

JVET-V0090

# AHG11: Neural network based temporal processing

AHG 11: Neural-network-based video coding

Byeongdoo Choi, Zeqiang Li, Wei Wang, Wei Jiang, Xiaozhong Xu, Shan Liu



# Outline

Motivation

Proposed Method

- Proposed Framework
- Proposed Network Model

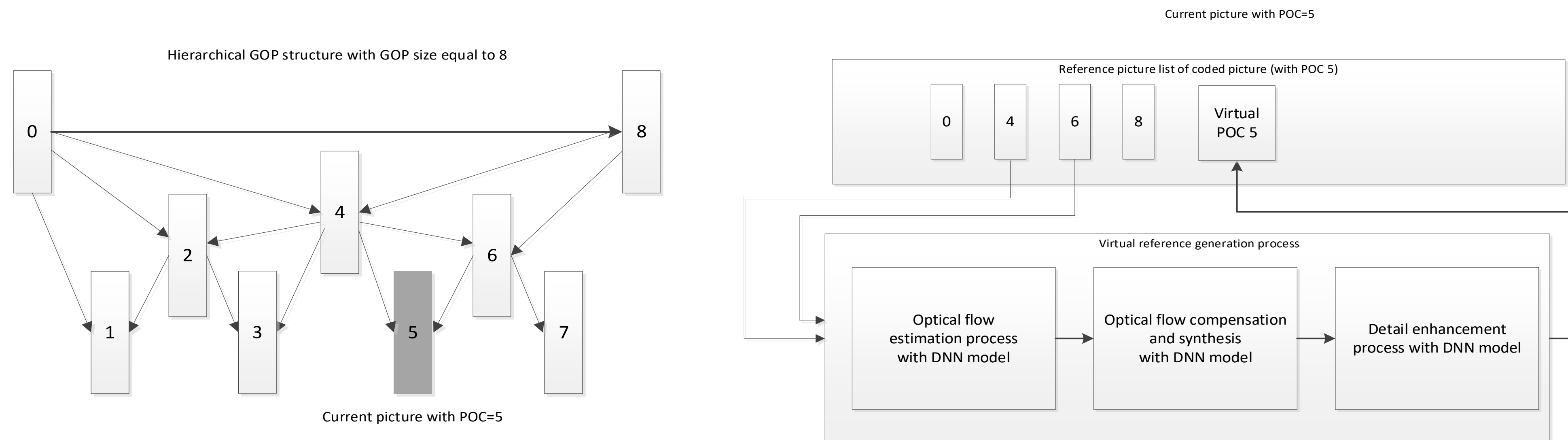
Network info for Training & Inference

Experimental Results

Conclusion

# Prior work

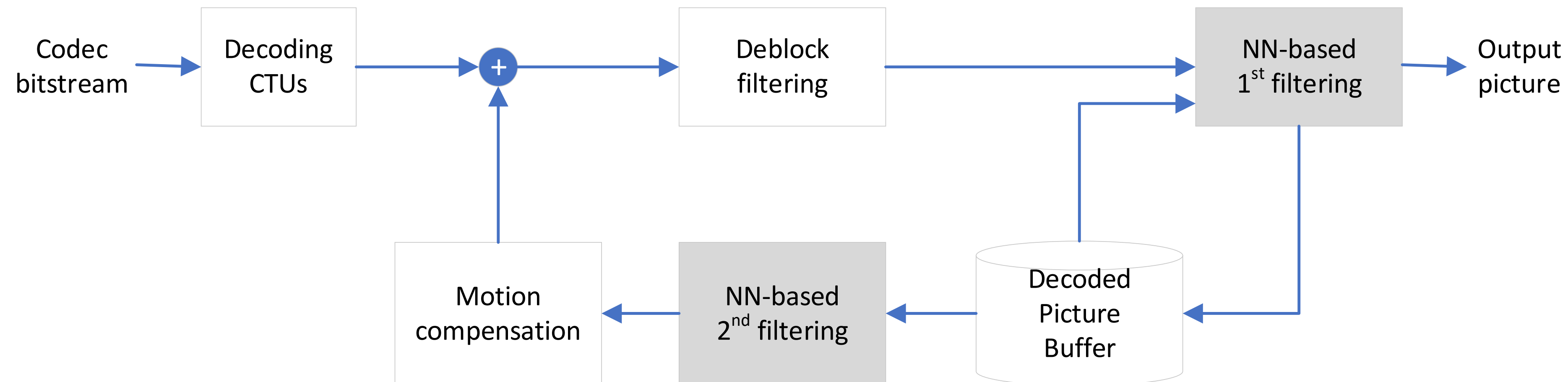
- Framework for virtual reference
  - Virtual reference is generated through NN-based temporal processing



Can we apply the NN-based temporal processing for improvements of both output pictures and reference pictures?

# Proposed Method

- Two step NN-based temporal processing
  - 1<sup>st</sup> step: detail enhancement of output pictures
  - 2<sup>nd</sup> step: prediction accuracy enhancement of motion-compensated prediction



# Proposed Method

## Two step NN-based temporal processing

### ■ - 1<sup>st</sup> step

Input: the current decoded picture and two neighboring reference pictures

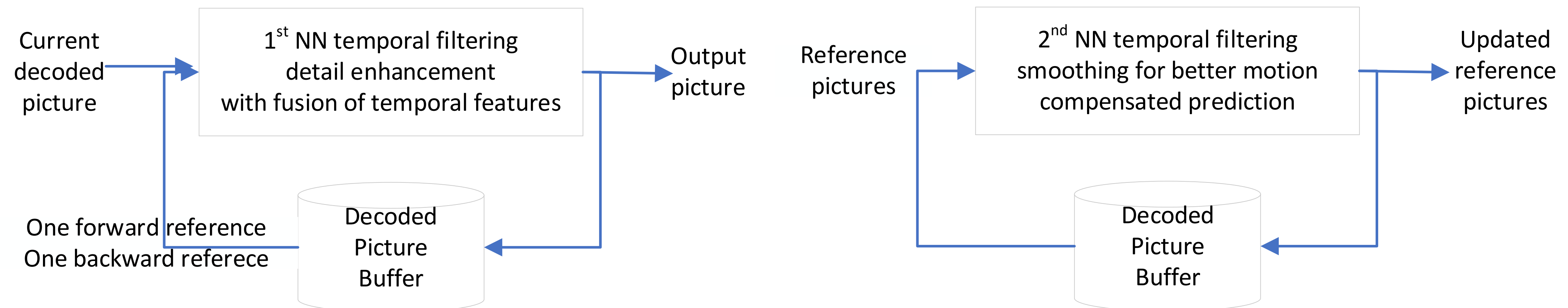
If appropriate reference pictures not present,  
the current picture is processed with image-based processing.

Output: an output picture with detail improvement

### - 2<sup>nd</sup> step

Input: two forward and one backward reference pictures

Output: a reference picture for motion compensation



# Network information for training and inference

## ■ Training Stage

| Network Information in Training Stage |                                                                                            |                                                                             |
|---------------------------------------|--------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| Mandatory                             | Device Type                                                                                | CPU: Intel(R) Xeon(R) CPU E5-2698 v4 @ 2.20GHz<br>GPU: Tesla V100-DGXS-32GB |
|                                       | Framework:                                                                                 | Pytorch 1.15.1 + CUDA 10.1 + cuDNN 7.6.4                                    |
|                                       | Number of GPUs per Task                                                                    | 4                                                                           |
|                                       |                                                                                            |                                                                             |
|                                       | Epoch:                                                                                     | 60                                                                          |
|                                       | Batch size:                                                                                | 4                                                                           |
|                                       | Training time:                                                                             | 6 days                                                                      |
|                                       | Training data information:                                                                 | BVI-DVC                                                                     |
|                                       | Training configurations for generating compressed training data (if different to VTM CTC): | Same as VTM CTC<br>QP=22, 27, 32, 37, 42                                    |
| Optional                              |                                                                                            |                                                                             |
|                                       | Number of iterations                                                                       |                                                                             |
|                                       | Patch size                                                                                 | 480x480, 64*64                                                              |
|                                       | Loss Functions                                                                             | L1Loss                                                                      |
|                                       | Optimizer:                                                                                 | ADAM                                                                        |
|                                       | Other information:                                                                         |                                                                             |
|                                       |                                                                                            |                                                                             |

## ■ Inference Stage

| Network Information in Inference Stage |                         |                                                           |
|----------------------------------------|-------------------------|-----------------------------------------------------------|
|                                        | Device Type             | CPU: Intel Core i9-9900x<br>CPU @ 3.60GHz, 32GB<br>Memory |
|                                        | Framework:              | Pytorch 1.15.1 + CUDA 10.1 + cuDNN 7.6.4                  |
|                                        | Number of GPUs per Task | 0                                                         |
|                                        | Number of model         | 5 x (Image-based model) +<br>5 x (video-based model)      |
|                                        | -                       | -                                                         |
|                                        | Memory Parameter (MB)   | 5 x 80.9MB                                                |
|                                        | MAC (Giga)              | 2.03e-3 Giga/pixel                                        |
|                                        |                         |                                                           |
|                                        |                         |                                                           |
| Optional                               |                         |                                                           |
|                                        | Total Memory (MB)       |                                                           |
|                                        | Batch size:             | 1                                                         |
|                                        | Other information:      | -                                                         |
|                                        |                         |                                                           |

# Experimental Results

## ■ Proposed NN-temporal processing versus VTM 11

BD-rate and Coding Time for NN-based Video Coding Tool Testing in Inference Stage (Tested 33frames with GOP=32 and patch size 256x256)

|         | Random Access Over VTM-11.0 |             |             |                     |                     |                     |                      |
|---------|-----------------------------|-------------|-------------|---------------------|---------------------|---------------------|----------------------|
|         | BD-rate (Y)                 | BD-rate (U) | BD-rate (V) | Encoding Time (CPU) | Decoding Time (CPU) | Encoding Time (GPU) | Decodin g Time (GPU) |
| Class B | -3.14%                      | -4.41%      | -6.13%      | 2299.43%            | 5597619.22%         |                     |                      |
| Class C | -4.69%                      | -8.07%      | -6.54%      | 3804.11%            | 4466192.03%         |                     |                      |
| Class D | -6.87%                      | -5.82%      | -6.09%      | 3611.28%            | 3361108.43%         |                     |                      |

---

# Conclusion

- The NN-based temporal processing, which had been developed for virtual reference, was applied for
  - 1) improvement of output picture details, and
  - 2) better prediction accuracy of motion compensation.
- The simulation results show some promising gains, as compared to the virtual reference.
- Further research is planned, considering the harmonization with other (NN) tools (e.g. NN-inloop filtering).

Thanks!  
Q&A