Fast Underwater Image Enhancement for Improved Visual Perception

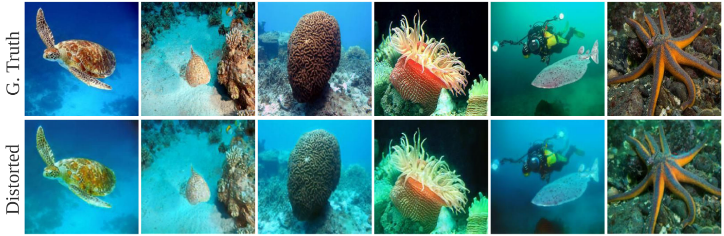
Chinchuthakun Worameth

January 17, 2022

Yamashita Laboratory
Department of Transdisciplinary Science and Engineering
School of Environment and Society
Tokyo Institute of Technology

Challenges of underwater imaging

- Light propagation under water is different than in the atmosphere
  - Basically **a different set of non-linear transformations** from real-world scenes to image (color) coordinate system
  - For example, underwater images usually have **higher green or blue hues** because red wavelengths are absorbed in deep water region



Paired instances from EUVP dataset [1]

Challenges of underwater imaging (2)

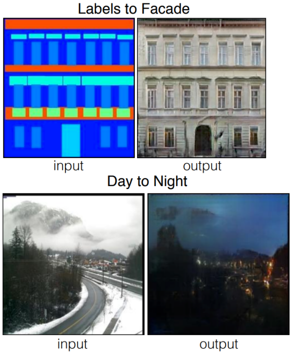
- We can perform **dehazing** and **color correction** using **physical models** with additional parameters, e.g. depth or optical water-quality measures
  - Not always available in robotic applications
  - **Multimodal models are computational expensive for real-time deployment**
- Several models based on **CNN** and **GAN** are proposed
  - Unfortunately, We could not achieve a good performance with deep learning
  - **Underwater data are expensive and difficult to acquire!**

Research overview

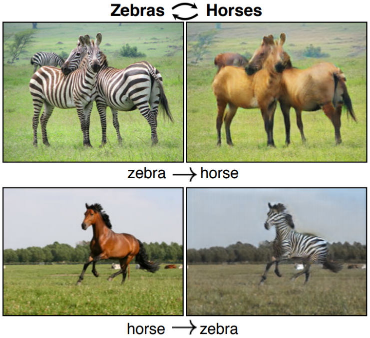
- Published a **EUVP dataset (Enhancement of Underwater Visual Perception)**
  - 12K paired and 8K unpaired underwater of poor and good quality
- Presented a deep learning model for fast underwater image enhancement
  - Framed the problem as **image-to-image translation**
  - Proposed a **conditional GAN architecture** to solve it
  - Evaluated on the proposed dataset

Background: Image-to-image translation

- Transform **style** of images in one domain to another domain
- Some famous examples — **Pix2Pix** [2] and **CycleGan** [3]



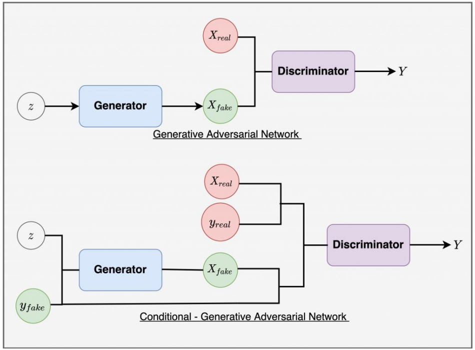
Examples from Pix2Pix [2]



Examples from CycleGAN [3]

Background: Recap of GAN

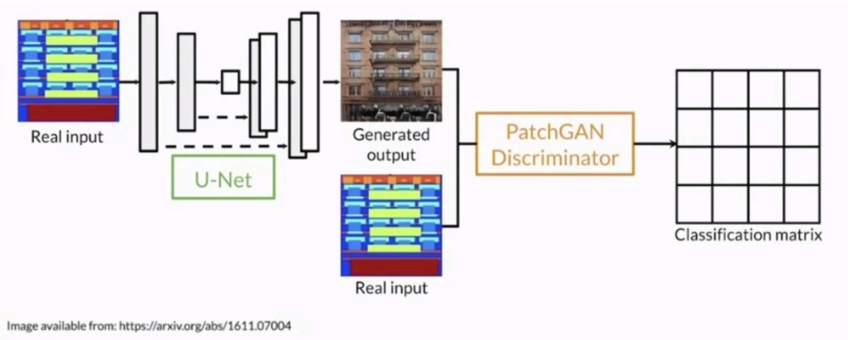
- A two-player min-max game between **generator** and **discriminator**
- **Conditional GANs** impose constrains such that generators produce samples belonged to a specified class



[1] Image available from `https://learnopencv.com/conditional-gan-cgan-in-pytorch-and-tensorflow/`

Background: Pix2Pix in a nutshell

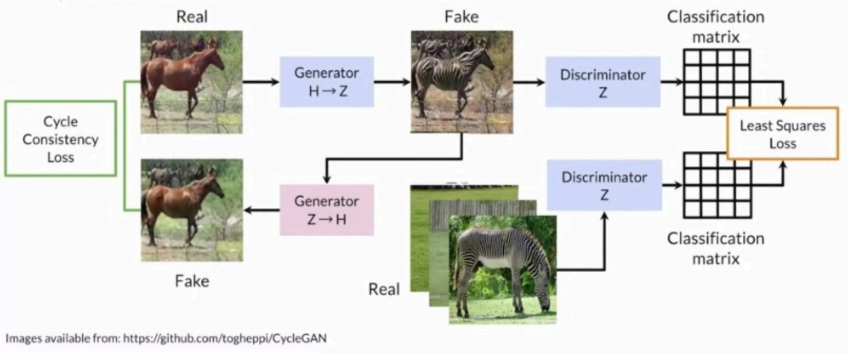
- Generator follows **U-Net architecture**
- Discriminator uses **PatchGAN**, predicting if patches of images are real or fake



[1] Image available from *https://www.haikutechcenter.com/2020/10/pix2pix-gan-architecture-for-image-to.html*

Background: CycleGAN in a nutshell

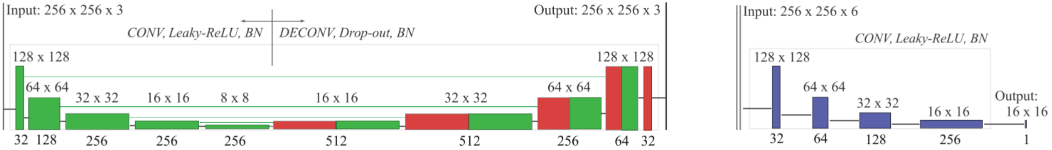
- Although Pix2Pix requires paired training data, we can solve this problem with **cycle-consistency loss**



[1] Image available from *https://www.haikutechcenter.com/2020/11/cyclegan-gan-architecture-for-learning.html*

Methodology: Network Architecture

- Following previous works, generator and discriminator are **mini U-Net** and **PatchGAN**, respectively.



Architecture of FUnIE-GAN [1]

Methodology: Loss Function

- **Paired image translation**
  Let X, Y be the source and target domain. Consider an input Z with a generator G and discriminator D:
  - **Adversarial Loss**

$$L_{cGAN}(G, D) = \mathbb{E}_{X,Y}[\log D(Y)] + \mathbb{E}_{X,Y}[\log(1 - D(X, G(X, Z)))]$$

  - **Global Similarity**

$$L_1(G) = \mathbb{E}_{X,Y,Z}[\|Y - G(X, Z)\|_1]$$

  - **Image content**

$$L_{con}(G) = \mathbb{E}_{X,Y,Z}[\|\Phi(Y) - \Phi(G(X, Z))\|_2]$$

    where Φ denotes the image content function, i.e. the high-level features extracted by a certain layer of a pre-trained **VGG-19 architecture**

Methodology: Loss Function (2)

- **Paired image translation (con't)**

$$G^* = \arg\min_G \max_D L_{cGAN}(G, D) + \lambda_1 L_1(G) + \lambda_c L_{con}(G)$$

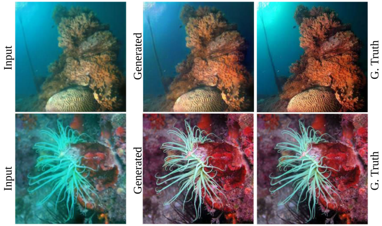
- **Unpaired image translation**
    - Since we do not have access to the ground-truth image, we only use adversarial loss and cycle-consistency loss
    - Let $G_F : \{X, Z\} \rightarrow Y$ and $G_R : \{Y, Z\} \rightarrow X$:

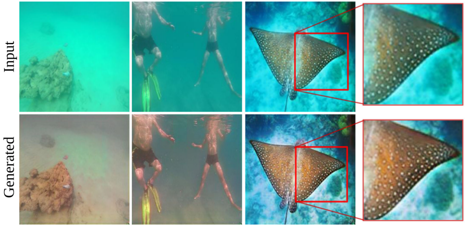
$$G_F^*, G_R^* = \arg\min_{G_F, G_R} \max_{D_Y, D_X} L_{cGAN}(G_F, D_Y) + L_{cGAN}(G_R, D_X) + \lambda_{cyc} L_{cyc}(G_F, G_R)$$

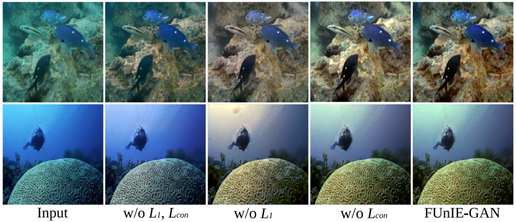
where

$$L_{cyc}(G_F, G_R) = \mathbb{E}_{X,Y,Z}[\|X - G_R(G_F(X, Z))\|_1] + \mathbb{E}_{X,Y,Z}[\|Y - G_F(G_R(X, Z))\|_1]$$

Experiments: Qualitative Analysis







Experiments: Quantitative Analysis

- Three metrics are considered:
  - **Peak Signal-to-Noise Ratio (PSNR)**: measure reconstruction quality

$$PSNR(\mathbf{x}, \mathbf{y}) = 10 \log_{10}\left(255^2 / MSE(\mathbf{x}, \mathbf{y})\right)$$

  - **Structural Similarity (SSIM)**: measure luminance, constrast, and structure

$$SSIM(\mathbf{x}, \mathbf{y}) = \left(\frac{2\mu_{\mathbf{x}}\mu_{\mathbf{y}} + c_1}{\mu_{\mathbf{x}}^2 + \mu_{\mathbf{y}}^2 + c_1}\right)\left(\frac{2\sigma_{\mathbf{xy}} + c_2}{\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{y}}^2 + c_2}\right)$$

  - **Underwater Image Quality Measure (UIQM)** [4]
    - Quantifies underwater image colorfulness, sharpness, and contrast according to human visual perceptions.

Experiments: Qualitative Analysis and User Study

- **Comparison with other state-of-the-arts**

Model	$PSNR(G(\mathbf{x}), \mathbf{y})$ Input: 17.27 ± 2.88	$SSIM(G(\mathbf{x}), \mathbf{y})$ Input: 0.62 ± 0.075
Uw-HL	18.85 ± 1.76	0.7722 ± 0.066
Mband-EN	12.11 ± 2.55	0.4565 ± 0.097
Res-WGAN	16.46 ± 1.80	0.5762 ± 0.014
Res-GAN	14.75 ± 2.22	0.4685 ± 0.122
LS-GAN	17.83 ± 2.88	0.6725 ± 0.062
Pix2Pix	20.27 ± 2.66	0.7081 ± 0.069
UGAN-P	19.59 ± 2.54	0.6685 ± 0.075
CycleGAN	17.14 ± 2.65	0.6400 ± 0.080
FUnIE-GAN-UP	21.36 ± 2.17	0.8164 ± 0.046
FUnIE-GAN	21.92 ± 1.07	0.8876 ± 0.068

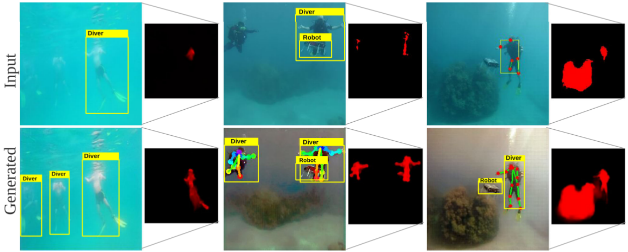
Model	**Paired data** Input: 2.20 ± 0.69 G. Truth: 2.91 ± 0.65	**Unpaired data** Input: 2.29 ± 0.62 G. Truth: N/A
Uw-HL	2.62 ± 0.35	2.75 ± 0.32
Mband-EN	2.28 ± 0.87	2.34 ± 0.45
Res-WGAN	2.55 ± 0.64	2.46 ± 0.67
Res-GAN	2.62 ± 0.89	2.28 ± 0.34
LS-GAN	2.37 ± 0.78	2.59 ± 0.52
Pix2Pix	2.65 ± 0.55	2.76 ± 0.39
UGAN-P	2.72 ± 0.75	2.77 ± 0.34
CycleGAN	2.44 ± 0.71	2.62 ± 0.67
FUnIE-GAN-UP	2.56 ± 0.63	2.81 ± 0.65
FUnIE-GAN	2.78 ± 0.43	2.98 ± 0.51

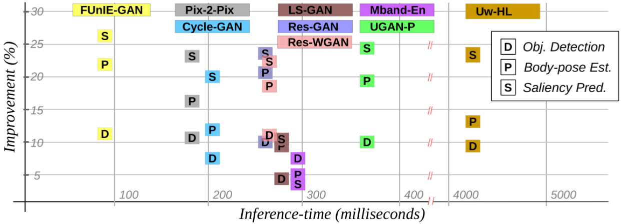
- **Human's preference on generated images**

Model	Rank-1 (%)	Rank-2 (%)	Rank-3 (%)
FUnIE-GAN	24.50	68.50	88.60
FUnIE-GAN-UP	18.67	48.25	76.18
UGAN-P	21.25	65.75	80.50
Pix2Pix	11.88	45.15	72.45

Experiments: Implication

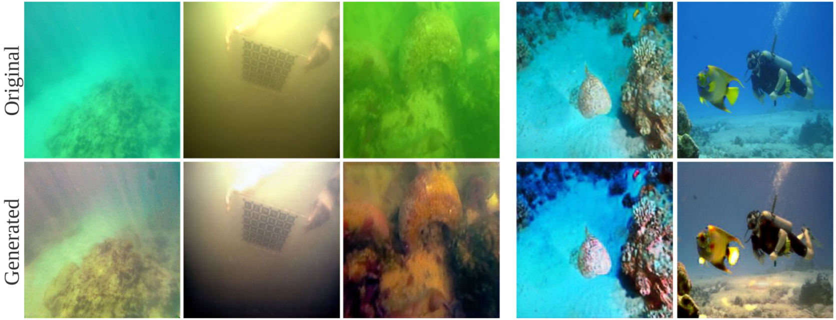
- Improve performance in downstream tasks





Limitations

- Not very effective in severely degraded and texture-less images
- Require careful hyperparameter tuning in unpaired image-to-image translation; otherwise, the discriminator will become too good too early



Issues with extremely low contrast and lack of textures *Inconsistent coloring by unstable training*

References i

[1] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3227–3234, 2020.

[2] P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," *CoRR*, vol. abs/1611.07004, 2016. [Online]. Available: http://arxiv.org/abs/1611.07004

[3] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *CoRR*, vol. abs/1703.10593, 2017. [Online]. Available: http://arxiv.org/abs/1703.10593

[4] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE Journal of Oceanic Engineering*, vol. 41, no. 3, pp. 541–551, 2016.