Exploring the Intersection of Computer Vision and Generative Adversarial Networks in Medical Image Synthesis

*By **Dr Emily Chen, Prof. Chien-Ming, Dr Steve Lockey, Dr Hassan Khosravi & Dr Nell Baghaei***

Professor, University of Queensland, Gatton Campus, Gatton, QLD, Australia

1. Introduction to Computer Vision and Generative Adversarial Networks

Pixel-wise similarity metrics, including Euclidean norms and mean squared errors, measure sample distortion and constitute a popular option in computer vision applications, such as super-resolution, style transformation, and image-to-image translation. However, this approach is not suitable for several domains, including psychology research. The sufficient discrimination between the types of image content that fulfill complex constraints, such as the segmentation of individual organs and vessels within the human body, can only be indicated by a learned distance metric. Consequently, recent state-of-the-art computer vision tasks in medical imaging are supported by learned data representations of deep neural networks. This chapter investigates the intersection between computer vision models and generative adversarial networks in the field of medical images. Specifically, we examine the performance and properties of conditional generative adversarial networks and their Wasserstein extension for real-time multi-organ segmentation of X-ray computed tomography.

The supervised training of machine learning models for computer vision tasks has become increasingly dependent on the availability of high-quality labeled data. However, annotations are often labor-intensive to obtain as they are typically requested from domain experts (e.g., physicians in the case of medical images) and thus generally limited in size. To overcome the restrictions imposed by the necessary presence of labeled data, generative adversarial networks were introduced for unsupervised learning and widely adopted by the research community. Generative image models aim to maximize the similarity between real and generated samples, enabling the production of synthetic data that emulates training samples. Such synthesized image datasets can then be employed to train a more generalized computer vision model.

1.1. Definition and Overview of Computer Vision

Computer vision is closely related to other fields. Multidisciplinary topics such as image and speech processing often focus on the same algorithms and issues that computer vision does. Semi-automated object recognition can also be seen as an area of artificial intelligence. Nonetheless, computer vision is not limited to artificial intelligence. It can be seen as a part of computer science in general, and as a topic of interest within it, for example: in machine learning, pattern recognition, database systems, and neural networks.

Computer vision is the field that focuses on how computers can gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the human visual system can do, including pattern recognition and decision-making irrespective of perspective, distortion or lighting. A significant part of computer vision is devoted to the topic of image review, that is, extracting high-level information from digital images. The image data can take many forms, such as laser scanning range images, light field images, sonar and radar images, and time-varying images. Processing may involve such tasks as classifying the image, detecting a particular object in the image, and segmenting the image.

1.2. Definition and Overview of Generative Adversarial Networks

Generative Adversarial Networks have two key characteristics that make them unique from other generative models. The main characteristics are that GANs have a G-generator as well as a D-discriminator component. The generator and discriminator components both use an adversarial process. This is the second unique aspect of GANs. The model's architecture is another characteristic that makes GANs unique. While other models apply a training approach to a pre-existing architecture or network that always outputs the same thing for a given set of inputs, GANs evolve and change over time through the adversarial process that involves the generator and the discriminator networks.

Generative adversarial networks (GANs) are a relatively new concept in machine learning. They were introduced for the first time in 2012 and developed by Ian J. Goodfellow and his colleagues in their 2014 research paper. Goodfellow, a Ph.D. student at the University of Montreal under Yoshua Bengio's supervision at the time, invented GANs during a party in Bengio's apartment. GANs, as well as the other specialized architectures we will discuss in later chapters, are the most intriguing topics to explore and understand.

1.3. Applications of Computer Vision and GANs in Medical Image Synthesis

Finally, by generating additional realistic examples, GANs can avoid overfitting in the training of machine learning classifiers.

In addition to traditional data augmentation methods aiming to increase the amount of available data by applying diverse image transformations such as affine transform, resizing, rotation, translation, and flipping, GANs can enhance data by creating permutations of existing images through label-preserving transformations. Moreover, GANs avoid the problem related to the increased size of data origin during augmentation and can create images with exceptional situational complexity that is desirable for training machine learning models, especially in medical diagnostic problems.

Many task-related domains become candidates where GANs methods can be efficiently used. Current use-cases of GANs for medical images traditionally include noise removal and acquisition noise level reduction, resolution enhancement, generating sequences from single images, missing anatomy or imaging artifacts restoring, virtual and augmented reality, domain adaptation, pre-processing before training a supervised classifier, and adversary-based training to generate realistic de-identified images. In their 2022 study, Menaga et al. describe a comprehensive approach for opinion mining and categorization utilizing both semantic knowledgebases and machine learning.

2. Fundamentals of Medical Image Synthesis

Furthermore, GANs are particularly attuned to providing a means of incorporating structured information and addressing problems where samples are high dimensional. Clinically, GANs can simulate identifiable distributed patterns in large and small scale multimodal data including positron emission tomography (PET) and MRI. Various extensions of the techniques have enabled clinically relevant tasks such as the estimation of 4D MRI when only select 2D images are available at each time point.

Generative adversarial networks (GANs) are an emerging field in machine learning. Unlike traditionally supervised machine learning problems, the GAN formulation involves a two-player game where a generator network can be used with examples drawn from a distribution of real data to construct artificial data. For images, the objective is to generate artificial data that is very similar to the real data. Because GANs are a type of unsupervised machine learning problem, they can be used in cases where labeled training data is unavailable. This is especially useful for semantic segmentation and detection tasks that incorporate large compound data sets.

Early work in medical image synthesis involved pattern recognition, search, and nearest neighbor-based techniques, which in some cases achieved impressive results given a large collection of medical images. However, not all medical image synthesis tasks have sufficient high-quality medical images available. Instead, a smaller collection of volunteer examples can be labeled and used to supervise the generation of new medical images. This makes the problem more similar to traditional image synthesis tasks.

Medical image synthesis is a technique for generating new, artificial medical images. The problem formulation requires that the generated images are realistic and appear similar to the medical images that are obtained via expensive imaging modalities such as magnetic resonance imaging (MRI) and computed tomography (CT). However, challenges arise because these imaging modalities provide very different types of image styles and data. To address these challenges, a variety of techniques have been developed.

2.1. Importance and Challenges of Medical Image Synthesis

For instance, unlike natural images of ImageNet, which contains more than 14 million images belonging to one of the 20,000 categories, the medical datasets' image numbers and disease categories are both severely limited. Distance learning from natural images to medical images, reduced data scale, and low image quality are three common issues existing in medical image processing. Distance learning from natural images to medical images, since the two datasets are completely different, the pre-training of CNN can clearly improve the extraction ability of the medical image feature domain. However, medical images have inherent properties that differentiate them from natural images because they portray the physiological structures within the human body and are acquired using a variety of sophisticated modalities, such as X-ray computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound (US).

Medical image synthesis is an important and challenging task in medical image analysis, in part because of limited public access to medical image datasets and the need for large annotated datasets. These constraints seriously restrict the use of medical imaging datasets and progress in the field as a whole. The training of machine learning algorithms usually requires a large number of annotated samples, while most current medical datasets are downsized because of limited raw data. Furthermore, in medical institutions, patients' rights and privacy are always considered, and patients' data can only be stored and used under strict regulations.

2.2. Types of Medical Imaging Modalities

Anatomical imaging is used to visualize the structure of an organ using high contrast resolution (spatial resolution). This helps in the discovery of various characteristics of organs that are important to identify normal as well as abnormal conditions in patients. Techniques which are used to provide information about tissue or metabolic activity (functional imaging) are generally non-contrast techniques. It may also require the use of contrast agents to help resolve additional details or areas of concerning activity. Computed Tomography (CT) imaging, positron emission tomography (PET) imaging, MRI imaging, and ultrasound imaging are commonly used for anatomical imaging. The modalities used in molecular imaging are PET-CT, PET-MRI, SPECT-CT, and radionuclide therapy.

Medical imaging is used to view the human body, its parts, and the diseases in them. An image of a human body or a body part is created through computer-processed data obtained from modalities that use relatively high levels of energy, e.g. X-ray, gamma rays (gamma camera), sound waves (ultrasound), radio waves (MRI), and radioactive tracers. Medical imaging can be generally categorized into anatomical imaging and molecular imaging. The image created essentially represents a physical picture of the body and all its parts, so that they can be studied and interpreted. It can be used for diagnosis, screening, treatment planning, and guidance for invasive procedures. A large majority of the imaging techniques are non-invasive. Imaging is often the first stage in a workflow which requires clinical knowledge. During this process, the images are checked and interpreted by a medical professional, often a radiologist.

2.3. Common Datasets and Benchmarks

For example, IDC-identifying the presence of invasive ductal carcinoma in breast cancer shapes is from 162 whole mount slide images of breast cancer specimens scanned at 40x. Currently, researchers typically acquire data from a small number of subjects for special tasks or manually label part of the larger datasets. Due to privacy protection concerns, the small dataset greatly impedes the development of CV research in medical imaging. It may seem that lack of training data also hinders the research for the synthesis from medical images. In this paper, we review, analyze the intersection of computer vision and generative adversarial networks in medical image synthesis, and present both qualitative and quantitative outcomes.

Owing to the progress of deep learning with big data, it is common in computer vision (CV) that researchers use the state-of-the-art models trained on large-scale benchmark datasets to initialize their own tasks. The ImageNet dataset contains 1.2 million training images and has greatly promoted the research in object classification and detection. In the biomedical imaging community, there are also some widely used datasets for validation purposes, such as camelyon16, 17 for pathological diagnosis and retina vessel segmentation datasets RITE 18 and IDRiD 19, 20, 21 for retinal vascular analyses.

3. Deep Learning Techniques for Image Synthesis

3.2. Generative Adversarial Networks in Synthesis of Images Generative Adversarial Networks have been proposed by Goodfellow et al. and consist of two networks, a Generative Network G, and a Discriminative Network D. The G network samples from a prior distribution and then transforms these samples into a sample that is supposed to be independent of other samples produced by x. The D network receives either a training example from the x dataset, or one of the samples from the results of the generator G. In either case, the goal of D is to learn to distinguish an x data with one of the generated samples. The adversarial networks are trained simultaneously, the objective of the G network is to minimize the probability of the Discriminative network to generate its samples. That is, the objective function for the G network is maxG log(D(G(z))). The D network objective, in turn, is to maximize the probability of assigning the correct labels to both incoming data samples from the x dataset and those produced by the G network, i.e. minG maxD log(D(x)) + log(1 - D(G(z))). When D and G train together, they reach a stage of equilibrium, where D is not able to distinguish between real and fake samples.

Overview of Deep Learning Techniques

Deep learning is a subdiscipline of machine learning and is based on learning representations of data. Learning is unsupervised, semi-supervised or supervised, and uses the discovered representations to classify similar data in the future. Deep learning models are constructed with deep learning algorithms that are forward-feeding artificial neural network that contain more than one hidden layer structure. This type of model has vastly improved the performance of machine perception problems, in particular with the use of large neural network architectures and the ability to learn from copious unsupervised data. There are several deep learning approaches that can inject visual realism in predefined image data that we describe in the section that follows.

3.1. Overview of Deep Learning in Image Synthesis

The versatility of GANs, along with an expansive array of flexible underlying architectures, means they are being used to synthesize medical images. Experts in the field know that these tasks are challenging, with data often having strong patterns underlying specific medical problems but generally being different enough to show unique microscope biologies or patient anatomies such that standard computer vision modeling does not work. As a result, the same features that have made GANs a powerful tool for non-medical data offer potential to really push over established boundaries. Furthermore, the flexible nature of GANs can provide interpretable explanations of how the models are making their decisions, which can be essential in patient treatment. However, simply slapping together a dataset and asking a GAN to create useful new medical images is not a surefire way to success.

Generative adversarial networks (GANs) are a type of generative model that have become widely popular both for the powerful images they can produce and their supporting paper by Ian Goodfellow and colleagues. GANs work by training two separate networks simultaneously to generate brand-new images that must be realistic. This realism comes from the fact that their output images can be mistaken for images of whatever type the training data consisted of - a form of "fooling" their opponent which gives the adversarial network its name. One network is trained to generate these realistic images, known as the generator, while the other network, called the discriminator, is trained to quickly judge whether an image is from the training data or a product of the generator. The competitions between these two networks throughout their joint training process ultimately lead to the generator producing extremely strong examples of synthetic data.

3.2. Convolutional Neural Networks (CNNs) in Image Synthesis

The rather efficient prevention of overfitting in CNNs has led the machine learning community to extensive exploration and use of general data augmentation and style transfer techniques. Augmentation, however, is limited by the explicit constraints and invariance built into convolutional networks. The results generated by CNNs are less sensitive to, dependent on, and inferior to the expressiveness and semantic understanding found in Generative Adversarial Networks when used to produce visual content (especially images). The principles behind GANs are heavily rooted in the dynamics of CNNs, although the style and outcomes in GANs are much more closely in line with how we process and interpret images, thereby allowing us to greatly extend the flexibility of CNNs used for the generation of visual content. Due to the examination of the flow and convolution in a neural network, some key results can be reinterpreted in a computationally simpler manner.

Generating realistic new content requires flexibility. Convolutional Neural Networks (CNNs) are drastically reduced in terms of this flexibility because they explicitly rely on translations, shifts, and dilation invariance. Invariances in the data might be readily exploited; however, this has also led to a general semantic loss: multiple outputs can easily be semantically identical, but different operations can lead to very different representations. CNNs intrinsically lack sufficient semantic understanding of the visual world, which is necessary for meaningful image synthesis. As a consequence, semantic and coherence constraints deemed essential to assure the realism of generated images are sparsely encoded in CNNs in an implicit or explicit manner. Layer activations at each location are usually mapped to specific image segments, resulting in non-informative, large feature maps with only a few or even no informative features.

3.3. Autoencoders and Variational Autoencoders (VAEs)

However, one critical stage in the training of the variational autoencoder is dictated by the latent space, which distributes the approximate inference posterior $P(Z|X)$ over the model input X. Hence, the challenge in the training of variational autoencoders comes from the distribution and model capacity issues of the zoning layer. Furthermore, because the traditional autoencoder model does not explicitly model the latent space or control the generation output, when given a random noise seed, the model may miss the input conditioning. Generally, exploring the full potential of the zoning space requires the demonstration and investigation of these models in the context of a real-world application. In this chapter, the applications of autoencoders and variational autoencoders to MNPs are discussed, showcasing their potential as an essential part of CVMNPs for deep learning model design.

One of the most commonly used deep learning models for unsupervised learning problems in MNPs is the autoencoder. An autoencoder is a deep learning model that is trained to encode input data into a compressed format or latent space representation and then decode or reconstruct the initial data as closely as possible to the original input. An autoencoder is always composed of an encoder and decoder model. The encoder model is responsible for encoding the input features into the compact format, while the decoder model tries to reconstruct the initial input from this compact representation. Variants of the autoencoder model, particularly the variational autoencoder, have found wide applications in computer vision as dimensionality reduction and in the generation of new synthetic data from an existing model.

Autoencoders and Variational Autoencoders (VAEs)

4. Generative Adversarial Networks (GANs) in Image Synthesis

Deep learning (DL) has had tremendous success in computer vision problems. Some of this success can be attributed to the development of large-scale datasets and the evolution of convolutional neural networks (CNNs), which are designed to automatically and adaptively learn spatial hierarchies of features. However, medical imaging is an area where such large-scale datasets are not readily available, due to various concerns with patient health data privacy, as well as risks of illegal data usage and/or distribution. Thus, medical imaging practitioners often do not have access to large datasets. Moreover, the increased complexity of the data required for more detailed studies, in addition to data scarcity, makes the problem of medical image synthesis a very challenging problem. This is likely to change in the coming years, given the accelerating pace of modern advances in AI and computer vision fields.

Generative adversarial networks (GANs) have proven to be powerful tools for generating data, which is especially useful in medical imaging since datasets can be very limited. Many examples have been presented in the literature, showing both the promise and the challenges that GANs bring. In this paper, we present a work in progress using a hybrid method that combines the great representational capabilities of convolutional neural networks with adversarial learning. We describe the several steps and choices, as well as the net design and adaptations; and most importantly, the results that have been achieved so far. This paper's main contribution is the detailed description of the process and the preliminary results. We show that the method has potential, provided proper adjustments are made in the training process, as tested by the good results achieved in 2D cardiac MRI data.

4.1. Introduction to GANs in Image Synthesis

The transformation of GANs by Deep Convolutional GAN (DCGAN), the inception of "Progressively Growing GANs" (PGGANs), and many other works that have expanded on what GANs can do have made GANs the backbone of modern computer vision research. Potential usage includes image generation, super-resolution reconstruction, style modification, inpainting, and image-to-image translation. While its implementation in art and computer science is inarguable, incorporating GANs in the sensitive field of medical image synthesis still necessitates reframing older GAN architectures as regular GAN use for scientific challenges has underperformed domain-specific convolutional neural networks.

Generative Adversarial Networks (GANs) provide a groundbreaking framework for creating computationally derived artifacts, which are especially useful in visual-related tasks. The creation of a GAN mainly involves defining a generator network designed to create the desired synthetic artifact (image/text/audio) and a discriminator network trained with a dataset of real images and the synthetic images produced. By playing a 'two-player game', both networks will improve iteratively, with the primary objective of the generator network being to produce images that are indistinguishable from the real dataset, so the discriminator network cannot differentiate the synthetic images from the real ones.

4.2. Architectures: DCGAN, WGAN, CycleGAN

For adversarial-like algorithms, one of the major challenges in using the above architectures in medical image synthesis is having sufficiently deep and wide model architectures to generate high-quality images. Generally, such adversarial-like models are applied to generate natural photographs, for which the input data has uncurated, high-quality annotations, and from which the images of interest are pulled from a relatively narrow range (e.g. human faces are all generally the same size and at a specific orientation facing the camera). Furthermore, photographs generally have narrow spatial ranges of pixel intensity values (since standard monitors and screens can only process images of 8 bits of depth/channel) and have relatively little noise in the images. However, medical images come in many different spatial dimensions, have relatively large differences in pixel intensity values, and have different specifications for different kinds of images.

Three key GAN architectures used in the initial batch of papers that used GAN models in medical image synthesis are DCGAN, WGAN, and CycleGAN. The Deep Convolutional GAN (DCGAN) architecture does not use fully connected layers, has all convolutional layers transposed in the generator, has LeakyReLU instead of ReLU, and has dropout. The Wasserstein GAN (WGAN) architecture used weight clipping, used a linear instead of ReLU activation between the layers of the critic, and trained using the Wasserstein Loss, instead of the log loss used in the original GAN. The CycleGAN architecture is applied in GAN models that use "unpaired" training data, does not have fully connected layers and does not use any convolution striding.

4.3. Training and Optimization Techniques

Further, the dynamics of the Adam optimizer can itself be altered to stabilize and optimize G, and this variant, called Adam with exponential moving averages of parameters, is discussed next.

To circumvent and address the inherent instability of GANs, usage of fully convolutional networks and batch normalization layers in G have been shown to be effective. Moreover, in training G, deep learning literature suggests that certain combinations of binary cross-entropy loss function, the optimizer, and the update step-size can generate competitive results. Specifically, usage of the Adam optimizer as the stochastic optimization algorithm together with the exponential learning rate decay has been shown to be effective. Similarly, discriminative learning rates can be used in GANs as they have been widely used in vision tasks such as image classification.

Second, if the architecture of G is not chosen properly, the generated images will be highly blurry, highlighting the fact that there is no explicit way to enforce that the generated samples come from the underlying data distribution.

In this section, we discuss the training and optimization of GANs. Training and optimizing GANs is challenging mainly because they are inherently unstable. Some of the reasons for the instability are: during training, the adversarial loss is minimized with respect to the generator G and the discriminator D, but in learning G, the discriminator D is changing continuously and D's change alters the dynamics of G. As a result, D and G can potentially enter a dizzying dance where the improvement of one comes at the cost of the other.

5. Medical Image Analysis and Diagnosis

The generated image is remarkably authentic that even the treating clinician would find it difficult to differentiate between real images. Such computer-aided synthesis of realistic images is tremendous in improving the deep learning model's generalizability by addressing the missing, mislabeled, or distorted ground truth at patch or pixel level in the training dataset. The ease of GANs use, its capability to learn data distribution, and their profound automation are shaping the future of medical image synthesis. Despite their advantages, the inherent manual interaction required for initialization of GANs with diseased patches of the host image, and the ever unresolved challenge of standardizing the hyperparameters are the challenges bestowed on GANs. Resolving the limitations and addressing the challenges are inevitable in achieving proficiency in GANs applications towards patient care.

Computational algorithms and artificial intelligence have emerged as critical tools in medical imaging, offering opportunities to improve clinical workflow efficiency, analyze complex multimodal data that are quantitatively and qualitatively superior to conventional human interpretations, and predict clinical outcomes. Medical image synthesis techniques customarily rely on predefined training data that need extensive human effort in collecting and accurately annotating large-scale databases a priori. As annotation costs and time increase, the need arises for new techniques to generate clinically realistic images independent of large bespoke datasets. The use of convolutional neural networks for automatic end-to-end learning by mimicking human cognition steadiness in medical image synthesis has broken new ground in this arena. Kudos to the Generative Adversarial Networks (GANs) for achieving novelty in medical image synthesis. The additional advantage that GANs have over other generative models is their ability to generate high-resolution, realistic-looking images.

5.1. Role of Computer Vision in Medical Image Analysis

In this subsection, our focus is on the role of computer vision in medical image analysis. Figure 2 shows some computer vision tasks related to medical image analysis. Localization and detection tasks are used for finding regions of interest (e.g., finding and localizing metastases in CT images or prostate in MRI images). Content-based image retrieval is about finding similar images when given an example image. Classification tasks are used for finding disease existence. Regression tasks are suitable for finding continuous parameters (e.g., estimating age, predicting disease stage or size). Segmentation tasks are used for finding precise localization of disease areas. Image enhancement and geometric transformation are used for improving image visualization. Image registration is to find corresponding regions of interest between two images, which are widely used for multi-modal image analysis, change detection, and tracking tasks. With the help of technical advance of GANs, it is easy to design and learn more advanced deep learning based computer vision algorithms related to these tasks.

5.2. Automated Diagnosis Systems

Using discrimination features (features of disease directly) is the first approach. Different types of deep learning models, such as CNN, sparse auto-encoders, and dictionaries, have been proposed for the recognition of discrimination features. They have high accuracy, strong robustness against noise, and high-speed computation. The second approach is tumor recognition methods, which aim at semantic segmentation in order to determine the area of interest (lesions or regions of interest) disease. The third approach concerns disease residual extraction methods, which aim to remove the influence of other diseases so that the system can probe into the base of a relevant disease. In these three methods, the CNN is the most popular model and exhibits the best performance in practice. In particular, a modification of CNN, called the deep neural network (DNN) model, has exhibited interesting prospects and high diagnostic performance.

Automated diagnosis systems attempt to mimic the work of radiologists by automatically interpreting medical images. There are unsupervised and supervised diagnosis systems. The supervised systems use labeled data to determine extracted feature patterns that differentiate between different categories of diseases. They usually consist of labeled training and testing datasets, conveniently extracted features, training classification systems, and patients' phenotypes as outputs. In contrast, unsupervised systems do not require labeled data. They usually consist of unlabeled training and testing datasets, automatically extracted high-level abstract features, neuron clustering, and deep unsupervised feature learning and algorithms. The usage of unsupervised methods may have a more efficient training and lower sensitivity to internal variation between phenotype categories. In particular, there are three approaches using deep learning.

5.3. Challenges and Ethical Considerations

Medical image synthesis is always limited by efficient spatial transformations and intensity changes to medical images. This is also true in the modern synthesis methods. Moreover, spatial information including part-location, boundary, and morphological variability could provide diagnostic and grading information. The incorporation of more interpretability and clinical relevancy information during training could be a feasible direction for future synthesis research. In addition to generating synthetic data, improving the generation of high-resolution images would enhance the accuracy and precision of all other models built for detection and diagnosis.

While medical image synthesis can provide useful image training data, there are potential ethical considerations worth highlighting. First, synthesized images can lack clinical relevance, causing networks to learn irrelevant clinical patterns. One potential solution is using synthesized results as the noisy clinical priors to improve network performance. Second, we cannot guarantee the real diagnosis labels of the synthesized images. To tackle this, we make a conclusion: in the study of transfer learning, a GAN can synthesize images that mimic a certain degree of imaging appearance, enabling a model to learn universal image features of large-scale, rich resources, therefore improving the transferability and generalization abilities of the model.

6. Applications of GANs in Medical Image Synthesis

The most popular medical image synthesis application of GANs involves the synthesis of cross-modality images/volumes, for example, in producing a T1-weighted high spatial resolution image from a T2-weighted input MRI for brain tumor segmentation, or histopathology images that correspond to breast lesion MRIs in order to remove the requirement for invasive tissue biopsies. This is also the most direct utilization of the versatility of GANs for generating diverse output images from the same input, thanks to the stochastic nature of GAN output.

Though many of the original applications and GAN variants were not directly targeted towards medical imaging, the inherent capability for GANs to generate unlimited, realistic synthetic images of a target modality, given a training source modality, makes GANs attractive for medical image synthesis. Many of the existing medical applications of GANs in image synthesis largely hinge on the same key advantages of GANs that were discussed in earlier sections (6.1): GANs can produce synthetic images in contrast to deterministic transformation maps of other conventional synthesis techniques, allowing it to enhance perceptual quality along with content preservation; and GANs do not rely on the computation of loss function terms as their regression-based counterparts in conditional image synthesis, allowing it to synthesize diverse images from an input image or volume.

6.1. Synthesizing X-ray and MRI Images

Two interesting examples of the application of image synthesis within this proposed hybrid modality alignment cascade are as follows. To bridge the modality gap between planning and alignment with daily acquired intra-image setup modality, image synthesis is first used to generate synthetic daytime and Fifield X-ray images of the patient being aligned. With the availability of large real-time radiotherapy X-ray image datasets, GANs are trained to artificially generate synthetic X-ray images directly from the patient's own breath-hold T2 weighted MRI. Despite the existence of high level spatial and intensity differences between X-ray and an individual MRI over all different patients in the intra-database, the investigated GANs successfully model the appearance to align with the X-ray color and edge-weighted appearance.

The application area most catered to in terms of medical imaging in the medical scientific community is magnetic resonance imaging (MRI) for data abstractions and pixel-level classification of labeled images. The innovations have had significant implications towards better medical decision support.

The technology has also moved towards X-ray images and imaging modalities of non-invasive radiotherapy. Given the vast research and publication of labeled datasets, a growing interest has been towards the question of unsupervised feature discovery and representation learning for these applications. Once synthesized, the architecture of X-ray to native CT is also adapted for use in the generation of synthetic X-ray and native MRI for radiotherapy-based alignment and setup verification. The undesirable low-level artifacts that exist within the generated synthetic images across datasets are considered when aligning later-stage sparse infilled GANs with active contours based on the target (X-ray).

6.2. Enhancing Image Quality and Resolution

To handle these problems, small organs are first extracted and generated using a dedicated network, and then results are fused by morphological operations. GANs have also been adopted to enhance nodules in CT after segmentation, which is another challenging problem due to the small size of nodules, that encourages us to explore solutions for small structures. Additionally, adjacent orthogonal 2D DSA images are presented based on a single moving 3D object. The DSA angiogram imaging suffers from the small number of projections and the spatial and contrast resolution limitations. The cycleGAN conversion of existing 2D images also suggests that GANs can generate missing orthogonal views at different angles for iterative reconstruction without additional radiation and data storage.

Medical images need to be of high quality in order for practitioners to make accurate decisions. In order to assist with the image quality concerns, GANs have been used to enhance extracted organs. Although several works have been implemented to generate high quality images, small and fine details are often missed. In multi-organ ultrasound images, small structures are inherently interconnected with the neighboring large organs which have similar surrounding tissues. Consequently, generated results can become blurrier where large organs are unable to allocate sufficient representation for small structures.

6.3. Data Augmentation and Imbalanced Datasets

Imbalanced datasets are common in medical image research problems, since one category could appear more frequently than another in real-world data. Criteria replacement generally leads to some bias in model performance. Mitigating class imbalance is critical for training accurate and realistic image synthesis models from unbalanced medical datasets. GAN training discriminates the real/fake sample pairs by applying the binary cross-entropy loss function to train the DisNet. In a real-world scenario, balancing the training of GANs with unequal numbers of samples representing various classes is a difficult problem. The goal of imbalanced image data is to incorporate prior knowledge about the differences in frequencies between classes so that the machine learning algorithm doesn't produce results that are biased against the majority class. Ensuring the generated images are accurately conditioned on the input labels can be accomplished by using the GAN model to generate the full set of unseen data. This may include additional samples from the under-represented classes.

As we discussed in Section 2.7, deep learning models often require enormous datasets to achieve superlative performance. This is especially true for generative models such as GANs, since they must capture the high-dimensional input space P(X) in an unsupervised manner. For medical imaging problems, we may have limited access to datasets due to patient privacy, the high cost of expert annotations, or underdiagnosis in some conditions. Data augmentation is a common technique to address the paucity of data during model training by artificially generating new training examples (realistic deformations of the original examples) in order to mitigate the risk of overfitting. Augmentation of 3D images is also vital to prevent overfitting during GAN training. Signal perturbation that preserves the original class labels/faux-real pairs is an essential property of the data augmentation process.

7. Evaluation Metrics for Image Synthesis

In the context of medical imaging, the situation is even worse. No abnormal image database is available, so no ImageNet-like score can be directly applied. Instead, each application of a generative model involves reconstructing some kind of input data at the test stage, and the quality of images is evaluated through these inputs (e.g., a diagnosis for a clinician, segmentation, object detection, or classification for a computer). This has deep implications on the types of synthesis models we should use: the models should be tailored to the different computer vision and deep learning tasks that we intend to perform with the synthetic data. Furthermore, it makes generalization a very challenging problem, since a model that has worked well in one domain does not necessarily generalize well in a different domain or with a different type of task.

Central to the training of an image synthesis model is the choice of performance metrics. These are the tools that guide the model towards the goal. For classical tasks such as classification and detection, error rates and invariances have been well studied and offer a direct means of drawing comparisons and making trade-offs for different models. In the context of image synthesis, however, the situation is quite different. Many of the properties that make two images "similar" depend on what one wants to use them for. For example, while high Inception Score means that the generative model can mimic an object classifier, it does not necessarily mean that the model can synthesize high-resolution natural images of a particular class, nor does it imply that the model can synthesize high-resolution, artifact-free textures.

7.1 Overview

7.1. Qualitative Metrics: SSIM, PSNR

Traditionally, these two metrics perform specifically well on contrast information such as resolution and noise, which is not necessarily the best approach for tasks that are visually similar at face value but challenge texture detail or have a trade-off between two independent metrics. Specifically for MRI, we consistently find real and synthesized images that have very low PSNR and less SSIM. Despite problems with visual quality, these low metrics may not necessarily reflect the desired synthesis goal. The performances of PSNR and SSIM are consistently low on FLAIR T1, Whole, and Dataset 3 more broadly, but especially low on segmentation classes normal.

Quantitatively, the similarity of images to ground truth is useful in medical image synthesis. We employ PSNR and SSIM. Peak signal-to-noise ratio (PSNR) is the most common distortion measure when pixel values are quantized, and its result is the de facto standard for synthesizing MRI acquired data. The metric represents the logarithm of the ratio of the data-based peak squared error and the average squared error of the difference of two inputs. In the context of PSNR, the size of the image and the sample space determine the relative metric. Structural similarity index (SSIM) considers three factors: pixel value contrast, structural similarity, and luminance normalization and multiscale structure.

7.2. Quantitative Metrics: FID, Inception Score

Introduced by Salimans et al., Inception Score evaluates how realistic the synthesized images of the GANs are and how diverse the outputs are. This is achieved by providing a pre-trained model as a feature extractor and computing the score based on the generated outputs given by the generator network. Since the generated images are meaningless to humans, the mean value of KL-divergence, having a lower bound to human assessment, between the network's output for each class should be less than that of the in-distribution training images. Karras et al. enhance Inception Score with Frechet Inception Distance and provide a more meaningful and distinguishing evaluation result in comparison to traditional GANs. Ke et al. go a step further by proposing MLP-Inception Score and c-MNIST Score to evaluate the class-conditional synthesis quality, which measures how well the model can generate quality images not only in the overall categories but also in certain classes. It is shown that KID and traditional IS fail to identify the weakness in generating sharp and distinguishable images. However, FID can still be meaningful when comparing different methods.

7.2.2. Inception Score

The Frechet Inception Distance (FID) serves as a top-performing unsupervised quality evaluation metric for GANs. The FID does not have a linear scale, and a comparison can only be drawn among related experimental settings. It is named after the Frechet distance, which measures similarity between two distributions. Kernel Inception Distance (KID), introduced by Binkowski et al., is essentially the same as FID but calculated with a kernel instead of a difference in means and covariances. Fixing the feature-generated dataset so that it does not change during training due to variations of the generator, a mathematical solution of handling the discrepancy in FID formulation has been derived. If there is a mode collapse, no lower bounds of FID are promised in vari dimensions of the generating space, where vari is the effective rank of the feature vector's covariance. Ayush et al. also propose two kinds of FIDs based on JSD and WD distance, respectively, and then take the minima among them if the underlying probability distributions are known.

7.3. Perceptual Metrics: LPIPS, FID

The proxy perceptual distance and LPIPS both trend well with mechanics-based model benchmarks and are appropriately influenced across disparate datasets by other qualitative metrics. Unfortunately, pre-trained neural feature extractors always make LPIPS only applicable to the dataset used for model training and other settings like data augmentation might lead to incorrect values. Besides, calculating LPIPS for each generated sample can be daunting if many generator networks have to be evaluated.

The Inception model is well-known to be state-of-the-art for generating feature representations of images for various tasks in deep learning. It is shown that the Euclidean distances over these representations correlate very well with perceptual quality. LPIPS exploits that to create a particularly perceptually meaningful feature representation for images from the layers of the Inception V3. Afterwards, it utilizes these features to create a simple perceptual metric that correlates greatly with human judgment while maintaining its utility across vastly different image domains.

8. Future Directions and Emerging Trends

8.1 Emerging Trends.

Following are the highlighted emerging trends and new directions: New Generation Models: Variational Auto-Encoders (VAEs) that map images to continuous latent spaces provide effective risk modeling, and faster training than GANs but they suffer from limited resolution and mode collapse. Latest progressive growing GANs have been successful in training with a wide range of resolutions up to 1024 x 1024. Encoders preserve the generative ability of the decoder in some models that collect the image from the latent variable distribution, hence enabling GAN-generated images to outperform

VAE-generated samples, although further work is needed to identify a cross-model pointer for different models.

8.2. Identified Opportunities. We highlight the key areas that we believe will help explore the concept of editing ability of commonly used models such as GANs, lead to a higher quality synthesized image, and bring this approach from research trials to practical clinic usage. Image quality assessment, representing a challenging task in the area of biomedical images, synergizes CV and GANs and may be included in the training pipeline, improving images plausibility. It assists model selection and optimization by ranking models, regions of interests, or optimization parameters selected. With enough human data, combining known expert ratings with machine-learned image features could be designed. Furthermore, it could determine if the images synthesized look realistic enough, enhancing practical applications. Subtle salient features are either of concern or completely overlooked by existing models. The clinical importance of the different features including bone marrow, muscle, spine, tissue density, and shape makes these features of particular interest, especially in problems where these features are straightforwardly designed.

In this section, we provide an analysis of the future opportunities and highlight existing problems that have not been resolved. In addition, we illustrate several emerging trends in the intersection of CV and GANs in medical image synthesis.

8.3. Explainable AI in Medical Image Synthesis

In the work described here, we generate synthetic medical images by training a GAN. The generator in the GAN uses the supervoxel model presented before to add one lesion object model to each slice. The discriminator in the GAN focuses specifically on allowing the generator to correctly add lesions to each slice and maintain model details within and around the synthetic lesions. We demonstrate that GANs trained in this new way are capable of generating synthetic results equivalent to those available by adding supervoxels to medical images. This research direction lays the groundwork for explainable AI application in the real aspects of medical imaging.

The methods for medical image synthesis described previously implicitly perform blending of objects in the generated images. However, for medical applications, more precise control over changes in the synthesized results from data manipulation is desired. The use of conditional control of generated images would allow fine control over these variations with neural networks. Generative Adversarial Networks (GANs) provide a means to generate images of any type, including control for generating specific outputs, and work with small datasets. The combined use of supervoxels and GANs for data augmentation, without the need to generate and use supervoxels, would be a strong step forward and is partly the focus of our discussion here on the generative adversarial network. This work extends prior generative work that has only focused on cross-modality synthesis and multi-label image synthesis with GANs and conditional control.

8.4. Multi-modal Image Synthesis

Also, presenting generative adversarial networks (GANs) as a multi-modal data multi-modal synthesis problem by using them in the context of generating any combination of images that can be collected together, generalizes the domain of deep MR-CT synthesis from a specific three accessing paired acquisition to a general connected multi-modal MR-CT generation task. With the recent surge in GAN research, we will likely continue to see advances that will unlock the potential to make 3D multi-modal synthesis across CT, MRI, and PET feasible and more available for applications in medical imaging.

CT-MR image synthesis is important, as MRI provides better anatomical information and reflects better signal contrast between different tissues. However, MR imaging has the limitation of long acquisition time, and due to its direct dependence on body water and molecular structures, inferior image contrast and anatomical structure information compared to CT images in bone imaging. Thus, much research has proposed to generate CT images similar to MR images obtained with fast imaging methods. These methods of single modal-to-single modal generation have been extended by integrating multi-modal data of every modality that is supposed to be generated.

Dealing with a data set of single-modal images is not common in medical imaging. Most medical data sets used in practice contain registered images of multiple modalities, and in 3D medical data sets, such multi-modal images are often associated with each other. T1-weighted (T1-w), T2-weighted (T2-w), and proton density (PD or DIRM for double inversion recovery image) are common magnetic resonance imaging (MRI) modalities that are often used together for clinical use. Computed tomography (CT) and positron emission tomography (PET) are also used together, as the T1-w, T2-w, and PD images of different MR modalities have discriminative information intrinsic to the body, and CT/PET imaging use different physical principles of signal generation.

8.5. Integration with Clinical Decision Support Systems

To begin benefiting from computer vision and GANs recent developments of medical image synthesis, the image generation process would be integrated into existing clinical decision support (CDS) applications. Where early generation results will aid existing clinicians' interaction by querying the generative model executed to modulate their visual representations as defined in the ethical guidelines, even if the image elements of the generative model query are not image-like.

In the traditional flow of medical imaging data, images are acquired by the modality, anonymized and shared, and analyzed independently by algorithms deployed in commercial or custom-built cloud-based execution engines that have a turn-around-time latency. The results, conclusions, and algorithms coding logic are then reviewed by human-expert diagnosticians that re-review the images, call for other medical tests, seek consultations, and sometimes in collaboration with patients, formulate management plans. These can be visualized with options depending on the diagnostic center's available services such as scheduling, education, personal electronic or hardcopy reports, and a forecast of overall medical manager's plan.

9. Conclusion

Given the obtained insights, we conclude with various practical guidelines and research opportunities in aligning the promise of models with actual capabilities and design choices for advancing the medical image processing forefront. Such suggested guidelines could overwrite challenges of scaling beyond the solved problem landscape and lead to artificially intelligent enhanced enabling future discoveries when implemented through careful engineering decisions and clinical fine-tuning.

Specifically, we showed connections and commonalities across various applications and training stability mechanisms, which can encourage new directions for a range of other applications to complement the works contributed thus far. With emphasis on several cross-pollination possibilities, we reviewed intrinsic properties of maps and visualization techniques to assess what these works can generate for infix functions, extrema, corpus lowering structures such as declarative memory surface.

From our findings, we found a meeting point of consensus in the literature and provided specific recommendations and a new taxonomy with a practical view. These findings should provide significant guidance to any investigators and encourage the well-reasoned development of follow-up advancements beyond the current architectural designs. The interest in these novel, diverse, bioinspired, and augmented models is indeed an indicator of the broad applicability of such integrated systems and suggests that biomedical image synthesis problems may soon be solvable.

In this work, we summarized 49 of the most significant contributions in medical image synthesis specifically in the domain of Biomedical and Health Informatics using GAN-based models as the foundation in a structured and systematic manner for every research direction. We explored the specific techniques, including the application domains, generating type, and loss function for analysis based on a recent comprehensive survey.

10. References

1. X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, “ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2097–2106.
2. H. R. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, and R. M. Summers, “Improving computer-aided detection using convolutional neural networks and random view aggregation,” IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1170–1181, 2016.
3. D. Nie, R. Trullo, C. Petitjean, S. Ruan, P. A. Shen, and D. Shen, “Medical image synthesis with context-aware generative adversarial networks,” in International Conference on Medical Image Computing and Computer-Assisted Intervention, 2017, pp. 417–425.
4. T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, “Spectral normalization for generative adversarial networks,” arXiv preprint arXiv:1802.05957, 2018.
5. A. Karnewar and M. Wang, “MSG-GAN: Multi-Scale Gradients GAN for stable and high-resolution image synthesis,” arXiv preprint arXiv:1903.06048, 2019.
6. Z. Liu, P. Luo, X. Wang, and X. Tang, “Deep learning face attributes in the wild,” in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3730–3738.
7. M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein generative adversarial networks,” in International Conference on Machine Learning, 2017, pp. 214–223.
8. M. Mirza and S. Osindero, “Conditional generative adversarial nets,” arXiv preprint arXiv:1411.1784, 2014.
9. J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232.
10. J. Zhao, M. Mathieu, and Y. LeCun, “Energy-based generative adversarial network,” in International Conference on Learning Representations, 2017.
11. T. Karras, T. Aila, S. Laine, and J. Lehtinen, “Progressive growing of GANs for improved quality, stability, and variation,” arXiv preprint arXiv:1710.10196, 2017.
12. S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in International Conference on Machine Learning, 2015, pp. 448–456.
13. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

14. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1125–1134.
15. X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz, “Multimodal unsupervised image-to-image translation,” in Proceedings of the European Conference on Computer Vision, 2018, pp. 172–189.
16. J.-Y. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, and E. Shechtman, “Toward multimodal image-to-image translation,” in Advances in Neural Information Processing Systems, 2017, pp. 465–476.
17. A. van den Oord, N. Kalchbrenner, L. Espeholt, O. Vinyals, A. Graves, et al., “Conditional image generation with PixelCNN decoders,” in Advances in Neural Information Processing Systems, 2016, pp. 4790–4798.
18. 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, 2014, pp. 2672–2680.
19. Pulimamidi, Rahul. "To enhance customer (or patient) experience based on IoT analytical study through technology (IT) transformation for E-healthcare." *Measurement: Sensors* (2024): 101087.
20. Pargaonkar, Shravan. "The Crucial Role of Inspection in Software Quality Assurance." *Journal of Science & Technology* 2.1 (2021): 70-77.
21. Menaga, D., Loknath Sai Ambati, and Giridhar Reddy Bojja. "Optimal trained long short-term memory for opinion mining: a hybrid semantic knowledgebase approach." *International Journal of Intelligent Robotics and Applications* 7.1 (2023): 119-133.
22. Singh, Amarjeet, and Alok Aggarwal. "Securing Microservices using OKTA in Cloud Environment: Implementation Strategies and Best Practices." *Journal of Science & Technology* 4.1 (2023): 11-39.
23. Singh, Vinay, et al. "Improving Business Deliveries for Micro-services-based Systems using CI/CD and Jenkins." *Journal of Mines, Metals & Fuels* 71.4 (2023).
24. Reddy, Surendranadha Reddy Byrapu. "Big Data Analytics-Unleashing Insights through Advanced AI Techniques." *Journal of Artificial Intelligence Research and Applications* 1.1 (2021): 1-10.
25. Raparthi, Mohan, et al. "Data Science in Healthcare Leveraging AI for Predictive Analytics and Personalized Patient Care." *Journal of AI in Healthcare and Medicine* 2.2 (2022): 1-11.