Adversarial Geometry-Aware Human Motion Prediction

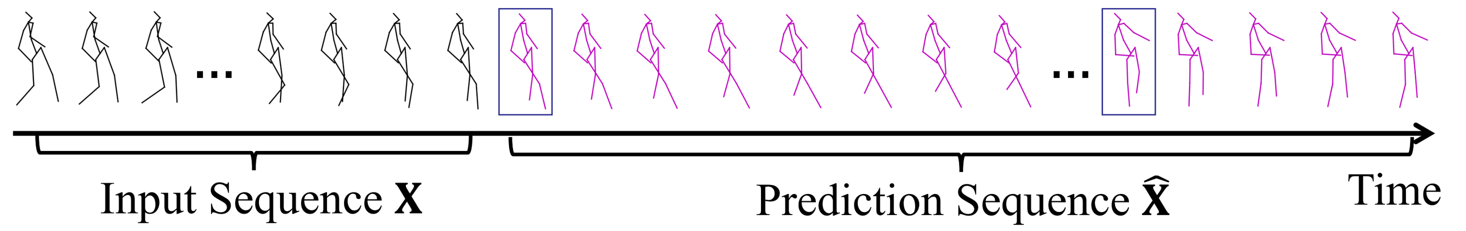
Liang-Yan Gui* Yu-Xiong Wang* Xiaodan Liang José M. F. Moura
Email: {lgui, yuxiongw, xiaodan1, moura}@andrew.cmu.edu

Carnegie Mellon University

ECCV 2018 European Conference on Computer Vision

Motivation

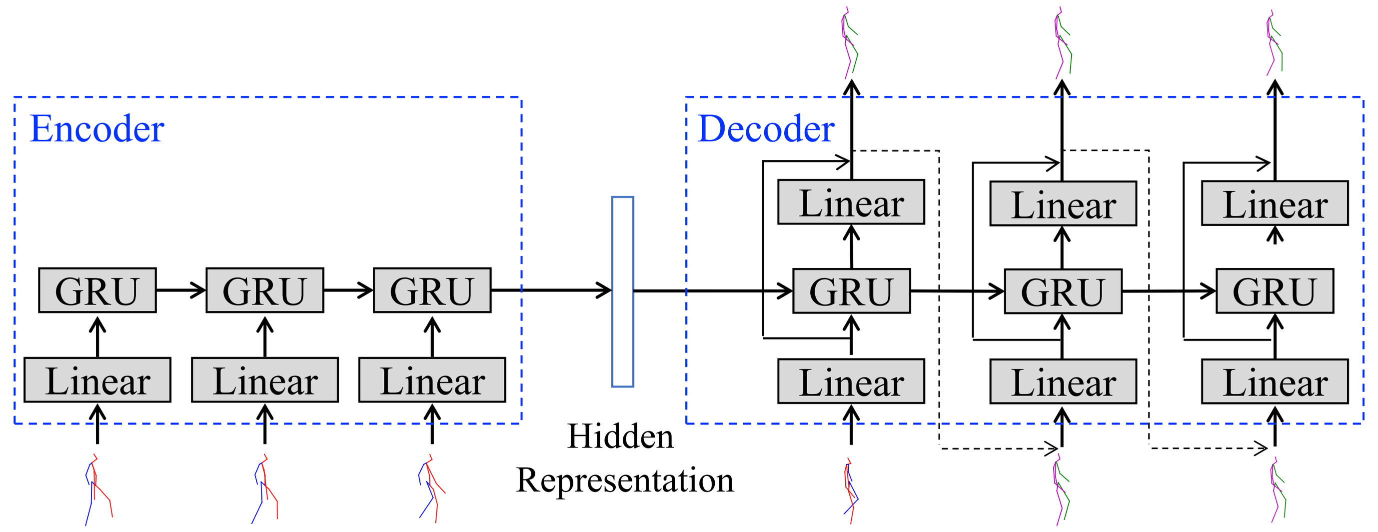
- **Problem statement:** human-like 3D motion prediction



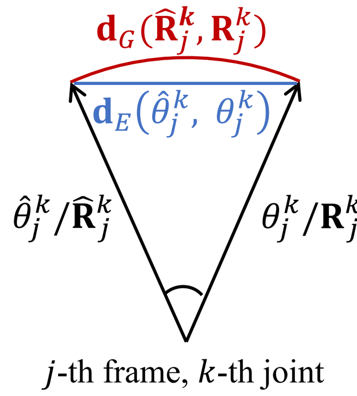
- **Limitation of SOTA encoder-decoder predictor:** discontinuity & unrealism
- **Key insight:** leverage structure of motion sequence
  - *Local* frame-wise geometric structure
  - *Global* sequence-level fidelity and continuity
- **Our approach:** adversarial geometry-aware encoder-decoder (AGED) model
  - **Geometric structure aware loss for 3D motion:** Euclidean → geodesic loss
  - **Adversarial training for validating prediction:** *two* global, complementary *recurrent* discriminators

Geometry-Aware Encoder-Decoder Predictor

- **Predictor $\mathcal{P}$:** sequence-to-sequence recurrent encoder-decoder network



- **Geodesic loss:** regress prediction $\widehat{\mathbf{X}}$ to groundtruth sequence $\mathbf{X}_{\text{gt}}$ frame by frame
  - **Motion frame:** 3D rotations of all joint angles
  - **3D rotation matrices R:** Special Orthogonal Group SO(3) with a Riemannian *manifold* structure
  - **Geodesic distance on the manifold:** angle between two rotation matrices

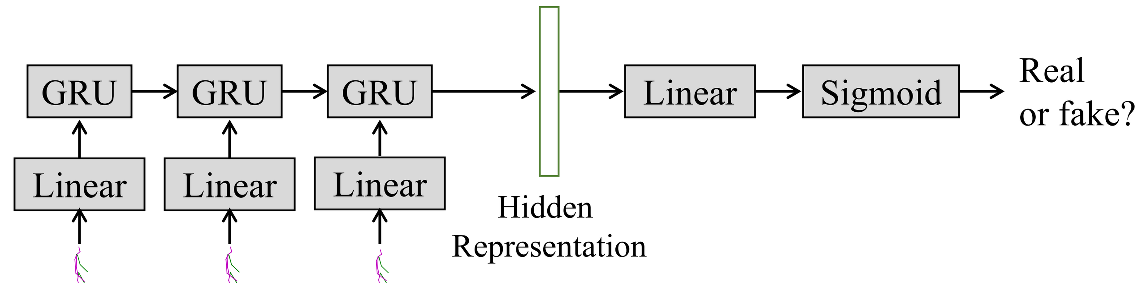


$$\mathbf{d}_G\left(\widehat{\mathbf{R}}_j^k, \mathbf{R}_j^k\right) = \left\|\log\left(\widehat{\mathbf{R}}_j^k \mathbf{R}_j^{k\,T}\right)\right\|_2, \quad \mathcal{L}_{\text{geo}}(\mathcal{P}) = \sum_j \sum_k \mathbf{d}_G\left(\widehat{\mathbf{R}}_j^k, \mathbf{R}_j^k\right)$$

Fidelity and Continuity Discriminators

- **Adversarial training:** predictor as a generator & two sequence-level *recurrent* discriminators
- **An unconditional, fidelity discriminator $\mathcal{D}_f$**
  - Examine the sequence-level plausibility of prediction
  - Distinguish between "short" sequences $\widehat{\mathbf{X}}$ & $\mathbf{X}_{\text{gt}}$

$$\mathcal{L}_{\text{adv}}^f(\mathcal{P}, \mathcal{D}_f) = \mathbb{E}\left[\log\left(\mathcal{D}_f(\mathbf{X}_{\text{gt}})\right)\right] + \mathbb{E}\left[\log\left(1 - \mathcal{D}_f(\mathcal{P}(\mathbf{X}))\right)\right]$$

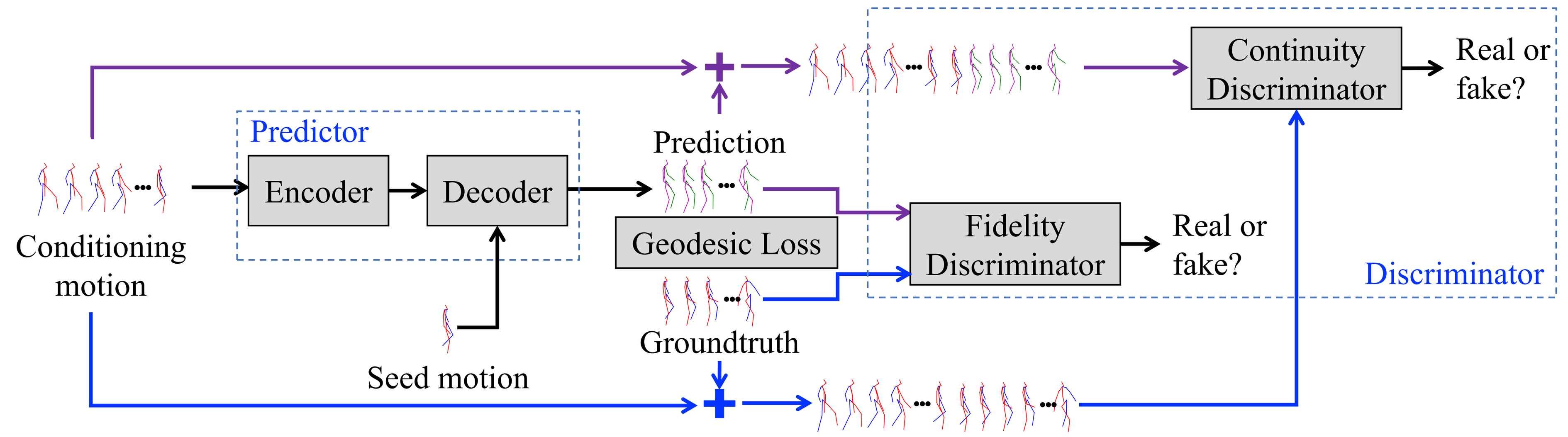


Recurrent discriminator architecture

- **A conditional, continuity discriminator $\mathcal{D}_c$**
  - Check the coherence of prediction with the input sequence
  - Distinguish between "long" concatenated sequences $\{\mathbf{X}, \widehat{\mathbf{X}}\}$ & $\{\mathbf{X}, \mathbf{X}_{\text{gt}}\}$

$$\mathcal{L}_{\text{adv}}^c(\mathcal{P}, \mathcal{D}_c) = \mathbb{E}\left[\log\left(\mathcal{D}_c(\{\mathbf{X}, \mathbf{X}_{\text{gt}}\})\right)\right] + \mathbb{E}\left[\log\left(1 - \mathcal{D}_c(\{\mathbf{X}, \mathcal{P}(\mathbf{X})\})\right)\right]$$

Adversarial Geometry-Aware Encoder-Decoder Model



- **Joint minimax optimization and adversarial training**
  - Integration of geodesic (regression) loss and two adversarial losses

$$\mathcal{P}^* = \arg\min_{\mathcal{P}} \max_{\mathcal{D}_f, \mathcal{D}_c} \lambda\left(\mathcal{L}_{\text{adv}}^f(\mathcal{P}, \mathcal{D}_f) + \mathcal{L}_{\text{adv}}^c(\mathcal{P}, \mathcal{D}_c)\right) + \mathcal{L}_{\text{geo}}(\mathcal{P})$$

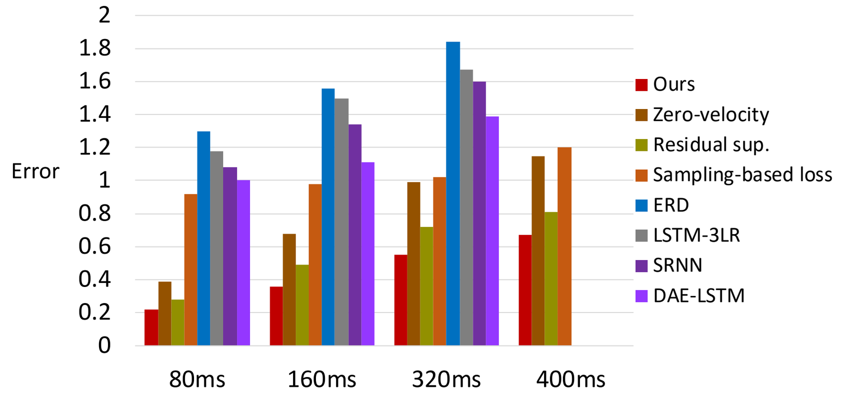
Human Evaluation and User Studies

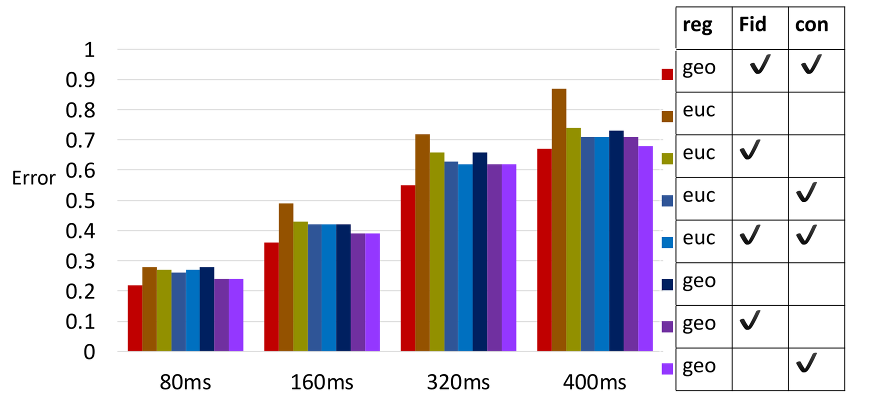
- **A/B testing and success rates:** 25 judges | choose more realistic ones from random pairs of videos

	Short-term				Long-term			
Model pair	Ours	Groundtruth	Sampling-based loss	Residual sup.	Ours	Groundtruth	Sampling-based loss	Residual sup.
Ours vs.	n/a	53.3%	98.6%	69.6%	n/a	48.7%	83.5%	93.1%
Groundtruth vs.	46.7%	n/a	99.7%	75.7%	51.3%	n/a	83.7%	94.9%

Quantitative Evaluation on H3.6M

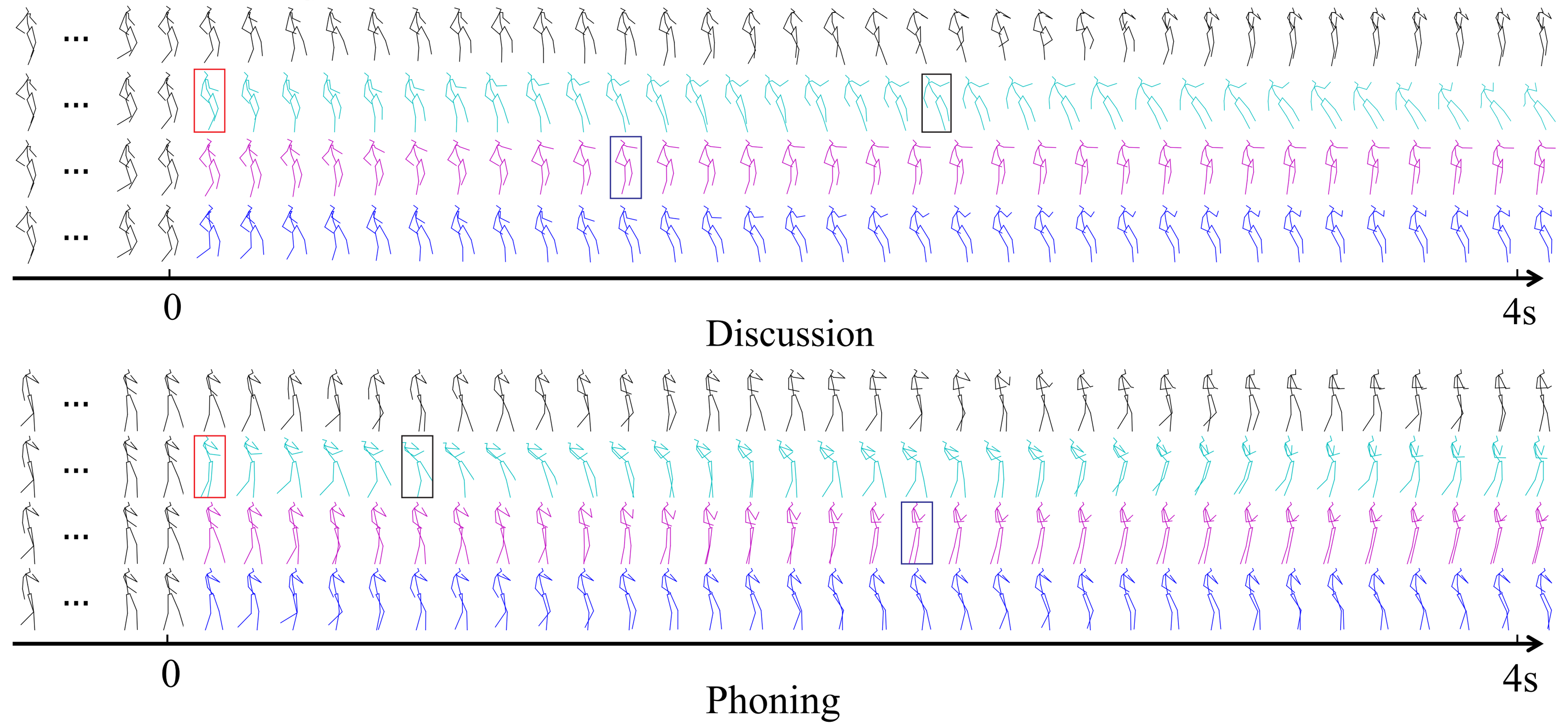
- **Comparison to prior work**
  - Representative short-term results of mean angle error for walking
  - Consistently outperform deep baselines *in all the scenarios*
  - Zero-velocity baseline: general difficulty of motion modeling
- **Unpacking the performance gain**
  - Geodesic loss is superior to Euclidean loss
  - Two discriminators are complementary to each other
  - Full model achieves the best performance





Qualitative Visualization

- **Our short/long-term predictions:** more realistic, continuous, and human-like



Top to bottom rows: groundtruth | sampling-based loss | residuls sup. | ours
Rectangles: Red-jump | Black-drift away to unrealistic motion | Blue-converge to mean pose

Conclusions

- **Geodesic loss:** geometrically meaningful, more precise distance measurement for motion frames
- **Recurrent discriminators:** adversarial learning for motion prediction and synthesis
- **Potential applications** in other motion modeling and analysis tasks