ASMNet: a Lightweight Deep Neural Network for Face Alignment and Pose Estimation

Ali Pourramezan Fard, Hojjat Abdollahi, and Mohammad Mahoor

Department of Electrical and Computer Engineering
University of Denver, Denver, CO

{Ali.pourramezanfard, hojjat.abdollahi, mohammad.mahoor}@du.edu

CVPR 2021 + AMFG workshop

Outline

- Introduction
- ASMNet Architecture
- ASM Assisted Loss Function
- Evaluation

Introduction

- ASMNet:

  it is a lightweight Convolutional Neural Network (CNN) which is designed to perform face alignment and pose estimation efficiently while having acceptable accuracy

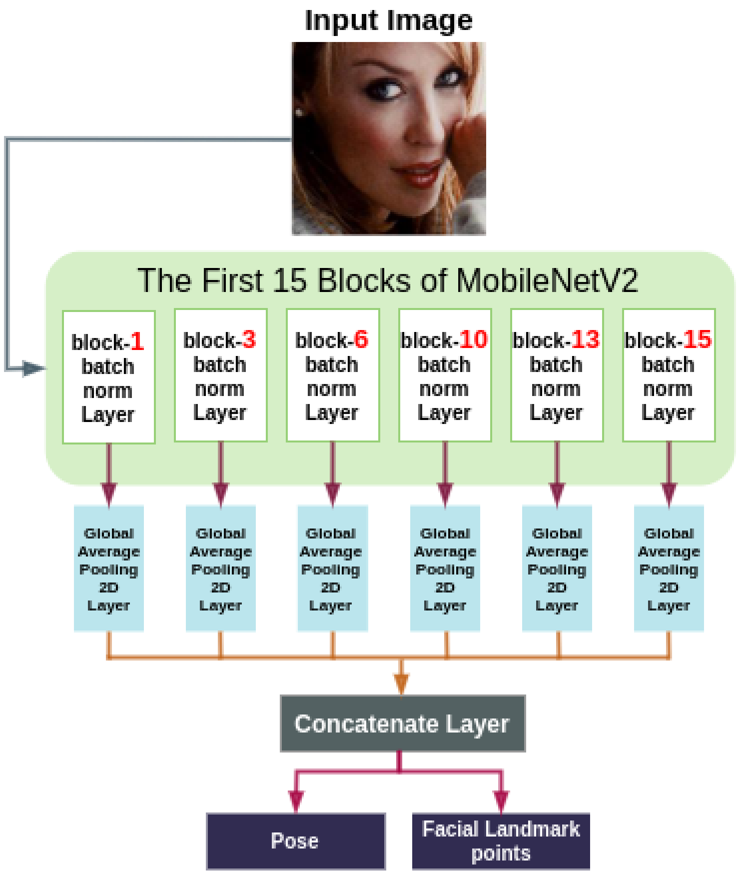
- Contributions:

  - Proposing a CNN inspired by MobileNetV2 while being about 2 times smaller in terms of number of parameters

  - Proposing a loss inspired by ASM to improve the accuracy of ASMNet

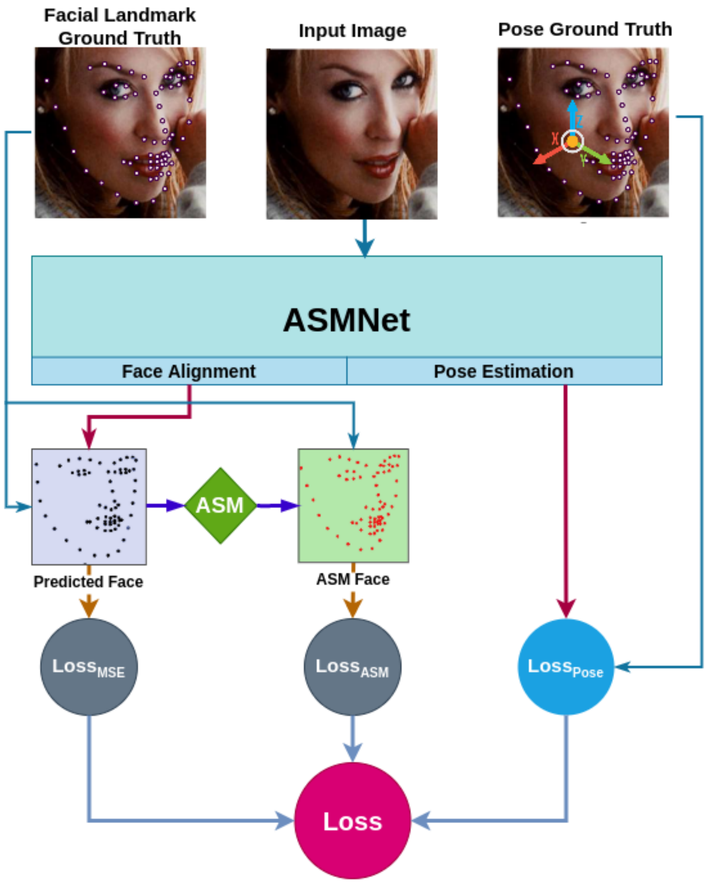
ASMNet Architecture

- Designed inspired by the architecture of MobileNetV2
- GlobalAveragePooling layers used in order to keep features from the very first layers to the last layer
- Only used the first 15 blocks of MobileNetV2 as there is no need for abstract features in the last block
- Designed ASMNet to perform face alignment as well as pose estimation



ASM Assisted Loss Function

- Proposed a new loss function called ASM-LOSS
- ASM-LOSS utilizes ASM to improve the accuracy of the network
- ASM-LOSS guides the network to first learn the smoothed distribution of the facial landmark points
- Then, ASM-LOSS leads the network to learn the original landmark points
- Estimate face pose with the assistant of smoothed facial landmark points



ASM Assisted Loss Function (cont.1)

1

$$G_{set} = \{(G_x^1, G_y^1), \ldots, (G_x^n, G_y^n)\}$$

$$P_{set} = \{(P_x^1, P_y^1), \ldots, (P_x^n, P_y^n)\}$$

2

$$A_{set} = \{(A_x^1, A_y^1), \ldots, (A_x^n, A_y^n)\}$$

$$\mathcal{ASM} : (G_x^i, G_y^i) \mapsto (A_x^i, A_y^i)$$

3

$$\mathcal{L}_{mse} = \frac{1}{N}\frac{1}{n}\sum_{j=1}^{N}\sum_{i=1}^{n}\|G_j^i - P_j^i\|_2$$

4

$$\mathcal{L}_{asm} = \frac{1}{N}\frac{1}{n}\sum_{j=1}^{N}\sum_{i=1}^{n}\|A_j^i - P_j^i\|_2$$

5

$$\mathcal{L}_{facial} = \mathcal{L}_{mse} + \alpha \times \mathcal{L}_{asm}$$

$$\alpha = \begin{cases} 2 & i < \frac{l}{3} \\ 1 & \frac{l}{3} < i < \frac{2l}{3} \\ 0.5 & i > \frac{2l}{3} \end{cases}$$

i : epoch number

l : Number of total epochs

ASM Assisted Loss Function (cont.2)

1

$$\mathcal{L}_{pose} = \frac{1}{N}\sum_{j=1}^{N}\frac{(y_j^p - y_j^t)^2 + (p_j^p - p_j^t)^2 + (r_j^p - r_j^t)^2}{3}$$

2

$$\mathcal{L} = \sum_{i=1}^{2}\lambda_{task_i}\mathcal{L}_{task_i}$$

T = { $\mathcal{L}_{facial}, \mathcal{L}_{pose}$ }

$$\lambda_{task} = \{1, 0.5\}$$

Comparison of Number of Parameters (in Million) and Flops (in Billion)

Method	NME		Params (M)	FLOPs (B)
	300W	WFLW		
mnv2	4.70	9.57	2.42	0.60
mnv2_r	4.59	9.41		
ASMNet_nr	6.49	11.96	1.43	0.51
ASMNet	5.50	10.77		

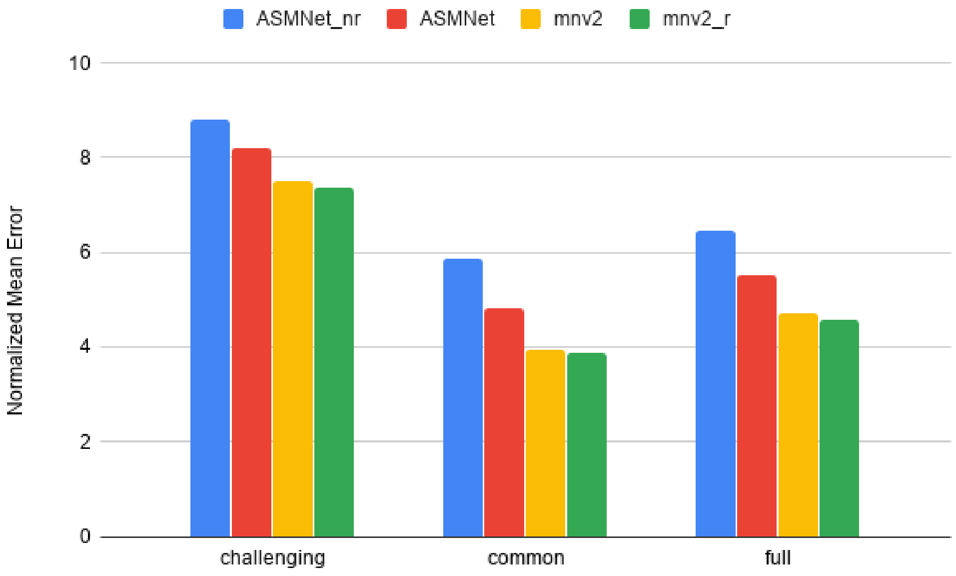
Method	Backbone	#Params (M)	FLOPs (B)
DVLN [45]	VGG-16	132.0	14.4
SAN [12]	ResNet-152	57.4	10.7
LAB [44]	Hourglass	25.1	19.1
ResNet50 (Wing + PDB) [15]	ResNet-50	25	3.8
ASMNet	MobileNetV2 [33]	1.4	0.5
MobileNetV2 [33]	-	2.4	0.6

Evaluation (cont.1)

Face Alignment Accuracy on 300W:

Table 2: Normalized Mean Error (in %) of 68-point landmarks localization on 300W [31] dataset.

Method	Normalized Mean Error		
	Common	Challenging	Fullset
RCN [16]	4.67	8.44	5.41
DAN [21]	3.19	5.24	3.59
PCD-CNN [22]	3.67	7.62	4.44
CPM [13]	3.39	8.14	4.36
DSRN [26]	4.12	9.68	5.21
SAN [12]	3.34	6.60	3.98
LAB [44]	2.98	5.19	3.49
DCFE [40]	2.76	5.22	3.24
mnv2	3.93	7.52	4.70
mnv2_r	3.88	7.35	4.59
ASMNet_nr	5.86	8.80	6.46
ASMNet	4.82	8.2	5.50

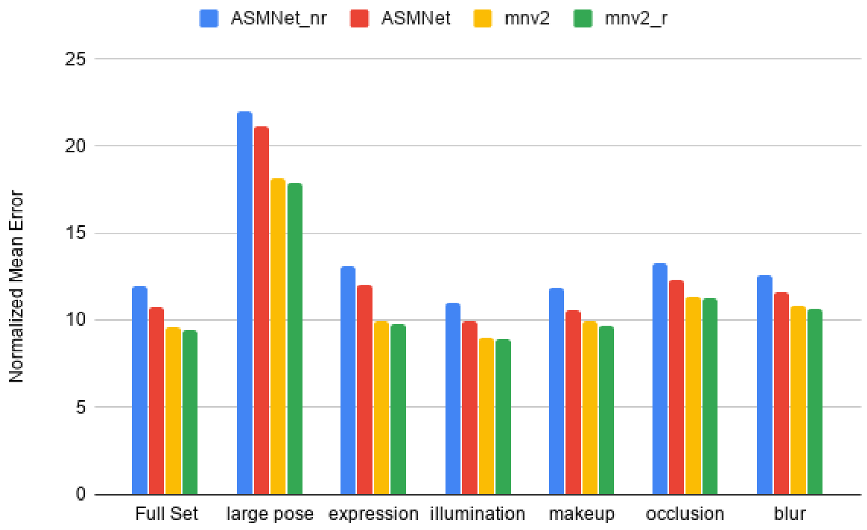


Evaluation (cont.2)

Face Alignment Accuracy on WFLW:

Table 3: Normalized Mean Error (in %), failure rate (in %), and AUC of 98-point landmarks localization on WFLW [44] dataset.

Metric	Method	Test set	Pose	Expression	Illumination	Make-Up	Occlusion	Blur
Mean Error (%)	ESR [5]	11.13	25.88	11.47	10.49	11.05	13.75	12.20
	SDM [47]	10.29	24.10	11.45	9.32	9.38	13.03	11.28
	CFSS [58]	9.07	21.36	10.09	8.30	8.74	11.76	9.96
	DVLN [45]	6.08	11.54	6.78	5.73	5.98	7.33	6.88
	LAB [44]	5.27	10.24	5.51	5.23	5.15	6.79	6.32
	ResNet50(Wing+PDB) [15]	5.11	8.75	5.36	4.93	5.41	6.37	5.81
	mnv2	9.57	18.18	9.93	8.98	9.92	11.38	10.79
	mnv2_r	9.41	17.86	9.78	8.90	9.67	11.25	10.66
	ASMNet_nr	11.96	21.95	13.08	11.02	11.84	13.24	12.60
	ASMNet	10.77	21.11	12.02	9.93	10.55	12.34	11.62



Pose Estimation Accuracy:

Table 4: Mean Absolute Error of pose estimation on 300W [31], WFLW [44] datasets compared to HopeNet[30].

Method		ASMNet_nr	ASMNet	mnv2	mnv2_r
300W [31]	yaw	2.41	1.62	1.75	1.71
	pitch	1.87	1.80	1.93	1.89
	roll	2.115	1.24	1.32	1.30
WFLW [44]	yaw	3.14	2.97	3.06	3.08
	pitch	2.99	2.93	3.03	2.94
	roll	2.23	2.21	2.26	2.22

Table 5: Mean Absolute Error of pose estimation on using ASMNet, JFA [48], and Yang*et. al* [50] on 300W [31].

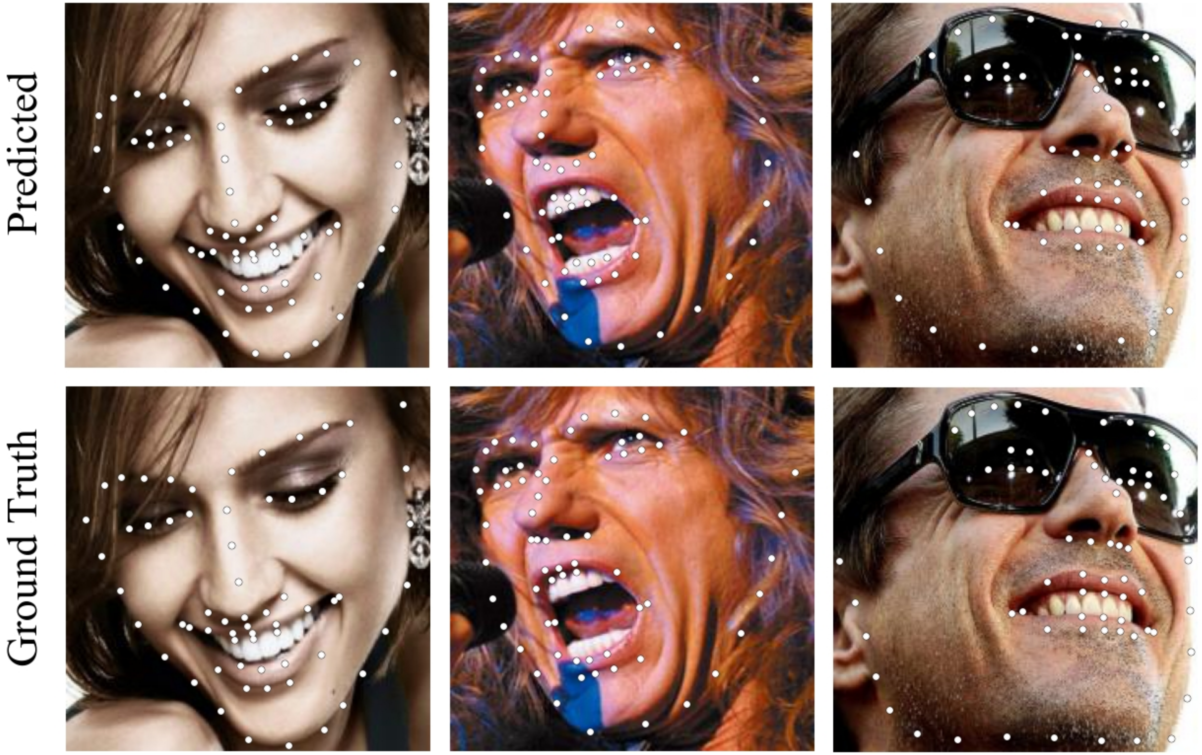
Method	Pitch	Yaw	Roll
Yang*et. al* [50]	5.1	4.2	2.4
JFA [48]	3.0	2.5	2.6
ASMNet	1.80	1.62	1.24

Evaluation (cont.4)

Evaluation of Visual Accuracy:

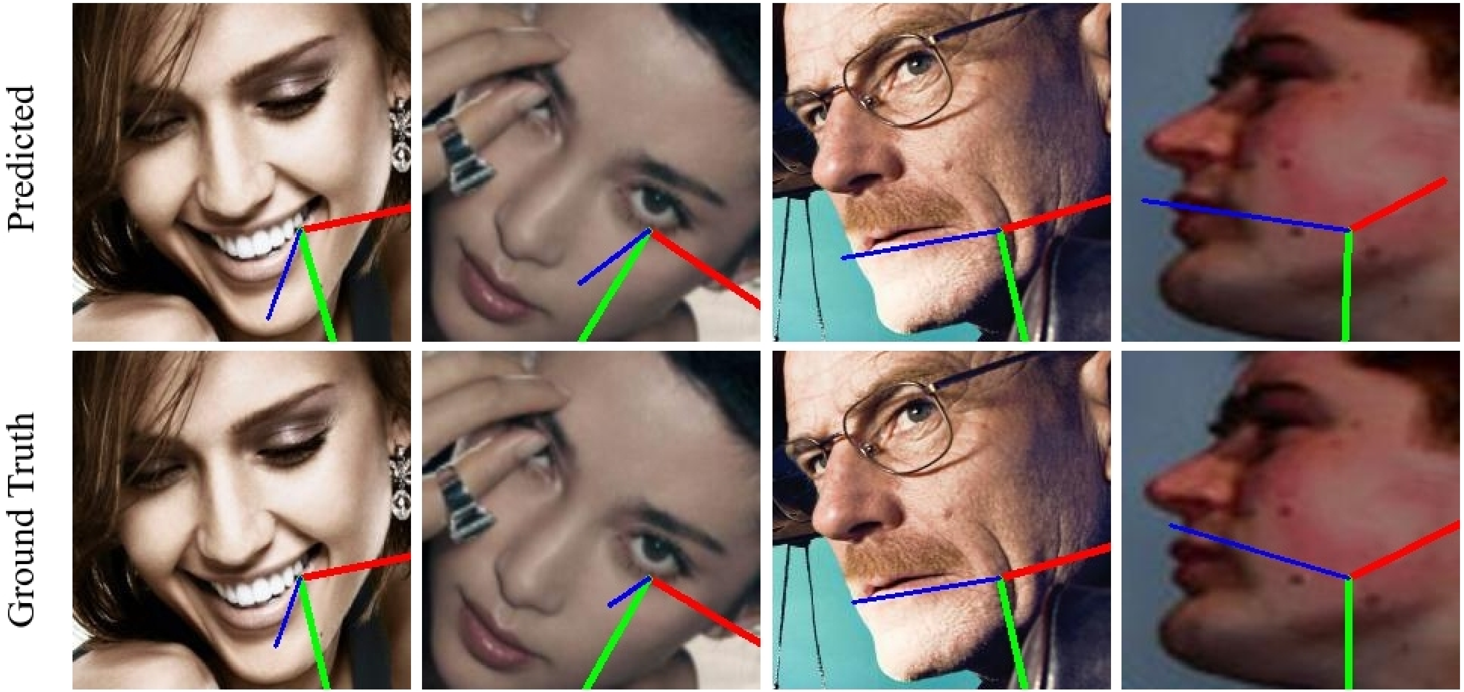
Evaluation (cont.5)

Evaluation of Visual Accuracy:



Evaluation (cont.6)

Evaluation of Visual Accuracy:



Thank You!