

# Video Future Frame Prediction

---

Vineeth S  
M. Tech Artificial Intelligence  
SR. No. 16543

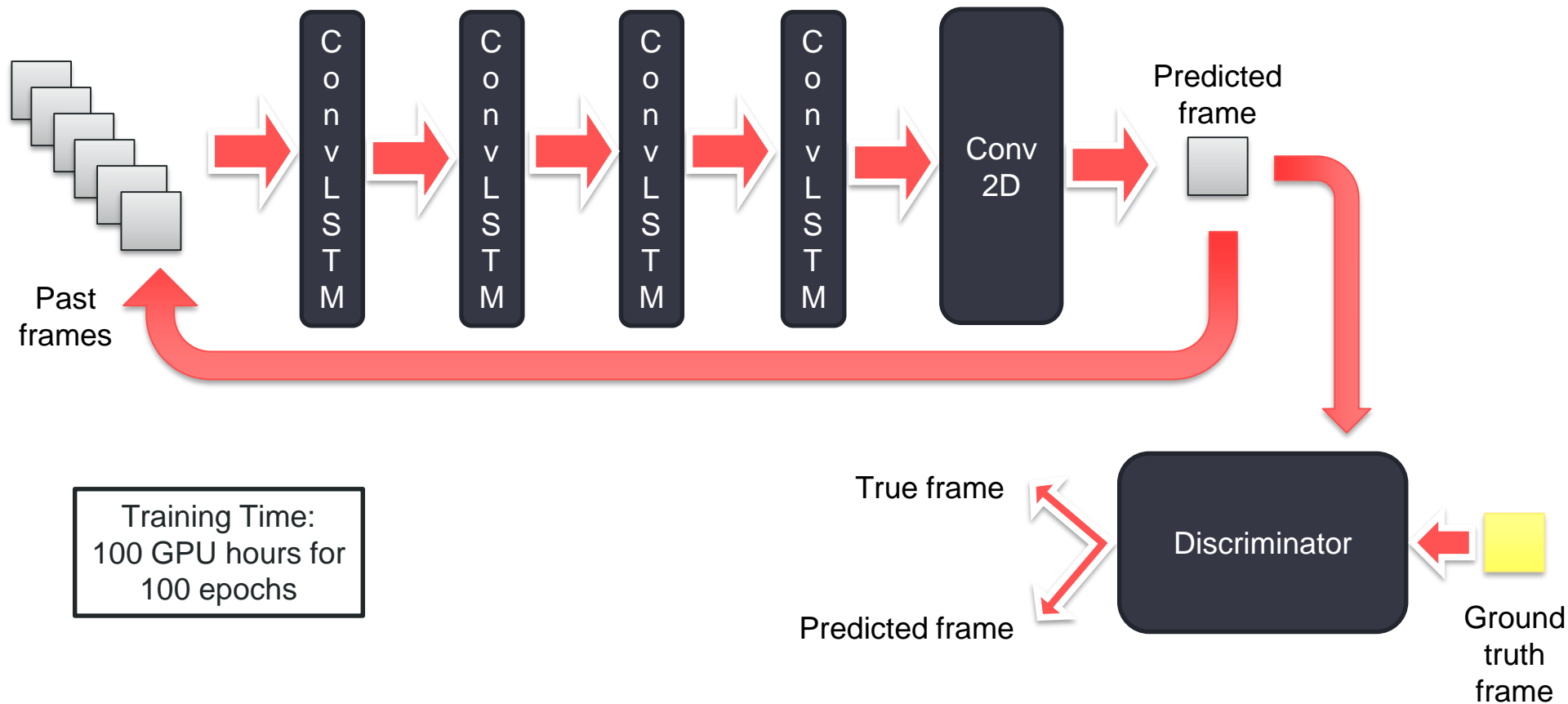
# Background

- Video Frame Prediction is the task of predicting future frames from past frames
- Caters to many applications such as autonomous navigation and self-driving
- Complexities such as occlusions, camera movement, lighting conditions, or clutter make this task difficult for a machine

## Proposed Approaches

- We present a novel Adversarial Spatio-Temporal Convolutional LSTM architecture to predict the future frames
- We evaluate the model on long-term future frame prediction and its performance of the model on out-of-domain inputs

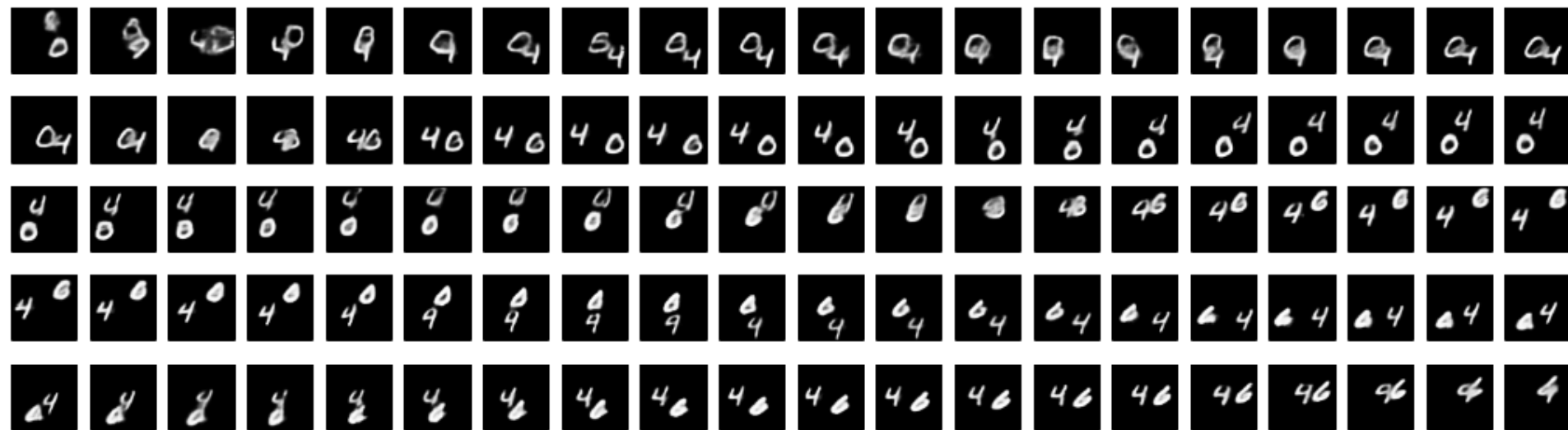
# Technical Details: Architecture



# Experiment 1: Long Term Frame Prediction



Ground truth sequence for the first 20 frames

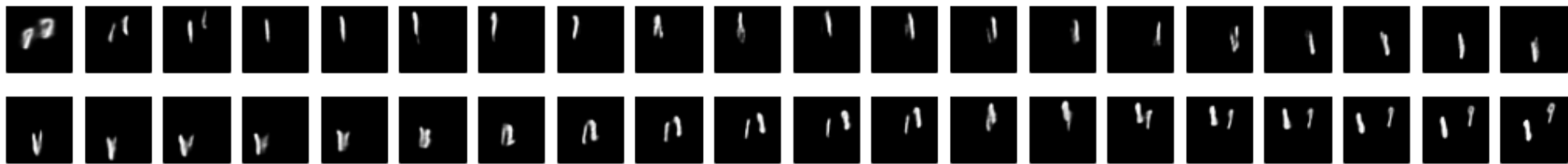


Predicted 100 frames from time instance 2 onwards

# Experiment 2: Out-of-Domain Input



Ground truth sequence for the first 20 frames



Predicted 40 frames from time instance 2 onwards

## Contributions (Novelty)

- We implement the whole architecture from scratch.
- The use of discriminator to improve frame quality in a ConvLSTM architecture

# Results & Conclusion

## Experiment 1

- The model is able to predict the frames until the numbers occlude in the video
- From the time of occlusion, the model chooses one of the multiple future pathways, since occlusion of different two digits could look similar
- Despite the prediction being wrong, we can still observe that the quality of the predicted frame is high with no significant deterioration

## Experiment 2

- We can observe that the model starts to hallucinate two numbers
- Possible explanation: The model could be treating the single input as an overlapping or occluded image of two digits