Simple Lines, Big Ideas: Towards Interpretable Assessment of Human Creativity from Drawing

Zihao Lin[1], Zhenshan Shi[1], Sasa Zhao[1], Hanwei Zhu[3], Lingyu Zhu[2], Baoliang Chen[1](✉), Lei Mo[1](✉)
[1] South China Normal University, [2] City University of Hong Kong, [3] Nanyang Technological University

Shanghai·2025

Introduction

Problems and Motivation

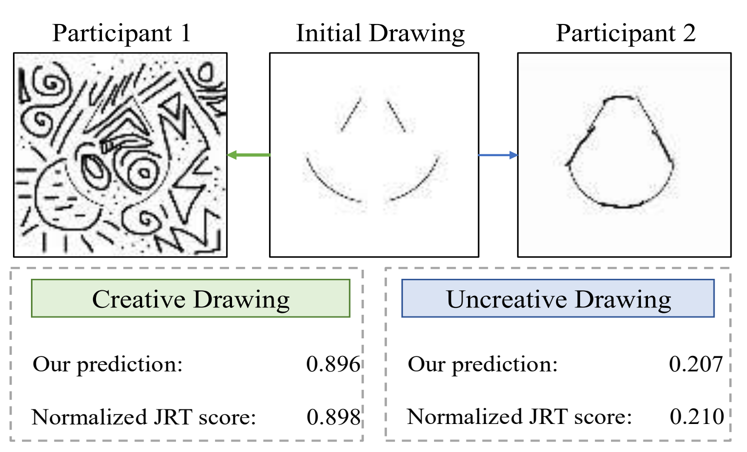

Fig1. The creativity of a drawing can be measured by a creativity score

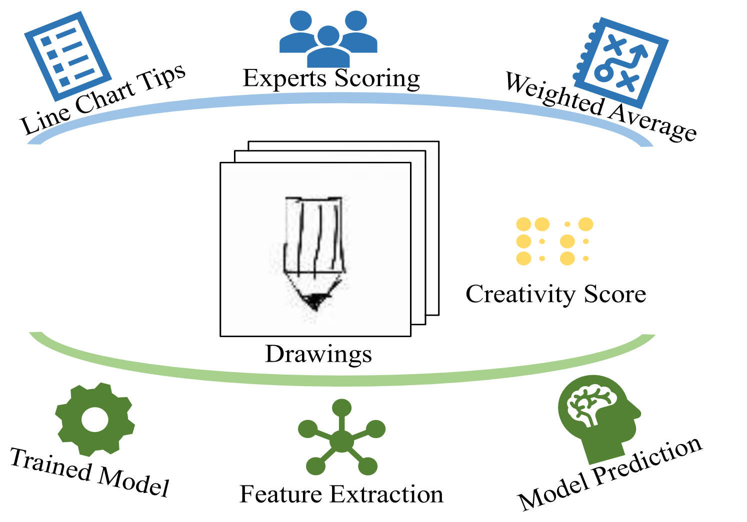

Fig2. Comparison of evaluation processes: the traditional(top) vs. our automated assessment model (bottom)

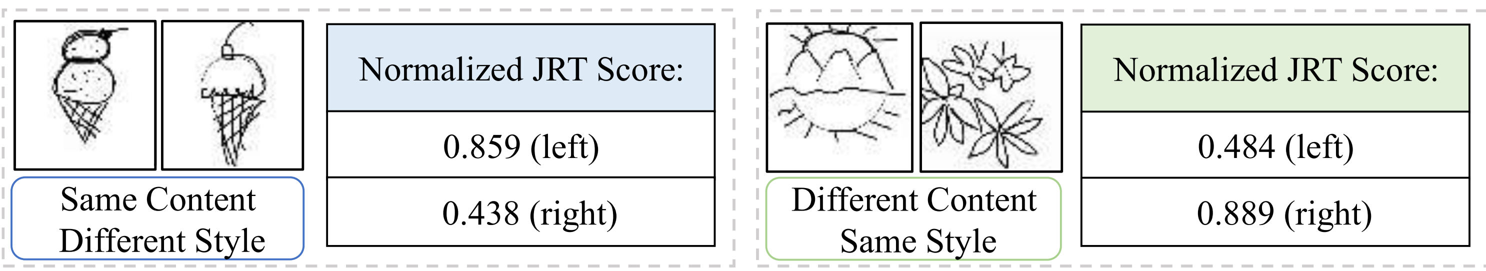

Fig3. Style and Content in Painting Creativity Assessment: JRT Score Comparison of Same Content-Different Style vs. Same Style-Different Content

Existing methods suffer from two major drawbacks:

- Remaining heavily reliant on labor-intensive and inherently subjective expert-based scoring
- Traditional automated models lack interpretability and have poor generalization ability.

Motivated by the cognitive understanding that creativity can emerge from both **what is drawn (content)** and **how it is drawn (style)**, we reinterpret the creativity score as a function of these two complementary dimensions.

Contributions

(1)Dataset Augmentation: Systematically enrich drawing datasets with **semantic category labels** to guide content-aware creativity learning.

(2)Interpretable Model: Propose a novel model for visual creativity assessment, decomposing drawings into **content & style components**.

(3)Learning Architecture: Develop a conditional multi-task architecture that adaptively focuses on **creativity-relevant signals** via content/style cues.

Approach

Overall Architecture

Built on CLIP (ViT-L/14 image encoder + text encoder), with three key modules for interpretable, adaptive learning:

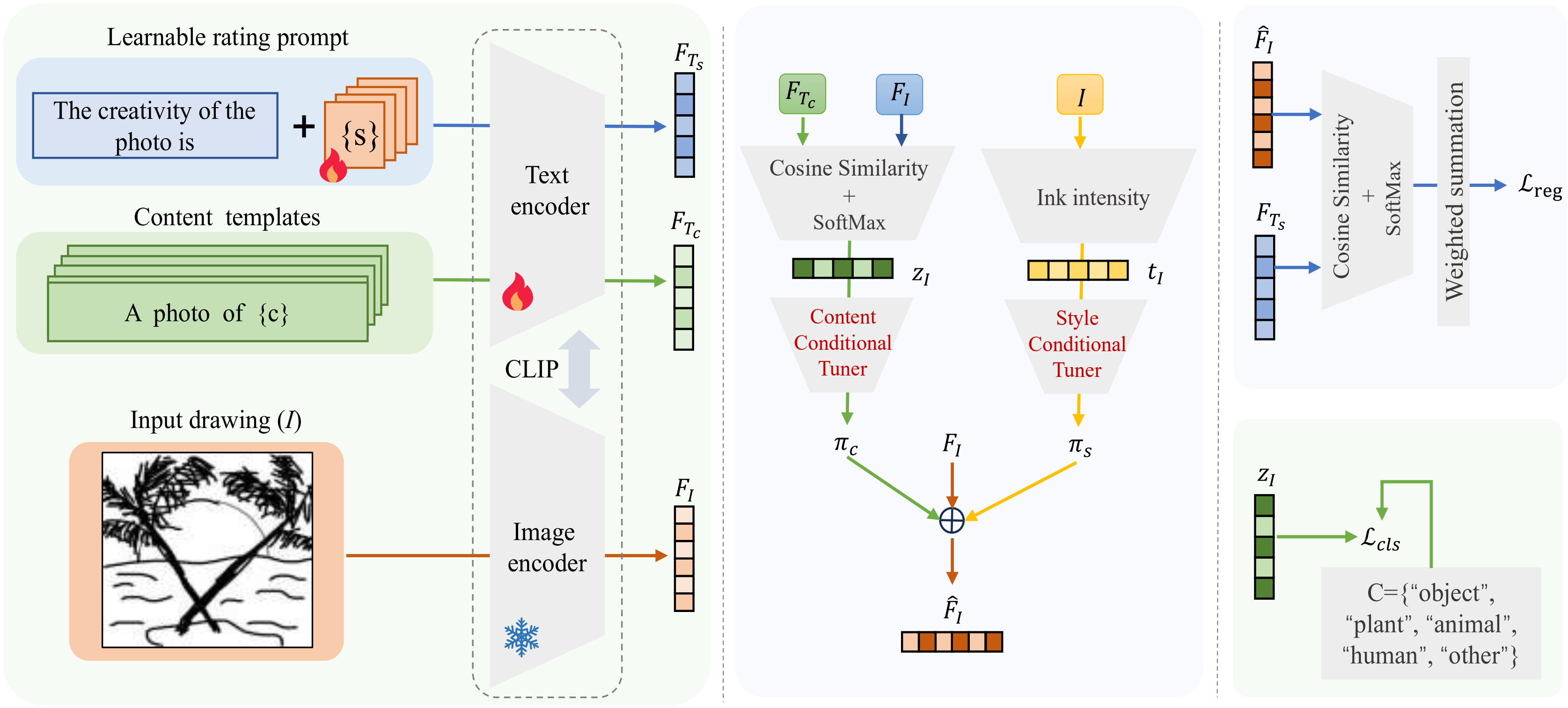

Fig4. Framework of our proposed Content-Style conditioned Creativity Assessment (CSCA) model

Key Modules

LCR Embeddings: 5 trainable tokens (δ_s) aligning creativity semantics.
Content Conditional Tuner (CCT): Modulates features via content category cues.
Style Conditional Tuner (SCT): Supplements with ink intensity (CLIP's blind spot).

Key Computation

Modulated Visual Embedding:

$$\hat{F}_I = F_I + \pi_c + \pi_s$$

Creativity Score Prediction:

$$\hat{p}(s|I) = \frac{\exp(sim(\hat{F}_I, F_{T_s})/\tau)}{\sum_{s=1}^{|S|}\exp(sim(\hat{F}_I, F_{T_s})/\tau)}$$

$$\hat{q}(I) = \sum_{s\in S}\hat{p}(s|I)\cdot w_s$$

Multi-Task Loss:

$$\mathcal{L} = MSE(score) + \lambda \cdot CrossEntropy(content\ cls)$$

Experiments and Results

Dataset

We enrich AuDrA-Drawings (13k+ drawing-rating pairs) with two critical annotations, namely **Content Label** and **Style Proxy** .

Subset	Drawing Type	#Samples	Raters	ICC
Primary Dataset	Abstract	11,075	50 trained undergraduates	>0.89
Rater Gen. 1	Abstract	670	3 new raters	0.73
Rater Gen. 2	Abstract	722	6 new raters	0.90
Rater & Task Gen.	Special Objects	679	3 new raters	0.63

Table1. Summary of AuDrA-Drawings datasets

Experiment Settings

Training: Batch Size = 16, Learning Rate=1e-5, Epochs=136, λ=1e-3
Metrics: SRCC (rank correlation), PLCC (linear correlation)
CLIP: Vit-L/14 with image encoder(frozen) and text encoder

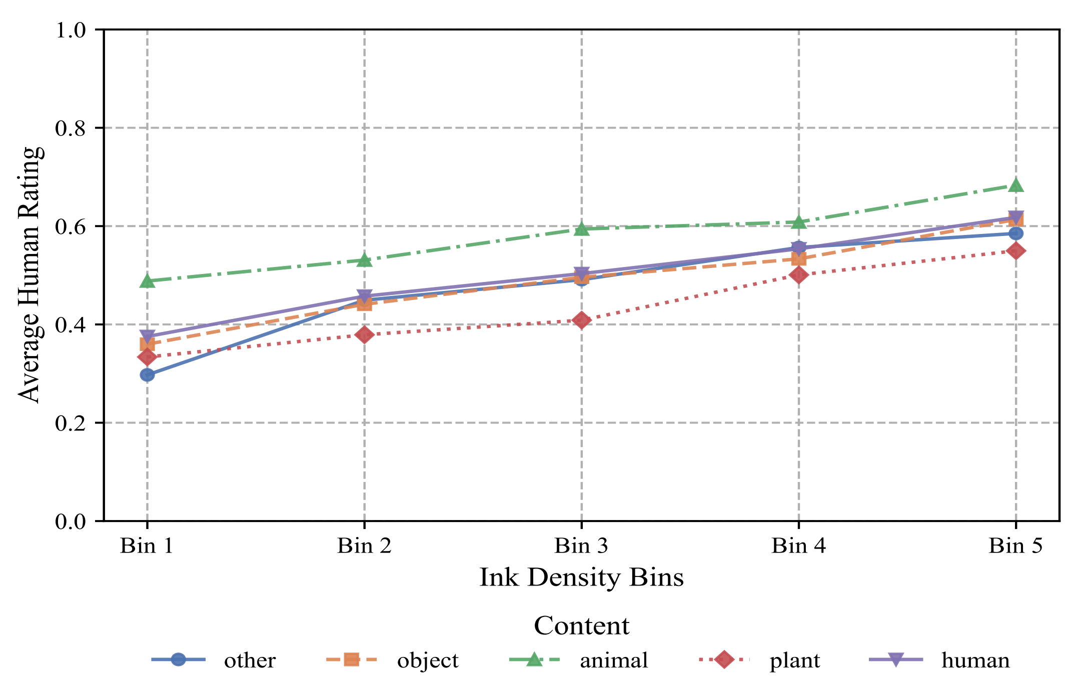

Fig5. Mean human creativity ratings across content categories, grouped by ink intensity levels

Results and Conclusion

Group	Model	Year	SRCC	PLCC
VGG	VGG16	2014	0.52	0.04
ResNet	ResNet34	2016	0.79	0.81
	ResNet50	2016	0.78	0.79
CLIPIQA+	CLIP(ViT-B/32)	2023	0.63	0.65
	CLIP(ViT-B/16)	2023	0.67	0.68
	CLIP(ViT-L/14)	2023	0.67	0.69
	CLIP(RN50)	2023	0.80	0.81
	CLIP(RN101)	2023	0.81	0.81
Others	Audra	2024	0.80	0.79
	AGIQA	2025	0.79	0.80
Ours	CSCA	2025	**0.86**	**0.87**

Table2. SOTA performance on Primary Dataset

Model	LCR	SCT	CCT	Primary	RG1	RG2	FG
CLIP Baseline	√	×	×	0.69	0.69	0.48	0.36
w/LCR	√	×	×	0.86	0.77	0.71	0.47
w/LCR + SCT	√	√	×	0.86	0.78	0.72	**0.51**
w/LCR + CCT	√	×	√	0.86	0.76	0.72	0.47
Our approach	√	√	√	**0.87**	**0.79**	**0.73**	0.49

Table3. Ablation studies results(PLCC)

SOTA Performance: Our method outperforms SOTA on Primary Dataset (Table 2), with consistent correlation metric improvements verifying the content-style conditioned creativity assessment framework's superiority.

Ablation studies: The data of the Table 3 verifies the necessity of the synergy among the LCR, CCT, and SCT modules for enhancing the model's accuracy and generalization, especially strengthening its robustness in cross-task and cross-rater scenarios.

Main conclusion :

In this study, we propose CSCA—an interpretable drawing creativity assessment framework integrating content and style. Experimental analysis shows:

- At the same ink density(style), "animal"-category drawings receive the highest creativity scores (owing to broader creative expression space).
- Across all content categories, scores increase with rising ink density (higher stylistic elaboration enhances creative perception).

These findings validate CSCA's design, which achieves SOTA on AuDrA-Drawings and replaces subjective expert scoring.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62401214, in part by grants from Project of Key Institute of Humanities and Social Science, MOE (16JJD880025).