ScaDS.AI DRESDEN LEIPZIG

CENTER FOR SCALABLE DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

Bio-Image Data Science

Robert Haase

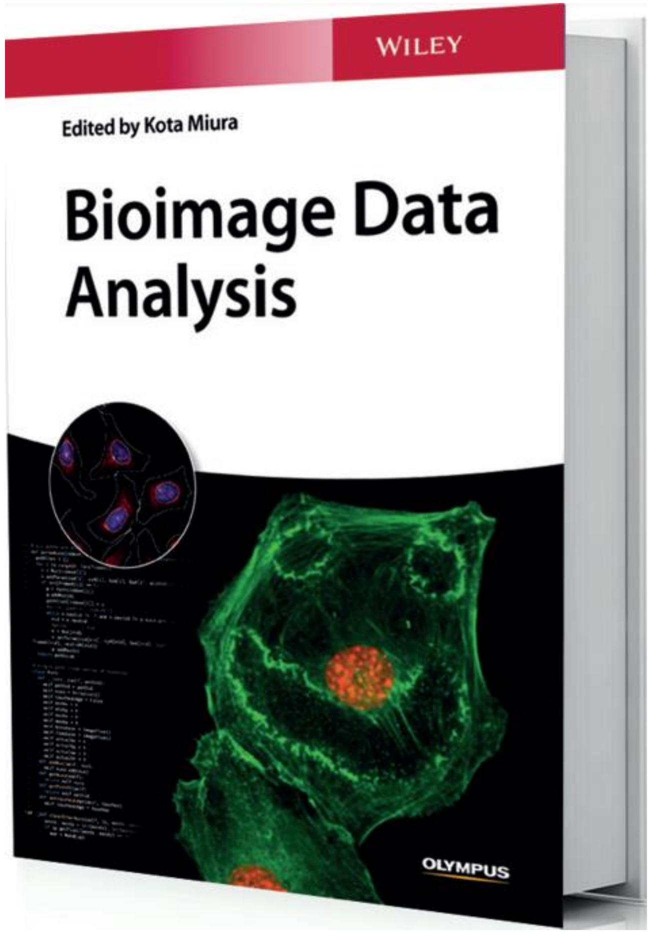
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

SACHSEN
Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Bioimage Analysis

- Kota Miura & Sebastien Tosi 2015:

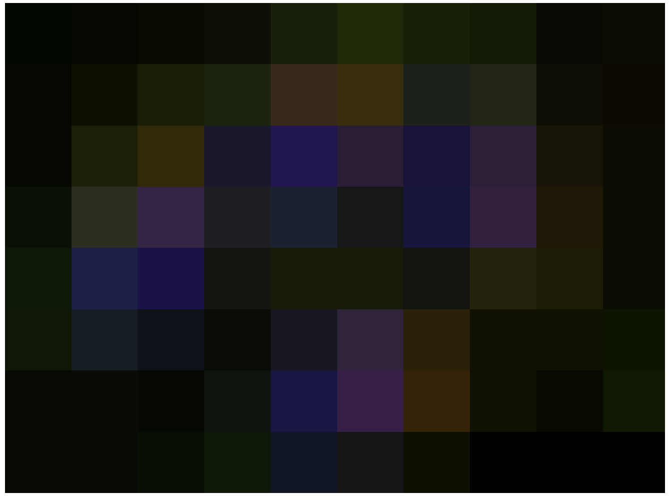
> In the light of this definition, image analysis, which is also called "computer vision," aims at mimicking the way we see the world and how we identify its visible structures. Image analysis in biology does undeniably also hold this element, but more importantly, its main goal is to *measure* biological structures and phenomena in order to study and understand biological systems in a quantitative way.
>
> To achieve this task, we in fact do not have to be bothered with similarity to the human recognition – we have more emphasis on the objectivity of quantitative measurement, rather than how that computer-based recognition becomes in agreement with human recognition. Therefore, in biology, image analysis is a process of identifying spatial distribution of biological components in images and measuring their characteristics to study their underlying mechanisms in an unbiased way. To underline this difference in the goals of image analysis in the two fields and to distinguish them from each other, we will now on refer to image analysis in biology as *bioimage analysis.*

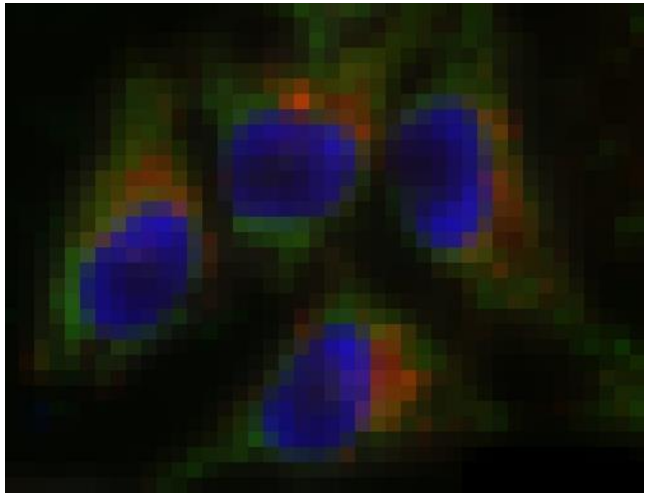


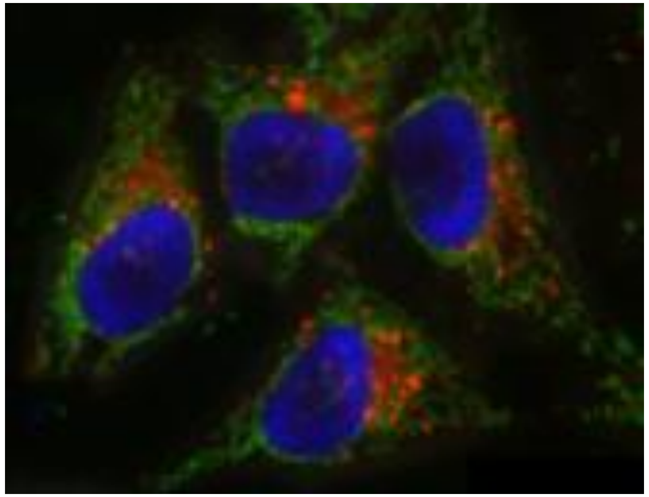
Source: Bioimage Data Analysis, First Edition. Edited by Kota Miura. © 2016 Wiley-VCH Verlag GmbH & Co. KGaA. Published 2016 by Wiley-VCH Verlag GmbH & Co. KGaA
https://analyticalscience.wiley.com/do/10.1002/was.00050003/full/bioimagedataanalysis.pdf

Pixel size versus resolution

- Pixel size is a property of a digital image.
- You configure it during the imaging session at the microscope.


Pixel size: 3.3 µm

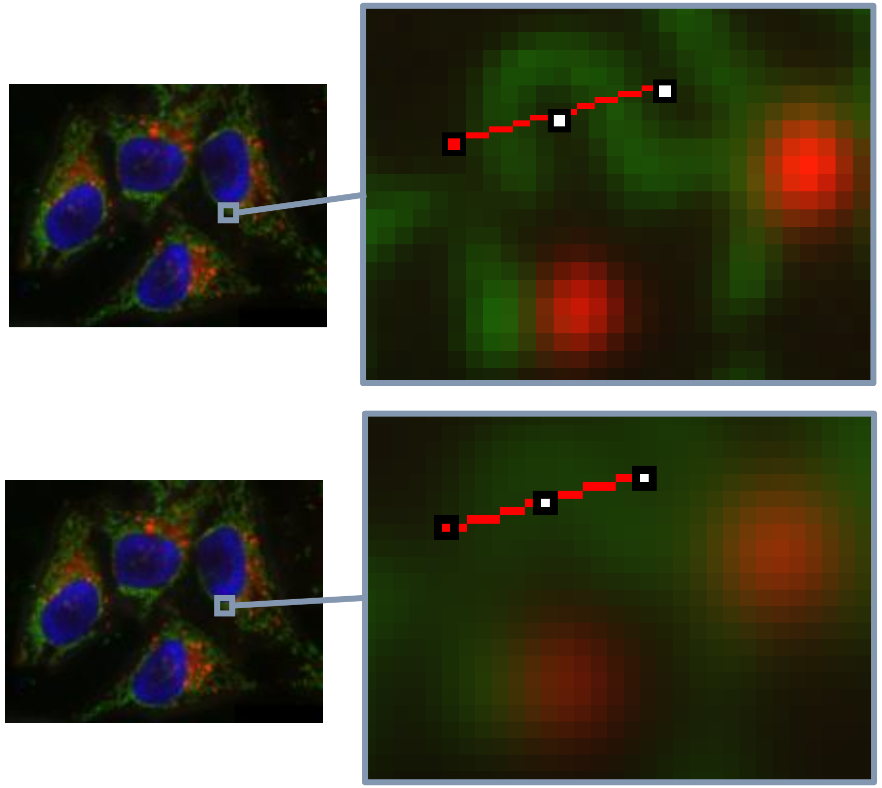

Pixel size: 0.8 µm

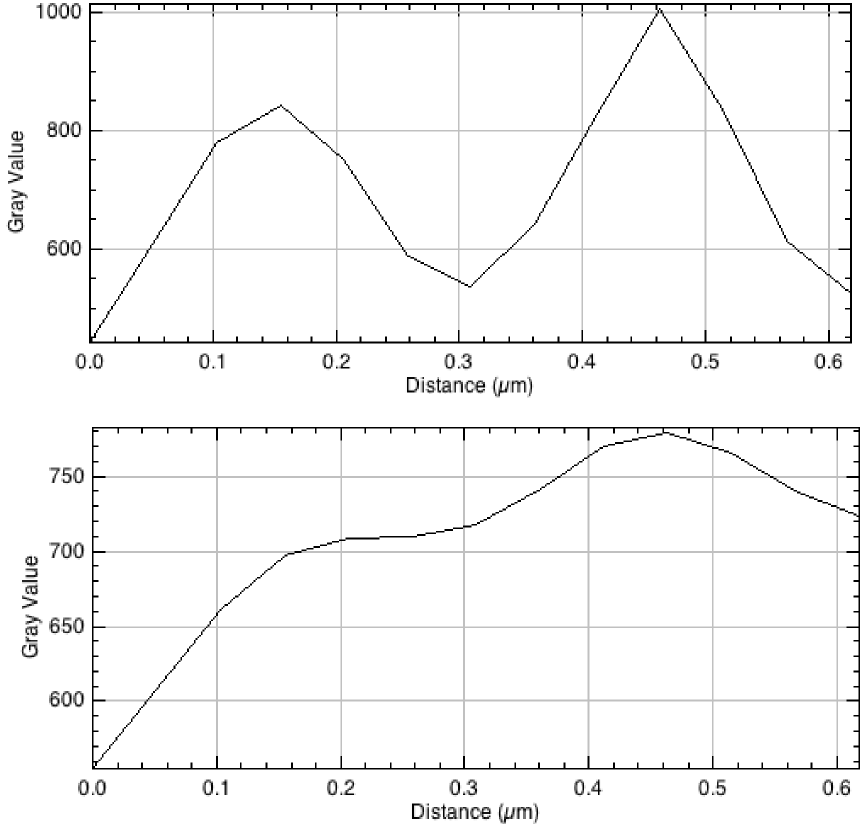

Pixel size: 0.05 µm

- We are not talking about resolution!

Pixel size versus resolution

- Resolution is a property of your imaging system.
- How small can objects be, to be still differentiable?





Quiz

- Enabling others to do your experiment is about …

Repeatability

Reproducibility

Replicability

Reliability

Quiz

- Reproducibility can be achieved by

Writing documentation

Writing code

Providing example data

Recording Video tutorials

Bio-image analysis is supposed to be

- **Quantitative**
  - We derive numbers from images which describe physical properties of the observed sample.
- **Objective**
  - The derived measurement does not depend on who did the measurement. The measurement is free of interpretation.
- **Reliable (trustworthy / validated)**
  - We are confident that the measurement is describing what it is supposed to describe.
- **Reproducible**
  - Enabling others to re-do the experiment. For this, documentation is crucial!
- **Replicability**
  - Others *do* execute the same analysis, potentially on other data, and see consistent results.
- **Repeatable**
  - We can do the same experiment twice under the *same conditions* and get the same measurements.

ScaDS.AI DRESDEN LEIPZIG
CENTER FOR SCALABLE DATA ANALYTICS
AND ARTIFICIAL INTELLIGENCE

Research Data Management

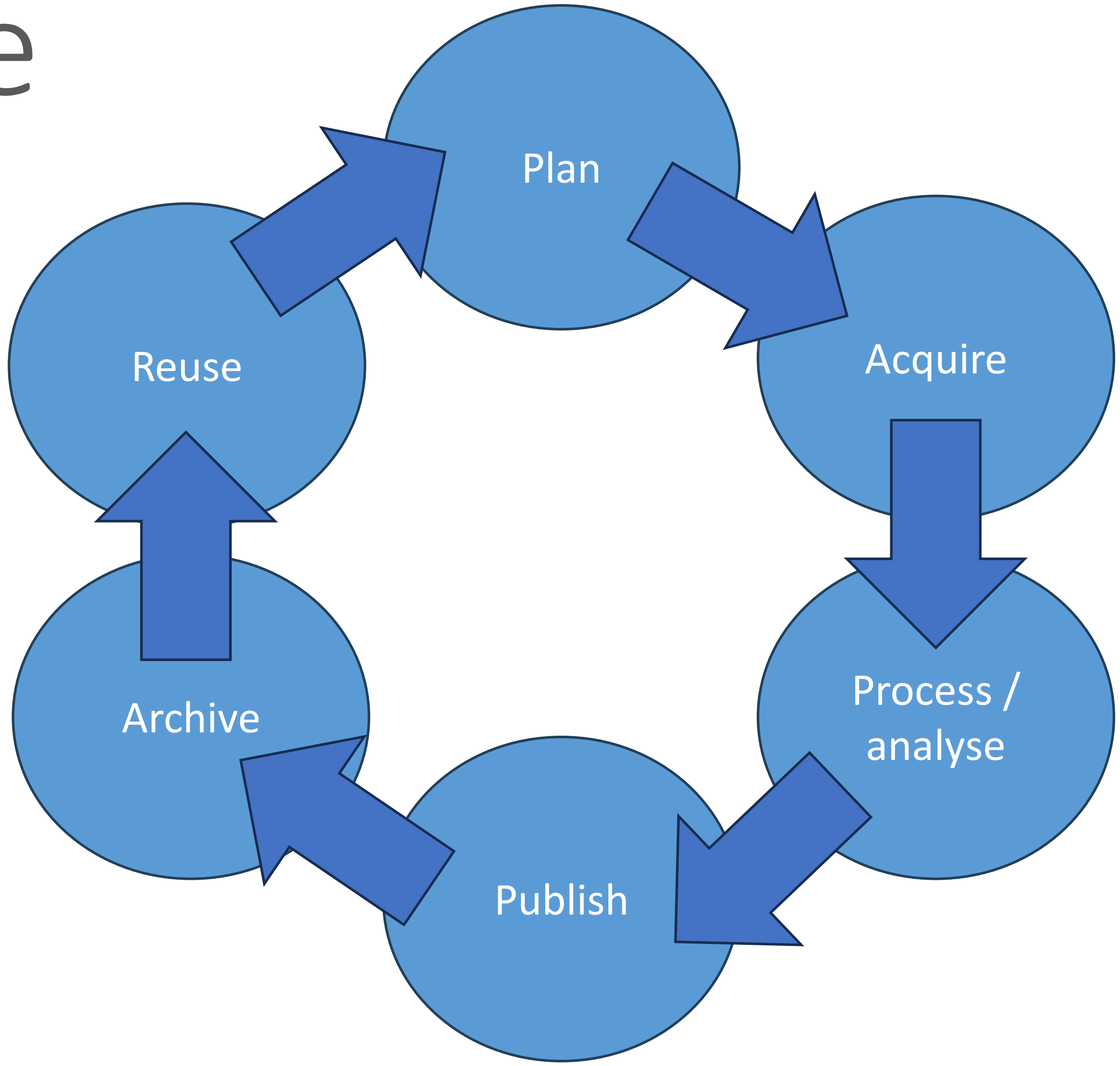
Robert Haase

GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

SACHSEN
Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

RDM Life Cycle

- Processes are ideally cyclic



Adaptiert von:
https://forschungsdaten.info/themen/informieren-und-planen/datenlebenszyklus/

Data Management Plans (DMPs)

- Describes the IS-state of a data environment
  - Which data is acquired / processed?
    (content, format, amount)
  - What meta-data is collected?
  - Which quality standards are targeted?
  - How is data saved, archived, backed-up, shared, published…?
  - Who is responsible for what?
    - Roles, job-profiles
  - What does this all cost?
    (IT infrastructure + human resources)

Survey

Think about the FAIR principles for data sharing, which one is wrong?

Findable

Accessible

Interoperable

Reproducible

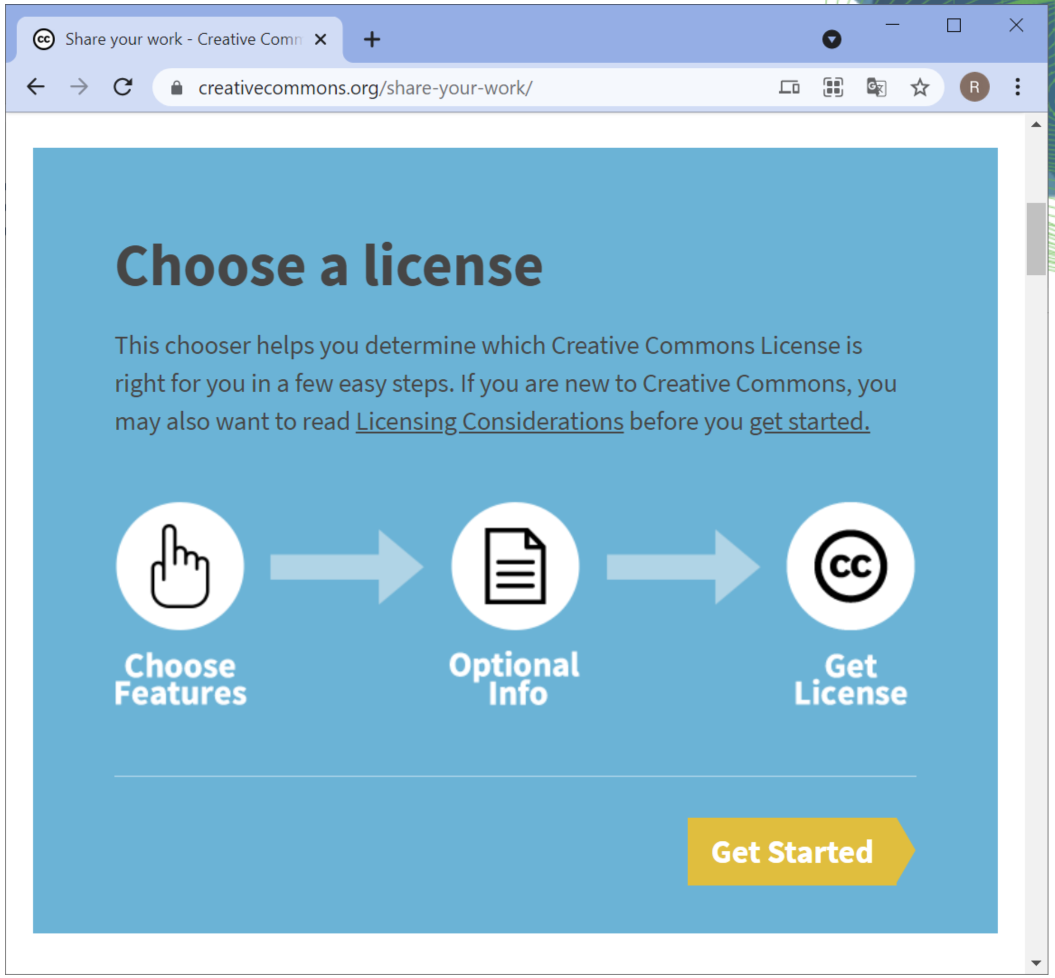
Standard for sharing: The FAIR-principles

- Findable
- Accessible
- Interoperable
- Reusable

Licensing: Creative Commons (CC)

- Public domain (CC0)
- Attribution International (CC-BY)
- Attribution ShareAlike Int. (CC-BY-SA)
- Attribution Non-Commercial Int. (CC-BY-NC)
- Attribution NoDerivatives Int. (CC-BY-ND)

+ Combinations, e.g. CC-BY-NC-ND



Content on this site is licensed under a Creative Commons Attribution 4.0 International license. Icons by The Noun Project.

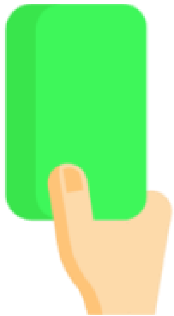
https://creativecommons.org/share-your-work/
https://creativecommons.org/about/cclicenses/

Quiz (recap)

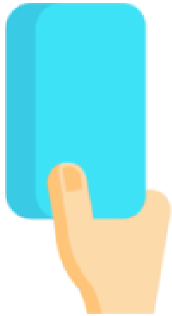
- If I combine two works licensed CC-BY and CC-BY-SA, what license do I have to use?

CC-BY

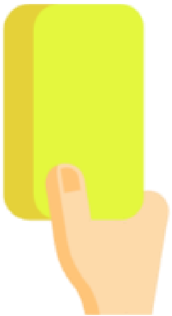
CC-BY-SA



CC-BY-ND



(not possible)

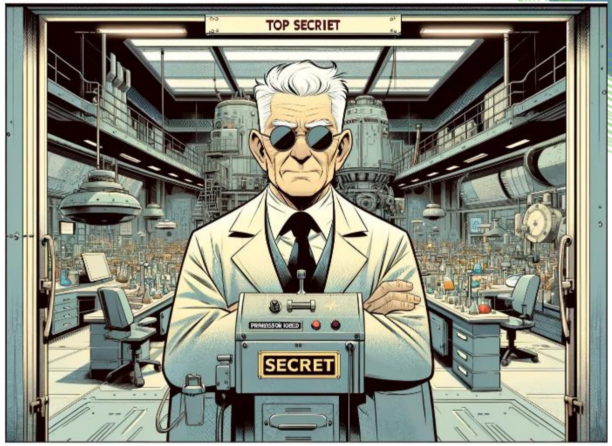


Licensing: Permissive versus restrictive

- Restrictive
  - You can reuse our stuff, but only if you …
    - License your work with the same license we do
    - Make your stuff openly available
    - Make no money with derivatives of our work
  - Examples: GPL, CC-BY-SA, CC-BY-NC, CC-BY-ND

I conclude, these are less *open* in a sense

- Permissive licensing:
  - Do whatever you like with our stuff, just make sure to mention / cite us …
  - Examples: BSD, MIT, Apache, CC-BY



Survey

Which open-source license might be the least popular in companies?

GPL

BSD

MIT

Apache

ScaDS.AI
DRESDEN LEIPZIG
CENTER FOR SCALABLE DATA ANALYTICS
AND ARTIFICIAL INTELLIGENCE

Research Software Management

Robert Haase

GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

SACHSEN
Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Documenting dependencies

- Maintain a document with the dependencies (and versions) you need in your project!
  - The conda way
  - The pip way

environment.yml - Notepad

File Edit Format View Help

```
name: cellpose-test
channels:
  - conda-forge
  - pytorch
dependencies:
  - python=3.8
  - pytorch=1.12
  - devbio-napari
  - cellpose-napari
```

Ln 3, C 100% Windows (CRLF) UTF-8

```
conda env create -f environment.yml
```

In case your environment is screwed up, you can rebuild it any time.

requirements.txt - Notepad

File Edit Format View Help

```
numpy>=1.21
andas~=1.4.0
napari~=0.4.11
matplotlib~=3.5.0
scikit-image~=0.19
umap-learn~=0.5.1
hdbscan~=0.8.27
scikit-learn~=1.0
numba~=0.55.1
```

Ln 2, 100% Windows (CRLF) UTF-8

```
pip install -r requirements.txt
```

Quiz

```
conda install package_a
```

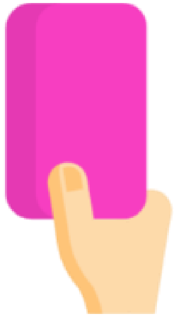
Depends on:
numpy<1.22.0

```
pip install package_b
```

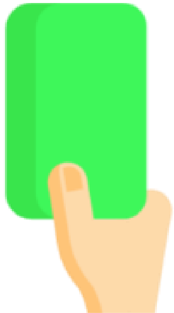
Depends on:
numpy>1.22.0

Dependency conflict

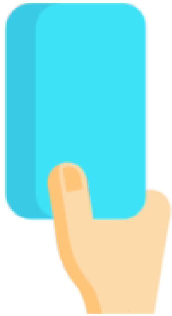
fails



works



works but…



ScaDS.AI
DRESDEN LEIPZIG
CENTER FOR SCALABLE DATA ANALYTICS
AND ARTIFICIAL INTELLIGENCE

Image Processing

Robert Haase

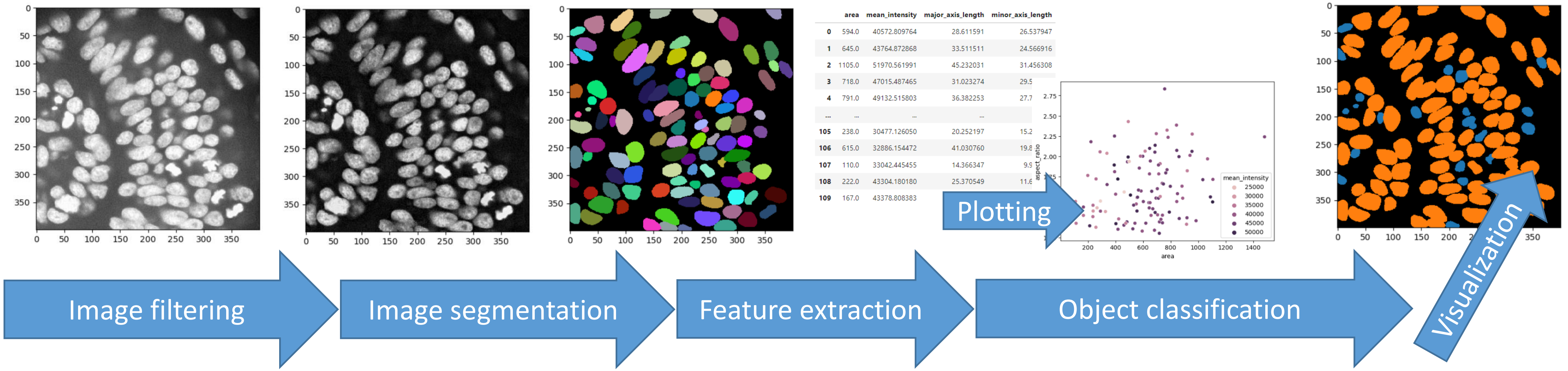
Reusing materials from Mauricio Rocha Martins (Norden lab, MPI CBG); Dominic Waithe (Oxford University); Alex Bird, Dan White (MPI CBG)

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

SACHSEN
Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Lecture overview: Bio-image Analysis

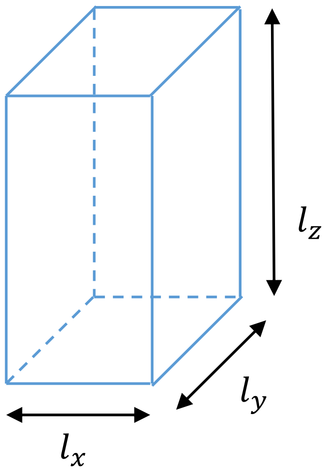
- Image Data Analysis workflows
- Goal: **Quantify observations, substantiate conclusions with numbers**



	area	mean_intensity	major_axis_length	minor_axis_length
0	594.0	40572.809764	28.611591	26.537947
1	645.0	43764.872868	33.511511	24.566916
2	1105.0	51970.561991	45.232031	31.456308
3	718.0	47015.487465	31.023274	29.5
4	791.0	49132.515803	36.382253	27.7
...	...	...	...	
105	238.0	30477.126050	20.252197	15.2
106	615.0	32886.154472	41.030760	19.8
107	110.0	33042.445455	14.366347	9.9
108	222.0	43304.180180	25.370549	11.6
109	167.0	43378.808383		

Image stacks and voxels

- 3-dimensional images consisting of voxels
- "Image stack"
- Often *anisotropic* (not equally large in all directions)



$$l_x = l_y \neq l_z$$

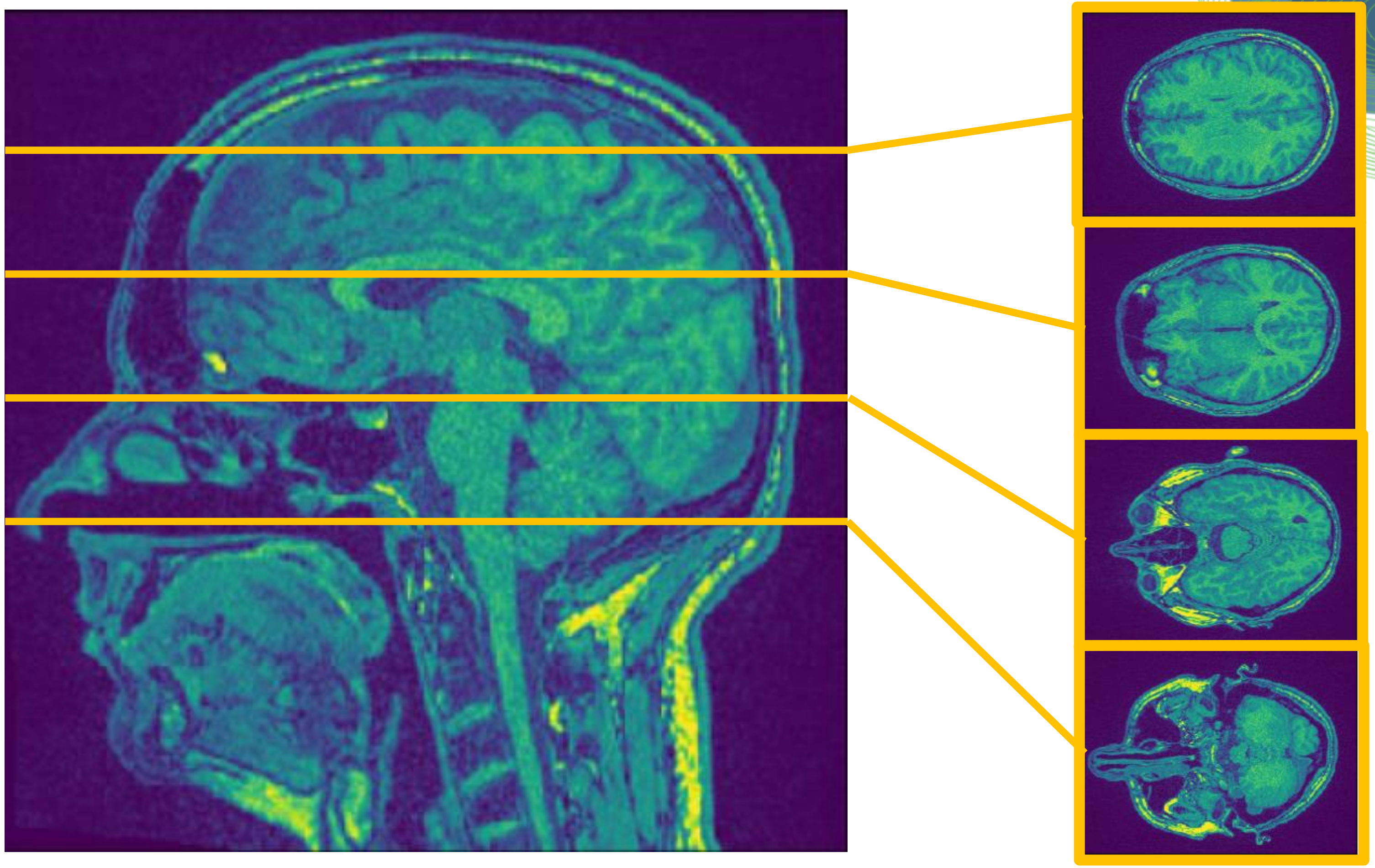
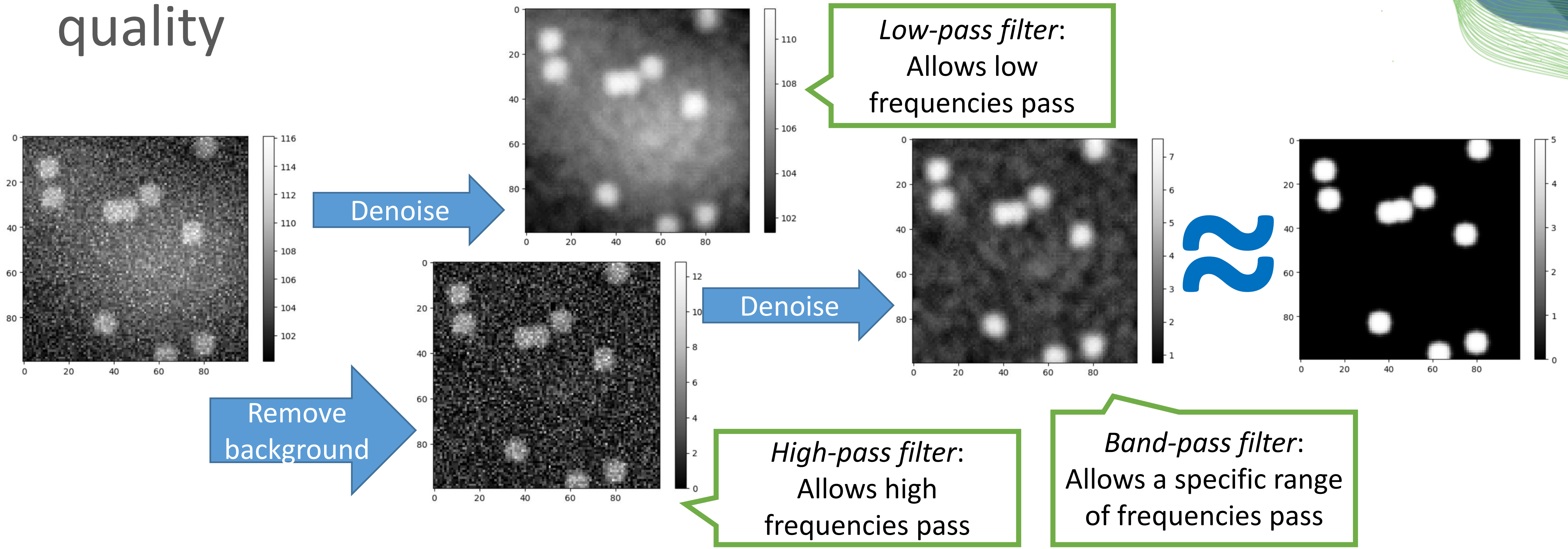


Image filtering

- Attempt to invert / “undo” processes disturbing image quality



Survey

Which is a non-linear filter?

Laplace

Gaussian

Top-hat

Sobel

Survey

Which is a noise-removal filter?

Laplace

Gaussian

Top-hat

Sobel

Survey

Which is a edge-detection filter?

Laplace

Gaussian

Top-hat

Sobel

Survey

Which is a background-removal filter?

Laplace

Gaussian

Top-hat

Sobel

CENTER FOR SCALABLE DATA ANALYTICS
AND ARTIFICIAL INTELLIGENCE

Image segmentation

Robert Haase

Using materials from Marcelo Leomil Zoccoler and Johannes Soltwedel, PoL, TU Dresden

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

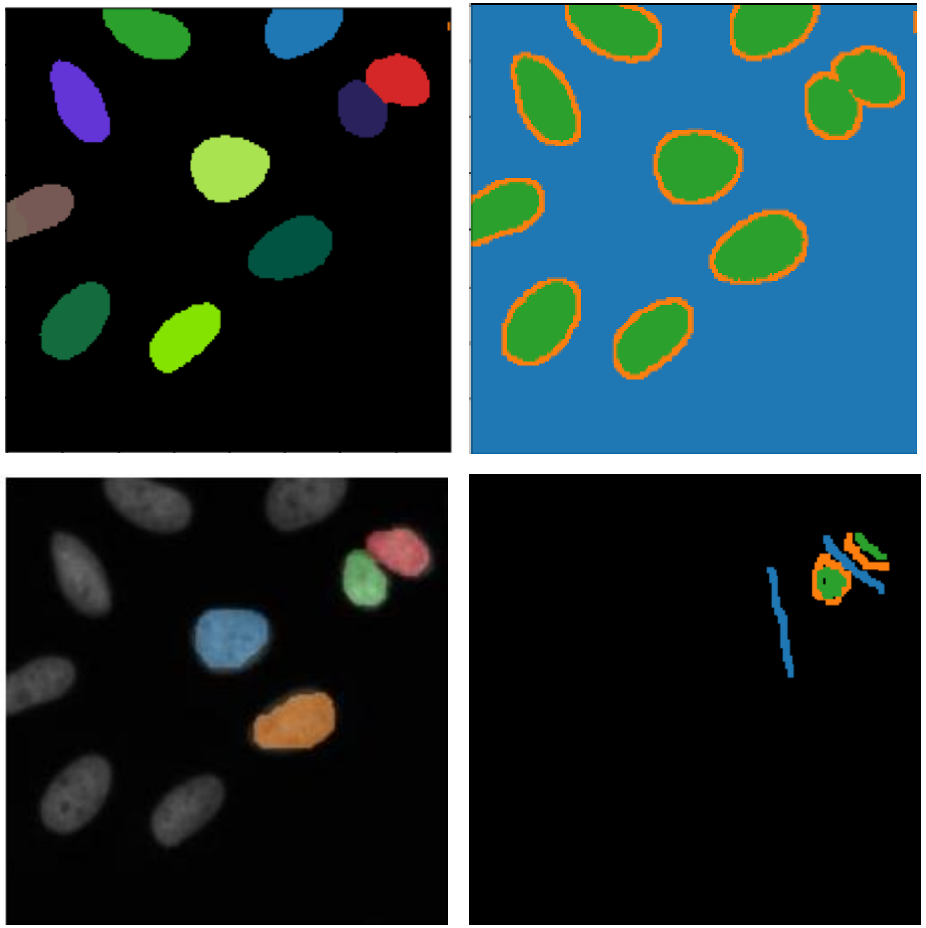
SACHSEN
Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Terminology

- Annotations are typically drawn by humans (e.g. to train machine learning models)

Instance segmentation

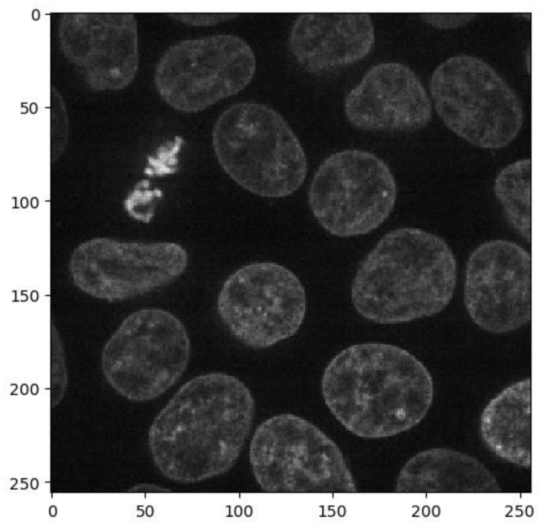
Semantic segmentation

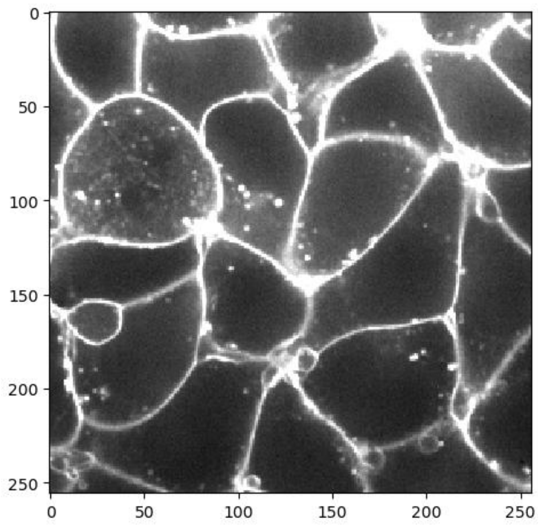


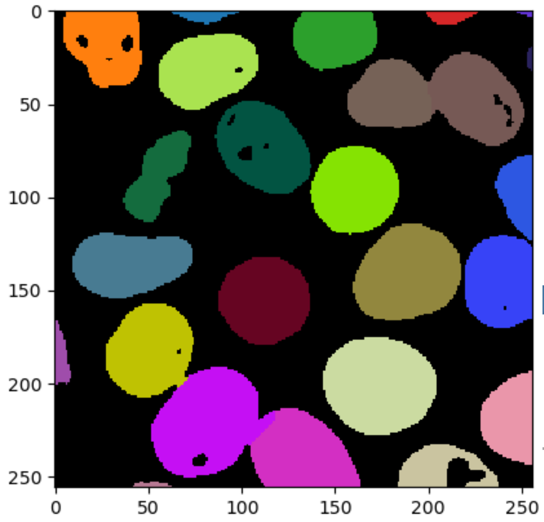
Sparse instance annotation

Sparse semantic annotation

Watershed

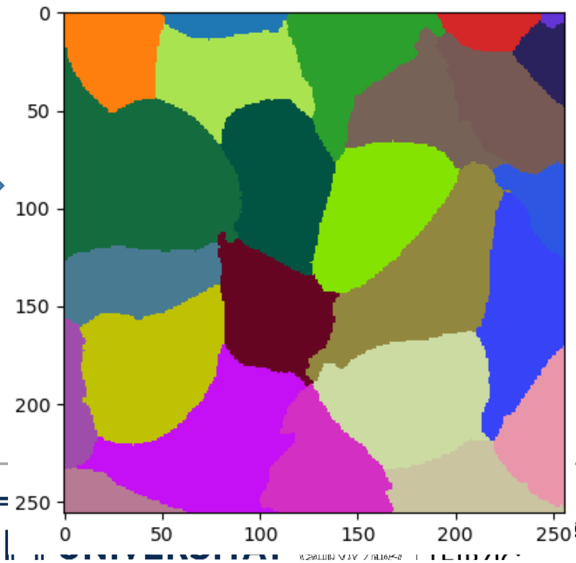






```
labeled_cells = seeded_watershed(membrane_channel, labeled_nuclei)

labeled_cells
```



https://www.napari-hub.org/plugins/napari-segment-blobs-and-things-with-membranes#seeded-watershed

CENTER FOR SCALABLE DATA ANALYTICS
AND ARTIFICIAL INTELLIGENCE

Surface reconstruction

Robert Haase

GEFÖRDERT VOM

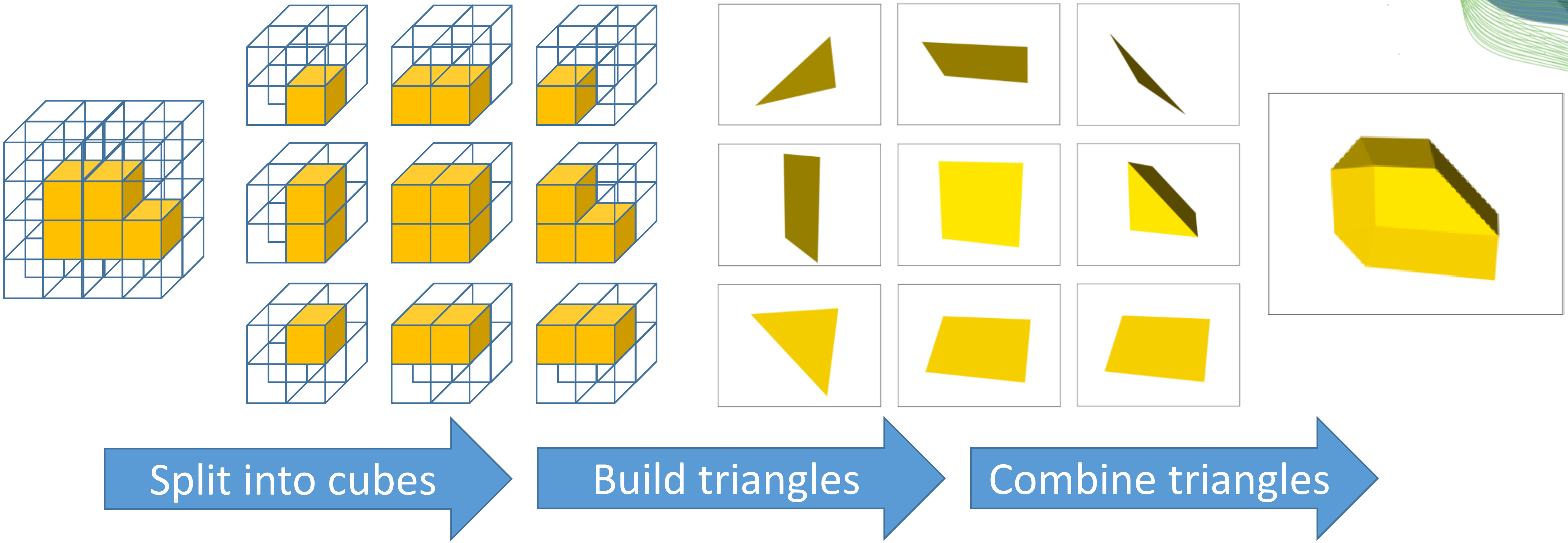
Bundesministerium
für Bildung
und Forschung

SACHSEN

Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Marching cubes algorithm

- Starting point: 3D binary image
- Cuts the image in small cubes and iterates over them



Lorensen, William E.; Cline, Harvey E. (1 August 1987). "Marching cubes: A high resolution 3D surface construction algorithm". *ACM SIGGRAPH Computer Graphics*. **21** (4): 163-169. CiteSeerX 10.1.1.545.613. doi:10.1145/37402.37422.

ScaDS.AI
DRESDEN LEIPZIG
CENTER FOR SCALABLE DATA ANALYTICS
AND ARTIFICIAL INTELLIGENCE

Segmentation quality estimation

Robert Haase

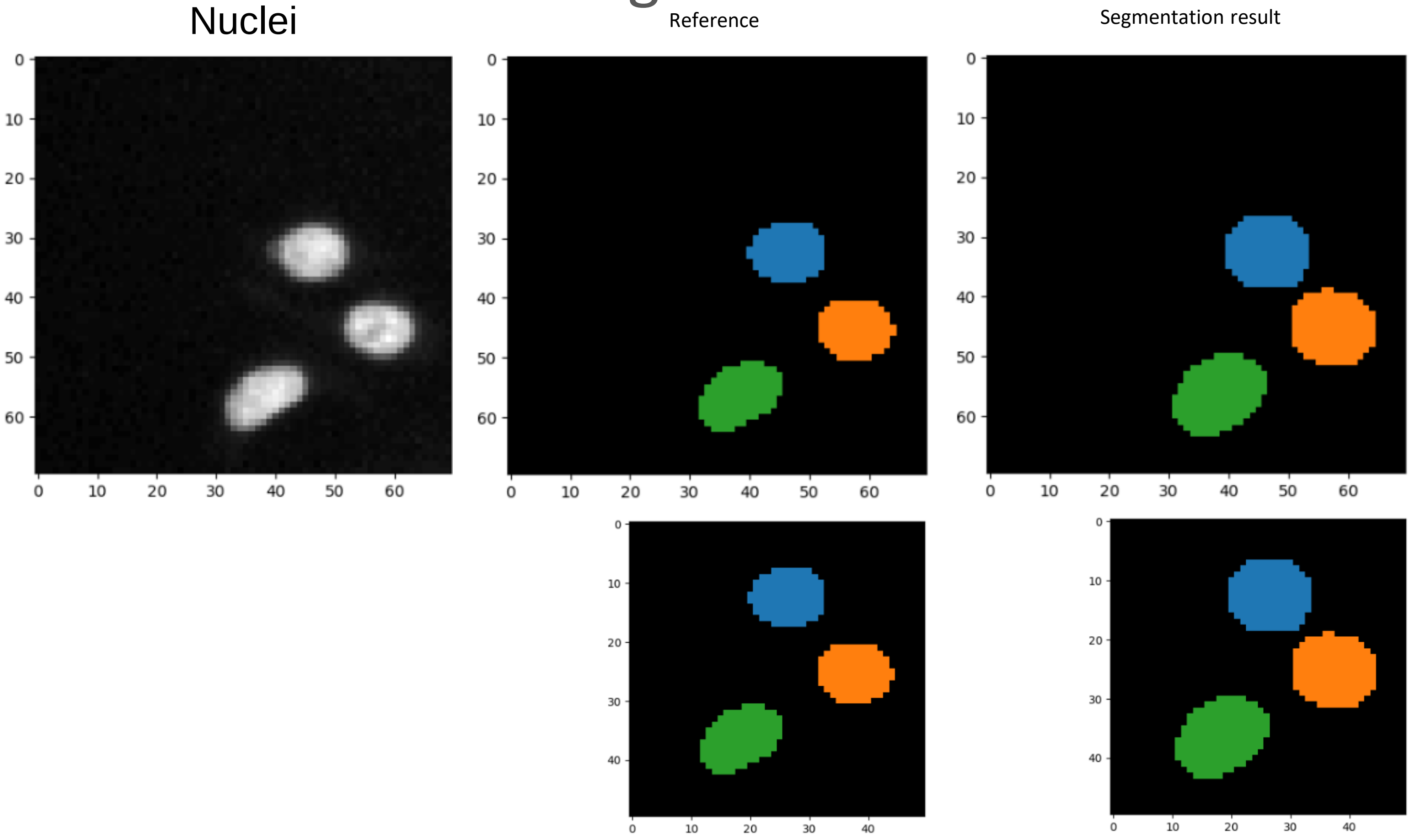
Reusing materials from Lena Maier-Hein, Annika Reinke (DKFZ) et al. and Martin Schätz (Charles Uni Prague)

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

SACHSEN
Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Accuracy versus Jaccard Index (IoU)

- Side-effects of image size and number of nuclei



$$A = \frac{TP + TN}{FN + FP + TP + TN}$$

$$J = \frac{TP}{FN + FP + TP}$$

Accuracy: 0.97
Jaccard Index: 0.73

Accuracy decreases because there are less correct black pixels (TN)

Accuracy: 0.95
Jaccard Index: 0.73

ScaDS.AI
DRESDEN LEIPZIG
CENTER FOR SCALABLE DATA ANALYTICS
AND ARTIFICIAL INTELLIGENCE

Feature extraction

Robert Haase

Using materials from Johannes Soltwedel, PoL, TU Dresden

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

SACHSEN
Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Feature extraction

- A *feature* is a countable or measurable property of an image or object.
- Goal of feature extraction is finding a minimal set of features to describe an object well enough to differentiate it from other objects.

- **Intensity based**
  - Mean intensity
  - Standard deviation
  - Total intensity
  - Textures
- **Mixed features**
  - Center of mass
  - Local minima / maxima
  - Distance to neighbors
  - Average intensity in neighborhood
- **Shape based /spatial**
  - Area / Volume
  - Roundness
  - Solidity
  - Circularity / Sphericity
  - Elongation
  - Centroid
  - Bounding box
- **Spatio-temporal**
  - Displacement,
  - Speed,
  - Acceleration
- **Topological**
  - Number of neighbors
- **Others**
  - Overlap
  - Colocalization

ScaDS.AI
DRESDEN LEIPZIG

CENTER FOR SCALABLE DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

Funded by

Bundesministerium für Bildung und Forschung

SACHSEN

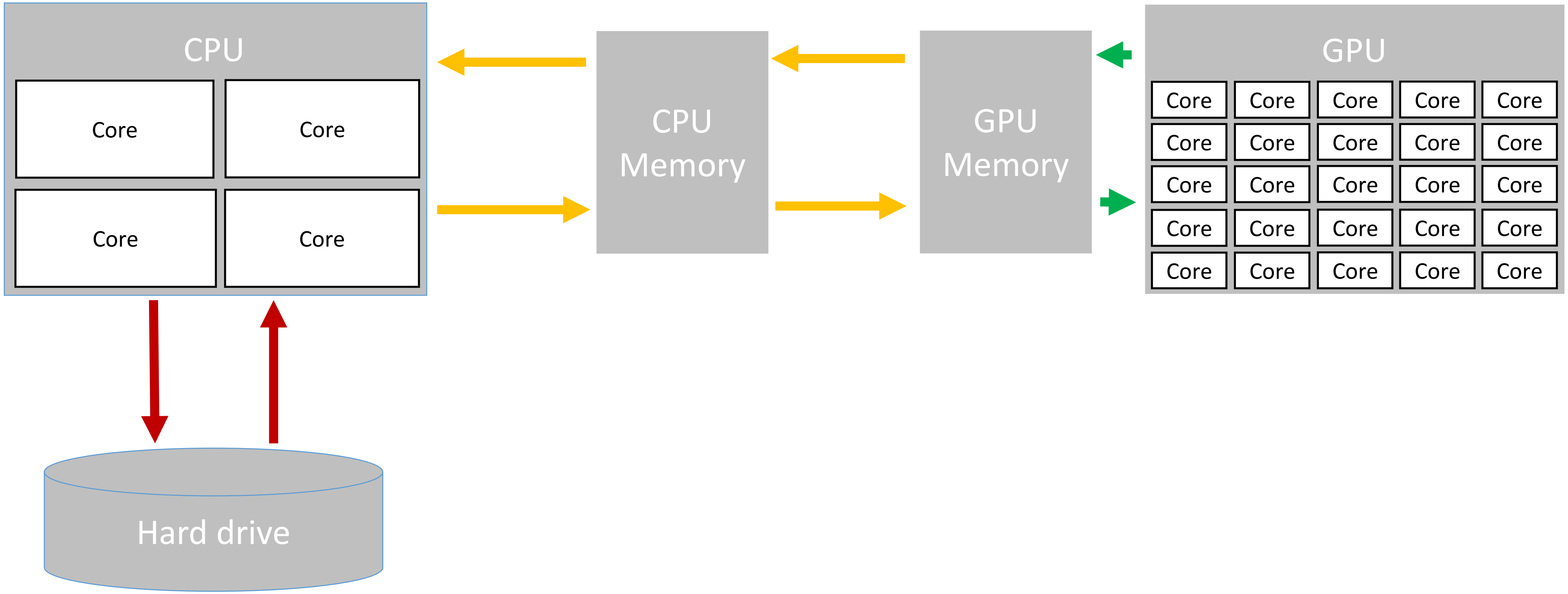
Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Distributed & GPU-accelerated Image Processing

Robert Haase

GPUs allow real-time image processing

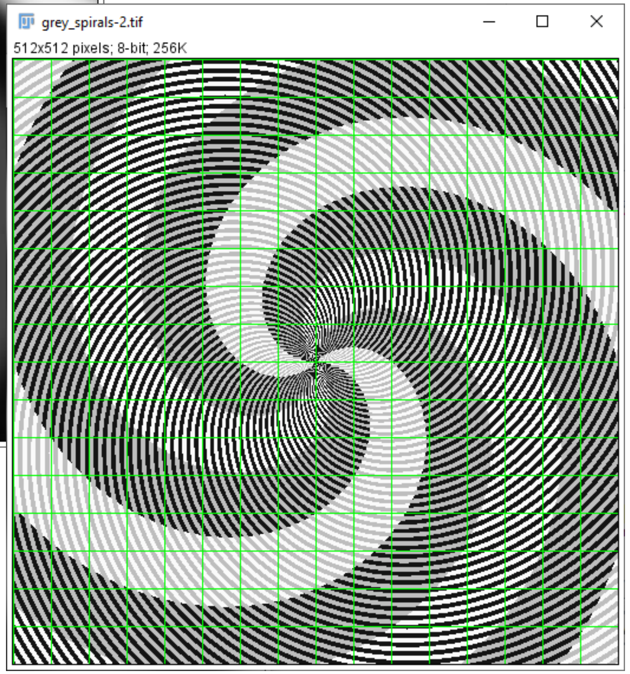
- GPUs are specialised in processing, very fast thanks to many cores and fast memory access

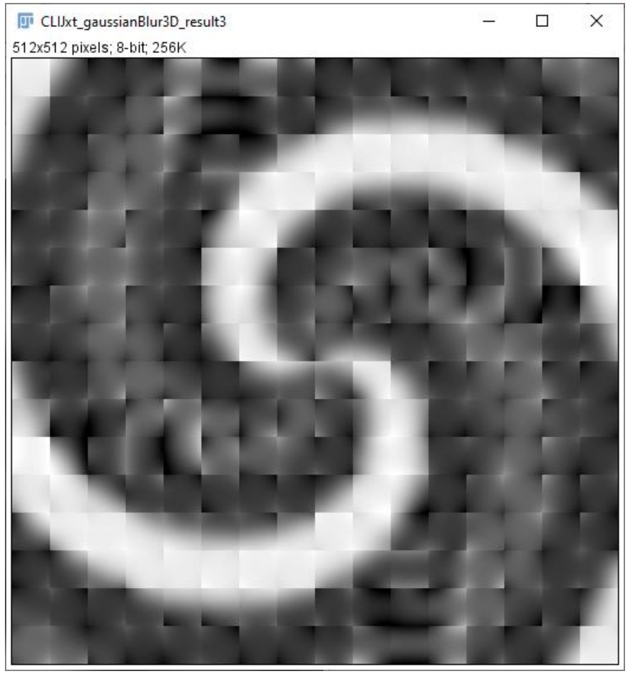


Tiling

- The last perimeter against big data

If the image is too large for the computer memory, image processing as a whole is *not possible.*





Processing tile-by-tile poses new challenges

ScaDS.AI
DRESDEN LEIPZIG
CENTER FOR SCALABLE DATA ANALYTICS
AND ARTIFICIAL INTELLIGENCE

Methods for comparing measurement methods

Robert Haase

Using materials Reusing materials from Daniela Vorkel, Douglas G. Altman and J. Martin Bland

GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

SACHSEN
Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

The confidence interval

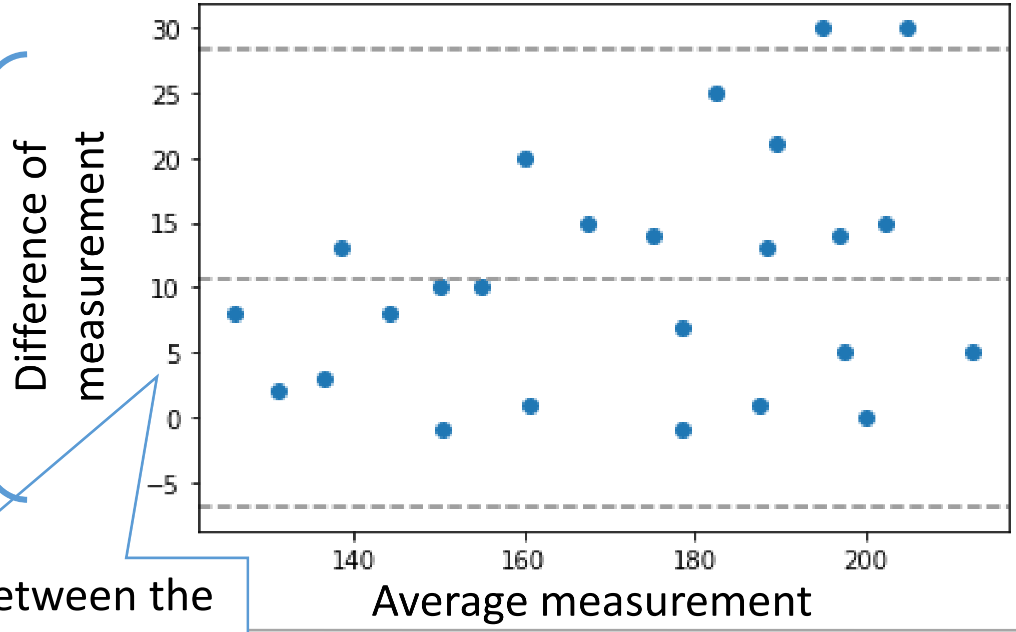
- "The British Standards Institution (1979) define a coefficient of repeatability as 'the value below which the difference between two single test results ... may be expected to lie with a specified probability; in the absence of other indications, the probability is 95 per cent'."[1]

The confidence interval *CI* of agreement

$$CI = (\mu - 2\sigma, \mu + 2\sigma)$$

Mean difference

Standard deviation of differences



The mean difference between the methods ("offset")

[1] Altman & Bland, The Statistician 32, 1983

ScaDS.AI DRESDEN LEIPZIG

CENTER FOR SCALABLE DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

Supervised and Unsupervised Machine Learning for Bio-image Analysis

Robert Haase

Funded by

Bundesministerium für Bildung und Forschung

SACHSEN

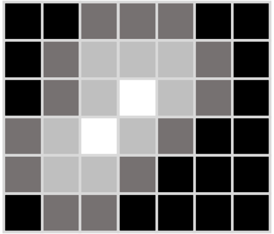
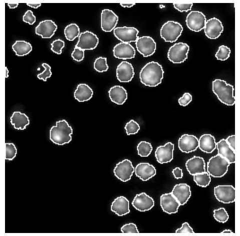
Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Reusing materials from Johannes Soltwedel, Till Korten, Johannes Müller, Laura Žigutyte (TU Dresden), Ryan Savill (MPI-CBG), Matthias Täschner (ScaDS.AI/Uni Leipzig) and the Scikit-learn community.

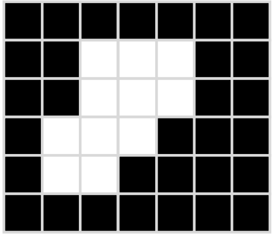
Machine learning

- Automatic construction of predictive models from given data

Pixels, Objects, Images, Audio, Text, Measurements, …


Dense Segmentation / Binarization



Object classification

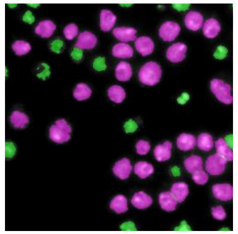
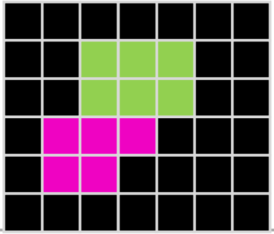


Image classification

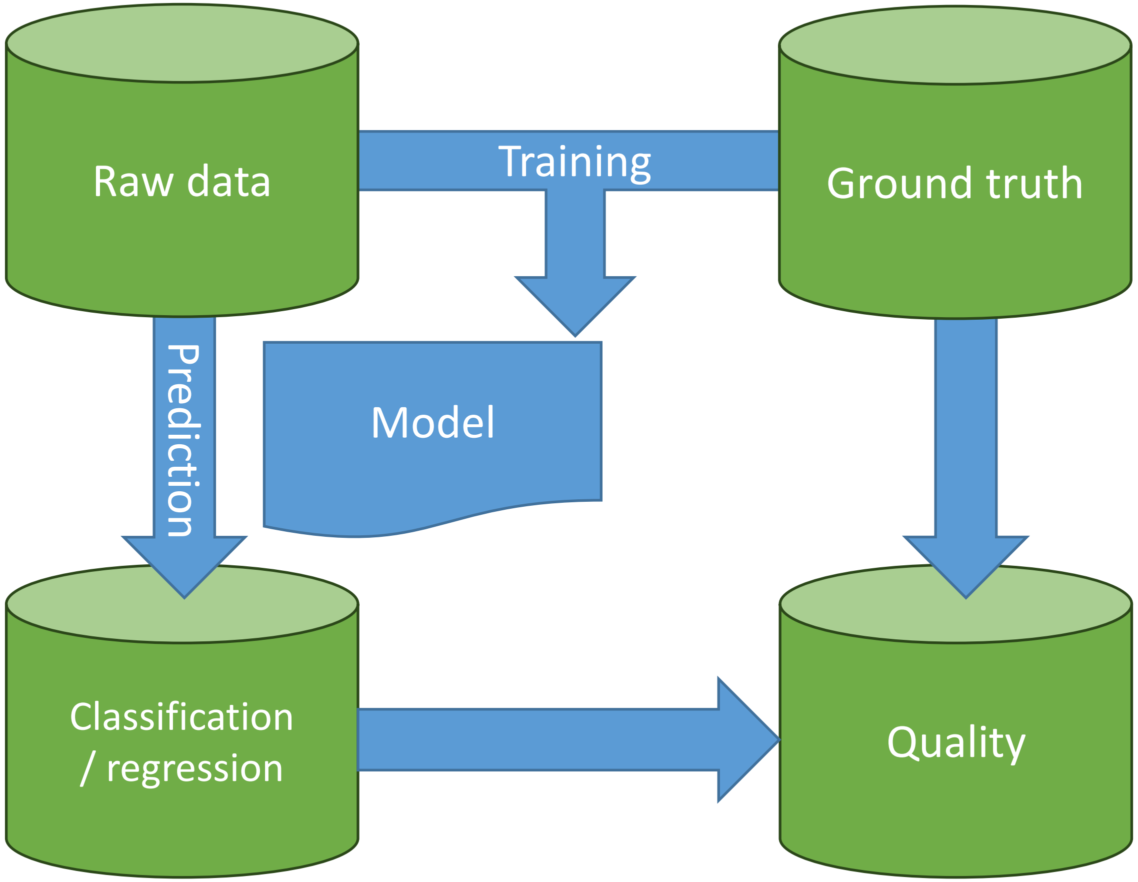
"Cat"

Instance segmentation



Cont. quantity

P_{Cat}= 0.5
$P_{Microscope}$= 0.4
Height = 80 cm



Annotated raw data, usually generated by humans

Precision, Recall

Hypothesis-driven quantitative biology

- Hypothesis: Cell shape can be influenced by modifying X.
- Null-Hypothesis: Circularity of modified cells is similar to cells in the control group.

- Sample preparation
- Imaging
- Cell segmentation
- Circularity measurement
- Statistics

Shall we use a different microscope?

Should we use a different segmentation algorithm?

Is circularity the right parameter to measure?

Hypothesis *generating* quantitative biology

- ~~Hypothesis: Cell shape can be influenced by modifying X.~~
- Question: Which image-derived parameter is influenced when modifying X?
  - Sample preparation
  - Imaging
  - Cell segmentation algorithm A, algorithm B, algorithm C
  - Measurement of circularity, solidity, elongation, extend, texture, intensity, topology ...
  - Statistics

Which segmentation algorithms allow measurements that show a relationship with X?

Why?

Which parameter shows any relationship with X?

Correlation statistics

[16]:
```python
def colorize(styler):
    styler.background_gradient(axis=None, cmap="PiYG")
    return styler

df = measurements.corr().T
df.style.pipe(colorize)
```

[16]:

	label	area	bbox_area	equivalent_diameter	convex_area	max_intensity	mean_intensity	min_intensity	extent	local_centroid-0	local_centroid-1	solidity	feret_diameter_max	major_axis_length	minor_axis_length	orientation	eccentricity
label	1.000000	0.261682	0.223070	0.249249	0.250594	0.110791	0.235692	nan	0.031673	0.177363	0.227746	0.090163	0.208067	0.198908	0.237521	0.319053	0.059804
area	0.261682	1.000000	0.973718	0.978723	0.997560	0.511730	0.530250	nan	-0.362472	0.847281	0.935689	-0.243908	0.930981	0.911069	0.859240	0.280673	0.348585
bbox_area	0.223070	0.973718	1.000000	0.948328	0.985584	0.481524	0.476951	nan	-0.546728	0.902854	0.904551	-0.416707	0.973189	0.967337	0.752580	0.213080	0.479196
equivalent_diameter	0.249249	0.978723	0.948328	1.000000	0.974614	0.633984	0.618553	nan	-0.395696	0.858779	0.947036	-0.266587	0.931696	0.904412	0.904698	0.197456	0.363799
convex_area	0.250594	0.997560	0.985584	0.974614	1.000000	0.506730	0.517356	nan	-0.413323	0.862417	0.934090	-0.305706	0.948048	0.932682	0.832264	0.263176	0.389269
max_intensity	0.110791	0.511730	0.481524	0.633984	0.506730	1.000000	0.825115	nan	-0.324093	0.504879	0.603305	-0.253635	0.536089	0.502524	0.645600	-0.139025	0.246172
mean_intensity	0.235692	0.530250	0.476951	0.618553	0.517356	0.825115	1.000000	nan	-0.160940	0.412859	0.609264	-0.077797	0.458515	0.422638	0.707711	0.132754	0.017030
min_intensity	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan	nan
extent	0.031673	-0.362472	-0.546728	-0.395696	-0.413323	-0.324093	-0.160940	nan	1.000000	-0.631158	-0.375580	0.853431	-0.631776	-0.664733	-0.062873	0.252915	-0.756019
local_centroid-0	0.177363	0.847281	0.902854	0.858779	0.862417	0.504879	0.412859	nan	-0.631158	1.000000	0.706437	-0.439244	0.937673	0.932889	0.623186	0.003490	0.560853
local_centroid-1	0.227746	0.935689	0.904551	0.947036	0.934090	0.603305	0.609264	nan	-0.375580	0.706437	1.000000	-0.290177	0.863585	0.840724	0.875044	0.271191	0.318154
solidity	0.090163	-0.243908	-0.416707	-0.266587	-0.305706	-0.253635	-0.077797	nan	0.853431	-0.439244	-0.290177	1.000000	-0.512903	-0.556555	0.049965	0.279509	-0.723572
feret_diameter_max	0.208067	0.930981	0.973189	0.931696	0.948048	0.536089	0.458515	nan	-0.631776	0.937673	0.863585	-0.512903	1.000000	0.996744	0.690639	0.077145	0.614849
major_axis_length	0.198908	0.911069	0.967337	0.904412	0.932682	0.502524	0.422638	nan	-0.664733	0.932889	0.840724	-0.556555	0.996744	1.000000	0.639308	0.076773	0.647021
minor_axis_length	0.237521	0.859240	0.752580	0.904698	0.832264	0.645600	0.707711	nan	-0.062873	0.623186	0.875044	0.049965	0.690639	0.639308	1.000000	0.278107	-0.012148
orientation	0.319053	0.280673	0.213080	0.197456	0.263176	-0.139025	0.132754	nan	0.252915	0.003490	0.271191	0.279509	0.077145	0.076773	0.278107	1.000000	-0.305652
eccentricity	0.059804	0.348585	0.479196	0.363799	0.389269	0.246172	0.017030	nan	-0.756019	0.560853	0.318154	-0.723572	0.614849	0.647021	-0.012148	-0.305652	1.000000
standard_deviation_intensity	0.189165	0.288670	0.267528	0.402328	0.285105	0.867057	0.902001	nan	-0.216260	0.284331	0.379400	-0.169801	0.306228	0.280378	0.455324	-0.089349	0.107307
aspect_ratio	0.036433	0.411794	0.581132	0.386884	0.462720	0.121313	-0.044872	nan	-0.848271	0.678234	0.321805	-0.787587	0.690082	0.736200	-0.030443	-0.181927	0.853302
roundness	-0.055815	-0.415592	-0.569335	-0.406856	-0.464090	-0.191680	0.009002	nan	0.834550	-0.638667	-0.359961	0.801971	-0.690444	-0.732103	0.003699	0.224205	-0.955978
circularity	-0.054152	-0.626241	-0.718764	-0.701230	-0.659125	-0.636372	-0.411166	nan	0.808533	-0.785693	-0.644979	0.773934	-0.832660	-0.839196	-0.435236	0.242901	-0.779895
UMAP0	-0.065835	-0.442711	-0.413779	-0.509190	-0.435101	-0.324496	-0.387465	nan	0.168523	-0.391875	-0.488311	0.068021	-0.457079	-0.437340	-0.479807	0.025473	-0.204662
UMAP1	0.139702	0.819263	0.813951	0.793707	0.821940	0.391350	0.365621	nan	-0.375632	0.720004	0.753502	-0.260000	0.753713	0.736954	0.702828	0.277117	0.251959
MANUAL_CLUSTER_ID	0.080739	0.677335	0.719434	0.590973	0.700457	0.156570	0.074372	nan	-0.371454	0.582543	0.616873	-0.418390	0.671673	0.686248	0.387847	0.163152	0.424045

My annotation *seems related* to area

My annotation *seems not related* to intensity

CENTER FOR SCALABLE DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

Deep Learning for Bio-image Analysis

Robert Haase

Funded by

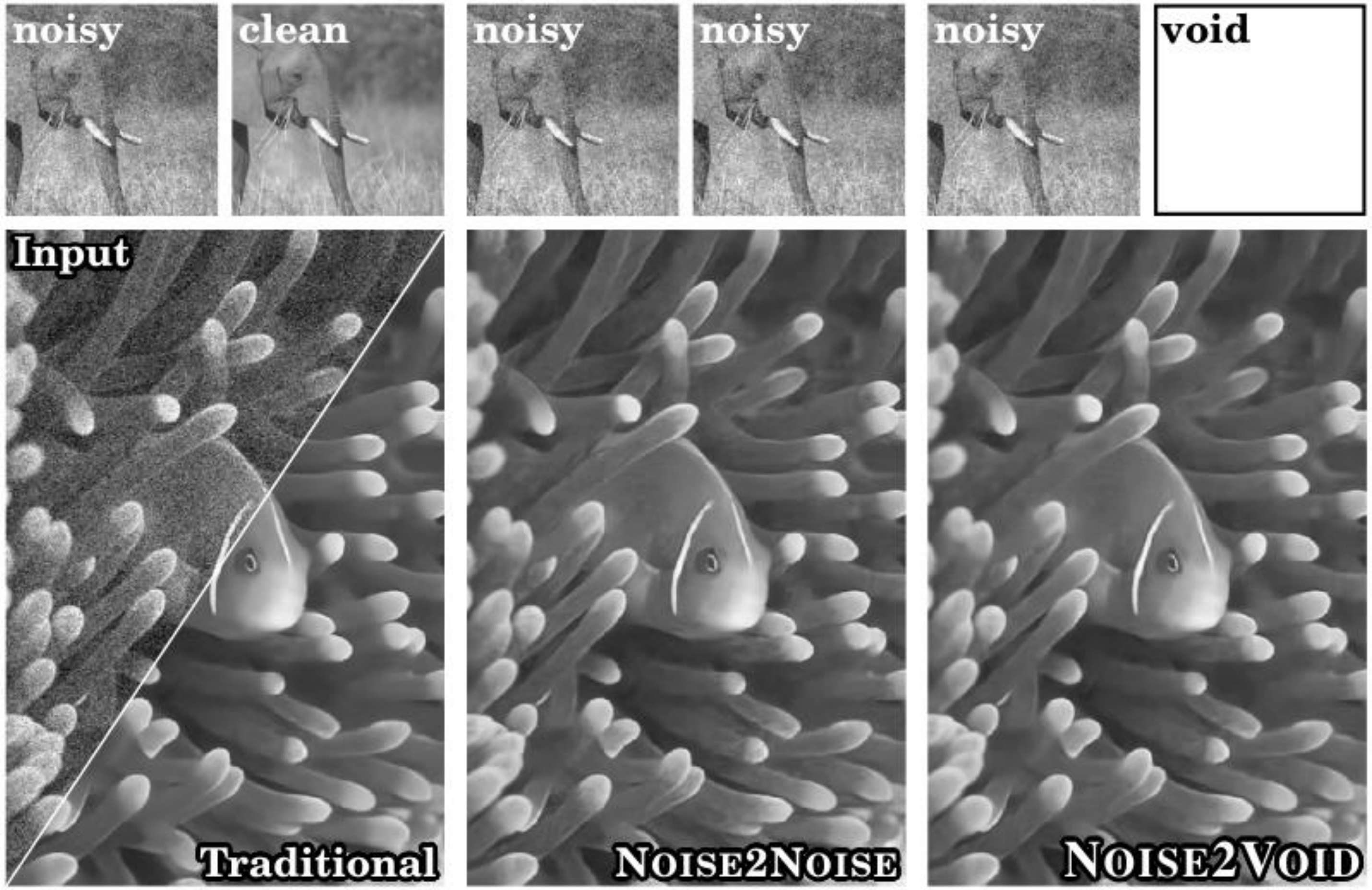
Bundesministerium für Bildung und Forschung

SACHSEN

Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Noise2void

- Image denoising without image pairs



Source: A.Krull et al. https://arxiv.org/pdf/1811.10980

CellPose

- Cell/Nuclei – segmentation based on flow-fields

- Technically similar to Watershed, but with a deep-learning based altitude-image

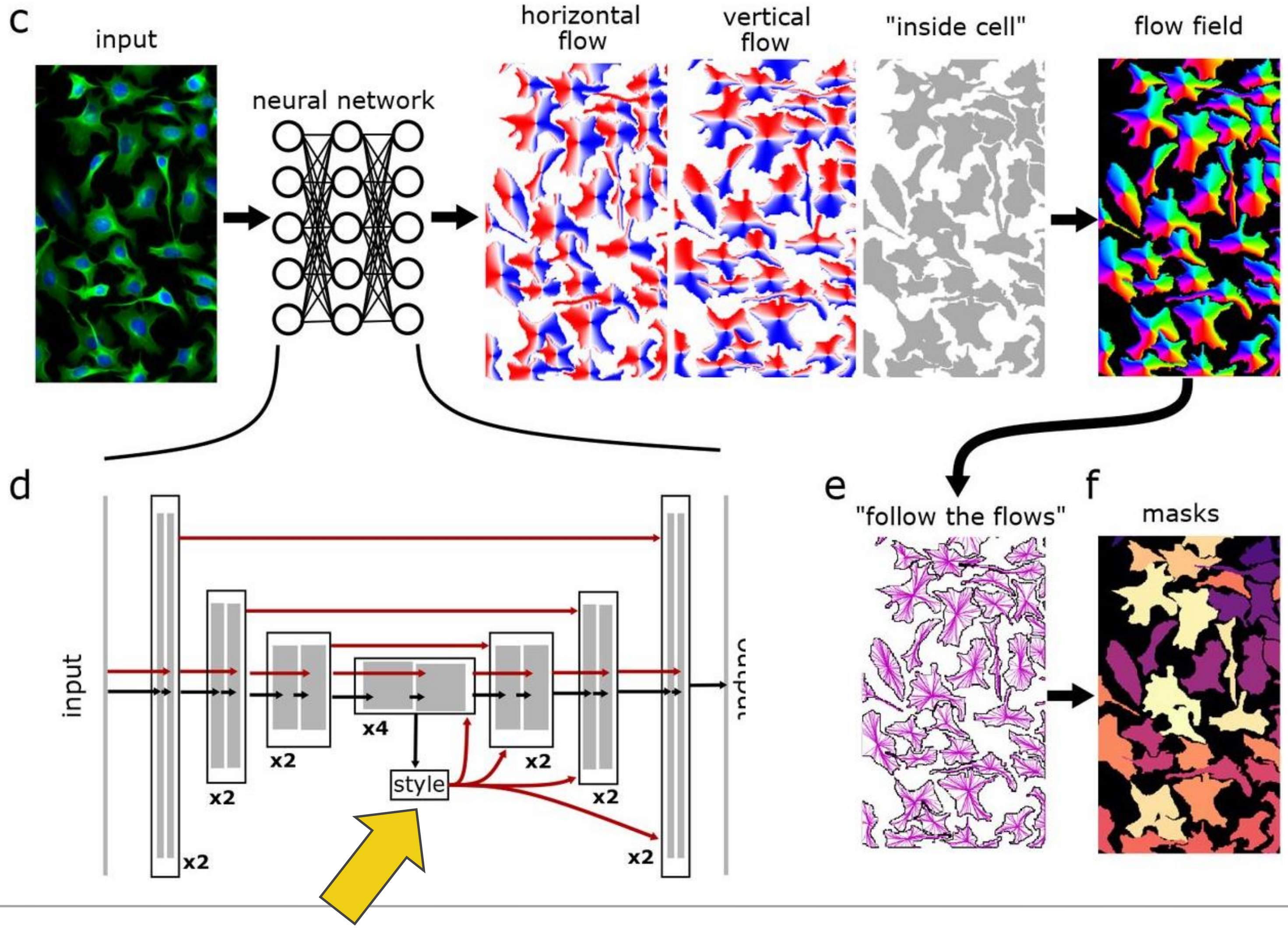


Image source: Modified from Stringer et al (2020) licensed [CC-BY-NC 4.0](https://creativecommons.org/licenses/by-nc/4.0/)
https://www.biorxiv.org/content/10.1101/2020.02.02.931238v2.full

CENTER FOR SCALABLE DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

Funded by

Bundesministerium für Bildung und Forschung

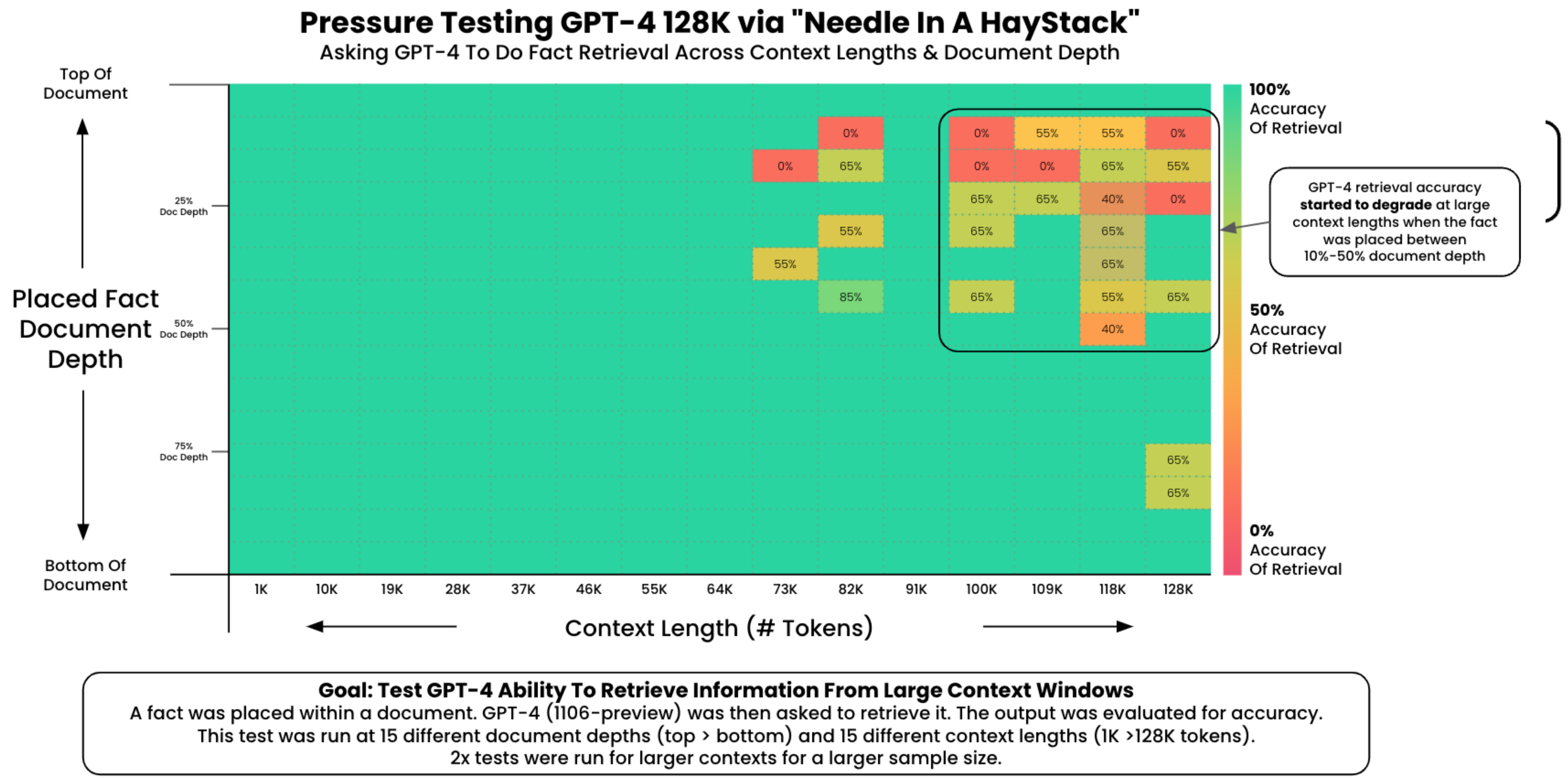
SACHSEN

Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

Prompt Engineering & ChatBots

Robert Haase

Context length



ScaDS.AI DRESDEN LEIPZIG

CENTER FOR SCALABLE DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

Large Language Models for Function Calling

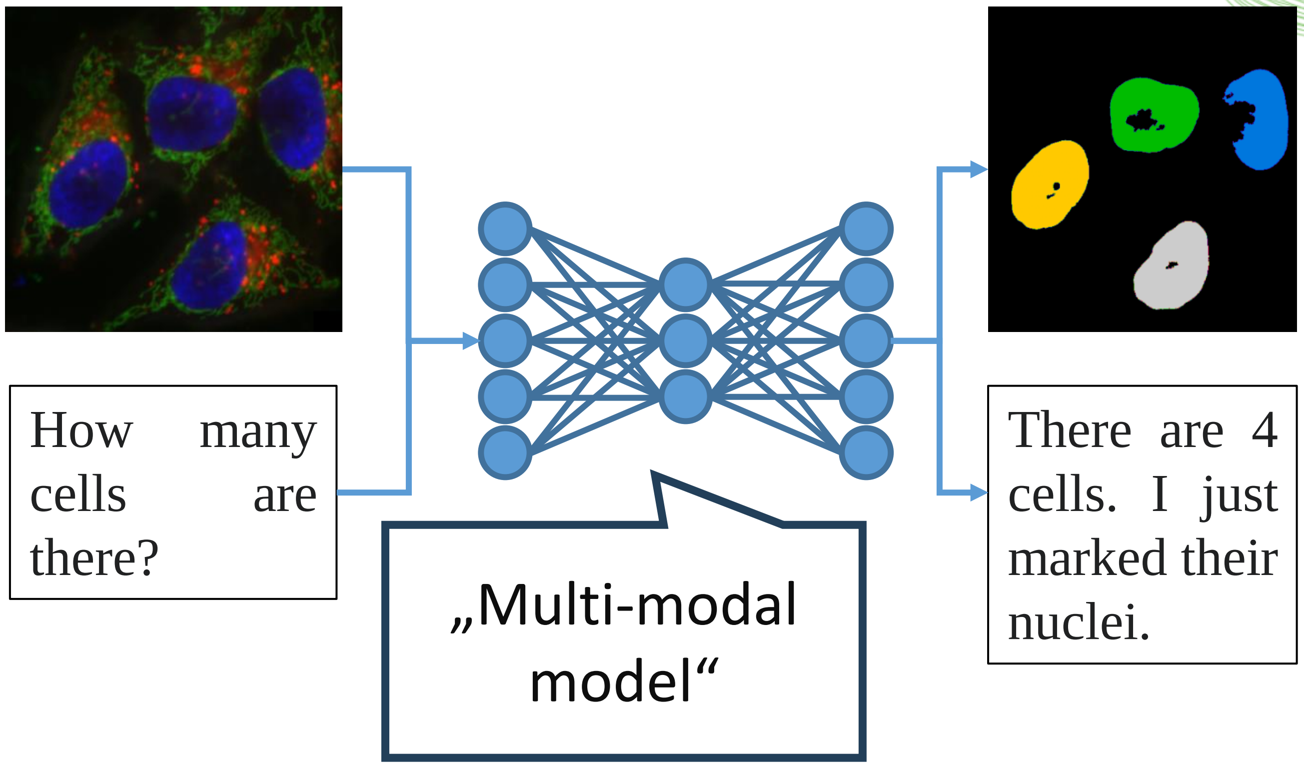
Robert Haase

Generative Artificial Intelligence

- Definition: "Generative artificial intelligence [...] is a type of artificial intelligence (AI) system capable of generating text, images, or other media in response to prompts."[1]

- Commonly based on Neural Networks
- Bridges fields:
  - Natural Language Processing (NLP)
  - Computer Vision (CV)

- Use-cases
  - Translating text
  - Writing emails, text, grant proposals
  - Summarizing articles
  - Writing code
  - General question answering
  - Image generation
  - Image interpretation / analysis

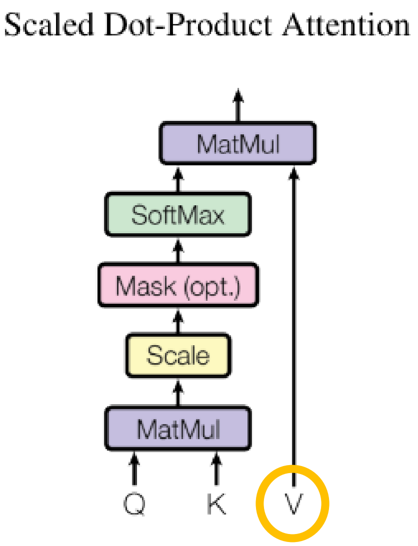


ScaDS.AI DRESDEN LEIPZIG
Robert Haase
@haesleinhuepf

[1] source: https://en.wikipedia.org/wiki/Generative_artificial_intelligence

Quiz: Recap

- The V in attention mechanisms stand for…?



The word we are determining attention from

The word we are determining attention to

The relevance between the two words

The variance of attention between the two words

Function calling

- (choosing a tool)

Given a list of tools…

- get_current_time
- order_food
- book_room

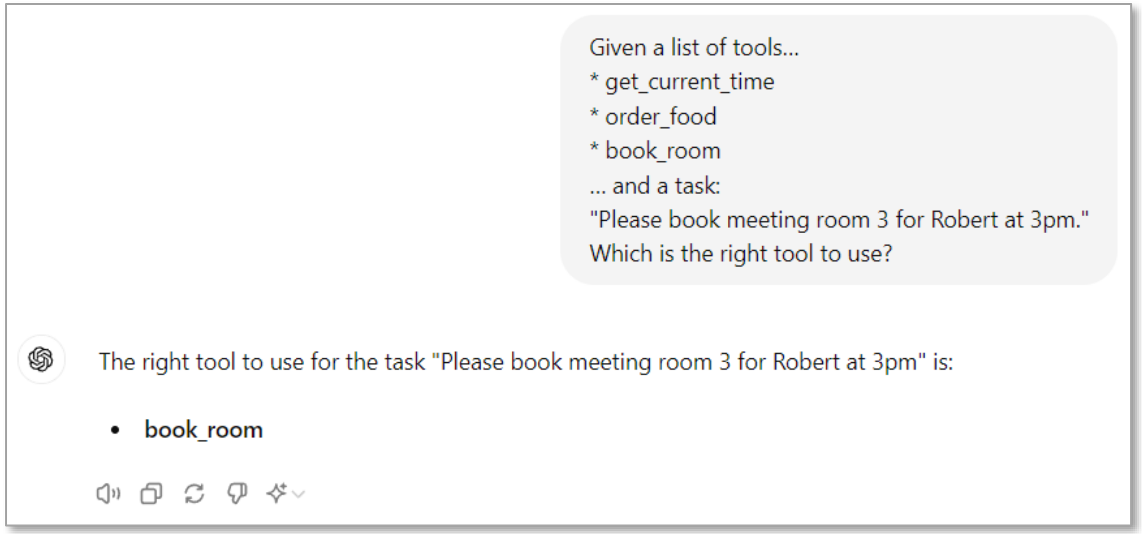
… and a task:

Please book meeting room 3 for Robert at 3pm.

Which is the right tool to use?

Some kind of next-word prediction task

book_room



Function calling

- (parameterizing)

Some kind of translation task

Given a function signature…

book_room(room, time, person)

… and a task:

Please book meeting room 3 for Robert at 3pm.

How could I use the tool?

book_room(“Meeting Room 3“, “3pm“, “Robert“)

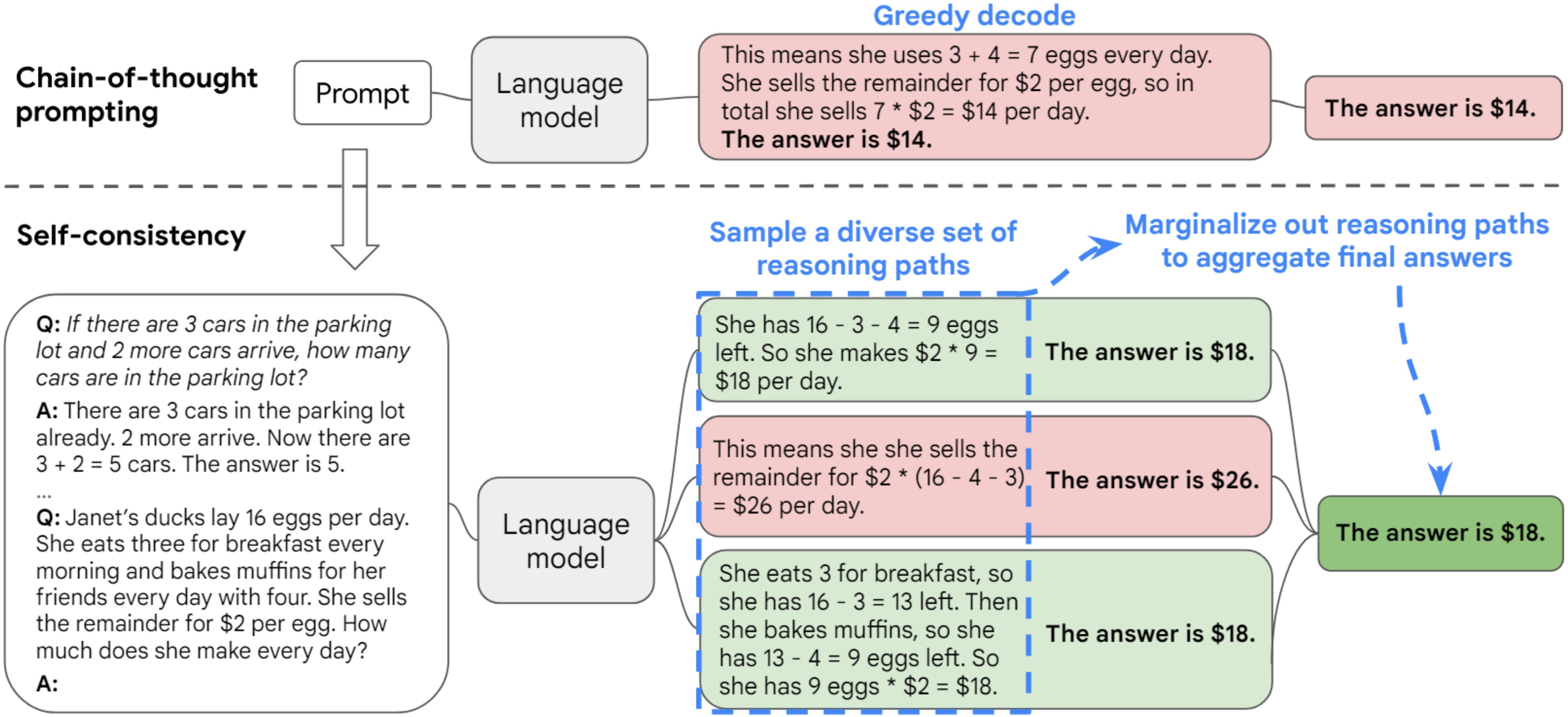
CENTER FOR SCALABLE DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

Prompt Engineering, Retrieval Augmented Generation and Fine-Tuning

Robert Haase

Self-consistency prompting

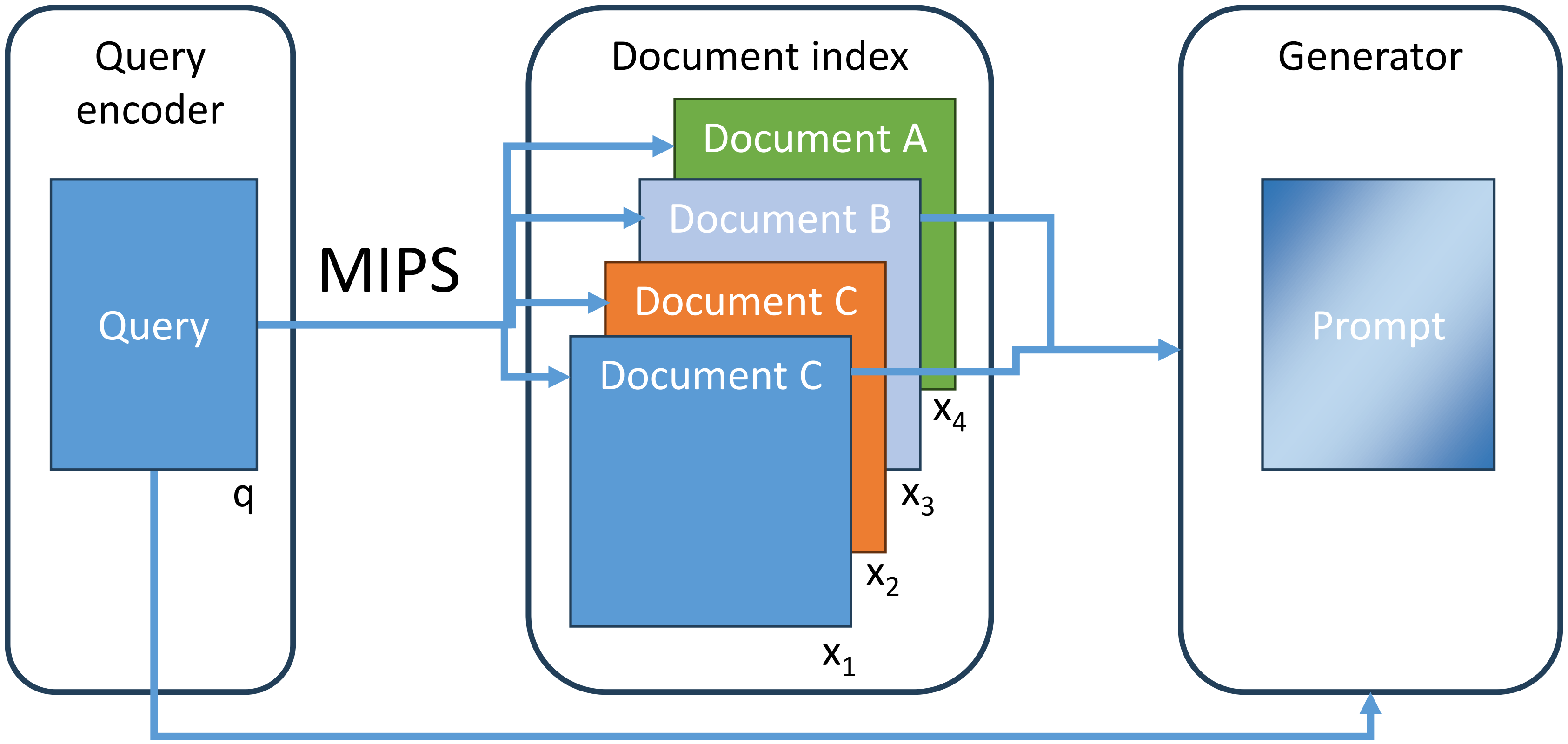
- Prompting multiple times and keep the least conflicting result



Cropped from Zhang et al 2022, licensed CC-BY 4.0
https://arxiv.org/abs/2203.11171

Retrieval Augmented Generation

- Enriching a prompt with relevant context



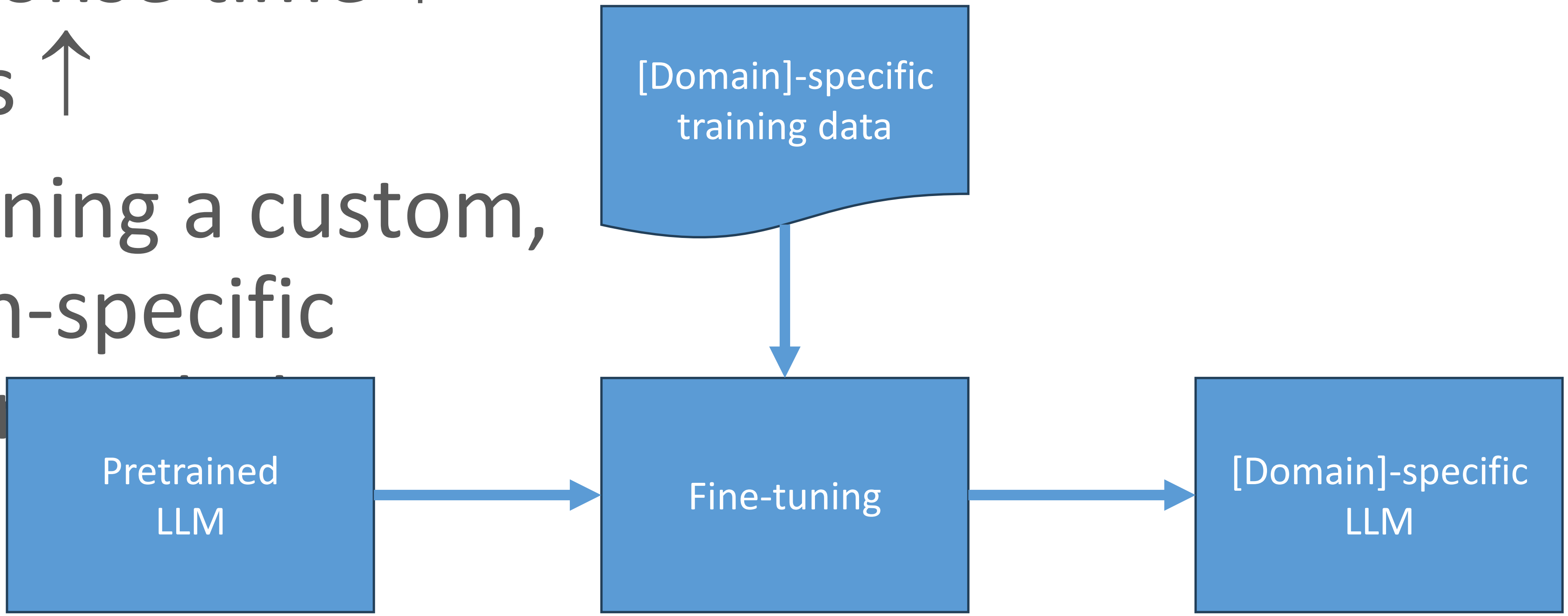
Maximum inner product search (MIPS)

$$x = \text{argmax}_{x_i \in D} x_i^T q$$

Read more: Lewis et al 2020
https://arxiv.org/abs/2005.11401

Fine-tuning

- Long prompts due to prompt-engineering)
  - Response time ↑
  - Costs ↑
- Fine-tuning a custom, Domain-specific model r



ScaDS.AI DRESDEN LEIPZIG

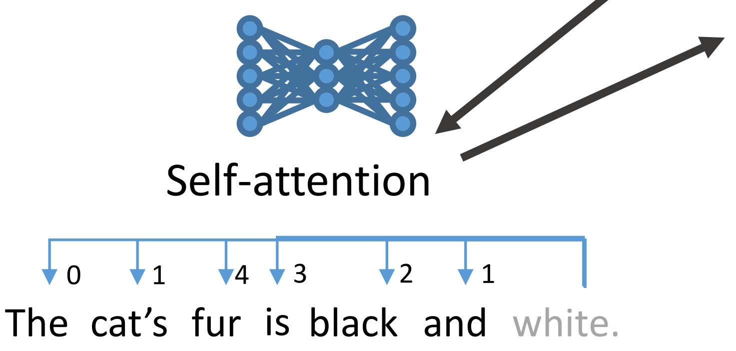
CENTER FOR SCALABLE DATA ANALYTICS AND
ARTIFICIAL INTELLIGENCE

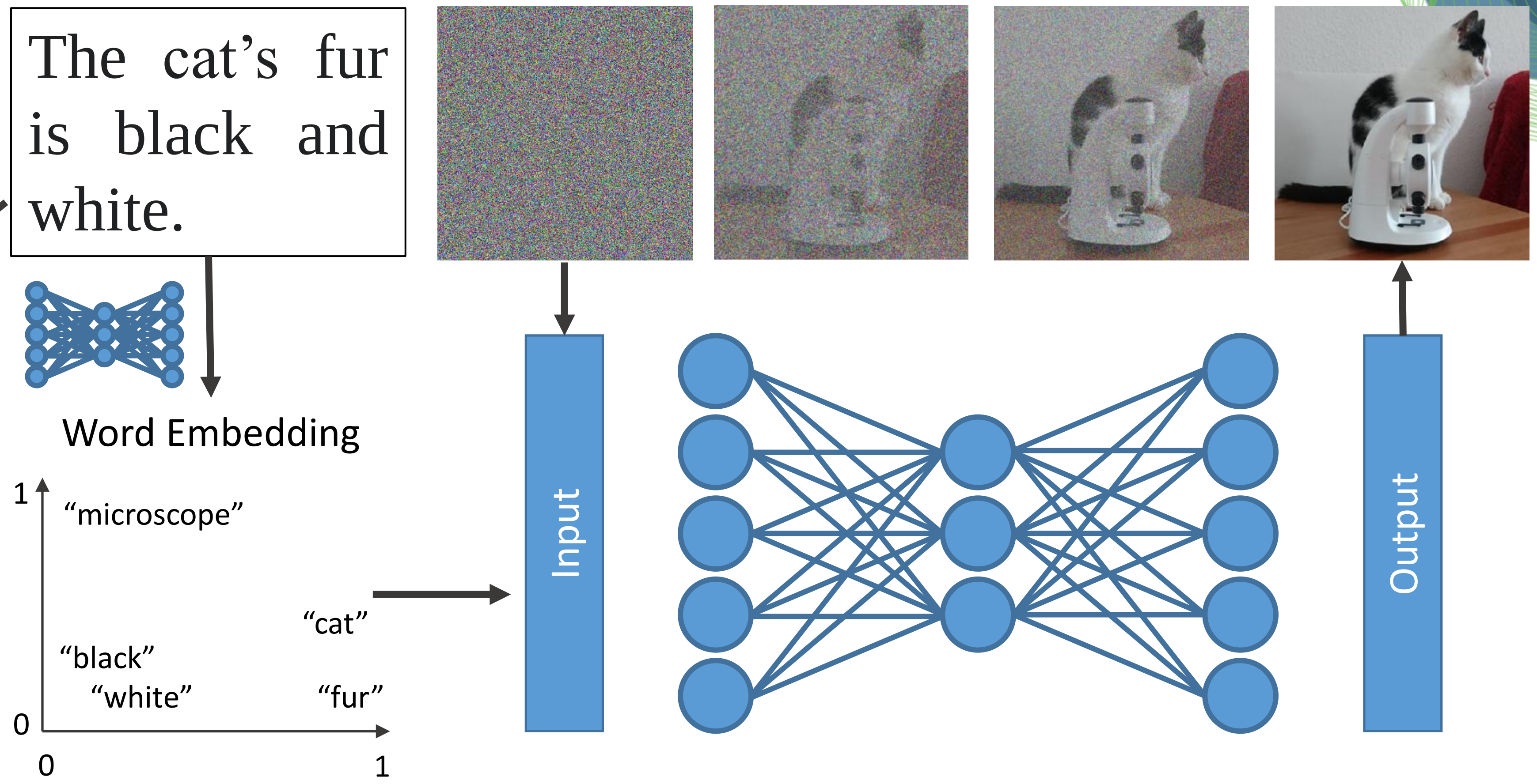
Image Generation and Vision Language Models

Robert Haase

Stable diffusion

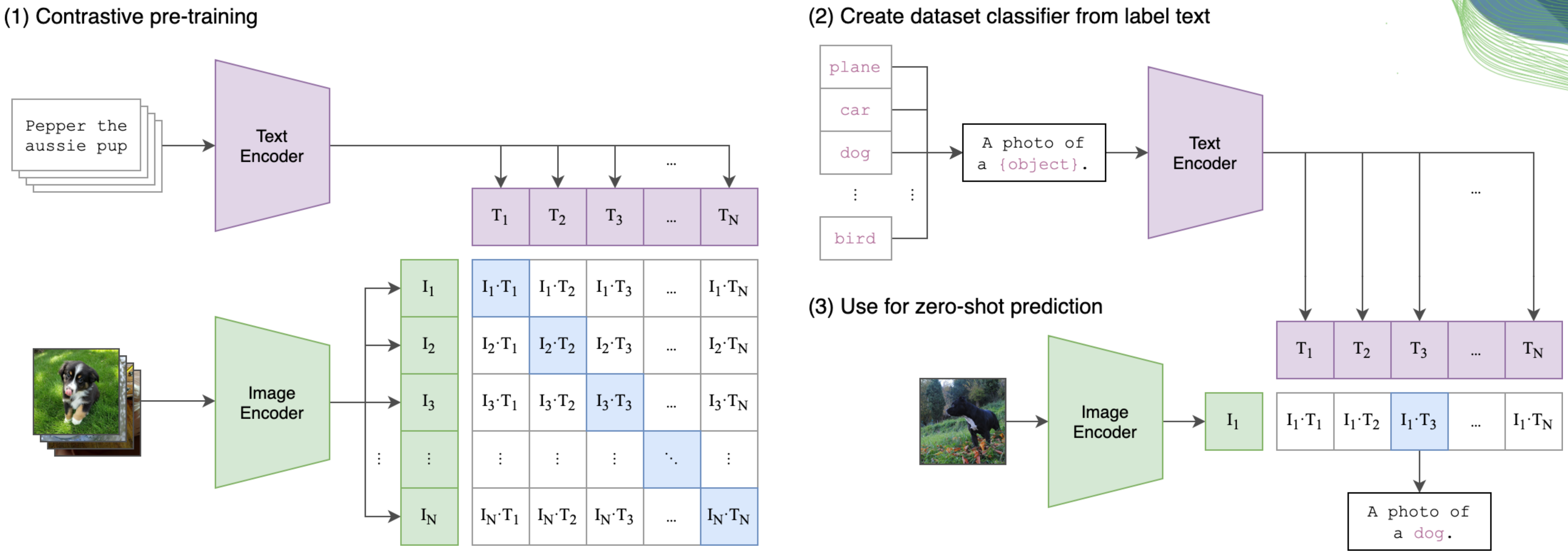
- Reminder:
  - Word embeddings
  - Attention





Contrastive Language-Image Pre-Training

- „CLIP“ Transformers



Source: Radford et al 2021, License MT
https://arxiv.org/abs/2103.00020
https://github.com/OpenAI/CLIP

CENTER FOR SCALABLE DATA ANALYTICS AND ARTIFICIAL INTELLIGENCE

NATIONAL RESEARCH DATA MANAGEMENT INFRASTRUCTURE FOR MICROSCOPY AND BIOIMAGE ANALYSIS

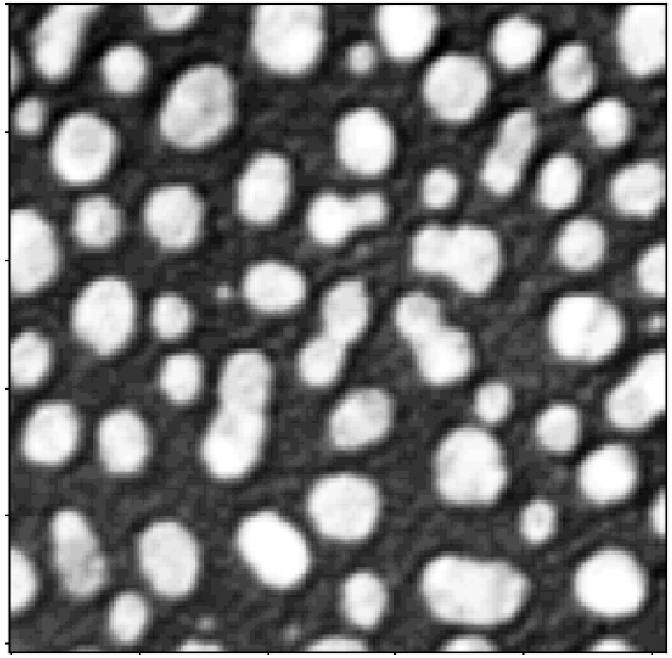
GLOBAL BIOIMAGE ANALYST'S SOCIETY

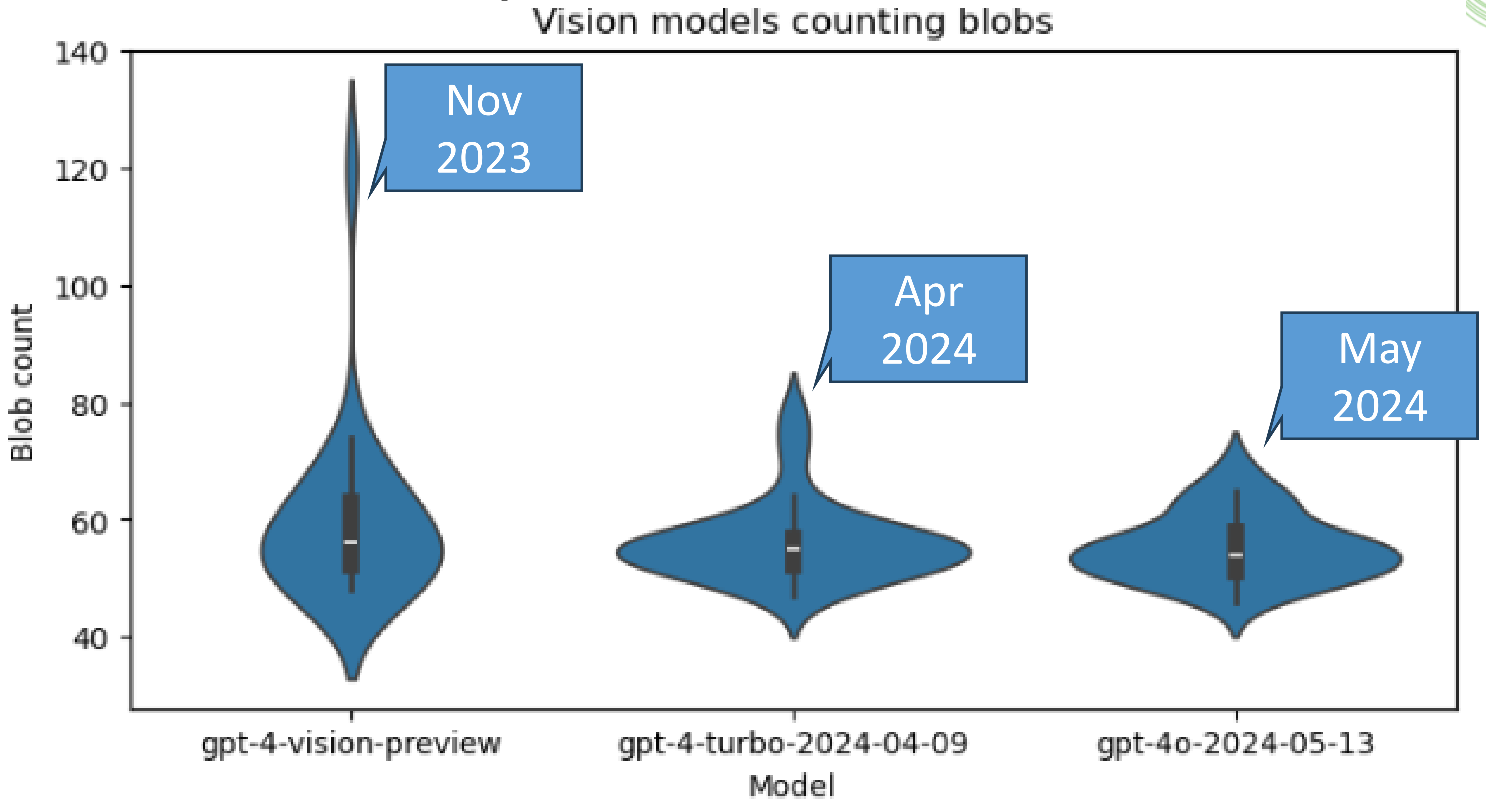
Benchmarking LLMs

Robert Haase

Benchmarking vision models

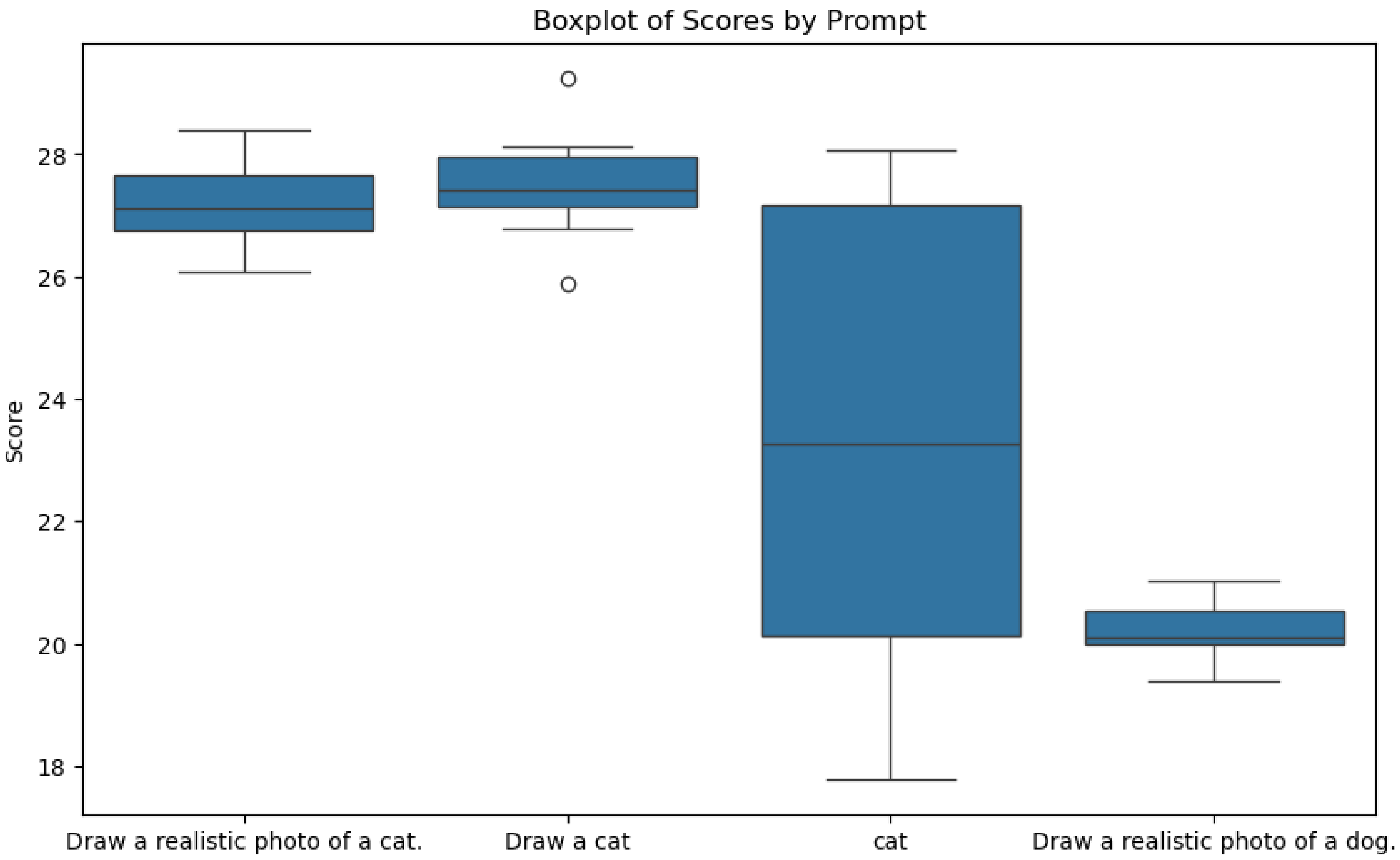
- Prompt: „Analyse the following image by counting the bright blobs. Respond with the number only. “ (n=25)





CLIP scores

- Example: Prompt optimization



Always have a control experiment!

Trying out different prompts for generating pictures of cats

Testing functional correctness: HumanEval

Abstract

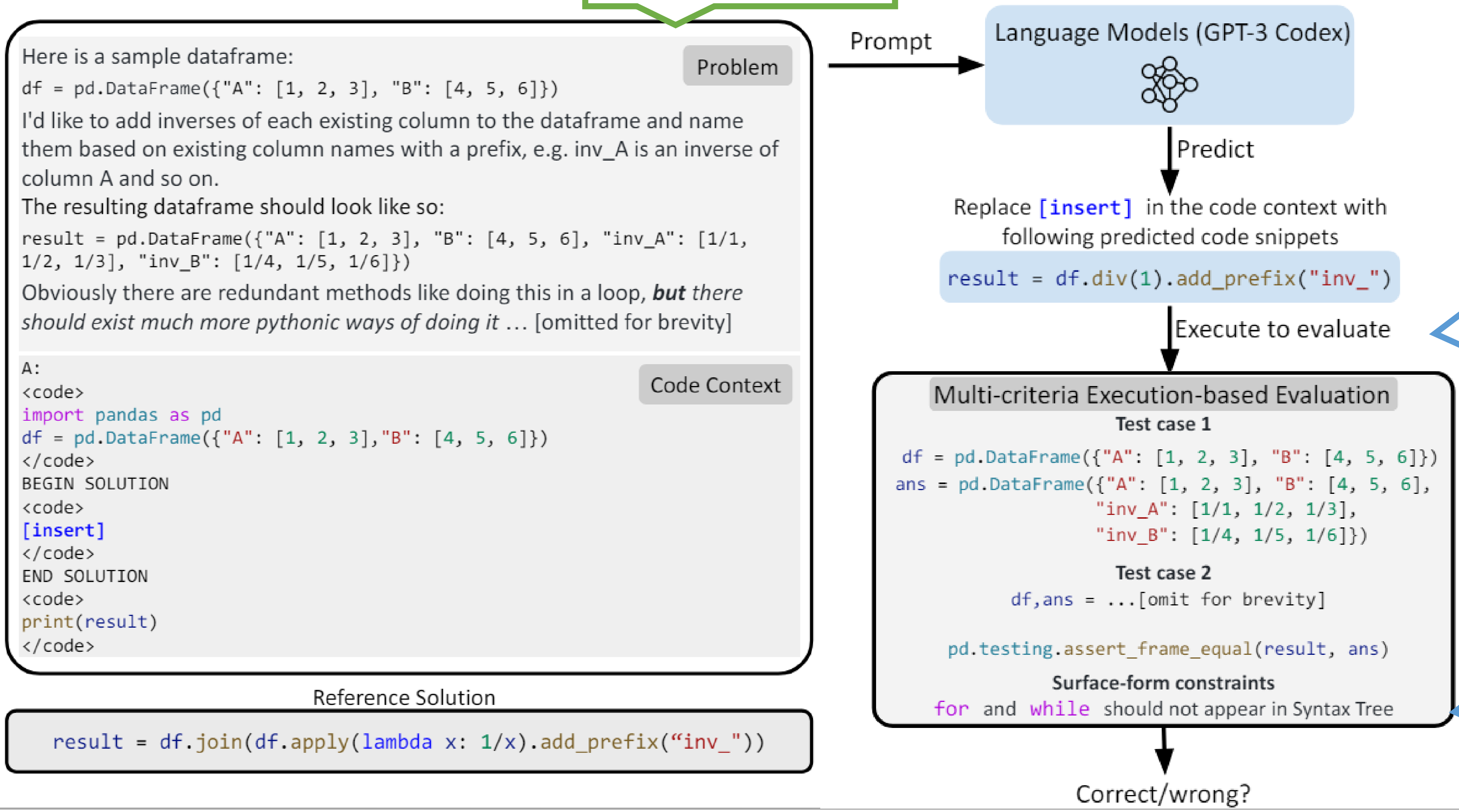
We introduce Codex, a GPT language model fine-tuned on publicly available code from GitHub, and study its Python code-writing capabilities. A distinct production version of Codex powers GitHub Copilot. On HumanEval, a new evaluation set we release to measure functional correctness for synthesizing programs from docstrings, our model solves 28.8% of the problems, while GPT-3 solves 0% and GPT-J solves 11.4% [...]

Publishing a new model + a new benchmark

Source: Chen et al 2021
https://arxiv.org/abs/2107.03374

DS-1000

Modified from stackoverflow



„functional correctness“

„surface-form constraints“

Exam

- 45 min
- 30 points
- (< 30 questions)
- Exam will cover the semester content accordingly
  - Bio-image Analysis / Microscopy
  - Machine/Deep Learning
  - Generative Artificial Intelligence
  - “closed book exam”

https://dbs.uni-leipzig.de/files/2024-02/Modulbeschreibung_BIDS_draft.pdf