On the generation and assessment of synthetic waste images

Nick Tsagarakis
Institute of Computer Science
Foundation for Research and Technology Hellas
Heraklion, Greece
nikolats@ics.forth.gr

Michail Maniadakis
Institute of Computer Science
Foundation for Research and Technology Hellas
Heraklion, Greece
mmaniada@ics.forth.gr

***Abstract*—In contemporary waste recycling, the assistance of autonomous robotic systems, equipped with machine learning capabilities, has become crucial for the identification and sorting of recyclable materials. The evolution of computer vision applications, reliant on machine learning, heavily hinges on extensive datasets employed for training intricate deep neural network models. Recently several works from various fields have explored techniques that facilitate the generation of big synthetic datasets starting from an initially limited set of images. This paper proposes a novel approach for generating synthetic waste images, which involves two main steps. The first regards the use of a neural network to implement a "critic" that can evaluate how realistic, synthetic images of recyclable objects may be. The second involves applying multiple random elastic deformations to images of individual recyclable objects to generate a large number of new appearances of the given objects. The critic evaluates the generated images, gauging their realism through a confidence score. We employ a rigorous confidence threshold to identify synthetic images with a notably realistic appearance. These individual object images are then utilized to craft composite synthetic images depicting multiple recyclable objects on a conveyor belt transporting recyclable waste in an industrial setting. The above summarized process facilitates the creation of expansive artificial datasets crucial for training neural networks. The efficacy of these datasets is assessed by examining their impact on the performance of trained detection models when applied to previously unseen and challenging industrial images. The obtained results show that the use of the synthetic datasets leads to better classification models in terms of both precision and accuracy, motivating more research in the field of artificially generated datasets.**

***Index Terms*—Synthetic Images, Generator, Discriminator, Waste Categorization.**

I. Introduction

Plastics are everywhere, yet a significant portion is utilized only once before being discarded, leading to environmental pollution and the loss of a valuable resource for the economy. According to Organisation for Economic Co-operation and Development estimates, merely 9% of the value of plastic packaging material remains within the economic cycle, with the majority being dumped to a landfill after a brief first-use.

The recovery and recycling of packaging materials from end consumers play a pivotal role in the growing popularity of the circular economy model. Thermoplastics, a category widely employed in packaging, are suitable for recycling due to their ability to be melted and reshaped multiple times while retaining their initial properties. Among thermoplastics, polyethylene terephthalate (PETE or PET) stands out as one of the most recyclable and most-used material in packaging applications, making it a primary focus for Material Recovery Facilities (MRFs).

In recent years, many MRFs have taken specific measures to enhance their productivity. This aims to not only mitigate environmental pollution but also boost profits by selling recovered plastics in the secondary market. To accomplish this goal, along with human resources and manual labour, MRFs have incorporated robotic arms equipped with advanced computer vision systems to conduct material recovery and boost processing capabilities. Detecting and categorizing recyclables in real-time as they traverse an industrial conveyor belt presents a considerable challenge for computer vision units, especially due to the high volume of congested material flow, and the fluctuations of light that may occur among the light source and the transparent or white surfaces of recyclables.

The prevailing strategy involves deploying deep neural networks trained to identify and categorize recyclables using extensive datasets comprising of images depicting waste on the conveyor belt. However, annotating these "belt-images" for training is a laborious and time-intensive process, demanding substantial manual image processing resources.

To address the labor-intensive image annotation process, we explore synthetic data generation methods, enabling the creation of large, automatically annotated datasets. In particular, we start by using single-object images that have been segmented from original belt-image through manual annotation (Figure 1). The method proposed in the current work is then utilized to generate multiple new/synthetic images of the initial object. These synthetic images can subsequently be overlaid on other belt-images, resulting in the creation of a sufficiently diverse dataset. Our focus is on implementing an easy-to-use, cost-effective approach capable of generating numerous synthetic images that introduce variability to the dataset features. Utilizing this augmented dataset for training deep neural networks is anticipated to enhance the efficiency of the resulting model compared to one trained with the original, limited-size dataset.

This work is financed by the European Union's "Horizon Europe" Research and Innovation program under project RECLAIM GA: 101070524.

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00184

(a) A belt-image.

(b) A single PETE object-image.

Fig. 1: An example of a belt-image taken in the MRF (left) and a segmented PETE object-image (right).

The present work adopts the generator-discriminator approach to develop new, realistic, artificial representations of recyclable objects. In short it consists of two primary steps, leveraging neural networks to enhance the realism of recyclable object representations. The first step involves the generation of new, artificial representations of recyclable objects by using random elastic deformations applied to individual object images to significantly diversify their appearances. At the second step, a neural network is trained to act as a critic that evaluates the generated images, assigning confidence scores to gauge their realism.

After setting a threshold value for confidence score, the chosen realistic individual object images are combined to craft composite synthetic belt-images. These composite images simulate a conveyor belt scenario in an industrial setting, where recyclable waste is transported for processing. The aim is to generate datasets that accurately represent the complexities of a real-world recycling environment. By utilizing realistic object images, our approach ensures that the resulting synthetic images maintain authenticity and mimic the challenges posed by the dynamic industrial setting of a MRF. Moreover, the current study explores the effect of the threshold that is set to the confidence score of the assessed artificial object images that are utilized in synthetic datasets, on the efficiency of resultant detecting models.

II. Literature review

The recent accomplishments in generative artificial intelligence have redirected some focus from data collection to data generation. Generative AI models can produce text, images, or other media by assimilating the patterns and structure of their input training data and subsequently generating new data with similar characteristics.

Generative Adversarial Networks (GANs) are a particularly popular approach for generating synthetic images [1]. At the core of GANs lies a generative model and a discriminative model that are tuned simultaneously through competitive training. The generative model tries to generate data samples that resemble real data, while the discriminative model aims to distinguish between real and generated samples. The adversarial nature of training fosters a competitive dynamic, leading to the refinement of both models over time.

The original GAN architecture has undergone numerous enhancements and variations to address challenges and improve performance. Notable architectures include Deep Convolutional GANs (DCGANs) [2], which introduced convolutional layers for image generation, and Wasserstein GANs (WGANs) [3], which modified the training objective to mitigate mode collapse and improve stability.

Another popular approach is generative Diffusion Models which learn intricate patterns and structures from input data, allowing them to generate new data with characteristics akin to the original [4]. Typically, Diffusion Models are used to transform noise into data through an iterative diffusion process. A third alternative is Variational Autoencoders (VAEs) which are able to generate new data by adopting a probabilistic approach [5]. In short, they consist of an encoder that maps input data to a probabilistic latent space, and a decoder that generates data from samples in this space.

Despite their successes, AI generative approaches face challenges such as mode collapse, training instability, difficult and laborious process of training neural networks, demand of high computational resources and finally the creation of biased outputs when the training data are limited.

The current work borrows from GANs the idea of implementing a discriminator module to assess the realism of artificial images. This approach offers a systematic mechanism to filter out images that lack realism and retain those closely resembling real waste objects.

As an alternative to generative AI methods, one may consider object deformation techniques to generate new waste-object instances. Ideally, geometric deformation can be defined in the 3D domain, which is reflected as a perturbation on the surface grid. The use of the grid provides the means to map perturbation into the whole solution domain. This is typically implemented by the mesh deformation approach. For mesh deformation, the interpolation and the spring analogy scheme [6], are widely used. However, besides being computationally expensive, 3d mesh deformation may frequently result in grid crossing and negative volumes. Moreover, 3D mesh deformation is particularly challenging to apply on waste images where the surface of the objects is unknown and their appearance can be significantly altered by dirt.

Turning to the 2D domain, image deformation techniques such as guided warping have proven to be valuable tools for the production of visual effects [7]. Most image warping techniques rely on the idea of treating the image domain as an elastic membrane, capable of deformation within specified constraints while maintaining a shape that is both natural and regular [8]. Grid based image deformation has been applied widely in several domains such as elastic image registration [9] and recently for data augmentation [10], [11].

Similarly, in the current work grid deformation is applied on single waste-object images to generate new appearances of the given objects, which are then used to create multiple synthetic

belt-images that are employed to train a waste identification and categorization Mask R-CNN model.

III. Methodology

As mentioned above, the approach proposed in the current work is inspired from Generative Adversarial Networks (GANs), that use(i) a generator neural network trained to generate random synthetic images and (ii) a discriminator neural network trained to distinguish between real and synthetic images, effectively discerning authentic from fabricated images. These two models engage in a competitive process, progressively enhancing their performance over time. When training GANs people face several issues with one of them being the time training takes to get a fully functional module [12]. In particular, the training of the GAN models requires a lot of resources that increase exponentially with the size of the processed images [13]. Given that most resources are devoted for generator training, to minimize resource consumption, we propose to substitute the training of the generator with a much simpler algorithmic procedure that employs random geometric transformations on the original image to randomly create new instances of the given object. The much lighter computational procedure that regards the training of the discriminator is preserved in the present work, similar to the GAN architecture. The proposed approach is graphically illustrated in Fig 2 and is described in more detail bellow.

A. Geometric deformation algorithm

We start by presenting the geometric deformation procedure applied on a limited collection of initial real object-images to generate a set of new synthetic object-images. This technique aims at applying geometric transformations to images by applying a grid of movable vertices over the image and utilizing inverse distance weighting interpolation [14] to make pixel-wise color adjustments. It assumes that the displacement of a grid point influences only the neighboring grid points within its vicinity. A weighting function governs the appearance within the grid, which is directly tied to the distance between the moved interior point and boundary points.

A randomly selected node from the grid is moved by a vector generated at random. The magnitude of this vector, indicating the extent of the movement, falls within the range of 10 to 30 pixels, and its orientation ranges from 0 to 360 degrees. Subsequently, following an inverse distance weighting interpolation, every other node of the mesh is moved, simulating a rigid body distortion.

The above procedure is applied multiple times to generate multiple local distortions on the original object image. The number of distortions is denoted with N and the sequence of distortions defines the shape of the newly obtained object image. Essentially, each iteration of this process introduces a fold or a curve to the object image, mimicking the physical distortion that a waste object may experience. Because of the stochastic way of applying these distortions, different runs of the algorithm over the same initial real image produce a different synthetic image as an outcome.

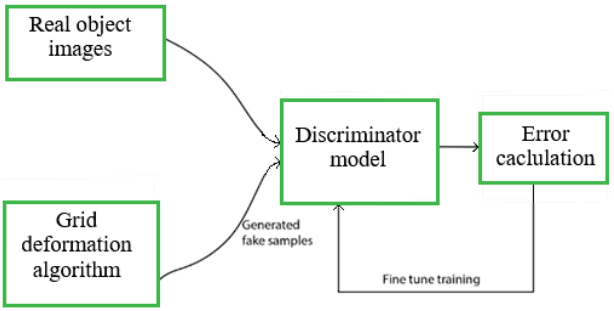

Fig. 2: Graphical representation of the architecture proposed in the current work.

B. Discriminator model

Additionally, a neural network is created, which aims at distinguishing between real and synthetic object-images. For that purpose, as described above, the proposed approach focuses on the development of a Discriminator model. That neural network receives as input an image, and returns a value between [0, 1], which represents the possibility of that image be real (values close to zero correspond to non-realistic images while values close to 1 correspond to highly realistic images). This value represents the "confidence" of the discriminator that a given image looks like a real waste object or not.

It is noted that the discriminator model is trained to deal with the unique characteristics of PETE bottles waste, which include (i) the original shape of PETE bottles and (ii) the expected distortions on bottle surface due to their compression in waste trucks. In other words, the training of the discriminator is geared towards enabling it to recognize not just generic non-realistic anomalies but also the PETE specific non-realistic anomalies that may have appeared due to the random geometric deformations (previous section). This aims to ensure that that, following the Discriminator filtering process, the deformations, folds, and curves introduced in the generated synthetic images closely resemble the real-world stresses that PETE bottles could experience.

The training of the Discriminator model, that is depicted in Fig 2, is performed by using a dataset comprising of 440 real PETE bottle images. The architecture for the Discriminator neural network consists of an input layer, three hidden layers of 1024, 512 and 256 neurons respectively with ReLU activation function and one output layer with sigmoid activation function, all fully connected. The images that are fed into that neural network must be RGB (3 channels) and resized to 256 x 256 pixels. Therefore, the input to the discriminator consists of an array of 3 x 256 x 256 neurons, with the corresponding values determined by normalizing RGB intensities from the range [0, 255] to the [-1, 1] range.

In each training step a batch of 8 randomly selected images is utilized, where 4 of them are real and the other 4 come from the grid-deformation algorithm that was described above. During training, the Discriminator is repeatedly adjusted in

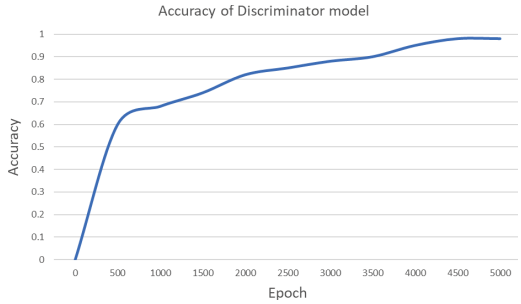

Fig. 3: The Discriminator accuracy evolution during training.

the direction that minimizes the sum of squared differences between predicted and actual realism of the examined images.

The rest of the training procedure is similar to the one followed in the training of Discriminator neural network of a typical GAN model. The main parameter values of the training process are: learning rate: 0.0002, learning momentum: 0.5, number of epochs: 5000, Batch size: 8 images.

As shown in Fig 3, even after 500 epochs, the Discriminator develops a broadly meaningful performance within the context of identifying realistic PETE appearance. After 4000 epochs the Discriminator gradually converges to a configuration that can successfully assess the realism of the examined images with approximately 98% accuracy.

C. Generation and evaluation of synthetic object-images

The next step entails the generation of large number of synthetic object-images, and the evaluation of their realism by using the Discriminator. For this purpose, 440 real object-images of PETE bottles are employed. In each real image, a distinct set of random grid deformations is applied 2500 times (see the procedure described in section III-A), resulting in a total of 1.1 million synthetic object-images. The Discriminator evaluates all these synthetic images and assigns them a confidence score (see Fig 4), which indicates their realism.

The chart illustrated in Fig 5, shows the distribution of the synthetic object-images' confidence score on logarithmic scale. Besides the prevalence of synthetic images being categorized as non-realistic (this was expected because of the random geometric deformations applied on original images), the method proposed has also the capability to generate a substantial quantity of new synthetic images that exhibit a high degree of realism. At the same time, by applying a threshold to the score of realism provided by the Discriminator, we can readily obtain distinct sets of images with varying degrees of deviation from the original data. More specifically, for the rest of the paper, we denote with D_x the set of synthetic object-images for which the Discriminator score is above x.

What is of interest in the present work, is to study the "value" of the newly generated object-images in improving the training of neural network models aiming at identifying PETE objects in challenging multi-waste industrial images. This is examined in the following paragraphs.

Dataset ID	Description of dataset
1	1000 real belt-images (Base dataset)
2	1000 real belt-images + 4000 synthetic belt-images using D_0
3	1000 real belt-images + 4000 synthetic belt-images using $D_{0.85}$
4	1000 real belt-images + 4000 synthetic belt-images using $D_{0.95}$

TABLE I: Description of datasets according to their training set differentiation.

D. Base and synthetic datasets of belt-images

To train a deep neural network model that recognizew PETE objects in images where many different types of waste are depicted, it is important to have a large dataset of belt-images in which the existing PETE objects have been clearly annotated. As discussed bellow, the current work considers the training of Mask R-CNN models which have proven very that have been shown to be very effective efficient in waste identification and categorization [15].

The synthetic object-images discussed in the previous section, are used to compile complex, automatically annotated belt-images with several PETE objects appearing on them. To this end, a random number of synthetic object-images are superimposed in random positions and orientations on top of real belt-images, thus forming a sufficiently large number of new, synthetic belt-images. To be able to draw conclusions about the effect of synthetic data on the training of the PETE recognition model it is necessary to formalize the experimental procedure. In particular, we make a comparative study that uses different datasets in training a PETE identification module.

As a reference dataset, a set of 1000 real belt-images is selected. Furthermore, to examine the effect of synthetic images, the reference dataset is enriched with 4000 synthetic belt-images, which are created using the object-images included in D_0, $D_{0.85}$, $D_{0.95}$, thus forming three additional synthetic datasets. This is summarised in Table I. It is mentioned that D_0 corresponds to absence of the procedure of Discriminator evaluation, as any object-image that is generated by grid-deformation algorithm can potentially be used in the newly created belt-images. The datasets are evaluated for their ability to train PETE detection Mask R-CNNn models which are then applied to real, previously unseen, belt-images.

IV. Results

Identifying waste in an industrial setting involves addressing the problem known as "instance segmentation." This is essential because the model must recognize and categorize multiple, potentially overlapping objects transported on the industrial belt. The established Mask Regional CNN (Mask R-CNN) that integrates a region-based convolutional neural network (R-CNN) with a mask prediction branch enabling it to provide pixel-level segmentation, has proven ability to effectively tackle waste identification and categorization tasks [15]. Hence, it is adopted in the current study to build the PETE identification module.

Initial image Confidence=0.34 Confidence=0.54 Confidence=0.86 Confidence=0.96

Fig. 4: Examples of confidence value for various single object images.

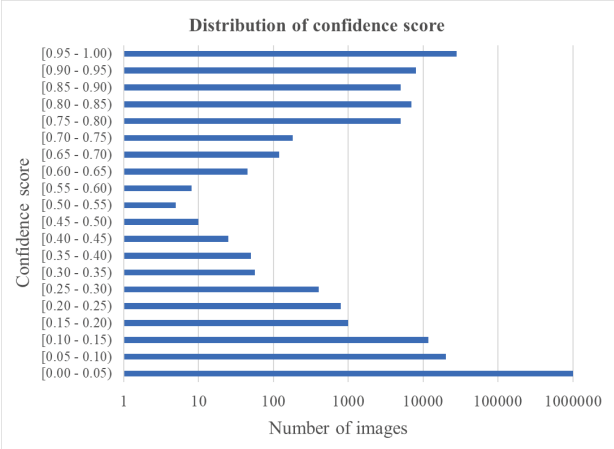

Fig. 5: The distribution of confidence scores, using a logarithmic scale for the images counting.

In this work, matterport package was used for the implementation of Mask R-CNN models [16]. All training parameters for the models remained consistent across all datasets considered in this study. Furthermore, the models were trained with the built-in data augmentation functionality enabled. For each dataset five different Mask R-CNN models have been trained each initialized with a different random seed. Then, the average performance of these five models is estimated based on a set of two hundred real, previously unseen images.

To assess the effectiveness of the generated solutions, we employ widely-used metrics in object detection models, specifically, Accuracy, Precision, and Recall rates. Accuracy is the measure of the percentage of correctly classified pixels overall that a model can predict. Considering the data imbalance where PETE objects occupy significantly less space than the prevalent background, we also compute the Balanced Accuracy, taking into account the more frequent appearance of the background. Furthermore, owing to the data imbalance, where PETE objects occupy significantly less space compared to the predominant background, we compute the Balanced Accuracy. This metric takes into account the disparity in frequency between background and PETE occurrences. Precision denotes the percentage of accurate predictions for a specific class (in this case, PETE) relative to all predictions. Recall represents the percentage of actual PETE pixels that the model can successfully identify.

The obtained results are summarized in Table II while

Dataset ID	Accuracy	Balanced Accuracy	Precision	Recall
1	94.69 %	76.57 %	67.68 %	55.14 %
2	97.88 %	80.70 %	74.52 %	57.24 %
3	98.27 %	87.41 %	83.23 %	60.17 %
4	95.38 %	78.62 %	80.22 %	58.73 %

TABLE II: The pixel-level performance of the Mask R-CNN models trained with different datasets.

indicative results are shown in Figure 6. In all training cases, the use of synthetic data (datasets 2, 3, and 4) consistently outperform the use of only the base dataset. This improvement is evident in both Precision (with an increase ranging from 6.5% to 12.5%) and Recall rates (with an increase ranging from 2% to 5%). In the case of Accuracy, the improvement is less. Still, the global picture reveals the ability of synthetic datasets to effectively generalize the characteristics of deformed waste objects, thereby contributing to the training of more effective Mask R-CNN based instance segmentation models.

Moreover, a comparison between dataset 2 and datasets 3 and 4 helps assess the actual effectiveness of the proposed methodology, in particular the effect of random geometric deformations and the role of the Discriminator. It is reminded that the synthetic dataset 2, lacks the evaluation step for inserted object-images and thus does not leverage the effect of the Discriminator. Even in that case we observe a clear improvement in the performance of the trained models. By considering how the other two datasets (3 and 4) have also affected the training of the model, the comparison reveals that the inclusion of "selected" object images, with realism verified by the Discriminator results in better performing models, showcasing an average increase of 7% and 2% in precision and recall rates, respectively in relation to dataset 2.

Focusing particularly on the dataset 3 (generated with object-images achieving realism score higher than 0.85) and contrasting it to the dataset 4 (generated with object-images achieving realism score higher than 0.95) we observe that dataset 3 leads to higher Mask R-CNN model performance, for all considered metrics. Intuitively, this implies that a less stringent selection of artificially generated object-images allows the trained models to more effectively generalize the expected features of PETE bottles, resulting in improved performance in challenging real-world images.

V. Conclusions and Future Work

The present study introduces a straightforward, easy to use methodology for generating valuable datasets of synthetic

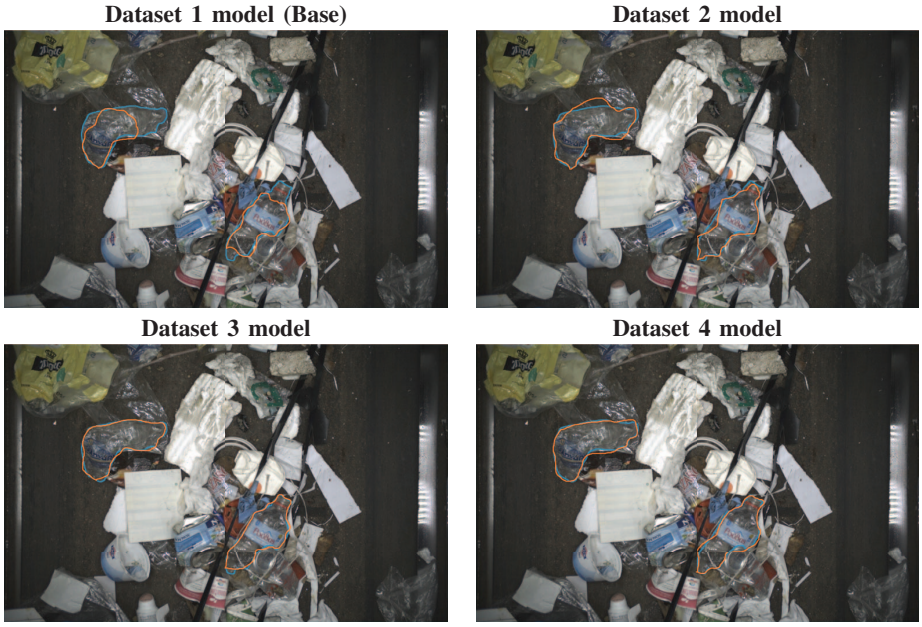

Fig. 6: Indicative identification of PETE bottles by the models trained with the different datasets considered in the present work. In all images model prediction polygon is shown in orange and groundtruth polygon is shown in blue.

images, particularly relevant to applications involving object identification and categorization.

The method involves manually segmenting the objects of interest in existing images, randomly applying geometric deformations to create new synthetic objects, training a neural network Discriminator to evaluate the realism of the generated objects, and finally using multiple synthetic objects to create additional images that enrich the original dataset. The obtained results suggest that the synthetic datasets produced through the proposed approach can significantly improve the training of instance segmentation, in our case Mask R-CNN models, outperforming those relying solely on the original data.

Our ongoing and future work focuses on the application of the proposed method to other types of recyclable materials such as aluminum and high density polyethylene (HDPE). Additionally, we are keen on evaluating its applicability in entirely different domains, such as food inspection.

REFERENCES

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, Eds., vol. 27. Curran Associates, Inc., 2014.

[2] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2016.

[3] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International Conference on Machine Learning*, 2017. [Online]. Available: https://api.semanticscholar.org/CorpusID:2057420

[4] A. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," *ArXiv*, vol. abs/2102.09672, 2021.

[5] Y. Pu, Z. Gan, R. Henao, X. Yuan, C. Li, A. Stevens, and L. Carin, "Variational autoencoder for deep learning of images, labels and captions," in *Neural Information Processing Systems*, 2016. [Online]. Available: https://api.semanticscholar.org/CorpusID:2665144

[6] M. M. Selim and R. P. Koomullil, "Mesh deformation approaches a survey," *Journal of Physical Mathematics*, vol. 7, 2016.

[7] Y. Liu, X. Lin, G. Shou, and H. S. Seah, "2d image deformation based on guaranteed feature correspondence and mesh mapping," *IEEE Access*, vol. 7, pp. 5208–5221, 2019.

[8] R. Setaluri, Y. Wang, N. Mitchell, L. Kavan, and E. Sifakis, "Fast grid-based nonlinear elasticity for 2d deformations," in *Symposium on Computer Animation*, 2015.

[9] J. Kybic and M. Unser, "Fast parametric elastic image registration," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1427–1442, 2003.

[10] S. Du, K. Hao, H. Zhang, X.-s. Tang, and B. Wei, "Patch elastic deformation: An effective data augmentation method," in *2022 China Automation Congress (CAC)*, 2022, pp. 2079–2084.

[11] E. Castro, J. S. Cardoso, and J. C. Pereira, "Elastic deformations for data augmentation in breast cancer mass detection," *2018 IEEE EMBS Int. Conf. on Biomedical & Health Informatics (BHI)*, pp. 230–234, 2018.

[12] S. Sinha, H. Zhang, A. Goyal, Y. Bengio, H. Larochelle, and A. Odena, "Small-gan: Speeding up gan training using core-sets," in *International Conference on Machine Learning*, 2019.

[13] A. Karwande, P. Kulkarni, T. Kolhe, A. Joshi, and S. Kamble, "Time efficient training of progressive generative adversarial network using depthwise separable convolution and super resolution generative adversarial network," 2022.

[14] Z. Liu, B. Xu, B. Cheng, and X. Hu, "Interpolation parameters in inverse distance-weighted interpolation algorithm on dem interpolation error," *Journal of Sensors*, vol. 2021, pp. 1–14, 12 2021.

[15] M. Koskinopoulou, F. Raptopoulos, G. Papadopoulos, N. Mavrakis, and M. Maniadakis, "Robotic waste sorting technology: Toward a vision-based categorization system for the industrial robotic separation of recyclable waste," *IEEE Robotics & Automation Magazine*, vol. 28, no. 2, pp. 50–60, 2021.

[16] W. Abdulla, "Mask r-cnn for object detection and instance segmentation on keras and tensorflow," 2017. [Online]. Available: https://github.com/matterport/Mask_RCNN