Generative Adversarial Networks: An Assessment on Transfer Learning

Brian Lee

Logan Eisenbeiser

NATIONAL SECURITY INSTITUTE
VIRGINIA TECH

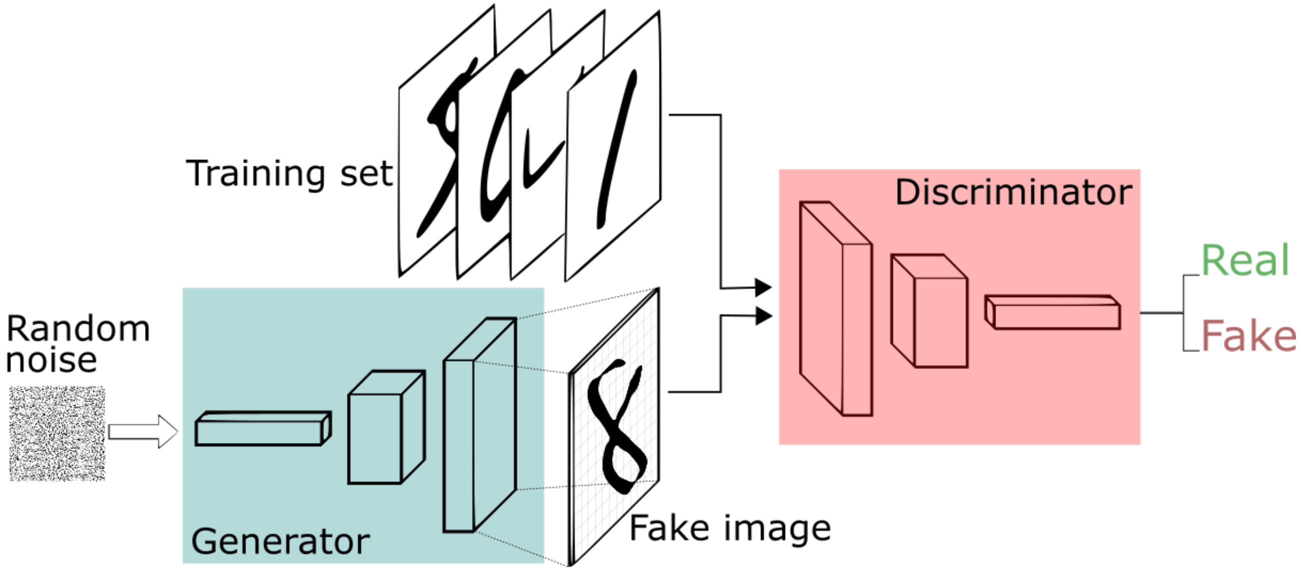
Generative Modeling

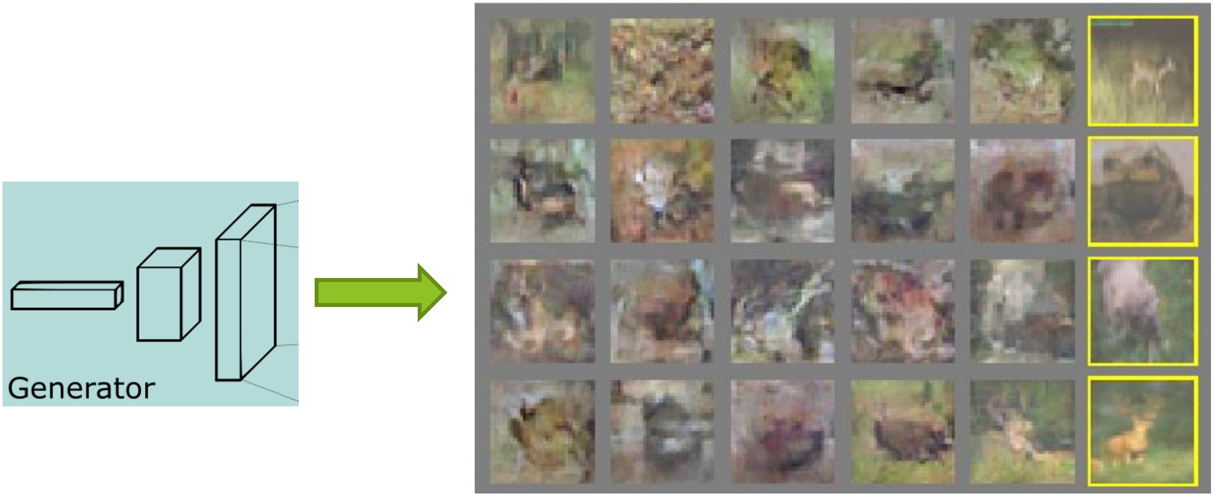
- **Learning an underlying distribution of data to generate new examples**

Overview of GANs

- Goodfellow et al. (2014)

- Generator
  - Generates new data
- Discriminator
  - Classifies given image samples as real or fake

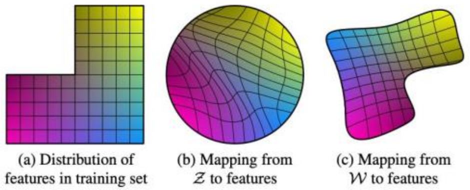


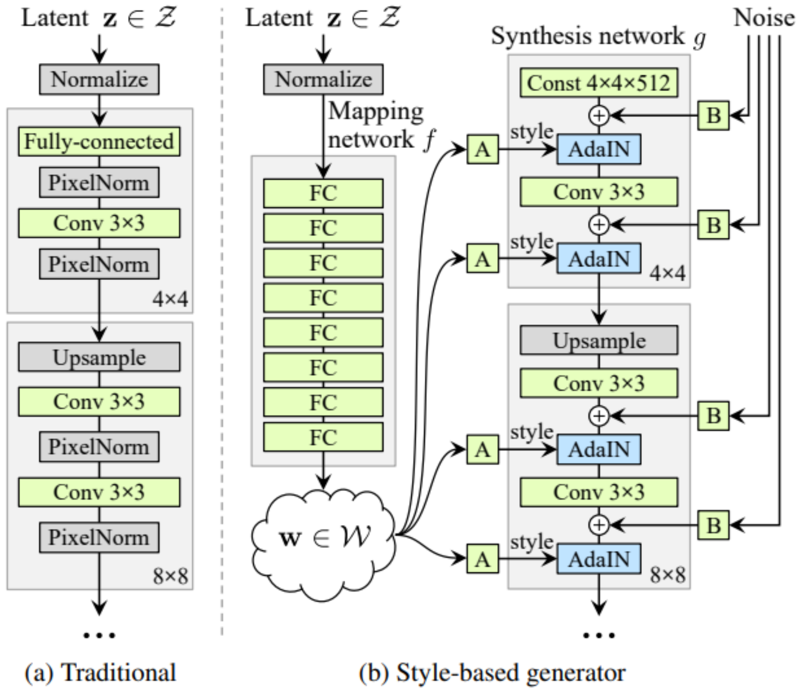


Goodfellow et al. https://arxiv.org/pdf/1406.2661.pdf

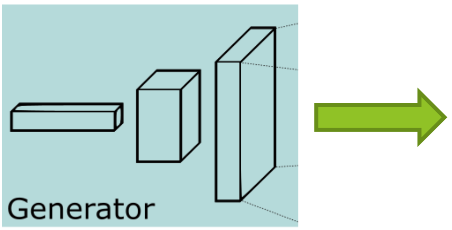
StyleGAN (2019)

- Karras et al. (2019-2021), NVIDIA
- Mapping network
  - Makes inputs easier to learn from
- Allows for application of styles
  - Layering and mixing of stylized features
  - Hair color, eye color, texture, etc.





- styleGAN3

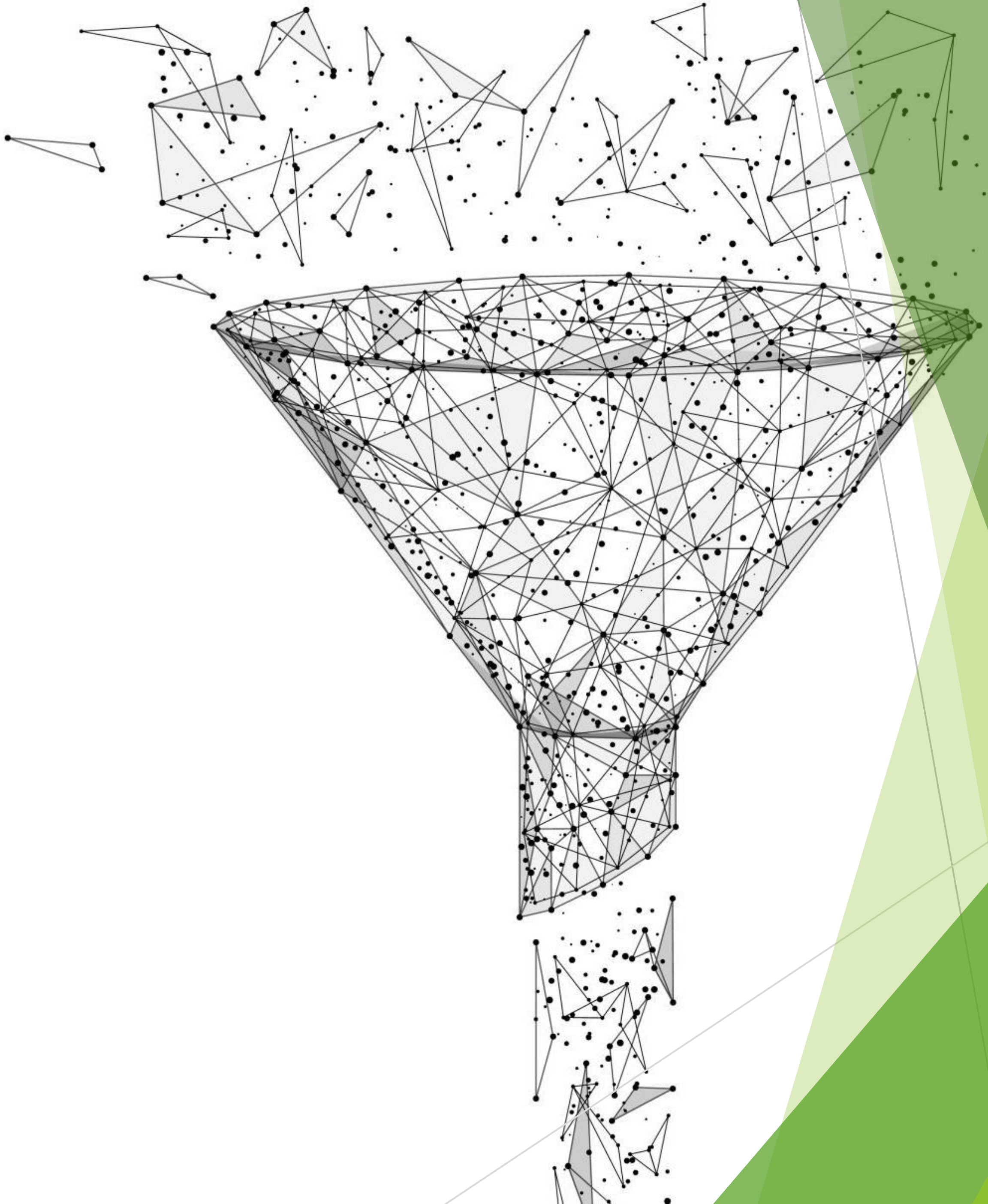


Research Objectives

1. Evaluate effectiveness of transfer learning in training generative adversarial networks
2. Assess the viability of small, created datasets

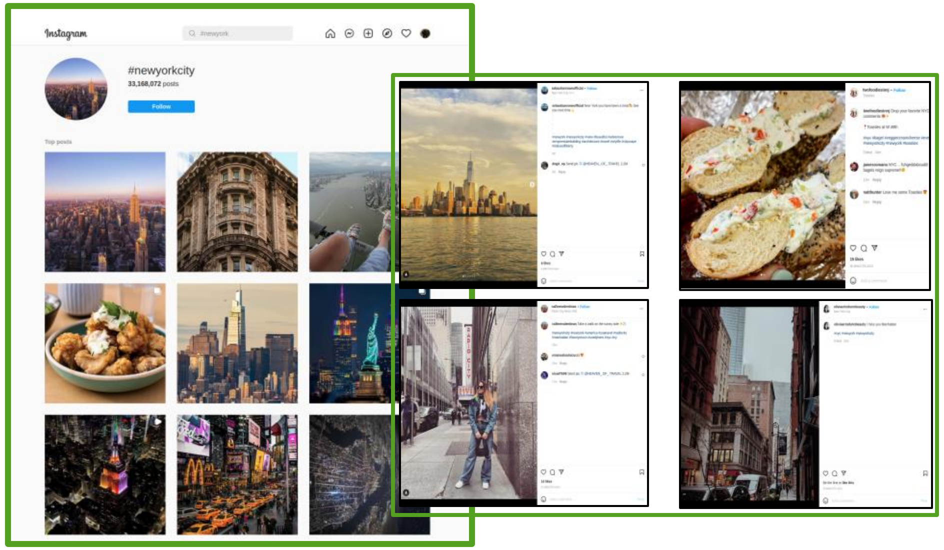
Dataset Assimilation

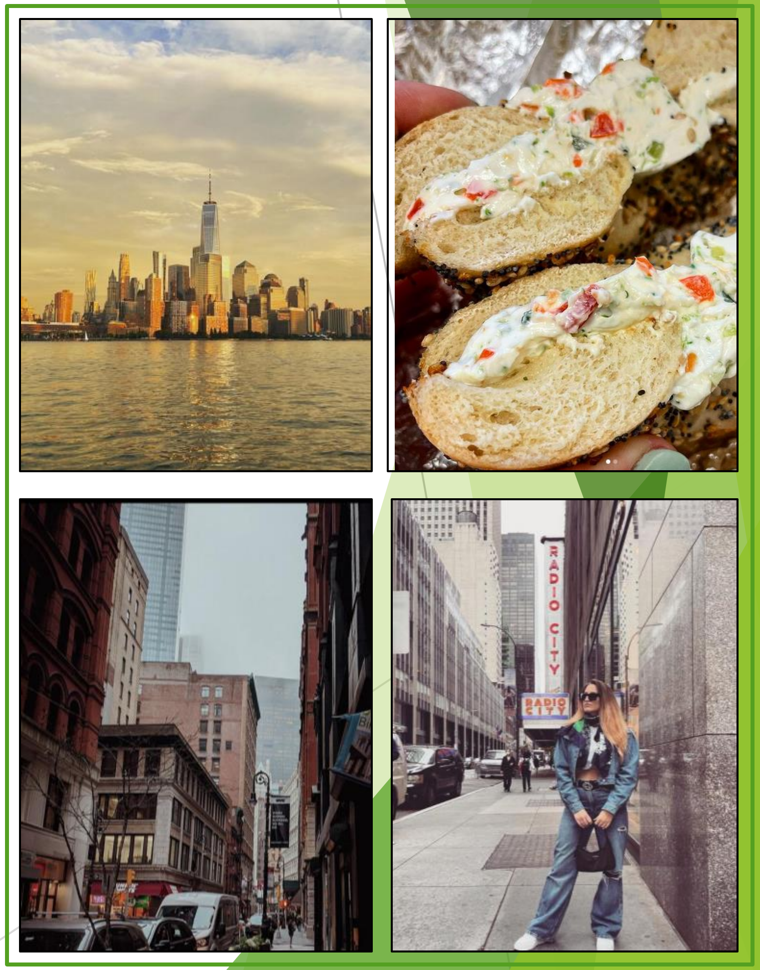
1. Data scraping
2. Data filtration



Data Scraping

- Instaloader package
  - Automates the extraction of media and their metadata directly from Instagram
  - Capabilities to scrape any public post by user profiles, hashtags, and a general search





NATIONAL SECURITY INSTITUTE
VIRGINIA TECH

- Examples of scraped images for "#freedomtowernyc:

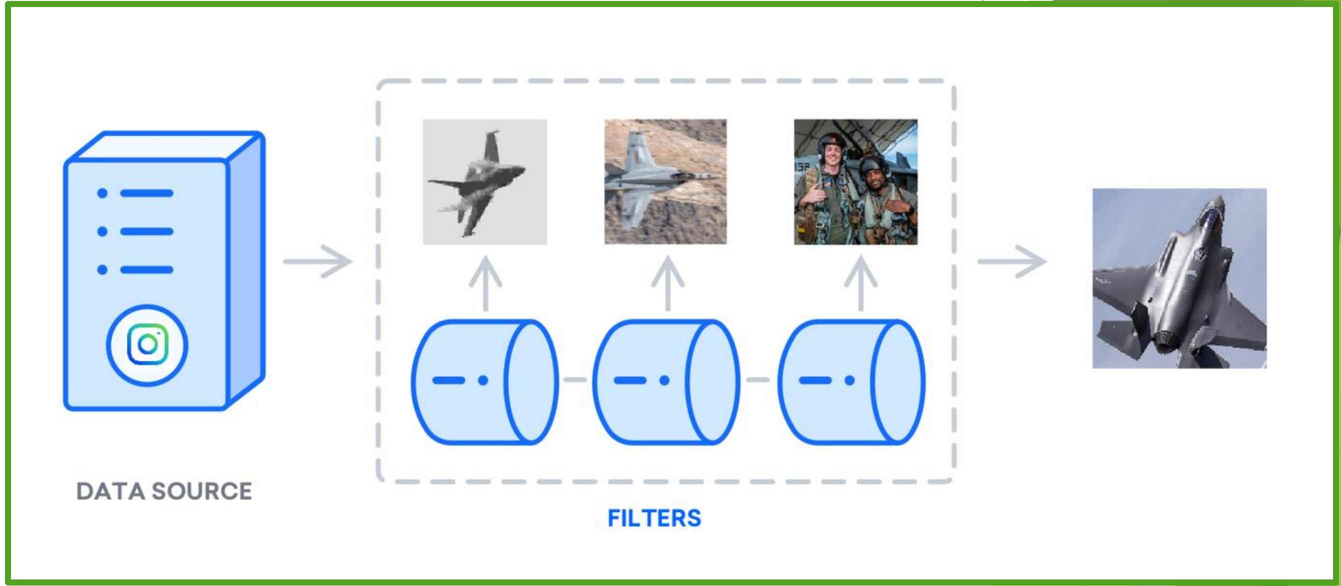
Instaloader

Raw Data Examples

- High variability in subjects and quality of collected photos
- Standardize raw data

Data Filtration

- Discard images that are not suitable for training
- Mimic manual classification
- Train with resulting dataset



Discard Criteria

- Filtering Criteria
  1. Any file without a valid image format (.jpg, .png, .webp, etc.)
  2. **Low-resolution images (80% of 1024px)**
  3. **Blurry images**
  4. Grayscale images
  5. **Images containing prominent text**
  6. **Images containing human faces**

Parameter Tuning

- Example: Misclassification of desired and undesired images

Text detection	Facial recognition	Blurriness
• False positive classification of text appearances. • Tradeoff between false positive classification and lenience	• Difficulty parsing out accessories (sunglasses, mask, hat, etc.) • 45-degree facial profiles	• Fine experimentation with how much blur is considered "blurry" • Blurry backgrounds

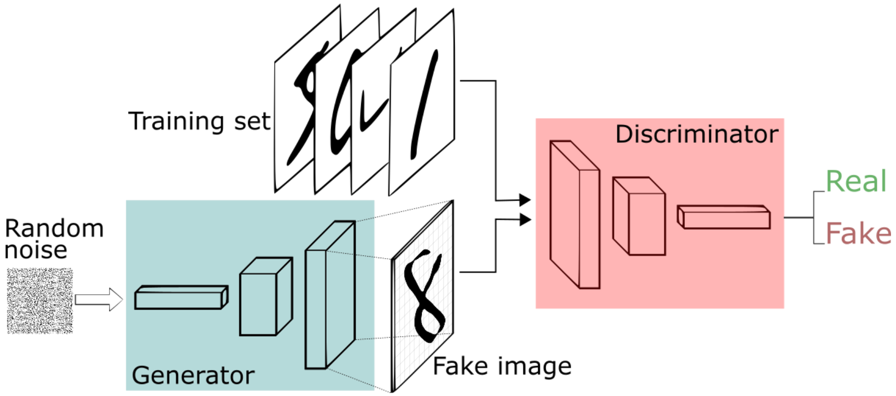
Training

1. GAN Training
   i. Training from small datasets
2. Transfer learning
3. Experiments
   i. Dataset changes
   ii. Network changes

NATIONAL SECURITY INSTITUTE
VIRGINIA TECH

GAN Training

- Advantages
  - High-quality modeling
  - Uses unlabeled data
  - Generator updates without data examples



- Disadvantages
  - Careful synchronization of G and D
  - Unstable training
    - Vanishing gradients
    - Mode collapse
    - Non-convergence

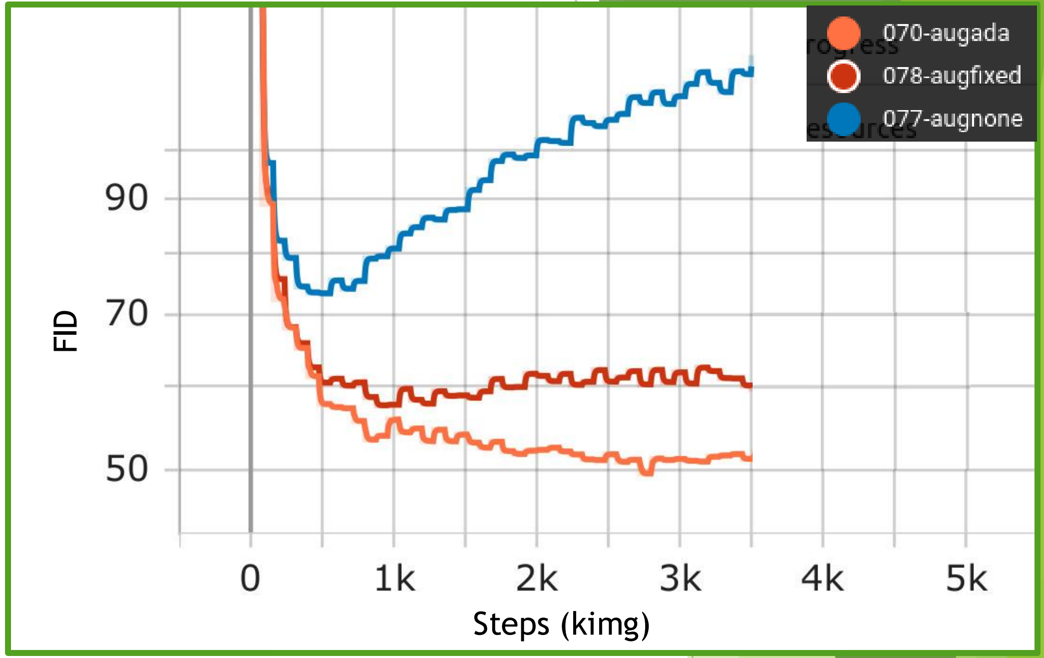
- Performance metric: Frechet Inception Distance (FID) score

Training Limited Datasets

- Overfitting when training from scratch on small datasets
  - Small Instagram datasets (500-1,500 images)
  - Poor quality or unrealistic results

- Solutions
  - Obtain larger dataset (50,000-100,000 images)
  - **Transfer learning**

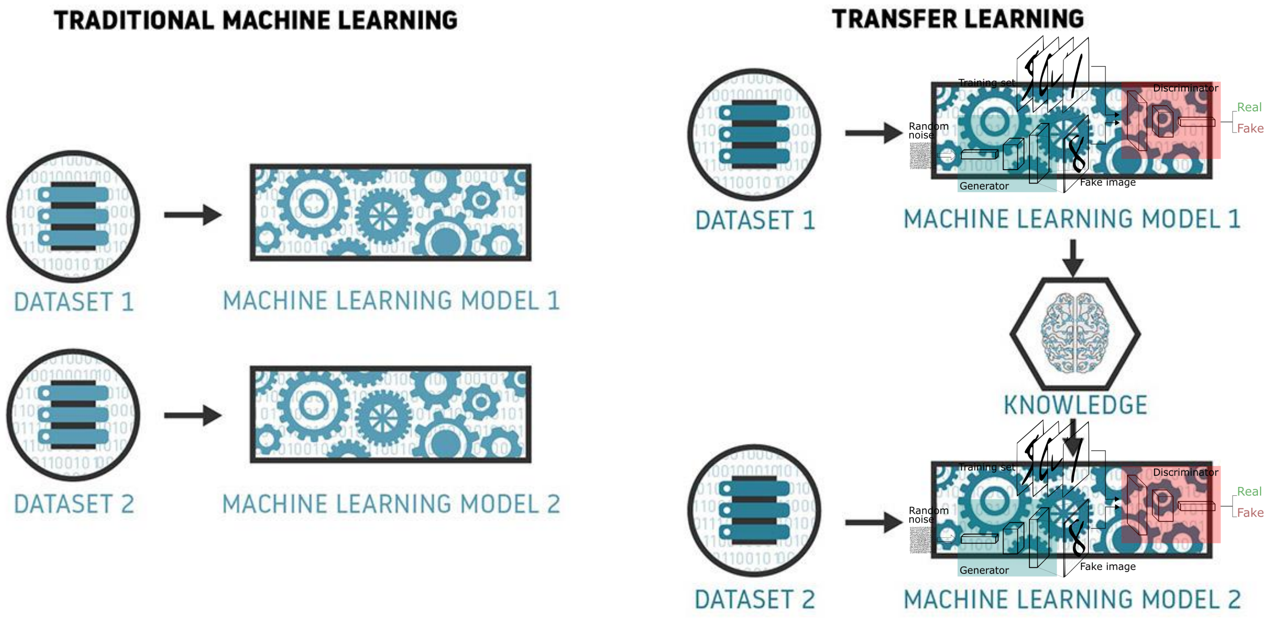
- FID of various data augmentation methods



Transfer Learning

- Applying a model pre-trained on one task to another task or domain
- Leverages learning done from the previous model

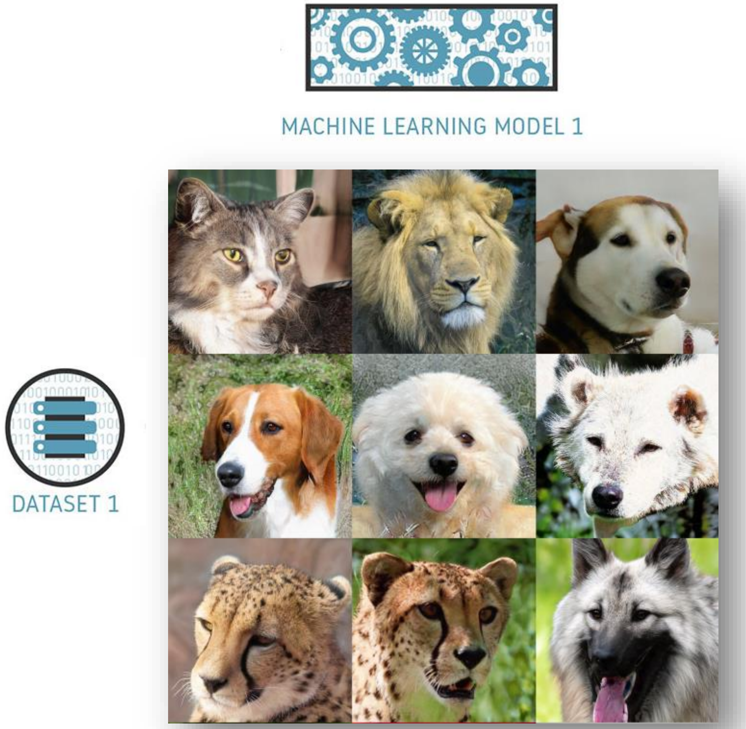
- Transfer learning approach
  - Cost efficient
  - Testing the limits of limited domains



NATIONAL SECURITY INSTITUTE
VIRGINIA TECH

Initial Experiments

- NVIDIA AFHQV2 Network
- Animal Faces-HQ (16,130 images)

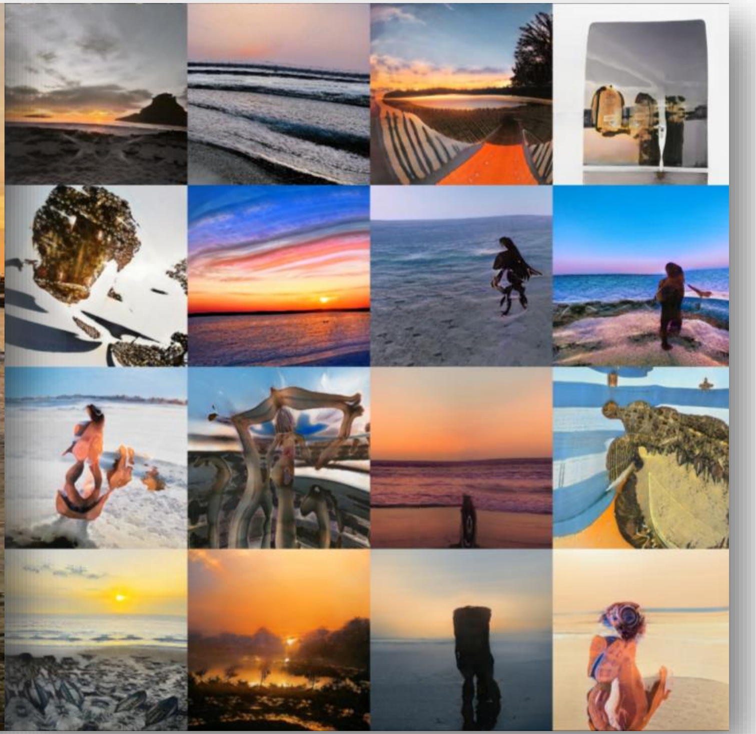


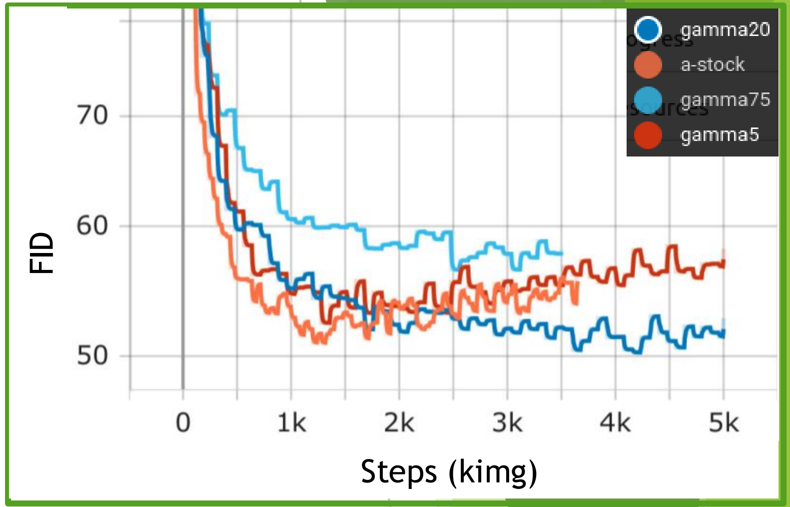
- NVIDIA AFHQV2 Network
- #beachsunset (640 images)

Initial Experiments

- Generated image

- Training snapshot



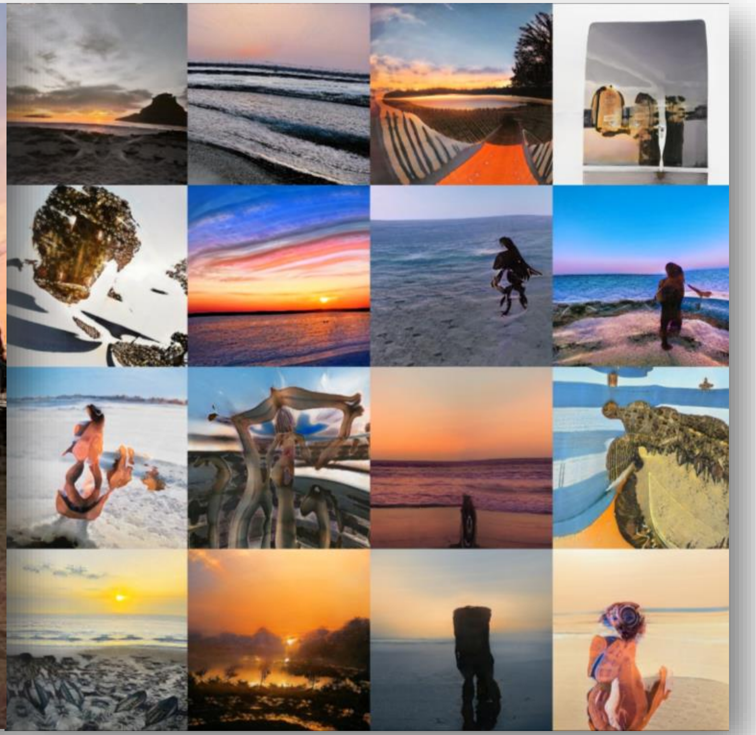


- Intelligible, passable examples after 3500 steps
- Network adaptation

Initial Experiments

- Generated image

- Training snapshot



- Generation difficulties
- High variance in foreground objects
- Confuses generator training

NATIONAL SECURITY INSTITUTE
VIRGINIA TECH

Domain Change

- Hypothesis: datasets easier to learn than others

- NVIDIA AFHQV2 Network
- Animal Faces-HQ (16,130 images)

MACHINE LEARNING MODEL 1

DATASET 1

- NVIDIA AFHQV2 Network
- #bettaphotography (811 images)

MACHINE LEARNING MODEL 1

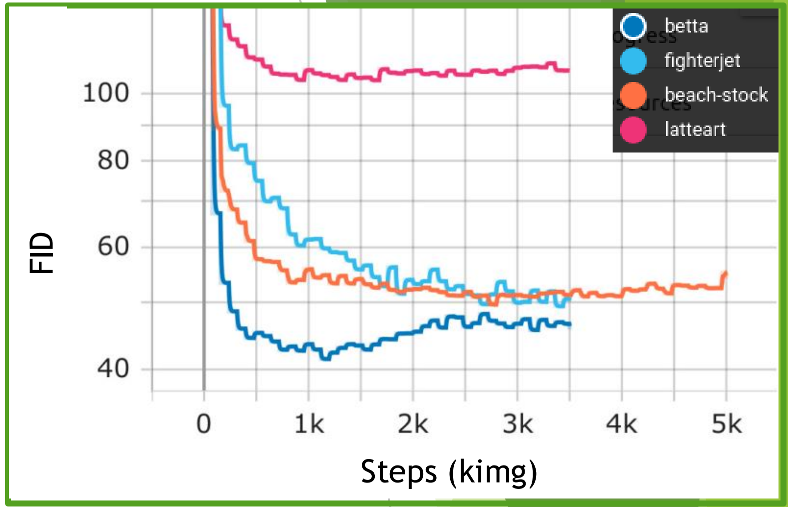
DATASET 2

Domain Change

- Generated image

- Training snapshot

- FIDs by dataset

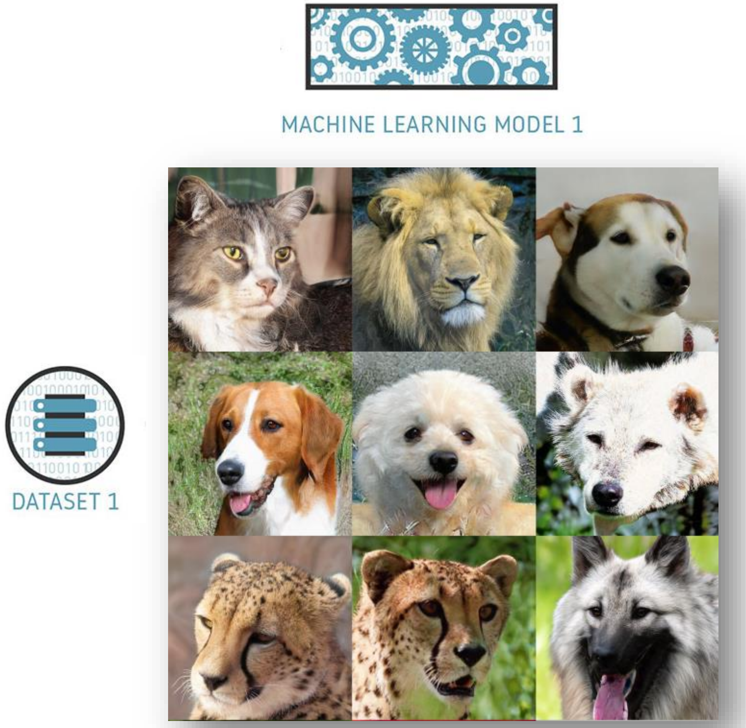


- Vast improvement in generation quality, objectivity
- Less frequent nonsense
- Minimum of 41.52 FID on betta fish dataset

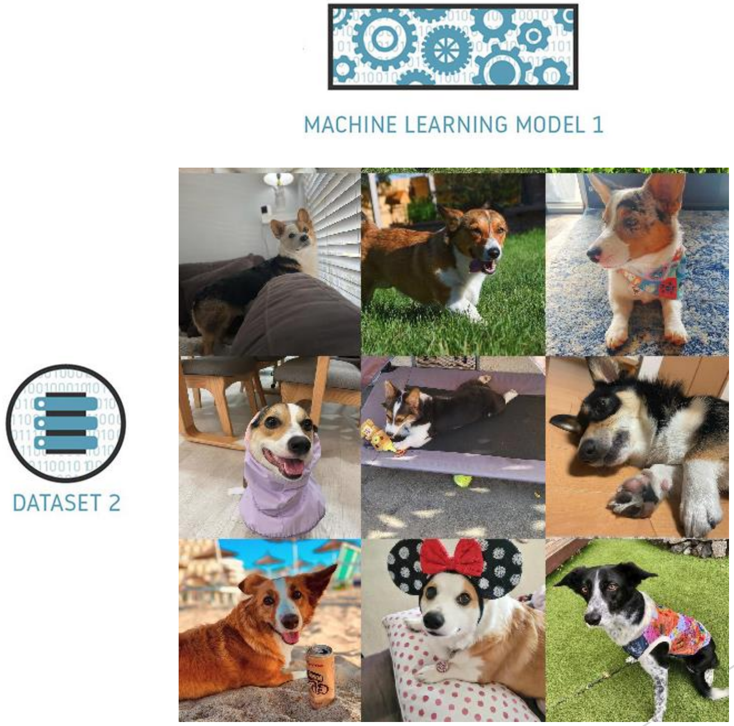
Domain Change

- Hypothesis: datasets easier to learn than others

- NVIDIA AFHQV2 Network
- Animal Faces-HQ (16,130 images)

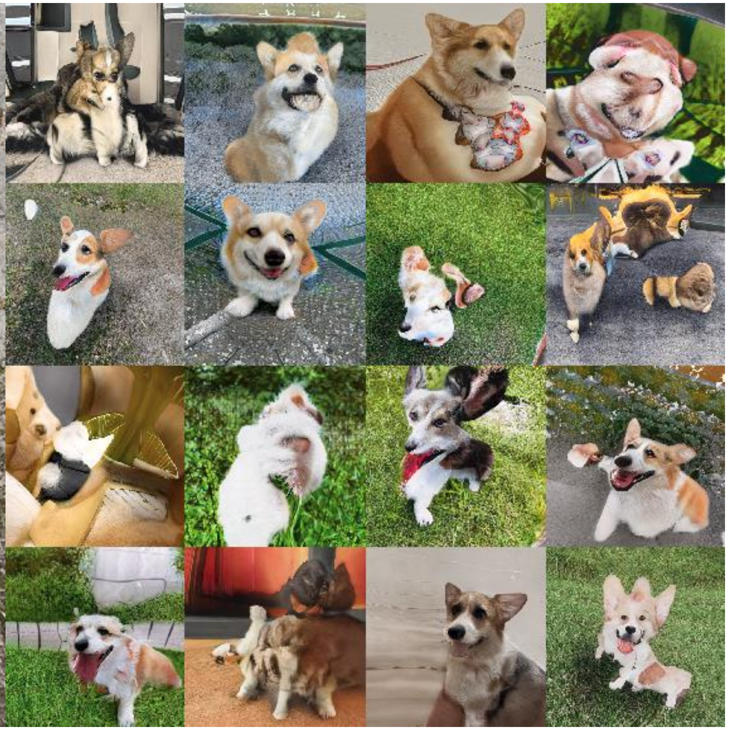


- NVIDIA AFHQV2 Network
- #corgi (529 images)

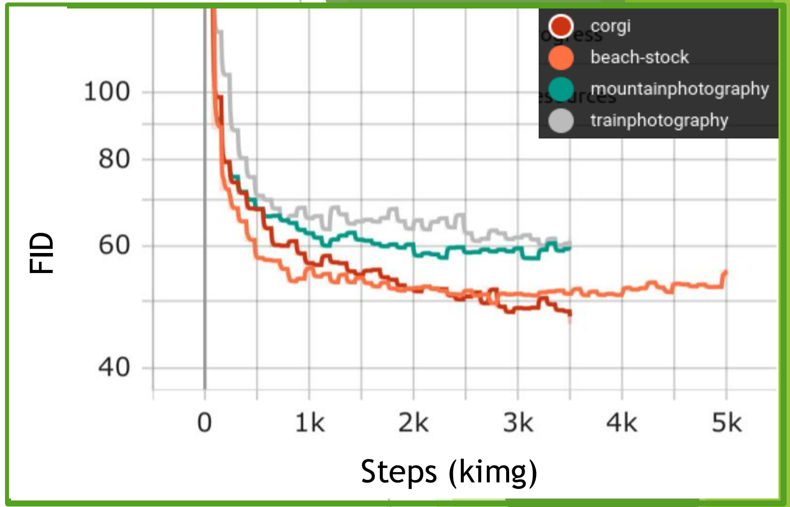


Domain Change

- Generated image
- Training snapshot



- FIDs by dataset



- More improvements in generation quality, objectivity
- Less frequent nonsense
- Further training possible

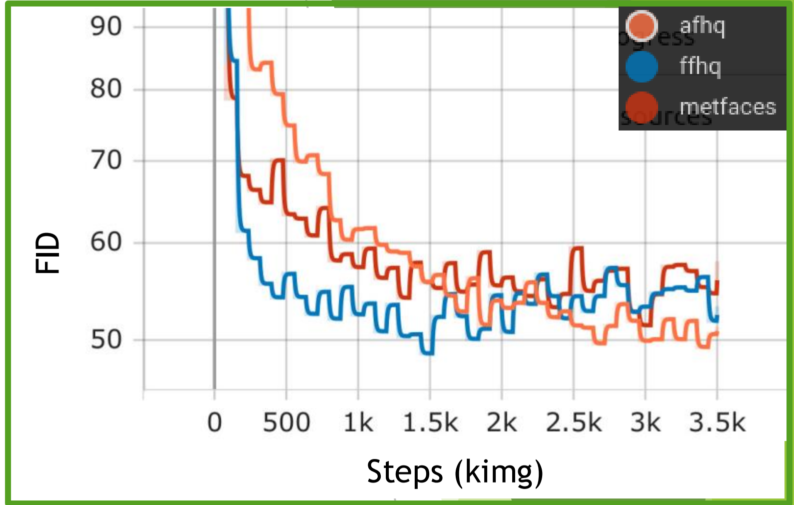
Network Change

- Top: NVIDIA FFHQ Network
  - Flickr Faces (70,000 images)
- Bottom: NVIDIA Metfaces Network
  - Metropolitan Museum of Art (1,336 images)

- NVIDIA FFHQ Network
- #fighterjet (650 images)

Network Change

- FIDs by network



- Generated images



- AFHQ-V2
- FFHQ
- Metfaces

- Generally similar quality of generations
- Oscillations around similar FID score
- Multiple networks viable on one dataset

Conclusions

- Transfer learning is promising method of training networks on limited datasets
  1. Domain of the dataset can have a large role
  2. Generation from a dataset can be robust across multiple pre-trained networks
  3. Augmentations are a necessity for small datasets, even in transfer learning

- Further work
  1. Explore specific measures optimal training datasets
  2. Test on a wider range of starting networks
  3. Comparison of training from scratch to transfer learning