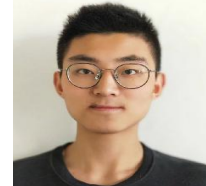
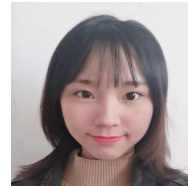
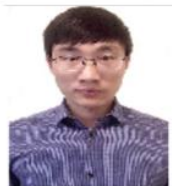


What We Believe

Something important in 2021

Dong Wang

2021/01/04



We need some belief

- Research resembles climbing a mountain
- Before someone raises his lights, nobody knows his height
- Tradeoff between safety and reward
- Here we only consider risk-takers...



We need some belief

- We are in dark, with limited knowledge
- The only thing we can resort to is a belief
- Belief is not bigotry



What kind of belief is good?

- It should be simple enough
- It should be general enough
- It should be intuitively correct



艾萨克·牛顿勋爵士埋葬在这里。他用数学方法揭示了行星的运动、彗星的轨迹、海洋和潮汐；他孜孜不倦地探究光线和颜色的本性；他解释了自然和万物的规律。他用自己的哲学证明了上帝的威严。

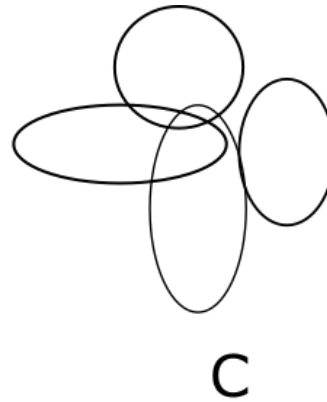
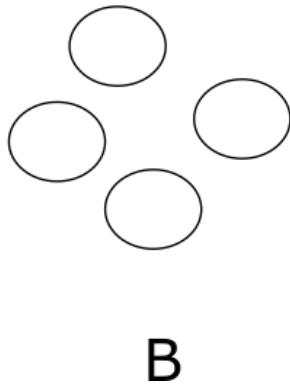
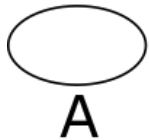
Our belief 1
Uncertainty pervasively exists

Uncertainty should be employed

- It should be modeled rather than ignored
 - For generation, it is the core.
 - For decoding, it is the knowledge.
- More data reduces the uncertainty in parameter, but not the uncertainty in prediction
- Powerful model reduces the uncertainty in prediction/inference/decoding

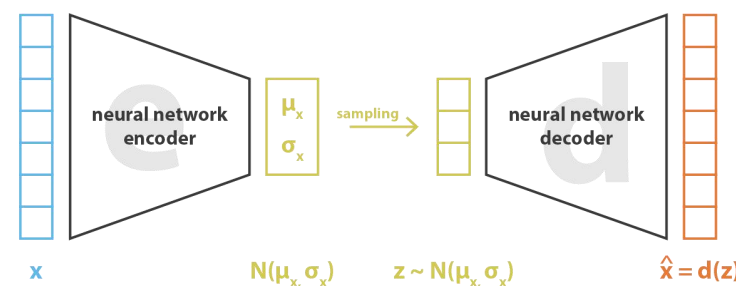
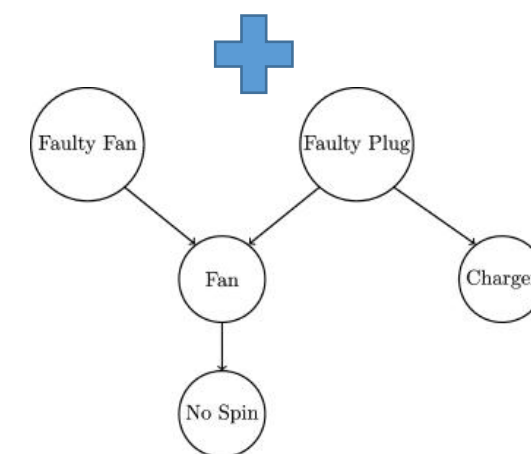
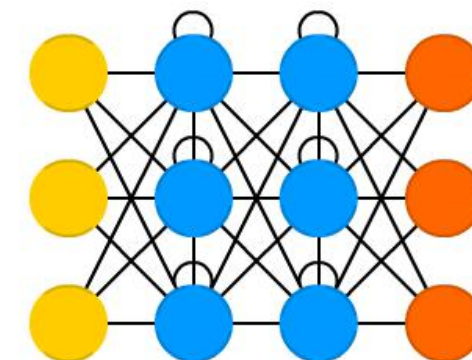
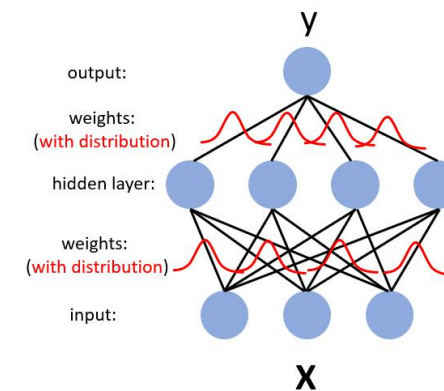
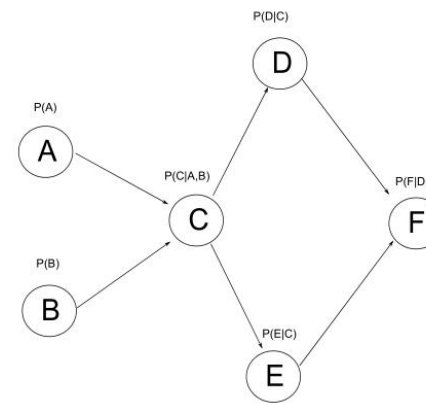
Uncertainty is transmitted

- Uncertainty transmission from one stage to another stage is typical in many data processing tasks.
- Inference is important



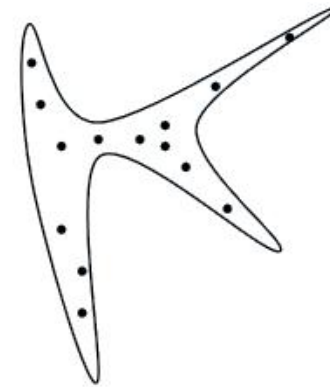
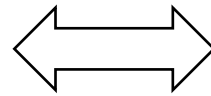
Uncertainty in Bayes and neural net

- Bayes net
 - Simple conditional and complex structure
- Neural net
 - Bayes neural net
- VAE
- Neural-Bayes



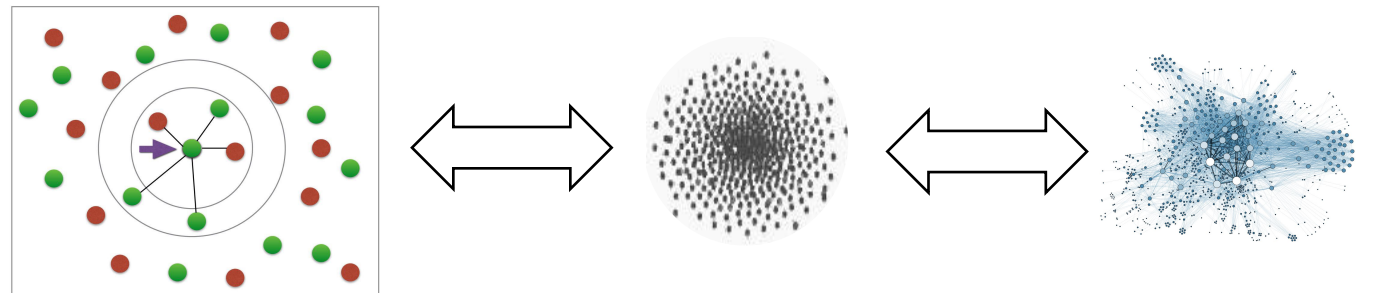
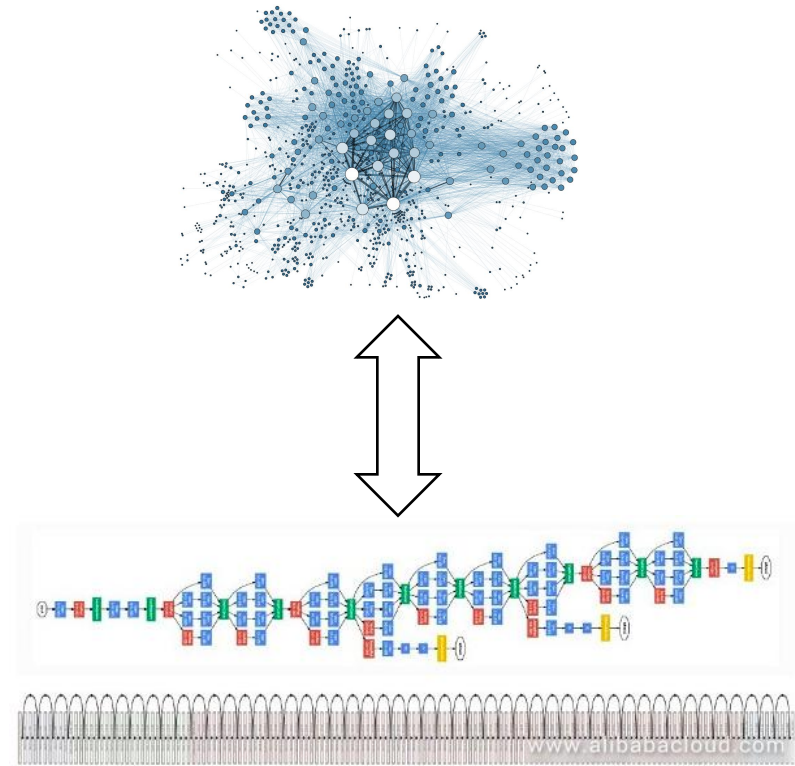
$$\text{loss} = \|x - \hat{x}\|^2 + \text{KL}[N(\mu_x, \sigma_x), N(0, I)] = \|x - d(z)\|^2 + \text{KL}[N(\mu_x, \sigma_x), N(0, I)]$$

Our belief 2
Data distribution can be manipulated



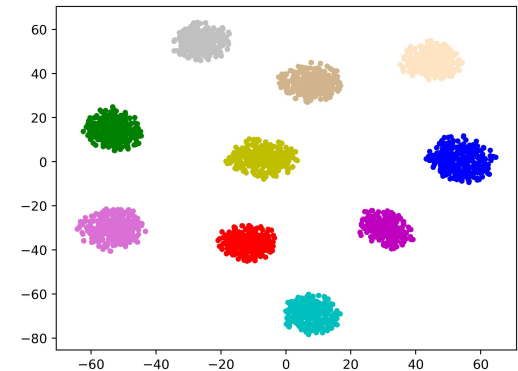
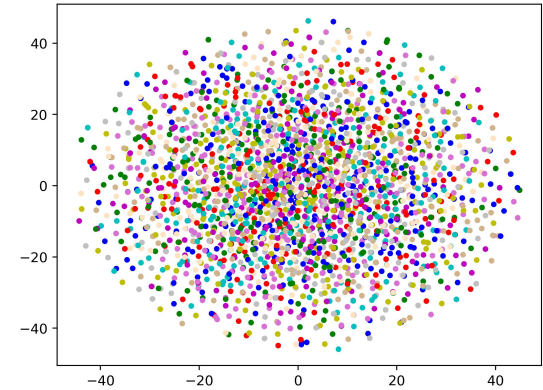
A new view for ML

- Data in nature is complex.
- Conventional approach design complex models to meet complex data
- We can map complex data to simple data, therein simple model being employed
- Understanding the data, then perform task



Many questions

- To what extent the distribution can be manipulated
 - e.g., can we squeeze the main entropy to a subspace? How to make different dimensions comparable?
- How to integrate discriminative information
- How to reduce dimensionality?



Many questions

- How to design more powerful invertible model?

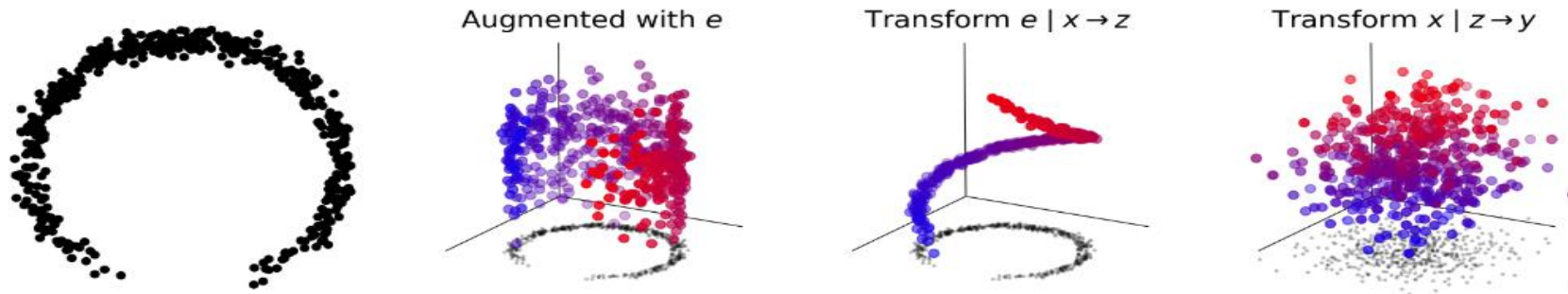
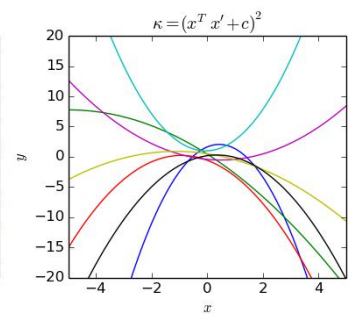
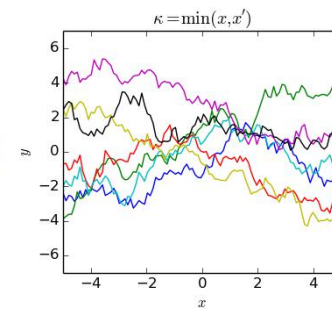
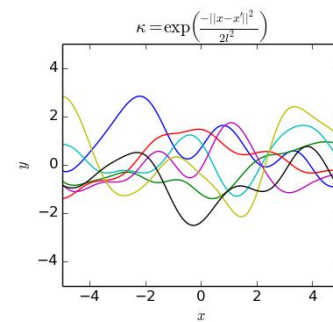


Figure 1. Transforming data x (left) via augmented normalizing flows: Black dots and blue dots are marginal and joint data points, respectively. *First step:* augment the data x with an independent noise e . *Second step:* transform the augmented data e conditioned on x into z . *Third step:* transform the original data x conditioned on z into y , resulting in a Gaussianized joint distribution of (y, z)

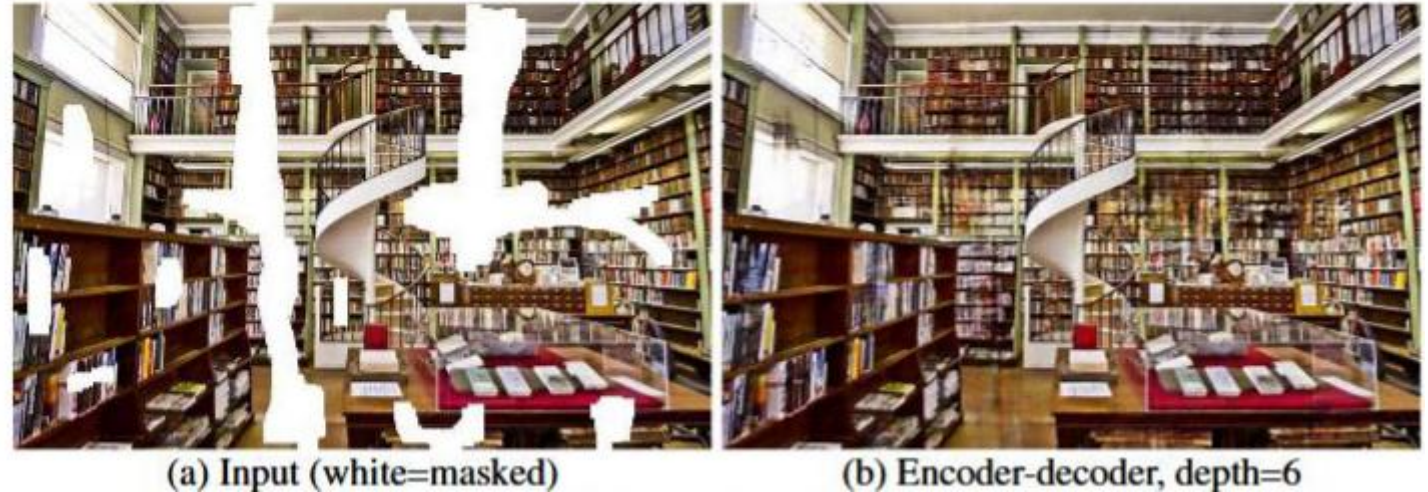
Our belief 3

Everything you need is a good prior

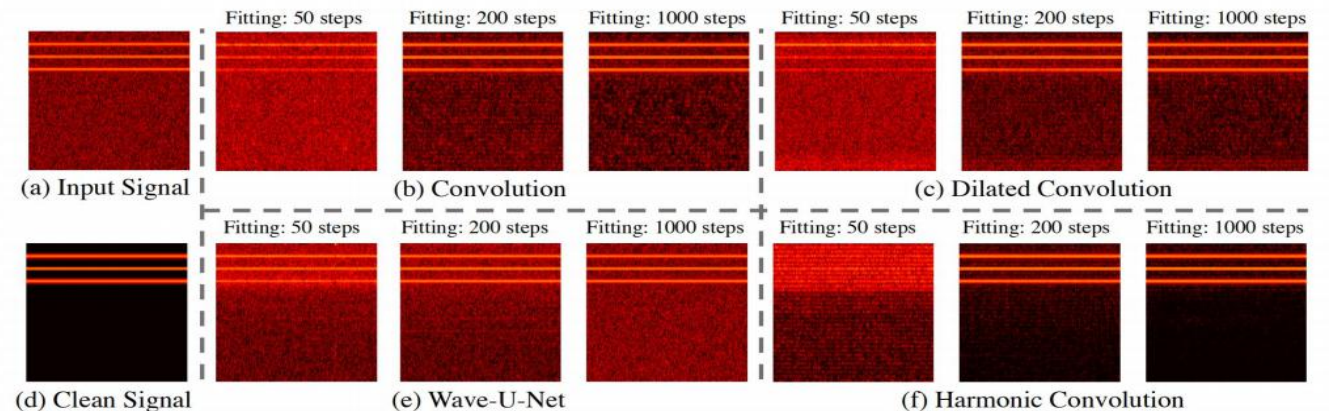


Prior is important

- Prior and data are two knowledge resource, though the latter is much more emphasized rightnow.
- A suitable prior is important



Ulyanov D, Vedaldi A, Lempitsky V. Deep image prior, CVPR, 2018.



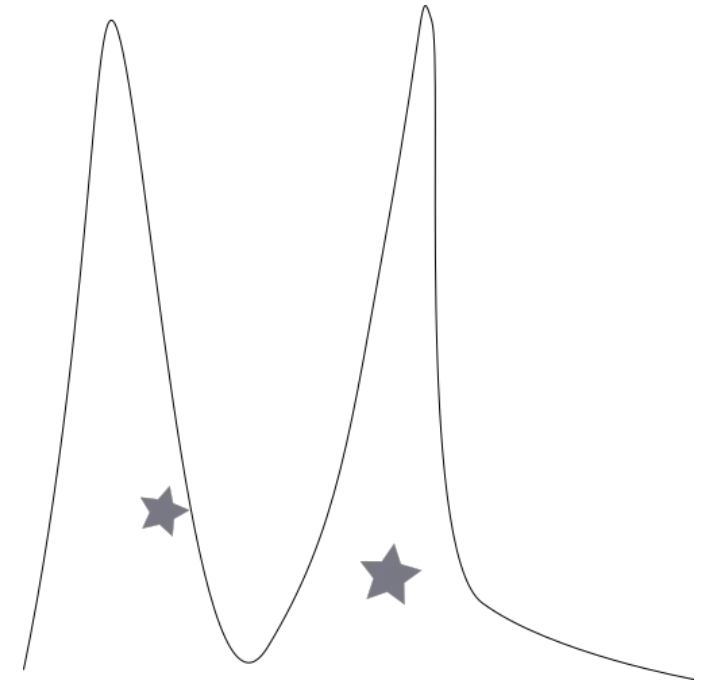
DEEP AUDIO PRIORS EMERGE FROM HARMONIC CONVOLUTIONAL NETWORKS, ICLR 2020

Implicit priors we are often using

- Classification with cross entropy loss
 - Empirical cross entropy
 - discrete, finite, categorical classes
- Regression with mean square error
 - Empirical cross (differential) entropy
 - Constrained by the Gaussian form plus limited (and identical) variation
- t-SNE with KL divergence
 - Constrained by the Gaussian form in the data space and Student's t in the projection space
-

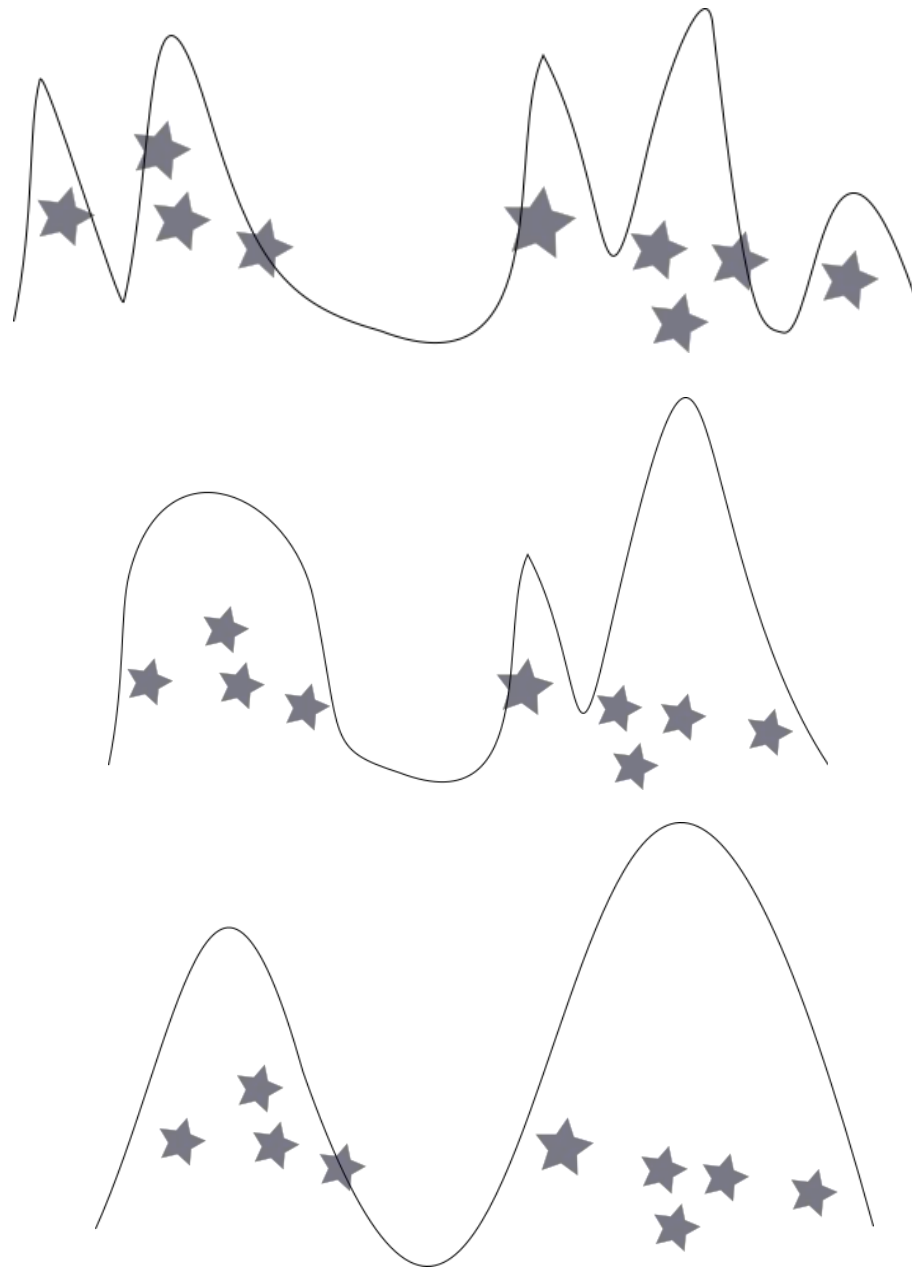
Prior is not only important, but essential for unbounded models

- Flow is an unbounded model
- In theory, the likelihood for the data can be infinitely high
- Reason: there is no prior how the data is distributed, and so the ML model is unbounded, if the model is strong enough.



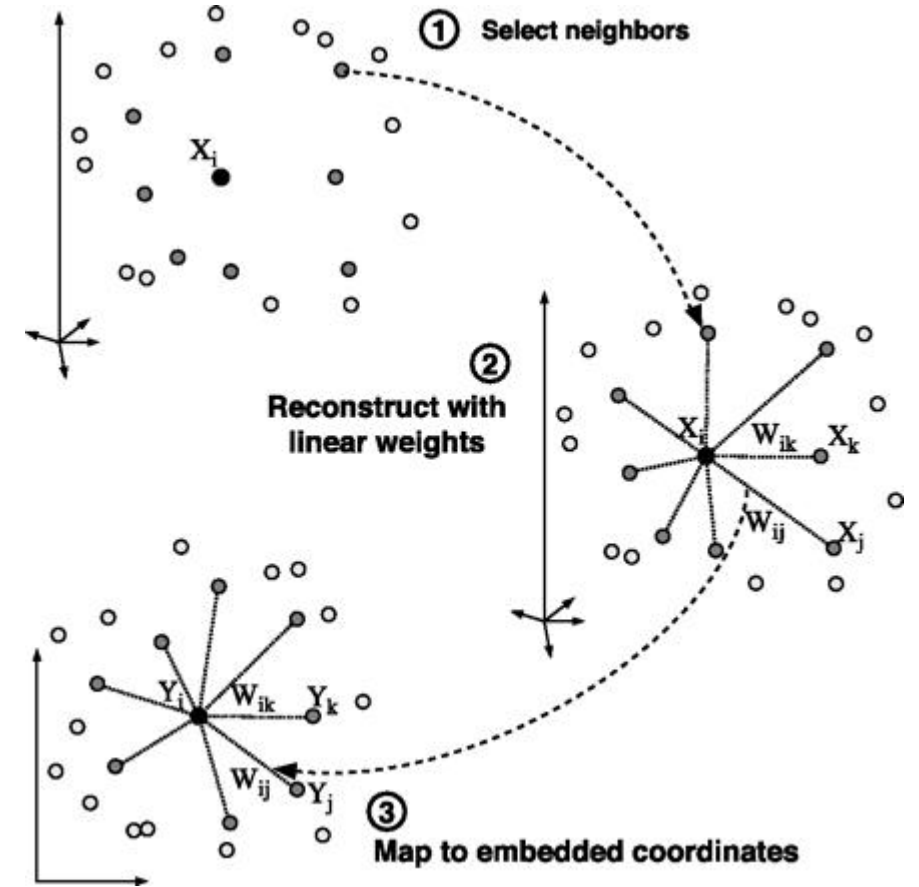
What is a good prior?

- On which scale the landscape is smooth
- More than Occam's razor --
Lantian's razor, Yunqi's razor, Bird's razor, airplane's razor...
- Occam's razor seems to find the largest patterns



Define prior on different scales

- The razor cannot control the structure
- We define the scale by neighbourhood size
- To solve non-linear and non-parametric, using linear estimation for local areas



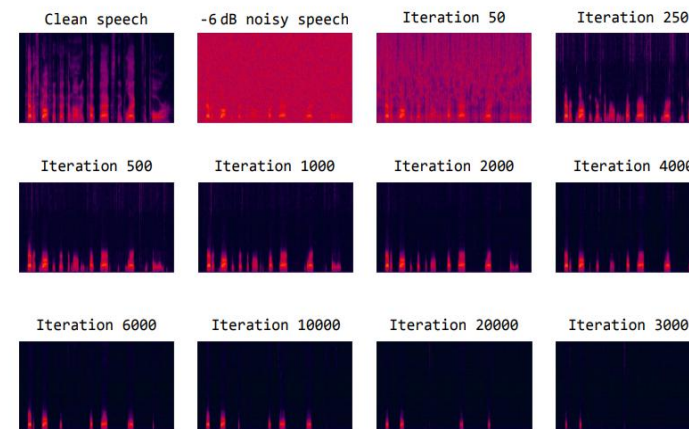
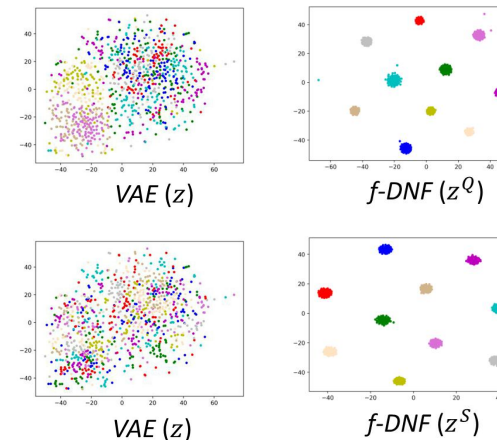
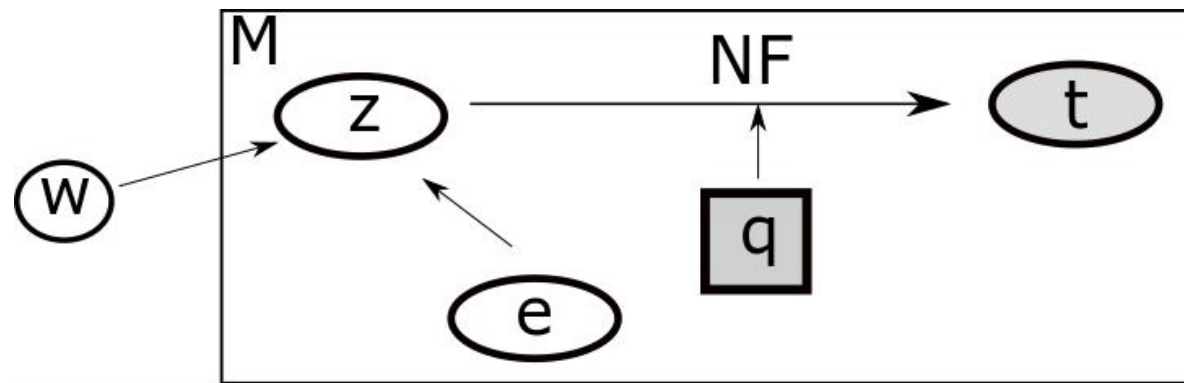
Sampling for non-parametric priors

- Sampling in the local linear space, to fill the 'holes' at the desired scale.
- With infinite samples, the ML will not diverge.



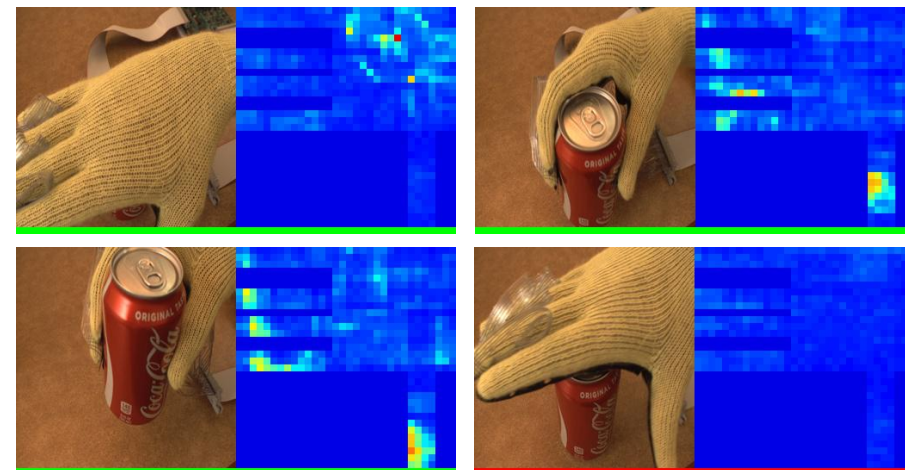
Focus in 2021: Understand speech

- Deep generative model for speech recognition and speaker recognition
- Deep generative model for speech factorization
- Speech enhancement with speech prior

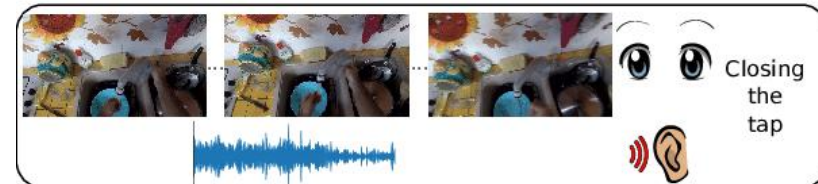


Focus in 2021: Multi-modality processing

- Bayes modality combination (product expert, Gaussian latent space...)
- Visual and tactile fusion
- Audio-visual fusion
- Thermal map and speech fusion



What is the person doing?



Focus in 2021: Promote popular AI

- Popularization AI education



The image shows a presentation slide with the following elements:

- Logo:** A stylized 'SLT' logo with 'GROUPING' written vertically to its right and 'Center for Speech and Language Technologies' written below it.
- Section Header:** '本期特邀老师' (Special Guest Teacher of This Issue).
- Guest Information:** A grey box containing the name '马少平' (Ma Shaoping), his titles '清华大学教授, 博士生导师' (Professor at Tsinghua University, PhD Supervisor), and '中国人工智能学会副监事长' (Vice Chairman of the Chinese Association of Artificial Intelligence).
- Portrait:** A small circular photo of Ma Shaoping.
- Question Card:** A blue-bordered card with a white robot illustration. The text reads '什么是图灵测试?' (What is the Turing Test?), '问' (Question) on the left, and 'Q1' on the right.
- Answer Arrow:** A blue arrow pointing downwards with the character '答' (Answer) in the center.



- Happy new year!