Learning Robust Landmark Detection via Hierarchical Structured Ensemble

Xu Zou, Sheng Zhong
School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan, China;

Introduction

Heatmap regression-based models have significantly advance the progress of landmark detection. However, the lack of structural constraints always generates inaccurate heatmaps resulting in poor landmark detection performance. While hierarchical structure modeling methods have been proposed to tackle this issue, they are all heavily rely on the manually designed tree structures. The designed hierarchical structure is likely to be completely corrupted due to the missing or inaccurate prediction of landmarks. To the best of our knowledge, no work before has investigated how to automatically model proper structures for landmarks, by discovering their inherent relations. In this paper, we propose a novel Hierarchical Structured Landmark Ensemble (HSLE) model for learning robust landmark detection, by using it as the structural constraints. Different from existing approaches of manually designing structures, our proposed HSLE model is constructed automatically via discovering the most robust patterns so HSLE has the ability to robustly depict both local and holistic landmark structures simultaneously. Our proposed HSLE can be readily plugged into any existing landmark detection baselines for further performance improvement. Extensive experimental results demonstrate our approach significantly outperforms the baseline by a large margin to achieve a state-of-the-art performance.

Keywords: Landmark Detection, Heatmap Regression, Hierarchical Structural Constraints, Pattern Discovery

Design and Implementation

The framework of the proposed approach is illustrated in Figure 1. The entire model can be jointly learned in an end-to-end fashion. The proposed HSLE model served as hierarchical structural constraints of facial landmarks.

Results

We compare our end-to-end trained model against state-of-the-art methods on 300W dataset. We report average point to point Euclidean errors normalized by both inter-pupil distance (ipd-norm) and inter-ocular distance (iod-norm), and median point-to-point Euclidean errors normalized by inter-ocular distance (iod-norm). The results are shown in Table 1. Experimental results demonstrate our approach consistently significantly outperforms 3 different state-of-the-art baselines by a large margin to achieve a comparable result against the state-of-the-art methods.

	common	challenge	full
	mean iod-norm error		
PCD-NN	3.67	7.62	4.44
SAN	3.34	6.60	3.98
DAN	3.19	5.24	3.59
LAB	2.98	5.19	3.49
DU-Net-BW-α	3.00	5.36	3.46
DU-Net	2.82	5.07	3.26
DCFE	2.76	5.22	3.24
HG*	3.30	5.69	3.77
HG-HSLE (ours)	**2.85**	**5.03**	**3.28**
DU-NET*	3.07	5.13	3.47
DU-NET*-HSLE(ours)	**2.88**	**5.01**	**3.30**
Merget *et al.* *	3.76	6.32	4.26
Merget*-HSLE(ours)	**3.21**	**5.69**	**3.70**

Table.1. Quantitative Results against state-of-the art methods and baselines on 300W dataset

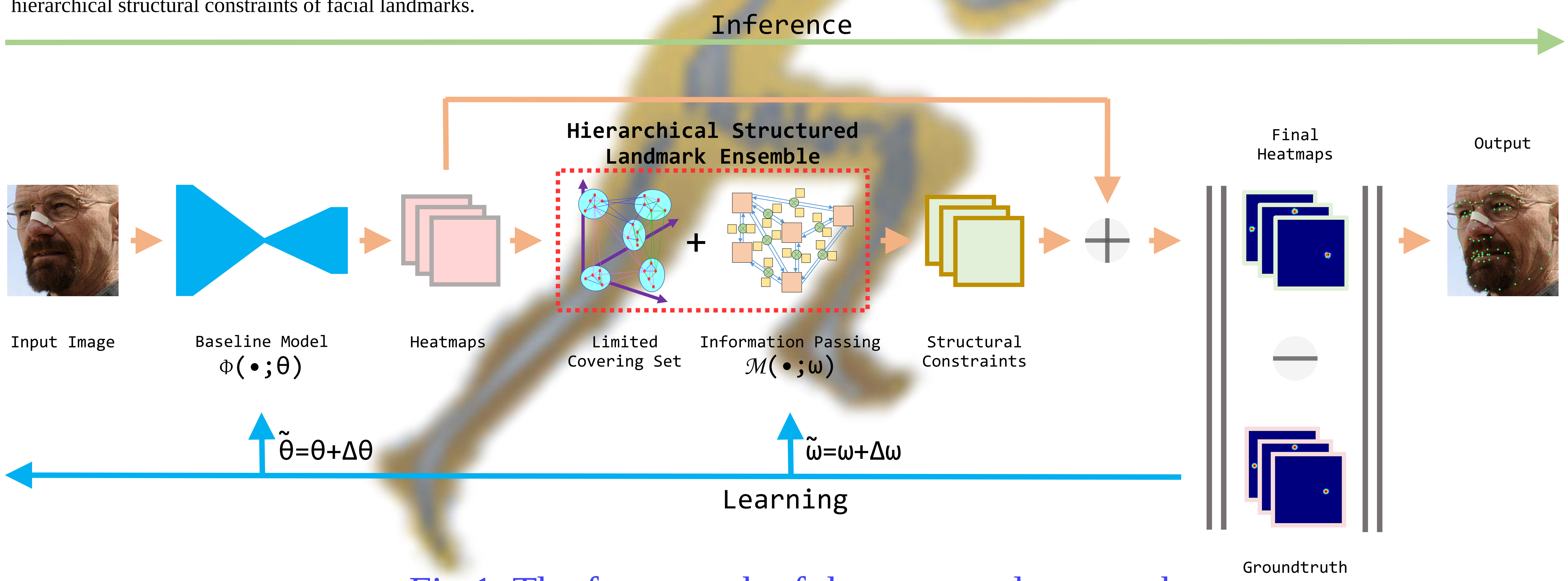

Fig.1. The framework of the proposed approach.

The Hierarchical Structured Landmark Ensemble (HSLE) model is first constructed automatically by discovering the most robust patterns. The HSLE model is used to represent holistic and local structural constraints of facial landmarks. Structural constraints, outputs of the HSLE, are expressed as a set of feature maps have the same 2D shape as heatmaps generated by the baseline model. In inference, the output of the entire model is a set of landmark coordinates directly derived from final heatmaps according to Equation $l_t^* = \arg\max \phi_t(\boldsymbol{I};\theta) + \tilde{\mathcal{H}}_t$

HSLE means clustering landmarks into different groups, connecting these landmarks within each group on the basis of specific structures, and passing information from one landmark to another through these structures. The HSLE is determined by:

$$\mathcal{C}^* = \arg\min_{\mathcal{C}} \sum_{i=1}^{N} \sum_{S_j^i \in \mathcal{C}_i} \kappa_{s_j^i}, \quad \left(\mathcal{C}_i^* = \arg\min_{\mathcal{C}_i} \sum_{S_j^i \in \mathcal{C}_i} \kappa_{s_j^i} \right)$$

$$s.t. \quad \begin{cases} \mathcal{K}(\mathcal{C}_i) \geq n_i - t_i \\ \mathcal{C}_i \subset T_i \\ \forall l_x \in \mathcal{C}_i, \sum_{j=1}^{m} \mathbf{1}\left(l_x \in S_j^i\right) \neq 0 \end{cases}$$

Conclusions

In this paper, we present a Hierarchical Structured Landmark Ensemble (HSLE) model for learning robust landmark detection. Due to the structural constraints propagated from the HSLE, the baseline landmark detector becomes more robust by trained jointly with the HSLE in an end-to-end fashion. The effectiveness of our idea has been verified by extensive experiments, indicates that landmark detection can be more robust via learning from hierarchical structural constraints.

Compared with the baseline model, the runtime of the proposed model for inference has increased by about 36ms on Intel i7-9700K (3.60GHz × 8) CPU and Nvidia GeForce GTX 1080Ti (11GB) GPU.

Acknowledgement

The authors acknowledge the school of Artificial Intelligence and Automation of Huazhong University of Science and Technology for providing funding.