Query Details

Back to Main Page

1. Please check and confirm if the authors and their respective affiliations have been correctly identified. Amend if necessary.

Confirmed: all affiliations are correct.

2. Please provide high-resolution source file for Fig. 1.

A high resolution file for Fig. 1 is uploaded (new_fig1.png).

Image-Based Pre- and Post-conditional Probability Learning for Efficient Situational Assessment and Awareness

Jie Wei ✉
Email : jwei@ccny.cuny.edu
Affiliationids : Aff1, Correspondingaffiliationid : Aff1

Weicong Feng
Email : wfeng@gradcenter.cuny.edu
Affiliationids : Aff1

Erik Blasch
Email : erik.blasch.1@us.af.mil
Affiliationids : Aff2

Erika Ardiles-Cruz
Email : erika.ardiles-cruz@us.af.mil
Affiliationids : Aff3

Haibin Ling
Email : hling@cs.stonybrook.edu
Affiliationids : Aff4

Aff1 Department of Computer Science, City College of New York, New York City, USA

Aff2 Air Force Research Lab, Arlington, USA

Aff3 Air Force Research Lab, Rome, USA

Aff4 Department of Computer Science, Stony Brook University, New York City, USA

Abstract

With the increasing amount and severity of environmental disasters, the importance of situational assessment and awareness for human assistance and disaster response is critical. AQ1 During natural disasters in populated regions, proper assistance and response efforts can only be planned and deployed effectively when the damage levels can be resolved promptly. Damage-level assessment is aided by aerial imagery. While human labeling of images provides a measure of credibility, in the presence of a large volume of image data, it takes a long time and great effort to achieve situational assessment and awareness which can significantly hamper the operational response time. Recently, extreme computational power has enabled Deep Learning to analyze large images. This paper presents an efficient and scalable humanitarian assistance and disaster response application for situational assessment and awareness using a method called *Image-based Pre- and Post-conditional Probability learning*, which matches the pre- and post-disaster images by effectively encoding one image that determines the damage levels through a deep learning method. Two example scenarios of humanitarian assistance and disaster response applications are examined: (1) pixel-wise semantic segmentation, and (2) contrastive learning patch-based damage classification, both showing promising results in the examined scenarios which motivates the application of the deep learning enhanced methods based on our new method to achieve computational efficiency while maintaining classification accuracy.

Keywords

Deep learning
Granular computing
Disaster management
Image classification

Image segmentation
Representational learning
Transfer learning

1. Introduction

To effectively deliver Human Assistance and Disaster Response (HADR) information requires efficient Situational Assessment and Awareness (SAA) [1]. It remains a considerable challenge to develop HADR SAA since it is typically a disastrous event with extremely few prior exemplars to resolve damage severity levels and even fewer near-real-time response recorded results. Usually, the HADR situation includes a large urban area or an infrastructure of crucial importance demanding critical region inspection and checking, prompt and accurate damage assessments (DA), resilient operating conditions analysis, and knowledge of highly hazardous and time-sensitive scenarios. To conduct HADR Object of Interest (OoI) analysis, the community designates (1) inanimate static objects as buildings, infrastructure, roads, and dwellings among others; (2) dynamic inanimate objects could be cars; and (3) dynamic animate objects would be people and animals. In all these situations, manual evaluation by human experts for real-time or near-real-time damage severity level assessments for all the OoIs is nearly impossible and unscalable. More specifically, the labeled OoIs are usually the static objects mostly associated with urban settings. The manual annotation procedure is exceedingly labor-intensive and time-consuming as each location of interest must be carefully inspected to make a call on the correct severity category where a great amount of domain knowledge on the object types and damage severity levels are demanded. The situation is even more severe for major disasters spanning a large region which slows down the HADR SAA endeavors. In response, (1) State-of-the-art hardware, such as GPU and high-performance computing, and (2) current software, e.g., Artificial Intelligence and Machine learning (AI/ML), and deep learning (DL) in particular [2], sensor fusion [3] approaches, and architectures such as edge and fog computing, should be called upon [4].

In this study, the xView2 dataset (https://xview2.org) is utilized. The xView2 dataset contains a large set of annotated high-resolution satellite images for building damage assessment. In the xView2 dataset, polygons and damage scores for each building are carefully annotated and reviewed by expert analysts. More than 850,000 building polygons from six different types of natural disasters, namely, *hurricanes, volcanoes, fire, tsunami, flooding, and earthquakes,* around the world are made available, covering more than 45,000 square kilometers in area. The ground truths of the xView2 dataset were carefully reviewed by experts for accuracy and utility for the specific task of automated building localization and damage assessment with four different damage severity levels: *no damage, minor, major, and total damages,* making it the premier source of high-quality labeled imagery for SAA and HADR. To ensure fair comparisons for different methods, the official train and test dataset split is provided in the publicly available testbed, where the number of images in train and test datasets is 2800 and 933 (with a ratio of 3 to 1), respectively. It must be pointed out that in the designated training dataset, the foreground regions, namely, the buildings, only occupy 5.92% of the regions on average resulting in about 94.08% (94.15%) of the image being background pixels. Furthermore, the four severity levels are *not balanced* with 4.37% (4.48%), 0.54% (0.49%), 0.68% (0.57%), and 0.32% (0.31%), for *no damage, minor, major,* and *total damage* severity levels, in the train (test) datasets, respectively, with most *no damage* regions and very few *total damage* regions. The unbalanced nature of different classes in the dataset calls for special care in devising effective classifiers and data augmentation approaches.

Winning teams from the xView2 data challenges [5], along with advanced AI/ML methods [6], continue to expand the analysis of overhead imagery exploitation. The advanced methods that offer computational and data efficiency should offer practical decision-making utility for HADR operations. Future methods should be computationally and data-efficient because:

(1) *Data Availability:* DA and HADR applications operate on a generally small amount of data which challenges DL training from scratch, and

(2) *Personnel Availability*: Most DA and HADR operators lack immense mobile computing power access to conduct on-site missions.

The *Image-based Pre- and Post-conditional Probability Learning* (IP2CL) encoding scheme effectively encodes OoIs, e.g., DA-interested regions/targets or HDAR buildings, in before- and after-disaster images that are to be used for damage level assessment. Two different scenarios where IP2CL is demonstrated include: (1) *pixel-wise semantic segmentation* that determines a damage-level classification for each OoI pixel, and (2) *patch-based damage level classification* that classifies a damage level as a small image patch centered around the OoI. Both frameworks support DA and HADR where the former uses a wide span of regions to select the most urgent regions; whereas the latter identifies one crucial OoI, e.g., a school or a hospital area, which requires immediate mission judgments to ensure a timely and safe response.

Current methods for HADR to encode the pre- and post-disaster information rely on contrastive learning (CL) methods as evidenced by the fact that almost all the xView2 challenge winning methods are based on a CL approach because it is the only DL method that can handle two images. However, as mentioned, the foreground regions only occupy about 5% of the area, and the crucial features learned by CL are likely more relevant to the background than the foreground; the CL's efficacy is thus seriously compromised as evidenced by the low damage level classification performances of all winning methods.

To address the aforementioned issues, *the objective of this work is to apply Machine learning, deep learning in particular, and granular computing techniques to develop a data and computing effective algorithm* so that the situation before and after a certain event or action can be encoded to be systematically analyzed and classified by the DL. The main contributions of this work are:

1. *Data representation* using IP2CL to encode the situation before and after events/actions for efficient analysis

2. *Data augmentation* via generative adversarial network (GAN) to effectively combat the data imbalance problem
3. *Transfer learning* and contrastive learning for semantic segmentation based on IP2CL
4. *Contrastive learning* for image-level classification using IP2CL

This chapter is organized as below: Sect. 2 presents related work; Sect. 3 expands on the new data representation IP2CL; Sect. 4 presents two important applications of the new IP2CL representation: Sect. 4.1 describes two generative approaches to tackle the data unbalance problems; Sect. 4.2 presents the IP2CL-based pixel-wise semantic segmentation and the associated results; Sect. 4.3 details the patch-based damage level classification based on IP2CL and the corresponding experimental results; finally, we conclude this paper with more remarks in Sect. 5.

2. Related Work in Machine Learning and Granular Computing Framework

The work developed here is inspired by the combination of powerful machine learning and the granular computing framework. Our method utilizes the emerging *granular computing* framework [7] as it deals with considerably different granular levels of data. Granular computing is of great utility in visual perception at various levels. In [8], granular computing clustering is conducted to segment images as a partition of sets. The classification and semantic segmentation of structured data were systematically addressed using the granular computing framework [9] in the graph domain. Granular computing served as a new framework behind deep learning that simulates the hierarchical problem-solving process [10] with useful applications in image processing and pattern recognition. More systematic studies of this valuable granular computing framework and its broad array of applications can be found in [11]. With the granular computing framework, in pixel-wise building segmentation, the pixels in the overhead images with granules of the lowest level are used to classify the local damage levels of buildings, which is a large cluster of semantically related pixels in the overhead images. The granules of much higher semantic level, by using DL techniques such as transfer learning and U-Net can provide knowledge of the general damage level. In patch-based damage level classification, the *entire image* is viewed as one huge granule to discern the global damage level assessment. The encouraging performances as reported in this chapter further confirm the powerful new philosophical granular computing framework in real-world applications.

Besides granular computing framework, Machine Learning, Deep learning in particular, is the enabling theme of this chapter, which has been widely employed in HADR. In [12], the authors used the xBD dataset to compare the performance of Siamese neural networks with the classical encoder-decoder architecture to achieve damage assessment from satellite images and find that the Siamese architecture outperforms the classical approach, especially for damage classification. The authors compared different encoders and decoders, including ResNet-18, -34, -50, -101, DenseNet-169, -201, SE-ResNext-50, and Inception-v4, and discovered the optimal model architecture. They also explored different loss functions, such as cross-entropy loss and weighted cross-entropy loss, to address the challenge of class imbalance in the damage detection task. Image augmentations, including spatial and color augmentations, were used to improve model accuracy.

A two-stage method was developed in [13] combining modified YOLOv4 and support vector machine (SVM) for post-earthquake building location and damage assessment using satellite remote sensing optical images. This method uses modified YOLOv4 for object detection and an SVM for classification, resulting in high accuracy in a binary damage assessment.

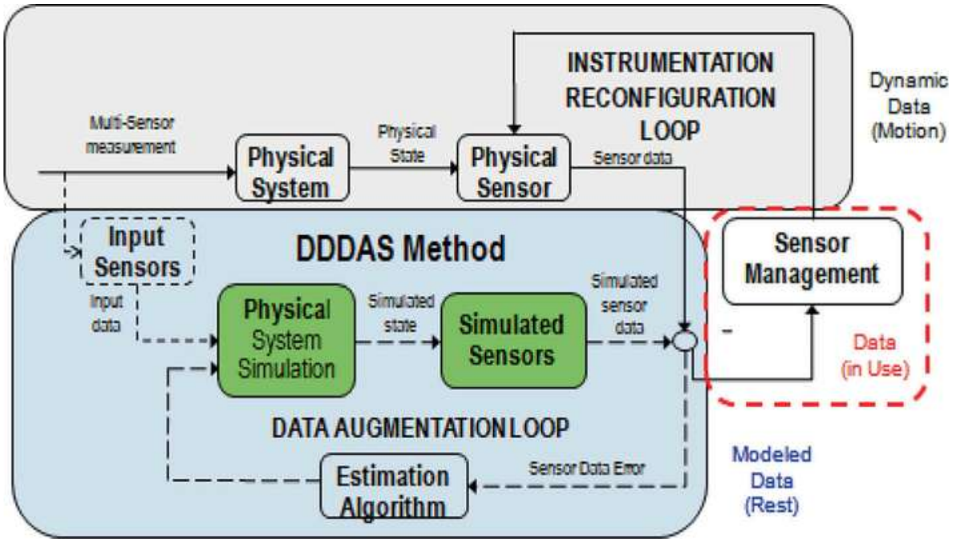
Due to the challenge of obtaining the pre- and post-disaster pair event images, some researchers turn to post-event only damage detection. A deep-learning framework called BDD-Net++ was developed in [14] for post-event-only building damage detection, which combines convolution blocks and transformer blocks. BDD-Net++ utilizes multiscale residual convolution blocks and self-attention blocks to improve performance. The framework is validated on two real-world datasets (the Haiti earthquake and the Bata explosion) and demonstrates high accuracy in binary classification.

Some researchers used more complicated architectures in damage assessment. For example, an architecture combining incremental learning, transfer learning, fusion learning, and generative adversarial learning was introduced in [15] to identify collapsed buildings in natural disasters. They use DREAM-B+, which is expanded from DREAM-B and xBD, and the translated images from them to train model, with an adaptive learning rate. On the other hand, they use CycleGAN [16] to translate images for incremental learning. The CycleGAN approach is validated by the Haiti earthquake and the Nepal earthquake data set. Due to the training data labels being modified significantly, whether this architecture can be generalized is still to be confirmed.

HADR and SAA approaches require effective designs of which the Dynamic Data Driven Applications Systems (DDDAS) paradigm construct has shown great promise (Fig. 1). AQ2 DDDAS [17] has demonstrated the ability to leverage DL for near-real-time situations in which the original DL-trained model is updated from continuous learning to support the effective labeling of SAA updates. SAA requires analysis from a variety of sensors such as wide-area motion imagery.

Fig. 1

The DDDAS diagrammatic chart



An unfortunately common problem, typically in HADR and SAA applications, is *the lack of balanced labeled data*. A recently popularized solution within the DL framework is through Generative Adversarial Networks (GAN), which is often used to address issues of unlabeled data, limited and unbalanced examples, and plausible scenarios. The Pix2Pix conditional GAN framework was proposed as a general approach for image-to-image translation [18] that can considerably increase images of similar statistical distributions.

In HADR, to classify buildings into different labels, the semantic segmentation technique is needed. U-Net [19] is a popularly used deep network architecture for semantic segmentation that consists of a series of contracting/down-sampling processes via a series of convolutional networks followed by an expansive/up-sampling process done by a de-convolution network, it has achieved great success in a broad array of successes in scenarios such as medical image segmentation, remote sensing applications in road detection, and 3D object detection [20]. Valuable pixel-wise classification performance can be achieved with the U-Net framework.

Given the extremely large number of parameters in any deep neural network, it is infeasible to assemble adequate training data to sufficiently estimate the huge volume of network parameters. *Transfer learning* [21] effectively tackles the problem of limited training data, where the reuse of a pre-trained network on a new problem takes advantage of the knowledge obtained from previous training data to elucidate knowledge on new problems. There are two different types of transfer learning, one is *fine-tuning* where all parameters of the pre-trained net are revised in the training process of the new training data; the other is *feature extraction* where the pre-trained network serves as the feature generator: a representation vector is created by feeding the new training data to the old model and a new fully-connected layer or a shallow convolutional network is trained over these feature vectors. By using transfer learning, valuable knowledge from a large image set, e.g., ImageNet, in ResNet or VGG deep nets used in this work, and wiki in chatGPT [22], can be effectively transferred to the new task, where the training data and computational resources are far less than those for the pre-trained network. Within transfer learning is *domain adaption*, where the knowledge of one source data is utilized for target data within the same domain, such as imagery collected in a region is learned and adapted to another region [23].

3. Materials and Methods for IP2CL

3.1. Dynamic Data-Driven Applications System Approach

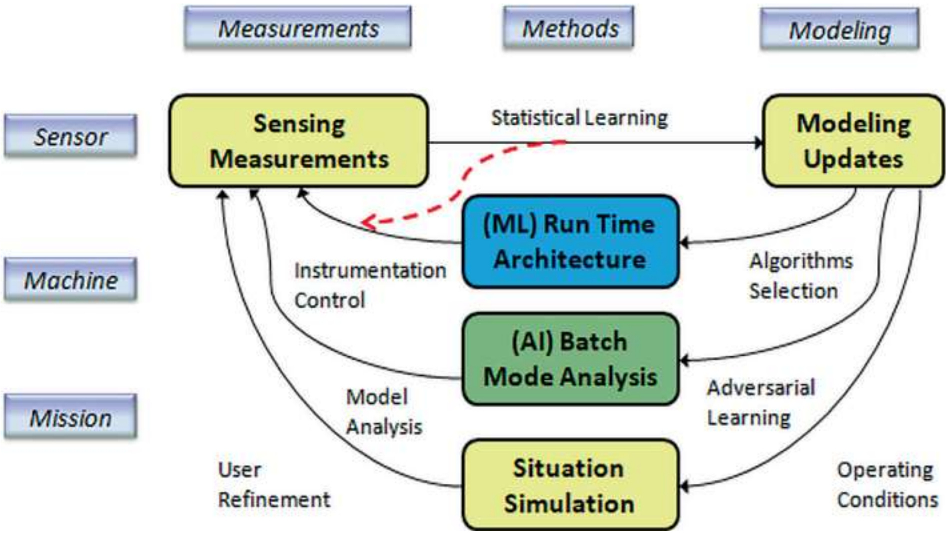
To design effective data and computing efficient HADR SAA methods, this chapter applies the DDDAS approach combined with the emerging utility of deep learning.

The overall chart of DDDAS is illustrated in Fig. 1 [24] wherein a standard signal processing of the instrumentation is augmented with data from the simulation. The instrumentation reconfiguration loop contains normal methods including computer vision tracking, internet of things processing, and environmental analysis where sensors are managed to understand the dynamic situation. The data augmentation loop includes first-principle physics that affords simulations such as using computer-aided design models of buildings to predict damage, weather predictions, and forecasting, as well as the movement of entities based on disaster evacuation scenarios. Hence, the loop is "as-is" while the data augmentation loop is "what-if" analytics and the loop is essentially a predictive digital twin.

During the past decades, the resurgence of ML methodologies using the DL framework engendered the DDDAS framework through powerful computers, copious amounts of data, and methods for training models for predictive analytics. For example, collecting a large volume of data for training purposes gives rise to insights into the physical world. Under the HADR framework, the natural terrain scene is modeled by the physical characteristics, and exploiting the prior knowledge analyzes a scene to be used to augment future data. When a natural disaster presents itself, the previous models should be adapted to facilitate the scene variations, e.g., a destructed building turns into a new scene, and demands evolved models within the framework of DDDAS. The common difficulty for DDDAS and DL is to provide the ground truths for the parameters from which computing and data-efficient algorithms are required during time-sensitive situations. Hence, as per the traditional definitions of a digital "twin", there is the static model of the building, but also the need for a real-time dynamic predictive model of the evolving damage from simulations (Fig. 2).

Fig. 2

The DDDAS concept integrating sensor measurements, model updates, and situation simulation augmented with AI/ML



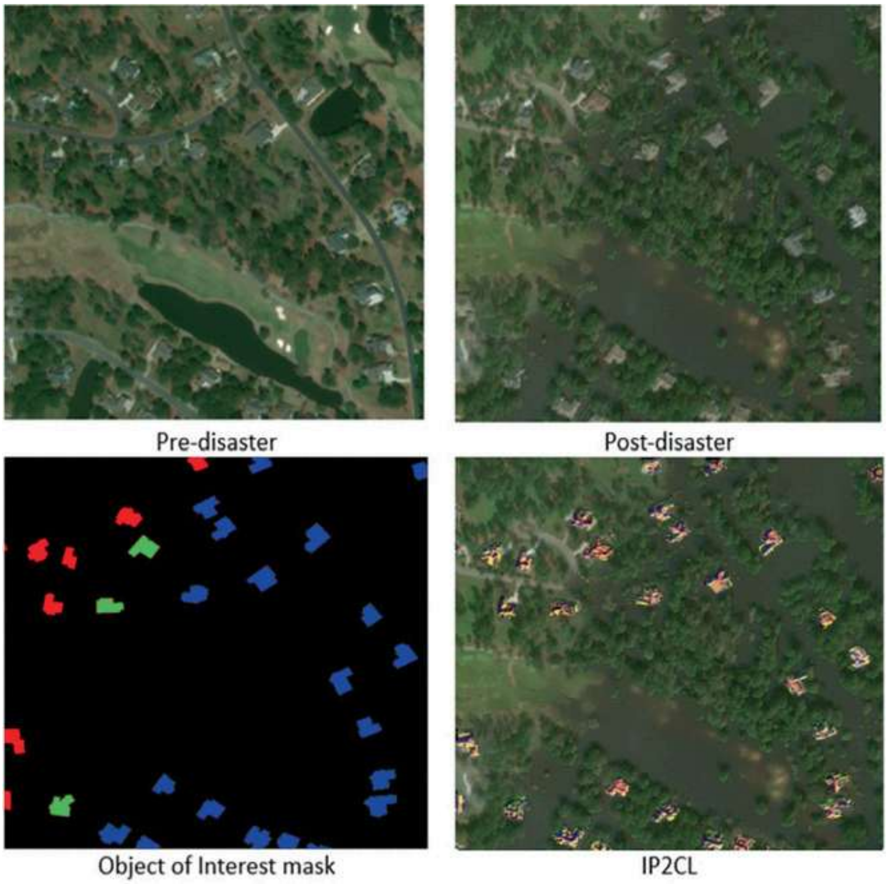
DDDAS methods utilize near-real-time simulations, where the original model trained by the DL process can be fine-tuned from adaptive learning and the efficient annotation of SAA changes. DDDAS is thus a framework of crucial interest to attain HADR objectives, which incorporates causal estimation analysis from both gathered data and data augmentation from models updated with DL methods, as shown in Fig. 2. Figure 2 highlights that the feedback associated with sensing the current situation and updating the knowledge based on the context affords three types of supporting data: ML run time analytics, AI batch learning such as deep learning methods, and situation simulations; that build a comprehensive model of the scenario. The ML analysis includes proven methods that quickly process the incoming data such as using a principle component analysis. When there is collected data of representative exemplars such as prior images from damaged buildings, then DL can be used to build a model where there are many unknown parameters. Finally, as the heart of the DDDAS method, the situation simulation can utilize first-principle modeling to develop the evolving analysis of such data as simulation methods. Contemporary example methods of each approach would be a feature-based unscented Kalman filter for an ML approach, a convolutional neural network (CNN) or recurrent neural network (RNN) for a DL batch analysis, or an Ensemble Kalman filter for data assimilation of the situation assimilation. The emerging challenge and opportunity is how to effectively and efficiently use these different levels of data granularity techniques together for a near-real-time response, especially for unknown dynamic scenarios.

3.2. Contrastive Learning Information Fusion Generative Adversarial Net (CLIFGAN)

Image damage level assessments taken by either satellites or airplanes, of the pre- and post-disaster/action for the OoI are normally available and well registered and matched, such as the xView2 dataset. In Fig. 3, a typical sample of the pre-, and post-disaster (“Hurricane Florence No. 409”) and the damage levels (**red**: no damage, **green**: minor damage, **blue**: major damage) of the structures and buildings from the training set are labeled and identified.

Fig. 3

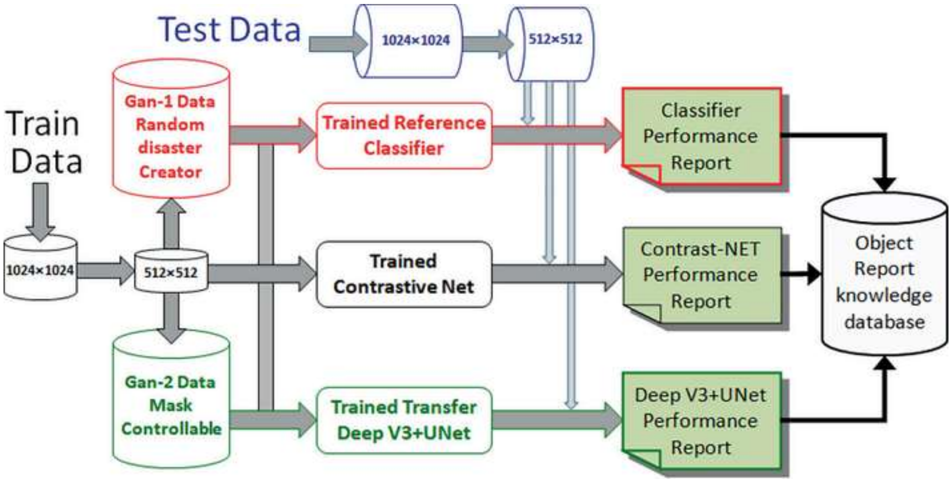
Sample disaster images from the XView2 dataset, Object of Interest (OoI) mask and the corresponding IP2CL representations. Row 1: one sample pair of *pre-disaster image* (left) and *post-disaster image* (right); Row 2: *the ground truths* of the damage severity level, where the background is black, red is no damage, green and blue are minor and major damage (left), and IP2CL (right) for the two images in Row 1



Historically, the CCNY/Air Force Research Lab team led by Drs. Wei and Blasch developed the 2021 leading Overhead Imagery Hackathon (OIH) approach by delivering the novel Contrastive Learning (CL) Information Fusion Generative Adversarial Net (CLIFGAN) system [5] as depicted in Fig. 4. CLIFGAN has the following remarkable components:

Fig. 4

Flowchart of CLIFGAN


Test Data
1024×1024
512×512
Train Data
1024×1024
512×512
Gan-1 Data Random disaster Creator
Gan-2 Data Mask Controllable
Trained Reference Classifier
Trained Contrastive Net
Trained Transfer Deep V3+UNet
Classifier Performance Report
Contrast-NET Performance Report
Deep V3+UNet Performance Report
Object Report knowledge database

(1) Providing imager reduced resolution from 1024 × 1024 pixels to 12 × 512 pixels to facilitate laptop computing;

(2) Innovating a generative adversarial network that enhances the training data by ~30%;

(3) Utilizing a CL-based classifier that highlights Siamese network differences in the pre- and post-disaster images; and

(4) Developing committee voting information fusion using the post-disaster images with transfer learning and DeepLabV3Plus.

Among all the results and developments, the CLIFGAN system results in the highest F1 score of 0.674, whereas previous fusion methods [25] only had an F1 score of 0.664. After comparing some of the submissions from the xView2 challenge winners [26], the top score was F1 of 0.617. Thus, the CLIFGAN-based approach was superior to the previous methods when addressing OIH image exploitation efficiency. After checking many of the XView2 submissions, CL [27] contributed to the exploitation methods with the two matching images to label objects of Interest. For example, the disaster scene buildings are labeled for pre and post-disaster. As a consequence, CLIFGAN ranked in the top three in the Overhead Imagery Challenge [28].

3.3. Contrastive Learning in HADR

CL seeks to label common data (e.g., imagery) with a similarity score while attributing higher weights for dissimilar identities. following equation to define the contrastive loss function of a CL embedding network [29]:

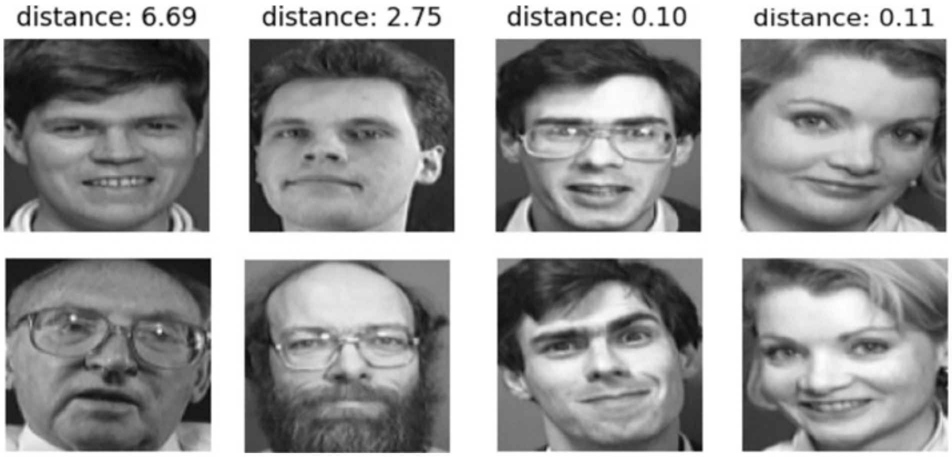
$$\begin{aligned} L_{CL} &= 1\left[y_i = y_j\right] \|\theta\left(x_i - x_j\right)\|_2^2 \\ &+ 1\left[y_i \neq y_j\right] \max\left(0, \epsilon - \|\theta\left(x_i - x_j\right)\|_2^2\right) \end{aligned} \quad (1)$$

where **1**[·] is the normal indicator function, x_i and x_j are two samples with associated labels y_i and y_j that will be contrasted, θ is the encoding contrastive deep net, a 3-layer Convolutional Neural Network (CNN) in our work, and ∊ is a hyper-parameter indicating the lower-bound distance for samples from different classes.

For the given training datasets, similar samples are created by applying data augmentation, such as performing image translation, flipping, shearing, rotation, and color jittering, to make more samples of the same label (y_i's) available, hence generating sufficient training/testing data to achieve CL training effectively. The embedding net θ learning by minimizing Eq. (1) can effectively pull objects of the same labels closer while pushing away those of different labels. As an intuitive visualization illustrating the power of CL, in Fig. 5 , after training a 2-D embedding net the distances among 4 pairs of images are shown, where it can be observed that the distances of different people in the first two columns have large distances; whereas the same people in the last two columns, even with glasses (Column 3) and different orientations and shades (Column 4), can still yield a small distance of 0.10 and 0.11, respectively.

Fig. 5

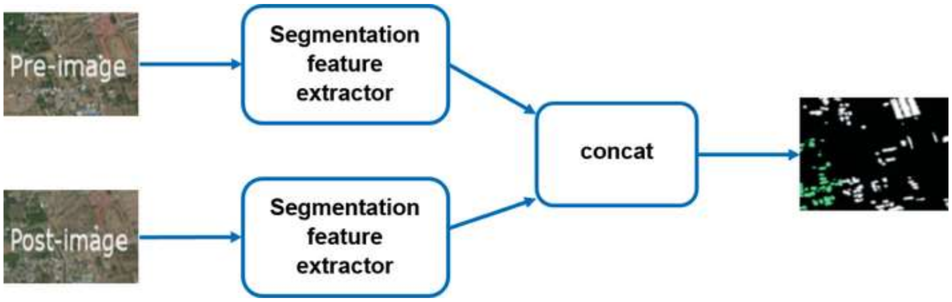
High-dimensional data are embedded by contrastive learning (CL) to low-dimensional data. The numbers in each column represent the Euclidean distances between each pair of images in the 2D embedding space. The CL embedded different people in the first two columns with large distances, while the same persons shown in the last two columns have very small distances



CL can thus find a broad spectrum of applications. In our CLIFGAN system, the Siamese network architecture as shown in Fig. 6 was employed. As is the case in Siamese networks, parameters of the same CNN are shared between the two branches. The features extracted from pre- and post-images are concatenated for damage label prediction.

Fig. 6

Siamese network used in contrastive learning approach for pre- and post-disaster images


Pre-image
Segmentation feature extractor
concat
Post-image
Segmentation feature extractor

The large scene covered by the XView2 imagery dataset challenges the CL which is commensurate for all HADR and DA SAA. For example, when using techniques in [30], the OoI only accounts for a small fraction of the whole scene, where more than 90% of the areas in the xView2 dataset are background pixels that do not contribute to the damage level assessment. Therefore, using CL may not result in HADR/DA solutions that support the use needs.

Using the recently popularized DL *attention* mechanism [31], computing processing power focuses on relevant pixels in the scene of interest. Transformers changed the field of Computer Vision and Natural Language Processing/Understanding [32]. Thus, there is a need to bring OoI attention to CL as evidenced in our CLIFGAN's relatively low classification performances (0.674), despite already being the best among all peers.

More and more research on damage assessment is based on synthetic aperture radar (SAR). In [33], the authors propose a statistical texture feature, G0-para, for SAR building damage assessment in various polarization modes. It is based on the statistical distribution of G0-para in SAR images to distinguish between collapsed and intact buildings. The proposed approach overcomes the limitations of traditional texture features by considering the statistical distribution of speckles in SAR images.

To improve the performance of the damage detection, fusion, and ensemble learning are used. Adriano et al. [34] provided a careful analysis of the building damage severity level recognition using data fusion and ensemble learning algorithms to achieve fast damage mapping tasks. The authors investigated three robust ensemble learning classifiers to achieve building damage recognition from SAR and optical remote sensing datasets. The proposed mapping framework successfully classified four levels of building damage with an overall accuracy of over 90% and an average accuracy of over 67% on the data set of the 2018 Sulawesi earthquake and tsunami. However, the paper does not provide detailed information about the data set and accessibility and what algorithms other than canonical correlation forests are used in their ensemble learning. Likewise, they do not define the overall accuracy and the average accuracy.

3.4. Probabilistic Disaster Modeling: IP2CL

IP2CL utilizes an image Z from a pre-image X and a post-image Y, which are viewed as random variables in [35]. The relations between the X and Y are captured by the conditional probability functions:

$$P\left(Z_0|X_0,Y_0\right)=Y_0 \tag{2}$$

$$P\left(Z_1|X_1,Y_1\right)=Norm\left(Y_1-X_1\right) \tag{3}$$

and by using an image indicator variable as subscript c: where 0 is for background and 1 for OoI; then X_c and Y_c correspond to the pre- and post-disaster image with values in the range of [0,1]. The new random variable Z_c takes its value from the same range of X and Y, then Norm(w) is defined by the following equation:

$$Norm\left(w\right)=\frac{w}{\max\left(\mathrm{w}\right)-\min(\mathrm{w})} \tag{4}$$

which normalizes w to the range of [0, 1]. The three steps as described in Eqs. (2–4) produce a new image Z with its background from the post image and its OoInormalized from the difference in the OoI region. Using the new image Z emphasizes the OoI pixel-wise differences. As shown in Fig. 3, the lower right panel illustrating a sample IP2CL frame, we observe that the OoI regions, including the buildings, are different from the background in terms of color variations. In summary, Z brings several benefits over the original signals X and Y:

(1) *Efficiency in data and computing*: reducing two color images that need to be processed to one significantly improves the computation and storage efficiency. By contrast, the CL employed in CLIFGAN requires much more RAM and GPU time for handling big sets of pre- and post-disaster images.

(2) *Flexibility in choosing DL methods*: using the red, green, and blue (RGB channels, the data to be fed to the deep nets are ordinary color images, thus more available deep nets trained on large data sets such as ImageNet and/or PASCAL visual objects can be employed. By contrast, far fewer nets can readily facilitate 6-channel inputs for data classification, in both image/object classification and semantic segmentation.

(3) *Saliency on Objects of Interest*: as described in previous paragraphs, using Z stresses OoI regions with larger values and variances, hence facilitating the training process.

(4) *Emulative of human annotation*: expert users manually label the pixel-wise difference in OoI for the damage level assessment.

(5) *Effectively uses of the context information*: post-disaster images present the crucial contextual information for OoI identification, as suggested by Eq. (2). The situations after natural disasters normally contain more valuable information regarding the OoI compared with those prior to the disasters. Directly using the pre-disaster image or removing the background, however, produced much worse classifications in our study. Consequently, we should use the post-disaster/action information, as dictated by Eq. (2), for valuable contextual information.

4. IP2CL Applications and Experimental Results

The IP2CL representation will be applied to two different scenarios of practical importance in this section: the first is to yield global semantic segmentation as originally conducted for HADR applications: obtain the damage severity levels for the whole image for all OoIs thus helping responders to make responses relative to the damage estimate. The second is for patch-based classification: in HADR or DA-SAA, the OoIs are of different importance, *e.g.*, hospitals or schools in HADR applications in DA are obviously of more importance, the local damage severity level evaluation for these regions of special importance should be conducted early with desirable performance. Before presenting the two approaches, the common problem of unbalanced data in the training set is tackled by two generative algorithms.

4.1. Two Generative Adversarial Networks for HADR Applications

Pre- and post-disaster pictures are crucial in existing state-of-the-art models to estimate the infrastructure damage level from overhead imagery. To remedy the class unbalance problem in the xView2 dataset, the pix2pix GAN is used. Given an input image, the generator component of a GAN model generates a transformed version of the input image. On the other end, the discriminator component in the GAN model, given an input image and a paired image (real or generated), determines whether the paired image is authentic or false, yielding a generator model to fool the discriminator model and minimize the loss between the generated image and the desired target image as well. The overall architecture permits the Pix2PixGAN to be trained over a diversified set of image-to-image conversion tasks, e.g., creating a post-disaster image corresponding to a pre-disaster one. In our work, we developed two novel Pix2PixGANs, named GAN-1 and GAN-2, to augment data with the desired size and nature.

Mask Controllable GAN: GAN-1

GAN-1 is a Pix2Pix GAN that takes a four-channel image as input and generates a post-disaster image as output. The four channels include the red, green, and blue (RGB) channels of the pre-disaster image together with one additional channel with the post-disaster labels. The GAN-1 outputs a generated post-disaster image with the RGB channels. GAN-1 can be used to produce as many post-disaster images as desired to balance the class labels. Specifically, GAN-1 works as follows. First, GAN-1 was trained on the xView2 images, which were reduced to the size of 512 × 512. To accelerate the training process, the images were resized and normalized to 256 × 256 pixels during training. The GAN can facilitate varying scenarios of disaster types and damage levels, be it earthquakes, fire, volcanos, hurricanes, etc. To augment data with desired labels and types, we use available pre-disaster images in the xView2 dataset and the associated labels of post-disaster images as inputs. The labels were then adjusted to obtain the desired damage severity level together with the pre-disaster images as input for the GANs. Finally, the resultant images were fed to the GAN as inputs to produce the post-disaster images of expected labels. The diagram of GAN-1 and 4 sample outputs corresponding to the 4 damage levels are illustrated in Figs. 7 and 8, respectively, where it can be observed that the generated images are visibly similar to their assigned labels. By using GAN-1, we can freely generate new pre- and post-disaster images as needed. For instance, in the original xView2 dataset, there are very few representations of ”major” and “total” damages, by using GAN-1, we can simply generate more of them to populate the training datasets such that the number of images for each of the four classes is roughly the same, thus effectively resolving the data unbalance problem, which is common for most practical HADR tasks.

Fig. 7

Flowchart of the GAN used to improve the data balance in the training dataset

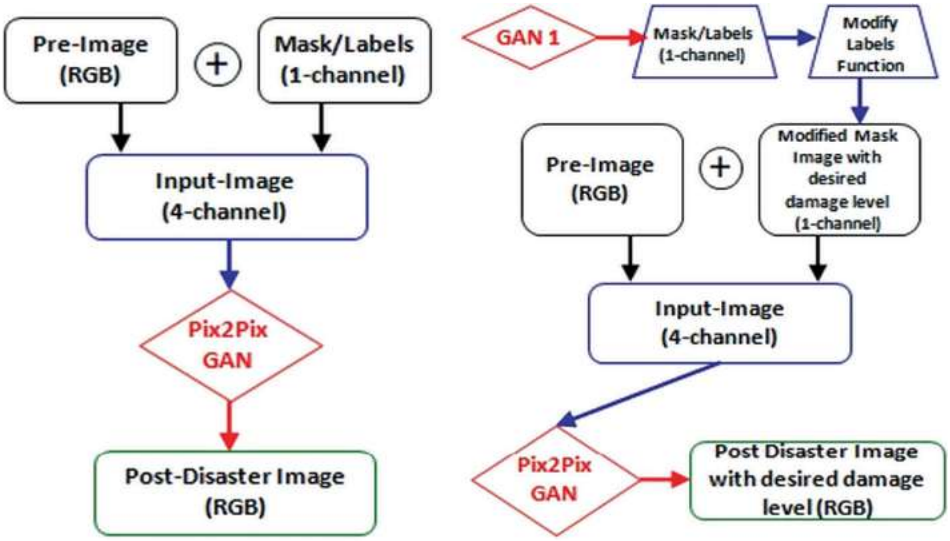
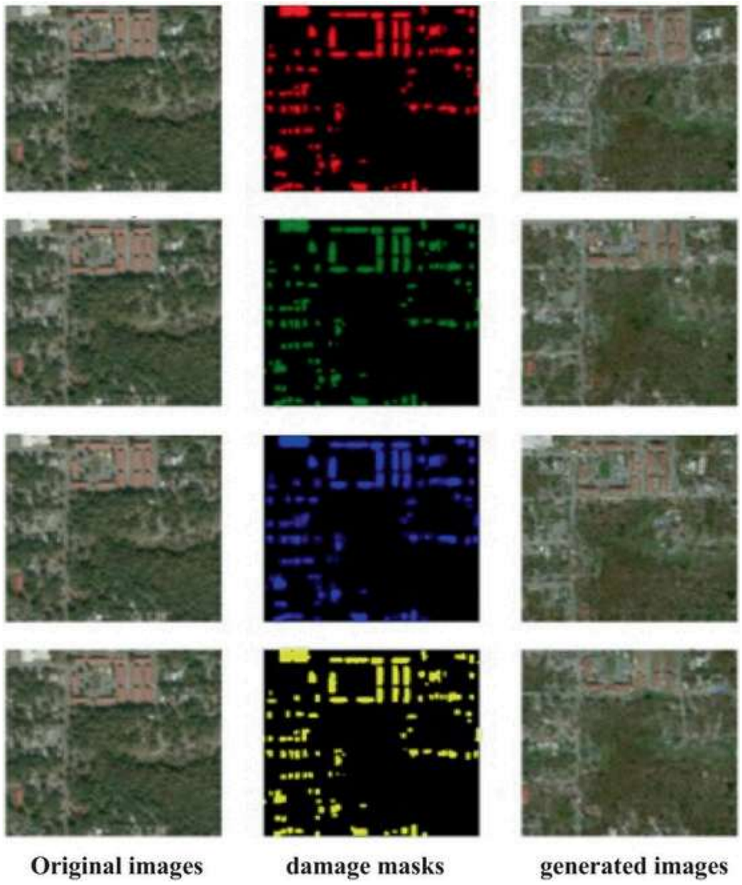


Fig. 8

Left column: original images; Middle column: controllable masks, regions in red, green, blue and yellow color correspond to infrastructures with no-, minor-, major- and total- damage, respectively, caused by the disaster; Right column: images generated by GAN-1

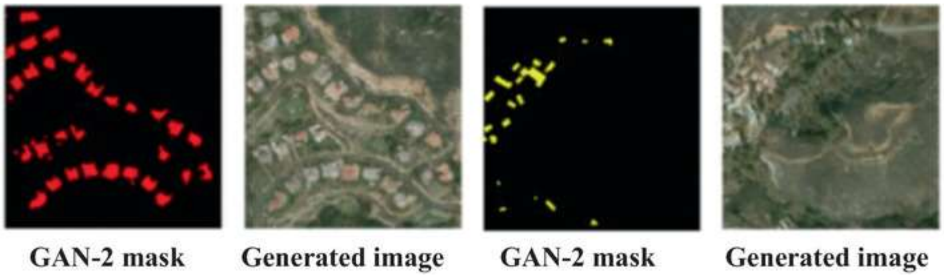


Random Disaster Creator GAN: GAN-2

GAN-2 is a Pix2Pix GAN similar to GAN-1 with the flowchart shown in Fig. 7, it differs from GAN-1 only in the input masks: GAN-2 uses pre-disaster images and *randomized* damage-level masks as input, and outputs post-disaster masks and images. The generator of GAN-2 takes as input a four-channel image consisting of a pre-disaster image, combined with a randomly drawn post-disaster mask with an arbitrary severity level. It then generates a four-channel image output, containing a three-channel RGB post-disaster image and its corresponding post-disaster mask. GAN-2 shares a similar pipeline with GAN-1 with some differences. It starts with a pre-processing stage that is the same as GAN-1. The Pix2Pix part of GAN-2 was trained by a four-channel image consisting of a pre-disaster image and its real post-disaster mask and generating an output four-channel image made up of a post-disaster image and the associated post-disaster mask. After training the Pix2Pix generator, similar to GAN-1, we developed a function to vary the severity level of the pre-disaster masks. Next, a damage severity level was randomly drawn to yield a new post-disaster mask label, and the label was combined with its pre-disaster image to give rise to the input images. After the input four-channel image was generated, it was given to the Pix2Pix GAN-2 generator to generate an output four-channel image, which can be split into a generated mask as well as its corresponding post-disaster image of randomized damage level. Two sample masks and images generated by GAN-2 are illustrated in Fig. 9, as can be observed that the randomly generated images and masks are visually similar to the real ones. We developed GAN-2 to ensure the randomness of the GAN-generated images. The crucial difference between GAN-1 shown in Fig. 8 and GAN-1 in Fig. 9 is that: the masks in GAN-1 (middle column of Fig. 8) are provided by a human operator, whereas the masks of GAN-2 (columns 2 and 4 of Fig. 9) are randomly generated by the deep net.

Fig. 9

Two Masks (same color code as Fig. 2) and images (Right column) generated by GAN-2



4.2. IP2CL in Semantic Segmentation

To handle images that cover a wide expanse (e.g., the leftmost panel of Fig. 8), we need to use a semantic segmentation algorithm to create dense damage labels for each pixel. It is nontrivial, however, to choose a *single* best segmentation algorithm. A strategy to address this issue is through robust fusion of multiple algorithms [6]. In our study, we found that the *U-Net* combined with *transfer learning*, as reviewed in Sect. 2, yields exceedingly powerful image segmentation.

In addition to CL, we observed that the combination of U-Net and transfer learning serves as another effective method to achieve data and computation-efficient classification for the HADR classification task in this work.

Using Matlab©'s *segnetLayers* methods or Python's segmentation models over the xView2 dataset, the recent DeepLab v3+ network generates competitive performance. The backbone of this U-Net-based architecture is chosen to the ImageNet-trained *Resnet* or *VGG*, where the crucial visual knowledge gathered from the large ImageNet could be handily migrated to the HADR images in the xView2 dataset. Within this U-Net architecture based on DeepLabv3+ and DL, depth separable convolutions are employed in the Atrous spatial pyramid pooling and decoding sub-nets. For data augmentation, we used various popular operations during training including scaling, rotations, flipping, and shearing.

Equipped with the aforementioned IP2CL techniques, the data fed to a deep net are essentially a set of normal color images highlighting the OoIs. Leveraging transfer learning methods using pre-trained deep nets, e.g., ResNet, AlexNet, and VGG, as feature generators, and exploiting DeepLabV3+ as the backbone for U-net-based pixel-wise segmentation, we conducted a dense pixel-wise semantic segmentation to assign all pixels in the HADR images to different labels. Table 1 presents the semantic segmentation performances resulting from our 2021 Overhead Imagery Challenge submission as well as the new IP2CL-based method. To paint a relatively complete picture of each method, the Precision, Recall, and F1 scores are provided. Furthermore, the processing time in terms of milliseconds (ms) of one image taken by each method is also reported, which is measured in a computer with an Intel i9-12,900 CPU, 64 GB RAM, and NVIDIA RTX 3090 GPU. The size of each net is also given. The IP2CL expressed more OoI information than post-only (with F1 of 66.4): just using the post-disaster images as inputs to the transfer learning and DeepLab v3+ based U-Net, and CL (with F1 of 67.4), resulting in better classification performance (0.697) at about the same runtime.

Table 1

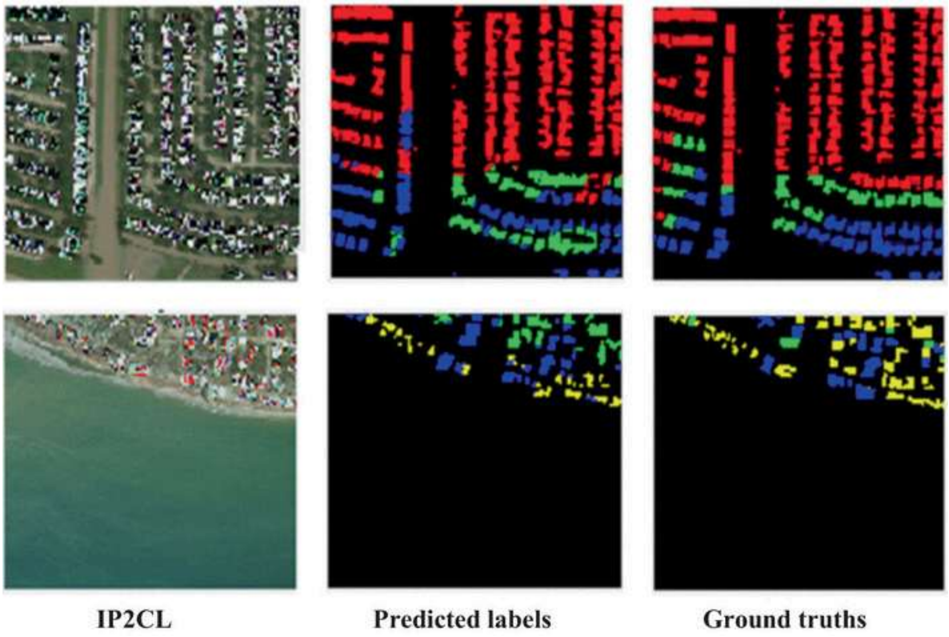
Performance of pixel-wise segmentation: precision, recall and F1 scores, run time (ms: millisecond), and net size (MB: Mega Bytes)

Algorithm	Method 1	Post-only	CCNY CL	IP2CL
Precision	64.2	65.6	67.2	**69.1**
Recall	59.3	67.2	67.6	**70.4**
F1 score	61.7	66.4	67.4	**69.7**
Run time (ms)	660.9	**102.6**	520.4	103.2
Net size (MB)	441	**9.7**	40	**9.7**

Furthermore, as the IP2CL is only an ordinary color image, like the *post-only* algorithm, the deep nets are considerably smaller than the CL-based method, i.e., 9.7 MB versus 40 MB. Method 1 is the result of reproducing the winning method in the original xView2 challenge. In consequence, the IP2CL-based segmentation method is more efficient in terms of net size and computational demands than our original methods. Two segmentation results are depicted in Fig. 10, the F1 scores of which are respectively 0.87 and 0.79.

Fig. 10

Two sample results of dense segmentation for each image pixels using IP2CL. Left column: IP2CL image; Middle column: pixel-wise labels predicted by IP2CL semantic segmentation; Right column: Ground truths (the color code scheme: red: no damage, green: minor damage, blue: major damage, yellow: total damage). The F1 scores by IP2CL-based segmentation algorithm for the first and second row are 0.87 and 0.79, respectively



A unique benefit of this transfer learning and information fusion method is its robustness to the size and behavior of the train data, while all other methods we have tried cannot effectively deal with small training data sizes and data augmented by GAN-1 and GAN-2. The two reproduced xView2 methods and our own CLIFGAN methods yielded considerably worse performance in these cases, the overall F1 scores were all reduced to less than 0.5, which were likely due to the use of Siamese nets that are seriously impacted by the even slight difference in GAN data and the inadequate training size. More investigations on the robust use of CL will be thoroughly investigated in the future. However, the transfer and fusion method demonstrated desirable resilience by yielding acceptable results. To stress-test the behavior of transfer learning, as shown in Table 2, with 10% of the xView2 training data, namely, merely 280 training overhead images are used to fine-tune the deep net, our methods can still deliver fairly valuable segmentation and classification performances. When using GAN-1 data to augment double the number of training sets to 560 images, the classification F1 score for 10% of data is improved by 2–3%. The results demonstrated that GAN-1 can help better increase the size of the training data. Also, note that a user of the CLIFGAN system can command the damage severity labels in GAN-1 to tackle the serious class unbalance trouble. The data created by GAN-2 does not seem to come from the same statistical dataset, thus when doubling the training set by adding 280 GAN-2 images reduces the performance. It can also be observed that although GAN-2 indeed creates visually relevant, or novel, images as shown in Fig. 11, these images enriched the overhead imagery dataset; however, its blind use in classification is problematic, as summarized in Table 2.

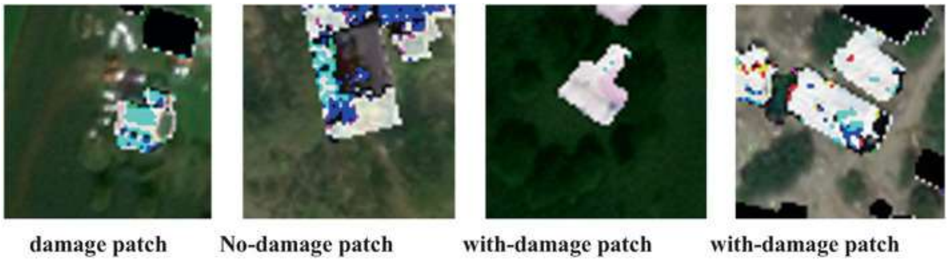
Table 2

Performances (precision, recall, and F1 scores) of cases with reduced data size and augmented samples (doubling the original training data size) by GAN-1 and GAN-2

	CCNY CL			IP2CL		
	Precision	Recall	F1	Precision	Recall	F1
Full data	67.2	67.6	67.4	69.4	70.1	69.7
10% data	55.9	58.1	57.0	56.8	60.5	58.6
10% data + GAN-1	59.2	60.4	59.8	61.0	61.6	61.3
10% data + GAN-2	53.9	48.9	51.3	55.4	45.2	49.8

Fig. 11

Four sample patches generated by our procedure used to train the CL net: Left two panels: without damage; Right two panels: with damage



4.3. Patch-Based Damage Level Classification

Since the pixel-wise dense damage classification presented in Sect. 3 was designed for large-area images, there is a need to develop methods that focus on a smaller area (i.e., patch). Examples of HADR/DA applications where a relatively small region requires more accurate results include responding to a school or a hospital, and whether it is damaged and demands prompt assistance. In these important first-responder scenarios, the damage severity level classification for each pixel, which is generally of relatively poor performance (with F1 < 0.7 with all state-of-the-art methods), can be discarded; conversely, where the OoI is located only one damage severity designation should be provided for the whole patch, where better classification performances are demanded and could be achieved if proper method is applied.

There are four different damage severity labels in the xView2 dataset. A major reason for the underperformance of existing classification performances is due to the considerably unbalanced nature of these labels: on average about 95% are background pixels, while for regions with damages, again ~74% are "*no damage*" (label 1), only ~5% with "*total damage*". Many image augmentation techniques were explored, e.g., GAN [36], different class importance, in an effort to reduce the data imbalance, but in vain. The IP2CL best F1 score by IP2CL is still at 0.70. Using the patch-based method can considerably boost the classification performance:

(1) Instead of categorizing damage severity levels from "*minor*" to "*total*", a binary label "*no damage*" and "*with damage*" is adequate, which avoids challenges associated with an exceedingly small number of "*major*" and "*total*" damage levels.

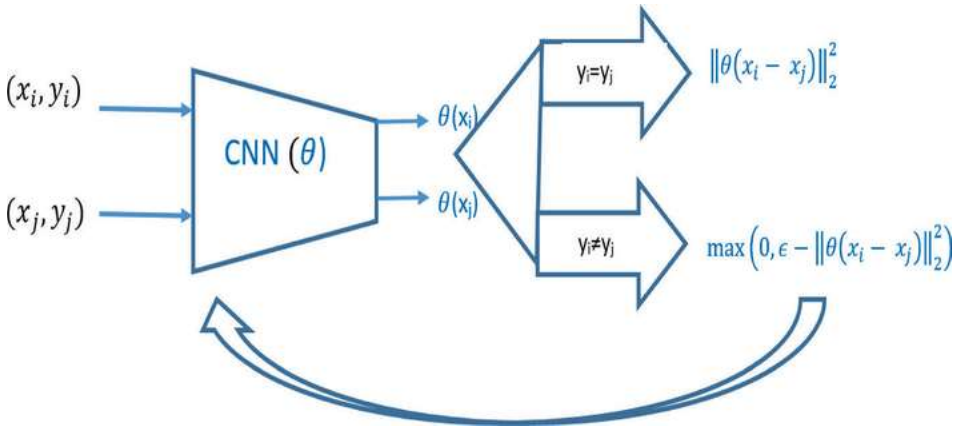
(2) When creating patches with "no damage" labels and "with damage" labels, we collected the statistics of different labels by varying the patch size in the range of 128 × 128, 100 × 100, 64 × 64, etc., and different OoI types, based on which the two thresholds (δ_1, δ_2) are obtained to determine the "with damage" and "no damage" types.

For a patch *p*, the "*no damage*" or "*with damage*" class label is decided if its maximal OoI has a size larger than $\delta_1/\delta_{2,}$ respectively, afterward the OoI regions in this patch belonging to the other class are eliminated to make sure the current patch only has one the unique label—the winning label to be identified—or else the OoI of the other class inside patch will confuse the upcoming training procedure, thus boosting the classification performance. The patch-based method generates patches with the "*with damage*" and "*no damage*" types that are roughly similar in size to nurture more effective classification in the future [7]. Figure 11 showcases two generated sample patches. After intensive offline processing of the xView2 datasets, with 64 × 64 patches and setting (δ_1, δ_2) to (0.12, 0.04), 21 k and 20 k training patches resulted for the two damage class type sets, as well as a test set for the two classes of 7.1 k and 6.8 k resulting a balanced data set.

After this processing, each IP2CL patch is an ordinary 64 × 64 color image with two labels: *with damage or without damage*. Since the patches in the same class look alike, while those in different classes are supposed to be very different, CL as introduced at length in Sect. 2 can be exploited to encode patches of the same class to a low-dimensional space in close vicinity, and distance those of differing labels. In our simple Siamese network, the share CNN is composed of merely three convolution layers followed by a fully connected layer to encode each 64 × 64 × 3 color image into a2D space by minimizing the contrastive loss dictated by Eq. (1) and illustrated in Fig. 12, the hyper-parameter ϵ is set at 2. To facilitate balanced training of the CLnet, each training data pair (x_i, y_i) and (x_j, y_j) as illustrated in Fig. 12 is drawn with probability 0.5 to come from the training set with y_i=y_j or y_i≠y_j, so that the pulling same labels and pushing of dissimilar labels are performed with roughly the same epochs.

Fig. 12

Diagram of the training process of the contrastive learning: for training pairs (x_i, y_i), (x_j, y_j), the 64 × 64 × 3 image data x are encoded by the siamese CNN to 2-D data $\theta(x)$, the loss function used to re-adjust the CNN differs according to the labels y



The optimizer used in our work is Adam with a learning rate of 5e-4. After training the CL by 50 epochs, which is observed to be adequately stabilized, the trained CNN θ is then employed to encode all training data *x*'s to a 2D space, subsequently a simple Support Vector Machine (SVM) was trained to be the resultant classifier in the training phase, that is,

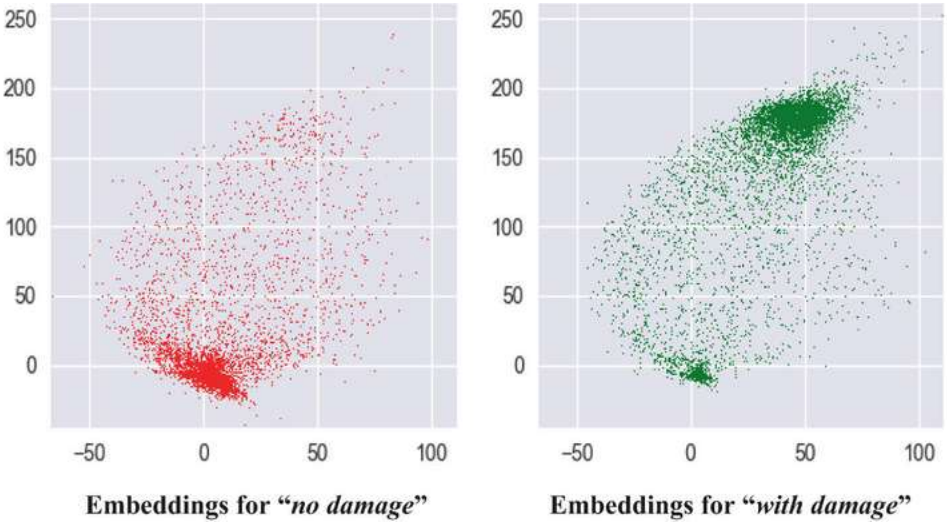
$$y_{pred}\left(x_i\right) = SVM\left(\theta\left(x_i\right)\right)$$

For the given training datasets and the CL net shown in Fig. 4, the predicted label for each patch x_i of the testing dataset can be obtained from Eq. (5). The resultant F1 score for the testing dataset is 95.9 with the confusion matrix $\begin{pmatrix} 0.98 & 0.02 \\ 0.06 & 0.94 \end{pmatrix}$.

As illustrated in Fig. 13, the two different damage classes in the test dataset are partitioned distinctly into mostly separated regions by the CL. We also tested different embedding space dimensionality different from 2, e.g., ranging from 3 to 10, the achieved performances were all not as good as the simple 2-D space by a large margin.

Fig. 13

The 2D embedding achieved by the contrastive learning method for IP2CL-based patches, "no damage" are shown as red points (left) and "with damage" green ones (right)



Due to the IP2CL representation, this patch-based damage severity classification resembles the original deep learning methods based on ImageNet, the readily available deep nets pre-trained on ImageNet can thus be used within the framework of transfer learning. We have used more than 10 different nets from Pytorch's *torchvision* models in our experimental studies, and six of them can deliver F1 scores above 0.90, as listed in Table 3. CL combined with IP2CL stands out as it not only yields the best F1 score (95.9%): better than the attention-based visual transformer by about 1 percentage (95.0%), and all is achieved by a net with a significantly smaller size: 9.5 versus 343 MB. The results are produced by using the ***fine-tuning*** option, not ***feature-extraction***, which yields far worse performances by more than 10%, likely since the IP2CL is considerably different from the images seen in ImageNet, where the features learned from ImageNet are of subpar performances. The Run time of each net to process one test image is also reported, based on the same computer for Table 1. Besides being able to deliver the best performance, it can be observed that the CL-based method delivers the second lowest run time (7.6 ms) only slightly longer than AlexNet (7.1 ms). It is thus a computing efficient approach as well.

Table 3

Patch-based damage severity classification performances, measured by the precision, recall, and F1 scores, using IP2CL, the processing time of one test image, and the net size

Algorithms	Precision	Recall	F1 score	Run time (ms)	Net size (MB)
Contrastive learning	**95.7**	**96.1**	**95.9**	7.6	8.6
Visual transformer	95.0	95.1	95.0	42.6	343
VGG19	94.2	94.3	94.2	26.5	558
ResNet152	93.0	93.0	93.0	32.2	233
ResNet 101	92.6	92.8	92.7	24.2	170
Inception v3	90.3	90.7	90.5	25.3	101
GoogleNet	90.2	90.4	90.3	11.2	23
VGG16	89.9	89.9	89.9	22.8	537
ResNet 18	89.6	89.9	89.8	8.4	45
AlexNet	89.6	89.8	89.7	**7.1**	229
mobileNet-v3	89.4	89.2	89.3	9.7	**6.2**
efficientNet-b7	84.3	84.7	84.5	41.7	258

CL together with IP2CL, therefore, delivered exceptionally valuable data and computing efficacy with excellent performances in this patch-based global image classification task, because the IP2CL intentionally changes the data representation and puts the analysis attention to the regions of special interest to take advantage of the great discrimination prowess of CL in content-based feature extractions.

5. Conclusion and Future Works

This chapter presents a new method within the machine learning and granular computing framework using image-based pre- and post-conditional probability learning (IP2CL) to emphasize the changes in the object of interest (OoI). Enhanced performance was shown on the xView2 dataset using pixel-wise damage classification based on semantic segmentation and patch-based damage classification. IP2CL combined with various dynamic data-driven applications systems augmented with deep learning methods, such as contrastive learning, generative adversarial network, U-Net, and transfer learning, achieved data as well as computationally efficient Situational Awareness and Assessment for Humanitarian Assistance and Disaster Response and Damage Assessment (HADR/DA) applications in both civilian and commercial applications.

In this work, the publicly available dataset xView2 dataset was used to validate our new representation and the corresponding deep-learning approaches in segmentation and classification tasks. At present, the fusion-based transfer learning method for the semantic segmentation case cannot deliver computational and data-efficient results yet. In the future, we will investigate more methods to achieve a more efficient and robust use of contrastive deep learning. Furthermore, other deep nets such as vision transformers [37] and the more recent attention-based nets, self-supervised representation learning methods [38] in contrastive, generative, and contrastive learning approaches, will also be examined to improve the computing and data efficiency. In addition, more re-adjustments and revisions will be examined to expand our method to more applications of crucial and practical importance to HADR and DA [39, 40].

References

1. Munir, A., Aved, A., Blasch, E.: Situational awareness: techniques, challenges, and prospects. AI **3**(1), 55–77 (2022)

2. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**, 9 (2015)

3. Blasch, E., et al.: Machine learning/artificial intelligence for sensor data fusion–opportunities and challenges. IEEE Aerosp. Electron. Syst. Mag. **36**(7), 80–93 (2021)

4. Wei, J., et al.: *Vehicle* engine classification using spectral tone-pitch vibration indexing and neural network. Int. J. Surveillance Monit. Res. Tech., Special issue on Machine learning and sensor fusion techniques **2**(3), 29 (2014)

5. Chen, H., et al.: Dual-tasks siamese transformer framework for building damage assessment. arXiv preprint arXiv:2201.10953 (2022)

6. Wei, J., et al.: NIDA-CLIFGAN: Natural infrastructure damage assessment through efficient classification combining contrastive learning, information fusion and generative adversarial networks. In: Artificial Intelligence in Human Assistance and Disaster Response workshop, p. 6. NeuralPS'21, arXiv preprint arXiv:2110.14518 (2021)

7. Yao, J.T., Vasilakos, A.V., Pedrycz, W.: Granular computing: perspectives and challenges. IEEE Trans. Cybern. **43**(6), 1977–1989 (2013)

8. Liu, H., Diao, X., Guo, H.: Quantitative analysis for image segmentation by granular computing clustering from the view of set. J. Algorithms Comput. Technol. **13**, 1748301819833050 (2019)

9. Bianchi, F.M., et al.: Granular computing techniques for classification and semantic characterization of structured data. Cogn. Comput. **8**, 442–461 (2016)

10. Hu, H., et al.: Perception granular computing in visual haze-free task. Expert Syst. Appl. **41**(6), 2729–2741 (2014)

11. Pedrycz, W., Chen, S.-M.: Granular Computing and Decision-Making: Interactive and Iterative Approaches, vol. 10. Springe (2015)

12. Eugene Khvedchenya, T.G.: Fully convolutional Siamese neural networks for buildings damage assessment from satellite images (2021). arXiv.org

13. Wang, Y., et al.: A two-stage seismic damage assessment method for small, dense, and imbalanced buildings in remote sensing images. Remote Sensing (2022)

14. Teymoor, S., et al.: BDD-Net+: a building damage detection framework based on modified coat-net. IEEE J. Selected Topics Appl. Earth Observ. Remote Sens. **16**, 4232–4247 (2023)

15. Ge, J., et al.: Rapid identification of damaged buildings using incremental learning with transferred data from historical natural disaster cases. ISPRS J. Photogramm. Remote Sens. (2023)

16. Zhu, J., Isola, P., Efros, A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2242–2251. Venice, Italy, 2017 (2017)

17. Blasch, E., Darema, F.: Introduction to the DDDAS2022 conference infosymbiotics/dynamic data driven applications systems. In: International Conference on Dynamic Data Driven Applications Systems. Springer (2022)

18. Isola, P., et al.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2017)

19. Siddique, N., et al.: U-net and its variants for medical image segmentation: a review of theory and applications. IEEE Access **9**, 82031–82057 (2021)

20. Çiçek, Ö., et al.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: 19th International Conference on Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016. Athens, Greece, October 17–21, 2016, Proceedings, Part II 19. Springer (2016)

21. Zhuang, F., et al.: A comprehensive survey on transfer learning. Proc. IEEE **109**(1), 43–76 (2020)

22. Ray, P.P., ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet of Things and Cyber-Phys. Syst. (2023)

23. Chen, H.-M., et al.: Targeted adversarial discriminative domain adaptation. J. Appl. Remote Sens. **15**(3), 038504–038504 (2021)

24. Blasch, E.P., et al.: Handbook of Dynamic Data Driven Applications Systems: Volume 1. Springer Nature (2022)

25. Blasch, E., Lambert, D.A.: High-Level Information Fusion Management and Systems Design. Artech House (2012)

26. Wei, J.: Small moving object detection from infra-red sequences. Int. J. Image Graph. **13**(03), 1350014 (2013)

27. Zhao, X., et al.: Contrastive learning for label efficient semantic segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (2021)

28. CCNY. CCNY places 3rd in international overhead imagery hackathon (2021)

29. Hadsell, R., Chopra, S., LeCun, Y.: Dimensionality reduction by learning an invariant mapping. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). IEEE (2006)

30. Blasch, E., et al.: Summary of methods in wide-area motion imagery (WAMI). In: Geospatial InfoFusion and Video Analytics IV; and Motion Imagery for ISR and Situational Awareness II. SPIE (2014)

31. Vaswani, A., et al.: Attention is all you need. In: Advances in Neural Information Processing Systems (2017)

32. Liu, Z., et al.: Swin transformer: hierarchical vision transformer using shifted windows. arXiv preprint arXiv:2103.14030 (2021)

33. Qihao Chen, H.Y., Li, L., Liu, X.: A novel statistical texture feature for sar building damage assessment in different polarization modes. IEEE J. Selected Topics Appl. Earth Observ. Remote Sens. **13** (2019)

34. Adriano, B., et al.: Multi-source data fusion based on ensemble learning for rapid building damage mapping during the 2018 Sulawesi Earthquake and Tsunami in Palu, Indonesia. Remote Sens. **11**(7), 886 (2019)

35. Wei, J.: Video content classification based on 3-D Eigen analysis. IEEE Trans. Image Process. **14**(5), 662–673 (2005)

36. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, vol. 27 (2014)

37. Dosovitskiy, A., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)

38. Jaiswal, A., et al.: A survey on contrastive self-supervised learning. Technologies **9**(1), 2 (2020)

39. Wei, J., et al.: Vehicle engine classification using normalized tone-pitch indexing and neural computing on short remote vibration sensing data. Expert Syst. Appl. **115**, 276–286 (2019)

40. Blasch, E., Ravela, S., Aved, A.: Handbook of Dynamic Data Driven Applications Systems. Springer (2018)