Multimedia Systems (2022) 28:1557–1558
https://doi.org/10.1007/s00530-022-00924-8

CORRECTION

Correction: STASiamRPN: visual tracking based on spatiotemporal and attention

Ruixu Wu[1,2] · Xianbin Wen[1,2] · Zhanlu Liu[1,3] · Liming Yuan[1,2] · Haixia Xu[1,2]

Published online: 13 May 2022

Correction to: Multimedia Systems
https://doi.org/10.1007/s00530-021-00845-y

The authors have corrected this article. After publication, concerns were raised about text overlap with an earlier preprint by Saribas et al. [1] in several paragraphs of the article without appropriate attribution. The authors have revised these paragraphs as follows:

However, temporal features are also critical in many computer vision applications where spatiotemporal information is adopted [1]. During the video tracking process, the tracking object will change continuously in the video frame, it will be in a static state or a moving state, its shape and size will also change, or there will be multiple similar objects in the video sequence. Therefore, it is necessary to capture longer temporal frames with more motion information.

…

Teng et al. [35] present a new deep architecture which incorporates the temporal and spatial information to boost the tracking performance.

…

STASiamRPN is proposed as a new visual object tracking method, its illustration is shown in Fig. 1. Our method STASiamRPN is based on SiamRPN [15]. It uses 3D CNN as the backbone network and integrates the cascade attention module (CAM).

…

In such cases, 3D information supplements the time dimension, with time changing, more motion information of the target can be obtained, and irrelevant background information will be greatly reduced. The 3D-CIResNet-22, proposed in this paper, adds a spatiotemporal dimension based on the 2D-CIResNet-22 and upgrades the backbone networks from a 2D CNN to 3D CNN.

…

During online tracking, we use 5 frames as a video sequence, which includes the first 4 frames and the last frame. We use the last frame as the tracking frame. If the video sequence was less than 5 frames, we repeated the first frame.

…

OTB2015 [40] is a widely used object tracking dataset, which contains 100 challenging video sequences. The tracking scenes involved in these video sequences can be divided into 11 annotation attributes, including fast motion (FM), background clutter (BC), motion blur (MB), deformation (DEF), illumination variation (IV), in-plane rotation (IPR),

The original article can be found online at https://doi.org/10.1007/s00530-021-00845-y.

✉ Xianbin Wen
xbwen@email.tjut.edu.cn

✉ Zhanlu Liu
lzl2005@email.tjut.edu.cn

Ruixu Wu
wrxwd@qq.com

Liming Yuan
yuanleeming@163.com

Haixia Xu
xuhaixia_xhx@163.com

[1] School of Computer Science and Engineering, Tianjin University of Technology, Tianjin 300384, China

[2] Key Laboratory of Computer Vision and System, Ministry of Education, Tianjin 300384, China

[3] Department of Physical Education, Tianjin University of Technology, Tianjin 300384, China

low resolution (LR), occlusion (OCC), out-of-plane rotation (OPR), out-ofview (OV), and scale variation (SV).

…

In order to show the impact of various components of the tracker, we conducted ablation experiments on the OTB2015 datasets. Table 3 shows the experimental results using different components. We replace the baseline SiamRPN backbone network AlexNet[17] with 3D-CIResNet-22 and analyzed their impact.

…

Finally, we used CAM for experiments and the results showed the highest experimental scores.

…

Reference

1. Saribas, H., Cevikalp, H., Köpüklü, O., & Uzun, B. TRAT: Tracking by Attention Using Spatio-Temporal Features. arXiv preprint arXiv:2011.09524 (2020).