Top-DB-Net: Top DropBlock for Activation Enhancement in Person Re-Identification

Rodolfo Quispe[†‡], Helio Pedrini[‡]

[†]Microsoft Corporation, USA — [‡]Institute of Computing, University of Campinas, Brazil

ABSTRACT

Person Re-Identification is a challenging task that aims to retrieve all instances of a query image across a system of non-overlapping cameras. Due to the various extreme changes of view, it is common that local regions that could be used to match people are suppressed, which leads to a scenario where approaches have to evaluate the similarity of images based on less informative regions. In this work, we introduce the Top-DB-Net, a method based on Top DropBlock that pushes the network to learn to focus on the scene foreground, with special emphasis on the most task-relevant regions and, at the same time, encodes low informative regions to provide high discriminability. The Top-DB-Net is composed of three streams: (i) a global stream encodes rich image information from a backbone, (ii) the Top DropBlock stream encourages the backbone to encode low informative regions with high discriminative features, and (iii) a regularization stream helps to deal with the noise created by the dropping process of the second stream, when testing the first two streams are used. Vast experiments on three challenging datasets show the capabilities of our approach against state-of-the-art methods. Qualitative results demonstrate that our method exhibits better activation maps focusing on reliable parts of the input images. The source code is available at: `https://github.com/RQuispeC/top-dropblock`.

METHODOLOGY – TOP-DB-NET

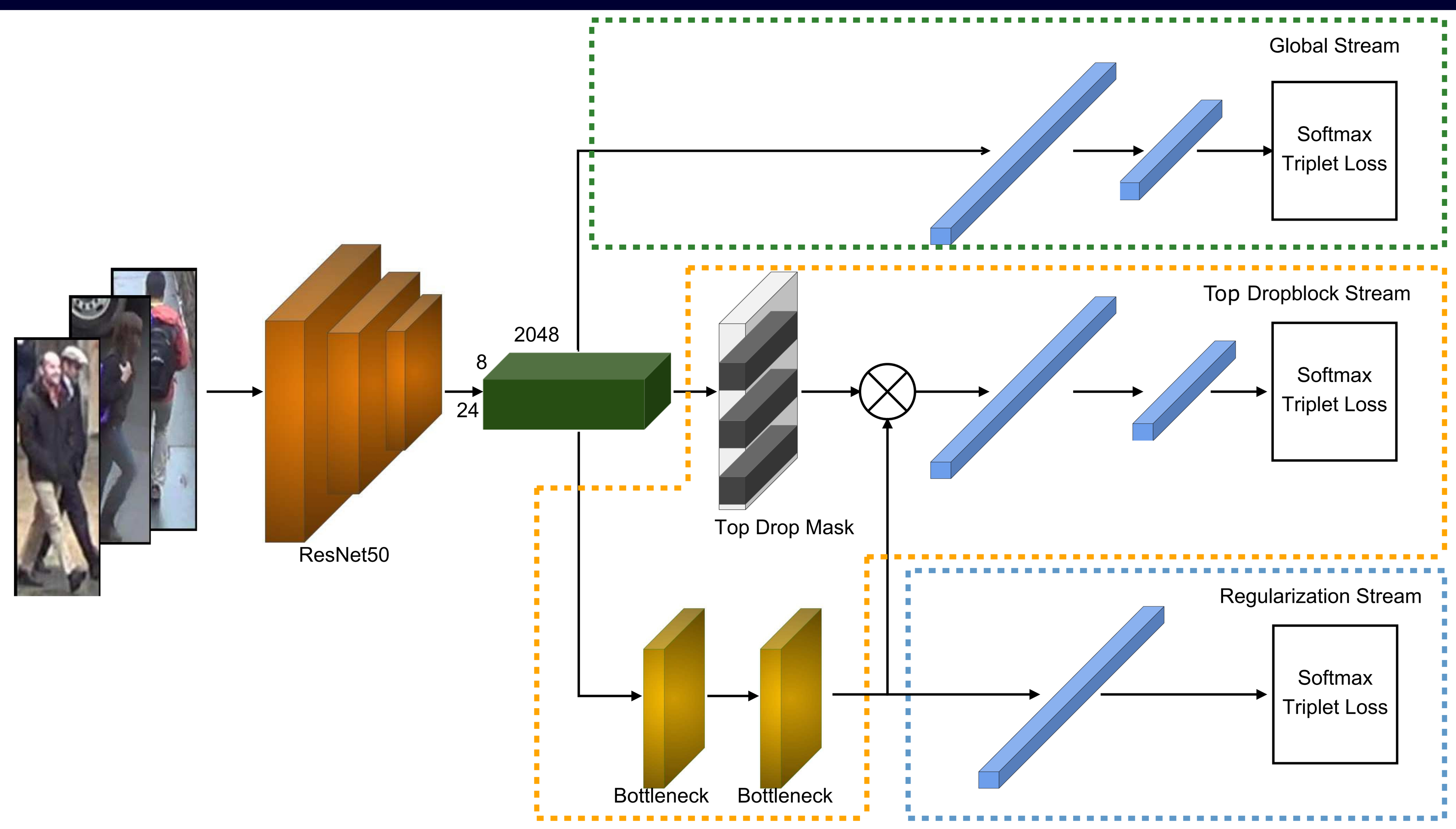

Figure 1: Proposed Top DropBlock Network (Top-DB-Net). It is composed of three streams that are able to focus on reliable parts of the input and encode low informative regions with high discriminative features for enhanced performance. It is trained using triplet loss and cross entropy. During the testing stage, the outputs of Global and Top DropBlock streams are concatenated.

METHODOLOGY – TOP DROPBLOCK

$$\mathrm{TDM}_{i,j,k} = \begin{cases} \mathbf{0}, & \text{if } \boldsymbol{r_j} \in \text{the largest values} \\ \mathbf{1}, & \text{otherwise} \end{cases}$$

$$\boldsymbol{r_j} = \frac{\sum_{k=1}^{w} \boldsymbol{A_{j,k}}}{\boldsymbol{w}}$$

$$\boldsymbol{A} = \sum_{i=1}^{c} |\boldsymbol{F_i}|^{\boldsymbol{p}}$$

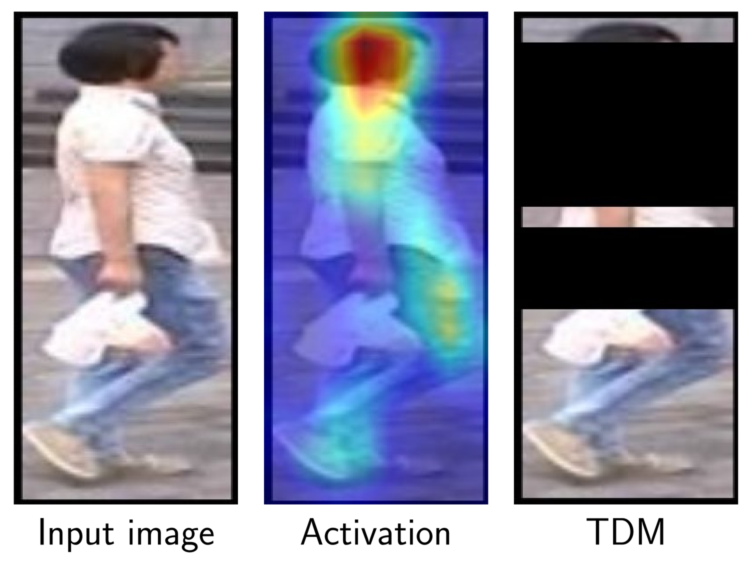

Figure 2: Input image, its activation map after epoch 120 and drop mask.

EXPERIMENTAL RESULTS

Table 1: Influence of Top-DB-Net streams and comparison with baseline.

	Market1501		DukeMTMC		CUHK03 (L)		CUHK03 (D)	
Method	mAP	R1	mAP	R1	mAP	R1	mAP	R1
no-drop Top-DB-Net	84.7	94.4	72.7	86.1	70.7	73.8	68.4	71.9
no-reg Top-DB-Net	83.9	93.9	71.1	86.1	71.4	74.6	69.4	73.5
Top-DB-Net	85.8	94.9	73.5	87.5	75.4	79.4	73.2	77.3
Baseline	85.2	94.1	73.2	85.6	72.2	74.7	70.3	73.7

Table 2: Comparison to the state-of-the-art approaches. RK stands for re-ranking. The sub-index indicates the ordinal position of this result.

	Market1501		DukeMTMC		CUHK03 (L)		CUHK03 (D)	
Method	mAP	R1	mAP	R1	mAP	R1	mAP	R1
BoT (Luo 2019)	85.9_5	94.5	76.4	86.4	–	–	–	–
PyrNet (Martinel 2019)	86.7	95.2_3	74.0	87.1	68.3	71.6	63.8	68.0
Auto-ReID (Quan 2019)	85.1	94.5	–	–	73.0	77.9	69.3	73.3
MGN (Wang 2018)	86.9_4	95.7_1	78.4_3	88.7_4	67.4	68.0	66.0	66.8
DenSem (Zhang 2019)	87.6_3	95.7_1	74.3	86.2	75.2	78.9	73.1	78.2_3
MHN (Chen 2019)	85.0	95.1_4	77.2	89.1_2	72.4	77.2	65.4	71.7
ABDnet (Chen 2019)	88.2_2	95.6_2	78.5_2	89.0_3	–	–	–	–
SONA (Xia 2019)	88.6_1	95.6_2	78.0	89.2_1	79.2_1	81.8_1	76.3_1	79.1_1
OSNet (Zhou 2019)	84.9	94.8	73.5	88.6_5	–	–	67.8	72.3
Pyramid (Zheng 2019)	88.2_2	95.7_1	79.0_1	89.0_3	76.9_2	78.9	74.8_2	78.9_2
Top-DB-Net (Ours)	85.8_6	94.9_5	73.5	87.5_6	75.4_3	79.4_2	73.2_3	77.3_4
BoT+RK (Luo 2019)	94.2_1	95.4	89.1_1	90.3_2	–	–	–	–
PyrNet+RK (Martinel 2019)	94.0	96.1_1	87.7	90.3_2	78.7_2	77.1_2	82.7_2	80.8_2
Auto-ReID+RK (Quan 2019)	94.2_1	95.4	–	–	–	–	–	–
Top-DB-Net+RK (Ours)	94.1_2	95.5_2	88.6_2	90.9_1	88.5_1	86.7_1	86.9_1	85.7_1

EXPERIMENTAL RESULTS

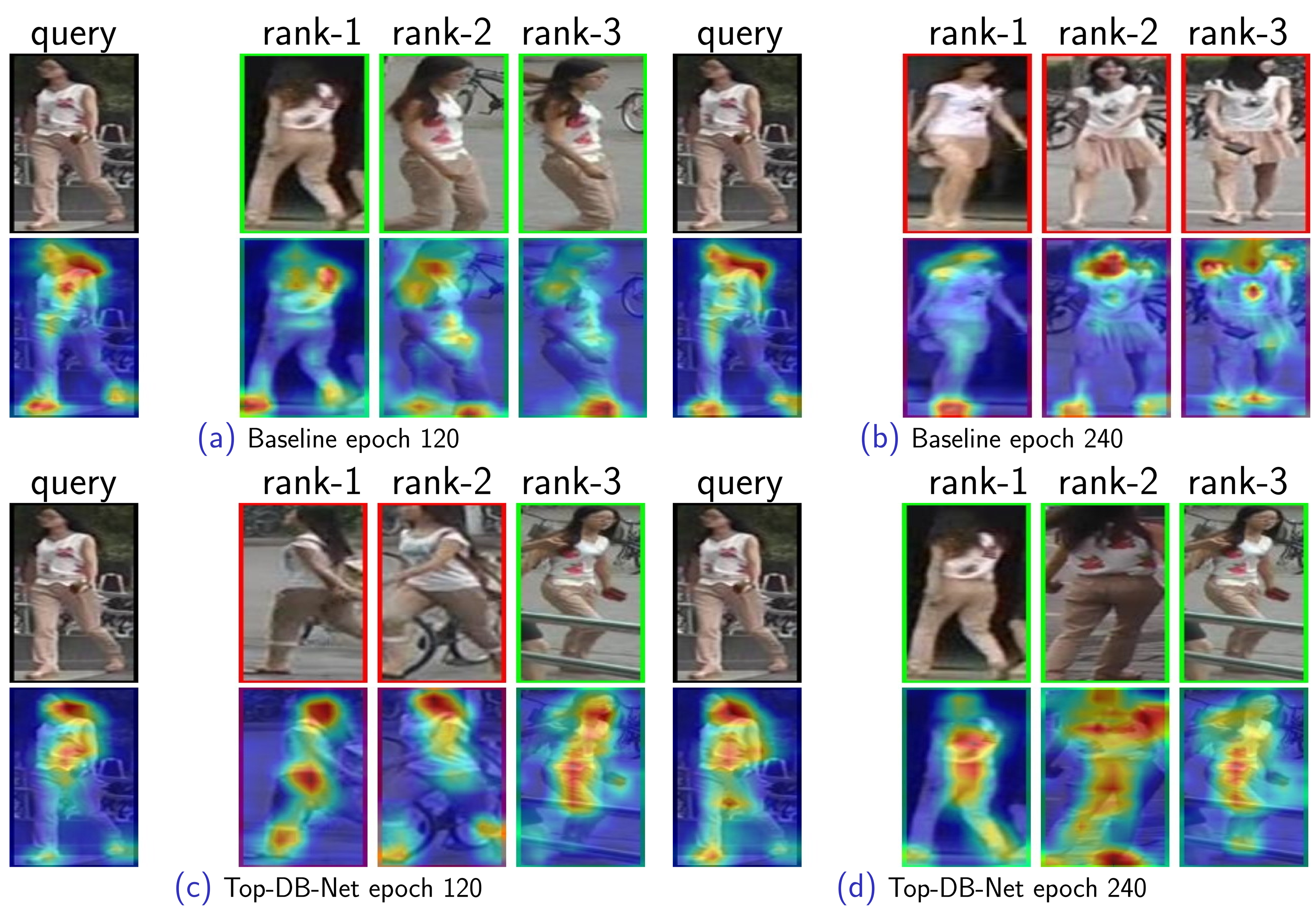

Figure 3: Comparison of activation and rank-3 evolution. The top and bottom sets show images for our baseline and our proposed method, respectively. We can see that using Top DropBlock, instead of Random DropBlock, makes the activations more spread out over the person, which helps to create a better feature representation. Correct results are highlighted in green, whereas incorrect results are highlighted in red.

CONCLUSIONS

We introduced Top-DB-Net, a network for the person re-identification problem based on Top DropBlock.

Top-DB-Net encourages the network to improve its performance by learning to generate rich encoding based on low informative regions.

Experiments on widely used datasets demonstrate the power of our method and its capability to improve activation maps.

Results suggest that methods proposed for interpretability of activation maps can help during training in ReID.