Supplementary Materials

A. Deep Evidential Regression

First, we provide the derivation of the posterior predictive distribution of the data, given the Normal-Inverse-Gamma prior parameters as seen in [1, 33].

$$
\begin{aligned}
p(y_i|\mathbf{m}) &= \int_{\boldsymbol{\theta}} p(y_i|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{m}) d\boldsymbol{\theta} \\
&= \int_{\sigma^2=0}^{\infty} \int_{\mu=-\infty}^{\infty} p(y_i|\mu,\sigma^2) p(\mu,\sigma^2|\mu_0,\nu,\alpha,\beta) d\mu d\sigma^2 \\
&= \int_{\sigma^2=0}^{\infty} \int_{\mu=-\infty}^{\infty} \left[\sqrt{\frac{1}{2\pi\sigma^2}} \exp\left\{ -\frac{(y_i-\mu)^2}{2\sigma^2} \right\} \right] \\
&\quad \left[\frac{\beta^\alpha \sqrt{\nu}}{\Gamma(\alpha)\sqrt{2\pi\sigma^2}} \left(\frac{1}{\sigma^2}\right)^{\alpha+1} \exp\left\{ -\frac{2\beta+\nu(\mu_0-\mu)^2}{2\sigma^2} \right\} \right] d\mu d\sigma^2 \\
&= \int_{\sigma^2=0}^{\infty} \frac{\beta^\alpha \sigma - 3 - 2\alpha}{\sqrt{2\pi}\sqrt{1+1/\nu}\,\Gamma(\alpha)} \exp\left\{ -\frac{2\beta + \frac{\nu(y_i-\mu_0)^2}{1+\nu}}{2\sigma^2} \right\} d\sigma^2 \\
&= \frac{\Gamma(1/2+\alpha)}{\Gamma(\alpha)} \sqrt{\frac{\nu}{\pi}} (2\beta(1+\nu))^\alpha (\nu(y_i-\mu_0)^2 + 2\beta(1+\nu))^{-\left(\frac{1}{2}+\alpha\right)} \\
p(y_i|\mathbf{m}) &= \mathrm{t}_{2\alpha}\left(y_i;\mu_0,\frac{\beta(1+\nu)}{\nu\alpha}\right)
\end{aligned}
\tag{S1}
$$

Consequently, the loss function is the negative log-likelihood of this posterior:

$$
\begin{aligned}
\mathcal{L}_{nll}(\theta) &= -\log p(y_i|\mathbf{m}) \\
&= -\log \mathrm{t}_{2\alpha}\left(y_i;\mu_0,\frac{\beta(1+\nu)}{\nu\alpha}\right) \\
\mathcal{L}_{nll}(\theta) &= \log\Gamma(\alpha) - \log\Gamma\left(\alpha+\frac{1}{2}\right) + \frac{1}{2}\log\left(\frac{\pi}{\nu}\right) - \alpha\log(2\beta(1+\nu)) + \left(\alpha+\frac{1}{2}\right)\log((y_i-\mu_0)^2 + 2\beta(1+\nu))
\end{aligned}
\tag{S2}
$$

with $\nu > 0, \alpha > 2, \beta > 0$.

B. Multivariate Deep Evidential Regression

For the multivariate case, we use the following construction of the multivariate Student's t-distribution:

$$
\mathrm{t}_\nu(\boldsymbol{y}_i;\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{\Gamma[(\nu+n)/2]}{\Gamma(\nu/2)\nu^{n/2}\pi^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \left[1 + \frac{1}{\nu}(\boldsymbol{y}_i-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu}) \right]^{-(\nu+n)/2}
\tag{S3}
$$

Where n is the matrix dimension and $\frac{\nu}{\nu-2}\boldsymbol{\Sigma}$ is the covariance matrix for $\nu > 2$ degrees of freedom. We go on to plug in the parameters of the Normal-Inverse-Wishart prior as derived by [33]:

$$
\begin{aligned}
\mathrm{t}_{\nu_0-n+1}\left(\boldsymbol{y}_i;\boldsymbol{\mu}_0,\frac{\kappa+1}{\kappa(\nu_0-n+1)}\boldsymbol{\Psi}\right) &= \frac{\Gamma[(\nu_0+1)/2]}{\Gamma[(\nu_0-n+1)/2](\nu_0-n+1)^{n/2}\pi^{n/2}|\frac{\kappa+1}{\kappa(\nu_0-n+1)}\boldsymbol{\Psi}|^{1/2}} \\
&\quad \left[1 + \frac{1}{\nu_0-n+1}(\boldsymbol{y}_i-\boldsymbol{\mu}_0)^\top \left(\frac{\kappa+1}{\kappa(\nu_0-n+1)}\boldsymbol{\Psi}\right)^{-1} (\boldsymbol{y}_i-\boldsymbol{\mu}_0) \right]^{-(\nu_0+1)/2} \\
&= \frac{\Gamma[(\nu_0+1)/2]}{\Gamma[(\nu_0-n+1)/2](\nu_0-n+1)^{n/2}\pi^{n/2}|\frac{\kappa+1}{\kappa(\nu_0-n+1)}\boldsymbol{\Psi}|^{1/2}} \\
&\quad \left[\frac{|\boldsymbol{\Psi} + \frac{\kappa}{\kappa+1}(\boldsymbol{y}_i-\boldsymbol{\mu}_0)(\boldsymbol{y}_i-\boldsymbol{\mu}_0)^\top|}{|\boldsymbol{\Psi}|} \right]^{-(\nu_0+1)/2}
\end{aligned}
\tag{S4}
$$

We construct the corresponding negative log-likelihood as follows:

$$
\begin{aligned}
\mathcal{L}_{nll}(\theta) &= -\log p(\boldsymbol{y}_i|\mathbf{m}) \\
&= -\log\Gamma[(\nu_0+1)/2] + \log\Gamma[(\nu_0-n+1)/2] + \frac{n}{2}\log\pi(\nu_0-n+1) \\
&\quad + \frac{1}{2}\log|\frac{\kappa+1}{\kappa(\nu_0-n+1)}\boldsymbol{\Psi}| + \frac{(\nu_0+1)}{2}\log|\boldsymbol{\Psi} + \frac{\kappa}{\kappa+1}(\boldsymbol{y}_i-\boldsymbol{\mu}_0)(\boldsymbol{y}_i-\boldsymbol{\mu}_0)^\top| - \frac{\nu_0+1}{2}\log|\boldsymbol{\Psi}| \\
&= -\log\Gamma[(\nu_0+1)/2] + \log\Gamma[(\nu_0-n+1)/2] + \frac{n}{2}\log\pi(\nu_0-n+1) \\
&\quad - \frac{\nu_0}{2}\log|\boldsymbol{\Psi}| + \frac{n}{2}\log\frac{\kappa+1}{\kappa(\nu_0-n+1)} + \frac{(\nu_0+1)}{2}\log|\boldsymbol{\Psi} + \frac{\kappa}{\kappa+1}(\boldsymbol{y}_i-\boldsymbol{\mu}_0)(\boldsymbol{y}_i-\boldsymbol{\mu}_0)^\top| \\
&= -\log\Gamma[(\nu_0+1)/2] + \log\Gamma[(\nu_0-n+1)/2] + \frac{n}{2}\log\frac{\pi(\kappa+1)}{\kappa} \\
&\quad - \frac{\nu_0}{2}\log|\boldsymbol{\Psi}| + \frac{(\nu_0+1)}{2}\log|\boldsymbol{\Psi} + \frac{\kappa}{\kappa+1}(\boldsymbol{y}_i-\boldsymbol{\mu}_0)(\boldsymbol{y}_i-\boldsymbol{\mu}_0)^\top| \\
&= -\log\Gamma[(\nu_0+1)/2] + \log\Gamma[(\nu_0-n+1)/2] + \frac{n}{2}\log\frac{\pi(\kappa+1)}{\kappa} \\
&\quad - \nu_0\operatorname{tr}\log\boldsymbol{L} + \frac{(\nu_0+1)}{2}\log|\boldsymbol{\Psi} + \frac{\kappa}{\kappa+1}(\boldsymbol{y}_i-\boldsymbol{\mu}_0)(\boldsymbol{y}_i-\boldsymbol{\mu}_0)^\top| \\
\mathcal{L}_{nll}(\theta) &= \log\Gamma\left[\frac{(\nu_0-n+1)}{2}\right] - \log\Gamma\left[\frac{(\nu_0+1)}{2}\right] + \frac{n}{2}\log\frac{\pi(\kappa+1)}{\kappa} \\
&\quad - \nu\operatorname{tr}\log\boldsymbol{L} + \frac{(\nu_0+1)}{2}\log|\boldsymbol{\Psi} + \frac{\kappa}{\kappa+1}(\boldsymbol{y}_i-\boldsymbol{\mu}_0)(\boldsymbol{y}_i-\boldsymbol{\mu}_0)^\top|
\end{aligned}
\tag{S5}
$$

In practice, we estimate the lower triangular matrix $\boldsymbol{L}$, where $\boldsymbol{\Psi} = \boldsymbol{L}\boldsymbol{L}^\top$, and employ the $\mathrm{softplus}(\cdot)$ activation function to ensure $\boldsymbol{L}_{ii} > 0$. Unlike [32], we use $\mathbb{E}[\boldsymbol{\Sigma}] = \boldsymbol{\Psi}/(\nu_0 - n - 1)$, i.e., the expectation of the $\boldsymbol{\Sigma}$ parameter, defined by the IW-prior, as our aleatoric variance, in line with the univariate interpretation of [1].

C. SIM dataset

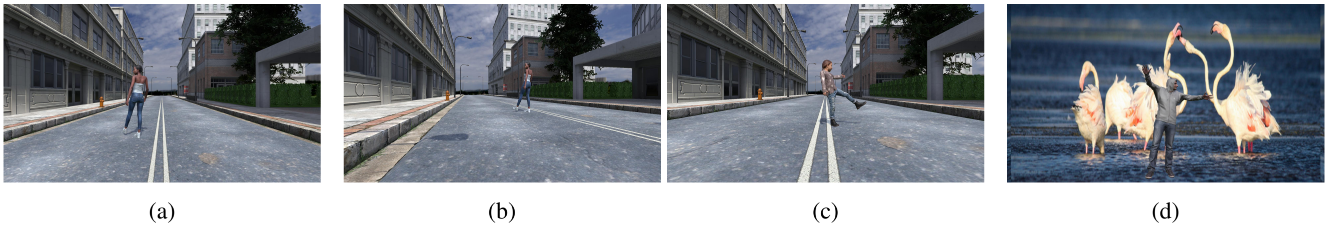
(a)
(b)
(c)
(d)

Figure S1. Samples from from the SIM dataset, showing different camera views (a/b), subjects (c), and backgrounds (d).

The SIM dataset was created internally and is subject to an earlier publication [5]. It is available online[1] but not a contribution of this work. The training subset features four distinct subjects performing a routine of varied motion-captured movements in a simulated street environment created with Unity. One of the subjects appears in front of randomly sampled MS-COCO images instead of the street environment. There are 11 camera views for each subject and scene resulting in close to 140.000 samples, 25% of which we use for training in the final 10 epochs. The accompanying validation set contains an additional 11 subjects, performing a simple walking-cycle animation in the same street scene. The validation set contains about 66.000 samples, 10% of which are used in the validation.

D. System details and execution speeds

All models were trained on the same machine and on a single RTX A6000 gpu. A full training run lasted anywhere from 8-24h. Training speeds increased from an average of $\sim$ 4.1it/s for MAP to $\sim$ 11.5it/s for DER, while inference speeds increased from about $\sim$ 7.9it/s to an average of $\sim$ 23it/s respectively. Naturally, the immense discrepancies in training and inference speeds are owed to the resampling step required for the estimation of epistemic uncertainty in MAP.

[1] See https://github.com/noboevbo/PedRec, in the ”Datasets” section.

E. Univariate Input Evidential Regression

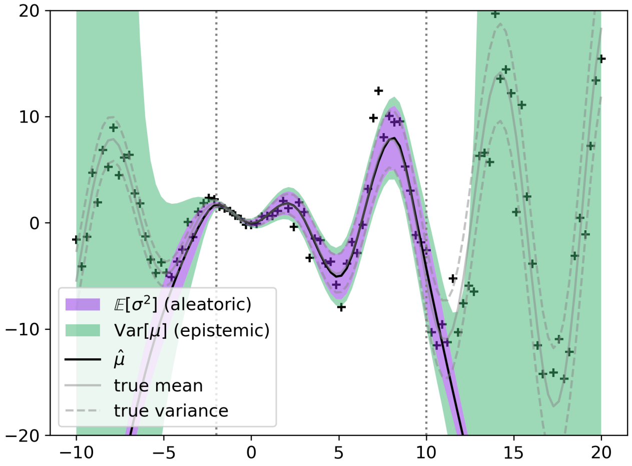

Figure S2. A simple deep evidential regression example featuring one input and one output variable. During training, the model sees only samples from the interval between the two dotted lines.

A simple one-independent, one-dependent regression problem serves as an illustrative example of the behavior of DER. Samples of the target function in the following example are corrupted by both heteroscedastic and homoscedastic noise:

$$\begin{aligned} f(x) &= x\sin(x) + x\epsilon + \epsilon \\ \text{where } \epsilon &\sim \mathcal{N}(0, 0.3) \end{aligned} \tag{S6}$$

Additionally, the training data contains only samples from the input interval $X \in [-2.0, 10.0]$, but the estimated function and its uncertainty estimates are evaluated in the interval $X \in [-10.0, 20.0]$. Note, that the estimated aleatoric variance (purple shaded area) captures the true variance almost perfectly, within the domain of the training data. In other words, in this simple example, there does not appear to be a problem of overconfidence. Outside of the training data, aleatoric estimates remain approximately stable, while epistemic uncertainty estimates (green shaded area) increase drastically.

F. MAP ablation

	(uni) MAP	(aleatoric-only) MAP	(epistemic-only) MCD
pck02 ↑			
MS-COCO	88.95	86.34	86.36
H36M	92.35	90.66	90.19
MPII	85.53	81.24	81.07
SIM	99.56	98.68	99.83
*BDD100k	65.77	63.75	65.04
mpjpe ↓			
H36M	74.41	82.54	80.61
SIM	14.39	31.06	19.32
ECE ↓	al./ep.		
MS-COCO	.033/.008	.031	.007
H36M	.037/.037	.036	.035
MPII	.033/.007	.027	.004
SIM	.039/.033	.042	.035
*BDD100k	.020/.044	.021	.039

Table S5. Performance metrics and expected calibration error for the univariate MAP model, combining MAP and MCD for simultaneous aleatoric and epistemic uncertainty estimation, and two ablation models, estimating either aleatoric or epistemic uncertainty. Note, that the ECE for the univariate model is given for both types of uncertainty, separated by a '/'.

Various dynamics in probabilistic machine learning may induce subpar final performance results. E.g., [37] suggest that the use of predictive variance could lead to premature convergence on hard-to-predict samples, as it is much easier for the model to predict a large variance than to accurately predict the output in some cases. Likewise, dropout-variational inference effectively increases the bias of the model through repeated sampling and aggregation, which may negatively impact performance. Hence, we conduct a series of ablation trials using MAP-only and MCD-only models to predict aleatoric and epistemic uncertainties exclusively and report ECE scores as well as 2d- and 3d-performance metrics.

We find that, contrary to our expectations, both ablation models perform worse in the absence of one of the two uncertainty quantification methods. Both models perform closer to the heatmap-based baseline model, except for a significant performance bump on the unseen BDD100k dataset, suggesting that the added bias improved generalization performance. Calibration scores for the respective uncertainty metrics remained largely unaffected, displaying only a slight increase in quality.

G. Spearman's Rank Correlation

		aleatoric $\mathbb{E}[\sigma^2]$				epistemic $\mathrm{Var}[\mu]$			
$\rho \uparrow$	joint	DER	(mv) DER	MAP	(mv) MAP	DER	(mv) DER	MAP	(mv) MAP
head	nose	**0.67**	0.64	0.65	0.61	**0.68**	0.67	0.45	0.44
	l-eye	**0.56**	**0.56**	0.55	0.48	**0.58**	0.50	0.39	0.33
	r-eye	**0.58**	0.55	0.56	0.53	**0.60**	0.52	0.39	0.37
	head-upper	0.34	0.30	**0.40**	0.30	**0.43**	0.40	0.25	0.21
	head-lower	0.28	0.25	0.25	**0.33**	**0.27**	0.23	0.25	0.25
	neck	0.35	**0.41**	0.27	0.31	0.35	**0.36**	0.24	0.26
arms	l-shoulder	0.51	0.50	**0.55**	0.54	**0.52**	0.50	0.38	0.36
	r-shoulder	0.52	0.52	**0.54**	0.53	**0.52**	0.51	0.35	0.36
	l-elbow	0.62	0.61	**0.63**	0.62	**0.62**	0.57	0.47	0.43
	r-elbow	0.60	0.62	**0.63**	0.60	**0.62**	0.55	0.45	0.41
	l-wrist	0.66	**0.67**	0.66	0.66	**0.67**	0.64	0.52	0.48
	r-wrist	0.65	0.62	**0.67**	0.65	**0.66**	0.60	0.51	0.44
torso	spine-center	0.44	0.49	0.44	**0.51**	0.43	**0.49**	0.35	0.38
	hip-center	0.40	0.45	0.43	**0.48**	0.44	**0.51**	0.36	0.40
legs	l-hip	0.49	0.53	0.51	**0.55**	0.51	**0.52**	0.41	0.42
	r-hip	0.50	0.52	0.51	**0.53**	0.51	**0.53**	0.42	0.45
	l-knee	0.60	0.59	0.60	0.60	**0.62**	0.58	0.51	0.50
	r-knee	0.59	0.59	0.59	0.59	0.60	**0.61**	0.51	0.50
	l-ankle	0.47	**0.51**	0.46	0.49	0.48	**0.51**	0.31	0.28
	r-ankle	0.44	0.43	**0.49**	**0.49**	**0.46**	0.45	0.33	0.29

Table S6. Per joint Spearman's rank correlation coefficient between each models predicted aleatoric and epistemic uncertainty and the corresponding residual. The data is averaged over all datasets and each predicted coordinate.

H. Per Dataset Examples

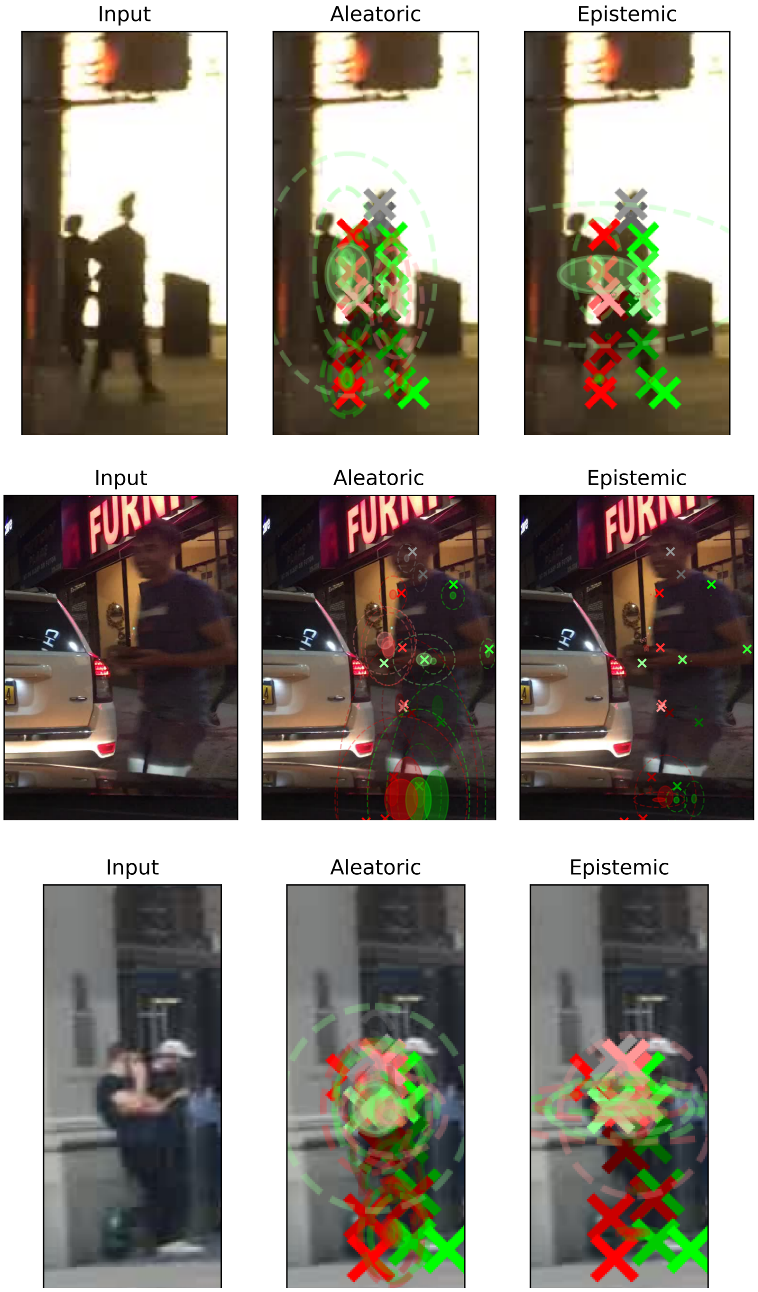

Figure S3. Qualitative results on samples from the **BDD100k** dataset. Xs mark the ground-truth position, shaded and dotted ellipses show the original and recalibrated predicted position areas. This dataset has not been seen during training, hence, epistemic uncertainty is highest across samples amongst all datasets used, as we would expect. The first example shows a case where the model confuses the orientation of the person, consequently flipping joints left-to-right. The second image, apart from illustrating higher aleatoric uncertainty due to motion blur, shows that in the absence of corresponding training data, occlusions do lead to increases in epistemic uncertainty. The final image exemplifies how the model struggles to disambiguate low-contrast proximity which might plausibly be alleviated through more training data (note that even the ground-truth labels appear to merge joints of both persons).

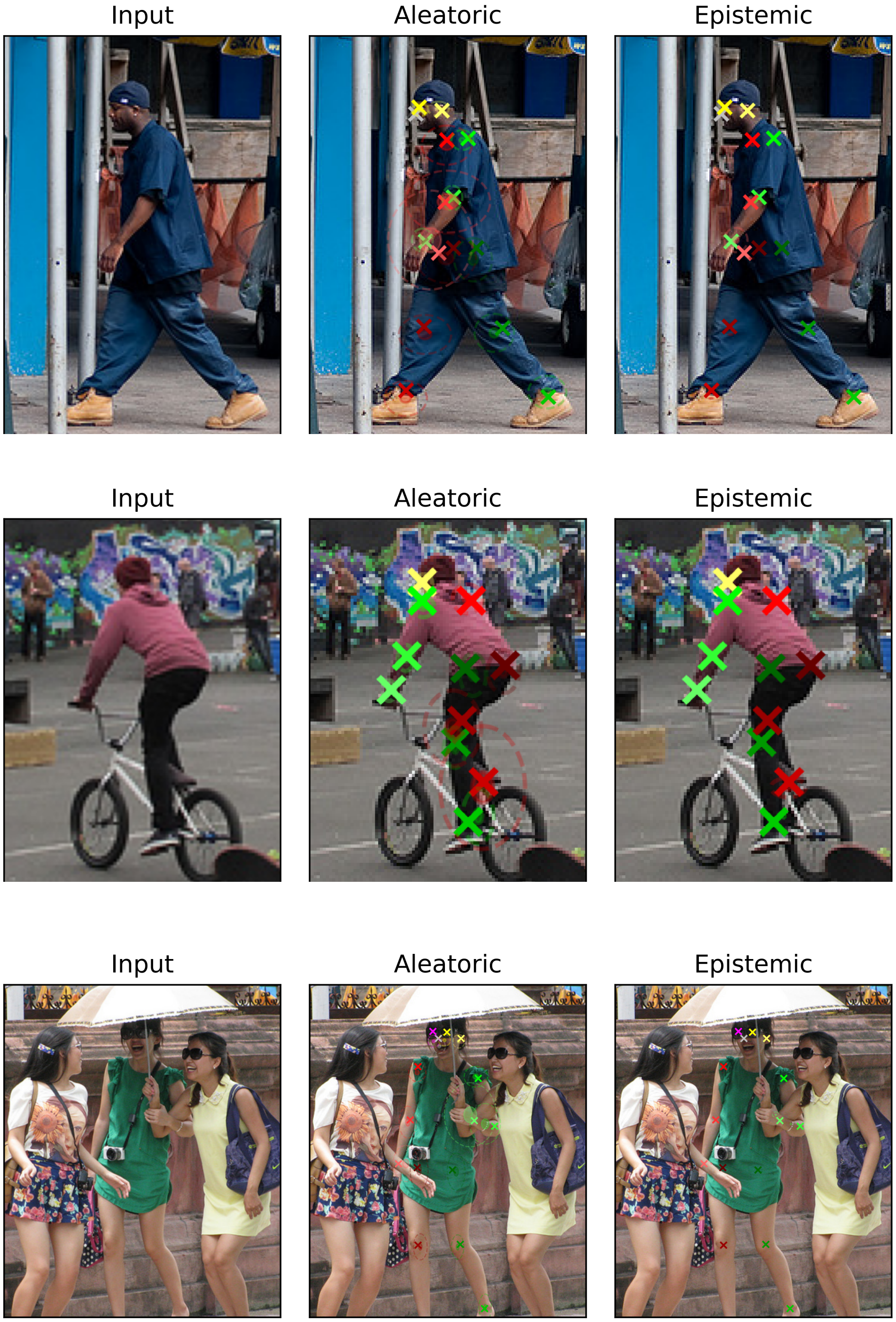

Figure S4. Qualitative results on validation samples from the **COCO** dataset. Occlusions (first two images) and ambiguous situations (arms, final image) are a visible source of aleatoric uncertainty. Epistemic uncertainty is generally low across all COCO validation samples.

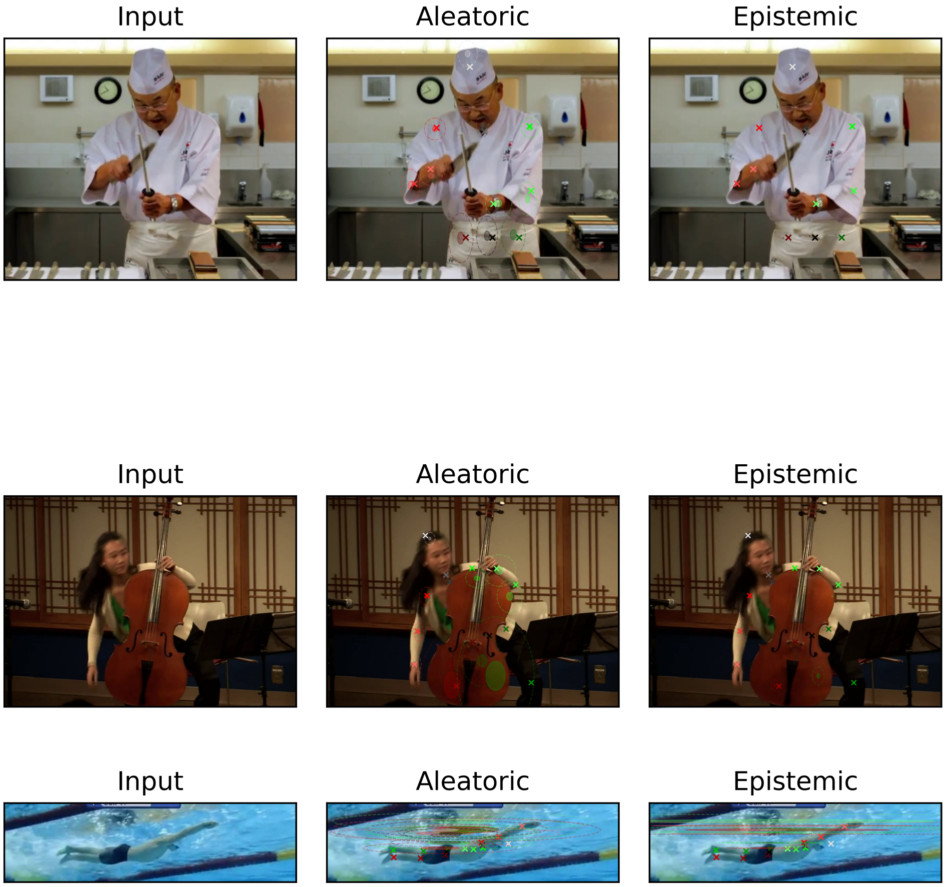

Figure S5. Qualitative results on validation samples from the **MPII** dataset. The model behaves much the same way as it does on the COCO dataset. Note, that low contrast in the second image leads to the model expecting the left leg behind the musical instrument. The final image illustrates the comparatively underrepresented case of a human body on its side. While aleatoric uncertainty is unexpectedly large, it is still outweighed by the plausible epistemic uncertainty.

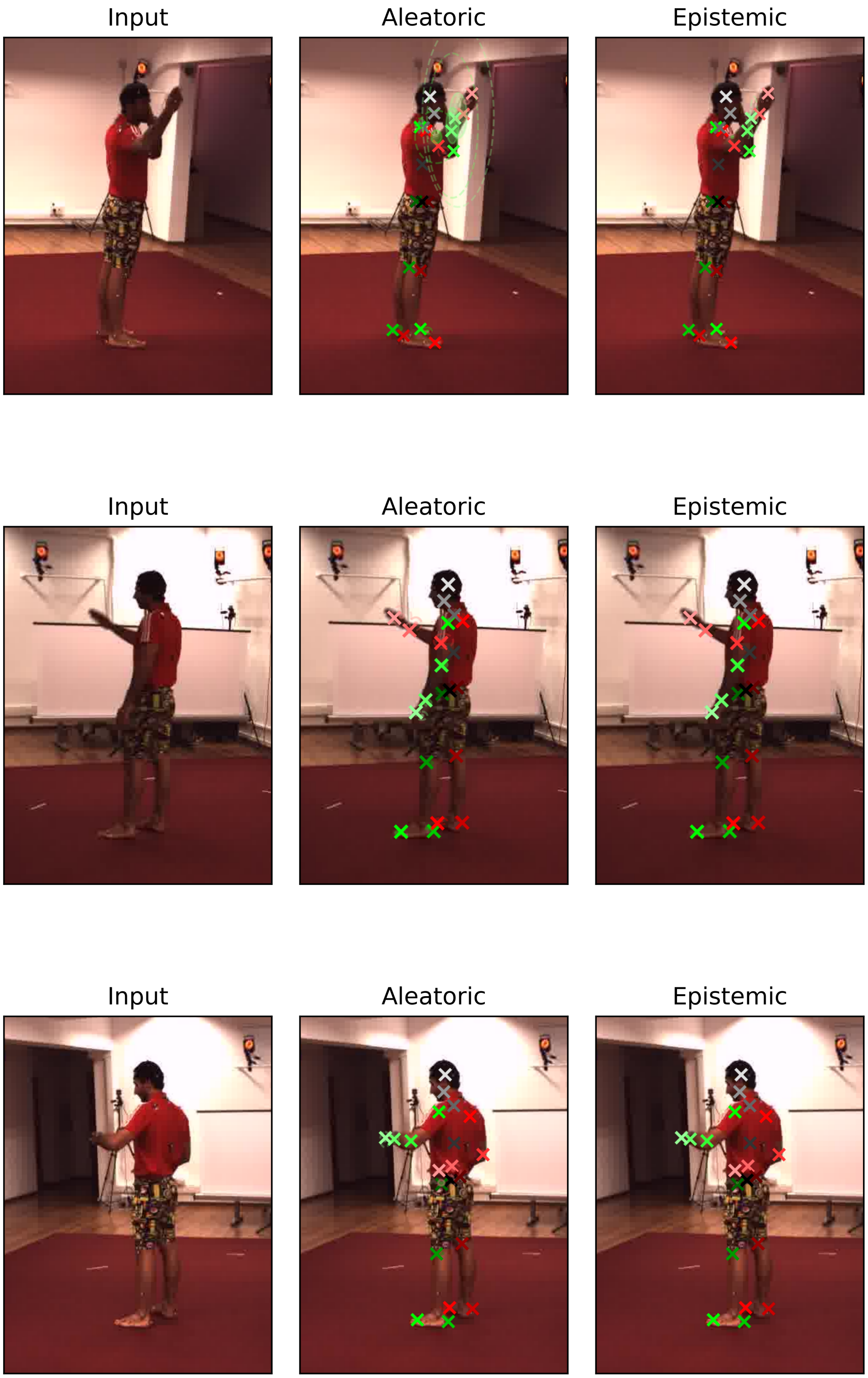

Figure S6. Qualitative results on validation samples from the **H36M** dataset. Given the unobstructed nature of the dataset, estimated uncertainties remain low. One of the most common sources for aleatoric uncertainty is self-occlusion.

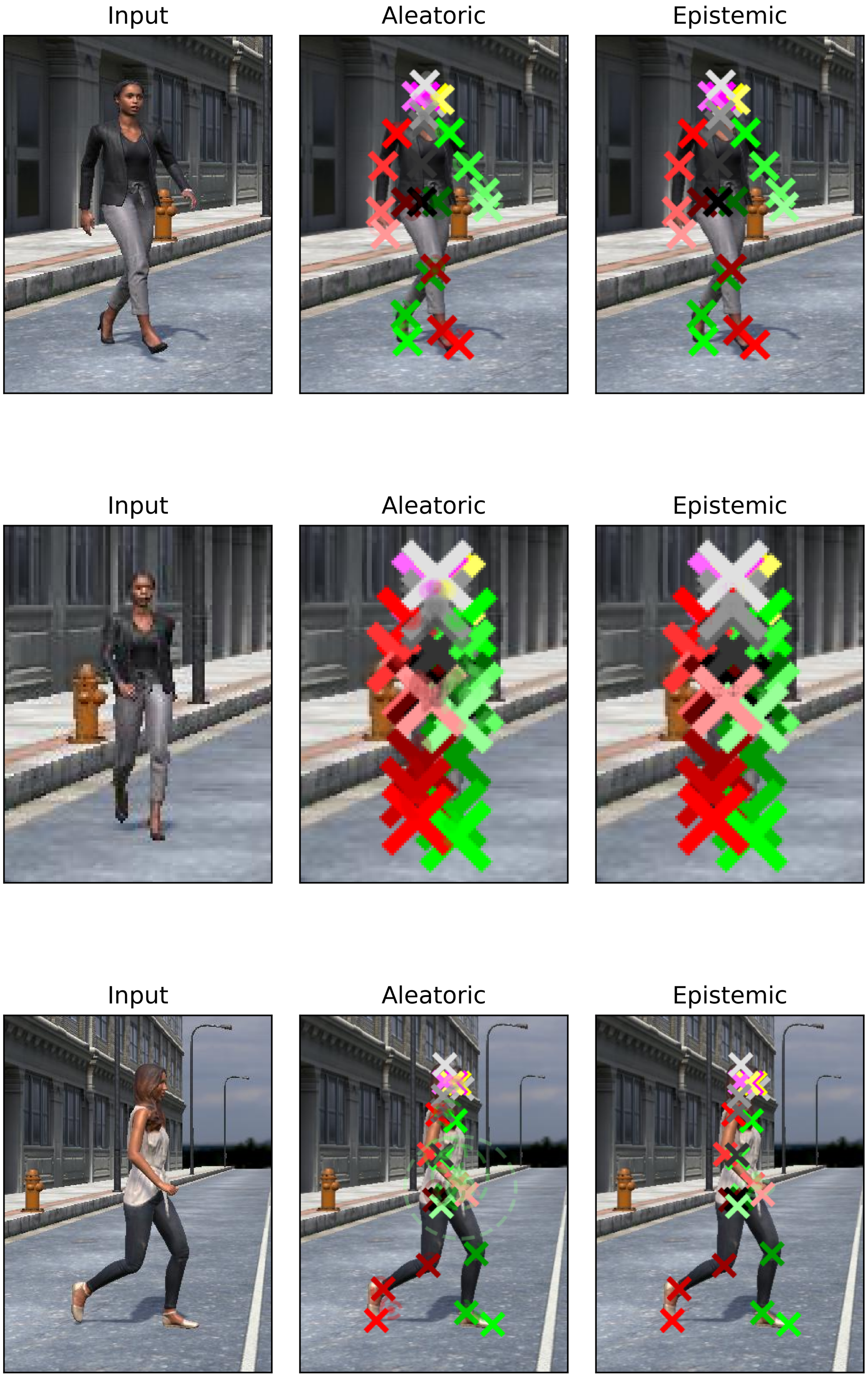

Figure S7. Qualitative results on validation samples from the **SIM** dataset. The estimated epistemic uncertainty in this dataset is very low due to the similarity of the samples, which only replace characters but do not feature very diverse poses. The most prevalent sources of aleatoric uncertainty are again self-occlusion and compression artifacts of very small appearing persons.