# Text-Driven Generative Domain Adaptation with Spectral Consistency Regularization

Zhenhuan Liu[1,2] Liang Li[1] Jiayu Xiao[1,2] Zheng-Jun Zha[3] Qingming Huang[1,2,4]

[1]Institute of Computing Technology [2]University of Chinese Academy of Sciences

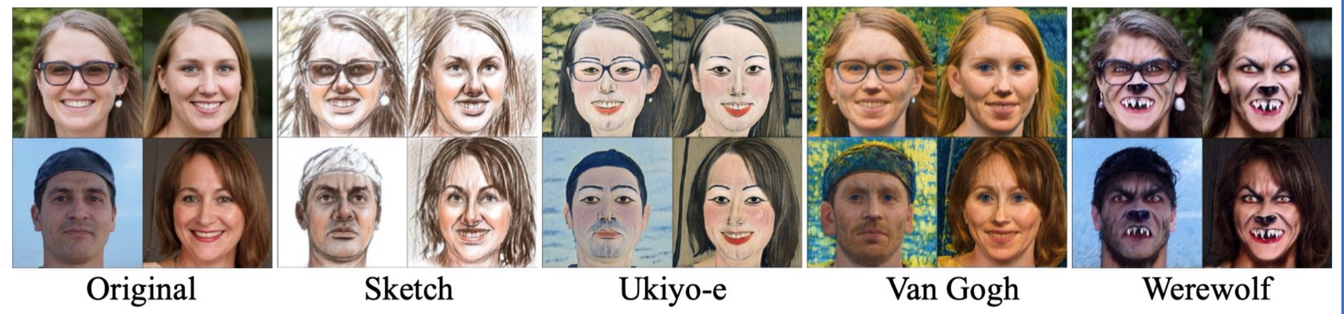
[3]University of Science and Technology of China [4]Peng Cheng Laboratory

ICCV23 PARIS

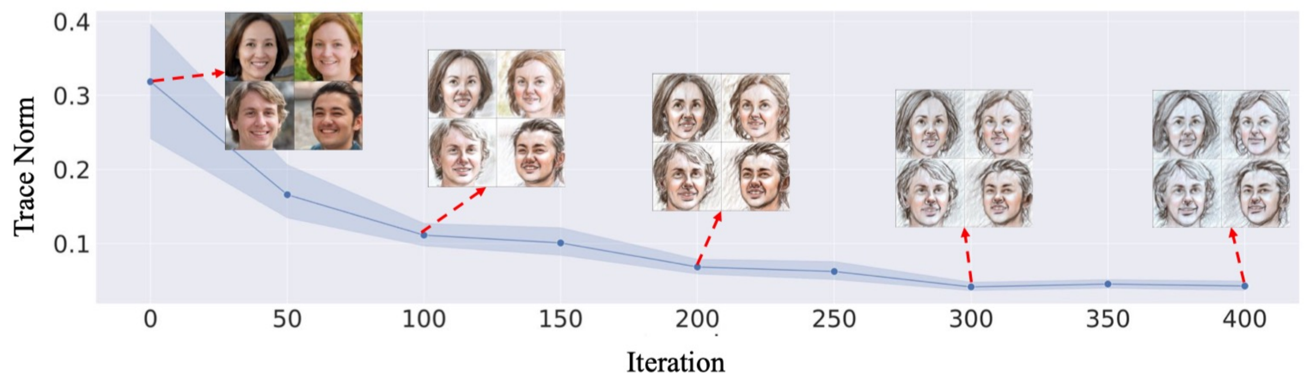
## Text-Driven Generative Domain Adaptation

### Text as Guidance of Target Image

Benefiting from Vision-Language models learning from large image-text pairs, text can be leveraged as a description of abstract visual semantics to guide generative domain adaptation instead of a collection of image samples in target domain.




### Mode Collapse in Adaptation Process




- Original identity attributes are collapsed after early training stage.
- Styles and diversity are entangled in the optimization process.

### Challenges

- How to measure the diversity of distribution quantitatively?
- How to regularize diversity without restricting style adaptation?

### Measuring Diversity from Geometric Point of View

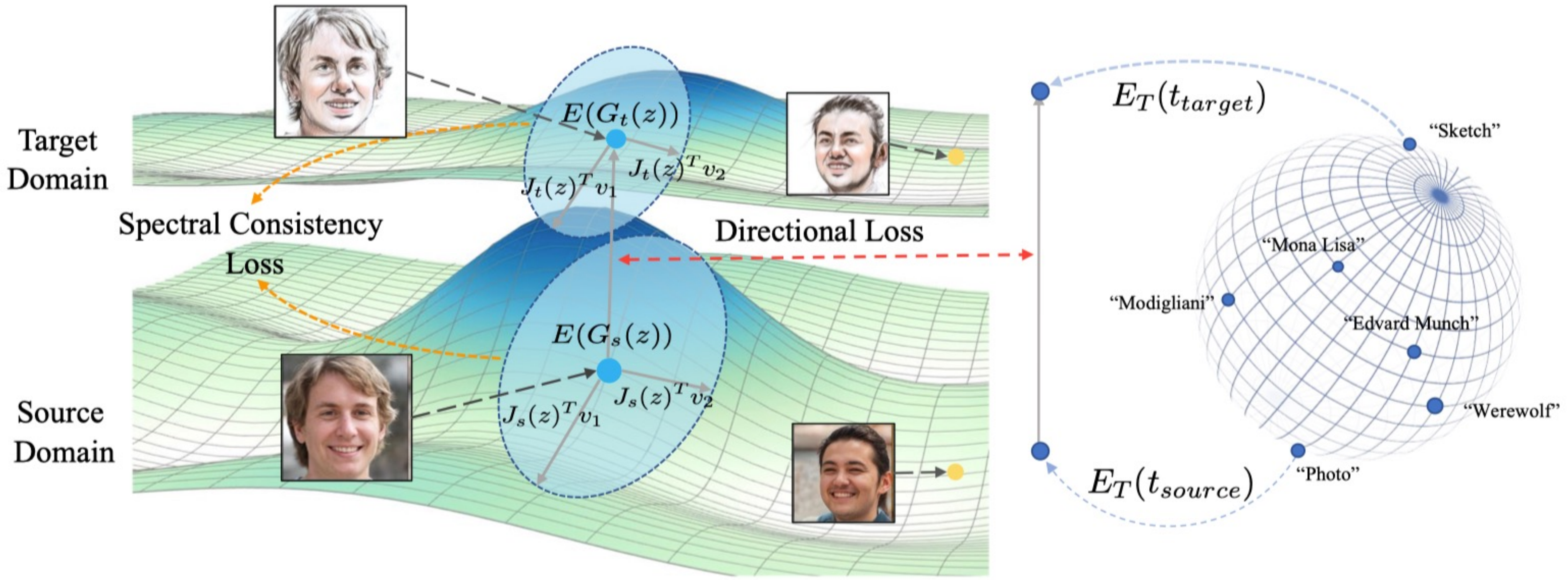
$$\lim_{\Delta z \to 0} d^2(G(z), G(z+\Delta z)) =$$

$$d^2(G(z), G(z)) + \frac{\partial d^2(G(z), G(z+\Delta z))}{\partial \Delta z} \cdot \Delta z +$$

$$\Delta z^T \cdot \frac{\partial^2 d^2(G(z), G(z+\Delta z))}{\partial \Delta z^2} \cdot \Delta z$$

1. Measuring distribution diversity via second order gradient.
2. Preventing absolute features to be restricted.

## Spectral Consistency Regularization

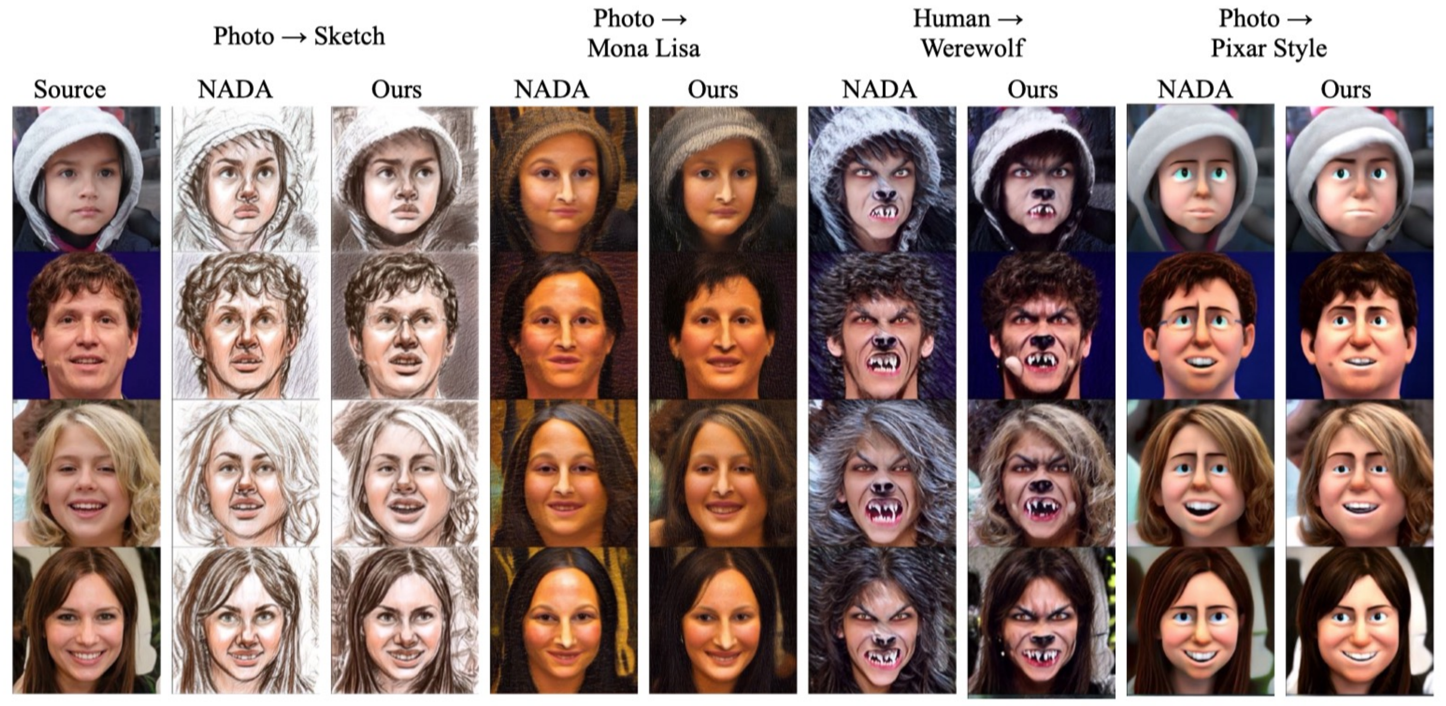



$$\mathcal{L} = \mathcal{L}_{dir} + \lambda_{spectral} \frac{\|\nabla_{G_L} \mathcal{L}_{dir}\|}{\|\nabla_{G_L} \mathcal{L}_{reg}\|} \mathcal{L}_{reg}$$

***Directional Loss***: Bring close the distance between generated images and target text description.

***Spectral Consistency Regularization***: Preserve intra-domain variations of source domain without restricting style effects of target generators.

## Diversity Preserving Domain Adaptation





## Experiments

### Quantitative Results Compared with Previous Methods

| Results | Photo → Sketch | | Photo → Mona Lisa | | Human → Werewolf | | Photo → Pixar | |
|---|---|---|---|---|---|---|---|---|
| | PPL | Trace | PPL | Trace | PPL | Trace | PPL | Trace |
| SCC | 547.34 | 0.526 | 485.70 | 0.521 | 428.45 | 0.343 | 440.17 | 0.535 |
| WDC | 378.25 | 0.090 | 363.17 | 0.191 | 311.34 | 0.116 | 345.31 | 0.167 |
| MS | 466.99 | 0.167 | 331.06 | 0.233 | 352.61 | 0.191 | 377.88 | 0.358 |
| PPL | 241.50 | 0.028 | 209.09 | 0.061 | 297.74 | 0.062 | 300.19 | 0.063 |
| CDC | 348.01 | 0.062 | 259.92 | 0.079 | 351.73 | 0.089 | 299.49 | 0.111 |
| NADA | 323.25 | 0.061 | 281.59 | 0.098 | 302.91 | 0.101 | 343.54 | 0.112 |
| Ours | 463.57 | 0.116 | 321.43 | 0.140 | 383.19 | 0.137 | 353.80 | 0.181 |

| Model Comparison | Quality | Style | Attributes |
|---|---|---|---|
| Ours vs. NADA [5] | 86.4% | 59.4% | 91.2% |
| Ours vs. SCC [35] | 64.2% | 86.2% | 65.0% |
| Ours vs. MS [14] | 59.4% | 83.6% | 54.2% |
| Ours vs. WDC [36] | 62.2% | 89.8% | 49.8% |
| Ours vs. PPL [11] | 92.4% | 71.4% | 95.0% |
| Ours vs. CDC [17] | 79.6% | 60.6% | 87.8% |

### Granularity-adaptive Regularization


## Downstream Applications