Beyond Hard Negative Mining:
Efficient Detector Learning via Block-Circulant Decomposition

J.F. Henriques, J. Carreira, R. Caseiro, J. Batista

Institute of Systems and Robotics
University of Coimbra
Motivation

• Setting: **object detection**

• Scan image with **learned template** of dense features (e.g., HOG, SIFT, CNN...)

• Core component of many approaches
Motivation

High performance usually requires Hard Negative Mining.
Motivation

High performance usually requires **Hard Negative Mining**.

1. Train initial model (e.g., SVM) with
   1. All positive samples (not shown)
   2. Random negative samples
Motivation

High performance usually requires **Hard Negative Mining**.

1. Train initial model (e.g., SVM) with
   1. All positive samples (not shown)
   2. Random negative samples

2. Scan negative images for **false-positives**

3. Re-train using **false-positives** as additional samples
Motivation

High performance usually requires **Hard Negative Mining**.

1. Train initial model (e.g., SVM) with
   1. All positive samples (not shown)
   2. Random negative samples

2. Scan negative images for **false-positives**

3. Re-train using **false-positives** as additional samples

(Repeat)

Several rounds are needed. Each round is **very expensive**.
Motivation

• Consider the **full set** of all potential samples.
Motivation

• Consider the **full set** of all potential samples.
Motivation

• Consider the **full set** of all potential samples.
Motivation

- Consider the **full set** of all potential samples.
Motivation

- Consider the **full set** of all potential samples.
Motivation

- Consider the **full set** of all potential samples.

- Hard Negative Mining avoids working on the full set by growing an **active set** of mined samples.
Motivation

Observation:

• Negative sets are highly redundant

• Pixels of overlapping windows are constrained to be the same

Questions:

• How does this influence a learning problem?
• Can we get rid of redundancies?
“Bold idea”

Let’s try to train with the full negative set.

Method:

• Collect base samples in a coarse grid.

• Train with the finer translations implicitly by using a Circulant Decomposition.
Cyclic shifts

• We need a **model of image translations**.

• Idea: Apply permutation matrix $P$ to base sample $\mathbf{x}$:

\[
\begin{bmatrix}
P \\
\end{bmatrix} \quad \times \quad \begin{bmatrix}
\mathbf{x} \\
\end{bmatrix} = \begin{bmatrix}
\mathbf{x}' \\
\end{bmatrix}
\]

$P$ represents a **cyclic shift**.
Cyclic shifts

• We need a **model of image translations**.

• Idea: Apply permutation matrix $P$ to base sample $\mathbf{x}$:

\[
\begin{bmatrix}
P
\end{bmatrix}
\begin{bmatrix}
\mathbf{x}
\end{bmatrix}
= \begin{bmatrix}
\end{bmatrix}
\]

$P$ represents a **cyclic shift**.
Cyclic shifts

- We need a **model of image translations**.

- Idea: Apply permutation matrix $P$ to base sample $\mathbf{x}$:

$$
\begin{bmatrix}
P
\end{bmatrix} \times \begin{bmatrix}
\mathbf{x}
\end{bmatrix} = \begin{bmatrix}
\mathbf{y}
\end{bmatrix}
$$

- Powers of $P$ shift by different amounts:

$$P^u \mathbf{x}, \ u \in \left\{ -\frac{\text{height}}{2}, \cdots, +\frac{\text{height}}{2} \right\}$$

- Represents a collection of fine translations of $\mathbf{x}$.

(Easy to generalize to horizontal + vertical)

$P$ represents a **cyclic shift**.
Cyclic shifts

- Goal: implicitly train with all shifts of all base samples.

shifts of base sample $x_1$

shifts of base sample $x_2$

shifts of base sample $x_3$
Cyclic shifts

- Goal: implicitly train with all shifts of all base samples.

- To see how shifted samples interact, analyze the Gram matrix:

\[ G_{ij} = \langle x_i, x_j \rangle \]

(Dot-products between pairs of samples)
### Gram matrix

**Dataset**

<table>
<thead>
<tr>
<th>Shift</th>
<th>Base sample</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ G \]

| dot-product |
Gram matrix

Dataset

<table>
<thead>
<tr>
<th>Shift</th>
<th>Base sample</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

$G$

dot-product
## Gram matrix

Dataset

<table>
<thead>
<tr>
<th>Shift</th>
<th>Base sample</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$G$  

dot-product
**Gram matrix**

- **Property #1:**
  
  \[ G \text{ is block-circulant} \]

⇒ Only 1 row of blocks is unique.

### Dataset

<table>
<thead>
<tr>
<th>Shift</th>
<th>Base sample</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1 2 3</td>
<td>1 2 3</td>
<td>1 2 3</td>
</tr>
<tr>
<td>0</td>
<td>1 2 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1 2 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1 2 3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ G \]
• **Property #1:**

   $G$ is **block-circulant**

   $⇒$ Only 1 row of blocks is unique.

• **Property #2:**

   Unique blocks contain the **cross-correlation** between all pairs of samples.

   $⇒$ Becomes simple product in the Fourier domain.
Proposed approach:

- **Fourier Transform** the samples (+ a small permutation)

\[ \iff \]

Projection on **Fourier Basis** with different frequencies.
• Each block of $G$ contains the projection on a different basis, or Fourier frequency.

 (# of frequencies = # of spatial cells of the samples)
We prove all off-diagonal blocks are zero.

Frequencies correspond to \textbf{independent} learning problems.

99.5\% for 18x10 HOG template
Circulant Decomposition

Base samples

(Negative and positive)

Feature extraction → Fourier Transform

1\textsuperscript{st} frequency

\begin{itemize}
  \item SVR
\end{itemize}

2\textsuperscript{nd} frequency

\begin{itemize}
  \item SVR
\end{itemize}

\begin{itemize}
  \item Inv. Fourier Transform
\end{itemize}

\vdots

s\textsuperscript{th} frequency

\begin{itemize}
  \item SVR
\end{itemize}

Split data by Fourier frequency

Concatenate trained weights from all Fourier frequencies

Template weights
Circulant Decomposition

- Equivalent to training with all shifts of the base samples.
- Surprisingly, easier than without shifts:
  - No shifts: one large SVR.
  - With shifts: many small SVR’s.
Circulant Decomposition

- **Closed-form**
- **Sub-problems:**
  - Can be solved in parallel
  - Use off-the-shelf SVR solvers
- **12 lines** of MATLAB code
Experiments

Single-template HOG object detection:

INRIA Pedestrians
• 1218 negative images
• $\sim 10^8$ potential samples

Caltech Pedestrians
• 4250 negative images
• $\sim 10^8$ potential samples
Experiments

Single-template HOG object detection:

INRIA Pedestrians
- 1218 negative images
- \(\sim 10^8\) potential samples

Caltech Pedestrians
- 4250 negative images
- \(\sim 10^8\) potential samples
Experiments

Single-template HOG object detection:

INRIA Pedestrians
- 1218 negative images
- $\sim 10^8$ potential samples

Caltech Pedestrians
- 4250 negative images
- $\sim 10^8$ potential samples
## Experiments

Single-template HOG object detection:

<table>
<thead>
<tr>
<th></th>
<th>Mining</th>
<th>Circulant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rounds</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Time (s)</td>
<td>7</td>
<td>35</td>
</tr>
<tr>
<td>AP</td>
<td>0.749</td>
<td>0.805</td>
</tr>
<tr>
<td>Time (s)</td>
<td>12</td>
<td>139</td>
</tr>
<tr>
<td>AP</td>
<td>0.165</td>
<td>0.380</td>
</tr>
</tbody>
</table>

$\sim14x$ speed-up
Experiments

Single-template HOG object detection:

<table>
<thead>
<tr>
<th></th>
<th>Mining</th>
<th>Circulant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rounds</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Time</td>
<td>INRIA</td>
<td>7</td>
</tr>
<tr>
<td>(s)</td>
<td></td>
<td>0.749</td>
</tr>
<tr>
<td>AP</td>
<td>Caltech</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.165</td>
</tr>
</tbody>
</table>

~14x speed-up
Experiments

Single-template HOG object detection:

- Swans
- Applelogos
- Bottles
- Giraffes
- Mugs

ETHZ Shapes
Conclusions

- Hard negative mining can be replaced with non-iterative training.
- There is a rich intrinsic structure in the problem.

Circulant Decomposition:
- Closed-form
- Parallel, small sub-problems
- Off-the-shelf SVR solvers
- 12 lines of MATLAB code

⇒
- ~14x speed-up
- Same/better performance

Theoretic development:
- Link between general learning algorithms and specialized Fourier signal processing.