Semantic Image Labelling as a Label Puzzle Game

Peter Kontschieder[1]
kontschieder@icg.tugraz.at
Samuel Rota Bulò[2]
srotabul@dsi.unive.it
Michael Donoser[1]
donoser@icg.tugraz.at
Marcello Pelillo[2]
pelillo@dsi.unive.it
Horst Bischof[1]
bischof@icg.tugraz.at

[1] Institute for Computer Graphics and Vision
Graz University of Technology
Austria
[2] Dipartimento di Scienze Ambientali, Informatica e Statistica
Università Ca' Foscari Venezia
Italy

In this work we present a novel solution to the semantic image labelling problem, *i.e*. the task of assigning object class labels to all pixels in a test image. We provide an interpretation in terms of a label puzzle game, where the final labelling is obtained by assembling discriminatively learned candidate sets of label puzzle pieces. Each label puzzle piece represents a topological and semantically plausible label configuration stemming from pixel-wise annotated training data. The puzzle game is generated by means of a modified random forest classifier [2], designed to learn the local, topological label-structure and hence the local context associated to the training data. To solve the puzzle game we propose an iterative optimization technique that maximizes an agreement function by alternatingly seeking for the best label piece per pixel and the corresponding, semantic labelling per image (see Figure 1). Our puzzle solving algorithm is simple and efficient to implement and we provide both, theoretical properties as well as experimental results. In the remainder of this abstract page, we provide brief definitions of the label puzzle game, the objective function of the puzzle game solver as well as a glance at our experimental findings on the challenging MSRCv2 and CamVid databases.

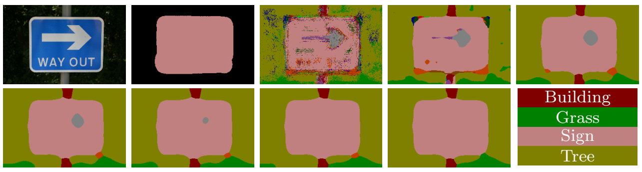

Figure 1: A semantic image labelling example of the proposed approach. From top left to bottom right: Image to be labelled, groundtruth labelling, initial random forest classification, labellings obtained by our approach after $t = 0, 5, 10, 20, 35, 50$ iterations, finally contained label captions.

The Label Puzzle Game considers semantic image labelling as the task of assembling possibly overlapping label puzzle pieces, where the pieces are label configurations obtained from a *label puzzle game generator*. Each puzzle piece $p \in \mathscr{P}$ is a function $p : \mathbb{Z}^2 \to Y \cup \{\perp\}$ mapping two-dimensional points to labels $Y = \{1, \ldots, k\}$ or to void ($\perp$), a special symbol indicating the absence of a label. A *puzzle configuration* is a function $z : D \to \mathscr{P}$ associating each pixel in $D \subseteq \mathbb{Z}^2$ with exactly one puzzle piece in $\mathscr{P}$. The set of puzzle configurations is denoted as $\mathscr{Z}$. A *labelling* for an image is a function $\ell : D \to Y$ mapping pixels in D to labels in Y. The sets of images, labellings and puzzle pieces are denoted by $\mathscr{I}$, $\mathscr{L}$ and $\mathscr{P}$, respectively.

We define the *agreement* of a puzzle piece $p \in \mathscr{P}$ located in $(i, j) \in D$ with a labelling $\ell \in \mathscr{L}$ as the number of corresponding pixels sharing the same label, *i.e.*

$$\phi^{(i,j)}(p, \ell) = \sum_{(u,v)\in D} [p(u-i, v-j) = \ell(u, v)] \,,$$

where $[Q]$ are the Iverson brackets yielding 1 if proposition Q is true and 0 otherwise. Given a puzzle configuration $z \in \mathscr{Z}$ and a labelling $\ell \in \mathscr{L}$, the *total agreement* $\Phi(z, \ell)$ of the image labelling puzzle is the sum of the agreements of each puzzle piece in z with the labelling ℓ according to

$$\Phi(z, \ell) = \sum_{(i,j)\in D} \phi^{(i,j)}(z_{i,j}, \ell) \,.$$

A *label puzzle game* for an image $f \in \mathscr{I}$ is a function π_f mapping each pixel $(i, j) \in D$ to a non-empty set of puzzle pieces $\pi_f(i, j) \subseteq \mathscr{P}$. This function restricts the possible choices of puzzle pieces per pixel and hence, also the set of admissible puzzle configurations to

$$\mathscr{Z}|_{\pi_f} = \{z \in \mathscr{Z} \,|\, z_{i,j} \in \pi_f(i, j)\} \,.$$

A solution of a label puzzle game π_f is a pair $(z^*, \ell^*) \in \mathscr{Z}|_{\pi_f} \times \mathscr{L}$ consisting of an admissible puzzle configuration and a labelling for f yielding the maximum total agreement

$$(z^*, \ell^*) \in \underset{(z,\ell)}{\arg\max} \left\{\Phi(z, \ell) \,\middle|\, (z, \ell) \in \mathscr{Z}|_{\pi_f} \times \mathscr{L}\right\} \,,$$

and can be obtained by iteratively updating the puzzle configuration $z^{(t+1)}$ at time $(t + 1)$ for a labelling $\ell^{(t)}$ by

$$z_{i,j}^{(t+1)} \in \underset{p}{\arg\max} \left\{\phi^{(i,j)}\left(p, \ell^{(t)}\right) \,\middle|\, p \in \pi_f(i, j)\right\} \,,$$

and producing the new labelling $\ell^{(t+1)}$ by taking a majority vote from all overlapping puzzle pieces according to

$$\ell^{(t+1)}(u, v) \in \underset{y}{\arg\max} \left\{ \sum_{(i,j)\in D} \left[z_{i,j}^{(t+1)}(u-i, v-j) = y\right] \,\middle|\, y \in Y \right\} \,.$$

Results at a Glance We evaluated our method on the MSRCv2 and CamVid databases and compared to results from plain random forest classification and the widely used conditional random field (CRF) model [1], using the random forest statistics as unary potentials and the standard contrast-sensitive Potts model as pairwise term. The results are listed in Table 1 and details to the experimental setup can be found in Section 5 of the paper.

Method	MSRCv2		CamVid	
	Global	Class Avg	Global	Class Avg
Unary	59	47	71	45
Unary + CRF	67	57	75	48
Unary + Puzzle	**70**	**60**	**82**	**50**

Table 1: Pixelwise classification scores of overall correctly classified pixels (Global) and per-class correctly classified pixels (Class Avg) in %.

Conclusion We proposed a novel solution to the semantic image labelling problem, formulated as a label puzzle game. First, the puzzle game is set up by assigning a set of discriminatively learned, structured class-label patches (puzzle pieces) to each pixel using a modified random forest classifier. To solve the label puzzle game, we introduced an optimization method that simultaneously selects puzzle pieces and assigns labels to pixels such that the agreement of the selected pieces with the underlying labelling is maximized. We found favourable results when directly comparing to a CRF which we explain by the fact that our method is restricted to selecting per-pixel puzzle pieces only from sets of semantically plausible label configurations rather than propagating arbitrary labellings in the associated graph.

[1] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *(PAMI)*, 2001.

[2] Peter Kontschieder, Samuel Rota Bulò, Horst Bischof, and Marcello Pelillo. Structured class-labels in random forests for semantic image labelling. In *(ICCV)*, 2011.