Déjà Vu Object Localization using IRF neural networks properties

Philippe Smagghe, Jean-Luc Buessler, and Jean-Philippe Urban

Abstract—This article introduces an original method of image detection and localization in a picture by scoring the output of a neural network to indicate an already seen (Déjà Vu) input. The classifier is a feedforward multi-layer perceptron adapted to supervised image recognition named Image Receptive Fields Neural Network (IRF-NN). It has interesting properties: fast and efficient training, as well as accurate classification skills on large learning sets. We show that a simple analysis of the neural response can be used to evaluate the probability that an input is known. This evaluation can be efficiently used to detect and localize objects in a picture using a sliding window approach. The generalization skills of the IRF-NN induce nice properties that improve the recognition process and the speed of the algorithm.

I. INTRODUCTION

OBJECT localization is the task of retrieving properties of objects from images, such as location, orientation, or shape. The main challenge is to perform the detection without being affected by changes in the visual appearance of objects or by the complexity of the scene. Research in this area is very active, as it is motivated by several fields of applications such as face detection [1], pedestrian detection [2] or robotic manipulation [3]. Most localization algorithms are based on the principle of a sliding window algorithm scanning the image combined with the evaluation of a probability-of-presence score.

This paper introduces an original method to score the detection window, based on the learning of a representative set of photos of objects by a neural network. The objects to be detected can be numerous (several hundreds) and completely different from each other, since the search is not feature-based. After training, the neural network is able to identify every learned object that appears in the sliding window. In this paper we focus on a complementary property of the supervised classification algorithm. It is possible to distinguish a known input from an unknown one by performing a simple statistical analysis of the neural response. A high score indicates a strong similarity between the input image and one of the training images. This score is obtained without comparing the input image to every reference.

This technique is based on a variant of feedforward multilayer perceptron (MLP) that we have recently introduced for image recognition, named Image Receptive Fields Neural Network (IRF-NN) [4]. As a Reservoir Computing [5] or Extreme Learning Machine (ELM) [6] approach, IRF-NN is likewise constituted of a large but constant hidden layer. The learning algorithm is only applied on the output layer. To adapt this approach for images, we introduce a specific structure of the weights for the input layer: neurons are stochastically initialized with a few degrees of freedom, so that their weights form random receptive fields in the image.

A recent communication [7] has presented our application of novelty detection with IRF-NN: simple criteria calculated with the neural response can detect unknown inputs. In the present article, these criteria are used to define the Déjà Vu (DV) score that evaluates a probability-of-presence of known inputs. This score, combined with a sliding window technique, can efficiently detect and precisely localize objects. The neural response can therefore identify the detected object. We name Déjà Vu Object Localization (DVOL) this search algorithm.

The properties of the DV score are illustrated using ALOI object photos [8]. ALOI views inserted into large pictures of rich content are easily detected. The generalization property of IRF-NN enables the detection of various views of objects. Moreover, this property allows increasing the sliding step, thus speeding up the search considerably.

This paper continues as follows: part 2 briefly introduces IRF-NN and its application of novelty detection. Part 3 presents a few existing detection algorithms and develops DVOL for a fast and robust multi-object localization. Part 4 evaluates the DV scores and experimentally studies their properties. Finally, part 5 presents experiments and discusses the localization results.

II. IRF-NN AND NOVELTY DETECTION

A. Structure of the IRF-NN algorithm

The proposed neural structure (IRF-NN) is presented in Fig. 1, illustrating the notion of image receptive fields. It uses the architecture of a MLP with one hidden layer [9]. The neuronal activation vector $\mathbf{H}$ is computed by taking the dot product of the raw image $I$ with weight vectors $\mathbf{g}_i$, designed as regular functions of the image space. As shown in Fig. 1, each weight vector $\mathbf{g}_i$ is represented as a bidimensional image to emphasize its Gaussian field structure,

$$
\mathbf{g}_i(x,y) = \gamma_i + \exp \left( -\frac{(x-\mu_{x_i})^2}{2\sigma_{x_i}^2} - \frac{(y-\mu_{y_i})^2}{2\sigma_{y_i}^2} \right)
$$

(1)

where $\sigma, \gamma, \mu$ are fixed random parameters, and $n_x, n_y$ the image sizes.
The color selectivity is given by random amplitude $\alpha_{lc}$ for each RGB component. The sigmoid output function ensures the non-linearity of the response. For a neuron $i$, with $x,y$ the coordinates of the pixels of the image, the activation is:

$$h_i = \tanh \left( \sum_l \sum_{x,y} \alpha_{lc} I_c(x,y) \cdot g_i(x,y) \right)$$

(2)

The non-linearity combined with the Gaussian structure of the weights induces a very selective response to stimuli, like receptive fields in biology. Every set of images can be accurately represented due to the random initialization and the very large number of neurons.

The training of the weights $W$ is computed without iteration, using a linear regression such as $W = S \cdot H^+$, with $S$ the desired outputs and $H^+$ the Moore-Penrose pseudo-inverse of the neuron activation $H$ of the training set. $S$ is set with a 1-of-$n$ encoding in classification applications. The class response $r$ is obtained by taking the maximum neural output such as:

$$r(l) = \arg \max (\hat{S}) = \arg \max (W \cdot H(l))$$

(3)

B. Properties and classification results

The neural network described above can be used to solve a supervised classification task using an image example set. The images are not pre-processed, and no feature extraction is needed. The algorithm has the advantage of a very fast training compared to other algorithms, with few and easily determined parameters.

The algorithm was tested on several image datasets. MNIST [10] is composed of images of isolated handwritten digits. ALOI-1000 and COIL-100 [11] are composed of various objects captured under different angles of view. For ALOI and COIL, one fourth of the images were used for training and the rest as test images. The classification results are presented in table 1. The algorithm shows very good performance in all classification tasks and excellent generalization skills for angle-of-view datasets like ALOI and COIL, with many classes and few training images. Training time for 18,000 ALOI views is only 20 minutes, compared to several hours in [13].

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Neurons</th>
<th>Classification Rate</th>
<th>State of Art Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNIST</td>
<td>7500</td>
<td>98.6 %</td>
<td>99.8 % [12]</td>
</tr>
<tr>
<td>ALOI</td>
<td>9000</td>
<td>99.8 %</td>
<td>89.6 % [13]</td>
</tr>
<tr>
<td>COIL</td>
<td>900</td>
<td>99.4 %</td>
<td>99.9 % [14]</td>
</tr>
</tbody>
</table>

C. Novelty detection criteria

We recently demonstrated [7] that the analysis of the response vector $\hat{S}$ for a new image allows to detect novelty, i.e. that the input image is too different from the learning set to be reasonably classified. The validation of the network response is an important concept. Indeed it is noticeable that the class response $r(l)$ from (3) is selected among learned classes, but does not encompass the lack of response or a reject class.

The specific structure of the IRF-NN network eases novelty detection with one or more simple criteria. We briefly recall here the principle and the proposed criteria, as our object localization is built on this technique.

It is well known that neural networks generate consistent results for entries similar to the training ones, and are inconsistent for unknown entries. It is usually difficult to evaluate the randomness of only one response. However, the specific configuration of IRF-NN allows an efficient evaluation. It takes advantage of two characteristics:

- The inner vector $h$ is designed as a generic high-dimensional representation for any image input, without training or adaptation on a specific subset. It is also continuous, as changes in activation are small for similar images.

- Using the network for classification into many classes with a 1-of-$n$ representation $S$ generates a high dimensional and high rank output space.
Several criteria $C$ that are relevant to discriminate novel objects were studied, using the statistical features of the response vector $\hat{S} = W \cdot h(I)$.

Considering $S^* \in \mathbb{R}^m$ the neural response sorted in descending amplitude, the criteria are defined as:

- $C_1 = \left| 1 - (S^*_1 - S^*_i) \right|$, an empirical classification margin;
- $C_2 = \left| S^*_m \right|$, the magnitude of the last response;
- $C_3 = \left| 1 - S^*_1 \right|$, the magnitude of the first response;
- $C_4 = \sigma(S) = \sqrt{\frac{1}{m-1} \sum_{i=2}^{m} (S^*_i - \overline{S})^2}$, the standard deviation.

Note that each $C_i$ value is near 0 for the images of the training set, as the expected responses are $S^*_i = 1, \forall i \in \{2..m\}$. For an unknown input, $C$ values are greater because $S$ is very stochastic in spite of $E(S_1) = \frac{1}{m} \forall i \in \{1..m\}$.

### III. Object Localization

This analysis of the IRF-NN response can also be used to detect and localize objects in a larger picture $P$. The novelty criteria are exploited to evaluate the probability that a sub-window of $P$ contains a known object. This probability-of-presence is named Déjà Vu score, because this value is not determined for a specific object, but rather for any of the objects of the learning set. When the DV score is sufficiently high, the class response $r(I)$ can be trusted to identify the object.

As illustrated in Fig. 2, the purpose of a localization algorithm is to detect and localize objects within images. This task can be difficult due to changes in the visual appearances of the searched objects. Main causes of changes are the objects poses, but also varying intensities, colors and shadows due to light sources, occultations, and blurred images.

![Fig. 2. Example of localization task](image)

5 objects from the ALOI dataset are included in this picture.

A. Existing detection algorithms

There is an extensive literature on object detection. The survey by Yang [1] depicts the plethora of algorithms published only for face detection. Hence we mention here only a few papers relevant to illustrate some aspects of our work.

A few approaches initiate detection with the search of salient features in the picture $P$, followed by a matching evaluation with object features stored in a database (e.g. [3]). However, matching is a complex task, and most authors use a sliding detection window (e.g. [2][15][16][17]).

To score the detection window, most techniques use a feature vector, with various principles such as SIFT [18], histograms [15] or sparse part-based representation [16]. Evaluation is then made on this vector, either with learning algorithms such as neural networks and SVM [15][17], or by comparison with a reference set, sometimes with interesting algorithmic optimizations [16][19].

A few algorithms nonetheless show that it may be efficient to directly work on images. Viola and Jones [2] achieve high detection rates on complex problems with cascades of classifiers. Deep Convolutional Neural Networks realize excellent recognition performances after training [12].

The IRF-NN approach [7] retains the principle of directly learning images, using a simple architecture with only a single feedforward layer. The detection technique introduced here needs only positive examples of objects to recognize, unlike most of the mentioned techniques.

B. Principle of Déjà Vu localization

We focus on the problem of multiple object retrieval, with invariance to the pose of the object. In this paper, the apparent size of the object is assumed known, and the size of the sliding window is fixed.

Let us consider first that the sliding window scans the whole picture for every pixel $k$. Each cropped image $I_k$ is evaluated by an IRF-NN trained with a fair number of objects, typically a hundred or more. The response $\hat{S}(I_k)$ is used to determine the DV score. A scoring map is formed by scanning the picture.

This score is based on the calculated novelty detection criteria (see Section II). When the neural response for an input yields a low DV value, this sub-image does not contain any object for which the algorithm was trained. Alternatively, sub-windows with large DV values are indicative of the presence of a known object. This value decreases when the object image is shifted, which allows a precise localization.

Since the DV score detects any known image, multiple different objects can be localized and recognized in the same scan, producing a multi-object detector. Two variants of detection techniques can be applied depending on the application. When the number of objects in the picture is known, a corresponding number of distinct maxima are retained from the scan. If not, the scoring map is thresholded.
\( C. \text{ Déjà Vu Score} \)

The DV score is defined as a function of the neural response \( \hat{S} \). It is calculated from the novelty criteria \( C_i \) described in part II.C, as a combination of normalized form \( D_i \), \( \forall i \in [1, A] \):

\[
D_i = s(C_i) = 0.5\left(1 + \tanh\left( a_i(1 - C_i - b_i) \right)\right) \quad (4)
\]

Parameters \( a_i \) and \( b_i \) of the sigmoid function are adjusted for each criterion. Each \( D_i \) has outputs in the [0 1] range, where 0 indicates unknown images, and 1 known images. The DV score is defined as the product of these values, which appears more robust than each criterion alone.

\[
DV = \prod_{i=1}^{A} D_i = \prod_{i=1}^{A} t_i(C_i) \quad (5)
\]

As a comparison, we study an alternative combination. The probability of presence \( P \) is expressed as a multi-variable logistic function with parameters \( \beta_i, i \in [0..A] \):

\[
P = \frac{1}{1 + \exp(-\beta_0 - \sum_{i=1}^{A} \beta_i C_i)} \quad (6)
\]

The parameters are identified by a classical logistic regression, which is supervised with known and unknown image sets.

\( D. \text{ Generalization and localization constraints} \)

The IRF neural network has excellent abilities for generalization for the task of image recognition [4]. Object detection and localization are also supported by this capacity of generalization, but localization has certain constraints in this regard:

1. The neural network is trained with a small but representative part of the views for each object. The DV score must allow the identification of one of these objects even when its appearance is different. This means that the DV score must remain constantly high for learning images as well as for variations of these images, such as rotated, shifted and/or blurred object images.

2. To precisely localize objects, the DV score must ensure a maximal value when the position matches the reference: its value must decrease when the distance augments.

These constraints can appear contradictory but we verify in Section IV.E that they are compatible. It is possible to combine precise localization and high generalization skills.

\( E. \text{ Scanning optimization} \)

The sliding window algorithm is extremely time-consuming when every pixel position in the picture is evaluated. One solution is to take advantage of the ability to recognize variations of the same images (constraint 1). In this case, it would be the ability to generalize to shifted images. If the DV score can still distinguish images that have been shifted x pixels from unknown images, then the scan would only need to consider images every 2x pixels instead of every pixel. To maintain the ability to precisely localize images as stated in constraint 2, the scanning process can be optimized as follows: instead of one global scan for every position, two scans can be used:

- A coarse scan with a maximum of 2x pixels step, which serves as an object detector. This scan will reveal the approximate locations of every object in the picture.
- A fine scan around the areas detected by the coarse scan, to precisely localize the objects at pixel precision.

The neural network used by the two scans does not have to be the same. The fine scan does not need to properly generalize to shifted images, so it can be simpler and faster.

The neural network of the coarse scan is trained with an extended learning set. Shifted variants of each image are added to induce required invariance to position. The DV score is high only in the appropriate scale as we verify in the next section.

IV. EVALUATION OF THE DV SCORE

A. Presentation of the Image Datasets

The experiments presented in this section use the ALOI image database [8], a collection of 72,000 photos of 1,000 objects recorded with systematical variation of viewing angle. To suit the experimental purposes of this article, we create two alternate versions, ALOI-\( \beta \), and ALOI-\( \gamma \).

ALOI-\( \beta \)

ALOI objects are surrounded with large dark backgrounds that are not relevant for the identification of the objects and detrimental to the network’s ability to generalize images. We therefore created ALOI-\( \beta \), where the bounding boxes of the objects are extracted, and all 72,000 images are then stretched to fit a size of 100x100. This new set is well adapted for object localization but is now more difficult to train because the object shape is often lost during the cropping procedure. IRF-NN presents a classification rate of 98.7% on this set.

ALOI-\( \gamma \)

ALOI-\( \gamma \) is an extension of ALOI-\( \beta \) that will be used as a generalization benchmark. This dataset is build with 50,000 images randomly selected from ALOI-\( \beta \) and distorted with a random combination of the following transformations:

- Image rotations of angles in range \([-10^\circ, 10^\circ]\]
- Circular shifts in random direction in range \([0,10]\) pixels
- Gaussian blur filters of sizes in range \([0,20]\)

ALOI-\( \gamma \) images
B. Parametrization of Déjà Vu scores

Parameters $a_i$ and $b_i$ in equation (4) are determined according to the following procedure.

An IRF-NN is trained with images of ALOI-$\beta$ and the criteria $C_i$ are computed for all images of ALOI-$\gamma$. For each criterion, parameters $a_i$ and $b_i$ are manually adjusted to ensure that all the $D_i$ have maximal sensitivity and similar dynamic.

Table 2 shows the resulting normalization parameters.

<table>
<thead>
<tr>
<th>CRITERIA NORMALIZATION PARAMETERS</th>
<th>$C_1$</th>
<th>$C_2$</th>
<th>$C_3$</th>
<th>$C_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>1.0</td>
<td>2.5</td>
<td>1.5</td>
<td>6.7</td>
</tr>
<tr>
<td>$b$</td>
<td>0.5</td>
<td>0.75</td>
<td>0.7</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Parameters $\beta$ of the probability of presence are determined with a logistic regression that requires unknown samples. We use 20,000 images from ALOI-$\beta$, and 70,000 unknown images from other datasets. The calculated parameters $\beta$ are shown in table 3.

<table>
<thead>
<tr>
<th>LOGISTIC REGRESSION PARAMETERS</th>
<th>$\beta_0$</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$-$86</td>
<td>12</td>
<td>14</td>
<td>-8</td>
<td>73</td>
<td></td>
</tr>
</tbody>
</table>

C. Scores comparison and analysis

The performance of these scores to distinguish novel images was tested with the ALOI-$\beta$ dataset. Three different image sets are used:

1. The neural network was trained to recognize objects with a learning set $L$ of 9,000 images from the first 500 ALOI-$\beta$ objects. A high DV score must be verified for every trained image.

2. The generalization set $G$ contains the 27,000 remaining views of the same objects. These images are variations of known inputs but not novelty. They are meant to be well classified by IRF-NN and recognized as known by the DV score.

3. The unrelated set $U$ is a collection of 70,000 very diverse images extracted from unrelated high-resolution pictures gathered from various websites. As these images are unknown, the DV score must be very low.

The statistical distributions of each score for each set are shown in Fig. 3: distributions are significantly different. ROC curves for each score are shown in Fig. 4 to assess thresholding efficacy. Fig. 4.a evaluates the ability to distinguish images from the Generalization set against those of the Learning set, and Fig. 4.b the images from the Generalization set against those of the Unrelated set. $D_2$ and $D_4$ are very discriminative for similar images. Meanwhile, $D_1$ and $D_3$ are better suited in the case of unrelated images.

The challenge for object detection is to discriminate Learning and Generalization sets from the Unrelated set. As the DV score is used both for detection and localization of objects, the selectivity of Déjà Vu criteria is improved by the combination in eq. 5. This will be confirmed with the tests of the whole algorithm in part V. Next sections study the sensitivity of this combination to detect image variants and to localize them.

D. Score sensitivity to image variations

To properly localize objects, the DV value needs to be high for known objects but also sensitive to the shift between the test window and the reference image. This shift sensitivity must persist for every image to localize, i.e. all images from the generalization set.

For the following test, an IRF-NN is trained with $L_1$, a set composed of a hundred objects from the ALOI-$\beta$, with 18 images per object, for a total of 1,800 images.

Figure 5.a presents the resulting DV values for all 7,200 images available for the 100 objects of $L_1$. For each figure, scores for images from the unrelated set $U$ are also evaluated for comparison. As the training set is designed to induce invariance to rotations, we are satisfied to observe a low variation of the DV values for intermediate angles. The scores remain very high compared to those of the Unrelated set (blue line).

For comparison, Fig. 5.b shows high selectivity for images that are rotated in the plane with angles ranging from -50 to 50°. This means there will not be any undesirable generalization if rotation examples are not included in the training set. Fig 5.c plots DV for blurred images.

Fig 5.d shows DV for images that are circularly shifted in a left-right direction. This illustrates a good sensitivity near
the optimal superposition and for training images. Even if a more detailed study is necessary, it confirms the possibility of an accurate localization.

E. Algorithm optimization using generalization skills

As described in part III.D, the exhaustive scan of the picture is time consuming, but it is possible to increase the step of the sliding window by adding shifted versions of the original images in the training set. This extension is tested with the ALOI-β dataset.

Fig. 5.e shows score DV with the same shifted images as in Fig. 5.d, but with a neural network trained with additional images shifted 5 and 10 pixels in left-right directions, for a total of 5 variants per image. The shifted views are now recognized with a large DV value, and will be detected in a picture. Such trained network can properly recognize known up to 10 pixels shifted images, thus one can use steps up to 20 pixels during the coarse scan. The computational gain is substantial even if a larger network is required: in the test presented in section V, using a step of 10 pixels reduces the detection time from 25s down to 2.5s.

V. LOCALIZATION EXPERIMENTS AND RESULTS

Our localization algorithm is tested on the first 100 objects from the ALOI-β dataset. Objects images are rotated with angles $\in \{-90^\circ, -45^\circ, 0^\circ, 45^\circ, 90^\circ\}$. Ten high resolution (1000x1000) photographs are chosen for their complexity and rich backgrounds. For each photograph, 1000 tests are realized with the superposition of 10 object images, which are selected and positioned randomly. The experiment counts a total of 100,000 detections.

The first neural network is trained with 100 classes of ALOI-β, with 1 image per class, 16 shifted variations (5 & 10 pixels in 8 directions) and 4 in-plane rotated variations ($\pm 45^\circ$ and $\pm 90^\circ$) for each image were added to the training, for a total of 5,400 images for 2,700 neurons. The second neural network for the fine scan is trained without additional shifted variations and only 200 neurons. Total training time is 38s with a 3.2GHz CPU. Testing times were 2.5s for the coarse test and 0.3s for the fine test.

Fig. 6.b shows the mapped results of the $D_t$ values from the first scan, for the image shown in Fig. 6.a. Fig. 6.c shows the mapped results of DV score for the first scan. Finally, Fig. 6.d shows DV for the second scan.

Global localization results are shown in table IV. The results of the first scan are noted $n$-detection and $t$-detection rates. $n$-detection is used when the number of objects to find is fixed, and only the maximum scores are considered for the second scan. $t$-detection is used when all scores above a certain threshold are taken. The localization rate is calculated as the percentage of tests with a 0 pixel localization error. Recognition is the proportion of the detected views that have been correctly identified by the network. The results are evaluated for the DV score, compared to product $D_1D_4$, and to probability $P$ calculated from the logistic regression.

![Fig. 5. Score responses DV for unknown images (blue), and a) varying angle-of view images b) in-plane rotated images c) blurred images](image)
shifted images, with standard training (d) and extended training (e) (red)

TABLE IV
LOCALIZATION RESULTS

<table>
<thead>
<tr>
<th></th>
<th>DV</th>
<th>$D_1, D_2$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>n-detection rate</td>
<td>99.52%</td>
<td>96.09%</td>
<td>94.57%</td>
</tr>
<tr>
<td>r-detection rate</td>
<td>97.01%</td>
<td>94.38%</td>
<td>94.23%</td>
</tr>
<tr>
<td>Localization rate</td>
<td>99.99%</td>
<td>99.99%</td>
<td>99.99%</td>
</tr>
<tr>
<td>Recognition rate</td>
<td>99.28%</td>
<td>99.28%</td>
<td>99.28%</td>
</tr>
</tbody>
</table>

Results in Table IV suggest that combination DV has better detection properties. Pixel localization is obtained in every situation: less than 10 pixels of cumulated error for 100,000 tests.

The excellent performances of the $n$-detection validate the technique used. The DV score map shown in Fig. 6.c has well-defined maxima that localize objects. It is more difficult to use a simple thresholding technique with $r$-detection. However, several algorithmic options are possible to improve detection and are currently studied.

Fig. 6, Localization examples (Colormap applies to all score maps)
a) Example of localization test image (Five ALOI-$\beta$ objects are included). b) Mapped results of $D_i$, coarse scan

1490
VI. DISCUSSION

In this article, we presented a new method of object localization named DVOL, based on the recognition of Déjà Vu images by IRF neural networks. The DV score is quickly computed using the neural network response. This technique benefits of the properties of the neural network. Classification is directly applied on images, without prior feature extraction. Training is fast and can easily process tens of thousands of photos with a very high recognition rate.

Therefore, the localization algorithm DVOL shows interesting features:

- False positives are exceptional: the complexity of the scene does not affect the detection or the localization process.
- Very fast training time for hundreds of objects, which is a huge improvement compared to classifiers such as Viola-Jones [20], which can take days for single object families.
- Testing times do not augment when multiple objects are searched. Although these times are longer than common algorithms, large optimizations in the scanning process are possible, hinting at real-time localization of multiple objects.

Several experiments have illustrated in this communication the particular importance of the generalization properties of IRF-NN. They are fundamental for the DVOL approach, and are easily adapted for the needs of an application. Works in progress further the experimental validation:

- The DV score decreases progressively when the characteristics of the tested images differ from the reference images, allowing a very precise spatial localization. This property leads to the possibility of scale estimation.
- Adding specific views to the training set allows to extend the recognition ability, and to create the wanted invariances for the detection of objects, such as blurred, distorted and rotated images.
- The same technique is already applied to speed up the scan of the image: a network trained with shifted views can efficiently detect objects, for example with a 10 to 20 pixels step.

Future research will be focused on the optimizations of DVOL, especially in terms of speed and robustness, and development of other applications such as image tracking, simultaneous face recognition and localization, or pedestrian detection.

REFERENCES