Creative AI

Garching, 05.08.2019

Felix Altenberger, Vadim Goryainov, Anastasia Litinetskaya, Jongwon Lee, Vitalii Mozin

Our Team

Felix Altenberger

Data Engineering and Analytics

Vadim Goryainov

Informatics

Anastasia Litinetskaya

Mathematics

Jongwon Lee

Data Engineering and Analytics

Vitalii Mozin

Data Engineering and Analytics

Agenda

- Introduction
- GANs for Image Generation from Text
- Demo Session
- System Architecture
- Summary and Q&A

Agenda

- **Introduction**
- GANs for Image Generation from Text
- Demo Session
- System Architecture
- Summary and Q&A

Project Description

This bird is red with white and has a very short beak

CreativeAI

Text Encoder -> GAN

Our Implementation

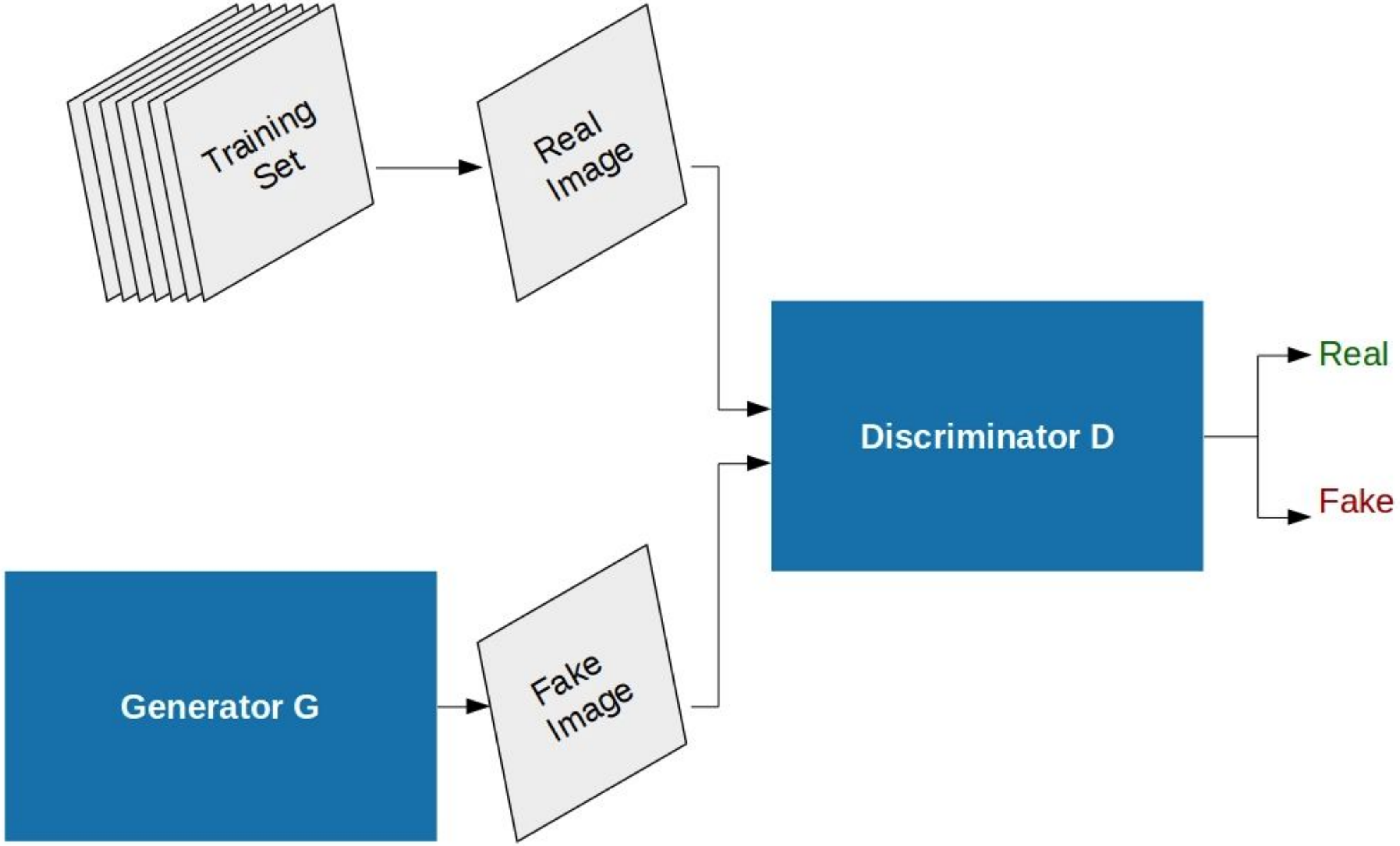
Dash Demo for CreativeAI

this bird is red with white and has a very short beak

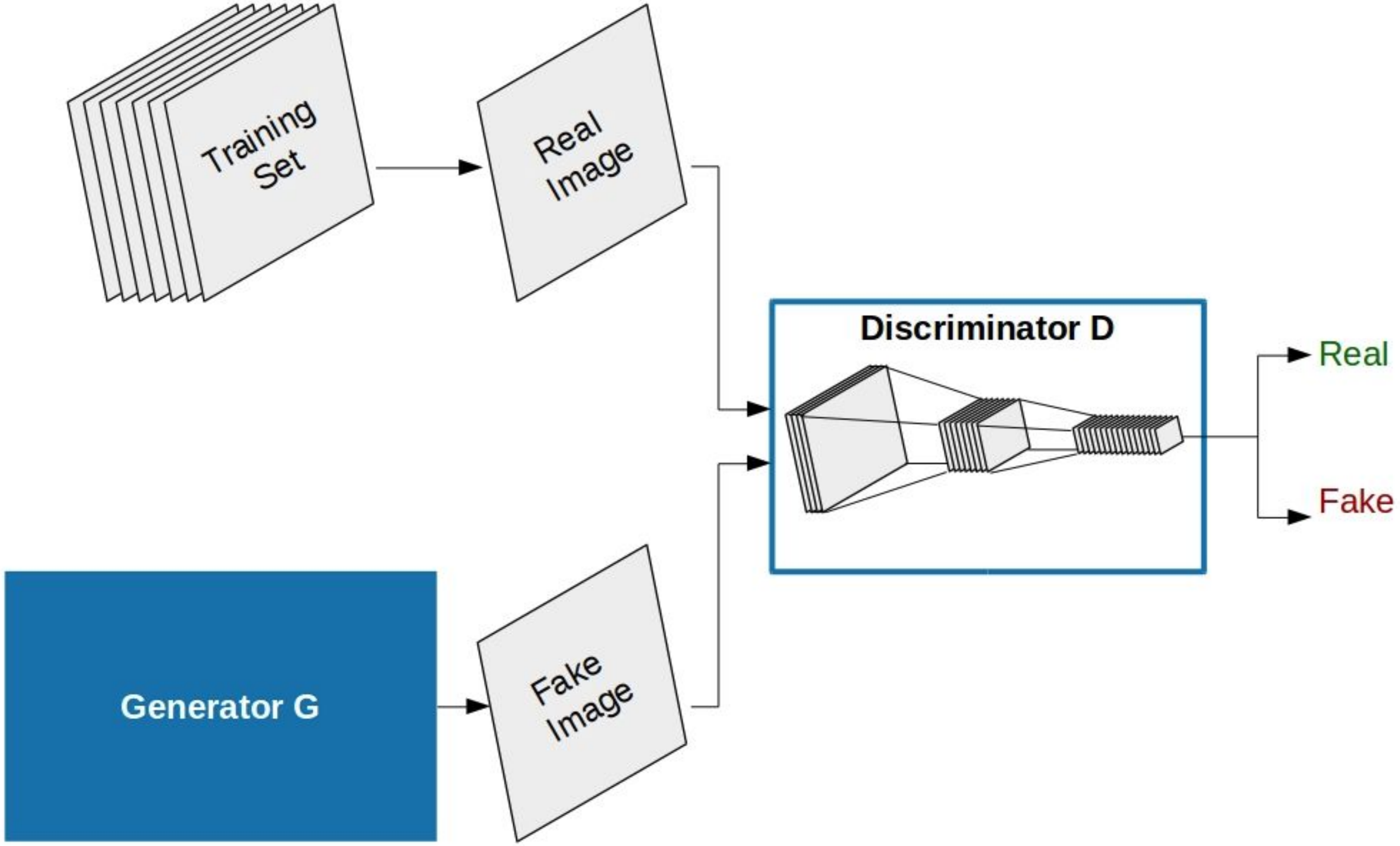
- AttnGAN
- StackGAN
- WGAN-CLS
- WGAN-CLS + Stage II of StackGAN
- StackGAN_v2

GENERATE

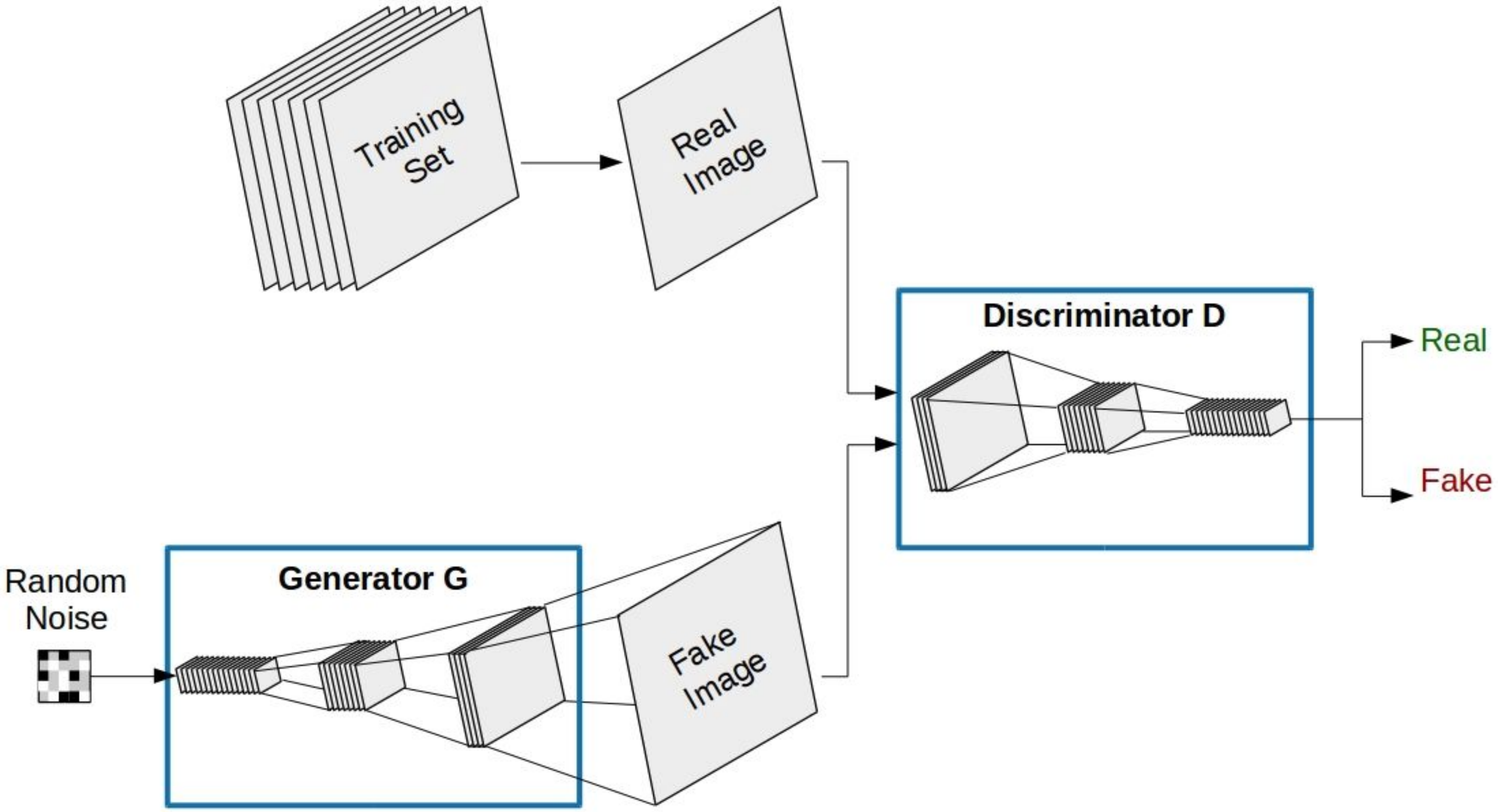
Generative Adversarial Networks



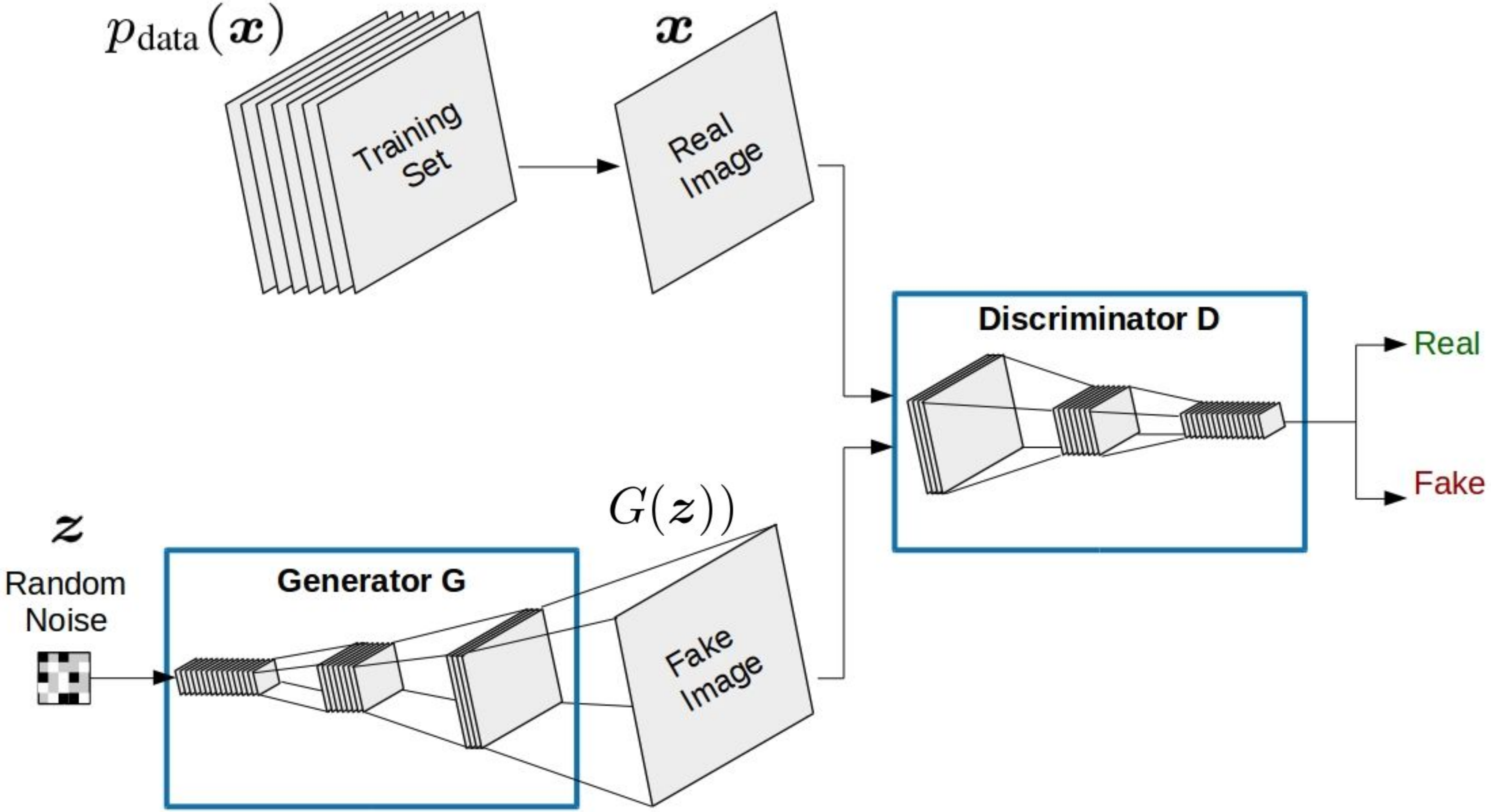
Generative Adversarial Networks



Generative Adversarial Networks



Generative Adversarial Networks



D(x)=1

D(x)=0

$$\min_G \max_D V(D, G) = \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z} \sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$$

Real or Fake?

Image Source: *Brock, Andrew, Jeff Donahue, and Karen Simonyan. "Large scale gan training for high fidelity natural image synthesis." arXiv preprint arXiv:1809.11096. 2018.*

Conditional GANs

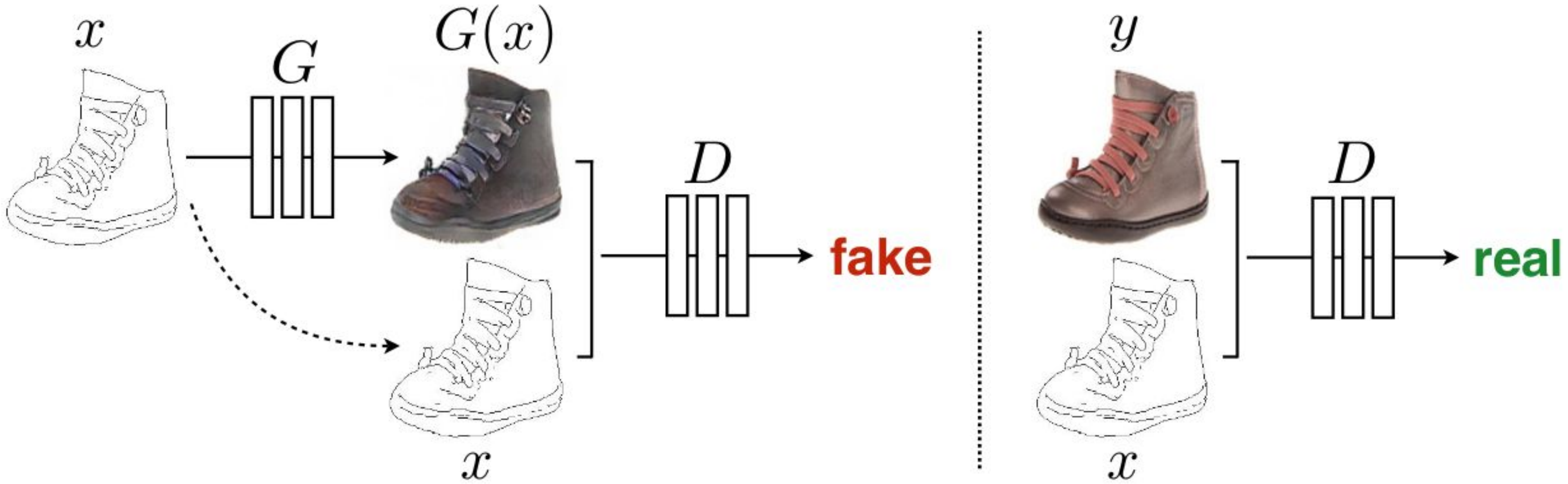
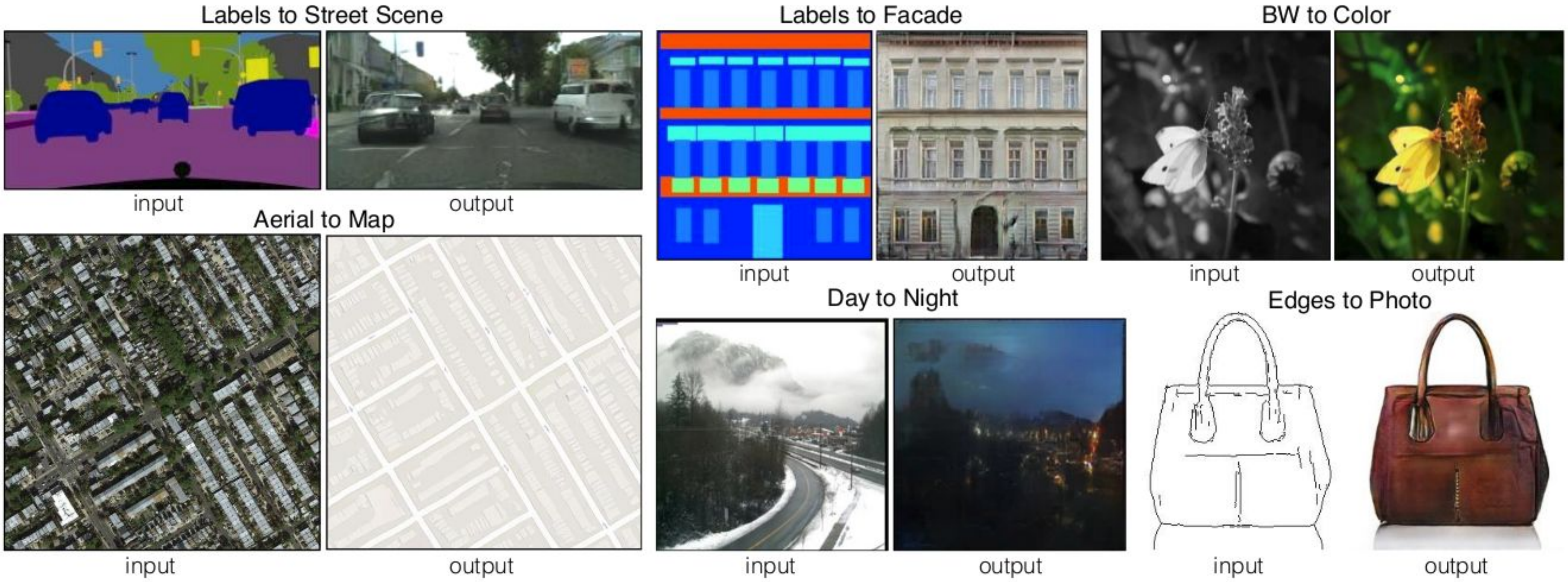


Image Source: *Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.*

Applications

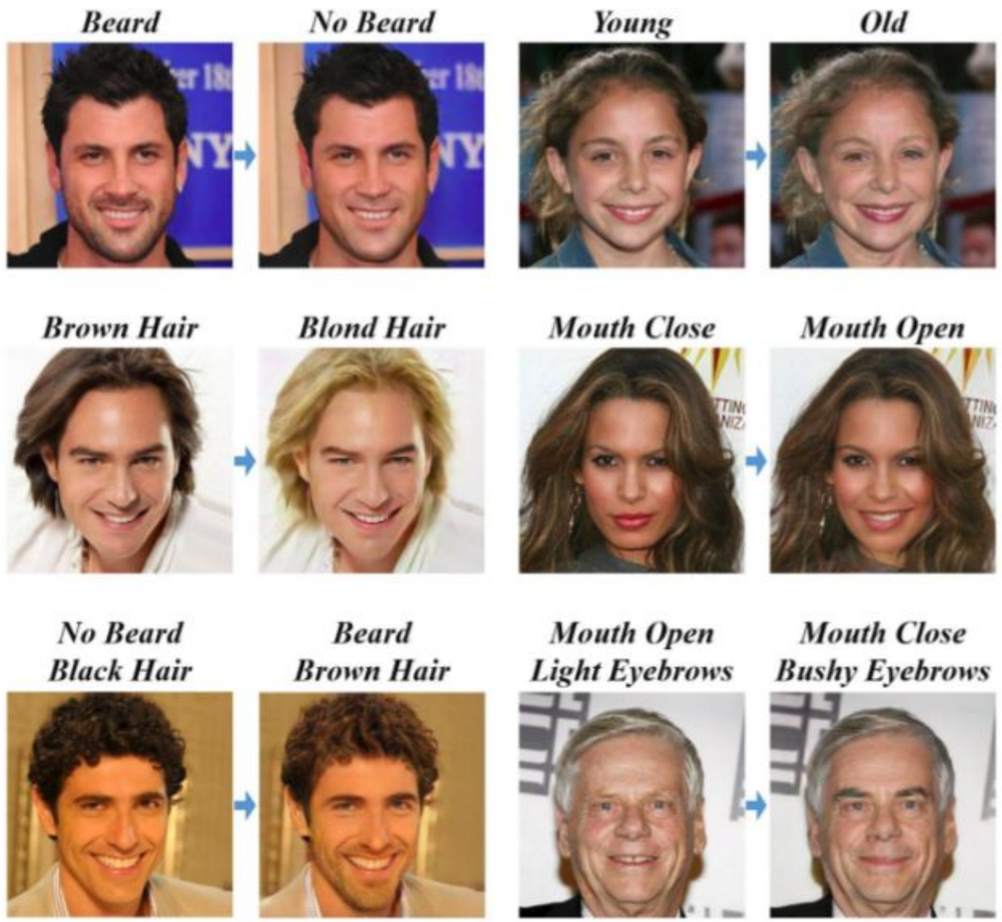
Pix2Pix: Image-to-Image Translation



Reference: *Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.*

Applications

AttGAN: Class-based image generation and modification



Reference: *He, Zhenliang, et al. "Attgan: Facial attribute editing by only changing what you want." IEEE Transactions on Image Processing. 2019.*

Applications

Image Generation from Text

- StackGAN
- StackGAN++
- AttnGAN

This bird is red with white and has a very short beak

Agenda

- Introduction
- **GANs for Image Generation from Text**
- Demo Session
- System Architecture
- Summary and Q&A

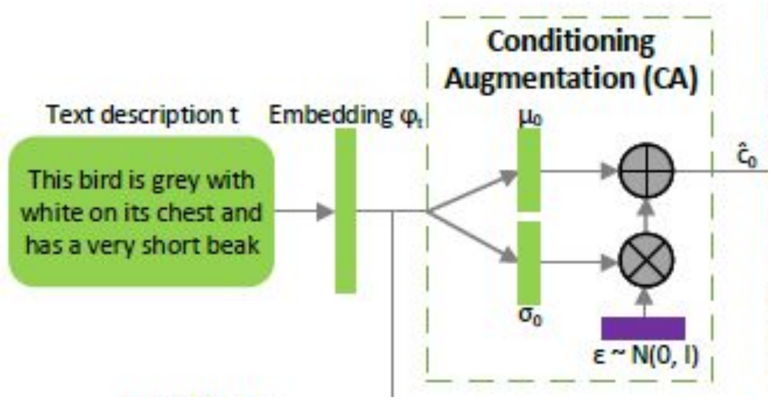
Dataset: Caltech-UCSD Birds 200-2011

- 11,788 images of birds
- includes annotations and image captions

Dataset available at: http://www.vision.caltech.edu/visipedia/CUB-200-2011.html

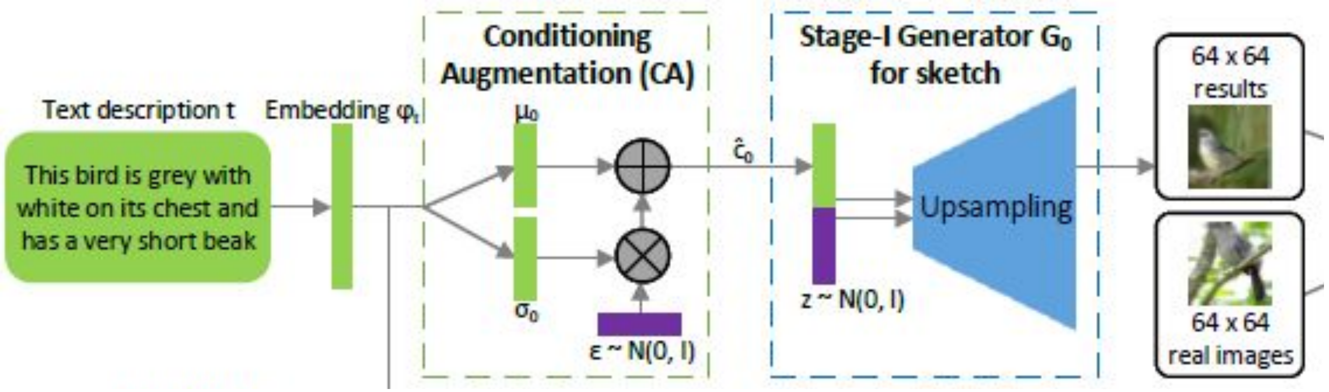
StackGAN

Idea: stack multiple GANs together!



StackGAN

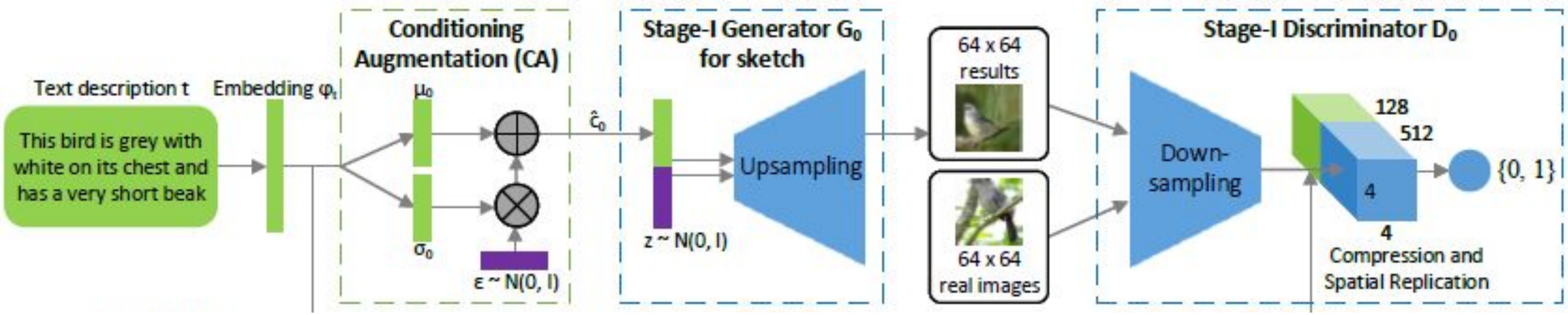
Idea: stack multiple GANs together!



Reference: *Zhang, Han, et al. "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks." Proceedings of the IEEE International Conference on Computer Vision. 2017.*

StackGAN

Idea: stack multiple GANs together!

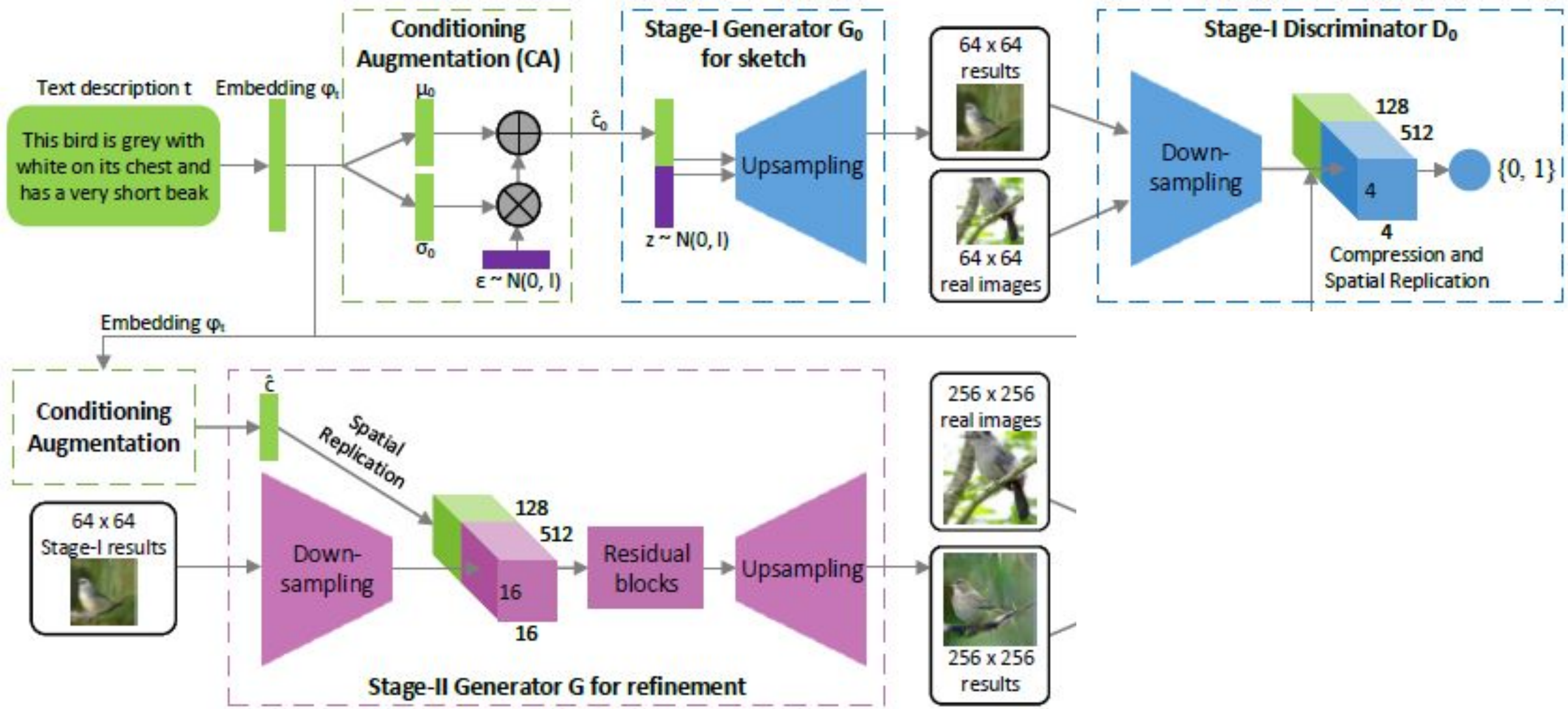


trained independently from the rest

Reference: *Zhang, Han, et al. "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks." Proceedings of the IEEE International Conference on Computer Vision. 2017.*

StackGAN

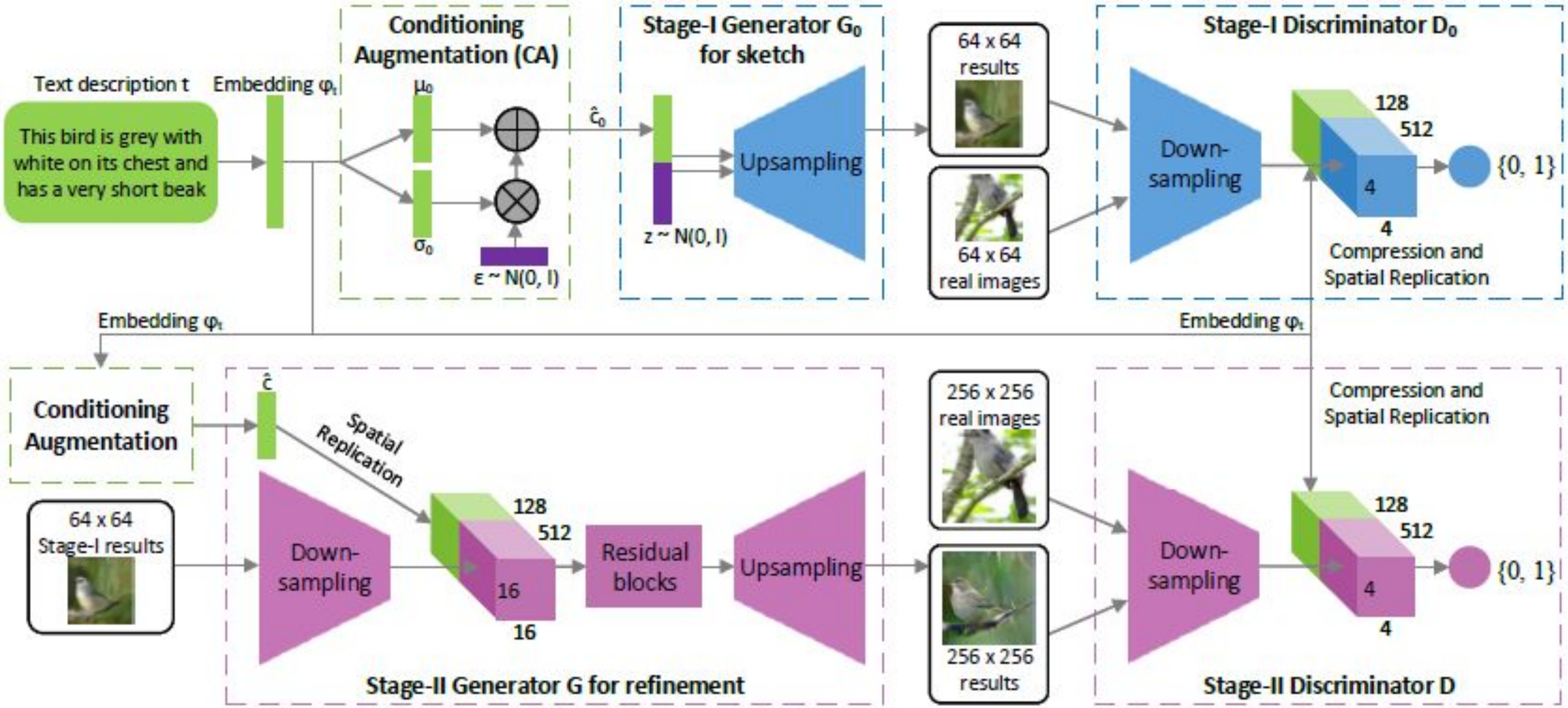
Idea: stack multiple GANs together!



Reference: *Zhang, Han, et al. "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks." Proceedings of the IEEE International Conference on Computer Vision. 2017.*

StackGAN

Idea: stack multiple GANs together!

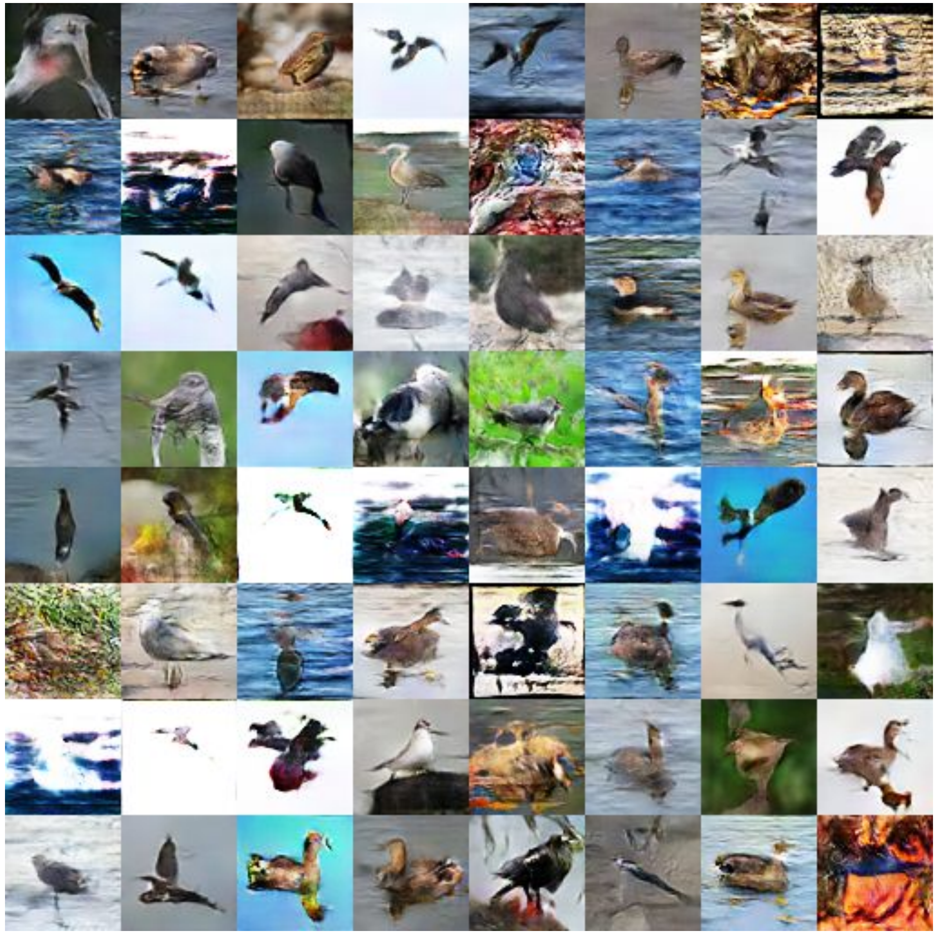


Reference: Zhang, Han, et al. *"Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks." Proceedings of the IEEE International Conference on Computer Vision. 2017.*

StackGAN: First Results

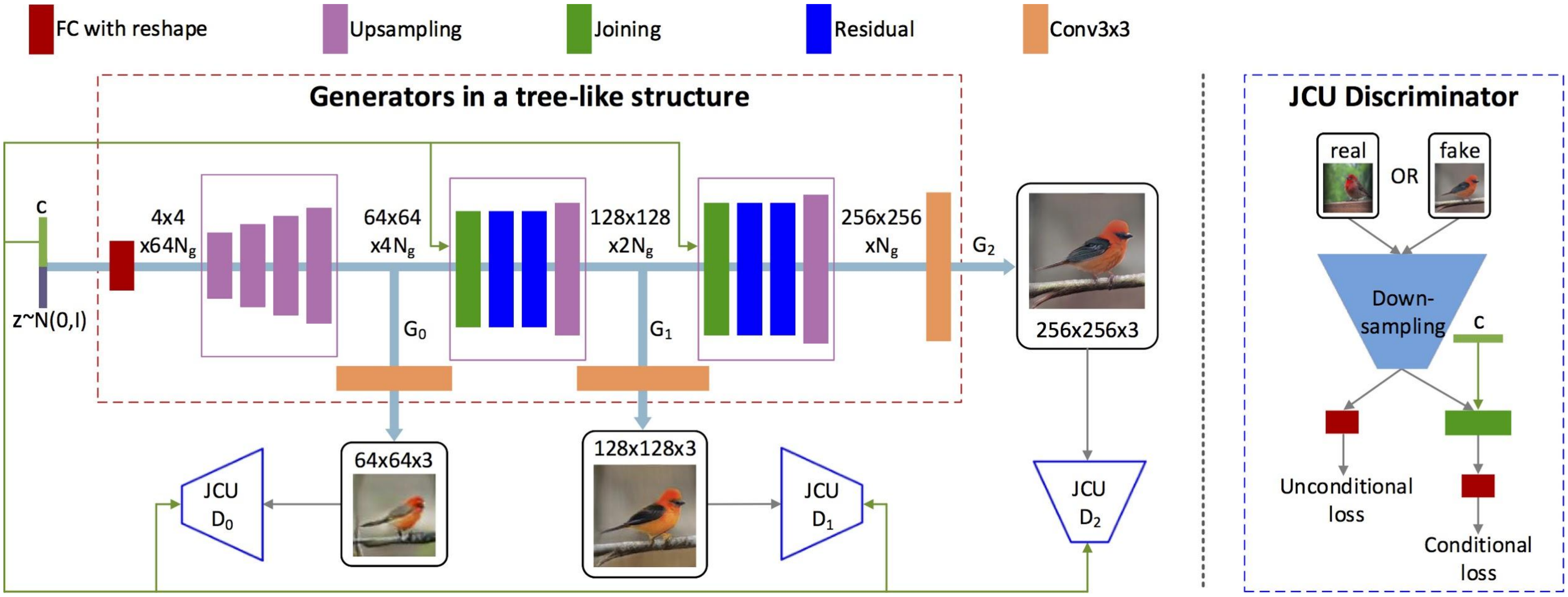
StackGAN Stage I	StackGAN Stage II
	

The Mode Collapse Problem

StackGAN Stage I	better would be something like...
	

StackGAN++

Idea: stack even more GANs together!

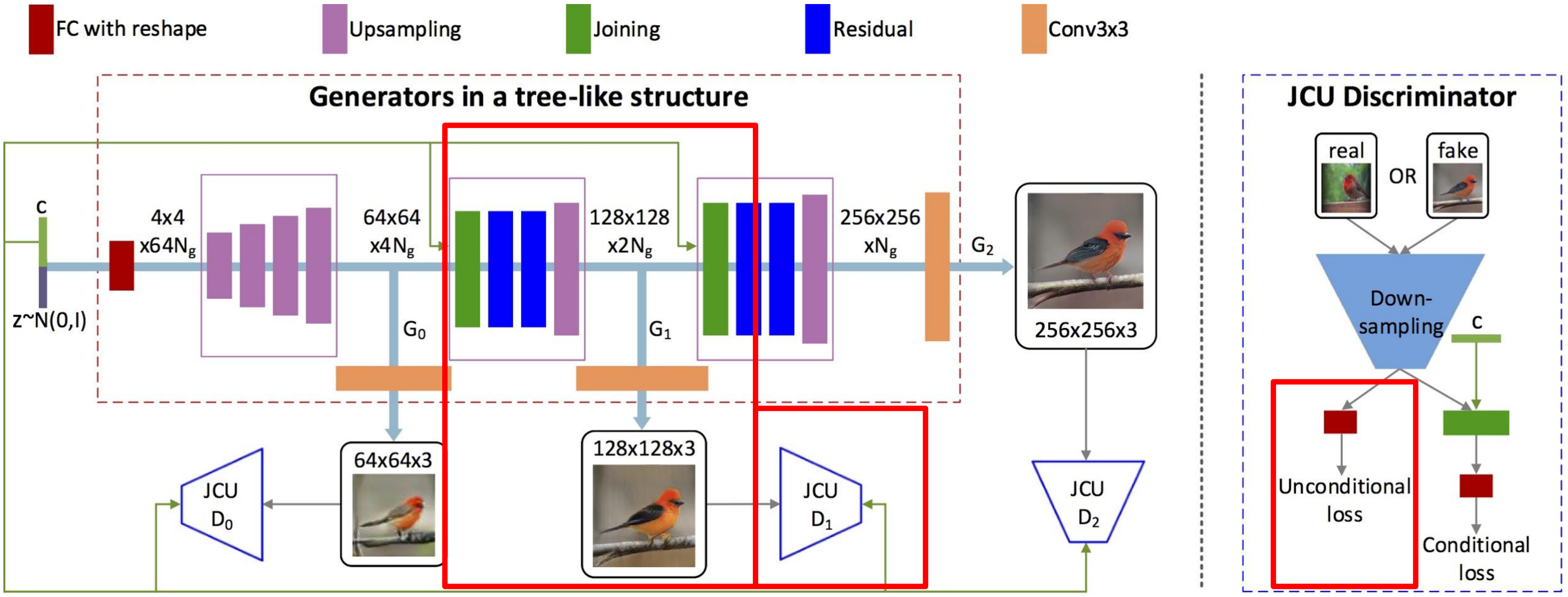


Reference: *Zhang, Han, et al. "Stackgan++: Realistic image synthesis with stacked generative adversarial networks." arXiv preprint arXiv:1710.10916. 2017.*

StackGAN++

Idea: stack even more GANs together!

all pairs trained together each step



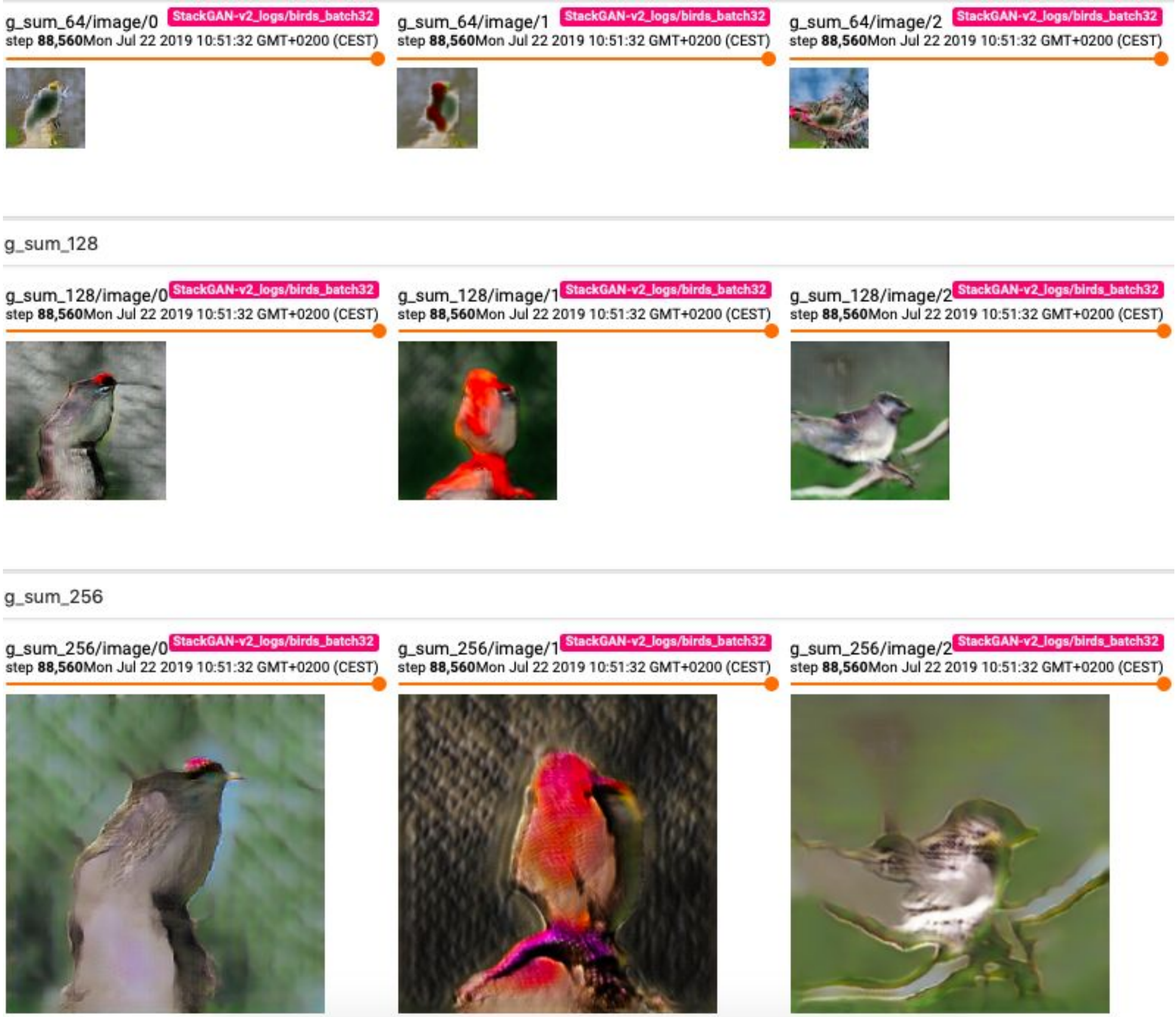
Reference: *Zhang, Han, et al. "Stackgan++: Realistic image synthesis with stacked generative adversarial networks." arXiv preprint arXiv:1710.10916. 2017.*

StackGAN++

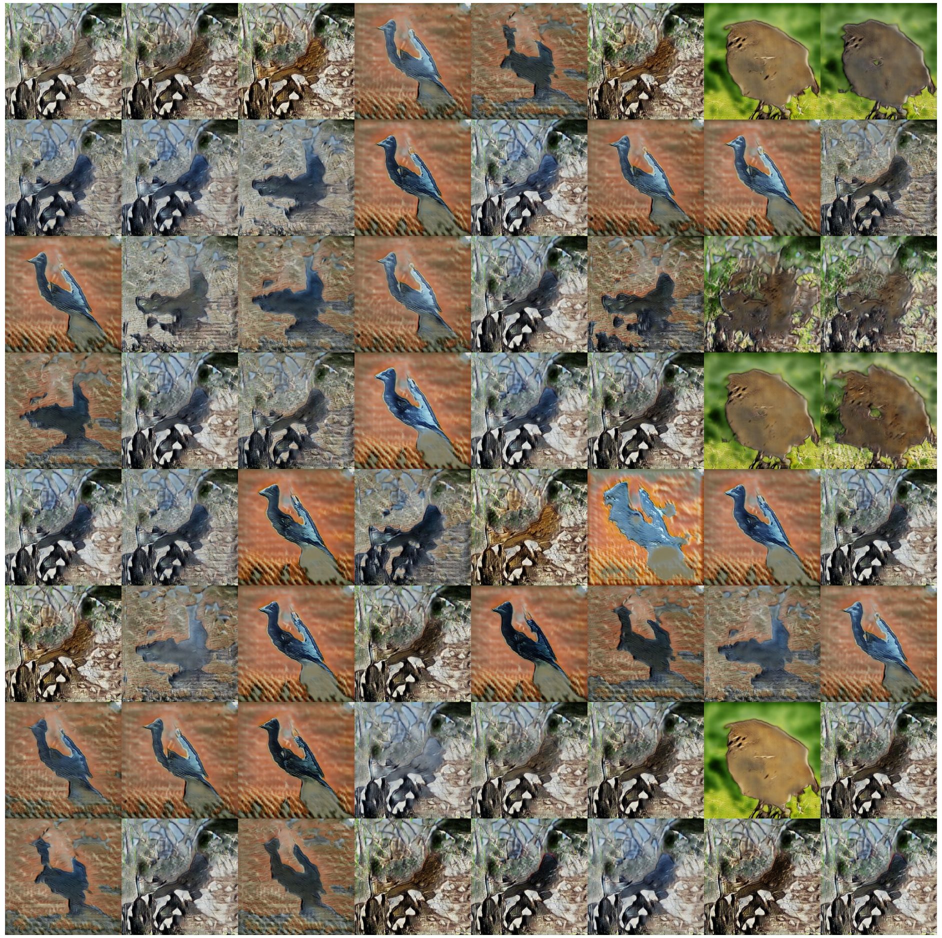
G_0: 64x64 image

G_1: 128x128 image

G_2: 256x256 image



StackGAN++: Results

StackGAN Stage II	StackGAN++
	

Wasserstein GAN (WGAN)

Idea: use a different loss function

$$\min_G \max_D \mathbb{E}_{x \sim P_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim P_z(z)}[\log(1 - D(G(z)))]$$

standard GAN loss function

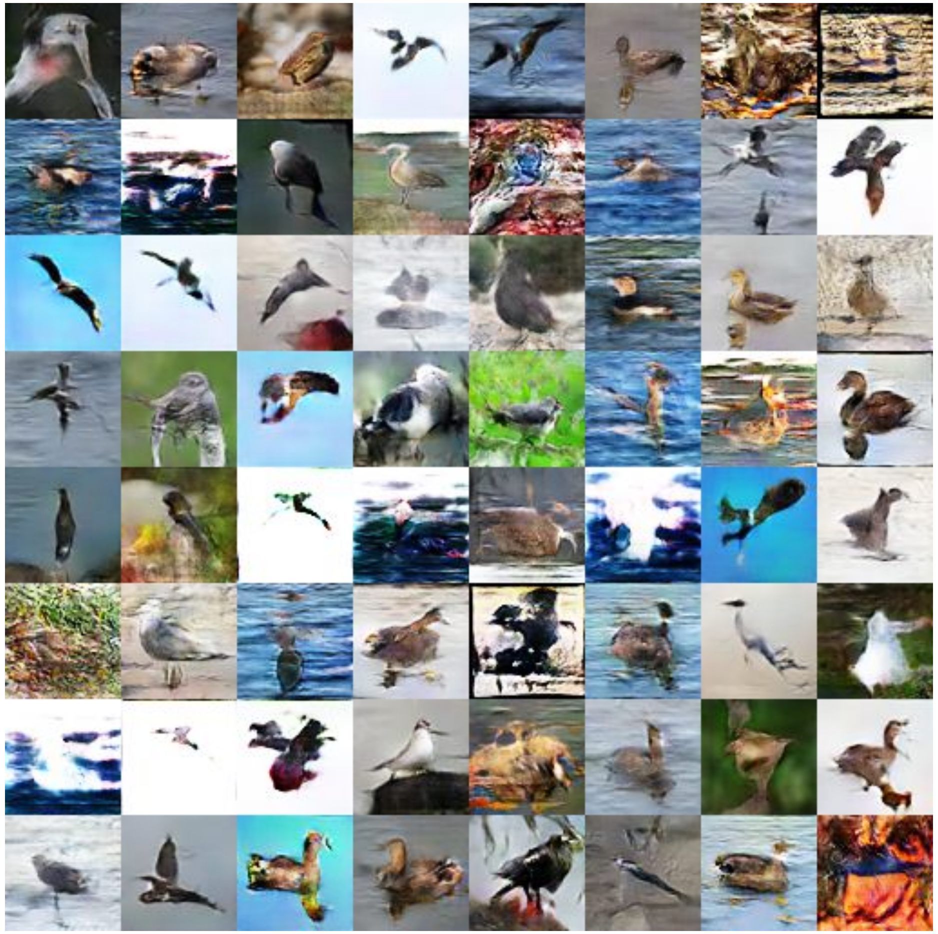
$$\downarrow$$

$$\min_G \max_{D \in \mathscr{D}} \mathbb{E}_{x \sim P_{data}(x)}[D(x)] - \mathbb{E}_{z \sim P_z(z)}[D(G(z)))]$$

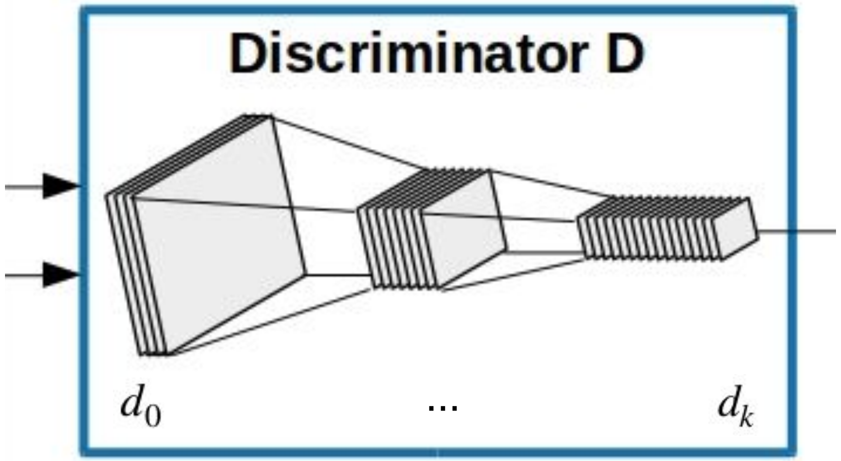
using Wasserstein-1 distance

Reference: *Arjovsky, Martin, Soumith Chintala, and Léon Bottou. "Wasserstein generative adversarial networks." International conference on machine learning. 2017.*

WGAN: Results

StackGAN Stage I	WGAN
	

Spectral Normalization



$$\|d_l\|_{Lip} = \sigma(W^l),\ W_{SN} := \frac{W}{\sigma(W)} \Rightarrow \|W_{SN}\|_{Lip} = 1$$

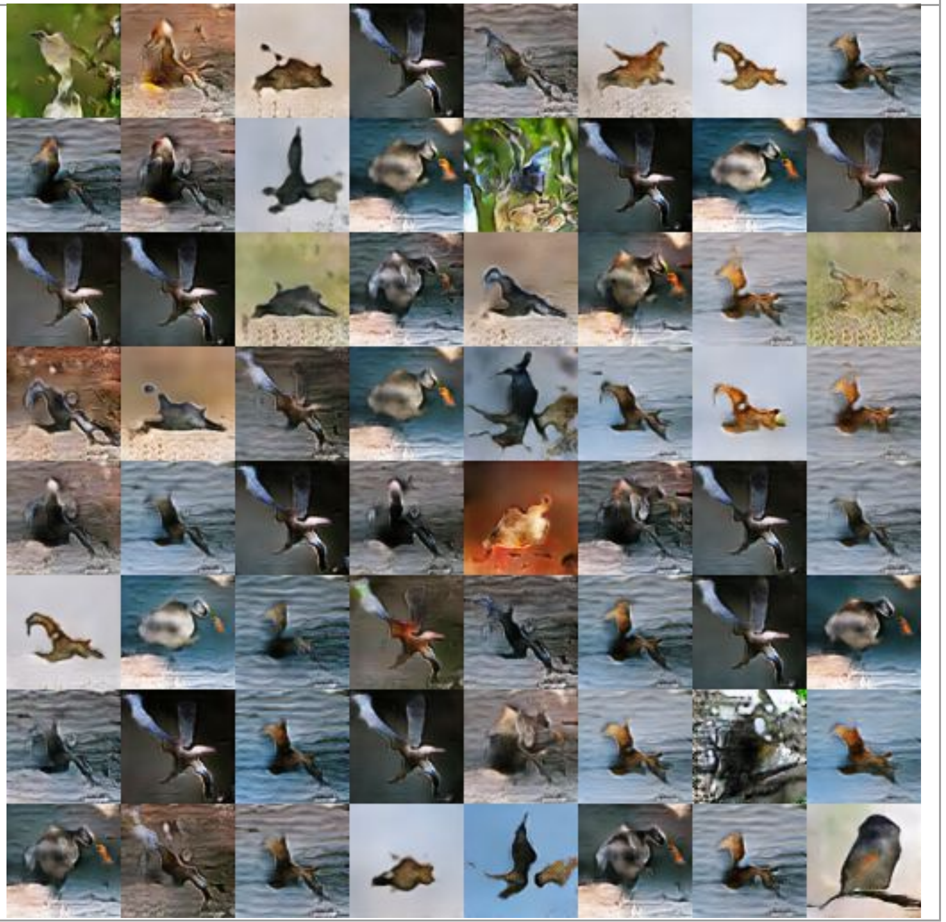
$$\|D\|_{Lip} \leq \prod_{l=0}^{k} \sigma(W^l) \Rightarrow \|D\|_{Lip} \leq \prod_{l=0}^{k} \sigma(W^l_{SN}) = 1$$

Reference: *Miyato, Takeru, et al. "Spectral normalization for generative adversarial networks." arXiv preprint arXiv:1802.05957. 2018.*

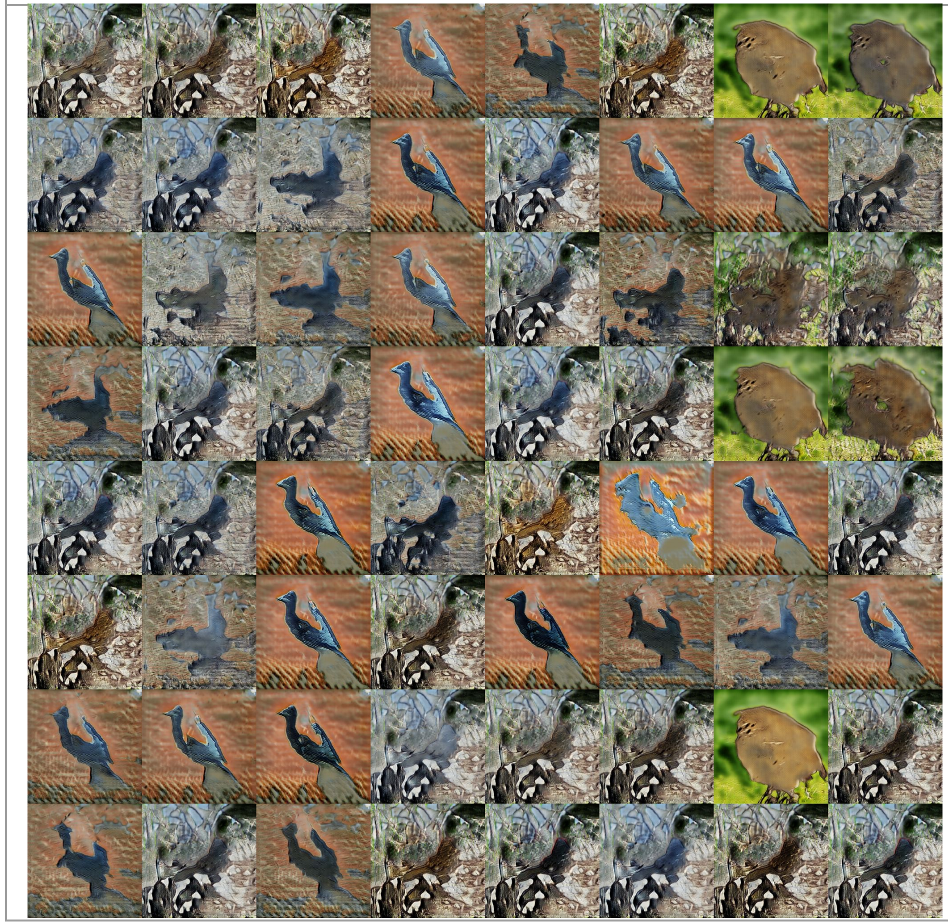
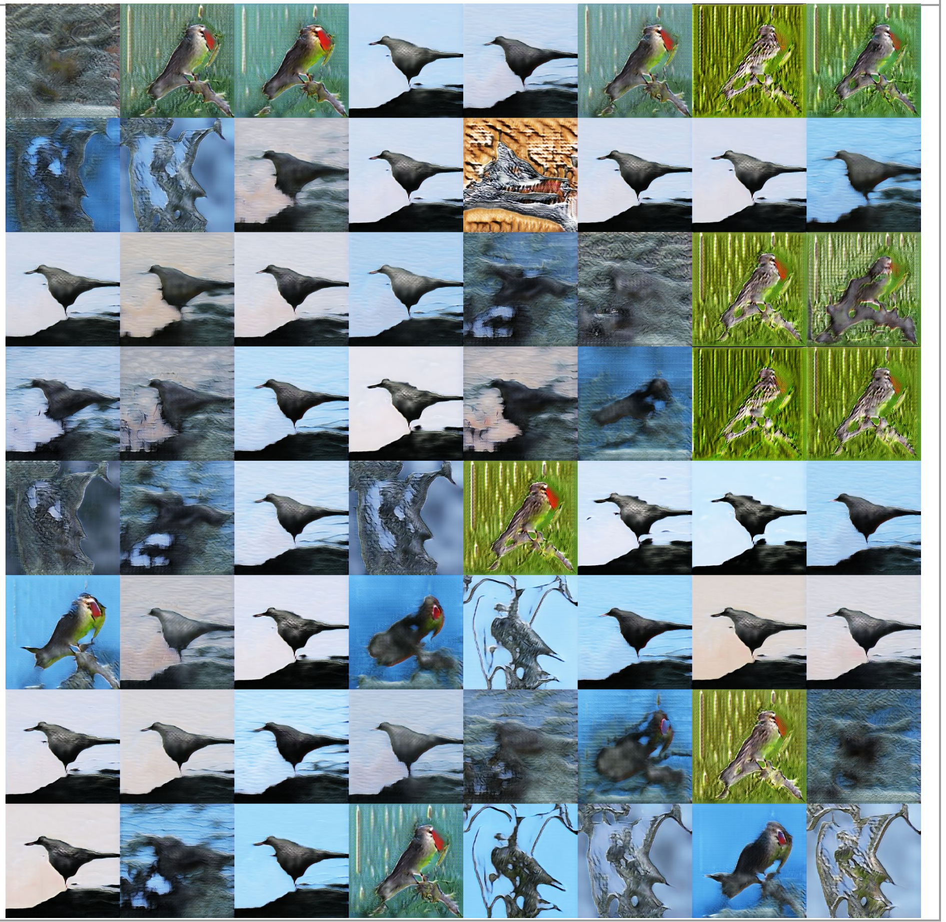
Spectral Normalization: Results

StackGAN Stage I	StackGAN Stage I, SN in D	StackGAN Stage I, SN in D and G
		

Different batch size

StackGAN Stage I, Batch Size 8	StackGAN Stage I, Batch Size 64
	

Different batch size

StackGAN++, Batch Size 8	StackGAN++, Batch Size 32
	

One-sided Label smoothing

Idea: smooth the label, i.e .9 and .1 instead of 1 and 0

- Smooth only positive labels as .9, and leave negative labels as 0 -> one-sided

$$\min_G \max_D \mathbb{E}_{x\sim P_{data}(x)}[\quad \log D(x)] + \mathbb{E}_{z\sim P_z(z)}[\log(1 - D(G(z)))]$$

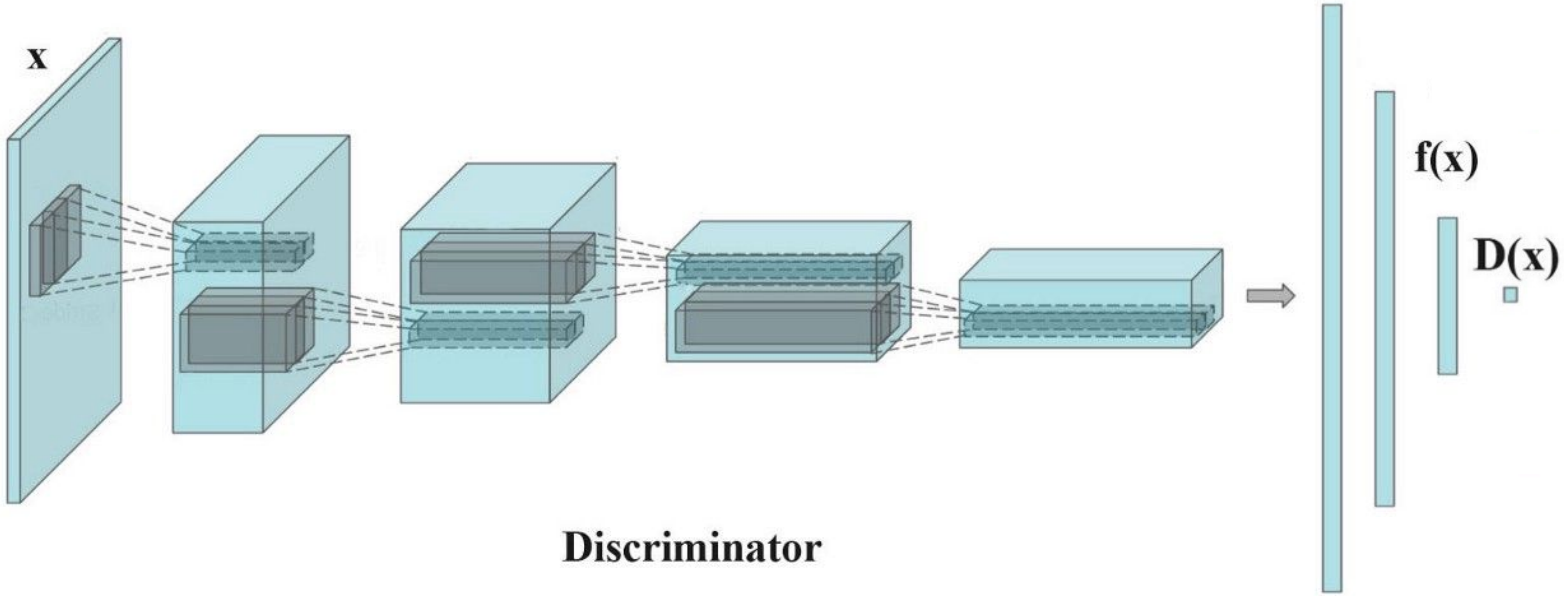
- It is valuable when the label cannot be believed 100%
- In this case, by smoothing the label, the model can consequently improve its robustness and performance

Reference: *Salimans, Tim, et al. "Improved techniques for training gans." Advances in neural information processing systems. 2016.*

Mini-batch discrimination

Idea: Check the similarity between samples in the same batch

- If the generator has mode collapse, the similarity should be higher than real image
- The mini-batch discrimination layer is not for just simple similarity calculation, but a trainable layer

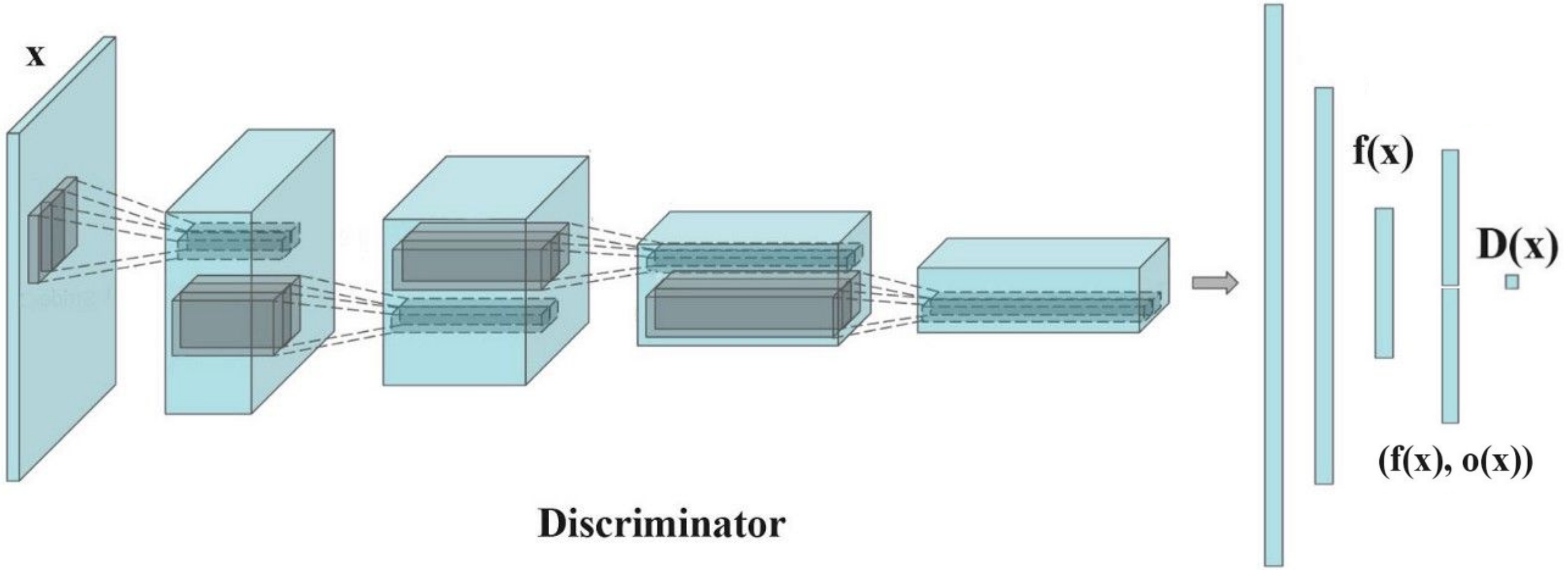


Reference: *Salimans, Tim, et al. "Improved techniques for training gans." Advances in neural information processing systems. 2016.*
Image Source: https://towardsdatascience.com/gan-ways-to-improve-gan-performance-acf37f9f59b

Mini-batch discrimination

Idea: Check the similarity between samples in the same batch

- If the generator has mode collapse, the similarity should be higher than real image
- The mini-batch discrimination layer is not for just simple similarity calculation, but a trainable layer

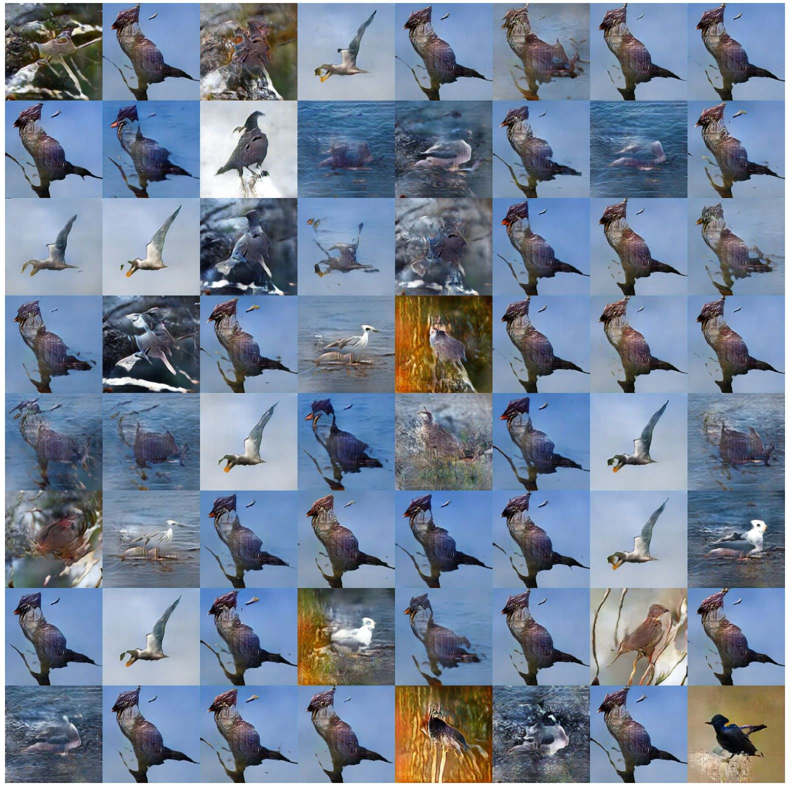


Reference: *Salimans, Tim, et al. "Improved techniques for training gans." Advances in neural information processing systems. 2016.*
Image Source: https://towardsdatascience.com/gan-ways-to-improve-gan-performance-acf37f9f59b

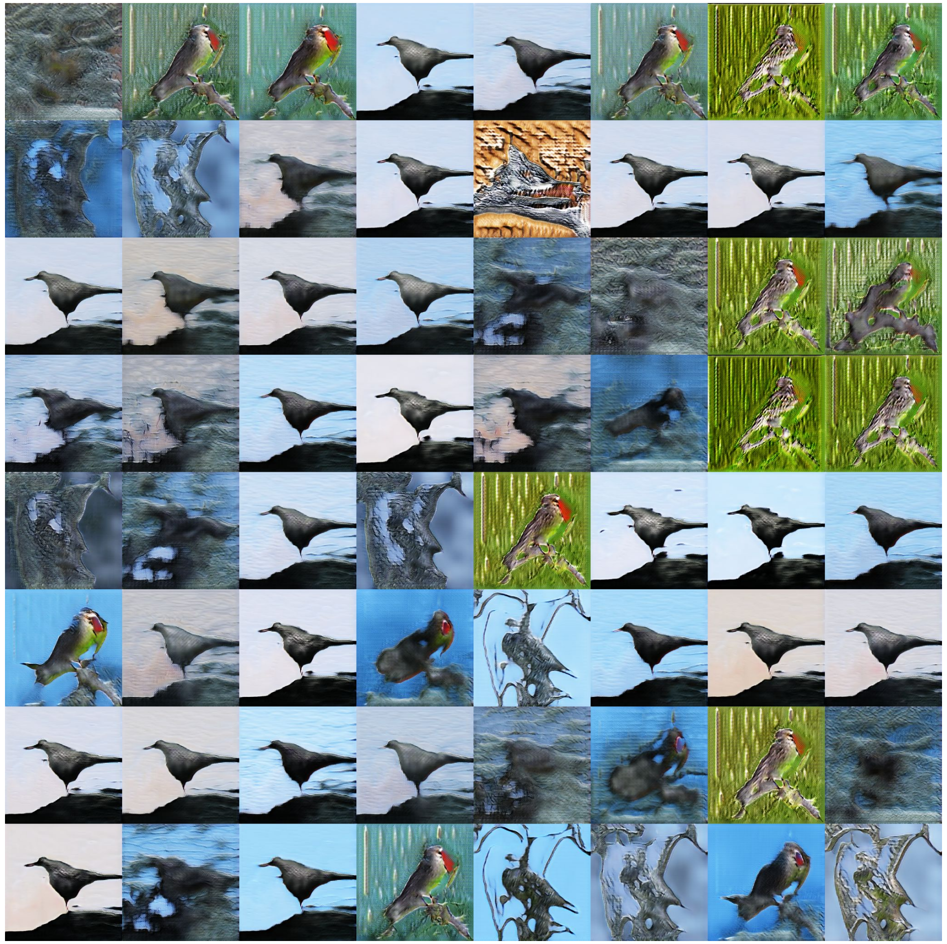
Mini-batch discrimination: Results

StackGAN Stage I, with residual layer	StackGAN Stage I, without residual layer
	

Mini-batch discrimination: Results

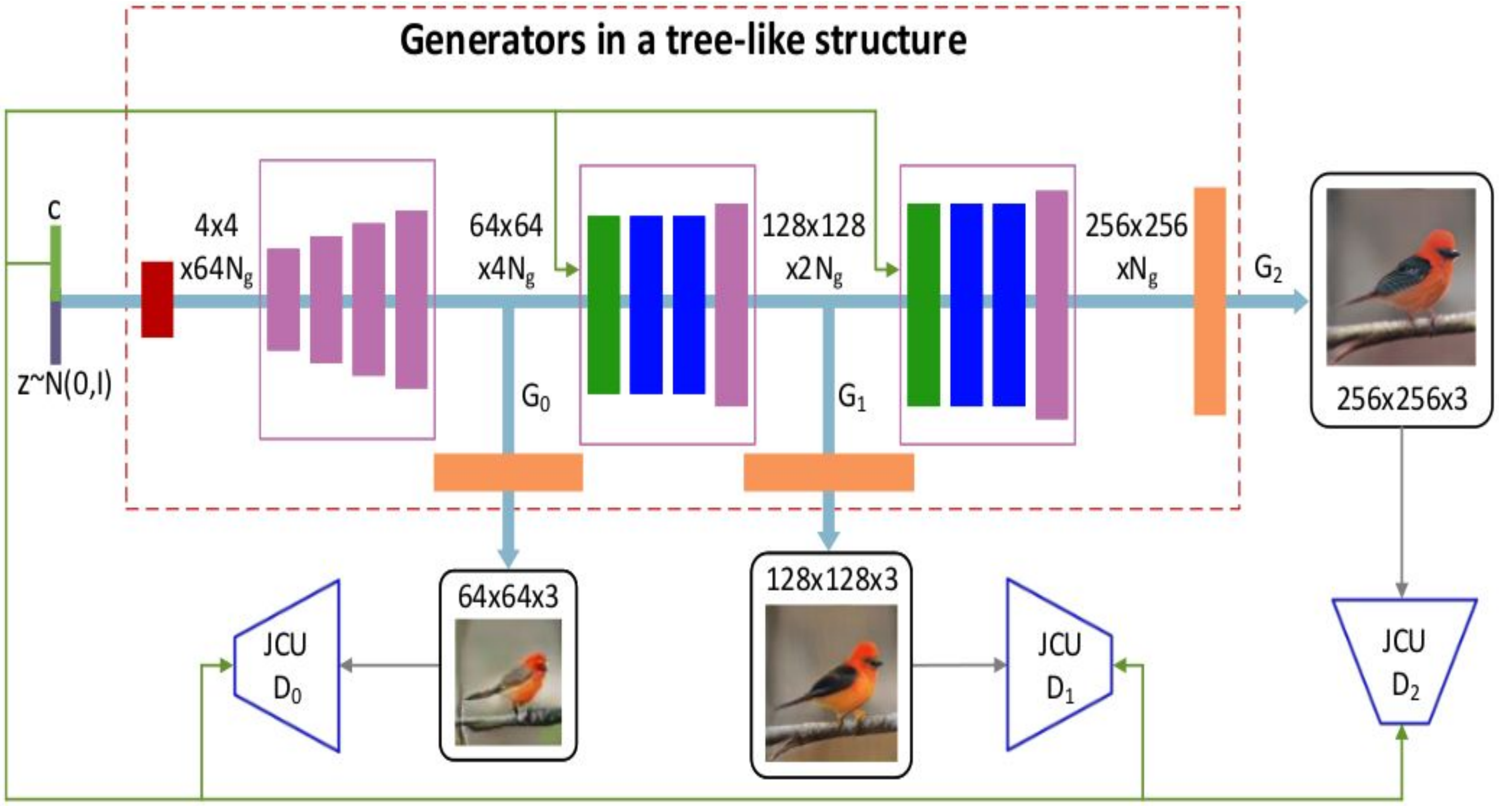
StackGAN Stage II, no residual layer, 8 conv layers, 8192 latent dim	StackGAN StageII, no residual layer, 7 conv layers, 1024 latent dim.	StackGAN StageII, no residual layer, 5 conv layers, 8192 latent dim
		

Mini-batch discrimination: Results

StackGAN++, without mini-batch discrimination	StackGAN++, with mini-batch discrimination
	

AttnGAN

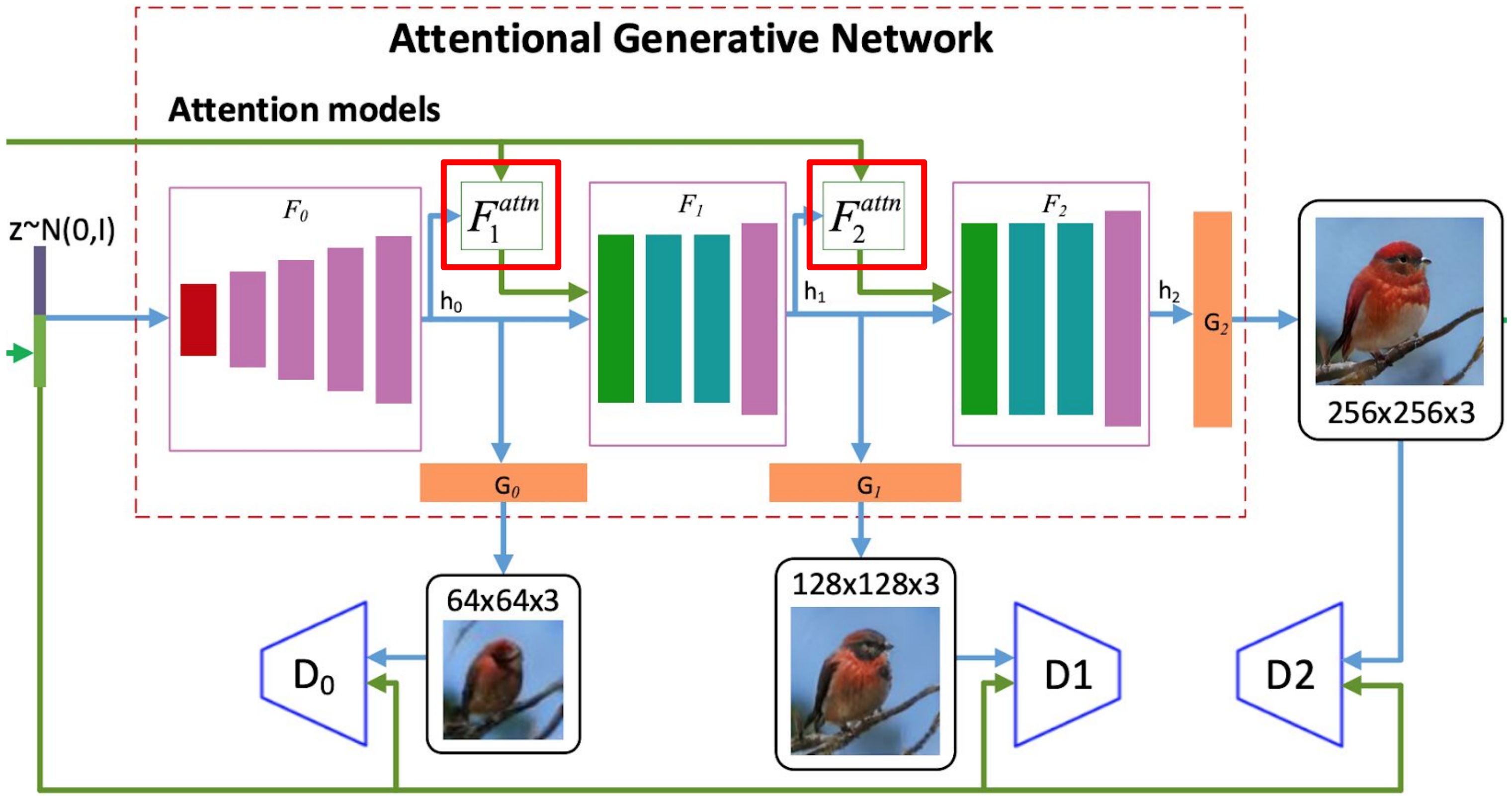
Idea: use StackGAN++ model as base architecture



Reference: *Zhang, Han, et al. "Stackgan++: Realistic image synthesis with stacked generative adversarial networks." arXiv preprint arXiv:1710.10916. 2017.*

AttnGAN

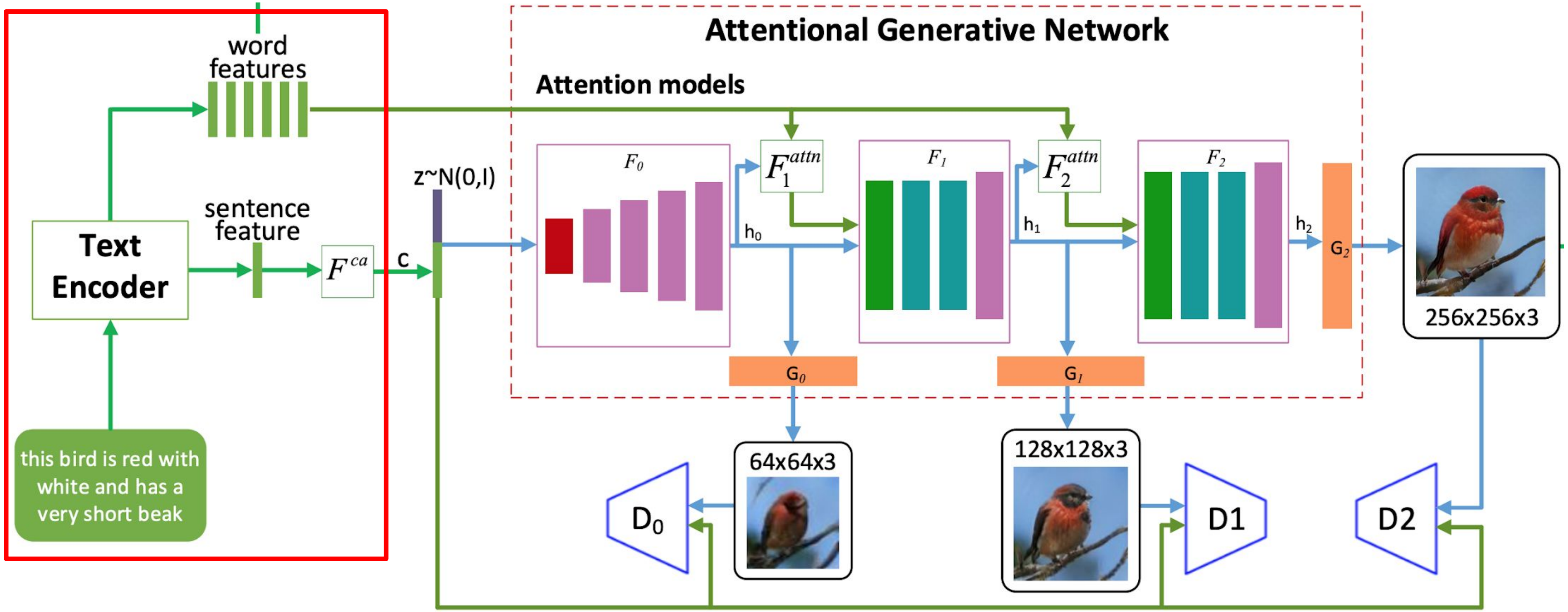
Idea: attention to words for image generation



Reference: *Xu, Tao, et al. "Attngan: Fine-grained text to image generation with attentional generative adversarial networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.*

AttnGAN

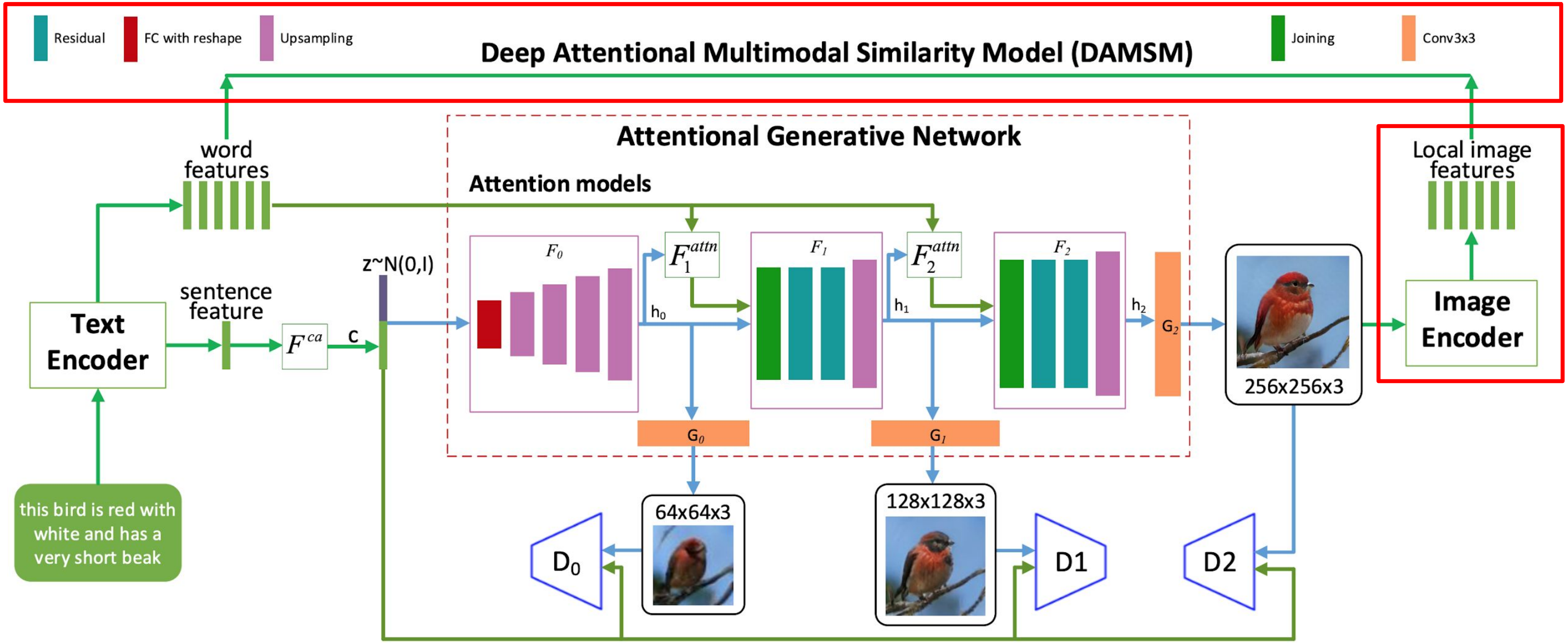
Idea: train own Text Encoder (Bidirectional LSTM)



Reference: *Xu, Tao, et al. "Attngan: Fine-grained text to image generation with attentional generative adversarial networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.*

AttnGAN

Idea: compare the relevance between final image and word features



Reference: Xu, Tao, et al. *"Attngan: Fine-grained text to image generation with attentional generative adversarial networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.*

AttnGAN: Results

a bird	a yellow bird with short beak	in the forest, there is a gray bird with red head who has a long beak

Agenda

- Introduction
- GANs for Image Generation from Text
- **Demo Session**
- System Architecture
- Summary and Q&A

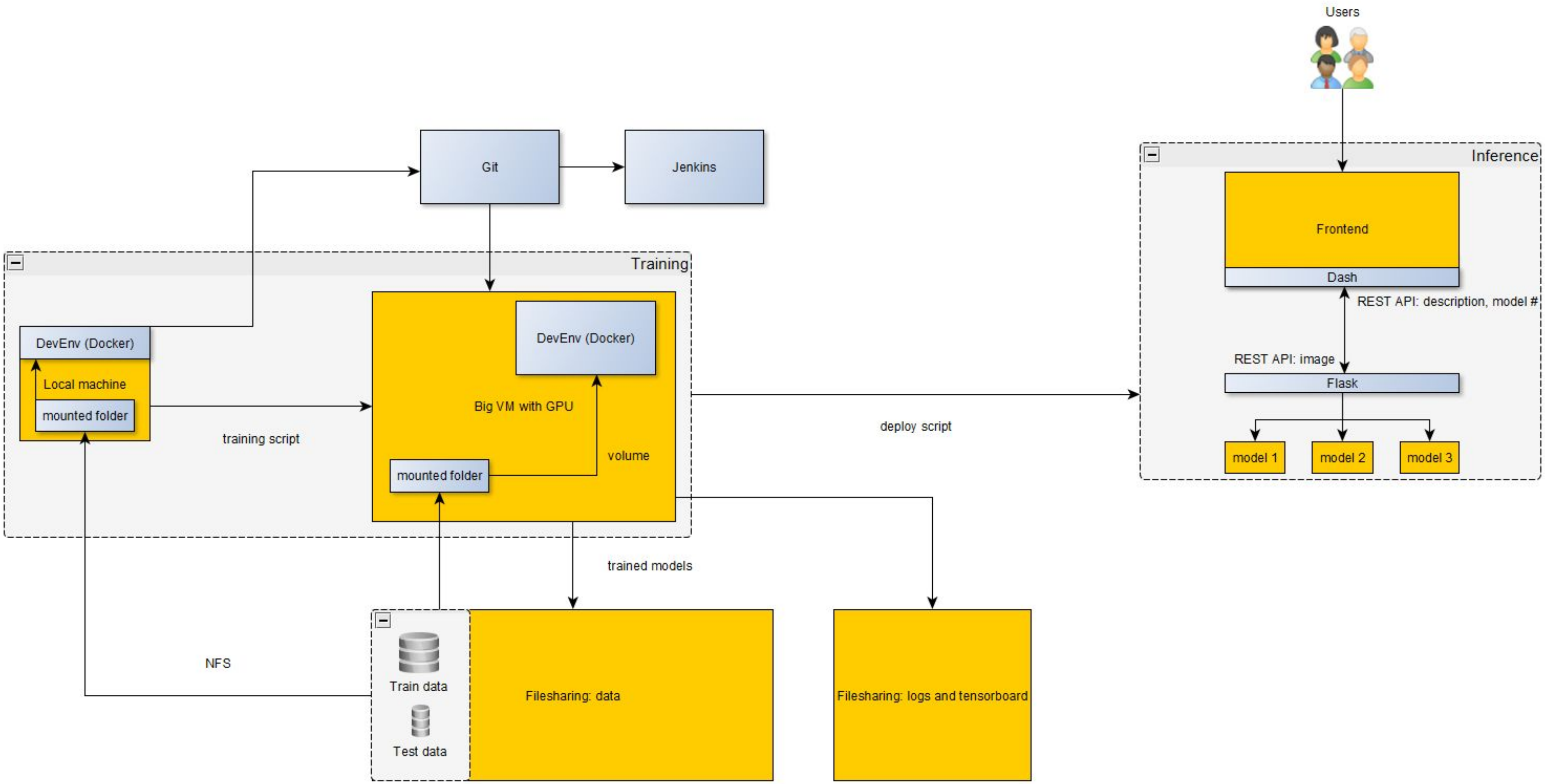
Demo Session

http://10.195.1.122:8050/

Agenda

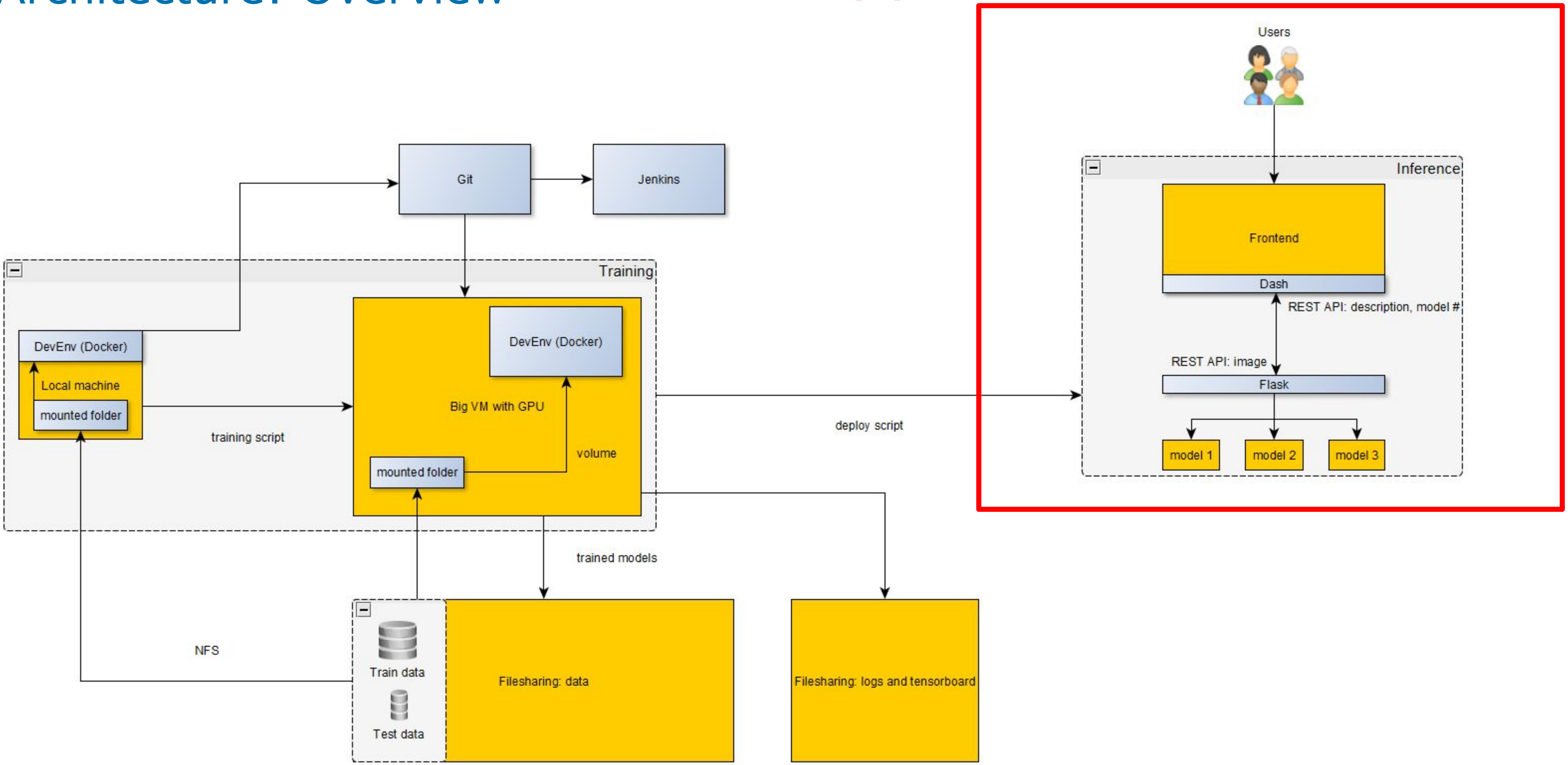
- Introduction
- GANs for Image Generation from Text
- Demo Session
- **System Architecture**
- Summary and Q&A

Architecture: Overview

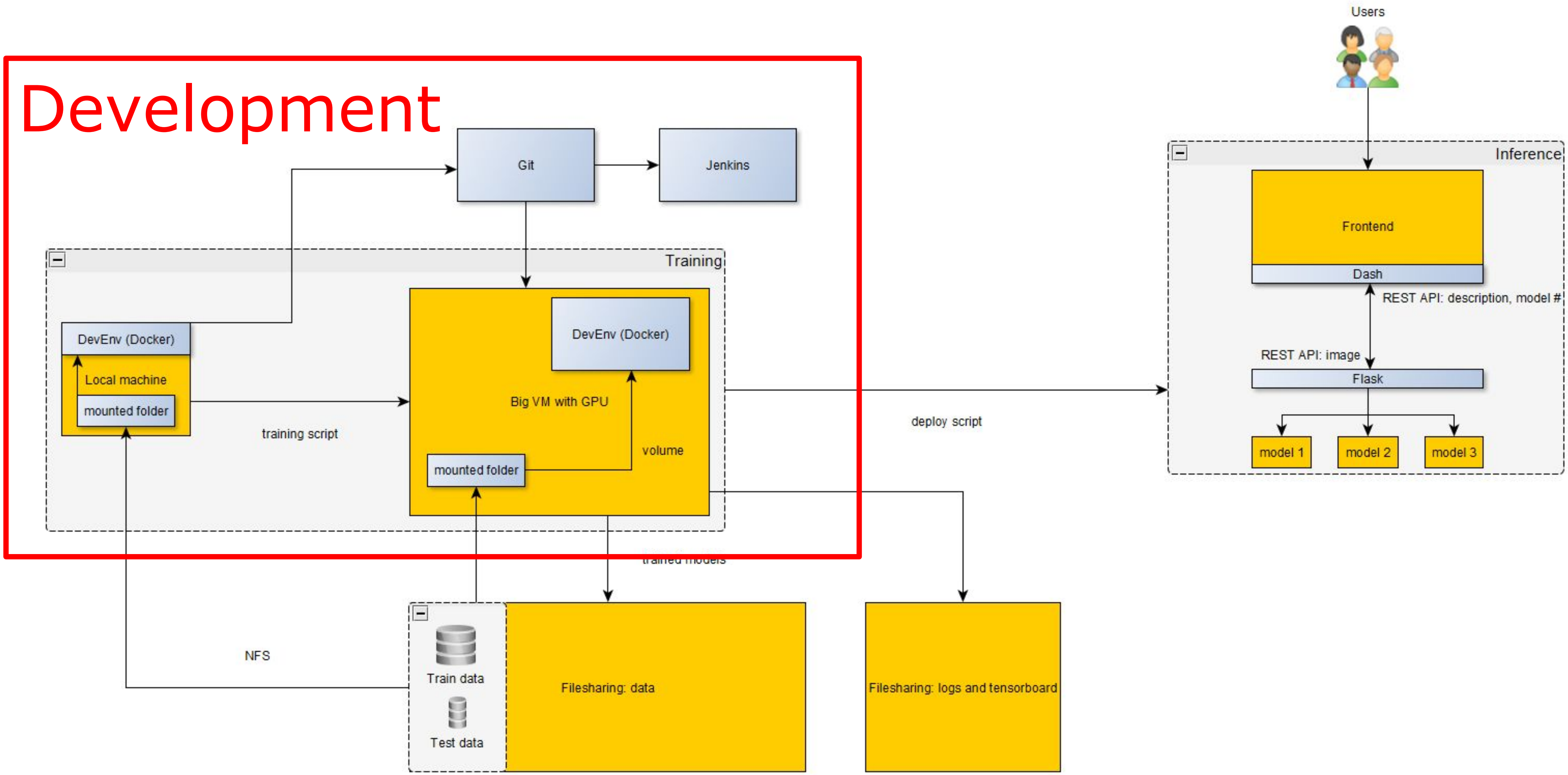


Architecture: Overview

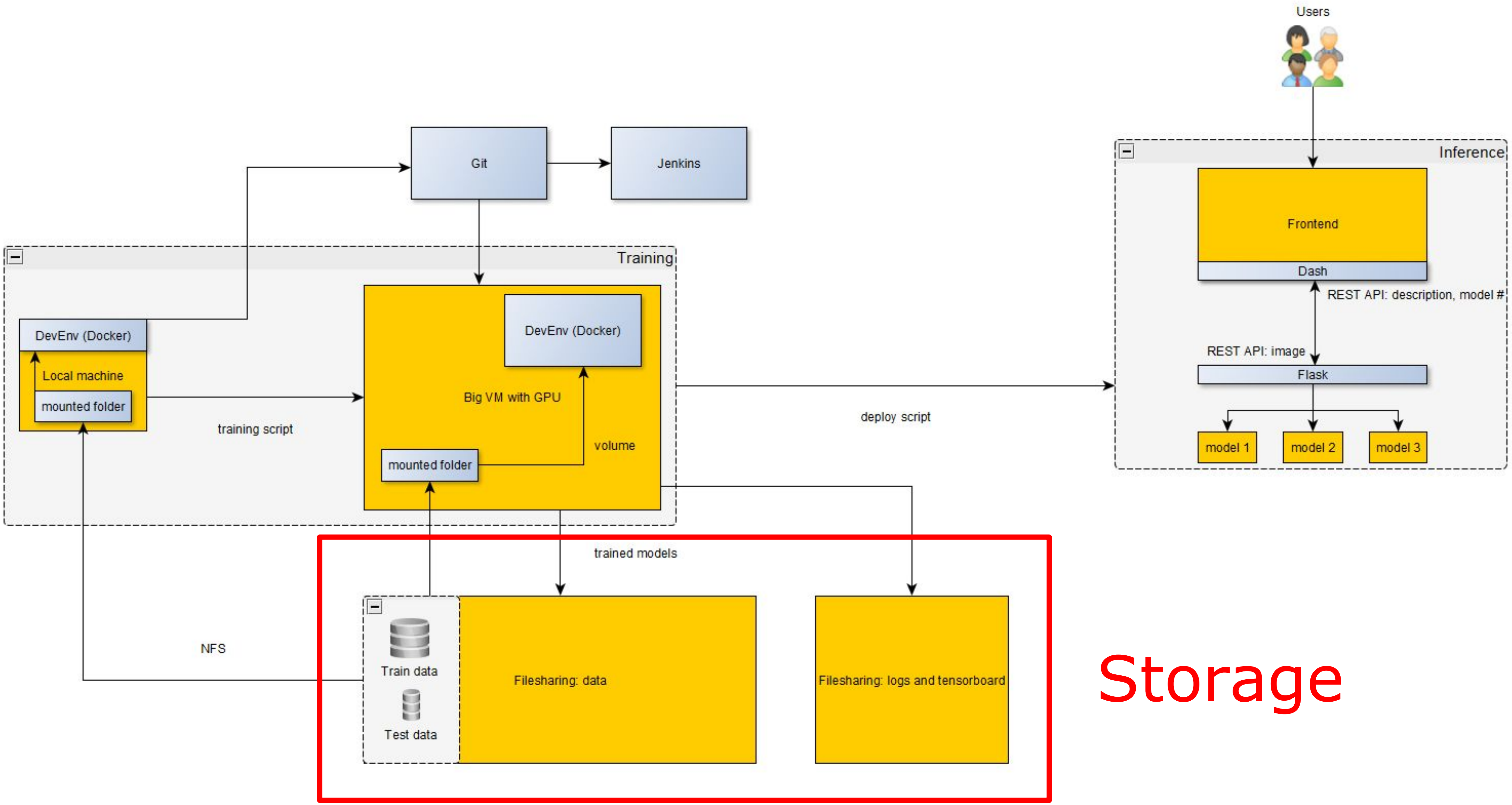
Application



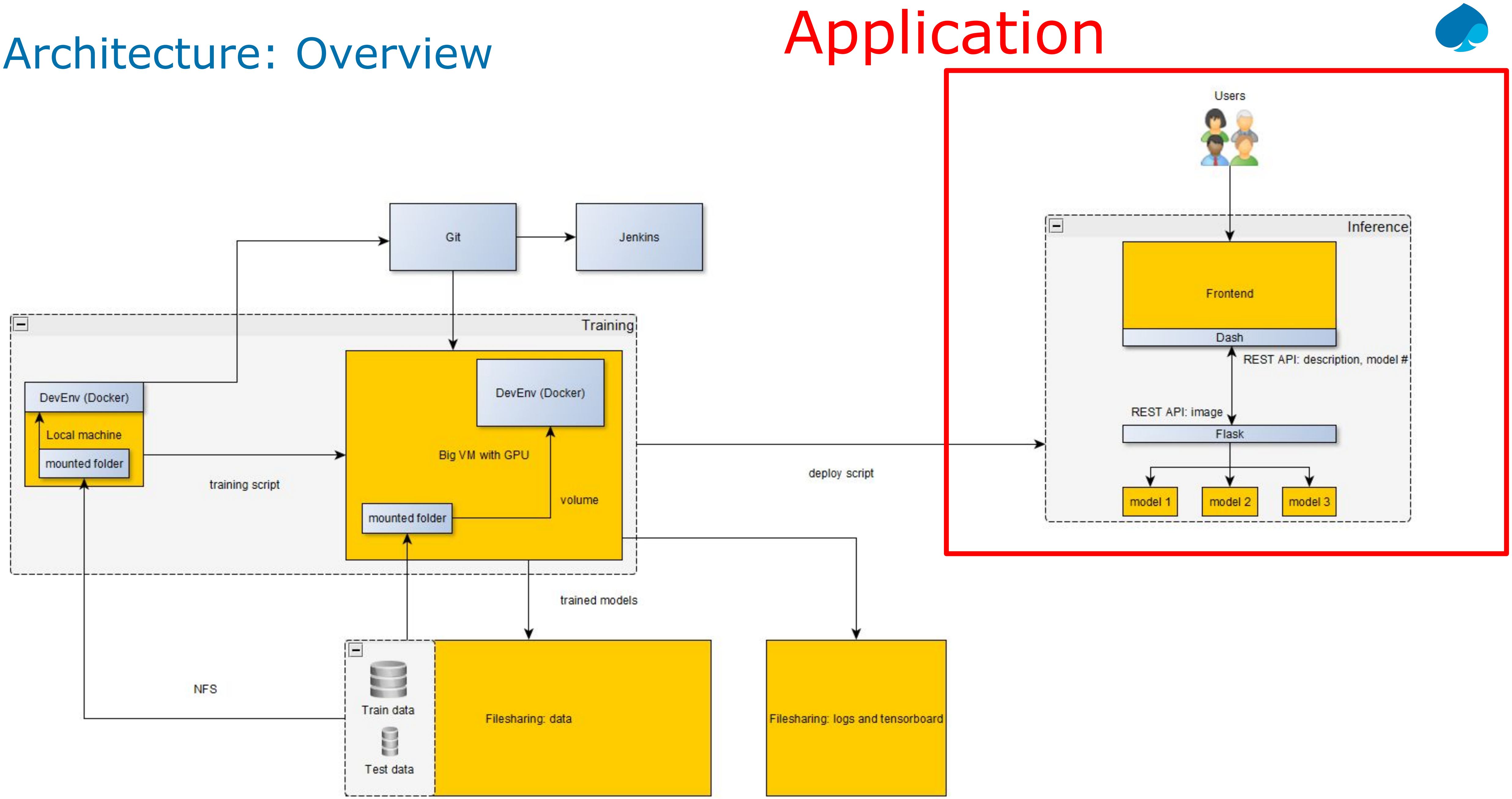
Architecture: Overview



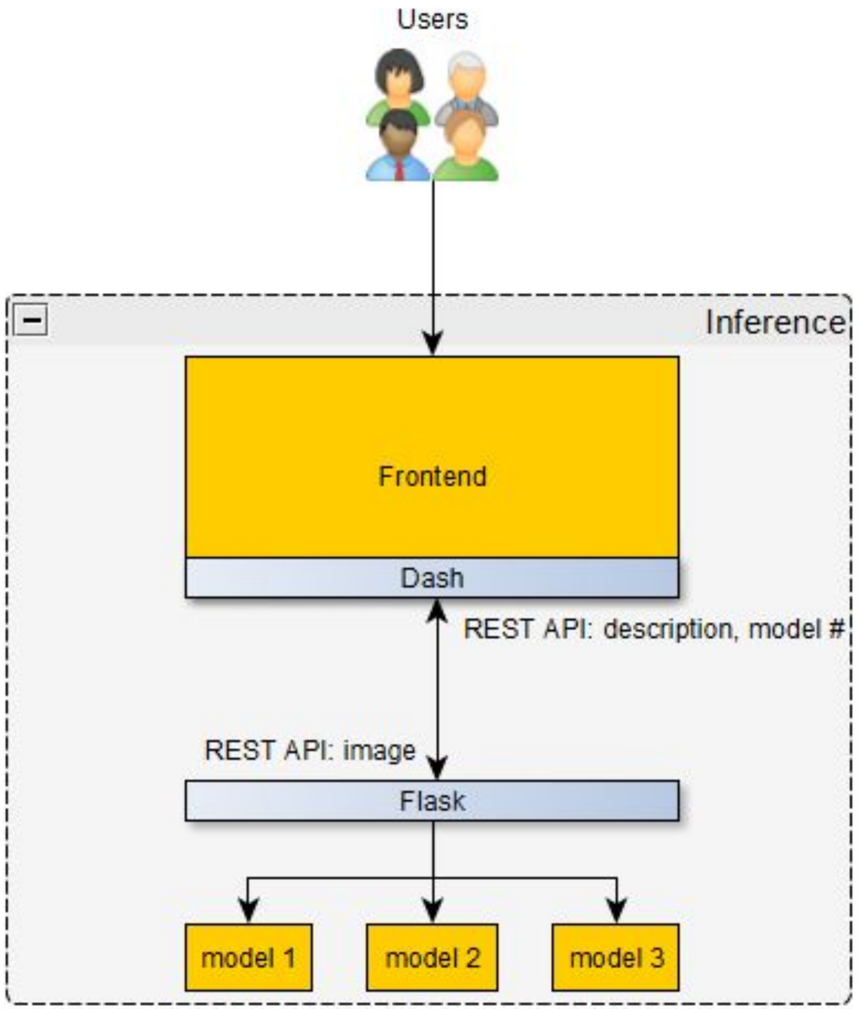
Architecture: Overview



Architecture: Overview



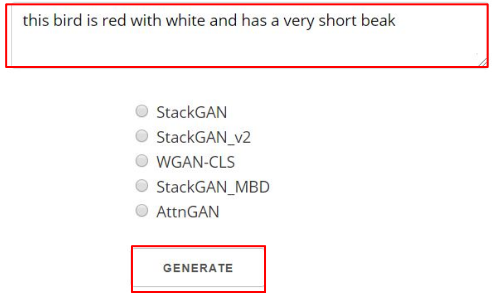
Architecture: Application

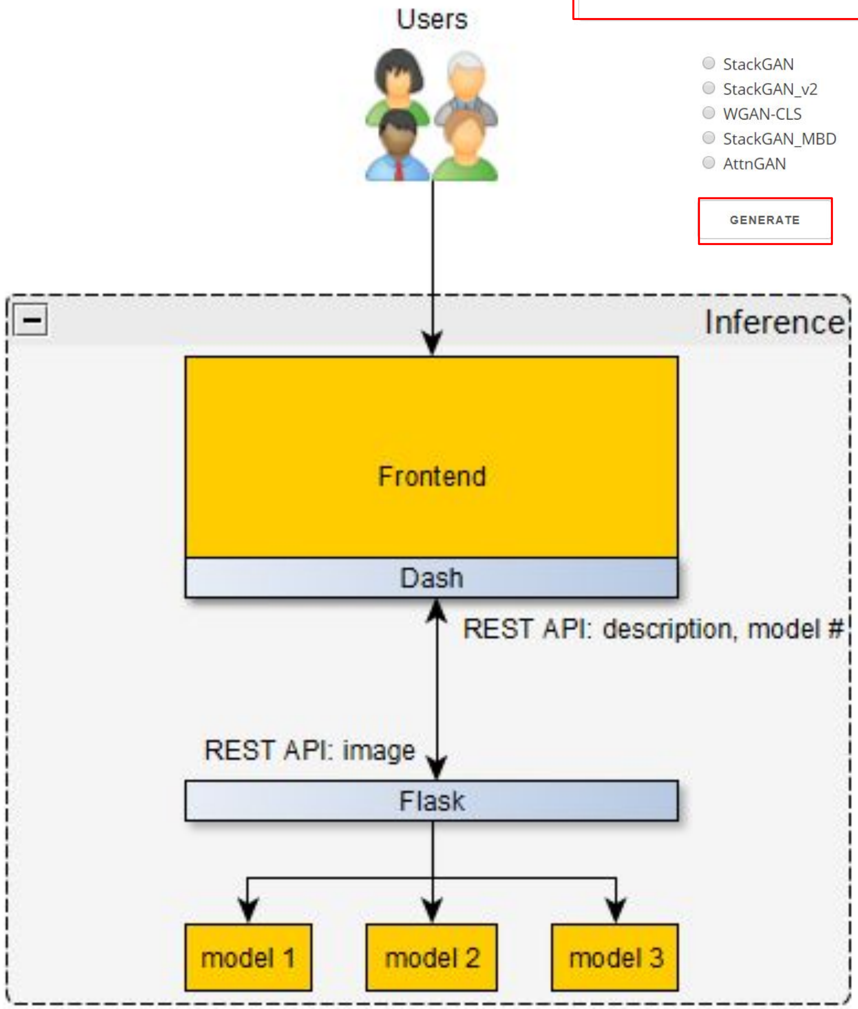


Dash

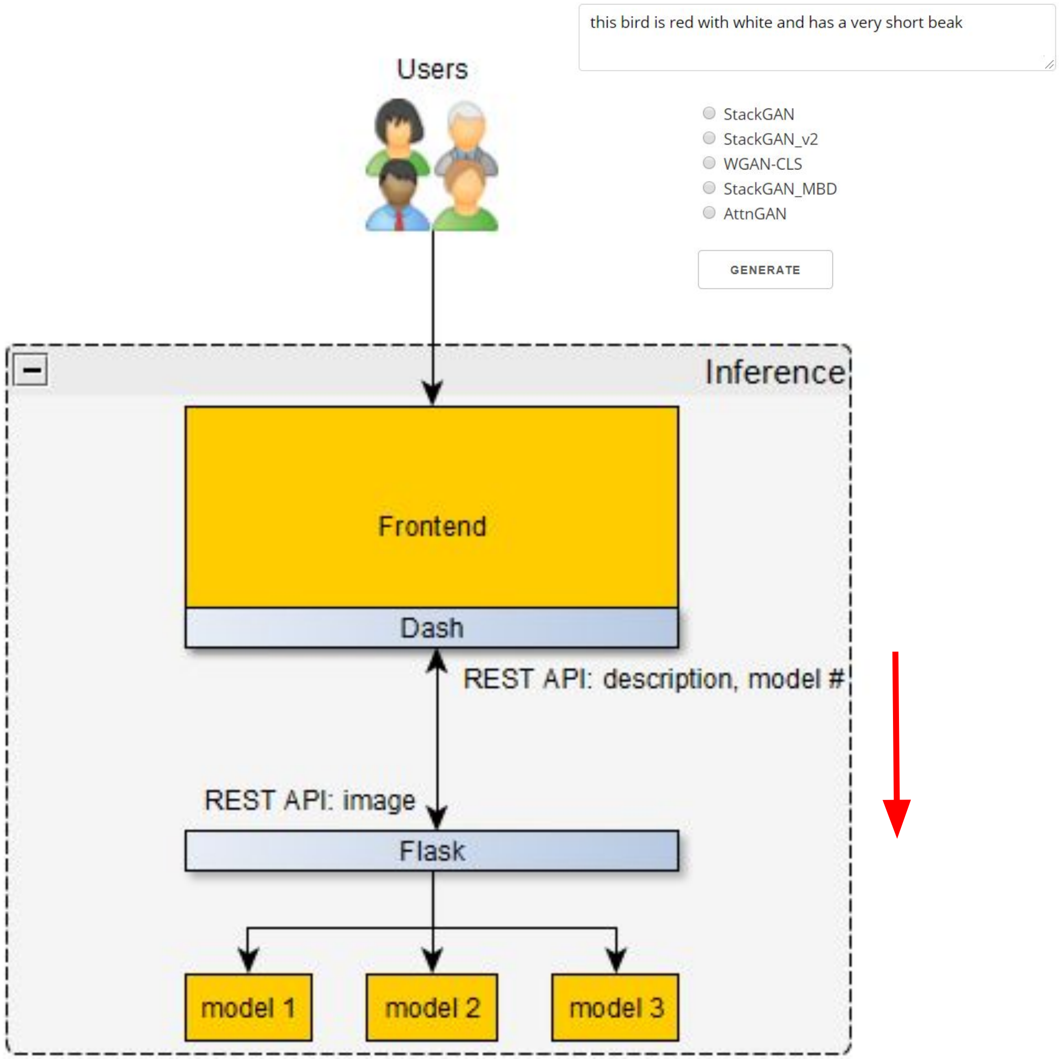
Flask
web development,
one drop at a time

Architecture: Flask



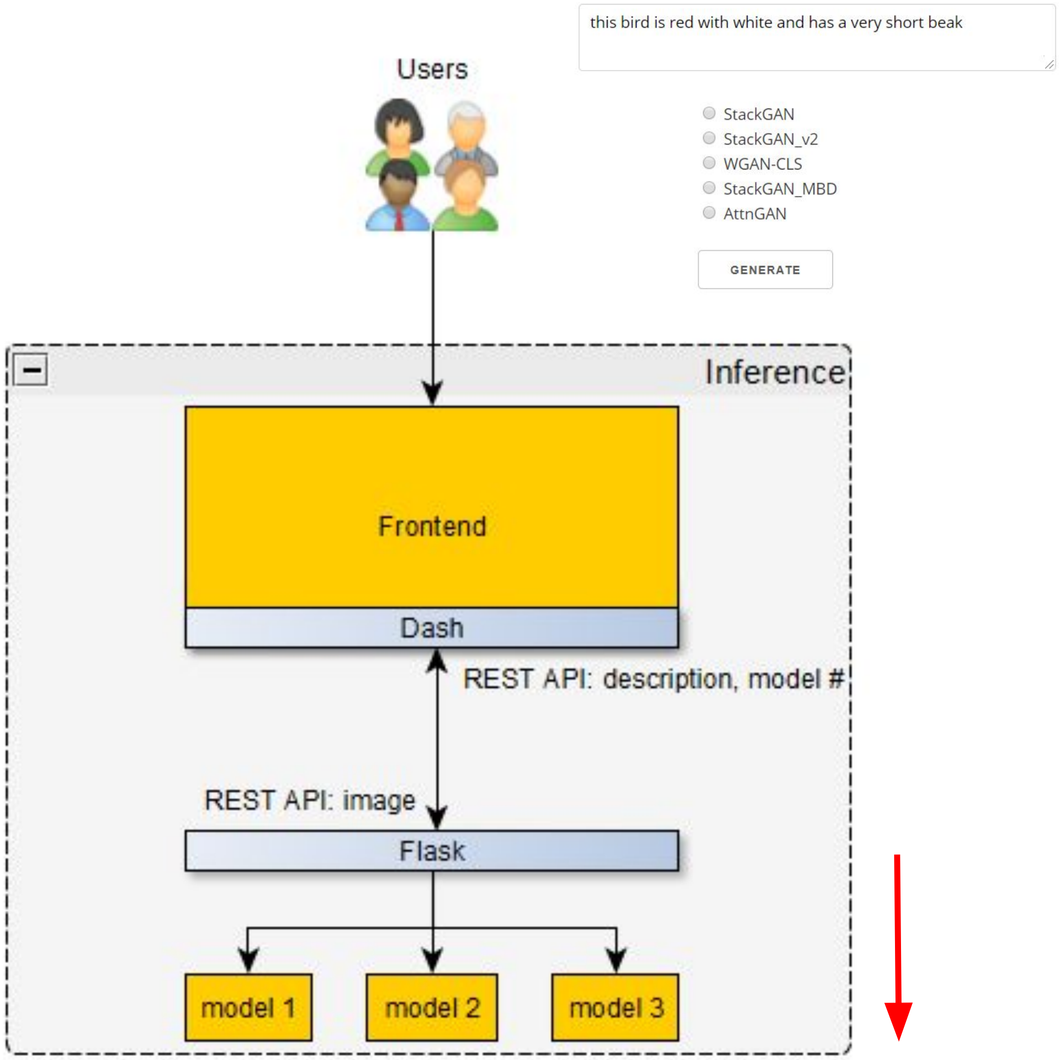


Architecture: Flask



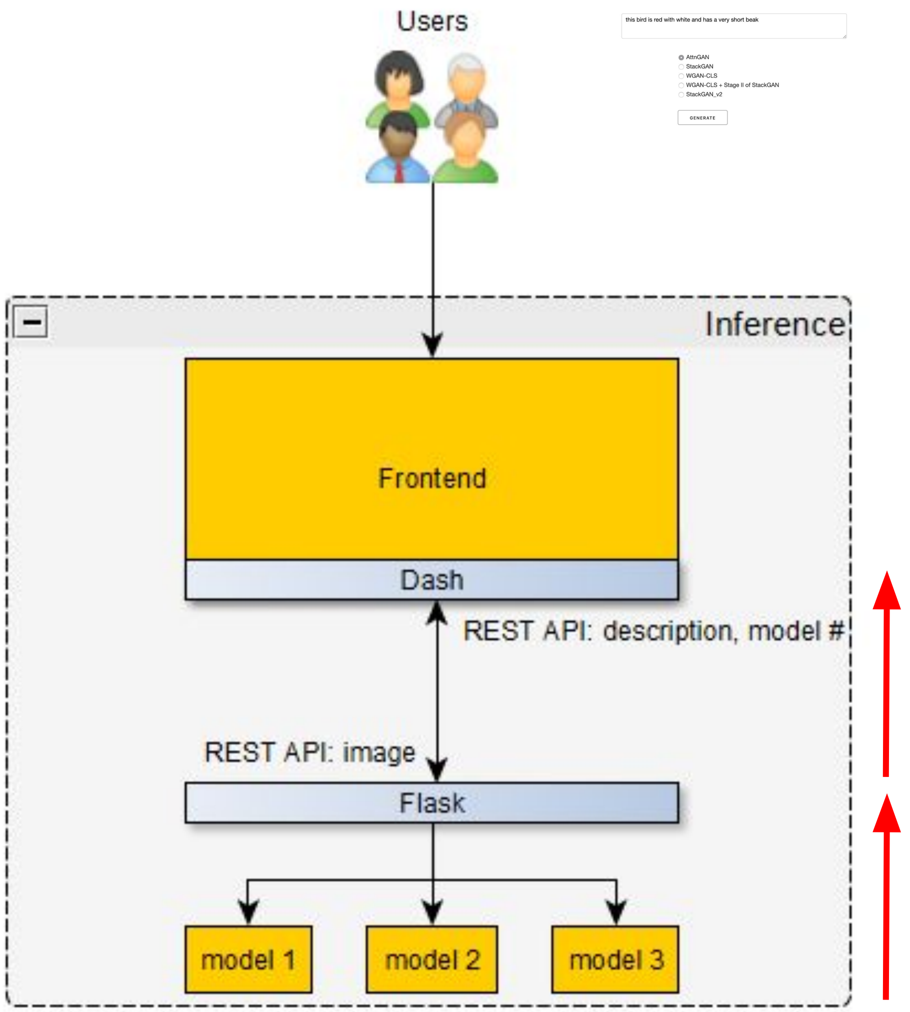
Dash: Give me the image for *%description%* produced by the *%model%*

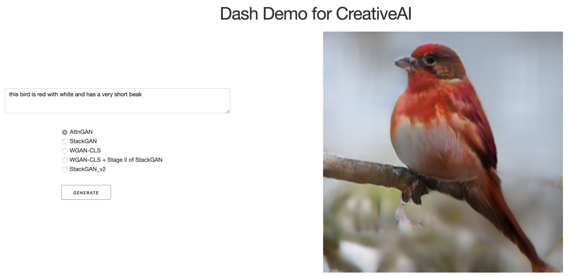
Architecture: Flask



Flask: one moment, I have %model% generate this image for you

Architecture: Flask





Flask: Done. Please take it as an array

Architecture: Docker

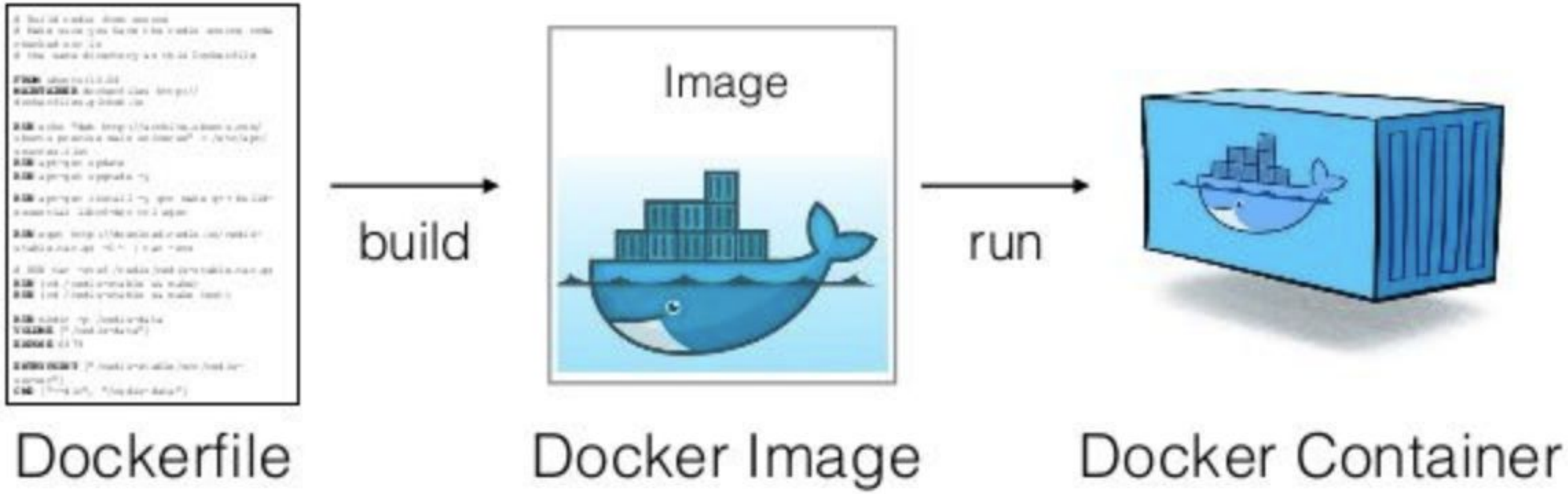
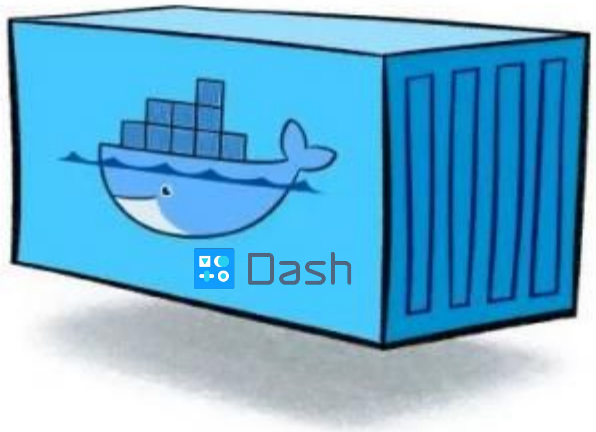
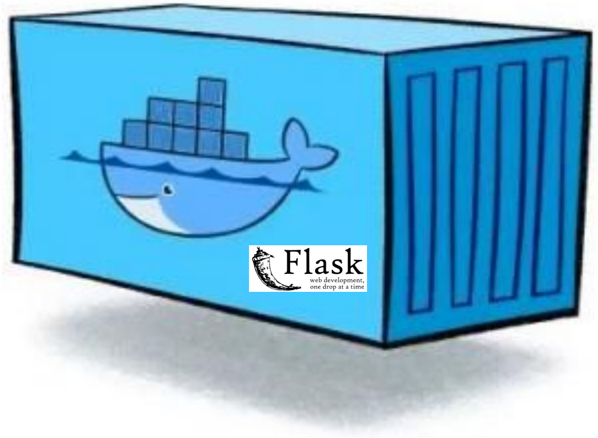


Image Source: https://medium.com/platformer-blog/practical-guide-on-writing-a-dockerfile-for-your-application-89376f88b3b5

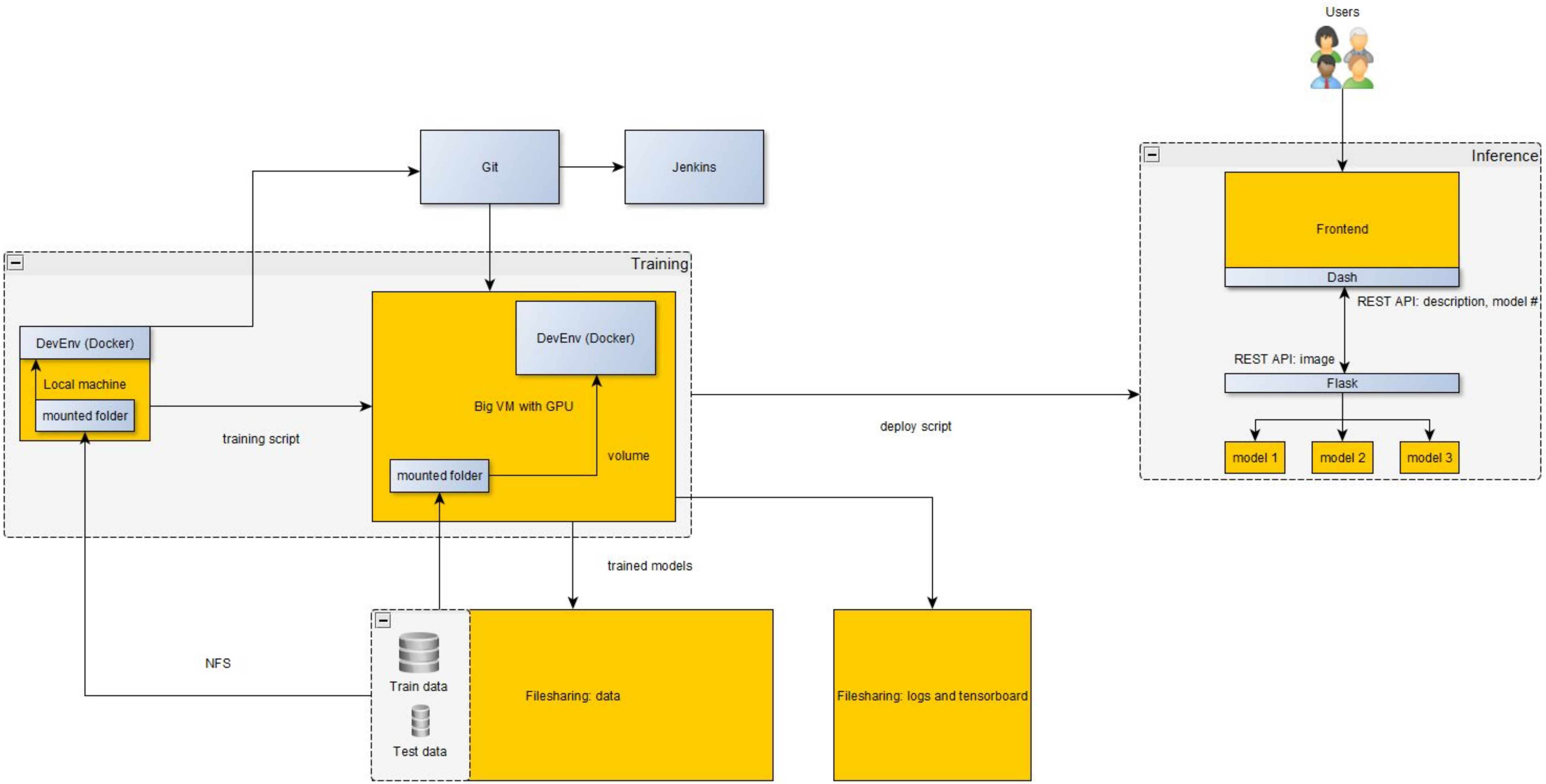
Architecture: Docker Compose



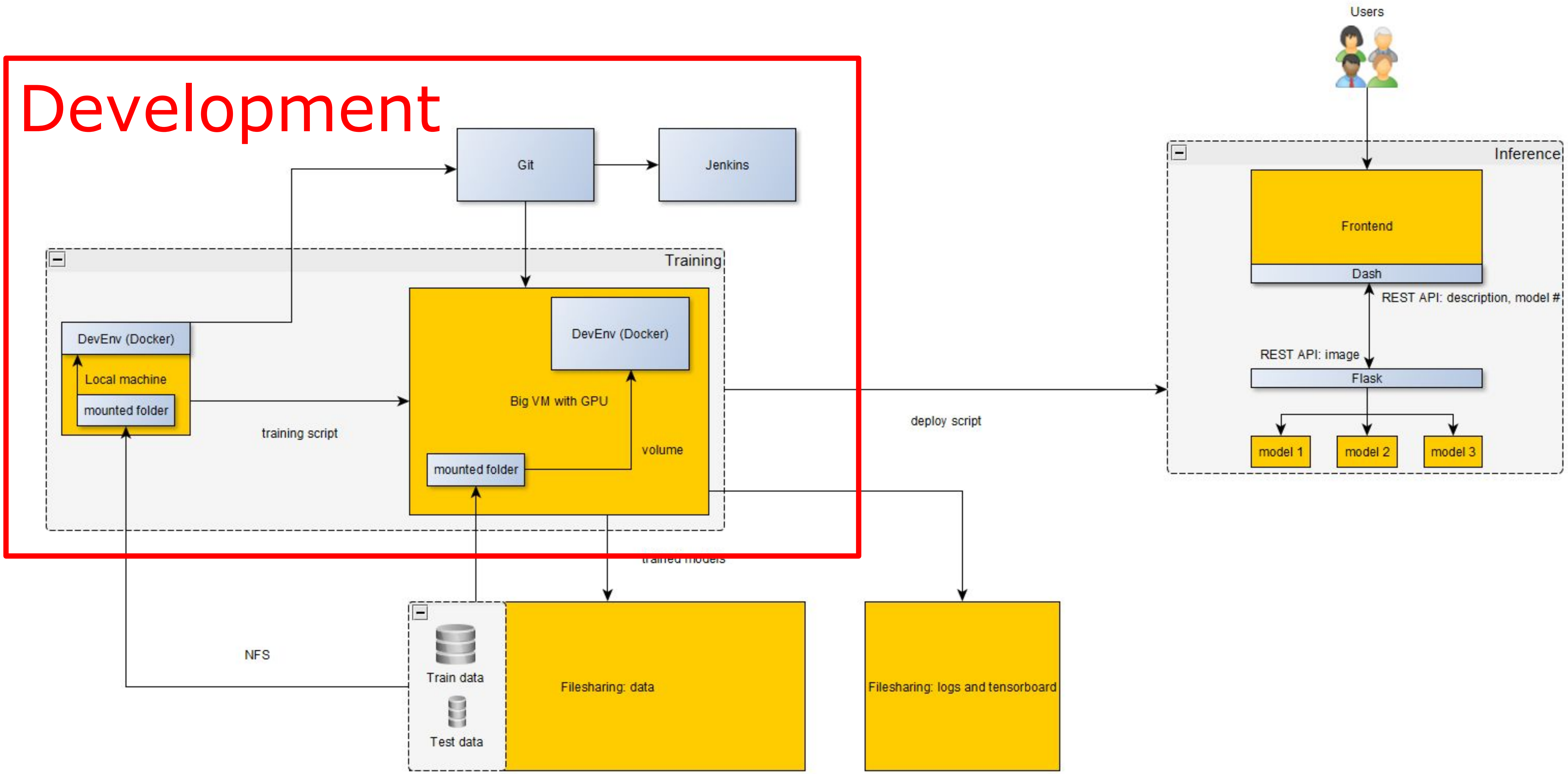
docker compose



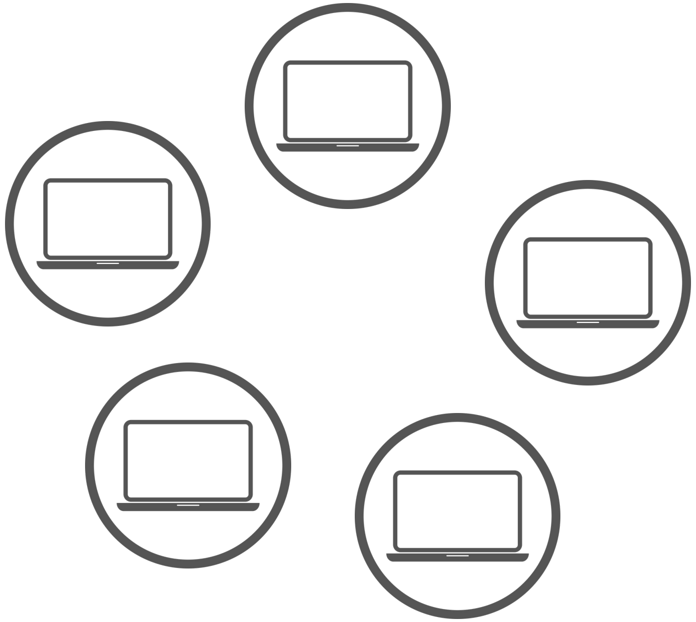
Architecture: Overview

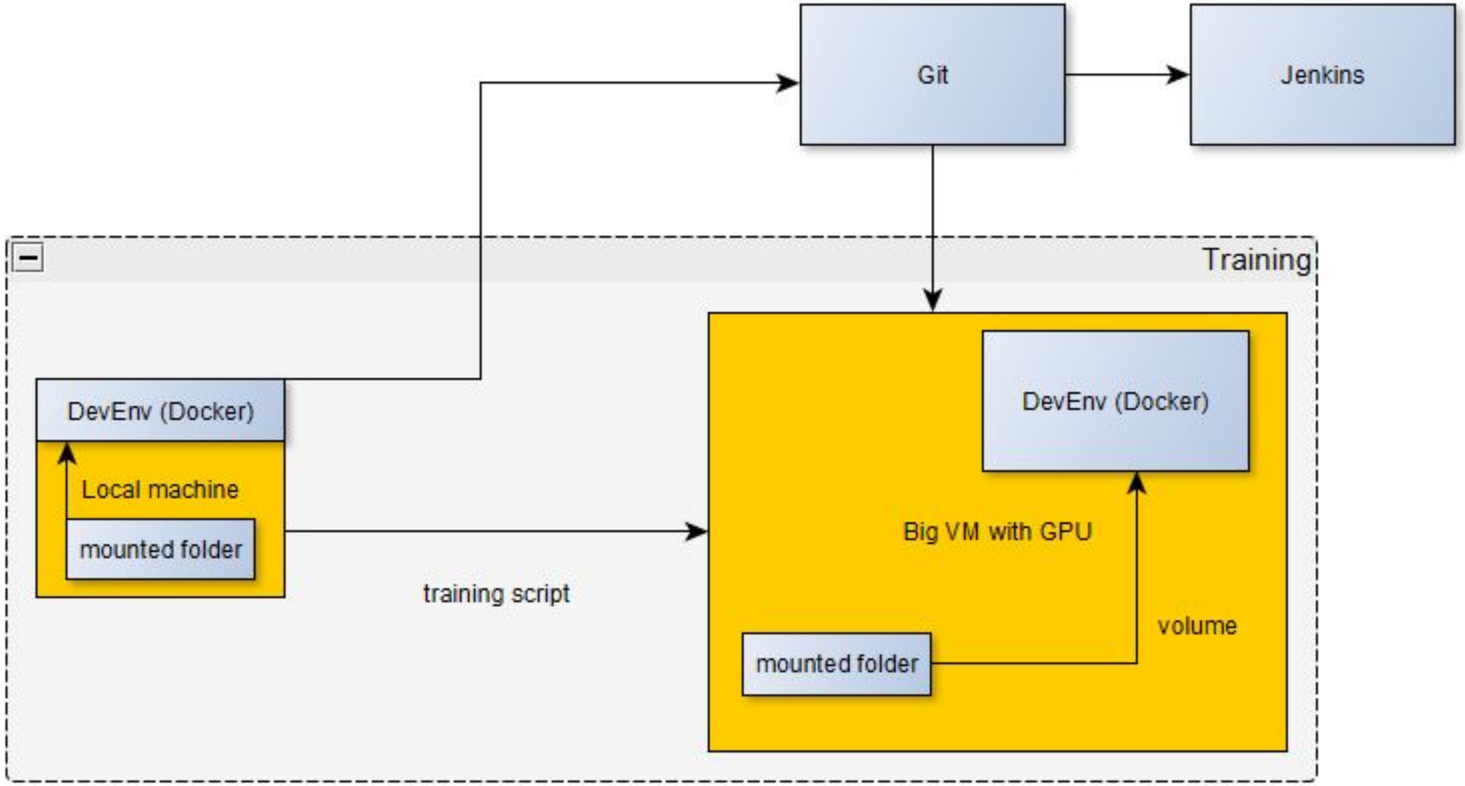


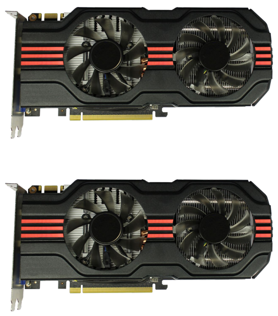
Architecture: Overview



Architecture: Development







Architecture: Git and Jenkins

Git Branches

- Master
- Feature/frontend/...
- Feature/backend/...
- Bug/…
- Fix/...

Jenkins

Automated software quality checks with four stages:

- **Build**: Code compiles?
- **Test**: Unit tests pass?
- **Lint**: Style guidelines not violated?
- **Black**: Similar formatting?

Stage View

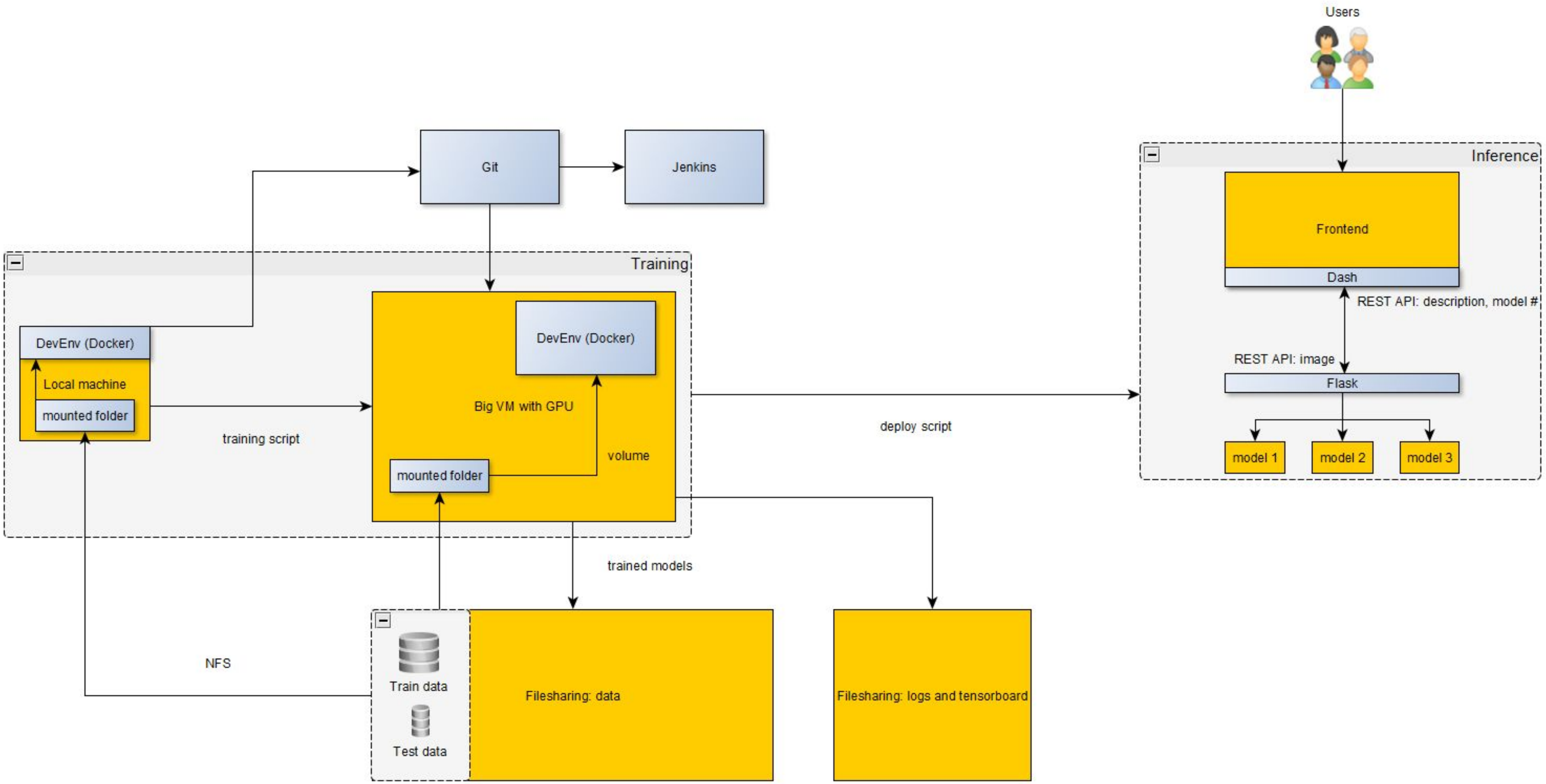
	Declarative: Checkout SCM	build	test	lint	black	Declarative: Post Actions
Average stage times: (Average full run time: ~28s)	640ms	20s	1s	3s	495ms	1s
#30 Jun 01 17:37 1 commit	581ms	18s	1s	3s	356ms	1s
#29 Jun 01 17:27 1 commit	714ms	20s	1s	3s	379ms	1s
#28 Jun 01 16:01 1 commit	626ms	20s	1s	3s	613ms	866ms
#27 Jun 01 15:46 1 commit	642ms	20s	1s	3s	634ms failed	935ms

pytest

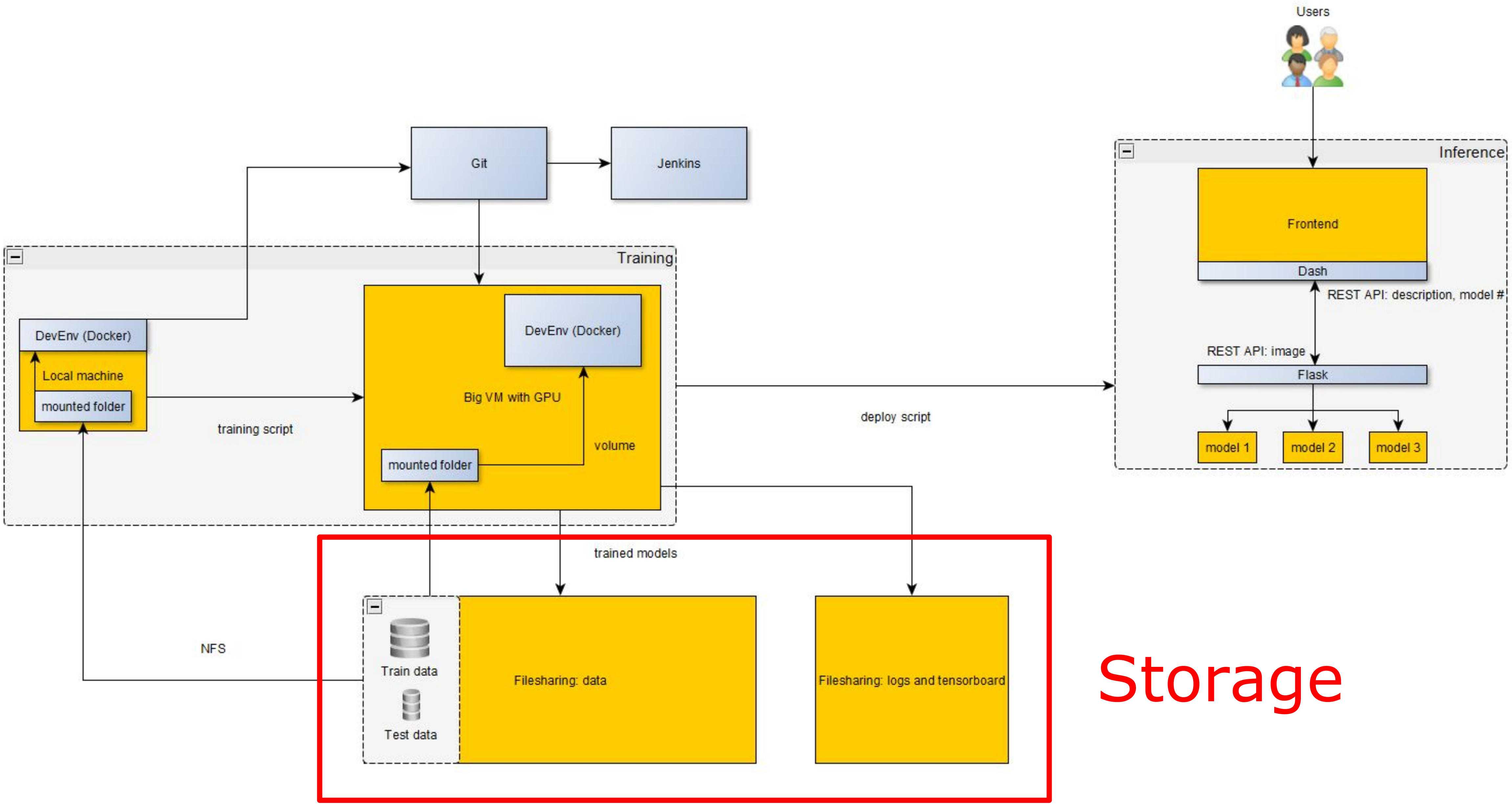
Pylint
Star your Python code!

Black

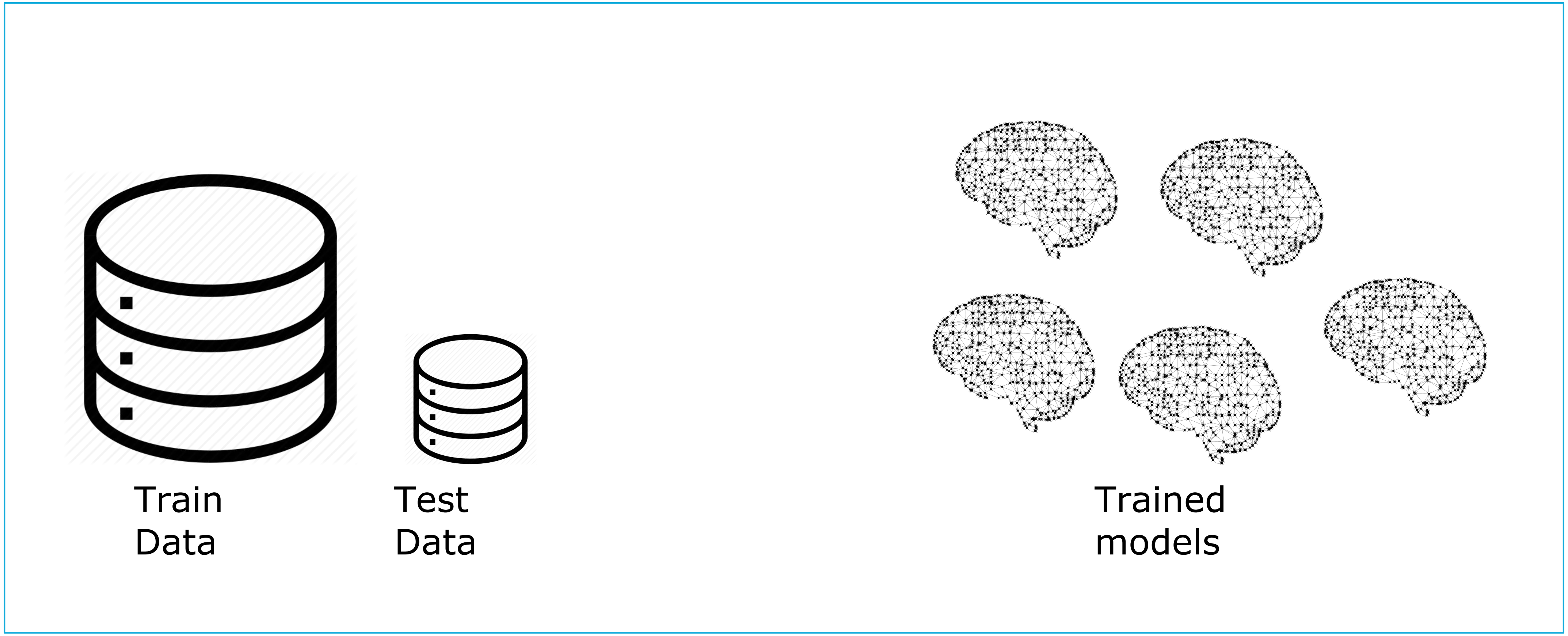
Architecture: Overview



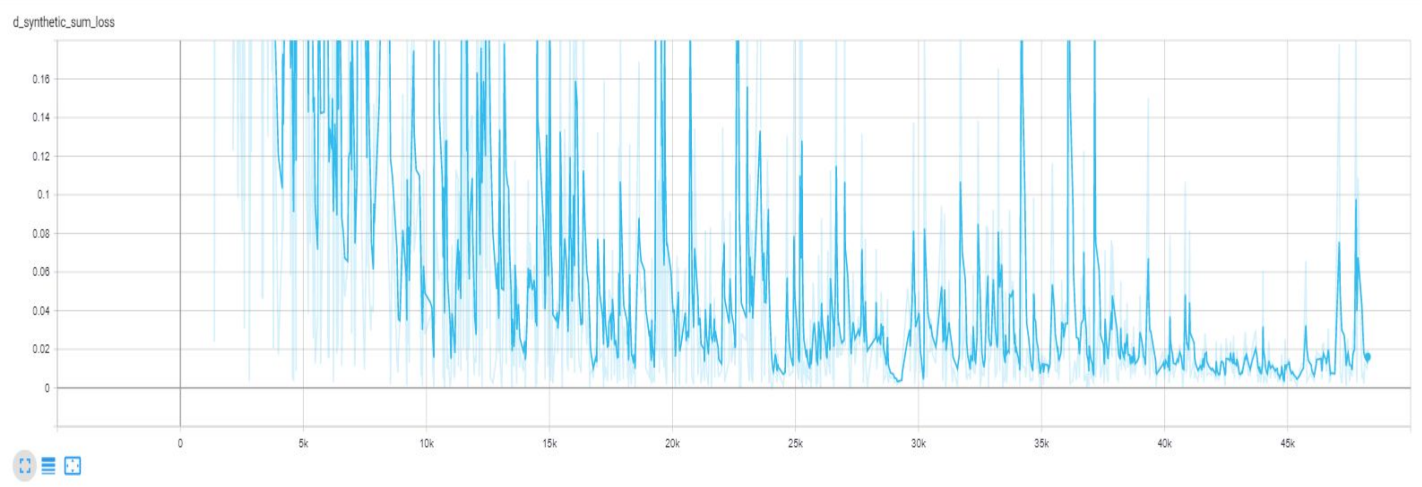
Architecture: Overview

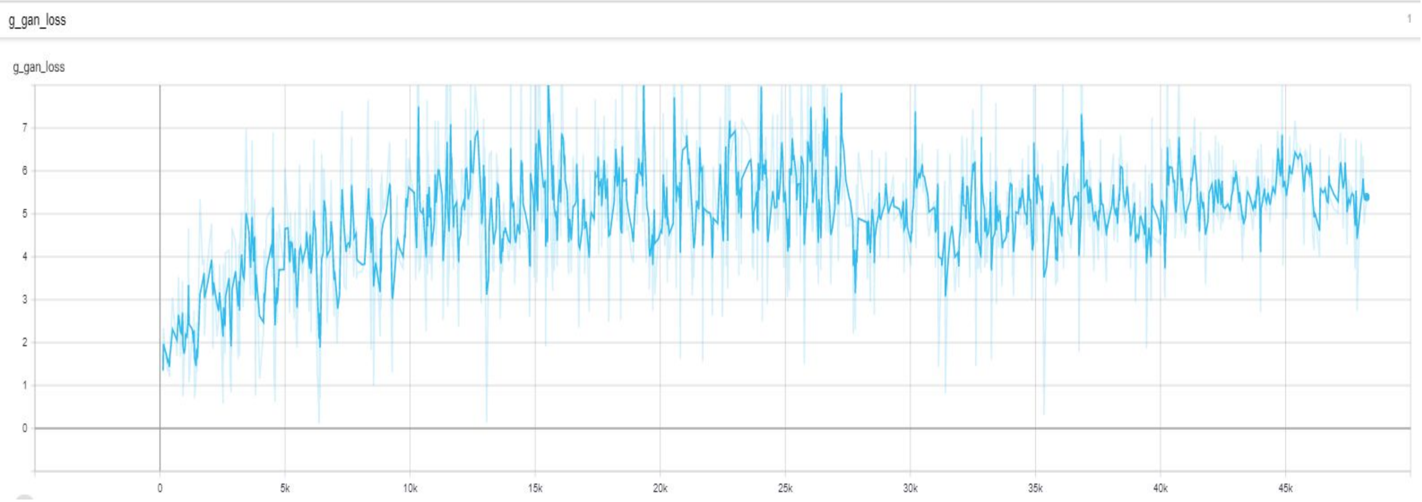


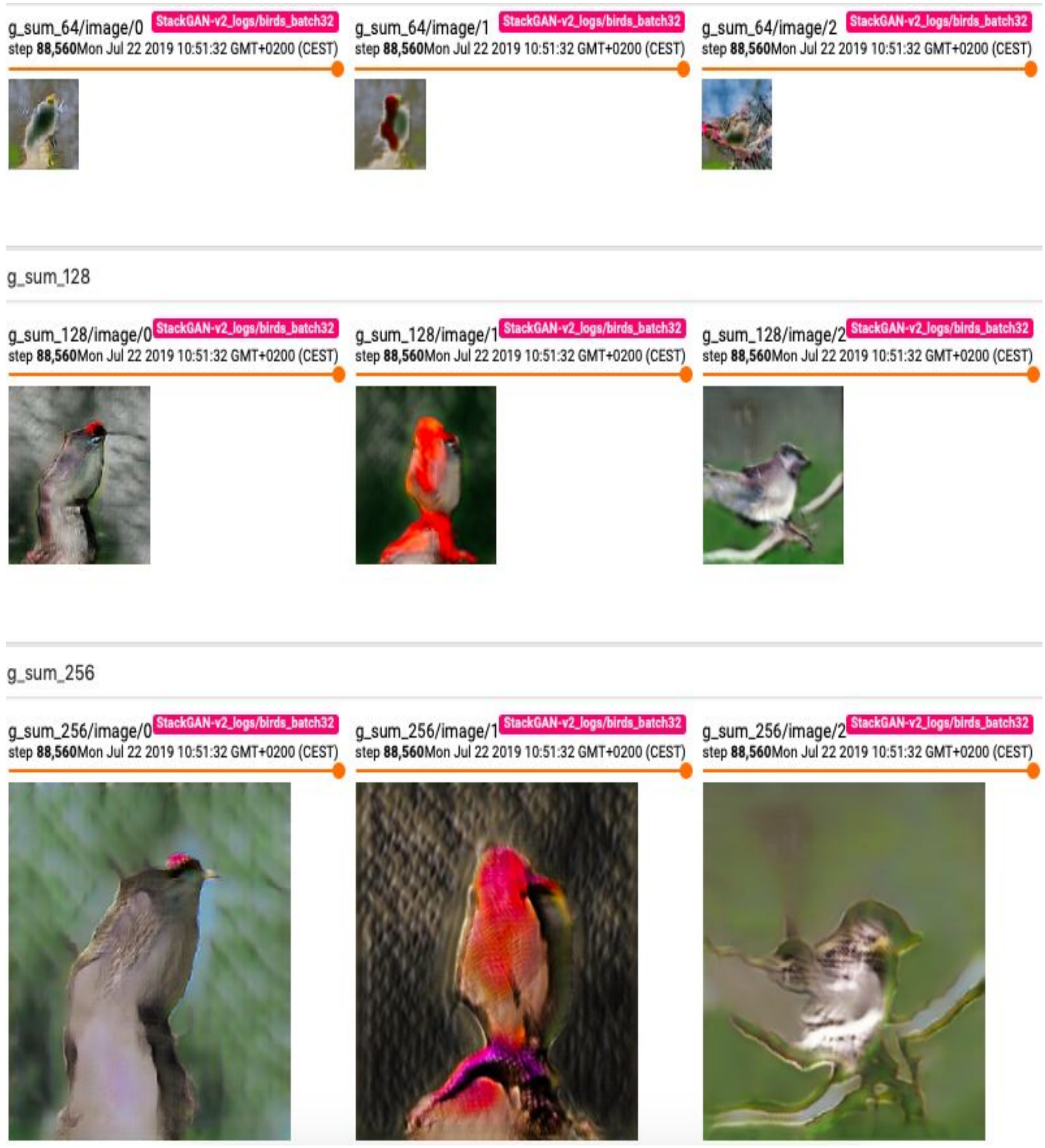
Architecture: Filesharing VM



Architecture: Tensorboard VM







Agenda

- Introduction
- GANs for Image Generation from Text
- Demo Session
- System Architecture
- **Summary and Q&A**

Summary

- Generative Adversarial Networks
- Models: StackGAN, StackGAN++, AttnGAN
- Techniques for GAN Training Stabilization
- Demo Session
- System Architecture

Sources

- Arjovsky, Martin, Soumith Chintala, and Léon Bottou. "Wasserstein generative adversarial networks." International conference on machine learning. 2017.
- Brock, Andrew, Jeff Donahue, and Karen Simonyan. "Large scale gan training for high fidelity natural image synthesis." arXiv preprint arXiv:1809.11096. 2018.
- He, Zhenliang, et al. "Attgan: Facial attribute editing by only changing what you want." IEEE Transactions on Image Processing. 2019.
- Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
- Miyato, Takeru, et al. "Spectral normalization for generative adversarial networks." arXiv preprint arXiv:1802.05957. 2018.
- Salimans, Tim, et al. "Improved techniques for training gans." Advances in neural information processing systems. 2016.
- Xu, Tao, et al. "Attngan: Fine-grained text to image generation with attentional generative adversarial networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.
- Zhang, Han, et al. "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks." Proceedings of the IEEE International Conference on Computer Vision. 2017.
- Zhang, Han, et al. "Stackgan++: Realistic image synthesis with stacked generative adversarial networks." arXiv preprint arXiv:1710.10916. 2017.

Thank you for your attention!

Q&A Time!

About Capgemini

A global leader in consulting, technology services and digital transformation, Capgemini is at the forefront of innovation to address the entire breadth of clients' opportunities in the evolving world of cloud, digital and platforms. Building on its strong 50-year heritage and deep industry-specific expertise, Capgemini enables organizations to realize their business ambitions through an array of services from strategy to operations. Capgemini is driven by the conviction that the business value of technology comes from and through people. It is a multicultural company of 200,000 team members in over 40 countries. The Group reported 2016 global revenues of EUR 12.5 billion.

Learn more about us at

www.capgemini.com

People matter, results count.

This message contains information that may be privileged or confidential and is the property of the Capgemini Group.

Copyright © 2019 Capgemini. All rights reserved.