Pyramid Self-contrastive Learning Framework for Test-time Ultrasound Image Denoising

Jiajing Zhang[a,1], Bingze Dai[a], Xi Zhang[a], Yue Xu[b] and Wei-Ning Lee[a,c,*]

[a]Department of Electrical and Computer Engineering, The University of Hong Kong, Pokfulam, Hong Kong
[b]Department of Biomedical Engineering, Duke University, North Carolina, United States
[c]School of Biomedical Engineering, The University of Hong Kong, Pokfulam, Hong Kong

ARTICLE INFO

Keywords:
Ultrasound
Denoising
Contrastive Learning
Test-Time Training
Self-supervision

ABSTRACT

The inherent electronic and speckle noise complicates clinical interpretation of ultrasound images. Conventional denoising methods rely on explicit noise assumptions whose validity diminishes under composite noise conditions. Learning-based methods require massive labeled data and model parameters. These pre-defined and pre-trained manners entail an inevitable domain shift in complex *in vivo* environments, so they are limited to a specific noise type and often blur structural details. In this study, we propose a pure test-time training framework for one-shot ultrasound image denoising and apply it to synthetic aperture ultrasound (SAU), which synthesizes transmit focus from sub-aperture transmissions. Our Aperture-to-Aperture (A2A) framework disentangles anatomical similarity and noise randomness from shuffled sub-apertures through self-contrastive learning in pyramid latent spaces. The clean image is then decoded from the anatomy space, while discarding the noise space. A2A is trained at test time on one noisy sample of SAU signals, so it fundamentally eliminates the domain shift and pretraining costs. Simulation experiments, including electronic noise levels of 0 to 30 dB and different inclusion geometries, demonstrated an improvement of 69.3% SNR and 34.4% CNR by A2A. The *in vivo* results showed 84.8% SNR and 25.7% CNR gains using only two aperture data of the heart in six echocardiographic views, liver, and kidney. A2A delivers clear images/signals across diverse imaging targets and configurations, paving the way for more reliable anatomical visualization and functional assessment by ultrasound.

1. Introduction

Ultrasound serves as an indispensable imaging modality in modern healthcare, offering unique advantages in dynamic and real-time assessment, surgical guidance, and point-of-care diagnostics without ionizing radiation. Conventional ultrasound imaging is based on single transmit focusing, so optimal spatial resolution is limited to the predefined foci. Achieving two-way focus throughout the field of view, synthetic aperture ultrasound (SAU) divides an array transducer into multiple sub-apertures and synthesizes transmit focusing by coherently compounding beamformed signals from individual sub-apertures [1]. SAU and its high frame-rate adaptations [2] have now become prevalent, especially for functional assessment of the heart and abdominal organs. Nevertheless, at high frame rates, SAU employs sub-aperture unfocused wave transmissions, exhibiting lower signal-to-noise ratio and spatial resolution.

Fig. 1(a) illustrates major noise sources in ultrasound images, including electronic noise, speckle noise, and side lobes. Electronic noise primarily arises from electromagnetic interference and electronics and appears as random fluctuations following a Gaussian distribution. As grainy patterns (green boxes in Fig. 1(a)), speckle noise results from coherent wave interferences among echoes scattered by sub-wavelength structures [3]; it typically follows Rayleigh, Homodyned K, or Nakagami distributions, depending on tissue microstructure [4]. Side lobes stem from low-energy and off-axis echoes and may mask hypoechoic structure.

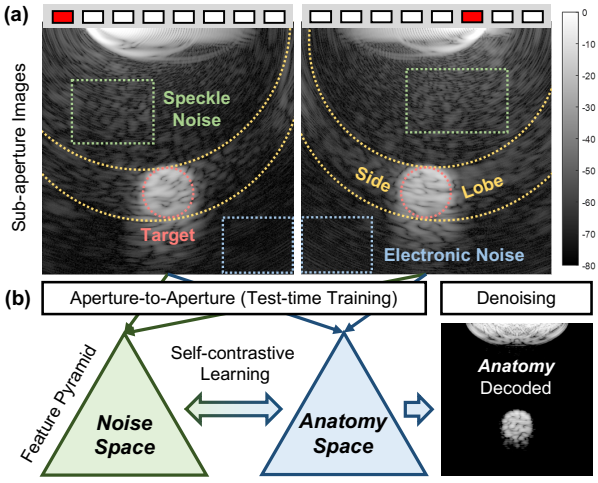

Figure 1: (a) Noise types in synthetic aperture ultrasound. (b) Aperture-to-Aperture (A2A) encodes sub-aperture images into a noise space and an anatomy space, and then decodes the anatomy space as a clean image.

These noises degrade image quality by obscuring anatomical boundaries, distorting fine structures, and reducing image contrast, thereby hindering downstream tasks such as tissue motion and blood flow estimation, strain and elasticity imaging, and computer-aided diagnosis. Unfortunately, these noise components are intrinsic to either the imaging system or wave-matter interaction.

*Corresponding author
wnlee@hku.hk (W. Lee)
ORCID(s): 0000-0001-8799-2492 (W. Lee)

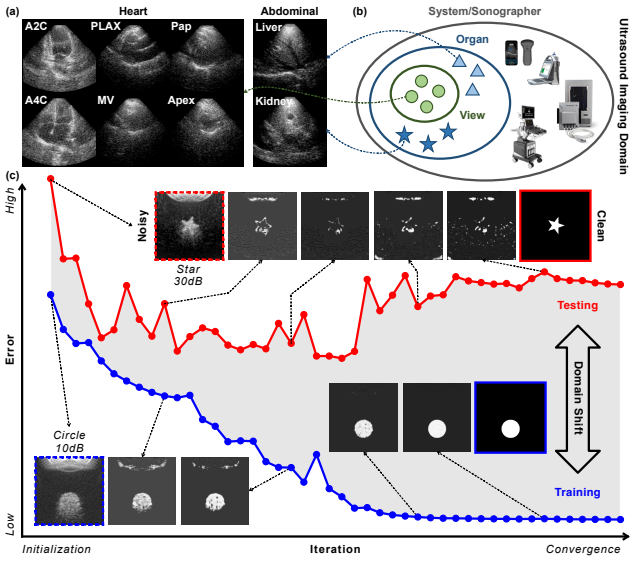

Figure 2: (a) Examples of *in vivo* ultrasound images sampled from (b) various imaging domains. (c) Illustration of domain shift under different SNRs and inclusion geometries. The dashed and solid boxes show noisy input and clean GT, respectively, for training (blue) and testing (red).

Despite extensive efforts in ultrasound denoising (Section 2), domain shift is still one of the most critical challenges. Conventional methods based on filters [5, 6, 7], local [8] or non-local [9, 10] similarity [11, 12, 13], and coherence factors [14, 15, 16, 17, 18] rely on explicit noise models that may not generalize across complex and dynamic imaging conditions. In practice, electronic noise varies between systems with different driving voltages and transmission frequencies, yielding different signal-to-noise ratio (SNR) levels [19]. Speckle noise differs among organs due to tissue-dependent scattering, which precludes a universal probabilistic representation [4, 20]. Furthermore, wave attenuation and time-gain compensation (TGC) cause depth-dependent noise levels [21]. As illustrated in Fig. 2(a), these factors produce considerable variations across organs and scan views. Therefore, model- or coherence-based methods [22, 23, 24] can hardly adapt to diverse imaging domains with sensitive parameters, and images tend to be over-smoothed under domain shift.

Recently, learning-based methods have employed regression models to learn implicit noise representations. Such data-driven approaches address the problem of unknown noise priors, but at the expense of massive labeled data and model parameters. In practice, their training is often confined to a given data domain. For SAU, a modality prized for its imaging flexibility, a system must handle diverse organs, scan views, and sonographer preferences. All these factors imply significant domain shifts (Fig. 2(b)).

Ultrasound noise is coupled with anatomical structures. Regression models learn a direct mapping from noisy to clean images rather than the underlying noise characteristics. Pre-trained models are favorable in the training domain, but they can degrade performance or even corrupt unseen anatomy and introduce artifacts when applied to real-world testing domains. This phenomenon is illustrated using simulated SAU images (Fig. 2(c)). A model trained until convergence on a dataset of circular inclusions with an SNR of 10 dB can fail on a test set of star inclusions with an SNR of 30 dB and severely break the boundaries. There remains an urgent need for a versatile ultrasound denoising method applicable to any imaging domain.

Data scarcity poses another challenge. Unlike CT or MR, where noisy–clean pairs can be acquired by low-high radiation dose or magnetic field, ultrasound can barely obtain paired data under *in vivo* conditions. For dynamic imaging, such as echocardiography, simultaneously capturing noisy and clean frames of a beating heart is infeasible. Hence, there is a pressing need for a self-supervised method that learns effectively from limited or even one-shot noisy data.

To tackle the aforementioned challenges, this paper proposes a self-contrastive test-time training framework for ultrasound signal denoising by learning the anatomical similarity and noise randomness among noisy samples from only one-shot imaging. The framework is first demonstrated in SAU and therefore termed as Aperture-to-Aperture (A2A). Our work provides the following contributions:

- **Self-contrastive learning** disentangles the anatomical similarity and noise randomness among apertures into two pyramid latent spaces;
- **A2A** forms a self-supervised proxy task of swapping shuffled noisy samples from multiple sub-aperture transmissions in SAU;
- **Pure test-time training** with only one-shot imaging eliminates the domain shift, pretraining costs, and labeled data reliance.

The proposed framework is architecture-agnostic and is validated through both simulation and *in vivo* experiments across various imaging targets, noise levels, and aperture configurations.

2. Related Works

2.1. Model-based Ultrasound Denoising

Ultrasound image denoising dates back to the 1970s [25] when global filters, including mean, median, and low-pass filters, were integrated into systems. Adaptive filters based on local noise statistics were subsequently introduced, including Lee, Wiener, and Frost filters, etc [6]. Speckle reducing anisotropic diffusion (SRAD) [8] applied anisotropic diffusion to despeckle. However, they tend to over-smooth images and discard fine structures. Wavelet-based methods combine the advantages of global and local methods [26] using various thresholding [27] in the wavelet domain. Non-local methods such as BM3D (block-matching and 3D filtering) [13] and WNNM (weighted nuclear norm minimization) [28] were adapted for ultrasound images [11, 12] under the low-rank prior [29]. OBNLM [10] combines the Gamma distribution to form a Bayesian non-local mean filter.

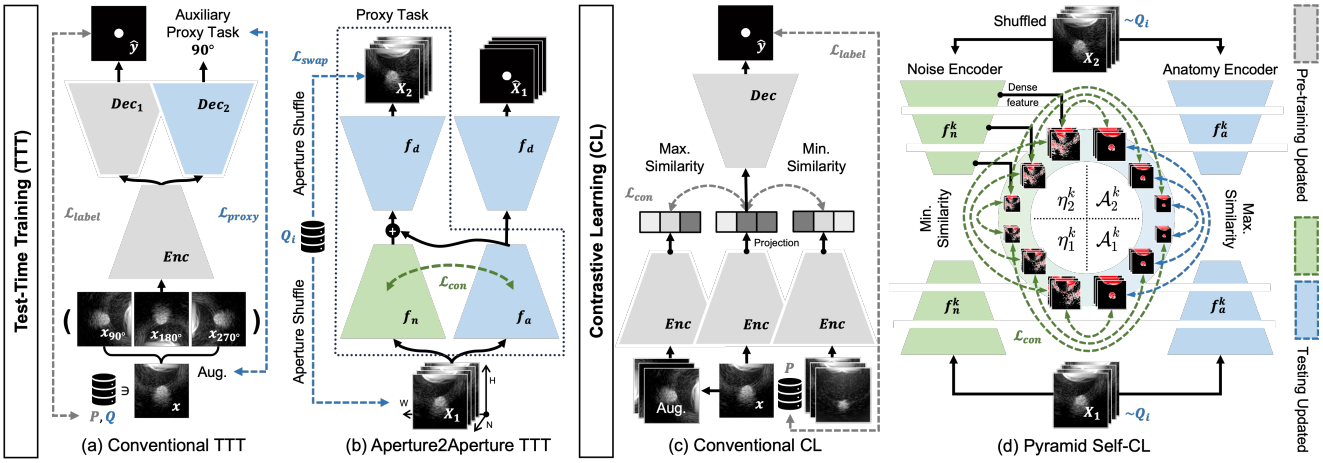

Figure 3: Comparison between (a) conventional test-time training and (b) our A2A strategy; comparison between (c) conventional contrastive learning and (d) our pyramid self-contrastive learning; Dashed lines indicate loss functions.

These methods are derived from explicit noise assumptions and are sensitive to parameters such as window size, threshold, and noise coefficients (intensity, variance, scale).

2.2. Coherence-based Ultrasound Denoising

Tissue inhomogeneities can cause phase distortion, resulting in focusing error and an increased level of side lobes. Coherence factor (CF) [14], which quantifies the coherent and incoherent signal energy components as a focusing quality factor, has been devised to weight aperture data to reduce the side lobes and correct phase aberration [14]. Based upon similar principles, generalized coherence factor (GCF) [15], phase coherence factor (PCF) [16], and other variants [16, 17, 18] were proposed. Alternatively, short-lag spatial coherence imaging [30] visualized the coherence values between adjacent elements as images [23, 24]. Besides, incoherent spatial compounding has been well established for the suppression of electronic and speckle noise and thus contrast enhancement [22].

These methods extend beyond the noise priors for B-mode images and measure coherence via predefined spatial covariance or correlation.

2.3. Learning-based Ultrasound Denoising

Learning-based methods formulate denoising as an image-to-image regression using RNN [31], UNet [32], GAN [33], autoencoder [34, 35], DDPM [36, 37], and customized attention mechanisms [38]. Their training paradigms can be categorized as supervised and unsupervised learning.

2.3.1. Supervised denoising

Supervised ultrasound denoising relies on large datasets of noisy-clean pairs to pre-train the model. Given numerous samples (x, y) drawn i.i.d. from a training distribution P, the optimization can be formulated as

$$\arg\min_{\theta} \mathbb{E}_{(x,y)\sim P}\left[\mathcal{L}\left(f_\theta(x), y\right)\right], \tag{1}$$

where x is a noisy input, and y is the clean or higher-quality counterpart of x, f_θ denotes the denoiser parameterized by θ, and $\mathcal{L}$ usually denotes an L_1 or L_2 distance.

Since collecting *in vivo* clean images is almost impractical, researchers have resorted to alternative label sources, including simulation [39, 40], conventional algorithms (e.g., WNNM) [41], images compounded from more tilted angles [37, 36], and other modalities (e.g., CT [42], ECG [43]). Some models were trained on natural images [44] before being transferred to simulated data [45]. However, none of these surrogates fully replicates the complex and dynamic *in vivo* conditions, thus limiting the variety of noise patterns learned. These methods inevitably suffer from domain shift and may introduce artifacts like fake textures.

2.3.2. Self-supervised denoising

The N2N [46] presents another paradigm that merely uses noisy images. Given noisy pairs (x_1, x_2) sampled from a clean image y from P, the objective is

$$\arg\min_{\theta} \mathbb{E}_{(x_1,x_2)\sim P}\left[\mathcal{L}\left(f_\theta(x_1), x_2\right)\right]. \tag{2}$$

If the noise in x_1 and x_2 is i.i.d. and zero-mean, y can be theoretically inferred under sufficient data and training.

Multi-angle plane-wave (PW) ultrasound acquires paired noisy images via transmissions with different steering angles. Huang et al. [47] and Asgariandehkordi et al. [48] divided PWs into odd and even groups to construct noisy pairs. Similarly, Jung et al. [49] utilized echoes from different angles as paired noisy observations of speckles. However, tissue dependency still violates the i.i.d. condition. Li et al. [20] proposed a multi-scale perturbation to induce variations in speckle patterns, thus creating noisy pairs. Yu et al. [50] generated noisy pairs by sampling B-mode images from overspreading chunks. The N2N was also extended to ultrasound annotation removal [51]. In addition to the N2N paradigm, Huh et al. [52] used unmatched 2D references to guide tunable 3D image enhancement; Muth et al. [53]

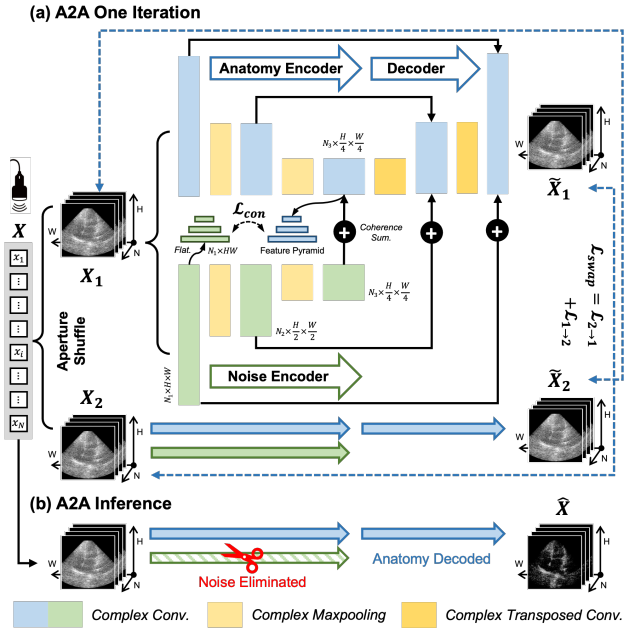

Figure 4: Pipeline of (a) one step of A2A iteration, (b) A2A inference (denoising) process.

achieved color-Doppler denoising via recursive dealiasing after segmentation.

Both supervised and self-supervised methods require large datasets for pretraining, which implies a potential domain shift out of the training distribution P. In fact, DIP (Deep image prior) [54] proved that a CNN could serve as an implicit denoiser without pretraining; N2S (Noise-to-Self) [55] leveraged independent noise across different measurements of one image. These methods made single-image denoising possible. In SAU, paired noisy signals are naturally acquired by sequential transmissions of sub-apertures. Nevertheless, no learning-based method has exploited signal coherence between sub-apertures.

3. Methodology

3.1. Problem Statement

This study considers ultrasound beamformed IQ signals. Given an SAU transmission with N sub-apertures, we construct the original IQ signal as $X \in \mathbb{C}^{N\times H\times W}$, where H denotes imaging depth and W denotes the number of elements in reception. A compounded noisy IQ signal $y \in \mathbb{C}^{H\times W}$ is the coherent sum of X along the aperture dimension. A B-mode image can be formed by log-compression after envelope detection on y. Let the clean IQ signal be $\hat{y}$ compounded from the clean $\hat{X}$. The relation between the clean and noisy sub-aperture signal can be modeled as

$$X = \hat{X} \odot \eta_{sp} + \eta_{sl} + \eta_e, \tag{3}$$

where additive $\eta_e \sim \mathcal{N}(0, \sigma^2)$ represents the electronic noise following a zero-mean Gaussian distribution; another additive η_{sl} represents the aperture-dependent side lobes; multiplicative η_{sp} represents the tissue-dependent speckle noise [56]; $\odot$ denotes a Hadamard product.

Unlike end-to-end regression that predicts a clean B-mode image, our goal is to denoise the sub-aperture IQ signal X. Compounding the denoised $\hat{X}$ approximates $\hat{y}$. $\hat{X}$ itself can facilitate functional assessment.

Table 1
Parameters used in *in vivo* imaging and k-Wave simulation.

	Imaging Parameter	Value
	Probe	P4-2 (64-element phased array)
	Virtual source	1.28 mm behind the array
	Transmit Wave	Diverging wave with $\frac{\pi}{2}$ opening angle
	Transmit frequency	2.5 MHz
In vivo	Number of sub-apertures	2~8
	Excitation	CaSA [57] driven at 4V
	Frame rate	1600 fps
Simulation	Number of sub-apertures	4, 8
	SNR	0, 10, 20, 30 dB

Table 2
Medium properties used in k-Wave simulation.

Parameter	Background Water	Inclusion Cardiac muscle [58]
Speed of sound: c_0 (m/s)	1540	1545
Density: ρ (kg/m^3)	1000	1060
Nonlinearity: B/A	8	7.1
Absorption: α_0 (dB/MHz/cm)	0.3	0.52
Geometry	/	Circle, star

3.2. Aperture-to-Aperture Strategy

3.2.1. Principle

From a statistical perspective, SAU can be regarded as sampling N times of the anatomy depicted by the same difference in acoustic impedance at tissue interfaces. As Fig. 1(a) shows, sub-aperture images in X are noisy observations of $\hat{y}$ with shared anatomical structure but different noise. The similarity between sub-apertures stems from a shared underlying $\hat{y}$, whereas their difference originates from random η_e, η_{sp} observed from different virtual source locations, and η_{sl} in different angular directions. Therefore, anatomical structures and noise are low-rank and high-rank components within the multi-aperture X, respectively.

Despite being intrinsically coupled in X, the low-rank anatomical and high-rank noise components become separable in a high-dimensional latent space, allowing a clean $\hat{X}$ to be derived exclusively from the low-rank part. Thus, we reformulate the denoising task as explicit decomposition and reconstruction of these low- and high-rank features extracted by neural networks.

Architecturally, A2A consists of three modules. (1) an anatomy encoder (f_a) encodes multi-aperture similarity from an input X; (2) a noise encoder (f_n) encodes the differences into a separate noise space; (3) a decoder (f_d) reconstructs another noisy $\tilde{X}$ combining both feature spaces as Eq. 4, whereas the clean $\hat{X}$ can be synthesized only from

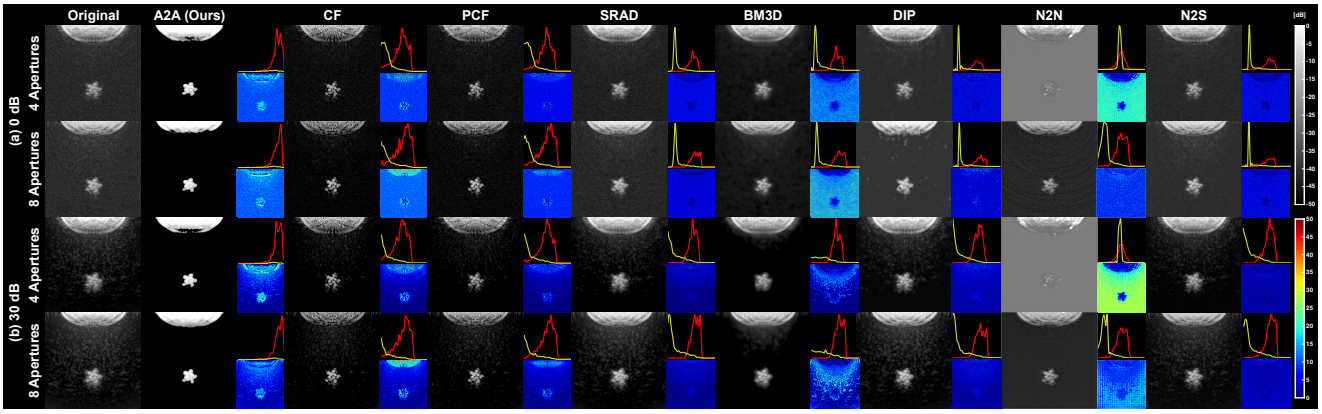

Figure 5: Performance comparison of denoising in simulations under SNR of (a) 0 dB and (b) 30 dB, and imaging configurations of four and eight sub-apertures. For each method, the denoised image is shown on the left side; the upper-right shows its intensity probability density functions (PDFs) of the signal (red curve) and noise (yellow curve) regions; the lower-right color image shows the difference map compared with the original image.

the anatomy space as Eq. 5:

$$\tilde{X} = f_d(f_a(X) + f_n(X)), \tag{4}$$

$$\hat{X} = f_d(f_a(X)). \tag{5}$$

3.2.2. Pure Test-Time Training (TTT)

In practice, the potential testing domain Q is unpredictable and cannot be fully covered by the pretraining domain P. A2A regards each testing environment as an independent data domain ($Q_i \subset Q$) and conducts initialized TTT on it. This pretraining-free method eliminates the premise that training and testing data are drawn from the same domain ($Q \approx P$), thereby fundamentally eliminating the domain shift.

Fig. 3 (a) illustrates the conventional TTT [59] pipeline. Firstly, it performs pretraining on labeled data $(x, y) \sim P$ and optimizes $\mathcal{L}_{label}$ as in Eq. 1. Secondly, it augments testing samples $x \sim Q$ to train an auxiliary proxy task (e.g., rotation angle prediction [59]) using $\mathcal{L}_{proxy}$ in the testing phase. The two tasks share an encoder but have separate decoders. When $\mathcal{L}_{label}$ and $\mathcal{L}_{proxy}$ have correlated gradients, the encoder keeps evolving during testing, thus adapting to Q. However, the decoder still relies on unavailable $\mathcal{L}_{label}$ supervised by $\hat{y}$.

In contrast, our A2A, in each step, randomly shuffles the original IQ signal $X \sim Q_i$ along the aperture dimension to create a noisy pair (X_1, X_2) in different aperture orders (Fig. 3(b)). A2A proposes a self-supervised swapping loop as the proxy task. As shown in Fig. 4(a), A2A forward propagates X_1 to generate $\tilde{X}_1$ approximating X_2; A2A then swaps the input and target, and uses X_2 to approximate X_1. One loop completes one iteration. A swapping loss is proposed as

$$\mathcal{L}_{swap}(X_1, X_2) = \mathcal{L}_{1\to2}(\tilde{X}_1, X_2) + \mathcal{L}_{2\to1}(\tilde{X}_2, X_1), \tag{6}$$

which is calculated by both coherence and incoherence L_2 distances. Using $\mathcal{L}_{1\to2}$ as an example,

$$\mathcal{L}_{1\to2}(\tilde{X}_1, X_2) = (\tilde{X}_1 - X_2)^2 + (|\tilde{X}_1| - |X_2|)^2, \tag{7}$$

where $|\cdot|$ calculates the magnitude of complex-valued data.

In Fig. 4 (b) and Eq. 5, the test-time trained A2A infers the clean $\hat{X}$ by discarding f_n and decoding exclusively from the anatomy space encoded by f_a.

3.3. Pyramid Self-contrastive Learning

The goal of A2A is to disentangle the low-rank anatomy space and high-rank noise space from the original X. Contrastive learning (CL) provides a tool for extracting similarities and differences among data [60]. As shown in Fig. 3(c), let $x \sim P$ be an anchor, conventional CL augments it to create positive samples and regards other data in P as negative samples [61]. A contrastive loss ($\mathcal{L}_{con}$), such as NCE [62] and InfoNCE [63], maximizes the similarity between projected features from (anchor, positive) pairs and minimizes that from the (anchor, negative) pairs. Even though $\mathcal{L}_{con}$ enhances the encoder's feature representation, the decoder still relies on unavailable $\mathcal{L}_{label}$ supervised by $\hat{y}$.

In our A2A process, a pyramid self-contrastive learning (PSCL) is proposed to extract similarities and differences between (X_1, X_2) pairs. f_a and f_n consist of K layers with progressively larger receptive fields. In one iteration step (Fig. 3(d)), f_a and f_n encode X_1 and X_2 into an anatomy space and a noise space:

$$\mathcal{A}_{1/2} = \left\{ \mathcal{A}^k_{1/2} \mid k = 1 \dots K \right\} = f_{\mathrm{N}}(f_{\mathrm{F}}(f_a(X_{1/2}))), \tag{8}$$

$$\eta_{1/2} = \left\{ \eta^k_{1/2} \mid k = 1 \dots K \right\} = f_{\mathrm{N}}(f_{\mathrm{F}}(f_n(X_{1/2}))), \tag{9}$$

where $f_{\mathrm{F}} : \mathbb{C}^{N_k \times H_k \times W_k} \mapsto \mathbb{R}^{N_k \times H_k W_k}$ flattens 2D complex feature maps with $H_k \times W_k$ shape into 1D modulus embeddings with $H_k W_k$ length; $f_{\mathrm{N}}(\cdot)$ aligns all embeddings into a unit hypersphere using L_2 normalization; $\mathcal{A}^k$ and η^k denote embeddings from the k-th layer in f_a and f_n, respectively; $\mathcal{A}_1$ and $\mathcal{A}_2$ represent two feature pyramids in the anatomy space, η_1 and η_2 represent two feature pyramids in the noise space.

As described in section 3.2.1, the objective of PSCL is to pull $\mathcal{A}_1$ and $\mathcal{A}_2$ closer while pushing η_1 and η_2 away. We

Table 3
Denoising performance of A2A against 7 comparison methods on simulation data under 4 electronic noise levels.

Metrics	Noise Level	Original Image	Denoising Methods (↑ or ↓ in Percentage)							
			A2A	CF	PCF	SRAD	BM3D	DIP	N2N	N2S
CNR	*0 dB*	6.56	9.29 (42% ↑)	4.23 (35% ↓)	5.04 (23% ↓)	8.06 (23% ↑)	8.33 (27% ↑)	7.77 (19% ↑)	7.34 (12% ↑)	7.51 (14% ↑)
	10 dB	7.01	8.42 (20% ↑)	4.85 (31% ↓)	5.78 (18% ↓)	7.78 (11% ↑)	7.71 (10% ↑)	7.75 (11% ↑)	7.40 (5% ↑)	7.19 (3% ↑)
	20 dB	7.00	9.58 (37% ↑)	4.93 (30% ↓)	5.87 (16% ↓)	7.48 (7% ↑)	7.60 (9% ↑)	7.36 (5% ↑)	7.21 (3% ↑)	7.05 (1% ↑)
	30 dB	6.99	9.62 (37% ↑)	4.93 (30% ↓)	5.86 (16% ↓)	7.45 (6% ↑)	7.60 (9% ↑)	7.39 (6% ↑)	7.17 (3% ↑)	7.05 (1% ↑)
gCNR	*0 dB*	0.845	0.896 (6% ↑)	0.742 (12% ↓)	0.776 (8% ↓)	0.877 (4% ↑)	0.897 (6% ↑)	0.868 (3% ↑)	0.852 (1% ↑)	0.849 (-)
	10 dB	0.848	0.866 (2% ↑)	0.761 (10% ↓)	0.791 (7% ↓)	0.877 (3% ↑)	0.896 (6% ↑)	0.873 (3% ↑)	0.852 (-)	0.844 (1% ↓)
	20 dB	0.848	0.899 (6% ↑)	0.762 (10% ↓)	0.792 (7% ↓)	0.876 (3% ↑)	0.896 (6% ↑)	0.862 (2% ↑)	0.853 (1% ↑)	0.841 (1% ↓)
	30 dB	0.848	0.895 (6% ↑)	0.762 (10% ↓)	0.792 (7% ↓)	0.876 (3% ↑)	0.895 (6% ↑)	0.869 (2% ↑)	0.852 (-)	0.844 (-)
SNR	*0 dB*	5.83	12.36 (112% ↑)	9.69 (66% ↑)	9.17 (57% ↑)	6.24 (7% ↑)	6.83 (17% ↑)	6.60 (13% ↑)	5.98 (3% ↑)	6.63 (14% ↑)
	10 dB	7.25	11.81 (63% ↑)	10.51 (45% ↑)	10.11 (39% ↑)	7.83 (8% ↑)	8.51 (17% ↑)	8.10 (12% ↑)	7.38 (2% ↑)	7.99 (10% ↑)
	20 dB	7.63	11.29 (48% ↑)	10.60 (39% ↑)	10.21 (34% ↑)	8.14 (7% ↑)	8.67 (14% ↑)	8.26 (8% ↑)	7.70 (1% ↑)	8.16 (7% ↑)
	30 dB	7.67	11.84 (54% ↑)	10.60 (38% ↑)	10.21 (33% ↑)	8.17 (6% ↑)	8.68 (13% ↑)	8.31 (8% ↑)	7.73 (1% ↑)	8.19 (7% ↑)
PSNR	*0 dB*	12.48	24.10 (93% ↑)	19.37 (55% ↑)	18.66 (50% ↑)	13.73 (10% ↑)	15.16 (21% ↑)	14.40 (15% ↑)	12.71 (2% ↑)	14.46 (16% ↑)
	10 dB	16.15	22.77 (41% ↑)	21.29 (32% ↑)	21.14 (31% ↑)	18.21 (13% ↑)	21.43 (33% ↑)	18.90 (17% ↑)	16.51 (2% ↑)	18.65 (15% ↑)
	20 dB	17.91	23.33 (30% ↑)	21.47 (20% ↑)	21.42 (20% ↑)	19.86 (11% ↑)	21.89 (22% ↑)	20.16 (13% ↑)	18.13 (1% ↑)	19.70 (10% ↑)
	30 dB	18.19	23.95 (32% ↑)	21.44 (18% ↑)	21.36 (17% ↑)	20.00 (10% ↑)	7.60 (58% ↓)	20.28 (11% ↑)	18.34 (1% ↑)	19.97 (10% ↑)
SSIM	*0 dB*	0.195	0.982 (403% ↑)	0.286 (46% ↑)	0.239 (22% ↑)	0.201 (3% ↑)	0.207 (6% ↑)	0.201 (3% ↑)	0.197 (1% ↑)	0.201 (3% ↑)
	10 dB	0.204	0.897 (339% ↑)	0.597 (192% ↑)	0.516 (152% ↑)	0.240 (17% ↑)	0.709 (247% ↑)	0.244 (19% ↑)	0.209 (2% ↑)	0.248 (21% ↑)
	20 dB	0.265	0.980 (269% ↑)	0.732 (176% ↑)	0.688 (159% ↑)	0.473 (78% ↑)	0.876 (230% ↑)	0.451 (70% ↑)	0.275 (4% ↑)	0.411 (55% ↑)
	30 dB	0.325	0.979 (201% ↑)	0.734 (126% ↑)	0.691 (113% ↑)	0.527 (62% ↑)	0.895 (176% ↑)	0.522 (61% ↑)	0.331 (2% ↑)	0.497 (53% ↑)

Table 4
Denoising performance of A2A against 7 comparison methods on simulation data with 2 inclusion geometries.

Inclusion Geometrics	Metrics	Original Image	Denoising Methods (↑ or ↓ in Percentage)							
			A2A	CF	PCF	SRAD	BM3D	DIP	N2N	N2S
Star	*CNR*	8.09	10.62 (31% ↑)	5.50 (32% ↓)	6.70 (17% ↓)	8.93 (10% ↑)	9.14 (13% ↑)	8.92 (10% ↑)	8.45 (4% ↑)	8.39 (4% ↑)
	gCNR	0.884	0.923 (4% ↑)	0.795 (10% ↓)	0.833 (6% ↓)	0.912 (3% ↑)	0.925 (5% ↑)	0.907 (3% ↑)	0.888 (-)	0.877 (1% ↓)
	SNR	8.33	12.82 (54% ↑)	11.87 (42% ↑)	11.47 (38% ↑)	8.98 (8% ↑)	9.76 (17% ↑)	9.26 (11% ↑)	8.45 (1% ↑)	9.16 (10% ↑)
	PSNR	17.22	24.72 (44% ↑)	22.53 (31% ↑)	22.24 (29% ↑)	19.30 (12% ↑)	18.28 (6% ↑)	19.83 (15% ↑)	17.48 (2% ↑)	19.54 (13% ↑)
	SSIM	0.261	0.984 (278% ↑)	0.635 (144% ↑)	0.580 (123% ↑)	0.390 (50% ↑)	0.725 (178% ↑)	0.387 (48% ↑)	0.267 (2% ↑)	0.371 (42% ↑)
Circle	*CNR*	5.69	7.83 (38% ↑)	3.97 (30% ↓)	4.57 (20% ↓)	6.46 (13% ↑)	6.48 (14% ↑)	6.21 (9% ↑)	6.11 (7% ↑)	6.01 (6% ↑)
	gCNR	0.811	0.855 (5% ↑)	0.718 (11% ↓)	0.742 (8% ↓)	0.841 (4% ↑)	0.867 (7% ↑)	0.829 (2% ↑)	0.816 (1% ↑)	0.813 (-)
	SNR	5.86	10.82 (85% ↑)	8.83 (51% ↑)	8.38 (43% ↑)	6.20 (6% ↑)	6.58 (12% ↑)	6.37 (9% ↑)	5.95 (2% ↑)	6.32 (8% ↑)
	PSNR	15.15	22.36 (48% ↑)	19.26 (27% ↑)	19.06 (26% ↑)	16.61 (10% ↑)	14.76 (-3% ↑)	17.04 (12% ↑)	15.36 (1% ↑)	16.84 (11% ↑)
	SSIM	0.235	0.936 (299% ↑)	0.539 (130% ↑)	0.487 (108% ↑)	0.330 (41% ↑)	0.618 (164% ↑)	0.323 (38% ↑)	0.240 (2% ↑)	0.307 (31% ↑)

formulate this process as dense contrastive learning [64] to facilitate local feature perception. In one A2A step, $(\mathcal{A}_1^k, \mathcal{A}_2^k)$ pair is naturally aligned because they come from the same encoder, so does the (η_1^k, η_2^k) pair. Let cosine similarity be $\mathrm{Cos}(\cdot,\cdot)$. The objective of one PSCL iteration becomes Eq. 10 and Eq. 11. For features belonging to the same $X_{1/2}$, PSCL also pushes $\mathcal{A}_{1/2}$ away from $\eta_{1/2}$ to reduce feature redundancy, as in Eq. 12.

$$\arg\max_{f_a} \sum_{k=1}^{K} \mathrm{Cos}(\mathcal{A}_1^k, \mathcal{A}_2^k) = \sum_{k=1}^{K} \frac{\mathcal{A}_1^k \cdot \mathcal{A}_2^k}{\left\|\mathcal{A}_1^k\right\| \cdot \left\|\mathcal{A}_2^k\right\|}, \tag{10}$$

$$\arg\min_{f_n} \sum_{k=1}^{K} \mathrm{Cos}(\eta_1^k, \eta_2^k) = \sum_{k=1}^{K} \frac{\eta_1 \cdot \eta_2}{\left\|\eta_1^k\right\| \cdot \left\|\eta_2^k\right\|}, \tag{11}$$

$$\arg\min_{f_a, f_n} \sum_{i=1}^{2} \sum_{k=1}^{K} \mathrm{Cos}(\mathcal{A}_i^k, \eta_i^k) = \sum_{i=1}^{2} \sum_{k=1}^{K} \frac{\mathcal{A}_i^k \cdot \eta_i^k}{\left\|\mathcal{A}_i^k\right\| \cdot \left\|\eta_i^k\right\|}. \tag{12}$$

A2A optimizes this feature-level dense CL by a customized cross-entropy loss:

$$\mathcal{L}_{con}(\mathcal{A}_1, \mathcal{A}_2, \eta_1, \eta_2) = \sum_{k=1}^{K} \frac{-1}{N_k} \log \left(\frac{e^{\mathrm{Cos}(\mathcal{A}_1^k, \mathcal{A}_2^k)}}{e^{\mathrm{Cos}(\mathcal{A}_1^k, \mathcal{A}_2^k)} + e^{\mathrm{Cos}(\eta_1^k, \eta_2^k)} + \sum_{i=1}^{2} e^{\mathrm{Cos}(\mathcal{A}_i^k, \eta_i^k)}} \right). \tag{13}$$

A2A's overall objective is then formulated as

$$\arg\min_{f_a, f_n, f_d} \mathbb{E}_{(X_1, X_2)\sim Q_i} \left[\mathcal{L}_{swap}(X_1, X_2) + \mathcal{L}_{con}(\mathcal{A}_1, \mathcal{A}_2, \eta_1, \eta_2) \right], \tag{14}$$

which enables disentangling low-rank anatomical components from X itself and reconstructing them into the underlying clean signals.

3.4. Lightweight Architecture

In theory, f_a, f_n, f_d are compatible with any architecture. We build a lightweight dual-head UNet to facilitate efficiency (Fig. 4(a)). f_a and f_n have three layers ($K = 3$) with channel numbers $N_1, N_2, N_3 = 16, 32, 64$. f_d is a

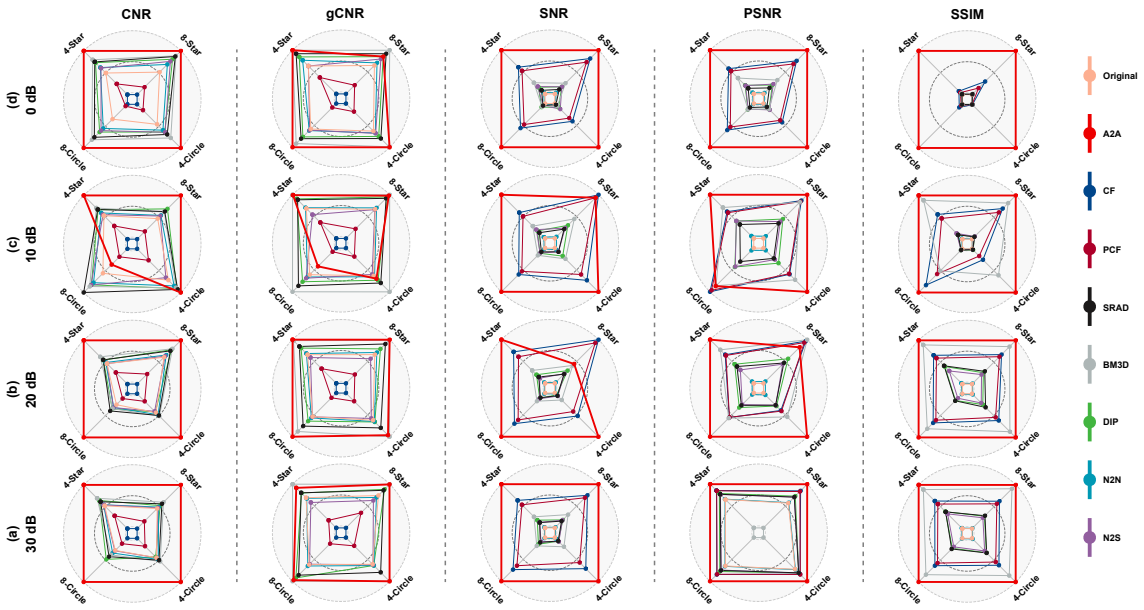

Figure 6: The radar plots of CNR, gCNR, SNR, PSNR, and SSIM on simulation data under four noise levels. For each metric, results of four or eight apertures and star or circle inclusion are shown at four corners, respectively.

decoder asymmetrical to f_a. A2A uses summation for skip connections, which connects $\mathcal{A}^k_{1/2} + \eta^k_{1/2}$ to implement Eq. 4, and connects only $\mathcal{A}^k_{1/2}$ to implement Eq. 5. A2A is built upon pure complex neural networks to process IQ data [65] using complex convolution and complex maxpooling, etc.

4. Experiments

4.1. Simulation Setup

We used the k-Wave toolbox [66] to simulate SAU images in the P4-2 phased array configuration. As Table 1, the full aperture of 64 elements was divided into four and eight sub-apertures to assess performance under different imaging configurations. SNR levels of 0, 10, 20, and 30 dB were simulated to evaluate robustness to electronic noise. To consider targets of different complexities, we simulated circular and star-shaped inclusions whose acoustic properties were with reference to those of the heart muscle (Table 2).

4.2. *In Vivo* Data Collection

We collected a large *in vivo* SAU dataset consisting of echocardiographic and abdominal images. The echocardiographic set includes six standard views (Fig. 2(a)): apical four-chamber (A4C), two-chamber (A2C), parasternal short-axis at the mitral valve (PSAX-MV), papillary muscle (PSAX-Pap), and apical (PSAX-Apex) levels, and parasternal long-axis (PLAX) views [67]. There were 16.4k, 16k, 13.6k, 16.48k, 16.24k, and 15.59k images for these views, respectively. The abdominal set includes 4.8k images of the liver and kidney. The cohort comprises 75 healthy subjects and 21 pathological subjects with hypertension and/or diabetes. Two sonographers performed the ultrasound scans. This diverse dataset supports evaluations under domain shifts across scan views, organs, subject conditions, and operators.

IQ signals were acquired using a Vantage 256 system (Verasonics, Kirkland, WA, USA) equipped with a P4-2 phased array[1]. The number of sub-apertures varied from two to eight, as summarized in Table 1. Each sub-aperture frame was scan-converged and resized to 512×512 pixels.

4.3. Implementation Details

A2A was developed with Python 3.10, PyTorch 2.1.2, CUDA 11.8, and deployed on an Intel Xeon CPU and an NVIDIA RTX 3090 GPU with <6029 MiB memory. A2A was trained online at test time using an Adam optimizer configured with L_2 regularization until $\mathcal{L}_{con}$ reached a plateau.

4.4. Comparison Methods

We compared A2A with seven representative ultrasound denoising methods described in related works (section 2). They include three types of techniques: (1) model-based SRAD [8] and BM3D [7]; (2) coherence-based CF [14] and PCF [16]; (3) learning-based and self-supervised DIP [54], N2N [46], and N2S [55]. Denoising performance was quantified by CNR, gCNR, SNR, PSNR, and SSIM metrics on diverse image domains.

5. Results

5.1. Simulation Results

5.1.1. Different noise levels

Table 3 summarizes the results of simulation data under four electronic noise levels. When the original images suffer from severe electronic noise, i.e., at 0 dB SNR as shown

[1]Human experiment protocols were approved by the Institutional Review Board of The University of Hong Kong (UW19-043, EA250292).

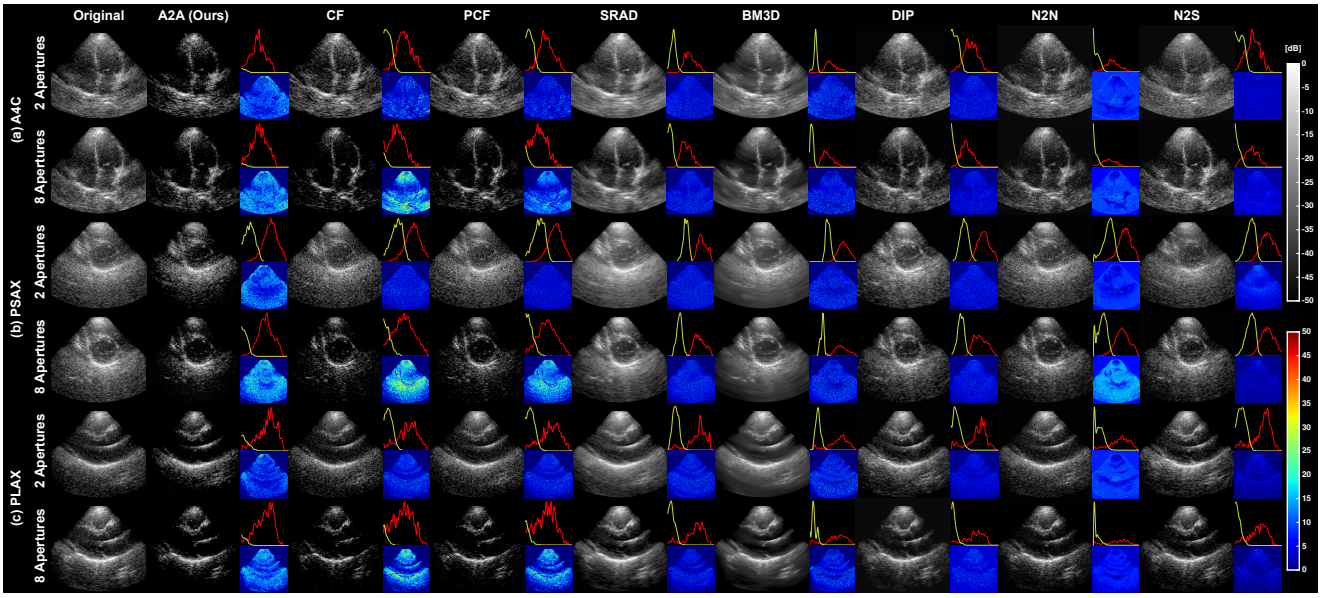

Figure 7: Denoising comparison on *in vivo* echocardiograms of (a) A4C, (b) PSAX, and (c) PLAX views. The imaging configuration includes two and eight apertures for each view. The visualization layout is identical to that in Fig. 5.

Table 5
The SNRs achieved by A2A and comparison methods on different *in vivo* organs or scan views with two or eight sub-apertures.

Aperture No.	Organ	Scan View	Original	A2A	CF	PCF	SRAD	BM3D	DIP	N2N	N2S
2 (min)	*Heart*	*A4C*	8.27	17.50 (112% ↑)	9.20 (11% ↑)	8.88 (7% ↑)	7.07 (14% ↓)	7.05 (15% ↓)	7.66 (7% ↓)	7.98 (4% ↓)	7.14 (14% ↓)
		A2C	8.03	13.04 (62% ↑)	8.75 (9% ↑)	8.48 (6% ↑)	7.00 (13% ↓)	7.07 (12% ↓)	7.97 (1% ↓)	7.65 (5% ↓)	8.36 (4% ↑)
		PSAX-MV	7.88	14.06 (78% ↑)	8.50 (8% ↑)	8.30 (5% ↑)	7.01 (11% ↓)	6.95 (12% ↓)	8.44 (7% ↑)	7.71 (2% ↓)	8.14 (3% ↑)
		PSAX-Pap	6.90	13.20 (91% ↑)	7.37 (7% ↑)	7.18 (4% ↑)	6.08 (12% ↓)	6.02 (13% ↓)	7.01 (2% ↑)	6.79 (2% ↓)	7.80 (13% ↑)
		PSAX-Apex	8.32	13.82 (66% ↑)	8.90 (7% ↑)	8.67 (4% ↑)	7.32 (12% ↓)	7.26 (13% ↓)	8.04 (3% ↓)	8.18 (2% ↓)	8.74 (5% ↑)
		PLAX	8.26	15.06 (82% ↑)	8.94 (8% ↑)	8.75 (6% ↑)	7.49 (9% ↓)	7.42 (10% ↓)	7.89 (5% ↓)	8.10 (2% ↓)	8.69 (5% ↑)
	Liver		12.99	24.45 (88% ↑)	13.80 (6% ↑)	13.60 (5% ↑)	11.79 (9% ↓)	11.47 (12% ↓)	14.42 (11% ↑)	12.43 (4% ↓)	14.37 (11% ↑)
	Kidney		10.56	19.78 (87% ↑)	11.58 (10% ↑)	11.20 (6% ↑)	8.97 (15% ↓)	8.97 (15% ↓)	10.23 (3% ↓)	10.33 (2% ↓)	10.20 (3% ↓)
8 (max)	*Heart*	*A4C*	10.66	16.64 (56% ↑)	18.52 (74% ↑)	17.30 (62% ↑)	9.82 (8% ↓)	10.07 (5% ↓)	11.13 (4% ↑)	9.95 (7% ↓)	10.70 (-)
		A2C	12.56	19.77 (57% ↑)	21.48 (71% ↑)	19.68 (57% ↑)	11.15 (11% ↓)	11.59 (8% ↓)	13.26 (6% ↑)	11.40 (9% ↓)	12.62 (1% ↑)
		PSAX-MV	11.81	17.82 (51% ↑)	19.47 (65% ↑)	18.05 (53% ↑)	10.57 (11% ↓)	10.67 (10% ↓)	12.45 (5% ↑)	11.20 (5% ↓)	11.93 (1% ↑)
		PSAX-Pap	10.39	17.56 (69% ↑)	17.57 (69% ↑)	16.19 (56% ↑)	9.37 (10% ↓)	9.39 (10% ↓)	10.69 (3% ↑)	10.01 (4% ↓)	10.66 (3% ↑)
		PSAX-Apex	11.15	17.18 (54% ↑)	18.31 (64% ↑)	17.02 (53% ↑)	10.18 (9% ↓)	10.20 (9% ↓)	11.73 (5% ↑)	10.74 (4% ↓)	10.99 (1% ↓)
		PLAX	11.58	17.45 (51% ↑)	18.89 (63% ↑)	17.71 (53% ↑)	10.72 (7% ↓)	10.81 (7% ↓)	11.77 (2% ↑)	11.04 (5% ↓)	11.43 (1% ↓)
	Liver		19.43	22.87 (18% ↑)	28.48 (47% ↑)	26.95 (39% ↑)	17.98 (7% ↓)	16.98 (13% ↓)	21.82 (12% ↑)	16.42 (16% ↓)	19.53 (-)
	Kidney		15.50	21.39 (38% ↑)	25.13 (62% ↑)	24.33 (57% ↑)	14.23 (8% ↓)	14.82 (4% ↓)	16.99 (10% ↑)	13.62 (12% ↓)	15.57 (1% ↑)

in Fig. 5(a), our A2A improved CNR by 41.6%, gCNR by 6% gCNR, SNR by 111.8%, PSNR by 93.1%, and SSIM by about fourfold. When the original images are dominated by the speckle noise, i.e., at SNR of 30 dB in Fig. 5(b), our A2A improved CNR by 37.5%, gCNR by 5.6%, SNR by 54.3%, PSNR by 31.7%, and SSIM by about twofold. At SNRs of 10 dB and 20 dB, the A2A also shows consistent performance and surpasses all comparison methods.

To evaluate the overall contrast improvement, CNR values were averaged from four noise levels. The ranking in the average increase in CNR was as follows: A2A > BM3D > SRAD > DIP > N2N > N2S > PCF > CF. From the perspective of denoised signal quality, A2A ranked first in SNR gain, followed by CF, PCF, BM3D, DIP, N2S, SRAD, and N2N. In terms of structural detail preservation, the SSIM increase was highest in the case of A2A, sequentially followed by BM3D, CF, PCF, SRAD, DIP, N2S, and N2N.

Fig. 5(a) shows that at 0 dB, the SRAD and BM3D still yielded noisy backgrounds, DIP suffered from random spots, and N2N generated radial textures. At 30 dB (Fig. 5(b)), CF, PCF, SRAD, DIP, and N2S could not effectively reduce strong speckles in the background. These differences highlighted the domain gap of these methods when encountering various noise levels. However, our A2A produced a clean background and was robust to different noise levels.

5.1.2. Different inclusion geometries

Table 4 summarizes the denoising performance across four different noise levels for the simulated star and circle geometries. Compared with the original images, A2A significantly improved CNR, SNR, PSNR, and SSIM. A2A outperformed the seven comparison methods and demonstrated its robust adaptation to both simple and complex targets, consistent with the findings from the electronic noise level analysis (Table 3). Regardless of the imaging target,

Table 6
CNR analysis of A2A and seven comparison methods across different *in vivo* organs or scan views with two or eight sub-apertures.

Aperture No.	Organ	Scan View	Original	A2A	CF	PCF	SRAD	BM3D	DIP	N2N	N2S
2 (min)	*Heart*	*A4C*	6.30	9.01 (43% ↑)	5.51 (13% ↓)	5.93 (6% ↓)	7.07 (12% ↑)	9.34 (48% ↑)	5.80 (8% ↓)	6.30 (-)	4.71 (25% ↓)
		A2C	5.50	6.84 (24% ↑)	4.09 (26% ↓)	4.71 (14% ↓)	8.71 (58% ↑)	8.43 (53% ↑)	6.18 (12% ↑)	5.46 (1% ↓)	6.07 (10% ↑)
		PSAX-MV	5.62	7.87 (40% ↑)	4.47 (20% ↓)	5.12 (9% ↓)	9.37 (67% ↑)	8.69 (55% ↑)	7.03 (25% ↑)	5.64 (-)	5.67 (1% ↑)
		PSAX-Pap	5.15	7.44 (44% ↑)	3.53 (32% ↓)	4.28 (17% ↓)	8.63 (68% ↑)	7.75 (50% ↑)	5.40 (5% ↑)	5.17 (-)	6.78 (32% ↑)
		PSAX-Apex	7.19	9.25 (29% ↑)	5.72 (20% ↓)	6.34 (12% ↓)	10.08 (40% ↑)	9.56 (33% ↑)	7.80 (8% ↑)	7.25 (1% ↑)	8.04 (12% ↑)
		PLAX	6.50	9.13 (41% ↑)	5.31 (18% ↓)	6.03 (7% ↓)	10.04 (55% ↑)	9.27 (43% ↑)	5.81 (11% ↓)	6.55 (1% ↑)	6.99 (12% ↑)
	Liver		6.60	6.94 (5% ↑)	5.64 (14% ↓)	6.19 (6% ↓)	9.65 (46% ↑)	9.28 (41% ↑)	7.83 (19% ↑)	8.30 (26% ↑)	7.44 (13% ↑)
	Kidney		6.92	8.58 (24% ↑)	6.07 (12% ↓)	6.38 (8% ↓)	9.62 (39% ↑)	9.66 (40% ↑)	7.61 (10% ↑)	8.76 (27% ↑)	6.57 (5% ↓)
8 (max)	*Heart*	*A4C*	6.28	7.54 (20% ↑)	3.82 (39% ↓)	4.67 (26% ↓)	9.82 (56% ↑)	9.61 (53% ↑)	7.39 (18% ↑)	6.14 (2% ↓)	5.80 (8% ↓)
		A2C	7.93	9.97 (26% ↑)	5.04 (36% ↓)	5.98 (25% ↓)	11.00 (39% ↑)	10.77 (36% ↑)	8.92 (13% ↑)	7.76 (2% ↓)	8.07 (2% ↑)
		PSAX-MV	7.79	9.54 (22% ↑)	4.28 (45% ↓)	5.41 (31% ↓)	10.58 (36% ↑)	10.24 (31% ↑)	9.31 (19% ↑)	7.70 (1% ↓)	8.21 (5% ↑)
		PSAX-Pap	7.45	9.22 (24% ↑)	4.69 (37% ↓)	5.73 (23% ↓)	10.34 (39% ↑)	9.79 (31% ↑)	8.30 (11% ↑)	7.42 (-)	8.15 (9% ↑)
		PSAX-Apex	8.22	10.14 (23% ↑)	5.07 (38% ↓)	6.29 (23% ↓)	11.11 (35% ↑)	10.71 (30% ↑)	9.53 (16% ↑)	8.20 (-)	8.52 (4% ↑)
		PLAX	7.75	9.09 (17% ↑)	4.14 (47% ↓)	5.57 (28% ↓)	10.57 (36% ↑)	9.99 (29% ↑)	8.34 (8% ↑)	7.68 (1% ↓)	7.58 (2% ↓)
	Liver		7.29	9.46 (30% ↑)	2.90 (60% ↓)	3.73 (49% ↓)	10.23 (40% ↑)	9.69 (33% ↑)	8.46 (16% ↑)	8.47 (16% ↑)	8.13 (11% ↑)
	Kidney		8.39	10.23 (22% ↑)	5.12 (39% ↓)	5.59 (33% ↓)	11.93 (42% ↑)	12.27 (46% ↑)	9.93 (18% ↑)	10.11 (21% ↑)	9.04 (8% ↑)

Table 7
gCNR analysis of A2A and seven comparison methods across different *in vivo* organs or scan views with two or eight sub-apertures.

Aperture No.	Organ	Scan View	Original	A2A	CF	PCF	SRAD	BM3D	DIP	N2N	N2S
2 (min)	*Heart*	*A4C*	0.871	0.917 (5% ↑)	0.812 (7% ↓)	0.832 (4% ↓)	0.974 (12% ↑)	0.971 (11% ↑)	0.835 (4% ↓)	0.868 -)	0.775 (11% ↓)
		A2C	0.825	0.848 (3% ↑)	0.750 (9% ↓)	0.768 (7% ↓)	0.948 (15% ↑)	0.944 (14% ↑)	0.836 (1% ↑)	0.821 (-)	0.844 (2% ↑)
		PSAX-MV	0.828	0.879 (6% ↑)	0.770 (7% ↓)	0.796 (4% ↓)	0.949 (15% ↑)	0.945 (14% ↑)	0.861 (1% ↑)	0.825 (-)	0.815 (2% ↓)
		PSAX-Pap	0.808	0.864 (7% ↑)	0.721 (11% ↓)	0.750 (7% ↓)	0.941 (16% ↑)	0.915 (13% ↑)	0.816 (1% ↑)	0.804 (-)	0.873 (8% ↑)
		PSAX-Apex	0.880	0.902 (2% ↑)	0.798 (9% ↓)	0.833 (5% ↓)	0.964 (10% ↑)	0.953 (8% ↑)	0.902 (3% ↑)	0.877 (-)	0.910 (3% ↑)
		PLAX	0.864	0.912 (6% ↑)	0.808 (6% ↓)	0.835 (3% ↓)	0.965 (12% ↑)	0.957 (11% ↑)	0.825 (5% ↓)	0.861 (-)	0.870 (1% ↑)
	Liver		0.877	0.896 (2% ↑)	0.838 (4% ↓)	0.857 (2% ↓)	0.974 (11% ↑)	0.983 (12% ↑)	0.922 (5% ↑)	0.944 (8% ↑)	0.906 (3% ↑)
	Kidney		0.881	0.918 (4% ↑)	0.836 (5% ↓)	0.846 (4% ↓)	0.974 (11% ↑)	0.995 (13% ↑)	0.879 (-)	0.941 (7% ↑)	0.836 (5% ↓)
8 (max)	*Heart*	*A4C*	0.853	0.871 (2% ↑)	0.772 (9% ↓)	0.800 (6% ↓)	0.967 (13% ↑)	0.971 (14% ↑)	0.894 (5)	0.850 (-)	0.824 (3% ↓)
		A2C	0.920	0.950 (3% ↑)	0.832 (10% ↓)	0.860 (6% ↓)	0.991 (8% ↑)	0.993 (8% ↑)	0.943 (3% ↑)	0.917 (-)	0.920 (-)
		PSAX-MV	0.907	0.935 (3% ↑)	0.799 (12% ↓)	0.836 (8% ↓)	0.984 (8% ↑)	0.994 (10% ↑)	0.963 (6% ↑)	0.905 (-)	0.923 (2% ↑)
		PSAX-Pap	0.900	0.933 (4% ↑)	0.796 (12% ↓)	0.831 (8% ↓)	0.983 (9% ↑)	0.984 (9% ↑)	0.927 (3% ↑)	0.896 (-)	0.922 (2% ↑)
		PSAX-Apex	0.908	0.930 (2% ↑)	0.798 (12% ↓)	0.839 (8% ↓)	0.982 (8% ↑)	0.985 (8% ↑)	0.944 (4% ↑)	0.905 (-)	0.921 (1% ↑)
		PLAX	0.907	0.927 (2% ↑)	0.790 (13% ↓)	0.840 (7% ↓)	0.982 (8% ↑)	0.981 (8% ↑)	0.926 (2% ↑)	0.904 (-)	0.895 (1% ↓)
	Liver		0.917	0.943 (3% ↑)	0.791 (14% ↓)	0.825 (10% ↓)	0.991 (8% ↑)	0.994 (8% ↑)	0.970 (6% ↑)	0.975 (6% ↑)	0.950 (4% ↑)
	Kidney		0.927	0.946 (2% ↑)	0.868 (6% ↑)	0.884 (5% ↑)	0.998 (8% ↑)	1.000 (8% ↑)	0.973 (5% ↑)	0.979 (6% ↑)	0.949 (2% ↑)

learning-based methods demonstrate moderate performance across all metrics, while coherence-based methods produced higher SNR but lower CNR, and model-based methods exhibit the opposite trend.

5.1.3. Different numbers of sub-apertures

Fig. 6 shows radar plots for simulation results of four and eight sub-apertures. A2A improved CNR by 42.1% and 50.5%, gCNR by 5.8% and 5.9%, SNR by 98.1% and 100.5%, PSNR by 84.8% and 96.1%, and SSIM both four times using four and eight apertures, respectively. Those results confirmed that A2A could adapt to flexible aperture settings.

As the first two rows in Fig. 5 show, raw images synthesized from fewer sub-apertures exhibited lower signal intensity and less pronounced speckles. This distribution shift caused N2N, which was pretrained on eight sub-aperture images, to fail on four sub-aperture ones. Conversely, A2A maintained robust performance, producing high-contrast outputs with clean backgrounds and sharp boundaries.

With all five metrics and four noise levels included, Fig. 6 demonstrates a balanced and superior performance of A2A regardless of the number of sub-apertures.

5.2. *In Vivo* Results

5.2.1. Different scan views

A2A consistently improved SNRs in six standard echocardiographic views. As listed in Table 5, using only two sub-apertures, the average SNR of A2A was 14.44 dB, with 81.8% higher than the original images. Using full eight sub-apertures, the average SNR achieved 17.74 dB with 56.2% higher than the original images. The box plots in Fig. 8 (a-f) illustrate the remarkably higher SNR distributions, and thus denoising capability, of A2A than the comparison methods.

A2A also exhibited consistent improvements in image contrast in six views. As listed in Tables 6 and 7, using only two sub-apertures, A2A's average CNR and gCNR were 8.25 dB and 0.887, which were 36.6% and 4.8% better than the original images. Using full eight sub-apertures, the average CNR and gCNR were 9.25 dB and 0.924, which were respectively 22.2% and 2.8% better than the original images. The CNR box plots in Fig. 8(g-i) show performance trends similar to those of gCNR in Fig. 8(m-r), where A2A overall manifested relatively higher CNR than other methods.

Fig. 7 visualizes the denoising results of representative A4C, PSAX, and PLAX views. Under full aperture settings, the coherence-based methods (CF and PCF) approached the performance of our A2A. However, their performance

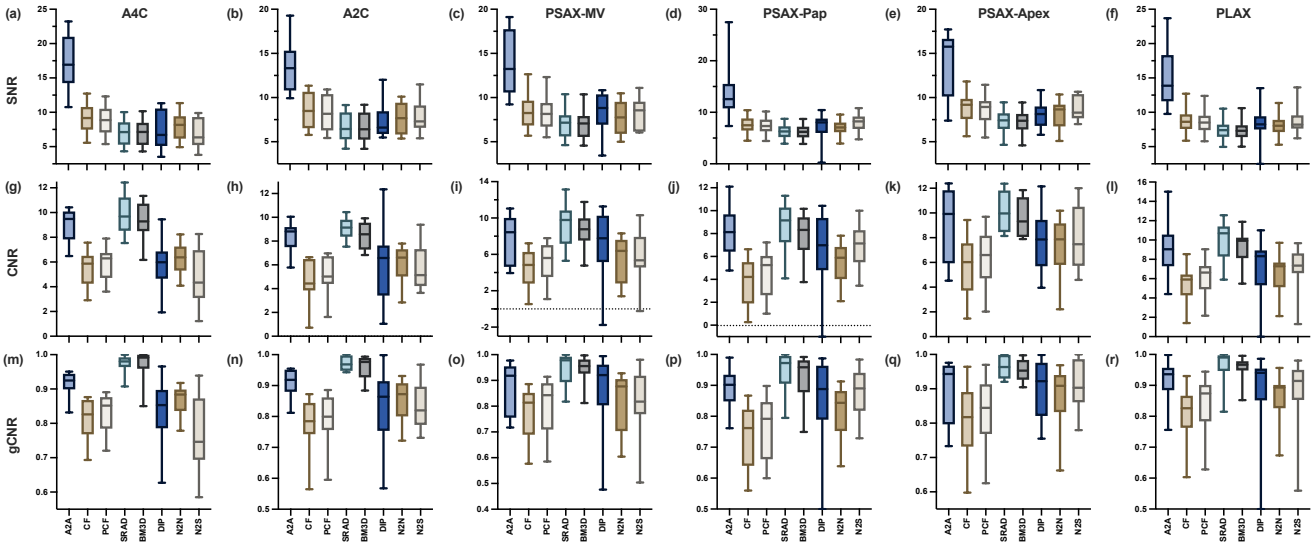

Figure 8: The box plots of (a-f) SNR, (g-l) CNR, and (m-r) gCNR calculated from the original image and the images obtained by A2A, and seven comparison methods on *in vivo* echocardiograms in six different views (A4C, A2C, PSAX-MV, PSAX-Pap, PSAX-Apex, and PLAX).

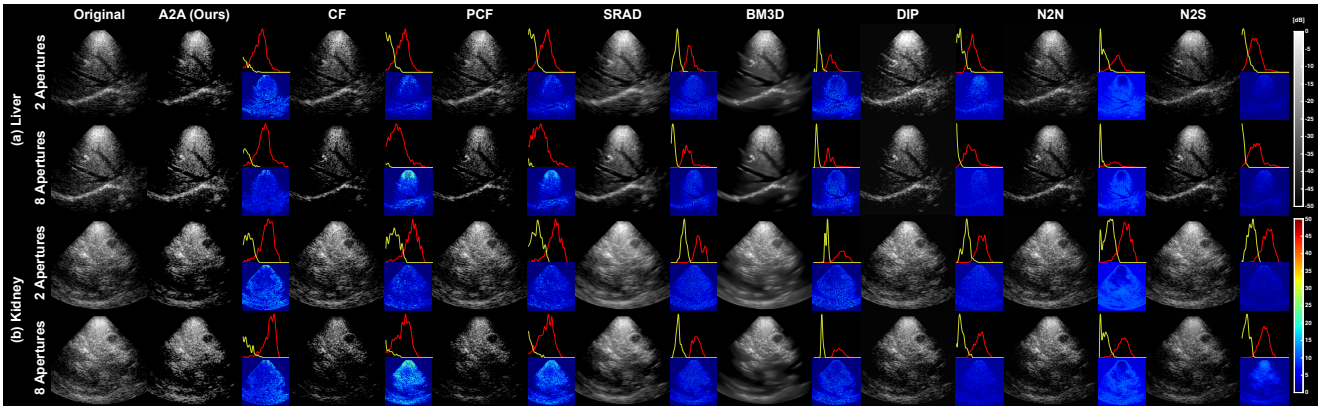

Figure 9: Denoising comparison on *in vivo* abdominal ultrasound of (a) liver and (b) kidney. The imaging configuration includes two and eight sub-apertures for each organ. The visualizing layout is identical to Fig. 5.

declined drastically as the number of sub-apertures dropped to 2. Consistent with the simulation results, model-based methods (SRAD and BM3D) blurred the entire structure and elevated the signal intensity in the heart chambers. Learning-based methods (DIP, N2N, N2S) showed some effectiveness in reducing Gaussian noise but failed to mitigate strong side lobes and speckles.

In contrast, A2A significantly suppressed noise in the cardiac chambers while preserving structural details of the myocardium. Especially in Fig. 7(a), A2A eliminated the side lobes around the atrial and ventricular walls, thereby revealing sharp anatomical boundaries. Electronic noise in deeper regions was also removed by A2A (Fig. 7(b-c)). These evidenced A2A's cross-domain generalization across diverse scan views.

5.2.2. Different imaging organs

In addition to echocardiograms, Tables 5, 6, and 7 summarize SNR, CNR, and gCNR calculated from the *in vivo* liver and kidney images. Under the two sub-aperture settings, A2A improved SNR, CNR, and gCNR on average by 87.8%, 14.9%, and 3.1%, compared to the original images, respectively. Using eight sub-apertures, the corresponding improvements were 26.7%, 25.6%, and 2.5%, respectively.

The exemplary liver and kidney image results in Fig. 9 show performance similar to echocardiograms in Fig. 7. The middle and left hepatic veins (MHV, LHV), as well as the internal structure in the liver ultrasound images, became clearer when subjected to A2A denoising. A2A reduced both electronic and speckle noise while preserving the signal intensity and detailed texture pertaining to important anatomy.

When eight sub-apertures were used, CF and PCF weakened the signal intensity in homogeneous tissue of the liver (Fig. 9 (a)) and almost diminished the fine structure inside the kidney (Fig. 9 (b)). However, they suppressed the clutter signal in the near field and focusing errors caused by phase aberration. SRAD and BM3D blurred the tissue texture and meanwhile generated circular artifacts (9 (b)). Learn-based

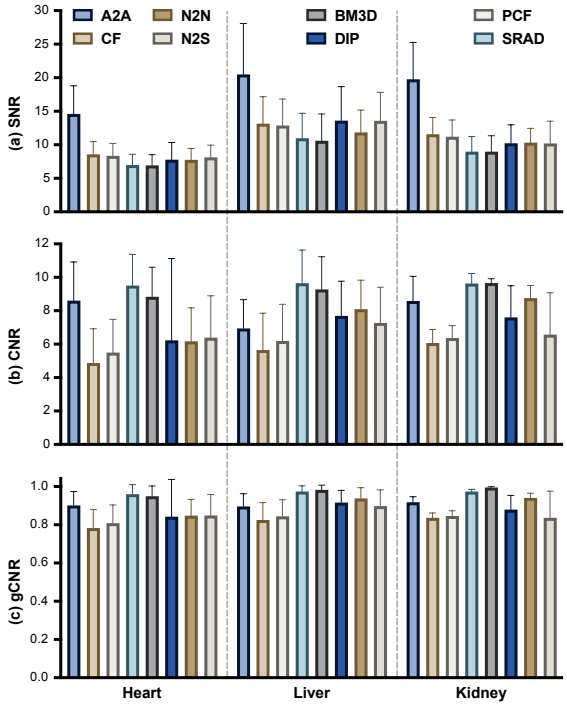

Figure 10: The (a) SNR, (b) CNR, and (c) gCNR among A2A and seven comparison methods on *in vivo* ultrasound images of different organs, including human heart, liver, and kidney.

methods still struggle to remove noise effectively in low SNR environment.

Fig. 10 summarizes the cross-organ results, in which A2A maintains a performance advantage over the comparison methods for *in vivo* heart, liver, and kidney images. These validated A2A's cross-domain generalization across various imaging organs.

6. Discussion

6.1. A2A Test-Time Training

A2A features a pure test-time training framework that is optimized for any test domain Q_i from scratch. DIP [54], likewise free from pretraining, leverages the implicit prior of CNNs that learn salient signals before learning noise. However, this assumption is inadequate for medical ultrasound. Diagnostic details are not necessarily salient in the images, as signals and noise are highly coupled. In contrast, A2A assumes that low-rank anatomical signals and high-rank noise are separable in a high-dimensional space. The convergence of this hypothesis is verified in Fig. 11. Through $\mathcal{L}_{con}$ optimization, f_a and f_n learn anatomy and noise separably without the risk of overtraining.

A2A's TTT is a generalized version of N2N [46] unbounded by the fixed pretraining domain, in which noisy pairs are randomly sampled from a single Q_i via sub-aperture shuffling. Our swapping loop provides self-supervision on output fidelity. In Fig. 11, $\mathcal{L}_{swap}$ converges to an f_d that can reconstruct the separated anatomy and noise features into arbitrary noisy signals $\sim Q_i$ with minimum errors. The

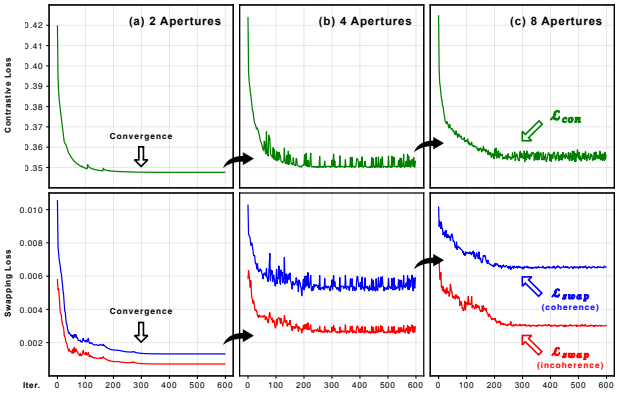

Figure 11: The $\mathcal{L}_{con}$ and $\mathcal{L}_{swap}$ monitored with A2A test-time training process using (a) 2, (b) 4, and (c) 8 apertures.

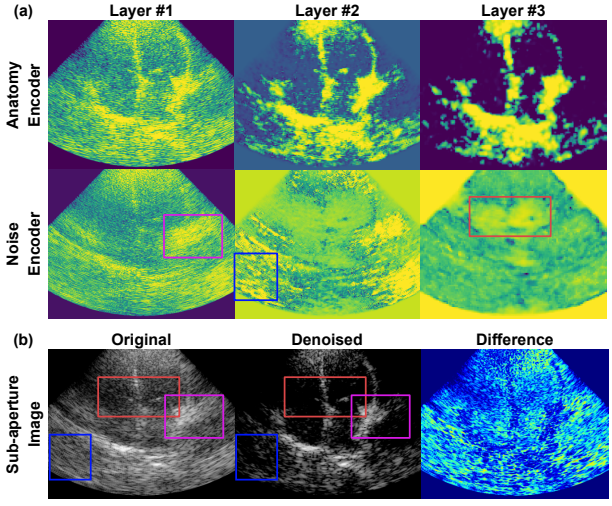

Figure 12: (a) The feature maps extracted by f_a and f_n. (b) The original and denoised image from a single sub-aperture.

clean signal is then reconstructed from the anatomy space using this f_d. Therefore, A2A relaxes the i.i.d condition and can address any noise type that shows distinct characteristics between sub-apertures.

6.2. Self-contrastive Learning (Self-CL)

In theory, the proposed Self-CL maximizes a lower bound on the mutual information (MI) between sub-apertures [68]. $\mathcal{L}_{con}$ converges to two feature spaces—an anatomy space with maximized MI and a noise space with minimized MI, as illustrated in Fig. 12. We visualized feature distributions via t-distributed stochastic neighbor embedding (t-SNE). As Fig. 13(a) shows, the latent space before Self-CL exhibited a random distribution where anatomy and noise features were intermingled. After Self-CL, the features in each layer aggregated into three clusters: (1) an anatomy cluster with overlapping $\mathcal{A}_1^k$ and $\mathcal{A}_2^k$, representing shared information among sub-apertures; (2) a noise cluster with η_1^k for unique noise components in shuffled sub-apertures X_1; (3) another noise cluster with η_2^k for X_2. As highlighted

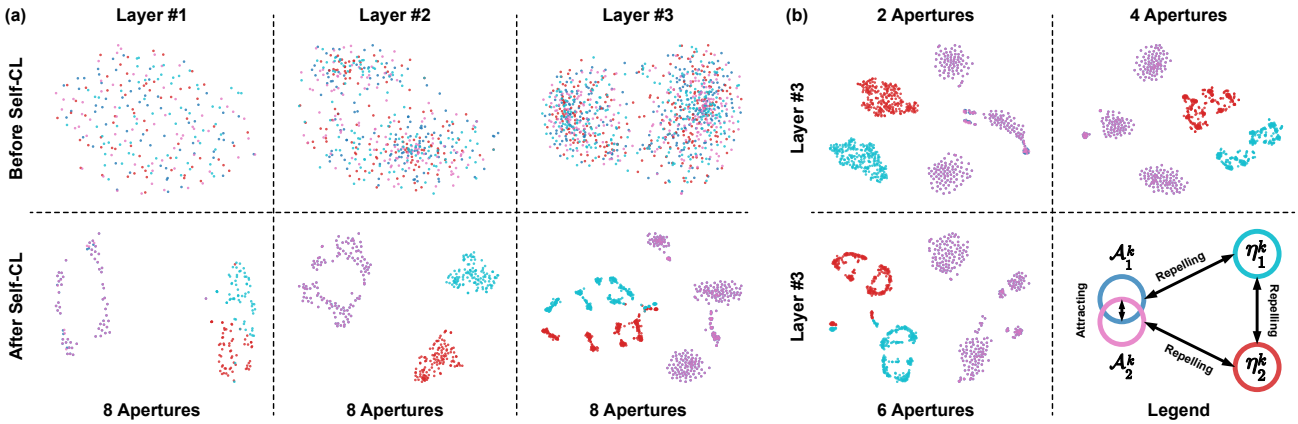

Figure 13: (a) The feature distributions visualized by t-SNE before and after PSCL in the 8 sub-aperture configuration. (b) The feature distributions learned from 2, 4, and 6 sub-apertures in the 3-rd layer. $\mathcal{A}_1^k, \mathcal{A}_2^k, \eta_1^k, \eta_2^k$ are coded in blue, pink, cyan, and red, respectively.

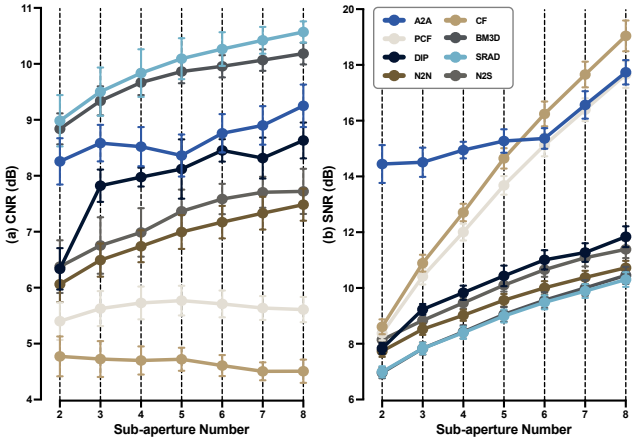

Figure 14: The (a) CNR and (b) SNR of *in vivo* ultrasound images under different sub-aperture settings.

in the legend, the attraction between $(\mathcal{A}_1^k, \mathcal{A}_2^k)$ is optimized by Eq. 10, while the repulsion between (η_1^k, η_2^k) and between $(\mathcal{A}_{1/2}^k, \eta_{1/2}^k)$ are optimized by Eq. 11 and Eq. 12, respectively.

6.3. Pyramid Learning Space

Unlike conventional CL that merely manipulates the deepest features after a projection head [63], the pyramid structure of our learning space enables multi-scale dense representations of anatomy and noise. This is critical for ultrasound images since three noise types manifest at distinct spatial scales. In Fig. 12(a), three layers of f_n hierarchically extract side lobes, speckle, and electronic noise. These noises are hence effectively reduced, as evidenced by the comparison regions in Fig. 12(b). The pyramid learning space of f_a facilitates a coarse-to-fine delineation of shared anatomy in Fig. 12(a), preserving both macro and micro structures in the denoised results.

Fig. 13(a) further reveals hierarchical behaviors across pyramid layers. In the shallowest layer, η_1^k and η_2^k clusters were in close proximity, indicating similar low-level noise (e.g., electronic noise) across sub-apertures. In contrast, the deepest layer exhibits a more intricate distribution that splits into multiple subgroups, suggesting that high-level noise is sensitive to different sub-apertures (e.g., side lobes at different tilt angles).

6.4. Number of Sub-apertures

Intuitively, a greater number of sub-apertures (N) for coherent compounding produces better image quality. However, increasing N is accompanied by a smaller sub-aperture size and thus a lower acoustic energy. It leads to worse initial SNR conditions in each sub-aperture signal and reduces the MI bound between sub-aperture signals, leading to higher convergence levels of $\mathcal{L}_{con}$ and $\mathcal{L}_{swap}$ in Fig. 11. Besides, a larger N implies more complicated patterns to learn. As shown in Fig. 13(b) ($N = 2, 4, 6$) and Fig. 13(a) ($N = 8$), the feature distribution became progressively sophisticated, especially for noise features.

Fig. 14 quantifies the effect of varying N from two to eight. For image contrast, coherence-based methods do not benefit from larger N, which amplifies speckle variation. Other methods show a positive correlation between CNR and N. Excluding SRAD and BM3D, which produced high contrast images through image blurring, A2A achieves the highest CNR consistently.

A2A achieves an overall high SNR, particularly when $N < 5$. The SNR of coherence-based methods improves drastically with increasing N and becomes comparable to A2A when $N > 5$. Model- and learning-based methods follow similar trends but remain notably inferior to A2A. Fig. 14 demonstrates A2A's superior denoising capability with few sub-apertures (e.g., $N = 2$).

7. Conclusion

This study presents an A2A framework for denoising ultrasound IQ signals across diverse image domains (imaging views, organs, noise levels, and aperture configurations).

The pyramid self-contrastive learning effectively differentiates low-rank anatomy information from high-rank noises among multiple sub-apertures. Operating in a pure test-time training paradigm, A2A learns from one-shot ultrasound imaging, thereby solving the domain shift challenge and releasing reliance on clean labels. The denoised IQ signals boost B-mode image quality and hold promise for more reliable functional assessment. Moreover, this method is compatible with any architecture, ensuring stronger model variants in the future.

8. Acknowledgements

This work was supported by the Hong Kong Research Grants Council General Research Fund (Grant No. 17205022).

9. Data and code availability

All data and code will be made publicly available at https://github.com/JustinJZhang/A2A/tree/main upon acceptance of this manuscript.

10. Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. A. Jensen, S. I. Nikolov, K. L. Gammelmark, M. H. Pedersen, Synthetic aperture ultrasound imaging, Ultrasonics 44 (2006) e5–e15.

[2] C. Papadacci, M. Pernot, M. Couade, M. Fink, M. Tanter, High-contrast ultrafast imaging of the heart, IEEE transactions on ultrasonics, ferroelectrics, and frequency control 61 (2) (2014) 288–301.

[3] K. Krissian, R. Kikinis, C.-F. Westin, K. Vosburgh, Speckle-constrained filtering of ultrasound images, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 2, IEEE, 2005, pp. 547–552.

[4] H. Xie, L. E. Pierce, F. T. Ulaby, Statistical properties of logarithmically transformed speckle, IEEE transactions on geoscience and remote sensing 40 (3) (2002) 721–727.

[5] T. Loupas, W. McDicken, P. Allan, Noise reduction in ultrasonic images by digital filtering, The British journal of radiology 60 (712) (1987) 389–392.

[6] M. Gupta, H. Taneja, L. Chand, Performance enhancement and analysis of filters in ultrasound image denoising, Procedia computer science 132 (2018) 643–652.

[7] K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, Image denoising by sparse 3-d transform-domain collaborative filtering, IEEE Transactions on image processing 16 (8) (2007) 2080–2095.

[8] Y. Yu, S. T. Acton, Speckle reducing anisotropic diffusion, IEEE Transactions on image processing 11 (11) (2002) 1260–1270.

[9] A. Buades, B. Coll, J.-M. Morel, Non-local means denoising, Image processing on line 1 (2011) 208–212.

[10] P. Coupé, P. Hellier, C. Kervrann, C. Barillot, Nonlocal means-based speckle filtering for ultrasound images, IEEE transactions on image processing 18 (10) (2009) 2221–2229.

[11] J. Yang, J. Fan, D. Ai, X. Wang, Y. Zheng, S. Tang, Y. Wang, Local statistics and non-local mean filter for speckle noise reduction in medical ultrasound image, Neurocomputing 195 (2016) 88–95.

[12] X. Li, T. Wang, Research on spinal ultrasound image denoising based on an improved bm3d algorithm, in: Fifth International Conference on Image Processing and Intelligent Control (IPIC 2025), Vol. 13782, SPIE, 2025, pp. 152–156.

[13] Y. Gan, E. Angelini, A. Laine, C. Hendon, Bm3d-based ultrasound image denoising via brushlet thresholding, in: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), IEEE, 2015, pp. 667–670.

[14] K. Hollman, K. Rigby, M. O'donnell, Coherence factor of speckle from a multi-row probe, in: 1999 IEEE Ultrasonics Symposium. Proceedings. International Symposium (Cat. No. 99CH37027), Vol. 2, IEEE, 1999, pp. 1257–1260.

[15] P.-C. Li, M.-L. Li, Adaptive imaging using the generalized coherence factor, IEEE transactions on ultrasonics, ferroelectrics, and frequency control 50 (2) (2003) 128–141.

[16] J. Camacho, M. Parrilla, C. Fritsch, Phase coherence imaging, IEEE transactions on ultrasonics, ferroelectrics, and frequency control 56 (5) (2009) 958–974.

[17] S.-L. Wang, C.-H. Chang, H.-C. Yang, Y.-H. Chou, P.-C. Li, Performance evaluation of coherence-based adaptive imaging using clinical breast data, IEEE transactions on ultrasonics, ferroelectrics, and frequency control 54 (8) (2007) 1669–1679.

[18] B. M. Asl, A. Mahloojifar, Minimum variance beamforming combined with adaptive coherence weighting applied to medical ultrasound imaging, IEEE transactions on ultrasonics, ferroelectrics, and frequency control 56 (9) (2009) 1923–1931.

[19] M. O'Donnell, Y. Wang, Coded excitation for synthetic aperture ultrasound imaging, IEEE transactions on ultrasonics, ferroelectrics, and frequency control 52 (2) (2005) 171–176.

[20] X. Li, N. Navab, Z. Jiang, Speckle2self: Self-supervised ultrasound speckle reduction without clean data, Medical Image Analysis (2025) 103755.

[21] K. J. Parker, Ultrasonic attenuation and absorption in liver tissue, Ultrasound in medicine & biology 9 (4) (1983) 363–369.

[22] P. C. Li, M. ODonnell, Evaluational spatial compounding, Ultrasonic imaging 16 (3) (1994) 176–189.

[23] M. A. L. Bell, R. Goswami, J. A. Kisslo, J. J. Dahl, G. E. Trahey, Short-lag spatial coherence imaging of cardiac ultrasound data: Initial clinical results, Ultrasound in medicine & biology 39 (10) (2013) 1861–1874.

[24] Y. Wang, C. Zheng, H. Peng, X. Chen, Short-lag spatial coherence combined with eigenspace-based minimum variance beamformer for synthetic aperture ultrasound imaging, Computers in Biology and Medicine 91 (2017) 267–276.

[25] R. B. Kuc, Application of kalman filtering techniques to diagnostic ultrasound, Ultrasonic Imaging 1 (2) (1979) 105–120.

[26] F. Y. Rizi, H. A. Noubari, S. K. Setarehdan, Wavelet-based ultrasound image denoising: Performance analysis and comparison, in: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2011, pp. 3917–3920.

[27] H. Rabbani, M. Vafadust, P. Abolmaesumi, S. Gazor, Speckle noise reduction of medical ultrasound images in complex wavelet domain using mixture priors, IEEE transactions on biomedical engineering 55 (9) (2008) 2152–2160.

[28] S. Khare, P. Kaushik, Speckle filtering of ultrasonic images using weighted nuclear norm minimization in wavelet domain, Biomedical Signal Processing and Control 70 (2021) 102997.

[29] L. Zhu, C.-W. Fu, M. S. Brown, P.-A. Heng, A non-local low-rank framework for ultrasound speckle reduction, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 5650–5658.

[30] M. A. Lediju, G. E. Trahey, B. C. Byram, J. J. Dahl, Short-lag spatial coherence of backscattered echoes: Imaging characteristics, IEEE transactions on ultrasonics, ferroelectrics, and frequency control 58 (7) (2011) 1377–1388.

[31] P. Kokil, S. Sudharson, Despeckling of clinical ultrasound images using deep residual learning, Computer Methods and Programs in Biomedicine 194 (2020) 105477.

[32] K. M. Mohamed, M. H. Ali, Ultrasound images enhancement using unet-deep learning according to resolution and speckle noise, International Journal of Mechanical Engineering 7 (5).

[33] L. Zhang, J. Zhang, Ultrasound image denoising using generative adversarial networks with residual dense connectivity and weighted joint loss, PeerJ Computer Science 8 (2022) e873.

[34] O. Karaoğlu, H. Ş. Bilge, I. Uluer, Removal of speckle noises from ultrasound images using five different deep learning networks, Engineering Science and Technology, an International Journal 29 (2022) 101030.

[35] H. Singh, A. S. Ahmed, F. Melandsø, A. Habib, Ultrasonic image denoising using machine learning in point contact excitation and detection method, Ultrasonics 127 (2023) 106834.

[36] H. Asgariandehkordi, S. Goudarzi, M. Sharifzadeh, A. Basarab, H. Rivaz, Denoising plane wave ultrasound images using diffusion probabilistic models, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control.

[37] Y. Zhang, C. Huneau, J. Idier, D. Mateus, Ultrasound image reconstruction with denoising diffusion restoration models, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2023, pp. 193–203.

[38] L. Yancheng, X. Zeng, Q. Dong, X. Wang, Red-mam: A residual encoder-decoder network based on multi-attention fusion for ultrasound image denoising, Biomedical Signal Processing and Control 79 (2023) 104062.

[39] D. Hyun, L. L. Brickson, K. T. Looby, J. J. Dahl, Beamforming and speckle reduction using neural networks, IEEE transactions on ultrasonics, ferroelectrics, and frequency control 66 (5) (2019) 898–910.

[40] X. Yu, S. Luan, S. Lei, J. Huang, Z. Liu, X. Xue, T. Ma, Y. Ding, B. Zhu, Deep learning for fast denoising filtering in ultrasound localization microscopy, Physics in Medicine & Biology 68 (20) (2023) 205002.

[41] S. Cammarasana, P. Nicolardi, G. Patanè, Real-time denoising of ultrasound images based on deep learning, Medical & Biological Engineering & Computing 60 (8) (2022) 2229–2244.

[42] T.-T. Zhang, H. Shu, K.-Y. Lam, C.-Y. Chow, A. Li, Feature decomposition and enhancement for unsupervised medical ultrasound image denoising and instance segmentation, Applied Intelligence 53 (8) (2023) 9548–9561.

[43] H. Wu, T. T. Huynh, R. Souvenir, Echocardiogram enhancement using supervised manifold denoising, Medical image analysis 24 (1) (2015) 41–51.

[44] P. Jarosik, M. Lewandowski, Z. Klimonda, M. Byra, Pixel-wise deep reinforcement learning approach for ultrasound image denoising, in: 2021 IEEE International Ultrasonics Symposium (IUS), IEEE, 2021, pp. 1–4.

[45] M. Jiang, C. You, M. Wang, H. Zhang, Z. Gao, D. Wu, T. Tan, Controllable deep learning denoising model for ultrasound images using synthetic noisy image, in: Computer Graphics International Conference, Springer, 2023, pp. 297–308.

[46] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, T. Aila, Noise2noise: Learning image restoration without clean data, arXiv preprint arXiv:1803.04189.

[47] L. Huang, J. Yin, J. Zhang, U.-W. Lok, R. M. DeRuiter, J. Jin, K. M. Knoll, K. E. Petersen, J. D. Krier, X.-y. Zhu, et al., Self-supervised deep learning for denoising in ultrasound microvascular imaging, arXiv preprint arXiv:2507.05451.

[48] H. Asgariandehkordi, M. Sharifzadeh, H. Rivaz, Lightweight physics-informed zero-shot ultrasound plane wave denoising, arXiv preprint arXiv:2506.21499.

[49] D. Jung, M. Kang, S. H. Park, N. Guezzi, J. Yu, Unsupervised speckle noise reduction technique for clinical ultrasound imaging, Ultrasonography 43 (5) (2024) 327–344.

[50] C. Yu, F. Ren, S. Bao, Y. Yang, X. Xu, Self-supervised ultrasound image denoising based on weighted joint loss, Digital Signal Processing 162 (2025) 105151.

[51] Y. Zhang, N. Jiang, Z. Xie, J. Cao, Y. Teng, Ultrasonic image's annotation removal: A self-supervised noise2noise approach, arXiv preprint arXiv:2307.04133.

[52] J. Huh, S. Khan, S. Choi, D. Shin, J. E. Lee, E. S. Lee, J. C. Ye, Tunable image quality control of 3-d ultrasound using switchable cyclegan, Medical Image Analysis 83 (2023) 102651.

[53] S. Muth, S. Dort, I. A. Sebag, M.-J. Blais, D. Garcia, Unsupervised dealiasing and denoising of color-doppler data, Medical image analysis 15 (4) (2011) 577–588.

[54] D. Ulyanov, A. Vedaldi, V. Lempitsky, Deep image prior, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 9446–9454.

[55] J. Batson, L. Royer, Noise2self: Blind denoising by self-supervision, in: International conference on machine learning, PMLR, 2019, pp. 524–533.

[56] A. M. Christensen, I. M. Rosado-Mendez, T. J. Hall, A systematized review of quantitative ultrasound based on first-order speckle statistics, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 71 (7) (2024) 872–886.

[57] Y. Zhang, Y. Guo, W.-N. Lee, Ultrafast ultrasound imaging with cascaded dual-polarity waves, IEEE Transactions on Medical Imaging 37 (4) (2017) 906–917.

[58] T. D. Mast, Empirical relationships between acoustic parameters in human soft tissues, Acoustics Research Letters Online 1 (2) (2000) 37–42.

[59] Y. Sun, X. Wang, Z. Liu, J. Miller, A. Efros, M. Hardt, Test-time training with self-supervision for generalization under distribution shifts, in: International conference on machine learning, PMLR, 2020, pp. 9229–9248.

[60] A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, F. Makedon, A survey on contrastive self-supervised learning, Technologies 9 (1) (2020) 2.

[61] T. Chen, S. Kornblith, M. Norouzi, G. Hinton, A simple framework for contrastive learning of visual representations, in: International conference on machine learning, PmLR, 2020, pp. 1597–1607.

[62] M. Gutmann, A. Hyvärinen, Noise-contrastive estimation: A new estimation principle for unnormalized statistical models, in: Proceedings of the thirteenth international conference on artificial intelligence and statistics, JMLR Workshop and Conference Proceedings, 2010, pp. 297–304.

[63] A. v. d. Oord, Y. Li, O. Vinyals, Representation learning with contrastive predictive coding, arXiv preprint arXiv:1807.03748.

[64] X. Wang, R. Zhang, C. Shen, T. Kong, L. Li, Dense contrastive learning for self-supervised visual pre-training, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 3024–3033.

[65] J. Zhang, B. Dai, H. Guan, W.-N. Lee, Msaf: A multi-level sensing adversarial framework for signal recovery in synthetic aperture ultrasound, IEEE Journal of Biomedical and Health Informatics.

[66] B. E. Treeby, J. Jaros, A. P. Rendell, B. T. Cox, Modeling nonlinear ultrasound propagation in heterogeneous media with power law absorption using a k-space pseudospectral method, The Journal of the Acoustical Society of America 131 (6) (2012) 4324–4336.

[67] C. Mitchell, P. S. Rahko, L. A. Blauwet, B. Canaday, J. A. Finstuen, M. C. Foster, K. Horton, K. O. Ogunyankin, R. A. Palma, E. J. Velazquez, Guidelines for performing a comprehensive transthoracic echocardiographic examination in adults: recommendations from the american society of echocardiography, Journal of the American Society of Echocardiography 32 (1) (2019) 1–64.

[68] F. Wang, H. Liu, Understanding the behaviour of contrastive loss, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 2495–2504.