Supplementary Material for "Frequency Decoupling for Motion Magnification via Multi-Level Isomorphic Architecture"

Fei Wang[1], Dan Guo[1,2]*, Kun Li[1], Zhun Zhong[1,3], Meng Wang[1,2]*
[1] School of Computer Science and Information Engineering, Hefei University of Technology, China
[2] Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, China
[3] School of Computer Science, University of Nottingham, NG8 1BB Nottingham, UK

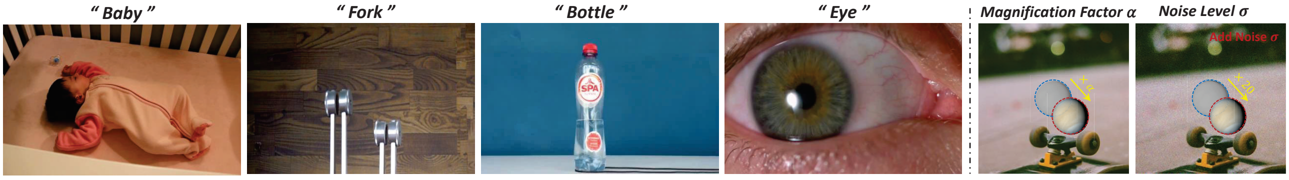


Figure A1. **Illustration of the demonstration video samples.** We randomly select four examples on the left named *baby*, *fork*, *bottle*, and *eye* videos from the Real-world Dataset and the two examples on the right (with magnification factor α and noise level σ, respectively) from the Synthetic Dataset as the demonstration video samples in the demo file FD4MM_demo.mp4.

The manuscript comprehensively analyzes and validates the proposed FD4MM through various qualitative and quantitative experiments. However, due to page space limitations, we do not extend the detailed descriptions of the data source and showcase more magnified results. Here, we offer a more intuitive explanation and demonstration of the data sources and instantiation examples. Furthermore, supplementary experiments are provided to demonstrate the superiority and effectiveness of FD4MM.

1. Summary of Contents

In this supplementary material, Section 2 introduces the widely-used Real-world Dataset and the newly proposed Synthetic Dataset in this work. Section 3 provides more instantiation results and compares our method with current learning-based methods [2, 4, 5] on these demonstration video samples. Please refer to FD4MM_demo.mp4 in the supplement folder for more detailed information. Furthermore, Section 4 supplements the additional experiments to conduct comprehensive validation and analysis from multiple perspectives.

2. Dataset Introductions

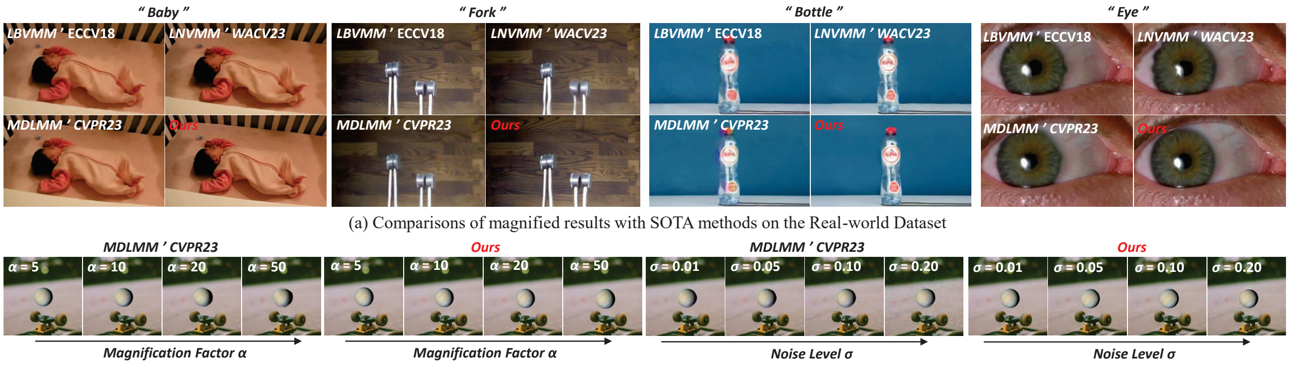
In this field, existing learning-based methods [2, 4, 5] are uniformly trained on the training dataset proposed by [2]. Note that all the methods are trained on the same dataset from [2] without any fine-tuning to ensure fair comparisons. The performance evaluation of the models is performed by cross-dataset testing. We will detail the test datasets used in the manuscript as follows.

Mode	Video	Time (s)	Resolution (pixels)	FPS
Static	Baby [2, 8]	10	960 × 544	30
	Fork [1]	4	544 × 640	30
	Drum [2, 8]	18	640 × 360	25
	Engine [8, 9]	12	1776 × 904	25
	Crane [2, 8]	8	720 × 1280	24
	Face [8, 9]	10	528 × 592	30
Dynamic	Gunshot [2, 5]	2	720 × 576	24
	Cattoy [2, 5]	20	640 × 360	30
	Eye [6, 10]	12	720 × 576	30
	Bottle [2, 10]	6	568 × 320	30
	Drill [4, 5]	2	1280 × 720	24
	Balloon [4, 5]	5	720 × 480	30

Table A1. **Video specification details of the Real-world Test Dataset.**

2.1. Real-world Dataset

The Real-world Dataset exhibits rich motion complexity and uncertainty, and it has been widely used in previous works [1, 2, 5–10] for cross-dataset testing in the motion magnification task. It consists of 12 videos from classic motion scenarios widely used in previous works, which can be split into two modes [2, 4, 5]: (1) static mode (inference on initial frame $I(x, 0)$ and current frame $I(x, t)$) and (2) dynamic mode (inference on continuous frames $\{I(x, t-1), I(x, t)\}$). In addition, we report the detailed time lengths, resolution sizes, and frame rates of all videos in Table A1.



(a) Comparisons of magnified results with SOTA methods on the Real-world Dataset

(b) Comparisons of magnified results with the SOTA method on the Synthetic Dataset

Figure A2. **Visualization samples from the demonstration video.** We capture the magnified images at a random timestamp of video for illustration visualization. In terms of static visualization, we can see our FD4MM exhibits higher magnification quality. These results are also verified by objective quantitative metrics such as LPIPS and MANIQA in the manuscript. The dynamic magnified results of these video samples are displayed in the appendix file FD4MM_demo.mp4.

2.2. Synthetic Dataset

The Synthetic Test Dataset consists of 10 videos of 1s with a resolution size of 640×640. Each video is synthesized with a foreground object from the public StickPNG library [1] and a background image from the DIS5K dataset [3]. As shown in the right side of Fig. A1, by controlling the magnitude of foreground object motion and simulating photographic noise, we evaluate the magnification accuracy at various magnification factors α and noise robustness at various noise levels σ. Here, the synthetic process of all synthetic data follows the rule of [2, 7], including the control of the magnification factor and the process of simulating noise.

3. Demonstration Examples

3.1. Real-world Examples

As shown in Fig. A2 (a), we provide four video examples from Real-world Dataset for intuitive visual demonstration and comparison, *i.e*., the “baby”, “fork”, “bottle”, and “eye” videos. Please refer to FD4MM_demo.mp4 in the supplement folder for more detailed information.

“Baby” Video. The video has subtle abdomen baby breathing motion and an almost fixed background, thus using the static inference mode for all methods. It is observed that at the magnification factor α of 20, all previous SOTA methods [2, 4, 5] can capture and magnify the breathing motion on the baby’s abdomen but generate significant flickering artifacts and introduce ringing artifacts and distortion at the baby’s abdomen. In contrast, our FD4MM effectively suppresses these issues with clearer magnification.

“Fork” Video. The ”fork“ video exhibits rapid and subtle vibrations along with some level of camera shake. In the static inference mode and with a magnification factor α of 20, our FD4MM, compared to previous SOTA methods [2, 4, 5], reduces ringing artifacts on the vibrating fork and is better compatibility with camera shake, thus avoiding the occurrence of flickering artifacts.

“Bottle” Video. The bottle in the video moves forward along the horizontal direction, which makes the bottle twisted, artifacts and distortion in the SOTA methods [2, 4, 5] in the dynamic inference mode and a magnification factor α of 10. There are especially serious ringing artifacts in the result of MDLMM [5]. In our result, the magnified result of the bottle has better spatial consistency and a more stable background in the video.

“Eye” Video. The eye in the video exhibits a turning pupil and numerous details, such as blood and eyelashes. In dynamic inference mode and with the magnification factor α of 15, previous SOTA methods [2, 4, 5] caused damage to the pupil structure and blurring of eyelash details, while our result preserves a more complete eye structure and clearer details with fewer flickering artifacts and distortion.

3.2. Synthetic Examples

The manuscript has provided sufficient quantitative experiments on the Real-world Dataset. Here, we supplement some Synthetic examples to provide an intuitive analysis of the magnified results of the synthetic videos. We validate the effectiveness of our proposed FD4MM at magnification factor α and noise level σ. As shown in Fig. A2 (b), the reference video is selected from the Synthetic Dataset with a skateboard image background from the DIS5K dataset [3] and the foreground of the white ball from the public StickPNG library. Please refer to FD4MM_demo.mp4 in the supplement folder for more detailed information.

Magnification Factor α Test. As the magnification factor α increases, achieving accurate magnification becomes more challenging. As shown in the demonstration video,

[1] https://www.stickpng.com/

Method	Subjective Evaluation	
	Magnified Effect ↑	Magnified Quality ↑
LBVMM [2]	3.65 ± 0.13	3.42 ± 0.14
LNVMM [4]	3.65 ± 0.14	3.40 ± 0.14
MDLMM [5]	3.96 ± 0.11	3.58 ± 0.13
Ours	**4.33 ± 0.12**	**4.25 ± 0.12**

Table A2. **User studies on the Real-world Test Dataset for magnified effects and quality are results of MOS with 95% confidence intervals.**

both our FD4MM and MDLMM method [5] successfully magnify the motion of the ball in the synthetic video at different α. However, it is evident that the results obtained with MDLMM exhibit undesired deformations and artifacts on the moving ball. In contrast, our results not only magnify the motion of the ball accurately but also better preserve the overall structure and details of the ball, reducing information loss caused by magnification.

Noise Level σ Test. When increasing the noise level σ in the video, motion magnification is susceptible to interference from the noise. Even though the MDLMM method [5] achieves magnification, the motion of the ball in the magnified results exhibits more artifacts and distortion when dealing with noise interference. However, our FD4MM demonstrates greater noise robustness in reducing artifacts and distortion caused by noise. Particularly, when the noise level σ is higher, our magnified results exhibit a more pronounced suppression of noise compared to MDLMM, and the motion of the ball is more accurate and complete.

4. Additional Experiments

4.1. User Study

To further understand the real visual effects of our method compared to others, we conduct a user study evaluation on the magnified results of each learning-based method [2, 4, 5] on the Real-world Test Dataset. Specifically, 36 participants were involved in this user study. Regarding the magnified effect, participants were asked, “Is the magnified motion more intense in amplitude?” In terms of magnified quality, participants were also asked, “Has the magnified motion better visual perceptual quality?” Users were required to rate each method on a scale of 5 to 1 (“Excellent,” “Good,” “Fair,” “Poor,” and “Bad”). Table A2 presents the Mean Opinion Scores (MOS) for magnification effect and quality. As the results show, our method significantly outperforms SOTA methods, exhibiting more natural and high-quality magnification.

4.2. Physical Accuracy Analysis

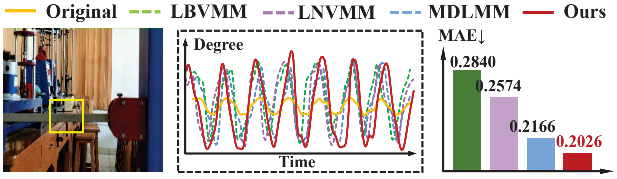


Figure A3. **Physical accuracy experiment for video motion magnification based on physical signal measurements.**

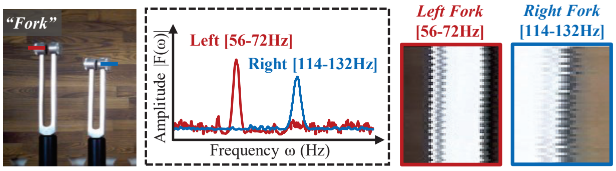


Figure A4. **Compatibility of our method for the temporal filter on motion in different frequency bands.** Taking the Fork video as an example, the left and right forks vibrate at 56-72 Hz and 114-132 Hz, respectively. We select the low-frequency features in different bands by using a temporal band-pass filter to amplify the motion in the interested frequency bands.

We conduct additional physical accuracy experiments, akin to [4, 5], to evaluate the magnification accuracy with different methods, as depicted in Figure A3. In the video, subtle vertical vibrations were generated in a mechanical rod of a universal vibration device. Using motion signals recorded by ultrasonic sensors and cameras as input videos, similar to [4, 5], motion signals were extracted from magnified videos (magnification factor $\alpha = 20$) for comparison with ultrasonic sensor signals, computing the Mean Absolute Error (MAE). Following the methodology of [5], traditional optical flow methods were employed to calculate displacement signals from the videos, and both sensor measurements and computed magnified signals were normalized to a range of 0 to 1. The results of this experiment also demonstrate that our proposed FD4MM exhibits the optimal MAE value.

4.3. Compatibility with Temporal Filters

In this section, we discuss the compatibility issues between our learning-based method and traditional temporal filters. Inspired by [2], we first conduct frequency analysis on the left and right forks in the Fork video to obtain the corresponding motion frequency bands in Figure A4. Subsequently, to ensure compatibility with temporal filters, we used the low-frequency features extracted from the obtained consecutive frames as input for the temporal filter. As illustrated in Figure A4, we found that our FD4MM also accommodates the frequency band selection functionality of temporal filters to amplify the desired motion, but still has the inherent characteristic of smoothing the motion amplitude.

References

[1] Brandon Y Feng, Hadi Alzayer, Michael Rubinstein, William T Freeman, and Jia-Bin Huang. 3d motion mag-

nification: Visualizing subtle motions from time-varying radiance fields. In *ICCV*, pages 9837–9846, 2023. 1

[2] Tae-Hyun Oh, Ronnachai Jaroensri, Changil Kim, Mohamed Elgharib, Fr'edo Durand, William T Freeman, and Wojciech Matusik. Learning-based video motion magnification. In *ECCV*, pages 633–648, 2018. 1, 2, 3

[3] Xuebin Qin, Hang Dai, Xiaobin Hu, Deng-Ping Fan, Ling Shao, and Luc Van Gool. Highly accurate dichotomous image segmentation. In *ECCV*, pages 38–56, 2022. 2

[4] Jasdeep Singh, Subrahmanyam Murala, and G Kosuru. Lightweight network for video motion magnification. In *WACV*, pages 2041–2050, 2023. 1, 2, 3

[5] Jasdeep Singh, Subrahmanyam Murala, and G Kosuru. Multi domain learning for motion magnification. In *CVPR*, pages 13914–13923, 2023. 1, 2, 3

[6] Shoichiro Takeda, Kazuki Okami, Dan Mikami, Megumi Isogai, and Hideaki Kimata. Jerk-aware video acceleration magnification. In *CVPR*, pages 1769–1777, 2018. 1

[7] Shoichiro Takeda, Yasunori Akagi, Kazuki Okami, Megumi Isogai, and Hideaki Kimata. Video magnification in the wild using fractional anisotropy in temporal distribution. In *CVPR*, pages 1614–1622, 2019. 2

[8] Neal Wadhwa, Michael Rubinstein, Frédo Durand, and William T Freeman. Phase-based video motion processing. *ACM TOG*, 32(4):1–10, 2013. 1

[9] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM TOG*, 31(4):1–8, 2012. 1

[10] Yichao Zhang, Silvia L Pintea, and Jan C Van Gemert. Video acceleration magnification. In *CVPR*, pages 529–537, 2017. 1