Works on crowd counting in videos

Spatiotemporal Dilated Convolution with Uncertain Matching for Video-based Crowd Estimation Yu-Jen Ma, Hong-Han Shuai, Member, IEEE, and Wen-Huang Cheng, Senior Member, IEEE (2021)

<https://arxiv.org/pdf/2101.12439v1.pdf>

LOCALITY-CONSTRAINED SPATIAL TRANSFORMER NETWORK FOR VIDEO CROWD COUNTING Yanyan Fang a ,Biyun Zhan a , Wandi Cai a , Shenghua Gao b , Bo Hu a,∗ (2019)

<https://arxiv.org/pdf/1907.07911v1.pdf>

Spatiotemporal Modeling for Crowd Counting in Videos Feng Xiong Xingjian Shi Dit-Yan Yeung Department of Computer Science and Engineering Hong Kong University of Science and Technology (2017)

<https://arxiv.org/pdf/1707.07890v1.pdf>

Comparable work done on pictures

TBD

Avilable datasets

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Dataset | Total Annotated frames | Training/Test frames | Total Annotated pedestrian | Average pr frame | Color | Resolution |
| UCSD | 2000 | 800/1200 | 49885 | 24.9 | Grey | 158 x 238 |
| MALL | 2000 | 800/1200 | 62315 | 31.2 | RGB | 640 x 480 |
| Worldexpo’10 | 3980 | 3380/600 | 199923 | 50.2 | RGB | 576 x 720 |

Evaluation Metrics

MAE (mean absolute error)

MSE (mean squared error)