Applications of Generative Adversarial Networks in Single Image Datasets

Dylan E. Cramer
crame160@morris.umn.edu
Division of Science and Mathematics
University of Minnesota, Morris
Morris, Minnesota, USA

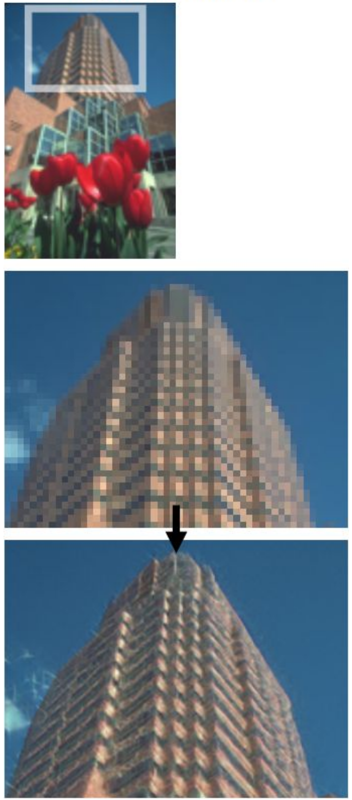
Introduction: Applications and Motivations

- Create new images
    - Learn what qualities make an image appealing

Introduction: Applications and Motivations

- Create new images
    - Learn what qualities make an image appealing
- Restore lost detail to an image with super resolution

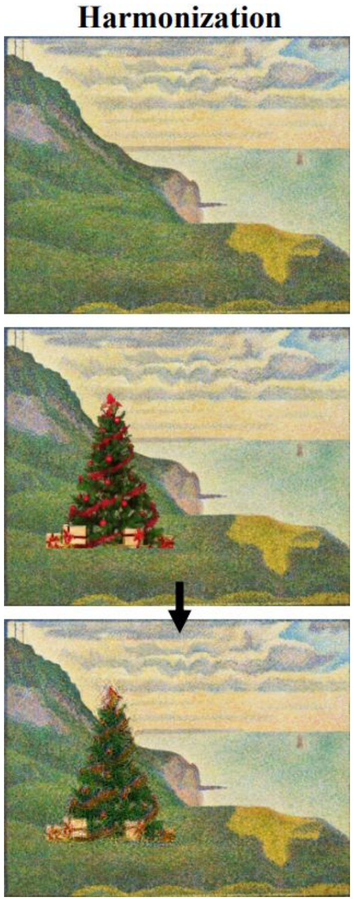
Super-resolution



Source: Shaham (2019)

Introduction: Applications and Motivations

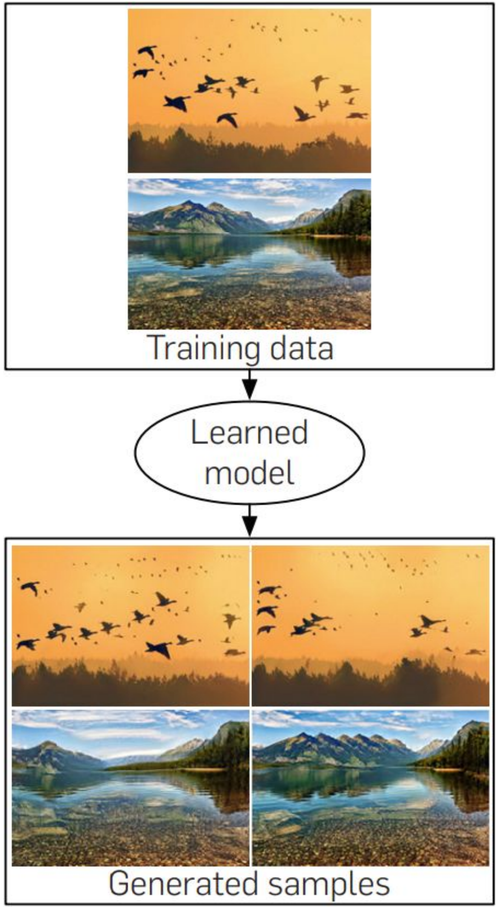
- Create new images
  - Learn what qualities make an image appealing
- Restore lost detail to an image with super resolution
- Image harmonization



Source: Shaham (2019)

Introduction: The Generative Problem

- Study a collection of training examples
- Learn the probability distribution that generated them
- Generate new examples that are indistinguishable from the real ones



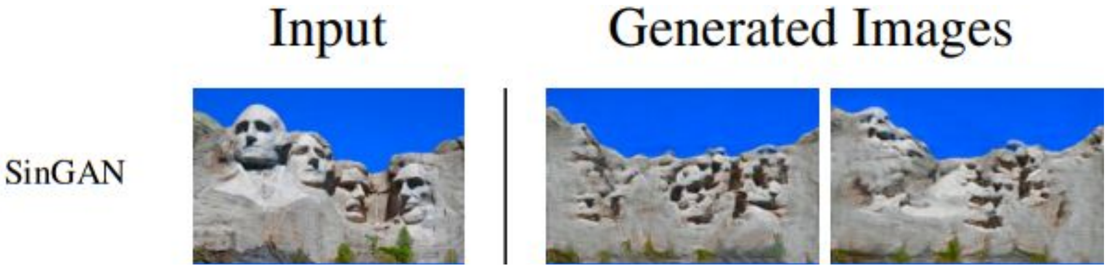
Source: Goodfellow (2020) & Shaham (2019)

Introduction: ConSinGAN

- Single image

Introduction: ConSinGAN

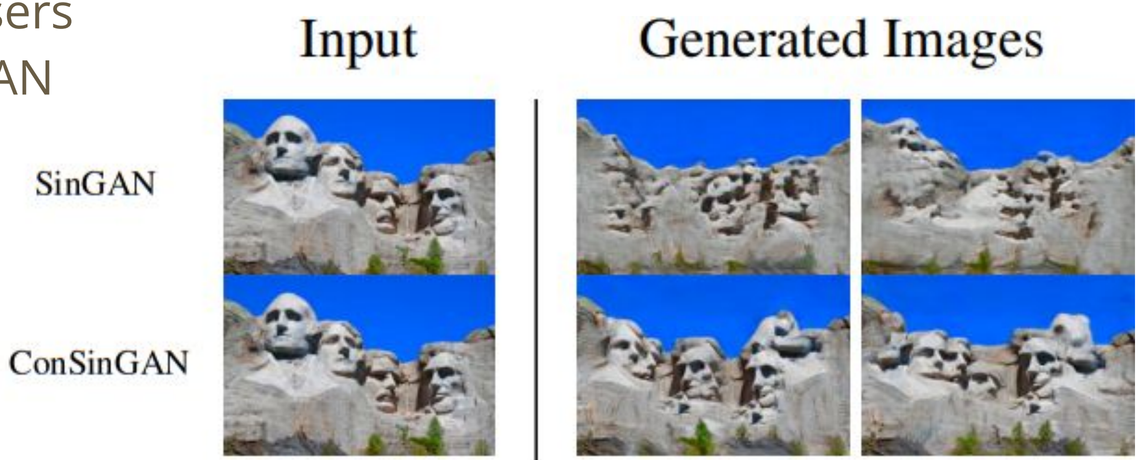
- Single image
- SinGAN (Single Image Generative Adversarial Network)



Source: Hinz (2020)

Introduction: ConSinGAN

- Single image
- SinGAN (Single Image Generative Adversarial Network)
- Concurrent SinGAN (ConSinGAN): faster and preferred by users over its predecessor SinGAN



Source: Hinz (2020)

Outline

1. Background
    a. Machine Learning
    b. Neural Networks
        i. Convolutional Neural Networks (CNNs)
    c. Generative Adversarial Networks (GANs)
2. ConSinGAN
    a. Image Generation
        i. Methods
        ii. Results
    b. Image Harmonization
        i. Methods
        ii. Results
3. Conclusion

Outline

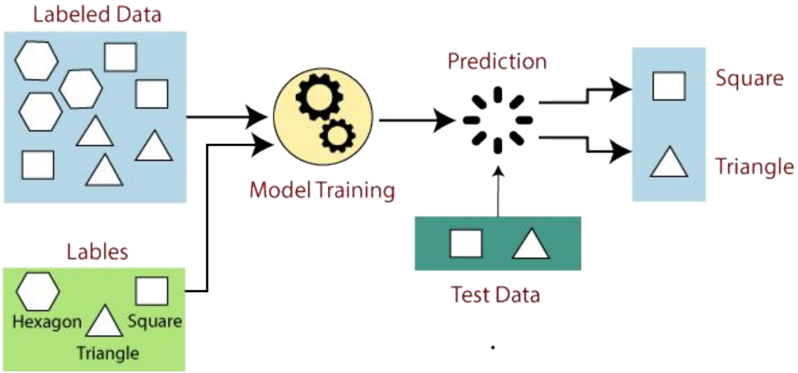
1. Background
    a. Machine Learning
    b. Neural Networks
        i. Convolutional Neural Networks
    c. Generative Adversarial Networks
2. ConSinGAN
    a. Image Generation
        i. Methods
        ii. Results
    b. Image Harmonization
        i. Methods
        ii. Results
3. Conclusion

Machine Learning Models

- A *model* makes predictions without being explicitly programmed
- Has *parameters* whose values start randomly
- Values adjusted through the process of training
- *Hyperparameters* control the rate of adjustment
  - *Learning rate* is an example of a hyperparameter

Supervised Learning vs Unsupervised Learning

- **Supervised Learning**
  - Requires a labeled dataset
  - Learns to associate the labels with the inputs
  - Used for problems such as classification

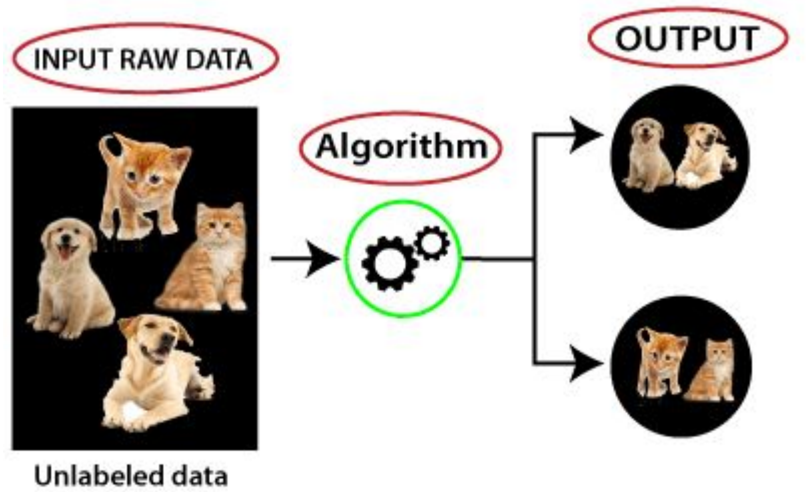


https://www.javatpoint.com/supervised-machine-learning

Supervised Learning vs Unsupervised Learning

- Supervised Learning
  - Requires a labeled dataset
  - Learns to associate the labels with the inputs
  - Used for problems such as classification

- **Unsupervised Learning**
  - Only requires the input
  - Learns patterns or properties of the data
  - Outputs associations between the dataset, such as its probability distribution



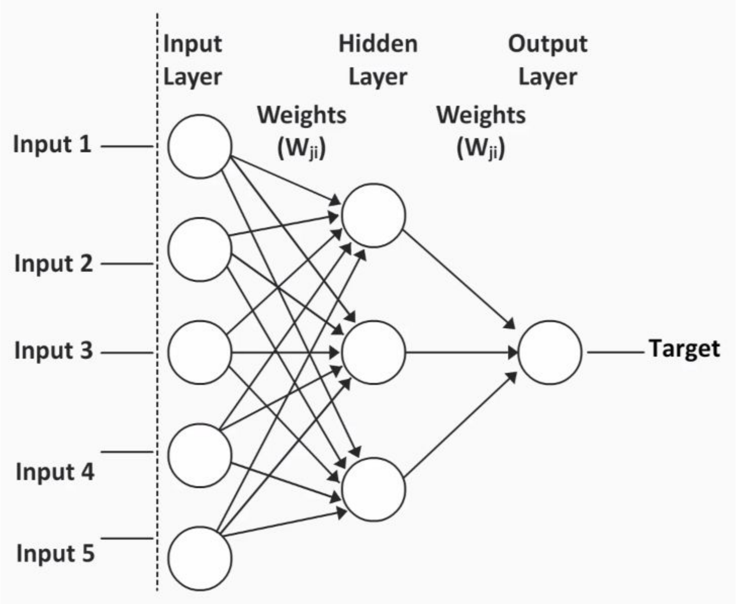
https://www.javatpoint.com/unsupervised-machine-learning

Outline

1. Background
   a. Machine Learning
   b. Neural Networks
      i. Convolutional Neural Networks
   c. Generative Adversarial Networks
2. ConSinGAN
   a. Image Generation
      i. Methods
      ii. Results
   b. Image Harmonization
      i. Methods
      ii. Results
3. Conclusion

Neural Networks

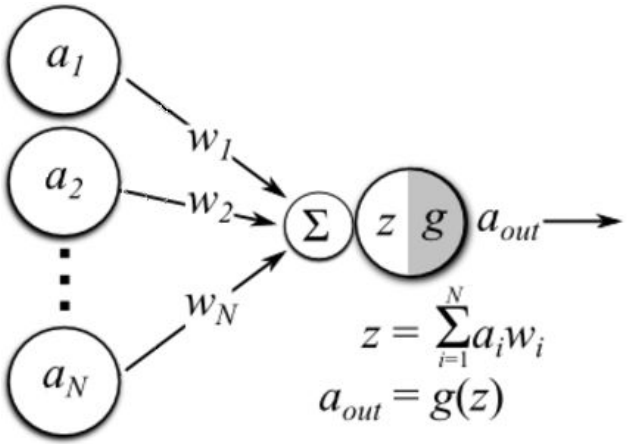
- An **artificial neural network (ANN)** is composed of layers of nodes:
    - An input layer
    - Any number of hidden layers
    - An output layer
- Edges of nodes defined by *weights*



https://subscription.packtpub.com/book/machine-learning/9781838828974/1/ch01lvl1sec05/a-sample-neural-network-model

Neural Networks

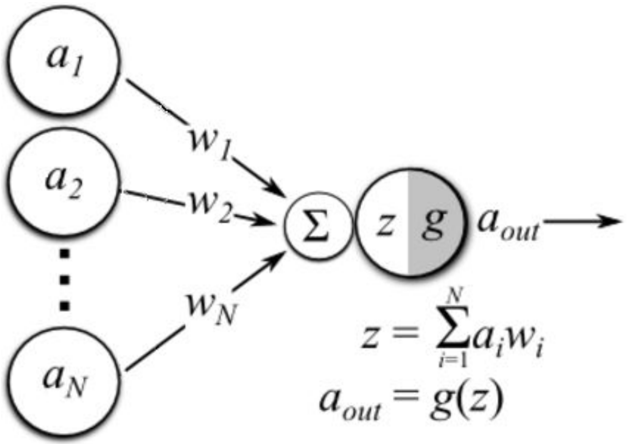
- An artificial neural network (ANN) is composed of layers of nodes:
  - An input layer
  - Any number of hidden layers
  - An output layer
- Edges of nodes defined by *weights*
- **Nodes**
  - Activation function determines how much the node effects the nodes it's connected to in the next layer



https://machine-learning.paperspace.com/wiki/weights-and-biases

Neural Networks

- An artificial neural network (ANN) is composed of layers of nodes:
  - An input layer
  - Any number of hidden layers
  - An output layer
- Edges of nodes defined by *weights*
- Nodes
  - Activation function determines how much the node effects the nodes it's connected to in the next layer
- **Weights adjusted after each iteration through a loss function**
  - Loss is calculated by the difference between the expected output and real output



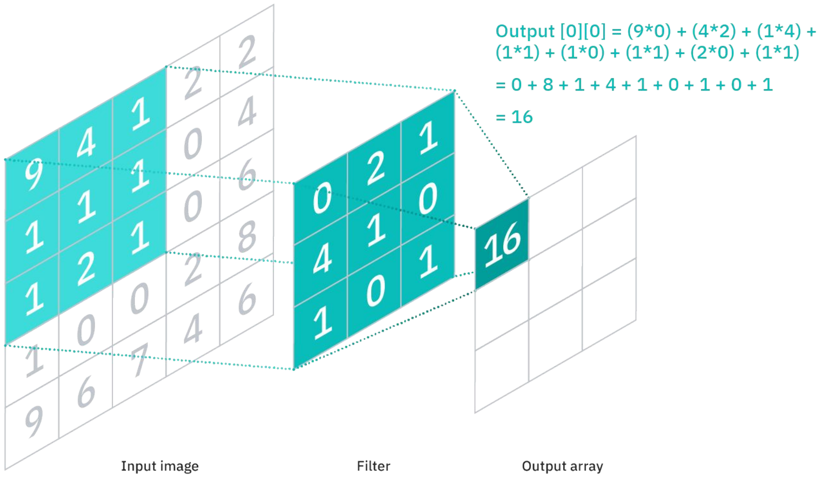
https://machine-learning.paperspace.com/wiki/weights-and-biases

Convolutional Neural Networks (CNNs)

- Type of ANN specialized for image classification and processing
- Three main types of hidden layers:
    - Convolutional layers
    - Pooling layers
    - Fully connected layer

Functions of CNN Layers

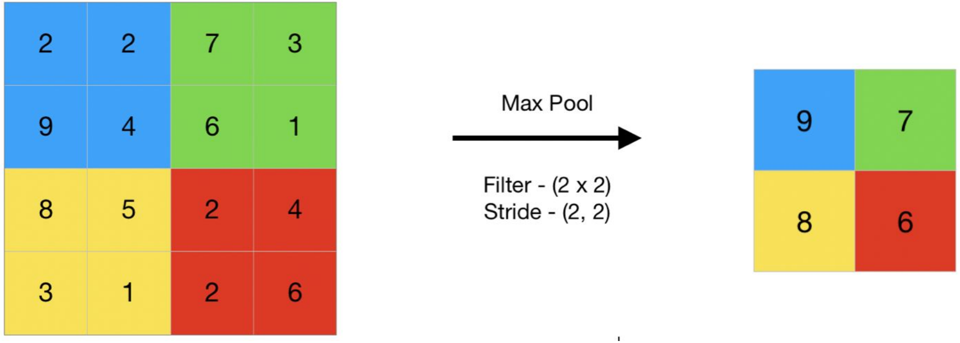
- **Convolutional layers**
  - Performs feature extraction
  - Filter strides across an image
  - Outputs feature maps



https://www.ibm.com/cloud/learn/convolutional-neural-networks

Functions of CNN Layers

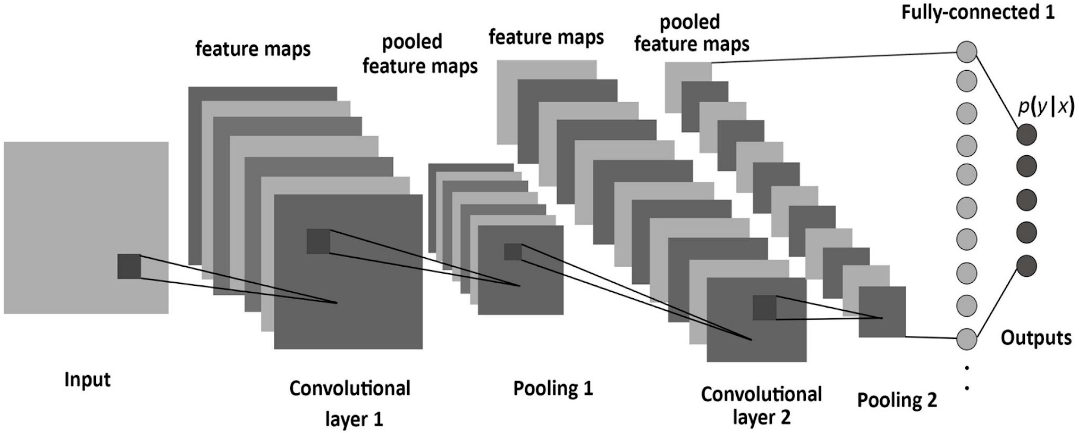
- Convolutional layers
  - Performs feature extraction
  - Filter strides across an image
  - Outputs feature maps
- **Pooling layers**
  - Combines feature maps
  - Dimensionality reduction using pooling



cnn-introduction-to-pooling-layer

Functions of CNN Layers

- Convolutional layers
  - Performs feature extraction
  - Filter strides across an image
  - Outputs feature maps
- Pooling layers
  - Combines feature maps
  - Dimensionality reduction using pooling
- **Fully connected layer**
  - Performs classification using the connections to previous layers



https://www.mdpi.com/1099-4300/19/6/242

Outline

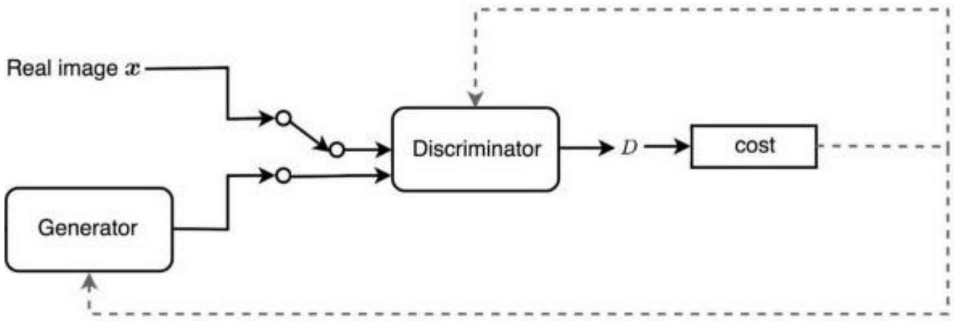
1. Background
    a. Machine Learning
    b. Neural Networks
        i. Convolutional Neural Networks
    c. Generative Adversarial Networks
2. ConSinGAN
    a. Image Generation
        i. Methods
        ii. Results
    b. Image Harmonization
        i. Methods
        ii. Results
3. Conclusion

Generative Adversarial Networks

- Applies game theory to the field of deep learning
- Composed of two parts:
  - Generator starting off with random noise
  - Discriminator trained on the dataset

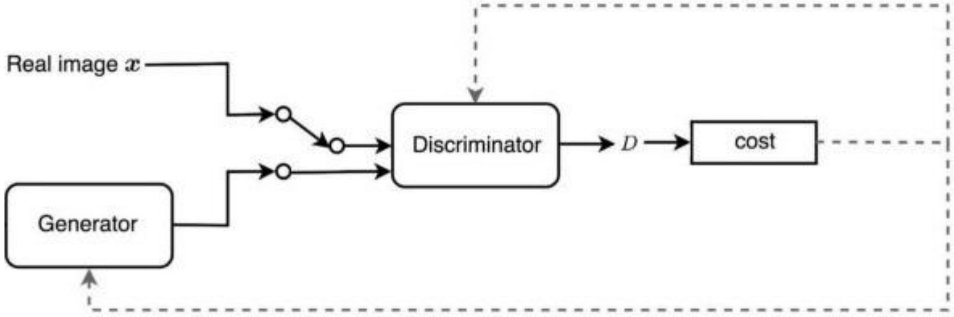
Generative Adversarial Networks

- Applies game theory to the field of deep learning
- Composed of two parts:
  - Generator starting off with random noise
  - Discriminator trained on the dataset
- One network's gain is the other's loss



Generative Adversarial Networks

- Applies game theory to the field of deep learning
- Composed of two parts:
    - Generator starting off with random noise
    - Discriminator trained on the dataset
- One network's gain is the other's loss
- Training is indirect, unsupervised

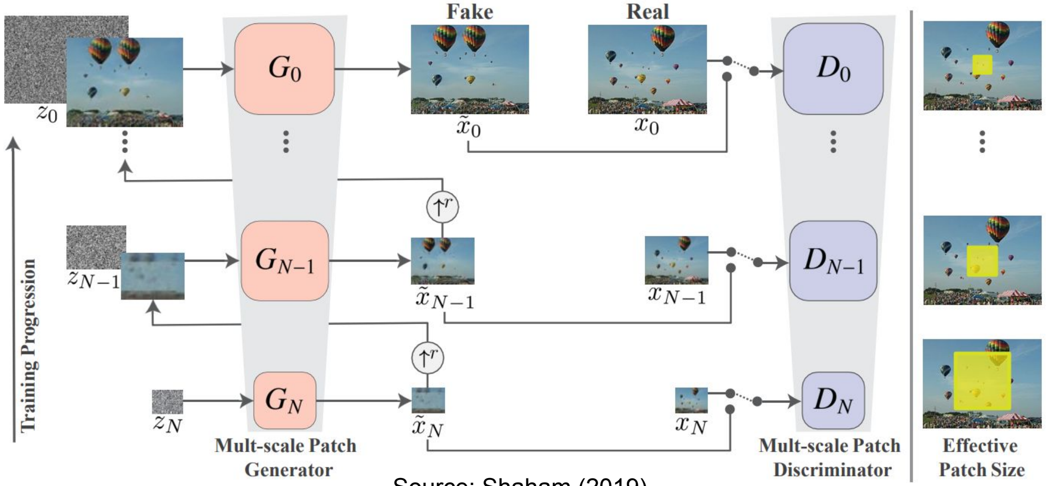


Outline

1. Background
    a. Machine Learning
    b. Neural Networks
        i. Convolutional Neural Networks
    c. Generative Adversarial Networks
2. ConSinGAN
    a. Image Generation
        i. Methods
        ii. Results
    b. Image Harmonization
        i. Methods
        ii. Results
3. Conclusion

Contextualizing Methods with SinGAN

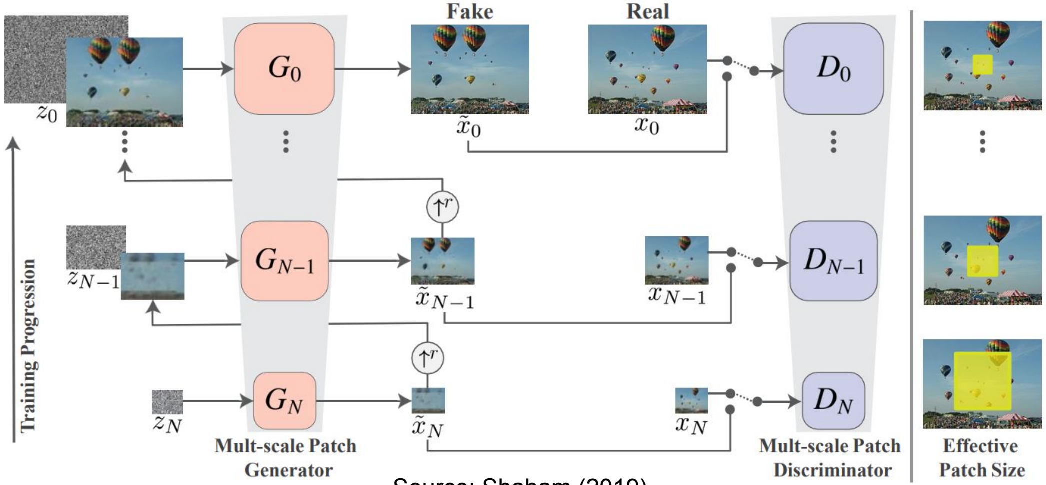
- Stages of Convolutional GANs that build a pyramid
  - First generator starts with random noise, small resolution
  - Up scaled result plus noise inputted into next stages generator



Source: Shaham (2019)

Contextualizing Methods with SinGAN

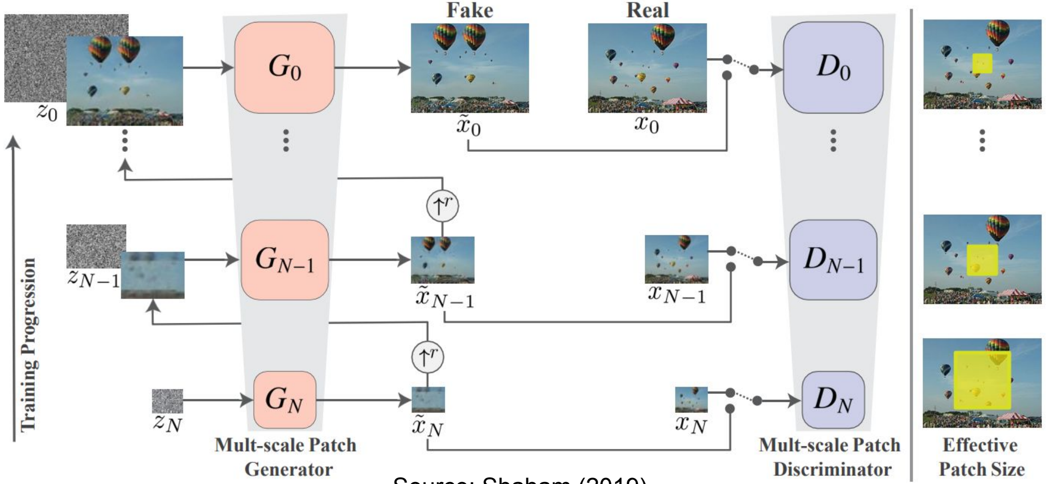
- Stages of Convolutional GANs that build a pyramid
  - First generator starts with random noise, small resolution
  - Up scaled result plus noise inputted into next stages generator
- Resolution increases with each stage, patch size (filter size) decreases
  - First stages learn the image structure, progressively finer details learned



Source: Shaham (2019)

Contextualizing Methods with SinGAN

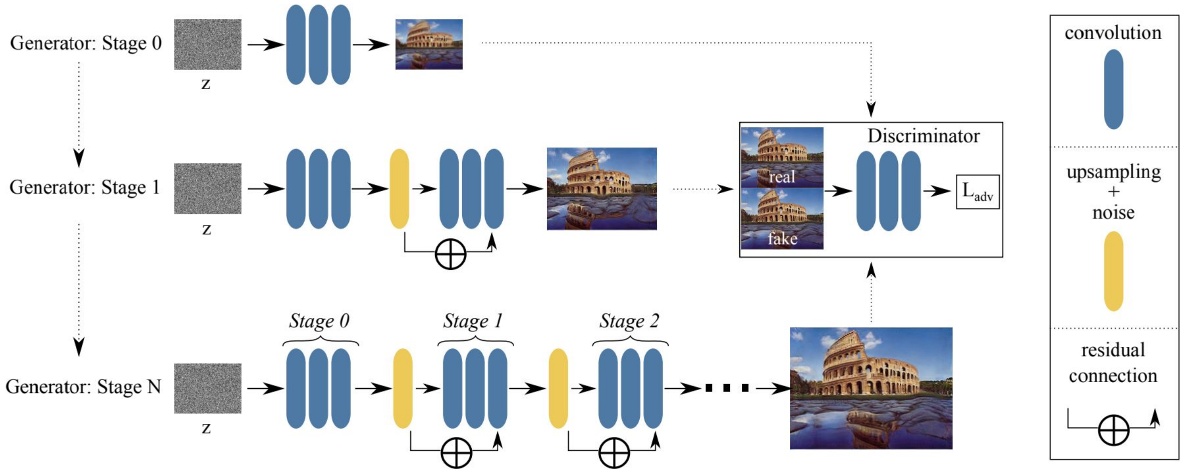
- Stages of Convolutional GANs that build a pyramid
  - First generator starts with random noise, small resolution
  - Up scaled result plus noise inputted into next stages generator
- Resolution increases with each stage, patch size (filter size) decreases
  - First stages learn the image structure, progressively finer details learned
- Weights of each completed stage are frozen



Source: Shaham (2019)

ConSinGAN Methods

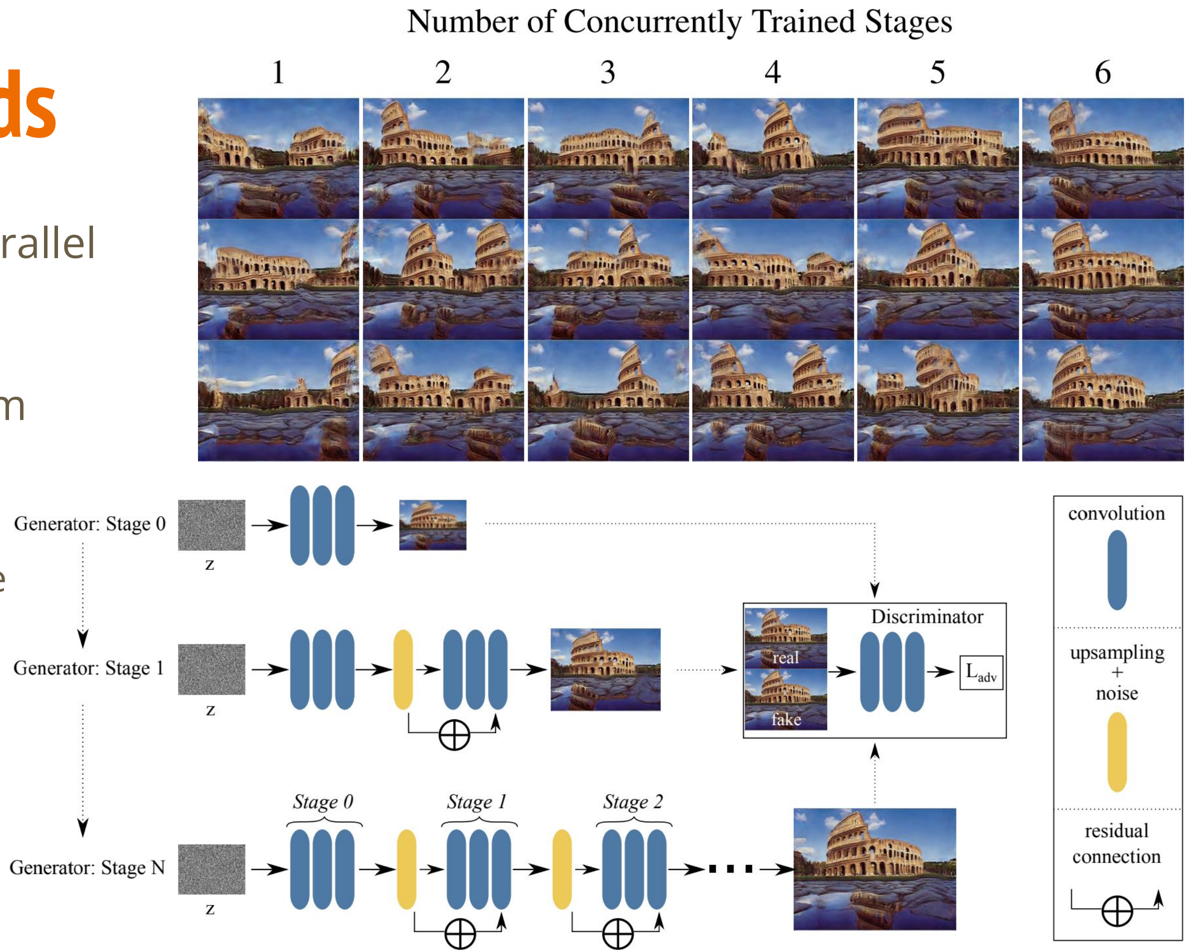
- Trains multiple stages in parallel
  - More realistic images
  - Less training time
- Scales up feature maps from previous stages



Source: Hinz (2020)

ConSinGAN Methods

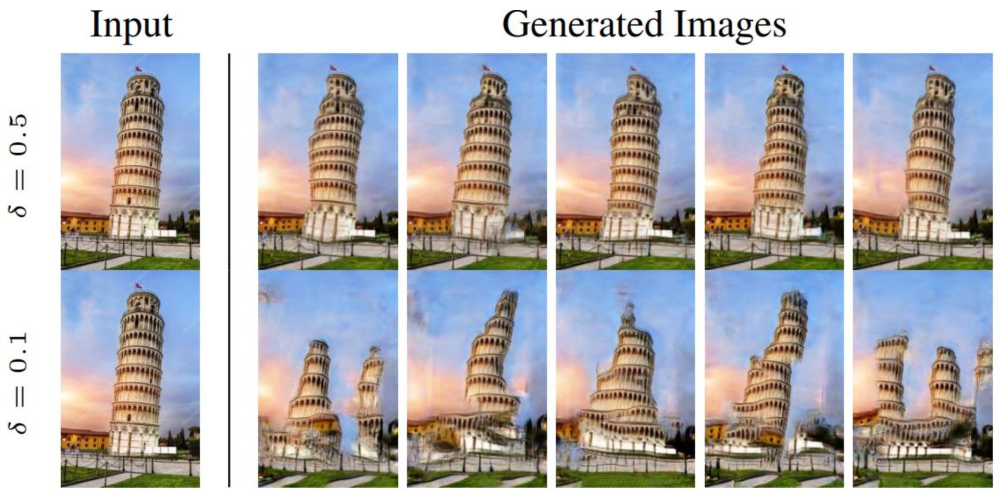
- Trains multiple stages in parallel
  - More realistic images
  - Less training time
- Scales up feature maps from previous stages
- **Allows for tradeoffs**
  - More stages, less variance

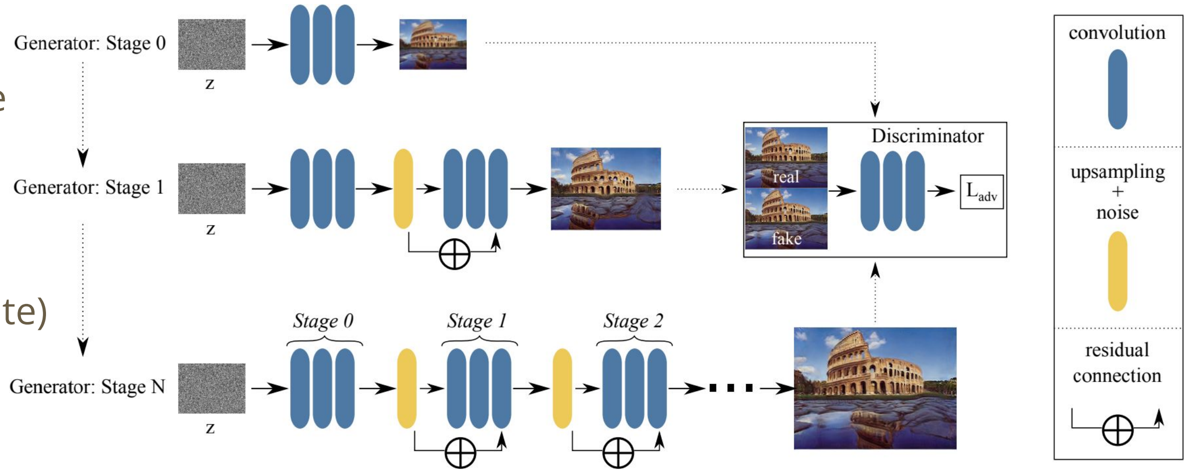


Source: Hinz (2020)

ConSinGAN Methods

- Trains multiple stages in parallel
  - More realistic images
  - Less training time
- Scales up feature maps from previous stages
- Allows for tradeoffs
  - More stages, less variance
- **Scales the learning rate**
  - Emphasises lower stages
  - Reduces overfitting
  - Scaled by (δ^N * learning rate)





Source: Hinz (2020)

ConSinGAN Results: SinGAN's Test

- 'Places' dataset
- Training image is shown next to the generated image
  - Users are asked to choose the real image
  - Confusion is the % of users that chose the generated image

Model	Confusion ↑	SIFID ↓	Train Time	# Stages	# Parameters
ConSinGAN	16.0% ± 1.4%	0.06 ± 0.03	24 min	5.9	~660k
SinGAN	17.0% ± 1.5%	0.09 ± 0.07	152 min	9.7	~1.34m

Table 1. Results of the user study and SIFID on images from the 'Places' dataset [3].

Source: Hinz (2020)

ConSinGAN Results: SinGAN’s Test

- ‘Places’ dataset
- Training image is shown next to the generated image
  - Users are asked to choose the real image
  - Confusion is the % of users that chose the generated image
- ***Single Image Frechet Inception Distance* (SIFID) used for quantitative analysis**
  - Compares the distribution of a pre-trained network’s activations between the sets of generated and real images
  - Lower scores shown to correlate with higher quality images

Model	Confusion ↑	SIFID ↓	Train Time	# Stages	# Parameters
ConSinGAN	16.0% ± 1.4%	0.06 ± 0.03	24 min	5.9	~660k
SinGAN	17.0% ± 1.5%	0.09 ± 0.07	152 min	9.7	~1.34m

Table 1. Results of the user study and SIFID on images from the ’Places’ dataset [3].

Source: Hinz (2020)

ConSinGAN's Test

- LSUN dataset
- Generated images from both models are shown next to each other
  - Users are asked to choose which image is better
  - First study chooses one image randomly from the set of generated images of SinGAN and ConSinGAN, likely from different training images
  - Second study pairs images from the same training image

Model	Random ↑	Paired ↑	SIFID ↓	Train Time	# Stages	# Parameters
ConSinGAN	56.7% ± 1.9%	63.1% ± 1.8%	0.11 ± 0.06	20 min	5.9	~660K
SinGAN	43.3% ± 1.9%	36.9% ± 1.8%	0.23 ± 0.15	135 min	9.1	~1.0M

Table 2. Results of the user studies and SIFID on images from the LSUN dataset [3].

Source: Hinz (2020)

Outline

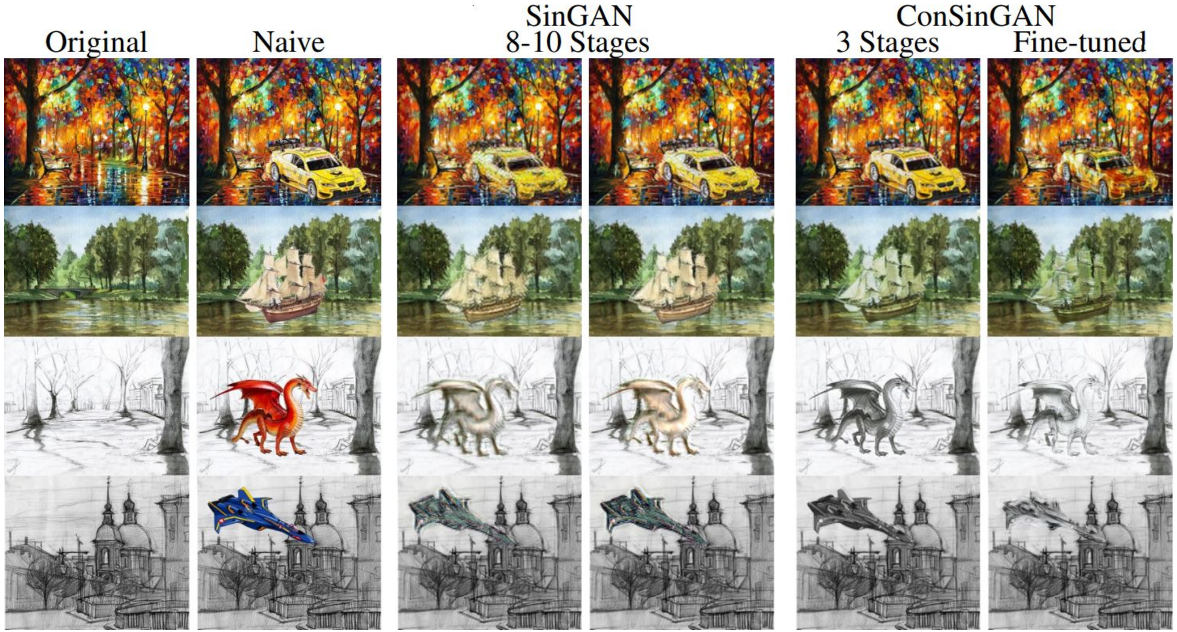
1. Background
    a. Machine Learning
    b. Neural Networks
        i. Convolutional Neural Networks
    c. Generative Adversarial Networks
2. ConSinGAN
    a. Image Generation
        i. Methods
        ii. Results
    b. Image Harmonization
        i. Methods
        ii. Results
3. Conclusion

Image Harmonization Methods

- Same architecture as image generation
- Trained for exactly three stages per image
- One thousand iterations per stage
- Random sample chosen every iteration
    - Combinations of additive noise and color transformations added

Image Harmonization Results

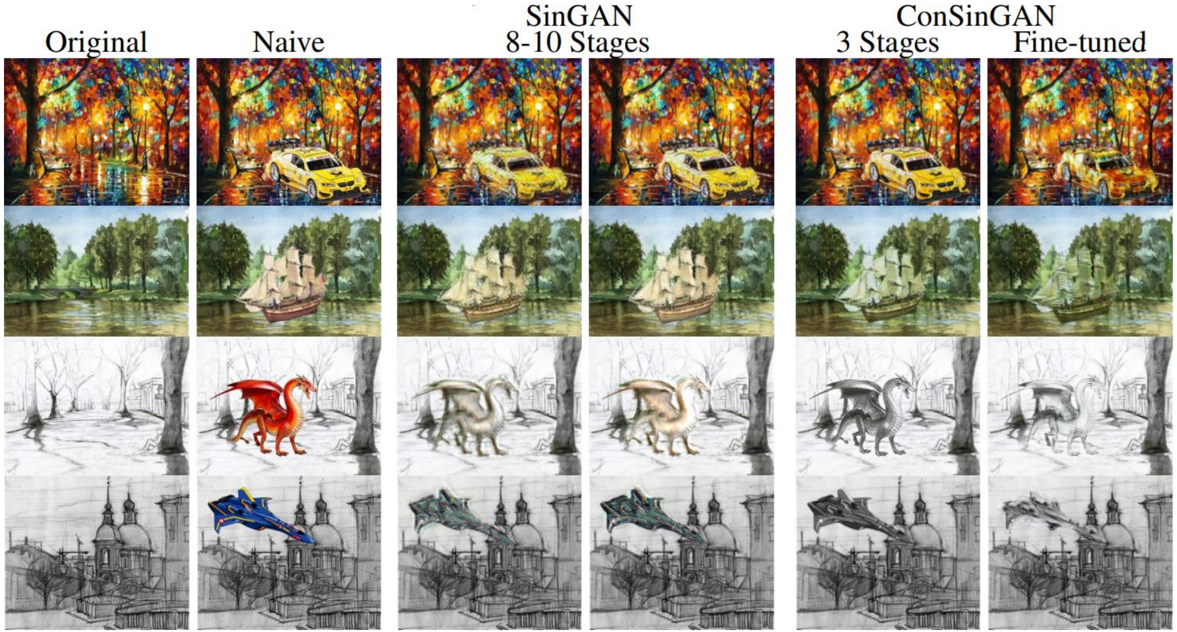
- SinGAN
  - Model needs to be fully trained
  - Artifacting
  - Colors not integrating



Source: Hinz (2020)

Image Harmonization Results

- SinGAN
  - Model needs to be fully trained
  - Artifacting
  - Colors not integrating
- ConSinGAN
  - Less than 10 minutes to train
  - Colors ‘absorbed’ into the image



Source: Hinz (2020)

Conclusion

- Improved upon the groundwork laid by SinGAN
- Drastically reduced training time
- Similar or better image generation results in both tested applications
  - Image Generation
  - Image Harmonization
- More testing needed

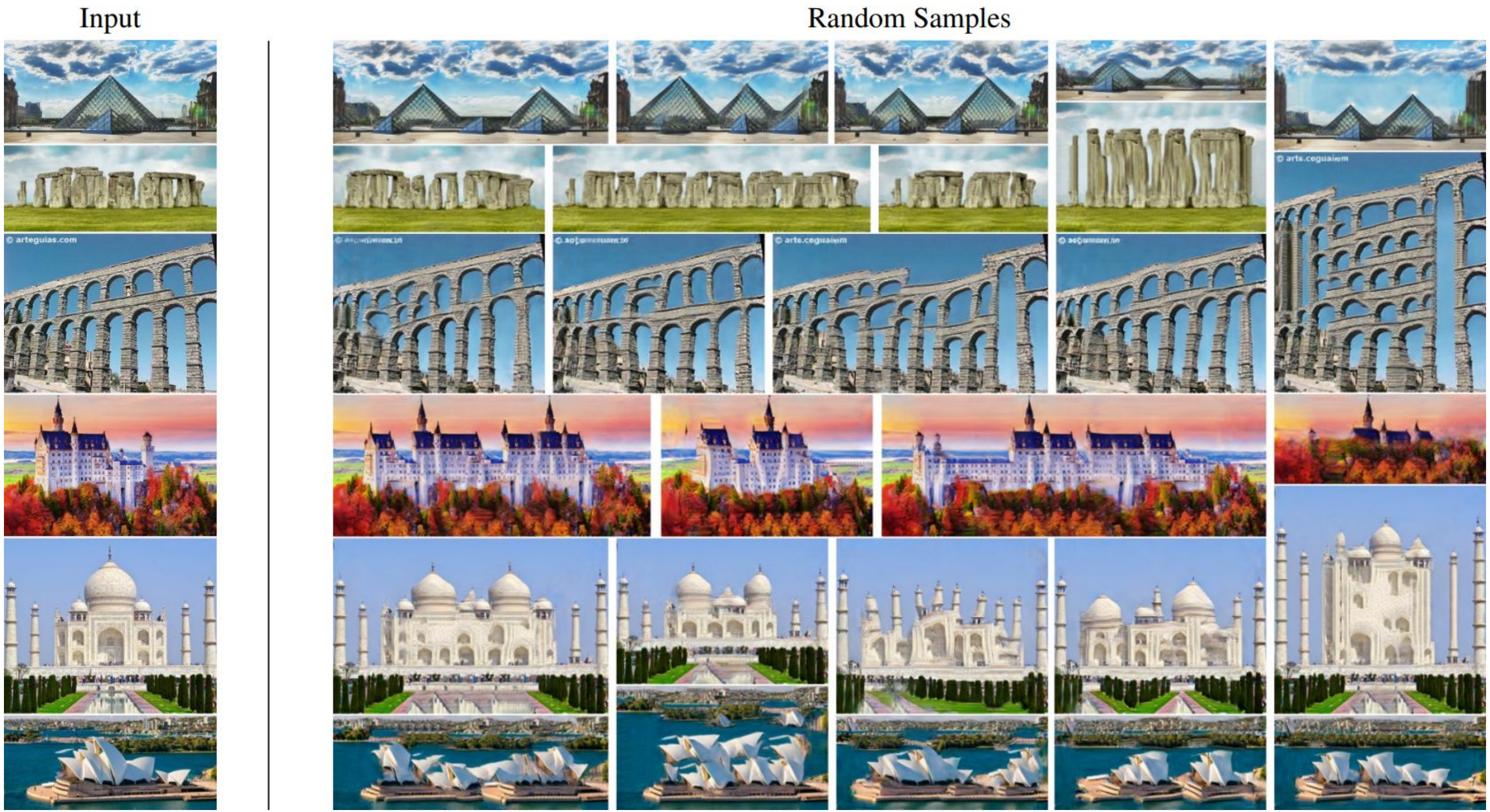
References

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative adversarial networks. Commun. ACM 63, 11 (November 2020), 139–144. https://doi-org.ezproxy.morris.umn.edu/10.1145/3422622

T. R. Shaham, T. Dekel and T. Michaeli, "SinGAN: Learning a Generative Model From a Single Natural Image," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 4569-4579, doi: 10.1109/ICCV.2019.00467.

Hinz, Tobias, et al. "Improved techniques for training single-image gans." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021. https://openaccess.thecvf.com/content/WACV2021/html/Hinz_Improved_Techniques_for_Training_Single-Image_GANs_WACV_2021_paper.html

Questions?



Source: Hinz (2020)