Robust Flow Component Identification for Blockwise SVD Clutter Filtering in High-Frame-Rate Ultrasound Using a Deeply Connected Neural Network

by

Cyrus Cerkauskas

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2022

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Ultrasound plays a critical role in the accurate and reliable observation of blood flow dynamics within the human body and is instrumental in the assessment of cardiovascular health and subsequent follow-up treatment. However, flow imaging innovations in ultrasound are dependent on the ability to remove the unwanted portion of the signal corresponding to static or slow moving tissues (or *clutter*) from the dynamic blood signal. If this filtering process cannot be done to a high degree of precision, any corresponding flow image will be of low quality, corrupted by non-flow signals. Filtering is typically performed on the basis of frequency, using a high-pass filter. This approach functions well when the velocity distributions of tissues and flow signals are distinct but fails entirely when the two spectra overlap.

New types of filters in high-frame-rate ultrasound (HiFRUS) making use of both temporal and spatial information use the singular value decomposition (SVD) have been proposed. These filters function by decomposing the input signal into a series of orthonormal basis vectors or components. In theory, as tissue and flow signals posses different signal statistics, they should be decomposed into different components and readily identified. The identified flow components can then be reconstituted to produce a filtered flow signal. In practise, flow and clutter component identification is a challenging task considering the adaptive nature of the SVD. Furthermore, flow and clutter signals may be mixed in the same components to varying degrees, making identification of flow signals a challenging task. While increased flow sensitivity and clutter rejection has been demonstrated by SVD filters, they currently lack the robustness required for clinical applications, often failing to perform in scenarios that challenge their innate assumptions of the flow signal decomposition.

The goal of this work is to develop a robust generalizable flow identification framework that produces high quality filtered flow images across challenging *in-vivo* flow scenarios where current SVD filters demonstrate inconsistency. A deeply-connected neural network (DNN) was trained on a variety of flow acquisitions to reproduce the area under the curve (AUC) value obtained after performing receiver operator characteristic (ROC) analysis on the segmented flow region using a variety of statistical quantities correlated to the presence of flow in each component of the decomposition. The use of the AUC metric and subsequent training of the DNN using multiple statistical factors represents the first attempt at using a supervised learning approach to identify the flow components of the decomposition using many statistical factors simultaneously. When the proposed model was applied to acquisitions of an *in-vitro* flow phantom, *in-vivo* brachial artery, and *in-vivo* femoral arteries, greater sensitivity and specificity (measured using contrast and AUC) were obtained when compared to literature SVD techniques. The proposed model was also sufficiently generalizable to identify small blood vessels in the *in-vivo* human kidney.

The proposed methodology demonstrates an improvement on the performance and consistency of SVD filters, helping to put this powerful technique in the hands of more users. Furthermore, the supervised training methodology developed here, using ROC analysis to obtain an AUC value for each component that describes the spatial distribution of its signal power, has the potential to be extended to other clutter filtering algorithms potentially leading to better feature identification than current unsupervised techniques.

Acknowledgements

I would like to thank all those who have helped me in my path to graduation. It has been a difficult few years and you have helped me on my way.

I would like to thank my supervisor Dr. Alfred Yu for his continued support and supervision these last few years.

I would like to thank the wonderful people at LITMUS from whom I have received nothing but support, guidance, and patience these last few years. You have made my time at LITMUS much more engaging and have given me a greater appreciation of the depth of the field of ultrasound. A special thanks to Adrian Chee, Billy Yiu, Jason Au, Di Xiao, and Hassan Nahas for the wonderfully stimulating conversations I've had over the years.

Finally, I would like to thank my parents and brother for their unwavering support in my studies throughout the good and the bad times. When I needed someone to talk to you were always there.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	
2D	Two-Dimensional
AUC	Area Under the Curve
CPU	Central processing Unit
CR	Contrast Ratio
CNR	Contrast to Noise Ratio
DAS	Delay-and-Sum
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DNN	Deep Neural Network
DR	Dynamic Range
FPR	False Positive Rate
GPU	Graphical Processing Unit
HiFRUS	High Frame Rate Ultrasound
HPF	High Pass Filter
ICA	Independent Component Analysis
MSE	Mean Squared Error
PCA	Principle Component Analysis
PRF	Pulse Repetition Frequency
RF	Radiofrequency
ROC	Receiver Operation Characteristic
ROI	Region of Interest
SSIM	Structural Similarity Index Measure
SSM	Spatial Similarity Matrix
SVD	Singular Value Decomposition
TOF	Time of Flight
TPR	True Positive Rate

Part I

Introduction

Ultrasound Flow Imaging as a Diagnostic Tool

Compared to other imaging modalities such as computer tomography (CT) or magnetic resonance imaging (MRI), ultrasound provides a safe, cheap, non-invasive method of imaging the internal structures of the human body in a point-of-care environment. As today's ultrasound scanners are low-cost mobile instruments, capable of being directly delivered to the patient and do not present any harmful ionizing radiation or potentially dangerous magnetic fields they are becoming increasingly common for disease treatment and diagnosis [1][2][3]. Furthermore, due to its inherent safety, ultrasound is the go-to standard imaging modality for applications such as fetal health assessment [4][5]. While ultrasound is typically associated with the imaging of physical structures within the human body, Doppler ultrasound provides a method of probing moving structures and even the flow of blood within the body, allowing for the visualization of blood flow dynamics within the arteries, veins, and heart and assessment of cardiovascular health [6][7][8].

As an imaging modality that provides resolution of blood flow dynamics, Doppler ultrasound is commonly used to diagnose cardiac disease. For instance, measurements of very high blood velocities in arteries may be indicative of plaque buildup (atherosclerosis) [9][10][11] or blood clots [12][13]. Doppler techniques may also be used to observe motion within the heart and help in the diagnosis of conditions such as arrhythmia [14][15].

Although many flow imaging techniques exist, flow imaging is always dependent on the separation or suppression of the signals originating from moving scatterers from the signals originating from static background part of the image such as the walls of the blood vessel and other surrounding tissue [16][17]. The process of suppressing the high-magnitude signals originating from stationary and slow moving tissue or *clutter* is called *clutter filtering* and is instrumental as a first step in all Doppler techniques for the formation of high quality ultrasound flow images [18].

Clutter filtering, as deployed in current clinical scanners, follows a major assumption — that tissue motion is relatively slow and blood flow speeds are relatively high. The corresponding frequency of signals originating from tissue and blood is therefore low and high respectively allowing for the suppression of tissue signals through the use of a high pass filter [16][19] and preset cutoff frequency. However, many scenarios exist in the body where blood and tissue movements are similar in velocity (such as in cardiac imaging), resulting in an inability to separate blood and tissue signals on the basis of frequency. If this is the case, the much higher magnitude tissue signals originating from large amounts of tissue dominate the much weaker flow signals resulting in an inability to visualize flow. Examples of particularly difficult scenarios where blood and tissue spectra overlap include cardiac applications where signals originating from the fast moving blood is difficult to isolate against the rapid motion of heart tissues and in micro-vessel imaging where the slow flow of blood through small vessels is obscured by random bodily movements. In these situations frequency filters fail to produce good flow images; more advanced clutter removal techniques are needed.

A Need for Robust Clutter Removal

Newer clutter filtering techniques involve the use of both spatial and temporal information to better detect and suppress tissue clutter and have been demonstrated in a variety of clinical applications. One such technique involves the use of the *singular value decomposition* (SVD) [20]. Spatiotemporal clutter filters utilizing the SVD have been used to measure flow within the brain [21][22][23], allowing researchers to identify regions of high blood flow and thus activity. Measurements of flow within skeletal muscles provides visualization of the transport and uptake of various blood-transported substances and their physiological role within the body [24]. For instance, using SVD filtering, blood flow, and hence insulin uptake could be observed with good fidelity when comparing healthy and obese insulin-resistant mice [25].

While SVD clutter filters have been shown to produce significantly better flow images than conventional clutter filters, both in traditional scanline imaging [26][27] as well as with ultra-fast plane wave transmissions [28]; a significant point of contention has remained; how can the elements of the decomposition which contain the blood flow signal be identified [29][1]?. The flow components, assumed to be present as a contiguous block in the decomposition, are generally identified through an analysis of the data matrices of the decomposition, revealing the estimated upper and lower boundaries of this blood *subspace*. Although investigations of the statistical parameters used to identify the blood subspace have been made [29], a holistic consideration of these parameters and their correlations between both themselves and the blood or clutter signals has yet to be performed and there is still much contention in the literature on how to best identify the boundaries of the flow subspace. Furthermore, the fundamental assumption that flow is always contained in an easily identifiable *subspace* with a well defined upper and lower threshold has only recently begun to be challenged [30][31].

Inherent in all clutter filtering algorithms is the fundamental trade-off between clutter removal and flow sensitivity (generically referred to as *specificity* and *sensitivity* in more generic separation algorithms [32]). For example, consider using a high pass filter on a signal with flow and clutter components overlapping in frequency (but with the flow signal distribution being higher in frequency than the clutter signal distribution). Here, setting a lower than optimal cutoff frequency will result in a filtered signal that contains most or all of the flow signal but will contain a significant amount of clutter as well. This filter is technically more sensitive to flow and will contain more of the flow distribution but is also much worse at clutter removal and may fail to produce a useful image as the magnitude of clutter signals is much higher than that of flow signals. Setting the cutoff frequency too high creates the inverse problem; more of the clutter signal is removed but at the cost of losing part of the flow signal. Furthermore the *optimal* cutoff may vary depending on application[2]; and so it is important to leave the dial balancing sensitivity to flow against clutter removal (*sensitivity* vs. *specificity*) in the hands of the medical practitioner.

This sensitivity and specificity trade-off applies to SVD filtering techniques where the decomposition of the flow and clutter signals into a discrete number of components is not perfect

[1] The details of SVD clutter filter design will be discussed in the next chapter.

[2] Although a 'mathematically correct' cutoff, however defined, may exist, it may be so that greater diagnostic value can be found in a flow image that is biased to greater flow sensitivity *OR* clutter removal.

and components often contain some mixture of the flow and clutter signals. Identifying the relative degree of flow and clutter in each component of the decomposition is a difficult operation to perform quantitatively, hence the difficulty in the literature of identifying the flow subspace thresholds. Most literature SVD algorithms only perform a binary classification on whether a component should be classified as flow or clutter but do not identify the degree to which that component contains flow or clutter signals, attempting to produce an upper and lower threshold of best fit. This automatic fit fails to allow for a tunable balance between sensitivity and specificity; techniques do exist that consider the relative flow and clutter compositions of SVD components [30][31] but do not produce a meaningful numerical result that allows for the tuning between flow sensitivity and specificity.

Another issue it that although SVD clutter filters have consistently been shown to outperform frequency filters in challenging *in-vivo* scenarios [33][34][35], their robustness has not been demonstrated at a level sufficient for clinical use. Demonstration has largely been restricted to single frames as algorithmic inconsistency in accurately and consistently identifying the flow subspace boundaries creates difficulty in producing filtered cineloops without flashing artifacts or inconsistent frame-to-frame performance [36]. Thus while SVD clutter filters may reduce or eliminate the 'flashing' artifacts so common when using frequency filters [37] they may introduce a new distinctive type of flashing artifacts if clutter is not consistently removed on a frame-to-frame basis.

The major issues with current implementations of SVD clutter filters can be summarized as follows.

1. **Clinical robustness:** A robust and sufficiently adaptable SVD algorithm that consistently outperforms 'best guess' manual operator input under varied and challenging imaging scenarios has not been fully demonstrated. While improved clutter filtering performance is frequently shown, this performance must hold up across all relevant imaging scenarios for entire cineloops and not simply selected frames. This leads to point 2).

2. **Simplistic techniques:** SVD filtering algorithms tend to use a limited amount of the overall information contained in the decomposition to identify flow components which can lead to inconsistency. Furthermore they often make specific assumptions, such as the assumed existence of flow in the image and will fail when these assumptions do not hold true.

3. **A focus on 'one size fits all' thresholding:** Current SVD filtering algorithms do not allow for meaningful user tuning between *sensitivity* and *specificity*. Instead the focus is on determining the best default thresholds. While this may be a mathematically sound approach, there is a significant difference in what constitutes the 'ideal' threshold across various imaging scenarios which may demand different balances of sensitivty and specificity. Furthermore, many current SVD algorithms do not quantitatively attempt to describe how flow and clutter signals are contained within individual SVD components.

Machine Learning and Ultrasound

While most literature SVD filtering techniques only utilize a fraction of the statistical information available in the decomposition [29] to identify the flow subspace thresholds, our goal is to simultaneously consider all relevant statistical information to arrive at a more accurate estimate. Our proposed framework will also allow us to numerically quantify the relative amount of flow in each component of the decomposition. With both *inputs* and *targets* available, it is possible to use a more complex *supervised learning* approach to identify the flow components of the decomposition.

In this thesis we propose to use a deeply-connected neural network (DNN) under a supervised learning approach to well-characterize the relationship between the SVD statistics and presence of flow. Supervised machine learning models and neural networks have been used extensively in ultrasound for clutter filtering applications [38][39][40][41]; however a fundamental issue has always been the limitations of the available target information. While literature networks have been used to perform clutter filtering, the target output of the networks were limited to the output of pre-existing clutter filtering algorithms and therefore literature networks were unable to advance the performance of the clutter filter beyond that of the input target training data. To address this, in this project we will use a combination of manual segmentation and *receiver-operator characteristic*[3] or *ROC* analysis to numerically quantify how the signal of each singular component is present in the identified flow region and train our network to numerically identify the flow-containing components of the decomposition.

Research Overview

The focus of this research project will be to develop a clutter filtering utilizing the SVD that addresses the points described above. We aim to accomplish this by first performing an in-depth analysis of the information contained in the SVD data matrices to identify and extract the statistical fingerprints indicative of the presence of flow. These statistical quantities will serve as the *inputs* for our framework to numerically quantify the presence of flow. We will then develop a method to numerically characterize the amount of flow signal contained within each component of the decomposition using a spatial segmentation technique, to obtain a prospective *target* for our framework to reproduce. We will then train a deep neural network using the input and target values obtained from large number of distinct and varied *in-vitro* and *in-vivo* acquisitions to give a numerical score representing the presence of flow within each component of the decomposition.

We will then evaluate the effectiveness of the proposed framework on unseen *in-vitro* and *in-vivo* acquisitions. In the case of *in-vivo* acquisitions, care will be taken to select complex flow scenarios such as the femoral bifurcation (in which the femoral vein is also present)

[3]ROC analysis has been used extensively in the literature to numerically quantify the performance of clutter filtering algorithms. Here, we propose to extend ROC analysis to each individual component of the decomposition.

and brachial artery with a comparison being made across the entire cardiac cycle. Both the performance of individual frames and the consistency of frame-to-frame output will be considered in accordance with standard literature assessment techniques.

Thesis Organization

This thesis thesis is organized into four chapters following this introduction. **Chapter 2** will present background information on ultrasound data acquisition and clutter filtering strategies. **Chapter 3** describes how the proposed clutter filtering framework will function, describing how input and target information is obtained and how the deep neural network is trained. **Chapter 4** describes how the proposed DNN SVD clutter filter performs against current literature algorithms in *in-vitro* and *in-vivo* scenarios. Finally in **Chapter 5** we interpret the results of the previous chapter and suggest future work in the field of SVD clutter filters.

Part II

Background: Ultrasound Flow Imaging

This chapter presents the basics of ultrasound data acquisition and image formation as groundwork for further discussions on clutter filtering and the considerations that must be made when performing Doppler flow imaging. The singular value decomposition and its use in spatiotemporal SVD clutter filtering is then presented. Current literature SVD filtering algorithms and their strengths and weakness are discussed. An introduction to supervised learning and finally a brief discussion on how neural networks function and are trained is then provided. The overall goal of this chapter is to provide a background and prepare for the introduction of the proposed SVD clutter filtering framework presented in Chapter 3.

Fundamentals of Ultrasound

Ultrasound imaging is a *pulse-echo* sensing technique where the echoes of ultrasonic pressure waves, reflecting off echogenic scatterers within a medium, are reconstructed to create a localized intensity map of the position of the scatterers. A simplified description of ultrasound imaging process proceeds as follows, piezoelectric crystals in a device known as a transducer are used to convert an electrical signal into acoustic pressures waves (1-15 Mhz) that penetrate into the body from the transducer-skin interface. In the process of travelling through the body, some of the acoustic energy is reflected whenever a 'scatterer' is encountered [42]. These returning waves travel back to the transducer where the piezoelectric crystals in the transducer head converts the intensity of the returning echoes into voltage measurements. These voltage measurements, known as *radiofrequency* (RF) data can then be processed to return a map of the location of acoustic scatterers in the field of view called a *b-mode* (brightness mode) image through time of flight calculations with the speed of sound within bodily tissues assumed to be a constant 1540 m/s. When applied to the human body, this mapping of acoustic scatters allows one to discern structural information such as the location and internal structure of different tissues such as blood, muscle, and bone. Figure 1 shows an b-mode image cross section of the human carotid where the highly echogenic muscles and tissues and less echogenic blood can be observed.

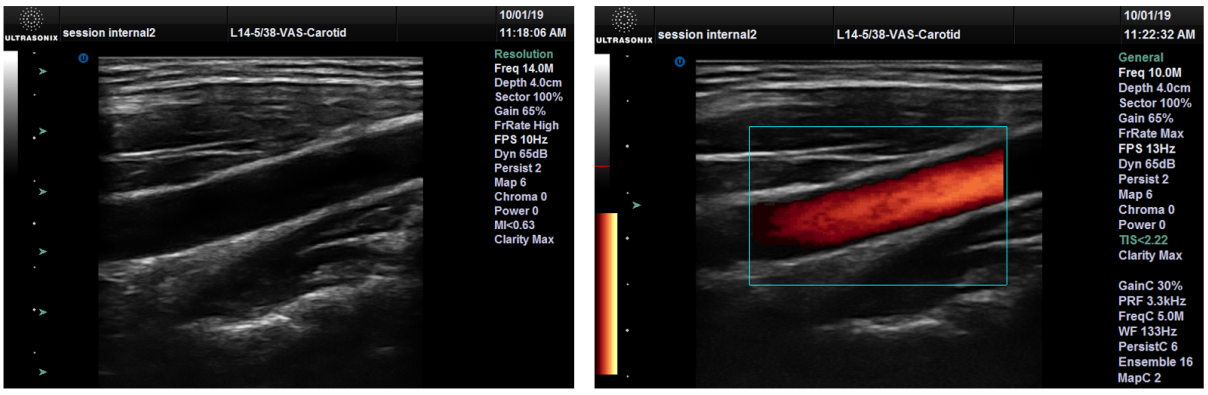

Figure 1: *Left*) A b-mode image of the human carotid artery acquired using conventional scanline ultrasound. *Right*) The corresponding power Doppler image showing the intensity of blood flow within the artery.

Data Acquisition and Beamforming

For the images presented in Figure 1 an ultrasound probe head containing a linear array of transducer elements was used; the use of multi-element probes is necessary to localize the echoes originating from an object to a specific point in space. In conventional ultrasound scanners, an approach called "scan-line" imaging is utilized, whereby an ultrasound pulses are emitted from sub-apertures of the array. This is shown in Figure 2a where a number of adjacent elements of the array are used for transmission and then beamformed on reception to create a single image line in the resulting b-mode image. The aperture window is then translated laterally and the next set of elements fired. This process continues until the b-mode image is complete and then repeated again from the beginning of the array for the next frame. A set number of firings per second are performed (this number is known as the *pulse repetition frequency* or *PRF*) — as many firings must be completed per frame the overall frame-rate is slow, $\sim$ 30 fps, although overall quality is high due to the use of focused transmissions.

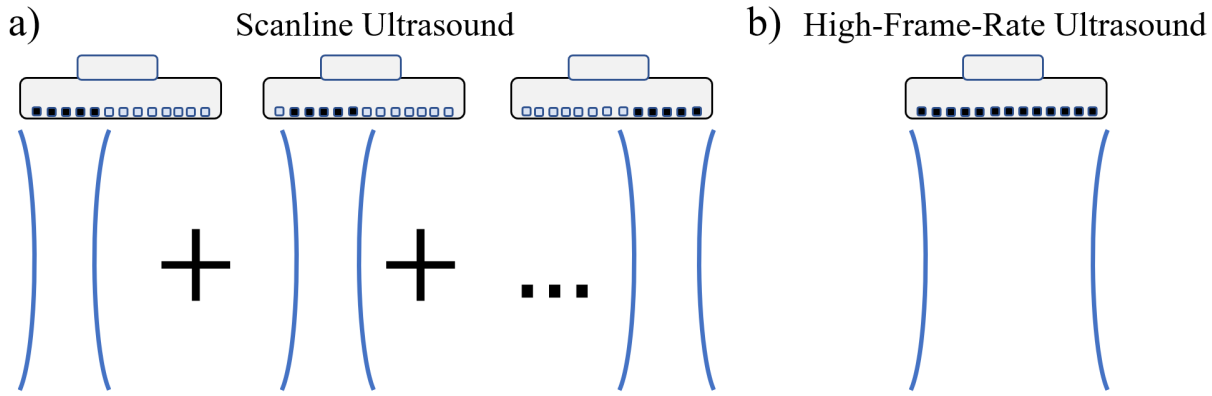

Figure 2: In scanline or conventional ultrasound *a*) a subset of array elements are used to form a focused transmission which upon reception is used to form a single *line* in the resulting image. *b*) In high-frame-rate ultrasound the entire imaging view is insonified to create an image of the entire imaging view at once.

In high-frame-rate ultrasound, the entire array of elements is fired simultaneously, creating a plane wave that insonifies the entire imaging region. All transducer elements are used both for both transmission and reception and frame-rate is limited effectively only by the PRF. However, high-frame-rate ultrasound suffers from a significant loss in quality due to a lower power intensity due to the insonification of the entire imaging region and a difficulty in localizing the exact position in space of scatterers. This can be corrected somewhat by compounding plane wave transmissions from different transmission angles [43].

Image formation in high-frame-rate ultrasound is most commonly performed using a process known as *delay and sum* beamforming [43]. Here *time-of-flight* (TOF) calculations (total time = transmit time + receive time) are performed for each pixel in the image, allowing for the localization in each transducer element's RF data that best represents the echoes originating from a given pixel. The resulting echo contributions can then be summed

together to give a value representing the intensity of echoes originating at each pixel's location in space. Displaying all of the summed pixel contributions under a logarithmic scale produces the final b-mode image. These b-mode images show structural anatomy (Figure 2) but do not reveal reveal the flow of blood within vessels; further signal processing is needed to detect the movement of scatterers.

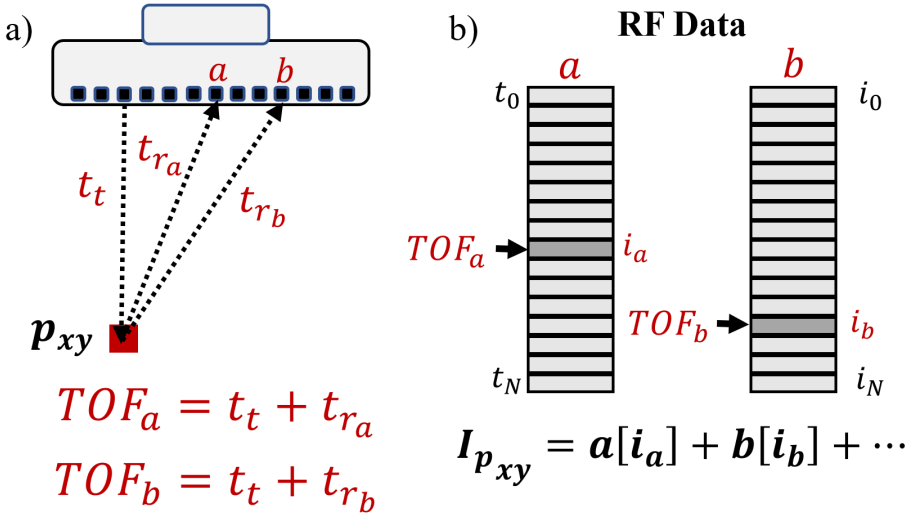

Figure 3: *a*) The time of flight for a plane wave reflecting off scatters at pixel xy can be calculated to two transducer elements a and b as the sum (TOF_a and TOF_b) of the transmit (t_t) and receive times (t_{r_a} and t_{r_b}) respectively. *b*) The RF data of elements a and b can then be inspected, and the indices (i_a and i_b) corresponding to times TOF_a and TOF_b identified. The final signal intensity of the red pixel ($I_{p_{xy}}$) is the sum of all relevant element contributions.

Principles of Flow Imaging

Although B-mode images show the location of echogenic structures within the body they contain only static structural information and therefore multiple transmissions are required to visualize blood movement, a time dependent phenomena. For flow imaging using a pulsed wave approach, multiple RF acquisitions are needed to calculate the change in position of scatters as a function of time.

Consider the scenario of a scatterer moving toward the transducer as shown in Figure 4. Here a pulse echo transmission occurs every $T_{PRI} = \frac{1}{PRF}$ where PRF is the pulse repetition frequency. As the scatterer travels toward the transducer the transducer receives the reflected signal earlier in the transmission cycle. The magnitude of the reflected signal for that particular depth (pixel) can be recorded for each slow time firing and the resulting sampled waveform used in the Doppler equation (1) to estimate the velocity of the scatterer toward or away from the transducer. Here f_D is the measured Doppler frequency, v is the velocity of the scatterer toward or away from the transducer, c is the speed of sound in tissue, f is the transmission frequency, and θ is the angle between the transducer and the direction of motion of the scatterer.

$$f_D = \frac{2vf}{c}\cos\theta \tag{1}$$

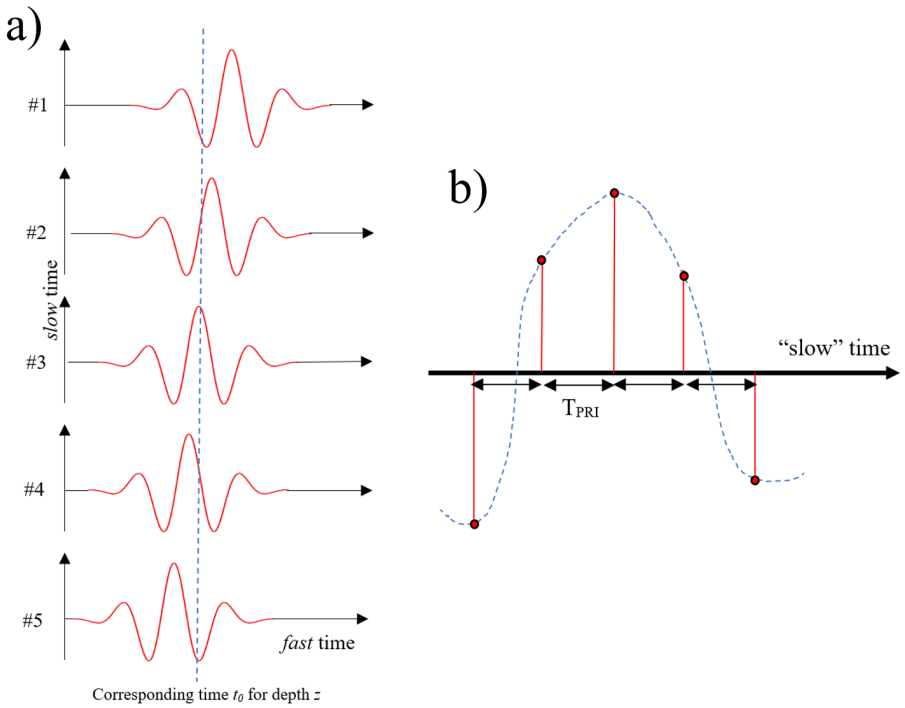

Figure 4: Measuring the Doppler frequency using pulsed-wave Doppler. *a*) A scatterer is observed moving towards the transducer over several slow time acquisitions. *b*) The intensity of the echoes returning from a given depth (the dashed line in the left-hand image) can be used to estimate the frequency of the scatterer motion in relation to the ultrasound transmission frequency ($f = \frac{1}{T_{PRI}}$).

This example only considers the case of a single scatterer. In reality many scatterers contribute to the signal of any given pixel, all of which may be moving at different velocities. It is therefore useful to compute the mean Doppler frequency of the ensemble of scatterers which can be performed through calculation of the *lag-one autocorrelation* [44]. In Equation 2, V represents the slow time series signal of length N for a given pixel. After computing the Doppler frequency one can then substitute into Equation 1 to compute the mean velocity of the given pixel.

$$R = \sum_{i=1}^{N-1} V^*(i) \cdot V(i+1) \tag{2}$$

$$f_D = \left(\frac{PRF}{2\pi}\right) \times \arctan\left(\frac{imag(R)}{real(R)}\right) \tag{3}$$

By displaying all pixels over a given magnitude threshold, we can form the colour Doppler image. By convention, flow toward the transducer is depicted as red while flow away from

the transducer is shaded blue. The power Doppler image, by contrast, is simply a map of the relative magnitude of motion in each pixel and contains no directional information. The b-mode image of the human carotid bifurcation is shown in Figure 5 along with its respective power and colour Doppler images.

As we are calculating the autocorrelation of adjacent samples in time, the maximum detectable frequency of pulsed wave Doppler is limited by the Nyquist limit [42]. Therefore, the maximum detectable frequency using pulsed Doppler techniques is one half the PRF. Should scatterers move faster than this limit their measured frequency appears to wrap around the spectrum and it appears as if these scatterers are moving backwards, similar to how a rotating helicopter blade appears to rotate backwards when the camera shutter speed is insufficient. This phenomenon is called *aliasing* and hinders flow visualization, making diagnosis difficult. Aliasing can be eliminating by sampling at a sufficiently high PRF, although this is not always possible in practice.

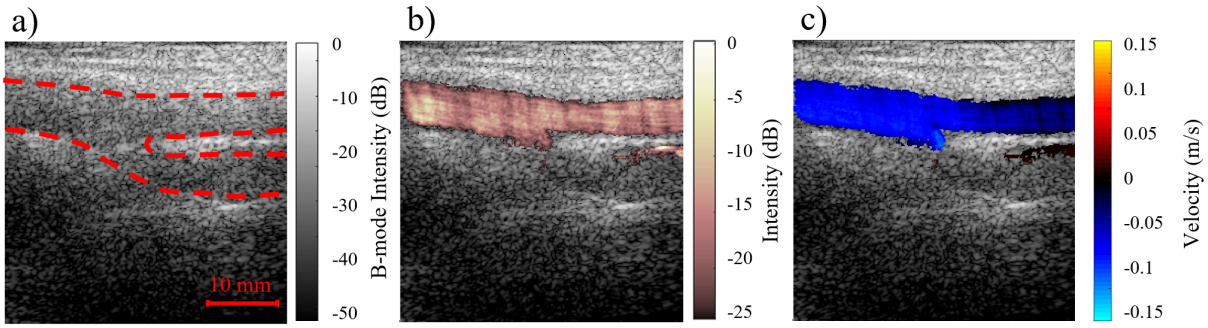

Figure 5: The femoral bifurcation. *a*) B-mode image. Red dashed lines indicate the arterial walls. *b*) Power Doppler image. *c*) Colour Doppler. Flow pixels are assigned a colour depending on whether they represent flow towards (red) or away from the transducer (blue). Only pixels with flow magnitudes over a defined threshold are assigned colours. At this moment in the cardiac cycle flow is only visible in the upper branch of the bifurcation.

Finally, we note that the number of firings used to create the sampled waveform is called the *ensemble size* and that while larger ensemble sizes may allow for better frequency measurements, temporal resolution is lost.

Clutter Filtering

In the creation of high quality flow images, it is necessary to suppress the signals resulting from the unwanted stationary or slow moving tissues that would otherwise dominate the image due to their large signal magnitudes. These unwanted signals are known as *clutter*. Figure 6 shows the origination of clutter in conventional scanline imaging, in which the signals originating from tissue regions where the grating lobes of the transducer array lie will be difficult to discriminate from the genuine flow signals due to the large amounts of energy (and hence echoes) being delivered to the grating lobe regions, even with a focused transmission. These unwanted tissue signals originating from the grating lobes will then

appear to originate along the main lobe where blood signals lie. In high-frame-rate imaging due to the use of a plane wave to insonify the entire imaging view, a given element will receive returning echoes from all equidistant regions simultaneously (same time of flight). Figure 6 shows that echoes originating from the green flow pixel will be received simultaneously with all other echoes along the dashed purple line at element a, including high magnitude red tissue pixels, resulting in distortion of the flow signal.

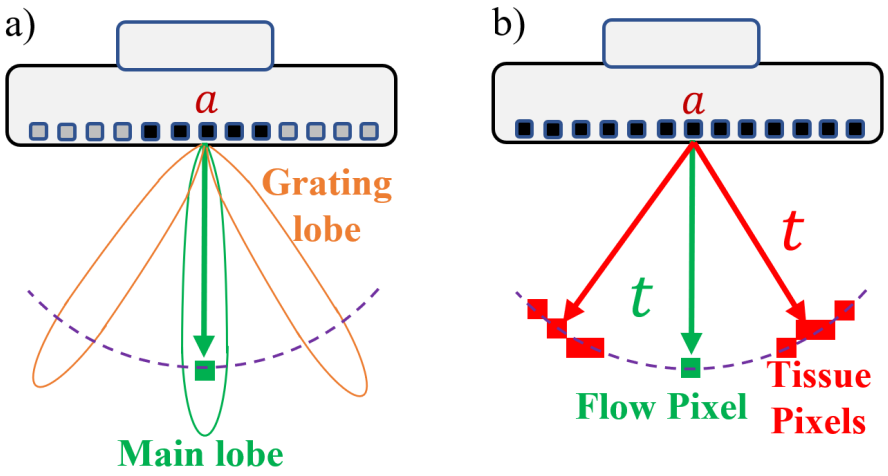

Figure 6: a) Clutter in conventional beam based ultrasound showing tissue signals originating from the grating lobes. b) In high-frame-rate ultrasound the entire region is insonified. All echoes from regions equidistant (with same time-of-flight t) to the transducer element a will be received simultaneously, corrupting potential flow signals.

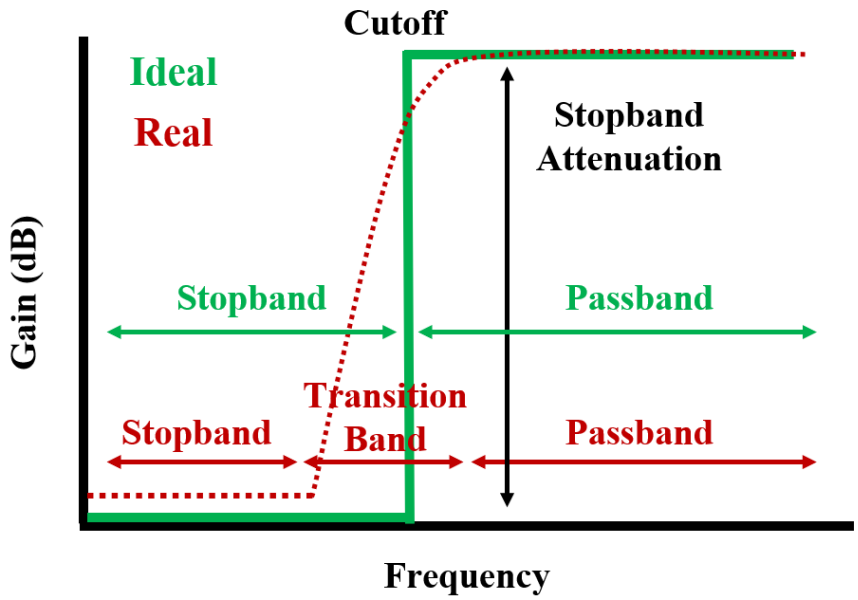

Figure 7: Frequency response of a typical high pass filter. The ideal filter as shown in *green* attenuates all signals below the cutoff frequency to zero magnitude. Filters constructed in practice shown in *red* are unable to attain a perfectly sharp transition between the stopband and passband and this region of changing attenuation is known as the transition band. Furthermore, practical filters are unable to perfectly remove all signals within the stopband, instead suppressing them by some finite amount. We also note that the exact shape of the frequency response will depend on the filter being used.

Flow images are generally formed through the use of a high pass filter (HPF), which works by suppressing some frequency components of the signal. We note that tissue clutter in ultrasound consists of signal components that are static or slow moving relative to blood. Therefore, if the low frequency components of the aggregate signal are suppressed, the resulting signal will consist mainly of blood. The name 'high-pass-filter' refers to the fact that high frequency components (passband) pass through the filter unscathed while low frequency components are suppressed (stopband). The frequency response of a typical high pass filter is shown in Figure 7.

While frequency-based clutter filtering achieves excellent results in a wide variety of scenarios and has been extensively implemented in medical scanners it requires that the blood and tissue signals possess separable frequencies. In some scenarios this separation is not guaranteed; for example in cardiac imaging the frequency spectra of fast flowing blood coincides with that of the quickly moving heart valves while in microvessel imaging where the blood flow through small vessels is so slow that it can be easily obscured by random tissue motions. In these scenarios, the frequency overlap between the blood and tissue spectra means that the two signal components cannot be well separated using a high pass frequency filter and attempts to do so will leave significant artifacts in the resulting image. This can be observed in Figure 8 which shows the spectrograms of separable and inseparable scenarios on the basis of frequency.

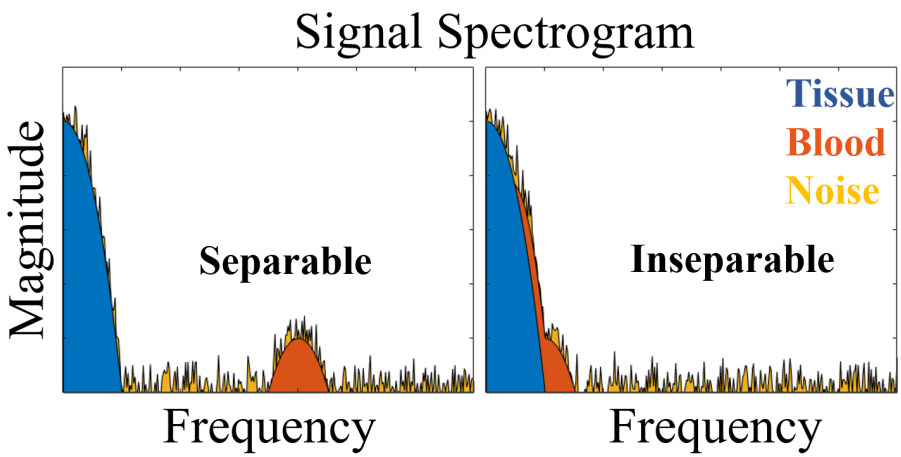

Figure 8: *Left.* Spectrogram of typical ultrasound signal. Tissue and blood components possess different frequencies and are easily separable on the basis of frequecy. *Right.* In certain circumstances such as slow blood flow the tissue and blood spectra overlap and blood and tissue signal components cannot be separated on the basis of frequency. Noise is distributed equally over the entire spectrum and is far more difficult to remove.

Spatiotemporal Filtering and the Singular Value Decomposition (SVD)

A possible solution to the problem of overlapping tissue and blood frequencies is to incorporate spatial information about the relative pixel positions of the blood and tissue signals into the picture. One such technique, *speckle tracking*, performs pattern matching between adjacent frames to track the motion of blood scatterers through space [45][46][47]. In this work we focus on the aforementioned singular value decomposition (SVD) which decomposes the raw signal into a number of self-consistent *basis vectors*.

It is assumed that the very different signal statistics of flow and clutter will result in the signal populations being decomposed into different sets of basis vectors. As the clutter signal is much larger in magnitude than the flow signal, it will be contained in the first elements of the decomposition. The flow signal is much weaker but still self-consistent and will be found in the middle elements of the decomposition. Finally, the last elements of the SVD will tend to contain uncorrelated high-frequency but low magnitude noise. The general goal of all SVD filters is therefore to identify the ideal boundary between these three signal subspaces.

Principles of SVD Filtering

The singular value decomposition (SVD) of a real or complex $m \times n$ matrix M is a factorization of the form $M = U\Delta V^\star$ where U is a $m \times m$ real or complex unitary ($UU^\star = I$) matrix, Δ is a real rectangular diagonal matrix ($m \times n$), and V is $n \times n$ unitary real or complex matrix [48].

$$M = U\Delta V^{\star} \tag{4}$$

Here we call the diagonal matrix Δ the *singular values* of M. Δ has a number of non-zero singular values equal to the rank of M. Conversely, the *columns* of U and V are known as the **left singular vectors** and **right singular vectors** of M respectively. By convention, the singular values Δ_{ii} are sorted in descending order.

$$\begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \\ M_{41} & M_{42} & M_{43} \end{pmatrix} = \begin{pmatrix} U_{11} & U_{12} & U_{13} & U_{14} \\ U_{21} & U_{22} & U_{23} & U_{24} \\ U_{31} & U_{32} & U_{33} & U_{34} \\ U_{41} & U_{42} & U_{43} & U_{44} \end{pmatrix} \begin{pmatrix} \Delta_{11} & 0 & 0 \\ 0 & \Delta_{22} & 0 \\ 0 & 0 & \Delta_{33} \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} V_{11}^{\star} & V_{13}^{\star} & V_{13}^{\star} \\ V_{21}^{\star} & V_{23}^{\star} & V_{23}^{\star} \\ V_{31}^{\star} & V_{33}^{\star} & V_{33}^{\star} \end{pmatrix} \tag{5}$$

In terms of computation it is often useful to remove elements in matrices U, Δ, or V which would be multiplied by zero and thus not possess any meaningful contribution. In this form Δ is a square diagonal matrix of size $rank(M) \times rank(M)$ with no 0 elements along its diagonal. This formulation decreases computation and memory requirements significantly. This form is known as the *reduced* SVD.

$$\begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \\ M_{41} & M_{42} & M_{43} \end{pmatrix} = \begin{pmatrix} U_{11} & U_{12} & U_{13} \\ U_{21} & U_{22} & U_{23} \\ U_{31} & U_{32} & U_{33} \\ U_{41} & U_{42} & U_{43} \end{pmatrix} \begin{pmatrix} \Delta_{11} & 0 & 0 \\ 0 & \Delta_{22} & 0 \\ 0 & 0 & \Delta_{33} \end{pmatrix} \begin{pmatrix} V_{11}^{\star} & V_{13}^{\star} & V_{13}^{\star} \\ V_{21}^{\star} & V_{23}^{\star} & V_{23}^{\star} \\ V_{31}^{\star} & V_{33}^{\star} & V_{33}^{\star} \end{pmatrix} \tag{6}$$

To perform clutter filtering using the SVD we first vectorize the ultrasound cineloop, turning it from a 3D array of K ensemble images of size N_x by N_y into a 2D matrix called a Casorati matrix with dimensions $N_x \times N_y$ by K. We can visualize the Casorati matrix as consisting of a vertical stack of pixel time series measurements as shown in Figure 9; in this fashion we note that time data is displayed along the horizontal direction while spatial data is distributed along the vertical. Therefore after computing the SVD, the columns of U will contain the *spatial singular vectors* while the columns of V will contain the *temporal singular vectors*.

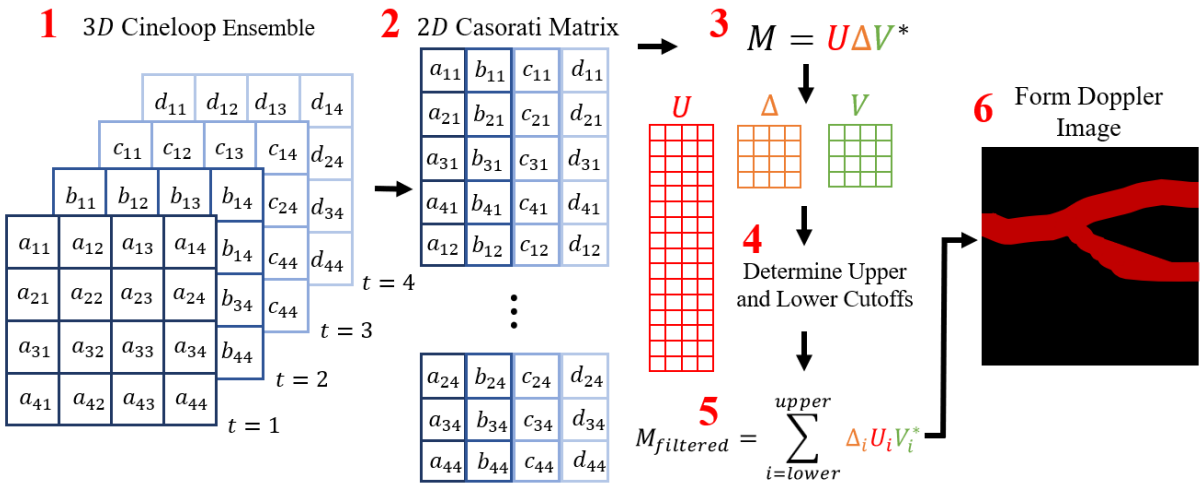

Figure 9: Clutter filtering using the singular value decomposition (SVD). **1)**The ultrasound cineloop is **2)** first vectorized into a 2D Casorati matrix, transforming the data structure from three to two dimensions. **3)** The SVD is then computed and **4)** some statistic(s) from the U, Δ, or/and V matrices are used to calculate the upper and lower cutoffs between which the blood signal is located. **5)** All singular values not between these cutoffs are replaced with 0's in the Δ matrix. **6)** The blood image is then reformed by multiplying together the U, Δ, and V matrices.

After computing the SVD of the vectorized ultrasound data we are left with the three U, Δ, and V data matrices. Let us make some observations. First for an initial ensemble size of K, we will have K values of Δ or singular values, i.e. $\Delta_1 ... \Delta_K$[4]. The i'th singular value is given by Δ_i while V_i and U_i refer to the i'th columns of the V and U matrices. We may also write the SVD in a slightly different form as shown in Equation 7.

$$M = \sum_{i=1}^{K} M_i = \sum_{i=1}^{K} \Delta_i U_i V_i^{\star} \tag{7}$$

Here the M_i's are referred to as the principle image components of M and can be calculated by taking the outer product of the U_i and V_i vectors. Although this is simply a rewriting of the definition of the SVD, this reformulation allows for the direct visualization of the signals contained in the individual image components aiding in the determination of the M_i's that contain blood.

The next step is to determine which of the K singular components contain the desired blood signal using a variety of parameters that can be extracted from the U, Δ, and V matrices. The Δ matrix contains the singular values of the decomposition, ordered from largest to smallest. Here, components which contain most of the total raw signal are readily identified. From the temporal singular vectors the mean Doppler frequency for each V_i may be calculated as described in Equation 2. From the spatial singular vectors the spatial

[4]This assumes that $K < N_x \times N_y$ which for relevant image sizes will always be true.

correlation matrix may be calculated (Equation 8) [29]. Here N represents the length of the vectors and $\bar{U_n}$ and σ_n represent the mean and standard deviation of the U_n indices. These *discriminating* factors are visualized in Figure 10.

$$C_{i,j} = \frac{1}{N^2} \sum_{i,j=1}^{K} \frac{(|U_i| - |\bar{U_i}|) \cdot (|U_j| - |\bar{U_j}|)}{\sigma_i \cdot \sigma_j} \tag{8}$$

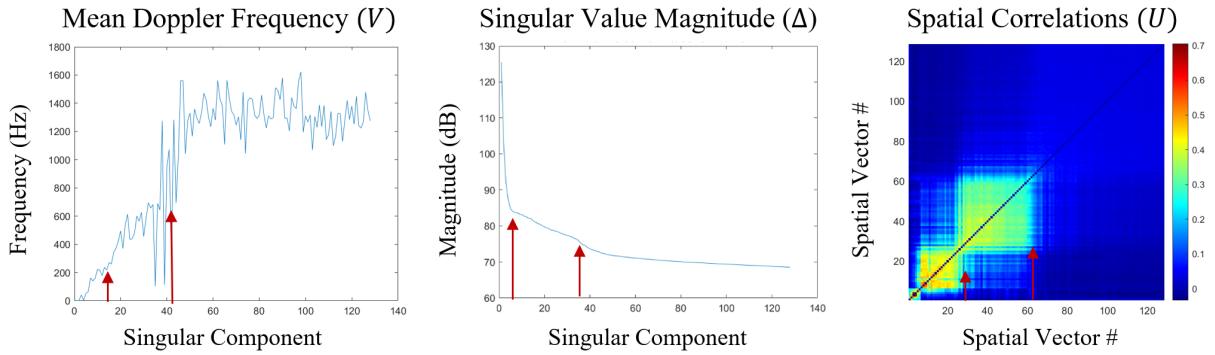

Figure 10: Discriminating factors for threshold selection. Upper and lower thresholds denoting the blood subspace may be calculated using temporal/frequency information derived from V, magnitude information from Δ, or spatial information from U. The red arrows show possible upper and lower thresholds using each discriminating factor.

Identification of the blood signal is generally dependant on a number of assumptions about the properties of blood, tissue, and noise signals. As follows;

1. That the *magnitude* of the tissue signal is far larger than that of the blood signal due to the tissue's far greater echogenicity and that generally far more of the viewing plane will contain tissue.

2. That the frequency of the tissue signal is generally low and that blood frequencies are generally higher.

3. That the blood or tissue signals will be strongly spatially correlated with themselves and poorly correlated with each other as they are contained in different pixels. Noise is uncorrelated.

Current SVD Filtering Algorithms and their Limitations

Current literature SVD filters use assumptions described above in conjunction with the SVD to define a lower threshold or cutoff in the singular values that separates the blood subspace from the tissue subspace. All singular components below this threshold are assumed to contain no blood signal and can be eliminated from consideration. Higher singular components are assumed to contain less and less blood signal and instead capture more and more random

noise. Therefore, a higher cutoff may be defined that separates the last of the blood signal from singular components that contain only noise. The final filtered signal may formed by replacing the Δ_i's above and below the cutoffs with zeros and multiplying through the U, Δ, and V matrices.

$$M_{filtered} = U\Delta_{filtered}V^{\star} = \sum_{i=\text{lower cutoff}}^{\text{upper cutoff}} \Delta_i U_i V_i^{\star} \tag{9}$$

Intuitively, an analogy can be made with reference to Figure 8 except that the x axis has been relabeled to 'Singular Component Number' with the SVD filter acting as a *bandpass* filter rather than a high pass filter and eliminating high and low frequency components[5]. It is extremely important to note that similarly to the 'inseparable' scenario shown in Figure 8 as it is not guaranteed that the blood and tissue signals will not mix, rather, it is hoped that the incorporation of spatial as well as temporal information into the filtering process adds an additional dimension of separation between the two signal types reducing potential mixing of signal types.

SVD clutter filters have been used extensively in the literature with scanline [26][27] ultrasound imaging although a modified approach is necessary due to the reduced ensemble sizes and sequential rather than simultaneous data acquisition. One such method, the *single-ensemble* SVD filter creates an Hankel matrix from the times series measurements of a single individual pixel. The SVD of the Hankel matrix is then computed and the frequencies of each singular component (calculated using the lag-one autocorrelation [44]) used to estimate and remove components containing clutter. Another scanline SVD filtering method, the *multi-ensemble* formulation, uses a data matrix constructed from the surrounding pixels as well as the sample pixel to obtain a better estimate of the blood components [26][27].

In high-frame-rate ultrasound, much larger ensemble sizes are achievable (by more than a factor of 10) and sampling takes place over the entire imaging plane. Therefore the data matrix is much larger, being of size $(M \times N) \times EnsembleSize$ where M and N are the dimensions of the ultrasound image.

The central tenet of SVD clutter filters is that the blood, tissue, and noise sub-spaces are separable from each other at a given threshold or point and that these thresholds can be identified using information contained within the U, V, and Δ SVD data matrices (Figure 10).

Common threshold estimators making use of the singular values (Δ) involve pre-selected fixed cutoffs [28][49], the turning point of the singular value curve [50][51] (indicating significant change in signal statistics), a proportion of the total signal strength [33] (assumption of a relative flow to clutter energy ratio), etc. Temporal information may also be leveraged and the threshold estimator set as an arbitrary predefined frequency (analogous to the cutoff frequency in a frequency filter) or turning point [50] (representing a significant change in frequency statistics). Spatial analysis has been performed using the spatial correlation matrix (Equation 8) and flow threshold identified through a block-matching algorithm [33] (clutter

[5]This is simply an analogy as the SVD does not function on the basis of frequency.

and flow signals possess a different spatial distribution). A good review of various threshold estimators and their relative performance can be found in Ref. [29] which extensively compares about a dozen theshold identification methods. Additionally, this work concluded that a spatial estimator using the subspace boundaries of spatial similarity index generally demonstrated the best clutter filtering performance.

This best-performing algorithm from Ref [29] using the spatial similarity matrix C (Equation 8) involves identifying that the blood and tissue subspaces should appear as two 'blocks' of correlations in C (Figure 10). This phenomena can be numerically quantified when two limit values within the ensemble size n_t are identified $1 \leq a \leq b \leq n_t$ (representing the lower and upper thresholds) such that the value of $\chi_N(a,b)$ is maximized as shown in Equation 10

$$\chi_N(a,b) = \frac{\chi(C, \alpha_{a,b})}{\sqrt{\chi(C,C) \cdot \chi(\alpha_{a,b}, \alpha_{a,b})}} \tag{10}$$

where

$$\chi(C, \alpha_{a,b}) = \frac{1}{n_t^2} \cdot \sum_{(n,m)} \big(C(n,m) - \bar{C}\big) \cdot \big(\alpha_{a,b}(n,m) - \bar{\alpha_{a,b}}\big) \tag{11}$$

Here n_t is the ensemble size and $\alpha_{a,b}$ is given by

$$\alpha_{a,b} = \begin{cases} 1, & \text{if } (n,m) \in [1,a) \\ 1, & \text{if } (n,m) \in [a,b] \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

Although time consuming to compute, this method has been frequently used in the literature [33][52] and is considered one of the more robust threshold identification methods although it only makes use of the spatial U information present in the SVD. While most threshold identification techniques in the literature consider only a single estimator to identify the blood subspace boundaries there are a few that do leverage multiple estimators to perform more robust subspace identification. Consider the following algorithm and its inherent assumptions to identify the upper and lower blood subspace boundaries from Ref. [50].

The lower clutter-flow threshold following Ref. [50] is obtained by first computing the turning point of the singular value curve. This turning point is then compared with the point at which the Doppler frequencies of the components exceeds some predefined cutoff, say 50 Hz. The lower threshold is defined as the greater of the two potential cutoffs. The upper flow-noise cutoff is defined in a similar way. As they are assumed to contain only noise [50], the Doppler frequency of the noise components is assumed to be be generally high and quite consistent; a potential 'pre-cutoff' threshold can be identified as the index at which this 'plateau' in the Doppler frequency disappears (this effect can be seen somewhat in Figure

10). It is assumed that for noise, the singular values (log-scaled) above the upper cutoff should follow a linear relationship with respect to the index of the decomposition; the points above the 'pre-cutoff' are used to identify the portion of the singular value magnitude curve used for the linear fit. The final upper flow-noise threshold is set as the index at which the singular value magnitude curve no longer maintains a linear relationship with the singular value index as a potential cutoff. This process is shown in Figure 11

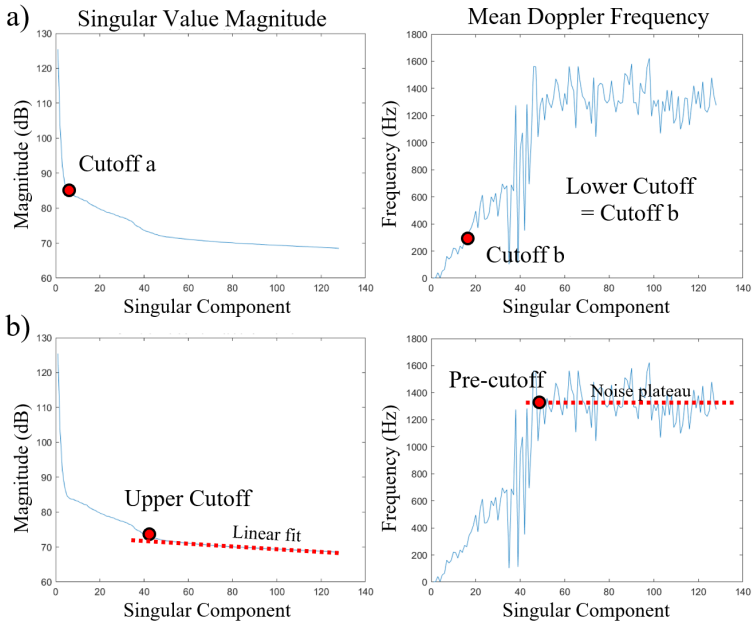

Figure 11: Hybrid upper and lower threshold estimator. *a*) Lower threshold identification. The lower threshold is defined as the higher of cutoffs a and b where cutoff a is calculated as the turning point of the singular value magnitude curve and cutoff b is defined as the singular index at which the mean Doppler frequency curve first crosses some predefined frequency threshold, here 300 Hz. *b*) Upper threshold identification. The upper threshold is calculated by first identifying the edge of the noise frequency ceiling as the pre-cutoff. A linear fit on the singular value magnitude curve using the indices above the pre-cutoff then gives the upper cutoff as the point at which the singular value curve begins to diverge away from the linear fit.

The two SVD filtering algorithms presented here represent the more robust literature SVD techniques and their performance under a variety of *in-vitro* and *in-vivo* imaging conditions will be discussed in the following chapters. However, both methods make many assumptions about the signal statistics of the input data matrix or how the SVD actually performs the decomposition. The breaking of these assumptions is shown in Figure 12. Here it is seen that the Doppler frequency plot of the decomposition or the spatial correlation matrix is often too noisy to extract meaningful information. Figure 12 also shows how both methods break down in a (relatively) simple *in-vitro* imaging scenario using a spiral flow phantom [53] with aliasing present. Finally, we remark that in the SSM in Figure 10 there are actually three 'blocks' of correlations present; the tiny clutter block (components 1 - 5) and the forward (components 6 - 30) and reverse flow (components 30 - 63) in an aliased environment. This

particular image was used illustrate how easy it is to draw incorrect conclusions from plots of discriminating factors, and the breadth of analysis required to 'certify' the accuracy and robustness of a filtering algorithm.

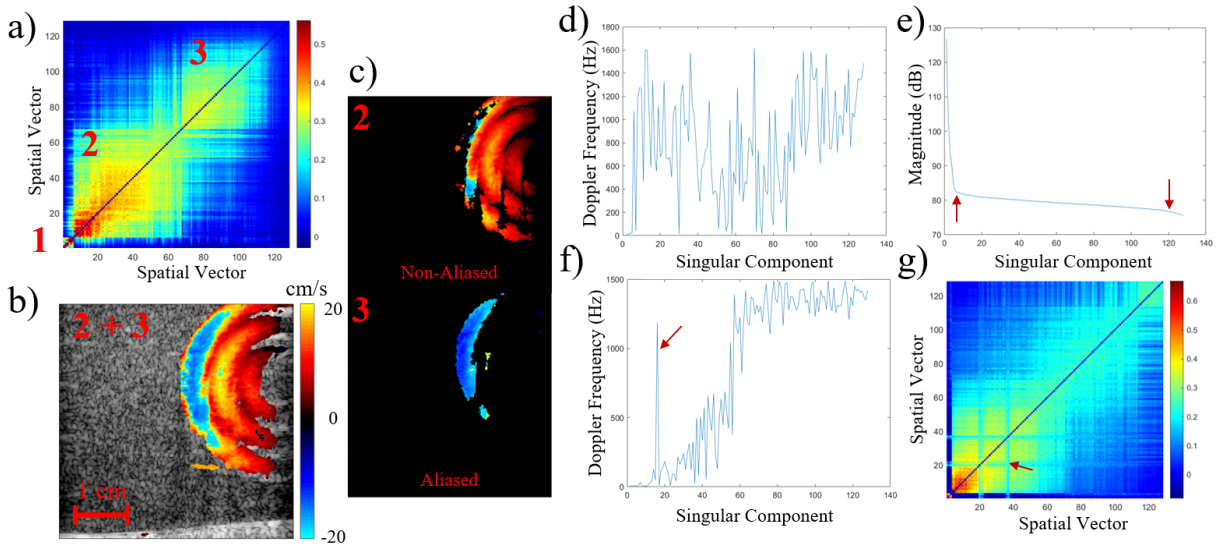

Figure 12: Limitations of the described SVD filters. *a*) - *b*) *In-vitro* color Doppler of a spiral flow phantom [53] offset to the edge of the imaging view. Three blocks of correlations are observed in the SSM. Block 1 corresponds to tissue clutter while blocks 2 and 3 correspond to the non-aliased and aliased flow as shown in *c*). *d*) - *e*) show the mean Doppler frequency curve and singular magnitude curve for the same spiral flow phantom adjusted to lie in the center of the imaging view. The difficulty in identifying the pre-cutoff in mean Doppler frequency curve is observed as there does not appear to be any noise plateau in the plot. The arrows in *g*) show the actual identified upper and lower thresholds – we can observe that the flow subspace encompasses nearly the entire decomposition. *e*) The Doppler frequency plot for an *in-vivo* acquisition of a cross-section of the human carotid. Here we can observe a huge peak in the frequency that will stymie many frequency-based threshold algorithms. Finally *g*) shows a scenario (*in-vitro* flow phantom) where significant sub-structure can appear in the SSM, making the application of the algorithm described in Equation 10 difficult.

However, it is very easy to see that not all blood singular image components between the upper and lower threshold contain the same 'quality' of blood image. This is observed in Figure 13 which shows the same femoral artery as seen in Figure 5, albeit at a different moment in the cardiac cycle. The final filtered signal is shown on the left while filtered images corresponding to selected singular components (labelled in red) are shown on the right. To enhance clarity we use a relative scale and show these images under 15 dB dynamic range. Under visual identification it is obvious that the first singular component contains tissue signal and the twentieth noise and that under a thresholding approach we would likely describe the blood signal as being contained between components 8 to 18. However, we note that the signal to noise ratio of component 15 is very low. Little blood signal is observed in this singular component and it appears to be dominated by some 'beamlike' noise artifacts.

With a vanishing amount of blood signal but significant noise, component 15 contributes little to the image and if removed will improve the quality of the final filtered signal.

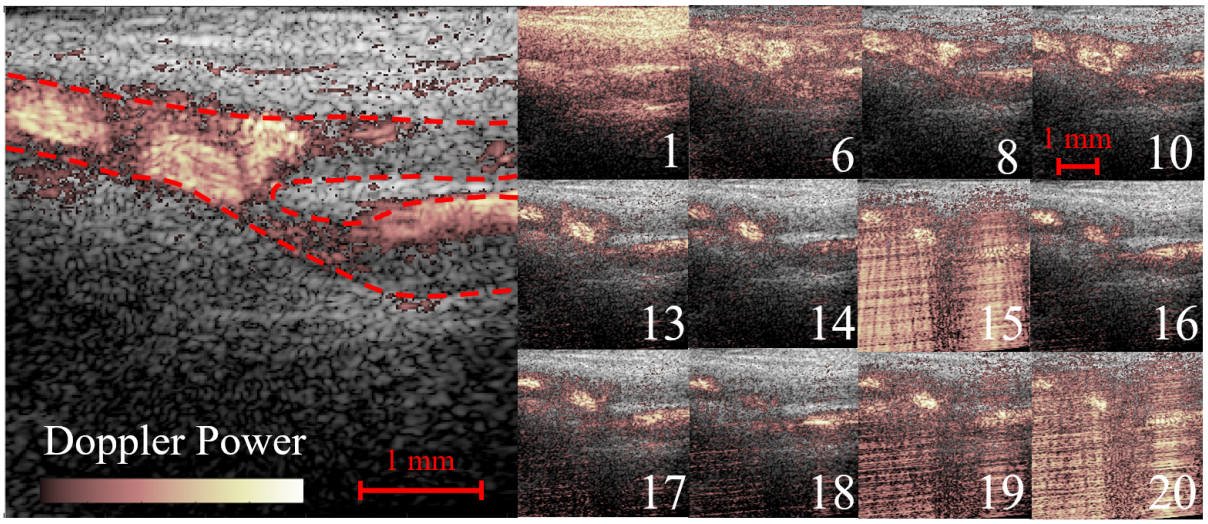

Figure 13: *Left* Filtered flow image of the femoral bifurcation created using SVD clutter filtering (20 *dB* dynamic range). The red dashes indicate the arterial walls. *Right* Image formed using the corresponding singular component (lower right). An ensemble size of 128 was used. Images here are shown under varying but appropriate dynamic range. Visually, tissue, blood, and noise are observed in the component images.

Previous work in the literature has explored the idea of *component selection* algorithms which do away with the notion of thresholds but instead attempt to classify singular components individually as containing primarily tissue, blood, or noise signals [30][31]. This method has significant merit, however it is not without significant caveats. In this algorithm, the singular value magnitude, the mean Doppler frequency, and the spatial correlation with respect to the first spatial singular vector (the values contained in the bottom row of the SSM) were first identified for each component in the decomposition. K-means clustering [54][55] was then performed with the clutter, flow, and noise signals forming clutters and allowing for the flow-containg components to be identified. Although signficant issues can be readily identified in the inherent assumptions being made using this approach[6] it nevertheless remains one of the few algorithms that attempts to identify components individually and does so using the statistics from the spatial, temporal, and magnitude data matrices simultaneously. Our proposed framework, discussed in the following chapters, will maintain these key properties.

It is also useful to discuss some of the other strategies that have been employed in SVD clutter filtering algorithms to enhance performance. In high-frame-rate ultrasound, much larger ensemble sizes are achievable than with scan-line imaging and sampling takes place

[6]Such as a presumption of three clusters – how would the algorithm perform if no flow was present? Furthermore the 'clusters' that were formed tended to be wormlike or otherwise irregularly shaped which does not lend itself well to K-mean clustering where clusters are assumed to be more spherical in shape. Additionally, K-means uses a distance-based metric for clustering meaning that the normalization on the discriminating factors needs to be determined or even adaptively set.

over the entire imaging plane. Many SVD clutter filters have been designed such that the entire beamformed image is used, perhaps 256 by 256 pixels, as the data matrix. We will call these types of SVD filters *global SVD*. Alternatively, the image may be further subdivided into overlapping *blocks*. In *blockwise* SVD, the beamformed image is divided into overlapping blocks of pixels, each of which undergoes the process described in Figure 9 with threshold identification being performed as described above. These filtered sub-images are then averaged to create a final image [50] encompassing the entire imaging view.

A blockwise approach to filtering was shown to greatly improve flow detection and contrast [50] due to the reduction of the overall 'signal bandwidth' present in the data matrix. As the data matrices of individual blocks contain only clutter and flow signals 'local' to the region, the total contained clutter and flow signal distributions are less complex, and the decomposition is better able to separate blood and tissue components within the defined ensemble length. For example, consider a scenario where the flow signal is highly localized to a region of the image (such as an artery or vein); under a blockwise approach, SVD clutter filtering can be performed for blocks that do and do not contain the flow signal, resulting in increased sensitivity in the flow blocks as flow/clutter discrimination can be performed without the inclusion of clutter signals from further regions of the image. By removing signals that possess different statistical properties from the data matrix, greater sensitivity can be obtained for the pixels in the region of interest.

One major issue with blockwise SVD is that *line-like* artifacts may be observed at the boundaries of the overlapping blocks[7] due to differences in the selected upper and lower thresholds [50]. One solution involves the use of randomized spatial downsampling whereby the 'blocks' are formed by constructing the data matrix from randomly selected pixels from the entire image [49][56]. The general SVD cluttering filtering process is then performed on these randomly created blocks. As pixels have been selected randomly, the line-like artifacts frequently observed in blockwise SVD are noticeably absent. It is important to note an interesting property of the Casorati matrix; due to the stacked arrangements of the times series of each pixel (i.e. each row in the Casorati matrix contains the time series of a given pixel) rows may be shuffled with no change in the end results of the decomposition; in other words it makes no difference how information is 'packed' into the Casorati matrix.

[7]These artifacts can generally be eliminated if appropriate 'block' parameters with sufficient overlap are chosen. This may greatly increase computational demand however.

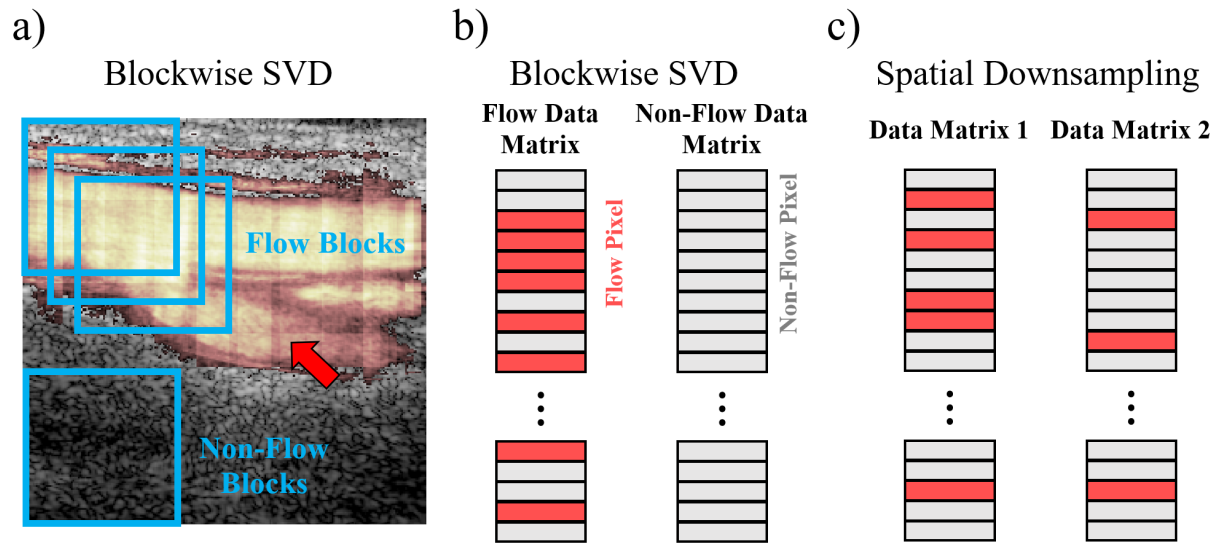

Figure 14: *a*) A power Doppler image of the femoral bifurcation computed using blockwise SVD (cyan squares represent the data matrices of each individual block) showing line-like artifacts (red arrow) due to insufficiently chosen block and overlap sizes. As the block size and overlaps are not commensurate, the artifacts are irregular. *b*) and *c*) The data matrices of flow and non-flow blocks. Under blockwise SVD flow and non-flow blocks contain only local pixels and the clutter filtering algorithm can identify the flow and clutter signals in the flow data matrix and that the non-flow matrix contains only clutter signals. If we were to randomly spatially sample the image to form the data matrices we would end up with a mix of flow and clutter signals in all, resulting in a less cohesive blood signal that is more difficult to identify – the flow signal is distributed across more data matrices.

However, randomized spatial downsampling is not without its own set of drawbacks. Randomized spatial downsampling does not perform any better than the more generic blockwise SVD; it is simply much faster as it does not require the very significant overlap between blocks that is required for blockwise SVD to eliminate the line-like artifacts. However, we remark that randomized spatial downsampling only performs well if blood and clutter signals are present and possess similar statistics across the entire image (applications such as kidney or liver imaging). If however, flow is present in only a small portion of the imaging view (imaging of an artery or vein) then using randomized spatial downsampling will result in data matrices with only a few flow pixels in each, resulting in very low flow to clutter ratio and poor clutter filtering performance (Figure 14).

Another common literature technique used to lessen the computational requirements of SVD clutter filtering remarks on the fact that generally the flow subspace is contained within the first half of the components of the decomposition. Therefore, one can save significant computational time by only calculating[8] the first half ranks of the SVD, as the higher index components would contain only noise and would be zeroed out anyway [56]. However, we remark that this approach is not generically robust – Figure 12 shows how in certain cases flow can be found in even the second last component of the decomposition.

[8]Actually an approximation.

The breaking of the global data matrix into smaller matrices that otherwise undergo the same clutter filtering process allows for advantages other than increased sensitivity in the region of interest. First it is trivially easy to parallelize the operation of multiple blockwise SVD calculations allowing for better performance on multi-threaded CPUs and GPUs [56]. Secondly, the computational demands of the reduced SVD of a m by n matrix generally scale as $O(m * n * min(m, n))$[57] and thus it is significantly more efficient to compute the SVD of several smaller matrices than one large matrix until there becomes overlap between blocks. We will also note that the main advantage of a blockwise approach is that the target data matrix is always divided into blocks of the same size making it easier to generalize classification algorithms or machine learning models to data acquired under different imaging scenarios.

In the framework proposed in this thesis we will elect to use blockwise SVD, due to its inherently better clutter filtering performance. The subdivision of the input signal into uniformly sized blocks further improves generalizability to machine learning models, allowing for greater independence with respect to imaging acquisition parameters. Although randomized spatial downsampling or reduced rank calculations have the potential to speed up computation, they do not increase filtering performance and can perform poorly under select conditions. Therefore, we will not implement them in our framework.

Different but *SVD-like* Clutter Filtering Algorithms

Although outside the focus of this thesis, clutter filters making use of other non-SVD decomposition methods also frequent the literature. We note that there are many forms of blind source separation that have been applied to the problem of separating out the clutter and flow signals. While SVD (or similar principle component analysis techniques) are very common in the literature, the use of independent component analysis has also been demonstrated to separate flow and clutter [58][59][60]. Here the key idea is that while principle component analysis techniques create a representation of the source signal in terms of mutually orthogonal bases corresponding to the direction of maximal variance, independent component analysis is not constrained by these restrictions and instead identify maximally independent bases into which the source signal is decomposed. Other techniques involving sparse and low-rank matrix decompositions[61][62] exist in the literature as well. A good review of difference blind source separation techniques and their role in ultrasound can be found in Ref [63].

Introduction to Machine Learning

Machine learning refers to the development of models or frameworks that automate the adaptive modelling process to improve performance on some task with minimal human input. These models can vary in complexity from simple clustering/classification and regression analysis to neural networks with millions of parameters capable of identifying street numbers from photographs to 95% accuracy[64]. The goal of this thesis work will be to produce a

machine learning framework that is capable of identifying the flow containing components of the singular value decomposition. The remainder of Chapter 2 will provide a brief general introduction to machine learning and neural networks with the details of the proposed framework being introduced later in Chapter 3.

Supervised and Unsupervised Learning

Machine learning models are largely broken up into two categories *supervised* and *unsupervised* learning.

Unsupervised learning attempts to analyze unlabeled datasets to identify hidden patterns or trends without human intervention. Here the ability to discover statistical similarities or differences in the dataset can be of great use in image or pattern recognition. For example, the multifactorial SVD algorithm discussed earlier [30][31] used K-means clustering to attempt to separate and identify the clutter, flow and noise signals. Figure 15 shows a reproduction of the general method employed in Ref. [30] to identify the clutter, flow, and noise clusters. Here the component elements of the decomposition self-separate into clusters (albeit poorly) based on their shared statistics. The clusters can then be labelled through knowledge of assumed signal statistics; the clutter cluster will be high in magnitude, strongly correlated, and low in frequency while the noise cluster should be low magnitude, high frequency, and uncorrelated. As discussed previously, k-means clustering is capable of successfully identifying the flow components of the decomposition, however, the clusters tend to be ill-formed, raising issues of normalization and robustness.

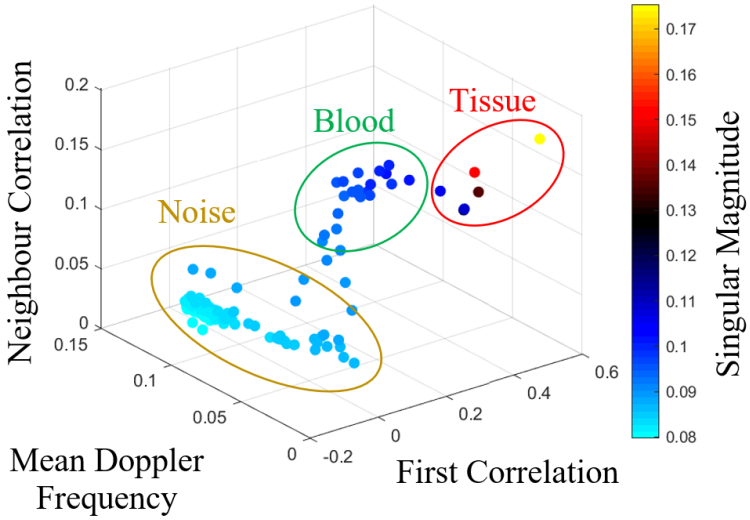

Figure 15: The result of k-means clustering performed on the signal statistics of each component of the decomposition. Here the mean Doppler frequency, magnitude, and two spatial correlation statistics (described in Chapter 3) were used for clustering.

Unsupervised learning is generally difficult as the output is not known in advance and can be difficult to 'force' into a desired prototype. For example, the above clustering has

been performed with number of clusters set equal to 3 ($k = 3$). However, what if no flow was present in the data matrix? Under these conditions we would expect only two clusters, one for the clutter/tissue signals and one for the noise to be present. As this cannot be directly communicated to the algorithm the algorithm must possess some means of determining whether to cluster with $k = 2$ or $k = 3$. This greatly increases difficulty and generally results in lower than desired accuracy.

An advantage of unsupervised learning is that it avoids having to label the dataset, which may be difficult or impossible. In this case here, we would need a method to identify or numerically quantify the amount of 'flow' in each component (which we discuss in Chapter 3) to obtain a label.

In **supervised** learning, the algorithm, provided with a corresponding set of labels to the input dataset, attempts to identify the relationship between the labels and the inputs. The trained model can then be used to assign labels to unseen data. In comparison to the k-means algorithm described above a supervised learning approach would provide the algorithm with the same statistical input data (such as the frequency, magnitude, and spatial correlations shown in Figure 15) as well as a corresponding labels (i.e. flow or non-flow component). The algorithm would then be tasked with identifying how the inputs relate to one another and the target labels and *self-adjust* itself to be able to generically identify flow and non-flow components when presented with similar but unseen new inputs. Supervised learning always functions through the use of a *loss* function, that relates the degree of similarity between the network's current output and the desired output. As the loss is minimized, the network's output more closely resembles the target output.

Supervised learning approaches can be further divided into two classes of problems: *classification* and *regression*. Classification algorithms are used when the labels are categorical (i.e. labels such as *flow* and *non-flow*) and regression when the outputs are continuous variables (i.e. probability that component is a flow component, or a 'score' on how much blood is contained within each component, etc.).

Neural Networks

A very common type of supervised learning approach is the neural network – groups of 'neurons' related by weighted connections to one another. Input fed into the first or input layer is broadcast through the network, propagating layer by layer and modified and controlled by the weighted connections, to arrive at the final output layer, which describes the 'decision' that the network has made. The final output layer could consist of a single neuron that is activated or deactivated, representing a binary decision (i.e. a 'yes' or 'no' model), or possess some continuous value (i.e. a probabilistic model) or consist of multiple neurons with discrete continuous values representing more complex models. A generic two layer network is shown in Figure 16. As every neuron in a given layer is connected to every neuron in every preceding or following layer, this is called a *deep* network.

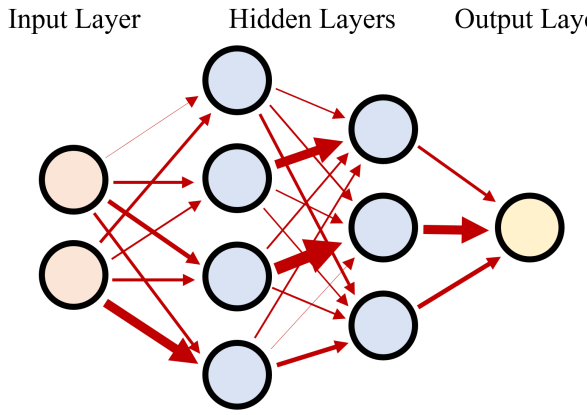

Figure 16: A two layer deep neural network. The width of the arrows connecting each layer of neurons represents the magnitude of the weights.

The foundational unit of the neural network is the neuron. As shown in Figure 16 a neuron accepts a number of inputs $(x_1, ..., x_n$, multiplied by weights $(w_1, ..., w_n)$ to produce an output y. However, there are two additional factors present, **1)** an additional bias term b, and **2)** an activation function f that serves to introduce non-linearities into the network. Without a non-linear activation function the network would be purely linear and unable to approximate more complex models. The output of an individual neuron is shown in Equation 13. Some examples of common activation functions are shown in Figure 17.

$$y = f\left(\sum_{i=1}^{n}\left[w_i x_i\right] + b\right) \tag{13}$$

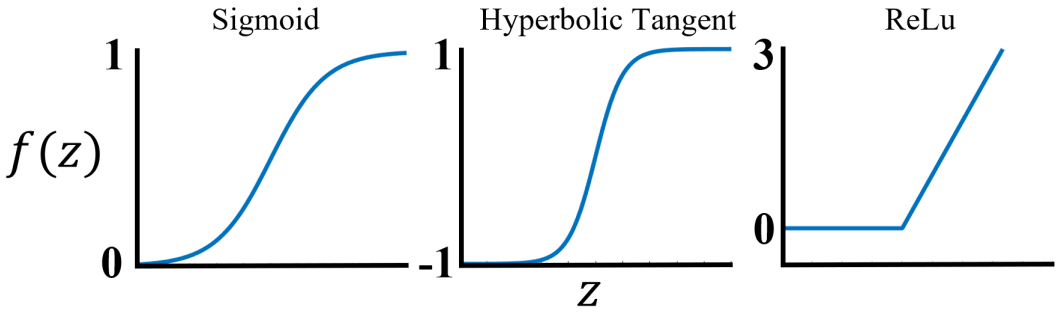

Figure 17: Some common activation functions. ReLu refers to rectified linear unit and serves as a thresholding function.

When the weights or biases of the neurons are adjusted, the outputs of the activation functions changes resulting in a different network output or prediction. To obtain correct output answers, the optimum weights and biases must be continuously updated in a process known as *training*.

Training Neural Networks and Other Considerations

Network training is composed of two fundamental techniques: *gradient descent* and *back-propagation*.

Gradient descent is the process of minimizing a loss function $\mathcal{L}$ with respect to the free parameters of the model. Loss functions take many forms but one of the most commonly used is the mean squared error (MSE) which is the average over n samples of the networks output y_i compared against the expected output y_i'.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i'\right)^2 \tag{14}$$

Mean squared error is used extensively in other applications; for instance the optimal fit when performing linear regression is the line that minimizes the summed distance squared of each data point from the line of best fit to the data. Similarly we can say that a model which has its weights and biases set to minimize the loss function (MSE) [65] will have much less prediction error than a model with different values for its weights and biases that gives a much higher loss.

Minimization of the loss function is achieved through gradient descent. The network output y_i is defined for a set of input weights w and biases b – i.e. we can consider the network as a function. Therefore, if the gradient of the loss function (the difference between the network output and the expected output) were to be obtained, the minimum of said function would represent the ideal set of weights and biases that would minimize the loss function. Here the gradient of the loss function would simply be the partial derivatives of the loss function with respect to each individual weight and bias: $\nabla f = \left(\frac{\partial \mathcal{L}}{\partial w_1}, ..., \frac{\partial \mathcal{L}}{\partial w_n}, \frac{\partial \mathcal{L}}{\partial b_1}, ..., \frac{\partial \mathcal{L}}{\partial b_n}\right)$. Successive 'steps' down this gradient will bring the network closer and closer to the minimum and therefore the set of weights and biases that would most accurately reproduce the input samples. Equation 15 shows how weights are updated ($w \rightarrow w'$) in each step down the gradient of the loss function. Here α refers to the *learning rate* or a measure of how large the steps down the gradient are. Figure 18 shows a high level overview of this process for a model with a single weight and bias (single neuron) in two dimensions.

$$w' = w - \alpha\left(\frac{\partial \mathcal{L}}{\partial w}\right) \tag{15}$$

For models with a large number of free parameters, the process of gradient descent, computed for each individual sample, can become quite computationally demanding. To address this, the gradient can be approximated over a number of training samples or a *batch*. The batch gradient can simply be computed as the sum of the gradients of each individual sample in the batch as the loss function is the sum over individual training samples. The process of gradient descent is repeated until all training samples have been used, a period known as an *epoch*. Training generally takes a number of epochs, depending on the model

and the training data, until the loss function has been suitably minimized – when further training no longer reduces the loss.

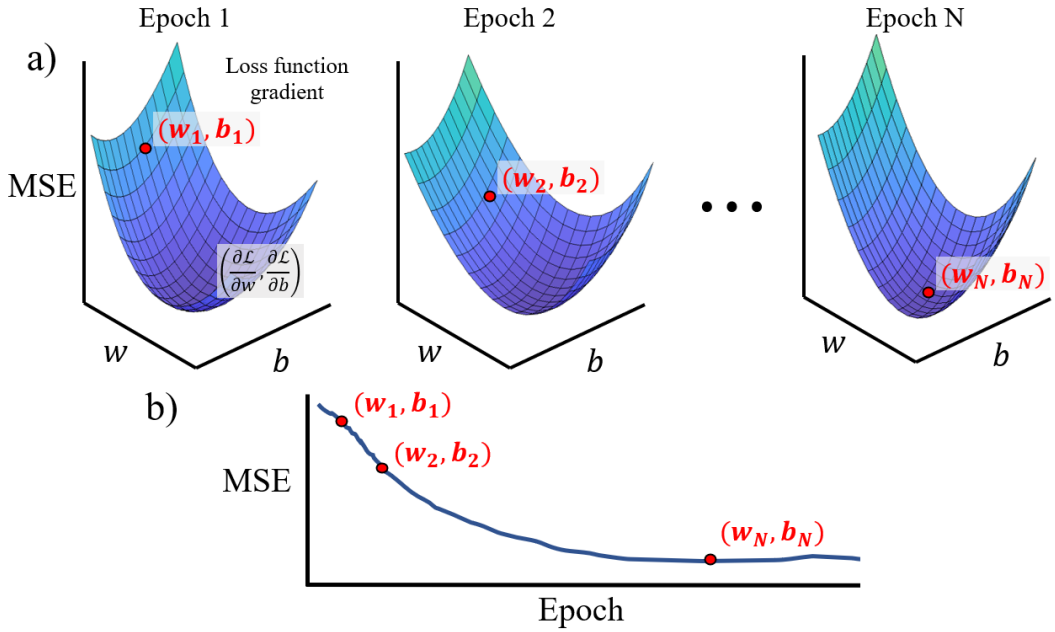

Figure 18: An example of gradient descent for a single neuron model. *a*) The initial weight w_1 and bias b_1 have significant loss or here, mean squared error. After an epoch of training a new set of weights and biases has been obtained, a new gradient computed, and a 'step' down the gradient performed. This process repeats itself until the set of weights and biases that results in a minimum of the updated loss function gradient has been obtained (Epoch N). *b*) Training is performed until the loss function shows no further improvement in decreasing the MSE.

To compute the gradient of the loss function, equivalently the partial derivative of the loss with respect to the free parameters of the model, a method called backpropagation [66] is used. Without going into too many details, backpropagation involves heavy use of calculus's chain rule to relate the partial derivatives of the loss function with respect to free parameters at the beginning of the network to those at the end of the network. An example of backpropagation for a very simple network with no biases or activation functions is shown in Figure 19.

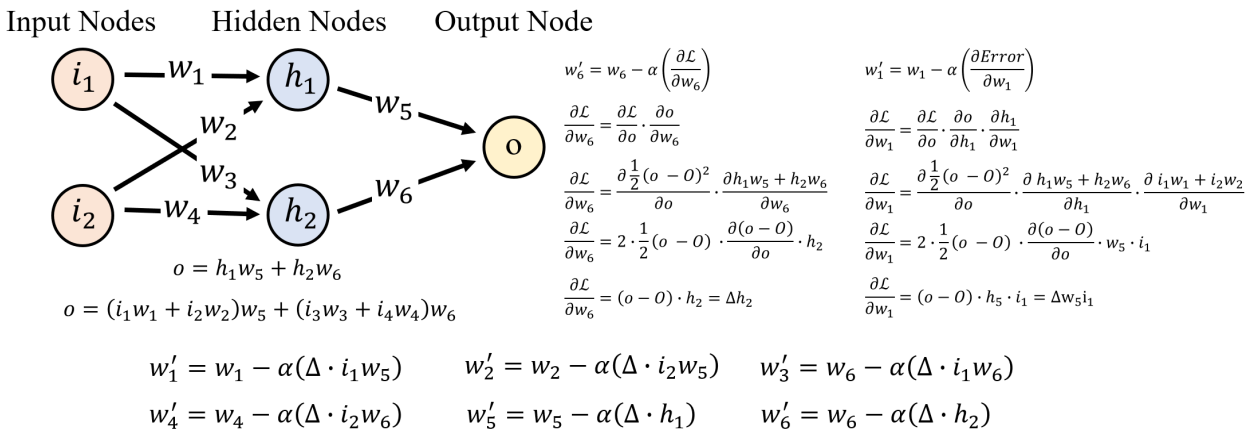

Figure 19: Backpropagation in an extremely simplified network (no biases or activation functions). Two input nodes, two hidden nodes, and one output node is present. The output of the network is given as o and the expected result is O. The MSE is used as the loss function (a $\frac{1}{2}$ term is added to simplify the derivatives) and the chain rule is used to represent the partial derivatives of free parameters at the beginning of the network to the output produced at the end of the network. If biases and activation functions were present they would simply appear as additional terms in the derivative (and changes in their values would be calculated exactly the same as for the weights).

The overall training process of a neural network is shown in Figure 20. Note the presence of a *validation set* in the data; it is good practice to withhold some of the dataset from the training set and use these samples to test the networks response to similar but different information to identify if the network is still maintaining sufficient generalizability, rather than adapting to the **exact** data samples being fed to it. Therefore, by tracking the loss on the validation dataset we can identify whether the network is generically improving performance or over-fitting on the training data by tracking the loss curves for the training and validation datasets. If the validation loss curve shows no improvement after training for a certain number of epochs while the training loss curve continues to decrease, we should stop training to avoid overfitting. Once we have identified the point at which training is no longer improving the loss training is halted and the network can be applied to the samples which we desire to make a prediction on. This *unseen* data should always be isolated from the network to avoid potentially biasing the network or the training process.

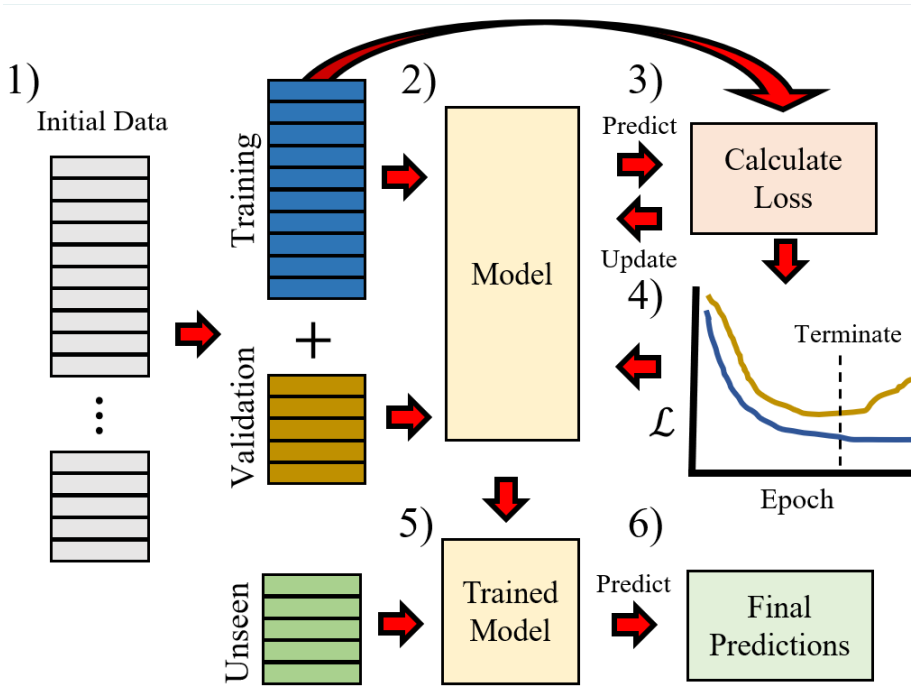

Figure 20: The general training process. *1*) The dataset is subdivided into training and validation subsets. *2*) - *3*) The network is trained using the training data. The loss is calculated and used to update the weights and biases of the network. *4*) During the training process, the loss of the training and validation sets are tracked. Training is halted when further training no longer results in a decrease in the loss. *5*) The trained model is then used to make predictions with the unseen data.

Part III

Flow Component Identification using a DNN

As described in the previous Chapter, our goal is to develop a neural network trained using statistics extracted from the U, Δ, and V SVD data matrices to output a target quantity that indicates the likelihood of a given component being a flow component. As labeled samples will be required to train the model, the target quantity must be well defined, posses physical meaning, and must be computed for all samples prior to model training. In this thesis we will use the *area under the curve* (AUC) of the *receiver operator classifier* (ROC) curve, a metric that has been comprehensively used to quantify the performance of clutter filters in the field of ultrasound [30][67][68], extending its use from assessment of the final filtered image to instead assess the presence of flow within each singular image component of the decomposition (Figure 13).

As inputs for the network, we will expand upon the discriminating parameters described in Refs. [30] and [31], investigating the SVD data matrices for additional 'fingerprints of flow' that can be used to increase the sensitivity to flow of our model. We will then describe the specific details of the data acquisition procedures, the subdivision of acquired data into the training, validation, and testing datasets and the required reprocessing and normalization before the data can be used for training. Finally we will discuss some of the performance and evaluation metrics used to evaluate clutter filters.

Proposed Model

Using the ROC to Quantify Flow within the Decomposition

The ROC curve describes the diagnostic performance of a binary classifier as the sensitivity threshold is varied and is created by plotting the true positive rate (TPR) or *sensitivity* against the false positive rate (FPR) or $(1-specificity)$ at various discrimination thresholds [69]. Different points on the ROC curve therefore represent the true positive vs false positive rate at different decision thresholds.

Figure 21 shows the performance of two classifiers. Note the difference in the overall structure of the two curves indicating different levels of predictive ability at different discrimination thresholds. The area under the curve (the two-dimensional region between the curve and the x-axis) can be used as an aggregate measure of performance across all classification thresholds. This AUC value ranges from 0, representing a perfect anti-classifier (always wrong), to 1, representing a perfect classifier. A value of 0.5 therefore is representative of a model that has no predictive ability.

There are several advantages and disadvantages to using the AUC. First, note that the area under the curve is scale invariant; it ranks classifiers on an ordinal scale but does not quantitatively describe which classifier performs better. While avoiding the issues of having to deal with absolute values, this means that the AUC can be difficult to utilize in terms of real physical meaning. Furthermore, scale-invariance is not always desirable; we may wish to prioritize one type of error over another depending on application. For instance, in some applications we may wish to have more sensitive flow detection, even if it means a lot of clutter will also be present. Nevertheless, the AUC remains one of the best ways to rank

classifiers as it measures the quality of the model's predictions irregardless of the chosen classification threshold.

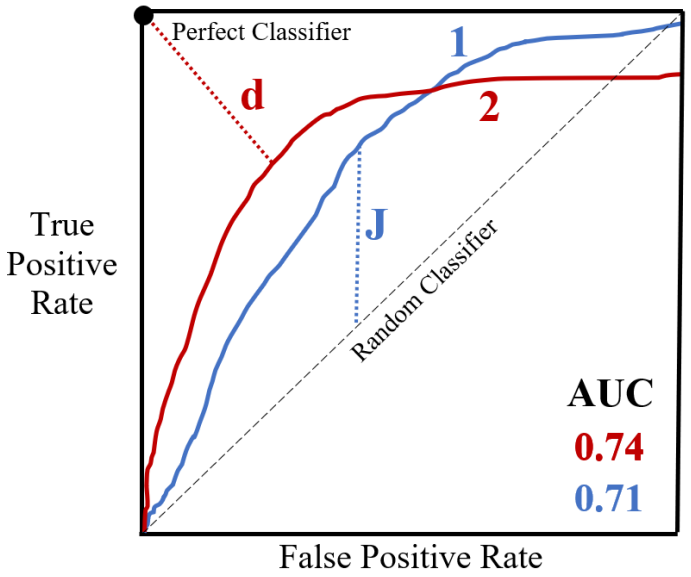

Figure 21: The ROC curve of two classifiers. Overall classifier 2 is superior to classifier 1 and indicated by the higher area under the curve value. A perfect classifier will have an AUC of 1 and pass through the upper left hand corner of the plot. The point giving the smallest distance d between this $(0, 1)$ point and the ROC curve d is one metric used to obtain the optimum cutoff. Another metric involves maximizing the vertical distance J between the diagonal random classifier line and ROC curve.

An area under the curve of 1 represents a perfect classifier (Figure 21). This curve will pass through the upper left hand corner of the plot with sensitivity = 100% and specificity = 100%. If the curve does not pass through the upper left hand corner (which will be true for almost all real-world classifiers), then we can find the point on the curve that minimizes the distance between the curve and the upper left hand corner i.e. $d^2 = (1-TPR)^2+(FPR)^2 = (1 - sensitivity)^2 + (1 - specificity)^2$. This threshold can be said to be the optimal cutoff or trade-off between minimizing the false positive rate while simultaneously maximizing the true positive rate. Another technique of determining the optimal threshold is performed by maximizing the vertical distance between the ROC curve and the diagonal 'chance' line (called the Youden's Index J [70]), maximizing the difference between the TPR and the FPR. We will use these thresholding techniques later in this thesis to determine the appropriate sensitivity thresholds for the power and color Doppler images of the proposed clutter filtering framework.

Clutter filters are assessed using the area under the ROC curve [68] as follows (Figure 22). First, the clutter filtered signal is obtained and its power Doppler computed. Next, the *region of interest* (ROI) which contains blood is defined. This segmentation can be automatically performed by segmentation algorithm, but we note that flow segmentation is difficult[71][72] especially as segmentation must also be performed during diastole when the flow signal is weak (due to the movements of tissues *in-vivo* during the cardiac cycle). For

this reason, manual segmentation we will use manual segmentation. We will then sweep intensity from the lowest to the highest measured pixel intensities in the image, recording the percentage of flow pixels (TPR) and tissue pixels (FPR) above the intensity threshold at each intensity value. These two values are plotted against one another to generate the ROC curve, and the AUC numerically computed. A higher AUC value indicates a better performing filter.

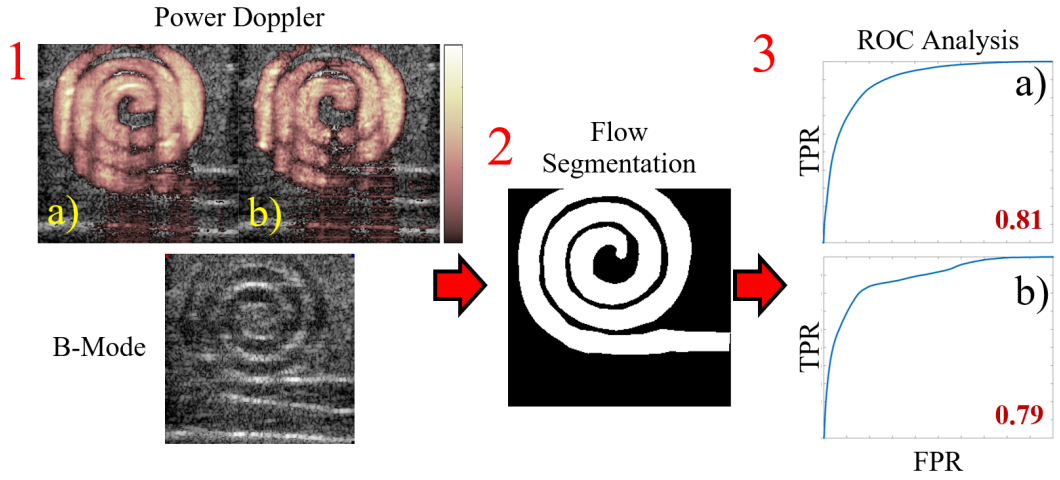

Figure 22: Filter performance evaluation through the ROC curve. *1*) The power Doppler (signal intensity) images produced through different clutter filters are obtained. Together, with the b-mode image the reference flow pixel region *2*) is identified. *3*) The ROC curves are then produced and the AUC is numerically computed from them. The filter with the greater AUC (filter a) is ranked as the best performing filter.

In this thesis, we propose to extend AUC analysis from assessment of the final filtered image to assessment of each of the singular image components of the SVD data matrix (Equation 7) which can be visualized in Figure 13. The obtained AUC values, one for each singular component, will be used as the training target value for the deep neural network proposed here. Just like ROC analysis for the final clutter filtered image, higher *component AUC* values indicate singular image components which possess signals that are more spatially constrained to the defined flow regions. We then propose to label the components with the highest AUC values as the flow components and use these components to form the final filtered flow images.

Figure 23 accomplishes this using a simple threshold cutoff[9]. We note that this threshold can be varied depending on application by the operator. Should, greater flow detection be desired at the cost of increased clutter the threshold can be dropped. If strong clutter rejection is required, the threshold can be increased. An example of the power Doppler of the femoral bifurcation created using this method is presented in Figure 25 a). Unlike

[9]For scenarios where flow flow is weak, such as during diastole, the AUC curve can be somewhat inconsistent and poorly formed. Therefore in addition to the threshold cutoff we will also require that the AUC curve possesses a concave-down structure. This was done by requiring that the actual cutoff be at least some $factor \times$ AUC of the first component.

other SVD clutter filtering techniques, filter sensitivity is tunable on a well defined metric. Furthermore, the defined AUC threshold may identify singular components that are not contiguous (Figure 23) leading to a framework that is better able to handle more complex mixed clutter and flow signals.

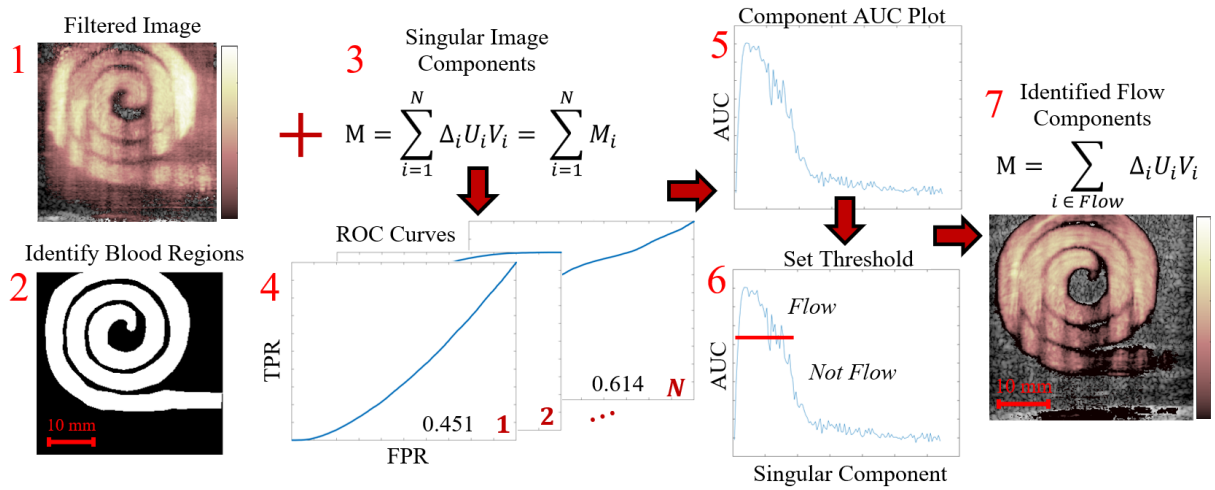

Figure 23: Determining flow components using the component AUC plot. A *1*) filtered[10] flow image and b-mode image (not shown here) are used to segment the *2*) flow region of the image. The *3*) N singular image components are computed and the *4*) ROC curves for **each** of the N singular image components is calculated. The areas under these ROC curves are then calculated and plotted. From the *5*) component AUC plot, the *6*) highest AUC components can be identified, after setting an AUC threshold and used to produce the *7*) optimal Doppler flow image for a given AUC threshold.

The effect of varying the AUC threshold is shown in Figure 24 where components within some percentage of the component with the highest AUC value have been labeled as flow and used to form the power Doppler image. If only a few components are used to form the power Doppler image, some patchiness is observed in the flow signal, despite little clutter being observed. As the AUC threshold is decreased, the flow signal grows in intensity (as does the clutter signal). If the AUC threshold is set too low, high magnitude clutter signals become included in the power Doppler image drowning out the flow signal.

[10] This reference image is needed to aid in segmentation. It does not have to be the best quality. The image in step 1 was obtained used a high-pass filter.

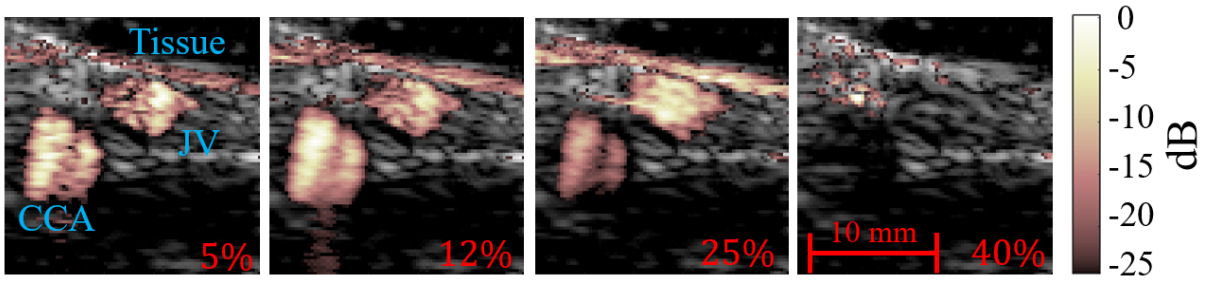

Figure 24: Demonstration of varying the threshold of acceptance for an *in-vivo* scenario showing the common carotid artery (CCA) and the jugular vein (JV). Power Doppler images were normalized relative to the maximum pixel intensity to better show the relative changes in flow and clutter pixel intensities. The threshold of acceptance (red values) is defined to include all singular components with an AUC that is within $x\%$ of the maximum measured component AUC. As the threshold is dropped, a more complete signal and less patchy signal is obtained up to the point where tissue components dominate the image.

One potential weakness of using ROC analysis to evaluate clutter filter performance and rank SVD components in terms of flow and clutter contents is the reliance on manual segmentation. However, as we have previously mentioned automatic segmentation algorithms from the literature are unlikely to perform well either due to changing flow statistics across the cardiac cycle. For instance it is much more difficult to detect flow during diastole, and indeed in some acquisitions it appears that no flow is present within certain parts of the artery vein. However, under ideal circumstances flow should be detected within these regions and they should be segmented as such. This operation is generally difficult due to the different flow intensities within what should be uniform flow regions. For example in Figure 13 poor flow detection is encountered at the branching of the femoral as the beam to flow angle approaches 90° while upstream flow detection is decent. In these cases *a priori* knowledge of the vessel structure allows a manual operator to segment flow regions to a high degree of accuracy. Figure 25 shows the effects on the component AUC curve for slightly different manual segmentation attempts. The effect of slight inaccuracies in the segmentation is minimal to large vessels as they represent changes in only small fractions of the flow and pixel populations.

Finally we note that the ROC curve and AUC are calculable and defined/normalized such that the absolute numbers of tissue and flow pixels will not affect the generation of the ROC curve, only how well/poorly the classifier does detect the true/false positives. Therefore, a block with only a few flow pixels may generate a similar AUC value as a block with a majority of flow pixels, despite both blocks containing very different signal statistics. To address this potential mismatch between signal statistics and AUC value we will only compute the AUC value for blocks with at least 10% flow pixels. For blocks with 10% or fewer flow pixels, we will set the AUC of all components to 0.5, or equivalent to a random classifier.

This choice satisfies two main goals. First the network is not being tasked with the problem of identifying clutter as well as identifying flow (an AUC of 0 indicates a perfect anti-classifier). Secondly the default AUC of tissue clutter is 0.5 NOT 0. Consider an improperly drawn ROI, indicating the presence of flow when there is none present in the

entire image. If the ROI is randomly drawn we would expect the singular image component of each singular component to more or less randomly include high and low intensity pixels inside and outside the ROI. Calculation of the ROC curve of each singular component would therefore yield values close to 0.5.

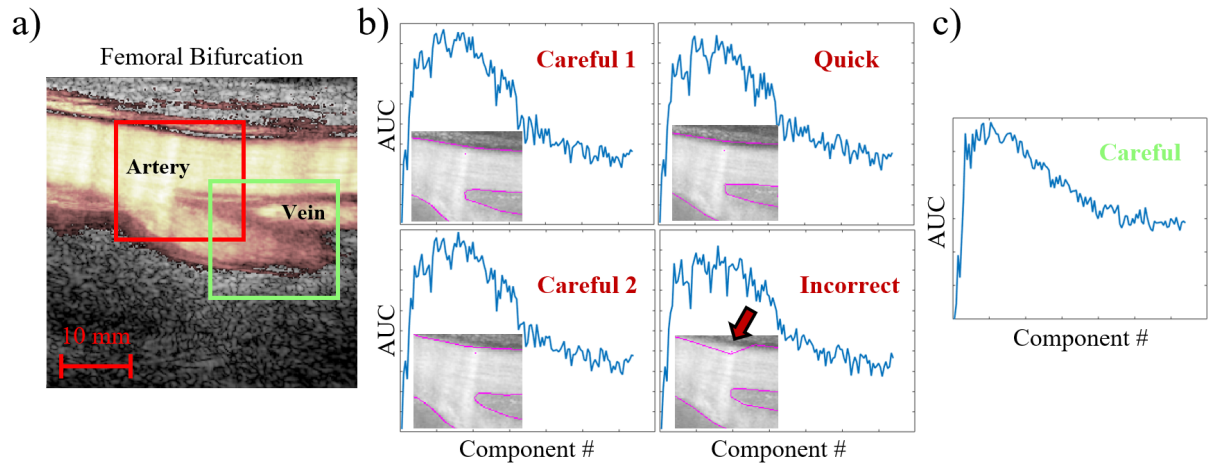

Figure 25: *a*) Power Doppler of the femoral bifurcation during systole created using the component AUC as described in Figure 23. The femoral vein can also be observed. *b*) The component AUC curves of the region in the red box (blockwise SVD is employed here) for 4 segmentation attempts: two careful attempts, one quick segmentation, and one attempt with an intentional error shown inset in grayscale (i.e. before power thresholding). The component AUC curves are very similar between the two separate careful attempts and only minor difference are observed with respect to the quick segmentation. The incorrect segmentation attempt is noticeably difference from the other but does maintain the same general structure of the AUC curve compared to the AUC curve for the green block in *c*). In general, provided care is taken in segmentation, the component AUC curve is quite reproducible.

Calculation of the component AUC curve is dependent on knowledge of the segmented flow region, something not available to the algorithm which is presented only with statistics extracted from the data matrices of the decomposition. Therefore, the goal of this framework will be approximate the component AUC curve using derived estimators from the SVD data matrices, and use the approximated component AUC curve to identify flow and non-flow components at the desired level of sensitivity. In the next section we will describe the estimators derived from the U, Δ, and V SVD data matrices and how they can be used to estimate the AUC.

Investigation of the SVD Data Matrices for Network Input

The Δ Matrix: Magnitude Information

Unlike the spatial or temporal matrices, the Δ matrix only contains information along its diagonal (N entries in total where N is the ensemble size) – the singular values, which are

arranged in order from largest to smallest representing the magnitude of each component of the decomposition. Equivalent to a size N vector, there is little to derive from the singular value curve. Furthermore, we note that the components are ordered from largest to smallest producing a constantly decreasing relatively smooth curve.

However, magnitude information is useful as a coarse method of classification as the diagonal of the Δ matrix represents all available magnitude information (as the U and V matrices are orthonormal). We will therefore use the singular magnitude information as a rough input to the network (with N elements, one for each singular component) to be tempered with additional information. Figure 26 shows how the singular value curve can be used as a very rough proxy for the flow subspace in flow and non-flow scenarios.

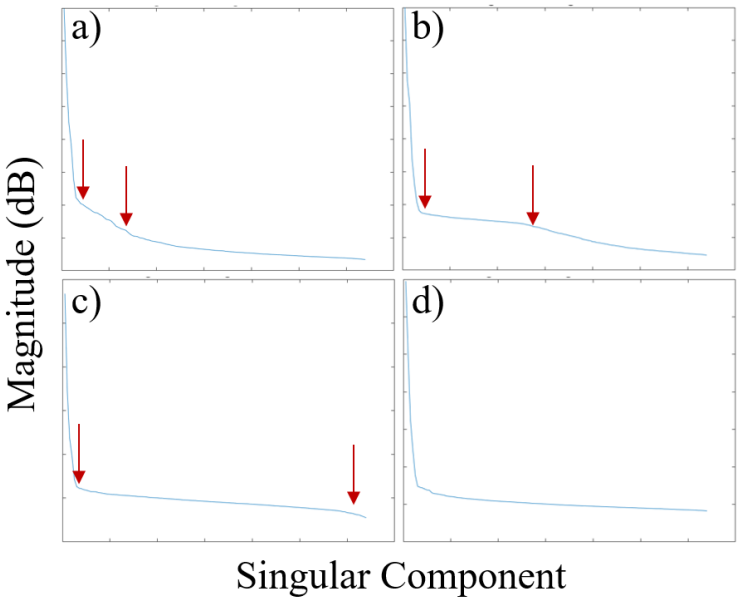

Figure 26: Singular magnitude curves for various imaging scenarios. *a-c*) (femoral bifurcation and *in-vitro* flow phantom) contain blood while *d*) (fleshy part of forearm) contains solely tissue. The red arrows show the estimated upper and lower threshold locations. We note that the turning point (first red arrow) appears present in every scenario and that it can be difficult to distinguish the fourth figure, with no blood signal, from the other three scenarios, particularly scenario *c*). If the relative amount of flow signal in the data matrix is low it becomes significantly more difficult to distinguish the flow threshold in the singular value curve.

The V Matrix: Temporal Information

A useful statistical estimator of the temporal characteristics of the data matrix is the lag-one autocorrelation as described in the previous chapter [44]. As discussed previously, there are many well-established methods of identifying threshold cutoffs from the frequency vs. singular component plot; we will use the mean Doppler frequency[11] of each singular component

[11]Similarly to literature algorithms we will actually use the absolute value of the mean Doppler frequency to measure flow speed rather than distinguishing between flow towards or away from the transducer (in

as one of our input variables.

However, one of the major issues encountered in our investigation of the temporal information is that the mean Doppler frequency curve frequently contains significant noise, often to the point that the general structure of the Doppler frequency curve no longer displays the general tendency to increase *relatively* smoothly with singular component index. This can be seen in Figure 27 where the expected frequency curve archetype as described in Ref [50] is only observed in panel a). Panes $b-d$) contain the frequency curves taken from flow blocks of the carotid cross section (Figure 24), femoral bifurcation (Figure 25), and the aliased spiral flow phantom[53] (Figure 12).

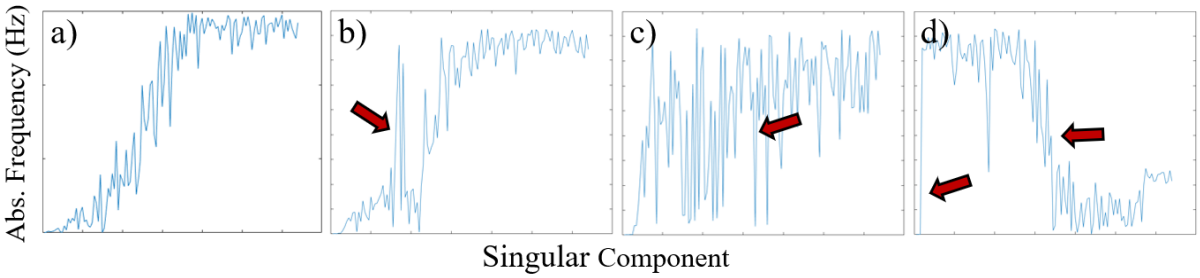

Figure 27: Mean Doppler frequency of each singular component ($N = 128$) for various imaging scenarios. Under many imaging scenarios the characteristic curve structure is not maintained. a) Femoral bifurcation. The expected ideal Doppler frequency curve as encountered *in-vivo*. As expected, the frequency increases with singular component number. b) The carotid cross section (containing the jugular vein). A large peak is encountered in the clutter components. These components could contain significant flow and should be selectively retained. c) A different region/block of same *in-vivo* femoral acquisition as a). The noise plateau is ill defined as the flow signal contains forward and backward flow and there is significant mixing of flow and clutter signals within individual components. d) Spiral flow phantom with aliasing; the aliased components *wrap* around the Nyquist limit. Furthermore it appears the there are only a few $(1-2)$ clutter components.

However chaotic the mean Doppler frequency appears to be, it does contain important information about the interaction of the flow and clutter signals and so we deem it a useful parameter for our proposed framework.

From the temporal data matrix we present another discriminating factor – the magnitude of the zero frequency component of the smoothed power spectral density of each singular component. We observe that for larger and larger singular component indices the smoothed power spectral density, initially with a peak at 0 Hz slowly becomes depleted as low frequencies are removed (a consequence of increasing the mean Doppler frequency) leaving a curious double-peaked feature. Figure 28 shows the power spectrograms of each singular component and the corresponding magnitude of the 0 Hz frequency component . We note that the zero frequency component generally contains less noise than Doppler frequency magnitude but can still be noisy. Across our investigation of many *in-vitro* or *in-vivo* imaging scenarios

either case blood flow is blood flow).

we find that the zero frequency magnitude is very consistent in its general structure and generally distinguishes between tissue clutter and flow fairly well. We will therefore use the magnitude of the zero frequency component as one of our input discriminating factors.

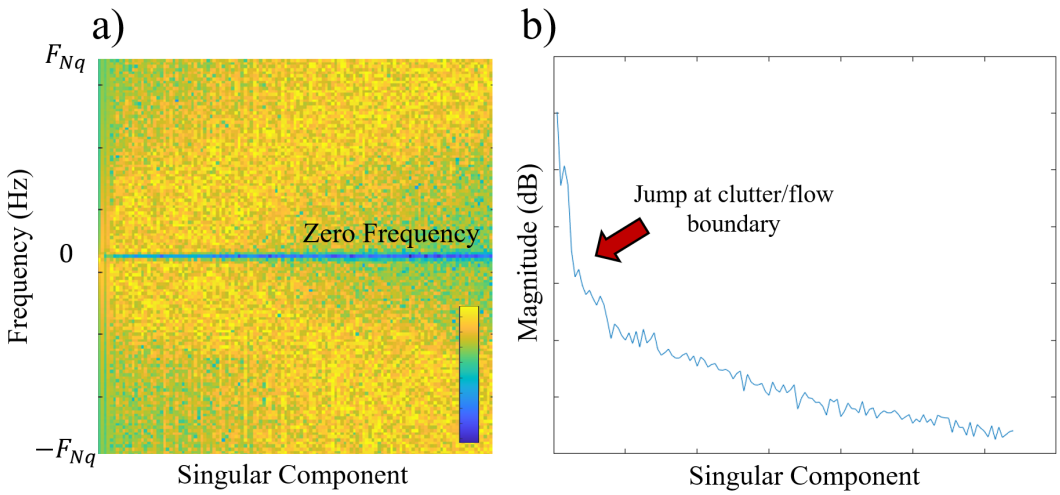

Figure 28: *a*) The power spectral density of each singular component. As the singular component index increases, we observe that the contents of each component shifts towards higher frequencies (absolute value) leading to the depletion of the zero frequency mode. *b*) A plot of the magnitude of the zero frequency component. A large jump is always observed between the tissue and flow subspace.

The U Matrix: Spatial Information

The spatial matrix is most commonly visualized through the spatial similarity matrix (Equation 8) which, as previously discussed, shows how well a given singular component is spatially correlated with another singular component. To avoid computation of the entire SSM, and to reduce the amount of input data to our model, we propose to reduce the SSM down to two quantities as shown in Figure 29.

From the spatial similarity matrix, we can extract the bottom row (or first column as the matrix is symmetric along the anti-diagonal). This quantity has been used before [30][31] but is relatively uncommon in the literature. This gives the spatial correlation of a given singular component with the first singular component or *first correlation* a quantity which should display a sharp boundary between tissue and blood components as the tissue and blood signals are only weakly spatially correlated.

We can also extract the one-off diagonal (or rather as shown in Figure 29, the one-off anti or sub-diagonal) in an attempt to measure how well a singular component is correlated with its neighbouring components. From this vector, we can compute a quantity we will call the *neighbour correlation*, which is given by the average correlation that particular singular component has with the next higher and next lower singular components. We note that the first and last of the N singular components will lack a neighbouring component. In this case we will just set the neighbour correlation to the correlation that the first or last

components have with their only neighbour (this is ultimately unimportant as the first and last components will not contain any flow signal).

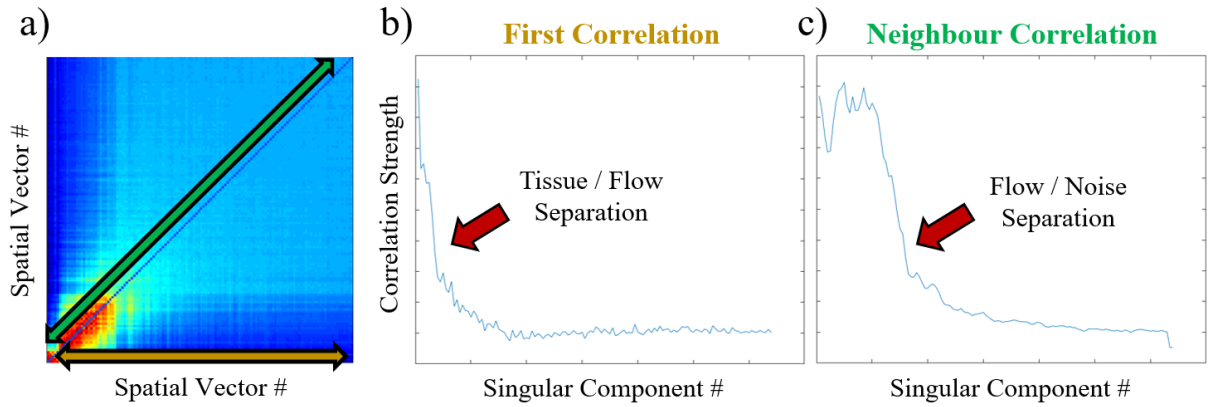

Figure 29: *a*) The spatial similarity matrix showing the first (first horizontal line shown in gold) and neighbour (1-off sub-diagonal shown in green) spatial correlations. A typical plot of these quantities is depicted in panes *b*) and *c*) showing how these quantities highlight the separation of the tissue and flow signals and flow and noise signals respectively.

While the spatial similarity matrix describes the correlations *between* singular components it does contain any information about the spatial correlations between the pixels within a singular component. For instance the speckle patterns characteristic of ultrasound, found in the first few singular components (which tend to generally resemble the b-mode image), while random, tend to possess some structure and length scale and are therefore very different from the Gaussian-distributed noise found in the higher singular image components were neighboring pixels are uncorrelated.

By simply computing the standard deviation of pixels inside a sliding window across each component image we observe that pixels in the blood components tend to be quite correlated and possess a high standard deviation while noise and tissue components are relatively uncorrelated. This is shown in Figure 30 in which the blood components are clearly distinct from noise components. The first few tissue components also tend to be locally spatially correlated, however, flow is never found in the first components and these components, with the aid of the other discriminating factors, should be safely discarded.

We will call this discriminating parameter the *local spatial correlation* as it refers to local pixel distributions. For instance, in Figure 30, blockwise SVD was performed with blocks of size 80×80 pixels; each singular image component is 80×80 pixels and the sliding window used in each singular image component was 20×20 pixels. Therefore, pixel spatial distributions are computed over a much smaller window (approximately a 3.2×3.2 mm region[12]), with the standard deviations across all sub-regions within a block being averaged together to produce a final aggregate value.

[12]These values were found to produce good consistent results across the imaging scenarios investigated.

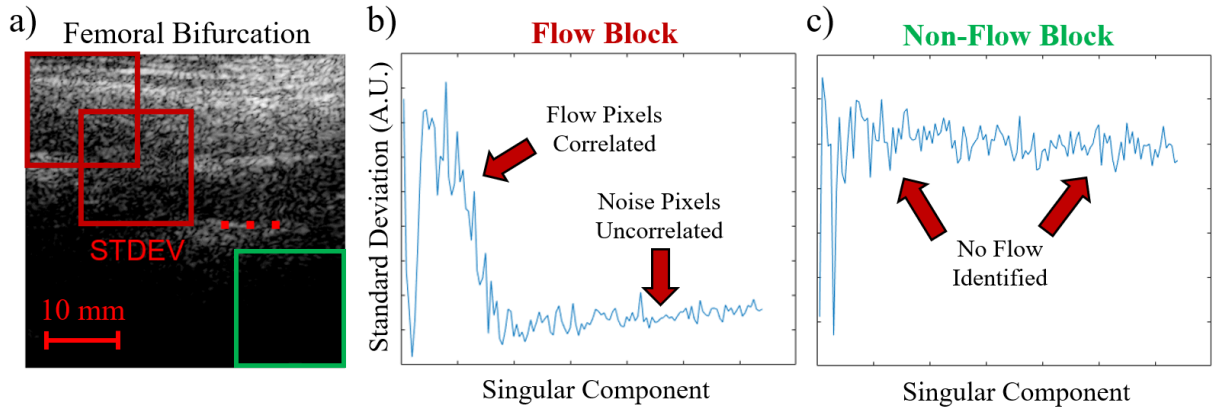

Figure 30: *a*) The femoral bifurcation. Blockwise SVD is performed with red blocks containing flow signal and green blocks containing no flow signal (Figure 13 shows the segmented flow region for this image). The average standard deviation of the pixels within a sliding window is computed for each of the N singular image components. *b*) and *c*) Plots of the standard deviation of pixels within each singular image component for a typical flow and non-flow block. Components with significant flow signal appear as highly correlated (high standard deviation) while components whose pixels contain just noise are more weakly correlated. The standard deviations presented in panes *b*) and *c*) have been min-max normalized; it is the relative change in standard deviation that we highlight.

While computing the spatial correlations tends to be computationally intensive due to the large size of the spatial data matrix, this inter-component spatial information provides significant insight into the type of signal present in each singular component and cannot be overlooked, especially due to its directly visually observable nature in the singular image components. Therefore, the local spatial correlation will be a useful discriminating factor for the proposed deep network model.

Data Acquisition and Preprocessing

As described in the previous section, our model attempts to reproduce the area under the ROC curve using 6 statistics as inputs. With an ensemble size of N, there will be $6N$ inputs and N approximated AUC values. In this thesis work, we have chosen an ensemble size of 128, long enough to obtain a good number of flow samples but short enough to maintain temporal resolution at a PRF of 3000 Hz or 3333 Hz. Therefore, our network will possess $128 \times 6 = 768$ total inputs and estimate 128 AUC values. The inputs and output of the proposed framework are summarized in Figure 1.

Table 1: Model Input and Output Parameters

Inputs	Details
Singular value magnitude	Magnitude of each component of the decomposition, diagonal of the Δ data matrix, log-scaled.
Mean Doppler frequency	Mean Doppler frequency of each component of the decomposition, lag-1 autocorrelation of each column of the V matrix.
Zero frequency magnitude	Magnitude of the 0 Hz bin for each singular component after taking the Fourier transform of each column of V.
First correlation	Spatial correlation between a given singular component and the first singular component, bottom row of the SSM.
Neighbour correlation	Spatial correlations between a given singular component i and its adjacent singular components $i\pm1$, 1-offset (sub) diagonal of the SSM.
Local spatial correlation	Standard deviation of pixel intensities computed across a 20×20 pixel sliding window on each singular image component.
Output	**Details**
Area under the Curve (AUC)	Area under the ROC curve computed for each singular component using a hand-segmented flow region. ROC curve is generated by computing the number of true positive flow pixels and false positive tissue pixels sweeping the intensity threshold from the lowest to highest pixel values.

Data Acquisition

Ultrasound scans were acquired using two scanners with similar acquisition settings. A SonixTouch research system (SonixTouch; Analogic Ultrasound; Peabody, MA, USA) was used for *in-vitro* acquisitions of ultrasound flow phantoms, while a US4US (USPlatform; us4us Ltd., Warsaw, Poland) scanner was used for all *in-vivo* acquisitions. The imaging acquisition parameters are shown in Table 2.

Table 2: Data Acquisition Parameters

Parameter	Details	
Ultrasound Scanner	SonixTouch	US4US
Ultrasound Probe	L14-5	SL1543
RF Data Range and Resolution	-2048 – 2047, 12 bit	-2048 – 2047, 12 bit
Number of Tx/Rx Channels	128	192
Array Pitch (mm)	0.3048	0.245
Transmit Frequency (MHz)	6	5
Transmit Angles[1]	-10, 0, 10	-10, 10
# Pulse Cycles	3	3
Sampling Rate (MHz)	40	25
Imaging Depth (mm)	60	47
Pulse Repetition Frequency (Hz)[2]	3333	3000

[1] Although acquisitions contained multiple angles, in this work **only** the 10° angle was used for processing as this resulted in a generally better beam to flow angle.

[2] Here the PRF refers to the frequency at which each individual angle is sampled. The total PRF for all angles is 10 MHz for the SonixTouch and 6 MHz for the US4US.

We note that although the acquisition parameters are slightly different between systems, this is unlikely to influence our framework, as our network inputs are higher order parameters and the framework is not exposed to the raw RF data or even the beamformed images.

Relatively generic beamforming parameters were chosen for both the SonixTouch and US4US systems, differing in only the prefilter passband. Blockwise SVD was performed identically on all acquisitions. Blocks of size 80×80 pixels with an overlap of 5 pixels were chosen as this combination tended to not produce the linear artifacts shown in Figure 14 while still being reasonably computationally cheap. The leftmost and bottom pixels were slightly cropped; this is not expected to have a significant effect on the results as image quality tends to be degraded near the edges of the imaging view due to reduced transmit power. Table 3 summarizes the beamforming and blockwise SVD filtering parameters.

The data acquisition scenarios are presented in the next section, along with their placement into the training, validation, or unseen datasets.

Table 3: Beamforming and SVD Processing Parameters

Beamforming Parameter	Details	
Ultrasound Scanner	SonixTouch	US4US
Prefilter Passband	4 – 8 MHz	3 – 7 MHz
Filter Design and Order	Equiripple (30^{th} Order)	
Apodization	Rectangular	
F-Number	1.25	
Assumed Speed of Sound (m/s)	1540	
Beamformed Image Size (mm)	4 mm × 4 mm	
Beamformed Image Size (pixels)	256 × 256	
SVD Parameter	**Details**	
Ensemble Size	128	
Window Size	16	
Block Size (pixels)	80 × 80	
Block Overlap (pixels)	5	
# Blocks per Frame	1156	
Final Image Size (pixels)	245 × 245[1]	

[1] Some of the image along the right side and bottom is cropped off during block-wise SVD.

Training, Validation, and Testing Datasets

To ensure robustness of the trained model, a large amount of flow data was acquired under various flow scenarios[13]. In total 30 *in-vitro* and *in-vivo* acquisitions were acquired, divided into groups of 17, 9 and 4 acquisitions for the training, validation, and unseen datasets respectively. Acquisitions were performed over the entire cardiac cycle (or a cardiac flow profile was used, in the case of the flow phantoms) with each acquisition taking at least 1.5s. Only every 5^{th} frame was used for training as flow signals generally did not appreciable change between frames, removing the highly correlated flow signals in adjacent frames. Furthermore, frames that could not be segmented, due to significant clutter or difficult to observe flow, were not included in the training data. The focus of this work was on *in-vivo* acquisitions, as *in-vitro* flow phantoms, with little movement, a consistent speed of sound, homogeneous tissue mediums and attenuation, are not particularly challenging sample cases. A few phantom acquisitions were included however, to provide a more diverse training and testing dataset. A summary of the acquisitions used for training is presented in Table 4.

[13]Some acquired as part of this work and some reused from previous experiments in this lab [73][74].

Table 4: Training and Testing Flow Acquisitions

Label	# Frames[1,2]	Type	Comments
F01	46	Training	Femoral bifurcation
F09_2	80	Training	Femoral bifurcation
F11	39	Training	Femoral bifurcation
F12	39	Training	Femoral bifurcation
SP_1	73	Training	Spiral flow phantom [53], carotid flow profile
HT_03[3]	73	Training	Helical toroidal phantom [75], carotid flow
HT_04[3]	73	Training	Helical toroidal phantom [75], carotid flow
J3_B	80	Training	Brachial artery
H5_B	43	Training	Brachial artery
HF_10	61	Training	Carotid bifurcation
HF_7	80	Training	Carotid artery, complex flow
CB_1	80	Training	Carotid artery, significant clutter
CC_2	76	Training	Carotid artery, clutter present
KC_2	80	Training	Carotid artery, strong wall reflection
RC_1	56	Training	Carotid artery, strong wall reflection
WC_1	57	Training	Carotid artery
RE_1	80	Training	Femoral Bifurcation
F07	45	Validation	Femoral bifurcation
HT_05[3]	73	Validation	Helical toroidal phantom [75], carotid flow
F04	80	Validation	Femoral bifurcation + Femoral vein
D2_L	63	Validation	Carotid artery
H1_L	41	Validation	Carotid artery, significant clutter present
KC_1	82	Validation	Carotid artery
WB_1	80	Validation	Brachial artery, strong flow signal
JF_1	53	Validation	Femoral Bifurcation, little flow in upper branch
YE_1	81	Validation	Femoral Bifurcation, clutter during systole
CF	270	Unseen	Bifurcation Phantom [76], carotid flow profile
D5B	340	Unseen	Brachial artery
F03	350	Unseen	Femoral Bifurcation, Little flow in upper branch
HB	340	Unseen	Femoral Bifurcation

[1] Every 5^{th} frame was processed. Total refers to the number of included frames NOT acquistion length

[2] Frames in which segmentation could not be clearly performed, were discarded.

[3] This flow phantom has a 3D structure. Acquisitions 03, 04, and 05 were taken along different 2D planes.

After acquisition, segmentation was performed on every 5^{th} frame. The ROC curve and subsequent AUC value was calculated for each of the 128 singular components in each of

the 1156 blocks in each frame. The statistical discriminating factors, as given in Table 1 were likewise computed for each of the 128 singular components in each block. Each frame, consisting of 1156 highly overlapping blocks, was then randomly sampled at 10% to reduce the amount of training data by removing (many) adjacent blocks with highly similar flow characteristics.

In total, 1113 distinct flow frames were used for training, each containing 115 randomly sampled blocks giving a total 129,995 training samples. The validation set consisted of 525 frames and 60,375 validation samples. While many machine learning frameworks do not use completely distinct training and validation datasets; choosing to aggregate all samples not part of the unseen dataset, preprocess and normalize, and then subdivide into training and validation datasets[77][78] we found that this method did not perform well in our case. Using the same subdivided dataset we found our framework produced similar loss curves for the training and validation sets. However, when the framework was applied to the unseen acquisitions, results were poor. Changes to the network structure continued to produce networks that appeared to perform well on the training and validation datasets but poorly on the unseen data. We expect that this is occurring because the input statistics likely differ significantly between data acquisitions necessitating the use of 1) large diverse training datasets that contain different flow profiles 2) a validation set that is distinct[14] from the training set so that the generalizable performance of the network on other unseen data can be assessed.

For this reason we have elected to use completely distinct acquisitions, although the same subject could be used (acquisitions taken at different positions/angles), for our training and validation datasets allowing us to estimate how the network would perform on other unseen data. Due to the large variation in flow profiles and signal statistics, a relatively large (and distinct) validation set was used relative to the training set.

Preprocessing and Normalization

The input parameters (Table 1) were first min-max normalized [79] (relative to the minimum and maximum in each individual 128 element block input) to equalize feature size and aid in training [80][81]. The physical motivation of using a *self-consistent* min-max normalization is that the inputs to the network will differ only in structure, not value (the plots of all inputs are scaled from 0 to 1 but maintain the same structure). If inputs were scaled relative to some arbitrary value (say correlations scaled to between -1 and 1) then the differences of inputs between samples would become relevant when they are generally not. For instance, observing the SSM, it is the relative correlations between components that delineates potential subspace boundaries, not the absolute values of the correlations.

[14] Due to overlap in space, blocks in the training and validations sets may have similar statistics. To some degree (i.e. 5 frames of separation) blocks can overlap in time as well. This violates the assumption that training samples are independent from one another.

Network Structure and Computational Details

The proposed model was implemented in Python (Ver. 3.9), using Tensorflow with GPU support (Ver. 2.6) using the Keras API frontend. It was found that a 3 layer deep neural network with 8,000 neurons per layer performed best (lowest loss values) with ReLU activation for every hidden layer, except the output layer which used a linear activation function. Additional layers did not appear to improve performance but significantly increased computational and memory requirements. An outline of the proposed model is shown in Figure 31.

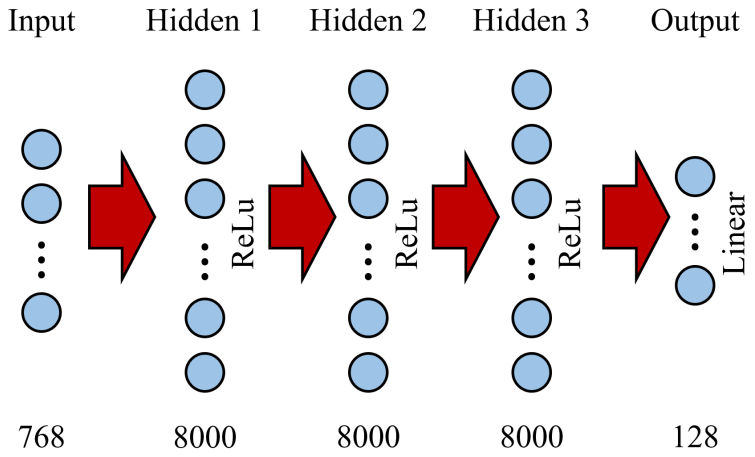

Figure 31: An overview of the proposed model. The numbers below each layer refer to the number of neurons in each layer. The connections between neurons are shown using a large red arrow for simplification.

To improve network generalizability it was proposed to use the Tensorflow *Dropout*(x) command, that randomly sets the weights in the layer to 0 at a rate of x. By forcing the network to make correct inferences using only a fraction of the possible weights, overfitting can be avoided. However, it was observed that networks trained using *Dropout*() were less able to accurately reproduce the sometimes unique structures of the AUC curves (Figures 34 - 36).

Other network structures were also explored. In summary, we suspect that the rather significant amount of noise in the discriminating factors and output AUC limit the final accuracy of the network as well as the rather simplistic model with its limited number of inputs[15]. For instance, we observed that a three layer network with 8,000 neurons in each layer was only slightly better than a single layer 8,000 neuron network.

We would also like to remark here that other classification models were tested using the same training and validation datasets such as a linear regression model and a random forest classifier. However, these methods did not perform anywhere near as well as the DNN.

[15]This is not a negative point – it means that a hugely complex framework not need be deployed.

All other calculations in this work such as computing the SVD of data matrices, discriminating parameters, Doppler calculations, etc. were performed using MATLAB (ver. 2021b; MathWorks, Natick, MA, USA).

Training and Validation

The network was trained on an RTX 3060 Mobile GPU (Nvidia, Santa Clara, CA, USA). The weights of each layer were initialized using the default Keras initialization (Glorot uniform initialization[82]) and trained using the Adam optimization algorithm[83] using the mean squared error (MSE) as the loss. Training was performed using a learning rate of 10^{-5} and a batch size of 32, running for 55 epochs on the training set. At the end of each epoch, the loss was computed on the validation set as well[16]. The training and validation losses are shown in Figure 32.

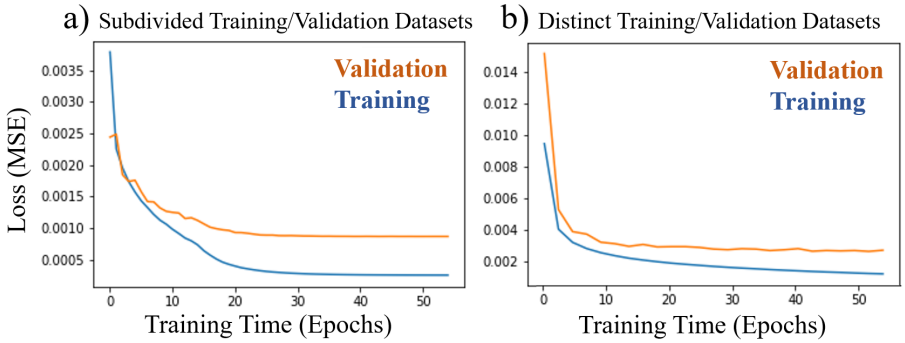

Figure 32: Loss (MSE) curves for the training and validation sets. *a*) The training and validation sets from Table 4 were combined and subdivided (using the *traintestsplit* function in Keras). As there is significant correlation between samples, the training and validation losses were quite comparable (given the scale of the plot). *b*) The network was trained using only the training acquisitions from Table 4 with the validation acquisitions held separate from the training data. In this scenario, the difference in MSE between the two datasets was much larger (given the scale of the plots) and the absolute MSE of the validation set significantly higher. It was found that changing the network structure would result in a small change in the MSE curves in pane *a*) but a large change in pane *b*); the change in pane *b*) was more indicative of final performance on the unseen acquisitions.

Figure 32 shows two loss curves, in the first the standard dataset splitting for the training and validation datasets was performed (i.e. all training and validation data from Table 4 was combined and split into 80% and 20% training and validation sets – non-isolated training and validation sets). In this case we observe that the loss curves of the training and validation data follow one another closely (looking at the absolute difference in MSE). In the other pane, isolated training and validation sets are used as described above. We found that

[16]As the validation set consisted of unseen data from completely distinct acquisitions, it tended to stop showing improvements in the loss, although we never observed significant regression.

due to the overlapping nature of our samples, subdividing a training set and attempting to optimize network structure based on a non-isolated validation set did not correlate well to performance on unseen samples. For this reason, we tracked the loss in the isolated validation set and optimized network design for a model that produced an isolated validation loss that **1)** never regressed (started increasing after some point) and **2)** produced the lowest absolute MSE after training.

Performance Metrics

Image Quality and Filtering Performance

The validation of any model involves testing the results of said model against a set of data that is known to be true – the *ground truth*. Unfortunately we do not possess the ground truth blood signal against which our clutter filtering algorithms can be bench-marked. This results in a clutter filter whose performance is very difficult to determine quantitatively. In the literature, clutter filter performance is most generally quantified using the contrast ratio (CR) and contrast to noise ratio (CNR) on reference blood and tissue/background patches [84][85][86]. This procedure requires us to identify which region(s) of the image contain blood and tissue pixels[17] and then perform the following calculations.

$$\begin{aligned} CR^1 &= \frac{\mu_b}{\mu_t} \\ CNR^1 &= \frac{|\mu_b - \mu_t|}{\sqrt{\sigma_b^2 + \sigma_t^2}} \end{aligned} \tag{16}$$

[1]As signal power is displayed on a logarithmic scale, these quantities are represented as the difference in [dB] between the flow and clutter pixels.

Here μ_b and μ_t refer to the mean signal power of blood and tissue pixels respectively and σ_b and σ_t to the variance. Higher CR and CNR values are indicative of filters with a greater ability to separate flow and clutter signals.

Figure 33 shows how the CR is calculated for two prospective filtered images. It is important to select relevant tissue and flow ROIs as the subsequent quantification of clutter filtering performance using these metrics is only true for the selected regions – and this numerical quantization is then applied to the entire frame as a whole. For instance in Figure 33 the wall artifacts obtained using method *b*) do not impact the calculation of the contrast ratio. Nevertheless, despite these caveats, we will use the CR as it is important metrics to assess filter performance. Due to the discrete nature of the SVD and potential use of further spatial filtering on the power and color Doppler images, the CNR was found to be of less utility.

[17]A blood and a tissue reference patch are used.

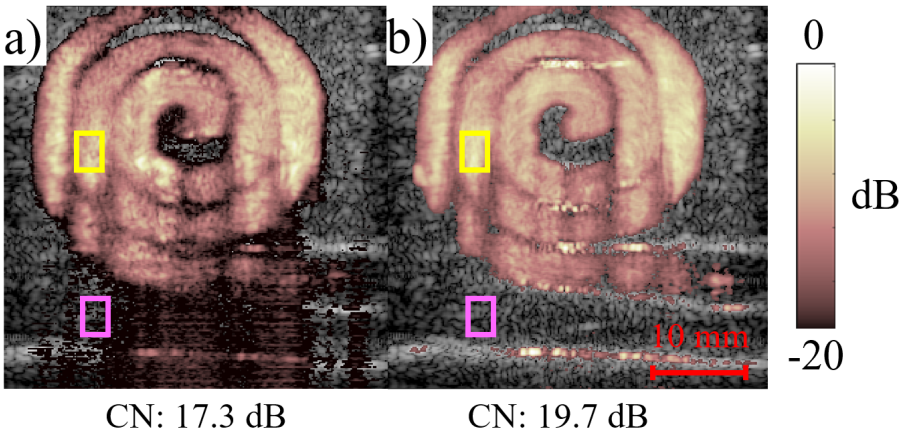

Figure 33: The calculated contrast ratio for the power Doppler image produced using two sample filtering techniques for the marked flow (yellow) and tissue ROI (pink) on a spiral flow phantom under carotid flow.

The CR and CNR are frequently used as they are easy to calculate and understand. However, they do not quantify how well the clutter filter performs in terms of sensitivity and specificity across various imaging scenarios as they are just a measure of pixel brightness and local consistency. While they will form a part of our analysis, we will additionally use the area under the ROC curve as described previously and found in Ref [68]. While the AUC does not possess a numerical interval scale, it will nevertheless allow us to identify the best and worst performing filters. ROC analysis is performed using the entire segmented flow region, giving this metric greater validity over the entire flow region,

One final issue with the contrast and contrast-to-noise ratios is that they are presented on an absolute scale in decibels. However, different clutter filtering techniques may stretch or scale the inherent signal data. For instance, consider what would happen if during the filtering operation[18] that was used to produce Figure 33 the signal intensity of all values was halved (halved in terms of the log-scaled values, i.e. a pixel with an intensity of 80 dB is now a pixel with an intensity of 40 dB). If the power Doppler of this filtered signal was compared with the original image and presented on the same scale it would appear significantly worse than the original unscaled image (with half the contrast ratio for instance) despite both images possessing the same diagnostic ability (and identical appearance if the dynamic range for the new power Doppler image was halved). Despite scaled values, both power Doppler images have the same pixel intensity distributions between the clutter and flow signals. In reality, it is not the amount of separation (in dB) of the flow and clutter statistics that creates a better filtered image, but the reduction in the overlap of the two distributions.

In practise, this effect is relatively minimal when comparing the performance of different SVD filtering techniques. However, when comparing to power Doppler images created through other filtering means such as through high-pass filtering, the ideal intensity scales of the SVD and HPF filtered images are very different and it is NOT appropriate to display or compute contrast values without adjustment. We propose the following method to determine

[18] We are abstracting this operation to that of a *black box*.

the most ideal/fair dynamic range of a power Doppler image using the ROC curve[19]

1. Segment flow region.
2. Create ROC curve.
3. Identify the point on the ROC curve closest to the top left corner and record the true positive rate X corresponding to that point.
4. Identify the intensity value of the X^{th} percentile flow pixel. This gives the bottom of the dynamic range relative to the max intensity flow pixel.
5. Dynamic Range: $\left[X^{th} \text{ flow pixel}, \text{max pixel value}\right]$

We will adopt this method to determine an appropriate dynamic range to compare results obtained through frequency filtering. However, as the singular value decomposition is unique for a given data matrix and the same data matrices are being used for each SVD filtering algorithm there is no inherent 'stretching' or 'shrinking' of the data scale occurring between the tested SVD filtering algorithms. Therefore we will present all SVD results using the same dynamic range.

However, we note that the most ideal dynamic range also possesses several caveats. **1)** It is different for each individual frame and thus unsuitable to quantify filtering performance across a cineloop. **2)** It is a range relative to the maximum pixel intensity, not a preset threshold and will not saturate. **3)** It requires flow segmentation. **4)** Most importantly, clinical scanners will not perform this analysis and will use a user-set dynamic range over the cineloop (or at least the part that is being examined).

Ultimately there is no perfect metric to absolutely assess image quality and filtering performance; every proposed method possesses some relative advantages and disadvantages.

Cineloop Consistency

While the consistently improved performance of SVD filters relative to frequency filters has frequently been demonstrated for still images, there has been little focus on the demonstration of consistent frame by frame performance. As the end user of any clutter filtering technology will be viewing multiple filtered frames in quick succession, it is important not only for the individual filtered frames to be consistently of high quality, but for individual frames to be consistent with one another. If this condition of consistency does not hold, flashing and strobing artifacts will be observed in the cineloop.

As SVD filters are *discrete* in nature, ultimately consisting of some number of singular components they are prone to produce flow images, that even if extremely good, may still possess significant variations between each other as signal features evolving in time are captured/mixed in different singular components which may or may not be used for adjacent

[19]This has the unfortunate effect of only being possible if the flow region is segmentable.

frames in the cineloop to form the final flow image. We will therefore have to track and make sure that variations between frames are minimized. This can be done using the Pearson correlation coefficient [87].

$$r = \frac{\sum_i (x_{1i} - \bar{x_1})(x_{2i} - \bar{x_2})}{\sqrt{\sum_i (x_{1i} - \bar{x_1})^2}\sqrt{\sum_i (x_{2i} - \bar{x_2})^2}} \tag{17}$$

Here r is the Pearson correlation coefficient between frames 1 and 2 where x_i refers to the intensity of the i^{th} pixel in frames 1 and 2 [88]. While the blood flow will evolve in time, we would expect frames to still be highly correlated with one another ($r > 0.95$) considering the relatively slow timescale of the cardiac cycle and the high sampling frequencies achievable with high-frame-rate ultrasound. Large drops in the correlation coefficient will therefore be indicative of highly inconsistent frame-by-frame output.

The Pearson correlation coefficient is calculated on a pixel-by-pixel basis. While minimizing pixel-to-pixel variations is important from an algorithmic perspective the human visual system is very sensitive to changes in orientation and contours or *structure* [89][90][91]. The structural similarity index (SSIM) is a metric frequently used to assess image quality which computes the similarity of a given image and its reference based on the distribution of pixels within a small window [92]. The SSIM is a weighted combination of three metrics: *luminance*, *structure*, and *contrast* and is shown in its most common form below in Equation 18 which compares between two images x and y. Here μ_x and σ_x and μ_y and σ_y represent the average and variance of pixels in x and y while σ_{xy} is the covariance. c_1 and c_2 are numerical values used to stabilize the division.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{18}$$

While the SSIM is generally used to compare a computed or modified image against its reference we aim to calculate the SSIM between subsequent frames in the filtered cineloop as a metric of playback 'smoothness'[93]. We will use the SSIM as a second metric in addition to the Pearson correlation as a metric to evaluate cineloop performance and quantify the presence of 'flashing' and frame-by-frame inconsistency in a manner consistent with the human visual system. As flow is evolving in time we would expect the SSIM of each frame pair to be less than 1; the purpose of computing the SSIM would be to identify large drops or inconsistencies in the SSIM that would indicate large structural differences in flow between frames.

High Pearson correlation or SSIM values are not valuable in isolation as they are only a metric of cineloop smoothness and not quality. A filter **can** perform consistently and poorly. We use these metrics primarily as a check on the consistency of the final filtered power Doppler images.

Part IV

Assessment of Filter Performance

Reproduction of Target AUC Curves

After training, the predicted AUC values were compared with the target AUC values. In general, good agreement is seen between the network output and the unseen AUC values. Figure 34 shows the reproduced and target AUC values for 9 random blocks in the D5B unseen dataset (brachial artery) at a random frame in the cineloop (neither systole nor diastole). The AUC values of blocks with flow are faithfully reproduced although the network output tends to be slightly smoothed. We note that the network exhibits much greater success during systole when the flow signal is stronger than during diastole where the flow signal is weaker. Additionally, blocks possessing fewer flow pixels tend to be reproduced worse than blocks containing significant flow signals.

However, in some cases the network reproduces the general structure of the AUC curve correctly albeit seemingly scaled down in magnitude. This can be seen in block 500 of Figure 35 where the reproduced AUC values are scaled up or down from the actual unseen values but the AUC curve structure is faithfully reproduced. We expect that this scaling is occurring because the calculated AUC is invariant to the overall amount of flow signal and in some cases the network is obtaining the correct relationships between discriminating factors to obtain the correct AUC curve structure, but lacking some piece of magnitude information to get the proportionality correct.

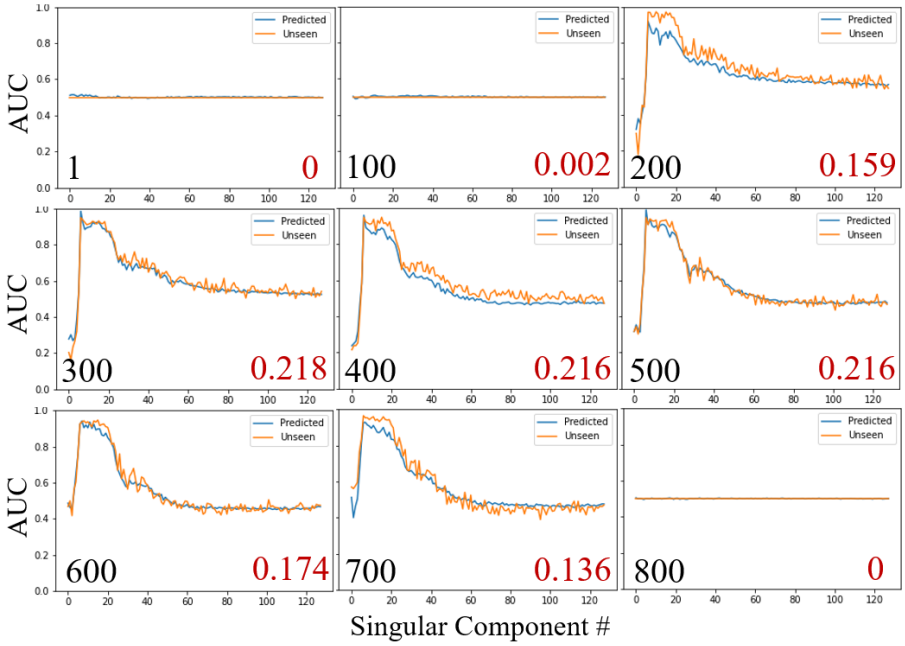

Figure 34: The predicted and actual AUC values for 9 blocks from frame 79 (midway between systole and diastole) of unseen acquisition D5B (**brachial artery**). The black numbers on the lower left hand side is the block # while the red numbers give the fraction of flow pixels contained within each block.

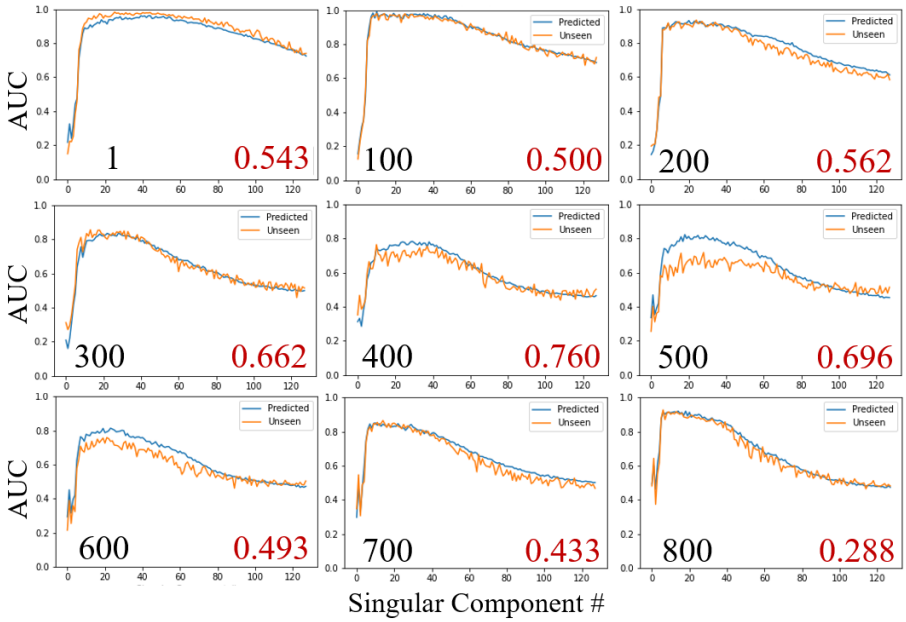

Figure 35: The predicted and actual AUC values for 9 blocks from frame 190 (systole) of unseen acquisition F03 (**femoral bifurcation**). The black numbers on the lower left hand side is the block # while the red numbers give the fraction of flow pixels contained within each block. All blocks contain flow, and the AUC values for components in most blocks are reproduced accurately.

While the reproduction of the AUC curve is reasonably robust for all unseen acquisitions when the amount of flow signal is significant, it does suffer significantly during diastole when the flow signal is weak. Figure 36 shows a few blocks acquired during diastole for several selected blocks in the unseen acquisitions (in the case of the brachial artery, AUC reproduction was generally quite good). In these cases reproduction of the target AUC curve is poor with not only non-matching curve structures being predicted by the network but also incorrect predictions on the presence of flow. However, in these scenarios no filtering method performs well (shown in the following section) and incorrect thresholds are chosen using the tested literature techniques (Table 5) as well.

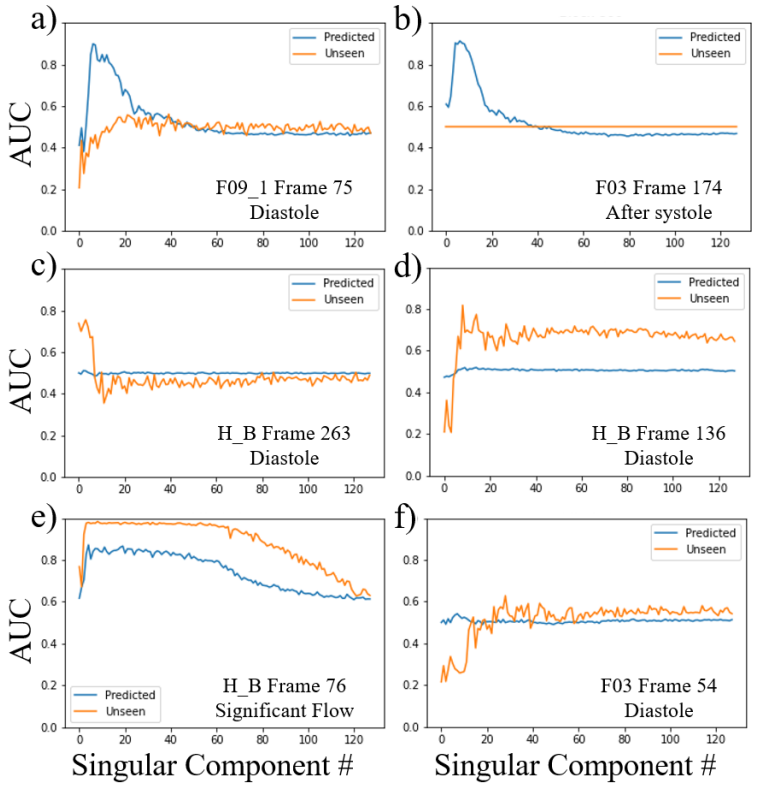

Figure 36: Some examples of incorrect AUC reproduction. *a*) A flow block without significant flow during diastole has components incorrectly marked as flow components. *b*) A non-flow block (according to segmentation) has components incorrectly marked as flow components. *c*) Something is incorrect with the unseen AUC calculation, the first component should never possess a high AUC value. The network predicts no flow present. *d*) Similar to pane *e*) but with a different structure. *e*) In the *in-vitro* flow phantom the flow signal is extremely strong and the AUC values of the flow components should be very high. Curiously, the network has predicted lower AUC values. *f*) Although the curves differ significantly, all component AUC values are low enough ($\lesssim 0.75$) that no components would be used to form the filtered image. F09_1 refers to another tested acquisition not presented here.

Clutter Filtering Performance

The proposed framework was compared to the most prevalent of the more complex SVD techniques used in the literature, specifically the hybrid estimator described in Ref. [50] and shown in Figure 11 and the SSM block correlation estimator from Ref. [29] described in Equation 10. Reference images produced using a generic high-pass FIR filter were also obtained. Clutter filter performance was assessed on the filtered power Doppler image using the contrast ratio and through ROC analysis (after flow region segmentation)[20]. This is

[20]The contrast ratio serves to quantify the relative difference in flow and background signal intensities at a specified point, while ROC analysis gives a measure of sensitivity and specificity across the entire segmented flow region. Alternatively as another performance metric we can compute the contrast ratio between the

summarized in Table 5.

Table 5: Tested Clutter Filters

Clutter Filtering Technique
1. Proposed Framework using AUC reproduction
2. Hybrid Method, Ref [50] (30 Hz Cutoff)
3. SSM spatial similarity, Ref [29]
4. High-pass filter[1]

[1] The entire data acquisition ($\sim$ 9000 frames) is first filtered using a high order FIR filter in Matlab (applied using Matlab's *filtfilt* function). Segments of the filtered signal (of size *Ensemble Length*) are used to form each frame in the Doppler cineloop. Filter parameters: normalized cutoff = 0.01, normalized passband = 0.05, attenuation = 100 dB, order = 169. The filter was created using Matlab's *firpm* function (Optimal Parks-McClellan FIR filter design tool). The chosen high-pass-filter was designed to represent the some of the best results that could be obtained with frequency filtering, with high stopband attenuation, a high order, and with the actual filtering being applied to the entire 18,000 frame cineloop rather than shorter frame-by-frame ensembles.

Although not a clutter filter itself, we will also make comparisons to the filtered results obtained using the actual calculated AUC curves using the same AUC thresholding values. This 'ground truth' AUC (G-AUC) result represents the ideal network output given our preprocessing and training technique using ROC analysis (these are the AUC values that would be used for training if this acquisition was not used for filter performance evaluation). By comparing to these 'ground truth' results we can ascertain whether potential poor performance is due to training and network configuration issues or the preprocessing and inherent methodology (specifically the use of ROC analysis on the segmented flow region and the AUC as a target metric).

AUC Sensitivity Thresholds

Several AUC thresholds were tested for each acquisition. As previously discussed AUC threshold selection was only weakly adaptive; statically set unless the AUC of the first singular component was very high (indicating an overlapping tissue signal or a lack of flow in the block). The formula used for setting an AUC threshold for the network output or the ground AUC values was given by

$$\text{Threshold} = max\big(\text{Set Threshold}, \text{Multiplier} \times AUC(1)\big) \tag{19}$$

total segmented flow and non-flow regions.

Where 'Set Threshold' is the predefined threshold AUC and 'Multiplier' is a defined proportion. The AUC threshold used for that particular block was set as the max of the predefined threshold or the estimated/computed AUC of the first singular component multiplied by the predefined value of 'Multiplier'. Including this additional condition helped select for AUC curves that were concave down (and therefore more likely to contain flow). Under this approach for instance, no components in the block shown in Figure 36*c*) would be identified as flow components.

Four sensitivity thresholds were tested, with their 'Set Threshold' and 'Multiplier' values shown in Table 6. Unless otherwise noted, all results presented in this work (network output and ground AUC calculations) utilized the third sensitivity threshold (**T3**) as it was generally found to be most optimal.

Table 6: Tested AUC Thresholds for the Proposed DNN SVD Filter

Filter Thresholds	Set Threshold	Multiplier
T1	0.65	1.2
T2	0.70	1.25
T3	0.75	1.4
T4	0.85	1.25

In-Vitro Carotid Flow Phantom

An in-house walled carotid bifurcation flow phantom [76] connected to a flow pump (setup details found in Ref. [94]) using blood mimicking fluid (nylon scatters – BMF-US; Shelley Medical Imaging, matched to the acoustic and viscous properties of human blood[95]) was used to assess the performance of the proposed filter *in-vitro*. At the beginning of the acquisition the flow pump was turned off. Shortly after beginning data acquisition, the flow pump was turned on. A carotid flow profile with a repetition frequency of 1 Hz (60 bpm) was generated with a volume flow rate of 12 mL/s. This volume flow rate during peak systole was sufficient to nearly generate two cycle aliasing in the upper branch of the phantom.

This simplistic scenario, with an *in-vitro* flow phantom, is included not to show how the proposed DNN filter is significantly more sensitive to flow and outperforms current literature filtering techniques (this is shown in the following sections), but rather as a test of robustness to illustrate the performance of our technique in a 'worst case' scenario – when the much more simplistic high-pass filter performs excellently and there is little to be gained and potentially much to be lost when using an SVD filter. In this situation, the fundamental limitations of the SVD and our component identification accuracy start to come into play, specifically how well the flow and clutter signals are distinctly decomposed into separate components by the SVD and how consistently flow and clutter components can be identified over multiple frames of 1156 blocks. The proposed DNN filter shows clear improvements over the tested literature filtering techniques in this scenario. Ultimately though, this scenario is included to demonstrate the lower bound performance of the proposed methodology; relatively little training data and a scenario where high-pass filtering shines.

We show the filtering performance for three representative frames in the acquisition. Figure 37 shows frame 53 from the beginning of the acquisition when there is little blood flow within the phantom. Strong, not aliased flow is present in frame 95 shown in Figure 38. Figure 39 of frame 200 shows filtering performance with high flow volume rates and significant aliasing. As flow velocities are high, we present the colour Doppler images in the bottom half of each figure described above as well to ascertain whether the proposed SVD filter is able to accurately produce correct velocity measurements. Table 7 summarizes the contrast and AUC measurements for the tested clutter filters in these scenarios.

We note that in the power Doppler images formed using the high-pass filter, significant reflections from the walls of the walled phantom are visible, especially as the high flow rate causes the walls of the phantom to move and generate tissue clutter. Along with the significant aliasing present in the upper branch, this is expected to produce a more challenging filtering scenario than typical *in-vitro* flow acquisitions. Due to the relatively small amount of *in-vitro* samples used in training (three total acquisitions) this is expected to be a challenging test for the proposed network, indicative of a scenario where limited training data is available[21].

Good performance was generally obtained using all tested filtering techniques. High pass filtering produced reasonable results although strong wall reflections were always observed (in testing the cutoff frequency of the high pass filter was adjusted but this did not appear to have an appreciable effect on the wall reflections). The SVD clutter filters had significantly greater success in removing these wall reflection artifacts with the DNN filter producing the best results, eliminating the reflections entirely or at worst, reducing them to intensities comparable to that of the flow pixels. All SVD filters suffered to some degree from incorrect component identification during some part of the cardiac cycle (Figures 37*c* and 38*c* for the SSM filter, Figure 38*a* for the proposed DNN filter, and Figure 39*h* for the hybrid filter). In general, we note that the proposed DNN filter produced generally good results when compared to the other tested filters offering improved contrast and AUC values when compared to the two literature SVD techniques. However, the proposed method encountered difficulty with clutter removal in the upper branch of the bifurcation and demonstrated a lack of sensitivity to flow in the lower branch. This same effect was observed to an even greater degree in the power Doppler images formed using the ground AUC values indicating an issue with the uniform AUC threshold set across all blocks (for the entire cineloop). It is possible that if adjustments in the training methodology and/or a more adaptive AUC threshold method was employed to correct the AUC values for potential bias between the upper and lower branches that this issue could be corrected.

Although, SVD filtering techniques are most frequently used to produce power Doppler measurements probing flow sensitivity, we have also chosen to generate the colour Doppler image to ascertain whether the proposed and literature SVD filtering techniques would bias velocity measurements due to choice of flow components of the decomposition. Doppler frequency measurements are presented for the entire frame without power thresholding as **1**). The colour Doppler measurements are intended solely as a evaluation of velocity estimation

[21] As discussed before, due to the lack of any sort of tissue structure, varying attenuations or speeds of sound, etc. *in-vitro* acquisitions generate very different statistics compared to *in-vivo* acquisitions.

and **2**). Velocities thresholded by poor power maps may inadvertently make accurate velocity estimations look visually poor and **3**). It is useful from a filtering perspective to observe what frequency components are retained or discarded in the non-flow regions (i.e. high-frequency noise elimination). We note that with the exception of the hybrid technique, which is not able to identify the aliased flow components, velocity measurements from all tested filters generally agree with one another. However, in some cases (Figure 38*g*) an incorrectly identified clutter component identified as flow created a minor artifact in the color Doppler images created using the DNN and SSM filtering techniques.

In summary however, this acquisition represents a scenario (strong flow *in-vitro*) where more than sufficient performance is obtained with the high-pass filter. While the tested SVD filters are able to remove the reflections originating from the vessel walls, they do have difficulty removing the clutter present on the left hand side of the image just above the vessel wall (yellow arrow in Figure 37a). While the proposed method clearly performs better than literature SVD techniques (Figures 37 – 38) the high pass filter performs extremely well in this scenario and already produces an excellent AUC of 0.969 ± 0.011. Rather, this acquisition serves to illustrate that **1)** the proposed method is more consistent and higher performing than literature SVD techniques, and **2)** the extremely difficult task of robustly identifying the flow components (ensemble size of 128) in 1156 blocks per frame for multiple frames can be achieved to a relatively high level using a network trained using limited *in-vitro* data. Ultimately however, all SVD filters are subject to the limitations of the decomposition[22]; if the level of mixing is greater than the separation of flow and clutter signals using a high-pass filter then no SVD filter will outperform the high-pass filter.

[22]Imperfect separation of flow and clutter signals into separate components without mixing.

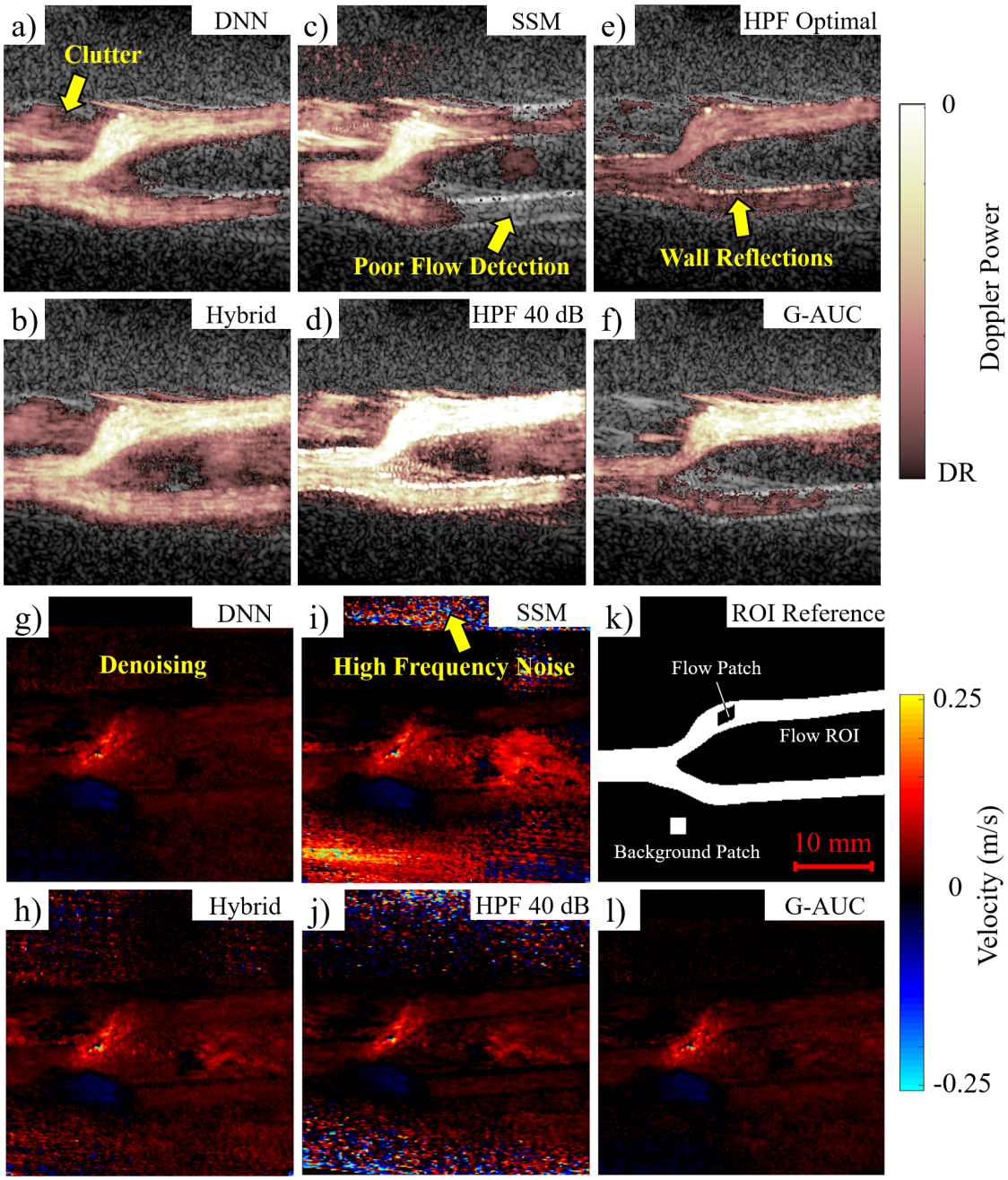

Figure 37: Power and colour Doppler images produced using the clutter filtering techniques described in Table 5 for frame 50 of acquisition CB (weak flow scenario). *a*), *g*) The proposed DNN filter. *b*), *h*) Hybrid technique. *c*), *i*) Spatial similarity technique. *d*), *j*) High-pass filtering under 40 dB dynamic range. *e*) High pass filtering using an 'optimally' set 42.5 dB dynamic range relative to the maximum pixel intensity. *f*), *k*) An example blood region ROI used for AUC calculations and the flow and tissue patch ROIs. *l*) The reference filtered images formed using calculated AUC values (based on a hand segmented blood ROI). All SVD filtered images are presented under 20 dB dynamic range. Compared to other filtering techniques it appears that the flow sensitivity of the G-AUC is too high in the upper branch and two low in the lower branch.

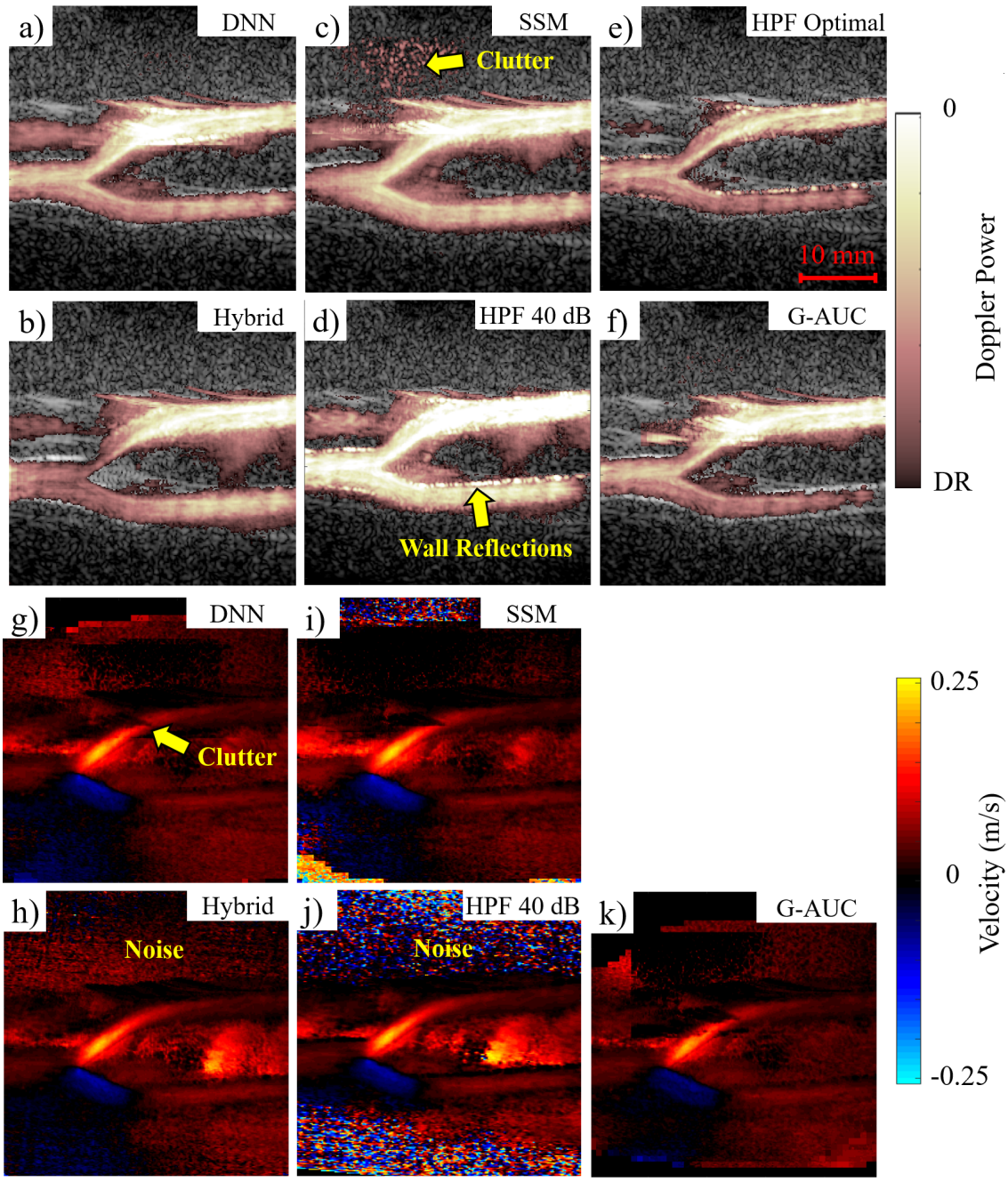

Figure 38: Power and colour Doppler images produced using the clutter filtering techniques described in Table 5 for frame 95 of acquisition CB (strong flow scenario). *a*), *g*) The proposed DNN filter. *b*), *h*) Hybrid method. *c*), *i*) Spatial similarity method. *d*), *j*) High-pass filtering under 40 dB dynamic range. *e*) High pass filtering using an 'optimally' set 34.3 dB dynamic range relative to the maximum pixel intensity. *f*), *k*) Ground AUC values (G-AUC). All SVD filtered images are presented under 20 dB dynamic range. Some clutter components in some data blocks were incorrectly classified as flow components resulting in the 'block-like' artifacts characteristic of blockwise SVD.

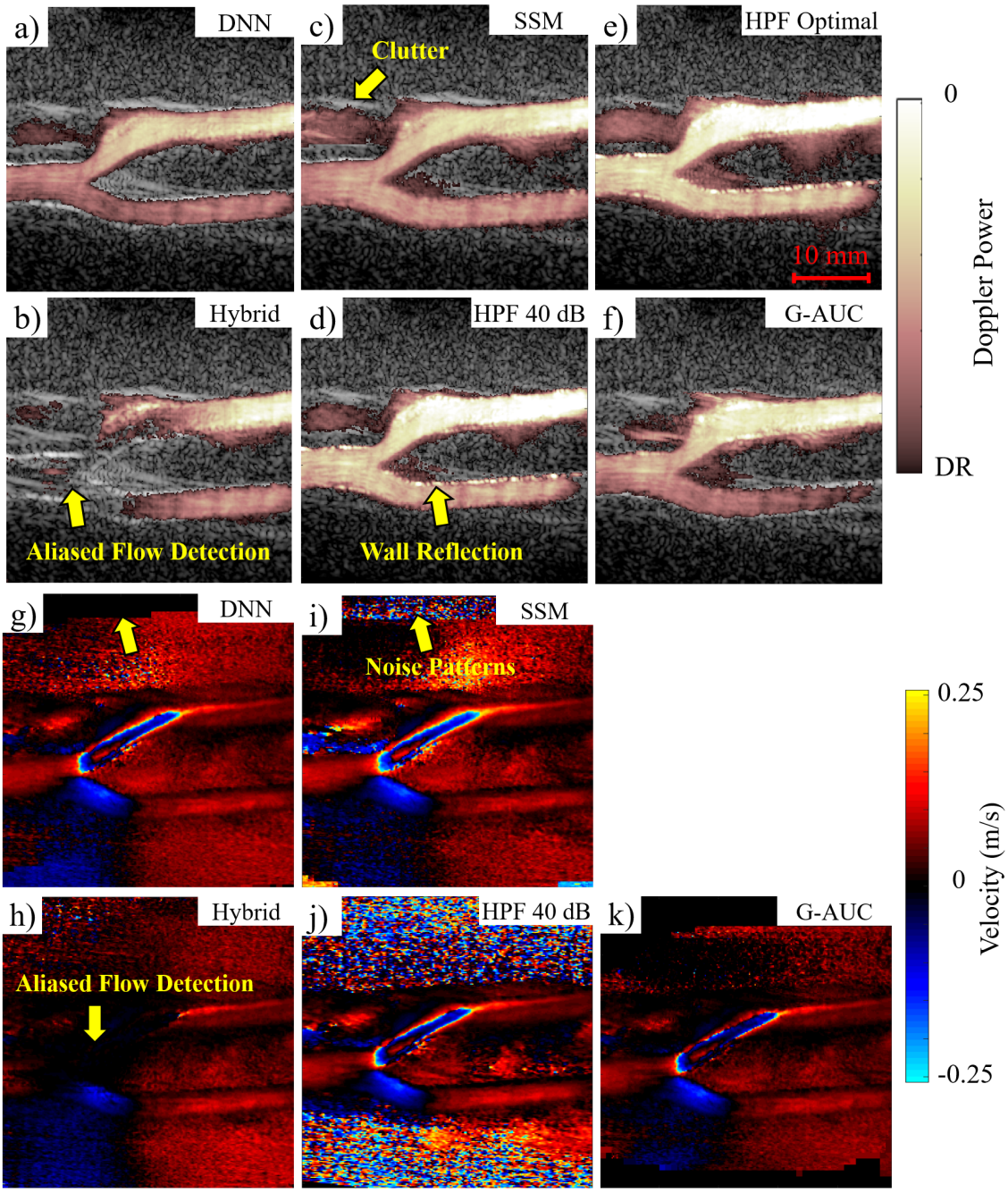

Figure 39: Power and colour Doppler images produced using the clutter filtering techniques described in Table 5 for frame 200 of acquisition CB (strong flow scenario). *a*), *g*) The proposed DNN filter. *b*), *h*) Hybrid method. *c*), *i*) Spatial similarity method. *d*), *j*) High-pass filtering under 40 dB dynamic range. *e*) High pass filtering using an 'optimally' set 31.2 dB dynamic range relative to the maximum pixel intensity. *f*), *k*) Ground AUC values (G-AUC). All SVD filtered images are presented under 20 dB dynamic range. The hybrid technique was unable to identify the aliased part of the flow signal.

Table 7: Filtering Performance: Carotid Bifurcation Phantom (CF)

	Filter			
Weak/No Flow (Frame 53)	DNN	Spatial	Hybrid	HPF
Contrast (Patch) [dB]	22.7	28.0	20.1	40.1[1]
Contrast (Region) [dB]	16.0	10.33	13.4	15.3[1]
AUC	0.945	0.854	0.968	0.949
Strong Flow (Frame 95)				
Contrast (Patch) [dB]	35.6	28.9	20.5	37.9[1]
Contrast (Region) [dB]	16.8	14.8	14.6	25.1[1]
AUC	0.972	0.966	0.964	0.978
Aliasing (Frame 200)				
Contrast (Patch) [dB]	31.1	18.54	17.7	35.9[1]
Contrast (Region) [dB]	17.3	14.6	12.5	23.9[1]
AUC	0.982	0.981	0.879	0.973
Cineloop Average[2]				
Contrast (Patch) [dB]	29.2 ± 3.2	20.7 ± 3.6	19.4 ± 5.4	37.8 ±2.9[1]
Contrast (Region) [dB]	17.2 ± 0.6	14.6 ± 1.0	12.8 ± 1.1	22.2 ±4.4[1]
AUC	0.975 ± 0.012	0.972 ± 0.022	0.848 ± 0.080	0.969 ± 0.011
Pearson Correlation	0.993 ± 0.009	0.994 ± 0.014	0.991 ± 0.021	0.985 ± 0.016
SSIM	0.869 ± 0.036	0.896 ± 0.039	0.867 ± 0.033	0.934 ± 0.015

[1] These are the default (unscaled) contrast ratios produced using the high-pass filter. These should not be compared directly to the contrast measurements produced through SVD filters and are presented here for completeness.

[2] Averaged over frames 50 - 280.

Figure 40 tracks the contrast of the patch and flow region shown in Figure 37k) and the AUC of the tested clutter filters across the entire 350 frame cineloop with the results over the entire cineloop (Frames 50-280) summarized in Table 7. We note that at the beginning of the cineloop where no (or little) flow is present all filters perform poorly and exhibit poor statistics (as we are trying to measure flow that isn't there). Note that due to the entirely different filtering technique employed by high-pass filtering the contrast values produced are not directly comparable to those produced by SVD filtering[23]. All filters produce reasonably consistent frame-to-frame results.

[23]In theory, the effective contrast between ideal flow and non-flow signals is scaled by the filter attenuation. As a high performance high-pass filter operating on the entire cineloop was used, the effective contrast between readily identifiable flow (above the cutoff frequency) and tissue signals (below the cutoff frequency) is very high.

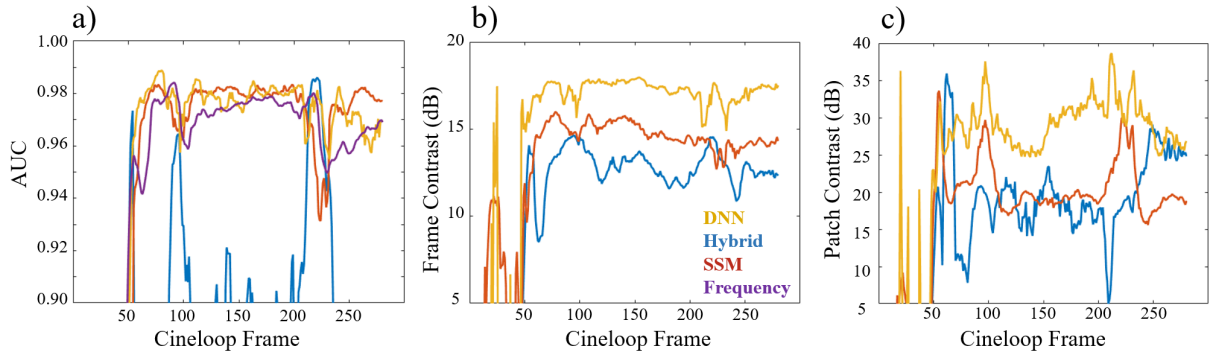

Figure 40: *a*) The area under the curve. All techniques exhibit poor flow detection before the flow pump is turned on (frame 50). After the pump is turned on most filters exhibit somewhat comparable performance with the exception of the Hybrid filter which does not perform well is aliasing flow is present. *b*) & *c*) The contrast between the flow and non-flow regions and between the blood and background ROIs. The DNN filter produces significantly better contrast compared with the other SVD techniques (results using the high pass filter are not shown here due to the different intensity scales).

Concluding Remarks

This imaging scenario was expected to be especially challenging for the proposed framework due to the lack of *in-vitro* data for training. However DNN clutter filtering performance performs better than the tested literature SVD techniques, both of which encounter serious difficulties (SSM method tends to incorrectly identify flow blocks while the hybrid method cannot deal with aliased signals). Observing the power and colour Doppler images produced using calculated AUC values we note that the proposed choice of using the AUC curve as a target works sufficiently for the network to reproduce accurate power and colour Doppler flow images. However, we note that the choice of a simplistic and uniform sensitivity threshold across the entire image is non-ideal as illustrated by the large amount of clutter in the upper branch of the phantom and the relatively weak flow sensitivity in the lower branch. This effect is moderated somewhat by the DNN (which has an averaging effect), however, sensitivity is still weak in the lower branch and additional clutter is present in the upper branch (as evidenced by comparisons to the other SVD clutter filters).

It appears, looking solely at the AUC values that little is gained in this scenario when comparing the DNN SVD filter to the high pass filter. The performance of the high-pass filter is generally quite good. However, the AUC is computed over the entire frame and the effect of small regions of incorrectly classified pixels (such as the highly reflective walls of the phantom) is relatively minor. The AUC is also not presented on an interval scale and differences in AUC between results may not scale as expected with visual identification.

Thus far there has been little discussion about filtering performance in the first 50 or so frames, before the flow pump is turned on. However, while the DNN could not separate the flow and the clutter components for this part of the cineloop, the flow within this region was identifiable to some degree in the SVD components. Figure 41 shows how the SVD filtering techniques tested in this work, fail to identify weak (perhaps residual) flow within

the carotid bifurcation phantom. These flow components, identifiable through ROC analysis, are present in the decomposition but could not be identified by any tested method, including the DNN clutter filter developed here that was trained using similar data. In the case of the proposed DNN filter, the inability to identify flow in these frames could be due to the previously-mentioned lack of *in-vitro* training data.

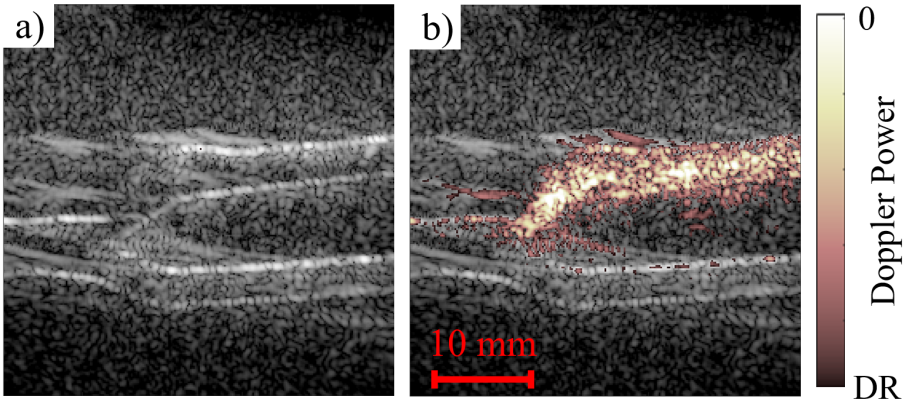

Figure 41: Power Doppler images for frame 15 of the carotid bifurcation acquisition acquired using the *a*) DNN SVD clutter filter and *b*) using the 'ground truth' AUC values for the same AUC sensitivity threshold. We see that flow is separated by the decomposition into some number of unmixed components. The power Doppler images produced using the hybrid and spatial correlation method produce equivalent power Doppler images to *a*).

Brachial Artery

The filtering results for an unseen acquisition of a volunteer's brachial artery (D5B in Table 4) is presented here. We note that in this acquisition flow within the artery was visible throughout the cineloop for every acquisition, including the high-pass filter, However, at certain points during the cardiac cycle, significant tissue clutter would appear for ∼20 frames. This clutter, along with strong wall reflections was extremely visible if a high pass filter was used. If an SVD filter was used, the clutter during these periods was greatly reduced (wall reflections completely removed). The proposed DNN SVD clutter filter was capable of removing almost all wall reflections and movement artifacts while still maintaining good sensitivity to flow.

Here we show the performance of the proposed methodology for two frames during diastole (slow flow) and systole (where significant clutter is present) against literature techniques. Due to a low beam to flow angle, flow velocities are difficult to measure accurately and we have elected to show only the power Doppler in the following figures.

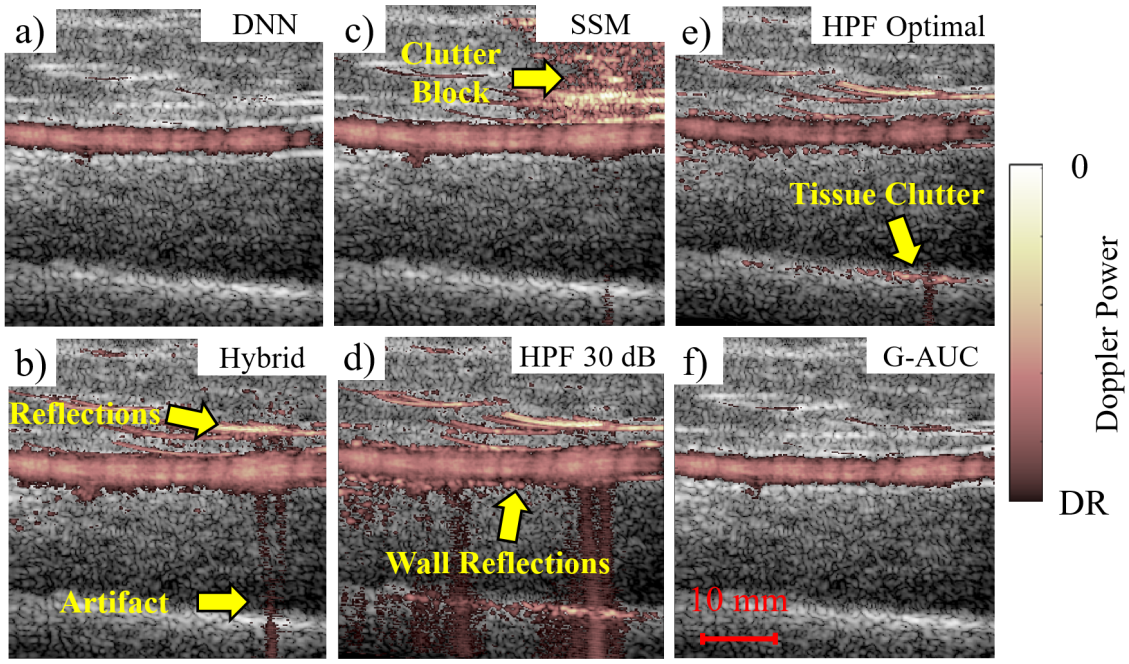

Figure 42: Power Doppler images produced using clutter filtering techniques (Table 5) during diastole where flow is present and there is little motion. (Frame 10 of D5B). Power Doppler images were obtained obtained using *a*) The proposed DNN SVD clutter filter, *b*) hybrid method, *c*) SMM correlation method, and *d*) - *e*) high-pass FIR filter. *f*) the 'ground truth' AUC values. Panes *a*) - *c*) are presented under 20 dB dynamic range. Pane *d*) under 30 dB dB dynamic range. Pane *e*) has been rescaled to a DR of 36.2 dB relative to max.

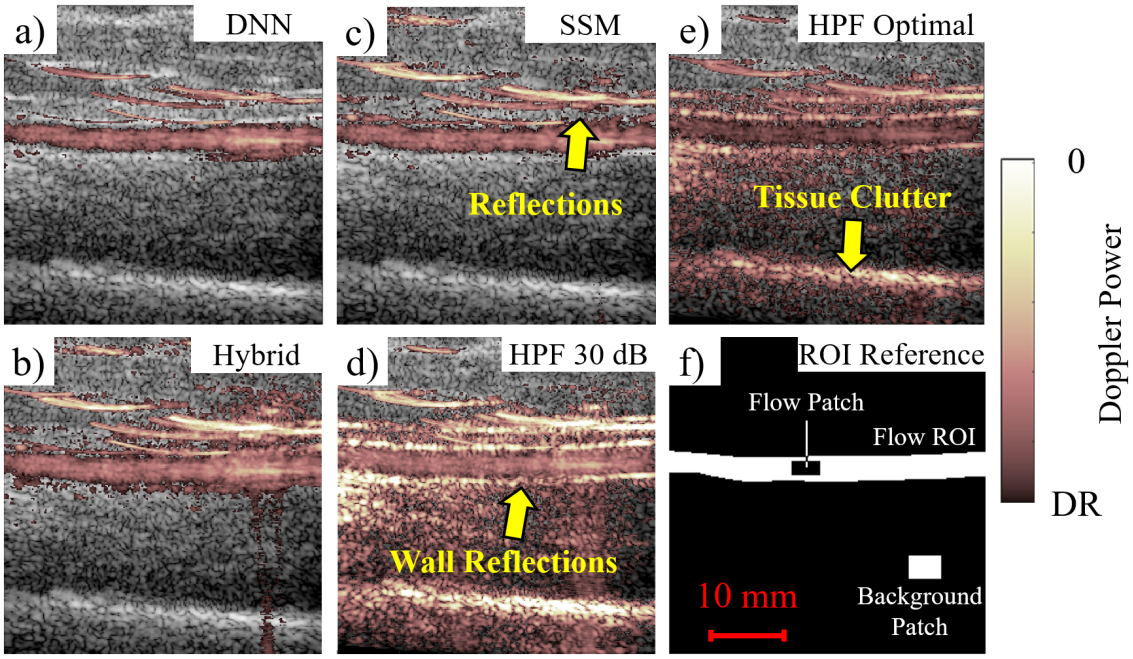

Figure 43: Power Doppler images produced using clutter filtering techniques (Table 5) during systole where significant patient motion is present in the brachial artery (Frame 100 of D5B). Power Doppler images were obtained obtained using *a*) The proposed DNN SVD clutter filter, *b*) hybrid method, *c*) SMM correlation method, and *d*) and *e*) high-pass FIR filter. Panes *a*) - *c*) are presented under 20 dB dynamic range. Pane *d*) is presented under 30 dB dynamic range. Pane *e*) has been rescaled to a DR of 56.6 dB DR relative to the maximum pixel intensity. Pane *f*) shows an example segmentation of the total flow region and the flow and background ROIs used to compute the values in Table 8.

While the SVD filters were able to remove the bright wall reflections (or at least reduce them down to the same intensity as flow within the vessel), the proposed model was additionally able to remove the beam-like artifacts present below the brachial artery in frame 10. In frame 100, the proposed model was able to remove significantly more of the surrounding clutter compared to the tested filters while maintaining the same flow signal intensity. Although not shown here, the improved power-map of the flow signal has the potential to significantly 'clean-up' color Doppler images due to better correlation with the actual flow region.

Table 8: Filtering Performance: Brachial Artery (D5B)

	Filter			
Frame 10	DNN	Spatial	Hybrid	HPF
Contrast (Patch) [dB]	30.6	9.7	8.5	11.9[1]
Contrast (Region) [dB]	17.0	8.9	8.4	10.7[1]
AUC	0.993	0.930	0.978	0.953
Frame 100				
Contrast (Patch) [dB]	35.1	7.7	6.6	5.1[1]
Contrast (Region) [dB]	20.2	9.2	6.3	-5.5[1]
AUC	0.970	0.933	0.915	0.738
Clean Signal[2]				
Contrast (Patch) [dB]	36.1 ± 1.7	11.0 ± 0.7	9.4 ± 0.6	13.3 ± 1.6[1]
Contrast (Region) [dB]	20.2 ± 0.5	11.5 ± 1.4	8.9 ± 1.3	11.0 ± 2.7[1]
AUC	0.995 ± 0.004	0.980 ± 0.019	0.973 ± 0.021	0.940 ± 0.025
Pearson Correlation	0.995 ± 0.002	0.966 ± 0.076	0.981 ± 0.017	
SSIM	0.881 ± 0.028	0.775 ± 0.038	0.769 ± 0.049	0.931 ± 0.024
Tissue Motion[3]				
Contrast (Patch) [dB]	33.8 ± 2.3	10.6 ± 1.1	9.2 ± 0.8	6.7 ± 5.7[1]
Contrast (Region) [dB]	19.7 ± 0.8	9.4 ± 1.6	7.4 ± 0.9	0.8 ± 6.3[1]
AUC	0.991 ± 0.005	0.960 ± 0.026	0.950 ± 0.183	0.798 ± 0.096
Pearson Correlation	0.993 ± 0.004	0.982 ± 0.023	0.990 ± 0.008	0.989 ± 0.016
SSIM	0.861 ± 0.038	0.729 ± 0.047	0.755 ± 0.040	0.904 ± 0.080

[1] These are the default (unscaled) contrast ratios produced using the high-pass filter.

[2] Average of frames 10 – 85.

[2] Average of frames 150 – 220.

For these particular frames, the proposed method produces a significantly cleaner image with better contrast and higher AUC compared to literature techniques. These results are consistent across the entire cineloop as shown in Figure 44 which tracks the patch contrast, frame contrast, and AUC across the entire 350 frame cineloop for the literature SVD filters and the proposed methodology.

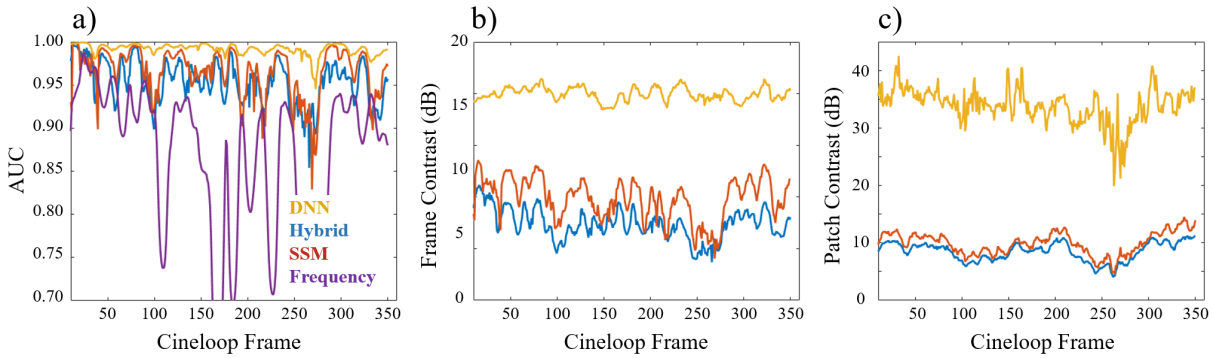

Figure 44: *a*) The area under the curve. The high pass filter struggles to remove tissue clutter throughout the acquisition. *b*) & *c*) The contrast between the flow and non-flow regions and between the blood and background ROIs. The DNN filter produces significantly better contrast compared with the other SVD techniques.

The effect of different sensitivity thresholds is shown in Figure 45 which shows how, regardless of choice of flow threshold[24], the proposed DNN clutter filter consistently outperforms other techniques and this performance is not due to choice of an ideal threshold. We observe that as the AUC threshold is increased, only the strongest flow components are observed. If the threshold is dropped sensitivity is decreased and more flow signal (and clutter due to the presence of mixed components) is present in the image.

[24]To within reasonable parameters. If increasingly inappropriate thresholds are chosen, incorrect components will start to be chosen in some blocks (see Figure 53), or virtually no signal will be observed.

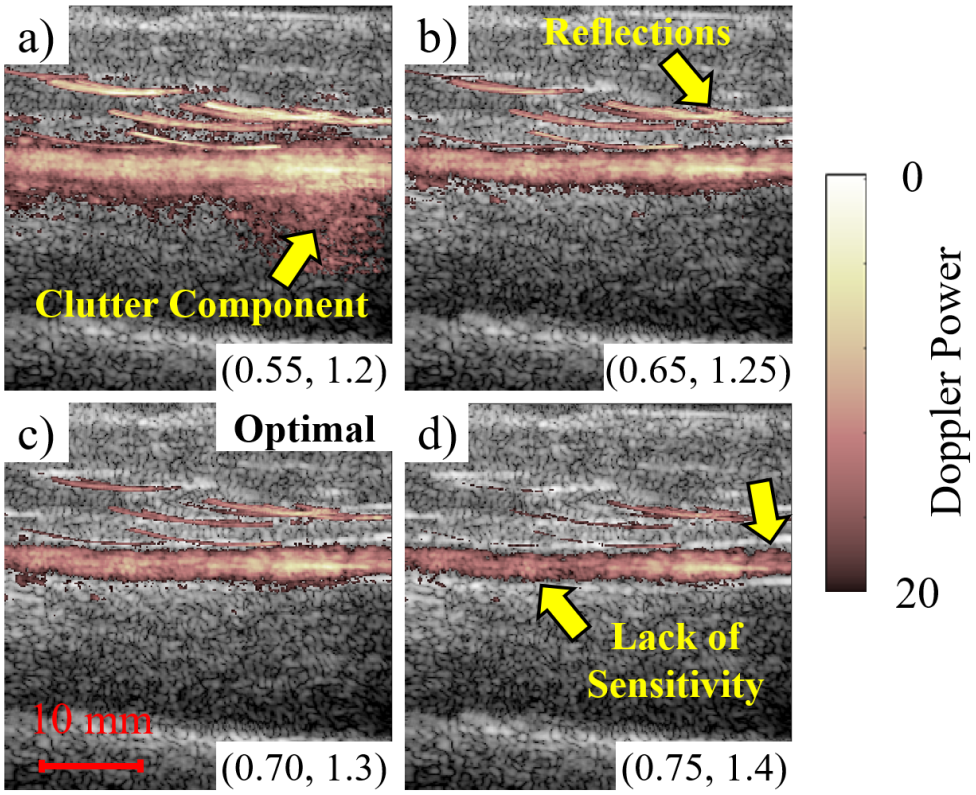

Figure 45: Power Doppler produced by the DNN for Frame 100 of the brachial acquisition for various sensitivity thresholds (*Set Threshold*, *Multiplier*). *a*) (0.55, 1.2) *b*) (0.65, 1.25) – The optimal threshold *c*) (0.70, 1.3) *d*) (0.75, 1.4). All images are presented under 20 dB dynamic range relative to the same fixed value. As the sensitivity threshold is modified, the trade-off between flow sensitivity and clutter removal is illustrated. In pane *d*), at a much higher sensitivity threshold, nearly all clutter is removed but flow detection at the edges of the lumen is correspondingly decreased. As the sensitivity threshold is decreased more flow is revealed, however, clutter starts to appear in the image. If the sensitivity is decreased too much as shown in pane *a*) additional clutter begins to appear.

Concluding Remarks

Compared to the high-pass filter, all SVD filters demonstrated lower frame to frame consistency due to the discrete nature of the singular value decomposition and the imposed restriction of 'choosing' predefined singular components. This is especially true if random tissue components are incorrectly identified as a flow component and included in frames of the cineloop. As these clutter components are much larger in magnitude than the flow components they will always appear in the log-scaled power Doppler image. The large number of blocks and singular components impose a very high accuracy threshold on SVD filters to consistently detect flow and reject clutter components that can be difficult to meet. Additionally, as high-pass filtering is performed on a sample-by-sample basis[25] the produced

[25]The filter order is larger than the effective ensemble size!

output is 'smoothed' whereas the SVD filters operate on a more granular ensemble by ensemble basis. Figure 46 shows the Pearson correlation and SSIM between adjacent frames across the cineloop. In general, all tested filtering techniques produced reasonably smooth cineloop playback[26], although consistency (but not filtering performance) was noticeably better with the high-pass filter.

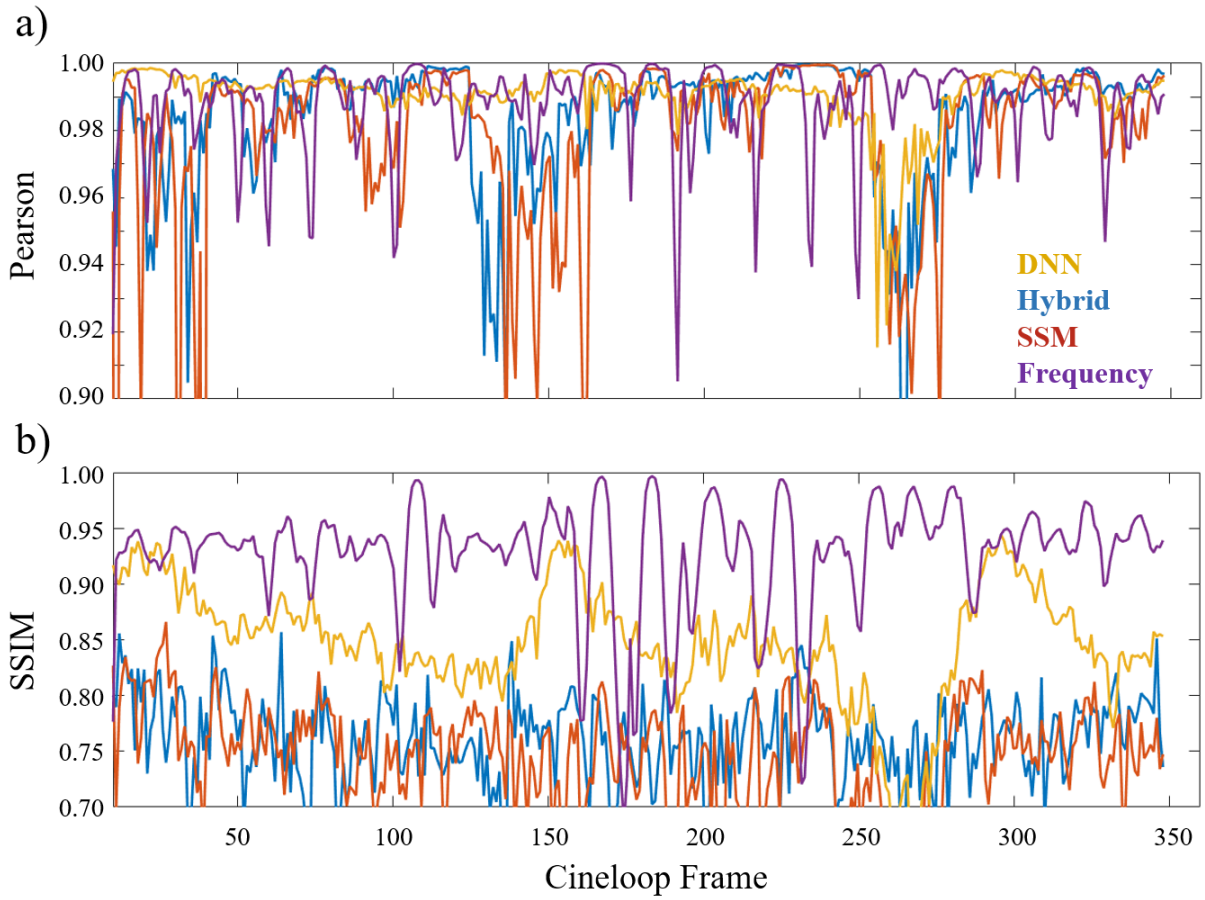

Figure 46: *a*) The Pearson correlation between frames for the tested filters. *b*) The structural similarity index measure (SSIM) between frames for the tested filters. The high-pass filter produced the most consistent frame to frame performance, however, the proposed model performed decently when compared to literature SVD filters as shown by the overall much higher SSIM over the entire cineloop as shown in pane *b*.

In summary, the proposed framework was able to consistently produce filtered power Doppler images with higher contrast and area under the ROC curve measurements than current literature filter in this particular acquisition of the brachial artery. Visually, the proposed filter was able to remove the beam-like artifacts that other filters could not and generally reduced the clutter artifacts that it could not wholly remove down to approximately the same intensity as the flow signal.

[26]Or 'flashing-type' phenomenon occurred over longer timescales i.e. on the order of multiple frames and thus could not be measured on a frame-to-frame basis.

Femoral Bifurcation

Two acquisitions of the femoral bifurcation were obtained from healthy volunteers and used for filter evaluation in more challenging flow conditions. In one acquisition not only was the femoral bifurcation visible but the femoral vein was also present, resulting in a scenario where pulsatile arterial flow and constant venous flow were present in the same image view (and even in the same block). These scenarios were chosen to assess filter performance under conditions of: **1)** significant tissue clutter originating from patient motion during systole, **2)** flow detection throughout the cardiac cycle, especially at low beam-to-flow angles and, **3**) consistency of filtering performance in complex flow scenarios on several individuals throughout the cardiac cycle.

Acquisition F03

In the F03 acquisition, significant clutter is observed during systole in the power and color Doppler images when using the high pass filter. This clutter is also difficult to remove with SVD filtering techniques. In addition, flow detection is poor during diastole in the upper branch as flow is almost perpendicular to the imaging axis. The proposed method was far more effective at removing clutter resulting from patient motion during systole than high pass filtering or the tested literature techniques as shown in Figure 47 for frame 77 of the acquisition. Here the DNN clutter filter was able to remove all clutter from the power Doppler image, producing a clear delineation of vessel boundaries. The reflection artifacts on the bottom of the main branch were mostly removed as well (reduced to about the same magnitude as the flow), allowing for accurate velocity estimation in this region (indicated by the yellow arrows). Table 9 shows the contrast of the patch, flow region and AUC for the segmentation shown in Figure 47k). The proposed filter shows a significant improvement in these quantities across the entire cineloop (Figure 48 and Table 9).

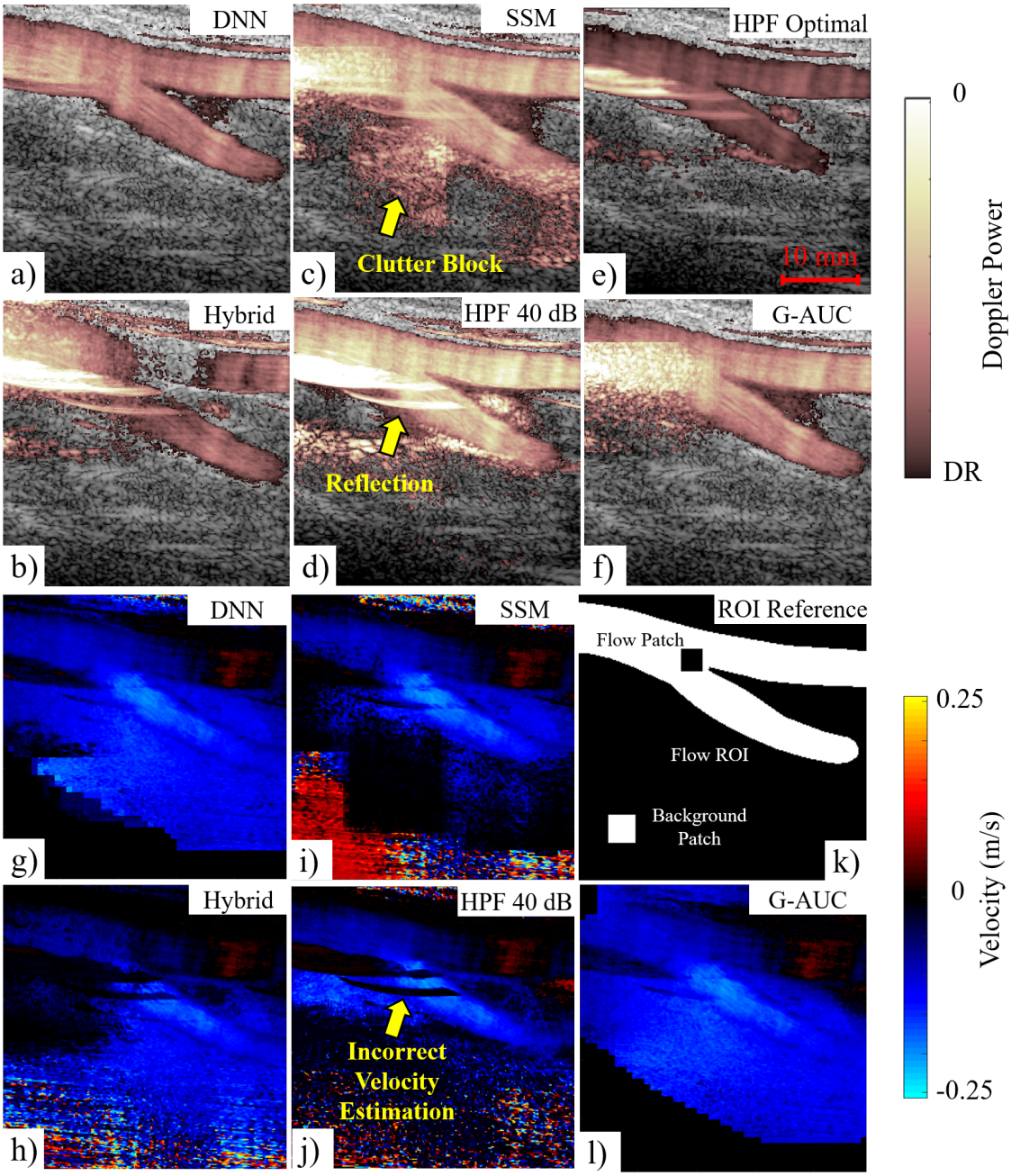

Figure 47: Power and colour Doppler images produced using the clutter filtering techniques described in Table 5 for frame 77 of acquisition F03 where significant tissue clutter is present. *a*), *g*) The proposed DNN filter. *b*), *h*) Hybrid method. *c*), *i*) Spatial similarity method. *d*), *j*) High-pass filtering under 40 dB dynamic range. *e*) High pass filtering using an 'optimally' set 62.2 dB dynamic range relative to the maximum pixel intensity. *f*), *k*) Ground AUC values (G-AUC). *l*) An example blood region ROI used for AUC calculations and the flow and tissue patch ROIs. All SVD filtered images are presented under 20 dB dynamic range. Only the proposed method was able to robustly identify flow and remove all tissue clutter from the power and color Doppler images.

Table 9: Filtering Performance: Femoral Artery (F03)

	Filter			
Frame 77	DNN	Spatial	Hybrid	HPF
Contrast (Patch) [dB]	70.0	11.7	13.8	26.6[1]
Contrast (Region) [dB]	15.6	8.2	5.1	4.7[1]
AUC	0.969	0.929	0.888	0.925
Systole[2]				
Contrast (Patch) [dB]	59.8 ± 8.8	15.99 ± 9.0	15.7 ± 2.4	26.4 ± 2.7[1]
Contrast (Region) [dB]	14.7 ± 1.5	7.9 ± 3.4	6.9 ± 1.3	10.9 ± 3.3[1]
AUC	0.951 ± 0.031	0.905 ± 0.116	0.917 ± 0.051	0.912 ± 0.149
Pearson Correlation	0.979 ± 0.045	0.981 ± 0.053	0.997 ± 0.003	0.985 ± 0.017
SSIM	0.840 ± 0.053	0.808 ± 0.044	0.826 ± 0.016	0.911 ± 0.042
Diastole[3]				
Contrast (Patch) [dB]	49.8 ± 9.4	23.5 ±1.6	16.9 ± 0.5	29.9 ± 1.5[1]
Contrast (Region) [dB]	11.5 ± 1.3	4.7 ± 1.7	5.6 ± 0.9	11.7 ± 1.4[1]
AUC	0.837 ± 0.024	0.796 ± 0.041	0.716 ± 0.053	0.613 ± 0.070
Pearson Correlation	0.991 ± 0.004	0.993 ± 0.013	0.992 ± 0.005	0.988 ± 0.003
SSIM	0.755 ± 0.024	0.792 ± 0.035	0.810 ± 0.016	0.924 ± 0.008

[1] These are the default (unscaled) contrast ratios produced using the high-pass filter.

[2] Average of frames 65 - 110.

[2] Average of frames 160 - 290.

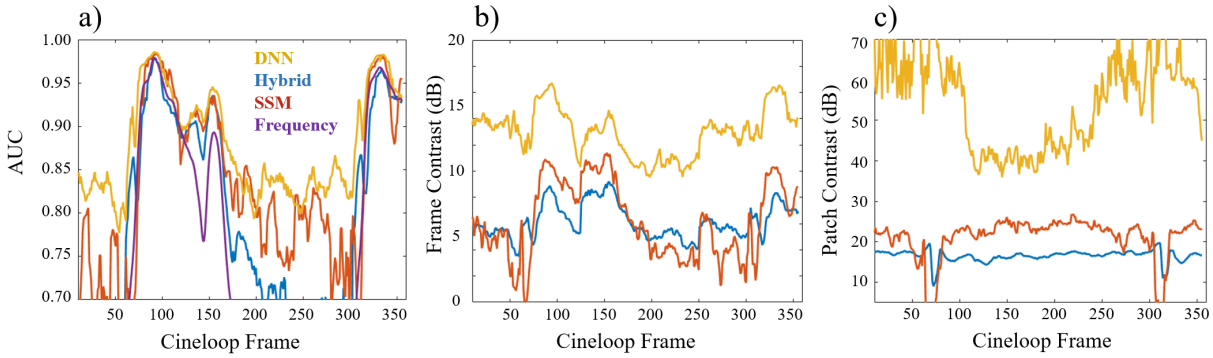

Figure 48: *a*) The area under the curve. All techniques struggle to detect flow in the upper branch during diastole resulting in poor AUC values. *b*) & *c*) The contrast between the flow and non-flow regions and between the blood and background ROIs. The DNN filter produces significantly better contrast compared with the other SVD techniques (results using the high pass filter are not shown here due to the different intensity scales).

A large part of the proposed filter's improved performance according to Figure 48 lies in significantly better clutter rejection rather than much better flow detection. Ultimately, the ability of an SVD clutter filter to detect flow is limited by the fundamental decomposition of the mixed clutter and flow signals into separate components. A decomposition of a data matrix that results in many mixed components will produce poorly filtered flow images no matter how inherently well designed and motivated the technique used to identify flow and clutter in the components of the decomposition is. While our proposed clutter filter is inherently better at separating flow from clutter (Figure 47) it is also significantly better at performing this identification for blocks which do not contain flow signal (i.e. no overlap on the segmented flow region). For these regions the network predicts a complete absence of flow (i.e AUC of 0.5, see Figure 34), resulting in an absence of signal in these regions.

Acquisition HB

This acquisition contains both the femoral bifurcation and the femoral vein. Figure 49 shows frame 215 of the acquisition, just after systole where flow is relatively weak and significant tissue motion is present. Due to the low velocities and high beam to flow angle the color Doppler images of this frame are not shown. In frame 215, the intense clutter band above the femoral artery is significantly reduced by our DNN filter when compared to high-pass filtering or literature SVD filters. Additionally, the beam-like artifacts have been removed by the proposed methodology (in general, the DNN was found to be quite adept at removing these types of artifacts). Flow detection in the lower branch was found to be poor using any filtering approach; however, all filtering techniques were able to resolve flow in the vein well.

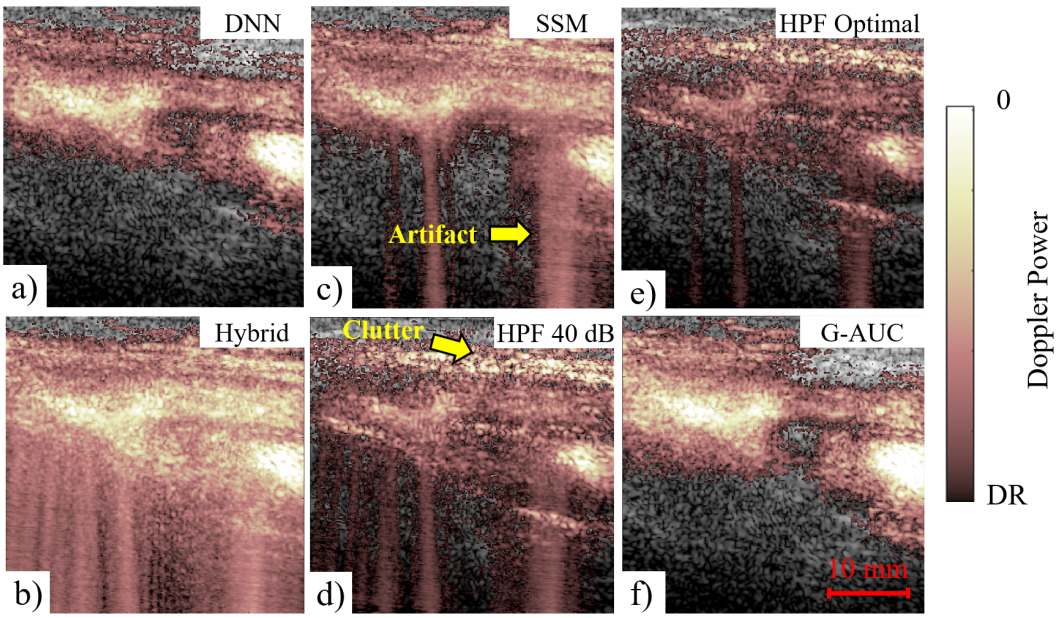

Figure 49: Power Doppler image of the femoral bifurcation and vein just after systole. Significant tissue motion is present and the flow signal is relatively weak (Frame 215). *a*) The proposed DNN SVD clutter filter, *b*) Hybrid method, *c*) SMM correlation method, and *d*) and *e*) high-pass FIR filter. Panes *a*) - *c*) are presented under 20 dB dynamic range. Pane *d*) is presented under 40 dB dynamic range. Pane *e*) has been rescaled to a DR of 42.6 dB DR relative to the maximum pixel intensity. *f*) ground AUC values (G-AUC) The proposed filter removes the majority of the surrounding clutter while maintaining good sensitivity to flow. Flow detection in the lower branch is generally poor for all filters.

In terms of sensitivity to flow during diastole, Figure 50 shows frame 254 of diastole where there is little flow in the femoral artery. At this point in the cardiac cycle, only the flow in the femoral vein is visible using the high-pass filter. We use this point in the cycle as a reference for sensitivity to slow flow, while still robustly identifying flow components in the femoral vein. In this scenario the proposed model (although using a slightly more sensitive AUC threshold), optimizes for the removal of clutter over the inclusion of heavily mixed clutter and flow (and predominantly clutter) components. Although this is likely the most statistically optimal choice for the network to make, it may not always be the most desired outcome. In some cases greater flow sensitivity may be more desirable. Using a theoretically more sensitive threshold (the T1 thresholds) did not result in the inclusion of these mixed components.

The results for frames 215 and 254 as well as the averages during systole and diastole are shown in Table 10. In general, all techniques performed well during peak systole when no clutter was present. However, performance during other parts of the cardiac cycle could be poor; cineloop statistics are shown in Figure 51.

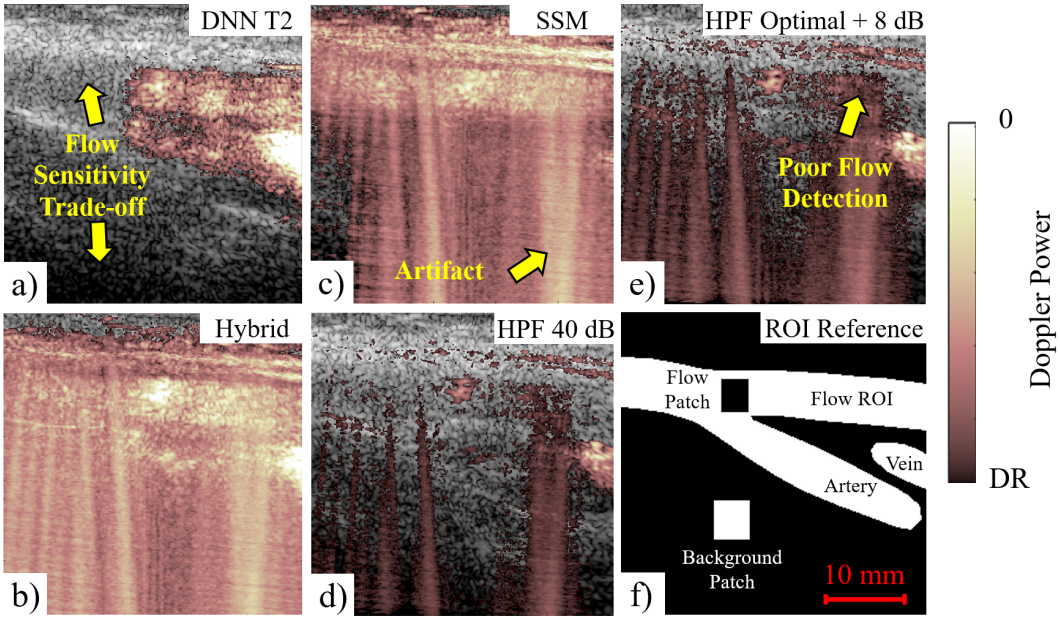

Figure 50: Power Doppler acquired during diastole where weak flow can just be resolved in the upper branch (Frame 254). Flow in the femoral vein is visible. *a*) The proposed DNN SVD clutter filter using the **T2** sensitivity threshold, *b*) Hybrid method, *c*) SMM correlation method, and *d*) and *e*) high-pass FIR filter. Panes *a*) - *c*) are presented under 20 dB dynamic range. Pane *d*) is presented under 40 dB dynamic range. Pane *e*) has been rescaled to a DR of 33.7 dB DR relative to the maximum pixel intensity. In this case an additional 8 dB of dynamic range was added to show the poor flow detection in the bifurcation of the high pass filter. Pane *f*) shows an example segmentation of the total flow region and the flow and background ROIs. The proposed filter correctly identifies the femoral vein and some flow in the upper and lower branches but fails to detect flow higher up the artery, marked by the yellow arrow.

As shown in Figure 51 all tested filters were able to detect flow in this femoral acquisition during systole fairly well. However, diastolic performance was significantly worse and no filter was able to detect flow well in either the femoral artery or vein during diastole, especially just before systole. This is unlikely to be related to filter performance and is much more likely to be related to the low SNR of the very weak flow during this portion of the cardiac cycle and the use of plane wave imaging rather than more sensitive conventional scan-line imaging. We will not include this portion of the cardiac cycle in our diastolic measurements in Table 10. The contrast measurements in Table 10 show very high values for the proposed framework. This is due to the greater clutter rejection performance of the network, in this case the elimination of the vertical beam-like artifacts.

Table 10: Filtering Performance: Femoral Bifurcation (HB)

	Filter			
Frame 215	DNN	Spatial	Hybrid	HPF
Contrast (Patch) [dB]	32.6	8.1	6.3	12.3[1]
Contrast (Region) [dB]	8.0	3.9	4.1	11.5[1]
AUC	0.861	0.786	0.844	0.730
Frame 254				
Contrast (Patch) [dB]	45.4	2.6	1.1	-3.7[1]
Contrast (Region) [dB]	6.9	1.5	1.5	-1.7[1]
AUC	0.808	0.686	0.692	0.531
Systole[2]				
Contrast (Patch) [dB]	44.7 ± 6.2	17.7 ±1.6	11.0 ± 0.6	17.0±3.2[1]
Contrast (Region) [dB]	14.8 ± 1.5	11.1 ± 1.4	8.6 ± 0.73	14.1 ± 3.9 [1]
AUC	0.960 ± 0.015	0.970 ± 0.017	0.968 ± 0.008	0.966 ± 0.012
Pearson Correlation	0.972 ± 0.028	0.989 ± 0.071	0.996 ± 0.003	0.988 ± 0.006
SSIM	0.797 ± 0.039	0.795 ± 0.042	0.827 ± 0.021	0.921 ± 0.010
Diastole[3]				
Contrast (Patch) [dB]	35.7 ± 7.8	7.7 ± 4.5	5.1 ± 3.9	9.3 ± 3.1[1]
Contrast (Region) [dB]	8.8 ± 4.4	4.9 ± 2.9	4.3 ± 2.6	6.0 ± 3.6[1]
AUC	0.873 ± 0.057	0.812 ± 0.106	0.820 ± 0.131	0.705 ± 0.192
Pearson Correlation	0.967 ± 0.028	0.974 ± 0.025	0.984 ± 0.005	0.978 ± 0.003
SSIM	0.732 ± 0.085	0.787 ± 0.029	0.778 ± 0.029	0.914 ± 0.049

[1] These are the default (unscaled) contrast ratios produced using the high-pass filter.

[2] Average of frames 160 - 200.

[2] Average of frames 215 - 250.

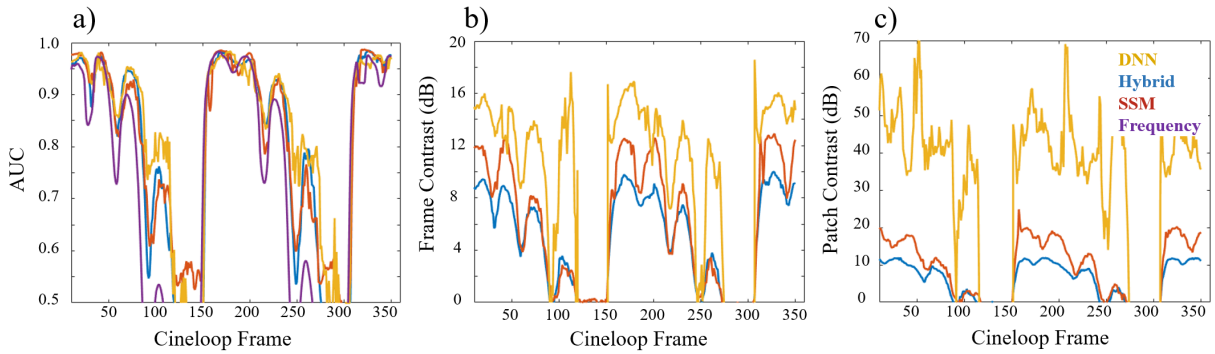

Figure 51: *a*) The area under the curve. All techniques perform well during systole and poor during diastole. During the end of diastole, no flow is identified within the flow region for any of the tested filters. *b*) & *c*) The contrast between the flow and non-flow regions and between the blood and background ROIs. Due to more robust removal of clutter elements the DNN filter produces significantly better contrast compared with the other SVD techniques

Concluding Remarks

The proposed DNN clutter filter produces a more ideal balance between clutter removal and flow detection than high-pass filtering or literature SVD filters as shown in Figures 49 and 50. The calculated AUC and contrast between flow and clutter regions was also significantly higher than the tested filter (Table 10) due to the proposed methods improved ability to identify and remove the predominantly clutter components from the decomposition. We will note that although the network does make the 'statistically optimal' identification of flow and clutter components it may be desirable to retain significantly more clutter if slightly better flow sensitivity can be obtained. However, we remark that much of the flow identification in the left-hand side of the upper branch in Figure 50 done by the eye is based on the presence of different 'textures' rather than different intensities. The images presented in Figure 50 are shown without any additional spatial filtering or additional post processing and thus possess a very 'raw' quality that would not be present in conventionally processed power Doppler images. If this additional spatial filtering, frame averaging, etc. were to be added, the textural information in panes *b*) and *c*) would be lost or greatly reduced.

Frame to frame consistency remains an issue for SVD filters in general due to the discrete nature of identifying the 'premade' flow components produced by the SVD. However, the proposed method is not significantly worse than current SVD filters and smoothness is only a desirable feature if the specified filter demonstrates good performance.

Proof of Concept: Small Vessel Imaging

To assess whether the proposed model, trained using only large segmentable flow vessels, was generalizable to small vessel imaging, the kidney of a healthy volunteer was imaged using the US4US scanner (same acquisition settings as described in Table 2). Although the signal to noise of the flow signal was weak due to the relative depth of the kidney and the use of

plane wave imaging, the branching arterial structure characteristic of kidney flow images was just visible in the lower right hand corner. Figure 52 shows the results of clutter filtering for a frame in the cineloop where flow could be observed. The proposed model accurately identifies flow within the kidney while rejecting the tissue clutter present in other parts of the imaging view.

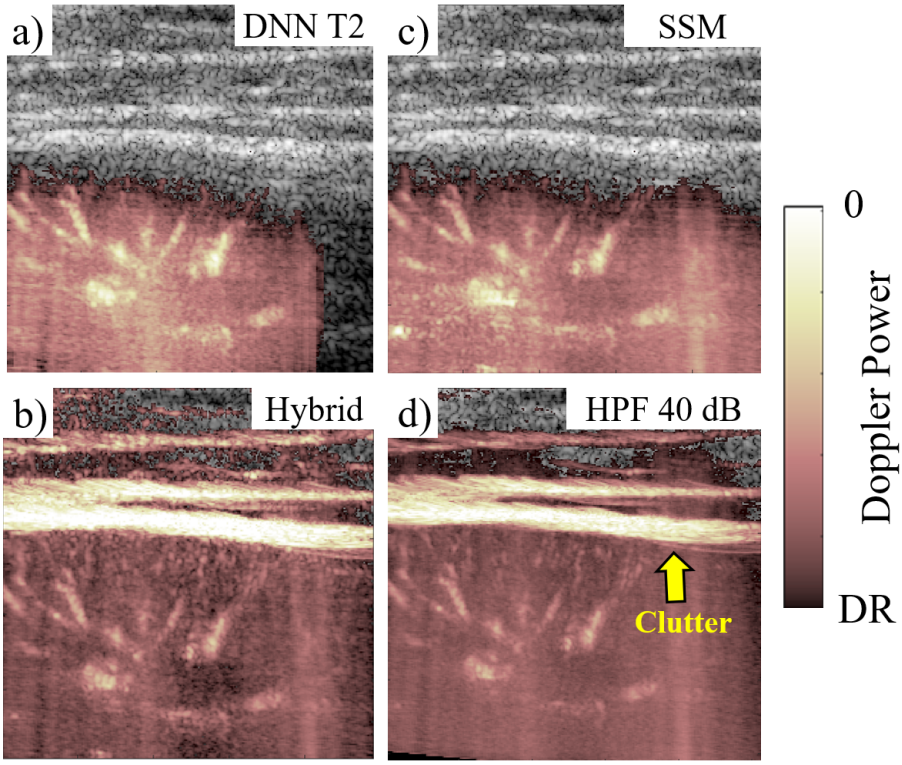

Figure 52: *a*) DNN with sensitivity threshold (0.6/1.0) *b*) Hybrid method *c*) SSM correlation method *d*) High-pass filter. SVD clutter filters are presented under 25 dB dynamic range[27]. The DNN filter rejects blocks with predominantly clutter, producing a similar result as the SSM filter although with more aggressive clutter removal.

Both the DNN filter and the SSM filter produce visually similar results and are able to identify and remove the bright reflections in the upper half of the image. In general, the DNN filter produces similar results although it appears to prioritize clutter removal over flow detection (adjusting sensitivity thresholds did not appear to improve this much).

[27]Saturation thresholds are individually set to best display the flow signal – this is necessitated by the large amount of high intensity clutter that is not removed by the high-pass filter and hybrid methods.

Further Comments

Implemented on a GPU, the DNN performs AUC predictions quite rapidly, outputting over 10 frames ($> 11,000$ total blocks) per second, significantly faster than the SSM block-fitting algorithm. At this stage, no rigorous optimization has been performed; we remark here that the inclusion of a neural network does not slow the SVD filtering procedure down by any significant degree[28] as other operations in filtering pipeline take far longer and by far are the *rate-limiting steps*. However, we remark that the acquisition of some of the statistical factors such as the local spatial correlation can take a significant amount of time. If different, optimized discriminating factors were used as inputs to the DNN, computational time could be further reduced. Although we have made no effort to check its feasibility, real time blockwise SVD filtering is likely not feasible for the current framework due to the significant computation required to compute the full rank SVD and the numerous discriminating parameters.

Currently, the AUC threshold cutoffs are relatively static and only weakly adaptive, based on the absolute values of the AUC curve. Adjustments to the AUC threshold can be made but this adjustment is global to all blocks in the frame, and not an individual adjustment to a specific block. Despite this, good results are obtained in most imaging scenarios as shown in the previous subsection. However, we note that when decreasing the sensitivity threshold the transition into a more cluttered image can happen abruptly due to mispredictions about clutter components in specific blocks. Figure 53[29] shows this effect in which the power Doppler image can be obtained with correctly set thresholds (pane a)), however, reducing the AUC sensitivity threshold below optimal induces an abrupt transition as clutter components in some blocks abruptly appear (pane b)). Good results are obtained before this effect appears but ideally the effects of dropping the sensitivity threshold would be gradual; it appears that a more adaptive method of setting the sensitivity threshold would be beneficial.

[28] Using Matlab's default *svd* function with the 'econ' flag and an 8 core AMD Ryzen 7 5800H CPU (Advanced Micro Devices, Santa Clara, CA, USA) it takes over 94 seconds (almost $100\times$ longer)to compute the SVD of 11,000 (80×80) $\times$ 128 data matrices (randomly initialized) using single precision floating point operations).

[29] As the largest components dominate a log-scaled image and clutter is much larger in magnitude than flow, these clutter blocks always dominate the power Doppler image.

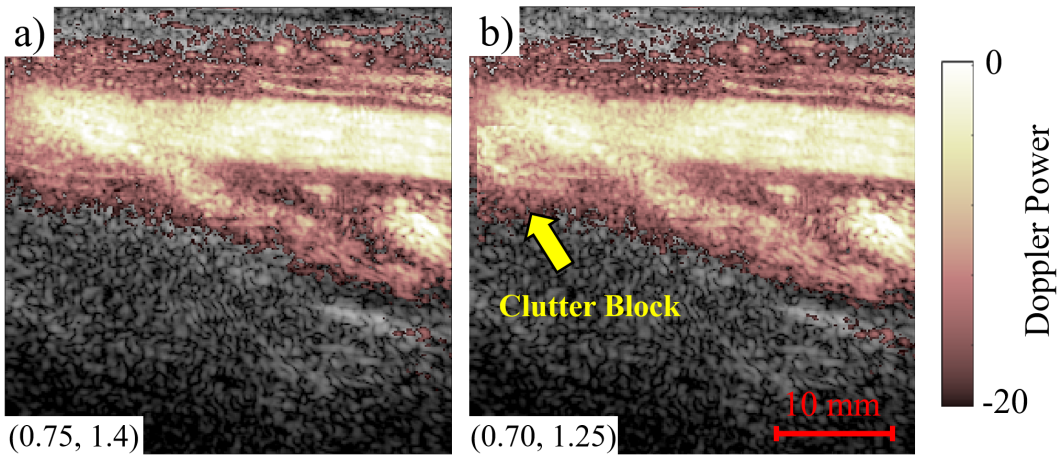

Figure 53: *a*) Femoral bifurcation (F03 Frame 205) where thresholds have been correctly set. The femoral bifurcation and vein are clearly visible. *b*) The same frame with the sensitivity threshold set too low. Rather than a gradual transition, a clutter block are abruptly appears. Results are presented using the same 20dB dynamic range. Sensitivity thresholds are shown in the lower left corner.

Finally, we note that while the proposed framework outperforms conventional literature techniques in most cases, the results obtained with the high pass filter can sometimes appear visually superior, specifically in terms of pixel-by-pixel 'sharpness'. As high-pass filtering is performed on a pixel-by-pixel basis, it is inherently capable of producing extremely sharp boundaries between clutter and flow. In contrast, the SVD filters used here take as input a block of pixels ($80 \times 80 = 6400$ pixels) and output combinations of singular image components, each of which contains an intensity value for each of the 6400 pixels. Although the SVD decomposes the input signal into mutually orthogonal basis vectors, the number of components is limited and often this decomposition is 'imperfect' (for our intended use) resulting in 'mixed' components containing both flow and tissue signals. In these cases, the vessel boundary tends to appear less sharp than would be desired; the flow signal appears to possess a blooming type effect around it. This is an inherent property of the decomposition when compared with a high-pass filter – the correlations between neighbouring pixels can be exploited to better identify flow at the cost of losing the ability to independently process each pixel.

Part V

Discussion and Future Work

Robust Flow Component Identification using Supervised Learning

The robust identification of the flow components in the data matrix after it has undergone the singular value decomposition extended across varied flow acquisitions is no easy task and has limited the use of SVD filters as a generic clutter filtering solution, requiring calibration or even different SVD filter designs to be applicable to specific imaging scenarios. Our proposed deeply connected neural network attempts to link the presence of flow to the reproduction of the area under the curve value obtained using ROC analysis on the segmented flow region by identifying what correlations in the network input statistics are representative of flow signals. If the components of the SVD with the highest reproduced AUC values are labeled as flow, accurate power Doppler images can be obtained that outperform current literature techniques. Furthermore, the proposed framework's ordinal prediction of the presence of flow in each component allows for the variation of the sensitivity threshold, allowing for 'adjusted' flow components that are more relevant to the diagnosis needed for the given imaging scenario in an easy to understand manner.

The proposed methodology was able to reproduce the AUC curve of flow blocks accurately (albeit with the scaling issue discussed in the previous chapter) while also able to identify blocks without any flow present (with an AUC of 0.5), generalizable to multiple *in-vivo* and *in-vitro* imaging scenarios across the cardiac cycle. The trained network was independently applied to the statistics of each of the 1156 blocks in each frame, which varied significantly, especially across the cardiac cycle, and consistently produced results that exceeded (or at least matched) current literature techniques. Some issues were encountered in reproducing AUC values during diastole but the resulting power Doppler images were comparable or better than those produced using other filtering techniques. In these cases, despite the statistical 'fingerprints' of flow being difficult to identify,[30] the proposed model, with access to multiple flow estimators, was able to better identify flow components.

Importantly, this work has demonstrated the utility of the component AUC curve, a target quantity obtainable through segmentation that can be used to ordinally rank the components of the SVD in terms of how their signal content lies within or outside the segmented flow region as a proxy for the amount of flow within each component. These values are a good approximation for the components of the decomposition that should be identified as flow components in any SVD clutter filtering algorithm. Critically, the methodology described here provides an accurate, reproducible *target* value that identifies flow components, a (somewhat) ground-truth,[31] that makes supervised learning techniques for SVD clutter filters feasible.

[30] Note that if the blood flow signal is weak, the flow signal will tend to be *mixed* in with clutter signals. As SVD filtering consists of picking some subset of elements of the decomposition (and the decomposition is unique) there is an upper limit on the performance of any SVD algorithm (at least without changing how the data matrix is created).

[31] The AUC is a good metric but not perfect – it has a number of limitations including the requirement of segmentation. Furthermore, it simply describes *spatially* where signal is present but contains no information about that signal content.

Limitations of the Proposed Framework

Although its stability was demonstrated in Figure 25, the *manual* segmentation required to create the component AUC curve will nevertheless be an issue of contention. As discussed in Chapter 3, automatic segmentation algorithms would encounter many difficulties and struggle to outperform manual segmentation across the large variety of training acquisitions (at least within the time constraints of this project). However, there is some contention on what *should* be segmented. As discussed in Chapter 3, we have chosen to segment based on where we know flow should exist (i.e. within the walls of blood vessels) even if no flow is currently observed in that region of the filtered flow image (during diastole in some vessels) – we are assuming that our current filtering technique is unable to isolate the flow signal but as there is blood within that region it does exist. We then assume that if no flow is present in the flow region, the spatial distribution of signal power should be such that there are no singular components with signals predominantly within the segmented flow region (noting that the interior of the lumen does not possess tissue scatterers) and the resulting AUC curves should be close to 0.5. However, we found that often the component AUC curves for these regions often possess significant noise; varying wildly around a value of 0.5. These swings may be caused by the identification and placement of spatially small unique correlations in the **very weak** identified flow signal in unique singular components by the decomposition. If these components were summed, the resulting flow signal would be more cohesive, however, spatially they only cover part of the segmented flow region at various levels of signal strength (see Figure 13) and so result in a very uneven inconsistent AUC curve[32]. This will be especially pronounced when included tissue signals (the flow signal is weak and mixing may occur) outside the segmented region are significant and inconsistently present in different singular components (component 15 in Figure 13 for example).

This large amount of noise and inconsistency in the AUC may make it difficult for the network to interpolate, leading to poor flow identification in blocks which do possess a more significant and identifiable flow signal. In other words, the larger amount of noise in the AUC values in blocks with no or little flow signal may be making it more difficult for the network to interpolate AUC values for networks with more moderate amounts of flow signal.

We have previously mentioned that a major difficulty with this project was the sheer amount of unique training data required to accurately train the DNN while still maintaining generalizability. Obtaining a labelled dataset for training is an unfortunate downside to supervised learning approaches. In this project (and in potential future works following this general procedure) this is a major issue as a tremendous amount of data acquisitions were needed. Many acquisitions were too noisy or low quality to be used and resulting in some difficulty in obtaining a sufficiently large training set. Compared to preliminary testing on a much reduced dataset (about half of the training and validation acquisitions listed in Table 4), the addition of more unique training and validation samples significantly improved the end clutter filtering performance of the proposed model. In particular, our DNN clutter

[32]This is a signal acquisition issue, not an inherent failing of the proposed method. The AUC curve produced through ROC analysis of the flow region still gives the optimal components that possess more signal within the flow region and less signal outside of it. Rather there is little flow signal at all.

filter performed significantly better in *in-vivo* than in the much simpler *in-vitro* flow phantom likely because only two *in-vitro* acquisitions were used in training.

The use of only two flow phantom training acquisitions further illustrates difficulties in obtaining a robust training set. If a flow phantom[53] is used for training under a carotid flow profile that phantom and any of its derivatives (even under different flow profiles) cannot be used in the validation or unseen datasets as the training and unseen samples would not be independent (flow of nearly every velocity is present in each region of the spiral at some point during the cardiac flow cycle). Here we also note that block statistics are very different *in-vitro* due to the use of homogeneous tissue-mimicking materials (for instance, the blood subspace is large and very clearly defined – see Figures 26c and 12) compared with *in-vivo* acquisitions resulting in poor generalizability between acquisition types. It is unsurprising perhaps, that the proposed model performed much better *in-vivo*.

A major advantage of our DNN clutter filtering model is that the difficulty in choosing appropriate thresholds, often very abstractly described such as the edges of 'a block of correlations in the SSM', is repackaged into a intuitive sensitivity threshold selection device. However, despite the AUC cutoff being much more intuitive and easy to adjust than the 'knobs' in other adaptive algorithms[33] it could be argued that the proposed method simply repackages the threshold identification problem into another threshold identification problem due to the many subtleties in choosing the optimal AUC cutoff.

This argument is valid. While we would argue that threshold adjustment in our model is intuitive and not inherently any different from tuning the cutoff frequency of a high-pass filter, it does represent another 'knob' that must be adjusted. Having two knobs, one that represents dynamic range and another sensitivity to flow makes things more complicated for the clinician. Furthermore, in our work we have not provided any mechanism to identify a default or useful 'presets' for various imaging scenarios. This would have to be addressed in future projects. Ultimately, there are advantages and disadvantages to algorithms that attempt to pick optimum thresholds and algorithms that put that choice in the hands of the user.

Addressing this current limitation by adaptively identifying a default AUC threshold (the user would still be able to set specific thresholds by applying an offset to this default value) would improve the accuracy of the current model. Perhaps this default should be individually identified for each block. An identified AUC threshold on a block-by-block basis could have the ability to address the magnitude scaling issues observed in Figures 34 and 35 where the DNN output possesses the same general structure as the input AUC value but is scaled up or down in magnitude.

A critical issue with using the AUC as a target is its independence of the relative amounts of flow and tissue pixels while the data matrix discriminating factors are heavily affected by the ratios of flow and tissue pixels present within the data matrix[34]. A final method of

[33]For instance, while the lower frequency threshold in the hybrid threshold estimator[50] appears to be intuitive, the inherent noisiness of the Doppler frequency curve (Figure 27) can cause alterations to the cutoff to cause erratic changes in the power Doppler image.

[34]These tend to be correlated but it may be useful to make a more strict distinction between blocks with large numbers of low intensity flow pixels and blocks with a few very high intensity flow pixels. The arbitrary

potentially improving network prediction may be to include the proportion of flow pixels in each block as an additional output of the network. This flow pixel proportion could perhaps be used as a weight in adjusting or setting an optimal AUC threshold as discussed in the previous paragraph.

Future Directions

Alternative Supervised Models

This project began as an attempt to improve on the results presented in Refs. [30] and [31] with the key idea to use repeated density-based clustering attempts (varying the hyper-parameters of the DBSCAN algorithm [96]) to identify the clustered points most likely to form part of the blood cluster. Although this was an improvement over the results obtained through K-Means clustering, it was by no means consistent across different acquisitions. Once the method of using the component AUC curve for supervised learning approaches was developed, preliminary investigations of several other supervised learning approaches such as a linear regression model [97] and a random forest classifier [98] were performed in parallel with DNN development. A brief investigation (results not shown here) found that these models performed adequately but not amazingly.

It would be interesting to explore other supervised learning approaches to this same problem. By no means is a neural network the guaranteed optimal solution to the problem of predicting the AUC values of each singular component. Our preliminary investigations into alternative clustering and classification models was quite brief and limited in scope and it is quite likely that other techniques could yield comparable or better results, especially if methodology changes are made. Furthermore, while the area under the receiver operator characteristic curve has shown itself to be a useful identifier of flow within the components of the decomposition, it possesses several major flaws and could be potentially replaced in this study with a hypothetical superior target metric.

Improving or Not Using Discriminating Factors

A natural extension of this work would be to further refine the chosen discriminating parameters, removing those that are computationally expensive to compute (such as the *local spatial correlation*) and add new quantities that further improve the robustness of the AUC estimate. There is a significant amount of information in the U, Δ, and V data matrices and their derived quantities; perhaps another robust estimator could be derived from the structure of the power spectral density plots of each singular component (Figure 28). Further analysis could also be performed on the singular image components.

While additional and improved discriminating factors could be proposed, each additional factor represents additional computational demand and ultimately slower filtering. As each

10% pixel cutoff for AUC calculation was a weak attempt to address this issue.

discriminating factor is simply some quantity derived from the SVD data matrices, why not simply use the raw U, Δ, and V data matrices as the input to the network? This would remove the need to compute discriminating factors, reduce preprocessing and potentially allow more calculations to be performed on the GPU which is significantly more powerful than the CPU.

Some preliminary testing using the raw data matrices as input was explored in this work. The principle image components M_i of blocks (several are shown in Figure 13) $M_i = \Delta_i U_i V_i^\star$ were calculated and formed into a 3D matrix (effectively a stack of the 2D principle image components). A 3D *convolutional* neural network [99][100] was then proposed to process input of this form. Convolutional neural networks have been used extensively in the field of ultrasound, for b-mode imaging [101][102][103], channel count interpolation [104], and clutter filtering [105][106], etc.[35]. However, the liability of using the raw SVD information is its size and/or the time taken to compute the SVD. In the proposed framework the discriminating factors are precomputed and stored. With 6 discriminating factors and an ensemble size of 128, only 768 elements need to be stored per block. Stored as 32-bit floating point values, it takes only 3 KB ($6 \times 768 \times 4$ bytes each = 3072 bytes) to store the inputs for each block (potentially less with compression). If the principle image component stack is used as input, over 3.2 MB are needed to store a given block ($80 \times 80 \times 128 \times 4$ = 3,276,800 bytes), a ratio of about 1000× more. The storage and memory requirements of storing approximately 190,000 training and validation samples is excessive and although perhaps addressable on some systems, was found to be unfeasible for this project. Alternatively, it may be possible to compute the SVD on the fly using optimized computational techniques while still achieving somewhat reasonable training times (See following sections). Although a difficult challenge, using the raw data matrices as input would greatly simplify the processing pipeline and provide significantly more raw information for network inference, representing a logical next step for this project.

Data Matrix Adjustments and Alternatives to the SVD

The singular value decomposition of a given data matrix is unique. This means that every potential SVD filtering algorithm that takes as input the same U, Δ, and V data matrices should ultimately identify the same flow components of the decomposition. While this property controls the scope of the problem, it produces an upper bound on the performance of any equivalent SVD algorithm (i.e. for any definition of 'optimal flow component' there is a set of components that will produce the best filtering performance) that represents an upper bound to filtering performance. In other words the potential mixing of the clutter and flow signals **within** the components of the decomposition cannot be addressed by any filtering technique. This limitation motivates changes to the input data matrix or the use of alternative or modified decompositions.

[35]Here networks have generally been used to reproduce the filtered images produced using known clutter filtering techniques. These algorithms almost always attempt to reproduce known results (i.e. they are trained using clutter filtered images) and therefore function as an easy and efficient reproduction mechanism rather than producing fundamental improvements in filtering performance. (The issue of the ground truth flow signal not being available presents inself).

Experimental results obtained using the *higher order singular value decomposition* (HOSVD), an extension of the SVD [107] that takes as input the 3D tensor cineloop (without repackaging data into a Casorati matrix) and outputs a central tensor and three 2D matrices, has shown improved clutter filtering performance[36] over conventional SVD filtering using the 2D Casorati matrix [108][109]. Furthermore as discussed in Section II the use of other decompositions and blind-source separation algorithms abound in the literature. It would be interesting to see if the current framework could be extended in these directions.

Towards Real-time Implementation and Clinical Use

Currently to the best of our knowledge, no clinical system uses SVD filtering techniques for flow detection. A major reason to this is firstly, the arbitrariness and difficulty in robustly identifying flow components and secondly, the tremendous computational demands required to perform blockwise SVD in real time, especially at a time when the major focus is towards more portable systems.

It is possible to perform the singular value decomposition on the much more powerful GPU [110][111][37]. However, it is difficult to implement the decomposition efficiently on highly parallel GPUs and most algorithms do not exhibit significant speedups over implementations on multicore CPUs for small matrices [111][112][38]. Nevertheless, an efficient GPU implementation would represent a large step toward complex SVD filtering techniques being implementable in real time.

Many SVD filters in the literature have embraced numerical 'approximations' to reduce computational requirements (such as randomized spatial downsampling [49][56]). Another common technique in the literature is to perform a limited rather than full rank decomposition, as the higher rank singular components rarely contain flow signal. Ref. [56] showed that calculating only an approximation to the first m singular components required significantly less computational time (up to 6× faster if only a few ranks needed to be computed) and delivered similar results to what was obtained using the full SVD. These calculations can also be efficiently implemented on the GPU [113].

The current framework requires the full rank singular value decomposition to be performed to identify flow components. One of the reasons for this choice was the realization that in certain *in-vitro* scenarios, flow may be found even in the high rank components of the decomposition (i.e. see Figures 26c and 12a) where typically only noise signals are found. In some *in-vivo* acquisitions we also note the presence of flow in higher rank components of the decomposition (i.e. Figures 34 and 35). Nevertheless, if this issue could be addressed and computational requirements decreased, it would represent a major step towards real-time SVD clutter filtering. Making use of these computational optimizations such as reduced

[36] In both cases flow thresholds were manually tuned.

[37] Present on current versions of Matlab.

[38] In our testing using Matlab, a RTX 3060 Mobile proved to be significantly slower (1.5 - 2× depending on matrix dimensions) than an 8 core AMD Ryzen 7 5800H mobile CPU for randomly initialized single precision matrices of sizes 6400 × 128 - the dimensions of the data matrices of 80× 80 pixel blocks with an ensemble size of 128.

rank SVD calculation and random spatial downsampling, SVD clutter filtering has been demonstrated in real time [49] – a real time implementation of the current framework would represent a major step towards widespread clinical adoption.

References

[1] A. Carovac, F. Smajlovic, and D. Junuzovic. Application of ultrasound in medicine. *Acta Informatica Medica*, 19(3):168, 2011.

[2] S. Sippel, K. Muruganandan, A. Levine, and S. Shah. Use of ultrasound in the developing world. *International Journal of Emergency Medicine*, 4(1):1–11, 2011.

[3] G. Ter Haar. Therapeutic applications of ultrasound. *Progress in Biophysics and Molecular Biology*, 93(1-3):111–129, 2007.

[4] T. Kiserud, G. Piaggio, G. Carroli, M. Widmer, J. Carvalho, L. Neerup Jensen, D. Giordano, J. Cecatti, H. Abdel Aleem, and S. Talegawkar. The world health organization fetal growth charts: a multinational longitudinal study of ultrasound biometric measurements and estimated fetal weight. *PLoS Medicine*, 14(1):e1002220, 2017.

[5] Z. Alfirevic and J. Neilson. Doppler ultrasound for fetal assessment in high risk pregnancies. *Cochrane Database of Systematic Reviews*, (4), 1996.

[6] J David Spence. Ultrasound measurement of carotid plaque as a surrogate outcome for coronary artery disease. *The American Journal of Cardiology*, 89(4):10–15, 2002.

[7] S. Nicholls, A. Hsu, K. Wolski, B. Hu, O. Bayturan, A. Lavoie, K. Uno, E. Tuzcu, and S. Nissen. Intravascular ultrasound-derived measures of coronary atherosclerotic plaque burden and clinical outcome. *Journal of the American College of Cardiology*, 55(21):2399–2407, 2010.

[8] R. Wyman, M. Mays, P. McBride, and J. Stein. Ultrasound-detected carotid plaque as a predictor of cardiovascular events. *Vascular Medicine*, 11(2):123–130, 2006.

[9] A. Nair, B. Kuban, E. Tuzcu, P. Schoenhagen, S. Nissen, and G. Vince. Coronary plaque classification with intravascular ultrasound radiofrequency data analysis. *Circulation*, 106(17):2200–2206, 2002.

[10] T. Yamamoto, Y. Ogasawara, A. Kimura, H. Tanaka, O. Hiramatsu, K. Tsujioka, M. Lever, K. Parker, C. Jones, and C. Caro. Blood velocity profiles in the human renal artery by Doppler ultrasound and their relationship to atherosclerosis. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 16(1):172–177, 1996.

[11] Y. Qiu, Y. Dong, F. Mao, Q. Zhang, D. Yang, K. Chen, S. Shi, D. Zuo, X. Tian, and L. Yu. High-frame rate vector flow imaging technique: Initial application in evaluating the hemodynamic changes of carotid stenosis caused by atherosclerosis. *Frontiers in Cardiovascular Medicine*, 8:119, 2021.

[12] A. Lensing, P. Prandoni, D. Brandjes, P. Huisman, M. Vigo, G. Tomasella, J. Krekt, J. ten Cate, M. Huisman, and H. Büller. Detection of deep-vein thrombosis by real-time b-mode ultrasonography. *New England Journal of Medicine*, 320(6):342–345, 1989.

[13] M. Mattos, G. Londrey, D. Leutz, K. Hodgson, D. Ramsey, L. Barkmeier, E. Stauffer, D. Spadone, and D. Sumner. Color-flow duplex scanning for the surveillance and diagnosis of acute deep venous thrombosis. *Journal of Vascular Surgery*, 15(2):366–376, 1992.

[14] N. Bravo-Valenzuela, L. Rocha, L. Nardozza, and E. Júnior. Fetal cardiac arrhythmias: Current evidence. *Annals of Pediatric Cardiology*, 11(2):148, 2018.

[15] S. Yuan and Z. Xu. Fetal arrhythmias: prenatal evaluation and intrauterine therapeutics. *Italian Journal of Pediatrics*, 46(1):1–7, 2020.

[16] S. Bjærum, H. Torp, and K. Kristoffersen. Clutter filter design for ultrasound color flow imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 49(2):204–216, 2002.

[17] A. Oglat, M. Matjafri, N. Suardi, M. Oqlat, M.and Abdelrahman, and A. Oqlat. A review of medical Doppler ultrasonography of blood flow in general and especially in common carotid artery. *Journal of Medical Ultrasound*, 26(1):3, 2018.

[18] S Bjaerum, H Torp, and K. Kristoffersen. Clutter filters adapted to tissue motion in ultrasound color flow imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 49(6):693–704, 2002.

[19] H. Torp. Clutter rejection filters in color flow imaging: A theoretical approach. *IEEE tTansactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 44(2):417–424, 1997.

[20] D. Kruse and K. Ferrara. A new high resolution color flow system using an eigendecomposition-based adaptive filter for clutter rejection. *IEEE transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 49(10):1384–1399, 2002.

[21] C. Demene, J. Baranger, M. Bernal, C. Delanoe, S. Auvin, V. Biran, M. Alison, J. Mairesse, E. Harribaud, M. Pernot, and T. Tanter. Functional ultrasound imaging of brain activity in human newborns. *Science Translational Medicine*, 9(411), 2017.

[22] B.F. Osmanski, S. Pezet, A. Ricobaraza, Z. Lenkei, and M. Tanter. Functional ultrasound imaging of intrinsic connectivity in the living rat brain with high spatiotemporal resolution. *Nature Communications*, 5(1):1–14, 2014.

[23] C. Demené, J. Robin, A. Dizeux, B. Heiles, M. Pernot, M. Tanter, and F. Perren. Transcranial ultrafast ultrasound localization microscopy of brain vasculature in patients. *Nature Biomedical Engineering*, 5(3):219–228, 2021.

[24] D. Ghosh, F. Xiong, S. Sirsi, P. Shaul, R. Mattrey, and K. Hoyt. Toward optimization of *in-vivo* super-resolution ultrasound imaging using size-selected microbubble contrast agents. *Medical Physics*, 44(12):6304–6313, 2017.

[25] D. Ghosh, J. Peng, K. Brown, S. Sirsi, C. Mineo, P. Shaul, and K. Hoyt. Super-resolution ultrasound imaging of skeletal muscle microvascular dysfunction in an animal model of type 2 diabetes. *Journal of Ultrasound in Medicine*, 38(10):2589–2599, 2019.

[26] A. Yu and R. Cobbold. Single-ensemble-based eigen-processing methods for color flow imaging-part i. the Hankel-SVD filter. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 55(3):559–572, 2008.

[27] A. Yu and L. Lostakken. Eigen-based clutter filter design for ultrasound color flow imaging: A review. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 57(5):1096–1111, 2010.

[28] C. Demené, T. Deffieux, M. Pernot, B. Osmanski, V. Biran, J. Gennisson, L. Sieu, S. Franqui, J. Correas, and M. Cohen. Spatiotemporal clutter filtering of ultrafast ultrasound data highly increases Doppler and fUltrasound sensitivity. *IEEE Transactions on Medical Imaging*, 34(11):2271–2285, 2015.

[29] J. Baranger, B. Arnal, F. Perren, O. Baud, M. Tanter, and C. Demené. Adaptive spatiotemporal SVD clutter filtering for ultrafast doppler imaging using similarity of spatial singular vectors. *IEEE Transactions on Medical Imaging*, 37(7):1574–1586, 2018.

[30] S. Waraich, A. Chee, D. Xiao, B. Yiu, and A. Yu. Auto SVD clutter filtering for US Doppler imaging using 3d clustering algorithm. In *International Conference on Image Analysis and Recognition*, pages 473–483. Springer, 2019.

[31] S. A. Waraich. Robust eigen-filter design for ultrasound flow imaging using a multivariate clustering. Master's thesis, University of Waterloo, 2020.

[32] J. Chubak, G. Pocobelli, and N. Weiss. Tradeoffs between accuracy measures for electronic health care data algorithms. *Journal of Clinical Epidemiology*, 65(3):343–349, 2012.

[33] M. Mozumi, R. Nagaoka, and H. Hasegawa. Utilization of singular value decomposition in high-frame-rate cardiac blood flow imaging. *Japanese Journal of Applied Physics*, 58(SG):SGGE02, 2019.

[34] M. Mozumi, R. Nagaoka, and H. Hasegawa. Singular value decomposition filtering in high-frame-rate cardiac vector flow imaging. *Bulletin of Electrical Engineering and Informatics*, 9(1):171–179, 2020.

[35] F. Vignon, Jun S. Shin, S. Huang, and J. Robert. Adaptive ultrasound clutter rejection through spatial eigenvector filtering. In *2017 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4. IEEE, 2017.

[36] J. Voorneveld, P. Kruizinga, H. Vos, Frank J. Gijsen, E. Jebbink, A. Van Der Steen, N. De Jong, and J. Bosch. Native blood speckle vs ultrasound contrast agent for particle image velocimetry with ultrafast ultrasound-in vitro experiments. In *2016 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4. IEEE, 2016.

[37] D. Cheung, H. Chiu, L. Zhang, C. Hu, K. Shung, and A. Yu. Adaptive clutter filter design for micro-ultrasound color flow imaging of small blood vessels. In *2010 IEEE International Ultrasonics Symposium (IUS)*, pages 1206–1209. IEEE, 2010.

[38] R. Van Sloun and Y. Cohen, R.and Eldar. Deep learning in ultrasound imaging. *Proceedings of the IEEE*, 108(1):11–29, 2019.

[39] H. Wang, S. Gao, M. Mozumi, M. Omura, R. Nagaoka, and H. Hasegawa. Preliminary investigation on clutter filtering based on deep learning. *Japanese Journal of Applied Physics*, 60(SD):SDDE21, 2021.

[40] O. Solomon, R. Cohen, Y. Zhang, Y. Yang, Q. He, J. Luo, R. van Sloun, and Y. Eldar. Deep unfolded robust PCA with application to clutter suppression in ultrasound. *IEEE Transactions on Medical Imaging*, 39(4):1051–1063, 2019.

[41] M. Tabassian, X. Hu, B.a Chakraborty, and J. D'hooge. Clutter filtering using a 3D deep convolutional neural network. In *2019 IEEE International Ultrasonics Symposium (IUS)*, pages 2114–2117. IEEE, 2019.

[42] R. Cobbold. *Foundations of Biomedical Ultrasound. Oxford University Press*, 2006.

[43] G. Moltaldo, M. Tanter, C. Clarisse, N. Benech, and N. Fink. Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 56(3):489–506, 2009.

[44] K. Kasai, C.and Namekawa, A. Koyano, and R. Omoto. Real-time two-dimensional blood flow imaging using an autocorrelation technique. *IEEE Transactions on Sonics and Ultrasonics*, 32(3):458–464, 1985.

[45] L. Bohs, B. Geiman, M. Anderson, S. Gebhart, and G. Trahey. Speckle tracking for multi-dimensional flow estimation. *Ultrasonics*, 38(1-8):369–375, 2000.

[46] A. Swillens, P. Segers, H. Torp, and L. Lovstakken. Two-dimensional blood velocity estimation with ultrasound: Speckle tracking versus crossed-beam vector Doppler based on flow simulations in a carotid bifurcation model. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 57(2):327–339, 2010.

[47] S. Nyrnes, S. Fadnes, M. Wigen, L. Mertens, and L. Lovstakken. Blood speckle-tracking based on high–frame rate ultrasound imaging in pediatric cardiology. *Journal of the American Society of Echocardiography*, 33(4):493–503, 2020.

[48] V. Klema and A. Laub. The singular value decomposition: Its computation and some applications.

[49] U.W. Lok, P. Song, R. Trzasko, J.D. Daigle, E.A. Borisch, P. Huang, C. Gong, S. Tang, W. Ling, and Chen. S. Real time SVD-based clutter filtering using randomized singular value decomposition and spatial downsampling for micro-vessel imaging on a Verasonics ultrasound system. *Ultrasonics*, 107:106163, 2020.

[50] P. Song, A. Manduca, J.D. Trzasko, and S. Chen. Ultrasound small vessel imaging with block-wise adaptive local clutter filtering. *IEEE Transactions on Medical Imaging*, 36(1):251–262, 2016.

[51] C. Chang, P. Chen, H. Huang, and C. Huang. *In-vivo* visualization of vasculature in adult zebrafish by using high-frequency ultrafast ultrasound imaging. *IEEE Transactions on Biomedical Engineering*, 66(6):1742–1751, 2018.

[52] H. Ikeda, S. Yoshizawa, M. Maeda, S. Umemura, and Y. Saijo. Blood flow imaging using singular value decomposition filter during high-intensity focused ultrasound exposure. *Japanese Journal of Applied Physics*, 58(SG):SGGE15, 2019.

[53] B. Yiu and A. Yu. Spiral flow phantom for ultrasound flow imaging experimentation. *IEEE transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 64(12):1840–1848, 2017.

[54] S. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.

[55] J. Hartigan and M. Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108, 1979.

[56] P. Song, J.D. Trzasko, A. Manduca, B. Qiang, R. Kadirvel, D.F. Kallmes, and S. Chen. Accelerated singular value-based ultrasound blood flow clutter filtering with randomized singular value decomposition and randomized spatial downsampling. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 64(4):706–716, 2017.

[57] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users' Guide. Society for Industrial and Applied Mathematics*, Philadelphia, PA, third edition, 1999.

[58] J. Tierney, J.and Baker, D. Brown, D. Wilkes, and B. Byram. Independent component-based spatiotemporal clutter filtering for slow flow ultrasound. *IEEE Transactions on Medical Imaging*, 39(5):1472–1482, 2019.

[59] S. Zobly and Y. Kadah. A new clutter rejection technique for doppler ultrasound signal based on principal and independent component analyses. In *2012 Cairo International Biomedical Engineering Conference (CIBEC)*, pages 56–59. IEEE, 2012.

[60] G. Wahyulaksana, L. Wei, J. Schoormans, J. Voorneveld, A. Van der Steen, N.o De Jong, and H. Vos. Independent component analysis filter for small vessel contrast imaging during fast tissue motion. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 69(7):2282–2292, 2022.

[61] M. Ashikuzzaman, C. Belasso, M. Kibria, A. Bergdahl, C. Gauthier, and H. Rivaz. Low rank and sparse decomposition of ultrasound color flow images for suppressing clutter in real-time. *IEEE Transactions on Medical Imaging*, 39(4):1073–1084, 2019.

[62] M. Bayat and M. Fatemi. Concurrent clutter and noise suppression via low rank plus sparse optimization for non-contrast ultrasound flow Doppler processing in microvasculature. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1080–1084. IEEE, 2018.

[63] R. Wildeboer, F. Sammali, R. Van Sloun, Y. Huang, P. Chen, M. Bruce, C. Rabotti, S. Shulepov, G. Salomon, and B. Schoot. Blind source separation for clutter and noise suppression in ultrasound imaging: Review for different applications. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 67(8):1497–1512, 2020.

[64] I. Goodfellow, Y. Bulatov, J. Ibarz, S. Arnoud, and V. Shet. Multi-digit number recognition from street view imagery using deep convolutional neural networks. *arXiv preprint arXiv:1312.6082*, 2013.

[65] J. Anderson. *An introduction to neural networks. MIT Press*, 1995.

[66] R. Hecht-Nielsen. Theory of the backpropagation neural network. In *Neural Networks for Perception*, pages 65–93. Elsevier, 1992.

[67] D. Leonov, N. Kulberg, V. Fin, V. Podmoskovnaya, L. Ivanova, A. Shipaeva, A. Vladzimirskiy, and S. Morozov. Comparison of filtering techniques in ultrasound color flow imaging. *Biomedical Engineering*, 53(2):97–101, 2019.

[68] A. Chee and A. Yu. Receiver-operating characteristic analysis of eigen-based clutter filters for ultrasound color flow imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 65(3):390–399, 2017.

[69] K. Hajian-Tilaki. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine*, 4(2):627, 2013.

[70] W. Youden. Index for rating diagnostic tests. *Cancer*, 3(1):32–35, 1950.

[71] K. Saini, M. Dewal, and M. Rohit. Ultrasound imaging and image segmentation in the area of ultrasound: a review. *International Journal of Advanced Science and Technology*, 24, 2010.

[72] C. Loizou. A review of ultrasound common carotid artery image and video segmentation techniques. *Medical & Biological Engineering & Computing*, 52(12):1073–1093, 2014.

[73] J. Au, B. Yiu, and A. Yu. Case studies in physiology: Visualization of blood recirculation in a femoral artery "trifurcation" using ultrasound vector flow imaging. *Journal of Applied Physiology*, 127(6):1809–1813, 2019.

[74] H. Nahas, J. Au, T. Ishii, B. Yiu, A.. Chee, and A. Yu. A deep learning approach to resolve aliasing artifacts in ultrasound color flow imaging. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 67(12):2615–2628, 2020.

[75] A. Chee, T. Ishii, B.Y. Yiu, and A.C.H. Yu. Helical toroid phantom for 3d flow imaging investigations. *Physics in Medicine & Biology*, 66(4):045029, 2021.

[76] A. Chee, C. Ho, and A. Yiu, B.and Yu. Walled carotid bifurcation phantoms for imaging investigations of vessel wall motion and blood flow dynamics. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 63(11):1852–1864, 2016.

[77] P. Refaeilzadeh, L. Tang, and H. Liu. Cross-validation. *Encyclopedia of Database Systems*, 5:532–538, 2009.

[78] Y. Xu and R. Goodacre. On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. *Journal of Analysis and Testing*, 2(3):249–262, 2018.

[79] C. Bishop. *Neural networks for pattern recognition. Oxford University Press*, 1995.

[80] J. Sola and J. Sevilla. Importance of input data normalization for the application of neural networks to complex industrial problems. *IEEE Transactions on Nuclear Science*, 44(3):1464–1468, 1997.

[81] L. Huang, J. Qin, Y. Zhou, F. Zhu, L. Liu, and L. Shao. Normalization techniques in training DNNs: Methodology, analysis and application. *arXiv preprint arXiv:2009.12836*, 2020.

[82] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.

[83] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[84] Y. Desailly, A.M. Tissier, J.M. Correas, F. Wintzenrieth, M. Tanter, and O. Couture. Contrast enhanced ultrasound by real-time spatiotemporal filtering of ultrafast images. *Physics in Medicine & Biology*, 62(1):31–42, 2016.

[85] A. Rodriguez-Molares, O.M. Rindal, J. D'hooge, S.E. Måsøy, A. Austeng, M.A.L. Bell, and H. Torp. The generalized contrast-to-noise ratio: a formal definition for lesion detectability. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 67(4):745–759, 2019.

[86] L. Lovstakken, S. Bjaerum, and H. Torp. Optimal velocity estimation in ultrasound color flow imaging in presence of clutter. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 54(3):539–549, 2007.

[87] K. Pearson. X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302):157–175, 1900.

[88] J. Lee Rodgers and W. Nicewander. Thirteen ways to look at the correlation coefficient. *The American Statistician*, 42(1):59–66, 1988.

[89] H. Gerhard, F. Wichmann, and M. Bethge. How sensitive is the human visual system to the local statistics of natural images? *PLoS Computational Biology*, 9(1):e1002873, 2013.

[90] D. Field, A. Hayes, and R. Hess. Contour integration by the human visual system: evidence for a local "association field". *Vision Research*, 33(2):173–193, 1993.

[91] Z. Wang, A. Bovik, and L. Lu. Why is image quality assessment so difficult? In *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages IV–3313. IEEE, 2002.

[92] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[93] G. Chen, Y. Shen, F. Yao, P. Liu, and Y. Liu. Region-based moving object detection using SSIM. In *2015 4th International Conference on Computer Science and Network Technology (ICCSNT)*, volume 1, pages 1361–1364. IEEE, 2015.

[94] C. Ho, A. Chee, B. Yiu, A. Tsang, K. Chow, and Y. Alfred. Wall-less flow phantoms with tortuous vascular geometries: Design principles and a patient-specific model fabrication example. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 64(1):25–38, 2016.

[95] K. Ramnarine, D. Nassiri, P. Hoskins, and J. Lubbers. Validation of a new blood-mimicking fluid for use in Doppler flow test objects. *Ultrasound in Medicine & Biology*, 24(3):451–459, 1998.

[96] M. Ester, H. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *1996 Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, volume 96, pages 226–231, 1996.

[97] D. Freedman. *Statistical Models: Theory and Practice*. cambridge university press, 2009.

[98] T. Ho. Random decision forests. In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, volume 1, pages 278–282. IEEE, 1995.

[99] K. Fukushima and S. Miyake. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and Cooperation in Neural Nets*, pages 267–285. Springer, 1982.

[100] J. Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61:85–117, 2015.

[101] B. Luijten, R. Cohen, F. de Bruijn, H. Schmeitz, M. Mischi, Y. Eldar, and R. van Sloun. Adaptive ultrasound beamforming using deep learning. *IEEE Transactions on Medical Imaging*, 39(12):3967–3978, 2020.

[102] A. Nair, M. Gubbi, T. Tran, A. Reiter, and M. Bell. A fully convolutional neural network for beamforming ultrasound images. In *2018 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4. IEEE, 2018.

[103] S. Khan, J. Huh, and J. Ye. Adaptive and compressive beamforming using deep learning for medical ultrasound. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 67(8):1558–1572, 2020.

[104] D. Xiao, W. Pitman, B. Yiu, A. Chee, and Y. Alfred. Minimizing image quality loss after channel count reduction for plane wave ultrasound via deep learning inference. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 69(10):2849–2861, 2022.

[105] M. Tabassian, X. Hu, B. Chakraborty, and J. D'hooge. Clutter filtering using a 3D deep convolutional neural network. In *2019 IEEE International Ultrasonics Symposium (IUS)*, pages 2114–2117. IEEE, 2019.

[106] K. Brown, D. Ghosh, and K. Hoyt. Deep learning of spatiotemporal filtering for fast super-resolution ultrasound imaging. *IEEE transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 67(9):1820–1829, 2020.

[107] L. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.

[108] M. Bayat, A. Alizad, and M. Fatemi. Multi-rate higher order singular value decomposition for enhanced non-contrast ultrasound doppler imaging of slow flow. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1178–1181. IEEE, 2018.

[109] K. Ozgun and B. Byram. A channel domain higher-order SVD clutter rejection filter for small vessel ultrasound imaging. In *2020 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4. IEEE, 2020.

[110] S. Lahabar and P. Narayanan. Singular value decomposition on GPU using CUDA. In *2009 IEEE International Symposium on Parallel & Distributed Processing*, pages 1–10. IEEE, 2009.

[111] M. Gates, S. Tomov, and J. Dongarra. Accelerating the SVD two stage bidiagonal reduction and divide and conquer using GPUs. *Parallel Computing*, 74:3–18, 2018.

[112] T. Dong, A. Haidar, S. Tomov, and J. Dongarra. Accelerating the SVD bidiagonalization of a batch of small matrices using GPUs. *Journal of Computational Science*, 26:237–245, 2018.

[113] Y. Lu, I. Yamazaki, F. Ino, Y. Matsushita, S. Tomov, and J. Dongarra. Reducing the amount of out-of-core data access for gpu-accelerated randomized SVD. *Concurrency and Computation: Practice and Experience*, 32(19):e5754, 2020.