

Lightweight Network For Video Motion Magnification

Supplementary Material

1 Problem Statement

Wu *et al.* [1] and Wadhwa *et al.* [2] have defined and initiated the problem of motion magnification. Let, an image $I(x, t)$, where x is spatial coordinates and t is time, can be expressed as the function of motion $\delta(x, t) : I(x, t) = f(x + \delta(x, t))$ such that $I(x, 0) = f(x)$. Then the magnified image $\tilde{I}(x, t)$ is defined as

$$\tilde{I}(x, t) = f(x + (1 + \alpha)\delta(x, t)) \quad (1)$$

where, α decides the amount of magnification. The aim is to magnify motion signal, while ignoring the noise. To select the signal of interest $\tilde{\delta}(x, t)$, [3], [2] and [4] used the hand-crafted bandpass filters $B(\cdot)$, such that $\tilde{\delta}(x, t) = B(\delta(x, t))$.

Hand designed filter based methods [3], [2], [4] are unable to ignore noise and other motion signals of large amplitude in the bandpass region. Also, they require the prior information about the signal of interest. By using learnable filters, we can process $\delta(x, t)$ with $\tilde{B}(\cdot)$ signal selector (which learns its properties based on input, output relation). Let, F_t and F_{t-1} are the input images which contain $\delta(x)$ motion signal such that $F_t = F_{t-1}(x + \delta(x))$. As we consider only two frames at a time, in $\delta(x, t)$, t parameter is ignored and the output F_o can be generated using Eq. (2).

$$F_o \approx F_{t-1}(x + (1 + \alpha)\tilde{B}(\delta(x))) \quad (2)$$

Two frame input architecture reduces the network complexity and number of parameters [5].

2 Additional Experiments

When the deep learning method is not trained directly with temporal filters, using temporal filters on intermediate features can produce incorrect results [5]. So to avoid that, video is first pre-processed with temporal filter to suppress unwanted motion. For this, [2] method's output at small magnification factor (magnification factor=4) is given as an input to our method. Visual results are shown in supplementary video. Figure S 1, shows the intermediate features which highlight the motion parts. As different temporal filter inputs features, highlight different motion parts.

Table 1: Parameters used for result generation. All the results are generated with variables and steps given by the respective authors. Source code and pre-trained model are downloaded from their official page, click here [25] [19] [18] [23].

Methods	Video	Magnification Factor	Frequency
Ours (M_1, M_2)	Cat toy	15	N/A
Ours (M_1, M_2)	Gun	15	N/A
Ours (M_1, M_2)	Drill	10	N/A
Ours (M_1, M_2)	Balloon	10	N/A
Ours (M_1, M_2)	baby	20	N/A
Ours (M_1, M_2)	guitar	4	N/A
Ours (M_1, M_2)	Physical Accuracy	10	N/A
Ours(M_1, M_2)	Circle videos with different backgrounds	50	N/A
Oh et al	Cat toy	10	N/A
Oh et al	Gun	10	N/A
Oh et al	Drill	10	N/A
Oh et al	Balloon	10	N/A
Oh et al	baby	20	2.5 Hz
Oh et al	Physical Accuracy	5	N/A
Oh et al	Circle videos with different backgrounds	60	N/A
Jerk-Aware	Cat toy	10	3
Jerk-Aware	Gun	10	20
Jerk-Aware	Drill	25	3
Jerk-Aware	Balloon	25	3
Jerk-Aware	baby	50	2.5
Jerk-Aware	Physical Accuracy	20	15
Jerk-Aware	Circle videos with different backgrounds	200	15
Anisotropy	Cat toy	100	3
Anisotropy	Gun	100	20
Anisotropy	Drill	100	3
Anisotropy	Balloon	100	3
Anisotropy	baby	150	2.5
Anisotropy	Physical Accuracy	200	3
Anisotropy	Circle videos with different backgrounds	400	15
Acceleration	Cat toy	4	3
Acceleration	Gun	10	20
Acceleration	Drill	4	3
Acceleration	Balloon	4	3
Acceleration	baby	100	2.5
Acceleration	Physical Accuracy	20	15
Acceleration	Circle videos with different backgrounds	200	15

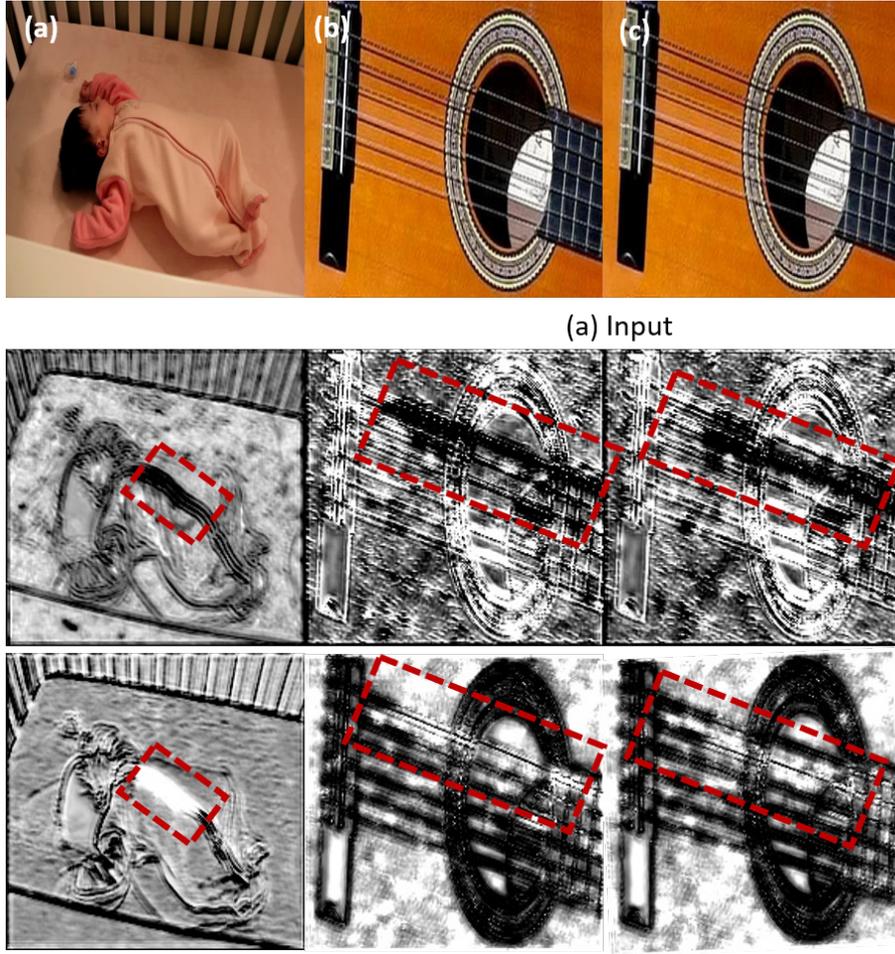


Figure S 1: Intermediate features (*row-2,3*) of our proposed lightweight network, highlighting the motion in the red bounding box. (a) shows the baby video motion features. Baby have minute chest motion while breathing. (b) and (c) highlight the motion part E-string (82Hz), A-string(110Hz) of guitar. Hence, the proposed method is able to capture the minute motions of baby video and temporal filter guitar video.

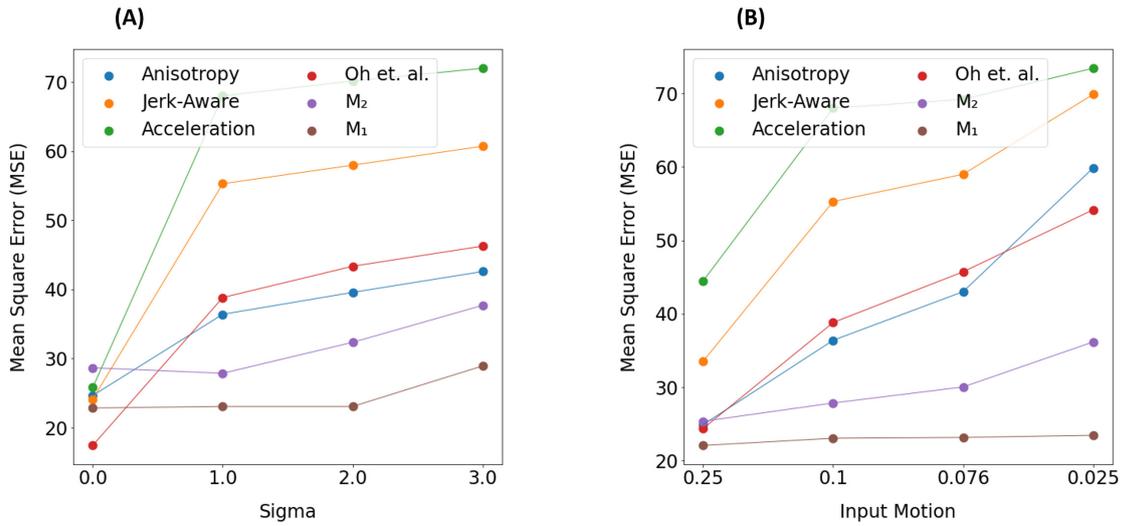


Figure S 2: **Noise Test and Sub-pixel Motion Test.**(A) shows the variation across increase of noise value (sigma) in input (noise test). (B) shows the effects of decrease in input motion (sub-pixel motion test). Magnification factor is changed such that it produces the same amount of magnification. In both cases, average mean square error (MSE) is computed across the predicted output and ground truth, over 25 different videos. Comparison is done with Anisotropy method, Jerk-aware method, Acceleration method, Oh *et al.* method, the proposed base model M_1 and lightweight model M_2 .

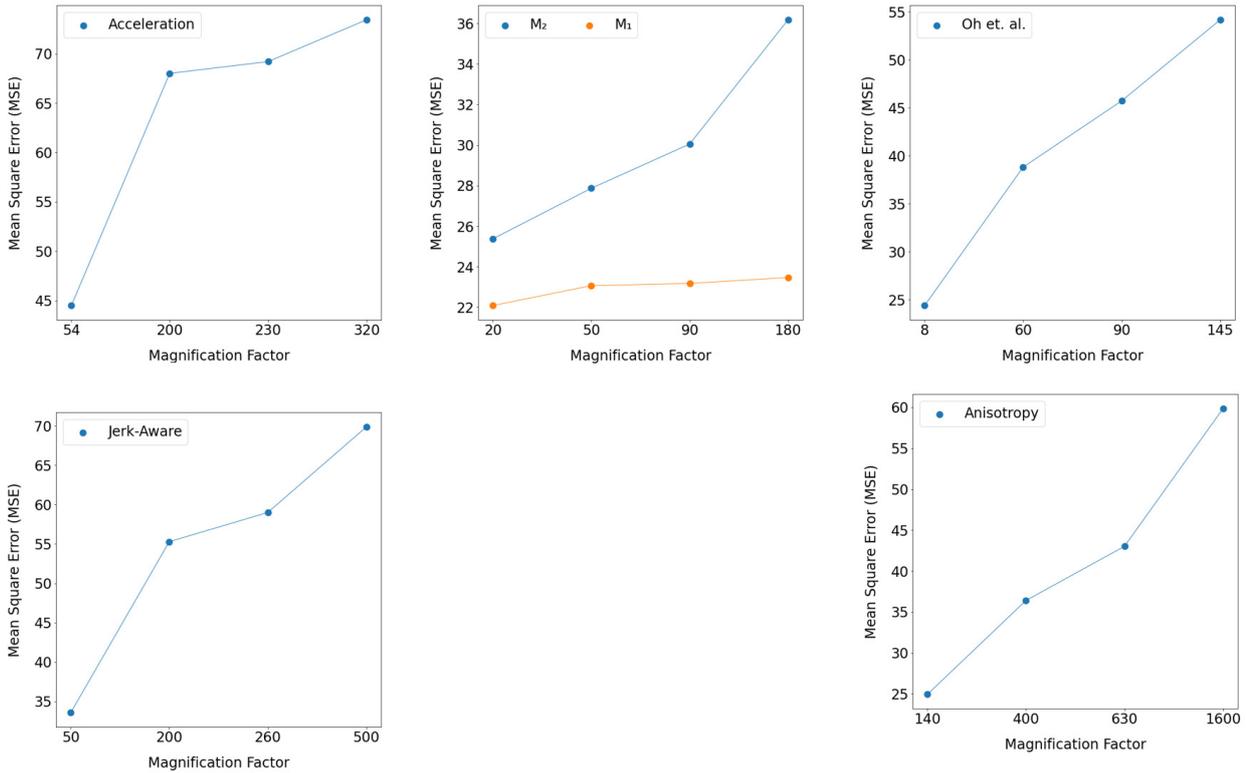


Figure S 3: **Effects of change in Magnification Factor** on Acceleration based method [4], Jerk-Aware method [3], Anisotropy method [6], Oh *et al.* [5], Ours Base model (M_1), and Ours lightweight model (M_2). The magnification factor, has different meaning in each respective method [5]. The values of magnification factor is chosen such that, they produce same amount of output motion with different input motion (the respective input motion values with output MSE are shown in Figure S 2 (B)). Average mean square error (MSE) is computed across the predicted output and ground truth, over 25 different videos.

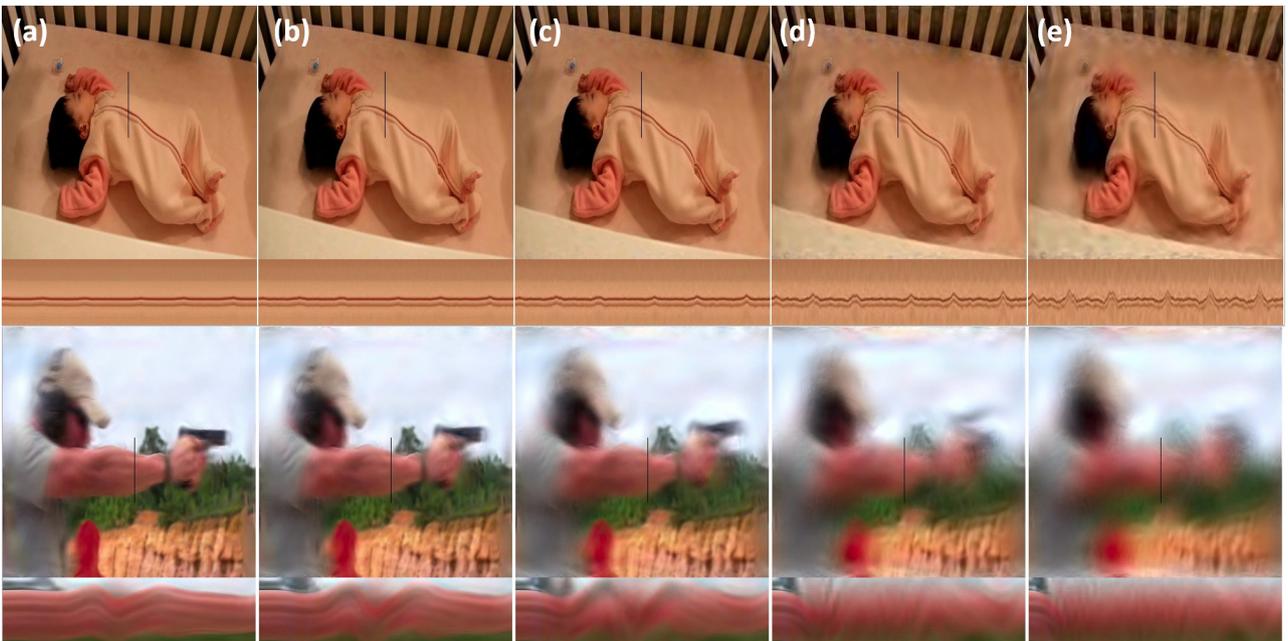


Figure S 4: **Effects of change in Magnification Factor.** Figure illustrates Acceleration method [4] output. Different values of magnification factor in increasing order from (a) 20, (b) 40, (c) 60, (d) 100, and (e) 200 are used to generate output shown in respective column. As, increase in magnification factor leads to more increase in distortions than small increment in magnification, especially in dynamic scenarios.

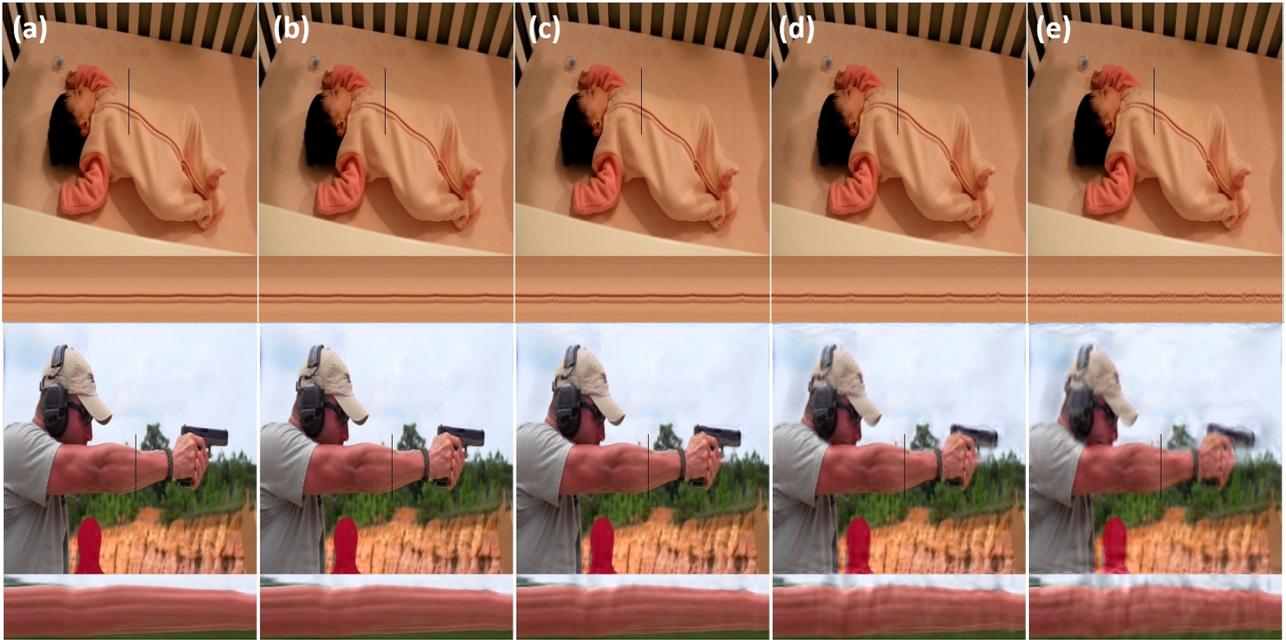


Figure S 5: **Effects of change in Magnification Factor.** Figure illustrates Anisotropy method [6] output. For columns, (a) 50, (b) 100, (c) 200, (d) 500, and (e) 1000 , respective magnification factor values are used to generate magnified output. As visible from the figure, with increase in magnification factor, there are minute changes in magnification while increments in distortions (especially in dynamic scenarios).

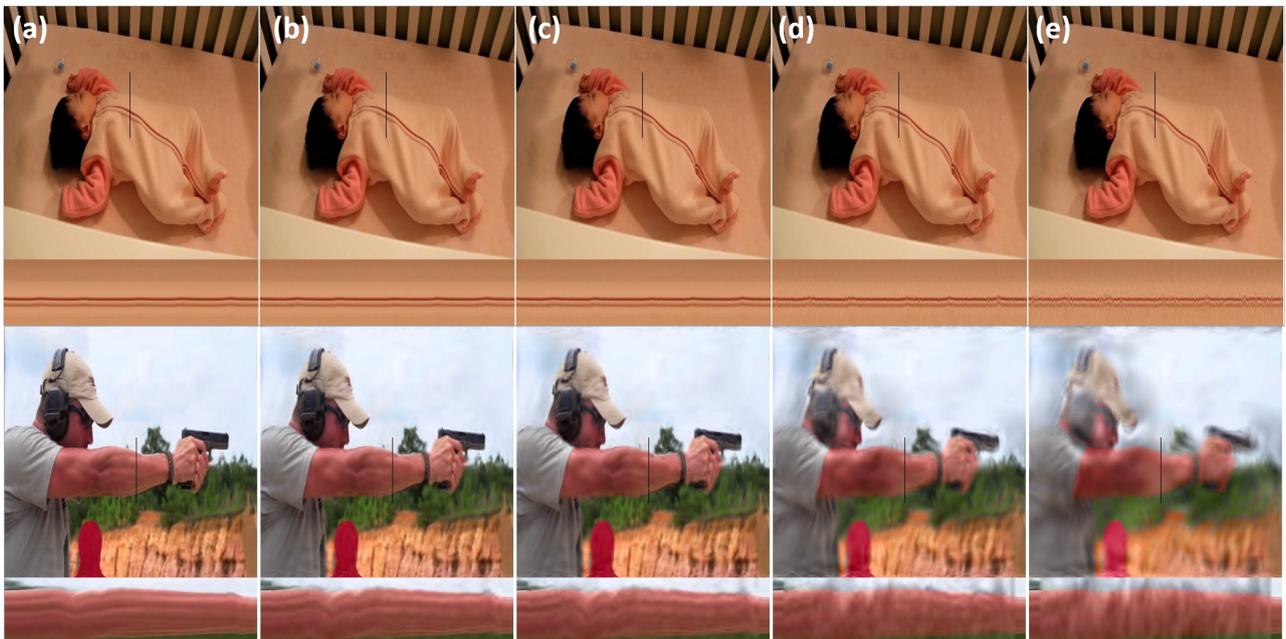


Figure S 6: **Effects of change in Magnification Factor.** Figure illustrates Jerk-Aware method [3]. Different values of magnification factor in increasing order from (a) 10, (b) 30, (c) 50, (d) 150, and (e) 400 are used to generate output shown in respective column. As visible from the figure, with increase in magnification factor, there are minute changes in magnification while increments in distortions (especially in dynamic scenarios).

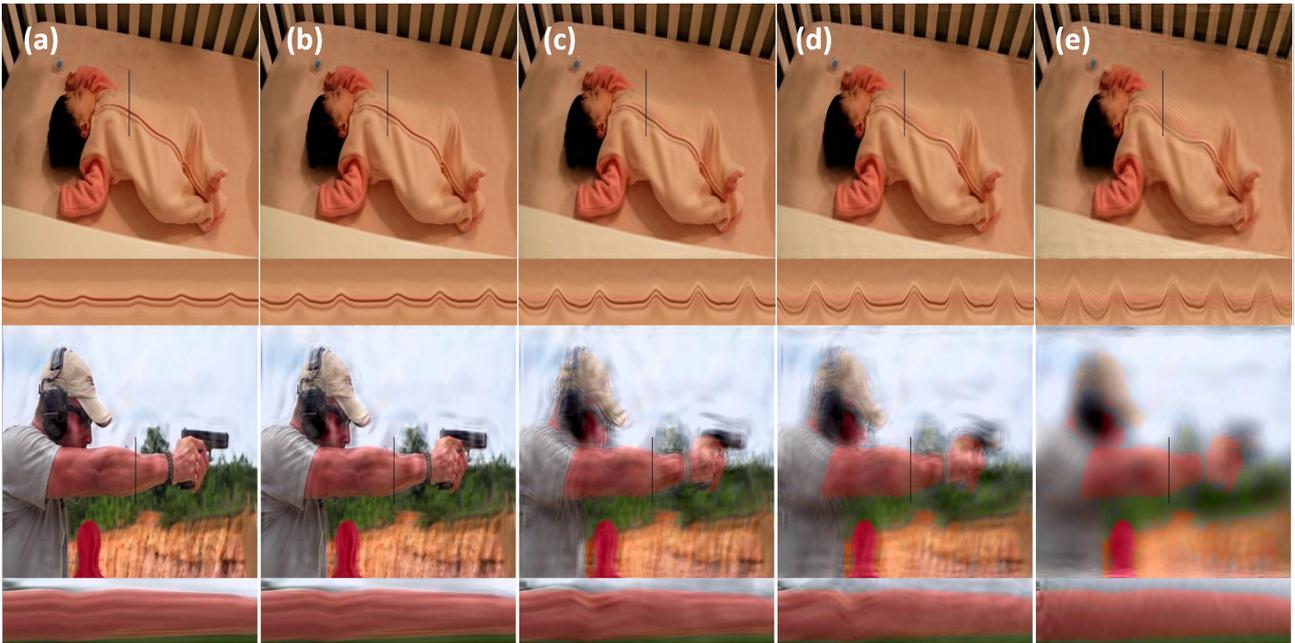


Figure S 7: **Effects of change in Magnification Factor.** Figure illustrates Phase based method [2] output. Different values of magnification factor in increasing order from (a) 10, (b) 20, (c) 40, (d) 60, and (e) 100 for baby video and (a) 1, (b) 2, (c) 5, (d) 10, and (e) 100 for gun video are used to generate output shown in respective column (note:- for other methods same values are used for both the videos). The linear methods, are not suitable for dynamic scenarios, as they are unable to ignore dynamic motion. So, they produces large distortions in gun video (dynamic scenarios). Where as in static scenario (baby videos), with increase in magnification factor their is increment in both, amount of magnification and ringing artifacts (visible as lines overlapping the edges of motion objects) in static scenario.

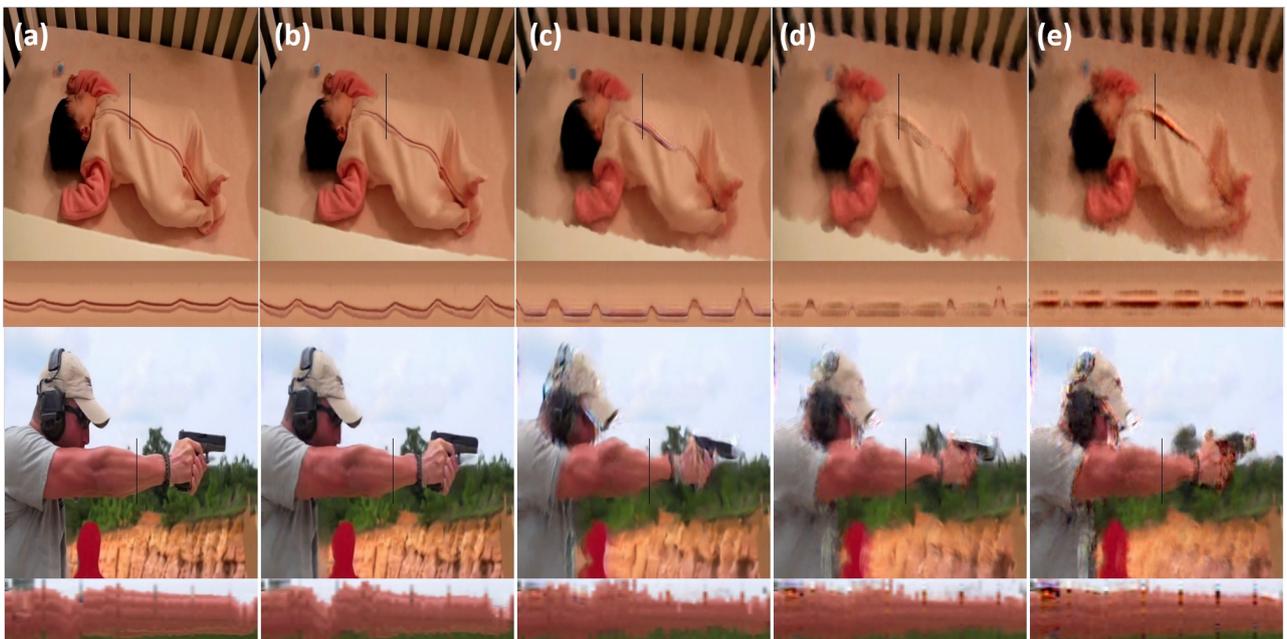


Figure S 8: **Effects of change in Magnification Factor.** Figure illustrates Oh *et al.* method [5] output. Different values of magnification factor in increasing order from (a) 10, (b) 20, (c) 50 (d) 100 and (e) 200 are used to generate output shown in respective column. It produces more magnification, (both in static and dynamic scenarios), but it also produces some unwanted motion (visible as large spikes in the temporal slice) and blurry distortions in the video. Distortions are increased with inclemently in the magnification factor.

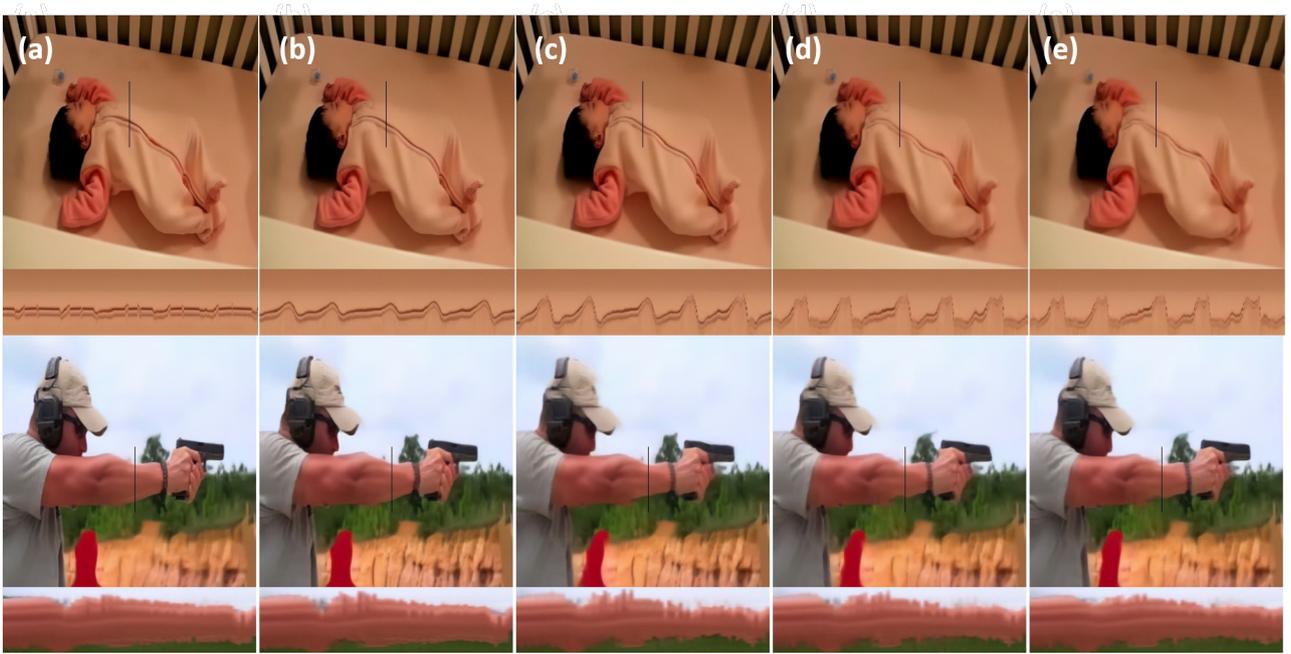


Figure S 9: **Effects of change in Magnification Factor.** Figure illustrates ours Base model (M_1) output. Different values of magnification factor in increasing order from (a) 10, (b) 20, (c) 50 (d) 100 and (e) 200 are used to generate output shown in respective column. M_1 shows fewer distortions while increasing the amount of magnification as compare to other SOTA methods, both in static and dynamic scenario.

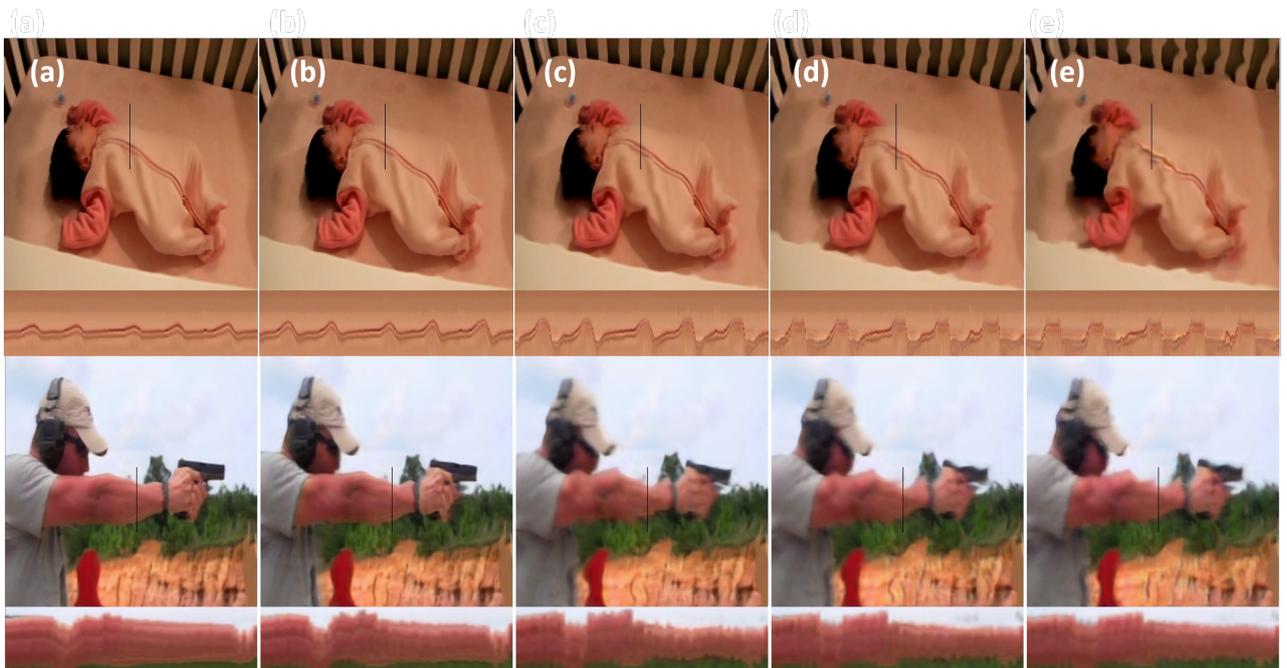


Figure S 10: **Effects of change in Magnification Factor.** Figure illustrates ours lightweight model (M_2) output. Different values of magnification factor in increasing order from (a) 10, (b) 20, (c) 50 (d) 100 and (e) 200 are used to generate output shown in respective column. M_2 also shows good amount of magnification, but with increase in magnification factor its performance degrades as compare to M_1 . This is expected as M_2 has much fewer parameters than M_1 , so their performance gap between become observable at extreme scenarios.

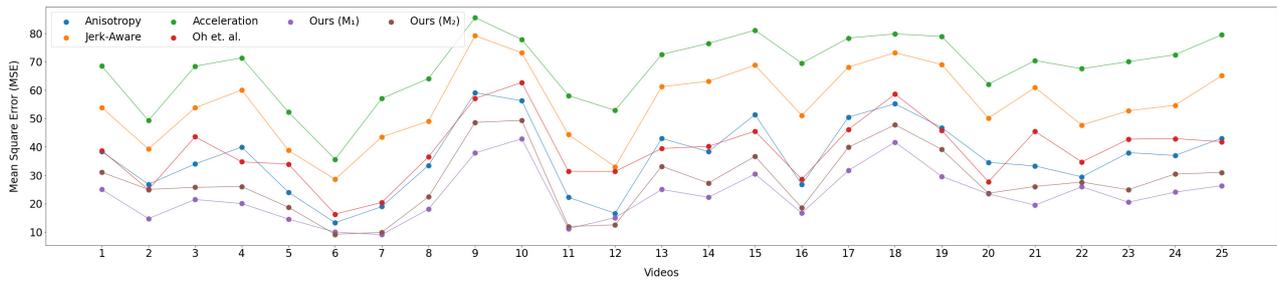


Figure S 11: Mean Square Error (MSE) of Anisotropy method, Jerk-aware method, Acceleration method, Oh *et al.* method, and the proposed methods on 25 synthetically generated videos containing different subtle motion of circles with various backgrounds.



Figure S 12: Different backgrounds used for generation of different synthetic videos for quantitative analysis. Video 1-10



Figure S 13: Different backgrounds used for generation of different synthetic videos for quantitative analysis. Video 11-20.



Figure S 14: Different backgrounds used for generation of different synthetic videos for quantitative analysis. Video 20-25.

References

- [1] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, “Eulerian video magnification for revealing subtle changes in the world,” *ACM transactions on graphics (TOG)*, vol. 31, no. 4, pp. 1–8, 2012.
- [2] N. Wadhwa, M. Rubinstein, F. Durand, and W. T. Freeman, “Phase-based video motion processing,” *ACM Transactions on Graphics (TOG)*, vol. 32, no. 4, pp. 1–10, 2013.
- [3] S. Takeda, K. Okami, D. Mikami, M. Isogai, and H. Kimata, “Jerk-aware video acceleration magnification,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1769–1777.
- [4] Y. Zhang, S. L. Pinteá, and J. C. Van Gemert, “Video acceleration magnification,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 529–537.
- [5] T.-H. Oh, R. Jaroensri, C. Kim, M. Elgharib, F. Durand, W. T. Freeman, and W. Matusik, “Learning-based video motion magnification,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 633–648.
- [6] S. Takeda, Y. Akagi, K. Okami, M. Isogai, and H. Kimata, “Video magnification in the wild using fractional anisotropy in temporal distribution,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1614–1622.