Adaptive L2 Regularization in Person Re-Identification

Xingyang Ni, Liang Fang, Heikki Huttunen

Tampere University, Finland

Tampere University

Introduction

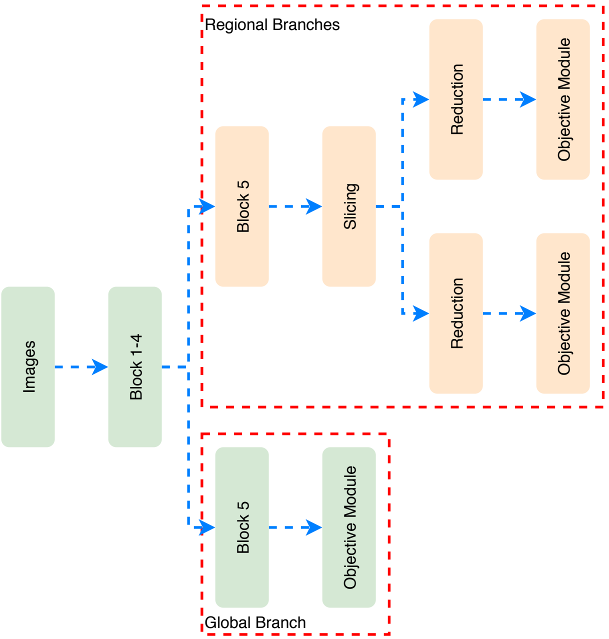
Person re-identification involves retrieving corresponding samples from a gallery set based on the appearance of a query sample across multiple cameras. It is a challenging task since images may differ significantly due to variations in factors such as illumination, camera angle and human pose.

L_2 regularization imposes constraints on the parameters of neural networks and adds penalties to the objective function during optimization. It is a commonly adopted technique which can improve the model's generalization ability. Existing approaches assign constant values to regularization factors in the training procedure, and such hyperparameters are hand-picked via hyperparameter optimization which is a tedious and time-consuming process.

In this work, our major contributions are twofold:

- We introduce an adaptive L_2 regularization mechanism, which optimizes each regularization factor adaptively as the training procedure progresses.
- With the proposed framework, we obtain state-of-the-art performance on MSMT17, which is the largest dataset for person re-identification.

Topology



Conventional L2 regularization

A neural network consists of a set of N distinct parameters,

$$P = \{\boldsymbol{w}_n \mid n = 1, \ldots, N\}, \tag{1}$$

with P containing all trainable parameters. Each $\boldsymbol{w}_n$ is an array which could be a vector, a matrix or a 3rd-order tensor.

Conventional L_2 regularization imposes an additional penalty term to the objective function, which can be formulated as follows:

$$L_\lambda(P) = L(P) + \lambda \sum_{n=1}^{N} \|\boldsymbol{w}_n\|_2^2, \tag{2}$$

where $L(P)$ and $L_\lambda(P)$ denote the original and updated objective functions, respectively. In addition, $\|\boldsymbol{w}_n\|_2^2$ refers to the square of the L_2 norm of $\boldsymbol{w}_n$, and the constant coefficient $\lambda \in \mathbb{R}_+$ defines the regularization strength.

Adaptive L2 regularization

One may wish to add penalties in a different way. Thus, it is possible to generalize even further, *i.e.*, defining a unique coefficient for each $\|\boldsymbol{w}_n\|_2^2$:

$$L_\lambda(P) = L(P) + \sum_{n=1}^{N} \left(\lambda_n \|\boldsymbol{w}_n\|_2^2\right), \tag{3}$$

where each parameter $\boldsymbol{w}_n$ is associated with an individual regularization factor $\lambda_n \in \mathbb{R}_+$. Obviously, it is infeasible to manually fine-tune those regularization factors λ_n for $n = 1, \ldots, N$ one by one, since N is in the order of 100 for models trained with ResNet50.

A straightforward extension is obtained by replacing the pre-defined constant λ_n with scalar variables which are trainable through backpropagation. However, such an approach without any constraints on λ_n will fail. Namely, setting negative values for λ_n allows naively increasing $\|\boldsymbol{w}_n\|_2^2$ so that $L_\lambda(P)$ decreases sharply.

To address the collapse problem, we apply the hard sigmoid function f which assures that the regularization factor λ_n would always have non-negative values. The regularization factor λ_n is obtained by applying the hard sigmoid on the raw parameters as

$$\lambda_n = f(\theta_n), \tag{4}$$

where $\theta_n \in \mathbb{R}$ $(n = 1, \ldots, N)$ are the trainable scalar variables. Furthermore, we introduce a hyperparameter $A \in \mathbb{R}_+$ which represents the amplitude. Hence, we get

$$\lambda_n = A f(\theta_n). \tag{5}$$

Combining Equation (3) and (5) gives

$$L_\lambda(P) = L(P) + \sum_{n=1}^{N} \left(A f(\theta_n) \|\boldsymbol{w}_n\|_2^2\right). \tag{6}$$

Results

Method	Backbone	Market mAP	Market R1	DukeMTMC mAP	DukeMTMC R1	MSMT17 mAP	MSMT17 R1
IANet	ResNet50	83.1	94.4	73.4	87.1	46.8	75.5
DGNet	ResNet50	86.0	94.8	74.8	86.6	52.3	77.2
OSNet	OSNet	84.9	94.8	73.5	88.6	52.9	78.7
BAT-net	GoogLeNet	87.4	95.1	77.3	87.7	56.8	79.5
RGA-SC	ResNet50	88.4	96.1	-	-	57.5	80.3
SCSN	ResNet50	88.5	95.7	79.0	91.0	58.5	83.8
Conventional L2	ResNet50	87.2	94.6	78.9	88.0	57.7	79.1
Adaptive L2	ResNet50	88.3	95.3	79.9	88.9	59.4	79.6
	ResNet101	88.6	94.8	80.6	89.2	61.9	81.3
	ResNet152	88.9	95.6	81.0	90.2	62.2	81.7
	ResNet152†	94.4	96.0	90.7	92.2	76.7	84.9

Histogram