Second Reading Pass - 24.10.21

Data

CUHK-SYSU

CUHK-SYSU (Xiao et al. 2017) is a large scale person search dataset containing 18,184 scene images and 96,143 annotated BBoxes, which are collected from two sources: street snap and movie. All people are divided into 8,432 labeled identities and other unknown ones. The training set contains 11,206 images and 5,532 different identities. The test set contains 6,978 images and 2,900 query people. The training and test sets have no overlap on images and query people. For each query, different gallery sizes from 50 to 4000 are pre-defined to evaluate the search performance. If not specified, the gallery size of 100 is used by default.

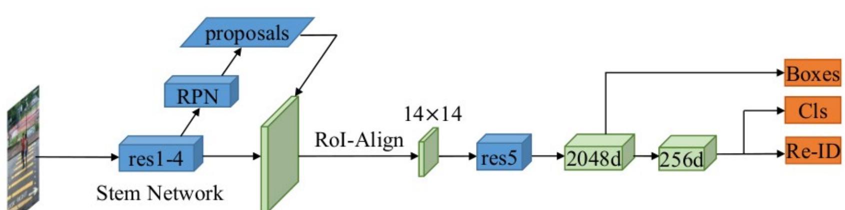
PRW

PRW is another widely used dataset (Zheng et al. 2017) containing 11,816 video frames captured by 6 cameras in Tsinghua university. 34,304 BBoxes are annotated manually. Similar to CUHK-SYSU, all people are divided into labeled and unlabeled identities. The training set contains 5,704 images and 482 different people, while the test set includes 6,112 images and 2,057 query people. For each query, the gallery is the whole test set, i.e., the gallery size is 6112.

Baselines

Multi-task network NAE is the baseline of this article.

The baseline provided is as below figure.



(a) Baseline

Upper Bounds

Instead of an upper bound, the motivation is determined in the paper. Here is the shared motivation:

Previous multi-stage Faster R-CNN is proposed to achieve better detection performance. However, the method aims at solving the detection and re-ID sequentially with a jointly optimized network to extract more discriminative features.

State of the Art

	Method	CUHK-SYSU		PRW	
		mAP	top-1	mAP	top-1
two-stage	DPM(Girshick et al. 2015)	-	-	20.5	48.3
	MGTS(Chen et al. 2018)	83.0	83.7	32.6	72.1
	CLSA(Lan, Zhu, and Gong 2018)	87.2	88.5	38.7	65.0
	RDLR(Han et al. 2019)	93.0	94.2	42.9	70.2
	IGPN(Dong et al. 2020b)	90.3	91.4	47.2	87.0
	TCTS(Wang et al. 2020)	93.9	95.1	46.8	87.5
end-to-end	OIM(Xiao et al. 2017)	75.5	78.7	21.3	49.9
	IAN(Xiao et al. 2019)	76.3	80.1	23.0	61.9
	NPSM(Liu et al. 2017)	77.9	81.2	24.2	53.1
	RCAA(Chang et al. 2018)	79.3	81.3	-	-
	CTXGraph(Yan et al. 2019)	84.1	86.5	33.4	73.6
	QEEPS(Munjal et al. 2019)	88.9	89.1	37.1	76.7
	HOIM(Chen et al. 2020a)	89.7	90.8	39.8	80.4
	BINet(Dong et al. 2020a)	90.0	90.7	45.3	81.7
	NAE(Chen et al. 2020b)	91.5	92.4	43.3	80.9
	NAE+(Chen et al. 2020b)	92.1	92.9	44.0	81.1
	OIM(ours)	87.1	88.5	34.0	75.9
	OIM+SeqNet(ours)	93.4	94.1	45.8	81.7
	OIM+SeqNet+CBGM(ours)	94.3	95.0	46.6	84.9
	NAE+SeqNet(ours)	93.8	94.6	46.7	83.4
	NAE+SeqNet+CBGM(ours)	**94.8**	**95.7**	**47.6**	**87.6**

Table 6: Comparison of mAP and top-1 accuracy with the state-of-the-art methods on CUHK-SYSU and PRW. Our models are shown in italics.

Ablation Study

Different detectors and re-identifiers;

SeqNet comes from better detection or more discriminative features. Person Search task is separated into two stages as detection stage with different detectors and re-ID stage with different re-identifiers.

SeqNet achieves better detection than NAE in overall. It is mainly because each head (RPN head/Faster R-CNN head/baseline head) of SeqNet will perform regression to BBoxes, which makes the model more selective against false positives.

SeqNet is more discriminative for re-ID because of the inconsistency of NAE , i.e., trained by low-quality proposals but tested by high-quality detected BBoxes.

Detection is not the performance bottleneck, SeqNet gains very little from better detection, and future research should focus on how to achieve a better re-ID.

SeqNet has two key components: FCS to improve the classification ability, NMS to accelerate the inference speed.

SeqNet is integrated into the existing end-to-end framework to verify the universality of the method. OIM is chosen as the base network. SeqNet improves the mAP of OIM by 6.3% and 11.8% on CUHK-SYSU and PRW benchmarks respectively, which demonstrates that the method is insensitive to base network.