

# Rotation Axis Focused Attention Network (RAFA-Net) for Estimating Head Pose

Ardhendu Behera<sup>[0000–0003–0276–9000]</sup>, Zachary Wharton<sup>[0000–0001–5983–6468]</sup>,  
Pradeep Hewage<sup>[0000–0002–6909–3546]</sup>, and Swagat Kumar<sup>[0000–0001–7405–3445]</sup>

## Supplementary Material

Tables 5 and 6 show the quantitative outputs of our model trained on 300W-LP and tested on the respective AFLW2000 and BIWI datasets. This has been presented in section 4.5 (Ablation Study) and Figure 7 of the main paper.

Similarly, Table 7 represents the quantitative outputs of our model trained and tested on BIWI dataset. Figure 8 visualises these outputs using the effect of bounding box margin (control with parameter  $0 \leq \gamma \leq 0.5$ ) on the average pose estimation error in degrees (y-axis) on different angle ranges (x-axis) for yaw, pitch and roll.

Table 5: The average pose estimation error in degrees using our RAFA-Net, which is trained on 300W-LP dataset and tested on AFLW2000 dataset. The effect of the bounding box margin (control with parameter  $0 \leq \gamma \leq 0.5$ ) on pose estimation error in degrees on different angle ranges for yaw, pitch and roll.

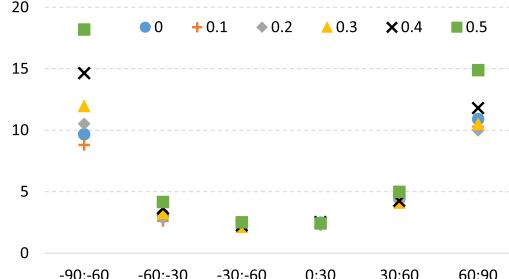
Angle Bin	Yaw ( $0 \leq \gamma \leq 0.5$ )					Pitch ( $0 \leq \gamma \leq 0.5$ )					Roll ( $0 \leq \gamma \leq 0.5$ )							
	0.0	0.1	0.2	0.3	0.4	0.5	0.0	0.1	0.2	0.3	0.4	0.5	0.0	0.1	0.2	0.3	0.4	0.5
-90:-60	5.8	4.6	4.9	5.6	6.0	6.1	34.6	33.1	32.4	31.5	32.8	37.5	35.8	30.3	31.5	32.6	35.2	37.5
-60:-30	5.1	4.3	4.1	4.1	4.2	4.3	10.1	9.1	9.2	8.8	9.3	11.5	10.2	9.4	9.3	9.3	10.0	11.5
-30:0	3.5	3.0	2.9	2.9	2.9	3.1	3.7	3.5	3.5	3.5	3.6	2.6	3.0	2.7	2.6	2.5	2.6	2.6
0:30	3.5	3.1	2.9	2.8	2.8	2.9	6.0	5.5	5.4	5.2	5.2	2.8	3.2	2.9	2.9	2.8	2.8	2.9
30:60	5.2	4.8	4.5	4.8	4.6	4.5	20.8	18.6	17.9	18.2	18.9	11.9	12.1	10.5	10.2	10.5	11.2	11.9
60:90	6.1	6.1	6.2	6.5	6.4	6.3	39.3	38.6	41.8	43.4	42.6	48.2	52.6	49.1	48.6	48.2	47.6	48.2

Table 6: The average pose estimation error in degrees using our RAFA-Net, which is trained on 300W-LP dataset and tested on BIWI dataset. The effect of the bounding box margin (control with parameter  $0 \leq \gamma \leq 0.5$ ) on pose estimation error in degrees on different angle ranges for yaw, pitch and roll.

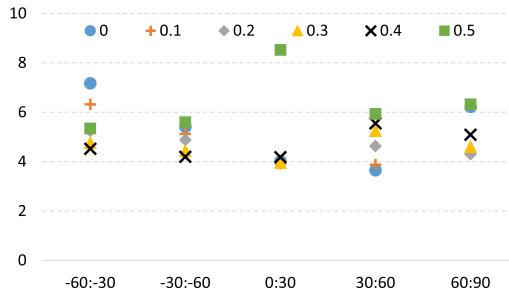
<b>Angle Bin</b>	<b>Yaw (<math>0 \leq \gamma \leq 0.5</math>)</b>					<b>Pitch (<math>0 \leq \gamma \leq 0.5</math>)</b>					<b>Roll (<math>0 \leq \gamma \leq 0.5</math>)</b>							
	0.0	0.1	0.2	0.3	0.4	0.5	0.0	0.1	0.2	0.3	0.4	0.5	0.0	0.1	0.2	0.3	0.4	0.5
-90:-60	4.1	4.1	4.2	4.3	4.6	4.5	19.6	20.5	18.4	18.6	18.8	20.8	-	-	-	-	-	-
-60:-30	9.1	8.7	8.3	8.0	7.7	7.9	7.5	6.9	6.4	6.4	6.5	7.1	13.6	12.2	10.8	10.4	10.7	12.6
-30:0	10.2	7.9	6.8	6.5	6.7	7.1	3.4	3.2	3.3	3.1	3.0	3.1	4.1	3.8	3.5	3.4	3.5	3.6
0:30	9.2	5.9	5.0	4.9	4.9	5.0	6.0	5.6	5.4	5.4	5.4	5.5	3.4	3.2	3.0	2.9	2.8	2.9
30:60	6.4	5.2	5.1	4.9	4.9	5.1	13.4	9.7	8.6	8.4	8.8	10.1	17.8	13.1	11.6	11.8	12.7	14.1
60:90	6.0	6.1	6.4	6.8	7.3	7.8	52.9	48.8	42.9	40.8	40.0	42.2	35.0	30.1	28.1	26.4	29.6	29.2

Table 7: The average pose estimation error in degrees using our RAFA-Net, which is trained and tested on BIWI dataset. The effect of the bounding box margin (control with parameter  $0 \leq \gamma \leq 0.5$ ) on pose estimation error in degrees on different angle ranges for yaw, pitch and roll.

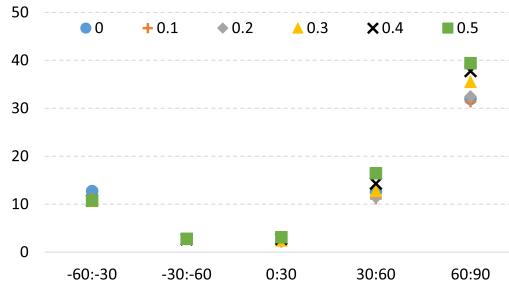
<b>Angle Bin</b>	<b>Yaw (<math>0 \leq \gamma \leq 0.5</math>)</b>					<b>Pitch (<math>0 \leq \gamma \leq 0.5</math>)</b>					<b>Roll (<math>0 \leq \gamma \leq 0.5</math>)</b>							
	0.0	0.1	0.2	0.3	0.4	0.5	0.0	0.1	0.2	0.3	0.4	0.5	0.0	0.1	0.2	0.3	0.4	0.5
-90:-60	9.7	8.8	10.5	12.0	14.6	18.2	-	-	-	-	-	-	-	-	-	-	-	-
-60:-30	3.1	2.6	2.8	3.3	3.6	4.2	7.2	6.3	5.2	4.8	4.5	5.3	12.8	11.2	10.7	10.7	10.8	10.7
-30:0	2.4	2.2	2.2	2.1	2.3	2.5	5.4	5.1	4.9	4.4	4.2	5.6	2.8	2.8	2.8	2.8	2.7	2.8
0:30	2.5	2.4	2.3	2.4	2.5	2.4	4.0	3.9	4.0	4.0	4.2	8.5	2.4	2.2	2.2	2.5	2.8	3.1
30:60	4.5	4.2	4.3	4.1	4.3	5.0	3.7	3.9	4.6	5.2	5.6	5.9	12.6	11.2	11.4	12.7	14.3	16.5
60:90	10.9	10.3	10.0	10.5	11.8	14.9	6.2	4.3	4.3	4.6	5.1	6.3	32.0	31.4	32.5	35.5	37.8	39.4



(a) Yaw: error in degrees (y-axis) on different angle ranges (x-axis)



(b) Pitch: error in degrees (y-axis) on different angle ranges (x-axis)



(c) Roll: error in degrees (y-axis) on different angle ranges (x-axis)

Fig. 8: Using RAFA-Net, the effect of bounding box margin (control with parameter  $0 \leq \gamma \leq 0.5$ ) on the average pose estimation error in degrees (y-axis) on different angle ranges (x-axis) for yaw, pitch and roll. This model was trained on BIWI and tested on BIWI. It is observed that the yaw and roll lean towards inconsiderate to the  $\gamma$  values for smaller angles, however they tend to be sensitive for larger angles. For this dataset, it is also observed that pitch is sensitive to the  $\gamma$  values for a wide range of angles.

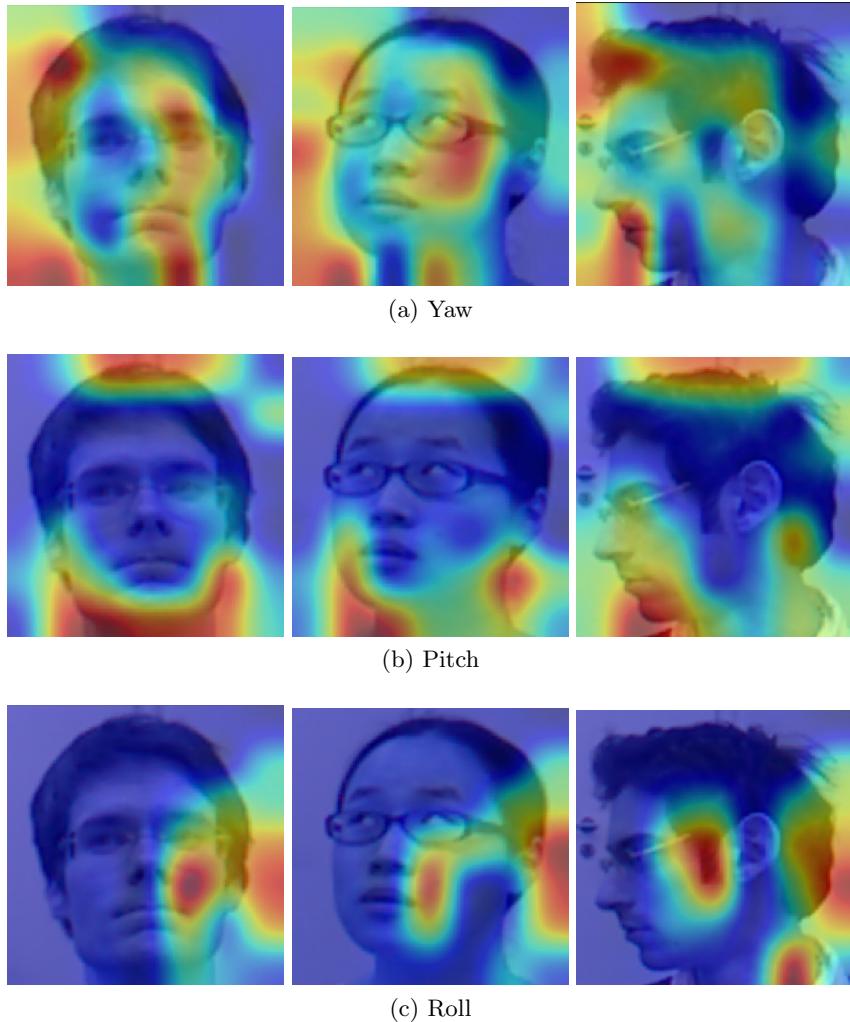


Fig. 9: More examples demonstrating the visualization of the proposed attentional spatial pooling using rotation axis specific class activation map. RAFA-Net is trained on BIWI and tested on BIWI. It shows the rotation axis specific representative features are used for head pose angle estimation. The trend is similar to example images of AFLW2000 (trained on 300W-LP and tested on AFLW2000) which are shown in the paper (Fig. 5).