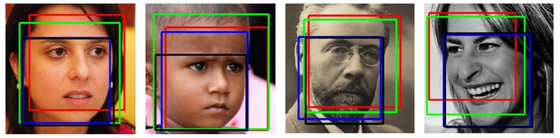
A Deep Regression Architecture with Two-Stage Re-initialization for High Performance Facial Landmark Detection

Jiangjing Lv[1*], Xiaohu Shao[1*], Junliang Xing[2], Cheng Cheng[1], Xi Zhou[1]
[1]Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences
[2]National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences

IEEE 2017 Conference on Computer Vision and Pattern Recognition
CVPR July 21-26 2017 HONOLULU

Observations

- Most of facial landmark detection algorithms are sensitive to the initialization brought by face detection results.
- Different face detectors often return various face bounding boxes with different scales and center shifts.



Motivations

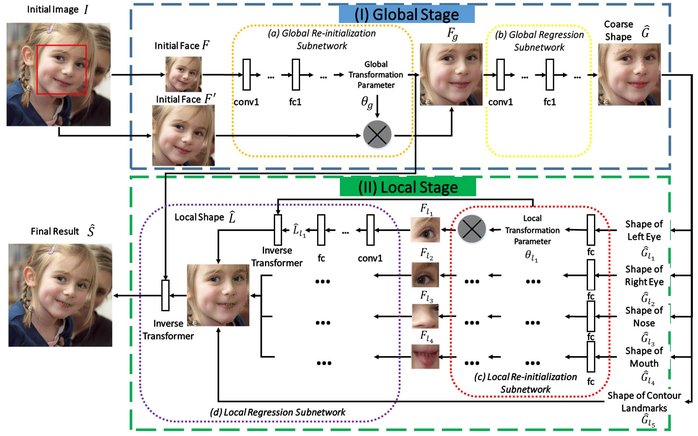
- Face boxes bounding by the ground truth shapes can boost the accuracy of facial landmark detection.
- Good performance of STN [14] on learning instance-specific transformations of training samples.

Contributions

- An end-to-end deep regression with two-stage re-initialization architecture.
- Supervised spatial transformer learning.
- Good robustness to different initialization.
- State-of-the-art performances on 300-W and AFLW.

Our Approach

- Our two-stage re-initialization model successively re-initializes a deep regression model from coarse to fine, and global to local, to substantially boost the landmark detection performance.
- Global Stage:
  - Global Re-initialization Subnetwork
  - Global Regression Subnetwork
- Local Stage:
  - Local Re-initialization Subnetwork
  - Local Regression Subnetwork



Global Stage

- The best canonical state for initialization of following regression is obtained by the supervised transformer subnetwork.

$$F_g = \mathbf{A}(F', T_{\theta_g})$$

- The target of the global regression is the transformed ground truth shape:

$$\mathbf{G}^* = T_{\theta_g}^{-1} \begin{pmatrix} \mathbf{S}^* \\ \mathbf{1} \end{pmatrix}$$

Local Stage

- Different parts of the face shape are further separately re-initialized to their own canonical states.

- Local regression to minimize the loss function of shape increment.
- Inverse Transformer: $\hat{\mathbf{S}}_{l_n} = T_R T_{\theta_g} T_{\theta_{l_n}} \begin{pmatrix} \hat{\mathbf{L}}_{l_n} \\ \mathbf{1} \end{pmatrix}$

Experiments

We first evaluate the robustness of our approach for various initialization, and then compare it with other state-of-the-art methods on the benchmark datasets.

Detectors	Common Subset	Challenging Subset	Full Set
VJ_{B_1}	8.90	14.39	9.98
$Dlib_{B_1}$	6.88	12.40	7.96
OD_{B_1}	5.43	8.97	6.12
GT_{B_1}	5.24	7.65	5.71
VJ_{B_2}	6.19	10.15	6.96
$Dlib_{B_2}$	5.30	9.13	6.05
OD_{B_2}	5.03	8.43	5.69
GT_{B_2}	5.04	7.64	5.55
VJ_{P^-}	4.95	8.36	5.62
$Dlib_{P^-}$	4.87	8.30	5.55
OD_{P^-}	4.56	8.16	5.27
GT_{P^-}	4.43	7.08	5.05
VJ_P	**4.50**	**7.89**	**5.16**
$Dlib_P$	**4.42**	**7.80**	**5.09**
OD_P	**4.36**	**7.56**	**4.99**
GT_P	**4.36**	**7.42**	**4.96**

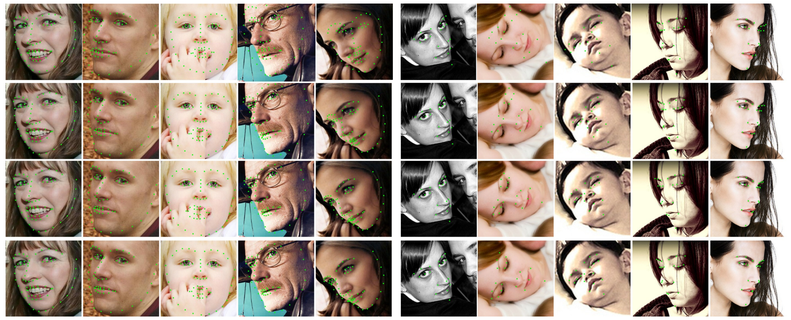
Robustness to Various Initialization

Method	Common Subset	Challenging Subset	Full Set
RCPR [2]	6.18	17.26	8.35
SDM [30]	5.57	15.40	7.52
ESR [5]	5.28	17.00	7.58
CFAN [35]	5.50	16.78	7.69
DeepReg [23]	4.51	13.80	6.31
LBF [21]	4.95	11.98	6.32
CFSS [38]	4.73	9.98	5.76
TCDCN [36]	4.80	8.60	5.54
DDN [34]	-	-	5.59
MDM [25]	4.83	10.14	5.88
$Baseline_1$	5.43	8.97	6.12
$Baseline_2$	5.03	8.43	5.69
$Proposed^-$	4.56	8.16	5.27
$Proposed$	**4.36**	**7.56**	**4.99**

Comparison with the State-of-the-arts on 300-W

Method	CDM [33]	RCPR	SDM	ERT [16]	LBF	CFSS	CCL [39]	$Baseline_1$	$Baseline_2$	$Proposed^-$	$Proposed$
NME	5.43	3.73	4.05	4.35	4.25	3.92	2.72	2.99	2.68	2.33	**2.17**

Comparison with the State-of-the-arts on AFLW



Comparison with the Baselines on 300-W and AFLW