Polarimetric Depth Estimation

Group 1 - Final Presentation

Kağan Küçükaytekin

Witold Pacholarz

Tobias Preintner

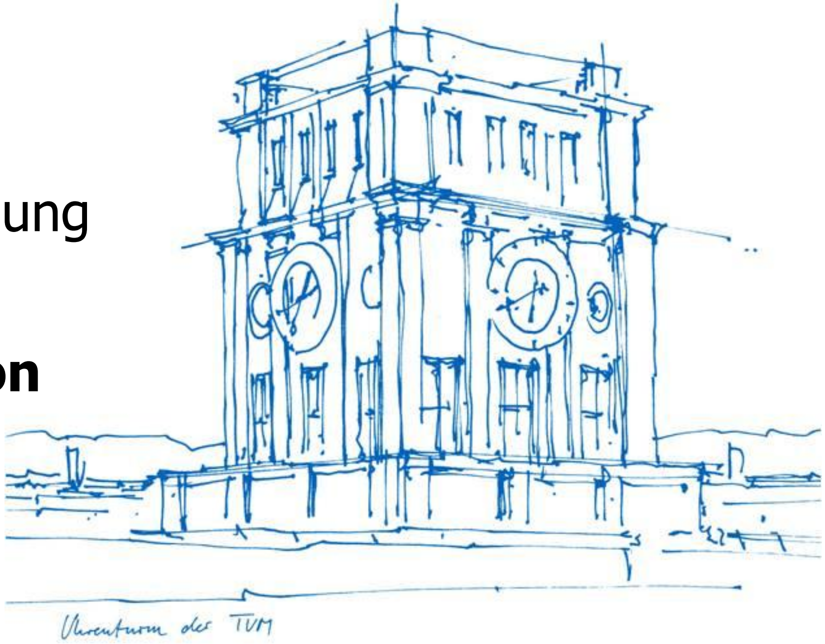
Ragıp Volkan Tatlıkazan

Supervisors: Patrick Ruhkamp, HyunJun Jung

Advanced Topics in 3D Computer Vision

Technical University of Munich

Munich, 28.07.2022

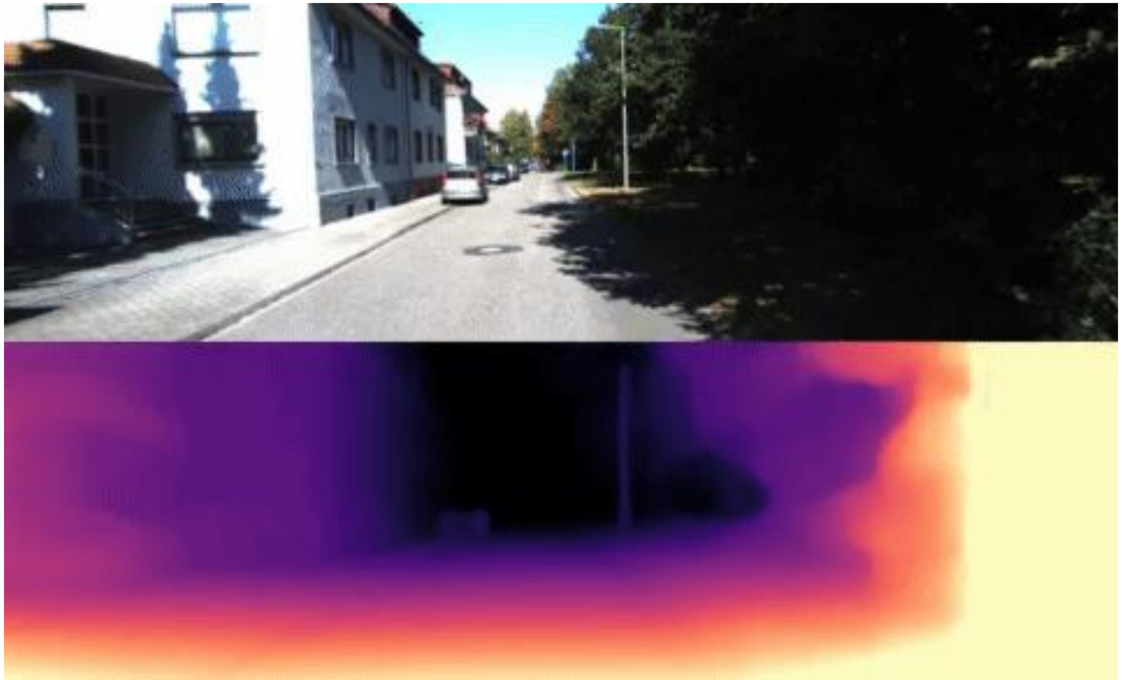


Agenda

- Motivation
- Related Work
- Polarimetric characteristics
- Architecture and losses
- Results and analysis
- Demonstration
- Limitations
- Conclusions
- Future development

Motivation

- What is monocular depth estimation?



"Digging Into Self-Supervised Monocular Depth Estimation"
Clément Godard, Oisin Mac Aodha, Michael Firman, Gabriel Brostow; ICCV 2019

Motivation

- What is monocular depth estimation?
- Where is it applied?

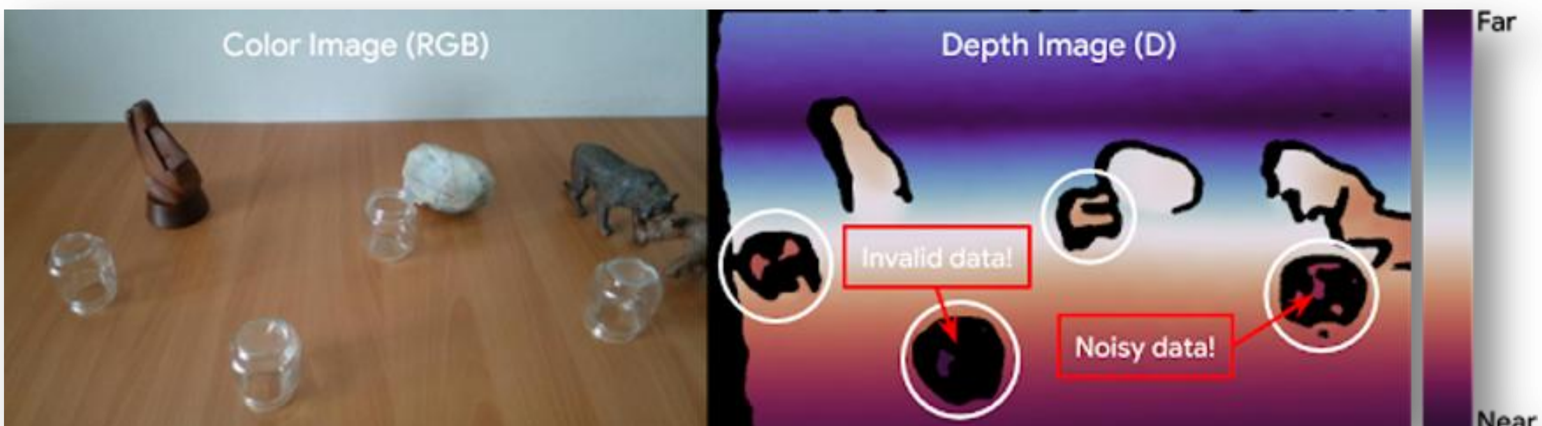
www.robots.ieee.org

www.theverge.com

www.uk.pcmag.com

Motivation

- What is monocular depth estimation?
- Where is it applied?
- Why polarised images?



www.ai.googleblog.com

Goal

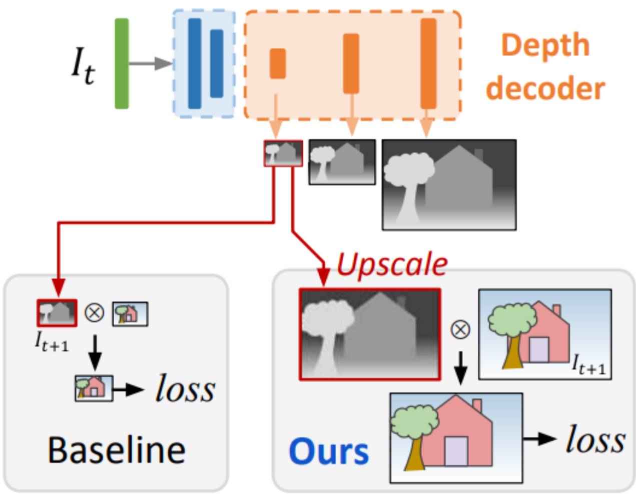
Quantitative and qualitative improvement
of the supervised monocular depth estimation
for photometrically challenging objects
by leveraging polarimetric characteristics of light

www.robots.ieee.org

Related work

Monodepth2

- Popular baseline for depth estimation
- Sequential frames as train data
- Depth loss using downscaled predictions

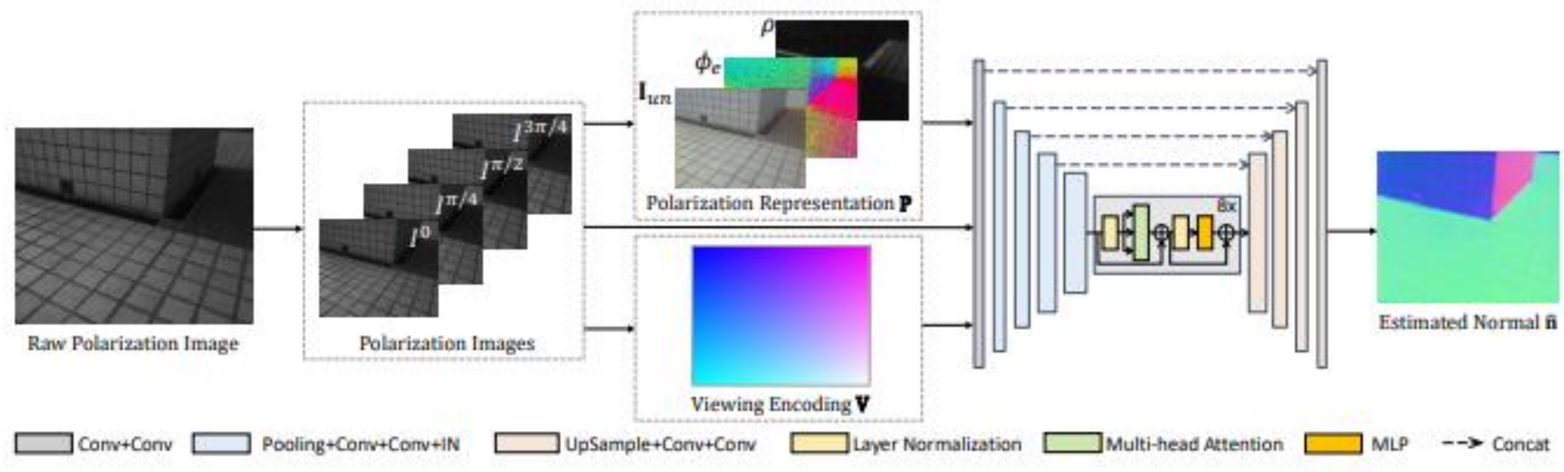


"Digging Into Self-Supervised Monocular Depth Estimation"
Clément Godard, Oisin Mac Aodha, Michael Firman, Gabriel Brostow; ICCV 2019

Related work

Shape from Polarization for Complex Scenes in the Wild

- Normals estimation for full scenes
- Encoder-decoder structure
- Attention in bottleneck

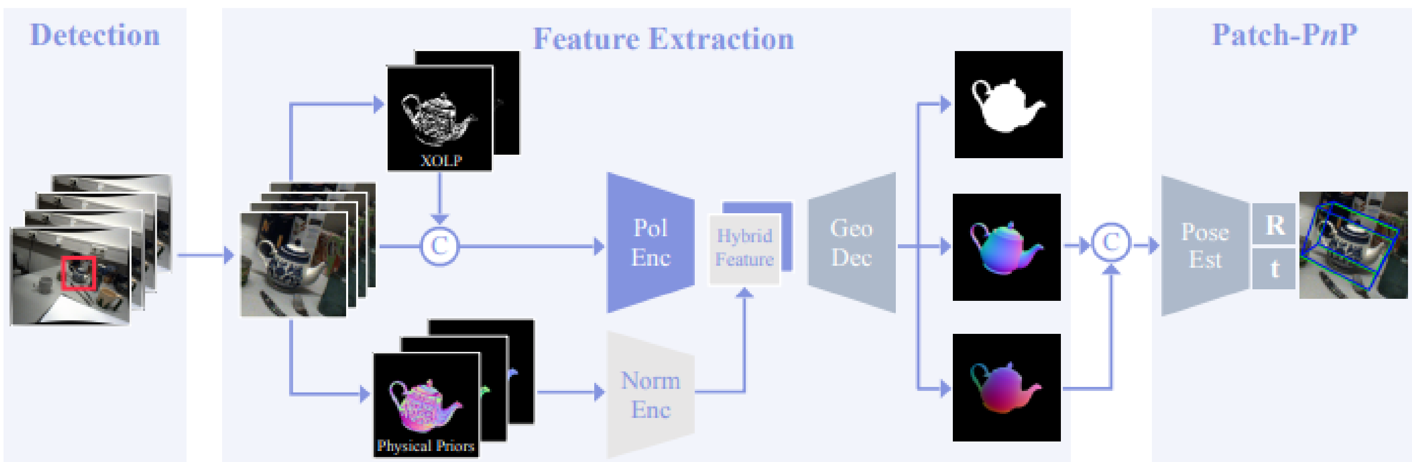


"Shape from Polarization for Complex Scenes in the Wild"
Chenyang Lei, Chenyang Qi, Jiaxin Xie, Na Fan, Vladlen Koltun and Qifeng Chen; CVPR 2022

Related work

Polarimetric Pose Prediction

- Prediction of 6D pose
- Utitizes a normal encoder

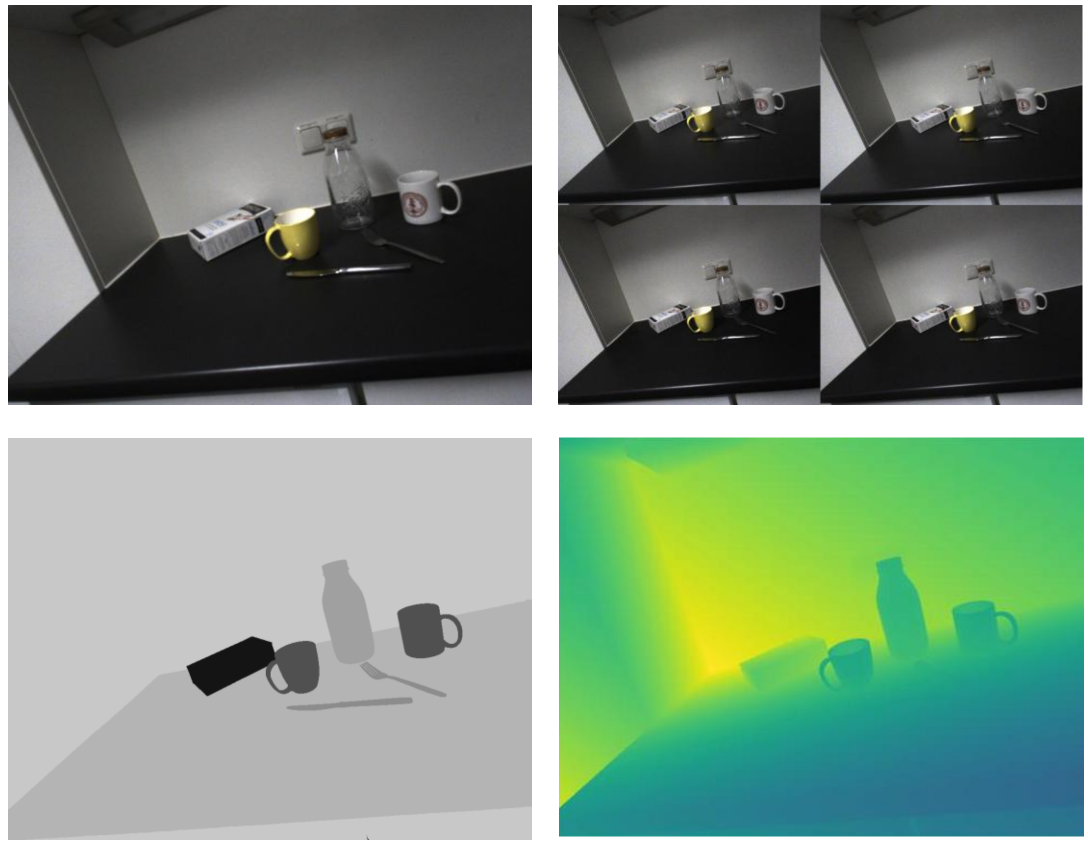


"Polarimetric Pose Prediction"
Daoyi Gao, Yitong Li, Patrick Ruhkamp, Iuliia Skobleva, Magdalena Wysocki, HyunJun Jung, Pengyuan Wang, Arturo Guridi, Nassir Navab, Benjamin Busam (ECCV 2022)

Dataset

HAMMER

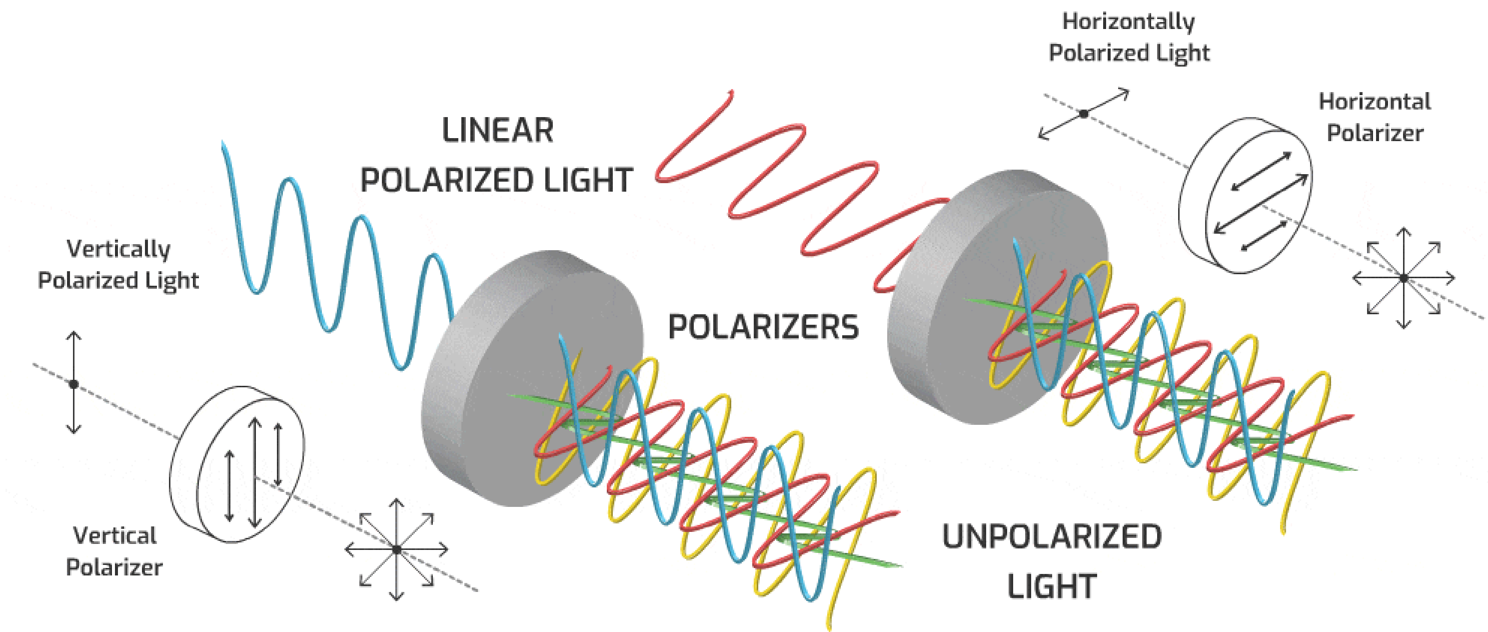
- RGB images
- 4 polarised images
  (0, 45, 90 and 135 deg)
- Instance masks
- Depth ground truth



"Is my Depth Ground-Truth Good Enough? HAMMER - Highly Accurate Multi-Modal Dataset for DEnse 3D Scene Regression"
HyunJun Jung, Patrick Ruhkamp, Guangyao Zhai, Nikolas Brasch, Yitong Li, Yannick Verdie, Jifei Song, Yiren Zhou, Anil Armagan, Slobodan Ilic, Aleš Leonardis, Benjamin Busam; 2022

Light polarisation

- Initially: a beam of light with multiple directional waves (unpolarised)
- After going through a polariser: some particular rays remain (polarised)



www.noetic.org

Polarimetric characteristics

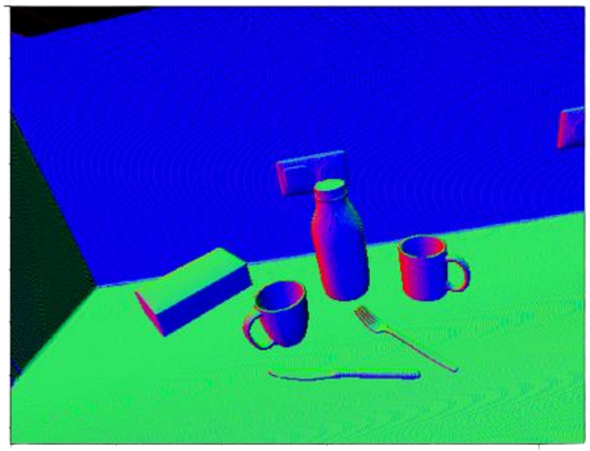
polarised images

DOLP

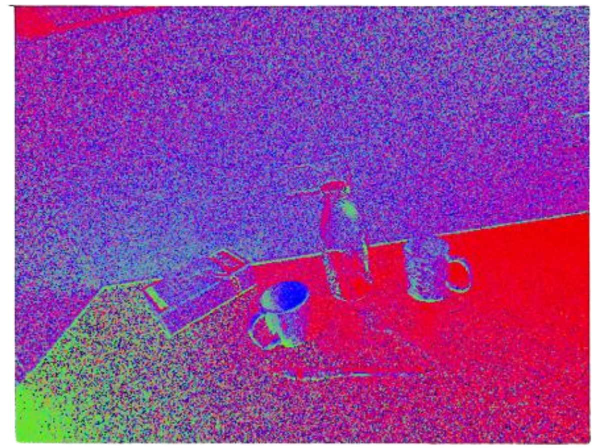
AOLP

Polarimetric characteristics

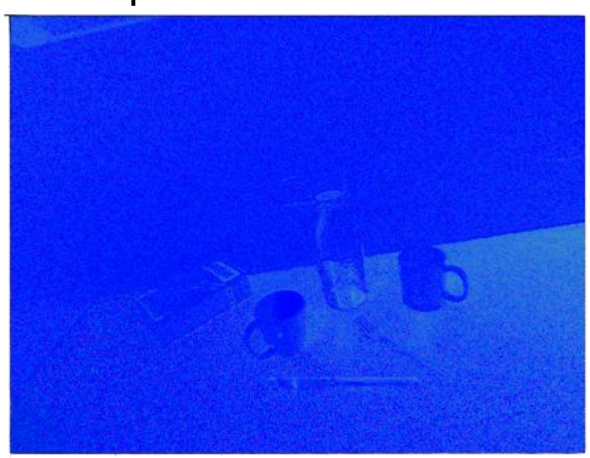
GT normals



diffuse normals



specular normals 1



specular normals 2

Data preprocessing

- Augmentation:
  - horizontal flip
  - random brightness, contrast, saturation, hue jitter

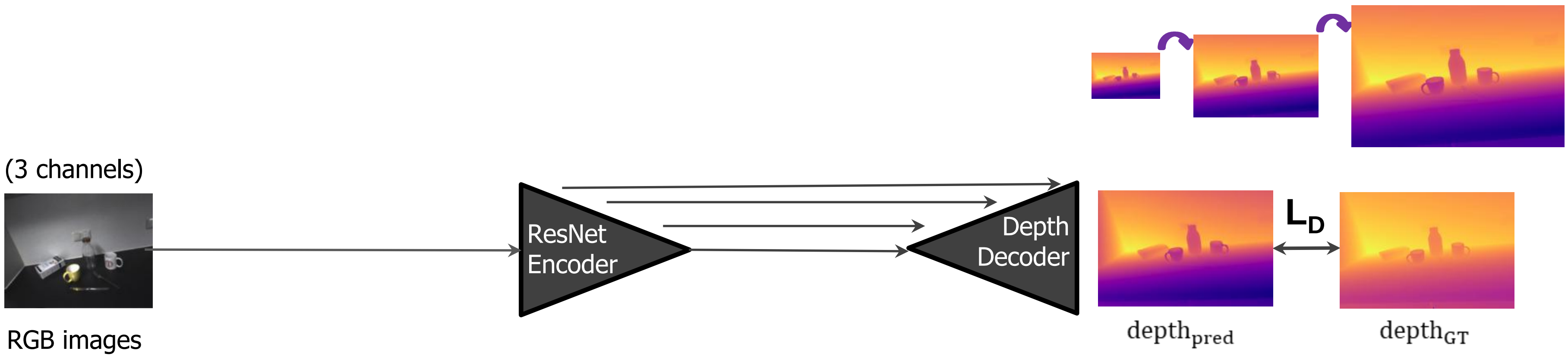
- Downscaling images for loss calculations

- Speeding up training with pre-splitting polarised images

- Standardisation of RGB and XOLP encoder inputs

Architecture development

Baseline architecture



Loss Functions

- Depth Regression

$$\mathcal{L}_1(y, \hat{y}) = |y - \hat{y}|$$

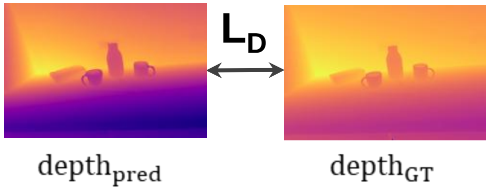
- Smoothing Loss (d stands for disparity):

$$\mathcal{L}_s(d, \hat{I}) = \frac{\partial d}{\partial x}\, e^{-\partial \hat{I}/\partial x} + \frac{\partial d}{\partial y}\, e^{-\partial \hat{I}/\partial y}$$

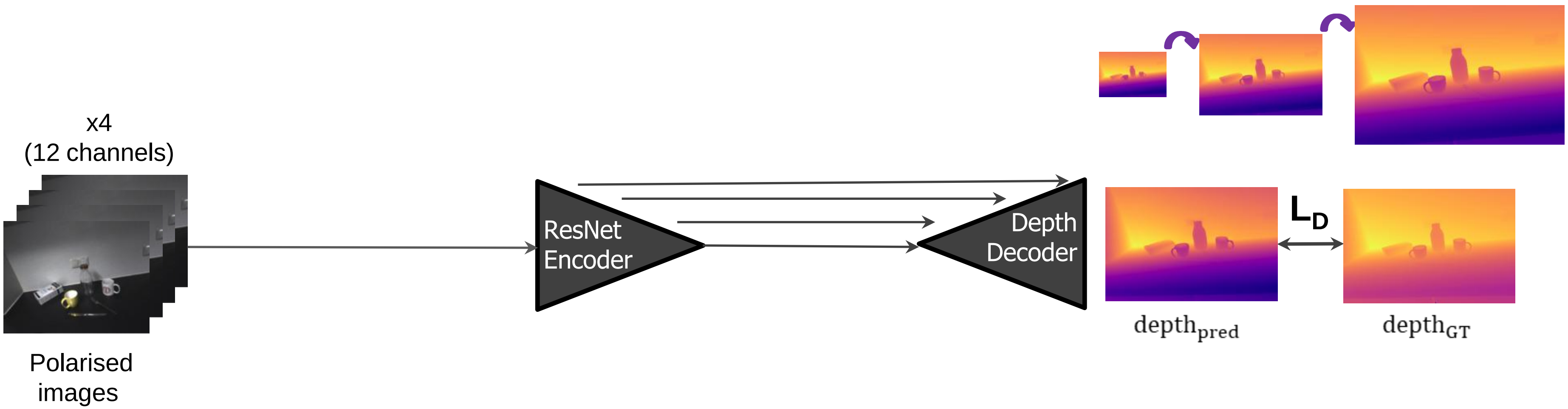
  - Discourage shrinking of the estimated depth

- Overall:

$$\mathcal{L} = \alpha \mathcal{L}_1 + \beta \mathcal{L}_s$$

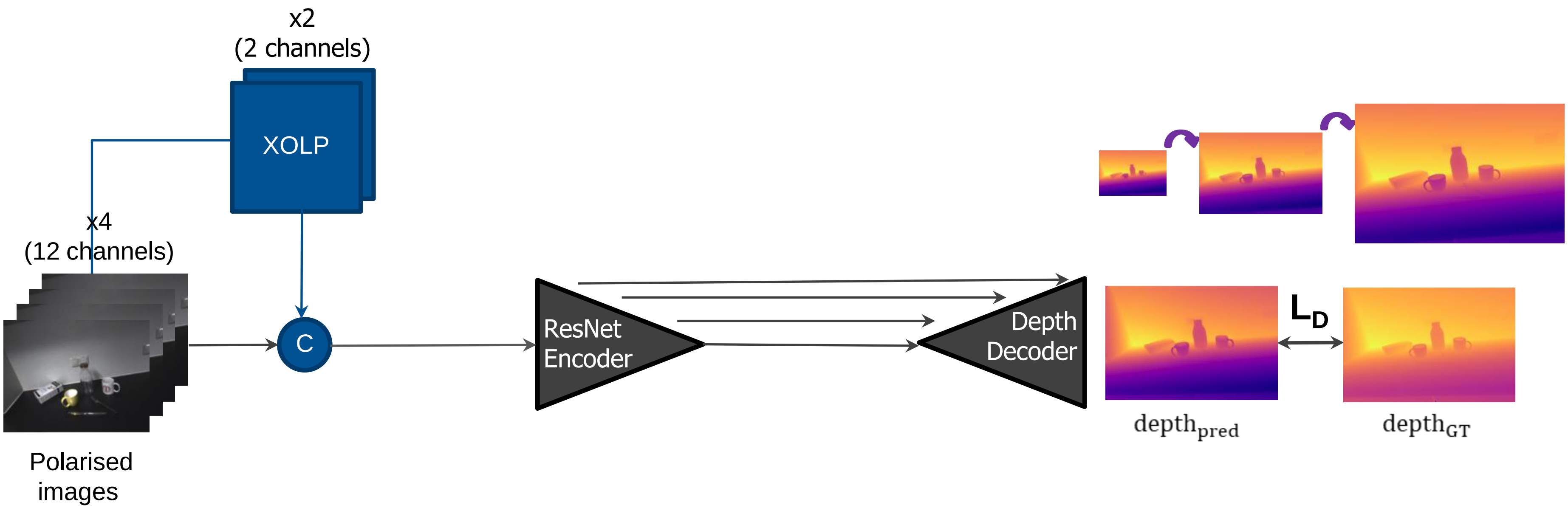


Architecture - proof of concept



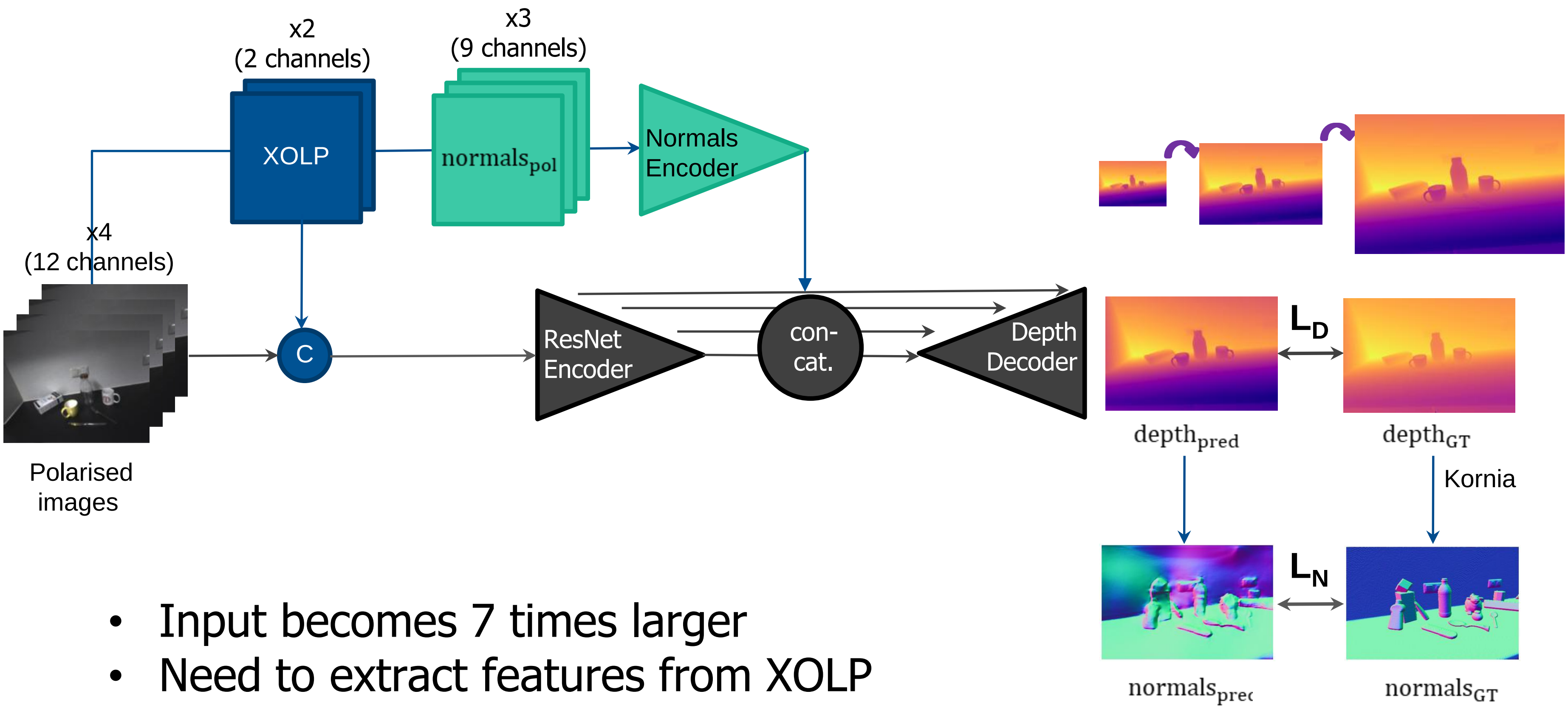
- Performs better
- Polarization contains significant information

Architecture - blending the priors



- Adding XOLP alone did not increase the results
- Can still add more information

Transition into the final architecture



- Input becomes 7 times larger
- Need to extract features from XOLP

Loss Functions

- Depth Regression

$$\mathcal{L}_1(y, \hat{y}) = |y - \hat{y}|$$

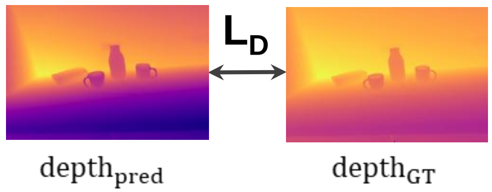
- Smoothing Loss (d stands for disparity):

$$\mathcal{L}_s(d, \hat{I}) = \frac{\partial d}{\partial x}\, e^{-\partial \hat{I}/\partial x} + \frac{\partial d}{\partial y}\, e^{-\partial \hat{I}/\partial y}$$

  - Discourage shrinking of the estimated depth

- Overall:

$$\mathcal{L} = \underbrace{\alpha \mathcal{L}_1 + \beta \mathcal{L}_s}_{L_D}$$



Loss Functions

- Depth Regression

$$\mathcal{L}_1(y, \hat{y}) = |y - \hat{y}|$$

- Smoothing Loss (d stands for disparity):

$$\mathcal{L}_s(d, \hat{I}) = \frac{\partial d}{\partial x}\, e^{-\partial \hat{I}/\partial x} + \frac{\partial d}{\partial y}\, e^{-\partial \hat{I}/\partial y}$$

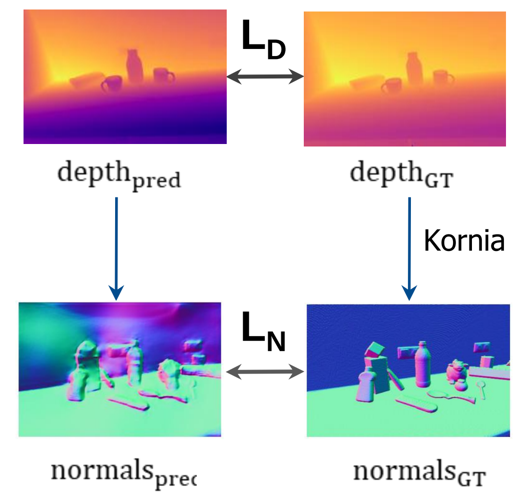
  - Discourage shrinking of the estimated depth

- Normals Loss

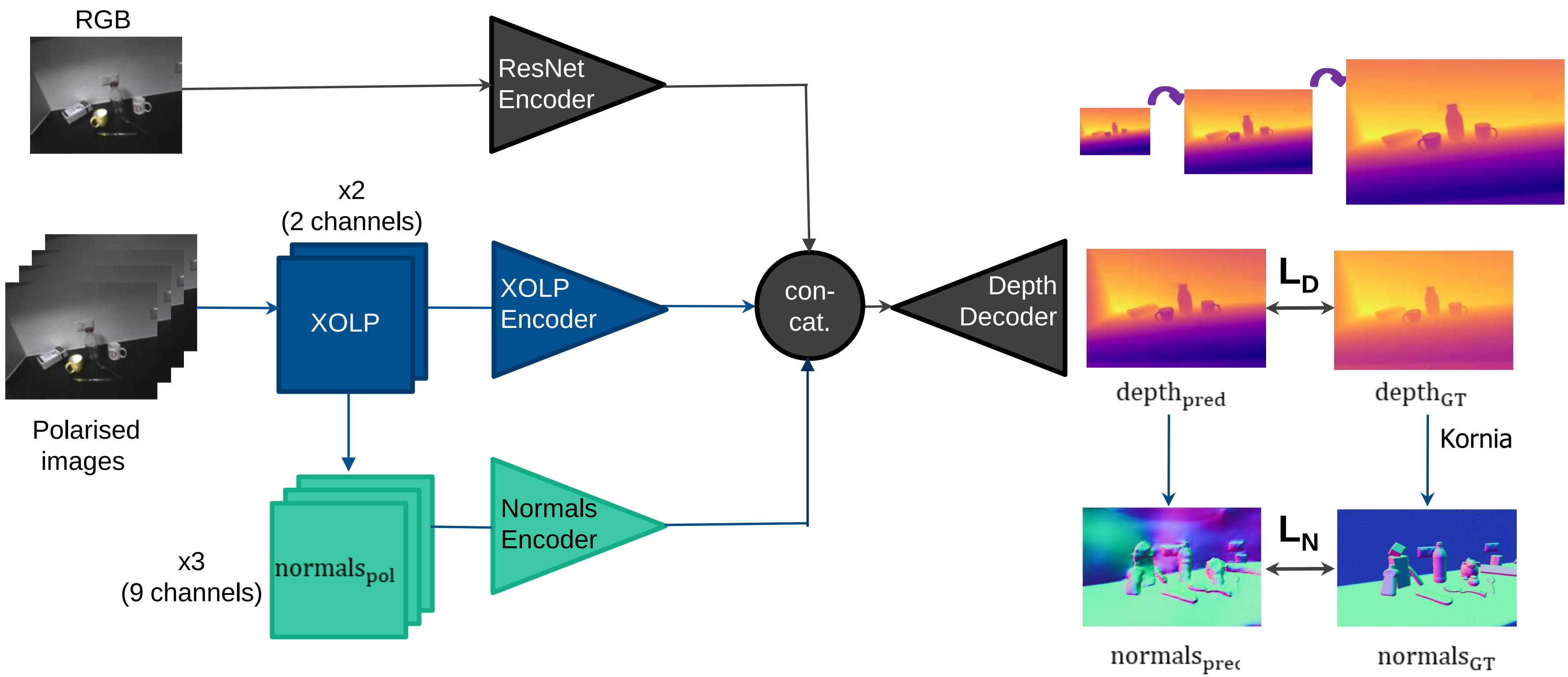
$$\mathcal{L}_n(n, \hat{n}) = 1 - cos(\angle(n, \hat{n}))$$

- Overall:

$$\mathcal{L} = \underbrace{\alpha \mathcal{L}_1 + \beta \mathcal{L}_s}_{\mathbf{L_D}} + \underbrace{\theta \mathcal{L}_n}_{\mathbf{L_N}}$$

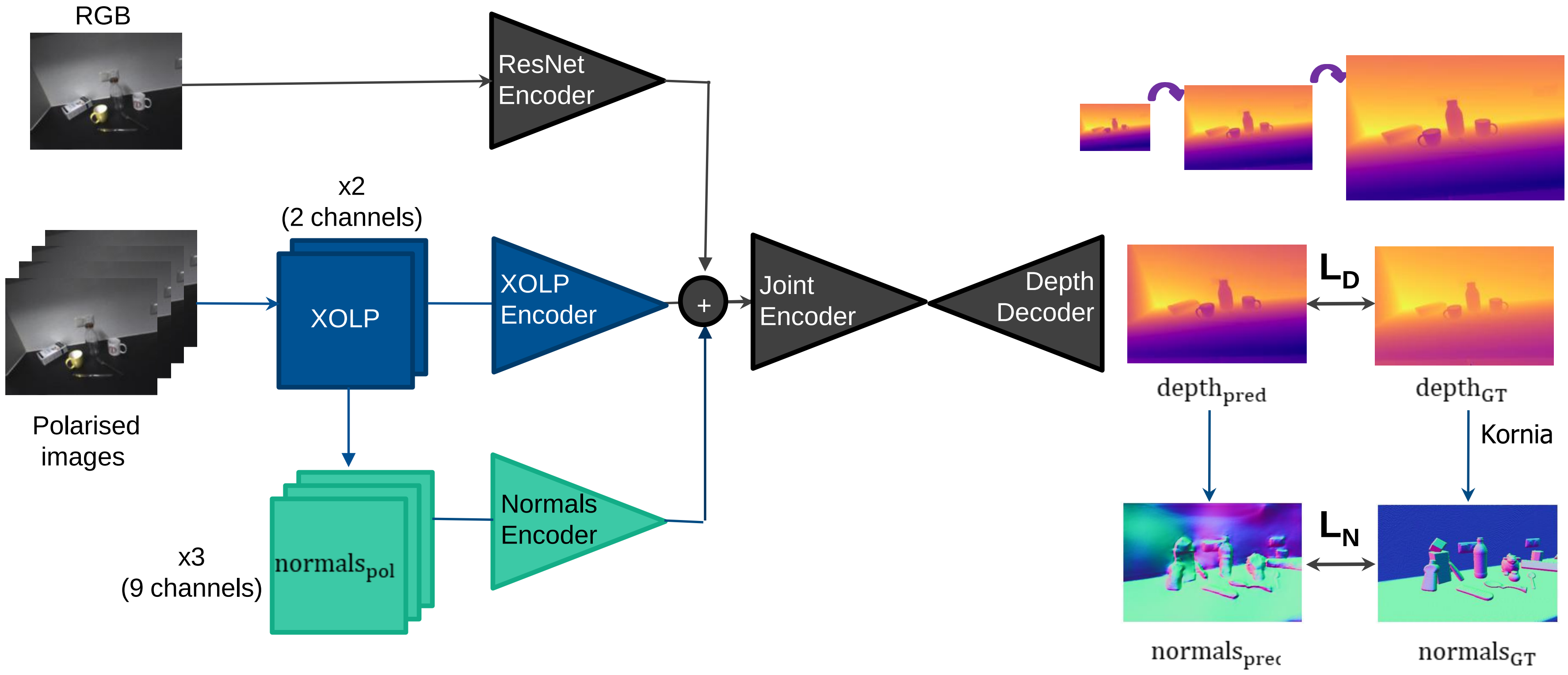


Transition into the final architecture

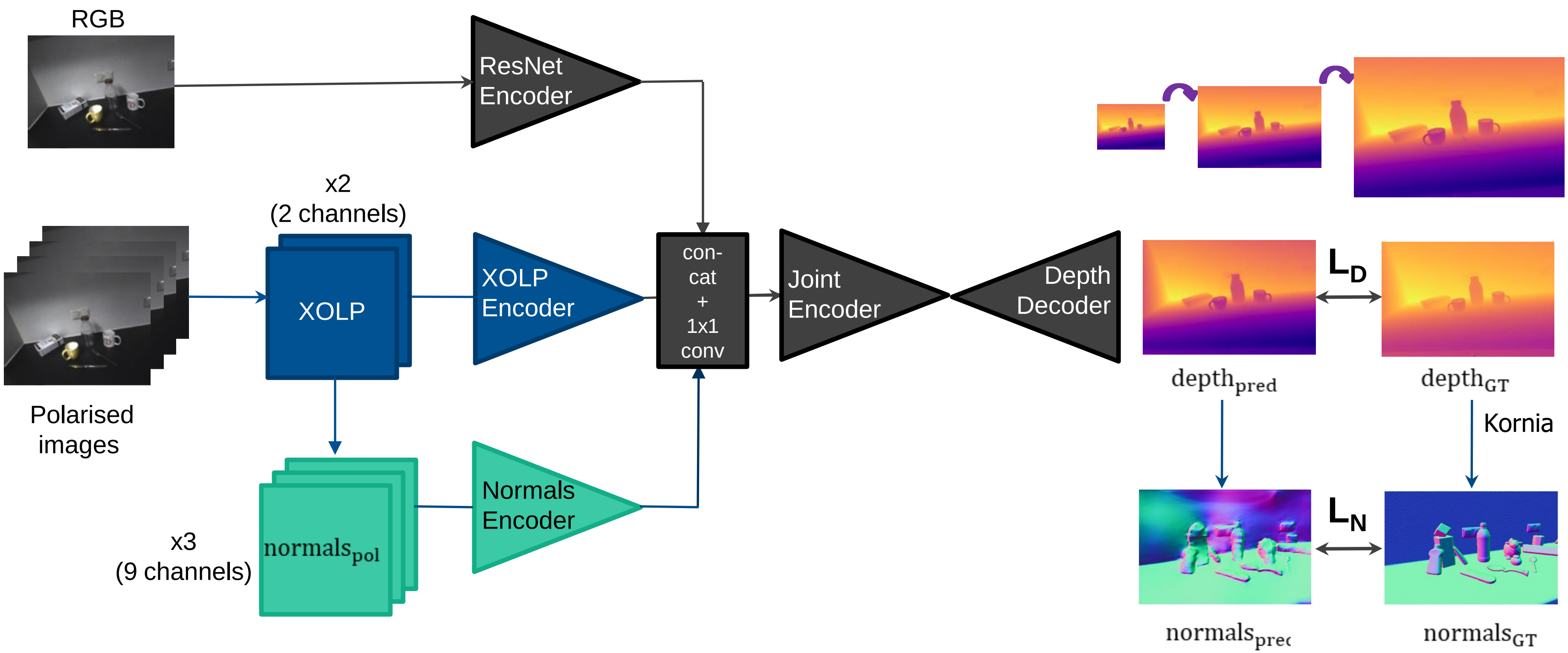


- Too much load on the decoder

Transition into the final architecture

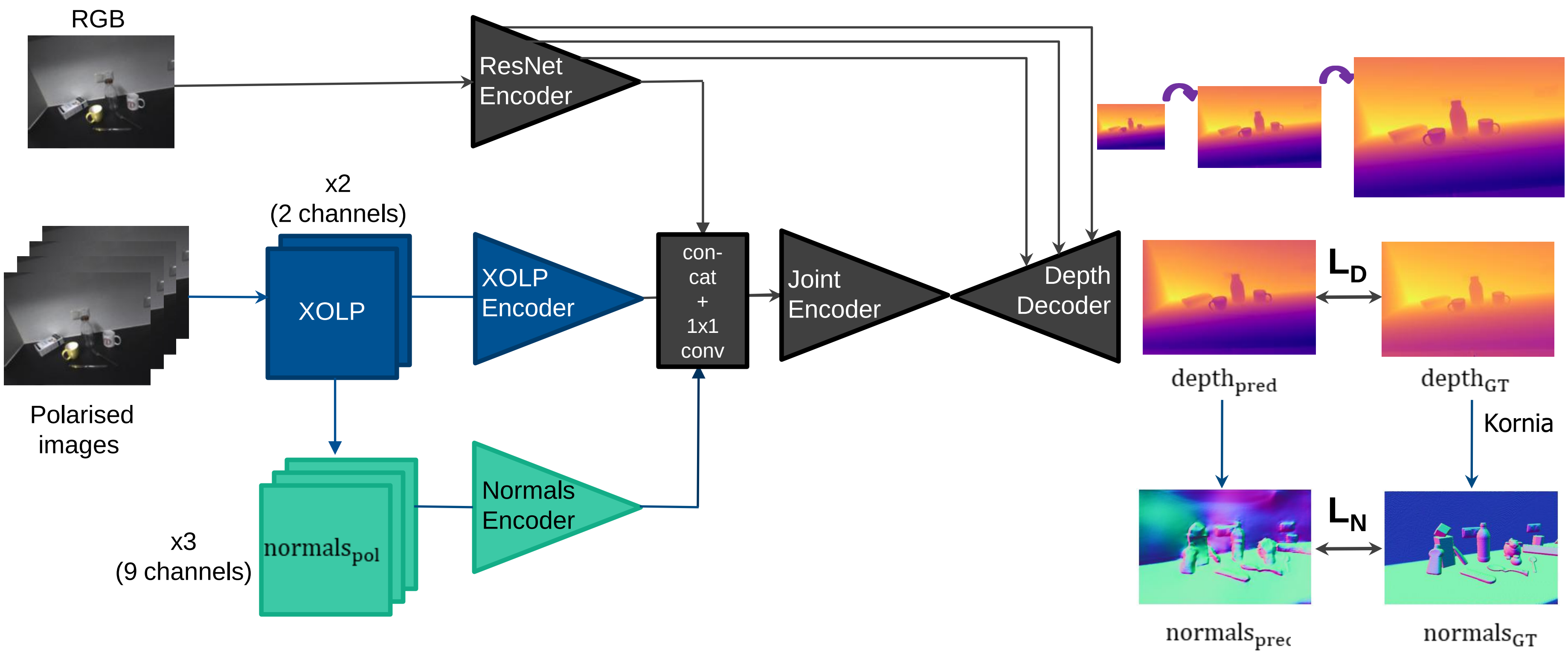


Transition into the final architecture



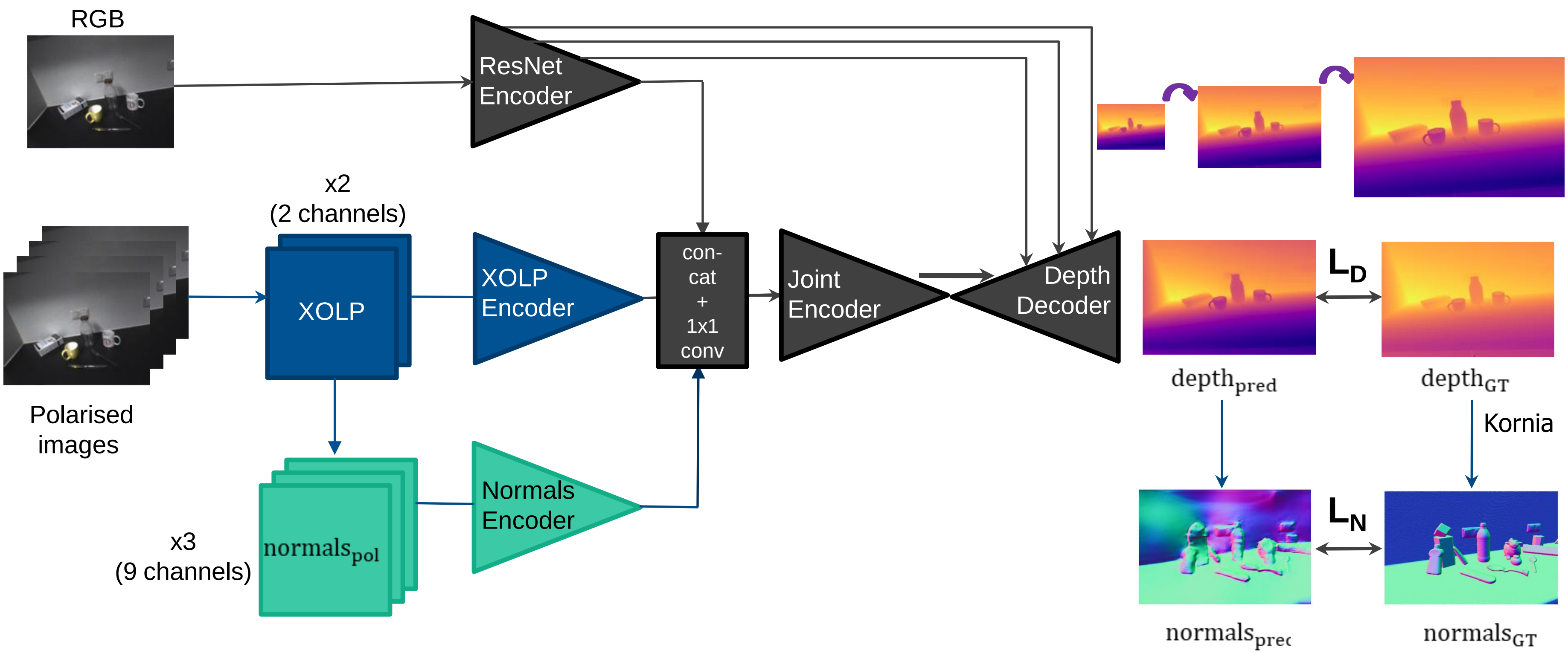
- Combine features channel-wise

Transition into the final architecture



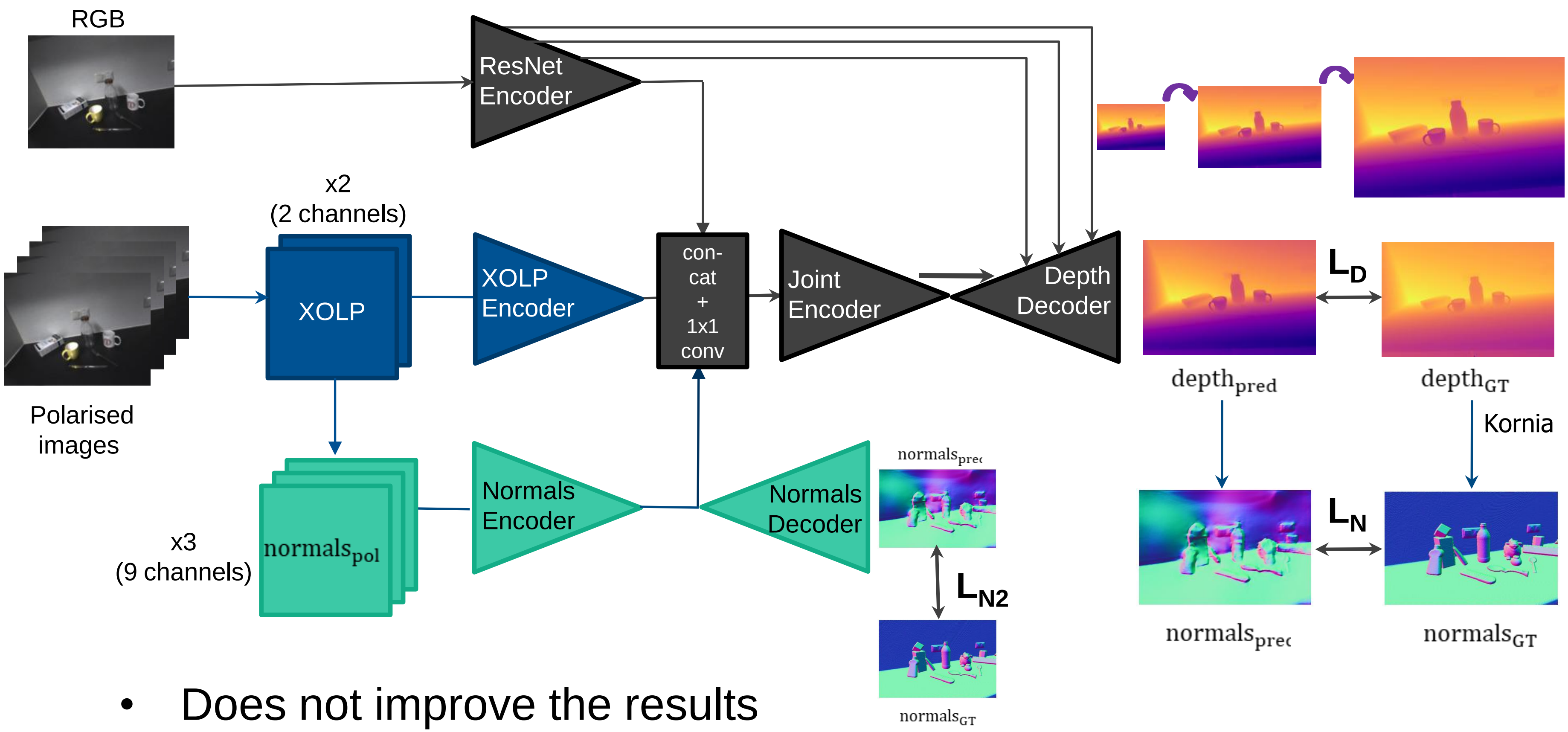
- Skip connections for high resolution depth estimation

Transition into the final architecture



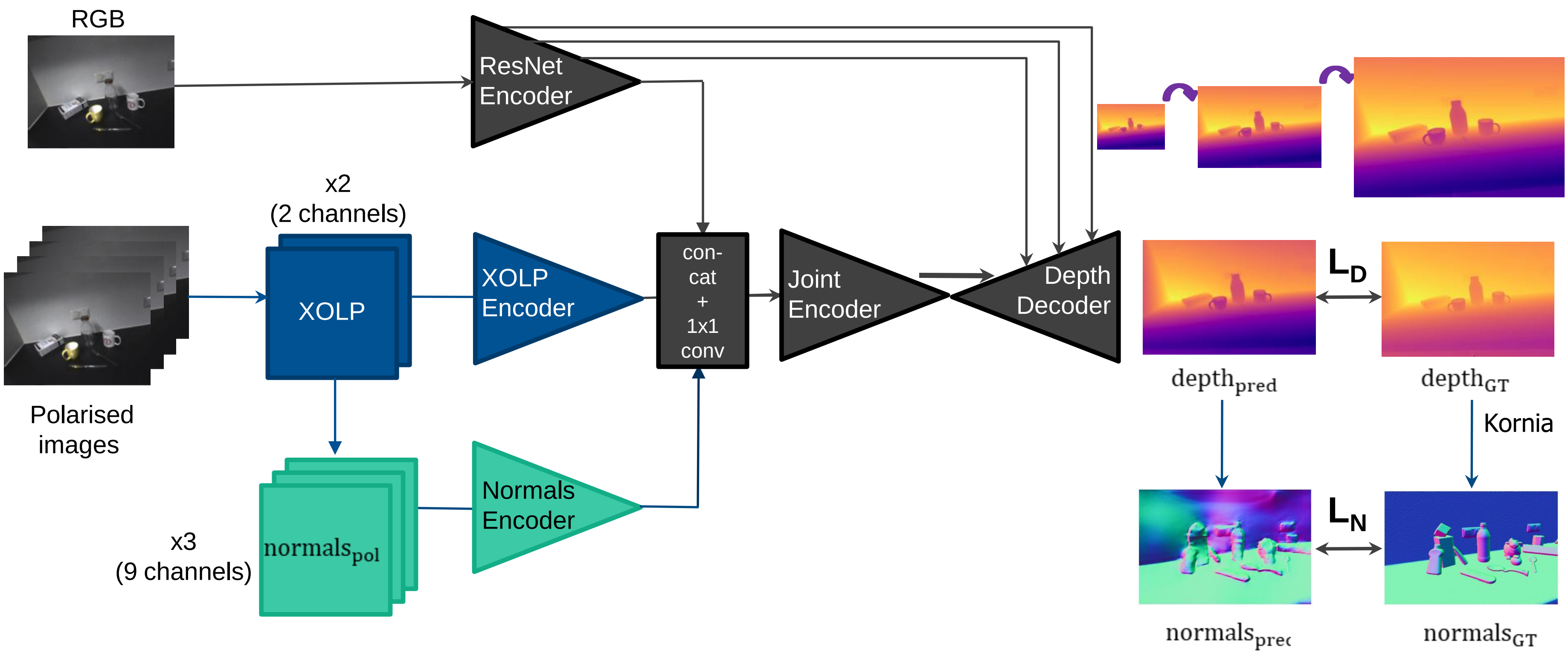
- Skip connections from the joint encoder boost results for non-Lambertian objects

Transition into the final architecture



- Does not improve the results

Final architecture



Ablations

Ablations – losses

$$\mathcal{L} = \alpha\mathcal{L}_1 + \beta\mathcal{L}_s + \theta\mathcal{L}_n$$

GLASS	a1	abs_rel	log_rms	rms	sq_rel
θ = 0	0.9575	0.08241	0.09517	0.06184	0.00762
θ = 0.35	**0.9896**	**0.08115**	**0.09086**	**0.05855**	**0.00652**
θ = 1	0.9566	0.09645	0.10460	0.06821	0.00930
β = 0	0.9628	0.09977	0.10770	0.07119	0.00988
losses only at scale 0	0.9456	0.08838	0.10110	0.06576	0.00825

Initial setup: α = 1, β = 1, θ = 0.35; scales: 0, 1, 2, 3. The table indicates changes of the specified values.

Ablations – losses

$$\mathcal{L} = \alpha\mathcal{L}_1 + \beta\mathcal{L}_s + \theta\mathcal{L}_n$$

Initial setup: α = 1, β = 1, θ = 0.35; scales: 0, 1, 2, 3. The table indicates changes of the specified values.

GLASS	a1	abs_rel	log_rms	rms	sq_rel
θ = 0	0.9575	0.08241	0.09517	0.06184	0.00762
θ = 0.35	**0.9896**	**0.08115**	**0.09086**	**0.05855**	**0.00652**
θ = 1	0.9566	0.09645	0.10460	0.06821	0.00930
β = 0	0.9628	0.09977	0.10770	0.07119	0.00988
losses only at scale 0	0.9456	0.08838	0.10110	0.06576	0.00825

METAL	a1	abs_rel	log_rms	rms	sq_rel
θ = 0	0.9113	0.1278	0.1335	0.08742	0.01389
θ = 0.35	**0.9767**	**0.1135**	**0.1218**	**0.07607**	**0.01006**
θ = 1	0.9021	0.1419	0.1476	0.09614	0.01609
β = 0	0.8377	0.1586	0.1607	0.10720	0.02086
losses only at scale 0	0.8700	0.1495	0.1492	0.09899	0.01765

Ablations – polarimetric characteristics

GLASS	a1	abs_rel	log_rms	rms	sq_rel
RGB	0.9417	0.10610	0.11730	0.07361	0.009677
RGB + XOLP	**0.9908**	**0.06904**	**0.08126**	**0.05221**	**0.004954**
RGB + normals	0.9723	0.08807	0.09818	0.06648	0.008598
RGB + XOLP + normals	0.9896	0.08115	0.09086	0.05855	0.006523

METAL	a1	abs_rel	log_rms	rms	sq_rel
RGB	0.9700	**0.09246**	0.1103	0.07016	0.008178
RGB + XOLP	**0.9974**	0.09275	**0.1029**	**0.06387**	**0.006764**
RGB + normals	0.9884	0.11570	0.1237	0.07703	0.009956
RGB + XOLP + normals	0.9767	0.11350	0.1218	0.07607	0.010060

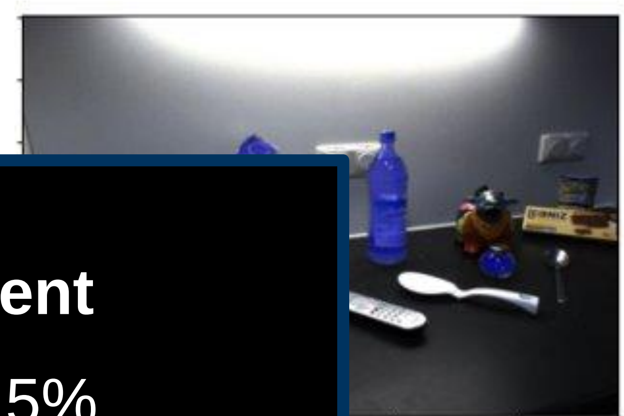
Ablations – polarimetric characteristics

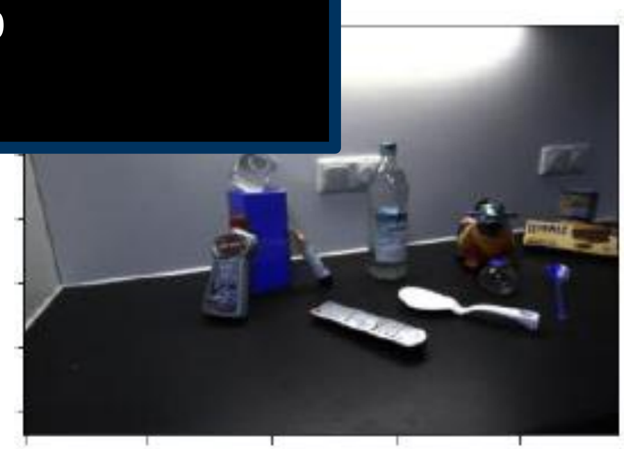
GLASS	a1	abs_rel	log_rms	rms	sq_rel
RGB	0.9417	0.10610	0.1172	0.07361	0.009677
RGB + XOLP					
RGB + normals					
RGB + XOLP + normals					

Relative improvement

transparent objects: 5%
metal objects: 3%

METAL					
RGB	0.9700	**0.09246**	0.1103	0.07016	0.008178
RGB + XOLP	**0.9974**	0.09275	**0.1029**	**0.06387**	**0.006764**
RGB + normals	0.9884	0.11570	0.1237	0.07703	0.009956
RGB + XOLP + normals	0.9767	0.11350	0.1218	0.07607	0.010060





Ablations – polarimetric characteristics

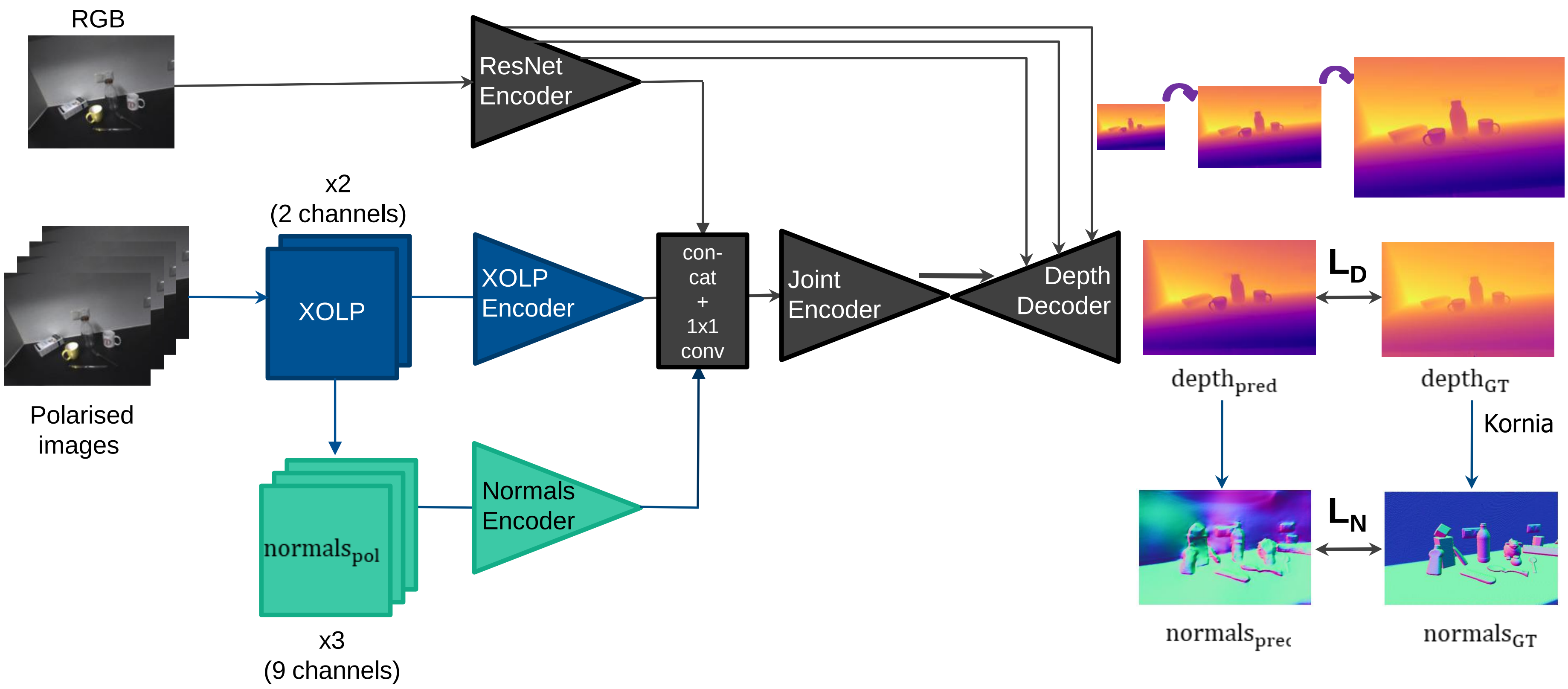
OBJECTS	a1	abs_rel	log_rms	rms	sq_rel
RGB	**0.93670**	0.10331	**0.12908**	**0.08602**	**0.0123**
RGB + XOLP	0.91344	**0.09740**	0.13261	0.09063	0.0129
RGB + normals	0.91583	0.10257	0.13679	0.09447	0.0140
RGB + XOLP + normals	0.92258	0.10347	0.13486	0.09236	0.0132

Ablations – polarimetric characteristics

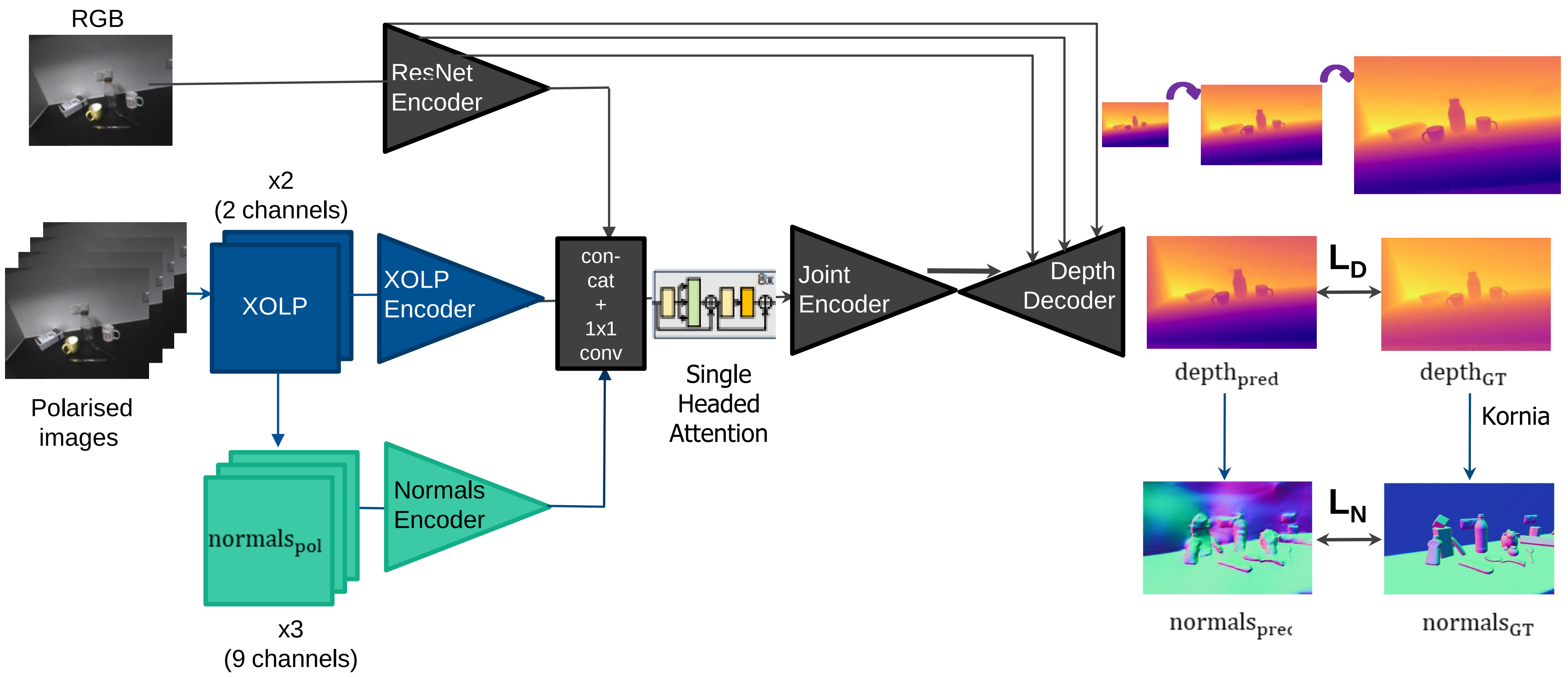
OBJECTS	a1	abs_rel	log_rms	rms	sq_rel
RGB	**0.93670**	0.10331	**0.12908**	**0.08602**	**0.0123**
RGB + XOLP	0.91344	**0.09740**	0.13261	0.09063	0.0129
RGB + normals	0.91583	0.10257	0.13679	0.09447	0.0140
RGB + XOLP + normals	0.92258	0.10347	0.13486	0.09236	0.0132

OBJECTS WITHOUT BOX	a1	abs_rel	log_rms	rms	sq_rel
RGB	0.95123	0.09958	0.12432	0.07799	0.01144
RGB + XOLP	0.97161	0.08409	**0.11243**	**0.07149**	**0.00892**
RGB + normals	0.96648	0.09153	0.11942	0.07777	0.01055
RGB + XOLP + normals	**0.97417**	**0.09036**	0.11605	0.07379	0.00943

Final architecture



Final architecture – with attention

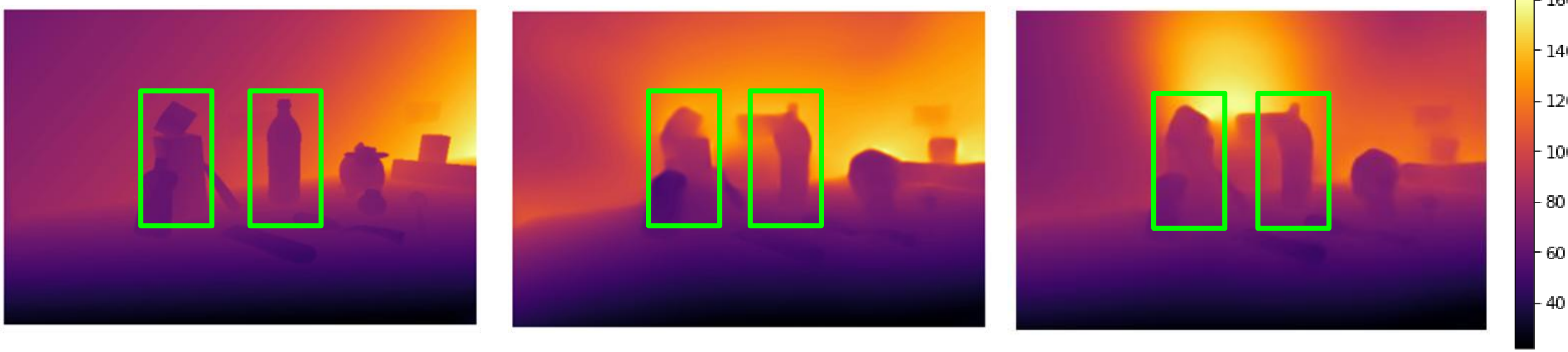


Ablations – polarimetric characteristics

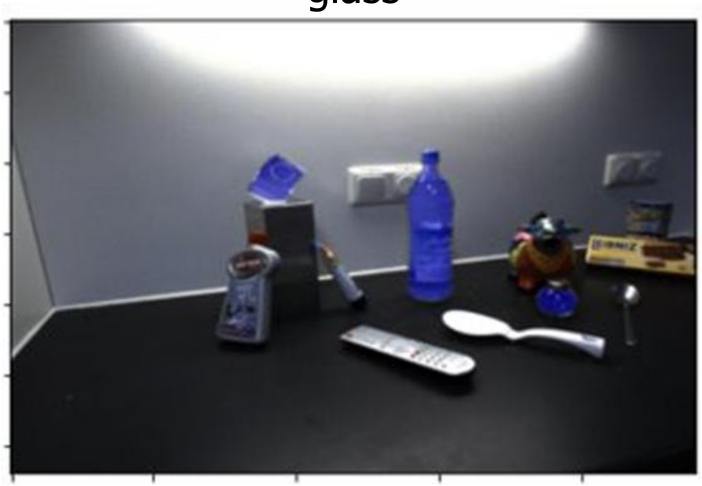
OBJECTS	a1	abs_rel	log_rms	rms	sq_rel
RGB	0.93670	0.10331	0.12908	0.08602	0.0123
RGB + XOLP	0.91344	0.09740	0.13261	0.09063	0.0129
RGB + normals	0.91583	0.10257	0.13679	0.09447	0.0140
RGB + XOLP + normals	0.92258	0.10347	0.13486	0.09236	0.0132
RGB + XOLP + normals + attention	**0.96769**	**0.08841**	**0.11351**	**0.07738**	**0.0010**

Qualitative analysis – polarimetry

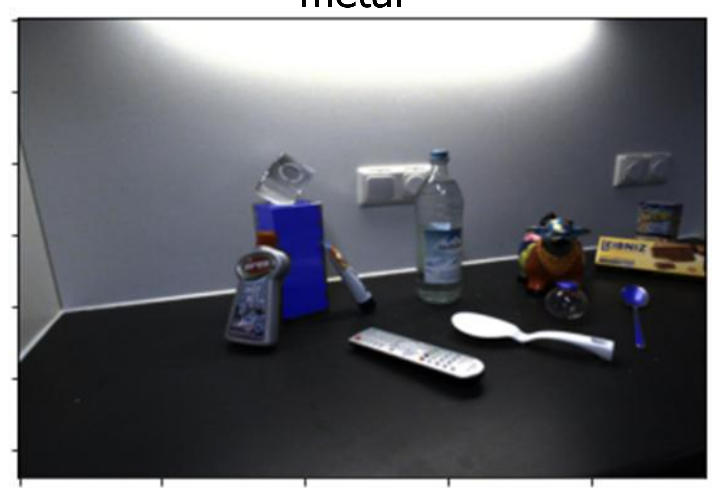
GT RGB RGB + XOLP + normals



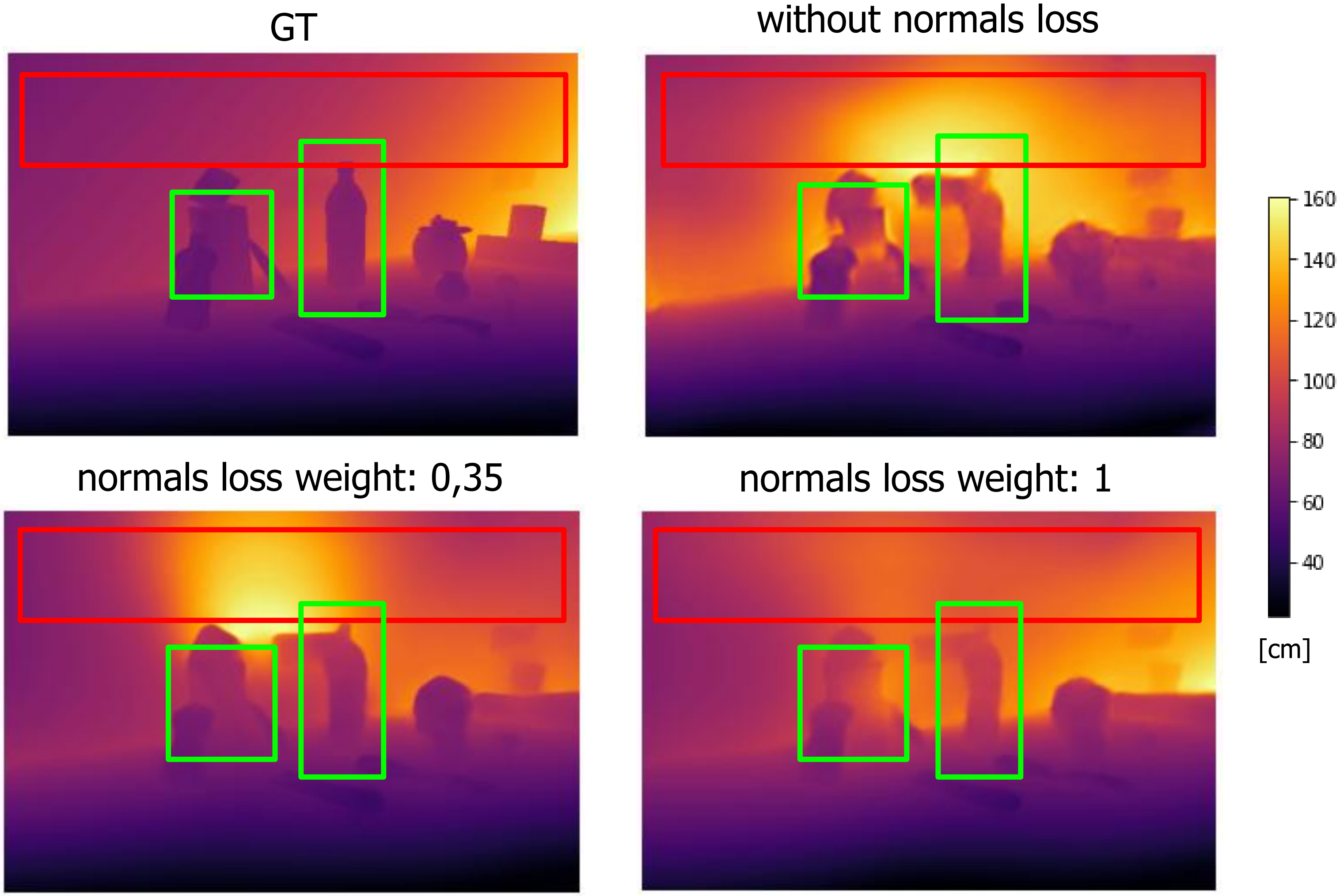
glass



metal



Qualitative analysis – normals loss

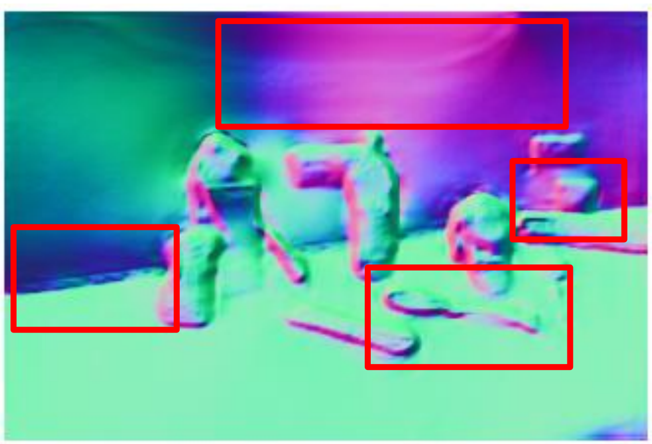


Qualitative analysis – normals loss

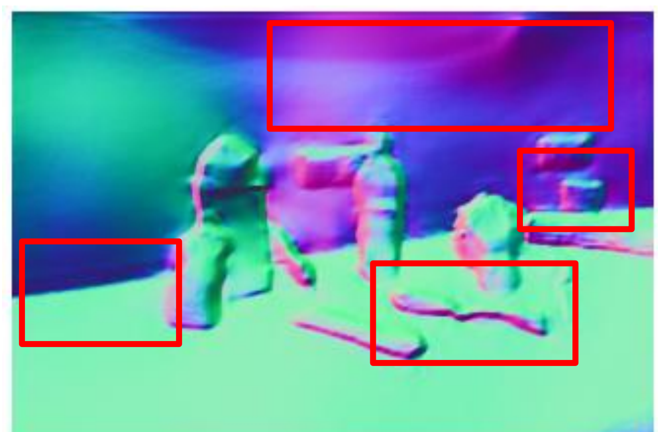
GT

without normals loss

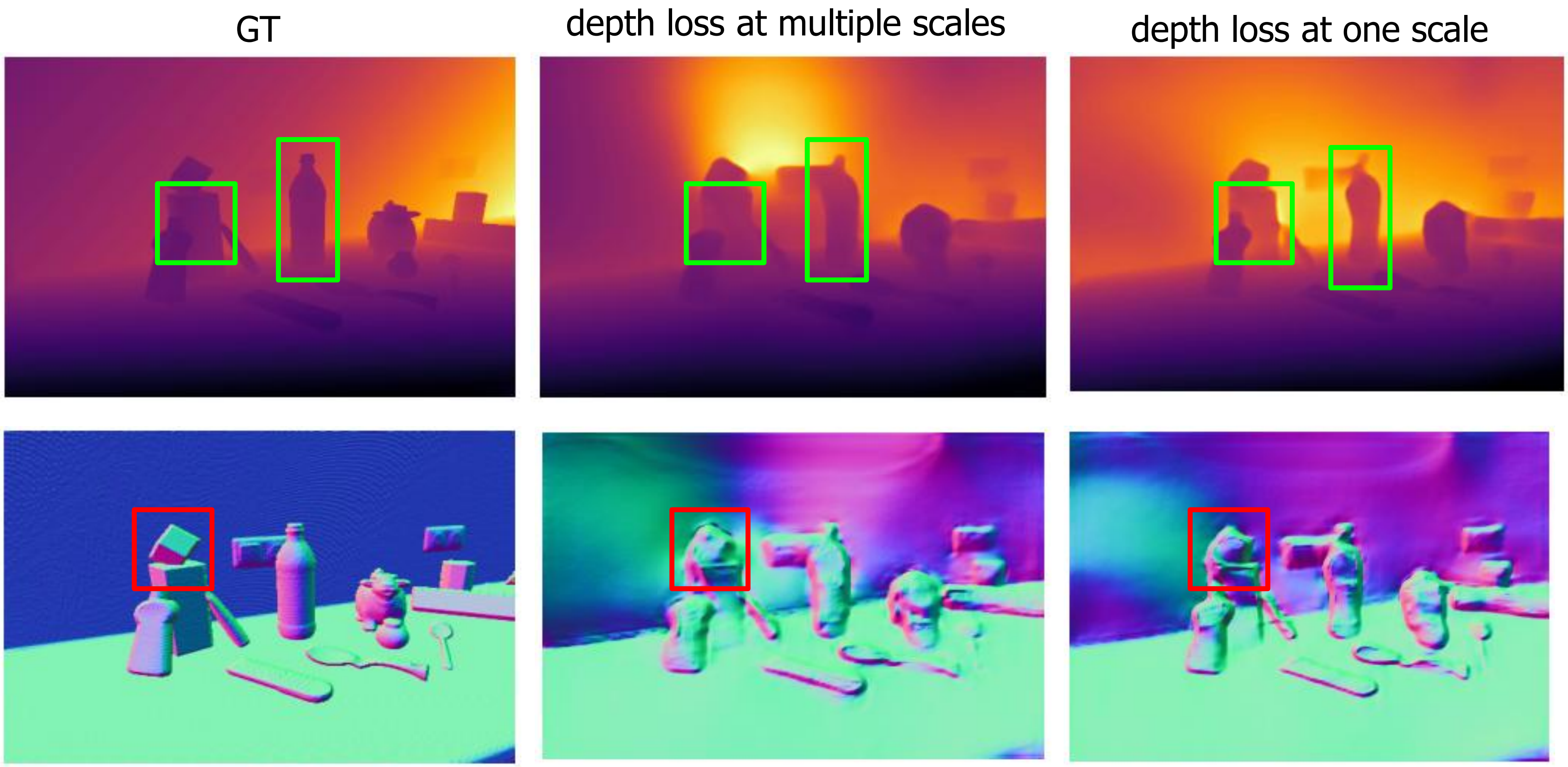
normals loss weight: 0,35



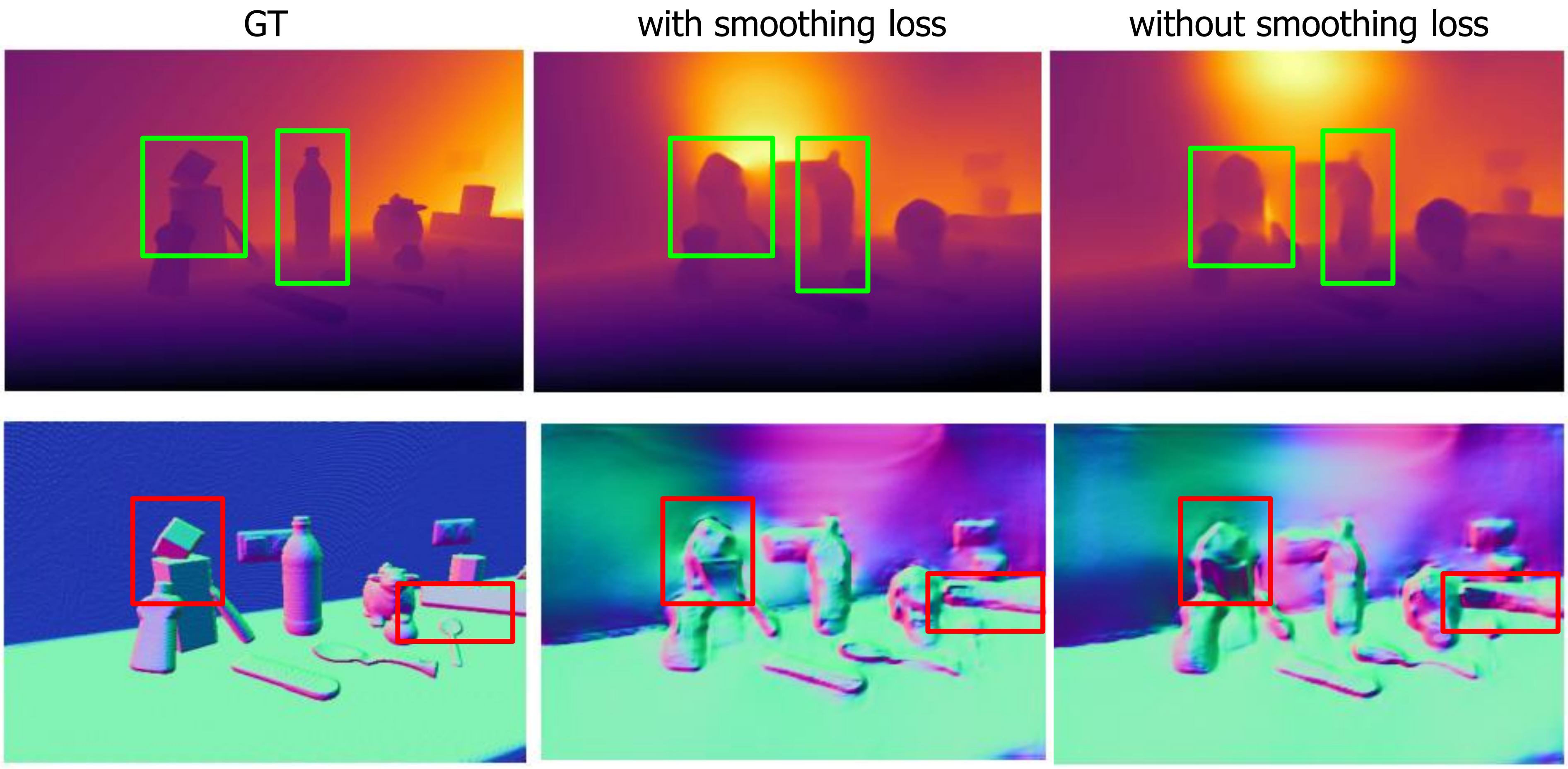
normals loss weight: 1



Qualitative analysis – loss at multiple scales



Qualitative analysis – smoothing loss

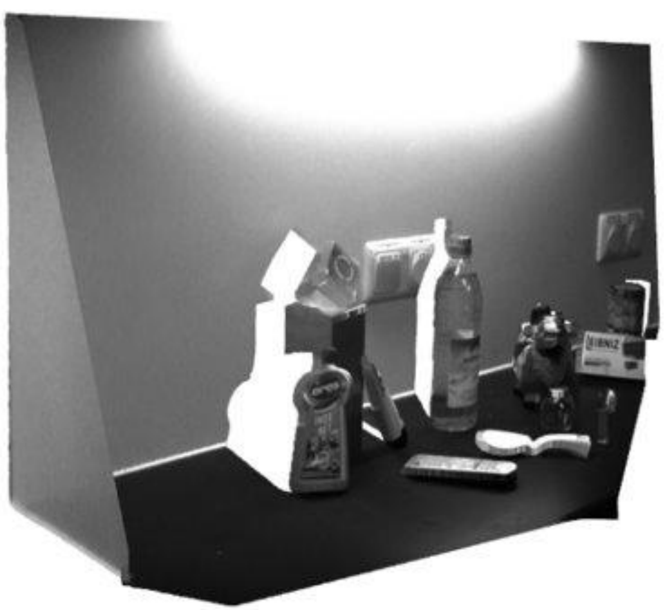


Point Cloud

GT

final architecture's prediction

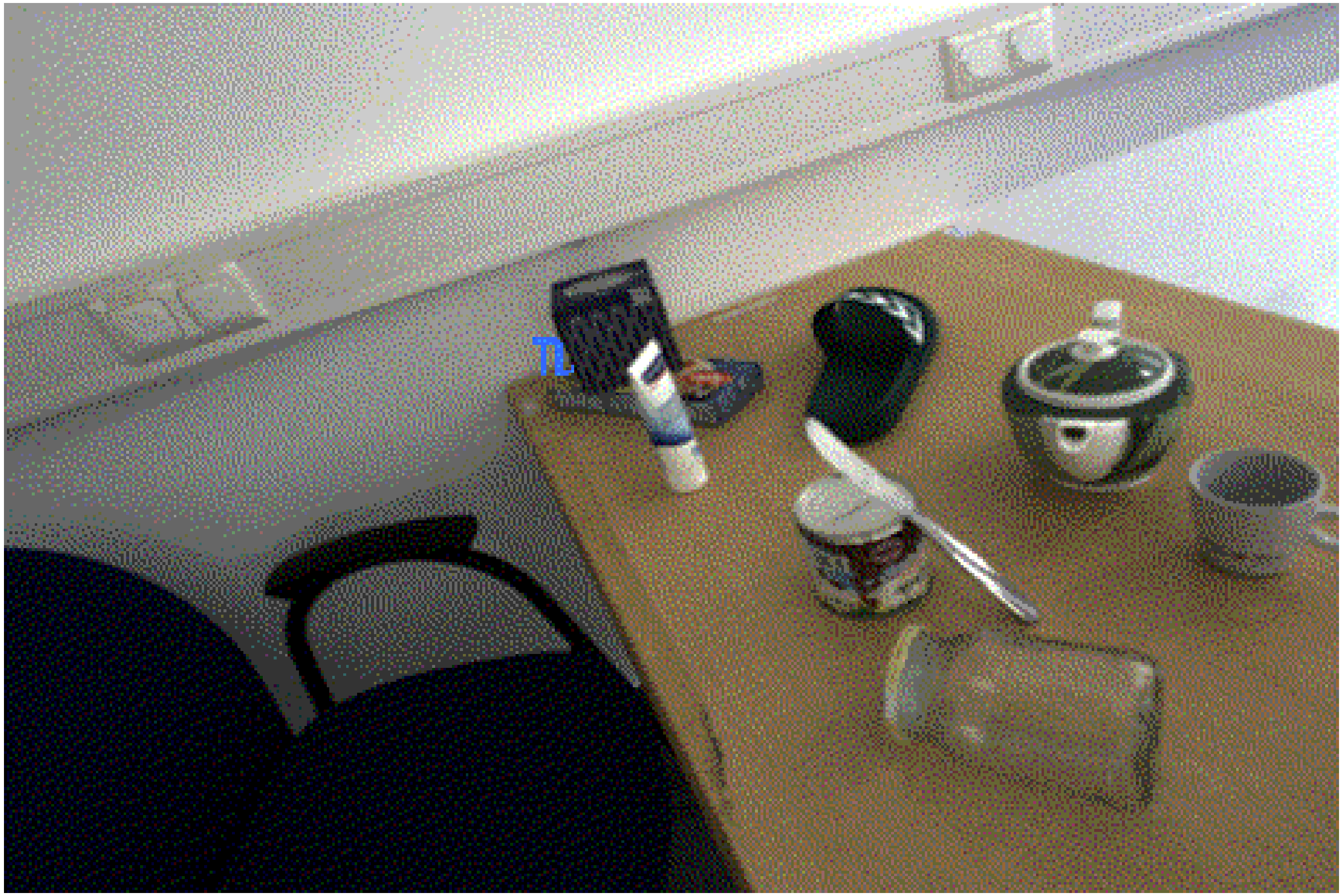
Point Cloud

GT	final architecture's prediction
	

Demonstration with the RGB prediction

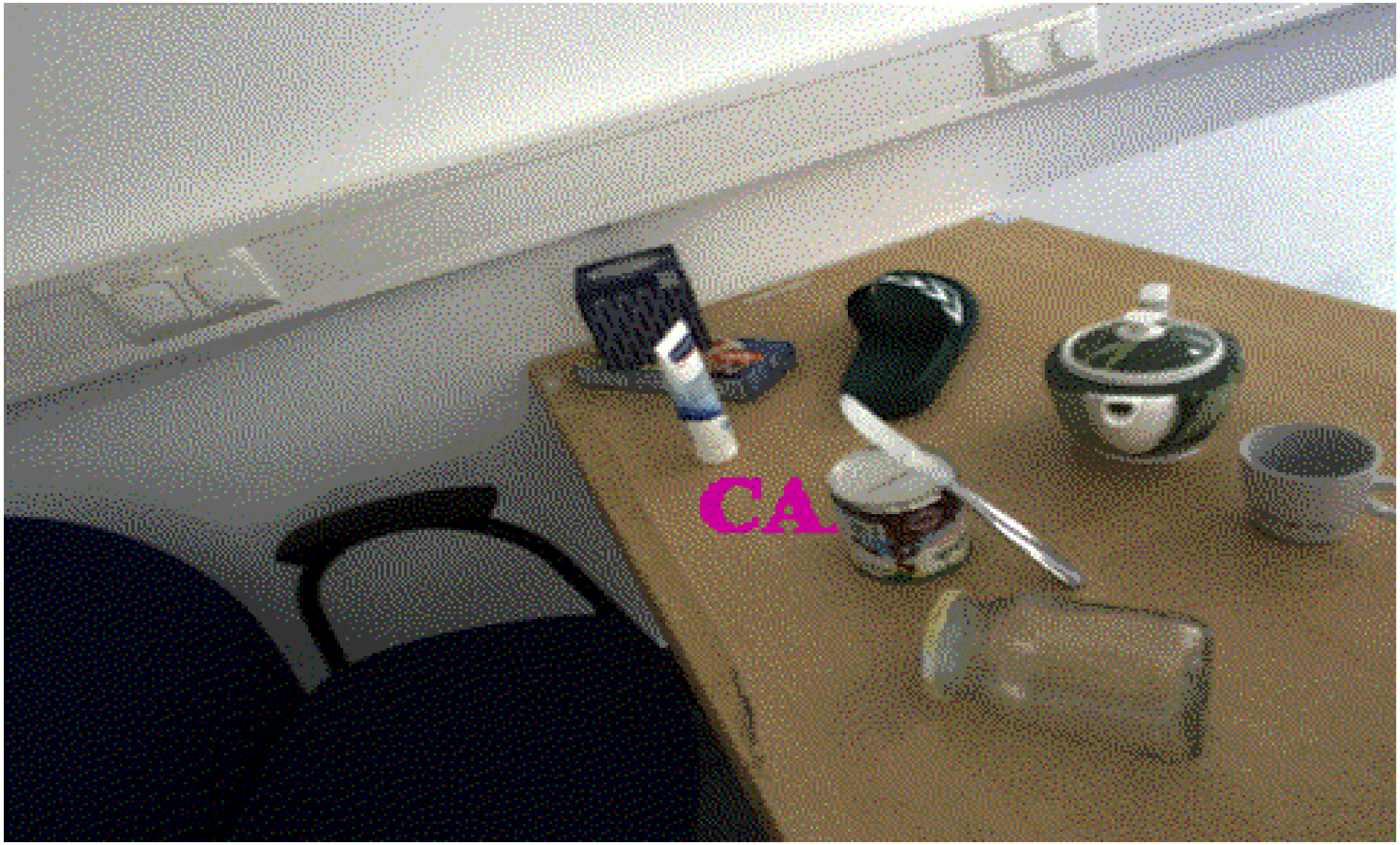
Demonstration with the final prediction

Demonstration with the final prediction



Demonstration with the final prediction

Demonstration with the final prediction

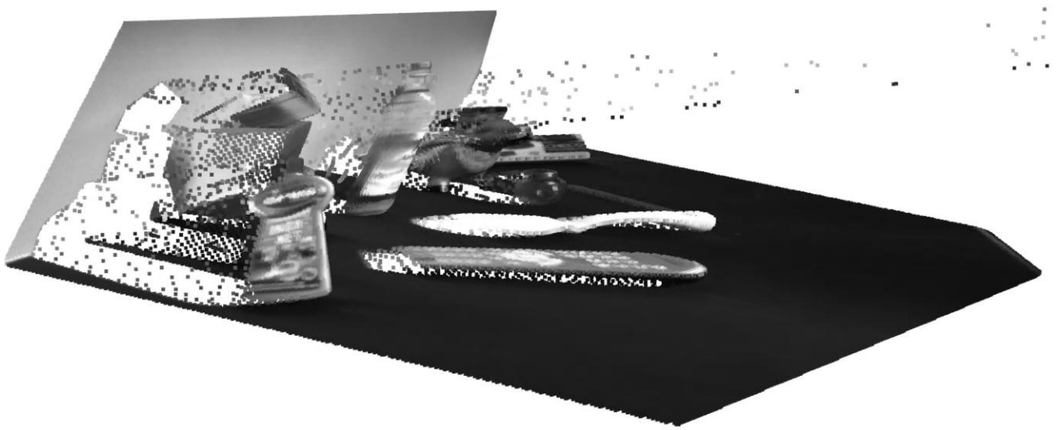


Limitations

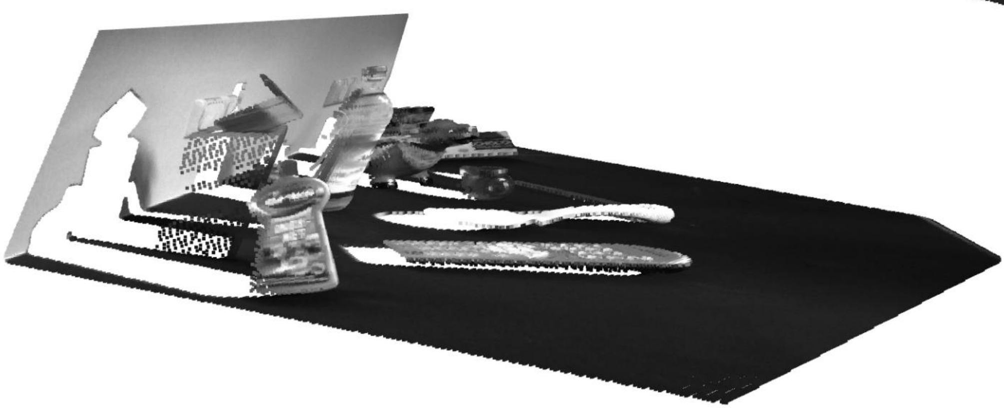
Depth artifacts at object edges influenced by:

- the used interpolation type
- convolutions

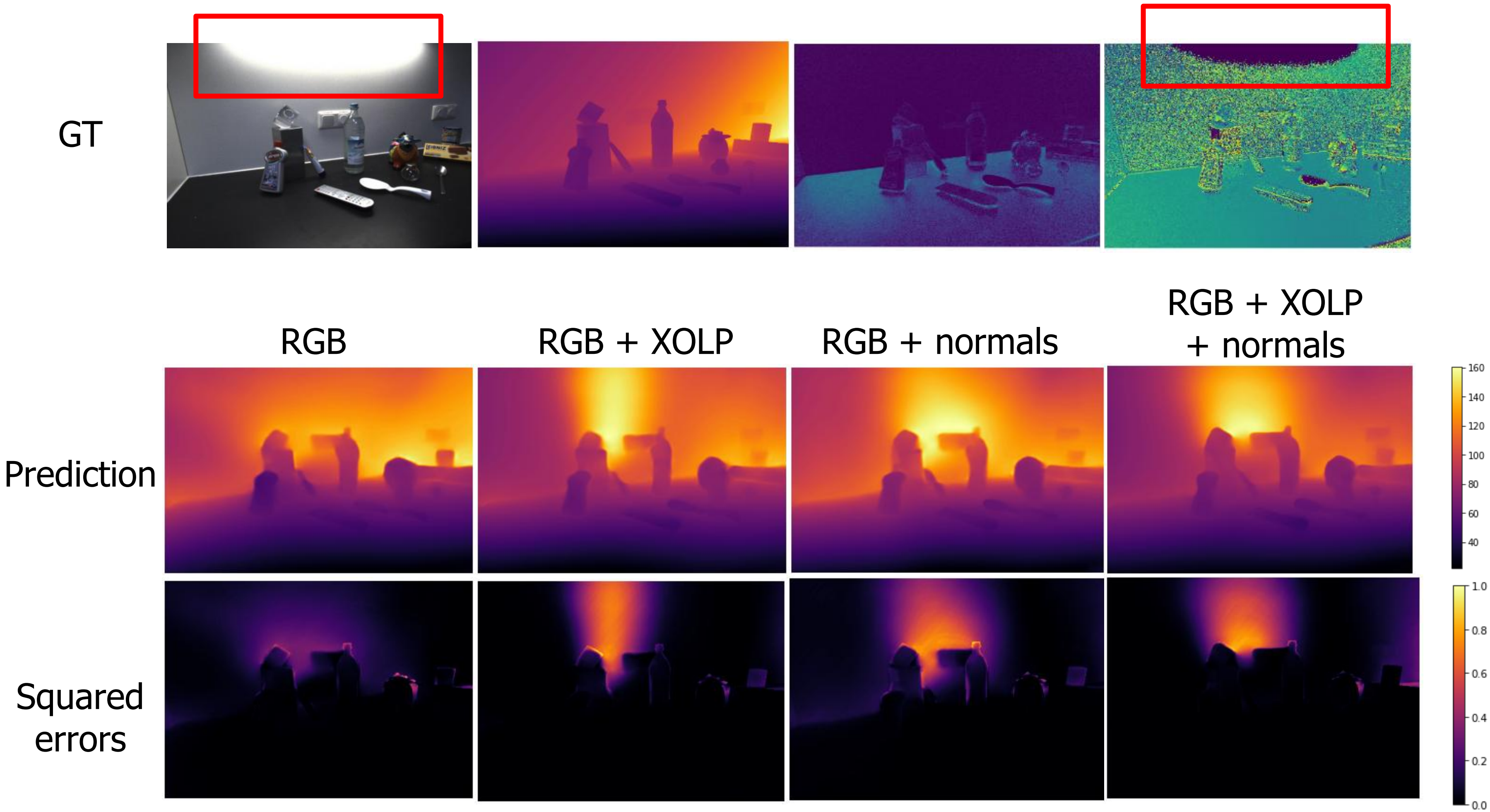
depth GT + bilinear interpolation



depth GT + nearest interpolation



Limitations



Limitations

- Drop of performance on the whole scene when using polarimetric characteristics, mostly because of the background influence

WHOLE SCENE	a1	abs_rel	log_rms	rms	sq_rel
RGB	**0.9690**	**0.0817**	**0.1050**	**0.0835**	**0.0085**
RGB + XOLP	0.9023	0.0978	0.1298	0.1138	0.0131
RGB + normals	0.8653	0.1018	0.1405	0.1094	0.0159
RGB + XOLP + normals	0.8977	0.0995	0.1330	0.1094	0.0138

RGB

RGB (error)

RGB + XOLP + normals

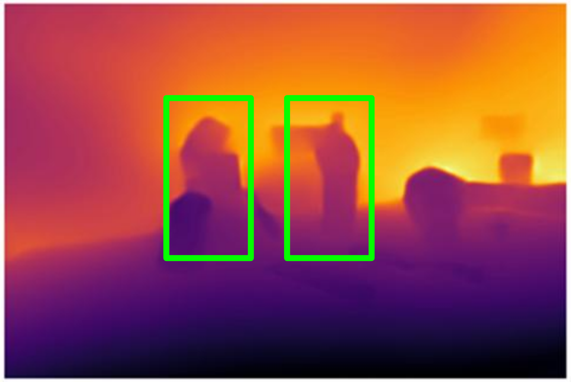
RGB + XOLP + normals (error)

Conclusions

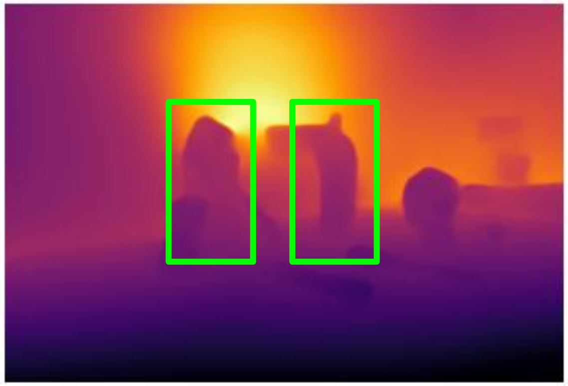
- We implemented a supervised monocular depth prediction model leveraging polarimetric characteristics

- Our depth estimation for photometrically challenging objects outperforms the plain RGB model

RGB

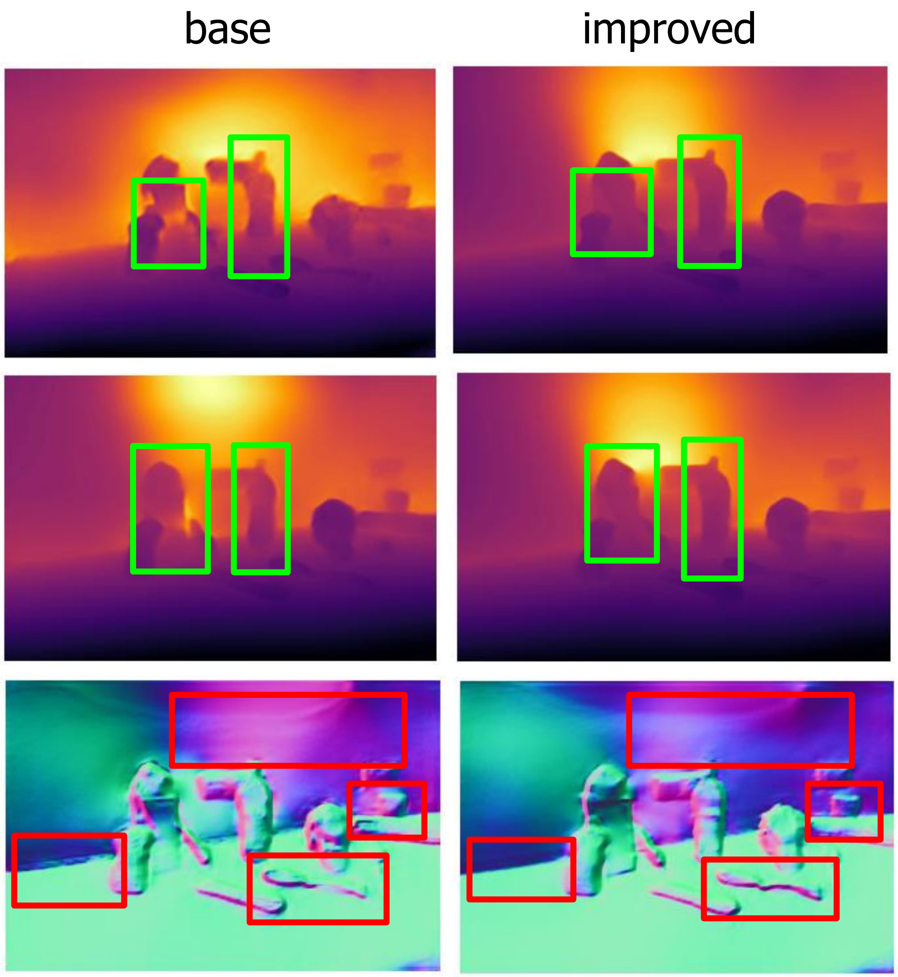


RGB + XOLP + normals



Conclusions

- Normals loss increases object smoothness and sharpness
- Smoothing loss and loss calculation at multiple scales prevent shrinking of objects
- Higher normals loss weight improves normals predictions



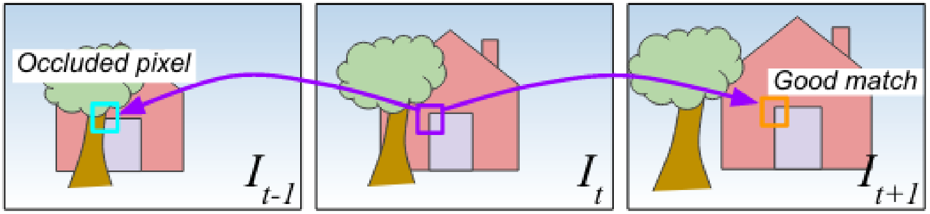
Conclusions

The developed model serves as

a solid base for further advancements

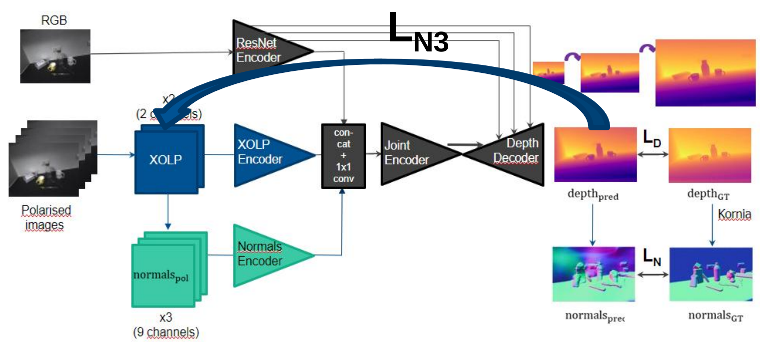
Future development

- Self-supervision (inspired by Monodepth2) as HAMMER includes trajectories



"Digging Into Self-Supervised Monocular Depth Estimation"
Clément Godard, Oisin Mac Aodha, Michael Firman, Gabriel Brostow; ICCV 2019

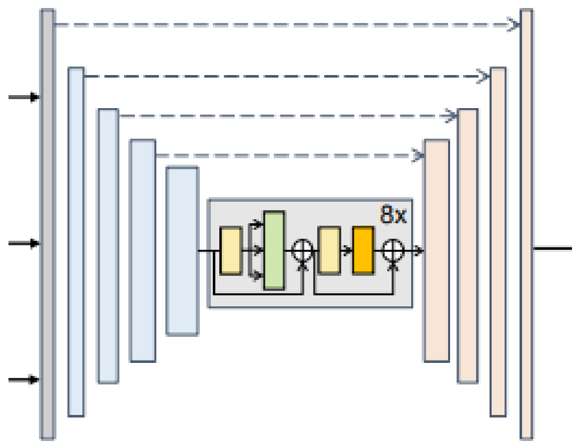
- Self-supervised losses leveraging inverse polarimetric transformations (as in CroMo)



"CroMo: Cross-Modal Learning for Monocular Depth Estimation."
Yannick Verdié, Jifei Song, Barnabé Mas, Benjamin Busam, Aleš Leonardis, Steven McDonagh; CVPR 2022

Future development

- Our architecture making use of attention gives best results – potential for further improvements



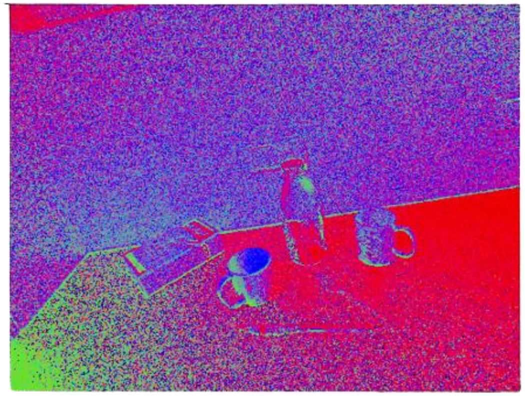
"Shape from Polarization for Complex Scenes in the Wild"
Chenyang Lei, Chenyang Qi, Jiaxin Xie, Na Fan, Vladlen Koltun and Qifeng Chen; CVPR 2022

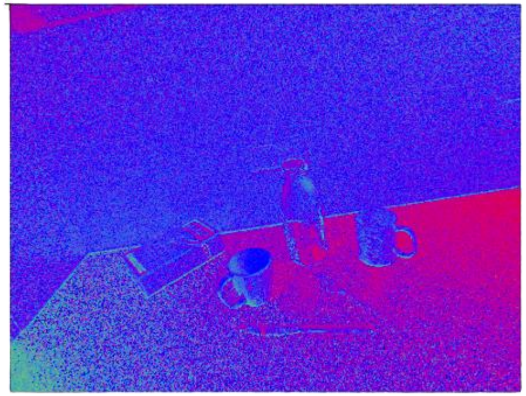
Future development

- Making use of additional modalities
  as HAMMER has more to offer:
  - ToF
  - real depth images

"Is my Depth Ground-Truth Good Enough? HAMMER - Highly Accurate Multi-Modal Dataset for DEnse 3D Scene Regression"
HyunJun Jung, Patrick Ruhkamp, Guangyao Zhai, Nikolas Brasch, Yitong Li, Yannick Verdie, Jifei Song, Yiren Zhou, Anil Armagan, Slobodan Ilic, Aleš Leonardis, Benjamin Busam; 2022

- Verifying influence of refractive indices (e.g. by using attention)






Q8