Project Management and Software Development for Medical Applications

Spatio-Temporal Depth Estimation in Real-Time

HyunJun Jung
(Supervisor : Benjamin Busam)

Technische Universität München + Johns Hopkins Whiting School of Engineering

Spatio-Temporal Depth Estimation in Real-Time

Background and Motivation

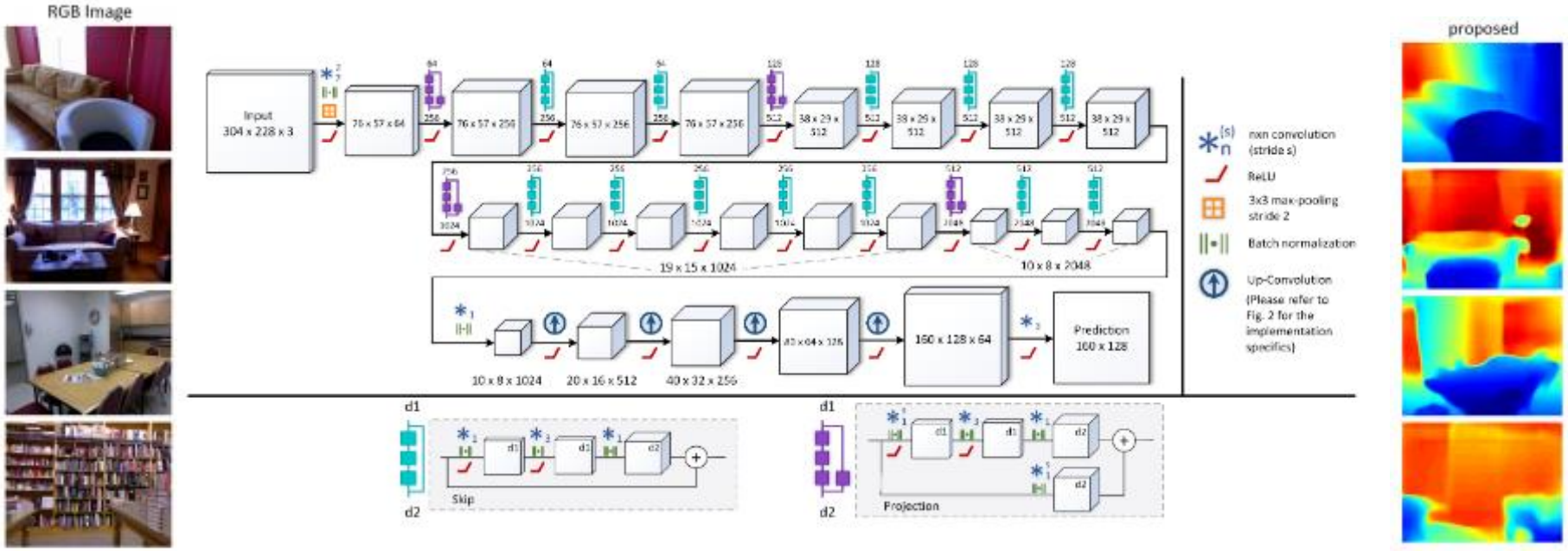
Technische Universität München

+

Johns Hopkins
Whiting School of Engineering

I want depth

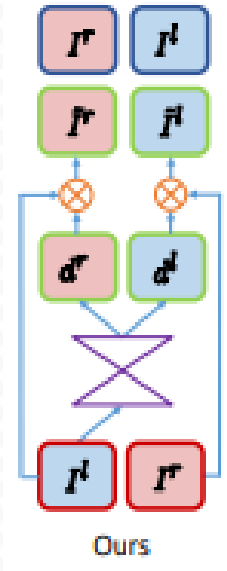
Well… Network can predict depth with an image !



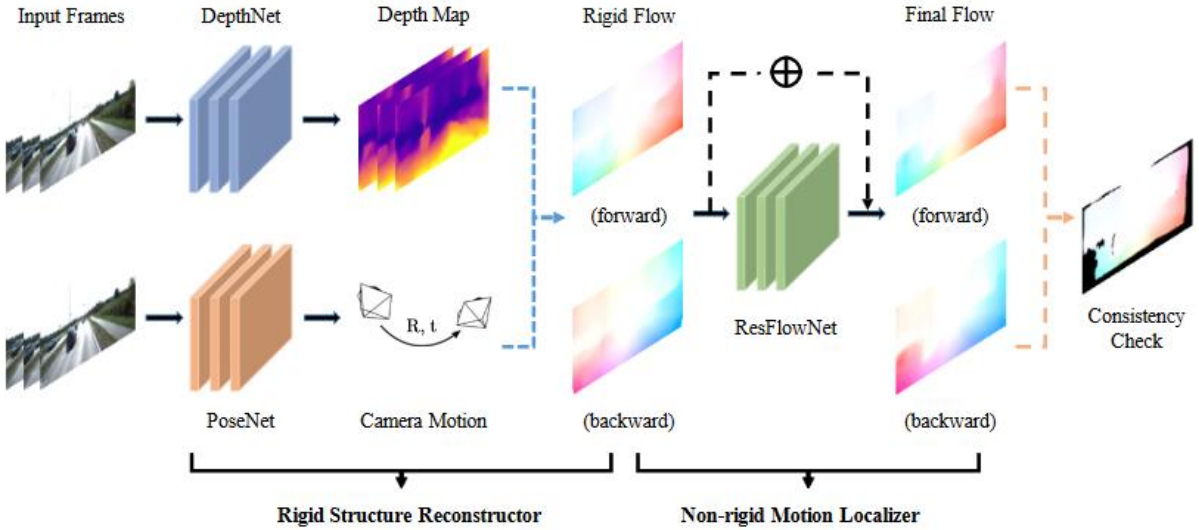
Deeper Depth Prediction with Fully Convolutional Residual Networks
Laina et al. 3DV 2016

I want better depth

Hmmm… more views gives better result !



Train with two views [1]



Train with multi views [2]

[1] Unsupervised Monocular Depth Estimation with Left-Right Consistency Godard et al. CVPR 2017
[2] GeoNet: Unsupervised Learning of Dense Depth, Optical Flow and Camera Pose, Yin, Shi. CVPR 2018

Spatio-Temporal Depth Estimation in Real-Time

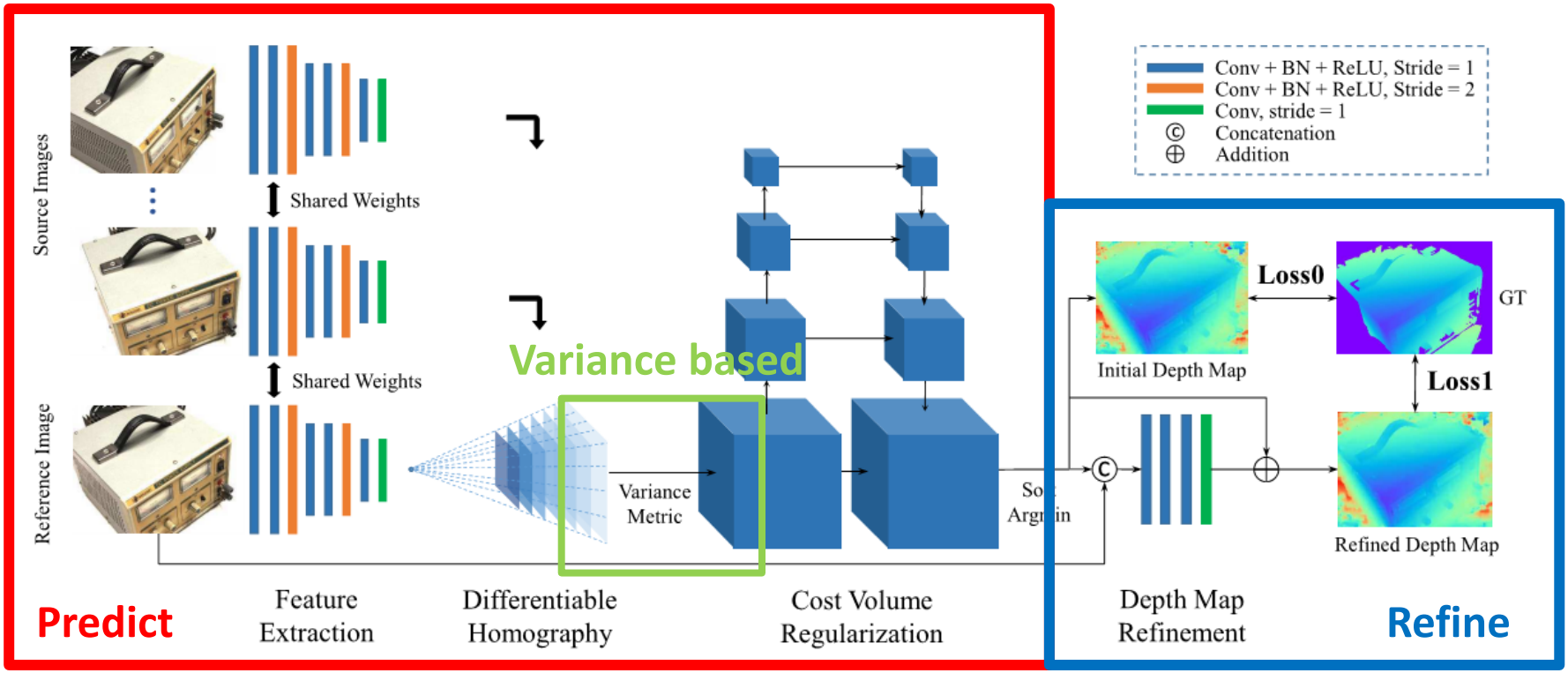
Technische Universität München
+
Johns Hopkins
Whiting School of Engineering

Spatio-Temporal Depth Estimation in Real-Time

Is.. Based on **Active Stereo Net** and **Multi View Stereo Net**

Let's talk about them first !

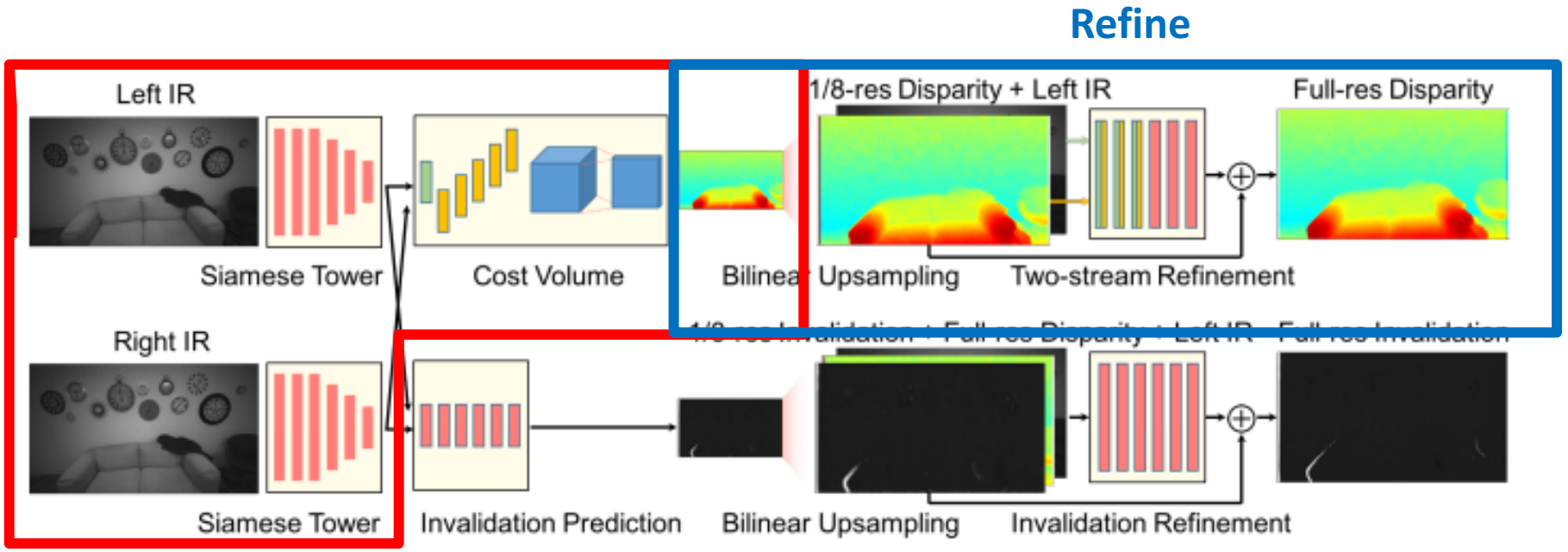
Multi View Stereo Net



- It uses Multiview information to get very accurate depth map
- Variance metric for cost volume makes it possible to use arbitrary number of views

MVSNet: Depth Inference for Unstructured Multi-view Stereo, Yao et al. ECCV 2018

Active Stereo Net



- It upsamples without upconvolution (fast)
- Network trains in unsupervised manner (no depth GT required)

ActiveStereoNet: End-to-End Self-Supervised Learning for Active Stereo Systems
Zhang et al. ECCV 2018

Spatio-Temporal Depth Estimation in Real-Time

Is.. Based on **Active Stereo Net** and **Multi View Stereo Net**

MVSNet Part :

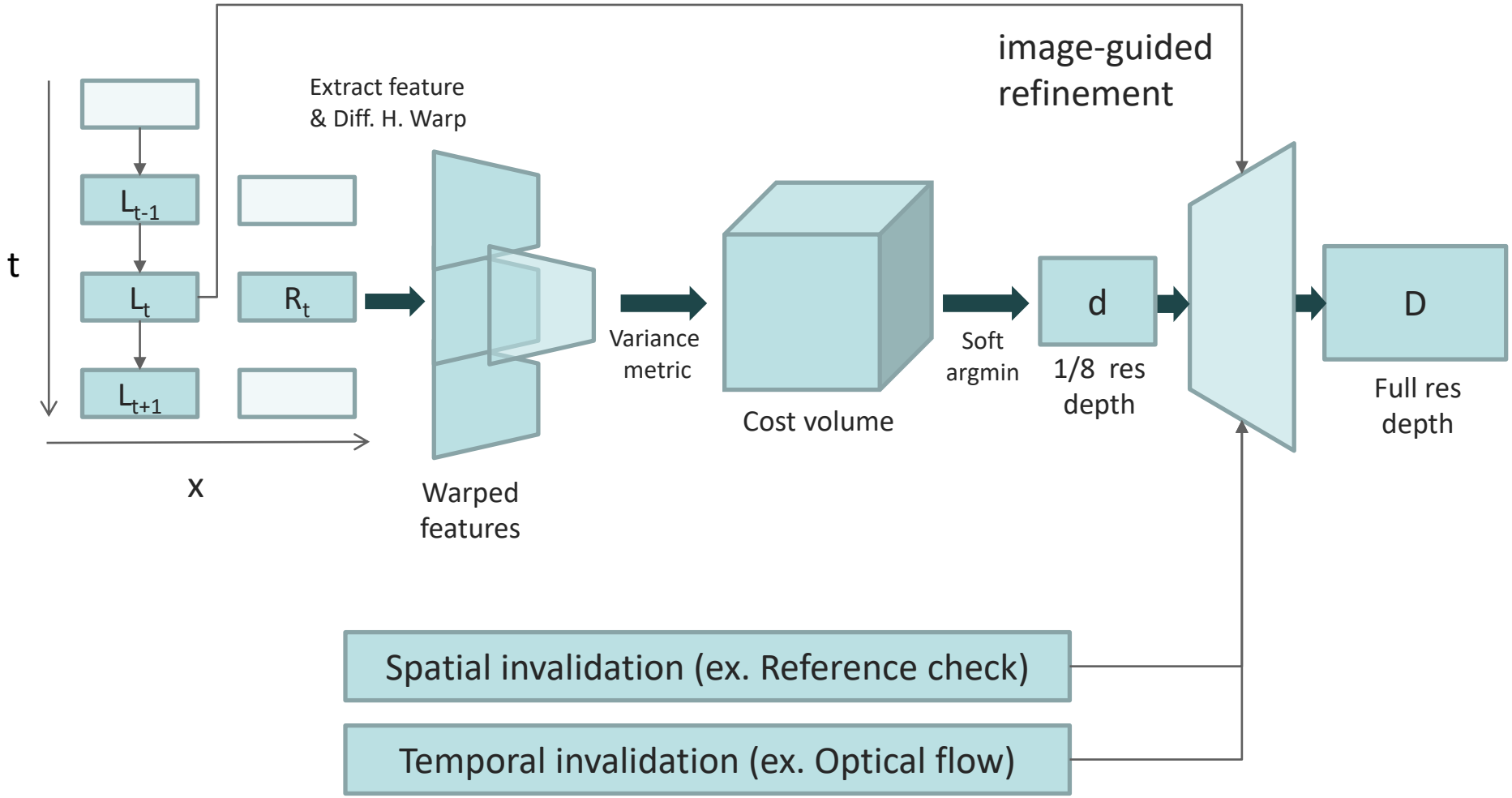
- Using Multiview input image to predict coarse level depth

ASN Part :

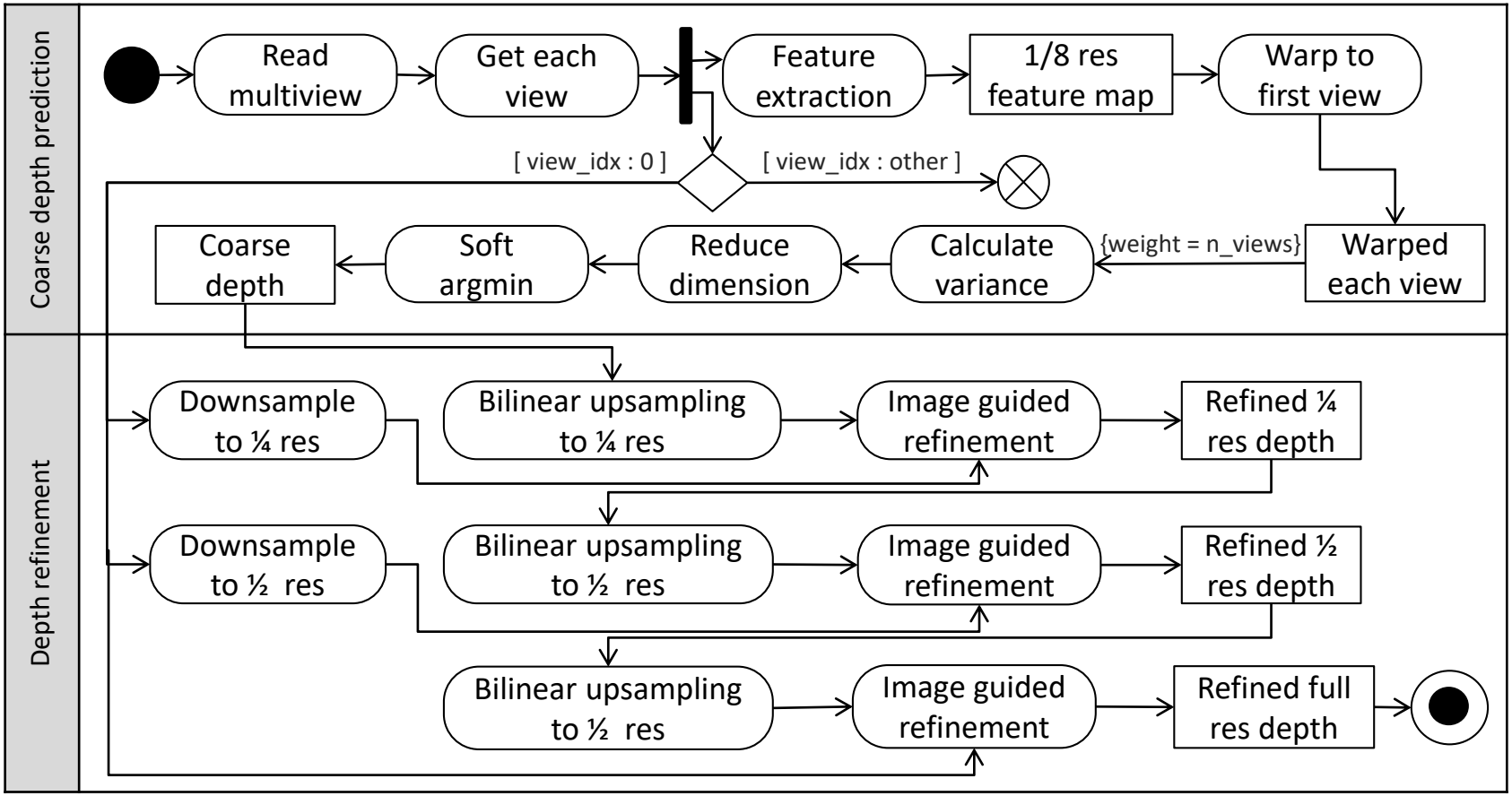
- Bilinear interpolation & guided refinement will upsample the depth (**much faster than upconvolution**)
- Self supervised training scheme will be applied at last.

Let's predict good depth with Multiview, in real time

Spatio-Temporal Depth Estimation in Real-Time



Spatio-Temporal Depth Estimation in Real-Time



Final Goal

Train multi view monocular camera to predict depth.

Input

- Realtime multiview monocular view
- R & sT from monocular SLAM (s = unknown scale)

Output

- Realtime depth map
- Scaling factor for Translation

Possible application

- Whenever we are limited to use monocular camera but need tracking (ex, endoscopy camera)

Spatio-Temporal Depth Estimation in Real-Time

Plans, Timeline

Technische Universität München

+

Johns Hopkins
Whiting School of Engineering

Plan for Implementation

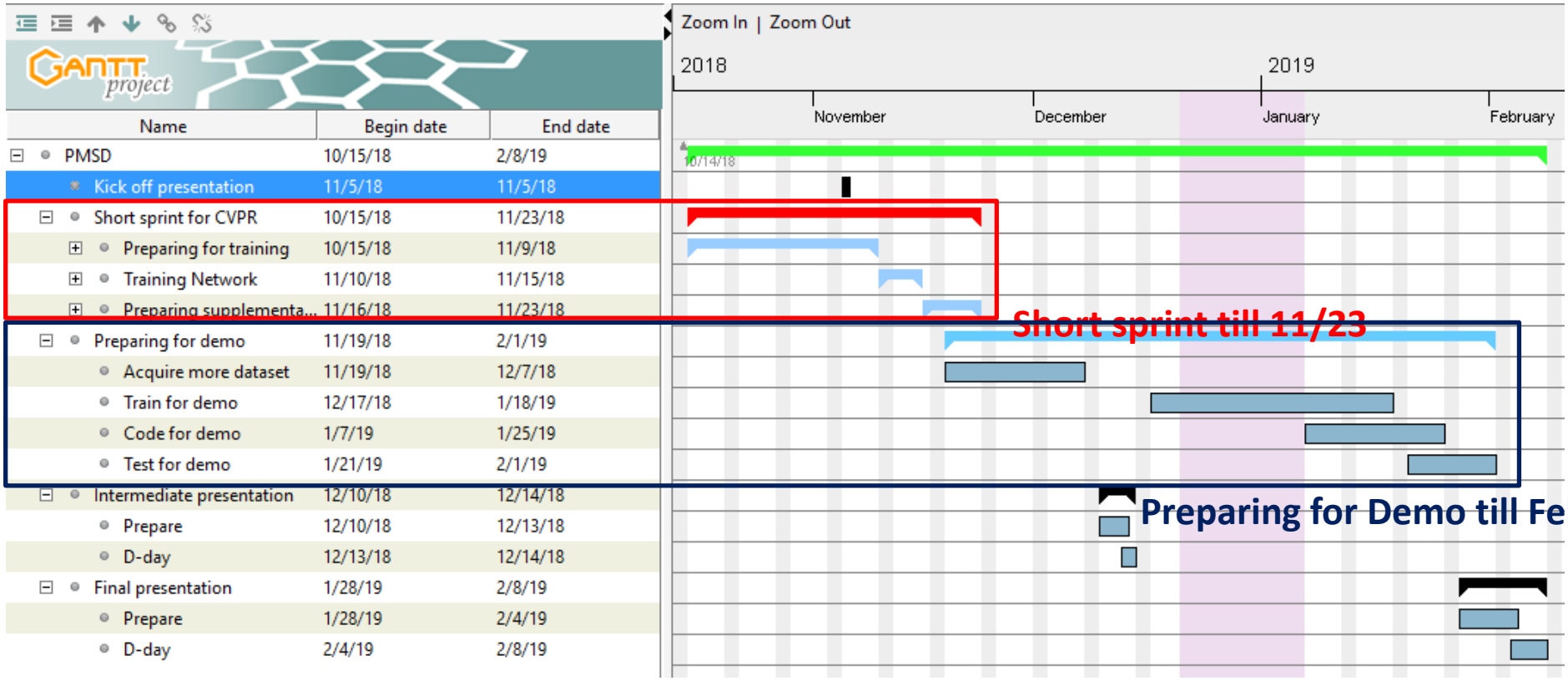
Stereo Camera Phase

- Supervised training with exact extrinsic (R, t) from simulation
- Supervised training with estimated extrinsic from SLAM
- Self-supervised training with estimated extrinsic

Mono Camera Phase

- Supervised training with extrinsic (R, scaled t).
  (scaling factor will be also predicted)
- Self-supervised version with same condition.

Timeline for Project



Spatio-Temporal Depth Estimation in Real-Time

Q & A

Technische Universität München

+

Johns Hopkins
Whiting School of Engineering