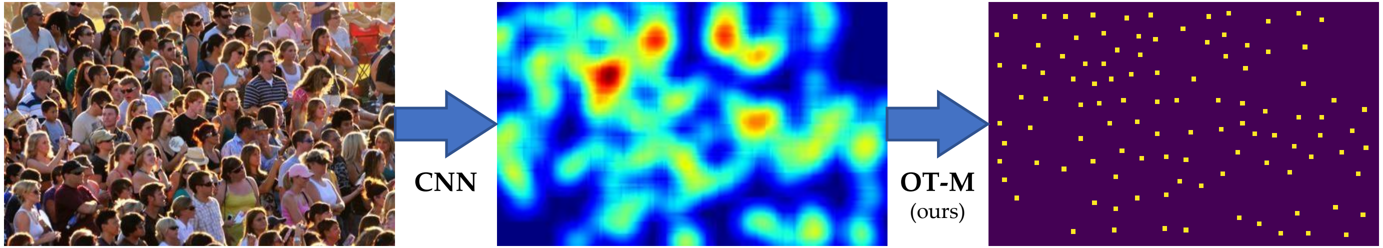
Optimal Transport Minimization: Crowd Localization on Density Maps for Semi-Supervised Counting

Wei Lin, and Antoni B. Chan

Department of Computer Science, City University of Hong Kong
elonlin24@gmail.com, abchan@cityu.edu.hk

Saturday, May 27, 2023

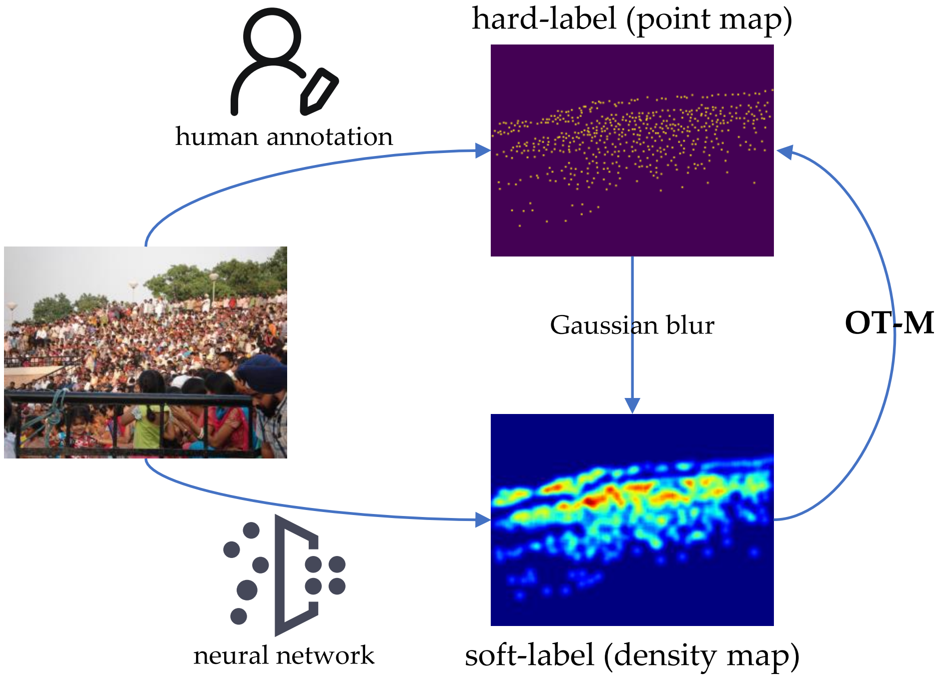
Motivation



Most crowd counting methods pay attention on density map prediction, few consider how to perform localization on it.

- Optimal Transport Minimization (OT-M) algorithm is proposed to estimate the locations of objects from density maps;
- OT-M is applied to produce *hard pseudo-labels* for semi-supervised counting, which conforms with schemes in other semi-supervised tasks.
- A Confidence-weighted Generalized Loss (C-GL) is proposed to reduce the influence of inaccurate pseudo-labels.

Optimal Transport Minimization



- **Objective:** estimate a hard label from a soft density map by minimizing the *entropic optimal transport cost* (Sinkhorn distance) between them.

$$\hat{\mathcal{B}} = \underset{\mathcal{B}=\{\boldsymbol{y}_j\}_{j=1}^m}{\arg\min} \ \mathcal{L}^{\varepsilon}(\mathcal{A}, \mathcal{B})$$

$$\mathcal{L}^{\varepsilon}(\mathcal{A}, \mathcal{B}) = \min_{\mathbf{P}\in\mathbf{U}(\boldsymbol{a},\boldsymbol{b})} \langle \mathbf{C}, \mathbf{P} \rangle - \varepsilon \mathcal{H}(\mathbf{P}),$$

$$= \sum_{i,j} C_{ij} P_{ij} + \varepsilon \sum_{i,j} P_{ij} \log(P_{ij})$$

$$C(\boldsymbol{x}_i, \boldsymbol{y}_j) = \|\boldsymbol{x}_i - \boldsymbol{y}_j\|^2$$

Optimal Transport Minimization

OT-M algorithm follows an *alternating scheme* that estimates the optimal transport plan from the current point map (the OT-step), and updates the point map by minimizing their transport cost (the M-step).

- Optimal Transport Step (OT-Step): the optimal transport plan $\mathbf{P}^{(k)}$ is computed while holding the cost matrix fixed:

$$\mathbf{P}^{(k)} = \operatorname*{argmin}_{\mathbf{P} \in \mathbf{U}(\boldsymbol{a},\boldsymbol{b})} \langle \mathbf{C}(\mathcal{B}^{(k-1)}), \mathbf{P} \rangle - \varepsilon \mathcal{H}(\mathbf{P})$$

- Minimization-Step (M-Step): the optimal cost matrix, parametrized by the points $\mathcal{B} = \{\boldsymbol{y}_j\}_{j=1}^{m}$ is computed while holding the transport plan fixed:

$$\mathcal{B}^{(k)} = \operatorname*{argmin}_{\mathcal{B} = \{\boldsymbol{y}_j\}_{j=1}^{m}} \langle \mathbf{C}(\mathcal{B}), \mathbf{P}^{(k)} \rangle - \varepsilon \mathcal{H}(\mathbf{P}^{(\mathbf{k})})$$

Optimal Transport Minimization

OT-Step

$$\mathbf{P}^{(k)} = \underset{\mathbf{P}\in\mathbf{U}(\boldsymbol{a},\boldsymbol{b})}{\text{argmin}} \langle \mathbf{C}(\mathcal{B}^{(k-1)}), \mathbf{P} \rangle - \varepsilon\mathcal{H}(\mathbf{P})$$

- The solution of optimal transport can be formulated as:

$$\mathbf{P} = \text{diag}(\mathbf{u})\mathbf{K}\text{diag}(\mathbf{v}), \quad \mathbf{K} = \exp(-\mathbf{C}/\varepsilon)$$

- Sinkhorn algorithm[1] repeats the following iterations to find **u** and **v** until convergence:

$$\mathbf{u}^{(l+1)} = \frac{\boldsymbol{a}}{\mathbf{K}\mathbf{v}^{(l)}}, \quad \mathbf{v}^{(l+1)} = \frac{\boldsymbol{b}}{\mathbf{K}^\top \mathbf{u}^{(l+1)}}$$

M-Step

$$\mathcal{B}^{(k)} = \underset{\mathcal{B}=\{\boldsymbol{y}_j\}_{j=1}^m}{\text{argmin}} \langle \mathbf{C}(\mathcal{B}), \mathbf{P}^{(k)} \rangle - \varepsilon\mathcal{H}(\mathbf{P}^{(\mathbf{k})})$$

- Plugging in the cost function ($\|\cdot\|^2$), each $\boldsymbol{y}_i$ can be optimized independently:

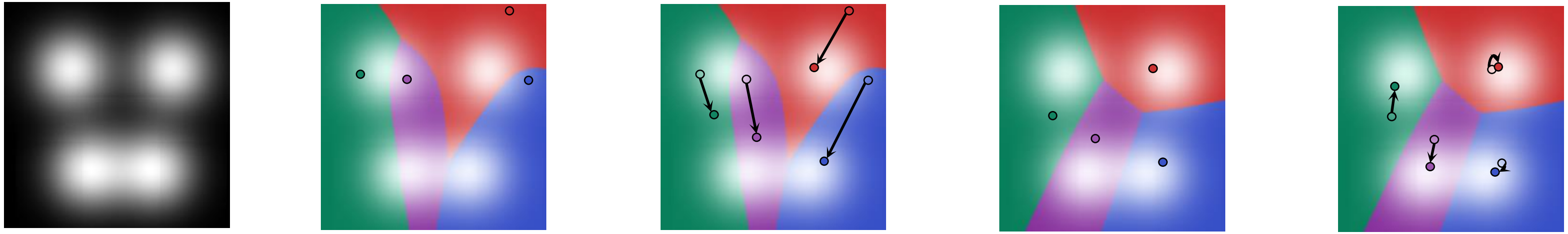
$$\boldsymbol{y}_j^{(k)} = \underset{\boldsymbol{y}_j}{\arg\min} \sum_{i=1}^{n} P_{ij}^{(k)} \|\boldsymbol{x}_i - \boldsymbol{y}_j\|^2$$

- Letting its derivative equal to zero:

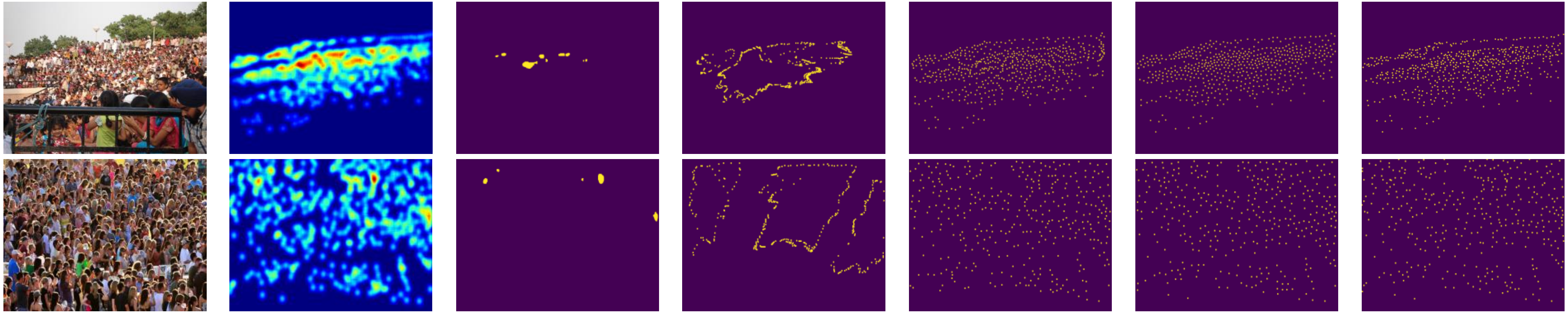
$$\frac{\partial}{\partial \boldsymbol{y}_j} \sum_{i=1}^{n} P_{ij}^{(k)} \|\boldsymbol{x}_i - \boldsymbol{y}_j\|^2 = 0 \Rightarrow \boldsymbol{y}_j^{(k)} = \frac{\sum_{i=1}^n P_{ij}^{(k)} \boldsymbol{x}_i}{\sum_{i=1}^n P_{ij}^{(k)}}$$

[1] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.

Optimal Transport Minimization

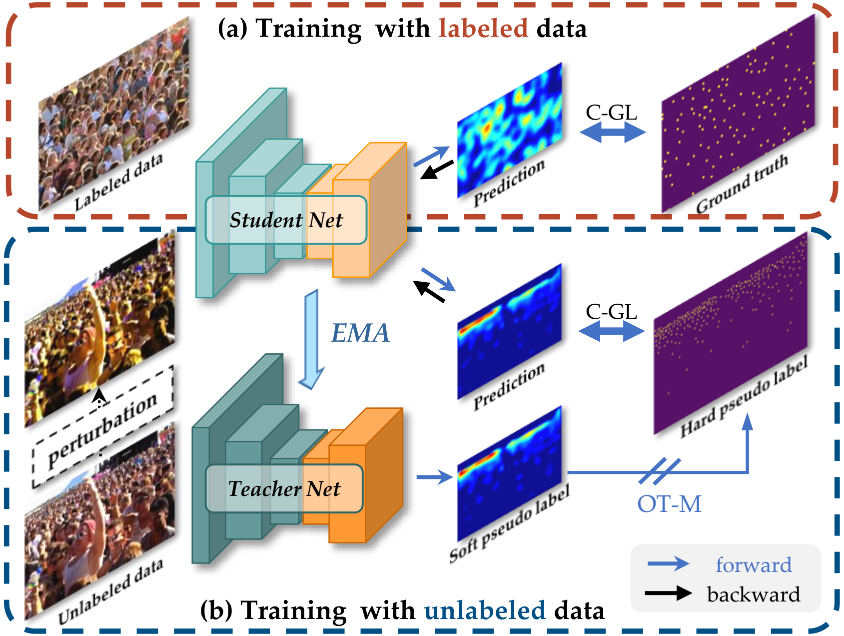


OT-step (1) M-step (1) OT-step (2) M-step (2)



input image density map iteration = 1 iteration = 2 iteration = 4 iteration = 8 Ground Truth

OT-M for Semi-Supervised Counting



- Labeled images: A student net is trained with fully-supervised learning on the GT point maps.
- Unlabeled images: A teacher net is used to generate a *soft pseudo-label (density map)* for perturbed input, and OT-M is applied to produce a *hard pseudo-label (point map)*.
- Mean-teacher: An exponential moving average (EMA) is used to update the parameters in the teacher net.
- C-GL: Confidence-weighted generalized loss is used to reduce the effect of inconsistent (noisy) pseudo-labels.

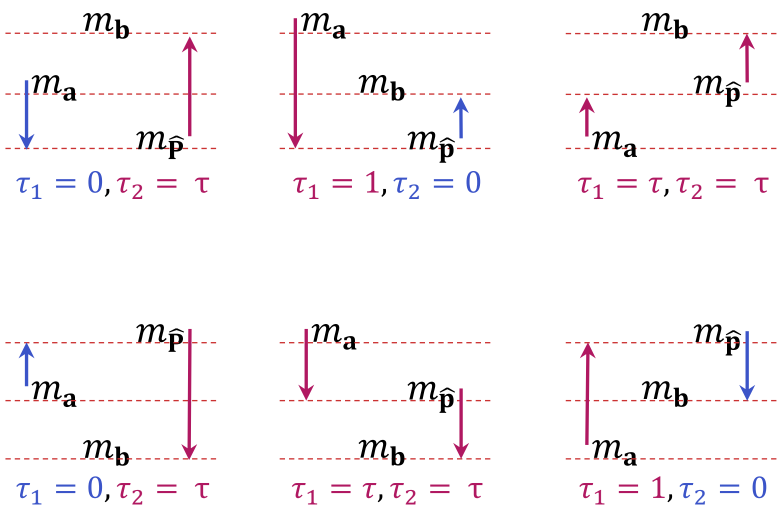
OT-M for Semi-Supervised Counting

Generalized Loss w/ Gating

$$L_{gl}^{\varepsilon,\tau} = \mathbf{a}^\top \mathbf{f}^* + \mathbf{b}^\top \mathbf{g}^* - \varepsilon \mathcal{H}(\widehat{\mathbf{P}}) + \tau_2 \|\widehat{\mathbf{P}} \mathbf{1}_m - \mathbf{a}\|_2^2 + \tau_1 \|\widehat{\mathbf{P}}^\top \mathbf{1}_n - \mathbf{b}\|_1$$

$$\tau_1 = \begin{cases} 0, & m_{\mathbf{a}} < m_{\mathbf{b}} < m_{\widehat{\mathbf{P}}}, \\ 0, & m_{\widehat{\mathbf{P}}} < m_{\mathbf{b}} < m_{\mathbf{a}}, \\ \tau, & \text{otherwise.} \end{cases} \quad \tau_2 = \begin{cases} 0, & m_{\mathbf{b}} < m_{\mathbf{a}} < m_{\widehat{\mathbf{P}}}, \\ 0, & m_{\widehat{\mathbf{P}}} < m_{\mathbf{a}} < m_{\mathbf{b}}, \\ \tau, & \text{otherwise.} \end{cases}$$

- $(\mathbf{f}^*, \mathbf{g}^*)$ and $\widehat{\mathbf{P}}$ are the gradients of $(\mathbf{a}, \mathbf{b})$ and the transport plan while applying Sinkhorn algorithm to KL-UOT.
- (τ_1, τ_2) is used to mask harmful case caused by the bias of Sinkhorn algorithm.



OT-M for Semi-Supervised Counting

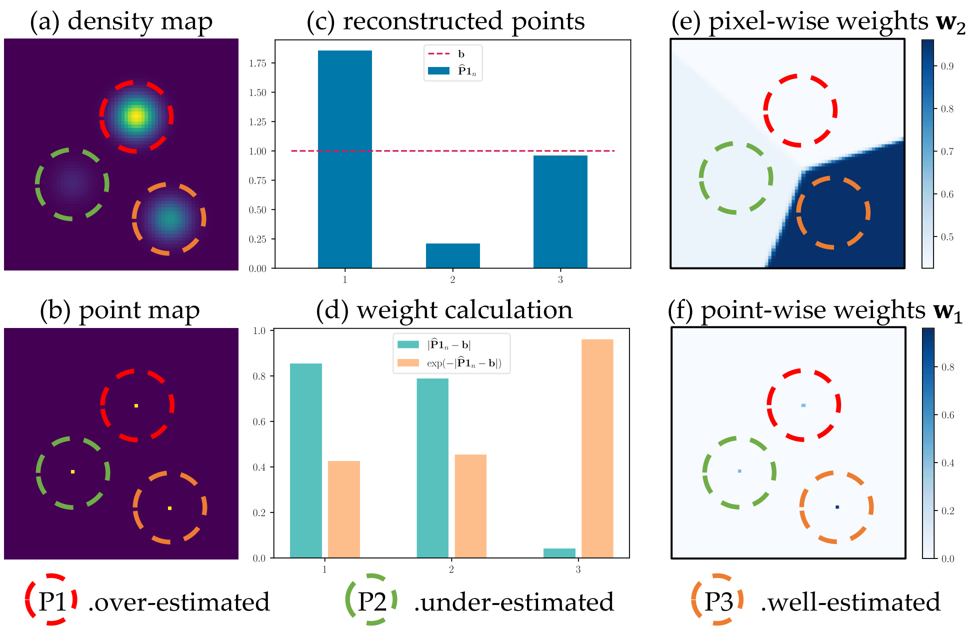
Confidence-weighted Generalized Loss

Confidence is computed by measuring the consistency of $\widehat{\mathbf{P}}$ and $\mathbf{b}$.

$$\mathbf{w}_1 = \exp\left[-\gamma\big(\mathrm{diag}(\mathbf{b})^{-1} |\widehat{\mathbf{P}}^\top \mathbf{1}_n - \mathbf{b}|\big)\right]$$

$$\mathbf{w}_2 = \mathrm{diag}(\widehat{\mathbf{P}}\mathbf{1}_m)^{-1}\widehat{\mathbf{P}}\mathbf{w}_1$$

$$L_{c\text{-}gl}^{\varepsilon,\tau,\gamma} = \mathbf{a}^\top \mathbf{W}_2 \mathbf{f}^* + \mathbf{b}^\top \mathbf{W}_1 \mathbf{g}^* - \varepsilon \mathcal{H}(\widehat{\mathbf{P}}) + \tau_2 \|\mathbf{W}_2(\widehat{\mathbf{P}}\mathbf{1}_m - \mathbf{a})\|_2^2 + \tau_1 \|\mathbf{W}_1(\widehat{\mathbf{P}}^\top \mathbf{1}_n - \mathbf{b})\|_1$$

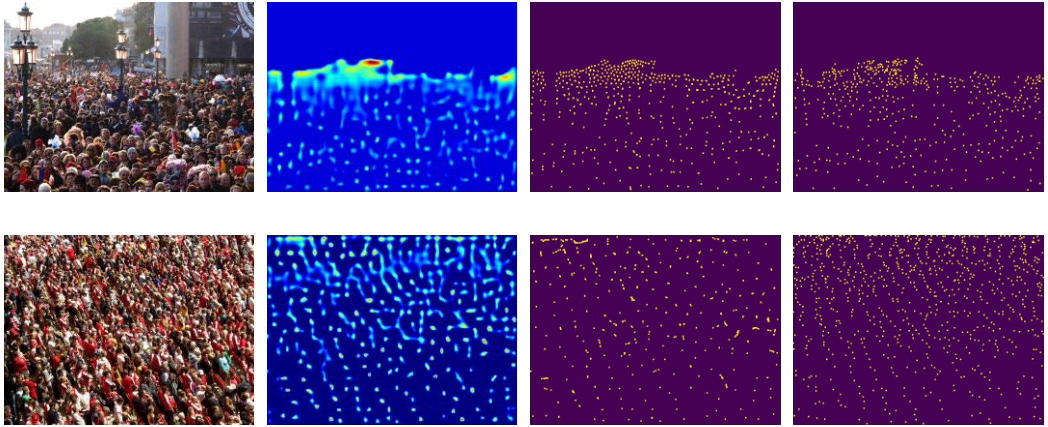


Experiments on Localization

Density Map	Localization	Precision	Recall	F-measure
ground-truth density map	LM [58]	0.892	0.736	0.807
	GMM [14]	0.842	0.838	0.840
	OT-M (ours)	**0.914**	**0.910**	**0.912**
GL [58] cvpr'21	LM [58]	0.782	0.748	0.765
	GMM [14]	0.750	0.728	0.739
	OT-M (ours)	**0.804**	**0.783**	**0.793**
MAN [27] cvpr'22	LM [58]	0.624	0.483	0.544
	GMM [14]	0.749	0.732	0.736
	OT-M (ours)	**0.772**	**0.755**	**0.760**
ChfL [47] cvpr'22	LM [58]	0.812	0.571	0.671
	GMM [14]	0.755	0.740	0.747
	OT-M (ours)	**0.780**	**0.765**	**0.772**

LM: Local Maximum
GMM: Gaussian Mixture Model

Method			Prec.	Rec.	F-meas.
box	Faster RCNN [42]	cvpr'15	0.958	0.035	0.068
density map	RAZNet [28]	cvpr'19	0.666	0.543	0.599
	GL+LM [58]	cvpr'21	**0.800**	0.562	0.660
	GL+OT-M(ours)		0.710	**0.658**	**0.683**
point	P2PNet [54]	iccv'21	0.729	0.695	0.712
	CLTR [23]	eccv'22	0.694	0.676	0.685



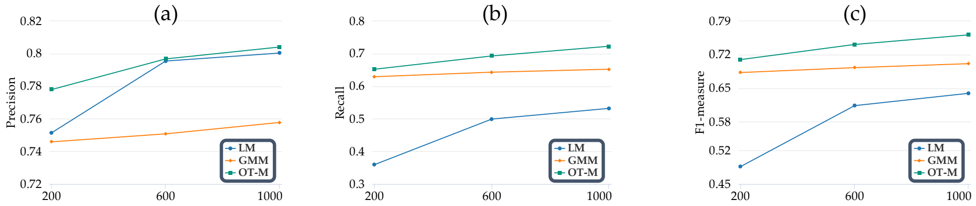
input image density map OT-M results ground truth

Experiments on Semi-Supervised Counting

Label Percentage	Methods	ST-A MAE	ST-A MSE	ST-B MAE	ST-B MSE	UCF-QNRF MAE	UCF-QNRF MSE	JHU++ MAE	JHU++ MSE
5%	DAC [26]	92.9$\pm$3.4	148.6$\pm$10.3	13.4$\pm$2.2	24.6$\pm$6.7	122.7$\pm$7.8	218.9$\pm$14.0	81.2$\pm$2.4	313.7$\pm$12.2
	OT-M (ours)	**86.0$\pm$2.2**	**132.7$\pm$3.3**	**12.8$\pm$1.4**	**22.0$\pm$4.5**	**120.1$\pm$7.3**	**208.9$\pm$11.7**	**80.9$\pm$3.1**	**303.1$\pm$9.5**
10%	DAC [26]	84.8$\pm$4.5	140.9$\pm$11.3	11.1$\pm$0.5	18.9$\pm$1.9	110.5$\pm$5.9	196.0$\pm$16.3	76.0$\pm$2.0	293.8$\pm$10.4
	OT-M (ours)	**81.6$\pm$2.6**	**127.1$\pm$3.8**	**10.9$\pm$0.5**	**18.1$\pm$1.4**	**107.9$\pm$4.1**	**180.6$\pm$7.8**	**75.5$\pm$1.6**	**287.9$\pm$11.1**
40%	DAC [26]	71.6$\pm$2.0	120.8$\pm$5.6	**9.0$\pm$0.3**	14.6$\pm$0.5	**91.8$\pm$4.7**	161.4$\pm$12.4	**64.1$\pm$3.0**	270.6$\pm$9.3
	OT-M (ours)	**70.0$\pm$2.2**	**113.0$\pm$6.9**	9.0$\pm$0.4	**14.2$\pm$0.7**	93.4$\pm$5.4	**157.5$\pm$7.8**	66.5$\pm$3.1	**268.2$\pm$9.5**

Label Pct.	Methods	ST-A MAE	ST-A MSE	ST-B MAE	ST-B MSE	UCF-QNRF MAE	UCF-QNRF MSE	JHU++ MAE	JHU++ MSE
5%	MT [55]	104.7	156.9	19.3	33.2	172.4	284.9	101.5	363.5
	L2R [29]	103.0	155.4	20.3	27.6	160.1	272.3	101.4	338.8
	GP [49]	102.0	172.0	15.7	27.9	160.0	275.0	98.9	355.7
	DAC [26]	85.2	135.0	**12.5**	22.1	123.5	207.3	83.9	308.8
	OT-M (ours)	**83.7**	**133.3**	12.6	**21.5**	**118.4**	**195.4**	**82.7**	**304.5**
10%	MT [55]	94.5	115.5	15.6	24.5	145.5	250.3	90.2	319.3
	L2R [29]	90.3	115.5	15.6	24.4	148.9	249.8	87.5	315.3
	IRAST [31]	86.9	148.9	14.7	22.9	135.6	233.4	86.7	303.4
	DAC [26]	82.5	123.2	10.9	19.1	115.1	193.5	74.0	297.1
	OT-M (ours)	**80.1**	**118.5**	**10.8**	**18.2**	**113.1**	**186.7**	**73.0**	**280.6**
40%	MT [55]	88.2	151.1	15.9	25.7	147.2	249.6	121.5	388.9
	L2R [29]	86.5	148.2	16.8	25.1	145.1	256.1	123.6	376.1
	SUA [38]	**68.5**	121.9	14.1	20.6	130.3	226.3	80.7	290.8
	DAC [26]	71.1	119.7	8.1	13.6	**96.8**	168.2	**66.3**	276.6
	OT-M (ours)	70.7	**114.5**	**8.1**	**13.1**	100.6	**167.6**	72.1	**272.0**

Method	MAE	MSE
Label only	138.52$\pm$10.65	242.26$\pm$16.62
LM [58]	148.53$\pm$9.53	270.25$\pm$23.67
GMM [14]	126.67$\pm$7.41	217.00$\pm$16.17
OT-M (ours)	**120.13$\pm$7.34**	**208.87$\pm$11.65**

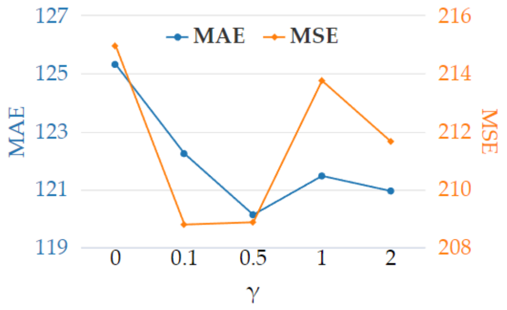


Ablation Study & Limitation

Data	gate	confidence	MAE	MSE
label only			145.59	257.31
	✓		144.48	255.33
	✓	✓	**138.52**	**242.26**

Data	loss for unlabeled data	MAE	MSE
label+unlabel	L2 loss	137.17	239.52
	L2 w/ confidence	135.88	233.19
	GL	125.32	214.96
	GL w/ confidence (C-GL)	**120.13**	**208.87**

gate	confidence	MAE	MSE
		127.69 ± 4.52	216.50 ± 11.45
	✓	123.85 ± 5.92	212.23 ± 12.02
✓		125.32 ± 7.62	214.96 ± 12.57
✓	✓	$\mathbf{120.13 \pm 7.34}$	$\mathbf{208.87 \pm 11.65}$



OT-M algorithm is limited by its efficiency.

- the total runtime for an image of 384×576 is 0.080s:
  - density map estimation :0.013s
  - OT-M :0.067s
- input images are cropped into 512×512 in semi-supervised counting. Average training time is 0.34s per sample.

Conclusion

- Optimal Transport Minimization algorithm, a parameter-free method for crowd localization on density map. OT-M alternates between two steps:

  - OT-step: the transport plan between the current point map and the input density map is estimated;
  - M-step: the point map is updated using the transport plan computed in the OT-step.

- OT-M is applied to semi-supervised counting via a teacher-student framework.

- A confidence-weighted generalized loss (C-GL) is proposed to reduce confirmation bias introduced by noisy predictions for unlabeled data.

電腦科學系
Department of
Computer Science

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Thanks

Wei Lin, and Antoni B. Chan

Department of Computer Science, City University of Hong Kong
elonlin24@gmail.com, abchan@cityu.edu.hk

Saturday, May 27, 2023