# Supplementary file for the main article A Subspace-based Method for Facial Image Editing

We use an iterative manner to modulate the parameters of the StyleGAN2 generator layer by layer, e.g., Eq.1.

$$g(\mu,\sigma)=\begin{cases}\mu_i^n+\sigma_i^n>\mu_i^{n-1}, \text{if} \quad i=1,2 \\ \mu_i^n+\sigma_i^n<\mu_i^{n-1}, \text{if} \quad i=3,4,5,6,7\end{cases} \tag{1}$$

Eq.1 as AWM's convergent criterion, where $i$ is an index of a metric in Table I, e.g., $\mu_1$ has the same meaning as that of $\mu_{\text{MSE}}$. $\mu_i^n$ and $\sigma_i^n$ denote the mean and variance of the $i'\text{th}$ metric with the number of iterations $n$. ↓ indicates that the small value corresponds to high image quality, while ↑ means the opposite in Table I. AWM converges if it meets Eq.1. By fully considering the mean and variance of multiple metrics, we eliminate oscillation effects with different batch input images. The proposed algorithm becomes more robust than those using an individual metric only.

Table I The convergent criterion is based on the following metrics. ↑ denotes that the higher, the better. ↓ denotes that the lower, the better.

| **Configuration** | **MSE** | **RMSE** | **PSNR** | **UQI** | **SSIM** | **MS-SSIM** | **VIF** |
|---|---|---|---|---|---|---|---|
| **index** | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| **trend** | ↓ | ↓ | ↑ | ↑ | ↑ | ↑ | ↑ |

We further provided a deep analysis of the convergence criterion and conducted an extra experiment to show that Eq.1 is a valid and effective criterion. The details are as follows:

1) It has diversified image evaluation metrics. We quantitatively analyze the image generated from full-space and sub-space in different aspects. Mean Square Error (MSE), Root Mean Square Error (RMSE), and Peak Signal Noise Ratio (PSNR), all of which can be used to evaluate the differences between the full-space and sub-space images from a global perspective. They treat all pixels in an image equally while not reflecting the full range of human visual characteristics. The Universal Quality Index (UQI) measures image quality by relevance, brightness, and contrast differences. The higher the value, the better the image quality. However, it cannot correlate all subjective evaluations, leading to its instability.

2) As a full-reference image quality evaluation metric, Structural Similarity Measure (SSIM) measures the similarity of images in brightness, contrast, and structure. It outperforms PSNR in terms of image denoising and similarity evaluation. It calculates different blocks and averages the results to ensure stability and satisfy a human visual system (HVS) for local information. Multi-Scale Structural Similarity Index (MS-SSIM) calculates the structural similarity index by combining multiple images at different scales. It can be more robust than SSIM when the observation condition changes. Visual Information Fidelity (VIF) is an image quality evaluation metric based on statistical image models, image distortion models, and human visual system models. Compared with PSNR and SSIM, it has higher consistency in subjective vision. The higher the VIF value, the better the image quality.

3) Generalizability. Eq.1 is a general criterion that can be used to modulate the parameters for multi-generators with different architectures, e.g., PGGAN, StyleGAN, StyleGAN2, and the third-party StyleGAN2 generators. All the sub-spaces of these generators are derived using Eq.1.

4) Robust sub-space. We can obtain the sub-space for generators when AWM meets Eq.1. Noteworthy, each sub-space is robust, which keeps rich knowledge in face generators and generates fine faces as done in full space. The results are shown in Fig.1.

5) Sufficient ablation studies. We divide the image evaluation metrics into three parts: 1) pixel-based metrics, i.e., MSE, RMSE, and PSNR. 2) image-based metrics, i.e., UQI, SSIM, and MS-SSIM. 3) model-based metrics, i.e., VIF. According to the three parts, we set three ablation studies. Removing pixel-based metrics while retaining other metrics, labeled as Config-1. Removing image-based metrics while retaining other metrics, labeled as Config-2. Removing model-based metrics while retaining other metrics, labeled as Config-3.

Table II The ablation studies for verifying Eq.1, $\sigma$ denotes the singular value when AWM convergences.

| StyleGAN2 | Offcial | Baby | Celebrity | Star | Model |
|---|---|---|---|---|---|
| Config-1 | $\sigma = 190$ | $\sigma = 95$ | $\sigma = 90$ | $\sigma = 100$ | $\sigma = 110$ |
| Config-2 | $\sigma = 70$ | $\sigma = 60$ | $\sigma = 50$ | $\sigma = 55$ | $\sigma = 50$ |
| Config-3 | $\sigma = 175$ | $\sigma = 95$ | $\sigma = 125$ | $\sigma = 80$ | $\sigma = 75$ |
| Eq.(3) | $\sigma = 165$ | $\sigma = 90$ | $\sigma = 115$ | $\sigma = 65$ | $\sigma = 95$ |

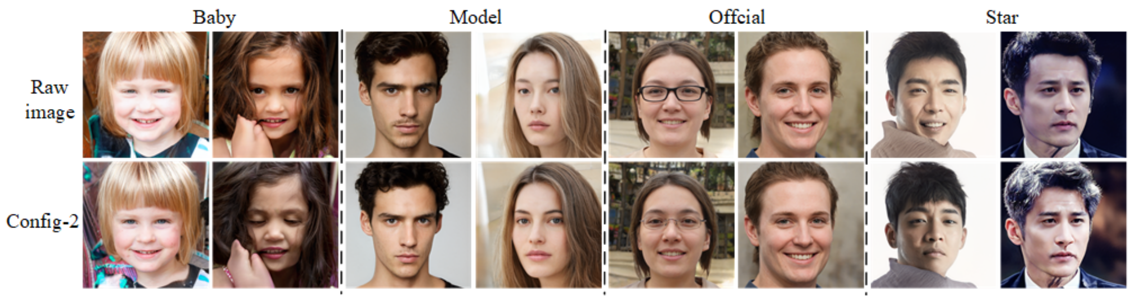


Fig.1 The comparison results between the raw image and generated image. The subspace is derived using Config-2.

We perform ablation studies on five different face generators. The experimental results are shown in the following Table II. The results show that config-1 and config-3 tend to get larger singular values while config-2 tends to get smaller singular values in comparison to Eq.(3). The larger singular values can not yield an optimal subspace, while the smaller singular values can not generate fine faces as done as in full space, e.g., losing facial identity, and producing image artifacts. The corresponding comparison results of Eq.1 are shown in Fig.1.

In addition, Eq.1 eliminates oscillation effects with different batch input images. To summarize, the above technical support and experimental results provide strong evidence to show the effectiveness of Eq.1.