



Research Paper Interpretation and Interconnection

Charles Joseph, Tianren Zhang, Jiaxing He

April 26, 2024

Contents

1	Background	2
2	Motivations	2
3	Data Collection & Preparation	3
4	Model	3
5	Application Demo	5
6	Visualization Design	7
7	Appendix	8
7.1	API Appendix	8
7.2	Code Appendix	9
8	Paper Contributions	9
9	Reference	9

1 Background

In the expansive realm of academic research, scholars often find themselves navigating a vast sea of literature to uncover relevant studies and insights. The sheer volume of available research papers presents a daunting challenge, particularly when delving into multifaceted and complex topics that require a nuanced understanding. As highlighted by recent studies, the task of identifying pertinent literature is not merely about access but also about efficiently sifting through and synthesizing extensive databases to find the most relevant information.

The importance of efficient literature search mechanisms is underscored by the increasing specialization within academic disciplines, which further complicates the process of literature review. Despite significant advancements in digital library technologies and search engines, the effective utilization of these resources often demands considerable time, expertise, and a methodical approach. Researchers are frequently required to engage in iterative search processes, refining their strategies as they go, which can be both time-consuming and labor-intensive.

2 Motivations

The motivation behind this project was Homework 3. Coming from a background with surface level machine learning and computer vision experience made it hard to properly understand what the paper was talking about. In addition, when trying to find similar research papers, it was difficult to find something that was directly related to the topic that we were presenting. Seeing this as a problem, we decided to build a solution. We wanted to build something that provided utility and can be used by anyone. Our hope is that aspiring researchers and students could find use in our application, and expedite their research and learning process.

We implemented a full-stack web application based on React and Python. We design a user-friendly front end, with a flask backend. It could take the title of the paper, and the question regarding the paper, and return relative information for the paper, give recommendations, and answer the questions. We use Semantic API and CrossRef API to get information for the paper, and provided precise recommendations. With the power of ChatGPT 4, we could answer the questions regarding the paper which give a great overall view of the paper and the industry.

3 Data Collection & Preparation

We designed this application to allow users to type the name of the paper that they are interested in into the search bar of our application and it will retrieve all of the authors of that paper, the date the paper was published, other publications by the first author, and recommended research paper related to the paper that the user searched up. We took all of our data from the Semantic Scholar API. Semantic Scholar is an application that has access to thousands of research papers and we made API requests to retrieve all of this information. We also wanted to allow our application to let users dive deeper into a particular topic or paper. Not only did we receive direct recommendations from the papers they searched, but we digged even further by exploring recommendations from those initial suggestions. In essence, our application operates on three layers: the first layer comprises the paper initially searched, followed by all related recommended papers, and finally, recommendations stemming from those secondary suggestions. API requests were made to get the initially recommended papers and then a second request was made for each of the related recommended paper to get our secondary set of recommendations.

For our interface, we wanted a visually appealing front end that is easy and intuitive to use. We did our front end in React and our back-end was done in Python. We had to establish API endpoints between the front-end and the back-end so they can properly communicate with each other. The challenge in doing this was that our back-end was in Python and our front-end was written in JavaScript, so we had to incorporate cross code functionality. This was something that was new to us and took some time to figure out. In addition, all of our graphs were made using Plotly. Plotly allowed us to build interactive graphs that enhanced our visualizations.

4 Model

Our intrinsic model of application is based on the paper[1].

Pipeline of model

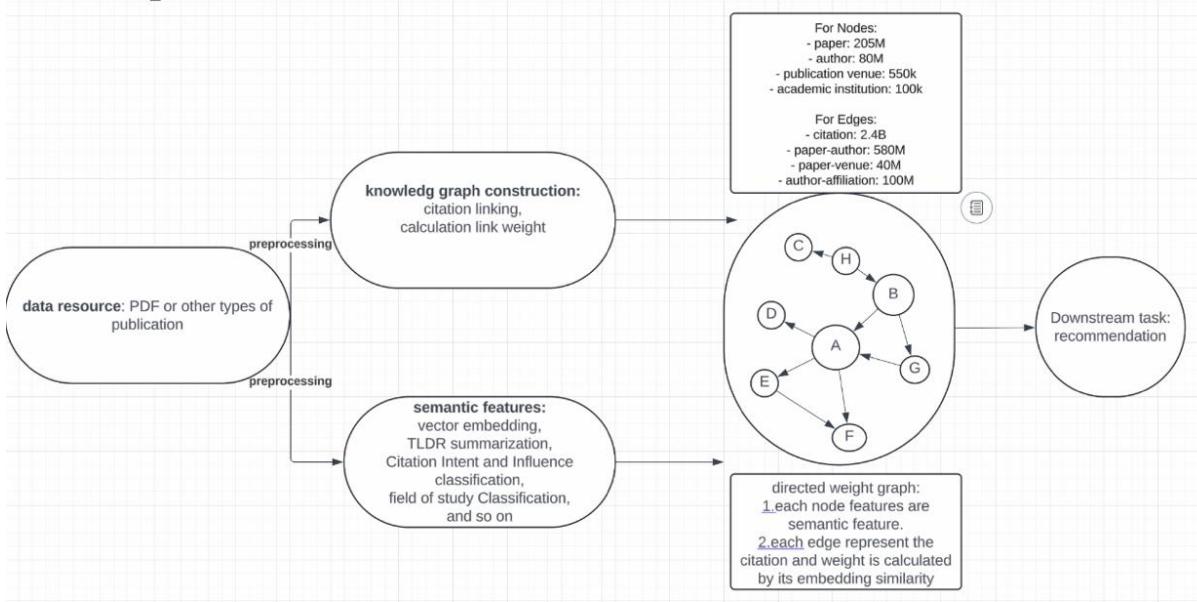


Figure 1: Model pipeline

To begin, the PDF or other publication data types are acquired, followed by pre-processing into a knowledge graph. This graph is constructed based on citations within the documents to calculate the similarity between related papers. This similarity serves as the link weight within the graph. Additionally, semantic features are extracted through various models [1] for embedding or classification purposes. Once pre-processing is complete, a directed graph is constructed to represent publication relationships. This graph can subsequently be utilized for paper recommendation purposes.

5 Application Demo

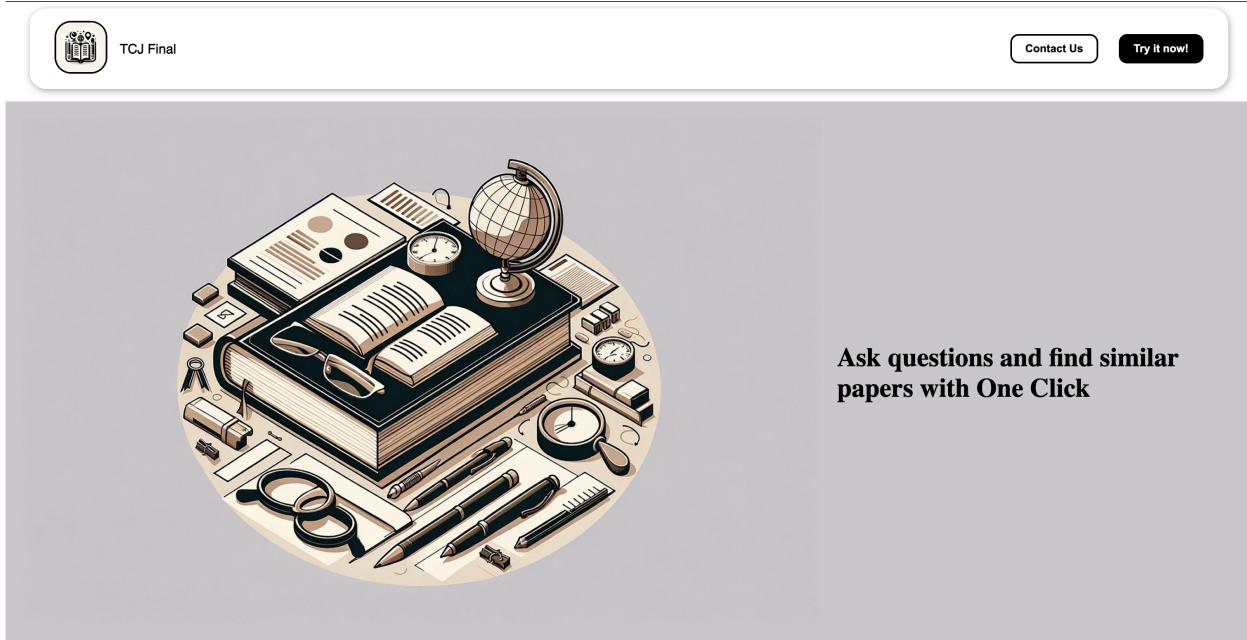


Figure 2: Application Landing Page



Figure 3: User Prompt

GlueGen: Plug and Play Multi-modal Encoders for X-to-image Generation

DOI: 10.1109/iccv51070.2023.02110

Authors:

- Can Qin - first - Northeastern University
- Ning Yu - additional - Salesforce AI Research
- Chen Xing - additional - Salesforce AI Research
- Shuai Zhang - additional - Salesforce AI Research
- Zeyang Guo - additional - Salesforce AI Research
- Stefano Ermon - additional - Stanford University
- Yun Fu - additional - Northeastern University
- Caiming Xiong - additional - Salesforce AI Research
- Ran Xu - additional - Salesforce AI Research

Published On: 2023-10-1

Conference: 2023 IEEE/CVF International Conference on Computer Vision (ICCV) in Paris, France, 2023-10-1 to 2023-10-6

Access Paper License References: 75

Part of References:

- Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation.
- Cogview: Matching text-to-image generation via cross-transformers.
- Beyond english-centered multilingual machine translation.
- Dominant-adversarial training of neural networks.
- Generative adversarial nets.
- Gans trained by a two-time-scale update rule converge to a local nash equilibrium.
- Denoising diffusion probabilistic models.
- Conditional adversarial domain adaptation.
- Rethinking network design and local geometry in point cloud: A simple residual mlp framework.
- Pytorch: An imperative style, high-performance deep learning library.
- Pointdan: A multi-scale 3d domain adaption network for point cloud representation.
- Learning transferable visual models from natural language supervision.
- Exploring the limits of transfer learning with a unified text-to-text transformer.
- Zero-shot text-to-image generation.
- Generative adversarial text-to-image synthesis.
- D2c: Diffusion denoising module for fast shot conditional generation

Figure 4: Paper Information

Can Qin

Momentum is All You Need for Data-Driven Adaptive Optimization

Adaptive gradient methods, e.g., ADAM, have achieved tremendous success in data-driven machine learning, especially deep learning. Employing adaptive learning rates according to the gradients, such methods are able to attain rapid training of modern deep neural networks. Nevertheless, they are observed to suffer from compromised generalization capacity compared with stochastic gradient descent (SGD) and tend to be trapped in local minima at an early stage during the training process. Intriguingly, we discover that the issue can be resolved by substituting the gradient in the second raw moment estimate term with its exponential moving average version in ADAM. The intuition is that the gradient with momentum contains more accurate directional information, and therefore its second-moment estimation is a more preferable option for learning rate scaling than that of the raw gradient. Thereby we propose ADAMS^Δ(3)^ΔS as a new optimizer reaching the goal of training quickly while generalizing much better. Extensive experiments on a variety of tasks and models demonstrate that ADAMS^Δ(3)^ΔS exhibits state-of-the-art performance and superior training stability consistently. Considering the simplicity and effectiveness of ADAMS^Δ(3)^ΔS, we believe it has the potential to become a new standard method in deep learning. Code is provided at <https://github.com/wyzjack/AdaMS>.

Why is the State of Neural Network Pruning so Confusing? On the Fairness, Comparison Setup, and Trainability in Network Pruning

The state of neural network pruning has been noticed to be unclear and even confusing for a while, largely due to a lack of standardized benchmarks and metrics^[3]. To standardize benchmarks, first, we need to answer: what kind of comparison setup is considered fair? This basic yet crucial question has barely been clarified in the community, unfortunately. Meanwhile, we observe several papers have used (severely) sub-optimal hyper-parameters in pruning experiments, while the reason behind them is also elusive. These sub-optimal hyper-parameters further exacerbate the distorted benchmarks, rendering the state of neural network pruning even more obscure. Two mysteries in pruning represent such a confusing status: the performance-boosting effect of a larger finetuning learning rate, and the no-value argument of inheriting pretrained weights in filter pruning. In this work, we attempt to explain the confusing state of network pruning by demystifying the two mysteries. Specifically, (1) we first clarify the fairness principle in pruning experiments and summarize the widely-used comparison setups; (2) then we unveil the two pruning mysteries and point out the central role of network trainability, which has not been well recognized so far; (3) finally, we conclude the paper and give some concrete suggestions regarding how to calibrate the pruning benchmarks in the future. Code: <https://github.com/mingsun-tse/why-the-state-of-pruning-so-confusing>.

HIVE: Harnessing Human Feedback for Instructional Visual Editing

Incorporating human feedback has been shown to be crucial to align text generated by large language models to human preferences. We hypothesize that state-of-the-art instructional image editing models, where outputs are generated based on an input image and an editing instruction, could similarly benefit from human feedback, as their outputs may not adhere to the correct instructions and preferences of users. In this paper, we present a novel framework to harness human feedback for instructional visual editing (HIVE). Specifically, we collect human feedback on the edited images and learn a reward function to capture the underlying user preferences. We then introduce scalable diffusion model fine-tuning methods that can incorporate human preferences based on the estimated reward. Besides, to mitigate the bias brought by the limitation of data, we contribute a new 1M training dataset, a 3.6K reward dataset for rewards learning, and a 1K evaluation dataset to boost the performance of instructional image editing. We conduct extensive empirical experiments quantitatively and qualitatively, showing that HIVE is favored over previous state-of-the-art instructional image editing approaches by a large margin.

Global Aligned Structured Sparsity Learning for Efficient Image Super-Resolution

Efficient image super-resolution (SR) has witnessed rapid progress thanks to novel lightweight architectures or model compression techniques (e.g., neural architecture search and knowledge distillation). Nevertheless, these methods consume

Figure 5: Other Papers Written by First Author

Recommended Papers:

LCM-Lookahead for Encoder-based Text-to-Image Personalization

Authors: Rinon Gal, Or Lichten, Elad Richardson, Or Patashnik, Amit H. Bermano, Gal Chechik, Daniel Cohen-Or

Recent advancements in diffusion models have introduced fast sampling methods that can effectively produce high-quality images in just one or a few denoising steps. Interestingly, when these are distilled from existing diffusion models, they often maintain alignment with the original model, retaining similar outputs for similar prompts and seeds. These properties present opportunities to leverage fast sampling methods as a shortcut-mechanism, using them to create a preview of denoised outputs through which we can backpropagate image-space losses. In this work, we exploit the potential of using such shortcut-mechanisms to guide the personalization of text-to-image models to specific facial identities. We focus on encoder-based personalization approaches, and demonstrate that by tuning them with a lookahead identity loss, we can achieve higher identity fidelity, without sacrificing layout diversity or prompt alignment. We further explore the use of attention sharing mechanisms and consistent data generation for the task of personalization, and find that encoder training can benefit from both.

Reference Can Play: An Alternative Approach to Conditional Generation via Model Inversion

Authors: Xuantong Liu, Tianyang Hu, Wenjia Wang, Kenji Kawaguchi, Yuan Yao

As a dominant force in text-to-image generation tasks, Diffusion Probabilistic Models (DPMs) face a critical challenge in controllability, struggling to adhere strictly to complex, multi-faceted instructions. In this work, we aim to address this alignment challenge for conditional generation tasks. First, we provide an alternative view of state-of-the-art DPMs as a way of inverting advanced Vision-Language Models (VLMs). With this formulation, we naturally propose a training-free approach that bypasses the conventional sampling process associated with DPMs. By directly optimizing images with the supervision of discriminative VLMs, the proposed method can potentially achieve a better text-image alignment. As proof of concept, we demonstrate the pipeline with the pre-trained BLIP-2 model and identify several key designs for improved image generation. To further enhance the image fidelity, a Score Distillation Sampling module of Stable Diffusion is incorporated. By carefully balancing the two components during optimization, our method can produce high-quality images with near state-of-the-art performance on T2I-Compcbench.

MaxFusion: Plug & Play Multi-Modal Generation in Text-to-Image Diffusion Models

Authors: Nithin Gopalakrishnan Nair, Jeya Maria Jose Valanarasu, Vishal M. Patel

Large diffusion-based Text-to-Image (T2I) models have shown impressive generative powers for text-to-image generation as well as spatially conditioned image generation. For most applications, we can train the model end-to-end with paired data to obtain photorealistic generation quality. However, to add an additional task, one often needs to retrain the model from scratch using paired data across all modalities to retain good generation performance. In this paper, we tackle this issue and propose a novel strategy to scale a generative model across new tasks with minimal compute. During our experiments, we discovered that the variance maps of intermediate feature maps of diffusion models capture the intensity of conditioning. Utilizing this prior information, we propose MaxFusion, an efficient strategy to scale up text-to-image generation models to accommodate new modality conditions. Specifically, we combine aligned features of multiple models, hence bringing a compositional effect. Our fusion strategy can be integrated into off-the-shelf models to enhance their generative prowess.

StyleGAN-Fusion: Diffusion Guided Domain Adaptation of Image Generators

Authors: Kunpeng Song, Lirong Han, Bingchen Liu, Dimitris Metaxas, A. Elgammal

Figure 6: List of Recommended Papers

Question:

Why GlueGen? Could you tell me more about this paper as much as possible?

Answer:

GlueGen is a novel architecture designed for plug-and-play multi-modal encoders for X-to-image generation tasks. It is a framework that allows for the seamless integration of different modalities (such as text, audio, or other data types) into a single encoder-decoder network for image generation. GlueGen is built upon the principles of modularity and flexibility, allowing for easy customization and extension to adapt to various data types and tasks.
One key advantage of GlueGen is its ability to handle multiple modalities simultaneously, enabling it to generate images from a diverse range of input data sources. This makes it particularly useful for tasks that involve combining information from different modalities, such as generating images from text descriptions or audio inputs.

Figure 7: ChatGPT Output of a Question Asked

6 Visualization Design

Recommendation Network

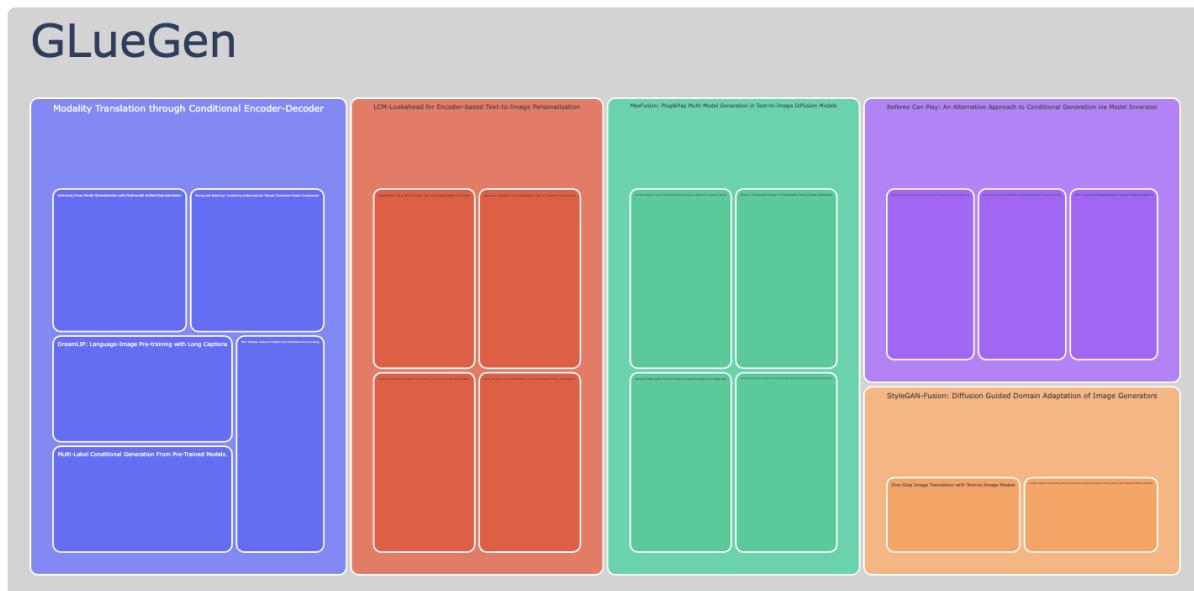


Figure 8: Treemap Visualization

Citations by Year

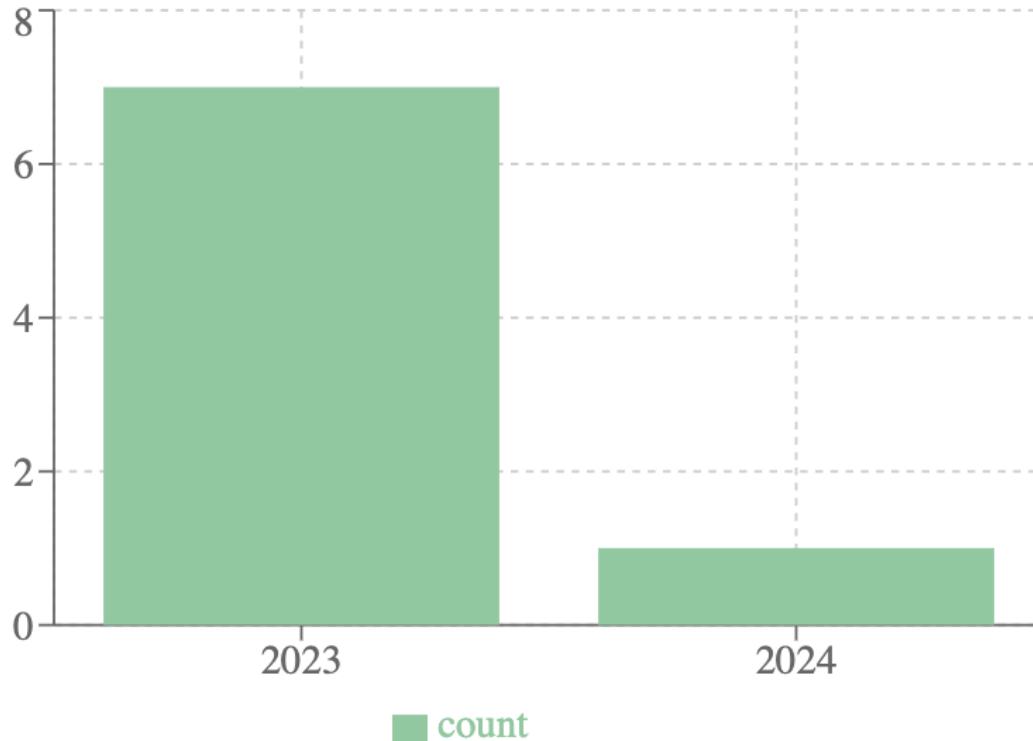


Figure 9: Citation Graph Visualization

7 Appendix

7.1 API Appendix

Semantic API Providing a reliable source of scholarly data for developers. The Semantic Scholar REST API allows you to find and explore scientific publication data about authors, papers, citations, venues, and more. The API is organized into the following services:

Academic Graph: Provides data about authors, papers, citations, venues, SPECTER2 embeddings, and more that allows you to link directly to the corresponding page on semanticscholar.org for more information.

Recommendations: Provides recommended papers similar to a given paper.

Datasets: Provides downloadable links of datasets in the academic graph.

Conference Peer Review: Provides utilities to help conference organizers with the problem of assigning reviewers to conference submissions. Includes detection of con-

flict of interest, based on co-author relationships, and computation of a matching score between a reviewer and a submission's topic, based on the reviewer's publication history.

CrossRef API offer a wide array of services to ensure that scholarly research metadata is registered, linked, and distributed. Crossref makes research objects easy to find, cite, link, assess, and reuse. It's a not-for-profit membership organization that exists to make scholarly communications better.

ChatGPT 4 is one of the most powerful AI model in the world. We can easily get instant answers, find creative inspiration, and learn something new.

7.2 Code Appendix

For the relative packages and source code, check out our public GitHub [front-end](#) and [backend](#) for all of our code. Please download the backend and the front end respectively, run them, and check the result!

8 Paper Contributions

Charles Joseph: Motivations, Data Collection and Preparation

Tianren Zhang: Application Demo

Jiaxing He: Background and Model

9 Reference

- [1] Kinney, Rodney Michael et al. "The Semantic Scholar Open Data Platform." ArXiv abs/2301.10140 (2023): n. pag.