ICASSP2026 Barcelona

Enhanced Video Compression with Context-Aware Dynamic Neural Adapter

Shaofan Sun[1], Shuangming Ma[2], Han Chen[2], Ling-Yu Duan[1], Jiaying Liu[1]

[1]Peking University

[2]Beijing Connected and Autonomous Vehicles Technology Co., Ltd

Outline

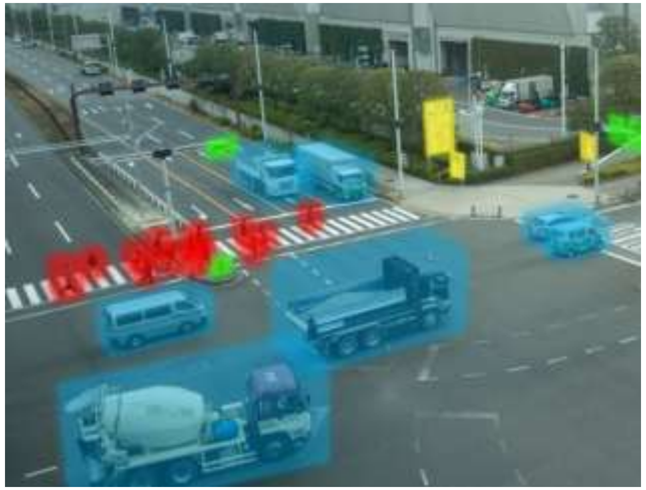
- Motivation
- Context-Aware Adapter
- Progressive Training
- Experiments
- Conclusion

Outline

- **Motivation**
- Context-Aware Adapter
- Progressive Training
- Experiments
- Conclusion

Motivation

- Growing demands for video processing and analyzing



Smart Traffic

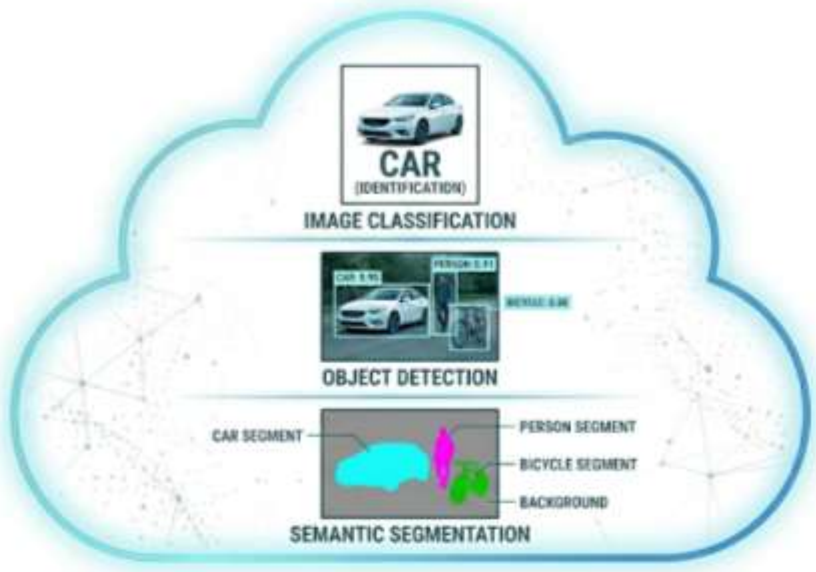
Automated Vehicle

Indoor Surveillance

Motivation

■ Machine analytics under edge-cloud collaboration

Channel



Edge

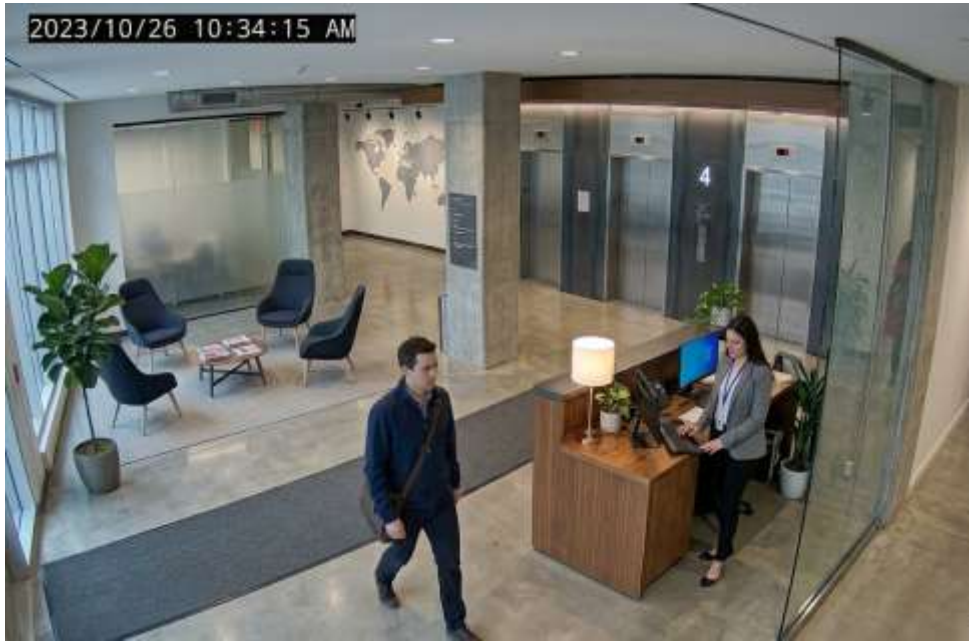
- Video data collection
- Simple computation
- Compression

Cloud

- Post-processing
- Complex analytics

Motivation

- **Simultaneous human and machine vision demands**



Human Vision

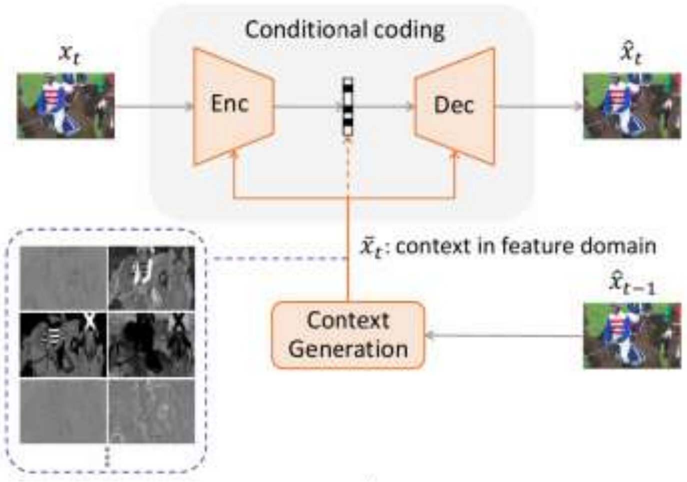
For Inspection & Judgment

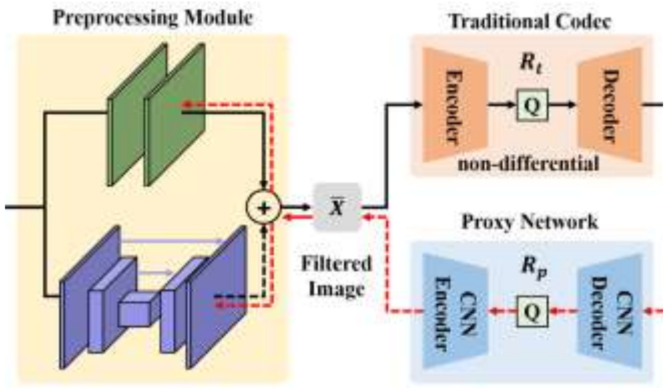
Machine Vision

For Automated Processing

Motivation

■ Various video compression methods have arisen, but...





Modern coding standards

- For pixel-level reconstruction
- Limited adaptability

End-to-end compression

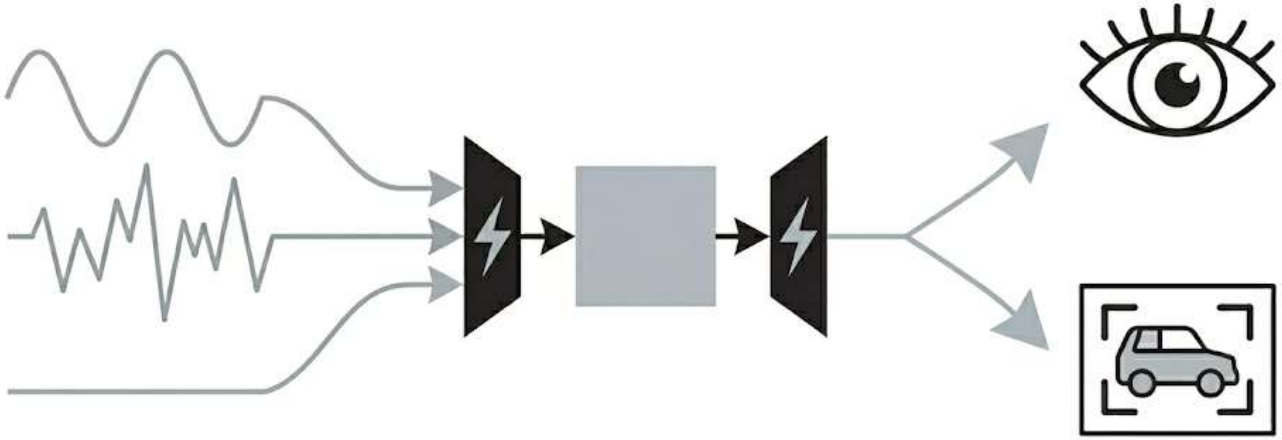
- High complexity
- Incompatibility

Deep neural network wrappers

- Limited to specific scenarios
- Poor task generalization

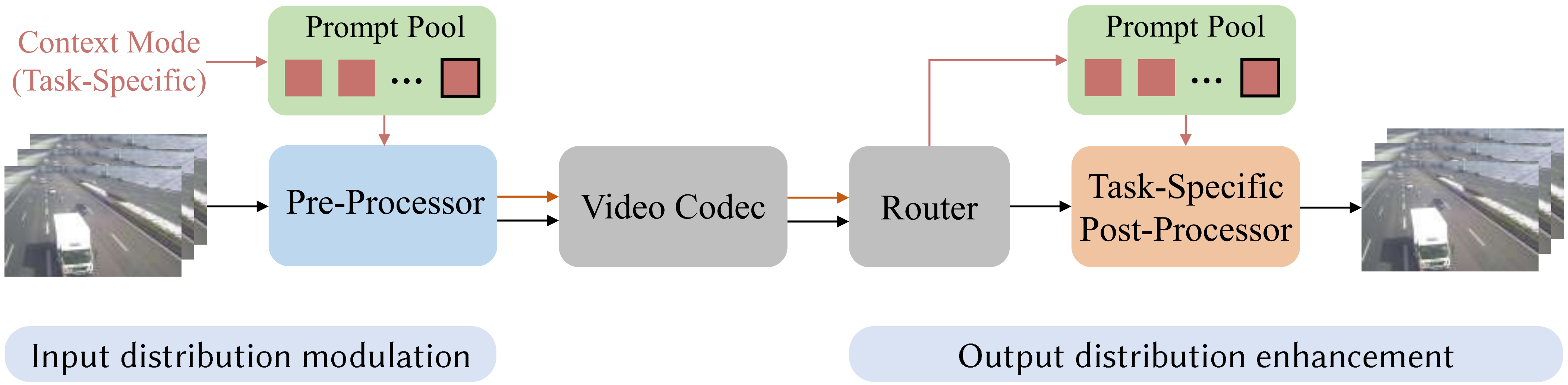
Motivation

- **A more flexible architecture is needed**
  - Adapt to various input distributions of signals
  - Handle different types of downstream tasks
  - Require low cost for edge device deployment
  - Be compatible with existing codecs



Motivation

- To decouple input/output distributions of signals from a core constant compression process, we propose **C**ontext-**A**ware dynamic neural adapter for enhanced **V**ideo **C**ompression (**CAVC**)



Outline

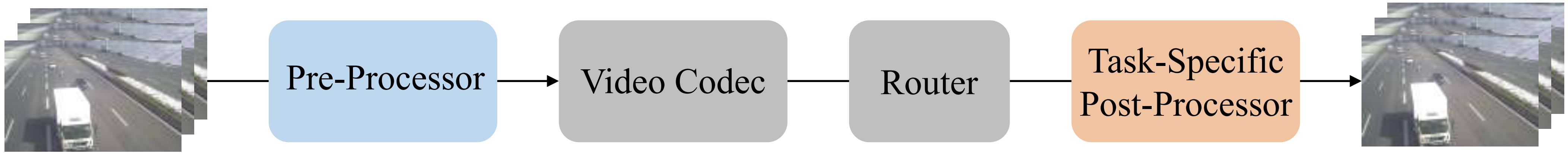
- Motivation
- **Context-Aware Adapter**
- Progressive Training
- Experiments
- Conclusion

Context-Aware Adapter

- **A dynamic neural adapter**
  - Plug-and-play with existing video codecs
    - e.g., AVC, HEVC, VVC, ...
  - No modifications required to the core compression process

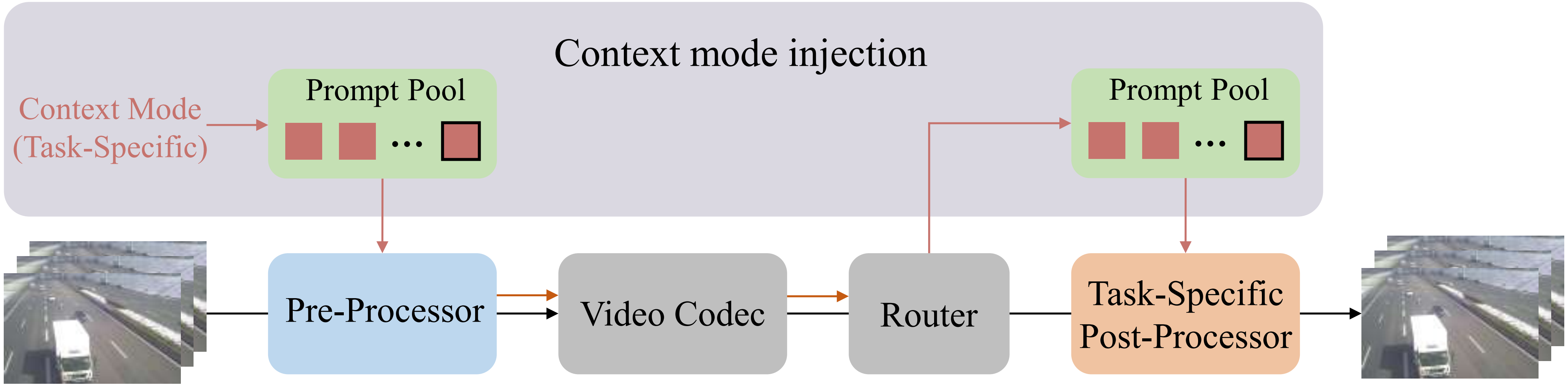
Context-Aware Adapter

- **A dynamic neural adapter**
  - Plug-and-play with existing video codecs
    - e.g., AVC, HEVC, VVC, ...
  - No modifications required to the core compression process



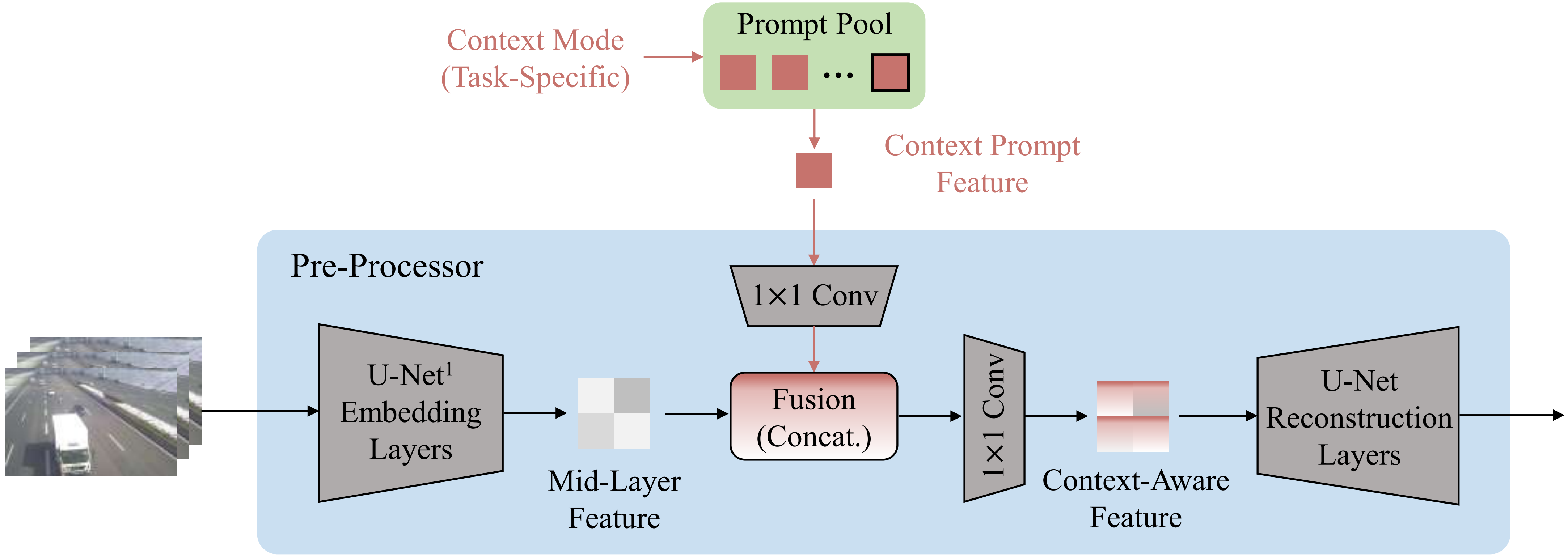
Context-Aware Adapter

- **How to control the Adapter?**
  - Context mode injection mechanism



Context-Aware Adapter

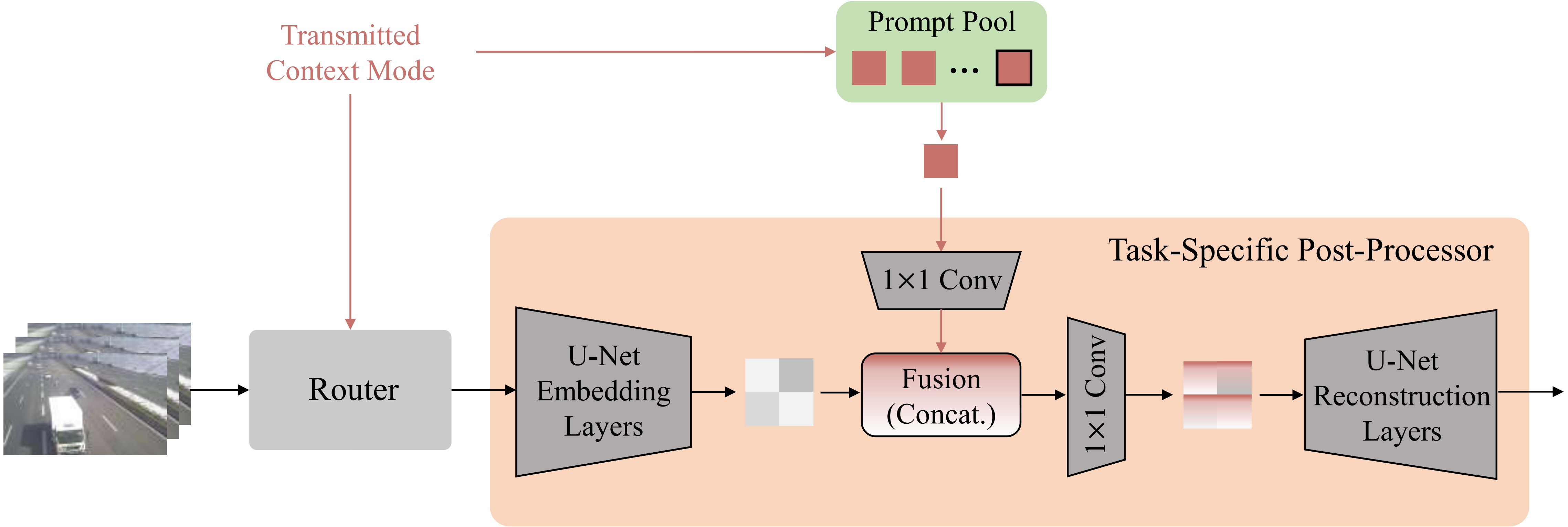
- **Context mode injection mechanism**
  - Mid-layer context feature fusion



[1]U-Net: Convolutional networks for biomedical image segmentation, MICCAI 15

Context-Aware Adapter

- **Context mode injection mechanism**
  - Mid-layer context feature fusion

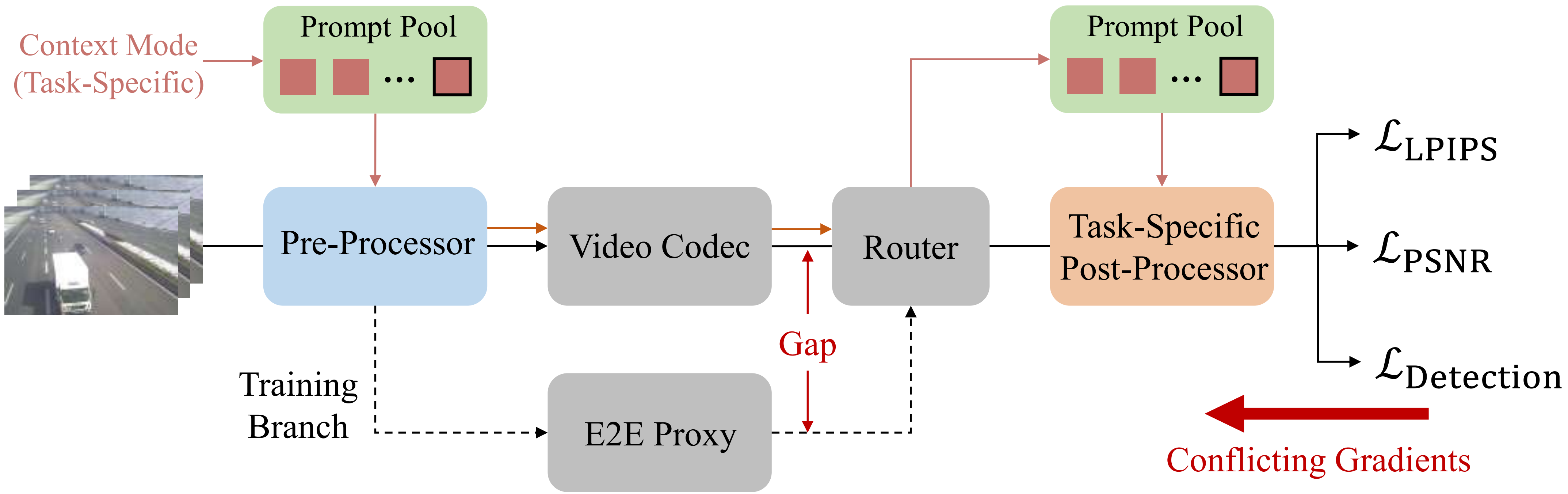


Outline

- Motivation
- Context-Aware Adapter
- **Progressive Training**
- Experiments
- Conclusion

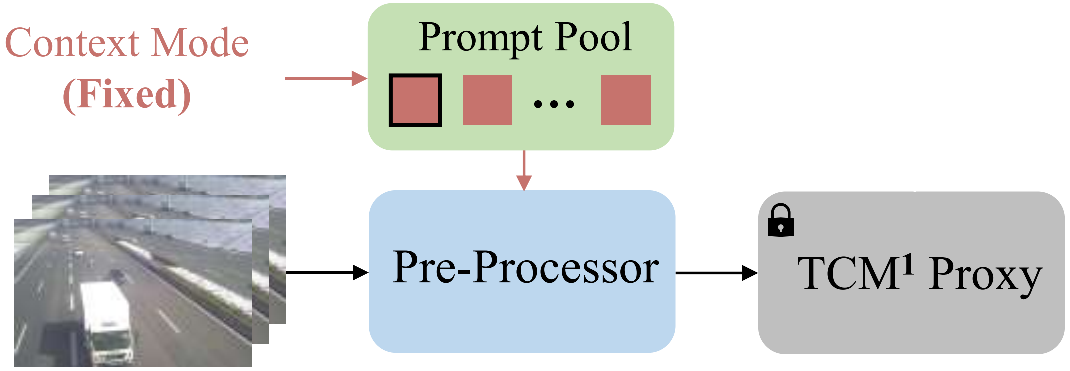
Progressive Training

- **Problems in practical training**
  - Joint multi-task training is unstable
  - Distribution gap between proxy and real codecs



Progressive Training

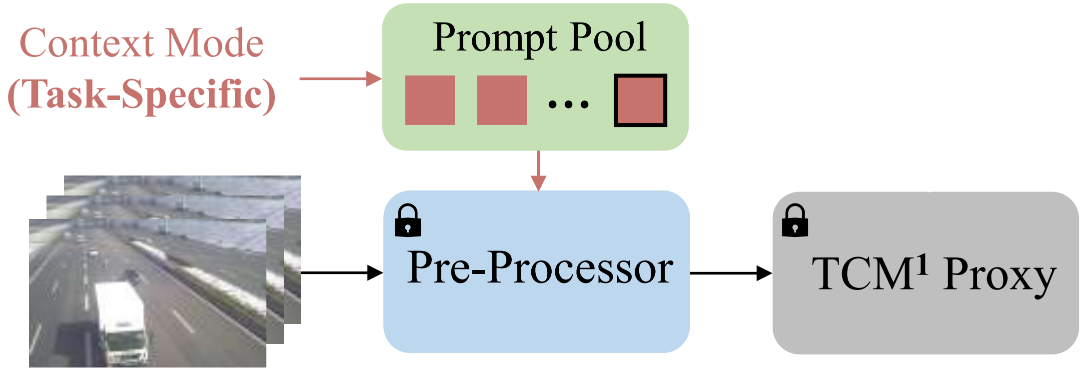
- **Joint multi-task training is unstable**
  - Initial Pre-Processor Training + Context Prompt Tuning



[1]Learned Image compression with mixed transformer-CNN architectures, CVPR 23

Progressive Training

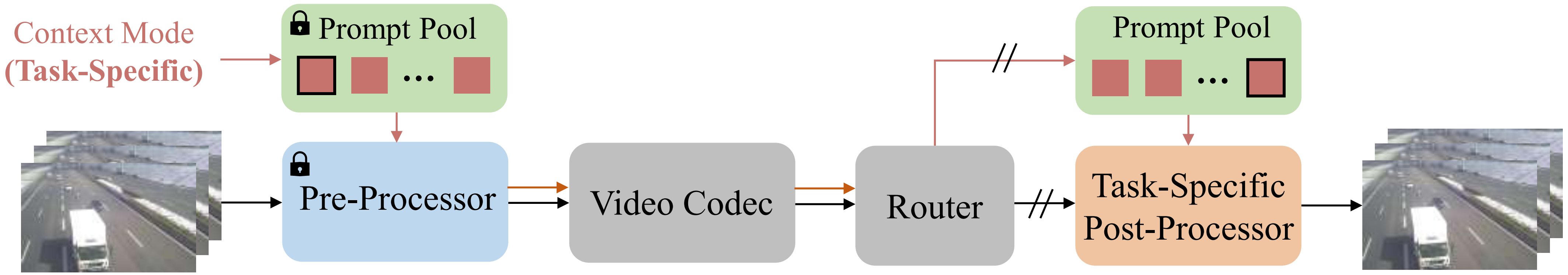
- **Joint multi-task training is unstable**
  - Initial Pre-Processor Training + Context Prompt Tuning



[1]Learned Image compression with mixed transformer-CNN architectures, CVPR 23

Progressive Training

- **Distribution gap between proxy and real codecs**
  - Post-training with real codecs



Outline

- Motivation
- Context-Aware Adapter
- Progressive Training
- **Experiments**
- Conclusion

Experiments

- **Baselines**
  - Standard video codecs
    - AVC

      Overview of the H.264/AVC Video Coding Standard, TCSVT 03
    - HEVC

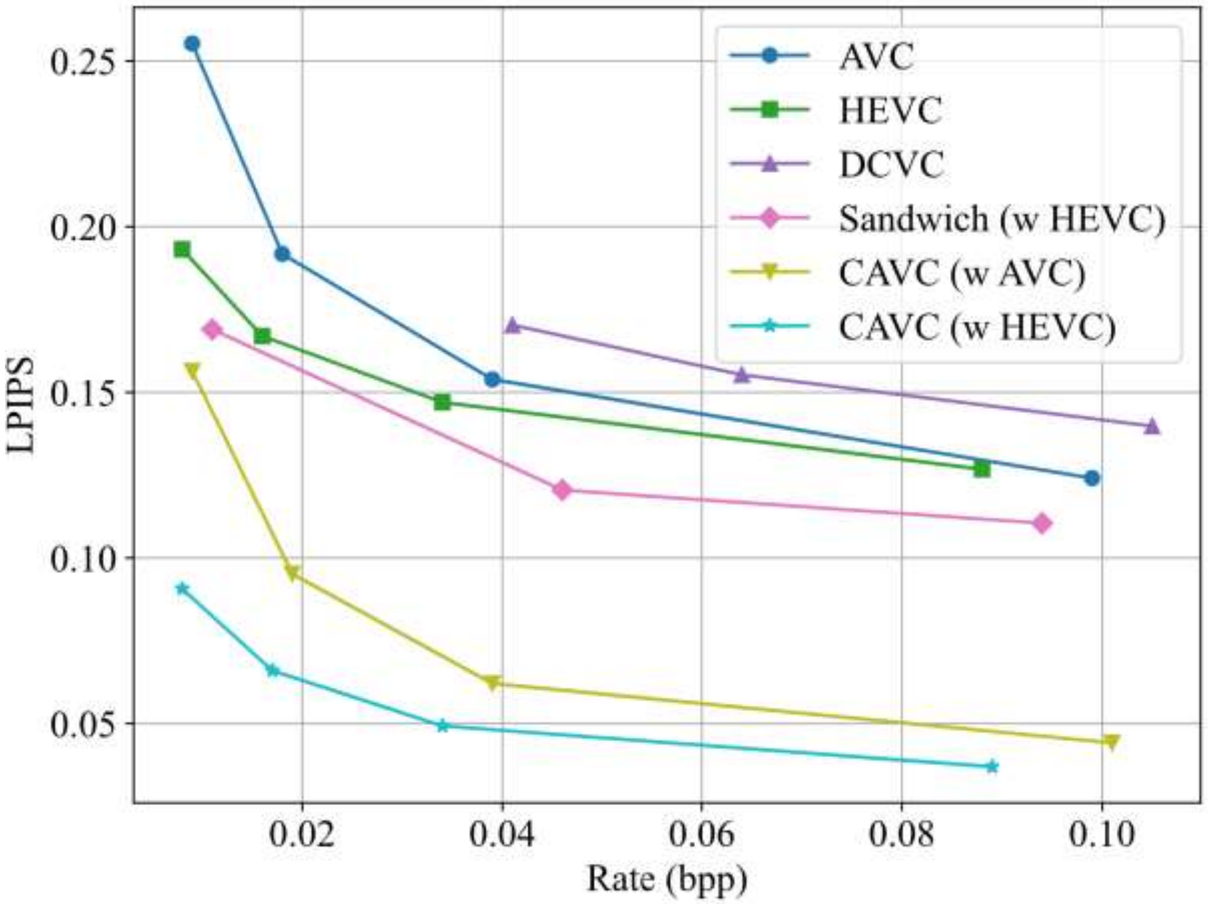
      Overview of the High Efficiency Video Coding (HEVC) Standard, TCSVT 12
  - End-to-end (E2E) method
    - DCVC

      Deep Contextual Video Compression, NeurIPS 21
  - Deep neural network wrapper
    - Sandwich

      Sandwiched Compression: Repurposing Standard Codecs with Neural Network Wrappers, arXiv 24

Experiments

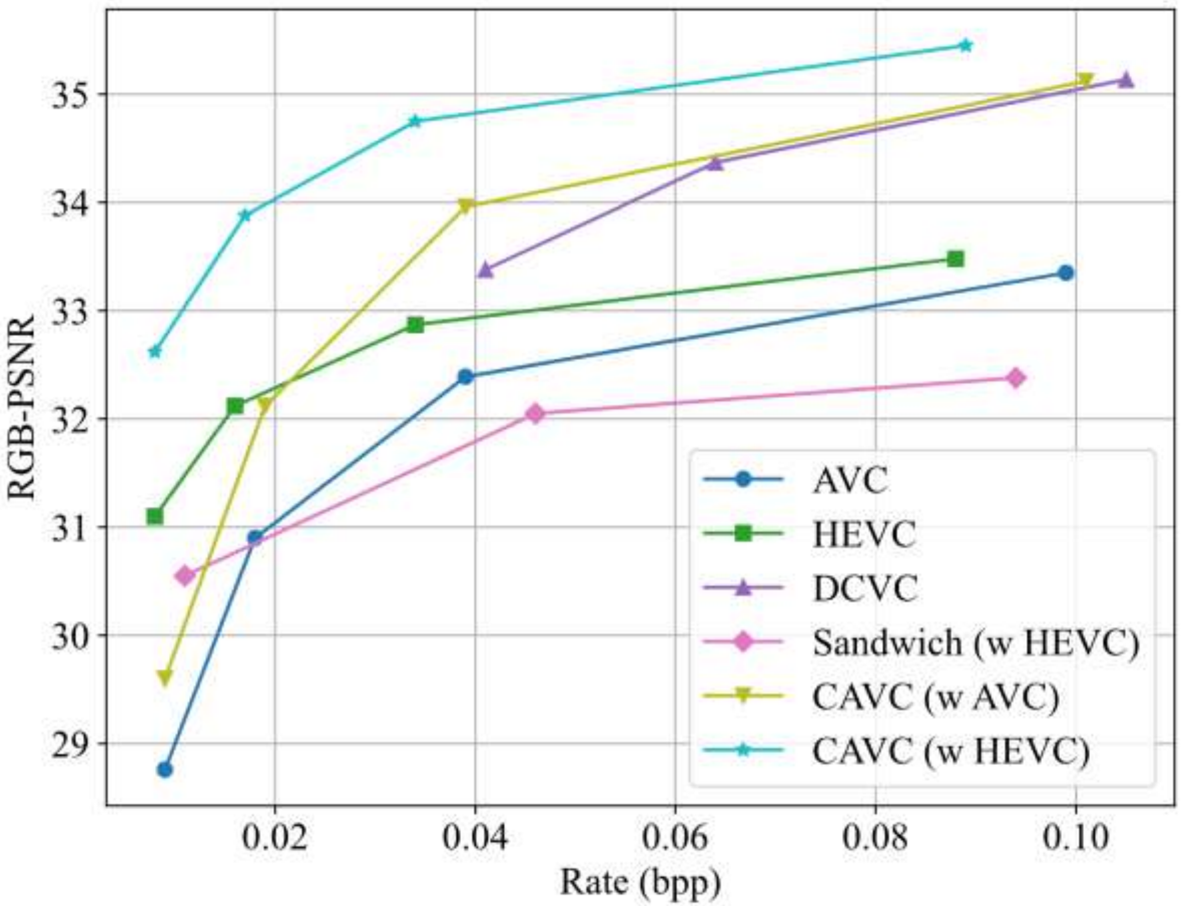
LPIPS-Alex on *UVG*[1]



[1]UVG dataset: 50/120fps 4k sequences for video codec analysis and development, ACM MMSys 20

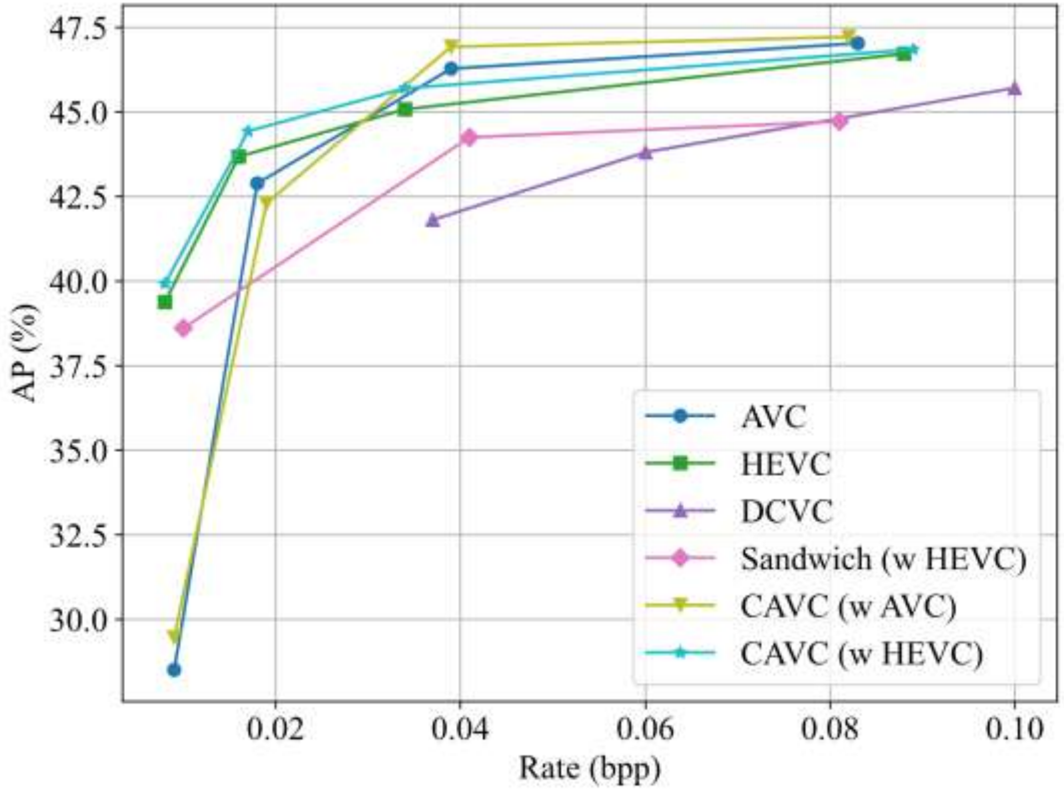
Experiments

- **PSNR on *UVG***



Experiments

■ Detection result (Average Precision) on *TUMTraf*[1]



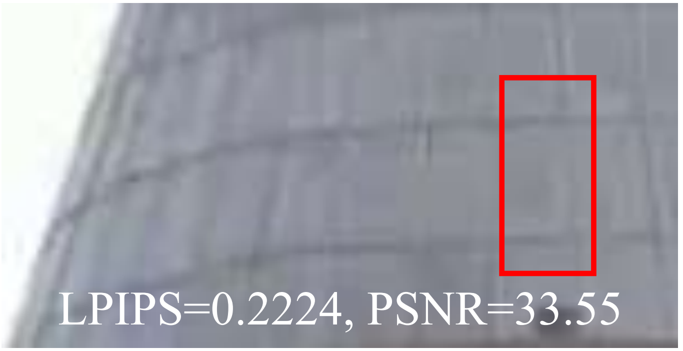
Highly Efficient for Edge Deployment!
↓ **49.9%** FLOPs (vs. DCVC) | +**1.3%** Params overhead for pre-processor

[1]A9-dataset: Multi-sensor infrastructure-based dataset for mobility research, IV 22

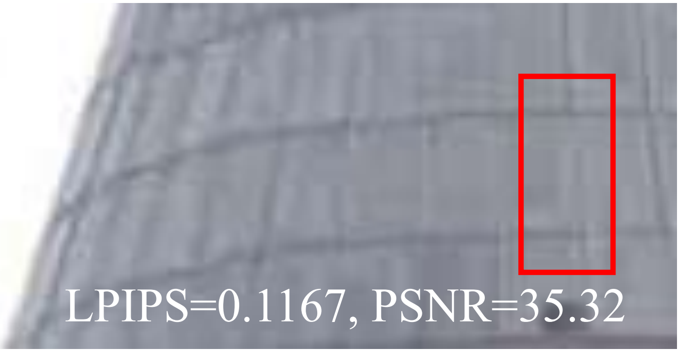
Experiments

- **Visual comparisons**
  - Human vision (LPIPS & PSNR)

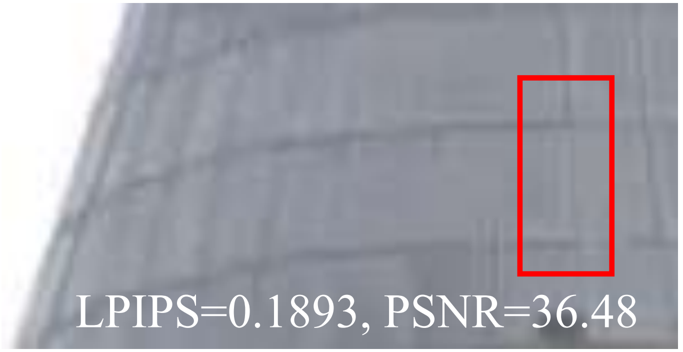
Ground Truth



HEVC bpp=0.038



CAVC (w HEVC, LPIPS) bpp=0.033

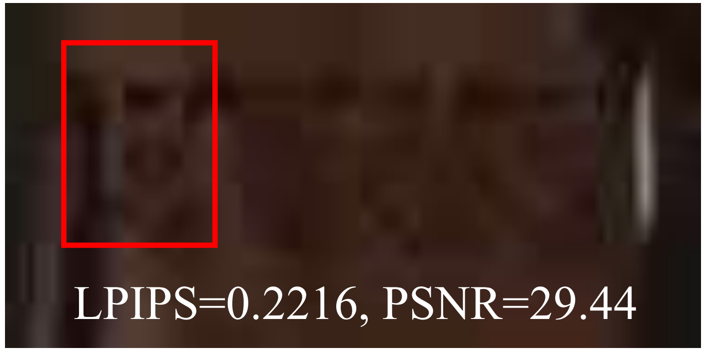


CAVC (w HEVC, PSNR) bpp=0.033

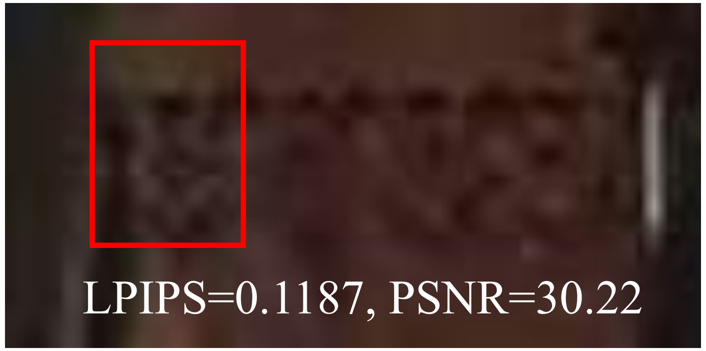
Experiments

- **Visual comparisons**
  - Human vision (LPIPS & PSNR)

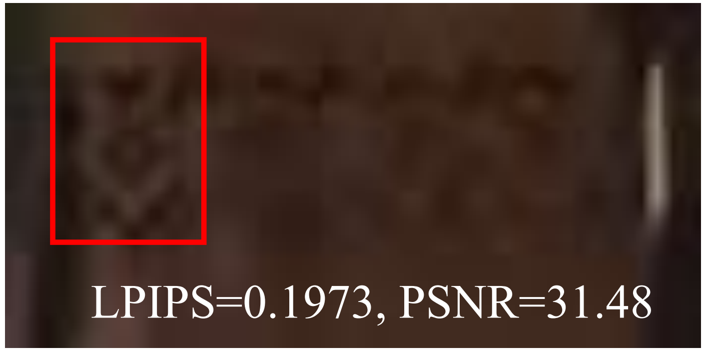
Ground Truth



HEVC bpp=0.038



CAVC (w HEVC, LPIPS) bpp=0.033



CAVC (w HEVC, PSNR) bpp=0.033

Experiments

- **Visual comparisons**
  - Machine vision (object detection)

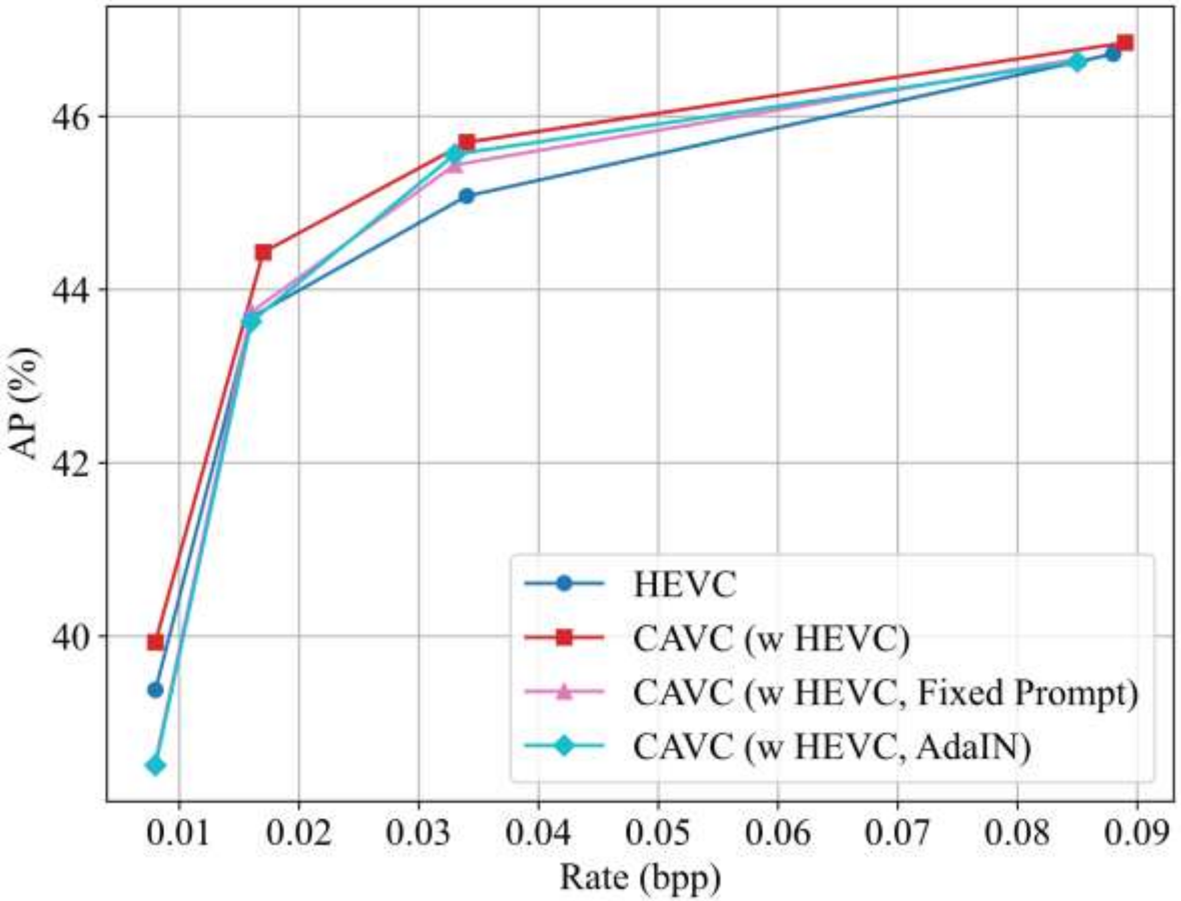
Ground Truth

HEVC bpp=0.016

CAVC (w HEVC) bpp=0.016

Experiments

- **Ablation study**
  - Role of context prompts
  - Context injection strategy



Outline

- Motivation
- Context-Aware Adapter
- Progressive Training
- Experiments
- **Conclusion**

Conclusion

- **A novel video compression framework with dynamic adapter**
  - **Flexible**

    Context Mode Injection mechanism for multi-task adaptation
  - **Compatible**

    Plug-and-play with standard video codecs
  - **Comprehensive**

    Superior performance for both human vision and machine tasks (e.g., BD-rate -76.4% on PSNR and -8.3% on detection for HEVC)
  - **Efficient**

    Reduces FLOPs by 49.9% (vs. DCVC) with 1.3% parameter overhead per task

北京大学
PEKING UNIVERSITY

北京车网
BJCW

ICASSP2026
Barcelona

Thank You!

Project Page

Group Page

Questions? Feel free to contact: **carefree_sun@stu.pku.edu.cn**