PDE-DRIVEN SPATIOTEMPORAL DISENTANGLEMENT

ICLR 2021

May 3rd to 7th, 2021

https://openreview.net/forum?id=vLaHRtHvfFp

Jérémie Donà,[1] Jean-Yves Franceschi,[1]
Sylvain Lamprier,[1] Patrick Gallinari[1,2]

[1]Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
[2]Criteo AI Lab, Paris, France

Fact

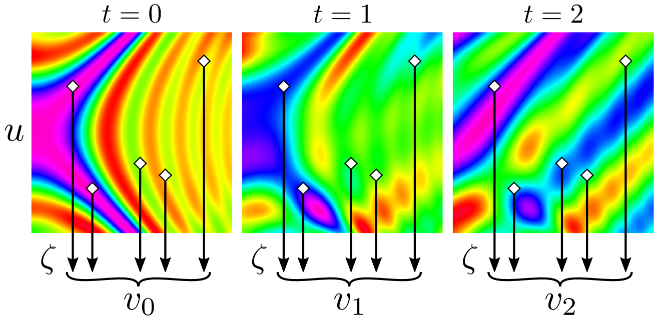
Disentanglement improves interpretability thanks to factorization

Hypothesis

A well designed spatiotemporal disentanglement would help prediction

- Prior spatiotemporal disentanglement works are often complex and do seldom analysis of its meaning
- We aim at grounding spatiotemporal disentanglement on stronger foundations

- Observations $v = (v_{t_0}, \ldots, v_{t_1})$
- State of the system $u_v = u\colon (x,t) \mapsto u_v(x,t)$
- We suppose that v is a measurement of u: $v_t = \zeta(u_v(.,t))$



Idea: Separation of Variables

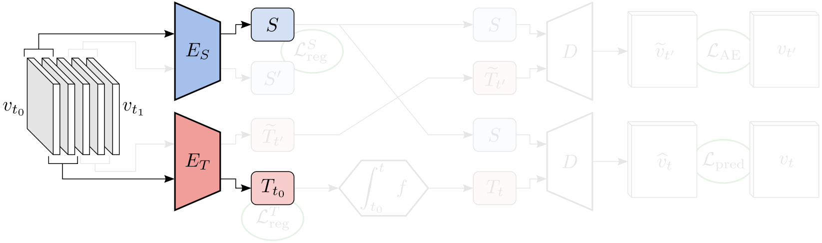
- We propose a novel interpretation of spatiotemporal disentanglement based on the separation of variables in PDEs
- Example with the heat equation:

$$\frac{\partial u}{\partial t} = c^2 \frac{\partial^2 u}{\partial x^2}, \quad u(0,t) = u(L,t) = 0, \quad u(x,0) = f(x),$$

with separable solutions:

$$u(x,t) = \underbrace{\mu \sin\left(\frac{n\pi}{L}x\right)}_{\phi(x)} \times \underbrace{\exp\left(-\left(\frac{cn\pi}{L}\right)^2 t\right)}_{\psi(t)} = \xi\big(\phi(x), \psi(t)\big)$$

- ϕ and ψ can be found by solving an ODE on t and a PDE on x

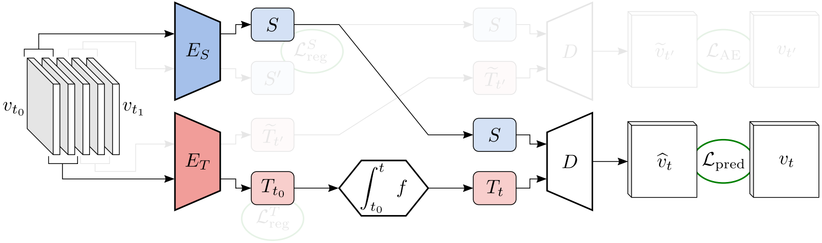


- We learn latent S, T_t and a decoder D such that:

$$\phi \equiv S \in \mathbb{R}^d, \qquad \psi \equiv T \colon t \mapsto T_t \in \mathbb{R}^p, \qquad D \equiv \zeta \circ \xi$$

- S and T_{t_0} are inferred with encoders E_S and E_T from conditioning frames:

$$V_\tau(t_0) = \left(v_{t_0}, \ldots, v_{t_0+\tau}\right)$$



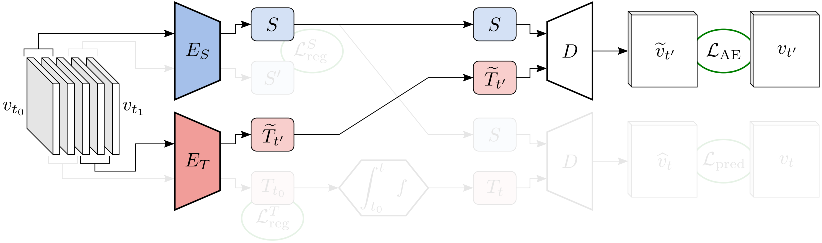
- $T \equiv \psi$ is driven by an ODE:

$$\frac{\partial T_t}{\partial t} = f(T_t)$$

- Forecasting and alignment losses:

$$\mathcal{L}_{\text{pred}} = \sum_t \|\widehat{v}_t - v_t\|_2^2, \quad \mathcal{L}_{\text{AE}} = \left\| D\Big(S, E_T\big(V_\tau(t')\big)\Big) - v_{t'} \right\|_2^2$$

SCIENCES SORBONNE UNIVERSITÉ

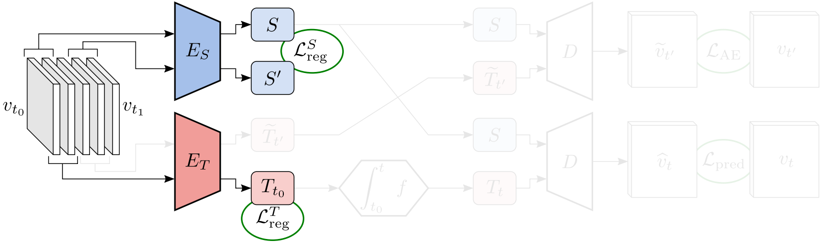


- $T \equiv \psi$ is driven by an ODE:

$$\frac{\partial T_t}{\partial t} = f(T_t)$$

- Forecasting and alignment losses:

$$\mathcal{L}_{\text{pred}} = \sum_t \|\widehat{v}_t - v_t\|_2^2, \quad \mathcal{L}_{\text{AE}} = \left\| D\Big(S, E_T\big(V_\tau(t')\big)\Big) - v_{t'} \right\|_2^2$$



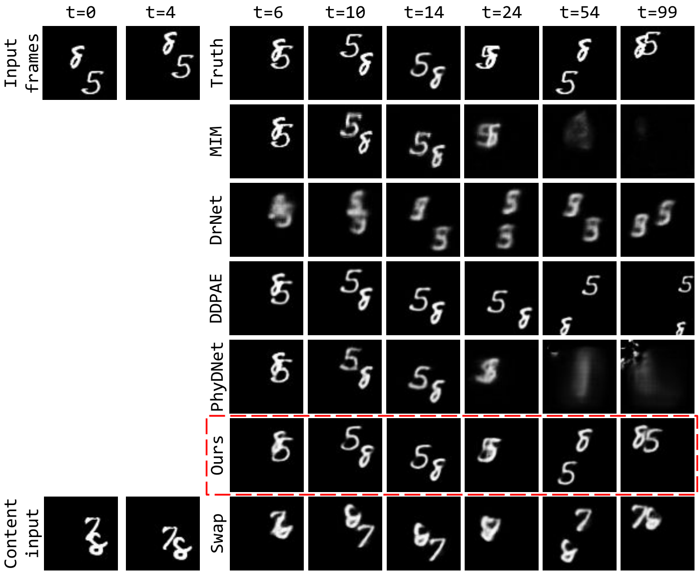
- From a strict S invariance constraint to a weaker one to take into account variations of observable content:

$$\frac{\partial E_S(V_\tau(t))}{\partial t} = 0 \Rightarrow \mathcal{L}^S_{\text{reg}} = \left\| E_S(V_\tau(t_0)) - E_S(V_\tau(t_1 - \tau)) \right\|_2^2$$

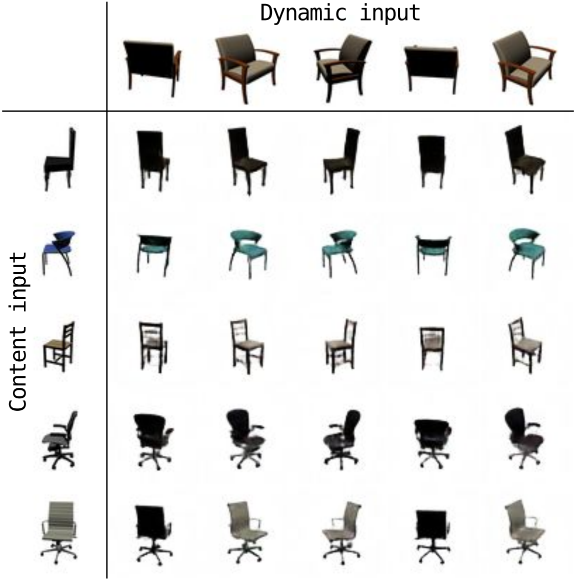
- Disentanglement loss:

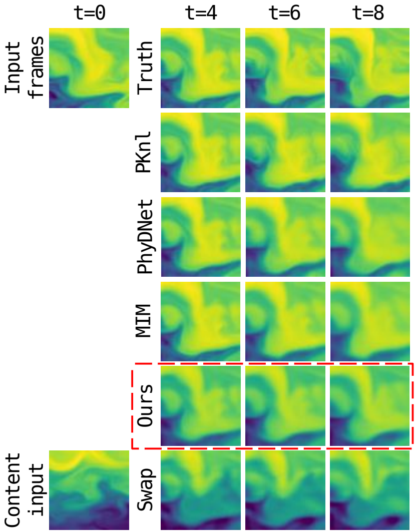
$$\mathcal{L}^T_{\text{reg}} = \left\| T_{t_0} \right\|_2^2 = \left\| E_T(V_\tau(t_0)) \right\|_2^2$$

Results (MNIST)



Results (Multiview)




t=0
t=4
t=6
t=8
Input frames
Truth
PKnl
PhyDNet
MIM
Ours
Content input
Swap

- We learn a simpler PDE acting on fewer variables thanks to the separation of variables inspiration
- This induces a simple model with better prediction and disentanglement performances
- Thorough experiments, comparisons and ablation studies confirm the benefits of the model.