LAMARR

Institute for Machine Learning and Artificial Intelligence

UNIVERSITÄT BONN

Juergen Gall

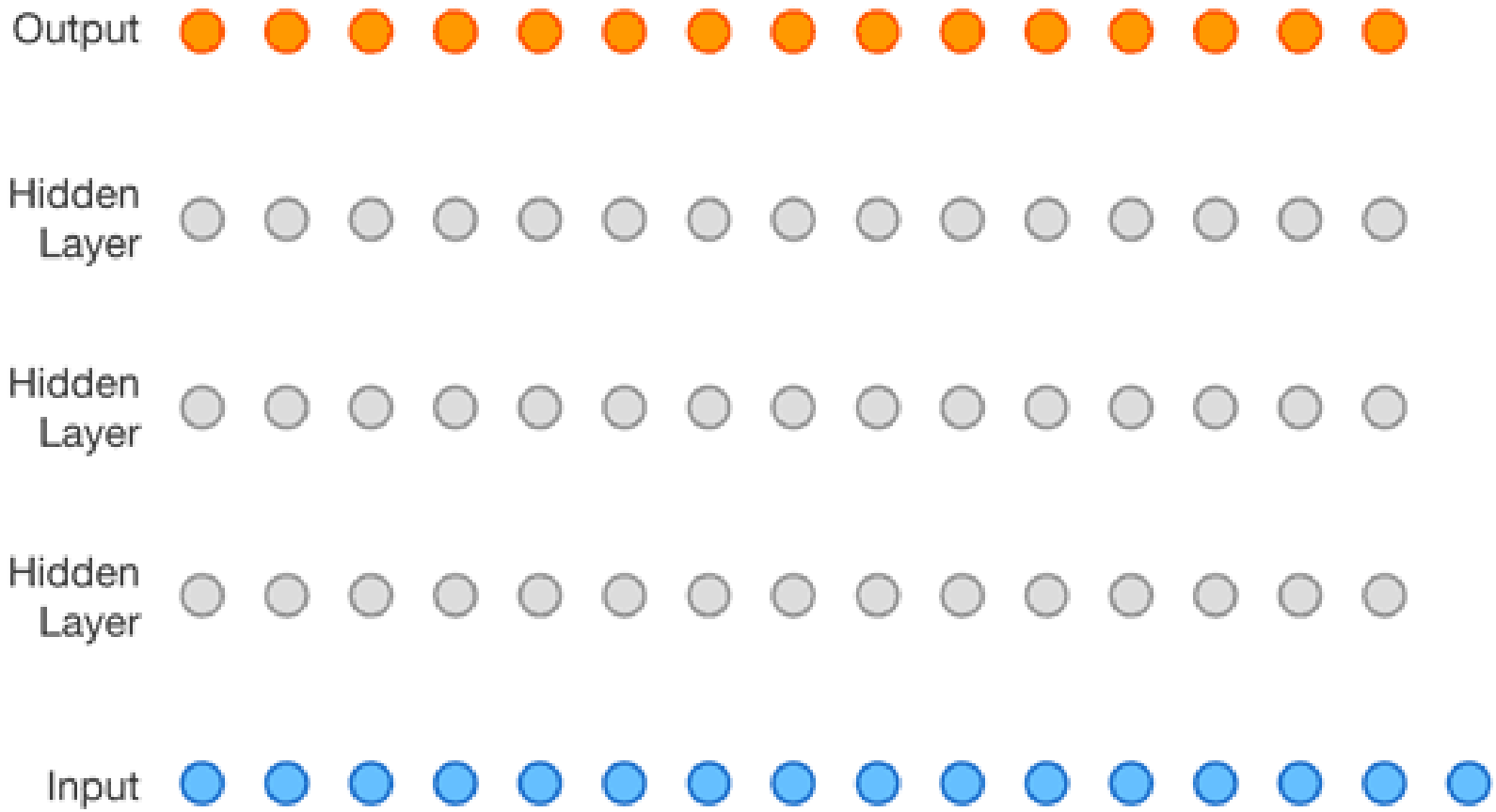
From Forecasting Human Behavior to Forecasting Extreme Weather and Climate Events

Temporal Action Segmentation

[E. Bahrami et al. **How Much Temporal Long-Term Context is Needed for Action Segmentation?** ICCV 2023]

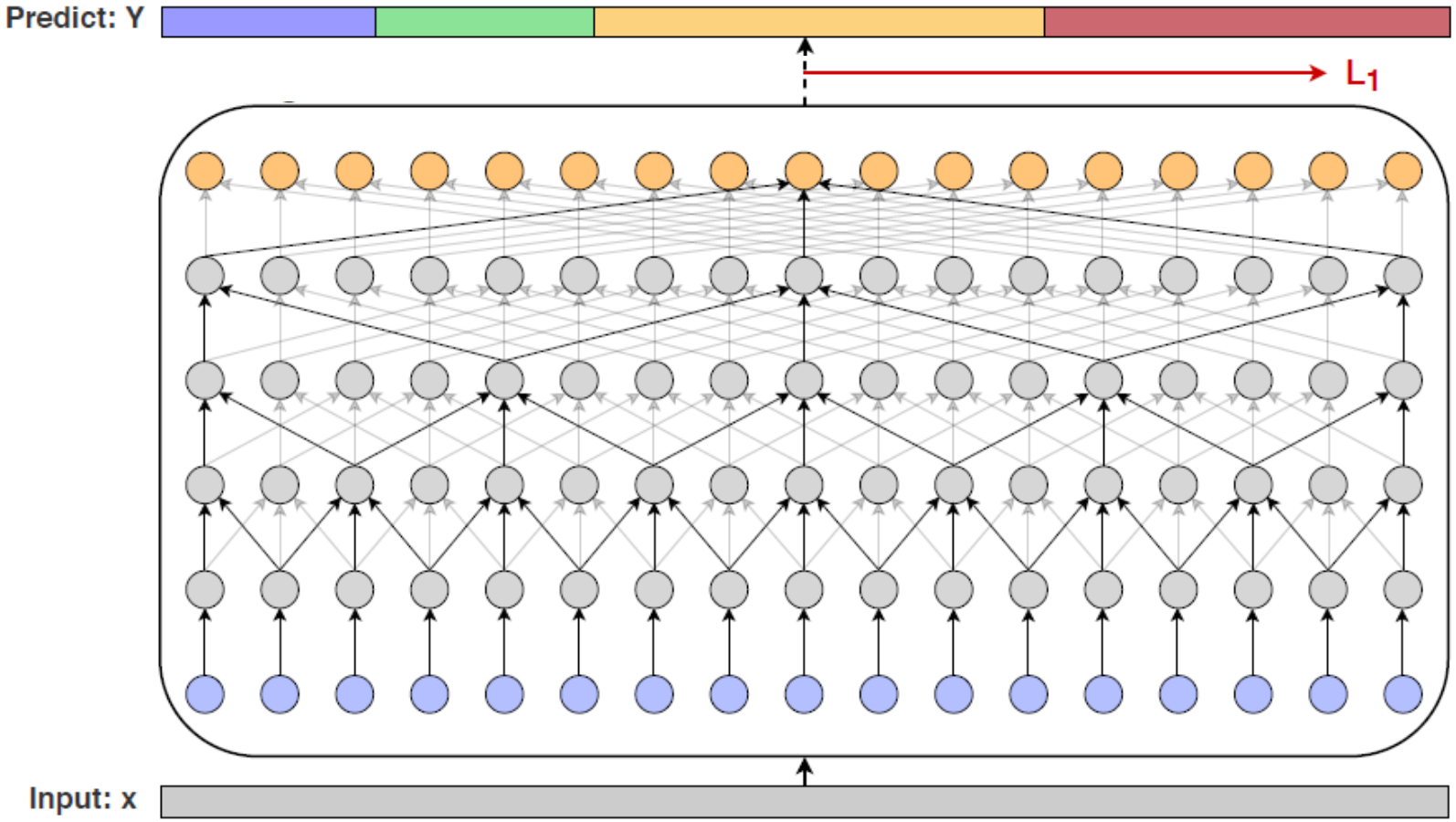
Temporal Convolutional Neural Network

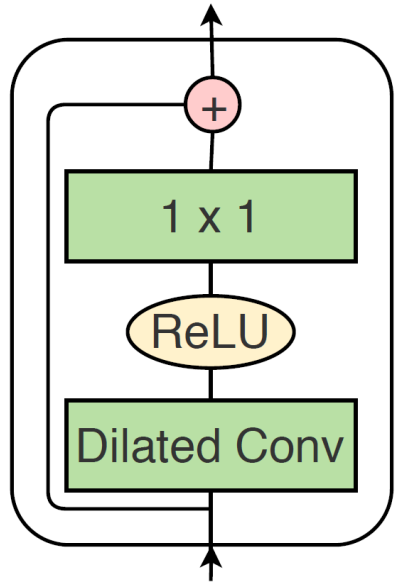
- Dilated convolutions for audio



[van den Oord et al. **WaveNet: A Generative Model for Raw Audio.** SSW 2016]

Multi-Stage Temporal Convolutional Network

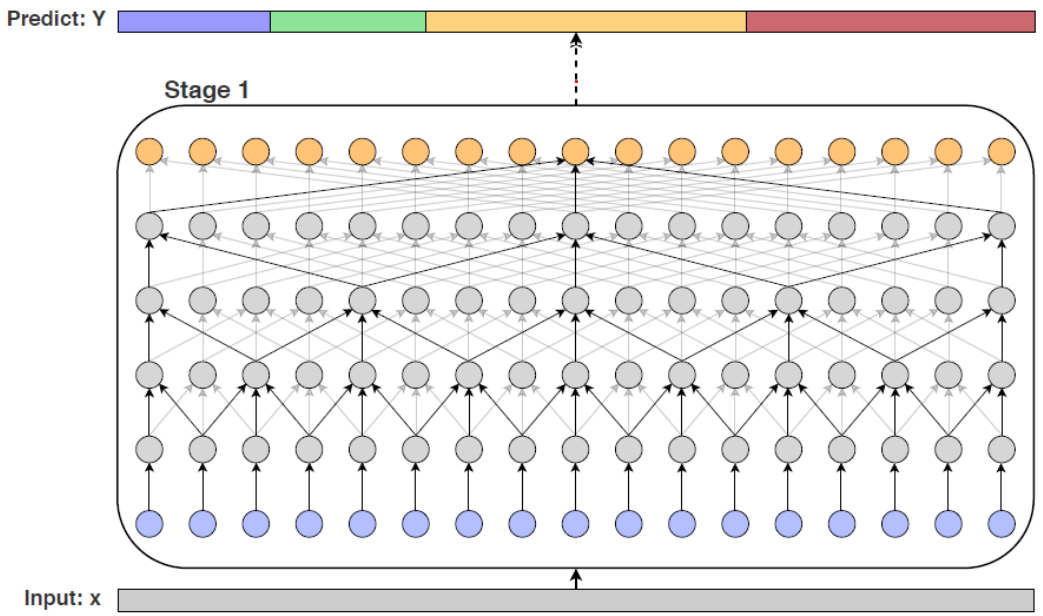


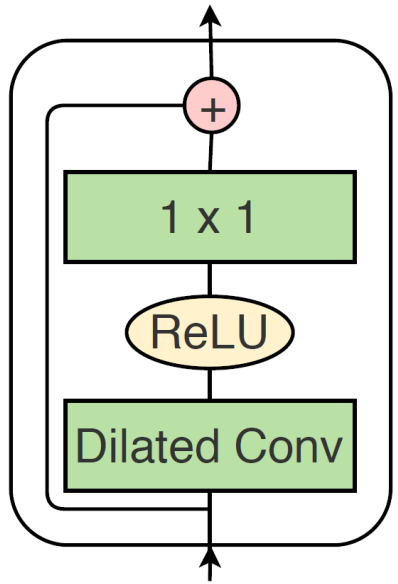


[Y. Abu Farha and J. Gall. **MS-TCN: Multi-Stage Temporal Convolutional Network for Action Segmentation.** CVPR 2019]

Multi-Stage Temporal Convolutional Network

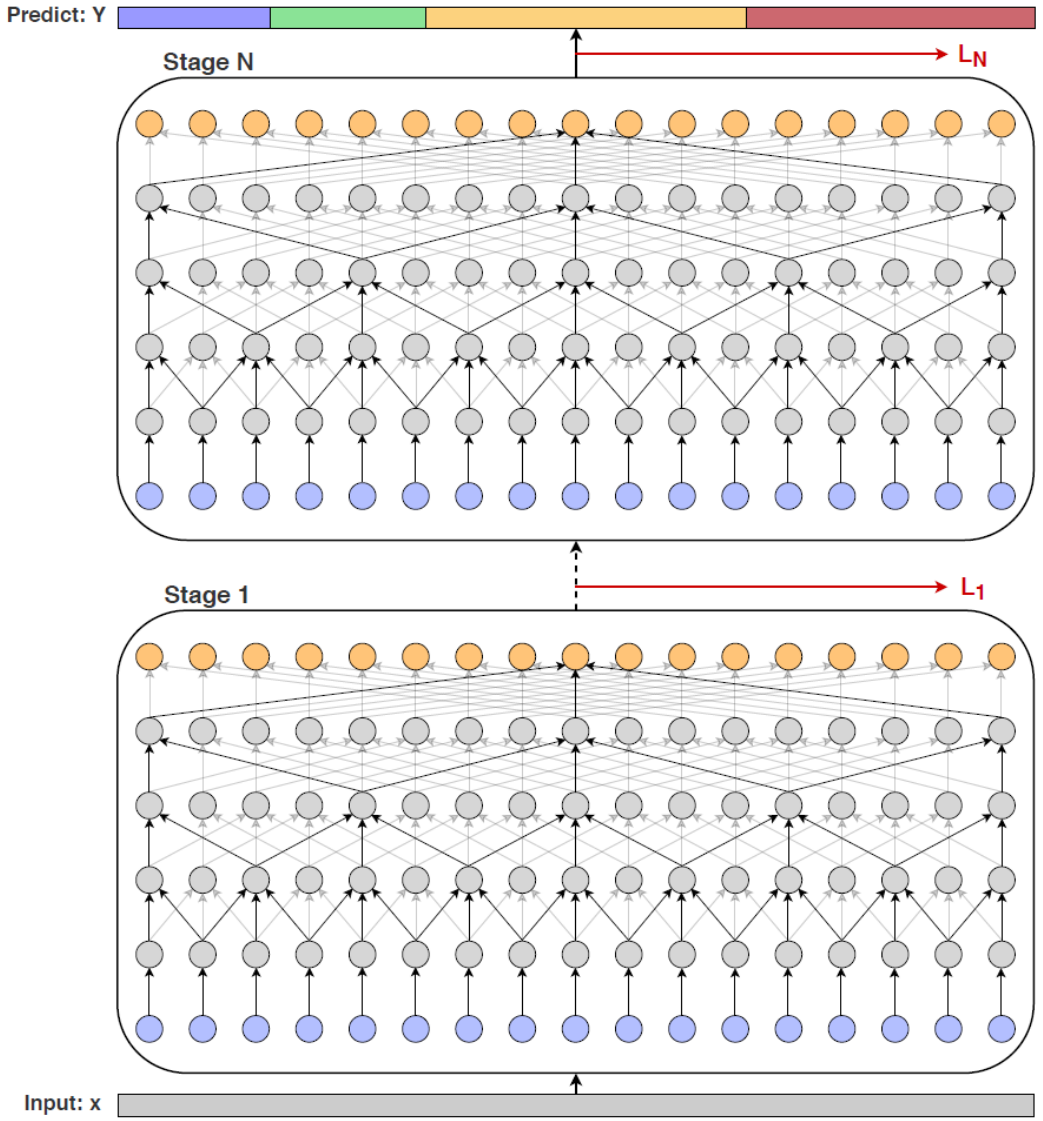
Loss: $\mathcal{L}_{cls} = \frac{1}{T}\sum_{t} -log(y_{t,c})$

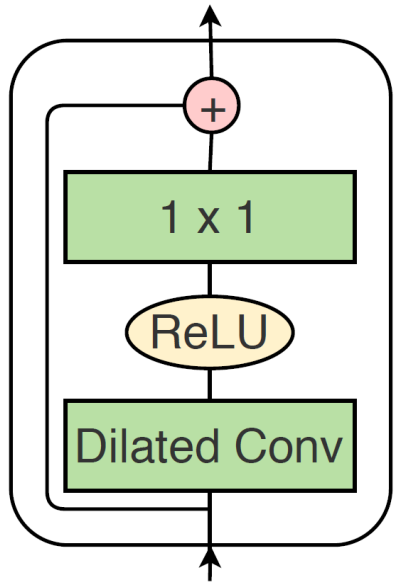




[Y. Abu Farha and J. Gall. **MS-TCN: Multi-Stage Temporal Convolutional Network for Action Segmentation.** CVPR 2019]

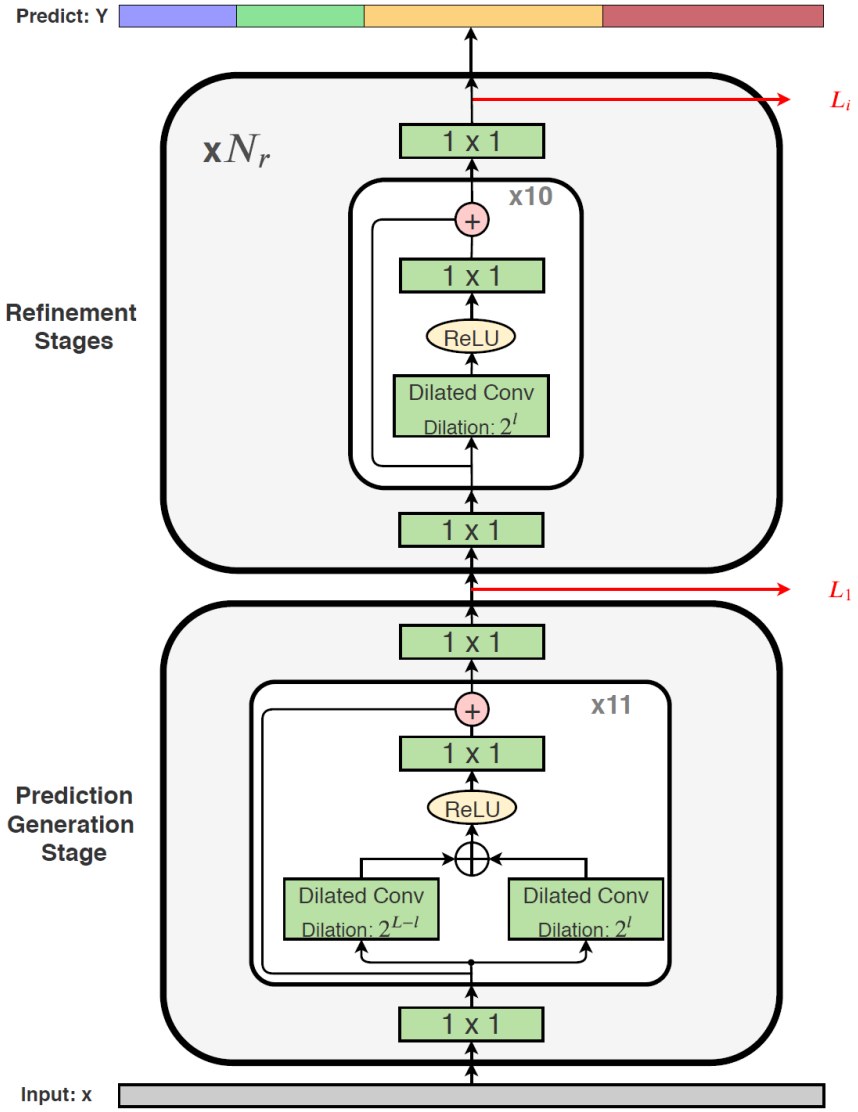
Multi-Stage Temporal Convolutional Network

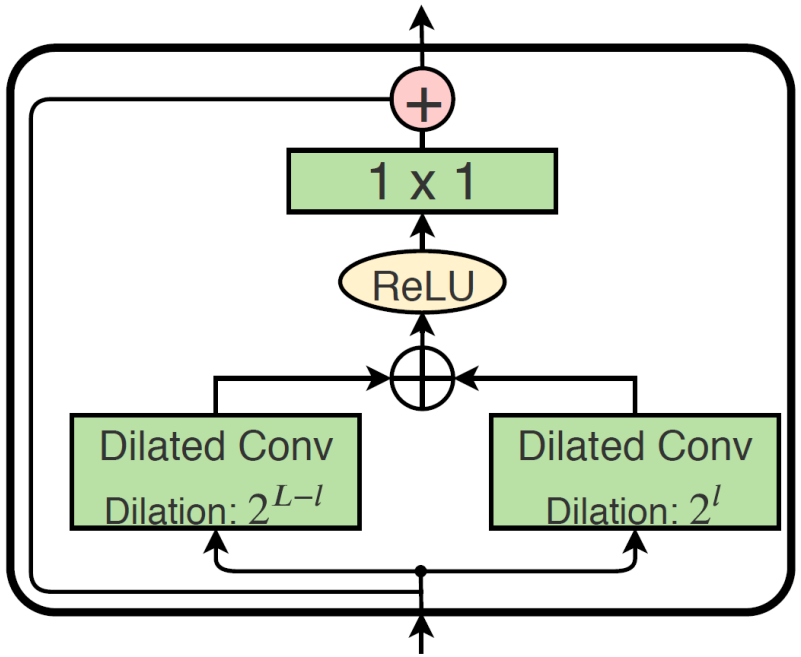




[Y. Abu Farha and J. Gall. **MS-TCN: Multi-Stage Temporal Convolutional Network for Action Segmentation.** CVPR 2019]

MS-TCN++



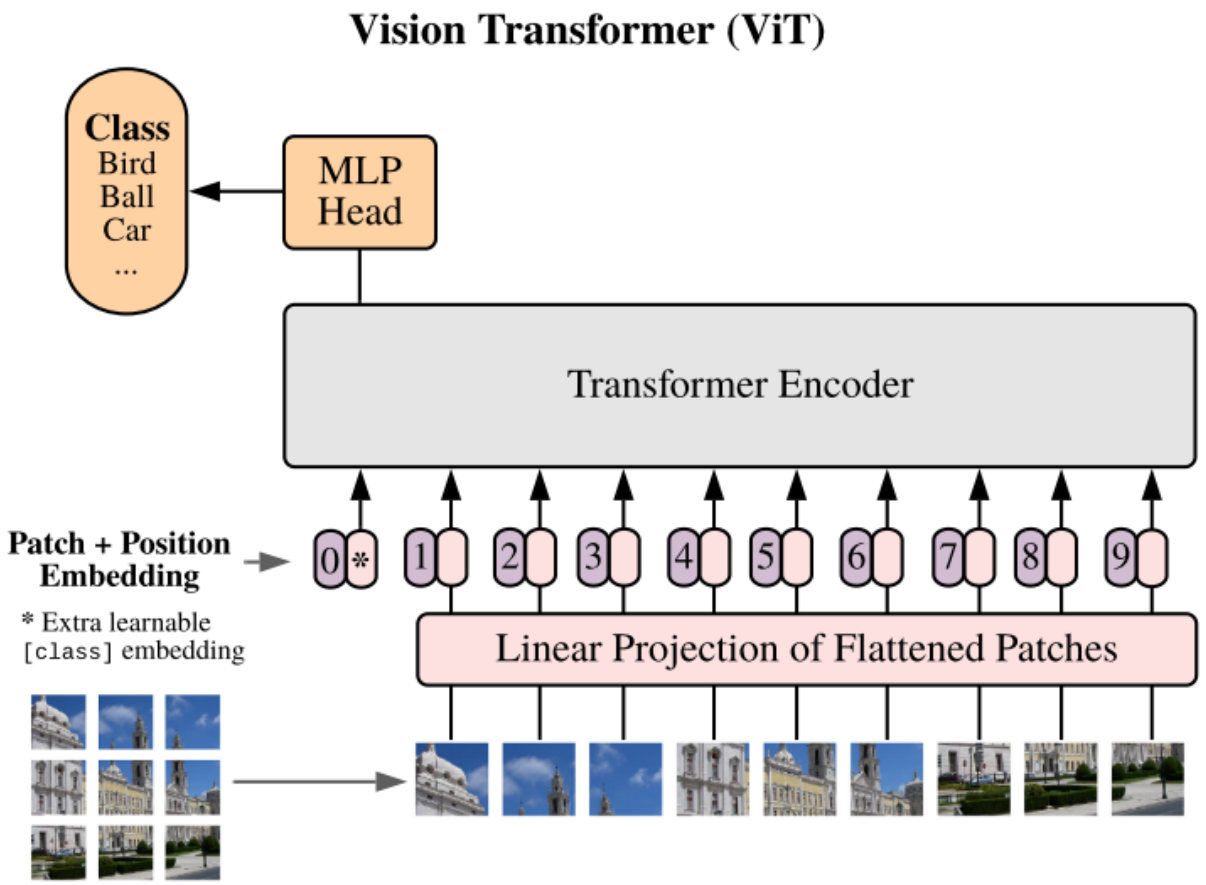


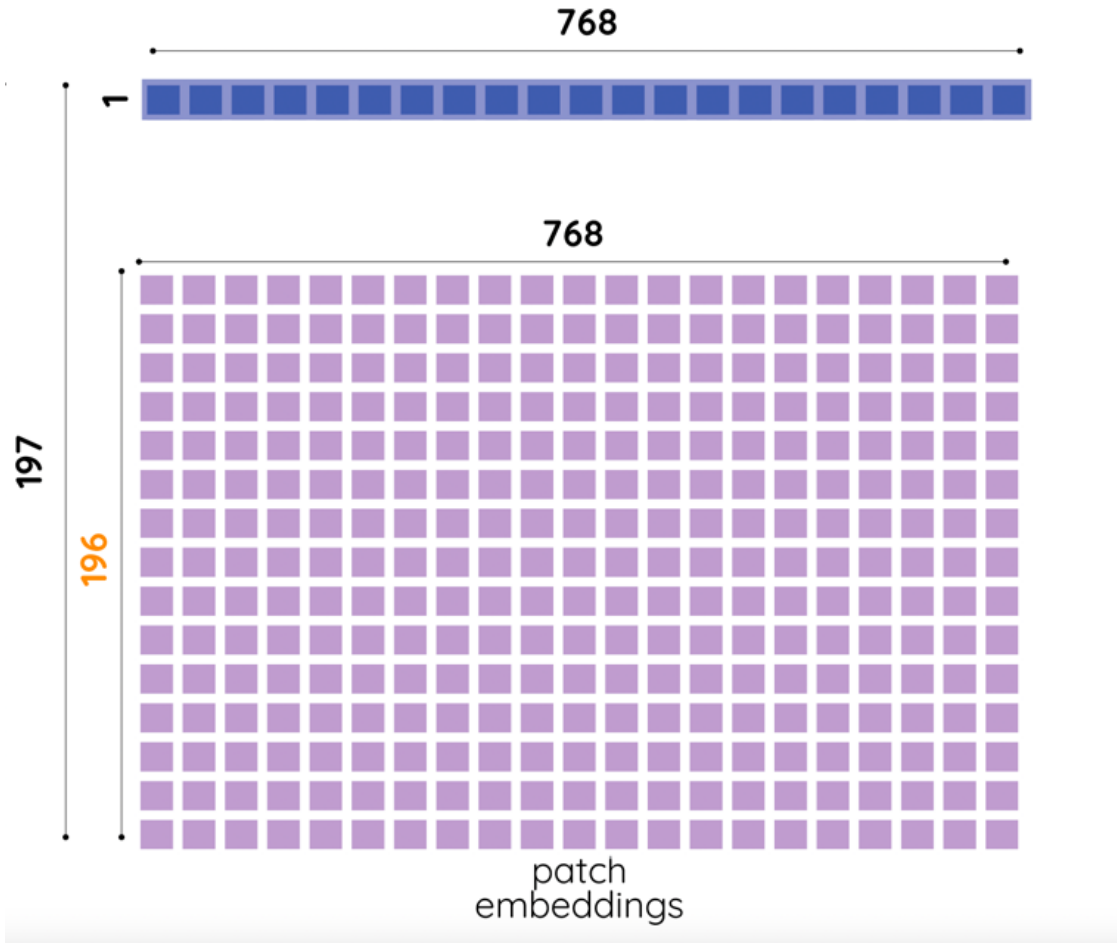
[S. Li et al. **MS-TCN++: Multi-Stage Temporal Convolutional Network for Action Segmentation.** TPAMI 2020]

MS-TCN++ vs. MS-TCN

[S. Li et al. **MS-TCN++: Multi-Stage Temporal Convolutional Network for Action Segmentation.** TPAMI 2020]

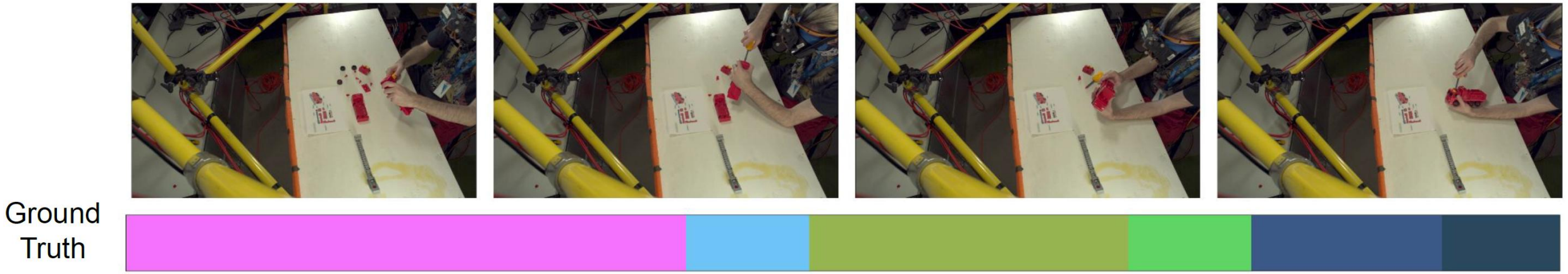
Vision Transformers for Images or Videos





[A. Dosovitskiy et al. **An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale.** ICLR 2021]

Vision Transformers for Action Segmentation?



[E. Bahrami et al. **How Much Temporal Long-Term Context is Needed for Action Segmentation?** ICCV 2023]

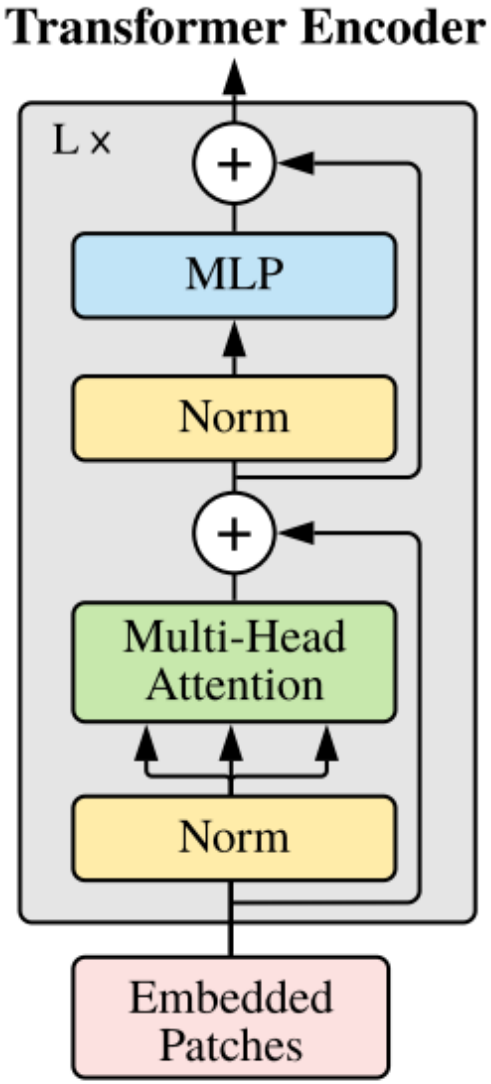
Transformer Block

- Self-Attention:

$$\mathcal{A} = \mathrm{Softmax}\left(\mathcal{Q}\mathcal{K}^T/\sqrt{d}\right)$$

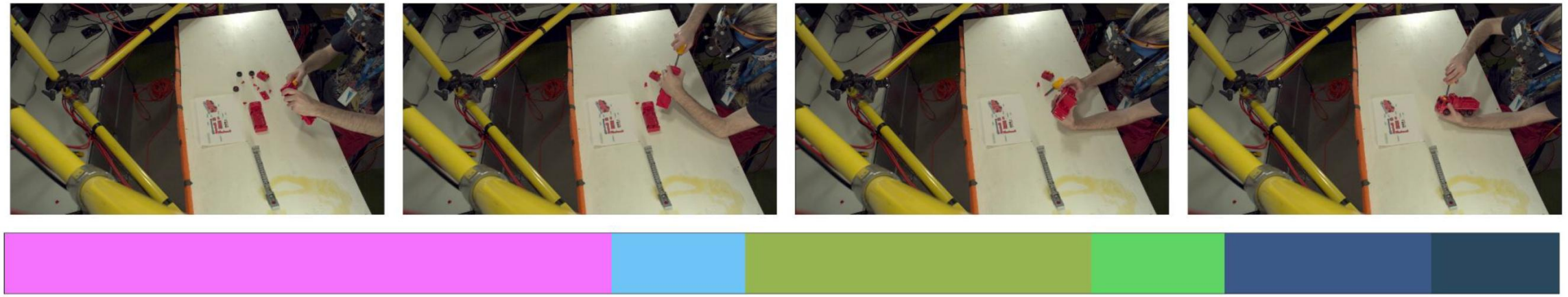
$$\mathcal{A} \in \mathbb{R}^{(N+1)\times(N+1)}$$

$$\mathcal{O} = \mathcal{A}\mathcal{V}$$



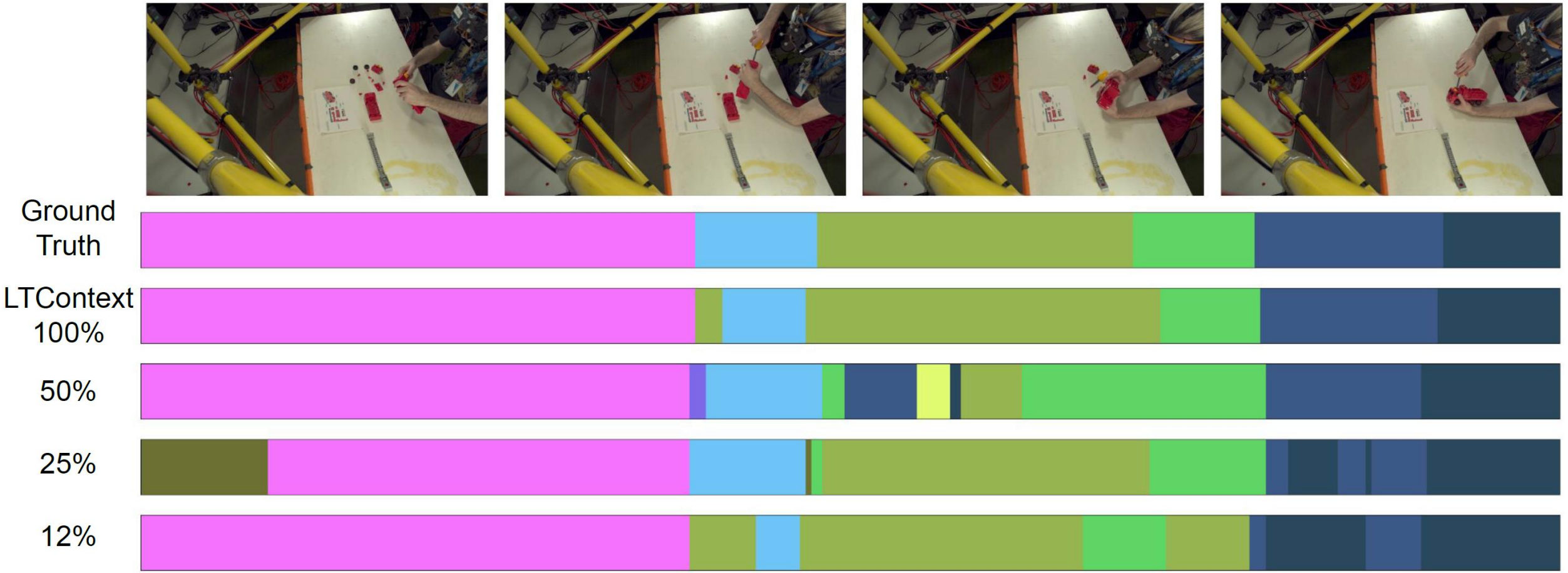
[A. Vaswani et al. **Attention Is All You Need.** NeurIPS 2017]

Vision Transformers for Action Segmentation



[E. Bahrami et al. **How Much Temporal Long-Term Context is Needed for Action Segmentation?** ICCV 2023]

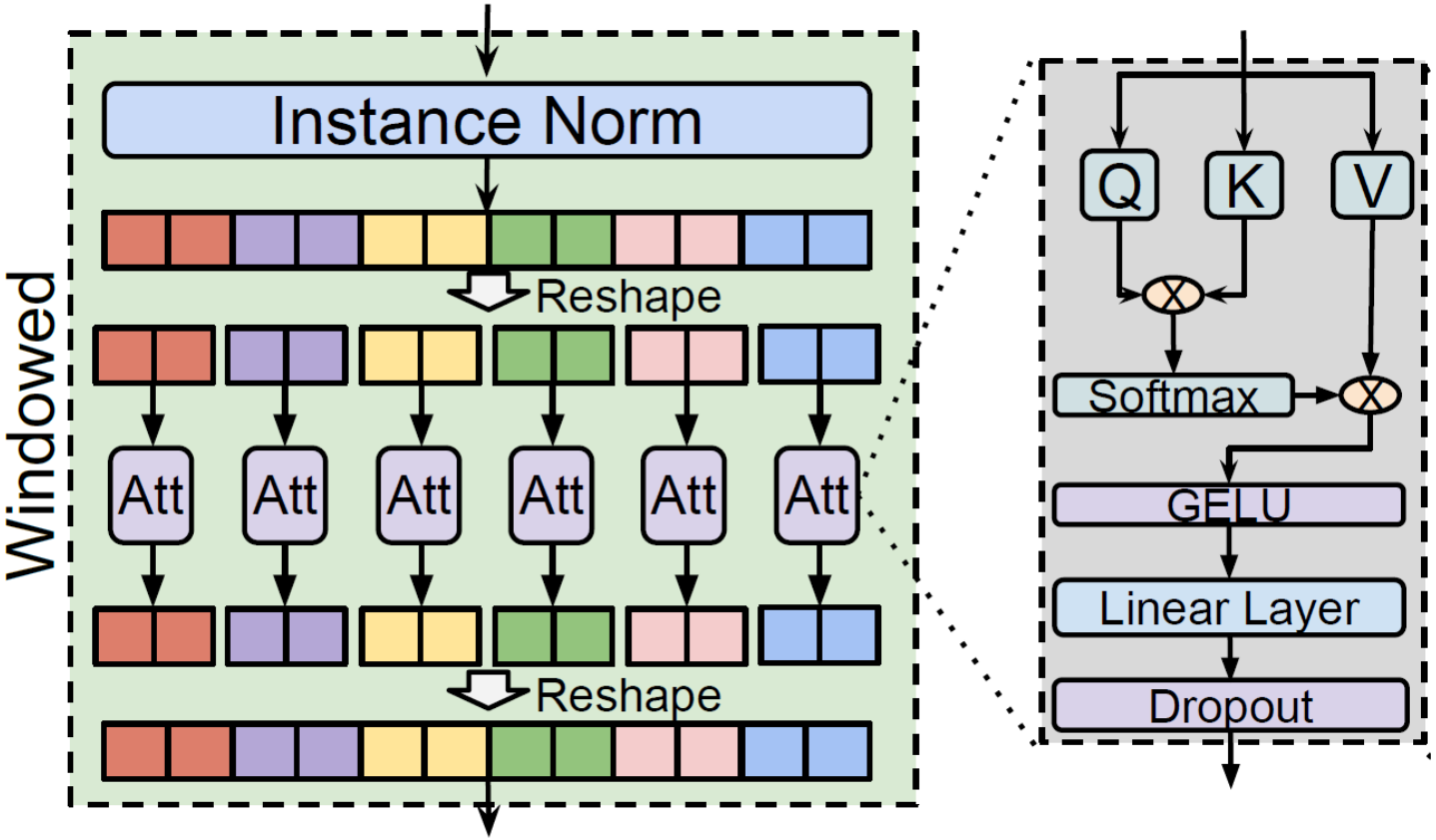
Vision Transformers for Action Segmentation



[E. Bahrami et al. **How Much Temporal Long-Term Context is Needed for Action Segmentation?** ICCV 2023]

Vision Transformers for Action Segmentation

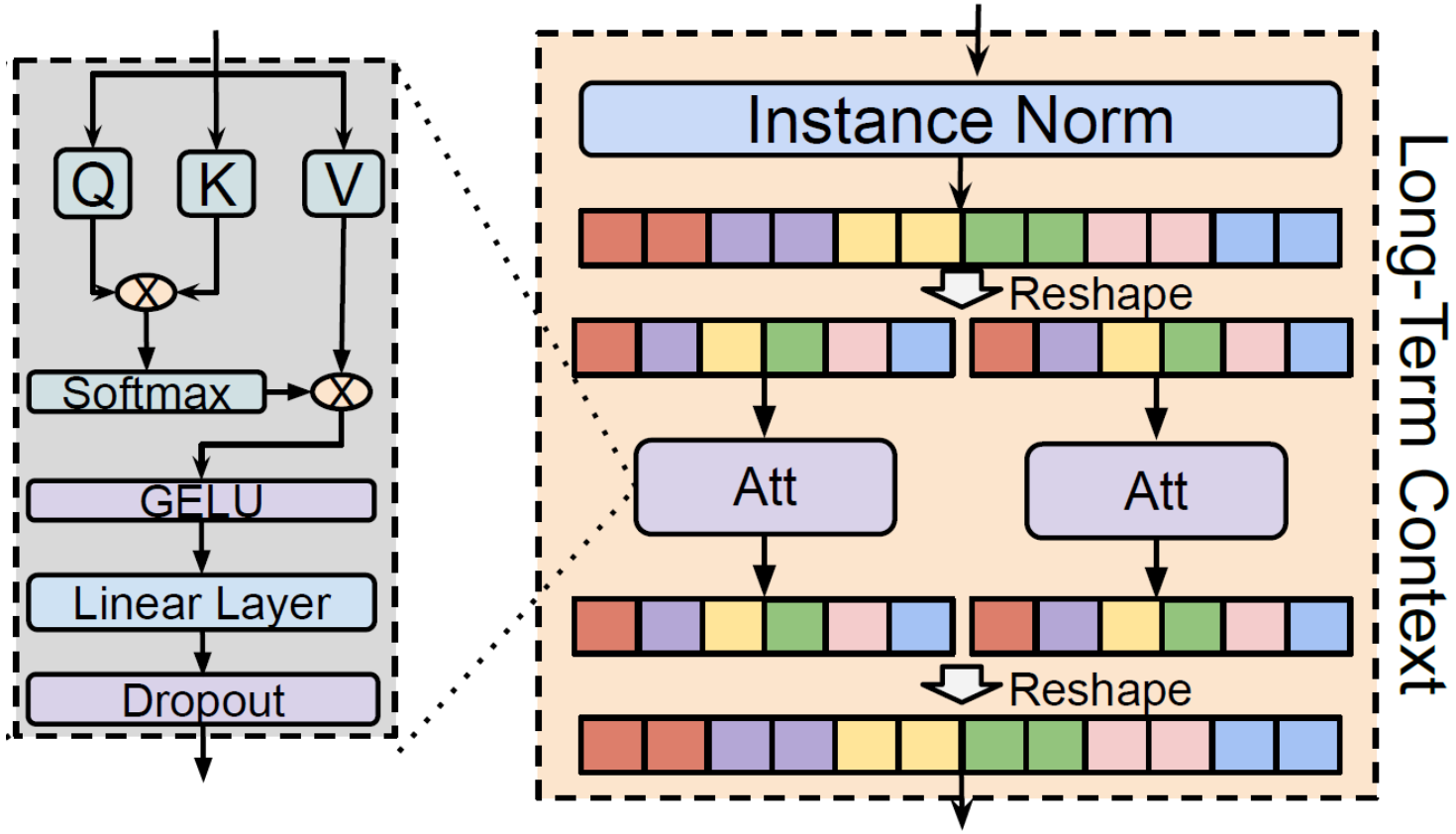
Attention over local window



[E. Bahrami et al. **How Much Temporal Long-Term Context is Needed for Action Segmentation?** ICCV 2023]

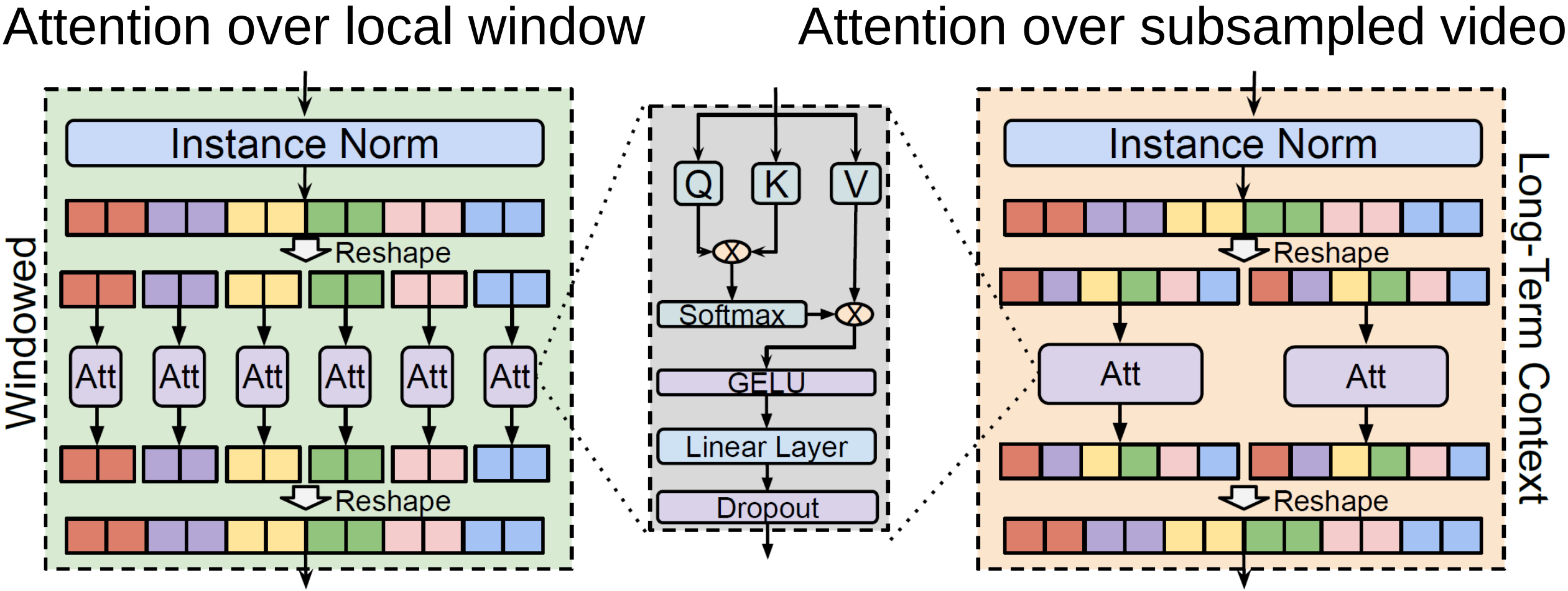
Vision Transformers for Action Segmentation

Attention over subsampled video



[E. Bahrami et al. **How Much Temporal Long-Term Context is Needed for Action Segmentation?** ICCV 2023]

Vision Transformers for Action Segmentation

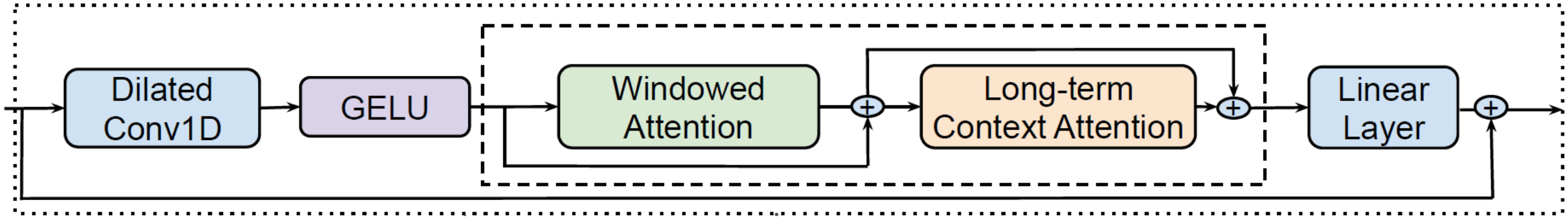


[E. Bahrami et al. **How Much Temporal Long-Term Context is Needed for Action Segmentation?** ICCV 2023]

Vision Transformers for Action Segmentation

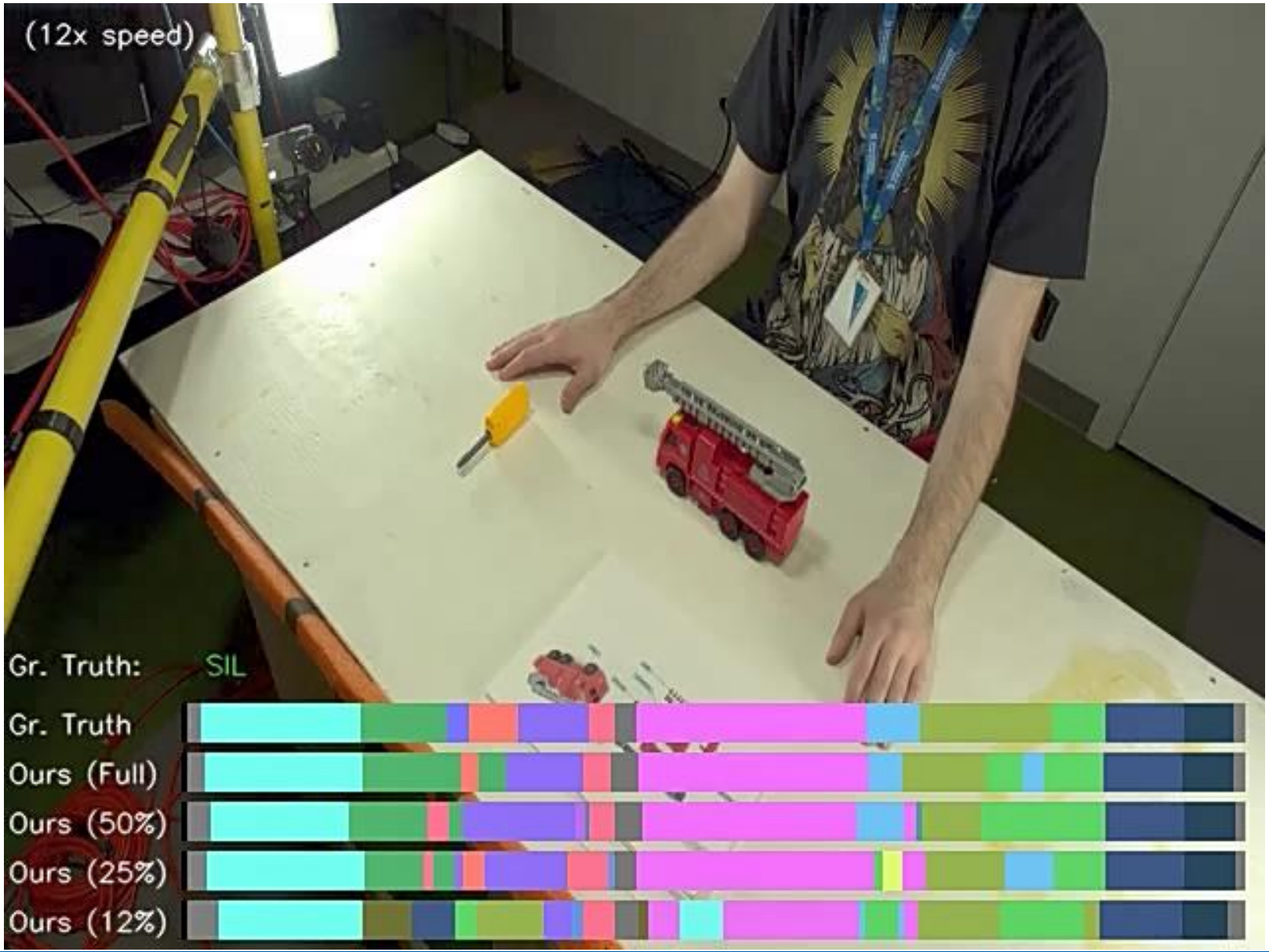
Attention over local window

Attention over subsampled video



[E. Bahrami et al. **How Much Temporal Long-Term Context is Needed for Action Segmentation?** ICCV 2023]

Vision Transformers for Action Segmentation



[E. Bahrami et al. **How Much Temporal Long-Term Context is Needed for Action Segmentation?** ICCV 2023]

Research Unit - Anticipating Human Behavior

BONN

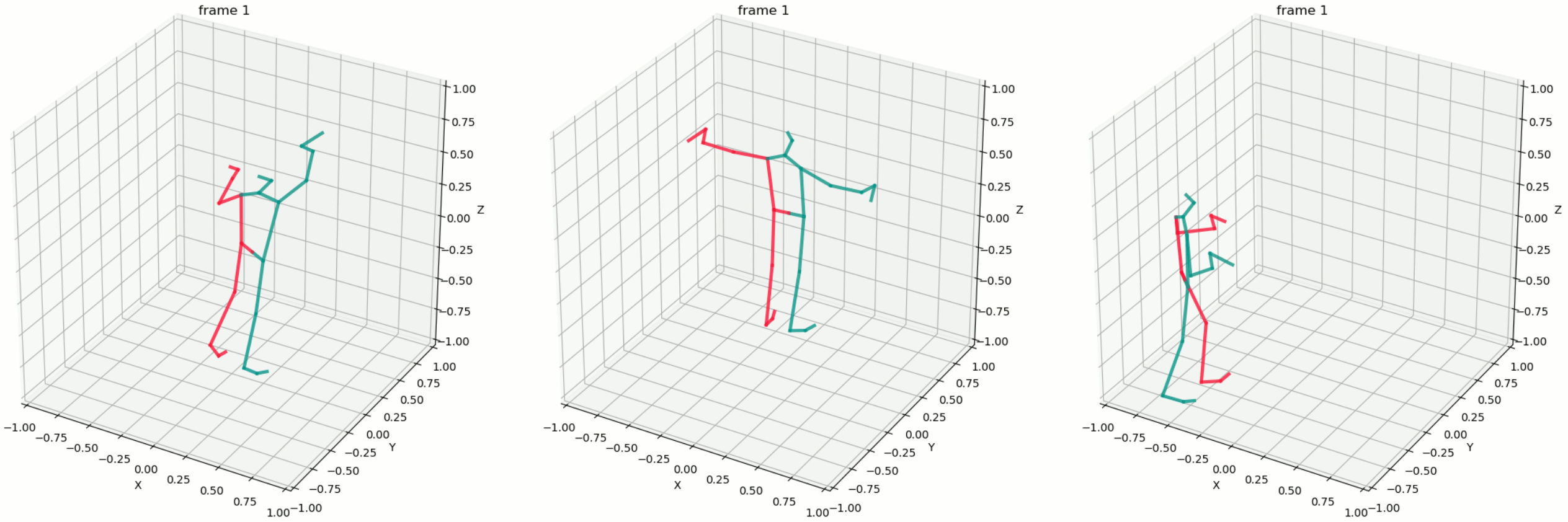
Human Anticipation for Human-Robot Collaboration

Anticipating Behavior

Past → Anticipate → Future

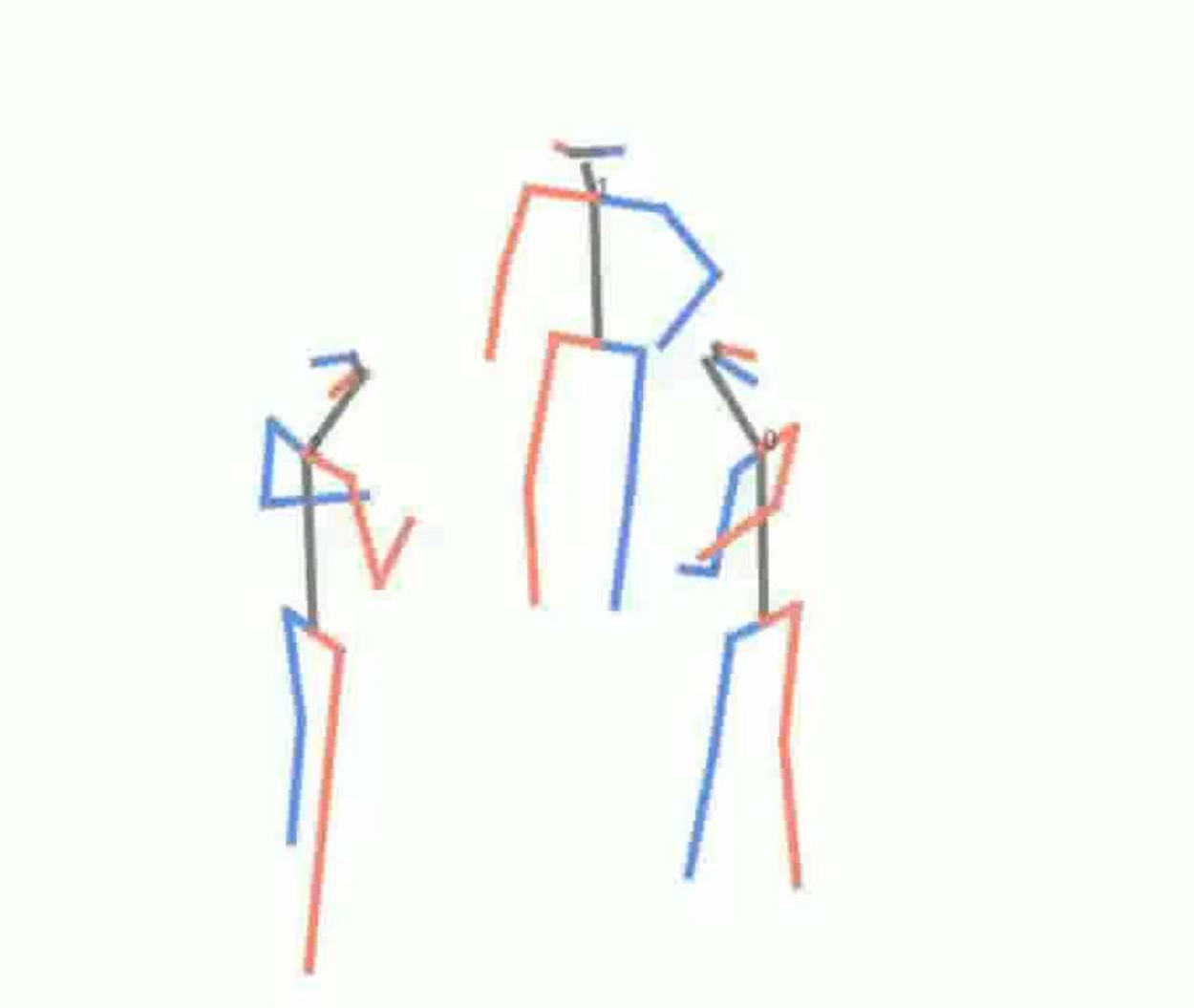
[Y. Abu Farha et al. **When will you do what? Anticipating Temporal Occurrences of Activities.** CVPR 2018]

Human Motion Anticipation



[J. Tanke et al. **Intention-based Long-Term Human Motion Anticipation.** 3DV 2021]

Multiple Human Motion Anticipation

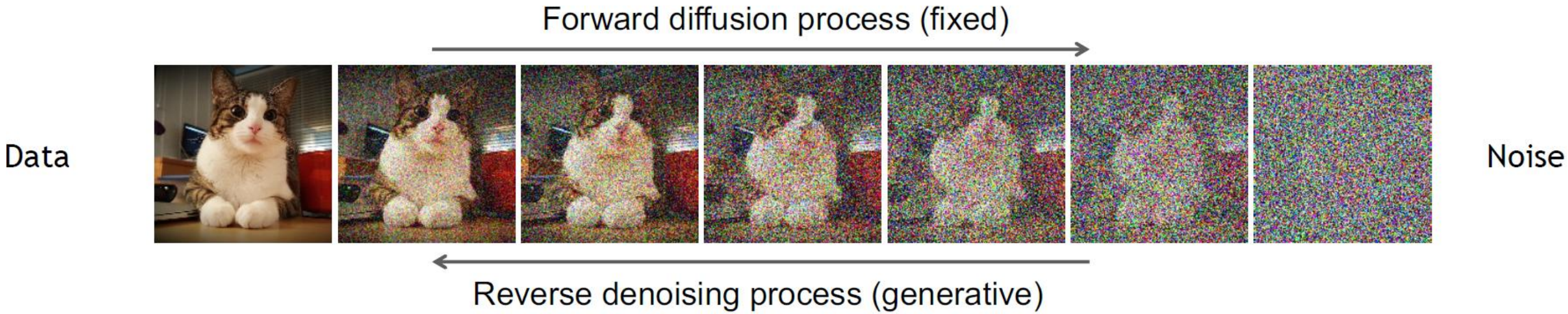


[J. Tanke et al. **Social Diffusion: Long-term Multiple Human Motion Anticipation.** ICCV 2023]

Diffusion Model

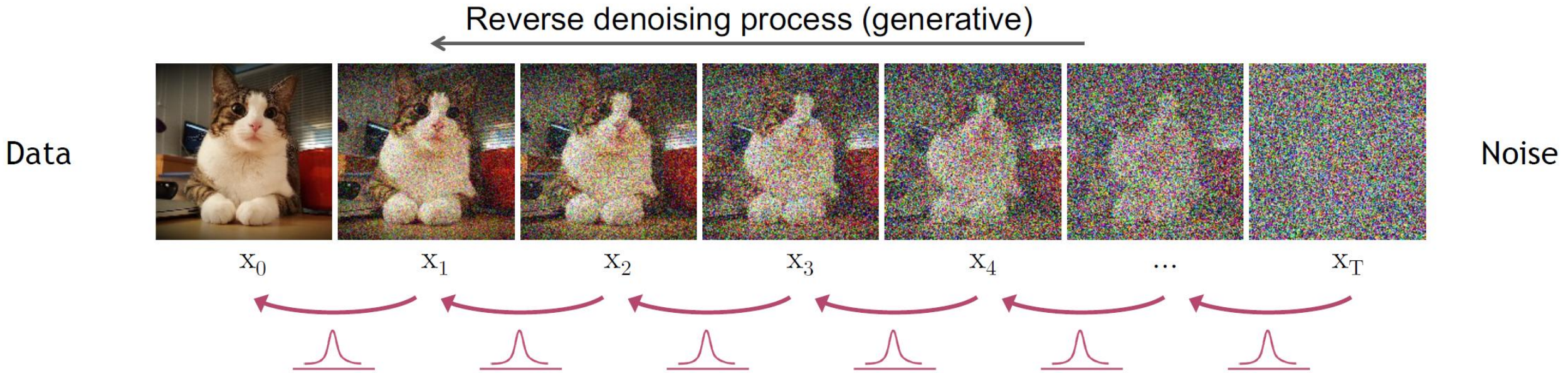
Diffusion models consist of two processes:

- Forward diffusion process that gradually adds noise to input
- Reverse denoising process that learns to generate data by denoising



Generative Process

Reverse denoising process



$$\hat{\mathbf{X}}_{0,t-1} = g(q(\hat{\mathbf{X}}_{t-1}|\hat{\mathbf{X}}_{0,t}), t-1)$$

g is a network

Generative Process

Reverse denoising process

$$\hat{\mathbf{X}}_{0,t-1} = g(q(\hat{\mathbf{X}}_{t-1}|\hat{\mathbf{X}}_{0,t}), t-1)$$

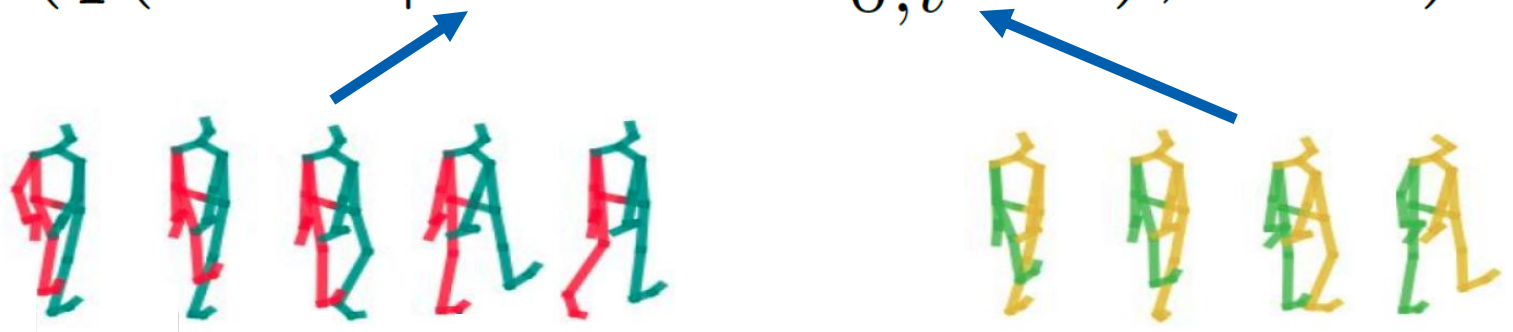
Generative Process

Reverse denoising process

$$\hat{\mathbf{X}}_{0,t-1} = g(q(\hat{\mathbf{X}}_{t-1}|\hat{\mathbf{X}}_{0,t}), t-1)$$

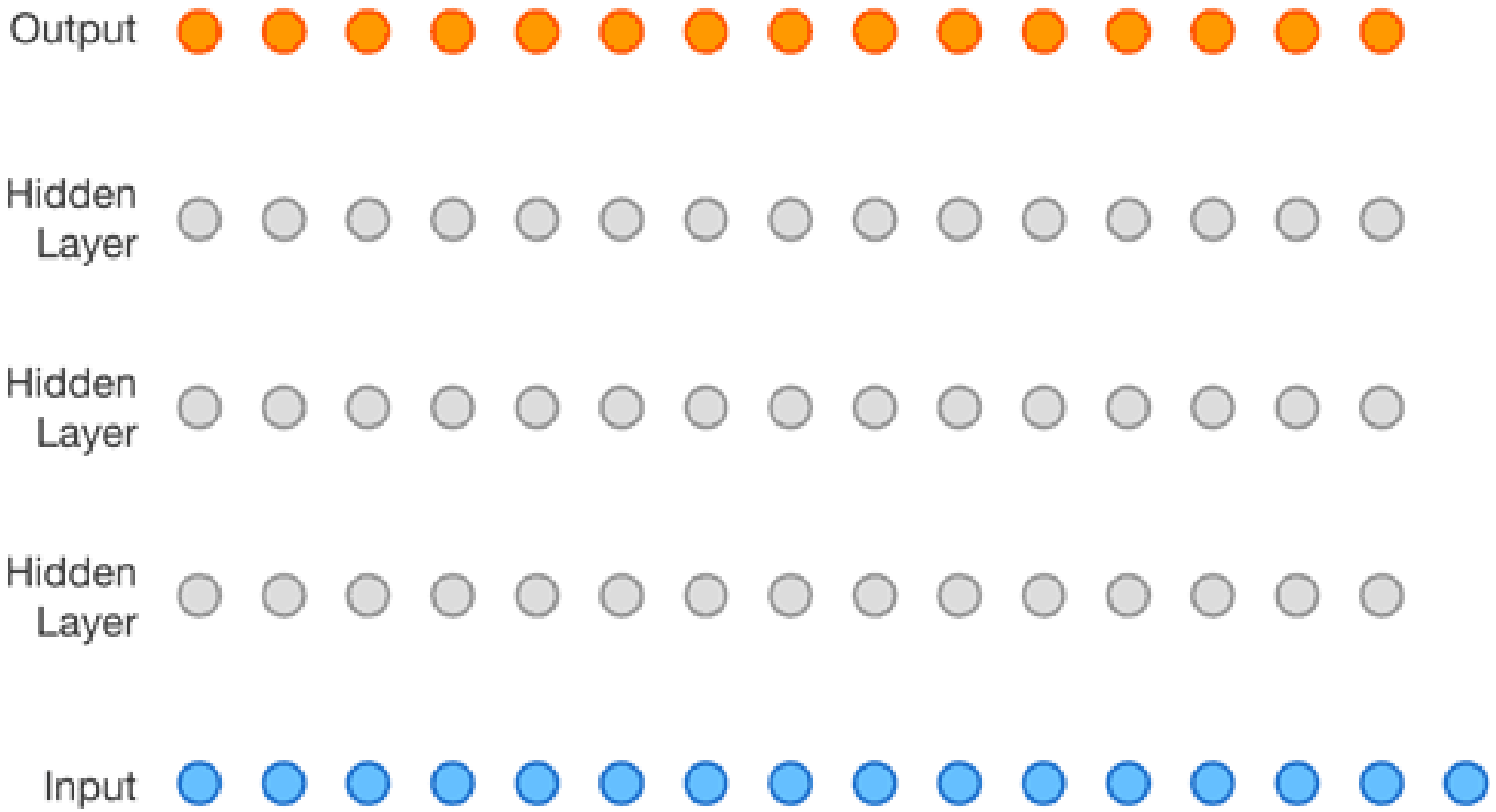
Condition denoising process on observation

$$\hat{\mathbf{X}}_{0,t-1} = g(q(\hat{\mathbf{X}}_{t-1}|\mathbf{X}^{1:n} \cup \hat{\mathbf{X}}_{0,t}^{n+1:N}), t-1)$$



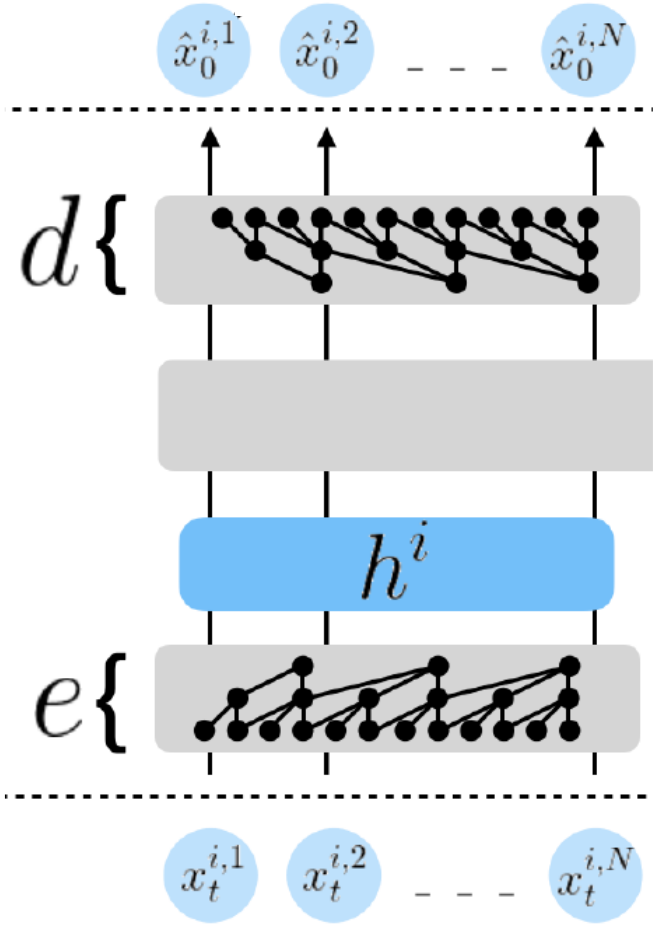
Generator – Single Person

Temporal convolutional network

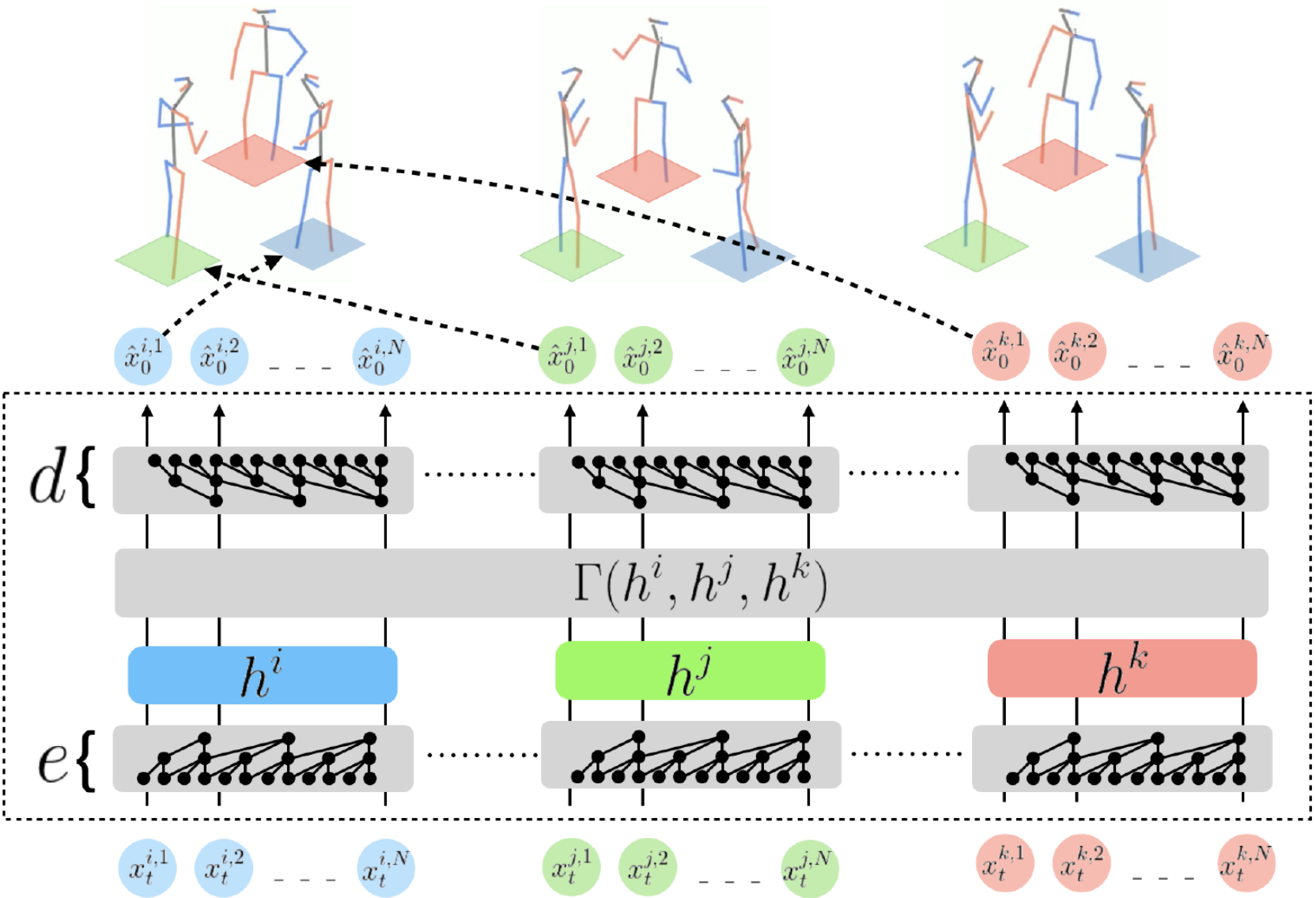


Generator – Single Person

Temporal convolutional network for encoder e and decoder d



Generator – Multiple Persons



Couple persons by aggregation function:

$$\Gamma_{\mathbb{E}}(\mathbf{H}) = \frac{1}{p}\sum_{i=1}^{p}\mathbf{h}^i$$

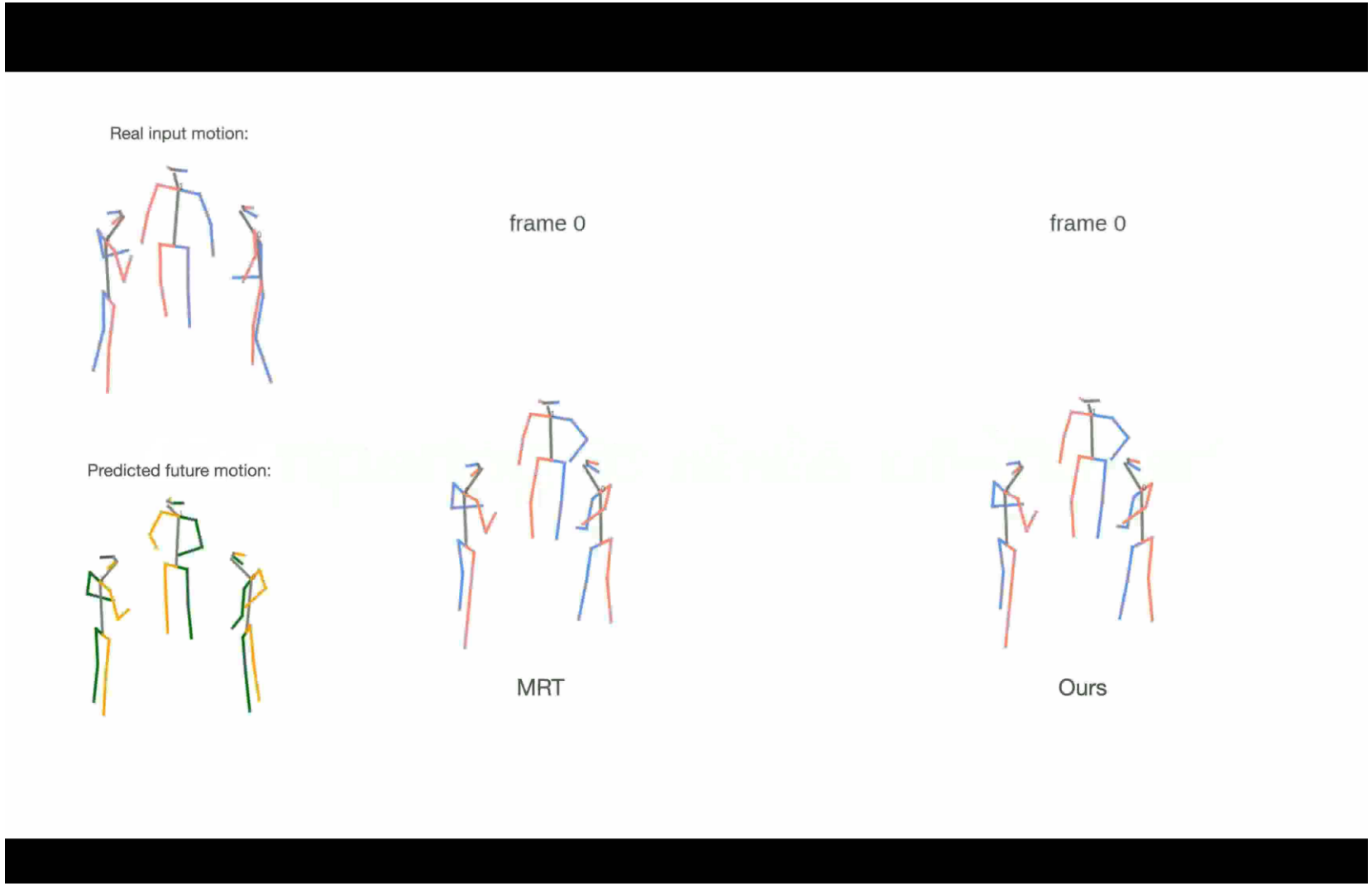
Haggling Dataset

Haggling Dataset

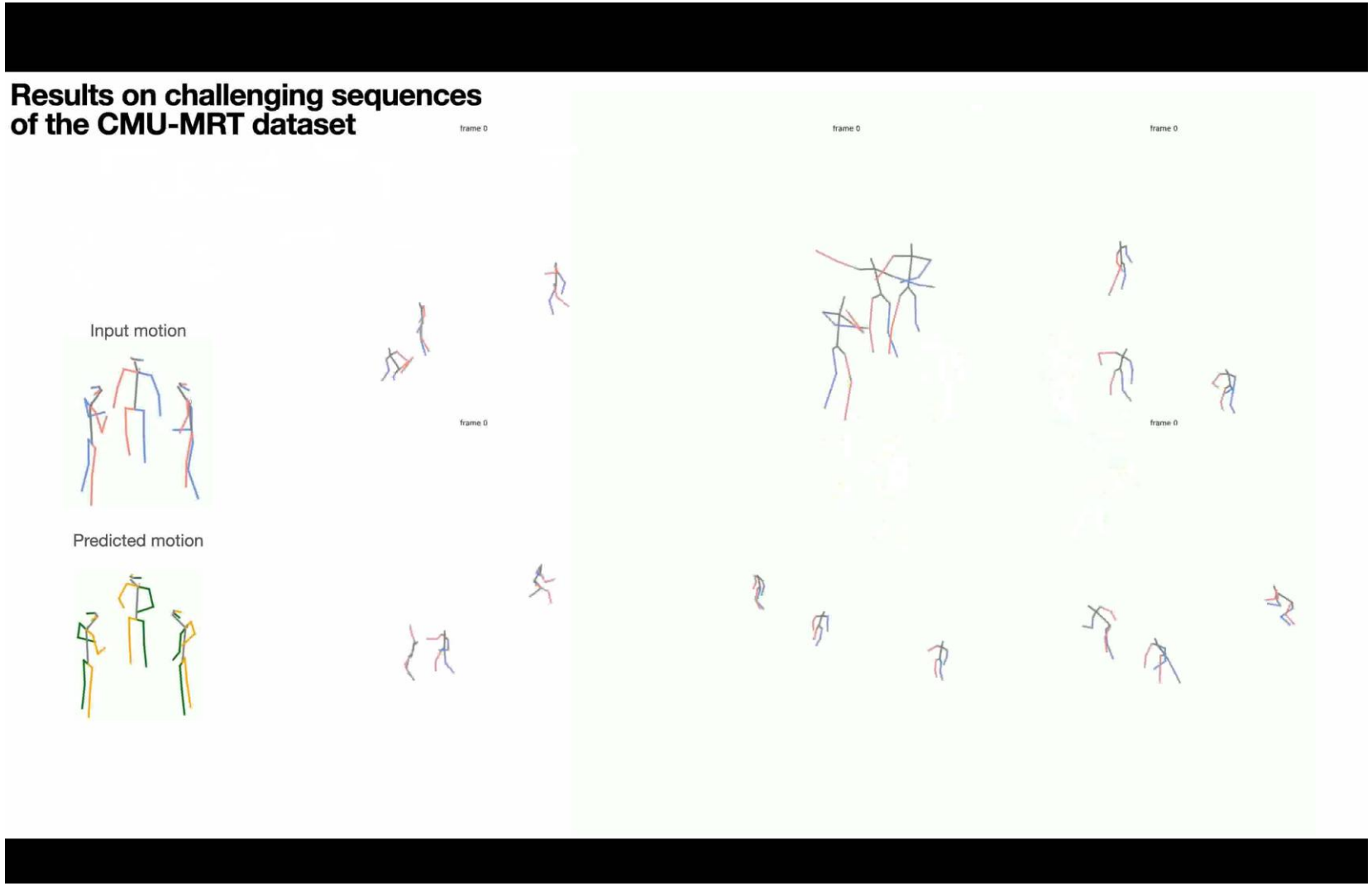
- Triadic human interaction between two sellers and one buyer

Towards Social Artificial Intelligence: Nonverbal Social Signal Prediction in Triadic Interaction, CVPR2019

Long-term Multiple Human Motion Anticipation

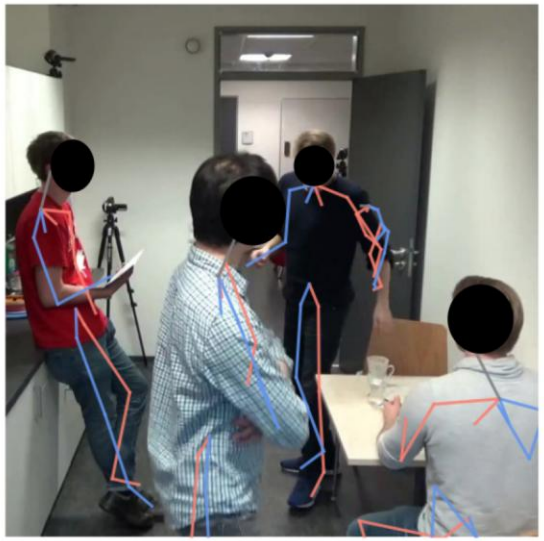


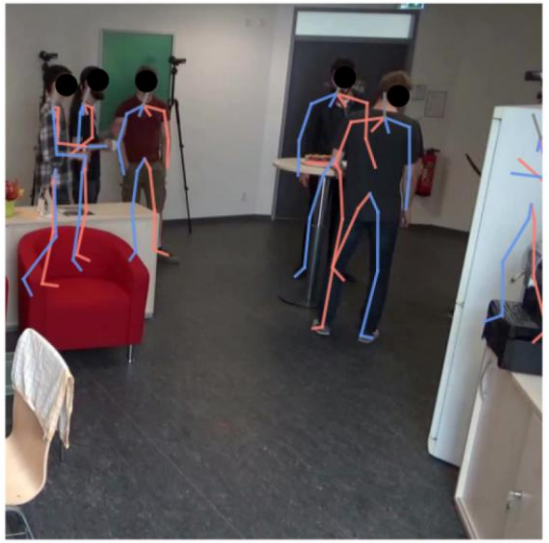
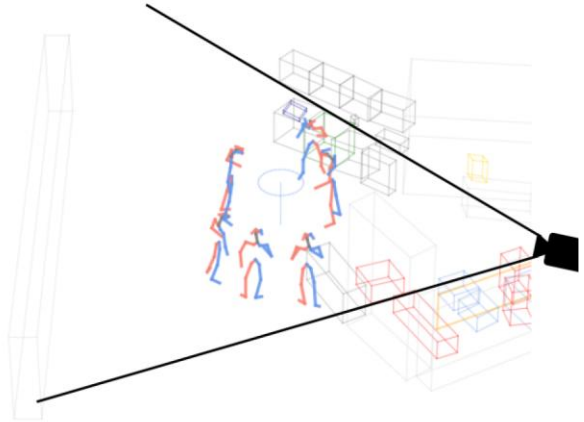
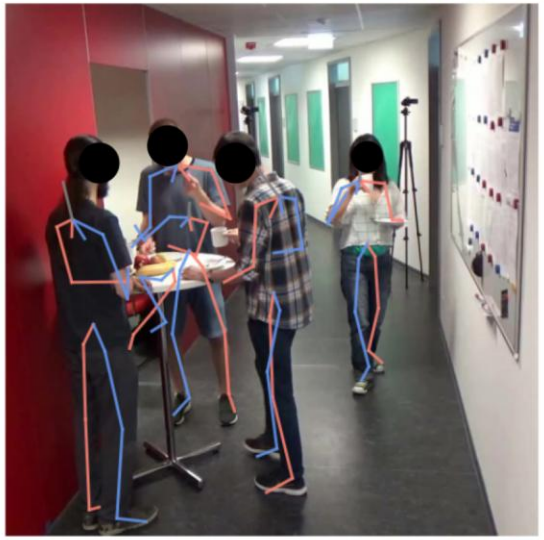
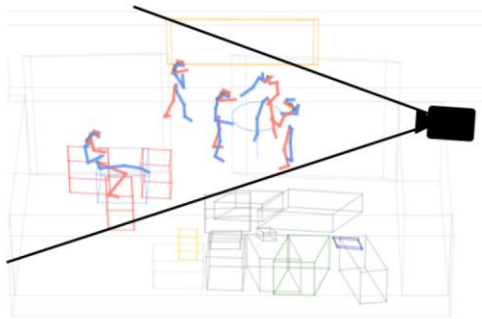
Long-term Multiple Human Motion Anticipation

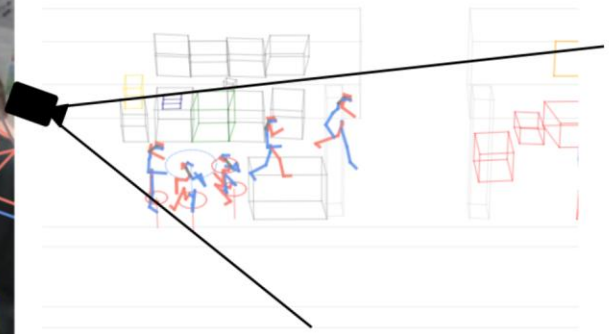


Humans in Kitchens: A Dataset for Multi-Person Human Motion Forecasting

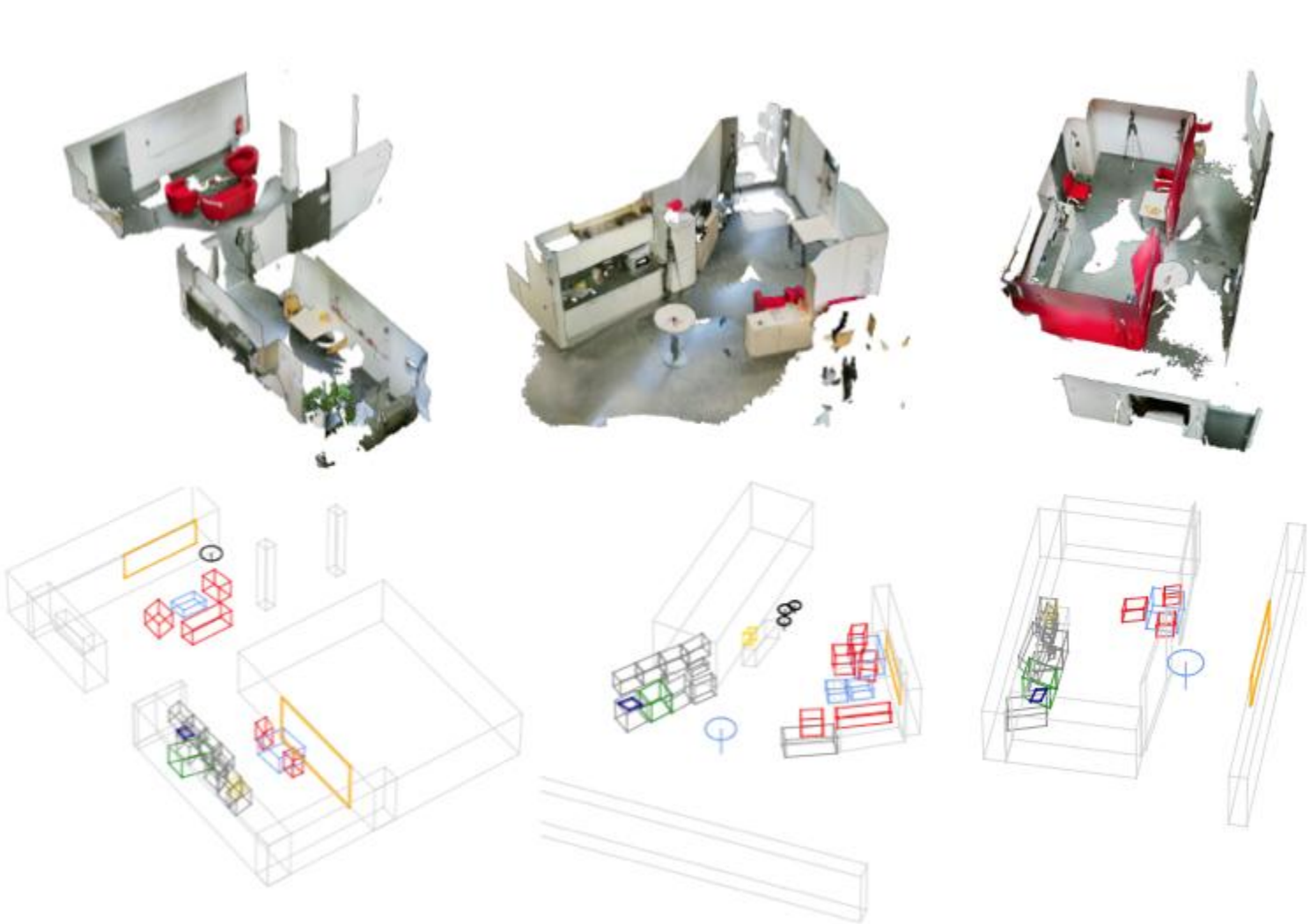
BONN

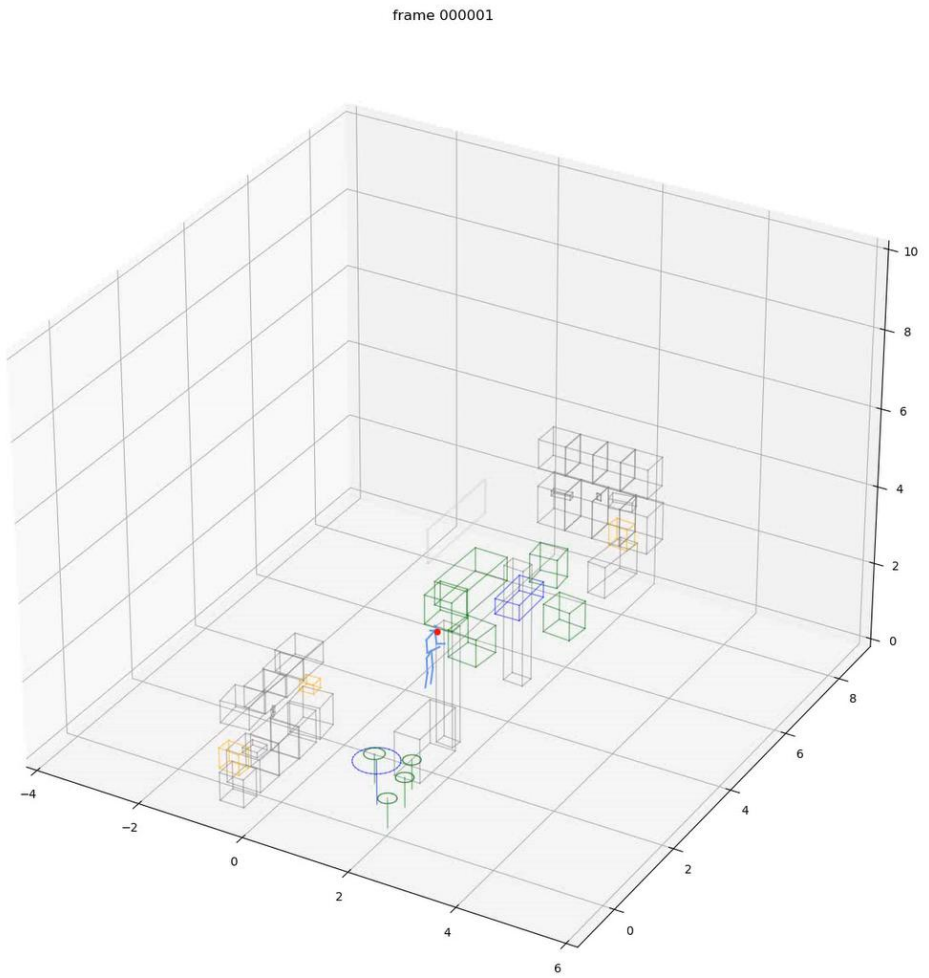




Dataset for Multi-Person Human Motion Forecasting



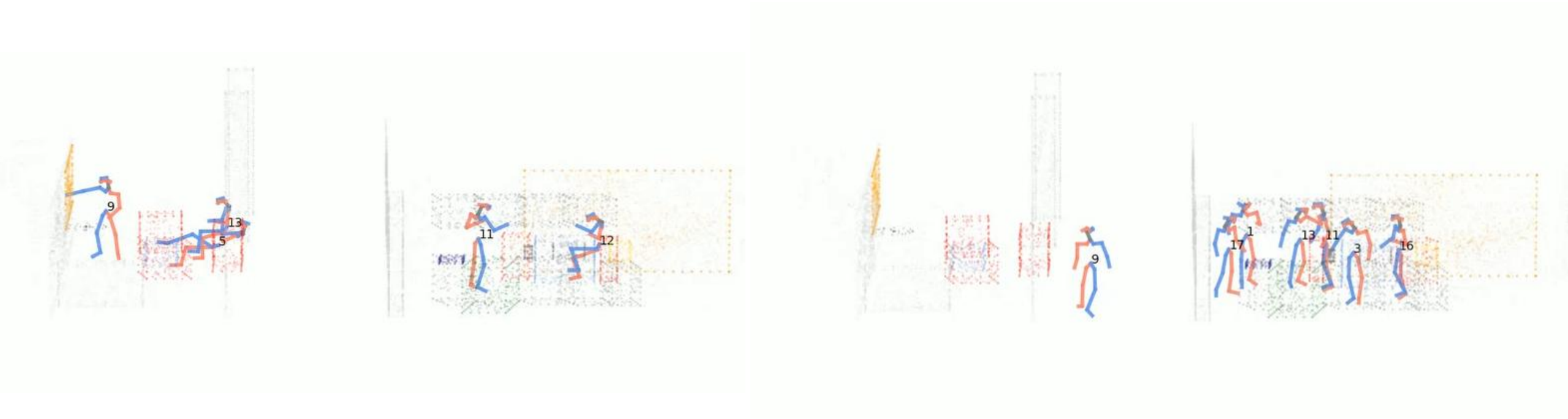


Dataset for Multi-Person Human Motion Forecasting

	A	B	C	D
# frames	128,959 (1.43h)	179,097 (1.99h)	175,392 (1.95h)	176,264 (1.96h)
# annotated poses	573,253	1,132,422	908,380	1,415,189
mean persons / frame	4.42	6.32	5.17	7.97
median persons / frame	4	6	5	7
max. persons / frame	9	14	9	16
# individuals	18	32	16	24
surface area	76.35m^2	57.22m^2	38.28m^2	80.40m^2
# camera views	11	11	12	12
# scene objects	37	40	29	50

[J. Tanke et al. **Humans in Kitchens: A Dataset for Multi-Person Human Motion Forecasting with Scene Context.** NeurIPS Datasets and Benchmarks Track 2023]

Dataset for Multi-Person Human Motion Forecasting



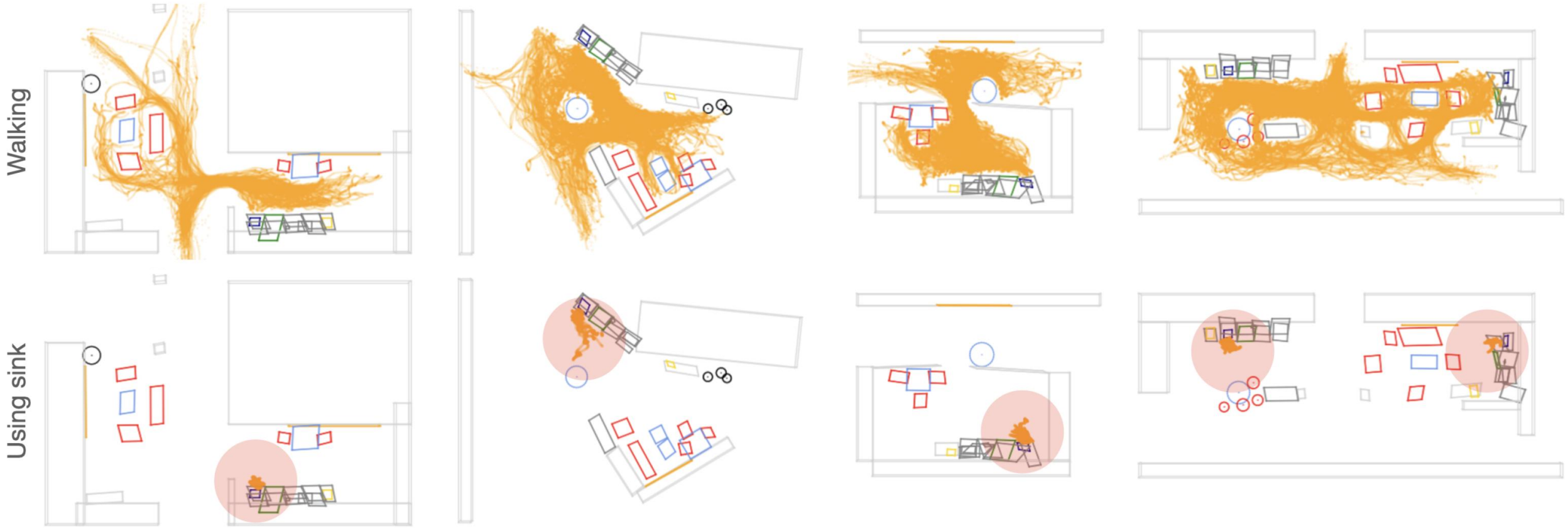
[J. Tanke et al. **Humans in Kitchens: A Dataset for Multi-Person Human Motion Forecasting with Scene Context.** NeurIPS Datasets and Benchmarks Track 2023]

Dataset for Multi-Person Human Motion Forecasting

[J. Tanke et al. **Humans in Kitchens: A Dataset for Multi-Person Human Motion Forecasting with Scene Context.** NeurIPS Datasets and Benchmarks Track 2023]

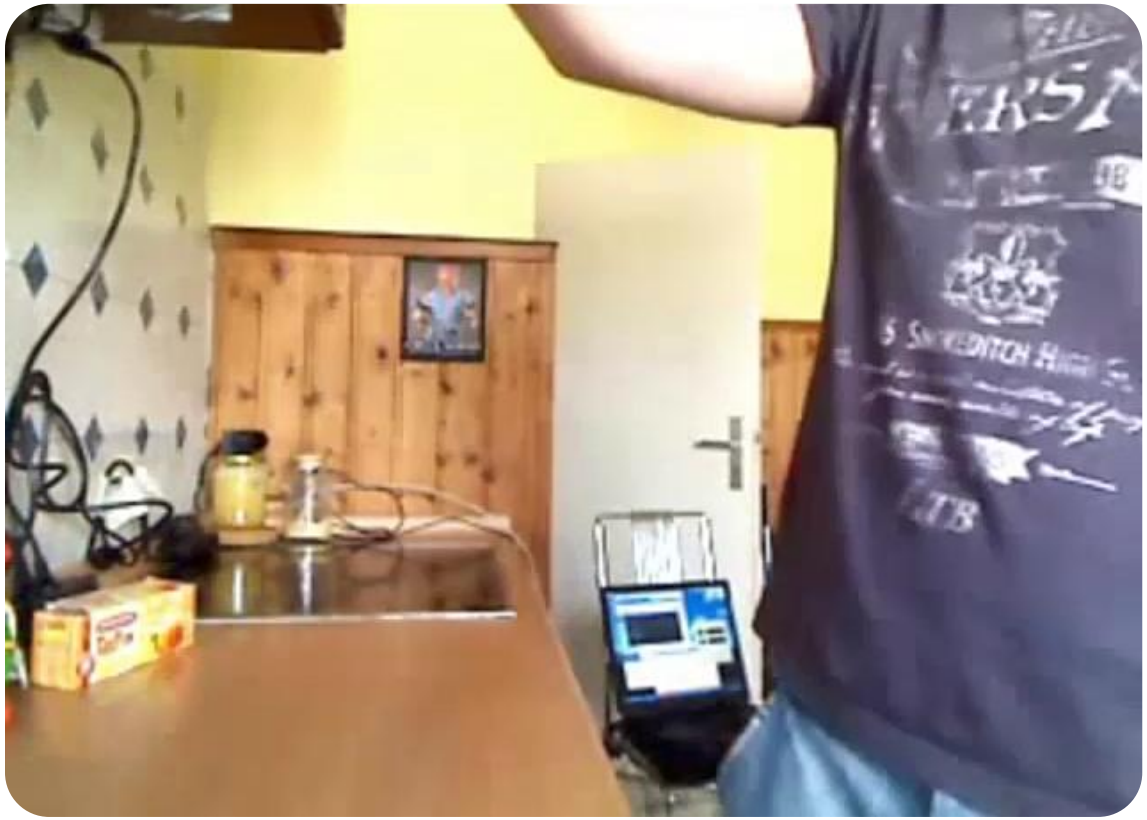
Dataset for Multi-Person Human Motion Forecasting

Activities



Anticipating Behavior

Past → Anticipate → Future



[Y. Abu Farha et al. **When will you do what? Anticipating Temporal Occurrences of Activities.** CVPR 2018]

Gated Temporal Diffusion for Anticipation

Model uncertainty of the future…

[O. Zatsarynna et al. **Gated Temporal Diffusion for Stochastic Long-term Dense Anticipation.** ECCV 2024]

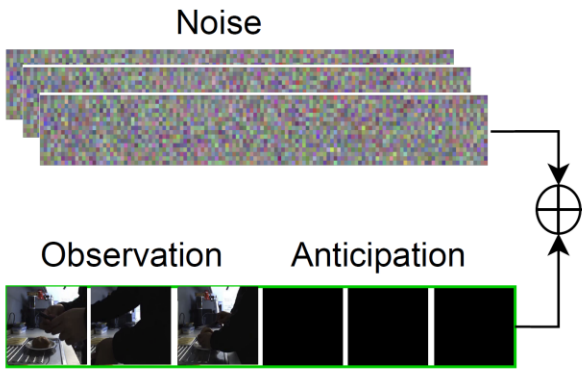
Gated Temporal Diffusion for Anticipation

Model not only uncertainty of the future…

… but also uncertainty of the observation!

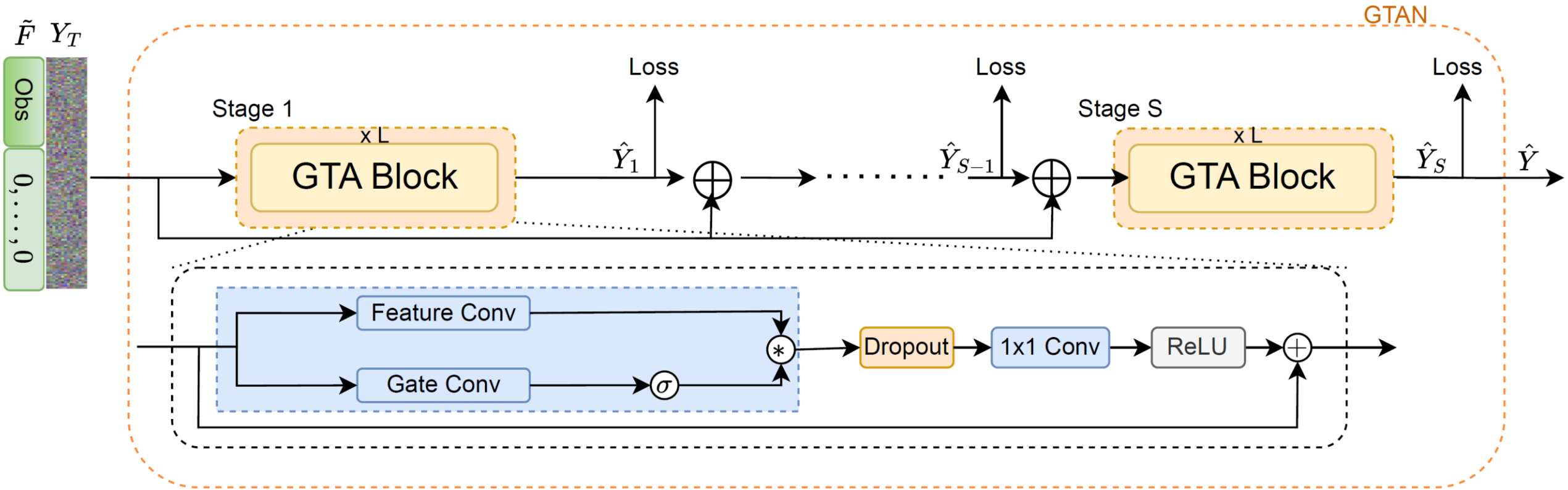
[O. Zatsarynna et al. **Gated Temporal Diffusion for Stochastic Long-term Dense Anticipation.** ECCV 2024]

Gated Temporal Diffusion for Anticipation



[O. Zatsarynna et al. **Gated Temporal Diffusion for Stochastic Long-term Dense Anticipation.** ECCV 2024]

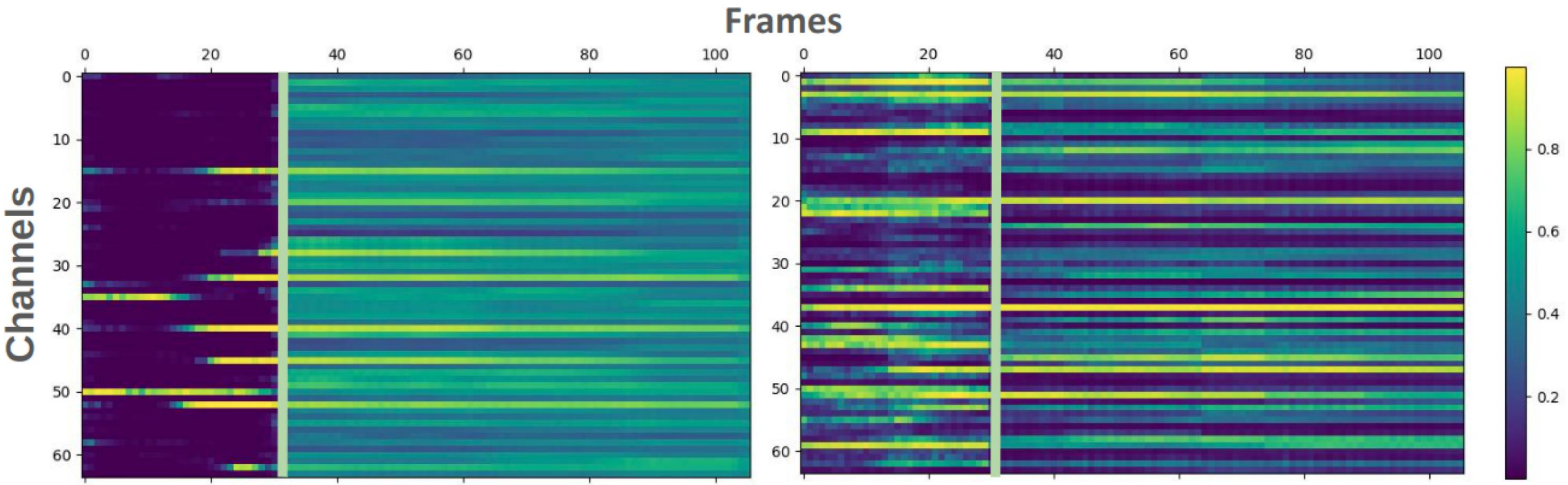
Gated Temporal Diffusion for Anticipation



$$\bar{H}_{s,l} = \sigma(W_g * \tilde{H}_{s,l-1} + b_g) \odot (W_f * \tilde{H}_{s,l-1} + b_f)$$

[O. Zatsarynna et al. **Gated Temporal Diffusion for Stochastic Long-term Dense Anticipation.** ECCV 2024]

Gated Temporal Diffusion for Anticipation



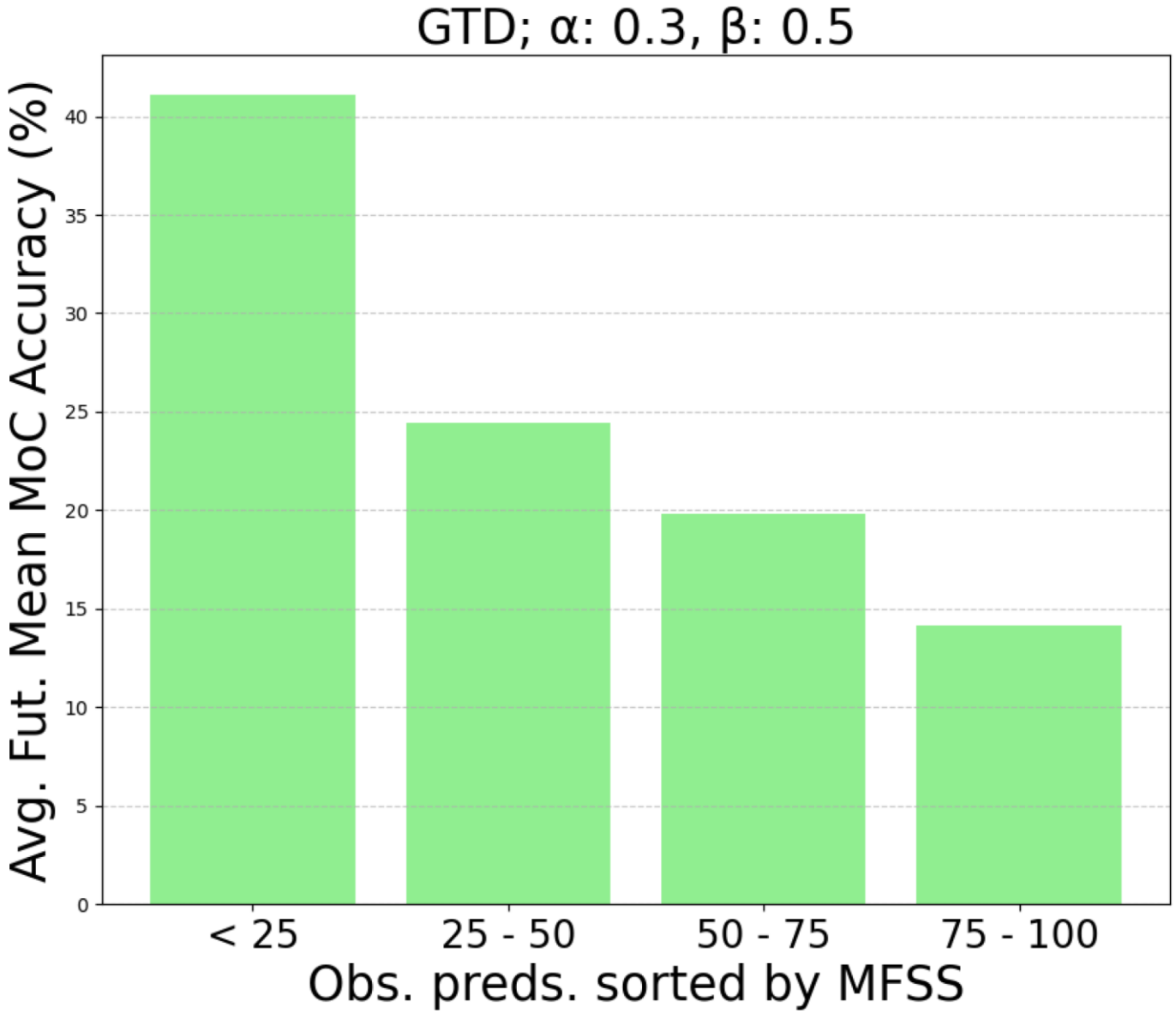
[O. Zatsarynna et al. **Gated Temporal Diffusion for Stochastic Long-term Dense Anticipation.** ECCV 2024]

Gated Temporal Diffusion for Anticipation

[O. Zatsarynna et al. **Gated Temporal Diffusion for Stochastic Long-term Dense Anticipation.** ECCV 2024]

Gated Temporal Diffusion for Anticipation

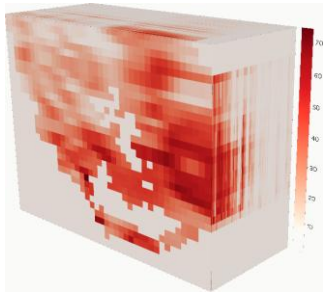
- Anticipation accuracy depends on uncertainty in observation



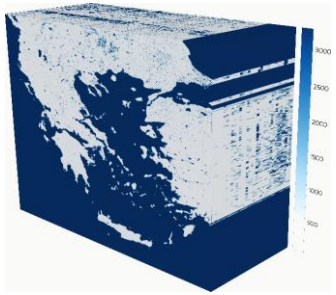
[O. Zatsarynna et al. **Gated Temporal Diffusion for Stochastic Long-term Dense Anticipation.** ECCV 2024]


Forecasting
Hundreds left homeless as Greek fires rage

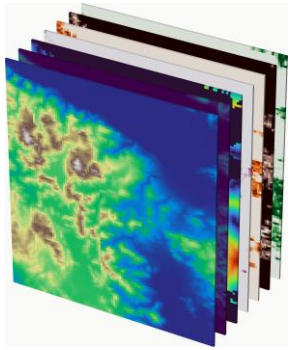
Wildfire Forecasting



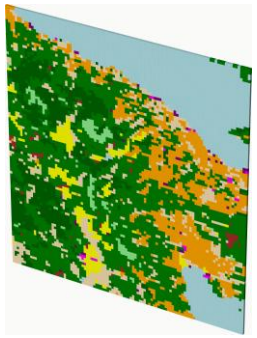
Meteorological data



Satellite products



Topographic variables

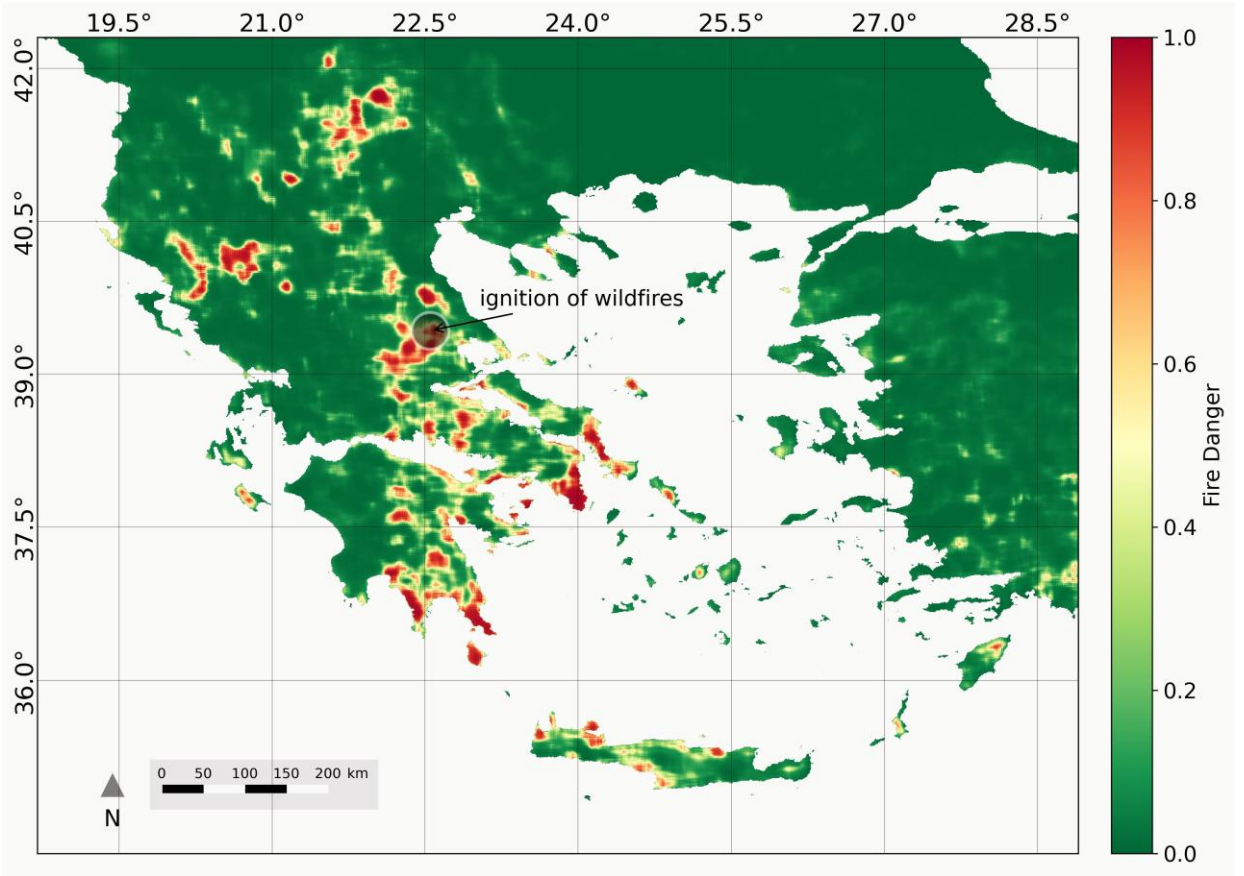


Land cover

Resolution:
1km × 1km × 1day

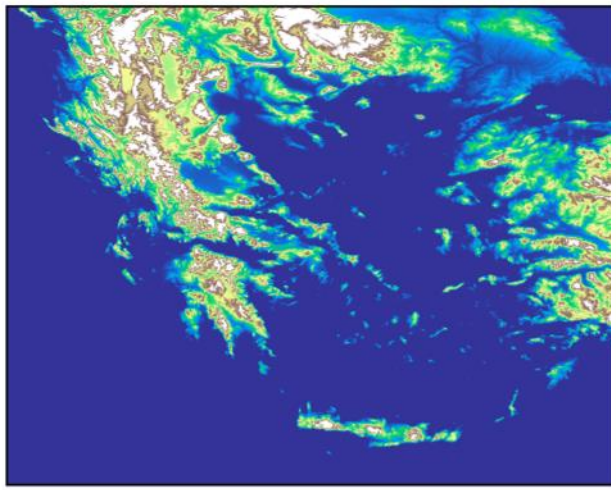
years - 2009-2021

Wildfire susceptibility map for 02/07/2021

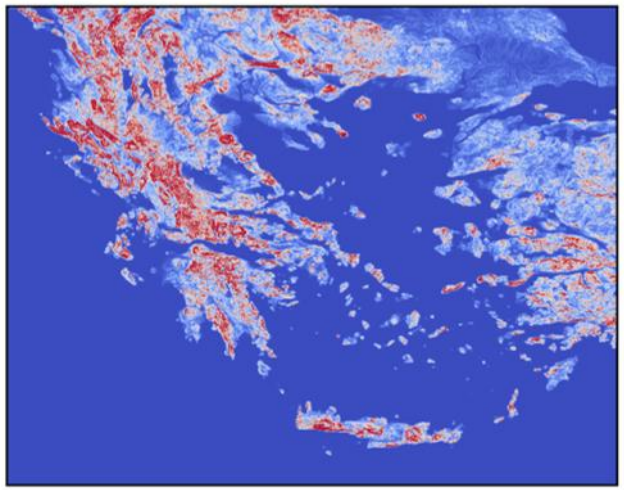


Static vs. Dynamic Variables

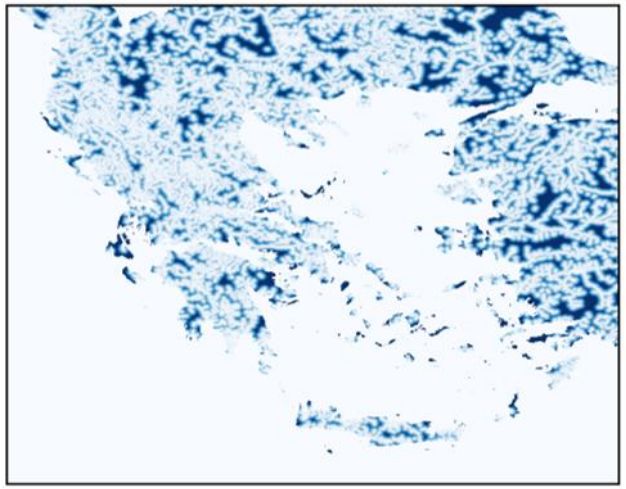
elevation



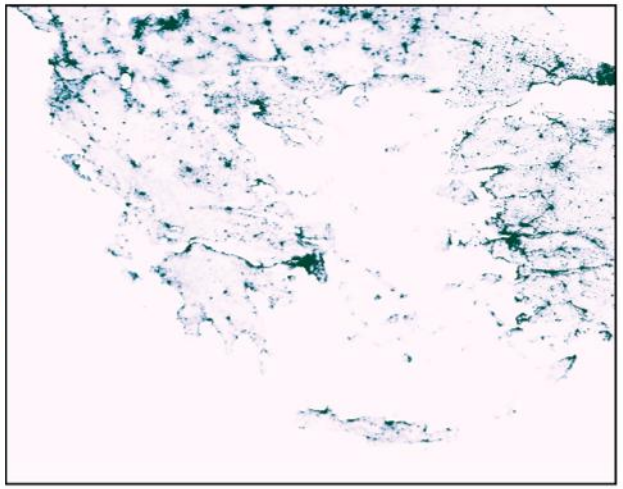
slope

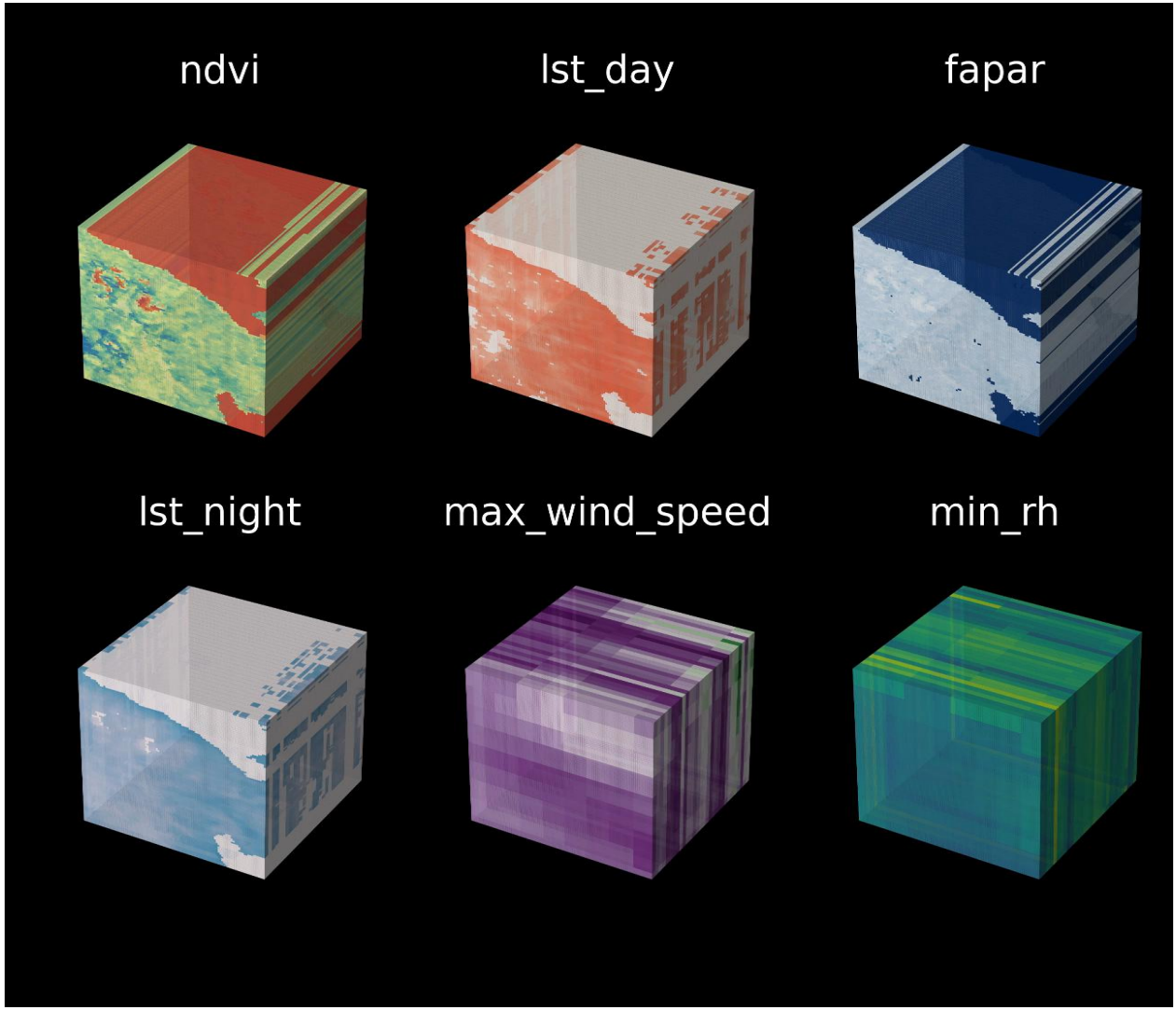


waterway distance

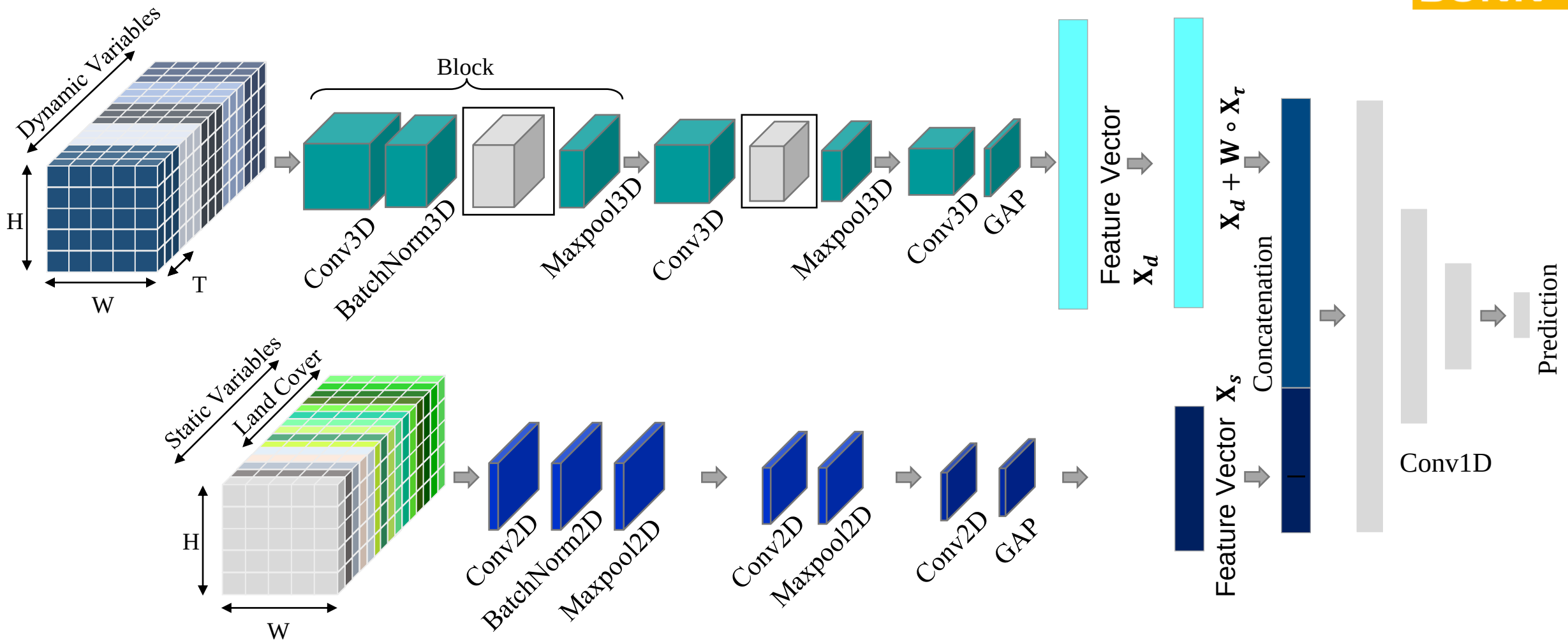


population density





Static vs. Dynamic Variables



[M.H.S. Eddin et al. **Location-Aware Adaptive Normalization: A Deep Learning Approach for Wildfire Danger Forecasting.** IEEE Transactions on Geoscience and Remote Sensing 2023]

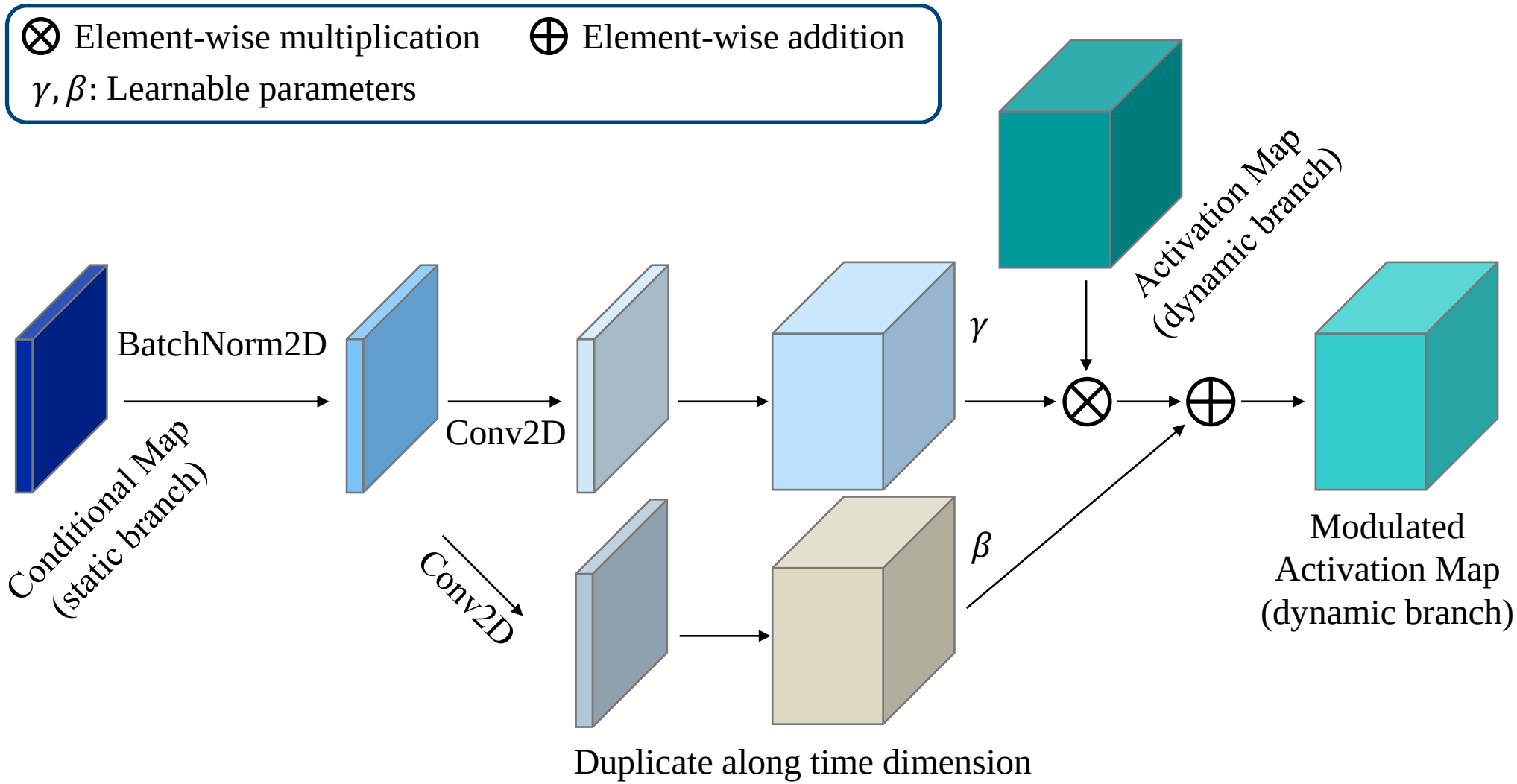
Location-aware Adaptive Normalization layer (LOAN)

- Dynamic variables often depend on static variables
- Modulate dynamic features by static features

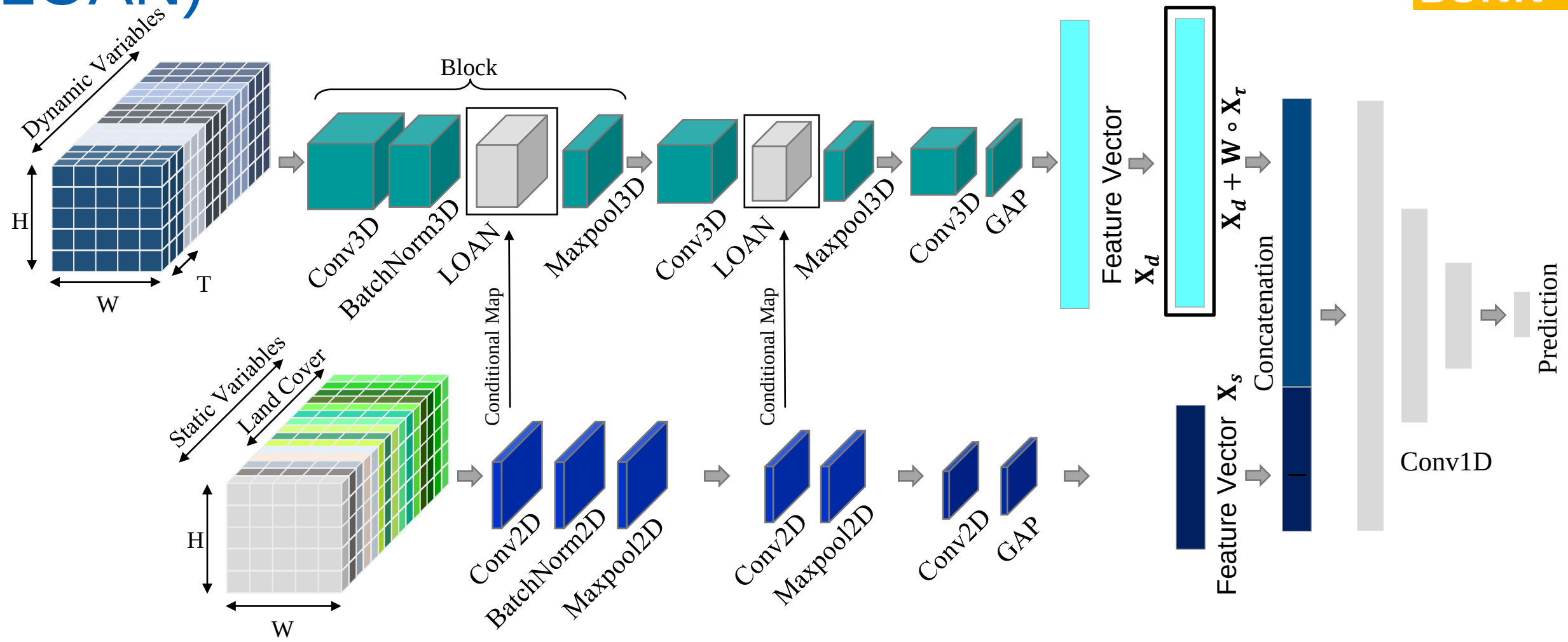
$$z_{d}^{i}{}_{(n,k,t,w,h)} \cdot \gamma_{s}^{i}{}_{(n,k,w,h)} + \beta_{s}^{i}{}_{(n,k,w,h)}$$

[M.H.S. Eddin et al. **Location-Aware Adaptive Normalization: A Deep Learning Approach for Wildfire Danger Forecasting.** IEEE Transactions on Geoscience and Remote Sensing 2023]

Location-aware Adaptive Normalization layer (LOAN)

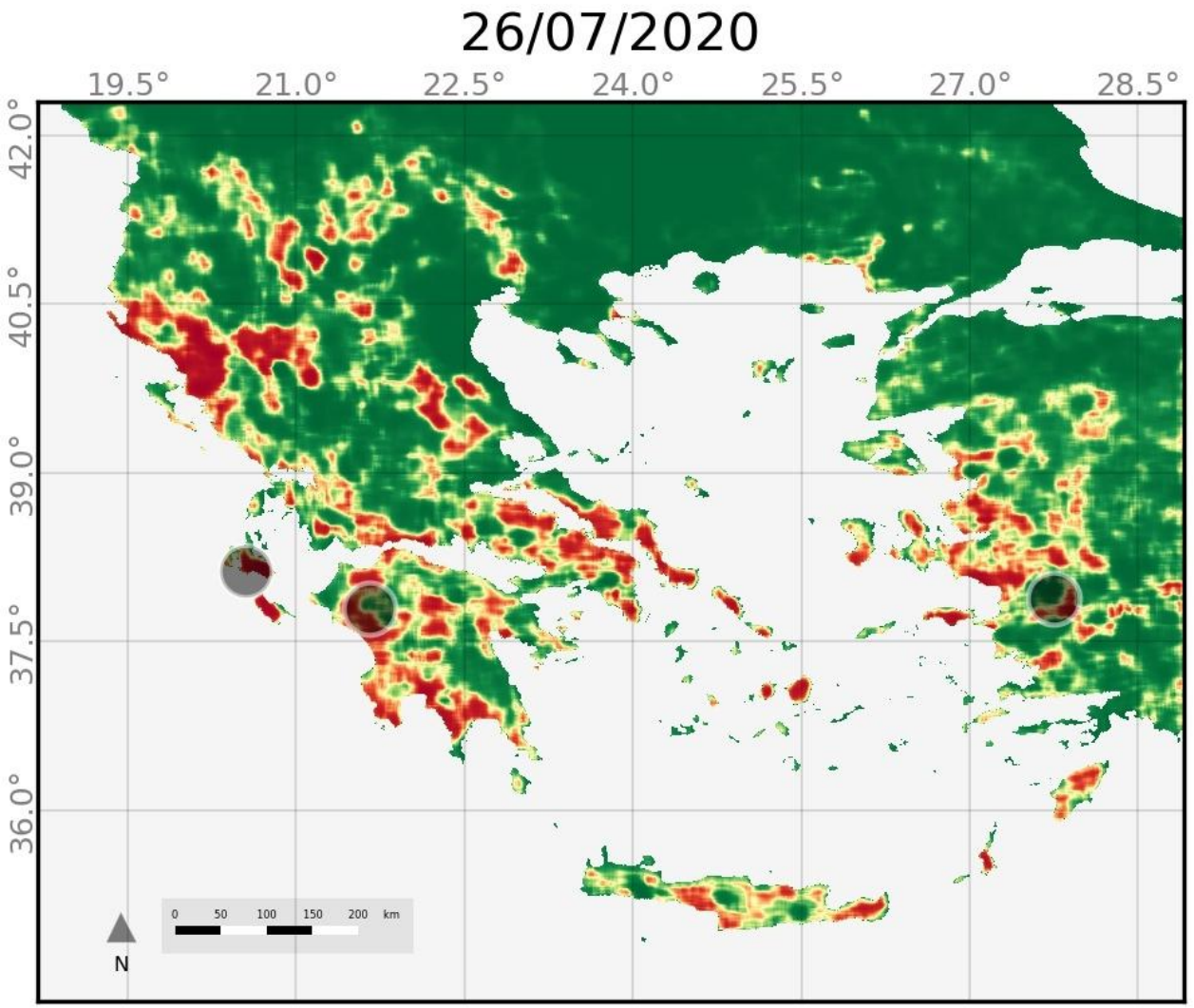


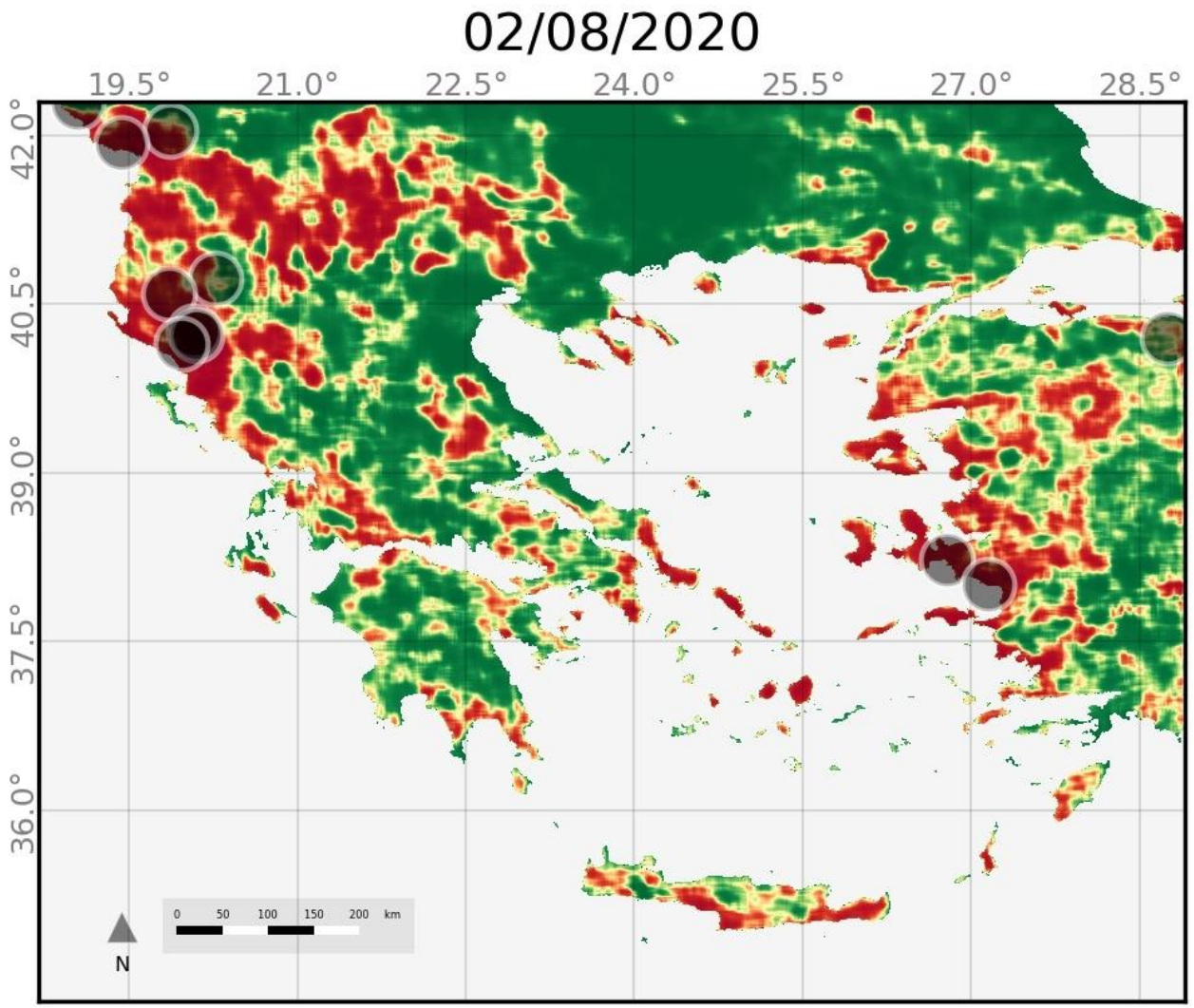
Location-aware Adaptive Normalization layer (LOAN)



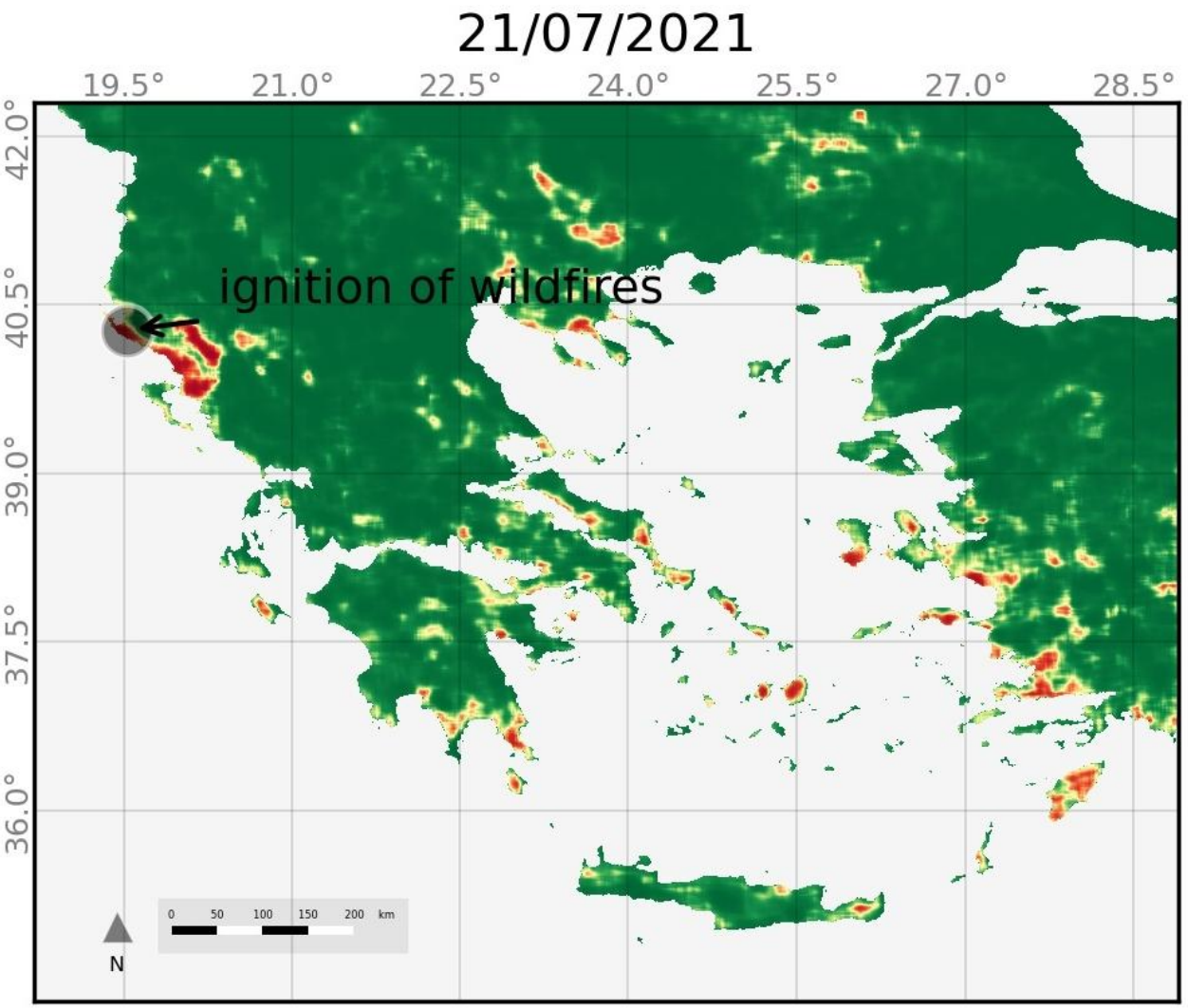
[M.H.S. Eddin et al. **Location-Aware Adaptive Normalization: A Deep Learning Approach for Wildfire Danger Forecasting.** IEEE Transactions on Geoscience and Remote Sensing 2023]

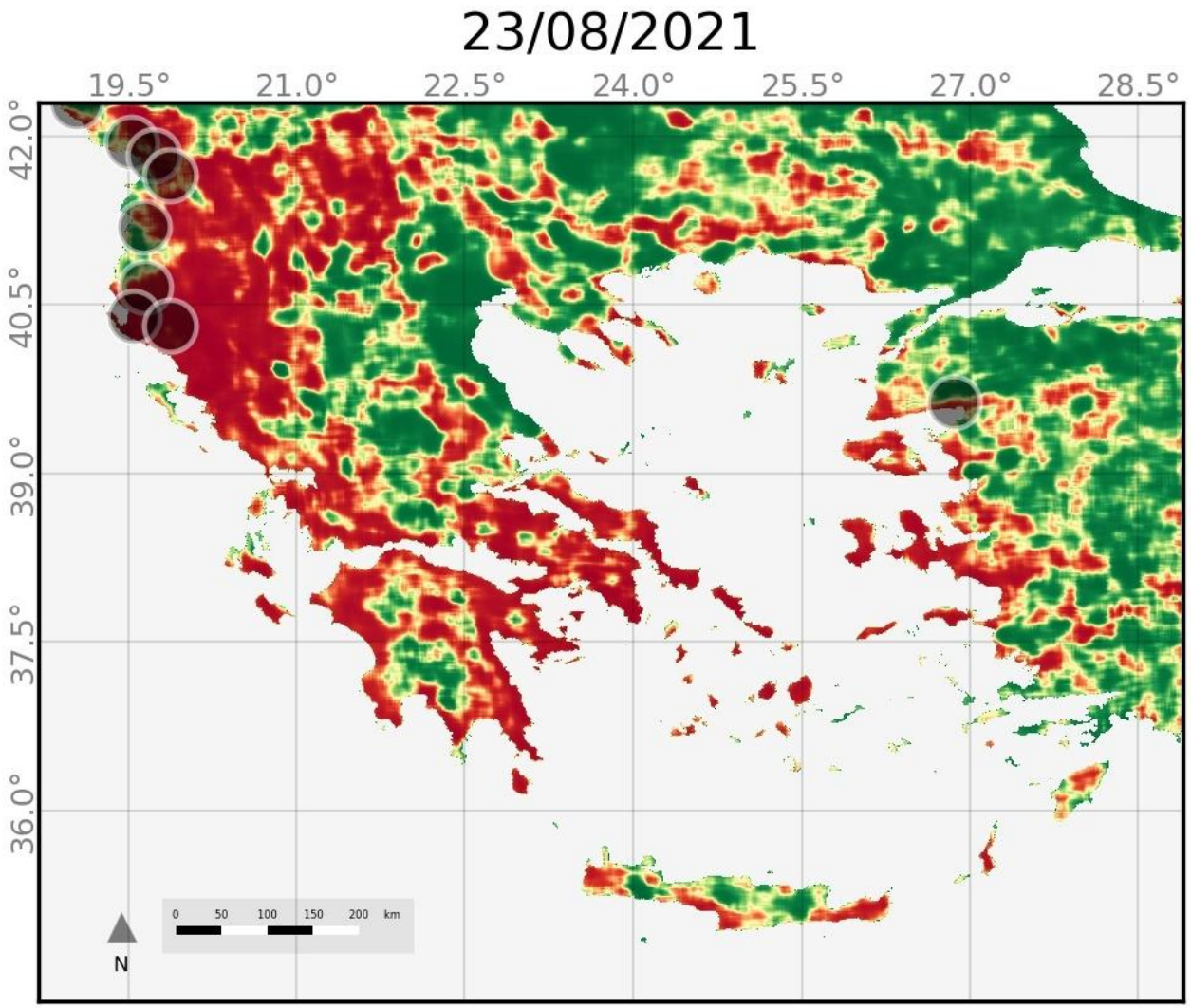
Qualitative Results





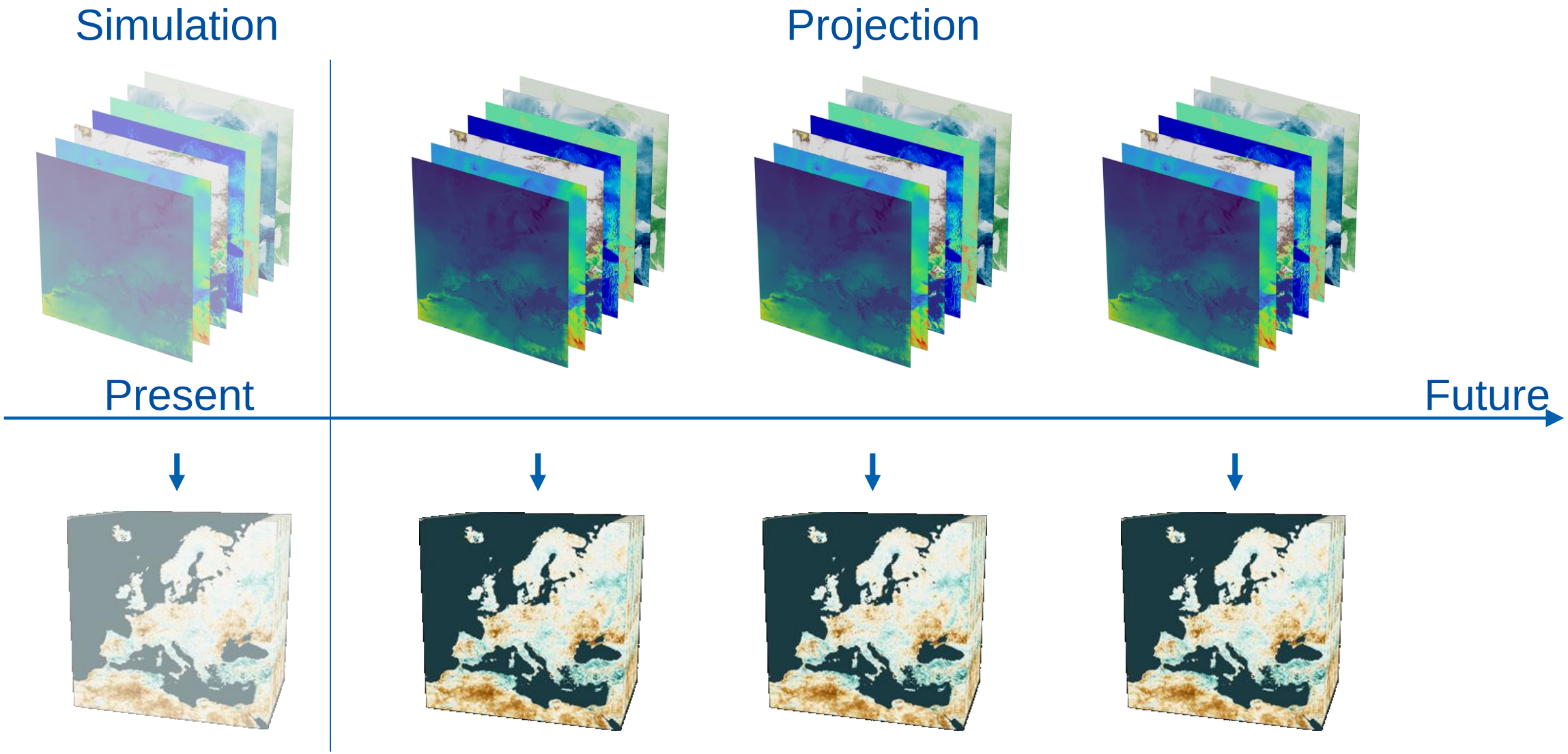
Qualitative Results



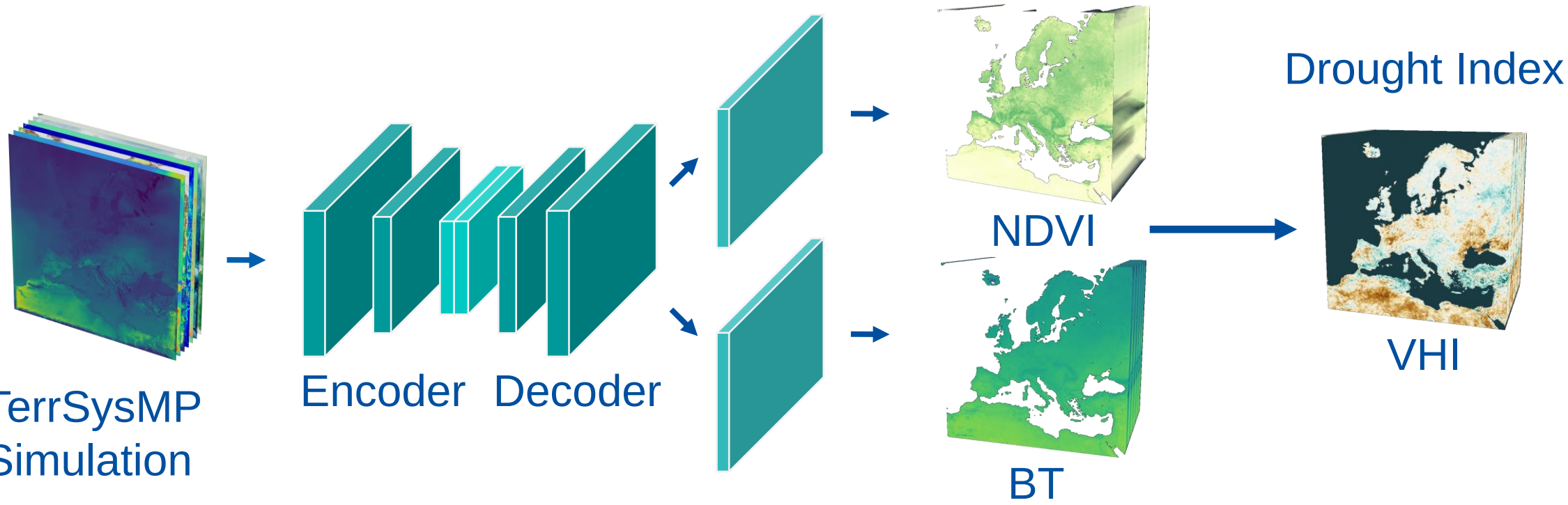


Forecasting Agricultural Drought

Forecasting Agricultural Drought

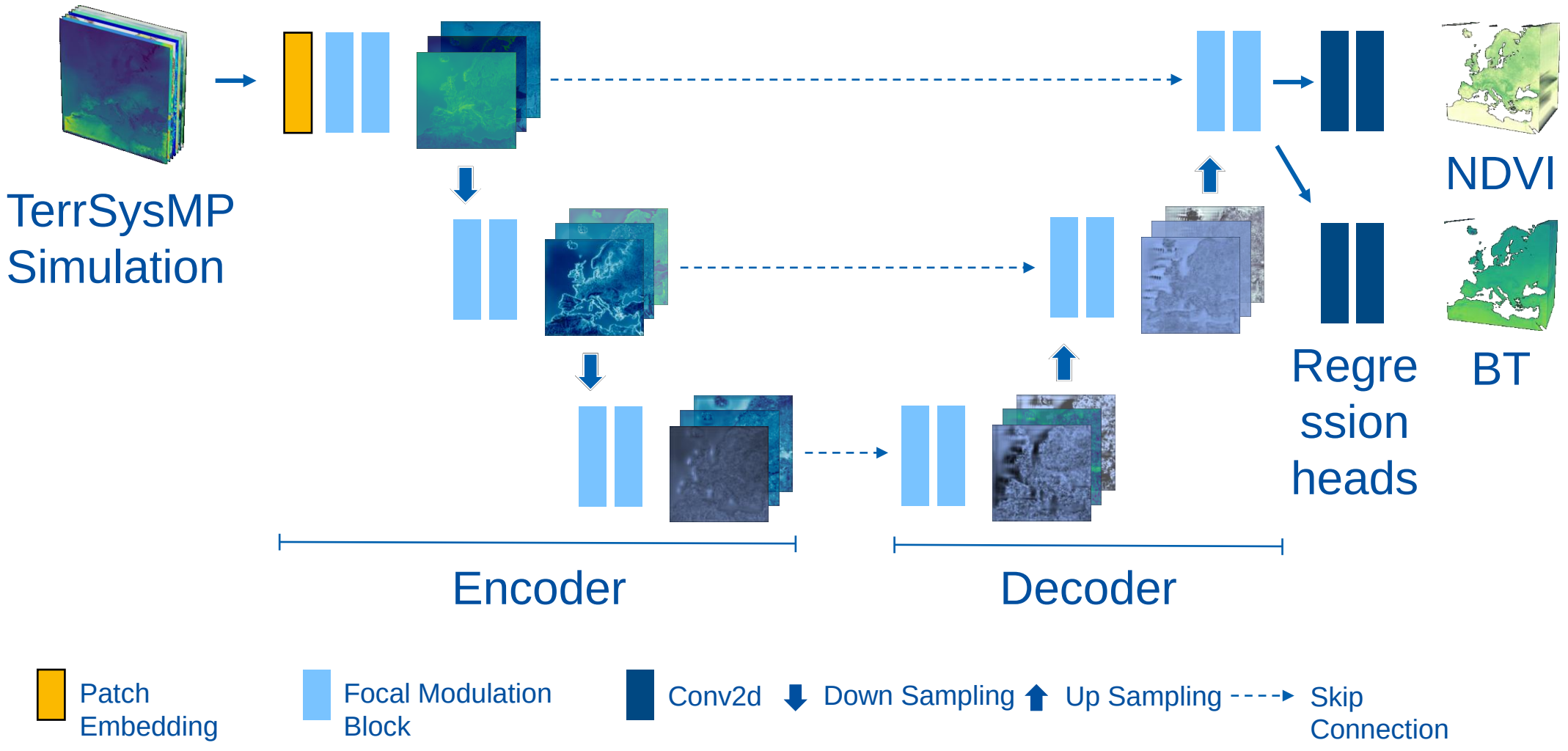


Forecasting Agricultural Drought

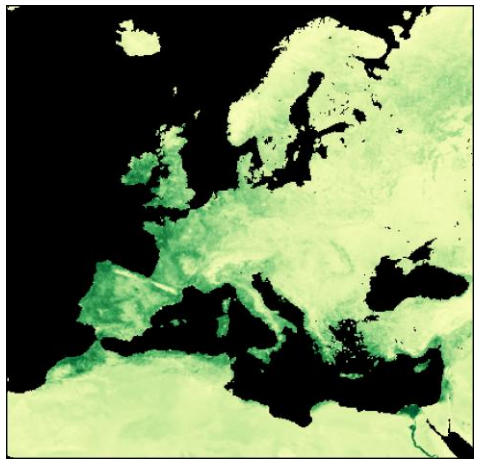


[M.H.S. Eddin et al. **Focal-TSMP: Deep learning for vegetation health prediction and agricultural drought assessment from a regional climate simulation.** EGUsphere preprint 2023]

Forecasting Agricultural Drought



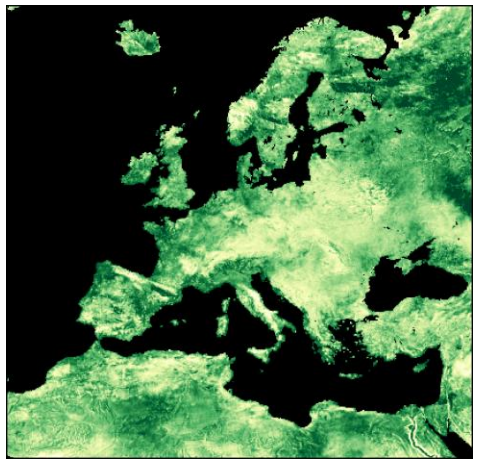
Agricultural Drought Indices



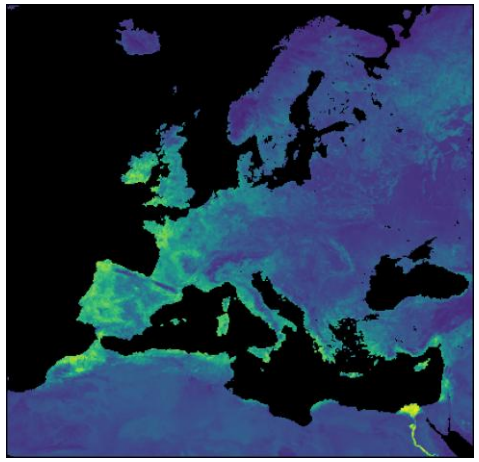
NDVI

$$VCI = 100\ \frac{NDVI\ - NDVI_{min}}{NDVI_{max} - NDVI_{min}}$$

Moisture Condition



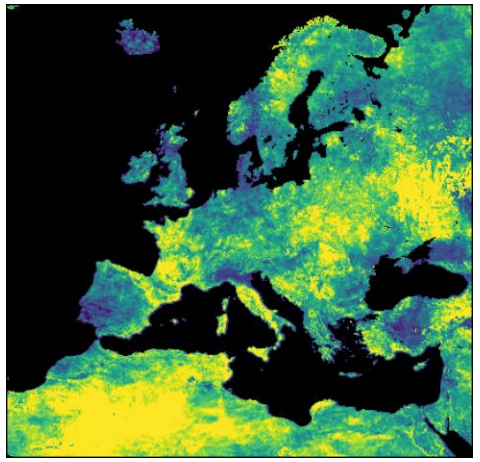
VCI



BT

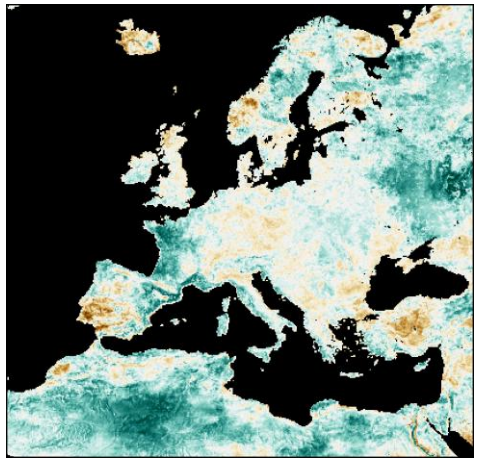
$$TCI = 100\ \frac{BT_{max}\ - BT}{BT_{max} - BT_{min}}$$

Thermal Condition



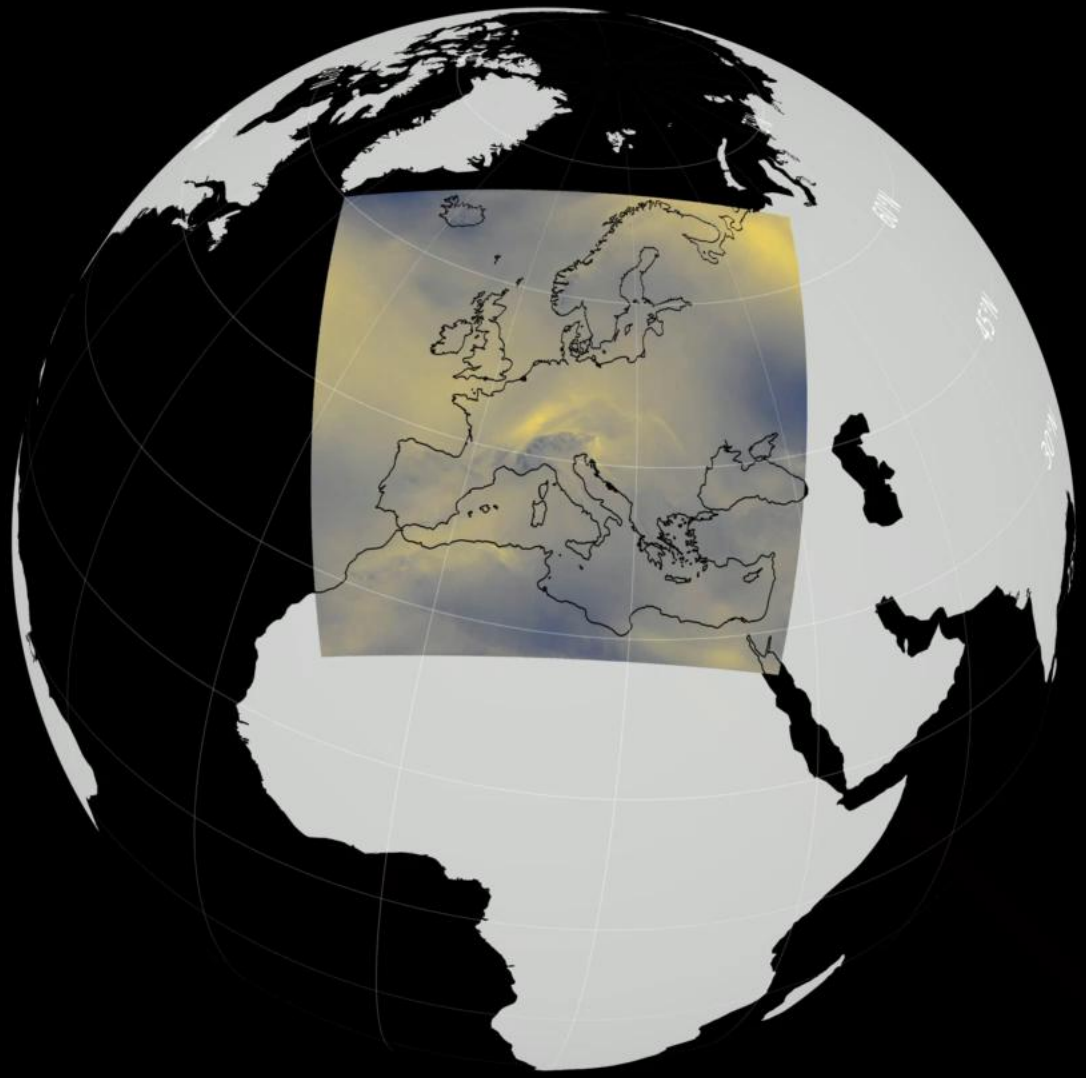
TCI

Vegetation Health Index



$$VHI = \alpha(VCI) + (1 - \alpha)TCI$$

Forecasting Agricultural Droughts



Thank you for your attention.

LAMARR

Institute for Machine Learning
and Artificial Intelligence

UNIVERSITÄT BONN