

Multi-View 3D Face Reconstruction in the Wild Using Siamese Networks - Submission to 3DFAW Challenge



3DFAW Challenge

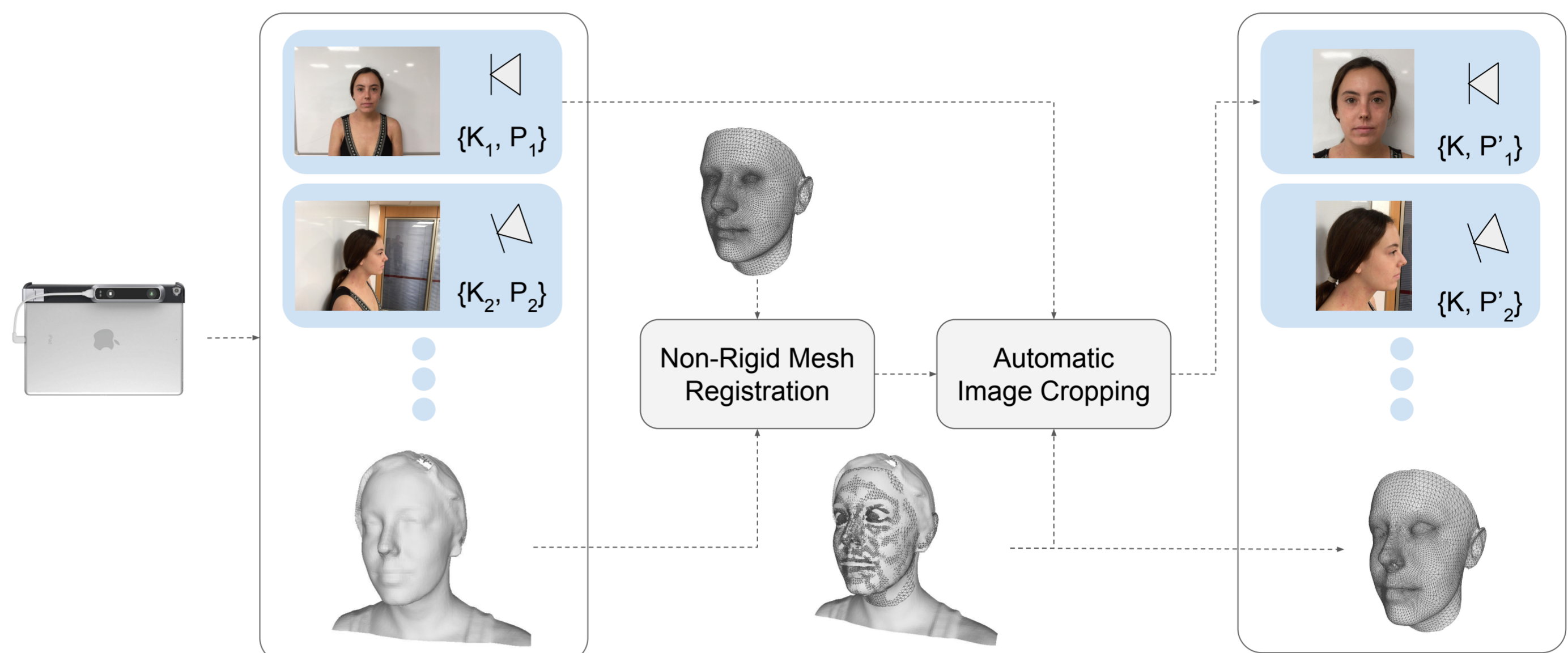
Given a side-to-side video capture of the face:



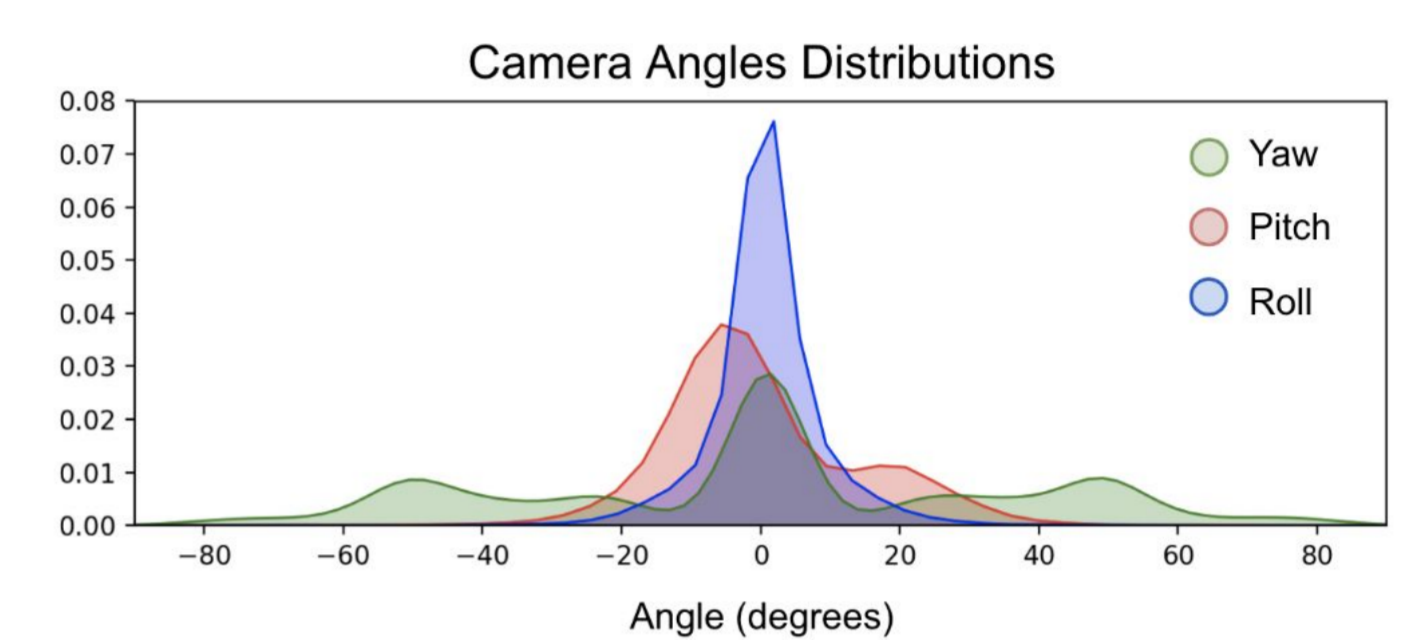
Video frames

Which is the 3D facial geometry of the input subject?

Building the Training Dataset

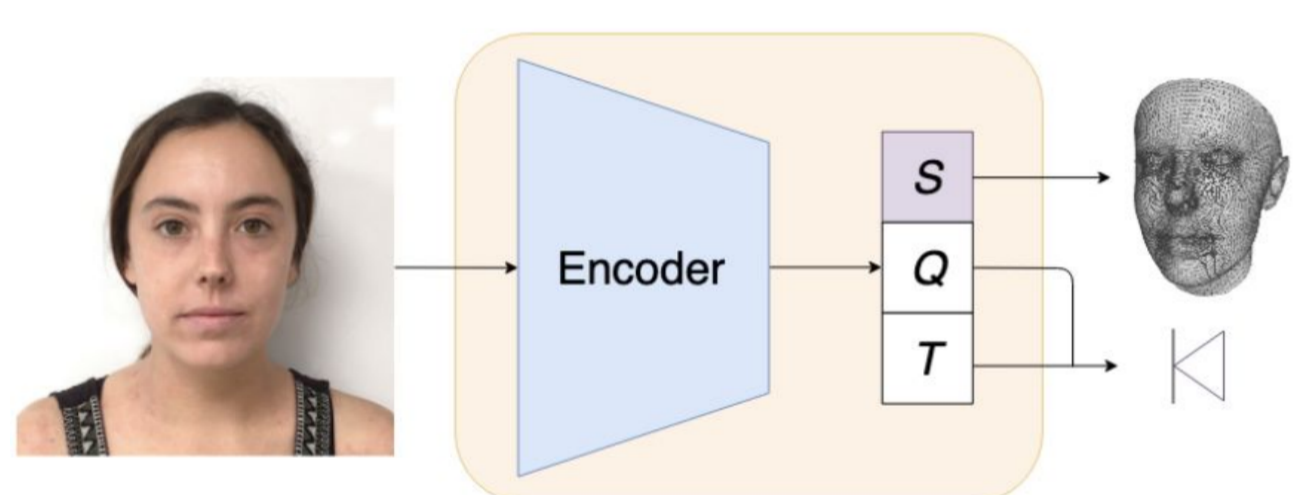


Split	# subjects	# images	Average views/subject
Train	4543	20349	4.4
Validation	675	2976	4.4
Test	1310	6347	4.8



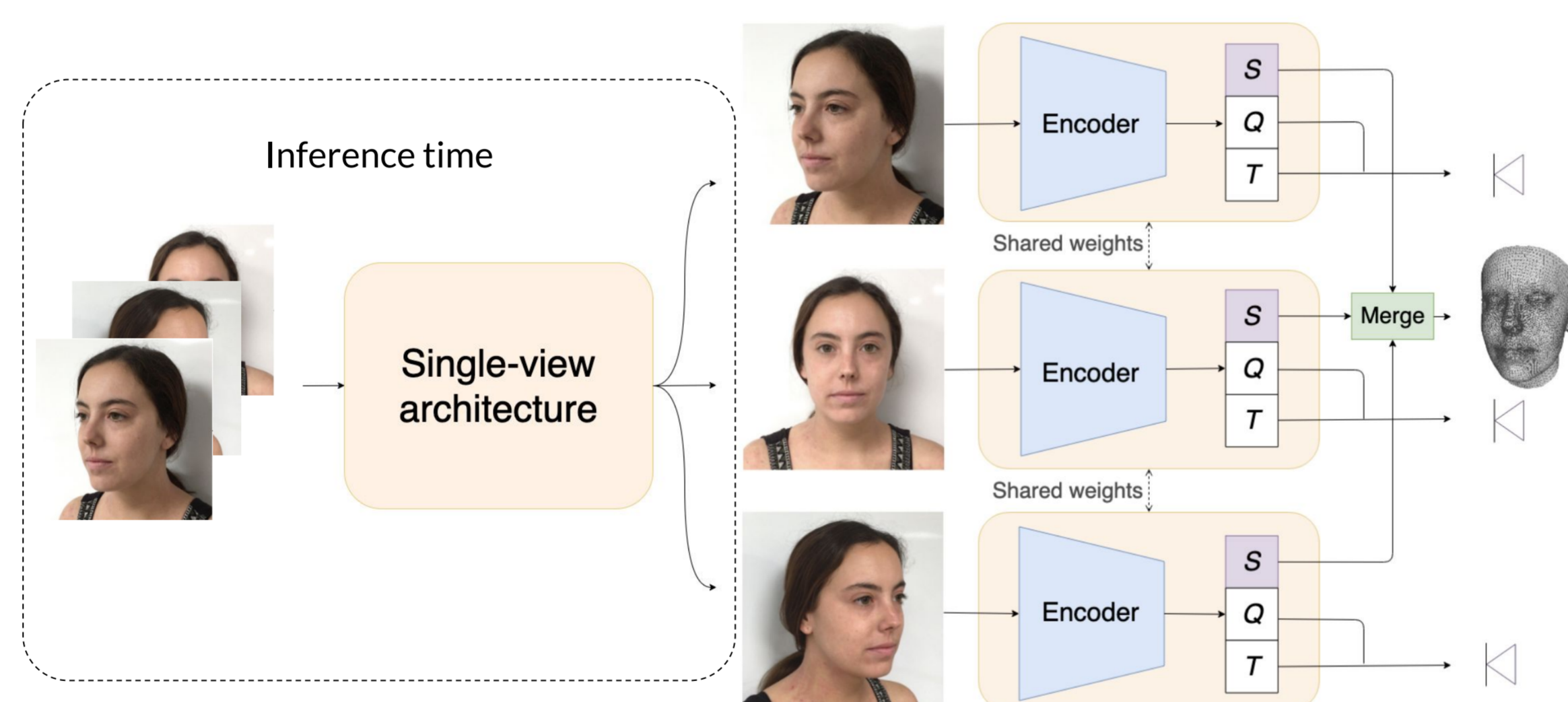
Architectures

Learning monocular 3D reconstruction (SV)



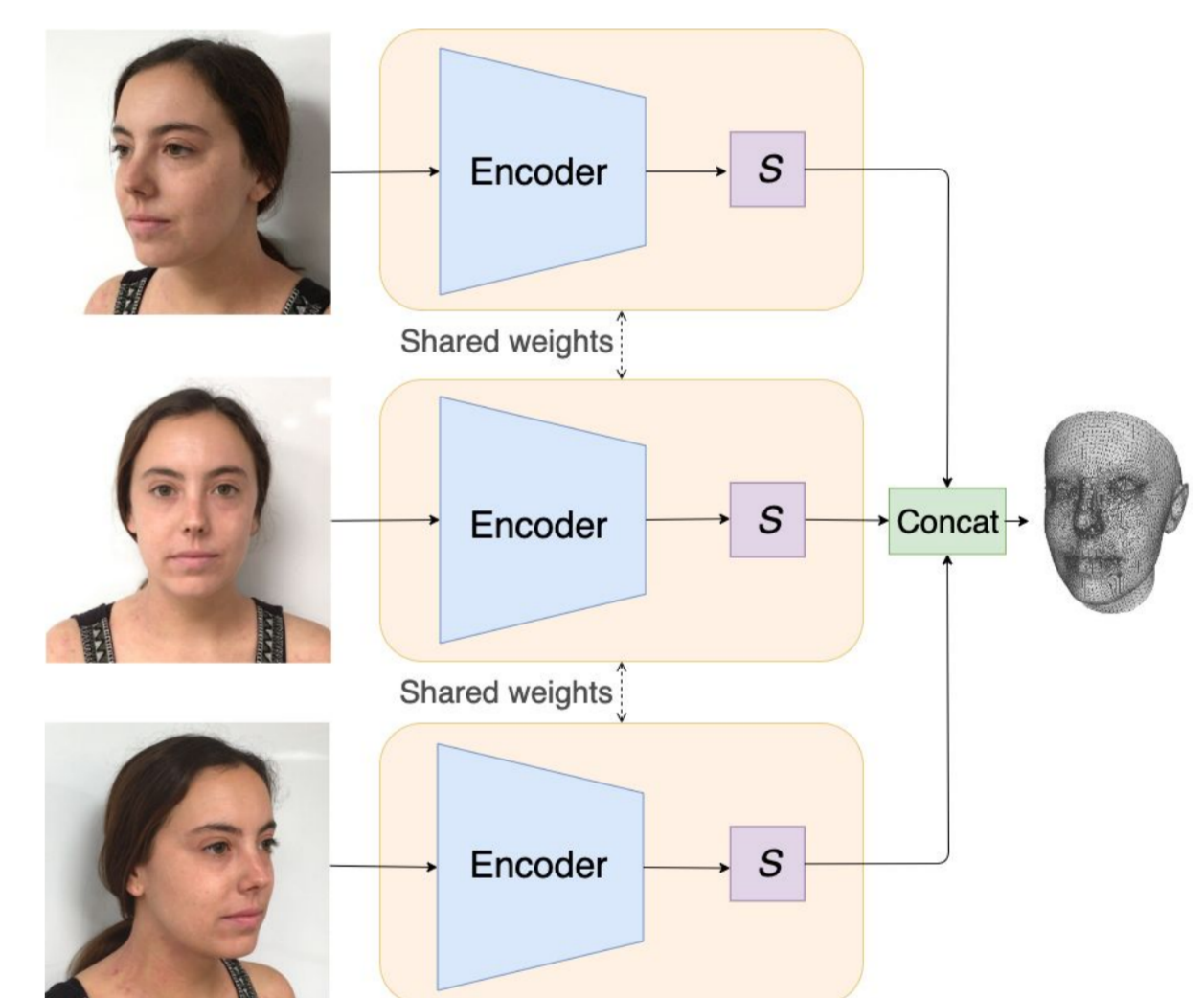
$$\mathcal{L}_{MRL} = \sum_{v=1}^V \|\mathcal{P}(q_v, t_v)(x_H) - \mathcal{P}(q_v, t_v)(D(\hat{x}_H))\|_1$$

Learning multiview 3D reconstruction (MV Add/Concat)



$$\mathcal{L} = \sum_{v=1}^V \|\mathcal{P}(q_v, t_v)(s_H) - \mathcal{P}(q_v, t_v)(\hat{s}_H)\|_2^2$$

Optimizing MV Concat for 3D shape regression (MV Shape-Concat)



$$\mathcal{L} = \sum_{n=1}^N \|s_n - \hat{s}_n\|_2^2$$

Results on 3DFAW Challenge

Quantitative results

Model	ARMSE (mm)
3DMM Mean	3.02
SV Frontal	2.62
SV Mean	2.51
MV Add	2.43
MV Concat	2.33
MV Shape-Concat	2.23
MV Shape-Concat Mean	2.14

Qualitative results

