

Objectives

Person Search

- finding the query person from large-scale surveillance video.
- requiring **person detection** and **re-identification (Re-ID)**.

Frameworks

- two-step: cascading a person detection model and a person Re-ID model.
- one-step: integrating person detection and Re-ID into a unified end-to-end trainable framework.

Motivation

Challenge

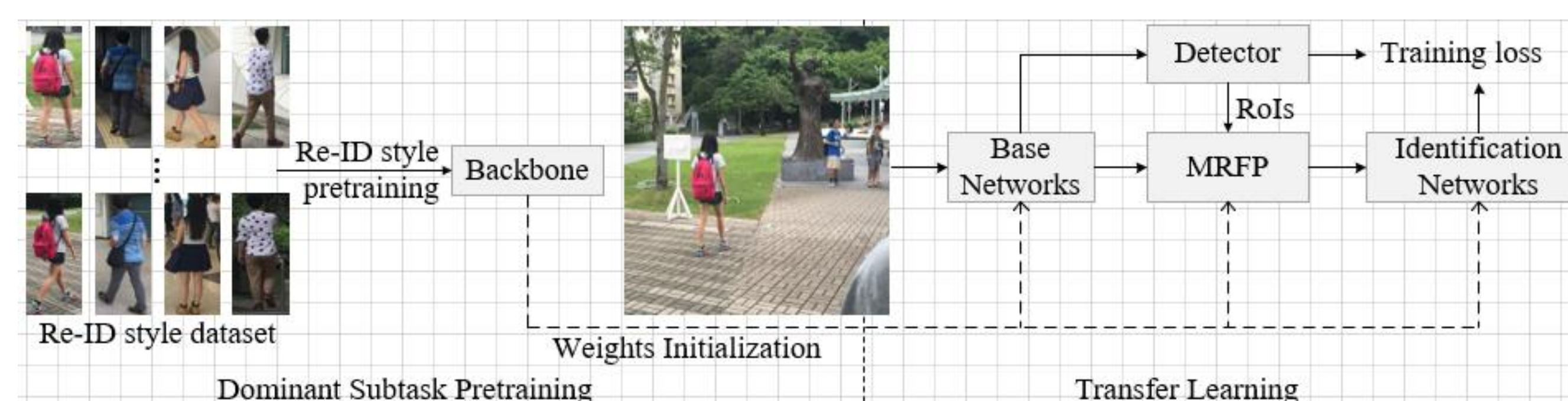
- ImageNet pretraining suffers from the large domain gap between the ImageNet dataset and target person search datasets.
- Important details to Re-ID subtask may be missing due to shared base networks.

What to do

- design a new pretraining method.
- keep more details for Re-ID subtask.

Method

Overview



Pipeline of the proposed SSP method and MRFP layer.

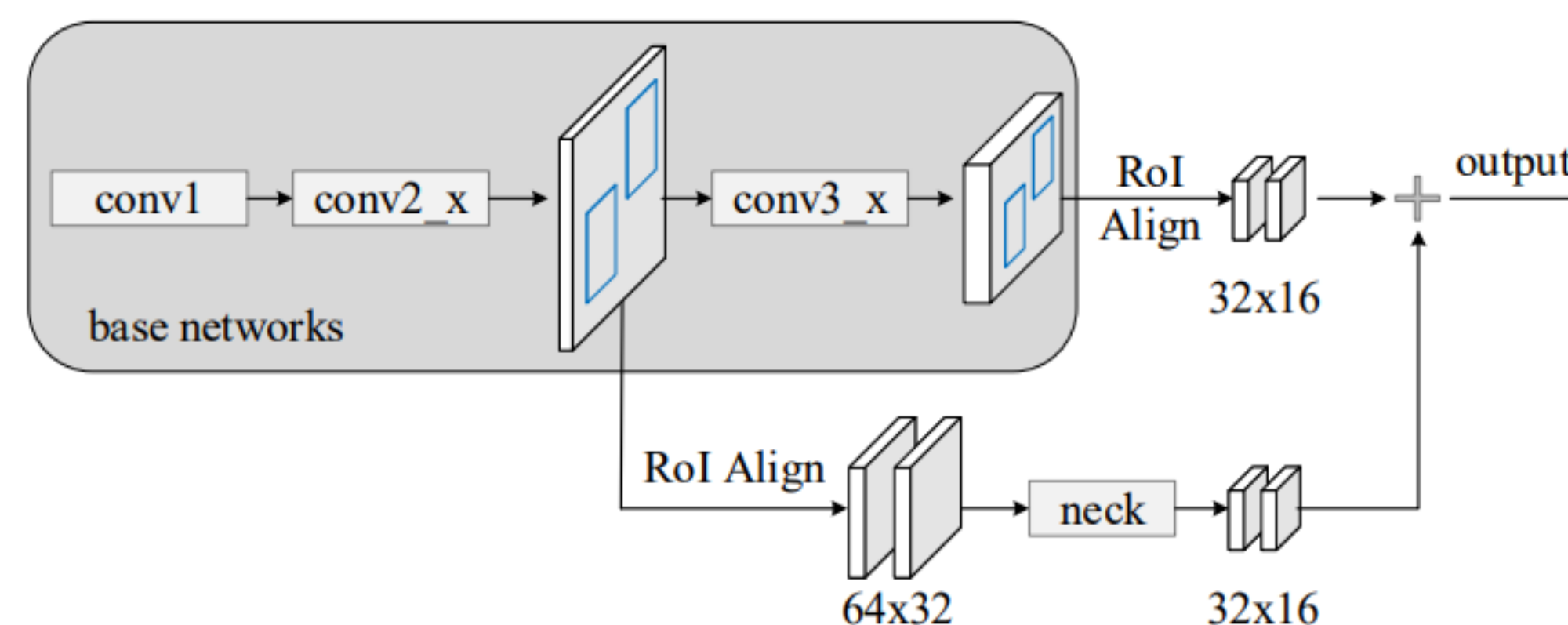


Illustration of the proposed MRFP layer.

Contributions

- Proposing a Subtask-dominated Supervised Pretraining (SSP) method to avoid the domain gap and improve person search performance.
- Proposing a Multi-level RoI Fusion Pooling to keep more important details to Re-ID subtask.

Results

Method	mAP (%)	Top-1 (%)
Baseline (random init)	27.6	75.1
Baseline (ImageNet)	48.0	86.1
Baseline (self-supervised)	46.9	85.9
Baseline (SSP)	54.0	88.0
Baseline (ImageNet)+MRFP	51.4	86.9
Baseline (SSP)+MRFP	59.6	89.7

Conclusion

- We propose the SSP method to avoid the domain gap between the source data and target data in the ImageNet pretraining method. The SSP method can make advantages of limited annotations to provide a better initialization to a CNN backbone and helps it to converge to a better final solution.
- We propose the MRFP layer to keep more details for Re-ID subtask by cropping and fusing multi-level feature maps in the shared networks.