Imperial College London

ICLR

May, 2021

# DINO: A Conditional Energy-Based GAN for Domain Translation

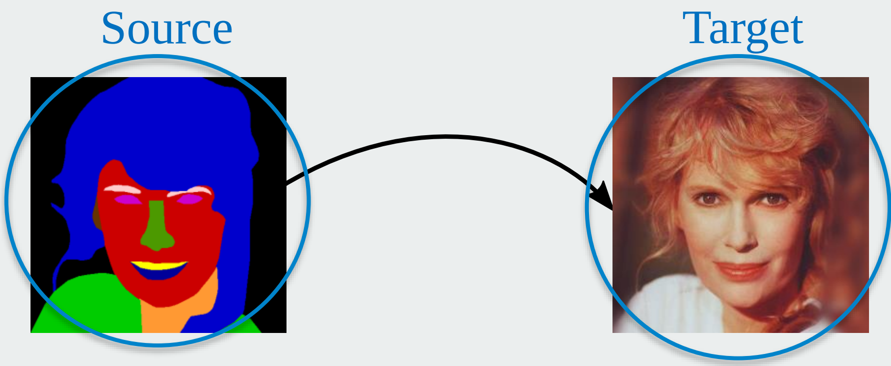
Konstantinos Vougioukas, Stavros Petridis & Maja Pantic

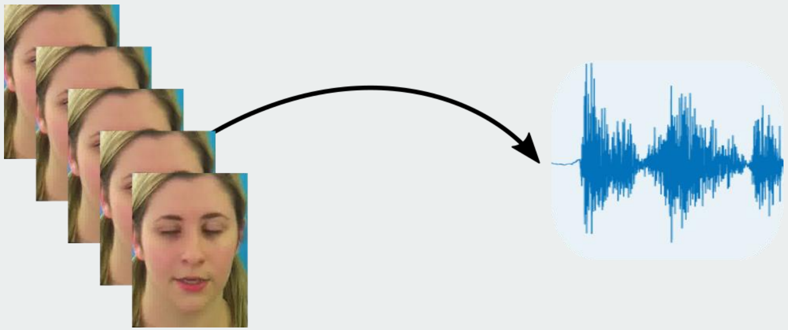
## Domain Translation Methods

- Domain translation should:
  1. Synthesize realistic data
  2. Maintain sample correspondence

- Adversarial training produces realistic results but can struggle to enforce correspondence.

- GAN-based translation systems often require:
  - Perceptual Losses
  - Reconstruction Losses (i.e., L1/L2)

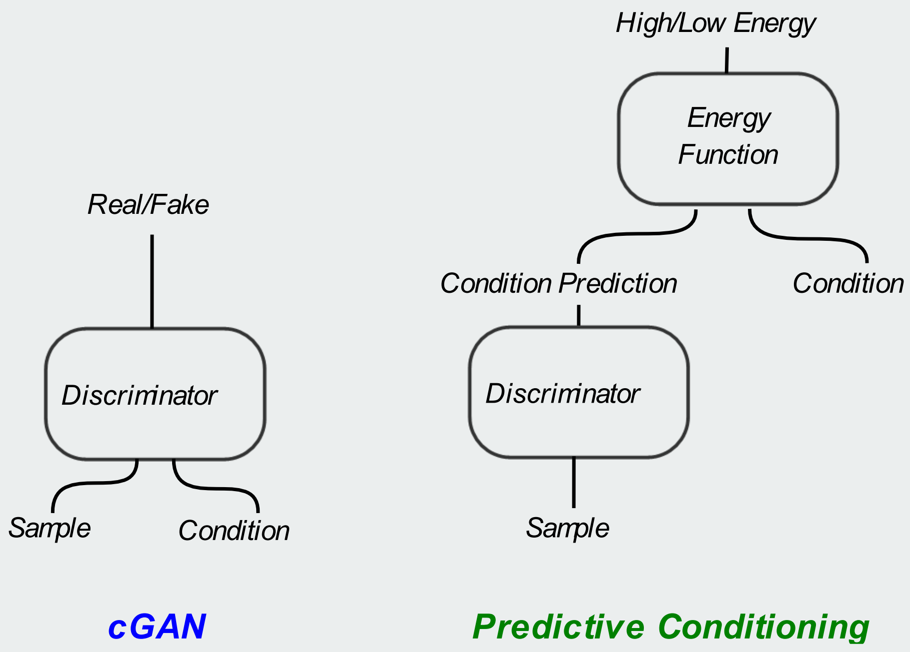


Imperial College London



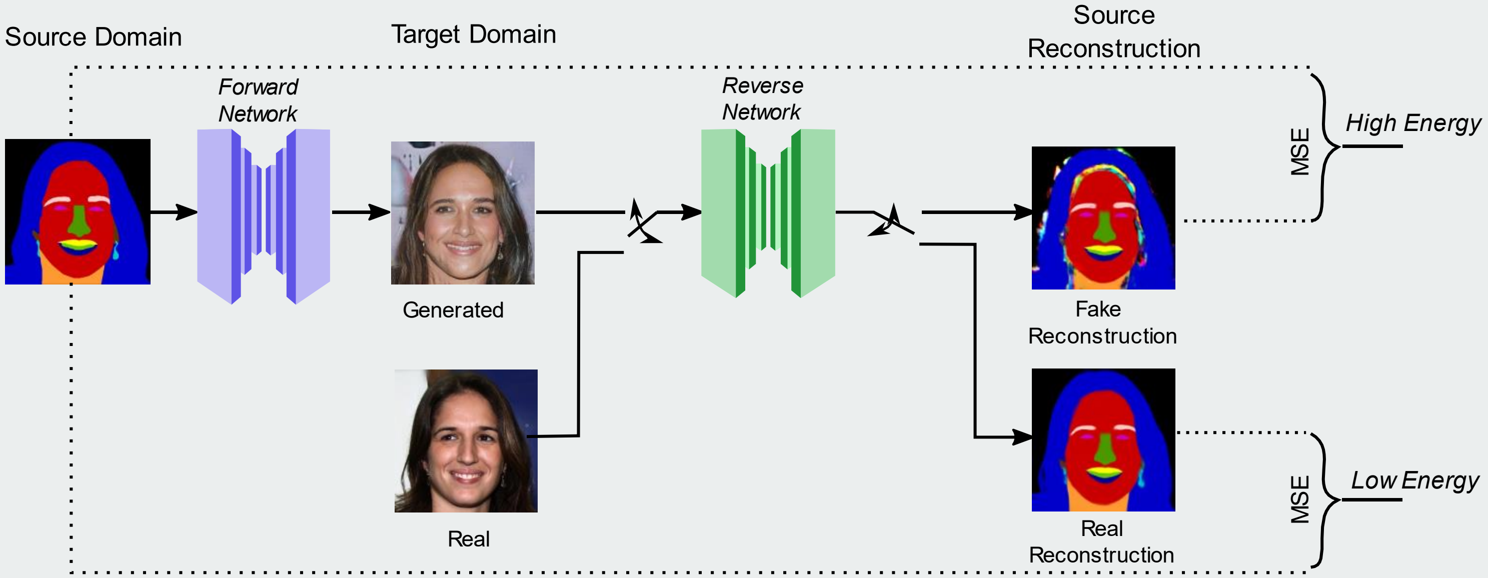
- The discrimination in cGANs **may** ignore the condition-sample correspondence.
- Predicting the condition ensures that the condition driving the generation can be recovered from samples.
- Predictive conditioning enforces condition-sample correspondence.



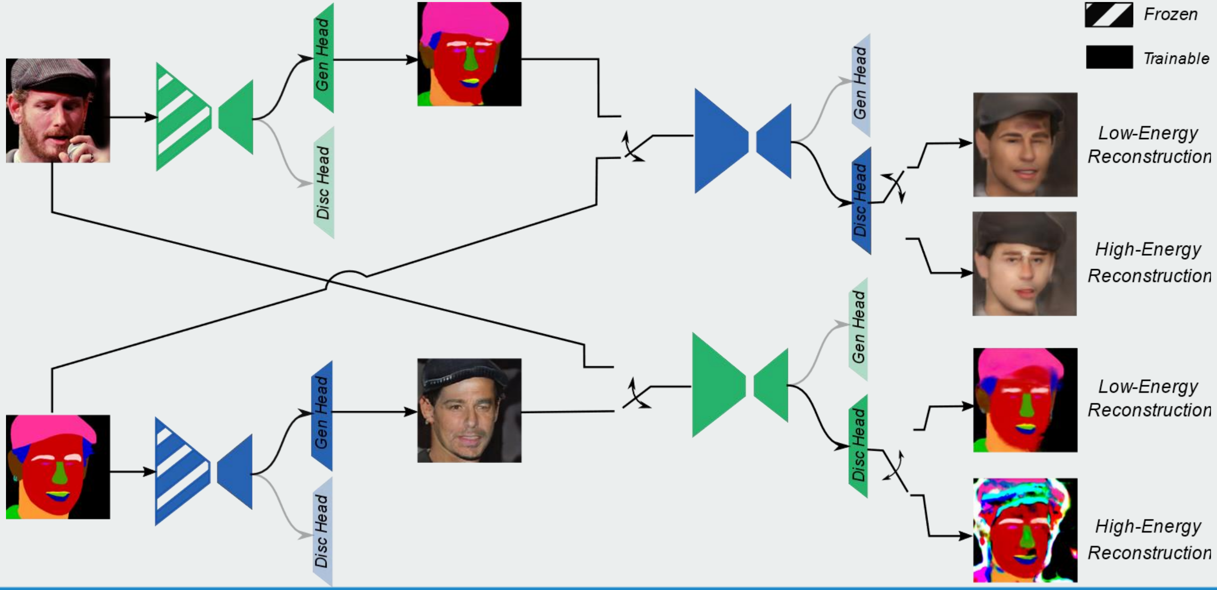
Imperial College London



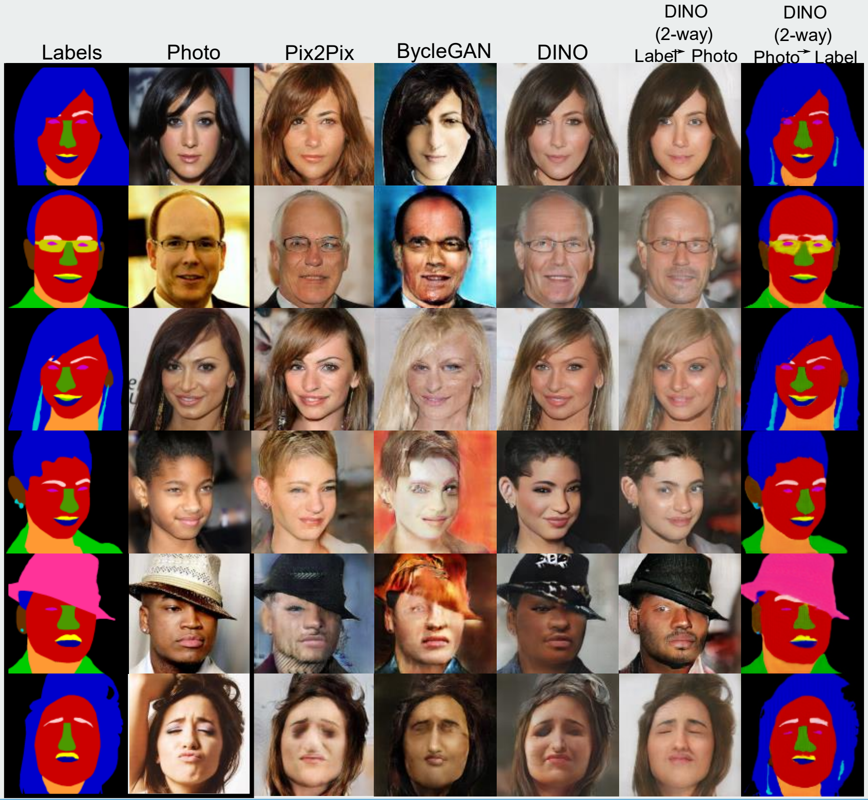
$$\min\max V(R,F) = \|R(y) - x\| - \|R(F(x)) - x\|$$



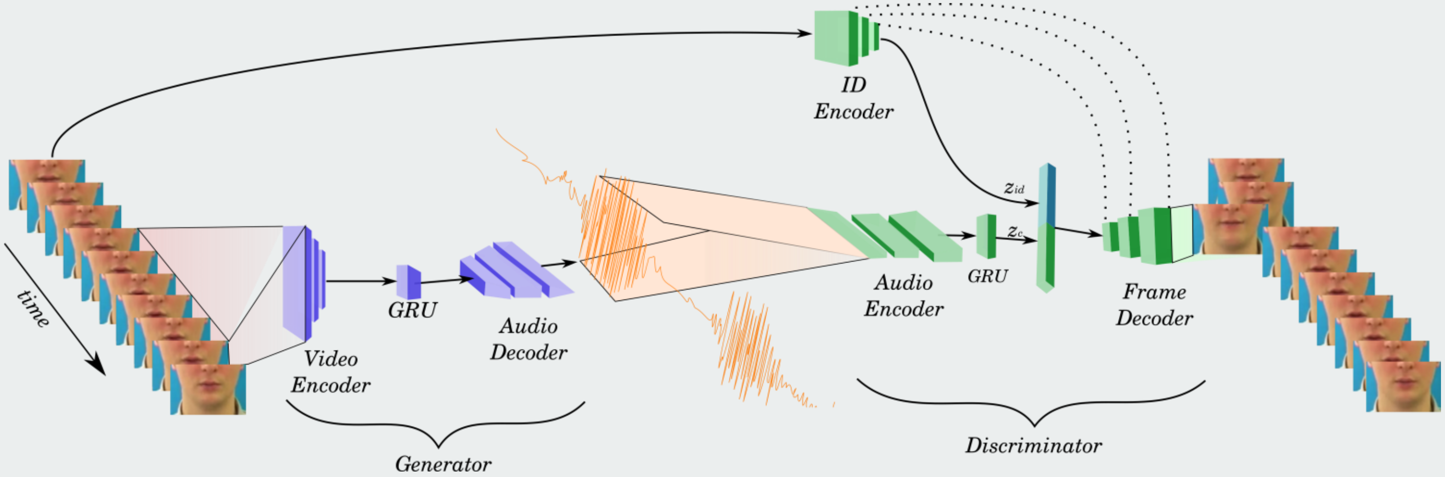
- Bidirectional training means both networks are treated as discriminator and generator.
- Training simultaneously for both objectives causes **instability** so training is decoupled. The encoder **is trained only for discrimination** and the decoder uses **different heads for generation and discrimination**.



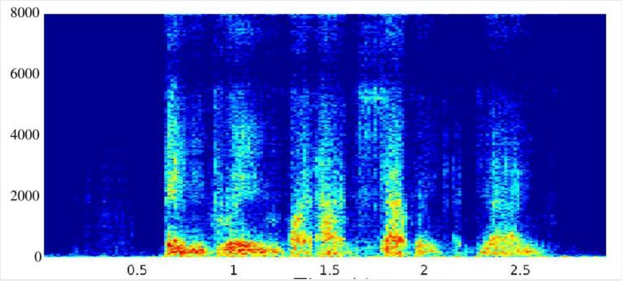
# Image-to-Image Translation




# Video-driven Speech Reconstruction




| Model | PESQ ↑ | STOI↑ | MCD↓ | WER↓ |
|---|---|---|---|---|
| Cond. WaveGAN | 0.96 | 0.37 | 43.7 | 81.7% |
| CycleGAN | 1.01 | 0.29 | 28.3 | 88.6% |
| Perceptual + GAN | **1.24** | 0.45 | 24.3 | 40.5% |
| DINO | 1.21 | **0.51** | **23.0** | **32.6%** |



Imperial College London

ICLR

https://github.com/DinoMan/DINO