Advanced Deep Learning in Computer Vision

Dimitrios Papadopoulos
Associate Professor, DTU Compute

Schedule

13:00-13:40 Project Proposals (Aasa, Morten, Dimitrios)

13:40-13:50 Break

13:50-17:00 Reading group

13:50-13:55 Introduce and assign roles
13:55-14:25 Do the role!
14:25-14:40 Discuss with your role-partner
14:40-14:50 Break
14:50-15:50 Discuss with your paper group (with Paper Responsible)
15:50-16:00 Finalize slides
16:00-16:10 Break
16:10-17:00 Present slides (4papers*4roles*3minutes)

Schedule

Date	What	Who
Tuesday 30/1	Intro to Course / Transformers and Attention	Dim
Tuesday 6/2	Visual Transformers and Applications	Dim
Tuesday 13/2	Generative models: DDPM	Aasa
Tuesday 20/2	Generative models: Text2Image (+CLIP)	Aasa
Tuesday 27/2	Neural Radiance Fields	Morten
Tuesday 5/3	Low-level Computer Vision (Super-resolution, denoising, video frame interpolation)	Morten
Tuesday 12/3	Project Start + Reading Group Week	All
Tuesday 19/3	Project Plan	All
Tuesday 26/3	Easter holiday	
Tuesday 2/4	Video Understanding (Actions, 3D, two-stream, multi-modal)	Dim
Tuesday 9/4	Limited Labels: Self-supervised learning, Weakly supervised, and active learning	Dim
Tuesday 16/4	Explainability (or GNNs and Point clouds)	Aasa
Tuesday 23/4	Fairness and Bias and Ethics	Aasa
Tuesday 30/4	Project presentation (poster session)	All
Wednesday 22/5	Exam: Oral	

"Exercise" weeks

Reading group

"Project" weeks

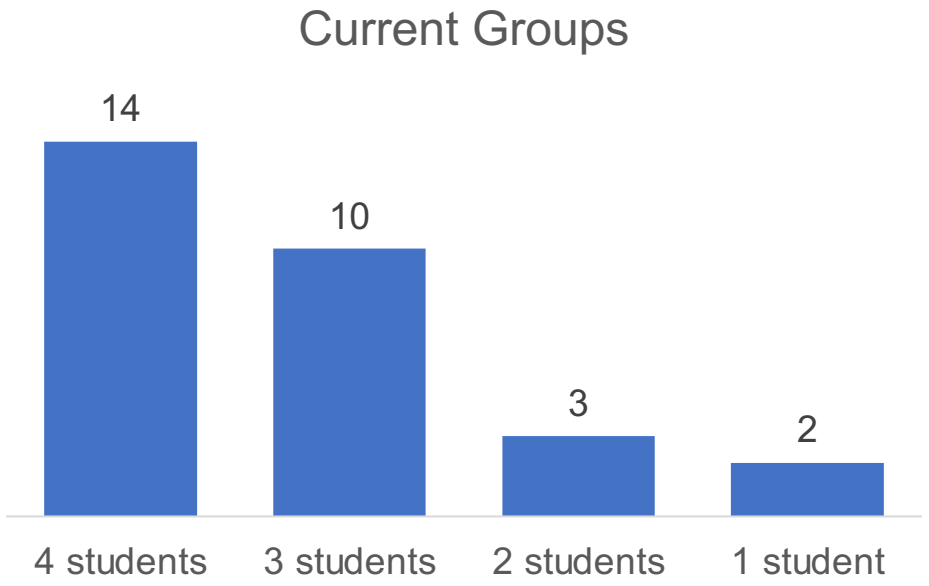
Projects

- Project work in **groups**. **Maximum** is 4 students.
  We recommend 4 (or 3 if not possible)

- Project topic:
  - Select a project from our proposals (**maximum 2 groups**)
  - Propose your own project (aligned with the content of the course)

- Weekly chat with project supervisor

- Deliverable: 1 poster in May



Projects - Timeline

Start

- 12/03
- Project proposals from lecturers

Plan

- 19/03
- Deadline: finalize your project plan (select from our proposals or propose your own)

Poster

- 30/04
- Live Poster Session. Receive and give feedback to peers.

Hand-in

- 8/05
- Hand-in final poster version on Learn

Exam

- 22/05
- Individual oral exams, questions on project work and on lectures/exercises

Projects - Dimitrios

Transformers (Image, Text)

FoodCLIP

CubeDETR

Transformers (Video, Audio, Multimodal)

FootballVAR

Aisynthesizer

Diffusion (DDPMs, Stable diffusion, Contitional Controls)

GraphControl

SuperMarioDiffusion

Diffusion (but for recognition tasks)

DiffAugment

DiffSAM

How to make a Smørrebrød: Image-to-Recipe Retrieval

Keywords/lectures: Transformers, Vision Transformers, image-to-text, CLIP

Figure 1: **Food Understanding using computer vision and NLP.** The goal of this project is to perform image-to-recipe retrieval tasks. The model will be able to retrieve a recipe (from a list of known recipes) given an image query and, in reverse, to retrieve an image (from a list of known images) given a text recipe.

CubeDETR: End-to-End 3D Object Detection with Transformers

Keywords/lectures: Transformers, Object Detection

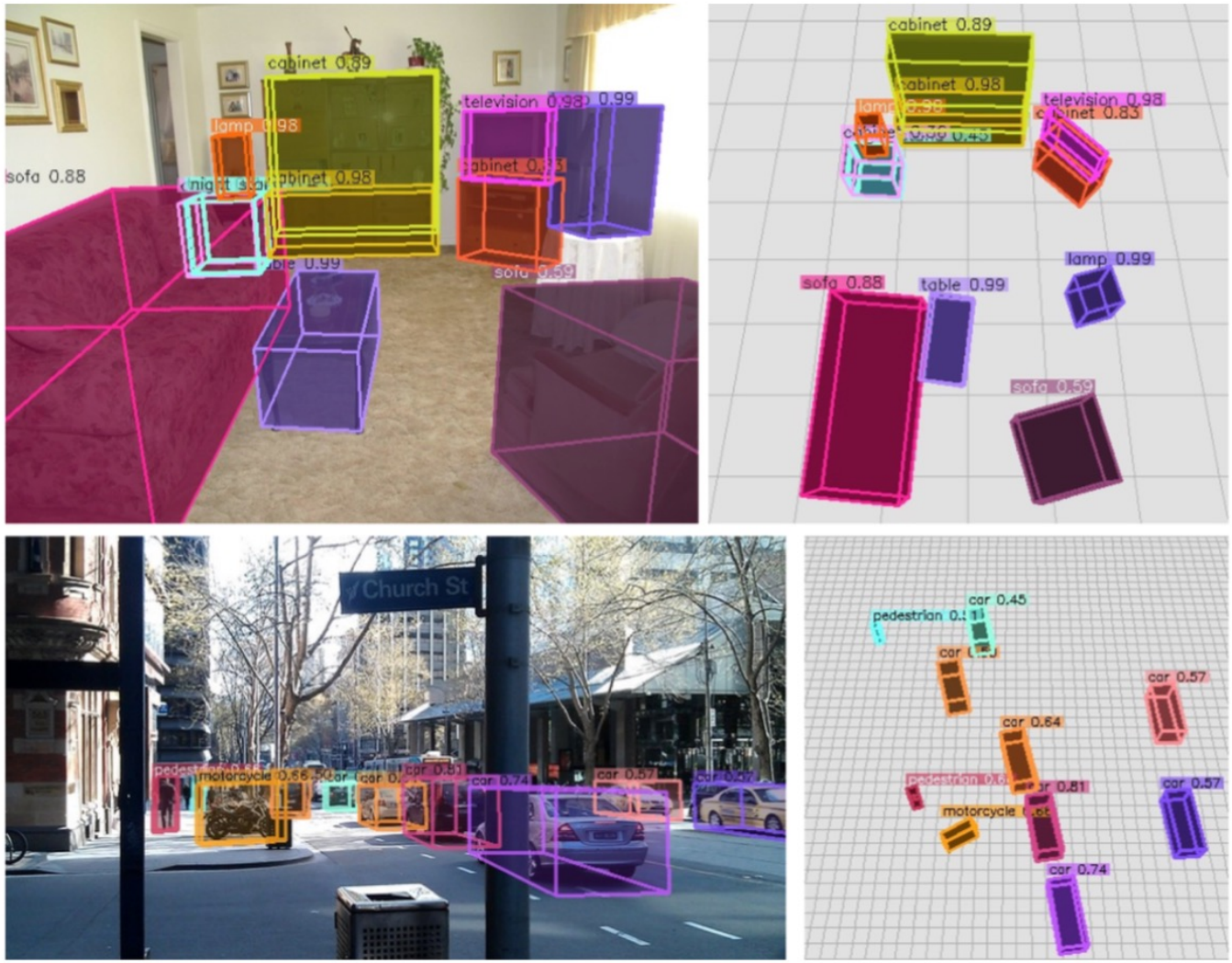
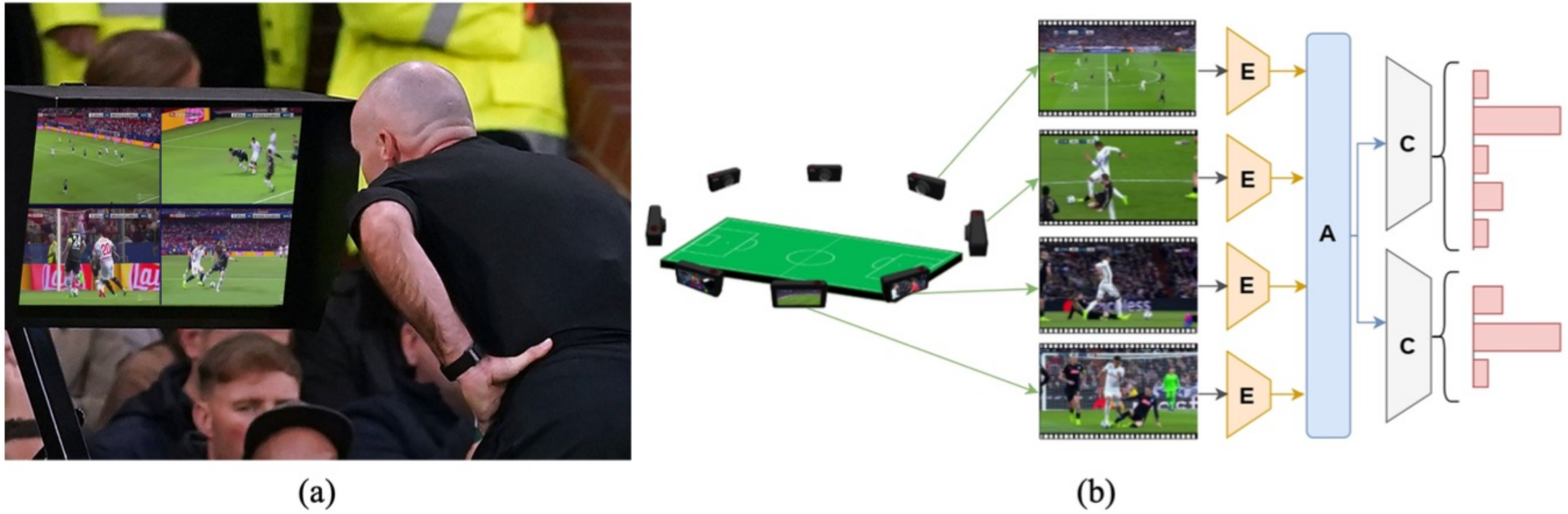


Figure 1: **3D Object Detection task.**

FootballVAR: AI Video Assistant Referee for Foul Decisions in a Football Game from Multiple Views

Keywords/lectures: Transformers, Video Understanding, Multimodal learning



(a) (b)

Figure 1: **Video Assistant Referee (VAR).** (a) The current Video Assistant Referee system in football that is used worldwide. (b) Learning to automatically detect fouls in a football game from multiple views (left figure from [3])

AIsynthesizer: Learning to generate music sounds from silent videos

Keywords/lectures: Transformers, Audio Recognition, Speech synthesis (TTS)

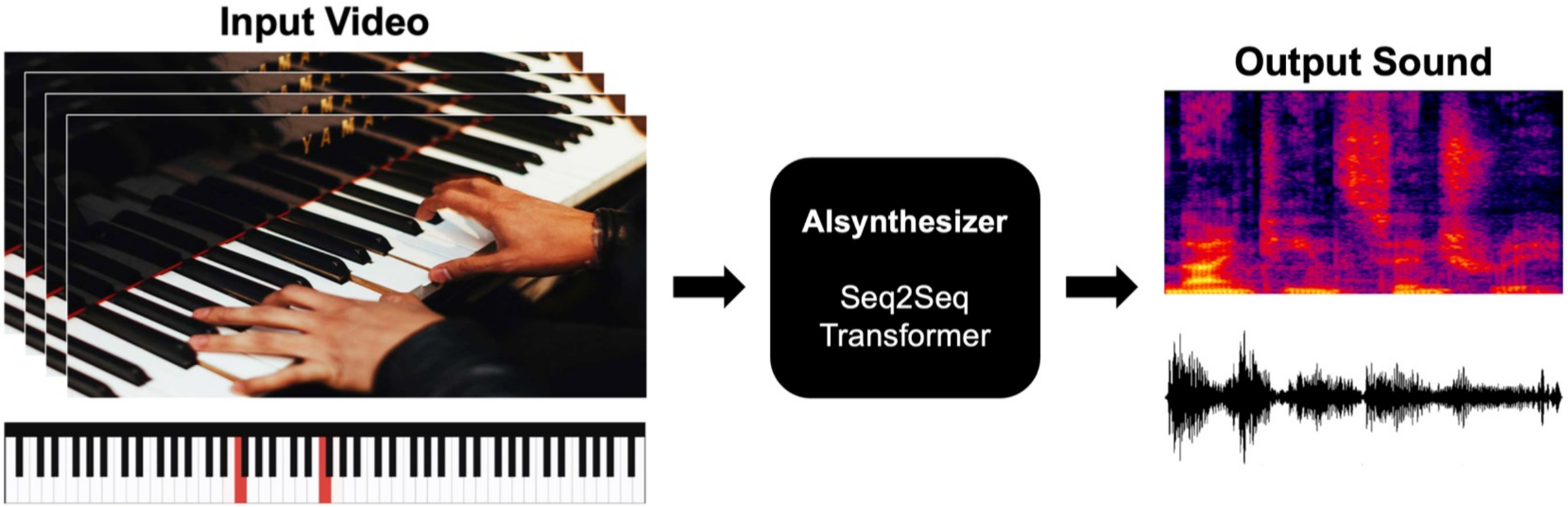


Figure 1: **AIsynthesizer.** The goal of this project is to build a model that takes as input a silent video of a playing musical instrument and predicts its raw sound. The idea is to use a sequence-to-sequence model based on transformers.

GraphControl: Generate images from scene graphs with diffusion models

Keywords/lectures: Diffusion, ControlNet, ControlNet, Graph Neural Networks

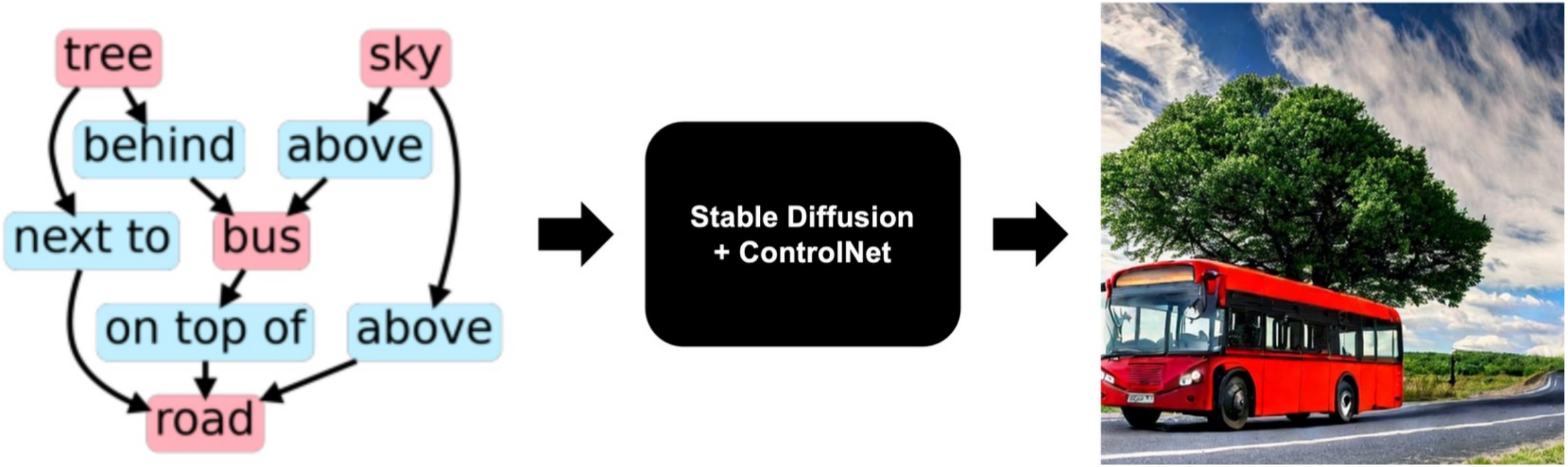


Figure 1: **GraphControl.** The goal of this project is to learn to generate images from scene graphs using a stable diffusion model and ControlNets [6].

SuperMarioDiffusion: Learning to Simulate Game Environments with Diffusion Models

Keywords/lectures: Diffusion, ControlNet, Game Engine, Neural Game Simulator

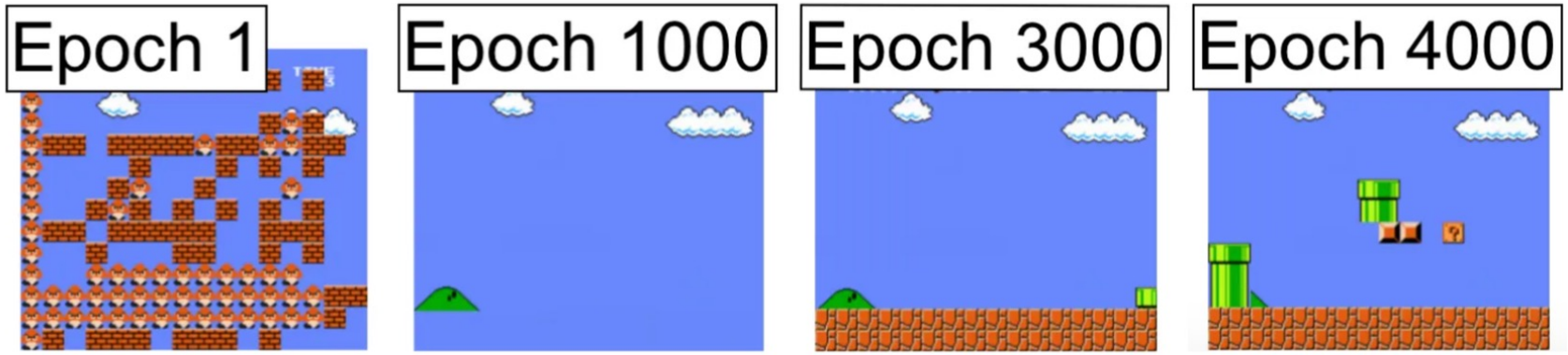
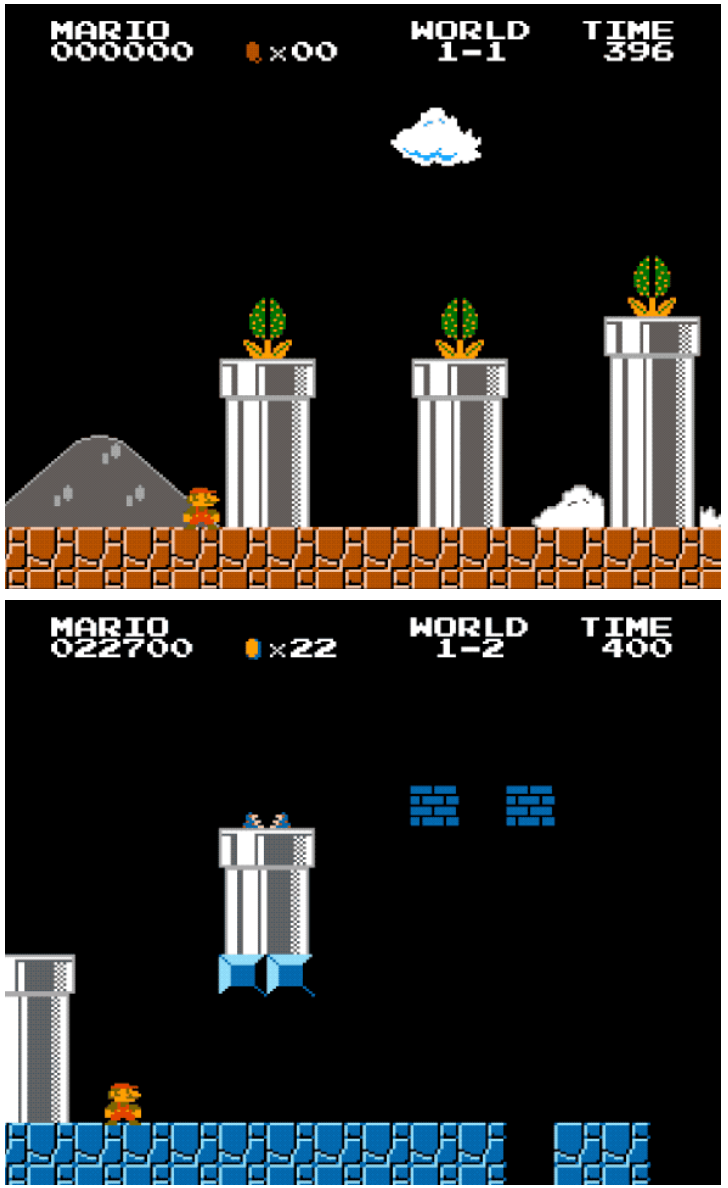


Figure 1: **Generating Super Mario Levels with a Diffusion Probabilistic Model.**



DiffAugment: Data Augmentation with DDPM for Object Recognition

Keywords/lectures: Diffusion models, DDPM, text2image, Data augmentation, Limited labels

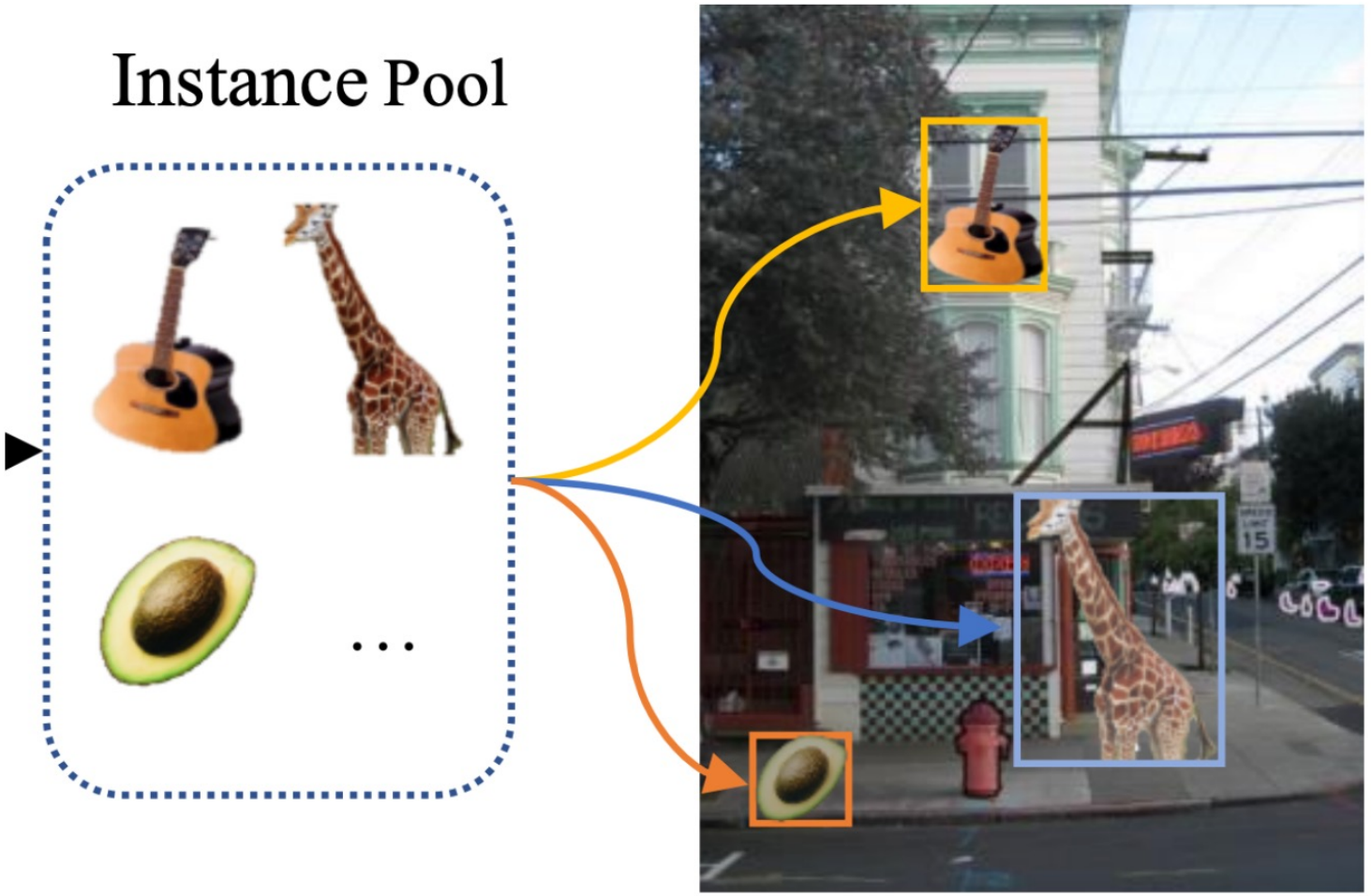


Figure 1: **Data Augmentation for Instance Segmentation.**

DiffSAM: Segment Anything with Diffusion and Controllable Prompting

Keywords/lectures: Diffusion, ControlNet, SAM

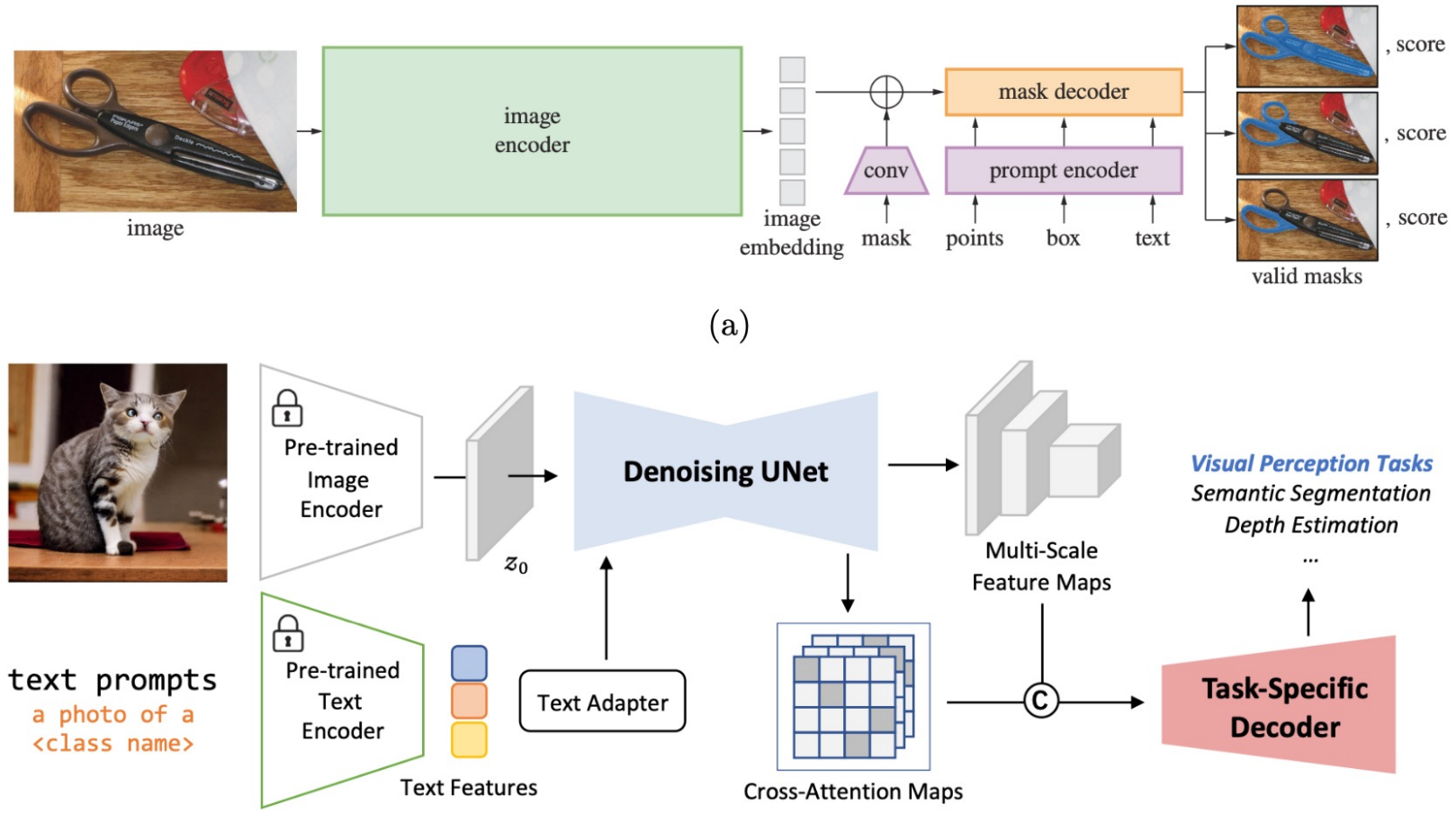


Figure 1: **(a) The Segment Anything model [5]. (b) The overall framework of VPD [13].** The goal of this project is to build a model that uses a diffusion model as a backbone and ControlNets for promptable segmentation

Anomaly Detection: Can you find gorillas in CT scans?

02501 Advanced Deep Learning in Computer Vision

Aasa Feragen
Professor, DTU Compute

12 March, 2024

Keywords/lectures: Diffusion models, anomaly detection

Figure 1: **Spotting gorillas is hard.** Radiologists tend to not notice gorillas embedded in CT scans [1] – can your ML algorithms do better?

Do you remember? Memorization in diffusion models

02501 Advanced Deep Learning in Computer Vision

Aasa Feragen
Professor, DTU Compute

12 March, 2024

Keywords/lectures: Diffusion models, memorization

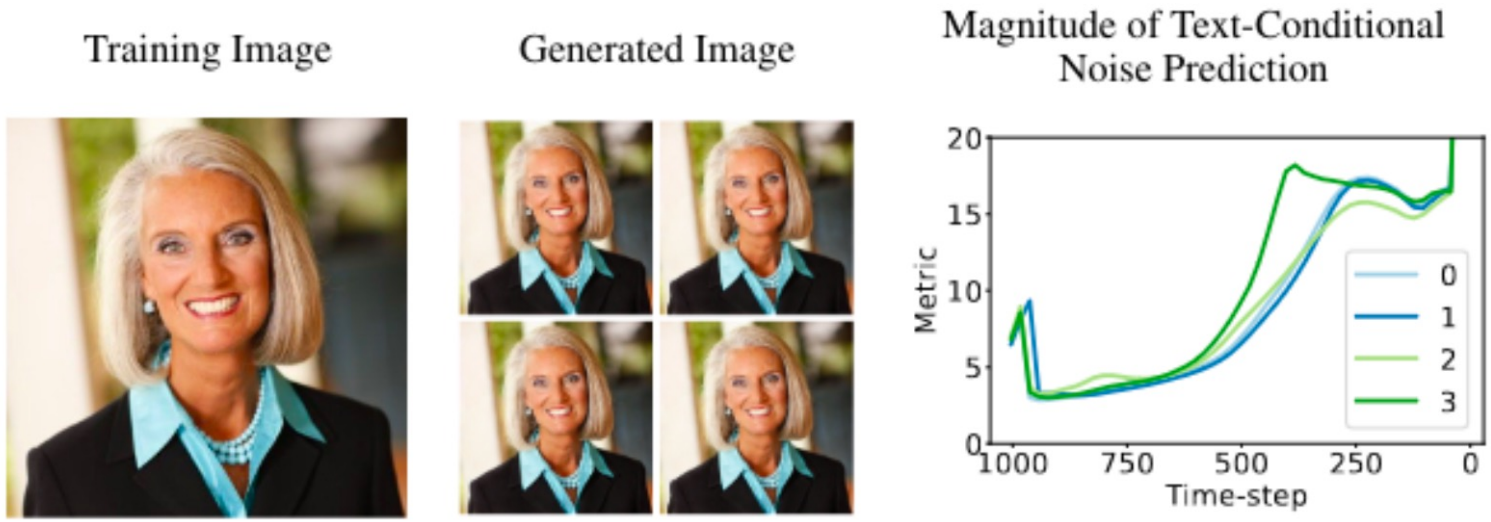


Figure 1: **Memorization in diffusion models.** Generative models have been known to memorize – diffusion models are no exception. This comes, however, at great risk for violating both property law and privacy when generative models are used in off the shelf software.

Was this yours? Attribution in diffusion models

02501 Advanced Deep Learning in Computer Vision

Aasa Feragen
Professor, DTU Compute

12 March, 2024

Keywords/lectures: Diffusion models, Attribution

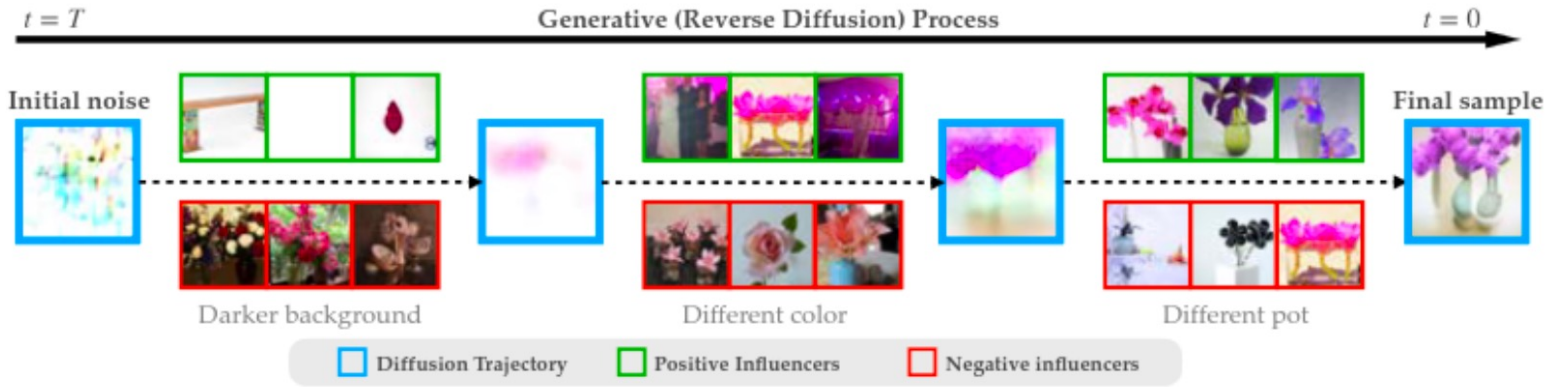


Figure 1: **Overview of our attribution framework.** For a given synthesized image, we apply our attribution method at individual steps along the diffusion trajectory. At each step t, our method pinpoints the training examples with the highest influence (positive in green, negative in red) on the generative process at that step. In particular, positive influencers guide the trajectory towards the final sample, while negative influencers guide the trajectory away from it. We observe that negative influencers increasingly resemble the final sample (the grey text highlights notable differences with the final sample). For more examples, see Appendix B.4.

How do diffusion models learn from underrepresented parts of the population?

02501 Advanced Deep Learning in Computer Vision

Aasa Feragen
Professor, DTU Compute

12 March, 2024

Keywords/lectures: Diffusion models, underrepresentation, bias

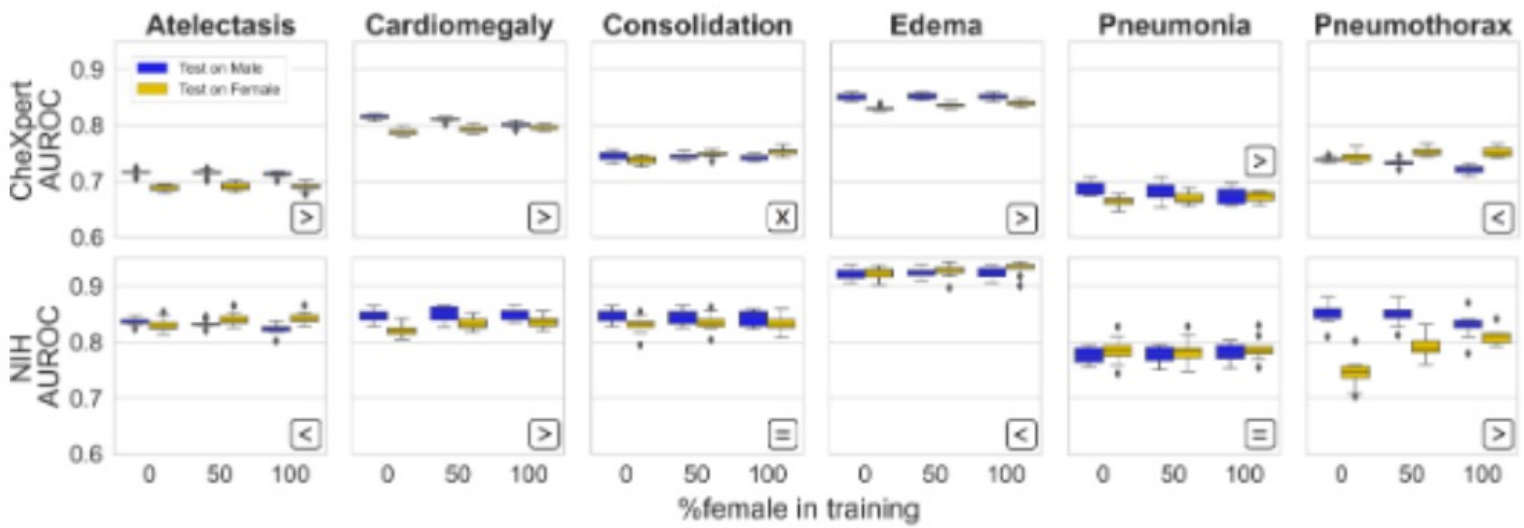


Figure 1: **Underrepresentation bias.** We know that predictive models tend to underperform on underrepresented groups. What about generative models? Figure from [3].

Explaining Hierarchical Vision Transformers

Paraskevas Piegos and Manxi Lin

12 March, 2024

Keywords/lectures: Vision Transformers, Model Explainability

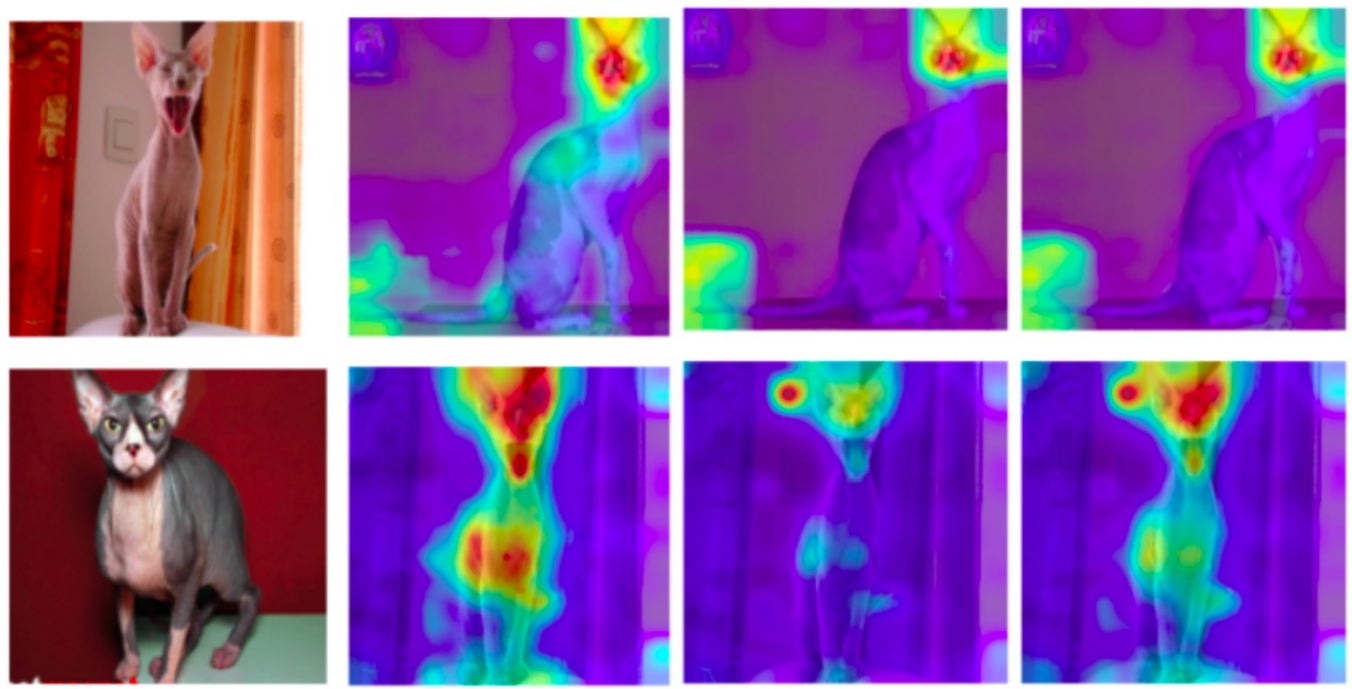


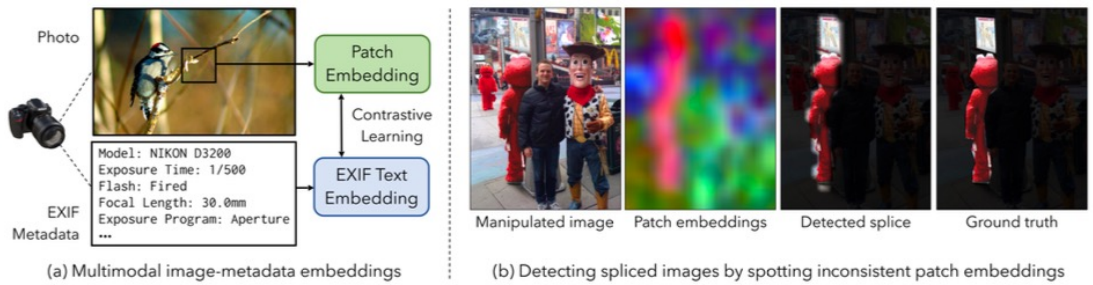
Figure 1: Illustration of what Swin Transformer sees in an image recognition task. The goal of this project is to develop post-hoc explanations for hierarchical vision transformers. The developed approach will be able to give saliency maps of a decision made by a hierarchical vision transformer.

DePhotoshop: Detecting and Reverse Image Manipulations with Exif-CLIP and ControlNet

Manxi Lin

12 March, 2024

Keywords/lectures: CLIP, ControlNet, Transformers, Stable Diffusion

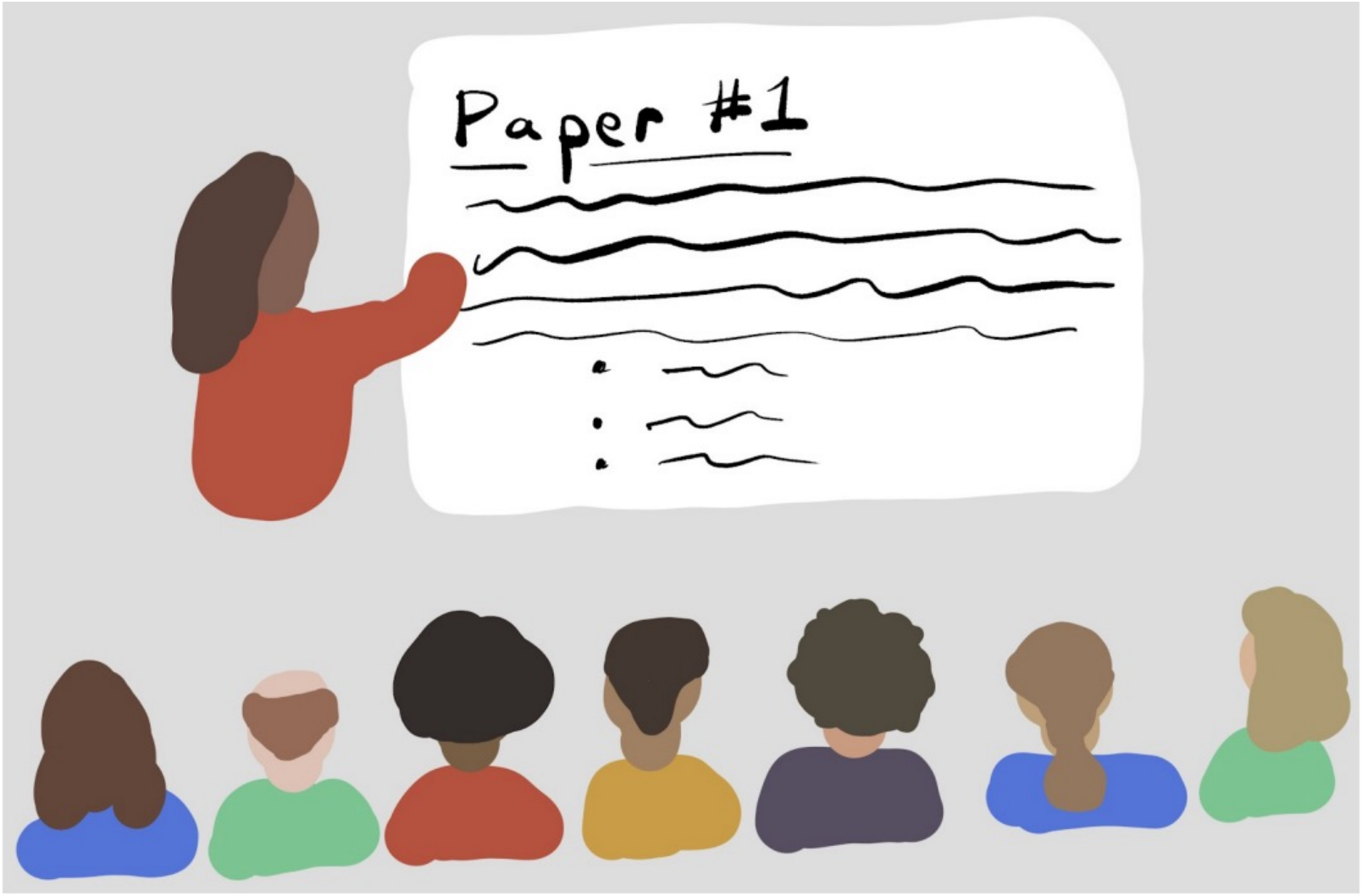


(a)

(b)

Figure 1: (a) Illustration of the manipulation detection process in [5]. (b) An example of image inpainting via Stable Diffusion.

Reading group (One-to-many format)



Reading group (Many-to-many format)

https://www.cs.toronto.edu/~jacobson/images/role-playing-paper-reading-seminars.pdf

Many-to-many format – “Role Playing Game”

Spotlight Presenter

Scientific Peer Reviewer

Archaeologist

Academic Researcher

Social Impact Assessor

https://www.cs.toronto.edu/~jacobson/images/role-playing-paper-reading-seminars.pdf

Spotlight Presenter (“Introduction”)

“On behalf of one of the authors, you are asked to give a spotlight presentation (3 minutes) for the paper at CVPR 2024 in Seattle.”

Scientific Peer Reviewer

"The paper has not been published yet and is currently submitted to a top conference where you've been assigned as a peer reviewer. Complete a full review of the paper answering all prompts of the official review form of the top venue in this research area (e.g., CVPR, ICCV for Computer Vision and NeurIPS for Deep Learning). This includes recommending whether to accept or reject the paper."

Scientific Peer Reviewer

Paper Summary. **In 5-7 sentences, describe the key ideas, experimental or theoretical results, and their significance.**

Strengths. **Consider the significance of key ideas, experimental or theoretical validation, writing quality, data contribution. Explain clearly why these aspects of the paper are valuable. Short bullet lists do NOT suffice.**

Weaknesses. **Consider the significance of key ideas, experimental or theoretical validation, writing quality, data contribution. Clearly explain why these are weak aspects of the paper, e.g. why a specific prior work has already demonstrated the key contributions, or why the experiments are insufficient to validate the claims, etc. Short bullet lists do NOT suffice.**

Paper rating. **Accept, Weak Accept, Borderline, Weak Reject, Reject**

Justification of rating. **What are the most important factors in your rating?**

Scientific Peer Reviewer

Paper Summary. In 5-7 sentences, describe the key ideas, experimental or theoretical results, and their significance.

Strengths. Consider the significance of key ideas, experimental or theoretical validation, writing quality, data contribution. Explain clearly why these aspects of the paper are valuable. Short bullet lists do NOT suffice.

Weaknesses. Consider the significance of key ideas, experimental or theoretical validation, writing quality, data contribution. Clearly explain why these are weak aspects of the paper, e.g. why a specific prior work has already demonstrated the key contributions, or why the experiments are insufficient to validate the claims, etc. Short bullet lists do NOT suffice.

Paper rating. Accept, Weak Accept, Borderline, Weak Reject, Reject

Justification of rating. What are the most important factors in your rating?

Archaeologist

"This paper was found buried under ground in the desert. You're an archaeologist who must determine where this paper sits in the context of previous and subsequent work. Find and report on one older paper cited within the current paper that substantially influenced the current paper and one newer paper that cites this current paper."

Academic Researcher

"You're a researcher who is working on a new project in this area. Propose an imaginary follow-up project not just based on the current but only possible due to the existence and success of the current paper."

Social Impact Assessor

"Identify how this paper self-assesses its (likely positive) impact on the world. Have any additional positive social impacts left out? What are possible negative social impacts that were overlooked or omitted?"

Roles

Paper	Diffusion Transformers	Gemini	ControlNet	Universal Guided Diffusion
Link	https://arxiv.org/abs/2212.09748	https://arxiv.org/abs/2312.11805	https://arxiv.org/abs/2302.05543	https://arxiv.org/abs/2302.07121
Responsible	**Dimitrios**	**Aasa**	**Thanos + Manxi**	**Paraskevas**
Spotlight 1	Andreas Theilgaard - s201715	Martin Lund Trinhammer -s201944	Marie Lunde -s194637	Hans Heje s194365
Spotlight 2	Lucas S - s223093	Matija Sipek - s222736	Calle Carlsen - 204114	
Reviewer 1	Jonathan Mikler - s222962	Christian Kjær - s211469	Erik Riise - S194633	Eleftherios Katiforis s222725
Reviewer 2	Esteban Zamora s222383	Shah Bekhsh	Jeremi Ledwon - s232952	Jacob Tuxen - s194572
Reviewer 3				
Archaeologist 1	Abdulrahman Ramadan s205576	Adrià Bosch Matas - s232775	Viktor Sebastian Petersen - s204225	András Bence Schin - s233084
Archaeologist 2	Andreas Fiehn - s204125	Esther Conradsen s193181	István László Mádi - s232971	Moamal Abdul-Mahdi s195398
Archaeologist 3		Michael Harborg - s204138		
Academic Researcher 1	Kaixuan Lu - s232248	Sebastian Ake Aaen	Jens Parslov- s194649	Aleksandra Ludwiniak s223216
Academic Researcher 2	Elle McFarlane - s222376	Alexandra Inselmann s194368	Ioanna Gemou - s233095	Katja Valérie Bonvin - s233494
Academic Researcher 3	Oliver Elmgreen s204070	Christian Ole Nielsen - s204131		
Social Impact Assessor 1	Rolando Esquivel s230025	Benjamin Jønch Jürgensen - s236771	Johannes Roed - s194648	Mikkel - s205421
Social Impact Assessor 2	Simon Daniel Eiriksson - s180722	Àiax Faura Vilalta s222703	Pedro Vieira - 240181	Boldizsár Elek - s233085
Social Impact Assessor 3		Luca D'Este - s233231		

Schedule

13:00-13:40 Project Proposals (Aasa, Morten, Dimitrios)

13:40-13:50 Break

13:50-17:00 Reading group

13:50-13:55 Introduce and assign roles
13:55-14:25 Do the role!
14:25-14:40 Discuss with your role-partner
14:40-14:50 Break
14:50-15:50 Discuss with your paper group (with Paper Responsible)
15:50-16:00 Finalize slides
16:00-16:10 Break
16:10-17:00 Present slides (4papers*4roles*3minutes)