NBTI: NN-BASED TYPOGRAPHY INCORPORATING SEMANTICS

Team P12: Dongjae Lee, Dongwoo Moon, Doojin Baek, Hanbee Jang, Min Kim

1. PROBLEM FORMULATION

Limited transformation
- Original losses (ACAP and Tone loss) were used to prevent excessive transformation to maintaining readability. However, those losses focus on only preserving the similarity of the **"shape"** of a letter itself.
- To address this, we introduced "**Embedding Loss**" to ensure that the transformed word is recognized as **the same letter**, rather than naively comparing shapes.

Not works for formless words
- Generating typography from formless words using the original model wasn't work.
- To overcome this, we added a model to convert **formless words** into **concrete ones**.

2. RELATED WORKS

Embedding Similarity
Recently, an encoder was utilized to find good candidates to replace the original letter (Tendulkar et al., 2019). We adopt this concept, but we use an encoder to prevent excessive transformation, not for finding good candidates.

LLM Fine-tuning
The success of large language models, such as GPT-3 (Brown et al., 2020), has resulted in the development of various downstream tasks through fine-tuning the pre-trained model. In this study, we harness the capabilities of GPT-3.5 to effectively generate concrete words from their original ambiguous forms, which the original model struggles to handle.

3. METHODS-EMBEDDING LOSS

Alphabet Characters Fonts Dataset (26 classes, 14900 fonts, 32x32 pixels, Total number of datasets 387400) was used to train Word Image Encoder.

Table 1. Hyperparameters for training

Batch size	Epoch	Learning Rate	Validation Ratio
512	100	0.001	30%

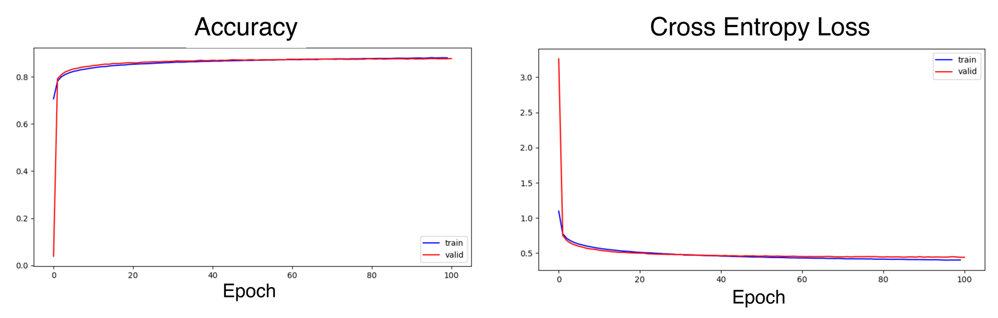
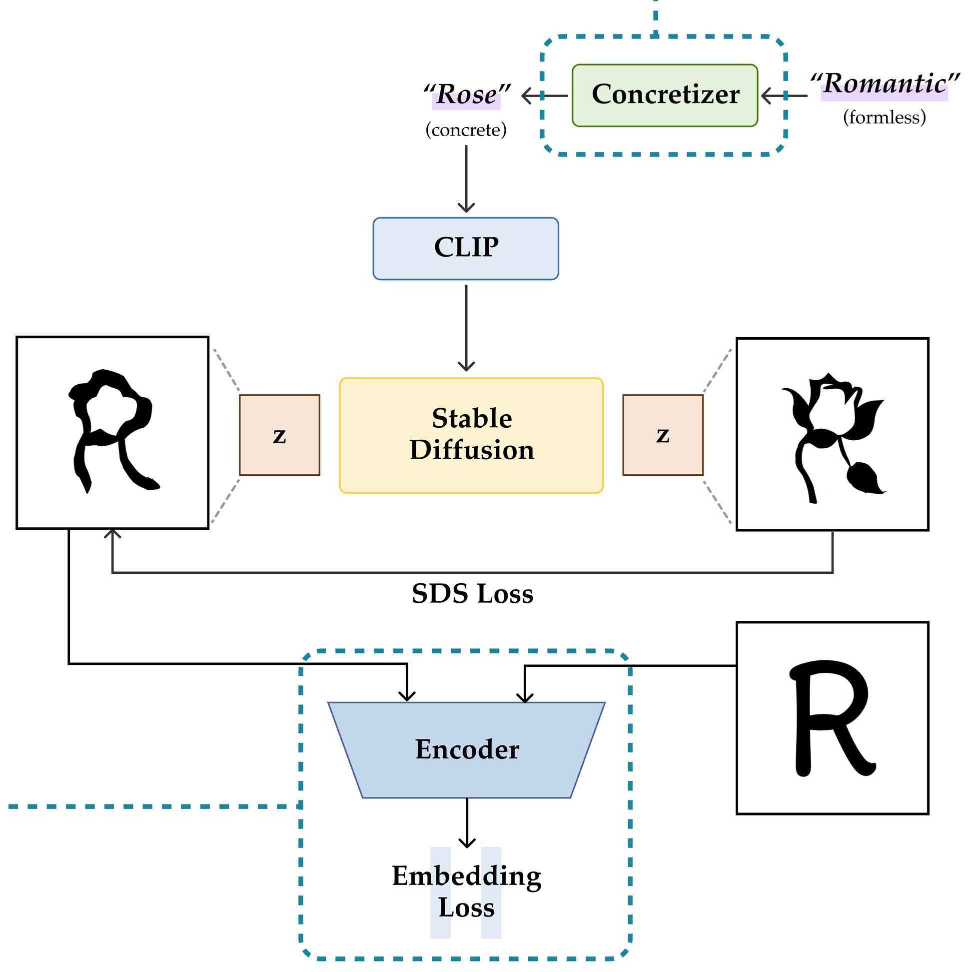


Table 2. Accuracy and loss of the encoder model

Validation Accuracy	Validation Loss
0.877	0.441

3. MODEL STRUCTURE



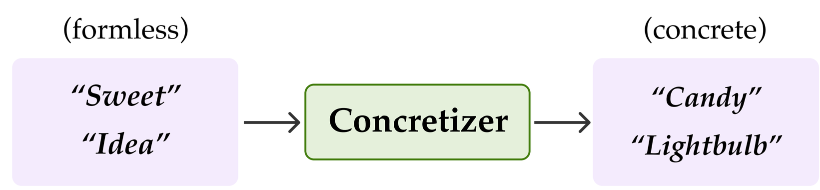
3. METHODS-CONCRETIZER

We used OpenAI's fine tuning API for the concretizer implementation. The dataset directly created a pair that matched formless word and concrete word (dataset size: 427).

Table 3. GPT-3.5 fine-tuning information

Model	Epoch	Batch size	Learning rate
GPT-3.5 text-davinci-003	4	0.2% of dataset	0.05

The converted words become the input of existing model and are reflected as semantics in the transformed characters through the transformation process.



4. EVALUATION

Although the evaluation criteria for typography may be subjective, we have defined two objectives indicating effectiveness of a typography. (1) The given concept of **meaning should be well expressed**, and (2) it should be **aesthetically beautiful**.
To verify this, three evaluations were conducted: perceptual study, ablation, and qualitative comparison.

4.1. Embedding loss - Ablation on loss function

In order to verify the effect of the changed loss function, we conducted ablation study by varying losses.

Table 4. The difference between the execution time and the number of hyperparameters according to loss functions

	Only SDS Loss	ACAP + Tone + SDS Loss (Word-as-Image)	Encoder + SDS Loss (ours)
Execution Time (min)	3.70	5.43	3.77
Hyperparameter #	0	30	1

Table 5. Qualitative comparison ma the loss functions

SDS Loss (No regulation)	LION	TENNIS	ENGINEER	MANGO	SUNFLOWER
ACAP + Tone + SDS Loss (Word-as-Image)	LION	TENNIS	ENGINEER	MANGO	SUNFLOWER
Encoder+SDS Loss (ours)	LION	TENNIS	ENGINEER	MANGO	SUNFLOWER

4.2. Concretizer - Qualitative comparison

Table 6. Qualitative comparisons between the original model and our model

Word-As-Image	SPICY	FANCY	WEIRD	STUDY	MINUTE	CREATIVE
NBTI (ours)	SPICY	FANCY	WEIRD	STUDY	MINUTE	CREATIVE

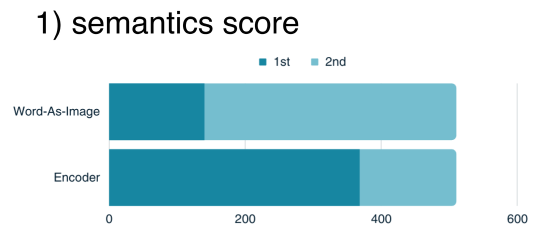
4.3. NBTI (Ours) - Perceptual study

We also conducted perceptual study to quantitatively evaluate the performance of our model.

We conducted a survey generated typography images for randomly selected 5 concrete words and 5 formless words on 108 students.

Figure 1. Perceptual study results on concrete words

1) semantics score



2) aesthetic score

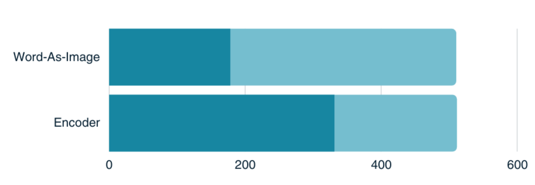
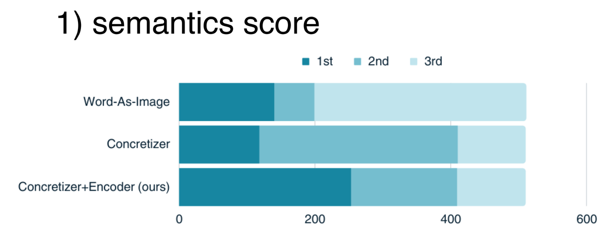
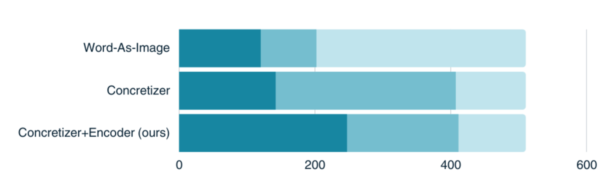


Figure 2. Perceptual study results on formless words

1) semantics score



2) aesthetic score



From the figure 1, we can observe the model with encoder got a significantly higher score (Semantics: 72.15% Aesthetic: 64.90%). From the figure 2, our model NBTI (concretizer + encoder) shows outstanding score among the 3 models.

5. CONCLUSION

We proposed an NN-based typography model NBTI that can visually represent letters, reflecting the meanings inherent in both concrete and formless words well.

1. NBTI showed a **44%** performance improvement in **execution time**.
2. NBTI offers **more freedom** in **transforming words** than other existing typography models.
3. NBTI excels at representing the **semantics of formless words** by finding alternative concrete words that symbolize them.

6. REFERENCES

[1] Iluz, Shir, et al. "Word-as-image for semantic typography." arXiv preprint arXiv:2303.01818 (2023).
[2] Brown, Tom, et al. "Language models are few-shot learners." Advances in neural information processing systems 33 (2020): 1877-1901.
[3] THOMAS "LIN Alphabet Characters Fonts Dataset" https://www.kaggle.com/datasets/thomasqazwsxedc/alphabet-characters-fonts-dataset
[4] Tendulkar, Purva et al. "Trick or TReAT : Thematic Reinforcement for Artistic Typography." ArXiv abs/1903.07820 (2019).