Tiny Head Pose Classification by Bodily Cues

Hasan I.[1], Tsesmelis T.[2], Galasso F.[3], Bue D. A.[2], and Cristani M.[1]

[1] University of Verona, Italy, [2] Istituto Italiano di Tecnologia Genoa, Italy, [3]Corporate Innovation OSRAM GmbH

Abstract

Head pose is an important cue for computer vision. Traditionally considered in human computer interaction applications, it becomes very hard to model in surveillance scenarios, due to the tiny head size. Here we present a framework based on Faster RCNN [1], which introduces a novel branch for head pose estimation. The key idea is to leverage the presence of human body, to better infer the head pose through a joint optimization process. Results on this novel benchmark and ablation studies on other task-specific datasets promote our idea and confirm the importance of the body cues to contextualize the head pose estimation.

1. Proposed Architecture

- Our model jointly detects people and estimates their viewing angle (where the people are looking at, which we call Head Pose Network HPN highlighted ellipse).
- HPN may work both with the whole detected people and after having localized their head.

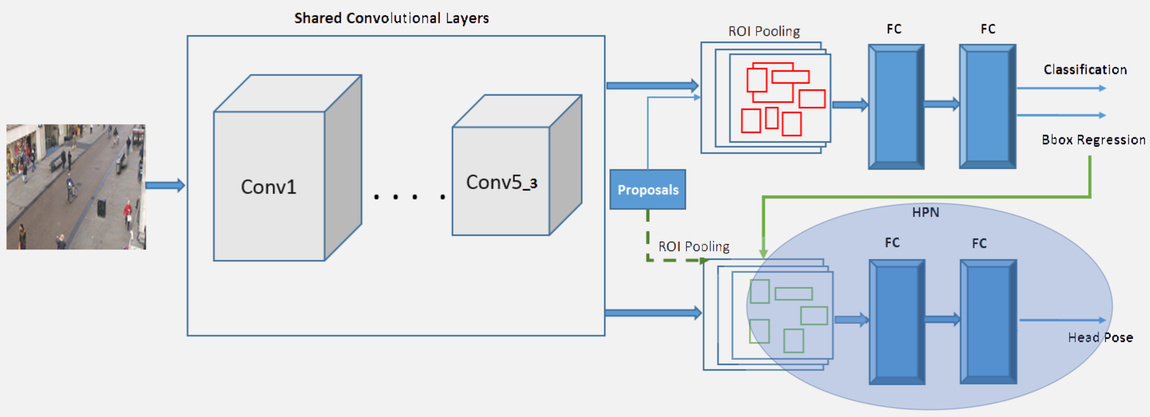

Fig. 1.Network Architecture. The figure illustrates the proposed Head Pose Classification Network (HPN). The green dotted line represents the filtered proposals at the training time and green solid represents the pedestrian detections at testing time.

2. Joint People Detection and Head Pose Estimation

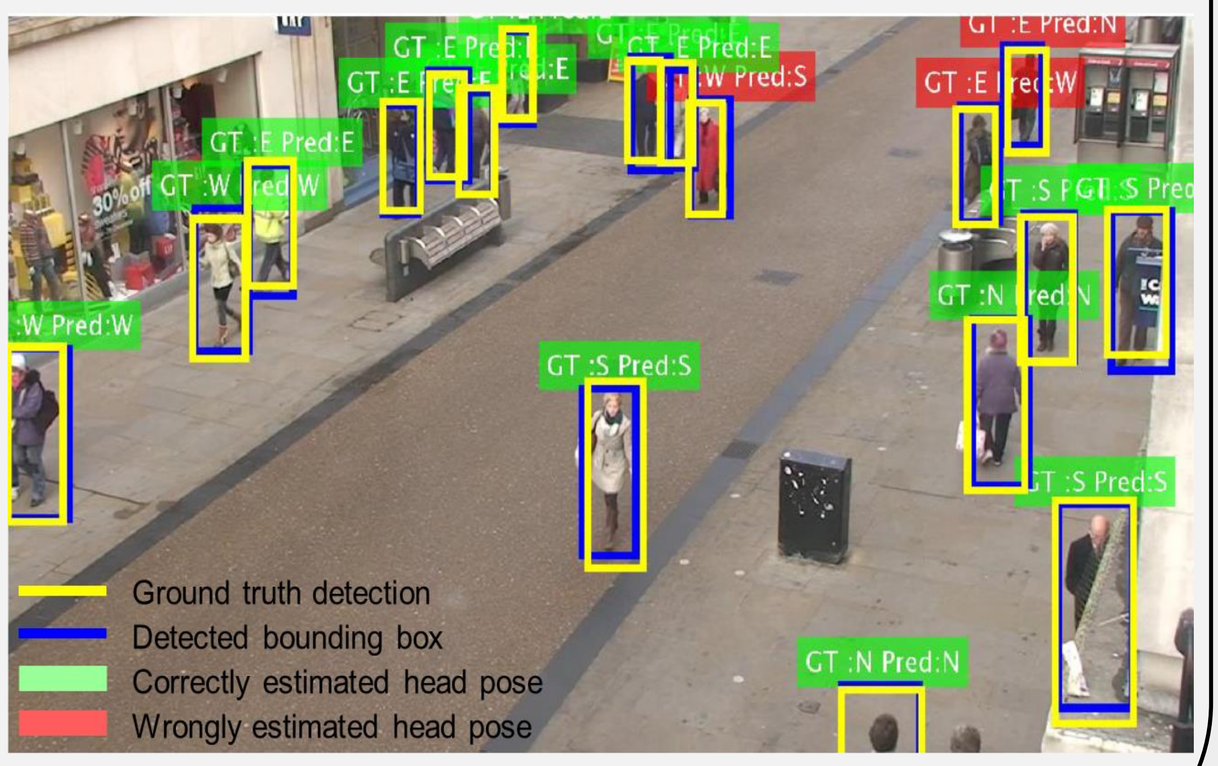

Fig. 2:Qualitative results of our proposed joint model.

3. Loss function

$$L(p, u, t^u, v, h, g) = L_{cls}(p, u) + \lambda[u = 1]L_{loc}(t_u, v) + \gamma[u = 1]L_{hp}(g, h)$$

- L_{cls} and L_{loc} are the original loss functions for background vs pedestrian classification and bounding box regression respectively.
- L_{hp} refers to the loss for the head pose. We are using softmax loss over K discrete directions.

4. Results

Table 1. Comparison with the state of the art head pose classification in regard to image scale variation.

Methods	Dataset	*HIIT*			*QMUL*			*QMULB*		
	Image Size	15x15	20x20	50x50	15x15	20x20	50x50	15x15	20x20	50x50
Frobenius	[2]	82.4	89.6	95.3	59.5	82.6	94.3	54.5	76.5	92
CBH	[2]	84.6	90.4	95.7	59.8	83.2	94.9	57	76.9	92.2
RPF	[3]	97.6	97.6	97.6	94.14	94.3	94.3	91.9	92.1	92.2
PSMAT	[4]	-			-		82.3	-		64.2
ARCO	[5]	-			-		93.5	-		89
HPN		**98.4**	**98.9**	**99.01**	**97.4**	**97.9**	**98**	**95.3**	**95.9**	**94.7**

4. Future Works

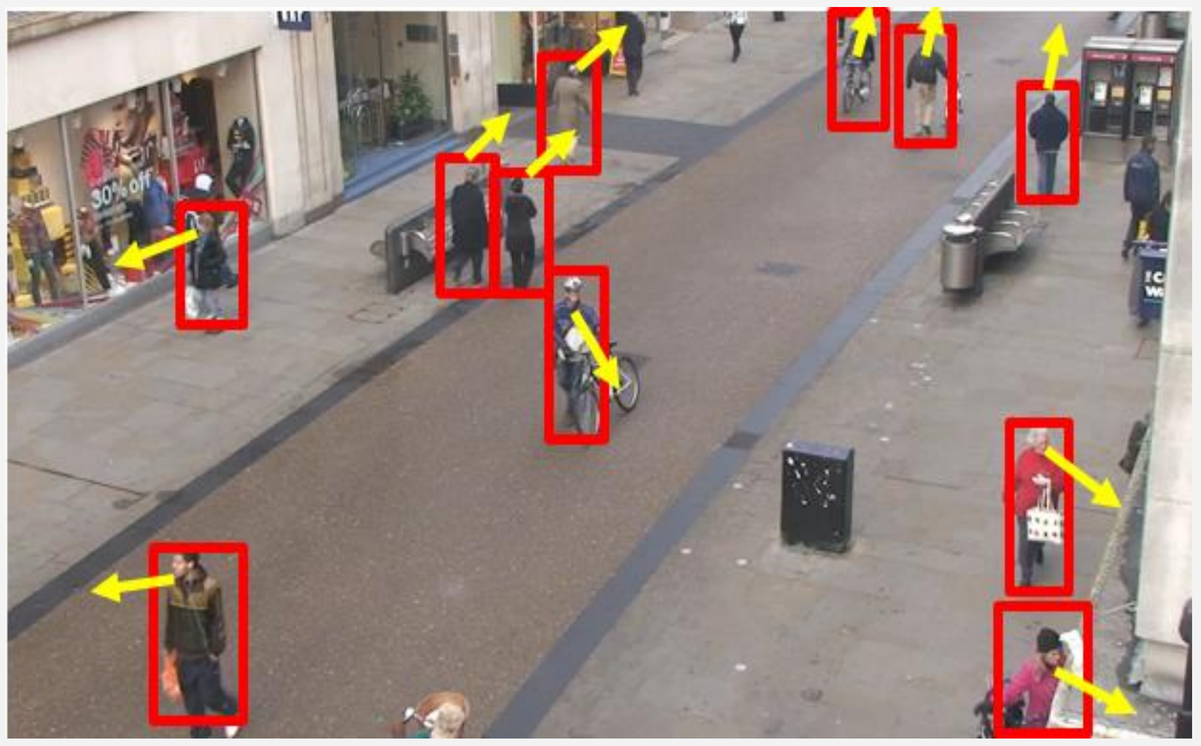

Fig. 3.Posing head pose estimation as a regression problem.

References

[1] Ren et al. Faster R-CNN: towards real time object detection. NIPS 2015
[2] Tosato et al. Characterizing humans on Riemannian manifolds. TPAMI 2013
[3] Lee et al. Fast and accurate head pose estimation via random projection forests. ICCV 2015
[4] Orozco et al. Head pose classification in crowded scenes. BMVC 2009.
[5] Tosato et al. Multi class classification on Riemannian manifolds for video surveillance. ECCV 2010

SceneUnderLight

ISTITUTO ITALIANO DI TECNOLOGIA VISUAL GEOMETRY AND MODELLING

UNIVERSITÀ DEGLI STUDI DI VERONA

OSRAM