Fine-Grained Head Pose Estimation Without Keypoints

Nataniel Ruiz, Eunji Chong, James M. Rehg

Georgia Institute of Technology

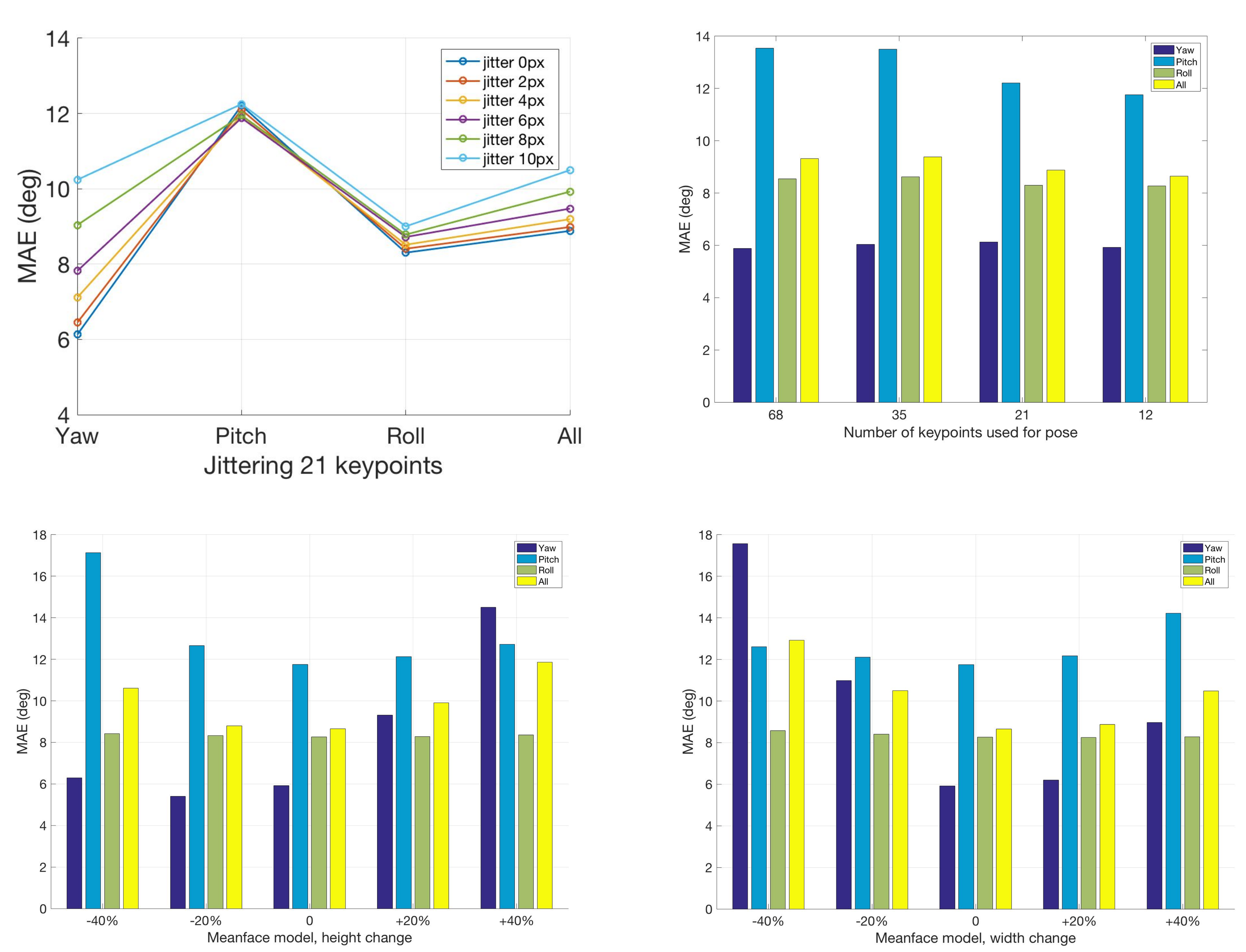
Contribution and take-home message

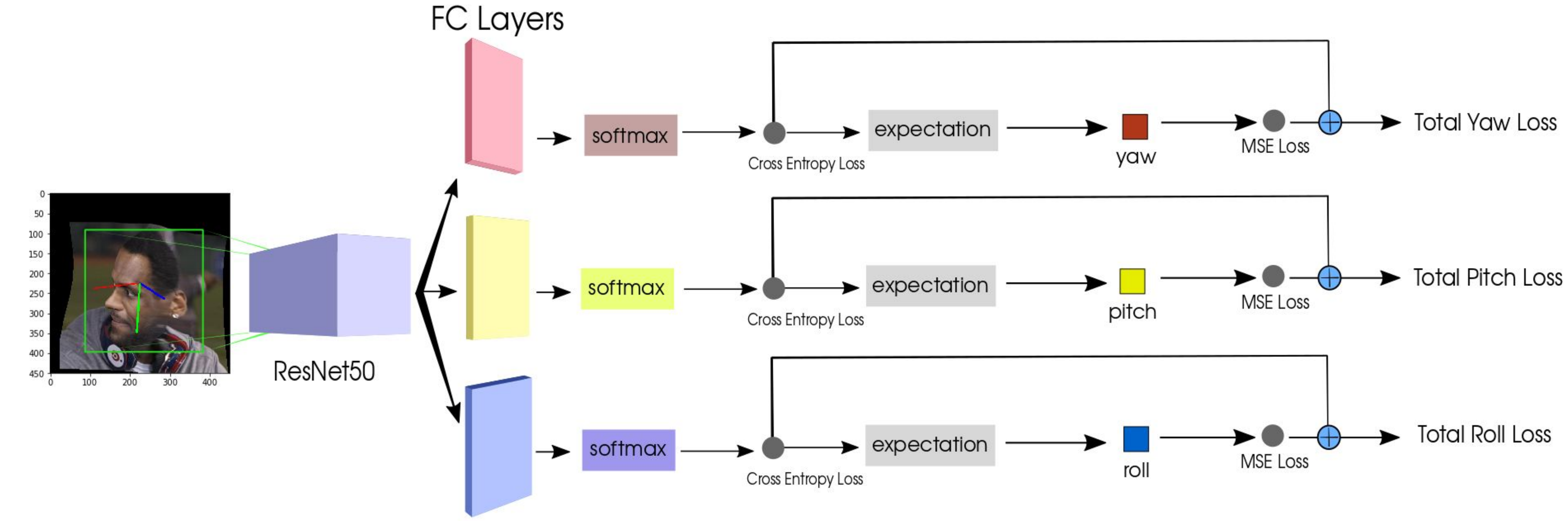
- Obtaining head-pose through keypoints is fragile and suboptimal.
- Using a deep network trained with a binned pose classification loss and a pose regression loss on a large dataset obtains state-of-the-art results which generalize to different datasets.
- Our method coupled with data augmentation is effective in tackling head pose estimation in low-resolution conditions.

On the Fragility of Landmark-To-Pose

Landmark to pose methods are sensitive to:

- Noise of landmark points
- Shape of mean head model
- Alignment algorithm
- Accuracy of landmark detector





Our Method

Different fully-connected layer for each Euler angle.

We use a linear combination of a binned pose classification loss and a regression loss.

$$\mathcal{L} = H(y, \hat{y}) + \alpha \cdot MSE(y, \hat{y})$$

Where H and MSE respectively designate the cross-entropy and mean squared error loss functions.

Examples

Experiments

	Yaw	Pitch	Roll	MAE
Multi-Loss ResNet50 ($\alpha = 2$)	5.167	6.975	3.388	5.177
Multi-Loss ResNet50 ($\alpha = 1$)	4.810	6.606	3.269	**4.895**
KEPLER [14]†	8.084	17.277	16.196	13.852
Multi-Loss ResNet50 ($\alpha = 1$)†	5.785	11.726	8.194	8.568
3DMM+ Online [33] *	2.500	1.500	2.200	**2.066**
FAN [2] (12 points)	8.532	7.483	7.631	7.882
Dlib [11] (68 points)	16.756	13.802	6.190	12.249
3DDFA [35]	36.175	12.252	8.776	19.068

Table 2. Mean average error of Euler angles across different methods on the BIWI dataset [6]. * These methods use depth information. † Trained on AFLW