# Deep Spatiotemporal Phase Unwrapping of Phase-Contrast MRI Data

Jiacheng Jason He[1], Christopher Sandino[1], David Zeng[1], Shreyas Vasanawala[1,2], and Joseph Cheng[2]

[1]Electrical Engineering, Stanford University, Stanford, CA, United States, [2]Radiology, Stanford University, Stanford, CA, United States

## Synopsis

**This work demonstrates the advantage temporal information provides for deep phase unwrapping of phase-contrast MRI data. Using a patch-based, three-dimensional ResNet architecture, our model performs better than state-of-the-art single-step algorithms. Our deep spatiotemporal phase unwrapping model continues the quest to lower $V_{enc}$ values to increase dynamic range and velocity-to-noise ratio (VNR) of 4D flow data by providing a robust method for phase unwrapping.**

## Introduction

Phase-contrast MRI encodes velocity information into the phase of images, giving clinicians access to more data for stronger diagnoses. The primary limiting factor for optimal data dynamic range and velocity-to-noise ratio (VNR) is the velocity encoding ($V_{enc}$) parameter, which is the maximum velocity (cm/s) that can be encoded in phase, from $-\pi$ to $\pi$. Any velocities that exceed the $V_{enc}$ will cause phase wrapping. A common clinical strategy is to overestimate the $V_{enc}$, since it can vary dramatically with anatomy differences and various pathologies, at the cost of lower velocimetric dynamic range and VNR. Meanwhile, a robust phase unwrapping algorithm would allow clinicians to choose a low $V_{enc}$ and obtain useable images with high dynamic range and VNR. Existing single-step algorithms[1,2] often only perform well on images with up to one cycle of phase wrapping. Building on existing deep phase unwrapping work,[3] we propose to use a modified patch-based 3D ResNet,[4] which exploits spatiotemporal correlations in phase-contrast data to improve phase unwrapping accuracy, especially for data with severe wrapping.

## Methods

With IRB approval, 4D flow data for 10 patients were retrospectively collected: 7 were used for training, 1 for validation, and 2 for testing. Data were acquired with a contrast-enhanced velocity-encoded volumetric RF-spoiled GRE sequence. Each set had approximately 300 slices in the sagittal axis, but the middle 219 were chosen to eliminate slices that do not cover the heart. Preprocessing was necessary because the data we started with was originally sampled at a $V_{enc}$ of 250 with no wrapping. We first found the phase difference between two serial flow-encoded images and concatenated that with the corresponding magnitude images. We then simulated wrapping by shifting, stretching, and normalizing the data to create wrapped images at random $V_{enc}$ values between 25 and 75. During training, to minimize the amount of noise seen by the model, the images were randomly cropped into patches in the spatial domain, to one-fifth the original size. This decision was made because noise phase is extremely different from anatomical phase, which can confuse the model when training. To increase the likelihood of cropping on a phase wrapped patch, the magnitude images were used to derive a probability density function for the crop selection. We hypothesized that magnitude image patches with higher signal intensity--which in our contrast-enhanced dataset can be correlated with blood--are more likely to have high flow and result in wrapping after $V_{enc}$ adjustments. All 30 timesteps were used. During validation and testing, the images are not patched to evaluate performance on full images.

The network architecture is shown in Figure 1, based on a ResNet model. Three-dimensional convolution layers were used, since the inputs were of size Nx20x30x30 at training time and size Nx100x150x30 at validation and testing time. The training lasted over 135 hours, training for 1000 epochs. A long training time was necessary due to the 12,320,000 parameters that needed to be updated.

## Results

Figure 2 compares the qualitative appearances of our model outputs with the high $V_{enc}$ ground truth, wrapped image, and Laplacian algorithm[1] output, which is used as a baseline. The animation scrolls through 5 $V_{enc}$ values, linearly spaced from 25 to 75. For quantitative comparison, we calculated the pixel-wise $L_1$ difference in the manually segmented aorta for each $V_{enc}$, shown in Figure 3.

## Discussion & Conclusions

Our deep phase unwrapping model clearly demonstrates a more robust method for unwrapping 4D flow MRI data. Since data is unwrapped slice-by-slice, this method can easily be generalized to 2D phase-contrast data. From the qualitative comparison in Figure 2, it is clear that our model outperforms the Laplacian baseline at lower $V_{enc}$ settings, which corresponds with our hypothesis, since the Laplacian algorithm is optimized for single-cycles of wrapping or less. The data shown in Figure 3 reinforces the observed qualitative results. One current limitation of our model, though, is the inability to accurately model eddy current-related background phase errors and flow-related signal dephasing which become more severe at lower $V_{enc}$ acquisitions. Future work will incorporate simulated training data to aid the model in compensating for these artifacts.

## Acknowledgements

## References

1. Loecher M, et al. Phase unwrapping in 4D MR flow with a 4D single-step Laplacian algorithm. J Magn Reson Imaging. 2016;43: 833-842.
2. Ying L. Phase unwrapping. Wiley Encyclopedia of Biomedical Engineering. 2006.
3. Johnson K. 3D Velocimetry Phase Unwrapping using Block-wise Classification with a Shift Variant Fully 3D Convolutional Neural Network. ISMRM 2018.
4. He K, et al. Deep residual learning for image recognition. arXiv:1512.03385 [cs.LG]. 2015.
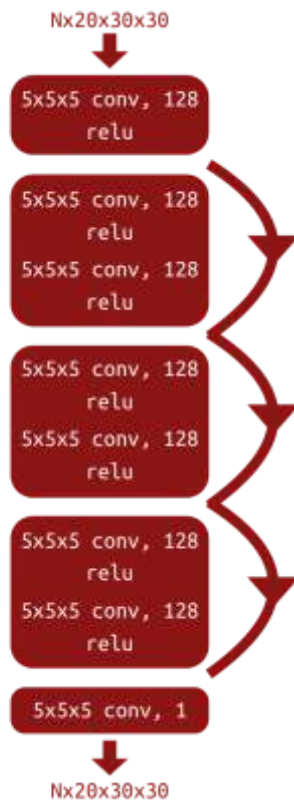
## Figures

Nx20x30x30

5x5x5 conv, 128
relu

5x5x5 conv, 128
relu
5x5x5 conv, 128
relu

5x5x5 conv, 128
relu
5x5x5 conv, 128
relu

5x5x5 conv, 128
relu
5x5x5 conv, 128
relu

5x5x5 conv, 1

Nx20x30x30

Figure 1. Our model is based on a 3D ResNet architecture. Note that although the images in the dataset are of size Nx100x150x30, the patching crops the images down to Nx20x30x30.
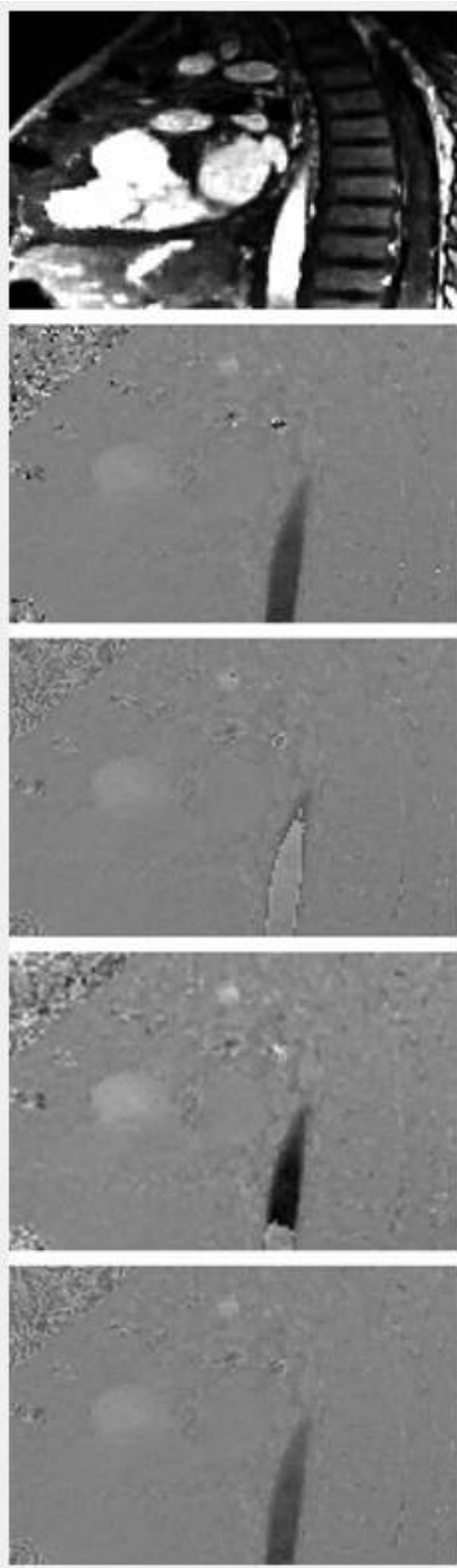
Figure 2. From top to bottom: magnitude-derived weights image, ground truth with a $V_{enc}$ of 250, simulated wrapping with a descending $V_{enc}$ from 75 to 25, the Laplacian algorithm output, and our model output.
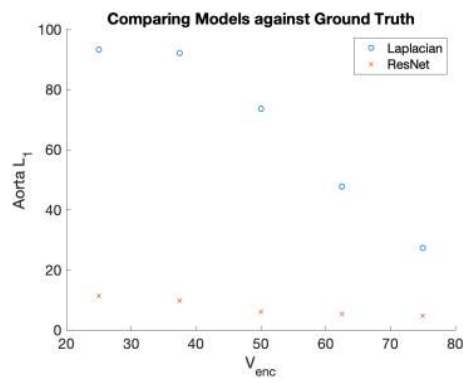
Figure 3. The blue points denote the pixel-wise $L_1$ difference between the Laplacian algorithm output and the ground truth at various $V_{enc}$ settings, while the orange points represent the same calculations for our model. Only the pixels in the aorta, as manually segmented, were used for this calculation, to disregard the noise phase.

# Deep Predictive Modeling of Dynamic Contrast-Enhanced MRI Data

Jiacheng Jason He[1], Christopher Sandino[1], Shreyas Vasanawala[1,2], and Joseph Cheng[2]

[1]Electrical Engineering, Stanford University, Stanford, CA, United States, [2]Radiology, Stanford University, Stanford, CA, United States

## Synopsis

**This work demonstrates the use of recurrent generative spatiotemporal autoencoders to predict up to fifteen future frames of abdominal DCE-MRI video data, starting with only three ground truth input frames for context. The objective is to predict what healthy patient video data and organ-specific contrast curves look like, to expedite anomaly detection and enable pulse sequence optimization. The model in this study shows promise; it was able to learn contrast changes without losing structural resolution during training time, and lays the foundation for future work.**

## Introduction

Dynamic contrast-enhanced MRI (DCE-MRI) provides an effective non-invasive assessment of the kidneys, livers, and other organs, and is particularly useful for oncological diagnosis. Two key challenges that diminish the clinical effectiveness of DCE-MRI are: the difficulty of optimizing for both spatial and temporal resolution, as well as local anatomy-of-interest resolution, due to already long scan times, and the lengthy manual data processing to quantify tissue perfusion and identify potential anomalies. Building on existing video prediction work,[1,2] we propose to use a recurrent generative spatiotemporal autoencoder model to predict contrast enhancement curves from initial pre-contrast images. By learning the contrast curves, the model enables real-time pulse sequence adjustments to optimize signal and contrast at key times and locations. Additionally, if the network can predict healthy behavior, the actual data can be compared with the prediction to quickly identify anomalies and compare tissue perfusion rates.

## Methods

Four-dimensional data from 100 patient volunteers are used; 70 patients for training, 10 for validation, and 20 for testing. Data was acquired using a $T_1$-weighted, 3D spoiled gradient recalled (SPGR) sequence on a GE 3T MR750 scanner, with a 32-channel cardiac coil.[3] Several preprocessing steps are taken: slices are zero-padded in the x- and y-directions so all images are the same dimensions, and the data is normalized. For initial proof of concept, only image magnitude is used. The first three ground truth frames are concatenated along the channel dimension for the first input. During training time, additional ground truth frames are used to predict each timestep. During validation time, predictions are concatenated to make the next predictions.

The autoencoder architecture is shown in Figure 1. The objective of the encoder is to embed the key features of the inputs, from which the decoder produces the next time frame. The encoder is based on a VGG16 model without fully-connected layers at the end.[4] The decoder upsamples and deconvolves, and adds a residual skip connection from the previous corresponding layers during the encoding network to assist with learning structural similarities. The loss function is a weighted mean squared error (MSE), with a higher weighting on the earlier frame predictions. In TensorFlow, we use the Adam optimizer to minimize dependence on learning rate tuning.

The three primary models are: a recurrent three-to-one (R3T1) model, a recurrent five-to-one (R5T1) model, and a non-recurrent three-to-one (N3T1) model. The R3T1 model is modelled after typical recurrent neural network (RNN) architectures, where the encoder and decoder share weights from timestep to timestep. The R5T1 model is an experiment to evaluate the benefits of more context frames. The N3T1 model is an unrolled model, which trains independent weights per timestep, which should theoretically improve performance at the cost of increased weights.

## Results & Discussion

Figure 2 shows the training results in sets of 3. The top set is from the R3T1 model, the middle set is the R5T1 model, and the bottom set is the N3T1 model. In each set, the top row is the ground truth frames, the middle row is the predicted frames, and the bottom row is the difference. All three models appear to perform well during training. Figure 3 shows the validation results for two slices over time. From left to right are the ground truth, R3T1 model, R5T1 model, and N3T1 model predictions. The early predictions are similar to the ground truth frames, but cases where one prediction diverges drastically from the ground truth, the following predicted frames deviate increasingly. This conclusion is evident in Figures 4 and 5. Figure 4 is the contrast enhancement curve, specifically for the liver. Figure 5 is the plot of MSE over time. The N3T1 model demonstrates the best performance over time as expected since each timestep trains unique weights. In practice, the predictive model will be combined with the data acquisition process and update the prediction as real data is acquired, so performance is expected to improve for all models.

## Conclusion

The evaluated models demonstrate the ability of a deep predictive model to learn contrast changes in abdominal DCE-MRI data. Different models are presented, each using time and context information differently, but ultimately drawing a similar conclusion: accurate prediction of each timestep is crucial to ensure no deviations that impair future predictions. One limitation with this study is that the data available to us came from patients with an assortment of abnormalities, since healthy pediatric patients are not often scanned. We believe, though, that training this network would only improve in accuracy if trained on healthy patient data, since there would be more anatomical consistency across patients.

## Acknowledgements

## References

1. Feng X, Meyer C. Accelerating cardiac dynamic imaging with video prediction using deep predictive coding networks. ISMRM. 2018.

2. Srivastava N, Mansimov E, Salakhutdinov R. Unsupervised learning of video representations using LSTMs. Proceedings of International Conference on Machine Learning (ICML). 2015.
3. Zhang T, Cheng JY, Potnick AG, et al. Fast pediatric 3D free-breathing abdominal dynamic contrast-enhanced MRI with high spatiotemporal resolution. J Magn Reson Imaging. 2015;41:460-473.
4. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 [cs.LG]. 2015.

# Figures



Figure 1. The encoder-decoder architecture used to generate predictions at each timestep. The encoder is based on the VGG16 architecture without fully-connected layers and produces the (N,6,7,1) embedding. The decoder deconvolves and upsamples encodings, using skip connections to improve structural preservation.
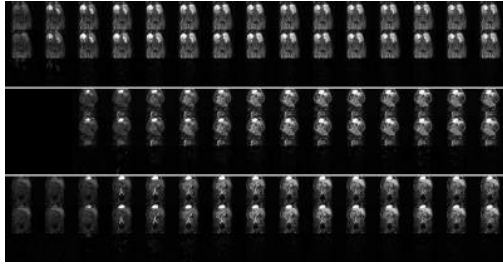


Figure 2. The top set is the R3T1 model, the middle set is the R5T1 model, and the bottom set is the N3T1 model, all at training time. In each set, the top row is the ground truth, the middle row is the output of the model, and the bottom row is the difference images.Since training data was randomly loaded, it is impossible to compare the same time series, but the difference images show that all three perform well during training.
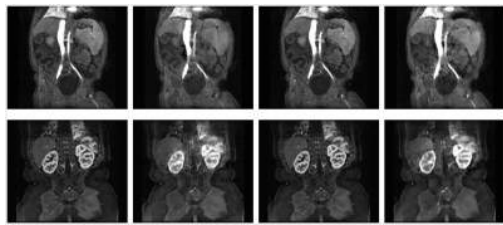


Figure 3. Results for two slices at validation time, one on the top row, one on the bottom. From left to right: ground truth, R3T1, R5T1, N3T1. The animation shows how each changes over time. All three deep predictive models perform well at first, but slowly they deviate from the ground truth. The N3T1 model preserves structure the best, since it had independently trained autoencoders for each timestep.
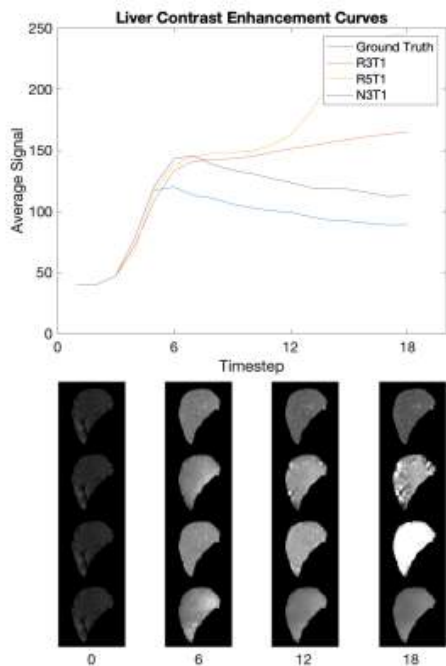
Figure 4. The contrast enhancement curve for the liver, with selected, segmented timesteps shown. The liver was manually segmented and the average signal was calculated at each timestep. Qualitatively, the shape of the curve for N3T1 looks best, albeit offset by a constant value.
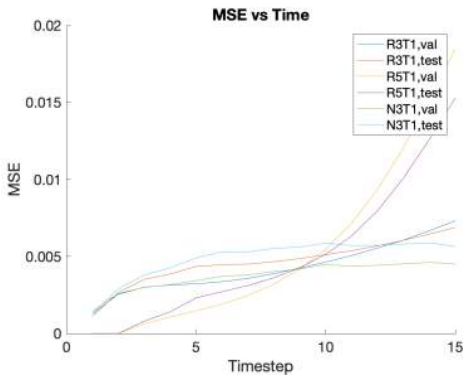


Figure 5. The validation and test MSE plotted over time for each model. This shows, quantitatively, that N3T1 was the best model.

# More Efficient Inductive Electric Vehicle Charger: Using Autonomy to Improve Energy Efficiency

Edward C. Fontana, Rick Barnett, Robert Catalano, James Harvey, Jiacheng He, George Ottinger, and John Steel
GE Critical Power

## ABSTRACT

Electric cars can help cities solve air quality problems, but drivers who live in apartments have no convenient way to charge daily, absent the well-controlled private garages where most electric vehicles (EVs) are currently charged each night. Environmentally robust, hands-free, inductive chargers would be ideal, but energy efficiency suffers. We asked whether the precise parking alignment provided by self-driving cars could be used to provide convenient inductive charging with improved charging efficiencies.

To answer this question, we split an inductor-inductor-capacitor (LLC) battery charger at the middle of the isolation transformer. The power factor correction, tank elements, and transformer primary windings are stationary, while the transformer secondary, rectifiers, and battery control logic are on the vehicle. The transformer is assembled each time the EV parks. A variety of transformers were tested for efficient energy transfer coincident with spacing to accommodate insulation on both the charger and vehicle side of the interface. Testing with different transformer parameters demonstrate a wall to battery energy efficiency of 95%, comparable to an onboard charger.

A hands-free, inductive, battery charger can deliver charging efficiencies comparable to galvanically connected onboard chargers – with no degradation in performance or safety when covered with a variety of contaminants. This shows promise for night charging of EVs at apartments, thus providing high public benefit with minimum public infrastructure expense.

**CITATION:** Fontana, E., Barnett, R., Catalano, R., Harvey, J. et al., "More Efficient Inductive Electric Vehicle Charger: Using Autonomy to Improve Energy Efficiency," *SAE Int. J. Alt. Power.* 6(2):2017, doi:10.4271/2017-01-1216.

## INTRODUCTION

Electric vehicles (EVs) provide a zero-emissions solution for transportation in cities. In 2013, sixty percent of over one thousand respondents were open to the idea of purchasing an EV [1]. Unfortunately, 48% of people surveyed do not have the ability to charge at home, due to lack of electrical outlet or space for a charger [1]. A wireless, hands-free charger that is durable enough to withstand environmental factors would be well suited to strategically important uncontrolled parking environments, like apartment parking lots and public garages [2], for overnight charging [3].

Modern inductive EV chargers provide low-clutter, ground-mount charging that is impervious to chemicals and electrically safe, as there are no exposed galvanic connections to create a shock or spark hazard. Today's systems are also hands-free. The existing designs available to consumers, though, are not energy efficient. Wall-to-battery inductive charging efficiencies are right at 85% [4, 5], with state-of-the-art research in the low nineties [6, 7, 8]. Galvanic chargers deliver 85 to 95% wall-to-battery efficiency [9, 10], and address hands-free operation by using a robot to make the high precision galvanic connections at waist level [11].

As vehicles become more autonomous and begin to park themselves with robot-precision [12], the need for a robot function in the charger is eliminated. The car can precisely drive into a docking station where the charging connections are made. We wondered, "Is there a way to use the precision of autonomous vehicle parking and modern circuit topologies to make ground-level inductive chargers as efficient as hot contact chargers?" GE Critical Power produces rectifier circuits that are 95 to 97% efficient [13]. The goal was to start there and use the transformer as the interface for inductive energy transfer to the vehicle—with single-owner apartment-home night charging as the use model. Figure 1 demonstrates our desired product.



Figure 1. CAD model of desired compact, efficient, inductive solution [14]

# APPROACH

We based our design on experience designing, producing and supporting over a million products using this circuit, references including [15] – [20] and two guiding principles.

The first, which is the foundation for the heart of this work is: "Everything is linear if you get close enough." We know this circuit delivers 95-96% efficiency with a separate resonant inductor and highly interleaved transformer. Our goal is to demonstrate a single transformer that can be assembled each time a vehicle parks that delivers conversion efficiencies that are comparable to that highly interleaved transformer that requires 20 minutes of assembly time to construct. To do that we needed to make all the key parameters close to those of our reference circuit.

The second guiding principle is: "Space allocation is key to product success." We had a team working to package the circuit in way that is practically useful. Heat dissipation, alignment, and customer acceptance are all important, if this product is to serve society well. These efforts act as bookends around the theoretical and experimental work that is the heart of this paper. These space allocation efforts form a blueprint for *future work* and appear only in that section.

## *Resonant LLC Circuit Topology*

Basing our design off of proven GE Critical Power rectifier topologies, we used the resonant LLC converter circuit with a half-bridge and output voltage doubling [21], Figure 2.
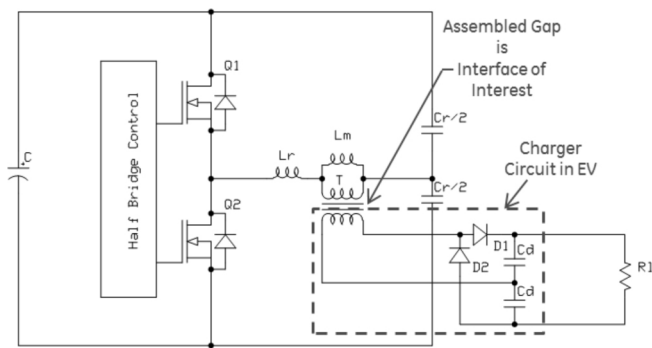


Fig. 2. Resonant LLC converter circuit with vehicle-side highlighted

For an environmentally robust design, we wanted to accommodate bumper covers and other plastic elements needed for 10,000 charging cycles of operation. Any separation in the magnetic structure caused by plastic covering prevents ideal coupling, creating leakage inductance. This intrinsic leakage inductance can be treated as beneficial. It allows us to forego the inclusion of a discrete resonant inductor and use the single assembled magnetic structure to establish resonance [21, 22].

## *Desired Characteristics*

With GE's in-house analysis tool, we calculated the First Harmonic Approximation for the circuit. The inputs were our desired input voltage, output power, and power efficiency, gathered from [10] for market needs. The tool produced the required turns ratio, magnetizing

inductance, resonant inductance—in our case, the leakage inductance, frequency range, and expected output voltage range. Table 1 shows our goals for the reference prototype; Figure 3 is an example of the tool's output as we work to hit the corners of the envelope.

Table 1. Design objectives for electric vehicle charging

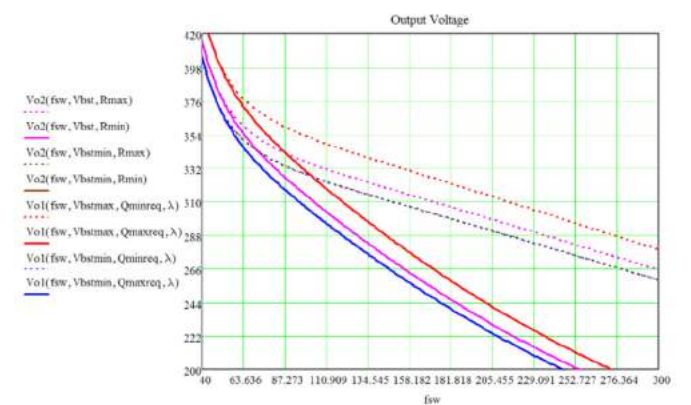| Characteristics | Value |
|---|---|
| Input voltage range | 390V – 420V |
| Nominal input voltage | 400V |
| Output voltage range | 280V – 420V |
| Maximum output power | 3kW |
| Desired efficiency | 95% |
| Desired magnetizing inductance ($L_M$) | 110µH |
| Desired leakage inductance ($L_L$) | 10µH |



Figure 3. Example simulation produced by in-house analysis tool

## *Resonant Frequency Range*

Efficiency and power loss is optimized when the circuit runs at or slightly above the resonant frequency—where the current running through the transformer is almost a perfect sinewave [21]. Our in-house analysis tool outputs a frequency range, with the minimum determined by the magnetizing inductance and the maximum determined by the leakage inductance. At high frequencies, current does not run efficiently through standard wire—the skin effect causes AC losses, and the proximity effect of the windings on each other reduce the efficiency, too. To decrease the effect of the variation of frequency on our transformer losses, we chose to use Litz wire for all of our windings.

## *Core Gap Determination*

Determining the transformer core gap, Figure 4, was an important decision at the start. The larger the core gap, the less the coupling [23]. For our environmental protection purposes, we wanted a large core gap. To find the optimal core gap, we first measured the effect of core gap on magnetizing and leakage inductance, and then found an operating point where variations in core gap would not create large variance in the inductances. We determined an insulator thickness of 1.5mm on each surface of the ground-side and vehicle-side would be durable and manufacturable with standard injection molding processes.
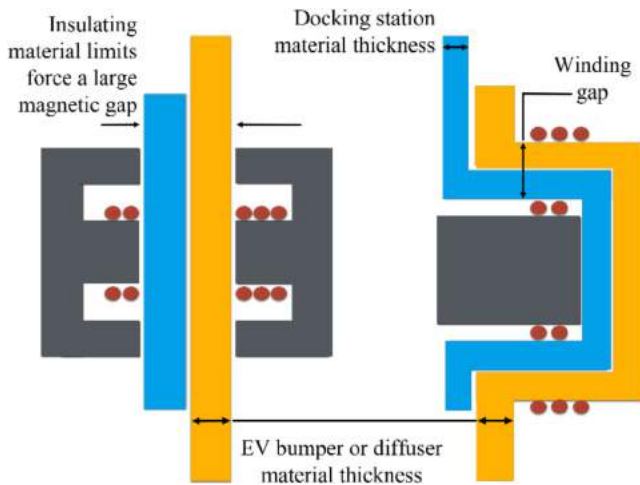
Figure 4. Model of transformers with core gaps

## INITIAL EXPERIMENTATION

We used a set-based concurrent engineering approach [24] to design the integrated transformer, so we wound and evaluated many transformers with a variety of geometries. For each geometry, we measured magnetizing inductance, leakage inductance, and AC resistance on the Keysight E4980A Precision LCR Meter.

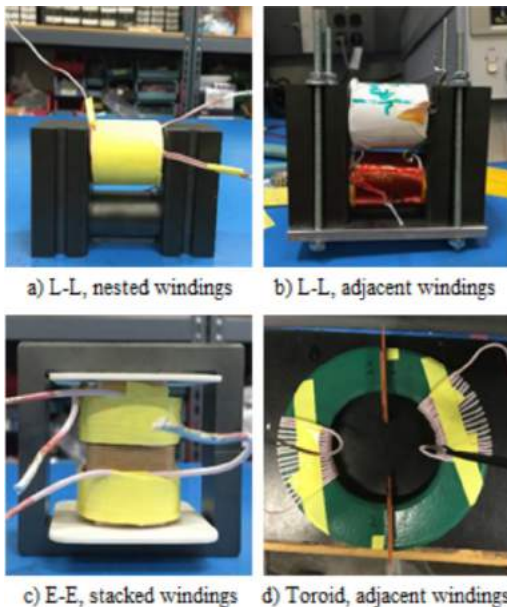The four primary geometries we began with are shown in Figure 5.



Figure 5. First designs tested for feasibility a) L-L, nested windings b) L-L, adjacent windings c) E-E, stacked windings d) Toroid, adjacent windings

Since the objective was to test a wide range of possibilities, we began with both "core" and "shell" type transformers [23]. Since theoretical calculation methods vary depending upon the type of core, we elected to wind and experiment instead of focusing on models and mathematics. The first geometry, Figure 5a, was an L-L core with the primary and secondary windings on the same leg. The primary winding nests into the secondary winding. From a theoretical

perspective, this winding should couple well, so the leakage inductance should be relatively low. The downside to this design is the accuracy necessary for the nesting of the windings. The second winding, Figure 5b, also uses an L-L core, but has the primary on one leg and the secondary on the other. Because there is no overlap of the windings at all, this winding was expected to have higher leakage and thus, lower magnetizing inductance. The third geometry, Figure 5c, was a "shell" type, an E-E core, with the primary winding on one end of the middle leg and the secondary on the other. The main benefit of the E-E core is the ability to have the windings close to each other, even though they do not overlap. Finally, we had a toroid core, Figure 5d, that was split in half, with windings on each side. Measurements were taken with the windings at a range of locations, from close together to far apart.

Additionally, we experimented with different configurations of core gap, turns ratio, number of turns per side, and frequency. We experimented from no core gap up to 13mm per leg. From our analysis tool, we theoretically wanted a turns ratio of approximately 1.25:1, but we tested more variations, too. Theoretically, the number of turns on the primary side effects the magnetizing and leakage inductances, so we wanted to see how that varied experimentally. Finally, because we chose to use Litz wire, we expected to see little variation at different frequencies. We tested frequencies from 20kHz to 120kHz to verify the effectiveness of the Litz wire.

Figure 6 shows a compilation of our initial magnetizing and leakage inductance measurements. The orange circle denotes the ideal values of 110μH and 10μH, respectively. This data shows that points from the same geometry were clustered closely, separated from the others. The clusters consist of the 102 configurations we tested. This suggests that core shape affects magnetizing and leakage inductance most.
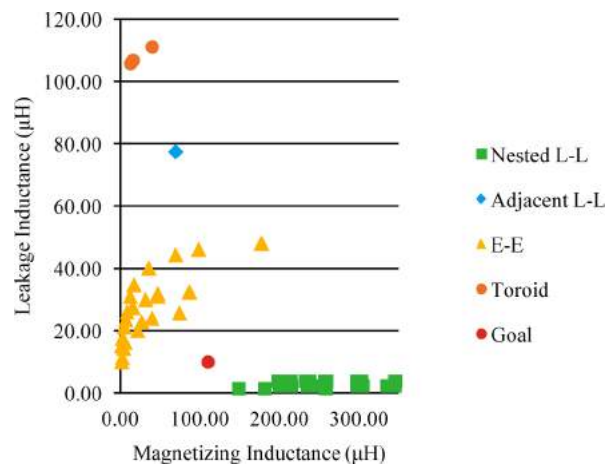


Figure 6. Effects of geometry on magnetizing and leakage inductance

Evidently, the adjacent windings on the L-L core and the split toroid core are not at all close to the ideal values. Meanwhile, some of the nested L-L core data points are close for magnetizing inductance, but all are too low in leakage inductance. The E-E core data points show a possibly ideal leakage, but low values for magnetizing inductance. W made the decision to focus on improving the nested L-L core and the E-E core [25].

# GEOMETRY IMPROVEMENTS

In the spirit of set-based concurrent engineering, we continued along with both the nested L-L and the E-E, since both had pros and cons. The nested L-L had more promising values, but required a more accurate alignment of the two sides. The E-E did not require nesting, but the low magnetizing inductance and high leakage inductance were concerning.

## *Improving the Nested L-L Core*

First, we wanted to increase leakage inductance in the nested L-L core geometry. E. C. Snelling's method for calculating leakage inductance is shown in (1), where N is the number of primary side turns, $l_w$ is the mean turn length, M is the number of section interfaces, Y is the winding overlap width, X is the overall winding height of affected windings, and $\sum X_\Delta$ is the sum of all interlayer section thickness. Note that all lengths and widths are in millimeters.

$$L_l = 4\pi10^{-4} \, (N^2 l_w / M^2 Y) \, (\sum X / 3 + \sum X_\Delta) \, \mu H$$

(1)

Meanwhile, (2) is the equation for magnetizing inductance in a core with an air gap, where $\mu_0$ is the permeability constant, A is the cross-sectional area of the core leg, and $l_{total\ gap}$ is the length of the air gap [23].

Although increasing the number of primary side windings would increase leakage inductance, it would also increase the magnetizing inductance. The nested L-L core already exceeds the desired magnetizing inductance, so we want to look at other factors that affect leakage inductance.

$$L_m = \mu_0 N^2 A / l_{total\ gap} \, \mu H$$

(2)

To only change the leakage inductance, we decided to shorten the width of the secondary winding. We accommodated this by making the winding taller. This decreases Y and increases X, $\sum X_\Delta$, and $l_w$. Therefore, we introduced a new design, keeping the L-L core and nested windings, but with a narrower, taller secondary, shown in Figure 7a.
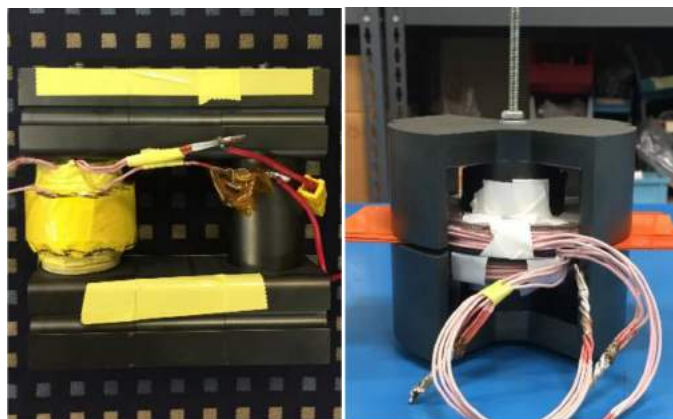


a. L-L, shorter nested          b. Pot core

Figure 7. Improved designs tested for feasibility

Upon adding the data from variations of the shorter nested L-L to the previous data, we see in Figure 8 that the leakage inductance increased from approximately 3µH to 10µH.
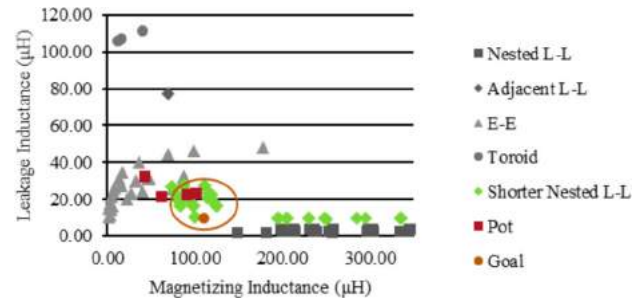


Figure 8. Introducing the shorter secondary nested L-L and pot core designs

## *Improving the E-E Core*

Next, we wanted to try to increase the magnetizing inductance on the E-E core. From (2), we knew the number of turns on the primary side would increase the magnetizing inductance. Therefore, keeping the same ratio, we wound an E-E core that was completely full with windings. Unfortunately, by filling up the core to maximize magnetizing inductance, the number of layers and thickness of the windings increased. This caused leakage inductance to increase greatly, too. The closest configuration to the desired values had between 80µH and 90µH magnetizing inductance, but at least 40µH leakage inductance. We concluded that changing the number of primary side windings was not the solution.

We wanted to keep the material—thus the permeability—the same, and we did not want to reduce the core gap further. Therefore, we decided to adjust the cross-sectional area of the core, Table 2. We found a large pot core, shown in Figure 7b, which had a larger cross-sectional area, so we measured its magnetizing and leakage inductances at varying core gaps and frequencies. The magnetizing inductance increased by a large margin, reaching over 100µH, while the leakage inductance was approximately 20µH. Although these numbers are not the exact desired values, we used GE's analysis tool to determine that it should still be usable if run at a lower resonant frequency of 85kHz.

Table 2. Transformer construction specifications

| Shape | L-L | E-E | Toroid | Pot |
|---|---|---|---|---|
| Core Number | EPCOS Stackable U-Cores | Magnetics Inc. 0R4820EC | DMR40 H160X9 0X20P | EPCOS PM 114/93 |
| Area (mm²) | 761 | 392 | 665 | 1720 |
| Magnetic Path Length (mm) | 170 | 184 | 392 | 200 |
| Volume (mm³) | 129,370 | 72,300 | 261,144 | 344,000 |
| Winding Leg Length (mm) | 45 | 56 | 392 | 63 |

Figure 8 also illustrates the performance of the pot core in context of the previous data. Along with the shorter secondary nested L-L core, the pot core was also much closer to the desired magnetizing and leakage inductance.

## MEASURING PEAK POWER EFFICIENCY

Moving forward with the shorter secondary nested L-L and the large pot core transformers, we took power measurements in an open loop with a fixed duty cycle. We used an electronic 400V source to simulate the output of the power factor corrected boost stage (PFC). Figure 9 shows our setup.

For our first prototype efficiencies, the shorter secondary nested L-L performed very well, achieving 97% efficiency across the resonant tank DC to DC converter. The pot core reached 91% resonant tank DC to DC efficiency, which did not meet our goal, but may have room for improvement. Additionally, the voltage supply that we used had a maximum output of 2.5kW, so we were not able to test up to the desired 3kW. From this initial data, though, we have already proven that it is possible for an inductive EV charger with a large core gap to be highly efficient. When viewed at a system level, with a 98% efficient PFC, wall-to-battery efficiency would be 95% for the shorter nested L-L core and 88% for the large pot core. This data is shown in Figure 10.
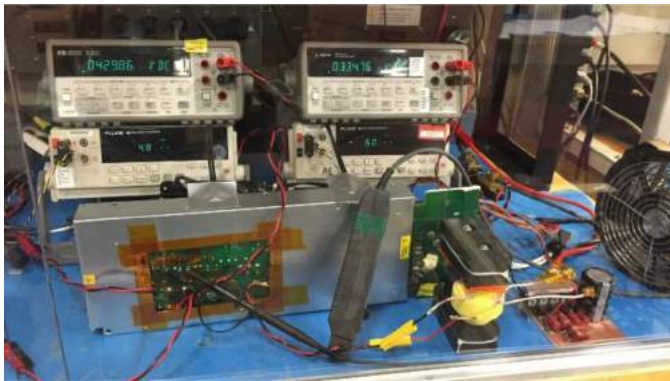


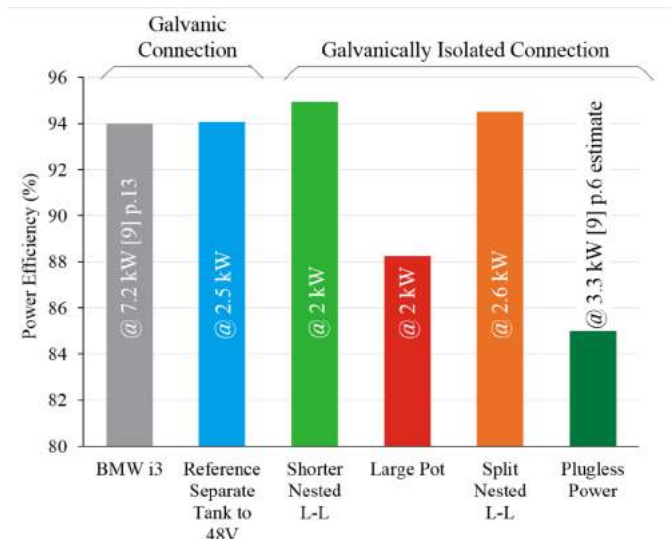Figure 9. Open loop power efficiency measurement setup



Figure 10. Peak wall-to-battery efficiency data comparison

Compared to the BMW i3's galvanic charger, Plugless Power's market inductive charger, and a production model transformer with a discrete resonant inductor for a market rectifier, the shorter nested LL core performed exceptionally. The solution does rely on the autonomous vehicle to precisely nest the two windings, though, so we wanted to see if we could understand and increase the margin for variance.

## DESIGN OPTIMIZATIONS

After achieving our 95% efficient design goal, we optimized our design for market-friendliness considering our use model.

### *Nested Winding Gap*

Our first optimization goal was to see how a gap between the nested primary and shorter secondary windings would affect the results. While keeping other factors constant, we tested winding gaps of 3, 4, 6, and 8mm and core gaps of 6, 8 and 12 mm. We used Nomex for 3mm of each gap as the plastic insulation; the rest of the winding gap was air.

Figure 11 shows the results of a parametric study of the influence of radial winding gap and magnetic path length core gap on the magnetizing and leakage inductances. Variables were kept constant except for winding and core gaps. The data show a linear relationship between winding gap and leakage inductance. Experimentally, although a 6mm gap creates a slightly too high leakage inductance, we can interpolate that a 5mm radial winding gap, to provide a leakage inductance of just over 20µH, would be acceptable. This 5mm radial winding gap translates into a 2 mm wall thickness for each of the stationary and vehicle sides of the interface and a 8mm gap in the magnetic path. A 2mm wall should provide good durability and a magnetic path length that delivers acceptable operational parameters for our LLC resonant circuit.
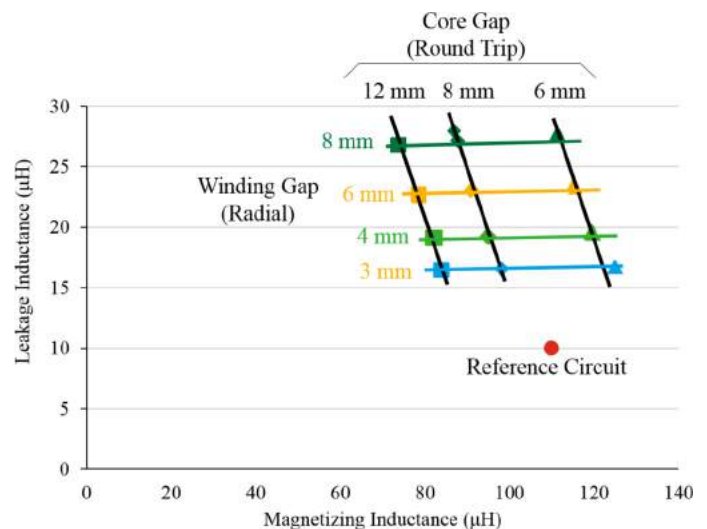


Figure 11. Effects of winding gap and core gap on magnetizing and leakage inductance

### Nested Winding Overlap

The second goal was to get uniform magnetizing inductance independent of the degree of nesting between the two winding elements. Figure 12 shows the apparatus we used. The secondary winding was indexed in 5mm increments from completely overlapped to not overlapped at all. Figure 13 shows the results.

Figure 13 shows that the useful region of overlap is between 0mm and 25mm, delivering leakage inductance that varies in a range from 7μH to 15μH.
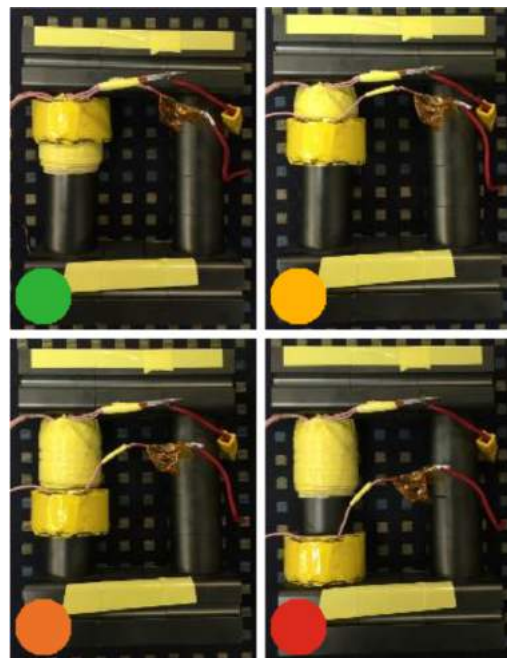


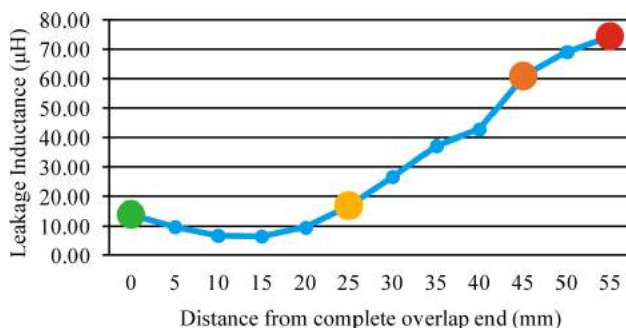Figure 12. Shifting of secondary to experiment with overlap distance



Figure 13. Effects of overlap distance on leakage inductance

### Split Nested L-L Transformer

To reduce conduction losses and improve efficiency further, we introduced a transformer design with the nested windings split on both legs of the L-L core, Figure 14. The efficiency of this transformer is almost as high as the single nested L-L, with 96.6% DC-DC efficiency at 2.64kW peak power. Figure 10 shows the wall-to-battery power efficiency data. This transformer topology did perform better than existing galvanic and inductive chargers.
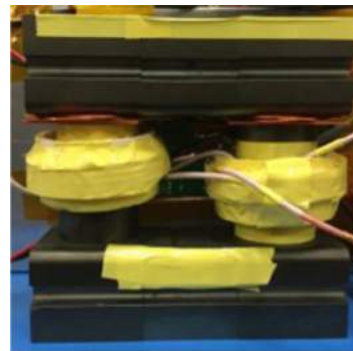


Figure 14. Split nested L-L transformer

## CONCLUSIONS

A hands-free, ground-level, EV docking, inductive battery charger can deliver charging efficiencies comparable to galvanically connected chargers—with no degradation in performance or safety when covered with a variety of environmental contaminants. Two different transformer constructions delivered winding and magnetic path length geometries to support large gaps—large enough to accommodate robust insulation on both ground-side and vehicle-side surfaces.

Ninety-five percent efficient, near ground level galvanically isolated inductive charging shows promise for night charging of EVs at apartments, thus providing high public benefit with minimum public infrastructure expense. Autonomous vehicle capability can indeed be applied to deliver hands free inductive charging while leaving a much smaller carbon footprint than do products on the market today.

## FUTURE WORK

With an understanding of the parameters that can be applied to deliver very efficient close coupled inductive charging, we can go forward on the size study working with a 2mm wall thickness for each of the insulators. Future focus is toward scaling this work up into a product that addresses market needs – similar to as shown in figure 1 – without losing any efficiency.
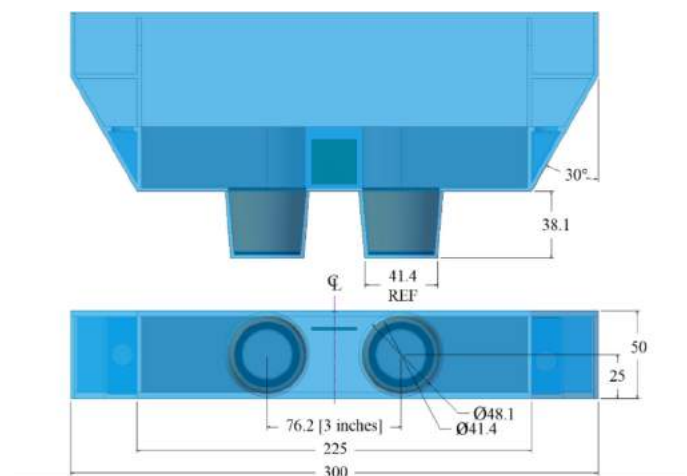


Figure 15. Stationary head dimension to deliver 7.2 kW (anticipated).

An accepted upper power level for home charging of Battery Electrical Vehicles is 10kW [26]. Interleaving two of these transformer circuits should deliver between and 5 and 10 kW – based largely on thermal limitations. Applying what we have learned here to redraw the stationary side of figure 4, using a nested approach with two transformers, we obtain an interleaved power train arrangement. This arrangement can be configured with either one or two power trains provisioned depending on customer preference. The primary side of this arrangement is shown in figure 15.

Additional volume is provided to deliver a self-alignment function that can be integrated into vehicles with a variety of industrial design languages. To achieve alignment without subjecting the nesting insulators to abrasion, the flat top surface first puts the docking head in planar alignment with the vehicle mounted *power receiver*. Then a 30-degree wedge followed by a flat on each side of the head locates the mating parts before nesting occurs. Only one degree of freedom remains for the last 40-50 mm of travel. Optical/Infrared communication ports verify alignment and provide redundant communication about battery bus voltage and current.

With the two nesting areas for windings biased toward the center, the return magnetic path is biased outwards and provided on the vehicle side of the interface. Energy transfer with 5 to 10 kW of power capability is provided by an active area 225mm wide and 50mm tall.

The need for magnetic return paths when applied to Figure 1, with the car subtracted, obtains Figure 16. We decided to configure the power receiver to simulate a dual exhaust - a survival feature that customers less familiar with electric cars will process without too much hesitation. This will be the lead concept and be used as an efficiency benchmark for alternative constructions.



Figure 16. Inductive Power Receiver Mounted in Diffuser

Our goal is 7.2kW delivered with the structure shown in Figure 17.

After confirming that efficiency is maintained in operation, we will confirm EMI performance in the GE Critical Power FCC Certified EMI chamber. Our hypothesis is the pot core will be better for EMI performance.

Another focus, for us to be able to bring this design into production, is cost. The large pot core was extremely expensive, and all the excess winding space is not necessary. The key element that makes it work is the large cross-sectional area. Therefore, we will experiment

with stacking E-E cores side-by-side to increase the cross-sectional area. Our thesis is that this will emulate the pot core and will provide the magnetizing inductance we need to deliver good battery charging performance at a much lower cost.
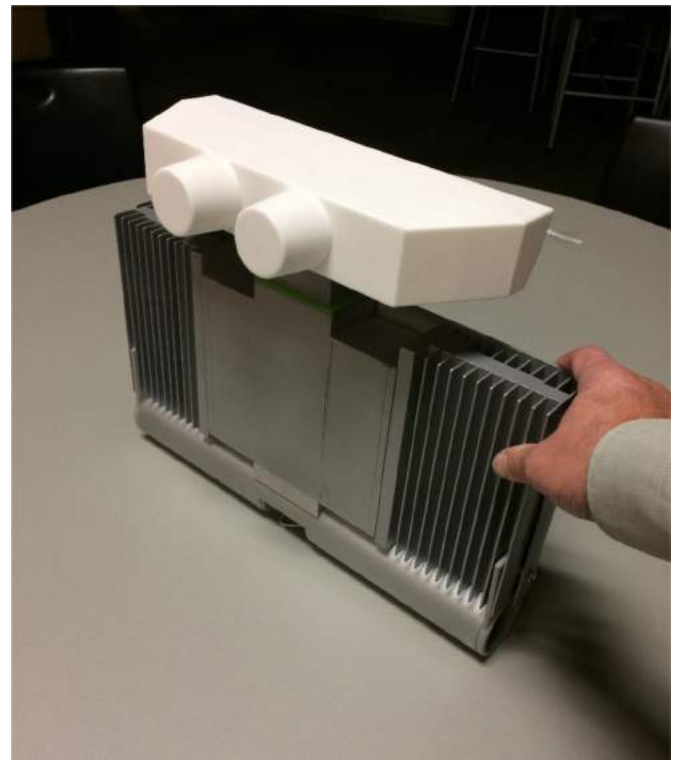


Figure 17. Thermal performance and efficiency test apparatus

# REFERENCES

1. Union of Concerned Scientists (UCS) and Consumer Reports, "Electric Vehicle Survey Methodology and Assumptions: American Driving Habits, Vehicle Needs, and Attitudes toward Electric Vehicles," 2013, http://www.ucsusa.org/evsurvey, accessed Aug. 2016.
2. Kettles, D., "Electric Vehicle Charging Technology Analysis and Standards," Florida Solar Energy Center, FSEC Report Number: FSEC-DR-1996-15, 2015, http://www.fsec.ucf.edu/en/publications/pdf/FSEC-CR-1996-15.pdf accessed July 2016.
3. Idaho National Laboratory, "Plug-in Electric Vehicle and Infrastructure Analysis," INL/EXT-15-35708, 2015, https://avt.inl.gov/sites/default/files/pdf/arra/ARRAPEVnInfrastructureFinalReportHqltySept2015.pdf accessed June 2016.
4. Onar, O.C. and Jones, P.T., "Wireless Charging of Electric Vehicles," Oak Ridge National Laboratory Presentation at U.S. DOE Vehicle Technologies Office 2015 Annual Merit Review and Peer Evaluation Meeting, Jun. 2015, http://energy.gov/sites/prod/files/2015/07/f24/vss103_onar_2015_o.pdf accessed July 2016.
5. Idaho National Laboratory, "PLUGLESS™ Level 2 EV Charging Systems (3.3kW) by Evatran Group Inc. – Results from Laboratory Testing off-board the vehicle," INL/MIS-14-33980, 2015, https://avt.inl.gov/sites/default/files/pdf/evse/PLUGLESSEvatranStandaloneTestResultsFactSheet.pdf accessed July 2016.
6. WiTricity®, "WiT-3300: WiT-3300 Development Kit," http://witricity.com/products/wit-3300/, accessed Jul. 2016.
7. Lu, F., Zhang, H., Hofmann, H., and Mi, C., "A Double-Sided LCLC-Compensated Capacitive Power Transfer System for Electric Vehicle Charging," *IEEE Transactions on Power Electronics* 30(11):6011–6014, 2015, doi:10.1109/TPEL.2015.2446891.
8. Onar, O.C., Miller, J.M., Campbell, S.L., Coomer, C. , "Oak Ridge National Laboratory Wireless Power Transfer Development for Sustainable Campus Initiative," Transportation Electrification Conference and Expo, 2013, doi:10.1109/ITEC.2013.6574506.

9. Carlson, B., "Idaho National Laboratory's Electric Vehicle Charging Infrastructure Laboratory," INL/MIS-15-36953, 2015, https://avt.inl.gov/sites/default/files/pdf/presentations/VSATTOctober2015EVIL.pdf accessed July 2016.

10. Deng, J., Li, S., Hu, S., Mi, C.C. , "Design Methodology of LLC Resonant Converters for Electric Vehicle Battery Chargers," IEEE Transactions on Vehicular Technology 63(4):1581–1592, 2014, doi:10.1109/TVT.2013.2287379.

11. EV Obsession, "Robotic Electric Car Charging Station From… Volkswagen?" http://evobsession.com/robotic-electric-car-charging-station-from-volkswagen/, accessed Jul. 2016.

12. Bosch Media Service, "Accident-free parking: Bosch technology makes anyone a professional parker," http://www.bosch-presse.de/pressportal/en/bosch-technology-makes-anyone-a-professional-parker-44808.html, accessed Jul. 2016.

13. GE Critical Power, "NE050AC48ATEZ Infinity Rectifier," Fact Sheet NE050AC48TEZ-FS, Rev. 06/2016, http://apps.geindustrial.com/publibrary/checkout/NE050AC48ATEZ-FS?TNR=Data%20Sheets%7CNE050AC48ATEZ-FS%7CPDF&filename=NE050AC48ATEZ-FS%206-3-16.pdf accessed Oct. 2016.

14. Fontana, E., "Passengers First Light Truck - A Modern Take on a Narrow Wake," SAE Technical Paper 2016-01-1333, 2016, doi:10.4271/2016-01-1333.

15. Dowell, P.L., "Effects of eddy currents in transformer windings," *Proceedings of the Institution of Electrical Engineers* 113(8):1387–1394, 1966, doi:10.1049/piee.1966.0236.

16. Perry, M.P., "Multiple Layer Series Connected Winding Design for Minimum Losses," *IEEE Transactions on Power Apparatus and Systems* 98(1):116–123, 1979, doi:10.1109/TPAS.1979.319520.

17. Vandelac, J.P. and Ziogas, P., "A novel approach for minimizing high-frequency transformer copper losses," *IEEE Transactions on Power Electronics* 3(3):266–277, 1988, doi:10.1109/63.17944.

18. Taguchi, G., "Introduction to Quality Engineering: Designing Quality into Products and Processes," (Quality Resources, 1986), ISBN 978-9283310846.

19. Jiang, H., Brazis, P., Tabaddor, M., and Bablo, J., "Safety Considerations of Wireless Charger For Electric Vehicles – A Review Paper," *IEEE Symposium on Product Compliance Engineering*, 2012, doi:10.1109/ISPCE.2012.6398288.

20. SAE International Surface Vehicle Information Report, "Wireless Power Transfer for Light-Duty Plug-In/Electric Vehicles and Alignment Methodology," SAE Standard J2954, Iss. May 2016.

21. ST Microelectronics, "An introduction to LLC resonant half-bridge converter," AN2644, 2008.

22. De Simone, S., Adragna, C., and Spini, C., "Design guideline for magnetic integration in LLC resonant converters," *International Symposium on Power Electronics, Electrical Drives, Automation and Motion*, 2008, doi:10.1109/SPEEDHAM.2008.4581225.

23. Rhodes, S.A., "The effects of separable cores on high power transformer design," M.S. thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1996.

24. Ward, A., Liker, J.K, Cristiano, J.J., and Sobek, D.K.II, "The Second Toyota Paradox: How Delaying Decisions Can Make Better Cars Faster," *Long Range Planning*, 1995, doi:10.1016/0024-6301(95)94306-J.

25. Pugh, S., "Concept Selection: A Method That Works," Proceedings of the International Conference on Engineering Design, 1981.

26. EV Obsession, "Electric Car Charging Capabilities — Comparison of 27 Models" September 8, 2015, http://evobsession.com/electric-car-charging-capabilities-comparison-of-27-models/, accessed Dec. 2016.

## CONTACT INFORMATION

For further information, please contact Ed Fontana by phone at 214-364-3184, or email at edward.fontana@ge.com

# PineappLeNet:
# Synthesizing Dynamic Contrast-Enhanced (DCE) Magnetic Resonance Data with a Generative Spatiotemporal Autoencoder

Jiacheng Jason He

Stanford University

450 Serra Mall, Stanford, CA 94305

jjhe@stanford.edu

## Abstract

The objective of this project was to generate temporal data for pediatric dynamic contrast-enhanced (DCE) magnetic resonance (MR) image data. Existing work on video prediction in other applications exist, but rely on fairly complex models. This paper proposes four recurrent models that take in limited context images to produces the rest of the timeseries. First, the data is zero-padded, normalized, and pruned to obtain timeseries where contrast changes are more frequent and noticeable. The first model takes in the first three timesteps and encodes and decodes them into the next timestep, which is then passed along for the next timestep prediction. The second model uses the first five timesteps. The third model uses an adaptive pruning technique to refine the data differently. The fourth model does not reuse weights and biases in the encoders and decoders between timesteps. The fourth model performed the best, producing the smoothest images and capturing nuances of the contrast changes.

## 1. Introduction

Magnetic resonance imaging (MRI) beautifully captures the magic of the Fourier Transform in a multi-dimensional domain. Data is collected over time, but after an inverse Fourier Transform, spatial data is produced. Unfortunately, this means that with a basic scan sequence, in order to improve image resolution, field-of-view, signal-to-noise ratio (SNR), and other desirable attributes, scan time must be increased, as detailed in [10].

One specific valuable imaging method is free-breathing dynamic contrast-enhanced (DCE) abdominal imaging, where some contrast agent is administered and images of the kidneys, liver, and other organs of the abdomen are rapidly acquired over a few minutes. This type of image data has proven to be primarily useful for detecting and characterizing lesions and tumors, as well as other clinical applications [17].

The two primary challenges of free-breathing DCE MRI highlighted in [17] are: the limited spatiotemporal resolution due to ephemeral contrast agent dynamics but lengthy 3D data acquisition time, and motion artifacts, particularly in pediatric populations. The thought process that followed, was: since DCE MRI structural information stays relatively constant–the primary change is in pixel values as contrast dynamics change–is it possible to use deep learning methods to produce temporal data from a limited number of time frames?

The predictions could be useful for a variety of clinical applications, such as: anomaly detection, normal or abnormal classification, and scan optimization.

Anomaly detection would take place as following: clinicians can take a free-breathing DCE MRI, then pass the first frame through PineappLeNet. If PineappLeNet has high enough accuracy, the results can be compared to the actual scan data to see if there are any anomalies in the scan data. This could lead to faster diagnoses.

Normal or abnormal classification would work if from the results of the network, we see a clear difference between healthy and unhealthy patient data. This could lead to faster diagnoses, as well.

Finally, scan optimization is possible if PineappLeNet can predict at which points contrast changes will occur. Scan operators can identify where big dynamic shifts occur and increase spatiotemporal resolution at those points, specifically, while sacrificing resolution at less critical points.

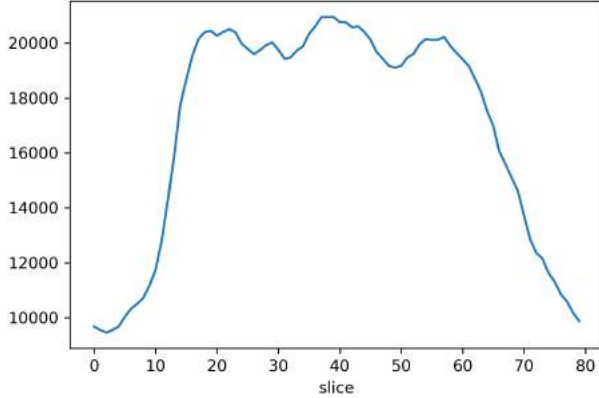Ultimately, the desired, ideal end-goal would be to

Figure 1. Example of mean differences plot.

input a single timestep and output the following seventeen timesteps. The limitations of the models studied, though, require some context such as the first three timesteps to produce the following fifteen.

## 2. Related Work

Interest in deep learning in the medical community has been growing rapidly, to tackle a variety of clinical needs [4] [3]. This shows that certain tasks, if accomplished with high enough accuracy, can be translated into and accepted by clinical practice.

In addition, this concept of producing temporal data from static or limited temporal data is not a foreign idea. Much work has been done in both interpolation [8], [9] and extrapolation [1], [2], [7], [8] [15], [16] of video frames from limited data. And expanding even more broadly, different ideas for generation of image data not necessarily over time and improvement of training generation networks have been attempted [5], [11], [13].

Specifically, in the area of video frame extrapolation, [7] appears to be one of the leading, state-of-the-art models. Their method proposes a combination of a variational autoencoder (VAE) and a generative adversarial network (GAN), drawing on the strengths of each. The VAE produces diverse results, while the GAN produces realistic, clearer images. Their model is based on their previous works [2], [1]. The strengths of their model are many, primarily the accuracy and crispness of images produced. The challenge, though, of replicating or building off their approach, was the sheer complexity of the system, since it involved many advanced neural network techniques, such as: recurrence, latent codes, warped pixels, and more.

## 3. Data

### 3.1. Dataset Information

This project builds directly on the work in [17]. The free-breathing DCE MRI dataset of 100 patients was provided by Joseph Y. Cheng and Chris Sandino of the Stanford Magnetic Resonance Systems Research Lab (MRSRL). 70 patients are used for training, 10 for validation, and 20 saved for testing.

The original study was conducted from June 2013 to July 2013, and 23 pediatric patients were scanned with a T1-weighted, multiphase 3D modified spoiled gradient recalled (SPGR) sequence on a GE 3T MR750 scanner, with a commercially-available 32-channel cardiac coil or a torso coil. There were 13 males and 10 females, with an age range from 6 weeks to 8.75 years old. After the conclusion of this study, 77 additional patients were scanned with the same sequence.

The dataset was initially formatted in the compressed file library (CFL) format, which could be read and imported into numpy arrays via functions written by Martin Uecker and Jonathan Tamir, from UC Berkeley.

The data was captured along the coronal plane with a varying numbers of samples along the phase encode (y direction of images) and readout (x direction of images) directions. The number of phase encodes ranges from 156 to 224; the number of readouts ranges from 154 to 192. Along the time axis, each dataset has 18 frames, although the frame rate varies. According to [17], for the 23 initial patients, the shortest frame rate ranged from 5.4 to 7.6 seconds, with an average of 6.5 seconds. This correlates to an average scan time of 117 seconds.

### 3.2. Preprocessing

Several preprocessing steps were taken. Slices were padded with zeros such that each slice was of height 224 and width 192. The number of slices (z direction) also varied, from 56 to 90. I did not zero-pad or crop data in the z direction because to greatly increase the amount of data, I train and test across only the x, y, and time axes, treating the number of slices as the batch size. The MRI data does not have RGB channels; rather, the data is complex and has real and imaginary parts. Based on experience from previous MRI deep learning projects, I took the magnitude image. I also normalized the data by finding the max image and dividing all images by the max, so images were within comparable ranges.

One final preprocessing technique I introduced during experimentation was deterministic and adaptive pruning. The idea came about because each patient's data contains some slices, such as the very front or very

| Input | Convolution | Max Pooling | Deconvolution | Output | Residual Connection |

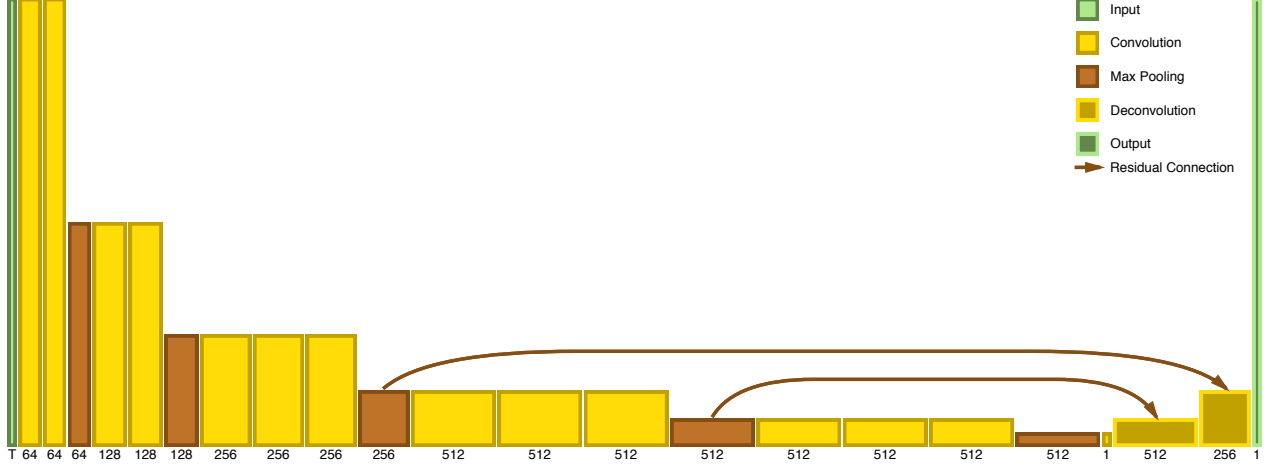| T | 64 | 64 | 64 | 128 | 128 | 128 | 256 | 256 | 256 | 256 | 512 | 512 | 512 | 512 | 512 | 512 | 512 | 512 | 1 | 512 | 256 | 1 |

Figure 2. Encoder-decoder architecture used in the working models.

back of the patient, where no meaningful changes are visible due to lack of contrast agent flow. I hypothesized that by training on data where the majority of timesteps showed no change, the networks would not effectively learn how to respond when there was supposed to be changes in contrast. Thus, I took the absolute differences between each timestep for each slice and reduced these down to a mean difference by averaging. The slices with more changes over time would have higher mean differences, while slices with few changes over time would have lower mean differences. An example of the mean differences plotted along the slice axis can be seen in Figure 1. For the deterministic pruning, I sorted the mean differences and took the top twenty slices per patient. For the adaptive pruning (AdaProon), I sorted the mean differences and took the top twenty-five percent per patient.

I used two different methods to split my data into features and labels. The first method, which I refer to as single-step, takes some set number of images concatenated along the channel dimension for the features and uses the next time step as the labels. The second method, which I refer to as timeseries, takes some set number of images along the time dimension for the features and uses the remaining timesteps in the whole timeseries as the labels. The purpose of the two data preparation methods will become clear in the discussions of the experiments.

## 4. Methods

### 4.1. Encoder

The general encoder-decoder based approach was devised based on the Fully Convolutional Network (FCN8) model described in [12]. The idea was to encode images into smaller embeddings, then decode them into the next time frame, as shown in Figure 2.

The encoder was based heavily on the VGG16 model as described in [14], but without the fully connected layers at the end. Instead, our model takes an image of size $(N, H, W, C)$ and passes it through a series of convolutions and max pooling layers, as shown in Figure 2. The convolutional layers are defined by the number of filters $(K)$, kernel size $(F)$, stride $(S)$, and zero-padding $(P)$. The equation to calculate the output of a convolutional layer is:

$$H_2 = \frac{H_1 - F + 2P}{S} + 1 \tag{1}$$

$$W_2 = \frac{W_1 - F + 2P}{S} + 1 \tag{2}$$

$$C_2 = K \tag{3}$$

Thus, for example, for my first layer, the input is $(N, 192, 224, T)$, where T is the arbitrary number of timesteps that are concatenated as an input. With a 3x3 kernel size, 64 filters, and padding of 1, the output size is $(N, 192, 224, 64)$.

The pooling layers are defined by the pooling size $(F)$ and the stride $(S)$. The equation to calculate the output of a pooling layer is:

$$H_2 = \frac{H_1 - F}{S} + 1 \tag{4}$$

$$W_2 = \frac{W_1 - F}{S} + 1 \tag{5}$$

$$C_2 = C_1 \tag{6}$$

By disregarding the fully connected layers in [14], the number of parameters in my encoder network decrease drastically. Calculating the number of weights

| Model | Epochs | Batch Size | LR Start | LR Decay | Pruning | Loss | Total Steps |
|-------|--------|------------|----------|----------|---------|------|-------------|
| R3T1 | 32 | 2 | 1e-4 | 0.96/1e4 | Deterministic | MSE | 22400 |
| R5T1 | 32 | 2 | 1e-4 | 0.96/1e4 | Deterministic | Weighted MSE | 22400 |
| RAP | 32 | 2 | 1e-4 | 0.99/1e4 | AdaProon | Weighted MSE | 21480 |
| RNR | 32 | 2 | 1e-4 | 0.99/1e4 | Deterministic | Weighted MSE | 22400 |

Table 1. Comparison of hyperparameters used with different models.

used in the encoder, I get approximately 14.7 million. Including the number of biases, this would increase slightly.

### 4.2. Decoder

The decoder was similar to the encoder, but going in the opposite direction. Rather than having many convolutional layers, though, I set it up to both upsample and deconvolve in the same layers. The amount of upsampling was dependent upon the stride and the kernel size was always double the stride; for example, a stride of 2 and kernel size of 4 would double the size. Between each deconvolutional layer, I added a residual skip connection from the previous corresponding layers during the encoding network. This was used in [12] and is useful because of the similarities in structure from timestep to timestep.

### 4.3. Loss

For all my loss calculations, I used some variant of the mean squared error (MSE) between the ground truth images and output predictions. The mean squared error for one timestep is calculated as such, where M and N are the dimensions of the image, $g$ is the ground truth image, and $\hat{g}$ is the predicted image:

$$\frac{1}{MN} \sum_{n=1}^{M} \sum_{m=1}^{N} [\hat{g}(n,m) - g(n,m)]^2 \qquad (7)$$

One flaw of using MSE is that it depends on the image intensity scaling, but since all the data was normalized to the same scaling, this flaw is mitigated.

To combine the MSEs of all the timesteps, one approach used was to just sum over all the timesteps. During experimentation, though, especially in the recurrent models presented later, the key really lies in predicting the first few timesteps correctly. If the first few are predicted incorrectly, then the following timesteps will undoubtedly be poor. Thus, I tried weighting the MSEs, so the loss would be higher if the first timestep was poorly predicted than if there was just some error in the later timesteps.

### 4.4. Optimization

For optimization of loss, I chose to use the Adam algorithm from [6] because of the low dependence on learning rate tuning. Adam first initializes the first and second moment vectors to zero, then it updates until it converges. The simplified update algorithm is as follows, where $m$ is the first moment, $v$ is the second moment, and $l$ is the learning rate:

$$m = \beta_1 m + (1 - \beta_1)dx \qquad (8)$$

$$v = \beta_2 v + (1 - \beta_2)dx^2 \qquad (9)$$

$$x = x - \frac{lm}{\sqrt{v} + \epsilon} \qquad (10)$$

### 4.5. Evaluation

For evaluation of my results, I relied on the absolute difference between ground truth timeseries and my predicted timeseries. This was a simple metric to use to be able to see which areas of the image the network was having trouble predicting.

$$d = |\hat{g} - g| \qquad (11)$$

The closer the absolute difference image was to all zeros the better. The advantage of using an image as a metric rather than reducing to a scalar was the preservation of spatial information and clearer understanding on what features the network was learning.

## 5. Experiments

I started by experimenting with single-step models, primarily to see how many previous timesteps the encoder-decoder would need to accurately predict the next timestep. From those preliminary results, I experimented with various recurrent models. The parameters I experimented with include: learning rate and decay, data pruning, and loss function weighting. Table 1 shows the primary recurrent model experiments.

### 5.1. Single-Step Models

From existing work on video prediction [1], [2], [7], [8] [15], [16], I learned that context from previous timesteps is important for accurate video prediction.

The question with my single-step experiments was: how much context is necessary, with regards to my data specifically?

I tried a two-to-one, a three-to-one, and a five-to-one. I concatenated two, three, or five timesteps along the channel dimension and passed it through the encoder-decoder. The two-to-one method performed poorly, but the three-to-one and five-to-one both had promising results. Thus, I moved forward with models based on those.

### 5.2. Recurrent Three-to-One (R3T1)

The three-to-one single-step model performed well from a qualitative observation, so I started with a recurrent model based on that design. Rather than using existing TensorFlow RNN cells, I manually unwrapped the recurrence I wanted. For similar behavior to typical RNN architectures, the encoder and decoder reused the same weights and biases from step to step.

During training, the model takes three timesteps from the ground truth data, concatenates, encodes, and decodes into the predicted next timestep. During validation and testing, only the first three ground truth steps are used. After that, the predictions are concatenated instead of the ground truth next steps. The model predicted 15 frames.

The MSE between the ground truths and predictions is calculated as described in the Methods section and minimized with the Adam optimizer. I started with a learning rate of 1e-4 and exponential decay of 0.96 per 1e4 steps. I tuned the learning rate some, but as described in [6], the Adam algorithm is not affected greatly by tuning.

The model was trained on deterministic pruned data, described in the Data section.

### 5.3. Recurrent Five-to-One (R5T1)

I also implemented a recurrent model based on the five-to-one single-step model. The design was similar to the R3T1 model, with the only difference being the model took five timesteps to predict the next. The advantage was the predictions had more context to work off of. The disadvantage was the model predicted 13 frames rather than 15.

For the loss calculation, I experimented with a weighted MSE, as described in the Methods section. The weights were: 12, 8, 8, 6, 6, 4, 4, 2, 2, 1, 1, 1, 1. I used an Adam optimizer with a starting learning rate of 1e-4 and exponential decay of 0.96 per 1e4 steps.

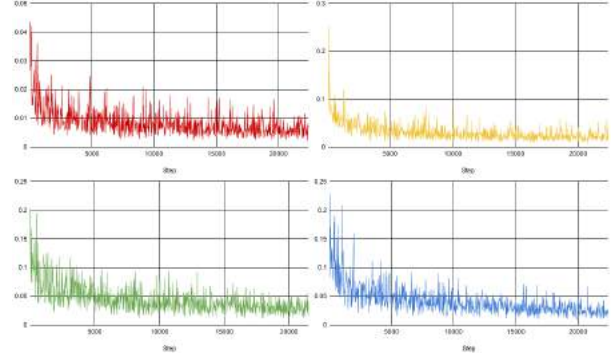The model was trained on deterministic pruned data, described in the Data section.



Figure 3. Comparison of training loss curves.

### 5.4. Recurrent AdaProon (RAP)

The next experiment was similar to the R3T1 model, but I pruned the training data with the AdaProon approach, described in the Data section. This resulted in less steps because less data was used for training. I also tried tuning the learning rate decay to decay less. The weights used for the loss calculation were: 16, 8, 8, 8, 6, 6, 6, 4, 4, 4, 2, 2, 2, 1, 1, 1, 1.

### 5.5. Recurrent NoReuse (RNR)

Finally, I experimented with an approach where weights and biases were not reused from timestep to timestep. This meant that each encoder and decoder would train a new set of weights and biases for each timestep. The advantage to this model is, presumably, the results would be better because rather than learning a generalized set of weights and biases for going from three steps to the next, it would learn weights and biases unique to each timestep. The disadvantage is it is less of a recurrent network and more just a large network with lots of weights and biases to train. The weights used for the loss calculation were the same as the RAP model.

## 6. Results & Discussion

### 6.1. Training

Figure 3 shows the training loss for the four models: R3T1 (red), R5T1 (yellow), RAP (green), and RNR (blue). While the loss curves were all quite noisy, they all generally had the optimal shape.

Figure 4 shows the training performance on a single dataset each. The top row of each set of three rows is the ground truth, the middle row is the training prediction, and the bottom row is the absolute difference. The first set of three rows is from R3T1, the second is R5T1, the third is RAP, and the fourth is RNR. Because I shuffled my data, I could not retrieve the same
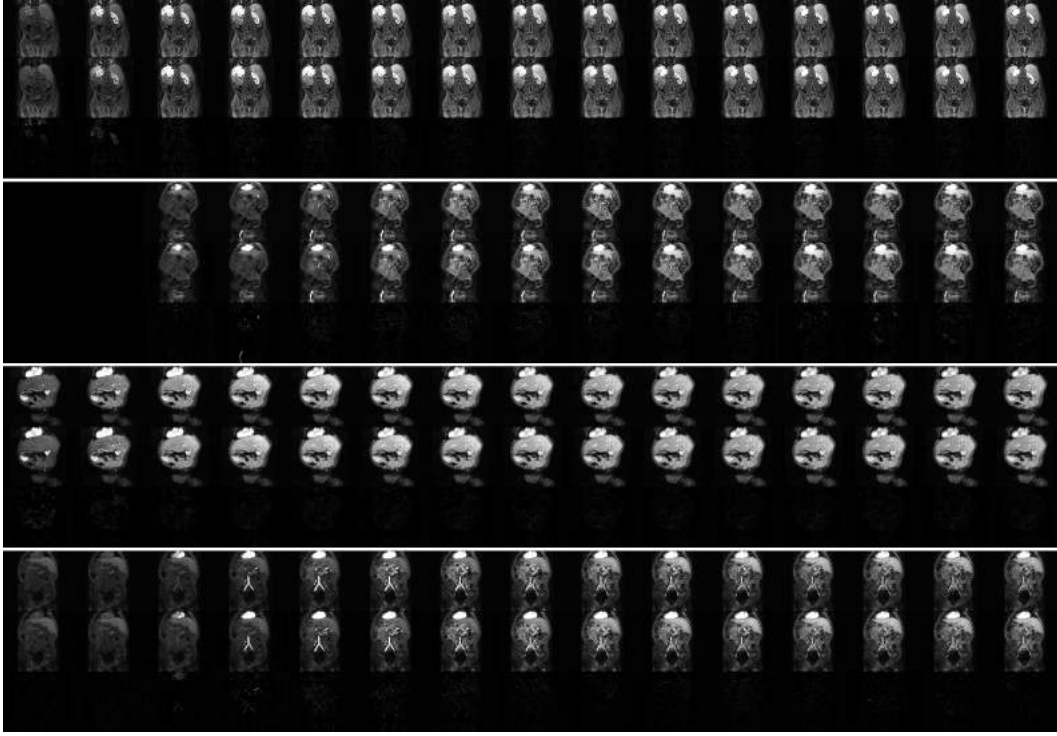
Figure 4. Example of training results for the models.

slice to compare, but overall, the training performance is qualitatively good. I had my colleagues, Joseph and Chris, look at the ground truth images and the predicted results blind for various slices, and they noted the similarity and indistinguishability. Thus, the models all trained well.

One key thing to note was that the model seemed to initially want to reproduce the most recent input it was given. For example, in the R3T1 model, after one epoch of training, the training results were not good and the model was not yet learning the next step. This gradually got better, especially in timeseries where the third input image and the next step were similar. The timeseries it struggled the most with were when the three input timesteps were all dim, but the contrast agent entered in the fourth timestep. It would start by predicting dim, but after many epochs of training, it would learn to go bright. This similar pattern was reflected in the other models, too.

### 6.2. Validation & Testing

A quantitative comparison of MSE calculations for each timestep can be seen in Figure 5. From these metrics across the whole validation and test datasets, we see that the RNR model performed the best from start to finish. Other models performed comparably
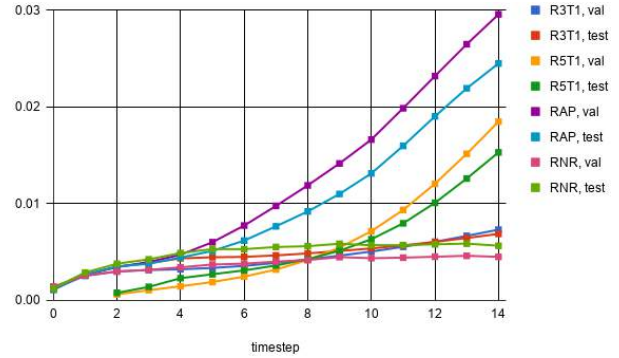


Figure 5. MSE over time for various recurrent models.

up through timestep 9, then they started to diverge.

For qualitative comparison, Figures 6 and 7 show a comparison for one timeseries. The first row is the ground truth for the predicted timesteps. The second row is the absolute difference between R3T1 and the ground truth, the third row is the absolute difference between R5T1 and the ground truth (with the first two timesteps as just black squares, since the R5T1 predicts less frames), the fourth row is the absolute difference between RAP and the ground truth, and the final row is the absolute difference between RNR and the ground truth.
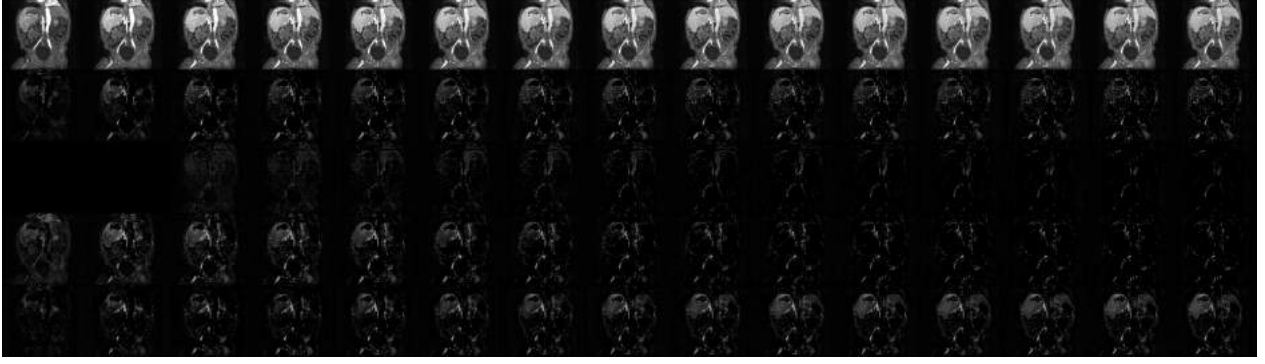
6

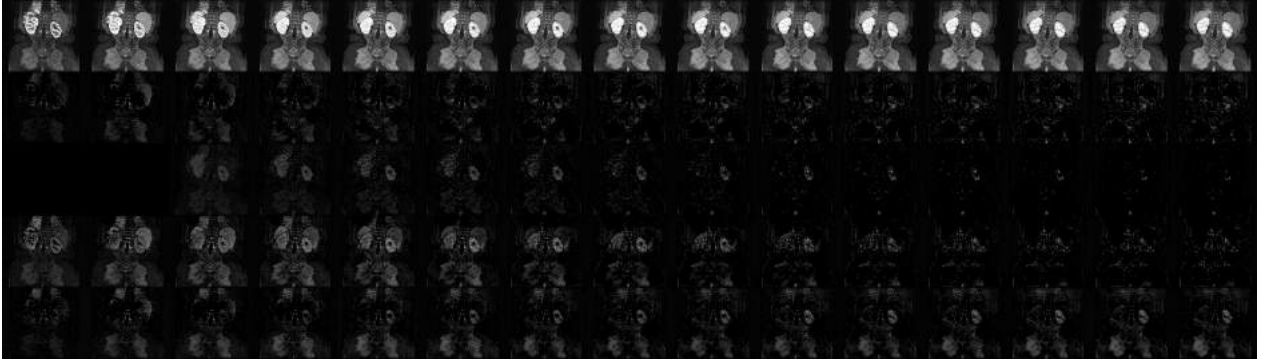Figure 6. Comparison of validation results between recurrent models.



Figure 7. Comparison of test results between recurrent models.
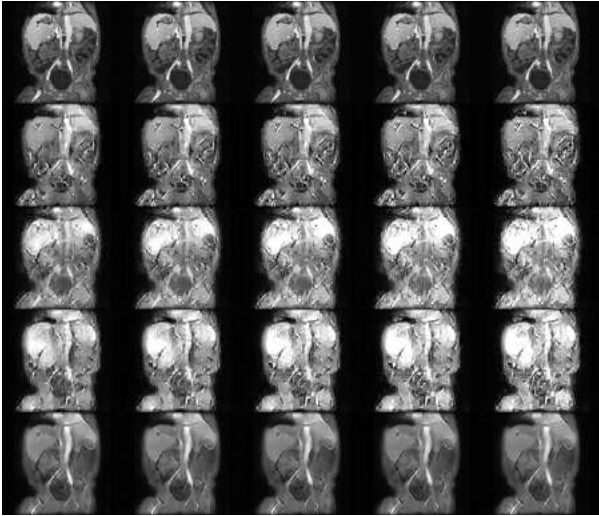


Figure 8. Latter timesteps of validation results.



Figure 9. Latter timesteps of test results.

The RAP model appeared to perform the worst, with the difference image being quite bright. The RNR model performed the best, with most of the smooth contrast areas correctly predicted, while lacking in some of the structural features. To see this result more clearly, Figures 8 and 9 show a more zoomed in view, particularly of the latter timesteps, for the output of each model compared to the ground truth. First

7

row is ground truth, second row is R3T1, third row is R5T1, fourth row is RAP, and fifth row is RNR. From a quantitative observation, it is clear that the RNR model produces the smoothest image, although all of them have fuzzy outputs by the last timesteps.

## 7. Conclusions & Future Work

Overall, my experiments showed that it is possible to use a neural network to predict time-resolved DCE MRI information, but with my current models, the performance will depend highly on the number of previous time frames provided and the quality of data used for training.

From a quantitative standpoint, the RNR model performed the strongest, with a weighted MSE-based loss equation. This conclusion correlates with my hypothesis and makes sense because the weights and biases are trained for their corresponding timesteps, not for a general prediction.

Qualitatively, though, all of the models showed promise but none were outstanding. All the models perform better when the bright spots start to occur early on, within the ground truth timesteps provided. In the cases where bright spots appear much later in the timeseries, the models all tend to do poorly. The RNR model produced the smoothest images, but conversely, it lost some of the structure.

For future work starting with these models, I would like to start by implementing more techniques to weight the loss so accurate prediction of early timesteps occurs. I have also been considering architectures where the number of timesteps passed forward is dynamic, for example: three ground truths are input to predict the fourth, then the three ground truths are concatenated with the prediction to produce the fifth, then all five are concatenated to produce the sixth, etc. The disadvantage with this architecture, though, would be, similar to the RNR model, reusing weights and biases for the encoders would not be possible for at least the first convolutional layer, since the input dimensions would vary each timestep. Theoretically, I predict it would improve the results of the RNR model.

Finally, I am also interested in attempting to implement the more complex models, such as the one proposed in [7]. Even starting with a simple GAN may produce more clear outputs which are less blurry or choppy.

## 8. Acknowledgements

## References

[1] M. Babaeizadeh, C. Finn, D. Erhan, R. H. Campbell, and S. Levine. Stochastical variational video prediction. CoRR, abs/1710.11252, 2017.

[2] C. Finn, I. Goodfellow, and S. Levine. Unsupervised learning for physical interaction through video prediction. In Proceedings of Neural Information Processing Systems (NIPS), October 2016.

[3] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan. Gan-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. CoRR, abs/1803.01229, 2018.

[4] E. Gibson, W. Li, C. Sudre, L. Fidon, D. I. Shakir, G. Wang, Z. Eaton-Rosen, R. Gray, T. Doel, Y. Hu, T. Whyntie, P. Nachev, M. Modat, D. C. Barratt, S. Ourselin, M. J. Cardoso, and T. Vercauteren. NiftyNet: a deep-learning platform for medical imaging. Computer Methods and Programs in Biomedicine, 158:113–122, 2018.

[5] T. Karras, T. Aila, S. Laine, and J. Lehtinen. Progressive growing of gans for improved quality, stability, and variation. In Proceedings of International Conference on Learning Representations (ICLR), 2018.

[6] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. CoRR, abs/1412.6980, 2014.

[7] A. X. Lee, R. Zhang, F. Ebert, P. Abbeel, C. Finn, and S. Levine. Stochastic adversarial video prediction. CoRR, abs/1804.01523, 2018.

[8] Z. Liu, R. Yeh, X. Tang, Y. Liu, and A. Agarwala. Video frame synthesis using deep voxel flow. In Proceedings of International Conference on Computer Vision (ICCV), October 2017.

[9] E. H. Meijering, W. J. Niessen, and M. A. Viergever. Quantitative evaluation of convolution-based methods for medical image interpolation. Medical Image Analysis, 5:111–126, 2001.

[10] D. G. Nishimura. Principles of Magnetic Resonance Imaging. 1.2 edition, 2016.

[11] X. Qi, Q. Chen, J. Jia, and V. Koltun. Semi-parametric image synthesis. In Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), April 2018.

[12] E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. In IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), May 2016.

[13] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W. kin Wong, and W. chun Woo. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. CoRR, abs/1506.04214, 2015.

[14] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556, 2015.

[15] N. Srivastava, E. Mansimov, and R. Salakhutdinov. Unsupervised learning of video representations using lstms. In Proceedings of International Conference on Machine Learning (ICML), July 2015.

[16] C. Vondrick, H. Pirsiavash, and A. Torralba. Generating videos wth scene dynamics. In Proceedings of Neural Information Processing Systems (NIPS), October 2016.

[17] T. Zhang, J. Y. Cheng, A. G. Potnick, R. A. Barth, M. T. Alley, M. Uecker, M. Lustig, J. M. Pauly, and S. S. Vasanawala. Fast pediatric 3D free-breathing abdominal dynamic contrast-enhanced MRI with high spatiotemporal resolution. Journal of Magnetic Resonance Imaging, 41:460–473, 2015.

**Deep Predictive Modeling of Dynamic Contrast-Enhanced MRI Data**

Magnetic resonance imaging (MRI) is a highly flexible imaging modality that allows visualization of soft tissue anatomy and function. However, although many advancements have been made in the field of MRI, inadequate spatiotemporal resolution and lengthy processing times continue to hinder accurate and rapid diagnosis of various life-threatening pediatric kidney and liver conditions, including Wilms tumors and renal cell carcinoma (RCC) in the kidneys and hepatoblastoma (HB) and hepatocellular carcinoma (HCC) in the liver [1]. The gold-standard of imaging methods used for these diagnoses is abdominal dynamic contrast-enhanced MRI (DCE-MRI), which produces a four-dimensional (x, y, z, time) dataset depicting tissue uptake of an injected Gadolinium contrast agent, as it flows through the kidneys, liver, and other organs. This data can be manually processed to measure tissue perfusion, a common biomarker for organ function. However, this task is time-consuming and susceptible to human bias and error [2].

I propose to use machine learning algorithms to expedite and enhance abdominal DCE-MRI data processing. The challenge lies in the fact that machine learning has proven to be a powerful tool for solving supervised learning tasks, where data is labeled, but success is scarce in the area of unsupervised tasks, where the objective is to learn structural information of unlabeled data. Recent work has demonstrated the effectiveness of deep predictive coding networks for unsupervised video prediction of complex natural scenes [3,4]. I believe deep models can be trained on healthy patient DCE-MRI data to predict how future frames would look, so actual data can be compared to the healthy prediction as a baseline to expedite diagnosis. Furthermore, if my model is able to predict the contrast enhancement curves per pixel, then it can segment anatomies to measure the local tissue perfusion of the organs of interest. Precise segmentation is critical for accurate perfusion calculations and disease diagnostics.

A consequential benefit of this model would be the ability to improve scan quality in real-time. Timing of data acquisition is crucial, as shown in [5]—for example, accurate capture of the late arterial phase is necessary for distinguishing between hypervascular and non-arterial lesions. If the model can predict where and when high contrast changes will occur as the scan sequence is running, parameters can be automatically adjusted to optimize for spatial or temporal resolution. Additionally, knowledge of contrast enhancement curves local to specific organs would allow for improved local anatomical resolution during data acquisition. Although advancements in MRI pulse sequence design have enabled some optimization for overall scan quality, machine learning techniques have the ability to optimize for local anatomies.

**Facilities & Collaborations**

Stanford has a rich legacy in MRI research, from Felix Bloch's fundamental discoveries in nuclear magnetic resonance to Albert Macovski's many contributions to various imaging techniques. I have spent two quarters learning from and working under Dr. Dwight G. Nishimura, Dr. Shreyas Vasanawala, and Dr. Joseph Y. Cheng, all of Stanford's Magnetic Resonance Systems Research Lab, which provides me with cutting edge imaging sequences and equipment provided by GE Healthcare. Our group's collaboration with Stanford Medicine provides access to patients and radiologists to label data. Finally, Stanford Computer Science and the Silicon Valley provide me access to the world's leading research groups in machine learning.
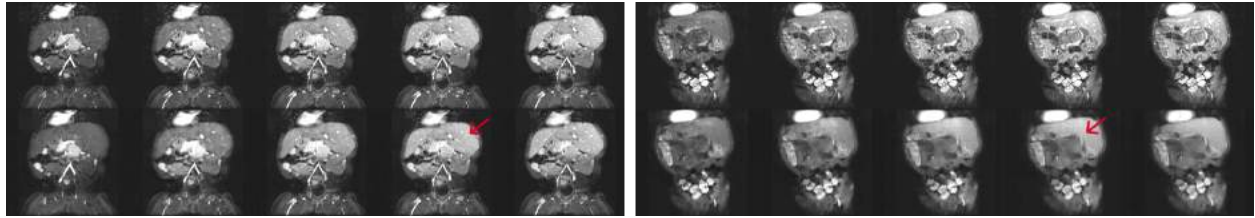
**Figure 1.** A subsection of the video prediction; the top row shows ground truth and the bottom shows predicted images. The left shows high performance on training data, but the right shows errors when tested on blind data.

### Aim 1: Training a predictive model

The first objective is to successfully create a generative model that can accurately reproduce latter frames in image sequences. In my Master's coursework, I learned about various neural network architectures and trained a recurrent convolutional spatiotemporal autoencoder that uses three frames to predict the next fifteen. The loss function used was a weighted mean-squared error. Results are shown in Figure 1. I will preprocess the sets to obtain sequences with high amounts of contrast change. The metrics for success will be mean squared error, peak signal-to-noise ratio, and radiologists scoring, which is the accepted metric in MRI.

### Aim 2: Segmenting and measuring GFR

The second objective is to produce and analyze pixel-wise contrast enhancement curves from the predictions, and train the model to segment organs of interest. Segmentation will be challenging because of undesired motion artifacts induced by respiration and natural anatomical variation in children of varying ages. Success will be determined by radiologist evaluation.

### Aim 3: Real-time MRI sequence optimization

The third objective is to work on pulse sequence optimization for both spatiotemporal and local anatomical resolution improvements. I will work with labmates who are experienced in pulse sequence programming, and I will integrate the deep predictive model into the process of acquiring the images. I envision this will start with the acquisition of ideally only one context frame, passing that into the predictive model, and automating changes to pulse sequence parameters based on prediction results. Success will be based on comparison of this modified pulse sequence with existing imaging sequences, as defined in [6].

### Broader Impacts

With my collaborators in Stanford Medicine, I will be able to facilitate clinical use and see immediate impact in expediting diagnoses. The research methodologies will be published in MRI and computer science journals, and the clinical results will be published in radiology and clinical journals. All software will be made available for shared use and development. Long-term, I hope this project will contribute to increased trust and collaboration between engineers and physicians, driving the future of medical imaging.

**References**
[1] I. Schmid, D. von Schweinitz, *J. Hepatocell. carcinoma*. **4**, 15–21 (2017). [2] F. Khalifa *et al.*, *Med. Phys.* **41**, 124301 (2014). [3] M. Mathieu, C. Couprie, Y. LeCun, (2015), *arXiv*, 1511.05440. [4] W. Lotter, G. Kreiman, D. Cox, (2016), *arXiv*, 1605.08104. [5] G. B. Chavhan, S. Shelmerdine, K. Jhaveri, P. S. Babyn, *RadioGraphics*. **36**, 1517–1532 (2016). [6] T. Zhang *et al.*, *J. Magn. Reson. Imaging*. **41**, 460–73 (2015).

# A Heuristic Approach to Automated Ostia Detection in Coronary CTA Volumes

**Abstract**

Coronary artery disease (CAD) threatens Americans as the leading cause of death. Coronary CT angiography (CTA) provides a noninvasive method for diagnosing CAD by creating 3D data of the heart. When using CTA volumes to diagnose CAD, cardiologists must first identify the coronary ostia. Current computer-aided methods have proven inefficient due to unnecessary algorithm complexity. The goal of this project is to create an algorithm that efficiently identifies the coronary ostia by using heuristics derived from CTA imaging mechanisms and heart anatomy. Three key steps to expedite the location of the coronary ostia are: ascending aorta detection, left coronary artery (LCA) ostium detection, and right coronary artery (RCA) ostium detection. The heuristics were identified from 19 training volumes. Algorithm evaluation was conducted on 38 volumes including the training volumes. The LCA and RCA ostia were detected with 100% accuracy in terms of the slice location. The accuracy of the intra-slice location of the ostia was 98.6%. This MATLAB-based algorithm, when run on an HP Pavilion dv7 laptop, averaged detection times of 11.44s over all 38 volumes. Using this algorithm, cardiologists will quickly start CAD diagnosis, which saves times, and thus, saves lives.

# A Heuristic Approach to Automated Ostia Detection in Coronary CTA Volumes

## Executive Summary

Coronary artery disease (CAD), the leading cause of death for both men and women in the US, kills over 400,000 Americans each year. To diagnose CAD, many cardiologists use coronary CT angiography (CTA), which creates 3D data of the heart. Cardiologists start the diagnosis by locating the coronary ostia, the sites where the ascending aorta opens to coronary arteries to supply the heart with oxygen-rich blood. Cardiologists use the coronary ostia as reference points to measure and document anomalies inside coronary arteries. A typical CTA volume is composed of 300 to 500 slices. An algorithm that automatically locates the coronary ostia and bookmarks the slices where ostia were captured will help cardiologists quickly start the diagnosis process. This paper presents a heuristic approach to the automated detection of coronary ostia in CTA volumes, which consists of three steps: ascending aorta detection, left coronary artery (LCA) ostium detection, and right coronary artery (RCA) ostium detection. The heuristics originate from a combination of CTA mechanisms and heart anatomy knowledge. For instance, in a sequence of slices the ascending aorta is captured consistently as a solid ellipse, and the LCA ostium always appears before the RCA ostium when navigating slices from head to toe. With the heuristics, the noise caused by non-uniformly distributed contrast material was largely suppressed and the search space for coronary ostia was substantially reduced. In the 38 CTA volumes available to this study, both the LCA and RCA ostia were detected with 100% accuracy in terms of the slice location and 98.6% accuracy of the intra-slice location. The average time of this MATLAB-based algorithm was 11.44s, on an HP Pavilion dv7 laptop. Using this algorithm, cardiologists will quickly start CAD diagnosis, which saves times, and thus, saves lives.

# 1. Introduction

In the United States, more than 400,000 people die of coronary artery disease (CAD) each year, making it the number one leading cause of death for both men and women [1]. CAD is caused by the build-up of a substance called plaque in the coronary arteries. These arteries supply oxygen-rich blood to the heart. The build-up of plaque is a condition called atherosclerosis, which plagues 80-90% of people above the age of 30 [2]. As the plaque builds up, it can harden or rupture, both of which can lead to detrimental effects. Hardened plaque within the coronary arteries limits the oxygen-rich blood flow to the heart, which can ultimately lead to heart failure or stroke [2]. If the plaque ruptures, a blood clot can form on its surface, which will also restrict the flow of blood through the coronary arteries. Ruptured plaque can also harden, leading to the narrowing of the coronary arteries. This condition is known as stenosis [1].

With the advancement of multidetector CT (MDCT) imaging technology, coronary CT angiography (CTA), as a noninvasive method for diagnosing atherosclerosis and stenosis, has been widely accepted by cardiologists over the past decade [3-5]. CTA involves the use of an MDCT scanner and intravenously administered contrast material to capture the details of heart vessels [3]. With a 64-slice MDCT, diagnostic quality can be achieved by closely following a well-defined protocol [4]. According to [5], CTA addresses the three questions that patients and physicians want answered the most: 1) Are there stenoses of the coronary arteries? CTA has the diagnostic accuracy to identify obstructive plaques, as well as high-risk non-obstructive plaques. 2) What is the best treatment strategy when stenoses are identified? CTA can assist in determining whether a patient should undergo revascularization or medical therapy. 3) Whatever the treatment approach, what is the prognosis? The evidence of the prognostic capabilities of CTA is beginning to accumulate.

When a CTA volume is given, it is a common practice that cardiologists start the diagnosis by locating coronary ostia. Coronary ostia are the sites where the ascending aorta opens to coronary arteries to supply the heart with oxygen-rich blood. Cardiologists use coronary ostia as reference points to record anomalies inside coronary arteries [6]. A typical CTA volume is composed of 300 to 500 slices; each slice is a 512 by 512 12-bit grayscale image [4,5]. When a large amount of data is presented, even well-trained experts can easily go blind [7]. An algorithm that automatically locates ostia and bookmarks the slices with ostia captured will help cardiologists quickly start the diagnosis process, which saves time, and thus, saves lives.

Segmentation of coronary arteries in CTA volumes has been actively studied for the past five years [8-10]. A region growing based approach was attempted by [8], in which fuzzy connected theory was introduced to deal with the variation of Hounsfield units (HU) caused by non-uniformly distributed contrast material. The seed for segmenting coronary arteries was selected by tracing the ascending aorta. A Hough transform based circle detection algorithm was used for detecting the ascending aorta. Both fuzzy connected theory and Hough transform are theoretically sound but computationally costly. In [9], a centerline tracking algorithm was applied to the detection of ostia using multi-scale medialness filters, which are very computationally expensive operations. Different from [8] and [9], [10] presented an approach that did not require the segmentation of the entire ascending aorta. It claimed to be a fast approach because searching for ostia was limited to a cube shaped bounding box. The bounding box, however, was created at the cost of using the marginal space learning (MSL) method that was proposed in previous work. The detection of ostia required aligning one side of the bounding box with the centerline of the aortic root, which implies the necessity of extracting the centerline of the aortic root.

This paper presents a heuristic approach to the automated detection of coronary ostia in CTA volumes. The approach consists of three steps: ascending aorta detection, left coronary artery (LCA) ostium detection, and right coronary artery (RCA) ostium detection. Section 2 briefly describes heart anatomy and CTA imaging mechanisms, and then gives a list of heuristics that were used for the methodology proposed in this research. Section 3 is dedicated to the description of the methodology. Experimental results are presented in Section 4. Section 5 discusses how the experimental results demonstrate the effectiveness of the methodology. Section 6 concludes the paper with discussion about the future work.

# 2. Materials

## 2.1 Heart Anatomy and Coronary CT Angiography (CTA)

The ascending aorta is the first section of the aorta, which starts from the left ventricle of the heart and extends to the aortic arch. Coronary ostia are the sites where the ascending aorta opens to LCA and RCA to supply heart muscle with oxygen-rich blood. Figure 1 illustrates the anatomical relationship between coronary ostia and the ascending aorta.
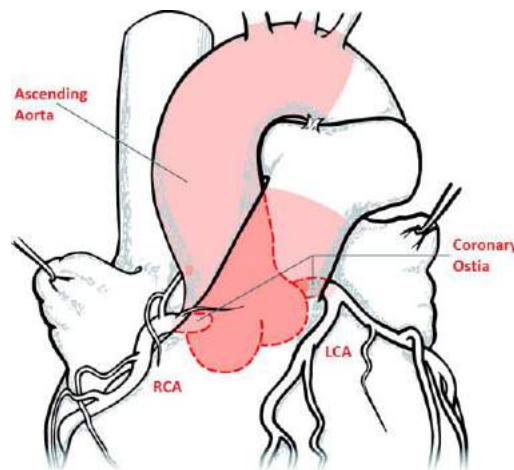


**Figure 1 A drawing that illustrates the anatomical relationship between the ascending aorta and LCA and RCA ostia. (Reproduced courtesy of http://circ.ahajournals.org/content/119/6/880/F1.large.jpg [11])**

3

CTA imaging involves the use of an MDCT scanner and intravenously administered contrast material in order to capture the details of heart vessels. The details about the protocol for creating CTA volumes can be found in [4]. A CTA volume is a collection of CT slices, with each slice a 512 by 512 12-bit grayscale image. The images are of the axial view of the chest, with the heart captured in the center of the image. The grayscale values in the images correspond to the Hounsfield scale (HU). The HU is a linear transformation of the original linear attenuation coefficient measurement into one in which the radiodensity of distilled water at standard pressure and temperature is defined as zero [12]. Figure 2 shows CT slices with the ascending aorta, LCA and RCA ostia captured. They are the 50th (ascending aorta), the 88th (LCA ostium), and the 116th (RCA ostium) slices of a CTA volume used as training data in this study.
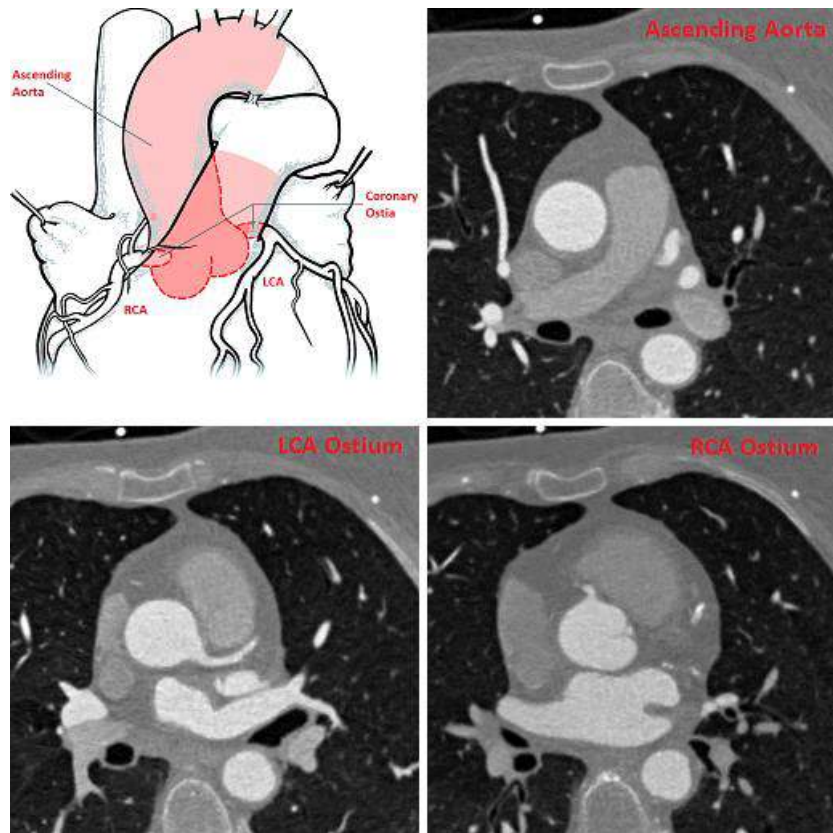


**Figure 2 Sample CT slices from a training volume with the ascending aorta and LCA and RCA ostia shown.**

## 2.2 CTA Volumes

In total, 38 CTA volumes were available for this study. Patient ages range from 34 to 86. Two-thirds of the volumes disclose gender, of which the gender ratio is approximately one to one. The number of slices per volume ranges from 301 to 507. Slice thickness ranges from 0.6 to 0.9 mm. Pixel spacing ranges from 0.2852 to 0.5371 mm. The volumes in this study are all indexed with via the LPI anatomical coordinate system [13], where x increases from right to left, y increases from anterior to posterior, and z increases from superior to inferior.
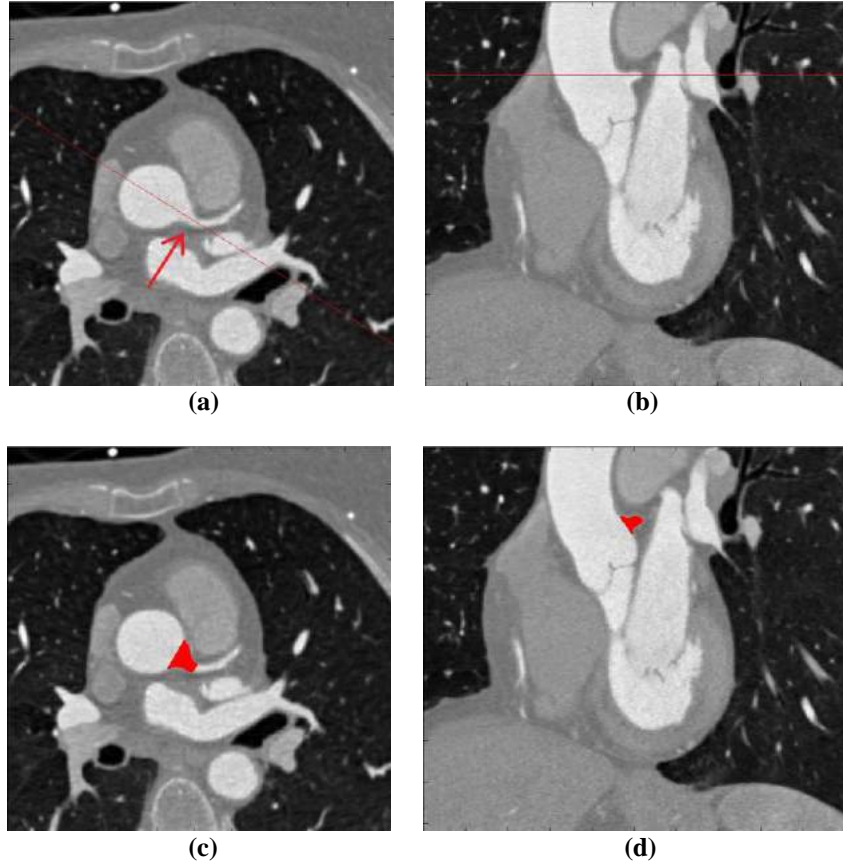


**Figure 3 (a) A slice with the LCA ostium captured. The red line and arrow indicate the way that the image in (b) was constructed. (b) An image that was constructed from the volume along the red line in (a). The red line on this image indicates the slice location of (a). (c) The area colored in red indicates the ground truth in the slice. (d) The area colored in red indicates the ground truth in the constructed image.**

The 38 volumes were randomly separated into two sets: training and testing, for the algorithm development and evaluation, respectively. Ground truths about the locations of the

5

coronary ostia in both the training and testing volumes were manually created. Figures 3 and 4 illustrate the method for creating the ground truths for LCA and RCA ostia, respectively. There were two testing volumes in which the ground truth for the RCA ostium could not be found.
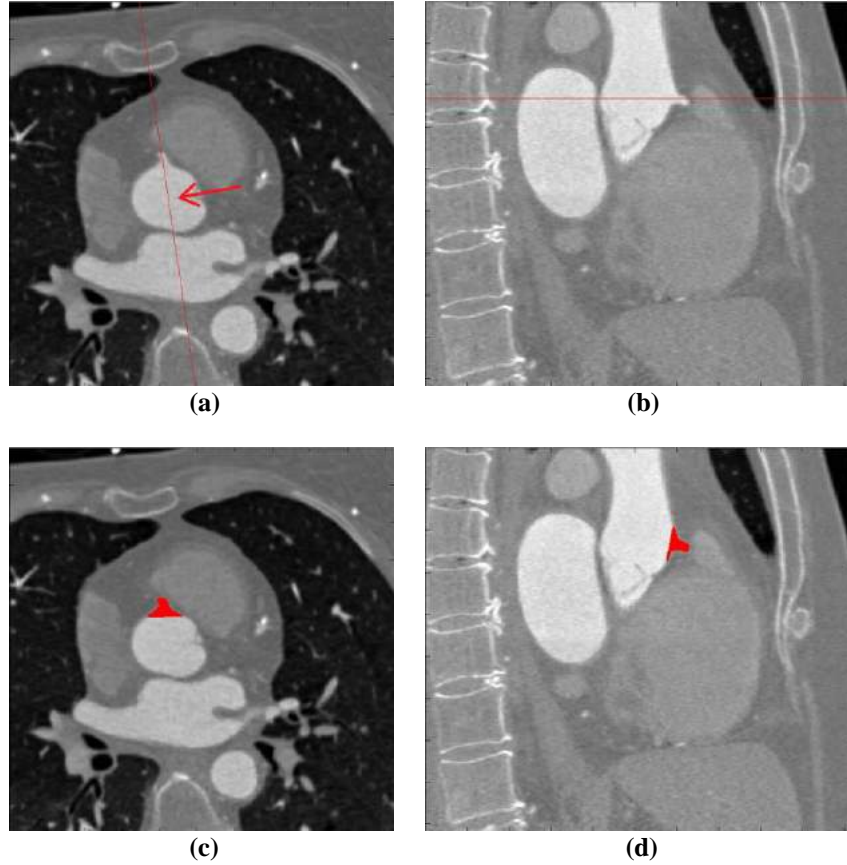


**Figure 4 (a) A slice with the RCA ostium captured. The red line and arrow indicate the way that the image in (b) was constructed. (b) An image that was constructed from the volume along the red line in (a). The red line on this image indicates the slice location of (a). (c) The area colored in red indicates the ground truth in the slice. (d) The area colored in red indicates the ground truth in the constructed image.**

## 2.3 Heuristics

The following heuristics were identified through the examination of the 19 training volumes.

- The ascending aorta is captured as a solid ellipse consecutively in the top half of the volume;

- The length of the minor axis of the ellipses, which is equivalent to the diameter of the ascending aorta, ranges from 25 mm to 38 mm;

6

- The center of the ellipses are always located in the top three fifths of the y-dimension of the slice image;

- The LCA ostium always appears before the RCA ostium when navigating slices from head to toe;

- The ascending aorta bends noticeably towards the left when navigating slices towards the LCA ostium slice;

- The RCA ostium slice is found no more than 55 mm away from the LCA ostium slice along the z-axis;

- The LCA and RCA ostia are anatomically attached to the vessel wall of the ascending aorta with the LCA ostium protruding towards the left side of patient's body and the RCA ostium protruding upwards within an angle of 135 degrees as shown in Figure 5;

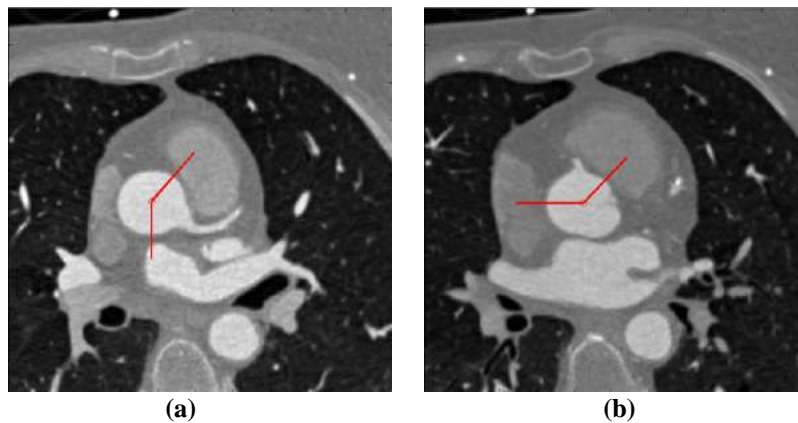- The HU values of contrast materials are between 250 and 900.



**(a)**            **(b)**

**Figure 5 The LCA ostium protrudes to the left side of patient's body within an angle of135 degree as shown in (a) and the RCA ostium protrudes upwards within an angle of 135 degree as shown in (b).**

According to [14], the HU values of contrast material in a typical CTA volume are usually 200 or higher, which confirms what was observed. Studies [15] and [16] examined CTA volumes with heart disease, while [17] examined CTA volumes of healthy asymptomatic adults. Meanwhile, [18] was conducted on non-contrast CT scans. They all reported similar statistics

regarding the diameter of the ascending aorta: male: 34+/-5.2 mm (mean+/-standard deviation), female: 30+/-3.7 mm (mean+/-standard deviation), which confirmed what was observed.

# 3. Methodology

The process of locating the coronary ostia requires a landmark because the LCA and RCA ostia locations vary tremendously across the 19 training volumes. The ascending aorta is a reasonable starting point because it remains relatively constant through the top half of each CTA volume and the coronary ostia are anatomically attached to the vessel wall of the ascending aorta.

## 3.1 Ascending Aorta Detection

According to the heuristics, limiting the search for the ascending aorta to the top half of a CTA volume will not miss the target.

One major problem to deal with was the noise caused by the non-uniformly distributed contrast material and the artifacts due to the CT reconstruction mechanisms. Noise suppression can be approached from many angles, such as by applying a Gaussian blurring filter across the slice. If blurring is attempted, however, edges of the ascending aorta and the surrounding vessels may become unclear, which will hurt the aorta detection performance in the long run. Therefore, the most effective approach was to take the mean of every pixel over a stack of slices and create a new slice that was assigned to the middle of that stack. The new slice was named the mean slice. Figure 6 shows the effectiveness of the noise suppression via the mean slice. The thicker the stack was, the more effectively the noise was suppressed in the mean slice. However, since the ascending aorta bends towards the left side of a patient's body, the ascending aorta on the mean slice may have appeared to have a blurred boundary if the stack was too thick. An optimal number that rounds up: 3mm/(slice thickness), was found to work well, experimentally, across

8

all the training volumes. For example, a volume of 300 slices and a slice thickness of 0.5 mm, would produce 25 mean slices. The number of 25 was calculated as follows. First, the number of slices per stack was calculated by dividing 3 mm by 0.5 mm, which was equal to 6. 150 slices in the top half of the volume were then divided by 6, which was equal to 25.
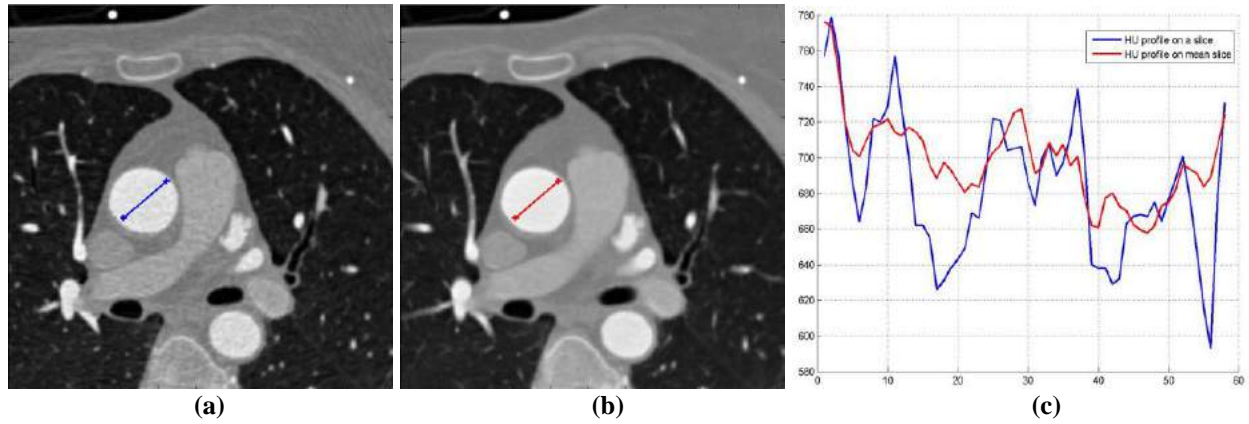


| (a) | (b) | (c) |

**Figure 6 Large variation of HU values caused by the non-uniformly distributed contrast material was effectively suppressed on the mean slice. (a) the 55th slice from a training volume; (b) an image constructed by taking the mean of slices 52 through 58 along z-axis; (c) the plot in blue corresponds to the HU values along the line in (a) and the plot in red corresponds to the HU values along the line in (b).**
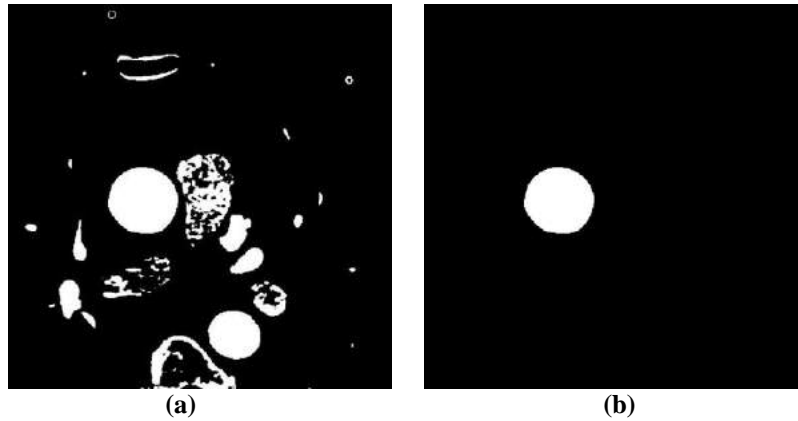


| (a) | (b) |

**Figure 7 (a) A binary image created from the mean slice in Figure 6; (b) A binary image, in which the white shape was identified as the ascending aorta.**

Although the mean slice method suppressed the noise within shapes, there was still an excessive amount of shapes with different ranges of HU values. Based on the heuristics, the HU values of the ascending aorta are between 250 and 900. Using these two values as lower and upper thresholds to create a binary image from the mean slice, a black and white image was

9

generated that effectively ruled out many irrelevant shapes. Figure 7 (a) shows the binary image created from the mean slice in Figure 6 (b).

Another disturbance when searching for the ascending aorta is the descending aorta that is also captured as a solid ellipse of similar HU values. Another way of reducing search space had to be implemented that would prevent the descending aorta from appearing. According to the heuristics, the center of the ascending aorta always exists in top three-fifths of each slice's y-axis. By applying this criterion within the algorithm, the search space for the ascending aorta was decreased, and the chance of accidentally locating the descending aorta was diminished.

Even after these methods of decreasing search space were applied, there was still a wide variety of shapes that remained. According to heuristics, the ascending aorta has a specific range that its diameter falls in. By applying the diameter thresholds to each of the shapes on the binary image, the variety of shapes was narrowed down further.

Since there was more than one mean slice, there was still a large collection of binary shapes that met all the criteria. The final step of the ascending aorta detection algorithm was the symmetry evaluation. A simple and effective method was implemented. The idea was to fold each shape in two ways: up and down and left and right. In each of the two folded images, the contrasting pixels were counted. The larger count out of the two folded images was recorded as the symmetry measure of the shape. The smaller the count was, the more symmetric the shape was. Among all the shapes in the collection, the one with the minimum symmetry measure was chosen as the ascending aorta. Figure 7 (b) shows the detected ascending aorta.

With the symmetry measure, the ascending aorta slice was always detected in front of the LCA ostium slice, which made the ascending aorta slice a good landmark.

## 3.2 LCA Ostium Detection

Since the ascending aorta was detected on the mean slice of a stack of slices, searching for the LCA ostium started from the slice right after the stack. As Figure 5 (a) in Section 2.3 shows, the LCA ostium is part of the ascending aorta; it protrudes towards the left side of patient's body within a certain angle. The LCA ostium detection problem can be solved by detecting the protrusion on the ascending aorta in the subsequent slices. The first slice found with the protrusion was chosen as the LCA ostium bearer.

The protrusion was mathematically evaluated by curvature changes along the boundary of the ascending aorta. The curvature was calculated with the formula as follows [19]:

$$k = \frac{x'y'' - y'x''}{\left(x'^2 + y'^2\right)^{3/2}} \tag{1}$$

where $x'$ and $y'$ and $x''$ and $y''$ are the first and second derivatives of $x$ and $y$, the coordinates of boundary points.

The ascending aorta that was detected in Section 3.1 was used as the reference for creating a circle mask in order to narrow down the search range and effectively rule out irrelevant shapes. Because the circle mask had to encompass all points of the ascending aorta, the maximum Euclidean distance between the boundary points and the centroid of the ascending aorta was calculated and recorded. Since the LCA ostium protrudes towards the left side of a patient's body, the circle mask was enlarged by an additional 25% to include the target: the LCA ostium. Because the ascending aorta bends, not only the center but also the length ratio of the major axis over the minor axis of the ellipse changes across the slices. A new circle mask had to be created each time for each slice by using the ascending aorta detected in the previous slice as the reference, thus creating a dynamic cylindrical search range. Figure 8 illustrates the idea of

11

using the circle masks to limit the scope of search for the LCA ostium, where both the center and

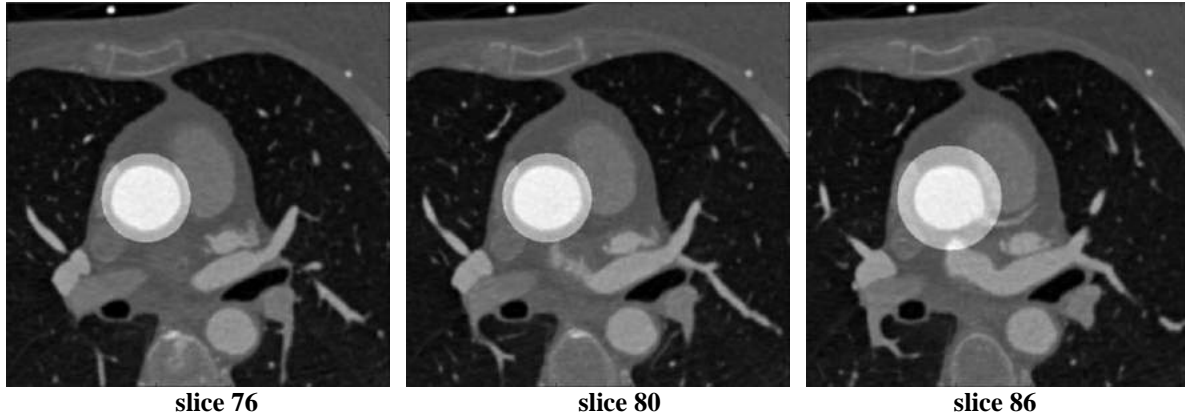the radius of the circle mask changed noticeably over the ten slices.



| slice 76 | slice 80 | slice 86 |

**Figure 8 A dynamically created circle mask is superimposed on the slice to illustrate the idea of limiting the search scope for the LCA ostium.**

The steps for identifying the ascending aorta in the slice being examined are as follows.

A binary image was first created with the same HU thresholds used in Section 3.1. A logical

"and" operation was then applied between the binary image and the circle mask image. A new

binary image was generated, in which the largest shape was selected as the ascending aorta.

As soon as the ascending aorta was identified, an 8-direction-connectedness based

boundary tracing algorithm was applied to the ascending aorta. Figure 9 (a) shows the boundary

tracing results. If formula (1) was directly applied to the boundary, the curvature signal would be

noisy. Therefore, the boundary was smoothed by taking the mean coordinate of a group of eleven

consecutive points and assigning the mean coordinate to the middle point of the group. This

process was applied to each boundary point, and the entire process was iterated five times. The

choice of eleven points and five iterations were both determined by experimentation. The

smoothed boundary is shown in Figure 9 (b). According to the heuristics shown in Figure 5 (a),

the LCA ostium is limited within a certain angle. Therefore, only the boundary segment within

the angle had to undergo the curvature calculation. Figure 9 (c) shows the boundary segment.

The curvature of the points on the boundary segment was assembled along the counterclockwise direction to create a 1-dimensional signal that is shown in Figure 9 (d). The mean and standard deviation of the curvature signal were calculated. It was found by experimentation that the LCA ostium occurred when the ratio of standard deviation over mean was larger than or equal to four.
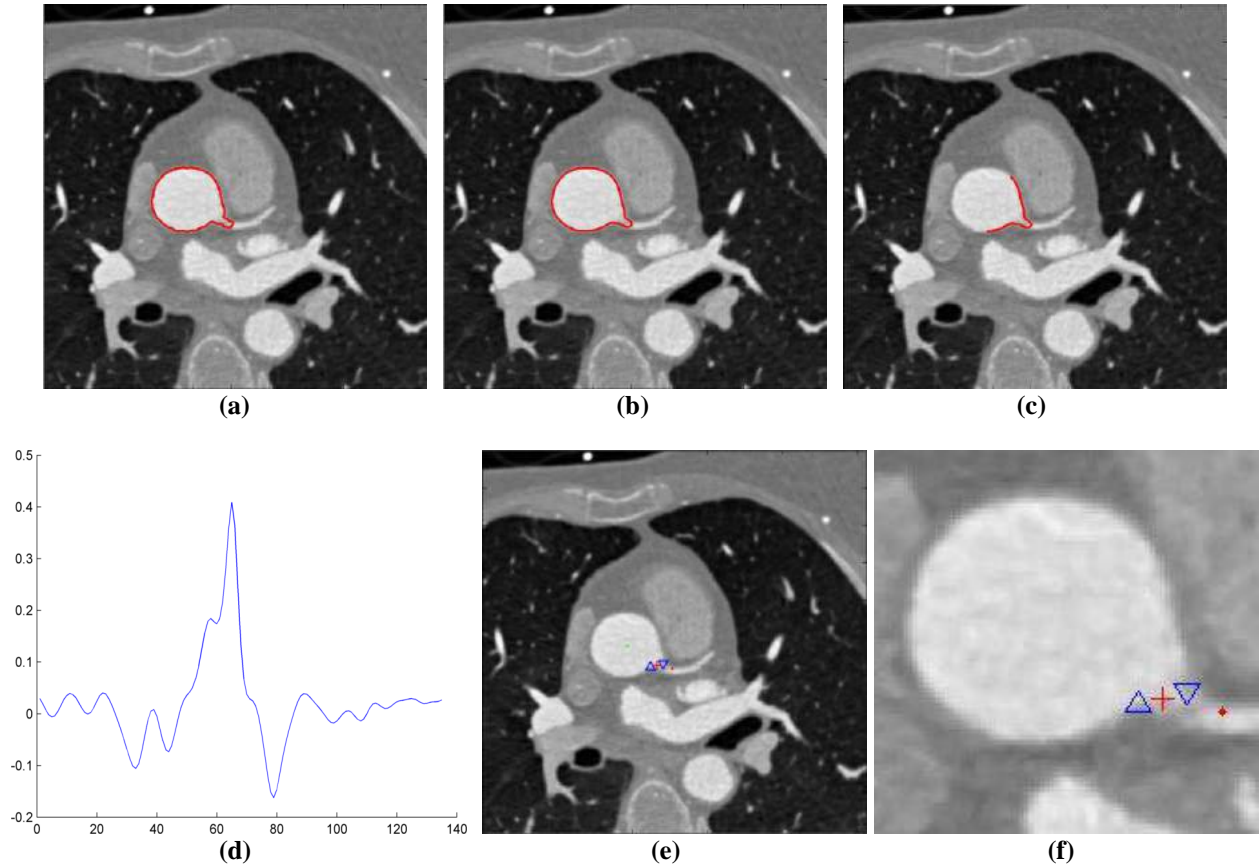


**Figure 9 (a) 8-direction-connectedness based boundary; (b) smoothed boundary; (c) boundary segment that is selected for curvature evaluation; (d) 1-D curvature signal created along the boundary segment in (c); (e) key points that are detected close to the LCA ostium; (f) the magnified view of the detection results in (e).**

Once the ratio of standard deviation over the mean met the criteria, three key points were recorded as shown in Figures 9 (e) and (f). The red dot corresponds to the peak point in Figure 9 (d). The blue upward triangle corresponds to the lowest valley in front of the peak point and the blue downward triangle to the lowest valley point after the peak point in Figure 9 (d). The red cross is a point calculated by taking the mean of the coordinates of the two blue triangle points.

13

## 3.3 RCA Ostium Detection

According to the heuristics, the RCA ostium slice appears after the LCA ostium slice. The slice

right after the LCA ostium is the starting point for searching for the RCA ostium.
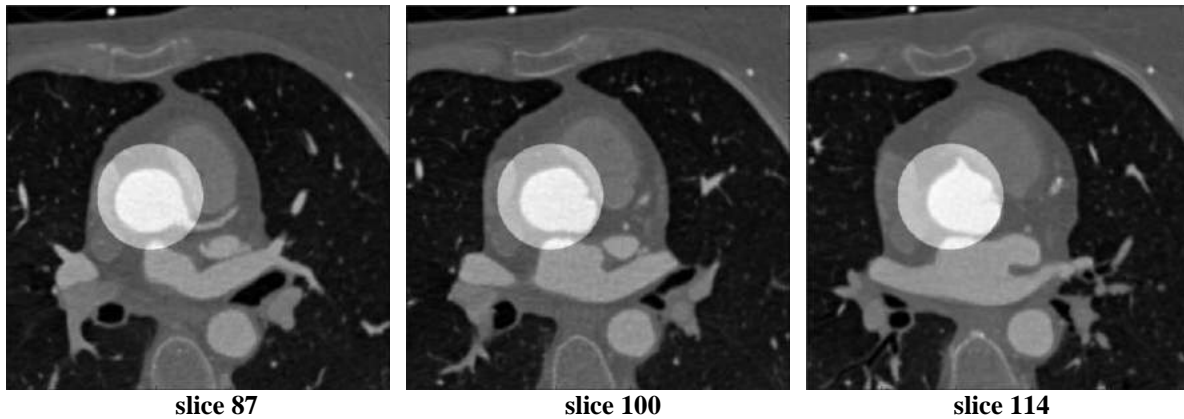


slice 87                                 slice 100                                 slice 114

**Figure 10 The same circle mask is superimposed on the slices to illustrate the idea of limiting the search scope for the RCA ostium.**

The process for detecting the RCA ostium is similar to that for detecting the LCA ostium.

But one significant difference is that the ascending aorta and the left ventricle were sometimes

captured at the same time on the RCA ostium slice, and they both have similar HU values

because the same contrast material flows inside. The ascending aorta bends to the left noticeably

up to the LCA ostium slice. A dynamic search range had to be implemented in the search for the

LCA ostium. On the contrary, the ascending aorta bends less between the LCA and RCA

because they are anatomically closer to each other; therefore, a fixed search range was able to

accomplish the goal more quickly and avoided introducing the left ventricle into the search range

unexpectedly. The ascending aorta detected in the LCA ostium slice was used as the reference

for creating the circle mask. Figure 10 illustrates the idea that the same circle mask was

repeatedly used to limit the search scope for the RCA ostium.

The rest of the steps for detecting the RCA ostium were similar to what was stated in

Section 3.2, with a few changes. One difference was the angle that was used to restrict the

curvature calculation. Another difference was the way of identifying the RCA ostium. As the slices approached the RCA ostium, noise increased, as observed in the training volumes. To prevent a false positive, a tally was set up to increase by one each time a consecutive slice that met curvature criteria was located: the largest curvature signal is three standard deviations above the mean. Once the tally reached three millimeters of slices in a row, it concluded the detection of the RCA ostium. If a slice failed to meet the criteria, the tally would restart.
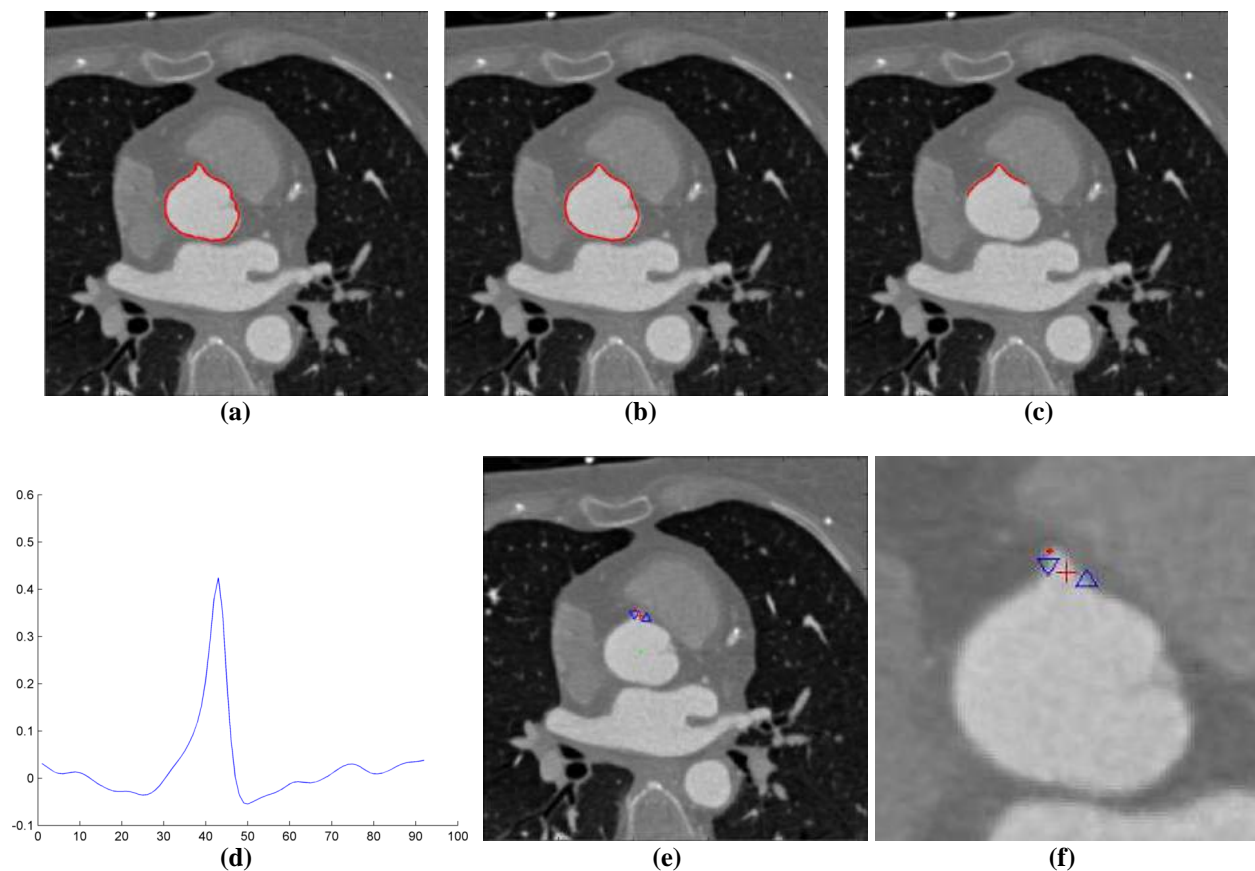


**Figure 11 (a) 8-direction-connectedness based boundary; (b) smoothed boundary; (c) boundary segment that is selected for curvature evaluation; (d) 1-D curvature signal created along the boundary segment in (c); (e) key points that are detected close to the RCA ostium; (f) the magnified view of the detection results in (e).**

Based on heuristics, the RCA ostium appears within 55 mm from the LCA ostium. If the RCA ostium was not found by that point, then the algorithm concluded that the CTA volume did not contain the RCA ostium. Figure 11 illustrates the key steps for detecting the RCA ostium.

# 4. Experimental Results

Experimental results on the training volumes are shown in Figure 12 (a), Figure 13 (a), and Figure 14 (a) and the ones on the testing volumes are shown in Figure 12 (b), Figure 13 (b), and Figure 14 (b). The images were manually created by assembling the cropped images.
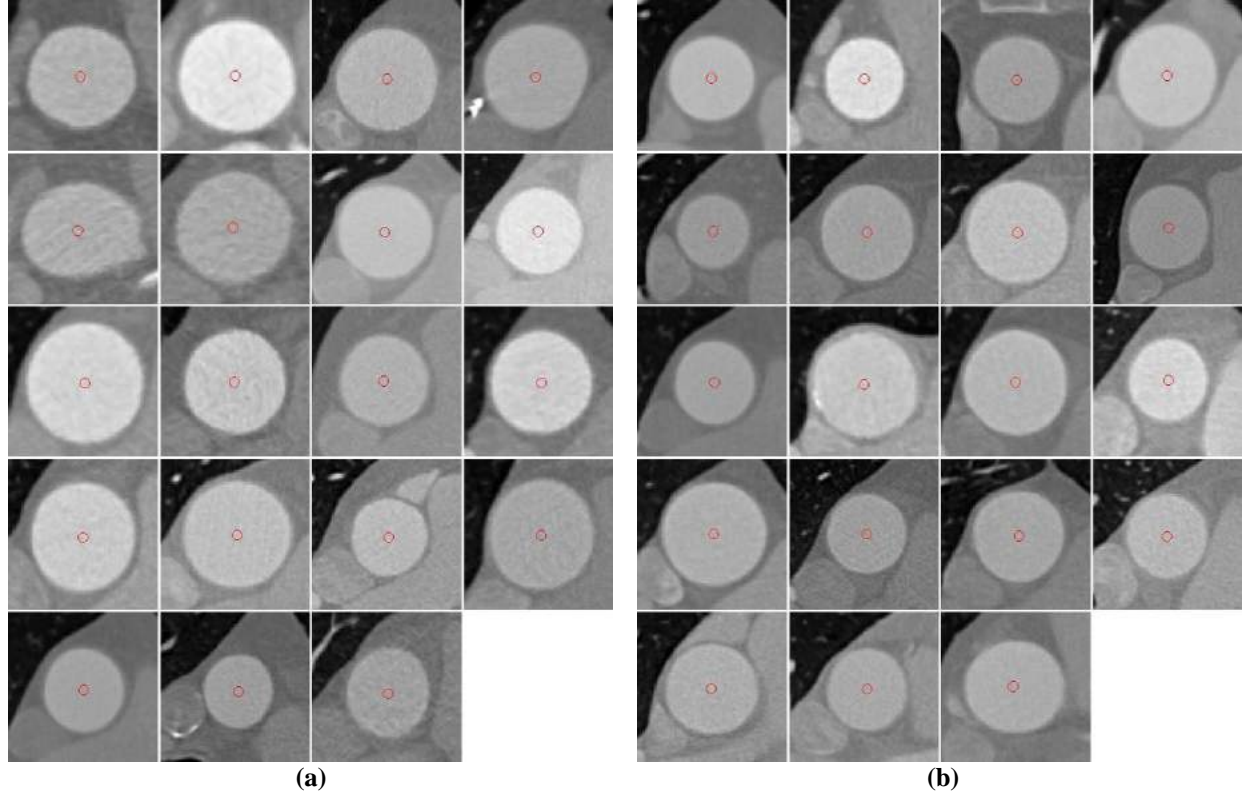


(a)                                 (b)

**Figure 12 The ascending aorta detected in (a) the 19 training volumes and (b) the 19 testing volumes.**

# 5. Discussion

The algorithms were implemented in MATLAB (R2007b) language. The system configuration is as shown in Table 1. With the 38 volumes, average time for detecting the ascending aorta, the LCA ostium, and the RCA ostium, was 4.3, 3.89, and 3.25 seconds, respectively.

Both the LCA and RCA ostia were detected 100% on the slices within the scope of the ground truths. The two testing volumes in which the RCA ostium was not found are the exact two volumes mentioned in Section 2.2. The RCA ostium detection overall performed well; at

least one of the three key points or the calculated point fell into the ground truths. The LCA ostium detection also worked fine except for the one at row three and column three in Figure 13 (b). The reason was the boundary noise caused by aorta movement at the time of the scan.

Here are the key features that made this methodology more efficient than the ones in the literature. The ascending aorta was used as a landmark, thus segmentation of the entire ascending aorta was unnecessary. The mean slice concept not only effectively suppressed noise, but also expedited the detection of the ascending aorta. Coronary ostia identification algorithms such as curvature evaluation were only applied to the slices that had a higher chance of being ostia bearers. Angle limitation on the curvature evaluation further improved the efficiency. Symmetry measurements proved to be equivalently effective and more efficient than the Hough transform method for the detection of the ascending aorta.

**Table 1 System configuration**

| Model | HP Pavilion dv7 Notebook PC |
|---|---|
| Processor | Intel(R) Core(TM)2 Duo CPU P8700 @ 2.53GHz |
| RAM | 6GB |
| System | Windows 7 Professional |

# 6. Conclusions

A heuristic approach to the automated detection of coronary ostia in CTA volumes was presented. The effectiveness and efficiency were proved with the experimental results on 38 CTA volumes available to this study. With this approach, cardiologists can jump to the ostia slices to quickly start the diagnosis process. To further assist the cardiologists, future work of this research is to study algorithms that detect plaques and stenoses and create the corresponding bookmarks in the CTA volumes. In order to do so, the immediate future work is to improve the current algorithms to refine the ostia locations so that the detected ostia can be used as reliable reference points for searching for the plaques and stenoses inside the coronary arteries.
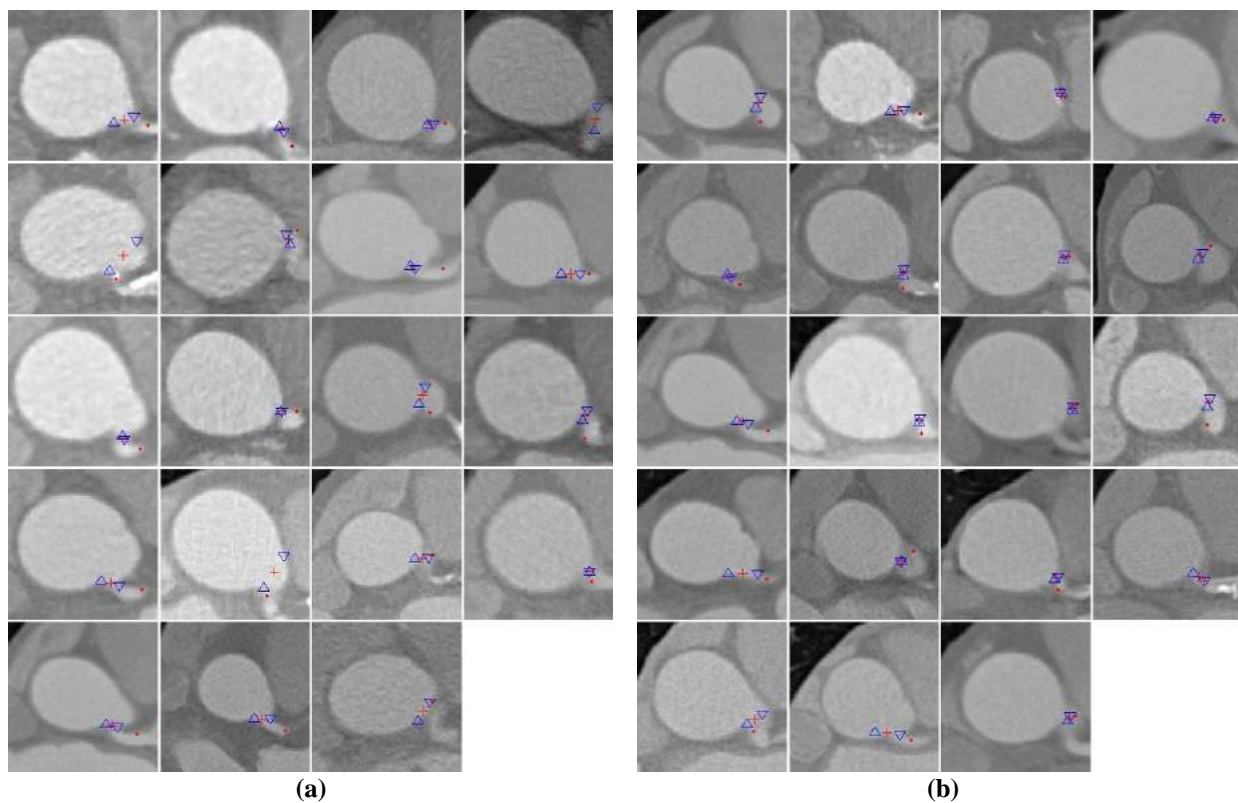
**Figure 13 The LCA ostium detected in (a) the 19 training volumes and (b) the 19 testing volumes.**
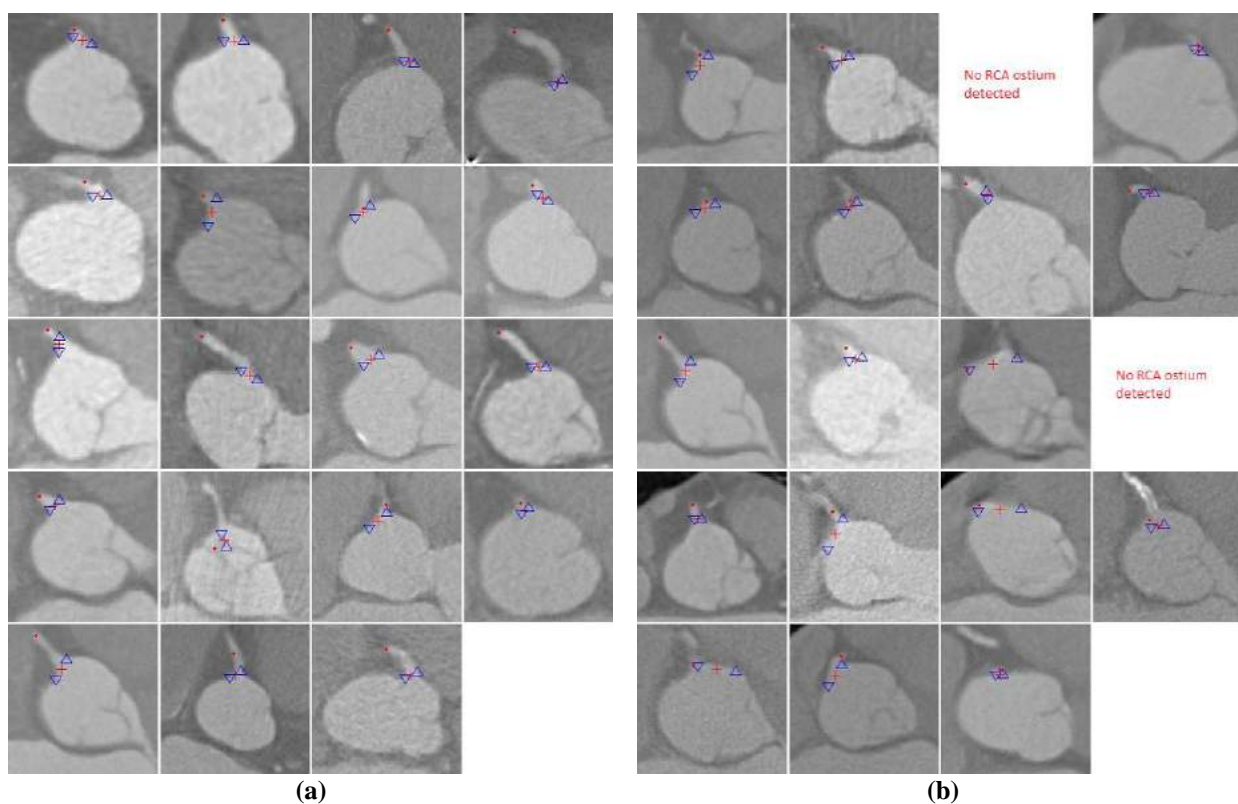


**Figure 14 The RCA ostium detected in (a) the 19 training volumes and (b) the 19 testing volumes.**

# References

1. "Who is at risk for coronary heart disease?" National Institutes of Health, (August 2012), Available HTTP: http://www.nhlbi.nih.gov/health/health-topics/topics/cad/atrisk.html.

2. "Atherosclerosis," University of Maryland Medical Center, Available HTTP: http://umm.edu/health/medical/altmed/condition/atherosclerosis.

3. B. Desjardins and E. A. Kazerooni, "ECG-Gated Cardiac CT," American Journal of Roentgenology, vol. 182, no. 4, pp. 993-1010, 2004.

4. U. Hoffmann, M. Ferencik, R. C. Cury, A. J. Pena, "Coronary CT angiography," The Journal of Nuclear Medicine, vol. 47, no. 5, pp. 797-806, 2006.

5. J. K. Min, "Coronary CTA versus cardiac catheterization: where do we stand today?" Supplement to Applied Radiology, pp. 32-40, 2006.

6. N. Funabashi, Y. Kobayashi, M. Perlroth, and G. D. Rubin, "Coronary artery: quantitative evaluation of normal diameter determined with electron-beam CT compared with cine coronary angiography – initial experience," Radiology, vol. 226, no. 1, pp. 263-271, 2003.

7. V. Hughes, "When experts go blind," National Geographic, (January 2013), Available HTTP: http://phenomena.nationalgeographic.com/2013/01/31/when-experts-go-blind/

8. C. Wang and O. Smedby, "An automatic seeding method for coronary artery segmentation and skeletonization in CTA," The Insight Journal, July 8, 2008.

9. H. Tek, M. A. Gulsun, S. Laguitton, L. Grady, D. Lesage, and G. Funka-Lea, "Automatic coronary tree modeling," The Insight Journal, August 14, 2008.

10. Y. Zheng, H. Tek, G. Funka-Lea, S. K. Zhou, F. Vega-Higuera, and D. Comaniciu, "Efficient detection of native and bypass coronary ostia in cardiac CT volumes: anatomical vs. pathological structures," Med Image Comput Comput Assist Interv, vol. 14, pt 3, pp. 403-410, 2011.

11. T. M. Tadros, M. D. Klein, and O. M. Shapira, "Ascending Aortic Dilatation Associated with Bicuspid Aortic Valve: Pathophysiology, Molecular Biology, and Clinical Implications," Circulation, Journal of the American Heart Association, vol. 119, pp. 880-890, 2009, Available HTTP: http://circ.ahajournals.org/content/119/6/880.

12. "Hounsfield Scale," Wikipedia: The Free Encyclopedia, Available HTTP: http://en.wikipedia.org/wiki/Hounsfield_scale.

13. "Coordinate Systems," Wikipedia: The Free Encyclopedia, Available HTTP: http://www.slicer.org/slicerWiki/index.php/Coordinate_systems.

14. D. Fleischmann, G. D. Robin, A. A. Bankier, and K. Hittmair, "Improved uniformity of aortic enhancement with customized contrast medium injection protocols at CT angiography," Radiology, vol. 214, no. 2, pp. 363-371, 2000.

15. R. Erbel and H. Eggebrecht, "Aortic dimensions and the risk of dissection," Heart, vol. 92, pp. 137-142, 2006.

16. H. J. Patel and G. M. Deeb, "Ascending and arch aorta: pathology, natural history, and treatment," Circulation, Journal of the American Heart Association, vol. 118, pp. 188-195, 2008, Available HTTP: http://circ.ahajournals.org/content/118/2/188.

17. S. S. Mao, N. Ahmadi, B. Shah, D. Beckmann, A. Chen, L. Ngo, F. R. Flores, Y. L. Gao, M. J. Budoff, "Normal thoracic aorta diameter on cardiac computed tomography in healthy asymptomatic adult; impact of age and gender," Acad Radiol., vol. 15, no. 7, pp. 827-834, 2008.

18. A. Wolak, H. Gransar, L. E. J. Thomson, J. D. Friedman, R. Hachamovitch, A. Gutstein, L. J. Shaw, D. Polk, N. D. Wong, R. Saouaf, S. W. Hayes, A. Rozanski, P. J. Slomak, G. Gemano, and D. S. Berman, "Aortic size assessment by noncontrast cardiac computed tomography: normal limits by age, gender, and body surface area," JACC: Cardiovascular Imaging, vol. 1, no. 2, pp. 200-209, 2008.

19. "Curvature," Wikipedia: The Free Encyclopedia, Available HTTP: http://en.wikipedia.org/wiki/Curvature