18th of February, 2020

# Generative Models in e-Commerce

Urs Bergmann

Zalando Research &
HLEG AI, European Commission

zalando research

# ZALANDO - FASHION E-COMMERCE AT A EUROPEAN SCALE

~ 6.5 billion EUR
revenue 2019

~ 145 million
orders in 2019

31 million
active customers

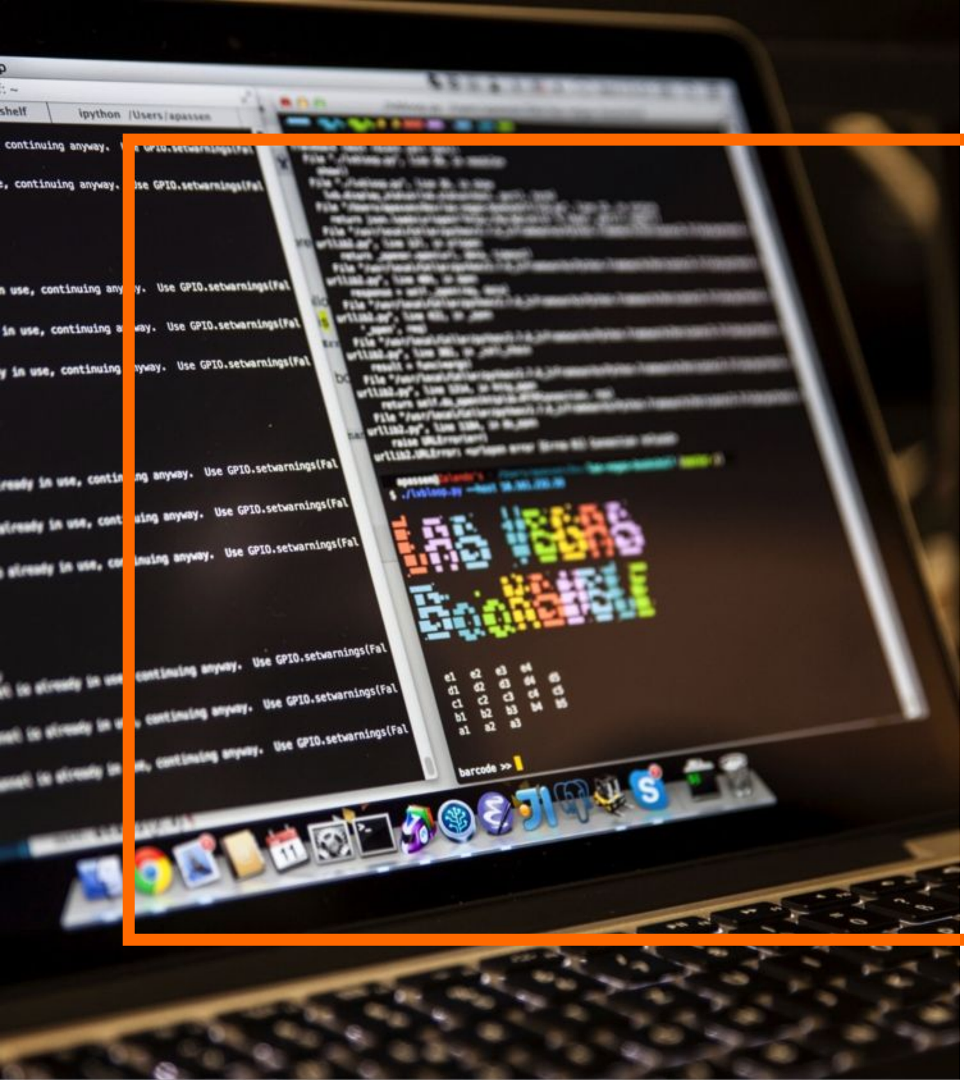
> 500,000
product choices

~ 14,000
employees in Europe

>70%
of visits via mobile devices

> 2,500
brands

17
countries



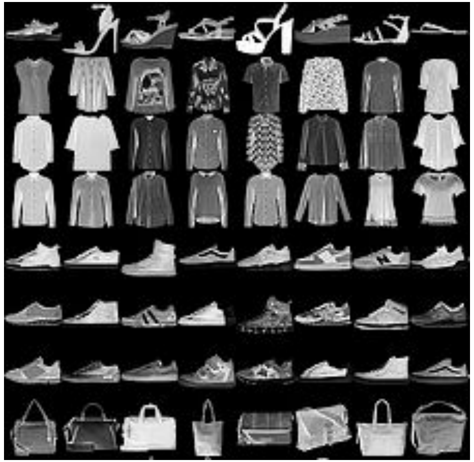
# Deep Learning

# AN EXEMPLARY LEARNING PROBLEM - FASHION MNIST

MNIST [LeCun et al., 1998]

## Supervised Learning Task

- learn on pairs of images and classes $\{(x_i, y_i)\}_{i \in \mathcal{D}}$
- predict unknown class for novel image



Fashion MNIST [Xiao et al., 2017]

## Proven Best Solutions: Deep Learning

- Define *architecture*
- Define objective / *loss function*

Optimize loss on data

## Generative Modeling: a difficulty with hand-engineered losses

- hand-engineered losses are problematic, e.g.

$$\mathrm{MSE} = \sum_i \left(f_i(x) - y_i\right)^2 = -\log \mathcal{L}$$

- MSE corresponds to isotropic Gaussian

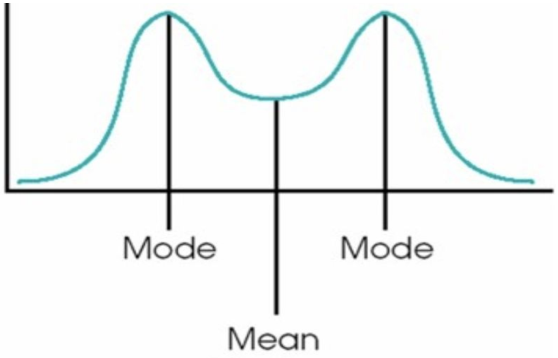
$$\mathcal{L} \propto \exp\left(-\sum_i \left(f_i(x) - y_i\right)^2\right)$$

- assumes independence and optimizes for mean!

  → hence results in **blurry images**!

### Possible Improvements

- VAE [Kingma&Welling, Rezende et al., 2014]
- AR models [Bengio&Bengio, 2000]
- Flows [Rezende&Mohamed, 2015]
- GANs [Goodfellow et al., 2014]




(a) Input context

(c) Context Encoder
($L2$ loss)

[Pathak et al. 2016]

## Generative Adversarial Networks [Goodfellow et al., 2014]

### Core Idea

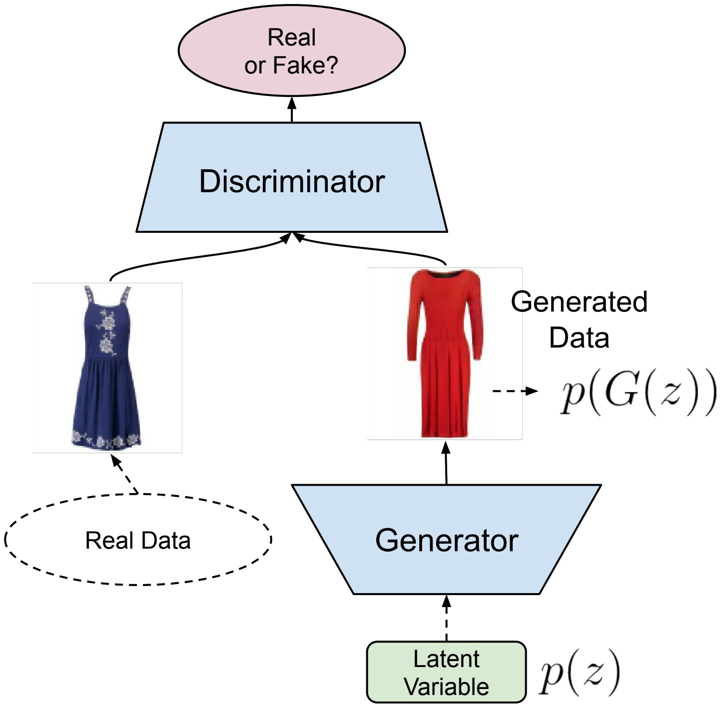
- simple noise prior $p(z)$
- target: transform prior to to data distribution

$$p(G(z)) \sim p(x)$$

- solved via two-player game:
  - G: generate data from distribution
  - D: estimate probability of generated vs. real data

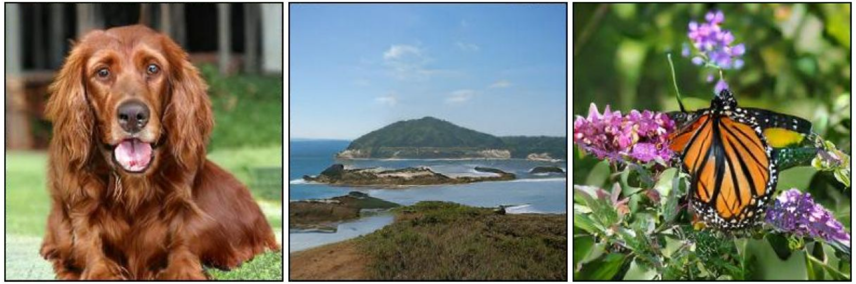
### Original Training Value Function

$$\min_G \max_D V(D, G) = \mathbb{E}_{z \sim p_z(z)} \left[\log\left(1 - D\left(G\left(z\right)\right)\right)\right]$$
$$= \mathbb{E}_{x' \sim p_r(x)} \left[\log D(x')\right]$$



# Generative Adversarial Networks

[Brock et al., 2019]



[Karras et al., 2019]

## Advantages

- no need for loss in image space
- state-of-the-art image sampling/modeling

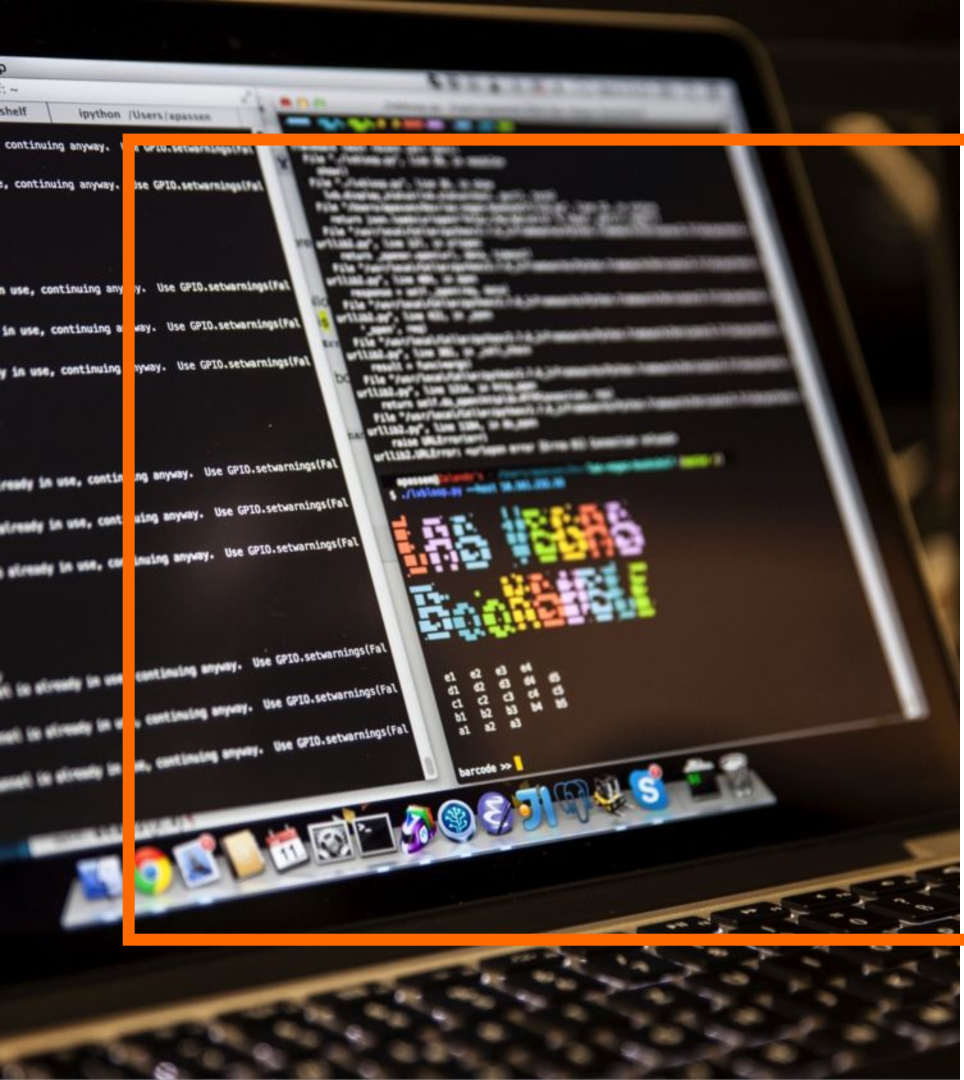
## Difficulties & Limitations

- implicit model - no likelihood & inference
- training hard & unclear best objective
- mode-collapse
- no fine-grained control over output
- fixed & limited output dimensionality

## Selected Improvements

- DCGAN [Radford et al., 2016]
- WGAN [Arjovski et al., 2017]
- WGAN-GP [Gulrajani et al., 2017]
- SAGAN [Zhang et al., 2018]
- SN-GAN [Miyato et al., 2018]
- StyleGAN [Karras et al., 2018]
- BIGGAN [Brock et al., 2019]
- StyleGAN2 [Karras et al., 2019]

Fashion Design



# Disentangling Input Conditionals

[Yildirim et al., 2018]

## Controls & Attributes

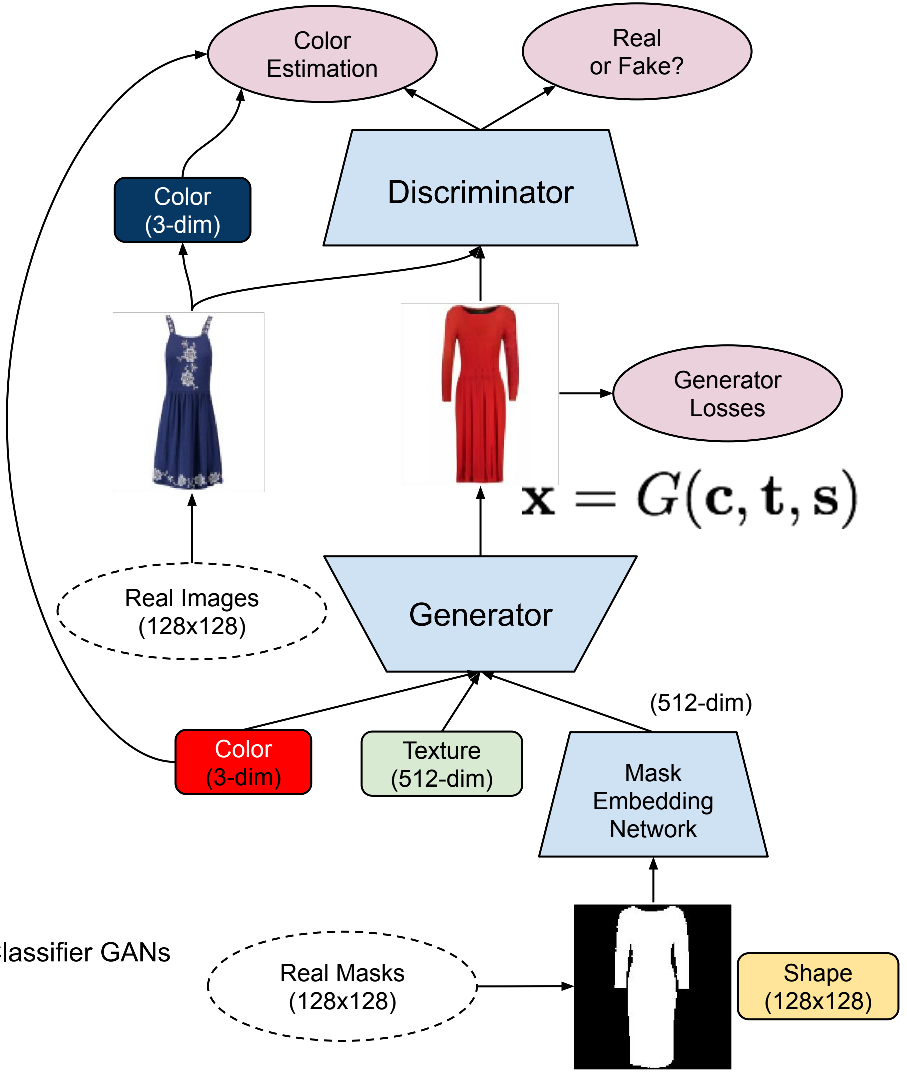
- Color $\mathbf{c}$
  - 3-dim, RGB
- Texture (local structure) $\mathbf{t}$
  - 512-dim
- Shape (mask) $\mathbf{s}$
  - embedded into 512-dim

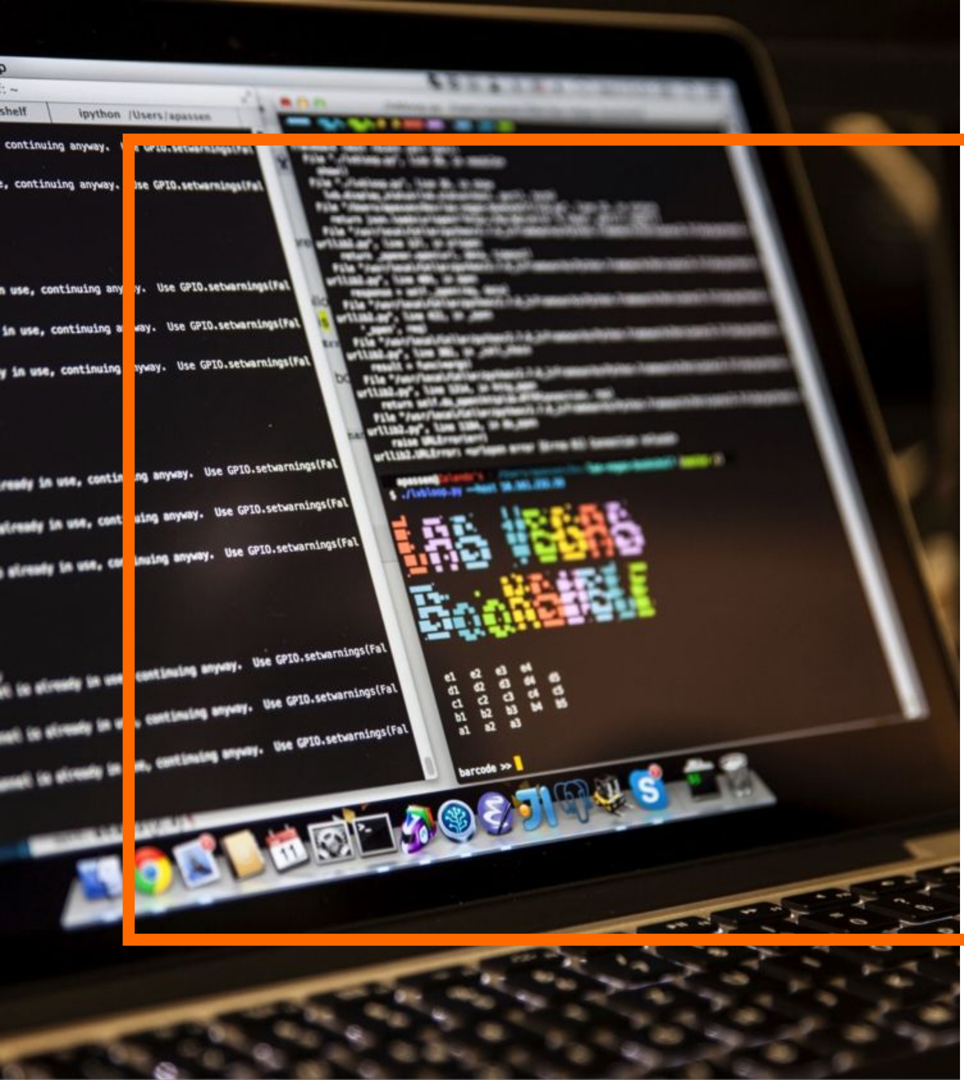
## Discriminator Loss

$$\max_D \boxed{\mathcal{L}_W - \lambda_{\text{gp}} \mathcal{L}_{\text{gp}}} - \boxed{\mathcal{L}_{\text{aux}}}$$

"Improved Training of Wasserstein GANs"
[Gulrajani et. al., 2017]

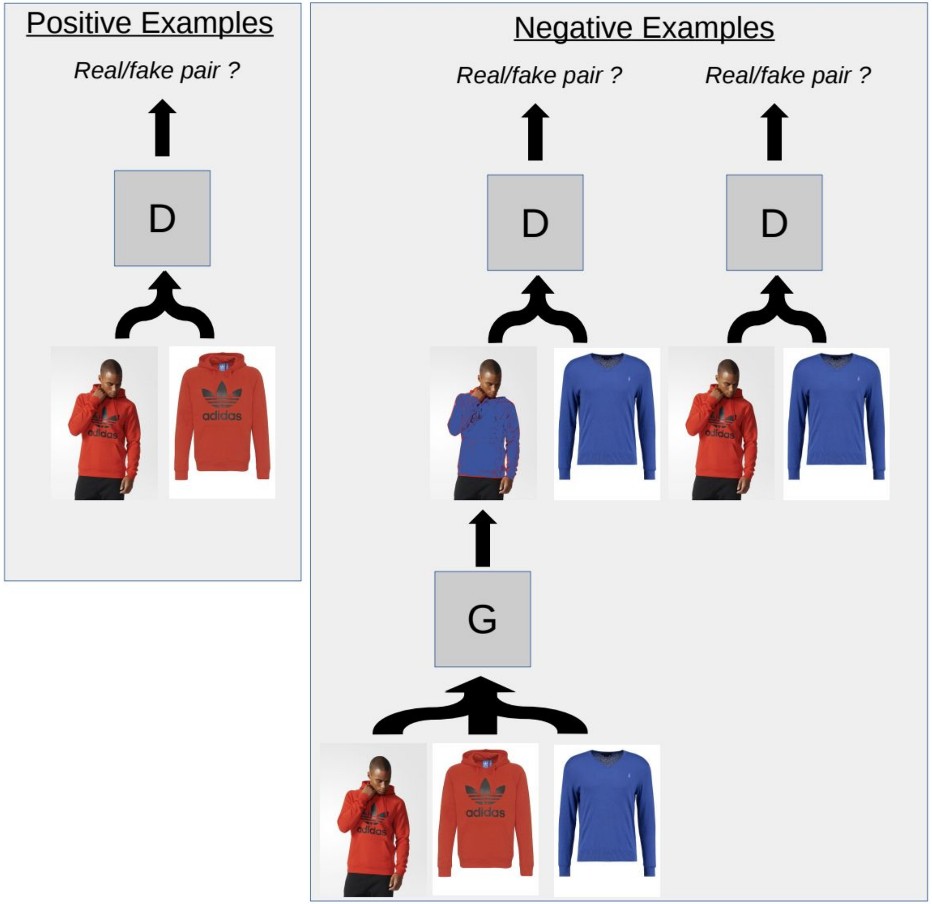
"Conditional Image Synthesis With Auxiliary Classifier GANs
[Odena et. al., 2017]





# Learning to Try On Clothes

# SWAP ARTICLES ON PEOPLE: CAGAN [Jetchev and Bergmann, 2017]

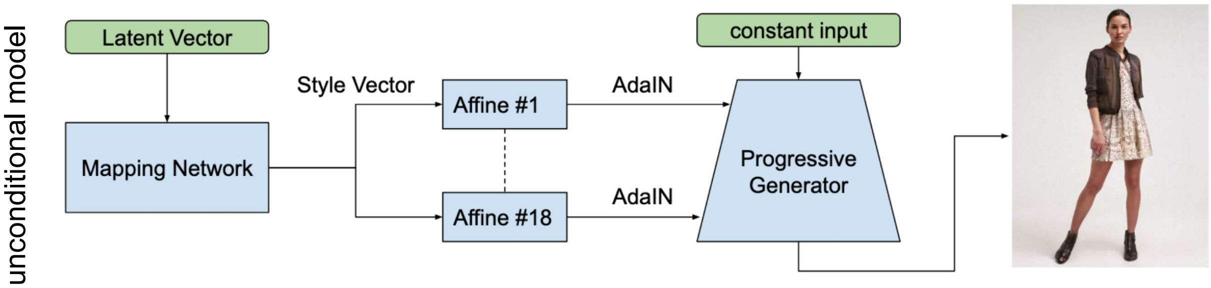
# SWAP ARTICLES ON PEOPLE: CAGAN [Jetchev and Bergmann, 2017]

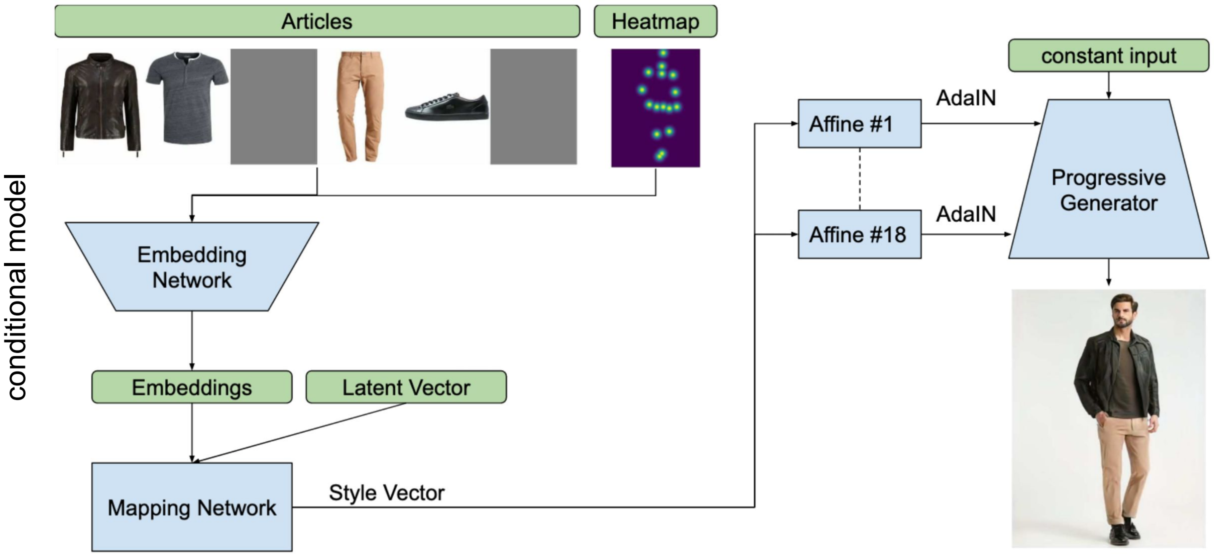
## Implicitly Learned Segmentation

## Limitation

- fine details hard to learn

# Configuring Pose & Outfits [Yildirim, Jetchev, Vollgraf, Bergmann, 2019]




## Method

- extend StyleGan [Karras et al, 2019]
- condition on article & pose embedding + random vector

## Data

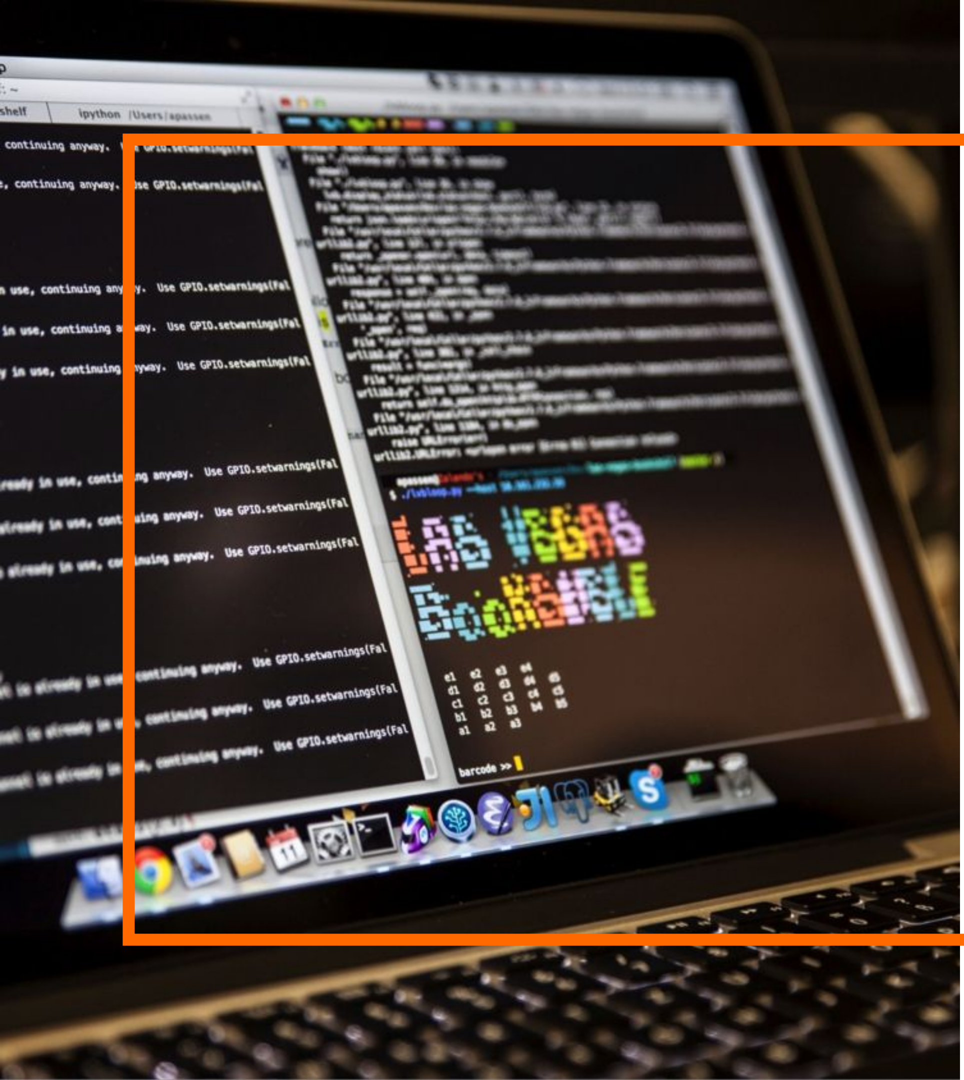
# Configuring Pose & Outfits - Unconditional Model Results

# Configuring Pose & Outfits - Conditional Model Results

Outfit #1

Outfit #2

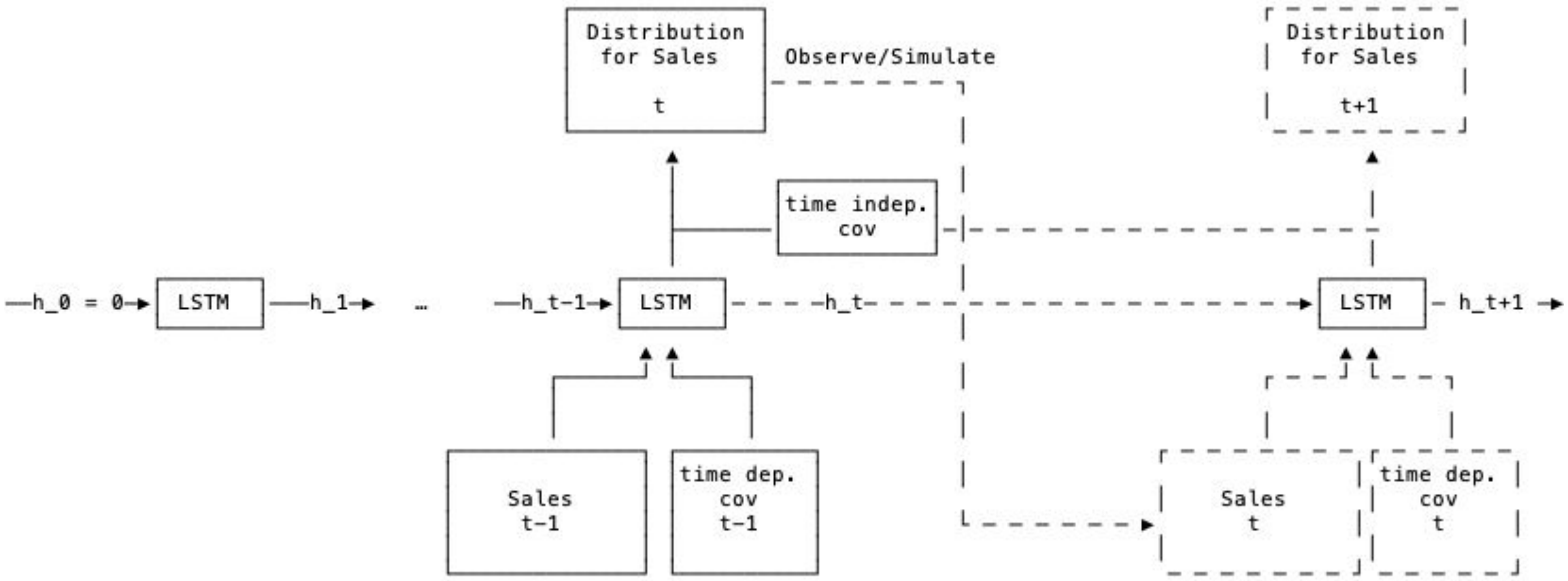
Urs Bergmann | Zalando Research




# Time Series Forecasting:

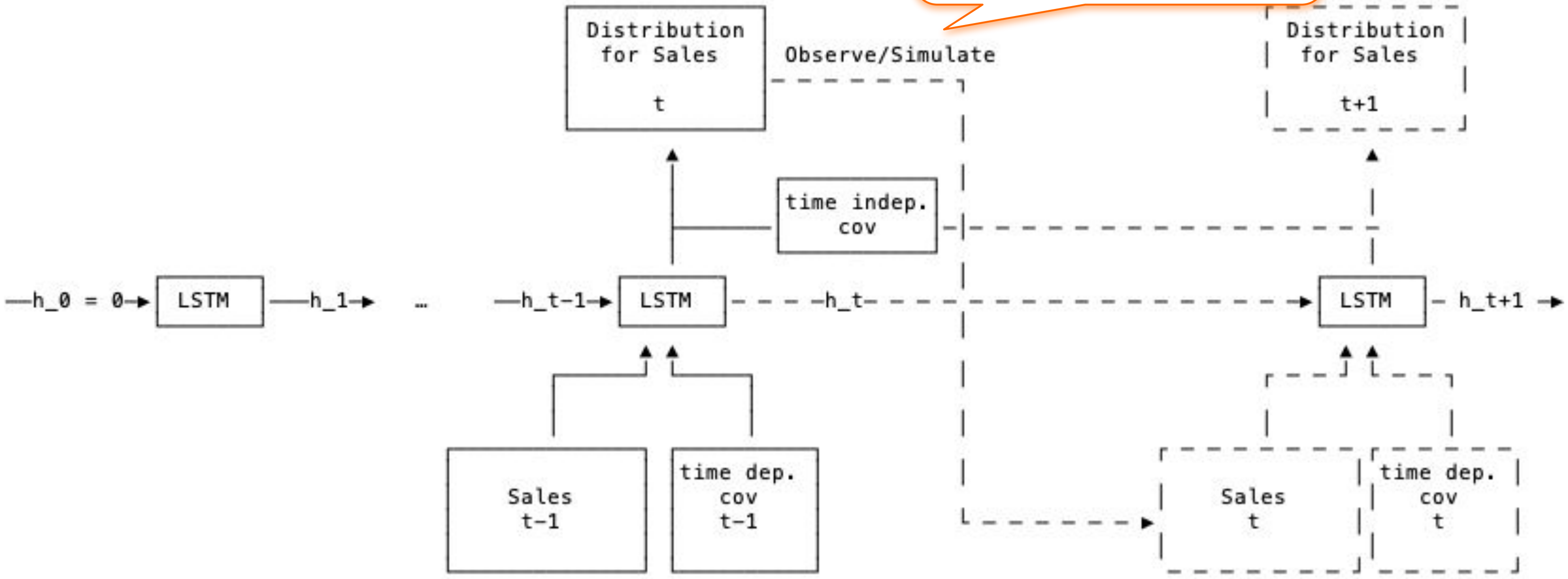
## Sales Prediction

# Starting out probabilistically: DeepAR model [Salinas et al., 2017]



# Starting out probabilistically: DeepAR model

training:
loss is log likelihood of
observed value

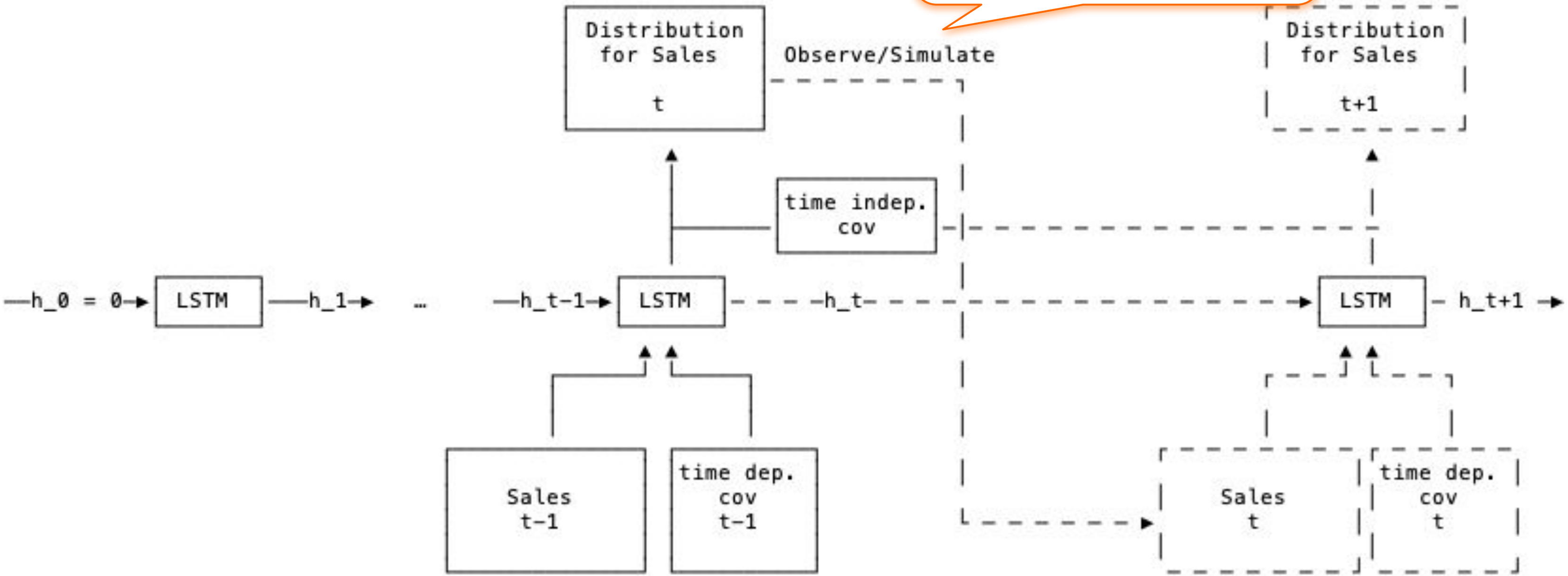





zalando research

# Starting out probabilistically: DeepAR model

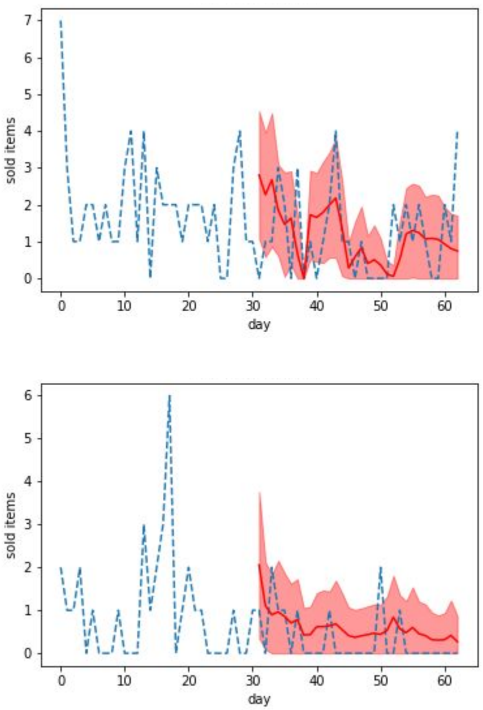
inference:
sample from distribution to
get quantiles etc.



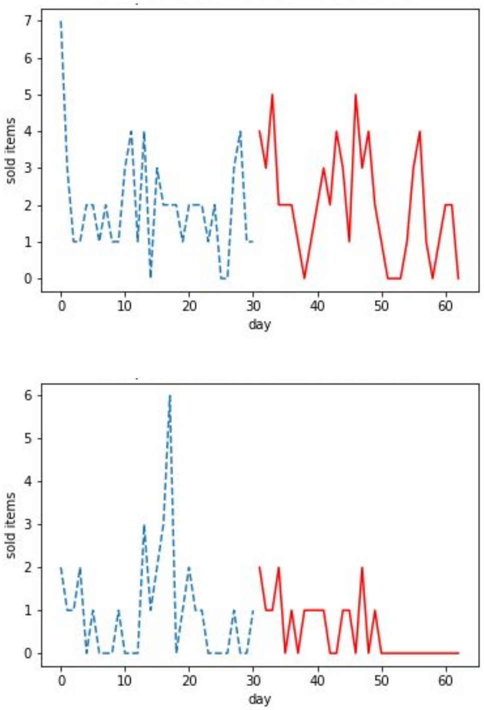



zalando research

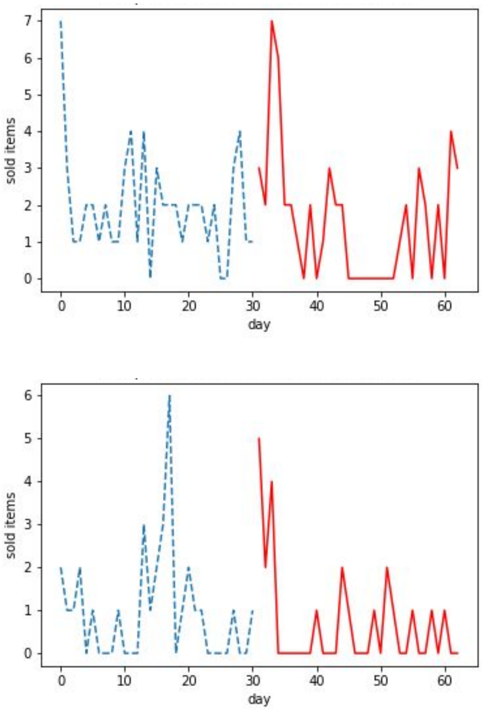
# Simulator runs for two articles




averaged simulations



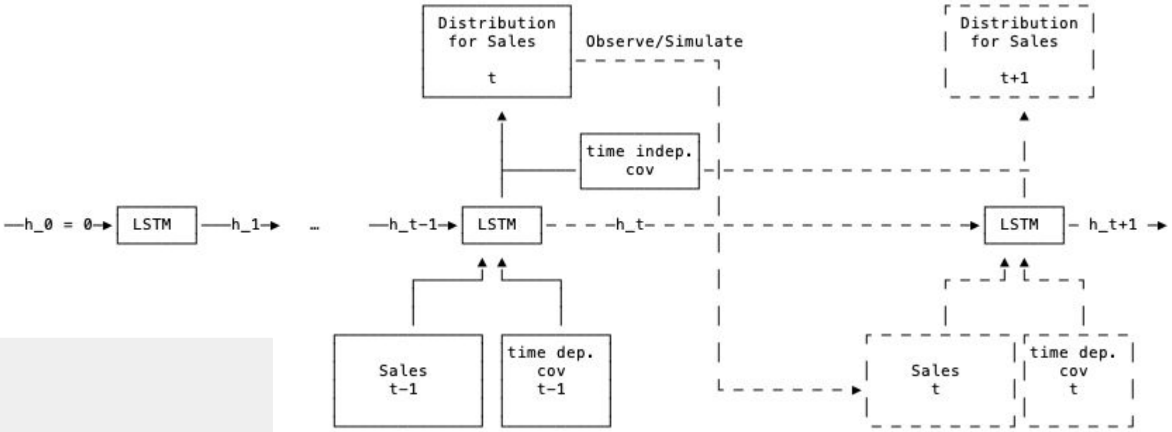

simulation 1




simulation 2

zalando research

# Starting out probabilistically: DeepAR model [Salinas et al., 2017]




## Limitations

- distribution needs to be specified
- time-series modeled independently
    - no article interactions!



zalando research

## Learning a multi-variate distribution: RealNVP [Dinh et al., 2016]

### Idea

- for all bijections f it holds

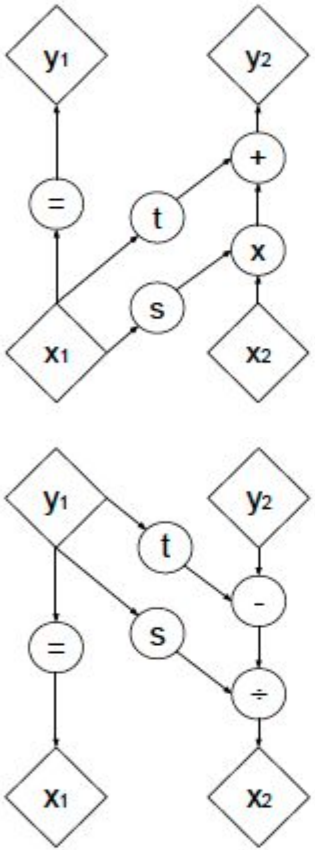
$$p_X(x) = p_Z(f(x)) \left| \det \left( \frac{\partial f(x)}{\partial x^T} \right) \right|$$

- affine coupling layer

$$\begin{cases} y_{1:d} & = x_{1:d} \\ y_{d+1:D} & = x_{d+1:D} \odot \exp\big(s(x_{1:d})\big) + t(x_{1:d}) \end{cases}$$

- inverse of coupling layer

$$\Leftrightarrow \begin{cases} x_{1:d} & = y_{1:d} \\ x_{d+1:D} & = \big(y_{d+1:D} - t(y_{1:d})\big) \odot \exp\big(-s(y_{1:d})\big) \end{cases}$$



Forward propagation

Inverse propagation

## RealNVP [Dinh et al., 2016]

### Why is this great?

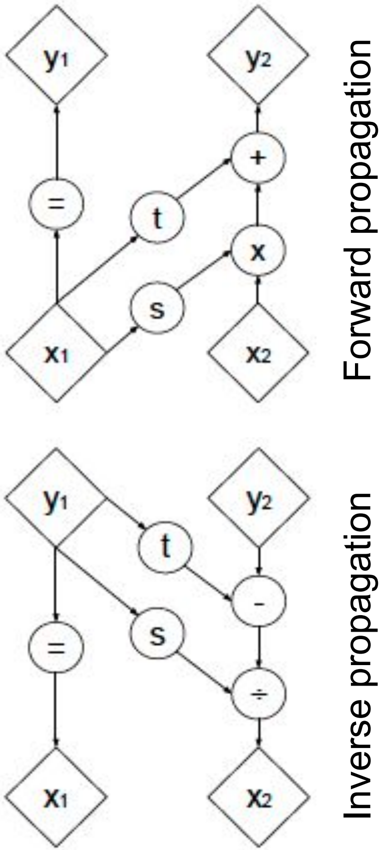
- affine coupling layer

$$\begin{cases} y_{1:d} &= x_{1:d} \\ y_{d+1:D} &= x_{d+1:D} \odot \exp\big(s(x_{1:d})\big) + t(x_{1:d}) \end{cases}$$

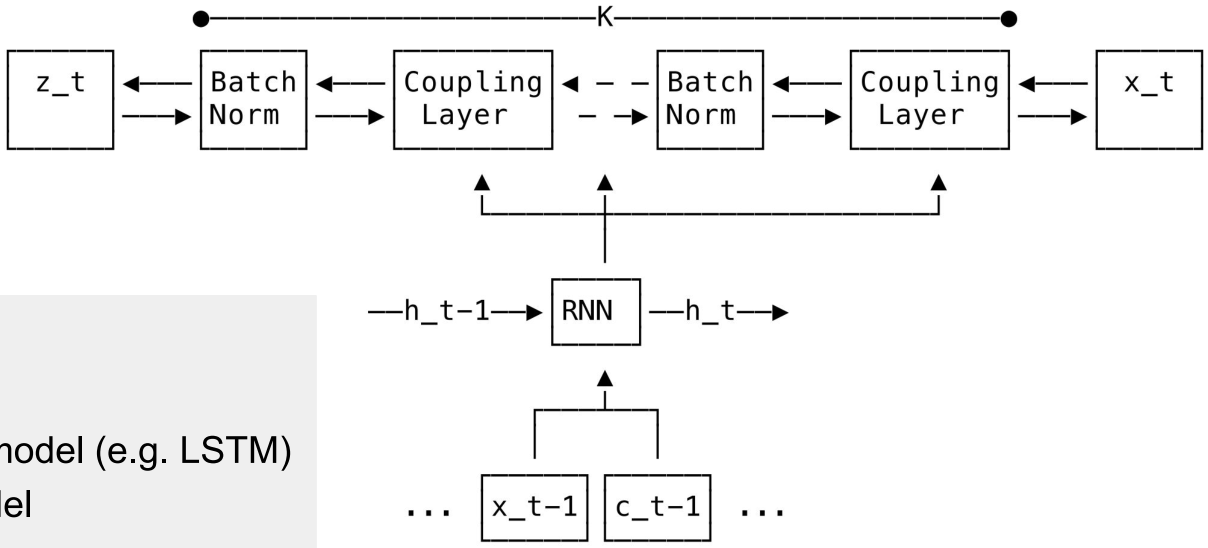
- has Jacobian

$$\frac{\partial y}{\partial x^T} = \left[ \begin{array}{cc} \mathbb{I}_d & 0 \\ \frac{\partial y_{d+1:D}}{\partial x^T_{1:d}} & \operatorname{diag}\left(\exp\left[s\left(x_{1:d}\right)\right]\right) \end{array} \right]$$

- *efficient to calculate!*



## Conditioned Temporal Flows [Rasul et al., 2020]




### Key Idea

- combine autoregressive model (e.g. LSTM) with conditional Flow model

### Advantages

- learns distribution model from data
- able to model interactions of time-series

# Performance on real-world data sets

*Table 1.* Test set $\text{CRPS}_{\text{sum}}$ comparison (lower is better) of models from (Salinas et al., 2019) and our models `GRU-Real-NVP`, `GRU-MAF` and `Transformer-MAF`. The *two* best methods are in bold where the mean and standard errors are obtained by re-running each method three times.

| Data set | Vec-LSTM ind-scaling | Vec-LSTM lowrank-Copula | GP scaling | GP Copula | GRU Real-NVP | GRU MAF | Transformer MAF |
|---|---|---|---|---|---|---|---|
| Exchange | 0.008±0.001 | 0.007±0.000 | 0.009±0.000 | 0.007±0.000 | **0.0064**±0.001 | **0.005**±0.001 | **0.005**±0.001 |
| Solar | 0.391±0.017 | 0.319±0.011 | 0.368±0.012 | 0.337±0.024 | 0.331±0.02 | **0.315**±0.023 | **0.301**±0.014 |
| Electricity | 0.025±0.001 | 0.064±0.008 | 0.022±0.000 | 0.024±0.002 | 0.024±0.001 | **0.0208**±0.000 | **0.0207**±0.000 |
| Traffic | 0.087±0.041 | 0.103±0.006 | 0.079±0.000 | 0.078±0.002 | 0.078±0.001 | **0.069**±0.002 | **0.056**±0.001 |
| Taxi | 0.506±0.005 | 0.326±0.007 | 0.183±0.395 | 0.208±0.183 | **0.175**±0.001 | **0.161**±0.002 | 0.179±0.002 |
| Wikipedia | 0.133±0.002 | 0.241±0.033 | 1.483±1.034 | 0.086±0.004 | 0.078±0.001 | **0.067**±0.001 | **0.063**±0.003 |

# CONCLUSION

Generative Image Modeling

- supervised losses + deep learning work excellent on many tasks
- extension of conditional GANs allow conditional image creation for real-time fashion exploration
- GANs allow generation of data even in face of lacking training data

*Time-Series Modeling*

- Combining Flows with AR models yields powerful multi-variate time-series models



zalando research

# Thanks to...

**Computer Vision**

- Christian Bracher
- Sebastian Heinz
- Siavash Haghiri
- Roland Vollgraf

**Intelligent Control**

- Kashif Rasul
- Ingmar Schuster
- Saboor Sheikh
- Calvin Seward

**Image Creation**

- Nikolay Jetchev
- Gokhan Yildirim

**Natural Language Processing**

- Josip Krapac

## THANK YOU!



zalando research

# We’re hiring!

**jobs.zalando.com**