

Rethinking the Joint Optimization in Video Coding for Machines: A Case Study



Changsheng Gao, Zhuoyuan Li, Li Li, Dong Liu*, Feng Wu

University of Science and Technology of China (USTC),
Intelligent Visual Data Coding Lab (iVC)

Data
Compression
Conference

1. Contributions

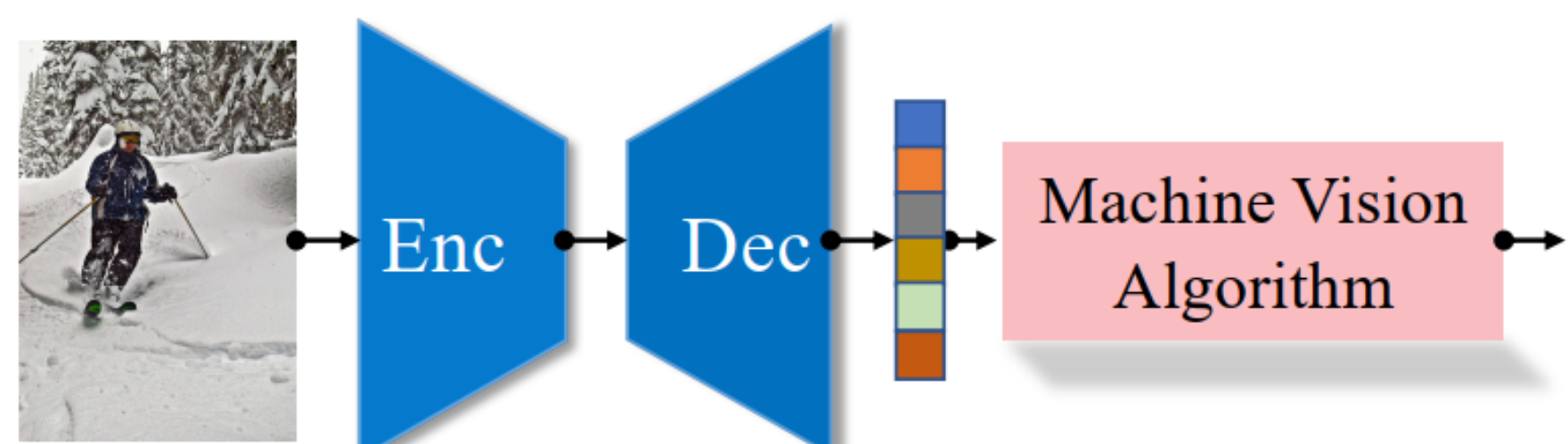
1. We unveil the counter-intuitive fact that joint optimization does not necessarily lead to performance improvement in VCM.
2. We formulate two joint optimization strategies and study their impacts on machine vision performance by analyzing the impact of compression on the extracted features.
3. We point out that tailed semantics parsing is the key to the performance improvement in VCM.

2. Motivation and Analysis

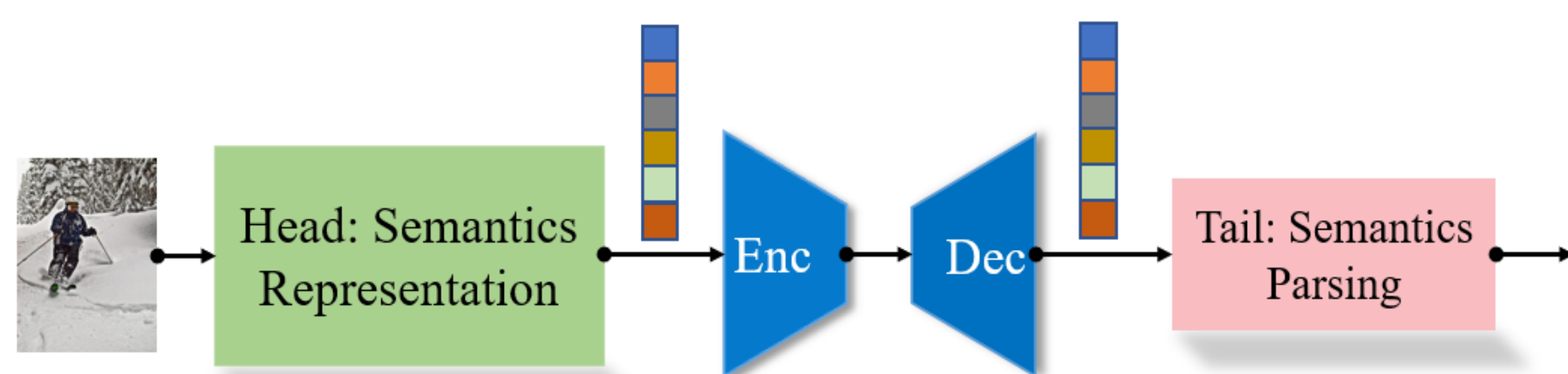
Illustration of joint optimization-based pipelines in VCM.



(a) First type of joint optimization: Machine vision algorithm is applied on the reconstructed image.



(b) Second type of joint optimization: Machine vision algorithm is applied on the reconstructed feature.



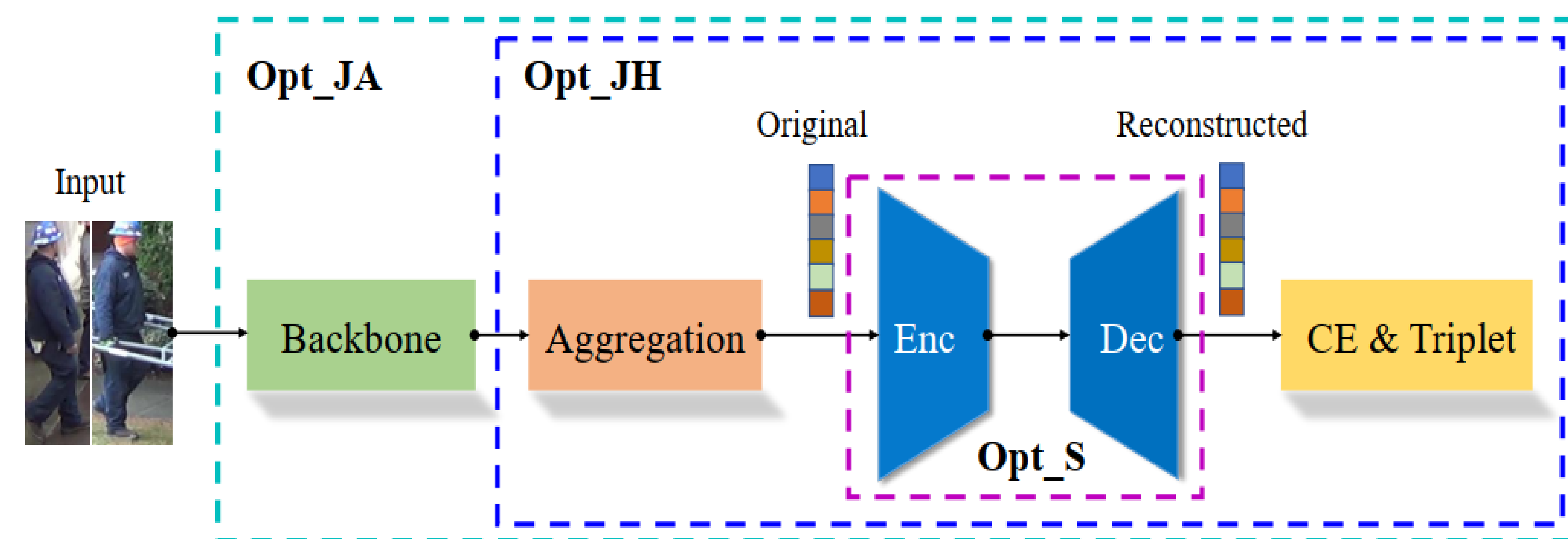
(c) Third type of joint optimization: Machine vision algorithm is divided into two parts (head and tail), and applied on the reconstructed feature of tailed semantics parsing.

In the existing joint optimization methods, the reconstructed images or features invariably undergo tailed semantics parsing to facilitate machine vision analysis, regardless of the employed joint optimization approach. **This naturally raises a question:**

where does the performance improvement originate in the existing methods—through joint optimization or tailed semantics parsing?

3. Feature Compression Methods

To figure this out, we investigate the joint optimization solely by eliminating the tailed semantics parsing module from the pipelines. To do so, we need to examine a machine vision task from which the tailed semantics parsing module can be removed without compromising joint optimization. Given this, we opt to base our investigation on the person ReID task.



Joint optimization framework of person ReID and feature compression

As the baseline, we introduce the separate optimization, denoted as **Opt_S**. For the first joint optimization approach, denoted as **Opt_JA**, we optimize all modules jointly. To evaluate the influence caused by the information capacity constraint, we propose the second optimization method, denoted as **Opt_JH**.

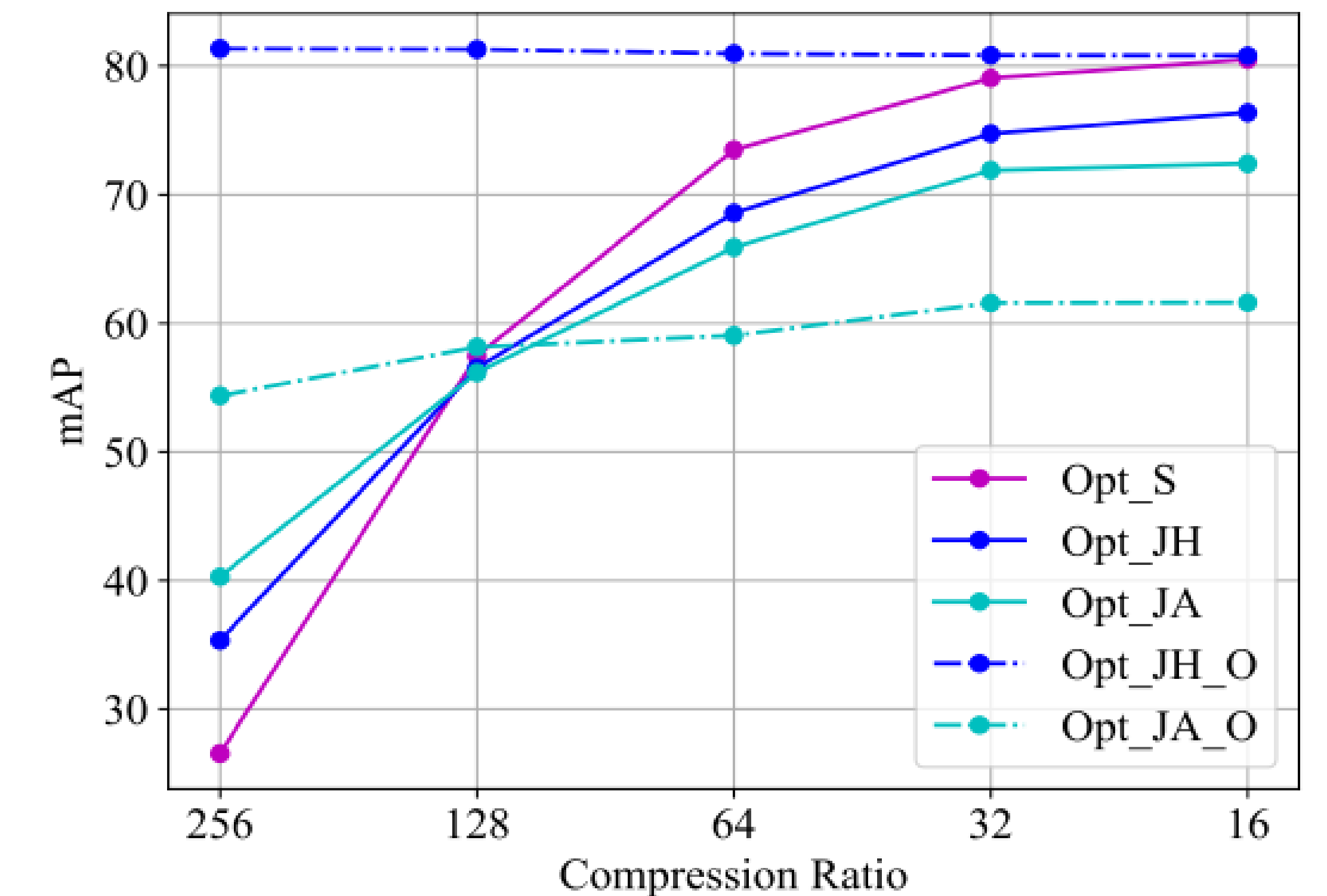
To investigate the impact of joint optimization, we introduce various information capacity constraints by adjusting the output channels of the encoder and input channels of the decoder.

4. Experiments

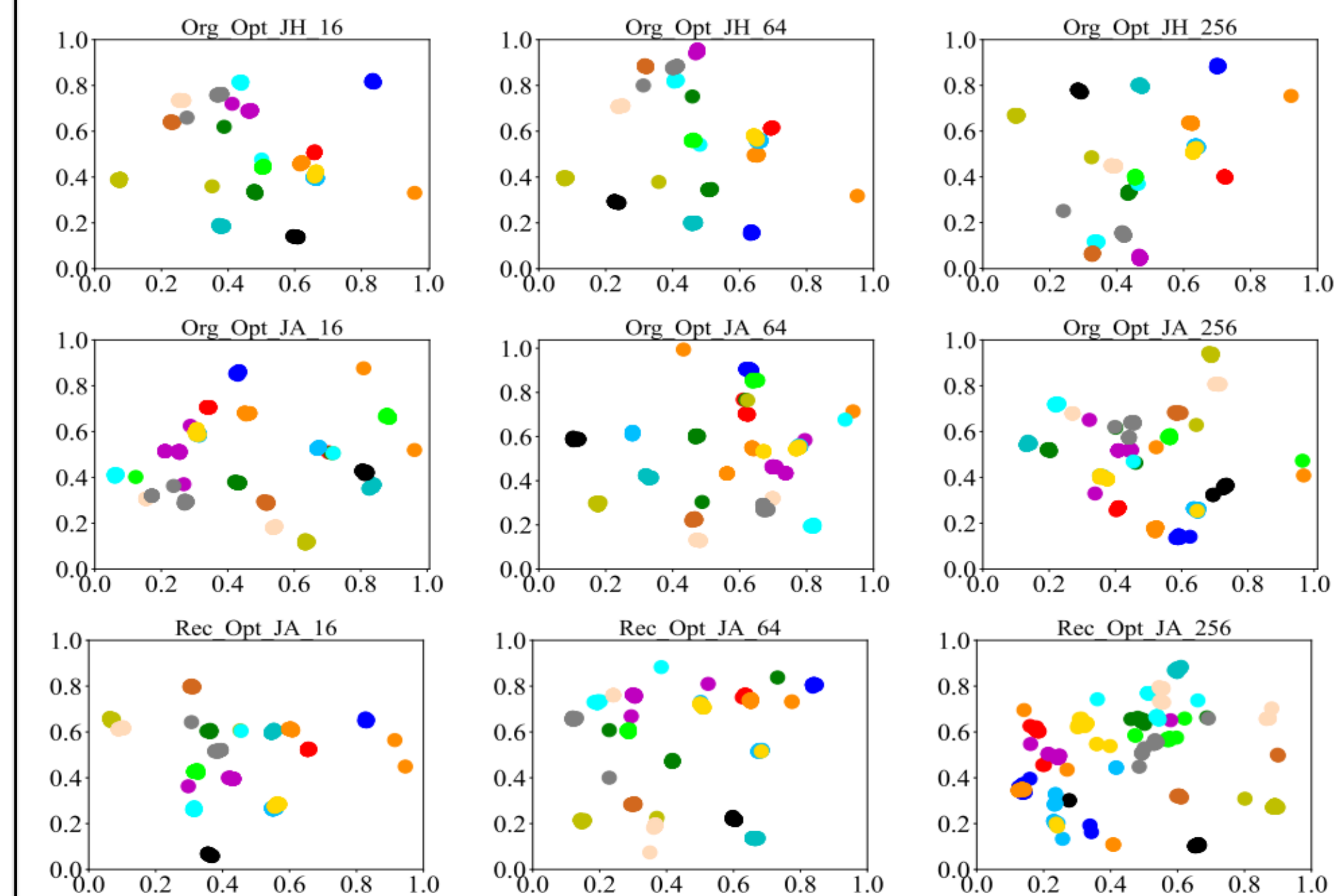
Influence of Feature Compression on Feature Extraction: We compare the performance of the original features. Based on the result, we can conclude that compression negatively impacts feature extraction and degrades the performance.

Feature Compression Analysis: Feature compression can improve the compactness and distinctiveness of original features. The phenomenon verifies decoder's ability in semantics parsing.

Person ReID performance comparison between different optimization strategies



t-SNE visualization of original features and reconstructed features for Opt_JA and Opt_JH



Performance comparison between the base decoder and the enhanced decoder

Decoder	Opt_S	Opt_JH	Opt_JA
Base	25.96	35.44	40.07
Enhanced	25.96	36.10	41.66

Tailed Semantics Parsing Analysis: The performance of different decoders is presented in Table. This finding emphasizes the significance of semantics parsing for performance improvement in joint optimization strategy.