DS-GA.3001 Embodied Learning and Vision

Mengye Ren

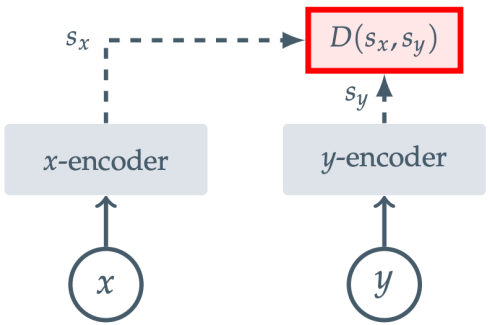
NYU

Spring 2026

elvcourse.org

Joint Embedding Predictive Architecture (JEPA)

- Predict what changes from one view to another.



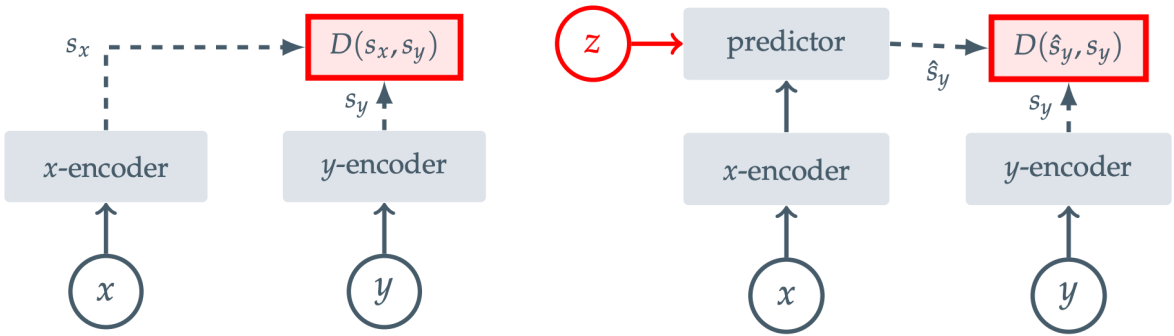
JEA

Assran et al. Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture. ICCV 2023.

NYU

Joint Embedding Predictive Architecture (JEPA)

- Predict what changes from one view to another.

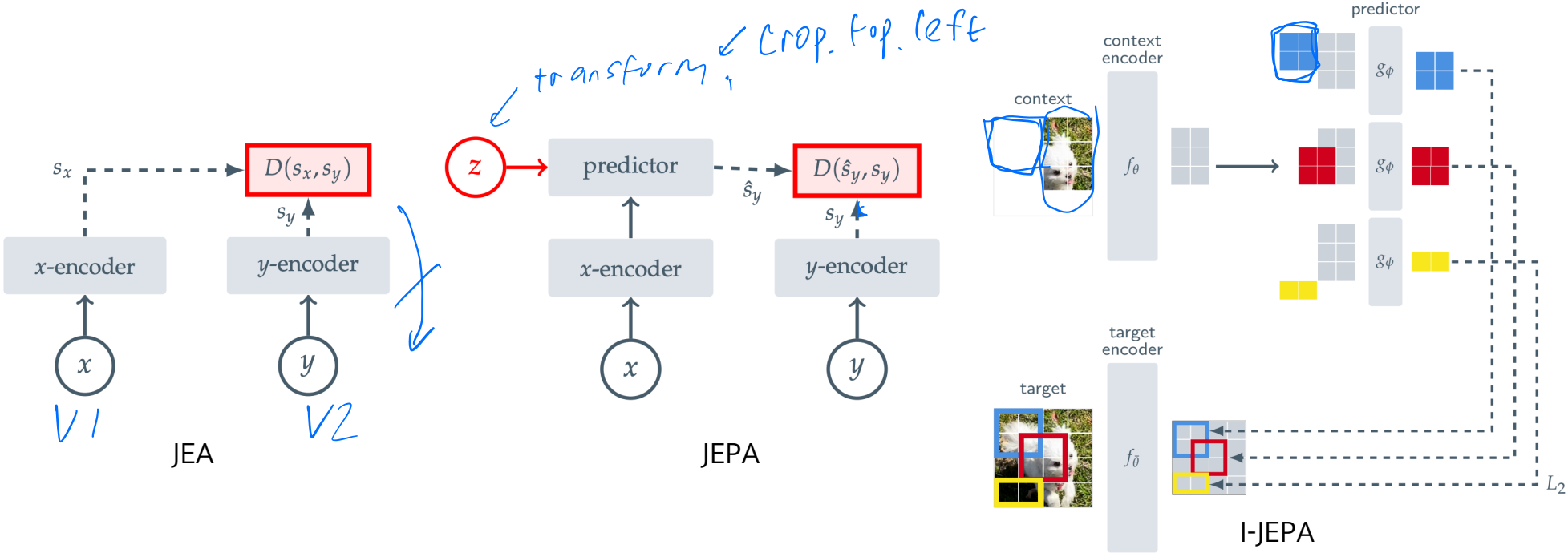


Assran et al. Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture. ICCV 2023.

NYU

Joint Embedding Predictive Architecture (JEPA)

- Predict what changes from one view to another.



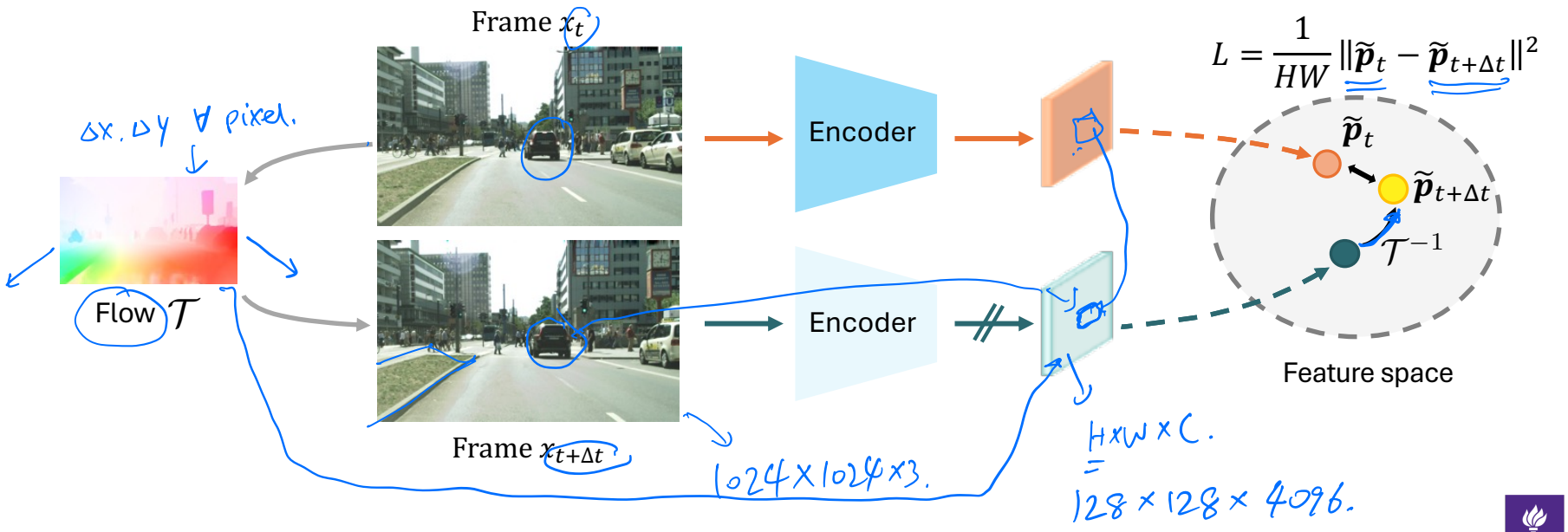
Assran et al. Self-Supervised Learning from Images with a Joint-Embedding Predictive Architecture. ICCV 2023.

NYU

Joint Embedding Prediction with Dense Motion

- Can we use adjacent video frames as self-supervision?
- Objects move densely throughout the image.

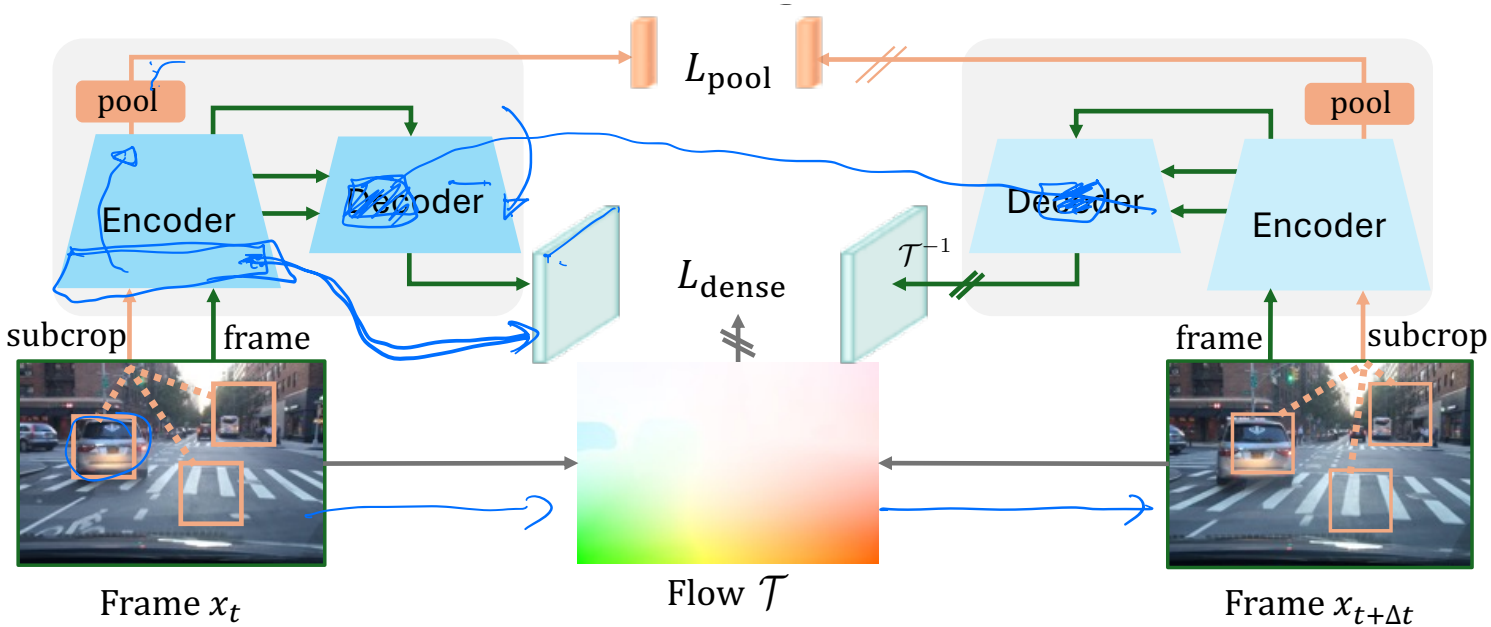
motion equivariance.



NYU

Joint Embedding Prediction with Dense Motion

- Perform SSL in multiple scales (small objects vs. big regions).

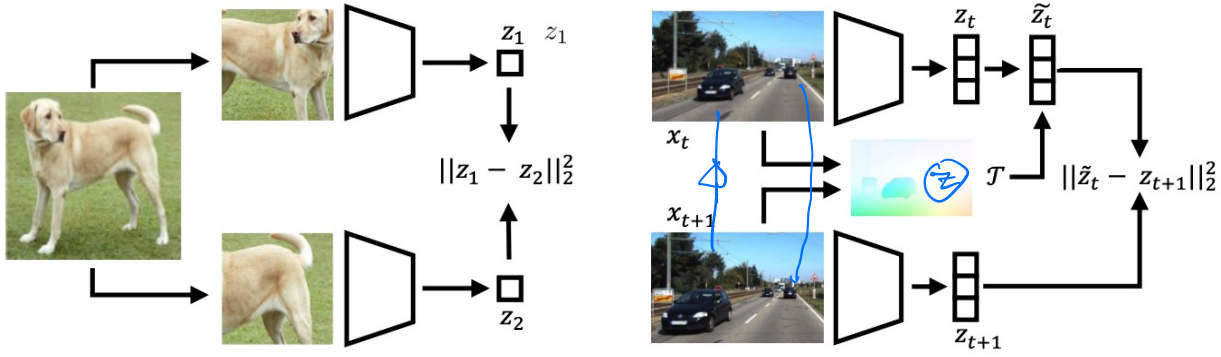


Wang et al. PooDLe: Pooled and Dense Self-Supervised Learning from Naturalistic Videos. ICLR 2025.

NYU

Joint Motion and Semantic Learning

- We either create or assume some external motion / action information to send to predictors.
- Additional motion needs to be inferred.



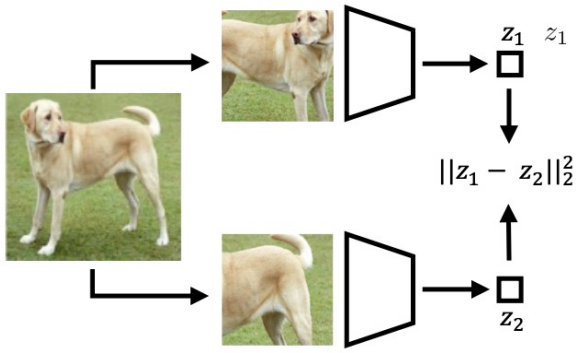
JEA

JEPA w/ dense motion

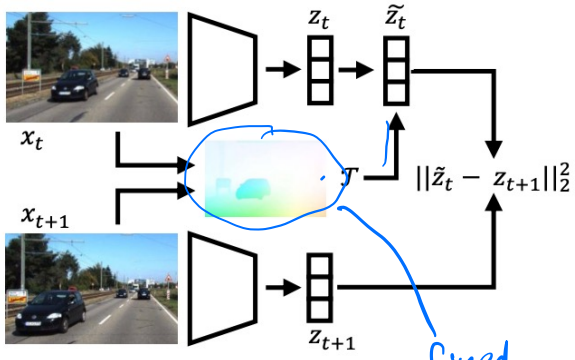
Hoang et al. Midway Network: Learning Representations for Recognition and Motion from Latent Dynamics.

Joint Motion and Semantic Learning

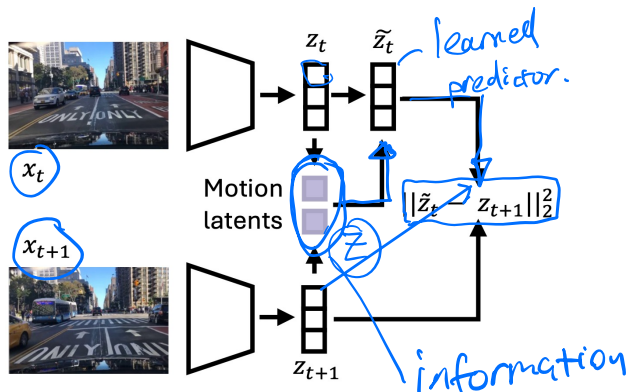
- We either create or assume some external motion / action information to send to predictors.
- Additional motion needs to be inferred.



JEA



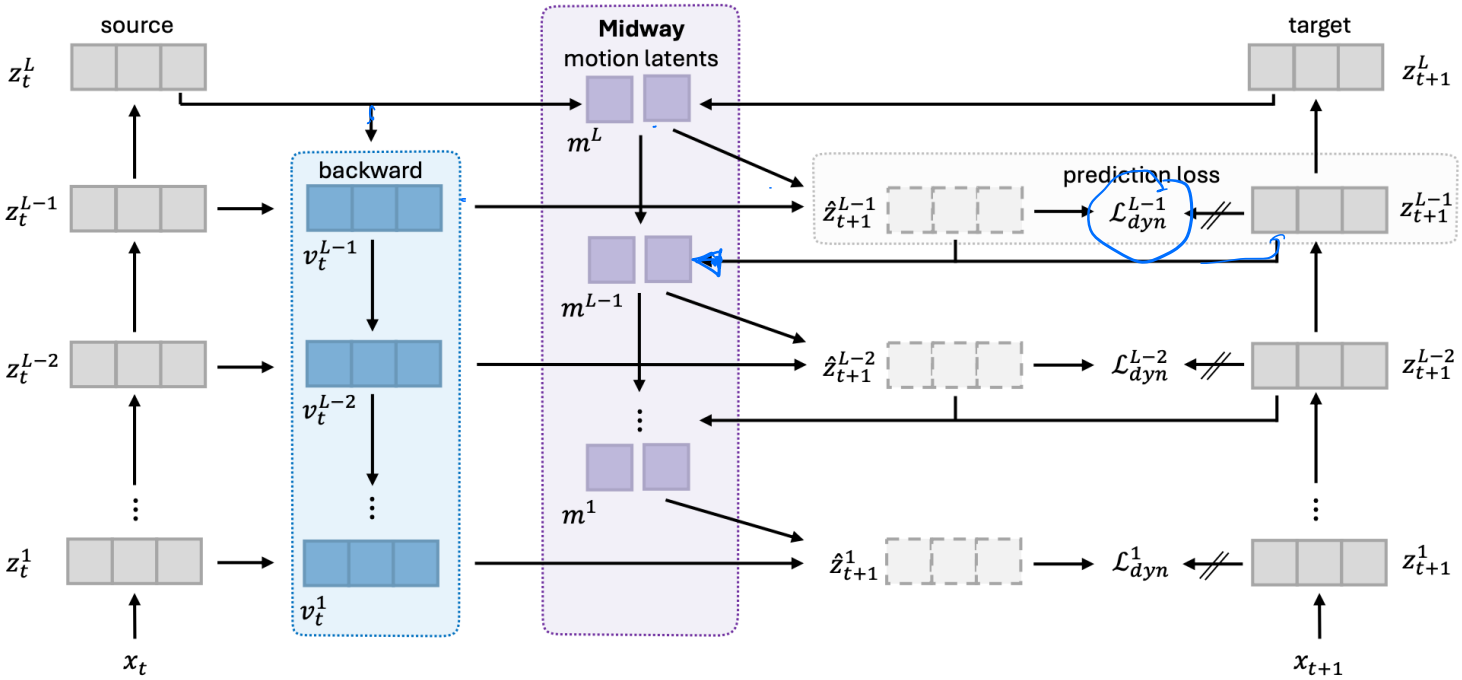
JEPA w/ dense motion



JEPA w/ inferred motion

Midway Networks

- Using backward refinement from Optical Flow networks.



Hoang et al. Midway Network: Learning Representations for Recognition and Motion from Latent Dynamics. ICLR 2026.

NYU

Midway Networks

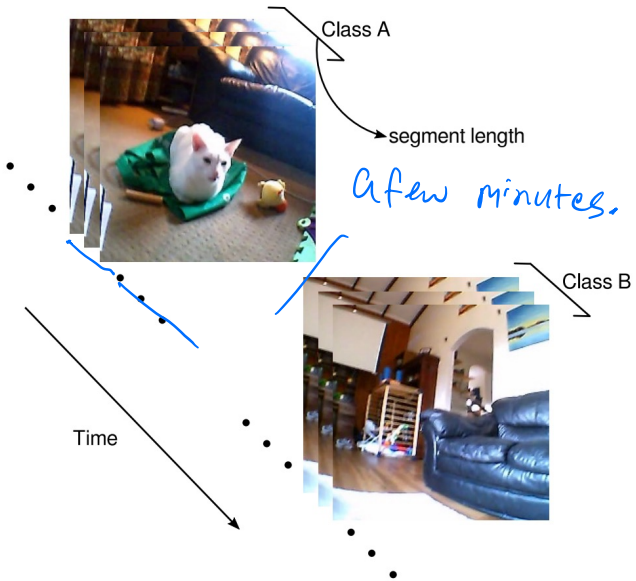
- Use perturbation to visualize the motion learned from motion tokens.

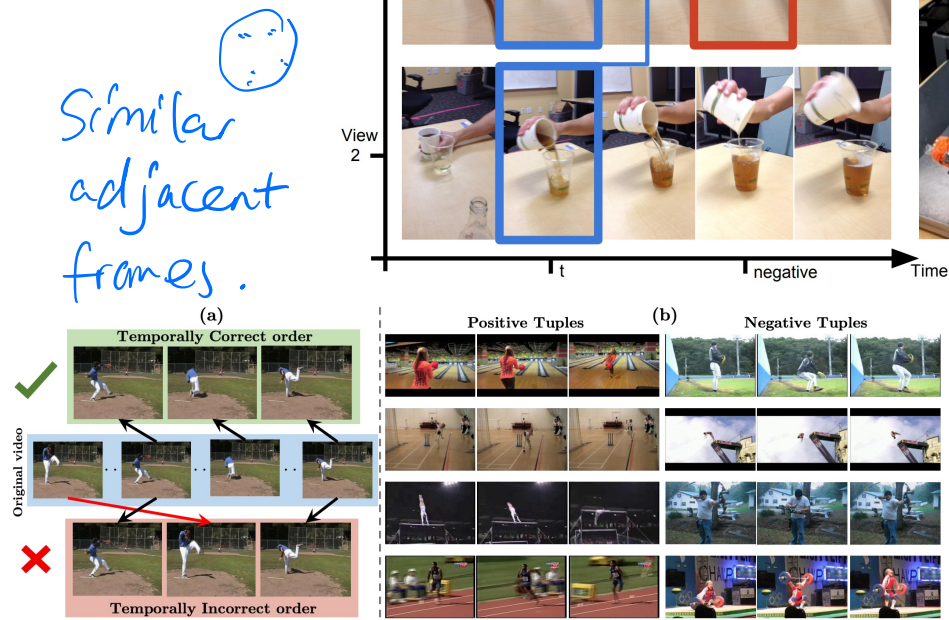
Hoang et al. Midway Network: Learning Representations for Recognition and Motion from Latent Dynamics. ICLR 2026.

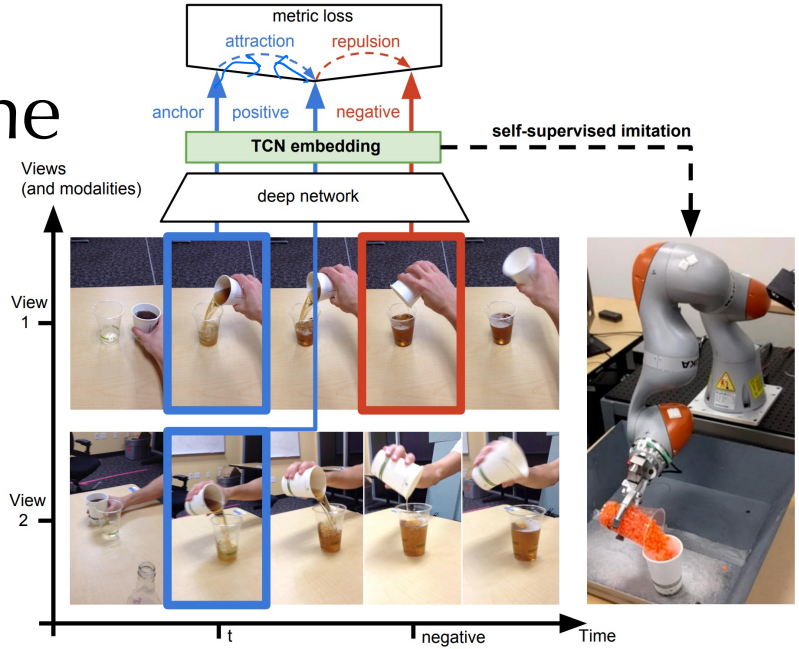
NYU

SSL with Time

- Use time as an additional source of supervision.





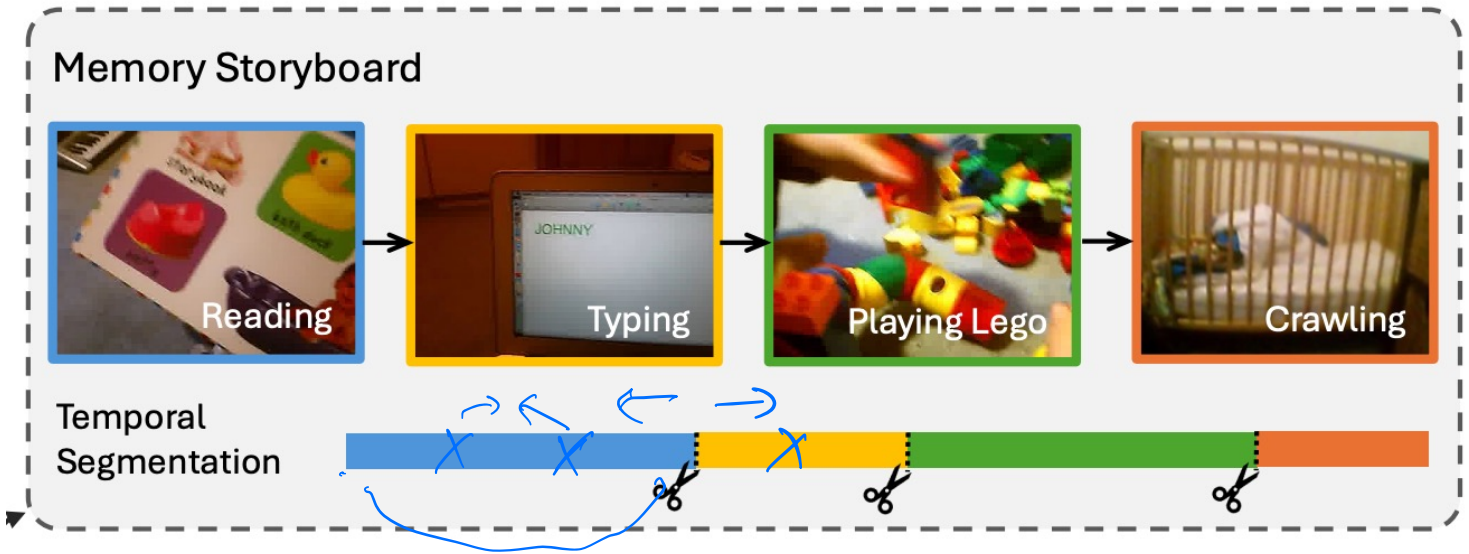


Misra et al. Shuffle and Learn: Unsupervised Learning using Temporal Order Verification. ECCV 2016.
Sermanet et al. Time-Contrastive Networks: Self-Supervised Learning from Video. ICRA 2018.
Orhan et al. Self-Supervised Learning through the Eyes of a Child. NeurIPS 2020.

NYU

SSL with Time

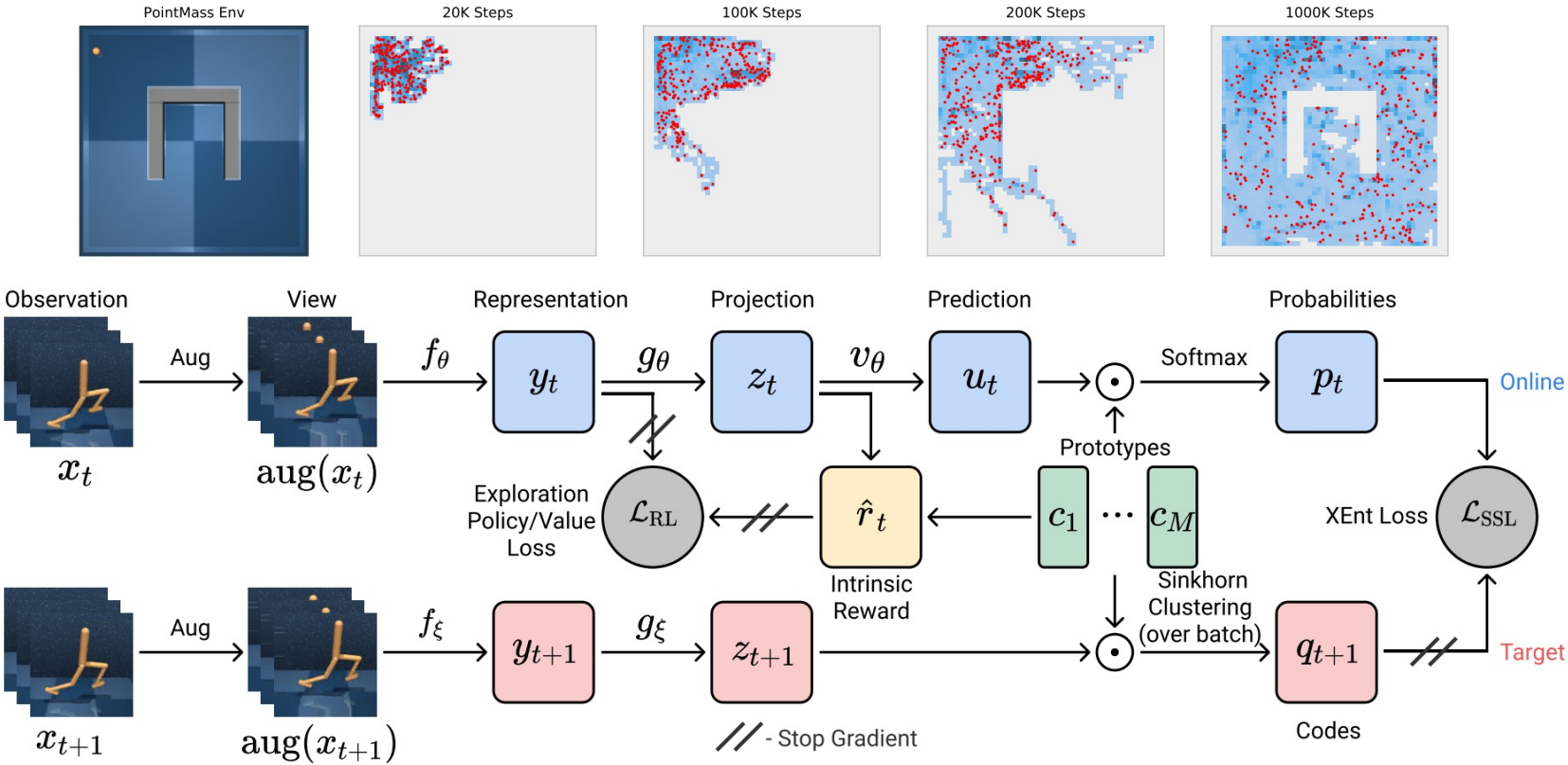
- We can segment videos into meaningful events.
- Leverage the spatiotemporal continuity structure.



Yang & Ren. Memory Storyboard: Leveraging Temporal Segmentation for Streaming Self-Supervised Learning from Egocentric Videos. CoLLAs 2025.

NYU

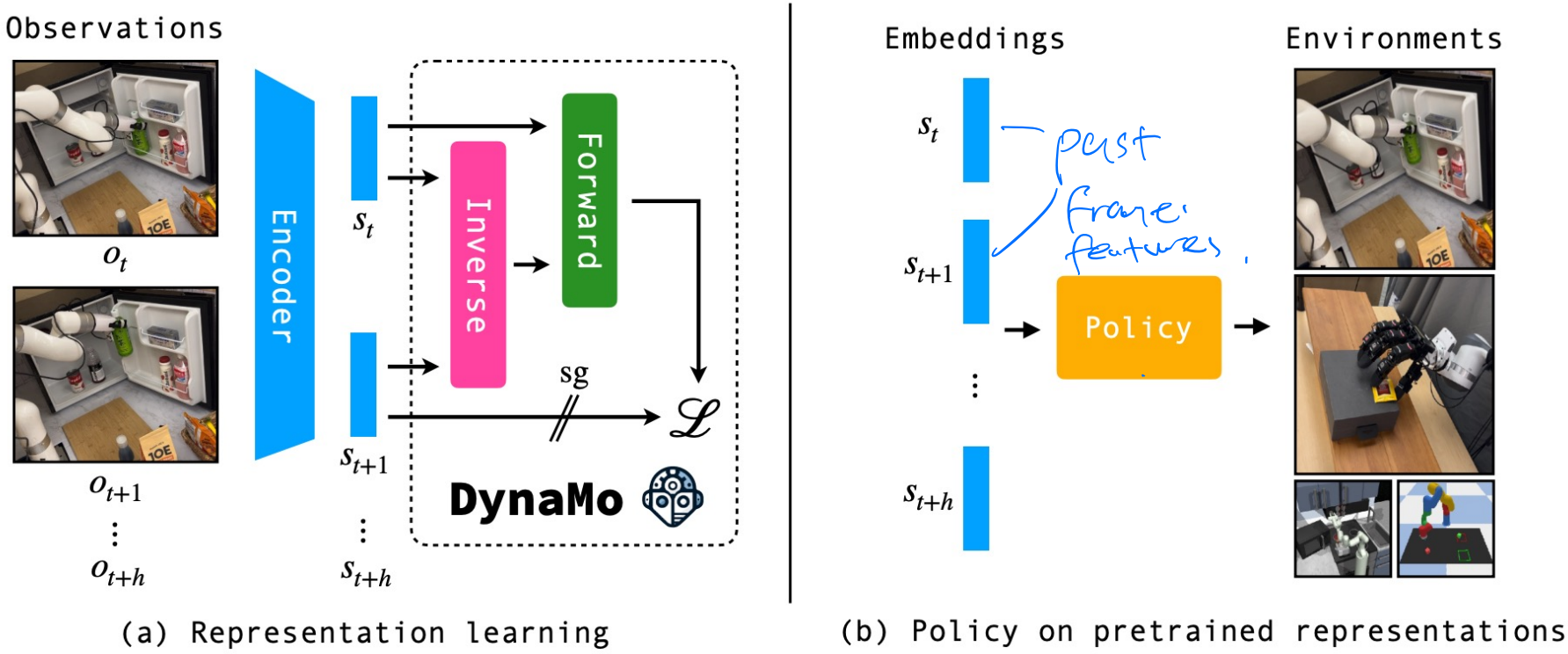
SSL for Visual Control



Yarats. Reinforcement Learning with Prototypical Representations. ICML 2021.

NYU

SSL for Visual Control

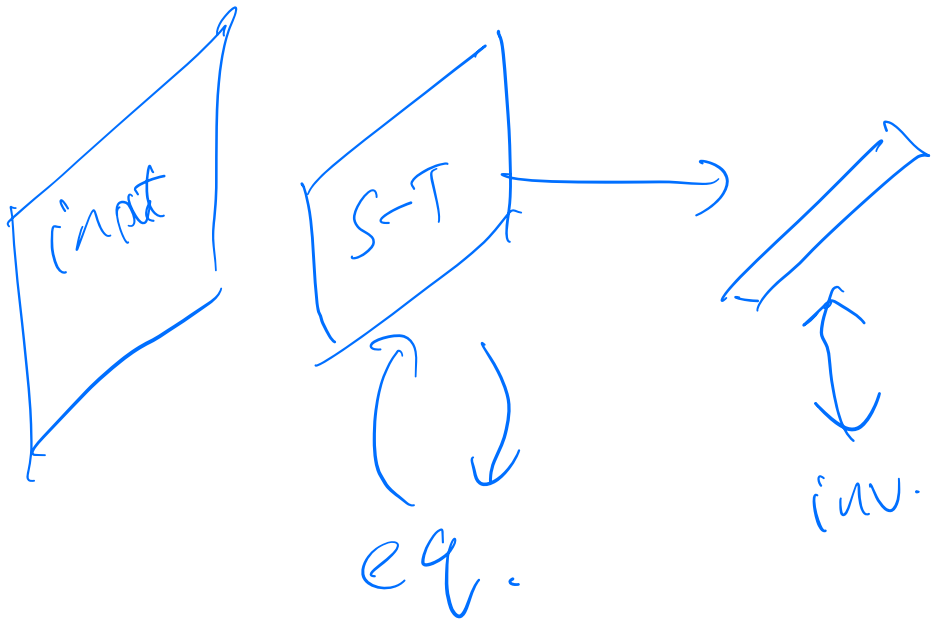


Cui et al. DynaMo: In-Domain Dynamics Pretraining for Visuo-Motor Control. NeurIPS 2024.

NYU

Summary

- Representation learning leverage the information in unlabeled data.
- Inductive biases matter
  - spatial-time invariance
  - motion equivariance



NYU

Summary

- Representation learning leverage the information in unlabeled data.
- Inductive biases matter
  - spatial-time invariance
  - motion equivariance
- A foundation for sensorimotor learning – pretrained representations as a starter for policy learning.

NYU

Emergent Attention, Object Discovery

- SSL representations show awareness of object classes and instance identities.

NYU

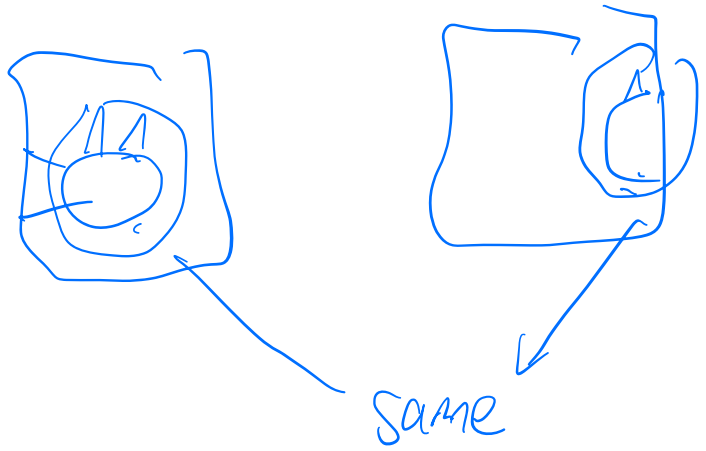
Emergent Attention, Object Discovery

- SSL representations show awareness of object classes and instance identities.
- Why does attention show awareness of objects?

NYU

Emergent Attention, Object Discovery

- SSL representations show awareness of object classes and instance identities.
- Why does attention show awareness of objects?
- The network is encouraged to associate different parts of the objects together in order to identify whether two inputs belong to the same image or not.



NYU

Emergent Attention, Object Discovery

- SSL representations show awareness of object classes and instance identities.
- Why does attention show awareness of objects?
- The network is encouraged to associate different parts of the objects together in order to identify whether two inputs belong to the same image or not.
- Attending to semantically similar parts facilitates the process.

NYU

Emergent Attention, Object Discovery

- SSL representations show awareness of object classes and instance identities.
- Why does attention show awareness of objects?
- The network is encouraged to associate different parts of the objects together in order to identify whether two inputs belong to the same image or not.
- Attending to semantically similar parts facilitates the process.
- The network is a hierarchical information processing pipeline – Lower layers integrate more granular and smaller neighborhood.

Weak-to-Strong Supervision

- General idea: Use self-supervised learning to learn good features, which allow us to generate low-quality masks.

NYU

Weak-to-Strong Supervision

- General idea: Use self-supervised learning to learn good features, which allow us to generate low-quality masks.
- Then use these masks as pseudo labels and supervise the network to predict these low-quality masks.

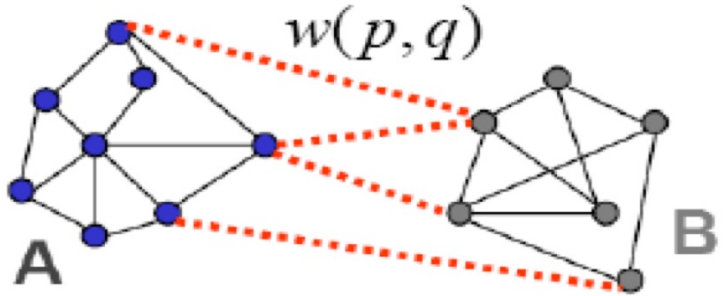
NYU

Weak-to-Strong Supervision

- General idea: Use self-supervised learning to learn good features, which allow us to generate low-quality masks.
- Then use these masks as pseudo labels and supervise the network to predict these low-quality masks.
- Question: how do we come up with masks? What loss is used to supervise the network?

Graph Cut

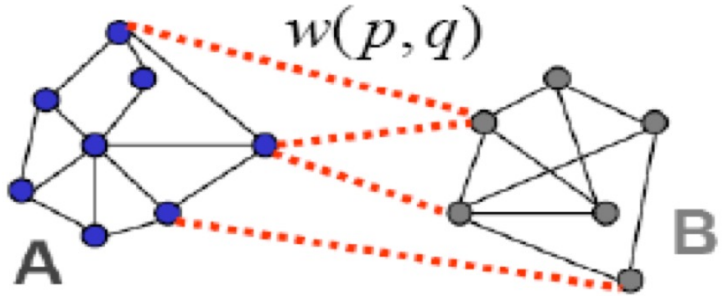
- Segmentation is essentially a clustering problem.



$$cut\,(A, B) = \sum_{p \in A, q \in B} w(p, q)$$

Graph Cut

- Segmentation is essentially a clustering problem.
- We can transform the clustering problem with the graph cut problem.

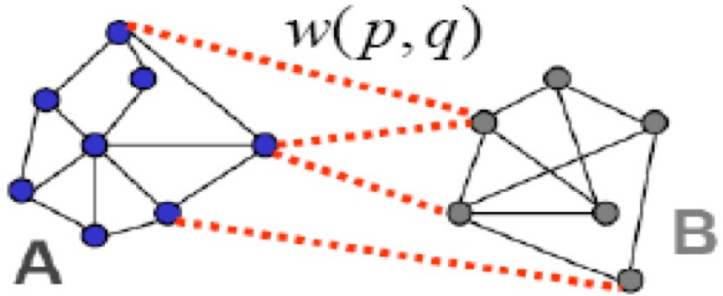


$$cut\,(A, B) = \sum_{p \in A, q \in B} w(p, q)$$

NYU

Graph Cut

- Segmentation is essentially a clustering problem.
- We can transform the clustering problem with the graph cut problem.
- Pixel = node.

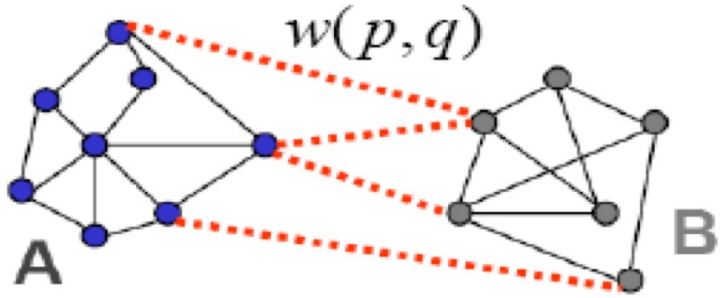


$$cut\,(A, B) = \sum_{p \in A, q \in B} w(p, q)$$

NYU

Graph Cut

- Segmentation is essentially a clustering problem.
- We can transform the clustering problem with the graph cut problem.
- Pixel = node.
- Affinity between the two pixels = edge value (flow).

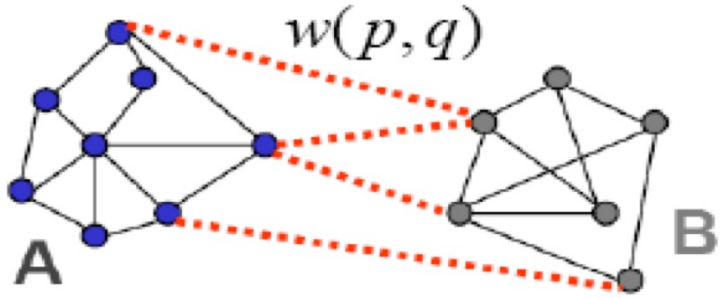


$$cut\,(A, B) = \sum_{p \in A, q \in B} w(p, q)$$

NYU

Graph Cut

- Segmentation is essentially a clustering problem.
- We can transform the clustering problem with the graph cut problem.
- Pixel = node.
- Affinity between the two pixels = edge value (flow).
- Objective: Cut the graph into disconnected components with a minimum sum of edge values.

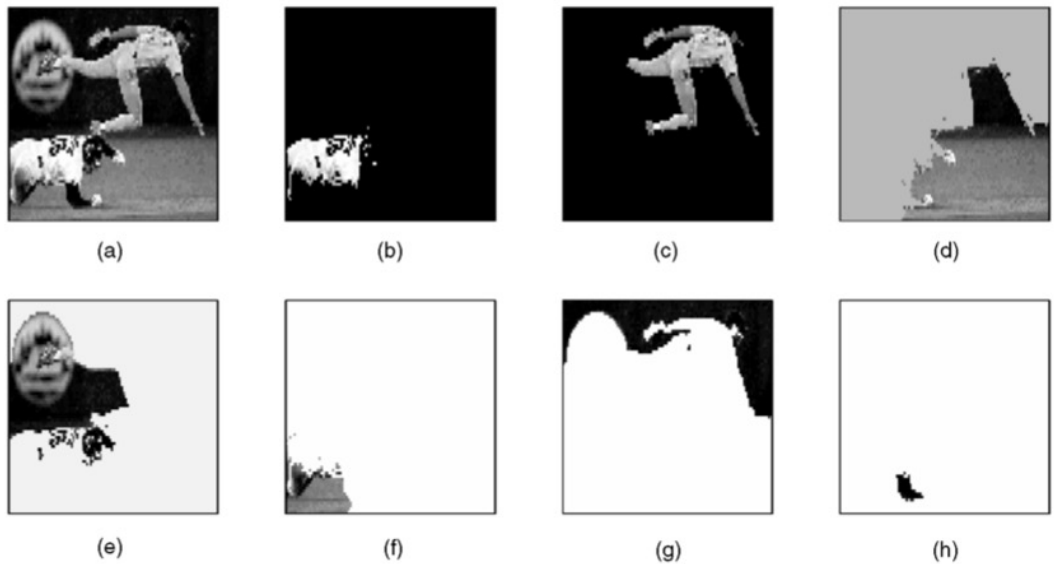


$$cut\,(A, B) = \sum_{p \in A, q \in B} w(p, q)$$

NYU

NCut

- Sort the eigenvectors from the smallest to the largest.

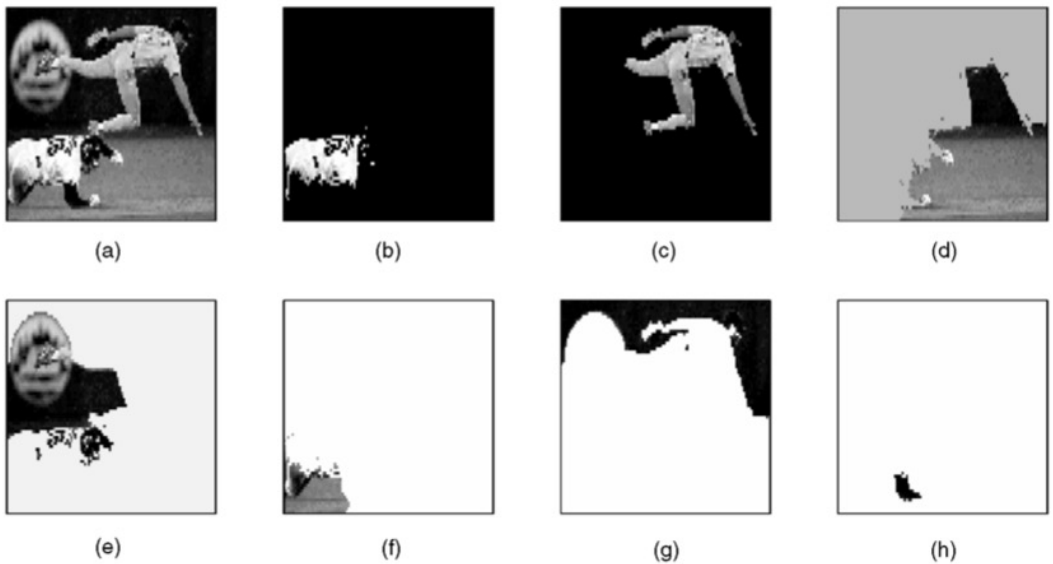


Shi and Malik. Normalized Cuts and Image Segmentation. TPAMI 2000.

NYU

NCut

- Sort the eigenvectors from the smallest to the largest.
- This was a classic image segmentation technique operating directly on image intensity.

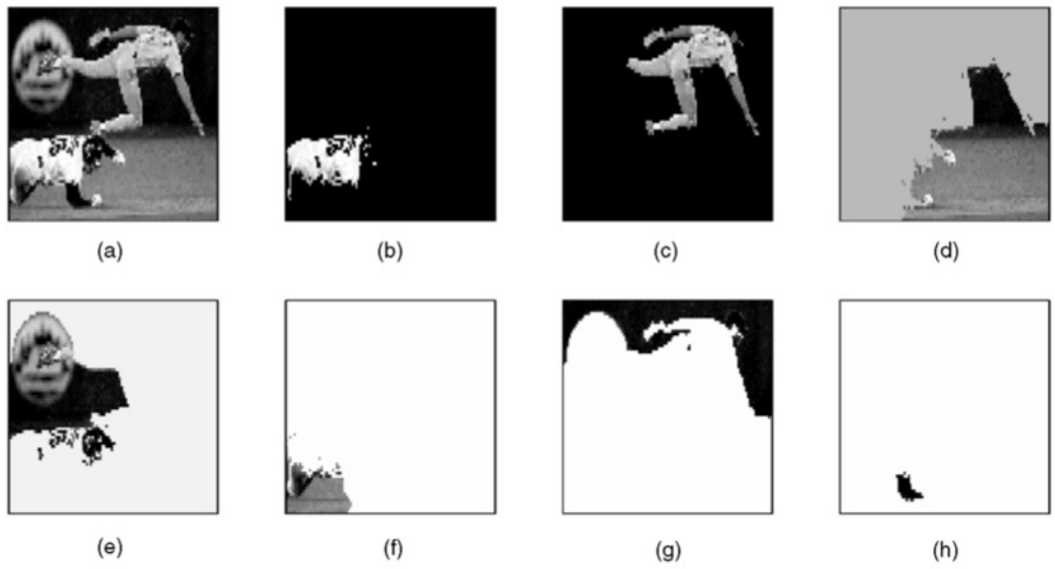


Shi and Malik. Normalized Cuts and Image Segmentation. TPAMI 2000.

NYU

NCut

- Sort the eigenvectors from the smallest to the largest.
- This was a classic image segmentation technique operating directly on image intensity.
- Now, instead of segmenting pixels, we can directly segment semantically meaningful representations from self-supervision.

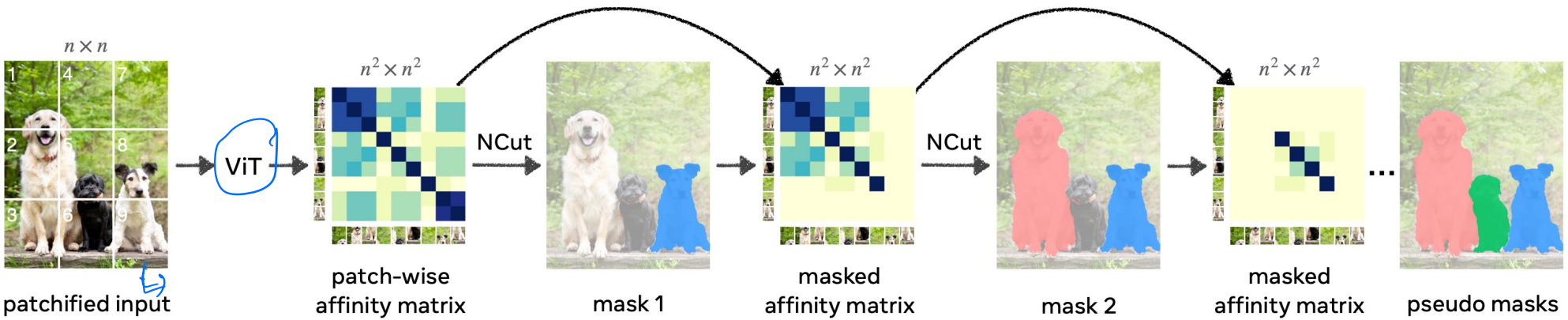


Shi and Malik. Normalized Cuts and Image Segmentation. TPAMI 2000.

NYU

MaskCut

- Use a pretrained DINO ViT network.

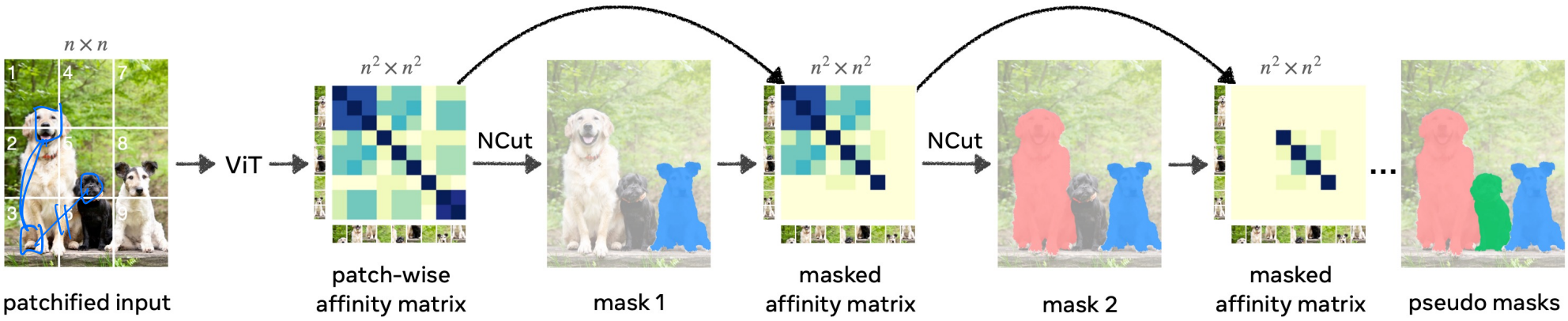


Wang et al. Self-supervised transformers for unsupervised object discovery using normalized cut. CVPR 2022.
Wang et al. Cut and Learn for Unsupervised Object Detection and Instance Segmentation. CVPR 2023.

NYU

MaskCut

- Use a pretrained DINO ViT network.
- Use the "key" features from the last attention layer: $W_{ij} = \frac{K_i K_j}{\|K_i\|_2 \|K_j\|_2}$



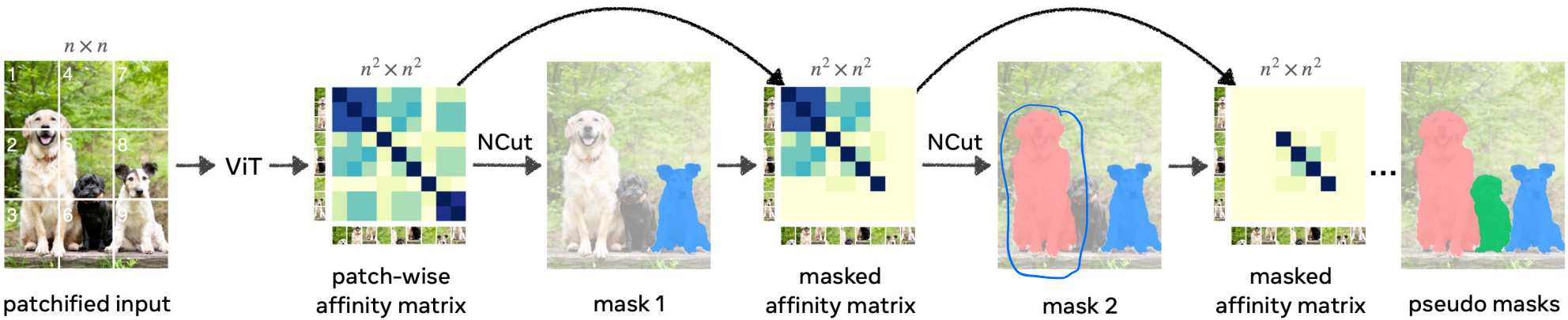
Wang et al. Self-supervised transformers for unsupervised object discovery using normalized cut. CVPR 2022.
Wang et al. Cut and Learn for Unsupervised Object Detection and Instance Segmentation. CVPR 2023.

NYU

MaskCut

- Use a pretrained DINO ViT network.
- Use the "key" features from the last attention layer: $W_{ij} = \frac{K_i K_j}{\|K_i\|_2 \|K_j\|_2}$
- Iterative NCut on the pairwise matrix by masking out the regions from previous stages.

Graph cut.

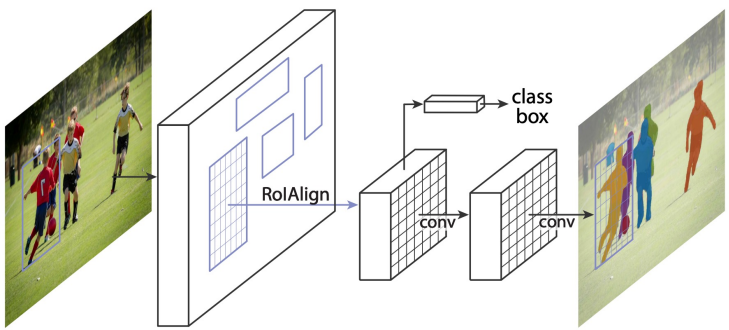


Wang et al. Self-supervised transformers for unsupervised object discovery using normalized cut. CVPR 2022.
Wang et al. Cut and Learn for Unsupervised Object Detection and Instance Segmentation. CVPR 2023.

NYU

Iterative Self-Training

- Now add a MaskRCNN structure on top of the pretrained network.

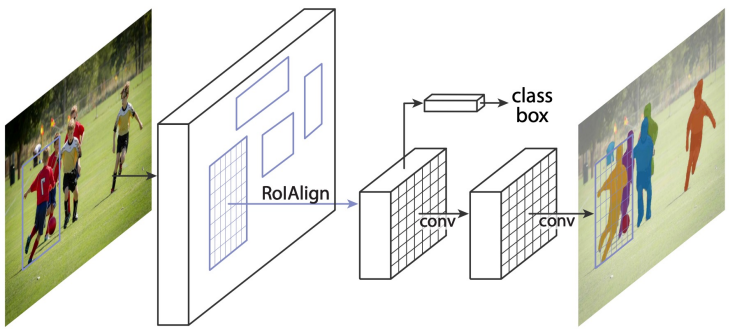


Wang et al. Cut and Learn for Unsupervised Object Detection and Instance Segmentation. CVPR 2023.

NYU

Iterative Self-Training

- Now add a MaskRCNN structure on top of the pretrained network.
- Select the predictions with the highest confidence score and use them as labels.

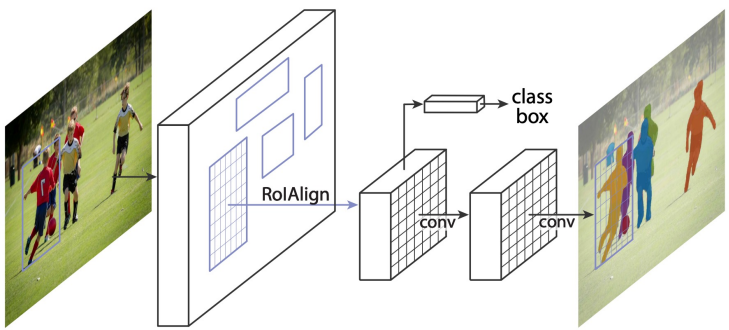


Wang et al. Cut and Learn for Unsupervised Object Detection and Instance Segmentation. CVPR 2023.

NYU

Iterative Self-Training

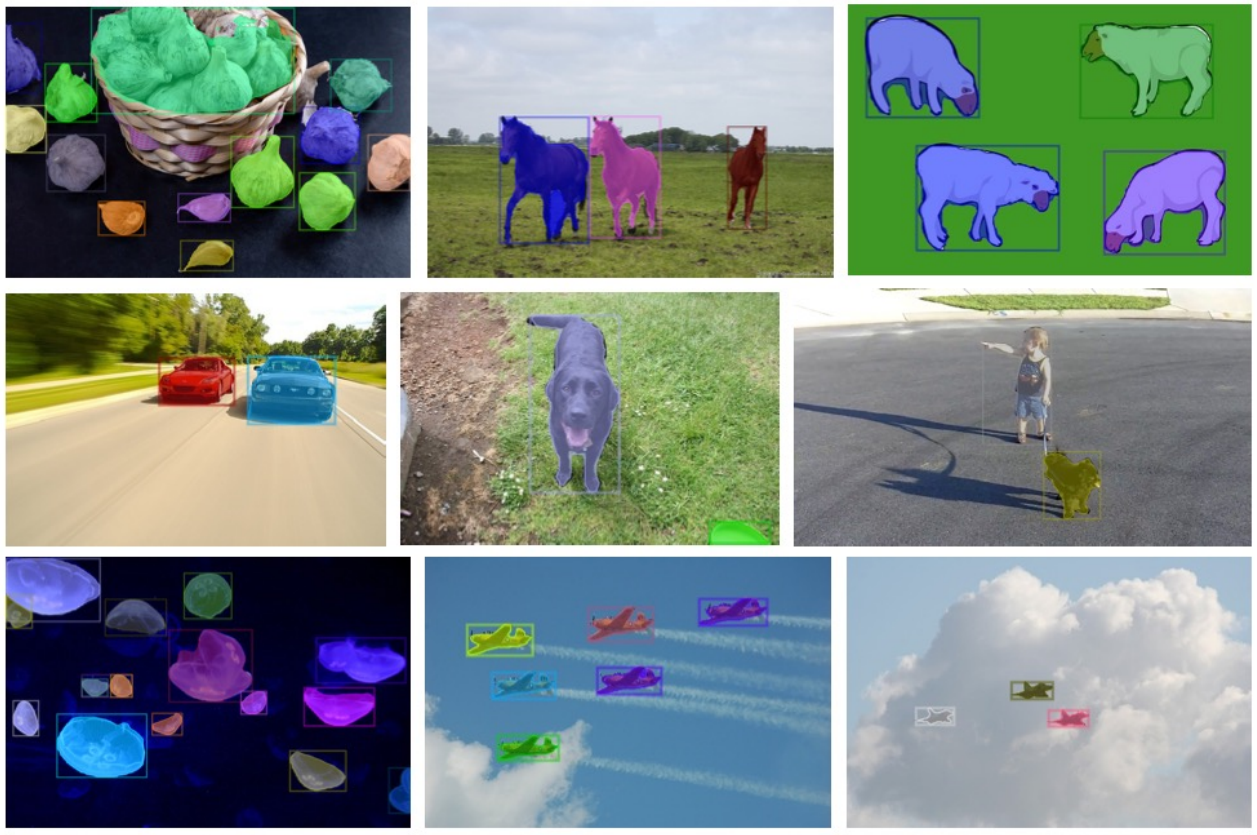
- Now add a MaskRCNN structure on top of the pretrained network.
- Select the predictions with the highest confidence score and use them as labels.
- Neural networks can learn from the noisy labels and output smoother predictions.



Wang et al. Cut and Learn for Unsupervised Object Detection and Instance Segmentation. CVPR 2023.

NYU

More Visualization

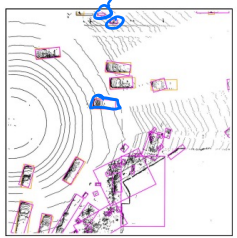


Wang et al. Cut and Learn for Unsupervised Object Detection and Instance Segmentation. CVPR 2023.

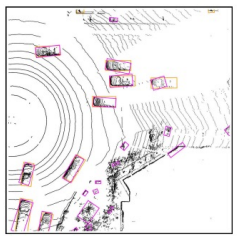
NYU

Pseudo Labels in 3D

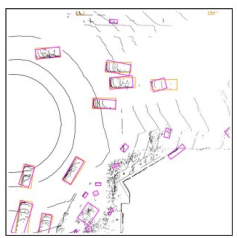
1 Point clustering pseudo-labels



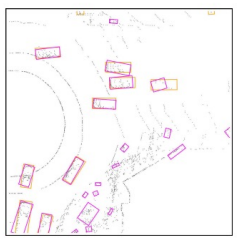
2 Filter out temporally inconsistent tracks



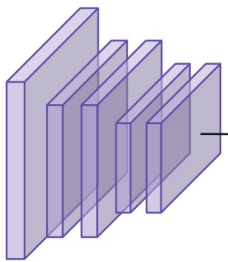
3 Randomly drop lidar beams



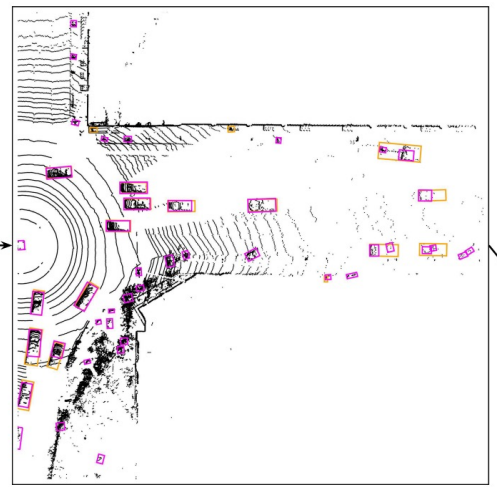
4 Randomly drop spherical rows/cols



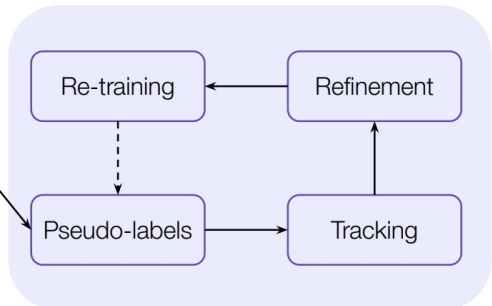
5 Train CNN in short range



6 Zero-shot generalization to long-range



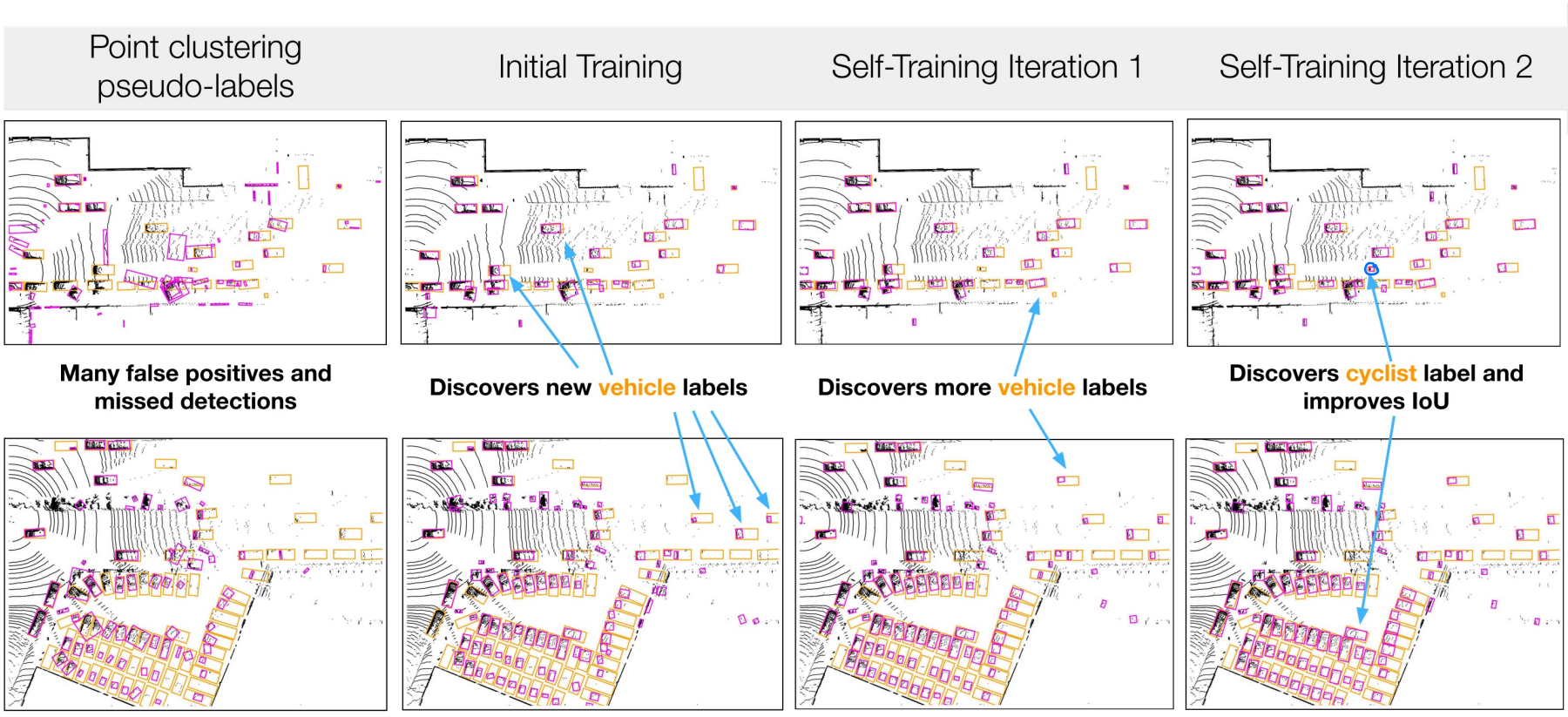
7 Self-training in long-range



Zhang et al. Towards Unsupervised Object Detection from LiDAR Point Clouds. CVPR 2023.

NYU

Iterative Refinement of Pseudo Labels

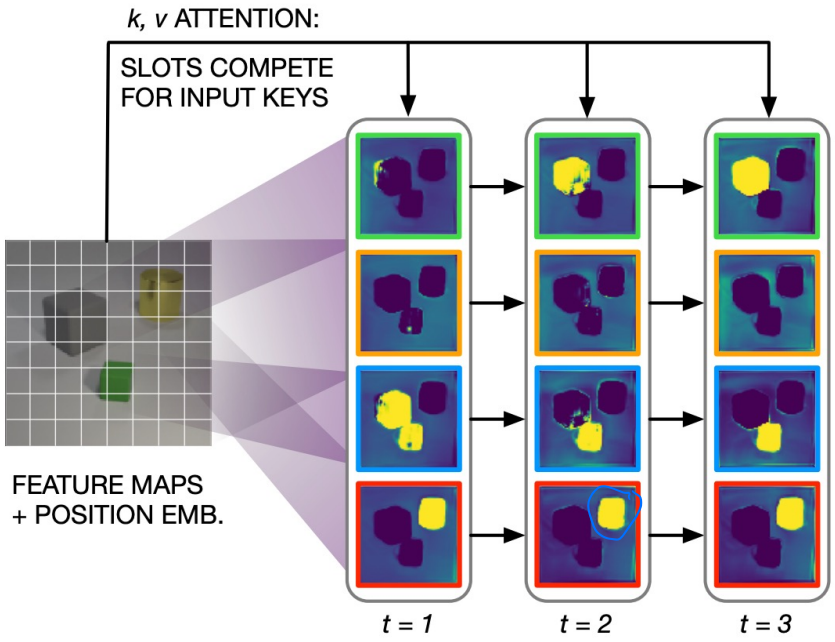


Zhang et al. Towards Unsupervised Object Detection from LiDAR Point Clouds. CVPR 2023.

NYU

Slot Attention Networks

- Can we learn clustering as an end-to-end operation?



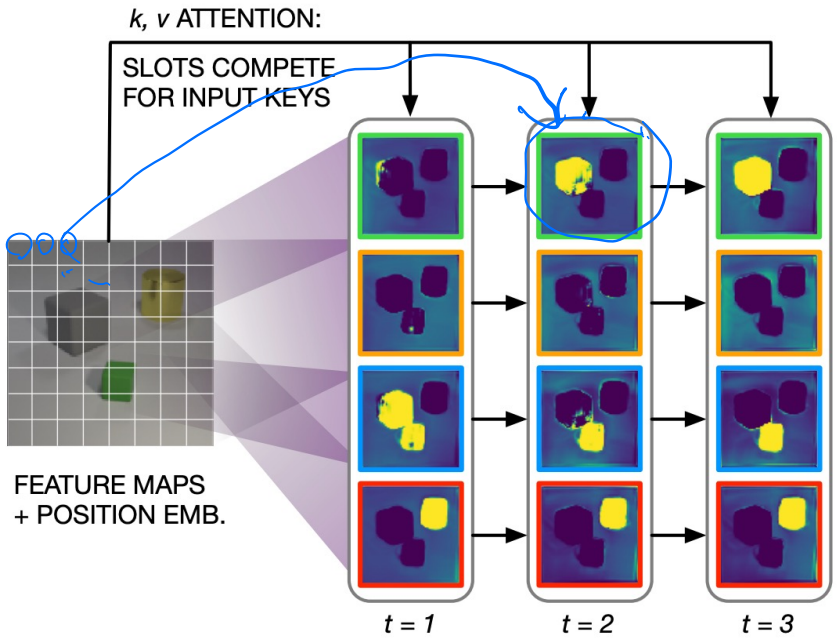
(a) Slot Attention module.

Locatello et al. Object-Centric Learning with Slot Attention. NeurIPS 2020.

NYU

Slot Attention Networks

- Can we learn clustering as an end-to-end operation?
- Slot attention is inspired by the success of the attention mechanism.



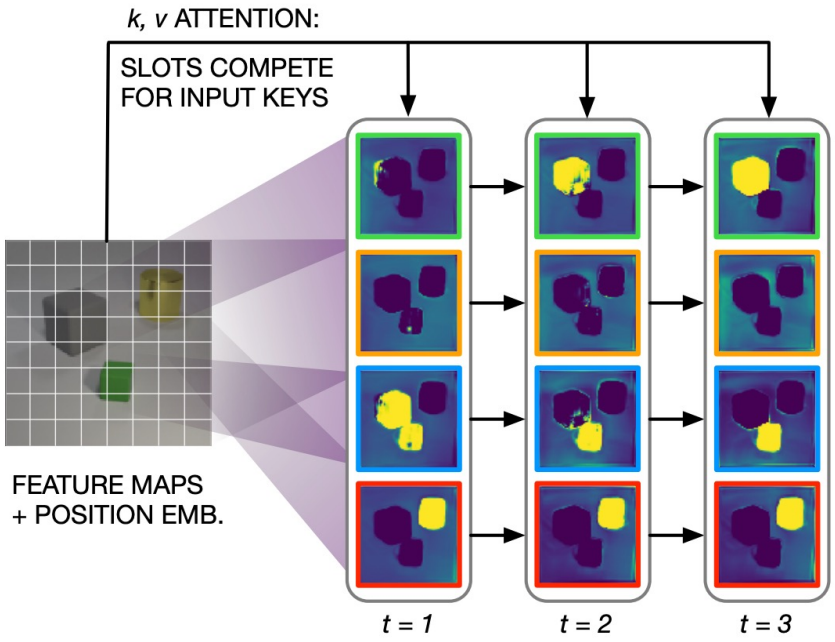
(a) Slot Attention module.

Locatello et al. Object-Centric Learning with Slot Attention. NeurIPS 2020.

NYU

Slot Attention Networks

- Can we learn clustering as an end-to-end operation?
- Slot attention is inspired by the success of the attention mechanism.
- Each “slot” attends to a region of the image and stores an object centric representation.



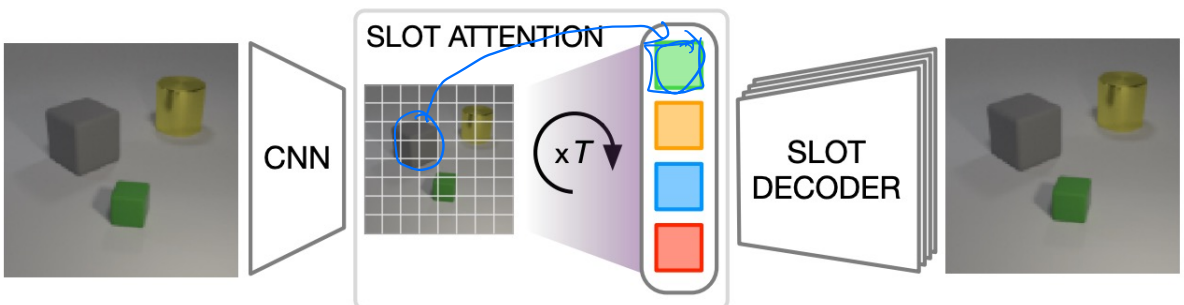
(a) Slot Attention module.

Locatello et al. Object-Centric Learning with Slot Attention. NeurIPS 2020.

NYU

Slot Attention Networks

- Goal: Reconstruct the image with a concise slot-based representation.



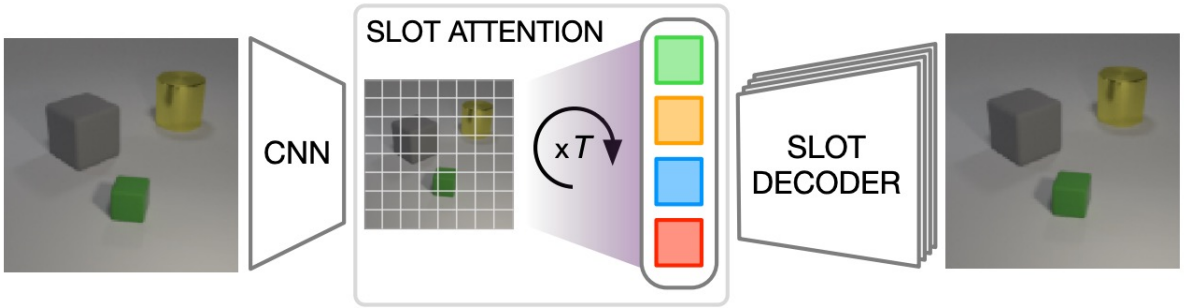
Locatello et al. Object-Centric Learning with Slot Attention. NeurIPS 2020.

NYU

Slot Attention Networks

- Goal: Reconstruct the image with a concise slot-based representation.
- Input: $x \in \mathbb{R}^{N \times D}$ (after encoder), Slots: $m_t \in \mathbb{R}^{M \times D}$. Normalize: $\tilde{m}_{t-1} = LN(m_{t-1})$.

total objects.

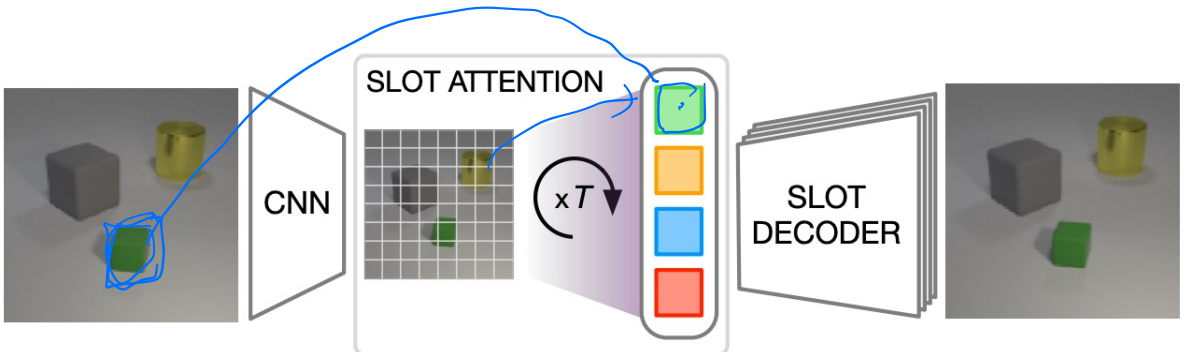


Locatello et al. Object-Centric Learning with Slot Attention. NeurIPS 2020.

NYU

Slot Attention Networks

- Goal: Reconstruct the image with a concise slot-based representation.
- Input: $x \in \mathbb{R}^{N \times D}$ (after encoder), Slots: $m \in \mathbb{R}^{M \times D}$. Normalize: $\tilde{m}_{t-1} = LN(m_{t-1})$.
- Attention over slots: $a_{t,i,j} = \frac{\frac{1}{\sqrt{D}} k(x_i) \cdot q(\tilde{m}_j)^\top}{\sum_j \frac{1}{\sqrt{D}} k(x_i) \cdot q(\tilde{m}_j)^\top}$.

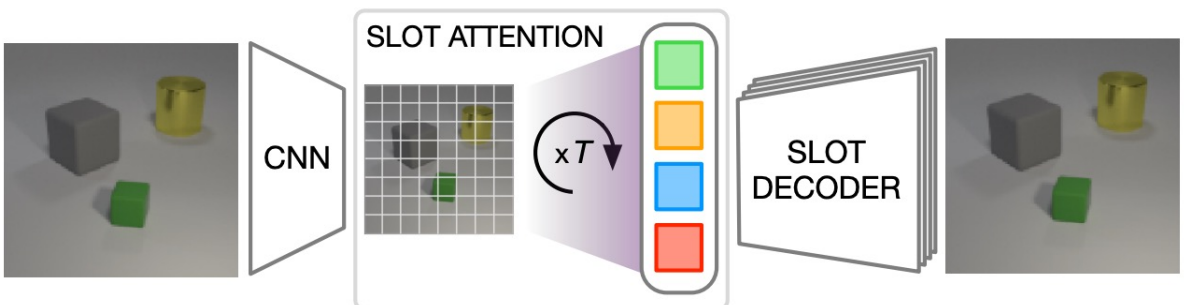


Locatello et al. Object-Centric Learning with Slot Attention. NeurIPS 2020.

NYU

Slot Attention Networks

- Goal: Reconstruct the image with a concise slot-based representation.
- Input: $x \in \mathbb{R}^{N \times D}$ (after encoder), Slots: $m \in \mathbb{R}^{M \times D}$. Normalize: $\tilde{m}_{t-1} = LN(m_{t-1})$.
- Attention over slots: $a_{t,i,j} = \frac{\frac{1}{\sqrt{D}} k(x_i) \cdot q(\tilde{m}_j)^\top}{\sum_j \frac{1}{\sqrt{D}} k(x_i) \cdot q(\tilde{m}_j)^\top}$.
- Updates: $u_{tj} = \sum_i a_{tij} v(x_i)$.

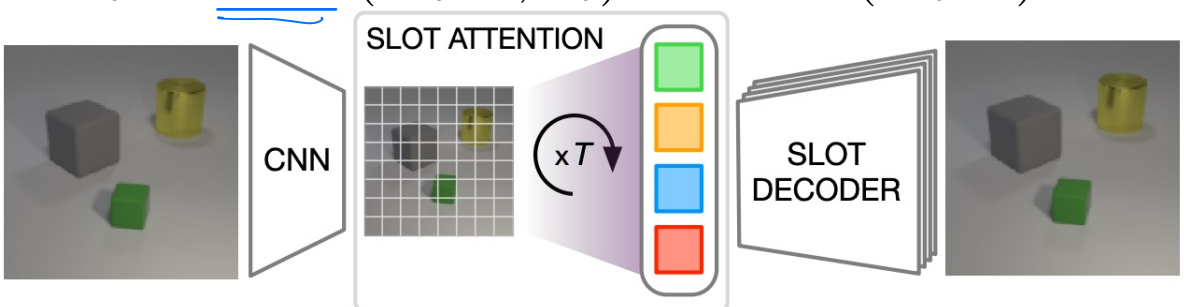


Locatello et al. Object-Centric Learning with Slot Attention. NeurIPS 2020.

NYU

Slot Attention Networks

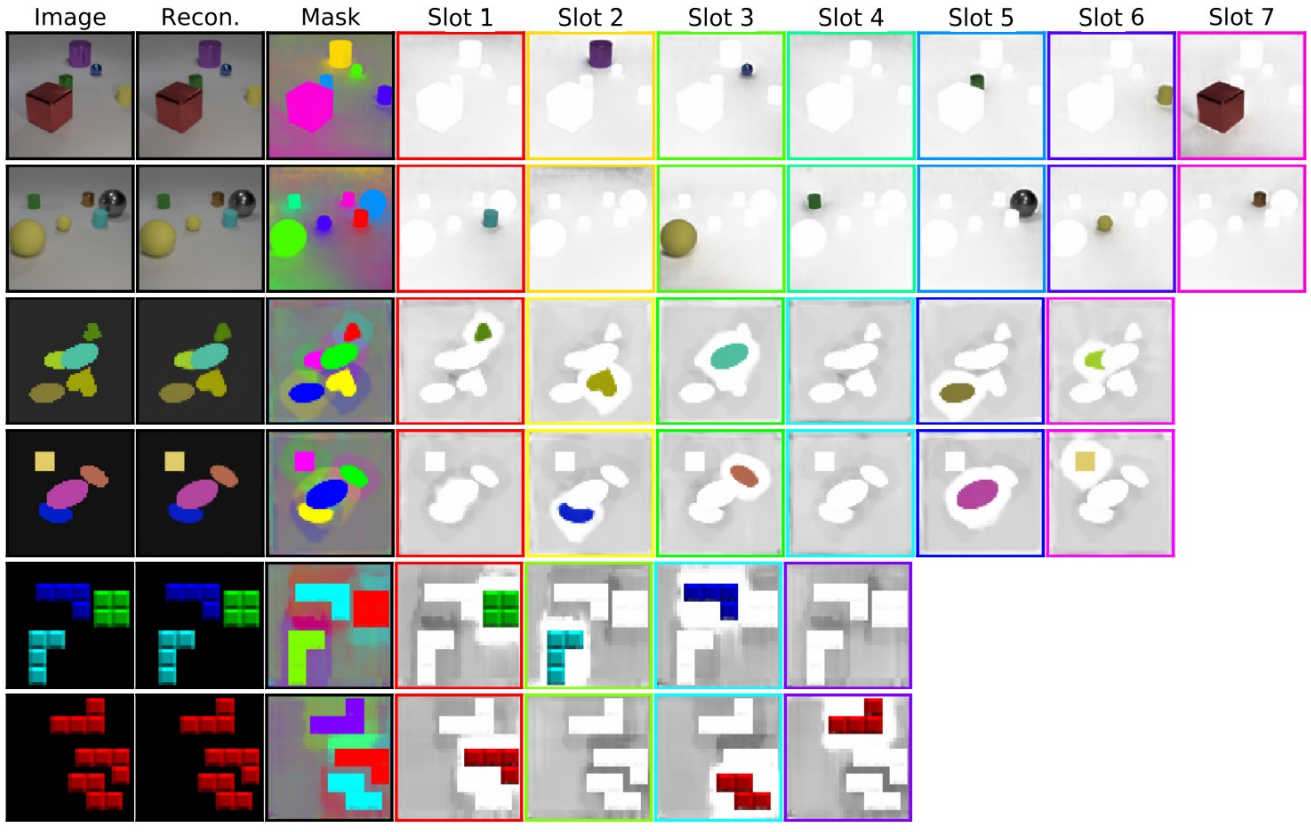
- Goal: Reconstruct the image with a concise slot-based representation.
- Input: $x \in \mathbb{R}^{N \times D}$ (after encoder), Slots: $m \in \mathbb{R}^{M \times D}$. Normalize: $\tilde{m}_{t-1} = LN(m_{t-1})$.
- Attention over slots: $a_{t,i,j} = \frac{\frac{1}{\sqrt{D}} k(x_i) \cdot q(\tilde{m}_j)^\top}{\sum_j \frac{1}{\sqrt{D}} k(x_i) \cdot q(\tilde{m}_j)^\top}$.
- Updates: $u_{tj} = \sum_i a_{tij} v(x_i)$.
- Write into slots: $m_t = GRU(m_{t-1}, u_t) + MLP(\tilde{m}_{t-1})$.



Locatello et al. Object-Centric Learning with Slot Attention. NeurIPS 2020.

NYU

Slot Attention Networks

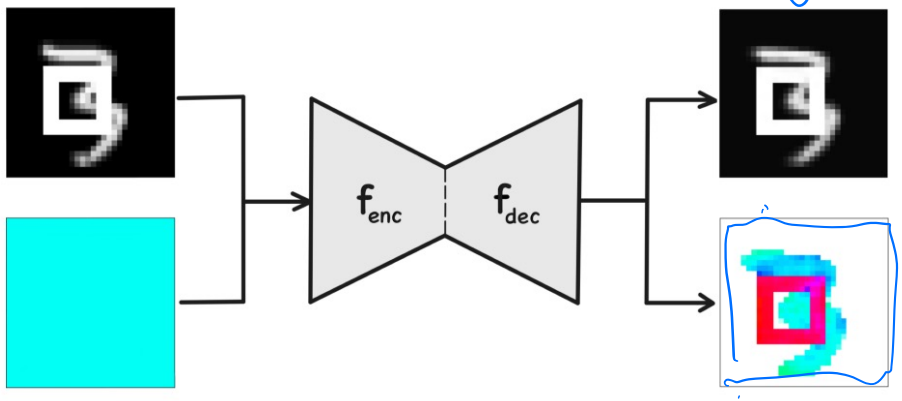


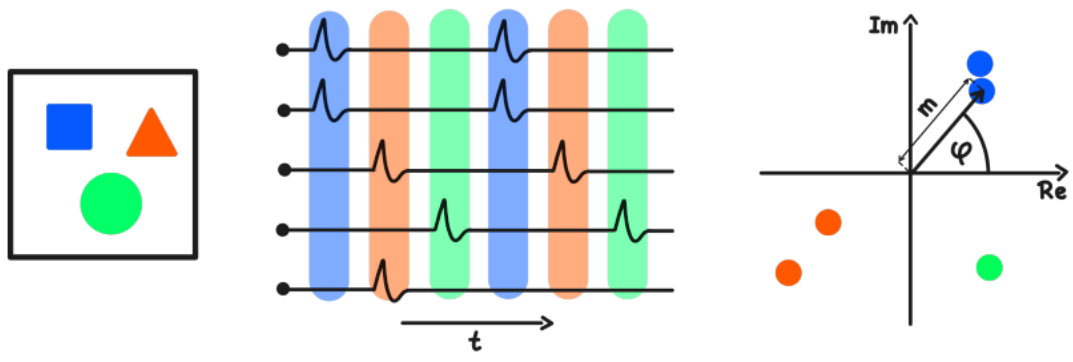
Locatello et al. Object-Centric Learning with Slot Attention. NeurIPS 2020.

NYU

Complex-Valued Autoencoders (CAEs)

- The complex number can represent magnitude and phase: $z = m \cdot e^{i\varphi} \in \mathbb{C}$.

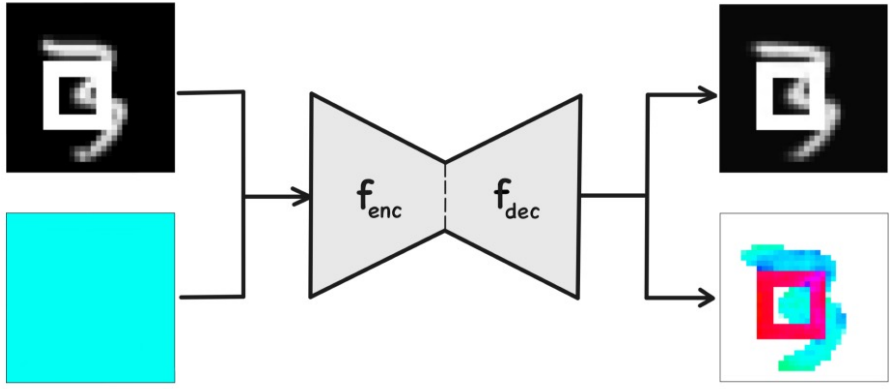


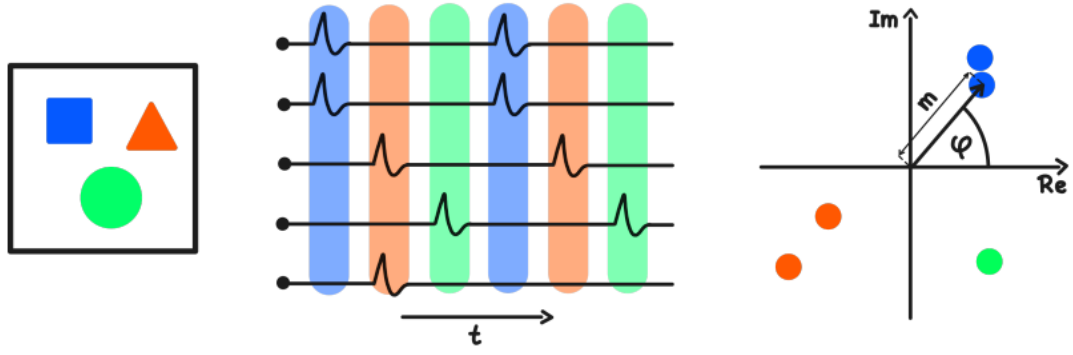


Löwe et al. Complex-Valued Autoencoders for Object Discovery. TMLR 2022.

Complex-Valued Autoencoders (CAEs)

- The complex number can represent magnitude and phase: $z = m \cdot e^{i\varphi} \in \mathbb{C}$.
- Each pixel starts with an initial phase 0.





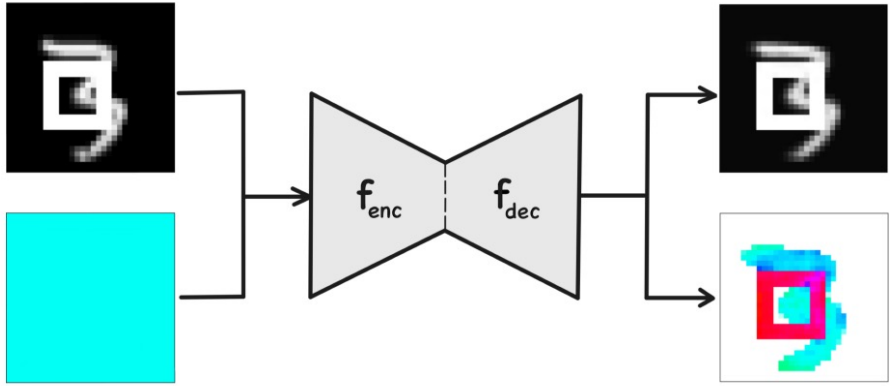
Löwe et al. Complex-Valued Autoencoders for Object Discovery. TMLR 2022.

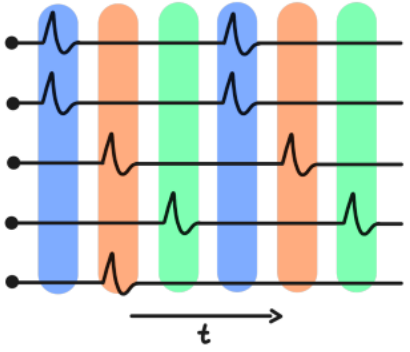
NYU

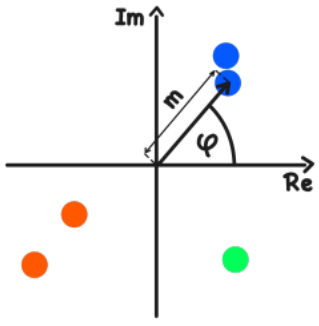
Complex-Valued Autoencoders (CAEs)

- The complex number can represent magnitude and phase: $z = m \cdot e^{i\varphi} \in \mathbb{C}$.
- Each pixel starts with an initial phase 0.

$$\hat{\mathbf{z}} = f_{\text{dec}}(f_{\text{enc}}(\mathbf{x})) \in \mathbb{C}^{h \times w}.$$







Löwe et al. Complex-Valued Autoencoders for Object Discovery. TMLR 2022.

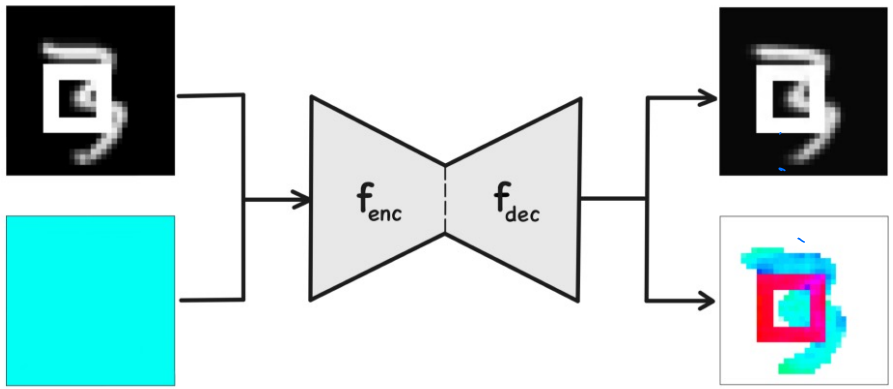
NYU

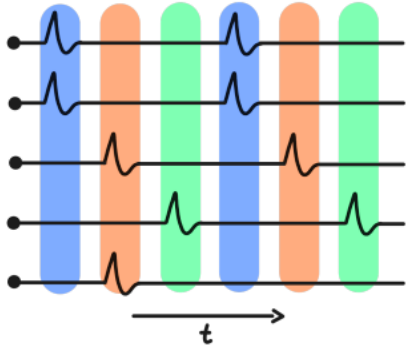
Complex-Valued Autoencoders (CAEs)

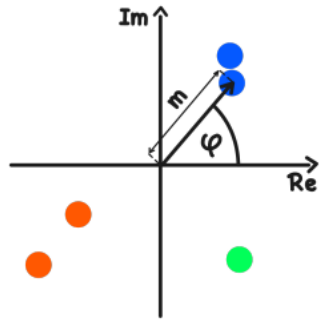
- The complex number can represent magnitude and phase: $z = m \cdot e^{i\varphi} \in \mathbb{C}$.
- Each pixel starts with an initial phase 0.

$$\hat{\mathbf{z}} = f_{\text{dec}}(f_{\text{enc}}(\mathbf{x})) \in \mathbb{C}^{h \times w}.$$

$$\hat{\mathbf{x}} = f_{\text{out}}(\hat{\mathbf{z}}) \in \mathbb{R}^{h \times w}.$$







Löwe et al. Complex-Valued Autoencoders for Object Discovery. TMLR 2022.

NYU

CAE: More Details

- Apply weights separately to real and imaginary:

$$\boldsymbol{\psi} = f_{\mathbf{w}}(\mathbf{z}) = f_{\mathbf{w}}(\mathrm{Re}(\mathbf{z})) + f_{\mathbf{w}}(\mathrm{Im}(\mathbf{z})) \cdot i \quad \in \mathbb{C}_{d_{\mathrm{out}}}$$

Shared network.

Löwe et al. Complex-Valued Autoencoders for Object Discovery. TMLR 2022.
Neuronal Synchrony in Complex-Valued Deep Networks. ICLR 2014.

CAE: More Details

- Apply weights separately to real and imaginary:

$$\boldsymbol{\psi} = f_{\mathbf{w}}(\mathbf{z}) = f_{\mathbf{w}}(\mathrm{Re}(\mathbf{z})) + f_{\mathbf{w}}(\mathrm{Im}(\mathbf{z})) \cdot i \quad \in \mathbb{C}_{d_{\mathrm{out}}}$$

- Bias on magnitude and phase:

$$\boldsymbol{m}_{\boldsymbol{\psi}} = |\boldsymbol{\psi}| + \boldsymbol{b}_{\boldsymbol{m}} \quad \in \mathbb{R}^{d_{\mathrm{out}}} \quad \boldsymbol{\varphi}_{\boldsymbol{\psi}} = \arg(\psi) + \boldsymbol{b}_{\boldsymbol{\varphi}} \quad \in \mathbb{R}^{d_{\mathrm{out}}}$$

Löwe et al. Complex-Valued Autoencoders for Object Discovery. TMLR 2022.
Neuronal Synchrony in Complex-Valued Deep Networks. ICLR 2014.

NYU

CAE: More Details

- Apply weights separately to real and imaginary:

$$\boldsymbol{\psi} = f_{\mathbf{w}}(\mathbf{z}) = f_{\mathbf{w}}(\mathrm{Re}(\mathbf{z})) + f_{\mathbf{w}}(\mathrm{Im}(\mathbf{z})) \cdot i \quad \in \mathbb{C}_{d_{\text{out}}}$$

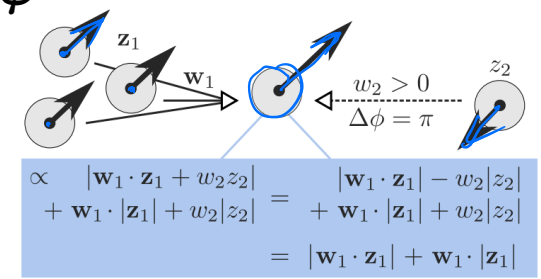
- Bias on magnitude and phase:

$$\boldsymbol{m}_{\boldsymbol{\psi}} = |\boldsymbol{\psi}| + \boldsymbol{b}_{\boldsymbol{m}} \quad \in \mathbb{R}^{d_{\text{out}}} \quad \boldsymbol{\varphi}_{\boldsymbol{\psi}} = \arg(\psi) + \boldsymbol{b}_{\boldsymbol{\varphi}} \quad \in \mathbb{R}^{d_{\text{out}}}$$

- Gating:

$$\boldsymbol{\chi} = f_{\mathbf{w}}(|\mathbf{z}|) + \boldsymbol{b}_{\boldsymbol{m}} \in \mathbb{R}^{d_{\text{out}}}$$

$$\boldsymbol{m}_{\mathbf{z}} = 0.5 \cdot \boldsymbol{m}_{\boldsymbol{\psi}} + 0.5 \cdot \boldsymbol{\chi} \in \mathbb{R}^{d_{\text{out}}}$$



Löwe et al. Complex-Valued Autoencoders for Object Discovery. TMLR 2022.
Neuronal Synchrony in Complex-Valued Deep Networks. ICLR 2014.

NYU

CAE: More Details

- Apply weights separately to real and imaginary:

$$\boldsymbol{\psi} = f_{\mathbf{w}}(\mathbf{z}) = f_{\mathbf{w}}(\mathrm{Re}(\mathbf{z})) + f_{\mathbf{w}}(\mathrm{Im}(\mathbf{z})) \cdot i \quad \in \mathbb{C}_{d_{\text{out}}}$$

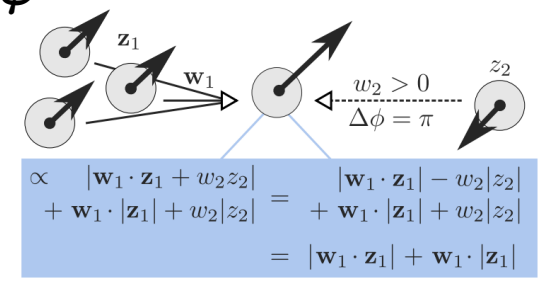
- Bias on magnitude and phase:

$$\boldsymbol{m}_{\boldsymbol{\psi}} = |\boldsymbol{\psi}| + \boldsymbol{b}_{\boldsymbol{m}} \quad \in \mathbb{R}^{d_{\text{out}}} \quad \boldsymbol{\varphi}_{\boldsymbol{\psi}} = \arg(\psi) + \boldsymbol{b}_{\boldsymbol{\varphi}} \quad \in \mathbb{R}^{d_{\text{out}}}$$

- Gating:

$$\boldsymbol{\chi} = f_{\mathbf{w}}(|\mathbf{z}|) + \boldsymbol{b}_{\boldsymbol{m}} \in \mathbb{R}^{d_{\text{out}}}$$

$$\boldsymbol{m}_{\mathbf{z}} = 0.5 \cdot \boldsymbol{m}_{\boldsymbol{\psi}} + 0.5 \cdot \boldsymbol{\chi} \in \mathbb{R}^{d_{\text{out}}}$$

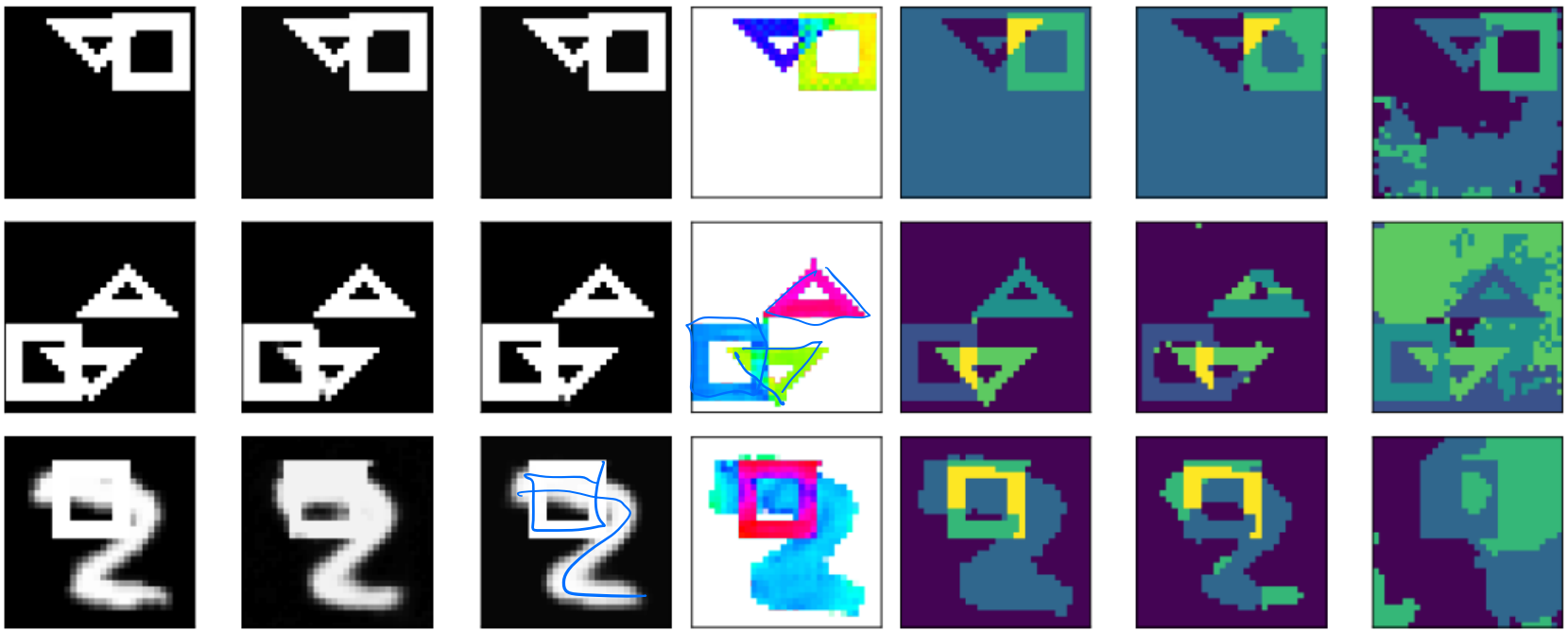


- Activation: $\mathbf{z}' = \mathrm{ReLU}(\mathrm{BatchNorm}(\boldsymbol{m}_{\mathbf{z}})) \cdot e^{i\boldsymbol{\varphi}_{\boldsymbol{\psi}}} \quad \in \mathbb{C}^{d_{\text{out}}}$

Löwe et al. Complex-Valued Autoencoders for Object Discovery. TMLR 2022.
Neuronal Synchrony in Complex-Valued Deep Networks. ICLR 2014.

NYU

Complex-Valued Autoencoders



Löwe et al. Complex-Valued Autoencoders for Object Discovery. TMLR 2022.

NYU

Summary: Object Discovery

- Combine deep features with clustering algorithms.

NYU

Summary: Object Discovery

- Combine deep features with clustering algorithms.
- Pseudo-labels to train detector networks.

Summary: Object Discovery

- Combine deep features with clustering algorithms.
- Pseudo-labels to train detector networks.
- Creative end-to-end learning-based solutions exist, but there are still plenty room for improvement.
  - Possible to train from scratch!

Summary: Object Discovery

- Combine deep features with clustering algorithms.
- Pseudo-labels to train detector networks.
- Creative end-to-end learning-based solutions exist, but there are still plenty room for improvement.
  - Possible to train from scratch!
- What do we make use of the discovered objects? Is it better to keep the awareness in the latent space?

NYU

Module 4:
World Models and End-to-End Planning

NYU

World Models: Predicting Future

3D, latent, → predict future ← → planning.

World Models: Predicting Future

- World models: Models that predict the future states based on the past sequences of states and actions.

World Models: Predicting Future

- World models: Models that predict the future states based on the past sequences of states and actions.
- Predicting future is a natural self-supervised learning objective.

World Models: Predicting Future

- World models: Models that predict the future states based on the past sequences of states and actions.
- Predicting future is a natural self-supervised learning objective.
- Why?

NYU

World Models: Predicting Future

- World models: Models that predict the future states based on the past sequences of states and actions.
- Predicting future is a natural self-supervised learning objective.
- Why?
  - Reward is sparse, rollout trajectories are abundant.

NYU

World Models: Predicting Future

- World models: Models that predict the future states based on the past sequences of states and actions.
- Predicting future is a natural self-supervised learning objective.
- Why?
  - Reward is sparse, rollout trajectories are abundant.
  - Learning task /planning aware representations.

NYU

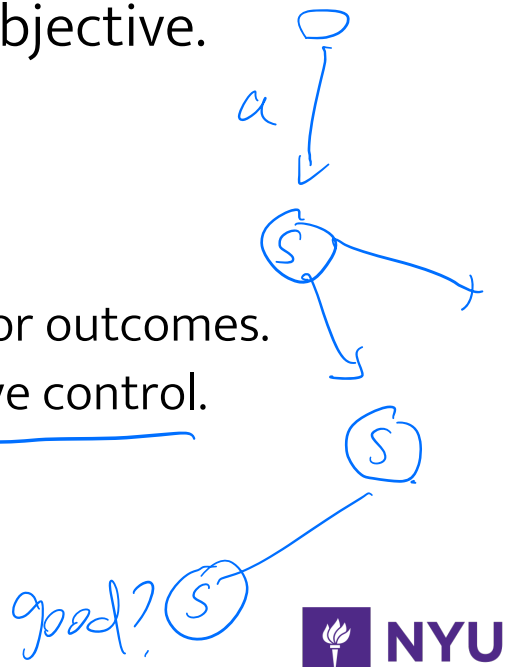
World Models: Predicting Future

- World models: Models that predict the future states based on the past sequences of states and actions.
- Predicting future is a natural self-supervised learning objective.
- Why?
  - Reward is sparse, rollout trajectories are abundant.
  - Learning task /planning aware representations.
  - Predicting future can “simulate” rollouts without querying for outcomes.

NYU

World Models: Predicting Future

- World models: Models that predict the future states based on the past sequences of states and actions.
- Predicting future is a natural self-supervised learning objective.
- Why?
  - Reward is sparse, rollout trajectories are abundant.
  - Learning task /planning aware representations.
  - Predicting future can “simulate” rollouts without querying for outcomes.
  - Can help planning by predicting at test time. Model predictive control.



NYU

Model-Predictive Control (MPC)

- A classic example of world models.
- Analytical forms, complete knowledge of the dynamical system.

Rawlings et al. Model Predictive Control: Theory, Computation, and Design.
https://engineering.purdue.edu/~zak/Second_ed/MPC_handout.pdf

NYU

Model-Predictive Control (MPC)

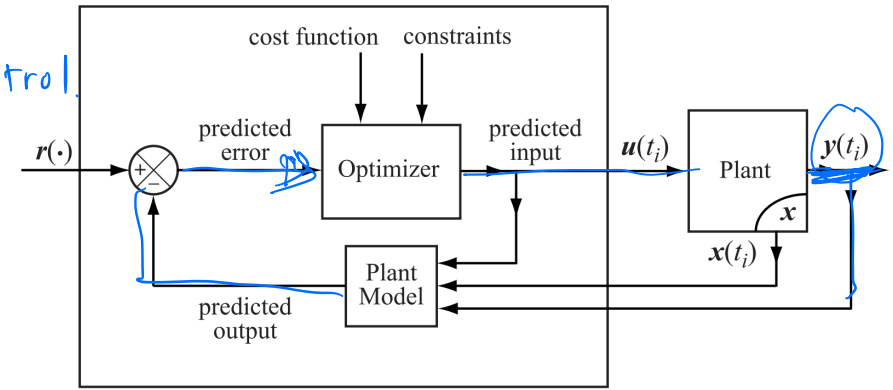
- A classic example of world models.
- Analytical forms, complete knowledge of the dynamical system.

General Form:

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}, \mathbf{u}, t)$$

$$\mathbf{y} = h(\mathbf{x}, \mathbf{u}, t)$$

$$\mathbf{x}(t_0) = \mathbf{x}_0$$



Model Predictive Controller

Rawlings et al. Model Predictive Control: Theory, Computation, and Design.
https://engineering.purdue.edu/~zak/Second_ed/MPC_handout.pdf

Model-Predictive Control (MPC)

- A classic example of world models.
- Analytical forms, complete knowledge of the dynamical system.

General Form:

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}, \mathbf{u}, t)$$
$$\mathbf{y} = h(\mathbf{x}, \mathbf{u}, t)$$
$$\mathbf{x}(t_0) = \mathbf{x}_0$$

Linear Form:

$$\frac{dx}{dt} = A\mathbf{x} + B\mathbf{u}$$
$$\mathbf{y} = C\mathbf{x} + D\mathbf{u}$$
$$\mathbf{x}(t_0) = \mathbf{x}_0$$

Rawlings et al. Model Predictive Control: Theory, Computation, and Design.
https://engineering.purdue.edu/~zak/Second_ed/MPC_handout.pdf

NYU

Model-Predictive Control (MPC)

- A classic example of world models.
- Analytical forms, complete knowledge of the dynamical system.

General Form:

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}, \mathbf{u}, t)$$
$$\mathbf{y} = h(\mathbf{x}, \mathbf{u}, t)$$
$$\mathbf{x}(t_0) = \mathbf{x}_0$$

Linear Form:

$$\frac{dx}{dt} = A\mathbf{x} + B\mathbf{u}$$
$$\mathbf{y} = C\mathbf{x} + D\mathbf{u}$$
$$\mathbf{x}(t_0) = \mathbf{x}_0$$

Optimize Value/Cost Function:

$$\min_{\mathbf{u}} J(\mathbf{x}(0), \mathbf{u})$$

Rawlings et al. Model Predictive Control: Theory, Computation, and Design.
https://engineering.purdue.edu/~zak/Second_ed/MPC_handout.pdf

NYU

Model-Predictive Control (MPC)

- A classic example of world models.
- Analytical forms, complete knowledge of the dynamical system.

General Form:

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}, \mathbf{u}, t)$$
$$\mathbf{y} = h(\mathbf{x}, \mathbf{u}, t)$$
$$\mathbf{x}(t_0) = \mathbf{x}_0$$

Linear Form:

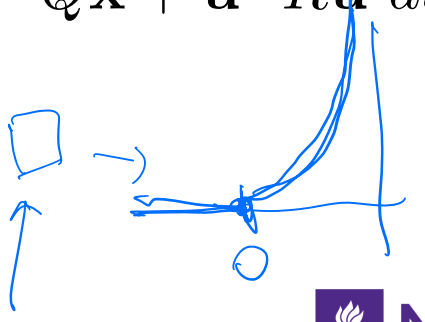
$$\frac{dx}{dt} = A\mathbf{x} + B\mathbf{u}$$
$$\mathbf{y} = C\mathbf{x} + D\mathbf{u}$$
$$\mathbf{x}(t_0) = \mathbf{x}_0$$

Optimize Value/Cost Function:

$$\min_{\mathbf{u}} J(\mathbf{x}(0), \mathbf{u})$$

Quadratic Form (LQR):

$$J = \int_0^\infty \mathbf{x}^\top Q\mathbf{x} + \mathbf{u}^\top R\mathbf{u}\, dt.$$



Rawlings et al. Model Predictive Control: Theory, Computation, and Design.
https://engineering.purdue.edu/~zak/Second_ed/MPC_handout.pdf

NYU

Model-Predictive Control (MPC)

- A classic exar
- Analytical for

al system.

General Form:

$$\frac{d\mathbf{x}}{dt} = f(:$$

$$\mathbf{y} = h(:$$

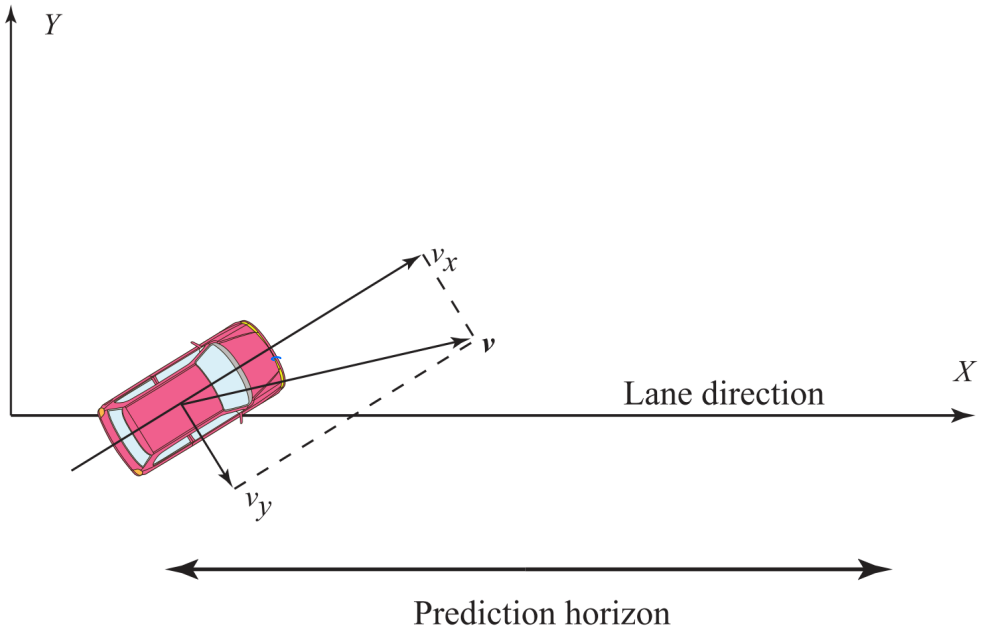
$$\mathbf{x}(t_0) = \mathbf{x}_0$$

alue/Cost Function:

$$J(\mathbf{x}(0), \mathbf{u})$$

Form (LQR):

$$\mathbf{x}^\top Q\mathbf{x} + \mathbf{u}^\top R\mathbf{u}\ dt.$$



- Example: Cars: x: position velocity angle angular velocity; u: jerk and angular accel.;

Rawlings et al. Model Predictive Control: Theory, Computation, and Design.
https://engineering.purdue.edu/~zak/Second_ed/MPC_handout.pdf

NYU

Several Types of World Models

- State
  - Explicitly defined state space
  - Needs a separately trained perception network

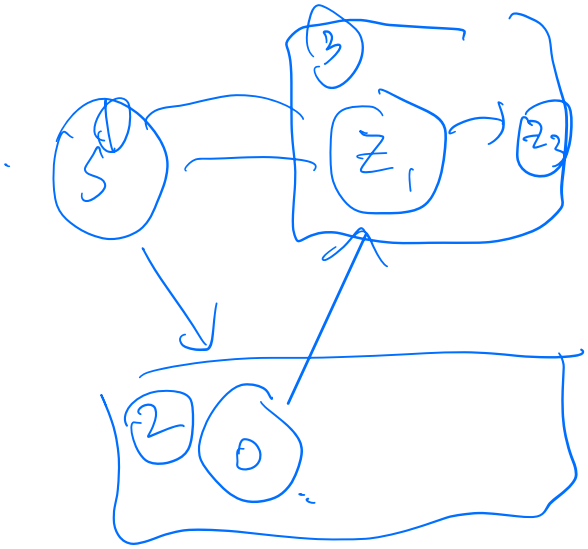
Several Types of World Models

- State
  - Explicitly defined state space
  - Needs a separately trained perception network
- Raw Input
  - Generate future camera video frames / point clouds

NYU

Several Types of World Models

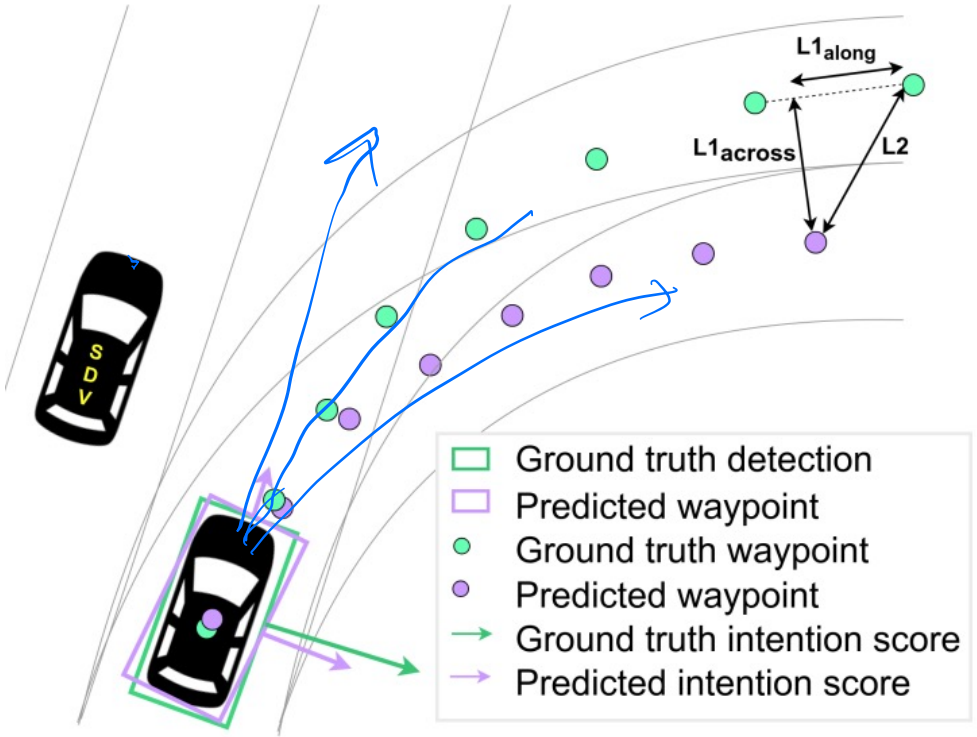
- State
  - Explicitly defined state space
  - Needs a separately trained perception network
- Raw Input
  - Generate future camera video frames / point clouds
- Latent
  - Generate future latent states
  - May go outside of the latent manifold



NYU

Trajectory Prediction as Object Detection

- We also need to predict external dynamic objects.

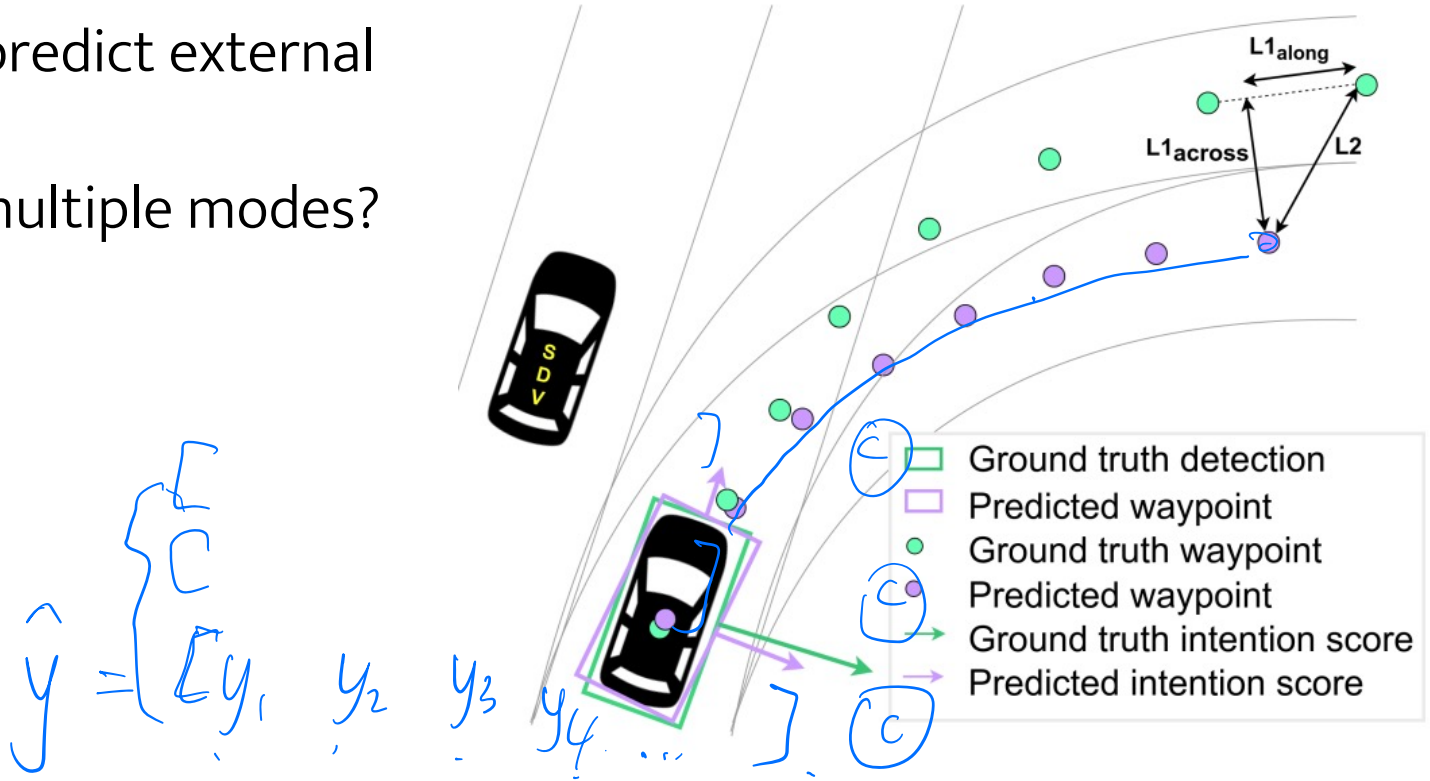


Casas et al. IntentNet: Learning to Predict Intention from Raw Sensor Data. CoRL 2018.

NYU

Trajectory Prediction as Object Detection

- We also need to predict external dynamic objects.
- How to capture multiple modes?

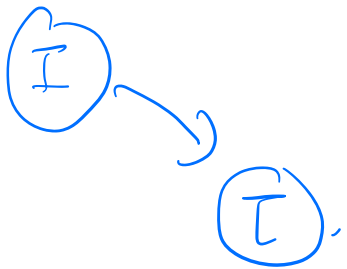


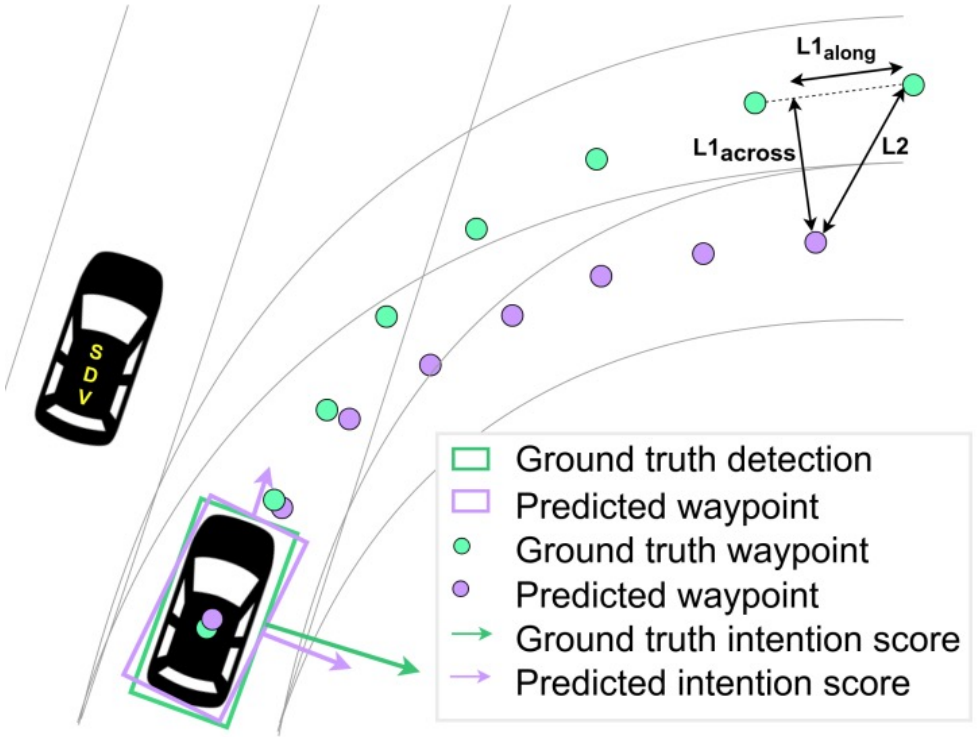
Casas et al. IntentNet: Learning to Predict Intention from Raw Sensor Data. CoRL 2018.

NYU

Trajectory Prediction as Object Detection

- We also need to predict external dynamic objects.
- How to capture multiple modes?
- Discrete intention prediction problem:
  - keep lane, turn left/right, left/right change lane, stopping, parked, etc.



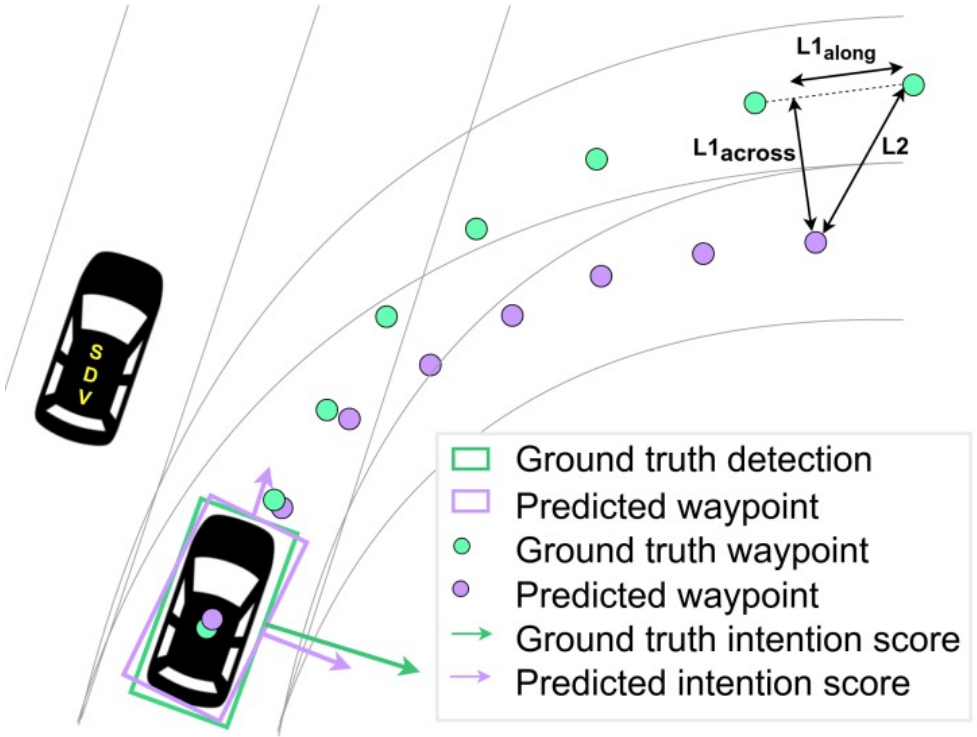


Casas et al. IntentNet: Learning to Predict Intention from Raw Sensor Data. CoRL 2018.

NYU

Trajectory Prediction as Object Detection

- We also need to predict external dynamic objects.
- How to capture multiple modes?
- Discrete intention prediction problem:
  - keep lane, turn left/right, left/right change lane, stopping, parked, etc.
- Output multiple trajectories

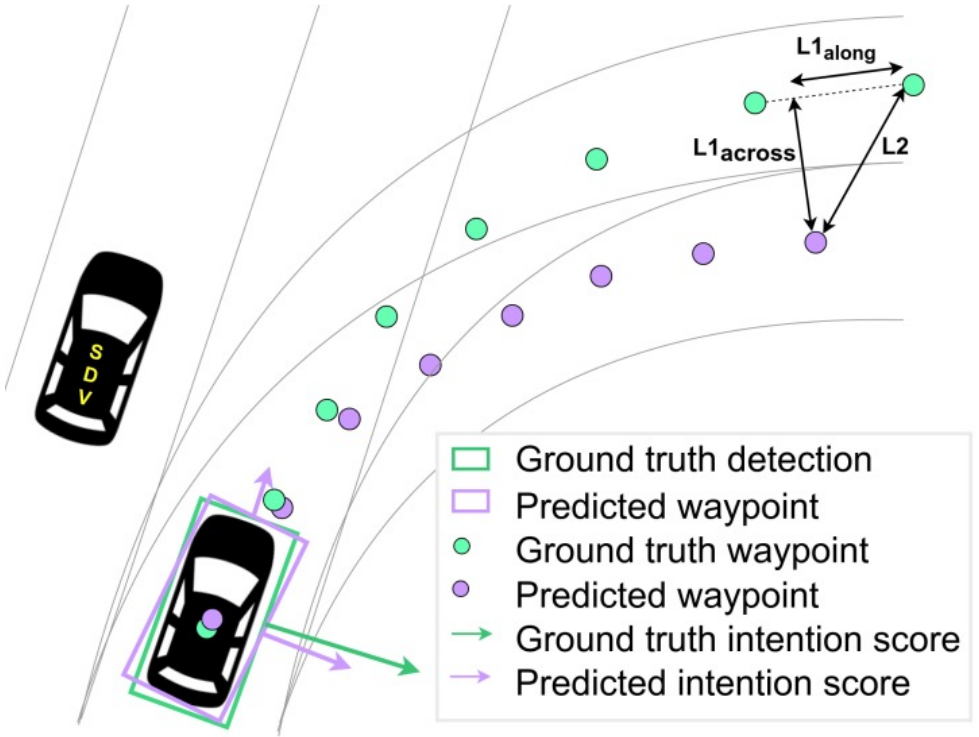


Casas et al. IntentNet: Learning to Predict Intention from Raw Sensor Data. CoRL 2018.

NYU

Trajectory Prediction as Object Detection

- We also need to predict external dynamic objects.
- How to capture multiple modes?
- Discrete intention prediction problem:
  - keep lane, turn left/right, left/right change lane, stopping, parked, etc.
- Output multiple trajectories
- Requires high-level action labels.



Casas et al. IntentNet: Learning to Predict Intention from Raw Sensor Data. CoRL 2018.

NYU

Semantic Occupancy Volume Prediction

- 4-D volume: H, W, T, C (class) $o_{i,j}^{t,c}$

num. timesteps.

Perceive, Predict, and Plan: Safe Motion Planning Through Interpretable Semantic Representations. ECCV 2020.

NYU

Semantic Occupancy Volume Prediction

- 4-D volume: H, W, T, C (class) $o_{i,j}^{t,c}$
- Doesn't grow with the increasing number of actors.

Perceive, Predict, and Plan: Safe Motion Planning Through Interpretable Semantic Representations. ECCV 2020.

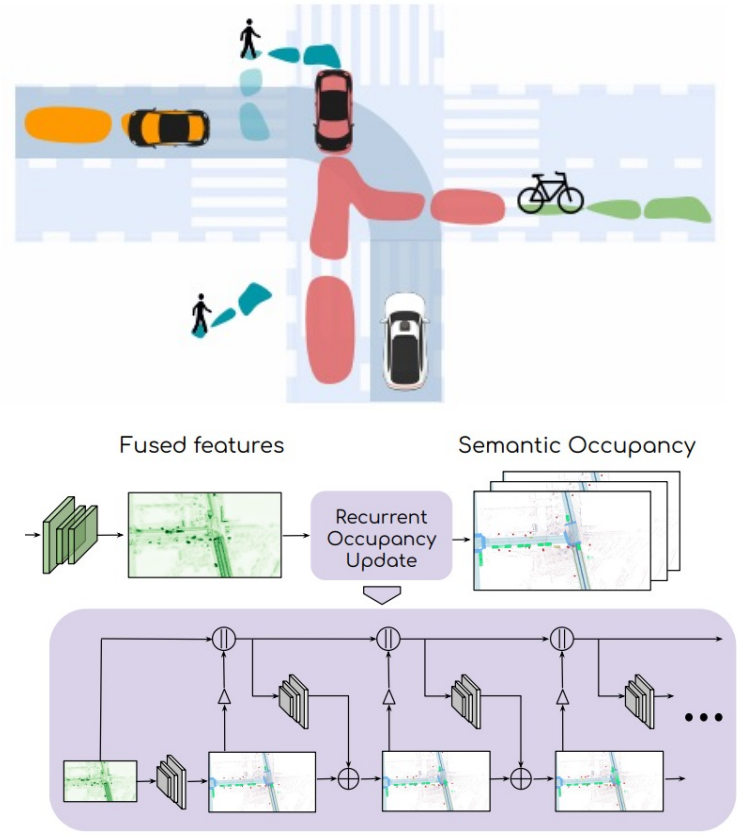
NYU

Semantic Occupancy Volume Prediction

- 4-D volume: H, W, T, C (class) $o_{i,j}^{t,c}$
- Doesn't grow with the increasing number of actors.
- Recurrent occupancy updates for further into the future.

$$l^{t,c} = l^{t-1,c} + \mathcal{U}_\theta^t(f_{\text{occ}}, l^{0:t-1,c})$$

↳ render Cost volume.

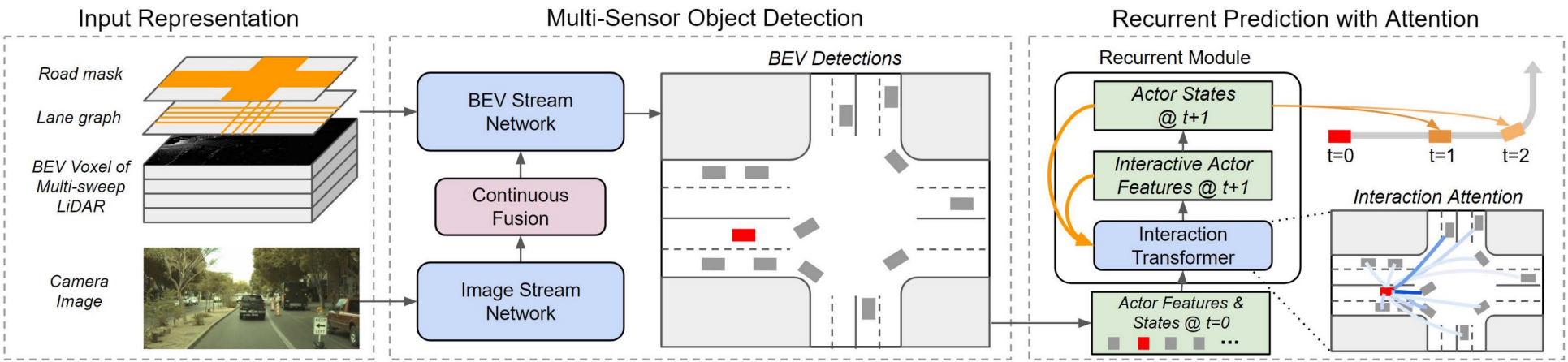


Perceive, Predict, and Plan: Safe Motion Planning Through Interpretable Semantic Representations. ECCV 2020.

NYU

Multi-Agents Joint Prediction

- Joint predict future trajectories by attending to other actors.

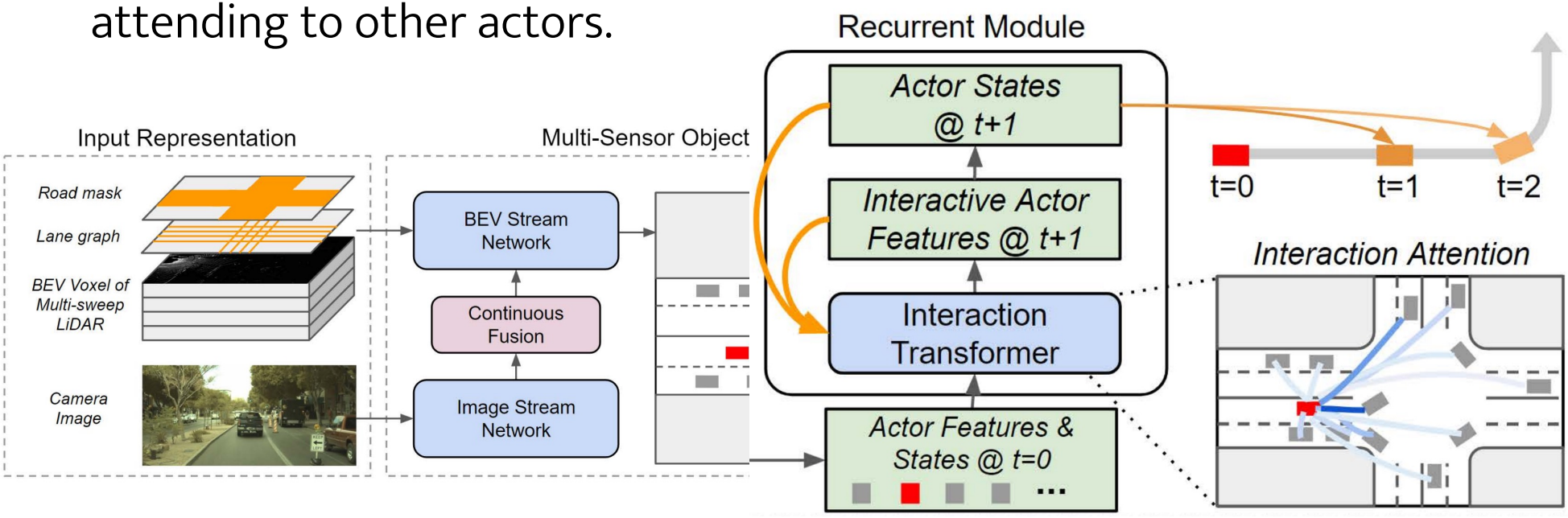


Li et al. End-to-end Contextual Perception and Prediction with Interaction Transformer. IROS 2020.

NYU

Multi-Agents Joint Prediction

- Joint predict future trajectories by attending to other actors.



Li et al. End-to-end Contextual Perception and Prediction with Interaction Transformer. IROS 2020.

NYU

Generating Raw Pixels

https://deepmind.google/blog/genie-3-a-new-frontier-for-world-models/

Genie 3: A new frontier for world models

NYU

Generating Raw Pixels

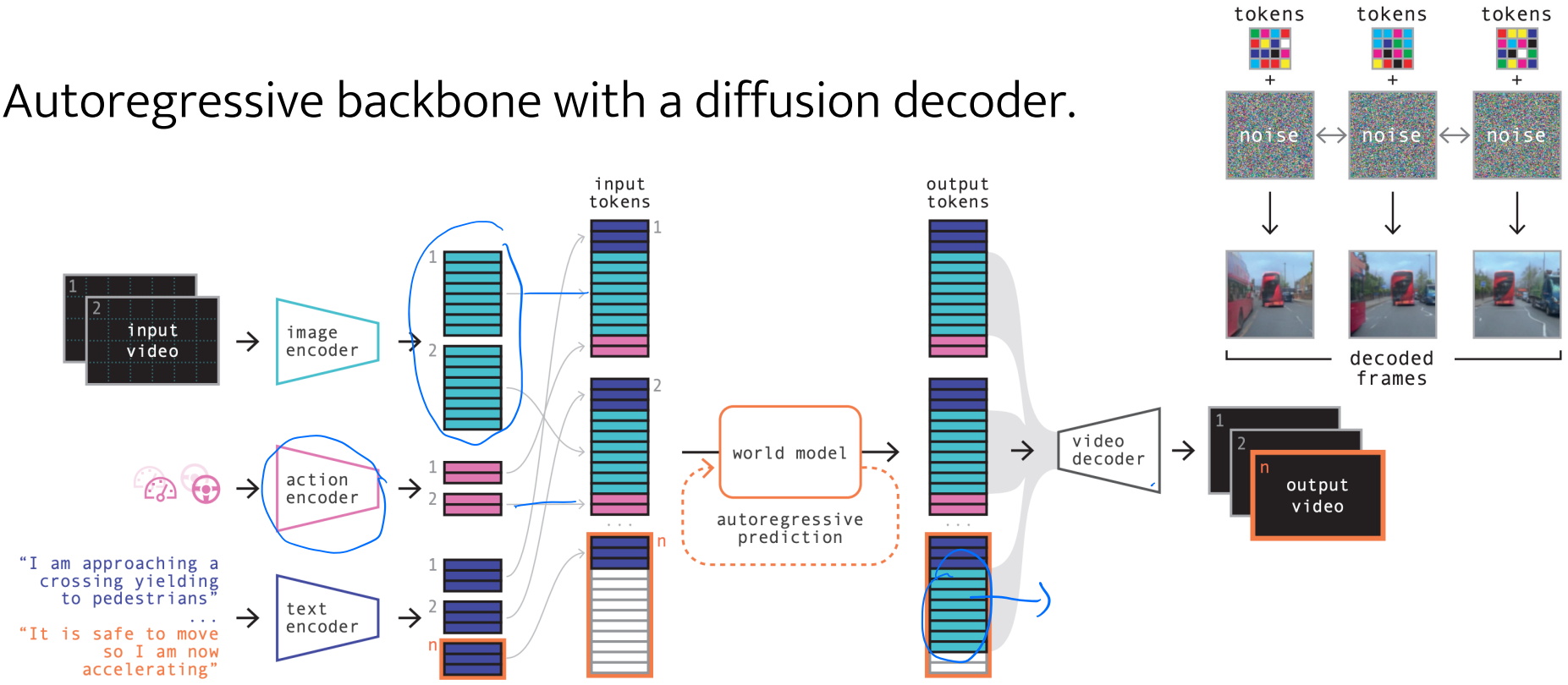
https://deepmind.google/blog/genie-3-a-new-frontier-for-world-models/

Genie 3: A new frontier for world models

NYU

Action-Conditioned Video Generation

- Autoregressive backbone with a diffusion decoder.



Hu et al. GAIA-1: A Generative World Model for Autonomous Driving. arXiv 2023.

NYU

https://www.youtube.com/watch?v=5Jx2QgEUZUI

Hu et al. GAIA-1: A Generative World Model for Autonomous Driving. arXiv 2023.

NYU

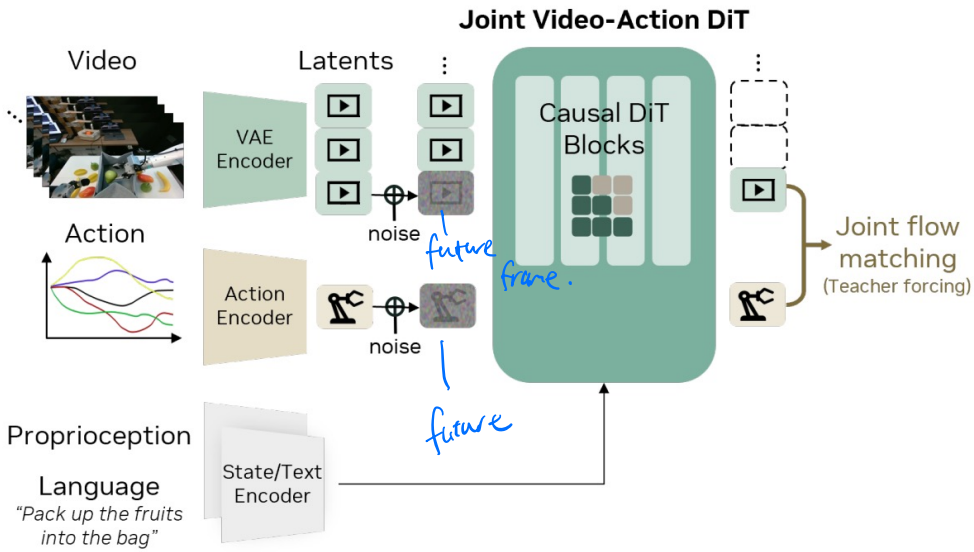
https://www.youtube.com/watch?v=5Jx2QgEUZUI

Hu et al. GAIA-1: A Generative World Model for Autonomous Driving. arXiv 2023.

NYU

Action-Conditioned Video Generation

- DiT backbone with teacher forcing flow matching objective.
- Single actor robotic videos.

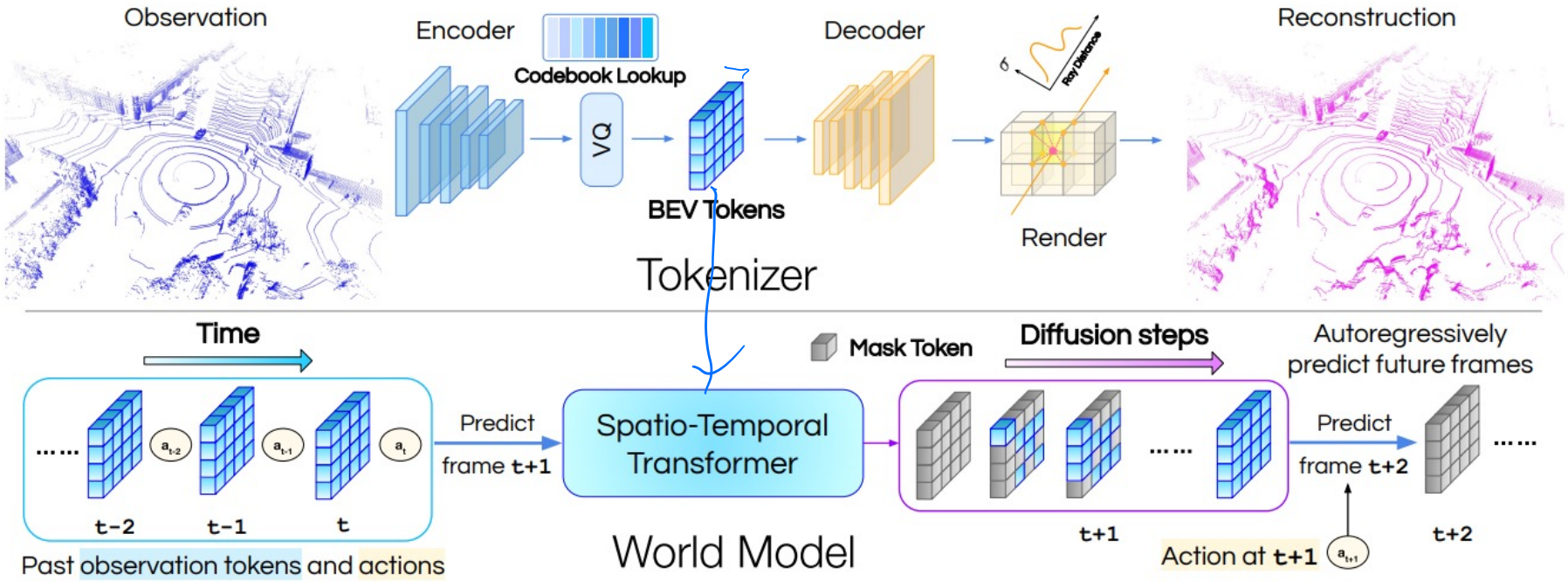


Ye et al. DreamZero: World Action Models are Zero-shot Policies. 2026

NYU

World Model in 3D volume prediction

- Autoregressively predict future 3D point clouds.



Zhang et al. Copilot4D: Learning Unsupervised World Models for Autonomous Driving via Discrete Diffusion. ICLR 2024.

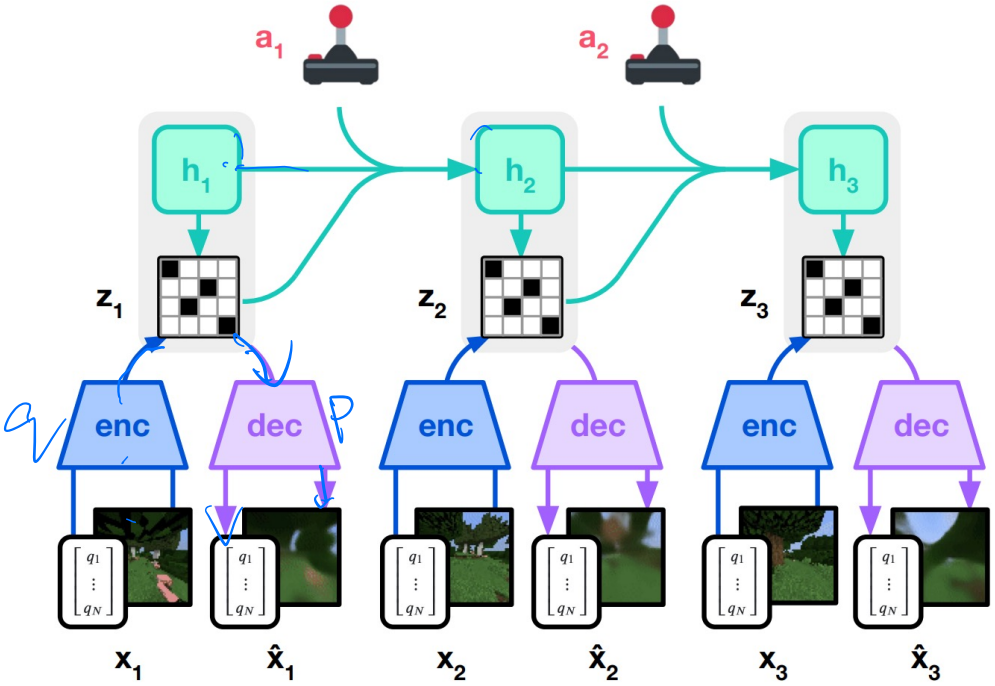
NYU

Latent Sequence World Model

- Autoencoder to ensure the latent representations are meaningful.

$$z_t \sim q_\phi(z_t \mid h_t, x_t)$$

$$\hat{x}_t \sim p_\phi(\hat{x}_t \mid h_t, z_t)$$



JEPA

Hafner et al. Mastering Diverse Domains through World Models. 2023.

NYU

Latent Sequence World Model

- Autoencoder to ensure the latent representations are meaningful.

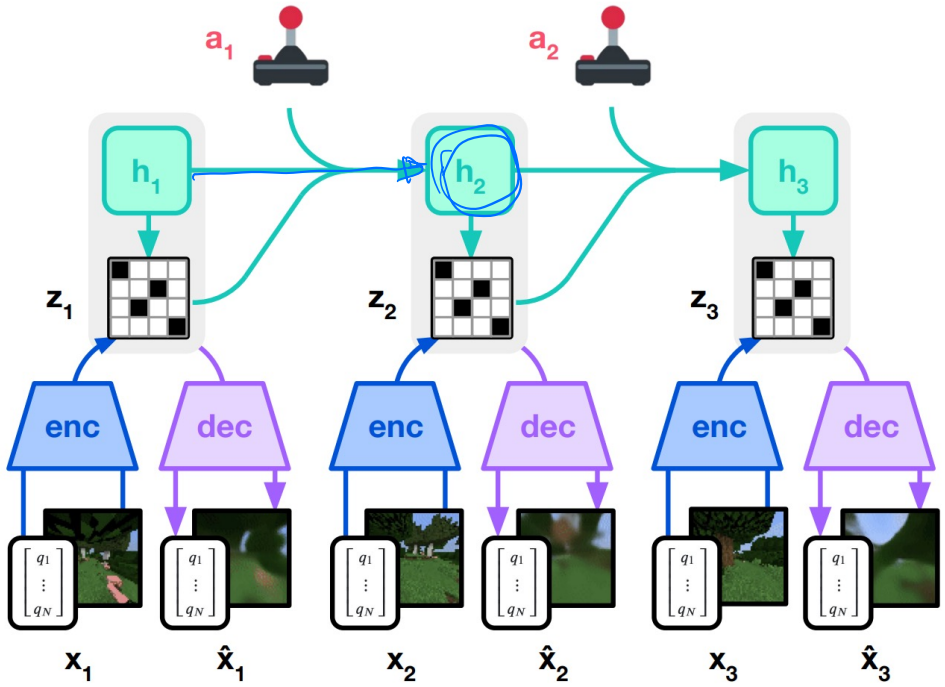
$$z_t \sim q_\phi(z_t \mid h_t, x_t)$$

$$\hat{x}_t \sim p_\phi(\hat{x}_t \mid h_t, z_t)$$

- Learn a sequence model to predict the latent conditioned on previous action.

$$h_t = f_\phi(h_{t-1}, z_{t-1}, a_{t-1})$$

$$\hat{z}_t \sim p_\phi(\hat{z}_t \mid h_t)$$



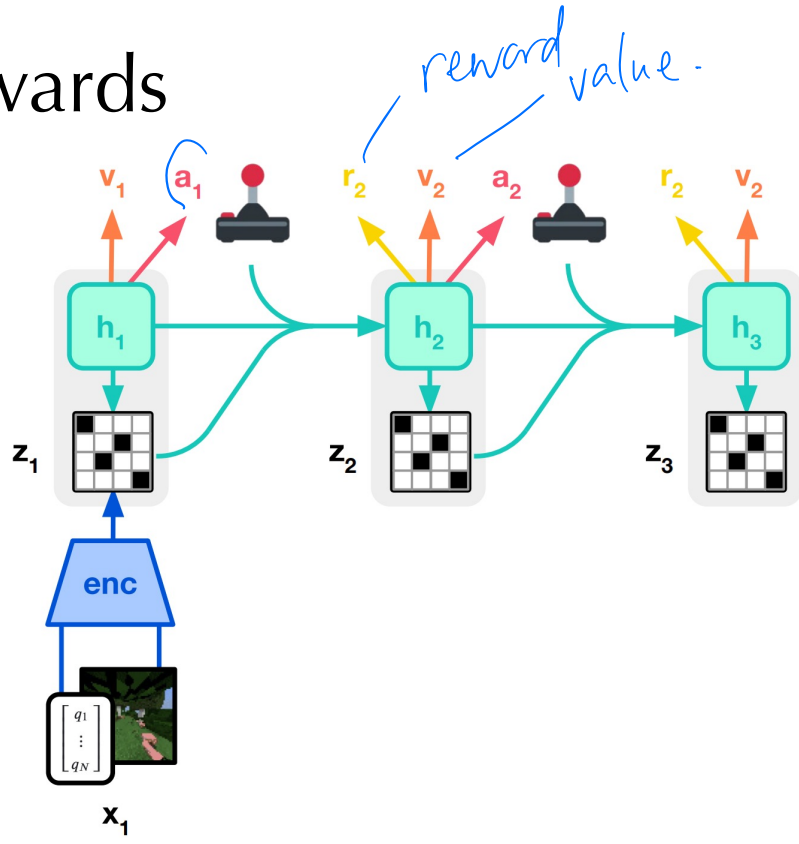
Hafner et al. Mastering Diverse Domains through World Models. 2023.

NYU

Incorporating Rewards

- Predicting reward

$$\hat{r}_t \sim p_\phi(\hat{r}_t \mid h_t, z_t)$$



Hafner et al. Mastering Diverse Domains through World Models. 2023.

NYU

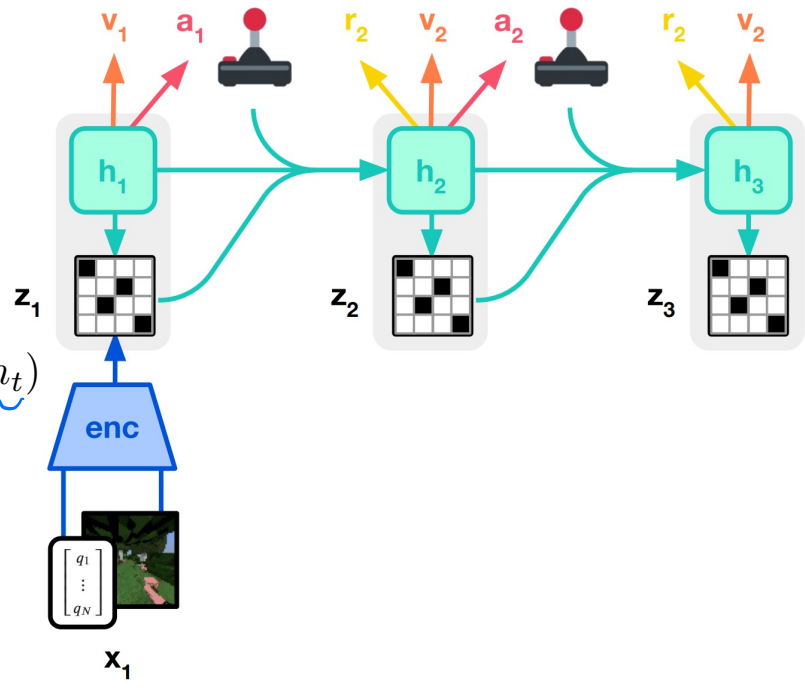
Incorporating Rewards

- Predicting reward

$$\hat{r}_t \sim p_\phi(\hat{r}_t \mid h_t, z_t)$$

- Reconstruction, reward, continue

$$\mathcal{L}_{\text{pred}}(\phi) = -\log p_\phi(x_t \mid z_t, h_t) - \log p_\phi(r_t \mid z_t, h_t) - \log p_\phi(c_t \mid z_t, h_t)$$



Incorporating Rewards

- Predicting reward

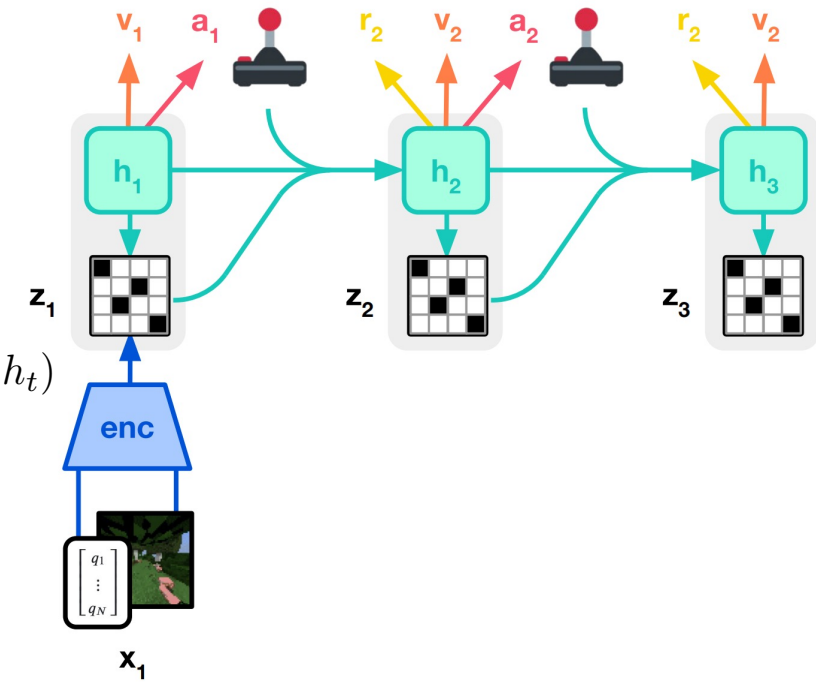
$$\hat{r}_t \sim p_\phi(\hat{r}_t \mid h_t, z_t)$$

- Reconstruction, reward, continue

$$\mathcal{L}_{\text{pred}}(\phi) = -\log p_\phi(x_t \mid z_t, h_t) - \log p_\phi(r_t \mid z_t, h_t) - \log p_\phi(c_t \mid z_t, h_t)$$

- Dynamics: Predicting future z

$$\mathcal{L}_{\text{dyn}}(\phi) = \max(1, \mathrm{KL}[\mathrm{sg}(q_\phi(z_t \mid h_t, x_t) \parallel p_\phi(z_t|h_t))]$$



Hafner et al. Mastering Diverse Domains through World Models. 2023.

NYU

Incorporating Rewards

- Predicting reward

$$\hat{r}_t \sim p_\phi(\hat{r}_t \mid h_t, z_t)$$

- Reconstruction, reward, continue

$$\mathcal{L}_{\text{pred}}(\phi) = -\log p_\phi(x_t \mid z_t, h_t) - \log p_\phi(r_t \mid z_t, h_t) - \log p_\phi(c_t \mid z_t, h_t)$$

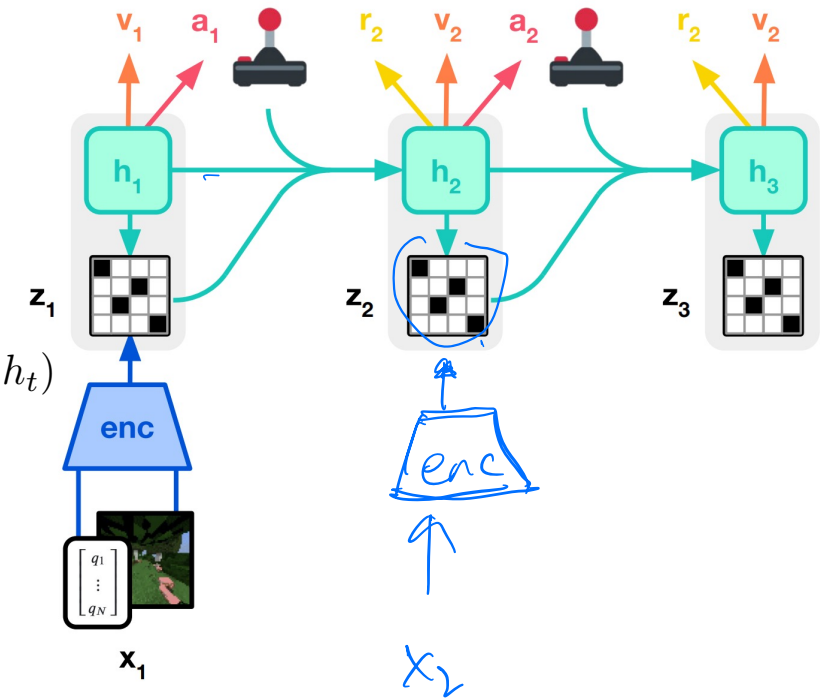
- Dynamics: Predicting future z

$$\mathcal{L}_{\text{dyn}}(\phi) = \max(1, \mathrm{KL}[\mathrm{sg}(q_\phi(z_t \mid h_t, x_t) \parallel p_\phi(z_t|h_t))]$$

- Align representation

$$\mathcal{L}_{\text{rep}} = \max(1, \mathrm{KL}[q_\phi(z_t|h_t, x_t) \parallel \mathrm{sg}(p_\phi(z_t|h_t))])$$

JEPA

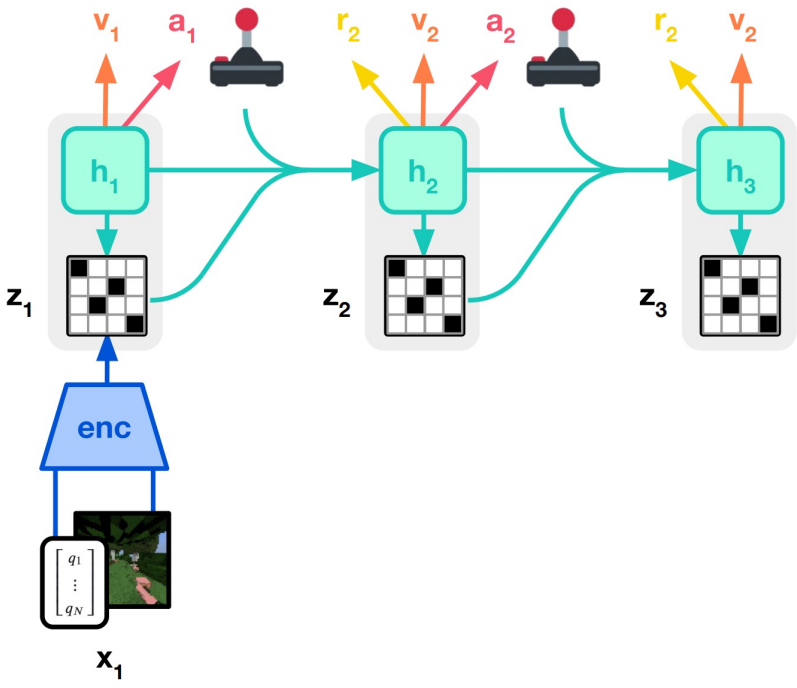


Hafner et al. Mastering Diverse Domains through World Models. 2023.

NYU

Incorporating Rewards

- How to use WM in planning? Predicting value by simulate a batch of trajectories.



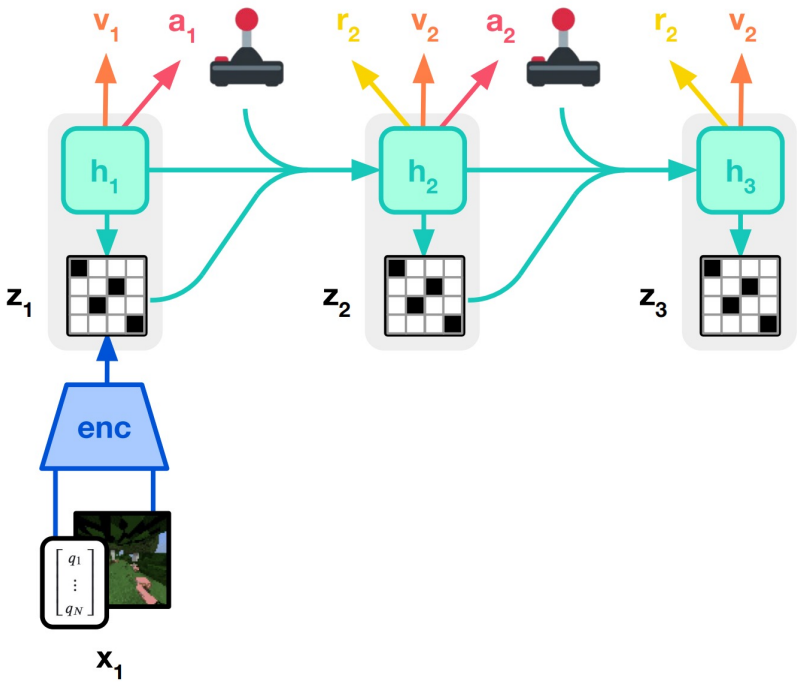
Hafner et al. Mastering Diverse Domains through World Models. 2023.

NYU

Incorporating Rewards

- How to use WM in planning? Predicting value by simulate a batch of trajectories.
- Actor-Critic RL:

$a_t \sim \pi_\theta(a_t \mid s_t)$ $\quad$ $v_\psi(R_t \mid s_t)$



Hafner et al. Mastering Diverse Domains through World Models. 2023.

NYU

Incorporating Rewards

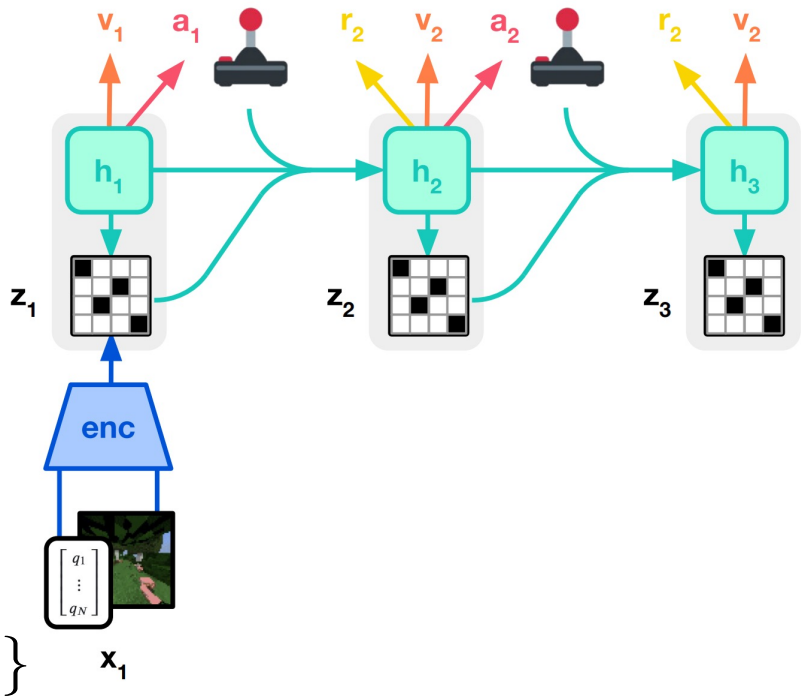
- How to use WM in planning?
  Predicting value by simulate a batch of trajectories.
- Actor-Critic RL:
  $a_t \sim \pi_\theta(a_t \mid s_t)$ $\quad$ $v_\psi(R_t \mid s_t)$
- Learning a critic:

Categorical distribution

$$\mathcal{L}_\phi = -\sum_{t=1}^{T} \log p_\phi(R_t^\lambda \mid s_t) \quad s_t = \{h_t, z_t\}$$

Sum of discounted future rewards

$$R_t^\lambda = r_t + \gamma c_t[(1-\lambda)v_t + \lambda R_{t+1}^\lambda] \qquad R_T^\lambda = v_T$$



Hafner et al. Mastering Diverse Domains through World Models. 2023.

NYU

Actor Learning

- Learn a policy network $a_t \sim \pi_\theta(a_t \mid s_t)$

Levine. Policy Gradients. Deep RL course. 2017.

NYU

Actor Learning

- Learn a policy network $a_t \sim \pi_\theta(a_t \mid s_t)$
- REINFORCE algorithm [Williams 1992]

Normalization

Entropy Regularizer

$$\mathcal{L}(\theta) = -\sum_{t=1}^{T} \mathrm{sg}((R_t^\lambda - v_\phi(s_t))/\max(1, S)) \log \pi_\theta(a_t \mid s_t) + \eta H[\pi_\theta(a_t \mid s_t)]$$

critic.

Levine. Policy Gradients. Deep RL course. 2017.

NYU

Actor Learning

- Learn a policy network $a_t \sim \pi_\theta(a_t \mid s_t)$
- REINFORCE algorithm [Williams 1992]

Normalization — Entropy Regularizer

$$\mathcal{L}(\theta) = -\sum_{t=1}^{T} \mathrm{sg}((R_t^\lambda - v_\phi(s_t))/\max(1, S)) \log \pi_\theta(a_t \mid s_t) + \eta H[\pi_\theta(a_t \mid s_t)]$$

- Notes on policy gradient:

Levine. Policy Gradients. Deep RL course. 2017.

NYU

Actor Learning

- Learn a policy network $a_t \sim \pi_\theta(a_t \mid s_t)$
- REINFORCE algorithm [Williams 1992]

$$\mathcal{L}(\theta) = -\sum_{t=1}^{T} \mathrm{sg}((R_t^\lambda - v_\phi(s_t))/\max(1, S)) \log \pi_\theta(a_t \mid s_t) + \eta H[\pi_\theta(a_t \mid s_t)]$$

Normalization

Entropy Regularizer

- Notes on policy gradient:

$$\pi_\theta(\tau) \nabla_\theta \log \pi_\theta(\tau) = \pi_\theta(\tau) \frac{\nabla_\theta \pi_\theta(\tau)}{\pi_\theta(\tau)} = \nabla_\theta \pi_\theta(\tau).$$

Levine. Policy Gradients. Deep RL course. 2017.

NYU

Actor Learning

- Learn a policy network $a_t \sim \pi_\theta(a_t \mid s_t)$
- REINFORCE algorithm [Williams 1992]

Normalization Entropy Regularizer

$$\mathcal{L}(\theta) = -\sum_{t=1}^{T} \mathrm{sg}((R_t^\lambda - v_\phi(s_t))/\max(1, S)) \log \pi_\theta(a_t \mid s_t) + \eta H[\pi_\theta(a_t \mid s_t)]$$

- Notes on policy gradient:

$$\pi_\theta(\tau)\nabla_\theta \log \pi_\theta(\tau) = \pi_\theta(\tau)\frac{\nabla_\theta \pi_\theta(\tau)}{\pi_\theta(\tau)} = \nabla_\theta \pi_\theta(\tau).$$

$$\pi_\theta(\tau) = \pi_\theta(s_1, a_1, \ldots, s_T, a_T) = p(s_1)\prod_{t=1}^{T} \pi_\theta(a_t|s_t)p(s_{t+1}|s_t, a_t).$$

Levine. Policy Gradients. Deep RL course. 2017.

NYU

Actor Learning

- Learn a policy network $a_t \sim \pi_\theta(a_t \mid s_t)$
- REINFORCE algorithm [Williams 1992]

Normalization — Entropy Regularizer

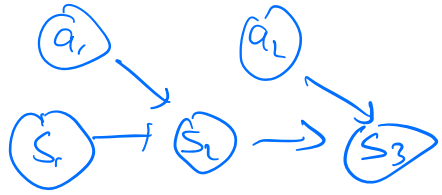
$$\mathcal{L}(\theta) = -\sum_{t=1}^{T} \mathrm{sg}((R_t^\lambda - v_\phi(s_t))/\max(1, S)) \log \pi_\theta(a_t \mid s_t) + \eta H[\pi_\theta(a_t \mid s_t)]$$

- Notes on policy gradient:

$$\pi_\theta(\tau)\nabla_\theta \log \pi_\theta(\tau) = \pi_\theta(\tau)\frac{\nabla_\theta \pi_\theta(\tau)}{\pi_\theta(\tau)} = \nabla_\theta \pi_\theta(\tau).$$

$$\pi_\theta(\tau) = \pi_\theta(s_1, a_1, \ldots, s_T, a_T) = p(s_1)\prod_{t=1}^{T} \pi_\theta(a_t|s_t)p(s_{t+1}|s_t, a_t).$$

$$\log \pi_\theta(\tau) = \log p(s_1) + \sum_{t=1}^{T} \log \pi_\theta(a_t|s_t) + \log p(s_{t+1}|s_t, a_t).$$



Levine. Policy Gradients. Deep RL course. 2017.

NYU

Actor Learning

- Learn a policy network $a_t \sim \pi_\theta(a_t \mid s_t)$
- REINFORCE algorithm [Williams 1992]

Normalization — Entropy Regularizer

$$\mathcal{L}(\theta) = -\sum_{t=1}^{T} \mathrm{sg}((R_t^\lambda - v_\phi(s_t))/\max(1, S)) \log \pi_\theta(a_t \mid s_t) + \eta H[\pi_\theta(a_t \mid s_t)]$$

- Notes on policy gradient:

$$\pi_\theta(\tau)\nabla_\theta \log \pi_\theta(\tau) = \pi_\theta(\tau)\frac{\nabla_\theta \pi_\theta(\tau)}{\pi_\theta(\tau)} = \nabla_\theta \pi_\theta(\tau).$$

$$\pi_\theta(\tau) = \pi_\theta(s_1, a_1, \ldots, s_T, a_T) = p(s_1)\prod_{t=1}^{T} \pi_\theta(a_t|s_t)p(s_{t+1}|s_t, a_t).$$

$$\log \pi_\theta(\tau) = \log p(s_1) + \sum_{t=1}^{T} \log \pi_\theta(a_t|s_t) + \log p(s_{t+1}|s_t, a_t).$$

$$\mathcal{L}(\theta) = \mathbb{E}_{\tau\sim\pi_\theta(\tau)}[r(\tau)] = \int \pi_\theta r(\tau) d\tau.$$

Levine. Policy Gradients. Deep RL course. 2017.

NYU

Actor Learning

- Learn a policy network $a_t \sim \pi_\theta(a_t \mid s_t)$
- REINFORCE algorithm [Williams 1992]

Normalization

Entropy Regularizer

$$\mathcal{L}(\theta) = -\sum_{t=1}^{T} \mathrm{sg}((R_t^\lambda - v_\phi(s_t))/\max(1, S)) \log \pi_\theta(a_t \mid s_t) + \eta H[\pi_\theta(a_t \mid s_t)]$$

- Notes on policy gradient:

$$\pi_\theta(\tau)\nabla_\theta \log \pi_\theta(\tau) = \pi_\theta(\tau)\frac{\nabla_\theta \pi_\theta(\tau)}{\pi_\theta(\tau)} = \nabla_\theta \pi_\theta(\tau).$$

$$\pi_\theta(\tau) = \pi_\theta(s_1, a_1, \ldots, s_T, a_T) = p(s_1)\prod_{t=1}^{T} \pi_\theta(a_t|s_t)p(s_{t+1}|s_t, a_t).$$

$$\log \pi_\theta(\tau) = \log p(s_1) + \sum_{t=1}^{T} \log \pi_\theta(a_t|s_t) + \log p(s_{t+1}|s_t, a_t).$$

$$\mathcal{L}(\theta) = \mathbb{E}_{\tau\sim\pi_\theta(\tau)}[r(\tau)] = \int \pi_\theta r(\tau) d\tau.$$

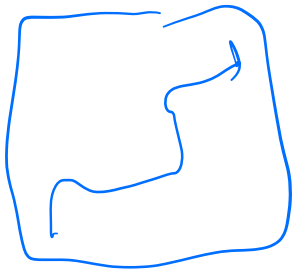
$$\nabla\mathcal{L}(\theta) = \int \nabla\pi_\theta r(\tau) d\tau = \int \pi_\theta(\tau)\nabla \log \pi_\theta r(\tau) d\tau = \mathbb{E}_{\tau\sim\pi}\nabla \log \pi_\theta r(\tau).$$

Levine. Policy Gradients. Deep RL course. 2017.

NYU

When Do We Need A Learned Actor?

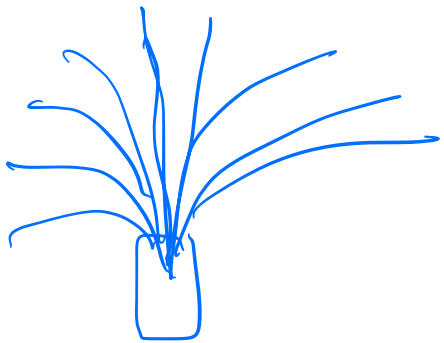
- For low dimensional or discrete problems, we can directly take the argmax of the value function.



NYU

When Do We Need A Learned Actor?

- For low dimensional or discrete problems, we can directly take the argmax of the value function.
- For problems with a good model, we can roll out and sample many future trajectories. Evaluation can be done in real time with GPU.



NYU

When Do We Need A Learned Actor?

- For low dimensional or discrete problems, we can directly take the argmax of the value function.
- For problems with a good model, we can roll out and sample many future trajectories. Evaluation can be done in real time with GPU.
- It compresses planning into a reactive policy.

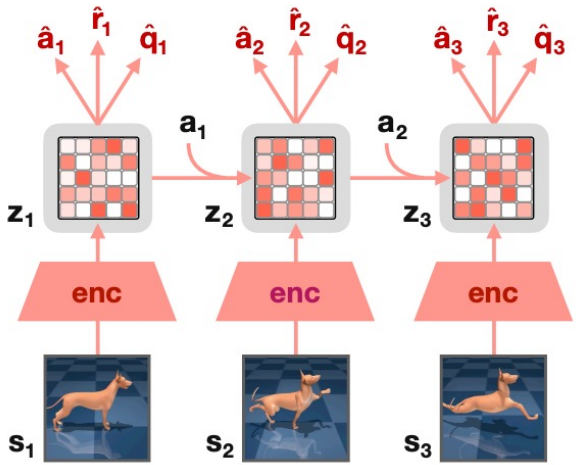
NYU

When Do We Need A Learned Actor?

- For low dimensional or discrete problems, we can directly take the argmax of the value function.
- For problems with a good model, we can roll out and sample many future trajectories. Evaluation can be done in real time with GPU.
- It compresses planning into a reactive policy.
- For general control problems, learning a separate actor can be a general solution without invoking domain knowledge.

Temporal Difference MPC

- Instead of learning an actor, we can directly update the actor network at test-time by optimizing the learned Q function from TD.



No reconstruction. JEPA.

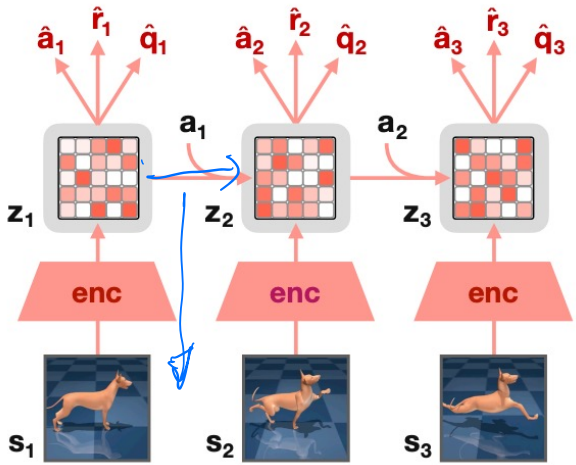
Hansen et al. Temporal Difference Learning for Model Predictive Control. ICML 2022.

NYU

Temporal Difference MPC

- Instead of learning an actor, we can directly update the actor network at test-time by optimizing the learned Q function from TD.
- Dynamics Learning:

$$z_t = h_\theta(x_t)$$
$$\hat{z}_{t+k} = d_\theta(\hat{z}_{t+k-1}, a_{t+k-1})$$



No reconstruction. JEPA.

Hansen et al. Temporal Difference Learning for Model Predictive Control. ICML 2022.

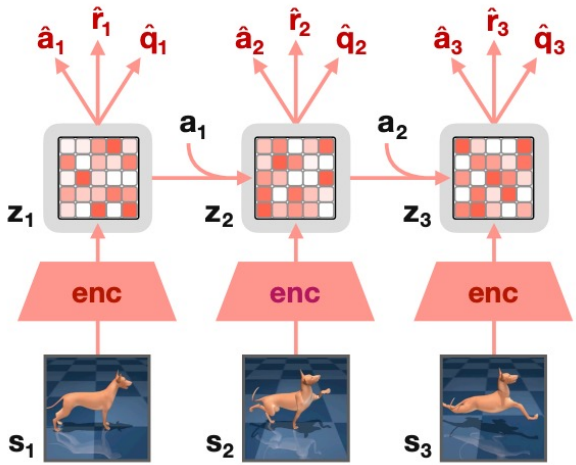
NYU

Temporal Difference MPC

- Instead of learning an actor, we can directly update the actor network at test-time by optimizing the learned Q function from TD.
- Dynamics Learning:

$$z_t = h_\theta(x_t)$$
$$\hat{z}_{t+k} = d_\theta(\hat{z}_{t+k-1}, a_{t+k-1})$$

- TD: $\hat{Q}(z_t, a_t) \approx r_t + \gamma \bar{Q}(z_{t+1}, a_{t+1})$



No reconstruction. JEPA.

Hansen et al. Temporal Difference Learning for Model Predictive Control. ICML 2022.

NYU

Temporal Difference MPC

- Instead of learning an actor, we can directly update the actor network at test-time by optimizing the learned Q function from TD.
- Dynamics Learning:

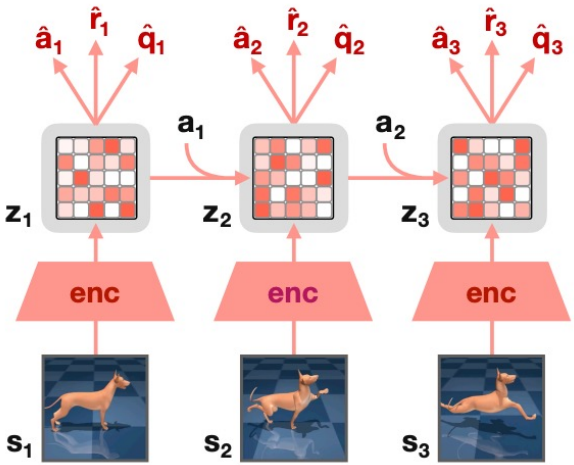
$$z_t = h_\theta(x_t)$$
$$\hat{z}_{t+k} = d_\theta(\hat{z}_{t+k-1}, a_{t+k-1})$$

- TD: $\hat{Q}(z_t, a_t) \approx r_t + \gamma \bar{Q}(z_{t+1}, a_{t+1})$
- Planning:

$$J(A) = \sum_{k=0}^{H-1} \gamma^k \hat{r}(z_{t+k}, a_{t+k}) + \gamma^H \hat{Q}(z_{t+H}, a_{t+H})$$

estimate reward within horizon.

$$\mu^* = \arg\min_{\mu} \left(\mathbb{E}_{A \sim \mathcal{N}(\mu, \Sigma)}[-J(A)] + \lambda D_{\mathrm{KL}}(\mathcal{N}(\mu, \Sigma) \parallel \mathbb{P}_{\mathrm{prior}}) \right)$$



No reconstruction. JEPA.

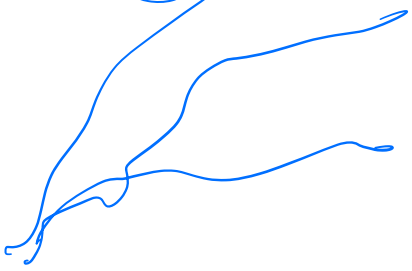
Hansen et al. Temporal Difference Learning for Model Predictive Control. ICML 2022.

NYU

Sample Diverse Trajectories From Learned Prior

- Initialize and sample from policy prior

$$a_{t.i} \sim \mathcal{N}(\mu_\theta(z_t), \sigma_\theta(z_t)^2)$$



Max-Entropy

$$\mathcal{L}_{\text{prior}} = -\mathbb{E}_{a \sim \pi_\theta(\cdot|z)} \left[\hat{Q}_\theta(z, a) - \alpha \log \pi_\theta(a|z) \right]$$

Hansen et al. TD-MPC2: Scalable, Robust World Models for Continuous Control. ICLR 2024.

NYU

Sample Diverse Trajectories From Learned Prior

- Initialize and sample from policy prior

$$a_{t.i} \sim \mathcal{N}(\mu_\theta(z_t), \sigma_\theta(z_t)^2)$$

Max-Entropy

$$\mathcal{L}_{\text{prior}} = -\mathbb{E}_{a \sim \pi_\theta(\cdot|z)} \left[\hat{Q}_\theta(z, a) - \alpha \log \pi_\theta(a|z) \right]$$

- Roll out

$$R(\tau_i) = \sum_{k=0}^{H-1} \gamma^k \hat{r}_{t+k,i} + \gamma^H \hat{Q}(\hat{z}_{t+H,i}, a_{t+H,i})$$

Hansen et al. TD-MPC2: Scalable, Robust World Models for Continuous Control. ICLR 2024.

NYU

Sample Diverse Trajectories From Learned Prior

- Initialize and sample from policy prior

$$a_{t.i} \sim \mathcal{N}(\mu_\theta(z_t), \sigma_\theta(z_t)^2)$$

Max-Entropy

$$\mathcal{L}_{\text{prior}} = -\mathbb{E}_{a\sim\pi_\theta(\cdot|z)}\left[\hat{Q}_\theta(z,a) - \alpha \log \pi_\theta(a|z)\right]$$

- Roll out

$$R(\tau_i) = \sum_{k=0}^{H-1} \gamma^k \hat{r}_{t+k,i} + \gamma^H \hat{Q}(\hat{z}_{t+H,i}, a_{t+H,i})$$

gradient-free.

- Weighted average smoothing and adaptation

$$w_i = \frac{\exp\left(\frac{1}{\lambda} R(\tau_i)\right)}{\sum_{j=1}^{N} \exp\left(\frac{1}{\lambda} R(\tau_j)\right)} \qquad a_t^{new} = \sum_{i=1}^{N} w_i a_{t,i}$$

Hansen et al. TD-MPC2: Scalable, Robust World Models for Continuous Control. ICLR 2024.

NYU

Latent MPC Using Gradient Descent

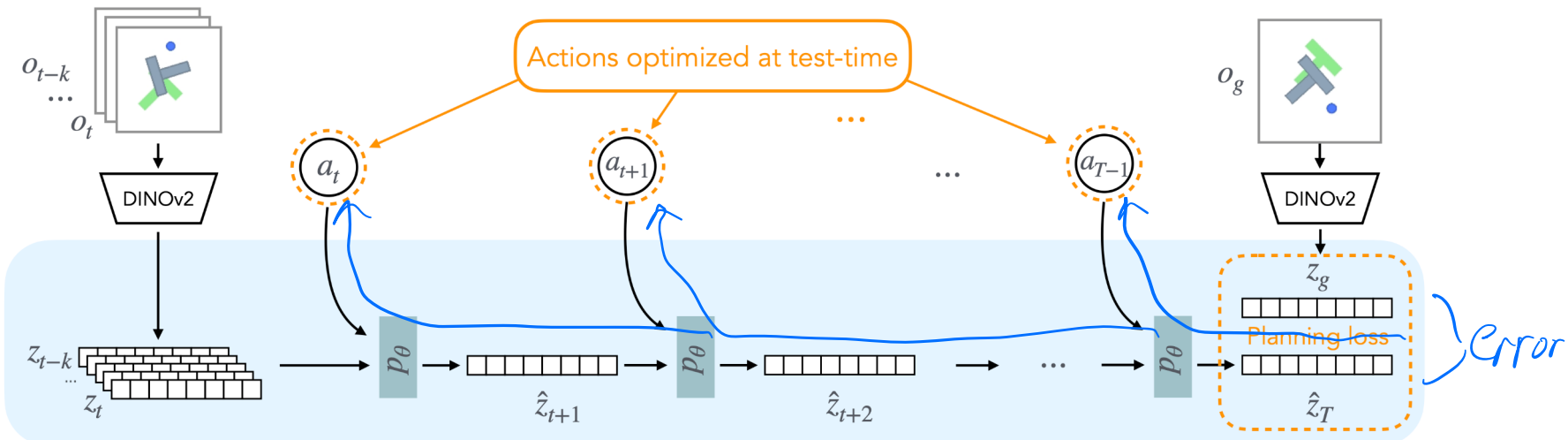
- MPC is a differentiable computation graph.
- We can directly optimize action variables.

Zhou et al. DINO-WM: World Models on Pre-trained Visual Features enable Zero-shot Planning. ICML 2025.

NYU

Latent MPC Using Gradient Descent

- MPC is a differentiable computation graph.
- We can directly optimize action variables.

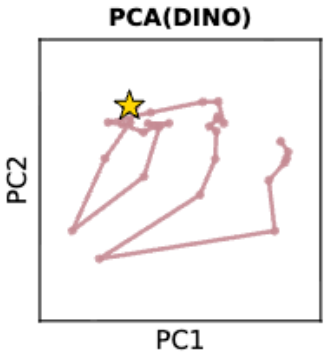


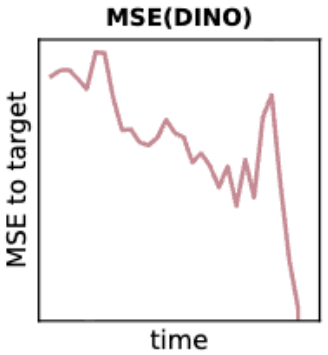
Zhou et al. DINO-WM: World Models on Pre-trained Visual Features enable Zero-shot Planning. ICML 2025.

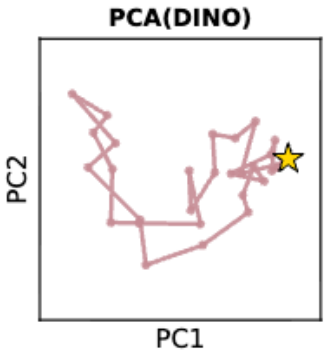
NYU

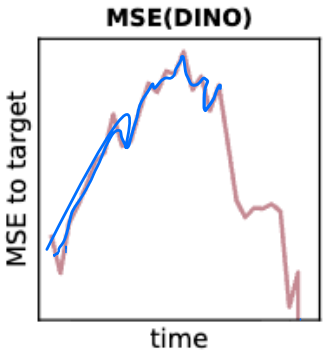
Learn Better Latent WM Representations

- The representation space is often jagged.
- Hard to optimize action sequences at test-time with gradient descent.







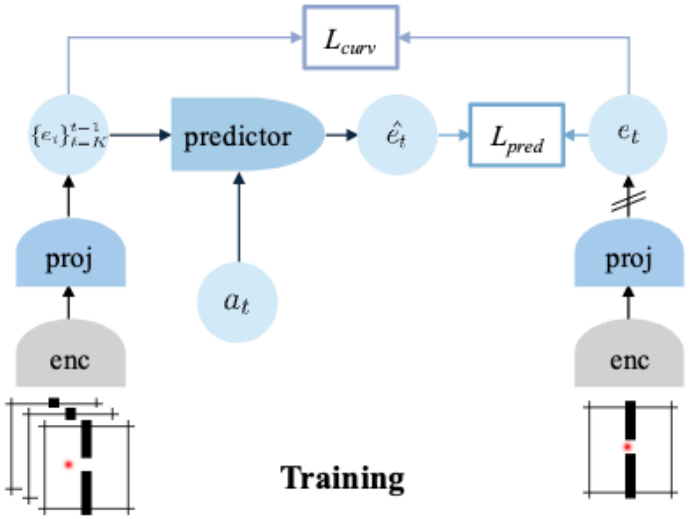


Wang et al. Temporal Straightening for Latent Planning. WMW 2026.

NYU

Straightening Objective

- In addition to predictors, maximizing cosine similarity across adjacent velocities (minimizing curvature).



$$\mathcal{L}_{pred} = ||\hat{z}_{t+1} - \text{sg}(z_{t+1})||_2^2.$$

JEPA

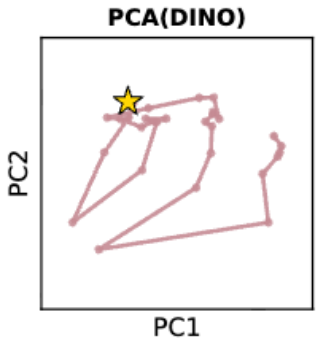
$$v_t = z_{t+1} - z_t, \quad v_{t+1} = z_{t+2} - z_{t+1}.$$

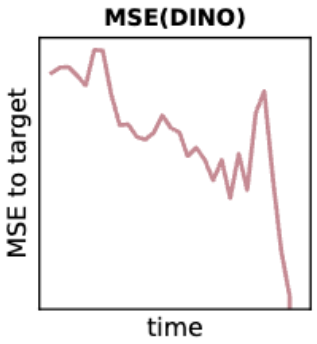
$$\mathcal{L}_{curve} = 1 - \frac{v_t \cdot v_{t+1}}{||v_t||_2 \cdot ||v_{t+1}||_2}.$$

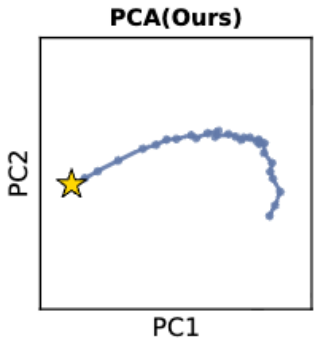
Wang et al. Temporal Straightening for Latent Planning. WMW 2026.

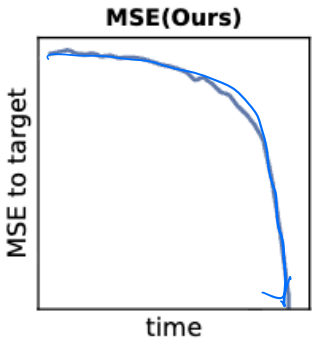
NYU

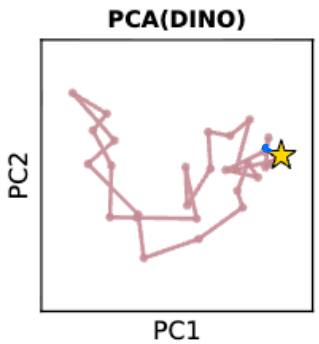
Straightened Representations

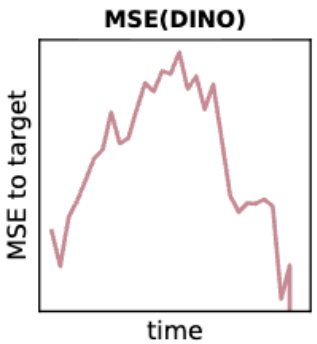


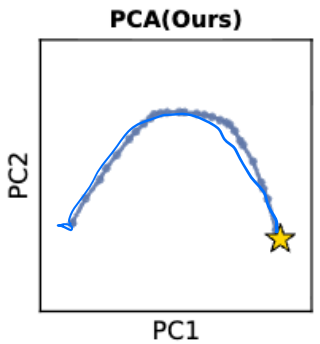


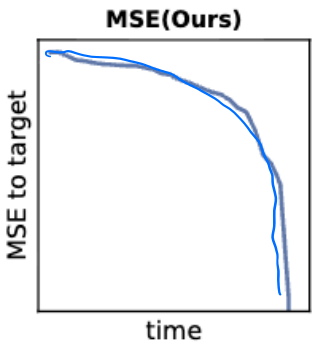












Wang et al. Temporal Straightening for Latent Planning. 2026.

NYU

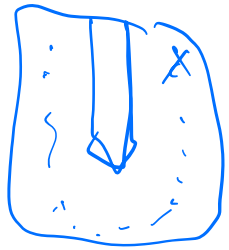
Why Straightening Helps?

- Encourages the predictor to be linear – more generalizable.

Wang et al. Temporal Straightening for Latent Planning. 2026.

NYU

Why Straightening Helps?

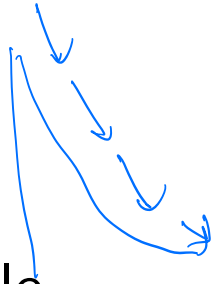


- Encourages the predictor to be linear – more generalizable.
- Shortest path in the state space (with obstacle) – shortest path in the latent space (without obstacle).

Wang et al. Temporal Straightening for Latent Planning. 2026.

NYU

Why Straightening Helps?



- Encourages the predictor to be linear – more generalizable.
- Shortest path in the state space (with obstacle) – shortest path in the latent space (without obstacle).
- Convexification effect:

$$\nabla_{\mathbf{a}} C(z_T(\mathbf{a})) = \left(\frac{\partial z_T}{\partial \mathbf{a}}\right)^T \nabla_z C(z_T)$$

$$H = \underbrace{\left(\frac{\partial z_T}{\partial \mathbf{a}}\right)^T \nabla_z^2 C(z_T) \left(\frac{\partial z_T}{\partial \mathbf{a}}\right)}_{\mathcal{T}_1\text{:Cost Convexity}} + \underbrace{\nabla_z C(z_T) \cdot \frac{\partial^2 z_T}{\partial \mathbf{a}^2}}_{\mathcal{T}_2\text{:Dynamics Curvature}}$$

Closer Zero.

Wang et al. Temporal Straightening for Latent Planning. 2026.

NYU

Why Straightening Helps?

- Encourages the predictor to be linear – more generalizable.
- Shortest path in the state space (with obstacle) – shortest path in the latent space (without obstacle).
- Convexification effect: $$\nabla_{\mathbf{a}} C(z_T(\mathbf{a})) = \left(\frac{\partial z_T}{\partial \mathbf{a}}\right)^T \nabla_z C(z_T)$$

$$H = \underbrace{\left(\frac{\partial z_T}{\partial \mathbf{a}}\right)^T \nabla_z^2 C(z_T) \left(\frac{\partial z_T}{\partial \mathbf{a}}\right)}_{\mathcal{T}_1\text{:Cost Convexity}} + \underbrace{\nabla_z C(z_T) \cdot \frac{\partial^2 z_T}{\partial \mathbf{a}^2}}_{\mathcal{T}_2\text{:Dynamics Curvature}}$$

O(1) wrt time.

- Conditioning effect: Convergence of convex gradient optimization depends on κ (condition number). For straight trajectories, condition number is constant.

$$\|\mathbf{a}^{(k)} - \mathbf{a}^*\| \leq \left(\frac{\kappa - 1}{\kappa + 1}\right)^k \|\mathbf{a}^{(0)} - \mathbf{a}^*\|$$

Wang et al. Temporal Straightening for Latent Planning. 2026.

NYU

Summary: World Models

- Explicit object representation
  - Traditional, light weight, instance-specific, hard to learn jointly

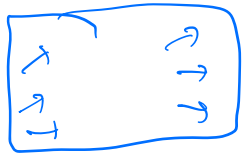
WM for external actors.
Low-dim spaces

WM for simple, single actor environment, visualization

WM for general planning and test-time MPC

NYU

Summary: World Models



- Explicit object representation
  - Traditional, light weight, instance-specific, hard to learn jointly
- Differentiable occupancy, motion field
  - Relatively heavy, spatially grounded, end-to-end learnable

WM for external actors. Low-dim spaces

WM for simple, single actor environment, visualization

WM for general planning and test-time MPC

NYU

Summary: World Models

- Explicit object representation
  - Traditional, light weight, instance-specific, hard to learn jointly
- Differentiable occupancy, motion field
  - Relatively heavy, spatially grounded, end-to-end learnable
- Raw video/3D prediction
  - Expensive, good for simulation and safety for short horizons
  - Good for single actor environments

WM for external actors. Low-dim spaces

WM for simple, single actor environment, visualization

WM for general planning and test-time MPC

NYU

Summary: World Models

- Explicit object representation
  - Traditional, light weight, instance-specific, hard to learn jointly
- Differentiable occupancy, motion field
  - Relatively heavy, spatially grounded, end-to-end learnable

WM for external actors. Low-dim spaces

- Raw video/3D prediction
  - Expensive, good for simulation and safety for short horizons
  - Good for single actor environments

WM for simple, single actor environment, visualization

- Global latent, RNNs
  - General-purpose, flexible
  - Planning-aware representations

Collect TD reward and train value function / critic.

Model predictive control.

WM for general planning and test-time MPC

NYU

Learnable Planning and Cost

NYU

Imitation Learning

- The explicit policy model, supervised learning (behavior cloning)

$$\hat{a} = f_\theta(x) \qquad \mathcal{L} = \min_i \|a_i - \hat{a}\|_2^2 \qquad \mathcal{L} = -\log \hat{a}_j$$

Imitation Learning

- The explicit policy model, supervised learning (behavior cloning)

$$\hat{a} = f_\theta(x) \qquad \mathcal{L} = \min_i \|a_i - \hat{a}\|_2^2 \qquad \mathcal{L} = -\log \hat{a}_j$$

- Energy-based (cost-based) approach

$$\tau^\star = \operatorname{argmin}_\tau E(x, \tau)$$

$$p(\tau \mid x) = \frac{\exp(E(x,\tau))}{\int_\tau \exp(E(x,\tau))}$$

NYU

Imitation Learning

- The explicit policy model, supervised learning (behavior cloning)

$$\hat{a} = f_\theta(x) \qquad \mathcal{L} = \min_i \|a_i - \hat{a}\|_2^2 \qquad \mathcal{L} = -\log \hat{a}_j$$

- Energy-based (cost-based) approach

$$\tau^\star = \operatorname{argmin}_\tau E(x, \tau)$$

$$p(\tau \mid x) = \frac{\exp(E(x,\tau))}{\int_\tau \exp(E(x,\tau))}$$

- Dataset Aggregation (DAgger)
  - Learned policy may deviate from experts
  - Need to collect more groundtruths

Initialize $\mathcal{D} \leftarrow \emptyset$.
Initialize $\hat{\pi}_1$ to any policy in Π.
for $i = 1$ **to** N **do**
  Let $\pi_i = \beta_i \pi^* + (1 - \beta_i)\hat{\pi}_i$.
  Sample T-step trajectories using π_i.
  Get dataset $\mathcal{D}_i = \{(s, \pi^*(s))\}$ of visited states by π_i and actions given by expert.
  Aggregate datasets: $\mathcal{D} \leftarrow \mathcal{D} \bigcup \mathcal{D}_i$.
  Train classifier $\hat{\pi}_{i+1}$ on $\mathcal{D}$.
end for
Return best $\hat{\pi}_i$ on validation.

Algorithm 3.1: DAGGER Algorithm.

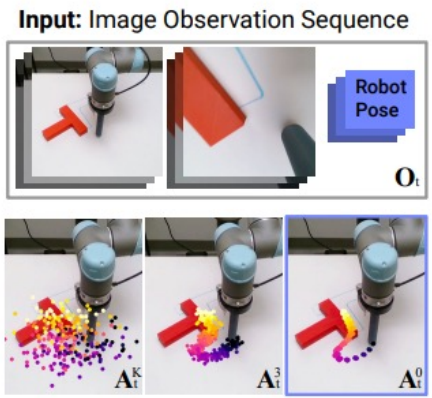
NYU

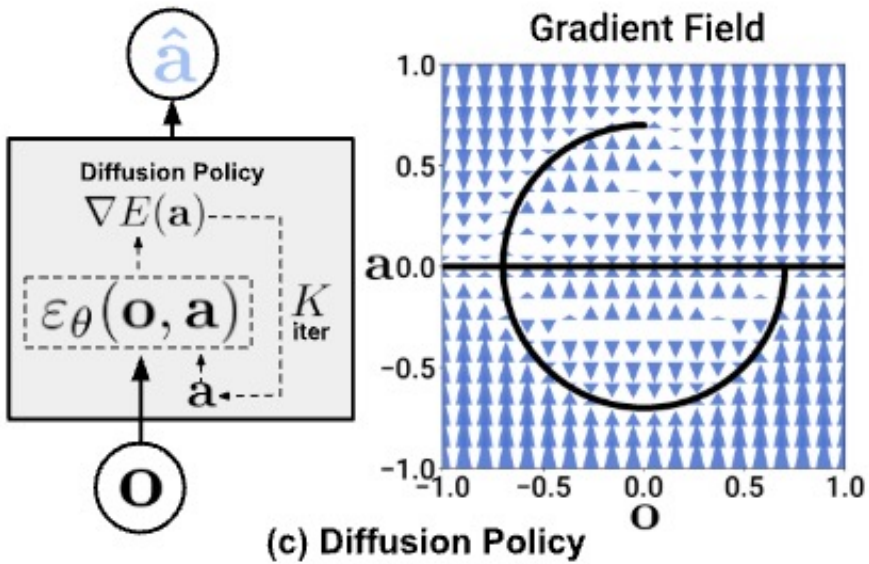
Direct Policy Learning from Diffusion

- Error prediction network is conditioned on observation features.

$$A_t^{k-1} = \alpha(A_t^k - \gamma\epsilon_\theta(O_t, A_t^k, k) + \mathcal{N}(0, \sigma^2 I)).$$

$$\mathcal{L} = MSE(\epsilon^k, \epsilon_\theta(O_t, A_t + \epsilon^k, k)).$$





Chi et al. Diffusion Policy: Visuomotor Policy Learning via Action Diffusion. RSS 2023

NYU

Learning Cost

- Typically, we have some rough ideas on what the cost should look like
  - E.g. Avoid obstacle with squared distance barrier.

NYU

Learning Cost

- Typically, we have some rough ideas on what the cost should look like
  - E.g. Avoid obstacle with squared distance barrier.
- But there are costs that are implicitly defined.
  - E.g. Human comfort

NYU

Learning Cost

- Typically, we have some rough ideas on what the cost should look like
  - E.g. Avoid obstacle with squared distance barrier.
- But there are costs that are implicitly defined.
  - E.g. Human comfort
- Can be learned from human demonstrations (also called IRL)

NYU

Max-Margin Planning with Explicit Cost Volume

- If we have an explicit cost volume, the cost of a trajectory can be directly queried.

Zeng et al. End-to-end Interpretable Neural Motion Planner. CVPR 2019.

NYU

Max-Margin Planning with Explicit Cost Volume

- If we have an explicit cost volume, the cost of a trajectory can be directly queried.
- We can use the max-margin objective to make the groundtruth trajectory have lower costs.

$$\underset{\theta}{\operatorname{argmin}} \sum_{\{(\hat{x}_i^t, \hat{y}_i^t)\}_{i=1\ldots N}} \max_i \sum_{t=1}^{T} C_\theta^t[x_t, y_t] - C_\theta^t[\hat{x}_i^t, \hat{y}_i^t] + d_i^t$$

Zeng et al. End-to-end Interpretable Neural Motion Planner. CVPR 2019.

NYU

Max-Margin Planning with Explicit Cost Volume

- If we have an explicit cost volume, the cost of a trajectory can be directly queried.
- We can use the max-margin objective to make the groundtruth trajectory have lower costs.
- Find the lowest cost trajectory among a batch of samples.

$$\underset{\theta}{\operatorname{argmin}} \sum_{\{(\hat{x}_i^t, \hat{y}_i^t)\}_{i=1\ldots N}} \max_i \sum_{t=1}^{T} C_\theta^t[x_t, y_t] - C_\theta^t[\hat{x}_i^t, \hat{y}_i^t] + d_i^t$$

Zeng et al. End-to-end Interpretable Neural Motion Planner. CVPR 2019.

NYU

Max-Margin Planning with Explicit Cost Volume

- If we have an explicit cost volume, the cost of a trajectory can be directly queried.
- We can use the max-margin objective to make the groundtruth trajectory have lower costs.
- Find the lowest cost trajectory among a batch of samples.
- Low-dimensional/known dynamics problems: External samplers

$$\underset{\theta}{\operatorname{argmin}} \sum_{\{(\hat{x}_i^t, \hat{y}_i^t)\}_{i=1\dots N}} \max_i \sum_{t=1}^{T} C_\theta^t[x_t, y_t] - C_\theta^t[\hat{x}_i^t, \hat{y}_i^t] + d_i^t$$

Zeng et al. End-to-end Interpretable Neural Motion Planner. CVPR 2019.

NYU

Max-Margin Planning with Explicit Cost Volume

- If we have an explicit cost volume, the cost of a trajectory can be directly queried.
- We can use the max-margin objective to make the groundtruth trajectory have lower costs.
- Find the lowest cost trajectory among a batch of samples.
- Low-dimensional/known dynamics problems: External samplers
- In general, needs to perform optimization (e.g. DP)

$$\underset{\theta}{\operatorname{argmin}} \sum_{\{(\hat{x}_i^t, \hat{y}_i^t)\}_{i=1\ldots N}} \max_i \sum_{t=1}^{T} C_\theta^t[x_t, y_t] - C_\theta^t[\hat{x}_i^t, \hat{y}_i^t] + d_i^t$$

Zeng et al. End-to-end Interpretable Neural Motion Planner. CVPR 2019.

NYU

EBM Planning

- Energy-based framework also needs negative samples.

Florence et al. Implicit Behavioral Cloning. CoRL 2021.

NYU

EBM Planning

- Energy-based framework also needs negative samples.
- “Pick” the groundtruth sample among others.

Florence et al. Implicit Behavioral Cloning. CoRL 2021.

NYU

EBM Planning

- Energy-based framework also needs negative samples.
- “Pick” the groundtruth sample among others.
- If there isn’t an external sampler, we can either use autoregressive energy of sampling one dimension at a time, or gradient-based Langevin MCMC.

Florence et al. Implicit Behavioral Cloning. CoRL 2021.

NYU

EBM Planning

- Energy-based framework also needs negative samples.
- “Pick” the groundtruth sample among others.
- If there isn’t an external sampler, we can either use autoregressive energy of sampling one dimension at a time, or gradient-based Langevin MCMC.

Langevin MCMC: $\tilde{\mathbf{y}}_i^k = \tilde{\mathbf{y}}_i^{k-1} - \lambda \left(\frac{1}{2} \nabla_{\mathbf{y}} E_\theta(\mathbf{x}_i, \mathbf{y}_i^{k-1}) + \omega^k \right), \omega^k \sim \mathcal{N}(0, \sigma).$

Florence et al. Implicit Behavioral Cloning. CoRL 2021.

NYU

EBM Planning

- Energy-based framework also needs negative samples.
- “Pick” the groundtruth sample among others.
- If there isn’t an external sampler, we can either use autoregressive energy of sampling one dimension at a time, or gradient-based Langevin MCMC.

Langevin MCMC: $\tilde{\mathbf{y}}_i^k = \tilde{\mathbf{y}}_i^{k-1} - \lambda\left(\frac{1}{2}\nabla_{\mathbf{y}} E_\theta(\mathbf{x}_i, \mathbf{y}_i^{k-1}) + \omega^k\right), \omega^k \sim \mathcal{N}(0, \sigma).$

Loss: $\mathcal{L} = \sum_i -\log(p_\theta(\mathbf{y}_i \mid \mathbf{x}, \{\tilde{\mathbf{y}}_i\}_j) \qquad p_\theta(\mathbf{y}_i \mid \mathbf{x}, \{\tilde{\mathbf{y}}_i\}_j = \frac{e^{-E_\theta(\mathbf{x}_i, \mathbf{y}_i)}}{e^{-E_\theta(\mathbf{x}_i, \mathbf{y}_i)} + \sum_j e^{-E_\theta(\mathbf{x}_i + \tilde{\mathbf{y}}_{i,j})}}$

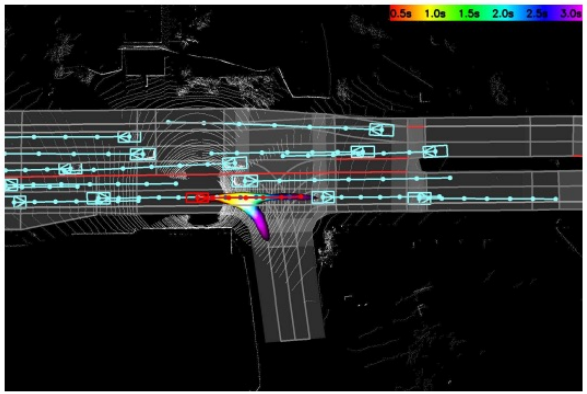
Florence et al. Implicit Behavioral Cloning. CoRL 2021.

NYU

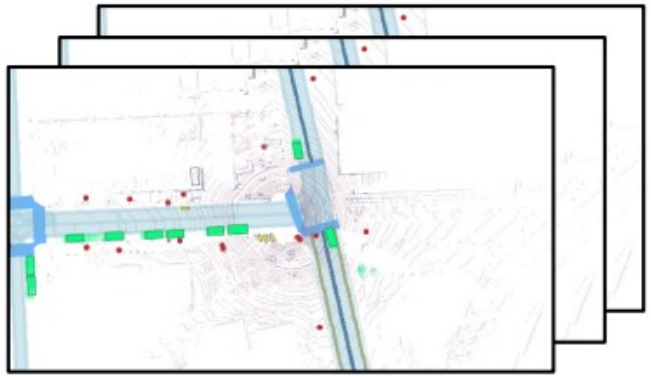
Non-Parametric Cost Volume for 2D Planning

- Interpretability (both costs and planner inputs)

Rasterization



Semantic Occupancy

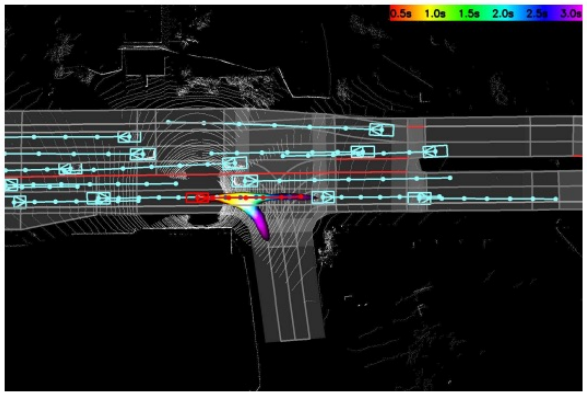


NYU

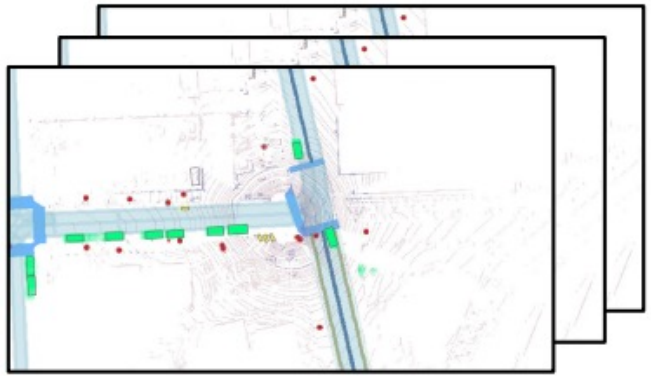
Non-Parametric Cost Volume for 2D Planning

- Interpretability (both costs and planner inputs)
- Use spatial geometry to form cost from explicit objects

Rasterization



Semantic Occupancy

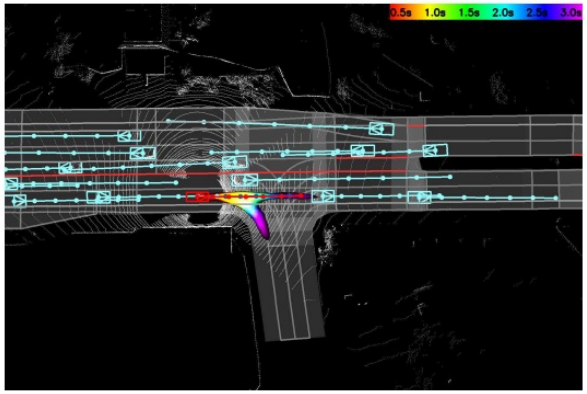


NYU

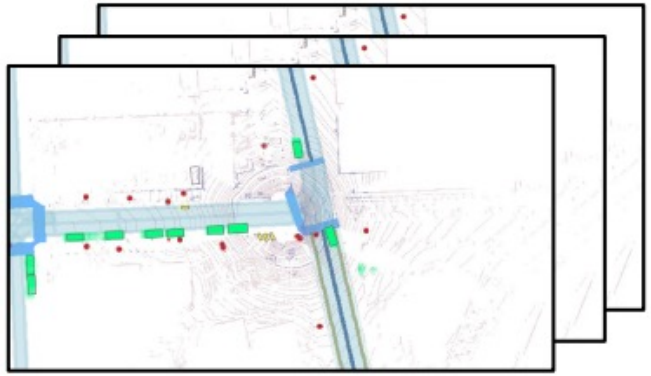
Non-Parametric Cost Volume for 2D Planning

- Interpretability (both costs and planner inputs)
- Use spatial geometry to form cost from explicit objects
- Predict spatial cost volume
  - Rasterize the scene for spatial inputs
  - Predict soft occupancy volumes (present and future)

Rasterization



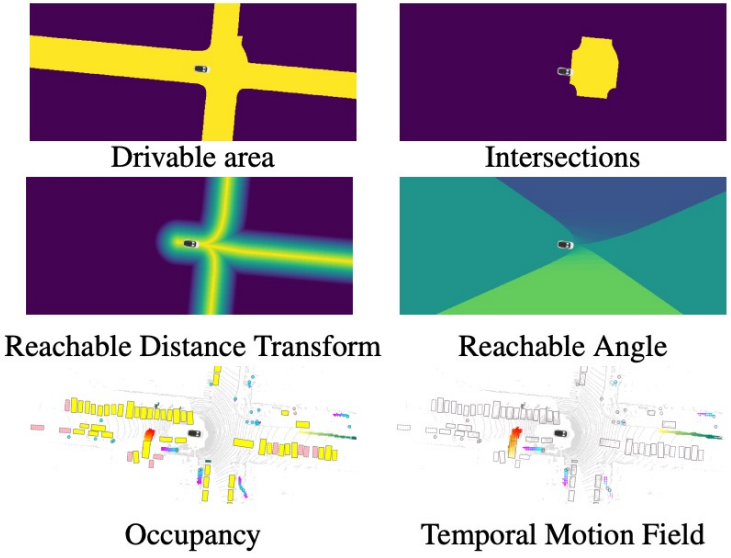
Semantic Occupancy

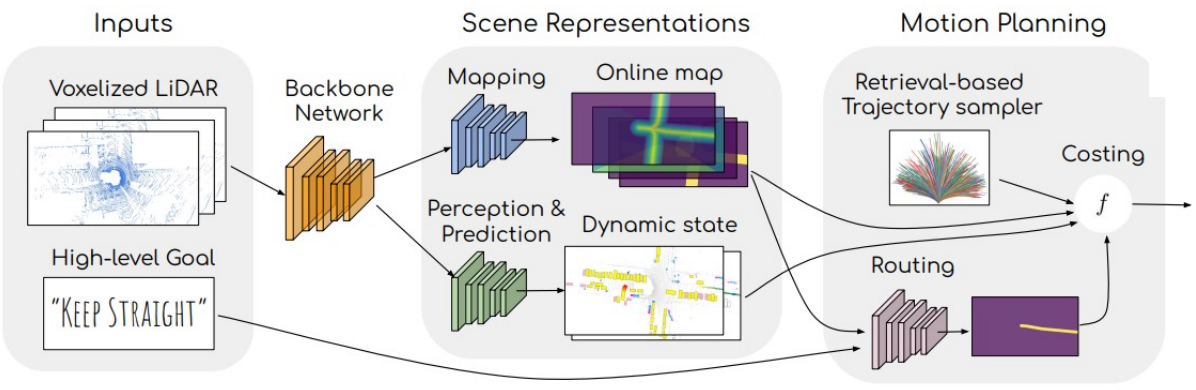


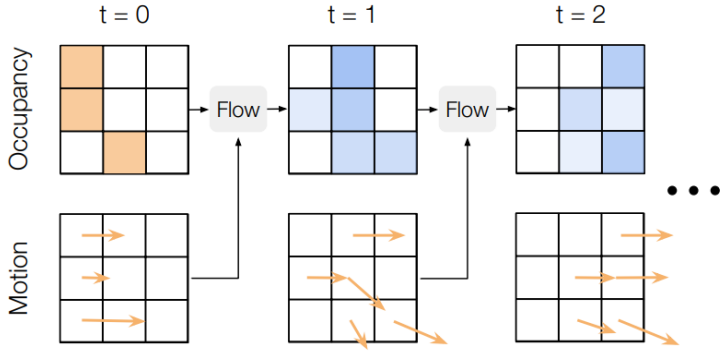
NYU

Learning Through Interpretable Predictions

- Semantic occupancy, motion field, mapping, etc. as intermediate predictions.



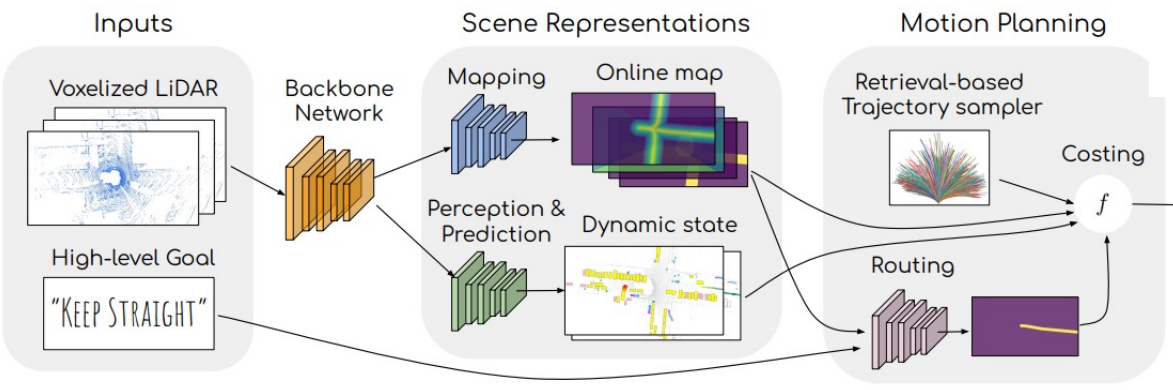


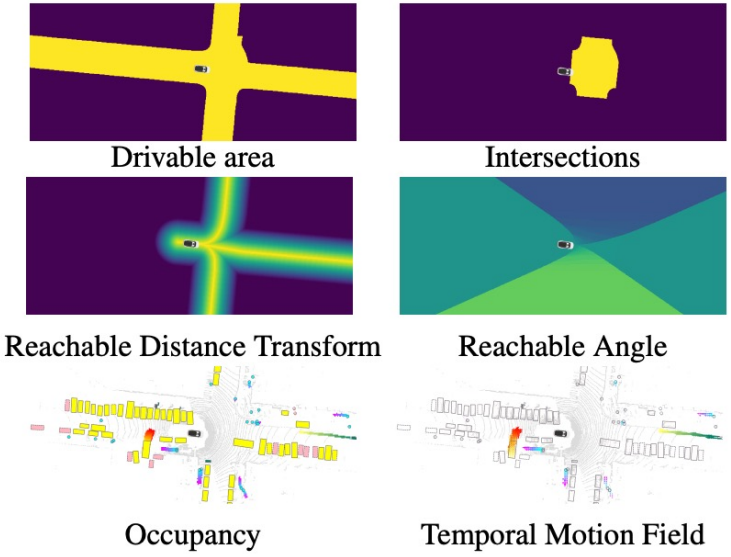


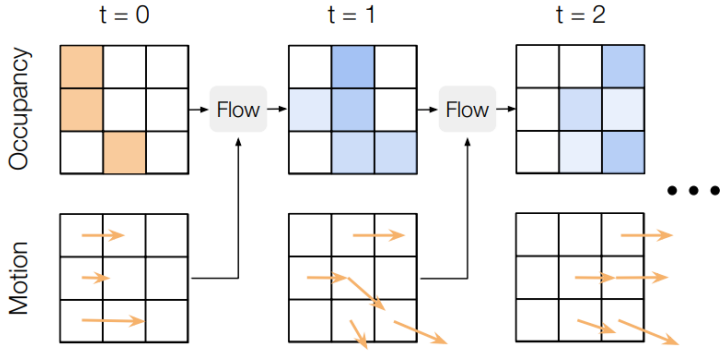
Casas et al. MP3: A Unified Model to Map, Perceive, Predict and Plan. CVPR 2021.

Learning Through Interpretable Predictions

- Semantic occupancy, motion field, mapping, etc. as intermediate predictions.
- Differentiable, supports end-to-end interpretable learning from perception to planning.







Casas et al. MP3: A Unified Model to Map, Perceive, Predict and Plan. CVPR 2021.

Value Iteration Networks

- A network design for predicting cost volumes that are grounded from the classic value iteration algorithm.

Tamar et al. Value Iteration Networks. NeurIPS 2016.

NYU

Value Iteration Networks

- A network design for predicting cost volumes that are grounded from the classic value iteration algorithm.
- Classic VI altorithm:

Tamar et al. Value Iteration Networks. NeurIPS 2016.

NYU

Value Iteration Networks

- A network design for predicting cost volumes that are grounded from the classic value iteration algorithm.
- Classic VI altorithm:

$$V^*(s) = \max_\pi V^\pi(s)$$

Tamar et al. Value Iteration Networks. NeurIPS 2016.

NYU

Value Iteration Networks

- A network design for predicting cost volumes that are grounded from the classic value iteration algorithm.
- Classic VI altorithm:

$$V^*(s) = \max_\pi V^\pi(s)$$

$$V^\pi(s) = \mathbb{E}^\pi \sum_{t=0}^{\infty} \gamma^t r(s_t, a_t)$$

Tamar et al. Value Iteration Networks. NeurIPS 2016.

NYU

Value Iteration Networks

- A network design for predicting cost volumes that are grounded from the classic value iteration algorithm.
- Classic VI altorithm:

$$V^*(s) = \max_\pi V^\pi(s)$$

$$V^\pi(s) = \mathbb{E}^\pi \sum_{t=0}^{\infty} \gamma^t r(s_t, a_t)$$

$$Q_n(s,a) = R(s,a) + \gamma \sum_{s'} P(s'|s,a) V_n(s')$$

Tamar et al. Value Iteration Networks. NeurIPS 2016.

NYU

Value Iteration Networks

- A network design for predicting cost volumes that are grounded from the classic value iteration algorithm.
- Classic VI altorithm:

$$V^*(s) = \max_\pi V^\pi(s)$$

$$V^\pi(s) = \mathbb{E}^\pi \sum_{t=0}^{\infty} \gamma^t r(s_t, a_t)$$

$$Q_n(s,a) = R(s,a) + \gamma \sum_{s'} P(s'|s,a) V_n(s')$$

$$V_{n+1}(s) = \max_a Q_n(s,a)$$

Tamar et al. Value Iteration Networks. NeurIPS 2016.

NYU

Value Iteration Networks

- A network design for predicting cost volumes that are grounded from the classic value iteration algorithm.
- Classic VI altorithm:

$$V^*(s) = \max_\pi V^\pi(s)$$

$$V^\pi(s) = \mathbb{E}^\pi \sum_{t=0}^{\infty} \gamma^t r(s_t, a_t)$$

$$Q_n(s,a) = R(s,a) + \gamma \sum_{s'} P(s'|s,a) V_n(s')$$

$$V_{n+1}(s) = \max_a Q_n(s,a)$$

$$\pi^*(s) = \operatorname{argmax}_a Q_\infty(s,a)$$

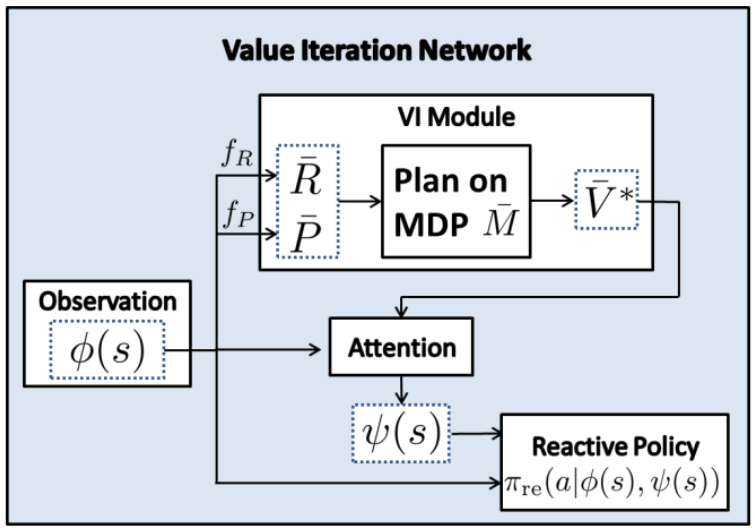
Tamar et al. Value Iteration Networks. NeurIPS 2016.

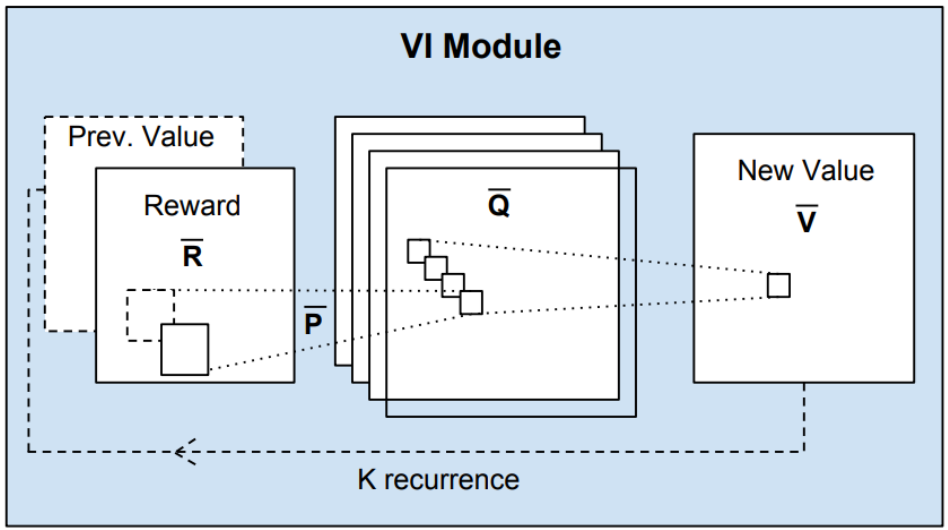
NYU

Value Iteration Networks

- Reward and previous value are fed into a network to generate Q of A channels. Transition matrix is convolutional kernel. Then Max-Pooling.

$$Q_n(s,a) = R(s,a) + \gamma \sum_{s'} P(s'|s,a) V_n(s') \quad V_{n+1}(s) = \max_a Q_n(s,a)$$





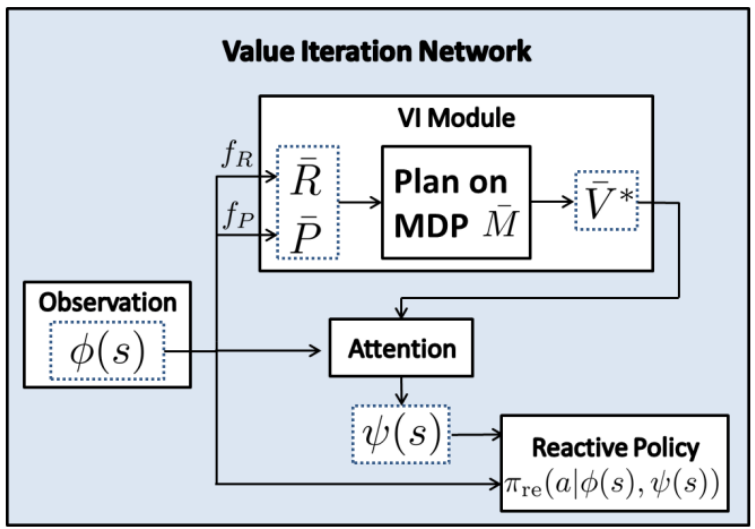
Tamar et al. Value Iteration Networks. NeurIPS 2016.

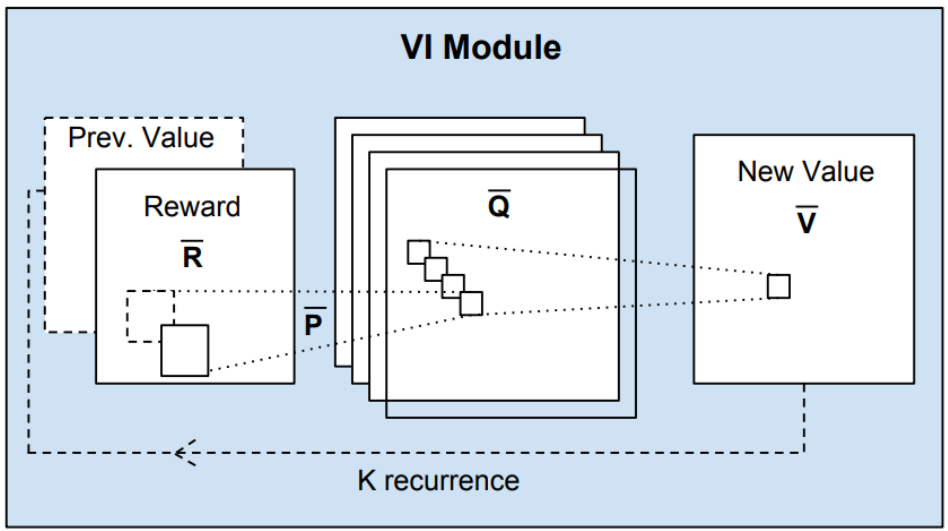
NYU

Value Iteration Networks

- Select the current state and choose an action from softmax.

$$\hat{a} \sim \mathrm{softmax}_a(Q(s,a))$$



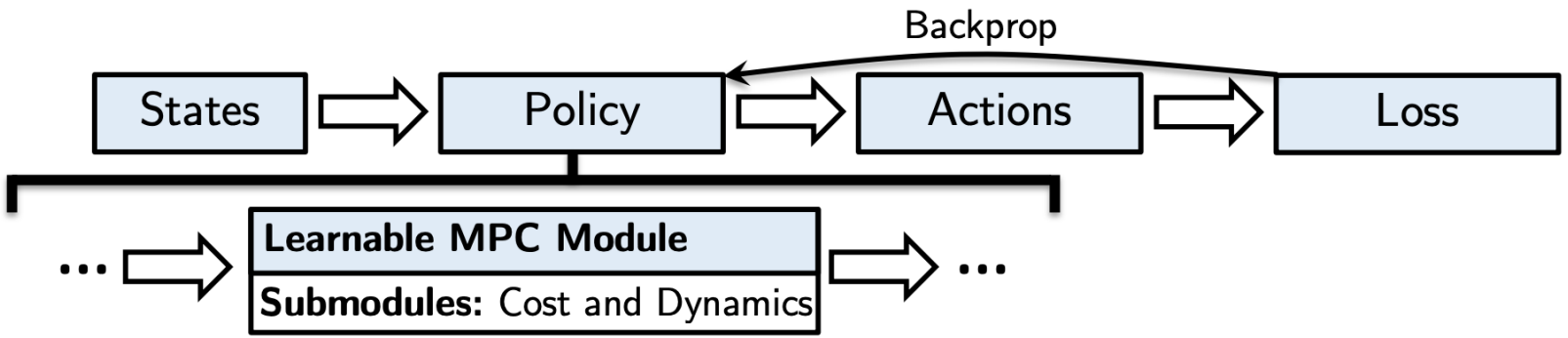


Tamar et al. Value Iteration Networks. NeurIPS 2016.

NYU

Backprop through Planning

- Treat planning as an end-to-end layer. Can be used for RL/Imitation.

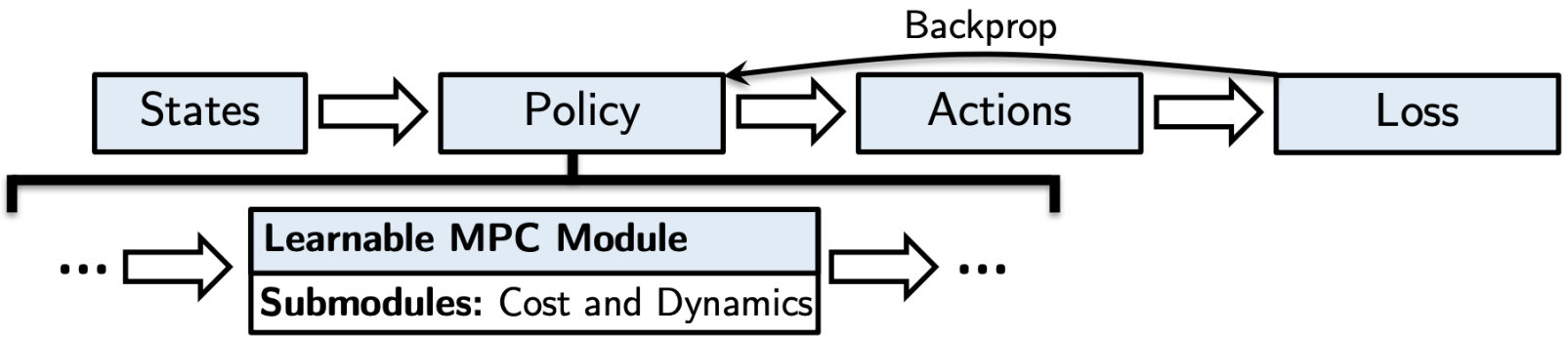


Amos et al. Differentiable MPC for End-to-end Planning and Control. NeurIPS 2018.

NYU

Backprop through Planning

- Treat planning as an end-to-end layer. Can be used for RL/Imitation.
- Option 1: Unrolling a finite number of steps

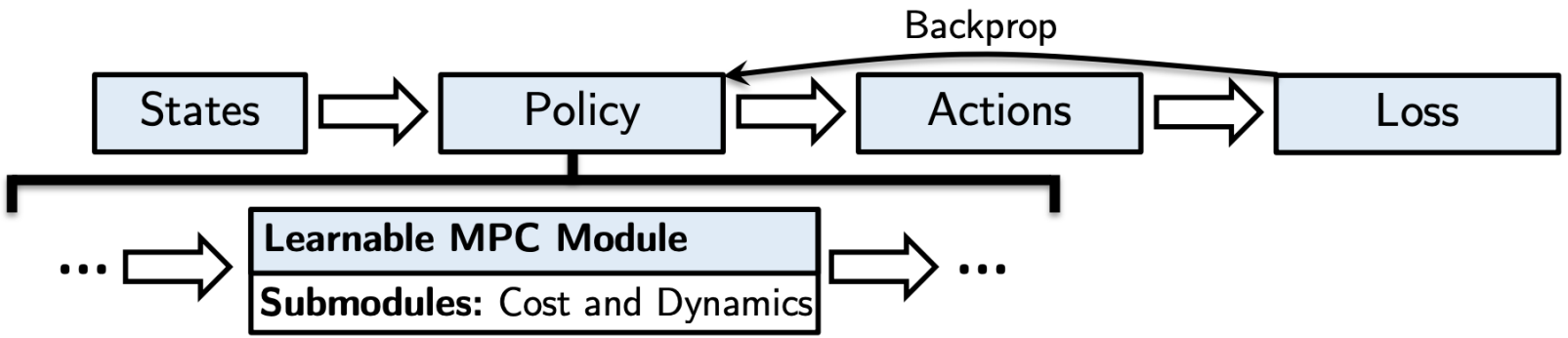


Amos et al. Differentiable MPC for End-to-end Planning and Control. NeurIPS 2018.

NYU

Backprop through Planning

- Treat planning as an end-to-end layer. Can be used for RL/Imitation.
- Option 1: Unrolling a finite number of steps
- Option 2: Solve till convergence, backprop for a finite step

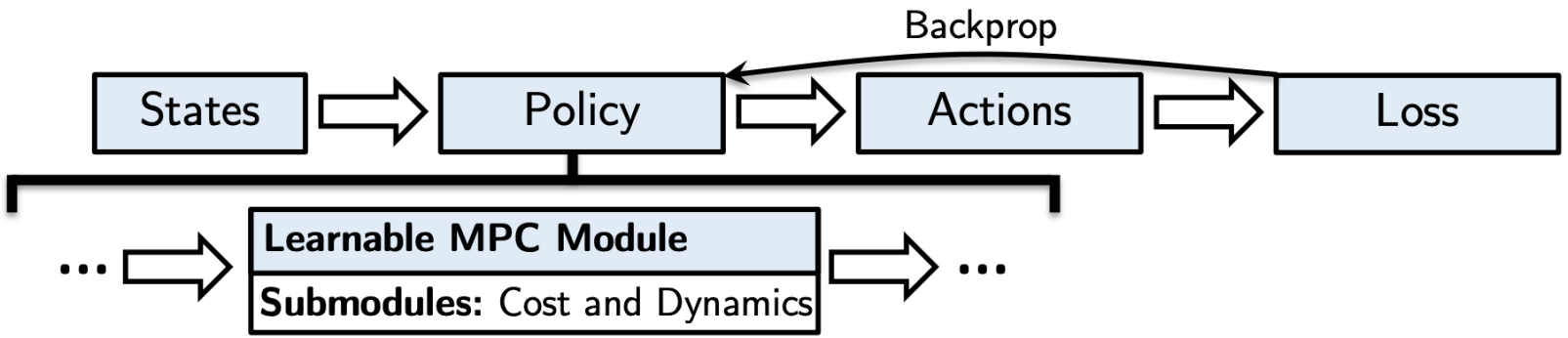


Amos et al. Differentiable MPC for End-to-end Planning and Control. NeurIPS 2018.

NYU

Backprop through Planning

- Treat planning as an end-to-end layer. Can be used for RL/Imitation.
- Option 1: Unrolling a finite number of steps
- Option 2: Solve till convergence, backprop for a finite step
- Option 3: Converged at fixed point: Implicit differentiation



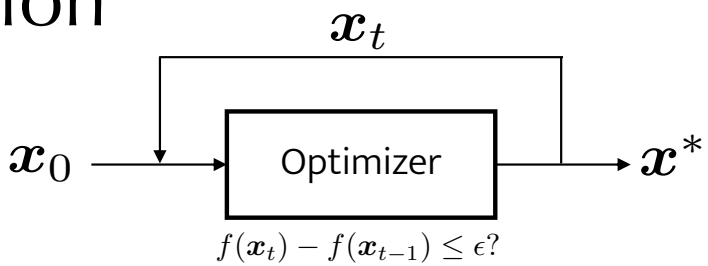
Amos et al. Differentiable MPC for End-to-end Planning and Control. NeurIPS 2018.

NYU

Implicit Differentiation

- Unconstrained case

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}}{\operatorname{argmin}} f(\boldsymbol{x}; \boldsymbol{\theta}).$$



Liao et al. Reviving and Improving Recurrent Back-Propagation. ICML 2018.

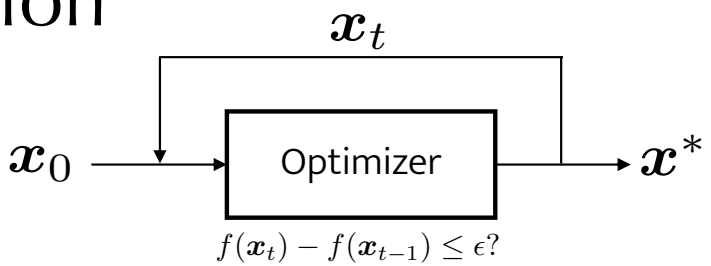
NYU

Implicit Differentiation

- Unconstrained case

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}}{\operatorname{argmin}} f(\boldsymbol{x}; \boldsymbol{\theta}).$$

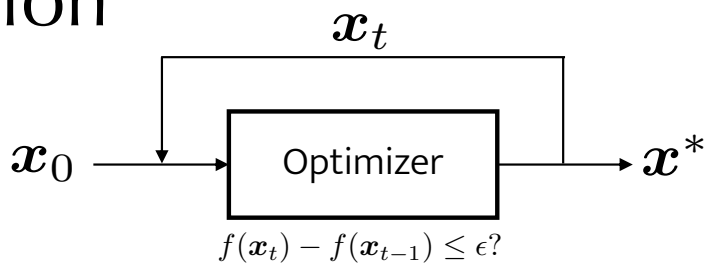
$$\mathbf{0} = \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}} J_{f,\boldsymbol{x}^*}(\boldsymbol{x}^*; \boldsymbol{\theta})$$



Liao et al. Reviving and Improving Recurrent Back-Propagation. ICML 2018.

NYU

Implicit Differentiation



- Unconstrained case

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}}{\operatorname{argmin}} f(\boldsymbol{x}; \boldsymbol{\theta}).$$

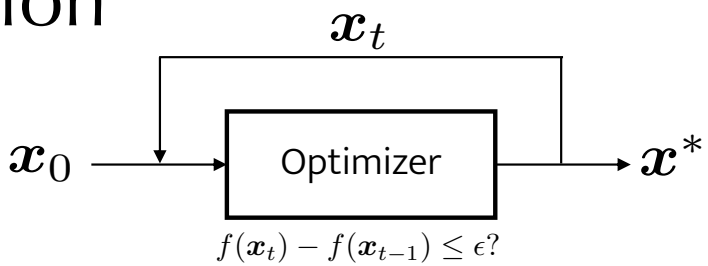
$$\mathbf{0} = \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}} J_{f,\boldsymbol{x}^*}(\boldsymbol{x}^*; \boldsymbol{\theta})$$

$$\mathbf{0} = \frac{\partial}{\partial \boldsymbol{x}^*} J_{f,\boldsymbol{x}^*}(\boldsymbol{x}^*; \boldsymbol{\theta}) \frac{\partial \boldsymbol{x}^*}{\partial \boldsymbol{\theta}} + \frac{\partial}{\partial \boldsymbol{\theta}} J_{f,\boldsymbol{x}^*}(\boldsymbol{x}^*; \boldsymbol{\theta})$$

Liao et al. Reviving and Improving Recurrent Back-Propagation. ICML 2018.

NYU

Implicit Differentiation



- Unconstrained case

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}}{\operatorname{argmin}} f(\boldsymbol{x}; \boldsymbol{\theta}).$$

$$\mathbf{0} = \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}} J_{f,\boldsymbol{x}^*}(\boldsymbol{x}^*; \boldsymbol{\theta})$$

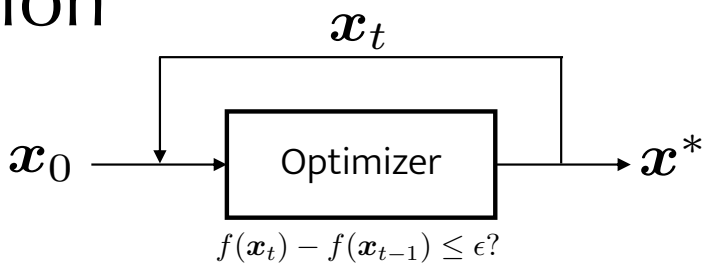
$$\mathbf{0} = \frac{\partial}{\partial \boldsymbol{x}^*} J_{f,\boldsymbol{x}^*}(\boldsymbol{x}^*; \boldsymbol{\theta}) \frac{\partial \boldsymbol{x}^*}{\partial \boldsymbol{\theta}} + \frac{\partial}{\partial \boldsymbol{\theta}} J_{f,\boldsymbol{x}^*}(\boldsymbol{x}^*; \boldsymbol{\theta})$$

$$\mathbf{0} = H_{f,\boldsymbol{x}^*}(\boldsymbol{x}^*; \boldsymbol{\theta}) \frac{\partial \boldsymbol{x}^*}{\partial \boldsymbol{\theta}} + \frac{\partial}{\partial \boldsymbol{\theta}} J_{f,\boldsymbol{x}^*}(\boldsymbol{x}^*; \boldsymbol{\theta})$$

Liao et al. Reviving and Improving Recurrent Back-Propagation. ICML 2018.

NYU

Implicit Differentiation



- Unconstrained case

$$\boldsymbol{x}^* = \underset{\boldsymbol{x}}{\operatorname{argmin}} f(\boldsymbol{x}; \boldsymbol{\theta}).$$

$$\mathbf{0} = \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}} J_{f,\boldsymbol{x}^*}(\boldsymbol{x}^*; \boldsymbol{\theta})$$

$$\mathbf{0} = \frac{\partial}{\partial \boldsymbol{x}^*} J_{f,\boldsymbol{x}^*}(\boldsymbol{x}^*; \boldsymbol{\theta}) \frac{\partial \boldsymbol{x}^*}{\partial \boldsymbol{\theta}} + \frac{\partial}{\partial \boldsymbol{\theta}} J_{f,\boldsymbol{x}^*}(\boldsymbol{x}^*; \boldsymbol{\theta})$$

$$\mathbf{0} = H_{f,\boldsymbol{x}^*}(\boldsymbol{x}^*; \boldsymbol{\theta}) \frac{\partial \boldsymbol{x}^*}{\partial \boldsymbol{\theta}} + \frac{\partial}{\partial \boldsymbol{\theta}} J_{f,\boldsymbol{x}^*}(\boldsymbol{x}^*; \boldsymbol{\theta})$$

$$\frac{\partial \boldsymbol{x}^*}{\partial \boldsymbol{\theta}} = H_{f,x^*}(\boldsymbol{x}^*; \boldsymbol{\theta})^{-1} \frac{\partial}{\partial \boldsymbol{\theta}} J_{f,x^*}(\boldsymbol{x}^*; \boldsymbol{\theta}).$$

Liao et al. Reviving and Improving Recurrent Back-Propagation. ICML 2018.

NYU

Implicit Differentiation

- How to compute Hessian inverse vector product?

Liao et al. Reviving and Improving Recurrent Back-Propagation. ICML 2018.

NYU

Implicit Differentiation

- How to compute Hessian inverse vector product?
- Conjugate gradient, solve $A\boldsymbol{x} = \boldsymbol{b}$

Conjugate Gradient Method

$\mathbf{r}_0 := \mathbf{b} - \mathbf{A}\mathbf{x}_0$
if $\mathbf{r}_0$ is sufficiently small, then return $\mathbf{x}_0$ as the result
$\mathbf{p}_0 := \mathbf{r}_0$
$k := 0$
repeat
$\alpha_k := \frac{\mathbf{r}_k^\mathsf{T}\mathbf{r}_k}{\mathbf{p}_k^\mathsf{T}\mathbf{A}\mathbf{p}_k}$
$\mathbf{x}_{k+1} := \mathbf{x}_k + \alpha_k\mathbf{p}_k$
$\mathbf{r}_{k+1} := \mathbf{r}_k - \alpha_k\mathbf{A}\mathbf{p}_k$
if $\mathbf{r}_{k+1}$ is sufficiently small, then exit loop
$\beta_k := \frac{\mathbf{r}_{k+1}^\mathsf{T}\mathbf{r}_{k+1}}{\mathbf{r}_k^\mathsf{T}\mathbf{r}_k}$
$\mathbf{p}_{k+1} := \mathbf{r}_{k+1} + \beta_k\mathbf{p}_k$
$k := k + 1$
end repeat
return $\mathbf{x}_{k+1}$ as the result

Liao et al. Reviving and Improving Recurrent Back-Propagation. ICML 2018.

NYU

Implicit Differentiation

- How to compute Hessian inverse vector product?
- Conjugate gradient, solve $A\boldsymbol{x} = \boldsymbol{b}$
- Neumann series (finite truncation)

$$(I - A)^{-1} = \sum_{k=0}^{\infty} A^k.$$

  - Same as backprop the last K steps (Option 2).
  - Memory savings.

Conjugate Gradient Method

$\mathbf{r}_0 := \mathbf{b} - \mathbf{A}\mathbf{x}_0$
if $\mathbf{r}_0$ is sufficiently small, then return $\mathbf{x}_0$ as the result
$\mathbf{p}_0 := \mathbf{r}_0$
$k := 0$
repeat
  $\alpha_k := \dfrac{\mathbf{r}_k^\mathsf{T}\mathbf{r}_k}{\mathbf{p}_k^\mathsf{T}\mathbf{A}\mathbf{p}_k}$
  $\mathbf{x}_{k+1} := \mathbf{x}_k + \alpha_k\mathbf{p}_k$
  $\mathbf{r}_{k+1} := \mathbf{r}_k - \alpha_k\mathbf{A}\mathbf{p}_k$
  if $\mathbf{r}_{k+1}$ is sufficiently small, then exit loop
  $\beta_k := \dfrac{\mathbf{r}_{k+1}^\mathsf{T}\mathbf{r}_{k+1}}{\mathbf{r}_k^\mathsf{T}\mathbf{r}_k}$
  $\mathbf{p}_{k+1} := \mathbf{r}_{k+1} + \beta_k\mathbf{p}_k$
  $k := k + 1$
end repeat
return $\mathbf{x}_{k+1}$ as the result

Liao et al. Reviving and Improving Recurrent Back-Propagation. ICML 2018.

NYU

Differentiable LQR

- Now add linear equality constraints on the dynamics and initialization.

$$\tau_{1:T} = \{x_t, u_t\}_{1:T}$$

$$\underset{\tau_{1:T}}{\text{argmin}} \sum_{t=1}^{T} \frac{1}{2}\tau_t^\top C_t \tau_t + c_t^\top \tau_t$$

$$\text{subject to } x_{t+1} = F_t \tau_t + f_t, x_1 = x_{\text{init}}.$$

Amos et al. Differentiable MPC for End-to-end Planning and Control. NeurIPS 2018.

NYU

Differentiable LQR

- Now add linear equality constraints on the dynamics and initialization.

$$\tau_{1:T} = \{x_t, u_t\}_{1:T}$$

- Chain rule: $\frac{\partial \ell}{\partial \theta} = \frac{\partial \ell}{\partial \tau^\star_{1:T}} \frac{\partial \tau^\star_{1:T}}{\partial \theta}$.

$$\underset{\tau_{1:T}}{\text{argmin}} \sum_{t=1}^{T} \frac{1}{2} \tau_t^\top C_t \tau_t + c_t^\top \tau_t$$

$$\text{subject to } x_{t+1} = F_t \tau_t + f_t, x_1 = x_{\text{init}}.$$

Amos et al. Differentiable MPC for End-to-end Planning and Control. NeurIPS 2018.

NYU

Differentiable LQR

- Now add linear equality constraints on the dynamics and initialization.

$$\tau_{1:T} = \{x_t, u_t\}_{1:T}$$

- Chain rule: $\frac{\partial \ell}{\partial \theta} = \frac{\partial \ell}{\partial \tau^\star_{1:T}} \frac{\partial \tau^\star_{1:T}}{\partial \theta}$.

$$\underset{\tau_{1:T}}{\text{argmin}} \sum_{t=1}^{T} \frac{1}{2} \tau_t^\top C_t \tau_t + c_t^\top \tau_t$$

$$\text{subject to } x_{t+1} = F_t \tau_t + f_t, x_1 = x_{\text{init}}.$$

General QP:

$$\boldsymbol{x}^* = \text{argmin} \frac{1}{2} \boldsymbol{x}^\top Q \boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{x}$$

$$\text{subject to } \; A\boldsymbol{x} = \boldsymbol{b}.$$

Amos et al. Differentiable MPC for End-to-end Planning and Control. NeurIPS 2018.

NYU

Differentiable LQR

- Now add linear equality constraints on the dynamics and initialization.

$$\tau_{1:T} = \{x_t, u_t\}_{1:T}$$

- Chain rule: $\frac{\partial \ell}{\partial \theta} = \frac{\partial \ell}{\partial \tau^\star_{1:T}} \frac{\partial \tau^\star_{1:T}}{\partial \theta}$.
- KKT:

$$\underset{\tau_{1:T}}{\text{argmin}} \sum_{t=1}^{T} \frac{1}{2} \tau_t^\top C_t \tau_t + c_t^\top \tau_t$$

$$\text{subject to } x_{t+1} = F_t \tau_t + f_t, x_1 = x_{\text{init}}.$$

General QP:

$$\boldsymbol{x}^* = \text{argmin} \frac{1}{2} \boldsymbol{x}^\top Q \boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{x}$$

$$\text{subject to } \; A\boldsymbol{x} = \boldsymbol{b}.$$

Amos et al. Differentiable MPC for End-to-end Planning and Control. NeurIPS 2018.

NYU

Differentiable LQR

- Now add linear equality constraints on the dynamics and initialization.

$$\tau_{1:T} = \{x_t, u_t\}_{1:T}$$

- Chain rule: $\frac{\partial \ell}{\partial \theta} = \frac{\partial \ell}{\partial \tau^\star_{1:T}} \frac{\partial \tau^\star_{1:T}}{\partial \theta}$.
- KKT:

$$\begin{bmatrix} Q & A^\top \\ A & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}^* \\ \boldsymbol{\lambda}^* \end{bmatrix} = \begin{bmatrix} -\boldsymbol{c} \\ \boldsymbol{b} \end{bmatrix}$$

$$K \begin{bmatrix} \boldsymbol{x}^* \\ \boldsymbol{\lambda}^* \end{bmatrix} = v.$$

$$\underset{\tau_{1:T}}{\text{argmin}} \sum_{t=1}^{T} \frac{1}{2} \tau_t^\top C_t \tau_t + c_t^\top \tau_t$$

$$\text{subject to } x_{t+1} = F_t \tau_t + f_t, x_1 = x_{\text{init}}.$$

General QP:

$$\boldsymbol{x}^* = \text{argmin} \frac{1}{2} \boldsymbol{x}^\top Q \boldsymbol{x} + \boldsymbol{c}^\top \boldsymbol{x}$$

$$\text{subject to } \; A\boldsymbol{x} = \boldsymbol{b}.$$

In classic LQR solver, the Riccati recursion solves this linear system.

Amos et al. Differentiable MPC for End-to-end Planning and Control. NeurIPS 2018.

NYU

Differentiable LQR

- Apply differentiation $\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}}\left(K\begin{bmatrix}\boldsymbol{x}^*\\ \boldsymbol{\lambda}^*\end{bmatrix}\right)=\frac{\mathrm{d}v}{\mathrm{d}\boldsymbol{\theta}}.$ $\frac{\mathrm{d}K}{\mathrm{d}\boldsymbol{\theta}}\begin{bmatrix}\boldsymbol{x}^*\\ \boldsymbol{\lambda}^*\end{bmatrix}+K\begin{bmatrix}\frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{\theta}}\\ \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{\theta}}\end{bmatrix}=\frac{\mathrm{d}v}{\mathrm{d}\boldsymbol{\theta}}$

Amos et al. Differentiable MPC for End-to-end Planning and Control. NeurIPS 2018.

NYU

Differentiable LQR

- Apply differentiation $\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}}\left(K\begin{bmatrix}\boldsymbol{x}^*\\ \boldsymbol{\lambda}^*\end{bmatrix}\right)=\frac{\mathrm{d}v}{\mathrm{d}\boldsymbol{\theta}}.\qquad \frac{\mathrm{d}K}{\mathrm{d}\boldsymbol{\theta}}\begin{bmatrix}\boldsymbol{x}^*\\ \boldsymbol{\lambda}^*\end{bmatrix}+K\begin{bmatrix}\frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{\theta}}\\ \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{\theta}}\end{bmatrix}=\frac{\mathrm{d}v}{\mathrm{d}\boldsymbol{\theta}}$

$$K\begin{bmatrix}\frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{\theta}}\\ \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{\theta}}\end{bmatrix}=K\begin{bmatrix}\frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{c}} & \frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{b}} & \frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}Q} & \frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}A}\\ \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{c}} & \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{b}} & \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}Q} & \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}A}\end{bmatrix}=\begin{bmatrix}-I & 0 & -x^* & -\lambda^*\\ 0 & I & 0 & -x^*\end{bmatrix}$$

Amos et al. Differentiable MPC for End-to-end Planning and Control. NeurIPS 2018.

NYU

Differentiable LQR

- Apply differentiation $\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}}\left(K\begin{bmatrix}\boldsymbol{x}^*\\ \boldsymbol{\lambda}^*\end{bmatrix}\right)=\frac{\mathrm{d}v}{\mathrm{d}\boldsymbol{\theta}}.\qquad \frac{\mathrm{d}K}{\mathrm{d}\boldsymbol{\theta}}\begin{bmatrix}\boldsymbol{x}^*\\ \boldsymbol{\lambda}^*\end{bmatrix}+K\begin{bmatrix}\frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{\theta}}\\ \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{\theta}}\end{bmatrix}=\frac{\mathrm{d}v}{\mathrm{d}\boldsymbol{\theta}}$

$$K\begin{bmatrix}\frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{\theta}}\\ \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{\theta}}\end{bmatrix}=K\begin{bmatrix}\frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{c}} & \frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{b}} & \frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}Q} & \frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}A}\\ \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{c}} & \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{b}} & \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}Q} & \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}A}\end{bmatrix}=\begin{bmatrix}-I & 0 & -x^* & -\lambda^*\\ 0 & I & 0 & -x^*\end{bmatrix}$$

$$K\frac{\partial\ell}{\partial\boldsymbol{z}^*}\begin{bmatrix}\frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{c}} & \frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{b}}\\ \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{c}} & \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{b}}\end{bmatrix}=\begin{bmatrix}-\frac{\partial\ell}{\partial\boldsymbol{x}^*}\\ \mathbf{0}\end{bmatrix}\qquad K\boldsymbol{d}^*=\begin{bmatrix}-\frac{\partial\ell}{\partial\boldsymbol{x}^*}\\ \mathbf{0}\end{bmatrix}$$

Amos et al. Differentiable MPC for End-to-end Planning and Control. NeurIPS 2018.

NYU

Differentiable LQR

- Apply differentiation $\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}}\left(K\begin{bmatrix}\boldsymbol{x}^*\\ \boldsymbol{\lambda}^*\end{bmatrix}\right)=\frac{\mathrm{d}v}{\mathrm{d}\boldsymbol{\theta}}.\qquad \frac{\mathrm{d}K}{\mathrm{d}\boldsymbol{\theta}}\begin{bmatrix}\boldsymbol{x}^*\\ \boldsymbol{\lambda}^*\end{bmatrix}+K\begin{bmatrix}\frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{\theta}}\\ \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{\theta}}\end{bmatrix}=\frac{\mathrm{d}v}{\mathrm{d}\boldsymbol{\theta}}$

$$K\begin{bmatrix}\frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{\theta}}\\ \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{\theta}}\end{bmatrix}=K\begin{bmatrix}\frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{c}} & \frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{b}} & \frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}Q} & \frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}A}\\ \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{c}} & \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{b}} & \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}Q} & \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}A}\end{bmatrix}=\begin{bmatrix}-I & 0 & -x^* & -\lambda^*\\ 0 & I & 0 & -x^*\end{bmatrix}$$

$$K\frac{\partial\ell}{\partial\boldsymbol{z}^*}\begin{bmatrix}\frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{c}} & \frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{b}}\\ \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{c}} & \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{b}}\end{bmatrix}=\begin{bmatrix}-\frac{\partial\ell}{\partial\boldsymbol{x}^*}\\ \mathbf{0}\end{bmatrix}\qquad K\boldsymbol{d}^*=\begin{bmatrix}-\frac{\partial\ell}{\partial\boldsymbol{x}^*}\\ \mathbf{0}\end{bmatrix}$$

- Equivalent QP:

$$\boldsymbol{d}^*=\underset{\boldsymbol{d}}{\operatorname{argmin}}\ \frac{1}{2}\boldsymbol{d}^\top Q\boldsymbol{d}+\frac{\partial\ell}{\partial\boldsymbol{x}^*}^\top\boldsymbol{d},$$
$$\text{subject to}\quad A\boldsymbol{d}=\mathbf{0}.$$

Amos et al. Differentiable MPC for End-to-end Planning and Control. NeurIPS 2018.

NYU

Differentiable LQR

- Apply differentiation $\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}}\left(K\begin{bmatrix}\boldsymbol{x}^*\\ \boldsymbol{\lambda}^*\end{bmatrix}\right) = \frac{\mathrm{d}v}{\mathrm{d}\boldsymbol{\theta}}. \qquad \frac{\mathrm{d}K}{\mathrm{d}\boldsymbol{\theta}}\begin{bmatrix}\boldsymbol{x}^*\\ \boldsymbol{\lambda}^*\end{bmatrix} + K\begin{bmatrix}\frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{\theta}}\\ \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{\theta}}\end{bmatrix} = \frac{\mathrm{d}v}{\mathrm{d}\boldsymbol{\theta}}$

$$K\begin{bmatrix}\frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{\theta}}\\ \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{\theta}}\end{bmatrix} = K\begin{bmatrix}\frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{c}} & \frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{b}} & \frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}Q} & \frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}A}\\ \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{c}} & \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{b}} & \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}Q} & \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}A}\end{bmatrix} = \begin{bmatrix}-I & 0 & -x^* & -\lambda^*\\ 0 & I & 0 & -x^*\end{bmatrix}$$

$$K\frac{\partial \ell}{\partial \boldsymbol{z}^*}\begin{bmatrix}\frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{c}} & \frac{\mathrm{d}\boldsymbol{x}^*}{\mathrm{d}\boldsymbol{b}}\\ \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{c}} & \frac{\mathrm{d}\boldsymbol{\lambda}^*}{\mathrm{d}\boldsymbol{b}}\end{bmatrix} = \begin{bmatrix}-\frac{\partial \ell}{\partial \boldsymbol{x}^*}\\ \mathbf{0}\end{bmatrix} \qquad K\boldsymbol{d}^* = \begin{bmatrix}-\frac{\partial \ell}{\partial \boldsymbol{x}^*}\\ \mathbf{0}\end{bmatrix}$$

- Equivalent QP:

$$\boldsymbol{d}^* = \underset{\boldsymbol{d}}{\operatorname{argmin}} \frac{1}{2}\boldsymbol{d}^\top Q\boldsymbol{d} + \frac{\partial \ell}{\partial \boldsymbol{x}^*}^\top \boldsymbol{d}, \quad \text{subject to } A\boldsymbol{d} = \mathbf{0}.$$

$$\frac{\partial \ell}{\partial Q} = \frac{1}{2}(\boldsymbol{d}_{\boldsymbol{x}}^* \otimes \boldsymbol{x}^* + \boldsymbol{x}^* \otimes \boldsymbol{d}_{\boldsymbol{x}}^*)$$

$$\frac{\partial \ell}{\partial A} = \boldsymbol{d}_{\boldsymbol{\lambda}}^* \otimes \boldsymbol{x}^* + \boldsymbol{\lambda}^* \otimes \boldsymbol{d}_{\boldsymbol{x}}^*.$$

Amos et al. Differentiable MPC for End-to-end Planning and Control. NeurIPS 2018.

NYU

Differentiable LQR

- The backward pass can also be formulated as a LQR problem.
- Swap c to $\nabla_{\tau^\star}\ell$ and f to 0.

Module 1 Differentiable LQR *(The LQR algorithm is defined in appendix A)*

Input: Initial state x_{init}
Parameters: $\theta = \{C, c, F, f\}$

Forward Pass:

1: $\tau^\star_{1:T} = \text{LQR}_T(x_{\text{init}}; C, c, F, f)$ ▷ Solve (2)
2: Compute $\lambda^\star_{1:T}$ with (7)

Backward Pass:

1: $d^\star_{\tau_{1:T}} = \text{LQR}_T(0; C, \nabla_{\tau^\star}\ell, F, 0)$ ▷ Solve (9), ideally reusing the factorizations from the forward pass
2: Compute $d^\star_{\lambda_{1:T}}$ with (7)
3: Compute the derivatives of ℓ with respect to C, c, F, f, and x_{init} with (8)

Amos et al. Differentiable MPC for End-to-end Planning and Control. NeurIPS 2018.

NYU

Differentiable MPC

- What about general MPC?

$$\underset{x_{1:T} \in \mathcal{X}, u_{1:T} \in \mathcal{U}}{\operatorname{argmin}} \sum_{t=1}^{T} C_t(x_t, u_t)$$

$$\text{subject to } x_{t+1} = f(x_t, u_t), x_1 = x_{\text{init}}.$$

Amos et al. Differentiable MPC for End-to-end Planning and Control. NeurIPS 2018.

NYU

Differentiable MPC

- What about general MPC?

$$\underset{x_{1:T} \in \mathcal{X}, u_{1:T} \in \mathcal{U}}{\operatorname{argmin}} \sum_{t=1}^{T} C_t(x_t, u_t)$$

$$\text{subject to } x_{t+1} = f(x_t, u_t), x_1 = x_{\text{init}}.$$

- Use Taylor expansion to approximate.

$$\tilde{C}^i_{\theta,t} = C_{\theta,t}(\tau^i_t) + {p^i_t}^\top (\tau_t - \tau^i_t) + \tfrac{1}{2}(\tau_t - \tau^i_t)^\top H^i_t (\tau_t - \tau^i_t).$$

Amos et al. Differentiable MPC for End-to-end Planning and Control. NeurIPS 2018.

NYU

Differentiable MPC

- What about general MPC?

$$\underset{x_{1:T}\in\mathcal{X},u_{1:T}\in\mathcal{U}}{\operatorname{argmin}} \sum_{t=1}^{T} C_t(x_t,u_t)$$

$$\text{subject to } x_{t+1}=f(x_t,u_t), x_1=x_{\text{init}}.$$

- Use Taylor expansion to approximate.

$$\tilde{C}^i_{\theta,t}=C_{\theta,t}(\tau^i_t)+{p^i_t}^\top(\tau_t-\tau^i_t)+\tfrac{1}{2}(\tau_t-\tau^i_t)^\top H^i_t(\tau_t-\tau^i_t).$$

- Fixed point iteration

$$\tau^{i+1}=\operatorname{argmin}_\tau \textstyle\sum_t^T \tilde{C}^i_t(\tau^i_t).$$

Amos et al. Differentiable MPC for End-to-end Planning and Control. NeurIPS 2018.

NYU

Differentiable MPC

- What about general MPC?

$$\underset{x_{1:T} \in \mathcal{X}, u_{1:T} \in \mathcal{U}}{\operatorname{argmin}} \sum_{t=1}^{T} C_t(x_t, u_t)$$

$$\text{subject to } x_{t+1} = f(x_t, u_t), x_1 = x_{\text{init}}.$$

- Use Taylor expansion to approximate.

$$\tilde{C}^i_{\theta,t} = C_{\theta,t}(\tau^i_t) + {p^i_t}^\top (\tau_t - \tau^i_t) + \tfrac{1}{2}(\tau_t - \tau^i_t)^\top H^i_t (\tau_t - \tau^i_t).$$

- Fixed point iteration

$$\tau^{i+1} = \operatorname{argmin}_\tau \textstyle\sum_t^T \tilde{C}^i_t(\tau^i_t).$$

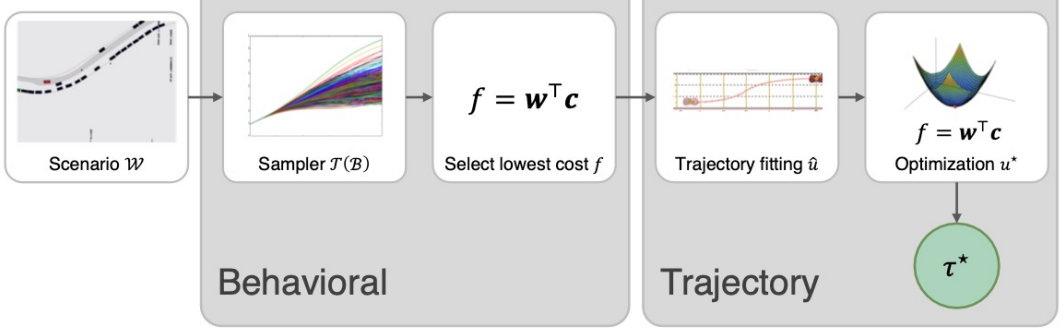
- Backward only depends on final quadratic approximation.

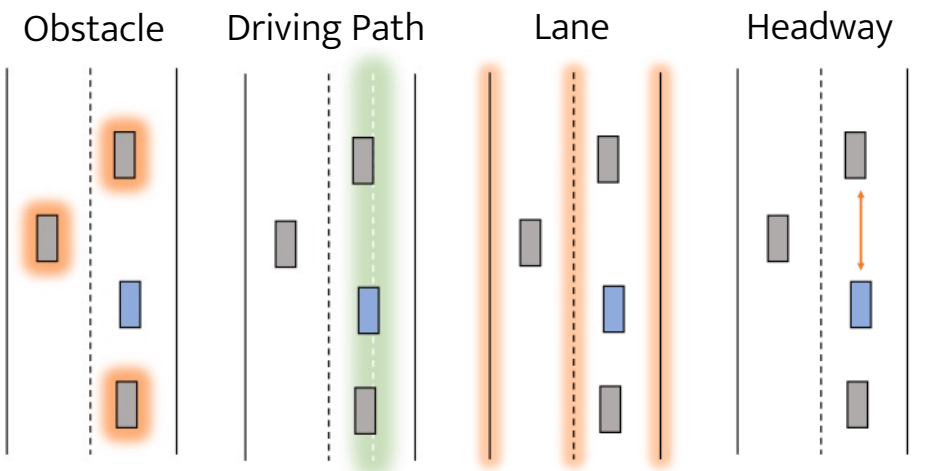
Amos et al. Differentiable MPC for End-to-end Planning and Control. NeurIPS 2018.

NYU

Behavioral vs. Trajectory Planning

- Gradient-based optimization provides a locally optimized trajectory.



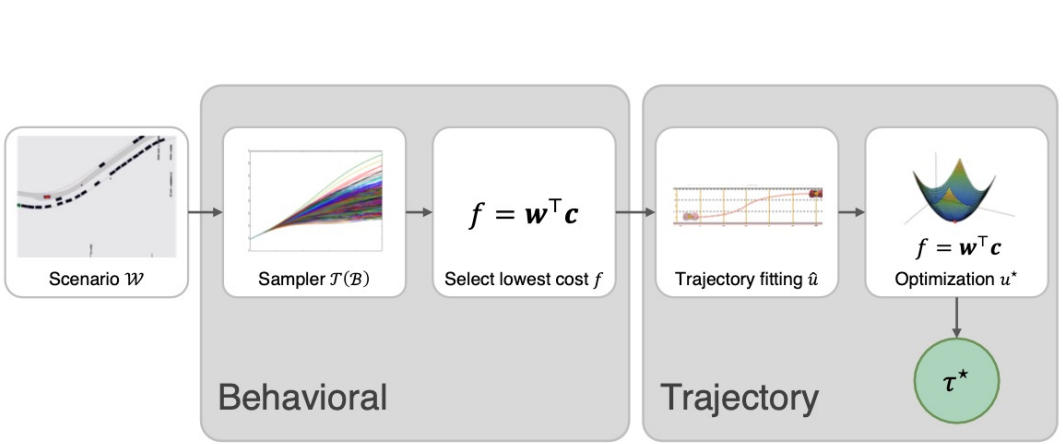


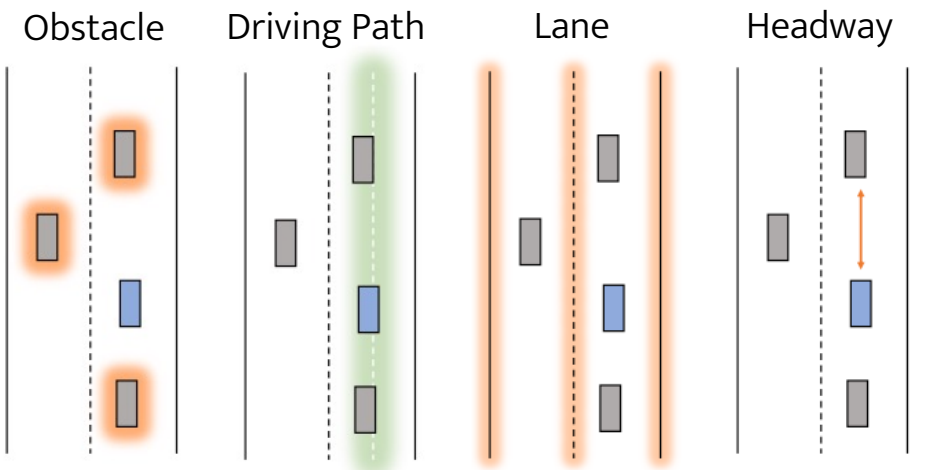
Sadat et al. Jointly Learnable Behavior and Trajectory Planning for Self-Driving Vehicles. IROS 2019.

NYU

Behavioral vs. Trajectory Planning

- Gradient-based optimization provides a locally optimized trajectory.
- Samples may be needed for reasoning global structure.



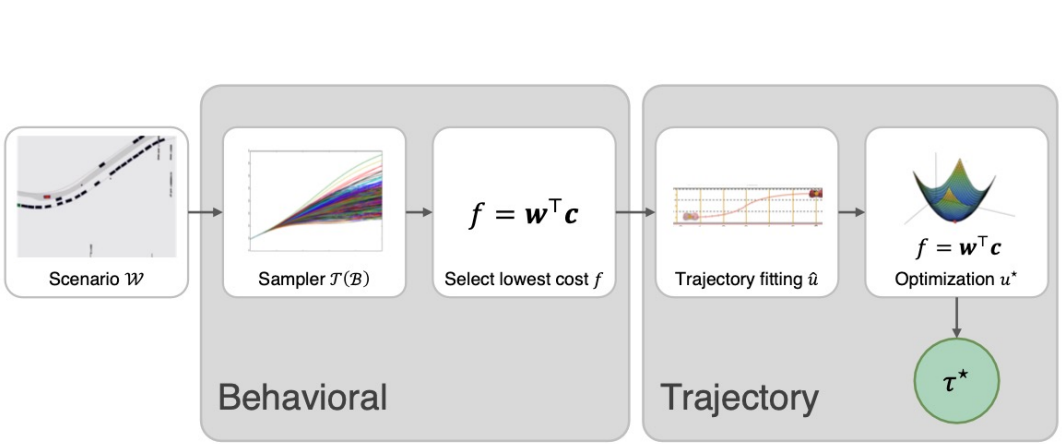


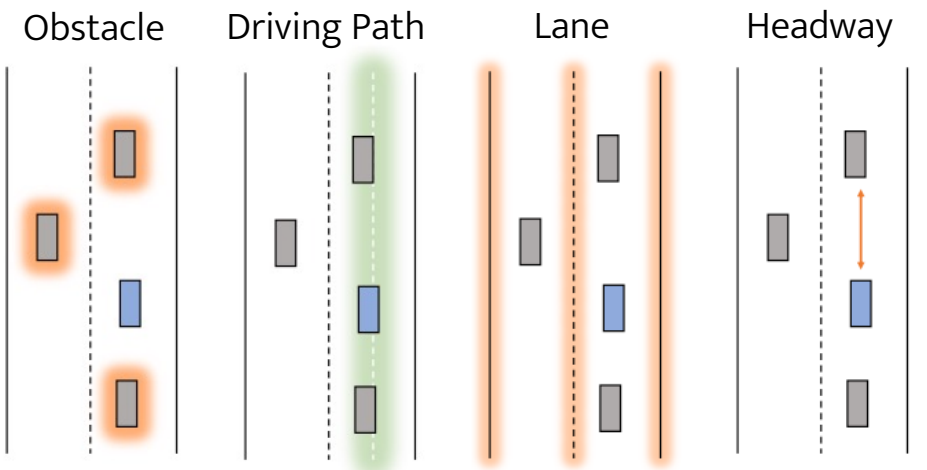
Sadat et al. Jointly Learnable Behavior and Trajectory Planning for Self-Driving Vehicles. IROS 2019.

NYU

Behavioral vs. Trajectory Planning

- Gradient-based optimization provides a locally optimized trajectory.
- Samples may be needed for reasoning global structure.
- Can learn together using the same learned costs.





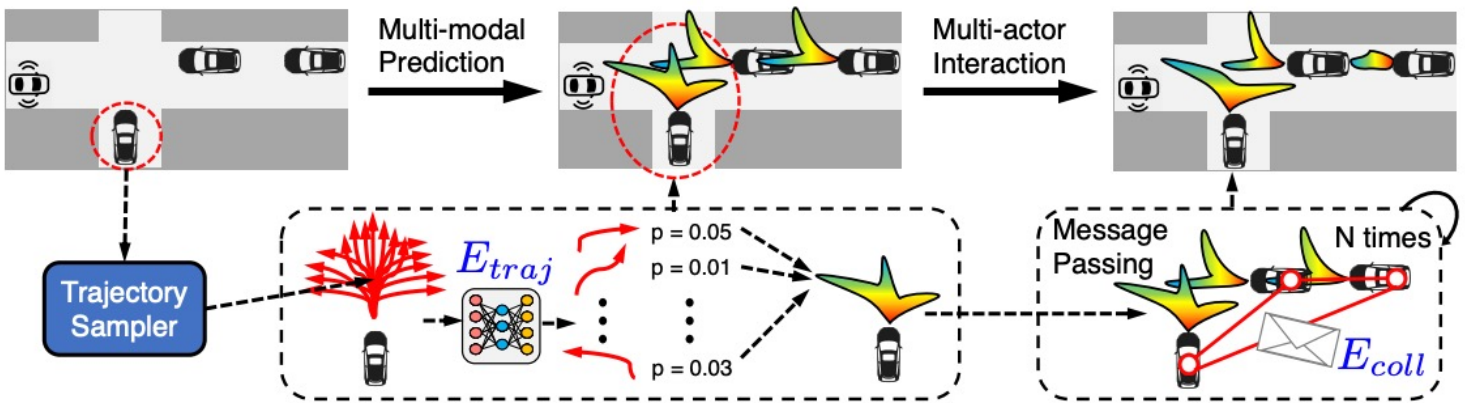
Sadat et al. Jointly Learnable Behavior and Trajectory Planning for Self-Driving Vehicles. IROS 2019.

NYU

Planning with Social Reasoning

- Jointly reason the future trajectories of multiple agents as an energy-based graphical model.

$$p(\mathbf{s}_1, \ldots, \mathbf{s}_N \mid \mathbf{X}) = \frac{1}{Z} \exp(-E_\theta(\mathbf{s}_1, \ldots, \mathbf{s}_N) \mid \mathbf{X})$$



Zeng et al. DSDNet: Deep Structured Self-Driving Network. ECCV 2020.

NYU

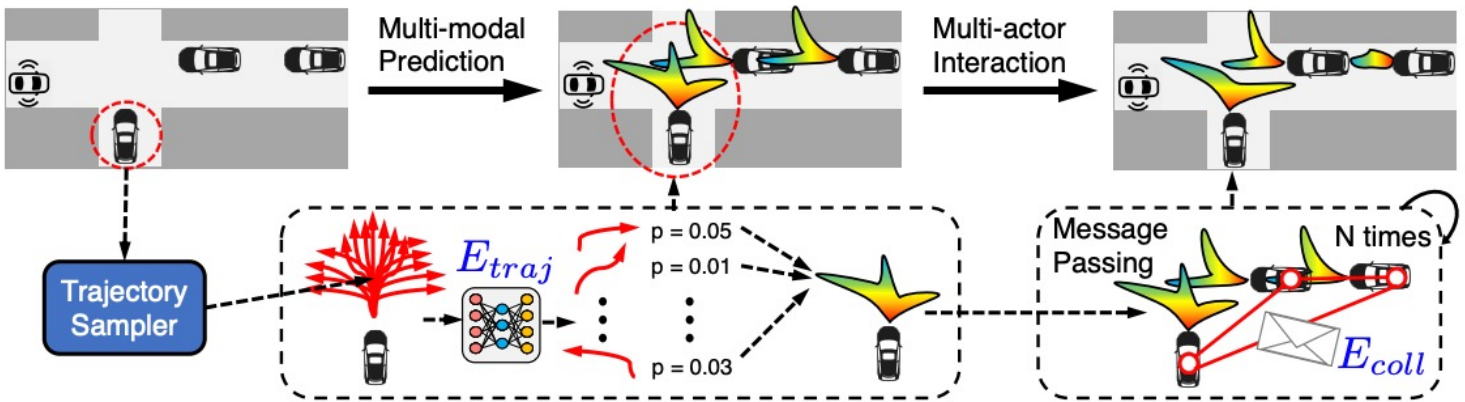
Planning with Social Reasoning

- Jointly reason the future trajectories of multiple agents as an energy-based graphical model.
- Trajectory Goodness + Collision.

$$p(\mathbf{s}_1, \ldots, \mathbf{s}_N \mid \mathbf{X}) = \frac{1}{Z} \exp(-E_\theta(\mathbf{s}_1, \ldots, \mathbf{s}_N) \mid \mathbf{X})$$

$$\sum_i E_\theta(s_i \mid X) + \sum_{i \neq j} E(\mathbf{s}_i, \mathbf{s}_j)$$

$$E(\mathbf{s}_i, \mathbf{s}_j) = \gamma \quad \text{if } \mathbf{s}_i \text{ collides } \mathbf{s}_j$$



Zeng et al. DSDNet: Deep Structured Self-Driving Network. ECCV 2020.

NYU

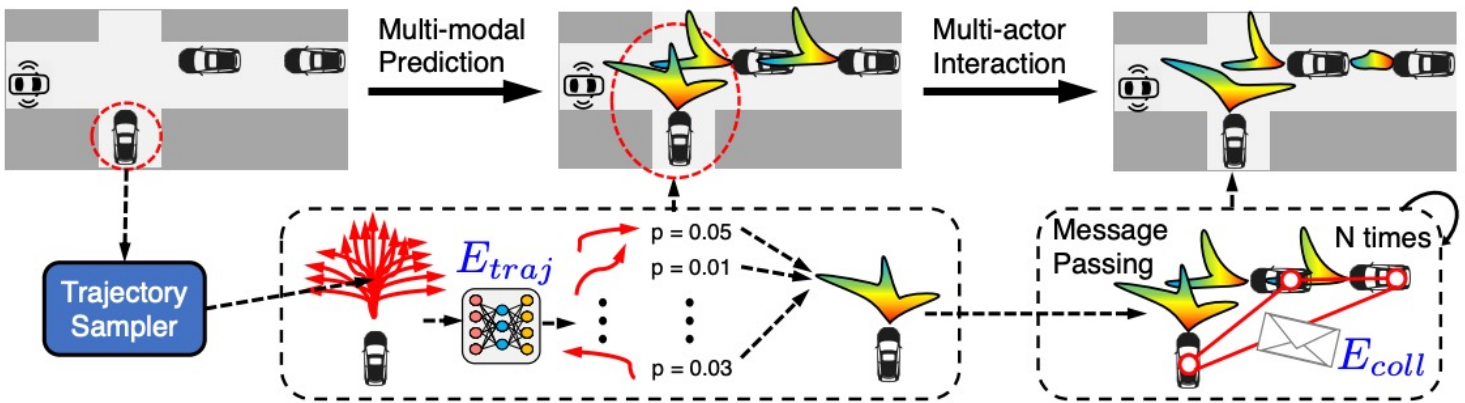
Planning with Social Reasoning

- Jointly reason the future trajectories of multiple agents as an energy-based graphical model.
- Trajectory Goodness + Collision.
- Batch of trajectory samples.

$$p(\mathbf{s}_1, \ldots, \mathbf{s}_N \mid \mathbf{X}) = \frac{1}{Z} \exp(-E_\theta(\mathbf{s}_1, \ldots, \mathbf{s}_N) \mid \mathbf{X})$$

$$\sum_i E_\theta(s_i \mid X) + \sum_{i \neq j} E(\mathbf{s}_i, \mathbf{s}_j)$$

$$E(\mathbf{s}_i, \mathbf{s}_j) = \gamma \quad \text{if } \mathbf{s}_i \text{ collides } \mathbf{s}_j$$



Zeng et al. DSDNet: Deep Structured Self-Driving Network. ECCV 2020.

NYU

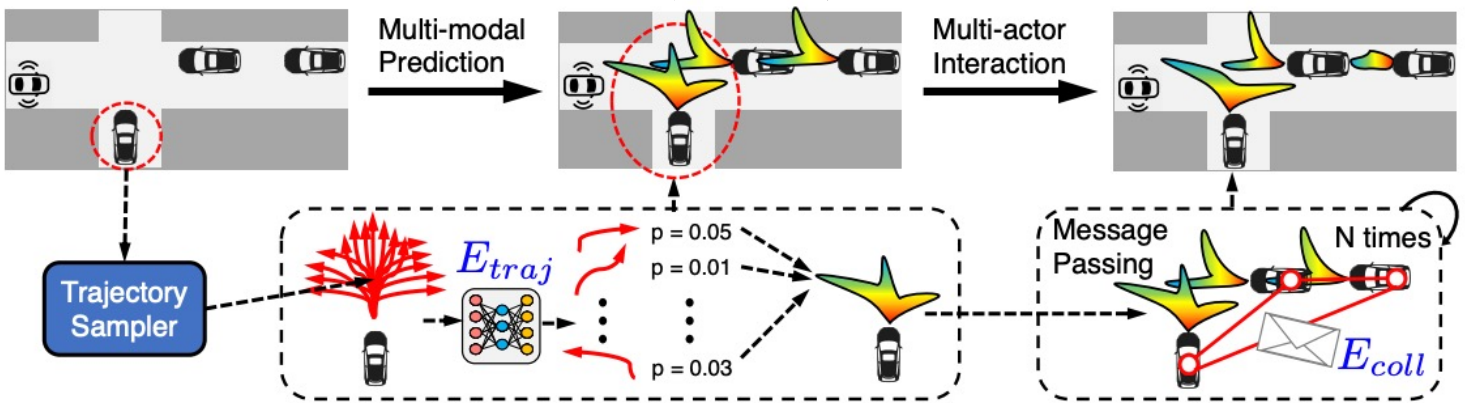
Planning with Social Reasoning

- Jointly reason the future trajectories of multiple agents as an energy-based graphical model.
- Trajectory Goodness + Collision.
- Batch of trajectory samples.
- Classification of groundtruth trajectory.

$$p(\mathbf{s}_1, \ldots, \mathbf{s}_N \mid \mathbf{X}) = \frac{1}{Z} \exp(-E_\theta(\mathbf{s}_1, \ldots, \mathbf{s}_N) \mid \mathbf{X})$$

$$\sum_i E_\theta(s_i \mid X) + \sum_{i \neq j} E(\mathbf{s}_i, \mathbf{s}_j)$$

$$E(\mathbf{s}_i, \mathbf{s}_j) = \gamma \quad \text{if } \mathbf{s}_i \text{ collides } \mathbf{s}_j$$



Zeng et al. DSDNet: Deep Structured Self-Driving Network. ECCV 2020.

NYU

Summary: End-to-End Planning and Control

- Direct Policy Prediction
  - Condition perception features into the model
  - Use of diffusion models

NYU

Summary: End-to-End Planning and Control

- Direct Policy Prediction
  - Condition perception features into the model
  - Use of diffusion models
- Cost Learning (IRL) from Experts
  - Max-margin, max-entropy/EBM
  - Need negative samples
  - Can be combined with efficient external samplers
  - Cost volume prediction: parametric + non-parametric

NYU

Summary: End-to-End Planning and Control

- Direct Policy Prediction
  - Condition perception features into the model
  - Use of diffusion models
- Cost Learning (IRL) from Experts
  - Max-margin, max-entropy/EBM
  - Need negative samples
  - Can be combined with efficient external samplers
  - Cost volume prediction: parametric + non-parametric
- Differentiable Planner
  - Backprop through local optimization, learnable cost
  - Can be memory efficient, implicit differentiation

NYU