Face alignment & landmark

CVPR, AAAI(2016-2019)

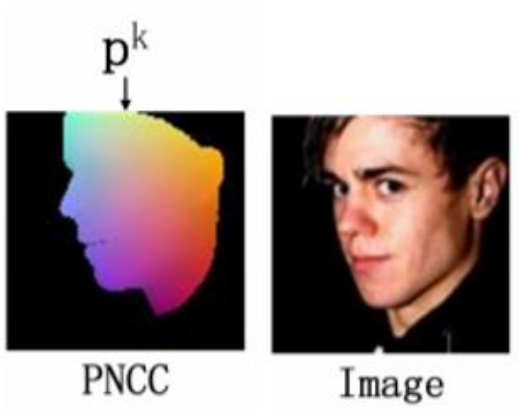
염지현

CVPR 2016 ~ 2018

Face Alignment Across Large Poses: A 3D Solution

Zhu, X., Lei, Z., Lius, X., Shi, H. and Li, S. Z., Face Alignment Across Large Poses: A 3D Solution, Proc. of CVPR 2016, PP. 146-155, 2016.

Input:



Output: Δp^k

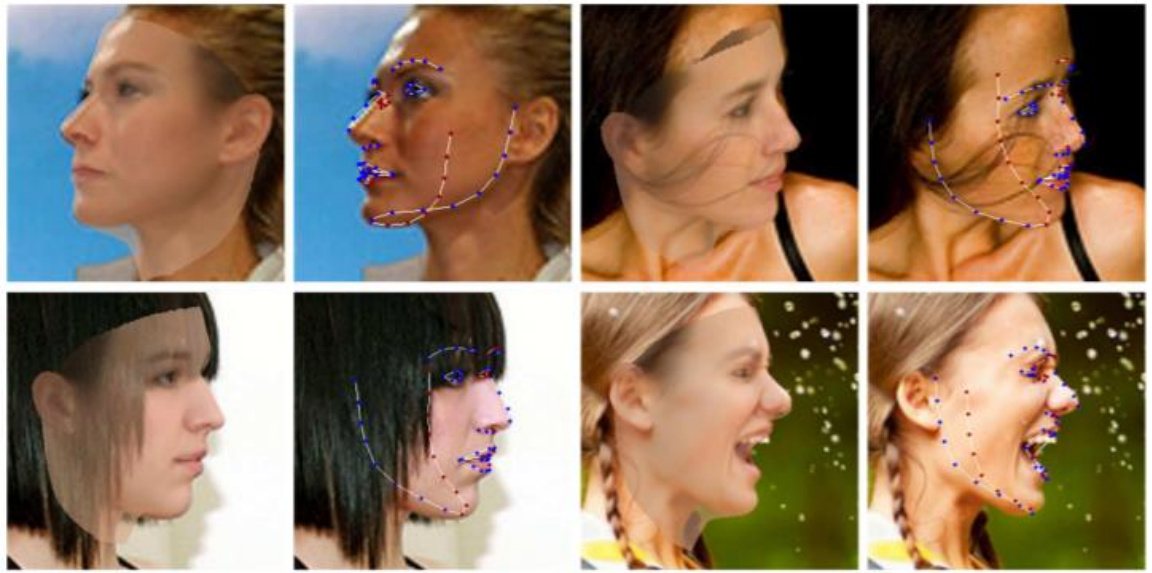


Figure 1. Fitting results of 3DDFA. For each pair of the four results, on the left is the rendering of the fitted 3D shape with the mean texture, which is made transparent to demonstrate the fitting accuracy. On the right is the landmarks overlayed on the 3D face model, in which the blue/red ones indicate visible/invisible landmarks. The visibility is directly computed from the fitted dense model by [21]. More results are demonstrated in supplemental material.

Face Alignment Across Large Poses: A 3D Solution

Zhu, X., Lei, Z., Lius, X., Shi, H. and Li, S. Z., Face Alignment Across Large Poses: A 3D Solution, Proc. of CVPR 2016, PP. 146-155, 2016.

- 하위 수준 feature 추출
- 가중치 공유

- 민감한 feature 추출
- 가중치 공유하지 않음

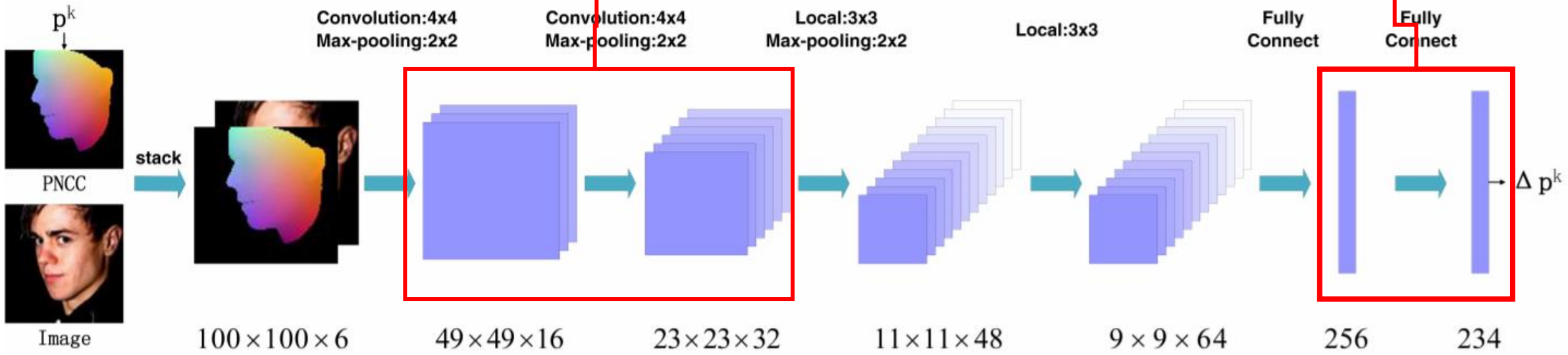


Figure 2. An overview of 3DDFA. At kth iteration, Net^k takes a medium parameter $\mathbf{p}^k$ as input, constructs the projected normalized coordinate code (PNCC), stacks it with the input image and sends it into CNN to predict the parameter update $\Delta\mathbf{p}^k$.

Face Alignment Across Large Poses: A 3D Solution

Zhu, X., Lei, Z., Lius, X., Shi, H. and Li, S. Z., Face Alignment Across Large Poses: A 3D Solution, Proc. of CVPR 2016, PP. 146-155, 2016.

Normalized mean error

Table 1. The NME(%) of face alignment results on AFLW and AFLW2000-3D with the first and the second best results highlighted. The bracket shows the training set. The results of provided alignment models are marked with their references.

	AFLW Dataset (21 pts)					AFLW2000-3D Dataset (68 pts)				
Method	[0, 30]	[30, 60]	[60, 90]	Mean	Std	[0, 30]	[30, 60]	[60, 90]	Mean	Std
CDM [49]	8.15	13.02	16.17	12.44	4.04	-	-	-	-	-
RCPR [7]	6.16	18.67	34.82	19.88	14.36	-	-	-	-	-
RCPR(300W)	5.40	9.80	20.61	11.94	7.83	4.16	9.88	22.58	12.21	9.43
RCPR(300W-LP)	5.43	6.58	11.53	7.85	3.24	4.26	5.96	13.18	7.80	4.74
ESR(300W)	5.58	10.62	20.02	12.07	7.33	4.38	10.47	20.31	11.72	8.04
ESR(300W-LP)	5.66	7.12	11.94	8.24	3.29	4.60	6.70	12.67	7.99	4.19
SDM(300W)	**4.67**	6.78	16.13	9.19	6.10	**3.56**	7.08	17.48	9.37	7.23
SDM(300W-LP)	**4.75**	5.55	9.34	6.55	2.45	3.67	4.94	9.76	6.12	3.21
3DDFA	5.00	**5.06**	**6.74**	**5.60**	**0.99**	3.78	**4.54**	**7.93**	**5.42**	**2.21**
3DDFA+SDM	**4.75**	**4.83**	**6.38**	**5.32**	**0.92**	**3.43**	**4.24**	**7.17**	**4.94**	**1.97**

Face Alignment Across Large Poses: A 3D Solution

Zhu, X., Lei, Z., Lius, X., Shi, H. and Li, S. Z., Face Alignment Across Large Poses: A 3D Solution, Proc. of CVPR 2016, PP. 146-155, 2016.

Table 2. The NME(%) of face alignment results on 300W, with the first and the second best results highlighted.

Method	Common	Challenging	Full
TSPM [56]	8.22	18.33	10.20
ESR [10]	5.28	17.00	7.58
RCPR [7]	6.18	17.26	8.35
SDM [45]	5.57	15.40	7.50
LBF [32]	**4.95**	11.98	6.32
CFSS [54]	**4.73**	**9.98**	**5.76**
3DDFA	6.15	10.59	7.01
3DDFA+SDM	5.53	**9.56**	**6.31**

Mnemonic Descent Method: A recurrent process applied for end-to-end face alignment

Trigeorgis, G., Snapc, P., Nicolaous, M. A., and Antonakos, E., Mnemonic Descent Method: A recurrent process applied for end-to-end face alignment, Proc. of CVPR 2016, PP. 4177-4187, 2016.

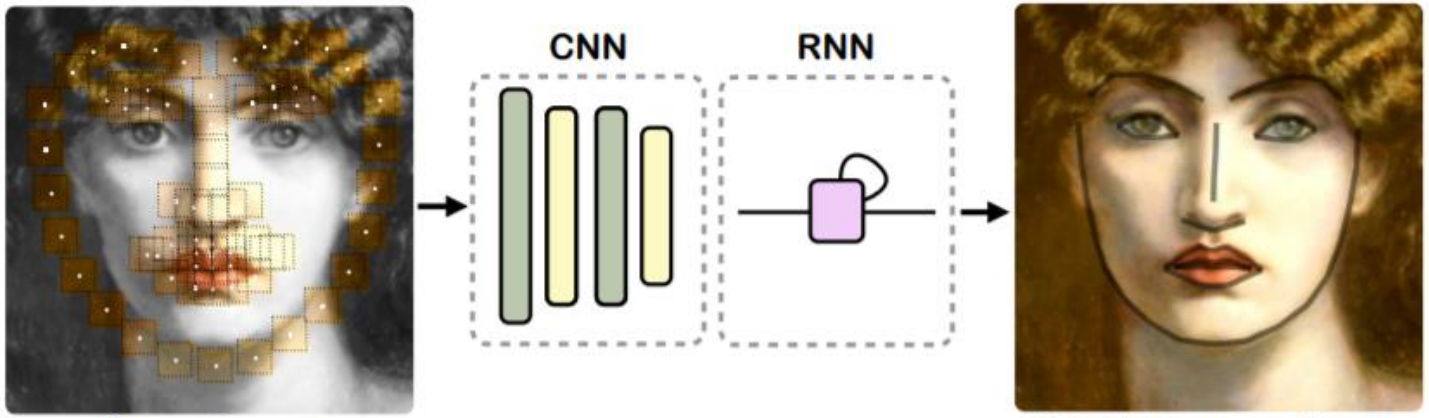


Figure 1: Mnemonic Descent Method learns to align a shape estimate to a facial image[1] in an end-to-end manner using a jointly learnt convolutional recurrent network architecture.

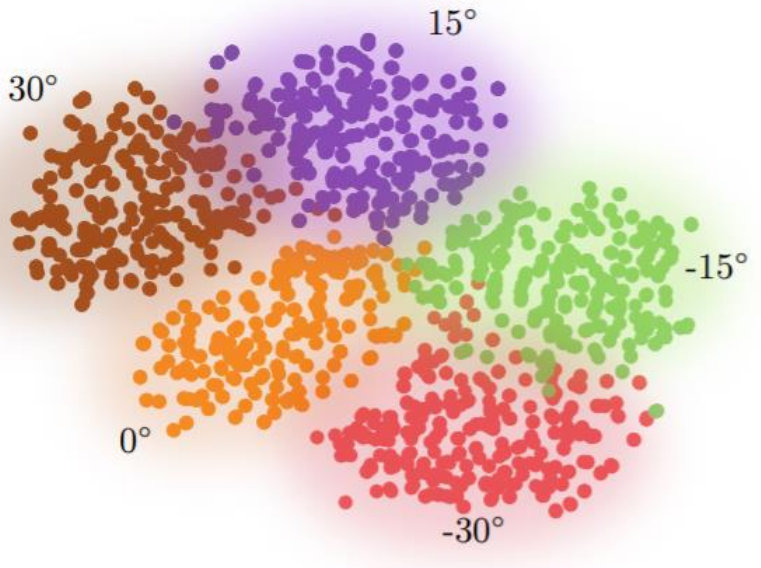


Figure 3: A t-SNE depiction of the internal states ($T = 1$) of MDM when asked to align 2000 randomly selected images of CMU Multi-PIE [24]. Each colour corresponds to a cluster of head pose. This visualisation demonstrates that MDM is effectively partitioning the input data based on the head pose. Best viewed in colour.

Mnemonic Descent Method: A recurrent process applied for end-to-end face alignment

Trigeorgis, G., Snapc, P., Nicolaous, M. A., and Antonakos, E., Mnemonic Descent Method: A recurrent process applied for end-to-end face alignment, Proc. of CVPR 2016, PP. 4177-4187, 2016.

다음 스텝에서 집중해야 하는 위치를 나타냄

Conv

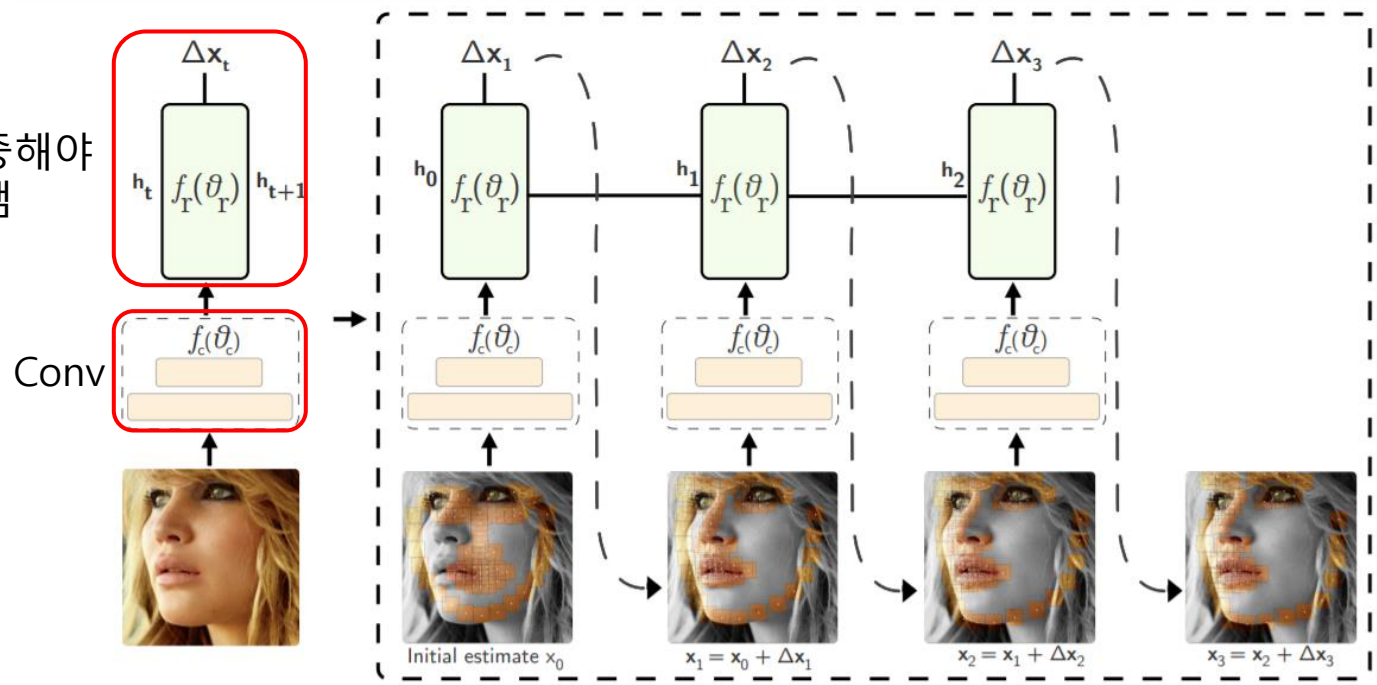


Figure 2: An illustrative example of MDM for a total of $T = 3$ time-steps. Initially the network input consists of a partial image observation, consisting of the patches extracted at the mean face $\mathbf{x}_0$. The extracted patches (30×30) at each time-step are passed through a subsequent convolutional network $f_c(\cdot;\theta_c)$, which in turn produces a representation that is robust to changes in appearance variation. Based on the current state $\mathbf{h}_t$, the mnemonic module (implemented as a recurrent network) generates a new state $\mathbf{h}_{t+1}$ and a new set of descent directions $\Delta\mathbf{x}_{t+1}$ that indicates where the network should focus next. After a total of $T = 3$ time-steps, MDM successfully estimates the landmark locations. An important distinction from the previous work on cascade models [56] is that the weights of the network $\theta = \{\theta_c, \theta_r, \mathbf{x}\}$ are *shared* across time.

Mnemonic Descent Method: A recurrent process applied for end-to-end face alignment

Trigeorgis, G., Snapc, P., Nicolaous, M. A., and Antonakos, E., Mnemonic Descent Method: A recurrent process applied for end-to-end face alignment, Proc. of CVPR 2016, PP. 4177-4187, 2016.

	51-points		68-points	
Method	**AUC**	**Failure (%)**	**AUC**	**Failure (%)**
ERT [26]	40.60	13.50	32.35	17.00
PO-CR [50]	47.65	11.70	–	–
Chehra [3]	31.12	39.30	–	–
Intraface [56]	38.47	19.70	–	–
Balt. et al. [5]	37.65	17.17	19.55	38.83
Face++ [63]	53.29	5.33	32.81	13.00
Yan et al. [58]	49.07	8.33	34.97	12.67
CFSS [64]	50.79	7.80	39.81	12.30
MDM	**56.34**	**4.20**	**45.32**	**6.80**

Table 1: Quantitative results on the test set of the 300W competition using the AUC (%) and failure rate (calculated at a threshold of 0.08 of the normalised error).

Mnemonic Descent Method: A recurrent process applied for end-to-end face alignment

Trigeorgis, G., Snapc, P., Nicolaous, M. A., and Antonakos, E., Mnemonic Descent Method: A recurrent process applied for end-to-end face alignment, Proc. of CVPR 2016, PP. 4177-4187, 2016.

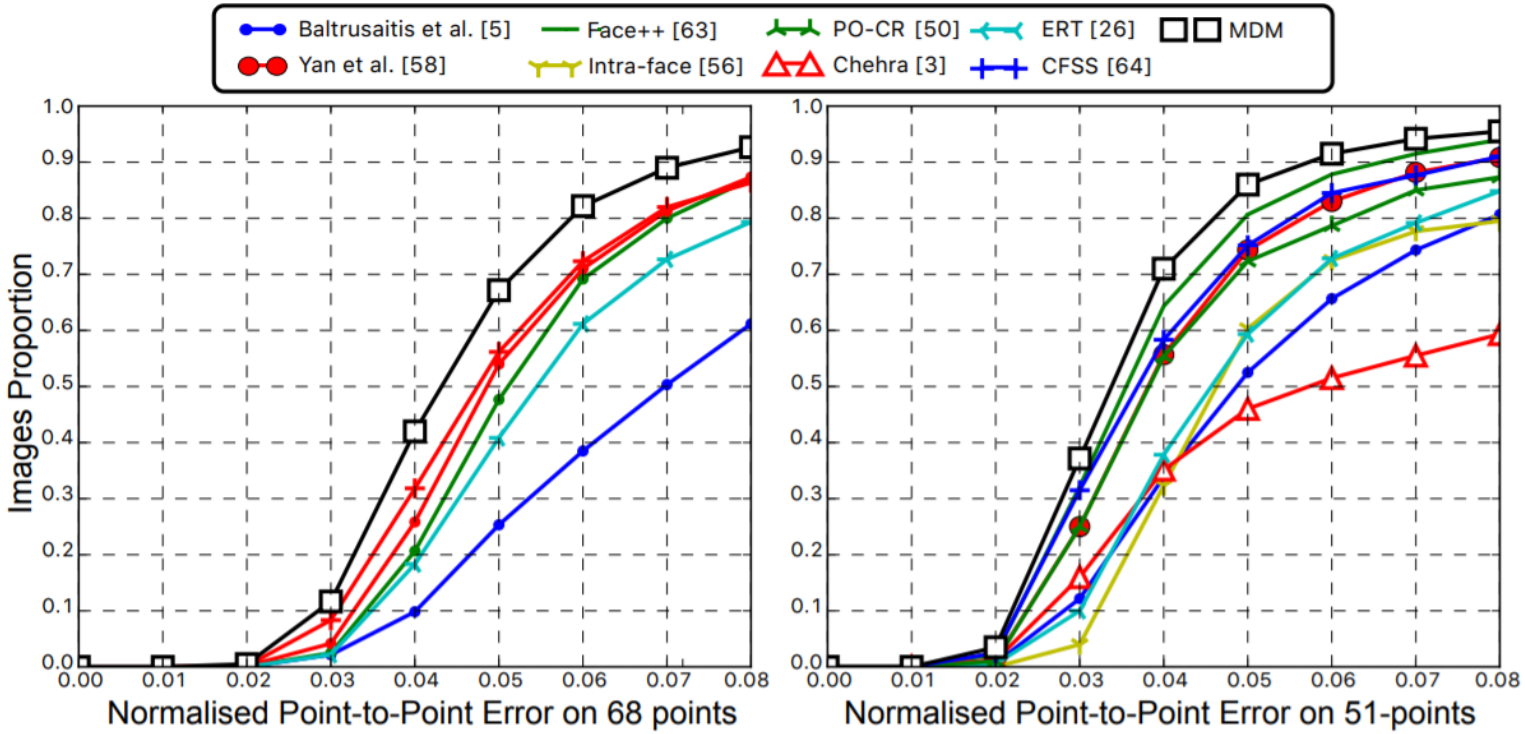


Figure 4: Quantitative results on the test set of the 300W competition (indoor and outdoor combined) [39] for both 68-point (left) and 51-point (right) plots. Only the top 3 performing results from the original competition are shown.

A Deep Regression Architecture With Two-Stage Re-Initialization for High Performance Facial Landmark Detection

Lv, J., Shao, X., Xing, J., Chen, C., and Zhou, X., A Deep Regression Architecture With Two-Stage Re-Initialization for High Performance Facial Landmark Detection, Proc. of CVPR 2017, pp. 3317-3326, 2017.

Input:

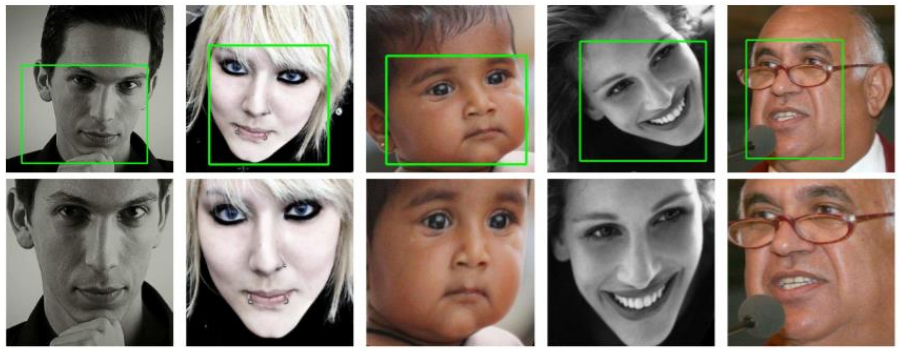
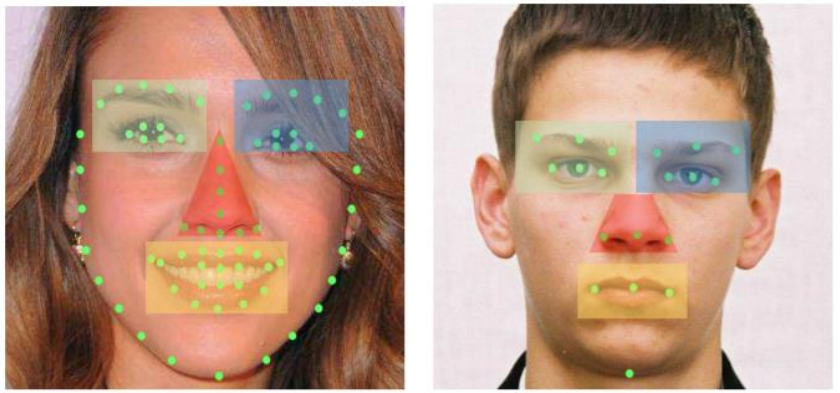


Figure 3. The results of the global re-initialization subnetwork. Top row: the input initial face images with initial face boxes. Bottome row: the transformed face images output by the global re-initialization subnetwork.

Output:



(a) (b)

Figure 4. Four parts of landmarks in the local re-initialization subnetwork, (a) 68 landmarks of 300-W, (b) 19 landmarks of AFLW.

A Deep Regression Architecture With Two-Stage Re-Initialization for High Performance Facial Landmark Detection

Lv, J., Shao, X., Xing, J., Chen, C., and Zhou, X., A Deep Regression Architecture With Two-Stage Re-Initialization for High Performance Facial Landmark Detection, Proc. of CVPR 2017, pp. 3317-3326, 2017.

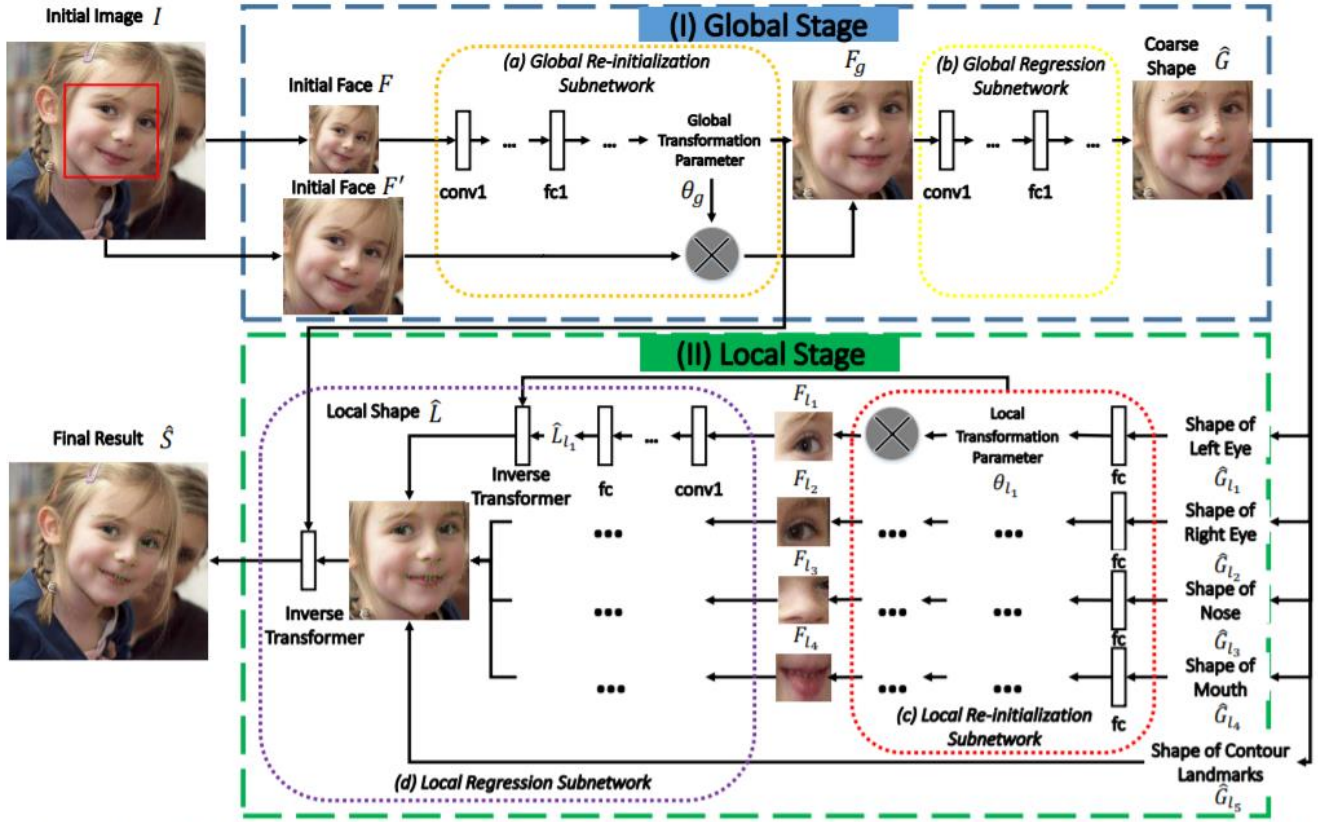


Figure 1. The pipeline of the proposed deep regression architecture with two-stage re-initialization for coarse-to-fine facial landmark detection. At the global stage (I), the face region is firstly re-initialized to a canonical shape state (a), and then regress a coarse shape (b). At the local stage (II), different face parts are further separately re-initialized to their own canonical shape states (c), followed by another regression subnetwork to get the final detection(d).

A Deep Regression Architecture With Two-Stage Re-Initialization for High Performance Facial Landmark Detection

Lv, J., Shao, X., Xing, J., Chen, C., and Zhou, X., A Deep Regression Architecture With Two-Stage Re-Initialization for High Performance Facial Landmark Detection, Proc. of CVPR 2017, pp. 3317-3326, 2017.

Table 2. The comparison of NME without and with using our proposed method on 300-W dataset based on different face extended scales (a), translations (b), rotations (c).

(a) Different Scales

Scale	0.1	0.2	0.3	0.4	0.5
$Baseline_1$	6.12	7.11	9.98	15.67	24.43
$Baseline_2$	5.69	6.27	7.05	9.59	13.54
$Proposed^-$	5.27	5.17	5.30	5.65	6.13
$Proposed$	**4.99**	**5.03**	**5.11**	**5.39**	**5.93**

(b) Different Translations

Translation	0.05	0.10	0.15	0.20	0.25
$Baseline_1$	6.29	6.96	8.51	11.86	18.67
$Baseline_2$	5.75	6.01	6.91	8.48	12.84
$Proposed^-$	5.28	5.46	5.61	5.96	6.46
$Proposed$	**5.01**	**5.15**	**5.26**	**5.36**	**5.77**

(c) Different Rotations

Rotation (°)	5	10	15	20	25
$Baseline_1$	6.35	6.98	7.91	9.35	11.71
$Baseline_2$	6.11	6.57	7.36	8.40	9.90
$Proposed^-$	5.48	5.60	5.75	6.03	6.43
$Proposed$	**5.13**	**5.24**	**5.42**	**5.77**	**6.20**

Table 3. The performance of our proposed method compared with other methods on 300-W dataset.

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**

A Deep Regression Architecture With Two-Stage Re-Initialization for High Performance Facial Landmark Detection

Lv, J., Shao, X., Xing, J., Chen, C., and Zhou, X., A Deep Regression Architecture With Two-Stage Re-Initialization for High Performance Facial Landmark Detection, Proc. of CVPR 2017, pp. 3317-3326, 2017.

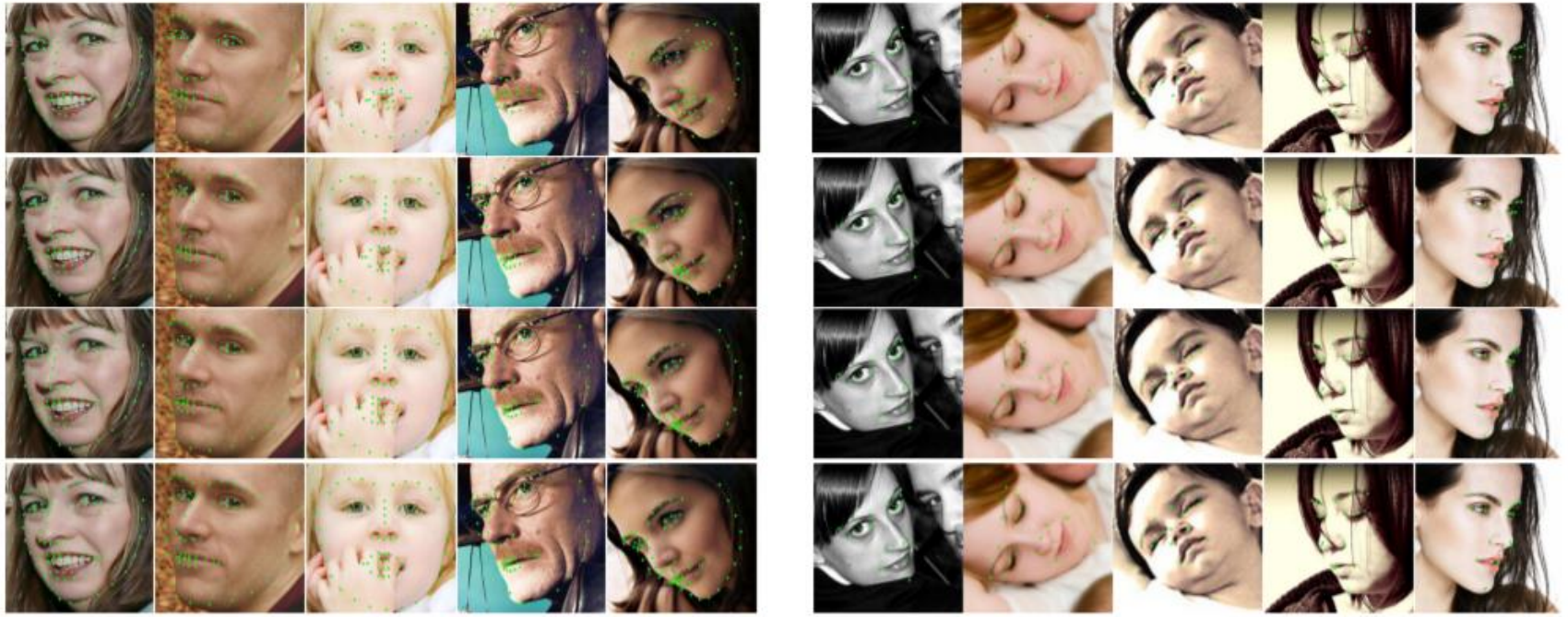


Figure 6. The comparison of facial landmark detection results on 300-W dataset (left) and AFLW dataset (right): The images are the results of $Baseline_1$ method, the $Baseline_2$ method, the $Proposed^-$ method and the $Proposed$ method from top to bottom of the single line.

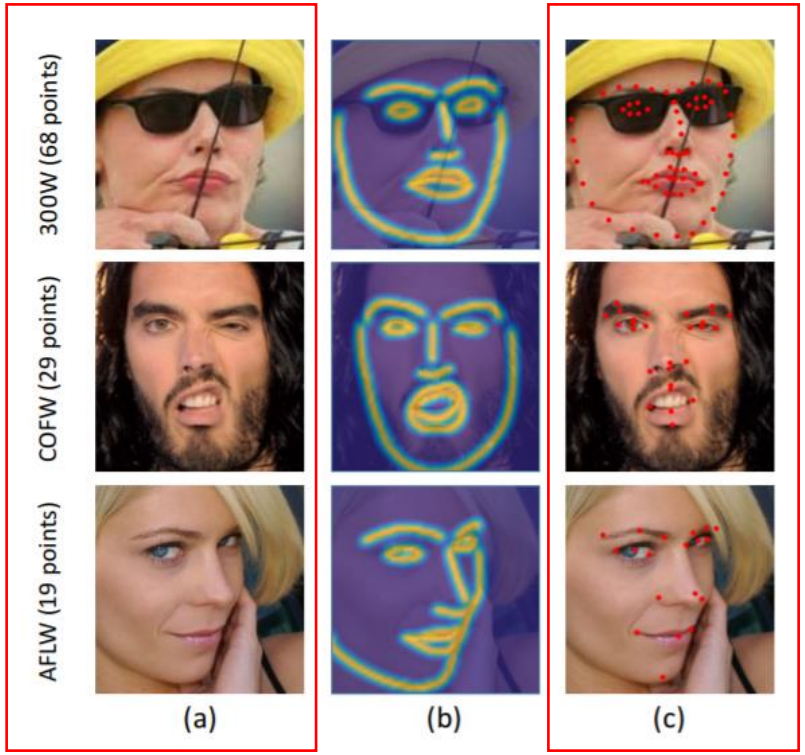
Table 4. Mean Error normalized by face size on AFLW dataset compared with other state-of-the-art methods .

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**

Look at Boundary: A Boundary-Aware Face Alignment Algorithm

Wu, W., Qian, C., Yang, S., Wang, Q., Cai, Y., and Zhou, Q., Look at Boundary: A Boundary-Aware Face Alignment Algorithm, Proc. of CVPR 2018, pp. 2129-2138, 2018.

input



output

Figure 1: The first column shows the face images from different datasets with different number of landmarks. The second column illustrates the universally defined facial boundaries estimated by our methods. With the help of boundary information, our approach achieves high accuracy localisation results across multiple datasets and annotation protocols, as shown in the third column.

Look at Boundary: A Boundary-Aware Face Alignment Algorithm

Wu, W., Qian, C., Yang, S., Wang, Q., Cai, Y., and Zhou, Q., Look at Boundary: A Boundary-Aware Face Alignment Algorithm, Proc. of CVPR 2018, pp. 2129-2138, 2018.

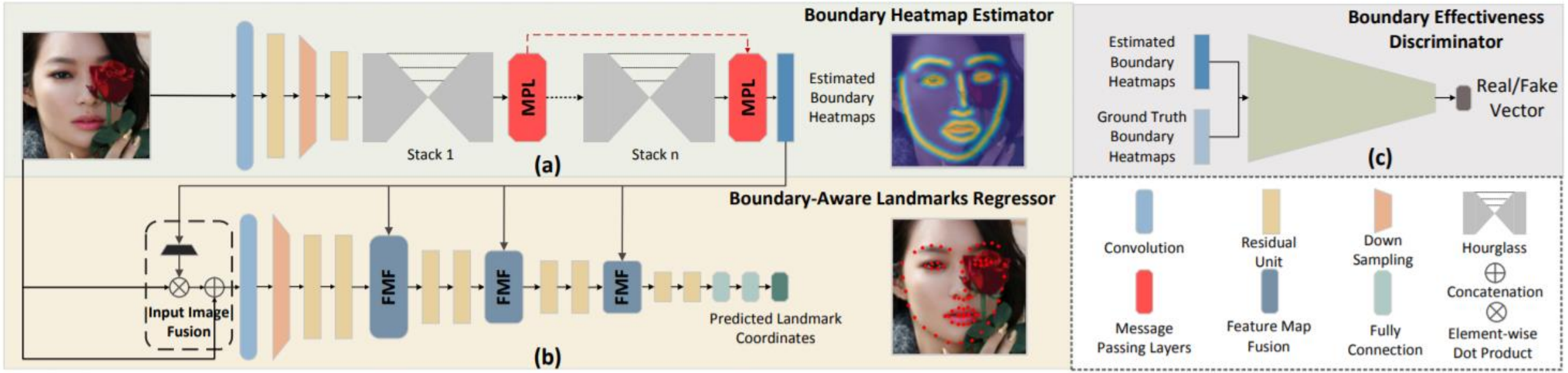


Figure 2: Overview of our Boundary-Aware Face Alignment framework. (a) Boundary heatmap estimator, which based on hourglass network is used to estimate boundary heatmaps. Message passing layers are introduced to handle occlusion. (b) Boundary-aware landmarks regressor is used to generate the final prediction of landmarks. Boundary heatmap fusion scheme is introduced to incorporate boundary information into the feature learning of regressor. (c) Boundary effectiveness discriminator, which distinguishes "real" boundary heatmaps from "fake", is used to further improve the quality of the estimated boundary heatmaps.

Look at Boundary: A Boundary-Aware Face Alignment Algorithm

Wu, W., Qian, C., Yang, S., Wang, Q., Cai, Y., and Zhou, Q., Look at Boundary: A Boundary-Aware Face Alignment Algorithm, Proc. of CVPR 2018, pp. 2129-2138, 2018.

Method	Common Subset	Challenging Subset	Fullset
Inter-pupil Normalisation			
RCPR [6]	6.18	17.26	8.35
CFAN [69]	5.50	16.78	7.69
ESR [7]	5.28	17.00	7.58
SDM [57]	5.57	15.40	7.50
LBF [38]	4.95	11.98	6.32
CFSS [72]	4.73	9.98	5.76
3DDFA [74]	6.15	10.59	7.01
TCDCN [70]	4.80	8.60	5.54
MDM [51]	4.83	10.14	5.88
RAR [55]	4.12	8.35	4.94
DVLN [53]	3.94	7.62	4.66
TSR [31]	4.36	7.56	4.99
LAB	**3.42**	**6.98**	**4.12**
LAB+Oracle	2.57	4.72	2.99
Inter-ocular Normalisation			
PCD-CNN [2]	3.67	7.62	4.44
SAN [59]	3.34	6.60	3.98
LAB	**2.98**	**5.19**	**3.49**
LAB+Oracle	1.85	3.28	2.13

Table 1: Mean error (%) on 300-W Common Subset, Challenging Subset and Fullset (68 landmarks).

Method	AUC	Failure Rate (%)
Deng *et al.* [14]	0.4752	5.5
Fan *et al.* [16]	0.4802	14.83
DenseReg + MDM [1]	0.5219	3.67
JMFA [15]	0.5485	1.00
LAB	**0.5885**	**0.83**
LAB+Oracle	0.7626	0.00

Table 2: Mean error (%) on 300-W testset (68 landmarks). Accuracy is reported as the AUC and the Failure Rate.

Look at Boundary: A Boundary-Aware Face Alignment Algorithm

Wu, W., Qian, C., Yang, S., Wang, Q., Cai, Y., and Zhou, Q., Look at Boundary: A Boundary-Aware Face Alignment Algorithm, Proc. of CVPR 2018, pp. 2129-2138, 2018.

Metric	Method	Testset	Pose Subset	Expression Subset	Illumination Subset	Make-Up Subset	Occlusion Subset	Blur Subset
Mean Error (%)	ESR [7]	11.13	25.88	11.47	10.49	11.05	13.75	12.20
	SDM [57]	10.29	24.10	11.45	9.32	9.38	13.03	11.28
	CFSS [72]	9.07	21.36	10.09	8.30	8.74	11.76	9.96
	DVLN [53]	6.08	11.54	6.78	5.73	5.98	7.33	6.88
	LAB	**5.27**	**10.24**	**5.51**	**5.23**	**5.15**	**6.79**	**6.32**
Failure Rate (%)	ESR [7]	35.24	90.18	42.04	30.80	38.84	47.28	41.40
	SDM [57]	29.40	84.36	33.44	26.22	27.67	41.85	35.32
	CFSS [72]	20.56	66.26	23.25	17.34	21.84	32.88	23.67
	DVLN [53]	10.84	46.93	11.15	7.31	11.65	16.30	13.71
	LAB	**7.56**	**28.83**	**6.37**	**6.73**	**7.77**	**13.72**	**10.74**
AUC	ESR [7]	0.2774	0.0177	0.1981	0.2953	0.2485	0.1946	0.2204
	SDM [57]	0.3002	0.0226	0.2293	0.3237	0.3125	0.2060	0.2398
	CFSS [72]	0.3659	0.0632	0.3157	0.3854	0.3691	0.2688	0.3037
	DVLN [53]	0.4551	0.1474	0.3889	0.4743	0.4494	0.3794	0.3973
	LAB	**0.5323**	**0.2345**	**0.4951**	**0.5433**	**0.5394**	**0.4490**	**0.4630**

Table 3: Evaluation of LAB and several state-of-the-arts on Testset and 6 typical subsets of WFLW (98 landmarks).

Look at Boundary: A Boundary-Aware Face Alignment Algorithm

Wu, W., Qian, C., Yang, S., Wang, Q., Cai, Y., and Zhou, Q., Look at Boundary: A Boundary-Aware Face Alignment Algorithm, Proc. of CVPR 2018, pp. 2129-2138, 2018.

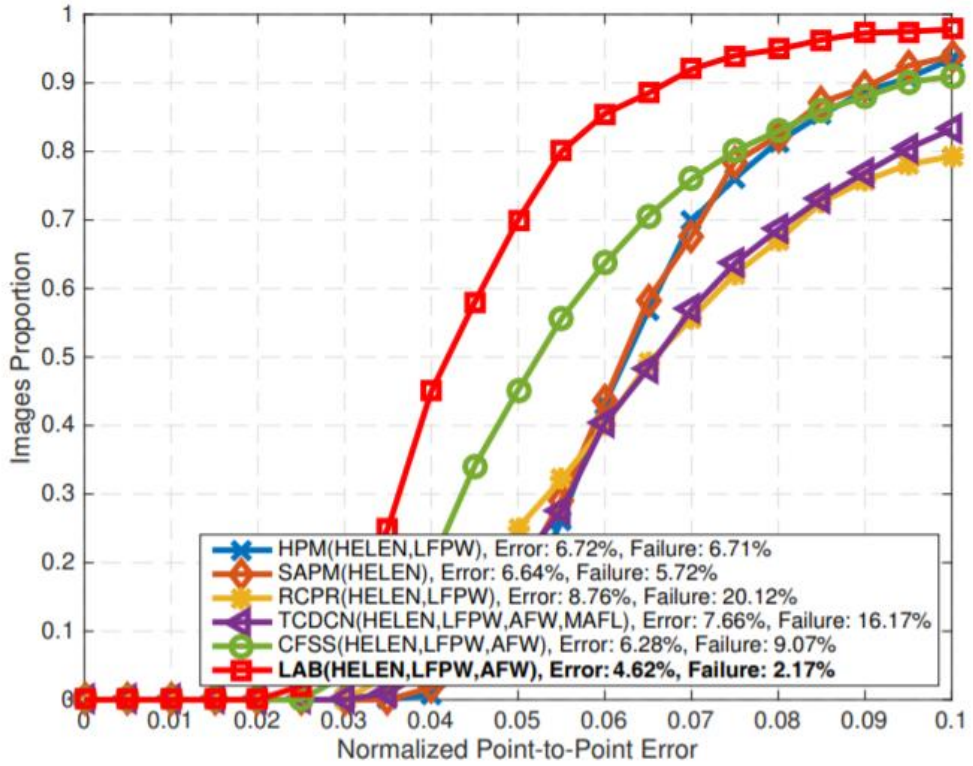


Figure 7: CED for COFW-68 testset (68 landmarks). Train set (in parentheses), mean error and failure rate are also reported.

Super-FAN: Integrated Facial Landmark Localization and Super-Resolution of Real-World Low Resolution Faces in Arbitrary Poses With GANs

Bulat, A., Tzimiropoulos, G., Super-FAN: Integrated Facial Landmark Localization and Super-Resolution of Real-World Low Resolution Faces in Arbitrary Poses With GANs, Proc. of CVPR 2018, PP. 109-117, 2018.

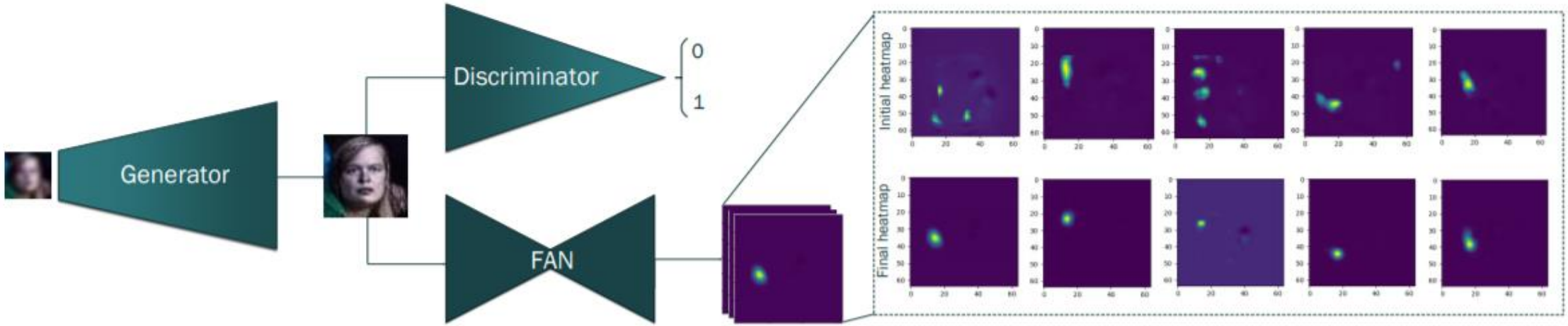


Figure 2: The proposed Super-FAN architecture comprises three connected networks: the first network is a newly proposed Super-resolution network (see sub-section 4.1). The second network is a WGAN-based discriminator used to distinguish between the super-resolved and the original HR image (see sub-section 4.2). The third network is FAN, a face alignment network for localizing the facial landmarks on the super-resolved facial image and improving super-resolution through a newly-introduced heatmap loss (see sub-section 4.3).

Super-FAN: Integrated Facial Landmark Localization and Super-Resolution of Real-World Low Resolution Faces in Arbitrary Poses With GANs

Bulat, A., Tzimiropoulos, G., Super-FAN: Integrated Facial Landmark Localization and Super-Resolution of Real-World Low Resolution Faces in Arbitrary Poses With GANs, Proc. of CVPR 2018, PP. 109-117, 2018.

Method	PSNR			SSIM		
	30	60	90	30	60	90
bilinear upsample (baseline)	20.25	21.45	22.10	0.7248	0.7618	0.7829
SR-ResNet	21.21	22.23	22.83	0.7764	0.7962	0.8077
SR-GAN	20.01	20.94	21.48	0.7269	0.7465	0.7586
Ours-pixel	**21.55**	22.45	23.05	**0.8001**	**0.8127**	**0.8240**
Ours-pixel-feature	21.50	22.51	23.10	0.7950	0.7970	0.8205
Ours-pixel-feature-heatmap	**21.55**	**22.55**	**23.17**	0.7960	0.8105	0.8210
Ours-Super-FAN	20.85	21.67	22.24	0.7745	0.7921	0.8025

Table 1: PSNR- and SSIM-based super-resolution performance on LS3D-W balanced dataset across pose (higher is better). The results are not indicative of visual quality. See Fig. 4.

Super-FAN: Integrated Facial Landmark Localization and Super-Resolution of Real-World Low Resolution Faces in Arbitrary Poses With GANs

Bulat, A., Tzimiropoulos, G., Super-FAN: Integrated Facial Landmark Localization and Super-Resolution of Real-World Low Resolution Faces in Arbitrary Poses With GANs, Proc. of CVPR 2018, PP. 109-117, 2018.

Method	[0-30]	[30-60]	[60-90]
FAN-bilinear	10.7%	6.9%	2.3%
FAN-SR-ResNet	48.9%	38.9%	21.4%
FAN-SR-GAN	47.1%	36.5%	19.6%
Retrained FAN-bilinear	55.9%	49.2%	37.8%
FAN-Ours-pixel	52.3%	45.3%	28.3%
FAN-Ours-pixel-feature	57.0%	50.2%	34.9%
FAN-Ours-pixel-feature-heatmap	61.0%	55.6%	42.3%
Super-FAN	**67.0%**	**63.0%**	**52.5%**
FAN-HR images	75.3%	72.7%	68.2%

Table 2: AUC across pose (calculated for a threshold of 10%; see [4]) on our LS3D-W balanced test set. The results, in this case, are indicative of visual quality. See Fig. 4.

Supervision-by-Registration: An Unsupervised Approach to Improve the Precision of Facial Landmark Detectors

Dong, X., Yu, X.-I., Weng, X., Wei, S.-E., Yang, Y., and Sheikh, Y., Supervision-by-Registration: An Unsupervised Approach to Improve the Precision of Facial Landmark Detectors, Proc. of CVPR 2018, PP. 360-368, 2018.

Supervision-by-Registration: An Unsupervised Approach to Improve the Precision of Facial Landmark Detectors

Dong, X., Yu, X.-I., Weng, X., Wei, S.-E., Yang, Y., and Sheikh, Y., Supervision-by-Registration: An Unsupervised Approach to Improve the Precision of Facial Landmark Detectors, Proc. of CVPR 2018, PP. 360-368, 2018.

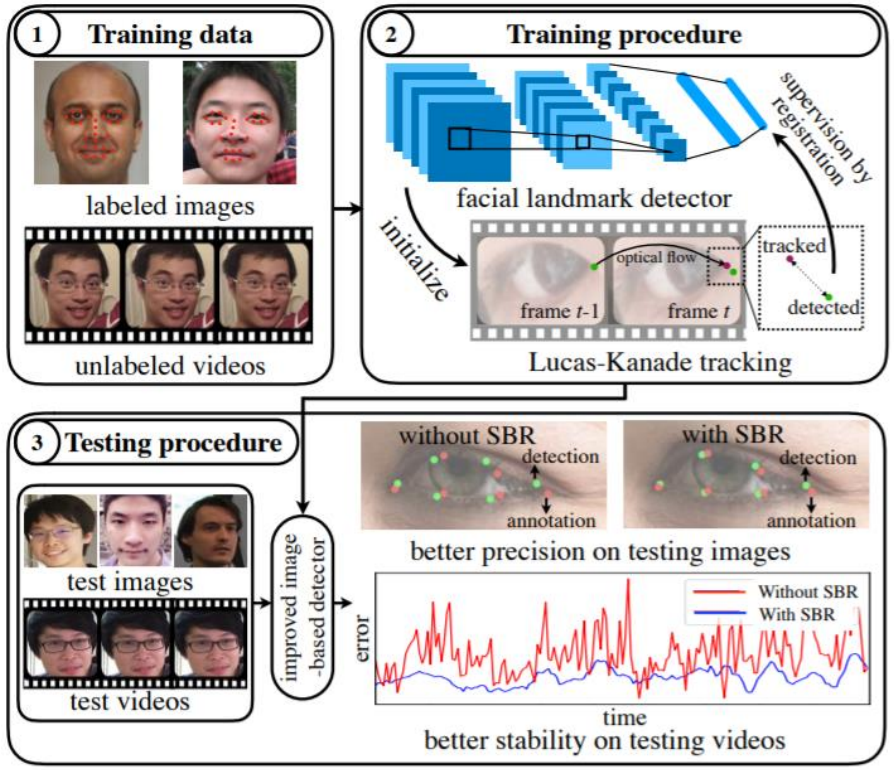


Figure 2. The **supervision-by-registration (SBR) framework** takes labeled images and unlabeled video as input to train an image-based facial landmark detector which is more precise on images/video and also more stable on video.

Supervision-by-Registration: An Unsupervised Approach to Improve the Precision of Facial Landmark Detectors

Dong, X., Yu, X.-I., Weng, X., Wei, S.-E., Yang, Y., and Sheikh, Y., Supervision-by-Registration: An Unsupervised Approach to Improve the Precision of Facial Landmark Detectors, Proc. of CVPR 2018, PP. 360-368, 2018.

Method	300-W			AFLW
	Common	Challenging	Full Set	
SDM [38]	5.57	15.40	7.52	5.43
LBF [25]	4.95	11.98	6.32	4.25
MDM [34]	4.83	10.14	5.88	-
TCDCN [39]	4.80	8.60	5.54	-
CFSS [40]	4.73	9.98	5.76	3.92
Two-Stage [19]	4.36	**7.56**	4.99	2.17
Reg	8.14	16.90	9.85	5.01
Reg + SBR	7.93	15.98	9.46	4.77
CPM	3.39	8.14	4.36	2.33
CPM + SBR	**3.28**	7.58	**4.10**	**2.14**

Table 1. Comparison of NME on 300-W and AFLW datasets.

Method	DGCM [13]	CPM	CPM+SBR	CPM+SBR+PAM
AUC@0.08	59.38	57.25	58.22	**59.39**

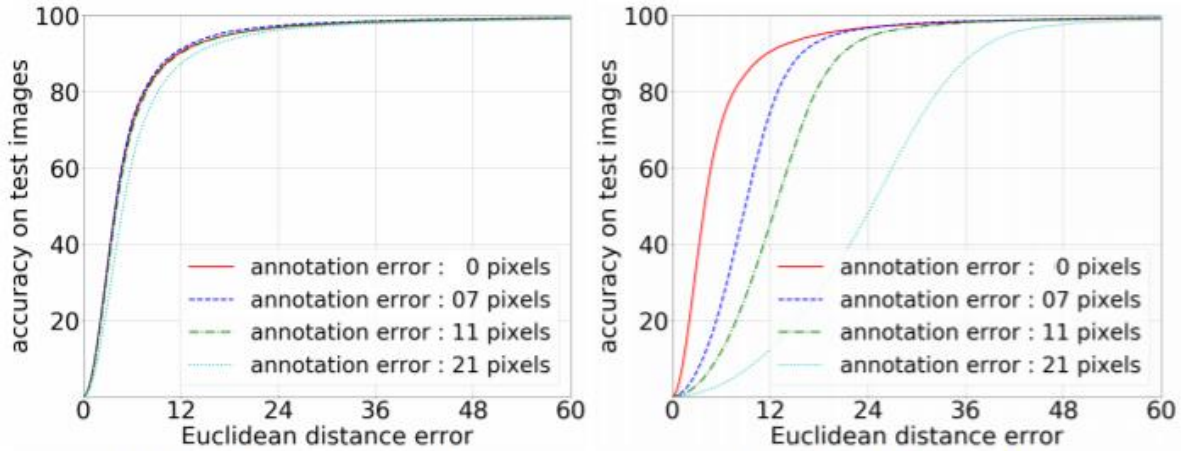
Table 2. AUC @ 0.08 error on 300-VW category C. Note that SBR and PAM do not utilize any additional annotations, but can still improve the baseline CPM and achieve the state-of-the-art results.

Method	SDM [38]	ESR [3]	RLB [25]	PIEFA [23]
NME	5.85	5.61	5.37	4.92
Ours	Reg	Reg+PAM	CPM	CPM+PAM
NME	10.21	9.31	5.26	**4.74**

Table 3. Comparisons of NME on YouTube Celebrities dataset.

Supervision-by-Registration: An Unsupervised Approach to Improve the Precision of Facial Landmark Detectors

Dong, X., Yu, X.-I., Weng, X., Wei, S.-E., Yang, Y., and Sheikh, Y., Supervision-by-Registration: An Unsupervised Approach to Improve the Precision of Facial Landmark Detectors, Proc. of CVPR 2018, PP. 360-368, 2018.



(a) evaluation results on clean testing data

(b) evaluation results on noisy testing data

Figure 8. **Effect of annotation error.** We add Gaussian noise to the annotations of SyntheticFace, and train the model on these noisy data. Different levels of Gaussian noise is indicated by different colors. **Left:** The models are evaluated on clean testing data of SyntheticFace. **Right:** The models are evaluated on noisy testing data of SyntheticFace, which has the same noise distribution as training.

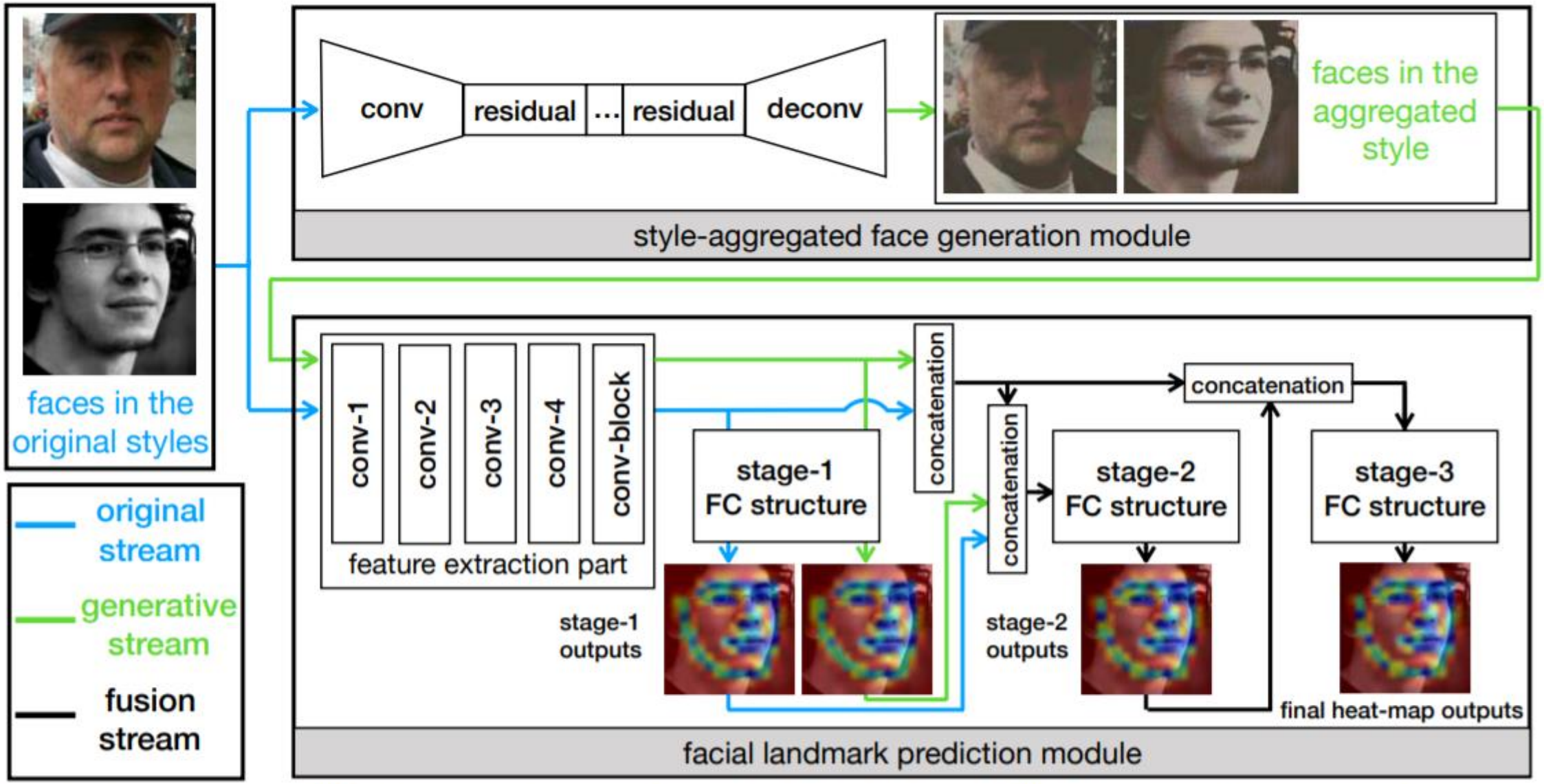


Figure 3. Overview of the SAN architecture. Our network consists of two components. The first is the style-aggregated face generation module, which transforms the input image into different styles and then combines them into a style-aggregated face. The second is the facial landmark prediction module. This module takes both the original image and the style-aggregated one as input to obtain two complementary features and then fuses the two features to generate heat-map predictions in a cascaded manner. "FC" means fully-convolution.

Style Aggregated Network for Facial Landmark Detection

Dong, X., Yan, Y., Ouyang, W., and Yang, Y., Style Aggregated Network for Facial Landmark Detection, Proc. of CVPR 2018, PP. 379-388, 2018.

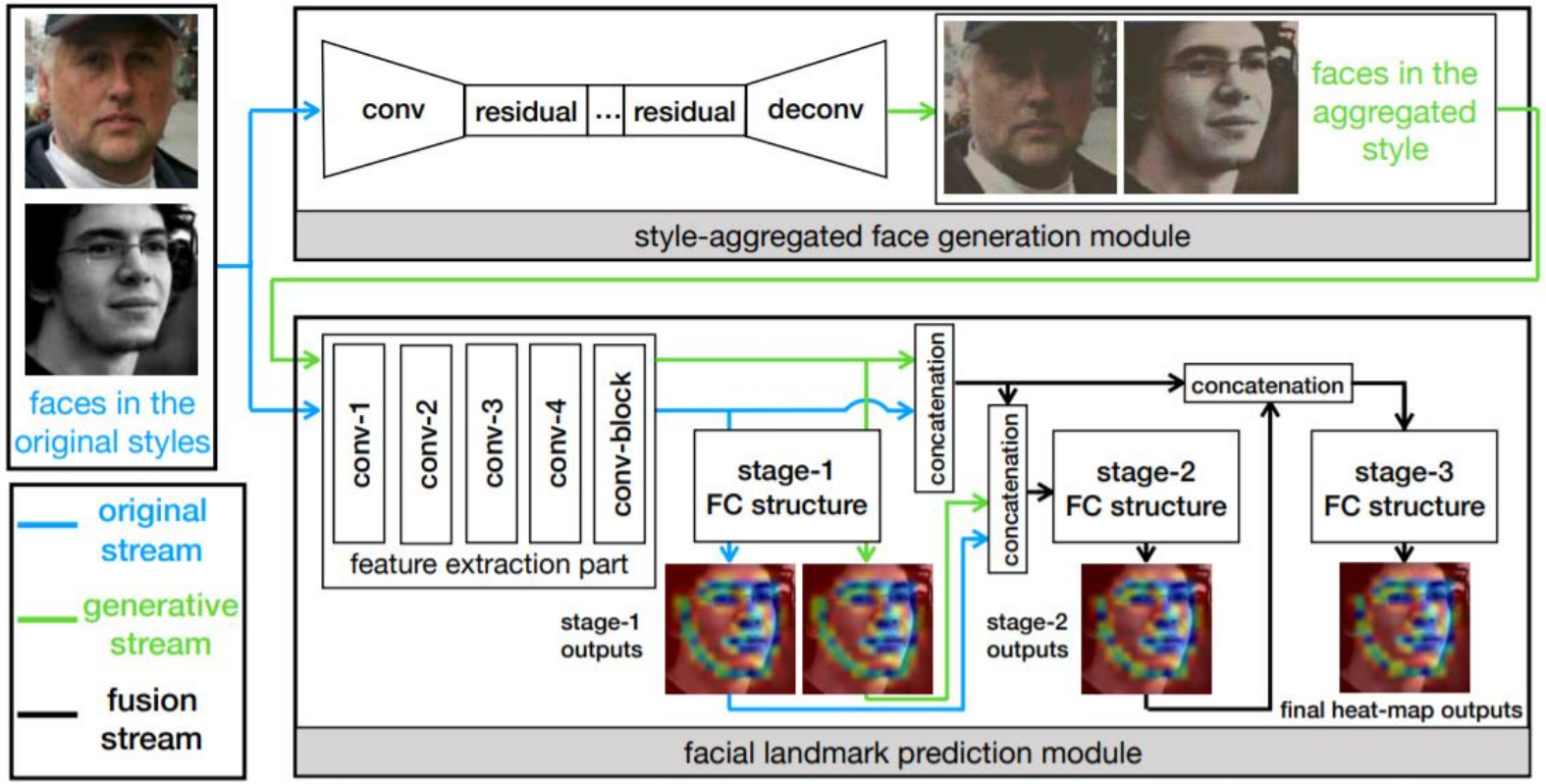


Figure 3. Overview of the SAN architecture. Our network consists of two components. The first is the style-aggregated face generation module, which transforms the input image into different styles and then combines them into a style-aggregated face. The second is the facial landmark prediction module. This module takes both the original image and the style-aggregated one as input to obtain two complementary features and then fuses the two features to generate heat-map predictions in a cascaded manner. "FC" means fully-convolution.

Style Aggregated Network for Facial Landmark Detection

Dong, X., Yan, Y., Ouyang, W., and Yang, Y., Style Aggregated Network for Facial Landmark Detection, Proc. of CVPR 2018, PP. 379-388, 2018.

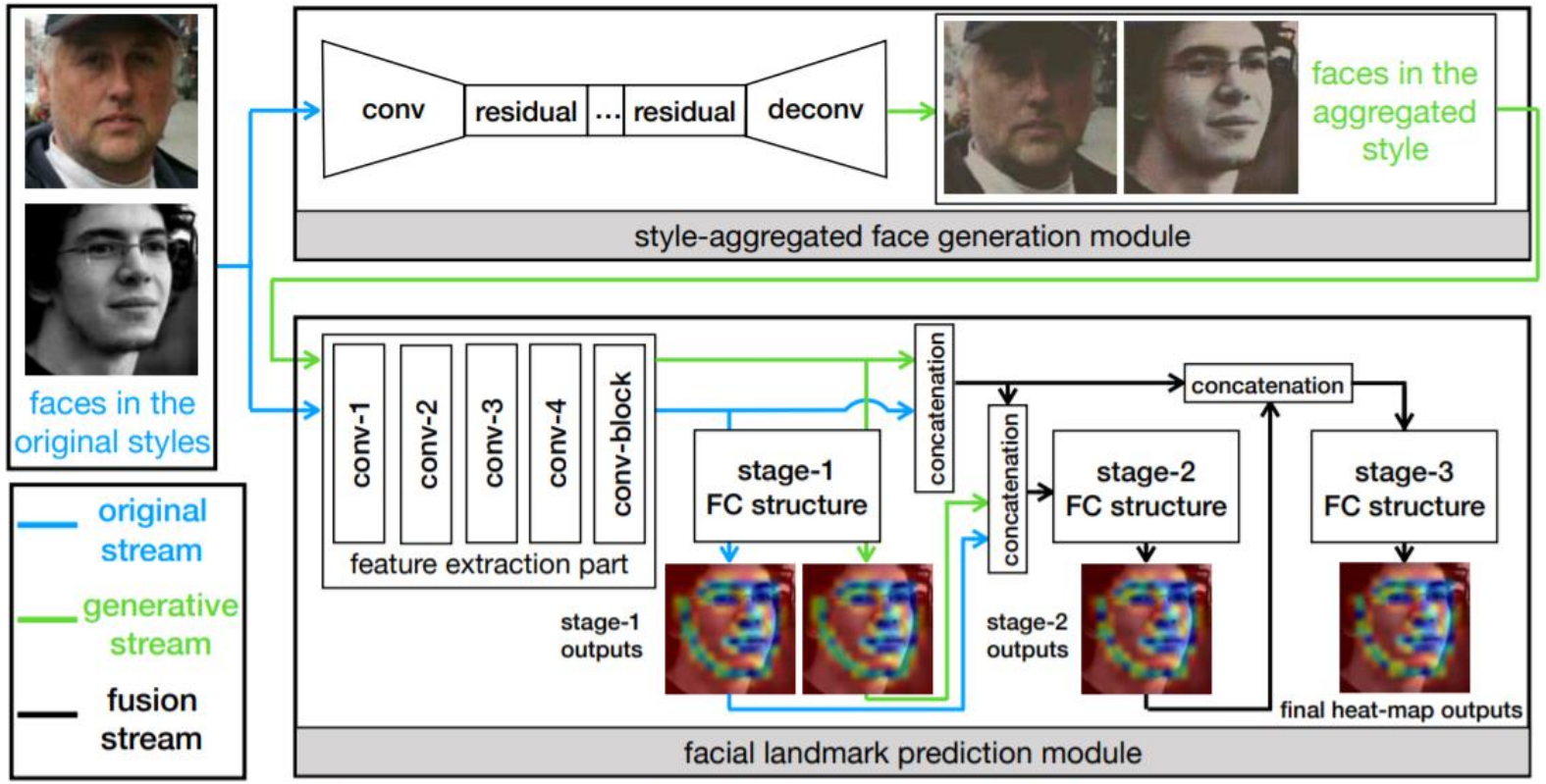


Figure 3. Overview of the SAN architecture. Our network consists of two components. The first is the style-aggregated face generation module, which transforms the input image into different styles and then combines them into a style-aggregated face. The second is the facial landmark prediction module. This module takes both the original image and the style-aggregated one as input to obtain two complementary features and then fuses the two features to generate heat-map predictions in a cascaded manner. "FC" means fully-convolution.

Style Aggregated Network for Facial Landmark Detection

Dong, X., Yan, Y., Ouyang, W., and Yang, Y., Style Aggregated Network for Facial Landmark Detection, Proc. of CVPR 2018, PP. 379-388, 2018.

Method	Common	Challenging	Full Set
SDM [64]	5.57	15.40	7.52
ESR [7]	5.28	17.00	7.58
LBF [43]	4.95	11.98	6.32
CFSS [72]	4.73	9.98	5.76
MDM [55]	4.83	10.14	5.88
TCDCN [68]	4.80	8.60	5.54
Two-Stage$_{OD}$ [31]	4.36	7.56	4.99
Two-Stage$_{GT}$ [31]	4.36	7.42	4.96
RDR [61]	5.03	8.95	5.80
Pose-Invariant[20]	5.43	9.88	6.30
SAN$_{OD}$	3.41	7.55	4.24
SAN$_{GT}$	**3.34**	**6.60**	**3.98**

Table 1. Normalized mean errors (NME) on 300-W dataset.

Robust Facial Landmark Detection via a Fully-Convolutional Local-Global Context Network

Merget, D., Rock, M., and Rigoll, G., Robust Facial Landmark Detection via a Fully-Convolutional Local-Global Context Network, Proc. of CVPR 2018, PP. 781-790, 2018.

Figure 5. Best viewed in the digital version. Qualitative results of our approach on the 300-W benchmark. The images are sorted according to their error (top left is best, bottom right is worst). All but the first two images shown are worse than the average case. More precisely, the mean and median errors of the images in rows 1 through 4 are in the 76.5% and 83.6% quantiles of the test set, respectively. The 5th row displays the 10 worst results. Note that the results are displayed in color, but our network only uses grayscale information.

Robust Facial Landmark Detection via a Fully-Convolutional Local-Global Context Network

Merget, D., Rock, M., and Rigoll, G., Robust Facial Landmark Detection via a Fully-Convolutional Local-Global Context Network, Proc. of CVPR 2018, PP. 781-790, 2018.

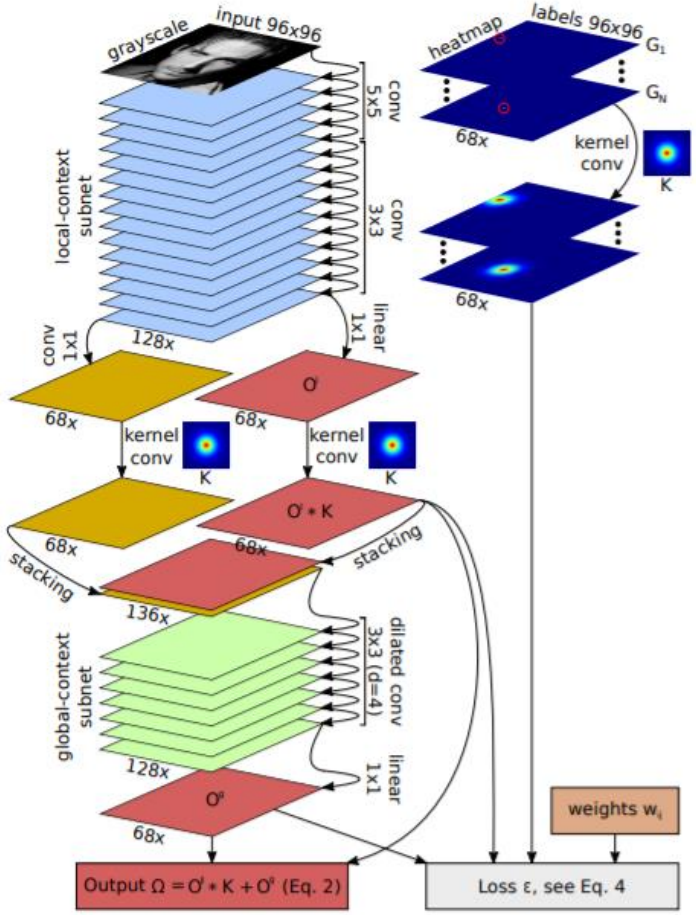


Figure 1. The network architecture used throughout this work.

Robust Facial Landmark Detection via a Fully-Convolutional Local-Global Context Network

Merget, D., Rock, M., and Rigoll, G., Robust Facial Landmark Detection via a Fully-Convolutional Local-Global Context Network, Proc. of CVPR 2018, PP. 781-790, 2018.

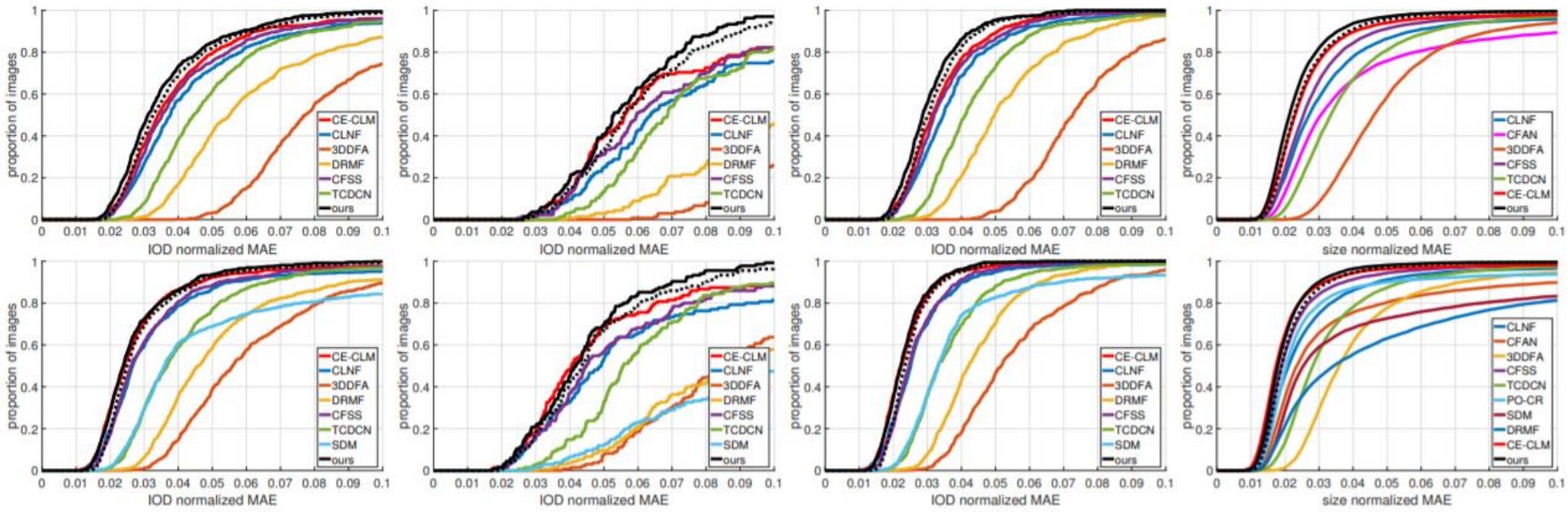


Figure 4. Best viewed in color. Quantitative results of our approach trained on the 300-W training set. Top/bottom row: With/without face outline. From left to right: 300-W [28], iBUG [29], LFPW [5] + HELEN [20], Menpo frontal train set [41]. The dashed line depicts the performance of our network without model fitting (*i.e.*, simple heatmap-wise maximum). Our approach is clearly more robust against outliers than other methods, most prominently on the challenging iBUG test set. Most of the time we are also more accurate.

Robust Facial Landmark Detection via a Fully-Convolutional Local-Global Context Network

Merget, D., Rock, M., and Rigoll, G., Robust Facial Landmark Detection via a Fully-Convolutional Local-Global Context Network, Proc. of CVPR 2018, PP. 781-790, 2018.

Method \ Data	iBUG [29]	LFPW [5] + HELEN [20]	Menpo [41] (frontal)
CLNF [4]	6.37/4.93	3.47/2.51	2.66/2.10
SDM [37]	– /10.73	– /3.31	– /2.54
CFAN [42]	8.38/6.99	– /–	2.87/2.34
DRMF [1]	10.36/8.64	4.97/4.22	– /3.44
CFSS [44]	5.97/4.49	3.20/2.46	2.32/1.90
TCDCN [43]	6.87/5.56	4.11/3.32	3.32/2.81
3DDFA [45]	12.31/8.34	7.27/5.17	4.51/3.59
PO-CR [35]	– /**3.33**	– /2.67	– /2.03
CE-CLM [40]	5.62/4.05	3.13/2.23	2.23/**1.74**
Ours (no model fit)	5.55/4.36	3.04/2.34	2.27/1.90
Ours	**5.29**/4.18	**2.86**/**2.21**	**2.14**/1.79

Table 1. Median IOD-normalized MAE with/without face outline for iBUG [29] and LFPW [5] + HELEN [20]. Median image size-normalized MAE with/without face outline for Menpo [41]. The best performance is highlighted in bold. While our approach does not achieve the best median performance on all datasets, the performance is very consistent.

Wing Loss for Robust Facial Landmark Localisation With Convolutional Neural Networks

Feng, Z.-H., Kittler, J., Awais, M., Huber, P., and Wu, X.-J., Wing Loss for Robust Facial Landmark Localisation With Convolutional Neural Networks, Proc. of CVPR 2018, PP. 2235-2245, 2018.

- 본 연구는 Wing Loss 연구로, CNN을 사용한 robust facial landmark 위치 파악을 위한 wing loss
- Landmark를 탐지하는 연구가 아니기 때문에 생략

AAAI 2018 ~ 2019

Brute-Force Facial Landmark Analysis With A 140,000-Way Classifier

Li, M., Jeni, L., and Ramanan, D. Brute-Force Facial Landmark Analysis With A 140,000-Way Classifier, Proc. of AAAI 2018, 2018.

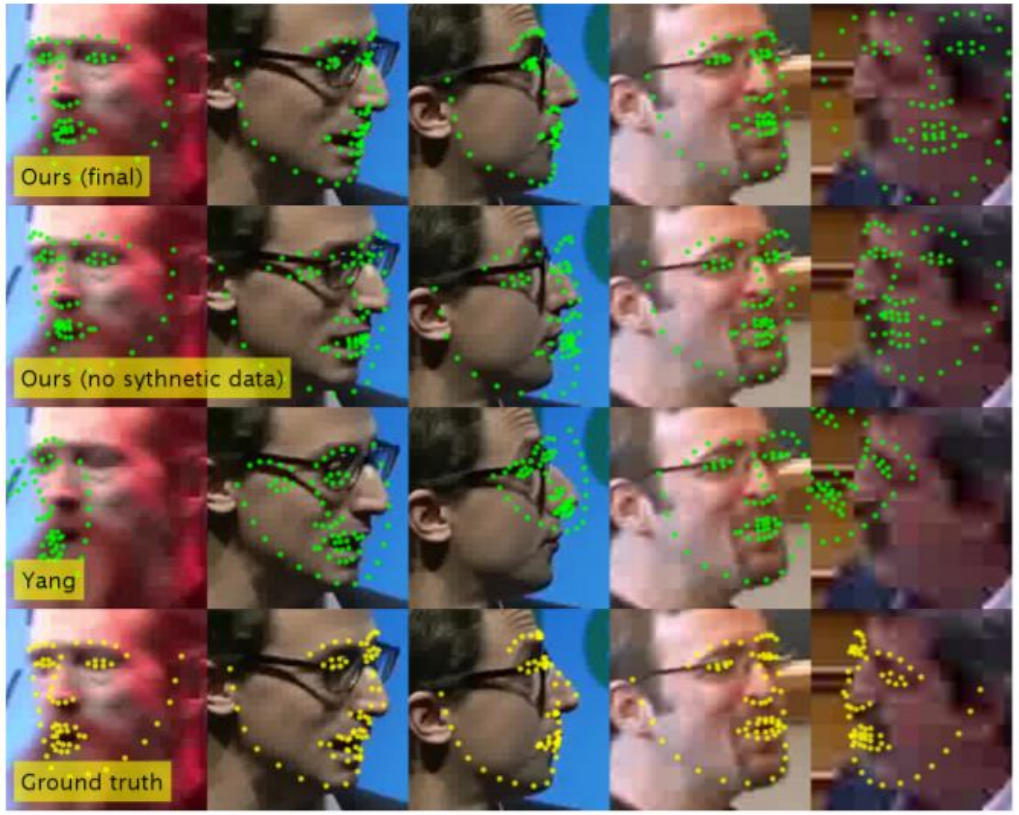


Figure 9: Qualitative results for Category 3-Hard frames on 300VW. We can see that by training on synthetic images (comparing Row 1 with Row 2), our method is robust to large pose variation. The last column shows a failure case.

Brute-Force Facial Landmark Analysis With A 140,000-Way Classifier

Li, M., Jeni, L., and Ramanan, D., Brute-Force Facial Landmark Analysis With A 140,000-Way Classifier, Proc. of AAAI 2018, 2018.

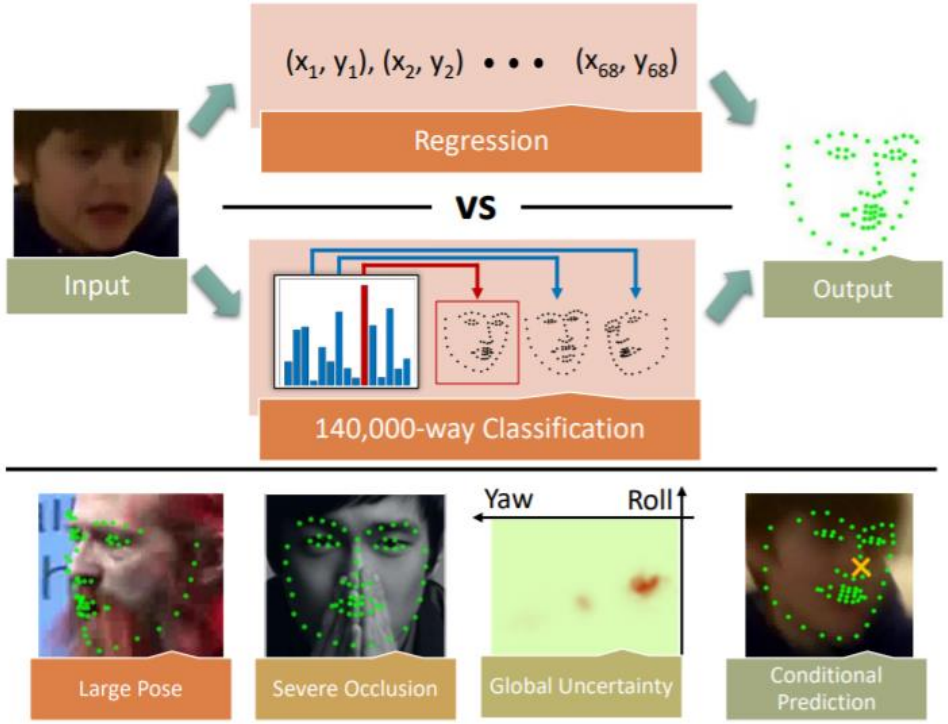
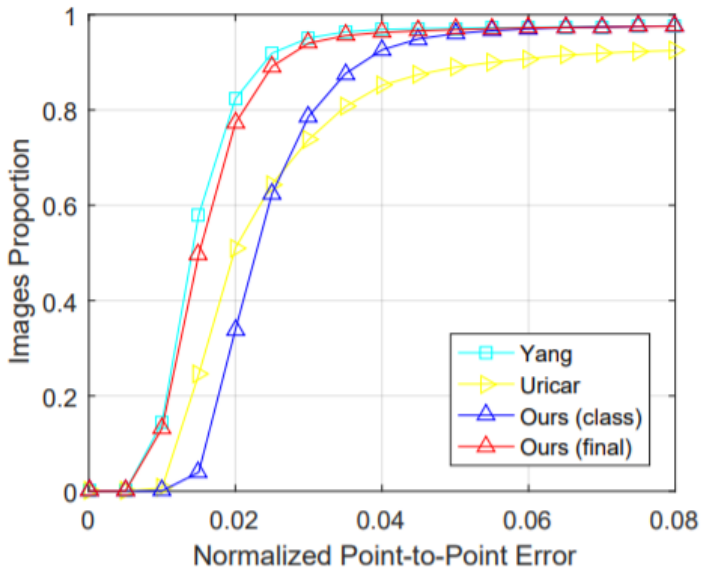


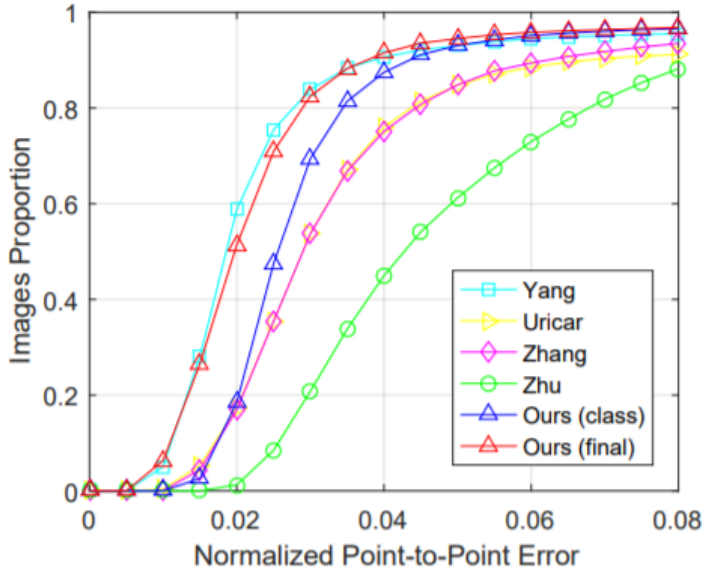
Figure 1: Face alignment is a regression problem, yet we solve it via large-scale classification. As shown in the bottom row, our model is able to handle severe occlusions and large pose variation and provide a global uncertainty estimate. Moreover, such uncertainty representation can used to produce conditional prediction in an interactive setup.

Brute-Force Facial Landmark Analysis With A 140,000-Way Classifier

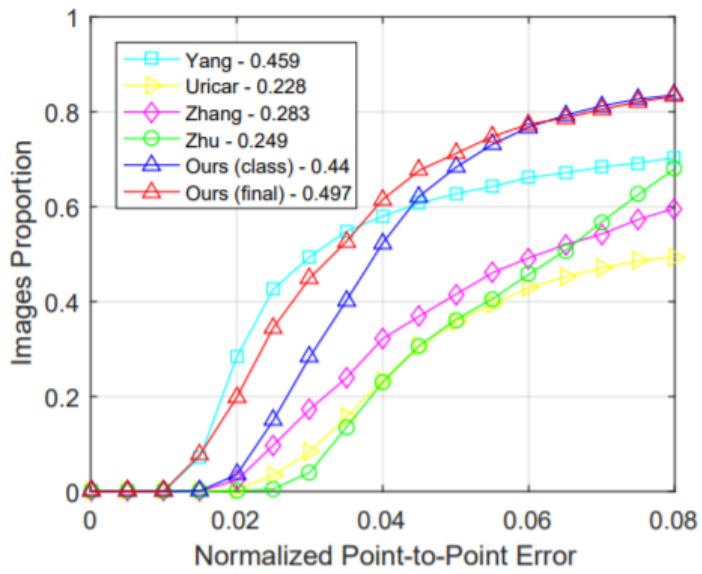
Li, M., Jeni, L., and Ramanan, D., Brute-Force Facial Landmark Analysis With A 140,000-Way Classifier, Proc. of AAAI 2018, 2018.



(a) Category 1 - naturalistic and well-lit.



(b) Category 3 - unconstrained



(c) Subset of cat 3 - hard frames only

Figure 7: The cumulative error distribution curves on the 300VW benchmark. The statistics for (a) and (b) are summarized in Tab 2 . The Area Under Curve (AUC) in (c) is show next to the names.

Brute-Force Facial Landmark Analysis With A 140,000-Way Classifier

Li, M., Jeni, L., and Ramanan, D., Brute-Force Facial Landmark Analysis With A 140,000-Way Classifier, Proc. of AAAI 2018, 2018.

Method	C1 AUC	C1 FR	C3 AUC	C3 FR
(Chrysos et al. 2017)	0.748	6.055	0.726	4.388
(Yang et al. 2015)	0.791	2.400	0.710	4.461
(Uricár, Franc, and Hlavác 2015)	0.657	7.622	0.574	7.957
(Xiao, Yan, and Kassim 2015)	0.760	5.899	0.695	7.379
(Rajamanoharan and Cootes 2015)	0.735	6.557	0.659	8.289
(Wu and Ji 2015)	0.674	13.925	0.602	13.161
(Zhang et al. 2014)	N/A	N/A	0.409	6.487
(Zhu et al. 2016)	N/A	N/A	0.635	11.796
Ours (classification)	0.678	2.398	0.635	3.431
Ours (+ regressor)	0.774	2.221	0.709	3.189
Ours (+ temp. smooth.)	0.777	2.462	0.718	3.298

Table 2: Comparing with existing methods on the 1st and 3rd category of 300VW benchmark. The 1st, 2nd and the 3rd place for each metric are color coded. Here AUC denotes the area under the CED curves in Fig 7 and FR denotes the failure rate in percentage.

Towards Highly Accurate and Stable Face Alignment for High-Resolution Videos

Tai, Y., Liang, Y., Liu, X., Duan, L., Li, J., Wang, C, Huang, F., and Chen, Y., Towards Highly Accurate and Stable Face Alignment for High-Resolution Videos, Proc. of AAAI 2019, 2019.

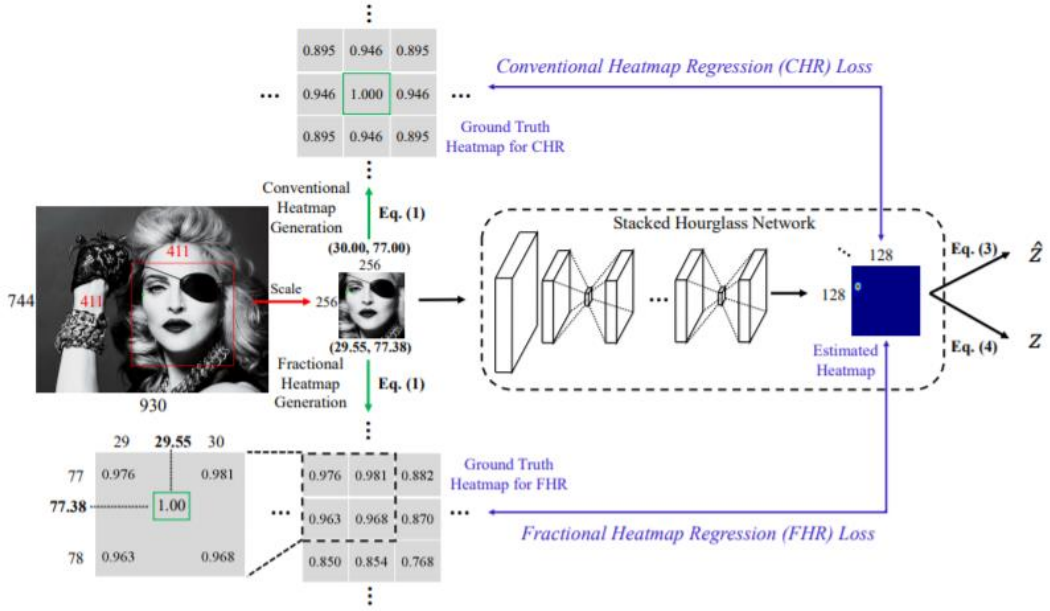


Figure 1: Comparisons between fractional regression heatmap and conventional heatmap regression. Our method differs conventional one in two aspects: 1) the ground truth heatmap for FHR maintains the precision of fractional coordinate, while the conventional one *discards* (e.g., from 29.55 to 30.00, 77.38 to 77.00); and 2) three sampled points on the heatmap analytically computes the fractional peak location of the heatmap (Eq. 4), while the conventional one only finds the maximum activated location (Eq. 3) that loses the fractional part and thus leads to *quantization error*.

Towards Highly Accurate and Stable Face Alignment for High-Resolution Videos

Tai, Y., Liang, Y., Liu, X., Duan, L., Li, J., Wang, C, Huang, F., and Chen, Y., Towards Highly Accurate and Stable Face Alignment for High-Resolution Videos, Proc. of AAAI 2019, 2019.

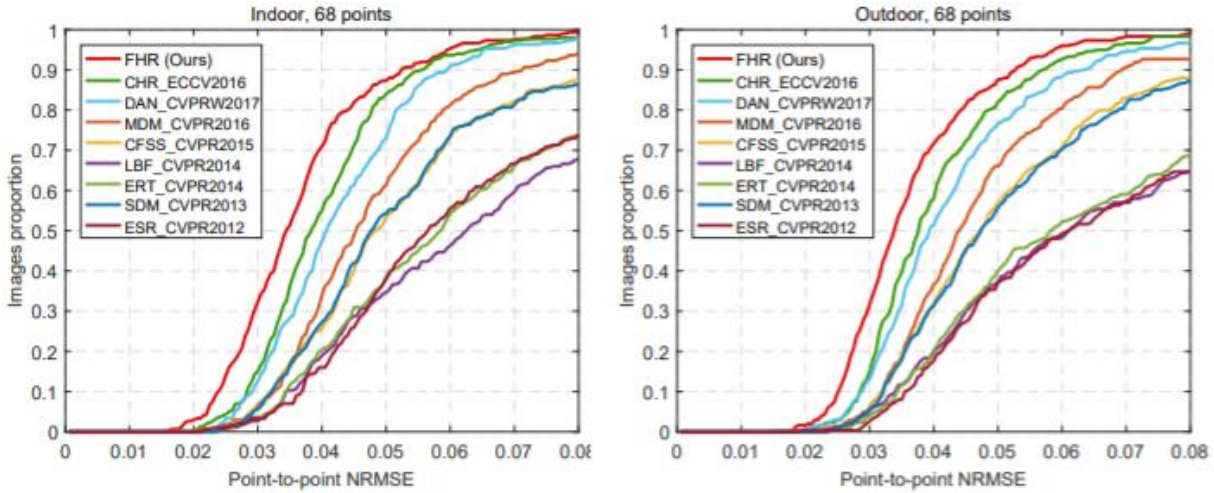


Figure 5: CED curves on 300W.

Table 4: NRMSE/stability comparisons with REDnet on 300-VW test set using 7 landmarks.

Methods	REDnet (2016)	FHR	FHR+STA
Scenario1	8.03/10.3	4.44/5.49	**3.93/3.04**
Scenario2	10.1/9.64	3.96/3.55	**3.82/3.44**
Scenario3	16.5/15.9	5.45/4.72	**4.91/4.45**

Towards Highly Accurate and Stable Face Alignment for High-Resolution Videos

Tai, Y., Liang, Y., Liu, X., Duan, L., Li, J., Wang, C, Huang, F., and Chen, Y., Towards Highly Accurate and Stable Face Alignment for High-Resolution Videos, Proc. of AAAI 2019, 2019.

Table 5: NRMSE comparison with state-of-the-art methods on Talking Face dataset using 7 landmarks.

Methods	CFAN (2014)	CFSS (2015)	IFA (2014)	REDnet (2016)	TSTN (2017)	CHR (2016)	FHR	FHR+STA
NRMSE	3.52	2.36	3.45	3.32	2.13	2.28	**2.06**	2.14

Table 6: NRMSE comparison with state-of-the-art methods on 300-VW test set using 68 landmarks.

Methods	TSCN (2016)	CFSS (2015)	TCDCN (2016)	TSTN (2017)	CHR (2016)	FHR	FHR+STA
Scenario1	12.5	7.68	7.66	5.36	5.44	5.07	**4.42**
Scenario2	7.25	6.42	6.77	4.51	4.71	4.34	**4.18**
Scenario3	13.10	13.70	15.00	12.80	7.92	7.36	**5.98**