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