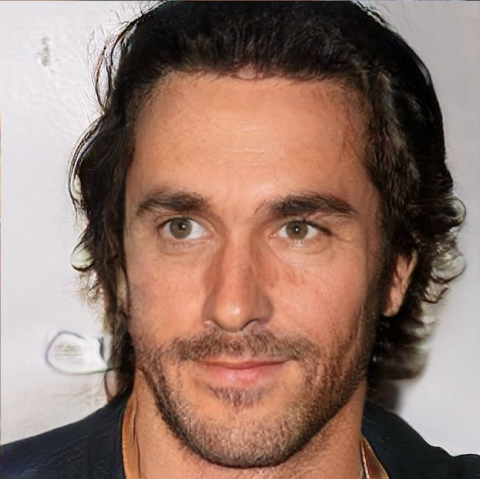
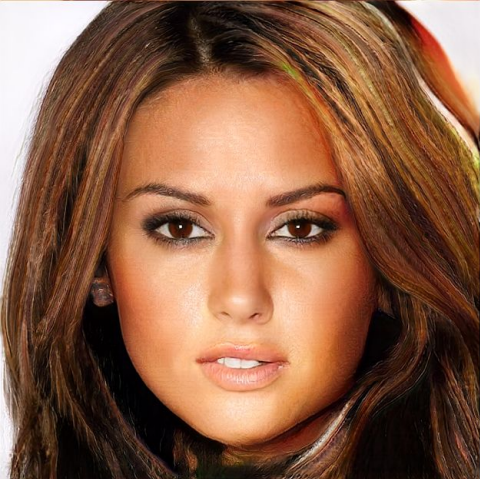


Progressive Growing of GANs

Tero Karras, Timo Aila, Samuli Laine, Jakko Lehtinen
Nvidia Research, Oct. 2017

Overview

This paper presents a new training methodology for Generative Adversarial Networks. Using this training methodology the authors were able to generate images of very high quality and resolution.



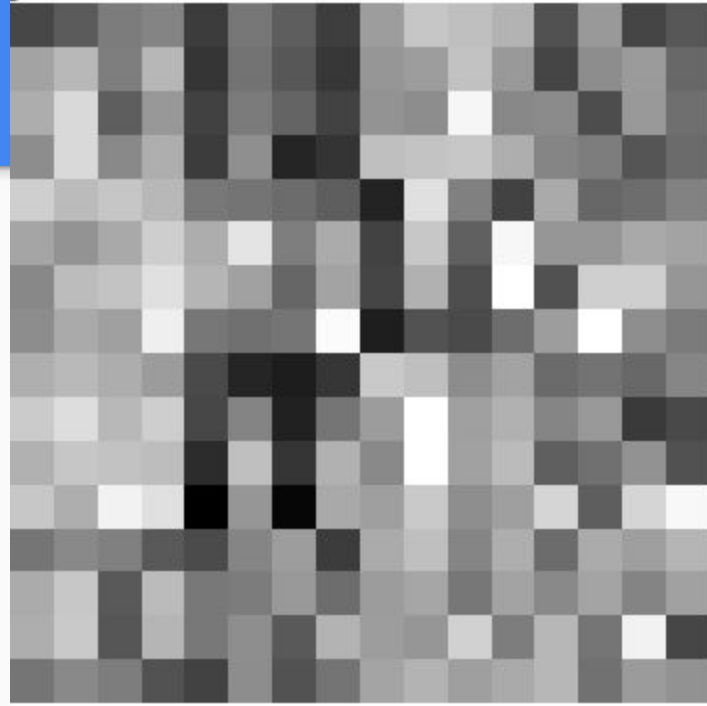
A B C D E

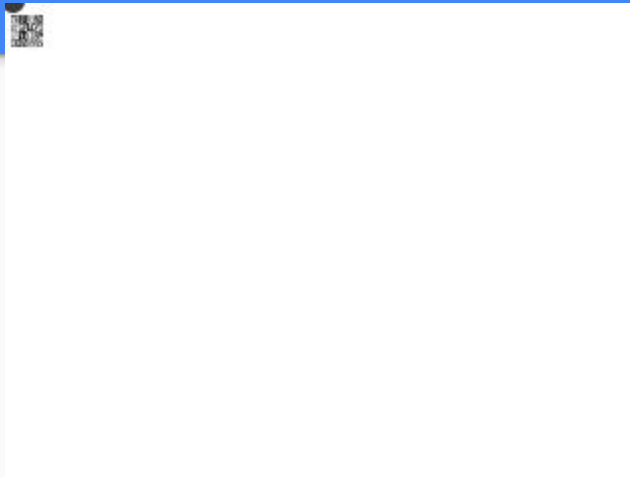
F G H I J

K L M N O

P Q R S T

U V W X Y





Talk Roadmap

- Generative Models
- Progressive Growing of GANs (PGGAN)
- Modeling Techniques
- Evaluating PGGAN
- Conclusions

Generative Models - overview

Autoregressive

Sharp images, slow to evaluate. No latent space.

VAE

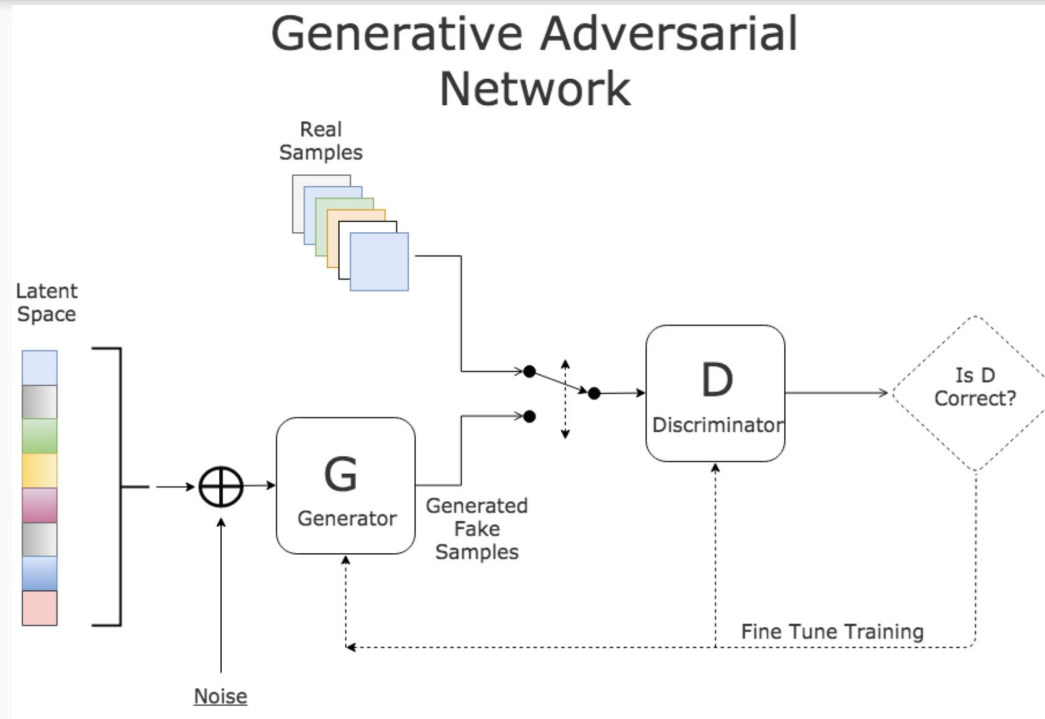
Fast to train, blurry images.

GANs

Sharp images, low resolution, limited variation, unstable training

Generative Models

Generative Models - GANs



Generative Models - GANs

Interested in learning $P_{\text{data}}(x)$ but can only sample from it so estimate it with $P_G(x)$. Learn $P_G(x)$ through adversarial training.

When G and D hit equilibrium then theoretically we should have $P_G(x) = P_{\text{data}}$.

Not that easy in practice...

Generative Models - GANs

Challenges

If not much overlap between training and generated distributions then gradients can point in random directions.

Little variation in results.

High resolution harder because easier to tell apart

High resolution requires smaller minibatches so training less stable

Progressive Growing of GANs (PGGAN)

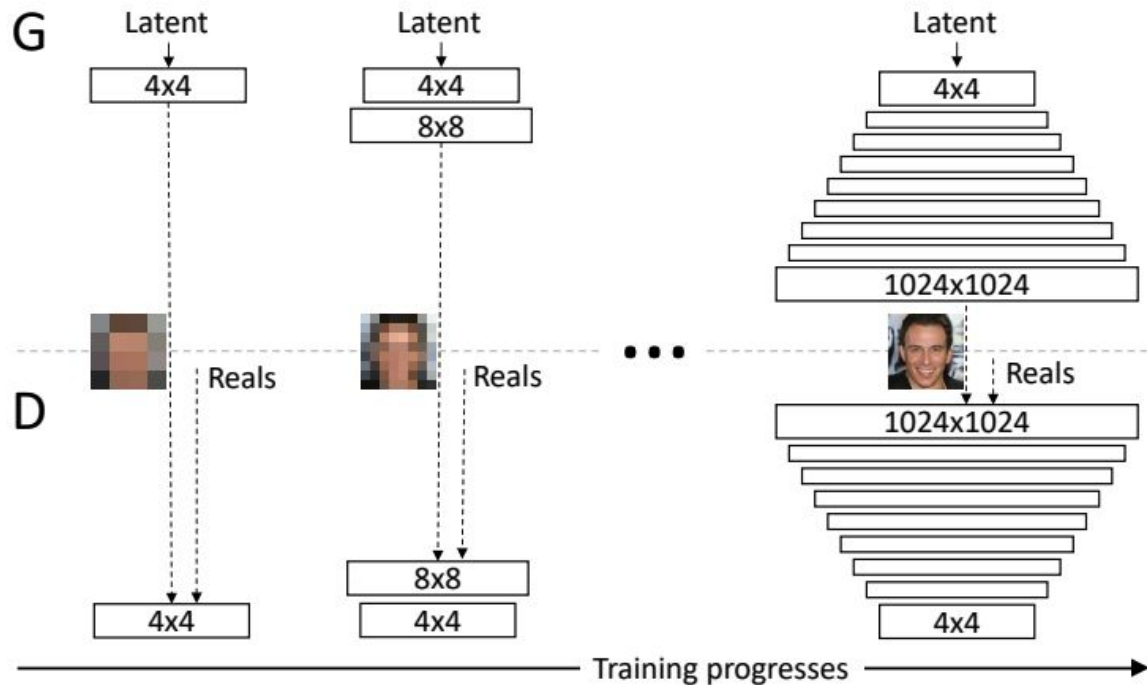
PGGAN - overview

“Our key insight is that we can grow both the generator and discriminator progressively, starting from easier low-resolution images, and add new layers that introduce higher-resolution details as the training progresses.”

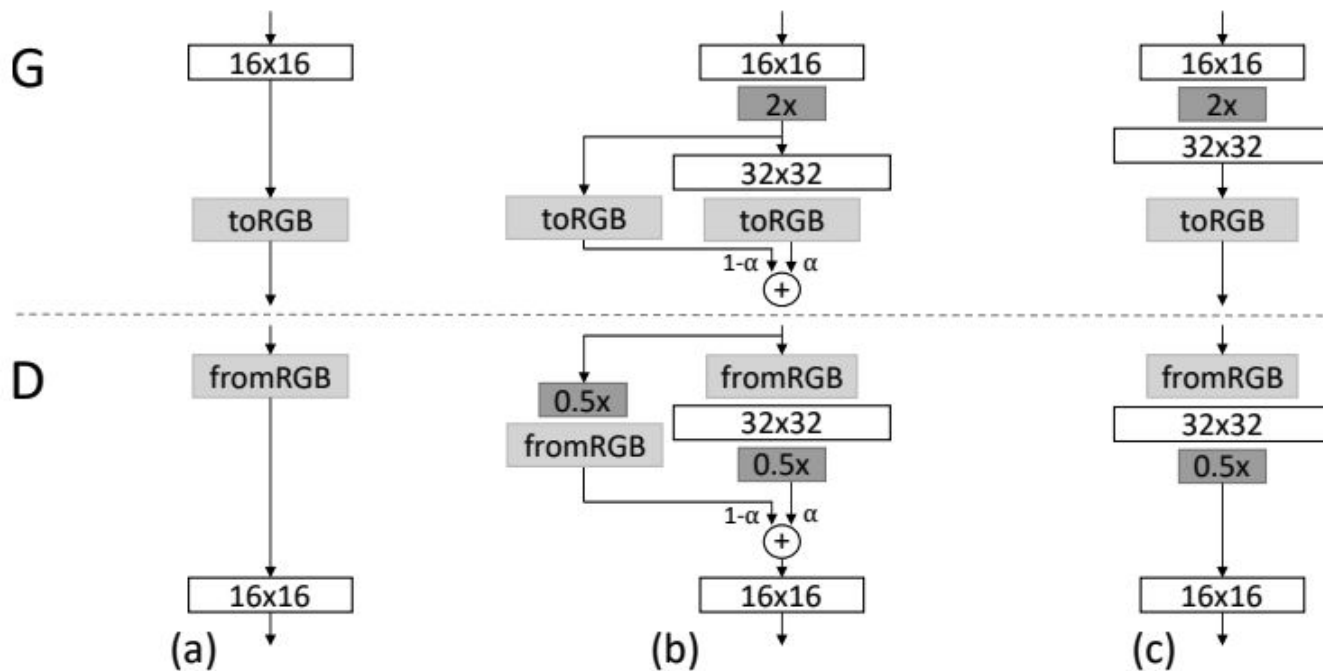
PGGAN - overview

- First discover large scale (low frequency) information and incrementally learn more fine scale (higher frequency) information.
- Generator and discriminator are mirrors and they're increased in synchrony.
- All layers are trainable throughout training
- New layers fade in smoothly

PGGAN - growing the GAN



PGGAN - fading in higher resolution layers



PGGAN - training example



Loss functions

- The author's say their work is independent of loss function
- Do experiments with both WGAN-GP and LSGAN

PGGAN - benefits

- Training avoids high resolution problem of too much divergence early on.
- Faster training, 2 - 6 x faster.
- Only use a single GAN instead of a hierarchy of GANs
- More stable training - more steps done at lower resolution with larger minibatches

Modeling techniques

Minibatch standard deviation

- Goal: encouraging the minibatches of generated and training images to show similar statistics
- Minibatch standard deviation simplifies this approach and improves variation.

Minibatch standard deviation

- Compute standard deviation for each feature in each spatial location
- Then average over all features and spatial locations to get a single value
- Replicate the value and concatenate it to all spatial locations and over the minibatch, yields one additional feature map

Normalization in Generator and Discriminator

- GANs are prone to increasing gradients due to mismatch between generator and discriminator
- The real issue is constraining signal magnitudes and competition
- They introduce “Equalized learning rate” and “Pixelwise Feature Vector Normalization in Generator”

Equalized learning rate

- With adaptive optimizers, dynamic range of parameters determines learning speed.
- Parameters with larger dynamic range can take longer to adjust and thus can have either too large or too small learning rate
- Use trivial $N(0,1)$ weight initialization and scale weights at runtime
- Scale at runtime to take advantage of scale invariance of adaptive optimization algos.

Pixelwise Feature Vector Normalization

- Prevent feature map magnitudes from getting too large
- Normalize the feature vector in each pixel to unit length in the generator after each convolutional layer

Evaluating PGGAN

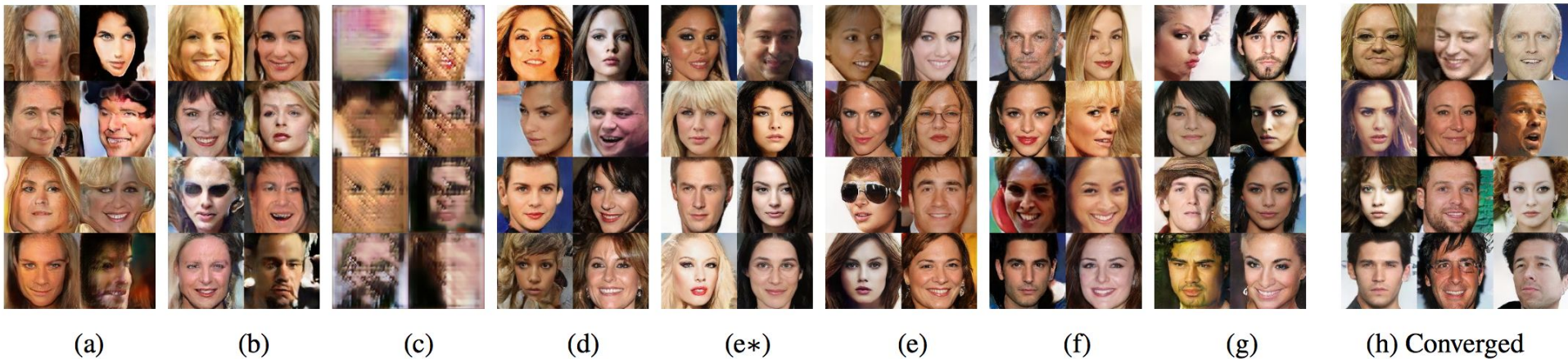
Evaluation metric

- MS-SSIM: Good at identifying global mode collapse not good for local mode collapse like on colors and textures
- Do MS-SSIM on local patches drawn from laplacian pyramid representations of generated and target images
- Sample 16,384 images and extract 128 descriptors from each level of the LP. Each descriptor is a 7x7 pixel neighbourhood with 3 color channels.
- Compute sliced wasserstein distance between samples. Smaller distance means that at that level of resolution training images and generator samples have similar variation.

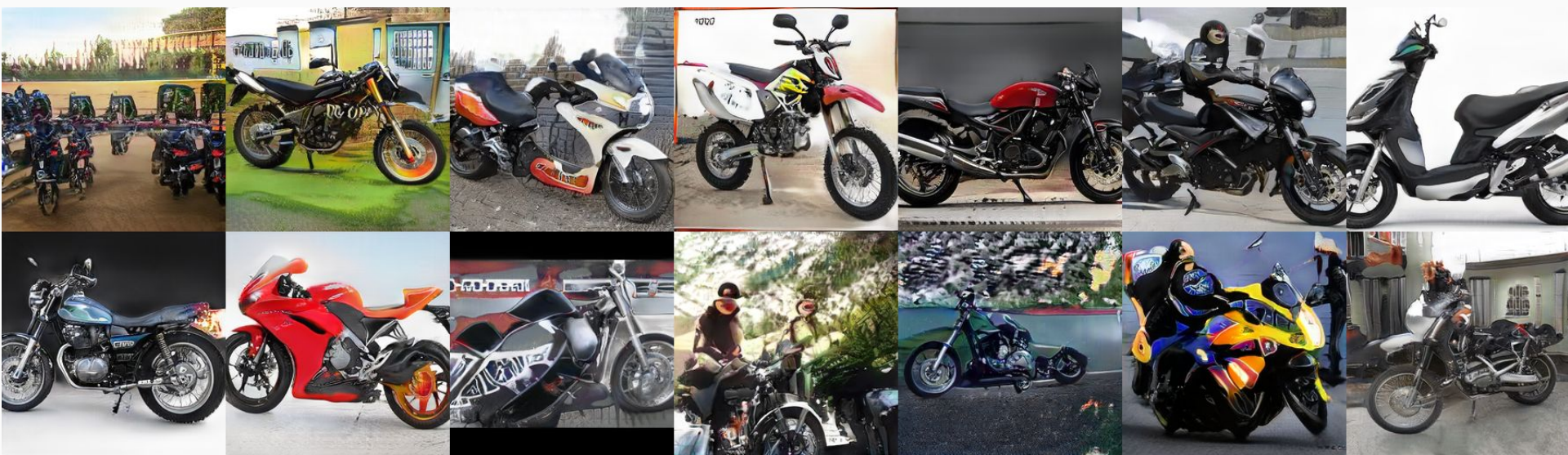
Ablation study - metrics

Training configuration	CELEBA						LSUN BEDROOM					
	Sliced Wasserstein distance $\times 10^3$					MS-SSIM	Sliced Wasserstein distance $\times 10^3$					MS-SSIM
	128	64	32	16	Avg		128	64	32	16	Avg	
(a) Gulrajani et al. (2017)	12.99	7.79	7.62	8.73	9.28	0.2854	11.97	10.51	8.03	14.48	11.25	0.0587
(b) + Progressive growing	4.62	2.64	3.78	6.06	4.28	0.2838	7.09	6.27	7.40	9.64	7.60	0.0615
(c) + Small minibatch	75.42	41.33	41.62	26.57	46.23	0.4065	72.73	40.16	42.75	42.46	49.52	0.1061
(d) + Revised training parameters	9.20	6.53	4.71	11.84	8.07	0.3027	7.39	5.51	3.65	9.63	6.54	0.0662
(e*) + Minibatch discrimination	10.76	6.28	6.04	16.29	9.84	0.3057	10.29	6.22	5.32	11.88	8.43	0.0648
(e) Minibatch stddev	13.94	5.67	2.82	5.71	7.04	0.2950	7.77	5.23	3.27	9.64	6.48	0.0671
(f) + Equalized learning rate	4.42	3.28	2.32	7.52	4.39	0.2902	3.61	3.32	2.71	6.44	4.02	0.0668
(g) + Pixelwise normalization	4.06	3.04	2.02	5.13	3.56	0.2845	3.89	3.05	3.24	5.87	4.01	0.0640
(h) Converged	2.42	2.17	2.24	4.99	2.96	0.2828	3.47	2.60	2.30	4.87	3.31	0.0636

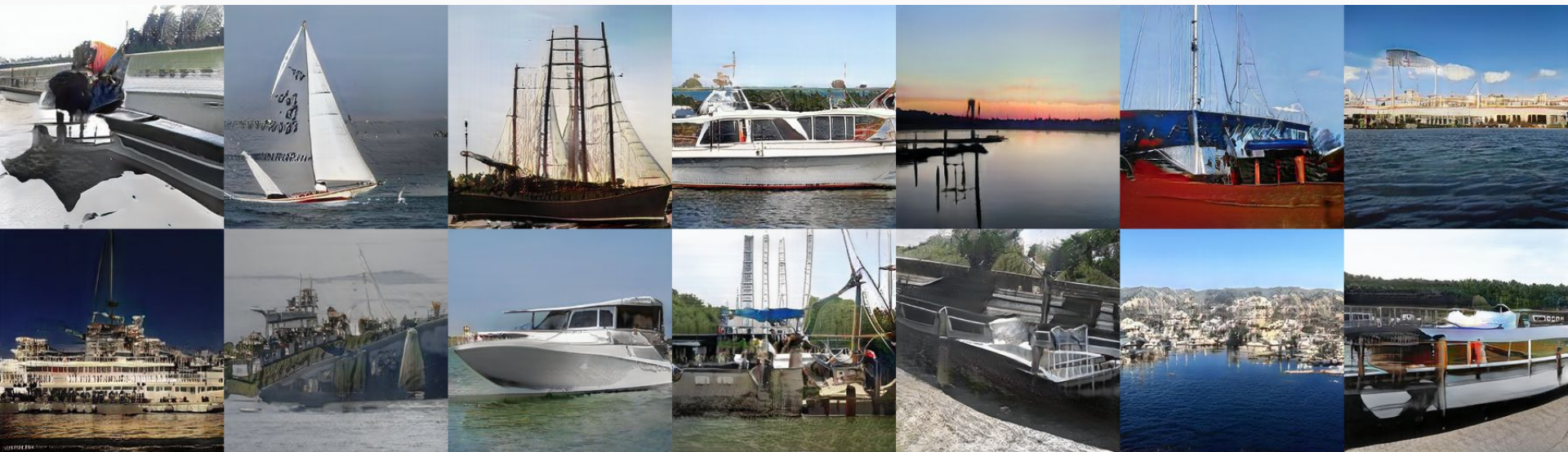
Ablation study - images



Results



Results



Conclusion

Progressive Growing of GANs

- Introduces a new training procedure that gradually increases the level of resolution as training progresses.
- Use minibatch standard deviation and two normalization techniques to increase variation in the results and stabilize training.
- Propose Sliced Wasserstein Distance as a more consistent alternative to MS-SSIM.
- Observe results that are at least as good as any in the literature for unconditional generation.

Points for discussion - criticizing paper

- What can we know (or speculate) about the impact of the progressive structure on modal collapse?
- A lot of structure is assumed. What can we know (or speculate) about the role of the pseudo-residual blocks and the training phase cycle structure of PGGAN
- A lot of structure is assumed. What can we know (or speculate) about the role of the pseudo-residual blocks and the training phase cycle structure of ProgressiveGAN

Thank you