Unsupervised Single Image Underwater Depth Estimation

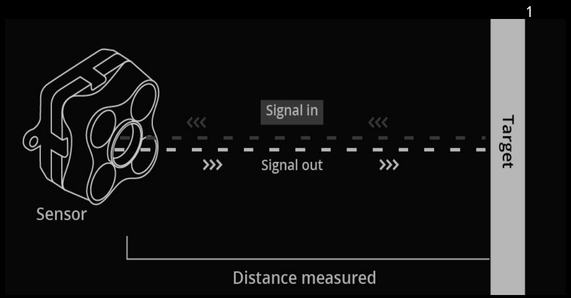
Honey Gupta and Kaushik Mitra
Computational Imaging Lab, IIT Madras, India

IEEE ICIP 2019

Underwater 3D imaging

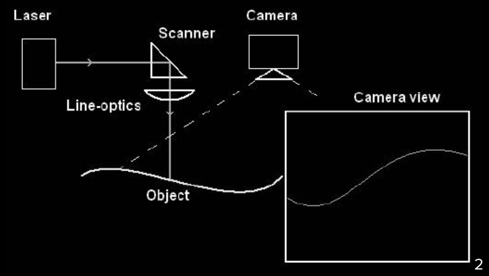
Active depth sensing techniques

Time-of-flight



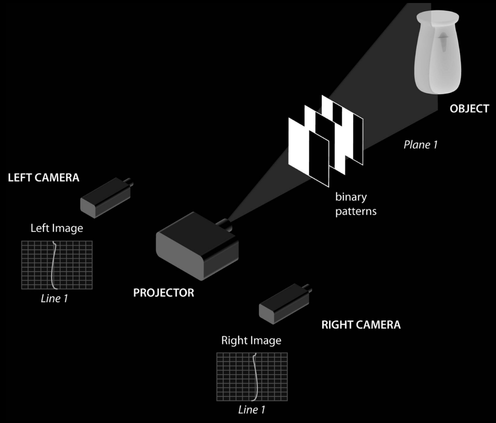
[Maccarone et al. 2015]

LiDAR



[McLeod et al. 2013]

Structured Light



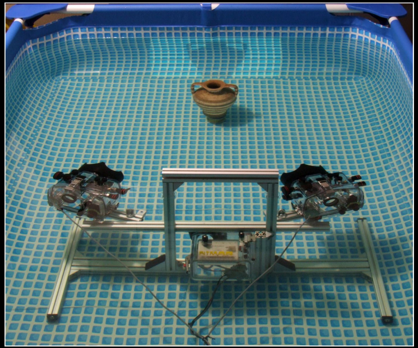
[Bruno et al. 2011]

Image courtesy:
[1] https://www.terabee.com/time-of-flight-principle
[2] A practical underwater 3D-Laserscanner

Underwater 3D imaging

Passive depth sensing techniques

Stereo



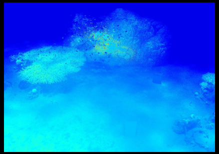
[1]

[Wu et al. (2013), Ferreira et al. (2016)]

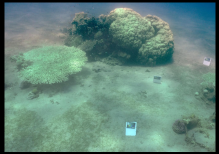
Single image

[2]

Input image



Transmission map



Enhanced image

$$t(p) = e^{-\eta d(p)}$$

Depth map

[Drews et al. (2016), Peng et al. (2015), Emberton et al. (2018), Berman et al. (2018)]

Image courtesy:
[1] Experimentation of structured light and stereo vision for underwater 3D reconstruction
[2] Berman et al. (2018)

Underwater 3D imaging

Passive depth sensing techniques

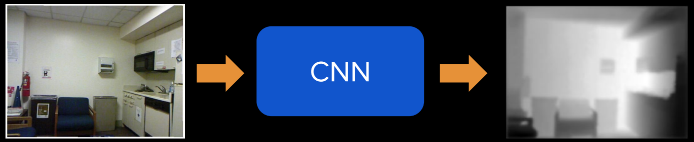
Focus of this paper

Single image

Image courtesy: [1] Berman et al. (2018)

Single image depth estimation

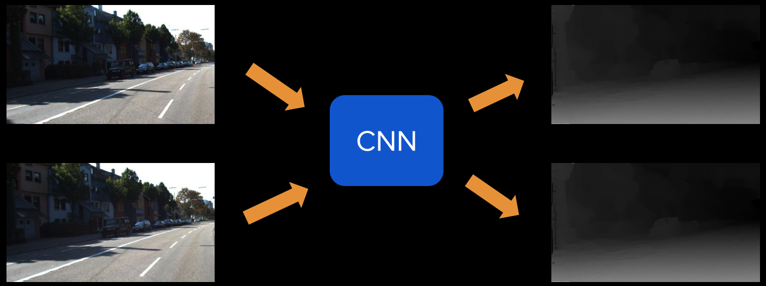
Supervised methods



[Eigen et al. (2014), Kendall et al. (2017), Li et al. (2018)]

Image courtesy: NYU Depth Dataset V2 (2012)

Unsupervised methods

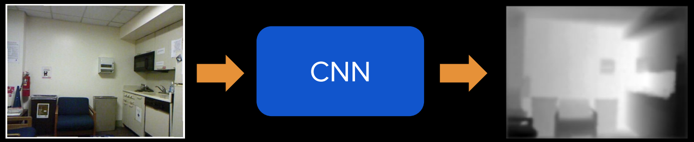


[Garg et al. (2016), Godard et al. (2017), Zhan et al. (2018)]

Image courtesy: KITTI stereo dataset (2015)

Single image depth estimation

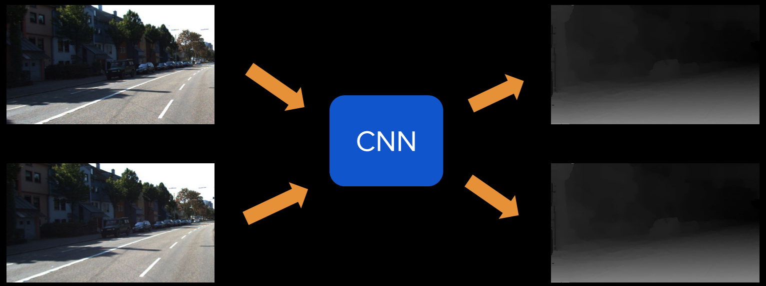
Supervised methods



Needs ground truth depth maps

Image courtesy: NYU Depth Dataset V2 (2012)

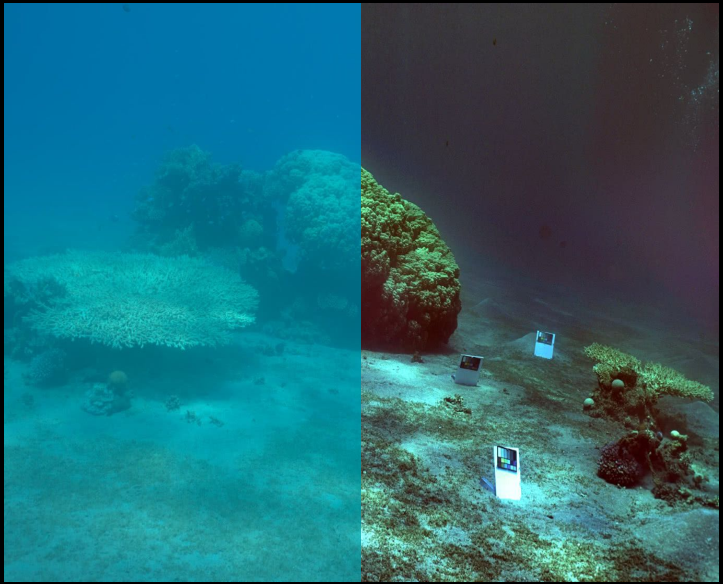
Unsupervised methods



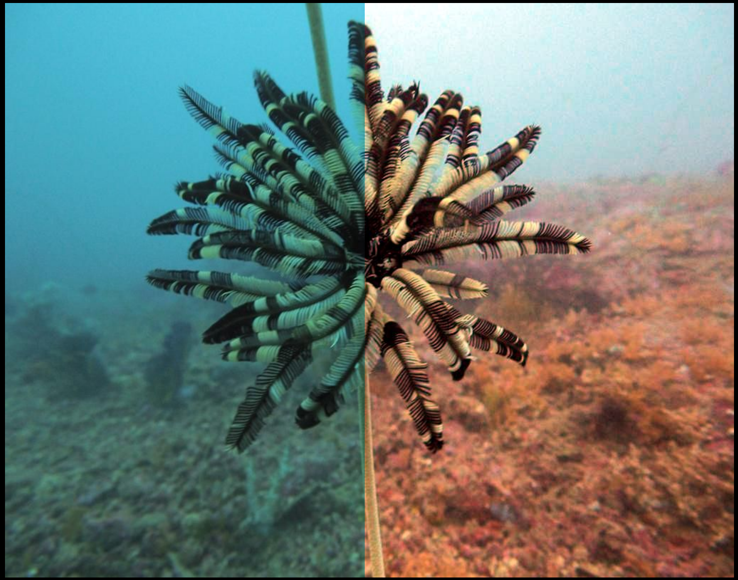
Needs a large stereo dataset

Image courtesy: KITTI stereo dataset (2015)

Challenges



Attenuation / haze



Color distortion

Source: Berman et al. (2018)

Source: http://puiqe.eecs.qmul.ac.uk (2018)

Challenges

- And the effect of these factors vary with
  - scene depth
  - illumination
  - turbidity

Reference | Depth: 5m | Depth: 15m

Source: Chiang et al. (2012)

How to build a robust model without supervision?

Unsupervised single underwater image depth estimation (UW-Net)

- We propose to use unsupervised style-transfer
  - learn mappings between underwater and above-water images
  - propose a network inspired from CycleGAN [Zhu et al. 2017]
- Exploit the haze information in underwater images
  - use haze as a cue for depth

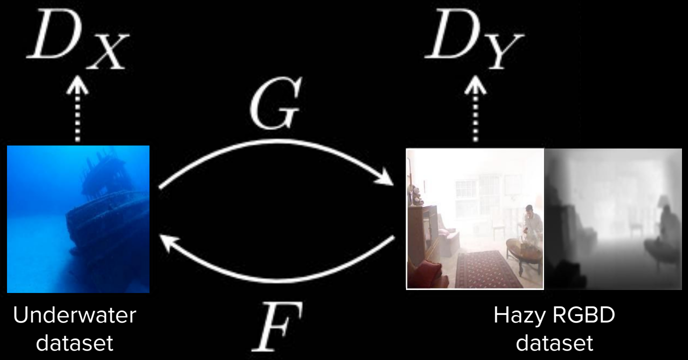


Image courtesy: Zhu et al. (2017), NYU Depth Dataset V2 (2012)

Motivation for our method

$$I(p) = J(p)t(p) + B_\infty(1 - t(p))$$

$$I(p) = J(p)t(p) + A(1 - t(p))$$

$$t(p) = e^{-\eta d(p)}$$

Similar image formation models = similar depth-dependent attenuation

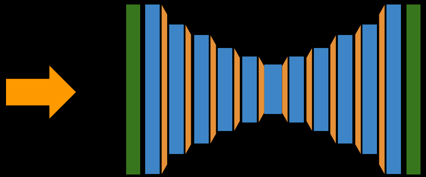
Image courtesy: Berman et al. (2018), https://thefella.com

Unsupervised single underwater image depth estimation (UW-Net)

Real underwater image

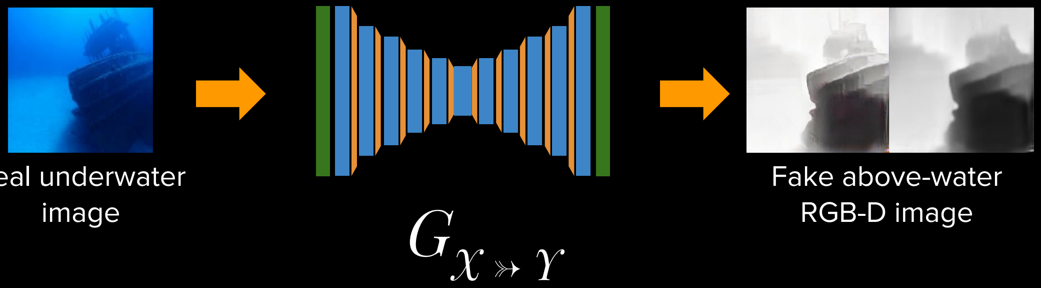
Unsupervised single underwater image depth estimation (UW-Net)

Real underwater image



$G_{\mathcal{X} \twoheadrightarrow \mathcal{Y}}$

Unsupervised single underwater image depth estimation (UW-Net)

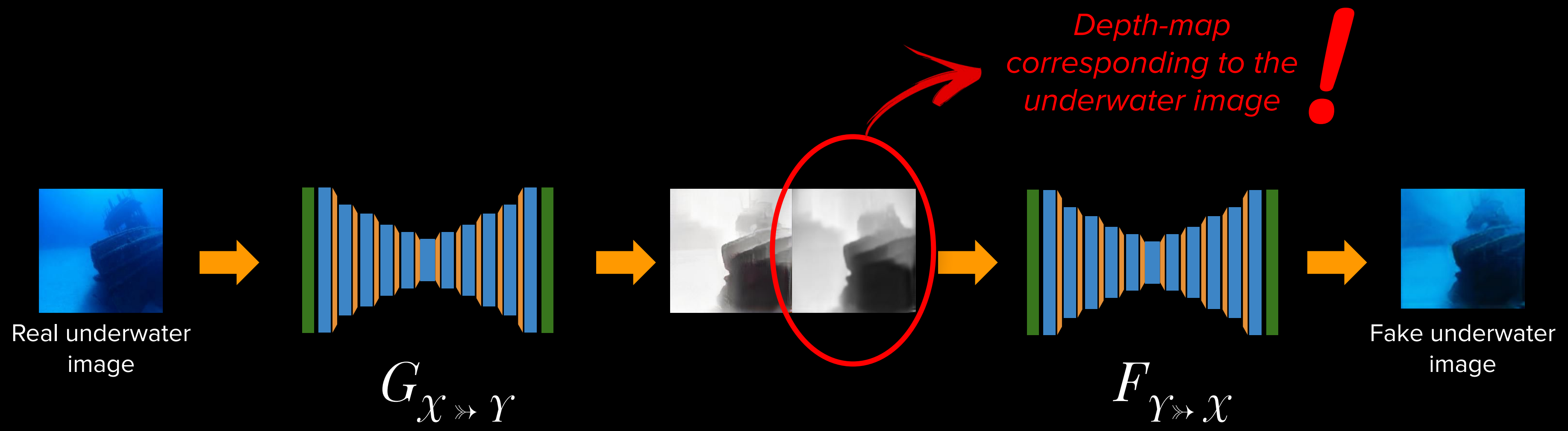


Unsupervised single underwater image depth estimation (UW-Net)

Underwater - cycle

Underwater - cycle

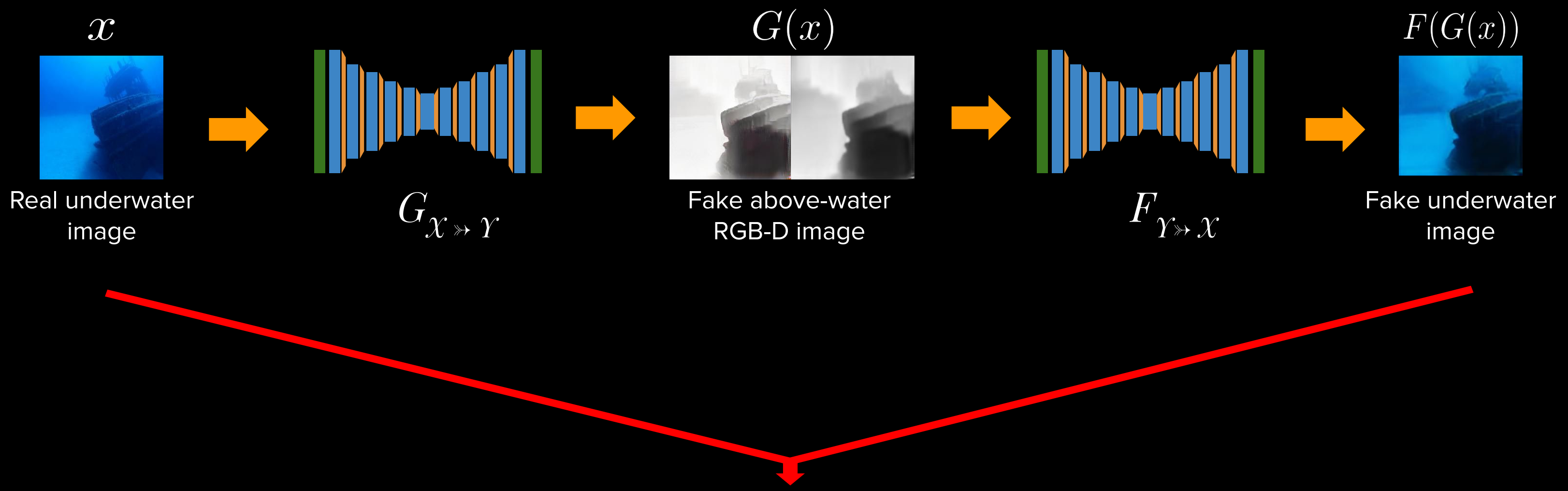
Underwater - cycle



RGBD - cycle


Real above-water RGB-D image
$F_{Y \twoheadrightarrow X}$
Fake underwater image
$G_{X \twoheadrightarrow Y}$
Fake above-water RGB-D image

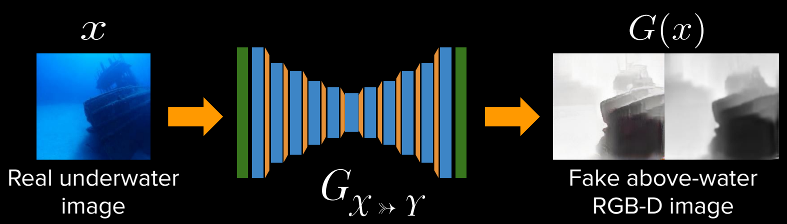
Reconstruction loss



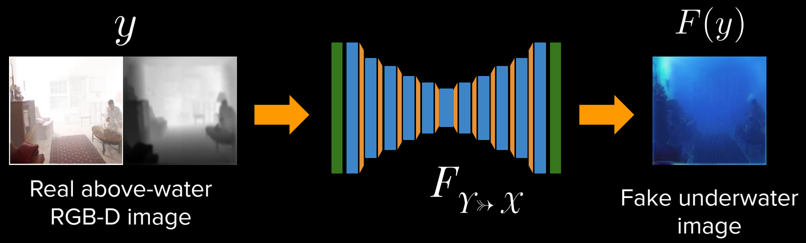
$$\mathcal{L}_{cyc}(G, F : x, y) = ||x - F(G(x))||_1 + ||y - G(F(y))||_1$$

How to avoid learning identity mapping?

Adversarial loss

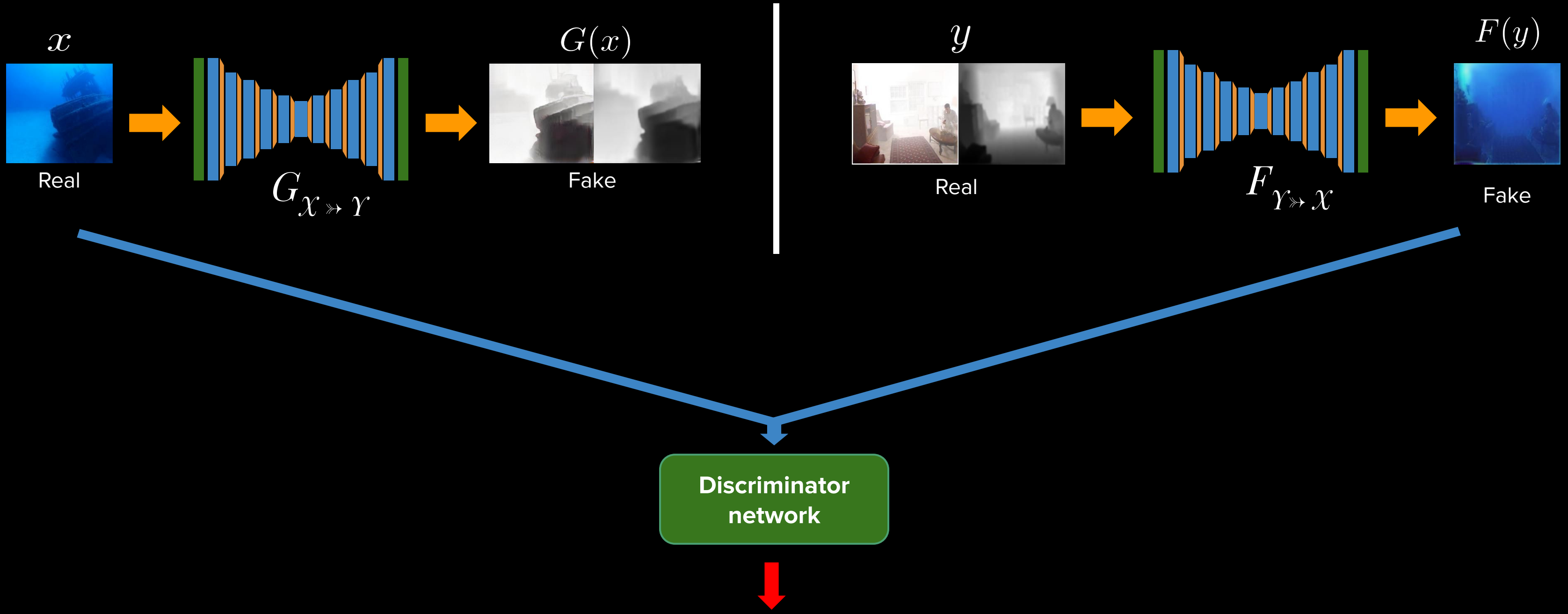


Underwater half-cycle



Above-water half-cycle

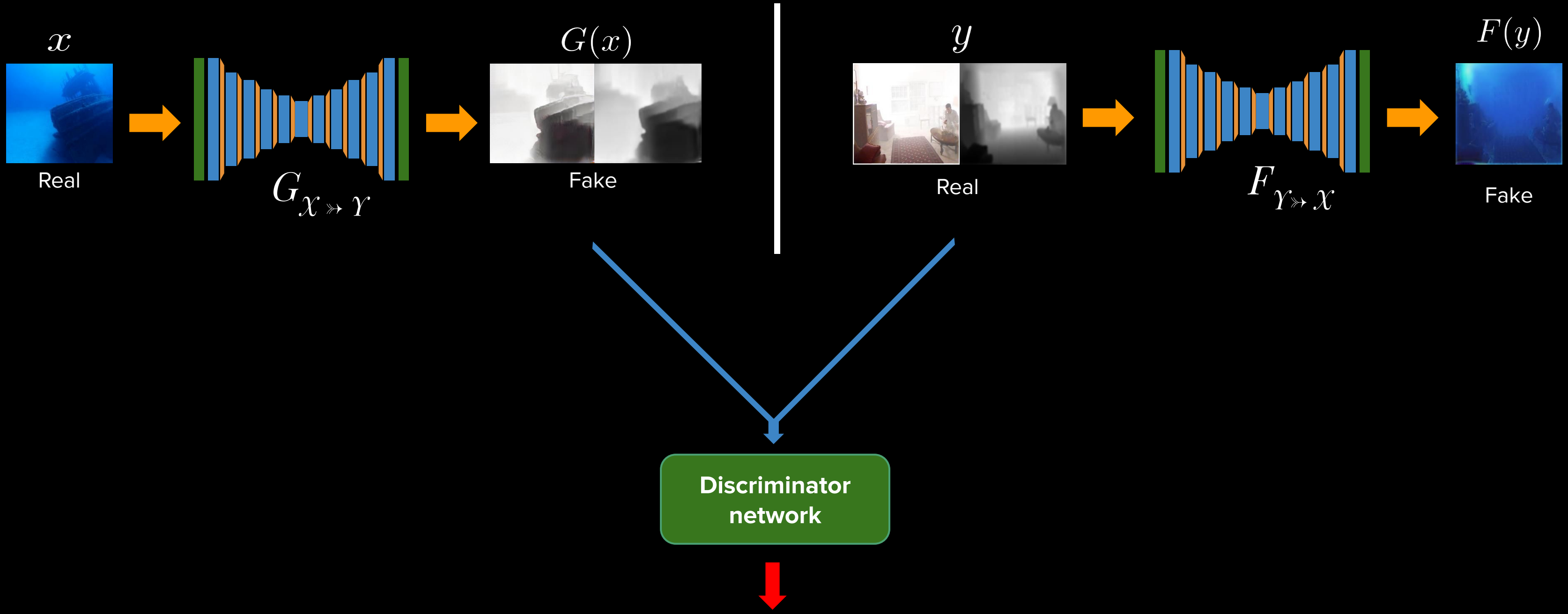
Adversarial loss



$$\mathcal{L}_{gan}(G, D_X, X, Y) = \mathbf{E}_x[(D_X(x) - 1)^2] + \mathbf{E}_y[D_X(F(y))^2]$$

Derived from: Mao et al. (2016)

Adversarial loss

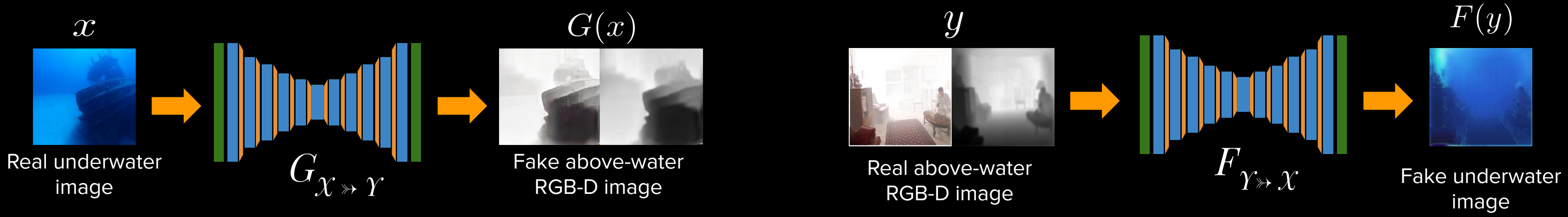


$$\mathcal{L}_{gan}(G, D_Y, X, Y) = \mathbf{E}_y[(D_Y(y) - 1)^2] + \mathbf{E}_x[D_Y(G(x))^2]$$

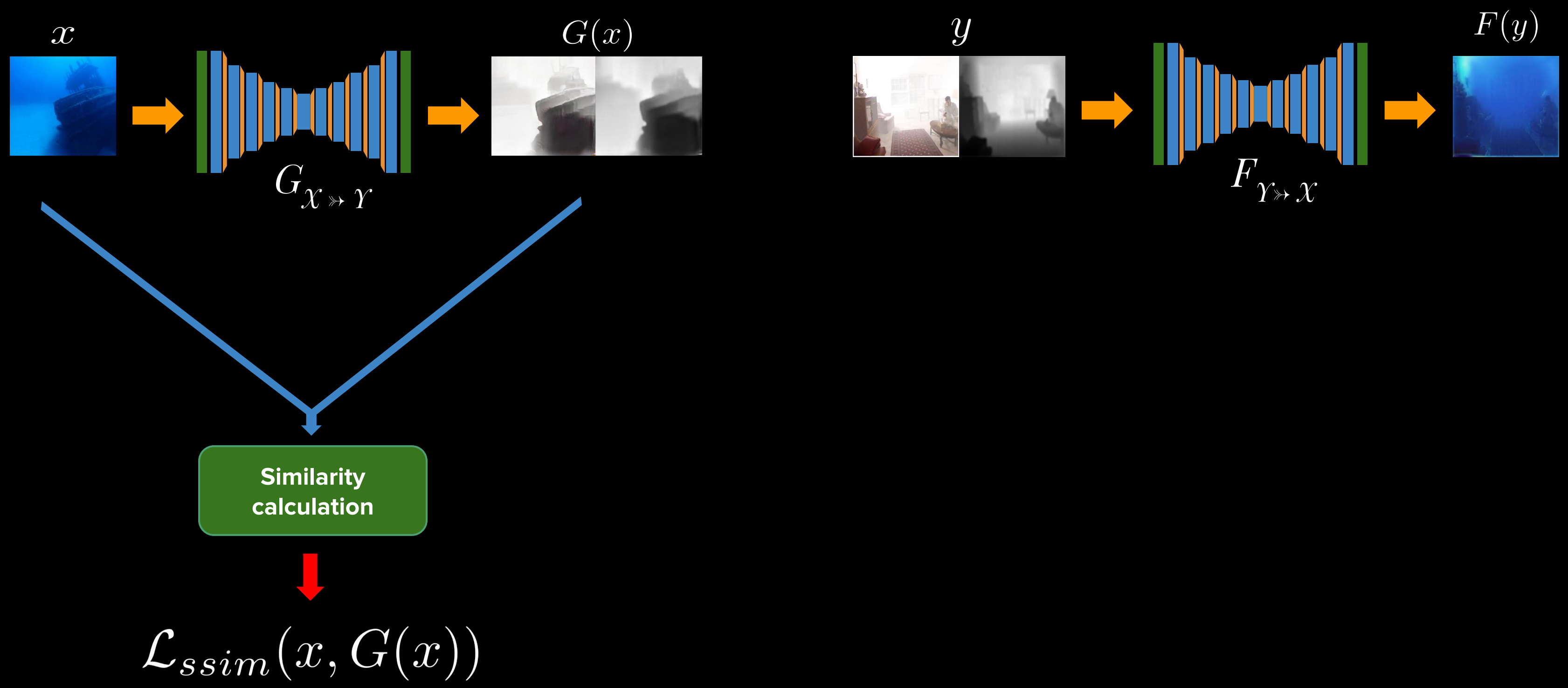
Derived from: Mao et al. (2016)

But is the adversarial loss good enough?

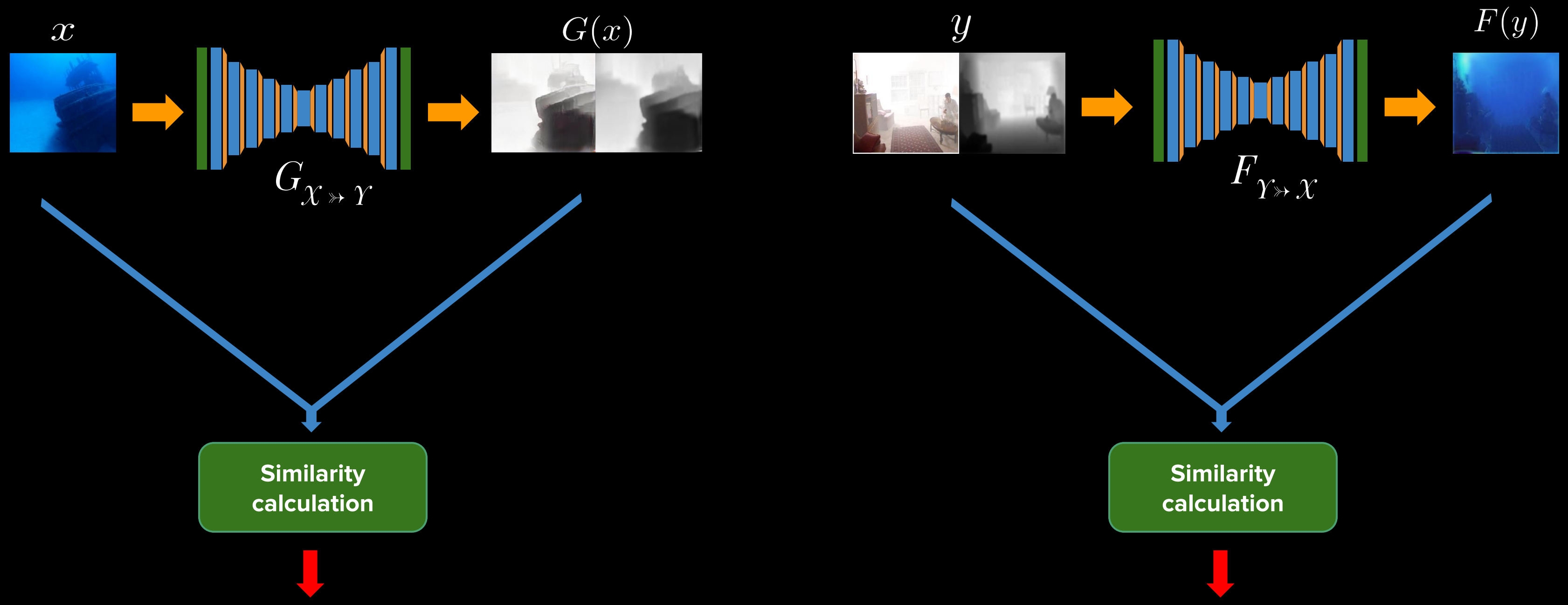
Structural loss



Structural loss



Structural loss



$$\mathcal{L}_{ssim}(x, G(x))$$

$$\mathcal{L}_{ssim}(y, F(y))$$

Overall training loss

$$\mathcal{L}_{total}(G, F, D_X, D_Y) =$$
$$\mathcal{L}_{cyc} + \gamma_{gan}\mathcal{L}_{gan} + \gamma_{ssim}\mathcal{L}_{ssim} + \gamma_{grad}\mathcal{L}_{grad}$$

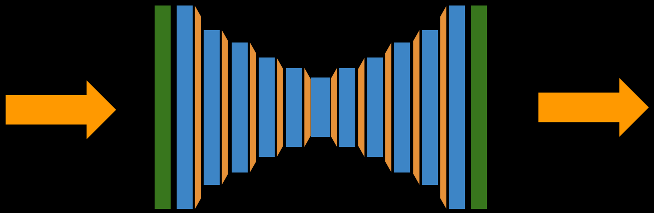
Overall training loss

$$\mathcal{L}_{total}(G, F, D_X, D_Y) =$$
$$\mathcal{L}_{cyc} + \gamma_{gan}\mathcal{L}_{gan} + \gamma_{ssim}\mathcal{L}_{ssim} + \gamma_{grad}\mathcal{L}_{grad}$$

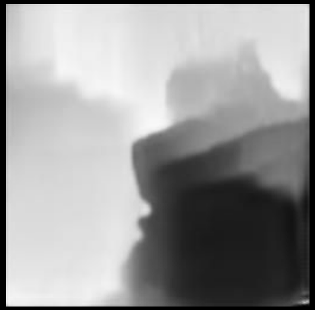
Gradient sparsity prior
for depth-map

Testing

Real underwater image



$G_{\mathcal{X} \twoheadrightarrow \mathcal{Y}}$

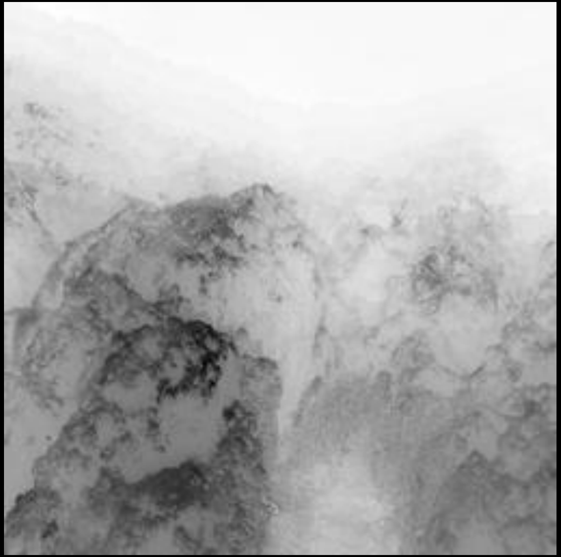


Predicted depth map

Comparison with traditional methods

Pearson correlation coeff. (ρ) /
Scale Invariant - MSE

Input underwater image



0.786 / 0.545

Peng et al. (2015)

0.903 / 0.145

UW-Net (Ours)

Comparison with traditional methods

ρ / SI-MSE

Input underwater image

0.782 / 0.411

Drews et al. (2016)

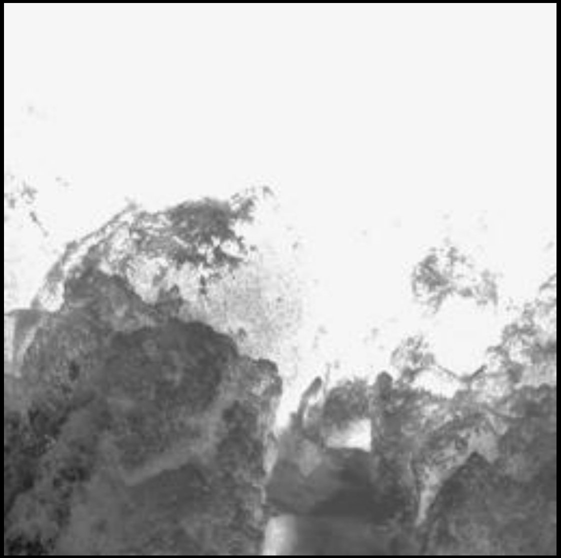
0.903 / 0.145

UW-Net (Ours)

Comparison with traditional methods

ρ / SI-MSE

Input underwater image



0.753 / 0.543

Berman et al. (2018)

0.903 / 0.145

UW-Net (Ours)

Comparison with deep-learning based method

ρ / SI-MSE

Input underwater image

0.809 / 0.419

Godard et al. (2017)
(pre-trained)

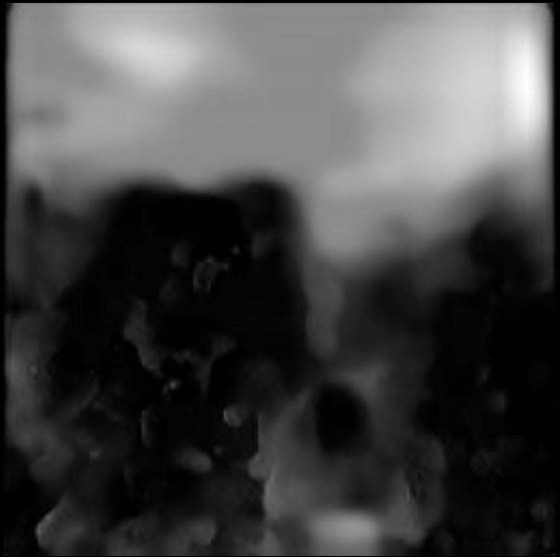
0.903 / 0.145

UW-Net (Ours)

Comparison with deep-learning based method

ρ / SI-MSE

Input underwater image



0.834 / 0.617

Godard et al. (2017)
(fine-tuned on underwater dataset)

0.903 / 0.145

UW-Net (Ours)

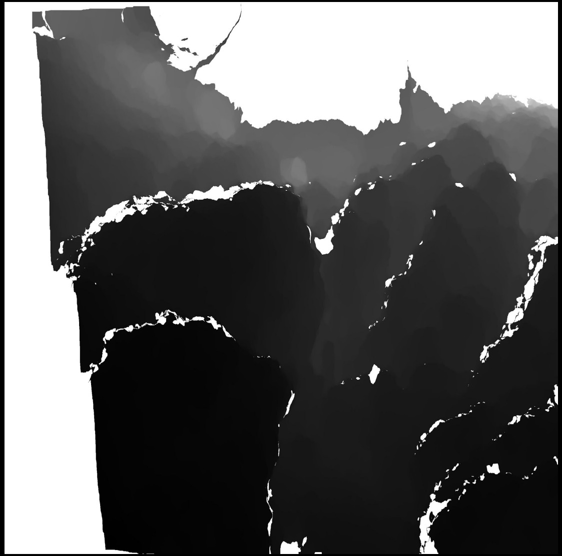
Comparison with ground-truth

ρ / SI-MSE

Input underwater image

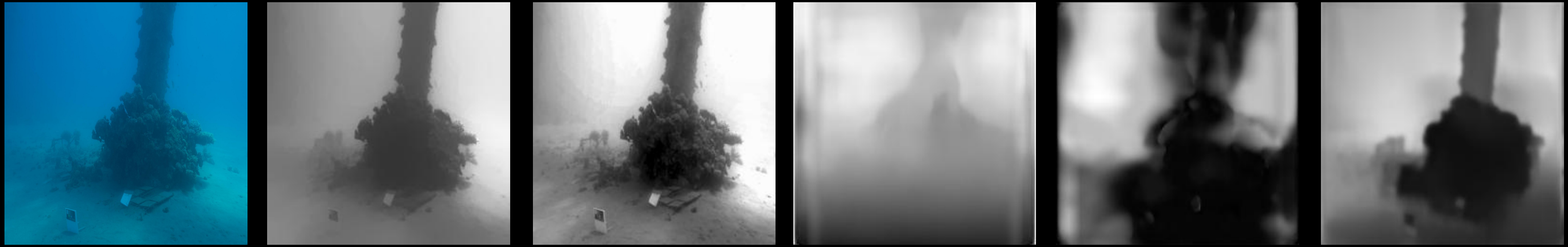
0.903 / 0.145

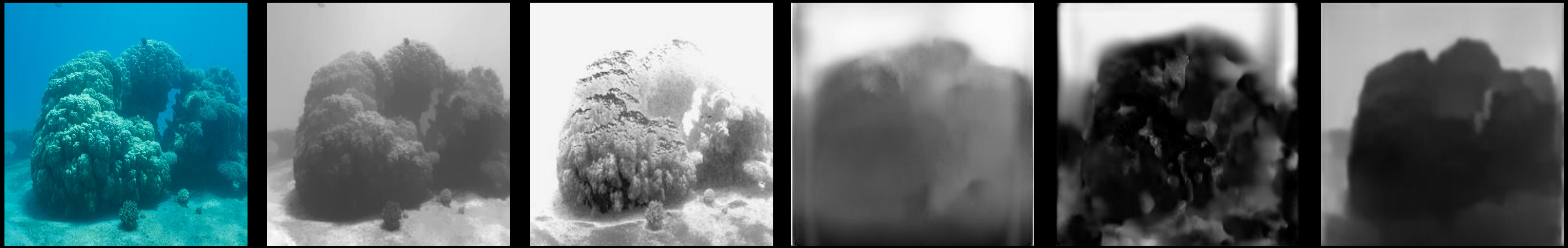
UW-Net (Ours)



Ground-truth

Comparison

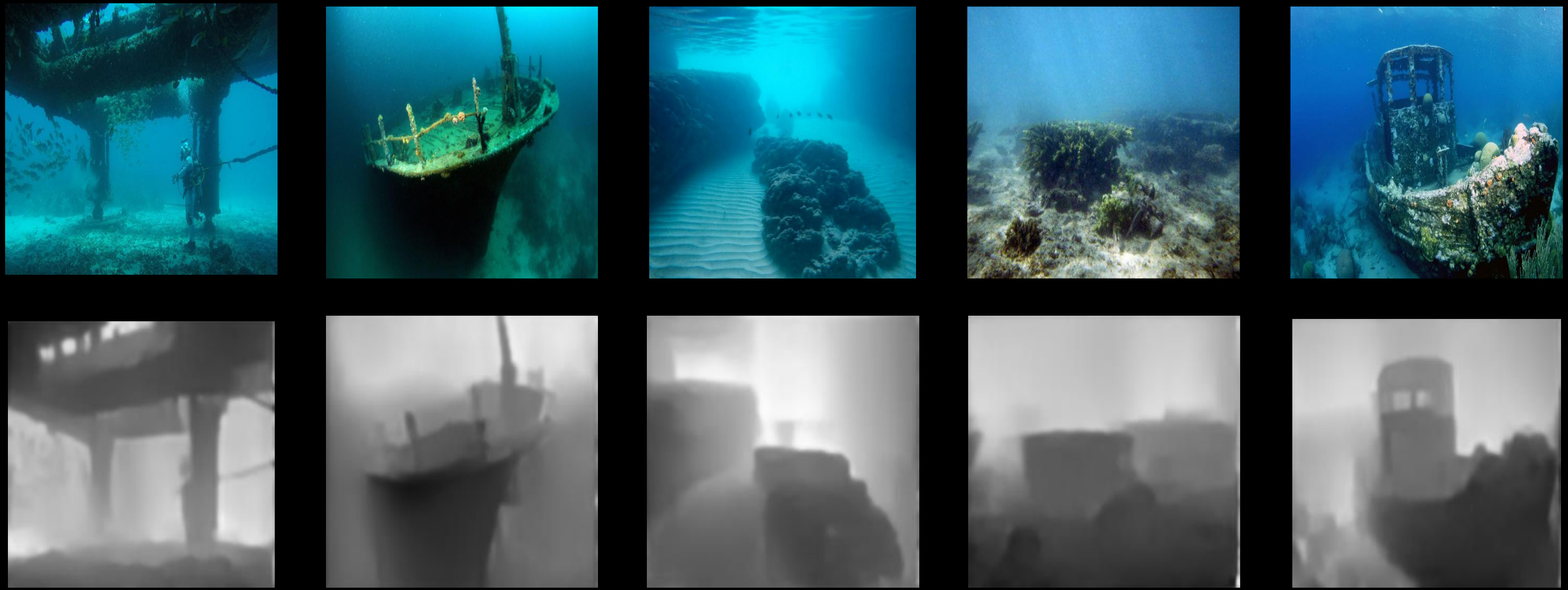




Input underwater image	UDCP	Berman et al.	Godard et al. without fine-tuning	Godard et al. with fine-tuning	UW-Net(Ours)
ρ / SI-MSE	0.651 / 0.513	0.650/0.595	0.716/0.406	**0.738/0.391**	0.791/0.340
ρ / SI-MSE	0.559 / 0.344	0.737/0.326	**0.759/0.206**	0.594/0.552	0.805/0.188

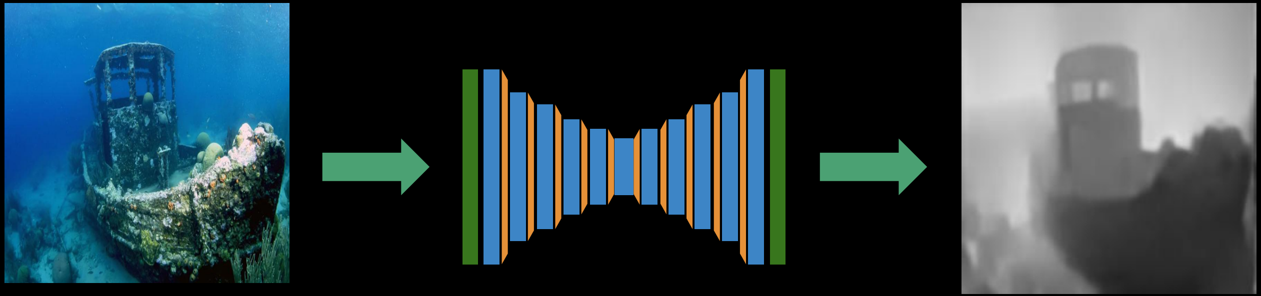
UW-Net on some underwater images

UW-Net on some underwater images



Conclusion

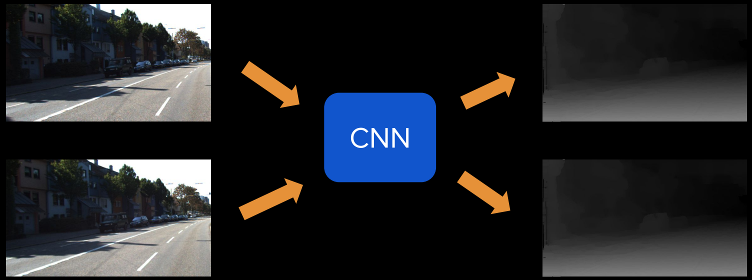
- We tackled the ill-posed problem of depth estimation from a single underwater image



Conclusion

- We tackled the ill-posed problem of depth estimation from a single underwater image
- Our method
  - is unsupervised and does not require ground-truth depth maps
  - doesn't require a large stereo dataset

Single image supervised methods



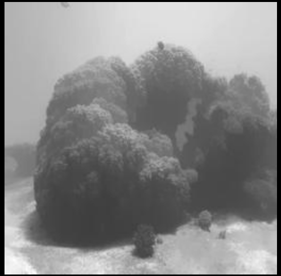
Single image unsupervised methods

Conclusion

- We tackled the ill-posed problem of depth estimation from a single underwater image
- Our method
  - is unsupervised and does not require ground-truth depth maps
  - doesn’t require a large stereo dataset
  - produces better depth maps for diverse underwater images

Input underwater image

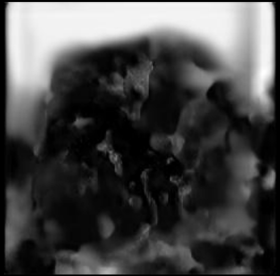
UDCP



Berman et al.

Godard et al. without fine-tuning

Godard et al. with fine-tuning



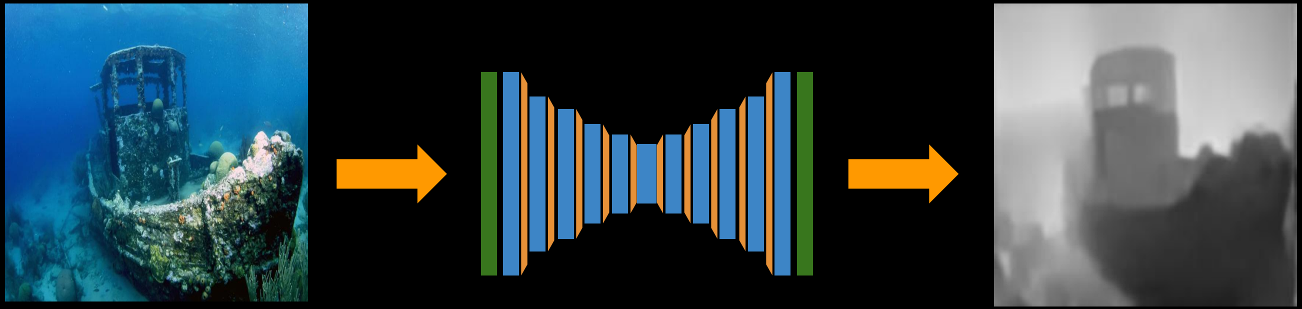
Ours

Conclusion

- We tackled the ill-posed problem of depth estimation from a single underwater image
- Our method
  - produces better depth maps for diverse underwater images
  - is unsupervised and does not require ground-truth depth maps
  - doesn’t require a large stereo dataset

Future directions:

The indirect learning method proposed can be used to solve a variety of ill-posed problems
- where obtaining ground-truth is very expensive
- capturing a large dataset is difficult
- Similarity between two image domains can be exploited



Project web-page
http://www.ee.iitm.ac.in/comp_photolab/project-underwater.html

Code available at https://github.com/honeygupta/UW-Net

[References]

[Maccarone et al.] Maccarone, A., McCarthy, A., Ren, X., Warburton, R.E., Wallace, A.M., Moffat, J., Petillot, Y. and Buller, G.S., 2015. Underwater depth imaging using time-correlated single-photon counting. Optics express, 23(26), pp.33911-33926.

[Bruno et al.] Bruno, F., Bianco, G., Muzzupappa, M., Barone, S. and Razionale, A.V., 2011. Experimentation of structured light and stereo vision for underwater 3D reconstruction. ISPRS Journal of Photogrammetry and Remote Sensing, 66(4), pp.508-518.

[Wu et al.] Wu, Y., Nian, R. and He, B., 2013, September. 3D reconstruction model of underwater environment in stereo vision system. In Oceans-San Diego, 2013 (pp. 1-4). IEEE.

[Ferreira et al.] Sánchez-Ferreira, C., Mori, J.Y., Farias, M.C. and Llanos, C.H., 2016. A real-time stereo vision system for distance measurement and underwater image restoration. Journal of the Brazilian Society of Mechanical Sciences and Engineering, 38(7), pp.2039-2049.

[McLeod et al.] McLeod, D., Jacobson, J., Hardy, M. and Embry, C., 2013, September. Autonomous inspection using an underwater 3D LiDAR. In Oceans-San Diego, 2013 (pp. 1-8). IEEE.

[Ancuti et al.] Ancuti, C.O., Ancuti, C., De Vleeschouwer, C., Neumann, L. and Garcia, R., 2017, September. Color transfer for underwater dehazing and depth estimation. In Image Processing (ICIP), 2017 IEEE International Conference on (pp. 695-699). IEEE.

[Drews et al.] Drews, P.L., Nascimento, E.R., Botelho, S.S. and Campos, M.F.M., 2016. Underwater depth estimation and image restoration based on single images. IEEE computer graphics and applications, 36(2), pp.24-35.

[Eigen et al.] Eigen, D., Puhrsch, C. and Fergus, R., 2014. Depth map prediction from a single image using a multi-scale deep network. In Advances in neural information processing systems (pp. 2366-2374).

[Kendall et al.] Kendall, A., Martirosyan, H., Dasgupta, S., Henry, P., Kennedy, R., Bachrach, A. and Bry, A., 2017. End-to-end learning of geometry and context for deep stereo regression. CoRR, vol. abs/1703.04309.

[Li et al.] Li, Z. and Snavely, N., 2018. Megadepth: Learning single-view depth prediction from internet photos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2041-2050).

[Garg et al.] Garg, R., BG, V.K., Carneiro, G. and Reid, I., 2016, October. Unsupervised cnn for single view depth estimation: Geometry to the rescue. In European Conference on Computer Vision (pp. 740-756). Springer, Cham.

[Godard et al.] Godard, C., Mac Aodha, O. and Brostow, G.J., 2017, July. Unsupervised monocular depth estimation with left-right consistency. In CVPR (Vol. 2, No. 6, p. 7).

[Zhan et al.] Zhan, H., Garg, R., Weerasekera, C.S., Li, K., Agarwal, H. and Reid, I., 2018, March. Unsupervised Learning of Monocular Depth Estimation and Visual Odometry with Deep Feature Reconstruction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 340-349).

[Zhu et al.] Zhu, J.Y., Park, T., Isola, P. and Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv preprint.

[Mao et al.] Mao, X., Li, Q., Xie, H., Lau, R.Y. and Wang, Z., 2016. Multi-class generative adversarial networks with the L2 loss function. arXiv preprint arXiv:1611.04076, 5.

[KITTI stereo dataset] Menze, M., Heipke, C. and Geiger, A., 2015. JOINT 3D ESTIMATION OF VEHICLES AND SCENE FLOW. ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences, 2.

[NYU Depth Dataset v2] Silberman, N., Hoiem, D., Kohli, P. and Fergus, R., 2012, October. Indoor segmentation and support inference from rgbd images. In European Conference on Computer Vision (pp. 746-760). Springer, Berlin, Heidelberg.

Haze as a cue for Depth

Depth-map obtained with normal color image as aerial input

Input aerial image

Input aerial depth map

Input underwater image

Estimated underwater depth map

Haze as a cue for Depth

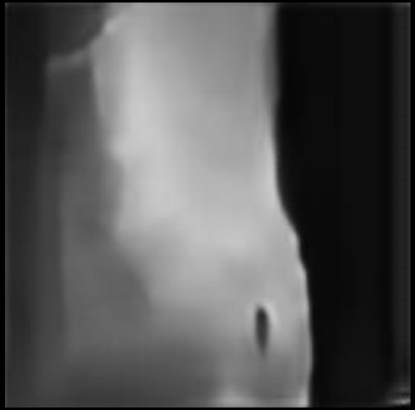
Depth-map obtained with hazy color image as aerial input

Input aerial image

Input aerial depth map

Input underwater image

Estimated underwater depth map



Training details

- Above-water dataset:
  - D-Hazy dataset
  - 1449 synthetic hazy images created from NYU Depth Dataset
  - Performed bilateral filtering on the depth maps for training.
- Underwater dataset
  - 1343 images collected from internet

  Downsampled to 256x256 and randomly cropped 128x128 for training

- Comparison with:
  - Non data-driven: Dark channel prior[6], UDCP [7]
  - Data-driven: Pretrained Unsupervised monocular depth estimation with left-right consistency [12], fine-tuned[12] on underwater stereo dataset (CADDY)