Uncertainty Quantification in Depth Estimation via Constrained Ordinal Regression (ECCV 2022)

2023.11.23

Contribution & Motivation

1. predictive variance = error variance (数据固有的不确定性，即输入图像无法解释的深度变化，也称为任意不确定性) + estimation variance (有限数据训练导致的网络参数的随机性，通常称为认知不确定性)

2. 离散深度值执行约束序数回归（ConOR）来学习无似然条件分布。可以将条件分布估计器的期望作为预测深度，将方差作为 error variance 的估计。

3. 通过根据重新采样的数据集计算的深度估计的样本方差来近似 estimation variance。

1. DORN (Deep Ordinal Regression Network) (CVPR, 2018)

2. What uncertainties do we need in bayesian deep learning for computer vision (NIPS, 2017) **TWO VARIANCES (NOT THE FIRST)**

3. Estimating depth from monocular images as classification using deep fully convolutional residual networks (IEEE Trans, 2017) **MCC**

4. Inferring Distributions Over Depth from a Single Image (IROS, 2019) **BC**

5. Estimating the mean and variance of the target probability distribution(ICNN, 1994) **GL, LGL**

5. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning (ICML, 2016) **MCD**

6. Simple and scalable predictive uncertainty estimation using deep ensembles (NIPS, 2017) **Deep Ensemble (DE)**

2

贝叶斯神经网络，简单来说可以理解为通过为神经网络的权重引入不确定性进行正则化（regularization），也相当于集成（ensemble）某权重分布上的无穷多组神经网络进行预测。

Review: MCC

$$L(\boldsymbol{\mu}|y^*) = -\sum_{k=1}^{K} \mathbb{1}(k = y^*) \log \mu_k(\mathbf{x}; \mathbf{w}).$$

简单的交叉熵函数无法捕捉越远应该惩罚越大'

可加系数

pred = [0.2, 0,7, 0.1]
gt = [0, 1, 0]

k是bin

Review: BC

$$L(\boldsymbol{\mu}|y^*) = -\sum_{k=1}^{K}[\tilde{q}(k;y^*)\log\mu_k(\mathbf{x};\mathbf{w}) + (1-\tilde{q}(k;y^*))(\log(1-\mu_k(\mathbf{x};\mathbf{w})))], \quad (4)$$

:re $\tilde{q}(k;y^*) = e^{\frac{\|k-y^*\|^2}{-2\sigma^2}}$ is an *unnormalized* version of soft

我们进一步将连续深度建模为 K 个独立伯努利随机变量 yk ~ B(1, μk) 的集合（不就是二项分布？），其中 μk 编码落入第 k 个深度区间的概率。

直接用交叉熵做分类的话，错误的情况无法区分。
确保远离真实标签的预测比接近真实标签的预测会受到更大的惩罚

Review: DORN

$$\mathcal{L}(\chi,\Theta) = -\frac{1}{\mathcal{N}}\sum_{w=0}^{W-1}\sum_{h=0}^{H-1}\Psi(w,h,\chi,\Theta),$$

$$\Psi(h,w,\chi,\Theta) = \sum_{k=0}^{l_{(w,h)}-1}\log(\mathcal{P}^k_{(w,h)}) + \sum_{k=l_{(w,h)}}^{K-1}(\log(1-\mathcal{P}^k_{(w,h)})),$$

$$\mathcal{P}^k_{(w,h)} = P(\hat{l}_{(w,h)} > k|\chi,\Theta),$$

$Y = \psi(\chi,\Theta)$ of size $W \times H \times 2K$

$$\mathcal{P}^k_{(w,h)} = \frac{e^{y_{(w,h,2k+1)}}}{e^{y_{(w,h,2k)}} + e^{y_{(w,h,2k+1)}}}, \quad (3)$$

where $y_{(w,h,i)} = \theta_i^T x_{(w,h)}$, and $x_{(w,h)} \in \chi$. Minimizing

$$\hat{d}_{(w,h)} = \frac{t_{\hat{l}_{(w,h)}} + t_{\hat{l}_{(w,h)}+1}}{2} - \xi,$$

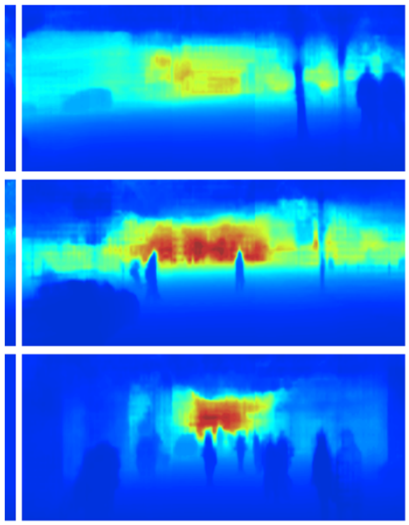
$$\hat{l}_{(w,h)} = \sum_{k=0}^{K-1}\eta(\mathcal{P}^k_{(w,h)} >= 0.5).$$

代表了预测的该像素点的深度值大于第k个阈值的概率。
l(w, h)则是GT label，损失函数希望让P(w,h)趋近于[1, 1, 1, 0, 0]，

5

直接用交叉熵做分类的话，错误的情况无法区分。
确保远离真实标签的预测比接近真实标签的预测会受到更大的惩罚

Review: DORN



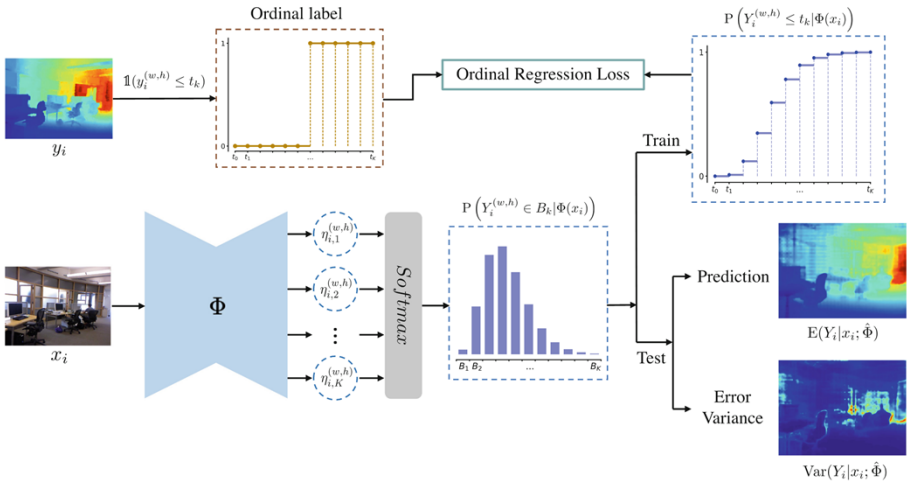
DORN

Method	cap	higher is better			lower is better			
		$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$	Abs Rel	Squa Rel	RMSE	RMSE_{log}
Make3D [51]	0 - 80 m	0.601	0.820	0.926	0.280	3.012	8.734	0.361
Eigen *et al.* [12]	0 - 80 m	0.692	0.899	0.967	0.190	1.515	7.156	0.270
Liu *et al.* [40]	0 - 80 m	0.647	0.882	0.961	0.217	1.841	6.986	0.289
LRC (CS + K) [19]	0 - 80 m	0.861	0.949	0.976	0.114	0.898	4.935	0.206
Kuznietsov *et al.* [33]	0 - 80 m	0.862	0.960	0.986	0.113	0.741	4.621	0.189
DORN (VGG)	0 - 80 m	0.915	0.980	0.993	0.081	0.376	3.056	0.132
DORN (ResNet)	0 - 80 m	**0.932**	**0.984**	**0.994**	**0.072**	**0.307**	**2.727**	**0.120**
Garg *et al.* [17]	0 - 50 m	0.740	0.904	0.962	0.169	1.080	5.104	0.273
LRC (CS + K) [19]	0 - 50 m	0.873	0.954	0.979	0.108	0.657	3.729	0.194
Kuznietsov *et al.* [33]	0 - 50 m	0.875	0.964	0.988	0.108	0.595	3.518	0.179
DORN (VGG)	0 - 50 m	0.920	0.982	0.994	0.079	0.324	2.517	0.128
DORN (ResNet)	0 - 50 m	**0.936**	**0.985**	**0.995**	**0.071**	**0.268**	**2.271**	**0.116**

6

直接用交叉熵做分类的话，错误的情况无法区分。
确保远离真实标签的预测比接近真实标签的预测会受到更大的惩罚

Method: ConOR



$$\ell\left(x_i, y_i^{(w,h)}, \Phi\right) = -\sum_{k=1}^{K}\left\{\mathbb{1}\left(y_i^{(w,h)} \le t_k\right)\log\left(\mathrm{P}(Y_i^{(w,h)} \le t_k|\Phi(x_i))\right)\right.$$
$$\left. + \left[1 - \mathbb{1}\left(y_i^{(w,h)} \le t_k\right)\right]\left[1 - \log\left(\mathrm{P}(Y_i^{(w,h)} \le t_k|\Phi(x_i))\right)\right]\right\},$$

Method: Predict: Error Variance Inference

$$\hat{g}^{(w,h)}(x_*) = \mathrm{E}\left[Y_*^{(w,h)}|x_*;\hat{\Phi}\right] = \sum_{k=1}^{K} \mu_k \mathrm{P}\left(Y_*^{(w,h)} \in B_k|\hat{\Phi}(x_*)\right),$$

$$\widehat{V}^{(w,h)}(x_*) = \mathrm{Var}\left[Y_*^{(w,h)}|x_*;\hat{\Phi}\right] \tag{11}$$

$$= \sum_{k=1}^{K} \left(\mu_k - \mathrm{E}\left[Y_*^{(w,h)}|x_*;\hat{\Phi}\right]\right)^2 \mathrm{P}\left(Y_*^{(w,h)} \in B_k|\hat{\Phi}(x_*)\right). \tag{12}$$

Hence our ConOR can predict the depth value together with error variance in the test phase.

$$y = \mathtt{argmin}_y F(y, \mathbf{p}) = \mathtt{argmin}_y \int |y - x| p(x; \mathbf{p}) dx.$$

Method: Estimation Variance Inference

$$\frac{1}{M-1}\sum_{m=1}^{M}\left(\mathrm{E}\left[Y_*^{(w,h)}|x_*;\hat{\Phi}_m\right]-\frac{1}{M}\sum_{j=1}^{M}\mathrm{E}\left[Y_*^{(w,h)}|x_*;\hat{\Phi}_j\right]\right)^2.$$

$$\upsilon_{m,i}^{(w,h)}=\hat{y}_i^{(w,h)}+\hat{\epsilon}_i^{(w,h)}\cdot\tau_{m,i}^{(w,h)}, \tag{15}$$

where $\tau_{m,i}^{(w,h)}$ is sampled from standard Gaussian distribution. For each replicate, we train the model on the new sampled training set:

$$\hat{\Phi}_m=\underset{\Phi}{\operatorname{argmin}}\sum_{i=1}^{n}\sum_{w=1}^{W}\sum_{h=1}^{H}\ell\left(x_i,\upsilon_{m,i}^{(w,h)},\Phi\right),\ \text{for } m=1,2,\ldots,M, \tag{16}$$

$$\hat{\Phi}_m=\underset{\Phi}{\operatorname{argmin}}\sum_{i=1}^{n}\sum_{w=1}^{W}\sum_{h=1}^{H}\omega_i^{(w,h)}\ell(x_i,y_i^{(w,h)},\Phi),\ \text{for } m=1,2,\ldots,M,$$

Experiment

Method	Prediction			Uncertainty: AUSE(ξ) ↓			Uncertainty: AURG(ξ) ↑		
	rmse↓	rel↓	δ_1 ↑	rmse	rel	$1-\delta_1$	rmse	rel	$1-\delta_1$
MCC [6,25]	3.011	0.081	0.915	0.180	0.421	0.566	0.673	0.248	0.460
BC [86]	2.878	0.078	0.919	0.179	0.292	0.304	0.674	0.398	0.658
GL+MCD [28]	3.337	0.102	0.875	0.111	0.216	0.137	0.726	0.456	0.787
GL+DE [51]	2.900	0.089	0.908	0.100	0.233	0.131	0.751	0.447	0.829
GL+WBS	3.064	0.083	0.906	0.095	0.243	0.132	0.739	0.433	0.818
GL+MBS	3.064	0.083	0.906	0.096	0.242	0.131	0.739	0.435	0.817
LGL+MCD [28]	3.219	0.158	0.836	0.160	0.531	0.452	0.558	0.146	0.558
LGL+DE [51]	2.852	0.132	0.873	0.159	0.548	0.397	0.538	0.132	0.601
LGL+WBS	2.965	0.132	0.870	0.212	0.528	0.396	0.559	0.130	0.602
LGL+MBS	2.965	0.132	0.870	0.158	0.524	0.384	0.557	0.131	0.597
ConOR+WBS	**2.709**	**0.075**	**0.928**	0.095	0.181	0.107	**0.754**	0.500	0.849
ConOR+MBS	**2.709**	**0.075**	**0.928**	**0.094**	**0.180**	**0.106**	**0.754**	**0.501**	**0.851**

为了进行比较，我们实现高斯似然（GL）和对数高斯似然（LGL）来估计error variance，并应用蒙特卡洛辍学（MCD）[28]和深度集成（DE）[51]来近似estimation variance。
继之前的工作[42,51]之后，我们在贝叶斯框架下设计的GL和LGL上采用MCD[28]和DE[51]。我们还在我们的框架中使用 WBS 和 MBS 实现了 Gaussian 和 Log Gaussian。
多类分类[6,25]（MCC）和二元分类[86]（BC），应用与我们相同的深度离散化策略。使用softmax置信度（MCC）和熵（BC）通常被视为完全不确定性[37]，因此它们不适合任何框架。我们确保重新实现的模型具有与我们相同的架构，但只有不同的预测头。

MC dropout 的 MC 体现在我们需要对同一个输入进行多次前向传播过程，这样在 dropout 的加持下可以得到“不同网络结构”的输出，将这些输出进行平均和统计方差，即可得到模型的预测结果及 uncertainty。而且，这个过程是可以并行的，所以在时间上可以等于进行一次前向传播。

神经网络产生的 softmax 概率不能表示 uncertainty?
但是，softmax 值并不能反应该样本分类结果的可靠程度。A model can be uncertain in its predictions even with a high softmax output. [1]
以 MNIST 分类为例，当模型在验证集上面效果很烂的时候，将一张图片输入到神经网络，我们仍然可以得到很高的 softmax 值，这个时候分类结果并不可靠；当模型在验证集上效果很好了，在测试集上甚至都很好，这个时候，我们将一

张图片加入一些噪声，或者手写一个数字拍成照片，输入到网络中，这个时候得到一个较高的 softmax 值，我们就认为结果可靠吗？我们这个时候可以理解为，在已知的信息中，模型认为自己做的挺好，而模型本身并不能泛化到所有样本空间中去，对于它没有见过的数据，它的泛化能力可能不是那么强，这个时候模型仍然是以已知的信息对这个没有见过的数据有很强的判断（softmax 某一维值很大），当然有时候判断很好，但有时候判断可能就有误，而模型并不能给出对这个判断有多少 confidence。
而 MC dropout 可以给出一个预测值，并给出对这个预测值的 confidence，这也就是贝叶斯深度学习的优势所在。

对于不确定性估计的比较，由于没有真实标签，我们遵循稀疏化误差的思想[39]。也就是说，当逐渐删除具有最高不确定性的像素时，误差应该单调减少。
因此，给定误差度量 xi，我们根据估计不确定性的降序排列像素迭代地删除子集 (1%)，并计算剩余像素的 Σ 以绘制曲线。
理想的稀疏化（oracle）是通过按真实误差的降序对像素进行排序来获得的；因此，我们通过稀疏化误差下的面积（AUSE）来衡量估计稀疏化和预言稀疏化之间的差异[39]。
我们还计算随机增益下的面积（AURG）[67]，它测量估计的稀疏化和没有不确定性建模的随机稀疏化之间的差异。我们采用均方根误差（rmse）、绝对相对误差（rel）和 1 − δ1 作为 Σ。为了公平比较，AUSE 和 AURG 都根据所考虑的指标进行归一化，以消除预测准确性的因素[39]。

Ablation

Dataset	Variance	AUSE(ξ) ↓			AURG(ξ) ↑		
		rmse	rel	$1-\delta_1$	rmse	rel	$1-\delta_1$
KITTI	Error	0.097	0.184	0.109	0.751	0.496	0.846
	Estimation (WBS)	0.103	0.188	0.132	0.745	0.493	0.823
	Estimation (MBS)	0.101	0.183	0.120	0.745	0.498	0.835
	Predictive (w/ WBS)	0.095	0.181	0.107	**0.754**	0.500	0.849
	Predictive (w/ MBS)	**0.094**	**0.180**	**0.106**	**0.754**	**0.501**	**0.851**
NYUv2	Error	0.305	0.350	0.349	0.333	0.235	0.544
	Estimation (WBS)	0.340	0.370	0.415	0.297	0.215	0.478
	Estimation (MBS)	0.326	0.365	0.396	0.311	0.220	0.497
	Predictive (w/ WBS)	**0.297**	**0.340**	**0.333**	**0.343**	**0.245**	**0.559**
	Predictive (w/ MBS)	**0.297**	0.343	0.336	0.340	0.243	0.557

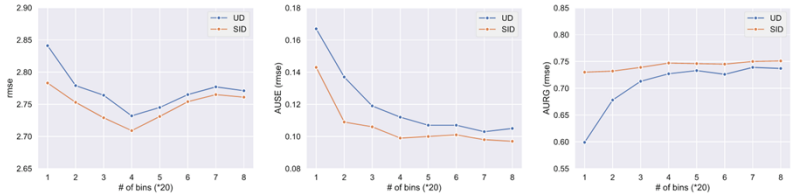
Dataset	Method	Prediction			AUSE(ξ) ↓			AURG(ξ) ↑		
		rmse ↓	rel ↓	δ_1 ↑	rmse	rel	$1-\delta_1$	rmse	rel	$1-\delta_1$
KITTI	GL	3.064	0.083	0.906	0.103	0.259	0.143	0.734	0.423	0.802
	LGL	2.965	0.132	0.870	0.157	0.540	0.427	0.557	0.135	0.602
	OR [24]	2.766	0.095	0.919	0.108	0.261	0.117	0.694	0.335	0.834
	ConOR	**2.709**	**0.075**	**0.928**	**0.097**	**0.184**	**0.109**	**0.751**	**0.496**	**0.846**
NYUv2	GL	0.534	0.171	0.770	0.344	0.413	0.528	0.258	0.167	0.330
	LGL	0.756	0.221	0.618	0.370	0.675	0.859	0.198	-0.150	-0.116
	OR [24]	0.509	0.146	0.814	0.314	0.392	0.411	0.289	0.172	0.468
	ConOR	**0.490**	**0.132**	**0.832**	**0.305**	**0.350**	**0.349**	**0.333**	**0.235**	**0.544**

这表明ConOR（任意不确定性）估计的误差方差已经可以解释大部分预测不确定性，并且我们的方法可以使用重采样方法进一步增强对不确定性的理解。这个结果是合理的，因为 KITTI [30] 和 NYUv2 [77] 训练集的样本量较大导致较低。

ConOR > OR基于有效条件分布进行统计推断的显着性

Ablation

Dataset	Method	AUSE(ξ) ↓			AURG(ξ) ↑		
		rmse	rel	$1-\delta_1$	rmse	rel	$1-\delta_1$
KITTI	ConOR	0.097	0.184	0.109	0.751	0.496	0.846
	ConOR+MCD	0.104	0.185	0.128	0.740	0.499	0.814
	ConOR+DE	0.096	0.181	0.112	0.749	0.500	0.848
	ConOR+WBS	0.095	0.181	0.107	**0.754**	0.500	0.849
	ConOR+MBS	**0.094**	**0.180**	**0.106**	**0.754**	**0.501**	**0.851**
NYUv2	ConOR	0.305	0.350	0.349	0.333	0.235	0.544
	ConOR+MCD	0.305	0.351	0.350	0.331	0.233	0.542
	ConOR+DE	0.303	0.351	0.343	0.327	0.229	0.557
	ConOR+WBS	**0.297**	**0.340**	**0.333**	**0.343**	**0.245**	**0.559**
	ConOR+MBS	**0.297**	0.343	0.336	0.340	0.243	0.557



(a) rmse (b) AUSE(rmse) (c) AURG(rmse)

这表明ConOR（任意不确定性）估计的误差方差已经可以解释大部分预测不确定性，并且我们的方法可以使用重采样方法进一步增强对不确定性的理解。这个结果是合理的，因为 KITTI [30] 和 NYUv2 [77] 训练集的样本量较大导致estimation variance较低。

DIG

Thanks