

# THÔNG TIN CHUNG CỦA BÁO CÁO

- Link YouTube video của báo cáo (tối đa 5 phút):  
<https://www.youtube.com/watch?v=W5Gy0123OyA>
- Link slides (dạng .pdf đặt trên Github):  
<https://github.com/Aph3li0s/CS519.PPLNCKH/blob/main/CS519.O21.KHTN.DeCuong.FinalReport.AIO.Slide.pdf>
- *Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới*
- *Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in*

<ul style="list-style-type: none"><li>● Họ và Tên: Phan Huỳnh Ngọc Trâm</li><li>● MSSV: 22521500</li></ul> 	<ul style="list-style-type: none"><li>● Lớp: CS519.O21.KHTN</li><li>● Tự đánh giá (điểm tổng kết môn): 9/10</li><li>● Số buổi vắng: 0</li><li>● Số câu hỏi QT cá nhân:</li><li>● Số câu hỏi QT của cả nhóm:</li><li>● Link Github: <a href="https://github.com/Aph3li0s">https://github.com/Aph3li0s</a></li><li>● Mô tả công việc và đóng góp cá nhân cho kết quả của cả nhóm:<ul style="list-style-type: none"><li>○ Lên ý tưởng đề tài</li><li>○ Viết nội dung, phương pháp báo cáo</li><li>○ Làm slide</li></ul></li></ul>
--	--

- Họ và Tên: Phạm Thạch Thanh Trúc
- MSSV: 22521551



- Lớp: CS519.O21.KHTN
- Tự đánh giá (điểm tổng kết môn): 9/10
- Số buổi vắng: 0
- Số câu hỏi QT cá nhân:
- Số câu hỏi QT của cả nhóm:
- Link Github:  
<https://github.com/KoryTheKorgy>
- Mô tả công việc và đóng góp cá nhân cho kết quả của cả nhóm:
  - Viết phần mở đầu, giới thiệu báo cáo
  - Chỉnh sửa báo cáo, slide
  - Làm video Youtube

# ĐỀ CƯƠNG NGHIÊN CỨU

## TÊN ĐỀ TÀI (IN HOA)

SÁNG TẠO KÍ TỰ CÁCH ĐIỀU TỪ HÌNH ẢNH TRONG THIẾT KẾ CHỮ NGHỆ THUẬT

## TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

CREATIVE LETTERING FROM ILLUSTRATED IMAGE IN ARTISTIC TYPOGRAPHY

## TÓM TẮT *(Tối đa 400 từ)*

Sự sáng tạo trong lĩnh vực thiết kế đồ họa không chỉ đòi hỏi kỹ năng chuyên môn và khả năng tư duy sáng tạo mà còn cần sự tinh tế trong việc kết hợp khéo léo và hài hòa giữa các yếu tố màu sắc, hình ảnh và kiểu chữ. Hiện nay, các mô hình tạo sinh đã đạt được nhiều thành tựu trong việc tạo ra các thiết kế độc đáo và đa dạng, từ việc tạo hình ảnh chân thực đến các tác phẩm nghệ thuật trừu tượng. Tuy nhiên, trong lĩnh vực sáng tạo chữ nghệ thuật (typography), các mô hình sinh ảnh hiện tại vẫn gặp nhiều khó khăn trong việc cách điệu các ký tự để phản ánh đặc điểm và ý nghĩa của hình ảnh mà chúng mô tả. Điều này dẫn đến sự hạn chế trong việc tạo ra các thiết kế chữ nghệ thuật có tính sáng tạo và độc đáo, đồng thời làm giảm khả năng truyền tải thông điệp một cách hiệu quả và tinh tế. Mặt khác, các mô hình tạo sinh ảnh hiện tại chỉ có thể xuất ảnh dưới dạng raster, tức là ảnh được định dạng bằng các điểm ảnh, vốn không phù hợp với yêu cầu của lĩnh vực thiết kế đồ họa. Để giải quyết vấn đề này, chúng tôi đề xuất phương pháp kết hợp khả năng sáng tạo của các mô hình trên để cách điệu ký tự, đồng thời chuyển đổi ảnh tạo ra sang định dạng vector. Các ký tự đầu vào sẽ được chuyển đổi thành phông chữ phù hợp và được trích xuất các điểm kiểm soát từ thư viện FreeType. Sau đó, chúng tôi fine-tune mô hình Stable Diffusion bằng cách tối ưu hóa các điểm kiểm soát này để cách điệu các ký tự. Kết quả là ảnh đầu ra sẽ được vector hóa để đạt được định dạng cần thiết cho thiết kế đồ họa. Các ký tự cách điệu do mô hình tạo ra giúp con người dễ dàng liên tưởng đến từ ngữ mà chúng

biểu thị, đồng thời vẫn giữ được hình dáng gốc ban đầu của chúng, giúp tăng cường khả năng nhận diện và truyền tải thông điệp một cách hiệu quả.

## GIỚI THIỆU (Tối đa 1 trang A4)

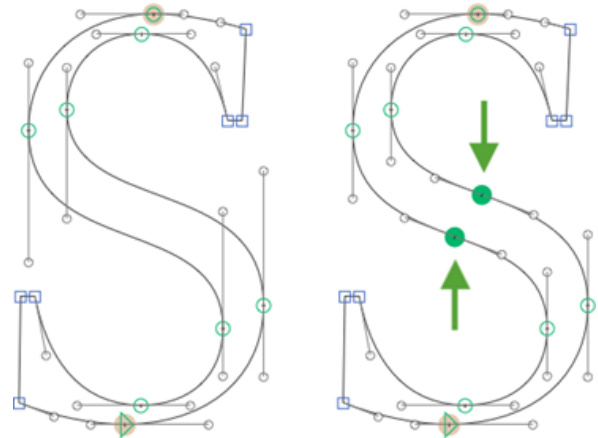
- Typography, hay còn gọi là nghệ thuật tạo hình chữ cái, đóng vai trò quan trọng trong thiết kế đồ họa và in ấn, đặc biệt trong việc tạo ra phong chữ độc đáo. Vì thế, yêu cầu được đặt ra đối với các nhà thiết kế là cần phải giữ được hình dáng của các kí tự ở mức độ nhất định để người xem có thể nhận ra kí tự ban đầu, nhưng vẫn có sự sáng tạo độc nhất. Việc này đã tạo ra các thách thức nhất định cho các mô hình sinh ảnh. Hiện nay, những mô hình sinh ảnh như Stable Diffusion [4] dù có thể đưa ra hình ảnh phù hợp với từ, nhưng chúng không thể thay đổi cấu trúc kí tự mà chỉ chèn hình ảnh vào từ được cho. Mặt khác, các phương pháp sáng tạo chữ nghệ thuật từ hình ảnh như [2] tuy có thể chèn hình ảnh và đồng thời giữ được hình dáng của kí tự nhưng chúng lại thiếu tính sáng tạo do giới hạn bởi số lượng hình ảnh và clipart. Hơn nữa, tất cả các phương pháp này đều tạo ra hình ảnh raster, vốn dễ mờ hoặc vỡ hình khi phóng to hoặc thu nhỏ. Vì thế, đồ họa vector được sử dụng thay thế bởi chúng cho phép thay đổi kích thước mà không mất đi chất lượng ban đầu, vì các hình dạng được vẽ lại chính xác theo các phép biến đổi vector.
- Vì thế, chúng tôi đề xuất một phương pháp giúp tạo ra các kí tự cách điệu độc đáo, mang màu sắc của từ ngữ đã cho và kết hợp hài hòa với hình ảnh mô tả. Kí tự cách điệu cần mang tính sáng tạo, nhưng phải duy trì hình dáng kí tự gốc nhằm đảm bảo tính dễ đọc. Ngoài ra, các chữ nghệ thuật này được định dạng ở đồ họa vector, giúp cho việc in ấn, chỉnh sửa đồ họa dễ dàng hơn.



Stable Diffusion v3 [4]	Phương pháp [2] kết hợp các vật thể với nhau nhằm tạo ra các kí tự thuộc chủ đề đồ ăn
<ul style="list-style-type: none"> <li>● <b>Input:</b> Một từ biểu thị ý nghĩa của một danh từ, và yêu cầu chỉnh sửa một kí tự thuộc từ đó cùng phong chữ được dùng để hiển thị từ.</li> <li>● <b>Output:</b> Từ được hiển thị bằng phong chữ trên dưới định dạng vector, trong đó kí tự yêu cầu sẽ được cách điệu với hình ảnh mà từ biểu thị sao cho kí tự đó mang sự kết hợp giữa hình ảnh mô tả và hình dáng đặc trưng của kí tự.</li> </ul>	
<div> <div> <b>Font:</b> Sans Serif  <b>Word:</b> BIRD  <b>Letter:</b> R </div> <div> </div> </div>	
<b>MỤC TIÊU</b> (Viết trong vòng 3 mục tiêu, lưu ý về tính khả thi và có thể đánh giá được) <ul style="list-style-type: none"> <li>● Tạo ra hai hình ảnh từ mô hình sinh ảnh: Hình ảnh minh họa cho từ khóa trong văn bản mô tả và hình ảnh hiển thị từ khóa dưới dạng phong chữ được yêu cầu.</li> <li>● Chuyển đổi hình ảnh đầu ra từ định dạng raster thành định dạng vector</li> <li>● Cách điệu kí tự bằng hình ảnh sao cho kí tự cách điệu mang được hình ảnh mô tả từ và giữ được hình dáng của kí tự gốc</li> </ul>	
<b>NỘI DUNG VÀ PHƯƠNG PHÁP</b>	

(Viết nội dung và phương pháp thực hiện để đạt được các mục tiêu đã nêu)

Trong đồ họa vector, các chữ cái sẽ được biểu diễn bằng các điểm kiểm soát (control points), giúp tạo thành các đường cong Bézier. Số lượng điểm kiểm soát càng nhiều, khả năng thể hiện các chi tiết phức tạp càng cao. Vì thế, các nhà thiết kế thường sẽ thực hiện chỉnh sửa các điểm kiểm soát nhằm tạo ra tác phẩm mong muốn. Đây chính là điểm mấu chốt trong phương pháp đề xuất của chúng tôi nhằm có thể cách điệu hóa các ký tự.



**Nội dung 1:** Fine-tune mô hình Stable Diffusion theo [6] nhằm có thể đưa ra các hình ảnh phù hợp dựa trên từ khóa được cung cấp qua văn bản mô tả, và phong chữ phù hợp của từ khóa.

- **Sử dụng mô hình pretrained Stable Diffusion (SD) [4] kết hợp với Contrastive Language-Image Pre-Training (CLIP)[5]:** Sử dụng mô hình đã được huấn luyện trước giúp tiết kiệm thời gian và tài nguyên, đồng thời đảm bảo chất lượng tốt hơn vì mô hình đã được học từ một lượng lớn dữ liệu văn bản và hình ảnh, đồng thời cung cấp nền tảng mạnh mẽ cho quá trình fine-tune. Nhờ đó, SD có thể cho ra các hình ảnh chất lượng, phản ánh chính xác ý nghĩa ngữ nghĩa của từ khóa được cung cấp từ người dùng.
- **Giới hạn ngữ cảnh bằng prompt:** Để đảm bảo rằng hình ảnh được tạo ra phù hợp với phong cách mong muốn, chúng tôi giới hạn ngữ cảnh của bài toán trên bằng cách mô tả trong input prompt text của Stable Diffusion. Vì vậy, các mô tả cần phải cụ thể và chi tiết. Chúng tôi giới hạn mô hình bằng cụm sau: *“minimal flat 2d vector icon, lineal color, on a white background, trending on artstation”*, nhằm hướng dẫn mô hình tạo ra hình ảnh phù hợp.
- **Sử dụng FreeType và kỹ thuật Masking:** Chúng tôi sử dụng thư viện

FreeType nhằm tạo ra phong chữ đúng với yêu cầu. Sau đó, kỹ thuật Masking được áp dụng trên phong chữ nhằm chỉ định kí tự cần thay đổi và bỏ qua các kí tự còn lại. Điều này đảm bảo rằng các kí tự còn lại vẫn duy trì được cấu trúc gốc trong quá trình cách điệu kí tự còn lại.

**Nội dung 2:** Áp dụng kỹ thuật được đề xuất bởi VectorFusion [6] nhằm chuyển đổi ảnh được sinh bởi Stable Diffusion [4] từ ảnh raster thành ảnh vector. Kỹ thuật này gồm 3 bước chính:

- Để tối ưu hóa các tham số, chúng tôi sử dụng kỹ thuật **Score Distillation Sampling (SDS)**[9]. Kỹ thuật này giúp tính toán gradient để có các điều chỉnh cần thiết nhằm giảm thiểu sự khác biệt giữa hình ảnh và văn bản mô tả, đồng thời đảm bảo hình ảnh phản ánh đúng ý nghĩa ngữ nghĩa. SDS giúp đảm bảo rằng ảnh raster được tối ưu hóa đúng cách trước khi chuyển đổi thành vector.
- Sau đó, **Layer-wise Image Vectorization (LIVE)**[8] được áp dụng để chuyển đổi hình ảnh này thành vector. Phương pháp phân tích hình ảnh raster và xác định các vùng có sai số cao. Sau đó, LIVE điều chỉnh các đường cong Bézier để tái tạo các chi tiết phức tạp của hình ảnh raster thành hình ảnh vector một cách chính xác. Quá trình này đảm bảo rằng các chi tiết nhỏ nhất của hình ảnh gốc được giữ nguyên và được tái tạo lại một cách chính xác trong hình ảnh vector.
- Cuối cùng, thư viện **DiffVG**[9] được sử dụng nhằm điều chỉnh các điểm kiểm soát. DiffVG đảm bảo rằng các đường cong được điều chỉnh liên tục và mượt mà, tránh các gãy khúc hay bất thường trong hình dạng trong quá trình biến đổi từ ảnh raster thành vector, từ đó đảm bảo ảnh vector mang chất lượng và độ chính xác cao.

**Nội dung 3:** Trích xuất các đường cong Bézier và điểm kiểm soát từ kí tự qua thư viện FreeType. Từ đó, chúng tôi đề xuất các phương pháp sau nhằm có thể cách điệu kí tự:

- Trong quá trình cách điệu, rất dễ xảy ra các trường hợp hình ảnh không mang hình dạng của kí tự. Vì thế, chúng tôi **tạo ra một tập hợp mới các điểm kiểm soát dựa trên khuôn khổ của các tam giác Delaunay**. Việc này giúp tạo ra một cấu trúc lưới nhằm dễ dàng quản lý và tinh chỉnh các đường cong Bézier một cách hiệu quả.
- Khi tinh chỉnh hình ảnh, ta cần bảo toàn hình dạng kí tự gốc. Sự biến dạng của điểm kiểm soát có thể thay đổi hình dáng kí tự, và thay đổi độ sáng/tông màu ảnh hưởng đến tính nhất quán và nhận diện. Do đó, chúng tôi áp dụng ràng buộc **biến dạng càng đồng dạng càng tốt (ACAP)[10]** nhằm duy trì hình dạng và cấu trúc của kí tự gốc đối khi thực hiện cách điệu.
- Sau khi có được kí tự cách điệu, ta cần đưa chúng vào từ ban đầu. Tuy nhiên, việc tinh chỉnh trên đã thay đổi hình dáng của kí tự cách điệu, qua đó, cần phải **điều chỉnh lại khoảng cách giữa các kí tự thường và kí tự cách điệu**. Để thực hiện điều này, chúng tôi đề xuất kỹ thuật Kerning nhằm đảm bảo rằng các kí tự cách điệu và các kí tự thường được đặt cách nhau một cách đồng đều và thẩm mỹ, giữ cho từ hoàn chỉnh trông gọn gàng và dễ đọc.

## KẾT QUẢ MONG ĐỢI

*(Viết kết quả phù hợp với mục tiêu đặt ra, trên cơ sở nội dung nghiên cứu ở trên)*

- Hình ảnh minh họa được tạo ra mô tả chính xác ý nghĩa của từ khóa, và hình ảnh của từ khóa được thể hiện bằng phong chữ yêu cầu.
- Kết quả đầu ra là ảnh vector, đảm bảo tính nhất quán và độ chính xác so với phiên bản ảnh raster.
- Kí tự được cách điệu cùng với hình ảnh minh họa, có sự sáng tạo nhưng vẫn giữ được hình dáng ban đầu của kí tự gốc. Ngoài ra, đối với đánh giá từ con người, kí tự cách điệu cần phải có mức độ nhận diện cao cả về mặt kí tự lẫn hình ảnh mô tả.



## **TÀI LIỆU THAM KHẢO** (*Định dạng DBLP*)

- [1] S. Iluz, Y. Vinker, A. Hertz, D. Berio, D. Cohen-Or, and A. Shamir, ‘Word-As-Image for Semantic Typography’, *ACM Trans. Graph.*, vol. 42, no. 4, p. 151:1-151:11, 2023.
- [2] J. Zhang, Y. Wang, W. Xiao, and Z. Luo, ‘Synthesizing Ornamental Typefaces’, *Comput. Graph. Forum*, vol. 36, no. 1, pp. 64–75, 2017.
- [3] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, ‘High-Resolution Image Synthesis with Latent Diffusion Models’, in *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022, 2022*, pp. 10674–10685.
- [4] P. Esser et al., ‘Scaling Rectified Flow Transformers for High-Resolution Image Synthesis’, *CoRR*, vol. abs/2403.03206, 2024.
- [5] A. Radford et al., ‘Learning Transferable Visual Models From Natural Language Supervision’, in *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event, 2021*, vol. 139, pp. 8748–8763.
- [6] A. Jain, A. Xie, and P. Abbeel, ‘VectorFusion: Text-to-SVG by Abstracting Pixel-Based Diffusion Models’, in *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023, 2023*, pp. 1911–1920.
- [7] B. Poole, A. Jain, J. T. Barron, and B. Mildenhall, ‘DreamFusion: Text-to-3D using 2D Diffusion’, in *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023, 2023*.
- [8] X. Ma et al., ‘Towards Layer-wise Image Vectorization’, in *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022, 2022*, pp. 16293–16302.
- [9] T.-M. Li, M. Lukáč, G. Michaël, and J. Ragan-Kelley, ‘Differentiable Vector Graphics Rasterization for Editing and Learning’, *ACM Trans. Graph. (Proc. SIGGRAPH Asia)*, vol. 39, no. 6, p. 193:1-193:15, 2020.

[10] K. Hormann and G. Greiner, ‘MIPS: An efficient global parametrization method’,  
Curve and Surface Design: Saint-Malo 1999, pp. 153–162, 2000.