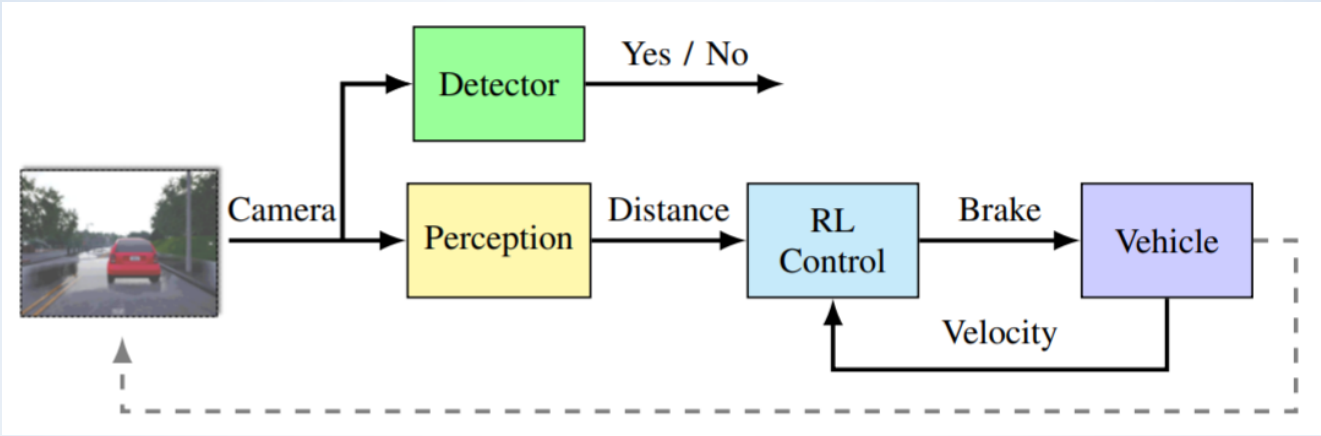
Assured Autonomy in Cyber Physical Systems Using Adversarial Autoencoders

Nicholas Potteiger, Feiyang Cai, Xenofon Koutsoukos

ISIS
Tel (615) 343-7472 Fax (615) 343-7440
1025 16th Avenue South|Nashville, TN 37212
www.isis.vanderbilt.edu

VANDERBILT UNIVERSITY

Motivations



Assurance monitoring based on inductive conformal anomaly detection

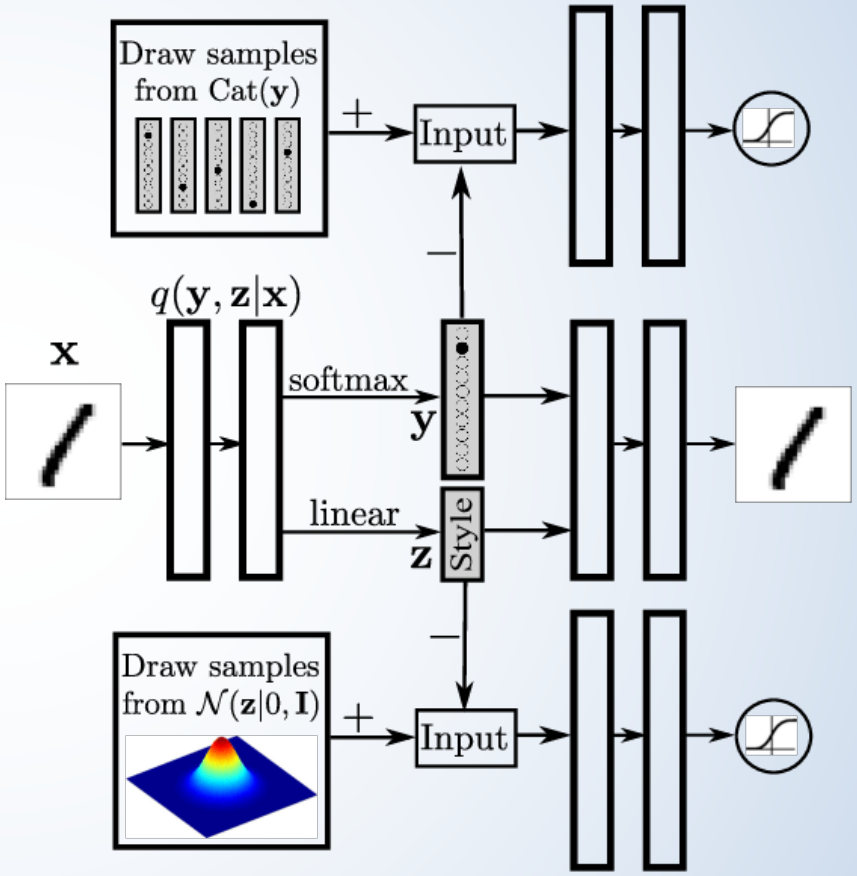
- Variational autoencoder (VAE)
- VAE for regression
- Adversarial Autoencoder (AAE)
- Deep support vector description (SVDD)

Evaluation using self-driving simulator and open datasets

- Advanced emergency braking system
- End-to-end self-driving controller

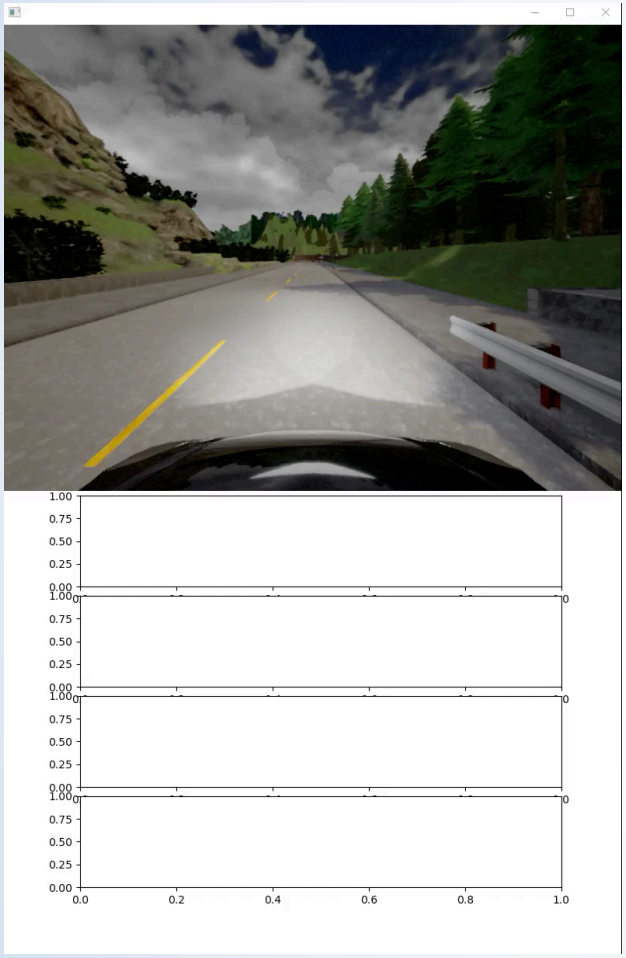
Adversarial Autoencoder Architecture

- Main idea: Train an appropriate neural network architecture
  - A nonconformity measure (NCM) to evaluate the degree to which a new input image disagrees from the distribution of the training data
  - Empirical *p*-values used for statistical significance testing
  - An assurance measure using a martingale process of the *p*-values

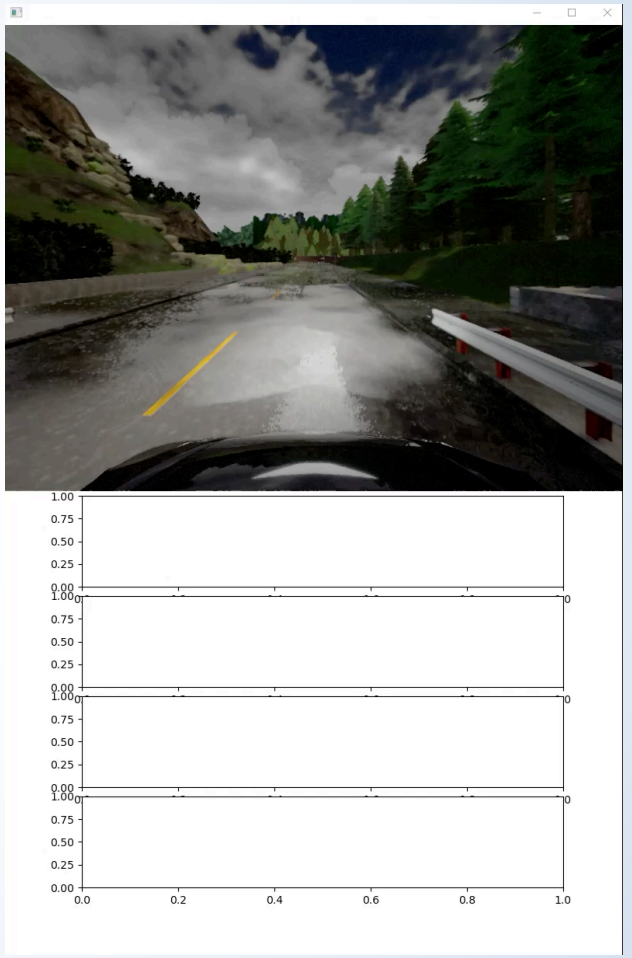


AAE in Action (Day)

In Distribution

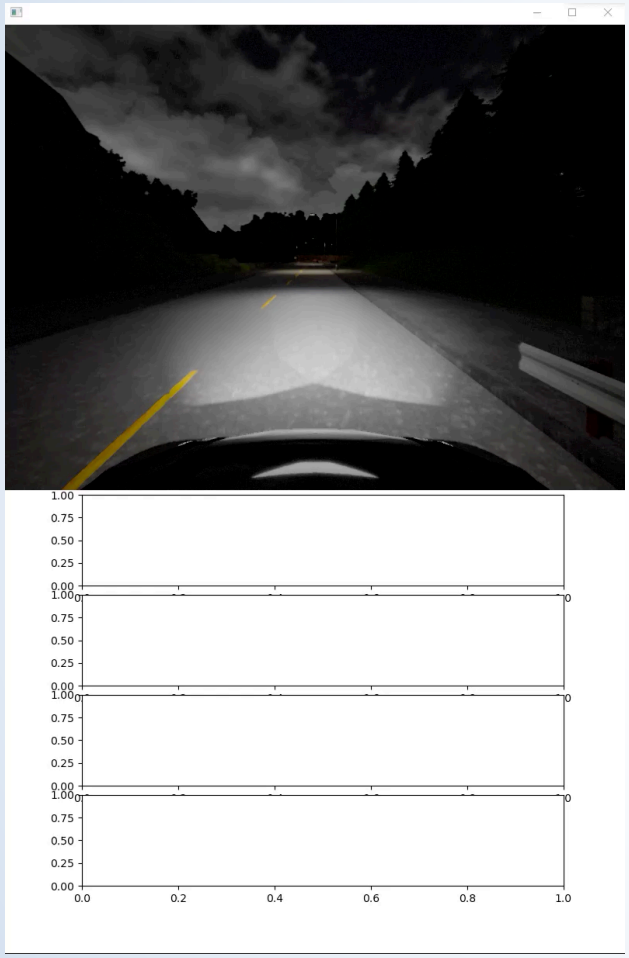


Out of Distribution

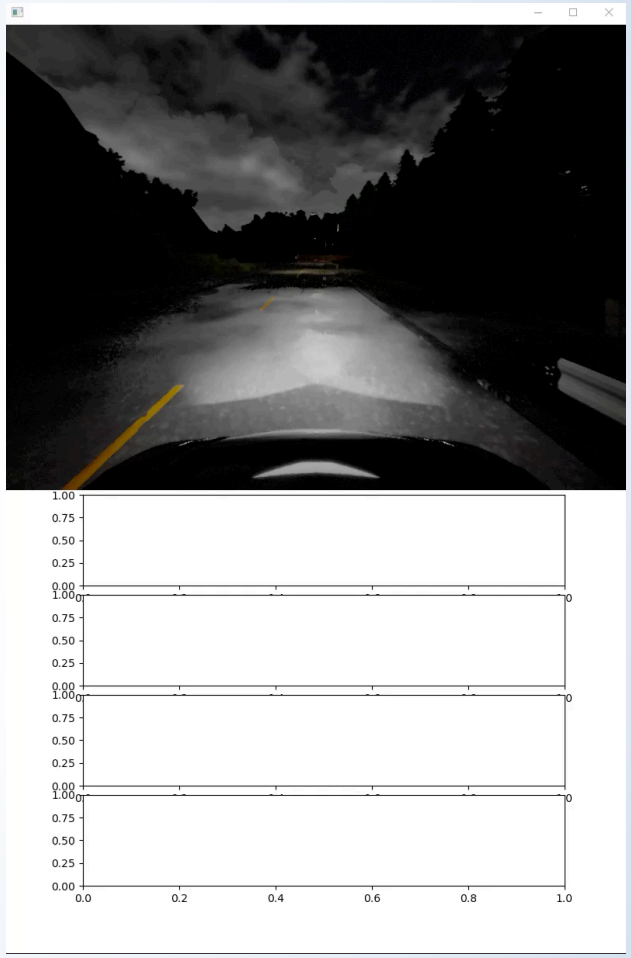


AAE in Action (Night)

In Distribution



Out of Distribution



Challenges & Future Work

- AAE network iterations
- Jumping into the code too early
- Real car images, more classes, higher resolution generated images
- Repeatability on multiple datasets