Blind Restoration of High-Resolution Ultrasound Video

Chu Chen[1,2✉], Kangning Cui[1,2], Pasquale Cascarano[3], Wei Tang[1,2], Elena Loli Piccolomini[4], and Raymond H. Chan[2,5]

[1] Department of Mathematics, City University of Hong Kong, Hong Kong
[2] Hong Kong Centre for Cerebro-cardiovascular Health Engineering, Hong Kong
[3] Department of the Arts, University of Bologna, Italy
[4] Department of Computer Science and Engineering, University of Bologna, Italy
[5] Department of Operations and Risk Management and School of Data Science, Lingnan University, Hong Kong
chuchen4-c@my.cityu.edu.hk

Abstract. Ultrasound imaging is widely applied in clinical practice, yet ultrasound videos often suffer from low signal-to-noise ratios (SNR) and limited resolutions, posing challenges for diagnosis and analysis. Variations in equipment and acquisition settings can further exacerbate differences in data distribution and noise levels, reducing the generalizability of pre-trained models. This work presents a self-supervised ultrasound video super-resolution algorithm called Deep Ultrasound Prior (DUP). DUP employs a video-adaptive optimization process of a neural network that enhances the resolution of given ultrasound videos without requiring paired training data while simultaneously removing noise. Quantitative and visual evaluations demonstrate that DUP outperforms existing super-resolution algorithms, leading to substantial improvements for downstream applications.

Keywords: Ultrasound · Video Super-resolution · Deep Image Prior · Self-supervised Learning · Ejection Fraction.

1 Introduction

Ultrasound (US) imaging is an essential diagnostic tool in modern medicine, widely utilized in various clinical applications, including cardiology [3,22], obstetrics [14,2], and musculoskeletal [21] imaging. Its ability to provide visualization of internal structures, coupled with its non-invasive nature, makes US a preferred imaging technique. Furthermore, the dynamic recording of US videos captures much more information about tissues than single images, making diagnosis heavily reliant on video analysis. Applications include breast lesion detection [28,31] and cardiovascular monitoring [11,9,24], where metrics like Ejection Fraction (EF) play a critical role. However, the quality of US data often suffers from both inherent and technical limitations. Acoustic noise, caused by scattering and absorption of sound waves in complex tissue structures, degrades image clarity,

while limited resolution, stemming from constraints in transducer technology and frequency ranges, affects the ability to capture fine anatomical details. These challenges not only affect diagnostic accuracy but also impact downstream tasks like EF prediction, where precise visualization of cardiac structures and motion is essential. Addressing these limitations requires continuous advancements in US technology to enhance visual quality and diagnostic reliability through super-resolution (SR) algorithms.

Related Works. The development of large training datasets has significantly advanced the research on SR algorithms for natural images [1] and videos [19,16]. These datasets enable supervised training of neural networks, resulting in models that achieve exceptional performance. Among all, deep convolutional neural networks (CNN) [15,30] have shown remarkable capabilities in single image SR (SISR) tasks. To ensure temporal correlation in video enhancement, techniques such as the flow-based [27] and attention-based method [32] are integrated with CNNs. However, significant shifts in data distributions make these pre-trained models struggle with generalization. Consequently, artifacts are often introduced in the upsampled videos.

In the field of US image SR, acquiring paired low-resolution (LR) and high-resolution (HR) datasets is particularly challenging, leading researchers to either create simulated datasets for supervised training or explore unsupervised/self-supervised approaches. Choi et al. [8] modified the SRGAN [12] to enhance B-mode US images with low lateral resolution, improving their similarity to HR images. A U-net-style network [26] is trained on simulated pairs for super-resolved US images. Liu et al. [17] proposed ZSSR-Cycle, a zero-shot and self-supervised generative adversarial network framework for US image SR.

One particularly intriguing development in image processing is the Deep Image Prior (DIP) method [25], which demonstrates that CNN structures can inherently capture and enhance image features without relying on any predefined training dataset. In fact, it has been shown that CNNs can more effectively replicate and enhance the structural characteristics present in images compared to arbitrary noise patterns [6]. In the field of SISR, DIP has achieved competitive performance compared to other supervised training models. Variations of DIP have also been proposed for video processing. The Deep Video Prior (DVP) [13] adopted DIP for individual video frame with iteratively reweighted training strategy to address the multi-modal inconsistencies. The Recursive Deep Prior Video (RDPV) [5] proposed a recursive updating rule for CNN optimization when dealing with each frame for the light Time-Lapse Microscopy video SR. While DIP-based methods effectively capture underlying image structures, their ability to fit fine details can also lead to overfitting, where noise and artifacts are misinterpreted as part of the high-frequency details that SR methods seek to enhance. These unwanted patterns adversely affect diagnostic and analytical applications based on the SR-enhanced US videos.

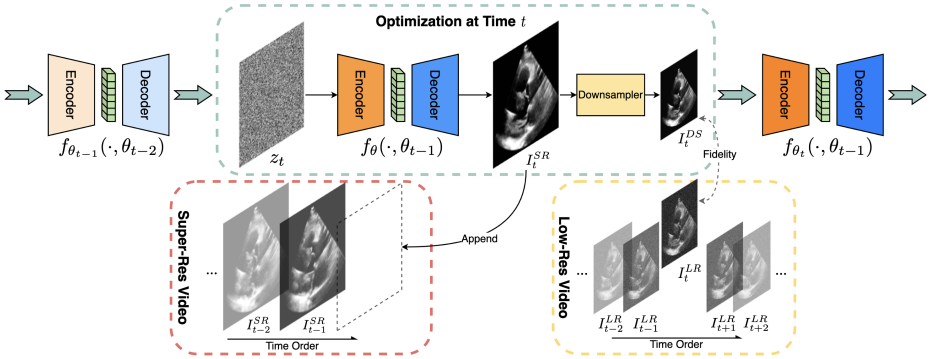

Fig. 1: Schematic of the DUP algorithm.

Contributions. US video restoration faces fundamental challenges from lacking paired data, heterogeneous distributions, and multi-factorial corruptions. We address these through a blind restoration method followed by extensive evaluations. The key contributions of this work are as follows:

- We present a self-supervised framework Deep Ultrasound Prior (DUP) for US video SR, eliminating the need for paired training data through video-specific optimization.
- DUP adopts a Weight Inheritance (WIn) strategy and dual regularization for CNN optimization, accelerating convergence and achieving information sharing among successive frames while removing noise and artifacts that appeared in US videos.
- We systematically compare DUP with existing SR techniques under various levels of noise, which demonstrates the superior performance and robustness of our approach in US video restoration.
- By evaluating EF prediction from restored cardiac videos, DUP achieves the lowest error, highlighting its potential to improve clinical diagnostic accuracy.

2 Method

2.1 Problem Formulation

The inverse problem of frame-wise video SR can be formulated as

$$I_t^{LR} = \mathcal{A} I_t^{HR} + e_t, \quad t = 1, 2, \ldots, T, \tag{1}$$

where $\mathcal{A}$ represents the measurement operator, $I_t^{LR} \in \mathbb{R}^{MN}$ is the vectorized tth LR video frame with spatial resolution of $M \times N$, T is the number of frames, $I_t^{HR} \in \mathbb{R}^{s^2 MN}$ is the underlying HR frame with upscaling factor $s \in \mathbb{Z}^+$, and $e_t \in \mathbb{R}^{MN}$ is the measurement error. In this work, we focus on the linear degradation

operator $\mathcal{A} \in \mathbb{R}^{MN \times s^2 MN}$. Since the SR problem is known to be strongly ill-posed [29], the reconstructed SR frame I_t^{SR} that solves this system is not unique while sensitive to the measurement error e_t. The common approach to address this problem is to add prior constraints on the desired I_t^{SR}, narrowing the space of the possible solution. We formulate the underdetermined systems in Eq. (1) into a set of variational regularized optimization problems:

$$I_t^{SR} = \underset{I_t}{\operatorname{argmin}} \frac{1}{2}||I_t^{LR} - \mathcal{A}I_t||_2^2 + \lambda R(I_t), \quad t = 1, 2, \ldots, T, \tag{2}$$

where $\{I_t\}_{t=1}^T$ are variables. In Eq. (2), the SR video $\{I_t^{SR}\}_{t=1}^T$ is estimated by minimizing the sum of the ℓ_2 fidelity term and the regularization term R for each frame, where λ is the regularization parameter that balances the two terms.

2.2 Deep Ultrasound Prior

The ability to reconstruct and augment detail highlights CNNs' potential for self-supervised learning techniques, improving the spatial resolution and clarity of US video. We now explain our proposed DUP method in detail.

This algorithm produces super-resolved US video in a frame-by-frame manner following the time order. Let integer $t \in [1, T]$ be an index of the video frames and $z_t \in \mathbb{R}^{s^2 MN}$ be a random image with targeted spatial resolution. DUP utilizes an encoder-decoder CNN treated as a mapping from the random image z_t to SR frame $I_t^{SR} = f_\theta(z_t, \omega) : \mathbb{R}^{s^2 MN} \times \mathbb{R}^{\Theta} \to \mathbb{R}^{s^2 MN}$, where the sub-index θ is the trainable parameters of the CNN and Θ is the total number of parameters, the second entry $\omega \in \mathbb{R}^{\Theta}$ defined initialization of CNN.

During the reconstruction process of SR images, we aim to enhance only the contrast details while preventing the amplification of random noise and artifacts and even removing them. A widely recognized approach for achieving this is the total variation (TV) model [23], which replaces the regularization term R in Eq. (2) with a TV regularization term R_1:

$$R_1(I_t) = \sum_{i=1}^{s^2 MN} (|(D_h I_t)_i| + |(D_v I_t)_i|) \tag{3}$$

where D_h and D_v are the first-order finite difference discrete operators along the horizontal and vertical directions, respectively.

However, optimizing with only total variation regularization can result in staircase effects appearing in the image. To address the piece-wise constant artifacts, we additionally introduce a higher-order (HO) [7,4] term R_2:

$$R_2(I_t) = \frac{1}{2} \sum_{i=1}^{s^2 MN} (|(D_h I_t)_i|^2 + |(D_v I_t)_i|^2). \tag{4}$$

Hence, the optimization problem in Eq. (2) is solved by optimizing the parameters of the CNN as

$$\theta_t = \underset{\theta}{\operatorname{argmin}} \frac{1}{2}||I_t^{LR} - \mathcal{A}f_\theta(z_t, \omega)||_2^2 + \lambda_1 R_1(f_\theta(z_t, \omega)) + \lambda_2 R_2(f_\theta(z_t, \omega)), \tag{5}$$

thereby the SR frames can be reconstructed by $I_t^{SR} = f_{\theta_t}(z_t, \omega)$.

The overview of the proposed DUP framework is shown in Fig. 1. We now explain in detail the technical advancement of DUP.

WIn Strategy For any frame-by-frame video enhancement algorithm, leveraging temporal correlation is essential for maintaining high-quality results. DUP employs a Weight Inheritance (WIn) strategy to utilize the continuity of adjacent frames for improved efficiency, allowing the CNN to converge faster while accumulating optimization iterations. The WIn strategy in DUP operates as follows: 1. The CNN is randomly initialized with parameter θ_0 as processing the first video frame. After optimization based on the first low resolution frame I_1^{LR}, the optimal parameter of the network is θ_1 and corresponding super-resolved image $I_1^{SR} = f_{\theta_1}(z_1, \theta_0)$; 2. For subsequent frames I_t^{LR} $(t \geq 2)$, CNN is initialized with the optimized parameters θ_{t-1} from the preceding frame I_{t-1}^{LR}, and the super-resolved image is generalized by $I_t^{SR} = f_{\theta_t}(z_t, \theta_{t-1})$. This WIn strategy enhances the computational efficiency in frame-by-frame video processing. By sharing network parameters, CNNs utilize information from neighboring frames, exploiting the temporal correlations and achieving faster convergence within a limited number of iterations. Furthermore, continuous weight inheritance means the CNN effectively trains on more samples as it processes additional frames, leading to better SR performance through extended iterations.

Early Stopping and Input Update Since DUP follows a self-supervised learning strategy, there is no need to consider generalization beyond the given video. However, excessive optimization can introduce unwanted artifacts into the image. Specifically, over-optimizing the fidelity term may lead to overfitting measurement errors, whereas excessive optimization of the regularization term can cause over-smoothing. To mitigate these issues, the iterative process is early-stopped based on a criterion that monitors the loss over a fixed window. Training stops when the loss value stabilizes. With the WIn strategy, the CNN typically requires fewer iterations after processing the first frame, allowing DUP to begin monitoring the loss after fewer training steps.

Furthermore, according to DIP, the input to the CNN is a randomly generated noise image z_1. Due to the strong similarities and periodic patterns exhibited among frames in a US video, using the same noise image as input for every frame (i.e., $z_t = z_1, t \geq 2$) may cause the CNN to undergo lazy training, leading to outputs that remain overly similar to the first frame. In response, DUP updates the noise image for each frame as follows: 1. $z_1 = n_1$; 2. $z_t = z_{t-1} + \sigma n_t, (t \geq 2)$, where each pixel of the random image $\{n_t\}_{t=1}^{T} \in \mathbb{R}^{s^2MN}$ is $i.i.d$ sampling from $\mathcal{N}(0, 1)$, and σ is the standard deviation of the added noise in each update. This formulation allows the CNN to learn a mapping from the random noise input to the residual between successive frames.

Further implementation details are provided in the section 3.2.

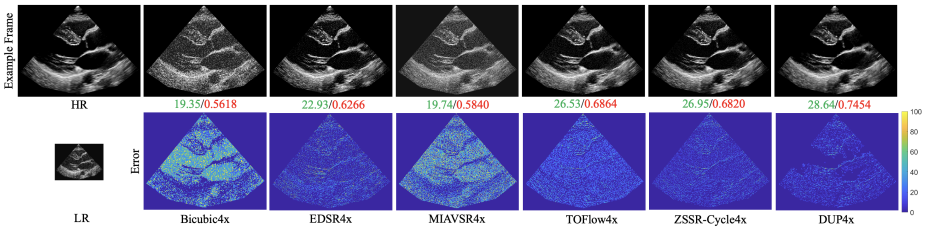

Fig. 2: Visual comparisons of 4× SR results. The paired HR and LR example frames are shown on the far left. The upper row displays the SR frames from various methods, with quantitative results (PSNR (dB)/SSIM) shown underneath, while the lower row is the corresponding error maps.

3 Experiments

3.1 Datasets

Our validation combines video quality assessment and downstream tasks using two datasets: 1) EchoNet-LVH [10] (1024×768 resolution, 90-200 frames/clip) for restoration evaluation. We simulate LR scenarios through normalization and downsampling with Gaussian noise of various standard deviations (std 0.01-0.1), comparing outputs against original HR videos. 2) EchoNet-Dynamic [20] (112×112 resolution, 28-1002 frames/clip) for application testing. We randomly selected 100 videos from each dataset for the respective evaluation scenarios.

3.2 Implement Details

In this part, we explain the implementation of our method and experiments. A Lanczos kernel is used for performing the degradation process $\mathcal{A}$, which aligns with the real-world US image degradation characteristics [18]. The CNN is an Encoder-Decoder-style architecture featuring skip concatenation. The architecture consists of four encoder and decoder base units each, which perform convolutions using 128 feature maps, along with batch normalization layers and Leaky ReLU activations. Downsampling is achieved via convolutional layers with a stride of two, while up-sampling is by a Lanczos operator. For the first video frame, we allow up to 3000 iterations, implementing early stopping after 2000 iterations with a patience of 100. For subsequent frames, we reduce the maximum to 2000 iterations, with early stopping commencing at 1000 iterations and patience of 50. σ is set to 0.03 and $\lambda_1 = \lambda_2 = 0.01$ in Eq. (5). For the downstream task, we employ EchoCLIP [9], a CLIP-like vision-language model trained on medical reports paired with US videos. The restored videos (each frame resized to 224×224) generated by different SR methods are fed into EchoCLIP to perform zero-shot EF prediction. All experiments were conducted in Python 3.8.17 on a PC equipped with Intel® Xeon® Silver 4210 Processor CPU 2.20GHz and Nvidia GeForce RTX 3090 GPU with 24G of memory.

Table 1: Ablation studies.

Method	*HO*	*TV*	PSNR (dB)↑	SSIM↑
DIP (w/o WIn)	✗	✗	25.53	0.7133
	✓	✗	27.16	0.7432
	✗	✓	28.49	0.7432
	✓	✓	28.63	0.7440
DUP	✗	✗	25.08	0.7069
	✓	✗	26.85	0.7409
	✗	✓	28.85	0.7722
	✓	✓	**28.99**	**0.7740**

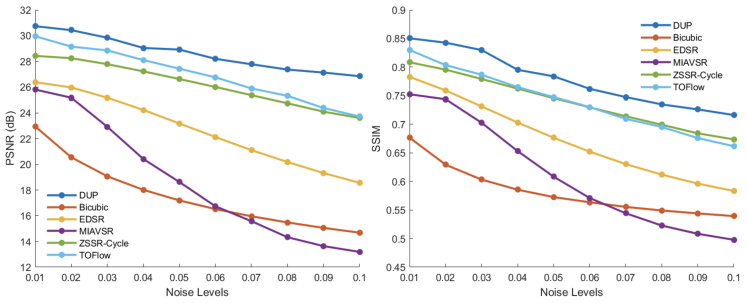

Fig. 3: Quantitative evaluations on 4×SR videos corrupted by multi-level noise.

3.3 Image-based Evaluations

We compared DUP against four categories of methods: (1) Bicubic interpolation, (2) SISR learning-based architectures (EDSR [15] and DIP [25]), (3) video SR-based methods (TOFlow [27] and MIAVSR [32]), and (4) SR framework for US (ZSSR-Cycle [17]). The visual comparison results in Fig. 2, based on an LR video with 0.05 std of additive noise, reveal that DUP notably outperforms other methods, specifically in reconstructing clear textures of the ventricles with significantly lower error. In contrast, the results from other methods exhibit noticeable errors after restoration, compromising image clarity. This visual superiority aligns with the quantitative metrics, where DUP attains the highest peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) values. To evaluate the robustness against noise as demonstrated in Fig. 3, DUP consistently achieves the best reconstruction performance under different levels of noise corruption, while other baselines degrade rapidly. From both quantitative and qualitative aspects, DUP yields a superior result with rich details and less noise. In order to analyze the role of regularizations and WIn, we conduct ablation studies on these setups. As listed in Table 1, the best SNR and structural detail are achieved when both HO and TV priors are incorporated into the target loss, regardless of the presence of WIn. Without WIn, DUP defaults to the same training strategy as DIP, resulting in more iterations(3000 iters/frame) and longer optimization

Table 2: EF Prediction Error

SR Method	Bicubic	EDSR	RDN	ZSSR-Cycle	DUP
Mean Error (%) ↓	10.26	8.69	8.46	7.60	**7.23**

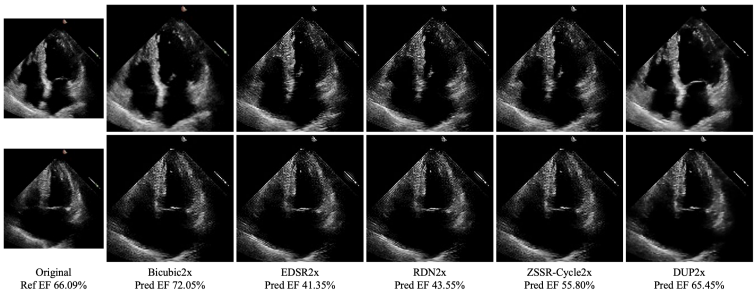

Fig. 4: Visualization of maximum diastole (upper) and minimum systole (lower) detected within cardiac cycles.

time. Despite this, DUP with the WIn strategy outperforms and achieves greater efficiency.

3.4 Downstream Evaluations

To evaluate the practical utility of our proposed method, we assess its performance on the downstream task of EF prediction. Since the EF prediction model only accepts 2× SR results and both MIAVSR and TOFlow are limited to 4×, we substituted these two baselines with RDN [30] for this part of the study. Table 2 compares the mean prediction error of EF across different SR preprocessing methods, where DUP achieves the lowest mean error of 7.23%. This demonstrates the potential of DUP in enhancing the quality of US videos for diagnostic tasks. Fig. 4 visualizes a sample EF prediction result with detected minimum systole and maximum diastole frames within cardiac cycles. EchoCLIP identified frames corresponding to the two heart phases that align with the manually annotated ground truth from the DUP SR results, while those extracted from other SR videos displayed discrepancies. Consequently, the improvement in visual quality directly correlates with the lower error in EF prediction observed in Fig. 4. The sample further validates the practical applicability of DUP in clinical settings.

4 Conclusion

In this work, we propose a self-supervised learning method, Deep Ultrasound Prior (DUP), for US video restoration, eliminating the need for HR video supervision. DUP increases spatial resolution while using CNNs' implicit regularization to capture detailed features accurately. By incorporating the WIn strategy

with an early stopping and input update mechanism, the network can perceive and share local frame information while improving convergence speed and performance. With the addition of TV regularization and HO terms, DUP effectively eliminates the measurement error (e.g., noise and artifacts) commonly found in US videos. We demonstrate the superiority of DUP by comparing the quality of super-resolved US videos and their robustness against noise with state-of-the-art methods. Ablation studies confirm the necessity of the WIn strategy and explicit regularization terms for effective video restoration. Furthermore, we validate DUP's capabilities as a pre-processing method in downstream tasks, showing that the restored videos enhance the accuracy of EF predictions for cardiac. This work establishes DUP as a comprehensive solution for US video restoration and overcoming the bottlenecks of its clinical applications.

Acknowledgments. This work is partially supported by HKRGC (grant number CityU11301120, C1013-21GF, CityU11309922, CityU9380162), ITF (grant number LU BGR 105824, MHP/054/22), and the InnoHK initiative of the Innovation and Technology Commission of the Hong Kong Special Administrative Region Government.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 126–135 (2017)
2. Alfirevic, Z., Stampalija, T., Medley, N.: Fetal and umbilical doppler ultrasound in normal pregnancy. Cochrane Database of Systematic Reviews (2015)
3. Aly, I., Rizvi, A., Roberts, W., Khalid, S., Kassem, M.W., Salandy, S., du Plessis, M., Tubbs, R.S., Loukas, M.: Cardiac ultrasound: an anatomical and clinical review. Translational Research in Anatomy **22**, 100083 (2021)
4. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. SIAM Journal on Imaging Sciences **3**(3), 492–526 (2010)
5. Cascarano, P., Comes, M.C., Mencattini, A., Parrini, M.C., Piccolomini, E.L., Martinelli, E.: Recursive deep prior video: a super resolution algorithm for time-lapse microscopy of organ-on-chip experiments. Medical Image Analysis **72**, 102124 (2021)
6. Cascarano, P., Franchini, G., Porta, F., Sebastiani, A.: On the first-order optimization methods in deep image prior. Journal of Verification, Validation and Uncertainty Quantification **7**(4), 041002 (2022)
7. Chan, T., Marquina, A., Mulet, P.: High-order total variation-based image restoration. SIAM Journal on Scientific Computing **22**(2), 503–516 (2000)
8. Choi, W., Kim, M., HakLee, J., Kim, J., BeomRa, J.: Deep cnn-based ultrasound super-resolution for high-speed high-resolution b-mode imaging. In: 2018 IEEE International Ultrasonics Symposium (IUS). pp. 1–4 (2018)
9. Christensen, M., Vukadinovic, M., Yuan, N., Ouyang, D.: Vision–language foundation model for echocardiogram interpretation. Nature Medicine pp. 1–8 (2024)

10. Duffy, G., Cheng, P., Yuan, N., He, B., Kwan, A., Shun-Shin, M., Alexander, K., Ebinger, J., Lungren, M., Rader, F., Schnittger, I., Ashley, E., Zou, J., Patel, J., Witteles, R., Cheng, S., Ouyang, D.: High-throughput precision phenotyping of left ventricular hypertrophy with cardiovascular deep learning. JAMA Cardiology **7**(4), 386–395 (2022)
11. Duffy, G., Cheng, P.P., Yuan, N., He, B., Kwan, A.C., Shun-Shin, M.J., Alexander, K.M., Ebinger, J., Lungren, M.P., Rader, F., et al.: High-throughput precision phenotyping of left ventricular hypertrophy with cardiovascular deep learning. JAMA Cardiology **7**(4), 386–395 (2022)
12. Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., Shi, W.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (July 2017)
13. Lei, C., Xing, Y., Ouyang, H., Chen, Q.: Deep video prior for video consistency and propagation. IEEE Transactions on Pattern Analysis and Machine Intelligence **45**(1), 356–371 (2022)
14. Leung, K.Y.: Applications of advanced ultrasound technology in obstetrics. Diagnostics **11**(7), 1217 (2021)
15. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 136–144 (2017)
16. Liu, C., Sun, D.: On bayesian adaptive video super resolution. IEEE Transactions on Pattern Analysis and Machine Intelligence **36**(2), 346–360 (2013)
17. Liu, H., Liu, J., Hou, S., Tao, T., Han, J.: Perception consistency ultrasound image super-resolution via self-supervised cyclegan. Neural Computing and Applications pp. 1–11 (2021)
18. Meijering, E.H., Niessen, W.J., Pluim, J.P., Viergever, M.A.: Quantitative comparison of sinc-approximating kernels for medical image interpolation. In: International conference on medical image computing and computer-assisted intervention. pp. 210–217. Springer (1999)
19. Nah, S., Baik, S., Hong, S., Moon, G., Son, S., Timofte, R., Mu Lee, K.: Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 0–0 (2019)
20. Ouyang, D., He, B., Ghorbani, A., Yuan, N., Ebinger, J., Langlotz, C., Heidenreich, P., Harrington, R., Liang, D., Ashley, E., Zou, J.: Video-based ai for beat-to-beat assessment of cardiac function. Nature **580** (04 2020)
21. Patil, P., Dasgupta, B.: Role of diagnostic ultrasound in the assessment of musculoskeletal diseases. Therapeutic Advances in Musculoskeletal Disease **4**(5), 341–355 (2012)
22. Pellicori, P., Platz, E., Dauw, J., Ter Maaten, J.M., Martens, P., Pivetta, E., Cleland, J.G., McMurray, J.J., Mullens, W., Solomon, S.D., et al.: Ultrasound imaging of congestion in heart failure: examinations beyond the heart. European Journal of Heart Failure **23**(5), 703–712 (2021)
23. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena **60**(1-4), 259–268 (1992)
24. Tang, W., Cui, K., Chan, R.H., Morel, J.M.: Bilateral signal warping for left ventricular hypertrophy diagnosis. arXiv preprint arXiv:2411.08819 (2024)
25. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018)

26. Van Sloun, R.J., Solomon, O., Bruce, M., Khaing, Z.Z., Eldar, Y.C., Mischi, M.: Deep learning for super-resolution vascular ultrasound imaging. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1055–1059. IEEE (2019)
27. Xue, T., Chen, B., Wu, J., Wei, D., Freeman, W.T.: Video enhancement with task-oriented flow. International Journal of Computer Vision **127**, 1106–1125 (2019)
28. Yoon, J.H., Kim, M.J., Lee, H.S., Kim, S.H., Youk, J.H., Jeong, S.H., Kim, Y.M.: Validation of the fifth edition bi-rads ultrasound lexicon with comparison of fourth and fifth edition diagnostic performance using video clips. Ultrasonography **35**(4), 318 (2016)
29. Yue, L., Shen, H., Li, J., Yuan, Q., Zhang, H., Zhang, L.: Image super-resolution: The techniques, applications, and future. Signal Processing **128**, 389–408 (2016)
30. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2472–2481 (2018)
31. Zhao, G., Kong, D., Xu, X., Hu, S., Li, Z., Tian, J.: Deep learning-based classification of breast lesions using dynamic ultrasound video. European Journal of Radiology **165**, 110885 (2023)
32. Zhou, X., Zhang, L., Zhao, X., Wang, K., Li, L., Gu, S.: Video super-resolution transformer with masked inter&intra-frame attention. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 25399–25408 (2024)