results show that the proposed model can produce the state-of-the-art results while consuming relatively low GPU memory.

*Corresponding author
Email address: h-yan12@mail.tsinghua.edu.cn (Han Yan)