

Table 1: **Quantitative Comparison on MHAD.** Scale is the resolution of image input to the 2D pose detector. The best results are highlighted in **bold**, and the second results are underlined. The results of the baseline comparison are in light blue , while the results of Dual-Diffusion are in dark blue .

Method	Venue	2D Pose Detector	Scale	MPJPE ↓ (mm)	BL ↓ (mm)	Sym ↓ (mm)	JDR ↑ (%)
TPPT [24]	ECCV'22	TPPT*	256	209.03	134.05	248.93	-
RSB-Pose50 [45]	arXiv'23	RSB-Pose50*	256	<u>32.10</u>	<u>10.21</u>	<u>12.13</u>	96.62
Epipolar_Tri [10]	CVPR'20	Epipolar_Tri*	256	90.73	33.67	34.21	-
Dual-Diffusion-Epi		Epipolar_Tri*	256	76.42 <u>14.31↓</u>	27.02 <u>6.65↓</u>	26.42 <u>7.79↓</u>	-
Tri-ViTpose	NeurIPS'22	ViTPose	256	70.84	42.55	48.43	95.83
Dual-Diffusion-ViT		ViTPose	256	61.02 <u>9.82↓</u>	37.90 <u>4.65↓</u>	30.09 <u>18.34↓</u>	95.88 <u>0.05↑</u>
Tri-ResNet50		ResNet50	256	60.04	23.68	36.65	95.95
Dual-Diffusion-ResNet50		ResNet50	256	54.51 <u>5.53↓</u>	18.09 <u>5.59↓</u>	24.64 <u>12.01↓</u>	98.86 <u>2.91↑</u>
Tri-RSB50		RSB-Pose50*	256	35.40	11.36	14.25	96.62
Dual-Diffusion-RSB50		RSB-Pose50*	256	30.96 <u>4.44↓</u>	9.60 <u>1.76↓</u>	11.53 <u>2.72↓</u>	98.94 <u>2.32↑</u>
Algebraic-Tri [14]	ICCV'19	ResNet152	384	51.69	27.11	45.69	95.95
RSB-Pose152 [45]	arXiv'23	RSB-Pose152*	384	29.33	8.70	9.94	97.40
AdaFuse [55]	IJCV'20	AdaFuse*	384	70.27	36.07	30.08	83.46
Dual-Diffusion-Ada		AdaFuse*	256	53.77 <u>16.50↓</u>	24.59 <u>11.48↓</u>	23.19 <u>6.89↓</u>	95.37 <u>11.91↑</u>
Tri-ResNet152		ResNet152	384	48.26	19.22	27.73	95.95
Dual-Diffusion-ResNet152		ResNet152	384	43.57 <u>4.69↓</u>	16.20 <u>3.02↓</u>	14.91 <u>12.82↓</u>	97.26 <u>1.31↑</u>
Tri-RSB152		RSB-Pose152*	384	29.78	9.84	11.61	97.40
Dual-Diffusion-RSB152		RSB-Pose152*	384	27.76 <u>2.02↓</u>	7.56 <u>2.28↓</u>	9.83 <u>1.78↓</u>	99.20 <u>1.80↑</u>

Table 2: **Quantitative Comparison on H36M.**

Method	Scale	Params (M) of Model other than Backbone	MPJPE (mm)	BL (mm)	Sym (mm)	JDR (%)
TPPT [24]	256	9.70	40.72	22.49	25.44	-
RSB-Pose50 [45]	256	9.25	<u>35.01</u>	<u>14.16</u>	<u>13.54</u>	<u>94.82</u>
Epipolar_Tri [10]	256	0.08	41.22	20.39	20.18	-
Dual-Diffusion-Epi	256	0.74	37.03 <u>4.19↓</u>	16.90 <u>3.49↓</u>	18.08 <u>2.10↓</u>	-
Tri-ViTpose [55]	256	-	41.49	18.09	20.75	93.33
Dual-Diffusion-ViT	256	0.74	35.20 <u>6.29↓</u>	16.02 <u>2.07↓</u>	19.66 <u>1.09↓</u>	95.77 <u>2.44↑</u>
Tri-RSB50	256	-	38.13	17.26	16.82	94.82
Dual-Diff-RSB50	256	0.74	33.17 <u>4.96↓</u>	12.29 <u>4.97↓</u>	11.75 <u>5.07↓</u>	94.91 <u>0.09↑</u>
Algebraic-Tri [14]	384	10.88	31.24	13.52	13.59	95.81
RSB-Pose152 [45]	384	9.25	<u>30.07</u>	<u>13.33</u>	<u>12.86</u>	<u>95.93</u>
AdaFuse [55]	384	1.02	30.27	15.23	14.36	94.25
Dual-Diffusion-Ada	256	0.74	29.17 <u>1.10↓</u>	13.85 <u>1.38↓</u>	13.57 <u>0.79↓</u>	96.06 <u>1.81↑</u>
Tri-RSB152	384	-	30.54	13.65	13.42	95.93
Dual-Diff-RSB152	384	0.74	28.67 <u>1.87↓</u>	12.06 <u>1.59↓</u>	12.35 <u>1.07↓</u>	95.97 <u>0.04↑</u>

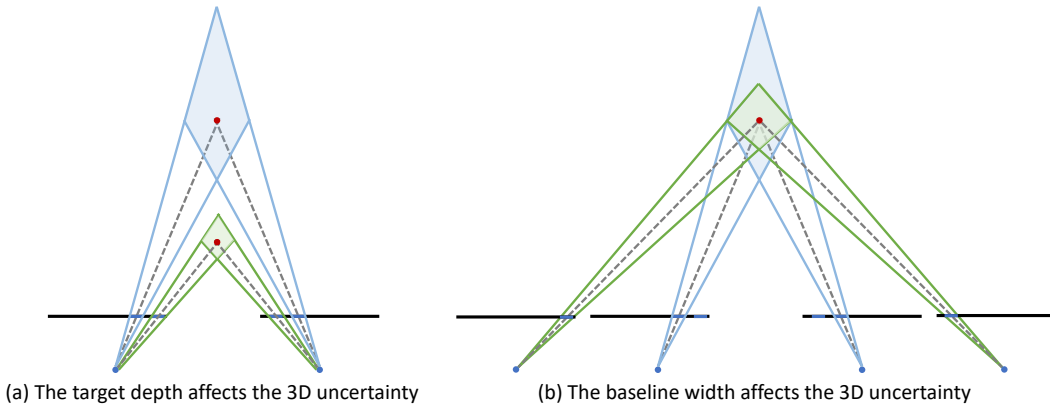


Figure 1: 3D uncertainty in a binocular setting is unstable due to factors such as (a) target depth and (b) baseline width.