JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Learning Geometry-aware Representations by *Sketching*

THU-PM-259

Hyundo Lee

Inwoo Hwang

Hyunsung Go

Won-Seok Choi

Kibeom Kim

Byoung-Tak Zhang

Seoul National University, AI Institute, Seoul National University

SEOUL NATIONAL UNIVERSITY

AIIS
Artificial Intelligence Institute
Seoul National University

Brief overview

Concept

Image

Geometric concepts

ex) position, shape, distance, orientation

z

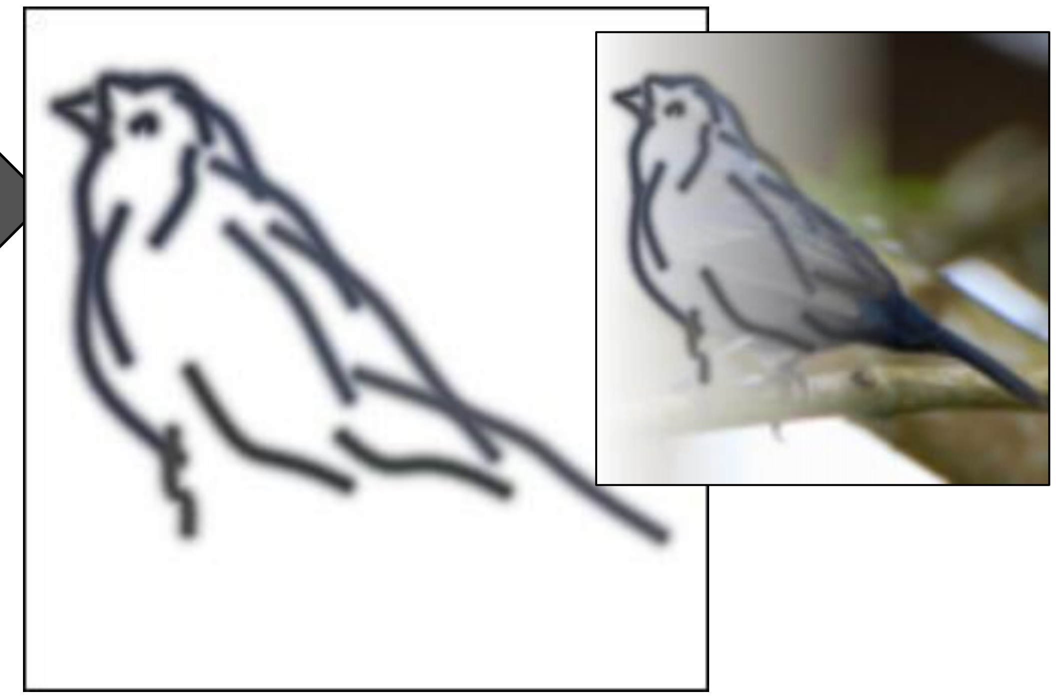
Feature (vector)

Brief overview

Concept

Stroke-based Sketching

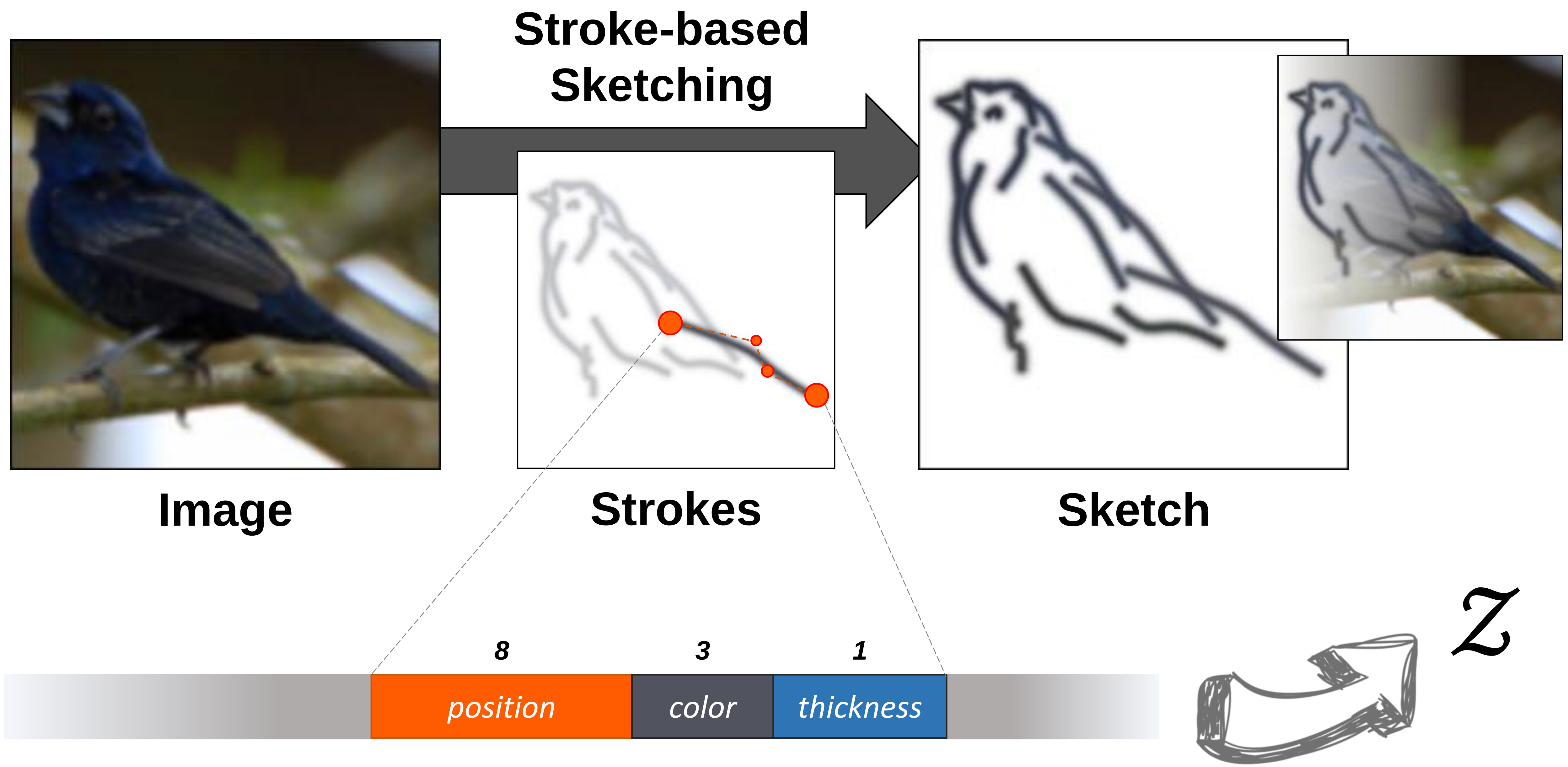
Image



Sketch

Brief overview

Concept



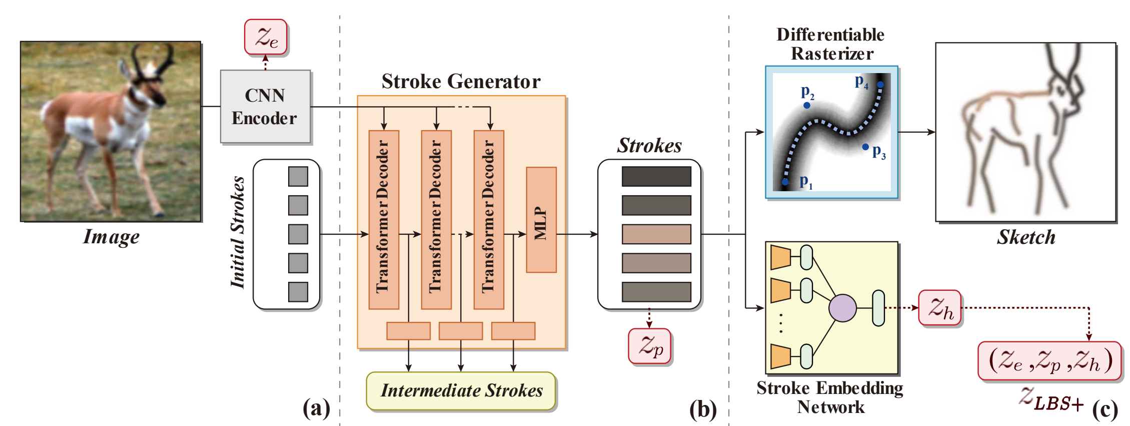
Brief overview

Contributions

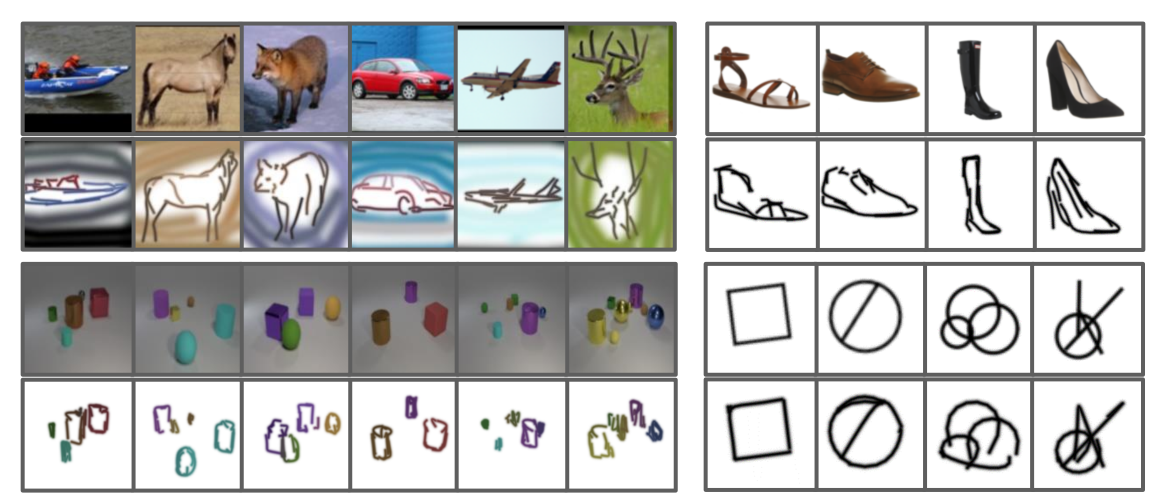
- Theoretically show that **strokes can convey geometric information**

- Design a sketch-based visual representation model,

  coined **LBS (Learning By Sketching)**



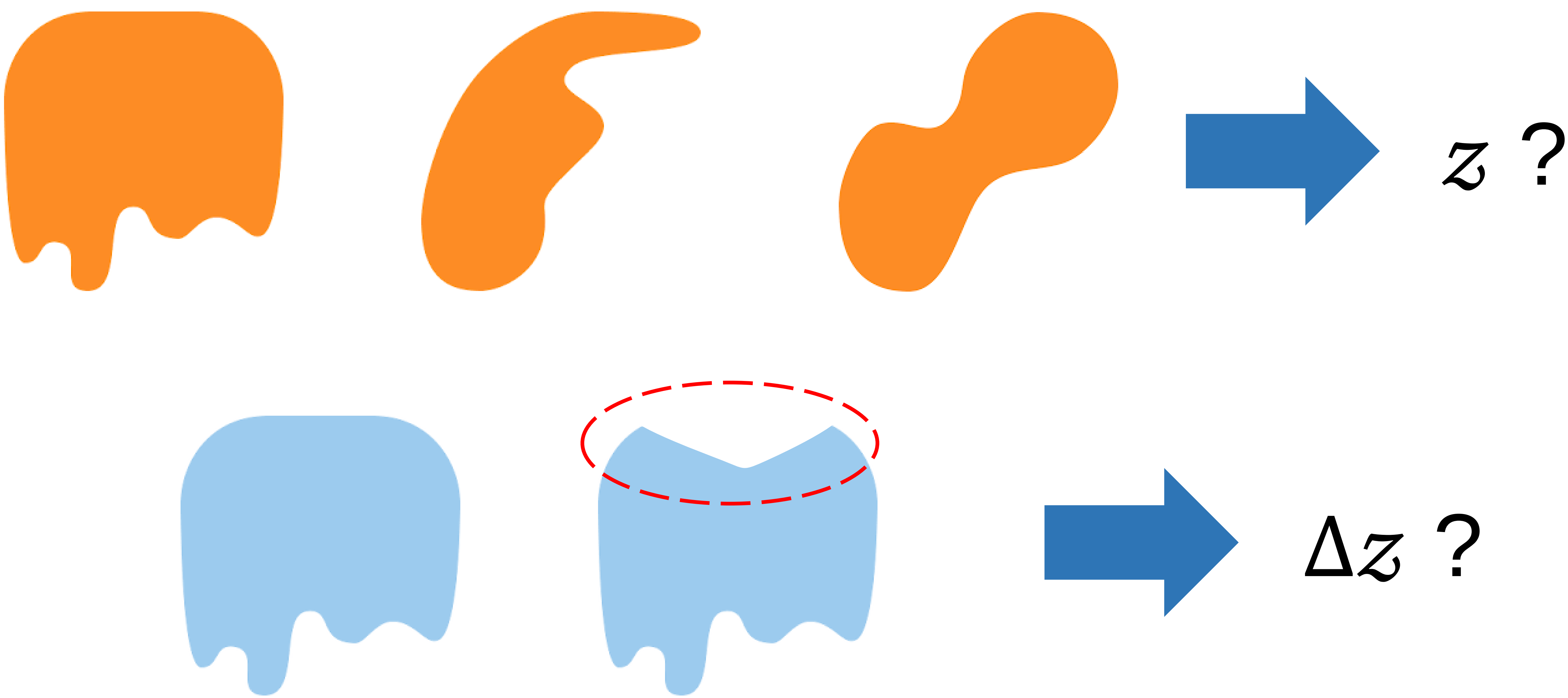
- Experiments for investigating the effectiveness of our approach in various domains and downstream tasks



Question

How do machines represent shape?

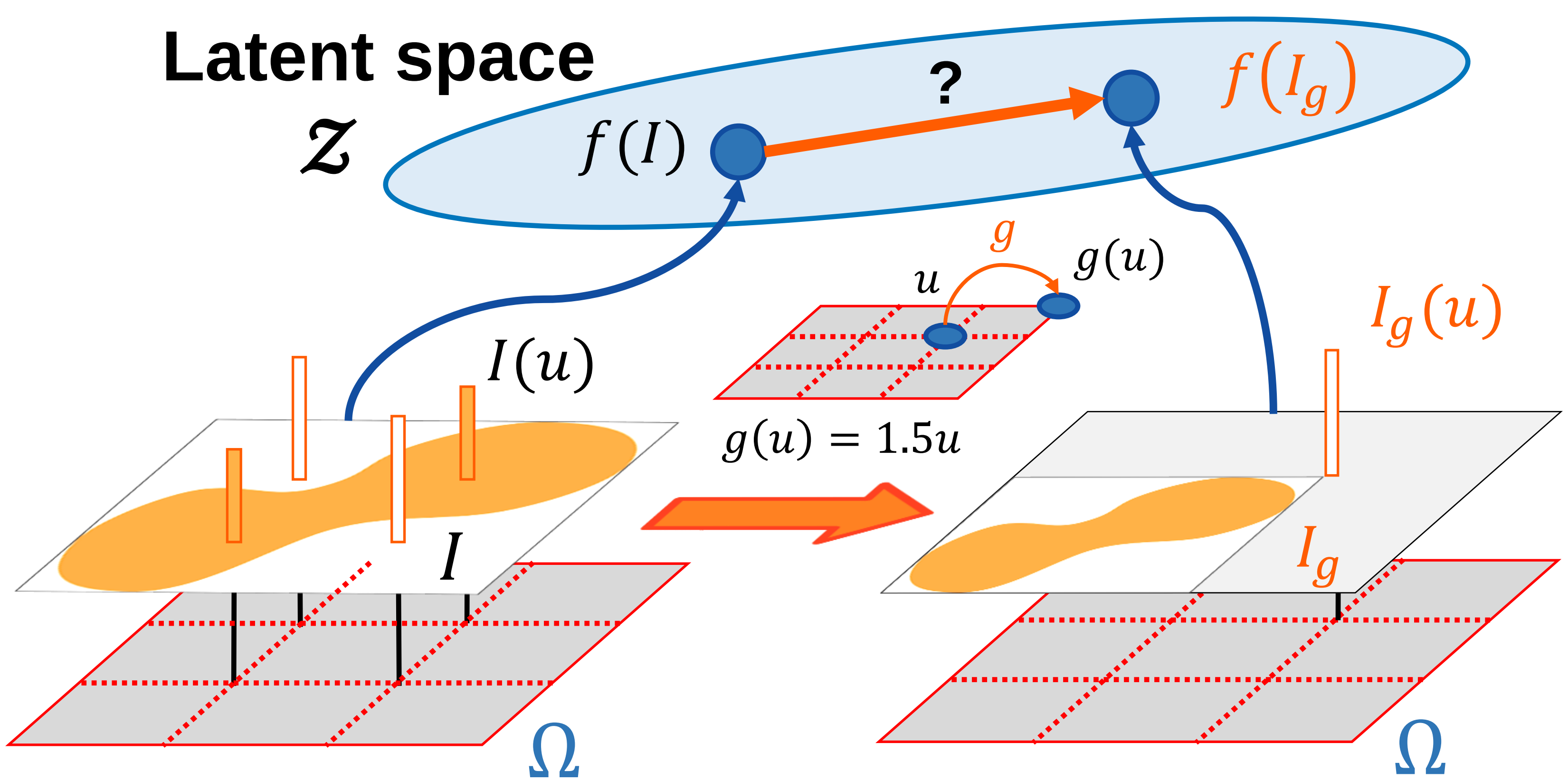
- Are they interpretable, well-aligned for downstream tasks?



Question

How do machines represent geometric concepts?

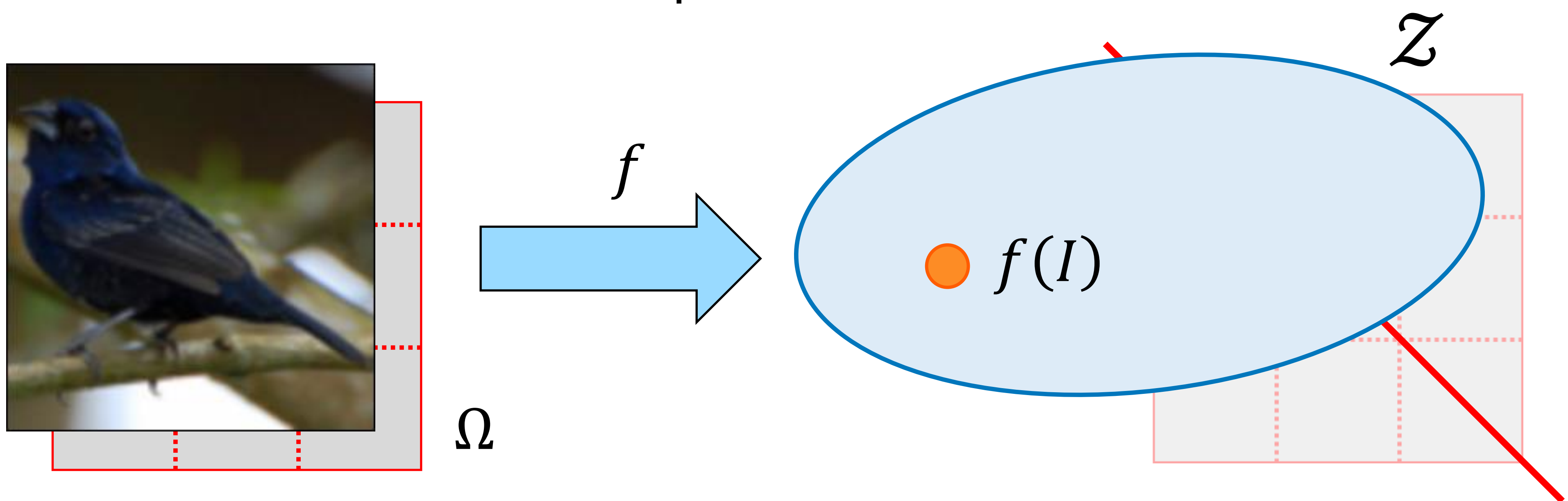
- Ex) position, shape, distance, orientation
- An image I belongs to a physical domain Ω: a 2D grid space
- For given geometric transformation $g: \Omega \rightarrow \Omega$,
  - The transformed image I_g, $I_g(u) = I\left(g^{-1}(u)\right)$
- With given representations $f(I)$ and $f\left(I_g\right)$, can we predict the transformation g?



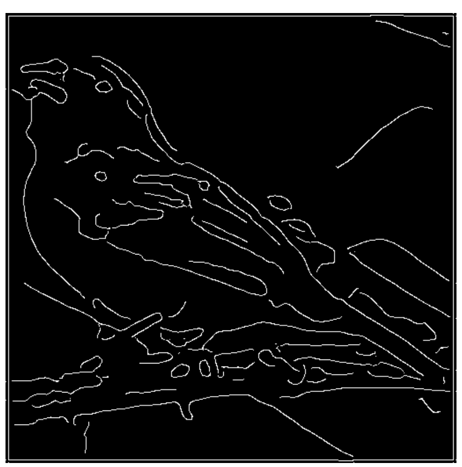
Background

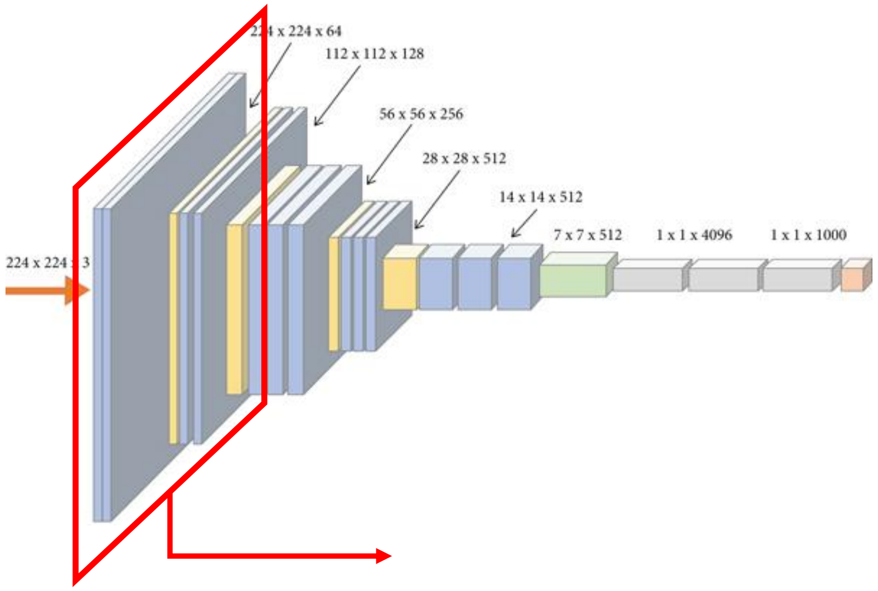
Previous studies

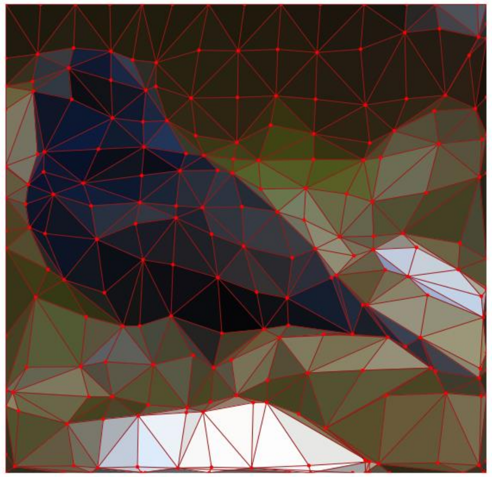
- Representations on semantic latent space



- Representations with same domain Ω





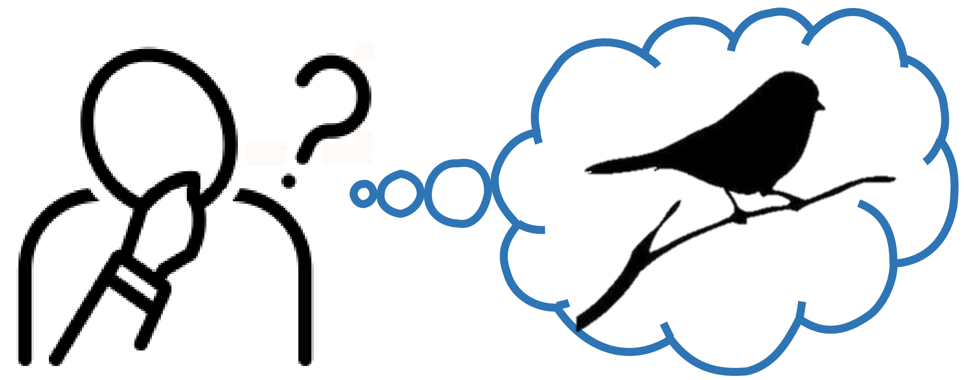


Motivation

How do human represent image?

- Language

Language
"A black bird sitting on a branch"



- Drawing/sketching

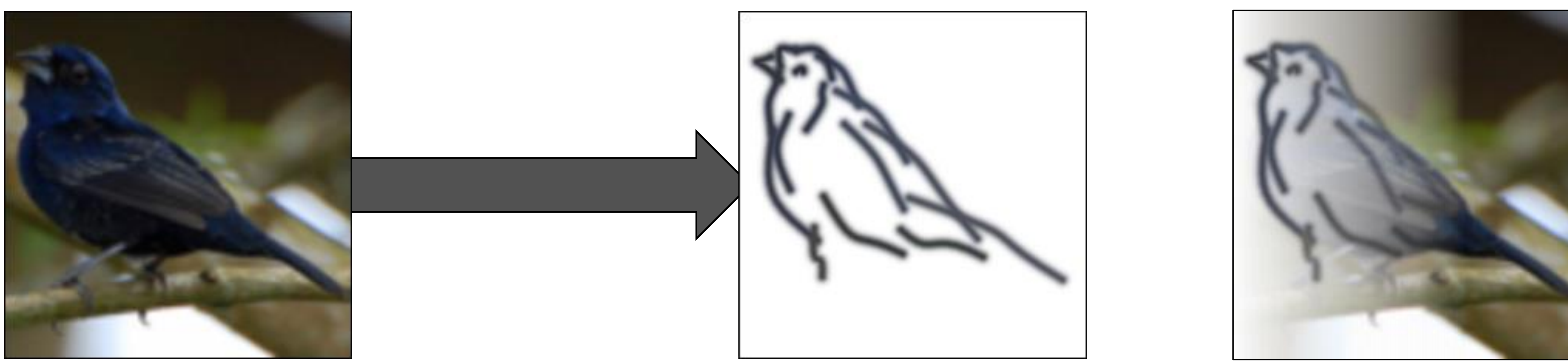
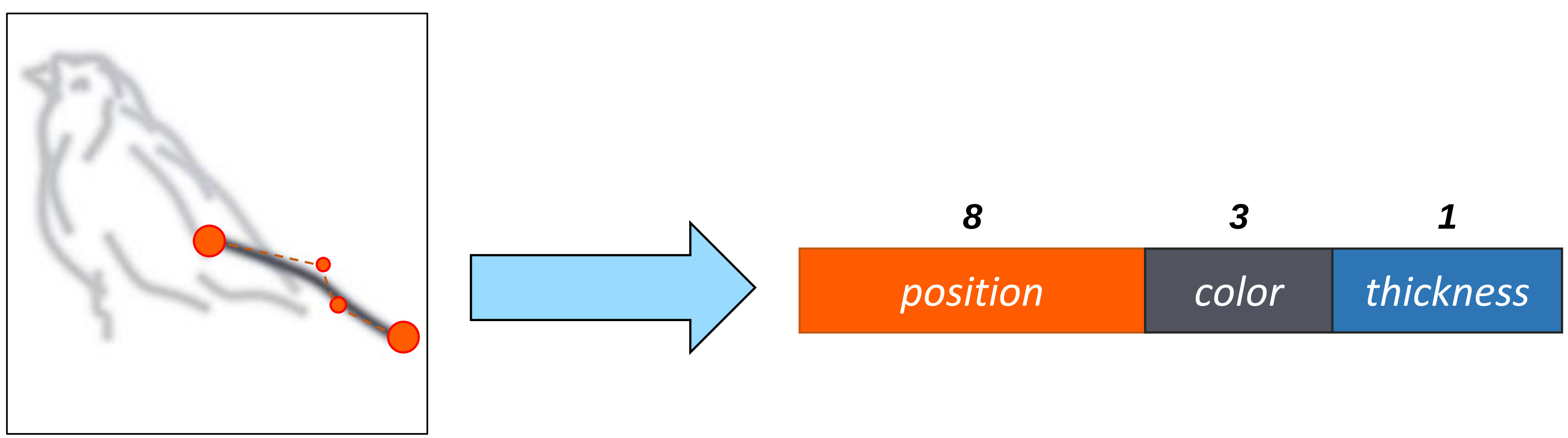


Image **Sketch**

Motivation

Core properties

- Sketching
  - salient features of an image ⇒ abstract image based on set of colored strokes
  - **Preserving geometry**: shares the same physical domain Ω
  - **Abstraction**: representation with limited # of strokes
  - **Compactness**: can be represented as a set of parametric curves



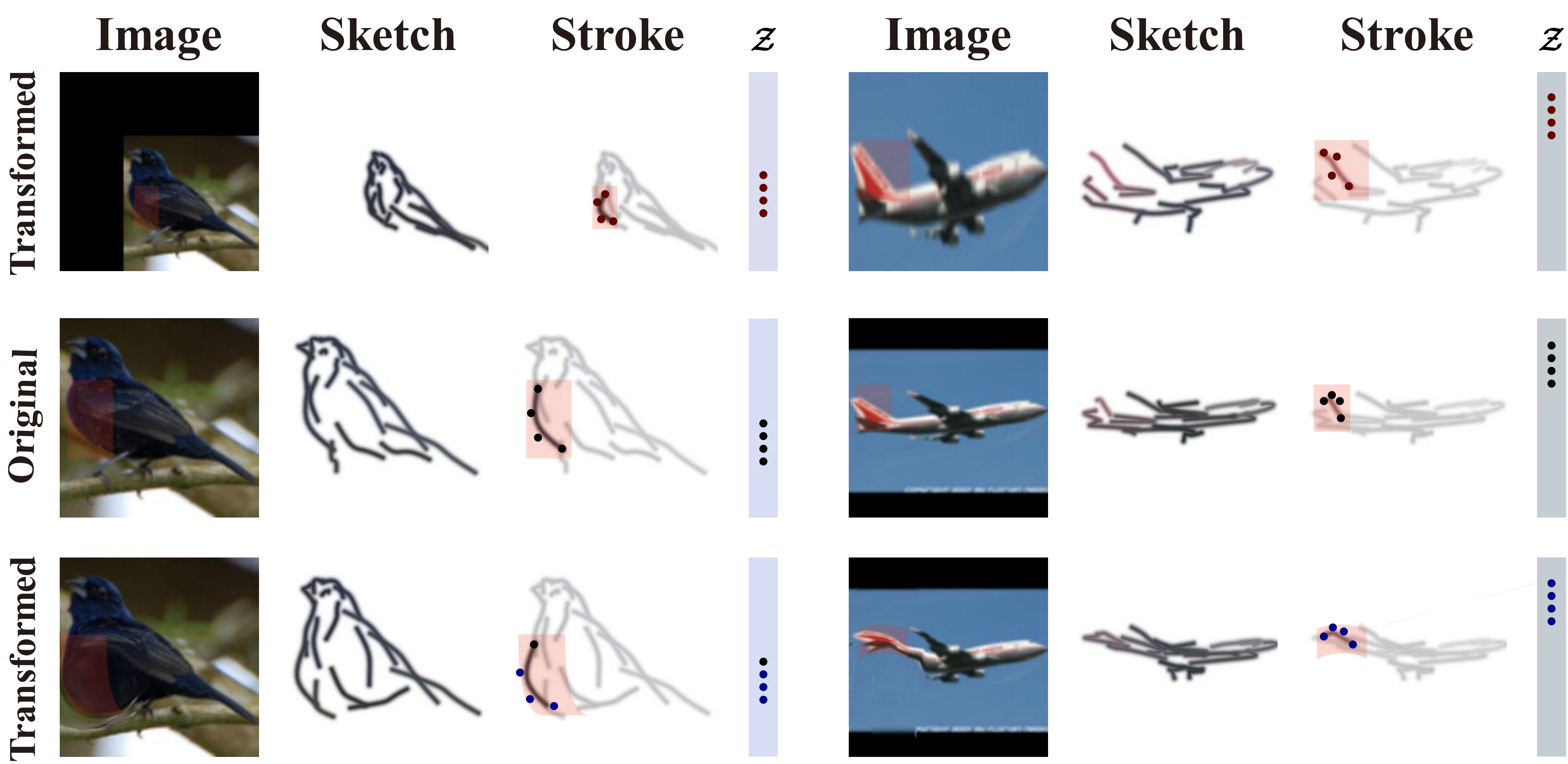
Stroke **Parameterization**

⇒ Based on these properties,
we use **strokes** as a **geometry-aware representation** for various downstream tasks

Mathematical framework

Does sketches & strokes represent geometric information?

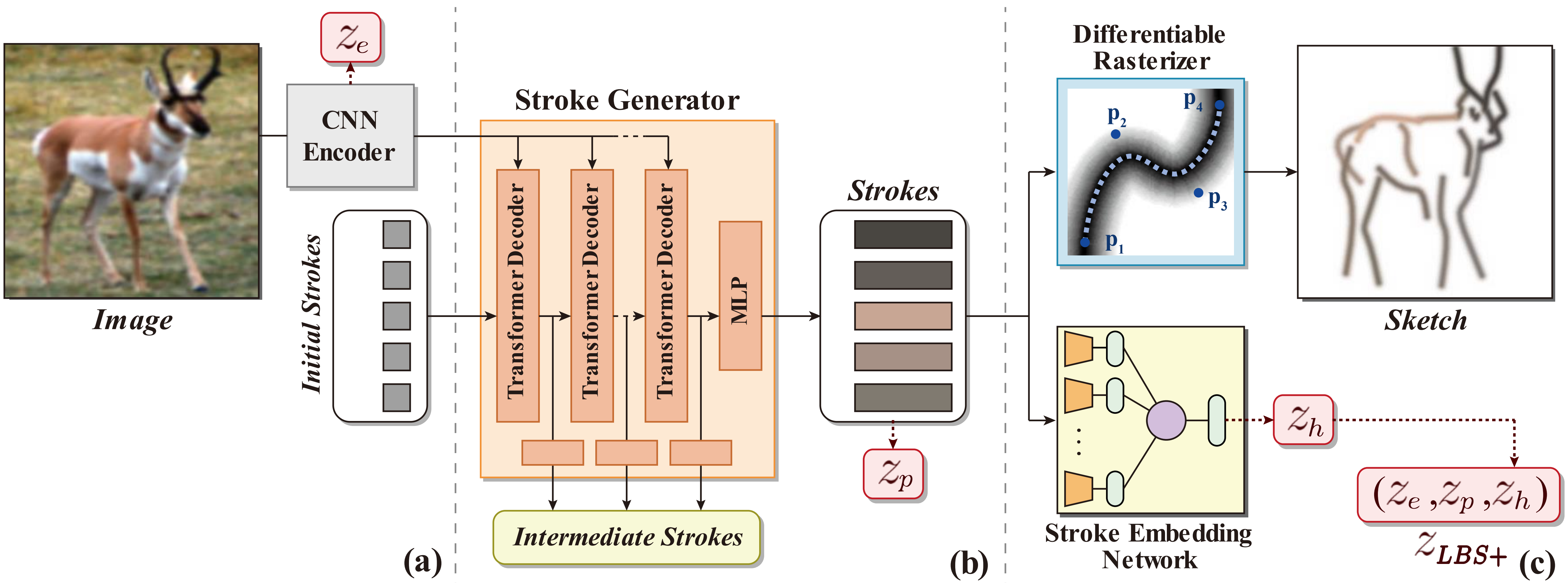
- With given representations $f(I)$ and $f(I_g)$, can we predict the transformation g?
- **$\mathcal{G}$-equivariance**: $\exists \rho' \; s.t. \;\; f(\rho(g) \cdot I) = \rho'(g) \cdot f(I)$
  - Sketching is equivariant to arbitrary geometric transformation $g \in \mathcal{G} \;\; (\rho' = \rho)$
  - Converting into stroke is equivariant to affine transformation $a \in \mathcal{A}$



Architecture

Learning By Sketching (LBS)

- Abstraction & reflecting geometric information in a short inference time
- Without using explicit sketch dataset.

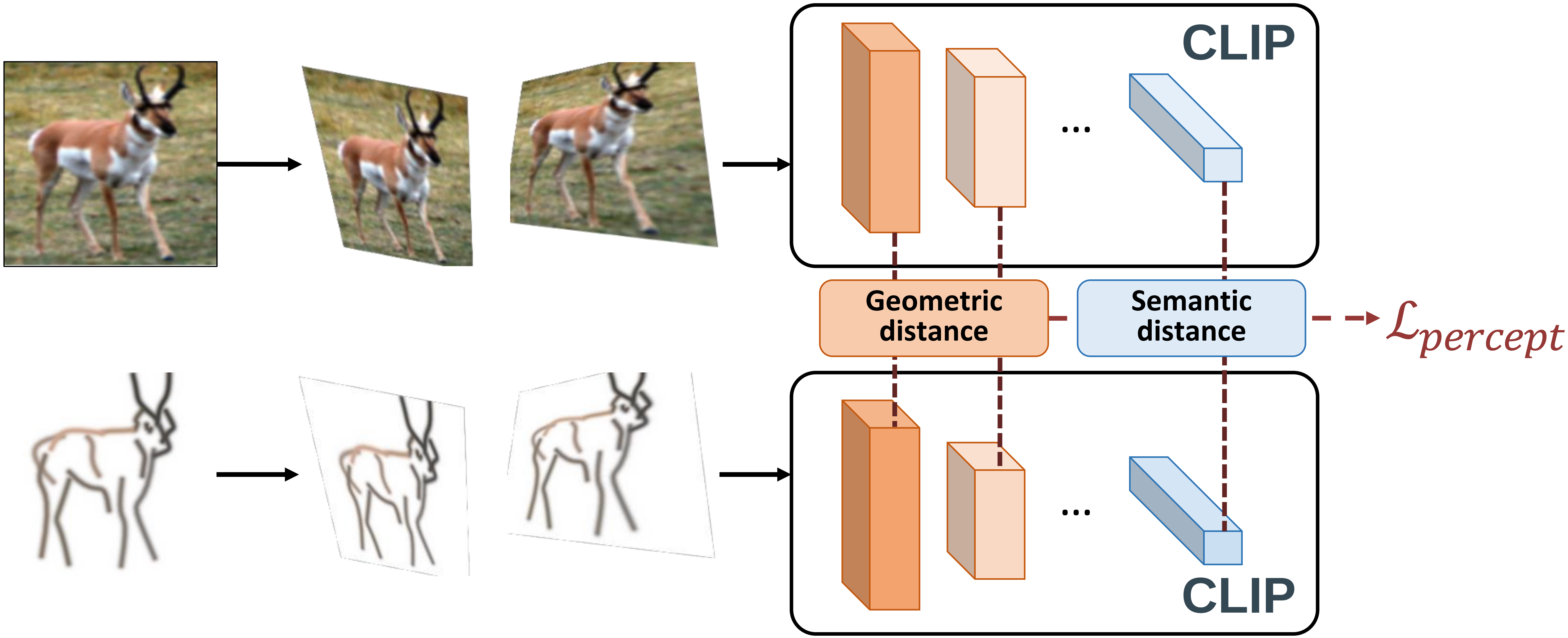


$$\mathcal{L}_{LBS} = \mathcal{L}_{percept} + \lambda_g \cdot \mathcal{L}_{guide} + \lambda_e \cdot \mathcal{L}_{embed}$$

Architecture

Learning By Sketching (LBS)

- $\mathcal{L}_{percept}$: CLIP-based perceptual loss [1]

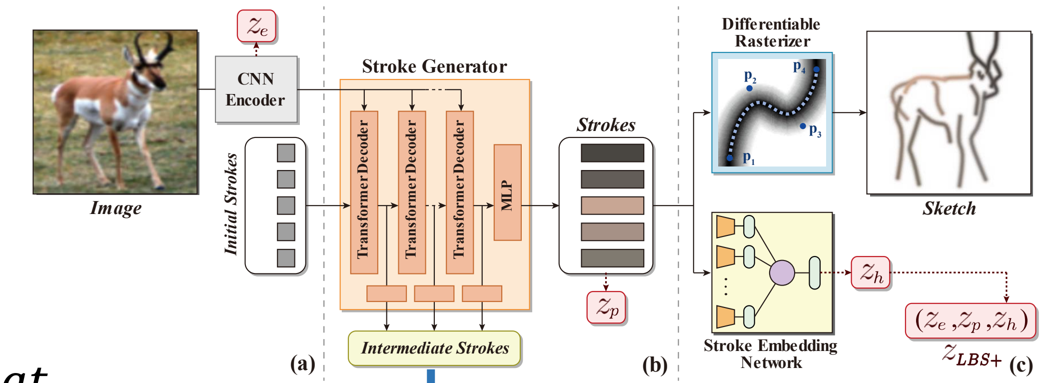


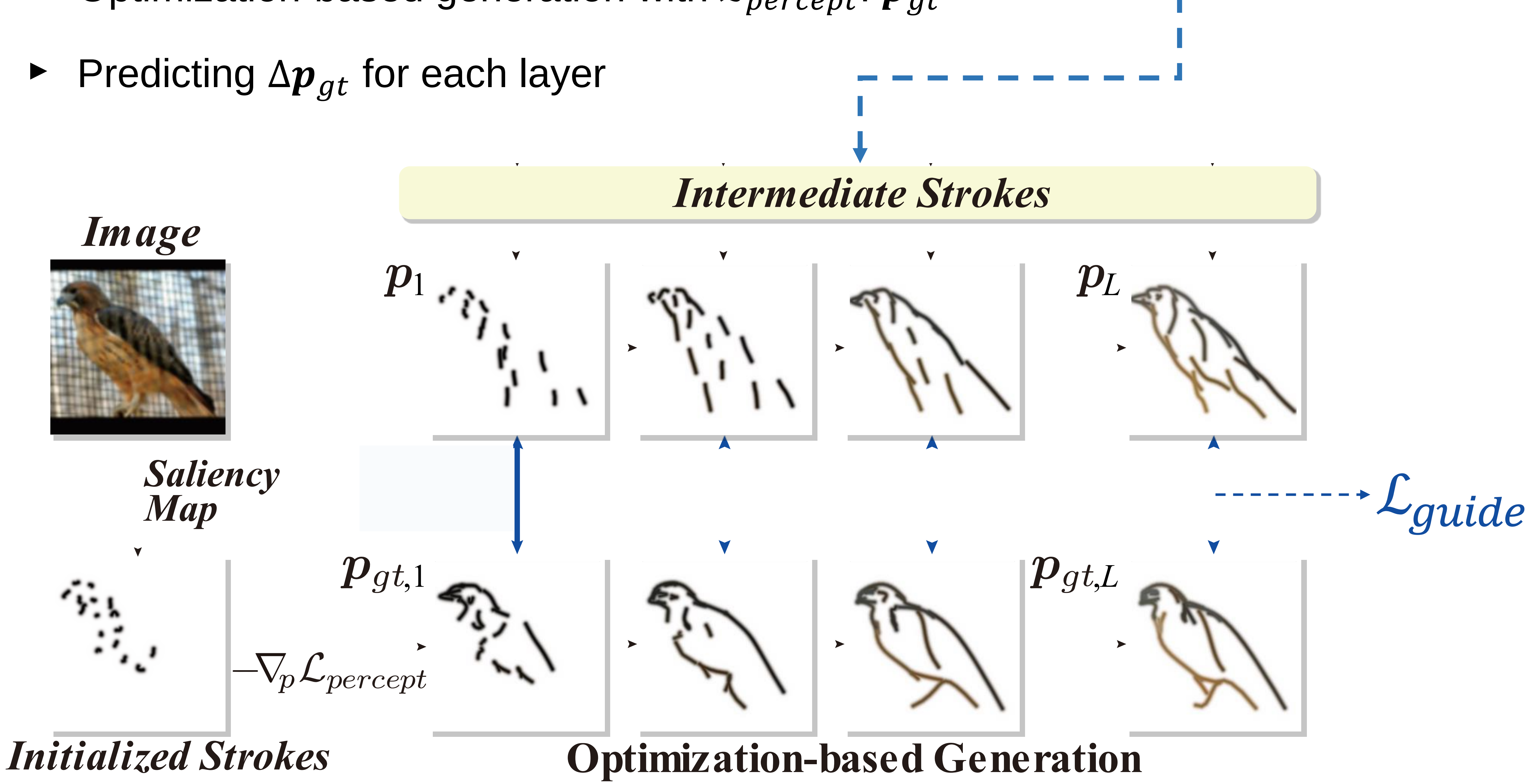
[1] Vinker, Yael, et al. "Clipasso: Semantically-aware object sketching." *ACM Transactions on Graphics (TOG)* 41.4 (2022): 1-11.

Architecture

Learning By Sketching (LBS)

- Guidance loss $\mathcal{L}_{guide}$
  - Optimization-based generation with $\mathcal{L}_{percept}$: $\boldsymbol{p}_{gt}$
  - Predicting $\Delta \boldsymbol{p}_{gt}$ for each layer

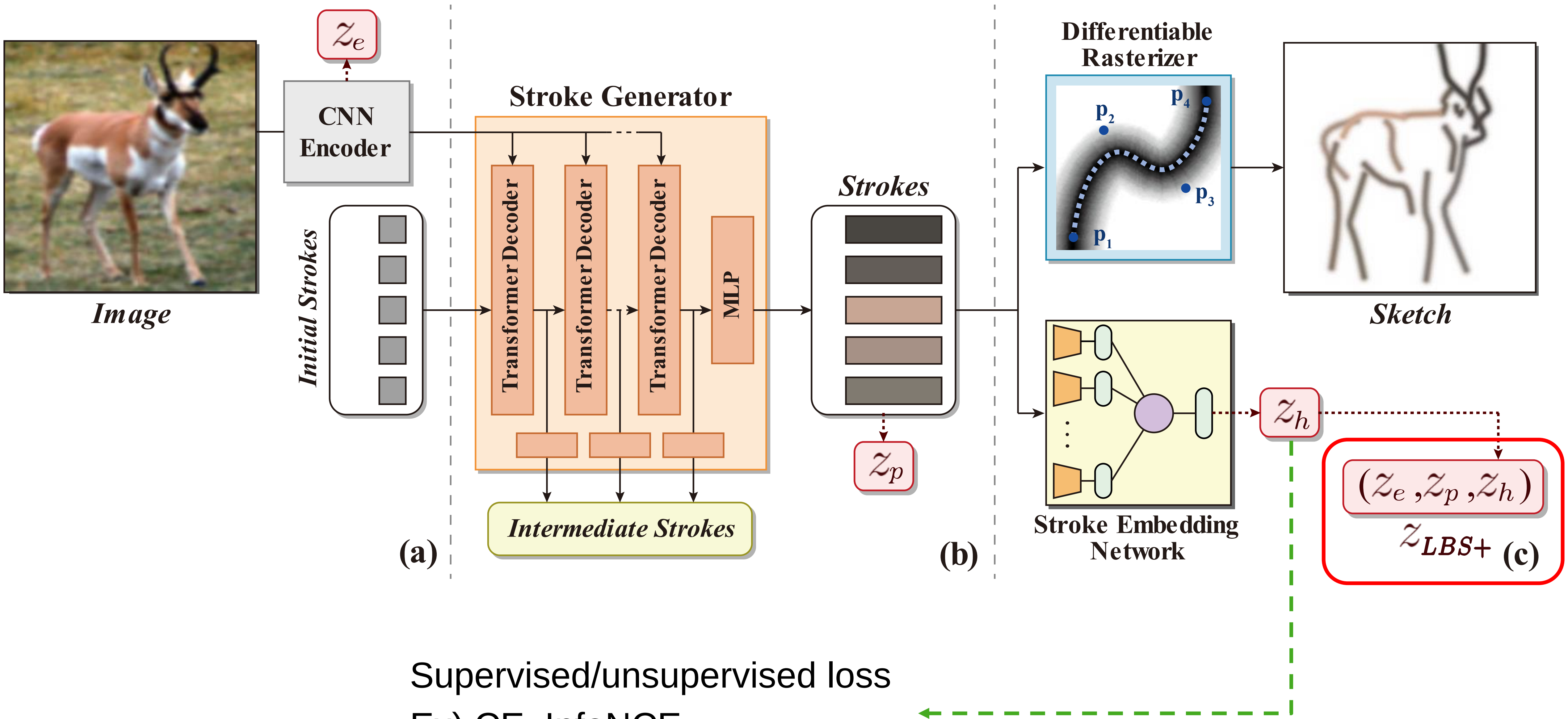




Architecture

Learning By Sketching (LBS)

- Stroke embedding loss $\mathcal{L}_{embed}$
  - z_h: combines the information of all strokes through a stroke embedding network



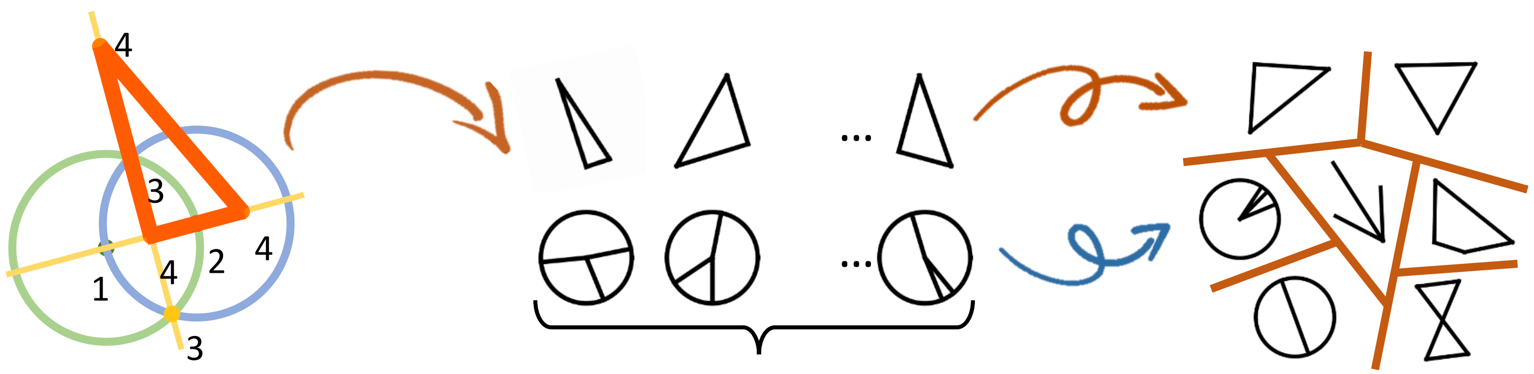
Supervised/unsupervised loss

Ex) CE, InfoNCE, …

Experiments

Quantitative results

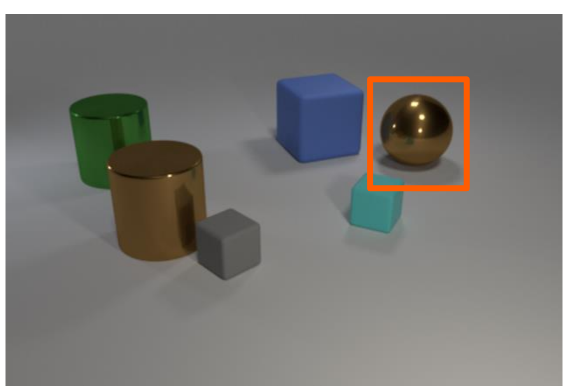
- Understanding geometric primitives & concepts



Euclidean geometry concepts | 10 realizations per class | concept classification

Label	Method	Geoclidean Constraints	Geoclidean Elements
✓	CE	53.89±1.58	70.57±4.29
	SupCon [38]	42.41±3.16	55.83±4.28
	LtD-diff [54]	57.26±2.19	69.47±2.11
	E(2)-CNN [73]	**71.03**±1.94	69.28±1.46
	LBS (CE)	50.01±1.58	**81.06**±3.14
✗	SimCLR [9]	32.04±0.64	65.14±4.11
	β-TCVAE [8]	17.18±1.35	33.82±1.64
	GeoSSL [55]	18.66±3.33	33.47±2.80
	HoG [14]	23.82	52.05
	LBS	**47.43**±1.34	**81.34**±0.16

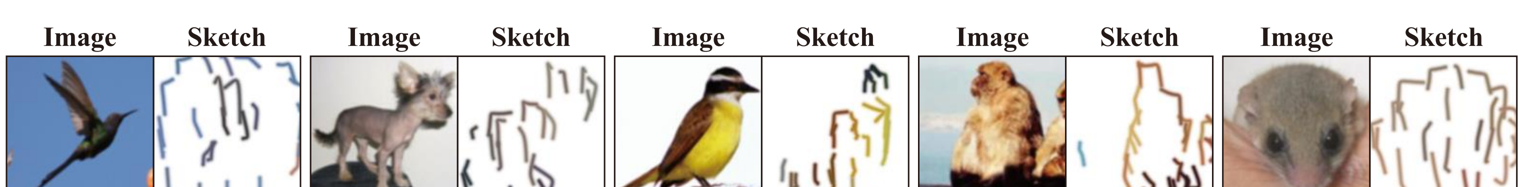
- Local geometric information & Simple spatial reasoning



Q) **Color** of the **leftmost** object?
Q) **Shape** of the rightmost object?
Q) Results after **shifting** the rightmost object?
…
Label: brown

Label	Method	RC	BC	Size	Shape	Third	Shift
✓	CE	98.71±0.10	76.02±0.90	92.51±0.40	49.97±0.29	40.66±0.20	**62.06**±1.62
	SupCon [38]	**98.75**±0.08	66.04±2.65	91.88±0.37	49.15±0.85	37.93±0.27	56.05±2.56
	LtD-diff [54]	62.29±0.48	15.84±0.43	63.98±3.38	43.96±3.05	16.47±0.59	17.21±0.29
	E(2)-CNN [73]	98.50±0.10	73.51±2.50	89.84±0.46	45.85±0.93	**41.95**±0.18	59.29±0.91
	LBS (CE)	97.49±0.22	**84.09**±0.84	**93.22**±0.29	**70.03**±0.68	38.23±0.25	51.56±0.16
✗	SimCLR [9]	60.61±1.24	63.77±2.29	83.35±0.60	41.95±0.33	33.42±0.55	43.05±0.55
	E(2)-CNN [73]	53.50±7.30	55.52±7.60	83.52±1.56	42.06±1.12	30.74±2.84	38.03±4.44
	β-TCVAE [8]	17.09±0.20	20.04±0.71	71.27±0.10	36.30±0.10	15.38±0.19	16.35±0.18
	GeoSSL [55]	20.16±0.63	21.61±0.67	73.79±0.78	44.08±1.10	15.39±0.16	16.94±0.34
	DefGrid [21]	73.81±0.91	73.38±0.80	81.50±0.22	46.34±0.77	24.90±0.27	36.28±0.13
	LBS	**84.31**±0.08	**83.00**±0.39	**92.66**±0.41	**70.01**±0.53	**37.41**±0.29	**49.32**±0.17
	CLIP [59]	37.39	54.98	77.51	66.91	34.75	34.80
	HoG [14]	56.83	58.69	81.73	61.14	24.28	33.25

- Domain transfer

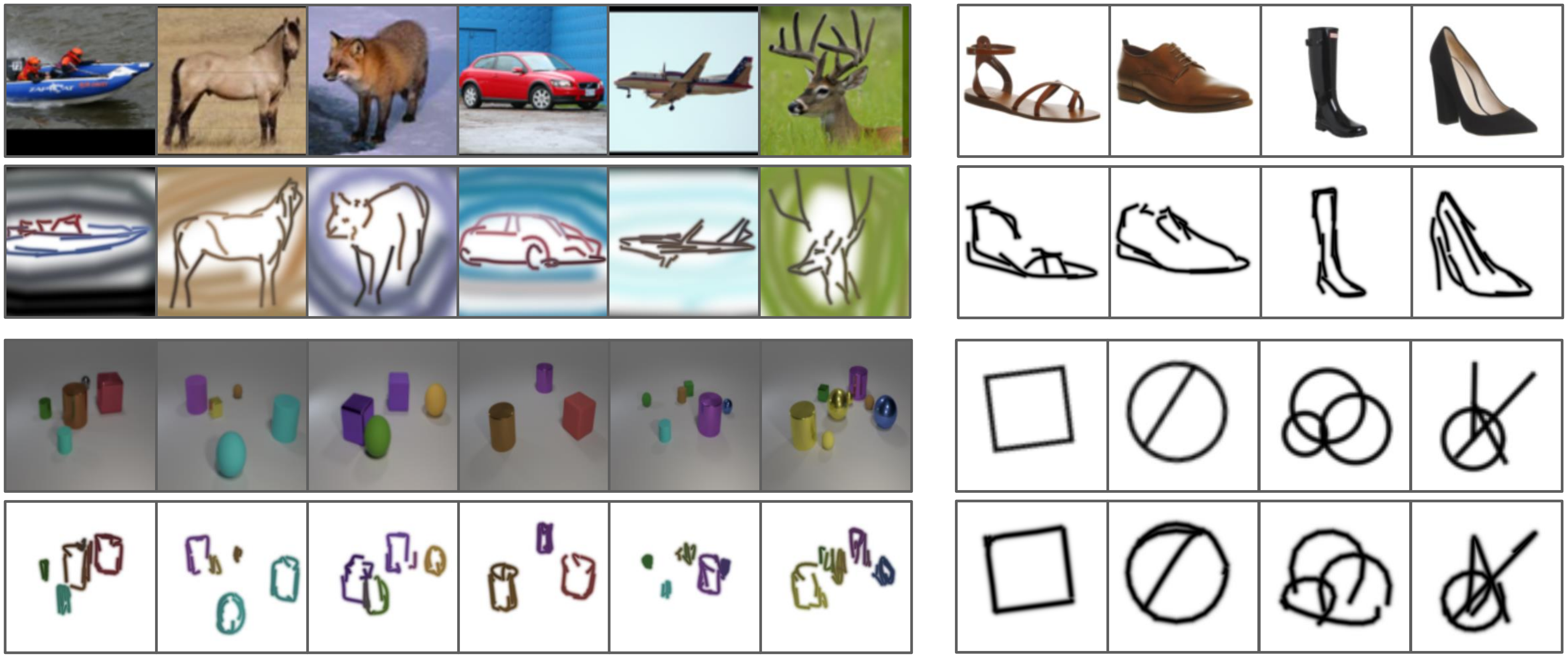


(b) CLEVR → STL-10

Dataset	Labeled Method	Labeled Accuracy	Unlabeled Method	Unlabeled Accuracy
CLEVR ↓ STL-10	CE	46.15±0.12	SimCLR [9]	41.68±0.05
	SupCon [38]	43.41±0.36	β-TCVAE [8]	27.35±0.38
	LtD-diff [54]	50.81±0.67	GeoSSL [55]	35.93±0.96
	E(2)-CNN [73]	45.19±0.84	E(2)-CNN [73]	38.50±0.49
	LBS (CE)	**56.48**±0.89	DefGrid [21]	33.13±0.17
			LBS	**55.35**±0.18

Experiments

Qualitative results



- Progressive optimization process

Thank you