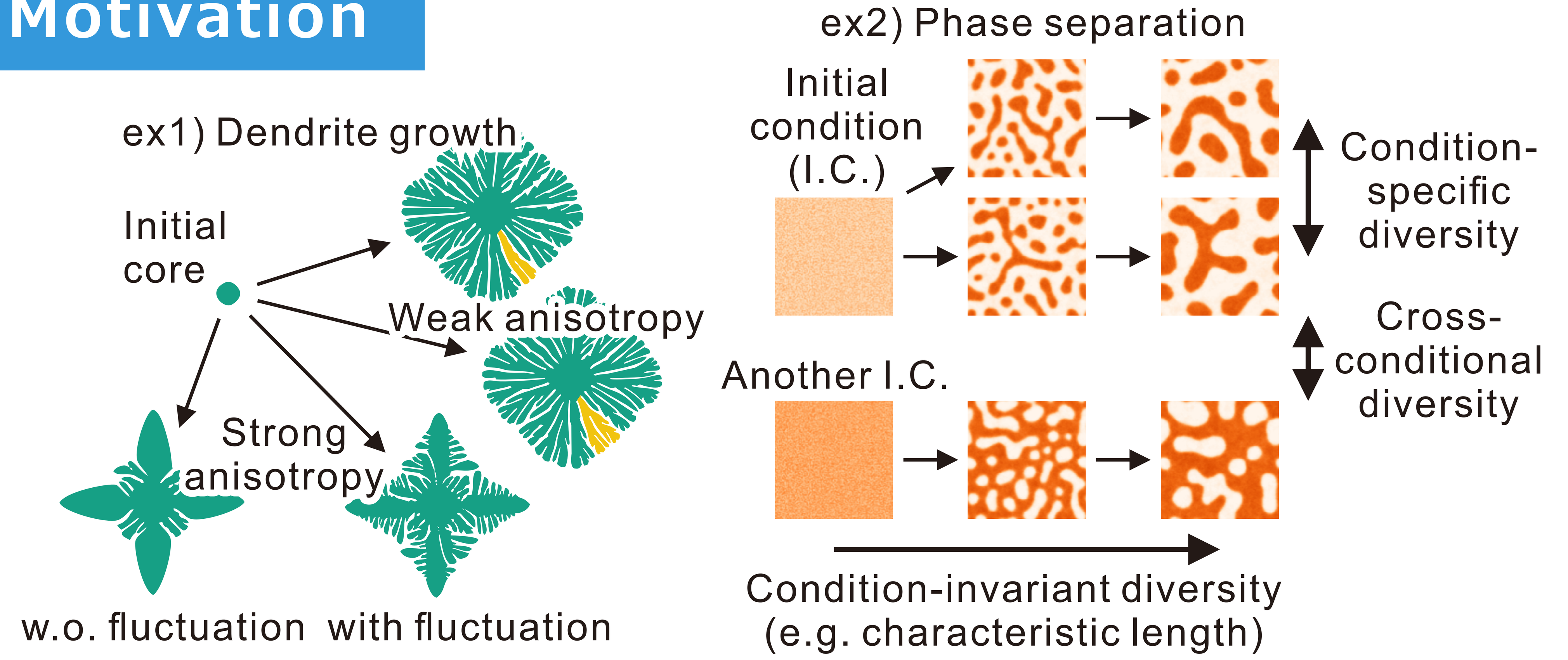
Flow-based Image-to-Image Translation with Feature Disentanglement

Ruho Kondo*, Keisuke Kawano, Satoshi Koide and Takuro Kutsuna (Toyota Central R&D Labs., Inc.) *r-kondo@mosk.tytlabs.co.jp

Motivation



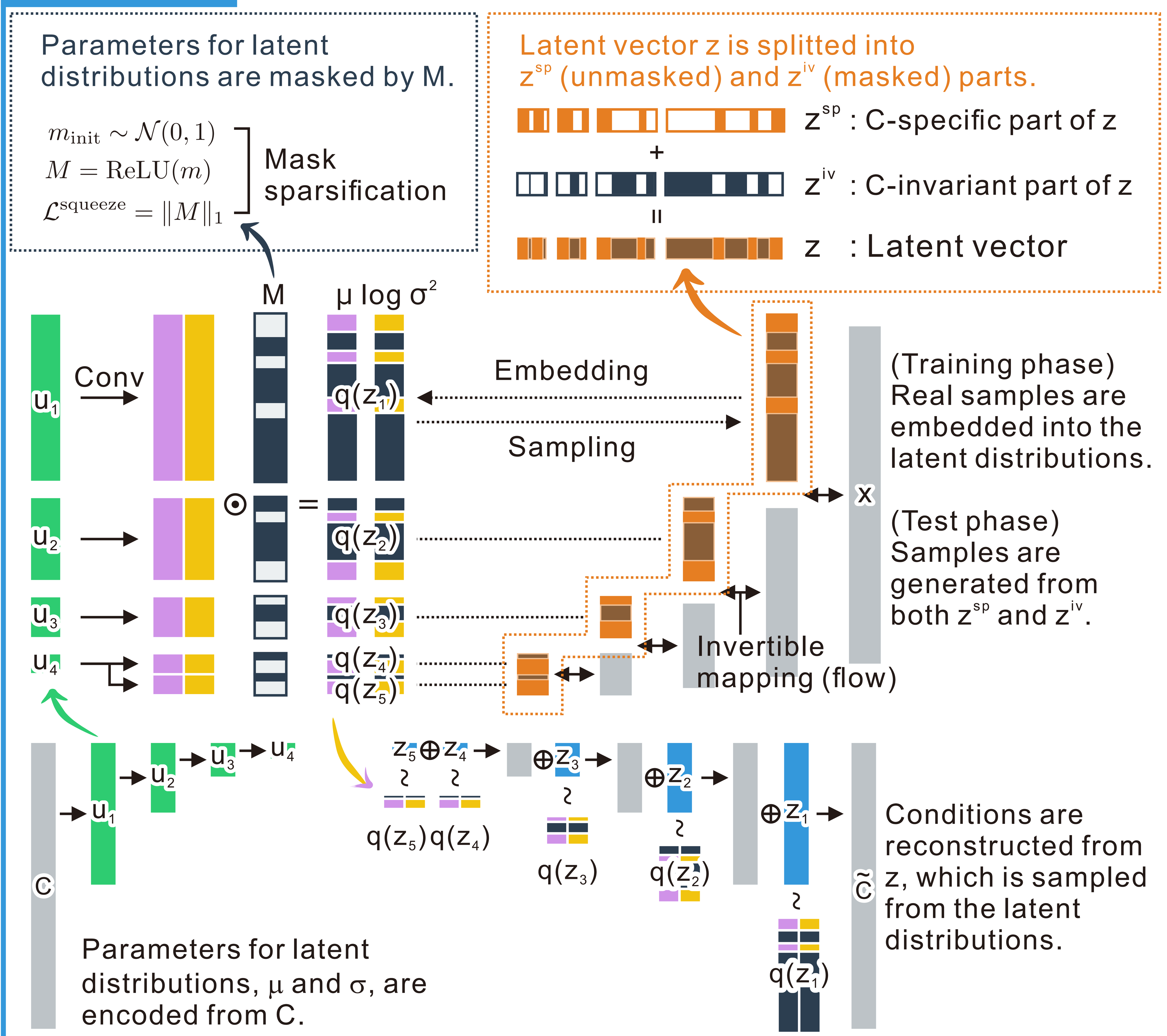
- Microstructure development can be seen as image-to-image translation.
- Some physical processes are stochastic and are induced by fluctuation.

⇒ **[Aim1] Generate diverse images conditioned on (initial) images.**

- Material properties strongly depends on microstructure.
- Analyzing the origin of uncertainty is important.

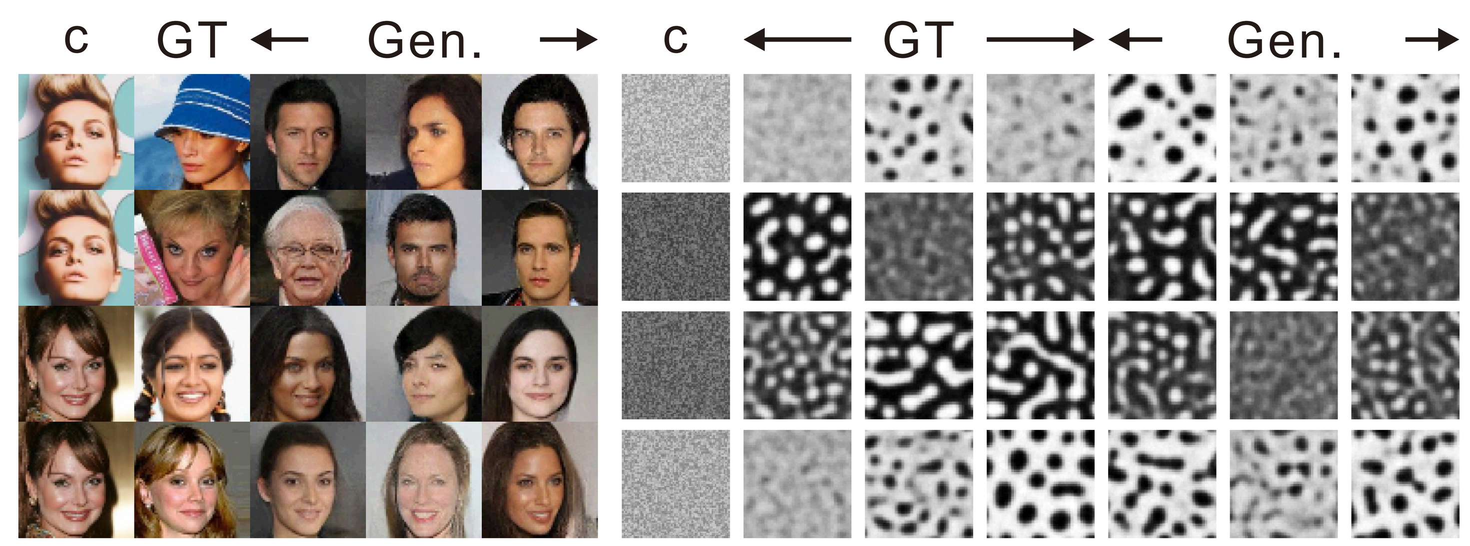
⇒ **[Aim2] Separate uncertainty into condition-specific/-invariant parts.**

Models

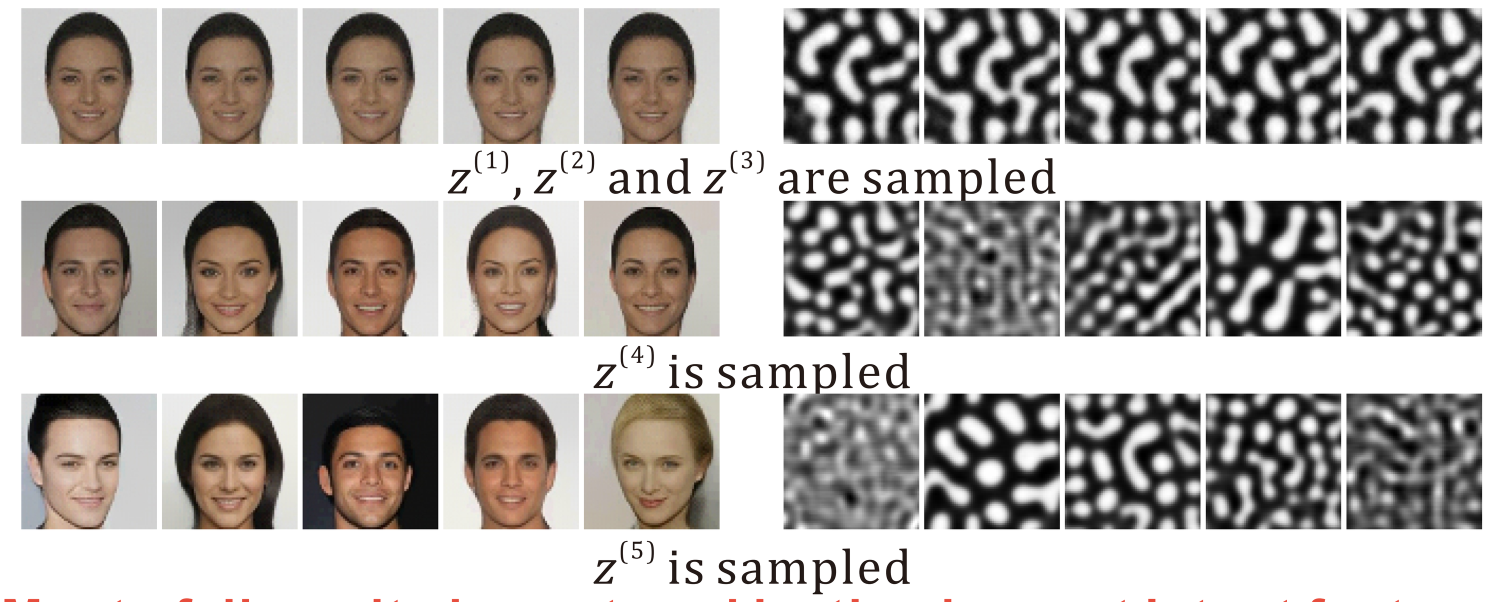


Results

Conditional generation

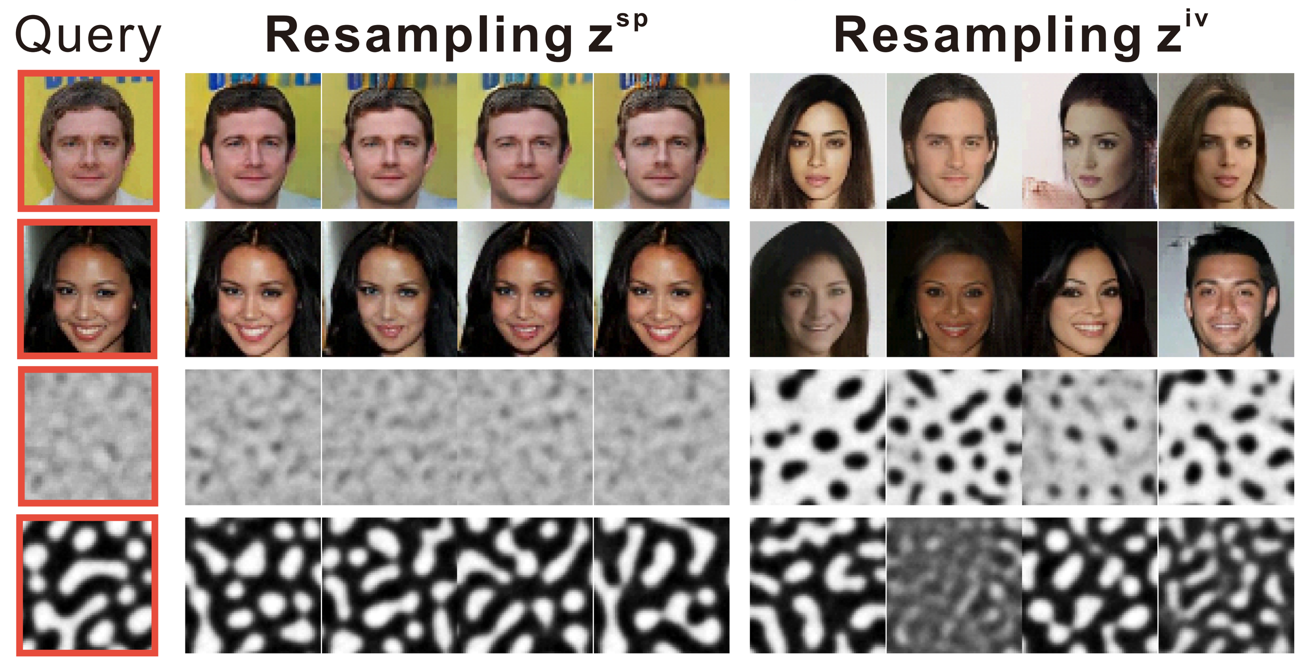


Multi-scale features

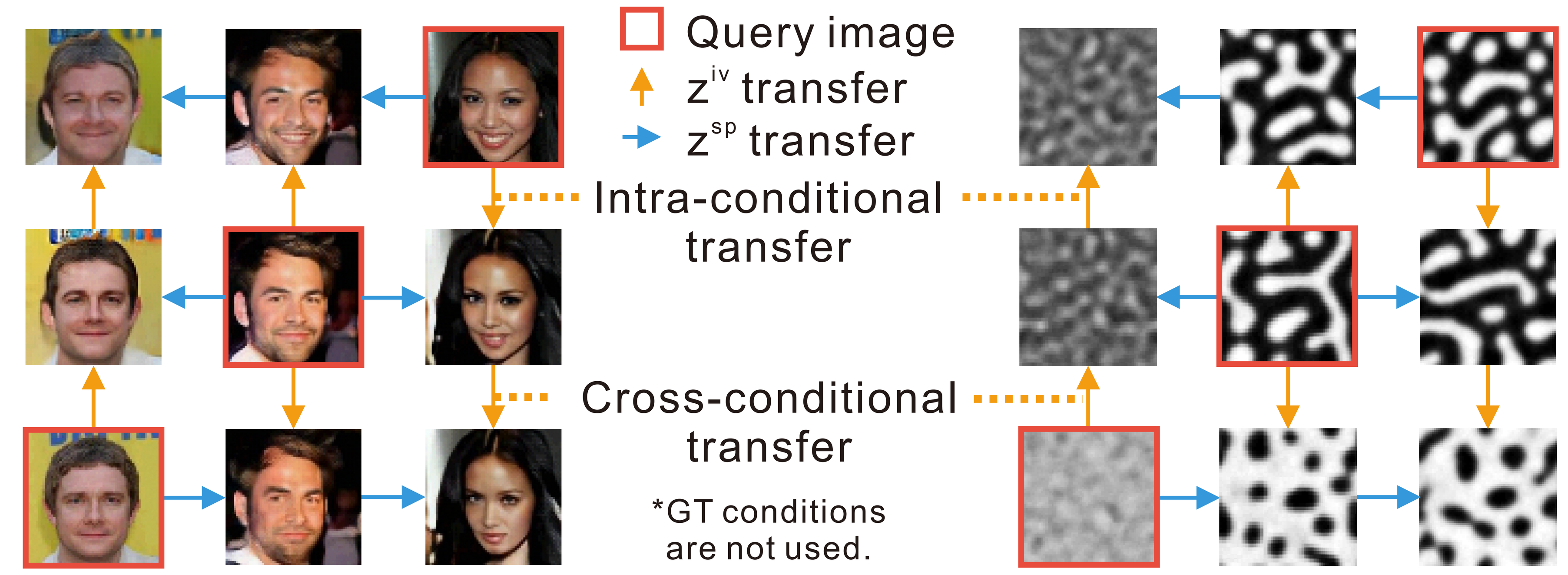


Most of diversity is captured by the deepest latent feature.

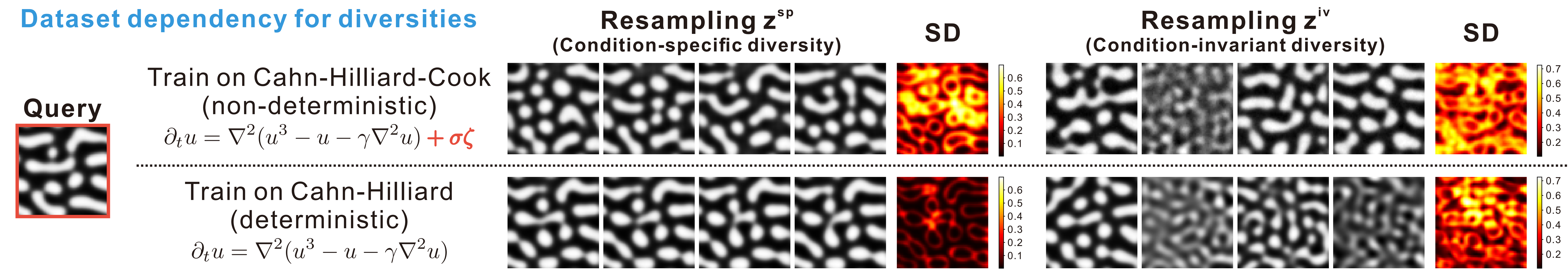
Feature disentanglement



Exchanging features between queries



Dataset dependency for diversities



Our model can learn whether there is condition-specific diversity (= environmental uncertaintity) or not.
[NOTE] Only one trajectory for each condition is used for training in both cases.

Ablation study

	CelebA		CHC	
	Acc.	dim($\mathbf{z}^{sp}$)	Err.	dim($\mathbf{z}^{sp}$)
$\mathcal{L}^{\mathrm{flow}}$	0.956	5,525	9.92	2,062
$+\mathcal{L}^{\mathrm{recons}}$	0.981	5,228	9.62	1,977
$+\mathcal{L}^{\mathrm{squeeze}}$	0.978	86	9.72	270
$+\mathcal{L}^{\mathrm{entropy}}$	0.977	113	9.67	111

$\mathcal{L}^{\mathrm{recons}}$ **: Improving $I(\mathbf{c};\mathbf{x})$**
$\mathcal{L}^{\mathrm{squeeze}}$ & $\mathcal{L}^{\mathrm{entropy}}$ **: Reducing $\dim(\mathbf{z}^{sp})$**

$$\mathcal{L}^{\mathrm{flow}} = \mathbb{E}_{x,c\sim p(x,c)}\left[-\log e_\phi(f_\theta(x)|c) - \sum_p \log J_p\right]$$

$$\mathcal{L}^{\mathrm{recons}} = \mathbb{E}_{c\sim p(c), z\sim e_\phi(z|c)}\left[-\log d_\psi(c|z)\right]$$

$$\mathcal{L}^{\mathrm{squeeze}} = \|M\|_1 \qquad \mathcal{L}^{\mathrm{entropy}} = \sum_{i=1}^{d_x} M_i \mathcal{H}[e_\phi(z_i|c)]$$

$$\mathcal{L}^{\mathrm{FUNS}} = \mathcal{L}^{\mathrm{flow}} + \mathcal{L}^{\mathrm{recons}} + \alpha\mathcal{L}^{\mathrm{squeeze}} + \beta\mathcal{L}^{\mathrm{entropy}}$$

Quantitative comparison

FID : Fréchet Inception Distance.
c-LPIPS : Intra-conditional Learned Perceptual Image Patch Similarity.
Acc.: Binary classification accuracy whether generated data is smile or not.
Err. : L^2 distance between predicted conditions from generated data and ground truth condition.

		CelebA			CHC		
		FID	c-LPIPS	Acc.	FID	c-LPIPS	Err.
VUNet	$T=1.0$	66.0±4.3	0.146	0.977	96.5±2.1	0.118	9.46
VUNet	$T=0.8$	81.7±3.7	0.103	0.981	164.0±5.7	0.113	**9.44**
PUNet	$T=1.0$	114.8±9.2	0.180	**1.000**	225.7±6.1	0.108	9.48
PUNet	$T=0.8$	117.2±5.5	0.146	**1.000**	227.9±6.0	0.088	9.48
Ours	$T=1.0$	39.6±3.9	**0.262**	0.973	**10.5 ± 2.0**	**0.157**	9.67
Ours	$T=0.8$	**29.5 ± 3.5**	0.256	0.967	11.1±2.5	0.155	9.81
Real data	-	-	0.284	0.924	-	0.169	9.54

Small FID / Large c-LPIPS / Comparable Acc. or Err.
⇒ Realistic diverse conditional image generation

*Err.=22.84 for random labels