ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

ΜΑΘΗΜΑ 9

ΑΝΤΙΣΤΡΟΦΑ ΠΡΟΒΛΗΜΑΤΑ

Inverse Problems

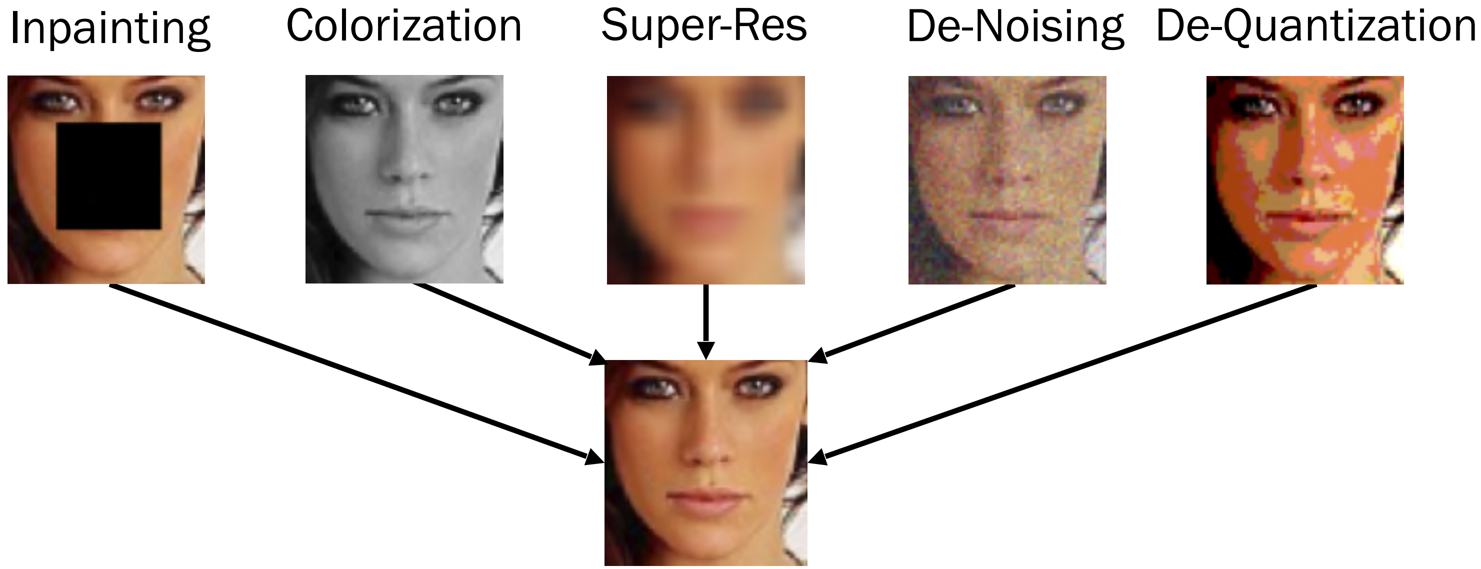
- Problem Definition
- Early Efforts
- Solutions Based on GANs
  - Add-Hoc
  - Applying Statistical Estimation
- Examples

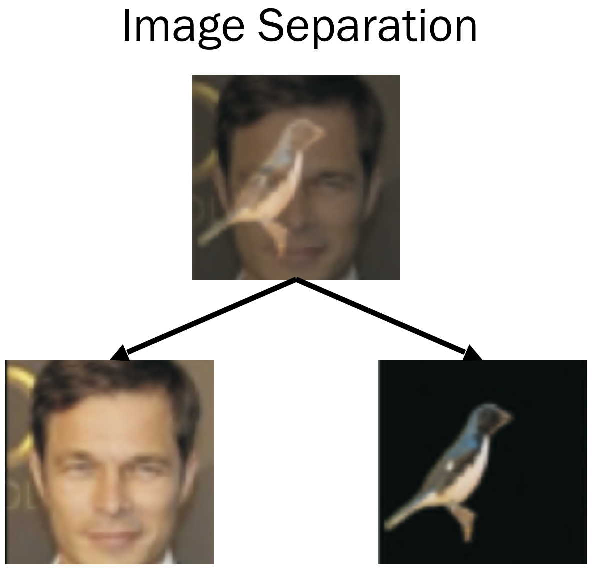
Problem Definition

We are given vector X (measurements) and interested in estimating vector Y

We assume $X = \mathsf{T}(Y) + W$ where $\mathsf{T}(Y)$ general (mostly) known transformation

Basic characteristic: $\boxed{\dim(X) \leq \dim(Y)}$





Early Efforts

Inpainting

USSR — Stalin

Diffusion based inpainting

No prior information, training data not very useful

Generative Models

X: ⋯

⋯ f(X)

Is it possible to generate synthetic data (realizations X_i) that follow f(X) ?

NOT an easy problem even if density f(X) is known!

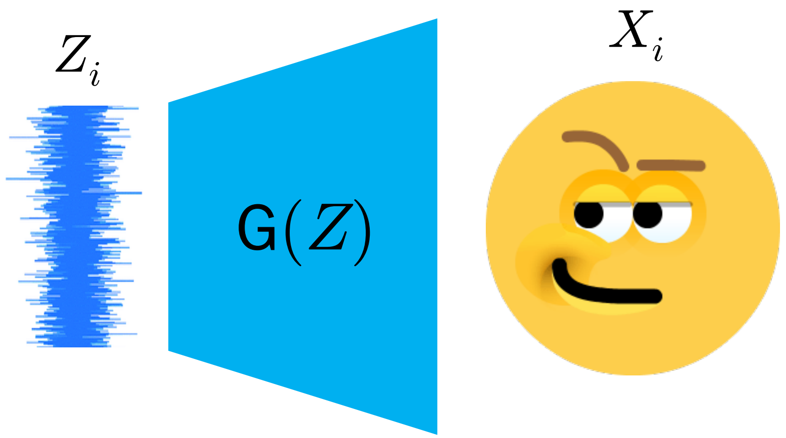
Begin with density h(Z): Simple to generate realizations Z_i

Find transformation G(Z): Such that $X_i =$ G(Z_i) follows f(X)

THEOREM: Under general conditions
a transformation G exists !!!

Pair $\{$G(Z),h$(Z)\}$ Generative model

G(Z) Generator



X follows f(X)

Z follows h(Z), design G(Z), so that Y = G(Z) follows f(Y)

THEOREM (Goodfellow et al. 2014): Z follows h(Z), define Y = G(Z) and cost

$$\mathrm{J}(\mathrm{G},\mathrm{D}) = \mathbb{E}_{\mathrm{f}}\left[\phi\left(\mathrm{D}(X)\right)\right] + \mathbb{E}_{\mathrm{h}}\left[\psi\left(\mathrm{D}\left(\mathrm{G}(Z)\right)\right)\right]$$

then the optimum solution to the adversarial problem

$$\min_{\mathrm{G}}\max_{\mathrm{D}} \mathrm{J}(\mathrm{G},\mathrm{D}) = \min_{\mathrm{G}}\max_{\mathrm{D}}\left\{\mathbb{E}_{\mathrm{f}}\left[\phi\left(\mathrm{D}(X)\right)\right] + \mathbb{E}_{\mathrm{h}}\left[\psi\left(\mathrm{D}\left(\mathrm{G}(Z)\right)\right)\right]\right\}$$

is such that Y = $\mathrm{G}_{\mathrm{o}}(Z)$ follows f(Y)

D(X) Discriminator G(Z) Generator

Data Driven Implementation

Highly likely realizations!

$\{X_1, X_2, \dots, X_n\}$ following f(X), $\{Z_1, Z_2, \dots, Z_m\}$ following h(Z)

D(X) approximated by neural network D(X,ϑ) (Discriminator)

Generator function G(Z) approximated by neural network G(Z,θ) (Generator)

$$\mathsf{J}(\theta,\vartheta) = \frac{1}{n}\sum_{i=1}^{n}\phi\big(\mathsf{D}(X_i,\vartheta)\big) + \frac{1}{m}\sum_{j=1}^{m}\psi\Big(\mathsf{D}\big(\mathsf{G}(Z_j,\theta),\vartheta\big)\Big)$$

Adversarial optimization becomes

$$\min_{\theta}\max_{\vartheta}\mathsf{J}(\theta,\vartheta) = \min_{\theta}\max_{\vartheta}\left\{\frac{1}{n}\sum_{i=1}^{n}\phi\big(\mathsf{D}(X_i,\vartheta)\big) + \frac{1}{m}\sum_{j=1}^{m}\psi\Big(\mathsf{D}\big(\mathsf{G}(Z_j,\theta),\vartheta\big)\Big)\right\}$$

$$\Rightarrow \{\theta_o, \vartheta_o\} \Rightarrow \theta_o \Rightarrow \mathsf{G}(Z,\theta_o)$$

IF Z follows h(Z) THEN $Y =$ G(Z,θ_o) follows f(Y)

Generative
Adversarial
Networks

Example (NVIDIA)

HD-CelebA (30 000 high definition images 1024 X 1024 of celebrities)

…  …

NVIDIA used progressive growing of GANs (4X4), (8X8),…,(1024X1024)

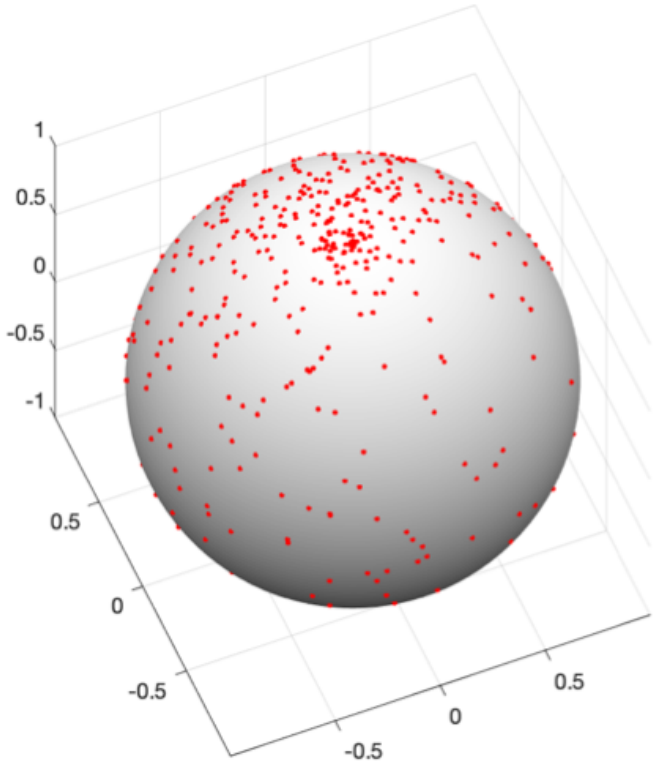
Y of size 3 X 10^6, Z Gaussian vector of length 500

Generative Models

vs

Probability Densities

Points in N-D space can be random and lie on a lower dimensional surface (manifold)



Example red points on sphere (2-D in 3-D space)

Points are random with coordinates $Y = [y_1, y_2, y_3]$ satisfying the deterministic equation

$$y_1^2 + y_2^2 + y_3^2 = r^2$$

Then density has the form

$$\mathrm{f}(y_1, y_2, y_3) = \delta(y_1^2 + y_2^2 + y_3^2 - r^2)\mathrm{h}(y_1, y_2)$$

Dirac $\delta(x)$ generalized function is defined as

$$\delta(x) = \begin{cases} 0 & x \neq 0 \\ \infty & x = 0 \end{cases}, \quad \int_{-\epsilon}^{\epsilon} \delta(x)\, dx = 1$$

Generative model would describe the random data with input density $h(z_1,z_2)$ and generator vector function $G(z_1,z_2)$

$$Y = G(z_1, z_2) \Rightarrow \begin{bmatrix} y_1 = G_1(z_1, z_2) \\ y_2 = G_2(z_1, z_2) \\ y_3 = G_3(z_1, z_2) \end{bmatrix} \Rightarrow \begin{bmatrix} y_1 = r\cos(2\pi z_1)\sin(\pi z_2) \\ y_2 = r\sin(2\pi z_1)\sin(\pi z_2) \\ y_3 = r\cos(\pi z_2) \end{bmatrix}$$

$h(z_1,z_2)$ defined on $[0,1]\times[0,1]$ and $G(z_1,z_2)$ is an ordinary function

Data are representable as $Y = G(Z)$, Z follows $h(Z)$. Many datasets satisfy

$$\dim(Z) \ll \dim(Y)$$

In HD CelebA: $\dim(Y) = 3 \times 1024 \times 1024 = 3 \times 10^6$

Input to Generator $G(Z)$: $\dim(Z) = 500$ (independent Gaussians)

Application to Inverse Problems

Solutions Based on GANs

Available training data $\{X_1,\ldots,X_n\}$

Design generator $\mathsf{G}(Z)$ so that when applied to Z with Z ~ $\mathsf{h}(\mathsf{Z})$ then $Y = \mathsf{G}(Z)$ has same density as training data

Generate $\{Z_1,\ldots,Z_m\}$ with density $\mathsf{h}(\mathsf{Z})$

$$\hat{\mathsf{J}}(\theta,\vartheta) = \frac{1}{n}\sum_{t=1}^{n}\phi\big(\underbrace{\mathsf{D}}_{\text{Discriminator}}(X_t,\vartheta)\big) + \frac{1}{m}\sum_{t=1}^{m}\psi\Big(\mathsf{D}\big(\underbrace{\mathsf{G}}_{\text{Generator}}(Z_t,\theta),\vartheta\big)\Big)$$

$$\min_{\theta}\ \max_{\vartheta}\ \hat{\mathsf{J}}(\theta,\vartheta)$$

Assume known Generative Model $\{\mathbf{G}(Z),\mathbf{h}(Z)\}$, also Discriminator $\mathsf{D}(X)$

General Problem

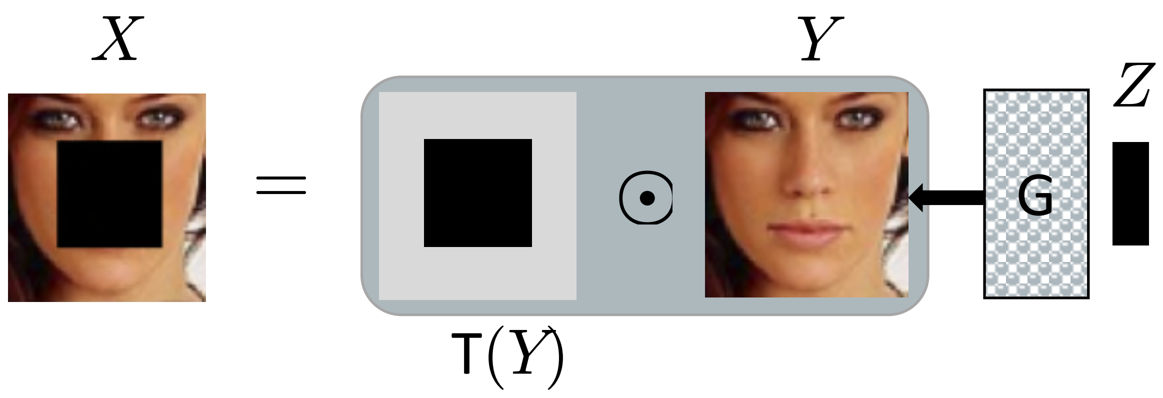
Given vector X (measurements) we are interested in estimating vector Y

with $X = \mathsf{T}(Y) + W$ and T(Y) known transformation

Y follows Generative Model

There exists Z following density $\mathsf{h}(Z)$ such that $\quad Y = \mathsf{G}(Z)$

Instead of estimating Y from X we estimate input to generative model Z



Measurements $X = \mathsf{T}(Y) + W$

Compute estimate $\hat{Y}$

Measurements $X = \mathsf{T}\big(\mathsf{G}(Z)\big) + W$

Compute estimate $\hat{Z}$

and let $\hat{Y} = \mathsf{G}(\hat{Z})$

Several image restoration problems in Computer Vision can be formulated as follows

Ideal

Measurement $X = \mathsf{T}(Y) + W$ more general $X = \mathsf{T}(Y, \alpha) + W$

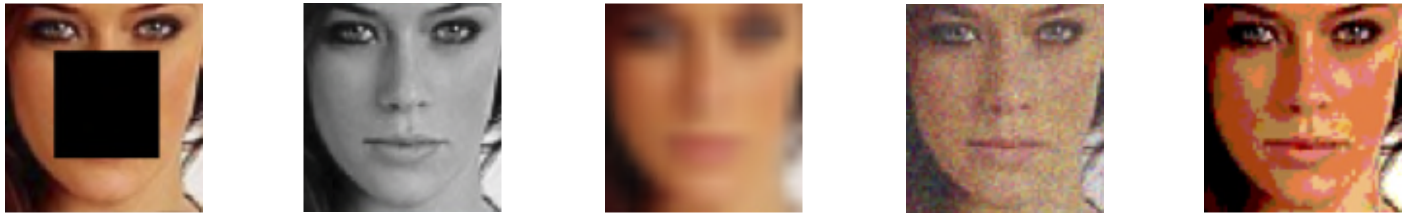
Known transformation Noise Unknown parameters

Problem: Recover (restore) ideal Y from measurements X

Ideal

Restore

Examples of transformations

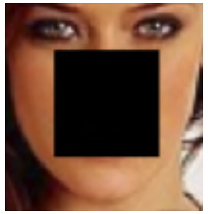


Inpainting Colorization Super-Res De-Noising De-Quantization

Recovering Y from measurements X is an ill posed problem

Y

X



Inpainting

More unknowns than equations

Classical approach: Impose “smoothness” constraints to obtain a (unique) solution

…  …

Available generative model $\{\mathrm{G}(Z), \mathrm{h}(Z)\}$: $Y = \mathrm{G}(Z)$

Since $Y = \mathrm{G}(Z)$, instead of estimating Y, estimate input to generator Z then recover Y as the output of the generator

Because $\dim(Z) \ll \dim(Y)$, significant computational gain and stable processing

1

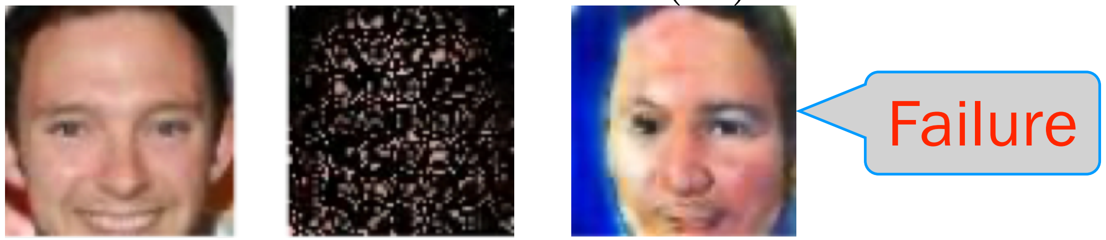
Ad-Hoc Approaches

Select Z so that measurement X and $\mathrm{T}(\mathrm{G}(Z))$ are "close"

$$\min_{Z} \|X - \mathrm{T}(\mathrm{G}(Z))\|^2 \Rightarrow Z_o \Rightarrow Y_o = \mathrm{G}(Z_o)$$

Well defined optimization, computationally stable

Y $\quad$ X $\quad$ $\mathrm{G}(Z_o)$



Generative model is a pair $\{\mathrm{G}(Z), \mathrm{h}(Z)\}$

Even for $\mathrm{T}(\mathrm{G}(Z_o))$ "close" to X, if likelihood $\mathrm{h}(Z_o)$ is very small

then $Y_o = \mathrm{G}(Z_o)$ is a bad solution

Must take into account input density $\mathrm{h}(Z)$

2

Yeh et al. (2017), (2018)

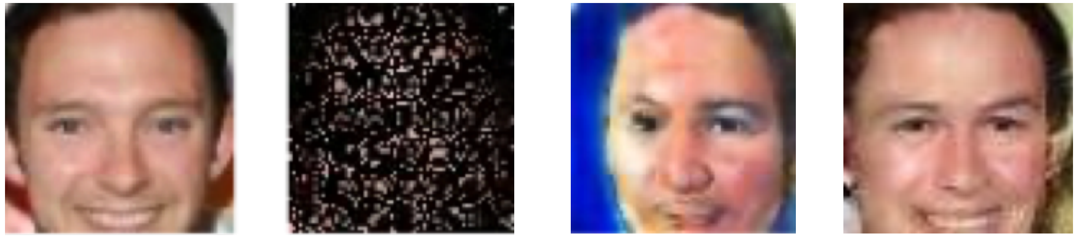
Regularizer

log $\mathsf{h}(Z)$

$$\mathsf{J}(Z) = \|X - \mathsf{T}\big(\mathsf{G}(Z)\big)\| + \lambda\Big\{ \log\Big(1 - \mathsf{D}\big(\mathsf{G}(Z)\big)\Big) - \log\Big(\mathsf{D}\big(\mathsf{G}(Z)\big)\Big) - \frac{1}{2}\|Z\|^2 \Big\}$$

Parameter needs tuning
Complicated

$$\min_Z \mathsf{J}(Z) \;\Rightarrow\; Z_o \;\Rightarrow\; Y_o = \mathsf{G}(Z_o)$$



Success ?

Asim et al. (2019)

$$\mathsf{J}(Z) = \|X - \mathsf{T}\big(\mathsf{G}(Z)\big)\|^2 + \lambda\|Z\|^2 \qquad \min_Z \mathsf{J}(Z) \;\Rightarrow\; Z_o \;\Rightarrow\; Y_o = \mathsf{G}(Z_o)$$

Both methods require exact knowledge of $\mathsf{T}(Y)$

Applying Statistical Estimation

Given densities $\mathsf{f}(X|Z)$ and prior $\mathsf{h}(Z)$, for measurement X estimate Z

MAP estimator is of the form

$$\hat{Z} = \arg\max_{Z} \mathsf{f}(Z|X) = \arg\max_{Z} \mathsf{f}(X,Z) = \arg\max_{Z} \mathsf{f}(X|Z)\mathsf{h}(Z)$$

Let $Z=\{Z_1,Z_2\}$ where there is prior for Z_1 but not for Z_2

Treat non-existing prior as degenerate uniform

$$\begin{aligned}
\arg\max_{Z_1,Z_2} \mathsf{f}(Z_1,Z_2|X) &= \arg\max_{Z_1,Z_2} \mathsf{f}(X,Z_1,Z_2) \\
&= \arg\max_{Z_1,Z_2} \mathsf{f}(X|Z_1,Z_2)\mathsf{h}(Z_1,Z_2) \\
&= \arg\max_{Z_1,Z_2} \mathsf{f}(X|Z_1,Z_2)\mathsf{h}_1(Z_1|Z_2)\mathsf{h}_2(Z_2) \\
\mathsf{h}_2(Z_2) \text{ degenerate uniform} \quad &= \arg\max_{Z_1,Z_2} \mathsf{f}(X|Z_1,Z_2)\mathsf{h}_1(Z_1|Z_2)
\end{aligned}$$

4

If interested in estimating Z_1 and Z_2 are nuisance parameters then

$$\hat{Z}_1 = \arg\max_{Z_1} \left\{ \max_{Z_2} \mathsf{f}(X|Z_1, Z_2)\mathsf{h}_1(Z_1|Z_2) \right\}$$

where $\mathsf{h}_1(Z_1|Z_2)$ prior of Z_1 given Z_2

If Z_1 does not depend on Z_2 then $\mathsf{h}_1(Z_1|Z_2)=\mathsf{h}_1(Z_1)$ and

$$\hat{Z}_1 = \arg\max_{Z_1} \left\{ \max_{Z_2} \mathsf{f}(X|Z_1, Z_2)\mathsf{h}_1(Z_1) \right\}$$

We are given vector X (measurements) and interested in estimating vector Y

We assume $X = \mathsf{T}(Y,\alpha) + W = \mathsf{T}(\mathsf{G}(Z),\alpha) + W$

$\mathsf{T}(Y,\alpha)$: transformation of known mathematical form possibly containing unknown parameters α. Can be different per measurement X

W: additive noise with density $\mathsf{g}_w(W,\beta)$ possibly containing unknown parameters β. Can be different per measurement X

Z: follows density $\mathsf{h}(Z)$ from generative model {$\mathsf{G}(Z)$, $\mathsf{h}(Z)$}

$Z_1 = Z,\quad Z_2 = \{\alpha,\beta\}\qquad \mathsf{f}(X|Z_1,Z_2) = \mathsf{f}(X|Z,\alpha,\beta) = \mathsf{g}_w\Big(X - \mathsf{T}\big(\mathsf{G}(Z),\alpha\big),\beta\Big)$

$$\hat{Z} = \arg\max_{Z}\left\{\max_{\alpha,\beta}\mathsf{f}(X|Z,\alpha,\beta)\mathsf{h}(Z)\right\}$$

$$= \arg\max_{Z}\left\{\max_{\alpha,\beta}\mathsf{g}_w\Big(X - \mathsf{T}\big(\mathsf{G}(Z),\alpha\big),\beta\Big)\mathsf{h}(Z)\right\}$$

$$\hat{Z} = \arg\max_Z \left\{ \max_\alpha \max_\beta \mathsf{g}_w\Big(X - \mathsf{T}(\mathsf{G}(Z),\alpha),\beta\Big)\mathsf{h}(Z) \right\}$$

W: additive noise is Gaussian mean 0 and covariance $\beta^2 I$

$$\max_\beta \mathsf{g}_w\Big(X - \mathsf{T}(\mathsf{G}(Z),\alpha),\beta\Big) = \max_\beta \frac{e^{-\|X - \mathsf{T}(\mathsf{G}(Z),\alpha)\|^2/2\beta^2}}{(\sqrt{2\pi\beta^2})^N}$$

$$= \frac{\mathsf{C}}{\|X - \mathsf{T}(\mathsf{G}(Z),\alpha)\|^N}$$

N: length of measurement vector X

$$\hat{Z} = \arg\max_Z \left\{ \max_\alpha \frac{\mathsf{C}\,\mathsf{h}(Z)}{\|X - \mathsf{T}(\mathsf{G}(Z),\alpha)\|^N} \right\}$$

$$= \arg\max_Z \frac{\mathsf{h}(Z)}{(\min_\alpha \|X - \mathsf{T}(\mathsf{G}(Z),\alpha)\|^2)^{N/2}}$$

Z: If input of generative model is Gaussian with mean 0 and covariance identity

$$\hat{Z} = \arg\max_{Z} \frac{\mathsf{C}'\, e^{-\|Z\|^2/2}}{(\min_\alpha \|X - \mathsf{T}(\mathsf{G}(Z), \alpha)\|^2)^{N/2}}$$

$$\boxed{\hat{Z} = \arg\min_{Z} \left\{ \log\left(\min_{\alpha} \|X - \mathsf{T}(\mathsf{G}(Z), \alpha)\|^2 \right) + \frac{1}{N}\|Z\|^2 \right\}}$$

$$\Leftrightarrow \quad \min_{Z,\alpha} \left\{ \log\left(\|X - \mathsf{T}(\mathsf{G}(Z), \alpha)\|^2 \right) + \frac{1}{N}\|Z\|^2 \right\}$$

If transformation satisfies

$$\mathsf{T}(Y,\alpha) = \alpha_1 \mathsf{T}_1(Y) + \cdots + \alpha_m \mathsf{T}_m(Y)$$

then $\mathsf{T}\big(\mathsf{G}(Z),\alpha\big) = \alpha_1 \mathsf{T}_1\big(\mathsf{G}(Z)\big) + \cdots + \alpha_m \mathsf{T}_m\big(\mathsf{G}(Z)\big) = \mathcal{G}(Z)A$

where $\mathcal{G}(Z) = \Big[\mathsf{T}_1\big(\mathsf{G}(Z)\big) \cdots \mathsf{T}_m\big(\mathsf{G}(Z)\big)\Big],\ A = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_m \end{bmatrix}$

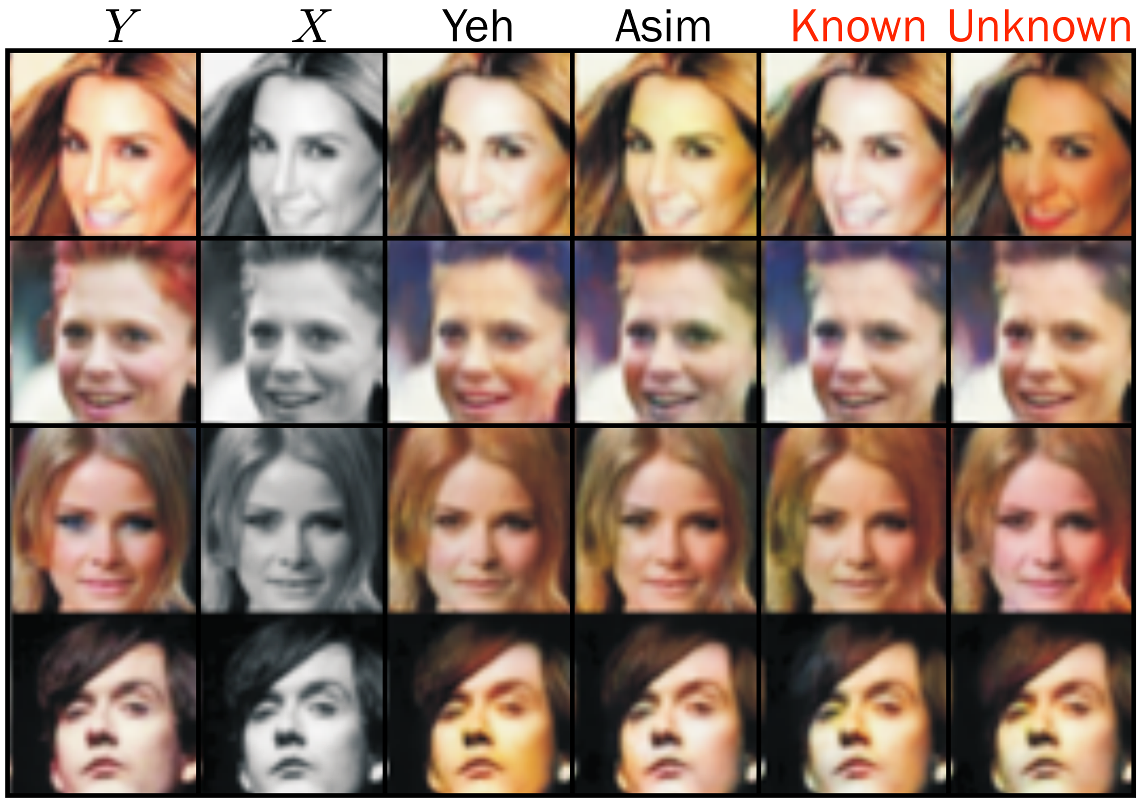
$$\min_A \|X - \mathcal{G}(Z)A\|^2 = \|X\|^2 - X^\intercal \mathcal{G}(Z)\big(\mathcal{G}^\intercal(Z)\mathcal{G}(Z)\big)^{-1}\mathcal{G}^\intercal(Z)X$$

$$\hat{Z} = \arg\min_Z \left\{\log\Big(\|X\|^2 - X^\intercal \mathcal{G}(Z)\big(\mathcal{G}^\intercal(Z)\mathcal{G}(Z)\big)^{-1}\mathcal{G}^\intercal(Z)X\Big) + \frac{1}{N}\|Z\|^2\right\}$$

Examples

Blurring with 3 X 3 mask

Colorization (green channel)



De-Quantization

2 levels per RGB channel, 8 colors

De-Quantization

3 levels per RGB channel, 27 colors

De-Quantization and Colorization

RGB → Gray → BW (2 levels)

DATA MIXTURES

$$Y_1 = G_1(z_1) \qquad z_1 \sim h_1(z)$$

$$Y_2 = G_2(z_2) \qquad z_2 \sim h_2(z)$$

$$X = \alpha_1 Y_1 + \alpha_2 Y_2 + W$$

If W Gaussian with unknown variance

If $h_1(z)$, $h_2(z)$ both Gaussian with identity covariance (not necessarily of same length!).

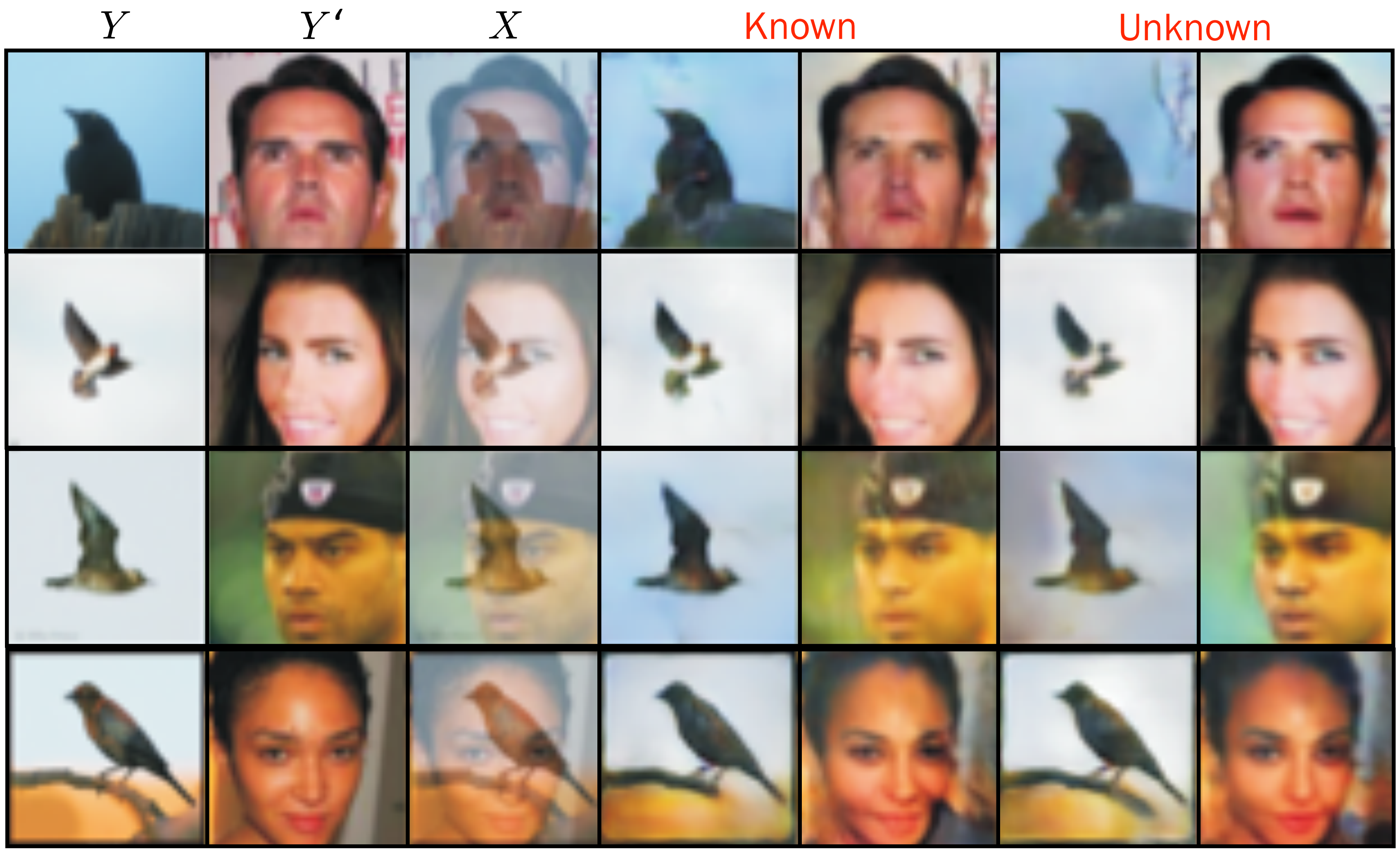
$$\min_{z_1, z_2} \left\{ \log \left(\underbrace{\min_{\alpha_1, \alpha_2} \| X - \alpha_1 G_1(z_1) - \alpha_2 G(z_2) \|^2}_{\|X\|^2 - X^T \mathcal{G}(z_1,z_2) \left(\mathcal{G}^T(z_1,z) \, \mathcal{G}(z_1,z_2) \right)^{-1} \mathcal{G}^T(z_1,z_2) X} \right) + \frac{1}{N} \left(\|z_1\|^2 + \|z_2\|^2 \right) \right\}$$

where

$$\mathcal{G}(z_1, z_2) = \left[G_1(z_1) \quad G_2(z_2) \right]$$

Data Mixtures

$$X = \alpha_1 Y_1 + \alpha_2 Y_2 + W$$



Nonlinear Data Mixtures

We assume two independent data vectors Y_1, Y_2 that are combined as

$$X = \mathsf{T}(Y_1, Y_2, \alpha) + W$$

Y_1: generative model $\{\mathsf{G}_1(Z_1), \mathsf{h}_1(Z_1)\}$
Y_2: generative model $\{\mathsf{G}_2(Z_2), \mathsf{h}_2(Z_2)\}$
W: additive noise with density $\mathsf{g}_w(W, \beta)$ containing unknown parameters β

$$\{\hat{Z}_1, \hat{Z}_2\} =$$

$$\arg\min_{Z_1, Z_2} \left\{ \log\Big(\min_{\alpha} \|X - \mathsf{T}\big(\mathsf{G}_1(Z_1), \mathsf{G}_2(Z_2), \alpha\big)\|^2\Big) + \frac{1}{N}\big(\|Z_1\|^2 + \|Z_2\|^2\big) \right\}$$

$$\Rightarrow \min_{Z_1, Z_2, \alpha} \left\{ \log\Big(\|X - \mathsf{T}\big(\mathsf{G}_1(Z_1), \mathsf{G}_2(Z_2), \alpha\big)\|^2\Big) + \frac{1}{N}\big(\|Z_1\|^2 + \|Z_2\|^2\big) \right\}$$