

# Stacked Hourglass Networks for Human Pose Estimation

代码开源. 第=页最底

Alejandro Newell, Kaiyu Yang, and Jia Deng

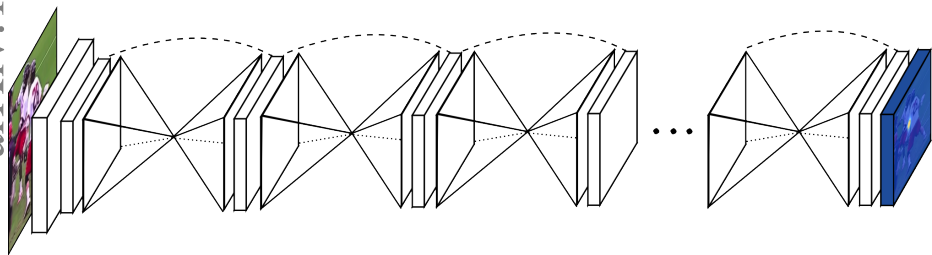
University of Michigan, Ann Arbor  
{alnewell, yangky, jiadeng}@umich.edu

**Abstract.** This work introduces a novel convolutional network architecture for the task of human pose estimation. Features are processed across all scales and consolidated to best capture the various spatial relationships associated with the body. We show how repeated bottom-up, top-down processing used in conjunction with intermediate supervision is critical to improving the performance of the network. We refer to the architecture as a “stacked hourglass” network based on the successive steps of pooling and upsampling that are done to produce a final set of predictions. State-of-the-art results are achieved on the FLIC and MPII benchmarks outcompeting all recent methods.

同样有

不断的 pooling 和上采样

**Keywords:** Human Pose Estimation



**Fig. 1.** Our network for pose estimation consists of multiple stacked hourglass modules which allow for repeated bottom-up, top-down inference.

## 1 Introduction

A key step toward understanding people in images and video is accurate pose estimation. Given a single RGB image, we wish to determine the precise pixel location of important keypoints of the body. Achieving an understanding of a person’s posture and limb articulation is useful for higher level tasks like action recognition, and also serves as a fundamental tool in fields such as human-computer interaction and animation.

对下游任务很重要