Generating Training Labels for Cardiac Phase-Contrast MRI Images

Anonymous Author(s)
Affiliation
Address
email

1 Introduction

Bicuspid aortic valve (BAV) is a highly prevalent malformation of the aortic valve that occurs in 1-2% of the population, where two leaflets of the aortic valve are present instead of the normal three. From the UK Biobank (UKBB)'s Allen et al. [2014] collection of 100,000 potential images Petersen et al. [2013], only 112 images have associated ground-truth labels, 12 of which (10.7%) are marked as depicting BAV, after being hand-labeled by collaborating cardiologists. The scarcity of training labels reflects the widespread problem of gathering enough training data to power state-of-the-art machine learning methods.

This is especially true for medical datasets, where the cost of hand-labeling by certified physicians is significantly higher than that of contracted human-intelligence workers such as the Amazon Mechanical Turk service. In our approach, we employ weak supervision based on generative models Varma et al. [2017] to address the issue of limited labels and remove the necessity for cardiologists to hand-label additional data.

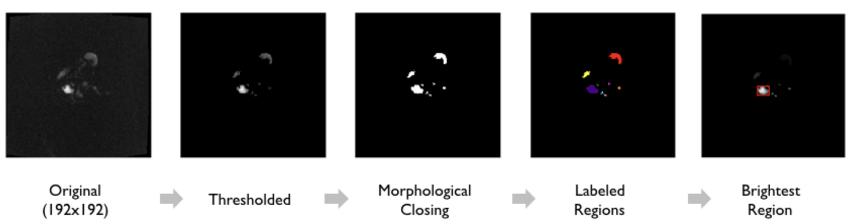

Figure 1: Pipeline from raw phase-contrast images to regions ready for feature extraction.

2 Methodology

We analyze phase-contrast images from the UKBB heart MRI dataset, which consists of 112 videos of the heart during the cardiac cycle, with 100 healthy and 12 BAV subjects. Since the phase-contrast images target the aortic valve, the point of concentrated blood flow during the cardiac cycle, the valve is the brightest portion of the image. We select the six brightest frames for our analysis, resulting in 600 frames of healthy patients and 72 frames of patients with BAV.

We experimented with several thresholding techniques to isolate regions of interest and denoise the image background. The resulting binary masks were applied to the original images so that the background was removed, and only primary regions of interest remained. Next, we filled in noisy holes of each thresholded image by computing the Otsu threshold to apply a morphological closing, such that each blob region would be treated as a discrete geometric shape for feature extraction Van der Walt et al. [2014]. We determined region labels by tracing connected neighbors of pixels in the scene, before finally selecting the region with the greatest intensity as our region of interest.

Submitted to 31st Conference on Neural Information Processing Systems (NIPS 2017). Do not distribute.

Table 1: Heuristic Function Evaluation Results

	Statistics			
Heuristic Function	Coverage	AUC	Accuracy	F1 score
`HF_Area`	0.4479	0.6325	0.8040	0.4158
`HF_Perimeter`	0.6235	0.6494	0.8186	0.4571
`HF_Eccentricity`	0.6190	0.5593	0.7380	0.2781
`HF_Intensity`	0.5446	0.5278	0.5574	0.2286

Table 2: Label Generation Evaluation Results

	Statistics			
Method	Coverage	AUC	Accuracy	F1 score
Majority Vote	0.7917	0.7026	0.6235	0.3077
Generative Model	**1.0**	**0.7089**	**0.8185**	**0.4020**

2.1 Weak Supervision

Physiologically, we expected the area and perimeter of BAV images to be smaller than those of normal images. After our preprocessing steps, however, we noticed that it was not uncommon for the region labeling to overestimate the size of the aortic valve. We suspect that this is a result of the irregularity in the shape of BAVs. We also expected eccentricity values to be greater based on physiological intuition, and intensity to be greater, because a smaller valve typically leads to more blood flow. For each of our features, we wrote a heuristic function (HF) to separate BAV and normally-classified images.

We use the Coral paradigm Varma et al. [2017] to learn dependencies and accuracies of the HFs. Coral uses a factor graph to encode the relationships between HFs, features, and class labels. It then uses this model to assign probabilistic training labels to the data, which can be used to train any end machine learning model.

3 Experimental Evaluation

We experimentally quantify the effectiveness of our approach by considering the coverage, AUC, accuracy, and F1 score of our predicted labels compared to ground truth labels. Because of the class imbalance of this task, we must consider optimizing for metrics that prioritize recall. Intuitively, it is crucial that we properly recall a high quantity of images with BAV for downstream classifiers. The class imbalance in our data translates into trade-offs when considering the marginal threshold for converting probabilistic labels into true labels for evaluation of the generative model. We compare the outputs of the Coral generative model, which learns accuracies and inter-heuristic dependencies, to a simple majority vote (MV) system and show the resulting performance in Table 2.

4 Future Steps

We have proposed a pipeline to preprocess phase-contrast images targeting aortic valves and generate relevant heuristic functions. In doing so, we lessen the semantic gap between cardiologists' diagnostic intuitions for labeling aortic valve data and a machine's ability to automate the label generation process. We have room to improve the output of our generative model by including additional prepreocessing steps to more accurately label valve regions. We can also normalize feature values, allowing the absolute thresholds of our HFs to operate more effectively. Finally, we plan to use the probabilistic labels of our generative model to train deep convolutional networks for the BAV classification task.

References

N. E. Allen, C. Sudlow, T. Peakman, R. Collins, et al. Uk biobank data: come and get it, 2014.

S. E. Petersen, P. M. Matthews, F. Bamberg, D. A. Bluemke, J. M. Francis, M. G. Friedrich, P. Leeson, E. Nagel, S. Plein, F. E. Rademakers, et al. Imaging in population science: cardiovascular magnetic resonance in 100,000 participants of uk biobank-rationale, challenges and approaches. *Journal of Cardiovascular Magnetic Resonance*, 15(1):46, 2013.

S. Van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu. scikit-image: image processing in python. *PeerJ*, 2:e453, 2014.

P. Varma, B. He, P. Bajaj, I. Banerjee, N. Khandwala, D. L. Rubin, and C. Ré. Inferring Generative Model Structure with Static Analysis. *ArXiv e-prints*, Sept. 2017.