Supplementary Material: Enhanced Correlation Matching based Video Frame Interpolation

1. Network Architectures

Figure 1 and 2 illustrate the detailed network architectures of proposed enhanced correlation matching based video frame interpolation (ECMNet) which are introduced in Sec.3 of the main paper. Each module is described with entire layer and each input and outputs. For the figure 1, the encoder network and feature downsampling network share the parameters for each input. Also, flow estimation network adopts recurrent pyramid structure with shared parameters. The upscailing ratio of each pyramid level is two except the last pyramid level which is four. The plural expression of input and output data means data for both sides. The activation function is ReLU [3] for flow estimation network and Leaky ReLU [2] for frame synthesis network and refinement network except for the last layer of each network. The importance mask and the blending mask are bounded by sigmoid function.

2. Video results

In addition to figure 5 in the main paper, we present more qualitative video results for the entire X4K1000FPS[4] video frame interpolation in the accompanied video. We compare our results with the existing methods, CAIN [1] and XVFI [4]. For the fair comparison, we fine-tuned the pre-trained model of CAIN [1] with the X4K1000FPS dataset. The results show that our method synthesizes visually plausible interpolation results with preserving context than previous methods.

References

[1] Myungsub Choi, Heewon Kim, Bohyung Han, Ning Xu, and Kyoung Mu Lee. Channel attention is all you need for video frame interpolation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 10663–10671, 2020. 1

[2] Andrew L Maas, Awni Y Hannun, Andrew Y Ng, et al. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, page 3. Citeseer, 2013. 1

[3] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010. 1

[4] Hyeonjun Sim, Jihyong Oh, and Munchurl Kim. Xvfi: extreme video frame interpolation. *arXiv preprint arXiv:2103.16206*, 2021. 1

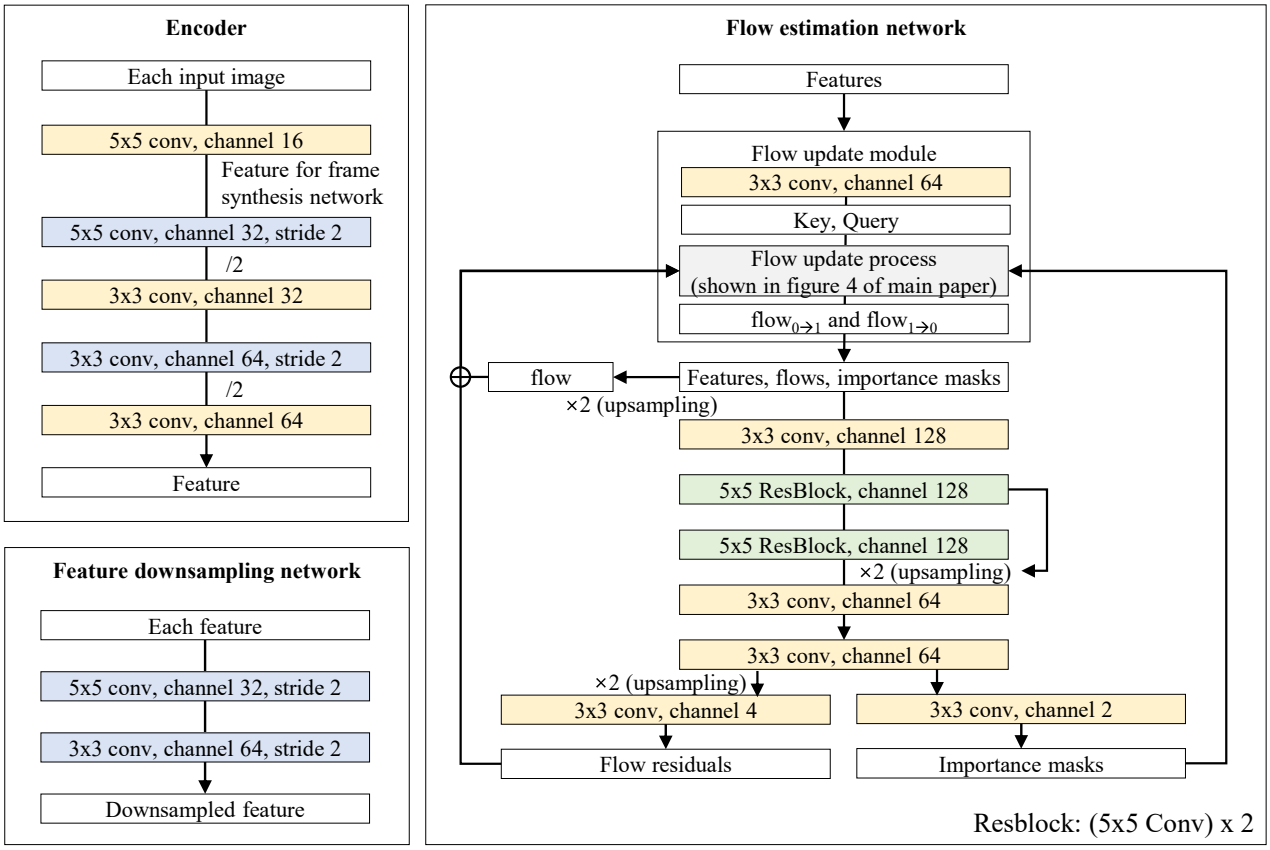

Figure 1: The detailed architecture of the encoder, feature downsampling network, and flow estimation network. The flow estimation network is repeated for each pyramid level.

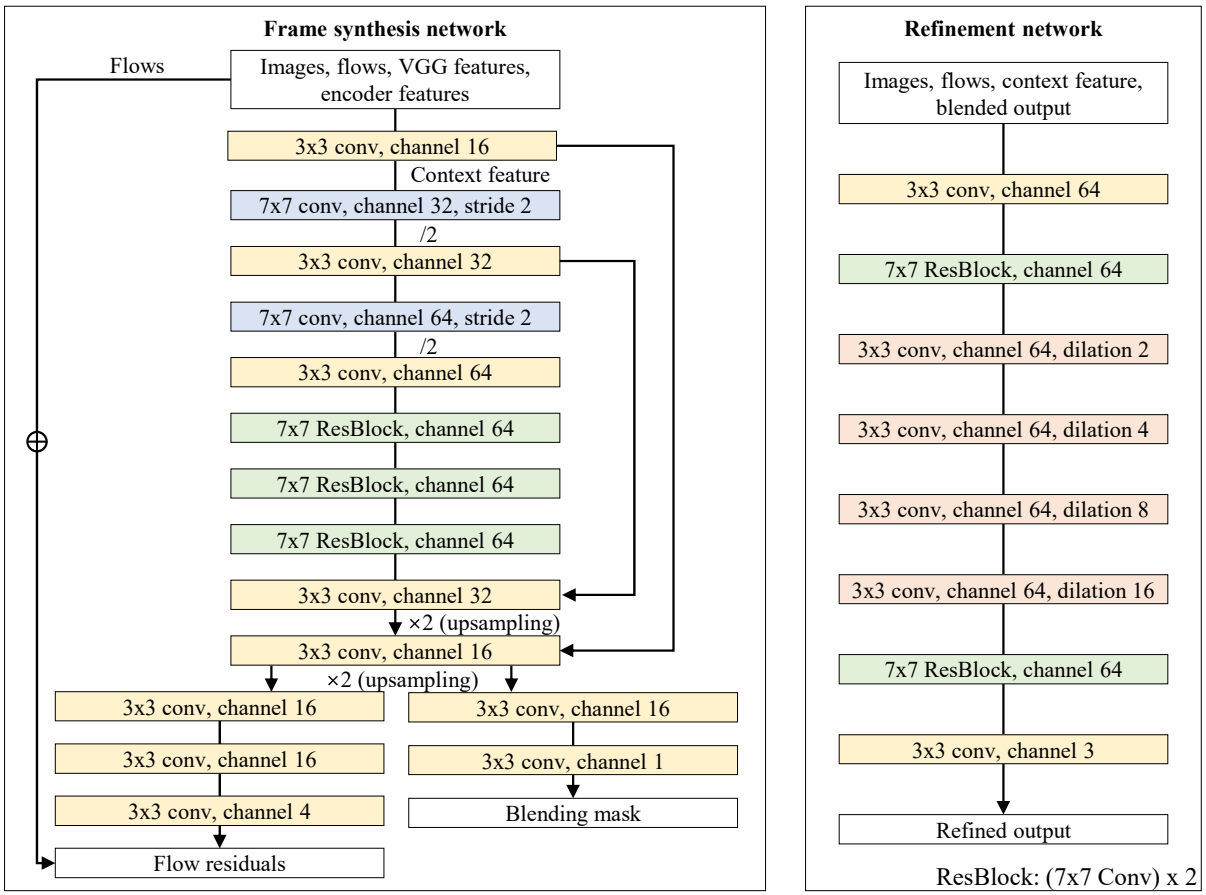

Figure 2: The detailed architecture of the frame synthesis network and refinement network.