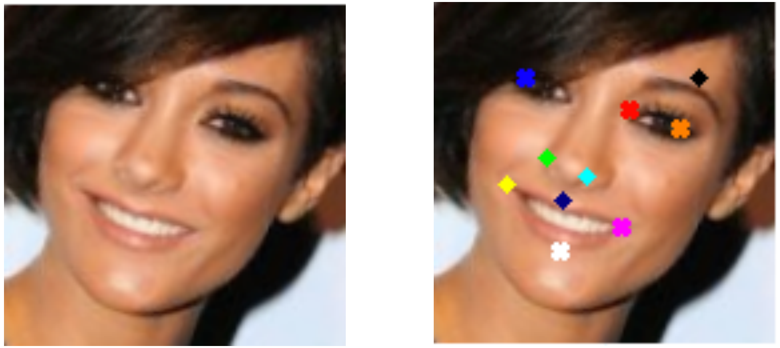
Unsupervised Learning of Landmarks based on Inter-Intra Subject Consistencies

Weijian Li, Haofu Liao, Shun Miao, Le Lu, and Jiebo Luo

Code Available: https://github.com/Weijian-li/unsupervised_inter_intra_landmark

Facial Landmark Localization

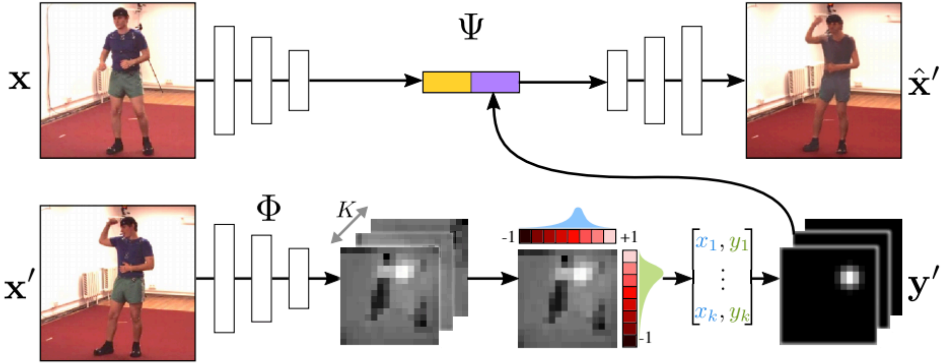
supervised learning

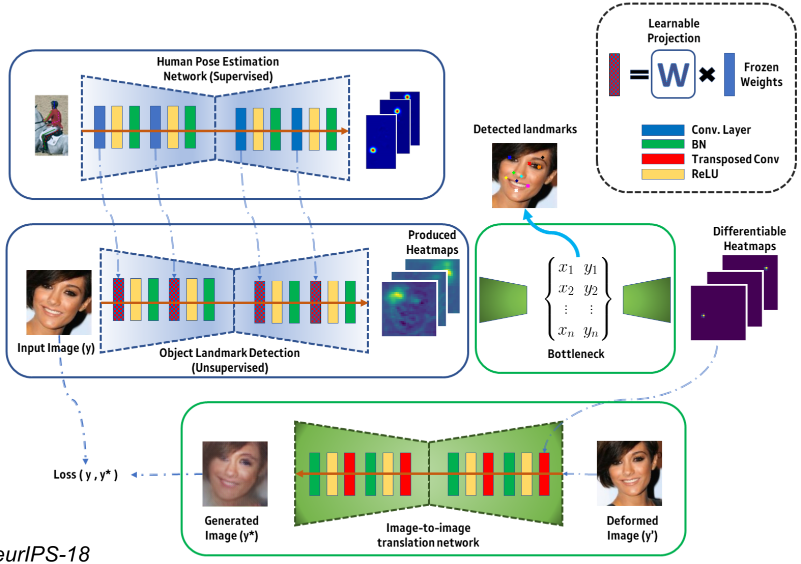


unsupervised learning

Left: Li et al., Structured Landmark Detection via Topology-Adapting Deep Graph Learning, *ECCV-20*
Right: Sanchez et al., Object Landmark Discovery Through Unsupervised Adaptation, *NeurIPS-19*

Unsupervised Landmark Localization



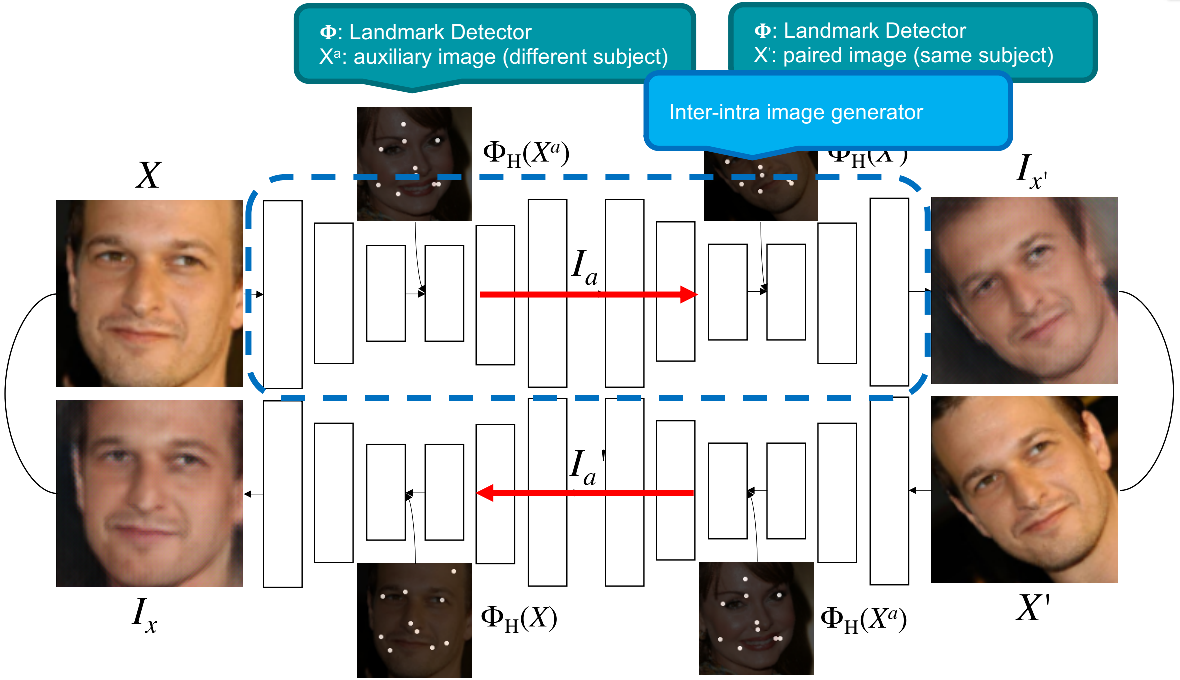


Left: Jakab et al., Unsupervised learning of object landmarks through conditional image generation, *NeurIPS-18*
Right: Sanchez et al., Object Landmark Discovery Through Unsupervised Adaptation, *NeurIPS-19*

Motivation

- Current approaches leverage reconstructing paired images for landmark learning
- Ensure position consistency across different subjects
  - Inject auxiliary landmarks
- Enrich position consistency for the same subject
  - Construct cycle-alike structure

Method



Landmark Detector $u_k = \frac{\sum_i exp(\beta S_k(i))i}{\sum_i exp(\beta S_k(i))}$ $\Phi_H(x;k) = exp(-\frac{1}{2\sigma^2}\|u-u_k\|^2)$

Inter-intra image generator

$$\mathcal{I}_a = \Psi(\mathcal{F}_s, \Phi_H(x^a)) = \Psi(\Phi_E(x), \Phi_H(x^a))$$

$$\mathcal{I} = \Psi(\mathcal{F}_t, \Phi_H(x\prime)) = \Psi(\Phi_E(\mathcal{I}_a), \Phi_H(x\prime))$$

Method

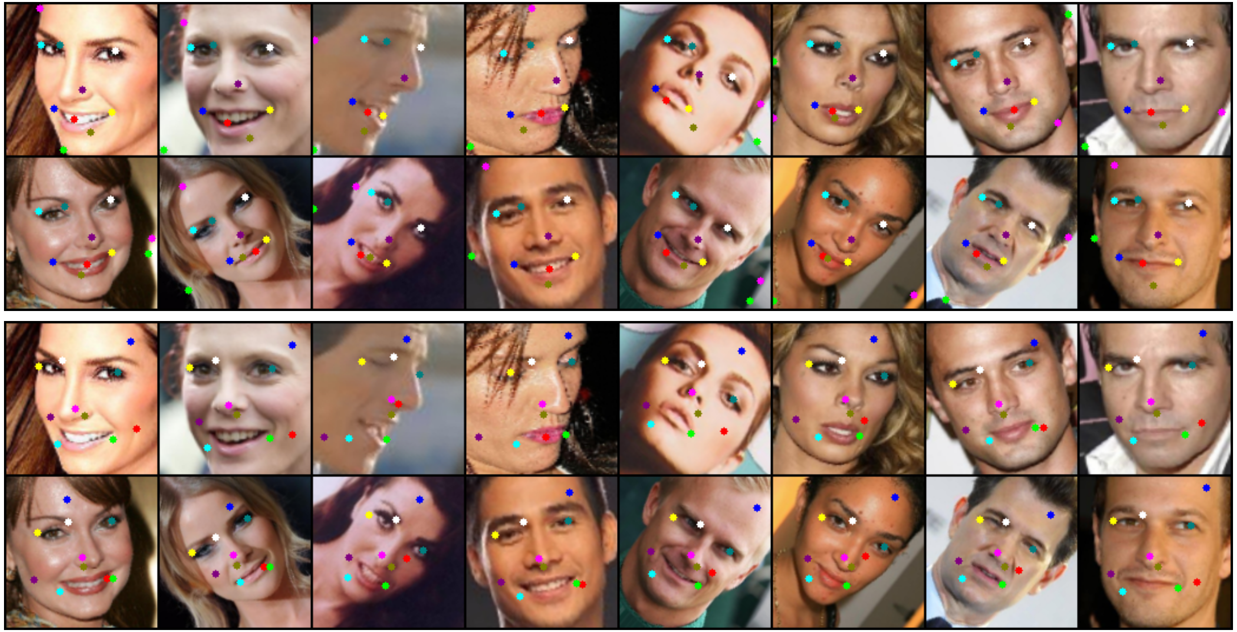
- Training
  - Reconstruction Loss
  - Perceptual Loss

$$\mathcal{L}_R(\mathcal{I}, \mathcal{I}_{gt}) = \|\mathcal{I} - \mathcal{I}_{gt}\|^2$$

$$\mathcal{L}_P(\mathcal{I}, \mathcal{I}_{gt}) = \sum_l \|VGG^l(\mathcal{I}) - VGG^l(\mathcal{I}_{gt})\|^2$$

$$\mathcal{L} = \mathcal{L}_R(\mathcal{I}_x, x) + \mathcal{L}_R(\mathcal{I}_{x\prime}, x\prime) + \mathcal{L}_P(\mathcal{I}_x, x) + \mathcal{L}_P(\mathcal{I}_{x\prime}, x\prime)$$

Results



Results

TABLE I: Normalized MSE evaluations on the public MAFL and AFLW dataset. Baseline*: our re-implementation of [12].

Method	K	MAFL	AFLW
Supervised			
TCDCN [25]		7.95	7.65
RAR [26]		-	7.23
MTCNN [27]		5.39	6.90
Unsupervised			
Thewlis [28]	-	5.83	8.80
Shu [18]	-	5.45	-
Sahasrabudhe [29]	-	6.01	-
Wiles [17]	-	3.44	-
Thewlis [10]	10	7.95	-
Sanchez [12]	10	3.99	6.69
Zhang [14]	10	3.46	7.01
Jakab [11]	10	3.19	6.86
Zhang [14]	30	3.15	6.58
Jakab [11]	30	2.58	6.31
Thewlis [13]	50	2.86	6.54
Jakab [11]	50	**2.54**	6.33
Baseline*	10	3.41	6.59
w. Inter-Subject	10	3.10	6.24
w. Cycle	10	3.12	6.28
Ours-All	10	3.08	6.20
Ours-All	30	2.89	6.08
Ours-All	50	2.85	**6.04**

TABLE II: Normalized MSE evaluations on the MAFL test-set for varying number (N) of supervised samples from MAFL training set used for learning the regressor. We use $K = 10$ intermediate landmarks.

N	Thewlis K=30 [10]	Sanchez K=10 [12]	Jakab K=30 [11]	Ours K=10
1	10.82	18.70	12.89	**9.03**
5	9.25	8.77	8.16	**7.50**
10	8.49	7.13	7.19	**7.09**
100	-	4.53	4.29	**3.71**
500	-	4.13	**2.83**	3.23
1000	-	4.16	**2.73**	3.17
5000	-	4.05	**2.60**	3.09
All	7.15	3.99	**2.58**	3.08

Discussion

- Semantically consistent but geometrically not
  - Include spatial constraints

Thank You!

Contact: weijian.li@rochester.edu

Weijian Li, Haofu Liao, Shun Miao, Le Lu, and Jiebo Luo

Code Available: http://github.com/Weijian-li/unsupervised_inter_intra_landmark