Programming Assignment 2: PoseRAC

Bingnan Li

2020533092

libn@shanghaitech.edu.cn

Abstract

In this assignment, I explored the state-of-the-art repeat action counting algorithm **PoseRAC** *and tried to reproduce the results in the original paper. Then I proved the necessity of Transformer encoder by canceling the encoder part and using Fully-Connected Layer only. Moreover, I also tried to explore the relationship between counting performance and the number of encoder layers together with number of heads in the multi-head attention module. The results show that the performance of PoseRAC barely improves when the number of encoder layers increases from 1 to 8, but the model will crush when the number of encoder layers is larger than 8 in my training setting. Besides, the performance of PoseRAC is not sensitive to the number of heads in the multi-head attention module. Finally, I replaced the triplet margin loss with* **contrastive loss** *and* **circle loss***, the results show that circle loss significantly boosts the model within 20 epoch, the evaluation metrics MAE and OBO improved from 0.2540 and 0.5395 to 0.2083 and 0.6053.*

1. Introduction

PoseRAC is the state-of-the-art repeat action counting algorithm proposed by [1]. This model achieves tremendous success in the repeat action counting task and improves the performance of the previous state-of-the-art model by a large margin. The novelty of PoseRAC is the new annotation method. The traditional method will annotate the start and the end frame of an action, models are forced the regress the locations or indices of an action. However, PoseRAC turns the annotation into two salient frames which indicates the most representative points of an action. Then PoseRAC utilizes a keypoint extractor to transform the salient frames into a series of keypoint with 3D coordinates. This operation enormously reduces the amount of data need to process and improves the effectiveness of information because it only uses transformer encoder layer to get the embedding information and a single FC layer to classify the embedding.

2. Formatting your paper

All text must be in a two-column format. The total allowable size of the text area is $6\frac{7}{8}$ inches (17.46 cm) wide by $8\frac{7}{8}$ inches (22.54 cm) high. Columns are to be $3\frac{1}{4}$ inches (8.25 cm) wide, with a $\frac{5}{16}$ inch (0.8 cm) space between them. The main title (on the first page) should begin 1 inch (2.54 cm) from the top edge of the page. The second and following pages should begin 1 inch (2.54 cm) from the top edge. On all pages, the bottom margin should be $1\frac{1}{8}$ inches (2.86 cm) from the bottom edge of the page for 8.5×11-inch paper; for A4 paper, approximately $1\frac{5}{8}$ inches (4.13 cm) from the bottom edge of the page.

3. Final copy

You must include your signed IEEE copyright release form when you submit your finished paper. We MUST have this form before your paper can be published in the proceedings.

Please direct any questions to the production editor in charge of these proceedings at the IEEE Computer Society Press: https://www.computer.org/about/contact.

References

[1] Ziyu Yao, Xuxin Cheng, and Yuexian Zou. Poserac: Pose saliency transformer for repetitive action counting. *arXiv preprint arXiv:2303.08450*, 2023. 1