Beyond Gaussian Pyramid: Multi-skip Feature Stacking for Action Recognition

Zhenzhong Lan, Ming Lin, Xuanchong Li, Alexander G. Hauptmann, Bhiksha Raj
School of Computer Science, Carnegie Mellon University
lanzhzh, minglin, xcli, alex, bhiksha@cs.cmu.edu

Abstract

This is the supplementary material for the paper entitled "Beyond Gaussian Pyramid: Multi-skip Feature Stacking for Action Recognition". The material gives the proof of theorem 1 and 2.

1. Proof

We present the proofs of the following two theorems.

Theorem 1. *Given a fixed time skip τ, with probability at least $1-\delta$, the condition number $\beta(PP^{\mathrm{T}})$ is bounded by*

$$\beta(PP^{\mathrm{T}}) \leq \frac{(1+c)\exp(-\gamma_1/\tau)+\Delta_\tau}{\exp(-\gamma_k/\tau)-\Delta_\tau} \tag{1}$$

$$\beta(PP^{\mathrm{T}}) \geq \frac{(1+c)\exp(-\gamma_1/\tau)-\Delta_\tau}{\exp(-\gamma_k/\tau)+\Delta_\tau} \,. \tag{2}$$

where

$$\Delta_\tau = 2\sqrt{k\frac{1}{T}(1+c)\log(2k/\delta)} \tag{3}$$

provided the number of feature points

$$T \geq \frac{1}{9(1+c)} k\log(2k/\delta) \,. \tag{4}$$

Theorem 2. *With probability at least $1-\delta$, the condition number of PP^{T} in the MIFS is bounded by*

$$\beta(PP^{\mathrm{T}}) \leq \frac{\sum_i \frac{T_i}{T} 2(1+c)\exp(-\gamma_1/\tau_i)+\Delta_\tau}{\sum_i \frac{T_i}{T} 2\exp(-\gamma_k/\tau_i)-\Delta_\tau} \,. \tag{5}$$

where

$$\Delta_\tau \leq 2\sqrt{k\frac{1}{\sum_i T_i}(1+c)\log(2k/\delta)} \,. \tag{6}$$

provided the number of feature points

$$T \geq \frac{1}{9(1+c)} k\log(2k/\delta) \,. \tag{7}$$

Our proofs are based on the following Matrix Bernstein's Inequaltiy.

Lemma 1 (Matrix Bernstein's Inequality)**.** *Let $\mathbf{x}_i \in \mathbb{R}^{p\times 1}$, $\|\mathbf{x}_i\|^2 \leq B$. $S = \mathbf{x}_1{\mathbf{x}_1}^{\mathrm{T}} + \cdots \mathbf{x}_n{\mathbf{x}_n}^{\mathrm{T}}$. Then with probability at least $1-\delta$,*

$$\|S - \mathrm{E}\{S\}\| \leq \sqrt{2B\,\|\mathrm{E}\{S\}\|\log(2p/\delta)} + \frac{B}{3}\log(2p/\delta) \,. \tag{8}$$

1.1. Proof of Theorem 1

Proof. For the i-th row, j-th column of P,

$$
\begin{aligned}
|P_{i,j}| =&[\boldsymbol{\alpha}_i(t_j+\tau)-\boldsymbol{\alpha}_i(t_j)] \le 2 && (9)\\
\|P_j\|^2 \le& 4k && (10)\\
\mathrm{E}\{P_{i,j}^2\} \le& 2(1+c)\exp(-\gamma/\tau) && (11)\\
\mathrm{E}\{P_{i,j}^2\} \ge& 2\exp(-\gamma/\tau) && (12)\\
\mathrm{E}\{P_{i,j}P_{k,j}\} =&0 \quad i\neq k && (13)\\
\lambda_{\max}\{\mathrm{E}\{P_jP_j{}^{\mathrm{T}}\}\} =&\frac{1}{T}\lambda_{\max}\{\mathrm{E}\{PP^{\mathrm{T}}\}\} \le 2(1+c)\exp(-\gamma_1/\tau) && (14)\\
\lambda_{\min}\{\mathrm{E}\{P_jP_j{}^{\mathrm{T}}\}\} =&\frac{1}{T}\lambda_{\min}\{\mathrm{E}\{PP^{\mathrm{T}}\}\} \ge 2\exp(-\gamma_k/\tau)\ . && (15)
\end{aligned}
$$

By Matrix Bernstein's inequality, with probability at least $1-\delta$, we have

$$
\begin{aligned}
4T\Delta_\tau \triangleq \left\|PP^{\mathrm{T}}-\mathrm{E}\{PP^{\mathrm{T}}\}\right\| \le& \sqrt{2\times 4k\times 2T(1+c)\exp(-\gamma_1/\tau)\log(2k/\delta)}+\frac{4k}{3}\log(2k/\delta)\\
=&4\sqrt{kT(1+c)\exp(-\gamma_1/\tau)\log(2k/\delta)}+\frac{4}{3}k\log(2k/\delta)\\
\le&4\sqrt{kT(1+c)\log(2k/\delta)}+\frac{4}{3}k\log(2k/\delta) && (16)
\end{aligned}
$$

When

$$T \ge \frac{1}{9(1+c)}k\log(2k/\delta) \qquad (17)$$

we have

$$\Delta_\tau \le 2\sqrt{k\frac{1}{T}(1+c)\log(2k/\delta)} \qquad (18)$$

Therefore, when T is large enough,

$$\beta(PP^{\mathrm{T}}) \le \frac{(1+c)\exp(-\gamma_1/\tau)+\Delta_\tau}{\exp(-\gamma_k/\tau)-\Delta_\tau}\ . \qquad (19)$$

A lower bound on $\beta(PP^{\mathrm{T}})$ could be given similarly by changing Δ_τ to $-\Delta_\tau$. □

1.2. Proof of Theorem 2

Proof. The proof is similar to Theorem 1, except that now we have P_i that is sampled from m different distributions. The i-th component of P_i is sampled from $i\tau$ skip with probability $T_i/\sum_j T_j = T_i/T$, where $T=\sum_j T_j$ is the total number of features. Follow the proof of Theorem 1, we have:

$$
\begin{aligned}
|P_{i,j}| \le& 2 && (20)\\
\|P_j\|^2 \le& 4k && (21)\\
\mathrm{E}\{P_{i,j}^2\} \le& \sum_i \frac{T_i}{T}2(1+c)\exp(-\gamma/\tau_i) && (22)\\
\mathrm{E}\{P_{i,j}^2\} \ge& \sum_i \frac{T_i}{T}2\exp(-\gamma/\tau_i) && (23)\\
\mathrm{E}\{P_{i,j}P_{k,j}\} =&0 \quad i\neq k && (24)\\
\lambda_{\max}\{\mathrm{E}\{P_jP_j{}^{\mathrm{T}}\}\} =&\frac{1}{T}\lambda_{\max}\{\mathrm{E}\{PP^{\mathrm{T}}\}\} \le \sum_i \frac{T_i}{T}2(1+c)\exp(-\gamma_1/\tau_i) && (25)\\
\lambda_{\min}\{\mathrm{E}\{P_jP_j{}^{\mathrm{T}}\}\} =&\frac{1}{T}\lambda_{\min}\{\mathrm{E}\{PP^{\mathrm{T}}\}\} \ge \sum_i \frac{T_i}{T}2\exp(-\gamma_k/\tau_i) && (26)
\end{aligned}
$$

By Matrix Bernstein's inequality, with probability at least $1-\delta$, we have

$$\begin{aligned}4T\Delta_\tau \triangleq \left\|PP^{\mathrm{T}} - \mathrm{E}\{PP^{\mathrm{T}}\}\right\| \leq& \sqrt{2\times 4k \times [\sum_i T_i 2(1+c)\exp(-\gamma/\tau_i)]\log(2k/\delta)} + \frac{4k}{3}\log(2k/\delta) \\ =& 4\sqrt{k[\sum_i T_i(1+c)\exp(-\gamma/\tau_i)]\log(2k/\delta)} + \frac{4}{3}k\log(2k/\delta) \\ \leq& 4\sqrt{k(1+c)T\log(2k/\delta)} + \frac{4}{3}k\log(2k/\delta) \end{aligned} \tag{27}$$

When

$$T \geq \frac{1}{9(1+c)}k\log(2k/\delta) \tag{28}$$

we have

$$\Delta_\tau \leq 2\sqrt{k\frac{1}{T}(1+c)\log(2k/\delta)} \tag{29}$$

$$\beta(PP^{\mathrm{T}}) \leq \frac{\sum_i \frac{T_i}{T}2(1+c)\exp(-\gamma_1/\tau_i) + \Delta_\tau}{\sum_i \frac{T_i}{T}2\exp(-\gamma_k/\tau_i) - \Delta_\tau}\ . \tag{30}$$

□