RISE RESEARCH INNOVATION SCHOLARSHIP ENTREPRENEURSHIP

2934
Computer and Information Sciences
Graduate

Northeastern University

Laplace Landmark Localization

Joseph P Robinson[1], Sergey Tulyakov[2] and Yun Fu[1]
[1]Northeastern University, USA [2]Snap Inc., USA

Problem Statement

Landmark localization (i.e. key-point detection) & alignment

- Essential for many vision tasks: Face recognition, pose estimation, expression analysis, much more
- Lots attention over years: revamped interest; DNNs push SOTA

Contribution 1

- Current SOTA landmark detectors have low confident mappings

Novel loss with high-order stats for *increase in confidence* **(Fig 1).**

Contribution 2

- Labeling is expensive, prone to human errors, and tedious; while an abundance of faces are available for free online.

Mitigate label costs with a *semi-supervised framework*.

Contribution 3

- Practical aspects: storage costs and speed on mobile device.

Minimize storage costs, while maximizing performance on CPU.

Fig 1 Heatmaps: SAM-based models (right) & our LaplaceKL (middle). Heatmaps are confidence scores that a pixel is a landmark. SAM-based are highly scattered (low in certainty), while our loss is concentrated (i.e. high in certainty). Importance of minimizing scatter shown experimentally (**Table 1**).

1. LaplaceKL Loss

Softargmax [1] (SAM), expected value over 2D normalized heatmap

$$\mathrm{softargmax}(\beta\mathbf{h}) = \sum_x \mathrm{softargmax}(\beta\mathbf{h_x})\cdot x \qquad (1)$$
$$= \sum_x \frac{e^{\beta\mathbf{h_x}}}{\sum_j e^{\beta\mathbf{h_j}}}\cdot x$$
$$= \sum_x p(x)\cdot x = \mathbb{E}_{\mathbf{h}}[x]$$

where K heatmaps (i.e., per landmark, $\mathbf{h}\in\mathbb{R}^{K\times h\times w}$).

Use higher-order statistics to learn heatmaps with greater confidence: Set $\tilde{\mathbf{s}} = \mathbb{E}_{\mathbf{h}}[\mathbf{x}]$, then $\mathrm{Laplace}(\mu, b=1)$ Assume Laplacian (i.e., α=1). Thus, $b = \mathbb{E}_{\mathbf{h}}[||\mathbf{x}-\mathbb{E}_{\mathbf{h}}[\mathbf{x}]||]$ for $\tau(\tilde{\mathbf{h}}) = \sum p(\mathbf{x})||\mathbf{x}-\tilde{\mathbf{s}}||_\alpha^\alpha$

Conveniently, KL has close-form solution for Laplacian [2]:

$$(2) \qquad \mathcal{L}_{\mathrm{KL}} = \mathbb{E}_{(\mathbf{d},\mathbf{s})\sim p(\mathbf{d},\mathbf{s})}\Big[\mathrm{D}_{KL}(q(\mathbf{s}|\mathbf{d})||p(\mathbf{s}|\mathbf{d}))\Big]$$

defies the proposed LaplaceKL loss (Labelled branch in **Fig. 2**).

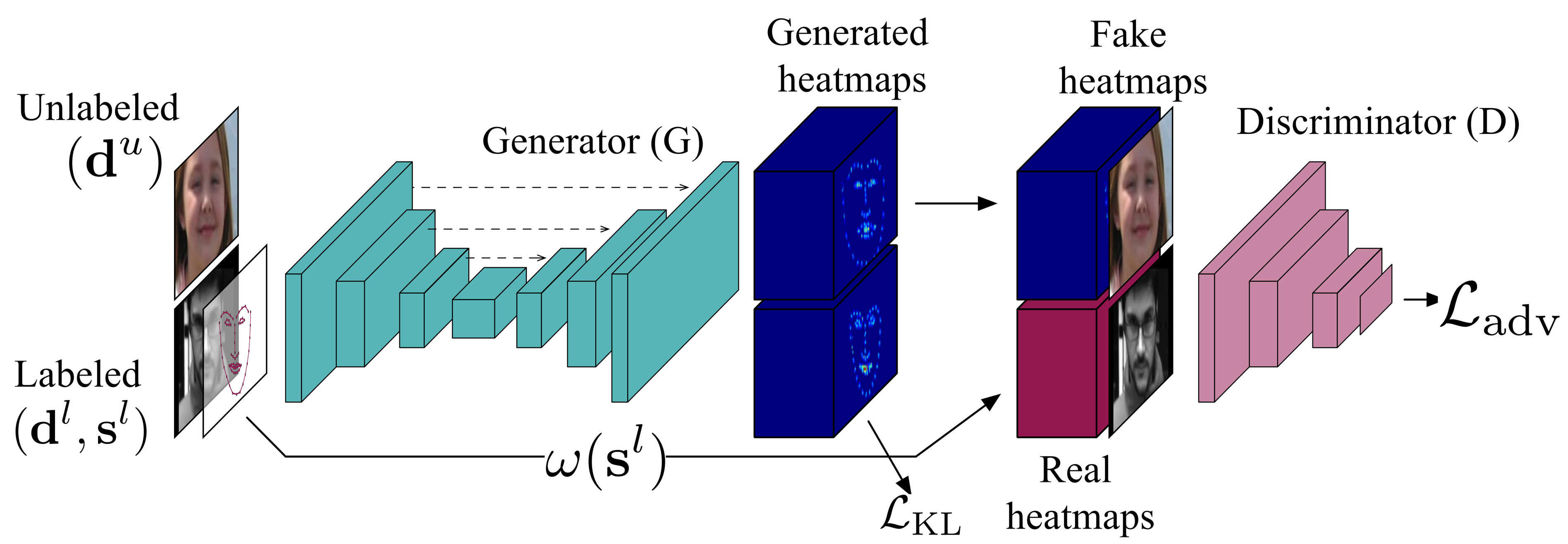

Fig 2 Semi-supervised framework for landmarks localization. Given input image, G makes K heatmaps, 1 per landmark. Labels generate real heatmaps ω(sl). G produces fake samples from unlabeled data. Source images are concatenated on heatmaps and passed to D.

Fig 3 Heatmaps predicted by our LaplaceKL+D(70K) (middle), SAM+D(70K) (right), and faces with ground-truth sketched in green (left). Colors set by value for heatmaps generated. Note our loss predicts with greater confidence, producing separated landmarks as seen in heatmap space–proposed minimizes spread; SAM-based landmarks smudge.

Table 1 NMSE on AFLW & 300W normalized by BB & interocular, respectfully.

	AFLW	300W		
		Common	Challenge	Full
SDM [Xiong et al]	5.43	5.57	15.40	7.52
CFSS [Lv et al]	2.17	4.36	7.56	4.99
RCSR [Wang et al]	-	4.01	8.58	4.90
RCN + (L+ELT) [Honari et al]	**1.59**	4.20	7.78	4.90
CPM+SBR [Dong et al]	2.14	3.28	7.58	4.10
SAM	2.26	3.48	7.39	4.25
SAM+D(10K)	-	3.34	7.90	4.23
SAM+D(30K)	-	3.41	7.99	4.31
SAM+D(50K)	-	3.41	8.06	4.32
SAM+D(70K)	-	3.34	8.17	4.29
LaplaceKL	1.97	3.28	7.01	4.01
LaplaceKL+D(10K)	-	3.26	6.96	3.99
LaplaceKL+D(30K)	-	3.29	6.74	3.96
LaplaceKL+D(50K)	-	3.26	**6.71**	3.94
LaplaceKL+D(70K)	-	**3.19**	6.87	**3.91**

Table 2 NMSE for nets 1/8, 1/4, 1/2, 1 the size (left-to-right). 2.8GHz Intel Core i7 CPU.

	No. of parameters, millions			
	0.0389	0.1281	0.4781	1.8724
SAM	6.86	4.83	4.35	4.25
+D(70K)	6.84	4.85	4.38	4.29
LaplaceKL	5.09	4.39	4.04	4.01
+D(70K)	**4.85**	**4.30**	**3.98**	**3.91**
Size (MB)	0.162	0.507	1.919	7.496
Speed (fps)	21.38	16.77	11.92	4.92

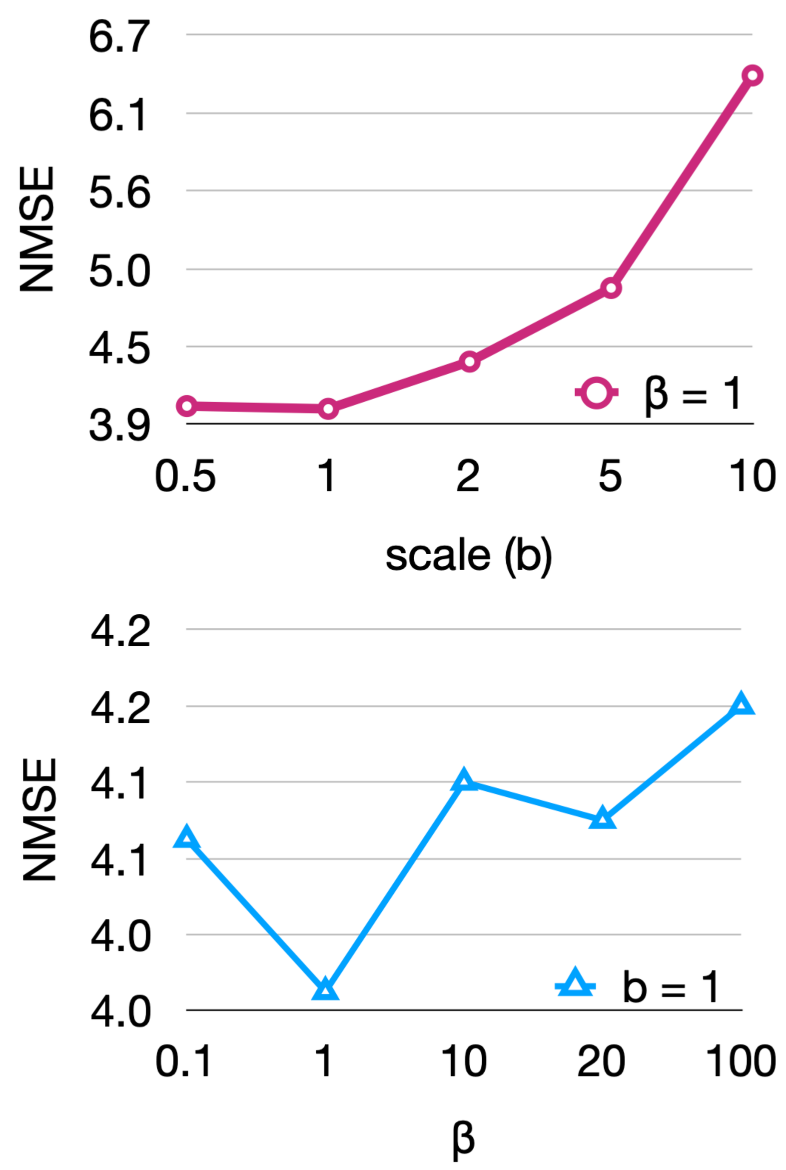

Fig. 4 Ablation study on LaplaceKL.

2. Semi-supervised Framework

- Implemented semi-supervised adversarial framework (**Fig 2**)

$$\mathcal{L}_{\mathrm{adv}} = \log D([\mathbf{D}_t^l, \mathbf{H}_{\mathrm{real}}]) + \log(1 - D([\mathbf{D}_t^u, \mathbf{H}_{\mathrm{fake}}]) \quad (4)$$

- Used unlabeled Megaface (*fake*) to boost performance Eq (2, 4):

$$\min_G \left(\max_D (\lambda\cdot\mathcal{L}_{\mathrm{adv}}(G,D)) + \mathcal{L}_{\mathrm{KL}}(G)\right) \qquad (5)$$

- More unlabeled data the better the performance (**Table 1**)
- More confident heatmaps (**Fig 3**); improved localization (**Fig 5**)

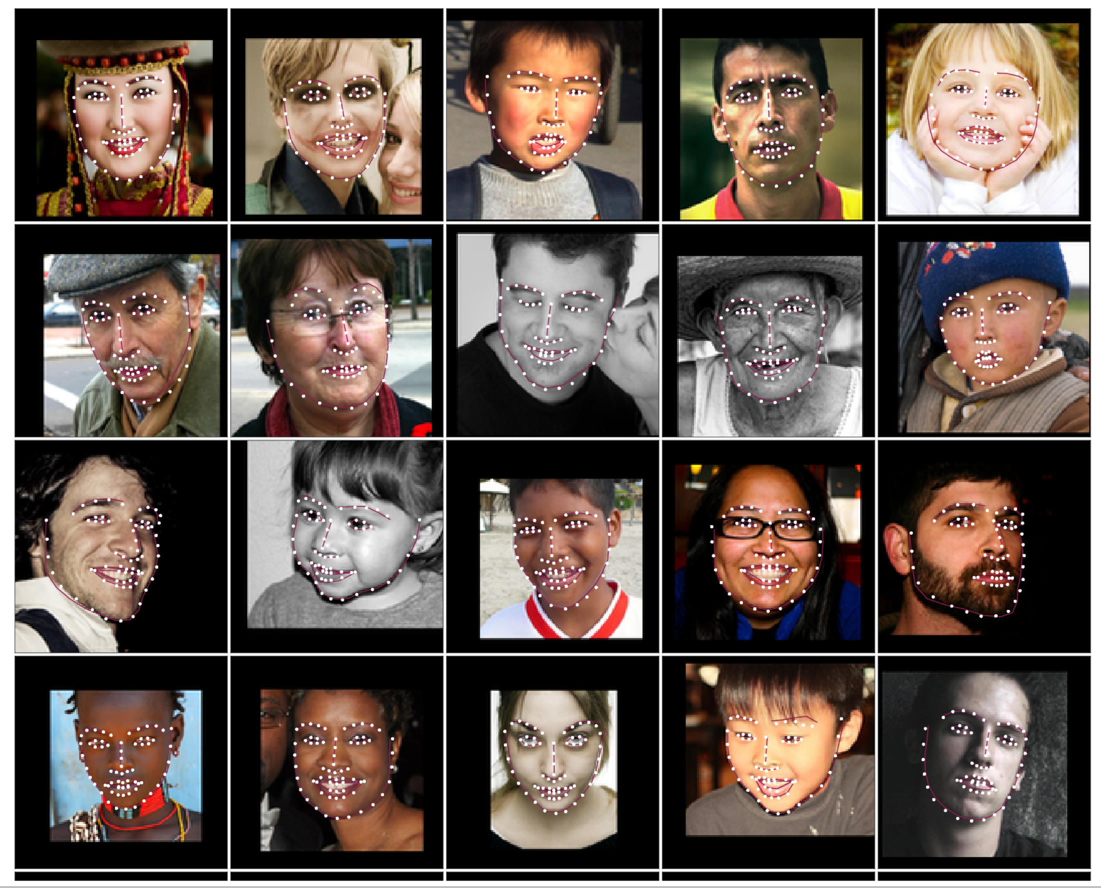

Fig 5 Samples of landmarks predicted with LaplaceKL (white), and ground-truth drawn as line segments (red). Notice the predicted tends to overlap with the ground-truth.

3. Practical Considerations

Conducted ablation studies on proposed loss:

- Reduced size by removing channels by factors of 2 (Table 2).
- Swept values of key parameters (**Fig 4**).

Summary

- Proposed loss function to minimize distribution of landmarks.
  - 1st to consider the “spread” of predicted heatmaps.
- Novel semi-supervised framework to leverage unlabeled data (i.e., face imagery) that is abundantly accessible.
- SOA on renown 300W dataset and 2nd to best on AFLW.
- Comparable performance in real-time with <400KB storage.

References

1. S Honari, et al. *Improving landmark localization with semi-supervised learning*. CVPR 2018.
2. M Hoffman, et al. *Stochastic variational inference*. The Journal of Machine Learning Research 2013.
3. X Xiong, F De la Torre. *Supervised descent and its applications to face alignment*. CVPR 2013.
4. J Lv, et al *Deep regression arch w 2stage for high performance landmark detection* CVPR 2017.
5. W Wang, et al *Recurrent convolutional shape regression*. TPAMI 2018.
6. Dong, et al. *Supervision-by-registration unsupervised approach to improve landmarks* CVPR 2018.
7. C Sagonas, et al. *300 faces in-the-wild : 1st face landmark localization challenge*. ICCVW 2013.
8. M Koestinger, et al. *AFW: large-scale database for face landmark localization*. ICCV 2011
9. A Nech, et al. *Level playing field for million scale face recognition*. CVPR 2017.
10. K Zhang, et al. *Joint face detection & alignment using multitask cascade CNNs*. IEEE SPL 2016.

Contact
jrobins1@ece.neu.edu
jrobsvision.com

Snap Inc.

Smile lab
Synergetic Media Learning Lab