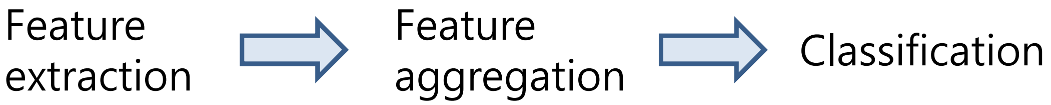
Learnable pooling with context Gating for video classification

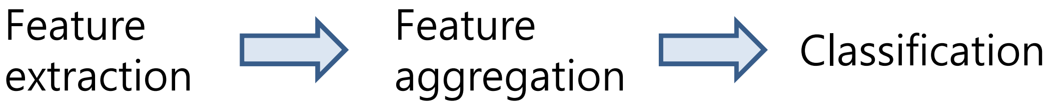
2017. 07. 08.

Video data classification



Feature extraction → Feature aggregation → Classification

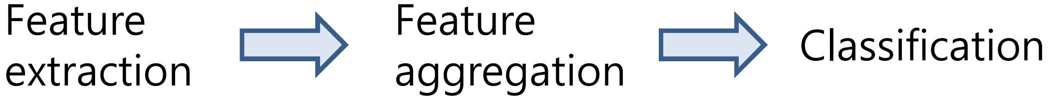
Video data classification(WILLOW)



Feature extraction → Feature aggregation → Classification

- Feature extraction : Given
- Feature aggregation : Use existing methods
    - LSTM, GRU, DBoF, VLAD, Fisher Vector encoding
- Classification : MoE

Video data classification(WILLOW)

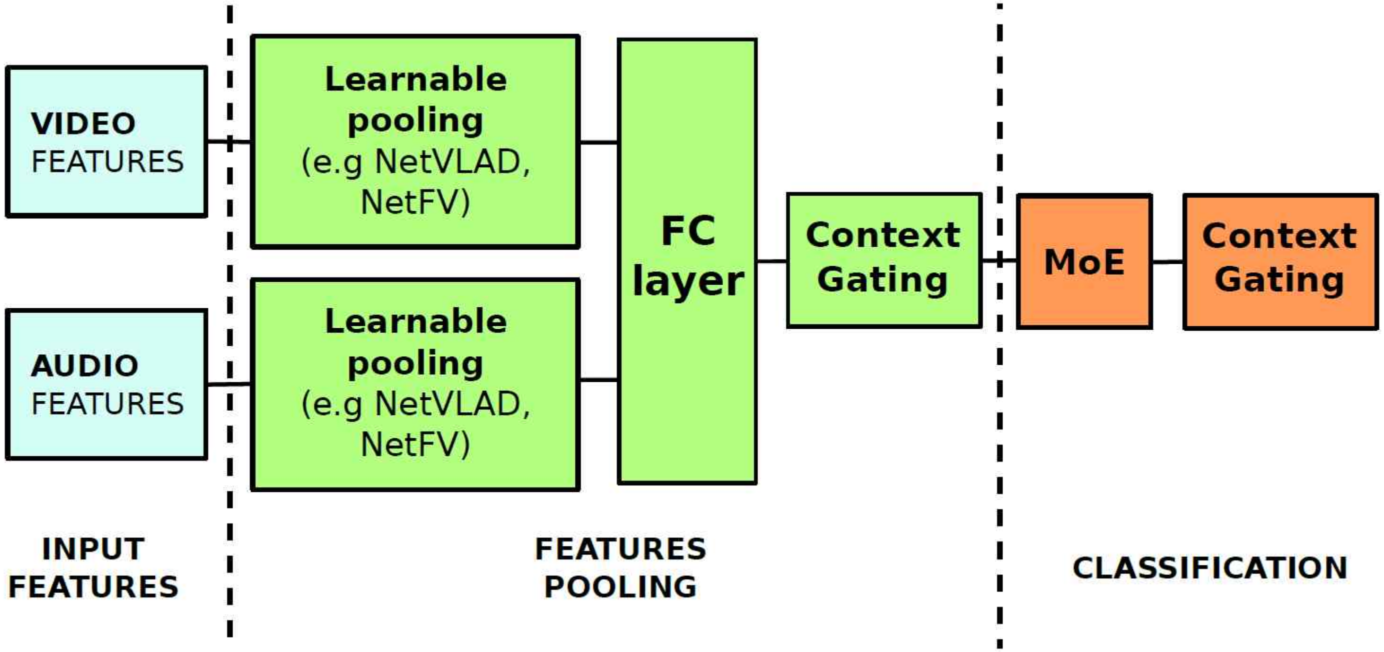


Feature extraction → Feature aggregation → Classification

- Feature extraction : Given
- Feature aggregation : Use existing methods
  - LSTM, GRU, DBoF, VLAD, Fisher Vector encoding
- Classification : MoE

+ Context Gating

Overview


VIDEO FEATURES
AUDIO FEATURES
Learnable pooling (e.g NetVLAD, NetFV)
Learnable pooling (e.g NetVLAD, NetFV)
FC layer
Context Gating
MoE
Context Gating
INPUT FEATURES
FEATURES POOLING
CLASSIFICATION

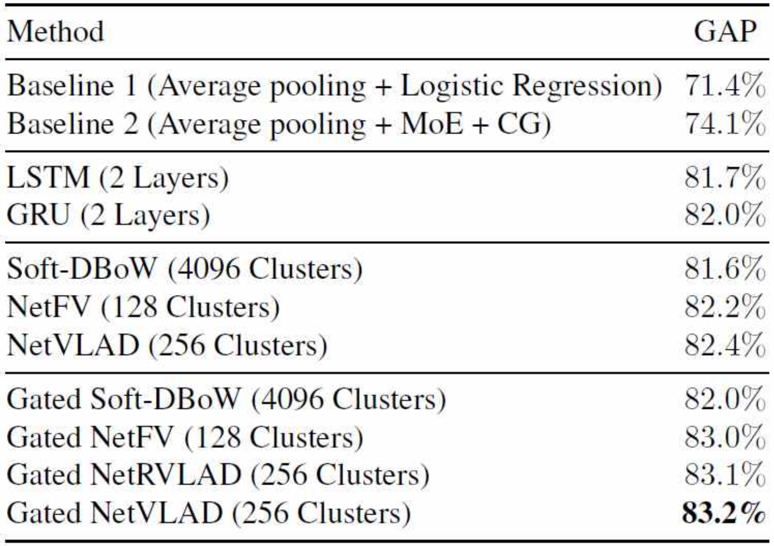
Context Gating

$$CG(X) = \sigma(W\ X + b)\ \cdot X$$

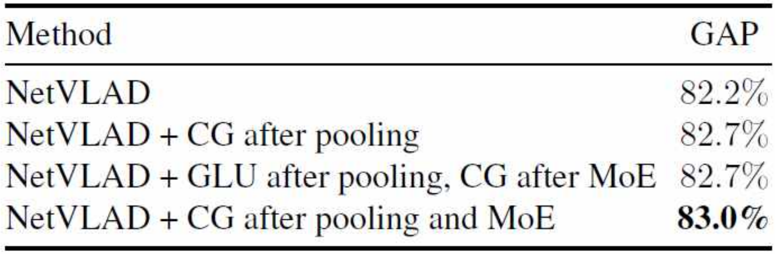
- Motivation
  1. Wish to introduce non-linear interactions among activations of the input representation
  2. Wish to recalibrate the strengths of different activations of the input representation through a self-gating mechanism

- Aim
  1. Capture dependencies among features
  2. Capture prior structure of output space

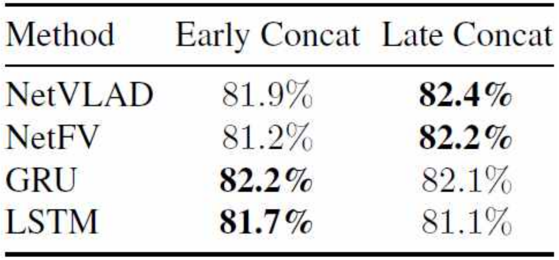
Experiments



Method	GAP
Baseline 1 (Average pooling + Logistic Regression)	71.4%
Baseline 2 (Average pooling + MoE + CG)	74.1%
LSTM (2 Layers)	81.7%
GRU (2 Layers)	82.0%
Soft-DBoW (4096 Clusters)	81.6%
NetFV (128 Clusters)	82.2%
NetVLAD (256 Clusters)	82.4%
Gated Soft-DBoW (4096 Clusters)	82.0%
Gated NetFV (128 Clusters)	83.0%
Gated NetRVLAD (256 Clusters)	83.1%
Gated NetVLAD (256 Clusters)	**83.2%**

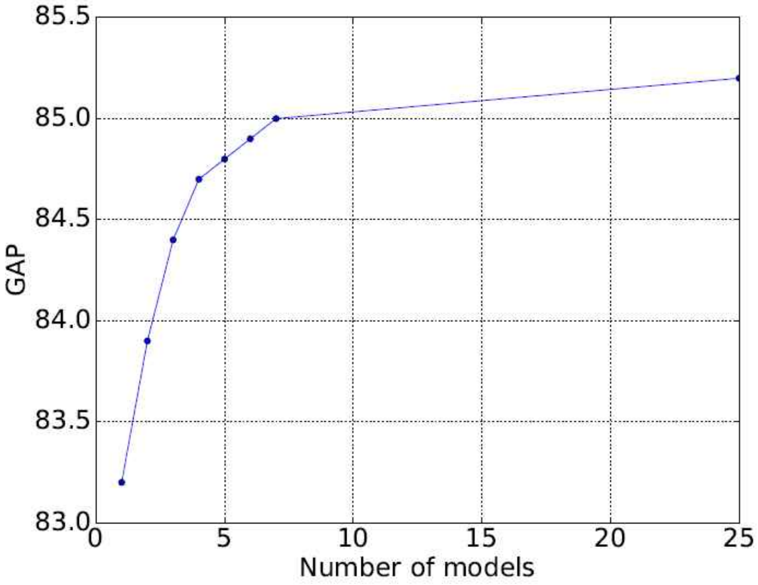


Method	GAP
NetVLAD	82.2%
NetVLAD + CG after pooling	82.7%
NetVLAD + GLU after pooling, CG after MoE	82.7%
NetVLAD + CG after pooling and MoE	**83.0%**



Method	Early Concat	Late Concat
NetVLAD	81.9%	**82.4%**
NetFV	81.2%	**82.2%**
GRU	**82.2%**	82.1%
LSTM	**81.7%**	81.1%

Ensemble



- The ensemble did not bring much when combining best but similar models
- Simple greedy approach
- The first seven models
  - Gated NetVLAD
  - Gated NetFV
  - Gated Soft-DBoW
  - Soft DBoW
  - Gated NetRVLAD
  - GRU
  - LSTM