

Pattern Recognition and Computer Vision

Instructor: Yao Yao

Fall 2024



Course Information



- **Course:** 模式识别与计算机视觉
- **Location:** 南雍楼东428
- **Instructor:** 姚遥
- **Instructor Office:** 南雍楼东509
- **Instructor Office hour:** Monday 2pm - 4pm
- **Instructor Email:** yaoyao@nju.edu.cn
- **Instructor Webpage:** <https://yoyo000.github.io/>



Yao Yao

- **Nanjing University**, Associate Professor *2023.06 - now*
- **Apple**, Senior Researcher *2020.04 - 2023.05*
- **Altizure**, Founding Member *2015.07 - 2020.04*

Research Interests

- 3D Computer Vision
- **Reconstruction**: 3D Reconstruction and Differentiable Rendering
- **Generation**: Generative Models for 3D Content Creation

Today's Topics



- Course Objective
- What is Pattern Recognition?
- Course Overview
- Introduction to Image Data



Course Objective



- **Grasp the basic knowledge** of pattern recognition and computer vision, including related problems and applications
- **Dive deep** into one or two recent topics related to PRCV
- With a focus on **computer vision** and **natural language processing**



Pre-requisite



- Machine Learning
- Image processing
- C++
- Python
- PyTorch



What is Pattern Recognition?

“The field of pattern recognition is concerned with the automatic discovery of **regularities** in **data** through the use of computer **algorithms** and with the use of these regularities to **take actions** such as classifying the data into different categories.”

Text

Image

Video

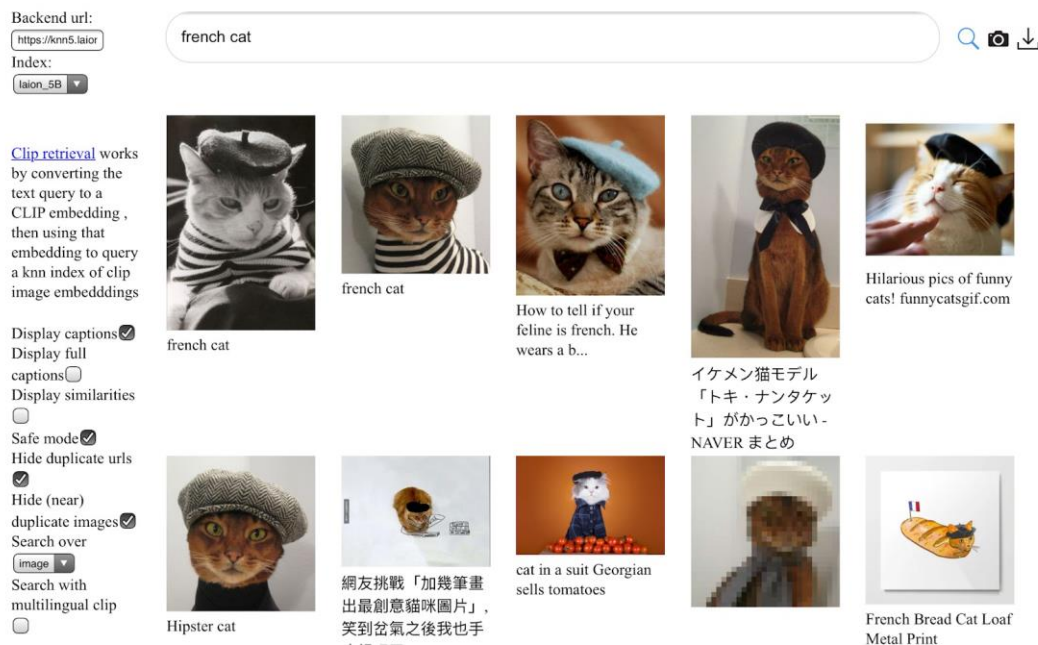
3D

...



- **How is the image captured?**
- **Where is the image captured?**
- **What is in the image?**

Data: Image



LAION 5B dataset

- What is the dataset about?
- Given this dataset, can you generate one new cat in the same pattern?



WikiPedia

-

Input

Andrew is free from 11 am to 3 pm, Joanne is free from noon to 2 pm and then 3:30 pm to 5 pm. Hannah is available at noon for half an hour, and then 4 pm to 6 pm. What are some options for start times for a 30 minute meeting for Andrew, Hannah, and Joanne?

Output

Andrew: 11 am - 3 pm

Joanne: 12 pm - 2 pm, 3:30 pm - 5 pm

Hannah: 12 pm - 12:30 pm, 4 pm - 6 pm

Common availability for a 30-minute meeting: 12 pm - 12:30 pm

ChatGPT

- Can you response to my words?
 - Predict the next word

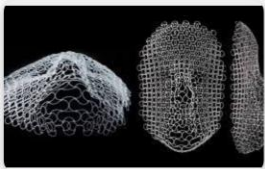


- 

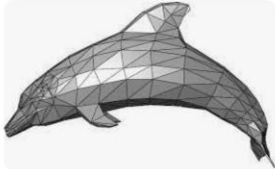


- Who is speaking?
- Can you translate the speech into text?
- Can you reduce the background noise?
- Can you tune my voice?

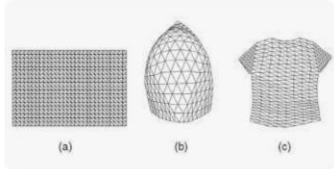
Data: Others



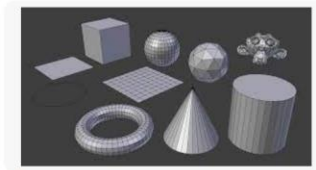
IMechE
This flat mesh automatically transform...



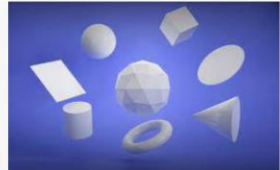
Wikipedia
Polygon mesh - Wikipedia



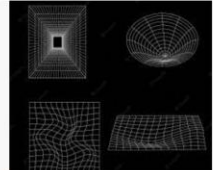
ResearchGate
Specifying the 3D shape of the rectangular mesh a...



Blender Documentation
Primitives - Blender Manual

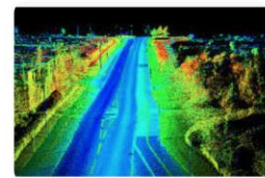


3D Studio
Introduction to 3D Polygon Mesh - 3D Studio

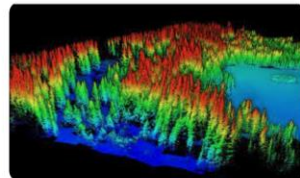


Freepik
Premium Vector | 3d wireframe sha...

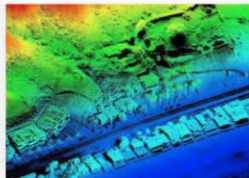
3D Shape



Analytics
Applications & Challenges with 3D P...



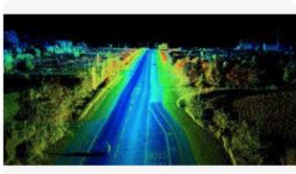
Geodetics
How LiDAR Point Cloud Data Can Affect Ef...



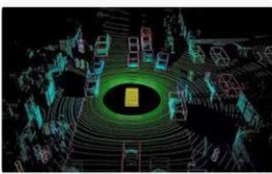
GIM International
Point Clouds: Photogrammetry or Lidar? | ...



Pix4D
LiDAR: what it is, and how it is useful fo...



Geospatial World
LizardTech awarded US patent for LiDAR p...

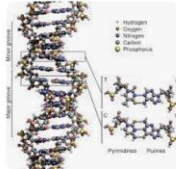


NVIDIA Developer
Webinar: Learn How NVIDIA DriveWork...

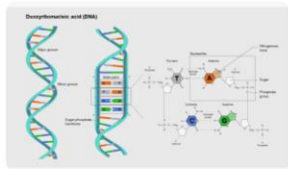
Lidar Points



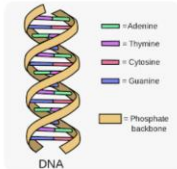
MedlinePlus
What is DNA?: MedlinePlu...



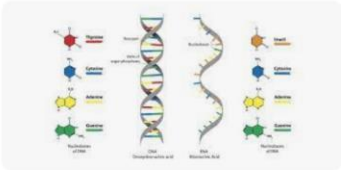
Wikipedia
DNA - Wikipedia



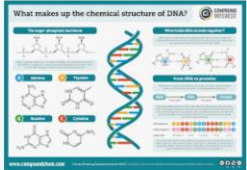
National Human Genome Research Institute
Deoxyribonucleic Acid (DNA)



ASHG
Building Blocks of the Gen...



Technology Networks
DNA vs. RNA - 5 Key Differences and Comparison | T...

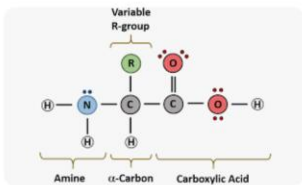


Compound Interest: Chemistry Infographics
Compound Interest: What makes up the Chem...

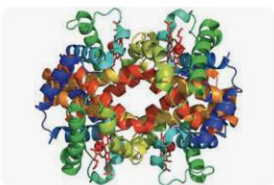


YourGenome
What is DNA? - YourGenome

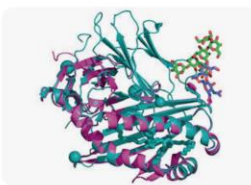
DNA



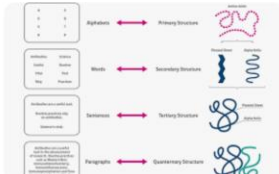
Western Oregon University
Chapter 2: Protein Structure - Chemistry



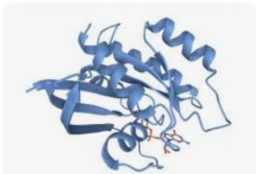
Institute for Advanced Study
Using Protein Sequences to Predict Stru...



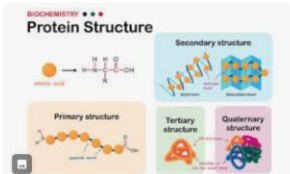
PNAS
Researchers turn to deep learning to ...



Proteintech
The Complexity of Proteins | Proteintech ...



EMBL-EBI
Levels of protein structure - tertiary |...



iStock
Biochemistry Diagram Show Levels Of Prot...

Protein



Feature

Image features: corners, edges, deep features

Text embeddings/tokenization

Point cloud descriptors/3D latent

...

Distribution

Distribution in 1D space

Distribution in N-D space

Distribution in high-dimensional N^2 -D image space

Distribution in latent space

...

How to Recognize Patterns



- Data collection
- Feature extraction / Data washing
- Modeling (+ Training)



Computer Vision

- Feature detector and descriptor
- Semantic segmentation
- Image Classification
- Image generation
- ...

Natural Language Processing

- Tokenization
- Grammatical Error Correction
- Relation extraction
- Text Classification
- ...

Others ...

模式识别 vs. 机器学习

- ✓ 机器学习在模式识别中有非常重要的作用
- ✓ 但是，模式识别具有更多的“系统”性
 - 数据获取
 - 提取特征
 - ...
- ✓ 主要的区别是：“数据” vs. “特征”
 - 但是，在深度学习中，这个区别不那么明显了

模式识别 vs. 计算机视觉

- ✓ 模式识别与计算机视觉(**computer vision**)的研究和应用有**非常多的重合**
 - 识别**recognition**是计算机视觉中最重要的问题之一
 - 模式识别中很大部分输入是图像
- ✓ 模式识别包含很多视觉以外的问题
 - 音频、雷达、文本、...
- ✓ 计算机视觉包括很多识别以外的问题
 - 如，超分辨率**super-resolution**
 - 三维重建**3D reconstruction**

模式识别与多媒体(multimedia)

- ✓ 都可能牵涉多种媒体
- ✓ 但多媒体更具有“系统”性
 - 多媒体不特别强调单个模块的性能
 - 更注重整个大系统的成功
 - 比模式识别更强调多种媒体之间的配合
 - 就算每个模块都采用了已有的技术，但是一个科学利用现有技术和多种媒体的系统仍然可以是很大的成功

Pattern Recognition vs. Others



- Overlapped with ML, CV, NLP
- Just focus on problems
- Do not set boundaries



Introduction:

- **W1** - Course Introduction

Low-level Vision:

- **W2** - Image features: image filters and edges
- **W3** - Image features: image features and matching

Mid-level Vision:

- **W4** - Camera calibration: camera model, structure from motion
- **W5** - Geometry reconstruction: stereo and multi-view stereo

High-level Vision:

- **W6** - Neural Network: layers, back propagation, computational graph
- **W7** - Backbone Networks: CNN, RNN; AlexNet - VGGNet - ResNet - MobileNet
- **W8** - Classification and Recognition: Methods

Natural Language Processing:

- **W9** - Tasks and Methods before LLM
- **W10** - Large Language Model: Transformer, BERT, and GPT

Advanced Topics:

- **W1x** - Aligning Vision and Language: Contrastive Language-Image Pretraining (CLIP)
- **W1x** - Bringing Vision and Graphics: Differentiable Rendering
- **W1x** - Modeling Distribution of Image: VAE, GAN, and 2D diffusion
- **W1x** - Modeling Distribution of Video: AR, Diffusion (UNet), and Diffusion Transformer (DiT)
- **W1x** - Modeling Distribution of 3D: lifting 2D diffusion to 3D and direct 3D diffusion

Exam:

- **W15** - Final Exam or Presentation

Lecture + Paper Reading

Recommended Courses:

- **Computer Vision**, Noah Snavely, Cornell University:

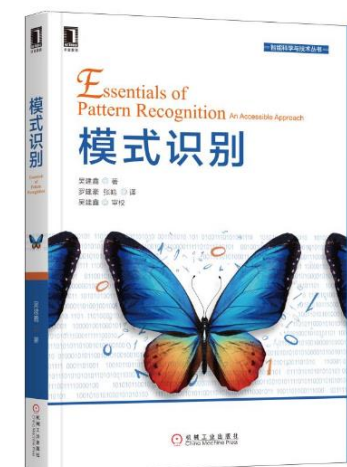
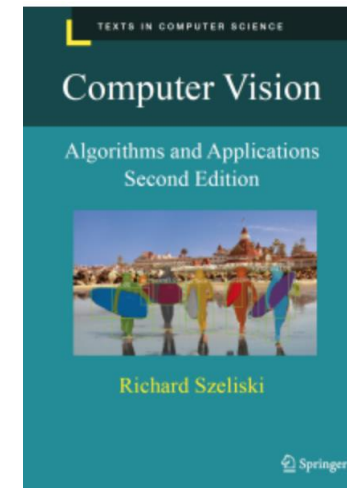
<https://www.cs.cornell.edu/courses/cs5670/2023sp/lectures/lectures.html>

- **Pattern Recognition**, Jianxin Wu, Nanjing University:

https://cs.nju.edu.cn/wujx/teaching_PR.html

Recommended Books:

- **Computer Vision: Algorithms and Applications, 2nd ed.** Richard Szeliski.
2022: <https://szeliski.org/Book/>
- **Essentials of Pattern Recognition: An Accessible Approach.**
Jianxin Wu. 2020: https://cs.nju.edu.cn/wujx/PR_Book_CN/PR_Book_CN.htm



Recommended Papers: 1-2 papers per week

- **W2 - Canny Edge Detector:** A Computational Approach to Edge Detection. John Canny. TPAMI 1986.
- **W2 - ControlNet:** Adding Conditional Control to Text-to-Image Diffusion Models. Lvmin Zhang. ICCV 2023.

- **Attendance:** $1\% \times 15 = 15\%$
- **Mid-term Project:** 35%
- **Final Project:** 50%

Mid-term Project: choose 1 below, **solo** work

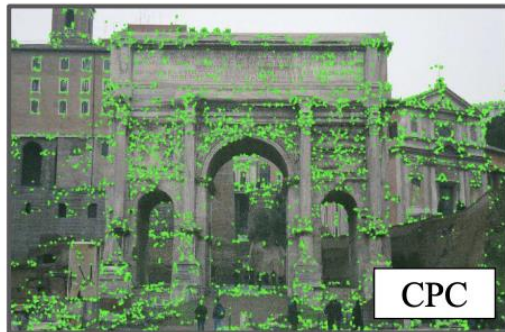
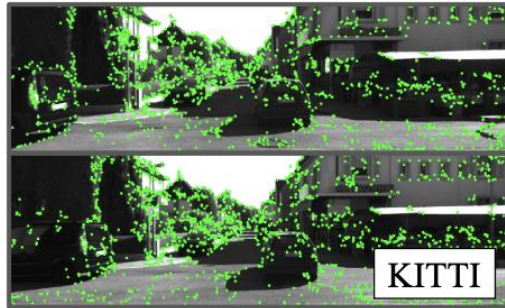
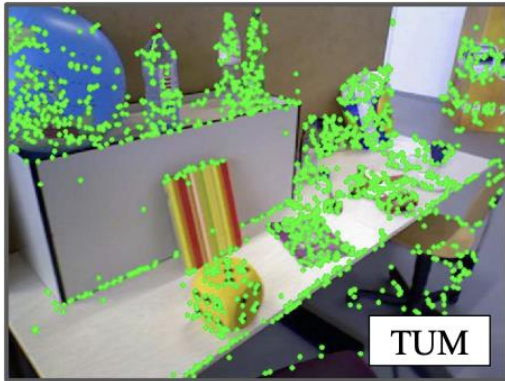
- Image Features: two-view matching
- Image Features: image-stitching for panoramas
- Point Cloud Features: point cloud registration

Final Project: choose 1 below, **pair** work

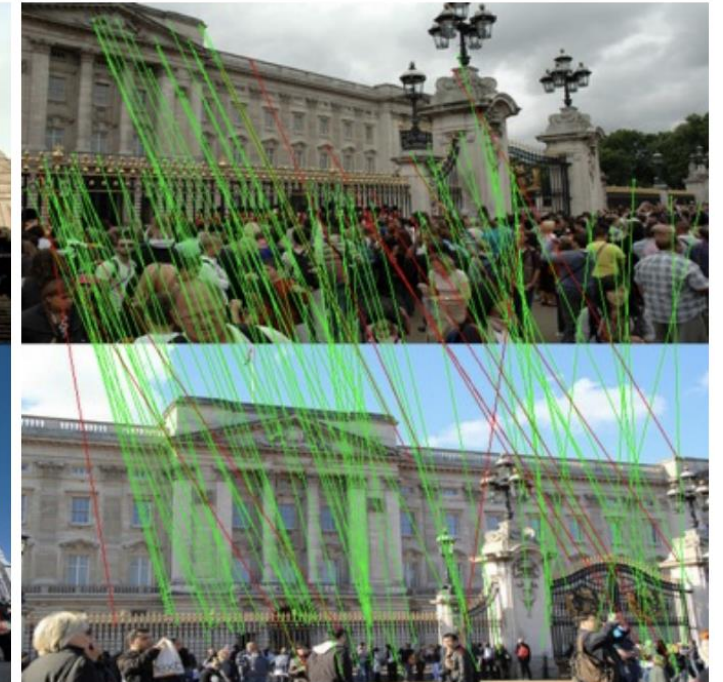
- Large Language Model: domain GPT
- Large Vision Model: domain diffusion
- Others... (discuss with Yao)

Open-source codes are allowed,
but make sure your delta is clear

Projects: Two-view Matching



ASLFeat. Luo et al, CVPR 2020



OANet. Zhang et al, CVPR 2019



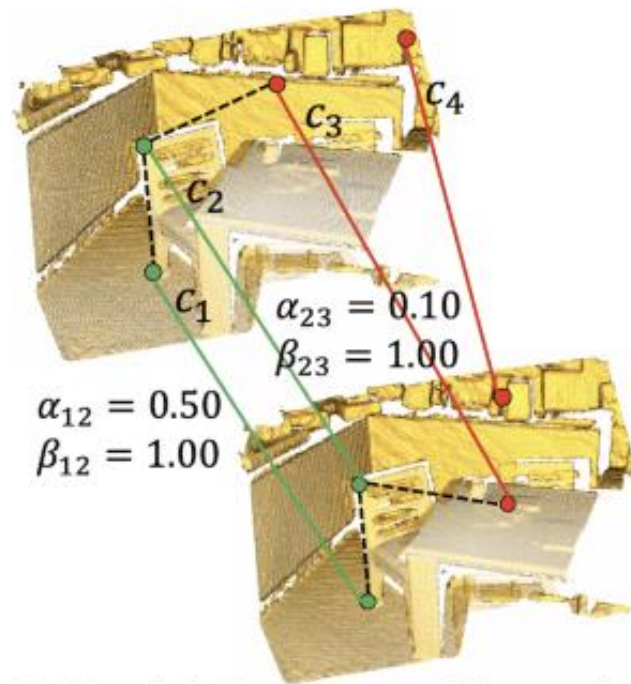
Projects: Image-stitching for Panorama



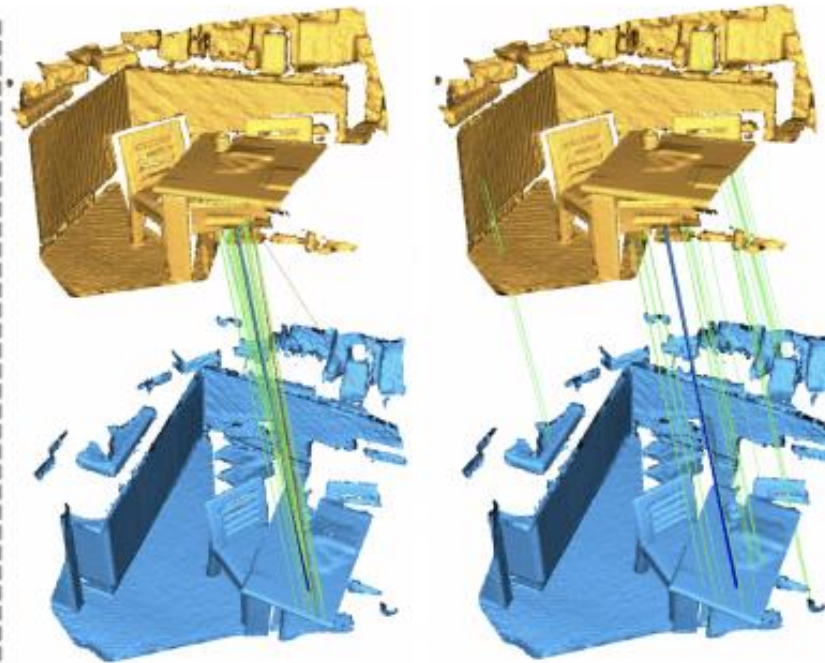
Mountains in Hong Kong (captured by Yao)



Projects: Point-cloud Registration



(a) Spatial Consistency Illustration



(b) Spatial kNN and Feature-space kNN

PointDSC [Bai et al, CVPR 2020]



Dataset Collection and Finetuning a Foundation Model

Projects: Domain GPT



GPT-4

Input

Andrew is free from 11 am to 3 pm, Joanne is free from noon to 2 pm and then 3:30 pm to 5 pm. Hannah is available at noon for half an hour, and then 4 pm to 6 pm. What are some options for start times for a 30 minute meeting for Andrew, Hannah, and Joanne?

Output

Andrew: 11 am - 3 pm

Joanne: 12 pm - 2 pm, 3:30 pm - 5 pm

Hannah: 12 pm - 12:30 pm, 4 pm - 6 pm

Common availability for a 30-minute meeting: 12 pm - 12:30 pm

ChatGPT

Could you tell me what are the key contributions of MVSNet?

Chat with any PDF

Join millions of students, researchers and professionals to instantly answer questions and understand research with AI



Drop PDF here

[Browse my Computer](#)

[From URL](#)

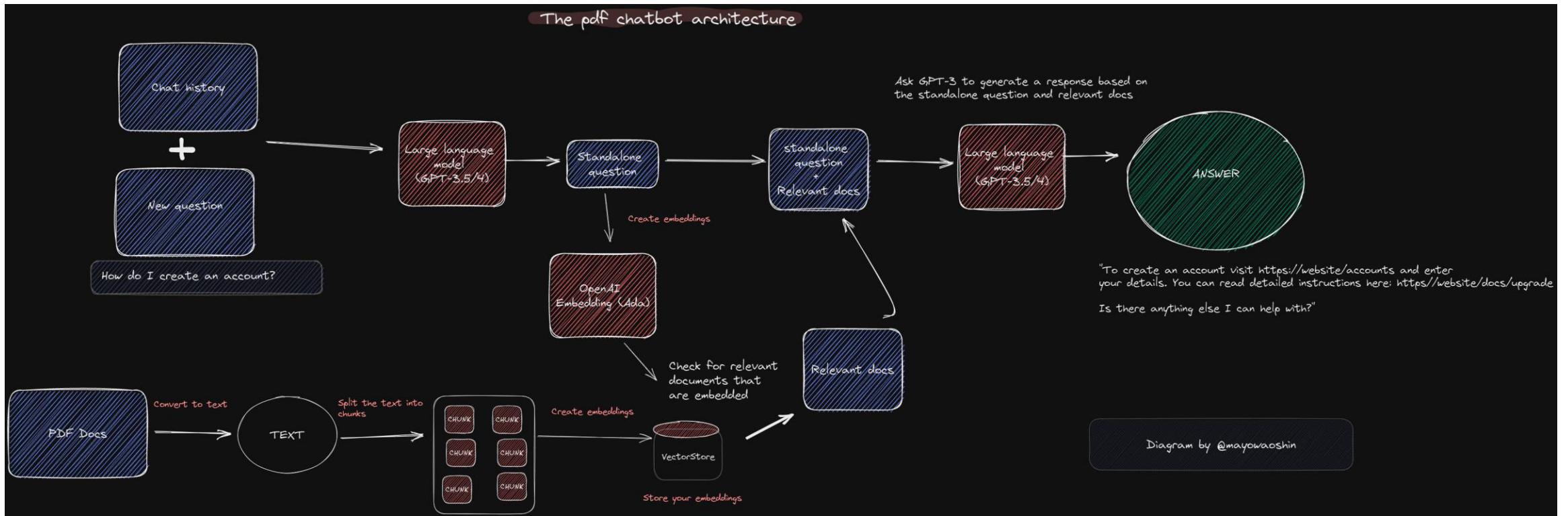
ChatPDF



Projects: Domain GPT



- External Database



PDF-ChatBot: <https://github.com/mayooear/gpt4-pdf-chatbot-langchain>



Projects: Domain GPT



- Finetuning on small LLM



Vicuna (generated by stable diffusion 2.1)

Vicuna 13B

☰ README.md

ChatGLM-6B

🌐 [Blog](#) • 🗨️ [HF Repo](#) • 🐦 [Twitter](#) • 📄 [\[GLM@ACL 22\] \[GitHub\]](#) • 📄 [\[GLM-130B@ICLR 23\] \[GitHub\]](#)

👉 加入我们的 [Slack](#) 和 [WeChat](#)

📍 在 [chatglm.cn](#) 体验更大规模的 ChatGLM 模型。

Read this in [English](#).

介绍

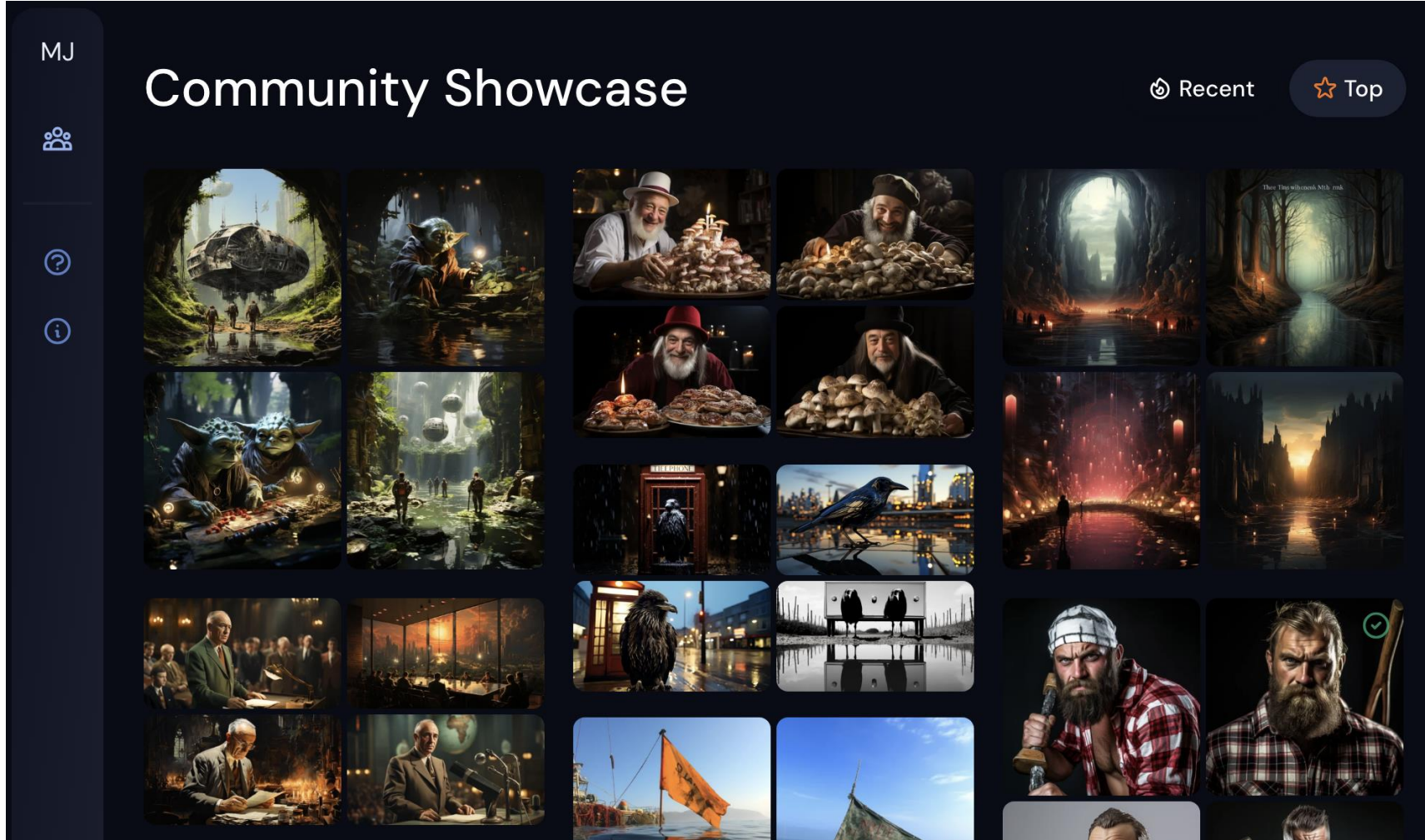
ChatGLM-6B 是一个开源的、支持中英双语的对话语言模型，基于 [General Language Model \(GLM\)](#) 架构，具有 62 亿参数。结合模型量化技术，用户可以在消费级的显卡上进行本地部署（INT4 量化级别下最低只需 6GB 显存）。ChatGLM-6B 使用了和 ChatGPT 相似的技术，针对中文问答和对话进行了优化。经过约 1T 标识符的中英双语训练，辅以监督微调、反馈自助、人类反馈强化学习等技术的加持，62 亿参数的 ChatGLM-6B 已经能生成相当符合人类偏好的回答，更多信息请参考我们的 [博客](#)。欢迎通过 [chatglm.cn](#) 体验更大规模的 ChatGLM 模型。

为了方便下游开发者针对自己的应用场景定制模型，我们同时实现了基于 [P-Tuning v2](#) 的高效参数微调方法 ([使用指南](#))，INT4 量化级别下最低只需 7GB 显存即可启动微调。

ChatGLM-6B 权重对学术研究完全开放，在填写 [问卷](#) 进行登记后亦允许免费商业使用。

ChatGLM 6B





Midjourney: <https://www.midjourney.com/showcase/top/>

Projects: Domain Diffusion



Stable Diffusion XL

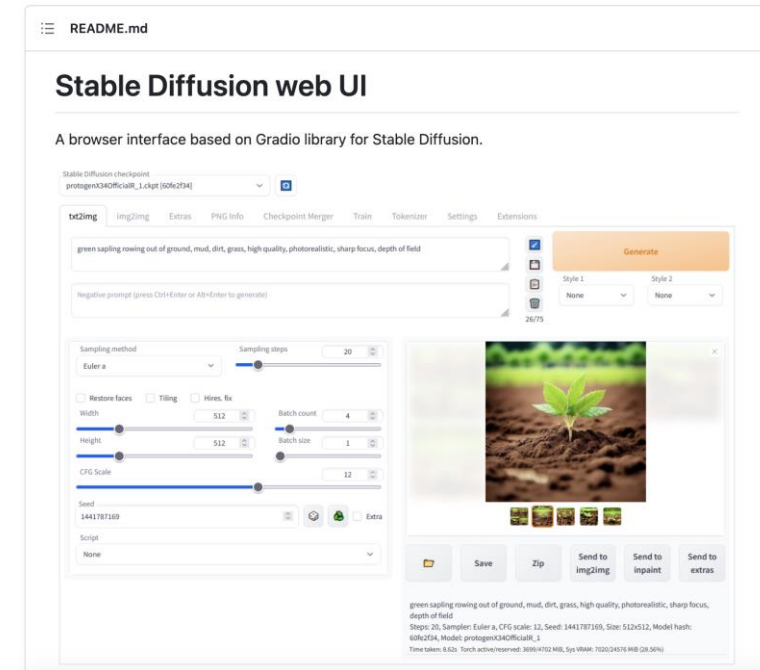
Create and inspire using the world's fastest-growing open source AI platform

With Stable Diffusion XL, you can create descriptive images with shorter prompts and generate words within images. The model is a significant advancement in image generation capabilities, offering enhanced image composition and face generation that results in stunning visuals and realistic aesthetics.

Try on Clipdrop



Stable Diffusion XL



Stable Diffusion web UI



Stable Diffusion with Lora

