# Recognizing cardiac magnetic resonance acquisition acquisition planes using finetuned convolutional neural networks

Jan Margeta, Antonio Criminisi, Daniel C. Lee, Nicholas Ayache

To appear in *Computer Methods in Biomechanics and Biomedical Engineering*
Vol. 00, No. 00, Month 20XX, 1–10

# Recognizing cardiac magnetic resonance acquisition acquisition planes using finetuned convolutional neural networks

J. Margeta[a*], A. Criminisi[b], D. C. Lee[c] and N. Ayache[a]

[a]*Asclepios, Inria Sophia Antipolis, 2004 Route des Lucioles, Sophia Antipolis, France*; [b]*Machine learning and perception, Microsoft research, Cambridge, UK*; [c]*Feinberg Cardiovascular Research Institute, Northwestern University, Feinberg School of Medicine, Chicago, USA*;

(*v0.1 released November 2014*)

In this paper we propose a convolutional neural network-based method to automatically wrangle missing or noisy cardiac acquisition plane information from magnetic resonance (MR) images. This is an important building block to organise and filter large collections of cardiac data prior to analysis. In addition it allows us to merge studies from multiple centers, to perform smarter image filtering, to select the most appropriate image processing algorithm, and to enhance visualisation of cardiac datasets in content based image retrieval.

We propose to use a finetuned convolutional neural network initially trained on a large natural image recognition dataset (Imagenet ILSVRC2012) to learn feature representations for better recognize cardiac views prediction and contrast this to a previously introduced method using classification forests and features learned from an augmented set of image miniatures.

We validated this algorithm on two different cardiac studies with 200 patients and 15 healthy volunteers respectively. Our new approach significantly improves the state of the art of image-based cardiac view recognition (97.66% F1 score). Despite the large number of the network's parameters, the algorithm does not overfit and performs quite well on another independent cardiac study.

## 1. Introduction

Optimal cardiac planes depend on global positioning of the heart in the thorax. This is more vertical in young individuals and more diaphragmatic in elderly. Instead of comonly used body planes (coronal, axial and sagital) the cardiac MR images are therefore acquired along several oblique directions aligned with the structures of the heart. Imaging in these standard cardiac planes ensures efficient coverage of relevant cardiac territories and enables comparisons across modalities, thus enhancing patient care and cardiovascular research.

Automatic recognition of this metadata is essential to appropriately select image processing algorithms, to group related slices into volumetric image stacks, to enable filtering of cases for a clinical study based on completeness, to help with interpretation and visualisation by showing the most relevant acquisition planes, and in content based image retrieval for automatic description generation. Although this orientation information is sometimes encoded within two DICOM image tags: *Series Description (0008,103E)* and *Protocol Name (0018,1030)*, it is not standardised, operator errors are frequently present, or this information is completely missing. Searching through large databases to manually cherrypick relevant views from the collections is therefore tedious.

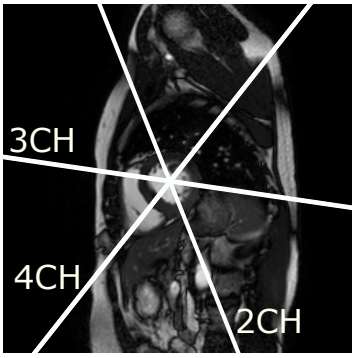
Difficulty with automated recognition - variability among the patients, different organs can be seen based on the plane orientation.

*Corresponding author. Email: jan@kardio.me

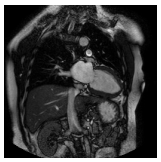
### 1.1 *Cardiac acquisition planes*

An excellent introduction to standard MR cardiac acquisition planes can be found in Taylor and Bogaert (2012). These planes are often categorized into two groups - the short and the long axis planes. In this paper we concentrate on the five most used cardiac planes acquired with steady-state free precession (SSFP) acquisition sequences (See fig. 1).

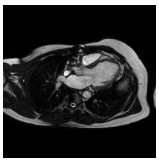
The left ventricular **short axis slices** (fig. 1(a)) are parallel to the mitral valve ring. These are acquired regularly spaced from the cardiac base to the apex of the heart, often as a cine 3D+t stack. These views are excellent for reproducible volumetric measurements or radial cardiac motion analysis but their use is limited in atrio-ventricular interplay or valvular disease study. The **long axis acquisition planes**: the 2-chamber, 3-chamber, and 4-chamber views (figs. 1(b) to 1(d)) are used to visualize different regions of the left atrium, mitral valve apparatus, and left ventricle. The 3-chamber and left ventricular outflow tract (fig. 1(e)) views provide visualization of the aortic root from two orthogonal planes. The 4-chamber view enables visualization of the tricuspid valve and right atrium.



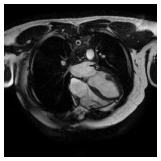

(a) Short axis (SAX)



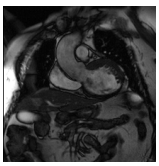
(b) Two chamber (2CH)



(c) Three chamber (3CH)



(d) Four chamber (4CH)



(e) Left ventricular outflow tract (LVOT)

Figure 1. Examples and mutual positioning of the short and the main left ventricular long axis cardiac MR views. See also fig. 9.

### 1.2 *Previous work*

The prior work on cardiac view recognition has been concentrated mainly on real-time recognition of cardiac planes for echography: Otey, Bi, et al. (2006), Park and Zhou (2007), Park, Zhou, et al. (2007), and Beymer and Syeda-Mahmood (2008). There exists some work on magnetic resonance by Zhou, Peng and Zhou (2012) and by Margeta, Criminisi, et al. (2014). The common methods are based on dynamic active shape models (Beymer and Syeda-Mahmood 2008), require to train part (Park, Zhou, et al. 2007) or landmark detectors (Zhou, Peng and Zhou 2012). Therefore any new view will require these extra annotations to be made. Otey, Bi, et al. (2006) avoided this limitation by training a ultrasound cardiac view classifier using gradient based image features. Margeta, Criminisi, et al. (2014) trained classification forests to predict the cardiac planes directly from cardiac MR image miniatures.

The state of the art in image recognition was heavily influenced by the seminal work of Krizhevsky, Sutskever and Hinton (2012) who trained a large (60 million parameters) convolutional neural network (CNN) on a massive dataset consisting of 1,2 million images and 1000 classes. They employed two major improvements: Rectified linear unit nonlinearity to improve convergence, and dropout to reduce the effect of overfitting.

Training a large network from scratch without a large number of samples still remains a challenging problem. Though a trained CNN can be adapted to a new domain by reusing already trained hidden layers of the network. It has been shown by Razavian, Azizpour, et al. (2014) that the classification layer of the neural net can be stripped, and the hidden layers can serve as excellent image descriptors for a variety of computer vision tasks (such as for photography style recognition by (Karayev, Trentacoste, et al. 2013)). Alternatively the prediction layer model can be replaced by a new one and the network can be finetuned through backpropagation.

## 2. Methods

Inspired by the recent success of deep convolutional neural network approaches we compare the previous classification forest based approach (Margeta, Criminisi, et al. (2014)) to a CNN trained on the ImageNet dataset and finetuned for cardiac view recognition. Similarly to the previous work, that learns an appropriate feature representation in order to recognize the acquisition planes directly from cardiac magnetic resonance images. Apart from the view annotation of the training set, no other information is necessary to train the classification models. To increase the number of the training samples we augment the dataset with small label preserving transformations such as translations, rotations and scales.

### 2.1 *From DICOM slice normals towards image-based features*

Both Zhou, Peng and Zhou (2012) and Margeta, Criminisi, et al. (2014) showed that where the *DICOM orientation (0020,0037)* tag is present we can use it to predict the cardiac image acquisition plane (See Fig. 2). It is sufficient to compute the plane normal vectors as a cross-product of the two image orientation vectors specified in the tag. We can feed these three-dimensional feature vectors into any classifier, in our case a classification forest. This method is represented in the results as *"DICOM orientation"*.

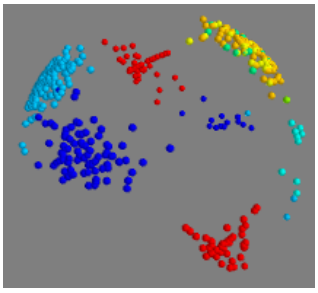


Figure 2. DICOM plane normals for different cardiac views (Margeta, Criminisi, et al. (2014)). In our dataset distinct clusters can be observed (best viewed in color). Nevertheless the separation might not be the case for a more diverse collection of images. Moreover as we cannot always rely on the presence of this tag an image-content based recognizer is necessary.

In the absence of the orientation tag we have to rely on the image intensity information only. We use $2D$ image slices individually rather than $3D$ or $3D + t$ volumes in order to allow the use of this recognizer on all view recognition dependent applications. In the following, our image based methods are presented.

### 2.2 *Classification forests and image miniature features (Margeta, Criminisi, et al. (2014))*

This method is posed as a standard image recognition framework where features are extracted from the images and are then fed to a classifier (a classification forest). Each image is therefore seen as a data point in a potentionally high-dimensional feature space.

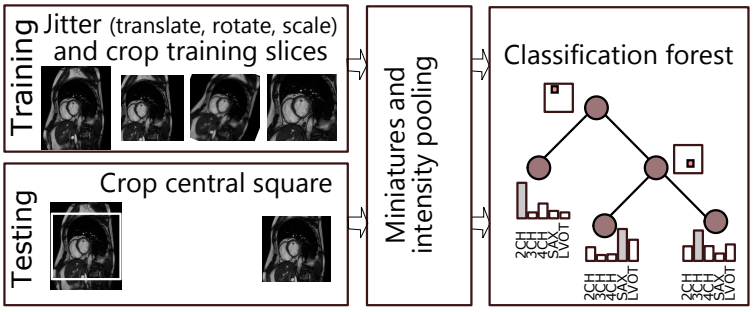



Figure 3. Classification forest pipeline for view recognition from image miniatures (Margeta, Criminisi, et al. (2014)). Discriminative pixels from image miniatures are chosen from a random pool as features for a classification forest. The training dataset is augmented to improve robustness to the differences in the acquisitions without the need for extra data and labels.

*Classification forests*

Classification forest (Breiman (1999)) is an ensemble learning method that constructs a set of randomized binary decision trees. Its advantage is computational efficiency and automatic selection of relevant features. At training time the tree structures are optimized by recursively partitioning the data points into the left or the right branches such that points with the same label points are grouped together and different labels are put apart. Each node of the tree sees only a random subset of the complete feature set and greedily picks the single most discriminative feature. This helps to make the trees in the forest different from each other. This leads to better generalization. At test time features chosen at the training stage are extracted and the images are passed through the trees and reach a set of leaves. Class distributions of all reached leaves are averaged across the forest and the most probable label is selected as the image view. See Criminisi, Shotton and Konukoglu (2011) for a detailed discussion on decision forests.

*Using image miniatures*

The radiological images are mostly acquired with the object of interest in the center. Therefore some rough alignment of structures can be expected (See fig. 4). Image intensity samples at fixed positions (even without registration) therefore provide strong cues about the position of different tissues (e.g. symmetric dark lungs or bright cavity in the center).

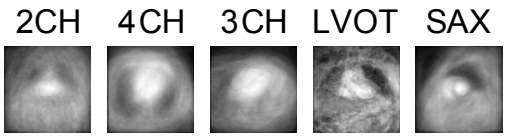



Figure 4. Averages of the DETERMINE dataset for each cardiac view after cropping the central square. A reasonable alignment of cardiac structures can be observed in the dataset. The cardiac cavities and main thoracic structures can be seen.

The only two preprocessing steps were therefore image central square extraction, resampling to 128x128 pixels and linear rescaling of the intensities to range between 0 and 255. We subsampled the cropped centers to two fixed sizes (20x20 and 40x40 pixels) and in addition we created a set of pooled image miniatures from each cropped center by dividing the image into non-overlapping 4x4 tiles and computing intensity minimum and maximum in each of these tiles (fig. 5).

The pooling added some invariance to small image translations and rotations (whose effect is within the tile size). The pixel values from these miniatures were then sampled at random positions and used directly as features. At each node of the tree 64 different locations and channels were tested and a simple threshold on this value is used to divide the data points into the left or the right partitions.

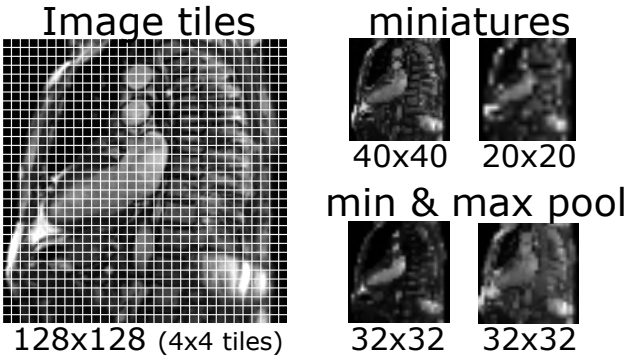



Figure 5. Image miniatures by min and max pooling local intensities (Margeta, Criminisi, et al. (2014)).

*Augmenting the dataset with jittering*

While the object of interest is in general placed at the image center, differences between various imaging centers and positioning of the heart on the image exist. The proposed miniature features are not fully invariant to these changes. To account for this the training set was augmented with extra images created by transforming the originals. These were translations (all shifts in x and y

between -10 and 10 pixels for a 5x5 grid), but also rotations (angles between -10 and 10 degrees with 20 steps) and scales (1-1.4 zoom factor with 8 steps while keeping the same image size). Note that the dataset augmentation extra expense is present mainly at the training time. The test time remained almost unaffected except that now a deeper forest could be learnt. Results using this method are presented in the evaluation as *"image forest"*.

### 2.3 *Convolutional neural network for view recognition*

The forest based method initially performed quite well but does not seem to improve with more data. In fact, it performs much worse. To further improve the performance we turned to the state of the art of image recognition, the convolutional neural networks. Their principle is rather simple. Results of several strided image convolutions are fed through a nonlinear function (in this case a rectifier), responses are locally aggregated with max pooling and finally fully connected logistic regression layers are stacked and connected to a multiway softmax in order to predict the target label. The role of the max pooling is similar to the forest based method i.e. to aggregate local responses and to allow some invariance of the input to small transformations. The parameters of the network such as parameters of the convolutional filters are optimized through backpropagation; using stochastic gradient descent the average batch prediction loss (soft-max) is decreased.

Dropout (Hinton, Srivastava, et al. 2012) was used during the training phase in the fully connected layers to reduce complex coadaptations of the neurons and to prevent overfitting. The output of each hidden neuron was set to 0 with probability 0.5. The dropped out neurons do not contribute to the forward pass and are skipped in the back-propagation. So for every input fed through the network, a different architecture of the network is sampled. At test time, all neurons are used but their output responses are multiplied by 0.5 to compensate for the difference in the topology.

#### *2.3.1 Extracting raw CaffeNet features*

We computed a new mean for our dataset (see fig. 6) and subtracted from all training and testing images prior to entering the CNN. We extracted the 4096-dimensional features from the last fully connected layer (fc7 - see fig. 7) of a slightly modified Krizhevsky's convolutional neural network model (bvlc_reference_caffenet in Caffe deep learning framework (Jia, Shelhamer, et al. 2014)) pretrained on ImageNet (Russakovsky, Deng, et al. 2014) and used them for classification by both random forests and linear support vector machines. The CaffeNet differs from Krizhevsky's AlexNet mainly in the order of local response normalization and max pooling operations.

Figure 6. The DETERMINE dataset mean from one of the training folds (inverted and stretched contrast). The short axis slices clearly dominate the dataset.

The cardiac images are resized to 256x256 squares regardless of their input dimensions and converted to rgb 256x256x3 by replication into the third dimension. The central (227x227x3) crop of the image is then fed forward through the network and the responses at fc7 are extracted. These features were adapted to a natural image recognition task but they never saw a cardiac MR image before. Similar to the work of Karayev, Trentacoste, et al. (2013) for photography style recognition we train a classifier with these features. These features already perform quite well for the cardiac view problem and are present in the validation as *CaffeNet fc7 Forest* and *CaffeNet fc7 SVM*.

### *2.3.2 Finetuning the network*

We compare the previous CNN approach with another one. Normally many examples are needed to train a large capacity neural network. However, in our case by starting from the weights of a pretrained network we can just finetune the network parameters with new data and adapt it to the new target domain. Here we used the pretrained CaffeNet (Jia, Shelhamer, et al. 2014) and replaced the last multinomial regression layer (1000-way) with a 5-nomial one (See fig. 7).

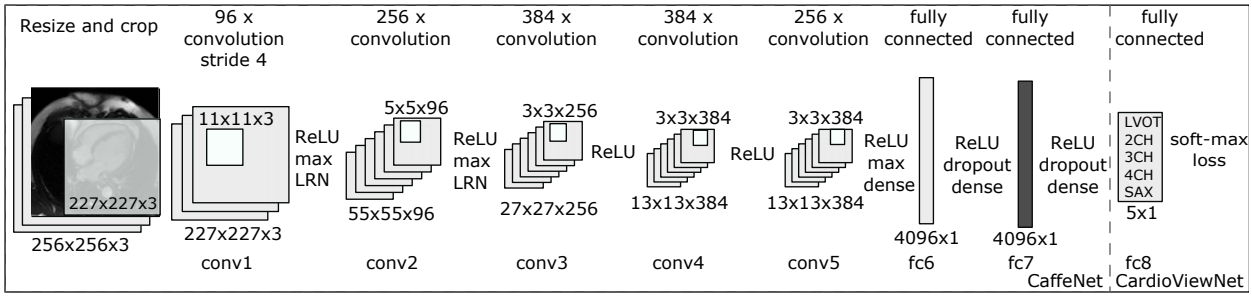
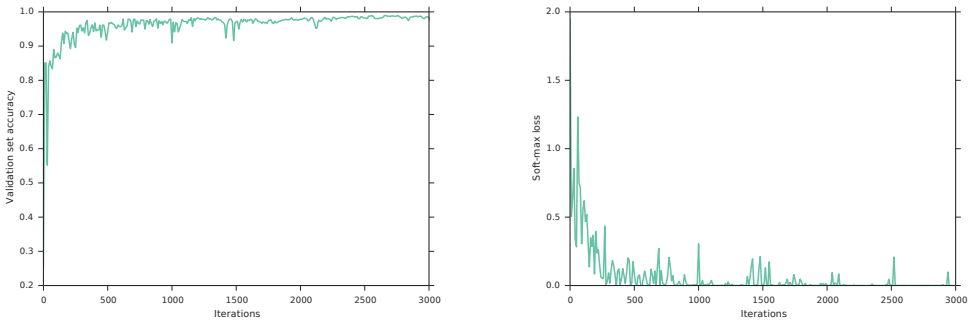



Figure 7. Our CardioViewNet is based on CaffeNet network structure and is adapted for cardiac view recognition. We initialized the network weights from a pretrained CaffeNet (Jia, Shelhamer, et al. 2014). We then replaced the last 1000-dimensional logistic regression layer (previous fc8) with a 5-dimensional to suit our view classification needs. Then we finetuned the network with our data. We also extracted features from the last 4096-dimensional fully connected layer fc7 (in dark gray) from both CaffeNet and our finetuned CardioViewNet and used them with a decision forest and a linear SVM classifier. We achieve the best performance with our finetuned network directly.

The net is finetuned with stochastic gradient descent with less aggressive parameters than those used to train with the full ImageNet dataset ($\gamma$: 0.1, stepsize: 20000, momentum: 0.9, and weight decay: 0.0005). A set of resized 256x256x3 images is used for training. At each iteration a batch of 32 random 227x227x3 crops is extracted from the resized cardiac slices and is fed forward through the network. Compared to the implementation of our forest based method, these are cheaply extracted on fly. This optimisation is run for 8000 iterations (See fig. 8. already after 3000 iterations the prediction error on the validation dataset reaches a plateau.).




(a) Mean prediction validation set accuracy

(b) Soft-max loss.

Figure 8. Finetuning our CardioViewNet model rapidly converges to its best performance.

The finetuning is quite efficient and takes approximately 4 hours on a single NVIDIA Tesla M2050 GPU. Results of this method are presented as *CardioViewNet*. Results for features extracted at fc7 from this finetuned net are listed as *CardioViewNet fc7 SVM* and *CardioViewNet fc7 Forest*.

## 3. Validation

We trained and validated this method on a dataset of slices from 200 patients from a multi-center study on post myocardial infarction hearts **DETERMINE** (Kadish, Bello, et al. (2009)).

We used only the steady state free precession magnetic resonance acquisitions. Compared to the previous work (Margeta, Criminisi, et al. (2014)) our dataset is roughly twice as big, comprising a total of 3268 image slices from 200 cardiac patients (2CH: 235, 3CH: 225, 4CH: 280, LVOT: 12, SAX:2516). The SAX slices cover the heart from the base to the apex. We also validated our models on the **STACOM motion tracking challenge** dataset (Tobon-Gomez, De Craene, et al. (2013)) containing 15 patients (2CH:15, 4CH:15, SAX:207). The LVOT views are severely underrepresented and serve as a test case for learning from very few examples.

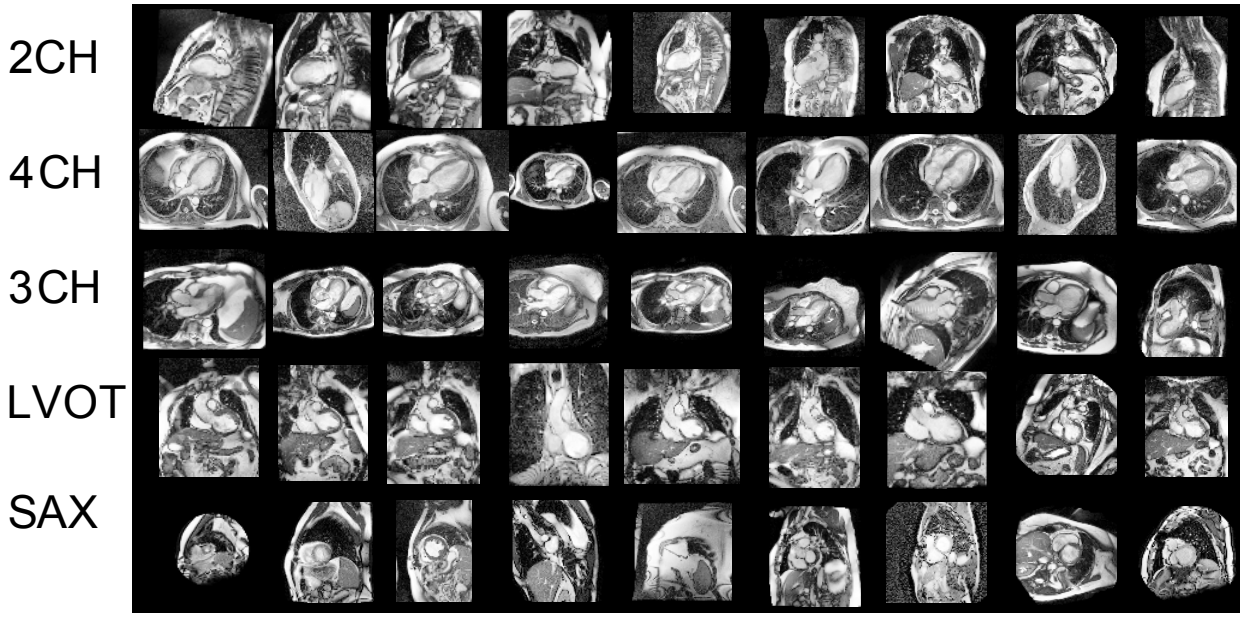



Figure 9. Typical examples of the training images from the DETERMINE dataset for different views. The acquisition and patient differences are visible. Moreover the short axis slices cover the heart from the apex to the base with quite different visual appearance.

We ran a randomized 10-fold cross validation by taking a random subset of 90% of the patients (rather then image slices) for training and used remaining 10% for validation. The patient splits guarantee that repeated acquisitions from the same patient that are occasionally present in the dataset never appear in both the training and the validation set and do not bias our results. Moreover we can place error bounds on the performance.

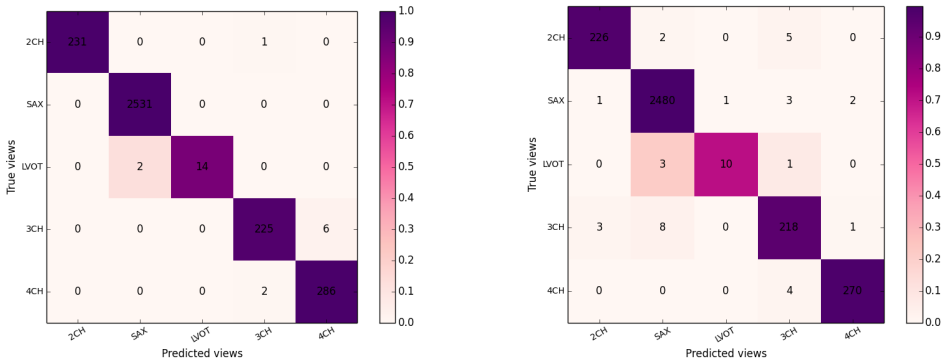
To study the robustness of the presented algorithms against the dataset bias, we trained recognizer models from different folds trained exclusively on the DETERMINE dataset (Kadish, Bello, et al. (2009)) and tested them on an independent dataset - the STACOM motion tracking challenge (Tobon-Gomez, De Craene, et al. (2013)) (KCL in short).

| | DETERMINE | KCL |
|---|---|---|
| DICOM orientation | $99.14 \pm 1.23$ | $99.08 \pm 0.46$ |
| Image forest | $59.33 \pm 4.15$ | $39.36 \pm 1.75$ |
| Image forest with jitter | $71.46 \pm 2.68$ | $48.87 \pm 2.02$ |
| CaffeNet fc7 + forest | $81.04 \pm 5.90$ | $91.44 \pm 0.86$ |
| CaffeNet fc7 + SVM | $93.54 \pm 2.85$ | $91.29 \pm 1.62$ |
| CardioViewNet fc7 + forest | $96.66 \pm 2.30$ | $92.90 \pm 2.80$ |
| CardioViewNet fc7 + SVM | $97.39 \pm 2.27$ | $88.40 \pm 88.40$ |
| CardioViewNet | $97.66 \pm 2.04$ | $91.01 \pm 3.29$ |
| CardioViewNet oversample | $97.53 \pm 2.06$ | $93.50 \pm 3.12$ |

Table 1. We computed average of individual view F1 scores in each fold (except for the underrepresented LVOT). Here we display means and standard deviations of these average F1 scores across all 10 folds. We failed to reproduce the previously published performance of the forest with the larger dataset.

## 4. Results and discussion

Here we present results for the method using DICOM based prediction and methods using image content. The mean average F1 scores are summarized in table 1 and total confusion matrices for the two best methods can be seen on fig. 10. F1 scores help us to concisely summarize both precision and recall of the models.




(a) Random forest on DICOM image normals only. (b) Finetuned CardioViewNet.

Figure 10. Sum of the confusion matrices over the 10 folds of our crossvalidation on the DETERMINE dataset for the two best models (one using DICOM normal information and the best image-based predictor using our finetuned neural network).

We confirm findings from the previous work that DICOM orientation can be used with confidence to predict cardiac views or serve as a prior (Zhou, Peng and Zhou 2012) if present. The larger DETERMINE dataset turned out to be more challenging for the forest-based method and it performed significantly worse than in the results published previously. On the contrary the performance of the CaffeNet features for cardiac view description is quite remarkable. These were not trained for cardiac MR images yet they perform better than most methods with handcrafted features. They most likely encode local texture statistics which helps with the prediction. Adding texture channels to the forest based method would probably improve the performance.

The quality of predictions using the finetuned CardioViewNet is almost on par with the approach using DICOM and significantly outperforms pure image based results of Zhou, Peng and Zhou (2012) while not requiring to train any extra landmark detectors. As features extracted from the CardioViewNet do a good job even when used with an external classifiers, they might be used to describe extra views without running another finetuning with those.

The predictions on the KCL dataset are naturally slightly worse as some differences between the studies exist. We have observed that test time oversampling (averaging predictions of the central and corner patches and their flips) improves the scores for this dataset although it does not improve the DETERMINE dataset predictions. This might indicate that a better thought oversampling strategy or image normalization and dataset augmentation might improve performance.

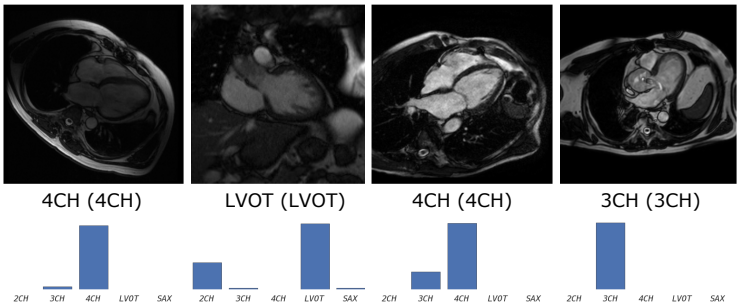



Figure 11. Examples of some of the least confident correct (the normal predictions are usually very peaky) predictions using CardioViewNet. Predicted and true (in parentheses) labels shown under the images.

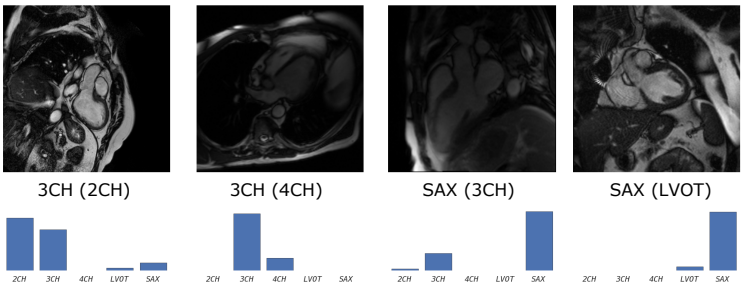



Figure 12. Examples of misclassifications using CardioViewNet. The failures usually happen with less typical acquisitions and truly ambiguous images. Extra data augmentation could probably help.

## Conclusion

Convolutional neural networks and features extracted from them seem to work remarkably well for medical images. As large datasets to train complex models are often not available, retargeting the domain of a previously trained model by finetuning to the new problem can help. Even using models trained for natural image scenes and object recognition might be a great baseline to start with. In our case doing network surgery and finetuning the pretrained model allowed us to make significant progress in cardiac view recognition from image content without handcrafting the features or training with extra annotations. Features extracted from our network should be useful as descriptors for new views and extend our method even to other pathology specific views (such as those used in congenital heart diseases) and acquisition sequences other than SSFP, and even to recognize the acquisition sequences themselves. This method makes an important contribution to the arsenal of tools for handling noisy metadata in our datasets and is already helping us to organize collections of cardiac images. We plan to use this method for semantic image retrieval and parsing of medical literature.

## Acknowledgments

We used data and infrastructure made available through the Cardiac Atlas Project (www.cardiacatlas.org - Fonseca, Backhaus, et al. (2011)). See Kadish, Bello, et al. (2009), Tobon-Gomez, De Craene, et al. (2013) for more details on the datasets. This work uses scikit-learn toolkit (Pedregosa, Varoquaux, et al. (2011)) for decision forests and Caffe deep learning framework (Jia, Shelhamer, et al. (2014)) for training of the convolutional neural network and the pretrained model.

## Funding

This work was supported by Microsoft Research through its PhD Scholarship Programme and ERC grant MedYMA 2011-291080. The research leading to these results has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 611823.

## References

Beymer D, Syeda-Mahmood T. 2008. Exploiting spatio-temporal information for view recognition in cardiac echo videos. 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops:1–8. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4563008.

Breiman L. 1999. Random forests-random features. Report no.:. Available from: http://oz.berkeley.edu/ breiman/random-forests.pdf.

Criminisi A, Shotton J, Konukoglu E. 2011. Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning. Foundations and Trends® in Computer Graphics and Vision. 7(2-3):81–227. Available from: http://www.nowpublishers.com/product.aspx?product=CGV&doi=0600000035.

Fonseca C, Backhaus M, Bluemke D, Britten R, Chung J, Cowan B, Dinov I, Finn J, Hunter P, Kadish A, et al. 2011. The Cardiac Atlas Project- an Imaging Database for Computational Modeling and Statistical Atlases of the Heart. Bioinformatics. 27(16):2288–2295.

Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. 2012. Improving neural networks by preventing co-adaptation of feature detectors:1–18. Available from: http://arxiv.org/abs/1207.0580.

Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T. 2014. Caffe: Convolutional Architecture for Fast Feature Embedding. arXiv preprint arXiv:14085093. Available from: http://arxiv.org/abs/1408.5093v1.

Kadish AH, Bello D, Finn JP, Bonow RO, Schaechter A, Subacius H, Albert C, Daubert JP, Fonseca CG, Goldberger JJ. 2009. Rationale and design for the Defibrillators to Reduce Risk by Magnetic Resonance Imaging Evaluation (DETERMINE) trial. Journal of cardiovascular electrophysiology. 20(9):982–7. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3128996&tool=pmcentrez&rendertype=abstract.

Karayev S, Trentacoste M, Han H, Agarwala A, Darrell T, Hertzmann A, Winnemoeller H. 2013. Recognizing Image Style:1–20. Available from: http://arxiv.org/abs/1311.3715.

Krizhevsky A, Sutskever I, Hinton G. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In: Advances in Neural Information Processing Systems. Available from: https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

Margeta J, Criminisi A, Lee DC, Ayache N. 2014. Recognizing cardiac magnetic resonance acquisition planes. In: Medical Image Understanding and Analysis. London, United Kingdom: Reyes-Aldasoro, Constantino Carlos and Slabaugh, Gregory. Available from: https://hal.inria.fr/hal-01009952.

Otey M, Bi J, Krishna S, Rao B. 2006. Automatic view recognition for cardiac ultrasound images. In: International Workshop on Computer Vision for Intravascular and Intracardiac Imaging. p. 187–194. Available from: http://www.engr.uconn.edu/ jinbo/doc/MICCAIworkshopCVII.pdf.

Park J, Zhou S. 2007. Automatic cardiac view classification of echocardiogram. ICCV 2007. Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4408867.

Park JH, Zhou SK, Simopoulos C, Otsuki J, Comaniciu D. 2007. Automatic Cardiac View Classification of Echocardiogram. IEEE 11th International Conference on Computer Vision:1–8. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4408867 http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4408867.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 12:2825–2830.

Razavian AS, Azizpour H, Sullivan J, Carlsson S. 2014. CNN Features off-the-shelf: an Astounding Baseline for Recognition. Available from: http://arxiv.org/abs/1403.6382.

Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, et al. 2014. ImageNet Large Scale Visual Recognition Challenge:37. Available from: http://arxiv.org/abs/1409.0575.

Taylor AM, Bogaert J. 2012. Cardiovascular MR Imaging Planes and Segmentation. In: Bogaert J, Dymarkowski S, Taylor AM, Muthurangu V, editors. Clinical cardiac mri se - 333. Springer Berlin Heidelberg; p. 93–107. Medical Radiology; Available from: http://dx.doi.org/10.1007/174_2011_333.

Tobon-Gomez C, De Craene M, McLeod K, Tautz L, Shi W, Hennemuth A, Prakosa A, Wang H, Carr-White G, Kapetanakis S, et al. 2013. Benchmarking framework for myocardial tracking and deformation algorithms: An open access database. Medical image analysis. 17(6):632–48. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23708255.

Zhou Y, Peng Z, Zhou X. 2012. Automatic view classification for cardiac MRI. In: 9th IEEE International Symposium on Biomedical Imaging (ISBI). Barcelona: IEEE; p. 1771–1774. Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6235924.