1 Convex Program for Two-Layer Convolutional Vector Output Neural Networks with Average Pooling

We can modify the convex formulation for CNN with average pooling from the polynomial activation NN paper into an architecture with vector outputs (with output dimension C) as follows. P is the pooling size, K is the number of patches.

$$
\begin{aligned}
&\min_{Z_k^{(t)}, Z_k'^{(t)}} \ell(\hat{Y}, Y) + \beta \sum_{t=1}^{C} \sum_{k=1}^{K/P} (Z_{k,4}^{(t)} + Z_{k,4}'^{(t)}) \\
&\text{s.t.} \quad \hat{Y}_{it} = a \frac{1}{P} \sum_{k=1}^{K/P} \sum_{l=1}^{P} x_{i,(k-1)P+l}^T (Z_{k,1}^{(t)} - Z_{k,1}'^{(t)}) x_{i,(k-1)P+l} + b \frac{1}{P} \sum_{k=1}^{K/P} \sum_{l=1}^{P} x_{i,(k-1)P+l}^T (Z_{k,2}^{(t)} - Z_{k,2}'^{(t)}) + \\
&\qquad\quad + c \sum_{k=1}^{K/P} (Z_{k,4}^{(t)} - Z_{k,4}'^{(t)}), \quad i \in [n], \, t \in [C] \\
&\qquad\quad \operatorname{tr}(Z_{k,1}^{(t)}) = Z_{k,4}^{(t)}, \ \operatorname{tr}(Z_{k,1}'^{(t)}) = Z_{k,4}'^{(t)}, \quad k \in [K/P], \, t \in [C] \\
&\qquad\quad Z_k^{(t)} \succeq 0, \ Z_k'^{(t)} \succeq 0, \quad k = [K/P], \, t \in [C] \,, \qquad (1)
\end{aligned}
$$

where the variables are the matrices $Z_k^{(t)}, Z_k'^{(t)} \in \mathbb{S}^{(f+1)\times(f+1)}$, $k = 1, \ldots, K/P$, $t = 1, \ldots, C$. f is the filter size. The output matrix is $Y \in \mathbb{R}^{n \times C}$. We have introduced the superscript (t) notation which goes from 1 to C, indicating the dimensions of the vector output.

Number of weights for this architecture is equal to

$$
\#weights = 2C \frac{K}{P} (f+1)^2 \,.
$$