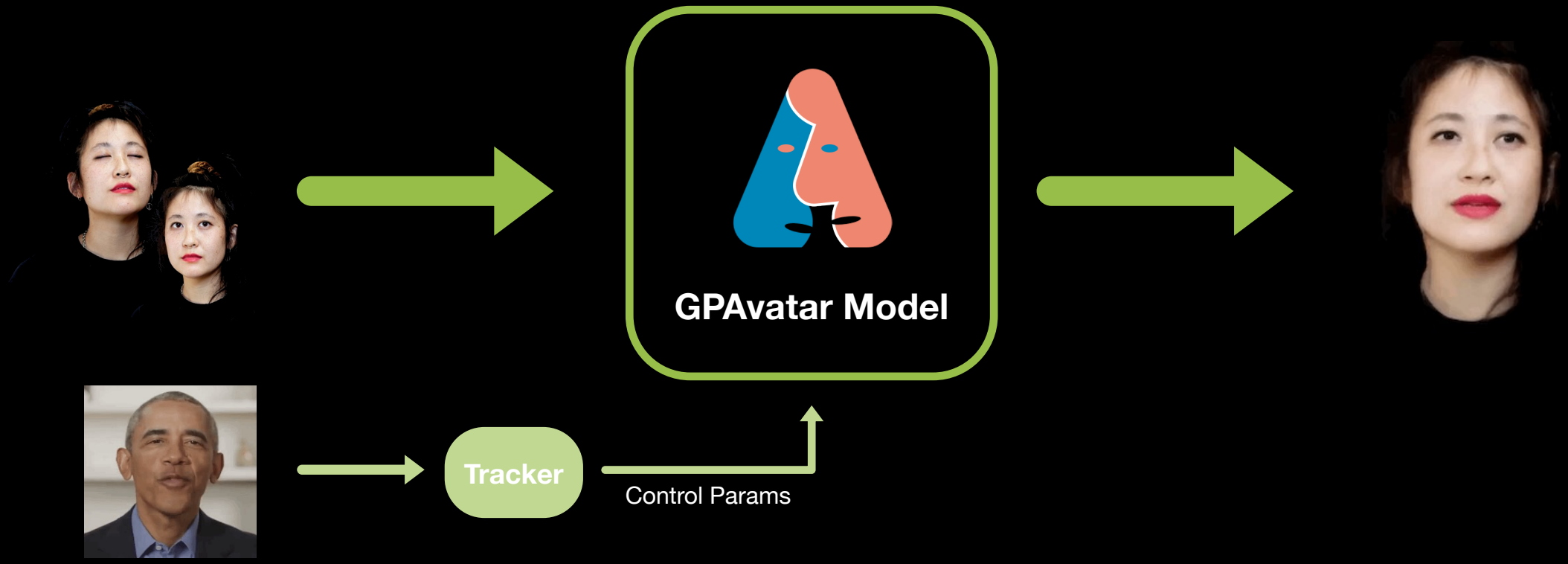
GPAvatar: Generalizable and Precise Head Avatar from Image(s)

Xuangeng Chu[1,2], Yu Li[2], Ailing Zeng[2], Tianyu Yang[2], Lijian Lin[2], Yunfei Liu[2], Tatsuya Harada[1,3]

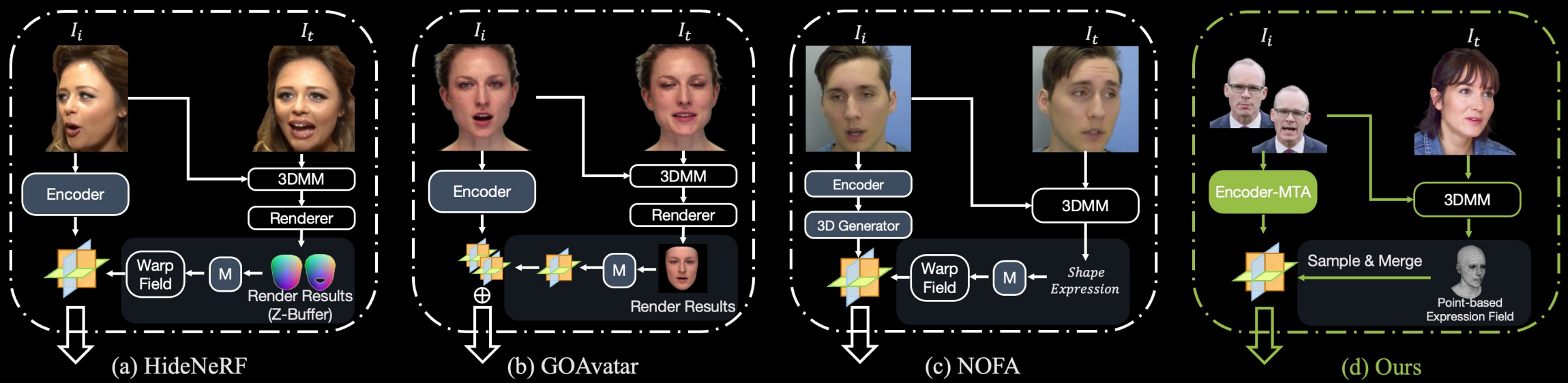
[1]The University of Tokyo, [2]International Digital Economy Academy (IDEA), [3]RIKEN AIP

Overview



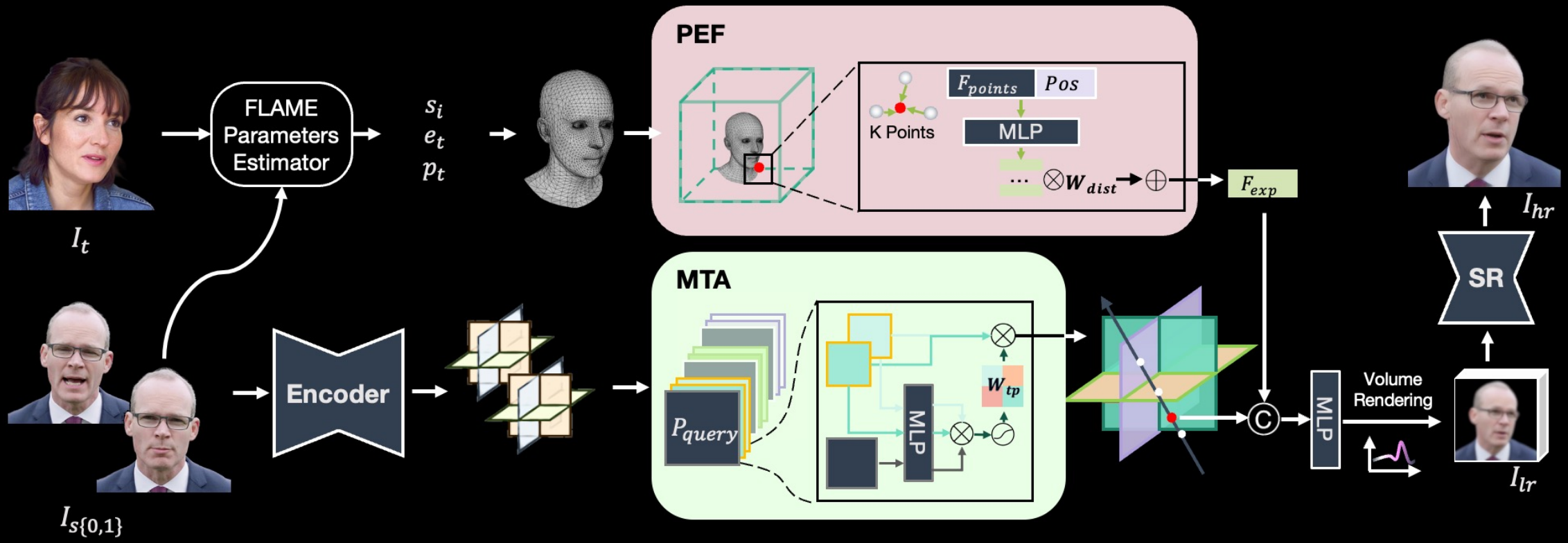
With one or several image inputs, we can reconstruct a controllable head avatar.

Different from Related Works



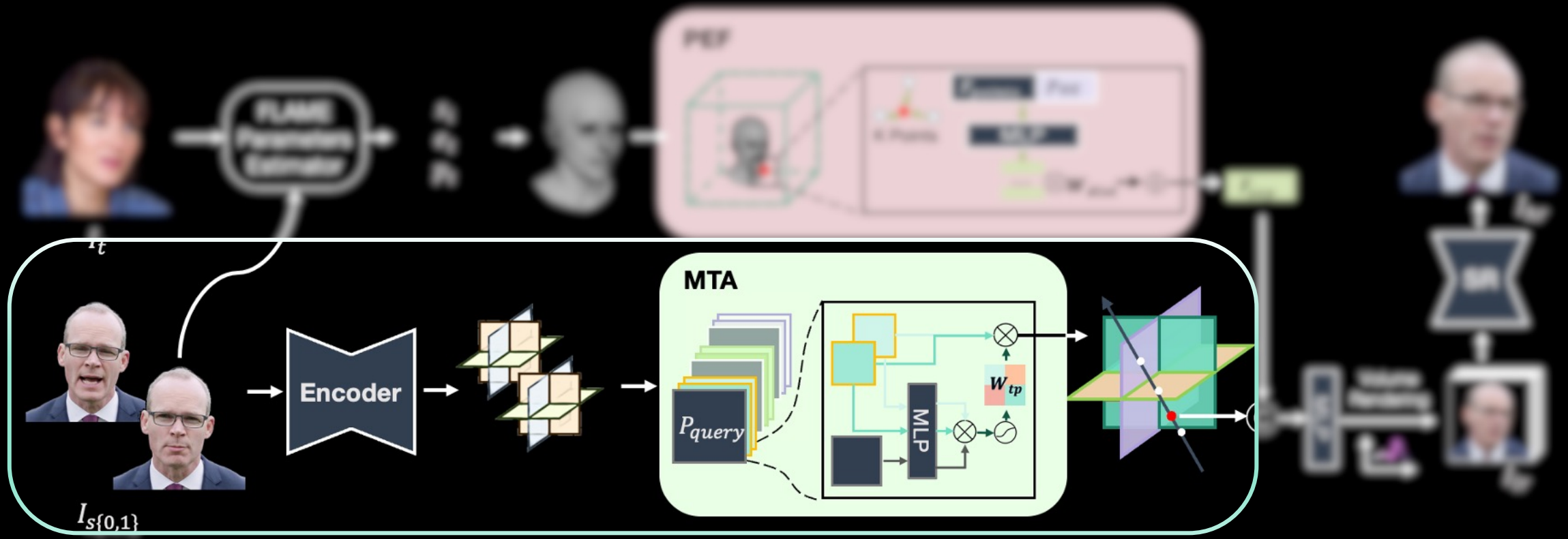
Compared with previous work, our method avoids the loss of expression details caused by over-processing.

Method Overview



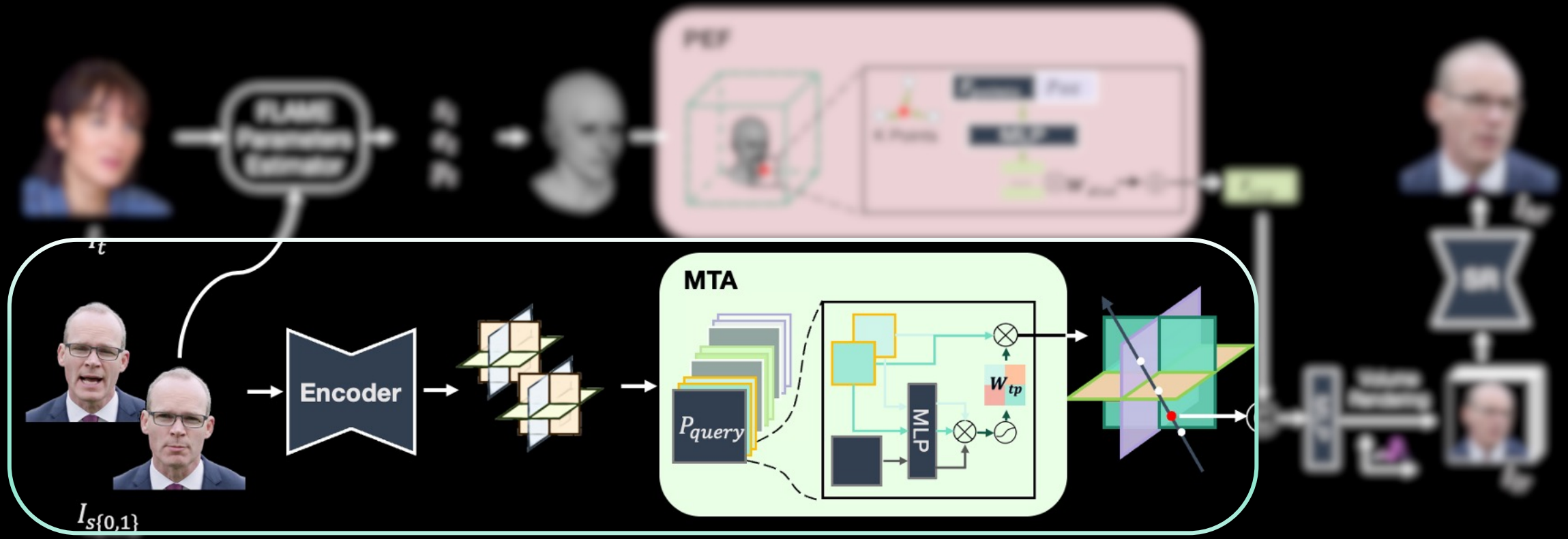
Here is an overview of our method, and we will introduce it step by step.

Canonical Feature Encoder



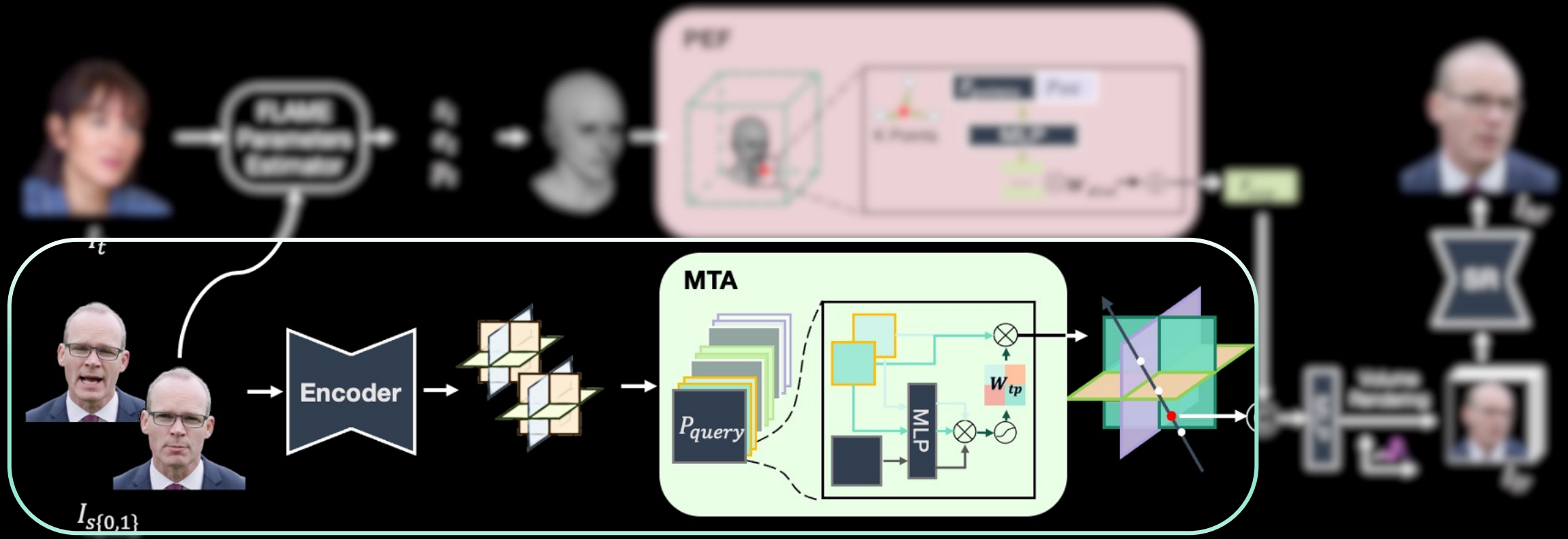
We use our Multi Tri-planes Attention (MTA) module to leverage information from multiple images with tri-planes.

Canonical Feature Encoder



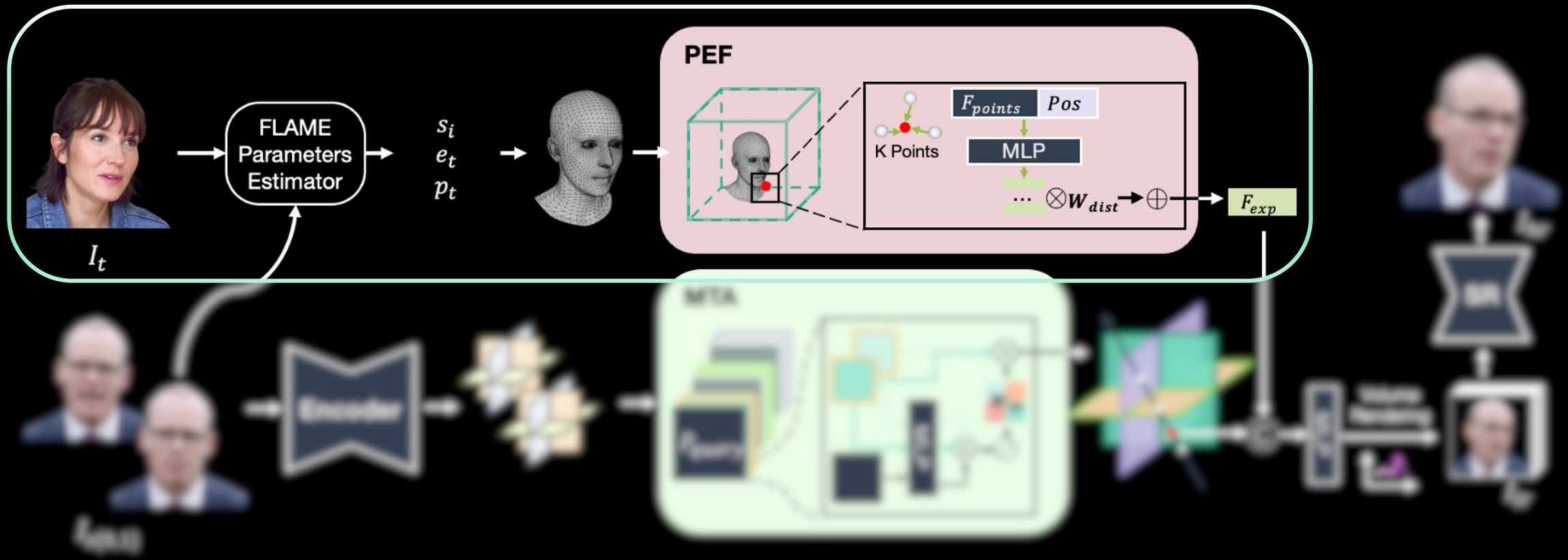
MTA can accept any number of images as input, including one image or multiple images.

Canonical Feature Encoder



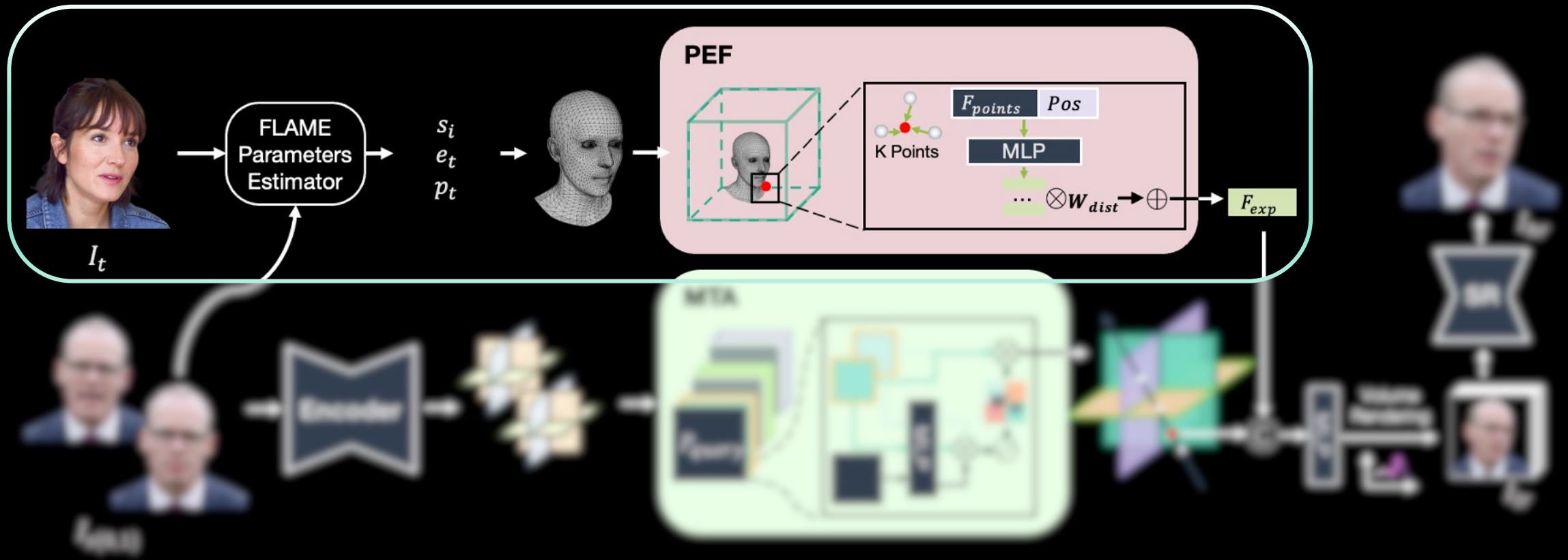
And use query planes to determine the weights of different planes and then merge them.

Point-based Expression Field



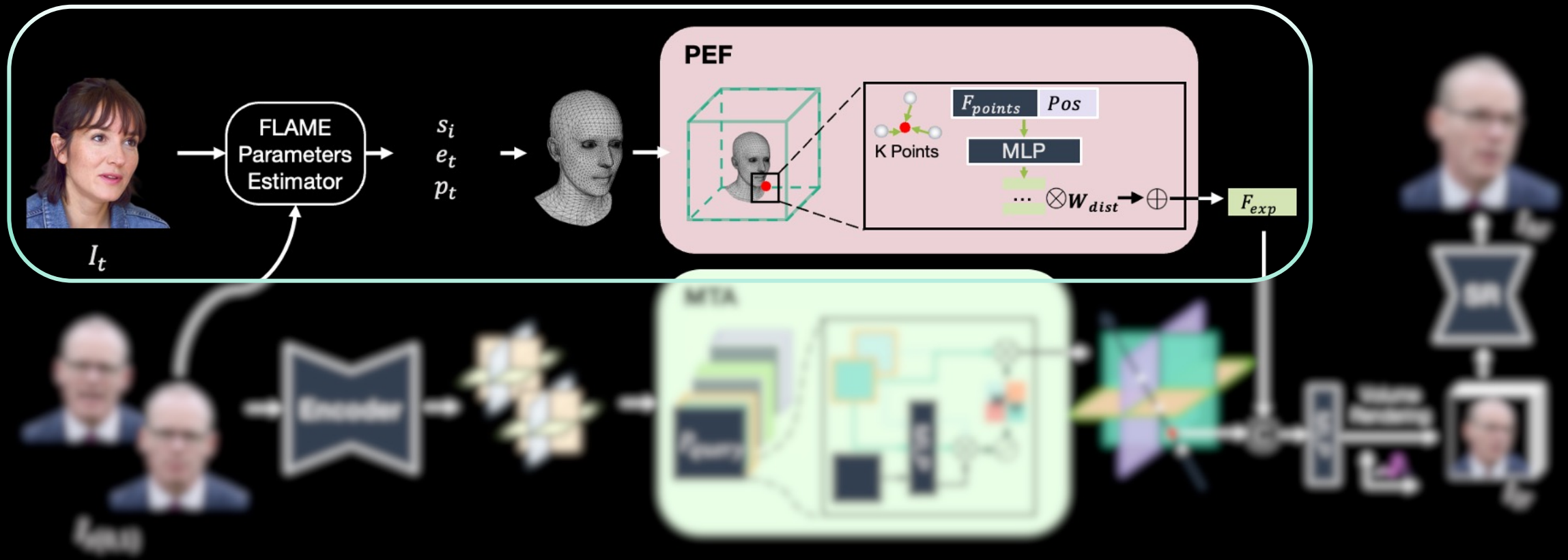
To achieve the re-driving capability, we design a point-based expression field (PEF).

Point-based Expression Field



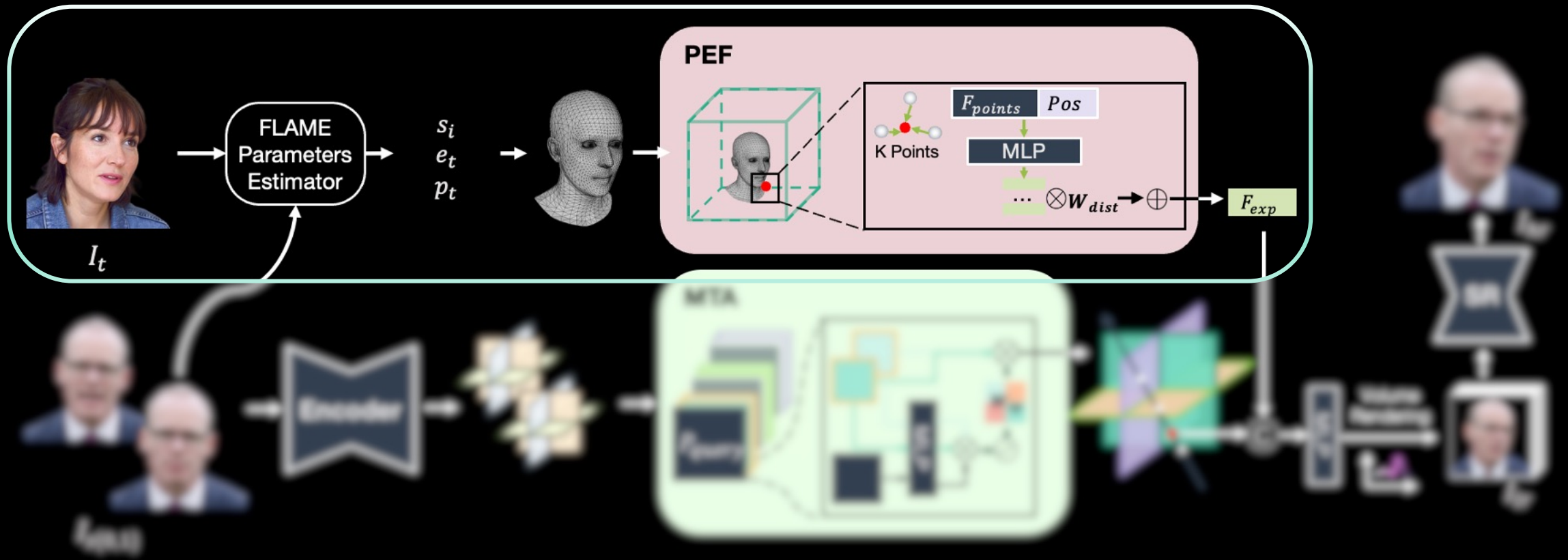
The point cloud is generated by FLAME and aligned with the canonical tri-planes feature space.

Point-based Expression Field



We associate learnable parameters with each point.

Point-based Expression Field



During the volume rendering, we query the nearest K points and merge features based on distance.

Rendering Module

Finally, we map canonical and expression features to RGB and density, perform volume rendering and SR.

Results

These results suggest that our method can capture subtle expressions effectively.

Results

Quantitative Results on VFHQ Dataset

Method	Self Reenactment								Cross-Id Reenactment		
	PSNR↑	SSIM↑	LPIPS↓	CSIM↑	L1↓	AED↓	APD↓	AKD↓	CSIM↑	AED↓	APD↓
ROME(Khakhulin et al., 2022)	19.88	0.735	0.237	0.679	0.060	0.497	0.017	4.53	0.531	0.936	0.026
StyleHeat(Yin et al., 2022)	19.95	0.738	0.251	0.603	0.065	0.593	0.024	5.30	0.506	0.961	0.038
OTAvatar(Ma et al., 2023)	18.10	0.600	0.346	0.660	0.092	0.734	0.035	6.05	0.514	0.962	0.059
Next3D(Sun et al., 2023)	19.95	0.656	0.281	0.631	0.066	0.727	0.026	5.17	0.482	0.996	0.036
HideNeRFLi et al. (2023a)	20.07	0.745	0.204	0.794	0.056	0.521	0.031	5.33	0.558	1.024	0.044
Ours One-in	22.08	0.765	0.177	0.789	0.039	0.434	0.017	3.53	0.558	0.910	0.034
Ours Two-in	22.86	0.779	0.169	0.771	0.035	0.411	0.017	3.44	0.551	0.907	0.034

* For a fair comparison, we compare and highlight all single-input methods. The multi-input method Ours-MT provided only for reference.

Quantitative results also support this conclusion.

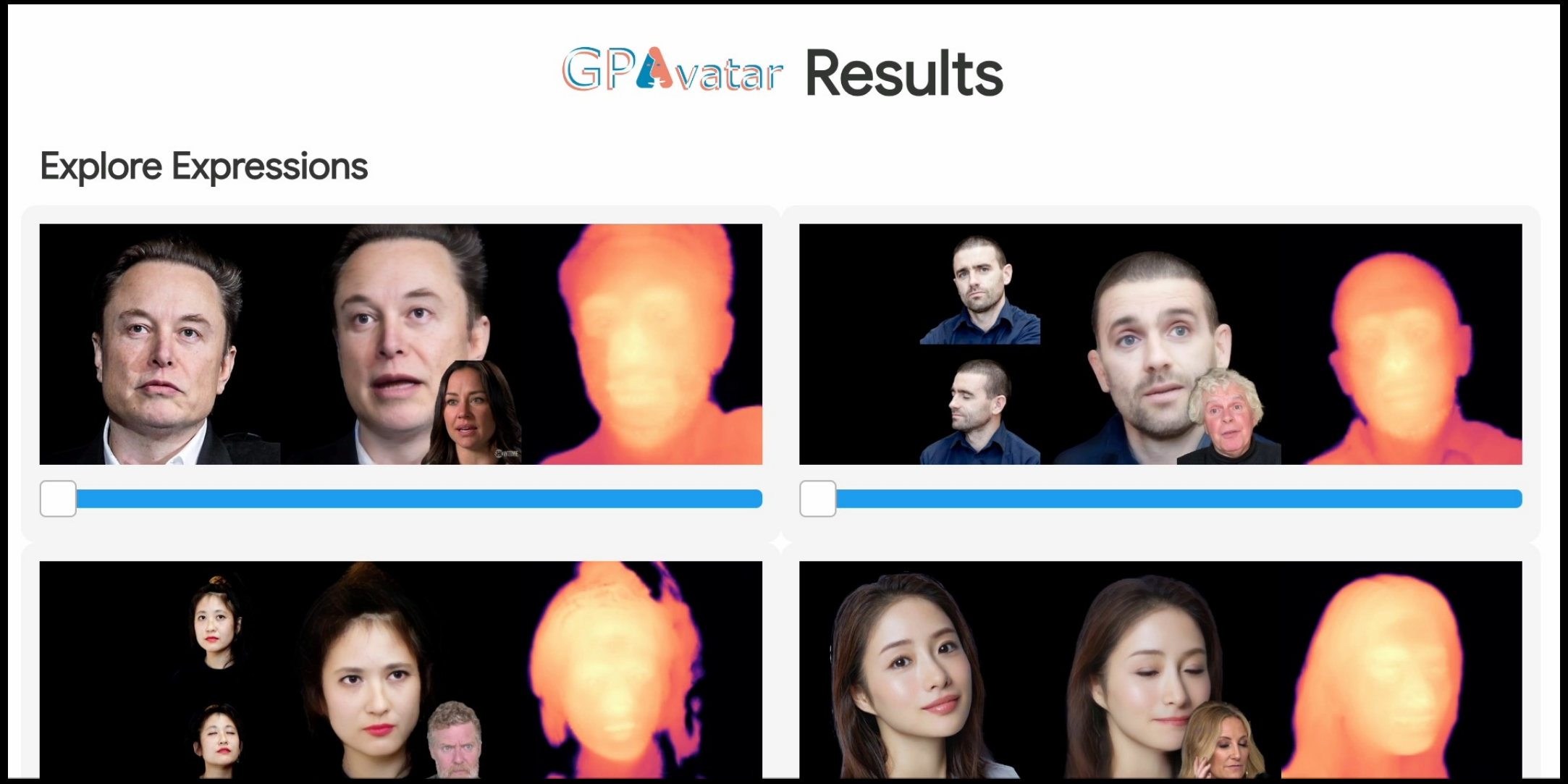
Results

Here are some results of our model.

Results

Here are some results of our model.

Results



More results can be found from our project website.

Thank you

GPAvatar

GPAvatar: Generalizable and Precise Head Avatar from Image(s)

東京大学 THE UNIVERSITY OF TOKYO idea RIKEN