Friedrich-Alexander-Universität
Technische Fakultät

FAU

Project Part 2

Computational Visual Perception (CompVP)

Bernhard Egger, Andreas Kist, Patrick Krauß, Tim Weyrich

How to complete module

- 7,5 ECTS
  - 5+2,5 ECTS
  - You need both!
  - You can’t get only 5 or only 2,5 ECTS

Overall project goal

- How much of vision do SotA generative video models “solve”?
- How well do they and other models work for corner cases of vision?

Paper 1 by October 27th :

Google DeepMind

2025-10-1

Video models are zero-shot learners and reasoners

Thaddäus Wiedemer[*1], Yuxuan Li[1], Paul Vicol[1], Shixiang Shane Gu[1], Nick Matarese[1], Kevin Swersky[1], Been Kim[1], Priyank Jaini[*1] and Robert Geirhos[*1]
[1]Google DeepMind

The remarkable zero-shot capabilities of Large Language Models (LLMs) have propelled natural language processing from task-specific models to unified, generalist foundation models. This transformation emerged from simple primitives: large, generative models trained on web-scale data. Curiously, the same primitives apply to today's generative video models. Could video models be on a trajectory towards general-purpose *vision* understanding, much like LLMs developed general-purpose *language* understanding? We demonstrate that Veo 3 can solve a broad variety of tasks it wasn't explicitly trained for: segmenting objects, detecting edges, editing images, understanding physical properties, recognizing object affordances, simulating tool use, and more. These abilities to perceive, model, and manipulate the visual world enable early forms of visual reasoning like maze and symmetry solving. Veo's emergent zero-shot capabilities indicate that video models are on a path to becoming unified, generalist vision foundation models.
Project page: https://video-zero-shot.github.io/

29 Sep 2025

1. Introduction

Paper 2 by November 27th :

frontiers in Psychology

ORIGINAL RESEARCH
published: 15 March 2018
doi: 10.3389/fpsyg.2018.00345

Check for updates

Illusory Motion Reproduced by Deep Neural Networks Trained for Prediction

Eiji Watanabe[1,2], Akiyoshi Kitaoka[3], Kiwako Sakamoto[4,5], Masaki Yasugi[1] and Kenta Tanaka[6]*

[1] Laboratory of Neurophysiology, National Institute for Basic Biology, Okazaki, Japan, [2] Department of Basic Biology, The Graduate University for Advanced Studies (SOKENDAI), Miura, Japan, [3] Department of Psychology, Ritsumeikan University, Kyoto, Japan, [4] Department of Physiological Sciences, The Graduate University for Advanced Studies (SOKENDAI), Miura, Japan, [5] Division of Integrative Physiology, National Institute for Physiological Sciences (NIPS), Okazaki, Japan, [6] Sakura Research Office, Wako, Japan

The cerebral cortex predicts visual motion to adapt human behavior to surrounding objects moving in real time. Although the underlying mechanisms are still unknown, predictive coding is one of the leading theories. Predictive coding assumes that the brain's internal models (which are acquired through learning) predict the visual world at all times and that errors between the prediction and the actual sensory input further refine

Structure of project

- Project Part 1: Play around with video diffusion models and reproduce paper
  (close to Paper 1)

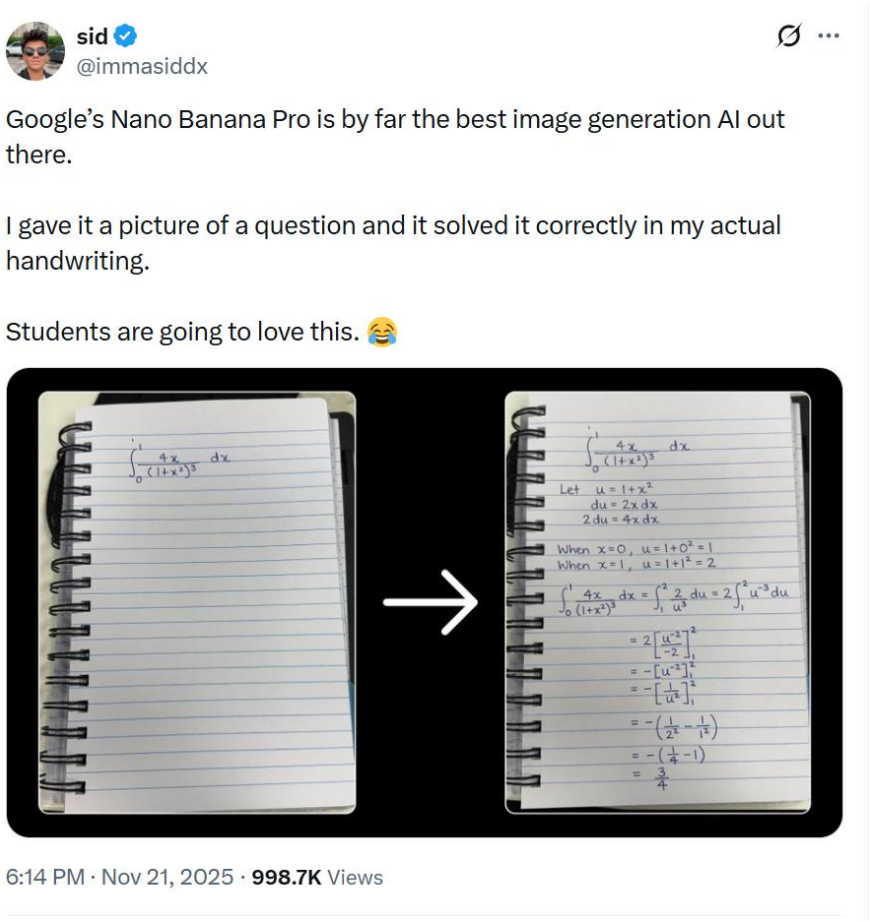
- Project Part 2: Experiment further with illusions, optical flow and depth estimation
  (close to paper 2)

- Project Part 3: Further Evaluation based on results of Part 1 and Part 2
  (close to paper 1 and 2)

What happened since?

What happened since?

What happened since?



What happened since?

What happened since?

What happened since?

frontiers in Psychology

ORIGINAL RESEARCH
published: 15 March 2018
doi: 10.3389/fpsyg.2018.00345

Check for updates

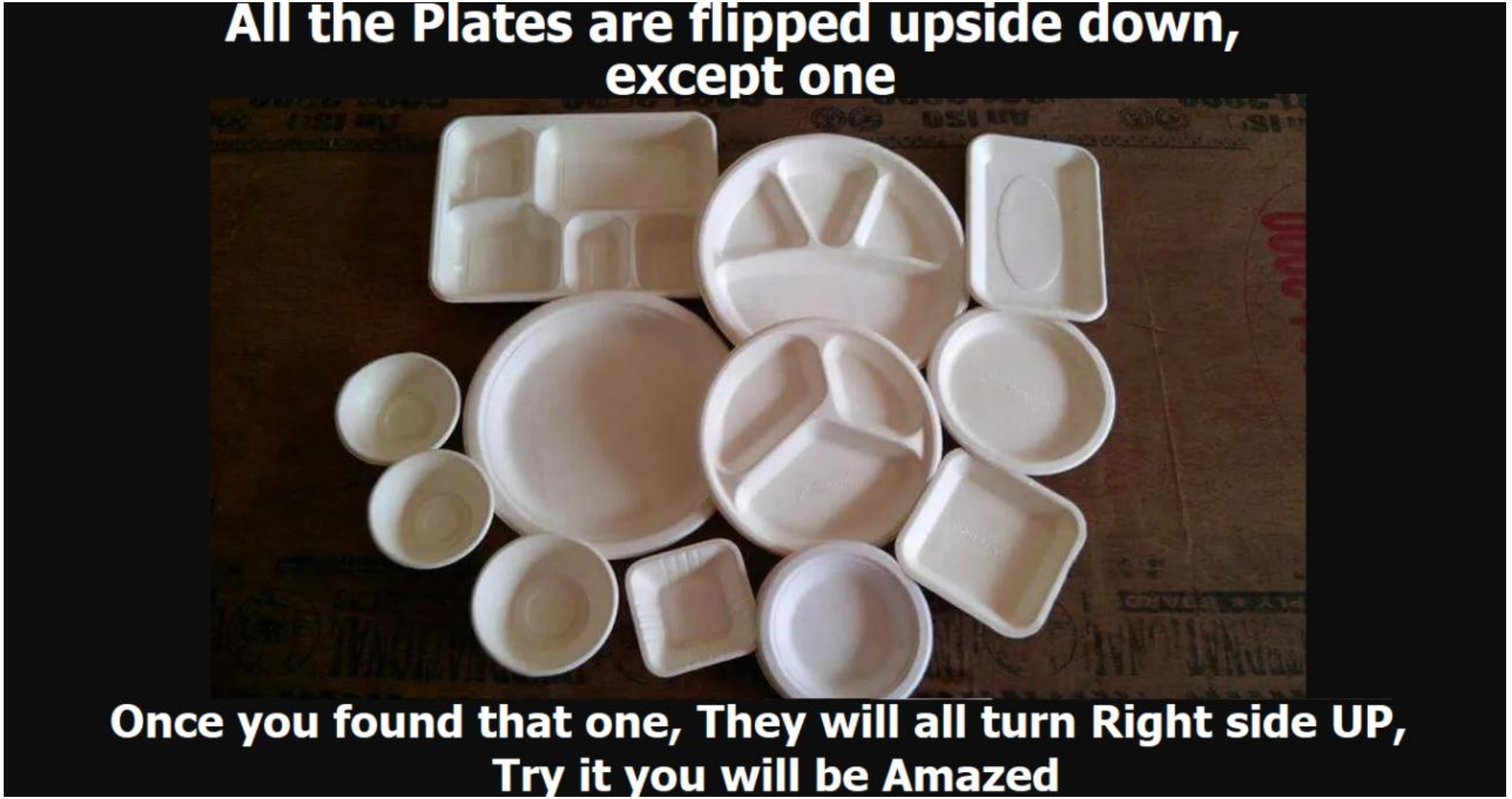
Illusory Motion Reproduced by Deep Neural Networks Trained for Prediction

Eiji Watanabe[1,2*], *Akiyoshi Kitaoka*[3], *Kiwako Sakamoto*[4,5], *Masaki Yasugi*[1] and *Kenta Tanaka*[6]

[1] *Laboratory of Neurophysiology, National Institute for Basic Biology, Okazaki, Japan,* [2] *Department of Basic Biology, The Graduate University for Advanced Studies (SOKENDAI), Miura, Japan,* [3] *Department of Psychology, Ritsumeikan University, Kyoto, Japan,* [4] *Department of Physiological Sciences, The Graduate University for Advanced Studies (SOKENDAI), Miura, Japan,* [5] *Division of Integrative Physiology, National Institute for Physiological Sciences (NIPS), Okazaki, Japan,* [6] *Sakura Research Office, Wako, Japan*

The cerebral cortex predicts visual motion to adapt human behavior to surrounding objects moving in real time. Although the underlying mechanisms are still unknown, predictive coding is one of the leading theories. Predictive coding assumes that the brain's internal models (which are acquired through learning) predict the visual world at all times and that errors between the prediction and the actual sensory input further refine

What happened since?



What happened since?

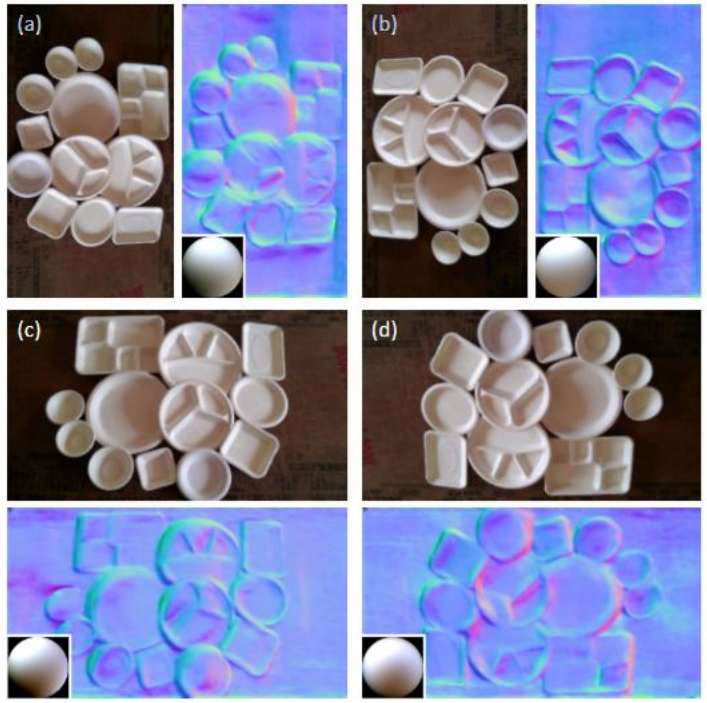


Fig. 16: Shape estimation results on the plates shading illusion. For four orientations of the image (a-d) we show the input image and the estimated normal map with estimated lighting inset.

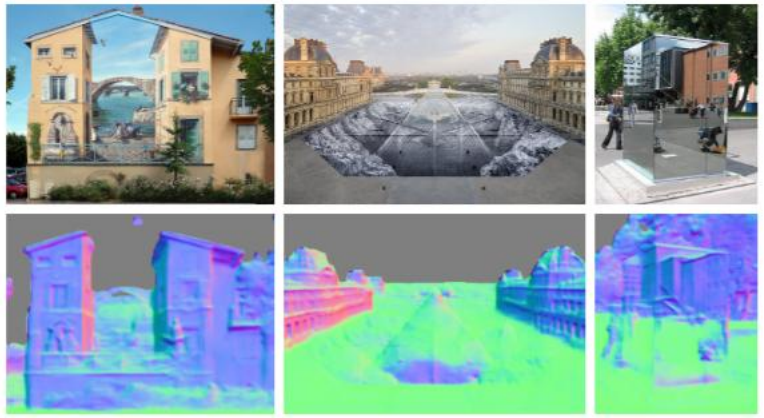
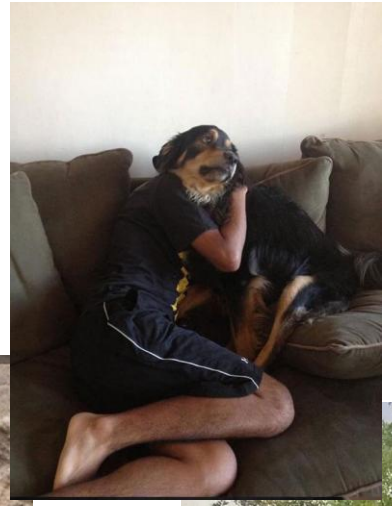


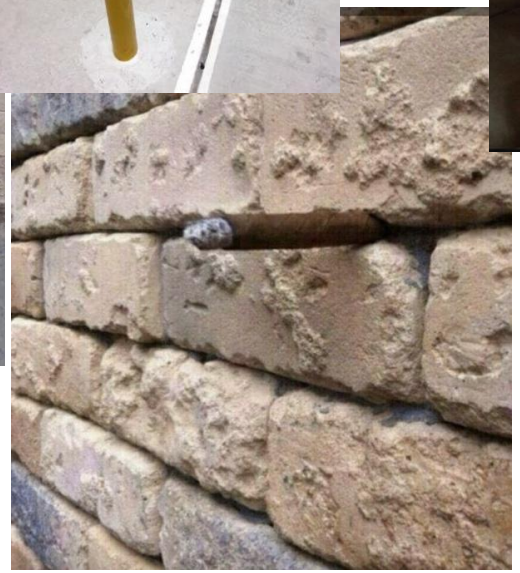
Fig. 18: Shape estimation in the presence of Trompe-l'œil and mirror illusions.

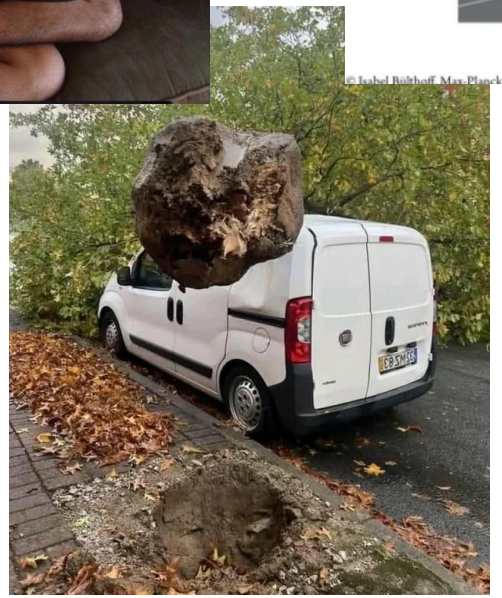
l'œil illusions and one for a mirror illusion. Trompe-l'œil use albedo variations (i.e. painted surface texture) to depict 3D scenes, giving the illusion of additional geometric variation. A mirror illusion is similar except that the albedo variation is a reflection of another part of the real scene. In all examples, the shape reconstruction is fooled by the illusion. In the example in the first column, both the sky segmentation network and our inverse rendering network are fooled. The shape of the painted windows, doors and

https://arxiv.org/pdf/2102.06591

What happened since?







https://twitter.com/PhotoInRealLife
https://www.realtimerendering.com/realartifacts/
https://bsky.app/profile/chrisoffner3d.bsky.social/post/3lcam7loiek2w

What happened since?

Akiyoshi's illusion pages

Akiyoshi KITAOKA, Professor, Psychology, Ritsumeikan University, Osaka, Japan
studying visual perception, visual illusion, optical illusion, trompe l'oeil, 3D, etc. ORCID

Since May 18, 2002; Updated November 23, 2025 Japanese, Serb, Portuguese, Chinese; Illusion calendar 2024

This work was partly supported by JSPS KAKENHI Grant Number 21H04426 awarded to Akiyoshi Kitaoka.

Warning: *Commercial abuse of my illusion images is prohibited. This page contains some works of "anomalous motion illusion", which might make sensitive observers dizzy or sick. Should you feel dizzy, you had better leave this page immediately.* More

Latest works (November 23, 2025) --- Newest page (November 23, 2025) --- Updated page (August 16, 2022) --- Illusion catalogue (June 3, 2014)
Page list of this site --- Books (September 23, 2019) --- Papers (February 11, 2021) --- JavaScript programs (October 13, 2024)

Journal of Illusion

The Journal of Illusion welcomes your submissions.

https://www.ritsumei.ac.jp/~akitaoka/index-e.html

How to pass

- Submission via Studon course:
  “Computational Visual Perception”

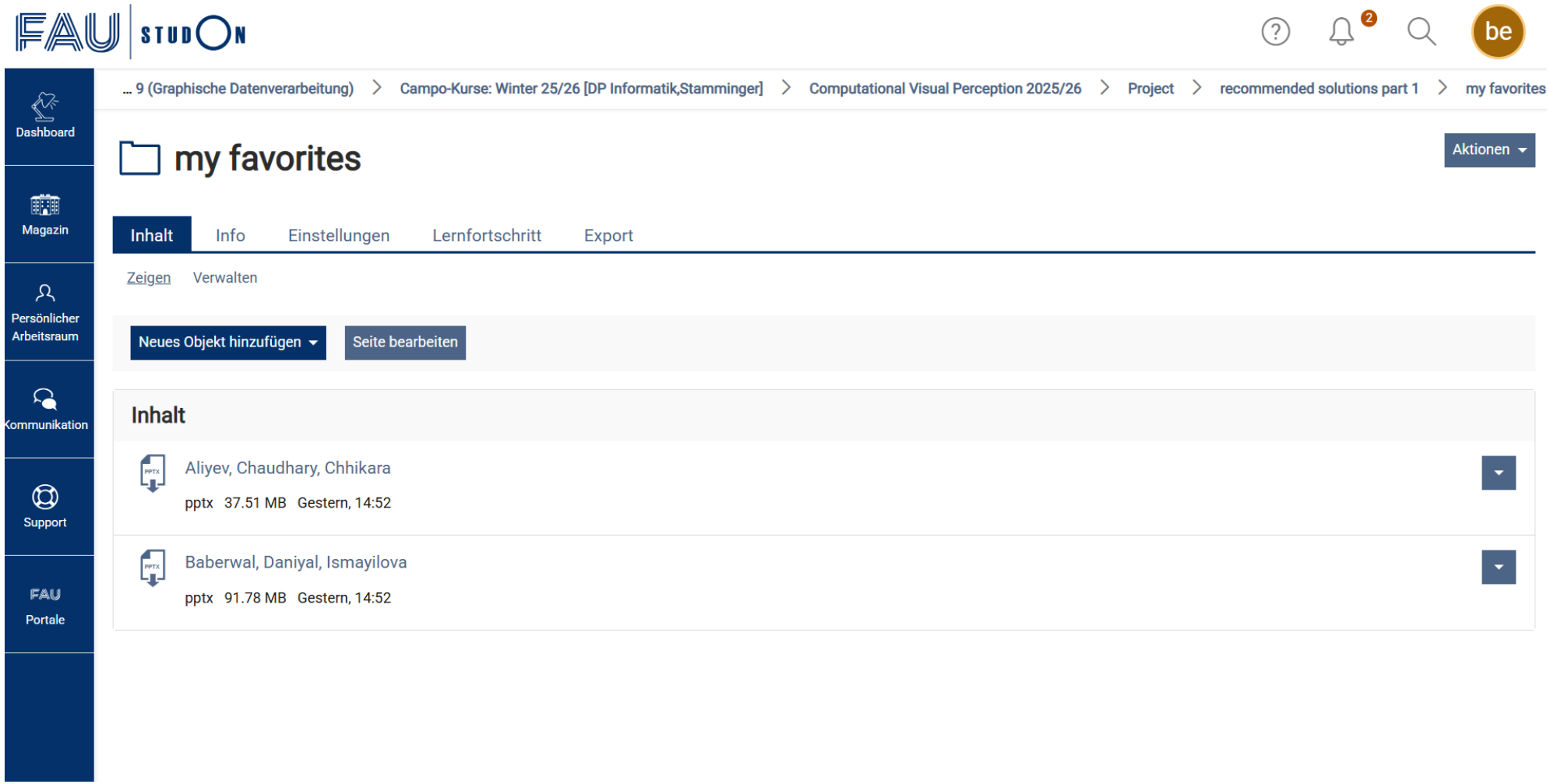
- Submission has to be in the exact format

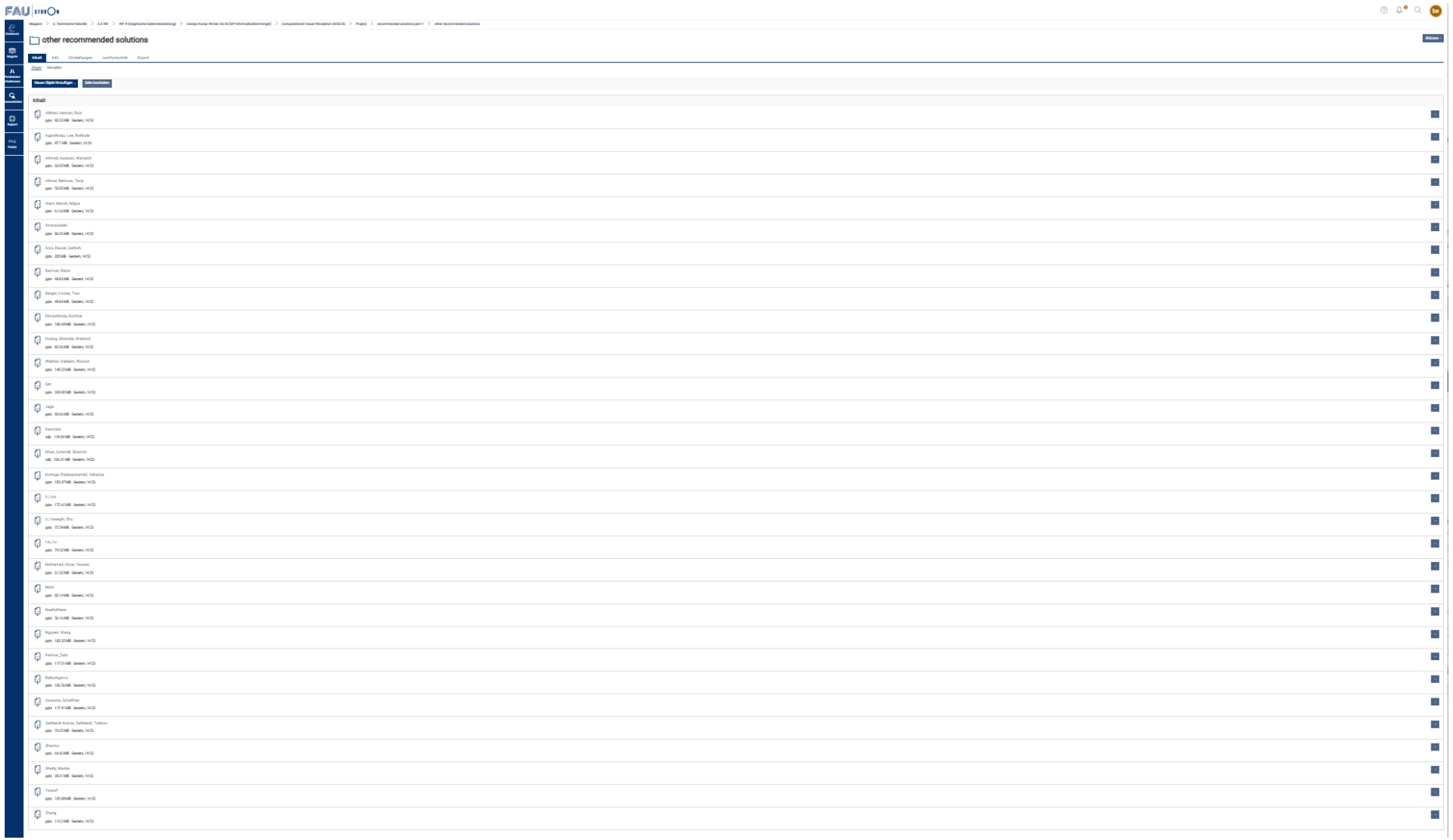
- Three strict project deadlines
  - November 21st,
  - January 2nd (feel free to submit early),
  - February 5th

How to pass

- Pass/Fail for each of the 3 parts,
- You need to pass 2 out of 3 parts

- The best solutions for each part will be released ~ 1 week after the deadline, to enable others to continue with the best solution of another team


FAU STUDON
be
Dashboard
Magazin
Persönlicher Arbeitsraum
Kommunikation
Support
FAU Portale
... 9 (Graphische Datenverarbeitung) > Campo-Kurse: Winter 25/26 [DP Informatik,Stamminger] > Computational Visual Perception 2025/26 > Project > recommended solutions part 1 > my favorites
my favorites
Aktionen
Inhalt
Info
Einstellungen
Lernfortschritt
Export
Zeigen
Verwalten
Neues Objekt hinzufügen
Seite bearbeiten
Inhalt
Aliyev, Chaudhary, Chhikara
pptx 37.51 MB Gestern, 14:52
Baberwal, Daniyal, Ismayilova
pptx 91.78 MB Gestern, 14:52


other recommended solutions
Inhalt

How to pass

- Scope of project ~ 150 hours per student
- Teams of 1-3 students
- Steps can be performed in new group

- If you are looking for a group, please stay after the class and talk to people who also stay
- If you are looking for a group and can only join virtually, please use the forum in “Computational Visual Perception” to team up

- Finding a group is your responsibility

Part 2 Options

- Play around with:
  - Image generation (based on input image, e.g. nanobanana)
  - Video generation (similar to Part1)
  - Segmentation (e.g. SAM3)
  - 3D Reconstruction (e.g. SAM3D or Gen3DSR)
  - 3D Reconstruction bodies (e.g. SAM3D or PLIKS))
  - Optical Flow (e.g. FLOWSEEK)
  - Depth/Normal Estimation (e.g. Marigold 2, DepthAnything)

  - Choose ≥ 2 and play around
  - Come up with your own research question(s)
  - Run it on your own (not just cloud-ui)

Part 2 deliverables

- Per project team 1 single pptx file (or powerpoint compatible)
- The pptx file contains:
  - Option 1 Formulated Research Question(s)
  - Experiments for your chosen option 1 for at least 5 images
  - 1 slide summarizing what you observed for your option 1
  - Option 2 Formulated Research Question(s)
  - Experiments for your chosen option 2 for at least 5 images
  - 1 slide summarizing what you observed for your option 2
  - Screenshot for both option 1 and option 2 how your interface works (e.g. terminal)

- (ideally all videos are set to play automatically)

Part 2 Upload

- In Studon course :
  “Computational Visual Perception”
- You will upload up to ~1GB (don’t!)
  Plan in internet speed, upload at university

  If you run into issues uploading, you send an md5 hash of your zip file **before** the deadline and you provide an alternative download link within 24h

Part 2 grading

- Correct format
- Option 1
- Formulated Research Question(s) for Part 1
- Summary text 1
- Option 2
- Formulated Research Question(s) for Part 2
- Summary text 2
- Proof for own interface

- Selected solution: all points fulfilled to full satisfaction
- Pass: at most one bullet point from the above missing

- Plagiarism will have serious consequences

Structure of project

- Project Part 1: Play around with video diffusion models and reproduce paper
  (close to Paper 1)

- Project Part 2: Experiment further with illusions, optical flow and depth estimation
  (close to paper 2)

- Project Part 3: Further Evaluation based on results of Part 1 and Part 2
  (close to paper 1 and 2)

Project Consulting

- You can ask questions in the forum
  “Computational Visual Perception”

- You come with concrete questions
- I’ll open a thread in the forum, where you can respond till Thursday each week if you want to meet
- I’ll distribute time slots each Friday

- No guarantee for any responses on the day of the deadline

How to complete module

- 7,5 ECTS
  - 5+2,5 ECTS
  - You need both!
  - You can’t get only 5 or only 2,5 ECTS

Don’t start late