

TricycleGAN

A Temporally Consistent CycleGAN for Unpaired Video-to-Video Translation

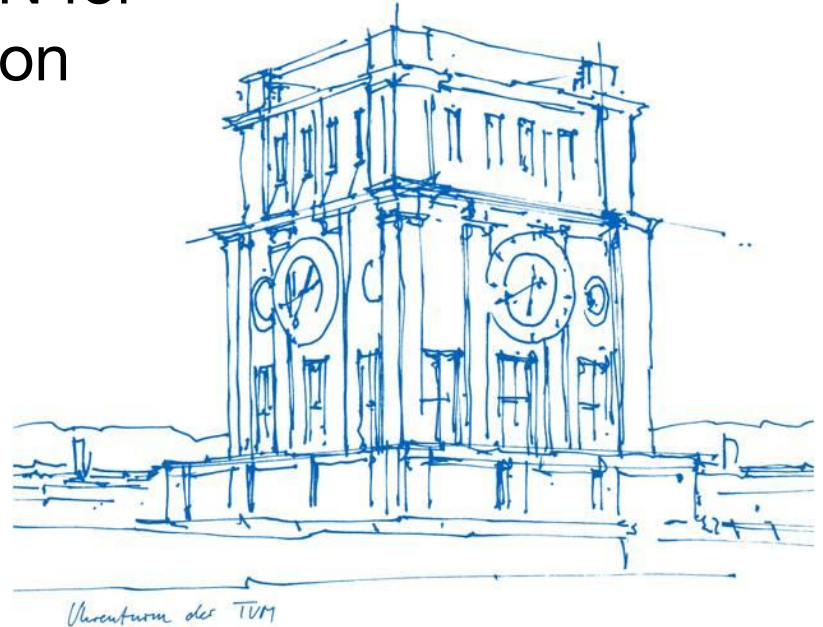
Master's Thesis Presentation

Author: Jonas Mayer

Supervisor: Prof. Dr. Nils Thuerey

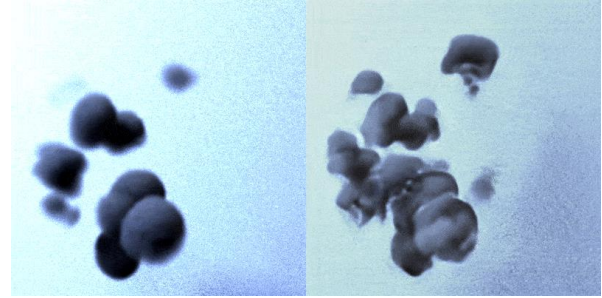
Advisor: Mengyu Chu

Garching, 28th August 2019



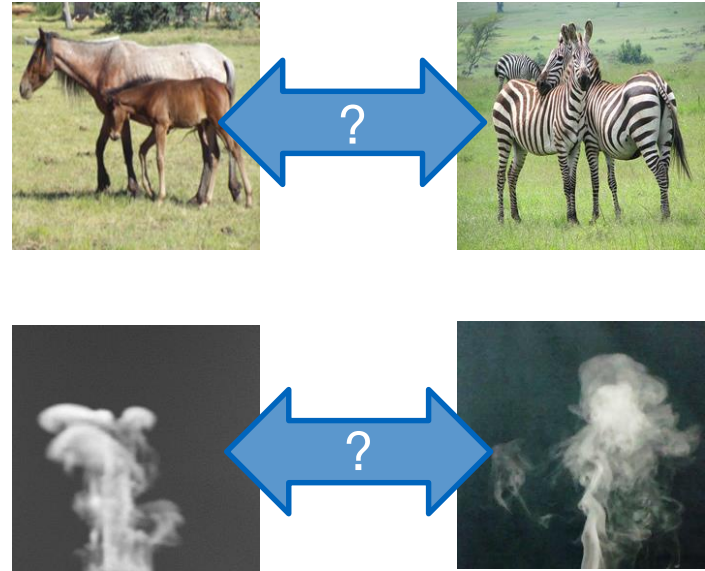
Motivation: Video-to-Video Translation

- Style Transfer
- Object-to-Object Translation
- Face-to-Face Translation
- Etc...

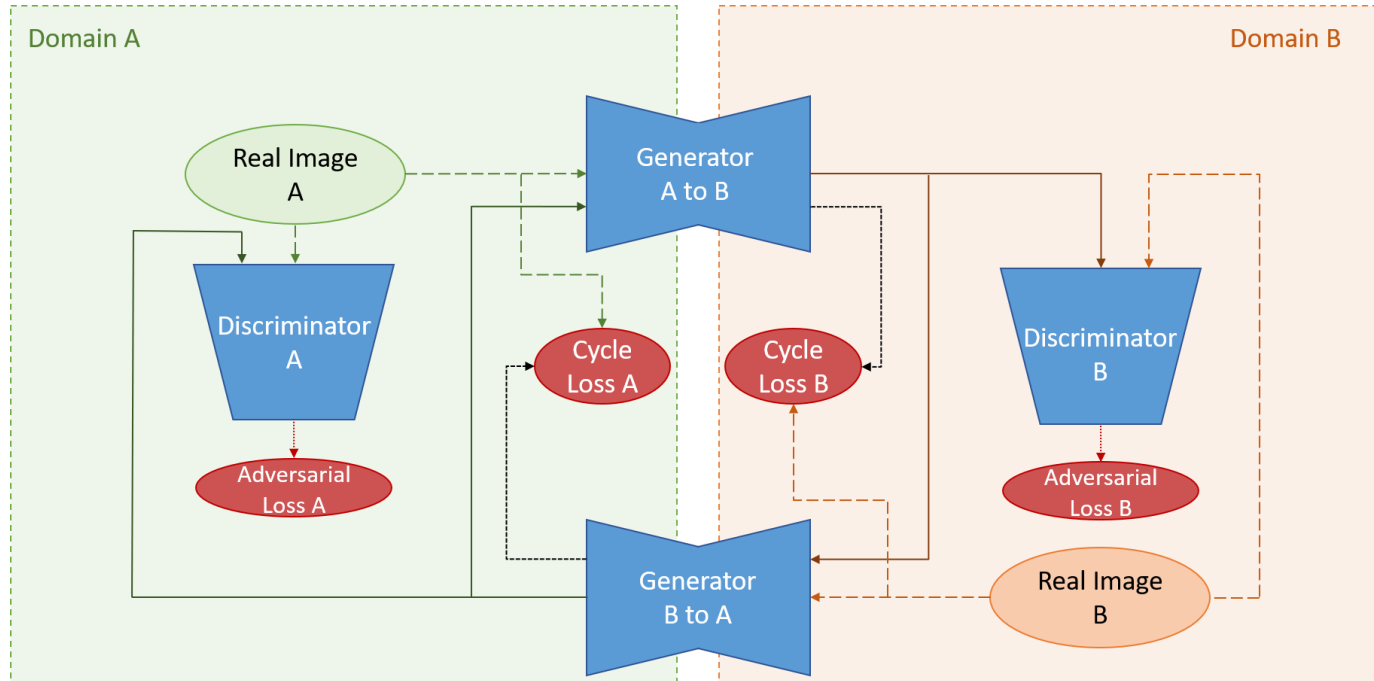


Motivation: Unpaired Data

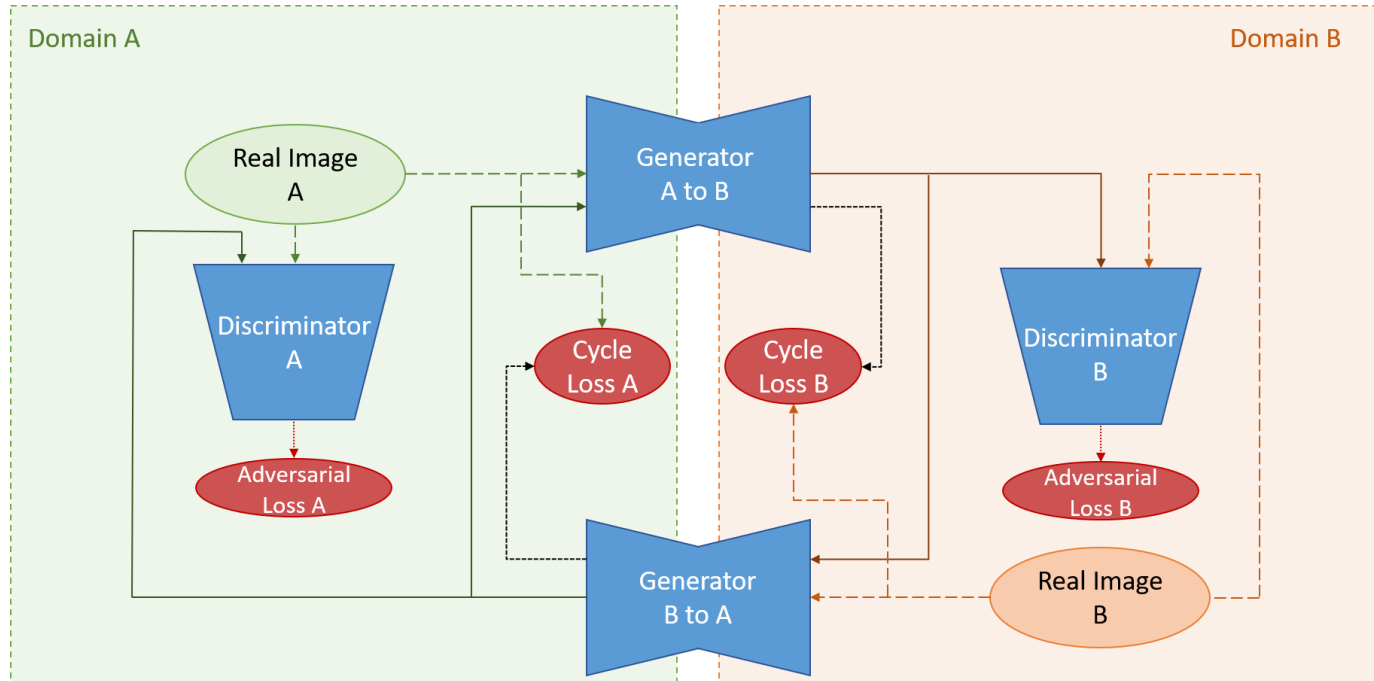
- Mapping between the two domains is unclear
- Data with matching input-output pairs is not available



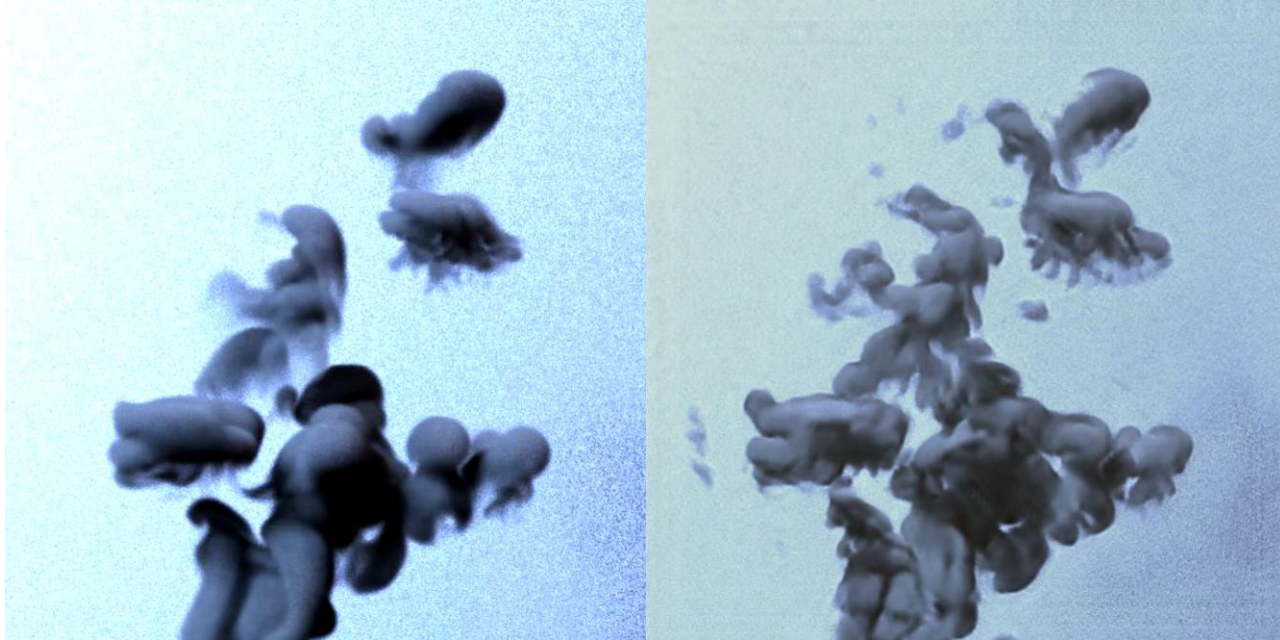
CycleGAN: Unpaired Image-to-Image Translation



CycleGAN: Unpaired Image-to-Image Translation



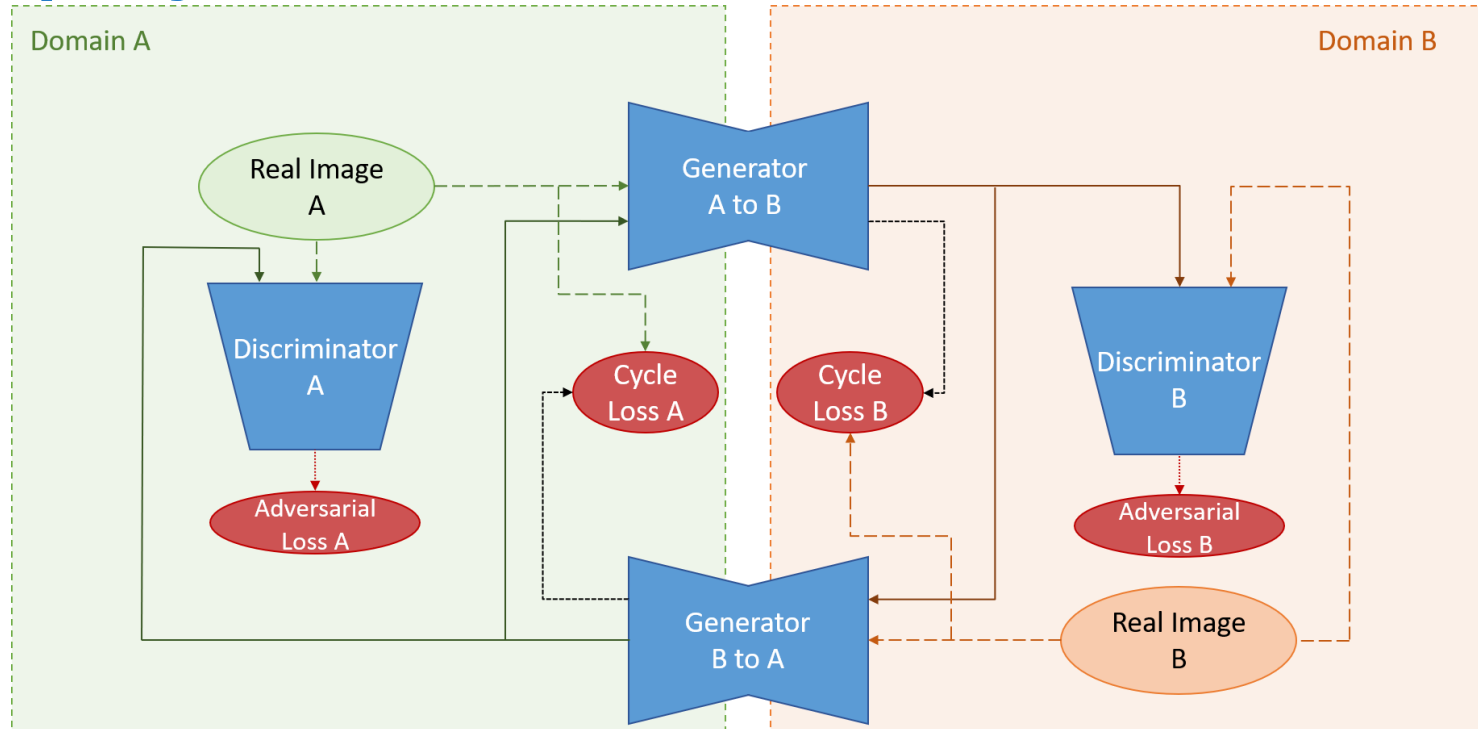
Problem: Temporal Inconsistency



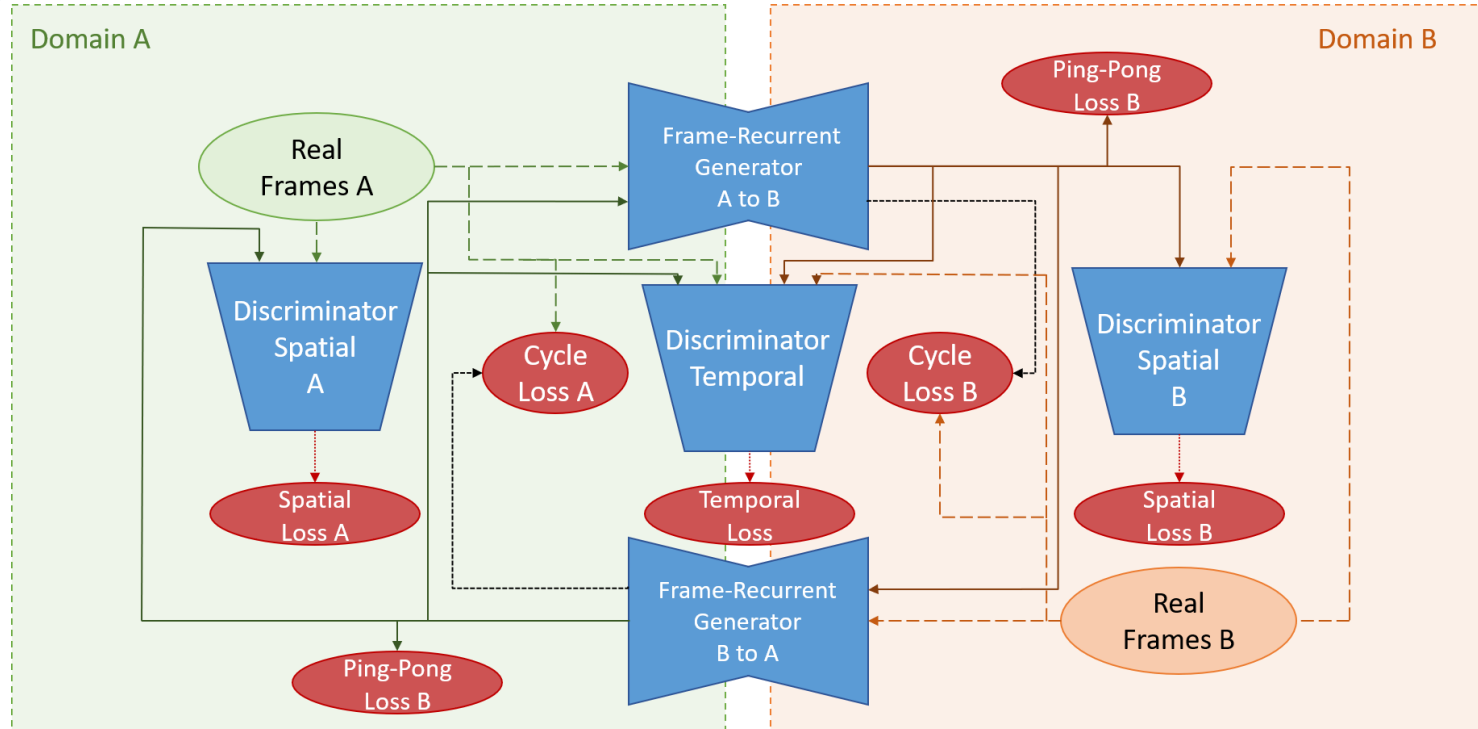
TricycleGAN

- Solution for Video-to-Video Translation
- Using unpaired Data
- Based on a cycleGAN

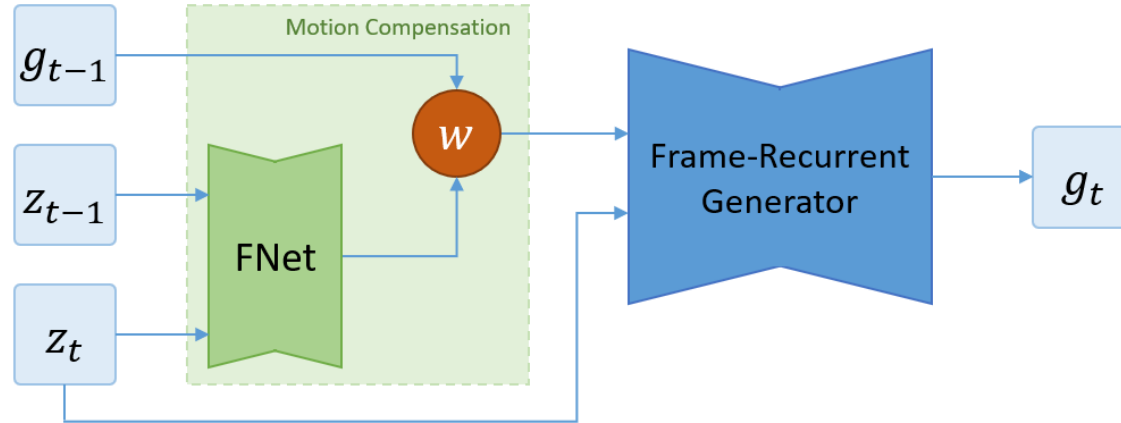
Recap: cycleGAN



TricycleGAN

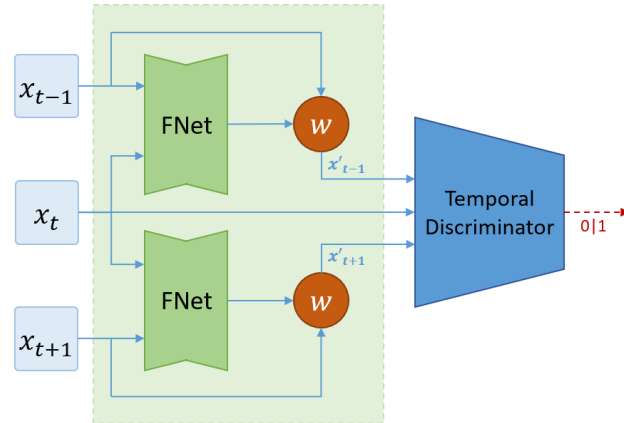


Frame-Recurrent Generator



- Manually fed with last frame
- Motion compensation with FNet

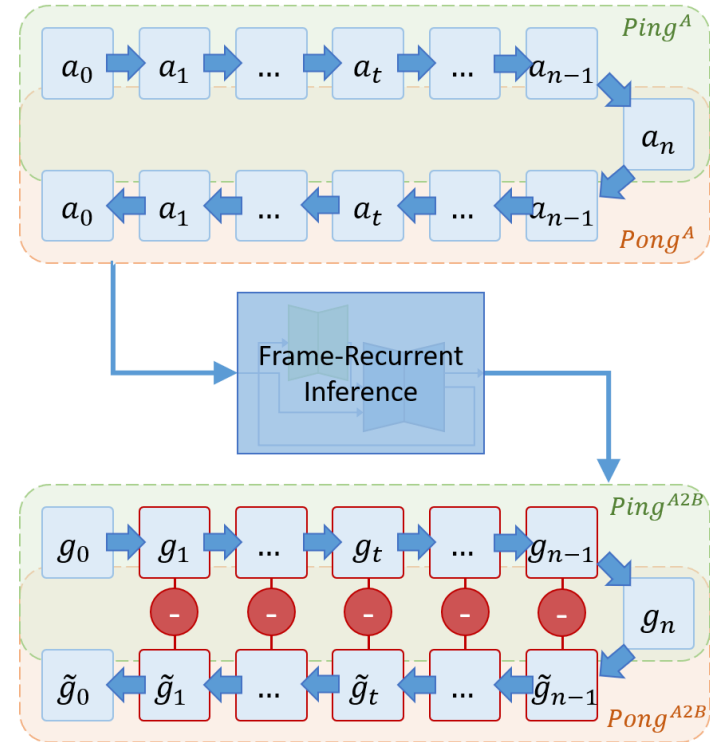
Temporal Discriminator



- Feed three consecutive frames
- Motion compensation with FNet

Ping-Pong Loss

- Long-term consistency
- **Premise:** Result independent of frame ordering

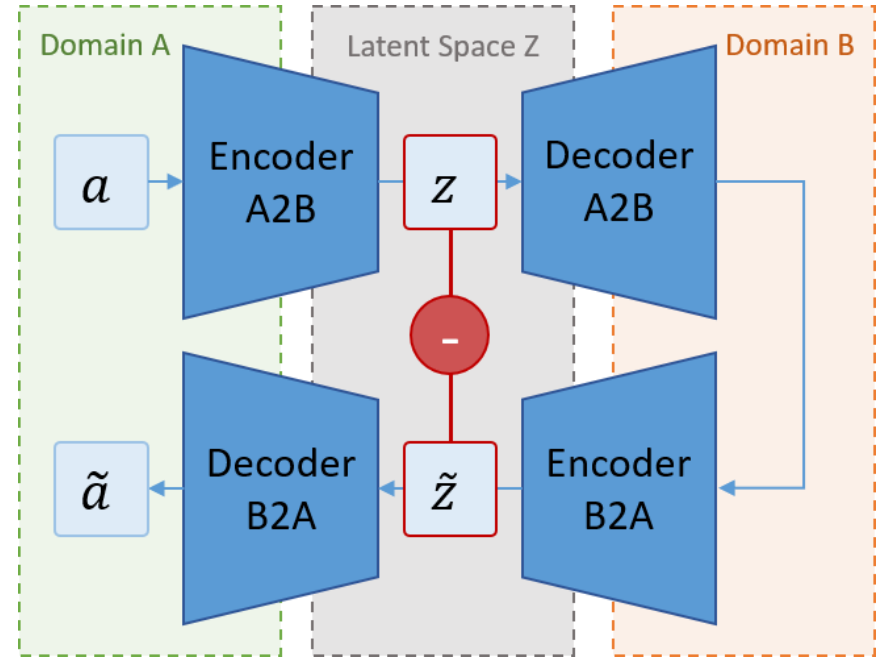


Problem: Training Time

- Typically convergence after ~30 hours
- Speeding up iterations only at cost of quality
- **Approach:** Speed up convergence by constraining training

Latent Space Consistency Loss

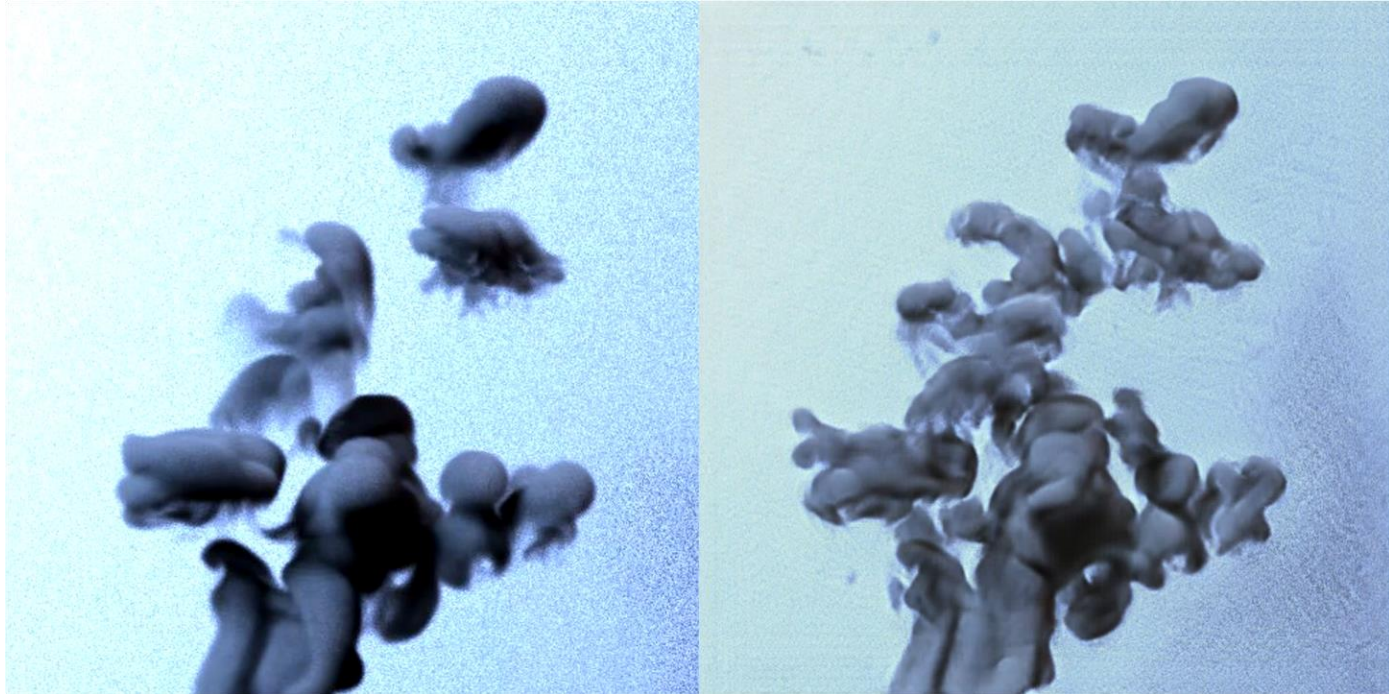
- Based on Encoder-Decoder Structure of Generators
- **Constraint:** Equal latent space representations



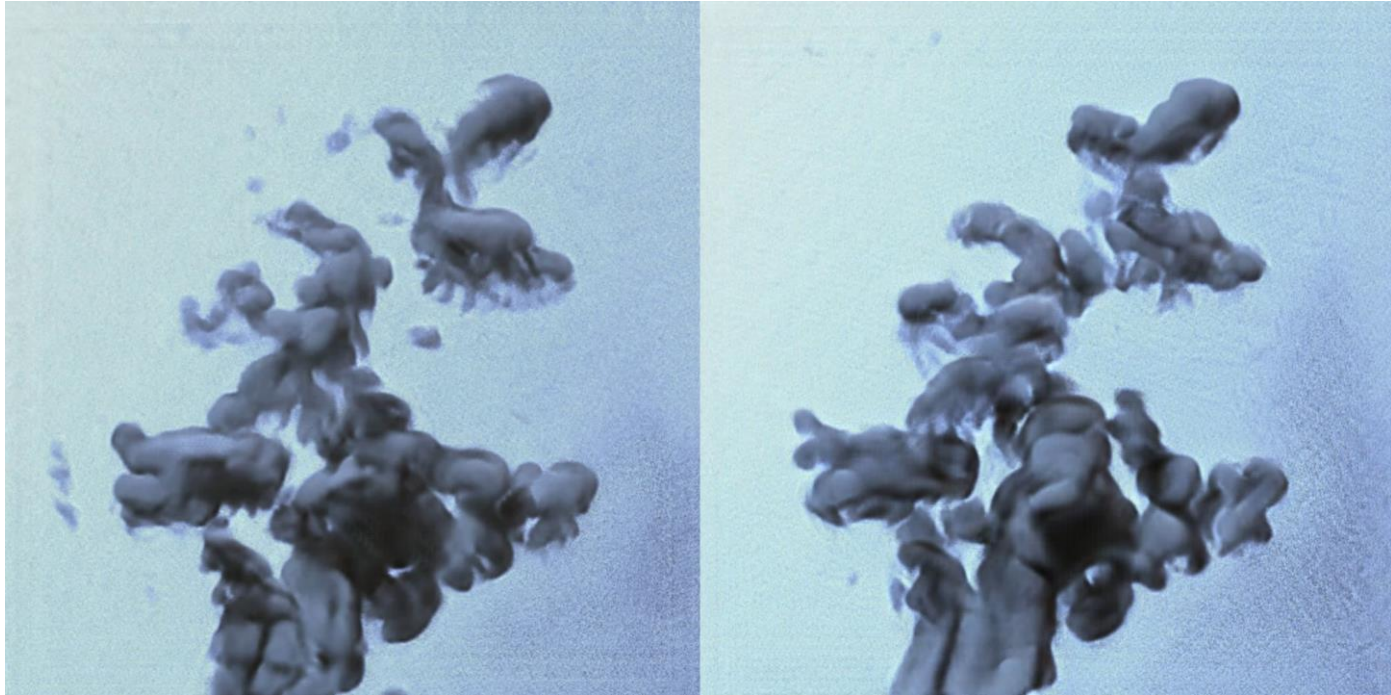
Discriminator Style Loss

- **Idea:** Generate same features as real data in discriminators
→ Feature loss using discriminators
- Due to unpaired data: spatial awareness not useful
→ Compute gram matrices instead

Results: Low-Res to High-Res Simulation

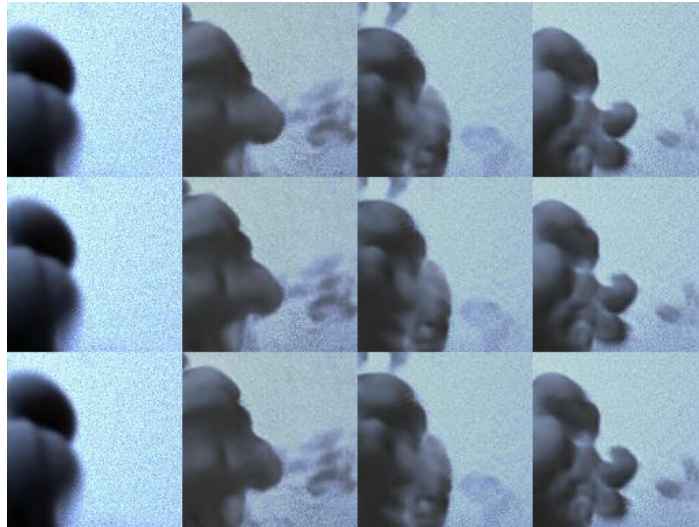


Results: Comparison CycleGAN



Results: Ablation Study

Results



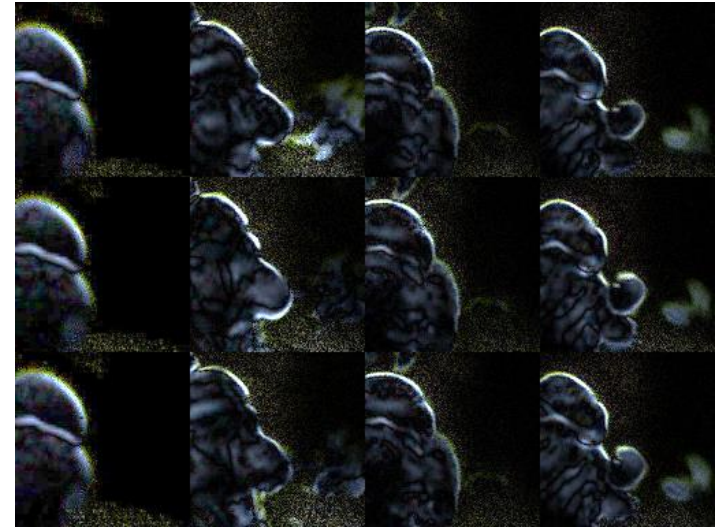
Input

CycleGAN

TricycleGAN light

TricycleGAN

Frame Differences



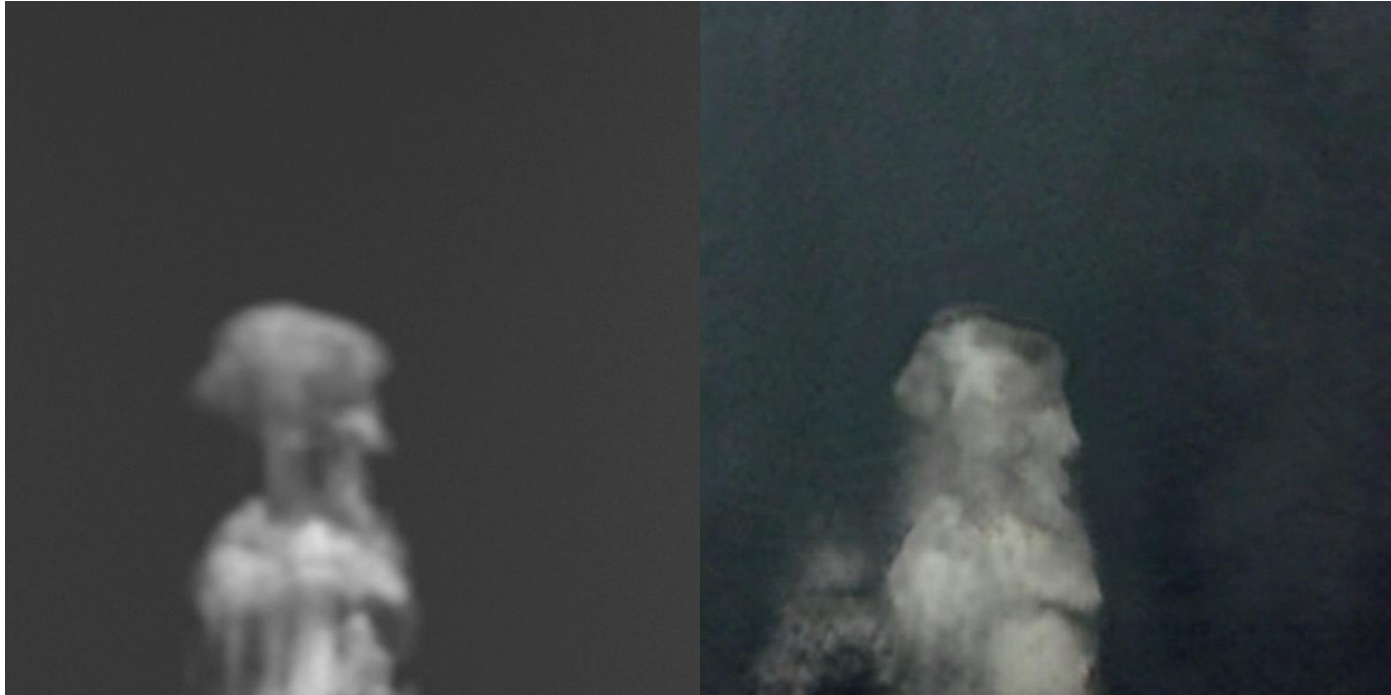
Input

CycleGAN

TricycleGAN light

TricycleGAN

Results: Simulation to Real Smoke



Results: Ping-Pong Loss



Results: Obama to Trump



Results: Trump to Obama



Results: Horse to Zebra



Limitations

- Generators and discriminators too simple
- Data quality
- L2 norm in Ping-Pong loss
- No domain-specific spatio-temporal properties

Conclusion

- Temporal Consistent CycleGAN
 - Temporal discriminator and frame-recurrent generators
 - Ping-Pong loss for long-term consistency
 - Novel discriminator style loss
- Generalizable Approach

Questions?

Appendix

Temporal Consistency: Previous Work

- Use optical flows to compensate for motion
- Compute pixelwise loss between frames
- **Problem:** produces blurry results

Implementation Details

- Implemented in Python/TensorFlow
- Trained on Nvidia Titan RTX