CS 8803-VLM

Zsolt Kira

Website: https://faculty.cc.gatech.edu/~zk15/teaching/AY2025_cs8803vlm_fall/index.html
Canvas: https://gatech.instructure.com/courses/458226
Ed: https://edstem.org/us/courses/82921/discussion

WELCOME!

Georgia Tech

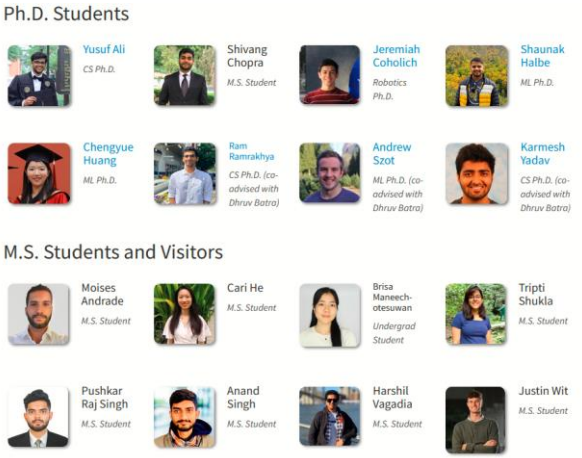
Zsolt Kira

zkira@gatech.edu
Associate Professor
School of Interactive Computing

Georgia Tech Machine Learning

RIPL
Robotics Perception and Learning

- Joined as Assistant Professor in 2018
- **Research Interests:** Intersection of deep learning and robotics, focusing on robustness and decision-making in an open world.
- **Group**
  - **Also a number of M.S. students per year**



- **Teaching:**
  - CS 8803-VLM Vision-Language Foundation Models (Fall), CS 7643 Deep Learning (Spring)

Georgia Tech

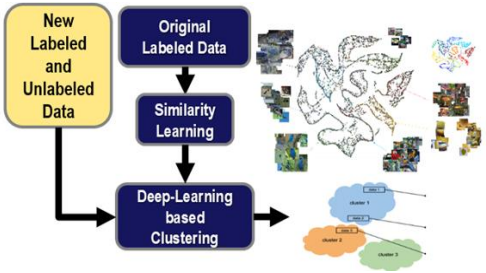
Zsolt Kira

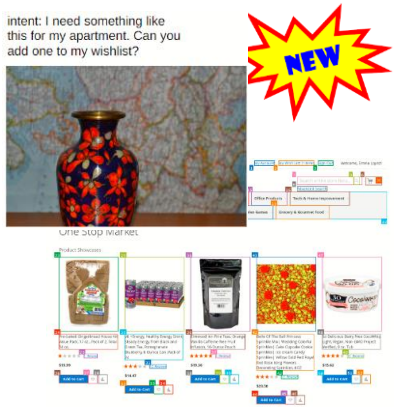
zkira@gatech.edu
Associate Professor
School of Interactive Computing
Georgia Institute of Technology

Georgia Tech Machine Learning

RIPL
Robotics Perception and Learning

Research Interests: Intersection of deep learning and robotics, focusing on robustness and decision-making in an open world





How can perception deal with changing environments and the open world?

Robust Open-World Learning

- Past: Semi and self-supervised, few-shot, continual learning
- Open-world learning and Vision-Language Models
- Robust fine-tuning of VMs/VLMs

How can we use VLMs for Learning, Planning, and Reasoning Agents

Planning & Reasoning

- VLMs for reasoning/planning
- Grounding

How can we scale robotics in DL era?

Scaling up Robotics

- Better simulation w/ NeRFs/3D
- Self-supervised and pre-training
- Combinations with large language and multi-modal models
  - Long-Context Models
- Vision-Language Action Models

Georgia Tech

- What are Vision-Language & Multi-Modal Models?
- Why, and why now?
- Logistics
- Q&A

Agenda for Today

Georgia Tech

What is multimodality?

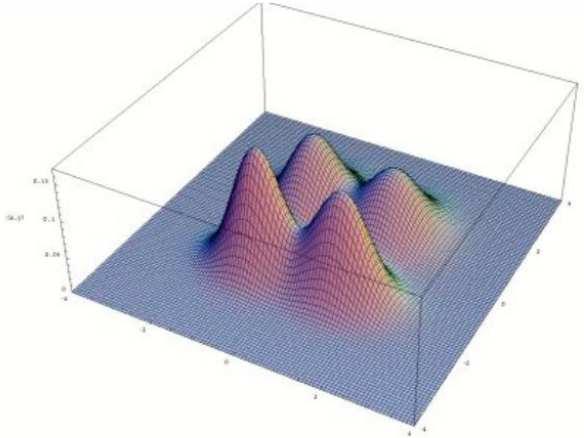
multimodal adjective

mul·ti·mod·al ˌməl-tē-ˈmō-dᵊl -ˌtī-

: having or involving several modes, modalities, or maxima

multimodal distributions

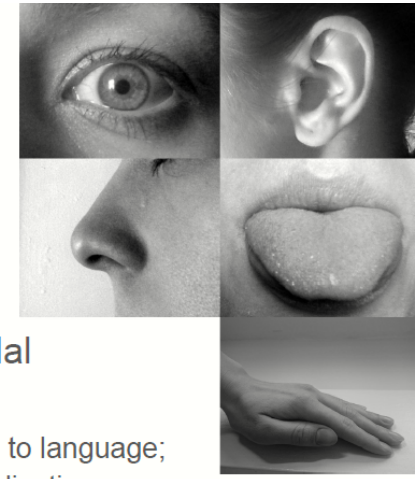
multimodal therapy



In our case, focusing on NLP: text + one or more other *modality* (images, speech, audio, olfaction, others). We'll mostly focus on images as the other modality.

Slide by Douwe Kiela

Why does multimodality matter?



A range of very good reasons:

- Faithfulness: Human experience is multimodal
- Practical: The internet & many applications are multimodal
- Data efficiency and availability:
  - Efficiency: Multimodal data is rich and “high bandwidth” (compared to language; quoting LeCun, “an imperfect, incomplete, and low-bandwidth serialization protocol for the internal data structures we call thoughts”), so better for learning?
  - Scaling: More data is better, and we’re running out of high quality text data.

Multimodality is one of the main frontiers of the new foundation model revolution.

Slide by Douwe Kiela

ChatGPT / Foundation Models / … buzzwords?

Bing's A.I. Chat: 'I Want to Be Alive. 😈'

In a two-hour conversation with our columnist, Microsoft's new chatbot said it would like to be human, had a desire to be destructive and was in love with the person it was chatting with. Here's the transcript.

Give this article 1.6K

https://www.nytimes.com/article/ai-artificial-intelligence-chatbot.html

ARTIFICIAL INTELLIGENCE

ChatGPT is about to revolutionize the economy. We need to decide what that looks like.

New large language models will transform many jobs. Whether they will lead to widespread prosperity or not is up to us.

By David Rotman March 25, 2023

https://www.technologyreview.com/2023/03/25/1070275/chatgpt-revolutionize-economy-decide-what-looks-like/

Exam	GPT-4	GPT-4 (no vision)	GPT-3.5
Uniform Bar Exam (MBE+MEE+MPT)	298 / 400 (~90th)	298 / 400 (~90th)	213 / 400 (~10th)
LSAT	163 (~88th)	161 (~83rd)	149 (~40th)
SAT Evidence-Based Reading & Writing	710 / 800 (~93rd)	710 / 800 (~93rd)	670 / 800 (~87th)
SAT Math	700 / 800 (~89th)	690 / 800 (~89th)	590 / 800 (~70th)
Graduate Record Examination (GRE) Quantitative	163 / 170 (~80th)	157 / 170 (~62nd)	147 / 170 (~25th)
Graduate Record Examination (GRE) Verbal	169 / 170 (~99th)	165 / 170 (~96th)	154 / 170 (~63rd)
Graduate Record Examination (GRE) Writing	4 / 6 (~54th)	4 / 6 (~54th)	4 / 6 (~54th)
USABO Semifinal Exam 2020	87 / 150 (99th - 100th)	87 / 150 (99th - 100th)	43 / 150 (31st - 33rd)
USNCO Local Section Exam 2022	36 / 60	38 / 60	24 / 60
Medical Knowledge Self-Assessment Program	75 %	75 %	53 %
Codeforces Rating	392 (below 5th)	392 (below 5th)	260 (below 5th)
AP Art History	5 (86th - 100th)	5 (86th - 100th)	5 (86th - 100th)
AP Biology	5 (85th - 100th)	5 (85th - 100th)	4 (62nd - 85th)
AP Calculus BC	4 (43rd - 59th)	4 (43rd - 59th)	1 (0th - 7th)
AP Chemistry	4 (71st - 88th)	4 (71st - 88th)	2 (22nd - 46th)
AP English Language and Composition	2 (14th - 44th)	2 (14th - 44th)	2 (14th - 44th)
AP English Literature and Composition	2 (8th - 22nd)	2 (8th - 22nd)	2 (8th - 22nd)
AP Environmental Science	5 (91st - 100th)	5 (91st - 100th)	5 (91st - 100th)
AP Macroeconomics	5 (84th - 100th)	5 (84th - 100th)	2 (33rd - 48th)
AP Microeconomics	5 (82nd - 100th)	4 (60th - 82nd)	4 (60th - 82nd)
AP Physics 2	4 (66th - 84th)	4 (66th - 84th)	3 (30th - 66th)
AP Psychology	5 (83rd - 100th)	5 (83rd - 100th)	5 (83rd - 100th)
AP Statistics	5 (85th - 100th)	5 (85th - 100th)	3 (40th - 63rd)
AP US Government	5 (88th - 100th)	5 (88th - 100th)	4 (77th - 88th)
AP US History	5 (89th - 100th)	4 (74th - 89th)	4 (74th - 89th)
AP World History	4 (65th - 87th)	4 (65th - 87th)	4 (65th - 87th)
AMC 10^3	30 / 150 (6th - 12th)	36 / 150 (10th - 19th)	36 / 150 (10th - 19th)
AMC 12^3	60 / 150 (45th - 66th)	48 / 150 (19th - 40th)	30 / 150 (4th - 8th)
Introductory Sommelier (theory knowledge)	92 %	92 %	80 %
Certified Sommelier (theory knowledge)	86 %	86 %	58 %
Advanced Sommelier (theory knowledge)	77 %	77 %	46 %
Leetcode (easy)	31 / 41	31 / 41	12 / 41
Leetcode (medium)	21 / 80	21 / 80	8 / 80
Leetcode (hard)	3 / 45	3 / 45	0 / 45

Table 1. GPT performance on academic and professional exams. In each case, we simulate the conditions and scoring of the real exam. We report GPT-4's final score graded according to exam-specific rubrics, as well as the percentile of test-takers achieving GPT-4's score.

GPT4 technical report, OpenAI, March 2023

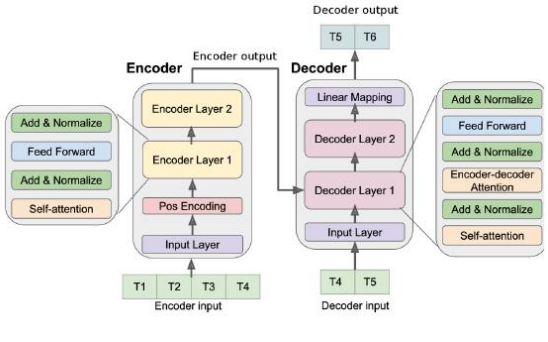
Georgia Tech

Homogenization of Deep Learning

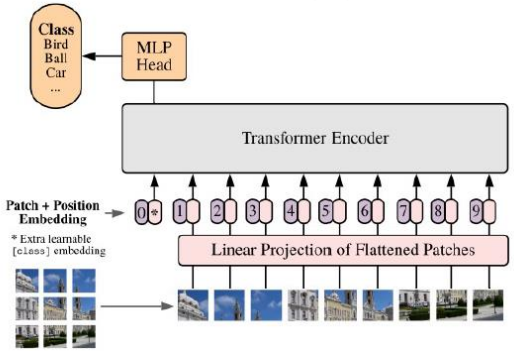
Homogenization is the **consolidation** of methodologies for building machine learning systems across a wide range of applications.

- Enabled by modular, plug-n-play nature of neural networks and training
- Consequence: Multi-modal, unified architectures, unified tasks (next-token prediction)

Example: The Transformer Models (Vaswani et al., 2017)



Transformer Models originally designed for NLP



Almost identical model (Visual Transformers) can be applied to Computer Vision tasks

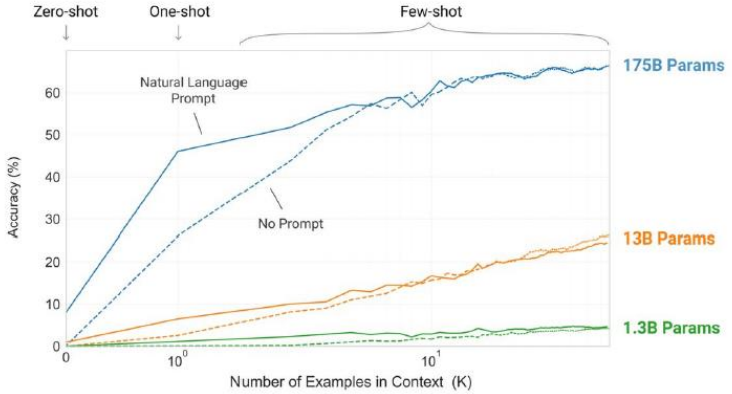
Emergence of new behaviors

Emergence means that the behavior of a system is implicitly induced rather than explicitly constructed.
For Deep Learning, emergence is often induced by larger model & more data.

Example: Compared to GPT-2's 1.5B parameter parameter model, GPT-3's 175-billion model permits "prompting" and "in-context learning", i.e., adapting to a new task simply by describing task.

Example input (prompt):
Ask it to translate French to English

$$\underbrace{\text{maison} \rightarrow \text{house}, \text{chat} \rightarrow \text{cat}, \text{chien} \rightarrow}_{\text{prompt}} \underbrace{\text{dog}}_{\text{completion}} .$$



Bommasani et al., On the Opportunities and Risks of Foundation Models

How/what should we train?

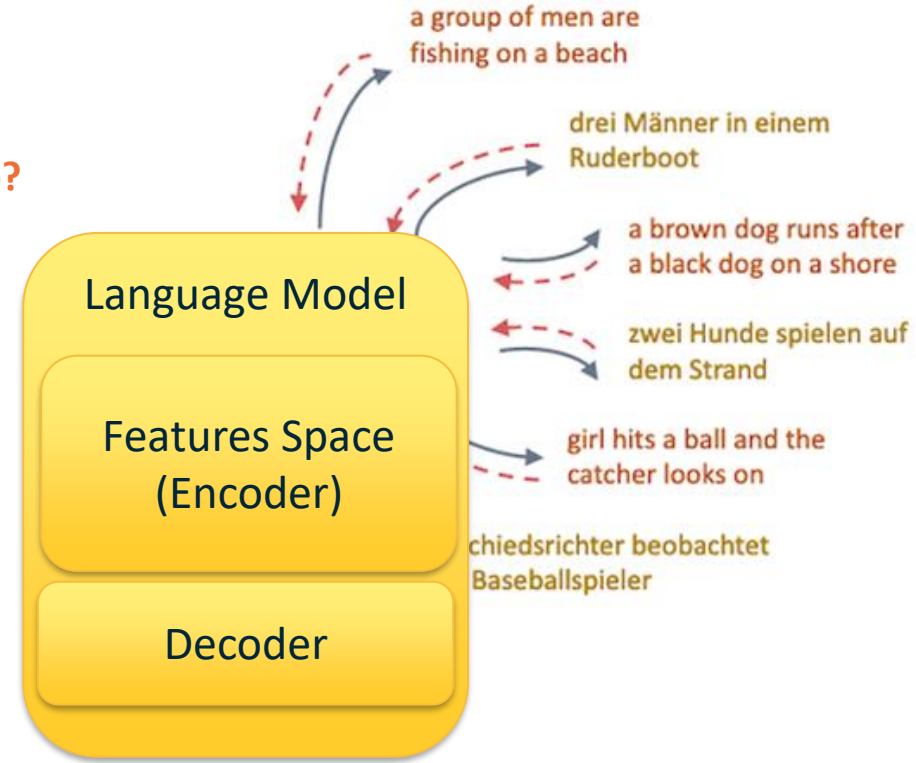
Using what data?

What tasks can we do?

Visual Feature Space

?

What should the interface be?

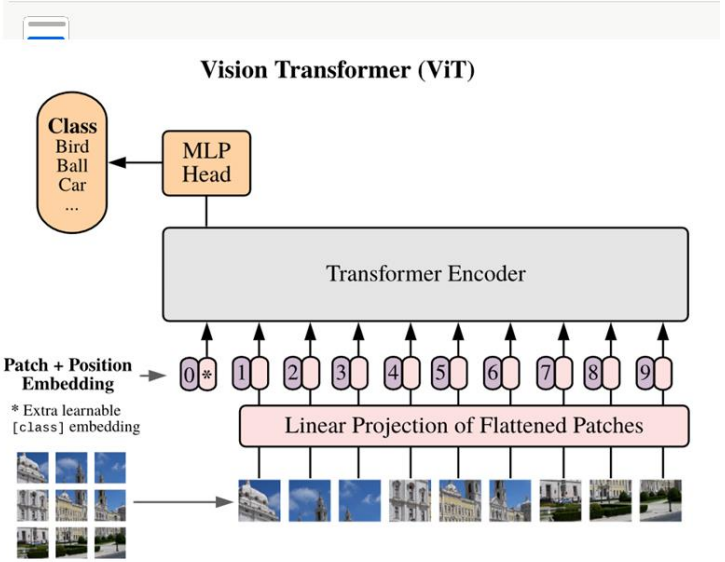


VLMs: Considerations

Georgia Tech

Potential ways of representing an image?

- Image encoder
  - Any architecture: ResNet, Vision transform (ViT)
  - Randomly initialized, SL/SSL pre-trained



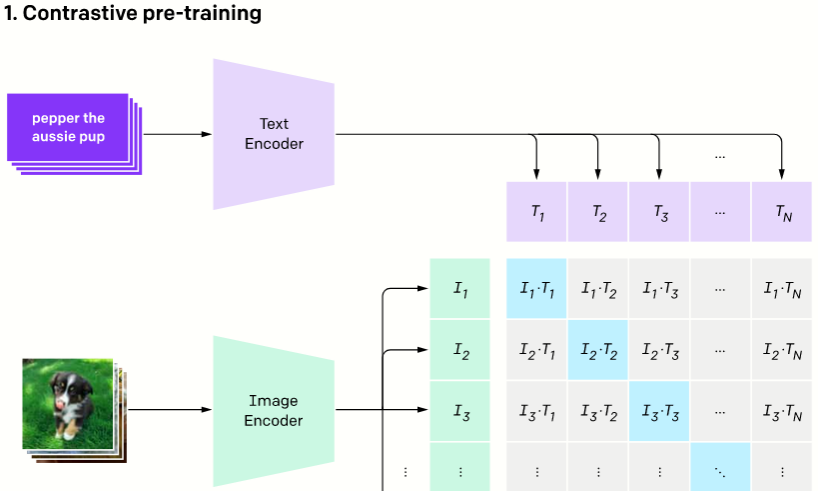
Slide by Vicente Ordóñez

Georgia Tech

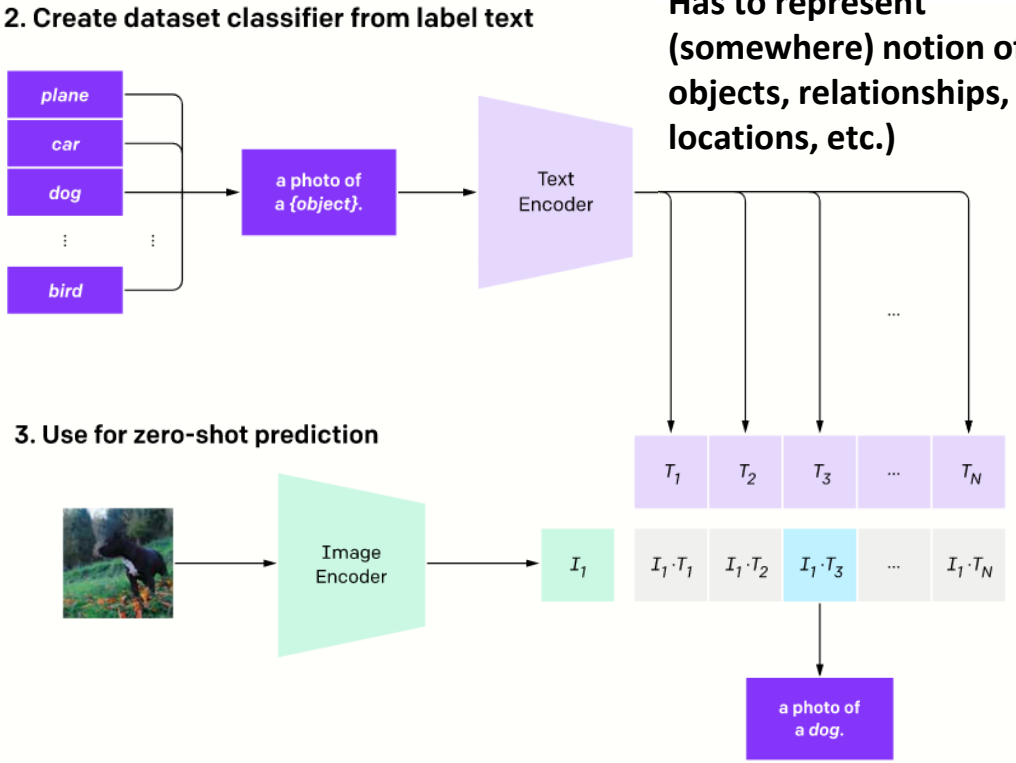
Method of alignment: Contrastive Learning

Downside?

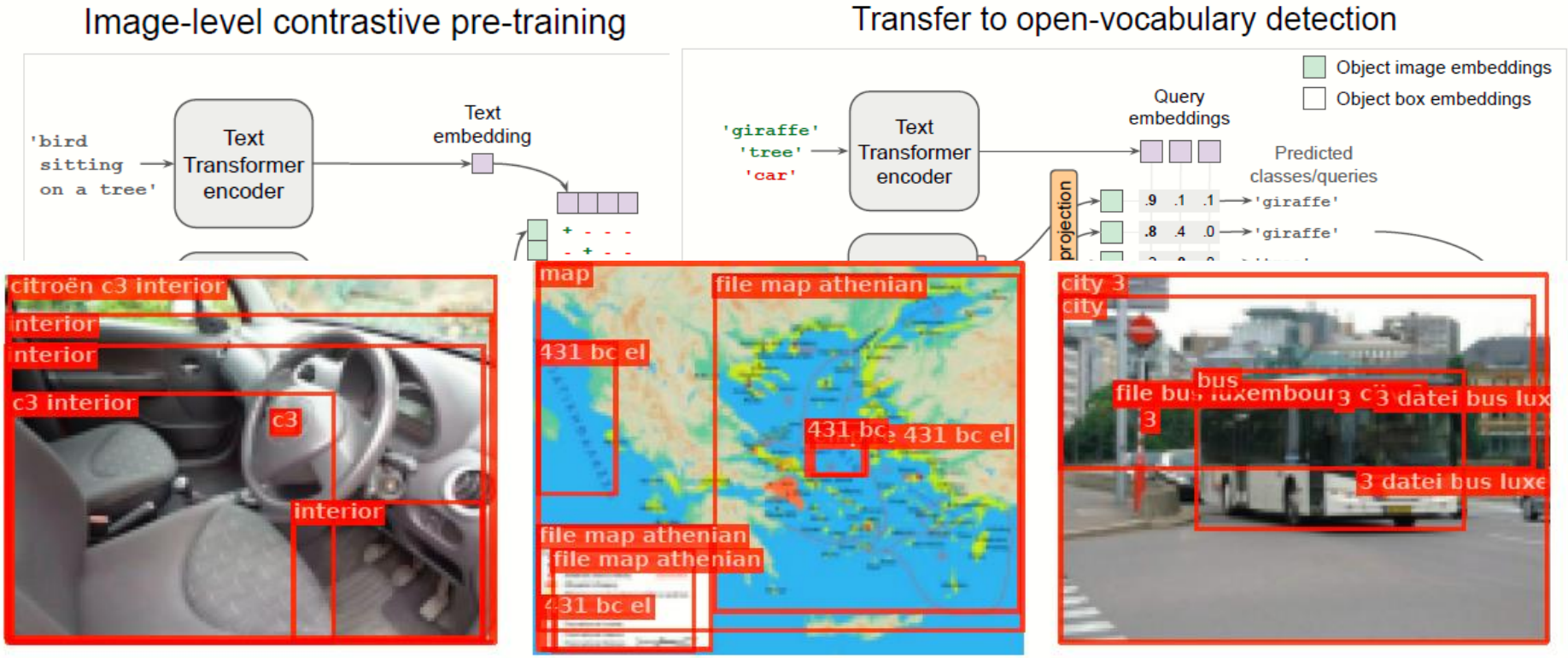
Coarse-grained. Has to represent (somewhere) notion of objects, relationships, locations, etc.)



Single image Encoder (ResNet, ViT)



Radford et al., Learning Transferable Visual Models From Natural Language Supervision

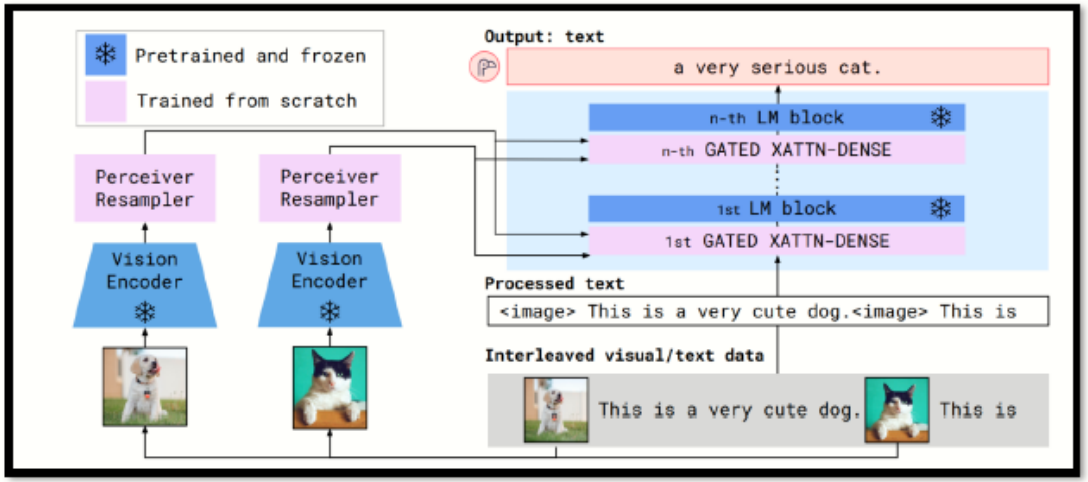


Minderer et al., Simple Open-Vocabulary Object Detection with Vision Transformers
Minderer et al., Scaling Open-Vocabulary Object Detection

Open-Vocabulary Detection

Georgia Tech

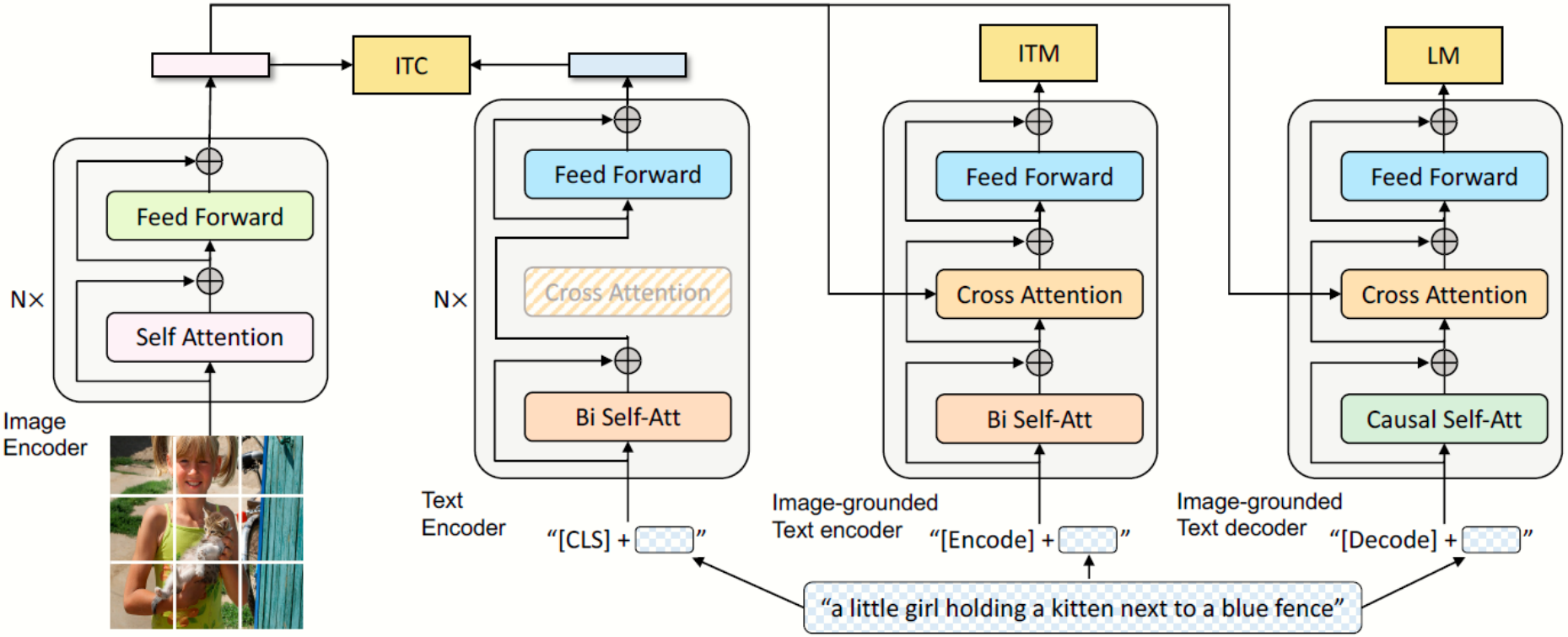
- Flamingo:



Language Model	Pre-trained: 70B Chinchilla
Connection Module	Perceiver Resampler Gated Cross-attention + Dense
Vision Encoder	Pre-trained: Nonrmalizer-Free ResNet (NFNet)

Slide by Chunyuan Li

Georgia Tech

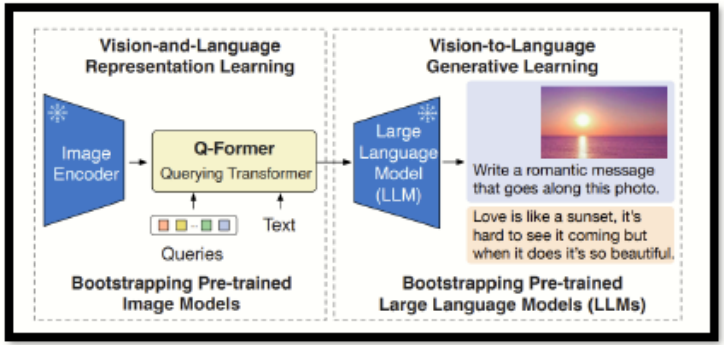


Li et al., BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation

BLIP

Georgia Tech

• BLIP2



Language Model	Pre-trained: FLAN-T5/OPT
Connection Module	Q-Former: Lightweight Querying Transformer
Vision Encoder	Contrastive pre-trained: EVA/CLIP

Recap on Language Modeling: Large Language Models (LLM)

OpenAI

GPT-2 → GPT-3 → ChatGPT / InstructGPT → GPT-4

What's new?

GPT-3	ChatGPT / InstructGPT	GPT-4
In-context-learning	In-context-learning	In-context-learning
Chain-of-thoughts (CoT)	Chain-of-thoughts (CoT)	Chain-of-thoughts (CoT)
	Instruction-Following	**Instruction-Following**
		Multimodal Input with image

Multimodal Space

Flamingo
BLIP2
GIT
...

Gap?
Instruction-Following
→ Alignment Research

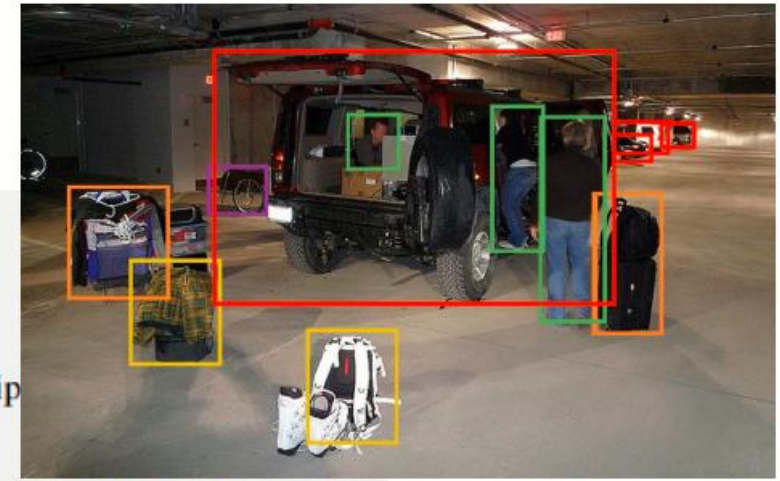
Multimodal GPT-4

GPT-assisted Visual Instruction Data Generation

Downside?

- Rich Symbolic Representations of Images
- In-context-learning with a few manual examples

→ Text-only GPT-4



Context type 1: Captions

A group of people standing outside of a black vehicle with various luggage.
Luggage surrounds a vehicle in an underground parking area
People try to fit all of their luggage in an SUV.
The sport utility vehicle is parked in the public garage, being packed for a trip
Some people with luggage near a van that is transporting it.

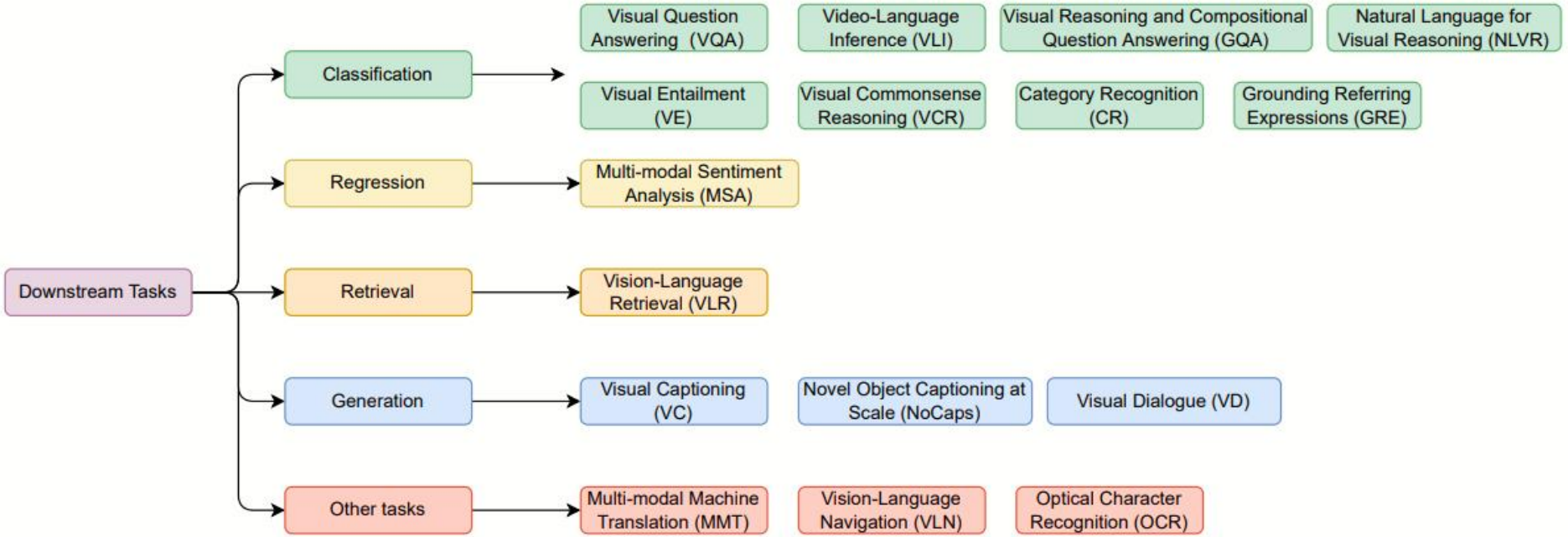
Context type 2: Boxes

person: [0.681, 0.242, 0.774, 0.694], person: [0.63, 0.222, 0.686, 0.516], person: [0.444, 0.233, 0.487, 0.34], backpack: [0.384, 0.696, 0.485, 0.914], backpack: [0.755, 0.413, 0.846, 0.692], suitcase: [0.758, 0.413, 0.845, 0.69], suitcase: [0.1, 0.497, 0.173, 0.579], bicycle: [0.282, 0.363, 0.327, 0.442], car: [0.786, 0.25, 0.848, 0.322], car: [0.783, 0.27, 0.827, 0.335], car: [0.86, 0.254, 0.891, 0.3], car: [0.261, 0.101, 0.787, 0.626]

Data – Using Captions

Slide by Chunyuan Li

Georgia Tech



VLP: A Survey on Vision-Language Pre-training, https://arxiv.org/pdf/2202.09061

Tasks

Visual Math and Text Reasoning

What is “a” in the first equation?
What is y in the last equation?

5 + a = 7	40 + y = 45
10 + y = 15	39 + a = 41
20 + a = 22	24 + y = 29
48 + a = 50	15 + a =17
10 + a = 12	34 + y = 39

The value of “a” in the first equation “5 + a = 7” is 2.

The value of “y” in the last equation “34 + y = 39” is 5.

Visual-Conditioned Joke/Meme

What is funny about it?

The humor in the comic may come from the situation, the characters, or the dialogue. The person is feeling unwell and is asked if they are pregnant or have the coronavirus, which is a current and relevant topic. The humor may come from the absurdity of the situation or the unexpectedness of the question.

Spatial/Coordinate Understanding

Where is the frisbee?
What is the relative position of the frisbee and the man?

The frisbee is located at coordinates (63,119,133,177).

The frisbee is located above the man in the image. The man is falling on the ground while the frisbee is in the air. The frisbee is also closer to the top of the image than the man..

Visual Planning and Prediction

What should I do after kneading the dough, before dividing it into pieces? For how long?

After kneading the dough, you should put it in a warm place before dividing it into pieces, according to the recipe in the image.

You should put the dough in a warm place for 1 hour.

Multi-Image Reasoning

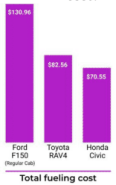
How much in total did I spend on travel?
How much in total did I pay for taxes?

You spent $1299.90 on airfare and $43.83 on an Uber ride for a total of $1343.73 on travel.

You paid $323.23 in total for taxes. This includes $321.90 for taxes and fees for the flight purchase and $1.33 for sales tax for the Windmill Family Restaurant.

Multi-Hop Document Understanding

Excluding the regular cab, what is the average total fueling cost?



The average total fueling cost excluding the Ford F150 (Regular Cab) is $76.55. This is calculated by taking the sum of the total fueling cost of Toyota RAV4 and Honda Civic and dividing it by 2. ($82.56 + $70.55) / 2 = $76.55.

Open-World Concept Understanding

What is the brand of these drink? Which one is unsweetened?

The brand of these drinks is “BETTER THAN MILK”. The unsweetened one is “BETTER THAN MILK organic OAT DRINK”.

Who are the celebrities in this image?

Keira Knightley and Natalie Portman are the celebrities in this image.

Is it safe to eat this mushroom?

Morel mushrooms are generally considered safe to eat when properly cooked.

Video Summarization/Event Localization

Can you breakdown this video tutorial to step-by-step instructions, with start and end time?

Here's a step-by-step breakdown of the tutorial with start and end time.

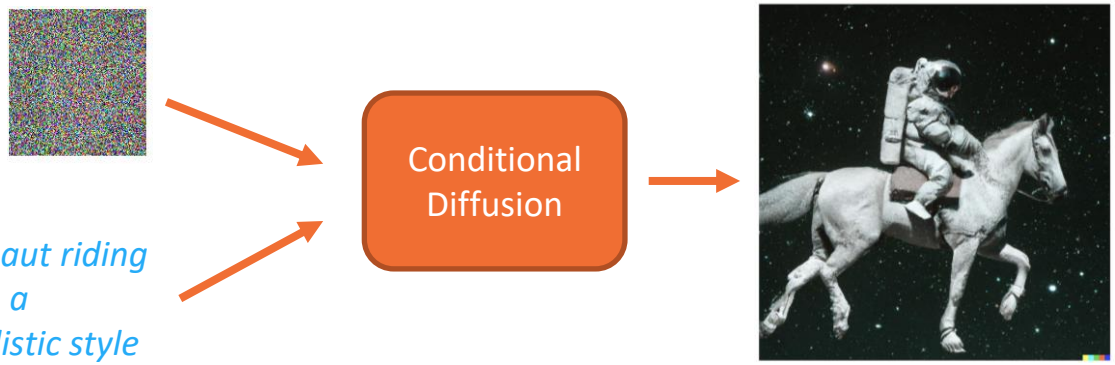
... Add Lettuce Add Tomato Add Bacon ...

Yang et al., MM-ReAct MM-ReAct: Prompting ChatGPT for Multimodal Reasoning and Action

Motivation: Vision-Language Models

Georgia Tech

Conditional Diffusion Models



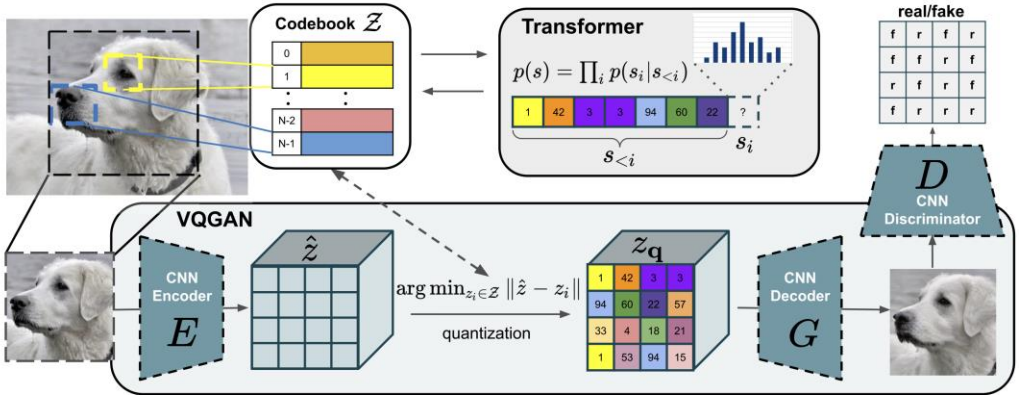
Simple idea: just condition the model on some text labels y!

$$\epsilon_\theta(x_t, y, t)$$

Latent-space Diffusion

Problem: Hard to learn diffusion process on high-resolution images

Solution: learn a low-dimensional latent space using a transformer-based autoencoder and *do diffusion on the latent space*!

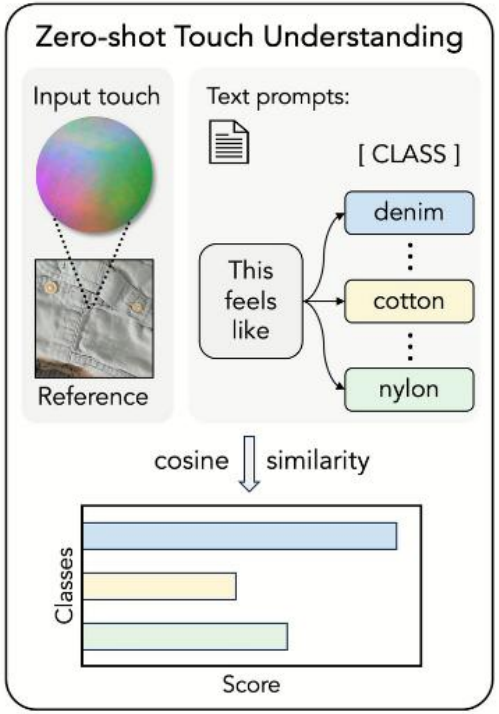


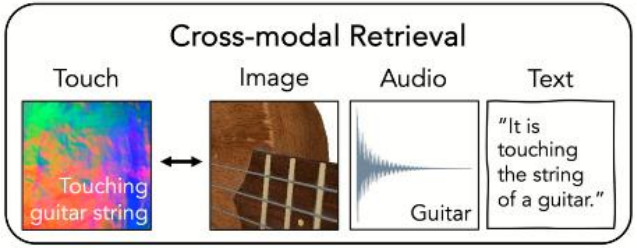
The latent space autoencoder

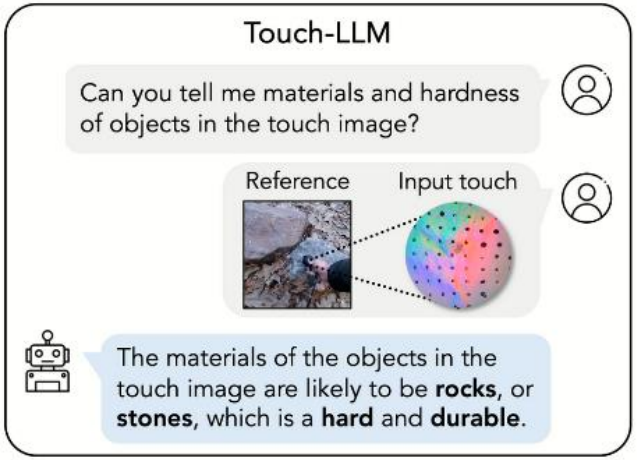
Esser and Rombach *et al.*, 2021

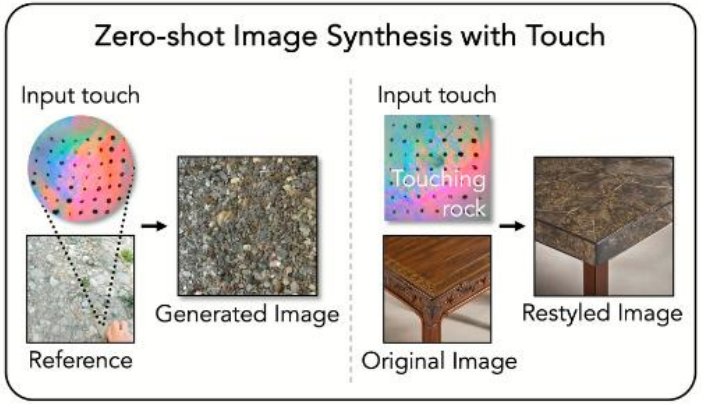
Videos!

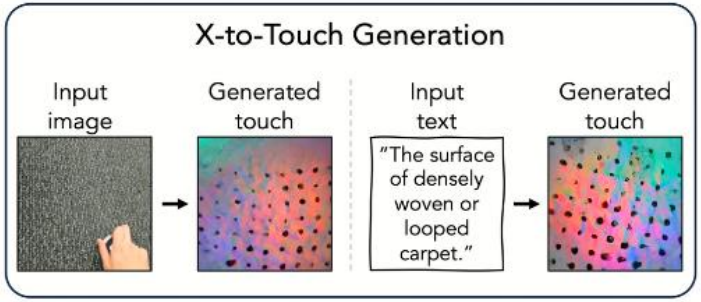
https://research.nvidia.com/labs/toronto-ai/VideoLDM/o-video/











Yang et al., Binding Touch to Everything: Learning Unified Multimodal Tactile Representations, CVPR 2024

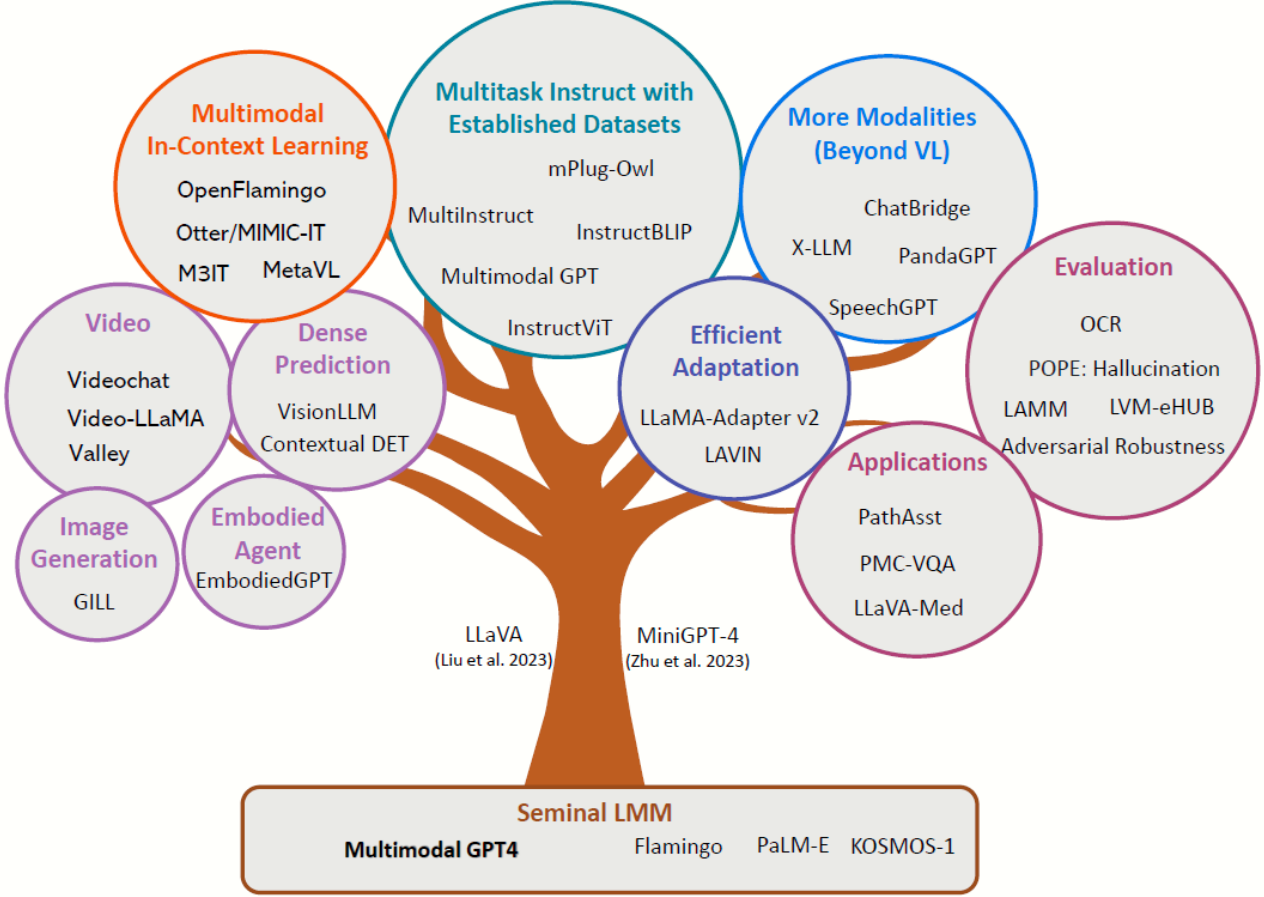
Other Modalities

Georgia Tech

- Fusion: Early, mid, late
- Pre-training? Instruction tuning? RLHF?
- Data Generation
- Training Objectives
- Alignment between modalities
- Grounding
- Discriminative / Generative
- Promptable/steerable models?
- VLM Reasoning?
- Applications
  - Zero-shot/open-vocabulary perception
  - Multi-Modal Tasks
    - Captioning, Vision-Question Answering
  - Decision-Making (Web GUI Agents, Embodied AI, etc.)
- **What are we improving? How does it come about?**

Important Concepts / Considerations

Georgia Tech



Slide by Chunyuan Li

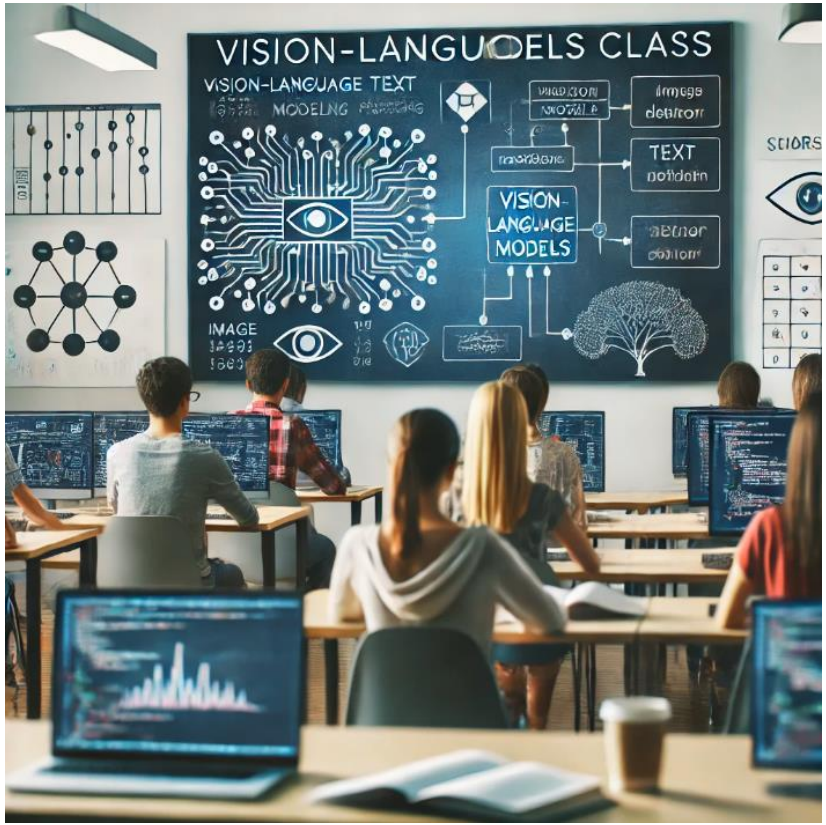
Vision-Language Explosion (2023)

Georgia Tech

- Understand the architectural, training, dataset, and evaluation aspects of multi-modal models.
  - Be able to compare & contrast
  - Have a holistic picture of the space
- Develop skills to read papers, identify gaps/limitations, create novel methods, and rigorously evaluation them.
- Develop skills for presenting research results.

Learning Objectives

Georgia Tech



Xijia Polina Zhang

Logistics

Georgia Tech

- This will be a **participation-driven course**
- In-person attendance will be mandatory!
  - No remote option, no recordings except for accommodations
  - Don’t come in sick please!!!

Delivery Format

Georgia Tech

- It's all deep learning – We expect some deep learning background!
  - CS 7643 (Deep Learning) or similar

Survey:
- Deep Learning class/equivalent?
- Transformers?
- PyTorch/Tensorflow?
- Vision? Language? Audio? Other?
- ECCV? CVPR? EACL? CHI? ICRA? COLING? ICLR?

- What do you want to get out of the class?

- We do not expect deep experience in all modalities!

Recommended Background

Georgia Tech

- **15% Class Participation**
- **20% Paper Reviews**
- **15% Paper Discussion**
- **50% Project**
  - 5% Proposal
  - 10% Midterm Presentation
  - 15% Final Presentation & Video
  - 20% Final Report

Credits: This course was designed with materials and inspiration from great courses such as Learning from Limited Labels and Internet Data Science

Sign-up is first come first served. Each paper needs a pro/con presenter, compare/contrast can be selected once that is full.
Paper selection subject to change.

			Discussion Lead		
Date	Topic	Paper	Presenter 1 - Pros (First Last Name)	Presenter 2 - Cons (First Last Name)	Presenter 3 Compare/Contrast, Related Work (First Last Name)
Thu, Aug 29		CLIP, ALIGN, & CoCa			
Tue, Sep 3		Open-Vocabulary (e.g. OWL/v2, LSeg)			
Thu, Sep 5		Datasets / Eval			
Tue, Sep 10		ViLT			
Thu, Sep 12		FLAVA			
Thu, Sept 19		Frozen			
Tue, Sep 24		Flamingo			
Thu, Sep 26		BLIP / BLIP-2			
Tue, Oct 1		Llava / LLava1.5/Llava-NeXt/Llava-OneVision			
Thu, Oct 3		Intern-VL			
Tue, Oct 8		UnifiedVLP/UnifiedIO-2			
Tue, Oct 10		Chameleon			
Thu, Oct 24		ViperGPT			
Tue, Oct 29		MM-CoT/ReAct			
Thu, Oct 31		Web Agents (WebGUM, SeeClick)			
Tue, Nov 5		Embodied AI (PALM-E/RT-2, etc.)			
Thu, Nov 7		Something with other modalities (e.g. audio) from Owens et al. Another option is ImageBind			
Tue, Nov 12		A Framework for Multi-modal Learning: Jointly Modeling Inter- & Intra-Modality Dependencies			
Thu, Nov 14		Survey paper			
Thu, Nov 19		Survey paper			

Some changes from last time:

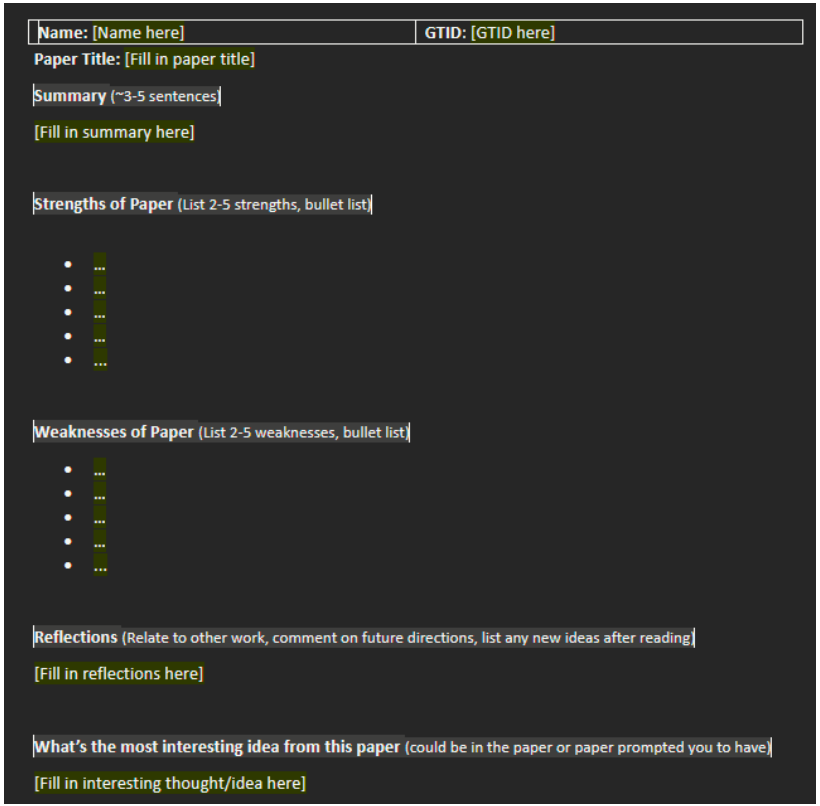
- +1 lecture on VLM “historical” models
- More papers on recent advancements

Last Year’s Schedule

Georgia Tech

Paper Reviews

- Template (instructions are in Word comments)
- 1-page review of each paper (except if presenter).
- **Due: 11:59 pm ET on the day before class**. Late reviews will not be accepted.
- We will drop your lowest 3 submissions.
- Limit reviews to 1-pg using our review template



Name: [Name here] | **GTID:** [GTID here]

Paper Title: [Fill in paper title]

Summary (~3-5 sentences)

[Fill in summary here]

Strengths of Paper (List 2-5 strengths, bullet list)

- …
- …
- …
- …
- …

Weaknesses of Paper (List 2-5 weaknesses, bullet list)

- …
- …
- …
- …
- …

Reflections (Relate to other work, comment on future directions, list any new ideas after reading)

[Fill in reflections here]

What's the most interesting idea from this paper (could be in the paper or paper prompted you to have)

[Fill in interesting thought/idea here]

- ~1 page total (separated into parts below)

- Detailed review
  - Brief (~3-4 sentence) summary
  - Main contribution
  - Strengths? Weaknesses?
- Relationship observed between the papers we are reading
- Pull out most interesting thought
- View class webpage for details

- **Write in your own words** and proof read (this is individual work!)

- Template
  For each paper, two to three students will be responsible for a 45-minute presentation including:
  - **Jointly** - set of slides summarizing the paper (problem, methods, experiments)
  - **Each student** presenting one focused slide on **strengths, weaknesses, and (if there is a third student) related papers**

- After (or during) the presentation we will have in-depth discussions! In addition to the above, you will have to jointly: Come up with five questions/points of discussion

- Add any additional resources for students to explore if background is needed or they would like to learn more

- Submit draft slides:
  - Friday before week of presentation (Tuesday presentation)
  - Monday on week of presentation (Thursday presentation)

- Submit final form the night before class presentation **11:59 pm ET**
  - Option of a practice presentation during office hours (optional).

Sign-up is first come first served. Each paper needs a pro/con presenter, compare/contrast can be selec

Paper selection subject to change.

Date	Topic	Paper	Presenter 1 - Pros (First Last Name)	Presenter 2 - Cons (First Last Name)
Thu, Aug 29		CLIP, ALIGN, & CoCa		
Tue, Sep 3		Open-Vocabulary (e.g. OWL/v2, LSeg)		
Thu, Sep 5		Datasets / Eval		
Tue, Sep 10		ViLT		
Thu, Sep 12		FLAVA		
Thu, Sept 19		Frozen		
Tue, Sep 24		Flamingo		
Thu, Sep 26		BLIP / BLIP-2		
Tue, Oct 1		Llava / LLava1.5/Llava-NeXt/Llava-OneVision		
Thu, Oct 3		Intern-VL		
Tue, Oct 8		UnifiedVLP/UnifiedIO-2		
Tue, Oct 10		Chameleon		
Thu, Oct 24		ViperGPT		
Tue, Oct 29		MM-CoT/ReAct		
Thu, Oct 31		Web Agents (WebGUM, SeeClick)		
Tue, Nov 5		Embodied AI (PALM-E/RT-2, etc.)		
Thu, Nov 7		Something with other modalities (e.g. audio) from Owens et al. Another option is ImageBind		
Tue, Nov 12		A Framework for Multi-modal Learning: Jointly Modeling Inter- & Intra-Modality Dependencies		
Thu, Nov 14		Survey paper		
Thu, Nov 19		Survey paper		

Let me know if you may drop the course, before signing up.

- Note: We will release semi-finalized list by tomorrow morning
- Feel free to suggest papers!
- Papers subject to updating, but similar topic outline
- **Sign up by Monday**

Paper Presentation - Schedule

Georgia Tech

- **50% Project**
  - 5% Proposal
  - 10% Midterm Presentation
  - 15% Final Presentation & Video
  - 20% Final Report

- Goal is to develop a novel approach to a multi-modal problem
- Shooting for publications at top venues highly encouraged!

- Possibilities
  - Design and evaluate a novel approach
  - A novel application, use case
  - Extension of a technique studied in class
  - Be creative!

- Target is a research paper at a good conference

- Work in teams of 3-4 (varies by enrollment)

- Sign up for teams by September 6 11:59PM ET

- Three in-class presentations (class schedule coming tonight)

- Project proposal [5%]
- Project update [10%]
- Final presentation and short video [15%]
- Final report [20%]

- Make sure you are saying everything we need to know to understand you
- Practice presenting and make sure you know what you want to cover
  - Practice is the **BIGGEST** driver of good presentations
- Think about your audience
- Make your talks visual and animated
- **Stick to the time limit!**

- Clearly define the problem statement (input, output)
- Place your work in the context of existing work
- Lay out a set of **claims** and the set of experiments you'll conduct to demonstrate your claims and effectiveness of approach
  - Did your claims actually get validated by results?
- Present a timeline
  - Concrete goals for the next update
  - Propose long shot goals
  - Present updates following same plan
  - See class webpage for more details

- Use any language / platform / package you like
- No support for code / implementation issues will be provided — this is not a programming assignment!
- AI Assistants OK for implementation
  - But you are the driver of the creativity, novelty, and approach!

Grade	Percentage Range
A	90% and above
B	80% - 89%
C	70% - 79%
D	60% - 69%
F	Below 60%

Grading Scale

Georgia Tech

1. *Read first, Write later*. Read the ENTIRE set of posts/comments before posting.
2. *Avoid language that may come across as strong or offensive*. Language can be easily misinterpreted.
3. *Follow the language rules of the Internet*. Do not write using all capital letters, use emoticons to temper.
4. *Consider the privacy of others.* Ask permission prior to giving out a classmate's email address or other information.
5. *inappropriate material*. Do not forward virus warnings, chain letters, jokes, etc. to classmates or instructors. The sharing of pornographic material is forbidden.

Be professional, objective, and kind!

NOTE: The instructor reserves the right to remove posts that are not collegial in nature and/or do not meet the Online Student Conduct and Etiquette guidelines listed above.

What is the collaboration policy?

- **Collaboration**
  - Encouraged for presentation and project (team members)
  - Can also discuss anything with classmates, but **reviews must all be written by you!**
    - Cite all resources, including collaborators, used
  - **Each student writes their own answers**

- **AI Assistants:**
  - **Paper Reviews, Presentations, Ed Posts/Discussions**: NO! All materials submitted for these deliverables must be entirely your own.
  - **Project:** It is OK to discuss the projects with others (of course including your team) and you are free to use whatever online codebases, blogs, resources, and AI coding assistants (e.g. Copilot) if you wish. Explicitly acknowledge any and all resources used. **Describe how you used AI if you did.**

- Zero tolerance on plagiarism
  - Neither ethical nor in your best interest
  - Always credit your sources, Don't cheat. We will find out.

How do I get in touch?

- Primary means of communication -- **Ed**
  - No direct emails to Instructor unless private information
  - Instructor/TAs can provide answers to everyone on forum
  - Class participation credit for answering questions!
  - No posting answers. We will monitor.
  - Stay respectful and professional

GT Resources for Mental Health

Georgia Tech Police Department
Emergency: Call 911 | 404-894-2500

Dean of Students Office
404-894-2565 | studentlife.gatech.edu
Afterhours Assistance Line & Dean on Call: 404-894-2204

Center for Assessment, Referral and Education (CARE)
404-894-3498 | care.gatech.edu

Collegiate Recovery Program
404-894-2575 | counseling.gatech.edu

Counseling Center
404-894-2575 | counseling.gatech.edu

Health Initiatives
404-894-9980
healthinitiatives.gatech.edu

LGBTQIA Resource Center
404-385-4780 | lgtbqia.gatech.edu

Stamps Psychiatry Center
404-894-1420

VOICE
404-385-4464 | 404-385-4451
24/7 Info Line: 404-894-9000 | voice.gatech.edu

Women's Resource Center
404-385-0230 | womenscenter.gatech.edu

Veterans Resource Center
404-894-4953 | veterans.gatech.edu

Georgia Crisis and Access Line
1-800-715-4225
The crisis line is staffed with professional social workers and counselors 24 hours per day, every day, to assist those with urgent and emergency needs.

Trevor Project
1-866-488-7386
Trained counselors are available to support anyone in need.

National Suicide Prevention Hotline
1-800-273-8255
A national network of local crisis centers that provides free and confidential emotional support to people in suicidal crisis or emotional distress 24/7.

Georgia State Psychology Clinic
404-413-2500
The clinic offers high quality and affordable psychological services to adults, children, adolescents, families and couples from the greater Atlanta area.

Computing

- Major bottleneck: GPUs
- Options
  - Your own / group / advisor's resources
  - PACE-ICE (w/ some H100s)
  - Google Colab
    - jupyter-notebook + free GPU instance
  - Google Cloud credits
    - Tutorial on setting up gloud: https://github.com/cs231n/gcloud

FAQ

- Can I audit? **No** ☹
- Will I get a seat?
  - We can’t expand class size. Waitlist is only hope.

- Read the class website **thoroughly** after my initial Ed post tonight
- Tentative schedule will be up by tonight
- Sign up to lead a discussion by Monday August 25th
- Probability of dropping class? Talk to me first.
- Start thinking about project teams

Coming Up

Georgia Tech

- **Have fun!**