Learning in Location Space
A New Framework for Object Detection

Damian Eads\textsuperscript{1,3}, David Helmbold\textsuperscript{1}, and Edward Rosten\textsuperscript{2}

\textsuperscript{1}Department of Computer Science
University of California, Santa Cruz

\textsuperscript{2}Department of Engineering
University of Cambridge

\textsuperscript{3}Space and Remote Sensing Sciences
Los Alamos National Laboratory

ISSDM Day
October 18, 2010
Problem

Find \((x, y)\) locations of objects.

Figure: Predictions marked in red.
Find \((x, y)\) locations of objects.

**Figure:** Predictions marked in red.
Sliding Window Detector

1. Train window detector
2. Slide window over image
   - all positions
   - all orientations
   - all scales
3. Arbitrate overlapping detections

Good for few large objects (face in a portrait)
How to find many smaller objects?
How do we find objects?

Pixel-based
find pixels of many objects

Beamer (2007-2009)

Location-based
make guesses of locations

HoS Boosting (late 2009-2010)
How do we find objects?

**Pixel-based**
find pixels of many objects

**Location-based**
make guesses of locations

Beamer (2007-2009)

HoS Boosting (late 2009-2010)
How do we find objects?

Pixel-based
find pixels of many objects

Location-based
make guesses of locations

Beamer (2007-2009)

HoS Boosting (late 2009-2010)
What is object detection anyway?

Is finding 50% of the pixels in all objects the same as finding 100% of the pixels in 50% of the objects?
Location Boosting

Radically different approach
- learns and predicts in \((x, y)\) location space
- combines ensemble of weak \((x, y)\) location predictors into strong predictor.

Contributions
- new kind of model: each weak hypothesis is a meta-object detector.
- new loss function: spatially motivated
- adaBoost variant: provably minimize loss function every iteration
Location Boosting

Radically different approach

- learns and predicts in \((x, y)\) location space
- combines ensemble of \textbf{weak} \((x, y)\) location predictors into \textbf{strong} predictor.

Contributions

- \textbf{new kind of model}: each weak hypothesis is a meta-object detector.
- \textbf{new loss function}: spatially motivated
- \textbf{adaBoost variant}: provably minimize loss function every iteration

Hit-or-Shift (HoS) Boosting
A **weak hypothesis** predicts locations on an image
\[ h = \{((x_1, y_1), c_1), ((x_2, y_2), c_2), \ldots, ((x_n, y_n), c_n)\}. \]

We filter away predictions with confidence lower than \( \theta \),
\[ h(\theta) = \{((x, y), c) \in h \text{ and } c \geq \theta\}. \]

**Figure:** Illustrates how \( h(\theta) \) changes as \( \theta \) is lowered.
Definition

A **weak hypothesis** predicts locations on an image

\[ h = \{((x_1, y_1), c_1), ((x_2, y_2), c_2), \ldots, ((x_n, y_n), c_n)\}. \]

We filter away predictions with confidence lower than \( \theta \),

\[ h(\theta) = \{((x, y), c) \in h \text{ and } c \geq \theta\}. \]

Figure: Illustrates how \( h(\theta) \) changes as \( \theta \) is lowered.
Definition

A **weak hypothesis** predicts locations on an image

\[ h = \{((x_1, y_1), c_1), ((x_2, y_2), c_2), \ldots, ((x_n, y_n), c_n)\}. \]

We filter away predictions with confidence lower than \( \theta \),

\[ h(\theta) = \{((x, y), c) \in h \text{ and } c \geq \theta\}. \]

---

**Figure:** Illustrates how \( h(\theta) \) changes as \( \theta \) is lowered.
Definition

A **weak hypothesis** predicts locations on an image:

\[ h = \{((x_1, y_1), c_1), ((x_2, y_2), c_2), \ldots, ((x_n, y_n), c_n)\}. \]

We filter away predictions with confidence lower than \( \theta \),

\[ h(\theta) = \{((x, y), c) \in h \text{ and } c \geq \theta\}. \]

\( \theta \geq 8 \)

**Figure:** Illustrates how \( h(\theta) \) changes as \( \theta \) is lowered.
Definition

A **weak hypothesis** predicts locations on an image

\[ h = \{((x_1, y_1), c_1), ((x_2, y_2), c_2), \ldots, ((x_n, y_n), c_n)\}. \]

We filter away predictions with confidence lower than \( \theta \),

\[ h(\theta) = \{((x, y), c) \in h \text{ and } c \geq \theta\}. \]

**Figure:** Illustrates how \( h(\theta) \) changes as \( \theta \) is lowered.
A **weak hypothesis** predicts locations on an image
\[ h = \{((x_1, y_1), c_1), ((x_2, y_2), c_2), \ldots, ((x_n, y_n), c_n)\}. \]

We filter away predictions with confidence lower than \( \theta \),
\[ h(\theta) = \{((x, y), c) \in h \text{ and } c \geq \theta\}. \]

**Figure:** Illustrates how \( h(\theta) \) changes as \( \theta \) is lowered.
A weak hypothesis predicts locations on an image 
\[ h = \{((x_1, y_1), c_1), ((x_2, y_2), c_2), \ldots, ((x_n, y_n), c_n)\}. \]

We filter away predictions with confidence lower than \( \theta \), 
\[ h(\theta) = \{((x, y), c) \in h \text{ and } c \geq \theta\}. \]

**Figure:** Illustrates how \( h(\theta) \) changes as \( \theta \) is lowered.
A weak hypothesis predicts locations on an image
\[ h = \{((x_1, y_1), c_1), ((x_2, y_2), c_2), \ldots, ((x_n, y_n), c_n)\}. \]

We filter away predictions with confidence lower than \( \theta \),
\[ h(\theta) = \{((x, y), c) \in h \text{ and } c \geq \theta\}. \]

**Figure:** Illustrates how \( h(\theta) \) changes as \( \theta \) is lowered.
**Definition**

A **weak hypothesis** predicts locations on an image $h = \{((x_1, y_1), c_1), ((x_2, y_2), c_2), \ldots, ((x_n, y_n), c_n)\}$.

We filter away predictions with confidence lower than $\theta$, $h(\theta) = \{(x, y), c) \in h \text{ and } c \geq \theta\}$.

**Figure**: Illustrates how $h(\theta)$ changes as $\theta$ is lowered.
The correlation function is given by

\[ C(u, v) = \int g \left( \frac{\|u - w\|_2}{r} \right) g \left( \frac{\|v - w\|_2}{r} \right) dw. \] (1)

We require \( 0 \leq C(u, v) \leq 1 \) and \( C(u, v) = C(v, u) \).

**Figure:** Plots of (a) two overlapping quadratic bumps with centers \( u \) and \( v \), (b) the truncated quadratic kernel \( g(\cdot) \) as a function of distance \( d = \|u - v\|_2 / r \), and (c) the correlation \( C(u, v) \) between \( u \) and \( v \).
The **objectness** of a location $u$ for $h$'s predictions with confidence at least $\theta$ is

$$f(u; \theta) = \sum_{v \in h(\theta)} C(u, v)$$  \hspace{1cm} (2)$$

where $C(u, v)$ quantifies the relatedness $u$ and $v$.

Objectness at $x$: 0.0

- **high** objectness on objects
- **low** objectness elsewhere
Objectness

Definition

The **objectness** of a location $u$ for $h$'s predictions with confidence at least $\theta$ is

$$f(u; \theta) = \sum_{v \in h(\theta)} C(u, v)$$  \hspace{1cm} (2)

where $C(u, v)$ quantifies the relatedness $u$ and $v$.

Objectness at $x$: 0.2

Ideally

- **high** objectness on **objects**
- **low** objectness elsewhere
Objectness

Definition

The **objectness** of a location \( u \) for \( h \)'s predictions with confidence at least \( \theta \) is

\[
f(u; \theta) = \sum_{v \in h(\theta)} C(u, v)
\]  

(2)

where \( C(u, v) \) quantifies the relatedness \( u \) and \( v \).

Objectness at \( x \): 0.5

Ideally

- **high** objectness on **objects**
- **low** objectness elsewhere
Objectness

Definition

The **objectness** of a location \( u \) for \( h \)'s predictions with confidence at least \( \theta \) is

\[
f(u; \theta) = \sum_{v \in h(\theta)} C(u, v)
\]

(2)

where \( C(u, v) \) quantifies the relatedness \( u \) and \( v \).

Objectness at \( x \): 0.8

Ideally

- **high** objectness on **objects**
- **low** objectness elsewhere
Training Algorithm

Boosting-based learning, each iteration:
- generate many weak hypotheses (grammar)
- pick best and add to master rule
- re-weight training data
- repeat
Issues

- false positives - "objectness" never reduced
- lack of detections suggests absence of objects
Issues
- false positives - "objectness" never reduced
- lack of detections suggests absence of objects

Solution
- let weak hypotheses predict negative objectness
- same "shift value" $s$ at all uncorrelated locations
- shift parameter $s$ found analytically
Hit-or-Shift (HoS) Framework

Definition

A **hit-or-shift (HoS) weak hypothesis** predicts positive carness when \( f(x; \theta) \) is positive, otherwise \(-s\).

\[
f'(x) = \begin{cases} 
\alpha f(x; \theta) & \text{if } f(x; \theta) > 0, \\
-s & \text{if } f(x; \theta) = 0.
\end{cases}
\]  

\( (3) \)

**Figure:** Hit-or-shift Weak Hypothesis
A hit-or-shift (HoS) weak hypothesis predicts positive carness when \( f(x; \theta) \) is positive, otherwise \(-s\).

\[
f'(x) = \begin{cases} 
\alpha f(x; \theta) & \text{if } f(x; \theta) > 0, \\
-s & \text{if } f(x; \theta) = 0.
\end{cases}
\] (3)

\textbf{Figure:} Hit-or-shift Weak Hypothesis
A hit-or-shift (HoS) master hypothesis is simply the cumulative objectness given by all weak hypotheses,

$$H_t(x) = \sum_{i=1}^{t} f'_i(x).$$  \hspace{1cm} (4)

**Figure:** Plots of (a) an image and its master hypothesis after (b) 10 iterations and (c) 100 iterations.
Loss: Object and Background

Each boosting iteration minimizes a two-part loss, the loss at **object locations**

\[ \mathcal{L}^{\text{obj}} = \sum_{x \in \text{obj}} \exp(-H_t(x)) \]  

(5)

and the loss at the **background**, 

\[ \mathcal{L}^{\text{bg}} = b \sum_{x \in \text{bg}} \max\{0, \exp(H_t(x)) - 1\} \],

(6)

where \( b \) is a trade-off parameter.
Experiments

**Car Detection**
- 12 large images of Phoenix, AZ
- 300 cars labeled
- split into three partitions

**Face Labeling**
- 1520 images of human faces
- label parts of face
- split into three partitions
Scoring is ill-posed

Pixel classifications in (a) and (b) have about the same number of correct pixels, is one better? (a) and (b) have many more pixels classified correctly than (c), but (c) finds the center of the object. Is one near miss (d) better than 10 near misses (e)? Are 5 correct hits (f) better than 1 near miss (d)?
Two scoring attributes:

1. **delineation boundary:** (a) object delineation (polygon) or (b) proximity delineation (circle)

2. **multiple detections penalty:** whether to treat **multiple detections** as false positives

For our scoring we used a **circular delineation boundary** and penalized **multiple multiple detections**.
Results: Arizona Test Set

**Quadratic Overlap (a)**

Arizona, Quadratic Overlap

![Precision/Recall Curve](image)

**Cylindrical (b)**

Arizona, Cylindrical

![Precision/Recall Curve](image)

**Average Precision**

<table>
<thead>
<tr>
<th></th>
<th>Rich</th>
<th>Haar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kink</td>
<td>0.355</td>
<td>0.138</td>
</tr>
<tr>
<td>Smooth</td>
<td>0.321</td>
<td>0.059</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Rich</th>
<th>Haar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kink</td>
<td>0.330</td>
<td>0.295</td>
</tr>
<tr>
<td>Smooth</td>
<td>0.256</td>
<td>0.148</td>
</tr>
</tbody>
</table>

**Figure:** Columns (a) and (b) show the precision/recall curves and average precisions for quadratic overlap and cylindrical kernels.
Results: Face Labelling

**Figure:** Shows precision recall curves using Haar features for labeling different parts of human faces.
Conclusions

- **location-based** approach more natural for **object detection**
  - easier data labeling
  - uniform treatment of weak/master hypotheses
  - uniform treatment of image (no subsampling)
- quickly learns and sifts through uninteresting background
- loss function directly tied to finding good locations
- **HoS weak hypotheses** are structured for efficient optimization
- can be used as a part detector
- **new algorithm**: *provably minimizes the loss* at every iteration given a new feature
Future Work

Our latest work involves significant adaptations for **large objects**
- new HoS detector based on SIFT features and vocabulary trees
- polar offset learning: exploits scale information to quickly learn offsets
- apply technique to PASCAL and CalTech data sets.