Image Generation using Continuous Filter Atoms

Dec 26, Jeong Hyun Jae, 논문 익는 마을 (KU BME JC)

Hyunjae Jeong

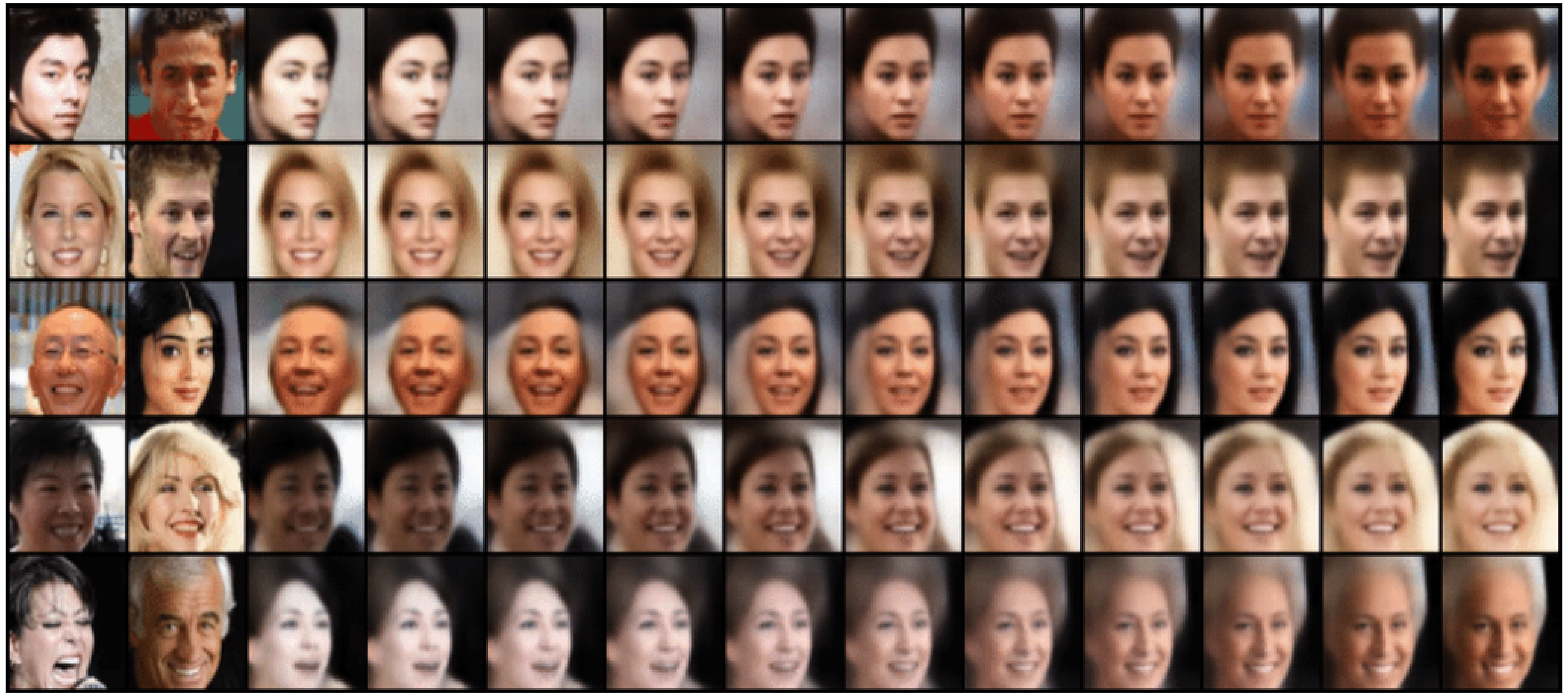
Yonsei University, Medical Life Systems Information Center (TAIL Lab)
Severance Hospital, Center for Clinical Imaging Data Science (CCIDS)
Severance Hospital, Radiology

CCIDS 의료영상데이터사이언스센터 Center for Clinical Imaging Data Science
tAILab.

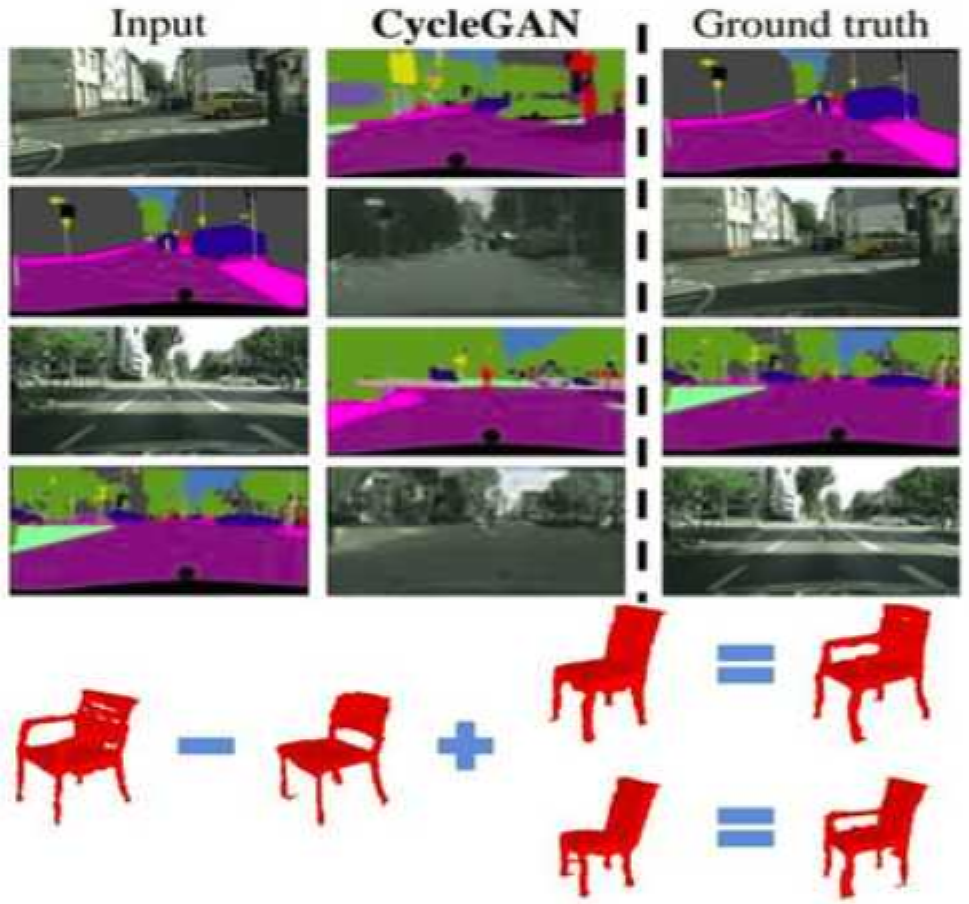
01 Introduction

Conditional Image Generation

- Conditional image generation has been widely studied in recent years due to their wide applications including i mage segmentation, style transfer, image inpainting, image super-resolution, image registration, and image sy nthesis
- Despite extensive research and application in these fields, limited progress has been made on conditional imag e generation using continuous or closely spaced labels due to difficulties pointed out in



Continuous Image Generation



Conditional Image Generation

Limitation of Image Generation and Diversity

- It remains challenging to encourage generation diversity, especially without heavy supervision, while maintaining output fidelity.
- explicit regularizations inevitably introduce additional hyperparameters, and sub-optimal hyperparameters can often cause either poor diversity or noticeable sacrifices in terms of generation quality and the correspondenc e to the input condition
- This direction is mainly restricted by the prohibitive cost of both modeling and sampling of the very high-dimen sional space of convolutional filters in modern image generative networks

↓

- The observation that a convolutional filter can be well approximated by a linear combination of low-dimensional filter atoms.
- Significantly reduces the cost of both parameter modeling and sampling. In this way, each sampling of the mod eled parameter subspace results in one deterministic transformation from the input condition to the desirable target domain, and the diverse outputs are achieved by sampling multiple times.

01 Introduction

Limitation of Low Dimensional Subspace Filters

- First, this method is sensitive to network configurations, and mode collapse, which results in a point estimation to the subspace of parameter, can often be observed.
- This method is incapable of modeling a continuous space, so that gradual changes to the generated parameter s cannot be obtained by simply interpolating samples; and such discontinuity will be further propagated to the output

↓

- We adopt a subspace view to convolutional filters by performing atom-coefficients decomposition
- Then we further model the filter subspace using neural ordinary differential equation (ODE), so that we are abl e to produce a continuous series of filter samples at arbitrarily fine resolution.

↓

Neural Ordinary Differential Equation
Ricky T. Q, Yulia Rubanova, Jesse Bettencourt, David Duvenaud ,NeurIPS,2018

01 Introduction

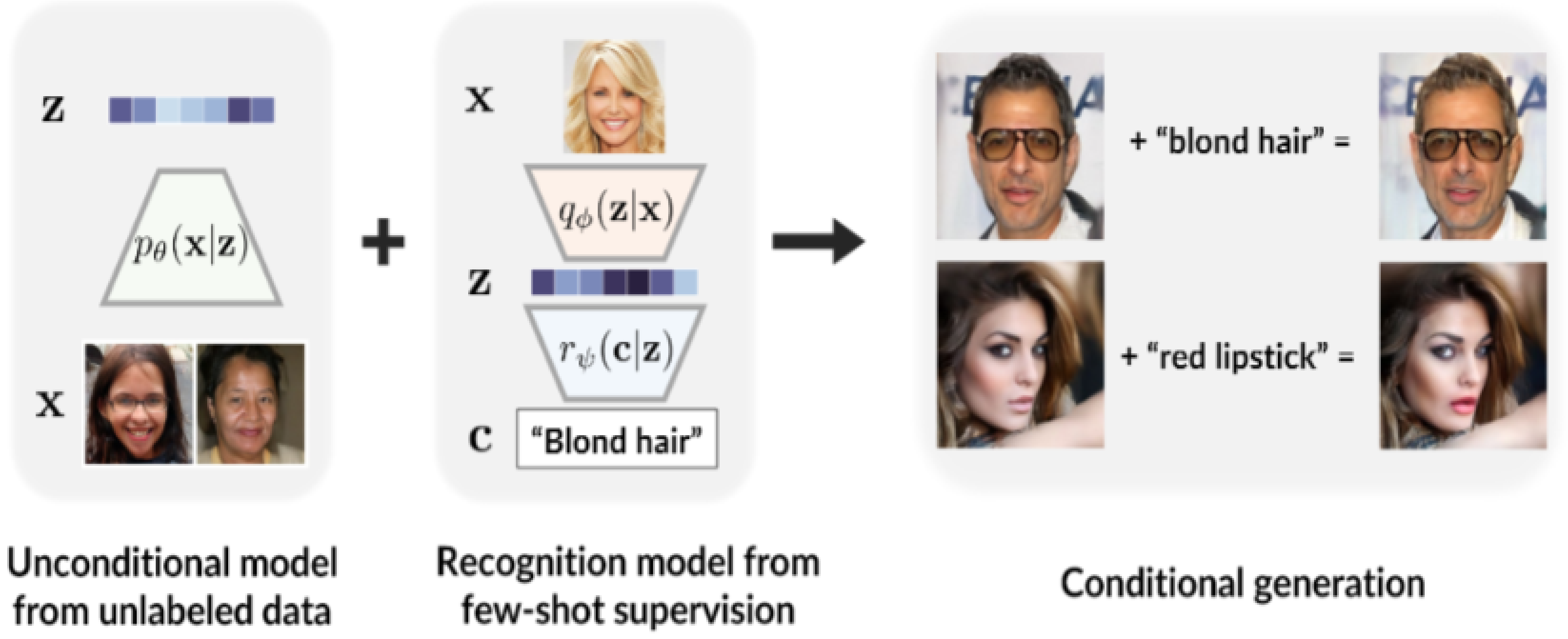
Contribution

- Continuous image synthesis that covers a wide range of gradually varying appearance with high fidelity a nd accurate correspondence to the input condition.
- Sequential image synthesis with explicitly specified starting and ending points to allow flexible appearanc e manipulation without heavy supervisions
- Interpolation of generated image appearance at arbitrarily fine resolution.
- An effective approach for generating images conditioned on continuous labels.

Related Work - image generation

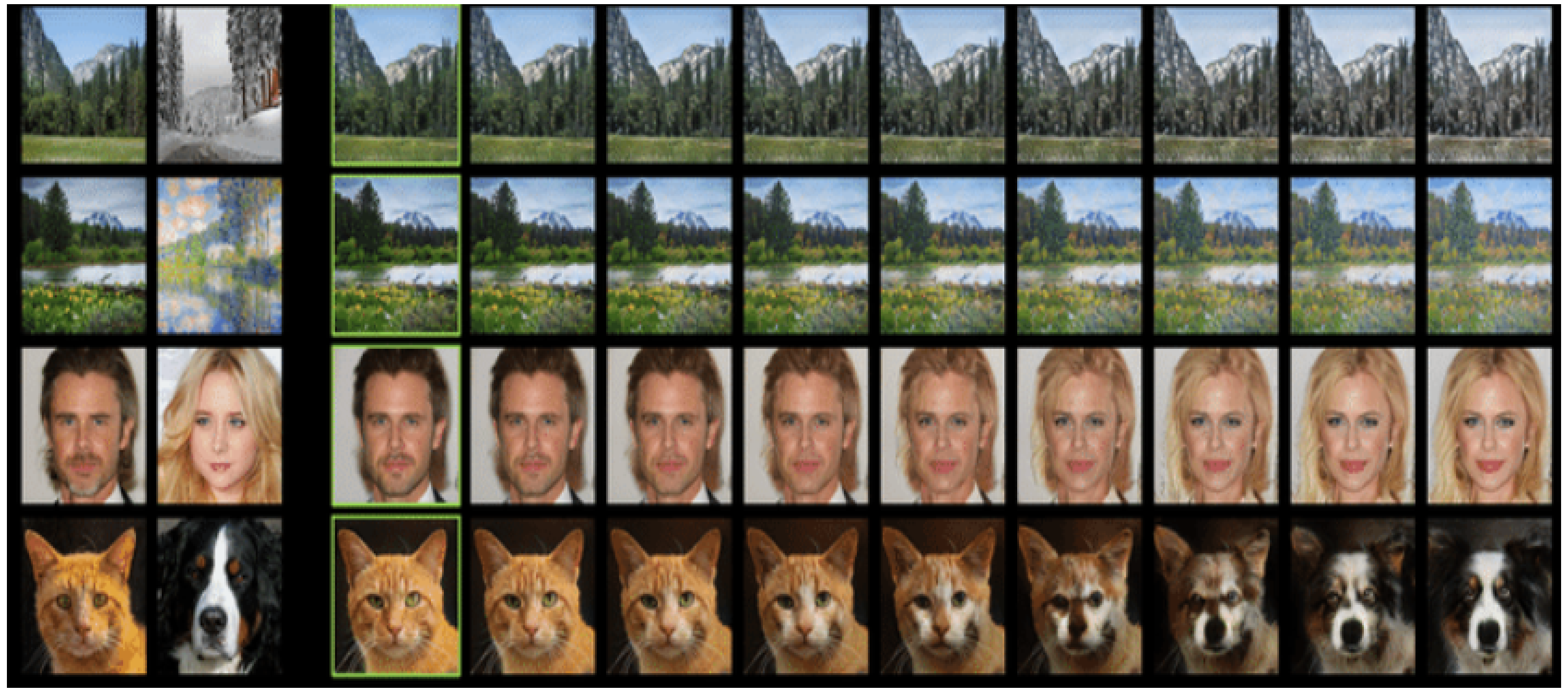
<Conditional image generation>

Recently, conditional image generation with condition o n continuous labels has attract attention, and new traini ng schemes are proposed to improve continuous condit ional image generation. From the perspective of the dy namic of parameter subspace, our method provides a si mple yet effective solution.



<Continuous image translation>

continuous image translation that aims to generate inte rmediate images between the source and the target im age is not yet widely studied. Previous works have mani pulated the attribute vectors either by interpolation or li near transformations to control the latent space

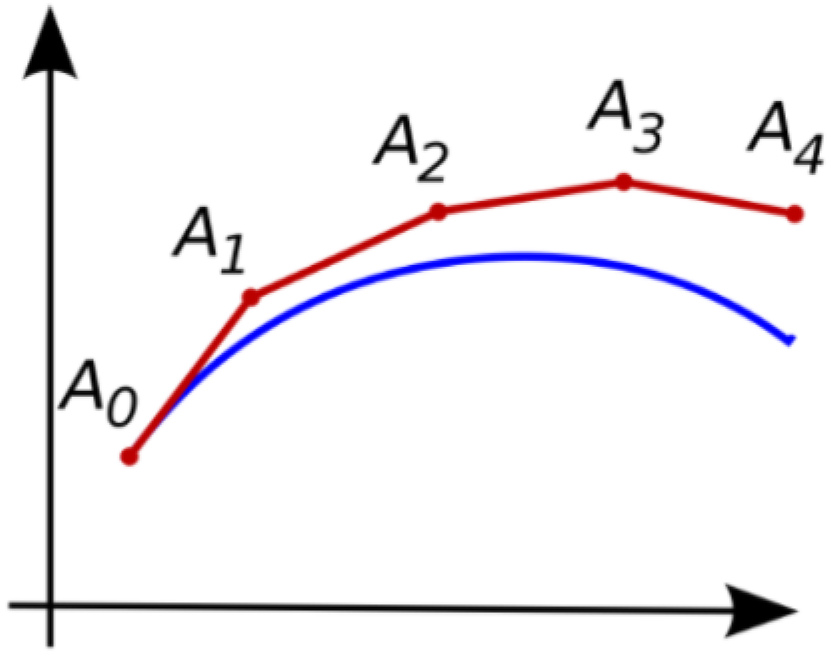


02 Methods

Continuous Filter Atoms - Neural ODE

- Neural ordinary differential equations (ODEs) are introduced in as an approach to interpret the neural network as a system of ODEs, where each ODE represents the underlying dynamics of the hidden elements
- Neural ODEs model a latent state $\mathbf{z}(t_s)$ as $\mathbf{z}(t_s) = \mathbf{z}(t_0) + \int_{t_0}^{t_s} f(\mathbf{z}(t), t; \boldsymbol{\theta})\, dt$, where $\frac{d\mathbf{z}(t)}{dt} = f(\mathbf{z}(t), t; \boldsymbol{\theta})$ is modeled as a neural network parametrized by $\boldsymbol{\theta}$. Then, the latent state at an arbitrary point t_s can be obtained as

<Euler's Method>



$$\mathbf{z}(t_s) = \mathrm{ODESolve}(\mathbf{z}(t_0), f, (t_0, t_s), \theta),$$

<t-th layer output h_t in Neural Network>

$$\mathbf{h}_t = \mathbf{h}_{t-1} + F(\mathbf{h}_{t-1})$$ - Resnet

$$\begin{bmatrix} \mathbf{a}_t \\ \mathbf{x}_{t+1} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_{t-1} \\ \mathbf{x}_t \end{bmatrix} + F\left(\begin{bmatrix} \mathbf{a}_{t-1} \\ \mathbf{x}_t \end{bmatrix}\right)$$ - RNN

<Runge-Kutta Method>

$$y_{n+1} = y_n + h \sum_{i=1}^{s} b_i k_i$$

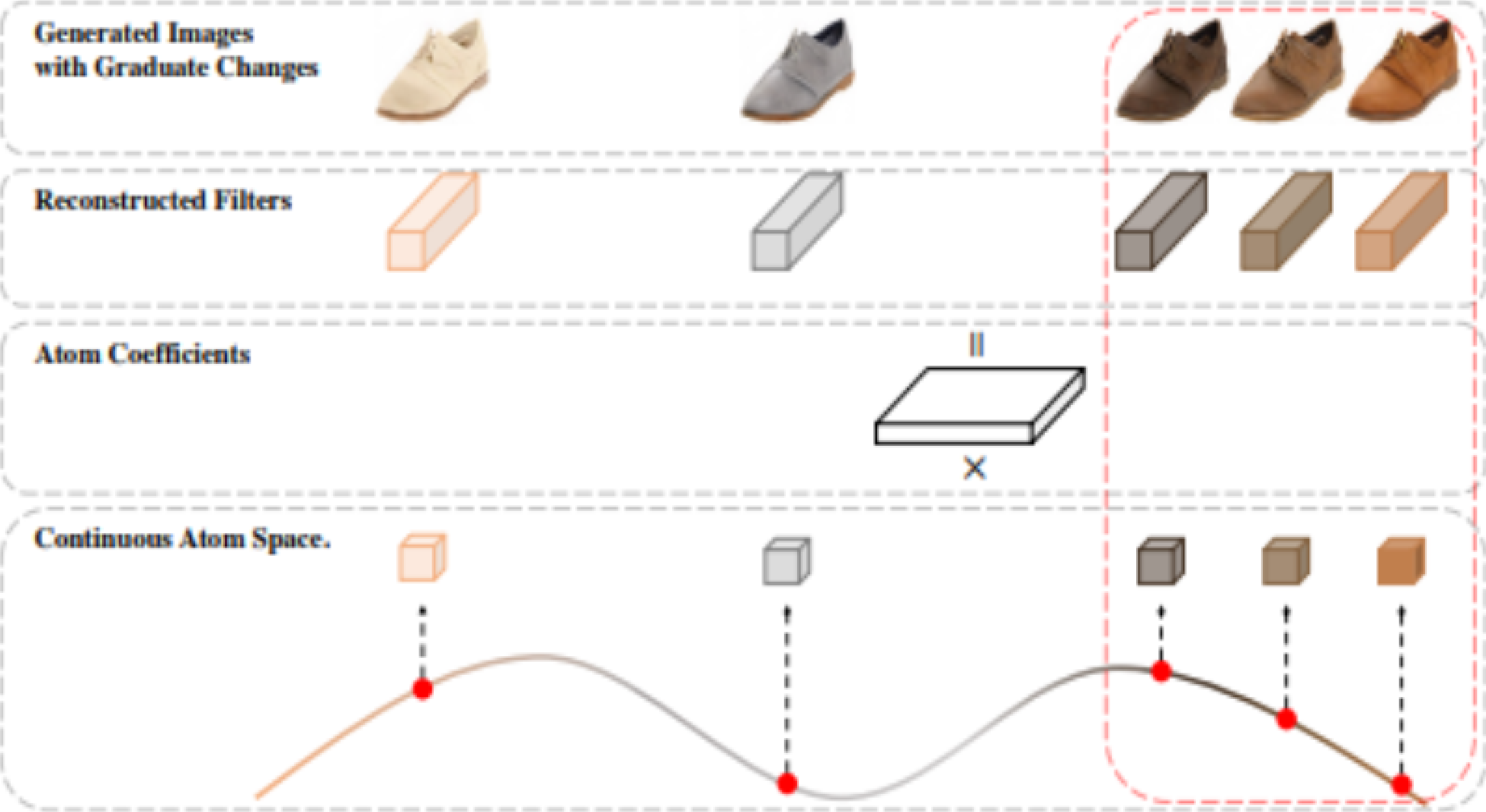
02 Methods

Continuous Filter Atoms - Convolutional Filter Decomposition

- Directly modeling the space of high-dimensional convolutional filters in an image generation network is practic ally prohibitive in terms of both computation and parameter scale
- Inspired by the observation that a convolutional filter can be well-approximated as a linear combination of filter bases - achieve efficient training and modeling of gradual changes in conditional image generations.

F = AD

<Continuous Atom Generator>



<Convolutional Filter Decomposition>

$$\mathcal{D}(t_s) = \mathcal{T}(t_s; \mathcal{D}(t_0), \theta) = \mathcal{D}(t_0) + \int_{t_0}^{t_s} f(\mathcal{D}(t), t; \theta) dt.$$

$$\frac{d\mathcal{D}(t)}{dt} = f(\mathcal{D}(t), t; \theta).$$

F = A l by l-sized convolutional filter
A = The filter subspace coefficients.
D = decompose the filter over a set of m atoms (respective atoms)
Tau = continuous atom Generator

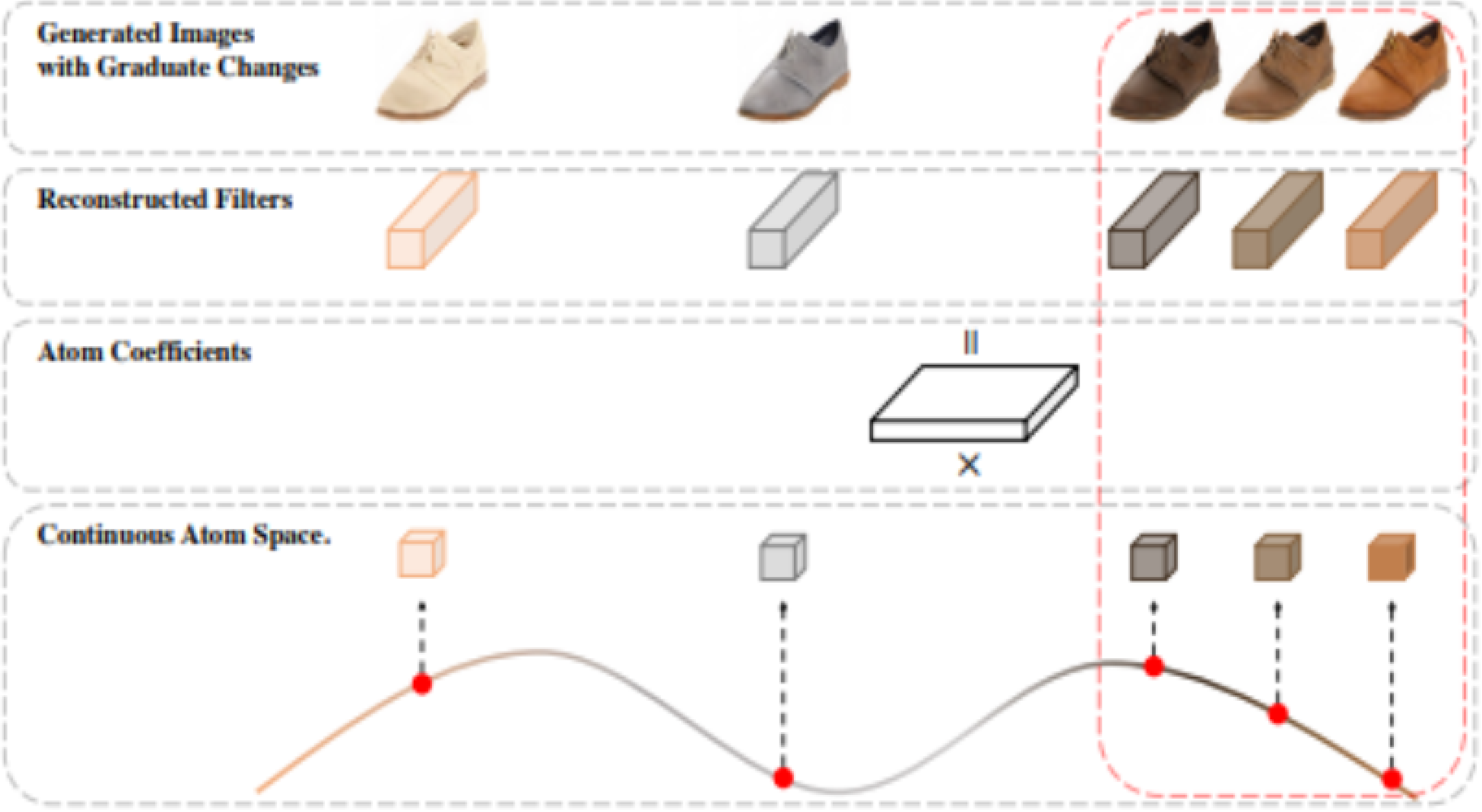
02 Methods

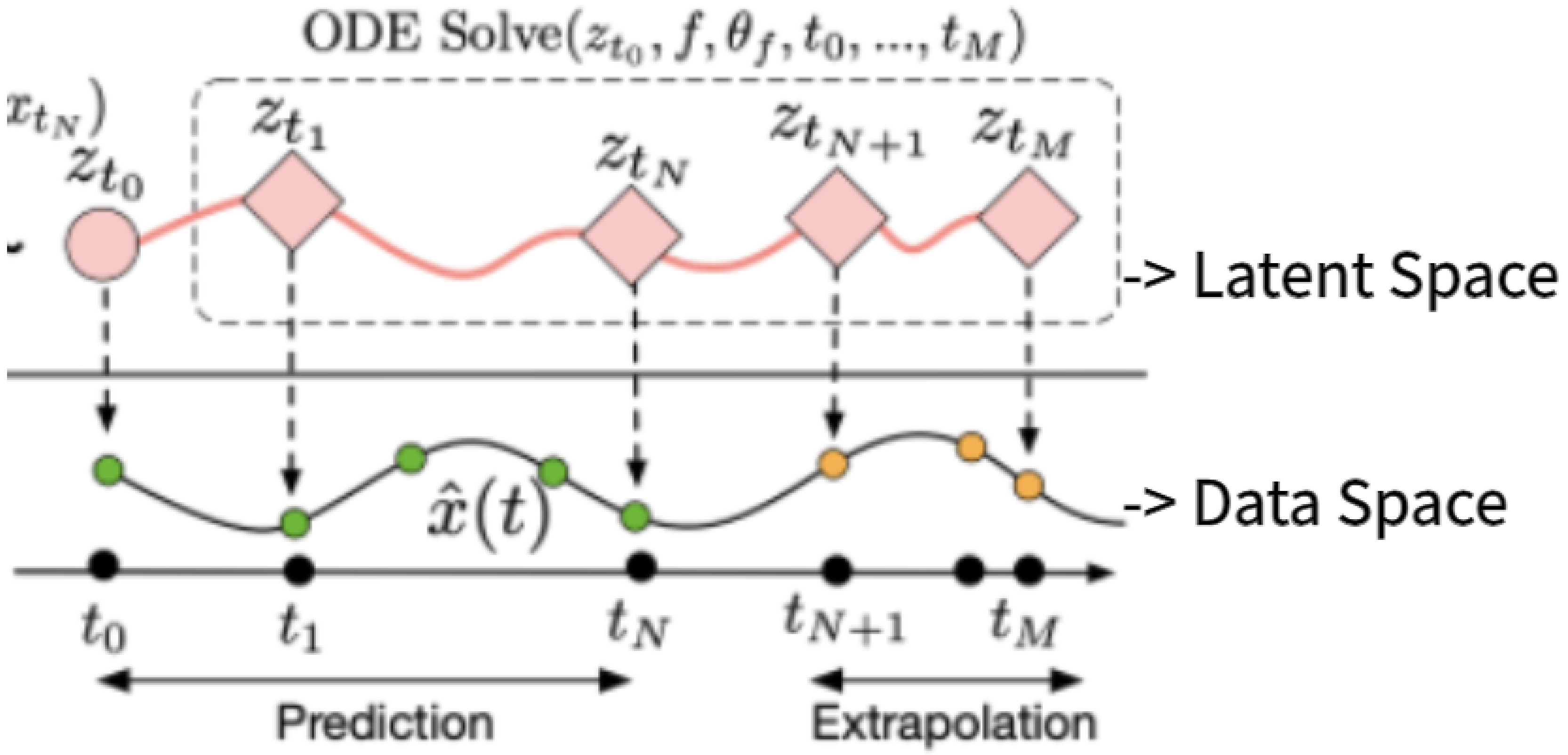
Continuous Atom Generation for Smooth Appearance Modeling

- we target for modeling gradual changes in generated images with respect to continuous input conditions.
  input conditions to the corresponding output with parameters instantiated from an underlying continuous spa
  ce, which subsequently leads to generated images with gradual changes

$$X^i(u) = \mathcal{F}^i(X^{i-1}, \mathbf{D}^i; \mathbf{A}^i) = \sigma(\sum_{k=1}^{m} \mathbf{A}_k^i \langle X^{i-1}, \mathbf{D}^{i,k} \rangle_{N_u} + b^i),$$

X(u) = Given an input image
Nu = local Euclidean grid centered at u
Fi = The i-th layer function

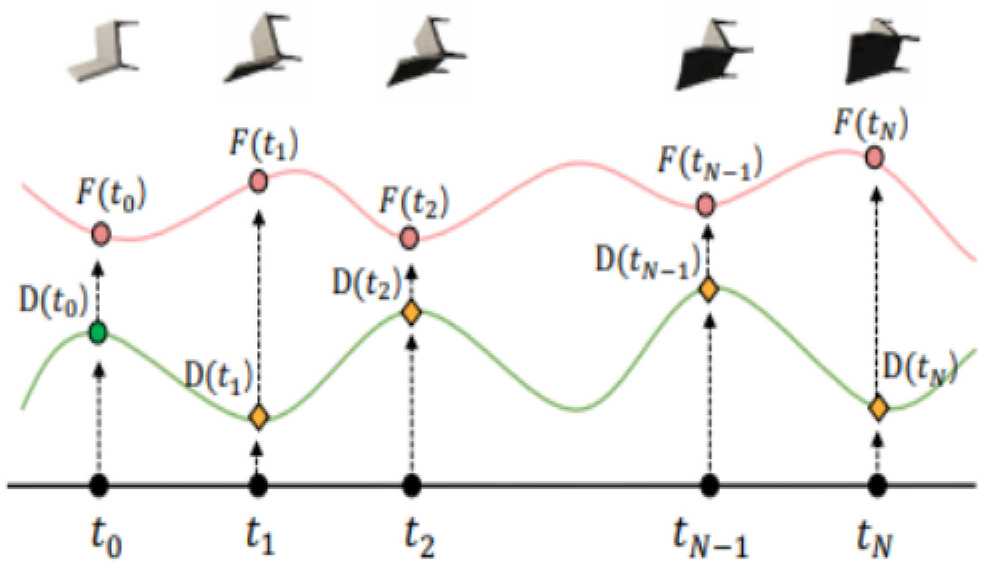


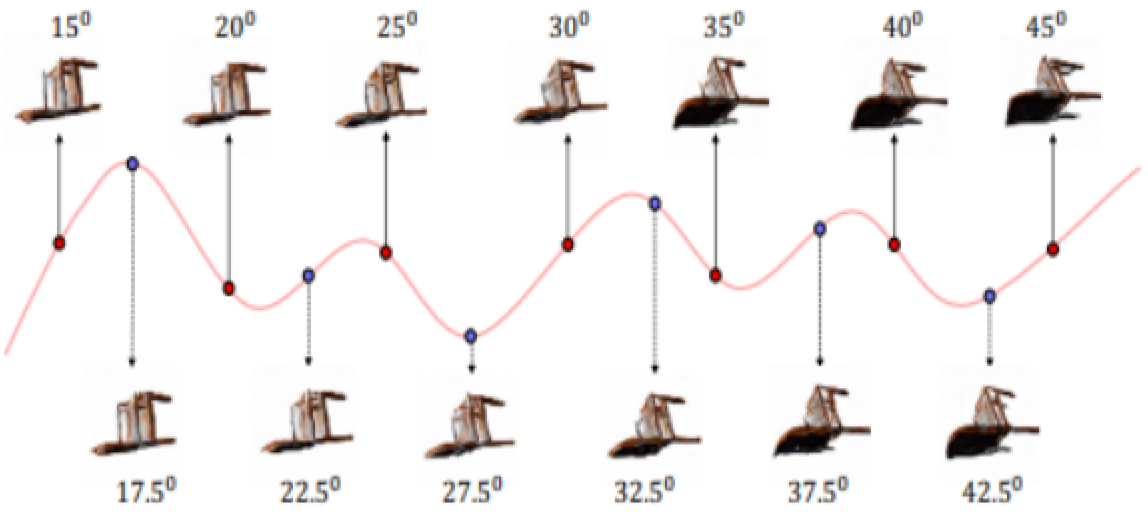


02 Methods

Continuous Atom Generation for Smooth Appearance Modeling

- we target for modeling gradual changes in generated images with respect to continuous input conditions.
  input conditions to the corresponding output with parameters instantiated from an underlying continuous spa ce, which subsequently leads to generated images with gradual changes





Theorem 1. *Suppose $\mathbf{D}_1^i$ and $\mathbf{D}_2^i$ are two continuously generated atoms, (i.e., $\exists M > 0, ||\mathbf{D}_1^i - \mathbf{D}_2^i||_2 \leq M|t_1 - t_2|$), and assume the activation function σ is non-expansive which holds for ReLU, then $\mathcal{F}^i$ is continuous in $\mathbf{D}^i$,*

$$||X_1^i - X_2^i||_2 \leq (||\mathbf{A}^i||_2 \lambda)\sqrt{|\mathcal{U}|} \cdot ||(\mathbf{D}_1^i - \mathbf{D}_2^i)||_2, \quad \textit{with } \lambda = \sup_{u \in \mathcal{U}} ||X^{i-1}||_{2,N_u}, \tag{5}$$

in which $X_1^i = \mathcal{F}^i(X^{i-1}, \mathbf{D}_1^i; \mathbf{A}^i)$, $X_2^i = \mathcal{F}^i(X^{i-1}, \mathbf{D}_2^i; \mathbf{A}^i)$ are outputs correspond to two atoms.

Proof. The proof is provided in the Appendix.

Ending point of the ODE integral

1)explicitly given to strictly control t he behavior of continuous dynamics

2) stochastically sampled from a pri or distribution, e.g., Uniform distrib ution or Gaussian distribution, to en hance the diversity of the generate d images.

Then, proposed method is much mo re straight-forward than the previo us conditional GAN methods

02 Methods

Conditional image generation with continuous atoms

- Conditional on images

$$\min_{\mathbb{G}} \max_{\mathbb{D}} \mathcal{L}(\mathbb{D}, \mathbb{G}) = \mathbb{E}_{\mathbf{x}\sim p(\mathbf{x}),\mathbf{y}\sim p(\mathbf{y}|\mathbf{x})}\left[\log\left(\mathbb{D}(\mathbf{x},\mathbf{y})\right)\right] + \mathbb{E}_{\mathbf{x}\sim p(\mathbf{x}),\boxed{t\sim \mathrm{Uniform}(t_0,t_{-1})}}\left[\log\left(1-\mathbb{D}\Big(\mathbf{x}, \mathbb{G}_{\phi,\theta,\boldsymbol{\mathcal{D}}_0}\left(\mathbf{x}; \mathcal{T}(t;\boldsymbol{\mathcal{D}}_0,\theta)\right)\Big)\right)\right],$$

→ empirical distribution p(x)
aims at p(ylx)
Time Series -> Unif(t0,t-1)

- Conditional on labels

$$\min_{\mathbb{G}} \max_{\mathbb{D}} \mathcal{L}(\mathbb{D}, \mathbb{G}) = \mathbb{E}_{x\sim p(x),\mathbf{y}\sim p(\mathbf{y}|x)}\left[\log\left(\mathbb{D}(x,\mathbf{y})\right)\right] + \mathbb{E}_{x\sim p(x),\boxed{\mathbf{z}\sim \mathcal{N}(0,\mathbf{I})}}\left[\log\left(1-\mathbb{D}\Big(x, \mathbb{G}_{\phi,\theta,\boldsymbol{\mathcal{D}}_0}\left(\mathbf{z}; \mathcal{T}(\alpha x;\boldsymbol{\mathcal{D}}_0,\theta)\right)\Big)\right)\right].$$

→ input vector, randomness by N(0,I)
alpha = predefined scalar, linearly scales the input continuous label

- Conditional on both images and labels

$$\min_{\mathbb{G}} \max_{\mathbb{D}} \mathcal{L}(\mathbb{D}, \mathbb{G}) = \mathbb{E}_{\mathbf{x}\sim p(\mathbf{x}),\mathbf{y}\sim p(\mathbf{y}|\mathbf{x},x)}\left[\log\left(\mathbb{D}(\mathbf{x},\mathbf{y})\right)\right] + \mathbb{E}_{\mathbf{x}\sim p(\mathbf{x}),\boxed{x\sim p(x)}}\left[\log\left(1-\mathbb{D}\Big(\mathbf{x}, \mathbb{G}_{\phi,\theta,\boldsymbol{\mathcal{D}}_0}\left(\mathbf{x}; \mathcal{T}(\alpha x;\boldsymbol{\mathcal{D}}_0,\theta)\right)\Big)\right)\right].$$

→ None of them can adopt continuous raw labels
the first work to present gradually changing both images and continuous labels.

Experiments

- Continuous image-to-image translation conditional on images

  - we target for modeling gradual changes in generated images with respect to continuous input conditions.
  input conditions to the corresponding output with parameters instantiated from an underlying continuous spa
  ce, which subsequently leads to generated images with gradual changes

Figure 4: Continuous image-to-image translation. The network is trained without any auxiliary loss functions or regularizations. From top to bottom, the image to image translation tasks are: nights → days, edges → handbags, and edges → shoes.

Table 1: Quantitative results on image generation conditional on images only.

Methods	Labels→Facedee		Map→Satellite		Night→Day		Edge→Shoe		Edge→Handbag	
	LPIPS ↑	FID ↓	LPIPS ↑	FID ↓	LPIPS ↑	FID ↓	LPIPS ↑	FID ↓	LPIPS ↑	FID ↓
BicycleGAN	0.1413	98.85	0.1150	145.78	0.103	120.63	0.139	72.49	0.184	96.28
MSGAN	0.1894	92.84	0.2189	152.43	0.176	107.90	0.167	**60.28**	0.228	89.96
BasisGAN	**0.2648**	88.70	**0.2417**	**35.54**	**0.184**	**102.56**	**0.242**	64.17	0.350	88.76
Ours	0.2274	**87.75**	0.2217	55.76	0.170	106.95	0.237	61.55	**0.361**	**85.57**

learned perceptual
image patch similarity (LPIPS)

1)focused on human visual percepti
on and perceptual distance metrics

2) l2, Euclidean distance, PSNR meas
ure assume pixel-wise independenc
e so that not proper to evaluate str
uctred outputs like images.

3) This metrics can assess perceptua
l similarity better.

03 Experiments

Experiments

- Continuous image-to-image translation conditional on images

  - we target for modeling gradual changes in generated images with respect to continuous input conditions.
  input conditions to the corresponding output with parameters instantiated from an underlying continuous spa
  ce, which subsequently leads to generated images with gradual changes

Datasets	RC49		RCA20		RA20		RL20		RT20	
Methods	Pix2Pix	Ours	Pix2Pix	Ours	Pix2Pix	Ours	Pix2Pix	Ours	Pix2Pix	Ours
LPIPS ↑	-	0.1127	-	0.0931	-	0.068	-	0.1089	-	0.1213
FID ↓	251.49	**195.01**	218.88	**85.06**	165.01	**73.82**	227.04	**124.83**	206.49	**70.92**

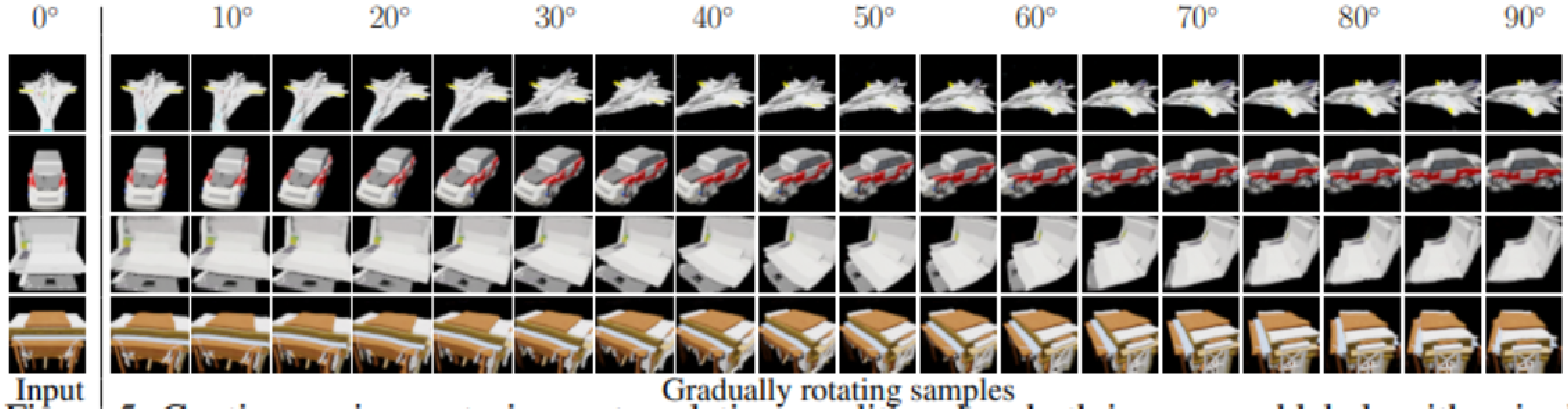


Figure 5: Continuous image-to-image translation conditional on both images and labels with paired samples. From top to bottom, the trained datasets are: RA-20, RCA-20, RL-20, and RT-20. The generated images of all datasets rotate from a degree 0.1 to 89.9 at the interval of 5 degree, along the yaw axis. All models are trained with Pix2Pix by plugging in our continuous atom generator.

learned perceptual
image patch similarity (LPIPS)

1)focused on human visual percepti
on and perceptual distance metrics

2) l2, Euclidean distance, PSNR meas
ure assume pixel-wise independenc
e so that not proper to evaluate str
uctred outputs like images.

3) This metrics can assess perceptua
l similarity better.

03 Experiments

Experiments

- Continuous image-to-image translation conditional on images

  - we target for modeling gradual changes in generated images with respect to continuous input conditions.
  input conditions to the corresponding output with parameters instantiated from an underlying continuous spa
  ce, which subsequently leads to generated images with gradual changes

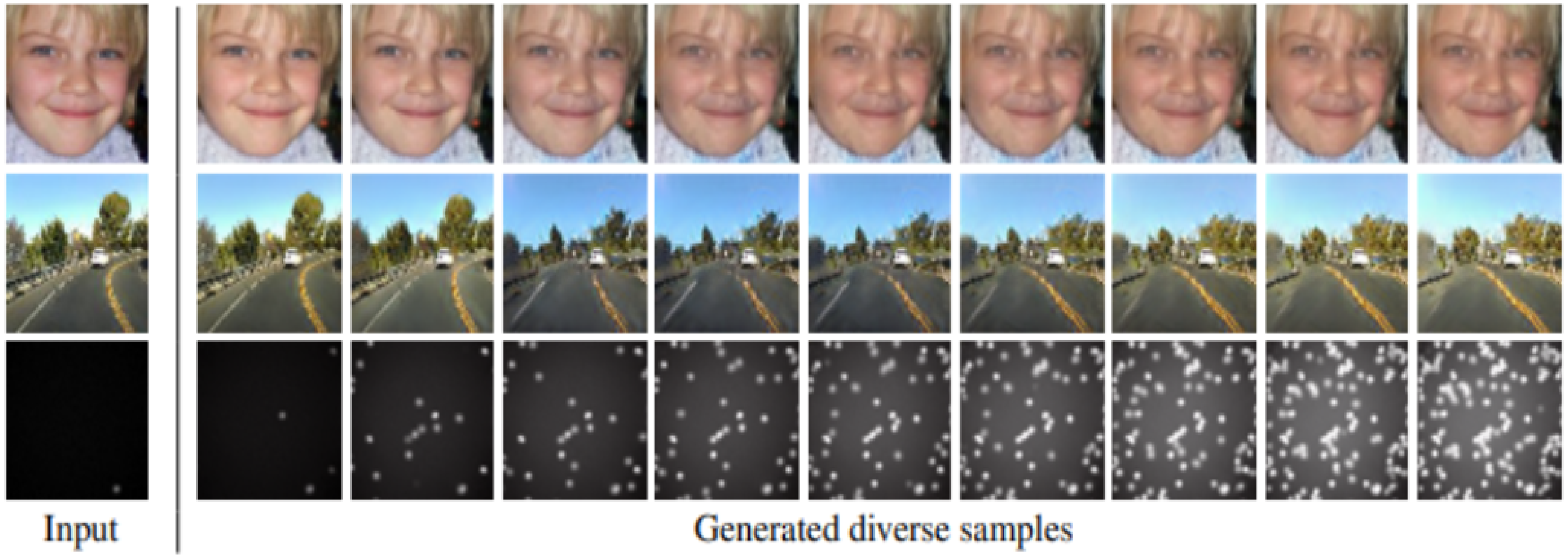


Figure 6: Continuous image-to-image translation conditional on both image and labels with unpaired samples. From top to bottom, the datasets are: UTKFace (young → old), SteeringAngle (steering angle at -80° → steering angle at 80°), and Cell200 (one cell → two-hundred cells). All models are trained with CycleGAN by plugging in our continuous atom generator.

learned perceptual
image patch similarity (LPIPS)

1)focused on human visual percepti
on and perceptual distance metrics

2) l2, Euclidean distance, PSNR meas
ure assume pixel-wise independenc
e so that not proper to evaluate str
uctred outputs like images.

3) This metrics can assess perceptua
l similarity better.

Experiments

- Continuous image-to-image translation conditional on regression labels

  - we target for modeling gradual changes in generated images with respect to continuous input conditions.
   input conditions to the corresponding output with parameters instantiated from an underlying continuous spa
  ce, which subsequently leads to generated images with gradual changes

Figure 7: Continuous image generation conditional on regression labels. From left to right, each column are images generated with the same atoms sampled from the continuous atom space and model the appearance of face at a particular age from 0 to 60. By fixing **z** in (7), each row shows the graduate appearance changes w.r.t. continuously sampled atoms.

Dataset	UTKFace		
Methods	cGAN	CcGAN	Ours
Intra-FID ↓	4.516	0.425	0.432
NIQE ↓	2.315	1.725	1.749
Diversity ↑	0.254	1.298	1.321
Label Score ↓	11.087	7.452	7.399

04 Conclusion

Conclusion

- we presented both theoretically and empirically that, continuous dynamics of convolutional filters can be effectively modeled in the filter subspace by neural ordinary differential equation, to subsequently achiev e conditional image generation with graduate appearance changes
- The introduced continuous filter atom generation enables continuous image generation conditional on im ages, labels, or both.
- We demonstrated its effectiveness using both superior quantitative and qualitative results.