HFSVQ: HIERARCHICAL VECTOR QUANTIZATION WITH RESIDUALS FOR FREQUENCY-SPECIFIC EMBEDDING - SUPPLEMENTARY MATERIALS

1. ALGORITHM FOR ALTERNATELY TRAINING STRATEGY

To ensure the stability of the training process and enhance performance, we propose the 'alternately training strategy'. During the training phase, we update F_L and F_H alternately. The specifics of the strategy are detailed in Algorithm 1. Note that the equations written in the algorithm are based on those in the main paper.

Algorithm 1 Alternately Training Strategy

X: the data used in the training stage.
** **Update** $F_L = \{E_L, Q_L, D_L\}$ **
1: $Z \leftarrow E_L(X)$
2: $Z_q \leftarrow Q_L(Z; C_L)$
3: $\hat{X}_l \leftarrow D_L(Z_q)$
4: Calculate Equation 9
5: Backward to F_L & Update
** **Update** $F_H = \{E_H, Q_H, D_H\}$ **
1: Freeze F_H
2: $\hat{X}_l \leftarrow F_H(X)$
3: $X_r \leftarrow X - \hat{X}_l$
4: $Z_r \leftarrow E_H(X_r)$
5: $Z_{q,r} \leftarrow Q_H(Z_r)$
6: $\hat{X}_r \leftarrow D_H(Z_{q,r})$
7: Calculate Equation 10
8: Backward to F_H & Update

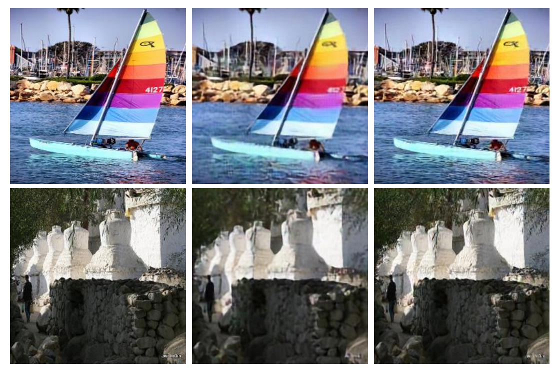
2. ADDITIONAL QUALITATIVE RESULTS FOR ABLATION STUDY

2.1. The Effect of Residual Image

Fig. 1 presents the qualitative results to evaluate the effectiveness of the residual image. It demonstrates that utilizing the residual image as input to F_H, rather than the whole output from F_L, better preserves image details.

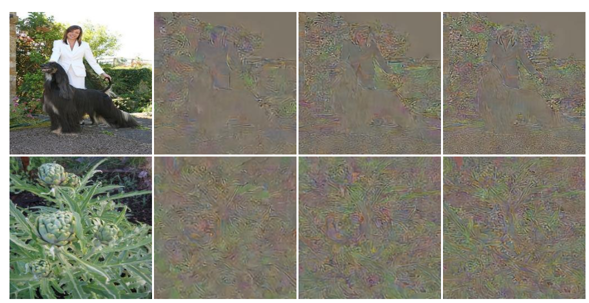
2.2. The Effect of Product Quantization

Fig. 2 demonstrates the relationship between the number of subspaces in C_H and the capture of high-frequency details. The results indicate that high-frequency details become clearer as the number of subspaces increases.



(a) (b) (c)

Fig. 1. Qualitative results to evaluate the effect of residual image. Each column visualizes (a) the input image, (b) the reconstructed image using whole image from F_L, and (c) the reconstructed image using the residual image.



(a) (b) (c) (d)

Fig. 2. Qualitative results to evaluate the capture of high-frequency details from (a) input image, based on the number of subspaces: (b) 1, (c) 2, and (d) 4.

3. DECISION FOR THE NUMBER OF CODES PER SUBSPACE

To determine the optimal number of codes for each subspace (*i.e.*, K_s), we conduct experiments across various settings of K_s, where K_s corresponds to (2, 4, 8, 16, 32, 64, 128). To

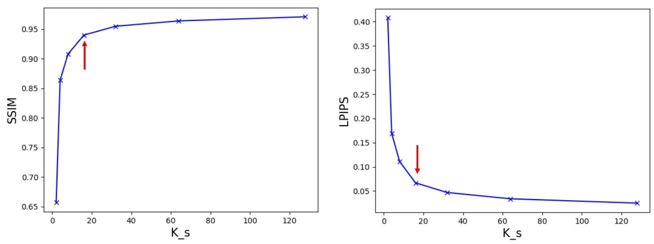


Fig. 3. Changes in SSIM and LPIPS according to the number of codes per subspace (K_s). The red arrows indicate the position $K_s = 16$, which we select.

evaluate which K_s is sufficient, we use SSIM and LPIPS metrics. As shown in Fig. 3, performance increases sharply until $K_s = 16$, but flattens out relatively after that. Since one of our goals is to minimize the resource allocation related to codebooks while balancing performance, we decide to assign 16 codes to each subspace.

4. ADDITIONAL EXPERIMENTAL RESULTS

We provide additional qualitative results for FFHQ (Fig. 4) and ImageNet (Fig. 5).

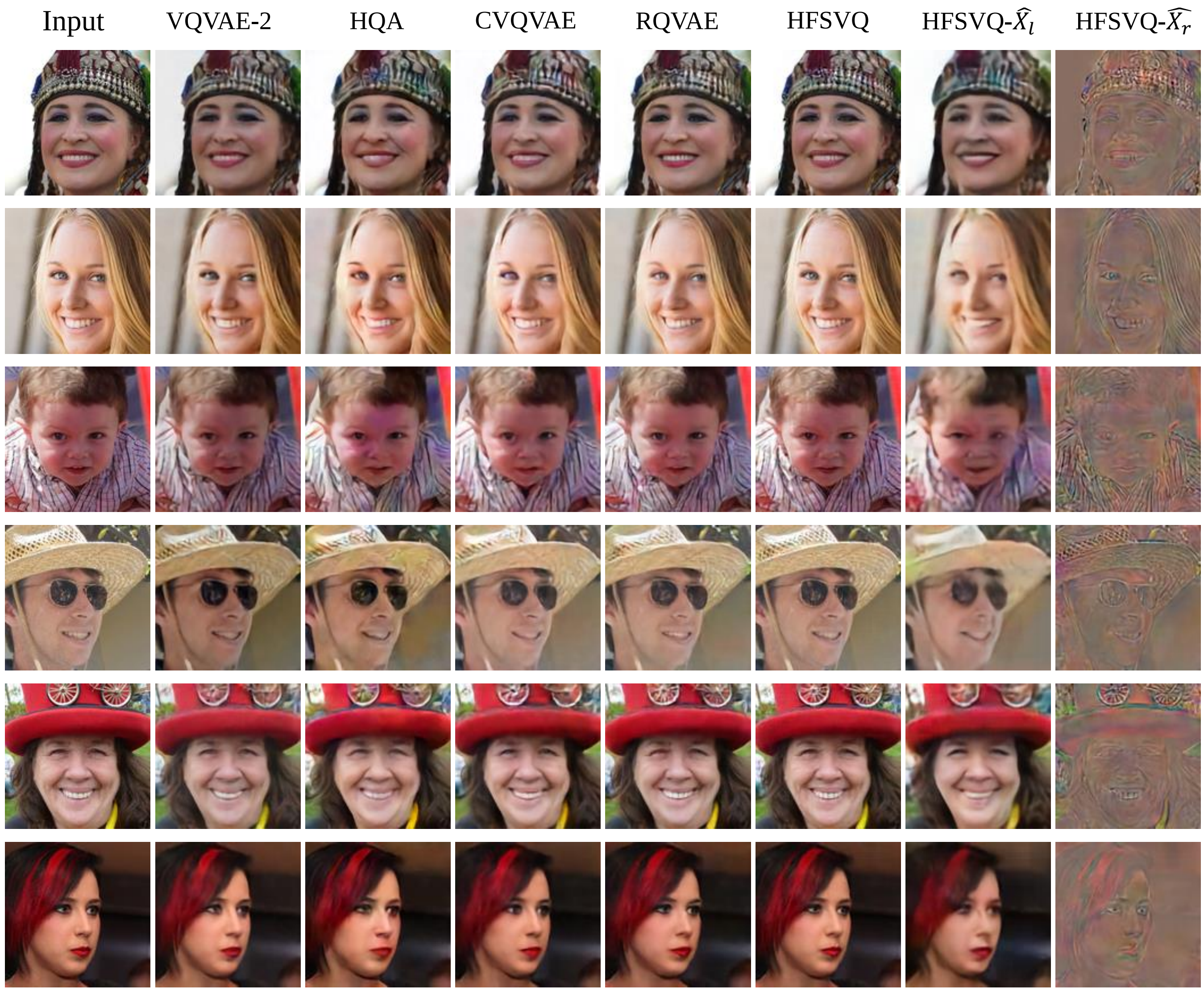


Fig. 4. Additional qualitative results for FFHQ.

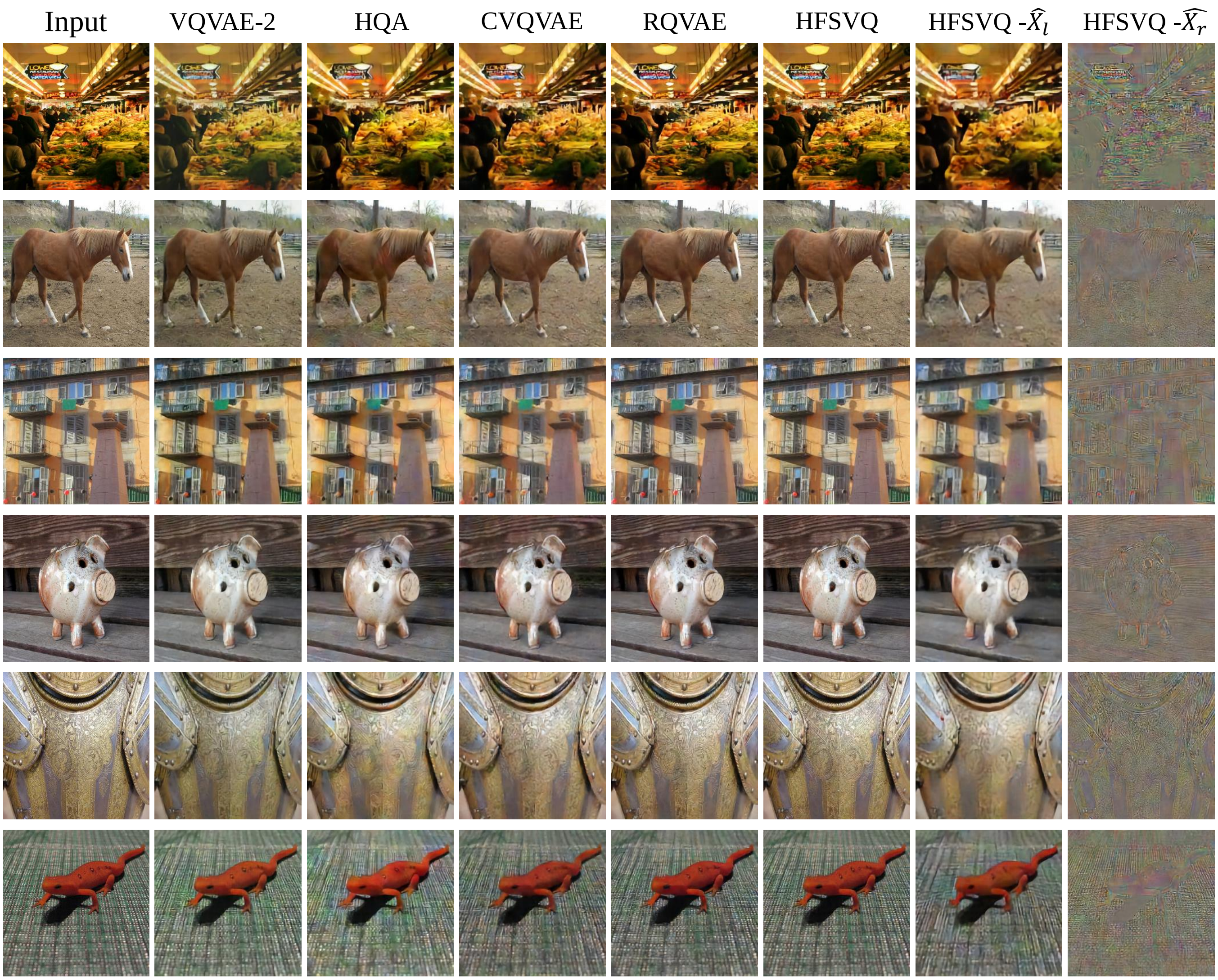


Fig. 5. Additional qualitative results for ImageNet.