Tampere University

Adaptive L_2 Regularization in Person Re-Identification

Xingyang Ni, Liang Fang, Heikki Huttunen

{xingyang.ni, liang.fang, heikki.huttunen}@tuni.fi

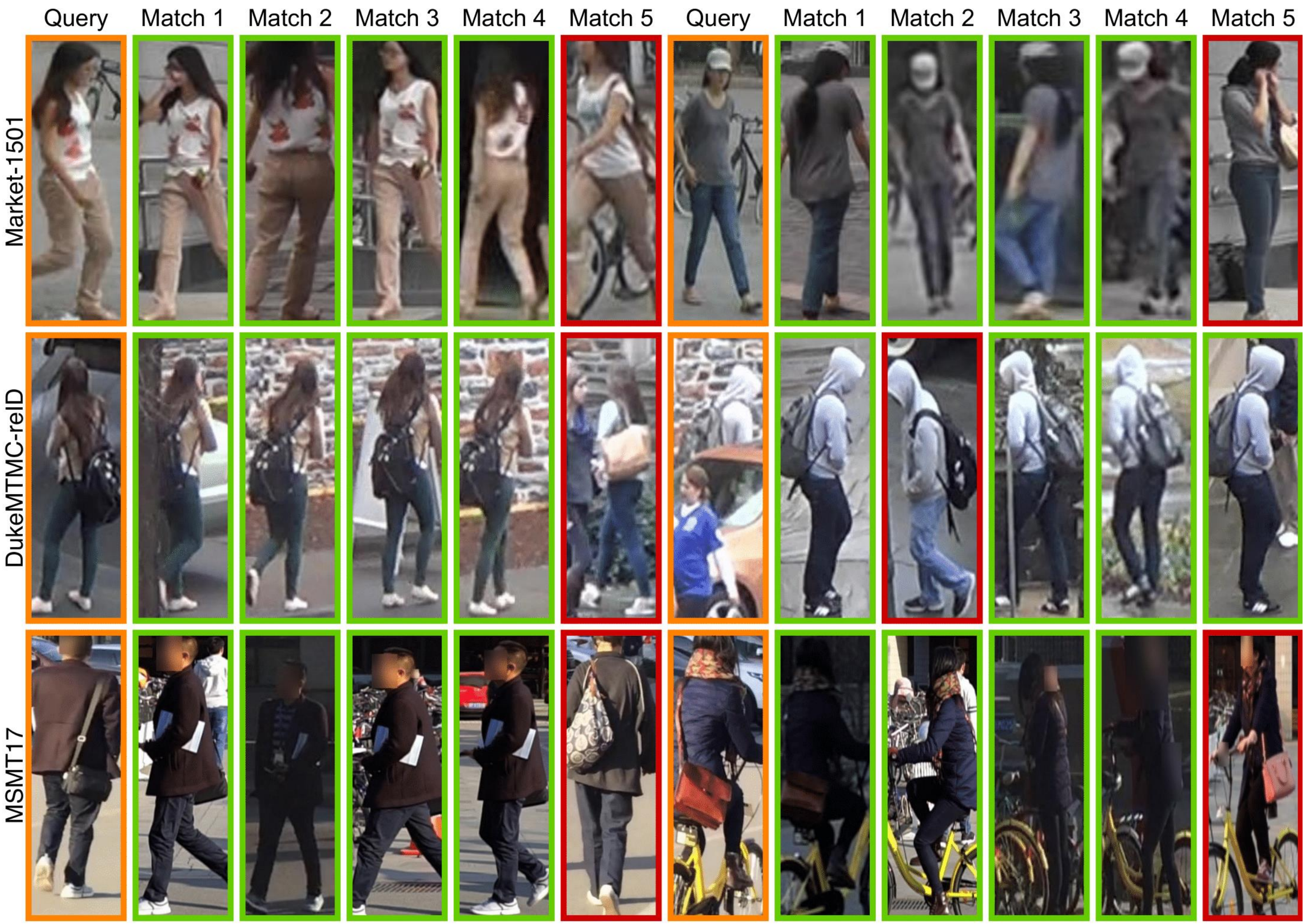
Tampere University, Finland

https://github.com/nixingyang/AdaptiveL2Regularization

Tampere University

Intro to Person Re-Identification

- It 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.



Conventional L_2 Regularization

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

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

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 L_2 Regularization

It can be generalized by defining a unique coefficient for each $\| \boldsymbol{w}_n \|_2^2$:

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

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.

Therefore, we treat them as any other learnable parameters and find suitable values from the data itself. 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. Thus the model collapses and would not learn useful feature embeddings.

Adaptive L_2 Regularization

To address the collapse problem, we apply the hard sigmoid function which assures that the regularization factor λ_n would always have non-negative values. The hard sigmoid function is defined as

$$f(x) = \begin{cases} 0, if\ x < -c \\ 1, if\ x > c \\ x/(2c) + 0.5, otherwise. \end{cases} \quad (4)$$

The regularization factor λ_n is obtained by applying hard sigmoid on the raw parameters as

$$\lambda_n = f(\theta_n), \quad (5)$$

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

$$\lambda_n = Af(\theta_n). \quad (6)$$

Combining these equations gives

$$L_\lambda(P) = L(P) + \sum_{n=1}^{N} (Af(\theta_n) \parallel \boldsymbol{w}_n \parallel_2^2)\,. \quad (7)$$

Key Takeaways

- In the literature, it is common practice to utilize hand-picked regularization factors which remain constant throughout the training procedure.
- Unlike existing approaches, the regularization factors in our proposed method are updated adaptively through backpropagation. This is achieved by incorporating trainable scalar variables as the regularization factors, which are further fed into a scaled hard sigmoid function.
- Extensive experiments validate the effectiveness of our framework. Most notably, we obtain state-of-the-art performance on MSMT17, which is the largest dataset for person re-identification.