Learnable pooling with Context Gating for Video Classification

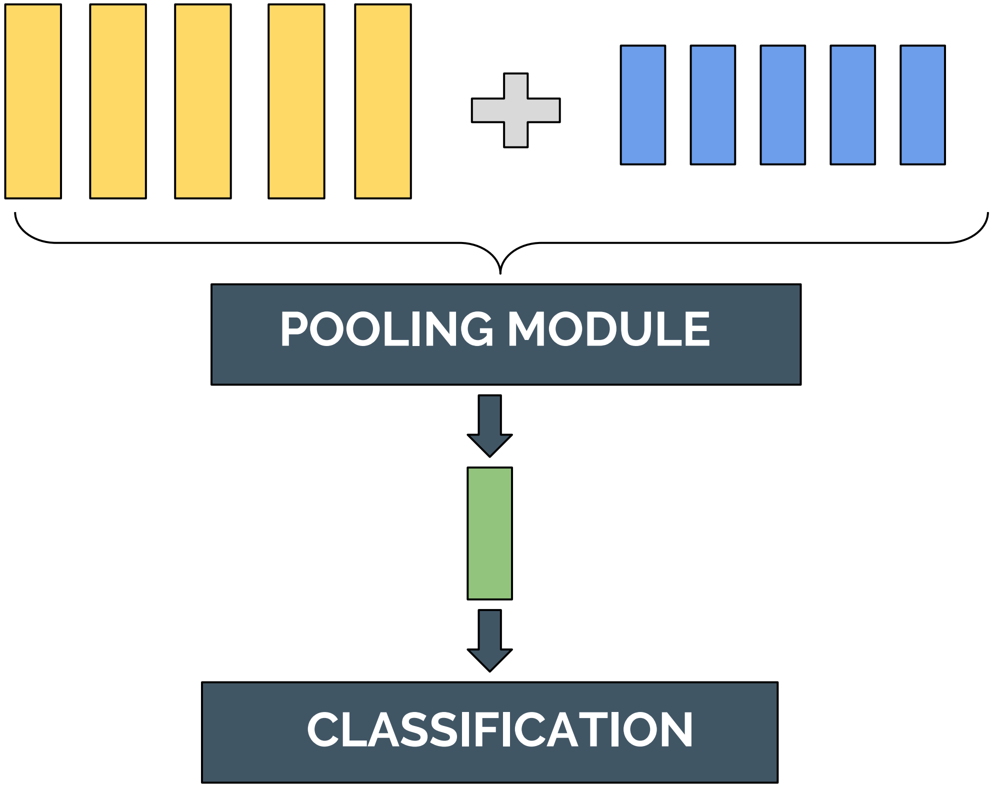
Antoine Miech, Ivan Laptev, Josef Sivic

ENS
ÉCOLE NORMALE SUPÉRIEURE

Inria
INVENTEURS DU MONDE NUMÉRIQUE

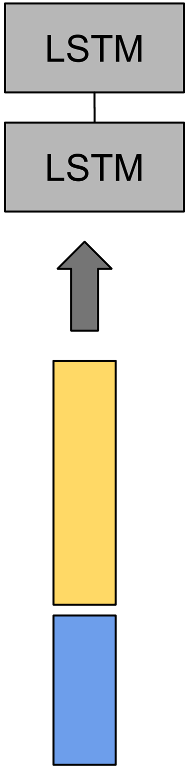
CZECH TECHNICAL
UNIVERSITY
IN PRAGUE

Goal: Multi-modal features pooling

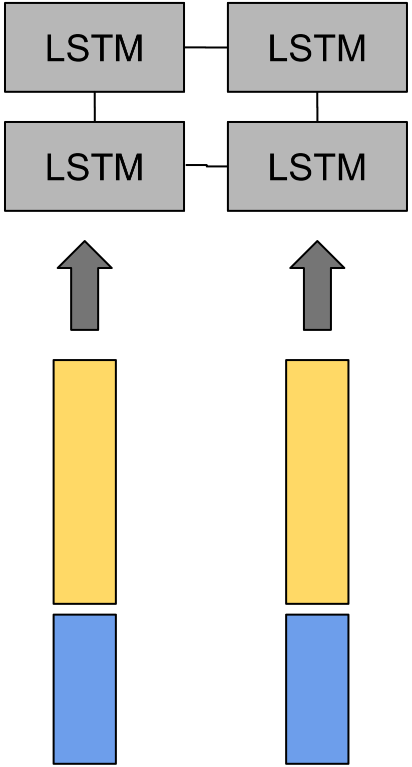


3

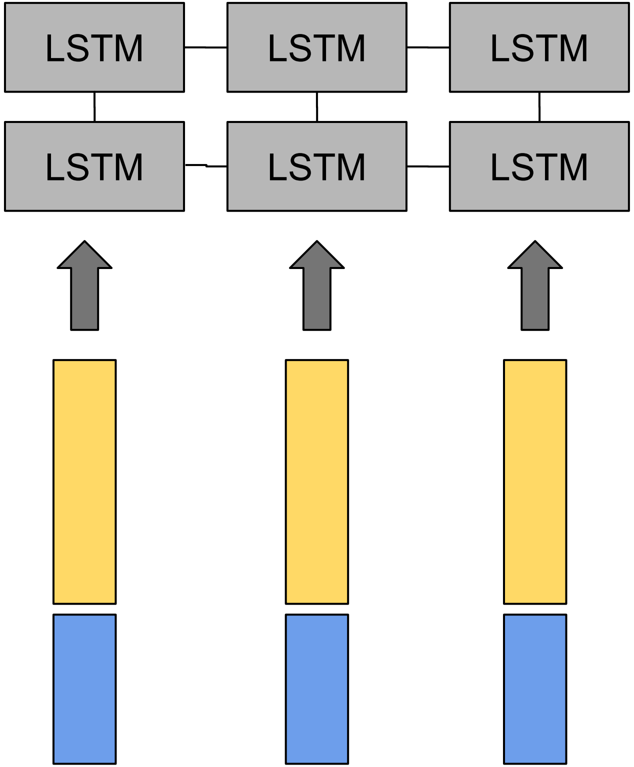
Recurrent model (e.g LSTM)



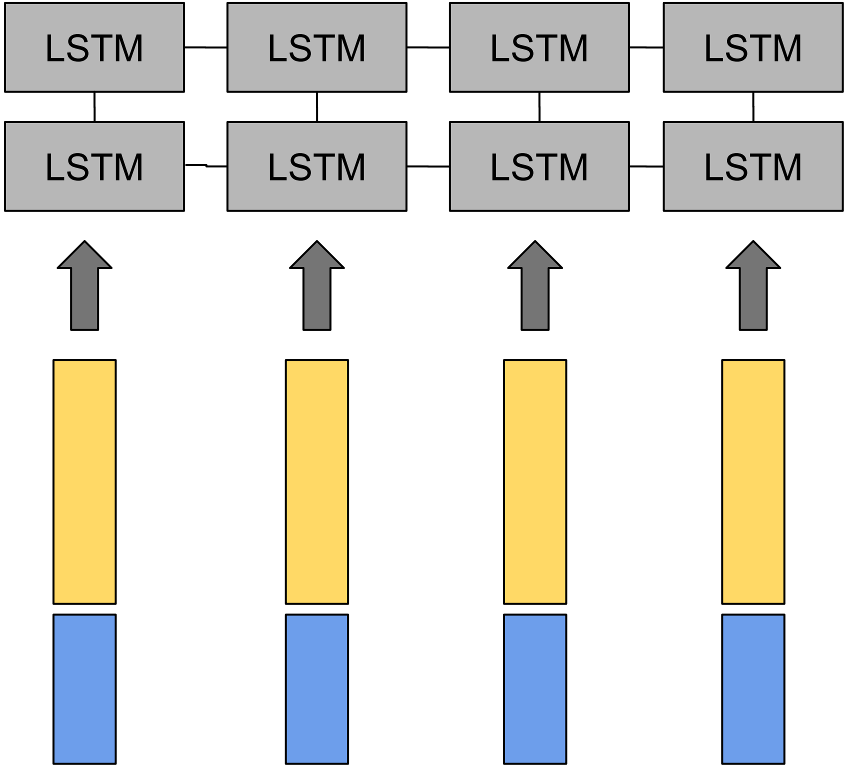
Recurrent model (e.g LSTM)



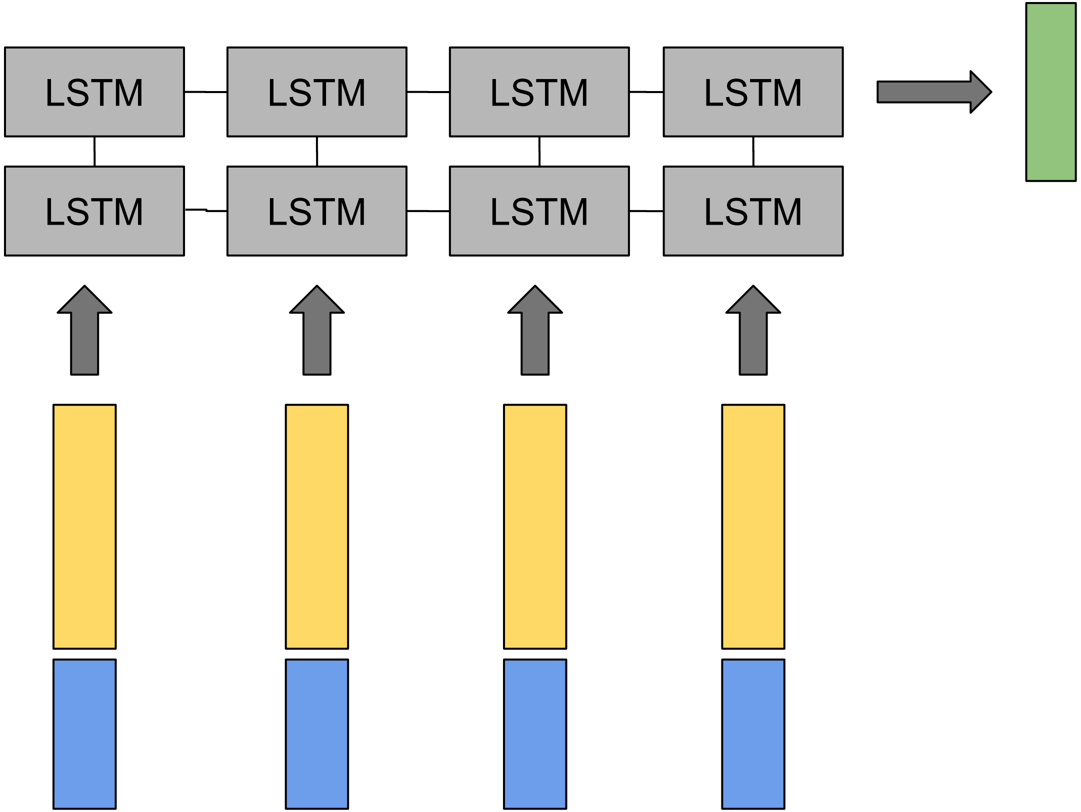
Recurrent model (e.g LSTM)



Recurrent model (e.g LSTM)



Recurrent model (e.g LSTM)

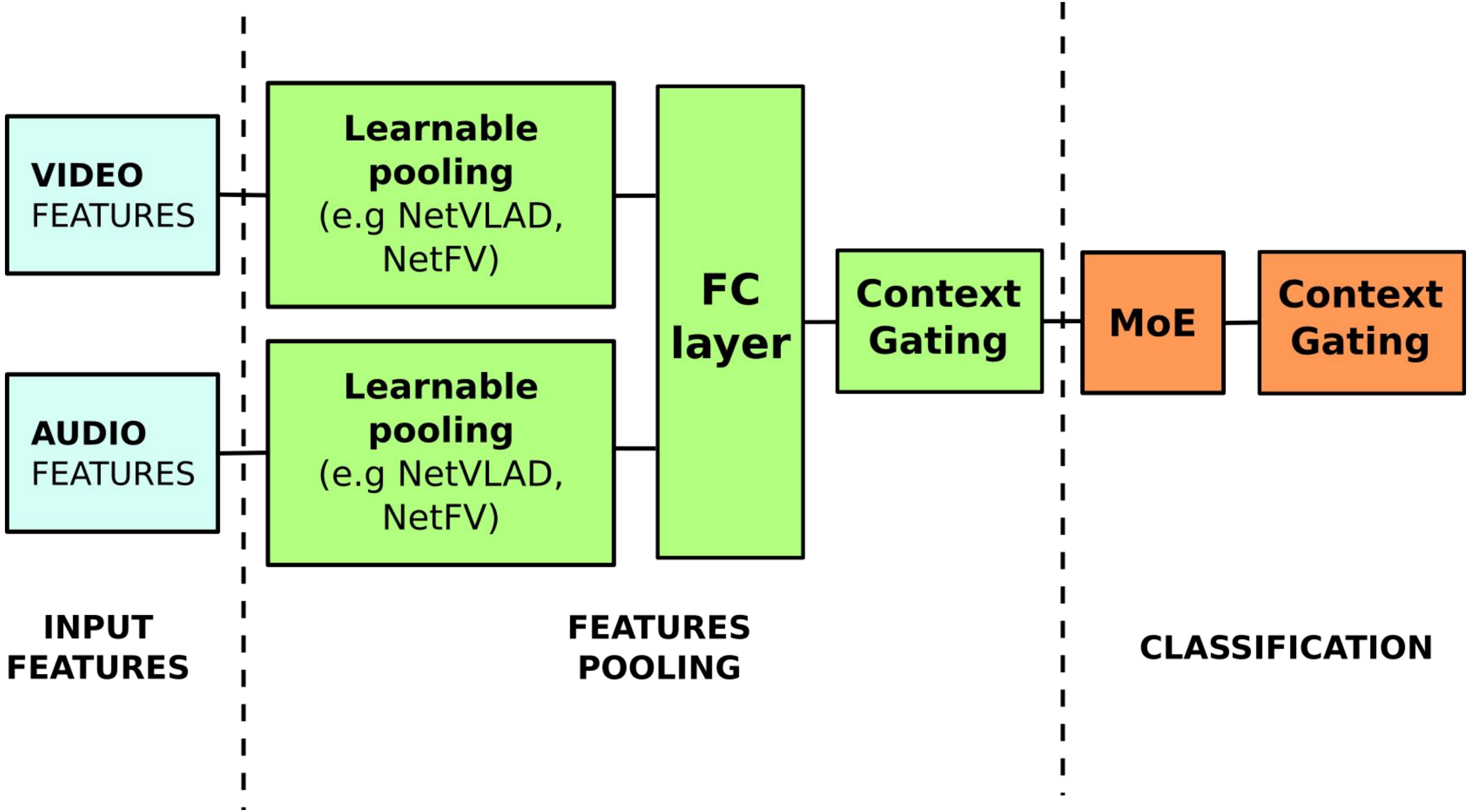


Problem ?

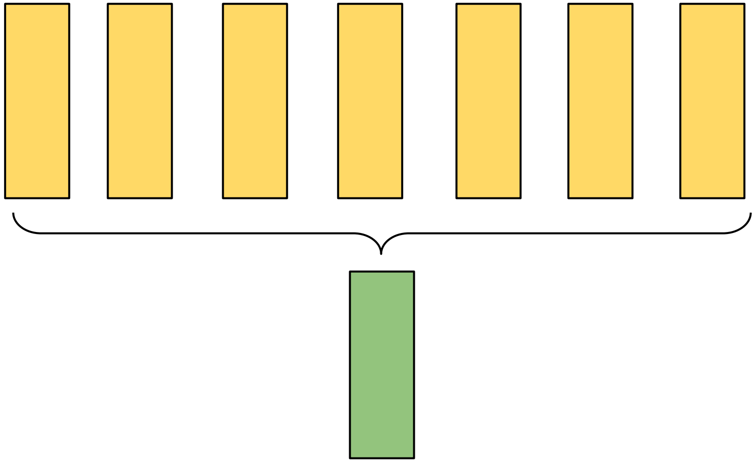
- Slow for inference/training
- This is NOT a sequential problem
- Needs lots of data for training
- How about very long videos ?

But surprisingly good results !

Our approach

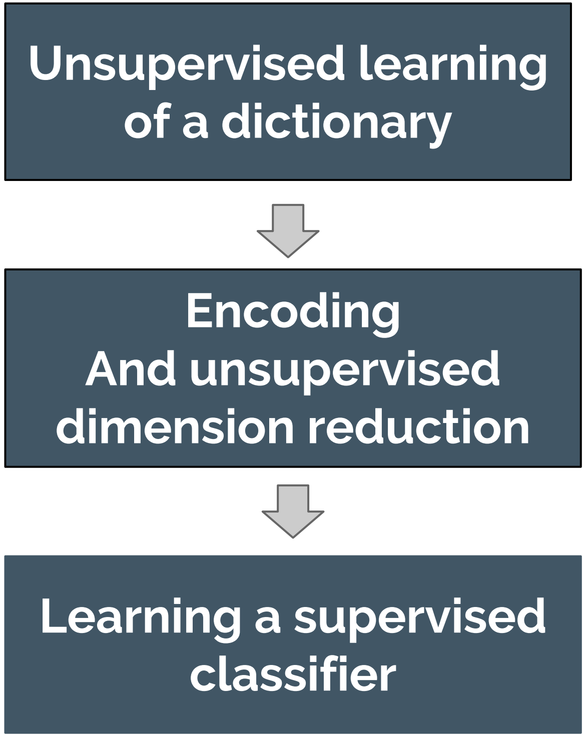


Traditional pooling



- **Bag-of-visual-words** [Sivic and Zisserman, 2003][Csurka et al., 2004]
- **VLAD** [Jégou et al., 2010]
- **Fisher Vector** [Perronnin et al., 2007]

Traditional Pipeline VS End-to-end Pipeline



Unsupervised learning of a dictionary

Encoding
And unsupervised dimension reduction

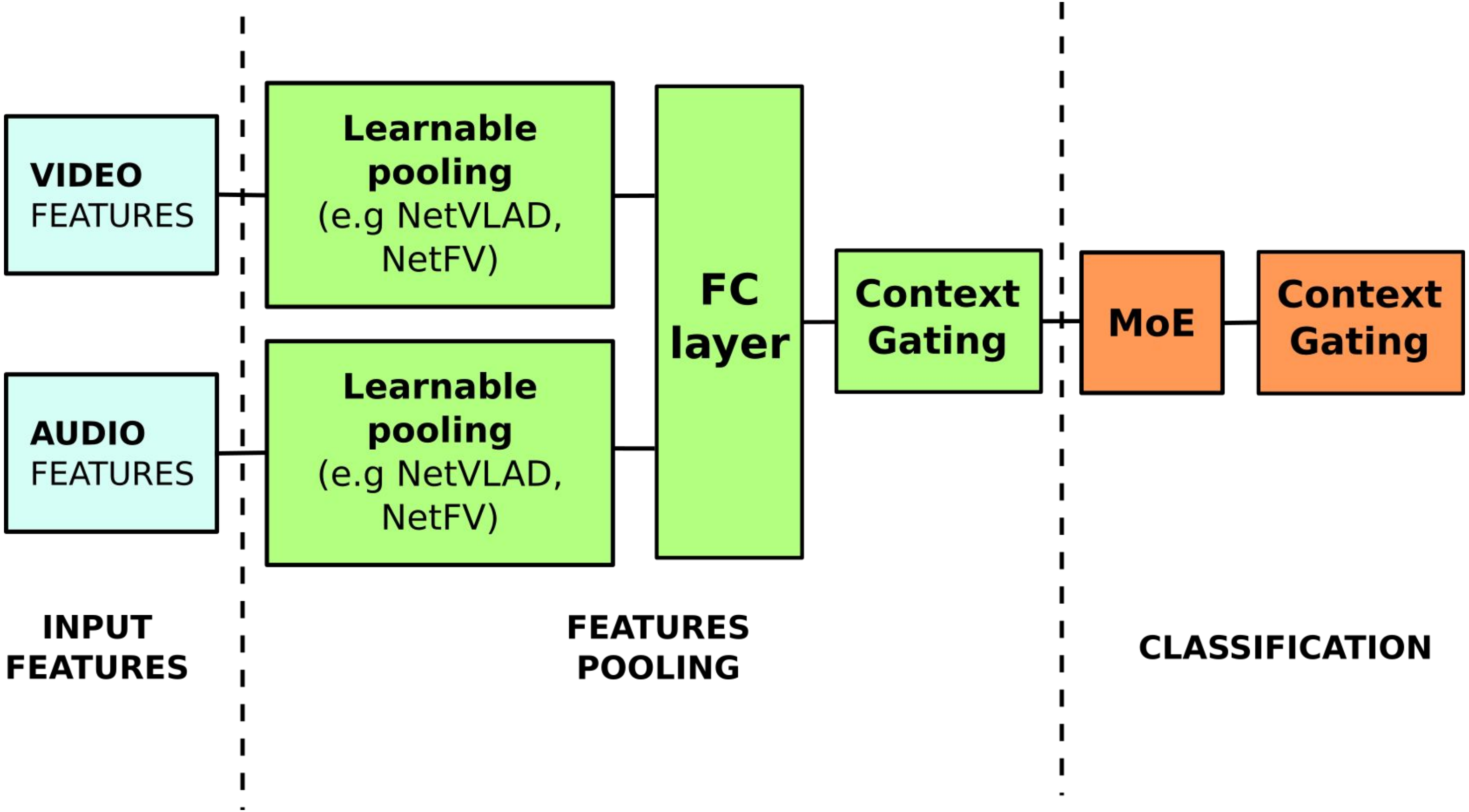
Learning a supervised classifier

Supervised learning of dictionary
+
Supervised dimension reduction
+
Learning a supervised classifier

Learnable pooling

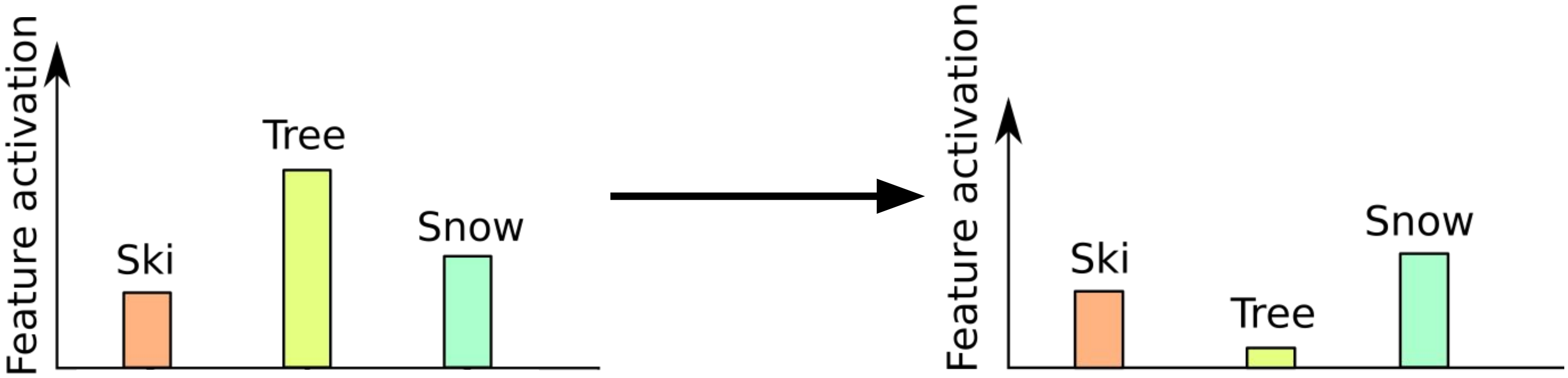
Unsupervised	End-to-End
Bag-of-visual-Words [Sivic and Zisserman, 2003]	Soft-DBoW
VLAD [Jégou et al., 2010]	NetVLAD [Arandjelović et al. CVPR 2016]
Fisher Vector [Perronnin et al., 2007]	NetFV

Model overview



Context Gating

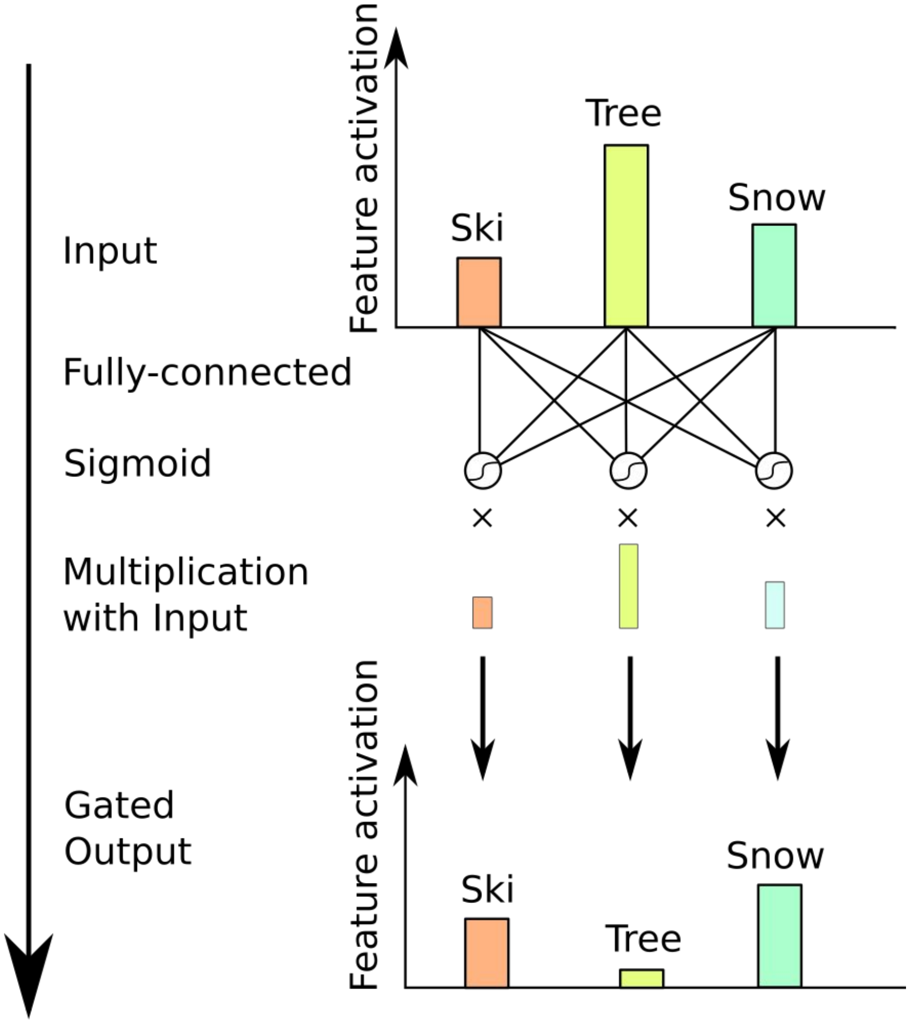
Context Gating



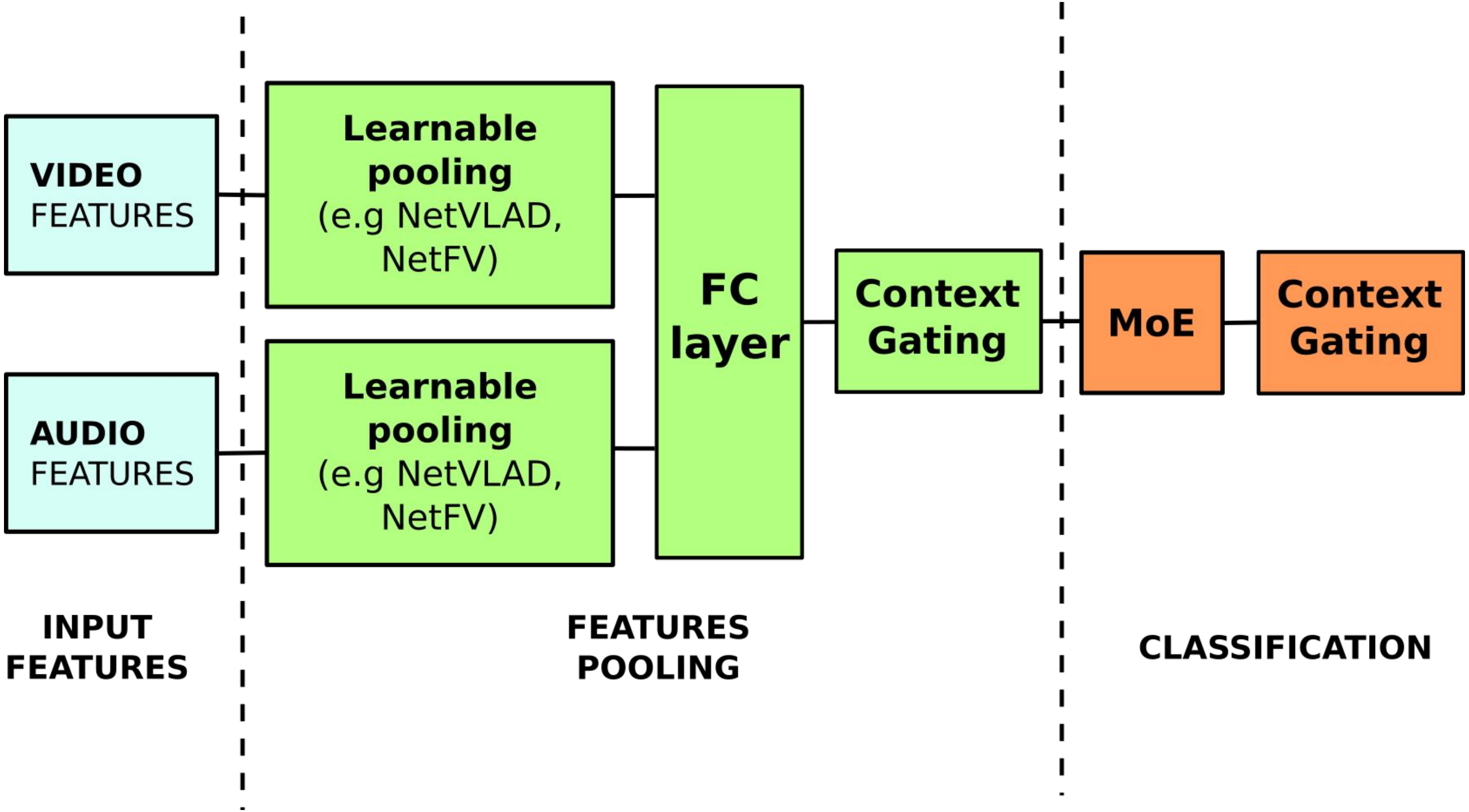
Context Gating

Equation:

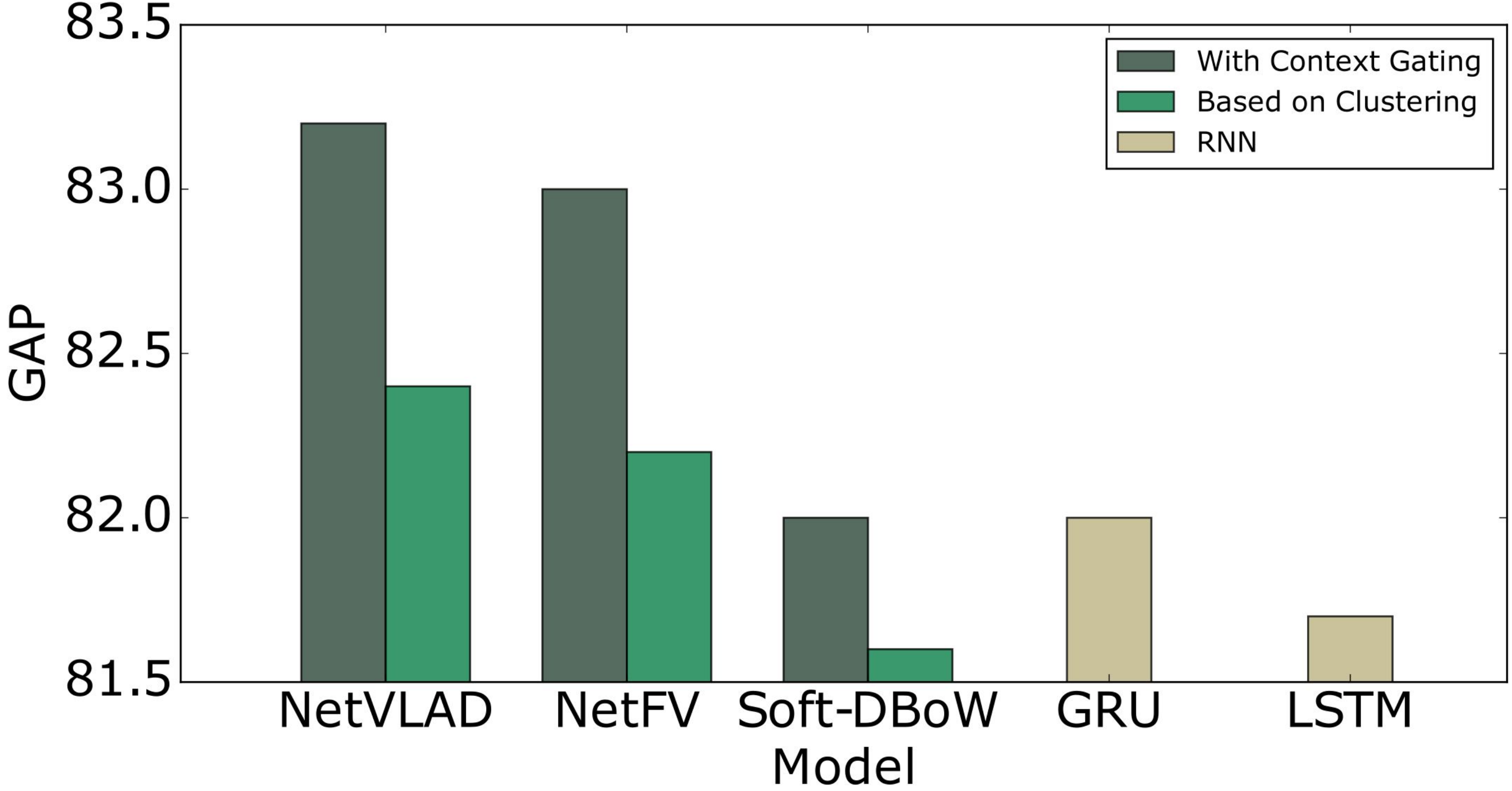
$$Y = \sigma(WX + b) \circ X$$



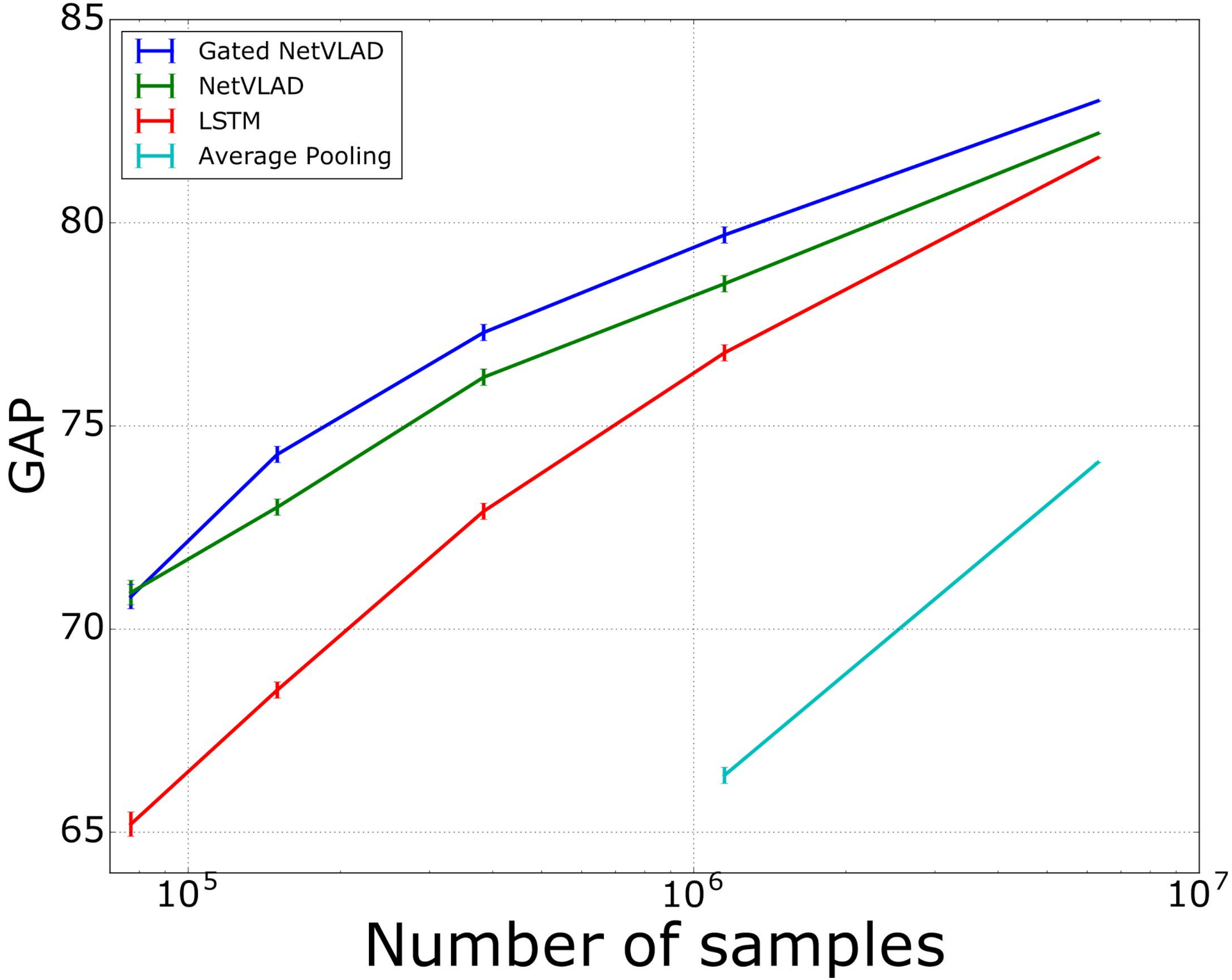
Model overview

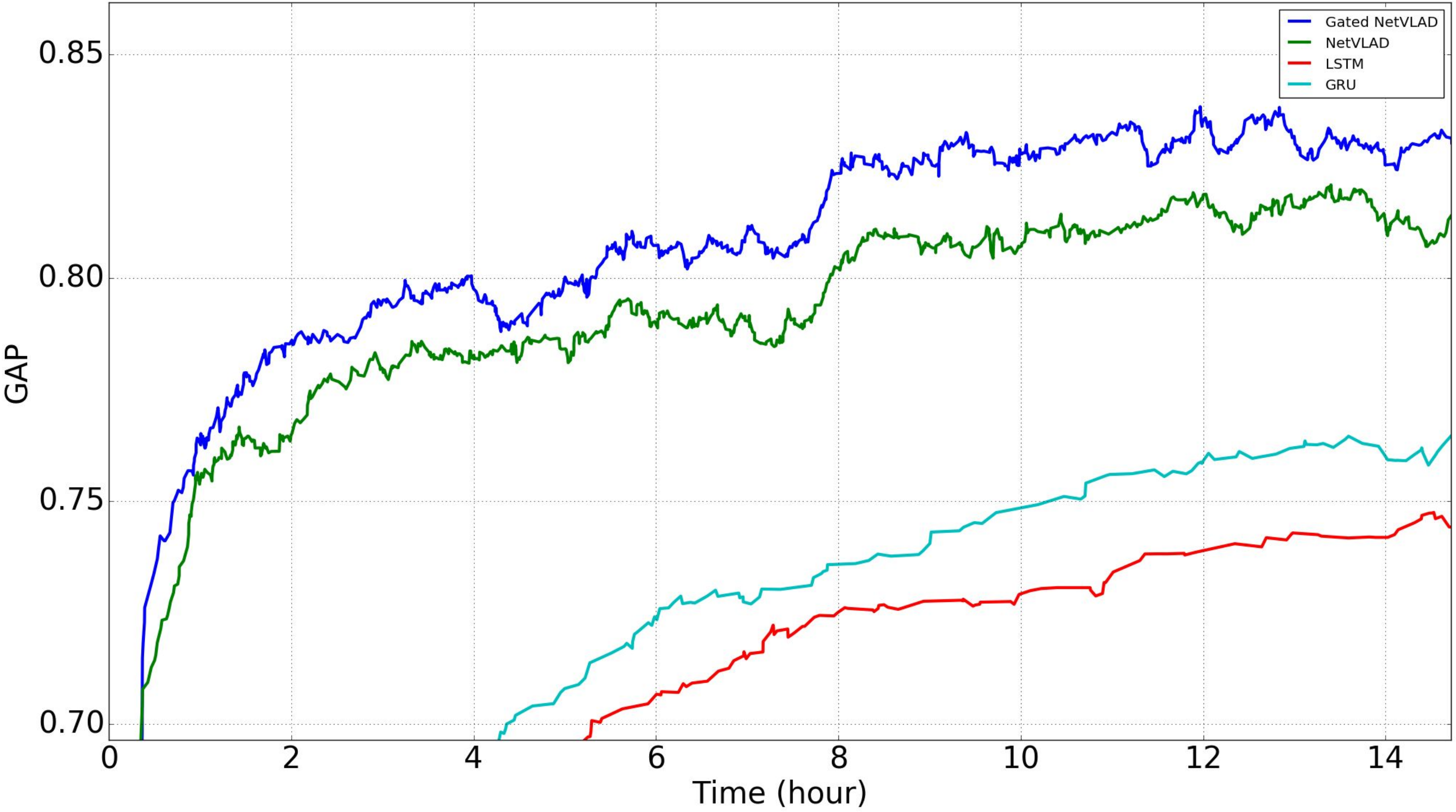


Results


83.5
83.0
82.5
82.0
81.5
GAP
With Context Gating
Based on Clustering
RNN
NetVLAD
NetFV
Soft-DBoW
GRU
LSTM
Model

Generalization

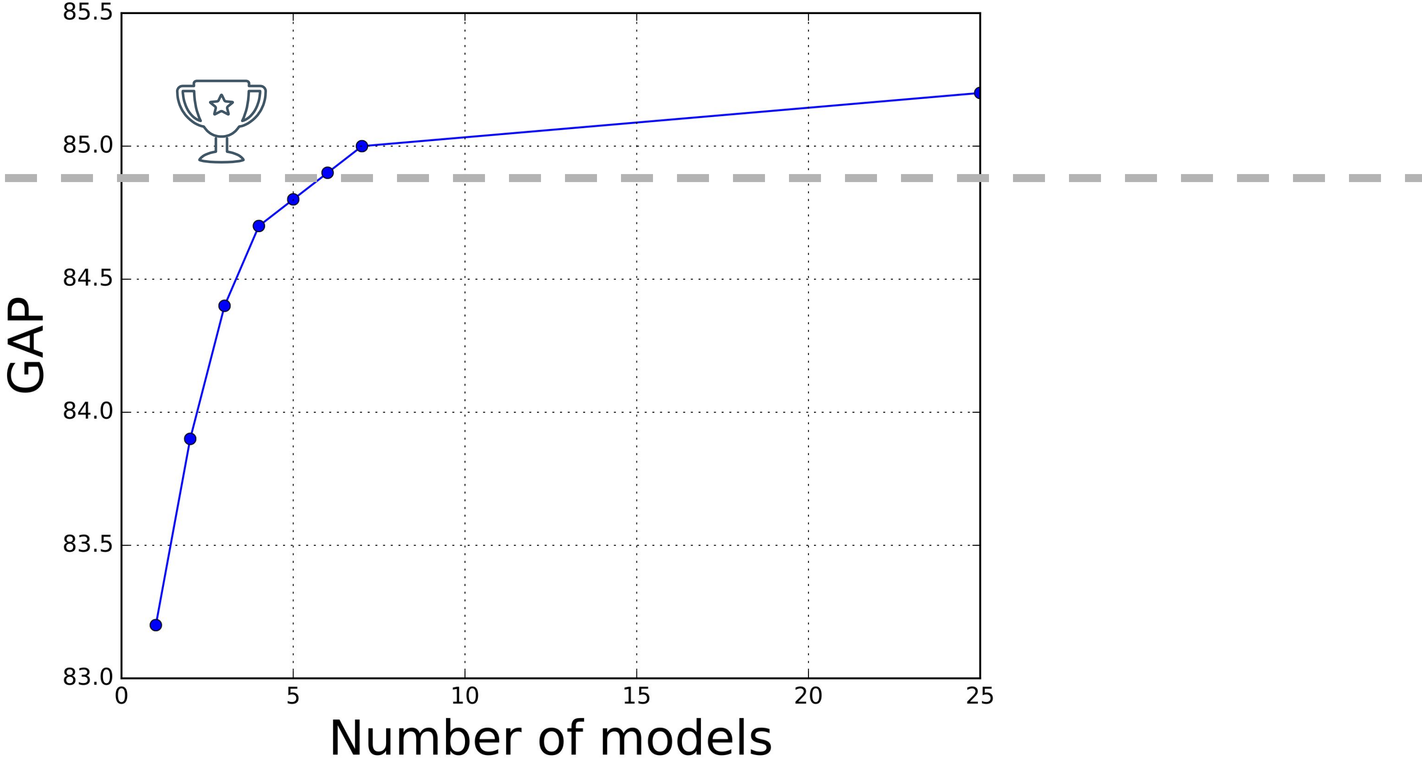



Gated NetVLAD
NetVLAD
LSTM
GRU
GAP
0.70
0.75
0.80
0.85
0
2
4
6
8
10
12
14
Time (hour)

Bonus

🏆 Winning the kaggle competition

Effects of ensembling



LOUPE Tensorflow toolbox

GITHUB LOUPE repo:

github.com/antoine77340/LOUPE

GITHUB Kaggle code repo:

github.com/antoine77340/Youtube-8M-WILLOW

Questions ?