

Deep Generative Models (AI314)

Master of Science in AI



Dr. Nibras Abdullah

*Assistant professor
Faculty of Informatics
Midocean University*



Lecture 7

Advanced Generative Models-2

Introduction



StyleGAN



Weight Modulation and Demodulation



Path Length Regularization



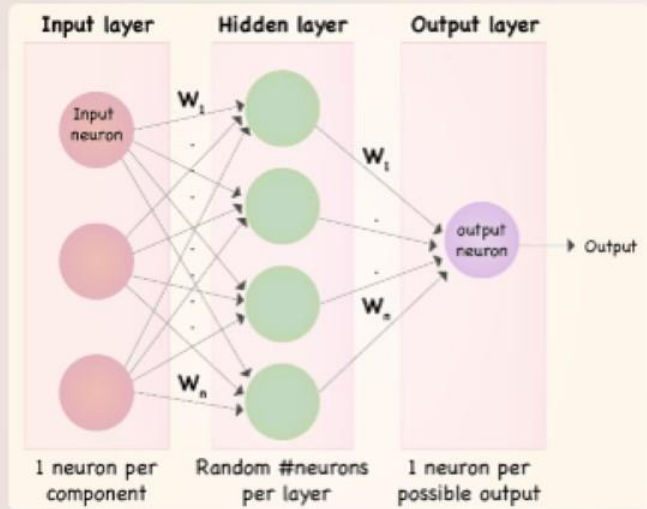
No Progressive Growing



Outputs from StyleGAN2



StyleGAN: A Deep Dive



StyleGAN3 is a GAN architecture from 2018 that builds on the earlier ideas in the ProGAN paper. In fact, the discriminator is identical; only the generator is changed. Often when training GANs it is difficult to separate out vectors in the latent space corresponding to high-level attributes—they are frequently entangled, meaning that adjusting an image in the latent space to give a face more freckles, for example, might also inadvertently change the background color. While ProGAN generates fantastically realistic images, it is no exception to this general rule. We would ideally like to have full control of the style of the image, and this requires a disentangled separation of features in the latent space.

StyleGAN's Disentangled Features

StyleGAN achieves this by explicitly injecting style vectors into the network at different points: some that control high-level features (e.g., face orientation) and some that control low-level details (e.g., the way the hair falls across the forehead).

1 High-Level Features

Control aspects like face orientation, overall shape, and general appearance.

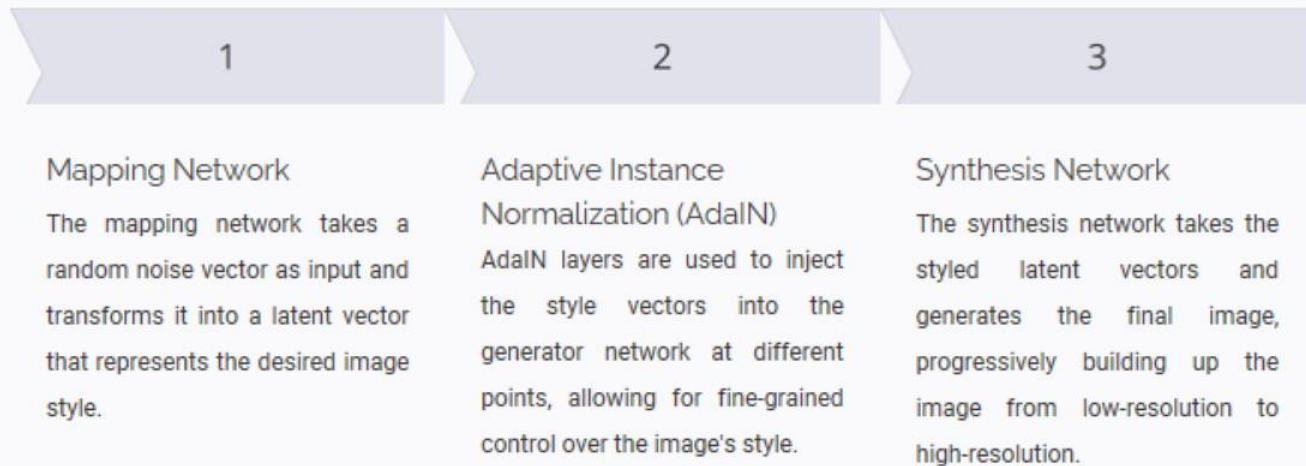
2 Low-Level Details

Influence specific details like hair texture, skin tone, and subtle facial expressions.

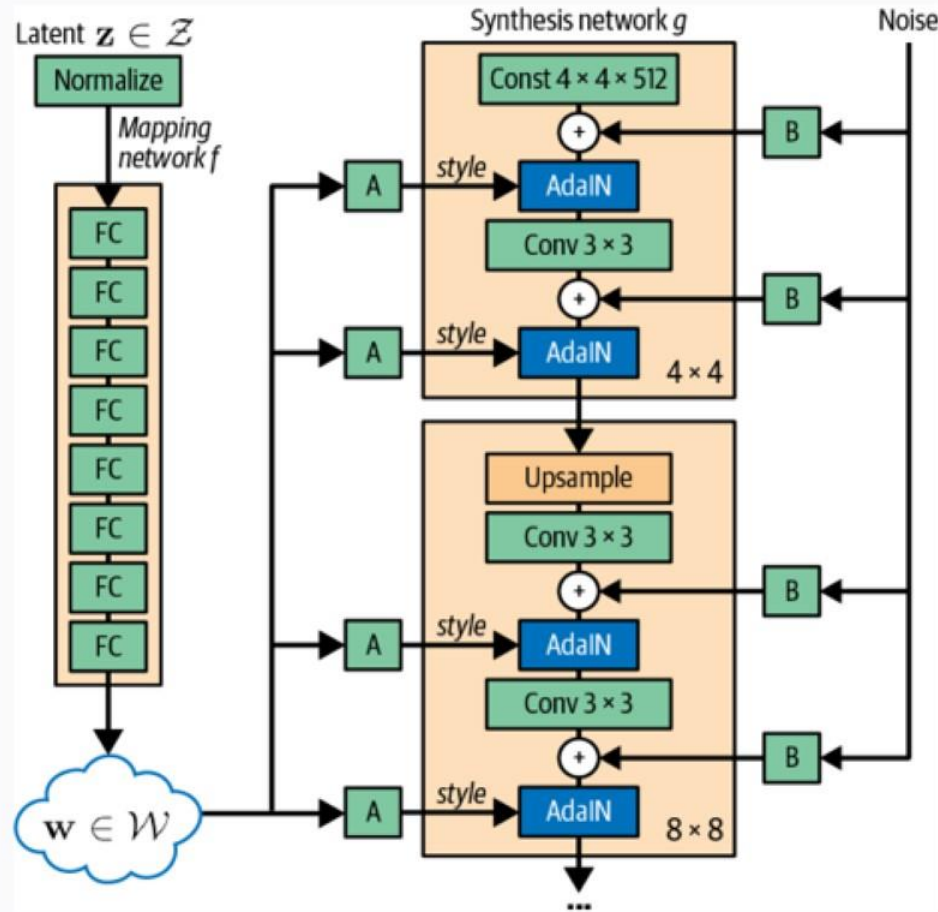


The StyleGAN Generator Architecture

The overall architecture of the StyleGAN generator is shown in the next Figure



The StyleGAN Generator Architecture



Mapping Network: The Foundation of Style

The mapping network is a series of fully connected layers that takes a random noise vector as input and transforms it into a latent vector that represents the desired image style. This latent vector is then used to control the style of the image at different points in the generator network.

Random Noise

The mapping network starts with a random noise vector. This vector is essentially a random set of numbers that serves as the initial input to the network.

Fully Connected Layers

The noise vector is then passed through a series of fully connected layers. These layers learn to transform the random noise into a meaningful latent vector that represents the desired image style.

Latent Vector

The output of the mapping network is a latent vector. This vector contains information about the desired image style, such as the overall shape, the presence of certain features, and the general appearance of the image.

Adaptive Instance Normalization (AdaIN): Injecting Style

AdaIN layers are used to inject the style vectors into the generator network at different points, allowing for fine-grained control over the image's style. AdaIN layers work by normalizing the input features to have zero mean and unit variance, and then scaling and shifting the normalized features using the style vector.

Layer	Function
Normalization	Ensures that the input features have a consistent distribution, regardless of the input image.
Scaling and Shifting	The style vector is used to scale and shift the normalized features, allowing for fine-grained control over the image's style.

Synthesis Network: Building the Image

The synthesis network takes the styled latent vectors and generates the final image, progressively building up the image from low-resolution to high-resolution. This progressive approach allows the network to learn complex image structures and generate highly realistic images.

1

Low-Resolution Image

The synthesis network starts by generating a low-resolution image. This initial image is a rough representation of the final image, with basic shapes and features.

2

Progressive Upsampling

The network then progressively upsamples the image, adding more detail and complexity at each stage. This process continues until the final, high-resolution image is generated.

3

High-Resolution Image

The final output of the synthesis network is a high-resolution image that is highly realistic and captures the desired style.



StyleGAN's Impact on Image Generation

StyleGAN has had a significant impact on the field of image generation. It has enabled the creation of highly realistic and diverse images, and it has opened up new possibilities for creative applications.

Artistic Expression

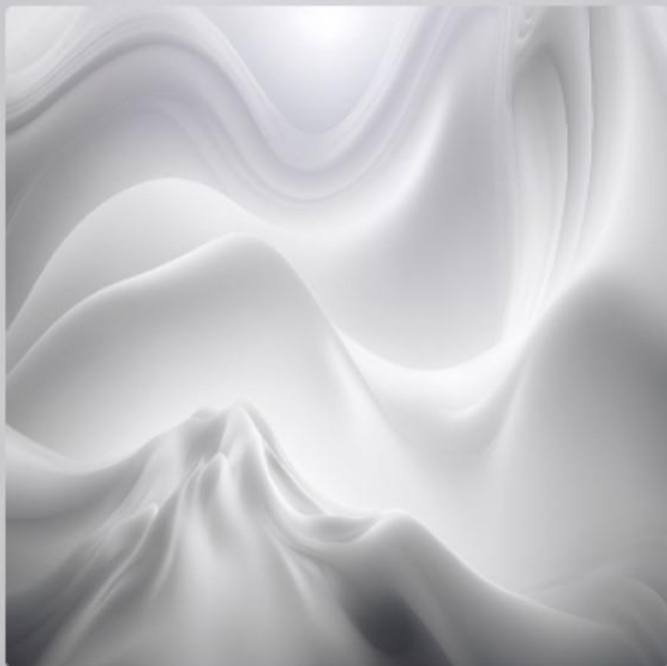
Artists can use StyleGAN to create unique and expressive artwork, exploring different styles and concepts.

Game Development

Game developers can use StyleGAN to generate realistic characters, environments, and assets, enhancing the visual fidelity of their games.

Film and Animation

Filmmakers and animators can use StyleGAN to create realistic characters, environments, and special effects, pushing the boundaries of visual storytelling.



StyleGAN: A Powerful Tool for Image Generation

StyleGAN is a powerful tool for image generation that has revolutionized the field. Its ability to disentangle features in the latent space and provide fine-grained control over image style has opened up new possibilities for creative applications.

Realistic Images

StyleGAN generates highly realistic images that are indistinguishable from real photographs.

Diverse Styles

StyleGAN can generate images in a wide range of styles, from photorealistic to abstract.

Control over Features

StyleGAN allows users to control specific features of the generated images, such as the pose, expression, and lighting.





The Future of StyleGAN

StyleGAN is a rapidly evolving technology, and its future is bright. Researchers are constantly working to improve its capabilities and explore new applications. As StyleGAN continues to develop, it is likely to have an even greater impact on the field of image generation and beyond.

1 Improved Realism

Researchers are working to improve the realism of StyleGAN-generated images, making them even more indistinguishable from real photographs.

2 New Applications

StyleGAN is being explored for a wide range of new applications, such as medical imaging, scientific visualization, and even art restoration.

3 Ethical Considerations

As StyleGAN becomes more powerful, it is important to consider the ethical implications of its use, such as the potential for misuse in creating deepfakes or generating harmful content.

Style Transfer: A Deep Dive



In the fascinating world of deep learning, style transfer has emerged as a powerful technique that allows us to seamlessly blend the artistic style of one image onto the content of another. This process often involves a two-step approach, employing separate networks for style extraction and content generation. Today, we delve into the architecture and mechanisms behind these networks, exploring the intricacies of style transfer and its remarkable capabilities.

The Mapping Network: Unraveling Style

The mapping network serves as the initial stage in style transfer, tasked with extracting and representing the stylistic essence of a given image. It takes as input a noise vector, often drawn from a random distribution like Gaussian noise. This noise vector acts as a starting point for the network, allowing for a degree of randomness and variability in the style representation.

This network's architecture typically consists of multiple layers, including fully connected layers and activation functions, that learn to map the input noise vector into a different latent space. This latent space, often referred to as the style space, captures the essential style features of the input image, transforming the random noise into a meaningful representation of the style.

Input Noise

The mapping network's journey begins with a noise vector drawn from a random distribution. This random input provides a foundation for the network's exploration of the style space.

Mapping Network Layers

The network's core lies in its layers, typically fully connected or convolutional, which learn to process the input noise vector. These layers gradually transform the random noise into a structured representation of the style.

Output Style Vector

The network's final output is a style vector, a compact representation of the extracted style. This vector captures the essential stylistic features of the input image, ready to be passed on to the synthesis network.

Disentanglement: Separating Style and Content

The mapping network plays a crucial role in disentangling the style from the content of an image. This disentanglement process is essential for allowing the style transfer process to selectively apply a specific style to any desired content image. The disentanglement achieved through the mapping network enables the separation of the style-choosing process from the image generation process.

Style Variation

- 1 The mapping network enables the exploration of a wide range of styles. By manipulating the input noise, we can generate different style vectors, corresponding to various artistic styles. This allows for flexibility in style selection, catering to diverse artistic preferences.

Independent Control

The mapping network facilitates independent control over style. The style vector, generated by the network, can be used to adjust the style of the output image without altering the content. This provides a powerful mechanism for manipulating the style of an image.

Content Preservation

By disentangling style from content, the mapping network enables the preservation of the original image's content while applying a new style. This ensures that the content of the image remains intact, while its visual appearance is transformed by the chosen style.



The Synthesis Network: Building Style-Infused Images

The synthesis network takes center stage in the style transfer process, responsible for generating the final image with the desired style. It receives the style vector from the mapping network and utilizes it to guide the image generation process. The synthesis network's core lies in its ability to incorporate the style vector into its internal operations, effectively transforming the content of the image while preserving its structural integrity.

The synthesis network often consists of multiple convolutional layers, each responsible for processing different levels of detail in the generated image. To inject the extracted style, the network incorporates adaptive instance normalization (AdaIN) layers. These layers, placed strategically within the synthesis network, allow for dynamic adjustment of the image's features based on the style vector.

Input Content

**1**

The synthesis network receives the content image as input. This content serves as the foundation upon which the desired style will be applied.

**2**

Style Injection

The style vector from the mapping network is injected into the synthesis network at specific layers, influencing the image's feature maps.

**3**

Synthesis Layers

The synthesis network's layers process the content image and the injected style, gradually building the final image with the desired style.

Adaptive Instance Normalization: Style Injection



Adaptive instance normalization (AdaIN) plays a pivotal role in style transfer, enabling the seamless integration of style into the image generation process. This layer allows for dynamic adjustment of the image's features based on the style vector, effectively transforming the image's appearance while preserving its underlying structure. AdaIN works by normalizing the features of the image, ensuring that they have a consistent distribution across different layers. This normalization is crucial for stabilizing the training process and preventing overfitting. However, the key to style transfer lies in the adaptive nature of AdaIN, which allows for adjustments to the normalized features based on the extracted style.

Normalizing Features

AdaIN first normalizes the features of the image, ensuring a consistent distribution across layers.

Scaling and Bias

The AdaIN layer then applies a scaling and bias operation based on the style vector, adjusting the normalized features to match the desired style.

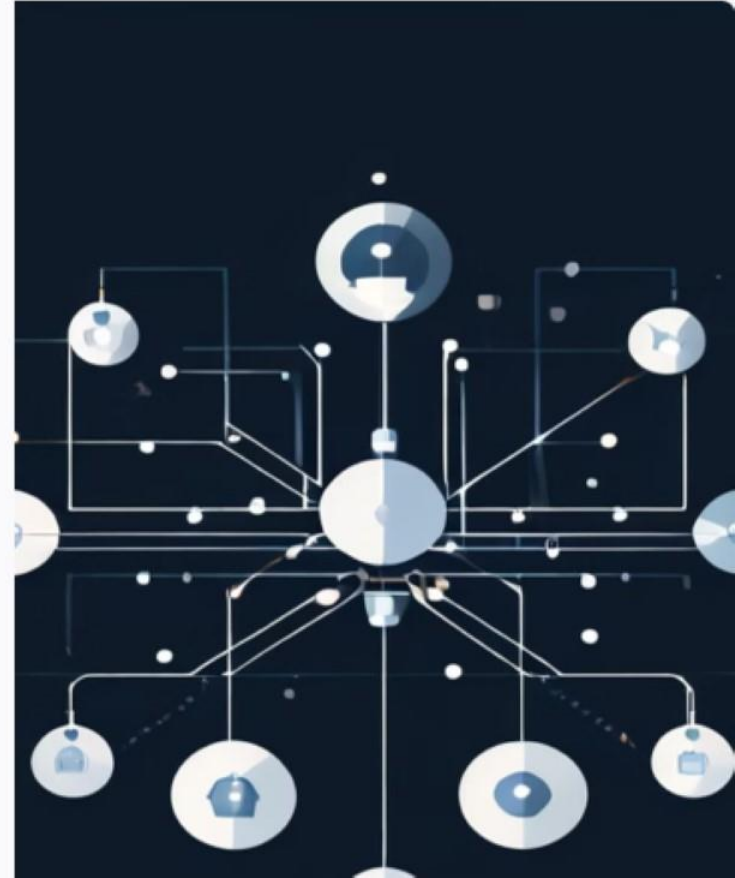
Style Transformation

The scaling and bias operations effectively transform the image's features, aligning them with the stylistic characteristics captured by the style vector.

Style Vector Injection: Shaping the Image's Features

The style vector, carrying the essence of the chosen style, is injected into the synthesis network at different points. These injection points are strategically placed within the network, targeting specific layers that process different levels of detail in the generated image. Each injection point corresponds to a differently densely connected layer, which generates two vectors: a bias vector and a scaling vector. These vectors define the specific style that should be injected at this point in the network.

The bias vector acts as a shift, adjusting the mean of the features, while the scaling vector acts as a stretch, adjusting the variance of the features. By manipulating these vectors, the style vector controls the specific style characteristics that are injected into the image at each layer. This approach allows for a fine-grained control over the style transformation, enabling the network to adapt the style to different content images and achieve a wide range of stylistic variations.



The Power of Style Transfer: Artistic Fusion

Style transfer holds immense potential in the realm of creative expression, enabling the fusion of artistic styles onto existing content. Its ability to seamlessly blend styles onto different images opens up a world of possibilities for artistic experimentation and exploration. The versatility of style transfer extends beyond traditional artistic styles, encompassing diverse visual effects and artistic interpretations.

Style transfer finds applications in various fields, from artistic creation and image editing to user interface design and content personalization. It empowers artists to experiment with new styles, photographers to enhance their images, and designers to create visually compelling user experiences. Style transfer is a powerful tool that bridges the gap between content and style, unlocking a world of artistic possibilities and fostering creative innovation.



The Power of Style Transfer: Artistic Fusion

Artistic Expression

Style transfer allows artists to experiment with different artistic styles, creating unique and captivating works of art.

Image Enhancement

Photographers can utilize style transfer to enhance their images, applying artistic filters and special effects to elevate their work.

User Interface Design

Designers can leverage style transfer to create visually appealing user interfaces, enhancing the user experience with unique aesthetics.

Content Personalization

Style transfer can be used to personalize content, allowing users to customize their visual experiences based on their individual preferences.

Beyond the Basics: Exploring Advanced Techniques

Style transfer, as a field of active research, continues to evolve with the emergence of advanced techniques that expand its capabilities and address various challenges. These advanced techniques aim to refine the style transfer process, improving its accuracy, efficiency, and creative potential.

1

Neural Style Transfer

Neural style transfer employs deep neural networks to learn the style from an image and apply it to a content image, offering a more sophisticated approach to style transfer.

2

Multi-Style Transfer

Multi-style transfer allows for the simultaneous application of multiple styles to a single image, creating intricate and visually captivating results.

3

Real-Time Style Transfer

Real-time style transfer enables the application of styles to images in real-time, facilitating interactive experiences and applications for live video processing.

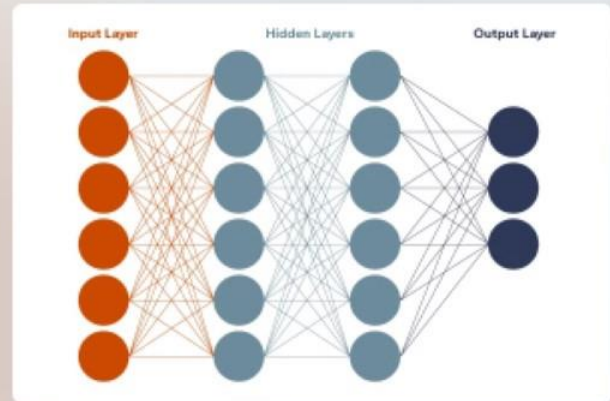


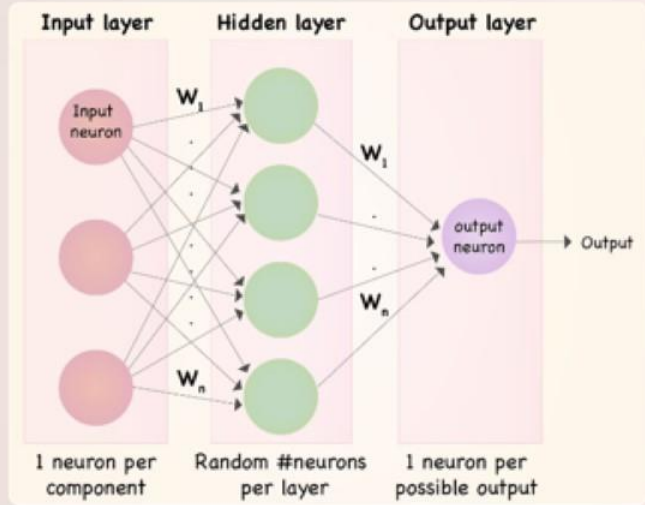
The Future of Style Transfer: Unlocking New Possibilities

The future of style transfer holds immense promise, with ongoing research and development pushing the boundaries of this powerful technique. Advancements in deep learning and computer vision are expected to further enhance the capabilities of style transfer, unlocking new possibilities for creative expression and technological innovation. Style transfer is poised to play a significant role in shaping the future of visual content creation, empowering artists, designers, and developers to explore uncharted territories of visual aesthetics. As the technique continues to evolve, we can anticipate exciting developments and applications that will revolutionize the way we create, manipulate, and interact with visual content.

Adaptive Instance Normalization (AdaIN) for Style Transfer

Adaptive instance normalization (AdaIN) is a powerful technique used in style transfer, enabling neural networks to effectively separate content and style information in images. This method allows for the manipulation of style in images while preserving the original content, leading to impressive results in image generation and artistic transformations.



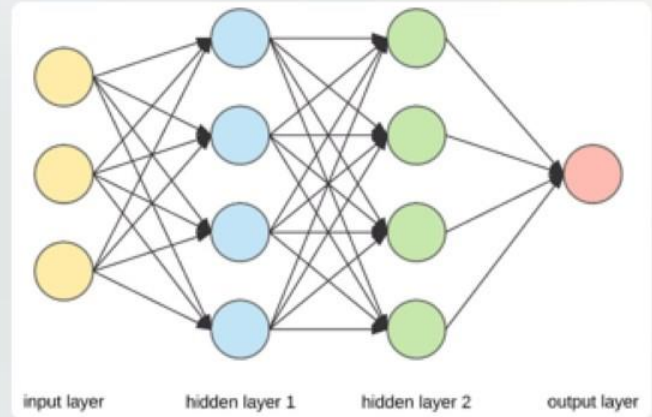


AdaIN Layer: Adjusting Mean and Variance

An AdaIN layer is a type of neural network layer that adjusts the mean and variance of each feature map i with a reference style bias b , μ and scale s , σ , respectively.⁴ Both vectors are of length equal to the number of channels output from the preceding convolutional layer in the synthesis network. The equation for adaptive instance normalization is as follows: AdaIN

Style Vector Injection: Controlling Feature Transformations

The adaptive instance normalization layers ensure that the style vectors that are injected into each layer only affect features at that layer, by preventing any style information from leaking through between layers. The authors show that this results in the latent vectors being significantly more disentangled than the original vectors.



Progressive Training: Controlling Style at Different Levels of Detail

Since the synthesis network is based on the ProGAN architecture, it is trained progressively. The style vectors at earlier layers in the synthesis network (when the resolution of the image is lowest— 4×4 , 8×8) will affect coarser features than those later in the network (64×64 to $1,024 \times 1,024$ -pixel resolution). This means that not only do we have complete control over the generated image through the latent vector, but we can also switch the vector at different points in the synthesis network to change the style at a variety of levels of detail.





Style Mixing: Breaking Correlations for Disentangled Styles

The authors use a trick known as style mixing to ensure that the generator cannot utilize correlations between adjacent styles during training (i.e., the styles injected at each layer are as disentangled as possible). Instead of sampling only a single latent vector, two are sampled (Z_1, Z_2), corresponding to two style vectors (W_1, W_2). Then, at each layer, either W_1 or W_2 is chosen at random, to break any possible correlation between the vectors.

AdaIN: A Key Component in Style Transfer

1

Content Preservation

AdaIN allows for the transfer of style without altering the original content of an image. This is achieved by separating the content and style representations within the network.

2

Style Control

The ability to inject style vectors at different layers of the network provides granular control over the style of the generated image, allowing for fine-grained adjustments.

3

Disentangled Representations

Style mixing techniques ensure that the style vectors are disentangled, preventing unwanted correlations and leading to more expressive and controllable style transfer.

4

Progressive Training

The progressive training approach allows for the control of style at different levels of detail, enabling the creation of images with both global and local style variations.



Applications of AdaIN

Artistic Style Transfer

Transforming images into different artistic styles, such as painting, photography, or graphic design.

Image Generation

Creating new images with specific styles and content, enabling the generation of realistic and stylized images.

Image Editing

Modifying the style of existing images, allowing for creative adjustments and enhancements.

Advantages of AdaIN

Improved Control

AdaIN provides fine-grained control over the style of generated images, allowing for precise adjustments and creative exploration.

Disentangled Representations

Style mixing techniques ensure that the style vectors are disentangled, leading to more expressive and controllable style transfer.

Efficient Training

AdaIN is computationally efficient and can be trained effectively on large datasets, enabling the development of powerful style transfer models.



Future Directions for AdaIN

Future research in AdaIN could focus on exploring new techniques for style disentanglement, improving the efficiency of style transfer, and developing novel applications for this powerful technique. The potential for AdaIN to revolutionize image generation and manipulation is vast, and ongoing research is likely to lead to even more impressive and creative results.

Stochastic Variation in Generative Adversarial Networks

Generative Adversarial Networks (GANs) have revolutionized the field of image synthesis, enabling the creation of remarkably realistic and diverse images. One key aspect of this success lies in the ability of GANs to incorporate stochastic variation into their image generation process. Stochastic variation introduces random elements into the network's output, contributing to the rich diversity and natural variability seen in real-world images.



The Role of Noise

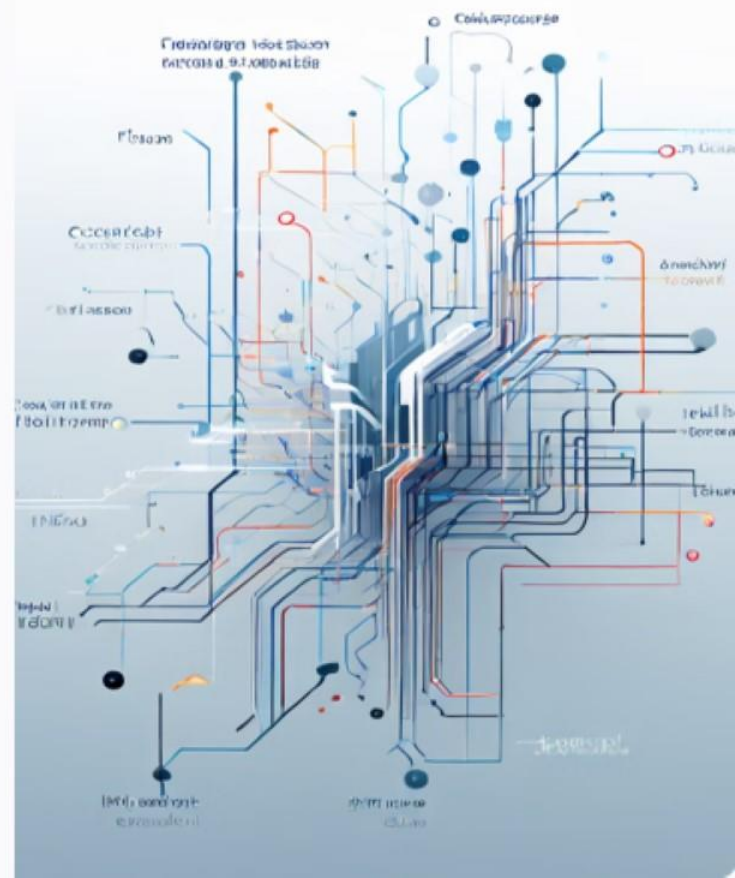
A core element of StyleGAN's stochastic variation strategy is the injection of noise after each convolutional layer. This noise, passed through a learned broadcasting layer (B), simulates the presence of minute, stochastic details in the real world. These details, often imperceptible to the naked eye, contribute to the subtle nuances and variations that make real-world images so compelling. Think of the subtle variations in hair placement, the uneven texture of a surface, or the randomness of a background. These details, collectively captured as "noise" by the network, are crucial in creating the realistic look and feel of generated images.

Noise Depth and Control

1 The depth at which noise is injected into the network directly affects the scale and impact of its influence. Deeper layers, responsible for capturing larger-scale features, experience noise that contributes to more significant variations, such as variations in pose or overall shape. On the other hand, noise injected at shallower layers, responsible for fine-grained details, primarily influences subtle aspects like texture and microstructure.

2 Learned Constant Inputs

The inclusion of noise at various levels within the network allows StyleGAN to start with a simple, learned constant as the initial input to the synthesis network. This constant serves as a starting point, and the stochasticity inherent in the style inputs and the noise inputs themselves provides enough variation to generate a wide range of distinct images.



StyleGAN in Action: Merging Styles

One of the most remarkable features of StyleGAN is its ability to merge styles between different images. This capability is illustrated in Figure 10-9, where two generated images, Source A and Source B, are created from distinct input vectors. The merging process involves passing the Source A vector through the synthesis network, but at a specific point, switching to the Source B vector. This switch point determines the level of detail at which the styles are merged.

Early Switch: Coarse Features

Switching later in the network, at higher resolutions, results in the transfer of only fine-grained details from Source B onto Source A. This might involve merging colors, texture variations, or microstructural features of the face, while preserving the coarse features of Source A. Essentially, this creates an image with the overall structure and shape of Source A but with subtle details borrowed from Source B.

Late Switch: Fine-grained Details

Switching early in the network, at resolutions like 4x4 or 8x8, merges coarse features from Source B onto Source A. This means that aspects like pose, face shape, and the presence of glasses from Source B are transferred, while Source A retains its original fine-grained details. The result is a blended image where the overall shape and structure are taken from Source B, but the finer details remain from Source A.

The Impact of Stochasticity: Realism and Diversity

The inclusion of stochastic variation plays a crucial role in the realism and diversity of images generated by StyleGAN. The random noise introduced at various levels in the network creates subtle variations that mimic the inherent randomness found in real-world images. This translates into a wider range of outputs, with each generated image exhibiting unique details

Hair Placement and Texture

Noise injection contributes to the realistic rendering of hair, capturing the subtle variations in hair placement, thickness, and texture that differentiate individual strands.

Surface Texture and Microstructure

The noise component helps create natural-looking textures on surfaces, capturing the fine-grained variations and irregularities that give objects a realistic appearance.

Background Complexity and Depth

Stochastic variation can also be used to generate complex and realistic backgrounds, simulating the random arrangement of objects and details that contribute to the depth and richness of a scene.



Controlling Stochasticity: Fine-tuning Image Generation

While stochastic variation is essential for generating realistic and diverse images, it's also crucial to have control over the level of randomness introduced into the generation process. StyleGAN provides mechanisms for fine-tuning the level of stochastic variation by allowing users to adjust the noise level at different layers within the network. This gives users a greater degree of control over the specific types of variations that are incorporated into the final generated images.

Noise Strength

Adjusting noise strength allows users to control the overall level of stochasticity introduced into the network. Higher noise levels lead to greater variations in the generated images, while lower noise levels result in more deterministic and less varied outputs.

Noise Filter

By applying filters to the noise, users can influence the types of variations introduced into the network. For example, filters can be used to emphasize certain frequency ranges within the noise, leading to specific types of variations in the generated image.

Noise Frequency

Adjusting the frequency of the noise allows users to control the scale of the variations introduced into the network. Higher frequencies result in finer-grained variations, while lower frequencies lead to larger-scale variations in the generated image.

Applications of Stochastic Variation in GANs

The ability to incorporate stochastic variation into GANs has opened up a wide range of applications across various domains. From creating realistic and diverse images for use in art, design, and entertainment to generating synthetic data for training machine learning models, stochastic variation in GANs has become an indispensable tool in many areas.

Image Synthesis and Manipulation

StyleGAN, with its ability to control stochasticity, has become a powerful tool for generating high-quality images, manipulating existing images, and creating new visual styles. This has led to applications in art generation, photo editing, and even the creation of personalized avatars.

Data Augmentation and Training

The ability to generate realistic synthetic data using GANs is a boon for machine learning, especially in domains where real data is limited or expensive to acquire. Stochastic variation ensures that the synthetic data generated is diverse and representative of real-world variations, leading to more robust and accurate machine learning models.

Artistic Exploration and Creativity

By introducing random elements into the creative process, stochastic variation has enabled artists and designers to explore new styles, create unique visuals, and push the boundaries of artistic expression. This has led to the development of innovative artistic tools and techniques powered by GANs.



Stochastic Variation: A Key to Realistic Image Synthesis

The inclusion of stochastic variation in GANs is a testament to the power of incorporating real-world randomness into generative models. By simulating the subtle variations present in real-world images, stochastic variation enables GANs to create remarkably realistic and diverse outputs. This capability has opened up exciting possibilities across various fields, from art and design to machine learning and data augmentation. As GAN research continues to evolve, we can expect to see even more sophisticated and powerful ways of leveraging stochastic variation to push the boundaries of image synthesis and generation.

Noise Injection

- 1 The introduction of noise at various levels within the network adds random variations, simulating the subtle details and imperfections present in real-world images.

Style Merging

- 2 The ability to merge styles from different images allows for the creation of unique combinations of features and attributes, leading to greater diversity and creativity in image generation.

Fine-grained Control

- 3 The ability to adjust noise levels and filter parameters provides fine-grained control over the level and types of stochastic variations introduced into the network.

Future Directions in Stochastic Variation

The field of stochastic variation in GANs is constantly evolving, with new research exploring ways to enhance the realism, diversity, and controllability of generated images. Future directions in this area include exploring new types of noise models, developing more sophisticated methods for style transfer and manipulation, and investigating the use of stochastic variation for generating high-resolution images and complex 3D scenes.

Area	Future Directions
Noise Models	Developing more advanced noise models that capture the intricate details and variations present in real-world scenes.
Style Transfer	Improving techniques for style transfer and manipulation, enabling users to seamlessly blend and transfer styles between different images.
High-Resolution Synthesis	Extending stochastic variation to generate high-resolution images with intricate details and realistic textures.
3D Scene Generation	Exploring the use of stochastic variation to generate complex 3D scenes, including realistic objects, environments, and lighting.



The Impact on Computer Vision and Beyond

The continued advancements in stochastic variation techniques within GANs have profound implications for computer vision and related fields. These advances are driving the development of more powerful and versatile tools for image synthesis, manipulation, and analysis. As the field continues to progress, we can expect to see even more transformative applications of stochastic variation in image generation, leading to new breakthroughs in art, design, entertainment, and scientific research.

1

Enhanced Realism and Diversity

The ability to generate more realistic and diverse images with finer-grained control over stochastic variation will lead to more compelling and immersive experiences in various applications, including virtual reality, gaming, and animation.

2

Data Augmentation and Synthetic Data Generation

The use of stochastic variation in GANs will continue to revolutionize data augmentation techniques, enabling the creation of synthetic datasets for training machine learning models in areas where real data is scarce or expensive to obtain.

3

New Artistic and Design Tools

The ability to control and manipulate stochastic variations in images will open up new possibilities for artists and designers, leading to the creation of innovative tools and



Thank You