

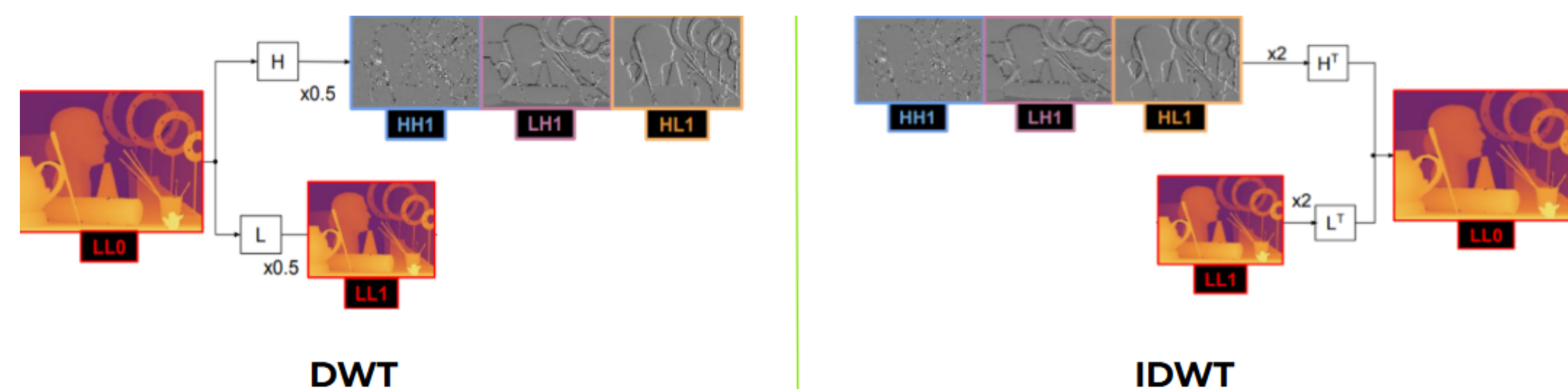
Wavelet Packing for Self-Supervised Monocular Depth Estimation

Ayoub Rhim¹, Lei Qin^{2*}, Rachid Benmokhtar² and Xavier Perrotton²

École des Ponts ParisTech¹ and Valeo Brain Division, Créteil, France²
ayoub.rhim@outlook.com, lei.qin@valeo.com, rachid.benmokhtar@valeo.com and xavier.perrotton@valeo.com

Motivation

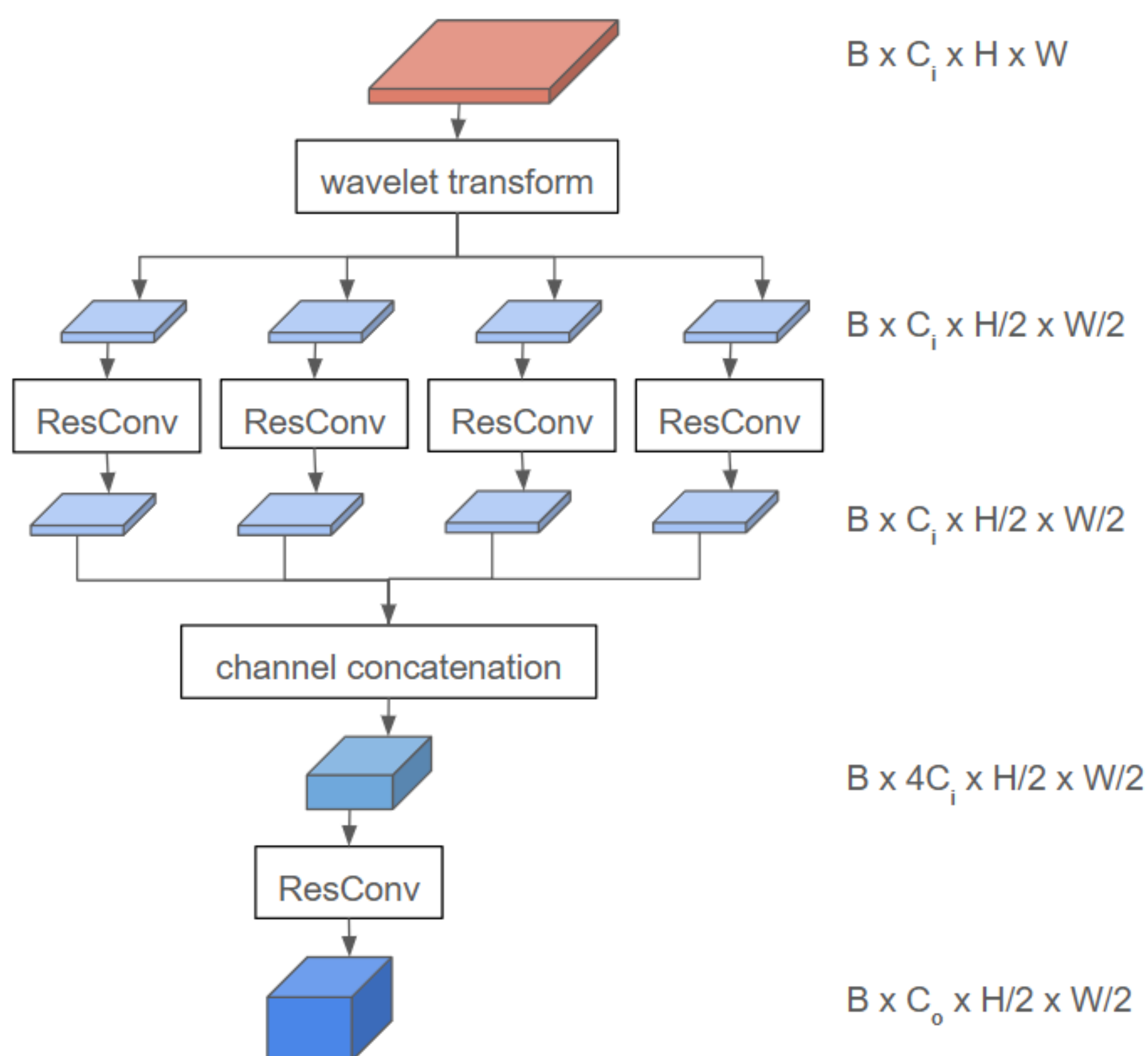
Dense depth prediction requires preserving detailed information in the network encoder and faithful reconstruction in the decoder.



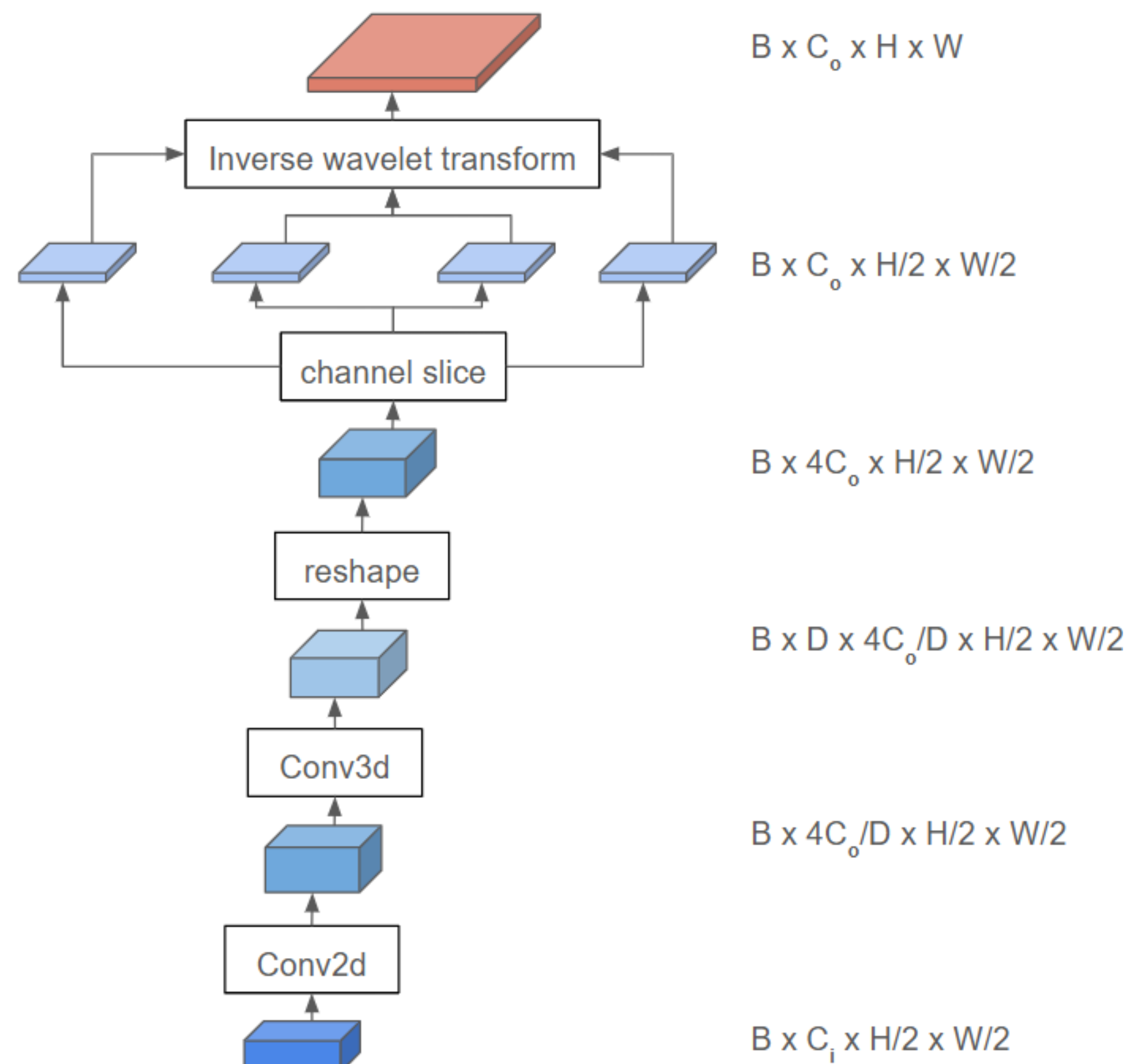
- DWT and IDWT involve only algebraic operations which are differentiable
- Use DWT for lossless information packing in the encoder
- Use IDWT for lossless information unpacking in the decoder

Method

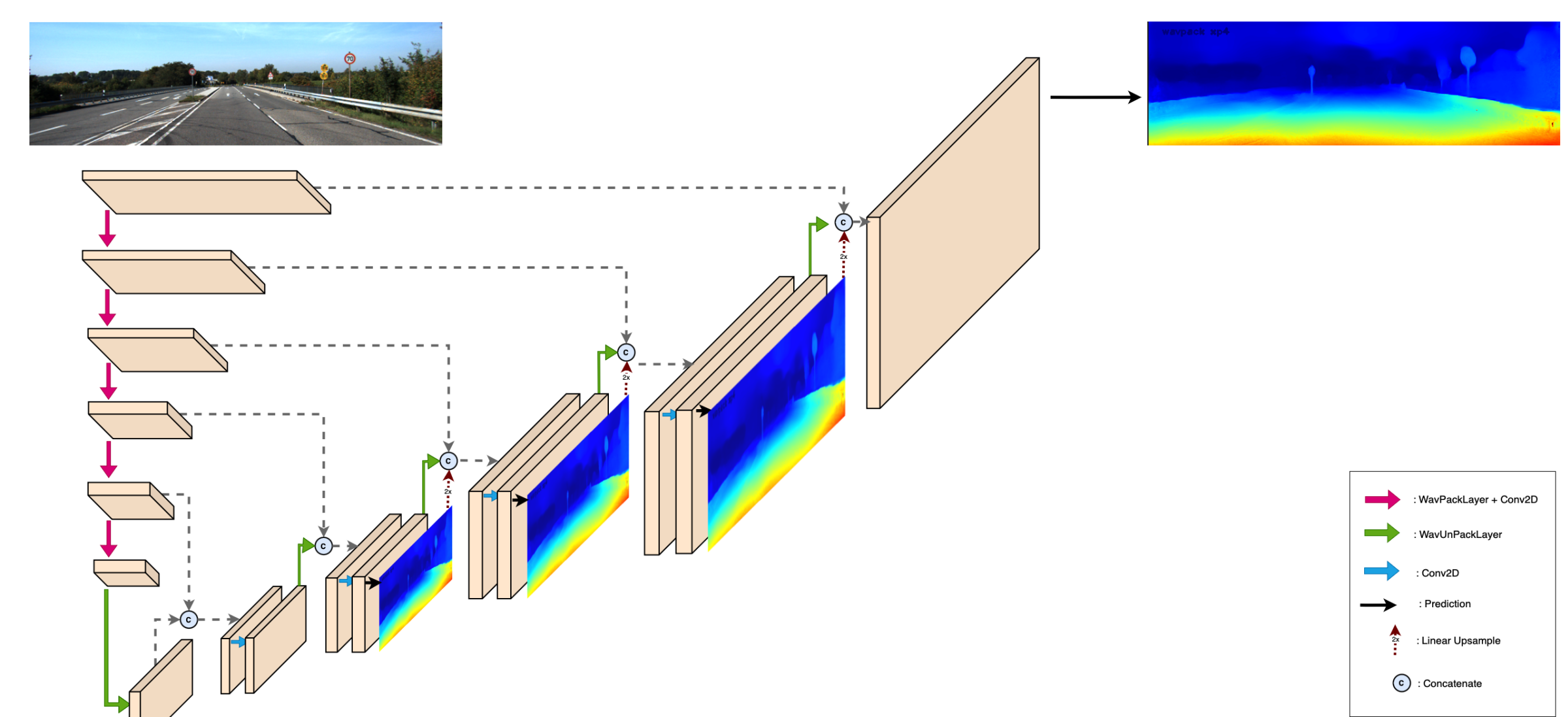
2.1 WavPacking Block



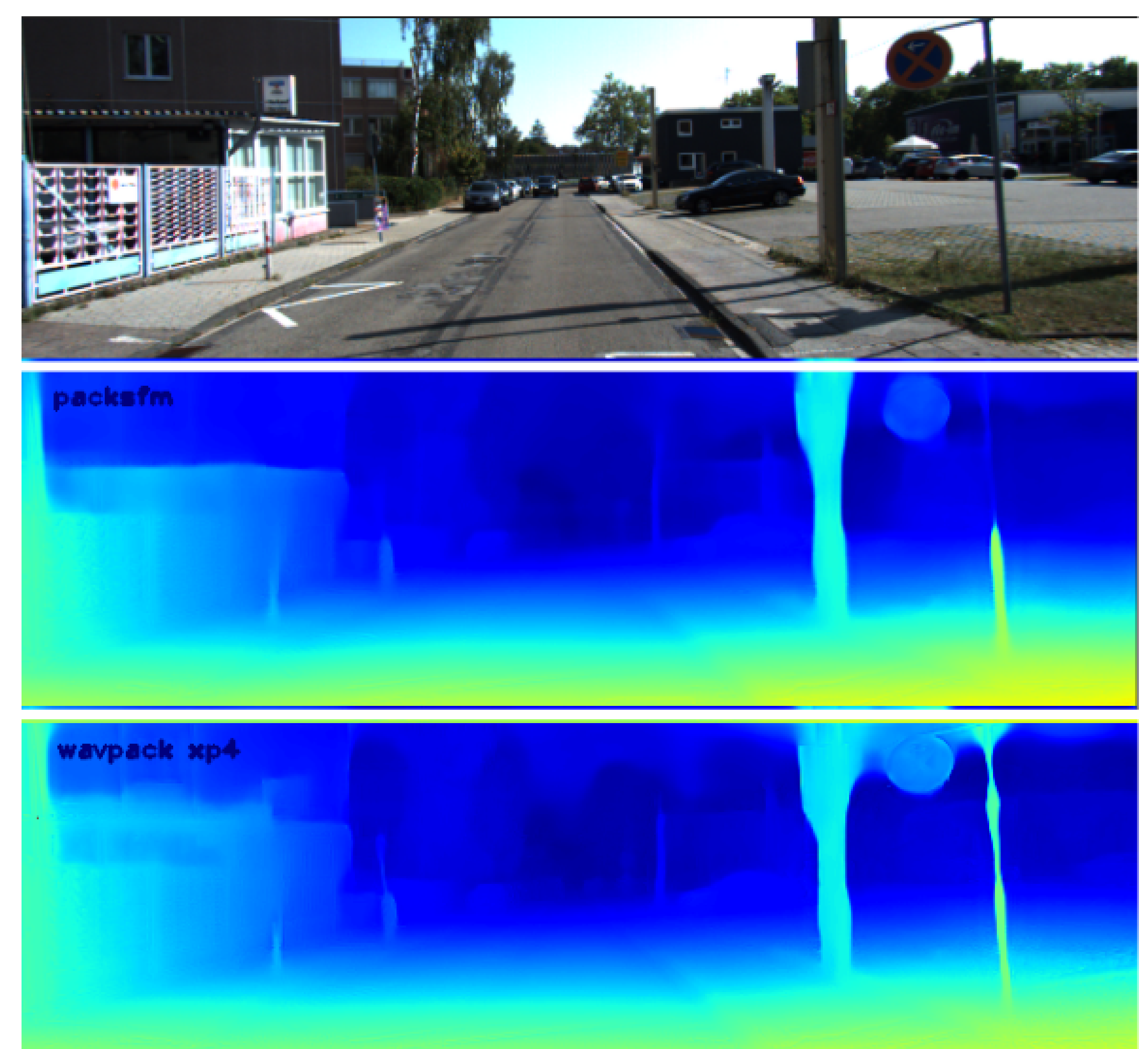
2.2 WavUnpacking Block



2.3 WavPacknet for Depth Estimation



Results



Methods	Parameters (millions)	GFLOPs	Training speed (1 A100 GPU)	Inference speed (1 RTX 2080 GPU)
3D PackNet	128.29	821.75	4.6 images/s	0.199 second/image
WavPackNet	68.65	308.76	7.1 images/s	0.102 second/image

Table 1: Network complexity and runtime comparison with 384x1280 input images

Conclusion

WavPackNet has approximately **half the complexity** and operates **twice as fast** as 3D PackNet, while matching or exceeding 3D PackNet in most configurations and evaluation metrics.

Resources



Code



Demo



Presentation