底层视觉任务中的

模型驱动

数据驱动

孟德宇

西安交通大学

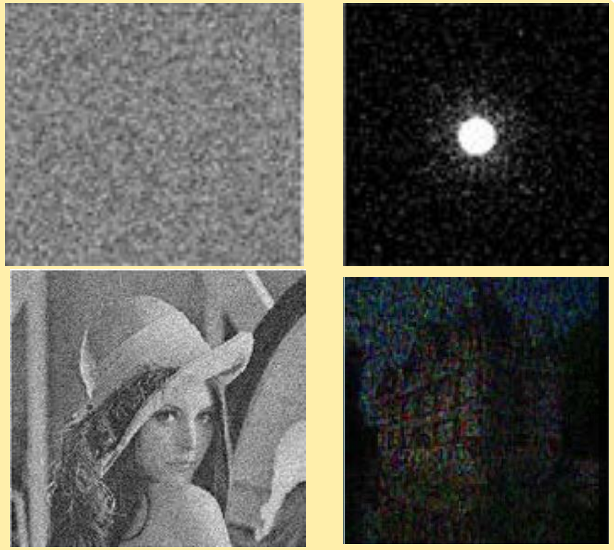
dymeng@mail.xjtu.edu.cn

http://gr.xjtu.edu.cn/web/dymeng

Image Restoration problem

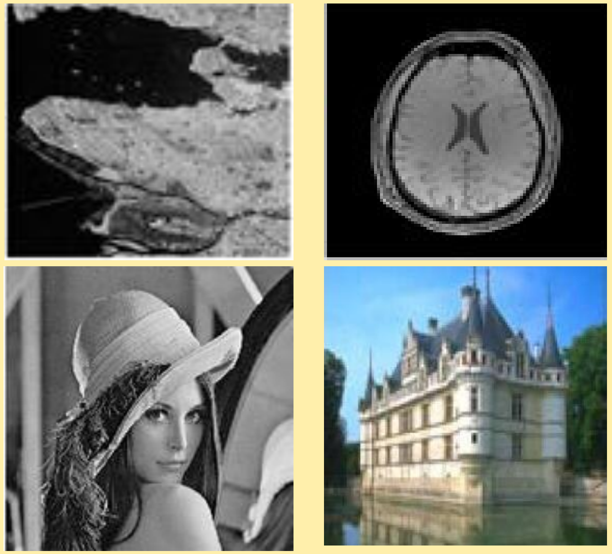
The inverse problem: $X = Y(W) + \varepsilon$

Observation X



- Image super-resolution
- Image denoising
- Image deblurring
- ……

Recovery Y(W)



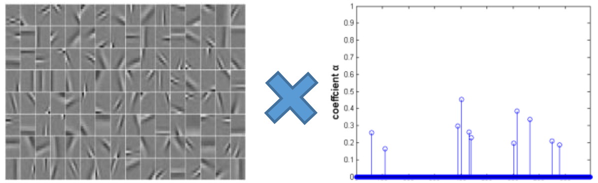
Model-driven Methodology

2010-2012

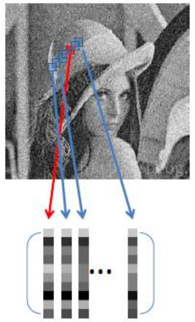
$$\arg\min_{Z,\theta} L_\theta(Y - Z) + R(Z) + R(\theta)$$

Low rank

Sparsity



Non-local Similarity



Model-driven Methodology: Generative Understanding

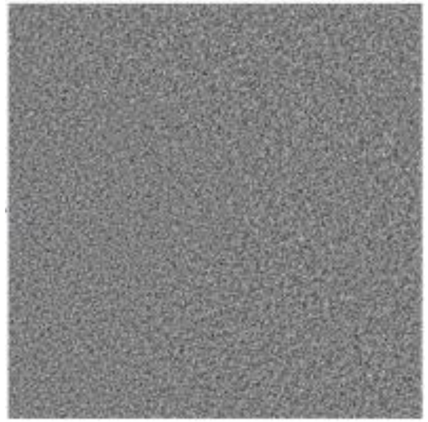
$$\arg\min_{Z,\theta} L_\theta(Y - Z) + R(Z) + R(\theta)$$

Y

=

Z

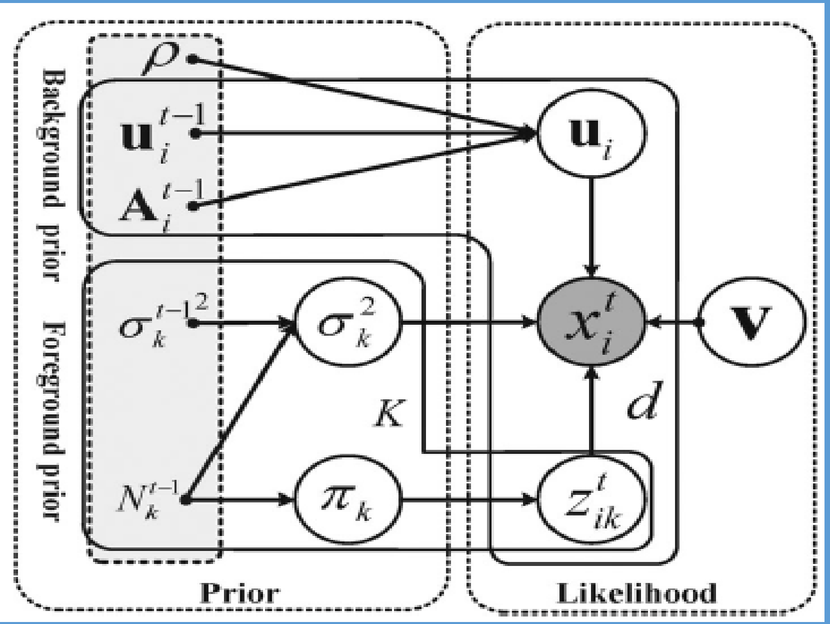
+



E

$z \sim p(z); e \sim p(e;\theta)$ → $p(z,e|y) \sim \text{likelihood}(y|z,e)p(z)p(e)$

Model-driven Methodology: Generative Understanding



$$p(\Pi,\Sigma,\mathrm{v},\mathrm{U}|\mathrm{x}^t,\Theta^{t-1}) \propto p(\mathrm{x}^t|\Pi,\Sigma,\mathrm{v},\mathrm{U})p(\Sigma|\Theta^{t-1})p(\Pi|\Theta^{t-1})p(\mathrm{U}|\Theta^{t-1})p(\mathrm{v})$$

Yong, Meng, Zuo, Zhang, TPAMI, 2018

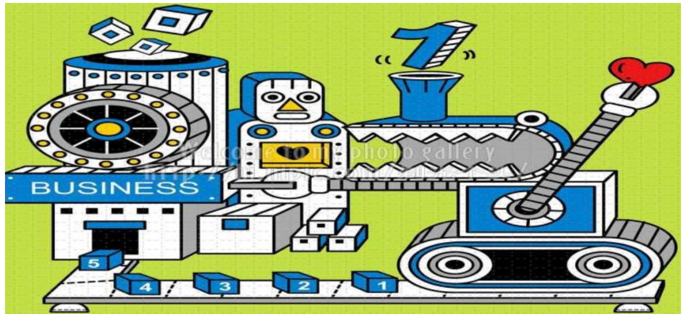
$$z\sim p(z);\ e\sim p(e;\theta) \quad \Rightarrow \quad p(z,e|y)\sim \text{likelihood}(y|z,e)p(z)p(e)$$

Model-driven Methodology

$$\arg\min_{Z,\theta} L_\theta(Y - Z) + R(Z) + R(\theta)$$

$$\arg\max_{Z,\theta}\ p(Z, \theta | Y)$$

Y $\rightarrow$ $Z* = \text{Algorithm}(Y)$

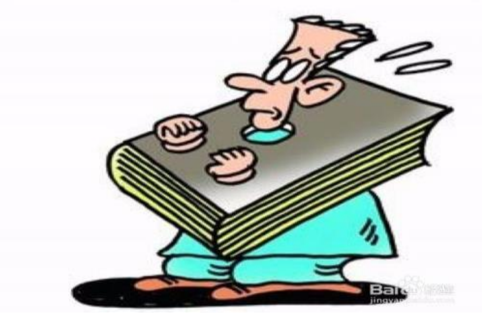


Model-driven Methodology

文雅之风

- 可解释性
- 小样本前提
- 模型针对性

- 预测速度慢
- 先验假设依赖
- 算法设计难

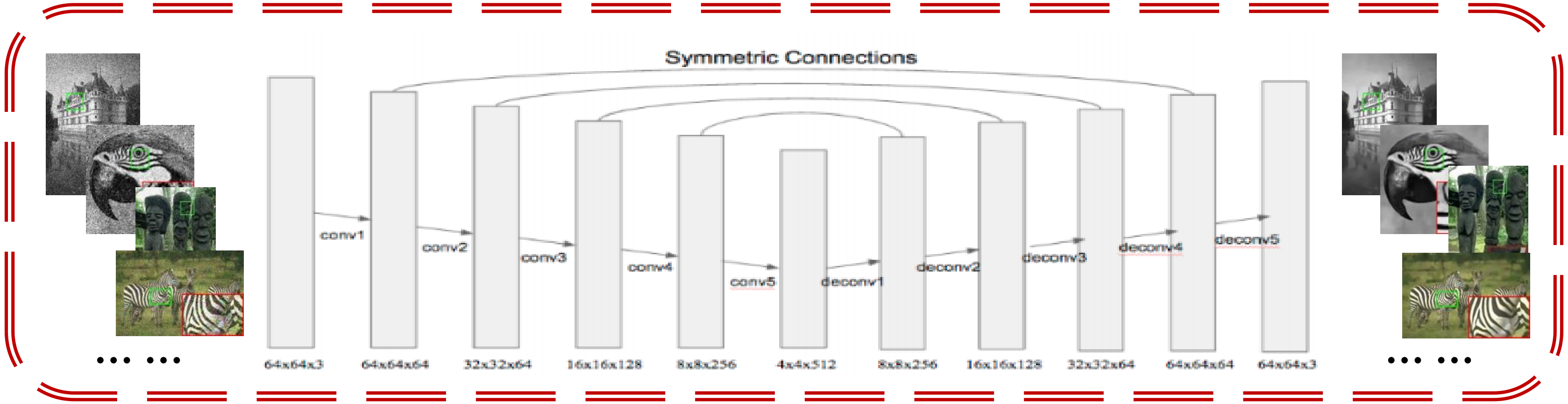


书呆子气

Data-driven Methodology: Learn Clean Image

2010-2012

$$\arg\min_{W} \left\|Z - \text{Network}_W(Y)\right\|_2$$

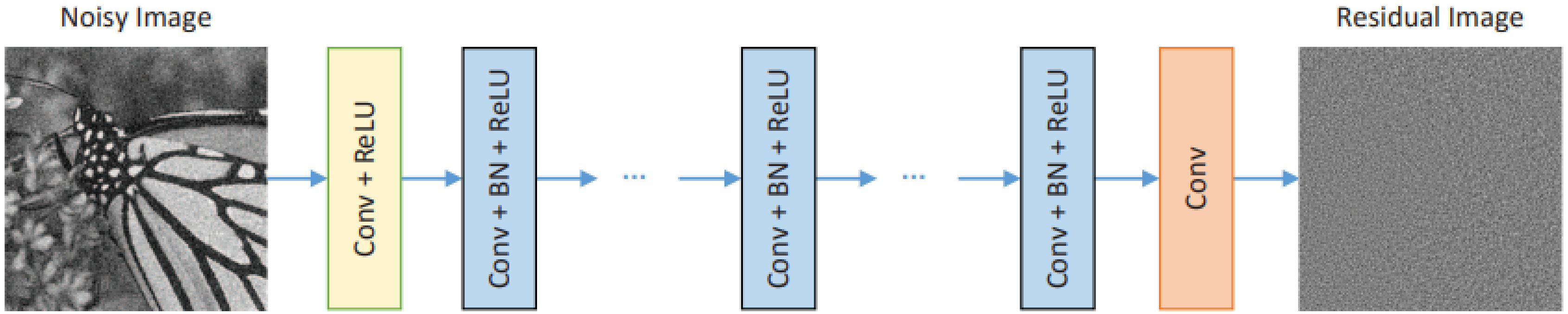


$$Y \longrightarrow Z_* = \text{Network}_{W_*}(Y)$$

Data-driven Methodology: Learn Noise

2010-2012

$$\arg\min_{W} ||E - \text{Network}_W(Y)||_2$$

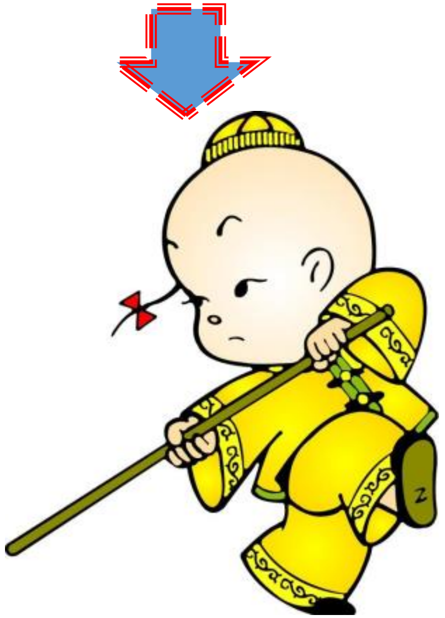


Residual Network

$$Y \quad \Rightarrow \quad E* = \text{Network}_{W*}(Y)$$

Zhang, Zuo, Chen, Meng, Zhang, TIP, 2017

Data-driven Methodology: Learn Noise



勇武之气

- 预测速度快
- 无须先验假设
- 无须设计算法

- 解释性差
- 依赖标注样本
- 结构难设计

鲁莽粗暴

Model-driven or/and Data-driven?

文雅之风

OR

勇武之气

- 可解释性
- 小样本前提
- 模型针对性

- 预测速度慢
- 先验假设依赖
- 算法设计难

- 预测速度快
- 无须先验假设
- 无须设计算法

- 解释性差
- 依赖标注样本
- 结构难设计

数模结合第一式

外练筋骨皮

模型

网络

（跳出模仿）在师傅指导下，进行有方向性的训练

（跳出有监督）在模型启发下，指导网络数据合理的梯度下降方向对网络参数进行训练

Semi-supervised Deep Learning

Unsupervised Data

$$\arg\min_{W} \; L(X - \mathrm{Network}_W(X)) + R(W)$$

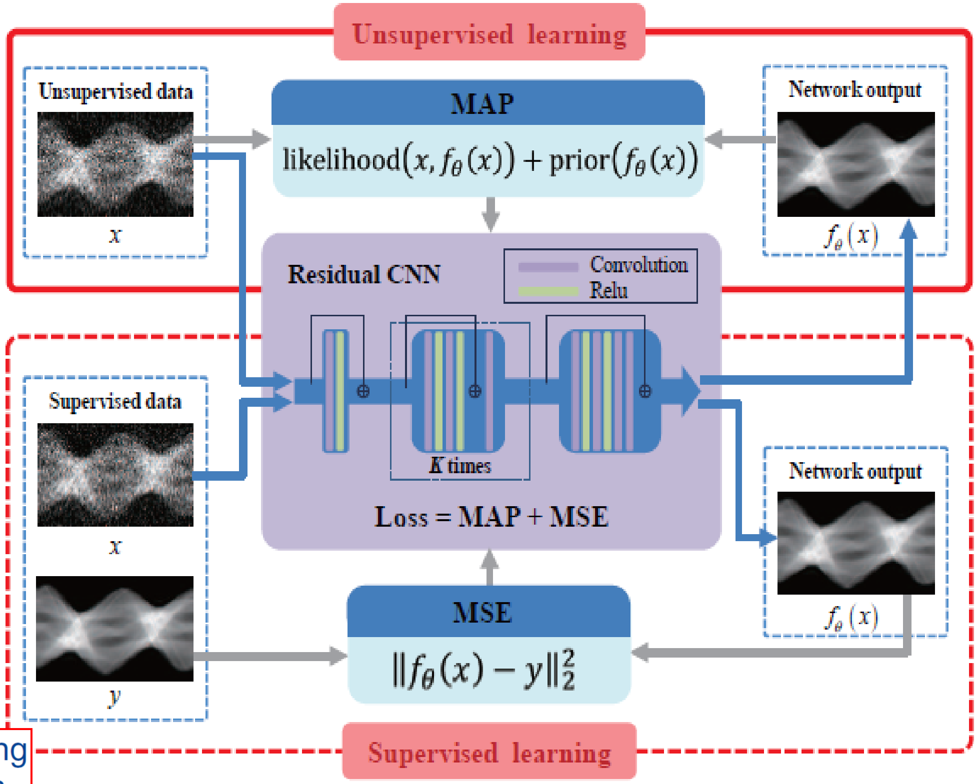
+

$$\arg\min_{W} \; \left|\left|Y - \mathrm{Network}_W(X)\right|\right|_2$$

Supervised Data

Attempt 1

Unsupervised/Semi-supervised Deep Learning for Low-dose CT Enhancement



Probabilistic understanding has been presented in our TMI18 paper

Robust Low-Dose CT Sinogram Preprocessing via Exploiting Noise-Generating Mechanism

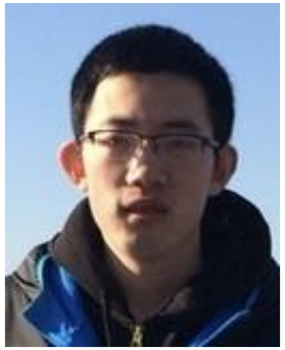
Qi Xie, Dong Zeng, Qian Zhao, Deyu Meng, Zongben Xu, Zhengrong Liang, and Jianhua Ma

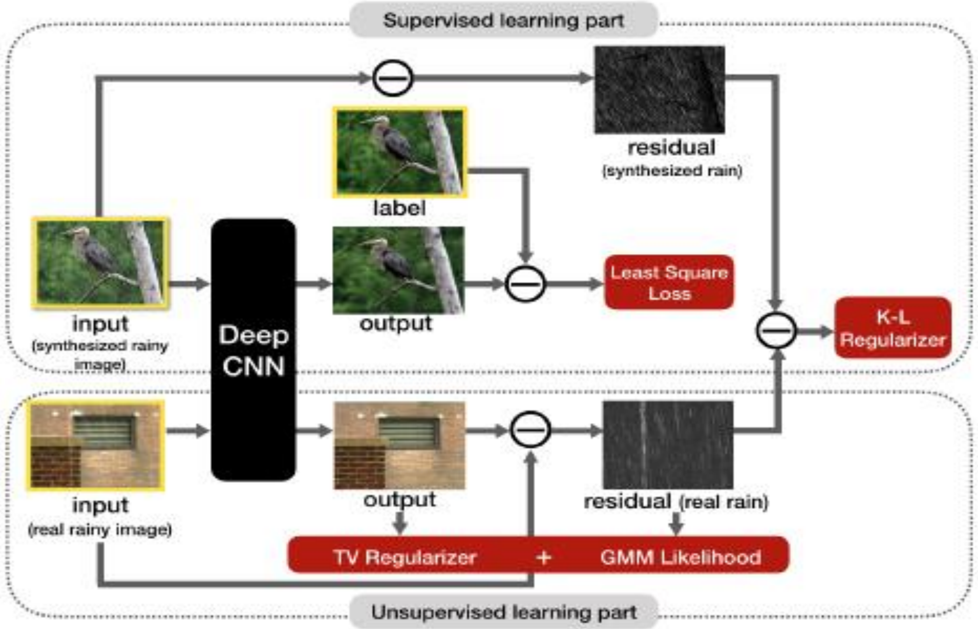
Geng, Deng, Zhao, Xie, Zeng, Ma, Zuo, Meng, arxiv, 2018

Attempt 2

Semi-supervised Transfer Learning for Image Rain Removal

Wei Wei[1,2], Deyu Meng[1]*, Qian Zhao[1], Zongben Xu[1], Ying Wu[2]
[1]School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China
[2]Department of Electrical and Computer Engineering, Northwestern University, IL, USA





Probabilistic understanding has been presented in our ICCV17 paper

Should We Encode Rain Streaks in Video as Deterministic or Stochastic?

Wei Wei[1], Lixuan Yi[1], Qi Xie[1], Qian Zhao[1,2], Deyu Meng[1,2,*], Zongben Xu[1,2]

Wei, Meng, Zhao, Xu,Wu, CVPR, 2019

数模结合第二式

内练一口气

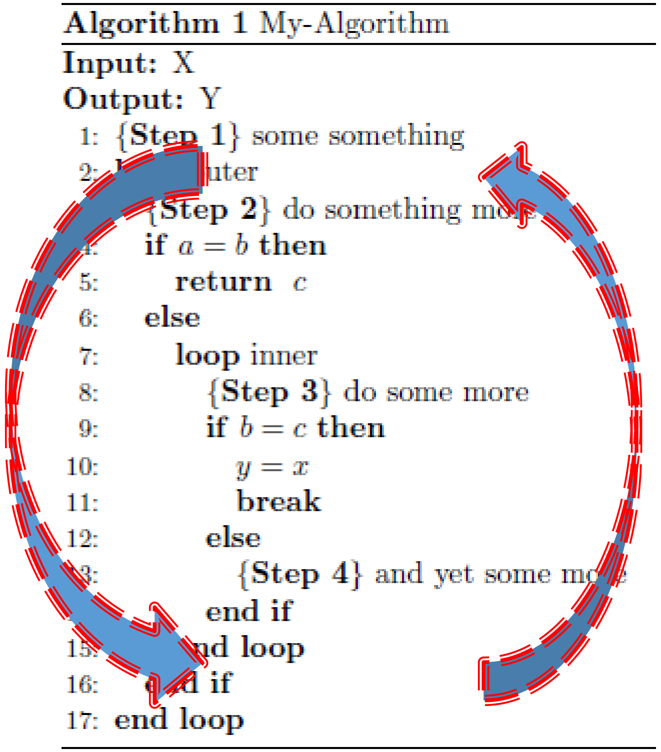
模型

网络

修炼内在，与环境一致

构建网络结构，与模型求解算法一致

Deep Unfolding



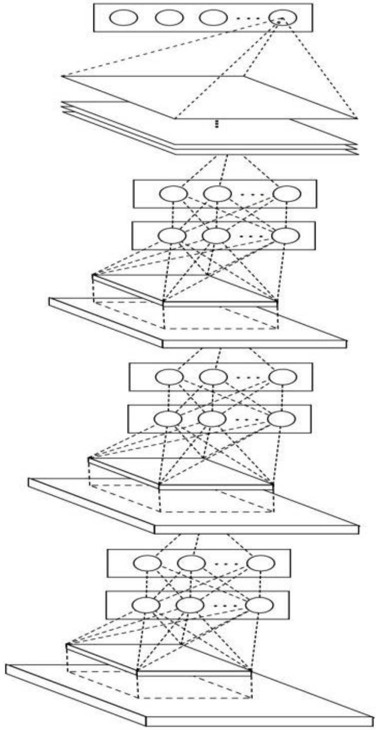
Algorithm 1 My-Algorithm

Input: X

Output: Y

{**Step 1**} some something
uter
{**Step 2**} do something mo
if $a = b$ **then**
return c
else
loop inner
{**Step 3**} do some more
if $b = c$ **then**
$y = x$
break
else
{**Step 4**} and yet some mo
end if
nd loop
d if
end loop

Algorithm



Network

Deep Unfolding

- Field of Experts (FoE) energy: $\min_{u} \sum_{i=1}^{N_k} \sum_{p=1}^{N} \rho_i((K_i u)_p)$

  Roth and Black, IJCV 2009

- Half-quadratic splitting $\min_{u,z} \sum_{i=1}^{N_k} \sum_{p=1}^{N} \left(\rho_i(z_p) + \frac{\beta}{2}(z - K_i u)_p \right)$

  Schmidt and Roth, CVPR 2014

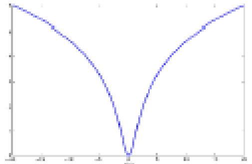
- Primal dual proximal scheme

  Wang et al., NIPS 2016

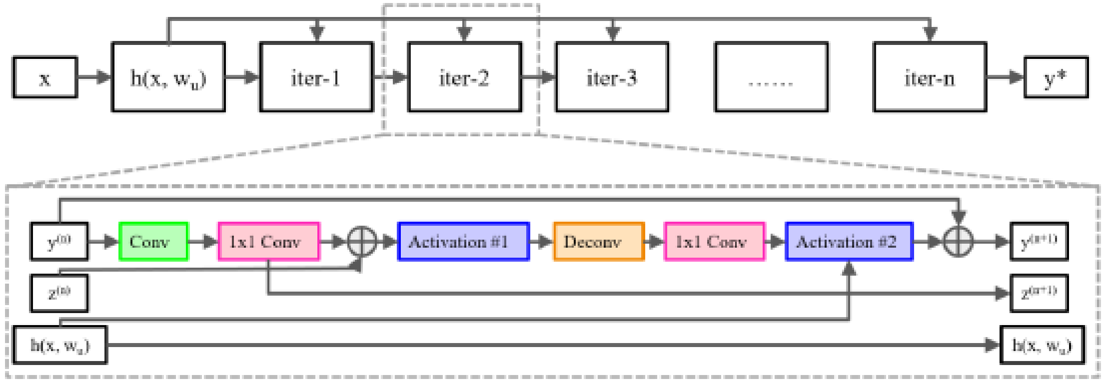
- Alternating direction method

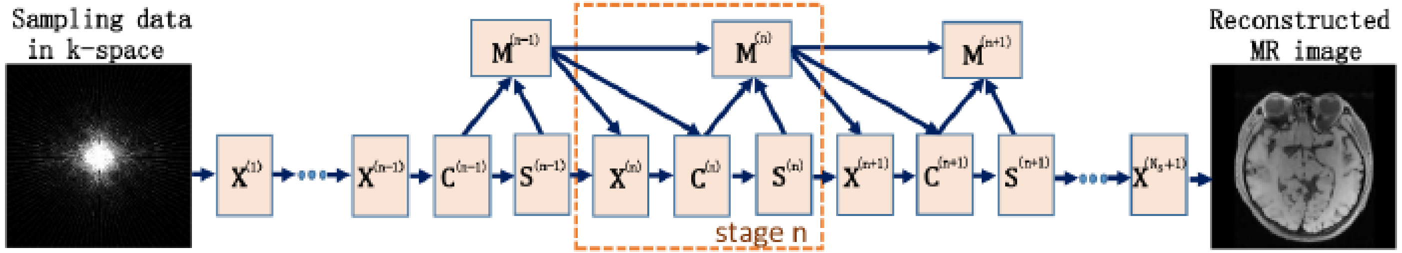
  Yang et al., NIPS 2016

$$\boxed{\boxed{\psi(u) = \nabla_u \mathcal{D}(u)} \quad \boxed{\rho'(z) \;=\; \phi(z)}}$$

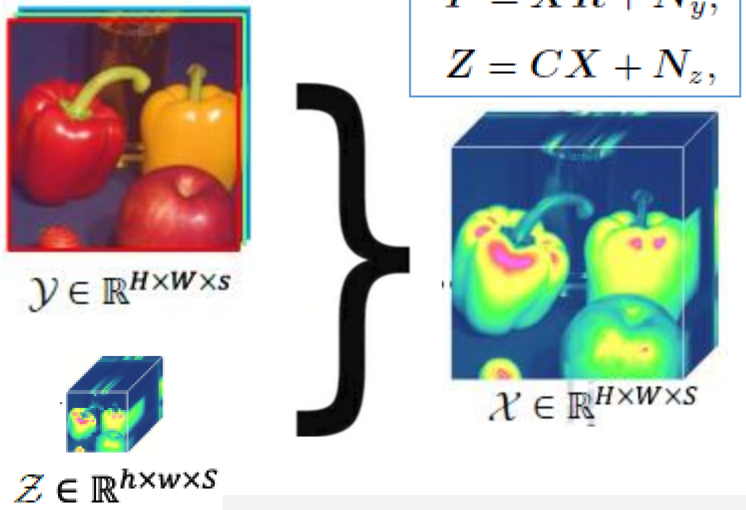


$\rho(z) = \log(1 + z^2)$

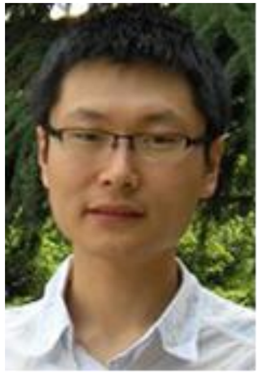




Our attempt: Hyper-spectral Image Fusion



$$Y = XR + N_y,$$
$$Z = CX + N_z,$$



$$\min_{\hat{X}} \|Y - \hat{X}R\|_F^2 + \lambda \|Z - C\hat{X}\|_F^2 + \beta P(\hat{X})$$

$$\Downarrow$$

$$\min_{\hat{X},\hat{Y},A,B,C} \|YA + \hat{Y}B - \hat{X}\|_F^2 + \alpha \|C(\hat{X}) - Z\|_F^2 + \beta P(\hat{Y})$$

- $\hat{X}_{k+\frac{1}{2}} = YA_{k+1} + \hat{Y}_k B_k$
- $E_k = C_\theta\left(\hat{X}_{k+\frac{1}{2}}\right) - Z$
- $\hat{X}_{k+1} = \hat{X}_{k+\frac{1}{2}} - \frac{\lambda}{2+L} C'_\theta(E_k)$
- $\hat{Y}_{k+1} = f_{P,B_k}(\hat{X}_{k+1} - YA_k)$

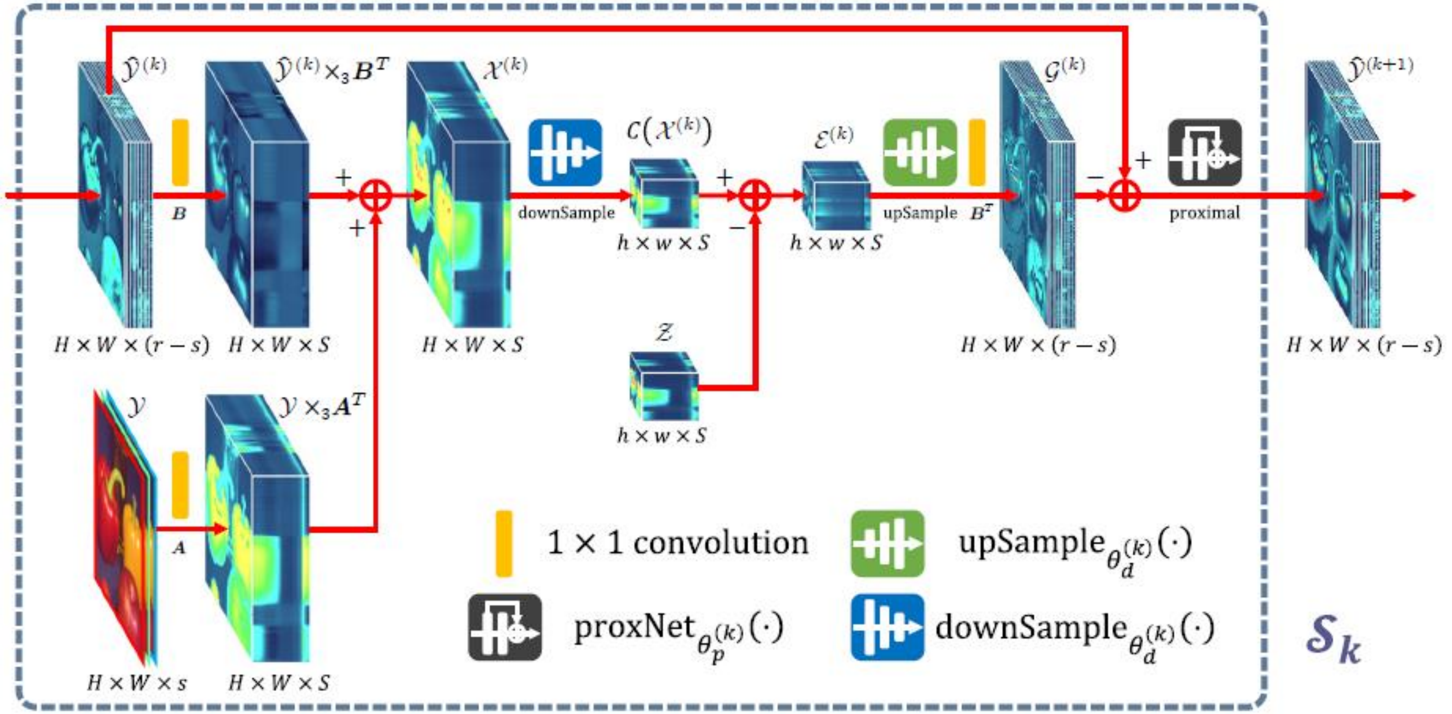
Xie, Zhao, Meng, Xu, CVPR, 2019

Our attempt: Hyper-spectral Image Fusion

Iterative optimization algorithm	Network design
For $k = 1{:}K$ do:	In stage $k = 1{:}K$ of the network do:
$\mathbf{X}^{(k)} = \mathbf{Y}\mathbf{A} + \hat{\mathbf{Y}}^{(k)}\mathbf{B}$ ⟶	$\mathcal{X}^{(k)} = \mathcal{Y} \times_3 \mathbf{A}^T + \hat{\mathcal{Y}}^{(k)} \times_3 \mathbf{B}^T$
$\mathbf{E}^{(k)} = \mathbf{C}\mathbf{X}^{(k)} - \mathbf{Z}$ ⟶	$\mathcal{E}^{(k)} = \mathrm{downSample}_{\theta_d^{(k)}}\left(\mathcal{X}^{(k)}\right) - \mathcal{Z}$
$\mathbf{G}^{(k)} = \eta\mathbf{C}^T\mathbf{E}^{(k)}\mathbf{B}^T$ ⟶	$\mathcal{G}^{(k)} = \eta \cdot \mathrm{upSample}_{\theta_u^{(k)}}\left(\mathcal{E}^{(k)}\right) \times_3 \mathbf{B}$
$\hat{\mathbf{Y}}^{(k+1)} = \mathrm{prox}_{\lambda\eta}\left(\hat{\mathbf{Y}}^{(k)} - \mathbf{G}^{(k)}\right)$ ⟶	$\hat{\mathcal{Y}}^{(k+1)} = \mathrm{proxNet}_{\theta_p^{(k)}}\left(\hat{\mathcal{Y}}^{(k)} -\right.$

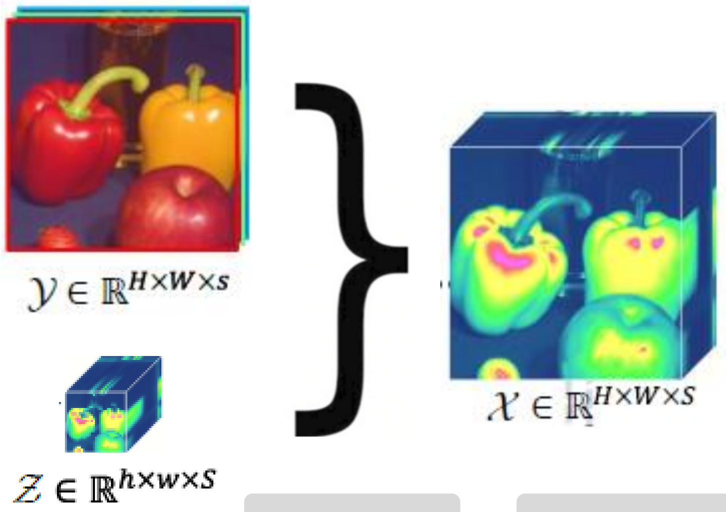
$$Y = XR + N_y,$$

$$Z = CX + N_z,$$



Xie, Zhao, Meng, Xu, CVPR, 2019

Our attempt: Hyper-spectral Image Fusion



$$Y = XR + N_y,$$
$$Z = CX + N_z,$$

Table 1. Average performance of the competing methods over 12 testing samples of CAVE data set with respect to 5 PQIs.

	ResNet	MHF-net with (K, L)			
		(4, 9)	(7, 5)	(10, 4)	(13, 2)
PSNR	32.25	36.15	36.61	36.85	**37.23**
SAM	19.093	9.206	8.636	7.587	**7.298**
ERGA	141.28	92.94	88.56	86.53	**81.87**
SSIM	0.865	0.948	0.955	0.960	**0.962**
FSIM	0.966	0.974	0.975	0.975	**0.976**

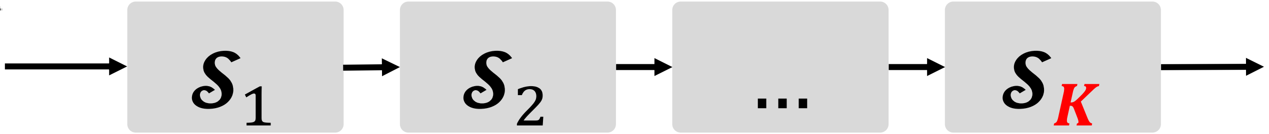
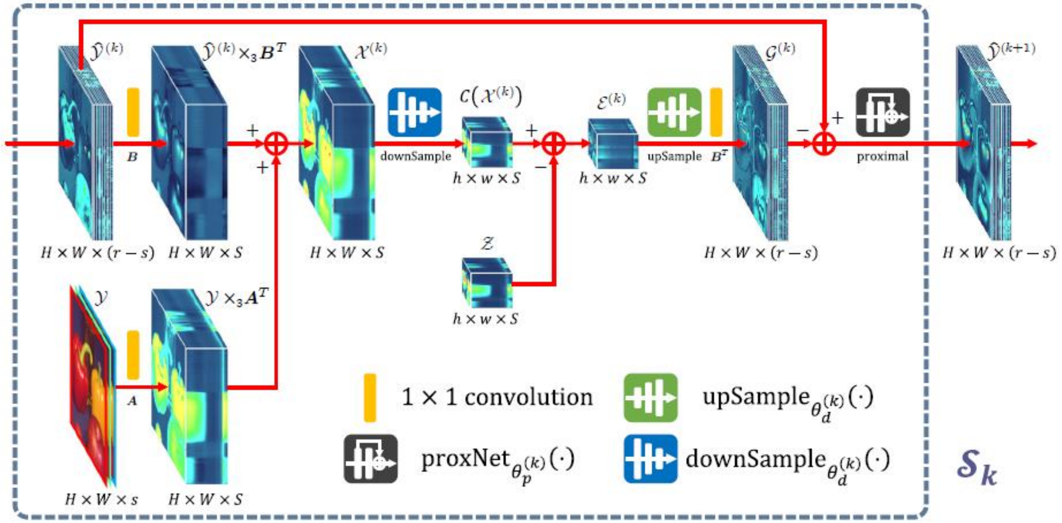


Table 3. Average performance of the competing methods over 16 testing samples of Chikusei data set with respect to 5 PQIs.

	PSNR	SAM	ERGAS	SSIM	FSIM
FUSE	26.59	7.92	272.43	0.718	0.860
ICCV15	27.77	3.98	178.14	0.779	0.870
GLP-HS	28.85	4.17	163.60	0.796	0.903
SFIM-HS	28.50	4.22	167.85	0.793	0.900
GSA	27.08	5.39	238.63	0.673	0.835
CNMF	28.78	3.84	173.41	0.780	0.898
M-FUSE	24.85	6.62	282.02	0.642	0.849
SASFM	24.93	7.95	369.35	0.636	0.845
PNN	24.30	4.26	157.49	0.717	0.807
3D-CNN	30.51	**3.02**	129.11	0.869	0.933
ResNet	29.35	3.69	144.12	0.866	0.930
MHF-net	**32.26**	**3.02**	**109.55**	**0.890**	**0.946**



Xie, Zhao, Meng, Xu, CVPR, 2019

More Attempts

Single image deraining

$$\min_{\Theta} \mathcal{L}(\Theta) = \| \mathcal{X} - \mathcal{H}^{\perp} \circ \mathcal{B} - \mathcal{H} \circ \mathcal{F} - R \|_F^2 + \lambda \| \mathcal{F} \|_{TV} + \alpha \| \mathcal{H} \|_{3DTV} + \beta \| \mathcal{H} \|_1 + b \sum_{k=1}^{K} \sum_{s=1}^{n_k} \| \mathcal{M}_{ks} \|_1$$

$$s.t. \quad \mathcal{B} = \mathrm{Fold}(U^T V)$$

$$\mathcal{R} = \sum_{k=1}^{K} \sum_{s=1}^{s_k} D_{ks} \otimes \mathcal{M}_{ks}, \quad \| D_{ks} \|_F^2 \leq 1,$$

Li, et al., CVPR, 2018

Low-dose CT Enhancement

$$\max_{Y,Q,b} \sum_{i=1}^{N} \left(-\frac{(P_i - Q_i)^2}{2\sigma^2} + Q_i \ln(I_{0i}) - Q_i Y_i - \ln(Q_i!) - I_{0i} e^{-Y_i} \right) - \frac{2}{b} \| D_2 Y \|_{1/2}^{1/2} - 2M \ln(b).$$

Xie, et al., TMI, 2018

Leision Detection

$$\min_{\Theta} - \sum_{i,j} \log \mathcal{N}((\mathbf{X}_1 - \mathbf{U}\mathbf{V}_1^T)_{ij} | 0, \sigma_1^2) - \sum_{i,\ell} \log \sum_{k=1}^{K} \pi_k \mathcal{N}((\mathbf{X}_2 - \mathbf{U}\mathbf{V}_2^T)_{i\ell} | 0, \sigma_k^2)$$

Wang, et al., TMI, 2019

数模结合第三式

万法归宗

网络

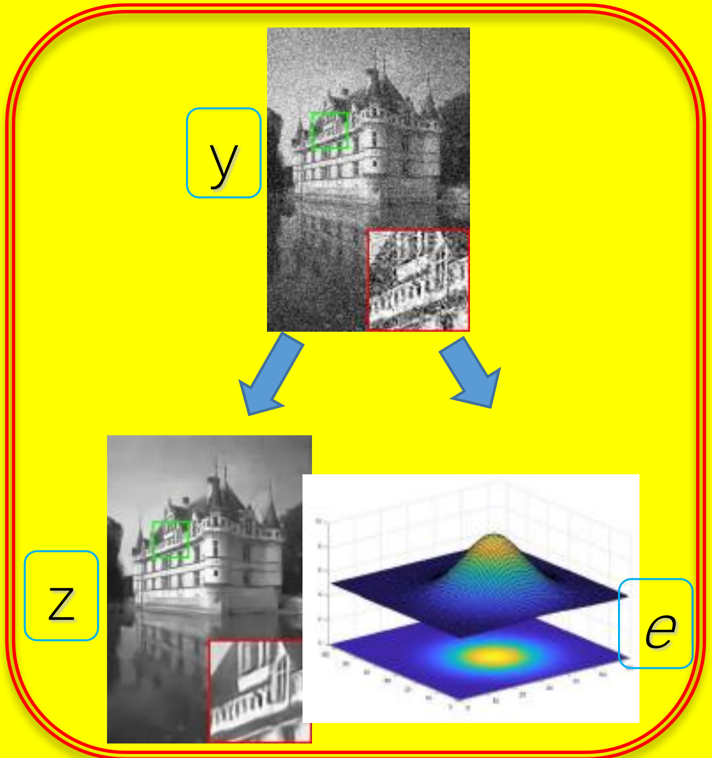
模型

艰难探索，获得自然规律

通过对参数化网络进行数据学习，获得共性推断模型

Model-driven Methodology: What we want?

$$q(z, \sigma^2 | y)$$



Model-driven Methodology: Generative Understanding

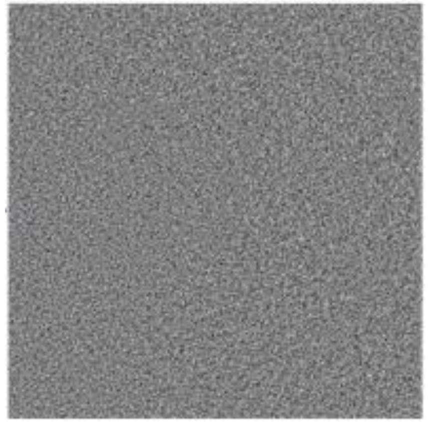
$$\arg\min_{Z,\theta} L_\theta(Y - Z) + R(Z) + R(\theta)$$

Y

=

Z

+



E

$z \sim p(z); e \sim p(e;\theta)$ $\rightarrow$ $p(z, e|y) \sim \text{likelihood}(y|z, e)p(z)p(e)$

Varational Posterior

$$p(z, \sigma^2|y) \Rightarrow q(z, \sigma^2|y) = q(z|y)q(\sigma^2|y)$$

$$q(z|y) = \prod_i^d \mathcal{N}(z_i|\mu_i(y; \boxed{W_D}), m_i^2(y; \boxed{W_D}))$$

D-Net

$$q(\sigma^2|y) = \prod_i^d \mathrm{IG}(\sigma_i^2|\alpha_i(y; \boxed{W_S}), \beta_i(y; \boxed{W_S}))$$

S-Net

Network parameters W_D and W_S are shared by posteriors calculated on all training data

Yue, Yong, Zhao, Meng, NeurIPS, 2019

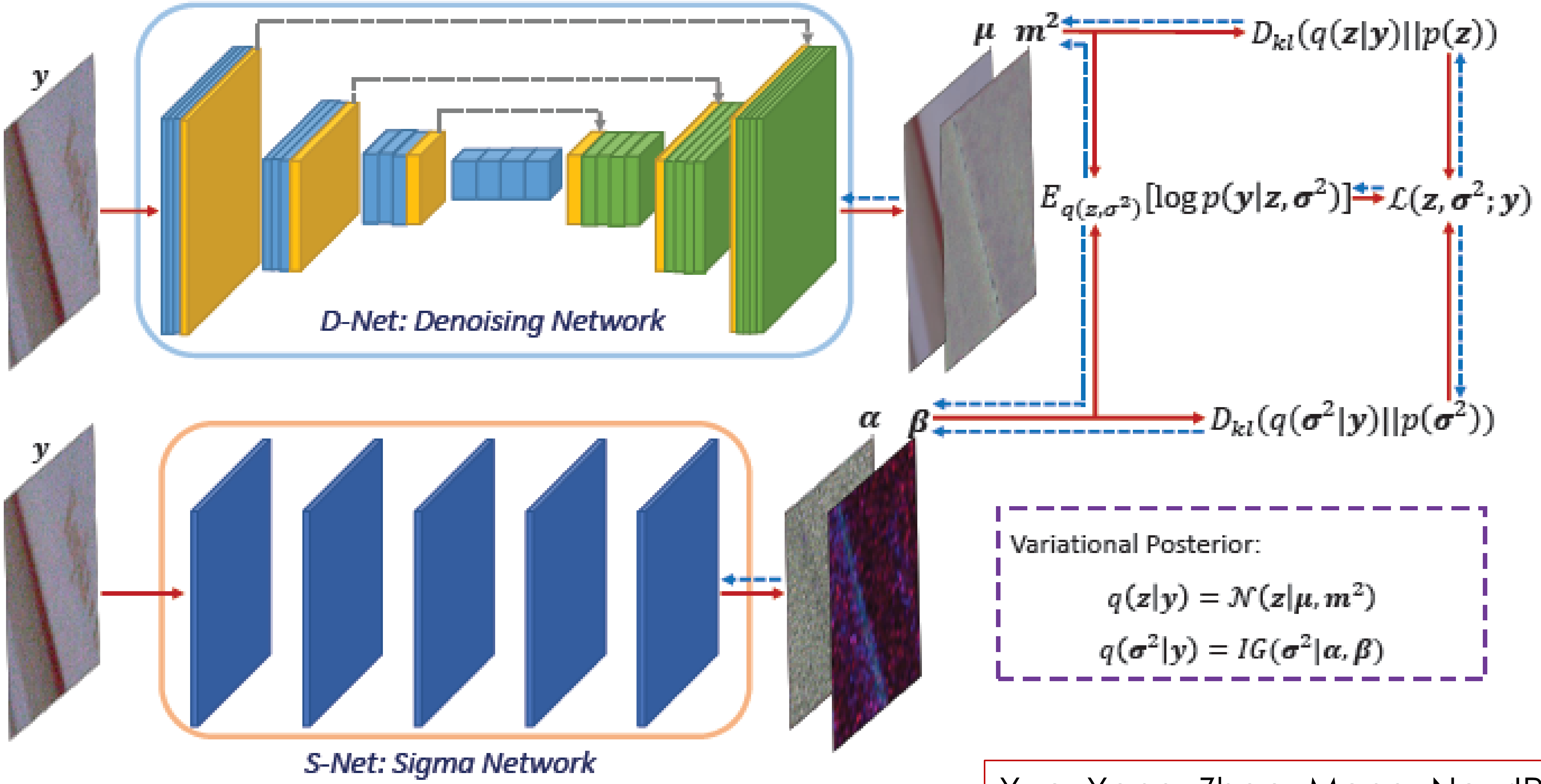
Objective: Minimizing KL Divergence

$$\min_{W_D, W_S} D_{KL}\left(q(z, \sigma^2 | y) || p(z, \sigma^2 | y)\right)$$

How?

Variational Inference!

Implementation Scheme



Yue, Yong, Zhao, Meng, NeurIPS, 2019

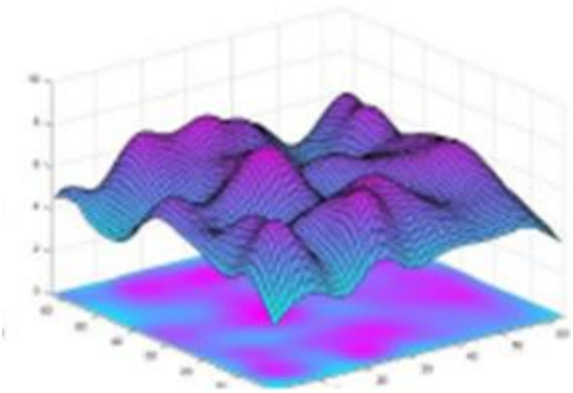
More Explanations on Rationality of This Objective

Variational Posterior:

$$q(\boldsymbol{z}|\boldsymbol{y}) = \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}, \boldsymbol{m}^2)$$

$$q(\boldsymbol{\sigma}^2|\boldsymbol{y}) = IG(\boldsymbol{\sigma}^2|\boldsymbol{\alpha}, \boldsymbol{\beta})$$

Restored Image



Extracted Noise Distribution

Yue, Yong, Zhao, Meng, NeurIPS, 2019

Summary

外练筋骨皮

内练一口气

万法终归宗

心齐泰山移