Convolutional Neural Networks for Automatic View Planning in Magnetic Resonance Imaging

Dmitry Korobchenko[1], Mikhail Sirotenko[1], Alexey Danilevich[1], Kirill Gavrilyuk[1], Praveen Gulaka[2], and Mikhail Rychagov[1]

[1] Samsung R&D Institute Russia, Moscow, Russia
[2] Samsung Electronics, Suwon, Korea

Abstract. View planning (scan planning) is the process of prescribing diagnostic slices in scout MRI volume. It is an important part of MRI investigation workflow. In this paper we present a novel framework for Automatic View Planning (AVP) consisting of a universal, deep multi-layer discriminative architecture for anatomical landmarks detection and several pre-processing and post-processing algorithms. The discriminative architecture is based on Convolutional Neural Network trained in two stages: unsupervised pre-training using convolutional Predictive Sparse Decomposition and supervised fine-tuning using classical techniques for neural networks training. To refine the positions of landmarks we introduce our own algorithm based on trained statistical model. The framework is applicable for a wide variety of MRI procedures such as brain, cardiac, knee and spine MRI. Verification on real clinical data showed acceptable competitive quality and speed for clinical applications.

1 Introduction

In an MRI study, view planning is one of the most important steps before diagnostics. The positions and orientations of planned slices depend on human body part under investigation. For example, typical cardiac view planning consists of obtaining 2-, 3-, 4-chamber and short axis views (Fig. 1). Typically view planning is performed manually by a doctor and suffers from several drawbacks: it is very time consuming, it is operator dependent, it requires qualified medical personnel. To overcome all these disadvantages, an automatic view planning system may be used to estimate desired view planes by analysis of 3D scout MR image. Such scout image is characterized by fast acquisition time and therefore low resolution and poor quality.

We propose a fully automatic view planning framework which is designed to be able to process various kinds of anatomies (brain, heart, spine, knee). Our approach is based on landmarks detection and supported by few anatomy specific pre-processing and post-processing algorithms. The key features of our framework are: deep learning methods for robust detection of landmarks in rapidly acquired low resolution scout images; unsupervised learning for overcoming the problem of small training dataset; position refinement of detected landmarks based on a statistical model.

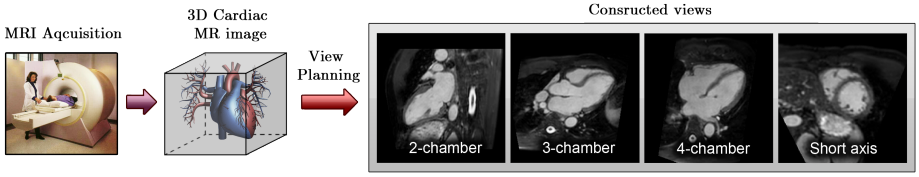


Fig. 1: Cardiac View Planning

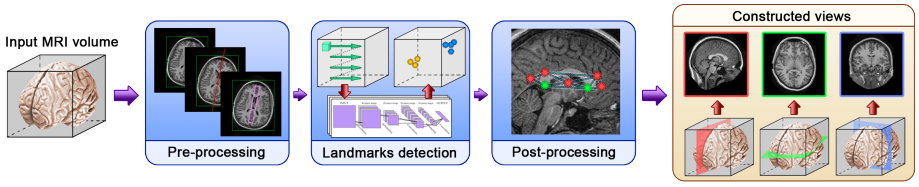


Fig. 2: Automatic View Planning workflow

1.1 Related work

Resent literature demonstrates application of various approaches for automatic view planning. Most authors consider only some specific MRI anatomy and create highly specialized algorithms aimed to perform view planning for this anatomy. For various anatomies (brain, cardiac, knee, spine) different authors use marginal space learning, segmentation, specific features for anatomical landmarks detection, deformable models, various anchor points detectors [1,2,5,6,8–11].

2 Automatic View Planning framework

The proposed AVP framework workflow consists of the following steps (Fig. 2): 3D scout MRI volume acquisition, pre-processing (including bounding box estimation, statistical atlas anchoring and anatomy specific operations), landmarks detection, post-processing (including landmarks' positions refinement and anatomy specific operations), estimation of view planes positions and orientations by landmarks positions.

2.1 Pre-processing

A commonly used operation for pre-processing stage is an estimation of bounding box, which bounds only essential part of body part under investigation. Such bounding box is helpful for preparation of working zones, local origins planning, etc. We estimate bounding box via integral projections.

On the next step, statistical atlas is applied to reduce search space. This atlas contains information about statistical distribution of landmarks' positions inside

certain part of body within local coordinate system connected to the bounding box. It is constructed on the base of a set of annotated volumetric medical images. The statistical atlas anchored to the volume defines search space boundaries.

For brain AVP we additionally perform an anatomy specific pre-processing, which consists in mid-sagittal plane (MSP) estimation. We apply our own approach based on brain longitudinal fissure detection. The estimated MSP is used as one of planned MRI views for brain AVP. Further, the MSP helps us to reduce search space for brain-specific landmarks detection.

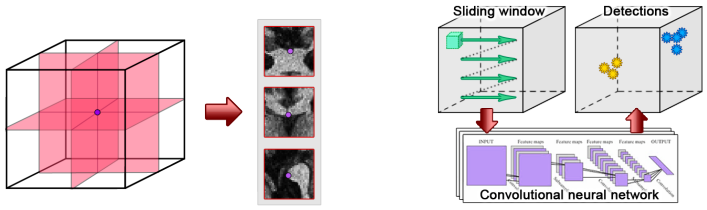
2.2 Anatomical landmarks detection

Landmarks detection is a process of estimation of anatomical landmarks' positions in medical image. For each MRI type (anatomy) a set of specific landmark types was chosen to be used for the computation of desired view planes. Unique landmarks of different types are used for brain, cardiac and knee anatomies. For spine MRI several landmarks of a certain type could be presented in medical image (anchored to different intervertebral disks).

Landmarks detection is equivalent to points classification or labels associating, such as: landmark L_1 or ... landmark L_N or *Background*. Background relates to the point type where no landmarks are located. Points to be classified (*search points*) are actually only a subset of all points in the volume: they are picked-up from search area defined by statistical atlas with some prescribed step (i.e. distance between neighboring points).

For classification of point its surrounding context is used (a portion of voxel data extracted from neighborhood of the search point). In our approach, we pick up a cubic subvolume around search point and extract three orthogonal slices of this subvolume passing through the search point (Fig. 3a). Landmarks detector scans the volume with a 3D sliding window and performs classification of each point by its surrounding context (Fig. 3b). Classification is done by means of multi-layer convolutional neural network (CNN) [3]. In inference mode trained network produces a vector corresponding to a probability distribution of appearance of certain landmark or *Background* in a respective point. Instead of absolute probability values we use a comparative measure which takes in account all classes probabilities relative to each other. Such measure is named *Landmark Candidate Score* (LCS) and is calculated for each point after CNN processing (Algorithm 1).

We filter all detected candidates by thresholds for LCS. Such thresholds are calculated in advance using a set of annotated MRI volumes. The calculation is performed by finding the thresholds which minimize loss function consisting of *false negative errors* (FN) and *false positive errors* (FP). Here the total of FNs is calculated as a number of ground-truths which have no detections within some sphere around it. The total of FPs is calculated as a number of all detections that are out of these spheres (too far from corresponding ground-truths).



(a) (b)

Fig. 3: Landmarks detection: (a) surrounding context of search point; (b) landmarks detector architecture

Input: CNN output vector $A = \{a_1, \ldots, a_N\}$
Output: LCS vector $Q = \{q_1, \ldots, q_N\}$
$c_{max} \leftarrow \arg\max A$
for $c \leftarrow 1$ ***to*** N **do**
  if $c = c_{max}$ **then** $q_c \leftarrow a_c - \max A \setminus \{a_c\}$
  else $q_c \leftarrow a_c - \max A$
end

Algorithm 1: Landmark Candidate Score Calculation

2.3 Post-processing

Statistical model for landmarks' positions refinement In some cases the output of landmarks detector could be ambiguous: there could be several competing landmarks candidates with high LCS measure. In order to eliminate this ambiguity, special algorithm is used based on statistics of landmarks' relative positions. The goal of the algorithm is to find configuration of landmarks with high LCS value minimizing the energy:

$$E(X, M_X, \Sigma_X) = \sum_{x_s \in X, x_t \in X} \Psi_{st}(x_s, x_t, M_X, \Sigma_X) \tag{1}$$

where $X \in \mathbb{R}^{3\times K}$ is a set of K vectors of coordinates of landmarks, called landmarks configuration; M_X - mean distances of each landmark from each other; Σ_X - landmarks coordinates distances covariance tensor; E - energy of the statistical model, lower values correspond to more consistent configurations; Ψ_{st} - spatial energy function measuring statistical consistency of two landmarks. In our implementation we define spatial energy as follows:

$$\Psi_{st}(x_s, x_t, M_X, \Sigma_X) = \frac{1}{2}(x_s - x_t - \mu_{st})^T \Sigma_{st}^{-1}(x_s - x_t - \mu_{st}) \tag{2}$$

where x_s and x_t correspond to 3-dimensional coordinate vectors from configuration X; μ_{st} - 3-dimensional mean distance-vector between landmarks s and t; Σ_{st}^{-1} - inverse covariance matrix of 3-dimensional vectors of distances between landmarks s and t.

Statistics μ_{st} and Σ_{st}^{-1} are computed once on an annotated dataset. Instead of computationally expensive direct optimization we use heuristic approach based on the assumption, that landmarks detections are mostly correct. On the first step of this algorithm from the plurality of landmark candidate points, a subset S_a is selected, consisting of M subsets $S_{a_1} \dots S_{a_M}$ of N candidate points having the greatest LCS for each of the M landmarks. Also a set S_b is selected from S_a consisting of M best candidates - one for each landmark. Next, the loop is defined for all candidates x_i from S_b. Then partial derivative $\frac{\partial E(X,M_X,\Sigma_X)}{\partial x_i}$ of the energy with respect to x_i is calculated. If the magnitude of this partial derivative is the greatest among the elements of S_b, then the nested loop is initialized for all x_j in S_{a_i}. On the next step, a new configuration $S_b{}'$ is defined by substitution of x_j in S_b instead of x_i. This new configuration is then used to compute the energy of the statistical model; if the energy is lower than the energy for S_b, then the $S_b{}'$ is assigned to S_b. This process is repeated until value of partial derivative magnitude is higher than predefined tolerance.

Spine AVP post-processing For spine AVP we additionally perform an anatomy specific post-processing which consists of detected landmarks clustering, spinal curve fitting and further refinement of intervertebral disks position within spinal curve.

2.4 Training landmarks detector

In our approach, we utilize CNN [3], which showed very promising results in various recognition tasks. Our network has several convolutional layers, pooling layers, rectification layers and fully-connected layers (similar to [3]). Output of CNN is a vector with number of elements equal to number of landmarks plus one (for *Background*). These output values correspond to pseudo-probabilities that current landmark is located in a current search point (or no landmarks are located here in case of Background).

We train our network in two stages. On the first stage we perform an unsupervised pre-training using convolutional Predictive Sparse Decomposition (PSD) [4]. Then we perform a supervised training (refining the pre-trained weights and learning other weights) using stochastic gradient descent with energy-based learning [7].

Pre-training At the first stage of learning, an unsupervised pre-training procedure initializes the weights of all convolutional (feature-extraction) layers of the CNN. Learning in unsupervised mode uses unlabeled (non-annotated) data: randomly picked samples from 3D medical images. This learning is performed via Sparse Coding and convolutional PSD techniques [4]. The learning process is done separately for each convolutional layer in stochastic style. Firstly, each input sample is encoded using *dictionary D* via convolutional sparse coding, which results in sparse feature maps (sparse codes). Secondly, weights of convolutional layer are adjusted by gradients in order to reduce divergence between features produced by the layer and features produced by sparse coding. For all layers except the first, the training set is formed as a set of the previous layer's outputs.

The dictionary D is obtained from training set in unsupervised mode (without using annotation labels). An advantage of unsupervised approach for the optimal dictionary finding is the fact that the dictionary is learned directly from data. So, found dictionary D optimally represents a hidden structure and specific nature of used data. An additional advantage of the approach is that it does not need a large amount of annotated input data for the dictionary training. After the unsupervised training is done, the entire CNN is ready to produce multilevel sparse codes which are good hierarchical feature representation of input data.

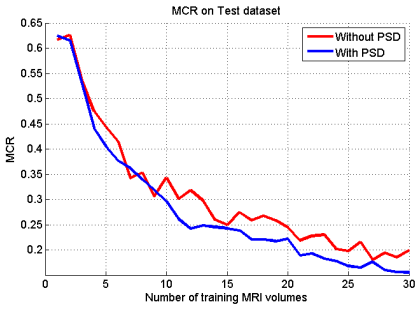
Fine-tuning On the next step a supervised training is performed using annotated data. For preparing the training data we use a set of annotated 3D medical images of certain anatomy. Training samples are randomly picked from the annotated volumes. Class label for a sample is a vector corresponding to probability distribution of landmark appearance in a respective point. Such target vector is calculated using energy-based approach. For each landmark class, the target value is calculated on the basis of a distance from current sample to the closest ground truth point of this class. We also added some extra samples with spatial distortions to train the system to be robust and invariant to noise. The notwork is tuned to produce outputs similar to target vectors via stochastic gradient descent. At the beginning of the procedure, some weights of feature extraction layers are initialized with the values computed at pre-training stage.

3 Results

For the algorithm testing and the landmarks detector training, a database of low resolution rapidly acquired 3D MRI images of various anatomies was collected and annotated (including 94 brain, 80 cardiac, 31 knee and 99 spine MRI volumes). In our implementation, we used a CNN with input size 32x32x3. Input volumes (at both learning and processing stages) were resized to spacing 2.0. The architecture of network is following: the 1st convolutional layer has 16 kernels 8x8 with shrinkage activation function; then goes abs-rectification, local contrast normalization and max-pooling with down-sampling factor 2; then we have the 2nd convolutional layer with 32 kernels 8x8 connected to 16 input feature maps with sparse connection matrix; lastly, a fully-connected layer with sigmoid activation function finalizes the net.

Used unsupervised pre-training is useful when we have only a few of labeled data. We demonstrate superiority of using PSD with few labeled MRI volumes. For such experiment, unsupervised pre-training of our network was performed in advance. Then, several annotated volumes were picked up for supervised fine-tuning. After training, misclassification rate (MCR) on test dataset was calculated. Plot on Fig. 4a shows performance of classification depending on various number of annotated volumes taking part in supervised learning.

Verification of the AVP framework was performed by comparing (by angle and generalized distance) automatically built views with views built on the base of ground-truth landmarks positions (Fig 4b). Examples of the constructed views are shown in Fig. 5. Computation time for our algorithm varies from 1 to 5



(a)

Anatomy	Dist (mm)	Ang (deg)
Brain	1.2±0.9	1.6±1.2
Cardiac	6.3±7.4	9.6±7.7
Knee	2.1±1.6	1.7±1.3
Spine	3.8±1.9	4.1±2.8

(b)

Fig. 4: Landmarks detection: (a) MCR plot calculated on test dataset using CNN learned with various number of annotated MRI volumes, red line - pure supervised mode, blue line - supervised mode with PSD initialization; (b) quality verification of AVP framework in terms of distance and angular error statistics on test datasets

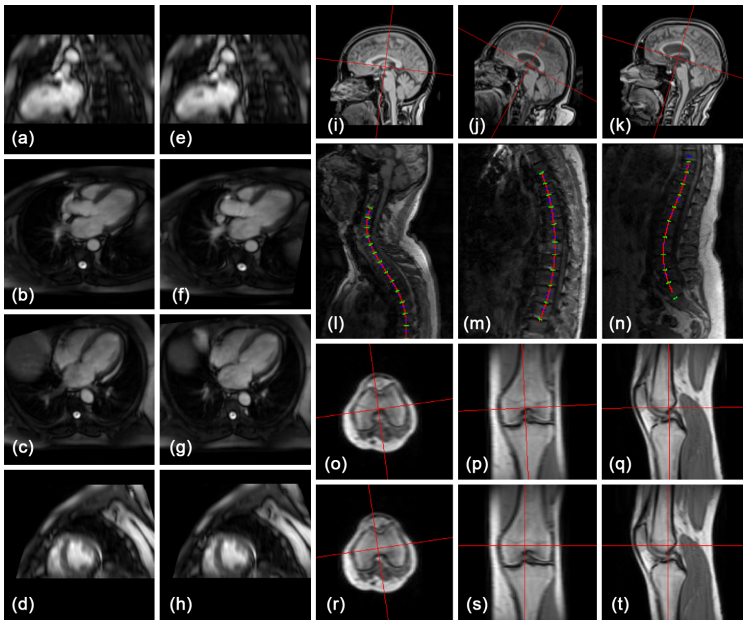


Fig. 5: AVP results: (a)-(d) ground-truth cardiac views; (e)-(h) automatically built cardiac views; (i)-(k) automatically built mid-sagittal views, red lines represents intersection with axial and coronal views; (l)-(n) positions and orientations of automatically detected disks(green) and spinal curve (red) for different spinal zones (cervical, thoracic and lumbar correspondingly); (o)-(q) ground-truth knee views; (r)-(t) automatically built knee views

seconds depending on anatomy. The results show that the developed technique provides acceptable and competitive quality for diagnosis, while the time spent for the operations (image acquisition and processing) is pretty short.

4 Conclusion

We have presented a novel automatic view planning framework based on robust landmarks detector, able to perform high accuracy view planning for the different anatomies using low quality rapidly acquired 3D scout images. The quality of proposed algorithmic solutions was verified using collected MRI datasets. Benchmarking of developed system demonstrate acceptable quality and high speed.

References

1. Bauer, S., Ritacco, L.E., Boesch, C., Nolte, L.P., Reyes, M.: Automatic scan planning for magnetic resonance imaging of the knee joint. Annals of biomedical engineering 40(9), 2033–2042 (2012)
2. Iskurt, A., Becerikli, Y., Mahmutyazicioglu, K.: Automatic identification of landmarks for standard slice positioning in brain mri. Journal of Magnetic Resonance Imaging 34(3), 499–510 (2011)
3. Jarrett, K., Kavukcuoglu, K., Ranzato, M., LeCun, Y.: What is the best multi-stage architecture for object recognition? In: Computer Vision, 2009 IEEE 12th International Conference on. pp. 2146–2153. IEEE (2009)
4. Kavukcuoglu, K., Sermanet, P., Boureau, Y.L., Gregor, K., Mathieu, M., Cun, Y.L.: Learning convolutional feature hierarchies for visual recognition. In: Advances in neural information processing systems. pp. 1090–1098 (2010)
5. Kelm, B.M., Zhou, S.K., Suehling, M., Zheng, Y., Wels, M., Comaniciu, D.: Detection of 3d spinal geometry using iterated marginal space learning. In: Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging, pp. 96–105. Springer (2011)
6. Law, M.W., Tay, K., Leung, A., Garvin, G.J., Li, S.: Intervertebral disc segmentation in mr images using anisotropic oriented flux. Medical image analysis 17(1), 43–61 (2013)
7. LeCun, Y., Chopra, S., Hadsell, R., Ranzato, M., Huang, F.: A tutorial on energy-based learning. Predicting structured data 1, 0 (2006)
8. Lu, X., Georgescu, B., Jolly, M.P., Guehring, J., Young, A., Cowan, B., Littmann, A., Comaniciu, D.: Cardiac anchoring in mri through context modeling. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010, pp. 383–390. Springer (2010)
9. Zhan, Y., Dewan, M., Harder, M., Krishnan, A., Zhou, X.S.: Robust automatic knee mr slice positioning through redundant and hierarchical anatomy detection. Medical Imaging, IEEE Transactions on 30(12), 2087–2100 (2011)
10. Zhang, L., Xu, Q., Chen, C., Novak, C.L.: Automated alignment of mri brain scan by anatomic landmarks. In: SPIE Medical Imaging. pp. 72592M–72592M. International Society for Optics and Photonics (2009)
11. Zheng, Y., Lu, X., Georgescu, B., Littmann, A., Mueller, E., Comaniciu, D.: Automatic left ventricle detection in mri images using marginal space learning and component-based voting. In: SPIE Medical Imaging. pp. 725906–725906. International Society for Optics and Photonics (2009)