

國立臺灣大學電機資訊學院資訊工程學系

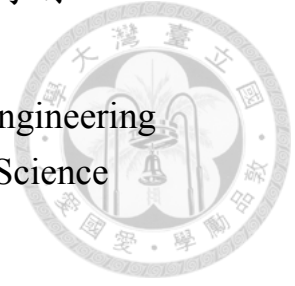
碩士論文

Department of Computer Science and Information Engineering

College of Electrical Engineering and Computer Science

National Taiwan University

Master Thesis



基於相位之心臟核磁共振超解析度成像

Efficient and Phase-aware Video Super-resolution for
Cardiac MRI

林智遠

Jhih-Yuan Lin

指導教授：徐宏民博士

Advisor: Winston Hsu, Ph.D.

中華民國 109 年 7 月

July, 2020

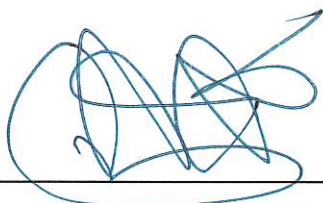
國立臺灣大學碩士學位論文
口試委員會審定書

基於相位之心臟核磁共振超解析度成像

Efficient and Phase-aware Video Super-resolution for
Cardiac MRI

本論文係林智遠君（學號 R07922021）在國立臺灣大學資訊工程學系完成之碩士學位論文，於民國 109 年 7 月 27 日承下列考試委員審查通過及口試及格，特此證明

口試委員：



（指導教授）

陳文進

葉木子

李志剛

蘇東弘

系主任

莊永裕



誌謝

碩士生涯即將劃下了句點，在這兩年當中，要感謝許多人的幫助和支持，我才能順利地完成碩士學位。首先要感謝父母從小到大的栽培和信任，讓我可以盡情地去探索自己的興趣，也很支持我做的決定。感謝指導教授徐宏民教授，每周都會和我進行研究上的討論並給予指導，也提供許多研究上所需要的資源，時常分享自己的經驗來藉此鼓勵我們，讓我們有勇氣面對一次又一次的實驗或是投稿的失敗。感謝和我一起合作的友誼，每周例行地討論研究的進度，腦力激盪出新穎的想法，不管是從研究方向、做實驗、寫程式到投稿會議，都和我參與了整個過程。也很感謝這兩年當中一起打拼的實驗室夥伴，旻昇、岳成、俞安、與晟、雅量、哲宇、宸晞、昱翔、東毅、均庭，無論是研究、修課和生活上，都給予我許多建議和幫助。再次謝謝所有的人，覺得自己很幸運能夠在人生當中的這兩年遇到你們。最後，請容許我將這本論文獻給我最親愛的家人：林百壽、高麗雪和林筱軒。



摘要

心臟核磁共振的技術被廣泛用於檢測心臟功能，受惠於其非侵入式的方式，檢查過程中不會對病人造成傷害。然而，取得高解析度的心臟核磁共振影像是一個耗時且昂貴的程序。為此，我們提出一個新穎的端到端 (end-to-end) 可訓練的神經網路來解決心臟核磁共振超解析度成像的問題，無須調整既有的硬體設備與掃描協定。我們妥善運用心臟的領域知識 (即心跳週期相位)，以週期性的函數來描述心跳週期，以對應心臟核磁共振所特有的循環特性。此外，我們提出殘差中的殘差學習機制 (residual of residual learning)，讓網路可以循序漸進地學習低解析度到高解析度的映射，透過這樣的機制，我們的方法能夠彈性地因應不同難度的問題。在大規模數據集上的定量與定性分析結果顯示我們的方法優於現有最佳的方法。

關鍵字：心臟核磁共振、超解析度成像、深度學習



Abstract

Cardiac Magnetic Resonance Imaging (CMR) is widely used since it can illustrate the structure and function of the heart in a non-invasive and painless way. However, it is time-consuming and high-cost to acquire high-quality scans due to the hardware limitation. To this end, we propose a novel end-to-end trainable network to solve CMR video super-resolution problem without the hardware upgrade and the scanning protocol modifications. We incorporate the cardiac knowledge into our model to assist in utilizing the temporal information. Specifically, we formulate the cardiac knowledge as the periodic function, which is tailored to meet the cyclic characteristic of CMR. Besides, the proposed residual of residual learning scheme facilitates the network to learn the LR-HR mapping in a progressive refinement fashion. This mechanism enables the network to have the adaptive capability by adjusting refinement iterations depending on the difficulty of the task. Extensive experimental results on large-scale datasets demonstrate the superiority of the proposed method compared with numerous state-of-the-art methods.

Keywords: Cardiac MRI, Video Super-resolution, Deep Learning



Contents

誌謝	iii
摘要	iv
Abstract	v
1 Introduction	1
2 Proposed approach	3
2.1 Overall architecture	3
2.2 Phase fusion module	4
2.3 Residual of residual learning	5
2.4 Loss function	6
3 Experiment	8
3.1 Data preparation	8
3.2 Evaluation metrics	8
3.3 Training details	9
3.4 Experimental results	9
3.5 Ablation study	10
4 Conclusion	12
Bibliography	13



List of Figures

1.1	Our main idea	2
2.1	Model overview	4
2.2	Proposed components	7
3.1	Qualitative results	10
3.2	Experimental analysis	11



List of Tables

3.1	Quantitative results	9
3.2	Ablation study	11



Chapter 1

Introduction

Magnetic Resonance Imaging (MRI) has been widely used to examine almost any part of the body since it can depict the structure inside the human non-invasively and produce high contrast images. Notably, cardiac MRI (CMR) assessing cardiac structure and function plays a key role in evidence-based diagnostic and therapeutic pathways in cardiovascular disease [25], including the assessment of myocardial ischemia, cardiomyopathies, myocarditis, congenital heart disease [26]. However, obtaining high-resolution CMR is time-consuming and high-cost as it is sensitive to the changes in the cardiac cycle length and respiratory position [21], which is rarely clinically applicable.

To address this issue, the single image super-resolution (SISR) technique, which aims at reconstructing a high-resolution (HR) image from low-resolution (LR) one, holds a great promise that does not need to change the hardware or scanning protocol. Most of the MRI SISR approaches [19, 3, 22] are based on the deep learning-based methods [5, 14], which learn the LR-HR mapping with extensive LR-HR paired data. On the other hand, several previous studies [11, 31] adapt the self-similarity based SISR algorithm [8], which does not need external HR data for training. However, straightforwardly employing the aforementioned methods is not appropriate for CMR video reconstruction since the relationship among the consecutive frames in CMR video is not well considered. Therefore, we adopt the video super-resolution (VSR) technique, which can properly leverage the temporal information and has been applied in numerous works [20, 10, 30, 27, 7], to perform CMR video reconstruction.

In this work, we propose an end-to-end trainable network to address CMR VSR problem. To well consider the temporal information, we choose ConvLSTM [28], which has been proven effective [6, 9], as our backbone. Moreover, we introduce the domain knowledge (*i.e.*, cardiac phase), which has shown to be important for the measurement of the stroke volume [13] and disease diagnosis [29], to provide the direct guidance about the temporal relationship in a cardiac cycle. Combined with the proposed *phase fusion module*, the model can better utilize the temporal information. Last but not the least, we devise the *residual of residual learning* inspired by the iterative error feedback mechanism [17, 2] to guide the model iteratively recover the lost details. Different from other purely feed-forward approaches [16, 10, 27, 30, 20], our iterative learning strategy can make the model easier in representing the LR-HR mapping with fewer parameters.

We evaluate our model and multiple state-of-the-art baselines on two synthetic datasets established by mimicking the acquisition of MRI [4, 31] from two publicly datasets [1, 24]. It is worth noting that one of them is totally for external evaluation. To properly assess the model performance, we introduce the cardiac metrics based on PSNR and SSIM. The experimental results turn out that the proposed network can stand out from existing methods even on the large-scale external dataset, which indicates our model has the generalization ability. To our best knowledge, this work is the pioneer to address the CMR VSR problem and provide a benchmark to facilitate the development in this domain.

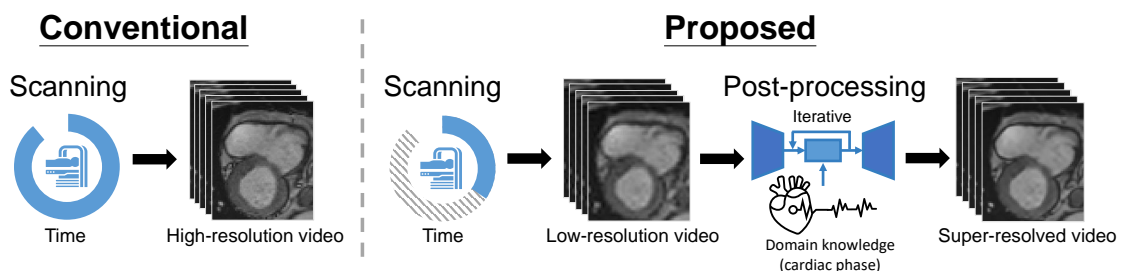


Figure 1.1: **Our main idea.** We present efficient post-processing to facilitate the acquisition of high-quality cardiac MRI (CMR) that is conventionally time-consuming, high-cost, and sensitive to the changes in the cardiac cycle length and respiratory position [21]. Specifically, we utilize the domain knowledge and iteratively enhance low-resolution CMR by a neural network, which can reduce the scan time and cost without changing the hardware or scanning protocol.



Chapter 2

Proposed approach

Let $I_{LR}^t \in \mathbb{R}^{H \times W}$ denote the t -th LR frame obtained by down-sampling the original HR frame $I_{HR}^t \in \mathbb{R}^{rH \times rW}$ with the scale factor r . Given a sequence of LR frames denoted as $\{I_{LR}^t\}$, the proposed end-to-end trainable model aims to estimate the corresponding high-quality results $\{I_{SR}^t\}$ that approximate the ground truth frames $\{I_{HR}^t\}$. Besides, \oplus refers to the element-wise addition.

2.1 Overall architecture

Our proposed network is illustrated in Fig 2.1. It consists of a feature extractor, a bidirectional ConvLSTM [28], a phase fusion module, and an up-sampler. The feature extractor (FE) first exploits the frame I_{LR}^t to obtain the low-frequency feature L^t . Subsequently, the bidirectional ConvLSTM [28] comprising a forward ConvLSTM ($ConvLSTM_F$) and a backward ConvLSTM ($ConvLSTM_B$) makes use of the low-frequency feature L^t to generate the high-frequency features H_F^t, H_B^t . With the help of its memory mechanism, the bidirectional ConvLSTM can fully utilize the temporal relationship among consecutive frames in both directions. In addition, we can update the memory cells in the bidirectional ConvLSTM in advance instead of starting with the empty states due to the cyclic characteristic of the cardiac videos. This can be done by feeding n consequent updated frames before and after the input sequence $\{I_{LR}^t\}$ to the network.

Furthermore, to completely integrate the bidirectional features, the designed phase

fusion module (PF) applies the cardiac knowledge of the $2N + 1$ successive frames from $t - N$ to $t + N$ in the form of the phase code $P^{[t-N:t+N]}$, which can be defined as $H_P^t = PF(H_F^{[t-N:t+N]}, H_B^{[t-N:t+N]}, P^{[t-N:t+N]})$, where H_P^t represents the fused high-frequency feature. After that, the fused high-frequency feature H_P^t combined with the low-frequency feature L^t through the global skip connection is up-scaled by the up-sampler (Up) into the super-resolved image $I_{SR}^t = Up(H_P^t \oplus L^t)$. We further define the sub-network (Net_{sub}) as the combination of PF , $ConvLSTM_F$ and $ConvLSTM_B$. The purpose of Net_{sub} is to recover the high-frequency residual $H_P^t = Net_{sub}(L^t)$. Besides, we employ the deep supervision technique [15] to provide the additional gradient signal and stabilize the training process by adding two auxiliary paths, namely $I_{SR,F}^t = Up(H_F^t \oplus L^t)$ and $I_{SR,B}^t = Up(H_B^t \oplus L^t)$. Finally, we propose the residual of residual learning that progressively restores the residual that has yet to be recovered in each refinement stage ω . To simplify the notation, ω is omitted when it equals to 0, e.g., L_F^t means the low-frequency feature of the t -th frame at the 0-th stage $L_F^{t,0}$.

2.2 Phase fusion module

The cardiac cycle is a cyclic sequence of events when the heart beats, which consists of systole and diastole process. Identification of the end-systole (ES) and the end-diastole (ED) in a cardiac cycle has been proved critical in several applications, such as the measure-

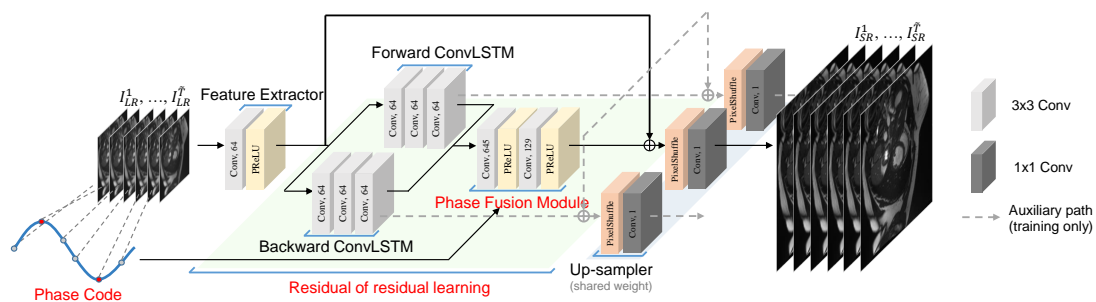


Figure 2.1: **Model overview.** The bidirectional ConvLSTM [28] utilizes the temporal information from forward and backward directions. The *phase fusion module* exploits the informative phase code to leverage the bidirectional features. With the *residual of residual learning*, the network recovers the results in a coarse-to-fine fashion. Auxiliary paths are adopted for stabilizing the training procedure.

ment of the ejection fraction and stroke volume [13], and disease diagnosis [29]. Hence, we embed the physical meaning of the input frames into our model with the informative phase code generated by projecting the cardiac cycle to the periodic Cosine function as depicted in Fig. 2.2a. Specifically, we map the process of the systole and the diastole to the half-period cosine separately:

$$P^t = \begin{cases} \text{Cos}(\pi \times \frac{t-ED}{ES-ED}), & \text{if } ED < t \leq ES \\ \text{Cos}(\pi \times (1 + \frac{(t-ES)\%T}{T-(ES-ED)})), & \text{otherwise} \end{cases} \quad (2.1)$$

where % denotes modulo operation and T is the frame number in a cardiac cycle.

The overview of the proposed phase fusion module is shown in Fig 2.2b. The features from the bidirectional ConvLSTM with the corresponding phase code are concatenated and fed into the fusion module. With the help of consecutive $2N + 1$ phase codes, it can link the same-position frames from different periods (inter-period). Besides, it can realize the heart is relaxing or contracting as the phase code is respectively increasing or decreasing (intra-period).

2.3 Residual of residual learning

In the computer vision field, the iterative error-correcting mechanism plays an essential role in several topics, such as reinforcement learning [17], scene reconstruction [18], and human pose estimation [2]. Inspired by this mechanism, we propose the residual of residual learning composing the reconstruction process into multiple stages, as shown in Fig. 2.2c. At each stage, the sub-network (Net_{sub}) in our model estimates the high-frequency residual based on the current low-frequency feature, and then the input low-frequency feature is updated for the next refinement stage. Let $L^{t,0}$ be the initial feature from the feature extractor (FE) and $L^{t,\omega}$ denote the updated feature at the iteration ω , the

residual of residual learning for Ω stages can be described as the recursive format:

$$L^{t,\omega} = \begin{cases} FE(I_{LR}^t), & \text{when } \omega = 0 \\ L^{t,\omega-1} \oplus Net_{sub}(L^{t,\omega-1}), & \text{if } 0 < \omega \leq \Omega \end{cases} \quad (2.2)$$



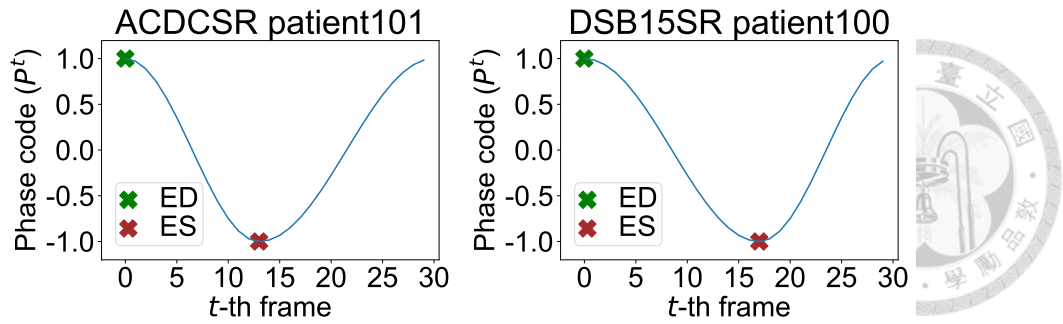
Then, the network generates the super-resolution result $I_{SR}^{t,\omega}$ based on the current reconstructed feature $L^{t,\omega}$, which can be written as:

$$I_{SR}^{t,\omega} = Up(L^{t,\omega} \oplus Net_{sub}(L^{t,\omega})) \quad (2.3)$$

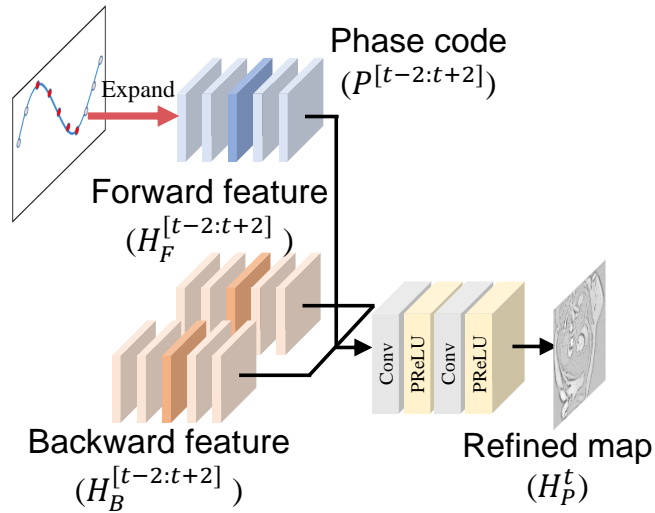
The model progressively restores the residual that has yet to be recovered in each refinement stage, which is so-called the residual of residual learning. Compared to other one-step approaches [16, 10, 27, 30, 20], the proposed mechanism tries to break down the ill-posed problem into several easier sub-problems in the manner of divide-and-conquer. Most notably, it can dynamically adjust the iteration number depending on the problem difficulty without any additional parameters.

2.4 Loss function

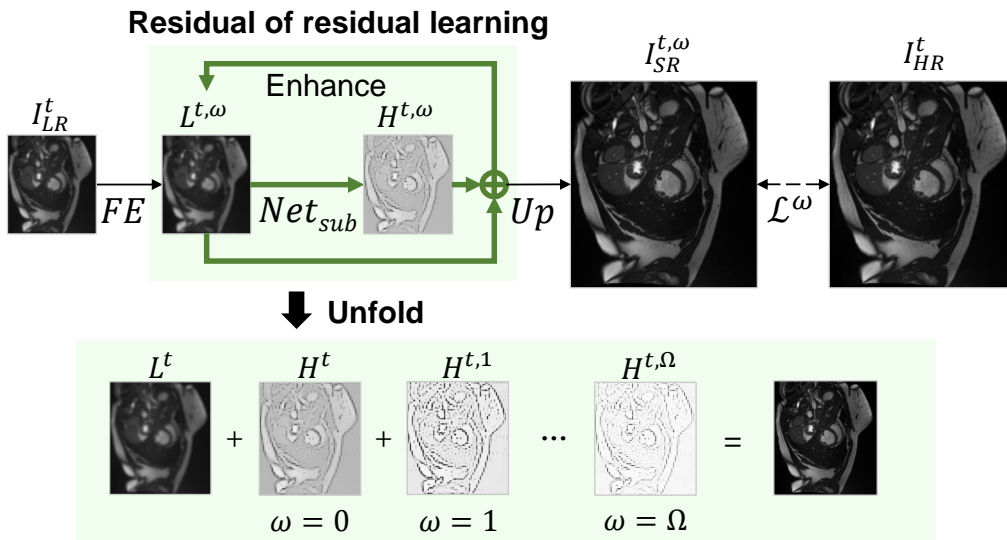
In this section, we elaborate on the mathematical formulation of our cost function. At each refinement stage ω , the super-resolved frames $\{I_{SR}^{t,\omega}\}$ are supervised by the ground-truth HR video $\{I_{HR}^t\}$, which can be formulated as $\mathcal{L}^\omega = \frac{1}{\tilde{T}} \sum_{t=1}^{\tilde{T}} \|I_{SR}^{t,\omega} - I_{HR}^t\|_1$, where \tilde{T} indicates the length of the video sequence fed into the network. We choose the L1 loss as the cost function since the previous works have demonstrated that the L1 loss provides better convergence compared to the widely used L2 loss [32, 16]. Besides, we apply the deep supervision technique as described in Sec. 2.1 by adding two auxiliary losses $\mathcal{L}_F^\omega = \frac{1}{\tilde{T}} \sum_{t=1}^{\tilde{T}} \|I_{SR,F}^{t,\omega} - I_{HR}^t\|_1$ and $\mathcal{L}_B^\omega = \frac{1}{\tilde{T}} \sum_{t=1}^{\tilde{T}} \|I_{SR,B}^{t,\omega} - I_{HR}^t\|_1$. Hence, the total loss function can be summarized as $\mathcal{L} = \sum_{\omega=0}^{\Omega} (\mathcal{L}^\omega + \mathcal{L}_F^\omega + \mathcal{L}_B^\omega)$, where Ω denoted as the total number of refinement stages.



(a) Phase code



(b) Phase fusion module



(c) Residual of residual learning

Figure 2.2: **Proposed components.** (a) *Phase code* formulated as the periodic function contains domain knowledge (*i.e.*, cardiac phase). (b) *Phase fusion module* can realize the phase of the current sequence with the cardiac knowledge to thoroughly integrate the bidirectional features. (c) *Residual of residual learning* aims at directing the model to reconstruct the results in a coarse-to-fine manner.



Chapter 3

Experiment

3.1 Data preparation

To our best knowledge, there is no publicly available CMR dataset for the VSR problem. Hence, we create two datasets named ACDCSR and DSB15SR based on the public MRI datasets. One is the Automated Cardiac Diagnosis Challenge dataset [1], which contains four dimension MRI scans of a total of 150 patients. The other is the large-scale Second Annual Data Science Bowl Challenge dataset [24] composed of 2D cine MRI videos that contain 30 images across the cardiac cycle per sequence. We use its testing dataset comprising 440 patients as the external assessment to verify the robustness and generalization of the algorithms. To more accurately mimic the acquisition of LR MRI scans [4, 31], we project the HR MRI videos to the frequency domain by Fourier transform and filter the high-frequency information. After that, we apply the inverse Fourier transform to project the videos back to the spatial domain and further downsample by bicubic interpolation with the scale factor 2, 3, and 4.

3.2 Evaluation metrics

PSNR and SSIM criteria have been widely used in previous studies to evaluate the SR algorithms. However, the considerable disparity of the proportion of the cardiac region to the background region in MRI images makes the results heavily biased towards the in-

Table 3.1: **Quantitative results.** The red and blue indicate the best and the second-best performance, respectively. We adopt CardiacPSNR/CardiacSSIM to fairly assess the reconstruction quality of the heart region. It is worth noting that the large-scale DSB15SR dataset is entirely for external evaluation.

Dataset	Scale	SISR			VSR				
		Bicubic	EDSR[16]	DUF[10]	EDVR[27]	RBPN[7]	TOFlow[30]	FRVSR[20]	Model (Ours)
ACDCSR	×2	33.0927 / 0.9362	37.3022 / 0.9681	37.4008 / 0.9688	- / -	37.5017 / 0.9694	36.6510 / 0.9641	- / -	37.5003 / 0.9696
	×3	29.0724 / 0.8472	32.8177 / 0.9201	32.7942 / 0.9203	- / -	32.9099 / 0.9225	32.4535 / 0.9136	- / -	32.9342 / 0.9231
	×4	26.9961 / 0.7611	30.2536 / 0.8631	30.2420 / 0.8621	30.2817 / 0.8655	30.3294 / 0.8653	30.0087 / 0.8538	30.1693 / 0.8592	30.4060 / 0.8668
DSB15SR	×2	34.1661 / 0.9597	40.1723 / 0.9815	40.3548 / 0.9822	- / -	40.3792 / 0.9824	39.5042 / 0.9794	- / -	40.4635 / 0.9821
	×3	29.1175 / 0.8854	33.9893 / 0.9424	33.9736 / 0.9428	- / -	34.1320 / 0.9445	33.6656 / 0.9386	- / -	34.2169 / 0.9451
	×4	26.5157 / 0.8065	30.6354 / 0.8907	30.7411 / 0.8918	30.8564 / 0.8949	30.7985 / 0.8933	30.3153 / 0.8836	30.5800 / 0.8889	30.9104 / 0.8956

significant background region. Therefore, we introduce CardiacPSNR and CardiacSSIM to assess the performance more impartially and objectively. Specifically, we employ a heart ROI detection method similar to [23] to crop the cardiac region and calculate PSNR and SSIM in this region. This can reduce the influence of the background region and more accurately reflect the reconstruction quality of the heart region.

3.3 Training details

For training, we randomly crop the LR clips of $\tilde{T} = 7$ consecutive frames of size 32×32 with the corresponding HR clips. We experimentally choose $n = 6$ and $\Omega = 2$ as detailed in Sec. 3.5, while $N = 2$ in the phase fusion module. We use the Adam optimizer [12] with learning rate 10^{-4} and set the batch size to 16. For other baselines, we basically follow their original settings except the necessary modifications to train them from the scratch.

3.4 Experimental results

To confirm the superiority of the proposed approach, we compare our network with multiple state-of-the-art methods, namely EDSR [16], DUF [10], EDVR [27], RBPN [7], TOFlow [30], and FRVSR [20]. We present the quantitative and qualitative results in Tab. 3.1 and Fig. 3.1 respectively. Our approach outperforms almost all the existing methods by a huge margin in all scales in terms of CardiacPSNR and CardiacSSIM. In addition, our method can yield more clear and photo-realistic SR results which subjectively closer to the ground truths. Moreover, the results on the external DSB15SR dataset are sufficiently

convincing to validate the generalization of the proposed approach. On the other hand, the comparison with regard to the model parameters, FPS, and the image quality in the cardiac region plotted in Fig. 3.2a demonstrates that our method strikes the best balance between efficiency and reconstruction performance.

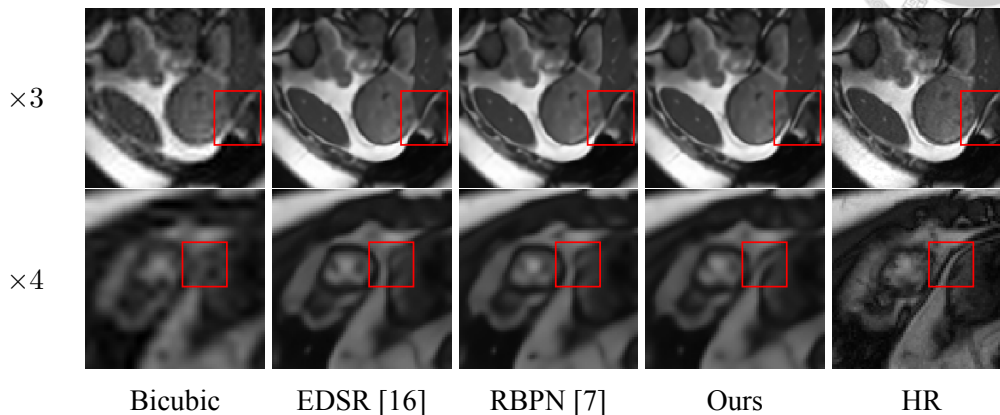


Figure 3.1: **Qualitative results.** Zoom in to see better visualization.

3.5 Ablation study

We adopt the unidirectional ConvLSTM as the simplest baseline. As shown in the Tab. 3.2, the temporal information is important since the model performance is worse when the memory cells in ConvLSTM are disabled. As the cardiac MRI video is cyclic, we can refresh the memory by feeding n successive frames. Accordingly, we analyze the relation between n and model performance. The result in Fig 3.2b turns out that the network significantly improves as the updated frame number increases. Moreover, the forward and backward information is shown to be useful and complementary for recovering the lost details. In Sec. 2.2, we exploit the knowledge of the cardiac phase to better fuse the bidirectional information. The result in Tab. 3.2 reveals that the phase fusion module can leverage the bidirectional temporal features more effectively. Besides, we explore the influence of the total number of refinement stages Ω in the residual of residual learning. It can be observed from Fig. 3.2c that the reconstruction performance is improved as the total refinement stages continue to increase. The possible reason for the saturation or degradation of the overall performance when Ω equals to 3 or 4 is overfitting.

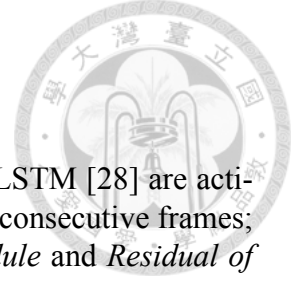
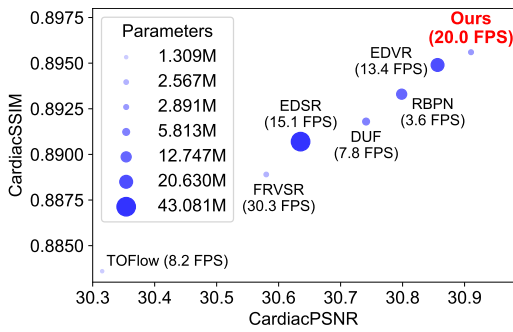
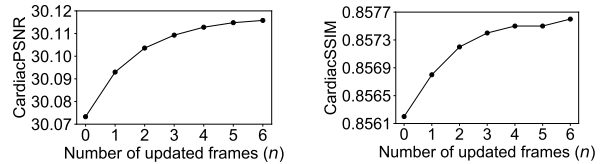


Table 3.2: **Ablation study.** *Memory*: the memory cells in the ConvLSTM [28] are activated; *Updated memory*: the memory cells are updated by feeding n consecutive frames; *Bidirection*: bidirectional ConvLSTM is adopted; *Phase fusion module* and *Residual of residual learning*: the proposed components are adopted.

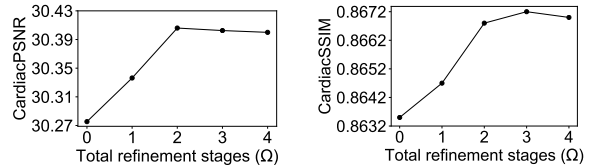
Memory	Updated memory ($n = 6$)	Bidirection	Phase fusion module	Residual of residual learning ($\Omega = 2$)	CardiacPSNR/CardiacSSIM
					29.7580 / 0.8458
✓					30.0733 / 0.8562
✓	✓				30.1790 / 0.8596
✓	✓	✓			30.2380 / 0.8623
✓	✓	✓	✓		30.2754 / 0.8635
✓	✓	✓	✓	✓	30.4060 / 0.8668



(a) Efficiency vs performance on DSB15SR dataset for scale $\times 4$. (FPS: processed frames per second)



(b) Analysis of the update frame number n .



(c) Analysis of total refinement stages Ω .

Figure 3.2: **Experimental analysis.** (a) Our network outperforms other baselines with fewer parameters and higher FPS. (b) The performance is progressively enhanced as n increases, which indicates that the prior sequence can provide useful information. (c) The performance can be improved with Ω increasing.



Chapter 4

Conclusion

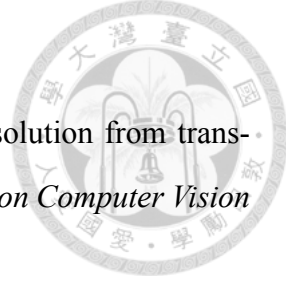
In this work, we define the cyclic cardiac MRI video super-resolution problem which has not yet been completely solved to our best knowledge. To tackle this issue, we bring the cardiac knowledge into our network and employ the residual of residual learning to train in the progressive refinement manner, which enables the model to generate sharper results with fewer model parameters. In addition, we build large-scale datasets and introduce cardiac metrics for this problem. Through extensive experiments, we demonstrate that our network outperforms the state-of-the-art baselines qualitatively and quantitatively. Most notably, we carry out the external evaluation, which indicates our model exhibits good generalization behavior. We believe our approach can be seamlessly applied to other modalities such as computed tomography angiography and echocardiography.



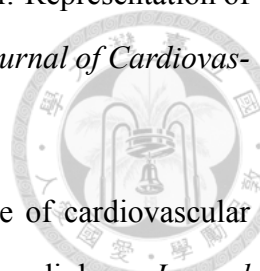
Bibliography

- [1] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, I. Cetin, K. Lekadir, O. Camara, M. A. G. Ballester, et al. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved? *IEEE transactions on medical imaging*, 37(11):2514–2525, 2018.
- [2] J. Carreira, P. Agrawal, K. Fragkiadaki, and J. Malik. Human pose estimation with iterative error feedback. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4733–4742, 2016.
- [3] Y. Chen, F. Shi, A. G. Christodoulou, Y. Xie, Z. Zhou, and D. Li. Efficient and accurate mri super-resolution using a generative adversarial network and 3d multi-level densely connected network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 91–99. Springer, 2018.
- [4] Y. Chen, Y. Xie, Z. Zhou, F. Shi, A. G. Christodoulou, and D. Li. Brain mri super resolution using 3d deep densely connected neural networks. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 739–742. IEEE, 2018.
- [5] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015.
- [6] C. Finn, I. Goodfellow, and S. Levine. Unsupervised learning for physical interaction through video prediction. In *Advances in neural information processing systems*, pages 64–72, 2016.

- [7] M. Haris, G. Shakhnarovich, and N. Ukita. Recurrent back-projection network for video super-resolution. *arXiv preprint arXiv:1903.10128*, 2019.
- [8] J.-B. Huang, A. Singh, and N. Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5197–5206, 2015.
- [9] Y. Huang, W. Wang, and L. Wang. Bidirectional recurrent convolutional networks for multi-frame super-resolution. In *Advances in Neural Information Processing Systems*, pages 235–243, 2015.
- [10] Y. Jo, S. Wug Oh, J. Kang, and S. Joo Kim. Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3224–3232, 2018.
- [11] A. Jog, A. Carass, and J. L. Prince. Self super-resolution for magnetic resonance images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 553–560. Springer, 2016.
- [12] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [13] A. Lalande, N. Salve, A. Comte, M.-C. Jaulent, L. Legrand, P. Walker, Y. Cottin, J.-E. Wolf, and F. Brunotte. Left ventricular ejection fraction calculation from automatically selected and processed diastolic and systolic frames in short-axis cine-mri. *Journal of Cardiovascular Magnetic Resonance*, 6(4):817–827, 2004.
- [14] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.
- [15] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu. Deeply-supervised nets. In *Artificial intelligence and statistics*, pages 562–570, 2015.



- [16] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 136–144, 2017.
- [17] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.
- [18] M. Montemerlo, S. Thrun, D. Koller, B. Wegbreit, et al. Fastslam: A factored solution to the simultaneous localization and mapping problem. *Aaai/iaai*, 593598, 2002.
- [19] C.-H. Pham, A. Ducournau, R. Fablet, and F. Rousseau. Brain mri super-resolution using deep 3d convolutional networks. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 197–200. IEEE, 2017.
- [20] M. S. Sajjadi, R. Vemulapalli, and M. Brown. Frame-recurrent video super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6626–6634, 2018.
- [21] M. Salerno, B. Sharif, H. Arheden, A. Kumar, L. Axel, D. Li, and S. Neubauer. Recent advances in cardiovascular magnetic resonance: techniques and applications. *Circulation: Cardiovascular Imaging*, 10(6):e003951, 2017.
- [22] J. Shi, Q. Liu, C. Wang, Q. Zhang, S. Ying, and H. Xu. Super-resolution reconstruction of mr image with a novel residual learning network algorithm. *Physics in Medicine & Biology*, 63(8):085011, 2018.
- [23] L. Tautz, O. Friman, A. Hennemuth, A. Seeger, and H.-O. Peitgen. Automatic detection of a heart roi in perfusion mri images. In *Bildverarbeitung für die Medizin 2011*, pages 259–263. Springer, 2011.
- [24] L. The National Heart and B. Institute. Data science bowl cardiac challenge data, 2015.

- 
- [25] F. von Knobelsdorff-Brenkenhoff, G. Pilz, and J. Schulz-Menger. Representation of cardiovascular magnetic resonance in the aha/acc guidelines. *Journal of Cardiovascular Magnetic Resonance*, 19(1):70, 2017.
- [26] F. von Knobelsdorff-Brenkenhoff and J. Schulz-Menger. Role of cardiovascular magnetic resonance in the guidelines of the european society of cardiology. *Journal of Cardiovascular Magnetic Resonance*, 18(1):6, 2015.
- [27] X. Wang, K. C. Chan, K. Yu, C. Dong, and C. Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.
- [28] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems*, pages 802–810, 2015.
- [29] H.-y. Xu, Z.-g. Yang, Y.-k. Guo, K. Shi, X. Liu, Q. Zhang, L. Jiang, and L.-j. Xie. Volume-time curve of cardiac magnetic resonance assessed left ventricular dysfunction in coronary artery disease patients with type 2 diabetes mellitus. *BMC cardiovascular disorders*, 17(1):145, 2017.
- [30] T. Xue, B. Chen, J. Wu, D. Wei, and W. T. Freeman. Video enhancement with task-oriented flow. *International Journal of Computer Vision*, 127(8):1106–1125, 2019.
- [31] C. Zhao, A. Carass, B. E. Dewey, and J. L. Prince. Self super-resolution for magnetic resonance images using deep networks. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 365–368. IEEE, 2018.
- [32] H. Zhao, O. Gallo, I. Frosio, and J. Kautz. Loss functions for neural networks for image processing. *arXiv preprint arXiv:1511.08861*, 2015.