A Co-occurrence Prior
for Continuous Multi-Label Optimization

Mohamed Souiai, Evgeny Strekalovskiy, Claudia Nieuwenhuis and Daniel Cremers

Technical University Munich, Germany

Abstract. To obtain high-quality segmentation results the integration of semantic information is indispensable. In contrast to existing segmentation methods which use a spatial regularizer, i.e. a local interaction between image points, the co-occurrence prior [15] imposes penalties on the co-existence of different labels in a segmentation. We propose a continuous domain formulation of this prior, using a convex relaxation multi-labeling approach. While the discrete approach [15] is employs minimization by sequential alpha expansions, our continuous convex formulation is solved by efficient primal-dual algorithms, which are highly parallelizable on the GPU. Also, our framework allows isotropic regularizers which do not exhibit grid bias. Experimental results on the MSRC benchmark confirm that the use of co-occurrence priors leads to drastic improvements in segmentation compared to the classical Potts model formulation when applied .

1 Introduction

1.1 Semantic Image Labeling

While traditional image segmentation algorithms have focused on separating regions based on homogeneity of color or texture, more recent methods have aimed at incorporating *semantic* knowledge into what is often called class-based image segmentation. Rather than simply grouping regions of similar color, the goal is to assign to each pixel of an image a semantic label such as "grass", "sky", "cow" or "horse", each of which does not necessarily share the same color model – horses may be white, brown or black for example. Such approaches allow to impose prior knowledge about which pairs of labels are likely to co-occur in a given image [15]. Figure 1 shows semantic labelings computed for the image of a cow on grass with and without a co-occurrence prior: While the color likelihood based data term has a slight preference for cat over cow, the co-occurrence additionally imposes the information that cows are more commonly observed on grass next to the ocean than cats.

A separate line of work has promoted the use of minimum description length (MDL) priors [16, 27, 5, 25] which impose a prior that favors a smaller number of labels in the final segmentation. In practice, the advantage of such MDL priors is that one can preserve a level of regularity while reducing the often over-smoothing boundary length regularization. While many experiments demonstrating the advantage of co-occurrence can often be reproduced with a simple MDL prior (that suppresses the emergence of undesired labels), for the example in Figure 1 co-occurrence is vital since the number of labels is in both cases the same.
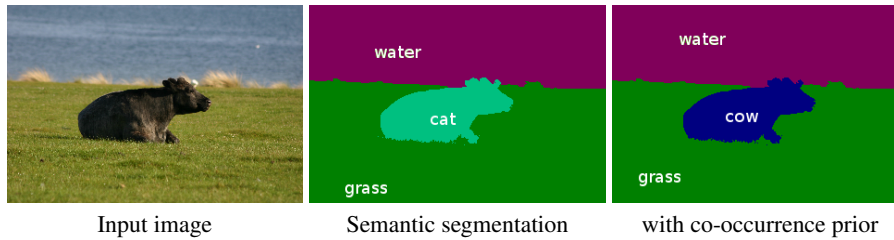
Fig. 1. We propose a convex relaxation for co-occurrence priors in spatially continuous semantic segmentation. Whereas purely data-driven semantic segmentation (middle) assigns the label 'cat' to the cow, co-occurrence priors (right) substantially improve the performance by imposing the knowledge that cows are more commonly encountered next to grass and ocean than cats.

Co-occurrence priors have been applied to the finite-dimensional discrete setting by means of combinatorial problems on a grid [15]. These problems can be cast as large scale integer linear programs which are solved approximatively using α-expansion [1] or the fast primal dual algorithm [14]. However, being defined on the grid these methods are inherently anisotropic. In addition the discrete methods used for solving such problems are sequential and thus only partially parallelizable [13]. In contrast, continuous formulations allow for isotropic regularizers and appropriate discretizations of such functionals do not show grid artefacts. Furthermore, the primal dual algorithm [4] used for solving the continuous saddle-point problem is defined point-wise and can thus be parallelized in a straight-forward manner and run in parallel using modern GPU's or other parallel architectures. For a detailed discussion see [19, 11].

Thus, a major challenge addressed in this paper is how to efficiently integrate co-occurrence priors into a convex continuous optimization approach, which allows for fast solutions independent of the initialization of the algorithm.

In contrast to common segmentation methods we refrain from using super-pixels [18, 7, 22, 15]. Super-pixels prevent pixels with similar colors from being assigned to different labels. As a consequence, elongated structures may be lost or larger chunks may be incorrectly assigned in the final solution – see the head of the sheep in Figure 2 which is assigned to the label 'cow'. To preserve elongated structures we use the non-local total variation formulation [24, 8].

1.2 Related work

The inspiration to this work predominantly draws from two lines of research, namely research on label configuration priors and research on convex relaxation techniques. On the one hand, there are a number of recent advances on label configuration energies for semantic image labeling, including the co-occurrence priors [15], MDL priors [5, 27, 25, 16], and hierarchical label cost priors [5].

On the other hand, there are a number of recent advances on convex relaxation techniques for spatially continuous multi-label optimization. These include relaxations for the continuous Potts model [2, 3, 17, 26], for the non-local continuous Potts model [24], for MDL priors [25], and for vector-valued labeling problems [9, 23].

| Input | no co-occ. | Ladicky et al. [15] | proposed |
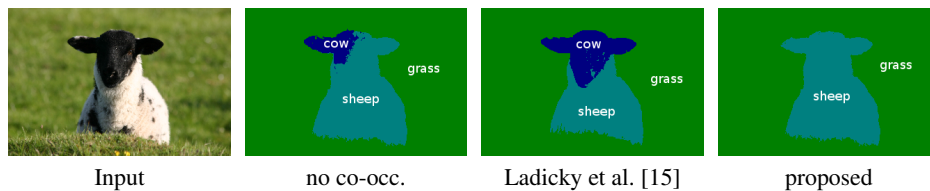
Fig. 2. Methods using super-pixel information for improving label consistency can produce mislabeling in the resulting segmentation as can be seen in the case of the approach in [15] which builds on the higher order CRF approach introduces by Kohli et al. [12].

1.3   Contributions

Our contributions are the following:

- We formulate the co-occurrence priors within a spatially continuous approach to semantic multi-label segmentation.

- We propose a convex relaxation of the co-occurrence based segmentation functional which can be solved optimally. This approach yields results independent of intialization and is - in contrast to discrete methods - straighfowardly parallelizable.

- The proposed integration of co-occurrence priors is done on a pixel level and therefore avoids the commonly used pre-segmentation into super-pixels.

2   Convex Multi-label Segmentation

Given a discrete label space $\mathcal{G} = \{1, ..., n\}$ with $n \geq 3$, the multi-labeling problem can be stated as a minimal partition problem. The image domain $\Omega \subset \mathbb{R}^2$ is to be segmented into n pairwise disjoint regions Ω_i which are encoded by the label indicator function $u \in BV(\Omega, \{0,1\})^n$

$$u_i(x) = \begin{cases} 1 & \text{if } x \in \Omega_i, \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

Here BV denotes the space of functions u for which the total variation

$$TV(u) := \sup_{\substack{p \in C_c^1(\Omega; \mathbb{R}^2) \\ |p(x)| \leq 1 \, \forall x}} \int_\Omega u(x) \, \text{div} \, p(x) \, dx \tag{2.2}$$

is bounded allowing for discontinuities. To ensure that each pixel is assigned to exactly one region a point-wise simplex constraint is imposed on u:

$$\sum_{i=1}^n u_i(x) = 1 \quad \forall x \in \Omega. \tag{2.3}$$

To find a solution to the minimal partition problem we minimize the general energy E which can be decomposed as follows:

$$E(u) = E_D(u) + E_S(u) + E_C(u), \tag{2.4}$$

The term E_D is called the data term, the expression E_S represents a regularization term usually requiring smoothness of the solution, and the term E_C is the global co-occurrence energy which will be introduced in Section 3.

The data term $E_D(u)$ assigns a cost $\varrho_i(x) : \Omega \to \mathbb{R}$ to each pixel x for belonging to region i (based on its color or texture). It can be written in terms of the indicator functions as

$$E_D(u) = \sum_{i=1}^{n} \int_{\Omega} u_i(x)\varrho_i(x)dx. \tag{2.5}$$

The regularization term $E_S(u)$ imposes a spatial smoothness which can be formulated by means of the Potts model.

The classical total variation based formulation of the Potts model, $\frac{1}{2} \sum_{i=1}^{n} TV(u_i)$, minimizes the length of the interface of each region which leads to an over-smoothing in images exhibiting objects with fine or elongated structures. To improve over this, Werlberger et al. [24] proposed a non-local variant of the Potts model, which improves the labeling quality on the boundaries. The key idea is that pixels x and y are likely to share the same label if they are spatially close *and* have a similar color. For each pair of pixels a weight w is defined, which measures this similarity:

$$w(x,y) = \exp\left[-\left(\frac{d_c(x,y)}{\alpha} + \frac{d_s(x,y)}{\beta} \right) \right]. \tag{2.6}$$

Here d_c and d_s denote the color and spatial distance scaled by the parameters α and β. The regularizer, finally, measures the weighted label differences for each pixel compared to its spatial neighborhood $\mathcal{N}_x$

$$E_S(u) = \sum_{i=1}^{n} \int_{\Omega} \left(\int_{\mathcal{N}_x} w(x,y) \, |u_i(y) - u_i(x)| \, dy \right) dx. \tag{2.7}$$

Introducing a dual variable p transforms the non-differentiable expression (2.7) to a fully differentiable one (for each fixed p):

$$E_S(u) = \sup_{p \in \tilde{\mathcal{K}}} \sum_{i=1}^{n} \int_{\Omega} \left(\int_{\mathcal{N}_x} p_i(x,y) \, (u_i(y) - u_i(x)) \, dy \right) dx \tag{2.8}$$

with the convex constraint set $\tilde{\mathcal{K}}$

$$\tilde{\mathcal{K}} := \left\{ p(x,y) \in C^1 \left(\Omega \times \Omega, \mathbb{R} \right)^n \big| \, |p_i(x,y)| \le w(x,y) \right\}. \tag{2.9}$$

Thus E_D and E_S above are both convex in u. Note that the above non-local regularizer (2.7), and with it also our energy, is not isotropic. One could also easily use its isotropic

version, where the L^1-norm in (2.7) is replaced by $\sqrt{\int_{\mathcal{N}_x} w(x,y)(u_i(y) - u_i(x))^2 \, dy}$ as in [8]. However, the results with both versions are almost the same, so that we have chosen the form (2.7) for computational efficiency. In general, our continuous formulation easily allows to incorporate isotropic regularizers, such as the isotropic Potts model relaxations [2, 3]. Additionally, in contrast to graph based methods there is no need for complex graph constructions when solving our optimization problem.

In order to obtain a convex optimization problem, the optimization domain must be convex as well. Therefore, we relax the binary constraints of the indicator function u and — together with (2.3) — obtain the convex set

$$\mathcal{S} := \left\{ u \in BV(\Omega, [0,1])^n \ \Big| \ \sum_{i=1}^n u_i(x) = 1 \quad \forall \, x \in \Omega \right\}. \tag{2.10}$$

After optimization of the relaxed problem the final pixel labeling $\mathcal{L} : \Omega \to \mathcal{G}$ can be recovered from the relaxed solution u^* by:

$$\mathcal{L}(x) = \arg \max_{1 \leq i \leq n} u_i^*(x) \tag{2.11}$$

3 A Continuous Co-occurrence Prior

In this section we introduce the global co-occurrence energy E_C in (2.4), which allows for the integration of semantic scene knowledge, such as for example that sheep and grass often appear together in the same image whereas sheep and wolves are rarely to be found. For each subset of labels $L \subseteq \mathcal{G}$ a specific penalty can be defined or learned from training data. Note that this penalty only depends on the simultaneous occurrence of specific labels in the image, not on their location or the size of their corresponding regions in the segmentation.

3.1 A Convex Formulation of Label Occurrences

In order to devise the co-occurrence prior it is necessary to model the occurrences of specific labels in the image. To this end, we introduce the label indicator function $l : \mathcal{S} \to \{0, 1\}^n$,

$$l_i(u) = \begin{cases} 1, & \text{if } \exists x \in \Omega : u_i(x) = 1, \\ 0, & \text{otherwise,} \end{cases} \tag{3.1}$$

which indicates for each label $i \in \mathcal{G}$ if it appears in the segmentation given by $u \in \mathcal{S}$. To obtain a convex formulation of the function l, we use the following relation which was already used by Yuan et al. [25]:

$$l_i(u) = \max_{x \in \Omega} u_i(x) \qquad \forall i \in \mathcal{G}. \tag{3.2}$$

where we use $\max$, instead of the formally correct $\mathrm{ess \, sup}$ operator, for readability. Note that the L^∞-norm on the left hand side of (3.2) couples the indicator functions of *all* pixels in the image domain and, thus, represents the key ingredient for the introduction

of the *global* co-occurrence prior. Due to the non-differentiability of the L^∞-norm and the non-convex range $\{0, 1\}$ of each label indicator function l_i, we relax l to map to the unit intervals $l : \mathcal{S} \to [0, 1]^n$ and replace (3.2) by the following convex constraint:

$$l_i(u) \geq u_i(x) \quad \forall x \in \Omega \quad \forall i \in \mathcal{G}. \tag{3.3}$$

The idea of defining a co-occurrence energy is that it should only depend on the label indicator functions l, which capture the global occurrence information:

$$E_C(u) = E_C(l(u)). \tag{3.4}$$

Provided that E_C has the following properties:

1. $E_C(l)$ is convex.
2. $E_C(l)$ is monotonically increasing w.r.t. l i.e.

$$l \preceq \tilde{l} \implies E_C(l) \leq E_C(\tilde{l}), \tag{3.5}$$

it can be easily shown that, replacing (3.2) by (3.3) and minimizing over l, one recovers the optimum of the optimization problem (2.4) with respect to constraint (3.2).

3.2 A Convex Formulation of the Co-occurrence Prior

We will now formulate the continuous *co*-occurrence prior. To this end, the concept of occurrence functions l in (3.2) for single labels is generalized to the occurrence of label *subsets* $L \subseteq \mathcal{G}$ by introducing the label subset indicator function $\delta_L : \mathcal{S} \to \{0, 1\}$:

$$\delta_L(u) = \begin{cases} 1 & \text{if } l_i(u) = 1 \ \forall i \in L, \\ 0 & \text{otherwise.} \end{cases} \tag{3.6}$$

The function δ_L indicates the simultaneous occurrence of *all* labels in the subset L in the image, so that it can be rewritten in the following way:

$$\delta_L(u) = \prod_{i \in L} l_i(u). \tag{3.7}$$

The co-occurrence prior is then defined as the sum over all possible label combinations of $\mathcal{G}$ (the elements of the power set of $\mathcal{G}$) weighted by the associated co-occurrence penalties $C(L) \geq 0$

$$E_C(l(u)) = \sum_{L \in \mathcal{P}(\mathcal{G})} \delta_L(u) \cdot C(L) = \sum_{L \in \mathcal{P}(\mathcal{G})} C(L) \prod_{i \in L} l_i(u). \tag{3.8}$$

Because of the product in (3.7), the term δ_L is not convex in terms of l for $|L| \geq 2$. Thus, a convex relaxation of the product is required to make the energy term convex. We relax the energy (3.8) term-wise for each subset L, i.e. by relaxing each individual addend

$$E_L(u) := C(L) \prod_{i \in L} l_i(u). \tag{3.9}$$

A convex formulation of this kind of energies was given in [23]. Let us briefly recall this approach for the convenience of the reader. The general considered energy is

$$E_0(v) = \sum_{\gamma_1 \in \Lambda_1, \ldots, \gamma_d \in \Lambda_d} c_\gamma \, v_{\gamma_1}^1 \cdot \ldots \cdot v_{\gamma_d}^d \tag{3.10}$$

with $d \geq 1$ finite label sets $\Lambda_1, \ldots, \Lambda_d$ and, for each $1 \leq i \leq d$, corresponding indicator variables $(v_\gamma^i)_{\gamma \in \Lambda_i} \in [0, 1]$ which satisfy the simplex constraint $\sum_{\gamma \in \Lambda_i} v_\gamma^i = 1$. The costs $c_{\gamma_1, \ldots, \gamma_d} \in \mathbb{R}$ can be arbitrary. The convex relaxation proposed in [23] is

$$E_0^{\text{rel}}(v) = \sup_{q \in \mathcal{Q}} \sum_{\gamma_1 \in \Lambda_1} q_{\gamma_1}^1 v_{\gamma_1}^1 + \ldots + \sum_{\gamma_d \in \Lambda_d} q_{\gamma_d}^d v_{\gamma_d}^d \tag{3.11}$$

with the convex set

$$\mathcal{Q} = \left\{ (q_{\gamma_i}^i)_{1 \leq i \leq d, \, \gamma_i \in \Lambda_i} \, \big| \, q_{\gamma_i}^i \in \mathbb{R}, \ q_{\gamma_1}^1 + \ldots + q_{\gamma_d}^d \leq c_\gamma \ \forall \gamma_1 \in \Lambda_1, .., \gamma_d \in \Lambda_d \right\}. \tag{3.12}$$

The convex energy E_0^{rel} is the *tightest possible* relaxation of E_0: In the recent journal version [10] of [23] it is shown that E_0^{rel} is the convex hull of E_0. For instance, this means that E_0^{rel} preserves the minimizers of E_0, i.e. minimizers of E_0 are also minimizers of E_0^{rel}, Furthermore, $E_0^{\text{rel}}(v)$ coincides with $E_0(v)$ for binary v ($v_{\gamma_i}^i \in \{0, 1\}$ for all i and $\gamma_i \in \Lambda_i$).

In our case, we have $d = |L|$ and only two labels per factor, i.e. $\Lambda_i = \{0, 1\}$ for all $1 \leq i \leq d$. The corresponding indicator variables are $v_0^i = 1 - l_i$ and $v_1^i = l_i$ for each i. Finally, the costs c_γ are given by

$$c_\gamma = \begin{cases} C(L), & \text{if } \gamma_1 = \ldots = \gamma_d = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{3.13}$$

Directly applying (3.11), and writing the dual variables as $q_0^i =: \varphi_L^i$ and $q_1^i =: \psi_L^i$ for each $1 \leq i \leq |L|$, we obtain the following convex formulation of E_C:

$$E_C(l(u)) = \sum_{L \in \mathcal{P}(\mathcal{G})} \left(\sup_{(\varphi_L, \psi_L) \in \mathcal{Q}_L} \sum_{i \in L} (1 - l_i(u))\varphi_L^i + l_i(u)\psi_L^i \right) \tag{3.14}$$

with the convex constraint set

$$\mathcal{Q}_L := \left\{ (\varphi_L, \psi_L) \big| \ \forall z \in \{0, 1\}^{|L|} \neq \mathbf{1} : \right.$$
$$\left. \sum_{i \in L} (1 - z_i)\varphi_L^i + z_i \psi_L^i \leq 0, \quad \sum_{i \in L} \psi_L^i \leq C(L) \right\}, \tag{3.15}$$

where $\mathbf{1}$ is a vector consisting of all ones. The terms $(1 - z_i)\varphi_L^i + z_i \psi_L^i$ in (3.15) arise from (3.12) by noting that $q_{z_i}^i = (1 - z_i)\varphi_L^i + z_i \psi_L^i$ for all i and $z_i \in \{0, 1\}$.

For $C(L) \to \infty$ in (3.8) for some subset L we obtain *hard constraints* on label configurations, i.e. $\delta_L(u) = 0$. Including this constraint prohibits segmentations containing

all labels of L: The energy will be infinite if the labels from L occur in the image simultaneously. This corresponds to simply dropping the constraint $\sum_{i \in L} \psi_L^i \leq C(L)$ in the above constraint set (3.15).

We can easily prove that $E(l(u))$ is monotonically increasing, i.e. that the requirement (2.5) is fulfilled. First, it is monotonic for binary l: It this case it is represented by the original formula (3.8). Since the label subset indicator functions $\delta_L(l)$ are clearly monotonous in l and the costs $C(L)$ are nonnegative, it follows that E_C is monotonous for binary l. This is also referred to as the principle of Occam's razor which states that among competing labelings the one with fewer labels should be favoured energetically. Second, we use the fact that (3.14) is the *convex hull* relaxation of $\delta_L(l)$ to the set of possibly non-binary l's. This and the monotonicity for binary l yields the monotonicy of $E_C(l)$ for general $l \in [0,1]^n$.

It follows that by means of the convex relaxation (3.3) we can recover the constraint (3.2) and thus globally minimize the overall energy (2.4). Since the power set of $\mathcal{G}$ is very large, we follow [15] and approximate the true costs $C(L)$ by taking only sets of two labels into account, with fixed costs which best approximate the original costs. For details see [15]. The resulting co-occurrence energy for sets of only two labels then reads as:

$$E_C(l(u)) = \sum_{1 \leq i < j \leq n} \Bigg(\sup_{\{\varphi_{ij}, \psi_{ij}\} \in \mathcal{Q}_{i,j}} (1 - l_i(u))\varphi_{ij,1} l_i(u)\psi_{ij,1}$$
$$+ (1 - l_j(u))\varphi_{ij,2} + l_j(u)\psi_{ij,2} \Bigg), \qquad (3.16)$$

where $\mathcal{Q}_{ij}$ is the convex constraint set of dual variables $\varphi_{ij}, \psi_{ij} \in \mathbb{R}^2$:

$$\mathcal{Q}_{ij} := \Bigg\{ (\varphi_{ij}, \psi_{ij}) \Big| \quad \psi_{ij,1} + \psi_{ij,2} \leq c_{ij}, \qquad (3.17)$$
$$\varphi_{ij,1} + \varphi_{ij,2} \leq 0, \ \varphi_{ij,1} + \psi_{ij,2} \leq 0, \ \psi_{ij,1} + \varphi_{ij,2} \leq 0 \Bigg\}.$$

Note that in the integer linear program (ILP) given in [15] the label subset indicator function $\delta_L(u)$ is realized by the following constraint ([15] equation (31)):

$$\delta_L(u) \geq \sum_{i \in L} l_i(u) - |L| + 1. \qquad (3.18)$$

In contrast to our framework with the tight convex relaxations (3.11) and (3.16), equation (3.18) would introduces additional trivial solutions after relaxing the ILP to allow $\delta_L(u)$ and $l_i(u)$ to be from $[0,1]$. For instance, (3.18) is always fulfilled for $\delta_L(u) = 0$ and any l such that $\sum_{i \in L} l_i \leq |L| - 1$, e.g. $l_i = 1/(|L| - 1)$. Therefore, our continuous formulation is a tighter relaxation than the linear relaxation given in [15].

4 Implementation

The overall saddle-point formulation of our optimization problem for label subsets containing only two labels can be summarized as follows:

$$\min_{\substack{u \in \mathcal{S} \\ l \in [0,1]^n}} \sup_{\substack{p \in \tilde{\mathcal{K}} \\ \varphi_{ij}, \psi_{ij} \in \mathcal{Q}_{i,j}}} \sum_{i=1}^{n} \int_{\Omega} \left(\int_{\mathcal{N}_x} p(x,y)(u_i(y) - u_i(x)) \, dy \right) dx + \sum_{i=1}^{n} \int_{\Omega} u_i(x) \varrho_i(x) dx +$$

$$\sum_{1 \leq i < j \leq n} \left((1 - l_i)\varphi_{ij,1} + l_i\psi_{ij,1} + (1 - l_j)\varphi_{ij,2} + l_j\psi_{ij,2} \right)$$

$$\text{s.t.} \quad l_i \geq u_i(x) \quad \forall x \in \Omega \quad \forall i \in \mathcal{G}. \tag{4.1}$$

In order to solve the above saddle-point problem we use the first order primal dual algorithm [21, 4] which is essentially a gradient descent in the primal variables and a gradient ascent in the dual variables with a subsequent computation of the respective proximity operators and an over-relaxation step for the primal variables. For the time steps we use recent preconditioning techniques introduced in [20]. This way there is no need to compute the Lipschitz constant of the underlying linear operator. The details of the implementation can be found in the appendix.

5   Experiments

The key contribution of this paper is the introduction of a co-occurrence prior into the continuous multi-label framework. In this section we evaluate the proposed continuous formulation of the segmentation problem with the co-occurrence prior on the MSRC database [6]. To preserve comparability to [15], we use their data term, which is based on texture boosting. As explained in Section 4 we approximate the exact co-occurrences by considering only binary label interactions. Overall we obtain $\binom{n}{2}$ possible binary label interactions, i.e. 210 for the 21 labels considered in the MSRC dataset. For each interaction we need to keep track of only eight scalar variables (dual variables and Lagrange multipliers), which is almost negligible compared to the remaining part of the optimization approach. Note that we obtain the co-occurrence weights c_{ij} from the training data as in [15].

   Some of the obtained segmentation results of the MSRC dataset are shown in Figure 3. A comparison of the segmentation accuracy of the proposed algorithm to the original formulation by Ladicky et al. on the whole benchmark can be found in Figure 4. The table indicates the numbers for each label separately as well as the average on the whole benchmark. The results show that the segmentations obtained with the proposed continuous formulation are comparable to those obtained by Ladicky et al. in terms of global accuracy (which neglects the number of images per label), and even outperforms them in terms of average accuracy over all benchmark images. Note that due to different training and evaluation sets of images the numbers in table 4 differ slightly from the numbers in [15].

   In order to evaluate the tightness of the optimization problem we computed the following relative optimality bound, which is an upper bound of the energy difference between the globally optimal binary solution and the computed binarized solution

$$B(u, \tilde{u}) = \frac{E(\tilde{u}) - E(u)}{E(u)}. \tag{5.1}$$

a) Input   b) Potts   c) Ladicky et al. [15]   d) Configuration Prior
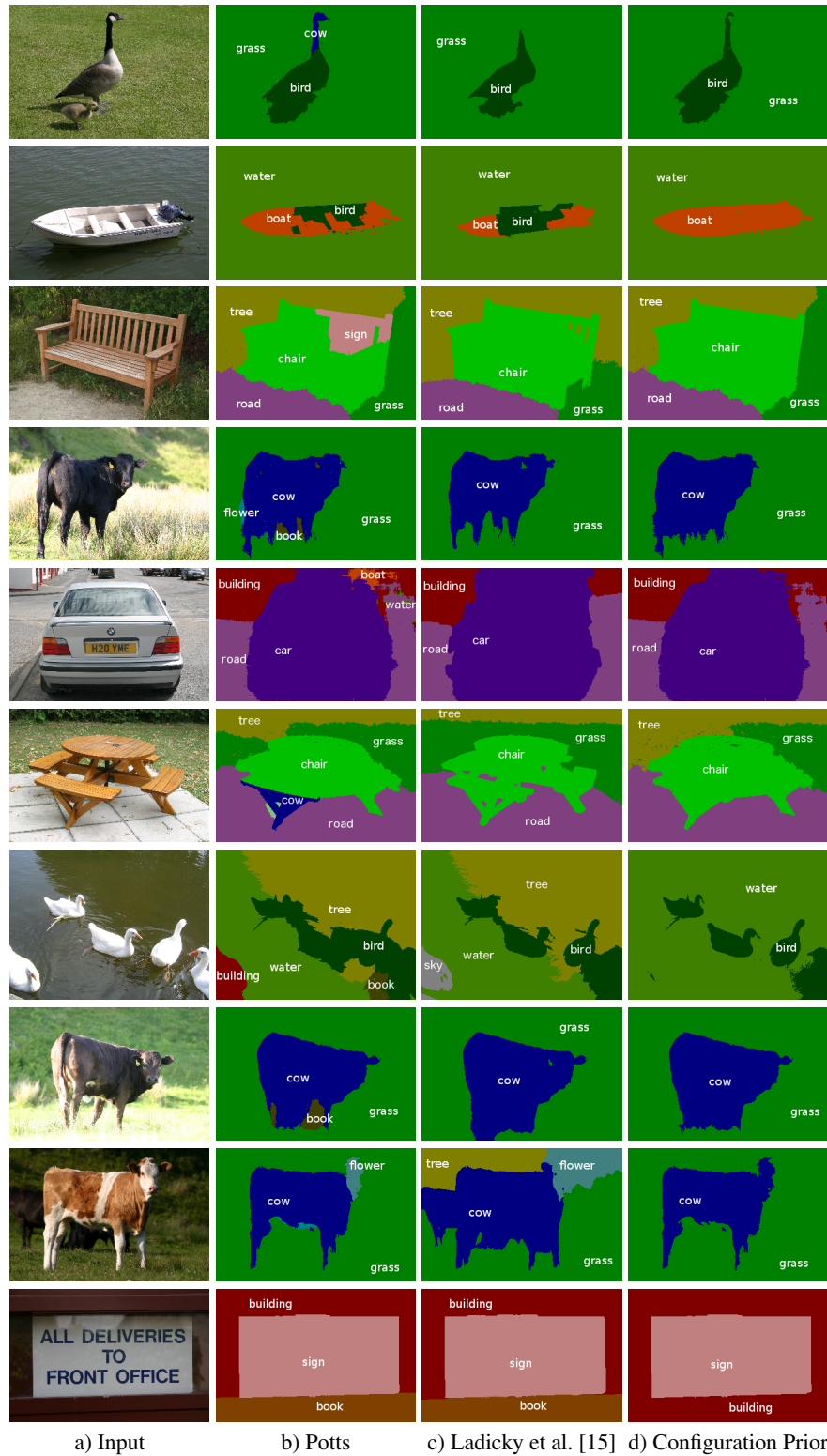
Fig. 3. Co-occurrence prior: Qualitative results on images taken from the MSRC database. The results show (a) the original benchmark image, (b) the segmentation result without co-occurrence prior, (c) the results by Ladicky et al. [15], (d) results from our continuous formulation of the segmentation algorithm incorporating the co-occurrence energy.

Here E is the energy given in (2.4), u the solution of the relaxed saddle point problem (4.1) and $\tilde{u}$ the binarized solution of u used for computing the final labeling, see (2.11). We obtained an average value of $B(u, \tilde{u}) = 0.17\%$ for the optimality bound, which means that our computed binary solutions are very close to the global optimum of the original optimization problem.

	Global	Average	Building	Grass	Tree	Cow	Sheep	Sky	Aeroplane	Water	Face	Car	Bicycle	Flower	Sign	Bird	Book	Chair	Road	Cat	Dog	Body	Boat
dataterm	83.99	77.18	67	97	91	85	86	95	88	81	90	82	94	81	62	42	91	66	86	79	54	72	31
Potts	84.80	77.88	69	97	91	86	86	96	86	82	90	81	93	83	62	42	91	68	86	80	56	72	28
proposed	85.99	**78.95**	72	97	91	87	86	97	86	84	90	83	93	83	64	44	93	73	87	81	59	74	25
[15]	**86.76**	77.78	76	99	90	77	84	99	82	88	88	80	90	90	71	47	94	68	90	73	55	77	15

Fig. 4. Segmentation accuracies of the daterterm, the pure Potts model, our approach using the continuous formulation of the co-occurrence energy, and the results by Ladicky et al. [15]. The scores for each label are defined as $\frac{\textit{True Positives} \cdot 100}{\textit{True Positives} + \textit{False Negatives}}$. While the score of the proposed method is slightly below that of [15] in the global score, it provides a better average performance.

6 Runtime

The proposed algorithm is based on a variational approach which allows for an implementation on graphics hardware. In practice the optimization scheme presented in Section 4 converges within 1000 iterations in terms of the maximal change of two successive iterations. For the experiments we used a NVIDIA Geforce GTX480 GPU and obtained average runtimes per image of 10 seconds for the co-occurrence segmentation. The runtimes are similar to the computation of the pure Potts model, since the overhead of computation compared to the classical multi-labeling problems is marginal as the configuration priors are defined on n scalar indicator variables compared to $\mathcal{O}(|\Omega|)$ variables for the indicator functions u.

7 Conclusion

We proposed a convex framework for continuous multi-label optimization which allows for the integration of semantic scene label information by means of co-occurrence priors. We formulated a variational approach together with a convex relaxation which can be optimized with fast primal-dual schemes. The approach compared favourably with respect to the discrete co-occurrence prior by Ladicky et al. [15] and optimality bounds demonstrate the tightness of our convex relaxations.

A Appendix

In the following we will give some details on the optimization scheme for solving saddle point problem (4.1). We start by introducing the Lagrange multipliers used for handling the convex constraints. Then we give a list of the update steps performed in our iterative minimization scheme.

In order to impose the label configuration priors, several kinds of constraints have to be implemented. The *simplex constraint* $\sum_i u_i(x) = 1$ in (2.3) can be implemented by introducing Lagrange multipliers $\lambda : \Omega \to \mathbb{R}$. The inequality constraints can also be easily implemented by introducing the following Lagrange multipliers:

Contraints	Lagrange Multipliers
$l_i(u) \geq u_i(x) \; \forall x \in \Omega$	$\alpha_i : \Omega \to \mathbb{R}^+ \quad \forall\, i \in \{1 \cdots n\}$
$\psi_{ij,1} + \psi_{ij,2} \leq c_{ij}$	$\beta_{ij} \in \mathbb{R}^- \quad \forall\, i < j \in \mathcal{G}$
$\varphi_{ij,1} + \varphi_{ij,2} \leq 0$	$\theta_{ij} \in \mathbb{R}^- \quad \forall\, i < j \in \mathcal{G}$
$\varphi_{ij,1} + \psi_{ij,2} \leq 0$	$\eta_{ij} \in \mathbb{R}^- \quad \forall\, i < j \in \mathcal{G}$
$\psi_{ij,1} + \varphi_{ij,2} \leq 0$	$\xi_{ij} \in \mathbb{R}^- \quad \forall\, i < j \in \mathcal{G}$

The update steps are performed point-wise for all $x \in \Omega$, $i \in \{1, ..., n\}$ and for all label pairs $1 \leq i < j \leq n$ which makes it possible to parallely implement above algorithm on modern graphics cards. Local constraints are tackled by a simple orthogonal projection Π into the respective convex sets which can be performed by simple truncation. Overall the update steps for solving saddle-point problem (4.1) are iterated in the following order:

Updates for dual variables:

$$p_i^{k+1}(x) = \Pi_{\bar{\mathcal{K}}} \left(p(x)_i^k + \frac{1}{2} (\nabla_w \bar{u}_i^k(x)) \right)$$

$$\lambda^{k+1}(x) = \lambda^k(x) + \frac{1}{n} \left(\sum_{i=1}^n \bar{u}_i^k(x) - 1 \right)$$

$$\alpha_i^{k+1}(x) = \Pi_{\mathbb{R}^+} \left(\alpha_i^k(x) + \frac{1}{2} (\bar{u}_i^k(x) - \bar{l}_i^k) \right)$$

$$\varphi_{ij}^{k+1} = \varphi_{ij}^k + \frac{1}{3} \left(\begin{matrix} 1 - \bar{l}_i^k + \bar{\theta}_{ij}^k + \bar{\eta}_{ij}^k \\ 1 - \bar{l}_j^k + \bar{\theta}_{ij}^k + \bar{\xi}_{ij}^k \end{matrix} \right)$$

$$\psi_{ij}^{k+1} = \psi_{ij}^k + \frac{1}{3} \left(\begin{matrix} \bar{l}_i^k + \bar{\beta}_{ij}^k + \bar{\xi}_{ij}^k \\ \bar{l}_j^k + \bar{\beta}_{ij}^k + \bar{\eta}_{ij}^k \end{matrix} \right)$$

Updates for primal variables:

$$u_i^{k+1}(x) = \Pi_{[0,1]} \left(u_i^k(x) - \frac{1}{2 + |\mathcal{N}(x)|} \left(\varrho_i(x) + \alpha_i^k(x) + \lambda^k(x) - \mathrm{div}_w \, p_i^k(x) \right) \right)$$

$$l_i^{k+1} = l_i^k - \frac{1}{|\Omega|}\left(\int_\Omega \alpha_i^k(x)\,dx\right)$$

$$\beta_{ij}^{k+1} = \Pi_{\mathbb{R}^-}\left(\beta_{ij}^k - \frac{1}{2}(\psi_{ij,1}^k + \psi_{ij,2}^k - C_{ij})\right)$$

$$\theta_{ij}^{k+1} = \Pi_{\mathbb{R}^-}\left(\theta_{ij}^k - \frac{1}{2}(\varphi_{ij,1}^k + \varphi_{ij,2}^k)\right)$$

$$\eta_{ij}^{k+1} = \Pi_{\mathbb{R}^-}\left(\eta_{ij}^k - \frac{1}{2}(\varphi_{ij,1}^k + \psi_{ij,2}^k)\right)$$

$$\xi_{ij}^{k+1} = \Pi_{\mathbb{R}^-}\left(\xi_{ij}^k - \frac{1}{2}(\psi_{ij,1}^k + \varphi_{ij,2}^k)\right)$$

Extrapolation steps:

$$\bar{u}_i(x)^{k+1} = 2u_i(x)^{k+1} - u_i(x)^k$$

$$\bar{l}_i^{k+1} = 2l_i^{k+1} - l_i^k$$

$$\bar{\beta}_{ij}^{k+1} = 2\beta_{ij}^{k+1} - \beta_{ij}^k$$

$$\bar{\theta}_{ij}^{k+1} = 2\theta_{ij}^{k+1} - \theta_{ij}^k$$

$$\bar{\eta}_{ij}^{k+1} = 2\eta_{ij}^{k+1} - \eta_{ij}^k$$

$$\bar{\xi}_{ij}^{k+1} = 2\xi_{ij}^{k+1} - \xi_{ij}^k$$

Note that for the differential operators we use a non-local version of the gradient ∇_w and its respective adjoint operator $-\operatorname{div}_w$ which are defined on a neighbourhood of x denoted by $\mathcal{N}(x)$. For details see [8], and for a complete introduction to non-local operators and their applications in computer vision, we refer to [24].

References

1. Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:2001, 2001.
2. A. Chambolle, D. Cremers, and T. Pock. A convex approach for computing minimal partitions. Tech. rep. TR-2008-05, University of Bonn, 2008.
3. A. Chambolle, D. Cremers, and T. Pock. A convex approach to minimal partitions. *SIAM Journal on Imaging Sciences*, 5(4):1113–1158, 2012.
4. A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *JMIV*, 40(1):120–145, 2011.
5. A. Delong, L. Gorelick, O. Veksler, and Y. Boykov. Minimizing energies with hierarchical costs. *International Journal of Computer Vision*, 100(1):38–58, 2012.
6. M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. Pascal (VOC2010) Results.
7. P. F. Felzenszwalb and D. P. Huttenlocher. Efficient graph-based image segmentation. *Int. J. Comput. Vision*, 59(2):167–181, Sept. 2004.

8. G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. *Multi-scale Modeling & Simulation*, 7(3):1005–1028, 2008.

9. B. Goldluecke and D. Cremers. Convex relaxation for multilabel problems with product label spaces. In *ECCV*, 2010.

10. B. Goldluecke, E. Strekalovskiy, and D. Cremers. Tight convex relaxations for vector-valued labeling. *SIAM Journal on Imaging Sciences*, 2013.

11. M. Klodt, T. Schoenemann, K. Kolev, M. Schikora, and D. Cremers. An experimental comparison of discrete and continuous shape optimization methods. In *ECCV*, Marseille, France, October 2008.

12. P. Kohli, L. Ladicky, and P. H. S. Torr. Robust higher order potentials for enforcing label consistency. *International Journal of Computer Vision*, 82(3):302–324, 2009.

13. N. Komodakis, N. Paragios, and G. Tziritas. Mrf optimization via dual decomposition: Message-passing revisited. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8, 2007.

14. N. Komodakis and G. Tziritas. Approximate labeling via graph cuts based on linear programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(8):1436–1453, 2007.

15. . Ladick, C. Russell, P. Kohli, and P. Torr. Inference methods for crfs with co-occurrence statistics. *International Journal of Computer Vision*, 103(2):213–225, 2013.

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

17. J. Lellmann, J. H. Kappes, J. Yuan, F. Becker, and C. Schnörr. Convex multi-class image labeling by simplex-constrained total variation. In X.-C. Tai, K. Mórken, M. Lysaker, and K.-A. Lie, editors, *SSVM 2009*, volume 5567 of *LNCS*, pages 150–162. Springer, 2009.

18. J. Malik, S. Belongie, T. Leung, and J. Shi. Contour and texture analysis for image segmentation. *Int. J. Comput. Vision*, 43(1):7–27, June 2001.

19. C. Nieuwenhuis, E. Toeppe, and D. Cremers. A survey and comparison of discrete and continuous multilabel segmentation approaches. *International Journal of Computer Vision*, 2013.

20. T. Pock and A. Chambolle. Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In *ICCV*, 2011.

21. T. Pock, D. Cremers, H. Bischof, and A. Chambolle. An algorithm for minimizing the piecewise smooth mumford-shah functional. In *IEEE International Conference on Computer Vision (ICCV)*, Kyoto, Japan, 2009.

22. B. C. Russell, W. T. Freeman, A. A. Efros, J. Sivic, and A. Zisserman. Using multiple segmentations to discover objects and their extent in image collections. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*, CVPR '06, pages 1605–1614, Washington, DC, USA, 2006. IEEE Computer Society.

23. E. Strekalovskiy, B. Goldluecke, and D. Cremers. Tight convex relaxations for vector-valued labeling problems. In *ICCV*, 2011.

24. M. Werlberger, M. Unger, T. Pock, and H. Bischof. Efficient Minimization of the Non-Local Potts Model. In *ICSSVMCV*, 2011.

25. J. Yuan and Y. Boykov. Tv-based multi-label image segmentation with label cost prior. In *BMVC*, pages 1–12, 2010.

26. C. Zach, D. Gallup, J.-M. Frahm, and M. Niethammer. Fast global labeling for real-time stereo using multiple plane sweeps. In *Vision, Modeling and Visualization*, October 2008.

27. S. C. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and bayes/mdl for multi-band image segmentation. *PAMI*, 18:884–900, 1996.