Object landmark discovery through unsupervised adaptation

Enrique Sanchez and Georgios Tzimiropoulos

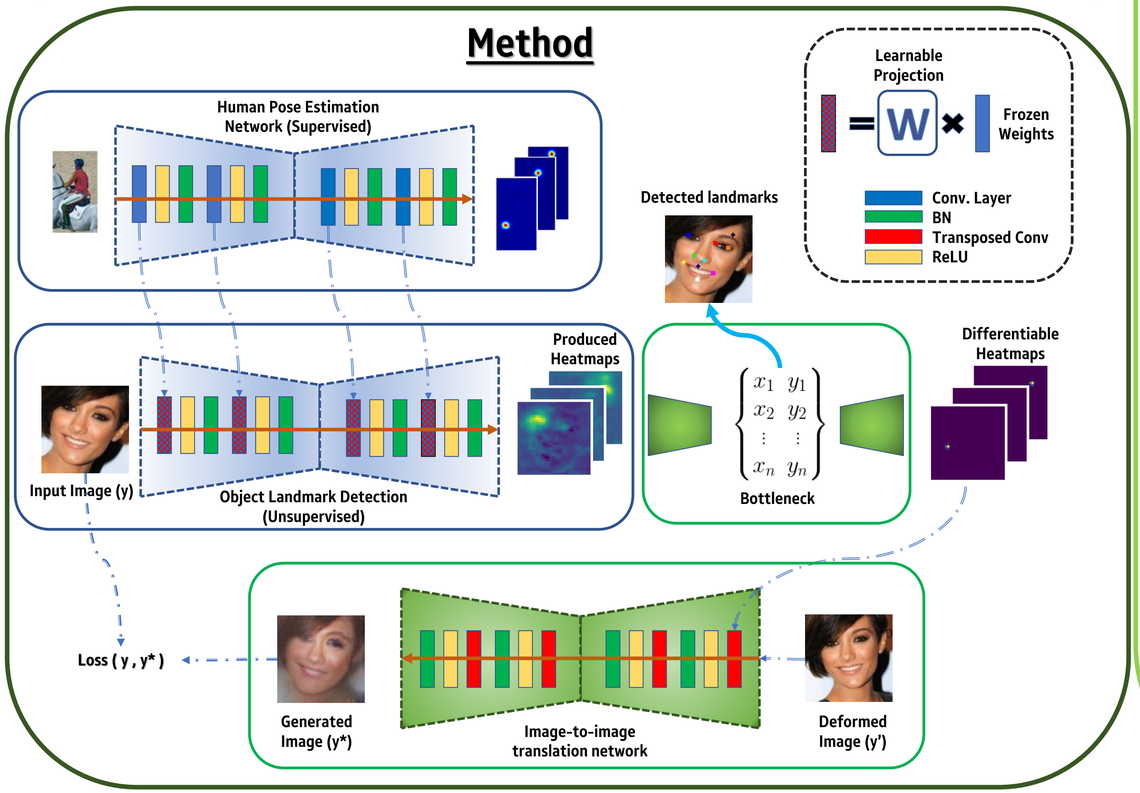
Samsung AI Centre Cambridge – University of Nottingham

https://github.com/ESanchezLozano/SAIC-Unsupervised-landmark-detection-NeurIPS2019

Summary

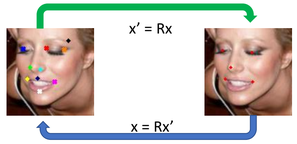
1. **Goal is** to learn an object detector w/o supervision through conditional image generation
2. We propose an **incremental learning** approach to unsupervised learning of object landmark detectors
3. Much constrained learning with **~10% parameters**
4. Novel evaluation that includes measuring the **consistency of** the discovered **landmarks**
5. We compare **three approaches** to unsupervised learning
   i. End-to-end training (scratch)
   ii. Fine-tuning from a pre-trained network
   iii. Incremental learning

Method



Evaluation

Forward

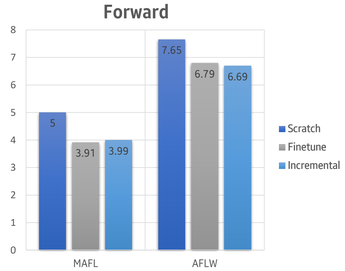


Backward

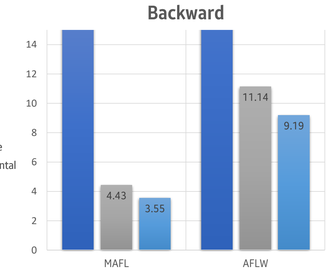
Consistency

$$e_i = \|\Psi^i_{\theta_{\mathcal{Y}}}(A(\mathbf{y})) - A(\Psi^i_{\theta_{\mathcal{Y}}}(\mathbf{y}))\|$$

Forward: Scratch/Finetune/Proposed



Backward: Scratch/Finetune/Proposed



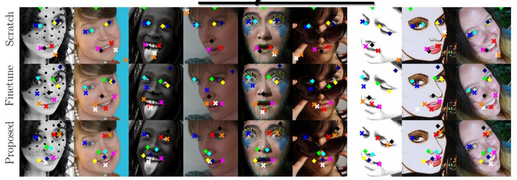
Forward: Comparison s.o.a.

Method	MAFL	AFLW
Supervised		
TCDCN [45]	7.95	7.65
MTCNN [44]	5.39	6.90
Unsupervised		
Thewlis [35]($K = 30$)	7.15	-
Jakab [13]†	3.32	6.99
Jakab [13]††	**3.19**	6.86
Zhang [43]($K = 10$)	3.46	7.01
Shu [31]	5.45	-
Sahasrabudhe [30]	6.01	-
Ours		
Baseline	5.00	7.65
Finetune	3.91	6.79
Proposed	3.99	**6.69**

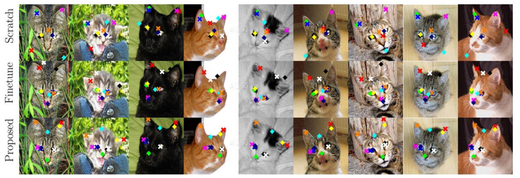
Consistency: Scratch/Finetune/Proposed

		1	2	3	4	5	6	7	8	9	10	Avg.
MAFL	Scratch	1.08	1.20	1.34	1.36	1.38	1.76	3.98	16.51	27.44	35.03	9.11
	Finetune	1.11	1.36	1.39	1.39	1.68	1.83	2.79	3.58	5.59	7.51	2.82
	Proposed	0.96	1.09	1.19	1.34	1.45	1.58	1.80	1.92	3.65	4.09	1.91
AFLW	Scratch	1.45	1.78	1.83	1.85	1.95	2.54	8.46	21.62	31.30	39.37	11.20
	Finetune	1.86	1.93	1.95	2.16	2.18	2.53	5.34	7.30	8.30	9.66	4.32
	Proposed	1.46	1.47	1.47	1.54	1.65	1.66	1.92	2.07	4.97	6.99	2.52
LS3D	Scratch	3.40	4.11	4.48	4.54	5.18	5.71	6.70	19.72	32.04	38.36	12.42
	Finetune	2.93	3.19	3.26	3.59	3.71	4.38	5.14	5.56	7.29	9.49	4.85
	Proposed	2.36	2.48	3.01	3.02	3.55	3.59	3.71	4.83	6.97	7.08	4.06
Shoes	Scratch	1.57	1.65	2.19	2.56	2.79	2.92	3.03	3.05	3.28	4.92	2.80
	Finetune	1.22	1.35	1.42	1.47	1.82	2.03	2.38	2.51	4.21	4.30	2.27
	Proposed	1.07	1.48	1.74	1.80	1.94	2.28	2.30	2.41	2.91	3.49	2.14
Cats	Scratch	1.27	1.44	1.61	1.82	2.30	3.37	3.46	4.44	27.13	28.11	7.50
	Finetune	1.27	1.48	1.81	1.82	1.82	1.84	1.89	5.48	5.93	7.14	3.05
	Proposed	1.00	1.01	1.25	1.60	1.65	1.79	3.57	3.60	3.64	5.29	2.44

Body -> Face



Body -> Cats



Body -> Shoes

Face -> Body

References

Thewlins et al. Unsupervised learning of object landmarks by factorized spatial embeddings. *ICCV'17*
Jakab et al. Unsupervised learning of object landmarks through conditional image generation. *NeurIPS'18*
Zhang et al. Unsupervised discovery of object landmarks as structural representations. *CVPR'18*