Meta-FC: Meta-Learning with Feature Consistency for Robust and Generalizable Watermarking

Supplementary Material

Table 6. Details of the experimental settings.

Details	Parameters
Operating System	Linux
GPU	NVIDIA RTX 3090
Memory	24 GB
Software libraries and frameworks	Python 3.10
	Albumentations 2.0.5
	Torch 2.6.0
	Pillow 11.2.1
	OpenCV-Python 4.11.0.86
	Scikit-Image 0.25.2

1. Detailed Training Configurations

This section provides the detailed training configurations. In the case of SepMark, only the encoder and a robust decoder are retained for training. For StegaStamp and DERO, we replace their distortion simulation modules with a combined distortion noise layer, while preserving the original encoder and decoder architectures. In particular, both SRD and Meta-FC are trained in the same settings to ensure a fair and consistent comparison. Details of the experimental setup and configurations are provided in Table 6.

2. Detailed Explanation of Distortions and Their Parameters in Experiments

In this section, we provide detailed descriptions of the distortions used in this paper, including their parameter settings.

Gaussian Blur (GB): Gaussian blur applies a Gaussian kernel to scan and modify each pixel in the image. The central pixel value in the kernel is replaced with the weighted average of neighboring pixels, where the weights follow a Gaussian distribution. In our experiments, the kernel size is set to 7×7, with a mean of 0 and a standard deviation of σ.

Salt-and-Pepper Noise (SPN): Salt-and-pepper noise randomly replaces a proportion ($p \in (0, 1)$) of the pixels in the watermarked image with either black or white pixels with equal probability.

Crop: Cropping involves replacing a rectangular patch in the watermarked image with zeros. The width and height of the patch are determined as a proportion ($p \in (0, 1)$) of the image dimensions.

JPEG Compression (JPEG): Since real JPEG compression is non-differentiable, we adopt a differentiable JPEG simulation layer following MBRS to emulate the compression process during training.

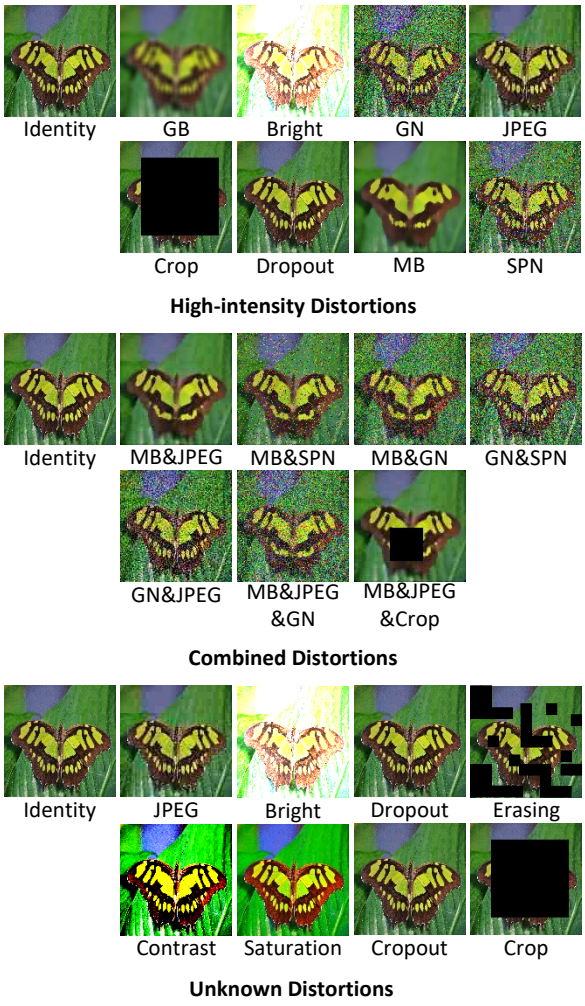

Figure 4. Visualization of known and unknown distortions for watermarked image.

Median Blur (MB): Median blur replaces each pixel with the median value of its neighboring pixels within a kernel. The only parameter is the kernel size, e.g., 5×5.

Gaussian Noise (GN): Gaussian noise refers to additive noise sampled from a Gaussian distribution. In our setup, the mean is set to 0 and the variance is σ.

Brightness (Bright): This distortion modifies the brightness of the watermarked image by multiplying all pixel val-

Table 7. ACC (%) comparison with per-SRD and PDL training strategies.

		Identity	GB	SPN	Crop	JPEG	MB	GN	Bright	Dropout	MB&JPEG	MB&SPN	MB&GN	GN&SPN	GN&JPEG	MB&SPN&GN	MB&JPEG&Crop
MBRS	per-SRD	**100**	97.72	98.04	92.82	84.67	91.36	93.73	92.90	98.80	73.87	89.66	71.05	94.37	78.98	68.77	76.27
	PDL	**100**	99.07	98.08	92.29	85.68	92.84	94.25	**94.79**	98.39	75.46	95.32	75.93	93.78	**80.05**	69.42	75.57
	Meta-FC	**100**	**99.33**	**98.44**	**95.78**	**86.79**	**95.25**	**96.01**	93.93	**99.75**	**84.18**	**99.23**	**82.57**	**96.51**	79.71	**75.28**	**82.44**
FIN	per-SRD	**100**	93.73	97.23	81.59	99.38	88.70	93.01	**89.96**	84.22	94.82	87.87	78.53	94.06	94.47	73.88	91.81
	PDL	**100**	96.07	97.33	82.51	99.84	87.48	94.60	89.31	84.49	95.10	87.52	78.39	94.27	95.04	74.27	93.80
	Meta-FC	**100**	**97.16**	**98.27**	**83.15**	**99.89**	**90.18**	**95.70**	89.32	**85.58**	**97.01**	**92.32**	**80.06**	**95.70**	**95.54**	**76.69**	**95.38**
DERO	per-SRD	**100**	**100**	98.50	96.34	87.57	99.85	93.02	91.88	95.69	95.96	99.71	93.07	93.64	89.78	89.89	94.76
	PDL	**100**	99.90	97.97	96.55	88.13	99.92	93.07	92.39	96.06	96.91	99.37	93.20	94.15	92.29	90.81	97.24
	Meta-FC	**100**	**100**	**99.61**	**97.44**	**91.78**	**100**	**95.04**	**92.86**	**98.10**	**98.23**	**99.97**	**95.48**	**96.08**	**93.75**	**92.38**	**97.53**

Table 8. ACC (%) comparison with the geometric distortion baseline.

	PSNR	Translate (0.5)	Shear (40)	Scale (0.6)	Rotate (60°)	Rotate (60°) & Shear (20)	Rotate (60°) & Translate (0.5)	Rotate (60°) & Scale (0.9)
SRD	36	**100**	94.94	94.71	**100**	90.37	99.95	91.31
Meta-FC	36	**100**	**96.39**	**95.12**	**100**	**92.44**	**100**	**93.19**

Table 9. ACC (%) comparison at PSNR = 38.

		PSNR	High-intensity	Combined	Unknown
FIN	SRD	38	90.51	73.29	80.88
	Meta-FC	38	**92.38**	**77.06**	**83.32**
DERO	SRD	38	90.51	83.95	89.41
	Meta-FC	38	**92.89**	**86.04**	**91.65**

ues by a brightness factor f. The value of f is randomly sampled from a predefined range, e.g., $f \in [0.85, 1.15]$, or fixed to a specific value such as $f = 4$.

Dropout: Dropout replaces a proportion ($p \in (0, 1)$) of pixels in the watermarked image with corresponding pixels from the original cover image. The dropout pixels are randomly selected across the entire image.

Erasing: Erasing randomly selects a proportion ($\sigma \in (0, 1)$) of pixels within the watermarked image and replaces all pixels inside this region with a constant value (0). This operation simulates occlusion or local corruption.

Contrast: Contrast adjusts the global contrast of the watermarked image by applying a linear intensity transformation. A contrast factor is sampled from a predefined range, e.g., ($f = 4$). Increasing (f) amplifies the intensity differences between bright and dark regions, whereas decreasing (f) suppresses these differences, leading to a flatter appearance.

Saturation: Saturation modifies the color vividness of the watermarked image by scaling the chromatic components in a perceptual color space. A saturation factor (f) is sampled from a specified interval, such as ($f = 3$).

3. Visualization of Distorted Images

In this section, we present the visualizations of the distortions used in this paper, including both known and unknown distortions, as illustrated in Figure 4.

4. Additional Experimental Analysis

4.1. Comparison with per-SRD and PDL.

We conduct additional experiments with per-SRD and PDL, as reported in Table 7. The results show that Meta-FC consistently achieves superior performance. Furthermore, Meta-FC is compatible with per-SRD and PDL training schemes and can be seamlessly integrated with them, leading to further improvements in optimization stability and robustness.

4.2. Comparison with Geometric Baseline.

We additionally evaluate SRD and Meta-FC based on the geometric distortion baseline [25], as shown in Table 8. The results further validate the generality and effectiveness of Meta-FC.

4.3. Comparison under High PSNR.

To evaluate performance under stricter imperceptibility constraints, we adjust the embedding strength to achieve a high PSNR of 38 dB. The corresponding results are reported in Table 9, demonstrating that Meta-FC maintains superior robustness even at high visual quality.

4.4. Detailed Ablation Study Results

In this section, we conduct a comprehensive ablation study to assess the individual contributions of the three core components of Meta-FC: the meta-training phase, the meta-testing phase, and the feature consistency loss. The study is carried out across a variety of watermarking models to evaluate the general effectiveness of each component. The robustness results are summarized in Table 10 and Table 11, where we compare the following configurations: **Mate-FC *w/o* meta-train & FC**, **Mate-FC *w/o* meta-test & FC**, **Mate-FC *w/o* meta-test**, **Mate-FC *w/o* FC** and **Mate-FC *w/o* meta-train & FC**. The results show that each component contributes complementary benefits. The complete

Table 10. Impact of different Meta-FC configurations on various models under known distortions.

Strategy	ACC (%)↑									
	StegaStamp		MBRS		FIN		SepMark		DERO	
	High-intensity	Combined	High-intensity	Combined	High-intensity	Combined	High-intensity	Combined	High-intensity	Combined
Mate-FC *w/o* meta-train & FC	94.72	98.18	94.75	77.44	91.66	88.57	94.62	80.44	95.96	94.57
Mate-FC *w/o* meta-test & FC	95.05	98.66	95.09	80.28	93.02	89.09	96.75	90.48	96.47	95.37
Mate-FC *w/o* meta-test	95.68	98.81	95.98	82.59	**93.84**	90.07	96.73	88.49	97.05	95.52
Mate-FC *w/o* FC	95.36	98.59	95.07	79.29	93.71	89.03	96.88	89.83	96.77	95.77
Meta-FC	**95.79**	**98.79**	**96.30**	**84.45**	93.43	**90.24**	**97.37**	**93.67**	**97.07**	**96.13**

Table 11. Impact of different Meta-FC configurations on various models under unknown distortions.

Strategy	ACC (%)↑				
	StegaStamp	MBRS	FIN	SepMark	DERO
Mate-FC *w/o* meta-train & FC	79.56	89.19	86.25	94.84	94.13
Mate-FC *w/o* meta-test & FC	80.61	90.07	86.47	93.49	95.32
Mate-FC *w/o* meta-test	80.45	90.13	86.46	95.74	95.23
Mate-FC *w/o* FC	81.92	**91.45**	87.18	**98.08**	**96.09**
Meta-FC	**82.07**	91.35	**87.29**	97.96	96.00

Meta-FC configuration consistently achieves the highest robustness under diverse distortion settings, demonstrating the effectiveness of integrating meta-learning with feature consistency loss.

5. Discussion with C^{3}hartMark

C^{3}hartMark achieves adaptability by sequentially fine-tuning on various noise layers. However, fine-tuning with new distortion will significantly change the embedded watermark feature, thus influencing the robustness against the previous distortion. Meta-FC adopts a meta-learning paradigm that simulates training on known distortions and testing on a held-out distortion within each batch, thereby narrowing the gradient gap across heterogeneous distortions. In addition, it introduces a feature consistency loss that aligns the last-layer decoder features of clean watermarked images and their distorted counterparts. This joint optimization encourages the model to learn stable and distortion-invariant representations, rather than relying on distortion-specific adaptations.