CS230

GAN Generated Images of Fruit Decay

Krithika Iyer
Department of Computer Science
Stanford University
ksiyer@stanford.edu

Abstract

About a third of all food produced in the world spoils and goes to waste. Computer vision models can help reduce such wastage. However, effectively training deep learning models requires large amounts of data. This project explores the feasibility of using two different types of GANs to generate large amounts of produce images with different levels of spoilage so that the synthetic images (along with natural images) can help develop an accurate classifier to evaluate the quality and freshness of produce. Style transfer type GAN is used to control the size of defects on the skin of the produce. Data Augmentation type GANs are used to generate many synthetic images that belong to the same class. Preliminary results indicate 12% improvement in classification accuracy based on the use of synthetic images.

1 Introduction

Given sufficient data, Deep Neural Networks have produced unprecedented performance. Lack of training data is an obstacle towards utilizing such an effective tool to solve large problems like food wastage. However, these same tools, i.e., deep neural nets, may be harnessed to generate synthetic training data to address the lack of natural training data. This project explores the feasibility of utilizing GANs for augmenting samples of images of produce of different quality.

2 Related work

CNNs for image classification is commonplace today [1]. Use of general purpose GANs [2] for augmenting images for classification has also been demonstrate[3]. The DAGAN [3] takes an image, encodes it, adds noise, and decodes it. In the process, the decoder learns a large family of class-invariant transformations for data augmentation. Auxiliary Classifier GAN (AC-GAN) has the discriminator perform classification in addition discriminating between real and synthetic data [4]. Although the generator and discriminator are competing on whether the generated image is real or synthetic, they are also cooperating on the classification part. Defects on the skin of produce indicate how bad or spoilt a produce is. Style transfer type GANs hold much potential for simulating such skin defects. Recent efforts [5, 6, 7] describe progress in this domain. Efforts to utilize style transfer type GAN to augment and create synthetic medical images are described in [8].

3 Methods

The approach is as follows:

1. Utilize Data Augmentation GAN (DAGAN) to learn transformations that do not change class for different classes of produce and utilize the same to produce augments images. [3]

2. Explore the feasibility of utilizing neural style transfer and cycle consistent GAN to alter the size of skin defects on a given produce and also transfer such skin defects from one produce to another, i.e., can skin defects (such as russeting, insect scar, etc) be transferred from an image of an apple to an image of pear or orange. [5, 6, 7]

The ultimate aim is to explore the feasibility of extracting and manipulating representations of produce defects in latent space to generate synthetic images with different levels of defects and spoilage in produce [7, 8]. One can utilize such synthetic images (with different classes of skin defects) to further generate additional synthetic images using the DA-GAN approach.

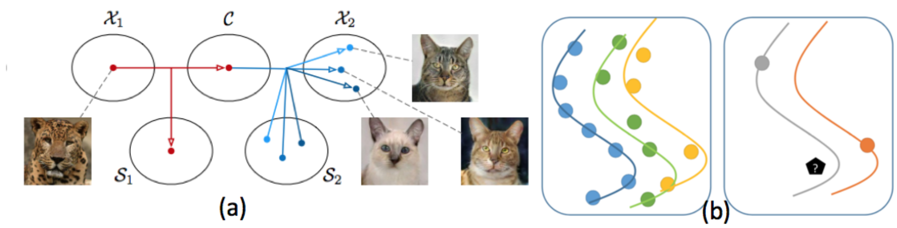

Figure 1: Illustration of our approach. (a) Translation of image from one domain (say Apples) to target domain (Oranges). Random skin defects of varying sizes (as style) can be incorporated into the target domain. (b) Learning a generative manifold for the classes in the source domain(left in (b)). The test point (pentagon) is near the orange point, but is actually closer to the learnt grey manifold. (Image credits: left [7], right [3])

Skin defects on produce are the tell-tale signs of spoilage. The defects are characterized by the percentage of surface area they cover (such as 10% or 15% of surface area) or by their size 1/8 of inch or 3/4 of an inch. Based on such skin-defect size, a visual property of the produce, the item may belong to different classes such as USDA-Grade 1, Grade-2, etc. The first step is to produce synthetic images belonging to different "visual" classes (of spoilage) using neural style transfer GAN. The second step is to utilize such synthetic images (along with natural images) as inputs to the DAGAN so the DAGAN can produce synthetic images within a particular class of produce. In preliminary investigation, the DAGAN exhibited superior results in generating synthetic images of different classes of produce: banana, apple, pear, etc. However, it did not produce good results for generating 5% spoilt apple, 10% spoilt apple etc. The defects on the skin are relatively small, compared to the size of the produce that the DAGAN could not learn to generate synthetic images with different sizes of skin defects. This is an area for further investigation in the future.

DAGAN used ResNet as the classifier in their research effort. The code released by the research group did not include the ResNet component. This effort utilized DenseNet [9] as the classifier component and integrated DenseNet and DAGAN code. DenseNet was selected because of its use of residual connections like the ResNet. The DAGAN and DenseNet were trained using the hyperparameters used in the original DAGAN study.

3.1 Datasets and Features

The three datasets utilized in the project are:

1. Fruits 360 dataset: A dataset containing over 90500 images of 131 fruits and vegetables [10]. They are 100 X 100 pixels in size.

2. Strawberry images dataset: collected by the author containing about 900 images of strawberries in different classes (unripe, ripe, overripe, decay, serious-defects, etc). The high resolution images were cropped to 256 X 256 pixels in size for use in this project.

3. USDA reference images: typically few (less than 5) images highlighting different defects of fruits and vegetables [11].

4 Experiments/Results/Discussion

Three sets of experiments were conducted as part of the project: DAGAN, Neural Style Transfer, and Cycle consistent GAN.

DAGAN: Typical data augmentation techniques use a very limited set of known invariances. DAGAN works different from other GANs. DAGAN leans a model of a much larger invariance space through training a form of conditional GAN in a domain called the source domain. DAGANs take a dataset consisting of different classes (such as apple, banana, pear, etc) and learn class-invariant data augmentations. The learned model (and representation) is used to generate and augment data (often based on single data points) in a target domain. For example, using the Fruits 360 dataset, a DAGAN may be trained on images of banana, apple, and watermelon. Once the DAGAN is trained, it can be used to generate synthetic images of pears. The architecture of DAGAN is shown in Figure 2.

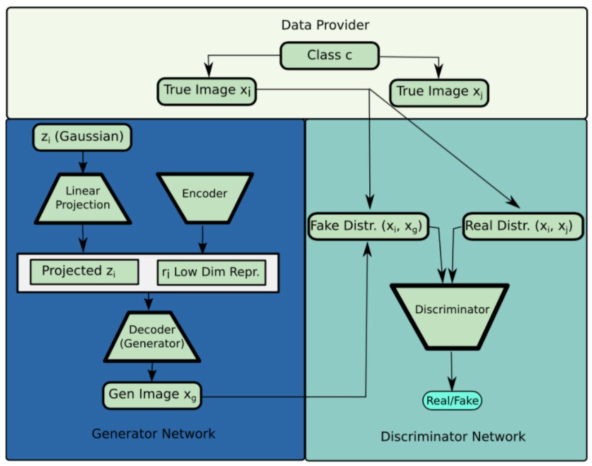

Figure 2: DAGAN Architecture. Left: Encoder takes an input image (from class c), creates a low dimensional representation (bottleneck). A random vector (zi) is transformed and concatenated with the bottleneck vector. Decoder generates an augmented image from the concatenated vectors. Right: Discriminator trained to discriminate between real and fake images. (Image from [3]).

Given a set of training data X = $\{x1, x2, \ldots..xm\}$ the learning by GAN takes the form: $z = N(0, I)$ and $v = f(z)$. where f is implemented via a neural network, v, the vectors being generated (whose distribution should match the distribution of input X), and z are the latent Gaussian variable that provides the variation in what is being generated. A GAN could also be used to map out a data augmentation manifold. Given a data point x, one can learn a representation of the input $r = g(x)$, along with a generative model as before. This model would take the form: $r = g(x), z = N(0, I), x = f(z, r)$, where the neural network f now takes on the representation r and the random z as inputs. Given a new x^*, we can: 1) Obtain a generatively meaningful representation of the datapoint $r^* = g(x*)$ that captures the information needed to generate other related data, and 2)Generate extra augmentation data $v1^*, v2^* \ldots$ that supplements the original x^*, which can be used in a classifier.

Description of Neural Style Transfer and Cycle Consistent GAN (along with associated Google Colab notebooks) are described in detail in Appendix A. Neural Style Transfer GAN was utilized to transfer skin defects from an apple to a pear. Both fruits exhibit a skin defect known as "russeting". Attempts were also made to transfer russeting-defect from a red apple to a green apple. Cycle GAN was utilized to transfer skin defects from strawberries to oranges. An interesting artifact of the effort is that the oranges have the shape of strawberries in the generated synthetic images. Since the aim of the experiment is study transfer of surface defects, in spite of the shape, the images convey useful data.

4.1 Results

Fruits 360 dataset was run with 100 classes as the source domain and 30 in the target domain. The DA-GAN was able to good generate synthetic images from the target domain. Additional images generated by DA-GAN is shown in Figure 9 (Appendix A). One can clearly see that DA-GANs can generate several variations of synthetic images that are with-in the same class. This can potentially solve the problem of not having enough training data.

Figure 3: Synthetic images generated by DA-GAN in the target domain. The first column is the natural image, rest of the columns are synthetic images. These classes of produce are from the target domain

The synthetic images generated by DAGAN were used to augment the training of image classification sets. Figure 8 (in Appendix A) shows an example of synthetic images generate by the DAGAN. Compared to a classifier trained only with natural images, the use of synthetic images improve classification accuracy by 12% for the Fruits 360 dataset and 5% for the strawberries dataset.

The neural style transfer efforts produced mixed results. While our goal was to transfer defects (like like features of "russetting", blemishes, bumps, holes, and other artifacts of the skin-defect), the color of the skin played a dominant role in style transfer. In Figure 4, one can see more "color" transfer than the transfer of russet line-like features.

Figure 4: Transferring "russeting" style

Compared to simple neural style transfer, Cycle consistent GAN shows improved results. From Figure 10, 11 and 12 (Appendix A), one can see the surface damages from the natural strawberry image (on the left) is transferred on to the synthetic orange image (on the right). The transfer results in exaggerated texture on the synthetic image. Manipulation of defect related features in latent space should further aid the transfer of such defects to the target domain.

DAGANs show the potential for generating synthetic images within the same image class. Because the situation with produce is unique, i.e., each produce class (banana, apple, etc) includes a second level of classification (i.e., Grade A, Grade B, based on its quality) the generation of synthetic images for use in produce quality estimation classifiers requires the use of two GAN approaches: style transfer type GANs for handling skin defects and augmentation type GAN for possible integration with classifiers. DAGAN by itself is not able to capture all of the skin defects on a sample image as the skin defects tend to be small.

5 Future Work

The immediate next steps include continuing with DAGAN experiments and utilizing synthetic images in image classification. We will also investigate the potential for investigating the extraction, de-entanglement and manipulation of features related to produce surface defects in latent space. More long-term, we see more granular control (in latent space) of the generation of synthetic images with different levels of skin defects via neural style transfer and consistent-cycle GAN. Such images may be used as inputs to AC-GAN or DA-GAN with integrated classifiers and meta-learning algorithms for accurately estimating produce quality. The Multimodal Unsupervised Image to Image Translation (MUNIT) framework could also help the effort aimed at latent space extraction and manipulation of defect features.

6 Contributions

I would like to acknowledge my TA mentor, Sharon, for her guidance during this project.

References

[1] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolu- tional neural networks. In Advances in neural information processing systems, pp. 1097–1105, 2012.

[2] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).

[3] Antoniou, A., Storkey, A., Edwards, H. (2018, October). Augmenting image classifiers using data augmentation generative adversarial networks. In International Conference on Artificial Neural Networks (pp. 594-603). Springer.

[4] Auxiliary Classification GAN. https://arxiv.org/pdf/1610.09585.pdf

[5] StyleGAN. https://arxiv.org/abs/1812.04948

[6] Cycle-Consistent Adversarial Networks. https://arxiv.org/abs/1703.10593

[7] Multimodal Unsupervised Image to Image Translation [MUNIT] framework. https://github.com/NVlabs/MUNIT

[8] Lukas Fetty, Mikael Bylund, Peter Kuess, Gerd Heilemann, Tufve Nyholm, Dietmar Georg, Tommy Löfstedt, Latent Space Manipulation for High-Resolution Medical Image Synthesis via the StyleGAN, Zeitschrift für Medizinische Physik, 2020.

[9] Gao Huang, Zhuang Liu, Kilian Q Weinberger, and Laurens van der Maaten. Densely connected convolutional networks. arXiv preprint arXiv:1608.06993, 2016.

[10] Fruits 360 dataset. Available at: https://www.kaggle.com/moltean/fruits

[11] US Department of Agriculture (USDA) Grades and Standards. Reference images available at: https://www.ams.usda.gov/grades-standards

Appendix A

Datasets:

Figure 5: Fruits 360 Dataset (100 X 100 pix)

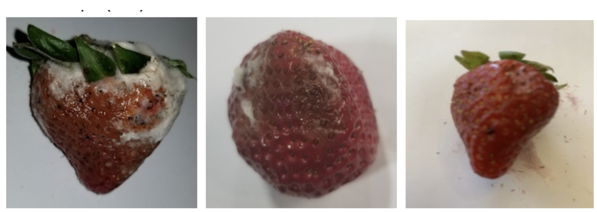

Figure 6: Strawberry Dataset (256 X 256 pix)- curated by author

Figure 7: USDA Produce Defect Reference Images

Results
Style Transfer (Figure 6)
Code: `https://colab.research.google.com/drive/1SreJehMLStnIA5tsEPTjZq20g5NzEWtz?usp=sharing`
The apple image on the left shows a skin defect ("russeting"). The experiment attempted to transfer such a skin defect to a pear. Image produced by different settings for content and style are shown in images (center and right).

DAGAN (Figure 7)
Code: `https://drive.google.com/file/d/10rkwNRcICszU3a0aihjX3kltn_cJOTX4/view?usp=sharing`

CycleGAN (Figure 8, 9 , 10)
Code: `https://colab.research.google.com/drive/1Eyz4XmLEjOJDue1FDKEWBHepJvNzZrsG?usp=sharing`
All images on the left are real defect conditions on strawberries. CycleGAN was used to transfer skin defect condition to an orange (images on the right are synthetic image generate by CycleGAN). One can clearly see exaggerated skin defects in the synthetic images.

Figure 8: Transferring "russeting" style

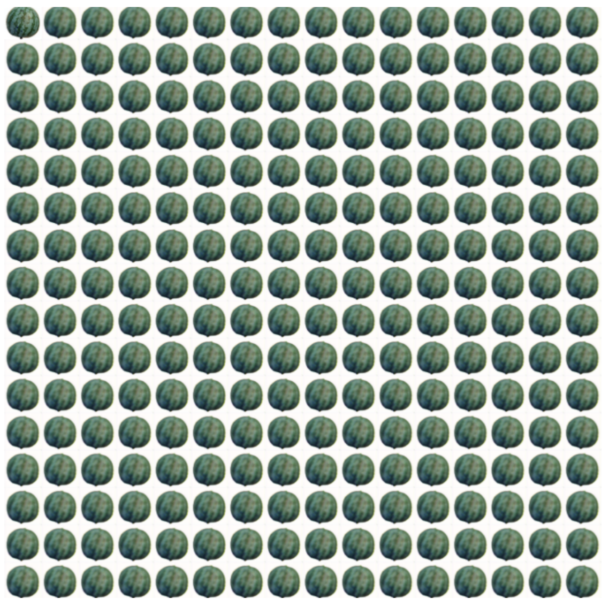

Figure 9: The image on the top left is the only real image (seed image). Rest of the images are generated by DAGAN.

Figure 10: Left: Fungus on strawberry, Right: transferred as breaks in the surface

Figure 11: Left: Skin decay on strawberry, Right: transferred as breaks in the surface

Figure 12: More examples of skin decay being transferred as a blemish on the right.