

PHẦN 1: GIỚI THIỆU (SLIDE 1 - 9)

Slide 1: Xin chào

Kính thưa Hội đồng bảo vệ, quý Thầy Cô và các bạn. Em tên là Trần Đình Khánh Đăng. Hôm nay, em xin phép trình bày khoá luận tốt nghiệp với đề tài: 'Tăng cường khả năng chuyển kiều chữ đa ngôn ngữ trong bài toán One-Shot bằng mô hình khuếch tán'.

Slide 2 & 3: Mục lục

Bài báo cáo của em sẽ đi qua 5 phần chính: Bắt đầu từ việc đặt vấn đề, phân tích các thách thức hiện tại, đi sâu vào phương pháp đề xuất, chứng minh bằng thực nghiệm và cuối cùng là kết luận.

Slide 4 & 5: Ứng dụng thực tế

Kính thưa hội đồng, chúng ta có thể thấy phong chữ hiện diện ở khắp mọi nơi, từ bao bì sản phẩm đến các biển hiệu quảng cáo. Nhu cầu về các bộ font chữ độc đáo, thẩm mỹ chưa bao giờ hạ nhiệt trong đời sống hiện đại.

Slide 6: Thách thức thiết kế truyền thống

Tuy nhiên, quy trình thiết kế font truyền thống đang gặp phải 3 rào cản rất lớn:

- Thứ nhất là Tốn kém:** Các nhà thiết kế phải vẽ thủ công từng nét, cực kỳ mất thời gian.
- Thứ hai là Quy mô:** Đặc biệt với hệ chữ tượng hình như Hán tự (CJK), số lượng ký tự lên tới hàng chục nghìn, việc vẽ tay toàn bộ là gần như bất khả thi.
- Và thứ ba là Rào cản đa ngữ:** Các font đẹp thường chỉ hỗ trợ tiếng Anh hoặc Trung, thiếu các ký tự mở rộng cho tiếng Việt hay tiếng Thái, gây khó khăn khi muốn đồng bộ thương hiệu toàn cầu.

Slide 7: Thông kê ký tự

Để hình dung rõ hơn về vấn đề quy mô, biểu đồ này cho thấy: dù chỉ cần khoảng 3.000 chữ để đọc hiểu văn bản, nhưng tổng số lượng Hán tự thực tế lên tới hơn 50.000. Đây là một khối lượng công việc khổng lồ nếu làm thủ công.

Slide 8 & 9: Giải pháp One-shot Font Generation

Chính vì vậy, giải pháp **One-shot Font Generation** ra đời. Mục tiêu của bài toán này là: AI chỉ cần nhìn **một mẫu duy nhất** (ví dụ chữ 'M') để học phong cách, sau đó nó tự động sinh ra toàn bộ các ký tự còn lại. Điều này giải quyết triệt để bài toán về tốc độ và chi phí.

Slide 10: Mục tiêu & Đóng góp

Tuy nhiên, đa số các mô hình hiện nay chỉ làm tốt trên đơn ngữ (ví dụ Hán sang Hán). **Mục tiêu của khoá luận** là xây dựng giải pháp **Cross-Lingual** (Đa ngôn ngữ), tập trung vào cản khó nhất là **Latin và Hán tự**. Đóng góp chính của em bao gồm:

1. Xây dựng pipeline dựa trên **Diffusion Model** thay vì GAN truyền thống.
 2. Đề xuất mô-đun mới tên là **CL-SCR** để xử lý sự khác biệt cấu trúc giữa hai hệ chữ này.
-

PHẦN 2: THÁCH THỨC & CƠ SỞ LÝ THUYẾT (SLIDE 11 - 14)

Slide 11: (Chuyển tiếp)

Tiếp theo, em xin phân tích sâu hơn về thách thức lý thuyết mà đề tài giải quyết.

Slide 12 & 13: Khoảng cách hình thái học (Morphological Gap)

Tại sao cặp Latin - Hán tự lại là thử thách lớn nhất? Như thầy cô thấy trên hình

Như thầy cô thấy trên hình:

- **Latin:** Cấu trúc tuyến tính, ít nét, mật độ thưa.
- **Hán tự:** Cấu trúc khối vuông, nét dày đặc và phức tạp.

Sự chênh lệch này tạo ra một 'Khoảng cách hình thái học' (Morphological Gap) rất lớn, khiến việc chuyển phong cách từ chữ này sang chữ kia rất dễ bị biến dạng.

Slide 14: Tiếp cận giải quyết

Các phương pháp cũ dựa trên GAN thường thất bại ở đây – ảnh sinh ra hay bị mờ hoặc vỡ nét – do cơ chế ánh xạ trực tiếp của GAN cố gắng 'ép' cấu trúc Latin vào khuôn Hán tự. Vì vậy, em chọn tiếp cận bằng **Diffusion Model**. Với cơ chế khử nhiễu dần dần, Diffusion cho phép kiểm soát tách biệt giữa **Cấu trúc** và **Phong cách**. Đây chính là chìa khoá để bắc cầu qua khoảng cách hình thái này.

PHẦN 3: PHƯƠNG PHÁP ĐỀ XUẤT (SLIDE 15 - 21)

Slide 15: (Chuyển tiếp)

Sau đây là chi tiết phương pháp đề xuất.

Slide 16: Tổng quan kiến trúc

Kính thưa hội đồng, kiến trúc của em được xây dựng dựa trên nền tảng FontDiffuser và huấn luyện qua 2 giai đoạn.

- **Giai đoạn 1 (Phase 1 - Bên trái):** Em giữ nguyên các mô-đun MCA và RSI. Nhiệm vụ của chúng là đảm bảo chữ sinh ra không bị mất nét hay méo mó.
- **Giai đoạn 2 (Phase 2 - Bên phải):** Đây là nơi chứa đóng góp cốt lõi của khoá luận. Nhận thấy module SCR cũ hoạt động kém trên dữ liệu đa ngữ, em đã thay thế hoàn toàn bằng mô-đun CL-SCR (Cross-Lingual Style Contrastive Refinement) để mô hình có thể học phong cách xuyên ngôn ngữ một cách chính xác."

Slide 17: Ý tưởng CL-SCR

Vậy CL-SCR có gì khác biệt? Nó được thiết kế lại để giải quyết vấn đề 'Domain Gap'. Thay vì chỉ lấy mẫu ngẫu nhiên, em mở rộng chiến lược lấy mẫu thành 2 luồng:

- **Intra-domain (Nội miền):** Lấy cặp ảnh cùng ngôn ngữ (Latin-Latin) để giữ sự ổn định nội tại.
- **Cross-domain (Xuyên miền):** Lấy cặp ảnh Latin và Hán tự. Đây là điểm mấu chốt để học cách chuyển giao phong cách sang miền đích.

Slide 18: Sơ đồ CL-SCR (Chi tiết)

Đi sâu vào sơ đồ hoạt động: Khi mô hình sinh ra một ảnh (x_{sample}), nó sẽ trích xuất đặc trưng và so sánh đồng thời trên hai luồng:

- **Luồng trên (Intra):** So khớp với chữ cái cùng hệ (ví dụ Latin với Latin).
- **Luồng dưới (Cross):** So khớp với chữ Hán đích (Target). Mục tiêu là kéo ảnh sinh lại gần phong cách đích, bắt cháp sự khác biệt về nội dung chữ.

Slide 19 & 20: Công thức Loss

Để hiện thực hóa điều này về mặt toán học, hàm Loss được xây dựng dựa trên nguyên lý InfoNCE:

- L_{intra} : Đảm bảo tính nhất quán trong nội bộ ngôn ngữ.
- L_{cross} : Kéo phong cách Latin lại gần phong cách Hán tự.

Tổng Loss CL-SCR là sự kết hợp có trọng số của hai thành phần này.

Slide 21: Hàm mục tiêu tổng quát

Tổng kết lại, hàm mục tiêu để huấn luyện toàn bộ mô hình là sự tổng hòa của:

- $\$L_{\text{MSE}}$: Để tái tạo điểm ảnh.
 - $\$L_{\text{percep}}$ & $\$L_{\text{offset}}$: Để giữ cấu trúc.
 - Và $\$L_{\text{CL-SCR}}$: Để tinh chỉnh phong cách đa ngôn ngữ."
-

PHẦN 4: THỰC NGHIỆM & ĐÁNH GIÁ (SLIDE 22 - 30)

Slide 22 & 23: Bộ dữ liệu

Chuyển sang phần thực nghiệm. Em kế thừa bộ dữ liệu chuẩn từ FTransGAN với 818 bộ font song ngữ. Đặc điểm của bộ này là sự đồng bộ tuyệt đối về phong cách giữa Latin và Hán tự, cung cấp ground-truth chính xác cho việc đánh giá.

Slide 24: Kịch bản đánh giá

Trong các kịch bản kiểm thử, em tập trung báo cáo kịch bản khó nhất là **UFSC (Unseen Font)** - tức là đưa vào một font chữ hoàn toàn mới, chưa từng gặp khi huấn luyện. Đây là thước đo chuẩn xác nhất cho khả năng tổng quát hoá của mô hình.

Slide 25: Cấu hình huấn luyện

Về cấu hình, mô hình được huấn luyện trên GPU Tesla P100. Đặc biệt ở Phase 2, em áp dụng kỹ thuật **Data Augmentation** (như Random Crop) cho module CL-SCR. Điều này giúp mô hình học được các đặc trưng phong cách bền vững hơn thay vì học vẹt vị trí pixel.

Slide 26: Thước đo đánh giá

"Việc đánh giá được thực hiện đa tầng:

- **Định lượng:** Dùng FID (độ chân thực) và LPIPS (độ tương đồng nhận thức).
- **Định tính:** So sánh bằng mắt thường và khảo sát người dùng (User Study)."

Slide 27: Kết quả Định lượng

Kính mời thầy cô nhìn vào bảng kết quả. Ở kịch bản khó nhất (UFSC), phương pháp của em (**Ours - dòng cuối**) đạt chỉ số FID là **13.55**. Con số này thấp hơn rất nhiều so với các phương pháp SOTA như DG-Font hay FontDiffuser gốc (vốn có FID trên 29). Điều này khẳng định chất lượng ảnh sinh ra có độ chân thực vượt trội.

Slide 28: Kết quả Định tính

Trực quan hơn, xin mời xem hình ảnh so sánh này: Thầy cô có thể thấy ở cột cuối cùng (Ours), các chữ Latin (như chữ c, d, e...) được tái tạo rất sắc nét. Đặc biệt, nó bắt chước chính xác độ đậm nhạt và các nét xước của chữ Hán mẫu (Reference). Trong khi đó, các phương pháp cũ thường bị mờ hoặc mất nét.

Slide 29: Đánh giá người dùng

Kết quả khảo sát mù trên 21 người dùng cũng cho thấy sự đồng thuận cao, với gần **70%** số phiếu bình chọn ảnh sinh ra từ mô hình của em có chất lượng tốt nhất.

Slide 30: Phân tích hiệu quả (Ablation Study)

Để trả lời câu hỏi 'Liệu CL-SCR có thực sự cần thiết?', bảng phân tích cắt giảm này cho thấy: Nếu bỏ thành phần Cross-Lingual đi, chỉ số FID tăng vọt (tức là xấu đi). Kết quả tốt nhất chỉ đạt được khi kết hợp đầy đủ cả hai luồng Intra và Cross trong mô-đun đề xuất.

PHẦN 5: KẾT LUẬN (SLIDE 31 - 35)

Slide 31: (Chuyển tiếp)

Cuối cùng là phần kết luận.

Slide 32: Tổng kết đóng góp

Tổng kết lại, khoá luận của em đã hoàn thành 3 mục tiêu lớn:

1. Giải quyết thành công bài toán One-shot Cross-Lingual.
2. Đóng góp về mặt kỹ thuật với mô-đun **CL-SCR** và cơ chế Loss hỗn hợp.
3. Chứng minh được hiệu quả thực nghiệm vượt trội so với SOTA hiện tại.

Slide 33: Hạn chế & Hướng phát triển

Tuy nhiên, em cũng nhìn nhận thắc mắc hạn chế lớn nhất là tốc độ suy diễn. Do bản chất phải khử nhiễu qua 20 bước, mô hình chậm hơn GAN khá nhiều. Hướng phát triển tiếp theo của em là áp dụng các kỹ thuật Chung cát (Distillation) để giảm số bước sinh ảnh, đồng thời mở rộng nghiên cứu sang sinh font dạng Vector (SVG) để ứng dụng thực tiễn tốt hơn.

Slide 34: Công bố liên quan

Về phần công bố liên quan, em đã có submit một bài báo lên The Journal of Supercomputing của Springer, hiện đang under review.

Slide 35: Lời cảm ơn

Em xin chân thành cảm ơn sự hướng dẫn tận tình của Tiên sĩ Dương Việt Hằng và sự lắng nghe của quý Thầy Cô trong Hội đồng. Em xin kết thúc phần trình bày tại đây và rất mong nhận được những câu hỏi, góp ý từ thầy cô.