WavePaint: Resource-efficient Token-mixer for Self-supervised Inpainting

Supplementary Material

1. Qualitative Results

The qualitative results of inpainting by WavePaint for wide, medium, and narrow masks are shown in Figure 1, Figure 2, and Figure 3, respectively. All the images were generated by WavePaint (12M parameters) trained only on wide masks. We see that WavePaint is able to fill in missing region with facial features. It is able to match the eye colours, eyebrows etc from the visible region to the generated parts.

In some images we observe the presence of fine-texture artifacts near the masked portions. We ran some experiments by training WavePaint in adversarial settings using a discriminator and observed that these artifacts disappear. For future work, we will create an adversarially trained WavaPaint model which would produce more realistic images and will be resource-efficient.

Additional qualitative comparison of images inpainted using WavePaint with those inpainted using LaMa [2] are shown in Figure 4.

2. WaveMix Block [1]

WaveMix blocks with one and three levels of 2D-discrete wavelet transform (2D-DWT) are shown in Figure 5 and Figure 6 respectively. WaveMix block having a single level of 2D-DWT is called WaveMix-Lite. Denoting input and output tensors of the WaveMix block by $\mathbf{x}_{in}$ and $\mathbf{x}_{out}$, respectively; level of the wavelet transform by $l \in \{1...L\}$, the four wavelet filters along with their downsampling operations at each level by $w^l_{aa}, w^l_{ad}, w^l_{da}, w^l_{dd}$ (a for approximation, d for detail); convolution, multi-layer perceptron (MLP), transposed convolution (upconvolution), and batch normalization operations by c, m, t, and b, respectively; and their respective trainable parameter sets by ξ, θ_l, ϕ_l, and γ_l, respectively; concatenation along the channel dimension by $\oplus$, and point-wise addition by $+$, the operations inside a WaveMix block can be expressed using the following equations:

$$\mathbf{x}_0 = c(\mathbf{x}_{in}, \xi); \mathbf{x}_{in} \in \mathbb{R}^{H\times W\times C}, \mathbf{x}_0 \in \mathbb{R}^{H\times W\times C/4} \quad (1)$$

$$\mathbf{x}_l = [w^l_{aa}(\mathbf{x}_0) \oplus w^l_{ad}(\mathbf{x}_0) \oplus w^l_{da}(\mathbf{x}_0) \oplus w^l_{dd}(\mathbf{x}_0)]; \quad \mathbf{x}_l \in \mathbb{R}^{H/2^l\times W/2^l\times 4C/4}, l \in \{1...L\} \quad (2)$$

$$\hat{\mathbf{x}}_l = [\mathbf{x}_l \oplus \tilde{\mathbf{x}}_{l+1}], \quad \hat{\mathbf{x}}_L = \mathbf{x}_L; \quad l \in \{1...L-1\} \quad (3)$$

$$\tilde{\mathbf{x}}_l = b(t(m(\hat{\mathbf{x}}_l, \theta_l), \phi_l), \gamma_l); \quad \tilde{\mathbf{x}}_l \in \mathbb{R}^{H/2^{l-1}\times W/2^{l-1}\times C_l} \quad \forall l > 1 \;\; C_l = C/2, \;\; C_1 = C, \;\; l \in \{1...L\} \quad (4)$$

$$\mathbf{x}_{out} = \tilde{\mathbf{x}}_1 + \mathbf{x}_{in}; \quad \mathbf{x}_{out} \in \mathbb{R}^{H\times W\times C} \quad (5)$$

3. Mask Generalization

In all the reported results, WavePaint was trained only on wide masks and tested on narrrow, medium and wide masks. For checking the generalisation ability of WavePaint to larger mask sizes, we trained WavePaint on medium masks and tested its performance on wide masks. The results are shown in Table 1. We see that WavePaint is providing good performance in wide masks even when trained of medium masks. This shows that WavePaint is able to generalize well on larger unseen masks during inference.

References

[1] Pranav Jeevan, Kavitha Viswanathan, Anandu A S, and Amit Sethi. Wavemix: A resource-efficient neural network for image analysis, 2023. 1, 3, 4

[2] Roman Suvorov, Elizaveta Logacheva, Anton Mashikhin, Anastasia Remizova, Arsenii Ashukha, Aleksei Silvestrov, Naejin Kong, Harshith Goka, Kiwoong Park, and Victor Lempitsky. Resolution-robust large mask inpainting with fourier convolutions, 2021. 1, 3

Figure 1. Wide-masked images from the CelebA-HQ dataset (top row) and their inpainted versions generated by WavePaint (bottom row) Inpainted images (bottom row)

Figure 2. Medium-masked images from the CelebA-HQ dataset (top row) and their inpainted versions generated by WavePaint (bottom row) Inpainted images (bottom row)

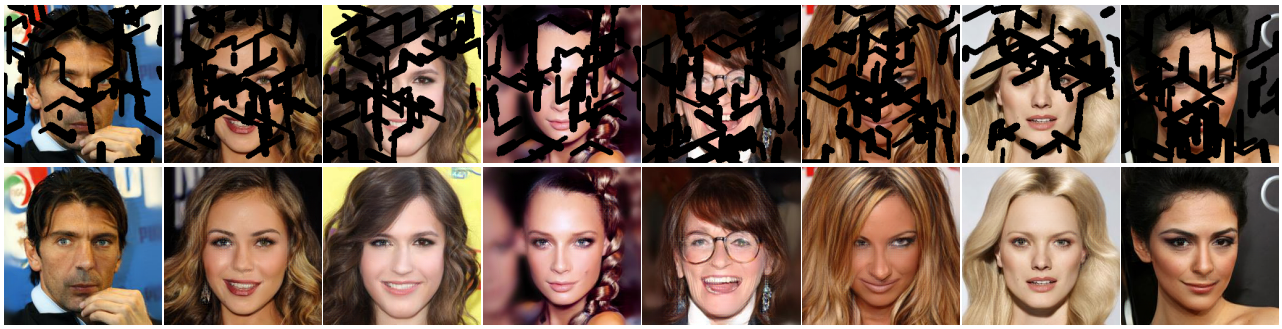
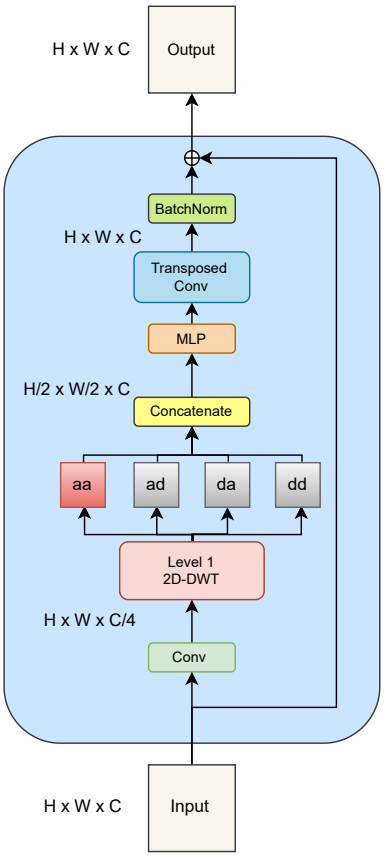


Figure 3. Narrow-masked images from the CelebA-HQ dataset (top row) and their inpainted versions generated by WavePaint (bottom row) Inpainted images (bottom row)

CelebA-HQ (256×256)							
Model	#Params↓	Narrow masks		Medium masks		Wide Masks	
		FID↓	LPIPS↓	FID↓	LPIPS↓	FID↓	LPIPS↓
WavePaint	12 M	4.68	0.073	4.98	0.082	6.69	0.108
WavePaint	13 M	4.44	0.070	4.79	0.080	6.27	0.104

Table 1. Mask generalization performance of WavePaint when using medium masks for training.

Figure 4. Qualitative comparison of inpainted images generated by WavePaint and LaMa [2]. The green arrows point to improper image completion.



WaveMix-Lite Block (1-level 2D-DWT)

Figure 5. WaveMix block architecture using level-1 2D-discrete wavelet transform. The image is take from [1]

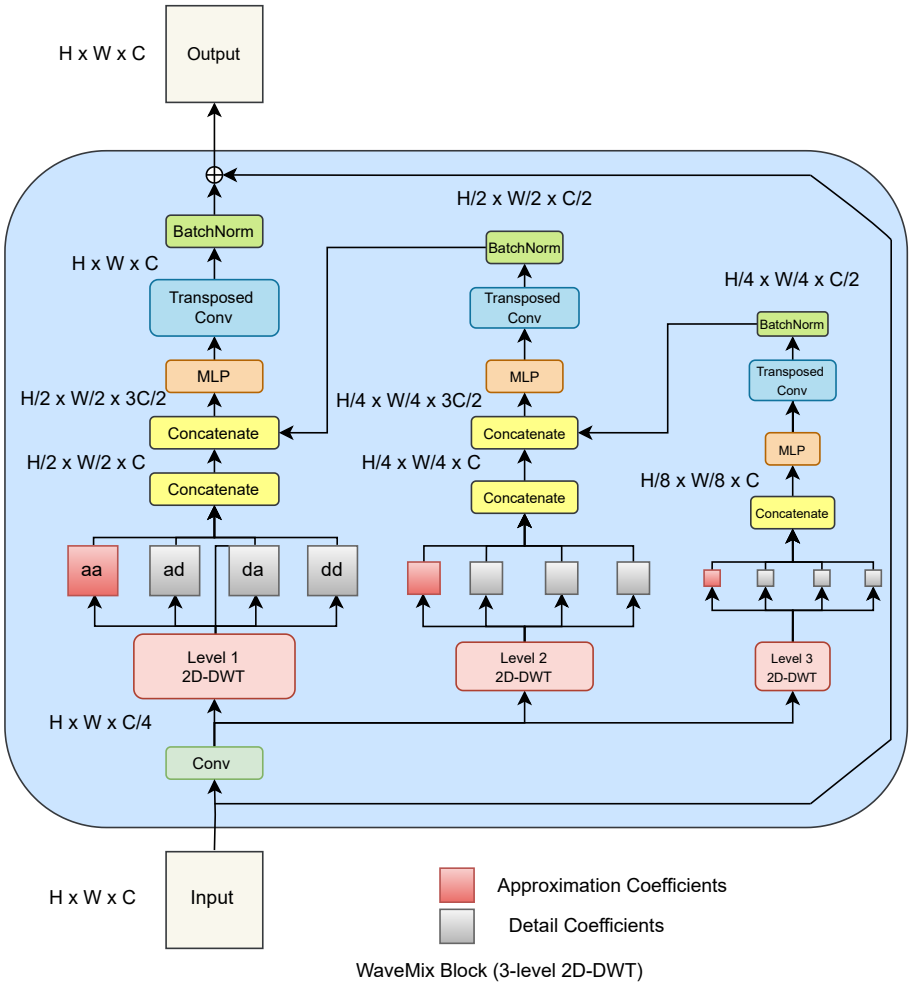


Figure 6. Details of the WaveMix block with 3 levels of 2D-DWT. The image is take from [1]