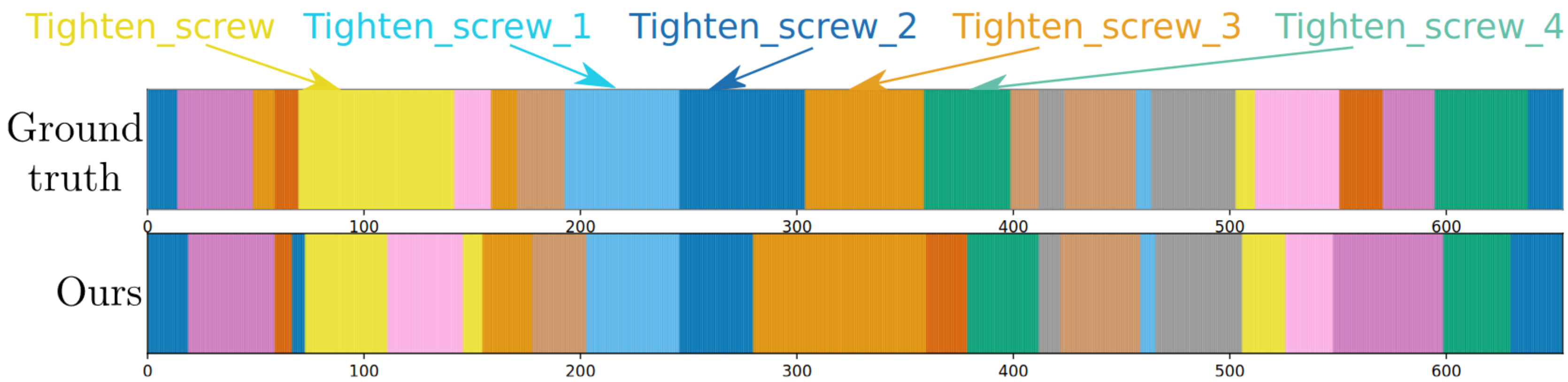
Deep kernel video approximation for unsupervised action segmentation

[1] S.L. Pintea and [2] J. Dijkstra

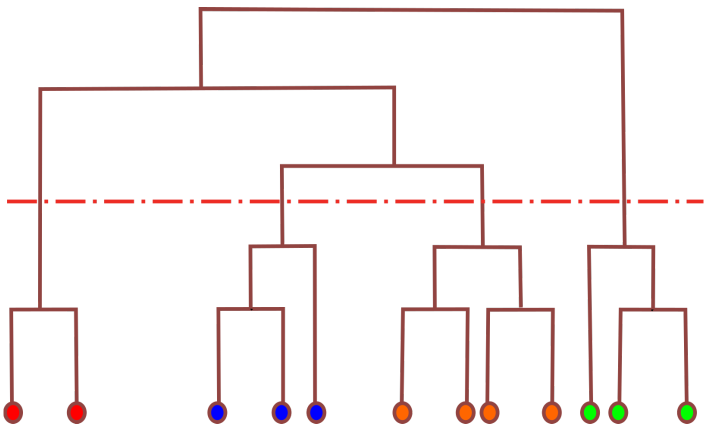
[1] Tilburg University, Netherlands
[2] Leiden University Medical Center, Netherlands

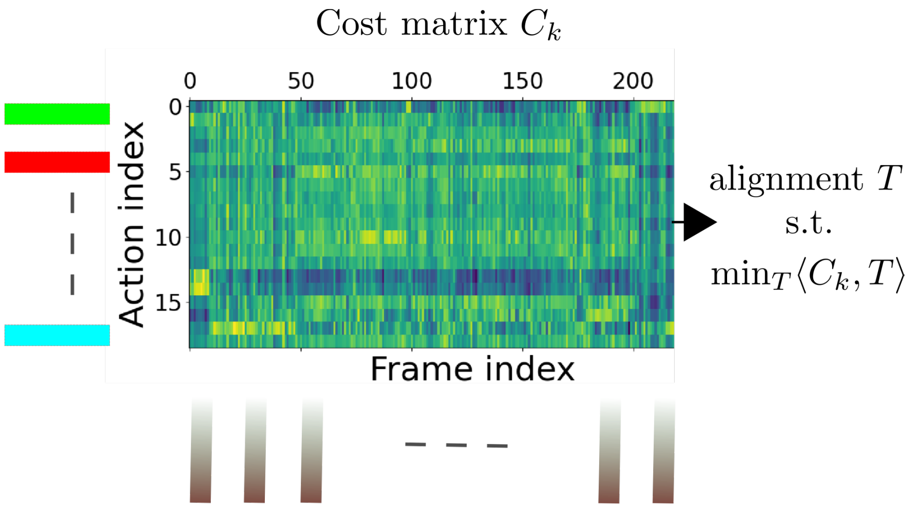
Unsupervised action segmentation



- **Aim**: Detect *unitary* actions in videos
  - How do we define what a *unitary* action is?

Prior work





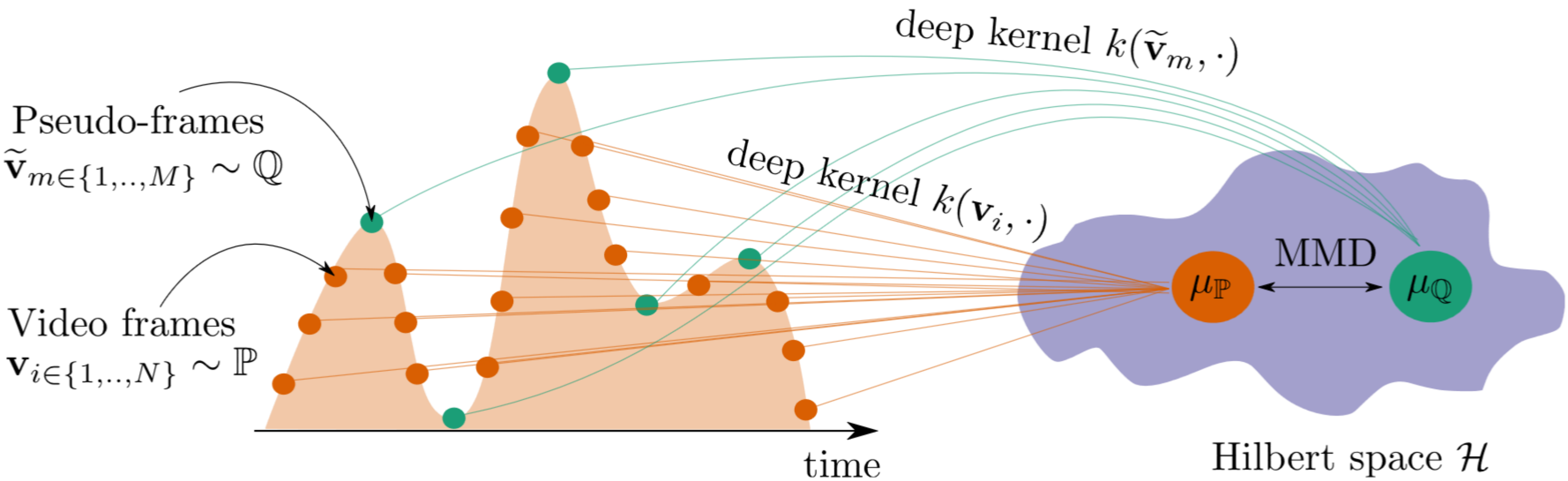
- S. Sarfraz, et. al., *“Temporally-weighted hierarchical clustering for unsupervised action segmentation”*, in CVPR, 2021.
- Z. Du, et. al., *“Fast and unsupervised action boundary detection for action segmentation”*, in CVPR, 2022.
- M. Xu and S. Gould, *“Temporally consistent unbalanced optimal transport for unsupervised action segmentation”*, in CVPR, 2024.
- E. Bueno-Benito and M. Dimiccoli, *“Clot: Closed loop optimal transport for unsupervised action segmentation”*, in ICCV, 2025.

Learning video approximations

- Single video processed at a time (sensitive data: e.g. hospitals)

Learning video approximations

- Single video processed at a time (sensitive data: e.g. hospitals)

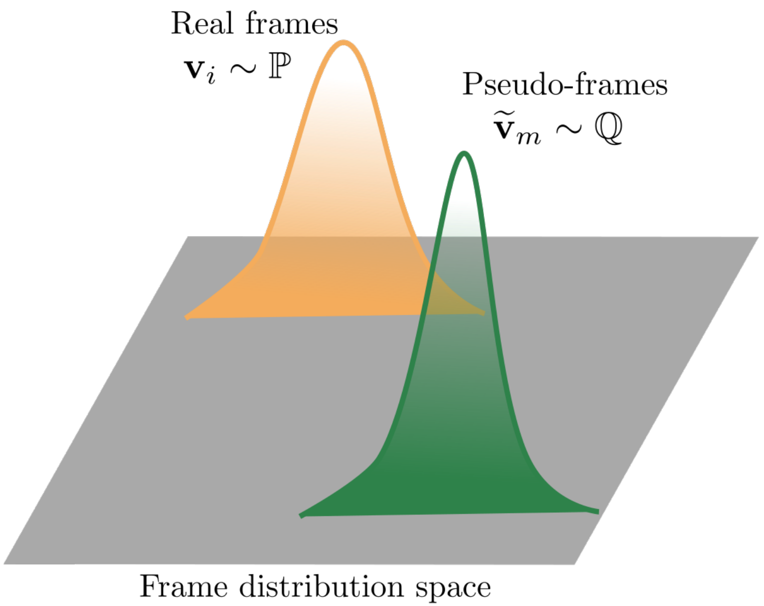


- Learn pseudo-frames $\widetilde{\mathbf{v}}_m \sim \mathbb{Q}$ such that they approximate *"as close as possible"* the distribution of real frames $\mathbf{v}_i \sim \mathbb{P}$

T. Nguyen, Z. Chen, and J. Lee, *“Dataset meta-learning from kernel ridgeregression”* in ICLR, 2020.

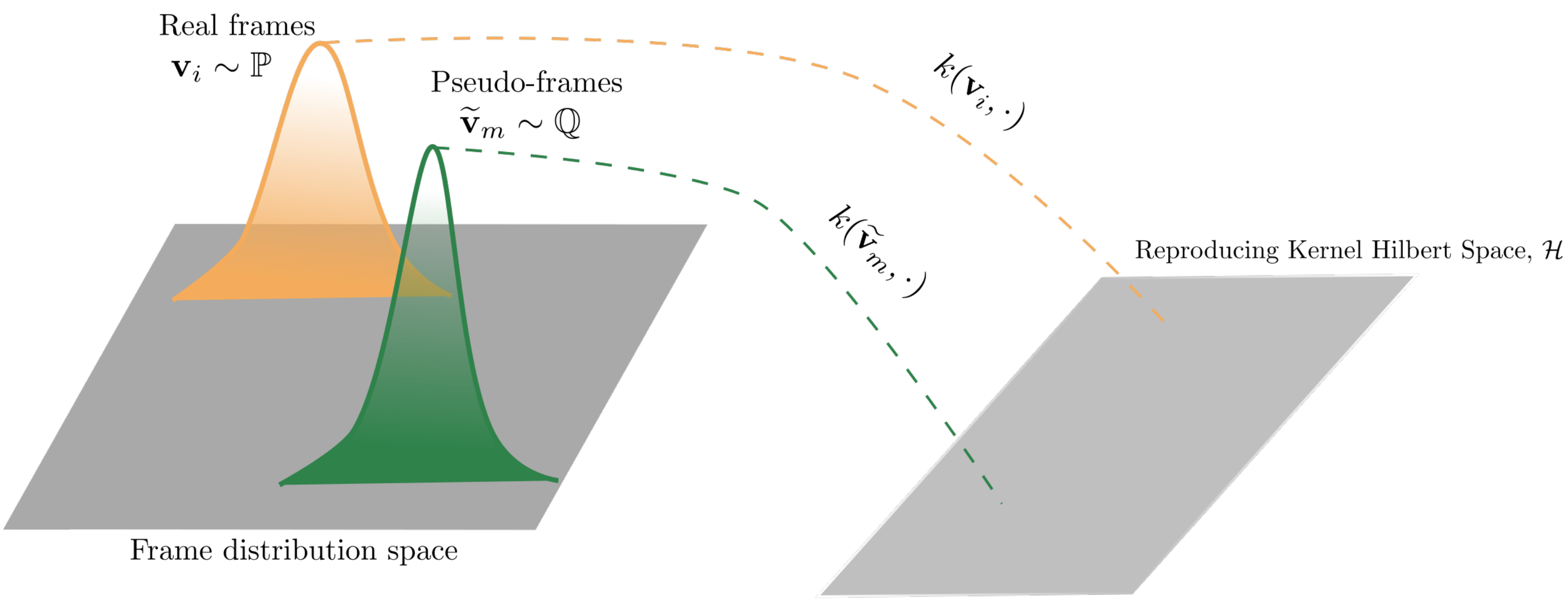
Metric choice

- *"As close as possible"* $\rightarrow$ maximum mean discrepancy (MMD)



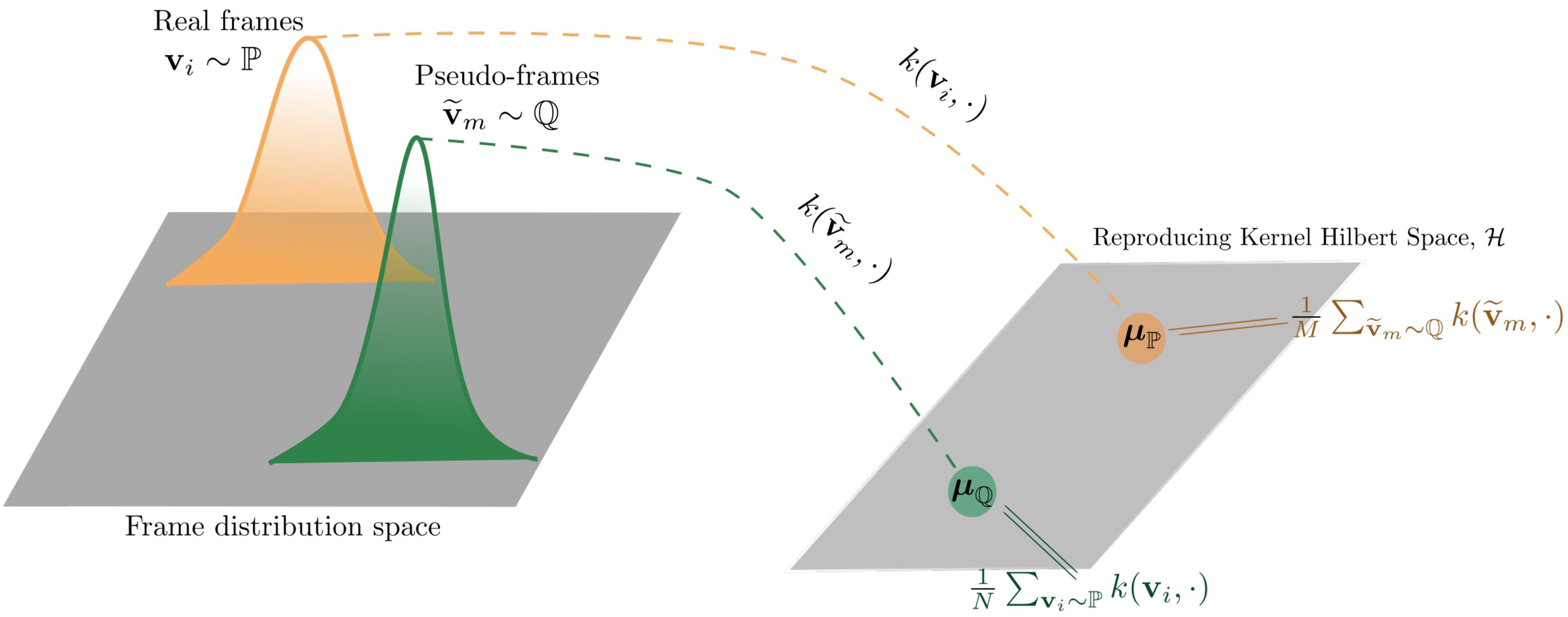
Metric choice

- *"As close as possible"* $\rightarrow$ maximum mean discrepancy (MMD)



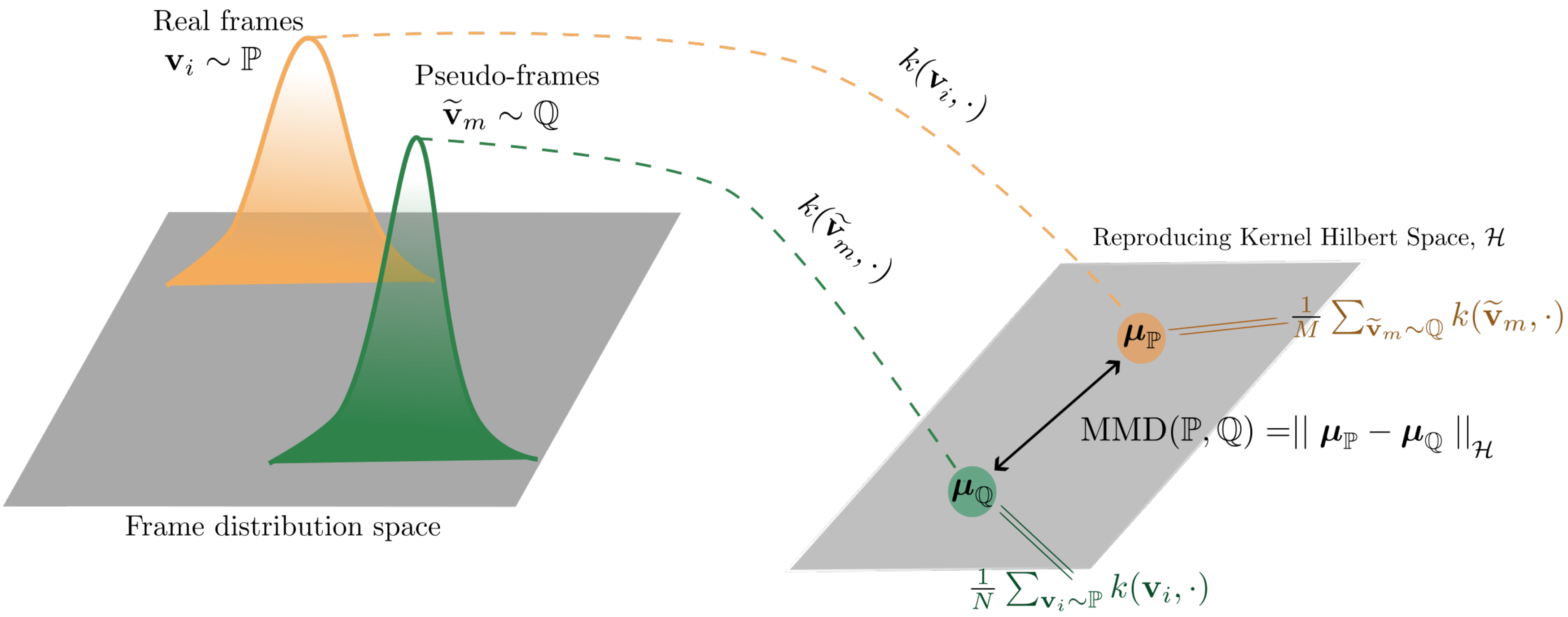
Metric choice

- *"As close as possible"* $\rightarrow$ maximum mean discrepancy (MMD)



Metric choice

- *"As close as possible"* $\rightarrow$ maximum mean discrepancy (MMD)



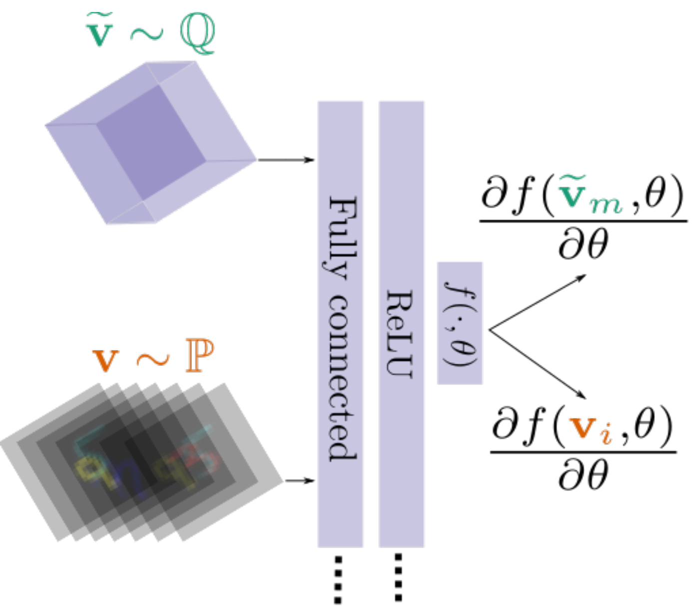
Kernel choice

- Catch: $\mu_{\mathbb{P}} = \mu_{\mathbb{Q}} \iff \mathbb{P} = \mathbb{Q}$ only holds for characteristic kernels (s.a. exponential family).

- Infinite NTKs *(S. Arora et. al., NeurIPS, 2019)* are more expressive than fixed exponential kernels (but not characteristic)

Kernel choice

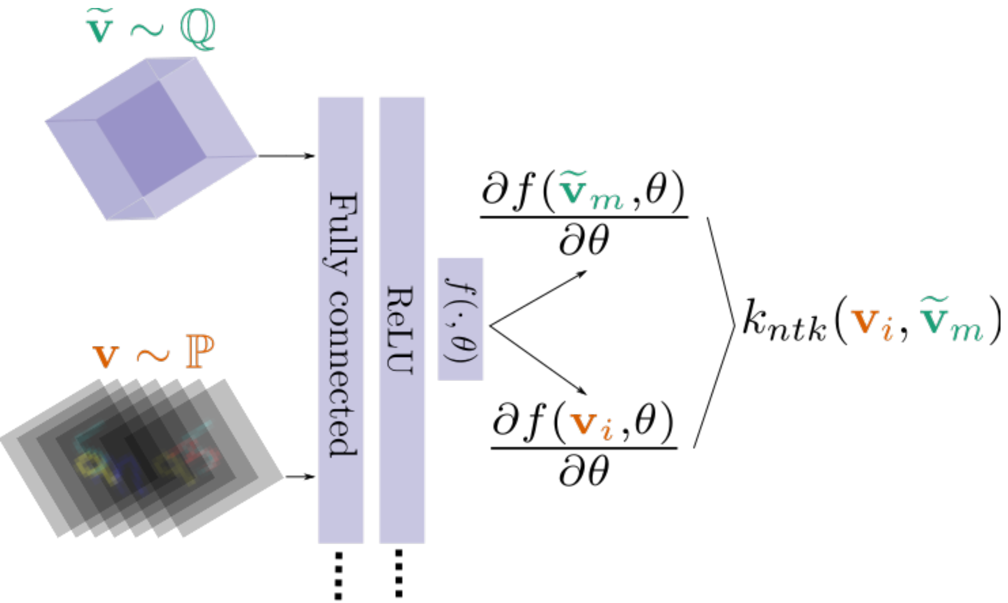
- Catch: $\mu_{\mathbb{P}} = \mu_{\mathbb{Q}} \iff \mathbb{P} = \mathbb{Q}$ only holds for characteristic kernels (s.a. exponential family).
- Infinite NTKs *(S. Arora et. al., NeurIPS, 2019)* are more expressive than fixed exponential kernels (but not characteristic)



Kernel choice

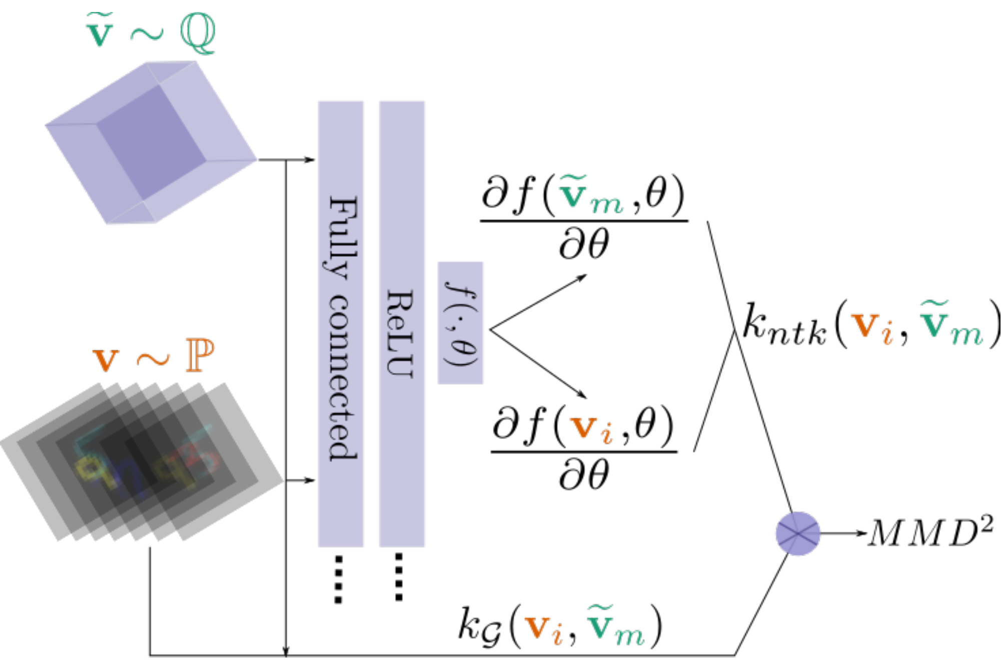
- Catch: $\mu_{\mathbb{P}} = \mu_{\mathbb{Q}} \iff \mathbb{P} = \mathbb{Q}$ only holds for characteristic kernels (s.a. exponential family).

- Infinite NTKs *(S. Arora et. al., NeurIPS, 2019)* are more expressive than fixed exponential kernels (but not characteristic)



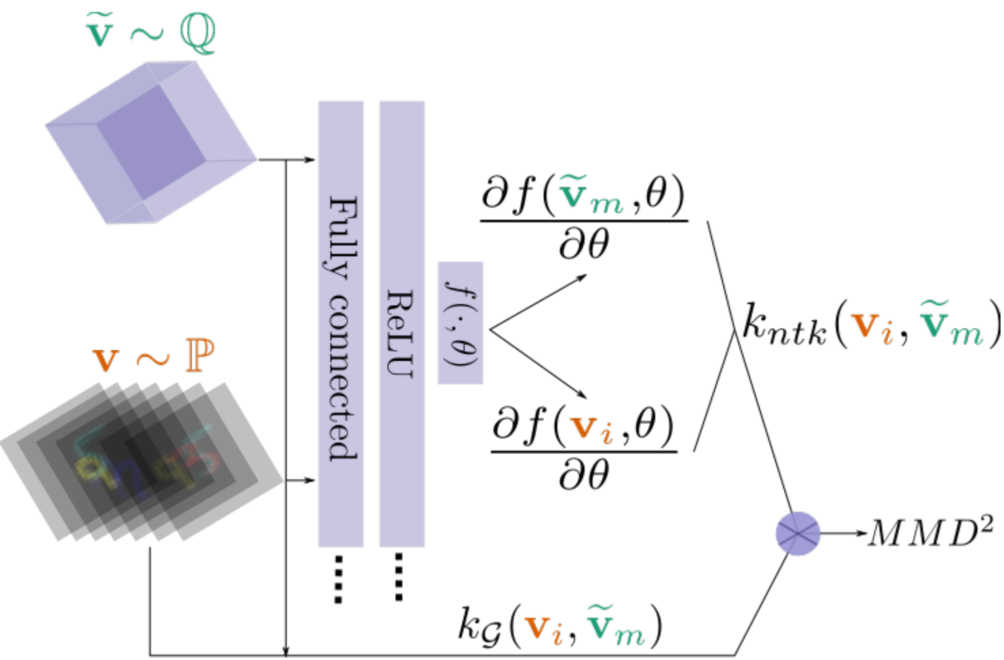
Kernel choice

- Catch: $\mu_{\mathbb{P}} = \mu_{\mathbb{Q}} \iff \mathbb{P} = \mathbb{Q}$ only holds for characteristic kernels (s.a. exponential family).
- Infinite NTKs *(S. Arora et. al., NeurIPS, 2019)* are more expressive than fixed exponential kernels (but not characteristic)



Kernel choice

- Catch: $\mu_{\mathbb{P}} = \mu_{\mathbb{Q}} \iff \mathbb{P} = \mathbb{Q}$ only holds for characteristic kernels (s.a. exponential family).
- Infinite NTKs *(S. Arora et. al., NeurIPS, 2019)* are more expressive than fixed exponential kernels (but not characteristic)

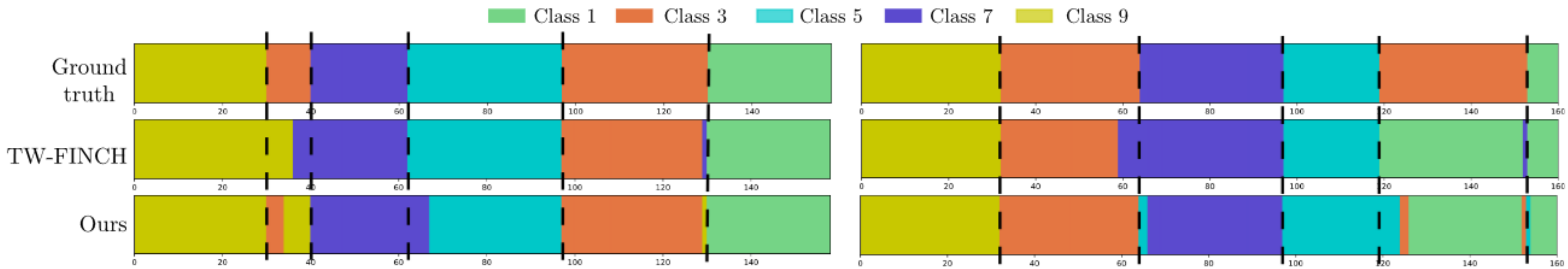


	MoF (↑)	IoU (↑)	F_1 (↑)
Gauss	62.1%	46.0%	56.7%
NNGP [48]	67.7%	46.2%	56.6%
NTK [30]	69.1%	42.3%	52.6%
NTK sphere [33]	67.1%	46.6%	57.3%
SNTK [34] [1]	62.7%	45.7%	56.5%
Gauss×NTK sphere	69.8%	44.6%	55.0%
Gauss×NTK (Ours)	70.1%	44.9%	55.3%

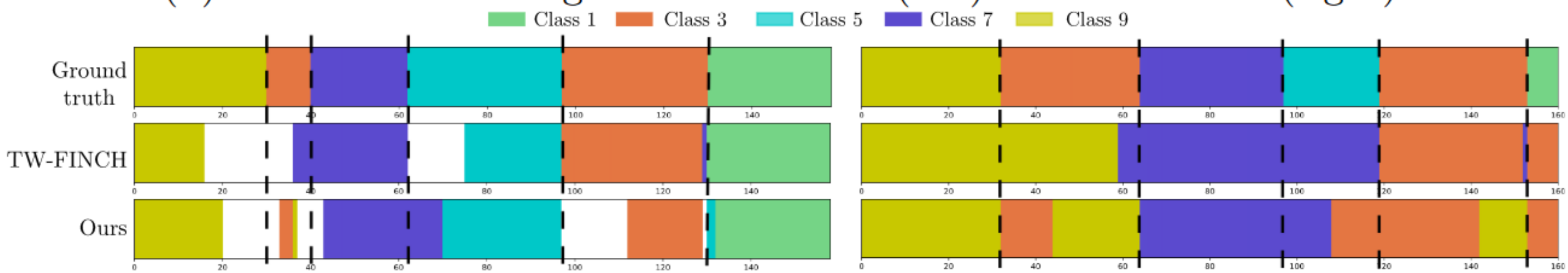
Does it work?

	True # actions			Random # actions	
	MoF (↑)	Boundary acc. (↑)		MoF (↑)	Boundary acc. (↑)
TW-FINCH [2]	72.5%	64.4%	TW-FINCH [2]	60.3%	58.8%
Ours	73.3%	51.2%	Ours	60.5%	62.4%

(a) Segmentation with true classes and random number of classes



(b) True number of segments: 5^{th} video (left) and 35^{th} video (right).



(c) Random number of segments: more (5^{th} video) and less (35^{th} video).

Thank you!

?