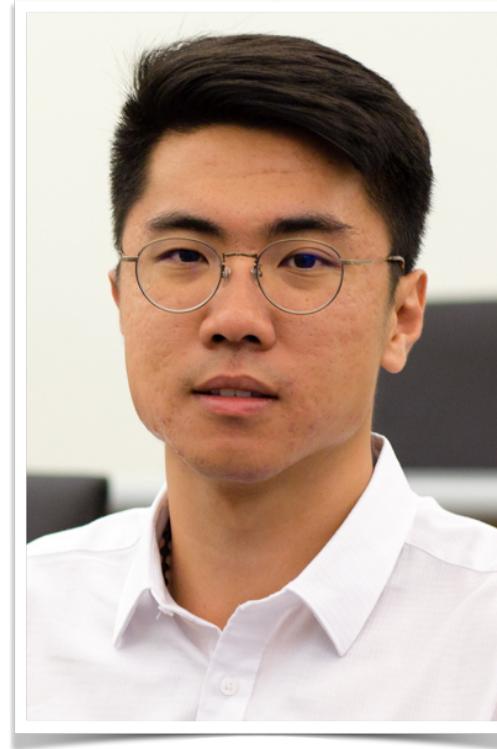
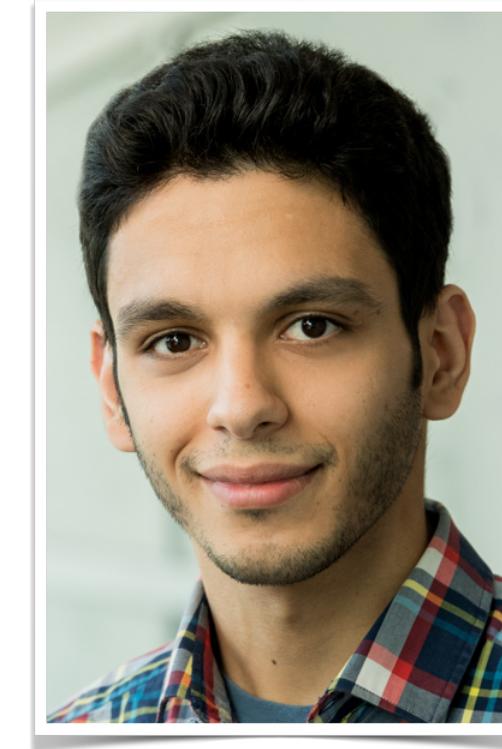


# Conjugate Energy-Based Models



Hao  
Wu\* [1]



Babak  
Esmaeili\* [1]



Michael  
Wick [2]



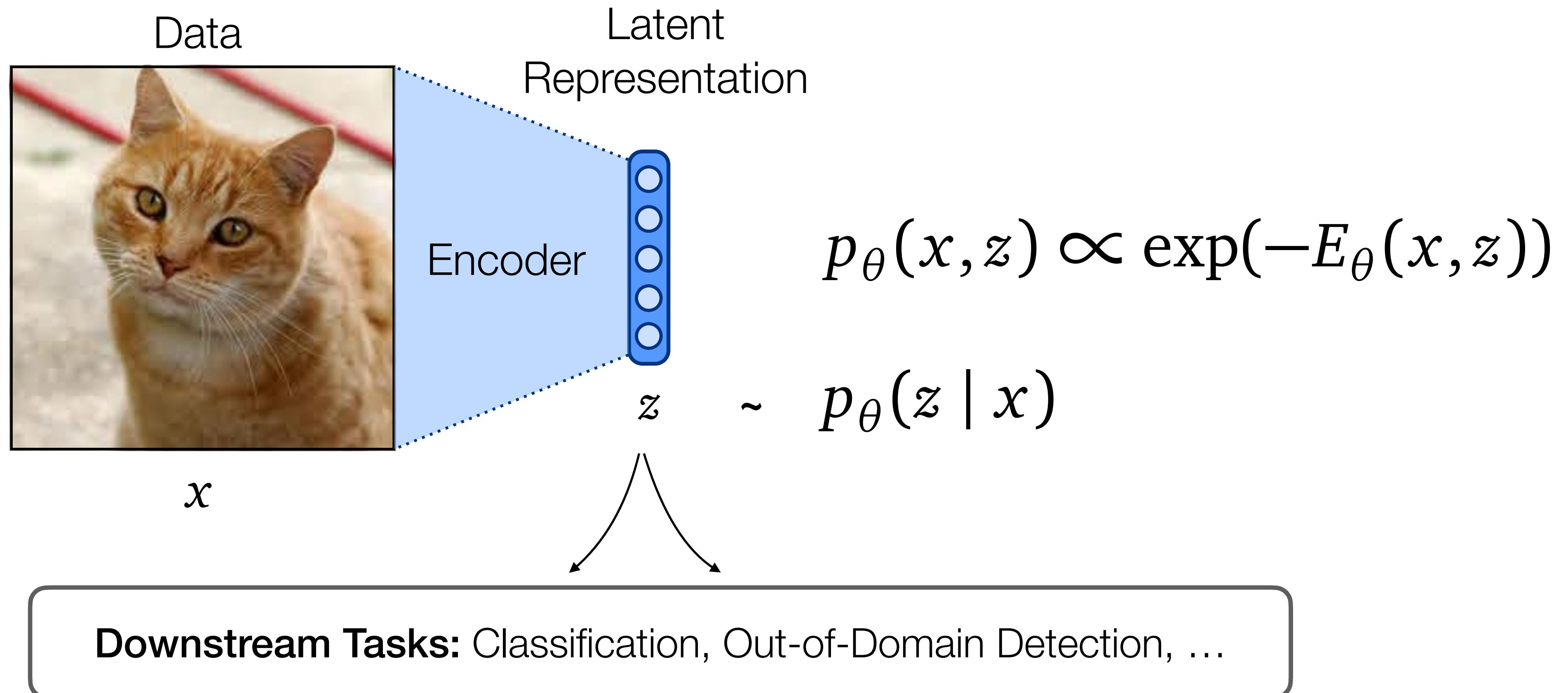
Jean-Baptiste  
Tristan [3]



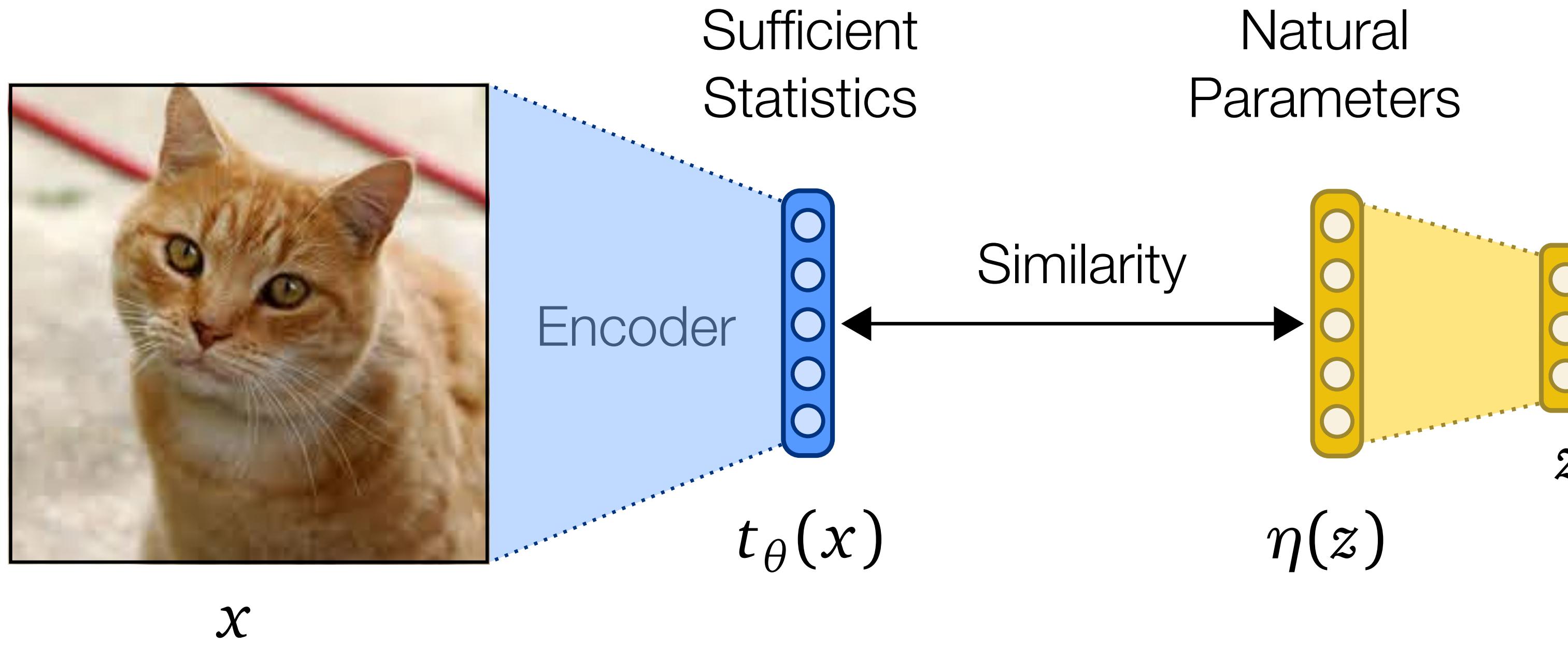
Jan-Willem  
van de Meent [1]

[1] Northeastern University, [2] Oracle Labs, [3] Boston College

# Learning Representation with Energy-Based Models

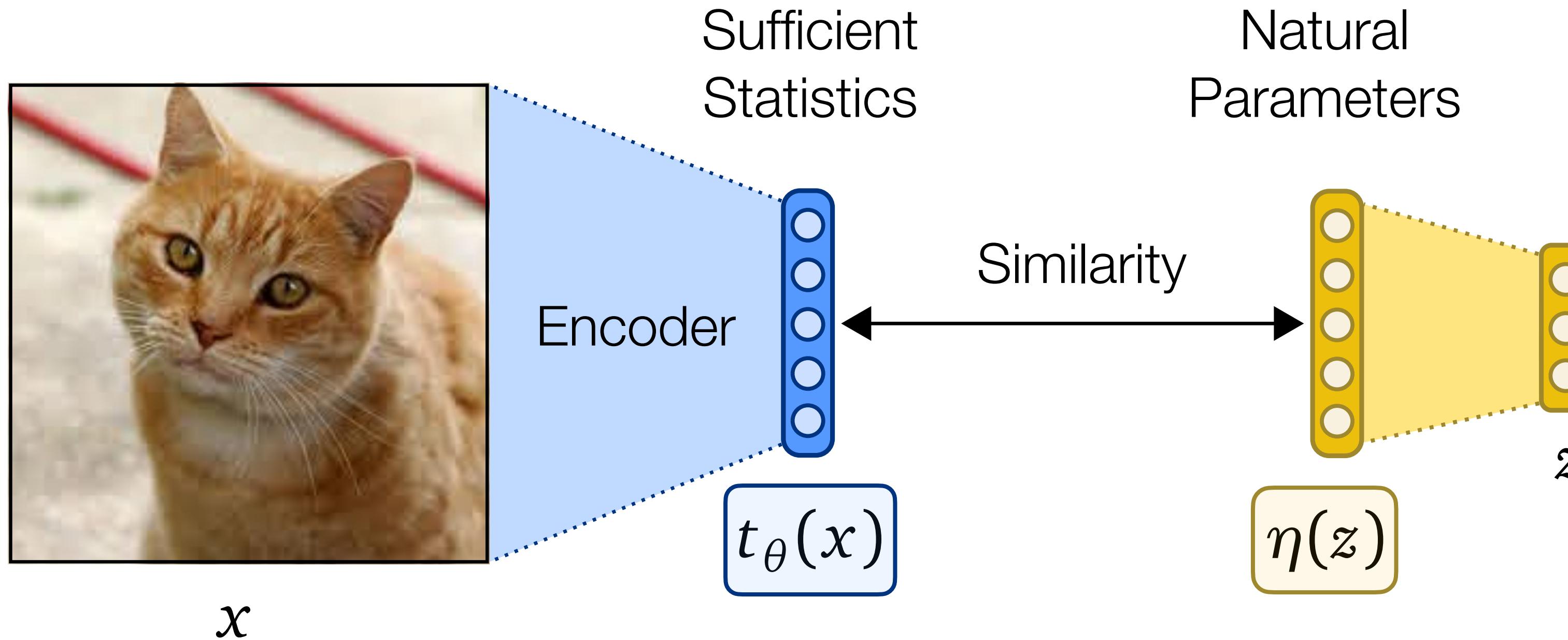


# Conjugate Energy-Based Models



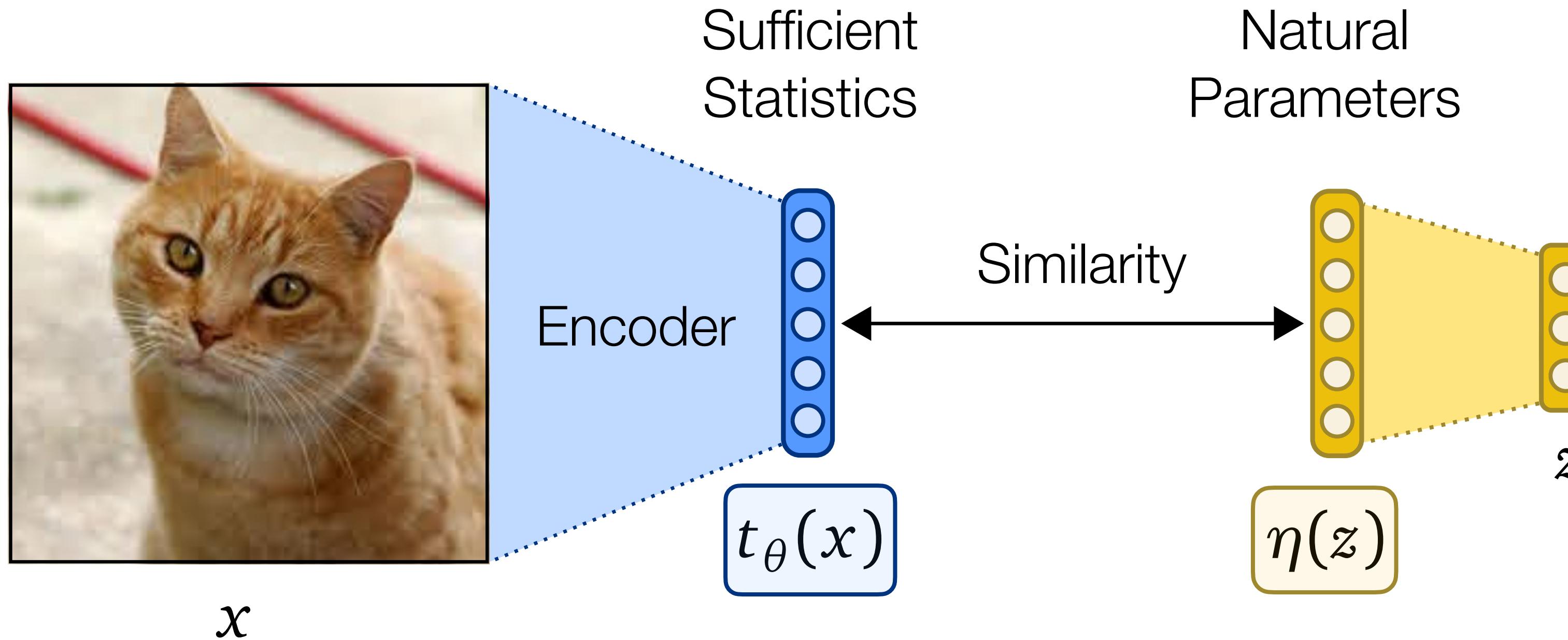
$$-E_\theta(x, z) = \underbrace{\langle t_\theta(x), \eta(z) \rangle}_{\text{Similarity (inner product)}} + \underbrace{\langle \lambda, \eta(z) \rangle - A(\lambda)}_{\text{Bias: } \log p(z | \lambda) \text{ (exponential family)}}$$

# Conjugate Energy-Based Models



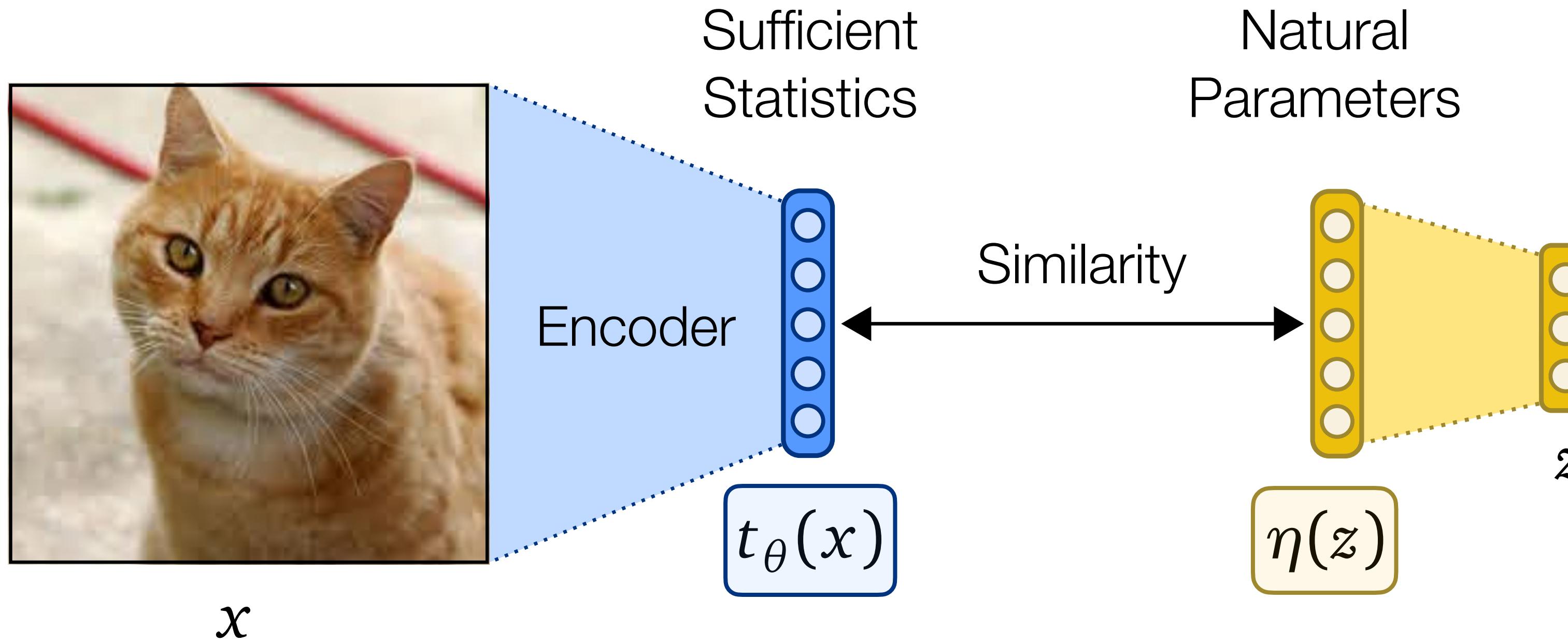
$$\begin{aligned} -E_\theta(x, z) &= \langle t_\theta(x), \eta(z) \rangle + \langle \lambda, \eta(z) \rangle - A(\lambda) \\ &= \langle t_\theta(x) + \lambda, \eta(z) \rangle - A(\lambda) \\ &\quad + A(t_\theta(x) + \lambda) - A(t_\theta(x) + \lambda) \end{aligned}$$

# Conjugate Energy-Based Models



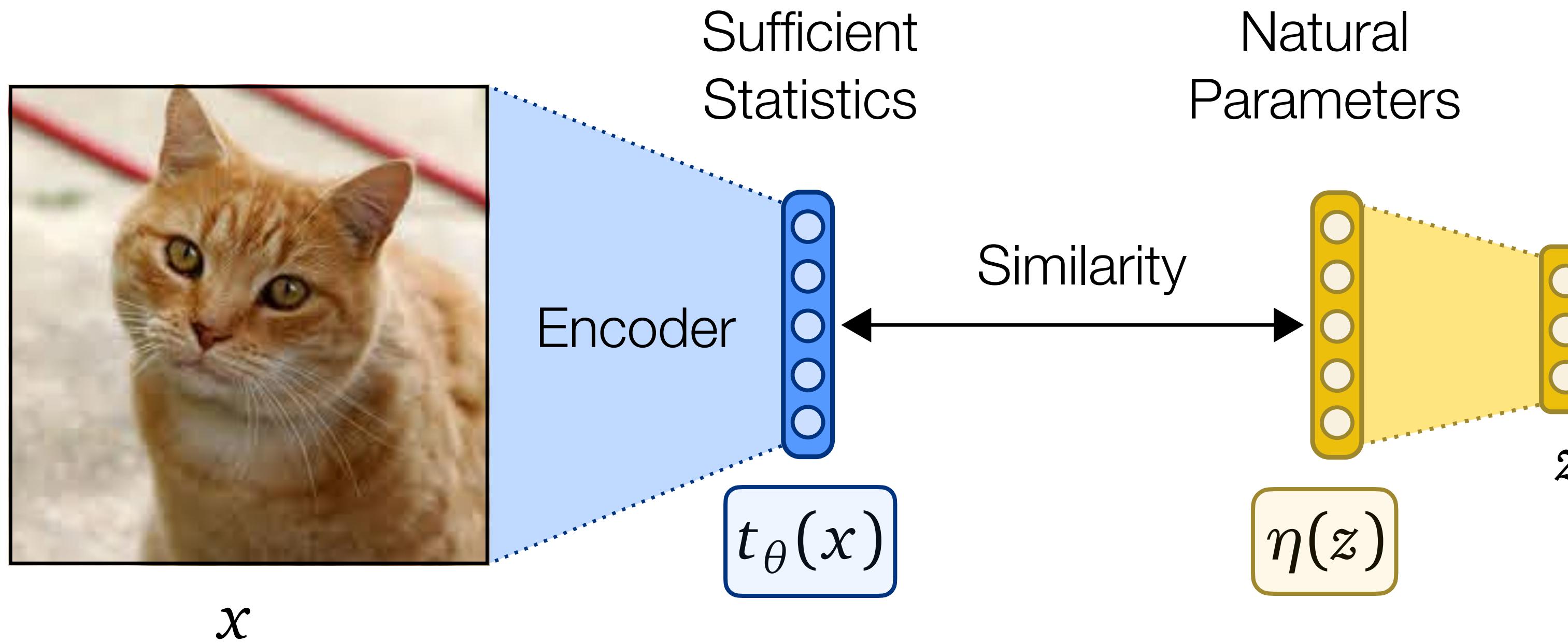
$$\begin{aligned} -E_\theta(x, z) &= \langle t_\theta(x), \eta(z) \rangle + \langle \lambda, \eta(z) \rangle - A(\lambda) \\ &= \langle t_\theta(x) + \lambda, \eta(z) \rangle - A(\lambda) \\ &\quad + A(t_\theta(x) + \lambda) - A(t_\theta(x) + \lambda) \end{aligned}$$

# Conjugate Energy-Based Models



$$\begin{aligned} -E_\theta(x, z) &= \langle t_\theta(x), \eta(z) \rangle + \langle \lambda, \eta(z) \rangle - A(\lambda) \\ &= \langle t_\theta(x) + \lambda, \eta(z) \rangle - A(t_\theta(x) + \lambda) \\ &\quad + A(t_\theta(x) + \lambda) - A(\lambda) \end{aligned}$$

# Conjugate Energy-Based Models



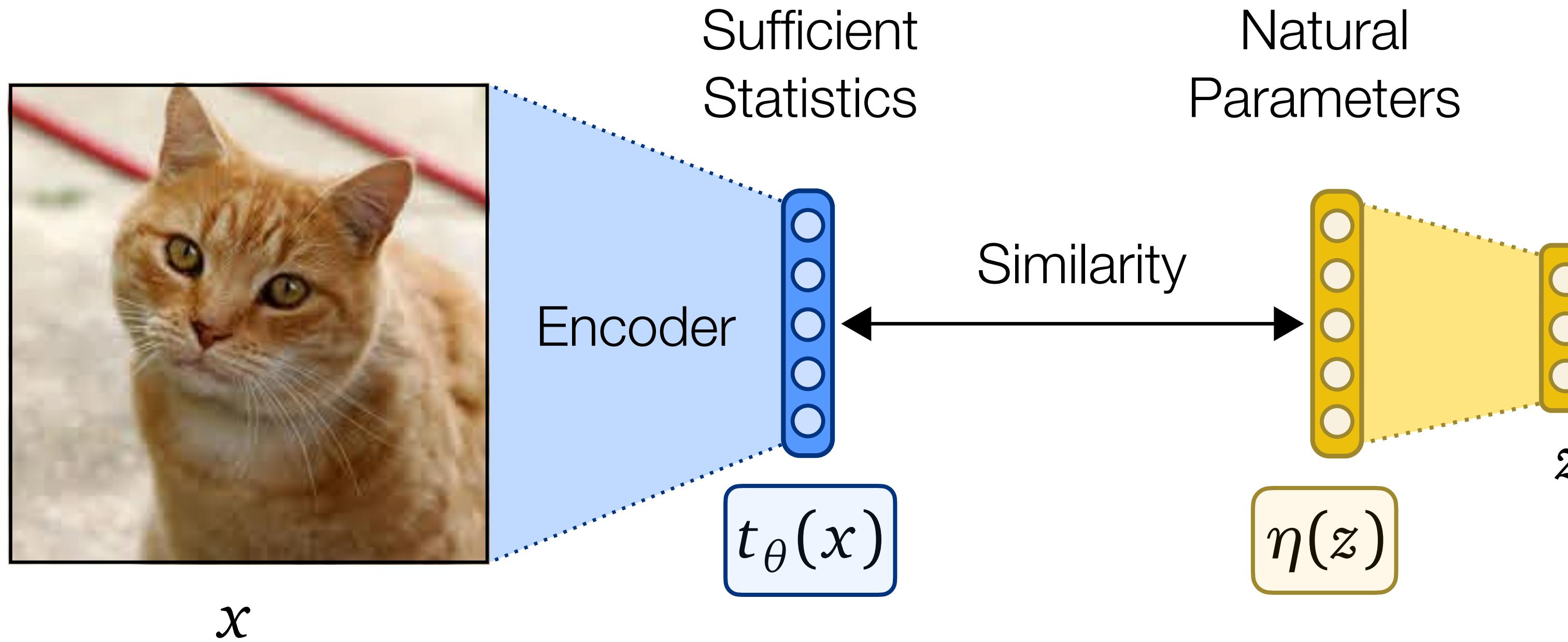
$$-E_\theta(x, z) = \langle t_\theta(x), \eta(z) \rangle + \langle \lambda, \eta(z) \rangle - A(\lambda)$$

$$= \langle t_\theta(x) + \lambda, \eta(z) \rangle - A(t_\theta(x) + \lambda)$$

$$+ A(t_\theta(x) + \lambda) - A(\lambda)$$

Conjugate Posterior  
 $\log p(z | t_\theta(x) + \lambda)$

# Conjugate Energy-Based Models



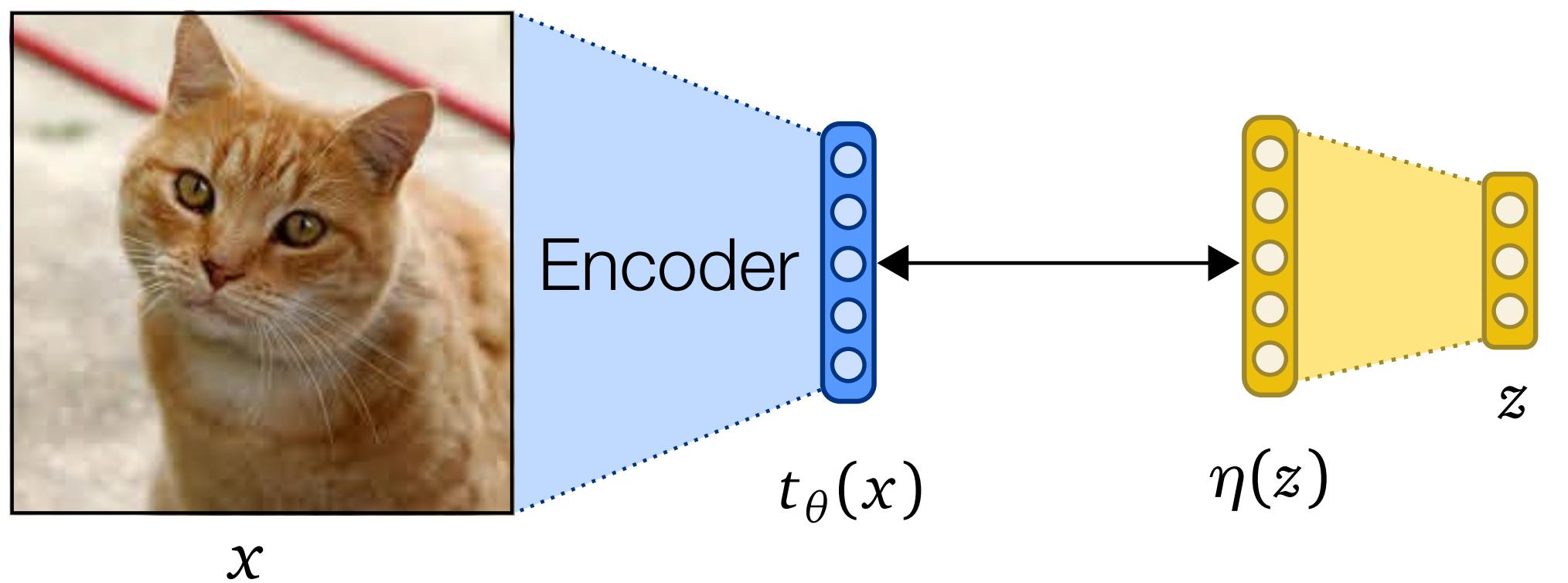
$$\begin{aligned} -E_\theta(x, z) &= \langle t_\theta(x), \eta(z) \rangle + \langle \lambda, \eta(z) \rangle - A(\lambda) \\ &= \langle t_\theta(x) + \lambda, \eta(z) \rangle - A(t_\theta(x) + \lambda) \\ &\quad + A(t_\theta(x) + \lambda) - A(\lambda) \end{aligned}$$

EBM (can train with standard methods)

Maximum marginal likelihood

No supervision

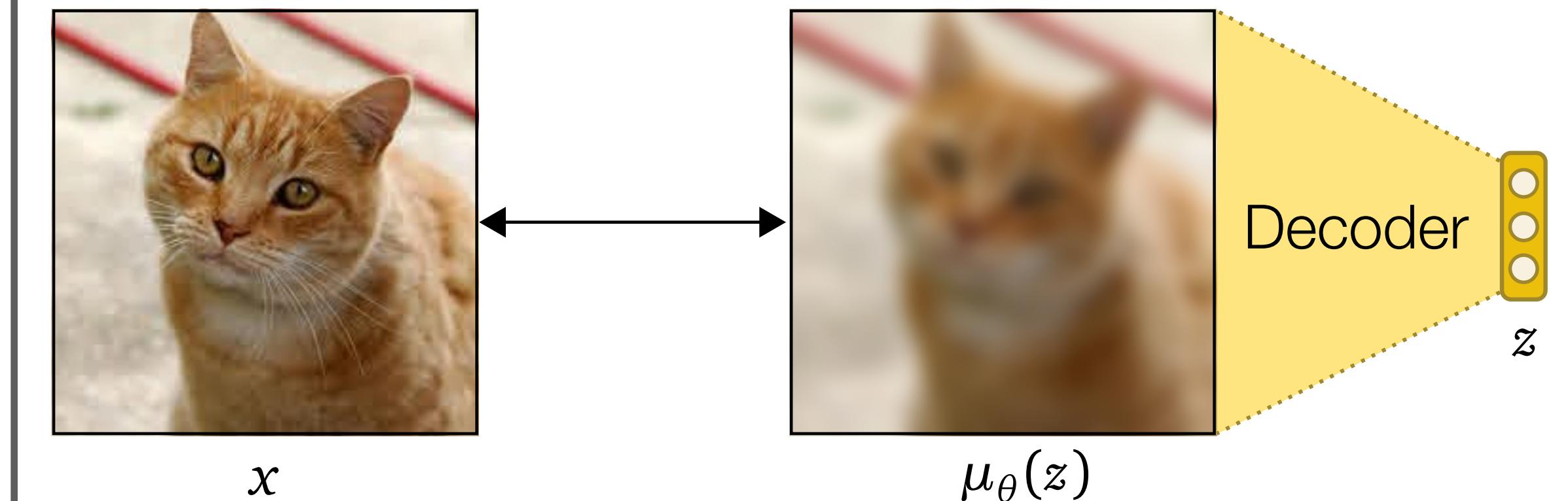
## Conjugate Energy-Based Model



Posterior  $p_\theta(z | x) = p(z | \lambda + t_\theta(x))$

Likelihood  $p_\theta(x | z) \propto \exp(\langle t_\theta(x), \eta(z) \rangle)$

## Variational Autoencoder



Posterior  $p_\theta(z | x) \approx q_\phi(z | x)$

Likelihood  $p_\theta(x | z) = p(x | \mu_\theta(z))$

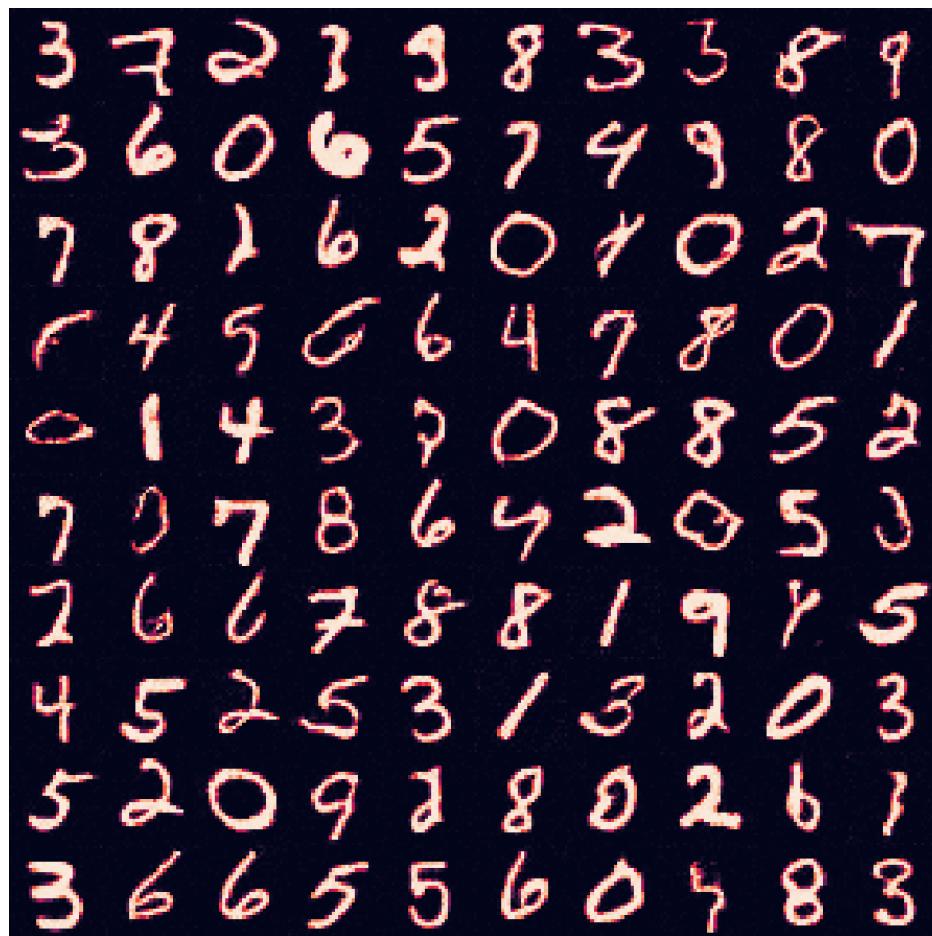
# Experiments

CEBM: Gaussian inductive bias

GMM-CEBM: GMM inductive bias

# Can CEBMs approximate the data distribution?

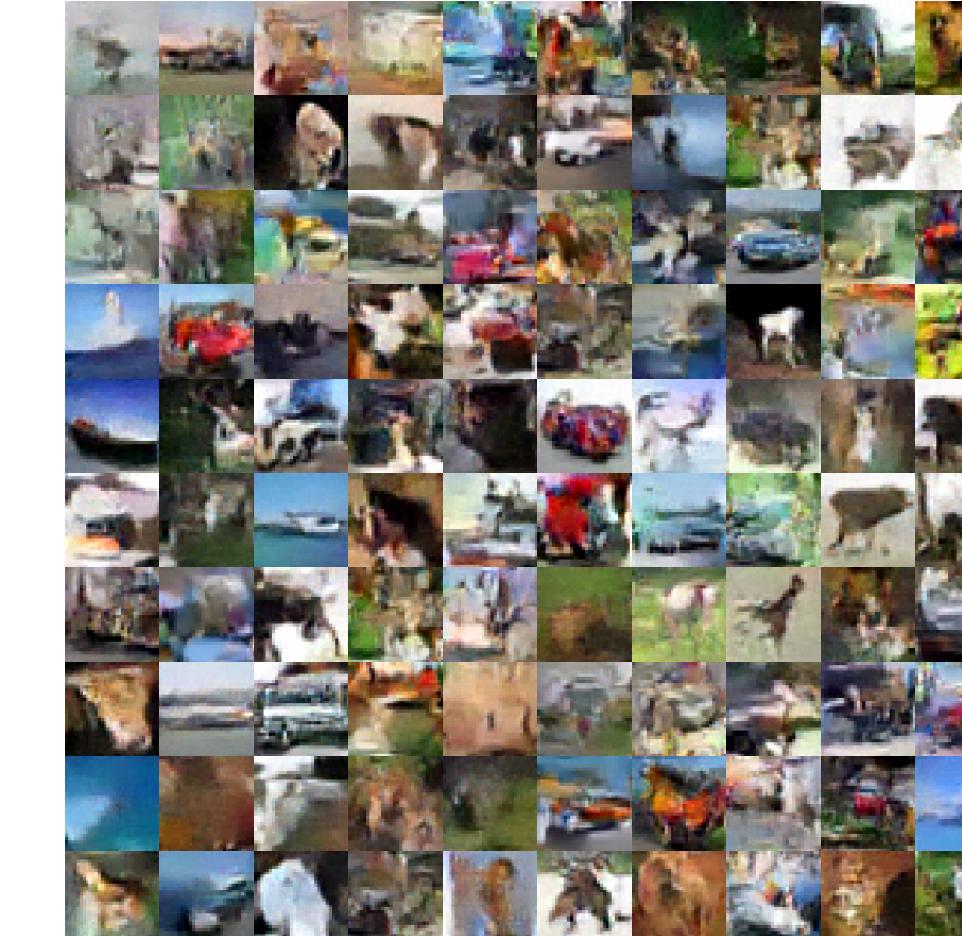
MNIST



F-MNIST



CIFAR-10

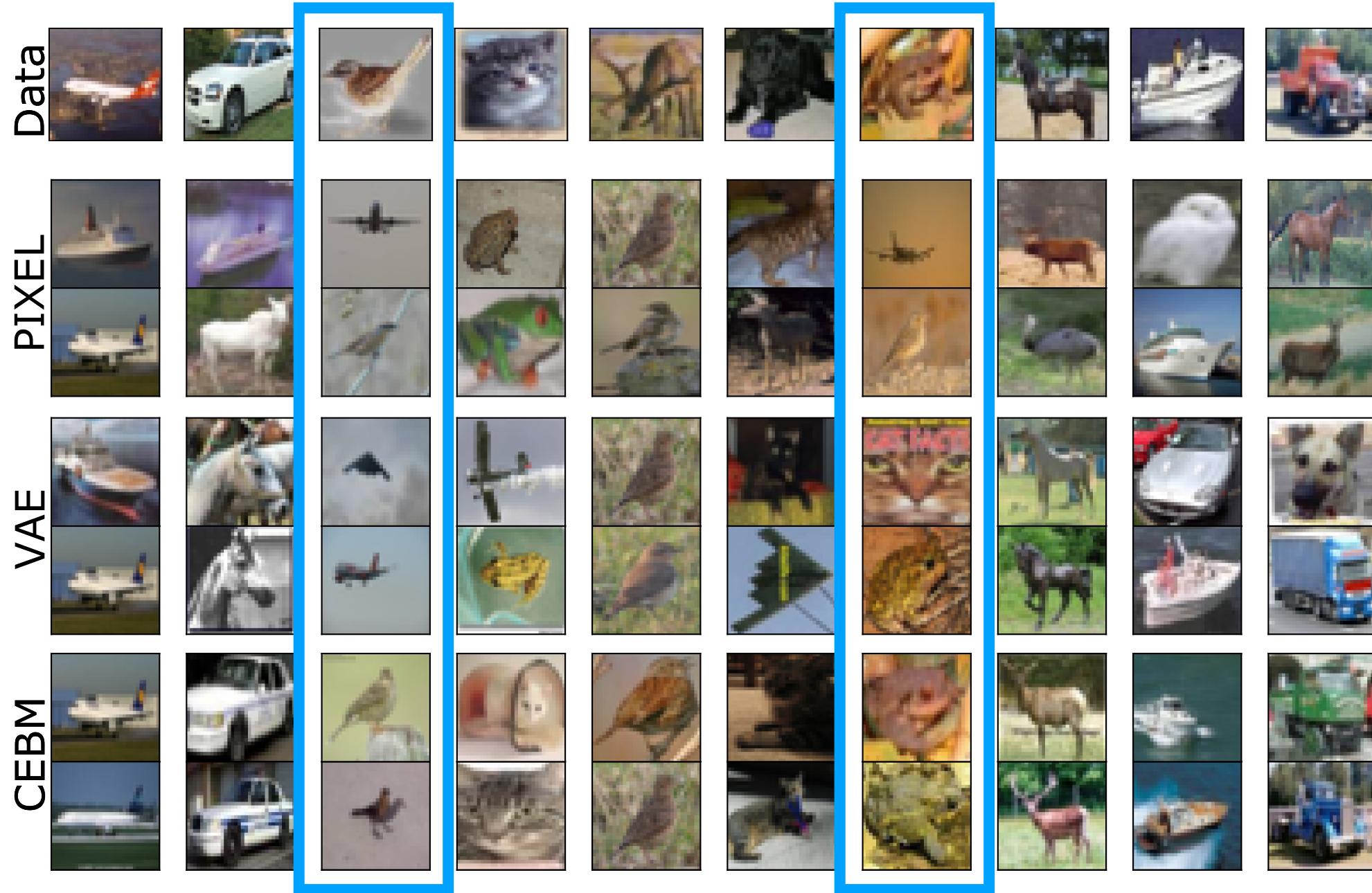


SVHN



# Alignment Between Learned Presentations and Class Labels

Nearest Neighbors



PIXEL

0.35

True Class Labels									
0	1	2	3	4	5	6	7	8	9
0 - 0.49	0.02	0.11	0.02	0.07	0.02	0.03	0.01	0.22	0.01
1 - 0.11	0.22	0.08	0.05	0.13	0.05	0.07	0.03	0.2	0.05
2 - 0.08	0.0	0.38	0.06	0.24	0.06	0.09	0.02	0.05	0.0
3 - 0.04	0.0	0.17	0.24	0.15	0.15	0.14	0.04	0.05	0.01
4 - 0.06	0.0	0.22	0.05	0.46	0.05	0.09	0.03	0.05	0.0
5 - 0.05	0.0	0.16	0.14	0.15	0.29	0.11	0.03	0.05	0.01
6 - 0.03	0.0	0.19	0.08	0.24	0.06	0.35	0.01	0.04	0.0
7 - 0.06	0.01	0.14	0.06	0.2	0.07	0.09	0.29	0.06	0.02
8 - 0.12	0.02	0.05	0.04	0.08	0.02	0.02	0.01	0.62	0.02
9 - 0.11	0.06	0.08	0.07	0.11	0.04	0.08	0.06	0.21	0.2

Confusion Matrices

VAE

0.38

True Class Labels									
0	1	2	3	4	5	6	7	8	9
0 - 0.49	0.04	0.09	0.03	0.05	0.03	0.05	0.05	0.14	0.03
1 - 0.07	0.41	0.05	0.05	0.05	0.05	0.05	0.04	0.06	0.09
2 - 0.08	0.02	0.34	0.09	0.16	0.08	0.11	0.05	0.04	0.03
3 - 0.04	0.04	0.12	0.23	0.1	0.16	0.15	0.08	0.04	0.04
4 - 0.05	0.02	0.18	0.08	0.32	0.07	0.14	0.09	0.02	0.03
5 - 0.04	0.02	0.11	0.14	0.12	0.3	0.12	0.1	0.03	0.03
6 - 0.03	0.02	0.11	0.14	0.12	0.08	0.44	0.05	0.01	0.03
7 - 0.05	0.04	0.1	0.06	0.12	0.07	0.07	0.4	0.04	0.05
8 - 0.13	0.07	0.05	0.04	0.05	0.02	0.03	0.04	0.5	0.07
9 - 0.09	0.15	0.05	0.04	0.06	0.05	0.06	0.06	0.09	0.36

CEBM

0.45

True Class Labels									
0	1	2	3	4	5	6	7	8	9
0 - 0.54	0.03	0.09	0.03	0.05	0.02	0.05	0.02	0.14	0.03
1 - 0.04	0.49	0.03	0.05	0.04	0.05	0.02	0.07	0.02	0.07
2 - 0.07	0.01	0.38	0.09	0.18	0.07	0.12	0.04	0.02	0.01
3 - 0.04	0.02	0.11	0.29	0.12	0.16	0.18	0.05	0.02	0.03
4 - 0.03	0.01	0.16	0.06	0.42	0.05	0.17	0.07	0.02	0.01
5 - 0.03	0.01	0.17	0.13	0.35	0.13	0.13	0.05	0.01	0.02
6 - 0.01	0.01	0.09	0.08	0.12	0.06	0.6	0.02	0.0	0.01
7 - 0.03	0.01	0.06	0.06	0.15	0.08	0.07	0.47	0.02	0.04
8 - 0.14	0.07	0.03	0.04	0.04	0.02	0.03	0.02	0.55	0.05
9 - 0.07	0.14	0.03	0.05	0.04	0.03	0.06	0.04	0.09	0.46

# Downstream Task: Few-label Classification

Models	MNIST				Fashion-MNIST				CIFAR-10				SVHN			
	1	10	100	full	1	10	100	full	1	10	100	full	1	10	100	full
VAE	42	85	92	95	41	63	72	81	16	22	31	38	<b>13</b>	13	16	36
GMM-VAE	53	86	93	97	49	68	79	84	<b>19</b>	23	33	39	<b>13</b>	14	23	56
BIGAN	33	67	85	91	46	65	75	81	18	<b>30</b>	<b>43</b>	52	11	20	42	56
IGEBM	63	89	95	97	50	<b>70</b>	79	83	16	26	33	42	10	16	35	49
CEBM	<b>67</b>	89	95	97	<b>52</b>	<b>70</b>	77	83	<b>19</b>	<b>30</b>	42	<b>53</b>	12	<b>25</b>	<b>48</b>	<b>70</b>
GMM-CEBM	<b>67</b>	<b>91</b>	<b>97</b>	<b>98</b>	<b>52</b>	<b>70</b>	<b>80</b>	<b>85</b>	16	29	42	52	10	17	39	60

Thank you!