Redesigned Skip-Network for Crowd Counting with Dilated Convolution and Backward Connection

S. Sooksatra, T. Kondo, P. Bunnun, and A. Yoshitaka

Abstract

Crowd counting is a challenging task dealing with object counting in a large number of objects. With an advanced technology, a regression-based method was utilized in a crowd counting to design an estimation network (CNN) and formulate a density map calculating the number of objects in the image [1]. In recent years, estimation networks with skip connections or forward connection, as shown in Fig. 1 (left) have been emphasized by existing studies, to improve high-level features represented as objects with large scale (high information on texture). Each layer extracts features in a different object scale and crowd density. However, crowded images generally consist of objects with small scales which can be extracted from shallow layers. Therefore, their networks are not suitable for crowd counting.

Fig. 1: The estimation network with skip connection consisting of forward (left) and backward connections (right).

In this research, skip-net has been modified for crowd counting. An estimation network with backward connections, reverse version of forward connections, as shown in Fig. 1 (right) is proposed by passing high-level features to shallow layers and emphasizing its low-level feature [2]. By this technique, shallow layers can acknowledge incoming information from higher layers and minimize a regression loss in convergence speed. Since an estimation network has a hierarchical structure, higher-level features can not be extracted without passing shallower layers, where a cyclic loop can be created in an estimation network. To solve this issue, two identical networks called Slave and Master networks are utilized. The Slave network has only function to formulate higher-level features for concatenating with lower-level features in Master network counting objects, where their weights were optimized separately. In addition, the pooling and up-sampling layers were replaced with dilated convolution to preserve semantic information or density map quality while increasing the size of receptive fields.

Our network was modified from U-net [3] as a backbone network and tested on three crowd counting datasets for counting humans and vehicles. The estimation network is evaluated by mean absolute error and mean square error indicating the accuracy and robustness of an estimation network, respectively. The empirical results show that our network has less error than other configurations of skip connections. For the effect of dilated convolution, more objects can be formulated in a predicted density map, reducing undercounting errors.

Keywords: surveillance system, crowd counting, regression-based approach, skip connection, dilated convolution.

Reference

[1] Sindagi, V. A., and Patel, V. M.l. "A survey of recent advances in cnn-based single image crowd counting and density estimation". *Pattern Recognition Letters*, 107, pp. 3-16, 2018.

[2] Sooksatra, S., Kondo, T., Bunnun, P., and Yoshitaka, A. “Redesigned Skip-Network for Crowd Counting with Dilated Convolution and Backward Connection”. *Journal of Imaging*, 6(5), pp. 28, 2020.

[3] Ronneberger, O., Fischer, P., and Brox, T. “U-net: Convolutional networks for biomedical image segmentation”. *In International Conference on Medical image computing and computer-assisted intervention*, pp. 234-241, 2015.