LOCAL ENTROPY-OPTIMIZED TEXTURE MODELS FOR SEMI-AUTOMATIC SPINE LABELING IN VARIOUS MRI PROTOCOLS

M. Wimmer, D. Major, A. A. Novikov, K. Bühler

VRVis Center for Virtual Reality and Visualization, Vienna, Austria

ABSTRACT

We present a novel pipeline for acquisition protocol independent spine labeling in volumetric Magnetic Resonance Imaging (MRI) data of the lumbar spine. Our learning-based system uses local Entropy-optimized Texture Models (ETMs) for reducing the intensity scale in clinical data to only a few gray levels. The task of intervertebral disc localization is then performed on the normalized data. The benefit of our method is, that we can deal with various MRI protocols, such as T1-weighted (T1w) and T2-weighted (T2w) scans. Using the entropy objective allows us furthermore to apply the algorithm to acquisition protocols which are not covered by the training set. We achieve high disc localization accuracies for both, MRI protocols which are covered and not covered by training. The approach can be easily extended to other modalities.

Index Terms— Spine labeling, MRI, Entropy-optimized Texture Models

1. INTRODUCTION

Labeling of the spinal column in MR sequences is an important task in clinical practice, as it serves the diagnosis and operation planing of spine-related pathologies. However, manual labeling is a time-consuming task for clinicians, hence automatic or semi-automatic approaches are in demand.

There is a wide range of different MR acquisition protocols nowadays available which have high variations in terms of appearance and have no standardized scale like Hounsfield for CT. In the literature, various recent works targeted spine labeling in *specific* MR sequences, e.g. on T1w [1], T2w [2], [3], Dixon protocol [4] or SPIR [5] data. However, all these methods are specific to a certain protocol. Thus, approaches which are able to localize the spinal parts without retraining for the different imaging parameters are of high interest. Lootus et al. [6] applied Deformable Part Models using the HOG descriptor combined with a graphical model in order to localize vertebrae in a set of T2w MRI scans. The authors claim that their method works without retraining also on CT data, although no extensive evaluation is provided. In a more recent work [7], they introduce a normalization based on the vertebral intensities, which makes their algorithm more robust to parameter changes and different vendors in T2w MRI scans. Zukić et al. [8] introduced an automated detection and segmentation framework for vertebrae using a boosted cascade of simple features for detection and the watershed method for segmentation. Their method is able to handle T1w, TIRM and T2w MRI data at once, although the extension to other sequence types requires a retraining of their detectors. Cai et al. [9] address the problem of labeling and segmenting the spine in a modality independent way. Restricted Boltzmann machines are used for landmark detection, followed by a global spine model matching algorithm. Finally, vertebrae are registered and segmented via local models. Their framework was evaluated on MR and CT data and is applicable without retraining to different modalities. However, it is much more complex than our proposed method.

Contribution. We present a novel *semi-automatic pipeline* for labeling of the spinal column, which can process 3D MR scans with a high intensity variability, like T1w and T2w scans, acquired on different scanners with varying scan parameters. We propose a *learning-based system*, where we train *ETMs* [10] for reducing the intensity scale in the clinical data. We build *general models*, which normalize data across different MR protocols. Thus, no separate models for different acquisition setups are required. Moreover, our method is applicable to sequences and protocols which are not covered by the training set. When labeling an unseen scan, the learned models are applied and intervertebral disc (IVD) centers are localized with an *adaptive approach* in the normalized data.

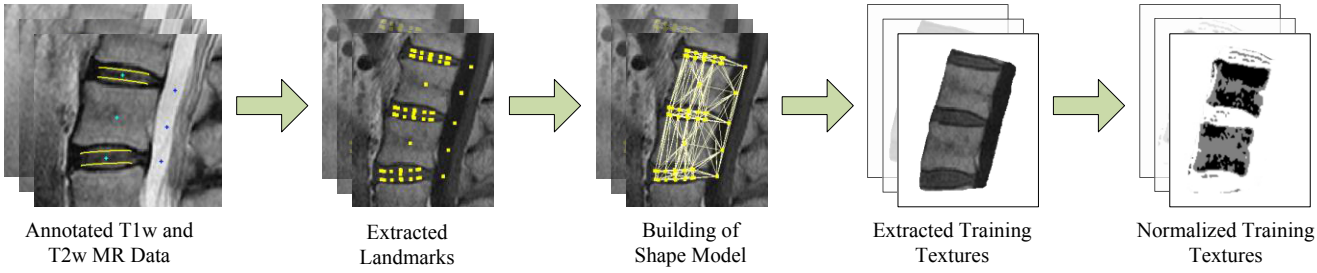
2. METHODS AND MATERIALS

The following section introduces ETMs in general, our proposed application within the scope of spine labeling and the suggested refinement scheme.

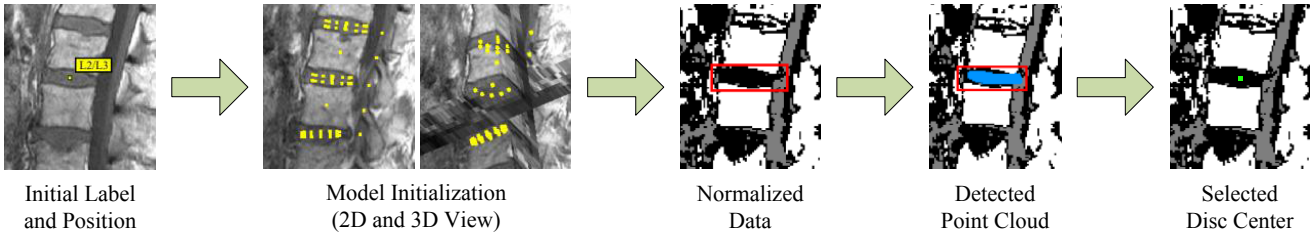
2.1. ETMs in General

Zambal et al. [10] introduced 2D ETMs for segmentation from dense landmarks at object boundaries. They propose

The competence center VRVis with the grant number 843272 is funded by BMVIT, BMWFW, and the Vienna Business Agency within the scope of COMET - Competence Centers for Excellent Technologies. The program COMET is managed by FFG. Thanks go to our project partner AGFA Healthcare for providing data and valuable input.



(a) Overview of the training of ETMs for data normalization: From an annotated set of MR data, corresponding landmarks are extracted and a shape model is built. Training textures are extracted and the texture transformations are then optimized iteratively [10].



(b) Basic steps when labeling an MR scan: Based on the initial position and label provided by a user, the corresponding model is placed in the scan and matched. On the normalized scan, the final center $\boldsymbol{d}_i^*$ is obtained with an adaptive refinement method.

Fig. 1: System overview describing the training of ETMs and the labeling of an unseen MR scan.

a novel texture model, that extends AAMs with a texture normalization approach. This enables the usage of the model in a modality independent way. It is built using m training textures T_k resampled by n texels and quantized to r_k source gray levels. The task is to find optimal mappings f_k for every training texture T_k which maps the r_k input values to a gray scale with only a few (s) target values. The mappings are found with iterative optimization of two entropy-based objectives: each texel's source gray value should be mapped to a certain target value. Thus, the uncertainty of all mappings is minimized, which is described by the *model entropy* H^{model}. On the other hand, a compensation term prevents degeneration of mappings f_k and hence drives the mapped textures towards maximum information content. This is described by the *image entropy* H^{tex}.

The model is matched to an *unseen image* by iteratively changing its shape parameters and texture mapping. Bayesian reasoning is used to assess the matching quality. It is based on the likelihood of the normalization result of the unseen texture given a particular shape and on the shape prior. The obtained mapping is then applied, which results in the desired intensity-reduced scan with only s target gray levels.

In summary, H^{model} ensures, that the same tissues are normalized to the same target level. More than that, the image entropy term H^{tex} guarantees that the contrast between different tissues is preserved. These are important and powerful characteristics, which enable us to apply such a model without retraining to images, that express similar intra-tissue homogeneity and inter-tissue contrast to those captured by the model.

2.2. ETMs for Spine Labeling

We extend the ETMs to 3D and propose to build *local models* from sparse landmarks from a mixed set of T1w and T2w MR volumes (see Fig. 1a). We build *three-disc-models* $\mathcal{M}_i$ around a *middle disc* $\boldsymbol{d}_i$ with anatomical label $\lambda_i = \{T12/L1, \ldots, L4/L5\}$, including also its adjacent *upper disc* $\boldsymbol{d}_{i-1}$ and *lower disc* $\boldsymbol{d}_{i+1}$. This has been already successfully applied in the scope of spine labeling in the literature [11]. Around each IVD center, we include sampled positions along the surface of a cylinder, which approximates the size of the IVD. Furthermore, we add the two vertebral body centers between $\boldsymbol{d}_{i-1}$ and $\boldsymbol{d}_{i+1}$ and spinal canal landmarks, which correspond to the IVD and vertebra centers.

2.3. Labeling of an Unseen Volume Dataset

Our semi-automatic method requires minimal input from a user when labeling an unseen scan: an initial click position $\boldsymbol{p}$ in the volume inside an IVD or vertebra and its corresponding anatomical label λ_i (see Fig. 1b).

ETM matching. An instance of the learned model $\mathcal{M}_i$, which corresponds to the user-assigned anatomical label λ_i is placed at position $\boldsymbol{p}$ and matched accordingly. As a result, the normalized scan is retrieved, as well as *candidate positions* for the landmarks, e.g. the middle disc $\boldsymbol{d}'_i$, upper disc $\boldsymbol{d}'_{i-1}$, lower disc $\boldsymbol{d}'_{i+1}$ and vertebrae centers. We then apply a refinement step for the candidate disc center position $\boldsymbol{d}'_i$.

Adaptive Disc Center Position Refinement. We span a bounding box $\mathcal{R}$ around the model-matched disc position $\boldsymbol{d}'_i$,

which defines our region of interest for the refinement (see Fig. 2). The size of $\mathcal{R}$ is based on the ground truth, from which we calculate the average dimension of discs in sagittal, axial and coronal direction. For every voxel inside $\mathcal{R}$ we decide if it belongs to the disc or not with an adaptive method inspired by *Haar-like features* [12]. We construct a filter with three regions: upper region $\mathcal{R}_U$, middle region $\mathcal{R}_M$ and lower region $\mathcal{R}_L$. The regions are then placed in the following way: $\mathcal{R}_M$ is placed at the current position $\boldsymbol{p}'$ in $\mathcal{R}$. $\mathcal{R}_U$ and $\mathcal{R}_L$ are displaced based on the IVD orientation vector $\boldsymbol{n}$ and the average disc thickness $\bar{t}$ estimated from the ground truth. The vector $\boldsymbol{n}$ is calculated based on the model-matched landmark positions. For every region $\mathcal{R}_U$, $\mathcal{R}_M$ and $\mathcal{R}_L$, the *intensity mode* is calculated: $\hat{m}_L$, $\hat{m}_M$ and $\hat{m}_U$. We consider the voxel as disc candidate if:

$$M(x,y,z) = \begin{cases} 1 & \text{if } \hat{m}_U \neq \hat{m}_M \wedge \hat{m}_L \neq \hat{m}_M \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

From the obtained binary IVD mask, we calculate the *centroid* as the refined center position $\boldsymbol{d}_i^*$.

Propagation. The labeling is performed in an iterative manner. From the model matched around the initial position $\boldsymbol{p}$ we also obtain candidate positions for the upper and lower IVD, i.e. $\boldsymbol{d}'_{i-1}$ and $\boldsymbol{d}'_{i+1}$. We continue the search first towards $L4/L5$ and then upwards to $T12/L1$ and repeat model matching, adaptive refinement and retrieval of the next IVD positions $\boldsymbol{d}'_{i-1}$ resp. $\boldsymbol{d}'_{i+1}$.

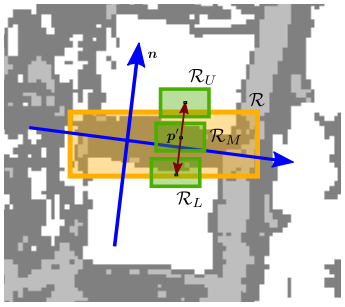


Fig. 2: Illustration of our proposed refinement method. Filter regions $\mathcal{R}_U$, $\mathcal{R}_M$ and $\mathcal{R}_L$ are shown in green and the search region $\mathcal{R}$ in orange. Displacement vectors are shown in red and the IVD orientation vectors in blue.

3. EVALUATION AND RESULTS

The following section provides an elaborate evaluation of our novel pipeline on lumbar MR volumes.

3.1. Data and Experimental Setup

Four datasets $\mathcal{S}_i$ of sagittal acquired lumbar MR scans from five different scanners are used for evaluation. The voxel sizes are highly anisotropic, with an in-plane resolution ranging from 0.59 mm^2 to 1.25 mm^2 and slice thickness from 2 to 6 mm. We reconstruct 62 volumes from the scans and consider seven IVDs from $T11/T12$ to $L5/S1$, which results in 434 IVDs in total.

Dataset $\mathcal{S}_1$: Our dataset with 13 T1w MR scans.

Dataset $\mathcal{S}_2$: Our dataset with 10 T2w MR scans.

Dataset $\mathcal{S}_3$: Challenge Dataset [3], consisting of T2w MR scans from 15 different subjects.

Dataset $\mathcal{S}_4$: Dixon protocol data [4]. Scans from eight subjects, where we use three different image channels: opposed-phase, fat and water saturated. We treat every channel separately, hence we obtain 24 volumes for testing.

In our datasets ($\mathcal{S}_1$, $\mathcal{S}_2$), magnetic field inhomogeneities are present and half of the scans exhibit at least one of the following pathologies: fractures, disc herniation, scoliosis and lumbar hyperlordosis.

We evaluated our method on three setups (see Table 1) on an Intel Xeon E5 PC. We split $\mathcal{S}_1$ and $\mathcal{S}_2$ into two subsets and perform two-fold-cross validation. Setup #2 and #3 demonstrate the generality of our method by evaluation on unseen MR sequences, which were *not included* in the training set.

	Training		Testing		
#1	$\mathcal{S}_{1,1}$	$\mathcal{S}_{2,1}$	$\mathcal{S}_{1,2}$	$\mathcal{S}_{2,2}$	
	$\mathcal{S}_{1,2}$	$\mathcal{S}_{2,2}$	$\mathcal{S}_{1,1}$	$\mathcal{S}_{2,1}$	
#2	$\mathcal{S}_1$	$\mathcal{S}_2$	$\mathcal{S}_3$	$\mathcal{S}_4$	
#3	$\mathcal{S}_3$		$\mathcal{S}_1$	$\mathcal{S}_2$	$\mathcal{S}_3$

Table 1: Evaluation setups

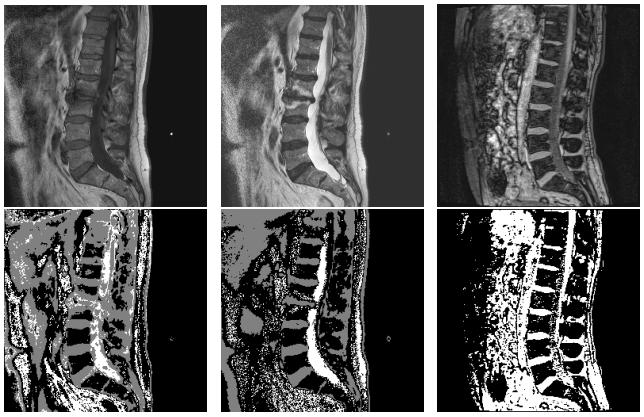
3.2. ETM Model Training

We trained models $\mathcal{M}_i$ for the middle IVDs from $T12/L1$ to $L4/L5$, which introduces an overlap of models and increases robustness. When choosing parameters for ETMs, the goal is to find rather homogeneous intensity mappings for IVDs, vertebrae and the spinal cord, without losing anatomical information. Best mapping results were obtained with $r = 110$ source and $s = 3$ target levels using leave-one-out cross-validation on our training data. This provides a good initial quantization, where we do not miss relevant intensity changes but also remove image noise. The time for model training depends on the number of training textures. We report 4.6 ± 1.5 min for setup #1 and #3 and 13 ± 2.3 min for setup #2 on average.

It is important to note, that we do not obtain the same mapping for one tissue in all sequences, e.g. for the mapping of disc intensities. The tissues are in a different range, but still homogeneous mappings are obtained (see Fig. 3).

3.3. Labeling Results

We start a full semi-automatic data normalization and labeling run inside every IVD. Seeds are placed within 10 mm to the ground truth center, which allows the models to converge towards the disc centers while matching. We provide the following measures:



(a) T1w (b) T2w (c) Dixon opp-phase

Fig. 3: Various mid-sagittal slices of MR scans from our data collection $\mathcal{S}$ (top row) and corresponding sample normalization results (bottom row).

IVD localization performance: p_ϵ explains, how many detected positions $\boldsymbol{d}_i^*$ lie within ϵ mm to the annotated IVD center $\boldsymbol{d}_i$, with $\epsilon \in \{2, 4, 6, 10\}$.

IVD center accuracy: The Euclidean distance from the localized position $\boldsymbol{d}_i^*$ to the ground truth $\boldsymbol{d}_i$ describes the mean position error $\bar{e}$ and standard deviation sd.

Highest IVD center position accuracy was achieved with setup #2 (see Fig. 4a), where we reach a mean error of 3.82 $\pm$ 2.47 mm with $p_{10} = 97.64\%$ (see Table 2). A high detection rate is also achieved on setup #3 with $p_{10} = 93.91\%$ and an avgerage distance of 4.45 $\pm$ 3.44 mm from the ground truth. Setup #2 and #3 were trained on only a *subset* of all available imaging protocols in $\mathcal{S}$. Thus we show, that our generalized approach is applicable to acquisition protocols, which were not included in the training, but exhibit similar contrast characteristics as captured by our learned ETMs (see Section 2.1). This is reflected in the results on $\mathcal{S}_4$. In general, we observe higher p_ϵ for T2w scans than for T1w scans, because of a better tissue contrast. Localization performance is lower for setup #1 due to pathologies and imaging artifacts. On disc-level, we obtain the best result for $L2/L3$ with 2.63 $\pm$ 1.52 mm (setup #2), because no severe abnormalities are present in this region. Highest error rates were observed for setup #1. Due to pathologies like herniation and lumbar hyperlordosis, we reach a low accuracy for $L5/S1$ (9.53 $\pm$ 3.67 mm) as well as for $T11/T12$ (5.64 $\pm$ 3.24 mm), where magnetic field inhomogeneities are present (see Fig. 4b).

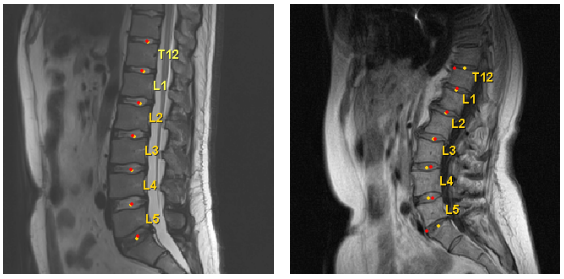
Overall processing time for an unseen MR scan is 12.6 $\pm$ 3.7 sec, which results in 1.8 $\pm$ 0.5 sec per IVD.

4. DISCUSSION AND CONCLUSION

We presented a novel, learning-based pipeline for semi-automatic labeling of lumbar MR volumes. Our main contribution lies in the generality of our method: We can process various imaging protocols and apply our approach also to unseen protocols, which were not covered by our training set. Furthermore, our method is significantly faster to train than recent deep learning approaches [9]. We successfully localized 84.99 % IVDs within 6 mm and 97.64 % within 10 mm to the ground truth center, which is competitive to localization measures of state-of-the-art methods. Zukić et al. [8] report a false negative rate of 7.1 % for automatic vertebrae detection. This method is closest to ours in terms of the variability of MR protocols (T1w, T2w, TIRM). Chen et al. [4] reach a position error of 1.3 $\pm$ 0.6 mm for all IVDs on Dixon protocol data, whereby all image channels are used for training and testing. We report 3.68 $\pm$ 1.94 mm (setup #2) and 3.44 $\pm$ 2.99 mm (setup #3), whereby we did not include Dixon data in the training. On an extended set of the Challenge Dataset $\mathcal{S}_3$, they achieve position errors between 1.8 $\pm$ 1.1 mm and 2.8 $\pm$ 6.5 mm for different cross-validation setups. We provided IVD center positions with a mean distance of 4.04 $\pm$ 3.14 mm to the expert-annotated ground truth position ($p_{10} = 94.97\%$). Overall, a higher deviation is present for $L5/S1$ and $T11/T12$, which we believe to decrease with a more enhanced refinement method. We plan to extend our framework into a fully automatic system and evaluate it also on cervical and thoracic scans.

	p_2 [%]	p_4 [%]	p_6 [%]	p_{10} [%]	$\bar{e} \pm sd$ [mm]
#1	**14.29**	**40.00**	**60.71**	**84.39**	**5.78 $\pm$ 3.76**
$\mathcal{S}_1$	13.88	37.96	57.14	77.76	6.27 $\pm$ 4.18
$\mathcal{S}_2$	14.69	42.04	64.29	91.02	5.29 $\pm$ 3.22
#2	**22.51**	**64.53**	**84.99**	**97.64**	**3.82 $\pm$ 2.47**
$\mathcal{S}_3$	27.76	64.49	80.14	94.97	4.04 $\pm$ 3.14
$\mathcal{S}_4$	19.21	64.56	88.04	99.32	3.68 $\pm$ 1.94
#3	**21.32**	**59.00**	**78.46**	**92.06**	**4.45 $\pm$ 3.44**
$\mathcal{S}_1$	6.31	40.45	60.11	80.52	6.24 $\pm$ 4.12
$\mathcal{S}_2$	11.84	43.06	71.63	94.49	4.93 $\pm$ 2.67
$\mathcal{S}_4$	32.14	74.15	89.71	96.34	3.44 $\pm$ 2.99

Table 2: Overall performance measures per evaluated setup (bold) and corresponding dataset-specific results.



(a) Successful case ($\mathcal{S}_3$) (b) Pathological scan ($\mathcal{S}_1$)

Fig. 4: Labeling results shown on two images: Our detected positions (red) and ground truth centers (yellow).

5. REFERENCES

[1] B.M. Kelm, M. Wels, S.K. Zhou, S. Seifert, M. Suehling, Y. Zheng, and D. Comaniciu, "Spine detection in CT and MR using iterated marginal space learning," *Medical Image Analysis*, vol. 17, no. 8, pp. 1283–1292, 2013.

[2] J.J. Corso, R.S. Alomari, and V. Chaudhary, "Lumbar Disc Localization and Labeling with a Probabilistic Model on Both Pixel and Object Features," in *MICCAI 2008*, vol. 5241 of *LNCS*, pp. 202–210. Springer, 2008.

[3] C. Chen, D. Belavy, W. Yu, C. Chu, G. Armbrecht, M. Bansmann, D. Felsenberg, and G. Zheng, "Localization and Segmentation of 3D Intervertebral Discs in MR Images by Data Driven Estimation," *IEEE TMI*, vol. 34, no. 8, pp. 1719–1729, 2015.

[4] C. Chen, D. Belavy, and G. Zheng, "3D Intervertebral Disc Localization and Segmentation from MR Images by Data-Driven Regression and Classification," in *Machine Learning in Medical Imaging*, vol. 8679 of *LNCS*, pp. 50–58. Springer, 2014.

[5] R.S. Alomari, J.J. Corso, and V. Chaudhary, "Labeling of Lumbar Discs Using Both Pixel- and Object-Level Features With a Two-Level Probabilistic Model," *IEEE TMI*, vol. 30, no. 1, pp. 1–10, 2011.

[6] M. Lootus, T. Kadir, and A. Zisserman, "Vertebrae Detection and Labelling in Lumbar MR Images," in *Computational Methods and Clinical Applications for Spine Imaging*, vol. 17 of *LNCV&B*, pp. 219–230. Springer, 2014.

[7] M. Lootus, T. Kadir, and A. Zisserman, "Automated Radiological Grading of Spinal MRI," in *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging*, vol. 20 of *LNCV&B*, pp. 119–130. Springer, 2015.

[8] D. Zukić, A. Vlasák, J. Egger, D. Hořínek, C. Nimsky, and A. Kolb, "Robust Detection and Segmentation for Diagnosis of Vertebral Diseases using Routine MR Images," *Computer Graphics Forum*, vol. 33, no. 6, pp. 190–204, 2014.

[9] Y. Cai, S. Osman, M. Sharma, M. Landis, and S. Li, "Multi-Modality Vertebra Recognition in Arbitrary Views using 3D Deformable Hierarchical Model," *IEEE TMI*, vol. 34, no. 8, pp. 1676–1693, 2015.

[10] S. Zambal, K. Bühler, and J. Hladůvka, "Entropy-optimized Texture Models," in *MICCAI 2008*, vol. 5242 of *LNCS*, pp. 213–221. Springer, 2008.

[11] D. Major, J. Hladůvka, F. Schulze, and K. Bühler, "Automated landmarking and labeling of fully and partially scanned spinal columns in CT images," *Medical Image Analysis*, vol. 17, no. 8, pp. 1151–1163, 2013.

[12] S. Pavani, D. Delgado, and A.F. Frangi, "Haar-like features with optimally weighted rectangles for rapid object detection," *Pattern Recognition*, vol. 43, no. 1, pp. 160–172, 2010.