SPEEDING-UP A CONVOLUTIONAL NEURAL NETWORK BY CONNECTING AN SVM NETWORK

J. Pasquet[1,2], M. Chaumont[1,3], G. Subsol[1], M. Derras[2]

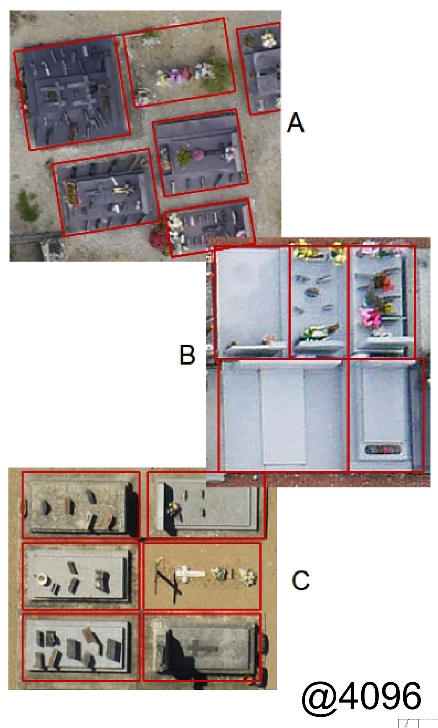
[1]LIRMM,University of Montpellier / CNRS, France
[2]Berger Levrault, Labège, France
[3]University of Nîmes, France

International Conference on Image Processing (ICIP) 2016

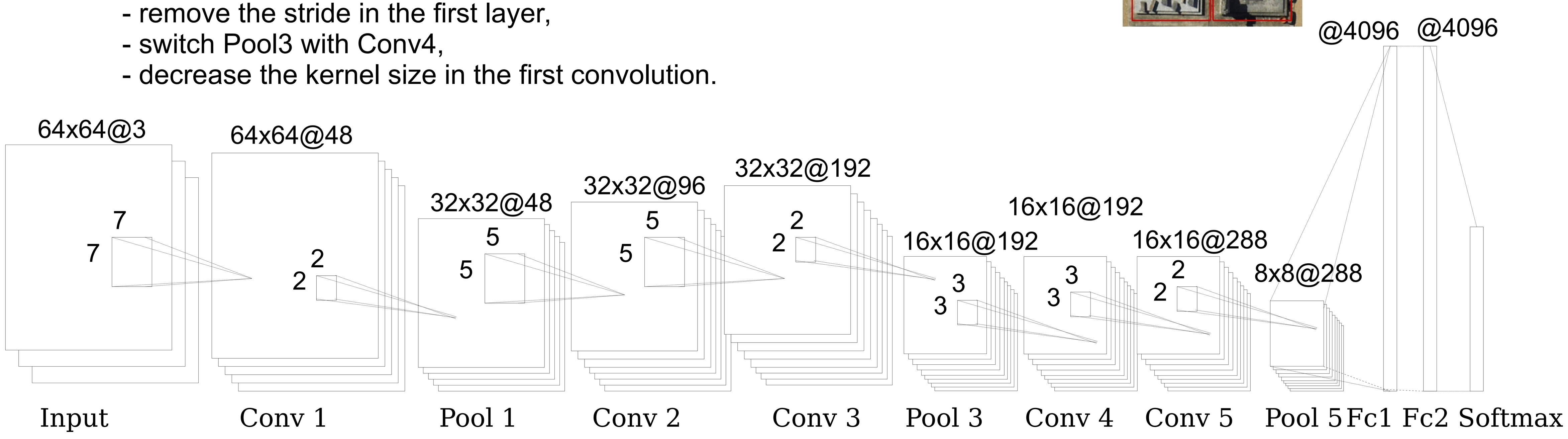
Context

- Detection of numerous and variable small urban objects in High-Definition aerial color images

- Applications : tombs in cemetery
  - Typical tombs size is 100x100 pixels
  - Tombs are very variable in shape and appearance
  - Image database: 24 images of 5,000 x 5,000 pixels (2.5 cm/pixel – 24 bits)



→ We adapt the AlexNet CNN [Krizhevsky et al., 2012] to process small images (64x64 pixels) :
- remove the stride in the first layer,
- switch Pool3 with Conv4,
- decrease the kernel size in the first convolution.



Problem during the testing step :

We use the sliding window process to analyse billions of positions

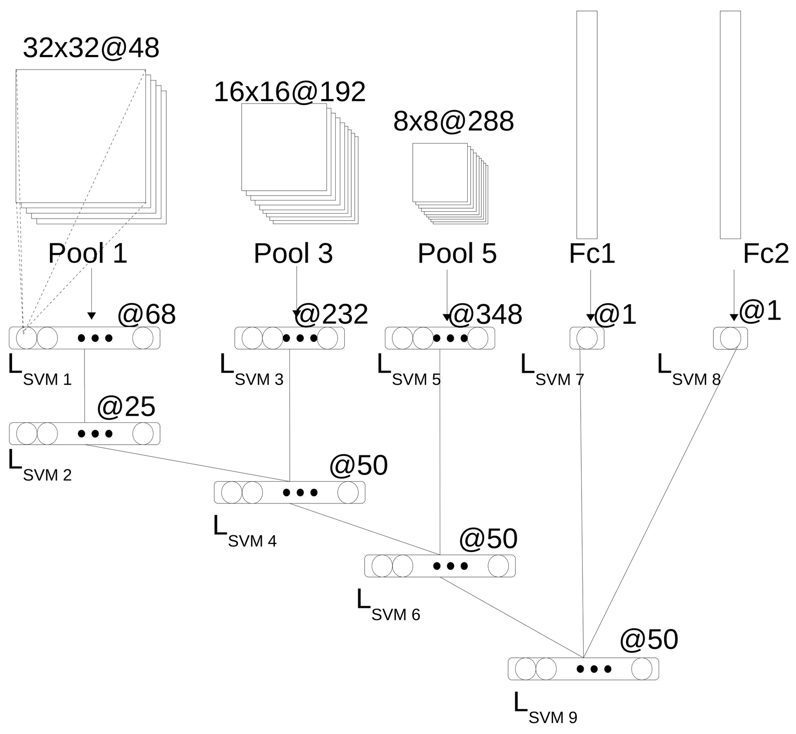
→ The computational cost is huge.

An approach to speed up a CNN

Connect an SVM network on the CNN in order to trigger early exits (as in cascade [Viola and Jones, 2001]) .

1. After the CNN training, we define the following SVM network architecture :

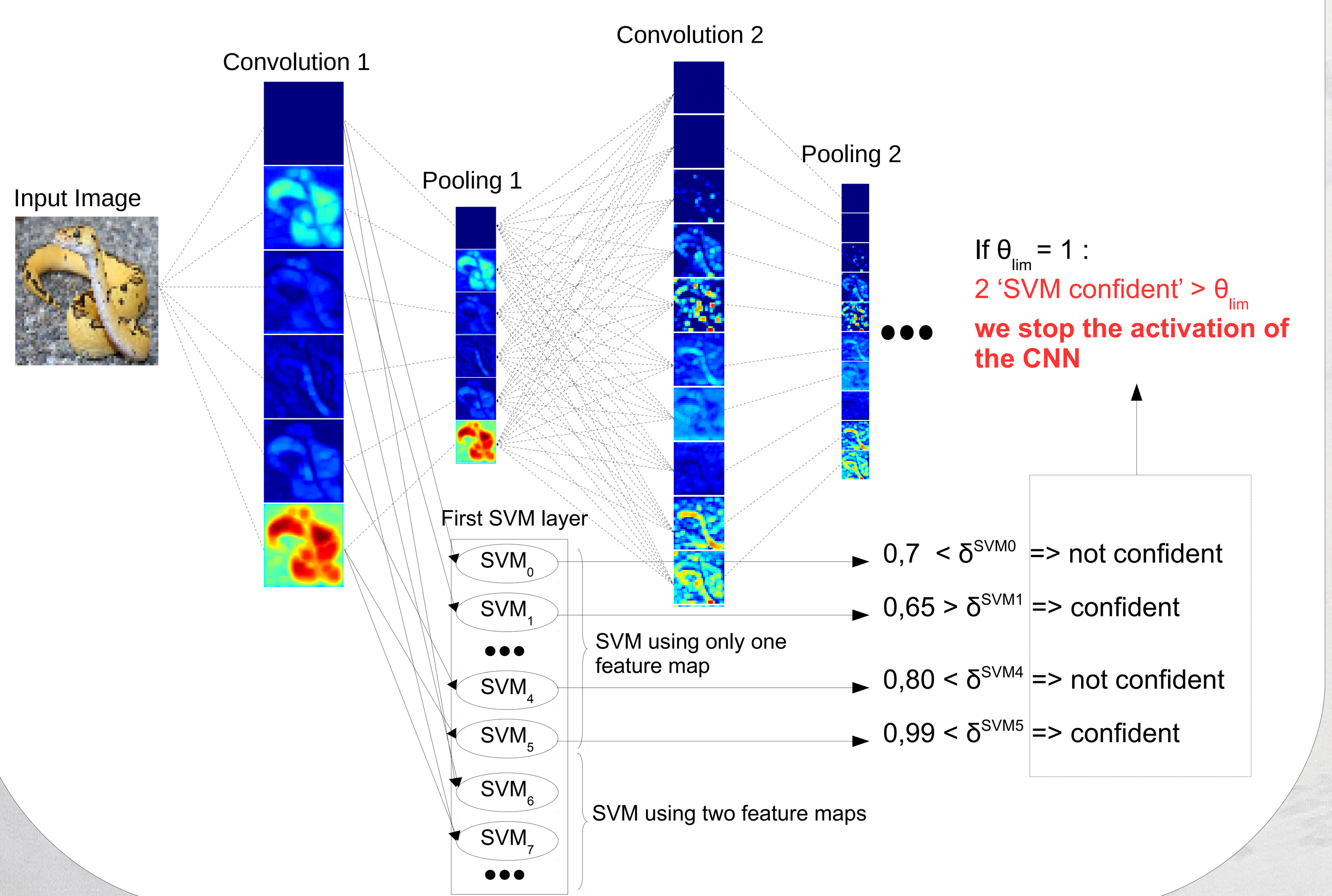
SVM Layer	Input layers	#SVM
SVM1	Pool1	68
SVM2	SVM1	25
SVM3	Pool3	232
SVM4	SVM1 & SVM3	50
SVM5	Pool5	348
SVM6	SVM4 & SVM5	50
SVM7	FC1	1
SVM8	FC2	1
SVM9	SVM6-SVM7	50



2. For each SVM i, we find the best threshold (noted δ^i) for which the precision is higher than an arbitrary value, using a validation database.

3. We sort all the SVM using their activation cost (see eq. 1) in such a way that the recall is maximal. This order defines the **activation path** [Pasquet et al., 2015].

4. During the testing step :
   - the CNN and the SVM network are activated **at the same time**,
   - if **score(SVM_i)≥δ^i** →SVM is **confident** ,
   - if **#(SVM confident)> θ_{lim}** → the network activation is **stopped**.



Results and conclusion

We define the computational cost as :

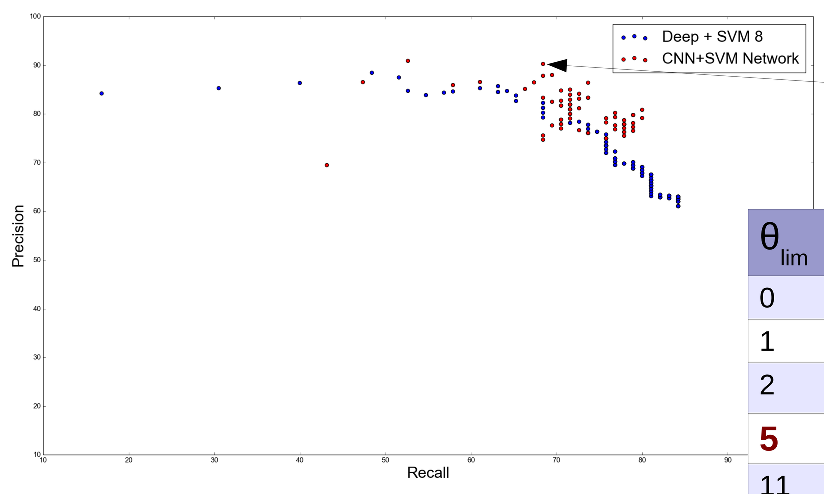
$$w^{(c)}=w^{(c-1)}+\begin{cases}(\text{kernel size})\times out^{(c-1)}\times N^{(c)} & \text{if convolution layer}\\ out^{(c-1)}\times N^{(c)} & \text{if FC layer}\\ 0 & \text{otherwise}\end{cases} \quad [1]$$

with $N^{(c)}$ the number of feature maps from the layer c and $out^{(c)}\begin{cases}N^{(c)} & \text{if FC layer}\\ height\times width & \text{otherwise}\end{cases}$

We model the computational cost using the relative cost, noted RC :

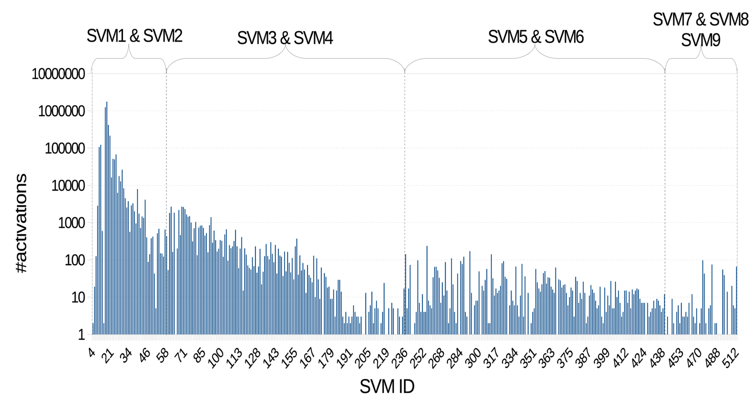
$$RC=\frac{w^{(c)}}{w^{(FC2)}}$$

The effectiveness of the proposed method :



θ_{lim}	Precision	Recall	RC (%)
0	69.49	43.15	1.38
1	86.53	47.56	1.59
2	90.90	52.53	1.75
5	**68.42**	**90.27**	**2.17**
11	82.7	70.52	2.83
26	80.95	71.52	4.74
27	77.90	70.52	4.94

Analysis of the proposed method :



The number of activations for each SVM

Average efficiency for groups of SVM layers with θ_{lim}

SVM Layer	#Activations	Error (%)	FP (%)	TP (%)
SVM1-2	3,946,411	0.38	0.38	0.02
SVM3-4	31,433	18.76	18.61	0.71
SVM5-6	4,724	50.70	49.03	5.89
SVM7-9	450	42.85	30.84	35.61

→ The activation path **massively reduces the computational cost**. In our experiments, for a recall of only 67%, an average of **97.8% of the network remains unused.**

→ **9% precision gain** for a recall set of 67%.

The authors would like to thank the Berger-Levrault group for supporting this research. Berger-Levrault is a French public regulation expert that addresses healthcare and local public administrations.

Website : www.berger-levrault.com

[1] Pasquet J., Chaumont M., Subsol G., and Derras M., "An efficient multi-resolution SVM network approach for object detection in aerial images," in Machine Learning for Signal processing, MLSP2015. IEEE, 2015

[2] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Computer Vision and Pattern Recognition, 2001. CVPR 2001. IEEE 2001

[3] Yichuan Tang, "Deep learning using support vector machines," ICML 2013 Challenges in Representation Learning Workshop, 2013.

[4] A. Krizhevsky, I. Sutskever, G. E. Hinton " ImageNet Classification with Deep Convolutional Neural Networks. "
in Advances in Neural Information Processing Systems 25, pp. 1097–1105. 2012.