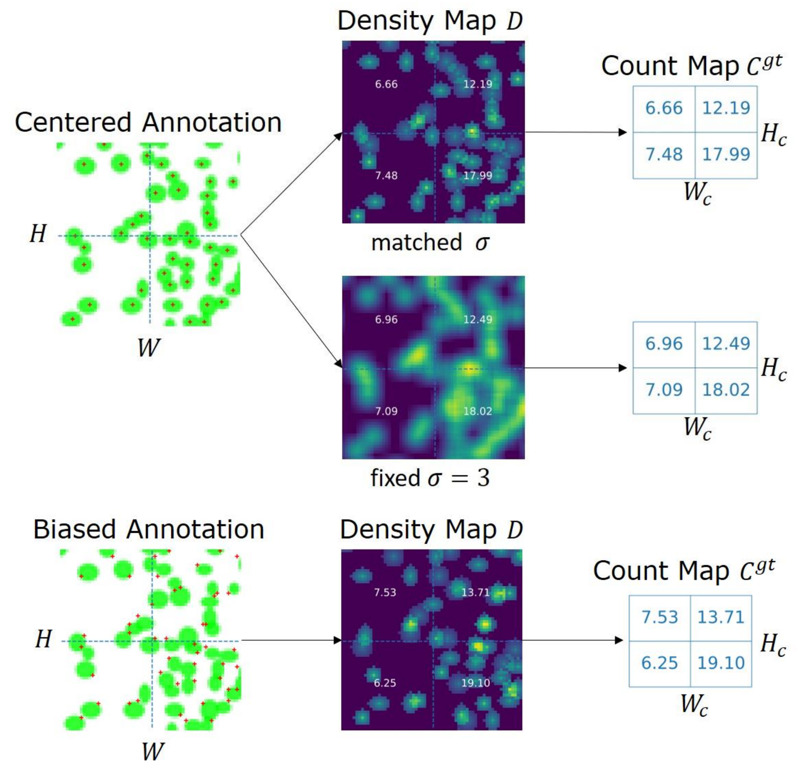
Discrete-Constrained Regression for Local Counting Models

Haipeng Xiong and Angela Yao
National University of Singapore

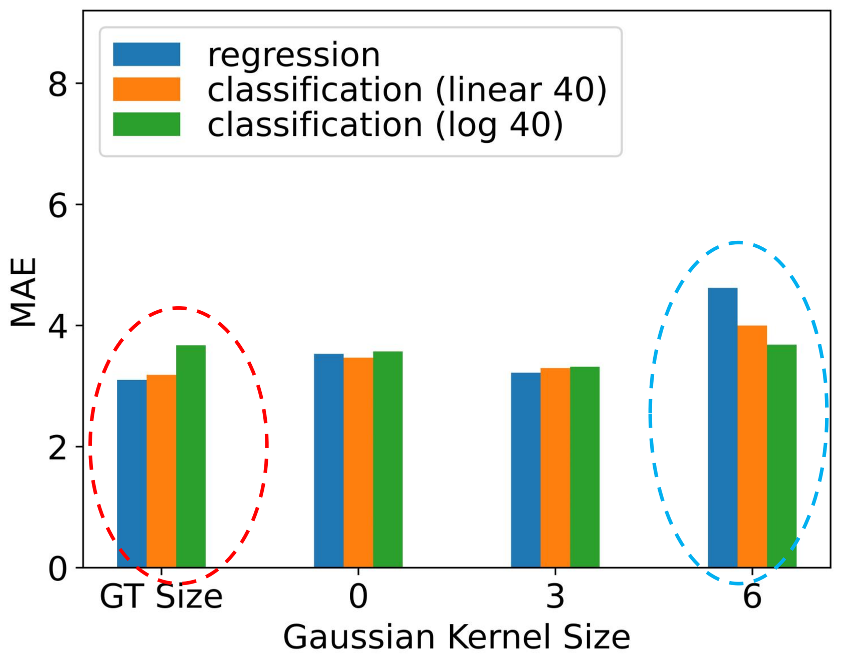
When will classification outperform regression in local counting models?

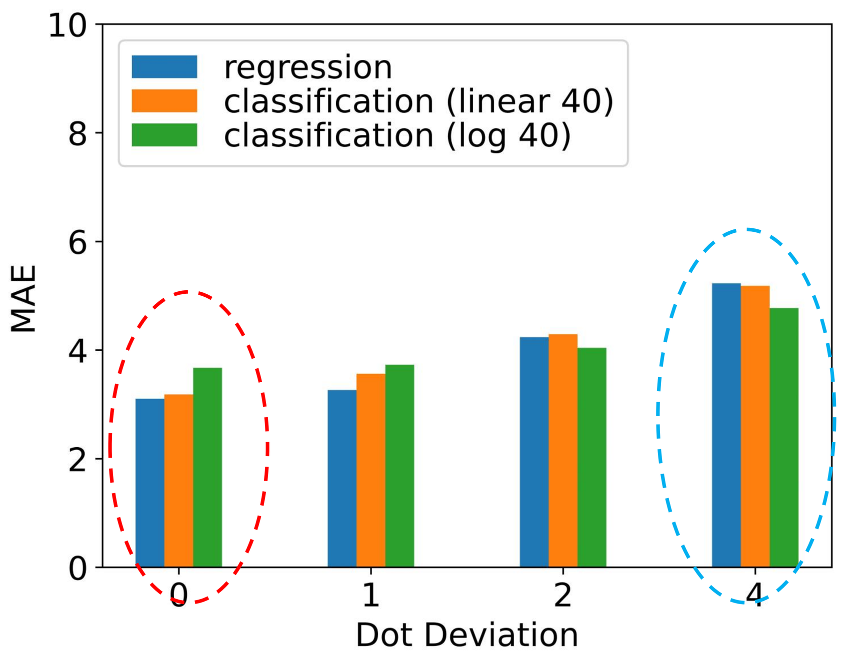
How to benefit regression from classification?

Observation



Ground truth local counts are not accurate, due to mismatched Gaussian kernel sizes and biased dot annotations.





Classification is not better than regression when GT is precise, but outperforms regression when GT is imprecise.

Discrete-constrained loss L_{dc}

Ground truth local counts are not accurate.

$$E(j,k) = C^{gt}(j,k) - C^{pre}(j,k)$$

$$E(j,k) = \overbrace{E^{true}(j,k)}^{\text{True error}} + \overbrace{\varepsilon(j,k)}^{\text{GT error}}$$

Gradient is correct when outside the interval $[C^{gt}(j,k) - |\varepsilon(j,k)|, C^{gt}(j,k) + |\varepsilon(j,k)|]$

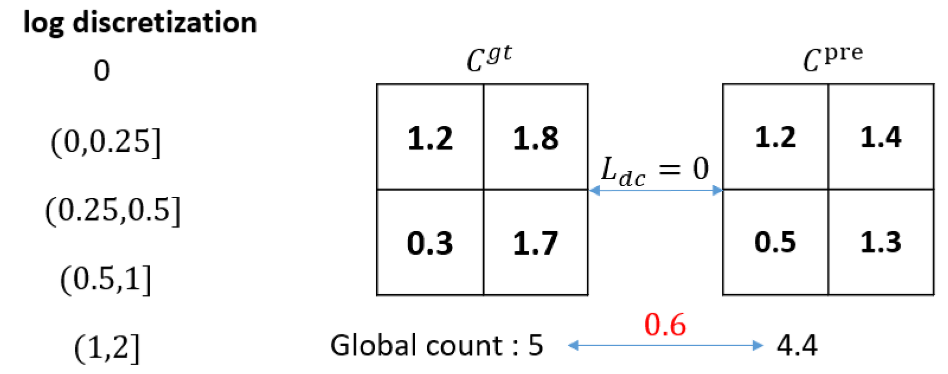
Such interval setting could be modelled via discretizing output range in classification. We adopt discretization like classification to tolerate the GT error.

Suppose $V_j < C^{gt} \leq V_{j+1}$,

$$L_{dc} = \begin{cases} 0, & \text{if } V_j < C^{pre} \leq V_{j+1} \\ |C^{pre} - C^{gt}|, & \text{otherwise} \end{cases}$$

Global Count loss L_{gc}

- simply using discrete regression loss L_{dc} would encounter discretization errors and make global counts imprecise



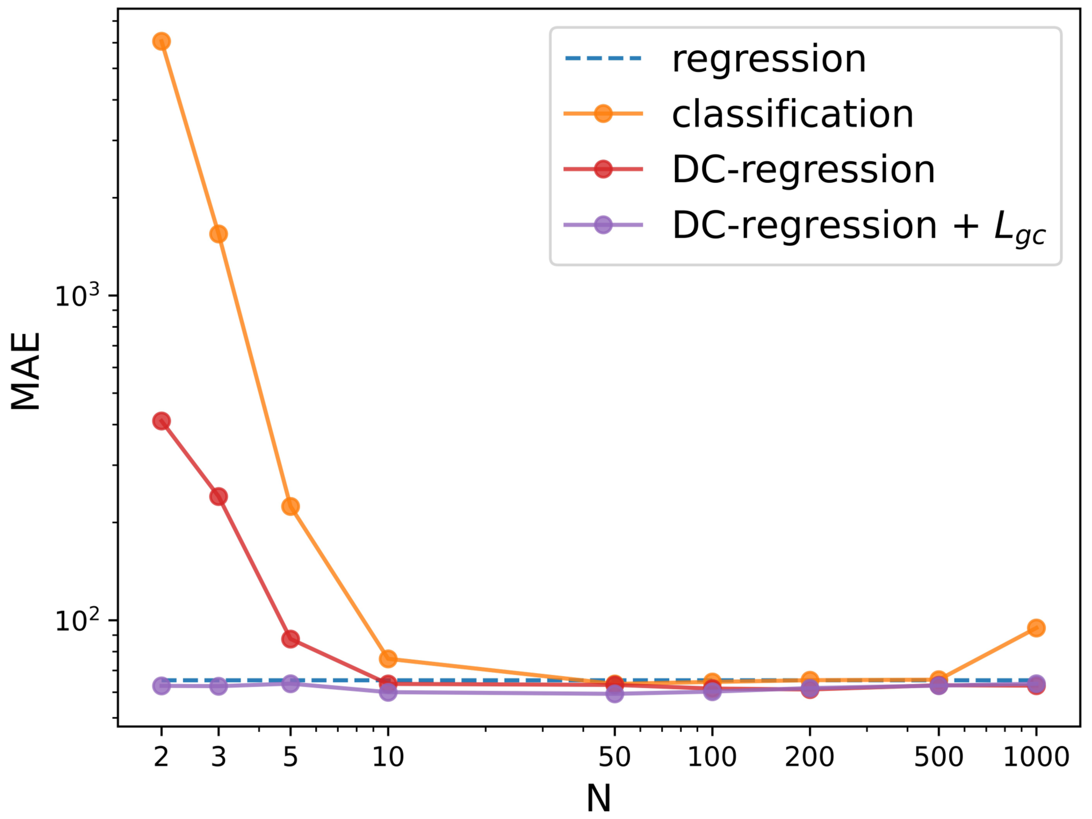
- we can add global count loss L_{gc} to eliminate discretization errors

Overall loss $L = L_{dc} + L_{gc}$

Conclusion

- In counting, classification outperforms regression when the learning target is imprecise. Regression is vulnerable to GT error, while classification can tolerate it via discretization.
- Discrete constraints could be adopted to prevent regression models overfit the label noise.
- A novel discrete-constrained (DC) regression model is proposed in this paper. DC-regression is applicable to both counting and age estimation tasks.

Results



Ablation Study on Interval Partitions

Method	Interval Partition	SHA MAE	SHA MSE	SHB MAE	SHB MSE
classification	linear	65.6	115.4	8.6	14.6
	log	64.3	112.3	8.7	17.2
	uep [23]	63.9	112.8	7.9	15.1
DC-regression	linear	62.5	106.0	7.5	12.2
	log	61.6	96.7	7.1	11.1
	uep [23]	61.9	104.2	7.4	12.2
DC-regression+L_{gc}	linear	**60.3**	**95.5**	**6.6**	11.0
	log	**60.3**	103.7	6.7	10.6
	uep [23]	61.3	97.5	6.7	**10.2**

Global count loss L_{gc} could eliminate the discretization error and make DC-regression robust to the number of intervals.

Global count loss L_{gc} makes DC-regression robust to interval settings.

Comparison with SOTA on crowd counting datasets

	Backbone	SHA MAE	SHA MSE	SHB MAE	SHB MSE	JHU MAE	JHU MSE	QNRF MAE	QNRF MSE
CSRNet [5]	VGG16	68.2	115.0	10.6	16.0	85.9	309.2	108.2	181.3
DRCN [17]	VGG16	64.0	98.4	8.5	14.4	82.3	328.0	112.2	176.3
BL [8]	VGG19	62.8	101.8	7.7	12.7	75.0	299.9	88.7	154.8
PaDNet [18]	VGG16	59.2	98.1	8.1	12.2	—	—	96.5	170.2
MNA [20]	VGG19	61.9	99.6	7.4	11.3	67.7	258.5	85.8	150.6
OT [22]	VGG19	59.7	95.7	7.4	11.8	68.4	283.3	85.6	148.3
UOT [9]	VGG19	**58.1**	95.9	**6.5**	**10.2**	60.5	**252.7**	**83.3**	**142.3**
Generalized Loss [21]	VGG19	61.3	95.4	7.3	11.7	**59.9**	259.5	84.3	147.5
regression (L_{reg})	VGG16	65.4	103.3	10.7	19.5	71.2	296.0	98.6	166.6
L_{BL}	VGG16	62.2	103.4	7.4	10.7	64.2	275.7	90.1	162.5
$L_{\text{bias}}^{\lambda}$	VGG16	62.9	108.5	7.8	12.0	68.6	289.4	93.3	160.8
classification	VGG16	64.6	106.7	8.7	17.2	67.8	261.6	97.6	163.2
S-DCNet (cls) [27]	VGG16	58.3	95.0	6.7	10.7	65.2	272.8	104.4	176.1
DC-regression	VGG16	61.6	96.7	7.1	11.1	67.2	288.2	91.4	157.5
DC-regression+$L_{\text{bias}}^{\lambda}$	VGG16	60.3	103.7	6.7	10.6	64.8	282.6	86.0	148.2
DC-regression+L_{BL}	VGG16	60.7	101.0	7.1	11.0	61.6	263.2	87.1	152.1
S-DCNet (dcreg)*	VGG16	59.7	**91.4**	7.0	11.6	60.0	269.9	86.9	159.3
S-DCNet (dcreg)†	VGG16	59.8	100.0	6.8	11.5	62.1	268.9	84.8	**142.3**

Comparison with SOTA on age estimation datasets

Method	Mega MAE	Mega RMSE	Mega CA3	Mega CA5	Mega CA7	MegaA MAE	MegaA RMSE	MegaA CA3	MegaA CA5	MegaA CA7
Posterior [30]	—	—	38.69	57.90	73.15	—	—	62.08	80.43	90.42
Xia *et al.* [24]	—	—	—	—	—	2.80	—	62.50	82.37	—
Yu *et al.* [19]	—	—	42.19	60.0	72.70	—	—	64.80	83.20	91.40
classification	5.57	7.15	39.72	57.10	71.45	2.91	4.14	68.19	84.82	93.03
regression	5.26	6.72	41.89	59.84	74.73	2.87	4.00	68.57	85.10	93.51
DC-regression	**5.15**	**6.58**	**42.36**	**61.31**	**75.26**	**2.80**	**3.97**	**69.40**	**85.98**	**93.79**