AFRL-AFOSR-JP-TR-2022-0061

Exploring deep learning based robot perception techniques for navigating outdoor terrains

HANSEOK KO
Korea University Research and Business Foundation
1-2 Anam-dong 5-ga Seongbuk-gu
Seoul, , 136713
KR

08/13/2022
Final Technical Report

Air Force Research Laboratory
Air Force Office of Scientific Research
Asian Office of Aerospace Research and Development
Unit 45002, APO AP 96338-5002

REPORT DOCUMENTATION PAGE

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE 20220813	**2. REPORT TYPE** Final	**3. DATES COVERED**	
		START DATE 20190430	**END DATE** 20220429

4. TITLE AND SUBTITLE
Exploring deep learning based robot perception techniques for navigating outdoor terrains

5a. CONTRACT NUMBER	**5b. GRANT NUMBER** FA2386-19-1-4001	**5c. PROGRAM ELEMENT NUMBER**
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Hanseok Ko

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Korea University Research and Business Foundation 1-2 Anam-dong 5-ga Seongbuk-gu Seoul 136713 KR	**8. PERFORMING ORGANIZATION REPORT NUMBER**

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AOARD UNIT 45002 APO AP 96338-5002	**10. SPONSOR/MONITOR'S ACRONYM(S)** AFRL/AFOSR IOA	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)** AFRL-AFOSR-JP-TR-2022-0061

12. DISTRIBUTION/AVAILABILITY STATEMENT
A Distribution Unlimited: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT

When autonomously navigating to its objectives, a ground robot encounters formidable challenges detecting and recognizing its surroundings and objects. From its sensory input, the robot's AI has to semantically segment the scenes such as terrain, vegetation, human-made structures, debris, water streams, etc. The on-board perception then has to assess intelligently and determine what parts of the scene the robot can traverse safely to the objective. The project goal is to develop a novel method of vision-based perception for assessing the navigability of terrains that an autonomous ground vehicle may encounter traversing in natural or structured environments. With the great success brought by the advances in deep learning, computer vision sometimes exceeds human-level performance in object recognition tasks. These algorithms, however, require a large size of class examples to perform accurately.

Although visual data are abundant, images relevant to ground navigation, especially those with class labels are scarce. Therefore, there is a need for a computer vision algorithm capable of high performance with small training sets and capable of recognizing novel objects. We propose to investigate the GAN-based data augmentation approach and the efficient scene understanding approach to tackle the data scarcity issue of the perception problem related to autonomous robotic maneuvers in previously unseen environments. Expected relevant robot maneuvering environments and scenes are typically unusual, and the data for the current paradigms of deep learning is either scarce or non-existent. Hence, it is expected that the GAN-based data augmentation approaches provide solutions to developing terrestrial robotic vehicles capable of perceiving and understanding novel environments.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT** SAR	**18. NUMBER OF PAGES** 18
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U		

19a. NAME OF RESPONSIBLE PERSON TONY KIM	**19b. PHONE NUMBER** *(Include area code)* 315-227-7008

Section 2: Technical Report

Accomplishments

1. Research Objectives:

We proposed to investigate the data augmentation approach and the few-shot learning approach to tackle the data scarcity issue of the perception problem related to autonomous robotic maneuvers in previously unseen environments. Ground robot's anticipated traversing environments and scenes are typically unusual, and data for the current paradigm of deep learning is either scarce or non-existent. The approaches developed are expected to provide the solutions to developing terrestrial robotic vehicles capable of perceiving and understanding novel environments. The research objectives accomplished are delineated as follows.

- **Objective 1:** Data collection and labeling for natural and unstructured environment to be used as ground work for robot perception platform (e.g. semantic segmentation).

- **Objective 2:** Investigate the deep learning-based image data augmentation methods to address the data scarcity problem in machine learning for perception.

- **Objective 3:** Investigate few-shot learning approaches to that small dataset can be used for training inferencing models.

- **Objective 4:** Investigate the deep learning−based semantic segmentation methods for enhanced scene perception.

2. Technical Approaches:

2.1. Objective 1: Data collection and labeling for the natural and unstructured environments to be used as groundwork for robot perception platform (e.g. semantic segmentation).

1) Collect public video datasets tailed for both structured and unstructured environment

2) Collect video datasets using commercial cameras (e.g. GoPro)

3) Collect public image datasets

4) Use the datasets for training, testing, and evaluation of various deep learning-based models to address the data augmentation, few-shot learning, and semantic segmentation objectives.

2.2. Objective 2: Investigate the deep learning-based image data augmentation methods to address the data scarcity problem in machine learning for perception.

1) Exploit a cycle-consistent generative adversarial network (CycleGAN) model to address the GAN-generated synthetic dataset issues: artifacts, shift color, and lost details.

2) Develop an unsupervised learning method based on a combination framework of CycleGAN and domain discriminator to address the data scarcity problem by training in an unpaired dataset condition and solve the SISR task to avoid checkerboard artifacts while preserving details.

3) Explore a mixed data sample augmentation strategy for training with time-frequency domain features.

2.3. Objective 3: Investigate few-shot learning approaches so that a small dataset can be used to train the inferencing models.

1) Develop a few-shot learning model to compensate for the sparseness of labeled data to utilize a large number of synthetic images and adjusts the networks from the source domain (synthetic data) to the target domain (real data) with a cross-domain training mechanism.

2) Explore an unsupervised adversarial training scheme in a domain adaptation framework to utilize the synthetic data and limited unlabeled real images to jointly train the segmentation network and effectively improve the segmentation network's generalization capability to the real domain.

3) Exploit a recurrent network architecture composed of a series of connected blocks in recurrent and feedback fashions for enhanced SR reconstruction. A recurrent network is exploited such that an SR image gets refined over a sequence of enhancement stages in a coarse-to-fine manner.

4) Explore a sparsity injecting method applied to the unsupervised learning approach for anomaly detection in video sequences.

2.4. Objective 4: Investigate the deep learning−based semantic segmentation methods to achieve enhanced scene for improved perception.

1) Develop a transformer architecture-based semantic segmentation network that incorporates a transformer (TrSeg) to adaptively capture the multi-scale information dependencies on the original contextual information by large margins.

2) Exploit a computationally efficient “Sketch-and-Fill Network (SFNet)” architecture with a three-stage Coarse-to-Fine Aggregation (CFA) module for semantic segmentation to alleviate the detail loss of the lower-resolution contextual information.

3) Develop a built-in memory module for semantic segmentation to overcome unexpected illumination changes.

2.5. Timeline for each objective.

No.	Date (yyyy.mm)	1st year efforts	Completion Status
1	2019.04 ~ 2019.06	Survey data augmentation and few-shot learning methods [Objective 2, Objective 3]	completed
2	2019.06 ~ 2019.09	Collect and label vision-based data in outdoor environment	completed

		[Objective 1]	
3	2019.09 ~ 2019.11	Implement conventional algorithms as baseline [Objective 2, Objective 3, Objective 4]	completed
4	2019.11 ~ 2020.02	Develop small data learning methods (e.g. GAN and few-shot learning) [Objective 3]	completed
5	2020.02 ~ 202-.04	Optimize the algorithms in accordance with low computations power [Objective 2]	completed

No.	Date (yyyy.mm)	2nd year efforts	Completion Status
1	2020.04 ~ 2020.07	Survey related works about state-of-art of simulation and few-shot learning based methods [Objective 3]	completed
2	2020.07 ~ 2020.08	Collect and label vision-based data in adverse outdoor environment [Objective 1]	completed
3	2020.08 ~ 2020.12	Develop a condition-invariant data augmentation algorithm based on GAN and few-shot learning [Objective 3]	completed
4	2020.12 ~ 2021.02	Evaluate and compare the performance of the proposed algorithms with other state-of-art methods [Objective 2, Objective 3, Objective 4]	completed
5	2021.02 ~ 2021.04	Optimize the algorithms in accordance with low computations power [Objective 2, Objective 4]	completed

No.	Date (yyyy.mm)	3rd year efforts	Completion Status
1	2021.04 ~ 2021.08	Optimize the layer structure and parameters of the GAN model [Objective 2, Objective 3, Objective 4]	completed
2	2021.08 ~ 2021.12	Optimize the layer structure of the few-shot learning model [Objective 3]	completed
3	2021.12 ~ 2022.02	Optimize the algorithms in accordance with low computations power [Objective 2, Objective 3, Objective 4]	completed
4	2022.02 ~ 2022.04	Evaluate and compare the performance of the proposed algorithms [Objective 2, Objective 3, Objective 4]	completed

3. Accomplishments:

3.1. Major Activity 1: Data collection and Labeling

Objective: Collect labeled datasets for the natural and unstructured environment to be used as ground work for robot perception platform (e.g. semantic segmentation).

Key outcomes:

1) We collected 2 public datasets, i.e., RUGD and Cityscapes for use in the semantic segmentation work, as well as collecting inhouse videos using GoPro cameras for evaluation. The inhouse Lab collected dataset is used for validating the effectiveness of the proposed algorithms. References [1, 2, 3, 4]

   - RUGD is a video dataset tailored for semantic segmentation in unstructured environments. It focuses on off-road autonomous navigation scenarios and contains

4,759/733/1,964 images for training, validation and testing. It has 24 categories with various scales of objects including tree, bush, bench, pole, and etc. The resolution of the images is 688 × 500.

- Cityscapes is a large-scale dataset for urban scene understanding containing 2,975/500/1,525 (5000 in total) images for training, validation and testing from multiple cities. It has 24 categories including car, bus, traffic sign, and etc. The resolution of the images is 2048 x 1024.

- In addition, we collected a few videos using GoPro cameras for evaluation. The videos are collected from a variety of outdoor areas including mountain paths, sidewalks and so on.

- Recorded semantic segmentation demo videos for RUGD test set, Cityscapes test set and videos collected by GoPro Cameras to validate the effectiveness of the project's proposed algorithms.

2) Collected dataset from the Places dataset.   The Places dataset [5] includes 5000 different underwater images. After voting by 5 researchers in the Lab, we categorized 937 clean images and 1014 turbid images are used for unpaired training set in the in the image-to-image translation work. Moreover, 890 underwater images are also collected as the benchmark dataset [6].

3) Collected NTIRE 2020 real world super resolution challenge dataset, which is composed of 2650 noisy low-resolution training images and 800 clean high-resolution images.

4) Collected TAU urban acoustic scenes for time-frequency data augmentation work: 2020 mobile benchmark [7], SECL_UMONS [8], and Voicebank + Diverse Environments Multichannel Acoustic Noise Database (DEMAND) benchmark [9]. TAU benchmark consists of 13965 training audio clips and 2970 test audio clips. SECL_UMONS dataset contains 2178 for training and 484 for test. DEMAND benchmark has 11572 training samples and 824 test samples.

5) Collected 829 lung CT slices from nine COVID-19 patients for training and 29 individual cases for test of the few-shot domain adaptation work. Also collected 10200 synthetic 2D

CT images generated inhouse at the ISPL Lab [10] as source domain and 1800 real CT slices as target domain.

6) Collected DIV8K dataset [11], which contains 1500 high resolution training samples and 100 high resolution validation samples.

7) Collected Ped2 [12] and Avenue [13] dataset. The Ped2 dataset consists of 16 training videos and 12 test videos with 12 anomalous events. The Avenue dataset contains 16 training videos and 21 test videos with 47 irregular events.

References for dataset efforts:

[1] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson,U. Franke, S. Roth, B. Schiele, The cityscapes dataset for semantic urban scene understanding, CoRR (2016)

[2] M. Wigness, S. Eum, J.G. Rogers, D. Han, H. Kwon, A RUGD dataset for autonomous navigation and visual perception in unstructured outdoor environments, in: International Conference on Intelligent Robots and Systems (IROS), 2019.

[3] B. Zhou, H. Zhao, X. Puig, T. Xiao, S. Fidler, A. Barriuso, and A. Torralba, ‘‘Semantic understanding of scenes through the ADE20K dataset,’’ 2016, arXiv:1608.05442. [Online]. Available: http://arxiv.org/abs/1608.05442

[4] B. Zhou, H. Zhao, X. Puig, T. Xiao, S. Fidler, A. Barriuso, and A. Torralba, ‘‘Semantic understanding of scenes through the ADE20K dataset,’’ 2016, arXiv:1608.05442. [Online]. Available: http://arxiv.org/abs/1608.05442

[5] Zhou, B.; Lapedriza, A.; Khosla, A.; Oliva, A.; Torralba, A. Places: A 10 million image database for scene recognition. *IEEE Trans. Pattern. Anal. Mach. Intell.* **2018**, *40,* 1452–1464.

[6] Li, C.; Guo, C.; Ren, W.; Cong, R.; Hou, J.; Kwong, S.; Tao, D. An underwater image enhancement benchmark dataset and beyond. arXiv 2019, arXiv:1901.05495

[7] A. Mesaros, T. Heittola, and T. Virtanen, “A multi-device dataset for urban acoustic scene classification,” in Proceedings of the Detection and Classification of Acoustic Scenes and Events 2018 Workshop (DCASE2018), November 2018, pp. 9–13. [Online]. Available: https://arxiv.org/abs/1807.09840

[8] M. Brousmiche, J. Rouat, and S. Dupont, “Secl-umons database for sound event classification and localization,” in ICASSP 2020- 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020, pp. 756–760.

[9] C. Valentini-Botinhao, X. Wang, S. Takaki, and J. Yamagishi, “Investigating rnn-based speech enhancement methods for noiserobust text-to-speech.” in SSW, 2016, pp. 146–152

[10] Jiang Y, Chen H, Loew MH, Ko H (2020) Covid-19 ct image synthesis with a conditional generative adversarial network IEEE Journal of Biomedical and Health Informatics

[11] Shuhang Gu, Andreas Lugmayr, Martin Danelljan, Manuel Fritsche, Julien Lamour, and Radu Timofte. Div8k: Diverse 8k resolution image dataset. In 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pages 3512–3516. IEEE, 2019.

[12] W. Li, V. Mahadevan, and N. Vasconcelos. Anomaly detection and localization in crowded scenes. IEEE transactions on pattern analysis and machine intelligence, 36(1):18–32, 2013.

[13] C. Lu, J. Shi, and J. Jia. Abnormal event detection at 150 fps in matlab. In Proceedings of the IEEE international conference on computer vision, pages 2720–2727, 2013.

3.2. Major Activity 2: Investigate data augmentation methods

Objective: Investigate the deep learning-based image data augmentation methods to address the data scarcity problem in machine learning for perception

Key outcomes:

1) **A cycle-consistent generative adversarial network (CycleGAN) model is developed** to address the GAN-generated synthetic dataset issues: artifacts, shift color, and lost details. Typical deep learning models for image enhancement get trained by paired synthetic datasets. As a result, these models are most effective for synthetic image enhancement tasks but less so for real-world images. In contrast, the cycle-consistent generative adversarial network (CycleGAN) models can be trained with unpaired datasets. On the other hand, the performance of the CycleGAN is highly dependent upon the dataset as it may generate unrealistic images with less content information than the original images. The key outcome of this research effort is that a novel solution is proposed that starts with a CycleGAN and adds a pair of discriminators to preserve the contents of the input image while enhancing the image quality. As a part of the solution, an adaptive weighting method gets added to limit the losses of the two types of discriminators to balance their influence and stabilize the training procedure. Extensive experiments show that the proposed method significantly outperforms the state-of-art methods on real-world underwater images. The process and results are open and published in the Journal of Marine Science and Engineering 2019.

2) **An unsupervised learning method is developed based on a combination of CycleGAN and domain discriminator to address the data scarcity problem** when training using unpaired datasets. The combined framework solves the SISR task and avoids checkerboard artifacts while preserving details. In previous approaches, paired datasets would be essential for training with assumed levels of image degradation. However, in real-world SR applications, training datasets are typically not of low-resolution and high-resolution image pairs, but only low-resolution images with unknown degradation would be made available as inputs. To address this issue, a cycle-in-cycle GAN is developed as an unsupervised learning model that uses unpaired datasets. In addition, several losses attributed to image contents, such as pixel-wise loss, VGG feature loss, and SSIM loss, are added for achieving stable learning and performance improvement. By combining a noise discriminator, texture discriminator, and color discriminator into a single domain discriminator, it becomes a unique structure that helps guide the generated images to follow the target domain distribution rather than the source domain. The effectiveness of the proposed model is validated in quantitative and qualitative experiments using the NTIRE2020 real-world SR challenge dataset. The process and results are published in the IEEE CVPRW 2020 Proceedings.

3) **A mixed data sample augmentation strategy is explored for** training with time-frequency domain features. A mixed data sample augmentation strategy is developed to enhance the performance of inferencing models for audio scene classification, sound event classification, and speech enhancement tasks. For these audio inferencing tasks, both time and frequency features would be helpful to capture the essential information for training. While several augmentation methods have shown effective in improving image classification performance, their efficacy toward time-frequency domain features on audio has not been assured. To aim for the audio aspect of the issue, a novel data augmentation approach “Specmix” is specifically designed for dealing with time-frequency domain features. The prescribed augmentation method essentially suggests mixing two different data samples by applying time-frequency masks shown effective in preserving the spectral correlation of each audio data sample. For effectiveness analysis, the experiments on acoustic scene classification, sound event classification, and speech enhancement tasks show that the proposed Specmix improves the inferencing performance of various neural network architectures by a maximum of 2.7%. The process and results are published in the Proceedings of Interspeech 2021.

3.3. Major Activity 3: Investigate few-shot learning approaches

Objective: Investigate few-shot learning approaches so that small dataset can be used for training inferencing models

Key outcomes:

4) **A few-shot learning model is developed to compensate for the sparseness of labeled data to utilize a large number of synthetic images** and adjusts the networks from the source domain (synthetic data) to the target domain (real data) with a cross-domain training mechanism. It is essentially a supervised domain adaptation-based few-shot learning method for CT scan images. The novelty of the proposed method consists in constructing a cross-domain training architecture by integrating a Siamese network and introducing two cross-domain training losses in addition to a classification loss. Siamese network-based architecture and the proposed cross-domain losses are demonstrated effective in handling the domain shift problem between the source and the target domains. Experimental results on the public CT dataset show that the proposed method outperforms the other state-of-the-art supervised domain adaptation methods on a few-shot diagnostic task. The procedure and results are published in the Proceedings of IEEE ICASSP 2021.

5) **An unsupervised adversarial training scheme is proposed in a domain adaptation framework to utilize the synthetic data and limited unlabeled real images to** jointly train the segmentation network and effectively improve the segmentation network's generalization capability to the real domain. In particular, the synthetic data and limited unlabeled real CT images are utilized to jointly train the segmentation network. Further, a novel domain adaptation module is proposed, wherein it is used to align the two domains and effectively improve the segmentation network's generalization capability to the real domain. Essentially, it is an unsupervised adversarial training scheme, which encourages the segmentation network to learn the domain-invariant feature, so that the robust feature can be used for segmentation. Experimental results demonstrate that our method achieves state-of-the-art segmentation performance on CT images in segmentation task. The procedure and results are published in the journal of Applied Intelligence, 2022.

6) **A recurrent network architecture composed of a series of connected blocks in recurrent and feedback fashions is developed for enhanced SR reconstruction such that more detailed information is produced at the output through the coarse-to-fine**

process. A recurrent network is exploited such that a SR image can be refined over a sequence of enhancement stages in a coarse-to-fine manner. Single image extreme Super Resolution (SR) is a difficult task as scale factor in the order of 10X or greater is typically attempted as such attempts often result fuzzy quality and loss in details in reconstructed images. By use of recurrent network, an SR image is refined over a sequence of enhancement stages in coarse to fine manner. Additionally, each stage involves back projection of SR image to LR images for continuously being refined during the sequence. The procedure and results are competed in the SR Challenge and published in the Proceedings of IEEE CVPRW, 2020

7) **A sparsity injecting method is explored and applied to the unsupervised learning approach** for anomaly detection in video sequences.

   As a way to minimize the learning with small size dataset, sparsity injection could be a solution approach to explore. Here we apply it to anomaly detection task in video which is a challenging problem in computer vision tasks. Deep networks recently have been successfully applied and achieved competitive performance in anomaly detection. Modern deep networks employ many modules which extract important features. However, these methods generally require a tremendous amount of computational load and training parameters. In this effort, a sparsity injecting module is proposed, which reinforces the feature representation of the existing model and presents the abnormality score function using sparsity. In experimental results, the proposed sparsity injecting module improves the performance of state-of-the-art methods without additional trainable parameters. The procedure and results are published in the Proceedings of IEEE AVSS, 2021

3.4. Major Activity 4: Deep learning based semantic segmentation

Objective: Investigate the deep learning-based semantic segmentation methods for enhanced scene perception

Key outcomes:

8) **A transformer architecture-based semantic segmentation network is developed** that incorporates a transformer (TrSeg) to adaptively capture multi-scale information with the

dependencies on original contextual information to capture multiscale information by large margins. Capturing the existence of objects in an image at multiple scales has been a challenging problem. This research effort addressed the semantic segmentation task based on transformer architecture. Unlike existing methods that capture multi-scale contextual information through infusing every single-scale piece of information from parallel paths, the proposed semantic segmentation network incorporates a transformer (TrSeg) to adaptively capture multi-scale information on original contextual information. Given the original contextual information as keys and values, the multi-scale contextual information from the multi-scale pooling module as queries is transformed by the transformer decoder. The experimental results show that the proposed TrSeg outperforms the other methods of capturing multiscale information by large margins. The proposed method and results are published in the Pattern Recognition Letters, 2021

9) **A computationally efficient "Sketch-and-Fill Network (SFNet)" architecture** is developed with a three-stage Coarse-to-Fine Aggregation (CFA) module for semantic segmentation to alleviate the detail loss of the lower-resolution contextual information. Recent efforts in semantic segmentation using a deep learning framework often require heavy computation, making them impractical to be used in real-world applications. There are two reasons that produce prohibitive computational costs: 1) heavy backbone CNN to create a high resolution of contextual information and 2) complex modules to aggregate multi-level features. In this research effort, we propose the computationally efficient architecture called ''Sketch-and-Fill Network (SFNet)'' with a three-stage Coarse-to-Fine Aggregation (CFA) module for semantic segmentation wherein the lower-resolution contextual information is first produced so that the overall computation in the backbone CNN is largely reduced. Then, to alleviate the detail loss of the lower-resolution contextual information, the CFA module forms global structures and fills fine details in a coarse-to-fine manner. To preserve global structures, the contextual information is passed without any reduction to the CFA module. Experimental results show that the proposed SFNet achieves significantly lower computational loads while delivering comparable or improved segmentation performance with state-of-the-art methods. Qualitative results show that the proposed method is superior to state-of-the-art methods in capturing fine detail while keeping global structures on Cityscapes, ADE20K, and RUGD benchmarks.

10) **A built-in memory module is developed for semantic segmentation to overcome unexpected illumination changes**. With the availability of many datasets tailored for

autonomous driving in real-world urban scenes, semantic segmentation for urban driving scenes achieves significant progress. However, semantic segmentation for off-road, unstructured environments is not widely studied. Directly applying existing segmentation networks often results in performance degradation as they cannot overcome intrinsic problems in such environments, such as illumination changes. In this research effort, a built-in memory module for semantic segmentation is proposed to overcome these problems. The memory module stores significant representations of training images as memory items. In addition to the encoder embedding-like items, the proposed memory module is specifically designed to cluster together instances of the same class even when there are significant variances in embedded features. Therefore, it makes segmentation networks better deal with unexpected illumination changes. A triplet loss is used in training to minimize redundancy in storing discriminative representations of the memory module. The proposed memory module is general so that it can be adopted in a variety of networks. Experiments were conducted on the Robot Unstructured Ground Driving (RUGD) dataset and RELLIS dataset, which are collected from off-road, unstructured natural environments. Experimental results show that the proposed memory module improves the performance of existing segmentation networks and contributes to capturing unclear objects over various off-road, unstructured natural scenes with equivalent computational cost and network parameters. As the proposed method can be integrated into compact networks, it presents a viable approach for resource-limited small autonomous platforms. The proposed method and results are published in the Proceedings of IEEE IROS, 2021.

4. Results disseminated:

The results are disseminated to largely scholarly communities: technical publications such as journals and prestigious conferences such as IROS, ICASSP, CVPR, and AVSS. Each publication is listed under the corresponding effort delineated in **2. Technical Approaches** section as follows.

1) A cycle-consistent generative adversarial network (CycleGAN) model is developed to address the GAN generated synthetic dataset issues: artifacts, shift color, and lost details.

(**Publication 1:** Jaihyun Park, David K. Kan, Hanseok Ko, "*Adaptive Weighted Multi-Discriminator CycleGAN for Underwater Image Enhancement*", Journal of Marine Science & Engineering, June 2019. (https://doi.org/10.3390/jmse7070200)

2) An unsupervised learning method is developed based on a combination framework of CycleGAN and domain discriminator to address the data scarcity problem by training in an unpaired dataset condition and solve the SISR task to avoid checkerboard artifacts while preserving details.

(**Publication 2:** Gwantae Kim, Jaihyun Park, Kanghyu Lee, Junyeop Lee, Jeongki Min, Bokyeung Lee, David K. Han, and Hanseok Ko, *"Unsupervise Real-World Super Resolution with Cycle Generative Adversarial Network and Domain Discriminator,"* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2020. Pg 456-457)

3) A mixed data sample augmentation strategy is explored for training with time-frequency domain features.

(**Publication 3:** Gwantae Kim, David Han, Hanseok Ko, "*SpecMix: A Mixed Sample Data Augmentation method for Training with Time-Frequency Domain Features*", Proceedings of Interspeech, 2021)

4) A few-shot learning model is developed to compensate for the sparseness of labeled data to utilize a large number of synthetic images and adjusts the networks from the source domain (synthetic data) to the target domain (real data) with a cross-domain training mechanism.

(Publication 4: Yifan Jiang, Han Chen, David Han, Hanseok Ko, "Few-Shot Learning for CT Scan Based Covid-19 Diagnosis." ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2021): 1045-1049.)

5) An unsupervised adversarial training scheme is proposed in a domain adaptation framework to utilize the synthetic data and limited unlabeled real images to jointly train the segmentation network and effectively improve the segmentation network's generalization capability to the real domain.

(**Publication 5:** Chen H, Jiang Y, Loew M, Ko H. “*Unsupervised domain adaptation based COVID-19 CT infection segmentation network*.” Applied Intelligence, 2022; Volume 52(6):6340-6353, doi: 10.1007/s10489-021-02691-x)

6) A recurrent network architecture composed of a series of connected blocks in recurrent and feedback fashions is developed for enhanced SR reconstruction such that more detailed information is produced at the output through the coarse-to-fine process.

(**Publication 6:** Junyeop Lee, Jaihyun Park, Kanghyu Lee, Jeongki Min, Gwantae Kim, Bokyeung Lee, Bonhwa Ku, David K. Han, Hanseok Ko, “*FBRNN: feedback recurrent neural network for extreme image super-resolution*” The IEEE/CVF Conference on Computer Vision & Pattern Recognition (CVPR) Workshops, 2020, pp. 488-489 (http://openaccess.thecvf.com/content_CVPRW_2020/papers/w31/Lee_FBRNN_Feedback_Recurrent_Neural_Network_for_Extreme_Image_Super-Resolution_CVPRW_2020_paper.pdf)

7) A sparsity injecting method is explored and applied to the unsupervised learning approach for anomaly detection in video sequences.

(Publication 7: B. Lee and H. Ko, "*Injecting Sparsity in Anomaly Detection for Efficient Inference*," 2021 17th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2021, pp. 1-5, doi: 10.1109/AVSS52988.2021.9663843.)

8) A transformer architecture based semantic segmentation network is developed that incorporates a transformer (TrSeg) to adaptively capture multi-scale information with the dependencies on original contextual information to capture multiscale information by large margins.

(**Publication 8**: Youngsaeng Jin, David Han, Hanseok Ko, “*TrSeg: Transformer for semantic segmentation*,” Pattern Recognition Letters, Volume 148, 2021, Pages 29-35)

9) A computationally efficient “Sketch-and-Fill Network (SFNet)” architecture is developed with a three-stage Coarse-to-Fine Aggregation (CFA) module for semantic segmentation to alleviate the detail loss of the lower-resolution contextual information.

(**Publication 9:** Youngsaeng Jin, S. Eum, David Han and Hanseok Ko, "*Sketch-and-Fill Network for Semantic Segmentation*," in IEEE Access, vol. 9, pp. 85874-85884, 2021, doi: 10.1109/ACCESS.2021.3088854.)

10) A built-in memory module is developed for semantic segmentation to overcome unexpected illumination changes.

(**Publication 10:** Youngsaeng Jin, David Han and Hanseok Ko, "*Memory-based Semantic Segmentation for Off-road Unstructured Natural Environments*," 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2021, pp. 24-31, doi: 10.1109/IROS51168.2021.9636620.)

Impacts

On the category of the development of the principal discipline(s) of the project:

This project investigated the data augmentation approach and the few-shot learning approach to tackle the data scarcity issue of the perception problem related to autonomous robotic maneuvers in previously unseen environments. Ground robot's anticipated traversing environments and scenes are typically unusual, and data for the current paradigm of deep learning is either scarce or non-existent. These approaches are expected to provide key solutions to developing terrestrial robotic vehicles capable of perceiving and understanding novel environments. The following are the key impacts of the findings as a result of the research efforts conducted during the research period.

The impact of the data augmentation efforts has been to verify and demonstrate that the deep learning-based image data augmentation methods can be used to address the data scarcity problem in machine learning for perception in unstructured scenes. In previous approaches, a paired dataset is required for training with assumed levels of image degradation. We developed the cycle-consistent generative adversarial network (CycleGAN) model and its variation to address the GAN-generated synthetic dataset issues: artifacts, shift color, and lost details, by unsupervised learning model so that more abundant unpaired datasets can be employed for the image enhancement task. By injecting conditional activations such as adaptive weighting or adjusting discriminators to balance their influence and stabilize the training procedure, the proposed methods significantly outperform the state-of-the-art methods on real-world images for enhancement or raising the quality of super-resolution images without the need for a paired dataset for training.

Another significant impact has been to show that it is possible to employ few-shot learning approaches so that small datasets can be used for training inferencing models to compensate for the sparseness of labeled data to utilize a large number of synthetic images and adjusts the networks from the source domain (synthetic data) to the target domain (real data) with a cross-domain training mechanism. Hence, this domain adaptation framework that utilizes the synthetic data and limited unlabeled real images to jointly train the segmentation network effectively improves the generalization capability to the real domain. Essentially, it is an unsupervised adversarial training scheme, which encourages the segmentation network to learn the domain-invariant feature, so that the robust feature can be used for segmentation. Experimental results demonstrate that the proposed method achieves state-of-the-art segmentation performance on CT images for semantic segmentation tasks.

Last but not least is the impact we can find from investigating the deep learning-based semantic segmentation methods for enhanced scene perception. By incorporating a unique design measure in architecture, it is possible to achieve preserving multi-scale information and maintain the network light as well as cope with unexpected illumination changes. This means that with the proper design of the network, we can generate a high-quality semantic segmentation in a highly unstructured scene. The proposed transformer architecture composed of the coarse-to-fine aggregation module, along with a built-in-memory, showed that it is indeed possible to capture fine detail while keeping global structures and improving the performance of existing segmentation networks. It is verified that the proposed network captures unclear objects over various off-road, unstructured natural scenes with equivalent computational cost and network parameters. As the proposed method can be integrated into compact networks, it presents a viable approach for resource-limited small autonomous platforms.

Changes

Changes in approach

None

Problems or delays

None

Expenditure Impacts

None

Significant changes in the use or care of human subjects, vertebrate animals and/or biohazards

Not Applicable

Changes to the primary place of performance from that originally proposed

None

Technical Updates

None