Joint Optimization of Cascaded Classifiers for Computer Aided Detection

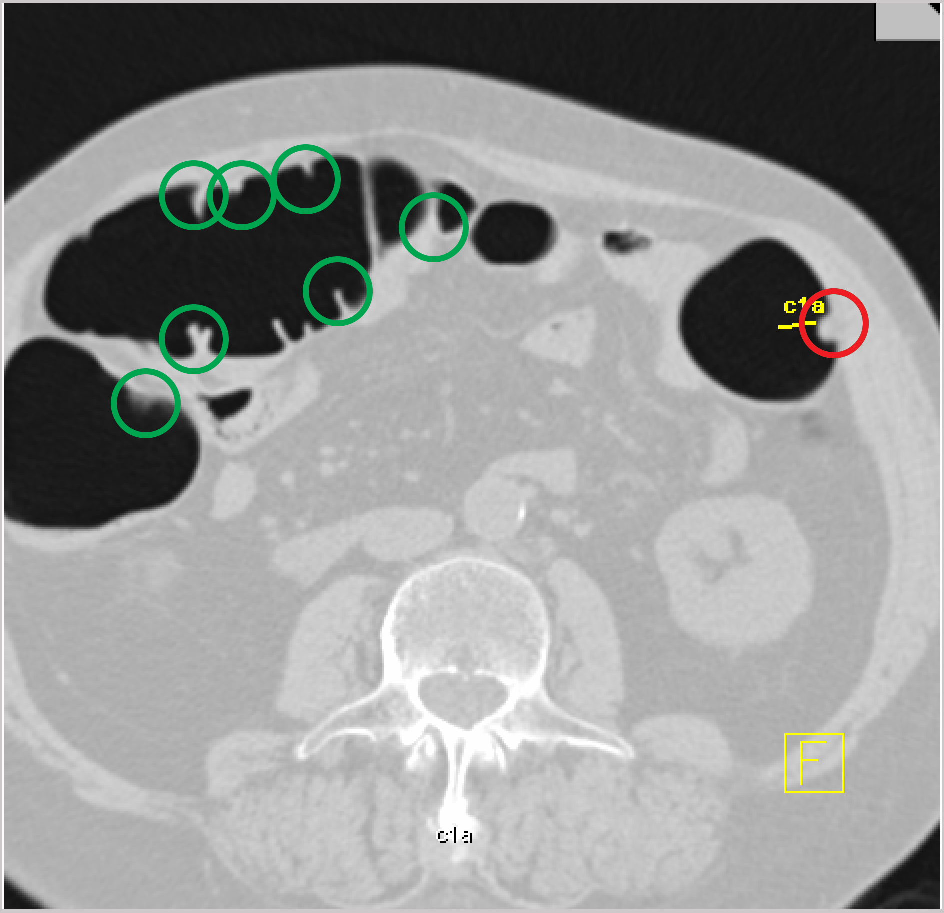
M. Murat Dundar and Jinbo Bi

Siemens Medical Solutions, Inc. USA
Malvern, PA 19355

SIEMENS

An Example of a CAD problem

- Identify suspicious regions (candidates)
- Extract features for each candidate
- Classify candidates as lesion or non-lesion

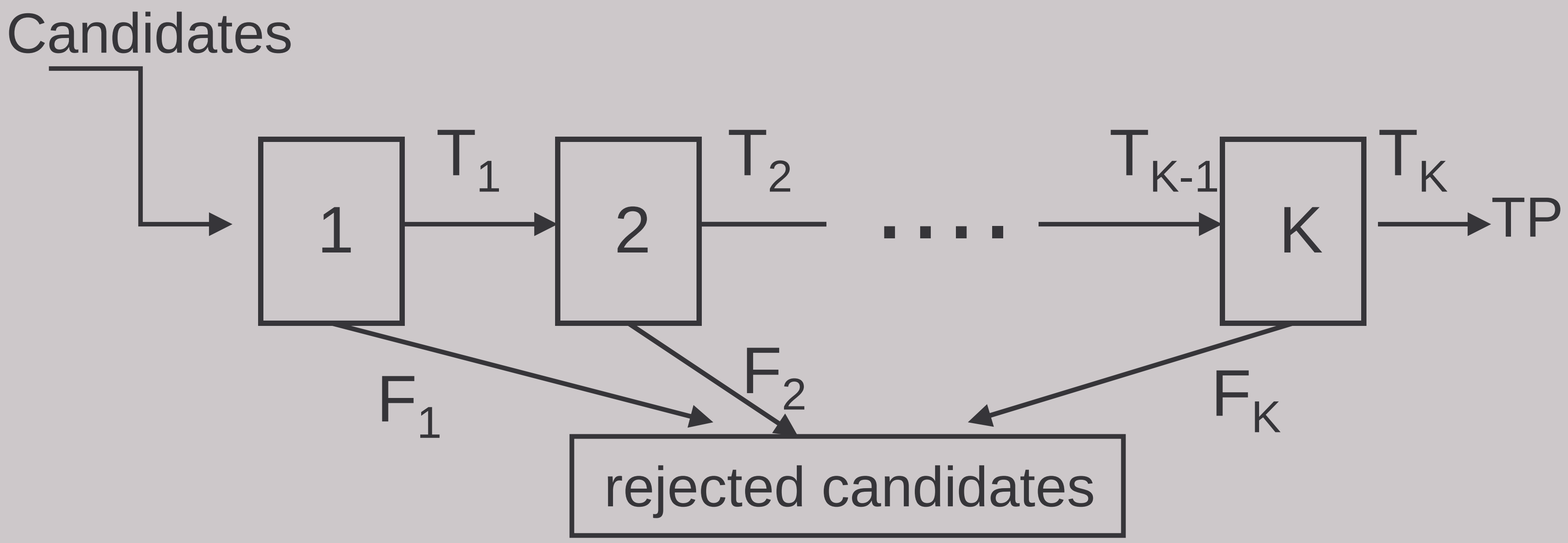


Challenges in Classification

- Extremely unbalanced between positive and negative classes

- Image features with different computational costs
  - Some cheap
  - Some very expensive

Traditional Cascaded Classifier



Training Sets: $S_1 \supset S_2 \supset \ldots\ldots \supset S_K$

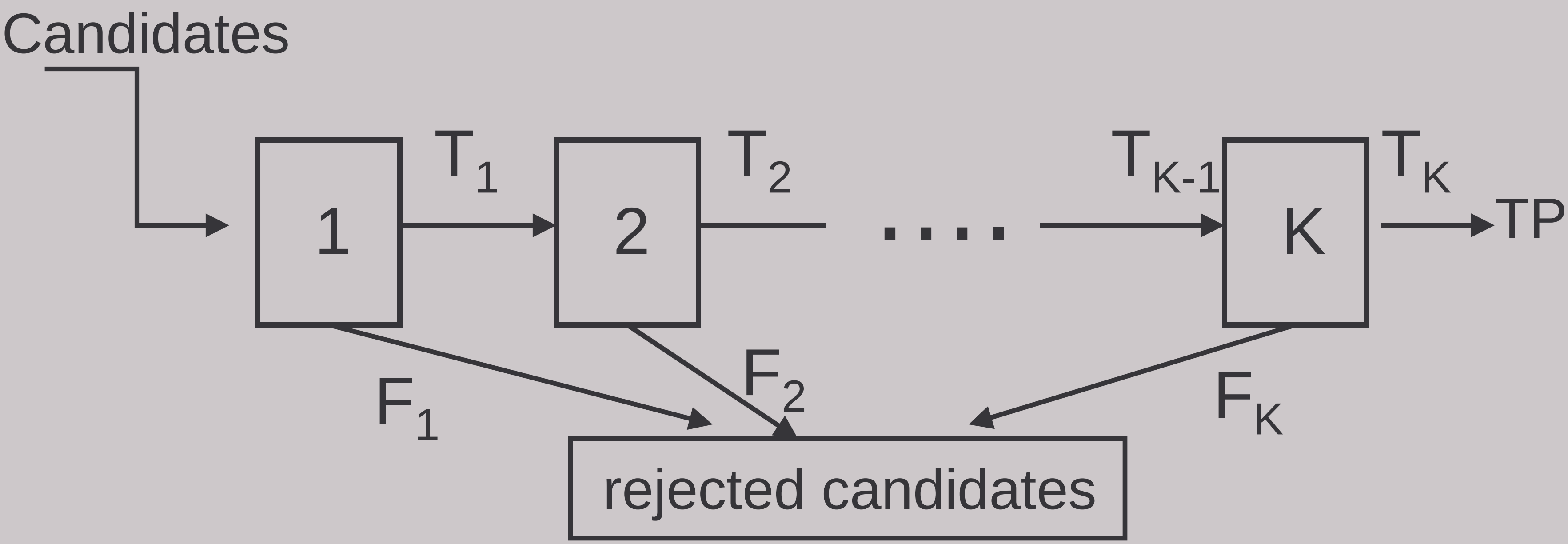
Some Issues

Existing Cascade Schemes

- The cascade is trained sequentially. Each classifier is optimized for the specific local stage.
- How to decide sensitivity each classifier should achieve
- It does not consider computational costs for features

SIEMENS

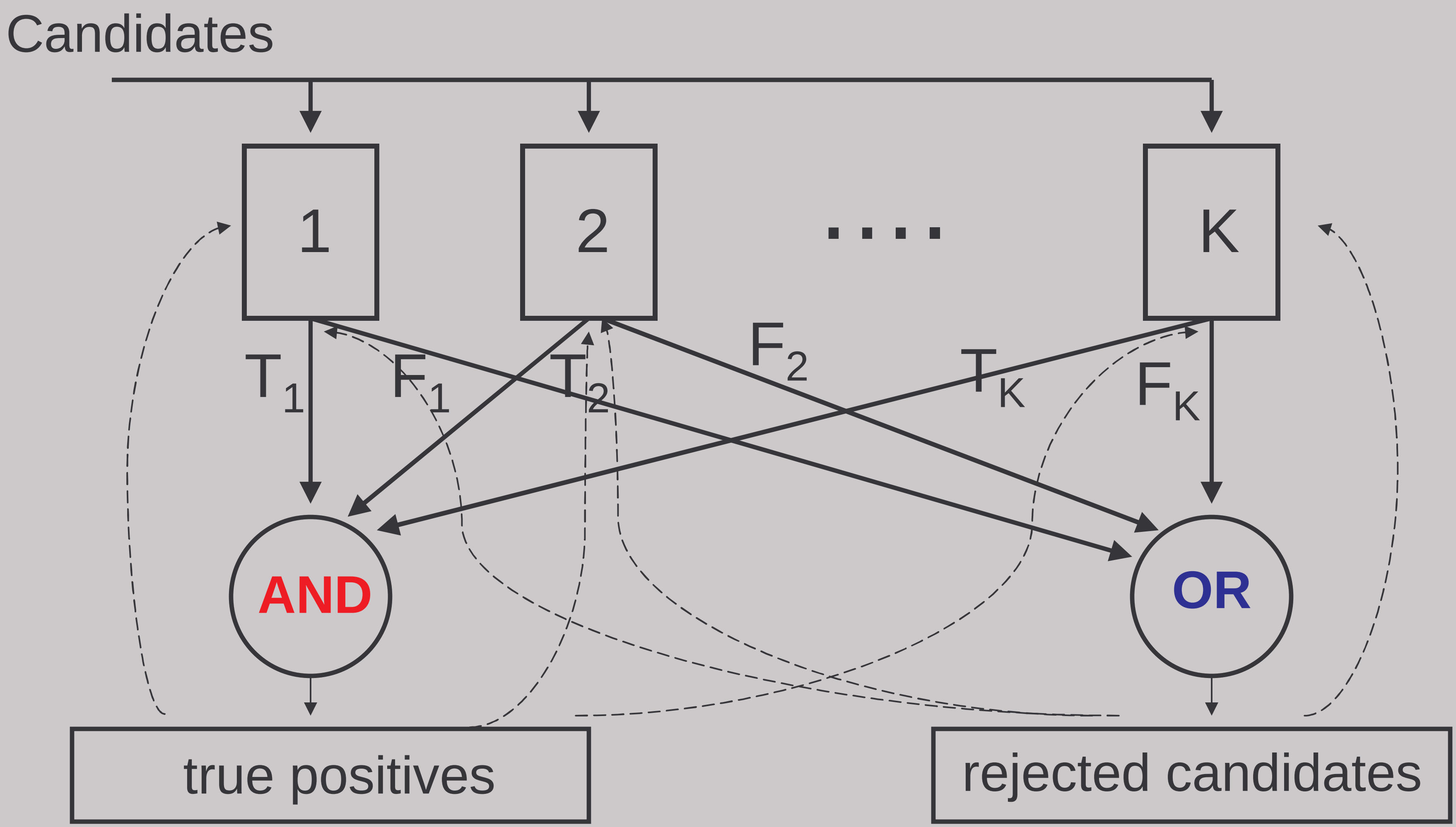
An AND-OR Framework for Offline Training



Positive example: all classifiers have to say “yes” -- “AND”

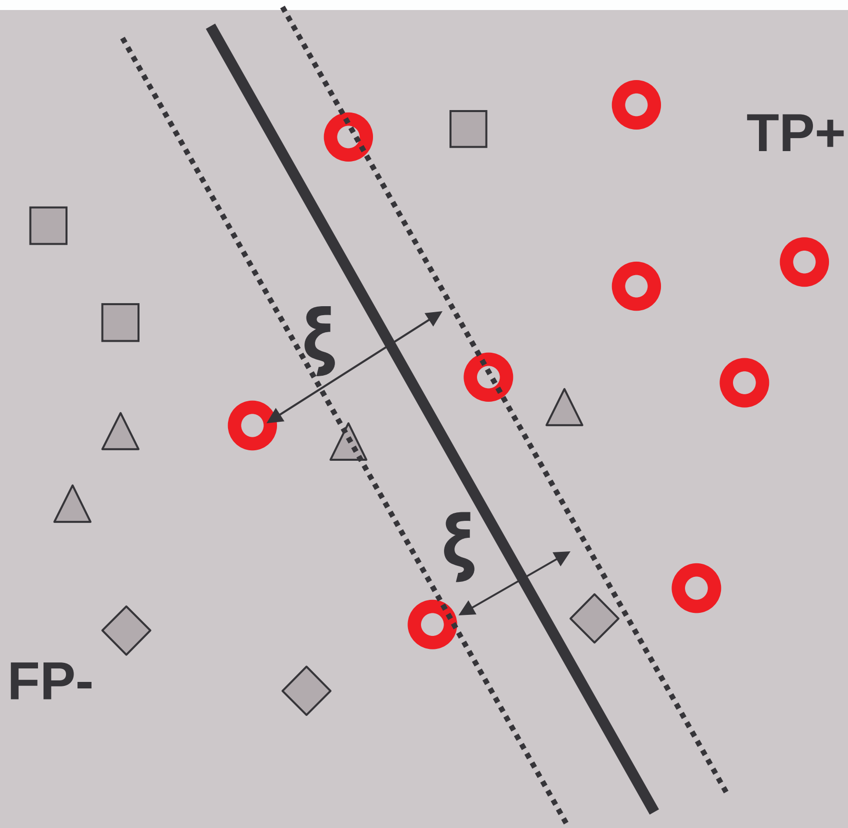
Negative example: any classifier says “no” – “OR”

An AND-OR Framework for Offline Training



Hinge Loss

$$\alpha^T x = 0$$



$$\xi_i = \max\{0,\ 1 - y_i \alpha^T x_i\}$$

SIEMENS

AND-OR Cascade Implementation

If the hinge loss = 0, the example is correctly classified,
If the hinge loss > 0, the example is mis-classified

Let ξ_{ik} be the hinge loss of i-th example induced by the classifier k

i-th Positive example: $\max(\xi_{i1}, \xi_{i2}, \cdots \xi_{iK})$ -- "AND"

i-th Negative example: $\prod_{k=1}^{K} \xi_{ik}$ – "OR"

SIEMENS

Objective Function in the AND-OR Framework

$$J(\alpha_1, \alpha_2, \cdots \alpha_K) = \nu_1 \sum_{i \in C-} \prod_{k=1}^{K} \xi_{ik}$$

Error on Negative Examples

$$+ \quad \nu_2 \sum_{i \in C+} \max(\xi_{i1}, \xi_{i2}, \cdots \xi_{iK})$$

Error on Positive Examples

$$+ \quad \sum_{k=1}^{K} P(\alpha_1)$$

Regularization to Control Complexity

Alternating Optimization Iterative Algorithm

Each iteration contains K steps, and each step optimizes a single classifier

At the k-th step,

Fix all classifiers (*α's*) but the classifier *k*

Minimize $J(\alpha_1,\ldots, \alpha_k ,\ldots \alpha_K)$ for optimal α_k

SIEMENS

Objective Function in the Sub-problem

$$J(\alpha_k) = \nu_1 \sum_{i \in C-} w_{ik} \xi_{ik}$$

$$+ \nu_2 \sum_{i \in C+} \max(\gamma_{ik}, \xi_{ik})$$

$$+ \quad P(\alpha_k)$$

$$w_{ik} = \xi_{i1} \cdots \xi_{i,k-1} \xi_{i,k+1} \cdots \xi_{iK}$$

$$\gamma_{ik} = \max(\xi_{i1}, \cdots \xi_{i,k-1}, \xi_{i,k+1}, \cdots \xi_{iK})$$

$$\boxed{\begin{array}{c} + \nu_2 \sum_{i \in C+} \xi_{ik} \\ \xi_{ik} \le \gamma_{ik} \end{array}}$$

Convex and can be solved by existing linear program or quadratic program solvers

Experiments – Automatic Polyp Detection

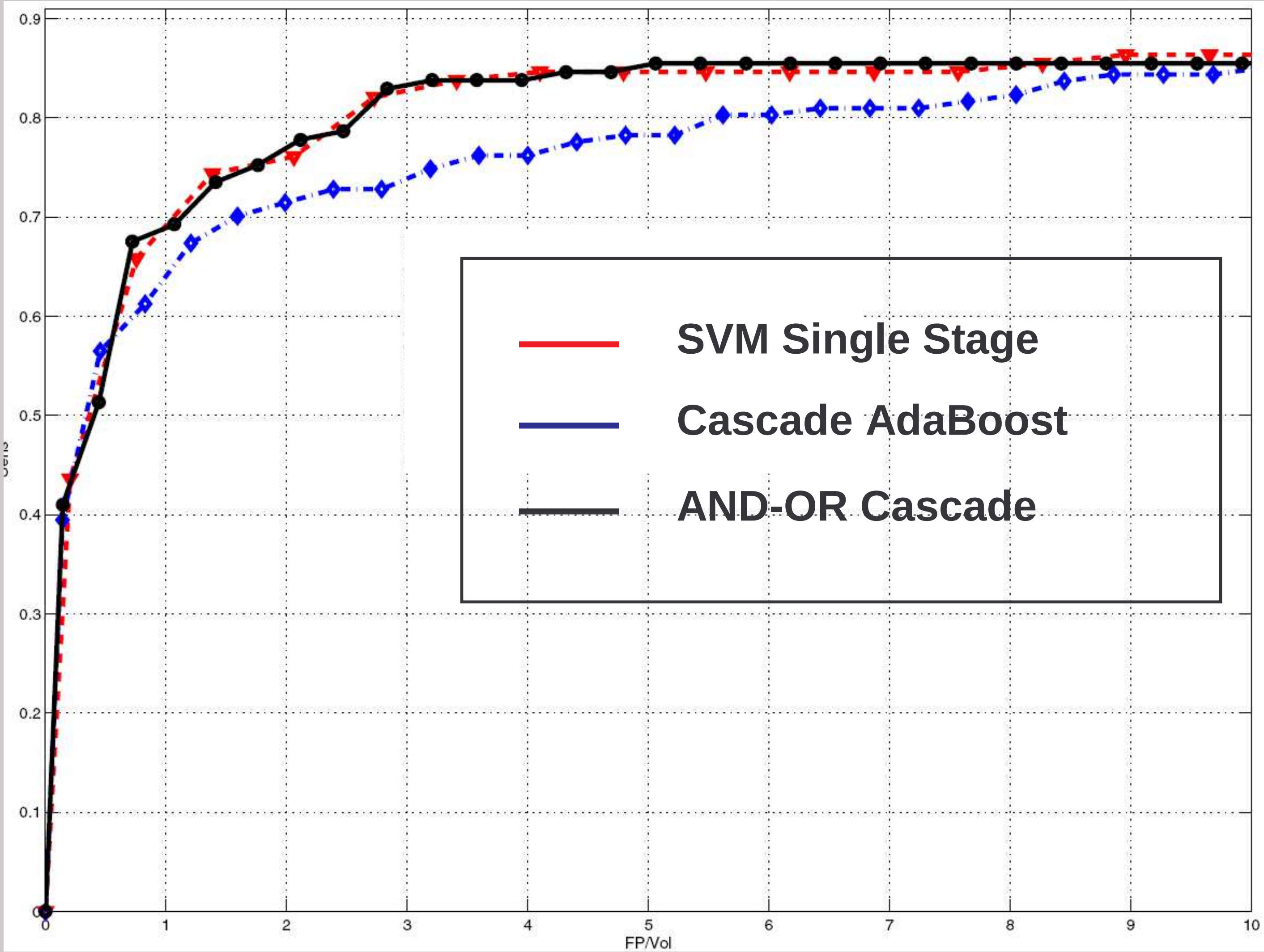
Data

	Volumes	Polyps	Candidates	FP/vol
Training	338	88	46,764	137.7
Test	396	106	55,497	139.4

46 numerical image features are computed,
and are in 4 groups of different computation costs

SIEMENS

ROC plots



Computational Cost

SIEMENS

	AND-OR Cascade			SVM	Cascade AdaBoost				
					Phase 1	Phase 2	Phase 3		
	C_1	C_2	C_3		C_1	C_2	C_3	C_4	C_5
Avg. CPU time per volume (sec.)	28	27.5	25.48	294.0	26.0	25.0		67.48	
number of polyps found (out of 106)	105	100	100	99	103	103	102	102	91
False positives per volume (initially 139.4)	55	18.2	5	5	50	48.2	30.8	20.6	5.0

Total Execution Time = 28 + 27.5 + 25.48 = 81 sec /volume

SIEMENS

Future Work

- This work studies an AND-OR offline training approach. It does not optimize the online test procedure. For example, how the resulting classifiers should be applied in the test phase.
- In the test phase, what is the optimal order to apply the obtained classifiers, such that the overall computational cost is minimum
- AND-OR design with other loss functions

Thank you! Questions and Comments