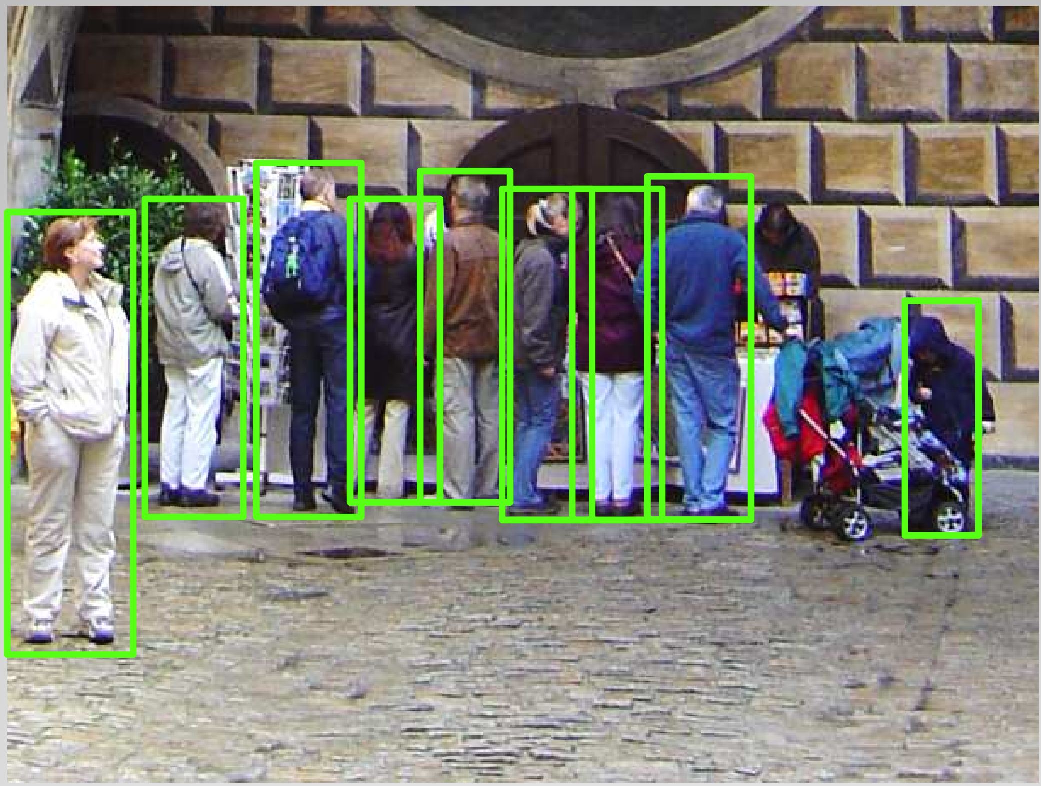
Beyond Hard Negative Mining:

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

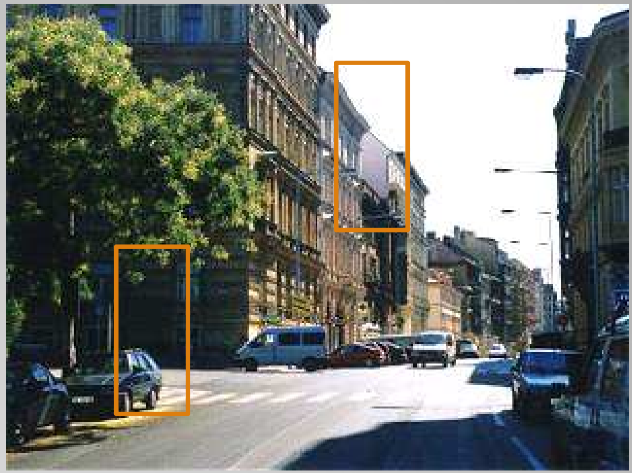
High performance usually requires
Hard Negative Mining.

⋮

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



⋮

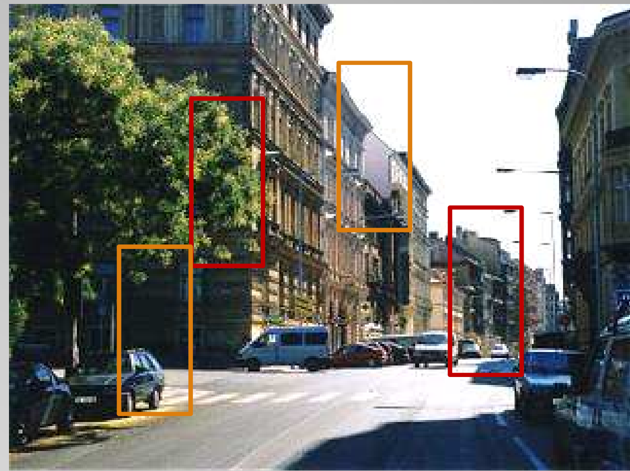
Motivation

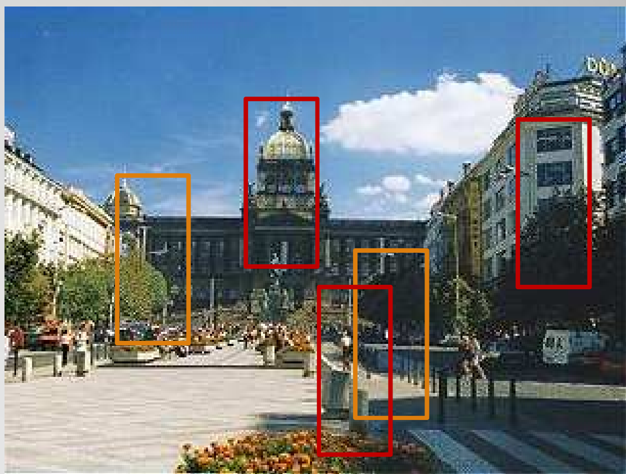
High performance usually requires
Hard Negative Mining.

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





⋮

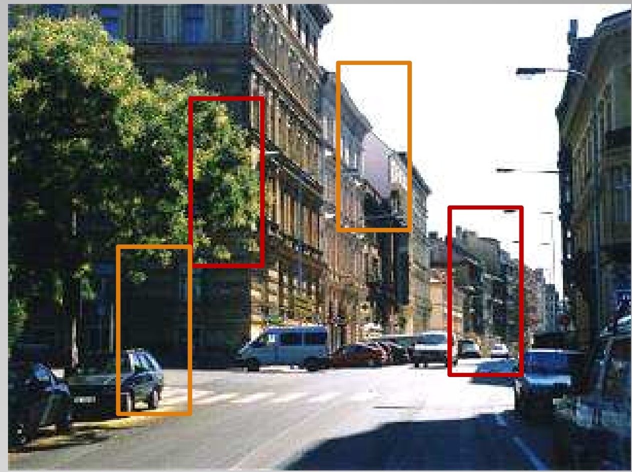
Motivation

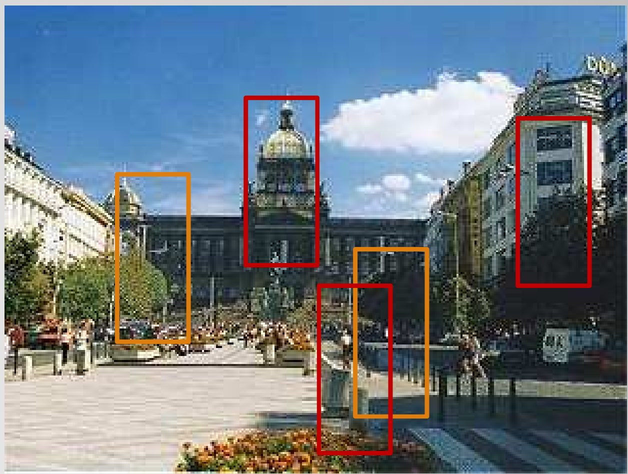
High performance usually requires
Hard Negative Mining.

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

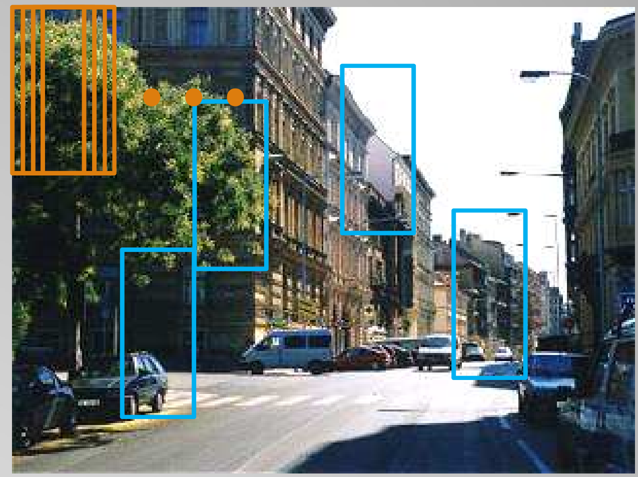
- Consider the **full set** of all potential samples.

⋮

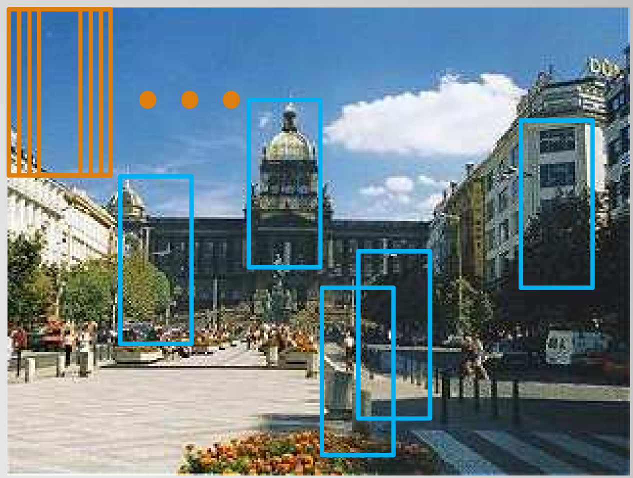
Motivation

- 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

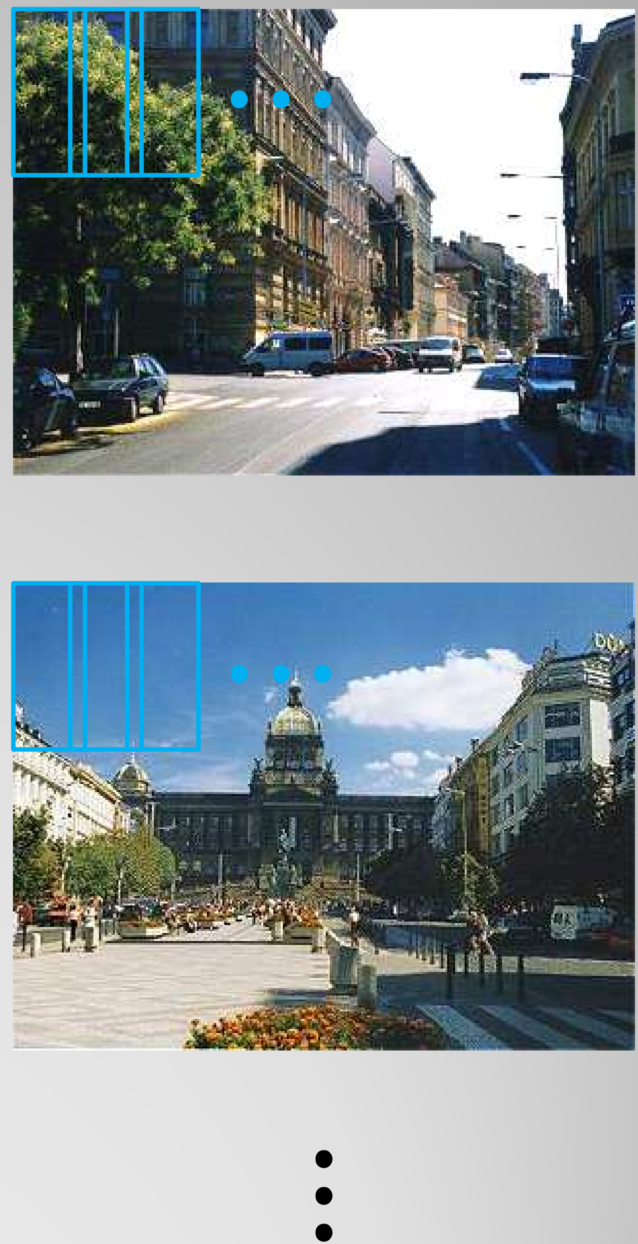
Observation:

- 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”



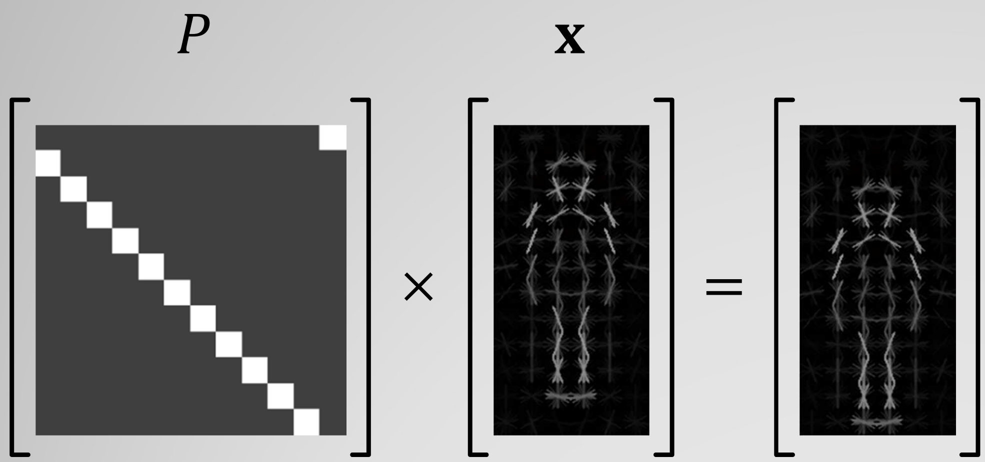
Let’s try to train with the **full negative set**.

Method:

- Collect base samples in a **coarse** grid.
- Train with the finer translations **implicitly** by using a **Circulant Decomposition**.

Cyclic shifts

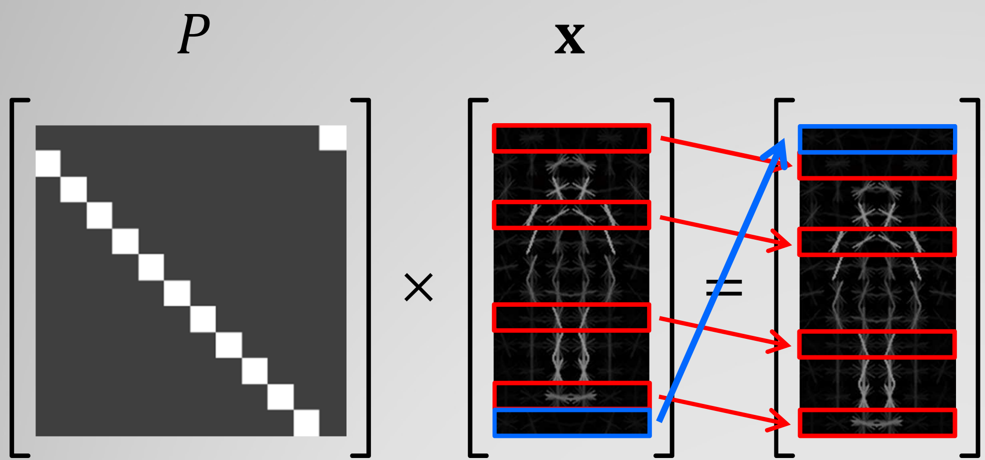
- We need a **model of image translations**.
- Idea: Apply permutation matrix P to base sample $\mathbf{x}$:



P represents a **cyclic shift**.

Cyclic shifts

- We need a **model of image translations**.
- Idea: Apply permutation matrix P to base sample $\mathbf{x}$:

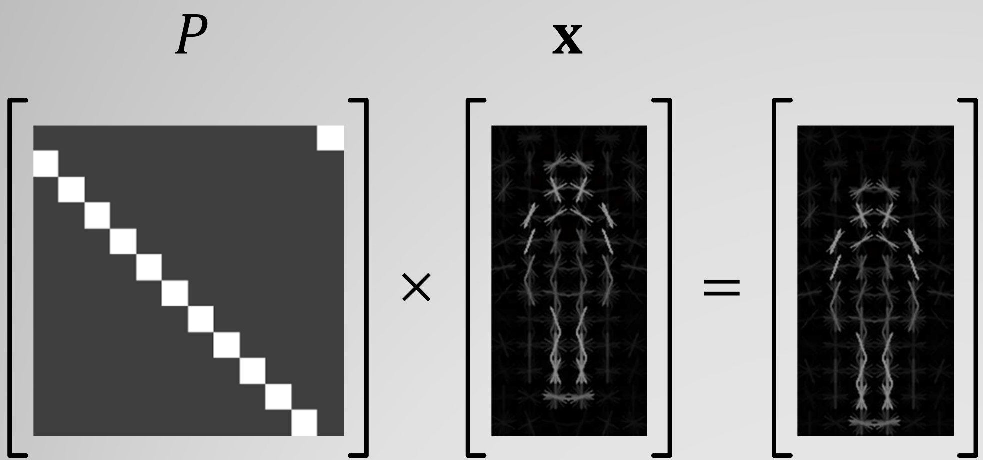


P represents a **cyclic shift**.

Cyclic shifts

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

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



P represents a cyclic shift.

- Powers of P shift by different amounts:

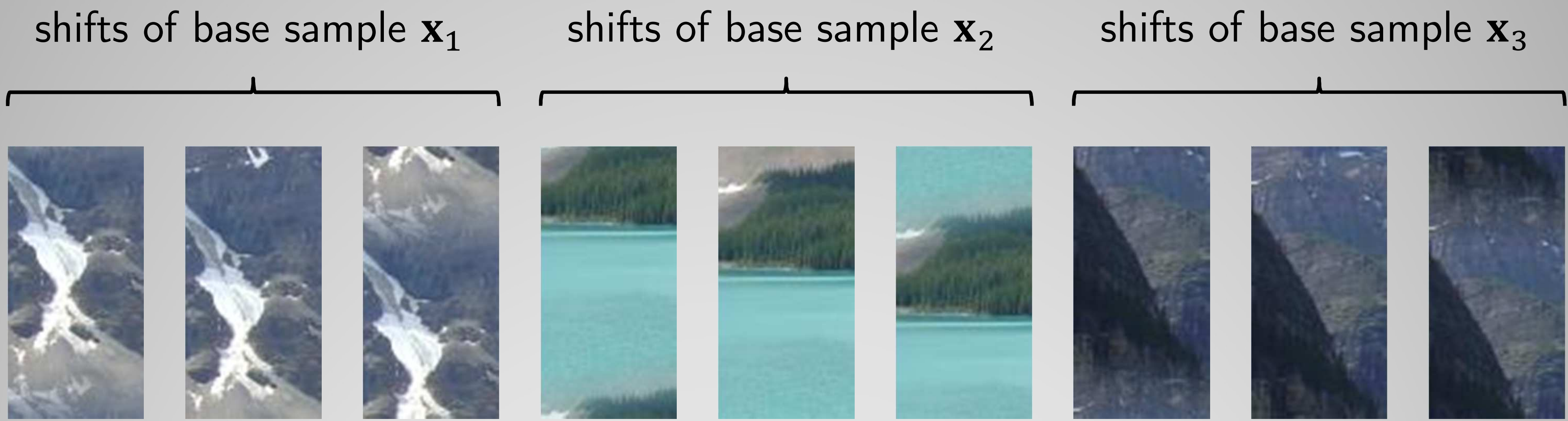
$$P^u\mathbf{x},\ u \in \left\{-\frac{\text{height}}{2}, \cdots, +\frac{\text{height}}{2}\right\}$$

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

(Easy to generalize to horizontal + vertical)

Cyclic shifts

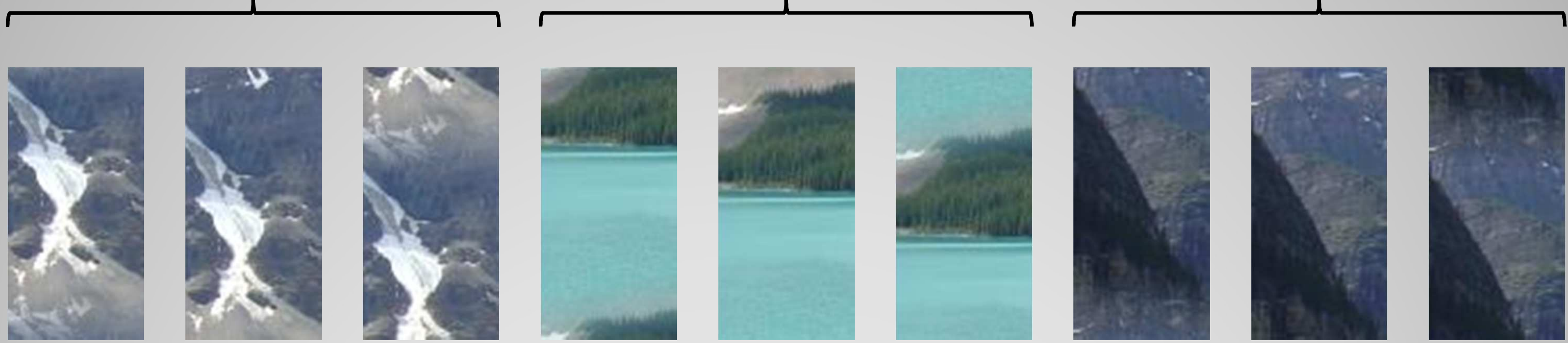
- Goal: implicitly train with **all shifts** of **all base samples**.



Cyclic shifts

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

shifts of base sample $\mathbf{x}_1$ | shifts of base sample $\mathbf{x}_2$ | shifts of base sample $\mathbf{x}_3$



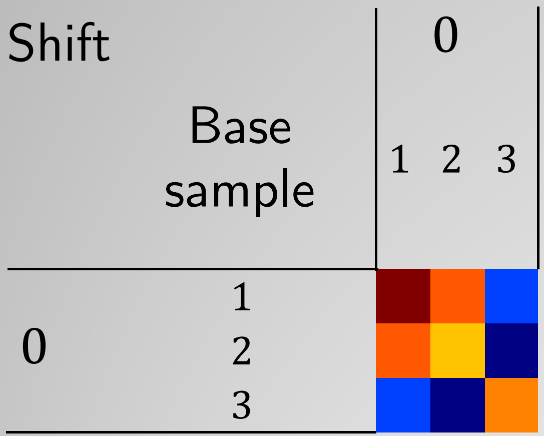
- To see how shifted samples interact, analyze the **Gram matrix**:

$$G_{ij} = \langle \mathbf{x}_i, \mathbf{x}_j \rangle$$

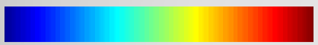
(Dot-products between **pairs** of samples)

Gram matrix

Dataset



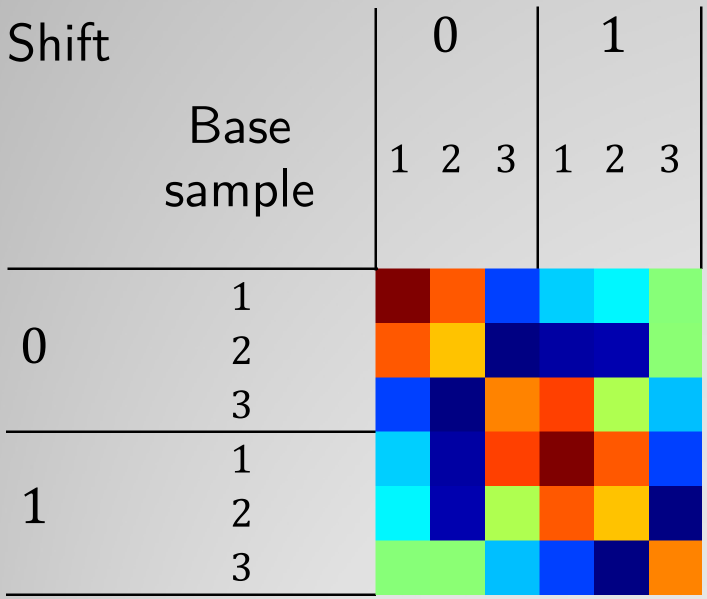
G



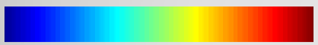
dot-product

Gram matrix

Dataset



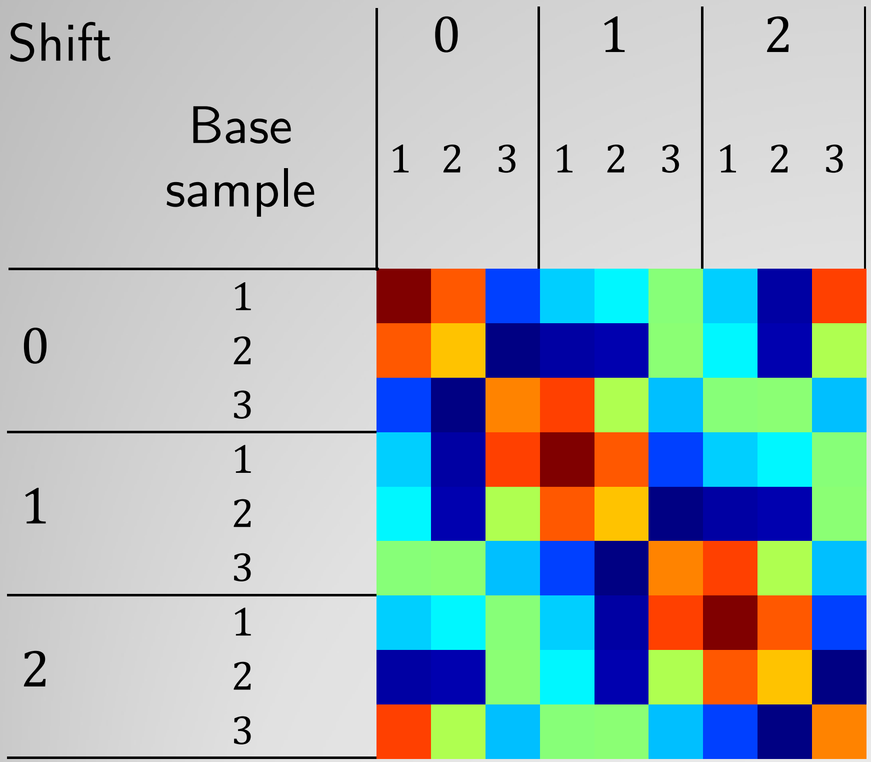
G

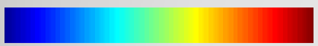


dot-product

Gram matrix

Dataset



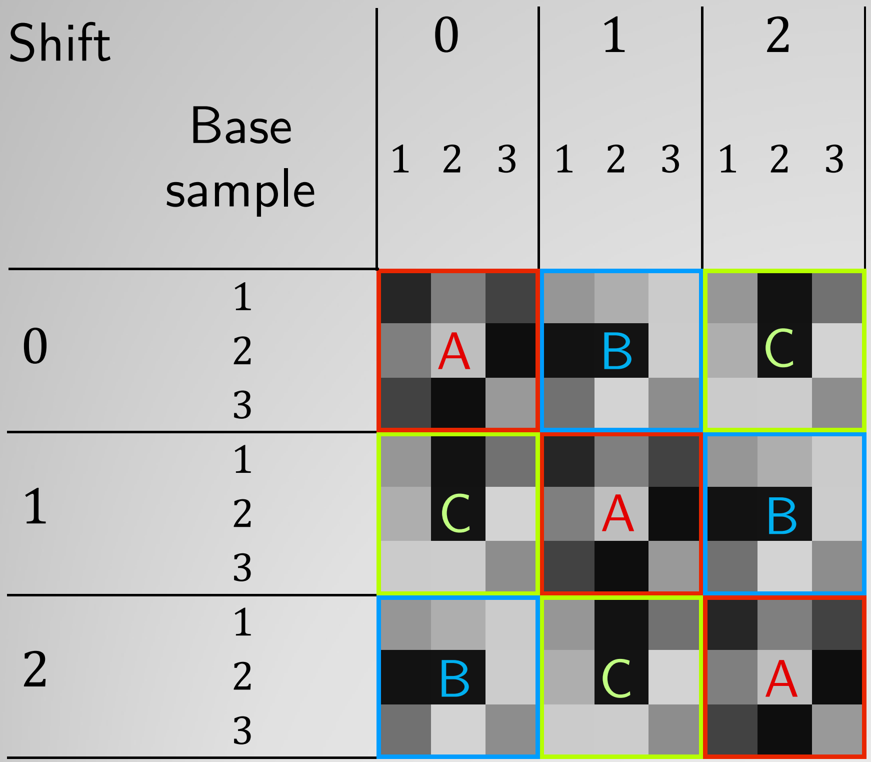


dot-product

G

Gram matrix

Dataset



G

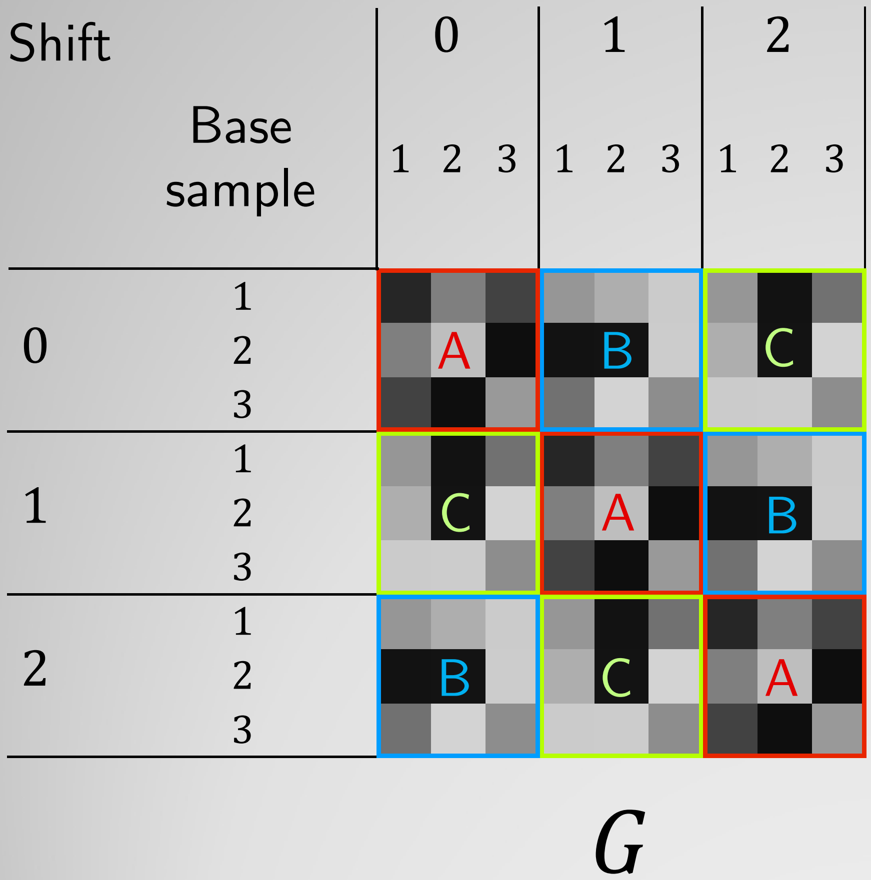
- *Property #1:*

G is **block-circulant**

⇒ Only 1 row of blocks is unique.

Gram matrix

Dataset



- *Property #1:*

  G is block-circulant

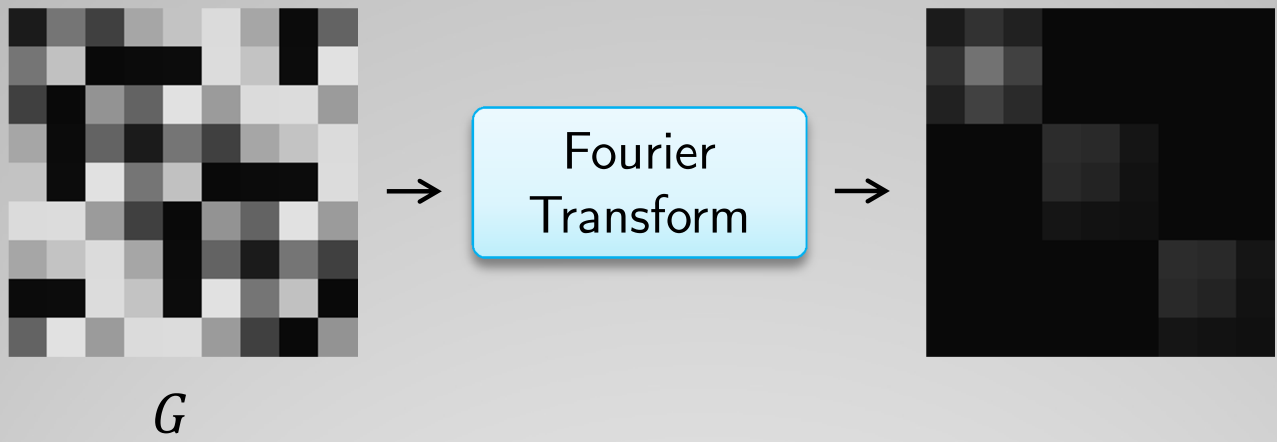
  ⇒ Only 1 row of blocks is unique.

- *Property #2:*

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

  ⇒ Becomes simple product in the Fourier domain.

Block-diagonalization

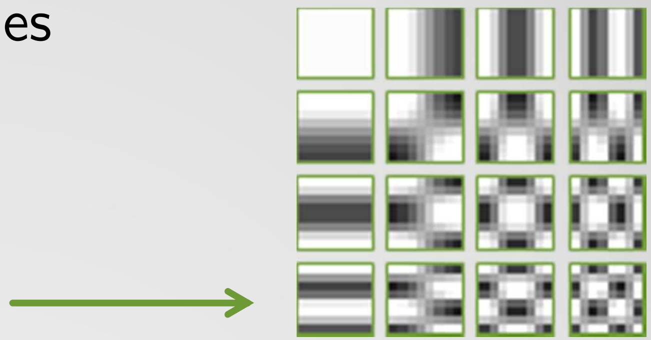


Proposed approach:

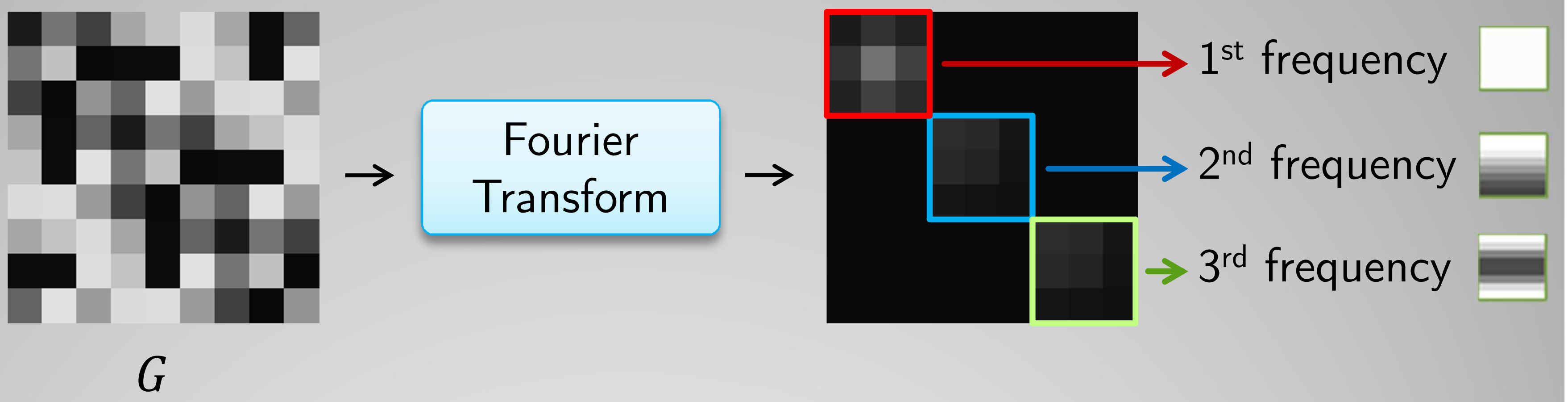
- **Fourier Transform** the samples (+ a small permutation)

$$\Leftrightarrow$$

Projection on **Fourier Basis** with different frequencies.



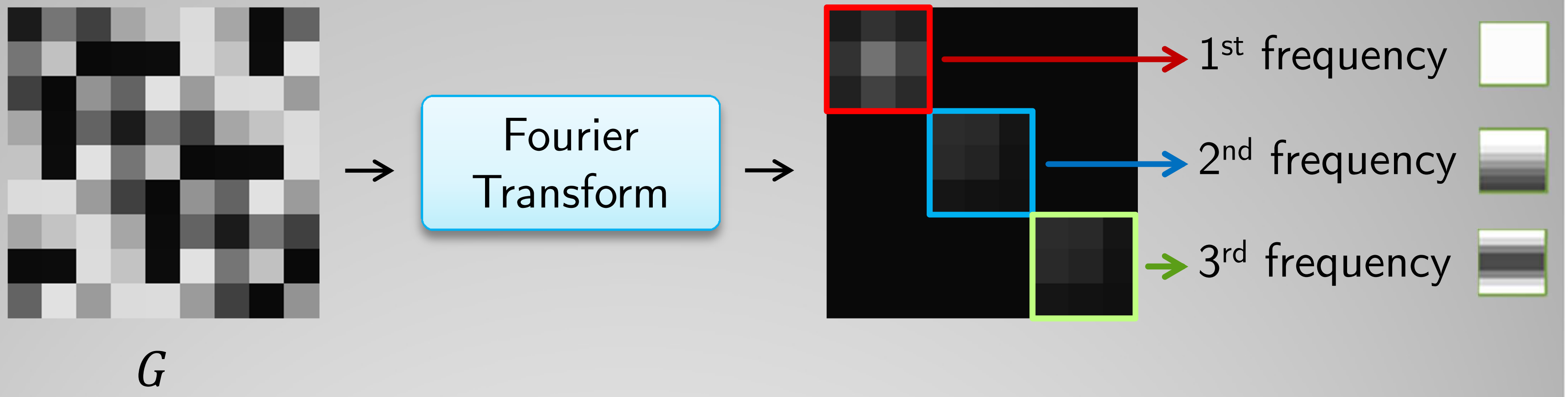
Block-diagonalization



- **Each block** of G contains the projection on a different basis, or **Fourier frequency**.

(# of frequencies $=$ # of spatial cells of the samples)

Block-diagonalization

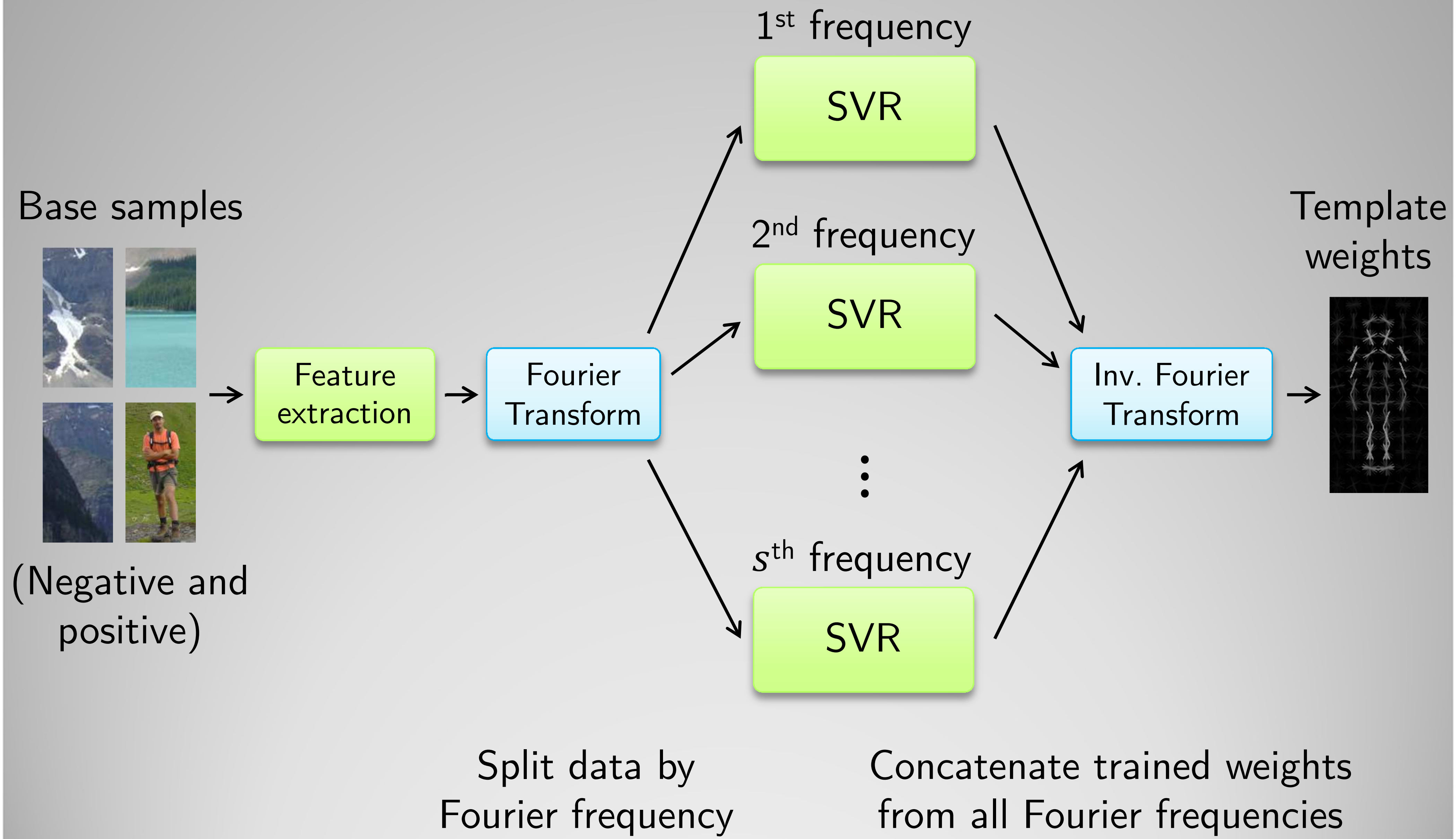


We prove all off-diagonal blocks are **zero**. → **99.5%** for 18x10 HOG template

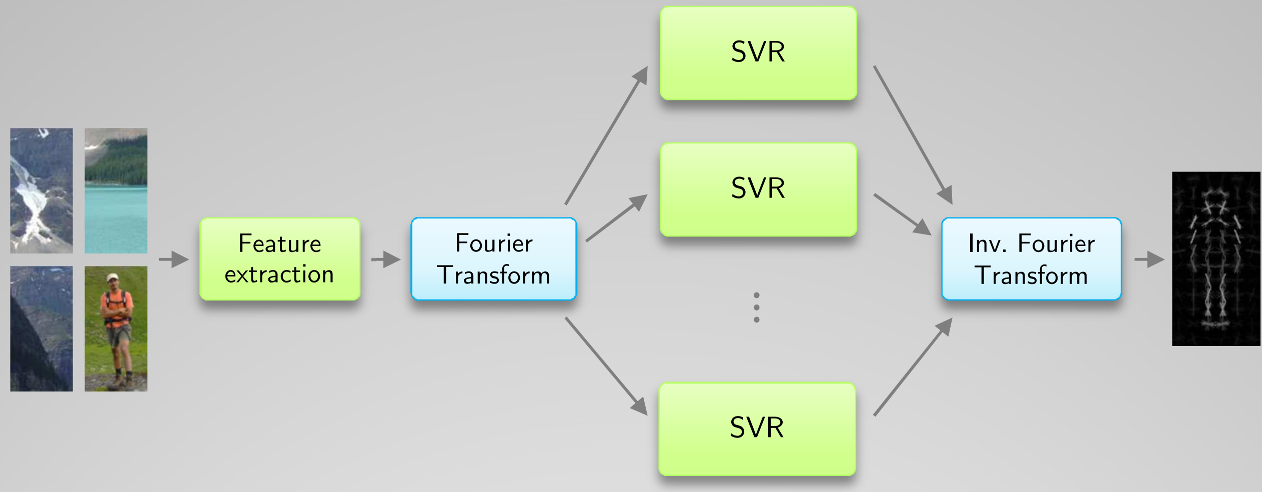
⇓

Frequencies correspond to **independent** learning problems.

Circulant Decomposition



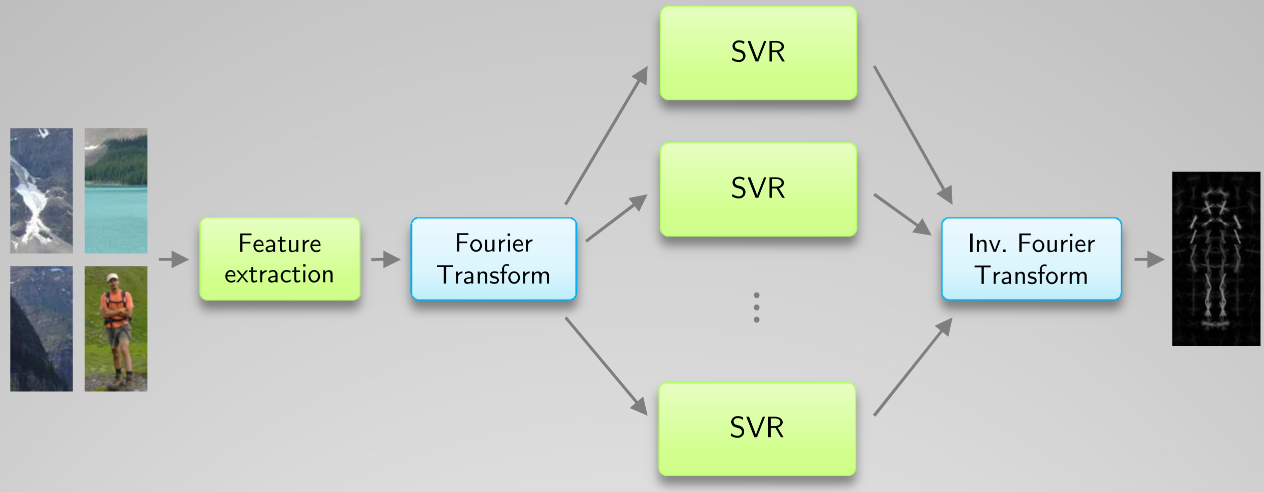
Circulant Decomposition



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

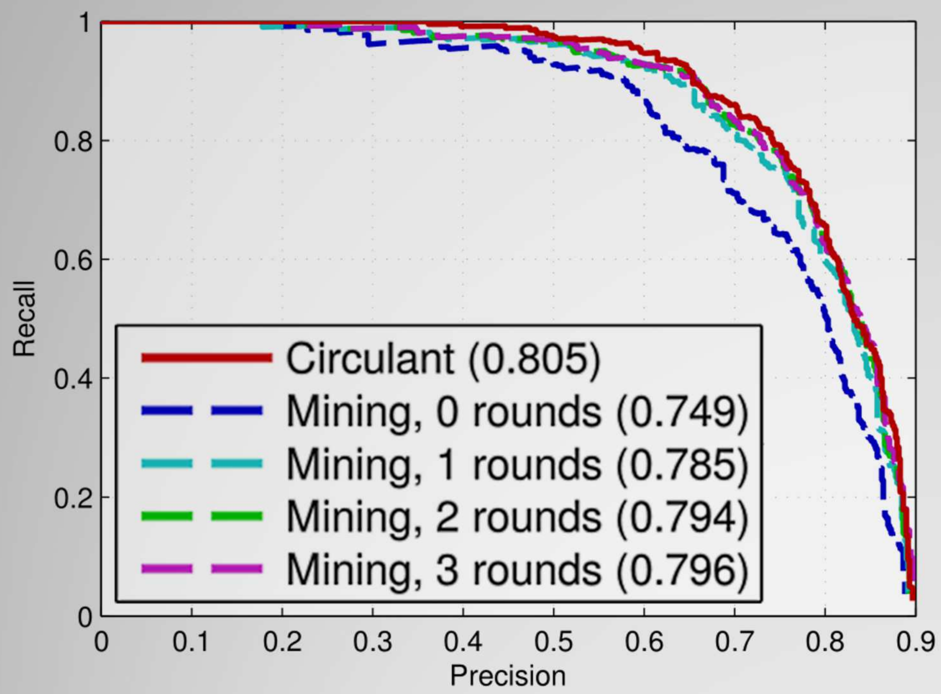


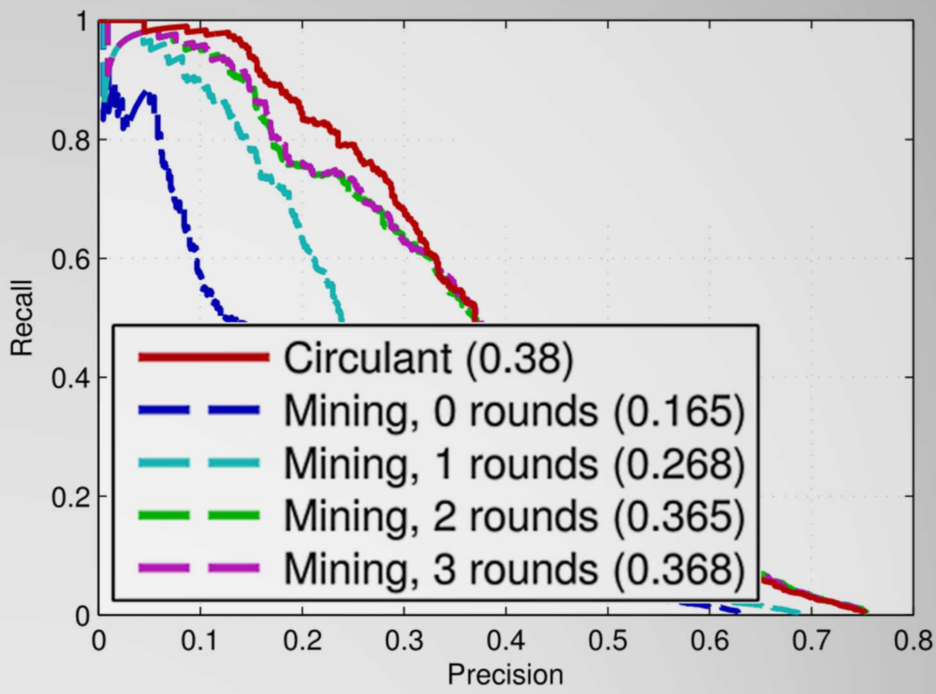
Circulant Decomposition

- **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

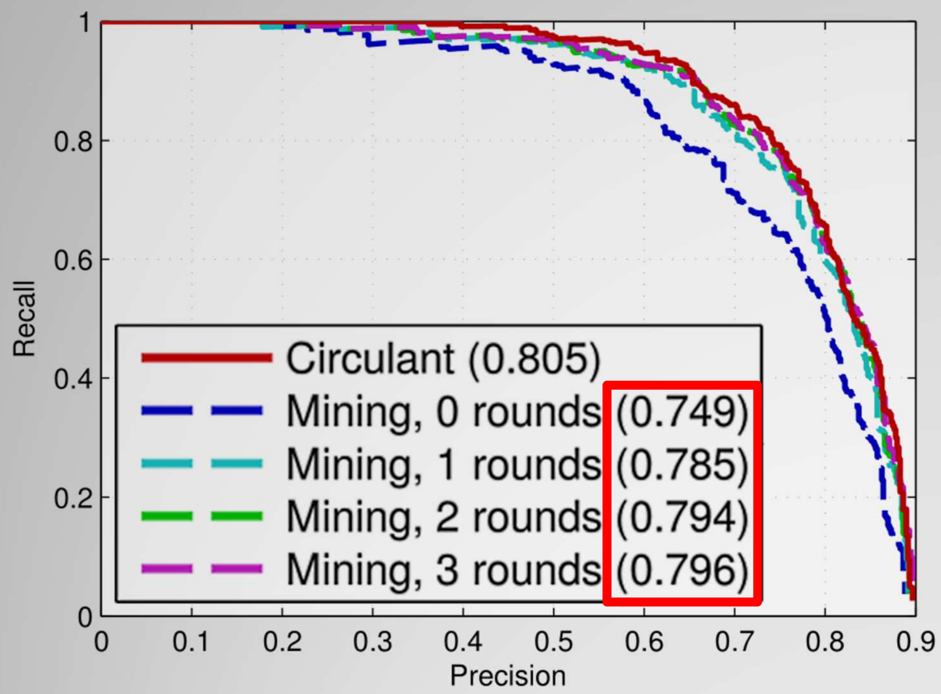
- 1218 negative images
- $\sim 10^8$ potential samples

Caltech Pedestrians

- 4250 negative images
- $\sim 10^8$ potential samples

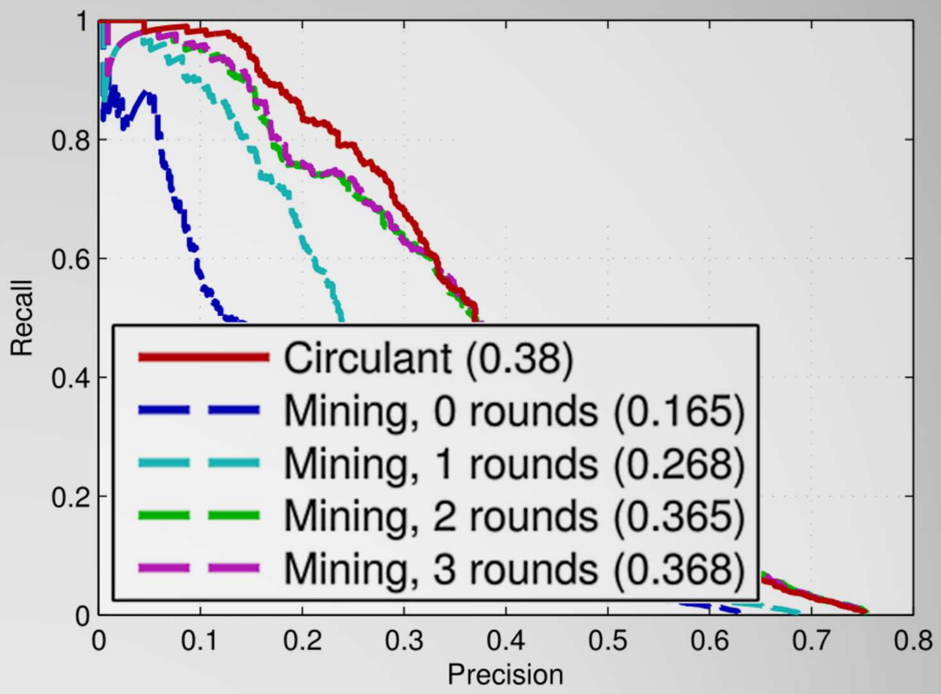
Experiments

Single-template HOG object detection:



INRIA Pedestrians

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

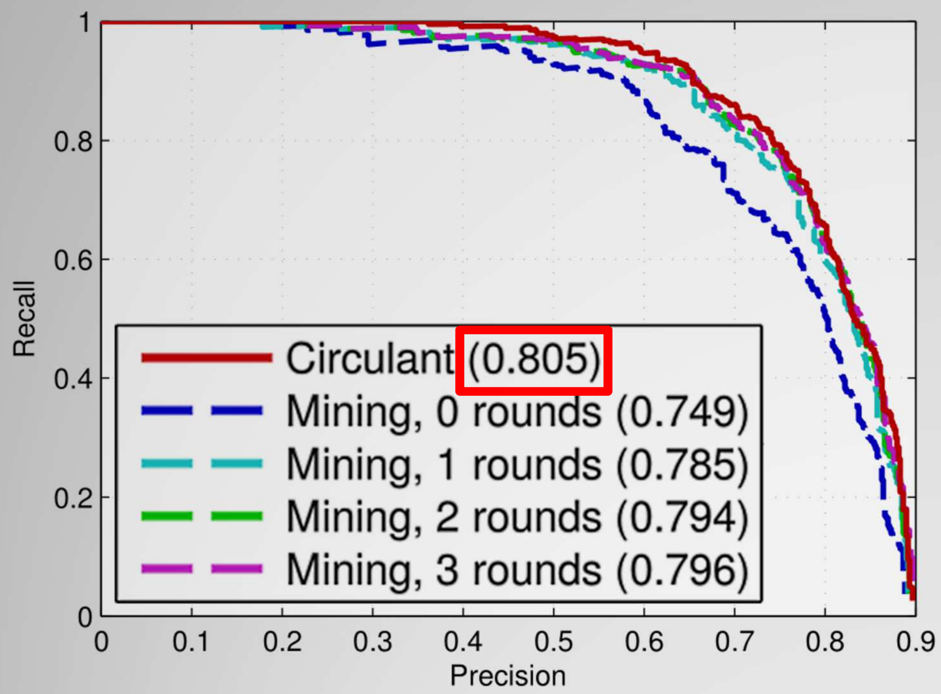


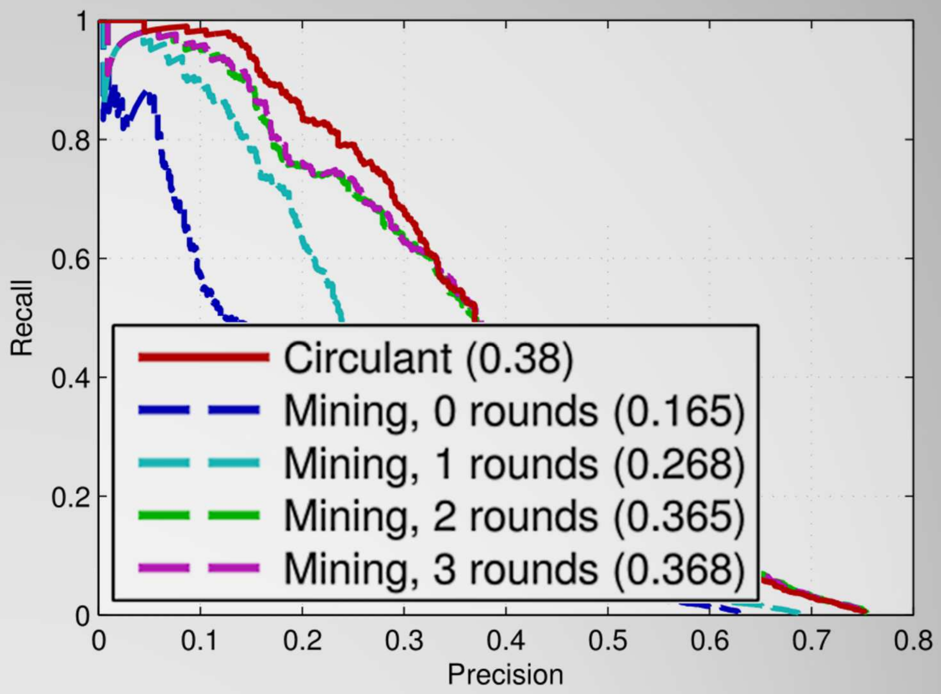
Caltech Pedestrians

- 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:

		Mining				**Circulant**
Rounds		0	1	2	3	**0**
Time (s)	INRIA	7	159	312	463	**35**
AP	INRIA	0.749	0.785	0.794	0.796	**0.805**
Time (s)	Caltech	12	646	1272	1901	**139**
AP	Caltech	0.165	0.268	0.365	0.368	**0.380**

~14x speed-up

Experiments

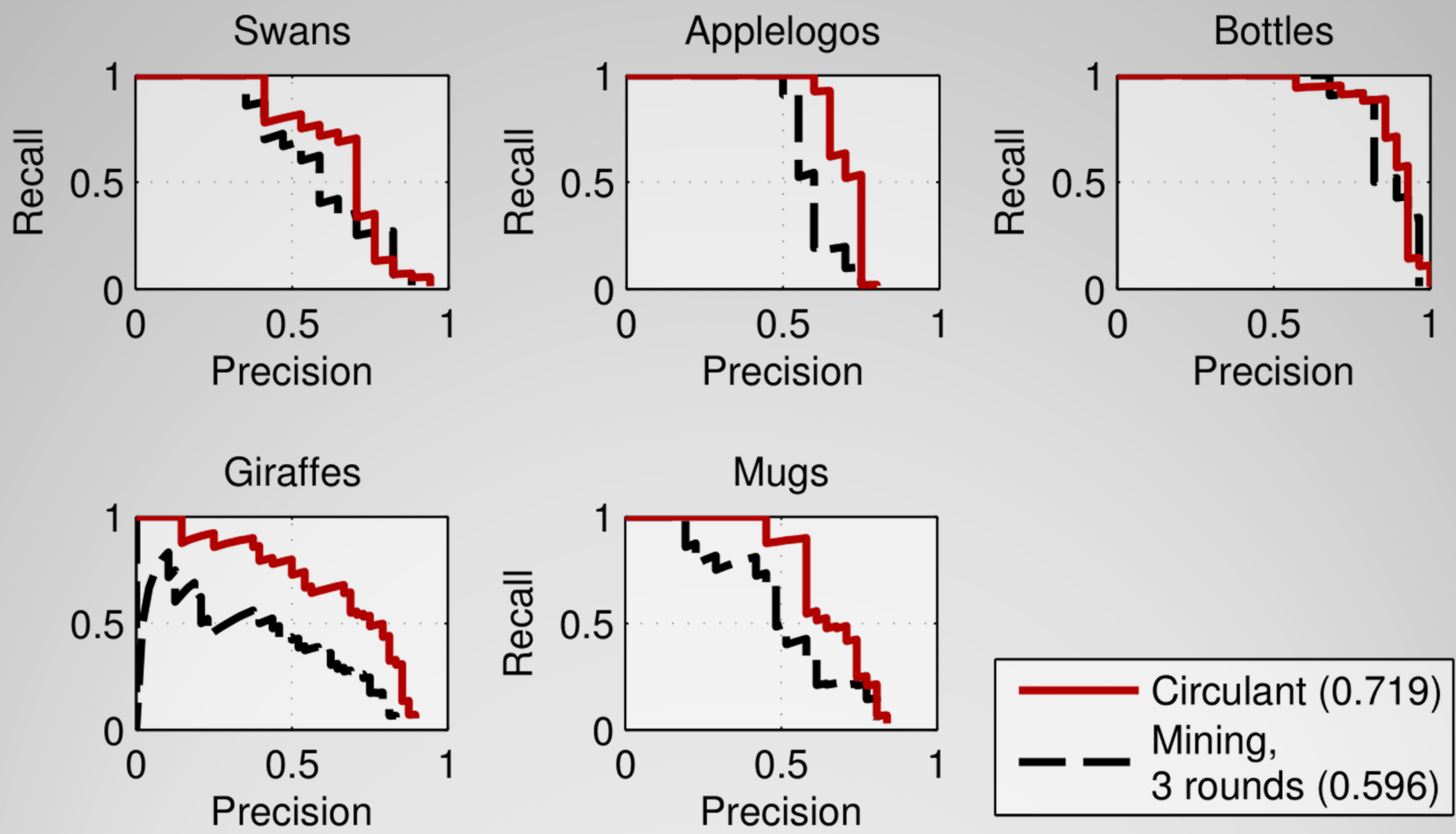
Single-template HOG object detection:

		Mining				**Circulant**
Rounds		0	1	2	3	**0**
Time (s)	INRIA	7	159	312	463	**35**
AP	INRIA	0.749	0.785	0.794	0.796	**0.805**
Time (s)	Caltech	12	646	1272	1901	**139**
AP	Caltech	0.165	0.268	0.365	0.368	**0.380**

~14x speed-up

Experiments

Single-template HOG object detection:



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

$\Rightarrow$
- **~14x speed-up**
- Same/better performance

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