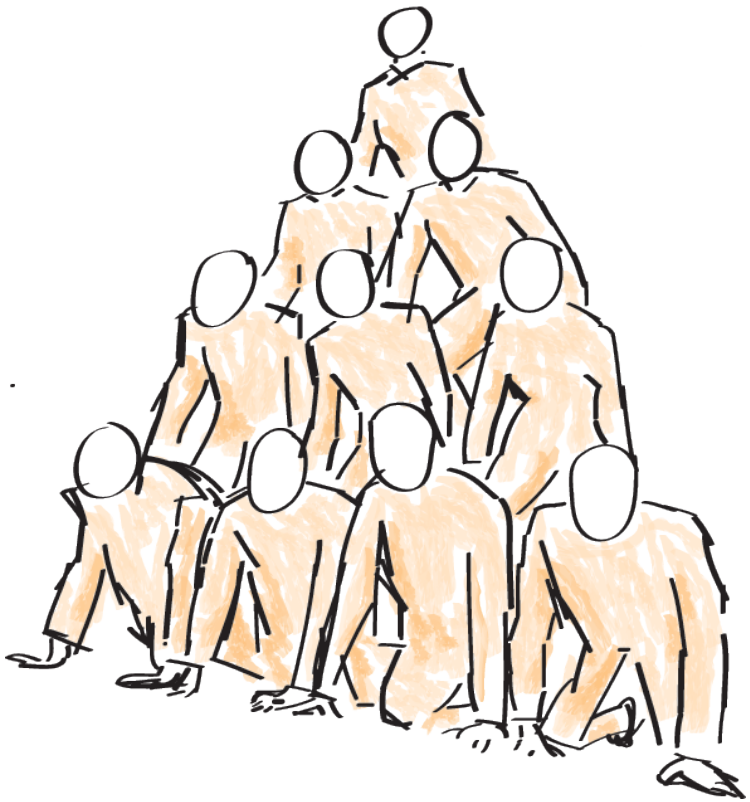
Standing on the Shoulder's of other Researchers - A Position Statement

Ulf Blanke[1], Diane Larlus[1], Kristof Van Laerhoven[2], Bernt Schiele[1,3]

[1]MIS TU Darmstadt, [2]ESS TU Darmstadt, [3]MPI INF Saarbrücken

TECHNISCHE UNIVERSITÄT DARMSTADT

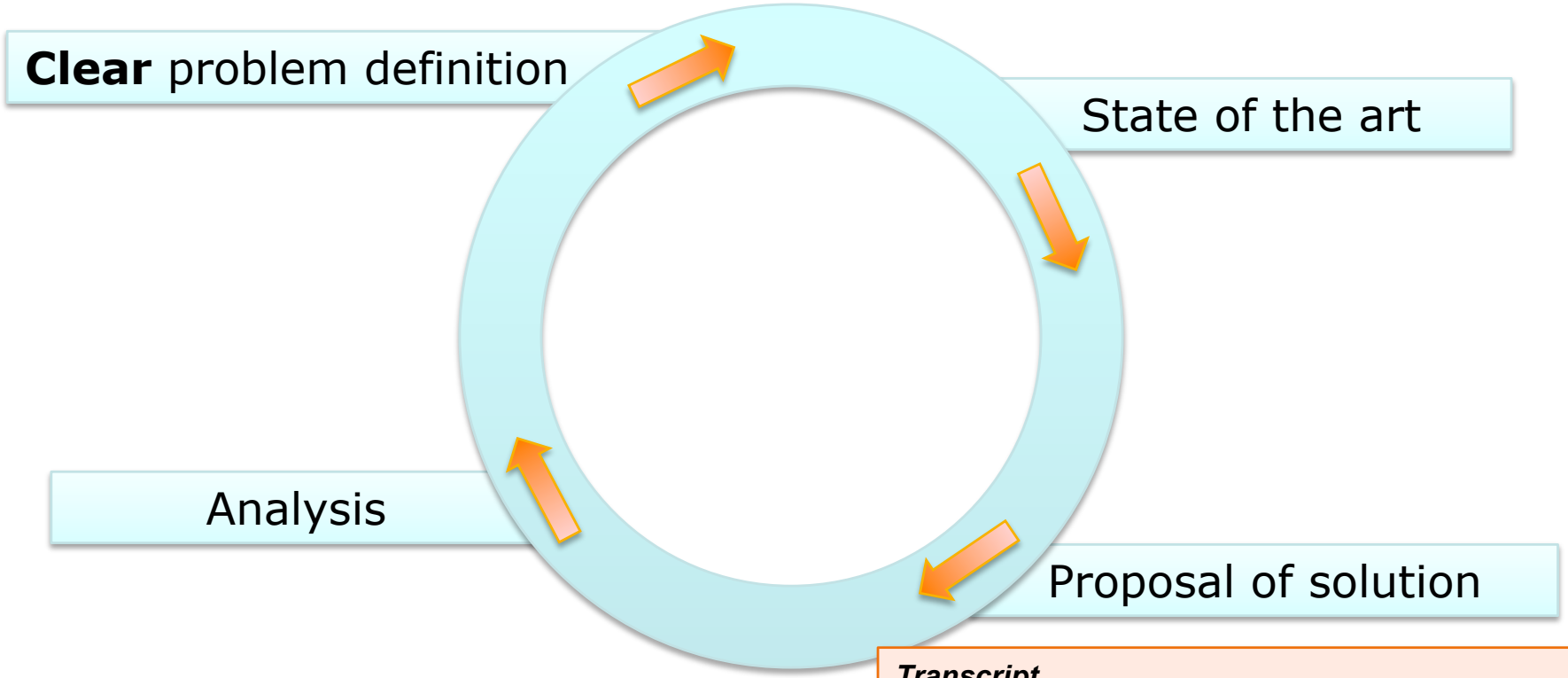


Transcript

- Research group's background:
- Two fields: Activity Recognition & Computer Vision
- We discussed altogether this workshop's topic…

Research Cycle

TECHNISCHE UNIVERSITÄT DARMSTADT



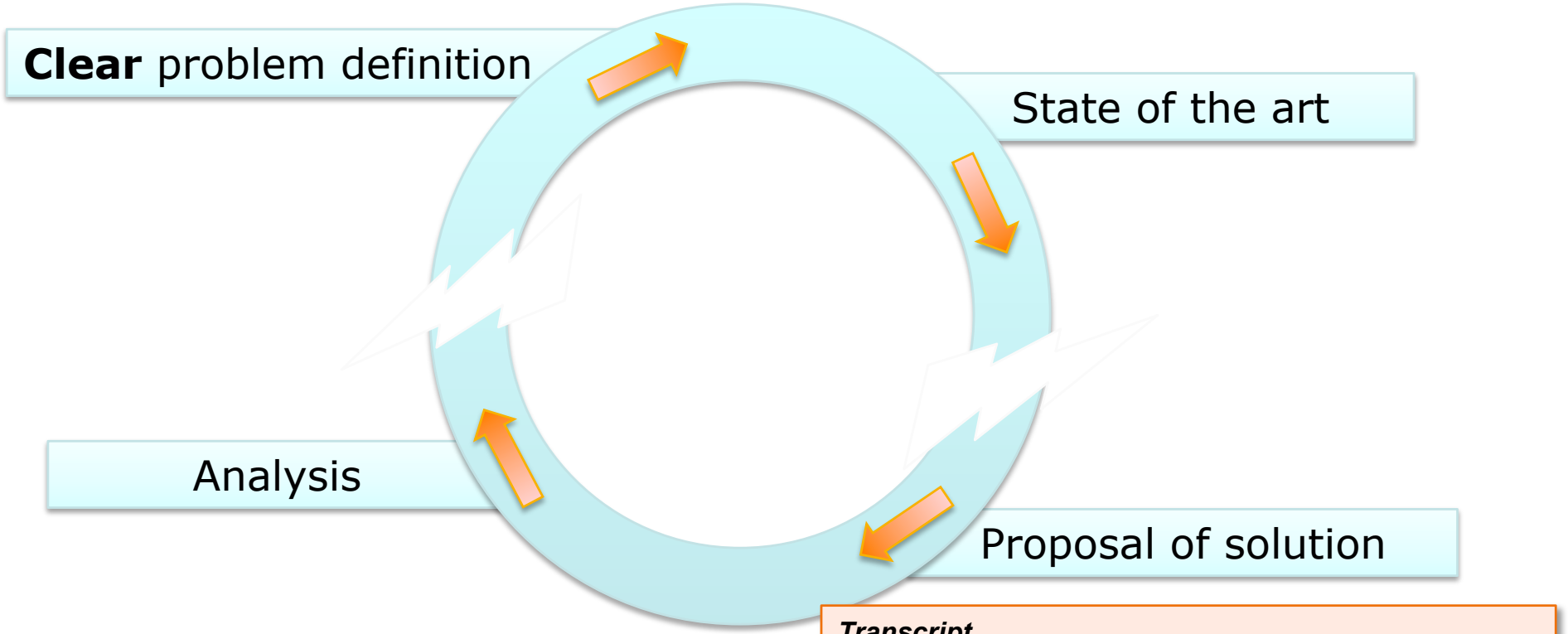
Transcript

- Usual way of thinking research; seems obvious.
- Researchers follow implicitly this process…***individually***

- Looking at bigger picture, i.e., the whole community, …

Research Cycle

TECHNISCHE UNIVERSITÄT DARMSTADT

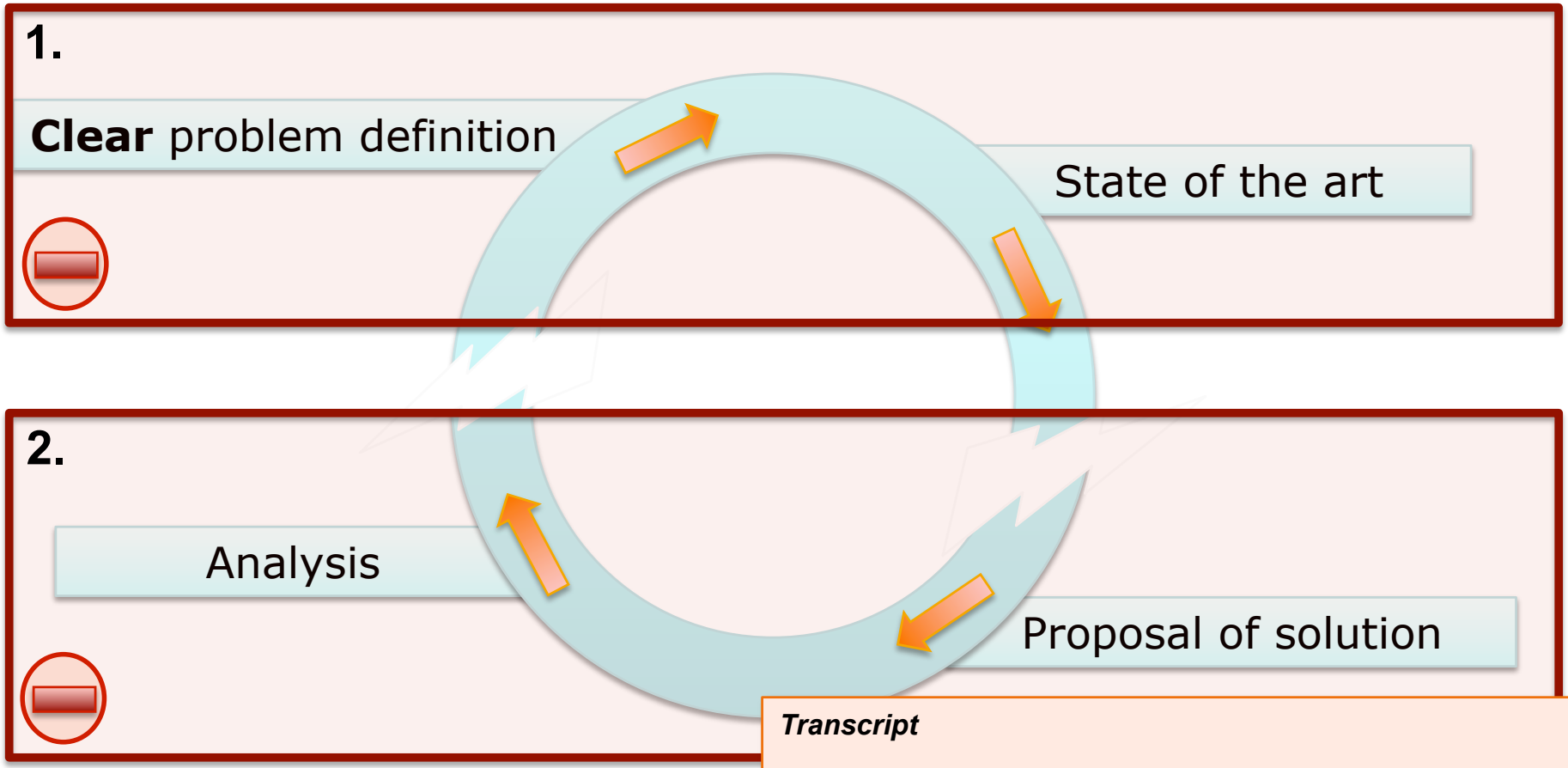


Transcript

…we can observe some ***breaches***.
More specifically, we identify two fronts…

Research Cycle

TECHNISCHE UNIVERSITÄT DARMSTADT



Transcript

1. **Problem** definitions are **not clear** enough
2. **Analysis** often **limited**, disallowing deriving new problems.
   (e.g., good results are emphasized, bad results are hidden → what's left to improve then)

To get the point, let's look at Activities of Daily Living…

1. Research Problems

TECHNISCHE UNIVERSITÄT DARMSTADT

Example: Activities of Daily Living (ADL)

- Definition given
- So far ***some*** of top level ADL well recognized

- Often subset of activities selected
  - Only 3 to 4 out of 6 top-level categories addressed
  - Different activities across different papers
- ADL aim also at assessing quality of activities performed

→ Motivation weak

→ problem statement unsharp

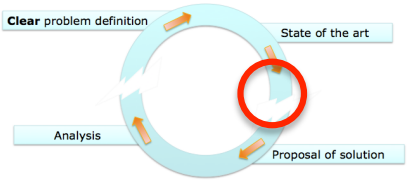
Transcript

…though given a **precise description** of **ADL**, **problem** statements remain **unsharp**, as consequence of inconsistent choice of activities (e.g., one author: brushing teeth, another: showering) and the uncertainty of how solutions impact the application (e.g., quality assessment not addressed, but important for application). Consequently, it's **difficult** to **evaluate solution** proposals to the **state of the art**…

2. Problem solution & analysis

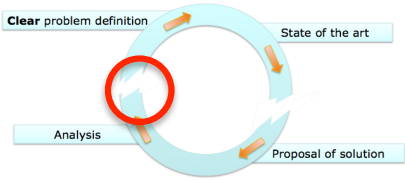
TECHNISCHE UNIVERSITÄT DARMSTADT

Gap at *state of the art* and *proposal of solution*



- hard to evaluate ← sluggish problem definition.
- What matter's?
  e.g., exact time span vs. event (activity spotting)?
- which metrics?

Gap at *analysis* and new *problem definitions*



- ***New algorithms*** often ***paired*** with ***new datasets***
  - shows feasibility but not generality
- strengths of approaches pushed, weaknesses often hidden
- datasets typically not shared

→ insight limited

→ improving approaches difficult

Additional notes

Evaluating state of the art:
Annotation disagreements can be handled ***(see slide 8).***

Analysis
More interesting: ***one*** algorithm ***multiple*** datasets, ***multiple*** algorithms on ***one*** datasets (and not too easy ones) ***(see slide 9)***

Next steps

TECHNISCHE UNIVERSITÄT DARMSTADT

Support ***reproducibility*** and ***comparitative*** results by…

… **clear** problem definition

…Sharing datasets

- Complex enough to allow improvements

…providing evaluation-methodologies

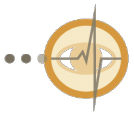
- precise description
- Even better evaluation source-code
- full source code ideal

  (-) clean up, (-) support, (+) others will compare to ***you***

  but not necessary as all you need is

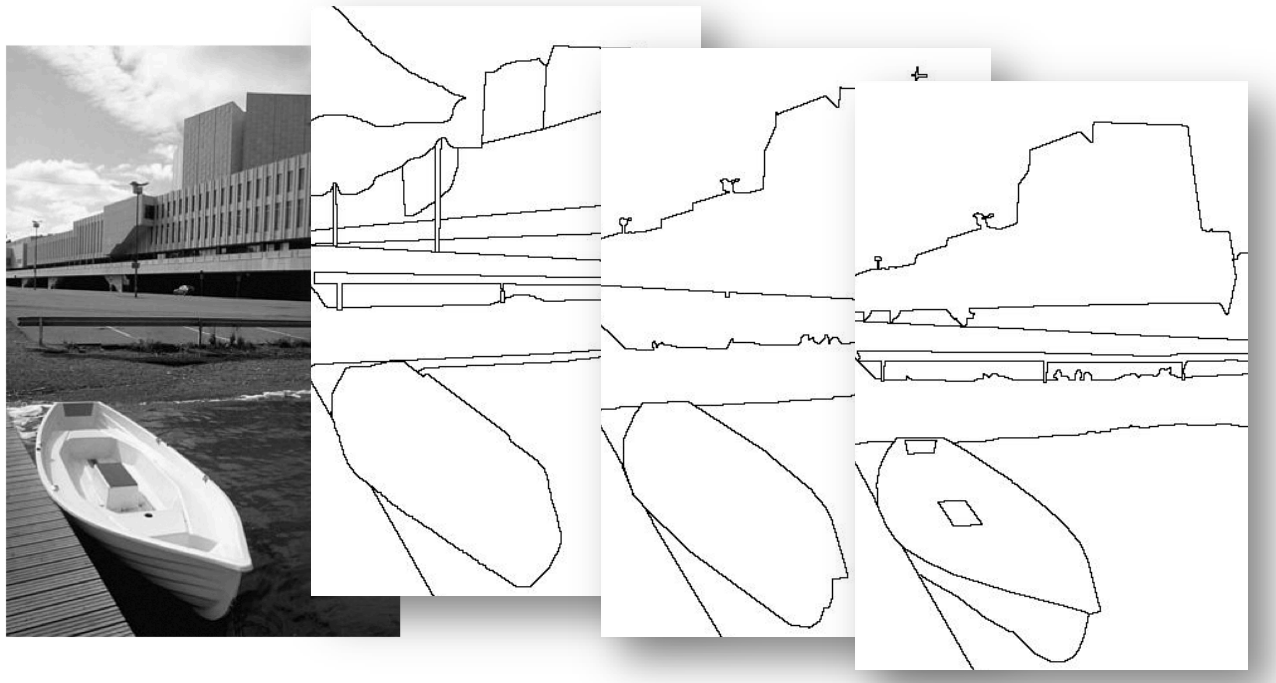
  evaluation method + ***previous results***

Additional notes
We know, we are not at scale to let things evolve implicitly.
We have to invest the effort of explicitly solve our community weaknesses



Other Communities
Example 1: Annotation Disagreement

TECHNISCHE UNIVERSITÄT DARMSTADT



Some activities with potential annotation disagreements:

- hammering, screwing, walking (repetitive)
- drinking, writing, drilling (unclear start/endpoint)

→ What about annotation ***versioning?***

Additional notes
•Different users, creating different annotations (subjective problem perception).
•Solution by creating weighted annotation, which can be evaluated by thresholding.

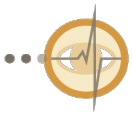
Other Communities
Example 2: Objectiveness by Challenge

TECHNISCHE UNIVERSITÄT DARMSTADT

- ***Host*** provides the evaluation tools
  - Evaluation metrics and scripts
  - labels, training data
  - verification set
  - test data
- ***Participant*** sends predictions

(+) bias-free evaluation criterion across approaches
(+) Establishing reference datasets as baseline for papers
(-) Lot of effort for host
(-) reward limited for host
(-) Bulletproof only with lots of test-data

Additional notes
How to reward the host?
How to reward the participant?



Where are we now…

TECHNISCHE UNIVERSITÄT DARMSTADT

…and what is good or bad?