VIEWPOINT ESTIMATION IN IMAGES BY A KEY-POINT BASED DEEP NEURAL NETWORK

Jiana Yang, Shilin Wang*, Senior Member, IEEE, and Gongshen Liu
School of Electronic Information and Electrical Engineering, Shanghai Jiaotong University

Introduction

Viewpoint estimation aims to determine the rotation angle of an object in its 3D space from a 2D image, as shown in Fig 1. It is challenging due to the great variations in the object's shape, appearance, visible parts, etc. To overcome the above difficulties, this paper proposed a new deep neural network, which employs the key-points of the object as a regularization term and semantic bridge connecting the raw pixels with object's viewpoint.

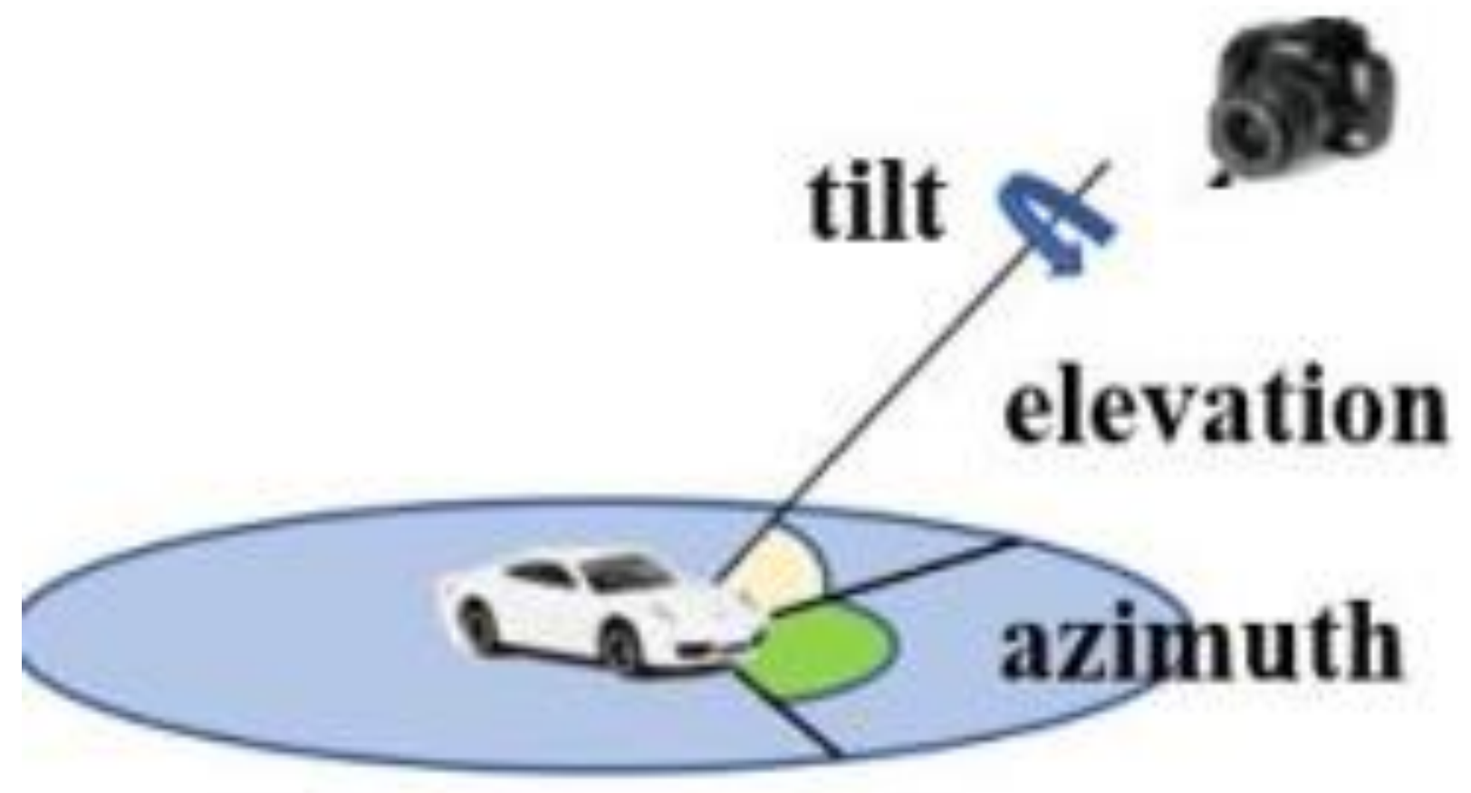

Fig.1 The 3D pose of an object

The proposed VE-Net

The new deep neural network proposed by this paper is called VE-Net. The overall architecture of VE-Net is given in Fig. 2 and it is composed of two parts, i.e. The key-point detection subnet (KD-Net) and The key-point encoding and viewpoint estimation subnet (KV-Net).

The KD-Net

The KD-Net extract key-points from object's bounding box by using stacked Hourglass structure (Fig 3).

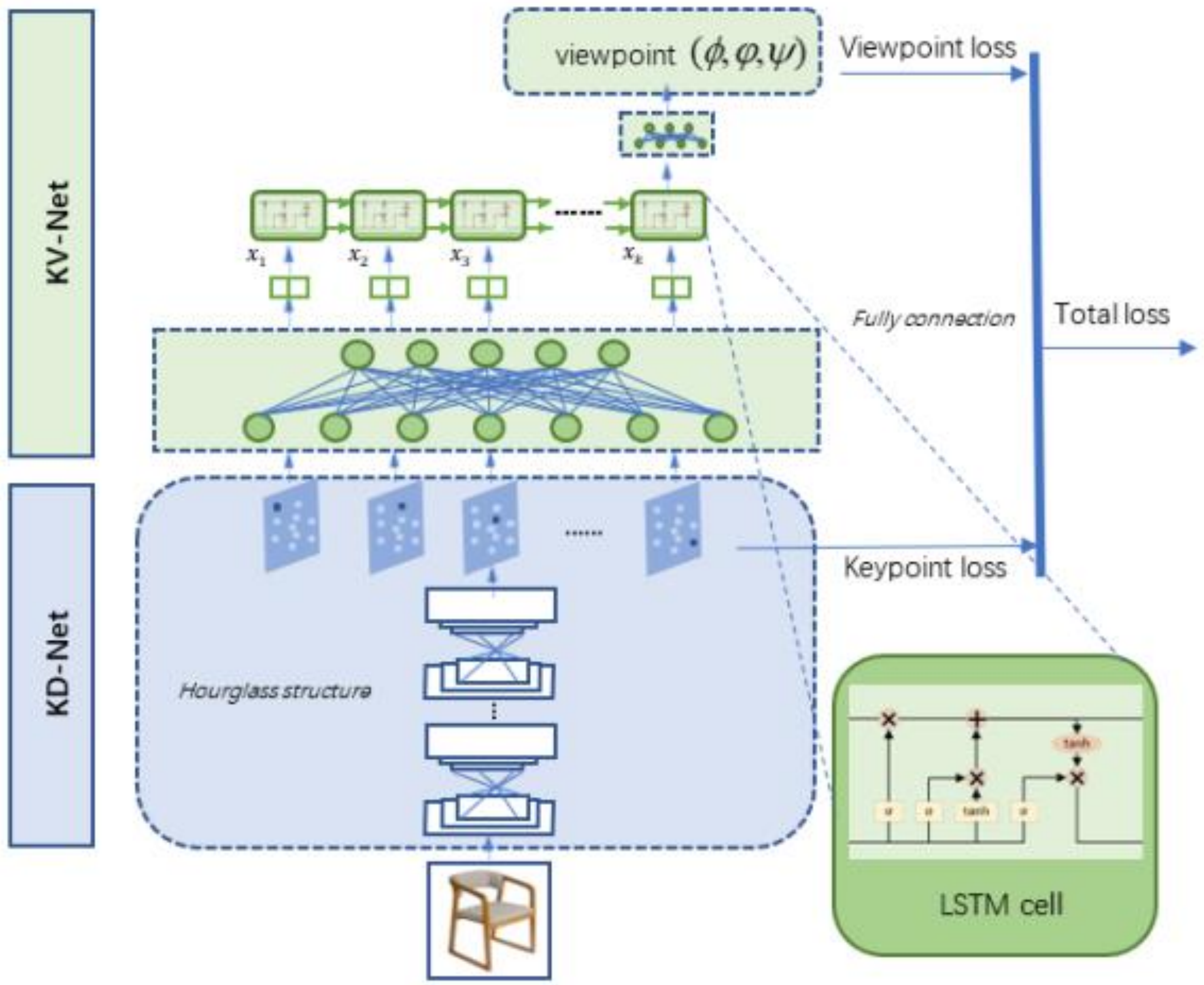

Fig.2 The overall structure of the VE-Net

The KV-Net

The key-point coordinates for all key-points detected by KD-Net are re-arranged according to a preset order and then fed into an LSTM network.

The output vector of the final LSTM state can be regarded as an encoded feature, which takes both the positions of all the extracted key-points and their intrinsic relationship into consideration and provides a more robust representation.

The encoded feature is then sent to fully connected layer which project the feature to the corresponding viewpoint class.

The KV-Net exploited the intrinsic spatial relationship among the key-points to relief the missing key-point problem caused by the noise/ambiguity in key-point detection and improve the accuracy and robustness of viewpoint estimation results.

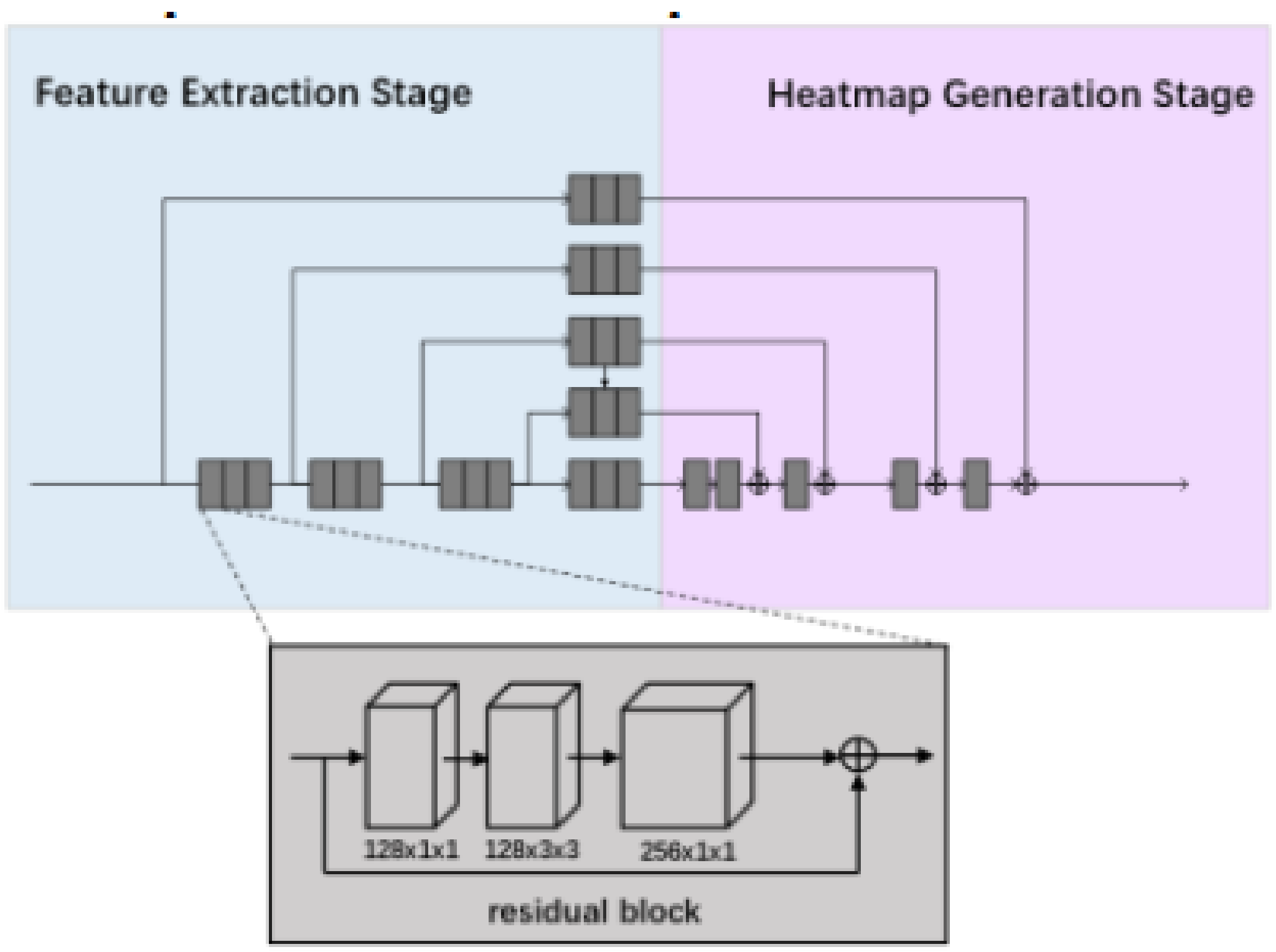

Fig.3 Illustration of the Hourglass structure

Experiment result

Category: Chair		[8]	[9]	[5]	[18]	Ours
Scenario I	Acc $\pi/6$	80	**86**	N/A	80	81
	Mederr	14.8	**9.7**	N/A	13.7	14.3
Scenario II	AVP24	17.5	7.4	4.4	N/A	**17.8**
	AVP $\pi/6$	27.8	13.8	11.4	N/A	**36.5**

Table.1 Viewpoint estimation results by various approaches.