MIT Computer Science and Artificial Intelligence Lab

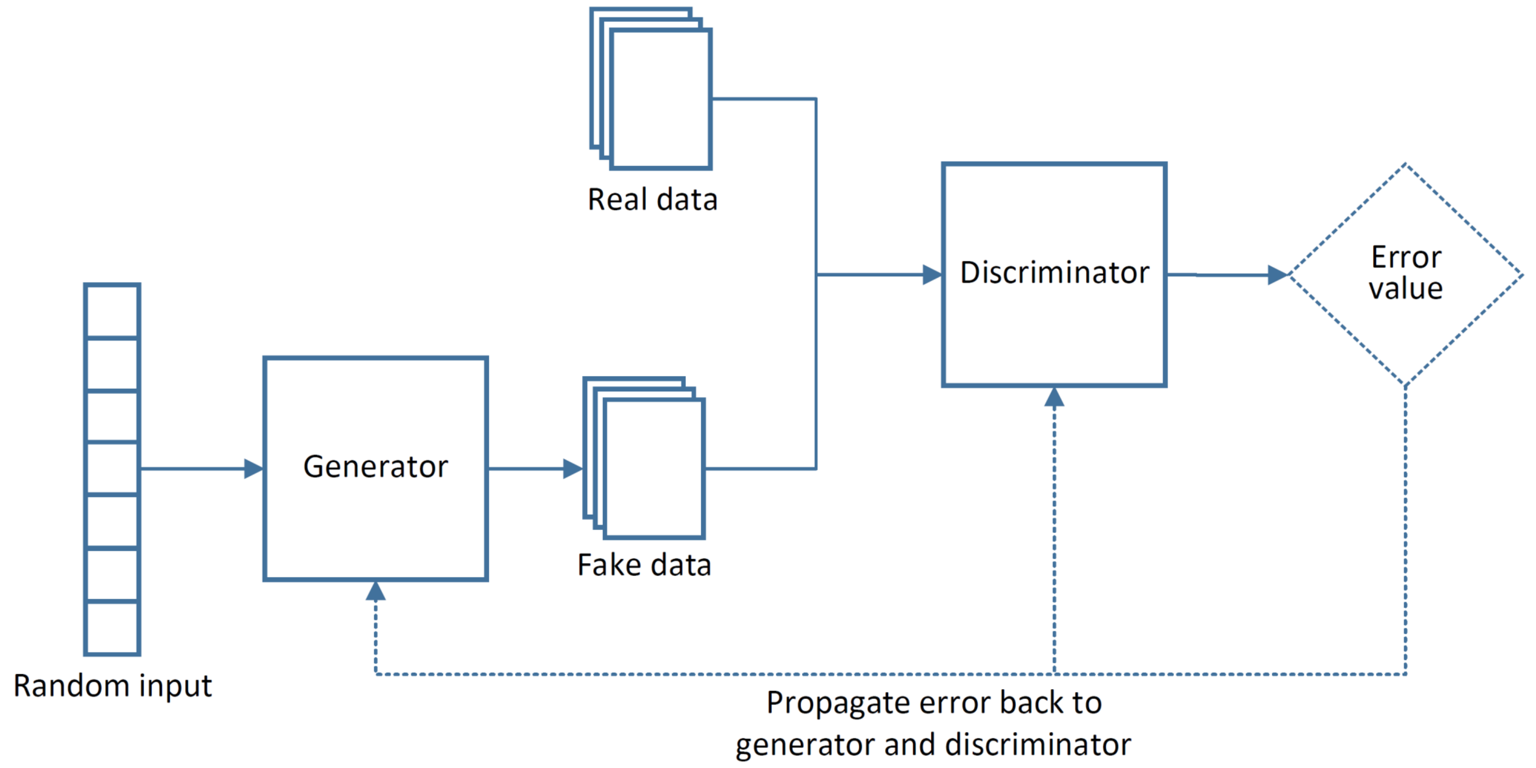
This is a caption for the example and it may be this long or longer

A System that Scales Robust Generative Adversarial Network Training

Jamal Toutouh, Tom Schmiedlechner, Ignavier Ng Zhi Yong, Abdullah Al-Dujaili, Erik Hemberg, Una-May O'Reilly

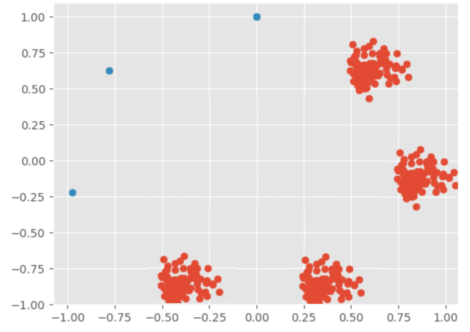
MOTIVATION

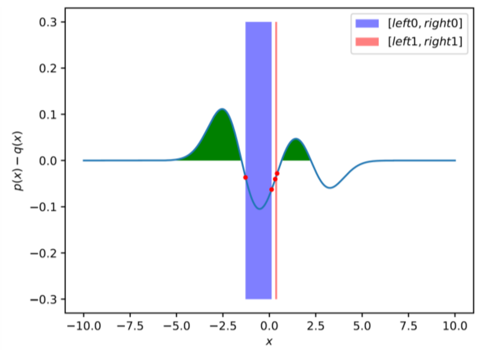
Generative Adversarial Networks



$$\min_{\mathcal{G}} \max_{\mathcal{D}} \mathcal{L}(\mathcal{D},\mathcal{G}) = \mathbb{E}_{x\sim p_{data}(x)}[\phi(\mathcal{D}(x))] + \mathbb{E}_{x\sim p_z(z)}[\phi(1-\mathcal{D}(\mathcal{G}(z)))]$$

- Simultaneous gradient updates in GAN training lead to **unstable dynamics**
- Similar degenerate behaviors have been studied by the coevolutionary computing for minimax optimization
  - **Mode Collapse ↔ Focusing**
  - **Discriminator Collapse ↔ Relativism**
  - **Vanishing Gradients ↔ Loss of Gradients**

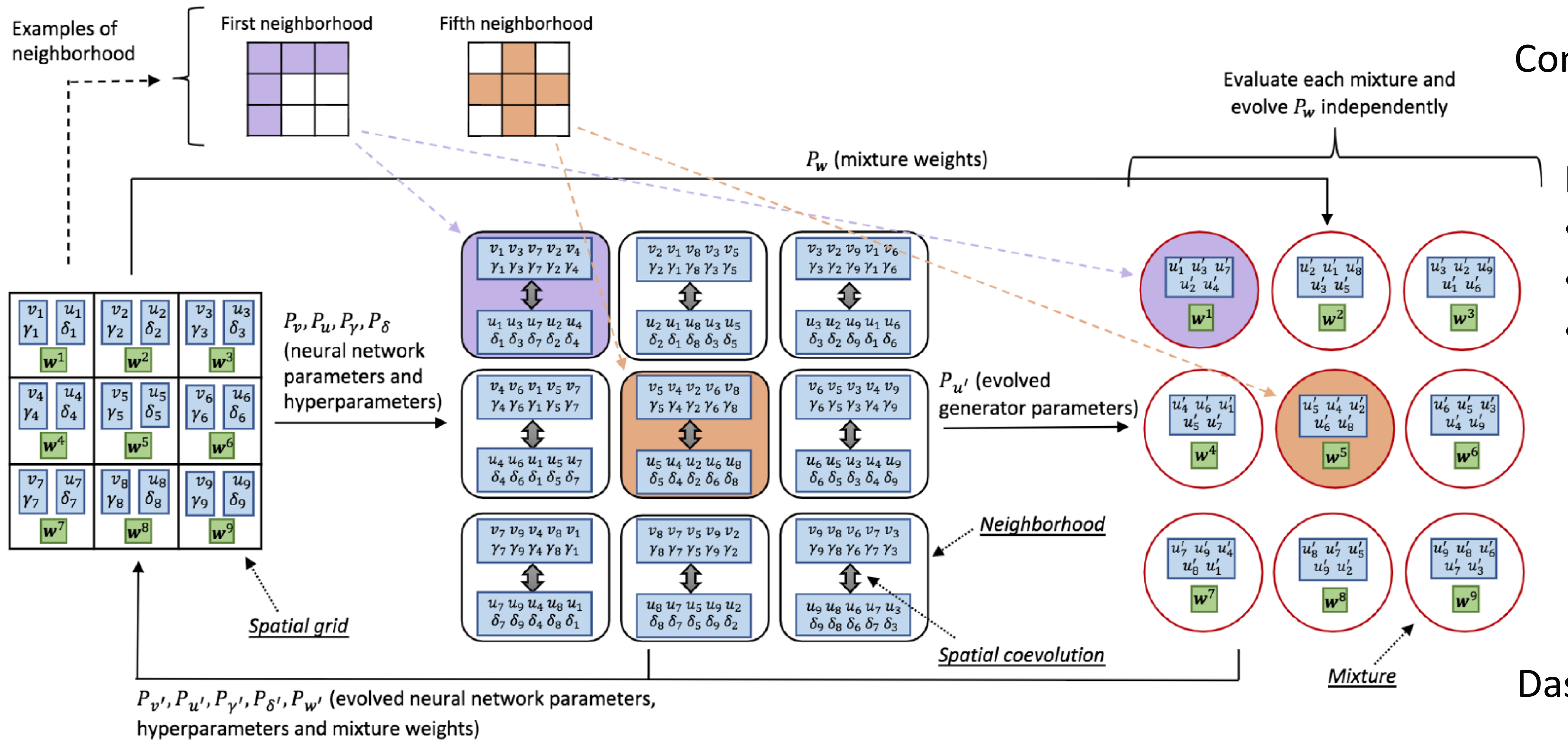




- Supplementing GAN training with established coevolutionary techniques proposed to address the unstable dynamics

LIPIZZANER: DISTRIBUTED COEVOLUTIONARY GANs

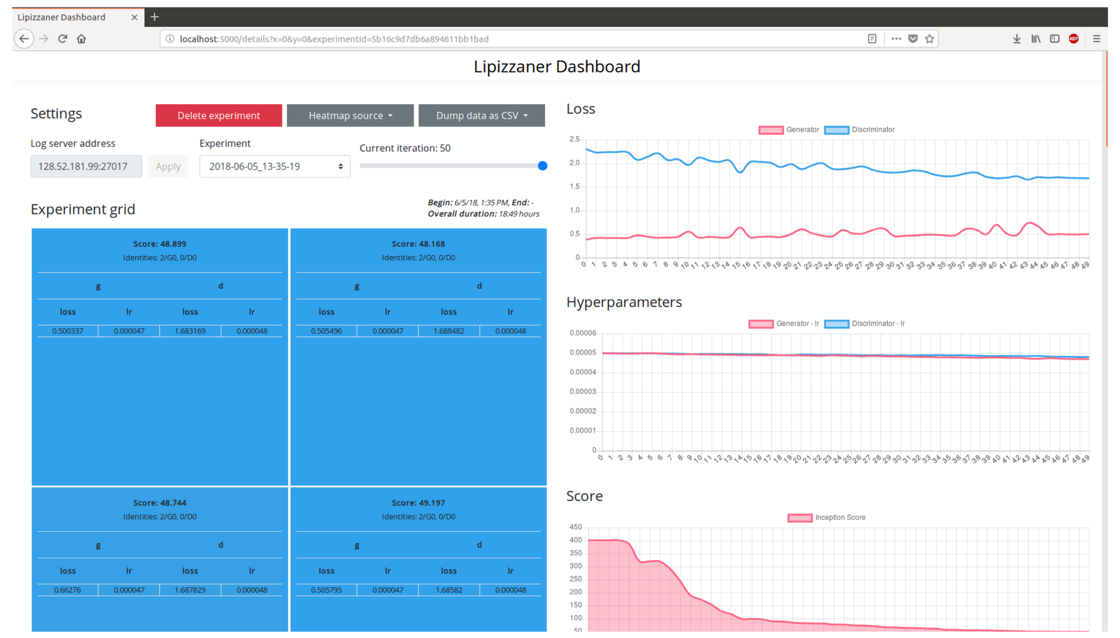
Lipizzaner is a distributed, coevolutionary framework to simultaneously train multiple GANs with gradient-based optimizers



Combines
- Comp Coev: Gradient-steps + Gaussian mutation
- Coop Coev: Evolve mixture weights for neighborhoods

Lipizzaner distributes populations over a spatial grid
- N grid cells represent population of training instances
- Cells communicate only with neighboring cell
- Computational benefits
  - **Linear** instead of quadratic scaling $N*neighborhood\ size$
  - Large datasets can be split onto different cells

Dashboard



- **Fast & improved convergence**
- **Diverse solutions**
- **Robustness**
- **Scalability**

RESULTS

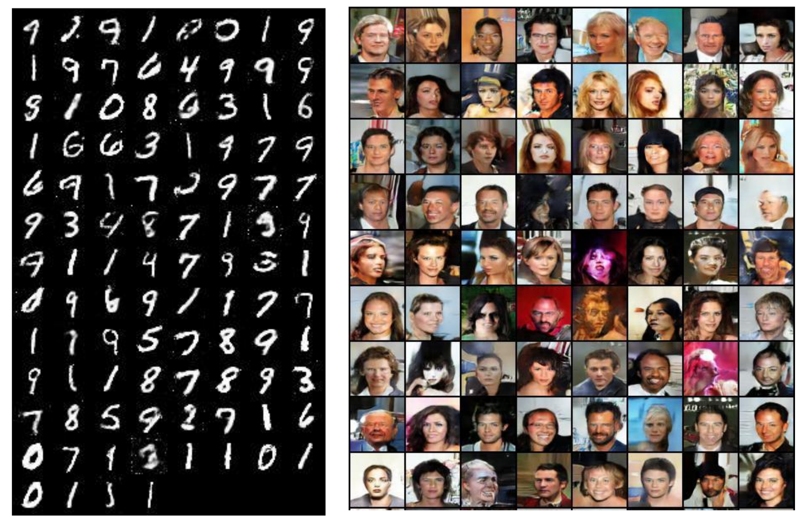


Figure: MNIST and Celeba Images generated by using Lipizzaner

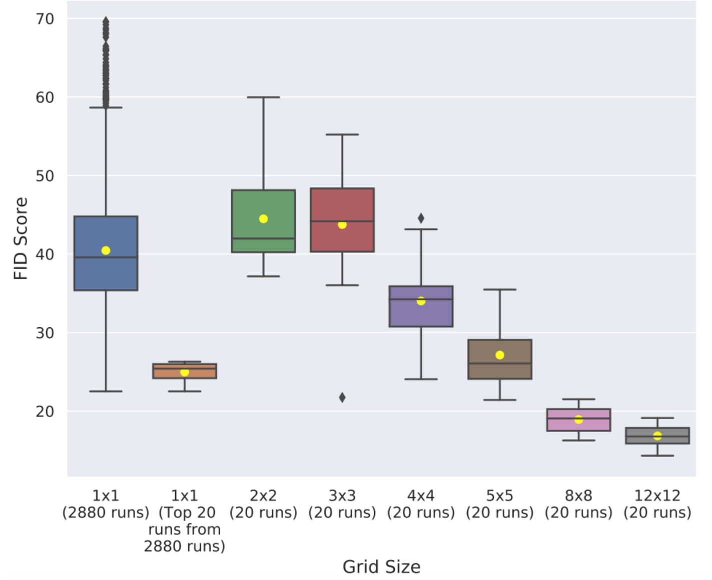


Figure: Box plot of the FID score for different grid sizes on MNIST.

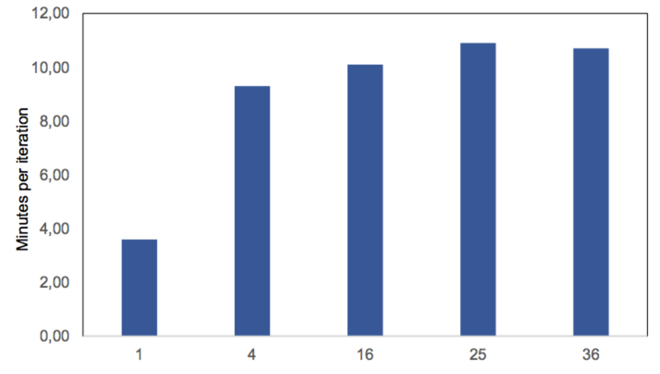
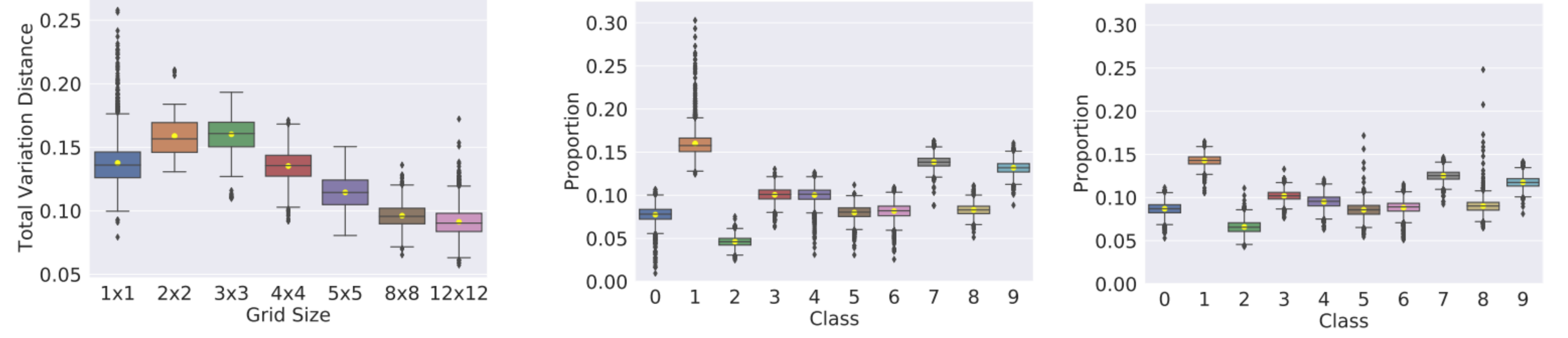


Figure: Near linear training times.



(a) TVD for different grid sizes (b) Classes distribution for 1×1 (d) Classes distribution for 12×12

Figure: The average TVD is shown in (a). The larger grid sizes have lower TVD. This is further supported by visualizing distribution of each classes of generated images for different grid sizes (b for 1x1 and d for 12x12). The distribution for 12×12 is the most uniform.

MIT Computer Science and
Artificial Intelligence Lab

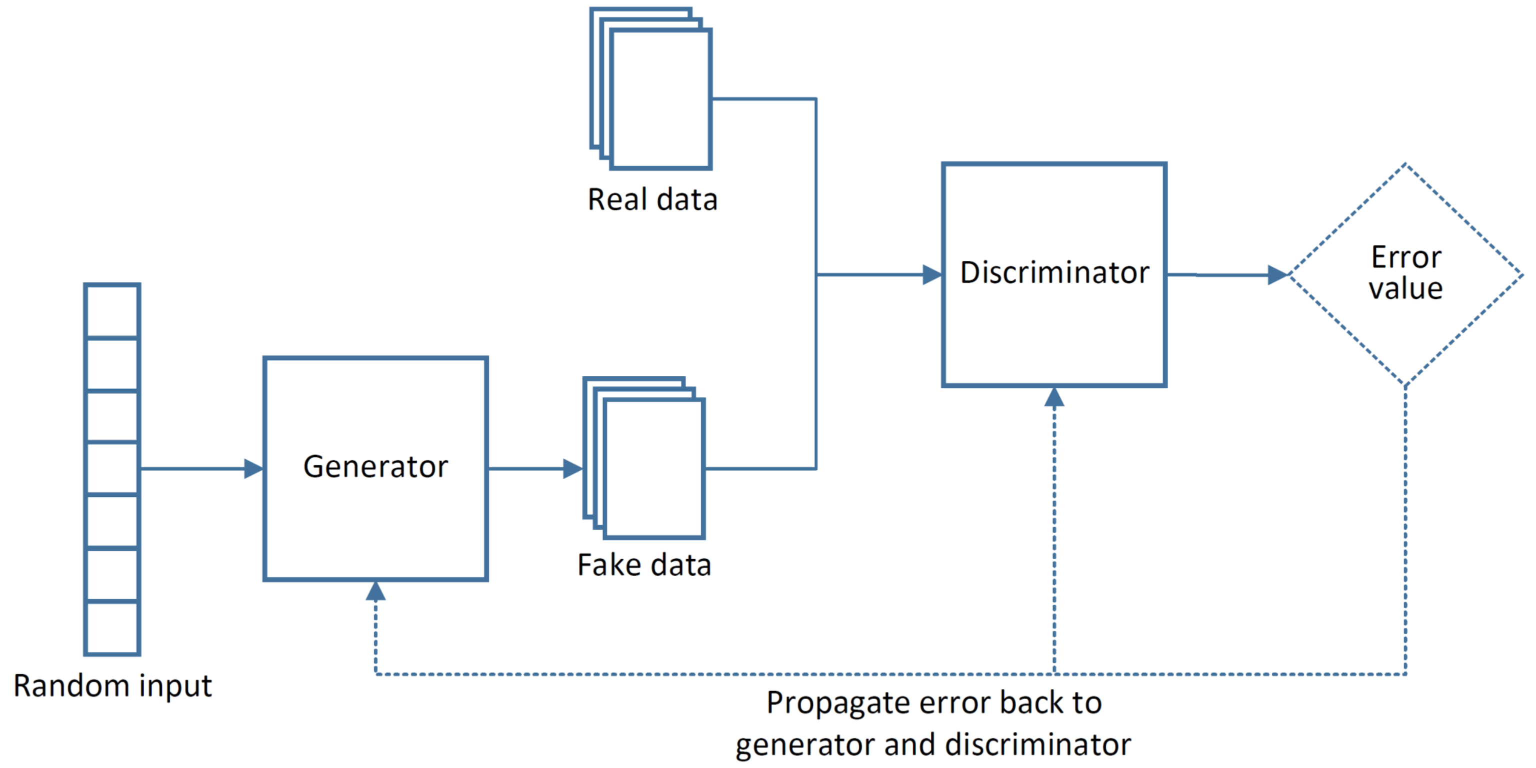
A System that Scales Robust Generative Adversarial Network Training

Jamal Toutouh, Tom Schmiedlechner,
Ignavier Ng Zhi Yong, Abdullah Al-Dujaili,
Erik Hemberg, Una-May O'Reilly

This is a caption for the example and it may be this long or longer

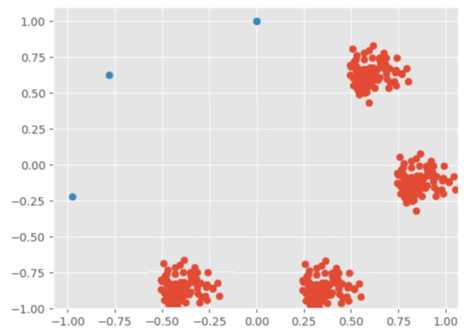
MOTIVATION

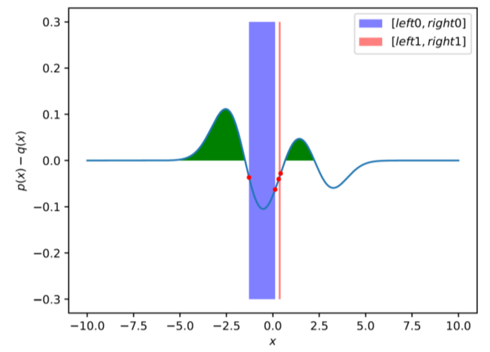
Generative Adversarial Networks



$$\min_{\mathcal{G}} \max_{\mathcal{D}} \mathcal{L}(\mathcal{D},\mathcal{G}) = \mathbb{E}_{x\sim p_{data}(x)}[\phi(\mathcal{D}(x))] + \mathbb{E}_{x\sim p_z(z)}[\phi(1-\mathcal{D}(\mathcal{G}(z)))]$$

- Simultaneous gradient updates in GAN training lead to **unstable dynamics**
- Similar degenerate behaviors have been studied by the coevolutionary computing for minimax optimization
  - Mode Collapse ↔ Focusing
  - Discriminator Collapse ↔ Relativism
  - Vanishing Gradients ↔ Loss of Gradients





- Our motivation is to supplement GAN training with established coevolutionary techniques proposed to address the unstable dynamics

LIPIZZANER

Lipizzaner is a distributed, coevolutionary framework to simultaneously train multiple GANs with gradient-based optimizers

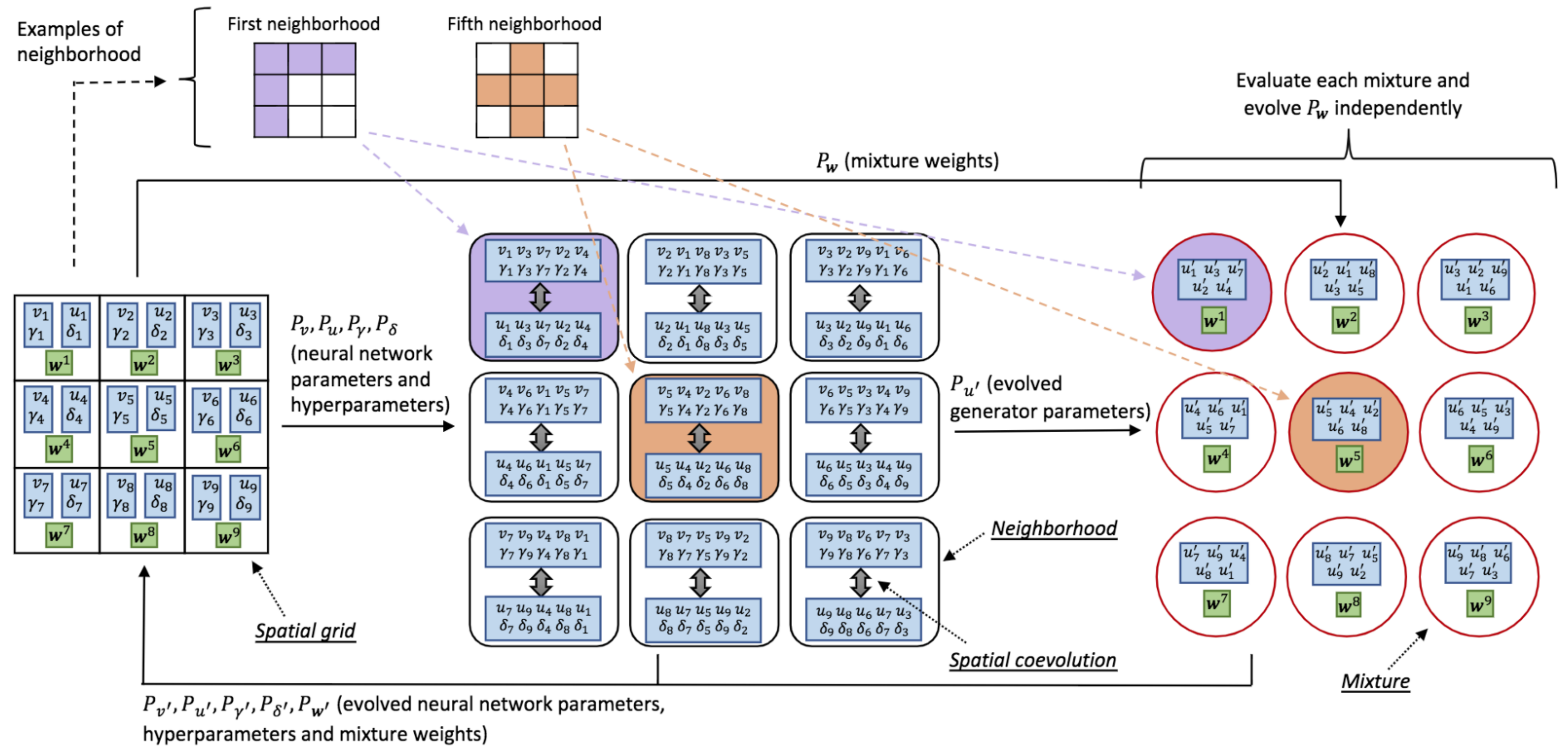
- It combines the advantages of both technologies:
  - Fast convergence due to gradient-based steps
  - Robustness to mode and discriminator collapse due to coevolution
  - Scalability due to a spatial distribution topology
  - Diverse solutions due to mixture evolution
  - Improved convergence due to hyperparameter evolution
- The training algorithm combines:

1) Competitive Coevolution:
  - Gradient-steps to evolve neural network weights and biases
  - Gaussian mutation to evolve optimizer hyperparameters

2) Cooperative Coevolution
  - Evolve spatial mixture weights for local neighborhoods

DISTRIBUTED COEVOLUTIONARY GANs

Lipizzaner distributes populations over a spatial grid
  - N grid cells represent population of training instances
  - Cells communicate only with neighboring cell
- Computational Benefits
  - **Linear** instead of quadratic scaling $N{*}neighborhood\ size$
  - Large datasets can be split onto different cells



RESULTS

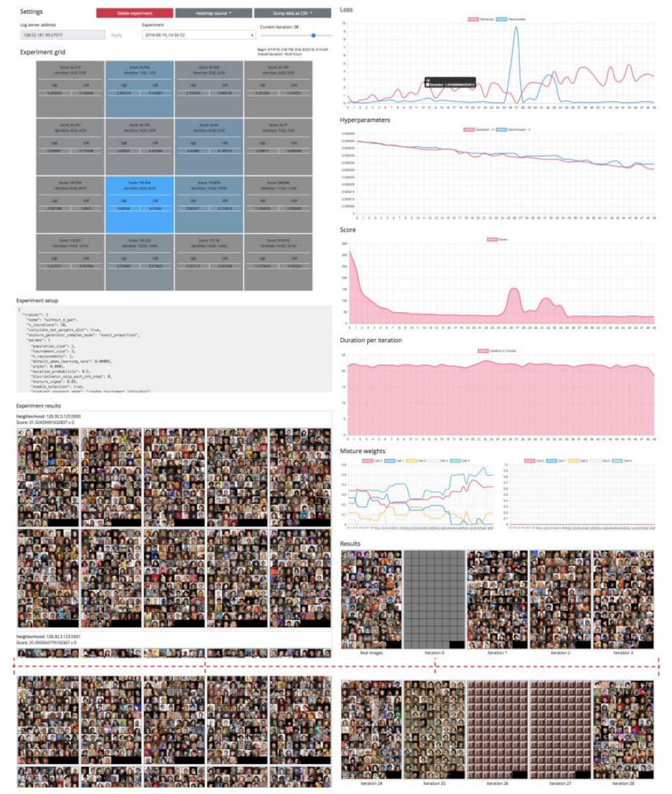


Figure: Lipizzaner Dashboard

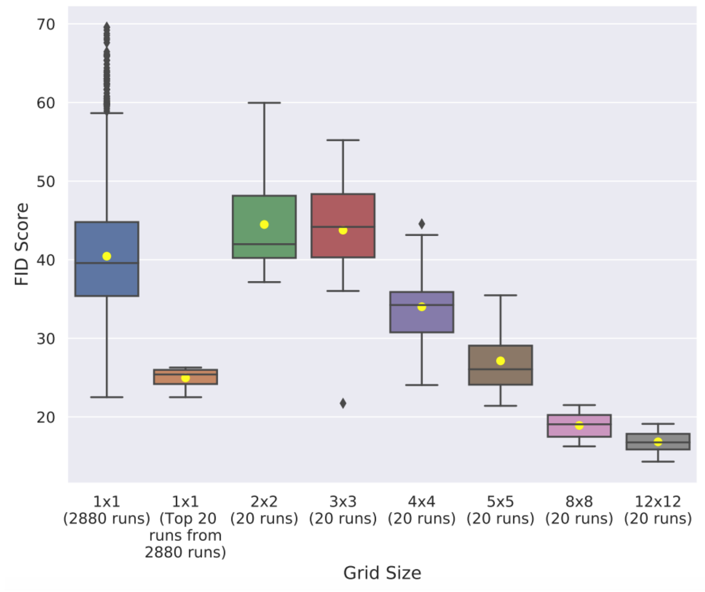


Figure: Box plot of the FID score for different grid sizes on MNIST.

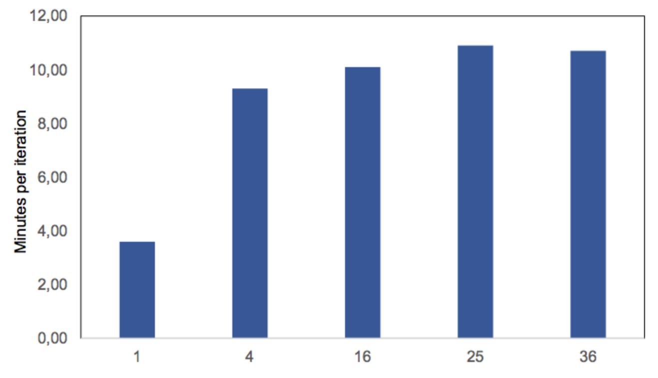
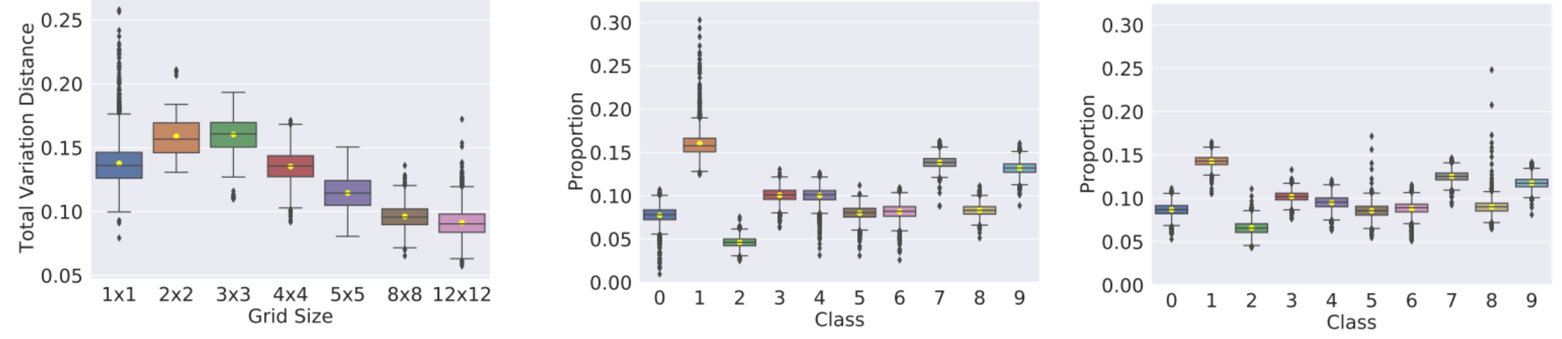


Figure: Near linear training times on AWS per iteration on the CelebA dataset, averaged over 30 iterations.



(a) TVD for different grid sizes (b) Classes distribution for 1 × 1 (d) Classes distribution for 12 × 12

Figure: Generator mixture distribution for MNIST. The average TVD is shown in (a). The larger grid sizes have lower TVD, which indicate that mixtures from larger grid sizes produce a more diverse set of images spanning across different classes. This is further supported by visualizing distribution of each classes of generated images for different grid sizes. The distribution of each classes of generated images for 1 × 1 is in (b). The distribution for 12 × 12,\ is the most uniform.