Computer science
Case study: Generative AI for image creation—a diffuse vision

For use in May and November 2027

Instructions to students

- Case study booklet required for higher level and standard level paper 1 computer science examinations.

8 pages
5527 – 4420

Blank page

Introduction

Many organizations around the world have started exploring the use of generative artificial intelligence (AI) in their industries. This case study highlights the application of generative AI in the design industry and explores the techniques and models used, as well as the unique challenges presented by generative AI.

Note: Sections **Generative adversarial networks (GANs)** (lines 48–59) and **Hybrid models** (lines 60–75) are for higher level only. All other sections are for both standard level and higher level.

Scenario

Visionary Studios, a creative design company, is evaluating generative AI models to streamline the production of high-quality images for advertising campaigns, concept art and digital media. As part of this initiative, the company is exploring advanced architectures such as *diffusion models*, *generative adversarial networks (GANs)* and *hybrid models*, each offering unique capabilities in *image generation*.

Image generation techniques

Generative AI models offer various approaches to creating images, each tailored to different creative and technical requirements. Visionary Studios is considering several key image generation models as part of its evaluation:

Text-to-image generation models

Models such as DALL·E and Stable Diffusion generate images from textual prompts, enabling creators to translate detailed descriptions into visual outputs. This approach is particularly useful for concept art and advertising, where specific themes or styles are required.

***Conditional* image generation models**

Conditional models generate images based on specific inputs, such as class labels, sketches and *segmentation maps*. This ensures that outputs align with predefined objectives or artistic styles.

- Class-conditional models can generate specific categories of images, such as animals or landscapes.
- *Image-to-image translation* transforms one image type into another, such as turning sketches into realistic rendering or converting black-and-white images into colour.

Unconditional image generation

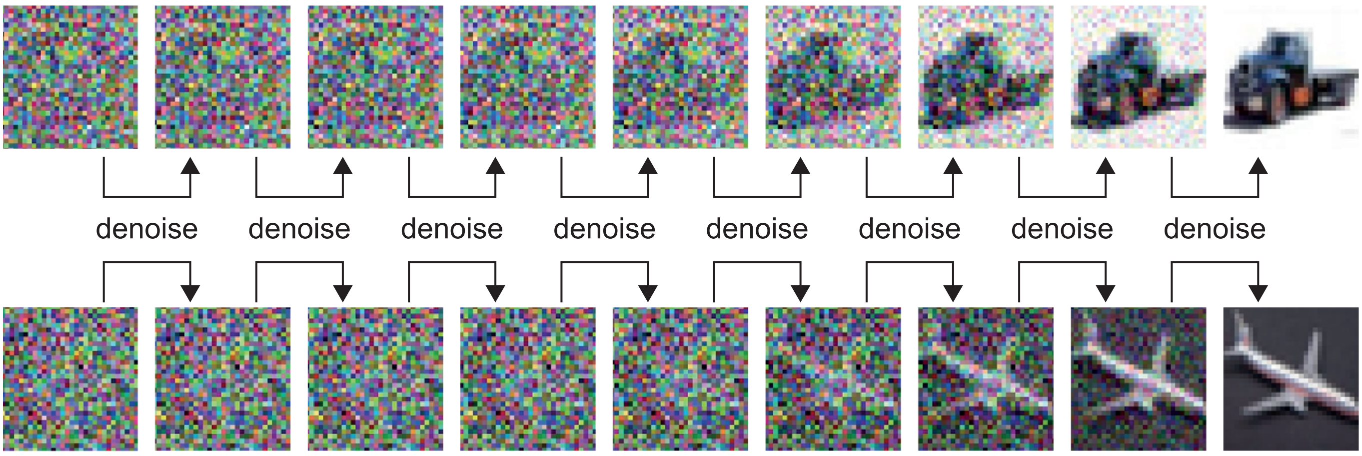
Unconditional models generate images without predefined inputs, relying solely on learned patterns from training data. This approach is useful for exploratory creativity, abstract art and generating novel, unexpected visuals. It is also valuable for creating synthetic datasets for AI training.

Diffusion models

Diffusion models generate images by leveraging an iterative denoising process, beginning with *noise injection*—introducing random noise as the starting point for image generation—which is refined over multiple steps into coherent outputs. The neural denoiser, typically a *convolutional neural network (CNN)*, gradually reconstructs features based on patterns learned during training.

The key to this process is that a denoiser can generate entirely new images by starting from noise (see **Figure 1**), with the nature of the generated content determined by the dataset used to train the model. For example, a model trained on animal images might reveal cats and dogs, while a model trained on images of cities might produce urban scenes. While diffusion models excel at producing photorealistic images, their iterative nature requires significant computational resources, making optimization of this process crucial for practical applications in creative industries.

Figure 1: The denoising process used in diffusion models

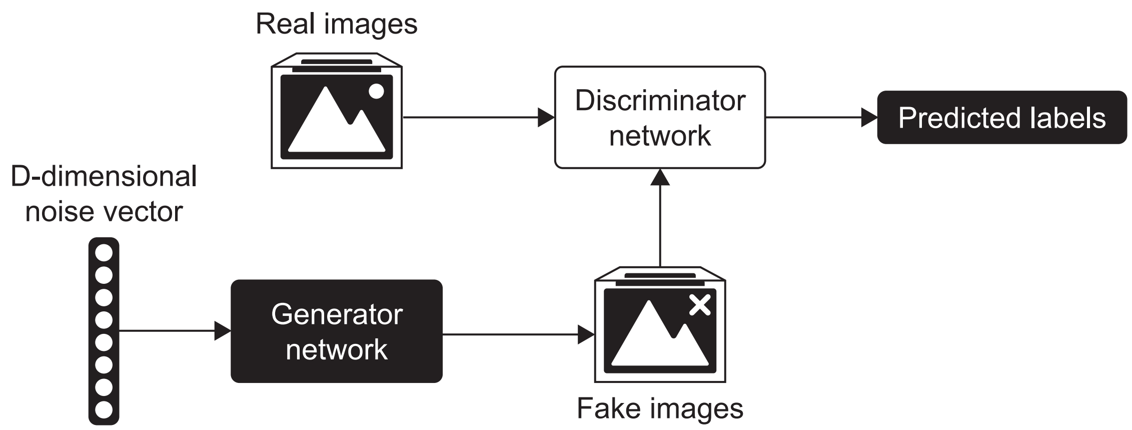


Modern diffusion models are based on the *denoising diffusion probabilistic model (DDPM)*, which formalizes the process of gradually adding and removing noise during training and generation.

Generative adversarial networks (GANs) (higher level only)

GANs create realistic images through an adversarial process involving two neural networks: a *generator* and a *discriminator*. The generator produces synthetic images by transforming a *D-dimensional noise vector* into candidate images, while the discriminator evaluates these outputs by comparing them to real images from the training dataset. The two neural networks are trained together in a competitive framework, where the generator improves by learning to deceive the discriminator and the discriminator becomes better at distinguishing real images from fake ones.

Figure 2: The GAN image creation process



This *adversarial dynamic* is key to a GAN's ability to produce sharp, high-quality images. The training process, however, is highly sensitive and prone to issues such as instability or *mode collapse*, where the generator produces limited variations instead of capturing the full diversity of the dataset. Various mitigation techniques have been developed to address this issue.

Hybrid models (higher level only)

Hybrid models combine different types of generative AI to overcome the limitations of using a single approach. By blending models together, hybrid models can produce higher-quality images and offer better control over output attributes and greater flexibility for different creative needs.

Some examples of models that might be included in a hybrid model are:

- *variational autoencoders (VAEs)*. These models learn to represent images in a *latent space*—a compressed and organized version of the original data—where similar images are placed closer together. VAEs make it possible to modify and generate images in different ways according to how the latent space is explored. VAEs are useful when flexibility, experimentation or artistic exploration is important.
- *flow-based models*. These models learn exact mapping between random noise and realistic images, and this mapping can be reversed without loss. This allows for the generation of new images and the ability to trace how each image was formed. Flow-based models are useful when precision, reproducibility or transparency in the generation process is important.

By combining the strengths of VAEs, flow-based models, GANs and diffusion models, hybrid models can provide more reliable, consistent and high-quality image generation systems.

Evaluating generative AI models

Visionary Studios aims to assess the feasibility and impact of integrating generative AI by considering several key factors:

- Output quality: How effectively can the model produce high-resolution, photorealistic, stylistically coherent images that align with project goals?
- Computational efficiency: Can the company's existing infrastructure support the training and deployment of the chosen model without significant delays or resource constraints?
- *Training stability*: Training generative models can be difficult to stabilize. Some models may produce repetitive or suboptimal outputs, while others train more reliably but require longer training times and greater computational resources.
- Flexibility and scalability: How adaptable is the model to different use cases, such as advertising campaigns, concept art and branded assets? How scalable is the model to the demands of future creative projects?
- Consistency: Ensuring consistent outputs is critical for maintaining brand identity and artistic coherence. This includes in:
    - *character consistency*. Models must be capable of generating uniform appearances for recurring characters or motifs across multiple images. Embedding-based approaches can help achieve this consistency.
    - style adherence. Conditional models, such as text-to-image generators, must ensure that generated images align with the intended artistic style or theme specified by prompts.

Ethical and legal considerations

The integration of generative AI raises ethical and legal considerations that Visionary Studios cannot overlook:

- *Dataset curation*: Training datasets must be carefully curated to avoid the inclusion of copyrighted material and ensure compliance with intellectual property laws.
- Bias and fairness: Visionary Studios must assess datasets for potential biases that could result in exclusionary or inaccurate representations in generated outputs.
- Transparency and AI disclosure: Visionary Studios should establish clear guidelines for disclosing the use of AI-generated content in its creative process to maintain trust with clients and audiences.

Challenges faced

The goal of Visionary Studios is to address the following challenges.

Standard level and higher level:
- Managing the iterative denoising process in diffusion models and addressing their computational demands for generating photorealistic images.
- Navigating the ethical considerations of using large datasets for training generative AI models, focusing on intellectual property and *bias mitigation*.

Higher level only:
- Balancing the performance of generator and discriminator networks in GANs to produce realistic images.
- Evaluating how a hybrid model could balance the strengths and weaknesses of VAEs, GANs, flow-based models and diffusion models.

Standard level students are not expected to know about GANs or hybrid models.

Higher level students are expected to focus deeply on diffusion models, GANs and hybrid models, while a broad understanding of VAEs and flow-based models is sufficient.

Additional terminology

Standard level and higher level

Bias mitigation
Character consistency
Convolutional neural network (CNN)
Dataset curation
Denoising
Denoising diffusion probabilistic model (DDPM)
Diffusion model
Embedding-based approach
Image generation
- Conditional
- Unconditional

Image-to-image translation
Noise injection
Segmentation map
Text-to-image generation
Training stability

Higher level only

Adversarial dynamic
D-dimensional noise vector
Generative adversarial network (GAN)
- Discriminator
- Generator

Hybrid model
- Flow-based model
- Variational autoencoder (VAE)

Latent space
Mode collapse

References:

For assessment purposes, all acknowledgements of third-party material are stored and published separately.