Curvature-guided Dynamic Scale Networks for Multi-view Stereo

Khang Truong Giang **Soohwan Song** **Sungho Jo**

School of Computing

KAIST, Korea

Background & Motivation

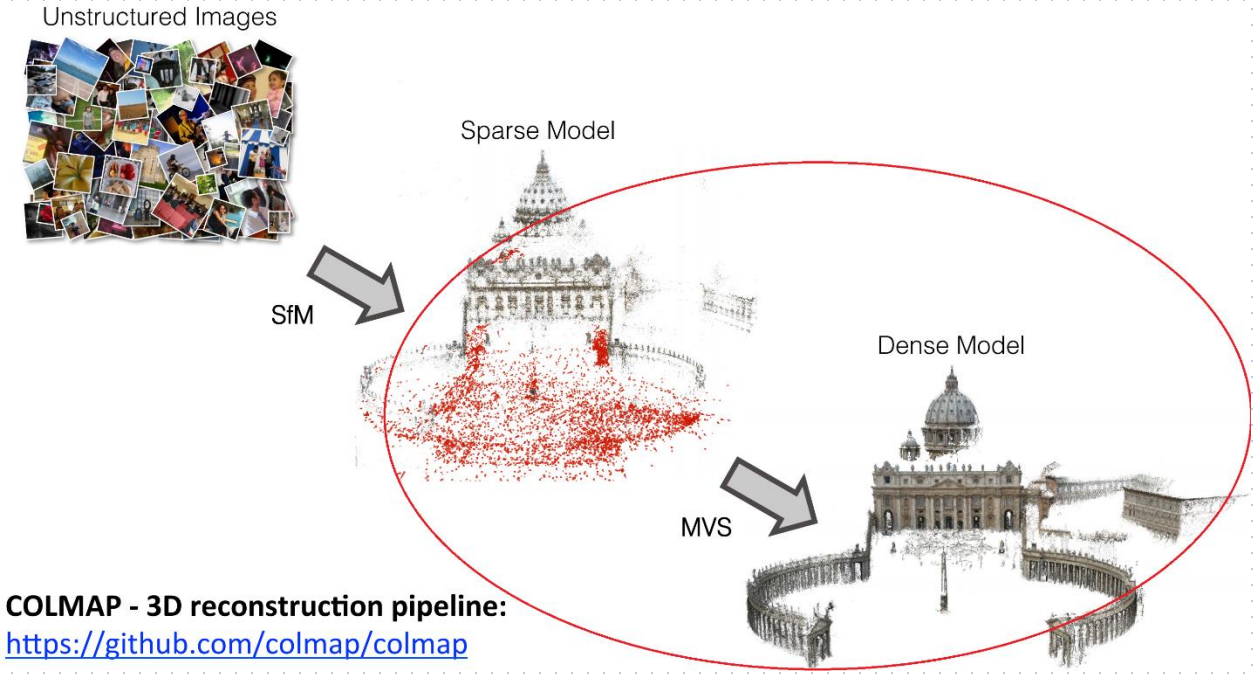
3D reconstruction:

reconstruct a 3D model from a collection of images

two main tasks: structure from motion (SfM) and multi-view stereo (MVS)

Multi-view Stereo

estimate dense correspondences (depth maps) for all images

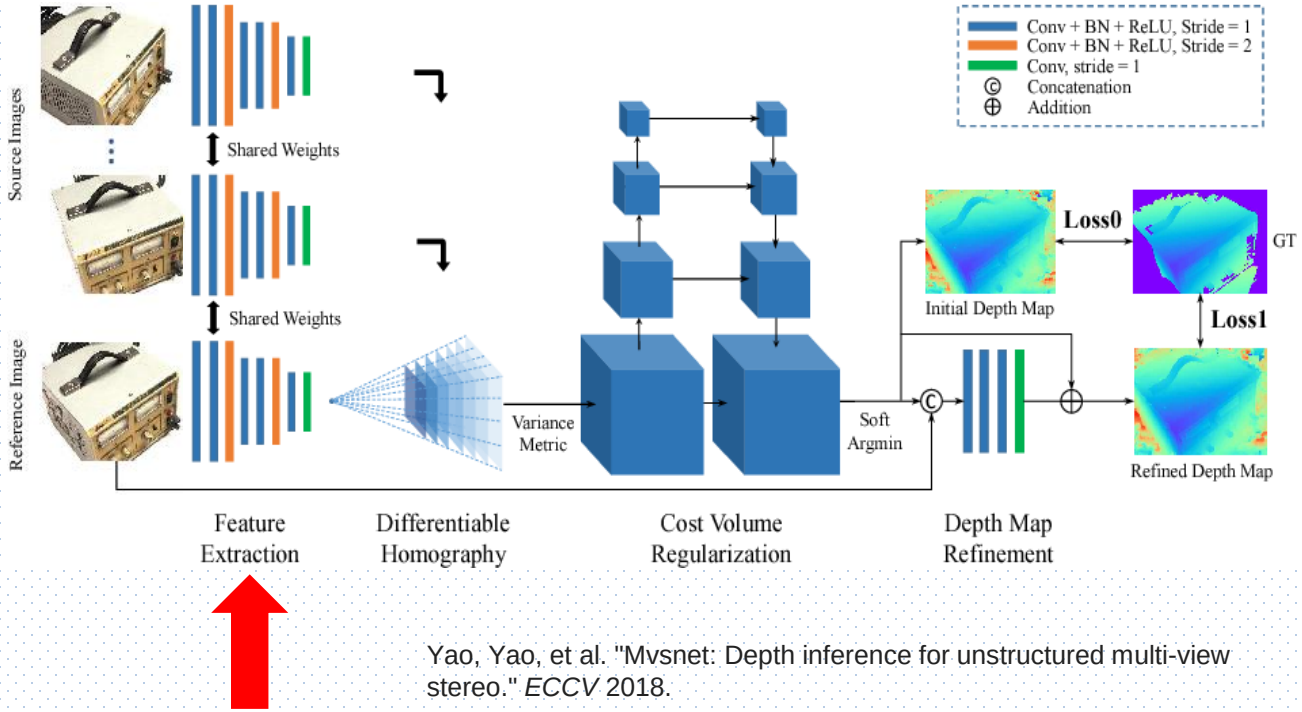


COLMAP - 3D reconstruction pipeline:
https://github.com/colmap/colmap

Related Work

Multi-view Stereo

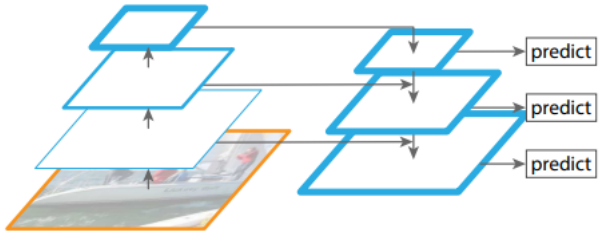
Traditional approach: COLMAP, openMVS…
Learning-based approach: MVSNet, CasMVSNet…



Yao, Yao, et al. "Mvsnet: Depth inference for unstructured multi-view stereo." *ECCV* 2018.

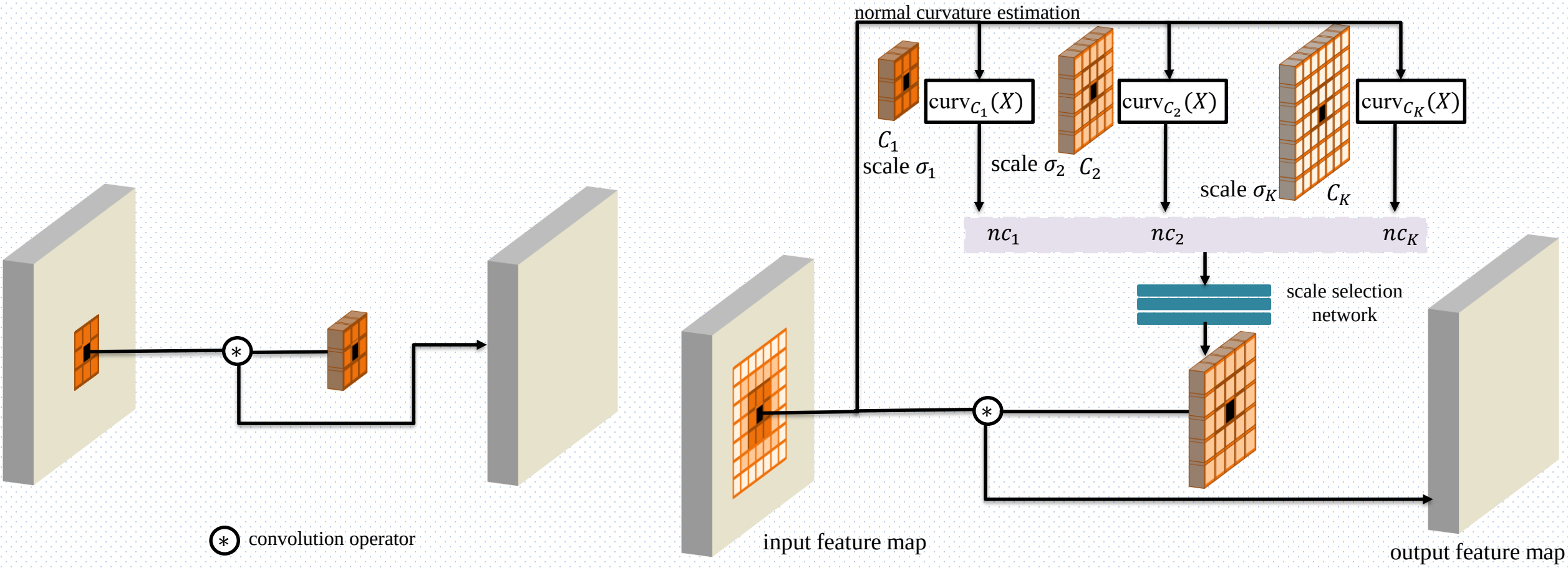
Multi-scale Feature Extraction

Standard CNN encoder: PMVSNet, AttMVSNet…
Feature pyramid network: VisMVSNet, CVPMVSNet…



Lin, Tsung-Yi, et al. "Feature pyramid networks for object detection." *CVPR* 2017.

Proposed Method

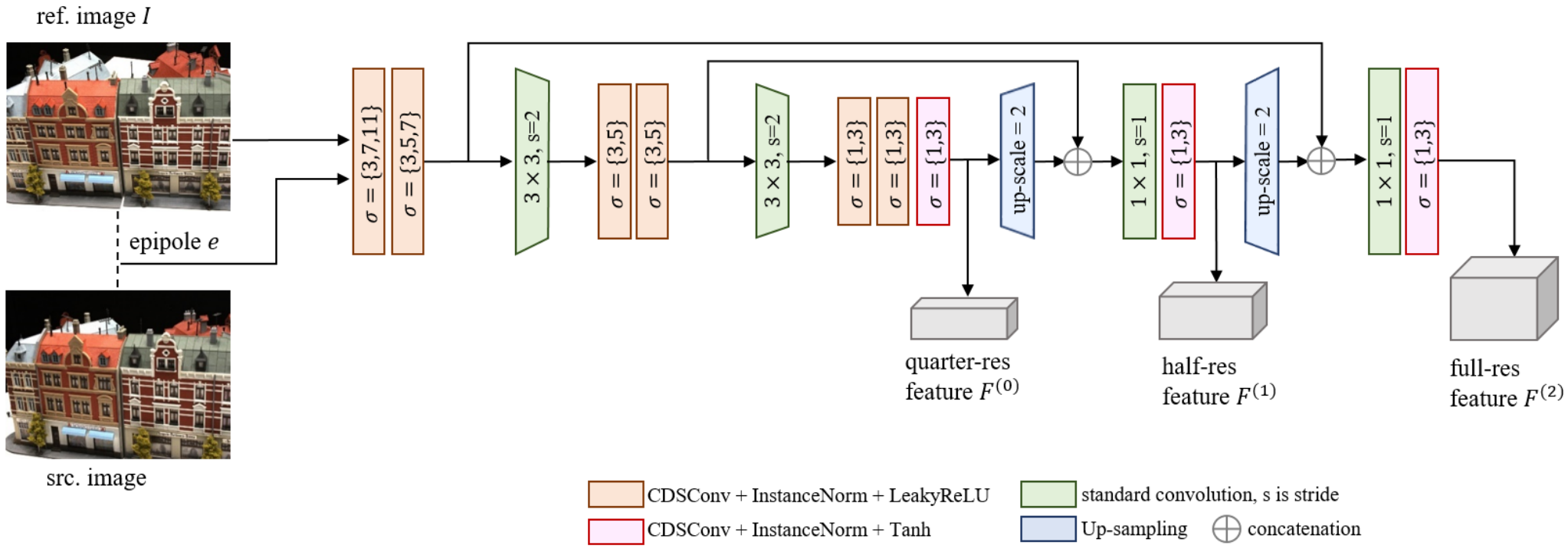


standard convolution

Curvature-guided dynamic scale convolution (CDSConv)

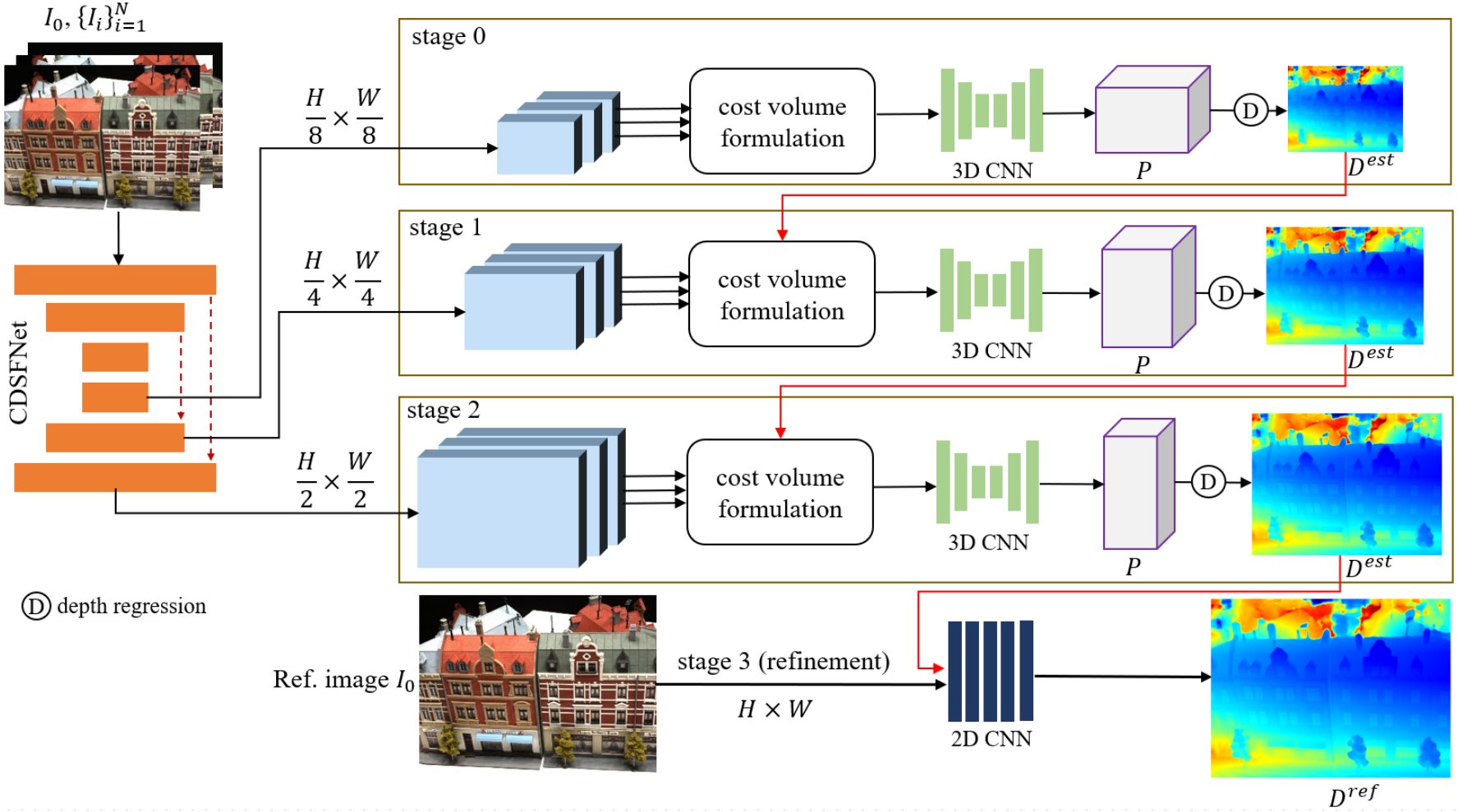
Proposed Method

An architecture for feature extraction: CDSFNet



Proposed Method

CDS-MVSNet architecture



Experimental Results

Evaluation of 3D reconstructed models on the DTU and Tanks & Temples datasets

	Tanks and Temples (F-score)									DTU (mm)		
	Mean	Family	Francis	Horse	Lighthouse	M60	Panther	Playground	Train	Acc.	Comp.	Overall
COLMAP	42.14	50.41	22.25	25.63	56.43	44.83	46.97	48.53	42.04	0.400	0.664	0.532
Vis-MVSNet	60.03	77.40	60.23	47.07	63.44	62.21	57.28	60.54	52.07	0.369	0.361	0.365
AttMVS	60.05	73.90	62.58	44.08	**64.88**	56.08	**59.39**	**63.42**	**56.06**	0.383	0.329	0.356
CasMVSNet	56.84	76.37	58.45	46.26	55.81	56.11	54.06	58.18	49.51	0.325	0.385	0.355
UCSNet	54.83	76.09	53.16	43.03	54.00	55.60	51.49	57.38	47.89	0.338	0.349	0.344
PatchMatchNet	53.15	66.99	52.64	43.24	54.87	52.87	49.54	54.21	50.81	0.427	**0.277**	0.352
PVA-MVSNet	54.46	69.36	46.80	46.01	55.74	57.23	54.75	56.70	49.06	0.379	0.336	0.357
CVP-MVSNet	54.03	76.50	47.74	36.34	55.12	57.28	54.28	57.43	47.54	**0.296**	0.406	0.351
BP-MVSNet	57.60	77.31	60.90	47.89	58.26	56.00	51.54	58.47	50.41	0.333	0.320	0.327
Ours	**61.58**	**78.85**	**63.17**	**53.04**	61.34	**62.63**	59.06	62.28	52.30	0.351	**0.278**	**0.315**

Table 1: The metrics used for DTU are accuracy, completeness and overall. The metric used for Tanks & Temples is F-score

Experimental Results

Qualitative results on DTU

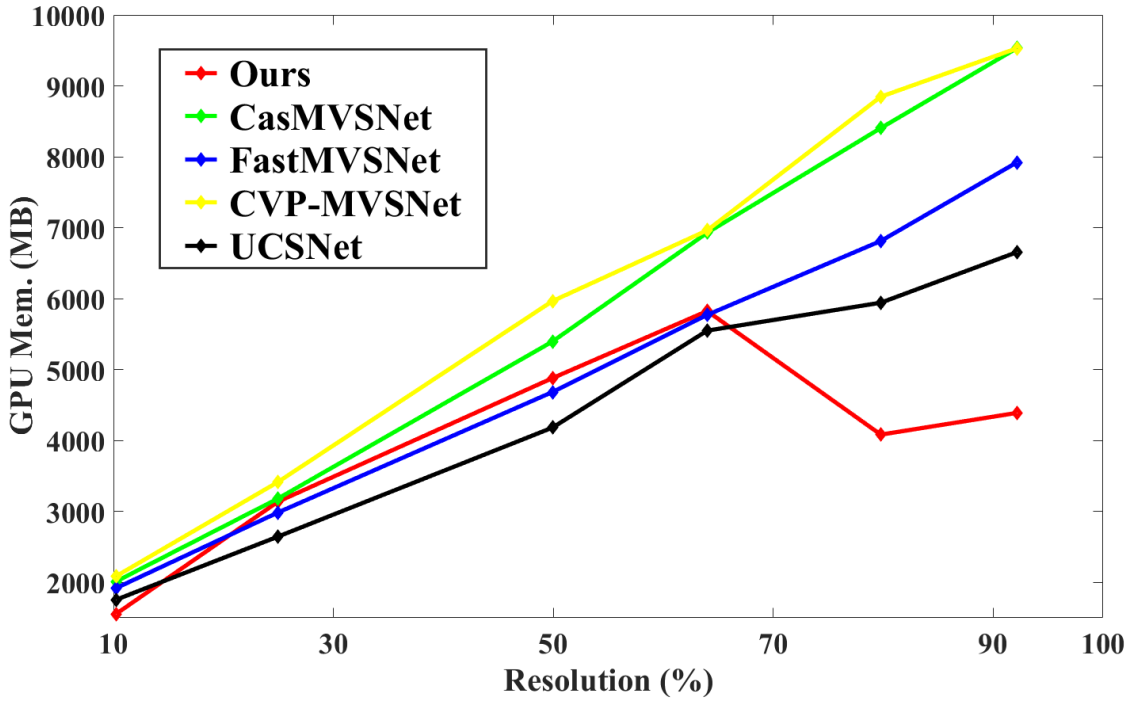
Experimental Results

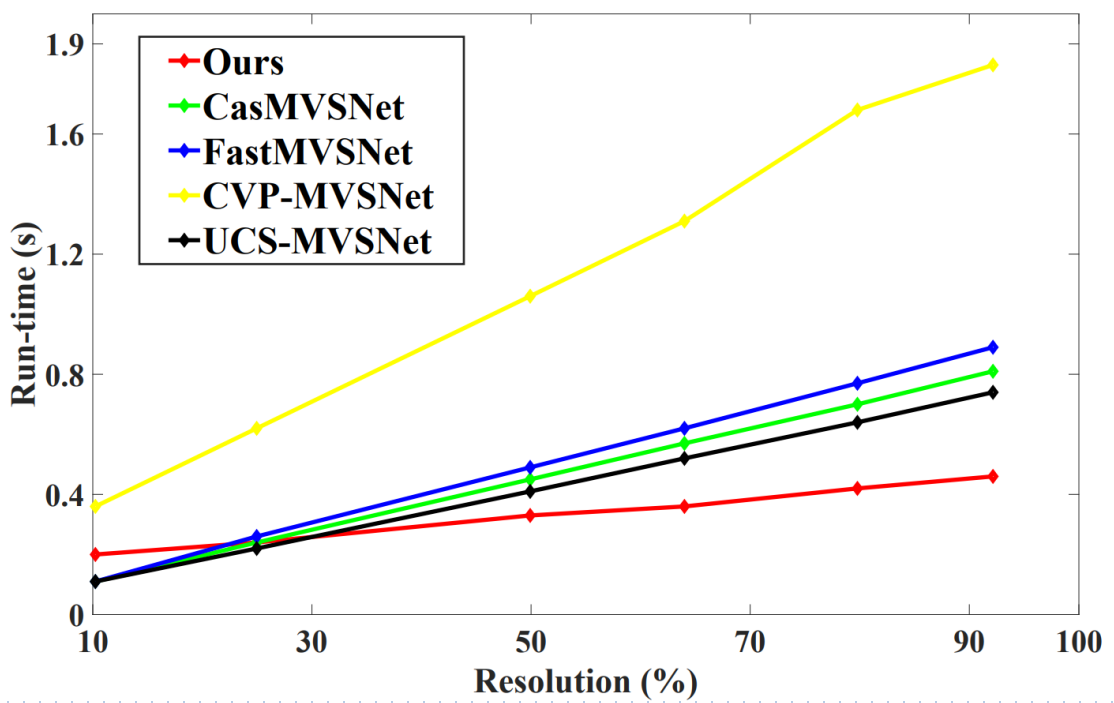
Qualitative Results on Tanks & Temples

Recall score

Experimental Results

Time and Memory





Ablation Study

#scales/layers	Prec. 2mm	Prec. 4mm	Prec. 8mm	MAE (mm)	Time (s)
1 (CDSFNet → FPN)	40.87	55.95	66.72	21.41	**0.114**
2	44.46	58.10	67.89	18.89	0.134
3	45.27	59.14	69.17	17.46	0.155
4	**46.77**	**60.63**	**70.61**	**16.10**	0.195

Table 2: Study on the impact of number of candidate scales in each CDSConv layer

Thanks for listening!