TUM

TECHNISCHE UNIVERSITÄT MÜNCHEN
INSTITUT FÜR INFORMATIK

Morphological Proximity Priors: Spatial Relationships for Semantic Segmentation

Julia Bergbauer, Claudia Nieuwenhuis, Mohamed Souiai, Daniel Cremers

TUM-I1222

Morphological Proximity Priors: Spatial Relationships for Semantic Segmentation

Julia Bergbauer, Claudia Nieuwenhuis, Mohamed Souiai and Daniel Cremers
Technical University of Munich, Germany
{julia.bergbauer,claudia.nieuwenhuis,mohamed.souiai,daniel.cremers}@in.tum.de

Abstract

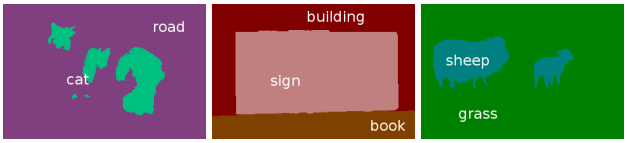
The introduction of prior knowledge into image analysis algorithms is a central challenge in computer vision. In this paper, we introduce prior knowledge about spatial relations of objects. Instead of using absolute *learned spatial priors, i.e. that 'sky' is always found in the upper half of the image, we introduce* relative *spatial geometric relationships, i.e. that specific objects are usually found above or below other objects or that they appear close to each other or far apart. We integrate such relative spatial object relationships into a variational approach by means of morphological operators and derive an exact convex relaxation which can be minimized globally. Extensive validations on an established semantic segmentation benchmark demonstrate the improved performance of the proposed method.*

1. Introduction

Image segmentation is an essential component in image content analysis and one of the most investigated problems in computer vision. The goal is to partition the image plane into 'meaningful' non-overlapping regions. Especially for complex real-world images, however, the definition of meaningful depends on the application or the user's intention. Typically, the desired segmentation consists of one region for each separate object or structure of the scene. Due to strongly varying texture and color models within and between different object classes, the segmentation task is very complex and requires additional prior information. For example animals such as horses, cows and sheep have similar color models and similarly textured fur. Since most segmentation algorithms only consider local color or texture information to assign each pixel to an object class, they often generate incorrect segmentations, where e.g. part of the cat is assigned the label 'road' as shown in Figure 1b).

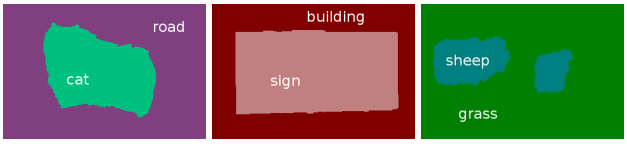
For humans the task of recognizing objects strongly relies on their context and inter-relations with other objects. Yet, up to now most segmentation approaches only contain

a) Input images



b) Co-occurrence prior by Ladicky *et al*. [4].

c) Non-metric prior by Strekalovskiy *et al*. [14].



d) Proposed geometric spatial proximity prior.

Figure 1: **Morphological Proximity Priors.** Statistics on object co-occurrence probabilities provide strong priors for semantic segmentation. Previous approaches fail since they either entirely neglect spatial object relations (b) or only consider directly adjacent object labels (c). The proposed proximity priors impose spatial geometric relations between objects such as distances, direction and relative location.

absolute priors on typical object locations in an image. We believe that especially *relative object relations* describing the distance and location of an object with respect to other objects are of interest.

To cope with the above mentioned difficulties we, therefore, learn statistics on spatial geometric relationships of objects, i.e. on their distance, direction and position with respect to each other. Examples of such relative relations are that cars do not appear above water or that cows and

sheep are usually not found in close proximity. The challenge we face in this paper is to find an efficient and convex optimization approach for multilabel segmentation with *relative* geometric spatial priors. Improved results based on geometric spatial object relationships in comparison to related approaches are shown in Figure 1.

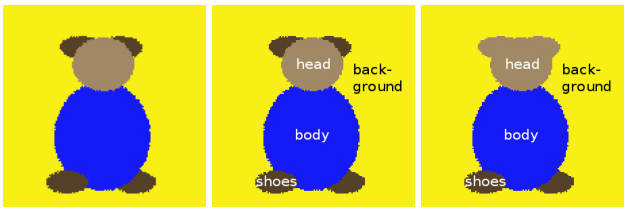
1.1. Related Work

In the past, the integration of label priors into multilabel segmentation approaches has been a successful means for improving segmentation results.

Global label priors impose prior knowledge on the union of all pixel labels in the image. Examples are the minimum description length (MDL) prior [5, 18, 17] and the connectivity prior [16, 9]. Ladicky *et al*. [4] proposed a global co-occurrence prior to penalize simultaneously occurring label sets in the image. The use of higher-order potentials for MRF optimization in this context was promoted by Kohli *et al*. [3]. Global priors disregard the spatial location of labels and the corresponding object sizes. As a consequence, if more pixels vote for a certain label then they may easily overrule penalties imposed by the label prior term – leading to segmentations such as the ones in Figure 1b) with large adjacent regions despite large co-occurrence cost for 'book' and 'building'.

Relative spatial label priors impose constraints on spatial relations between different objects. An example are the 'ordering constraints' for geometric scene labeling ('sky', 'ground', 'left', 'right', 'center') by Liu *et al*. [6] and by Strekalovskiy *et al*. [12]. Relative relations concerning the containment of regions within others or their exclusion were introduced into a Markov Random Field approach by Delong and Boykov [1]. For penalizing directly adjacent pixel labels Strekalovskiy *et al*. [14] introduced a local non-metric prior. The drawback of this approach is that the algorithm can avoid costly label transitions simply by introducing infinitesimal 'ghost labels' – see Figure 6. Furthermore, due to the strong locality the prior allows for regions to appear close to each other despite high co-occurrence penalties (see the labels 'sheep' and 'cow' in Figure 1c)). Considering more complex spatial relations between objects, i.e. *relative* spatial priors, will allow for a better integration of prior knowledge on label relations.

Learned relative location priors were already proposed by Gould *et al*. [2] in 2008. The authors formulated a two stage optimization problem, where they compute superpixels together with an occurrence based label likelihood in the first step. Based on the most likely label each superpixel then votes for labels at other superpixels in the image based on the relative location prior. In contrast we propose a single stage optimization problem, which optimizes presence and relative location likelihoods at the same time. We directly work on the pixel level without superpixels and exceed the results by Gould *et al*. [2] on the MSRC benchmark by more than 8%.



a) Original | b) Color-based segmentation | c) Desired segmentation

Figure 2: **Purely color-based segmentation often fails.** The ears of the bear are assigned the label 'shoes' instead of being combined with the head.

1.2. Contributions

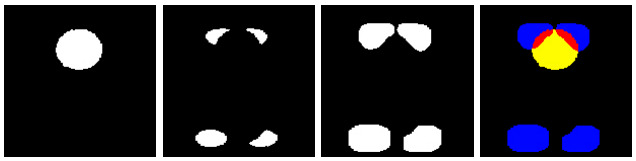
In this paper, we propose *morphological proximity priors* for semantic image segmentation. Specifically, we make the following contributions:

- We integrate learned geometric spatial relationships between different objects into a variational multi-label segmentation approach by means of morphological dilation operations.
- The specific geometric prior can be determined by simply defining the size and shape of the morphological structuring element for each label.
- We do not rely on prior superpixel partitions but directly work on the pixel level.
- We avoid the emergence of artificial 'ghost labels'.
- We give a convex relaxation which can be solved with fast primal-dual algorithms [11] in parallel on graphics hardware (GPUs).

2. Morphological Label Proximity Priors

We motivate the morphological proximity priors by means of the simple artificial teddy bear example in Figure 2a). Common segmentation approaches group pixels mainly according to their color, hence the ears of the bear are associated with the region 'shoes' (Figure 2b). The desired result, however, would rather connect the ears to the head instead of the shoes as shown in Figure 2c).

To obtain the desired solution, we make use of a *dilation*, an operation from mathematical morphology. To examine if two regions are close to each other in any direction we dilate one of the regions and compute the overlap between the dilation and the second region. For the teddy example, we enlarge the region 'shoes' and compute the overlap with the region 'head' as shown in Figure 3. We do not only consider



a) 'Head' b) 'Shoes' c) Dilation d) Overlap

Figure 3: **'Proximity' of two labels.** Proximity priors penalize the 'closeness' of two labels, in this example a) 'head' and b) 'shoes'. c) Dilation of the indicator function 'shoes', d) Overlap of the dilated region 'shoes' and the region 'head'. Appropriate penalties for such overlap introduce semantic information into the segmentation.

directly neighboring pixels as close but we consider proximity with respect to arbitrary neighborhoods. The size and shape of these neighborhoods is determined by the *structuring element* of the dilation and can thus be easily adapted.

2.1. A Continuous Formulation of the Dilation

Dilation is one of the basic operations in mathematical morphology. For general not necessarily binary images it can be defined as follows:

Definition 2.1 (Dilation of an image) *Let $I : \Omega \to \mathbb{R}^d$ be an image and $\mathcal{S} \subseteq \Omega$ a structuring element. The dilation of I by $\mathcal{S}$ is then given by*

$$(I \oplus s)(x) = \sup_{y \in \Omega} [I(y) + s(x-y)] = \sup_{z \in S} I(x+z),$$

$$\text{where } s(x) = \begin{cases} 0, & x \in \mathcal{S}, \\ -\infty, & otherwise. \end{cases} \quad (1)$$

Thus, the dilation result at a given location x in the image is the maximum value of the image within the window defined by the structuring element $\mathcal{S}$, when its origin is at x. The structuring element determines the type of geometric spatial relationship we want to penalize, see Figure 4. Symmetric structuring elements of specific sizes (Figure 4b)) consider the proximity of two labels without preference of a specific direction. By replacing the symmetric structuring element for example by a line we can penalize the proximity of specific labels in specific directions, e.g. the occurrence of a book below a sign by means of a vertical structuring element (compare Figure 4c)). The larger $\mathcal{S}$ the more pixels are considered as adjacent to x. The possibility to use structuring elements of different sizes and shapes is one of the major benefits of the proposed algorithm.

3. Variational Image Segmentation with Morphological Proximity Priors

Let $I : \Omega \to \mathbb{R}^d$ denote the input image defined on the image domain $\Omega \subset \mathbb{R}^2$. The general multilabel image segmentation problem with $n \geq 1$ labels consists of the partitioning of the image domain Ω into n regions $\{\Omega_1, \ldots, \Omega_n\}$ with $\Omega = \bigcup_{i=1}^{n} \Omega_i$ and $\Omega_i \cap \Omega_j = \emptyset \ \forall i \neq j$. This task can be solved by computing binary labeling functions $u_i : \Omega \to \{0, 1\}$ in the space of functions of bounded variation (BV) for $i = 1, \ldots, n$ indicating which of the n regions each pixel belongs to. We obtain $\Omega_i = \left\{ x \,\middle|\, u_i(x) = 1 \right\}$.

In order to assign each pixel to a specific label, we require some connection between the image and the different labels. We use color likelihoods $P(I(x) \,|\, x \in \Omega_i)$, which indicate the probability that a given pixel x with color $I(x)$ belongs to region Ω_i. In this paper, we use the same color likelihoods as Ladicky *et al*. [4].

Based on these color likelihoods, we compute a segmentation of the image by minimizing the following energy

$$E(\Omega_1, .., \Omega_n) = \frac{\lambda}{2} \sum_{i=1}^{n} \text{Per}_g(\Omega_i) + \sum_{i=1}^{n} \int_{\Omega_i} f_i(x)\, dx,$$
$$\text{where } f_i(x) = -\log P\left(I(x) \,\middle|\, x \in \Omega_i\right). \quad (2)$$

$\text{Per}_g(\Omega_i)$ denotes the perimeter of each set Ω_i, which is minimized in order to favor segments of shorter boundary. These boundaries are measured with either an edge-dependent or an Euclidean metric defined by the non-negative function $g : \Omega \to \mathbb{R}^+$. For example,

$$g(x) = \exp\left(-\frac{|\nabla I(x)|^2}{2\sigma^2}\right), \ \sigma^2 = \frac{1}{|\Omega|} \int_{\Omega} |\nabla I(x)|^2 dx$$

favors the coincidence of object and image edges.

To rewrite the perimeter of the regions in terms of the indicator functions we make use of the total variation:

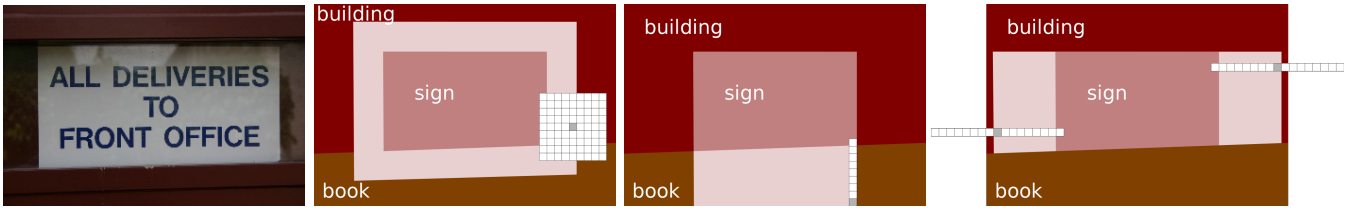
$$\text{Per}_g(\Omega_i) = \int_{\Omega} g(x) |Du_i| = \sup_{\xi_i : |\xi_i(x)| \leq g(x)} - \int_{\Omega} u_i \operatorname{div} \xi_i \, dx.$$

Since the binary functions u_i are not differentiable Du_i denotes their distributional derivative. Furthermore, $\xi_i \in \mathcal{C}_c^1(\Omega; \mathbb{R}^2)$ are the dual variables and $\mathcal{C}_c^1$ denotes the space of smooth functions with compact support. We can rewrite the energy in (2) in terms of the indicator functions $u_i : \Omega \to \{0, 1\}$:

$$E(u_1, .., u_n) = \sup_{\xi \in \mathcal{K}} \sum_{i=1}^{n} \int_{\Omega} (f_i - \operatorname{div} \xi_i)\, u_i \, dx, \quad (3)$$
$$\text{where } \mathcal{K} = \left\{ \xi \in \mathcal{C}_c^1(\Omega; \mathbb{R}^{2 \times n}) \,\middle|\, |\xi_i(x)| \leq \frac{\lambda g(x)}{2} \right\}.$$

3.1. The Novel Proximity Prior

To introduce the proximity prior into the optimization problem in (3), we define the proximity matrix $A \in \mathbb{R}_{\geq 0}^{n \times n}$. Each entry $A(i, j),\ i \neq j$ indicates the penalty for the occurrence of label j in the proximity of label i, which we



a) Original Image b) Symmetric SE c) Vertical SE d) Horizontal SE

Figure 4: **Possible dilations.** Different types of structuring elements (SE) incorporate different geometric spatial prior knowledge. Their origin is marked by the gray pixel. b) Symmetric elements only consider object distances, but are indifferent to directional relations. c) Vertical elements penalize vertical label co-occurrences such as the 'book' below the 'sign'. d) Horizontal elements penalize labels to the left and right. Using the symmetric element we obtain the result in Figure 1d).

denote by $i \sim j$. For $i = j$ we set $A(i,i) := 0$. The penalties can be computed from co-occurrence probabilities of training segmentations, e.g. by $A(i,j) = -\log P(i \sim j)$. An example for a proximity matrix is shown in Figure 5.

To detect if two regions i and j are close to each other, we compute the overlap between the dilation of the indicator function u_i denoted by d_i and the indicator function u_j. According to (1), the dilation of u_i can be formulated as:

$$d_i : \Omega \to \{0,1\} \quad d_i(x) = \max_{z \in \mathcal{S}} u_i(x+z) \quad \forall\, x \in \Omega \tag{4}$$

with the respective structuring element $\mathcal{S}$.

For each two regions i and j we can now penalize their proximity by means of the following energy term

$$\sum_{1 \le i < j \le n} \int_\Omega A(i,j)\, d_i(x)\, u_j(x)\, dx. \tag{5}$$

3.2. A Convex Relaxation

In the following we will propose a convex relaxation of the segmentation problem (3) combined with the proposed proximity prior in (5). To obtain a convex optimization problem, we require convex functions over convex domains.

Figure 5: **Proximity matrix** A**.** Learned penalty matrix for 21 regions and a symmetric structuring element (SE) of size 9×9. The lighter the color the more likely is the occurrence of the corresponding labels within the relative spatial context given by the SE, and the lower is the corresponding penalty.

Relaxation of the Binary Functions u_i The general multilabeling problem is not convex due to the binary region indicator functions $u_i : \Omega \to \{0,1\}$. To obtain a convex problem we will relax these conditions to $u_i : \Omega \to [0,1]$. Since each pixel should be assigned to exactly one label, optimization is carried out over the convex set

$$\mathcal{U} = \left\{ u \in BV(\Omega; [0,1]^n) \,\middle|\, \sum_{j=1}^{n} u_j(x) = 1 \;\; \forall\, x \in \Omega \right\}.$$

Relaxation of the Dilation Constraints The dilation constraints in (4) are not convex. To obtain a convex formulation we relax the constraints to

$$d_i(x) \ge u_i(x+z) \quad \forall\, x \in \Omega,\; z \in \mathcal{S}. \tag{7}$$

By simultaneously minimizing over the functions d_i we can assure that at the optimum d_i fulfills the constraints in (4) exactly. The inequality (7) can easily be included in the segmentation energy by introducing a set of Lagrange multipliers β_{i_z} and adding the following energy term:

$$\min_{d \in \mathcal{D}} \max_{\beta \in \mathcal{B}} \sum_{i=1}^{n} \sum_{z \in \mathcal{S}} \int_\Omega \beta_{i_z}(x) \left(d_i(x) - u_i(x+z) \right) dx, \tag{8}$$

$$\begin{aligned} \mathcal{B} &= \left\{ \beta_{i_z} \,\middle|\, \beta_{i_z} : \Omega \to [-\infty, 0] \;\; \forall\, z \in \mathcal{S},\; i = 1,..,n \right\}, \\ \mathcal{D} &= BV(\Omega; [0,1]^n). \end{aligned}$$

Relaxation of the Product in (5) The product of the dilation d_i and the indicator function u_j is not convex. A convex, tight relaxation of such energy terms was given by Strekalovskiy *et al.* [13]. To this end, we introduce additional dual variables q_{ij} and Lagrange multipliers α_{ij}:

$$\begin{aligned} \mathcal{Q} &= \left\{ q_{ij} \,\middle|\, q_{ij} : \Omega \to \mathbb{R}^4, 1 \le i < j \le n \right\}, \\ \mathcal{A} &= \left\{ \alpha_{ij} \,\middle|\, \alpha_{ij} : \Omega \to [-\infty, 0]^4, 1 \le i < j \le n \right\}. \end{aligned} \tag{9}$$

$$\min_{\substack{u\in\mathcal{U}\\ d\in\mathcal{D}\\ \alpha\in\mathcal{A}}} \max_{\substack{\xi\in\mathcal{K}\\ \beta\in B\\ q\in\mathcal{Q}}} \sum_{i=1}^{n} \Big\{ \int_\Omega (f_i - \operatorname{div}\xi_i) u_i \, dx + \sum_{z\in\mathcal{S}} \int_\Omega \beta_{i_z}(x) \left(d_i(x) - u_i(x+z)\right) dx \tag{6}$$
$$+ \sum_{j=i+1}^{n} \int_\Omega q_{ij}^1(1-d_i) + q_{ij}^2 d_i + q_{ij}^3(1-u_j) + q_{ij}^4 u_j$$
$$+ \alpha_{ij}^1 \left(q_{ij}^1 + q_{ij}^3\right) + \alpha_{ij}^2 \left(q_{ij}^1 + q_{ij}^4\right) + \alpha_{ij}^3 \left(q_{ij}^2 + q_{ij}^3\right) + \alpha_{ij}^4 \left(q_{ij}^2 + q_{ij}^4 - A(i,j)\right) dx \Big\}.$$

Resulting Optimization Problem After carrying out these relaxations we finally obtain the convex energy minimization problem in (6).

The projections onto the respective convex sets of ξ, d, β and α are done by simple clipping while that of the primal variable u is a projection onto the simplex in $\mathbb{R}^n$ [8].

4. Implementation

In order to find the globally optimal solution to this relaxed convex optimization problem, we employ the primal-dual algorithm published in [11]. Optimization is done by alternating a gradient descent with respect to the functions u, d and α and a gradient ascent for the dual variables ξ, β and q interlaced with an over-relaxation step on the primal variables. The step sizes are chosen optimally according to [10]. We stopped the iterations when the average update of the indicator function $u(x)$ per pixel was less than 10^{-5}.

By allowing the primal variables u_i to take on intermediate values between 0 and 1 we may end up with non-binary solutions. In order to obtain a binary solution to the original optimization problem, we assign each pixel x to the label L with maximum value after optimizing the relaxed problem:

$$L(x) = \arg\max_i \{u_i(x)\}, \; x \in \Omega. \tag{10}$$

During our experiments we observed that the computed relaxed solutions u are binary almost everywhere except at region boundaries, with more or less sharp transitions.

5. Experiments and Results

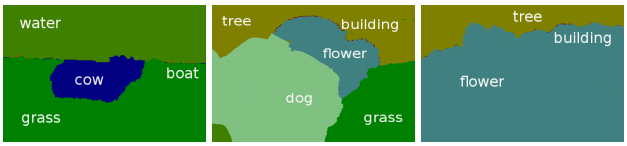
We have developed a novel approach for improving current segmentation algorithms by making use of morphological operations in order to introduce relative spatial relations between object labels. One of the major advantages of the proposed algorithm is that we can utilize structuring elements of different sizes and shapes which allow us to take into account larger neighborhoods of pixels in specific directions. In the following we will show results on the MSRC database and compare our segmentations to state-of-the-art approaches for semantic labeling. The penalty matrix A defined in Section 3.1 is learned from training data based on the relative frequencies of label occurrences within the local range defined by the structuring element. For a symmetric 9×9 element we obtain the penalty matrix in Figure 5.

5.1. Preventing Ghost Labels

'Ghost labels' denote thin artificial regions which are easily introduced if label distances are learned from training data, see for example [14].

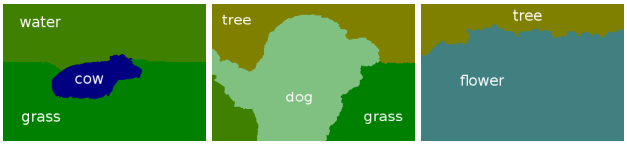
If the distance function does not obey the triangle inequality 'ghost labels' can appear. They reduce costs of direct label transitions by taking a 'detour' over a third, unrelated but less expensive label. Examples are given in Figure 6b) with a closeup in c). The segmentation result obtained by [14] e.g. contains very thin 'boat' regions at the edge of the 'grass' label, because the transition between the labels 'water' and 'boat' and 'boat' and 'grass' is in sum less costly than the direct transition between 'water' and 'grass'. Such 'ghost labels' can be prevented by considering larger neighborhood sizes of e.g. 15×15 pixels as done

a) Input images



b) Segmentation by Strekalovskiy *et al.* [14].

c) Zoom of b) showing ghost labels



d) Segmentations with proximity priors

Figure 6: **Proximity priors avoid ghost labels.** If the transition of two labels is cheaper via a third label artificial labels will be introduced into the segmentation as shown in b) and as closeup in c). The proposed proximity priors consider regions with more than one-pixel distance still as adjacent and thus avoid ghost labels.

by the proposed proximity priors in Figure 6d). The larger the neighborhood size the more expensive becomes the introduction of 'ghost labels'.

5.2. MSRC Segmentation Benchmark

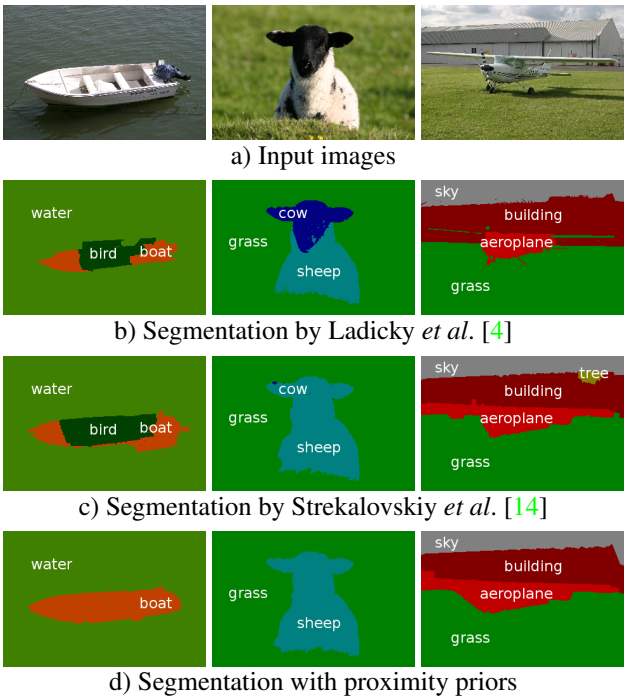
To evaluate the proposed segmentation algorithm we apply it to the task of object segmentation and recognition on the MSRC benchmark. This benchmark comprises 591 images which contain 21 different labels such as 'cow', 'book', 'building' or 'grass'. To conduct experiments on this benchmark, we follow Ladicky *et al.* [4] and divide the image set randomly into 60% training and 40% test images. The proximity matrix A is learned from the training set as described above in Section 3.1. For the experiments we chose a symmetric structuring element of size 9×9 and set λ equal to 0.3.

To evaluate the segmentation accuracy of the proposed method, in Table 1 we compare the benchmark scores of our method to the approaches by Gould *et al.* [2] with relative location priors, Ladicky *et al.* [4] with co-occurrence prior (with and without hierarchical prior), Lucchi *et al.* [7] for the data pairwise global and local models, Vezhnevets *et al.* [15] for the weakly and fully supervised approach and to Strekalovskiy *et al.* [14] with the non-metric distance functions for multi-label problems. The scores denote the average accuracy on the benchmark given as $\frac{\textit{True Positives} \cdot 100}{\textit{True Positives} + \textit{False Negatives}}$ per pixel and per class.

The results indicate that we outperform the other co-occurrence based methods in average class and pixel accuracy. The hierarchical co-occurrence prior by Ladicky *et al.* is added for completeness but not comparable since it uses potentials of order $|\Omega|$ and additional label hierarchy priors. Note that the high score of the approach by Strekalovskiy *et al.* [14] does not reflect the ghost label problem since these regions contain only very few pixels. However, the introduction of entirely unrelated objects, albeit small ones, is often problematic for subsequent applications.

The benchmark results in general suggest rather small improvements for the integration of geometric spatial priors. This is somewhat surprising since the images show strong improvements and the prior corresponds to typical human thinking. As already mentioned by Lucci *et al.* [7] who stated similar findings this is probably due to the rather crude ground truth of the benchmark with large unlabeled regions (see the black regions in Figure 8b)). These regions are not counted in the score, but nevertheless leave a lot of room for misclassification or improvements. Therefore, we think that the benchmark score should not be overstressed here.

Qualitative comparisons with the two best scoring of the above mentioned methods by Ladicky *et al.* [4] with hierarchical prior and by Strekalovskiy *et al.* [14] on the MSRC database are given in Figures 1 and 7. The results show



a) Input images

b) Segmentation by Ladicky *et al.* [4]

c) Segmentation by Strekalovskiy *et al.* [14]

d) Segmentation with proximity priors

Figure 7: **Improved results on the MSRC benchmark.** The results with proximity priors (d) are based on the introduction of geometric spatial label priors. We compare our results to other state-of-the-art methods based on co-occurrence priors.

that the proposed method reduces the number of mislabeled objects. For example, our approach is the only one which correctly detects the boat in Figure 7 without assigning part of it to the label 'bird'. Another example is the head of the sheep in the middle column of Figure 7 which is correctly labeled without any 'cow' pixels. The result of the cat in Figure 1 shows that we can avoid problems which appear due to prior superpixel segmentations.

5.3. Direction Dependent Proximity

Some object pairs only appear in specific spatial constellations, for example cars do not appear above water or books on top of buildings. Such relations can be encoded by applying directional structuring elements, e.g. horizontal or vertical lines. The corresponding penalties can either be defined or learned from training data. For the leftmost image in Figure 8 showing the head of the woman we used a learning based approach for a horizontal structuring element as shown in Figure 4d). We derive the penalty from the relative frequencies of objects appearing up to 40 pixels left and right of the label 'face' in the training images. For the bench and cat images we used a vertical structuring element directed up-/downwards (compare Figure 4c)) to penalize the label 'bird' below 'chair' and the label 'water' above 'street'.

	Avg. accuracy per pixel	Avg. accuracy per class	Building	Grass	Tree	Cow	Sheep	Sky	Aeroplane	Water	Face	Car	Bicycle	Flower	Sign	Bird	Book	Chair	Road	Cat	Dog	Body	Boat
Gould *et al*. [2] CRF + rel. loc.	76.5	64.38	72	95	81	66	71	93	74	70	70	69	72	68	55	23	82	40	77	60	50	50	14
Ladicky *et al*. [4] co-oc.	80	67.76	**77**	96	80	69	82	**98**	69	82	79	75	75	81	**85**	35	76	17	**89**	25	**61**	50	22
Lucchi *et al*. [7], DPG local	75	68.62	54	88	83	79	82	95	87	70	85	81	**97**	69	72	27	88	46	60	74	27	49	28
Lucchi *et al*. [7], DPG loc.+glob.	80	74.62	65	87	87	84	75	93	**94**	78	83	72	93	86	70	50	**93**	**80**	86	78	28	58	27
Vezhnevets *et al*. [15], weak sup.	67	66.52	12	83	70	81	**93**	84	91	55	**97**	87	92	82	69	**51**	61	59	66	53	44	9	**58**
Vezhnevets *et al*. [15], full sup.	72	71.71	21	93	77	86	**93**	96	92	61	79	**89**	89	**89**	68	50	74	54	76	68	47	49	55
Strekalovskiy *et al*. [14]	84.85	77.52	70	**97**	**92**	**89**	85	96	81	**83**	90	82	92	83	66	45	92	63	86	80	51	73	32
Proposed Proxmity Priors	**84.97**	**78.19**	69	**97**	**92**	87	87	97	87	82	91	83	94	84	62	44	**93**	67	86	**83**	57	**74**	26
Ladicky *et al*. [4] co-oc. + hier.	87	76.76	82	95	88	73	88	100	83	92	88	87	88	96	96	27	85	37	93	49	80	65	20

Table 1: **MSRC benchmark results.** We compare the segmentation accuracy to state-of-the-art segmentation algorithms with co-occurrence priors on the MSRC benchmark. The proposed proximity priors outperform the other methods with respect to average accuracy per pixel and per class. The approach by Ladicky *et al*. in the last row is added for completeness but is not comparable since it includes hierarchical label priors and uses potentials of the highest order $|\Omega|$ instead of order two as in our approach.

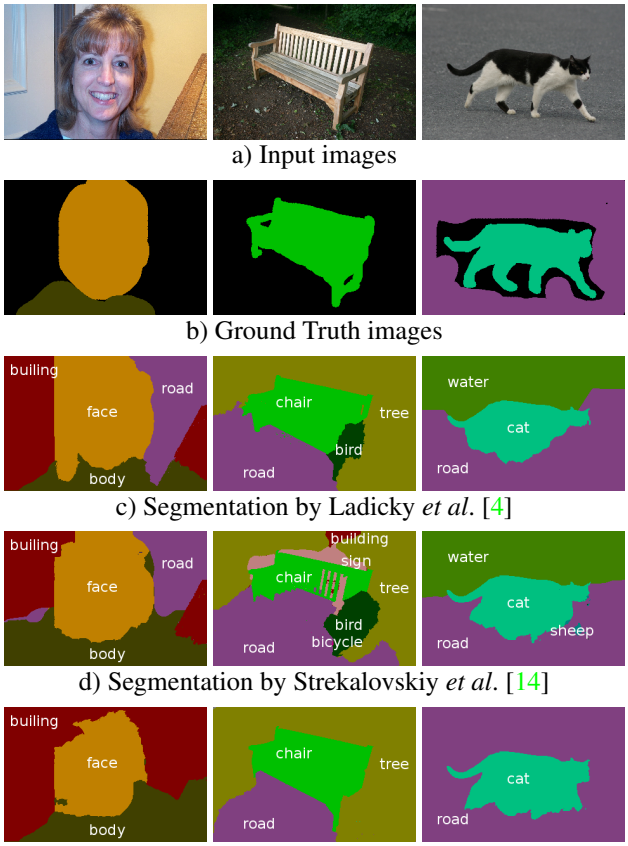


Figure 8: **Directional proximity priors.** The introduction of specific natural spatial relationships such as: 'bird' does not appear below 'chair' or 'water' does not appear above 'street' allows for improved segmentations of difficult images.

The results show strong improvements compared to previous approaches and thus show that proximity priors facilitate the segmentation of complex images.

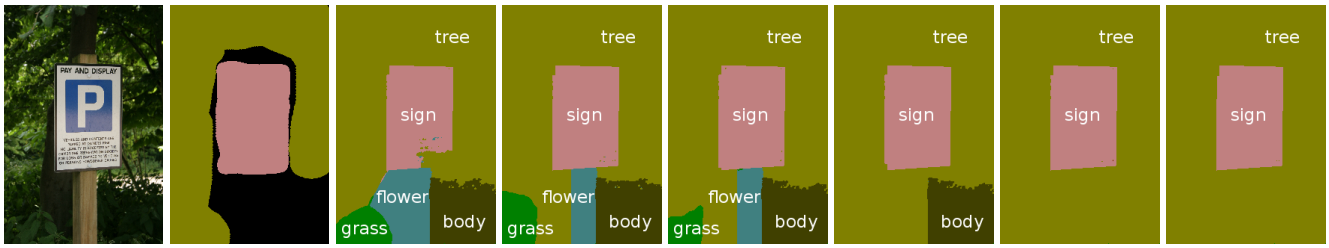
5.4. Runtimes

Due to the inherent parallel structure of the optimization algorithm [11] the approach can be easily parallelized and implemented on graphics hardware. We used a parallel CUDA implementation on an NVIDIA GTX 680 GPU.

Since the number of terms in the optimization functional depends on the size of the structuring element $\mathcal{S}$, the runtime increases for larger elements. To make the computation feasible we randomly selected only very few entries in the structuring element and neglected the others. Figure 9 shows the development of the solution for an increasing number of entries together with the corresponding runtime for a structuring element of size 15×15. We can conclude that already very sparse elements containing around ten entries yield results very similar to full structuring elements. The average runtime on the benchmark for a sparse element of 15×15 containing 15 entries is about three minutes. For comparison, the approach by Gould *et al*. [2] takes 31 seconds and the approach by Ladicky *et al*. [4] 16 seconds on average on the MSRC benchmark. Yet, these approaches require a prior partitioning of the image into a few hundred superpixels (200 to 400). Instead, our computations are carried out directly on the pixel level with image sizes of 320×240 pixels and thus tens of thousands of pixels.

6. Conclusion

We introduced proximity priors for semantic segmentation using morphological operations in a variational multilabel approach. Instead of introducing co-occurrence



Original	Ground Truth	$\lvert\mathcal{S}\rvert = 0$	$\lvert\mathcal{S}\rvert = 2$	$\lvert\mathcal{S}\rvert = 5$	$\lvert\mathcal{S}\rvert = 7$	$\lvert\mathcal{S}\rvert = 10$	$\lvert\mathcal{S}\rvert = 225$
		0.21 min	2.41 min	2.54 min	2.72 min	2.94 min	15.24 min

Figure 9: **Minimizing runtime.** To minimize runtime we use sparse structuring elements (SE). The development of the solution for an increasing number of entries in a structuring element $\mathcal{S}$ of size 15×15 shows that very few entries (here 10 entries in a 15×15 SE) are already sufficient to obtain accurate results. The runtimes denote the average runtime on the benchmark with the respective number of entries $|\mathcal{S}|$.

probabilities of label combinations proximity priors allow for the introduction of geometric spatial relationships of label pairs, i.e. direction and distance specific label pair probabilities. This prior is relative since it describes the spatial relationship between two labels (e.g. 'sky' appears above 'grass') in contrast to commonly used absolute spatial priors which impose knowledge on where a single label usually appears in an image (e.g. 'sky' appears in the upper half of the image). The specific spatial relationships are determined by the size and shape of the structuring element. Based on this geometric spatial prior both the computation of superpixels and the emergence of one pixel wide 'ghost labels' can be prevented. Furthermore, the label cost penalty is proportional to the size of the labeled regions and also affects object labels at larger spatial distances. We proposed a multilabel variational approach where neighborhood information is determined using morphological *dilation*. Subsequently, we derived an exact convex relaxation which can be solved optimally. The method outperforms previous co-occurrence based approaches on the MSRC benchmark.

References

[1] A. Delong and Y. Boykov. Globally optimal segmentation of multi-region objects. In *IEEE Int. Conf. on Computer Vision*, Kyoto, 2009. 2

[2] S. Gould, J. Rodgers, D. Cohen, G. Elidan, and D. Koller. Multi-class segmentation with relative location prior. *Int. J. of Computer Vision*, 2008. 2, 6, 7

[3] P. Kohli, L. Ladicky, and P. H. S. Torr. Robust higher order potentials for enforcing label consistency. *Int. J. of Computer Vision*, 82(3):302–324, 2009. 2

[4] L. Ladicky, C. Russell, P. Kohli, and P. Torr. Graph cut based inference with co-occurrence statistics. In *Proceedings of ECCV*, 2010. 1, 2, 3, 6, 7

[5] Y. Leclerc. Region growing using the MDL principle. In *Proc. DARPA Image Underst. Workshop*, pages 720–726, April 6-8 1990. 2

[6] X. Liu, O. Veksler, and J. Samarabandu. Order-preserving moves for graph-cut-based optimization. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 32(7):1182–1196, 2010. 2

[7] A. Lucchi, Y. Li, X. Boix, K. Smith, and P. Fua. Are spatial and global constraints really necessary for segmentation? In *IEEE Int. Conf. on Computer Vision*, 2011. 6, 7

[8] C. Michelot. A finite algorithm for finding the projection of a point onto the canonical simplex of $\mathbb{R}^n$. *Journal of Optimization Theory and Applications*, 50(1):195–200, July 1986. 5

[9] S. Nowozin and C. Lampert. Global connectivity potentials for random field models. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 2009. 2

[10] T. Pock and A. Chambolle. Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In *IEEE Int. Conf. on Computer Vision*, 2011. 5

[11] T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the Mumford-Shah functional. In *IEEE Int. Conf. on Computer Vision*, Kyoto, 2009. 2, 5, 7

[12] E. Strekalovskiy and D. Cremers. Generalized ordering constraints for multilabel optimization. In *IEEE Int. Conf. on Computer Vision*, 2011. 2

[13] E. Strekalovskiy, B. Goldluecke, and D. Cremers. Tight convex relaxations for vector-valued labeling problems. In *IEEE Int. Conf. on Computer Vision*, 2011. 4

[14] E. Strekalovskiy, C. Nieuwenhuis, and D. Cremers. Nonmetric priors for continuous multilabel optimization. In *European Conference on Computer Vision*, 2012. 1, 2, 5, 6, 7

[15] A. Vezhnevets, V. Ferrari, and J. M. Buhmann. Weakly supervised semantic segmentation with a multi-image model. In *IEEE Int. Conf. on Computer Vision*, 2011. 6, 7

[16] S. Vicente, V. Kolmogorov, and C. Rother. Graph cut based image segmentation with connectivity priors. In *Int. Conf. on Computer Vision and Pattern Recognition*, Anchorage, Alaska, June 2008. 2

[17] J. Yuan and Y. Boykov. TV-based multi-label image segmentation with label cost prior. In *British Machine Vision Conference*, pages 1–12, 2010. 2

[18] S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 18(9):884–900, 1996. 2