



ĐẠI HỌC QUỐC GIA TP.HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

oo

TĂNG CƯỜNG KHẢ NĂNG CHUYỂN KIẾU CHỮ ĐA NGÔN NGỮ TRONG BÀI TOÁN ONE-SHOT BẰNG MÔ HÌNH KHUẾCH TÁN

Sinh viên thực hiện:

Trần Đình Khánh Đăng

Giảng viên hướng dẫn:

TS. Dương Việt Hằng

Lớp khoá học:

KHMT2022.1

Khoa:

Khoa học máy tính

Mục lục

- 1. Giới thiệu**
- 2. Thách thức và Cơ sở lý thuyết**
- 3. Phương pháp đề xuất**
- 4. Thực nghiệm và Đánh giá**
- 5. Kết luận**

Mục lục

1. Giới thiệu

2. Thách thức và Cơ sở lý thuyết

3. Phương pháp đề xuất

4. Thực nghiệm và Đánh giá

5. Kết luận

Thiết kế phông chữ



TOP 4 FREE GOOGLE FONT COMBOS

by THE HUMANISTA CO.

Cormorant Garamond

SUBHEADING FUTURA FONT

This is the body text Futura font

FONT COMBO 1

Ovo

SUBHEADING WORK SANS FONT

This is the body text Ovo font

FONT COMBO 2

Poppins

Subheading Libre Bakersville Font

This is the body text Poppins font

FONT COMBO 3

Crimson Pro

SUBHEADING JOSEFIN SANS FONT

This is the body text Crimson Pro font

FONT COMBO 4

HOLISTIC DESIGN

WWW.THEHUMANISTA.CO

Ứng dụng rộng rãi của các phông chữ trong đời sống thực



Thách thức của thiết kế truyền thống

Mặc dù nhu cầu sử dụng phông chữ rất lớn, quy trình thiết kế truyền thống gấp nhiều trở ngại:

1. Tốn kém chi phí & thời gian:

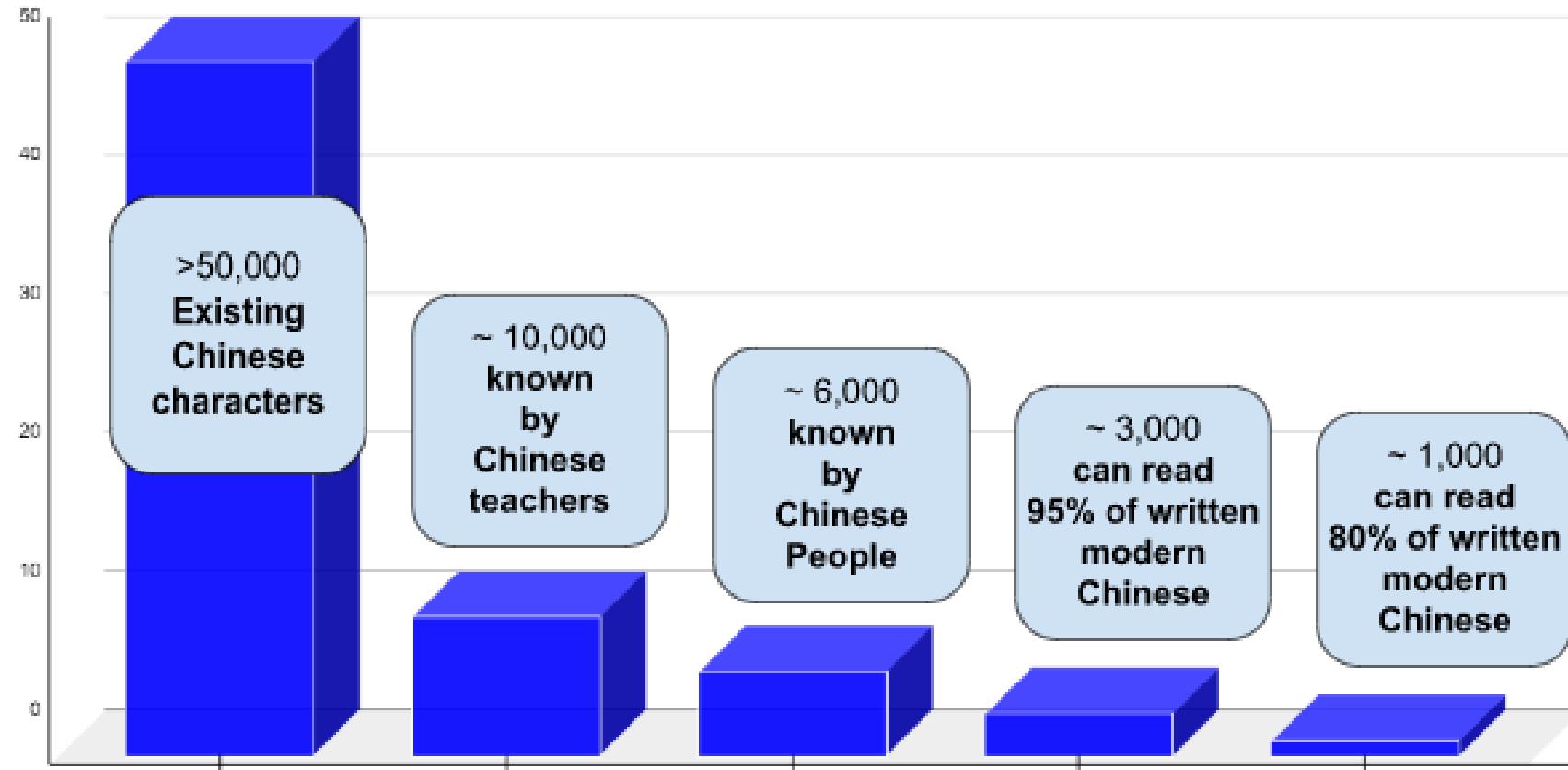
- Phải vẽ thủ công từng nét để đảm bảo tính thẩm mỹ.
- Quy trình lặp lại nhảm chán.

2. Thách thức về quy mô (Scale):

- Latin: Chỉ 52 ký tự.
- **CJK (Hán/Nôm):** Hàng chục nghìn ký tự. → **Rất tốn kém nếu làm thủ công hoàn toàn.**

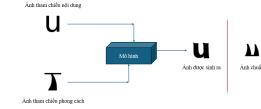
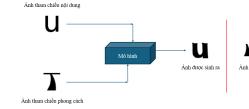
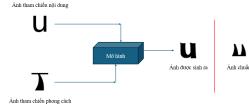
Thách thức của thiết kế truyền thống

NUMBER OF OFFICIAL CHINESE CHARACTERS



Giải pháp: One-shot Font Generation

Thay vì vẽ hàng ngàn đến hàng chục ngàn chữ, AI sẽ “học” phong cách từ **một chữ mẫu duy nhất**.



+

→

Nội dung

Phong cách mẫu

Kết quả

Lợi ích: Giảm chi phí thiết kế từ nhiều tháng xuống còn vài giây. Đồng thời ta có thể tăng cường các ký tự cho những ngôn ngữ ít được hỗ trợ như Thái, Việt, v.v...

Mục tiêu & Đóng góp của Khoa luận

Tuy nhiên, đa số mô hình hiện tại chỉ làm tốt trên đơn ngữ (VD: Hán → Hán).

Mục tiêu khoa luận: Xây dựng giải pháp **Cross-Lingual (Đa ngôn ngữ)** tổng quát.

→ **Phạm vi kiểm chứng (Scope):** Tập trung vào cặp **Latin - Hán tự**. (Lý do: Đây là cặp có cấu trúc khác biệt lớn nhất, đóng vai trò là trường hợp khó nhất để đánh giá khả năng của mô hình).

Đóng góp chính:

1. Xây dựng pipeline dựa trên **Diffusion Model** (thay vì GAN).
2. Đề xuất mô-đun **CL-SCR** để xử lý sự chênh lệch cấu trúc giữa hai hệ chữ này.

Mục lục

1. Giới thiệu

2. Thách thức và Cơ sở lý thuyết

3. Phương pháp đề xuất

4. Thực nghiệm và Đánh giá

5. Kết luận

Khoảng cách hình thái học (Morphological Gap)

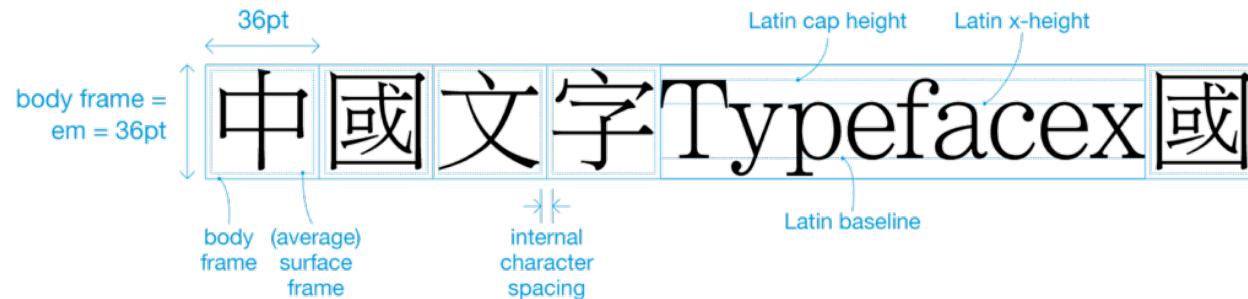
Tại sao cặp Latin - Hán tự lại là thách thức lớn nhất?

1. Latin (Đại diện hệ chữ cái):

- Cấu trúc tuyến tính (Linear).
- Ít nét, mật độ thưa.
- **Vấn đề:** Thiếu thông tin để suy diễn sang chữ phức tạp.

2. Hán tự (Đại diện hệ tượng hình):

- Cấu trúc khối vuông (Square block).
- Nét dày đặc, phức tạp.
- **Vấn đề:** Dễ bị biến dạng cấu trúc khi áp dụng phong cách lạ.



Khoảng cách hình thái học (Morphological Gap)

→ Khoảng cách (Gap) giữa hai nhóm này chính là rào cản lớn nhất mà mô hình cần vượt qua.

Tiếp cận giải quyết vấn đề

Với khoảng cách hình thái lớn như vậy, các phương pháp hiện tại xử lý ra sao?

1. Các phương pháp dựa trên GAN: 2. Tại sao chọn Diffusion Model?

(Ví dụ: DG-Font, FTransGAN)

- **Cơ chế:** Cố gắng học ánh xạ trực tiếp giữa hai miền ảnh.
- **Thất bại:** Do cấu trúc quá khác biệt, GAN thường sinh ra ảnh bị **Mode Collapse** (biến dạng) hoặc **Blur** (mờ) khi cố gắng “ép” chữ Latin thành khối vuông Hán tự.

- **Cơ chế:** Khử nhiễu dần dần (Denoising) từ trạng thái vô định hình.
- **Ưu điểm:** Cho phép kiểm soát cấu trúc (Structure) và phong cách (Style) tách biệt tốt hơn.

→ **Đây là chìa khóa để bắc cầu qua “Morphological Gap”.**

Mục lục

1. Giới thiệu

2. Thách thức và Cơ sở lý thuyết

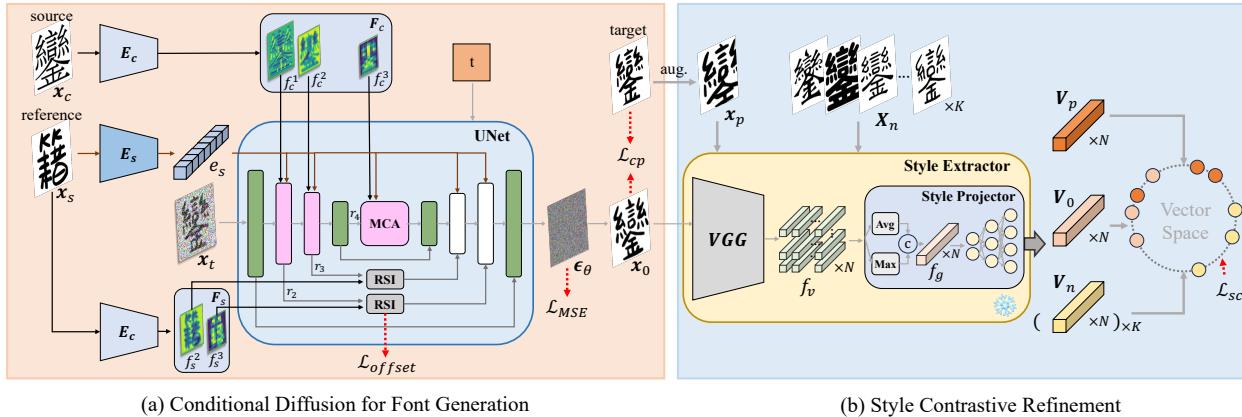
3. Phương pháp đề xuất

4. Thực nghiệm và Đánh giá

5. Kết luận

Tổng quan mô hình cải tiến

Kiến trúc FontDiffuser + CL-SCR Kế thừa FontDiffuser (AAAI'24) và đề xuất mô-đun mới cho đa ngôn ngữ.



3 Thành phần chính:

- MCA:** Tổng hợp nội dung đa tỷ lệ (Giữ nét).
- RSI:** Tương tác cấu trúc (Chống biến dạng).
- CL-SCR (New):** Tinh chỉnh phong cách xuyên ngôn ngữ.

Trọng tâm: Module CL-SCR

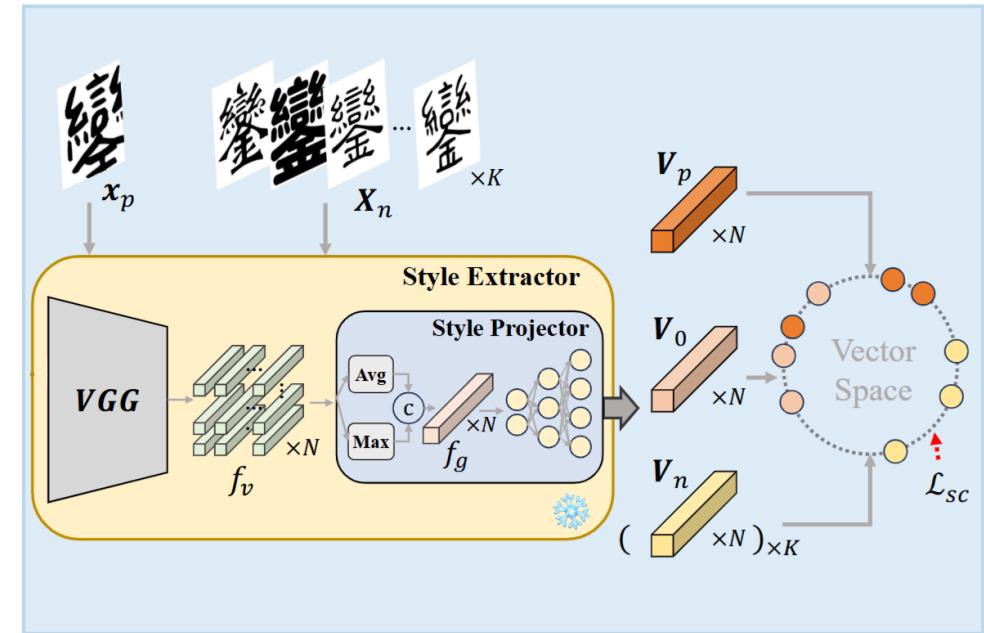
Cross-Lingual Style Contrastive Refinement Giải pháp cho vấn đề “Domain Gap” giữa hai ngôn ngữ.

Cải tiến cốt lõi:

- **Mở rộng mẫu âm (Negative Samples):** Lấy mẫu từ cả hai miền ngôn ngữ.
- **Chiến lược Loss hỗn hợp:**

$$L = \alpha \cdot L_{\text{intra}} + \beta \cdot L_{\text{cross}}$$

- Tăng trọng số $\beta = 0.7$ để ép mô hình học sự tương đồng phong cách bất kể ngôn ngữ nào.



Cơ chế tương phản đa miền

Mục lục

- 1. Giới thiệu**
- 2. Thách thức và Cơ sở lý thuyết**
- 3. Phương pháp đề xuất**
- 4. Thực nghiệm và Đánh giá**
- 5. Kết luận**

Thiết lập thực nghiệm

Dữ liệu & Kịch bản đánh giá

Bộ dữ liệu:

- 818 Font chữ (Bao gồm cả ký tự Latin và Hán tự).
- Chia tập: Train / Validation / Test.

Kịch bản khó nhất (Unseen Font):

- Đánh giá trên các font chữ **mô hình chưa từng thấy** trong quá trình huấn luyện.
- **Mục tiêu:** Kiểm chứng khả năng tổng quát hoá (Generalization).

Kết quả định lượng (Quantitative)

So sánh hiệu năng (FID Score - Thấp hơn là tốt hơn) Mô hình đề xuất vượt trội so với Baseline (FontDiffuser gốc) và các phương pháp GAN.

Method	Latin → Hán	Hán → Latin
FTransGAN (GAN)	45.20	82.15
DG-Font (GAN)	32.14	75.40
FontDiffuser (Baseline)	29.10	59.58
Ours (CL-SCR)	13.55 (Giảm 50%)	41.11 (Tốt nhất)

Nhận xét: CL-SCR giúp cải thiện đáng kể độ chân thực của ảnh sinh ra.

Kết quả định tính (Qualitative)

So sánh trực quan (Visual Comparison)

Chiều Latin → Hán:

- **Ours:** Tái tạo đúng nét cọ xước, đậm nhạt.
- **Baseline:** Nét đôi khi bị cứng hoặc sai độ đậm.

Chiều Hán → Latin:

- **Ours:** Giữ cấu trúc chữ rõ ràng.
- **DG-Font:** Bị lỗi “Content Leakage” (chữ Latin biến thành Hán).

Mục lục

1. Giới thiệu

2. Thách thức và Cơ sở lý thuyết

3. Phương pháp đề xuất

4. Thực nghiệm và Đánh giá

5. Kết luận

Tổng kết & Hướng phát triển

- ★ **Vấn đề:** Đã giải quyết bài toán chuyển đổi phong cách đa ngôn ngữ (Cross-Lingual) khó nhằn.
- ✓ **Giải pháp:** Đề xuất mô-đun **CL-SCR** với cơ chế Loss hỗn hợp.
- ✓ **Kết quả:** Vượt trội SOTA hiện tại (FID giảm 50% ở chiều E2C).

Hướng phát triển:

- Tối ưu tốc độ sinh ảnh (Fast Sampling/Distillation).
- Mở rộng sang tiếng Việt (Thư pháp/Quốc ngữ) và tiếng Nhật.

Công bố liên quan

Lời cảm ơn

**Xin cảm ơn Thầy Cô và Hội đồng
đã theo dõi và lắng nghe!**

Sinh viên thực hiện:

Trần Đình Khánh Đăng

Giảng viên hướng dẫn:

TS. Dương Việt Hằng

Lớp khoá học:

KHMT2022.1

Khoa:

Khoa học máy tính