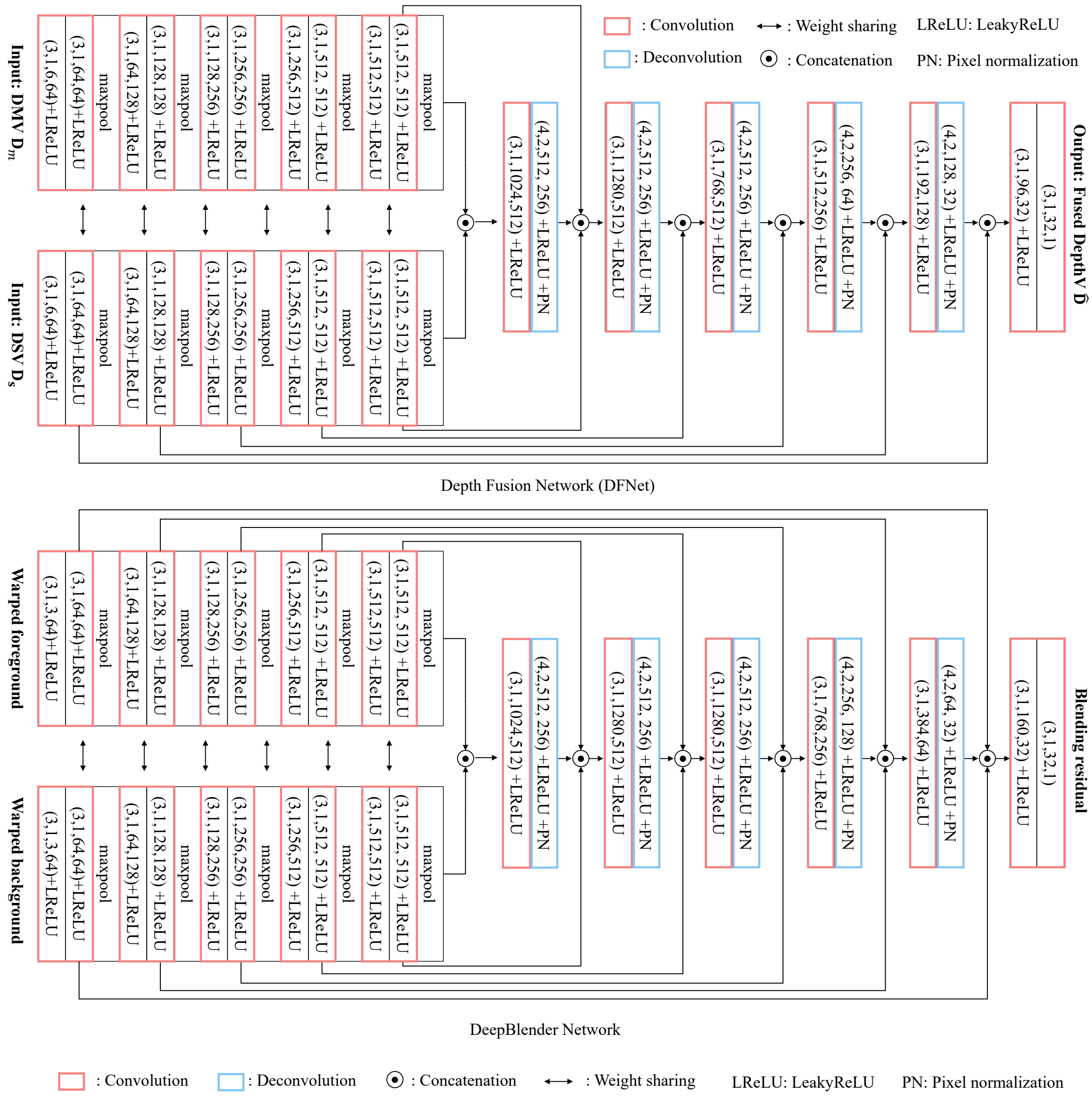
Novel View Synthesis of Dynamic Scenes with Globally Coherent Depths from a Monocular Camera

Jae Shin Yoon[†] Kihwan Kim[♯] Orazio Gallo[♯] Hyun Soo Park[†] Jan Kautz[♯]
[†]University of Minnesota [♯]NVIDIA

This supplementary document includes the implementation details of our networks, Depth Fusion Network (DFNet) and DeepBlender. Figure 1 describes the details of the network architecture. Both networks are trained with Stochastic Gradient Descent (SGD) with learning rate 1×10^{-4} for pre-training and 1×10^{-6} for self-supervision.



Depth Fusion Network (DFNet)

DeepBlender Network

Figure 1: The implementation details of our Depth Fusion Network (DFNet) and DeepBlender Network. In the convolutional and deconvolutional block, the filter property is defined as (filter size, stride size, input channel, output channel), where all intermediate inputs are zero-padded with one. We use 0.2 for LeakyReLU coefficient.