Supplementary File

Binghui Chen, Weihong Deng*, Jiani Hu
Beijing University of Posts and Telecommunications
chenbinghui@bupt.edu.cn, whdeng@bupt.edu.cn, jnhu@bupt.edu.cn

1. Proof for matrixes integration

$$\mathbf{a}(\mathbf{x}) = {\mathbf{P}^1}^T(\mathbf{w}^1 \odot \mathbf{x}) + \sum_{r=2}^{R} {\mathbf{P}^r}^T(\boldsymbol{\alpha}^r \odot \mathbf{z}^r) \tag{1}$$

$$\mathbf{a}(\mathbf{x}) = {\widehat{\mathbf{w}}^1}^T \mathbf{x} + \sum_{r=2}^{R} {\widehat{\boldsymbol{\alpha}}^r}^T \mathbf{z}^r \tag{2}$$

As described in paper, Eq.1 is integrated into Eq.2 via matrix integration. Below, we will provide a simple proof.

Proof. Now, we consider integrate $\{\mathbf{P}^1, \mathbf{w}^1\}$ into a new single matrix $\widehat{\mathbf{w}}^1$. We have $\mathbf{P}^1 \in \mathbb{R}^{C\times C}$, $\mathbf{w}^1 \in \mathbb{R}^C$, $\mathbf{x} \in \mathbb{R}^C$:

$$\mathbf{P}^1 = \begin{bmatrix} p_{1,1} & \cdots & p_{1,c} \\ \vdots & & \vdots \\ p_{c,1} & \cdots & p_{c,c} \end{bmatrix}, \ \mathbf{w}^1 = \begin{bmatrix} w_1 \\ \vdots \\ w_c \end{bmatrix}, \ \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_c \end{bmatrix}$$

computing ${\mathbf{P}^1}^T(\mathbf{w}^1 \odot \mathbf{x})$, we will have:

$${\mathbf{P}^1}^T(\mathbf{w}^1 \odot \mathbf{x}) = \begin{bmatrix} p_{1,1}w_1x_1 + \cdots + p_{c,1}w_cx_c \\ \vdots \\ p_{1,c}w_1x_1 + \cdots + p_{c,c}w_cx_c \end{bmatrix} \tag{3}$$

since $p_{i,j}$ and w_i are both learnable parameters, let $\widehat{w}^1_{i,j} = p_{i,j}w_i$, we will have a new matrix $\widehat{\mathbf{w}}^1 \in \mathbb{R}^{C\times C}$, where the i-th row, j-th column value is $\widehat{w}^1_{i,j}$. Then, ${\mathbf{P}^1}^T(\mathbf{w}^1 \odot \mathbf{x}) = {\widehat{\mathbf{w}}^1}^T\mathbf{x}$.

Learning the term $p_{i,j}w_i$ is equivalent to learning a single $\widehat{w}^1_{i,j}$, as a result, we can integrate $\{\mathbf{P}^1, \mathbf{w}^1\}$ into the single matrix $\widehat{\mathbf{w}}^1$. Similarly, $\{\mathbf{P}^r, \boldsymbol{\alpha}^r\}$ can be integrated into $\widehat{\boldsymbol{\alpha}}^r$. □

2. Ablation study on λ

To show the effect of hyper-parameter λ, we conduct experiments as in Tab.1 and Tab.2. One can observe that if using an appropriate value of λ, the performances of MHN-6(IDE/PCB) can be the best. And we set λ to be 1 and 1.4 for MHN-6(IDE) and MHN-6(PCB) respectively throughout our experiments.

	DukeMTMC-ReID		Market-1501	
	R-1	mAP	R-1	mAP
$\lambda = 0$	85.5	70.8	91.8	80.0
$\lambda = 1$	87.1	75.2	93.5	83.4
$\lambda = 1.4$	**87.5**	**75.2**	**93.6**	**83.6**

Table 1. Experiments with MHN-6 (IDE).

	DukeMTMC-ReID		Market-1501	
	R-1	mAP	R-1	mAP
$\lambda = 0$	87.7	75.4	93.9	83.2
$\lambda = 1.4$	88.7	76.5	94.7	84.4
$\lambda = 2$	**89.1**	**77.1**	**95.1**	**85.0**

Table 2. Experiments with MHN-6 (PCB).

3. Layer name

In Pytorch [1], the ResNet50 is composed of *conv1-bn1-relu-maxpool-layer1-layer2-layer3-layer4-avg*, where *layerX* contains several Res-blocks. In our MHN, the HOA modules are placed after *layer2*.

References

*Corresponding author

[1] https://github.com/layumi/Person_reID_baseline_pytorch