Improving Detection of Deformable Objects in Volumetric Data

Dominic Mai [1,3]
maid@informatik.uni-freiburg.de
Jasmin Dürr [2]
jasmin.duerr@biologie.uni-freiburg.de
Klaus Palme [2,3]
klaus.palme@biologie.uni-freiburg.de
Olaf Ronneberger [1,3]
ronneber@informatik.uni-freiburg.de

[1] Computer Science Department
University of Freiburg
Germany

[2] Institute of Biology II - Botany
University of Freiburg
Germany

[3] BIOSS Centre for Biological Signalling Studies
University of Freiburg
Germany

Abstract

In this paper, we investigate class level object detection of deformable objects. To this end, we aim for cell detection in volumetric images of dense plant tissue (*Arabidopsis Thaliana*), obtained from a confocal laser scanning microscope. In 3D volumetric data, the detection model does not have to deal with scale, occlusion and viewpoint dependent changes of the appearance, however, our application needs high recall and precision. We implement Felsenszwalb's Deformable Part Model for volumetric data. Corresponding locations for part training are obtained via elastic registration. We identify limitations of its star shaped deformation model and show that a pairwise connected detection model can outperform the star shaped Deformable Part Model in this setting.

1 Introduction

The analysis of biological or medical images is often done by fitting a model to the data. This model fitting is the basis for a further analysis, e.g. to interpret gene expression patterns in the correct anatomical context. If the location of the sought structures is unknown, the model fitting usually consists of a coarse localization step (*detection*) followed by a finer grained fitting to the data (*alignment*), e.g. [3]. A crucial part in this processing is the discriminative power of the detection model: If the wrong object is detected, the best alignment cannot correct for this error. Mai et al. [11] use a rigid detector based on histograms of oriented gradients (HOG) and an elastic registration to detect and align a certain type of cells in *Arabidopsis Thaliana*, in order to get a segmentation and reconstruction of the root. This approach, however, fails to produce good results for other layers with less distinctive cell shapes. The reason is the rigid detection model. While it can account for some local deformation of the cells, as illustrated in Fig. 1(b), false positive detections are frequent

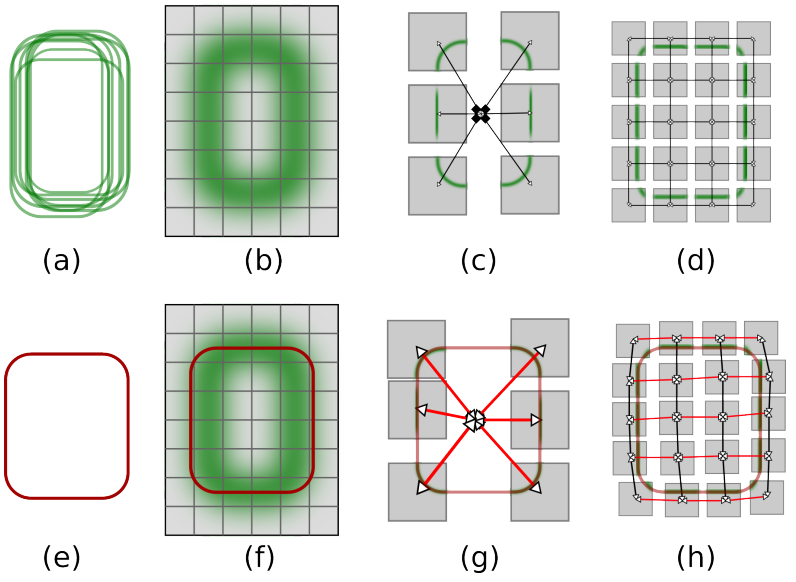


Figure 1: Illustration of different detection approaches and how they deal with deformation. (a) Overlay of the rigidly aligned positive training examples. (b) A rigid detection model allows for small local deformations due to the (soft-) binning of the gradients in the HOG cells. (c) The star shaped structure of the DPM allows parts to move independently. (d) Proposed model: The parts are connected pairwisely. (e) A Detection sample that is wider than most of the training examples. (f) The rigid filter barely detects the object. (g) Every part filter of the DPM has to pay a displacement penalty. (h) The parts of the proposed model only get penalties for horizontal displacements.

if multiple different interpretations of the data result in a similar score. An example is given in Fig. 5: Two small cells are mistaken for one big cell. The detection gets a high score, as the outer hulls of the two neighboring small cells receive high scores, and the context of the cells also fits well. Taking context into account can be beneficial as [9] have shown, but in this case it worsens the situation: in [10] the actual volume of a cell is represented by $\sim 15\%$ of the detection filter.

The *Deformable Part Model* (DPM) [9] augments the rigid detection filter with a set of high resolution part filters, that model parts of the object that move together when the object undergoes a deformation. These part filters are anchored to the root filter and can move independently, thus forming a star shaped deformation model. The gain in detection performance of the DPM over the rigid detection filter is drastic as Felsenszwalb et al. demonstrate on the PASCAL [8] dataset. We will show that the increase is also true in a 3D setting where the variation of the objects mostly stems from their deformable nature while occlusion, scale and viewpoint dependent appearance changes are not present.

However, a shortcoming of this model is illustrated in Fig. 1(g): When an object undergoes an elastic deformation, it is reasonable to assume that neighboring locations move in a similar manner. This means, that locally consistent displace-

ments should not be penalized, while locally inconsistent displacements should. The patches of the DPM can not account for this due to the star shaped structure of the deformation model.

The poselet detection framework [2] offers pairwise connectivity between poselet activations, however, their approach is more focused on modeling appearance changes based on viewpoint and articulation. The poselets are too big for modeling local deformations and the pairwise connections via keypoint predictions are used for a greedy clustering and rescoring.

The research on deformable image registration has produced powerful deformation models for the elastic alignment of data, e.g. [1, 12]. For an overview and a systematic classification of state of the art methods, see [14]. Many classes of deformations can be modeled, however, those models are not designed to have the discriminative power of above mentioned detection approaches. A comparison of alignments from different objects based on their registration score is therefore not very meaningful, as we show in the experiments section.

Contribution. We combine the ideas of discriminative detection and elastic registration by using a discriminative similarity measure with a pairwise deformation model, see Fig. 1(d,h). To this end, we show that deformable detection approaches can be formulated in a general elastic registration framework. We implement the proposed pairwise model and the DPM with a non-latent part training for 3D volumetric data. We show that we can improve precision and recall by rescoring with the alignment scores of the proposed model. We plan to make our software publicly available[1].

2 Modeling Deformations

In this section we introduce the formal framework used for deformable registration. We then show that detection of deformable objects is an instance of deformable registration by formulating two popular approaches in this framework and thereby motivate our approach.

2.1 Deformable Registration

The goal in non-linear deformable registration is to estimate a transformation $\theta : \mathbb{R}^3 \to \mathbb{R}^3$ that transforms the volumetric moving image M onto the fixed image F such that it is most similar to the fixed image with respect to a given similarity measure. We define a volumetric image I as a function:

$$I : \mathbb{R}^3 \to \mathbb{R}. \tag{1}$$

We can denote the transformation of the moving image M with the function composition operator $\circ$:

$$\tilde{M} = M \circ \theta \tag{2}$$

[1] http://lmb.informatik.uni-freiburg.de/people/maid/

Formally, we can now state the goal of the deformable registration as the optimization of the following energy functional:

$$E(F,M,\theta) = S(F,M\circ\theta) + \lambda R(\theta) \tag{3}$$

S measures the degree of alignment between the fixed image F and the transformed moving image $M\circ\theta$. R regularizes the transformation θ. λ controls the strength of this regularization. S is a mapping from the joint image space to the real values $(\mathbb{R}^3\to\mathbb{R},\mathbb{R}^3\to\mathbb{R})\to\mathbb{R}$.

2.2 Discriminative Detection

The goal in object class detection is to locate an object of a certain class in an image. We will first have a look at rigid detection, as illustrated in Fig. 1(b). This approach can be formalized in the elastic registration framework (Eqn. 3):

$$E_{\text{rigid}}(C,M,\mathbf{r}) = C\left(\mathbf{f}\left(M\circ\theta_{\mathbf{r}}\right)\right), \tag{4}$$

where $\mathbf{f}$ crops the image and transforms it into a vectorial feature-space. C is the classifier that obtains the similarity score of the observed image region and the object model. The transformation $\theta_{\mathbf{r}}$ is parameterized by $\mathbf{r}$. In rigid 2D detection for natural images it usually has three degrees of freedom: two translational (r_1,r_2) and one for scale (r_3). Transformations that result in local optima of E give the object locations. This approach was popularized by Dalal and Triggs [5] who introduced HOG features and use a linear support vector machine (SVM) as classifier:

$$C\left(\xi_{\mathbf{r}}\right) = \langle\mathbf{w},\xi_{\mathbf{r}}\rangle, \tag{5}$$

with $\mathbf{w}$ being the separating hyperplane. For simplicity of notation we define $\xi_{\mathbf{r}} := \mathbf{f}\left(M\circ\theta_{\mathbf{r}}\right)$. They optimize it by performing an exhaustive search over all locations and scales using a sliding window approach in a Gaussian scale-space. There is no penalty on the applied transformation ($\lambda = 0$), as all transformations are assumed to be equally likely. However, the ability of the model to cope with small local deformations is encoded in the HOG feature representation (Fig. 1(f)): No penalty is paid for a gradient that moves within a feature cell. Due to the soft binning, a gradient that enters a neighboring cell receives a lower score.

Felsenszwalb et al. [7] augmented the rigid detection model with a star shaped deformation model that is regularized by a quadratic penalizer. The concept of their *Deformable Part Model* is illustrated in Fig. 1(c): The Model consists of a root filter and several part filters, whose locations are "anchored" to the root filter. The root filter is a rigid detection filter as introduced by Dalal and Triggs and responsible for the coarse localization of the object. The part filters are much smaller than the root filter, but at twice the resolution. They are responsible for the precise localization of regions of the object that are assumed to move together when the object undergoes a deformation.

All filters $\mathbf{w}_i$ are realized as linear SVMs, with $\mathbf{w}_0$ being the root filter. During training, the relative mean locations of the parts $\mathbf{w}_1\ldots\mathbf{w}_n$ are learned along with the

costs for a displacement of the part from its most likely location:

$$E_{\mathrm{DPM}}(\mathbf{w_0},\ldots,\mathbf{w_n},M,\mathbf{r},\mathbf{t}_1,\ldots,\mathbf{t}_n) = \langle \mathbf{w}_0, \xi_{\mathbf{r}} \rangle + \sum_{i=1}^{n} \langle \mathbf{w}_i, \xi^i_{\mathbf{t}_i} \rangle - \lambda \sum_{i=1}^{n} R\left(\mathbf{t}_i\right) \tag{6}$$

The root filter location and scale $\mathbf{r}$ are searched exhaustively as in the rigid setting. For each location the optimal relative part-placements $\mathbf{t}_i, i \in \{1,\ldots,n\}$ are independent translations that can be found efficiently by using generalized distance transforms [8]. Note that the root filter location and the part placements also could have been expressed by a parameterized dense transformation $\theta(\mathbf{r},\mathbf{t}_1,\ldots,\mathbf{t}_n)$ in the sense of Eqn. 3.

Mai et al. [11] use elastic registration to generate a *sharp mean image* that represents a cluster of positive training images. They choose the image with minimal average bounding box distance to all other images in the cluster as reference image. All other images are elastically registered to this reference image (Fig. 2). We took over this strategy and use the deformation fields to establish correspondences on the training examples of every cluster. Therefore we do not need to treat the part locations as latent variables as we know the exact locations of the patches on every training image.

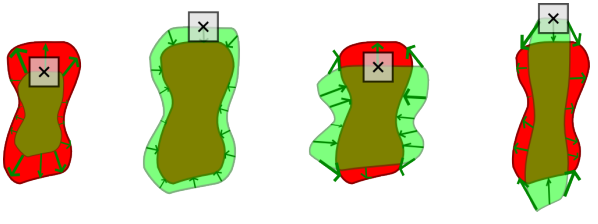


Figure 2: The training images (green) of one cluster are all registered to the reference image (red). The deformation fields (green arrows) can be used to obtain corresponding locations and the expected displacement of a patch.

3 Proposed model

As we have seen, we can formulate the DPM approach (Eqn. 6) in the elastic registration framework (Eqn. 3). We are now able to motivate our model from two perspectives:

1. Coming from the *DPM*, we change this model in the following way: We replace the star shaped sparse part model with a pairwisely connected dense part model. Therefore we do not need the anchoring to the root filter anymore. But as it would be computationally too demanding to optimize the model in a detection setting, with the majority of locations *not* containing the sought object, we also can motivate our approach

2. coming from the deformable registration framework point of view. Here we introduce a new discriminatively trained data term that can easily be integrated into existing frameworks as we show in the experiments section.

We propose a *Discriminative Deformable Model* (DDM):

$$E_{\text{prop}}(\mathbf{w}_1,\ldots,\mathbf{w}_{N_p},M,\mathbf{U}) = \underbrace{\sum_{i=1}^{N_p}\langle \mathbf{w}_i, \xi_{\theta(\mathbf{p}_i)}\rangle}_{\text{discriminative data term}} \underbrace{-\lambda \sum_{(i,j)\in\mathcal{E}} \frac{\left\|\mathbf{U}(\mathbf{p}_i)-\mathbf{U}(\mathbf{p}_j)\right\|_2}{\left\|\mathbf{p}_i-\mathbf{p}_j\right\|_2}}_{\text{pairwise deformation model}} \tag{7}$$

For the explanation of the symbols, have a look at Fig. 1(d): We discretely sample N_p control points $\mathbf{p}_i$ at equidistant image locations. Each control point is the center of a linear support vector machine $\mathbf{w}_i$ that encodes the appearance for the respective patch. We will train all the $\mathbf{w}_i$ jointly. $\mathbf{U}:\mathbb{R}^3\to\mathbb{R}^3$ is a displacement field. The relationship between a transformation and a displacement field is the following: $\theta(\mathbf{p})=\mathbf{U}(\mathbf{p})+\mathbf{p}$. The feature representation of an image patch extracted at the location $\mathbf{p}+\mathbf{U}(\mathbf{p})$ is represented by $\xi_{\theta(\mathbf{p})}$. The regularizer penalizes inconsistent movements of neighboring patches: The deviation vector $\mathbf{d}=\mathbf{U}(\mathbf{p}_1)-\mathbf{U}(\mathbf{p}_2)$ is $\mathbf{0}$ when two connected control points move with the same displacement. It is normalized by the initial edge length $\left\|\mathbf{p}_i-\mathbf{p}_j\right\|_2$. $\mathcal{E}$ is the connectivity set, i.e. all the edges (i,j) between the parts. λ controls the degree of regularization.

3.1 Training

In this section we describe how we train the patch detectors $\mathbf{w}_i$ and estimate λ which controls the degree of regularization. As the patch detectors are linear support vector machines, we need to present them with normalized positive and negative training examples. For the normalization, the effect of the transformation θ needs to be canceled out. For instance, to train a rigid detector, the positive examples need to be position and scale normalized (Fig. 1(a)). As we want to normalize for non linear deformations, we use the detection and alignment framework from [11] to mine shape normalized positive and negative training examples.

We describe the processing pipeline for a single detection hypothesis yielded from a single detector. We put the *sharp mean image* Z on the rotation normalized image B at the detection location. We compute the alignment $\tilde{Z}=Z\circ\theta$ using the elastic registration from [12]. Then we compare the corresponding warped segmentation mask $\tilde{S}_Z$ with the ground truth segmentation to classify the detection whether it will be the basis for a positive (+) or a negative (−) training example. Now we shape normalize the image B at the detection location by aligning it to the sharp mean image with the inverse transformation: $\tilde{B}=B\circ\theta^{-1}$. We create a set of N_p equidistant control points $\mathbf{p}_i$, such that the induced patches at the control point locations $\mathbf{p}_i$ form a partition of the sharp mean image Z. We compute 3D HOG features from the shape normalized image $\tilde{B}$ and extract the patches $\xi_{\mathbf{p}_i}$ for every control point.

We process all detections obtained from the training dataset in the described manner and end up with a positive training set $\mathcal{S}_i^+=\left\{\xi_{\mathbf{p}_i}^1,\ldots,\xi_{\mathbf{p}_i}^{N_+}\right\}$ and a negative training set $\mathcal{S}_i^-=\left\{\xi_{\mathbf{p}_i}^1,\ldots,\xi_{\mathbf{p}_i}^{N_-}\right\}$ for each control point $\mathbf{p}_i$.

Finally, we jointly train the patch detectors $\mathbf{w}_i$ using *libsvm* [4]. To this end we concatenate the features from all patches. A training vector $\mathbf{T}_i$ has the following

structure:

$$\mathbf{T}_i = \left[(\xi^i_{\mathbf{p}_1})^\intercal, \ldots, (\xi^i_{\mathbf{p}_{|\mathcal{P}|}})^\intercal \right], \quad i \in [1,N], \quad N = N_+ + N_- \tag{8}$$

After training, the joint separating hyperplane has the structure $\mathbf{W} = \left[\mathbf{w}_1^\intercal, \ldots, \mathbf{w}_{|\mathcal{P}|}^\intercal \right]^\intercal$ — the single patch detectors $\mathbf{w}_i$ can thus easily be obtained.

As the codomain of the $\mathbf{w}_p$ is not normalized, it is crucial to find a good value for the strength of the regularization λ. We achieve this by doing an exponential grid search with the goal of maximizing the intersection over union of the aligned segmentation masks with the ground truth segmentations.

3.2 Optimization

The structure of our model Eqn. 7 is identical to the model used in the *vibez* registration [12]. The only difference is that we replace the normalized cross correlation based data term with the discriminative patch detectors $\mathbf{w}_i$. We formulate it as a discrete labeling problem (*Markov Random Field*) with unary and binary costs. The sought label is the displacement $\mathbf{U}(\mathbf{p}_i)$ at a control point. This means that we have to evaluate a discrete set of displacement hypotheses $\mathbf{d}_j \in \mathbb{R}^3$ at every control point to obtain the *unary costs* $-\langle \mathbf{w}_i, \xi_{\mathbf{p}+\mathbf{d}_j} \rangle$. As in the vibez registration, we search along the axes and diagonals within a certain radius r, therefore we have $\mathbf{d}_j \in \{r \cdot (e_1,e_2,e_3)^\intercal \quad | \quad r \in \mathbb{N}, e_i \in \{-1,0,1\}\}$. The initial dense displacement field $\mathbf{U}$ is set to zero.

The *binary costs* are given by the regularizer of the model, as they couple neighboring control points: $\lambda \sum_{(\mathbf{p}_1,\mathbf{p}_2)\in\mathcal{E}} \frac{1}{\|\mathbf{p}_1-\mathbf{p}_2\|_2} \left\| \mathbf{U}(\mathbf{p}_1) - \mathbf{U}(\mathbf{p}_2) \right\|_2$.

We solve it with iterated graph cuts using the very efficient *fastPD* algorithm [9]. The values of $\mathbf{U}$ between the control points are obtained through a cubic spline interpolation. As the performance breaks down for very big search radii r, we need a coarse initialization. Therefore we run the optimization on top of the detection hypotheses of a rigid detector or a DPM detector.

4 Experiments

For the quantitative evaluation we had two roots (*r06, r14*) with ground truth segmentation available. We use one root for training and one for testing and vice versa. Each root contains about $\sim$2500 cells, so the generation of ground truth is cumbersome work. It is obtained by manually verifying a watershed segmentation computed on enhanced data [10]. For the available roots, this bottom up segmentation method gives good results, but it fails for data of worse quality [10], where the generation of reliable ground truth segmentations is nearly impossible. However, the goal of this section is to give a quantitative comparison between the different detection and alignment strategies discussed in this paper.

We chose a cell layer (Fig. 3), where the rigid detection followed by alignment approach from [10] fails to produce good results (Fig. 4). We follow their training procedure and split the training data into 15 clusters. For each cluster, we train a *rigid* detector and a *DPM* detector, as well as a *sharp mean image* for the gradient

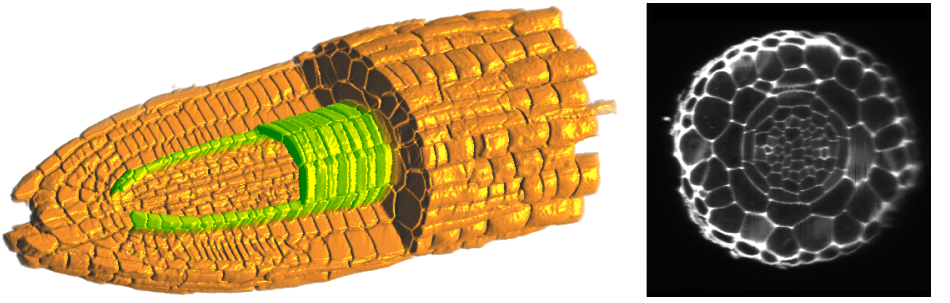
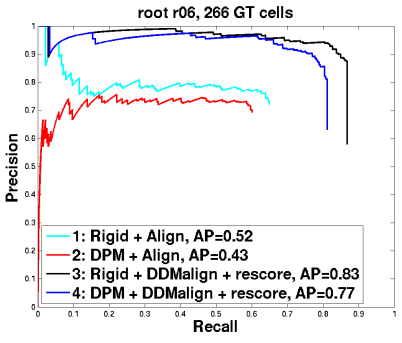


Figure 3: **(left)** A Volume rendering of the root *r06*, the layer used for training and detection is colored green. **(right)** A slice of the original raw data.

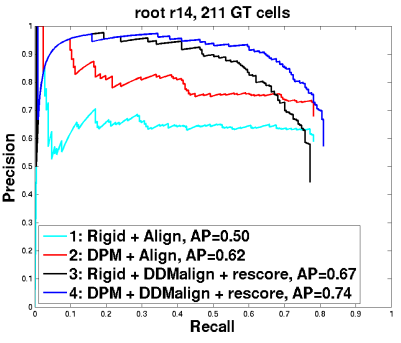
based alignment and the proposed *Discriminative Deformable Model* (DDM). The roots have a volume of $944 \times 413 \times 360$ (r14) and $1030 \times 433 \times 384$ (r06) voxels, each with a root coordinate system [13] attached for rotation normalization. As the sliding window detector cannot account for rotations, we extract 3×36 overlapping rotation normalized boxes along the main axis with an angle stepping of $10°$ and a size of $301 \times 101 \times 131$ voxels. All detectors work on 3D HOG features [11] with 20 bins for the gradient orientation, equally sampled on the unit sphere. The spatial pooling is realized by a triangular filtering with radius H_r and a subsampling with a factor of H_s. The rigid detectors have an average size of $54 \times 65 \times 43$ voxels, the HOG features are parameterized with $H_r = 5$ and $H_s = 2$. For the DPM detectors, the rigid detectors are augmented by 10 cubic parts with size of 12^3 voxels at full resolution ($H_s = 1$, $H_r = 3$) to enable a spatially more precise localization. The parts are placed evenly across the border of the cells: At those locations, the positive root filter energy is the highest and the displacement fields for the deformation learning (Fig. 2) are the most reliable. The patch size of the DDM is set to 7^3 voxels, as this is the best value for the vibez registration in this context. The patch-SVMs are also trained at full resolution ($H_s = 1$, $H_r = 3$) to allow for spatially precise localizations.

For testing, we run the 15 rigid and DPM detectors on the rotation normalized boxes of both roots: $\{r06, r14\} \times \{rigid, DPM\}$. We use a sliding window approach and we compute the convolutions efficiently in the Fourier domain. We perform a non-maximum suppression based on the volume of the segmentation mask corresponding to the detector. On a six core workstation, for one rotation normalized box, the rigid detection for 15 clusters takes $\sim 50s$, the DPM detection $\sim 835s$ while using $\sim 10\times$ more memory. The time is spent mostly on the computations of the score maps for the patches. For each detection hypothesis, we align the sharp mean image with the gradient based data term from [11] and the proposed *DDM*. The computation of the alignment takes $\sim 1.5s$ for either alignment strategy. This means, that during the time needed for one DPM detection, we can run the rigid detection and ~ 500 alignments. Note that the computations for detection and alignment are nearly perfectly parallelizable. Finally, we transform the aligned detections back to the original root. We do this in a greedy fashion, starting with the highest scoring detection.

We investigate 4 strategies (Fig. 4): 1. Rigid detection with alignment of the sharp mean image using the gradient based data term (cyan) 2. DPM detection with an



(a)



(b)

Figure 4: Precision-recall Graphs of the different detection strategies for the two roots (a) r06 and (b) r14. The alignment and rescoring with the proposed *Discriminative Deformable Model* (DDMalign, **black** curve and **blue** curve) produces the best results, independent of the underlying detector.

alignment of the sharp mean image using the gradient based data term (red) 3. Rigid detection with an alignment of the sharp mean image using the Discriminative Deformable Model and rescoring (black) 4. DPM detection with an alignment of the sharp mean image using the Discriminative Deformable Model and rescoring (blue).

We evaluate them using the *intersection over union* (IOU) measure. Therefore we compare the implied segmentation mask of the aligned detection hypothesis with the ground truth segmentation of the cell. For two volumes V and W the IOU is defined as

$$M_{iou}(V,W) = \frac{|V \cap W|}{|V \cup W|} \quad . \tag{9}$$

Following the PASCAL VOC criterion [6], We accept a detection as valid, iff IOU > 0.5. We visualize the results in a precision-recall graph (Fig. 4): The detections are sorted in descending order by thr detector score or the DDM alignment score. Then for every recall value $r \in [0,\ldots,r_{max}]$, the corresponding precision is plotted. The average precision (AP) is computed as the area below the precision-recall curve.

We can see that the result is relatively independent of the underlying detection approach. While the rigid detection based strategies perform better on r06 (Fig. 4(a)), the DPM detection based strategies are better on r14 (Fig. 4(b)). The reason why the DPM does not perform consistently better than the rigid detector can be found in the gradient based alignment step that is performed for every detection hypothesis. Prior to the alignment, the DPM detector has a better yield ratio of valid detections with an IOU > 0.5 over the rigid detector: 0.37 vs. 0.3 for r06 and 0.33 vs. 0.29 for r14. However, the alignment step that improves the precise localization of the detections drastically [11] seems to cancel out the advantage of the DPM prior to the alignment.

Both strategies that use a gradient based alignment without rescoring (1,2) produce a mean AP around ~ 0.52, with an average IOU of 0.70 for the valid detections. The strategies based on the *DDM* based alignment with rescoring outperform the

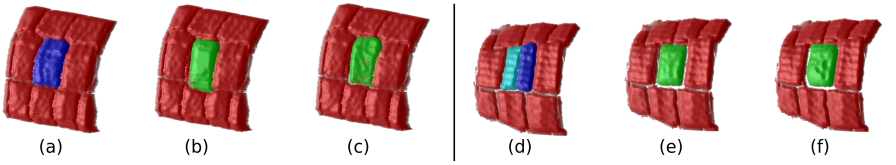


Figure 5: Two examples for the rescoring with the Discriminative Deformable Model. (a) Example for a ground truth cell (blue). (b) It is coarsely detected. (c) The alignment with the DDM results in a better localization and a better score for this detection. (d) Two ground truth cells in (cyan) and (blue). (e) A detector for a bigger cell produces a false positive detection. (f) The DDM alignment also encompasses both ground cells (false positive), but the produced alignment score is very low: In the greedy reconstruction, this detection is very unlikely to be considered, as its location will most likely already be taken by correct detections with higher alignment scores.

others by a margin of 0.23 percentage points with a mean AP of 0.75. The average IOU for the valid detections is 0.69. The reason for this substantial gap can be found in the more expressive scores produced by the DDM. We did not find an improvement by rescoring with the alignment scores of the gradient based data term. Two qualitative examples of the rescoring are given in Fig. 5.

5 Conclusions

In this paper, we introduced the *Discriminative Deformable Model* and showed that we can improve the average precision for volumetric deformable detection substantially. We achieved this by connecting the successful discriminative classification ideas from class level object detection with the powerful deformation models and highly efficient optimization methods from nonlinear deformable image registration.

Acknowledgements

This work was supported by the Excellence Initiative of the German Federal and State Governments: BIOSS Centre for Biological Signalling Studies (EXC 294) and the Bundesministerium für Bildung und Forschung (SYSTEC, 0101-31P5914).

References

[1] Brian B. Avants, Charles L. Epstein, Murray Grossman, and James C. Gee. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1):26–41, 2008.

[2] Lubomir Bourdev, Subhransu Maji, Thomas Brox, and Jitendra Malik. Detecting people using mutually consistent poselet activations. In *ECCV*, 2010.

[3] Thomas Brox, Lubomir Bourdev, Subhransu Maji, and Jitendra Malik. Object segmentation by alignment of poselet activations to image contours. In *IEEE*

International Conference on Computer Vision and Pattern Recognition (CVPR), 2011.

[4] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.

[5] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In Cordelia Schmid, Stefano Soatto, and Carlo Tomasi, editors, *International Conference on Computer Vision & Pattern Recognition*, 2005.

[6] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2):303–338, June 2010.

[7] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object Detection with Discriminatively Trained Part-Based Models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1627–1645, September 2010. ISSN 0162-8828. doi: 10.1109/TPAMI.2009.167.

[8] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Distance transforms of sampled functions. *Theory of Computing*, 8(19):415–428, 2012. doi: 10.4086/toc.2012.v008a019. URL http://www.theoryofcomputing.org/articles/v008a019.

[9] Nikos Komodakis, Georgios Tziritas, and Nikos Paragios. Performance vs computational efficiency for optimizing single and dynamic mrfs: Setting the state of the art with primal-dual strategies. *Computer Vision and Image Understanding*, 112(1):14–29, 2008.

[10] K. Liu, T. Schmidt, T.Blein, J. Dürr, K. Palme, and O. Ronneberger. Joint 3d cell segmentation and classification in the arabidopsis root using energy minimization and shape priors. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2013.

[11] Dominic Mai, Philipp Fischer, Thomas Blein, Jasmin Dürr, Klaus Palme, Thomas Brox, and Olaf Ronneberger. Discriminative detection and alignment in volumetric data. In Joachim Weickert, Matthias Hein, and Bernt Schiele, editors, *GCPR*, volume 8142 of *Lecture Notes in Computer Science*, pages 205–214. Springer, 2013. ISBN 978-3-642-40601-0.

[12] O Ronneberger, K Liu, M Rath, D Ruess, T Mueller, H Skibbe, B Drayer, T Schmidt, A Filippi, R Nitschke, T Brox, H Burkhardt, and W Driever. Vibez: A framework for 3d virtual colocalization analysis in zebrafish larval brains. *Nature Methods*, 9(7):735–742, 2012.

[13] T. Schmidt, M. Keuper, T. Pasternak, K. Palme, and O. Ronneberger. Modeling of sparsely sampled tubular surfaces using coupled curves. In *Pattern Recognition (Proc. DAGM)*, volume 7476 of *LNCS*, pages 83–92. Springer, 2012.

[14] Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. Deformable Medical Image Registration: A Survey. *IEEE Transactions on Medical Imaging*, 32(7): 1153–1190, May 2013. doi: 10.1109/TMI.2013.2265603. URL http://hal.inria.fr/hal-00858737.