Learned Trajectory Embedding for Subspace Clustering

Yaroslava Lochman[1] Carl Olsson[1,2] Christopher Zach[1]

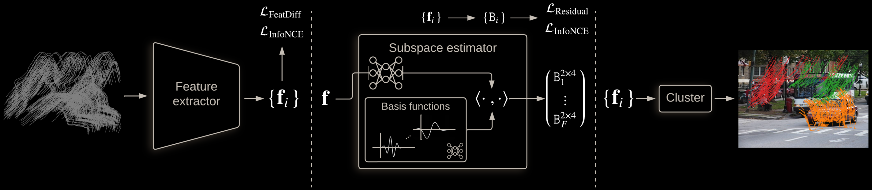
[1]Chalmers University of Technology [2]Lund University

March 12, SSBA 2024

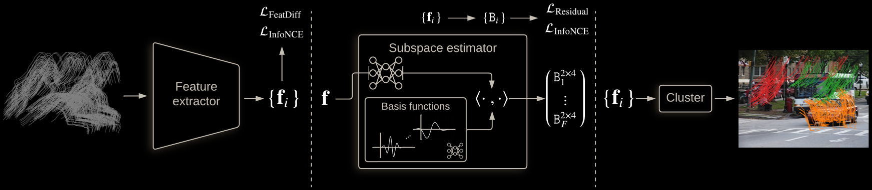
Outline

- Introduction: problem formulation, background
- Method: architecture, training, trajectory completion algorithm
- Results: invariance study, completion evaluation, benchmark
- Discussion: future work, Q&A

Problem Formulation

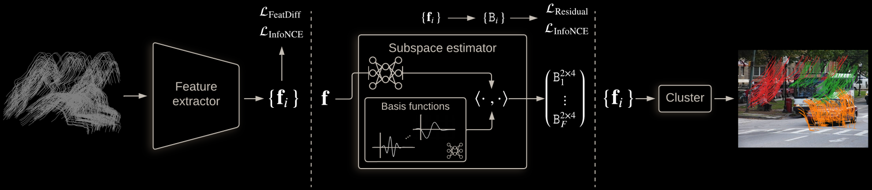


Problem Formulation



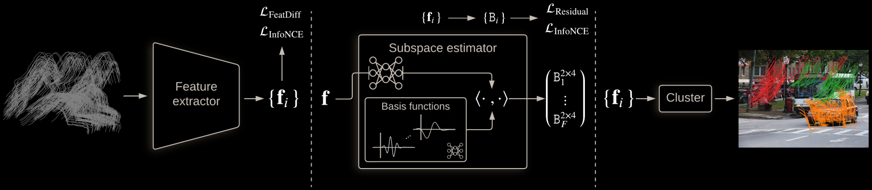
- Input: 2D point trajectories extracted from a video ($\mathtt{M}_{2F\times P}$)

Problem Formulation



- Input: 2D point trajectories extracted from a video ($\mathtt{M}_{2F\times P}$)
- Want to find: grouping with associated 3D rigid motions ($\mathtt{B}_1, \ldots, \mathtt{B}_c$)

Problem Formulation

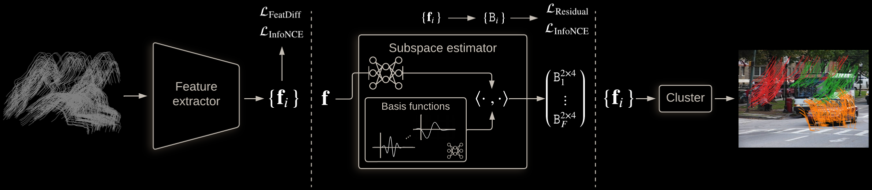


- Input: 2D point trajectories extracted from a video ($\mathtt{M}_{2F\times P}$)
- Want to find: grouping with associated 3D rigid motions ($\mathtt{B}_1, \ldots, \mathtt{B}_c$)
- Assuming affine projection

$$\mathtt{M}_{2F\times P}\mathtt{P}_\pi \approx \begin{bmatrix} \mathtt{B}_1\mathtt{C}_1^\top & \ldots & \mathtt{B}_c\mathtt{C}_c^\top \end{bmatrix}$$

where $\mathtt{P}_\pi$ — $P \times P$ permutation matrix

Problem Formulation



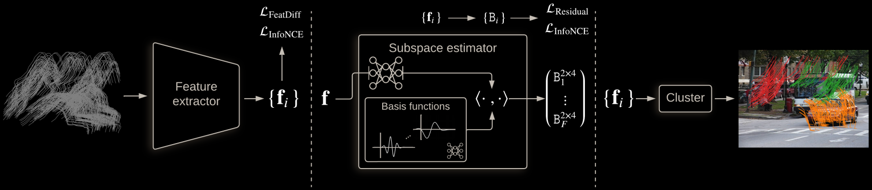
- Input: 2D point trajectories extracted from a video ($\mathtt{M}_{2F\times P}$)
- Want to find: grouping with associated 3D rigid motions ($\mathtt{B}_1, \ldots, \mathtt{B}_c$)
- Assuming affine projection

$$\mathtt{M}_{2F\times P}\mathtt{P}_\pi \approx \begin{bmatrix} \mathtt{B}_1\mathtt{C}_1^\top & \ldots & \mathtt{B}_c\mathtt{C}_c^\top \end{bmatrix}$$

where $\mathtt{P}_\pi$ — $P \times P$ permutation matrix

- Chicken-and-egg problem

Problem Formulation



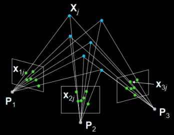
- Input: 2D point trajectories extracted from a video ($\mathtt{M}_{2F\times P}$)
- Want to find: grouping with associated 3D rigid motions ($\mathtt{B}_1, \ldots, \mathtt{B}_c$)
- Assuming affine projection

$$\mathtt{M}_{2F\times P}\mathtt{P}_\pi \approx \left[\mathtt{B}_1\mathtt{C}_1^\top \quad \ldots \quad \mathtt{B}_c\mathtt{C}_c^\top\right]$$

  where $\mathtt{P}_\pi$ — $P \times P$ permutation matrix
- Chicken-and-egg problem
- Expect high rates of occlusion in real scenarios

Background

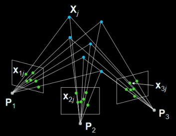
(Nonrigid) structure-from-motion



Background

(Nonrigid) structure-from-motion

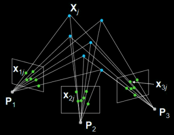
- For affine cameras, equivalent to subspace fitting



Background

(Nonrigid) structure-from-motion

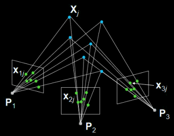
- For affine cameras, equivalent to subspace fitting
- SfM — too restricting, one rigid object



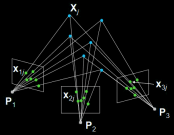
Background

(Nonrigid) structure-from-motion

- For affine cameras, equivalent to subspace fitting
- SfM — too restricting, one rigid object
- NRSfM — too general, deforming objects + gives an unconstrained solution

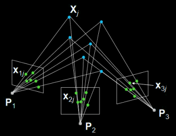


Background



Subspace clustering

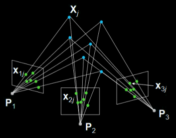
Background



Subspace clustering

- Works with data points in some Hilbert space

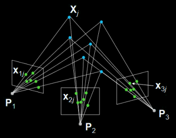
Background



Subspace clustering

- Works with data points in some Hilbert space
- Assumes the underlying model is the union of subspaces

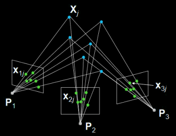
Background



Subspace clustering

- Works with data points in some Hilbert space
- Assumes the underlying model is the union of subspaces
- Aims to find: number, dimensionality and basis of each subspace + grouping

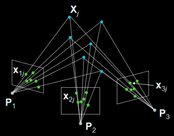
Background



Subspace clustering

- Works with data points in some Hilbert space
- Assumes the underlying model is the union of subspaces
- Aims to find: number, dimensionality and basis of each subspace + grouping
- Apply to our problem directly?

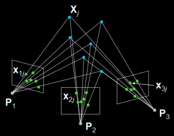
Background



Subspace clustering

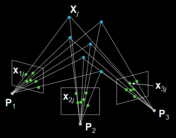
- Works with data points in some Hilbert space
- Assumes the underlying model is the union of subspaces
- Aims to find: number, dimensionality and basis of each subspace + grouping
- Apply to our problem directly? High-dimensional case $\Rightarrow$ slow/inefficient; does not exploit temporal information.

Background



RANSAC variations for multi-model fitting

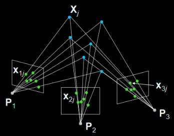
Background



RANSAC variations for multi-model fitting

- Robust statistical methods, good for low-dimensional data

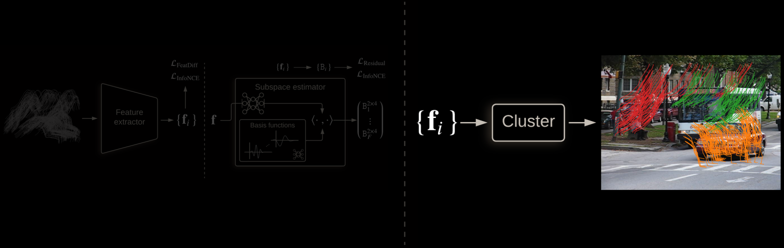
Background



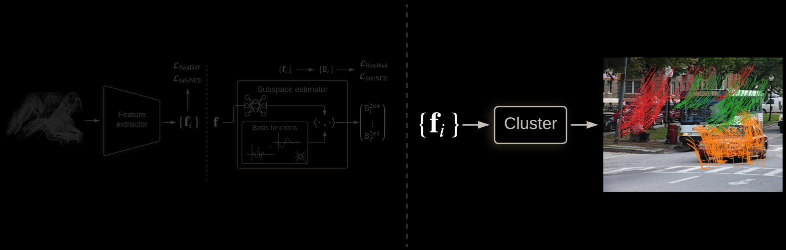
RANSAC variations for multi-model fitting

- Robust statistical methods, good for low-dimensional data
- Greedy $\Rightarrow$ inefficient; Joint (with energy minimization) $\Rightarrow$ slow

Learned Trajectory Embedding

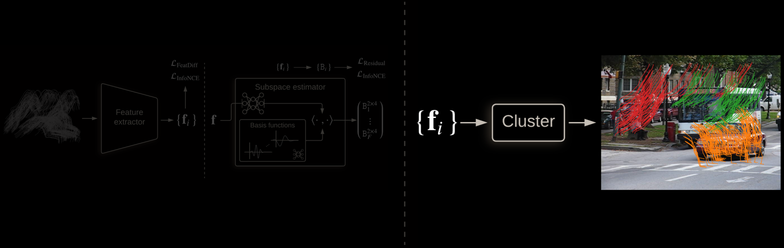


Learned Trajectory Embedding



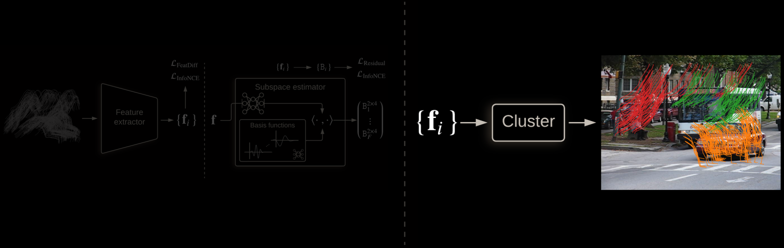
- Learn mapping from single trajectory $\mathbf{x}_i$ to feature representation $\mathbf{f}_i$

Learned Trajectory Embedding



- Learn mapping from single trajectory $\mathbf{x}_i$ to feature representation $\mathbf{f}_i$
- $\mathbf{f}_i$ fully identifies generating motion $\Rightarrow$ can be used for clustering

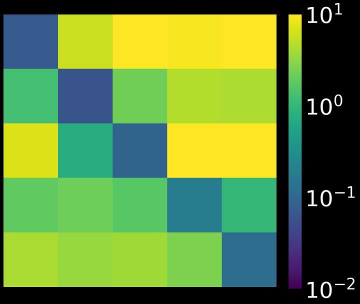
Learned Trajectory Embedding



- Learn mapping from single trajectory $\mathbf{x}_i$ to feature representation $\mathbf{f}_i$
- $\mathbf{f}_i$ fully identifies generating motion $\Rightarrow$ can be used for clustering
- Accurate and fast: no simultaneous grouping and motion estimation at test-time

Disjoint Subspace Assumption

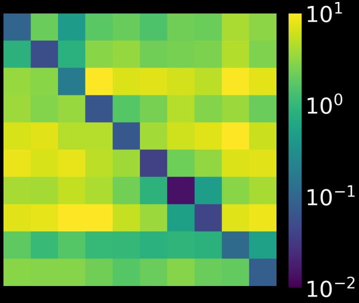
Re-using subspaces to explain trajectories in other clusters ⇒ higher errors.



Cluster-to-subspace errors for subsequences of length $F = 60$

Disjoint Subspace Assumption

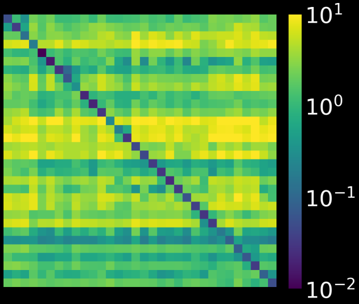
Re-using subspaces to explain trajectories in other clusters $\Rightarrow$ higher errors.



Cluster-to-subspace errors for subsequences of length $F = 40$

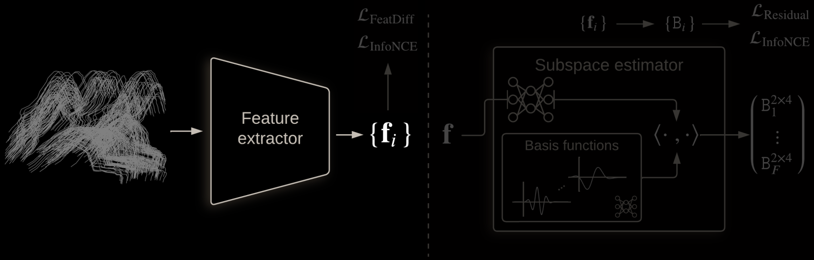
Disjoint Subspace Assumption

Re-using subspaces to explain trajectories in other clusters $\Rightarrow$ higher errors.

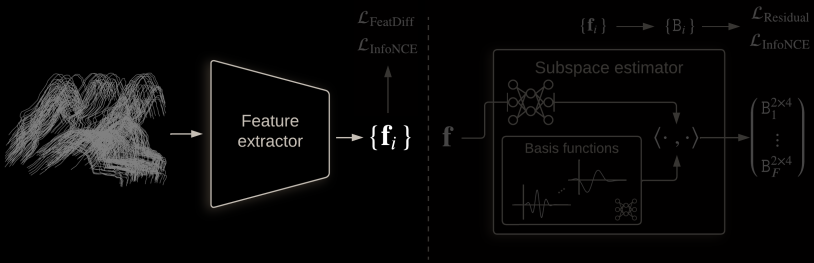


Cluster-to-subspace errors for subsequences of length $F = 30$

Feature Extraction

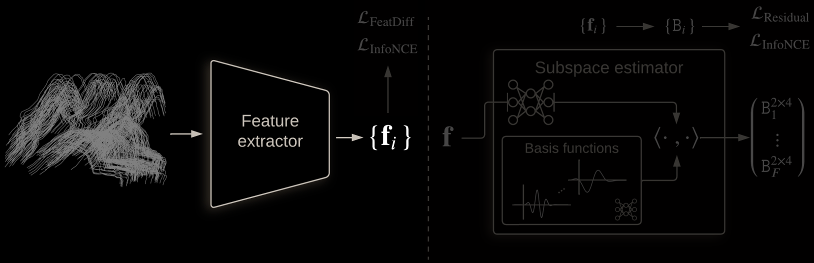


Feature Extraction



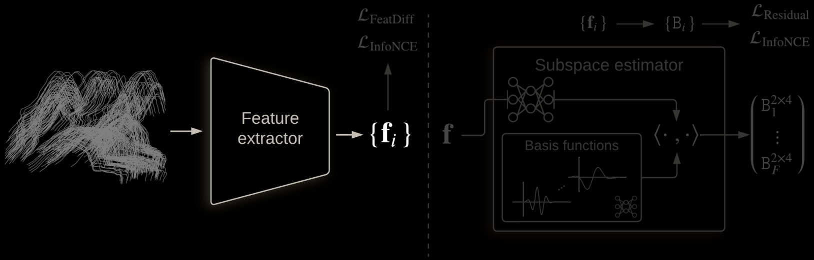
- PointNet style

Feature Extraction



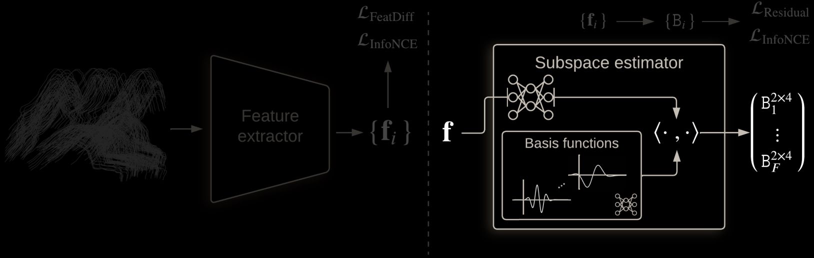
- PointNet style
- 1D convolutional in temporal domain

Feature Extraction

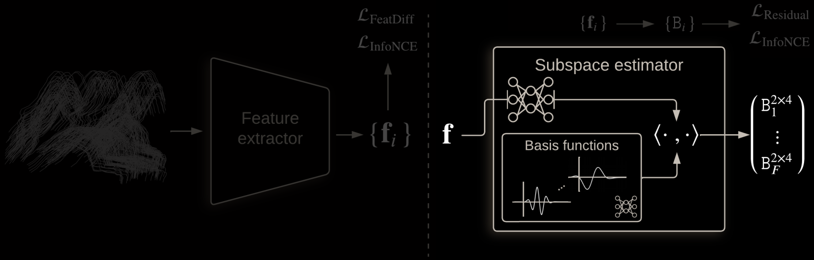


- PointNet style
- 1D convolutional in temporal domain
- No global context (e.g., spatial pooling)

Subspace Estimation

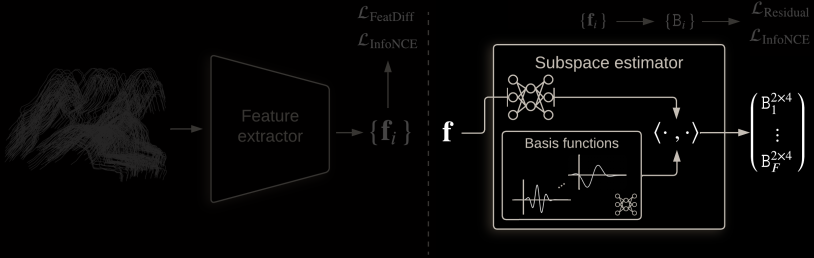


Subspace Estimation



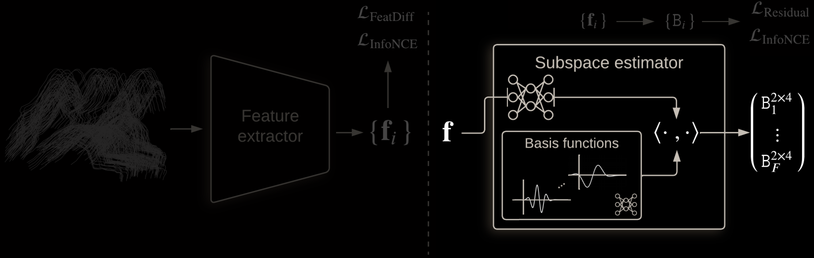
- Subspaces encode change of motion over time $\Rightarrow$ time-dependent basis

Subspace Estimation



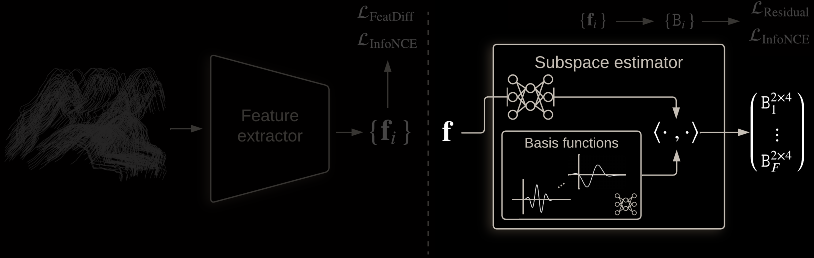
- Subspaces encode change of motion over time $\Rightarrow$ time-dependent basis
- Basis functions evaluated at time query t

Subspace Estimation



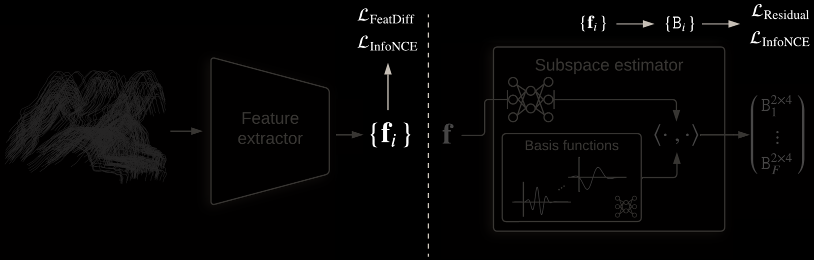
- Subspaces encode change of motion over time $\Rightarrow$ time-dependent basis
- Basis functions evaluated at time query t
- Basis coefficients inferred from features with an MLP

Subspace Estimation

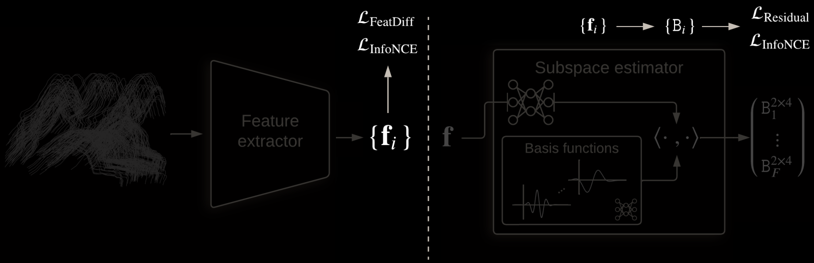


- Subspaces encode change of motion over time $\Rightarrow$ time-dependent basis
- Basis functions evaluated at time query t
- Basis coefficients inferred from features with an MLP
- Coordinate-MLP style (similar to conditional NeRFs)

Training

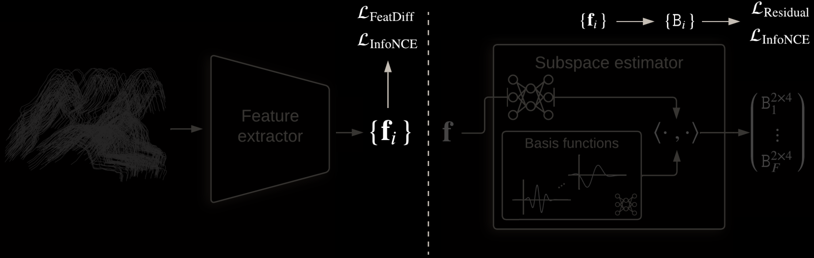

$\mathcal{L}_{\text{FeatDiff}}$
$\mathcal{L}_{\text{InfoNCE}}$
Feature extractor
$\{\mathbf{f}_i\}$
$\mathbf{f}$
$\{\mathbf{f}_i\} \longrightarrow \{\mathrm{B}_i\} \longrightarrow$
$\mathcal{L}_{\text{Residual}}$
$\mathcal{L}_{\text{InfoNCE}}$
Subspace estimator
Basis functions
$\langle \cdot, \cdot \rangle$
$\mathrm{B}_1^{2\times4}$
$\mathrm{B}_F^{2\times4}$

Training



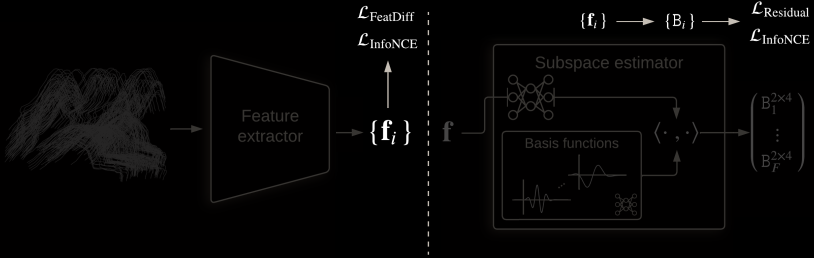
- Pre-train features via enforcing small within-cluster-distances and large between-cluster-distances

Training



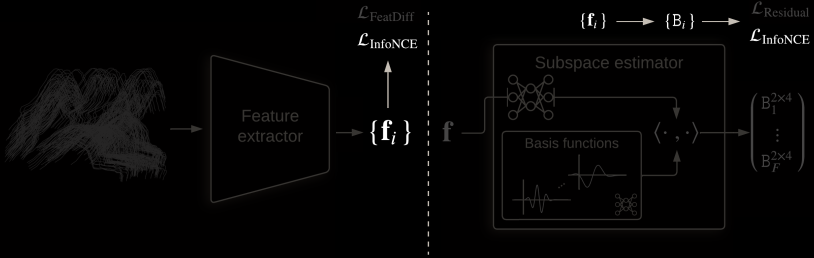
- Pre-train features via enforcing small within-cluster-distances and large between-cluster-distances
- Train subspace estimator via enforcing small residuals

Training



- Pre-train features via enforcing small within-cluster-distances and large between-cluster-distances
- Train subspace estimator via enforcing small residuals
- $+$ enforce feature closeness of original and reconstructed trajectories

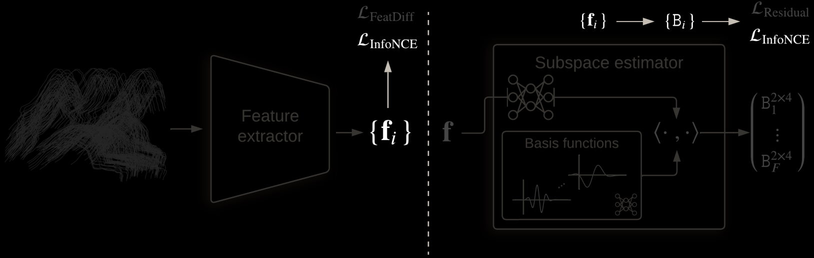
Losses



For f_θ — feature extractor, g_ϕ — subspace estimator:

$$\mathcal{L}_{\text{InfoNCE}} = \frac{1}{|\mathcal{Q}|} \sum_{(i,j,l,k)\in\mathcal{Q}} \log\left(\frac{p_{ij}}{p_{ij}+p_{lk}}\right) \qquad p_{ij} = \exp\left(-\frac{\|\mathbf{f}_i - \mathbf{f}_j\|_2^2}{T}\right)$$

Losses

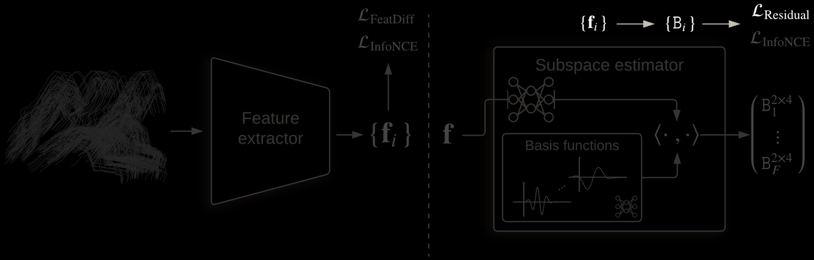


For f_θ — feature extractor, g_ϕ — subspace estimator:

$$\mathcal{L}_{\text{InfoNCE}} = \frac{1}{|\mathcal{Q}|}\sum_{(i,j,l,k)\in\mathcal{Q}} \log\left(\frac{p_{ij}}{p_{ij}+p_{lk}}\right) \qquad p_{ij} = \exp\left(-\frac{\|\mathbf{f}_i - \mathbf{f}_j\|_2^2}{T}\right)$$

$\Rightarrow$ approx. invariance of f_θ wrt cluster variation + smoothness of g_ϕ

Losses



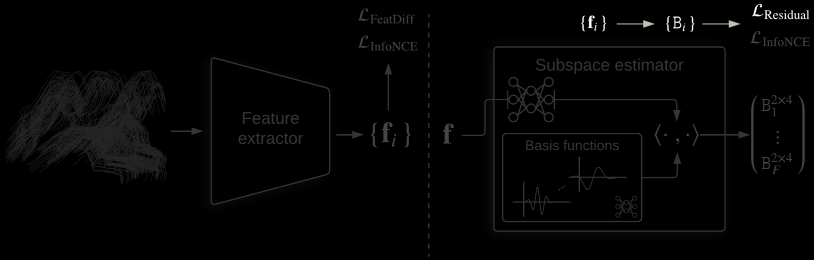
For f_θ — feature extractor, g_ϕ — subspace estimator:

$$\mathcal{L}_{\text{InfoNCE}} = \frac{1}{|\mathcal{Q}|}\sum_{(i,j,l,k)\in\mathcal{Q}} \log\left(\frac{p_{ij}}{p_{ij}+p_{lk}}\right) \qquad p_{ij} = \exp\left(-\frac{\|\mathbf{f}_i-\mathbf{f}_j\|_2^2}{T}\right)$$

$\Rightarrow$ approx. invariance of f_θ wrt cluster variation + smoothness of g_ϕ

$$\mathcal{L}_{\text{Residual}} = \sum_{\mathbf{x}} ||\mathbf{x} - \mathrm{B}\mathrm{B}^\dagger\mathbf{x}||_2^2$$

Losses



For f_θ — feature extractor, g_ϕ — subspace estimator:

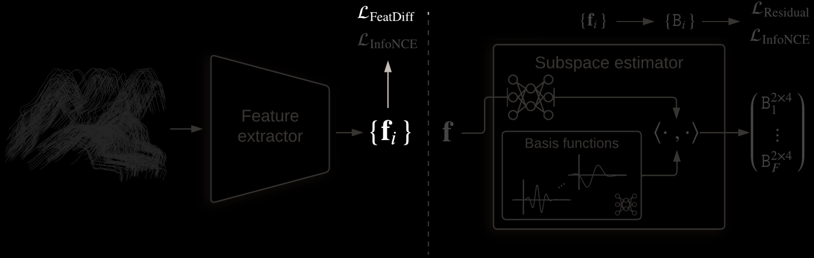
$\mathcal{L}_{\text{InfoNCE}} = \frac{1}{|\mathcal{Q}|} \sum_{(i,j,l,k)\in\mathcal{Q}} \log\left(\frac{p_{ij}}{p_{ij}+p_{lk}}\right) \qquad p_{ij} = \exp\left(-\frac{\|\mathbf{f}_i - \mathbf{f}_j\|_2^2}{T}\right)$

$\Rightarrow$ approx. invariance of f_θ wrt cluster variation + smoothness of g_ϕ

$\mathcal{L}_{\text{Residual}} = \sum_{\mathbf{x}} ||\mathbf{x} - \mathrm{BB}^\dagger \mathbf{x}||_2^2$

$\Rightarrow$ geometric consistency

Losses



For f_θ — feature extractor, g_ϕ — subspace estimator:

$\mathcal{L}_{\text{InfoNCE}} = \frac{1}{|\mathcal{Q}|}\sum_{(i,j,l,k)\in\mathcal{Q}} \log\left(\frac{p_{ij}}{p_{ij}+p_{lk}}\right) \qquad p_{ij} = \exp\left(-\frac{\|\mathbf{f}_i - \mathbf{f}_j\|_2^2}{T}\right)$

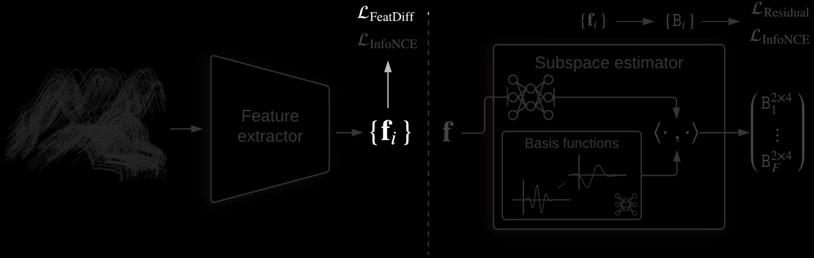
$\Rightarrow$ approx. invariance of f_θ wrt cluster variation + smoothness of g_ϕ

$\mathcal{L}_{\text{Residual}} = \sum_{\mathbf{x}} ||\mathbf{x} - \mathrm{BB}^\dagger\mathbf{x}||_2^2$

$\Rightarrow$ geometric consistency

$\mathcal{L}_{\text{FeatDiff}} = \sum_{\mathbf{x}} \|f_\theta(\mathbf{x}) - f_\theta\left(\mathrm{BB}^\dagger\mathbf{x}\right)\|_2^2$

Losses



For f_θ — feature extractor, g_ϕ — subspace estimator:

$$\mathcal{L}_{\text{InfoNCE}} = \frac{1}{|\mathcal{Q}|} \sum_{(i,j,l,k)\in\mathcal{Q}} \log\left(\frac{p_{ij}}{p_{ij}+p_{lk}}\right) \qquad p_{ij} = \exp\left(-\frac{\|\mathbf{f}_i - \mathbf{f}_j\|_2^2}{T}\right)$$

$\Rightarrow$ approx. invariance of f_θ wrt cluster variation + smoothness of g_ϕ

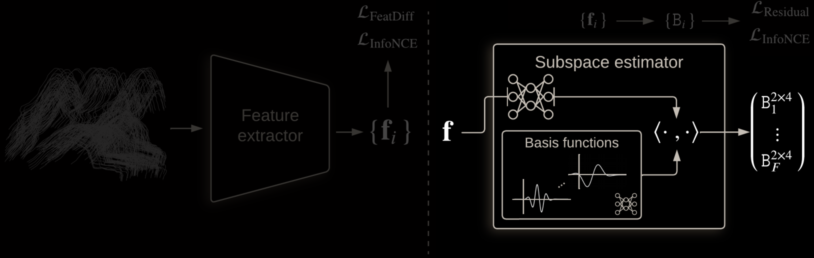
$$\mathcal{L}_{\text{Residual}} = \sum_{\mathbf{x}} ||\mathbf{x} - \mathrm{BB}^\dagger \mathbf{x}||_2^2$$

$\Rightarrow$ geometric consistency

$$\mathcal{L}_{\text{FeatDiff}} = \sum_{\mathbf{x}} \|f_\theta(\mathbf{x}) - f_\theta\left(\mathrm{BB}^\dagger \mathbf{x}\right)\|_2^2$$

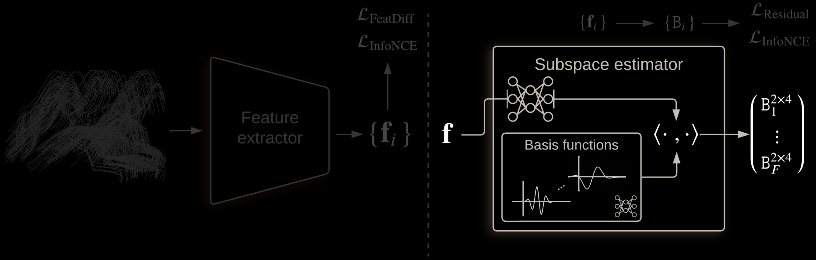
$\Rightarrow$ approx. invariance of f_θ wrt pixel noise + smoothness of f_θ

Basis Functions for Subspace Representation



Basis function can be:

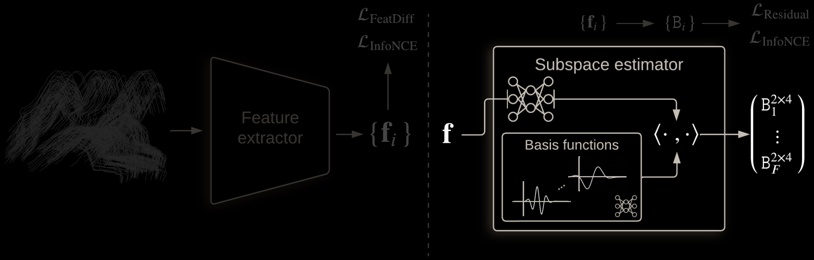
Basis Functions for Subspace Representation



Basis function can be:

- fully fixed (e.g., DCT) — too restrictive

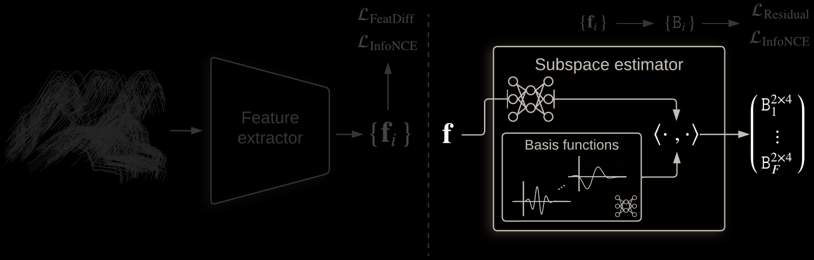
Basis Functions for Subspace Representation



Basis function can be:

- fully fixed (e.g., DCT) — too restrictive
- learned “non-parametric” (MLP) — can learn anything, too generic

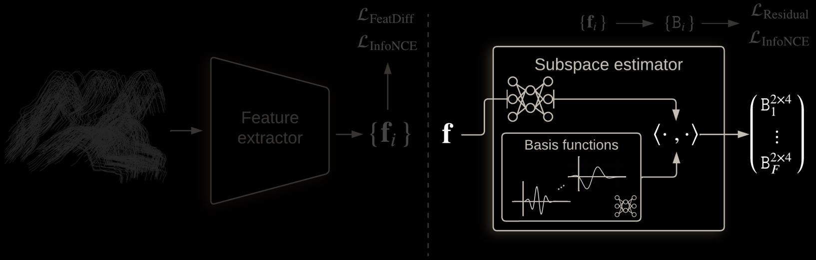
Basis Functions for Subspace Representation



Basis function can be:

- fully fixed (e.g., DCT) — too restrictive
- learned "non-parametric" (MLP) — can learn anything, too generic
- learned parametric (our choice) — trainable, but encodes temporal dependencies

Basis Functions for Subspace Representation



Basis function can be:

- fully fixed (e.g., DCT) — too restrictive
- learned "non-parametric" (MLP) — can learn anything, too generic
- learned parametric (our choice) — trainable, but encodes temporal dependencies

We use damped version of cosine basis

$$h_\psi^j(t) = e^{-(\alpha_j(t-\mu_j))^2}\cos(\beta_j t + \gamma_j)$$

Benchmark (fully visible trajectories)

	Hopkins155								
	2 motions			3 motions			All		
Method	Mean	Median	Time	Mean	Median	Time	Mean	Median	Time
RANSAC	5.56	1.18	175ms	22.94	22.03	258ms	9.76	3.21	194ms
GPCA	4.59	0.38	324ms	28.66	28.26	738ms	10.34	2.54	417ms
MSL	4.14	0.00	11h 4m	8.23	1.76	1d 23h	5.03	0.00	19h 11m
LSA	3.45	0.59	7.58s	9.73	2.33	15.96s	4.94	0.90	9.47s
ALC_5	3.03	0.00	-	6.26	1.02	-	3.76	0.26	5m 15s
ALC_{sp}	2.40	0.43	-	6.69	0.67	-	3.37	0.49	6m 11s
LRR	4.10	0.22	-	9.89	0.56	-	5.41	0.53	1.1s
SSC	0.82	0.00	-	2.45	0.20	-	2.45	0.20	920ms
RSIM	0.78	0.00	-	1.77	0.28	-	1.01	0.00	176ms
MultiCons	-	-	-	-	-	-	4.40	-	40ms
Ours	0.63	0.0	7ms	0.60	0.0	10ms	0.62	0.0	9ms

Benchmark (fully visible trajectories)

	Hopkins155								
	2 motions			3 motions			All		
Method	Mean	Median	Time	Mean	Median	Time	Mean	Median	Time
RANSAC	5.56	1.18	175ms	22.94	22.03	258ms	9.76	3.21	194ms
GPCA	4.59	0.38	324ms	28.66	28.26	738ms	10.34	2.54	417ms
MSL	4.14	0.00	11h 4m	8.23	1.76	1d 23h	5.03	0.00	19h 11m
LSA	3.45	0.59	7.58s	9.73	2.33	15.96s	4.94	0.90	9.47s
ALC_5	3.03	0.00	-	6.26	1.02	-	3.76	0.26	5m 15s
ALC_{sp}	2.40	0.43	-	6.69	0.67	-	3.37	0.49	6m 11s
LRR	4.10	0.22	-	9.89	0.56	-	5.41	0.53	1.1s
SSC	0.82	0.00	-	2.45	0.20	-	2.45	0.20	920ms
RSIM	0.78	0.00	-	1.77	0.28	-	1.01	0.00	176ms
MultiCons	-	-	-	-	-	-	4.40	-	40ms
Ours	**0.63**	**0.0**	7ms	**0.60**	**0.0**	10ms	**0.62**	**0.0**	9ms

Benchmark (fully visible trajectories)

	Hopkins155								
	2 motions			3 motions			All		
Method	Mean	Median	Time	Mean	Median	Time	Mean	Median	Time
RANSAC	5.56	1.18	175ms	22.94	22.03	258ms	9.76	3.21	194ms
GPCA	4.59	0.38	324ms	28.66	28.26	738ms	10.34	2.54	417ms
MSL	4.14	0.00	11h 4m	8.23	1.76	1d 23h	5.03	0.00	19h 11m
LSA	3.45	0.59	7.58s	9.73	2.33	15.96s	4.94	0.90	9.47s
ALC_5	3.03	0.00	-	6.26	1.02	-	3.76	0.26	5m 15s
ALC_{sp}	2.40	0.43	-	6.69	0.67	-	3.37	0.49	6m 11s
LRR	4.10	0.22	-	9.89	0.56	-	5.41	0.53	1.1s
SSC	0.82	0.00	-	2.45	0.20	-	2.45	0.20	920ms
RSIM	0.78	0.00	-	1.77	0.28	-	1.01	0.00	176ms
MultiCons	-	-	-	-	-	-	4.40	-	40ms
Ours	0.63	0.0	**7ms**	0.60	0.0	**10ms**	0.62	0.0	**9ms**

Trajectory Completion

- Let $\mathbf{x}$ contain missing values with pattern $\mathbf{w}$

Trajectory Completion

- Let $\mathbf{x}$ contain missing values with pattern $\mathbf{w}$
- $\hat{\mathbf{x}}(\bar{\mathbf{x}}) := \mathbf{w} \odot \mathbf{x} + \bar{\mathbf{w}} \odot \bar{\mathbf{x}}$

Trajectory Completion

- Let $\mathbf{x}$ contain missing values with pattern $\mathbf{w}$
- $\hat{\mathbf{x}}(\bar{\mathbf{x}}) := \mathbf{w} \odot \mathbf{x} + \bar{\mathbf{w}} \odot \bar{\mathbf{x}}$
- Objective of trajectory completion

$$\left\| \hat{\mathbf{x}}(\bar{\mathbf{x}}) - \mathtt{B}\mathtt{B}^{\dagger} \hat{\mathbf{x}}(\bar{\mathbf{x}}) \right\|^2 \to \min_{\bar{\mathbf{x}}} \qquad (\mathtt{B} = B_{\theta,\phi}(\hat{\mathbf{x}}, \mathbf{t}) \text{ – output of the network})$$

Trajectory Completion

- Let $\mathbf{x}$ contain missing values with pattern $\mathbf{w}$
- $\hat{\mathbf{x}}(\bar{\mathbf{x}}) := \mathbf{w} \odot \mathbf{x} + \bar{\mathbf{w}} \odot \bar{\mathbf{x}}$
- Objective of trajectory completion

$$\left\| \hat{\mathbf{x}}(\bar{\mathbf{x}}) - \mathtt{B}\mathtt{B}^{\dagger}\hat{\mathbf{x}}(\bar{\mathbf{x}}) \right\|^2 \to \min_{\bar{\mathbf{x}}} \qquad (\mathtt{B} = B_{\theta,\phi}(\hat{\mathbf{x}}, \mathbf{t}) \text{ – output of the network})$$

- Linear solution for a fixed $\mathtt{B}$

$$\bar{\mathbf{x}}^* = \mathtt{A}(\mathtt{B})\mathbf{x}$$

Trajectory Completion

- Let $\mathbf{x}$ contain missing values with pattern $\mathbf{w}$
- $\hat{\mathbf{x}}(\bar{\mathbf{x}}) := \mathbf{w} \odot \mathbf{x} + \bar{\mathbf{w}} \odot \bar{\mathbf{x}}$
- Objective of trajectory completion

$$\left\| \hat{\mathbf{x}}(\bar{\mathbf{x}}) - \mathtt{B}\mathtt{B}^{\dagger} \hat{\mathbf{x}}(\bar{\mathbf{x}}) \right\|^2 \to \min_{\bar{\mathbf{x}}} \qquad (\mathtt{B} = B_{\theta,\phi}(\hat{\mathbf{x}}, \mathbf{t}) \text{ -- output of the network})$$

- Linear solution for a fixed $\mathtt{B}$

$$\bar{\mathbf{x}}^* = \mathtt{A}(\mathtt{B})\mathbf{x}$$

- Yields iterative procedure

$$\begin{cases} \mathtt{B}_0 \leftarrow B_{\theta,\phi}(\mathbf{x}_{\mathsf{vis}}, \mathbf{t}) \\ \bar{\mathbf{x}}_i \leftarrow \mathtt{A}(\mathtt{B}_{i-1})\mathbf{x} \\ \mathtt{B}_i \leftarrow B_{\theta,\phi}(\mathbf{w} \odot \mathbf{x} + \bar{\mathbf{w}} \odot \bar{\mathbf{x}}_i, \mathbf{t}) \end{cases}$$

Trajectory Completion

- Let $\mathbf{x}$ contain missing values with pattern $\mathbf{w}$
- $\hat{\mathbf{x}}(\bar{\mathbf{x}}) := \mathbf{w} \odot \mathbf{x} + \bar{\mathbf{w}} \odot \bar{\mathbf{x}}$
- Objective of trajectory completion

$$\left\| \hat{\mathbf{x}}(\bar{\mathbf{x}}) - \mathtt{B}\mathtt{B}^{\dagger} \hat{\mathbf{x}}(\bar{\mathbf{x}}) \right\|^2 \to \min_{\bar{\mathbf{x}}} \qquad (\mathtt{B} = B_{\theta,\phi}(\hat{\mathbf{x}}, \mathbf{t}) \text{ – output of the network})$$

- Linear solution for a fixed $\mathtt{B}$

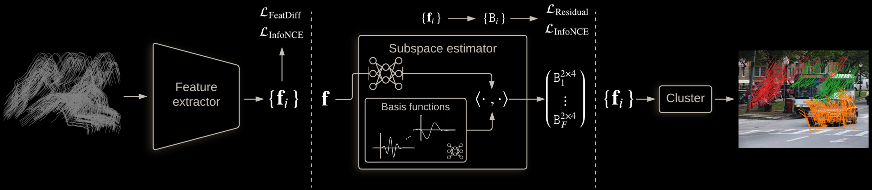
$$\bar{\mathbf{x}}^* = \mathtt{A}(\mathtt{B})\mathbf{x}$$

- Yields iterative procedure

$$\begin{cases} \mathtt{B}_0 \leftarrow B_{\theta,\phi}(\mathbf{x}_{\mathsf{vis}}, \mathbf{t}) \\ \bar{\mathbf{x}}_i \leftarrow \mathtt{A}(\mathtt{B}_{i-1})\mathbf{x} \\ \mathtt{B}_i \leftarrow B_{\theta,\phi}(\mathbf{w} \odot \mathbf{x} + \bar{\mathbf{w}} \odot \bar{\mathbf{x}}_i, \mathbf{t}) \end{cases}$$

- Approximate block-coordinate descent

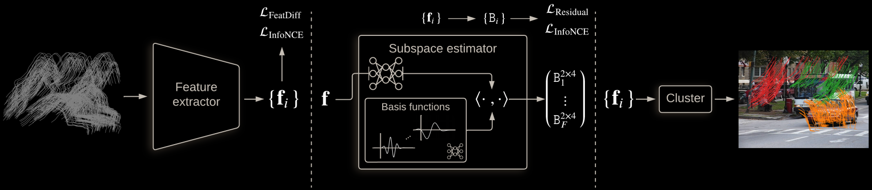
Framework



The network is trained on fully observed trajectories.

*ignoring uniform occlusions

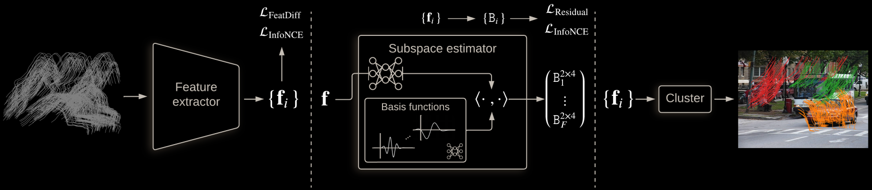
Framework



The network is trained on fully observed trajectories. During inference:

*ignoring uniform occlusions

Framework

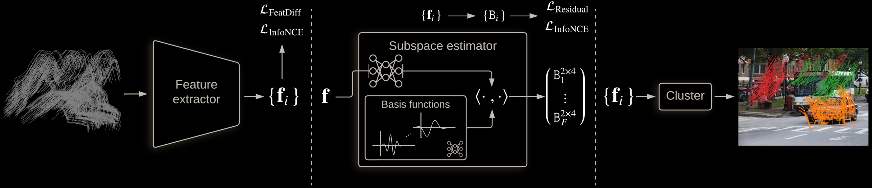


The network is trained on fully observed trajectories. During inference:

- Handling occlusions: full forward pass for the largest fully visible trajectory block* → initial subspaces B → iterative completion.

*ignoring uniform occlusions

Framework

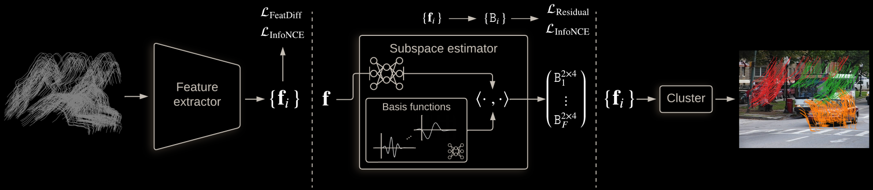


The network is trained on fully observed trajectories. During inference:

- Handling occlusions: full forward pass for the largest fully visible trajectory block* → initial subspaces B → iterative completion.
- Grouping: partial forward pass through f_θ, followed by clustering in the feature space of all scene trajectories.

*ignoring uniform occlusions

Framework

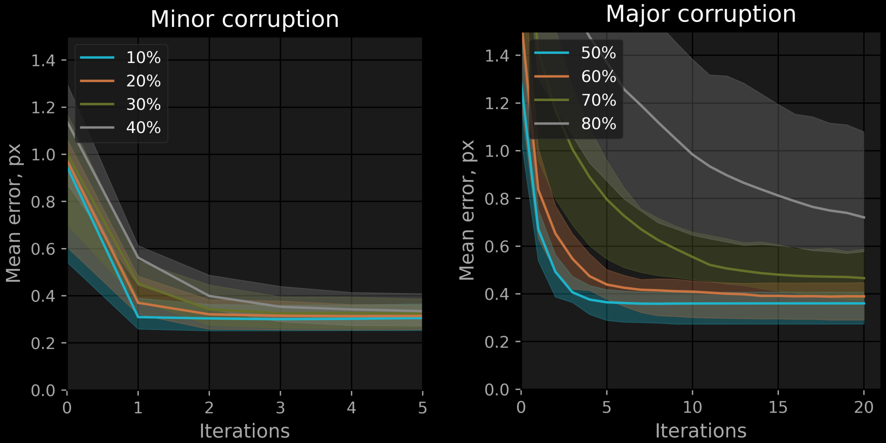


The network is trained on fully observed trajectories. During inference:

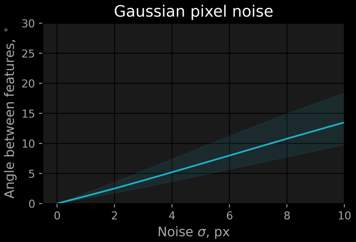
- Handling occlusions: full forward pass for the largest fully visible trajectory block* $\rightarrow$ initial subspaces B $\rightarrow$ iterative completion.
- Grouping: partial forward pass through f_θ, followed by clustering in the feature space of all scene trajectories.
- Model estimation: grouping, followed by linear subspace fitting.

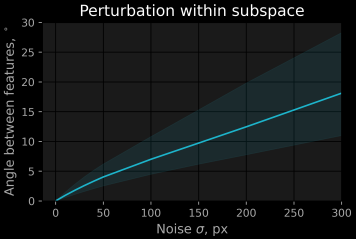
*ignoring uniform occlusions

Recovering from Uniform Occlusions

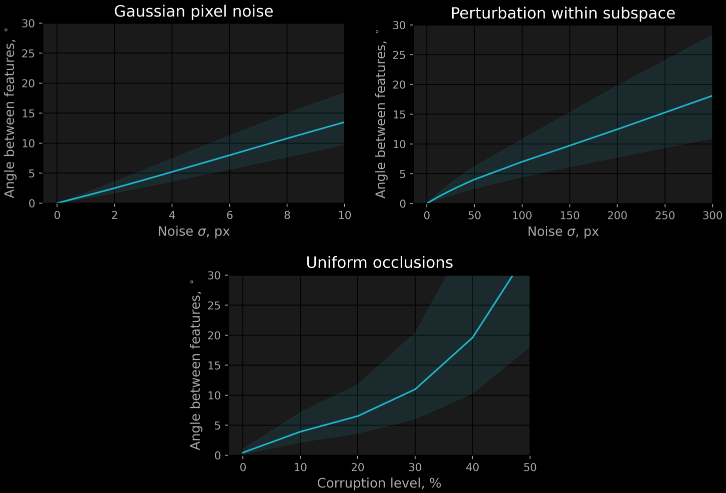


Approximate Invariances of f_θ

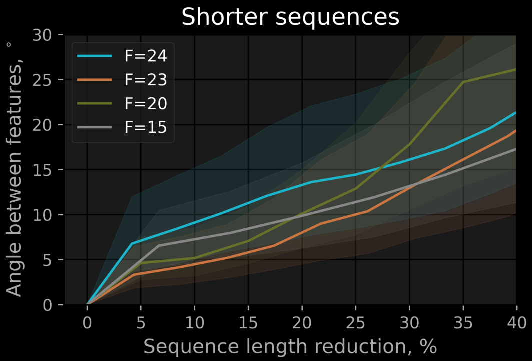




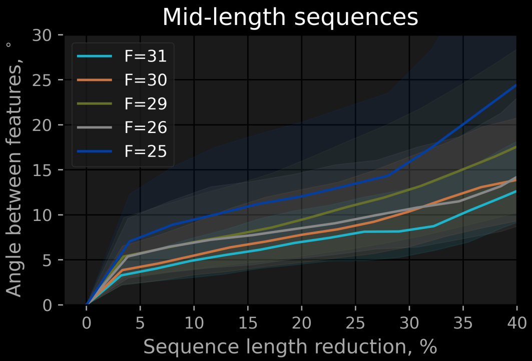
Approximate Invariances of f_θ



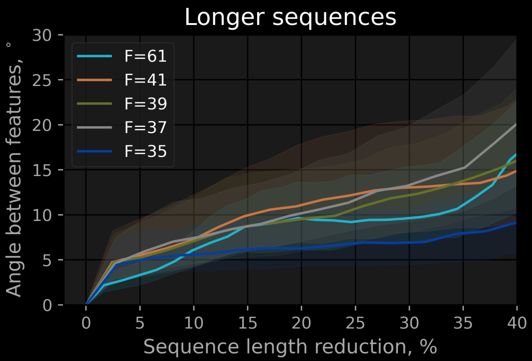
Synthesized Tracking Failure



Synthesized Tracking Failure



Synthesized Tracking Failure



Benchmark

	Hopkins155			Hopkins12		KT3DMoSeg	
Method	Mean	Median	Time	Mean	Median	Mean	Median
RANSAC	9.76	3.21	194ms	-	-	-	-
GPCA	10.34	2.54	417ms	-	-	34.60	33.95
MSL	5.03	0.00	19h 11m	-	-	-	-
LSA	4.94	0.90	9.47s	-	-	38.30	38.58
ALC_5	3.76	0.26	5m 15s	3.81	0.17	24.31	19.04
ALC_{sp}	3.37	0.49	6m 11s	1.28	1.07	-	-
LRR	5.41	0.53	1.1s	-	-	33.67	36.01
SSC	2.45	0.20	920ms	-	-	33.88	33.54
RSIM	1.01	0.00	176ms	0.68	0.70	-	-
MultiCons	4.40	-	40ms	-	-	-	-
Ours	0.62	0.0	9ms	5.12	2.04	5.85	0.80

Benchmark

	Hopkins155			Hopkins12		KT3DMoSeg	
Method	Mean	Median	Time	Mean	Median	Mean	Median
RANSAC	9.76	3.21	194ms	-	-	-	-
GPCA	10.34	2.54	417ms	-	-	34.60	33.95
MSL	5.03	0.00	19h 11m	-	-	-	-
LSA	4.94	0.90	9.47s	-	-	38.30	38.58
ALC_5	3.76	0.26	5m 15s	3.81	0.17	24.31	19.04
ALC_{sp}	3.37	0.49	6m 11s	1.28	1.07	-	-
LRR	5.41	0.53	1.1s	-	-	33.67	36.01
SSC	2.45	0.20	920ms	-	-	33.88	33.54
RSIM	1.01	0.00	176ms	0.68	0.70	-	-
MultiCons	4.40	-	40ms	-	-	-	-
Ours	**0.62**	**0.0**	9ms	5.12	2.04	5.85	0.80

Benchmark

	Hopkins155			Hopkins12		KT3DMoSeg	
Method	Mean	Median	Time	Mean	Median	Mean	Median
RANSAC	9.76	3.21	194ms	-	-	-	-
GPCA	10.34	2.54	417ms	-	-	34.60	33.95
MSL	5.03	0.00	19h 11m	-	-	-	-
LSA	4.94	0.90	9.47s	-	-	38.30	38.58
ALC_5	3.76	0.26	5m 15s	3.81	0.17	24.31	19.04
ALC_{sp}	3.37	0.49	6m 11s	1.28	1.07	-	-
LRR	5.41	0.53	1.1s	-	-	33.67	36.01
SSC	2.45	0.20	920ms	-	-	33.88	33.54
RSIM	1.01	0.00	176ms	0.68	0.70	-	-
MultiCons	4.40	-	40ms	-	-	-	-
Ours	0.62	0.0	**9ms**	5.12	2.04	5.85	0.80

Benchmark

	Hopkins155			Hopkins12		KT3DMoSeg	
Method	Mean	Median	Time	Mean	Median	Mean	Median
RANSAC	9.76	3.21	194ms	-	-	-	-
GPCA	10.34	2.54	417ms	-	-	34.60	33.95
MSL	5.03	0.00	19h 11m	-	-	-	-
LSA	4.94	0.90	9.47s	-	-	38.30	38.58
ALC_5	3.76	0.26	5m 15s	3.81	0.17	24.31	19.04
ALC_{sp}	3.37	0.49	6m 11s	1.28	1.07	-	-
LRR	5.41	0.53	1.1s	-	-	33.67	36.01
SSC	2.45	0.20	920ms	-	-	33.88	33.54
RSIM	1.01	0.00	176ms	0.68	0.70	-	-
MultiCons	4.40	-	40ms	-	-	-	-
Ours	0.62	0.0	9ms	**5.12**	**2.04**	5.85	0.80

Benchmark

	Hopkins155			Hopkins12		KT3DMoSeg	
Method	Mean	Median	Time	Mean	Median	Mean	Median
RANSAC	9.76	3.21	194ms	-	-	-	-
GPCA	10.34	2.54	417ms	-	-	34.60	33.95
MSL	5.03	0.00	19h 11m	-	-	-	-
LSA	4.94	0.90	9.47s	-	-	38.30	38.58
ALC_5	3.76	0.26	5m 15s	3.81	0.17	24.31	19.04
ALC_{sp}	3.37	0.49	6m 11s	1.28	1.07	-	-
LRR	5.41	0.53	1.1s	-	-	33.67	36.01
SSC	2.45	0.20	920ms	-	-	33.88	33.54
RSIM	1.01	0.00	176ms	0.68	0.70	-	-
MultiCons	4.40	-	40ms	-	-	-	-
Ours	0.62	0.0	9ms	5.12	2.04	**5.85**	**0.80**

Future work

- Generalization
  - Synthetic data generation

Future work

- Generalization
  - Synthetic data generation
- Model
  - Affine $\rightarrow$ pinhole camera model
  - Priors on the shape matrix $\mathtt{C}$
  - Temporal uncertainty

Future work

- Generalization
  - Synthetic data generation
- Model
  - Affine $\rightarrow$ pinhole camera model
  - Priors on the shape matrix $\mathtt{C}$
  - Temporal uncertainty
- Architecture
  - Incorporate global context
  - Transformers: better than convolutions? possibility of attention-based completion

Thank you!

- Q&A

Thank you!

- Q&A
- Email: `lochman@chalmers.se`

Project page

`ylochman.github.io/trajectory-embedding`