VFlow: More Expressive Generative Flows with Variational Data Augmentation

Jianfei Chen[*†], Cheng Lu[*], Biqi Chenli[*†], Jun Zhu[*†], and Tian Tian[*†]

[*]Tsinghua University
[†]RealAI

ICML 2020

Generative Flows

Invertible transformation

$$\mathbf{x} \xleftrightarrow{\mathbf{f}_1} \mathbf{h}_1 \xleftrightarrow{\mathbf{f}_2} \mathbf{h}_2 \cdots \xleftrightarrow{\mathbf{f}_L} \boldsymbol{\epsilon}.$$

Density

$$\log p(\mathbf{x}; \boldsymbol{\theta}) = \log p_{\boldsymbol{\epsilon}}(\boldsymbol{\epsilon}) + \log \left| \frac{\partial \boldsymbol{\epsilon}}{\partial \mathbf{x}} \right|.$$

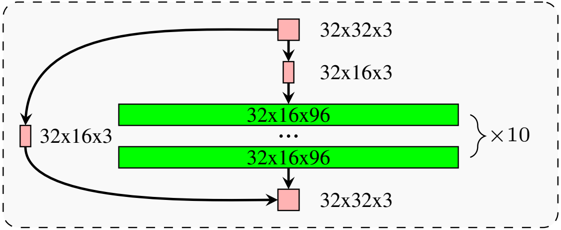
Invertible transformation step

- Forward $\mathbf{f}_l(\mathbf{h}_{l-1}; \boldsymbol{\theta})$
- Inverse $\mathbf{f}_l^{-1}(\mathbf{h}_l; \boldsymbol{\theta})$
- Jacobian $\left| \frac{\partial \mathbf{f}_l}{\partial \mathbf{h}_{l-1}} \right|$

- Tractable Jacobian (affine coupling layer)
- Free-form (FFJORD, residual flows)

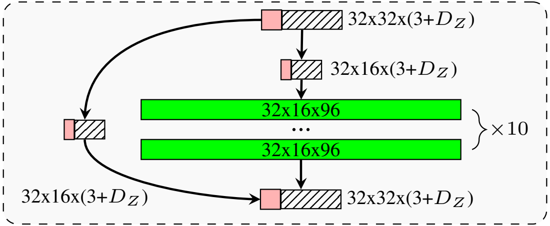
Bottleneck Problem

$$\mathbf{x} \overset{\mathbf{f}_1}{\longleftrightarrow} \mathbf{h}_1 \overset{\mathbf{f}_2}{\longleftrightarrow} \mathbf{h}_2 \cdots \overset{\mathbf{f}_L}{\longleftrightarrow} \boldsymbol{\epsilon}.$$



VFlow

$$(\mathbf{x}, \color{red}{\mathbf{z}}) \overset{\mathbf{f}_1}{\longleftrightarrow} \mathbf{h}_1 \overset{\mathbf{f}_2}{\longleftrightarrow} \mathbf{h}_2 \cdots \overset{\mathbf{f}_L}{\longleftrightarrow} \boldsymbol{\epsilon}.$$



Little computational and parameter overhead

Estimation

Let $\hat{p}(\mathbf{x})$ be the empirical data distribution, and

- $p_X(\mathbf{x}; \boldsymbol{\theta}_X)$ be the density of a vanilla flow,
- $p_{XZ}(\mathbf{x}, \mathbf{z}; \boldsymbol{\theta}_{XZ})$ be the density of an augmented flow, and
- $p_{XZ}(\mathbf{x}; \boldsymbol{\theta}_{XZ}) = \int_{\mathbf{z}} p_{XZ}(\mathbf{x}, \mathbf{z}; \boldsymbol{\theta}_{XZ}) \mathrm{d}\mathbf{z}$ be the marginal distribution
- Evidence lower bound (ELBO) $\log p_{XZ}(\mathbf{x}; \boldsymbol{\theta}_{XZ}) \geq \mathbb{E}_{q(\mathbf{z}|\mathbf{x};\boldsymbol{\phi})}[\log p(\mathbf{x}, \mathbf{z}; \boldsymbol{\theta}_{XZ}) - \log q(\mathbf{z}|\mathbf{x}; \boldsymbol{\phi})]$;
- $q(\mathbf{z}|\mathbf{x}; \boldsymbol{\phi})$ can be implemented with another conditional flow.

Maximum likelihood estimation for Flow

$$\max_{\boldsymbol{\theta}_X} \mathbb{E}_{\hat{p}(\mathbf{x})}[\log p(\mathbf{x}; \boldsymbol{\theta}_X)]$$

Maximum ELBO estimation for VFlow

$$\max_{\boldsymbol{\theta}_{XZ}, \boldsymbol{\phi}} \mathbb{E}_{\hat{p}(\mathbf{x})q(\mathbf{z}|\mathbf{x};\boldsymbol{\phi})}[\log p(\mathbf{x}, \mathbf{z}; \boldsymbol{\theta}_{XZ}) - \log q(\mathbf{z}|\mathbf{x}; \boldsymbol{\phi})],$$

where $\hat{p}(\mathbf{x})q(\mathbf{z}|\mathbf{x}; \boldsymbol{\phi})$ is the augmented data distribution.

Theoretical Guarantee

Theorem

$$\underbrace{\max_{\boldsymbol{\theta}_X} \mathbb{E}_{\hat{p}(\mathbf{x})}[\log p(\mathbf{x};\boldsymbol{\theta}_X)]}_{\textit{MLE of a Flow}} \leq \underbrace{\max_{\boldsymbol{\theta}_{XZ},\boldsymbol{\phi}} \mathbb{E}_{\hat{p}(\mathbf{x})q(\mathbf{z}|\mathbf{x};\boldsymbol{\phi})}[\log p(\mathbf{x},\mathbf{z};\boldsymbol{\theta}_{XZ}) - \log q(\mathbf{z}|\mathbf{x};\boldsymbol{\phi})]}_{\textit{Max ELBO estimation of a VFlow}}.$$

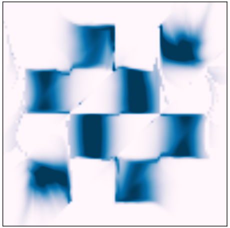
For any flow $\log p(\mathbf{x};\boldsymbol{\theta}_X)$

- Construct a special VFlow $\log p(\mathbf{x},\mathbf{z};\boldsymbol{\theta}_{XZ}) = \log p(\mathbf{x};\boldsymbol{\theta}_X) + \log p_{\boldsymbol{\epsilon}}(\mathbf{z})$;
- and a special variational distribution $\log q(\mathbf{z}|\mathbf{x};\boldsymbol{\phi}) = \log p_{\boldsymbol{\epsilon}}(\mathbf{z})$;
- The variational bound $\log p(\mathbf{x},\mathbf{z};\boldsymbol{\theta}_{XZ}) - \log q(\mathbf{z}|\mathbf{x};\boldsymbol{\phi}) = \log p(\mathbf{x};\boldsymbol{\theta}_X)$.

Toy Data

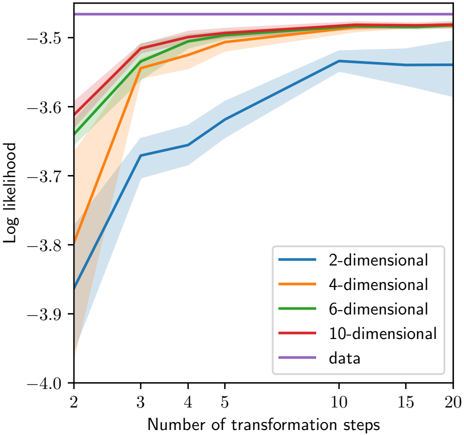
(a) Data (-3.47)

(b) 3-step, 10-dim VFlow (-3.51)

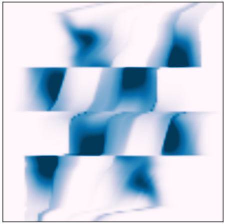


(c) 3-step Glow (-3.66)

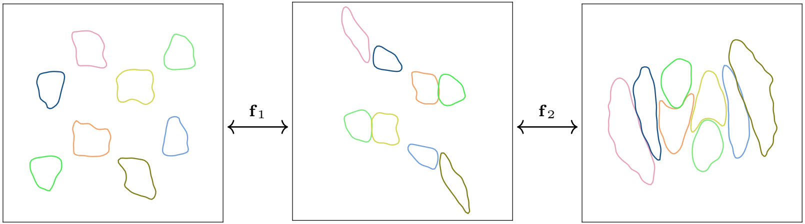
(d) 20-step Glow (-3.52)



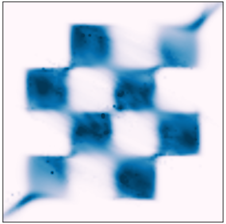
Visualization of Flow and VFlow



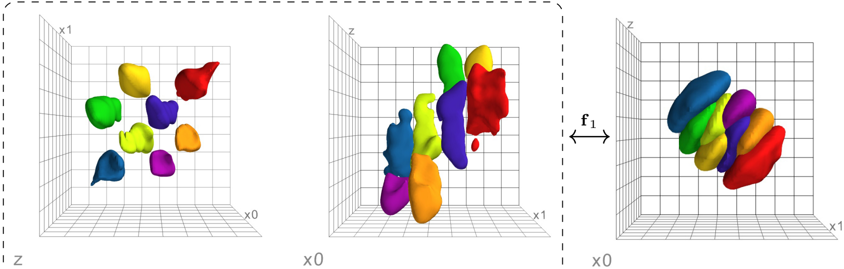
Model density (-3.80)



$\mathbf{x}$ $\mathbf{h}_1$ $\boldsymbol{\epsilon}$



Model density (-3.69)



$\mathbf{x}$ (top view) $\mathbf{x}$ (front view) $\boldsymbol{\epsilon}$

Image Density Modeling on CIFAR10

Model	bpd
Glow	3.35
FFJORD	3.40
Residual Flow	3.28
MintNet	3.32
Flow++	3.08
VFlow	**2.98**

Model	bpd	Parameters	Hidden channels	Steps
3-channel Flow++	3.08	31.4M	96	10
6-channel VFlow	**2.98**	37.8M	96	10
6-channel VFlow	3.03	16.5M	64	10
6-channel VFlow	3.08	**11.9M**	56	10

Conclusion

VFlow

- tackles the bottleneck problem of generative flows;
- can be easily combined with existing flows;
- fits in a variational data augmentation framework;
- is theoretically superior than vanilla flows;
- achieves state-of-the-art result for image density modeling.