

Deep Fashion Analysis with Feature Map Upsampling and Landmark-driven Attention

Jingyuan Liu^[0000–0003–1855–934X] and Hong Lu^[0000–0002–4572–2854]

Shanghai Key Lab of Intelligent Information Processing, School of Computer Science,
Fudan University, P. R. China
j i ngyuanl i u15, hongl u@fudan.edu.cn

Abstract. In this paper, we propose an attentive fashion network to address three problems of fashion analysis, namely landmark localization, category classification and attribute prediction. By utilizing a landmark prediction branch with upsampling network structure, we boost the accuracy of fashion landmark localization. With the aid of the predicted landmarks, a landmark-driven attention mechanism is proposed to help improve the precision of fashion category classification and attribute prediction. Experimental results show that our approach outperforms the state-of-the-arts on the DeepFashion dataset.

Keywords: Fashion Analysis · Landmark Detection · Clothing Category Classification · Attention Mechanism, Deep Learning

1 Introduction

Recent years, with the rapid growth of online commerce and fashion-related application, fashion image analysis and understanding have attracted increasing amount of attention in the community. Extensive studies have been conducted in this field, such as category classification, style or attribute prediction, fashion landmark localization, and fashion image synthesis.

In this paper, we study there core problems of fashion image analysis: landmark localization, category classification and attribute prediction. Previous works based on deep learning have shown much success in these fields [6][9][10][11][3][17][16]. However, most of them fail to further improve fashion analysis accuracy because of the low resolution of the predicted heatmaps after several pooling operations. It limits the prediction accuracy since fashion landmarks usually lie in the sharp corners or edges of clothes. In this paper, we address this problem by using transposed convolution to upsample the feature map. Thus, the predicted heatmaps are high-resolution and have the same size as the input fashion image, which will improve the accuracy of landmark localization.

For enhancing accuracy of category classification and attribute prediction, we also introduce a landmark-driven attention mechanism leveraging the predicted landmark heatmap. The landmark locations and the convolutional features are combined to form a new attention map, which gives our network a flexible way to

