

How to navigate through the ML research information flood

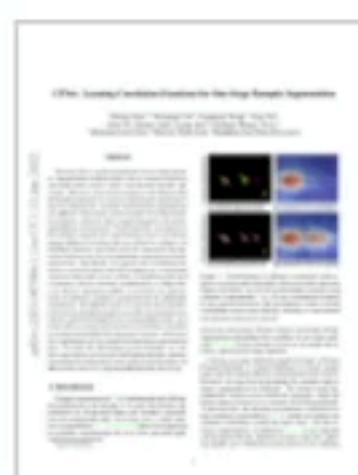
Dmytro Mishkin. Faculty of Electrical Engineering, CTU in Prague



CFNet: Learning Correlation Functions for One-Stage Panoptic Segmentation

Yifeng Chen, Wenqing Chu, Fangfang Wang, Ying Tai, Ran Yi, Zhenye Gan, Liang Yao, Chengjie Wa...

Recently, there is growing attention on one-stage panoptic segmentation methods which aim to segment instances and stuff jointly within a fully convolutional pipeline efficiently. However, most of the existing works directly feed the backbone features to various segmentation heads ignoring the demands for semantic and instance segmentation are different: The former needs semantic-level discriminative features, while the latter requires features to be distinguishable across instances. To alleviate this, we propose to first predict semantic-level and instance-level correlations among different locations that are utilized to enhance the backbone features, and then feed the improved discriminative features into the corresponding segmentation heads,





CFNet: Learning Correlation Functions for One-Stage Panoptic Segmentation

Yifeng Chen, Wenqing Chu, Fangfang Wang, Ying Tai, Ran Yi, Zhenye Gan, Liang Yao, Chengjie Wa...

Recently, there is growing attention on one-stage panoptic segmentation methods which aim to segment instances and stuff jointly within a fully convolutional pipeline efficiently. However, most of the existing works directly feed the backbone features to various segmentation heads ignoring the demands for semantic and instance segmentation are different: The former needs semantic-level discriminative features, while the latter requires features to be distinguishable across instances. To alleviate this, we propose to first predict semantic-level and instance-level correlations among different locations that are utilized to enhance the backbone features, and then feed the improved discriminative features into the corresponding segmentation heads,



Foreword

Foreword



Dmytro Mishkin
@ducha_aiki

...

I am preparing for my lectures on How to navigate CV/ML literature, how to read papers in an efficient way, re-implement them, etc.

Which issues in this area do you personally have?
Questions? What is the most misleading advice you ever get?

Please write in comments below

[Перекласти твіт](#)

5:13 пп · 15 січ. 2022 · Twitter for Android

||| Переглянути дії з твітом

7 Ретвітів 1 Цитувати твіт 135 Уподобань



Foreword



Dmytro Mishkin
@ducha_aiki

...

I am preparing for my lectures on How to navigate CV/ML literature, how to read papers in an efficient way, re-implement them, etc.
Which issues in this area do you personally have?
Questions? What is the most misleading advice you ever get?
Please write in comments below

[Перекласти твіт](#)

5:13 пп · 15 січ. 2022 · Twitter for Android

||| Переглянути дії з твітом

7 Ретвітів 1 Цитувати твіт 135 Уподобань



Foreword



Dmytro Mishkin
@ducha_aiki

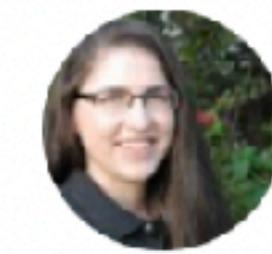
I am preparing for my lectures on How to navigate CV/ML literature, how to read papers in an efficient way, re-implement them, etc.
Which issues in this area do you personally have?
Questions? What is the most misleading advice you ever get?
Please write in comments below

[Перекласти твіт](#)

5:13 пп · 15 січ. 2022 · Twitter for Android

||| Переглянути дії з твітом

7 Ретвітів 1 Цитувати твіт 135 Уподобань



Amy Tabb
@amy_tabb

У відповідь @ducha_aiki

Misleading advice: do *whatever* exactly the way someone else does it, whether it be read 10 papers a day, or none.

Your strategy should be allowed to change depending on the needs of the research problem, your own constraints & preferences, etc., and that's ok.

[Перекласти твіт](#)

5:27 пп · 15 січ. 2022 · TweetDeck

1 Ретвіт 9 Уподобань



Foreword: trust noone in ML

I mean, really

Foreword: trust noone in ML

I mean, really

- You should develop your own style of reading ML papers, don't trust me

Foreword: trust noone in ML

I mean, really

- You should develop your own style of reading ML papers, don't trust me
- and never trust authors of the paper

Foreword: trust noone in ML

I mean, really

- You should develop your own style of reading ML papers, don't trust me
 - and never trust authors of the paper
 - ...even if it was published on CVPR

Foreword: trust noone in ML

I mean, really

- You should develop your own style of reading ML papers, don't trust me
- and never trust authors of the paper
- ...even if it was published on CVPR
- ...especially if it was published on CVPR

Foreword: trust noone in ML

I mean, really

- You should develop your own style of reading ML papers, don't trust me
- and never trust authors of the paper
- ...even if it was published on CVPR
- ...especially if it was published on CVPR

Naive beliefs about publication

- Published papers are right, and good
- Rejected papers are wrong, or bad

<https://www.youtube.com/watch?v=9Y7NCdKdNyE>



Foreword: trust noone in ML

I mean, really

- You should develop your own style of reading ML papers, don't trust me
- and never trust authors of the paper
- ...even if it was published on CVPR
- ...especially if it was published on CVPR
- OK, let's go

Naive beliefs about publication

- Published papers are right, and good
- Rejected papers are wrong, or bad



<https://www.youtube.com/watch?v=9Y7NCdKdNyE>

What do I want to achieve?

Why do you need to read papers?

Use-cases

Why do you need to read papers?

Use-cases

- (a) You are new to the field and want to get familiar with it

Why do you need to read papers?

Use-cases

- (a) You are new to the field and want to get familiar with it
- (b) You are new to the field and need to implement something ASAP

Why do you need to read papers?

Use-cases

- (a) You are new to the field and want to get familiar with it
- (b) You are new to the field and need to implement something ASAP
- (c) You are familiar with the field, but not the specific area in particular

Why do you need to read papers?

Use-cases

- (a) You are new to the field and want to get familiar with it
- (b) You are new to the field and need to implement something ASAP
- (c) You are familiar with the field, but not the specific area in particular
- (d) You know area well enough and want to keep yourself up-to-date

Why do you need to read papers?

Use-cases

- (a) You are new to the field and want to get familiar with it
- (b) You are new to the field and need to implement something ASAP
- (c) You are familiar with the field, but not the specific area in particular
- (d) You know area well enough and want to keep yourself up-to-date
- (e) I don't want, but I have to - I am CVPR reviewer

Why do you need to read papers?

Use-cases

- (a) You are new to the field and want to get familiar with it
- (b) You are new to the field and need to implement something ASAP
- (c) You are familiar with the field, but not the specific area in particular
- (d) You know area well enough and want to keep yourself up-to-date
- (e) I don't want, but I have to - I am CVPR reviewer

Each of this cases requires a little bit different strategy

You are new to the field and need to implement something ASAP (like, yesterday)

- That is really unfortunate situation.
- The only way to succeed is to be lucky and find ready implementation
- Check the following first:
 - Kaggle & DrivenData competitions
 - Papers With Code
 - fast.ai forums, pyimagesearch, etc.
 - Only then browse papers and github.

Kaggle

Kaggle
Competitions → keyword search → EDA kernels

Kaggle

Competitions → keyword search → EDA kernels

The screenshot shows the Kaggle homepage with the sidebar open. The sidebar includes links for 'kaggle', 'Create', 'Home', 'Competitions' (which is selected and highlighted in grey), 'Datasets', 'Code', 'Discussions', 'Courses', and 'More'. Below these are sections for 'Your Work', 'RECENTLY VIEWED' (with items like 'notebook_HPA_EDA', 'Human Protein Atlas', '1st place solution', and 'Google Landmark Retr...'), and 'RECENTLY EDITED' (with 'notebookd63442bf65'). At the bottom of the sidebar are 'View Active Events' and a 'Search' bar containing 'segmentation'.

The main content area is titled 'Competitions' and shows a search bar with 'segmentation' and a 'Filters' button. A 'Completed' filter is applied. The results are listed under 'Results' with a 'Recently Launched' dropdown. The results are as follows:

Competition	Description	Prize
Sartorius - Cell Instance Segmentation	Detect single neuronal cells in microscopy images SARTORIUS Featured · Code Competition · 1505 Teams · 17 days ago	\$75,000
HuBMAP - Hacking the Kidney	Identify glomeruli in human kidney tissue images HuBMAP Research · Code Competition · 1200 Teams · 8 months ago	\$60,000
Open Images Instance Segmentation RVC 2020 edition	Outline segmentation masks of objects in images Google Playground · 18 Teams · a year ago	Knowledge
iMaterialist (Fashion) 2020 at FGVC7	Fine-grained segmentation task for fashion and apparel iMaterialist Research · 56 Teams · 2 years ago	Knowledge
Open Images 2019 - Instance Segmentation	Outline segmentation masks of objects in images Google Research · 193 Teams · 2 years ago	\$25,000

Kaggle

Competitions → keyword search → EDA kernels



The screenshot shows the Kaggle interface. On the left is a sidebar with navigation links: kaggle, Create, Home, Competitions (which is selected), Datasets, Code, Discussions, Courses, More, Your Work, Recently Viewed, Recently Edited, and View Active Events. The main area is titled 'Competitions'. At the top right is a 'Your Work' button. Below it is a search bar with the query 'segmentation' and a 'Filters' button. A 'Completed' filter is applied. The results are listed under 'Results' with a 'Recently Launched' dropdown. There are five competition entries:

- Sartorius - Cell Instance Segmentation: \$75,000, Recently Launched, Sartorius logo, Detect single neuronal cells in microscopy images, Featured - Code Competition - 1505 Teams - 17 days ago.
- HuBMAP - Hacking the Kidney: \$60,000, Research - Code Competition - 1200 Teams - 9 months ago, Identify glomeruli in human kidney tissue images.
- Open Images Instance Segmentation RVC 2020 edition: Knowledge, Outline segmentation masks of objects in images, Playground - 18 Teams - a year ago.
- iMaterialist (Fashion) 2020 at FGVC7: Knowledge, Fine-grained segmentation task for fashion and apparel, Research - 56 Teams - 2 years ago.
- Open Images 2019 - Instance Segmentation: Knowledge, Outline segmentation masks of objects in images, Research - 193 Teams - 2 years ago.

Kaggle

Competitions → keyword search → EDA kernels

The screenshot shows a Kaggle competition page for 'Sartorius - Cell Instance Segmentation'. The page includes a sidebar with navigation links like Home, Competitions, Datasets, and Courses. A search bar at the top right allows filtering by 'segmentation'. The main area displays a list of competitions, with the top one being the current focus. The kernel 'EDA+EffDET [TF]' has 242 upvotes and 277 comments. It features a large code block with a redacted section containing the text '... RELEVANT DATAFRAME ROW - INDEX=42 ...'. Below the code is a table with columns: id, annotation, width, height, cell_type, plate_time, sample_date, sample_id, elapsed_timedelta, and im. The row for index 42 shows the following values:

	id	annotation	width	height	cell_type	plate_time	sample_date	sample_id	elapsed_timedelta	im
42	13325f865bb0	[195920 2 196623 5 197327 6 198030 7 198733 7 ...]	704	520	astro	09h00m00s	2020-09-14	astros[cereb]_G12- 2_Vessel- 361_Ph_4	0 days 09:00:00	/ki ce se

At the bottom, three images illustrate the segmentation process: 'Cell Image' (original grayscale microscopy image), 'Instance Segmentation Mask' (a binary mask where each cell is a different color), and 'Cell Image w/ Instance Segmentation Mask Overlay' (the original image with the colored masks overlaid).

Kaggle

If task & competition is relevant to you, check solutions

- Winning solutions
- Simple solutions

The screenshot shows the competition page for "Google Landmark Retrieval 2021". It features a large image of a bridge over water. The title "Google Landmark Retrieval 2021" is displayed, along with the subtitle "Given an image, can you find all of the same landmarks in a dataset?". Below the image, it says "Google · 263 teams · 3 months ago". The navigation bar at the top includes "Overview", "Data", "Code", "Discussion", "Leaderboard", and "Rules". The "Discussion" tab is currently selected.

The screenshot shows a post by "Dieter" titled "1st place solution", which was posted 3 months ago in the "landmark-retrieval-2021" discussion. The post begins with the text: "First, let me thank kaggle staff and google team for organ still overwhelmed and shocked by my result, which resulted in a 1st place on the leaderboar, an ultimate goal for many kagglers. My solut". To the right, there is a detailed description of the solution architecture and its performance.

The screenshot shows a post by "takuoko" titled "1st place solution", which was posted 12 days ago in the "sartorius-cell-instance-segmentation" discussion. The post starts with a thank you message to Kaggle and Sartorius. It then provides an "Overview" of the solution, stating that it is similar to the 2nd place solution. A flowchart diagram illustrates the detection and segmentation pipeline. The post concludes with a detailed explanation of the competition's validation metric and the validation scores achieved.

DrivenData

- Much smaller community than Kaggle, with much less forum life
- But contains some unique challenges

PHASE 1 | Facebook AI Image Similarity Challenge:
Matching Track
HOSTED BY FACEBOOK

TEAM

DISCUSSION

30

Problem description

Summary

Snowcast Showdown: Development Stage
HOSTED BY BUREAU OF RECLAMATION

HOME PROBLEM DESCRIPTION DEVELOPMENT STAGE REPORT TEMPLATE ABOUT

You will receive a reference set of 1 n images are derived from images in th

For this **Matching Track**, your task is one of the images in a large corpus o

User or team	Best public ↓ RMSE ⓘ	R ² ⓘ	Timestamp ⓘ	Trend (last 10)	# Entries
rasyidstat	3.4061	0.8305	2022-01-13 04:53:49		12
FBykov	3.5919	0.8115	2021-12-29 17:13:44		75
Team UA	3.6797	0.8021	2022-01-13 22:47:59		14
separate	3.7156	0.7982	2022-01-13 23:54:41		36
andrey1362010	3.8857	0.7793	2022-01-14 14:57:01		14
NxGTR	4.0837	0.7563	2022-01-14 13:25:01		16
Galeros93	4.1490	0.7484	2022-01-02 22:21:21		8

Email

Papers With Code



Browse SoTA > Computer Vision

Computer Vision

2594 benchmarks • 934 tasks • 1702 datasets • 22504 papers with code

Semantic Segmentation

Semantic Segmentation 206 benchmarks 2411 papers with code	Medical Image Segmentation 86 benchmarks 254 papers with code	Tumor Segmentation 1 benchmark 113 papers with code	Panoptic Segmentation 10 benchmarks 77 papers with code	3D Semantic Segmentation 8 benchmarks 76 papers with code
--	---	---	---	---

▶ See all 19 tasks

Image Classification

Image	Knowledge	Few-Shot Image	Fine-Grained	Semi-Supervised
-------	-----------	----------------	--------------	-----------------

Papers With Code

Caveat: sparsely populated

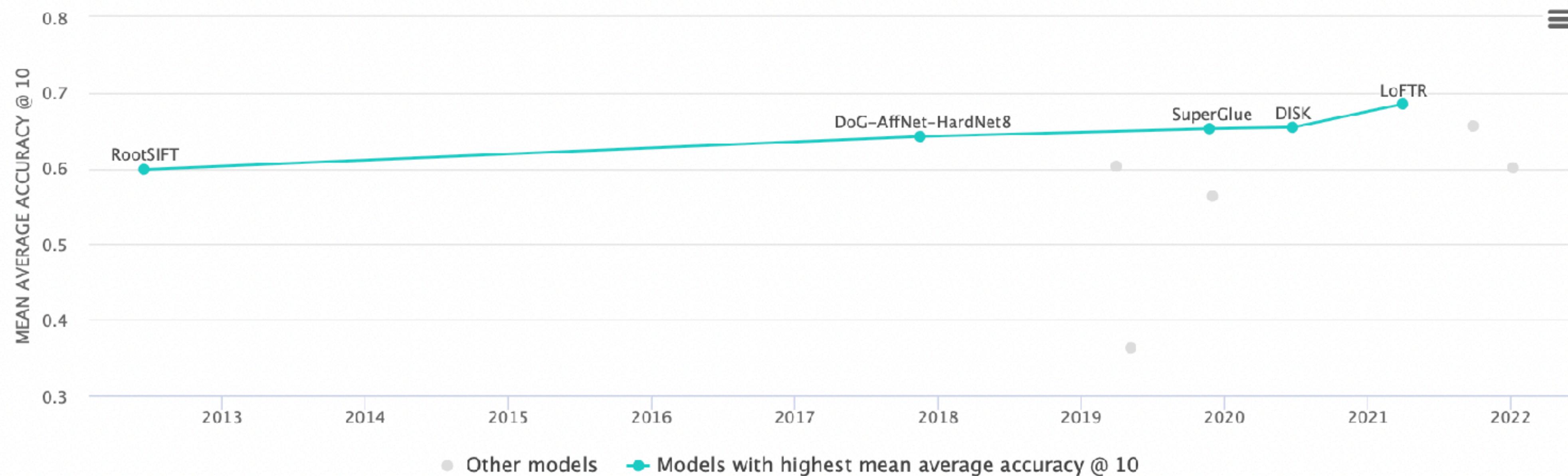
Image Matching on IMC PhotoTourism

Leaderboard Dataset

View mean average accuracy @ 10

by Date

for All models

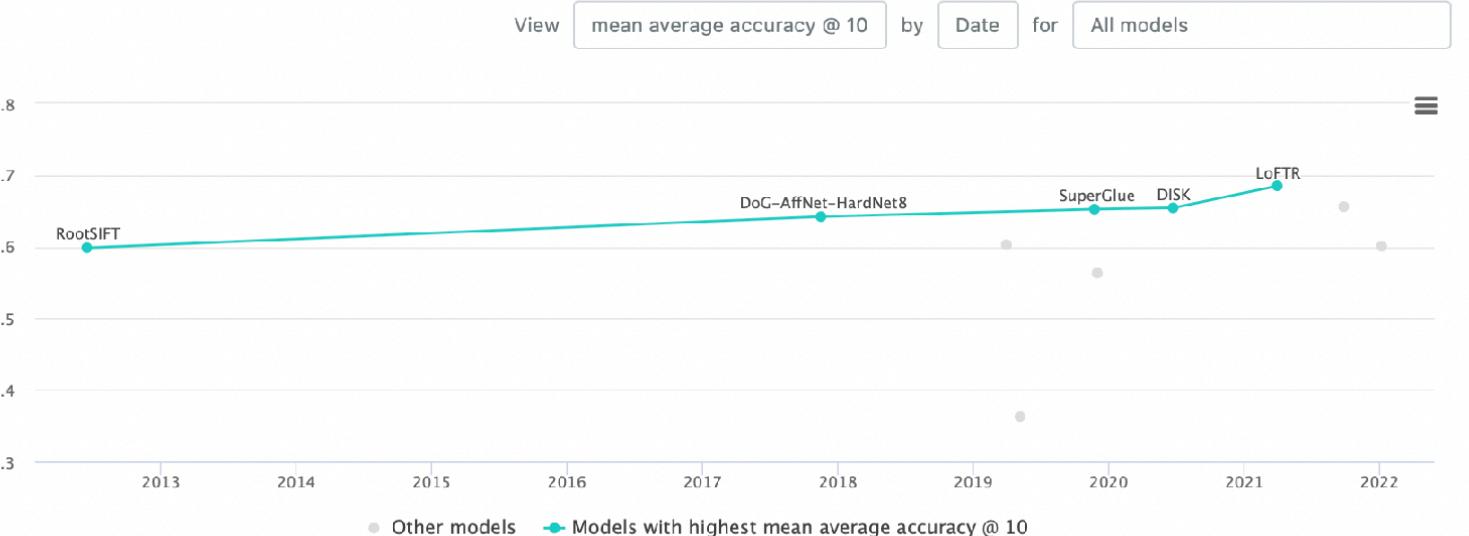


Papers With Code

Caveat: sparsely populated

Image Matching on IMC PhotoTourism

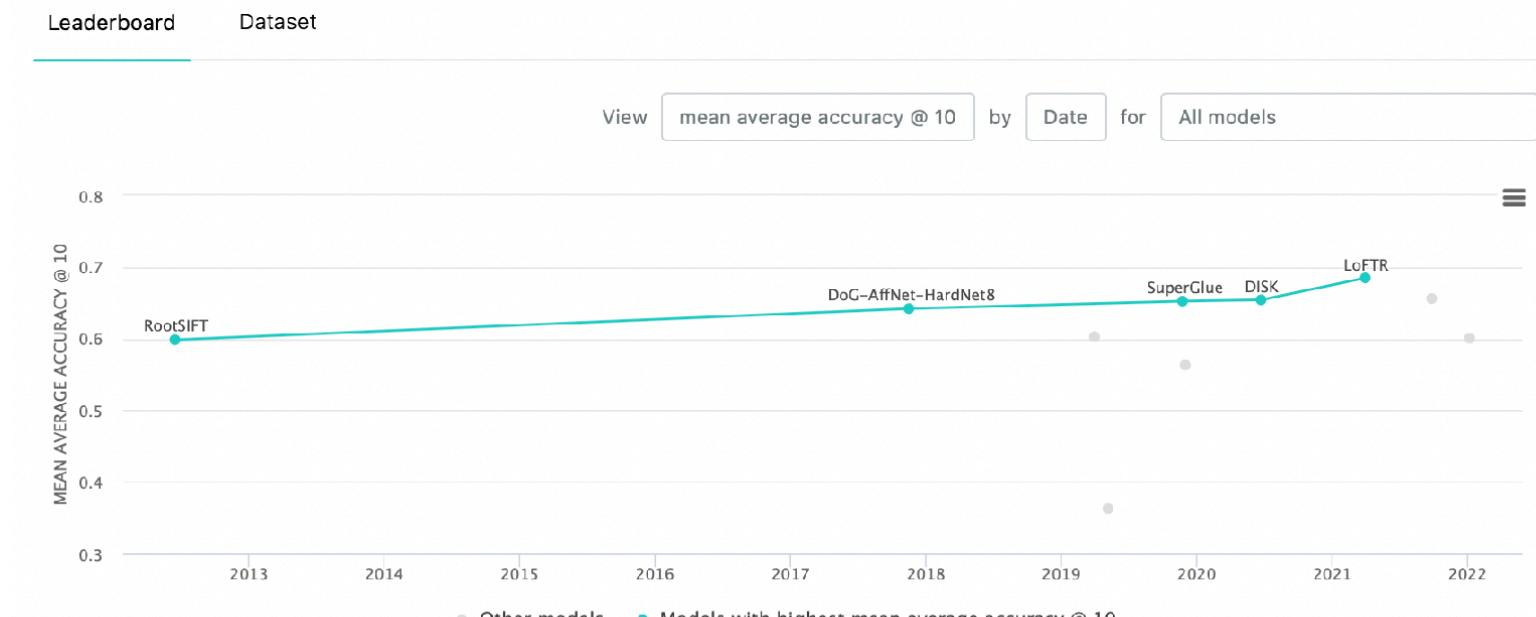
Leaderboard Dataset



Papers With Code

Caveat: sparsely populated

Image Matching on IMC PhotoTourism



Rank	Model	mean average accuracy @ 10	Extra Training Data	Paper	Code	Result	Year	Tags
1	LoFTR	0.68503	✓	LoFTR: Detector-Free Local Feature Matching with Transformers			2021	
2	HarrisZ improved, Blob DTM	0.65606	✗	HarrisZ⁺: Harris Corner Selection for Next-Gen Image Matching Pipelines			2021	
3	DISK	0.65435	✓	DISK: Learning local features with policy gradient			2020	
4	SuperGlue	0.65248	✓	SuperGlue: Learning Feature Matching with Graph Neural Networks			2019	
5	DoG-AffNet-HardNet8	0.64212	✓	Repeatability Is Not Enough: Learning Affine Regions via Discriminability			2017	
6	Key.Net-SOSNet	0.60285	✗	Key.Net: Keypoint Detection by Handcrafted and Learned CNN Filters			2019	
7	PoFeat	0.60072	✗	Decoupling Makes Weakly Supervised Local Feature Better			2022	
8	RootSIFT	0.59859	✗	Three things everyone should know to improve object retrieval			2012	
9	R2D2	0.56345	✓	R2D2: Reliable and Repeatable Detector and Descriptor			2019	

Papers With Code

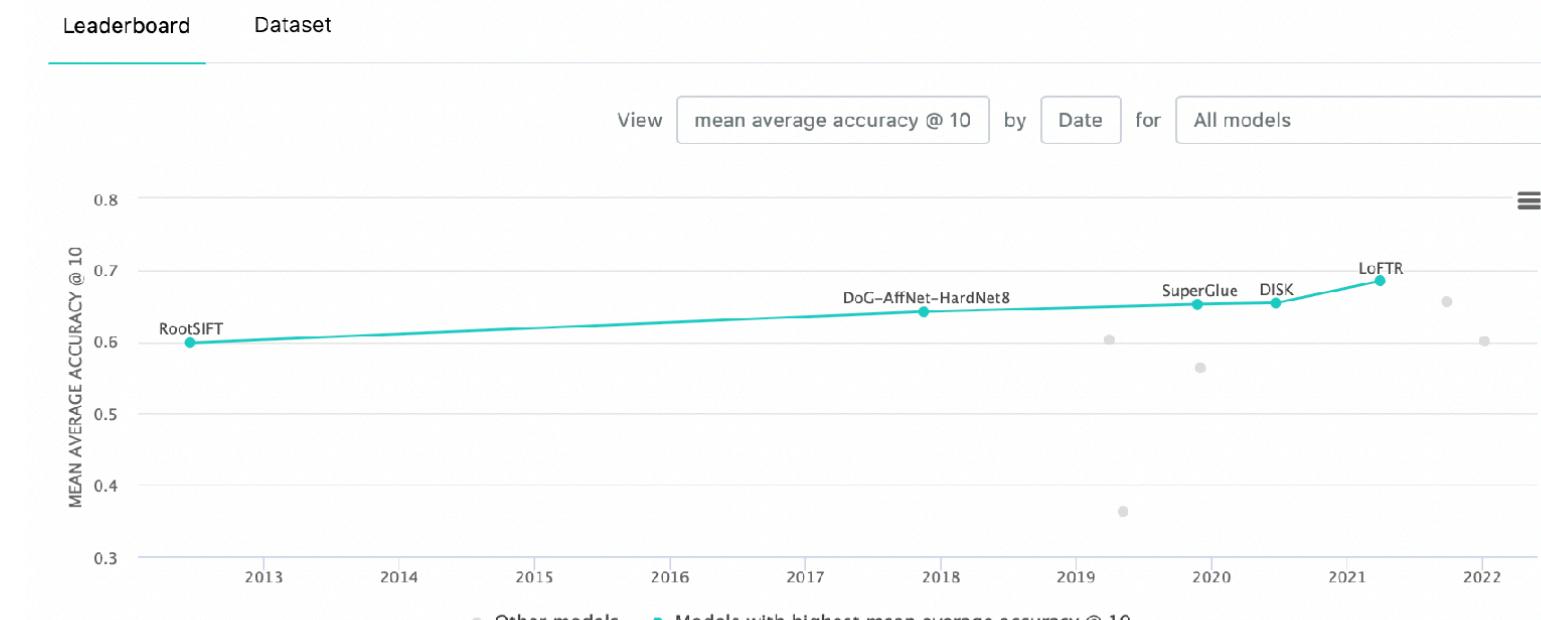
Caveat: sparsely populated

I have added this whole leaderboard

when preparing for this talk.

Before there was no leaderboard, and nobody checked it

Image Matching on IMC PhotoTourism



Rank	Model	mean average accuracy @ 10	Extra Training Data	Paper	Code	Result	Year	Tags
1	LoFTR	0.68503	✓	LoFTR: Detector-Free Local Feature Matching with Transformers			2021	
2	HarrisZ improved, Blob DTM	0.65606	✗	HarrisZ ⁺ : Harris Corner Selection for Next-Gen Image Matching Pipelines			2021	
3	DISK	0.65435	✓	DISK: Learning local features with policy gradient			2020	
4	SuperGlue	0.65248	✓	SuperGlue: Learning Feature Matching with Graph Neural Networks			2019	
5	DoG-AffNet-HardNet8	0.64212	✓	Repeatability Is Not Enough: Learning Affine Regions via Discriminability			2017	
6	Key.Net-SOSNet	0.60285	✗	Key.Net: Keypoint Detection by Handcrafted and Learned CNN Filters			2019	
7	PoSFeat	0.60072	✗	Decoupling Makes Weakly Supervised Local Feature Better			2022	
8	RootSIFT	0.59859	✗	Three things everyone should know to improve object retrieval			2012	
9	R2D2	0.56345	✓	R2D2: Reliable and Repeatable Detector and Descriptor			2019	

Papers With Code

Caveat: sparsely populated

 Methodology

Representation Learning

 Edit

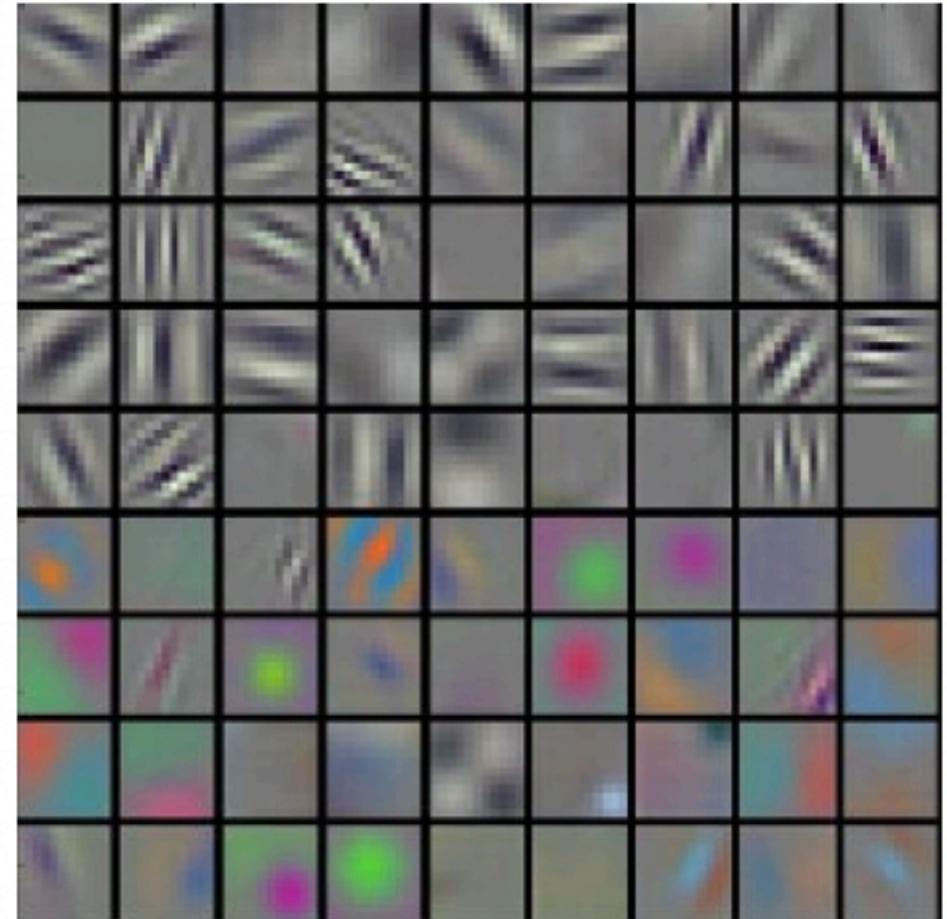
1643 papers with code • 2 benchmarks • 2 datasets

Representation learning is concerned with training machine learning algorithms to learn useful representations, e.g. those that are interpretable, have latent features, or can be used for transfer learning.

Deep neural networks can be considered representation learning models that typically encode information which is projected into a different subspace. These representations are then usually passed on to a linear classifier to, for instance, train a classifier.

Representation learning can be divided into:

- **Supervised representation learning:** learning representations on task A using annotated data and used to solve task B



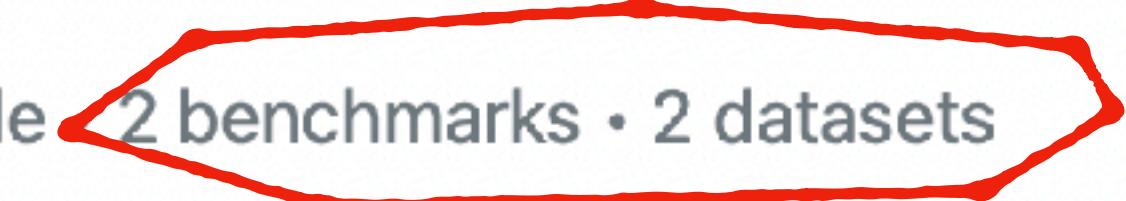
Content

Papers With Code

Caveat: sparsely populated

 Methodology

Representation Learning

1643 papers with code  2 benchmarks • 2 datasets

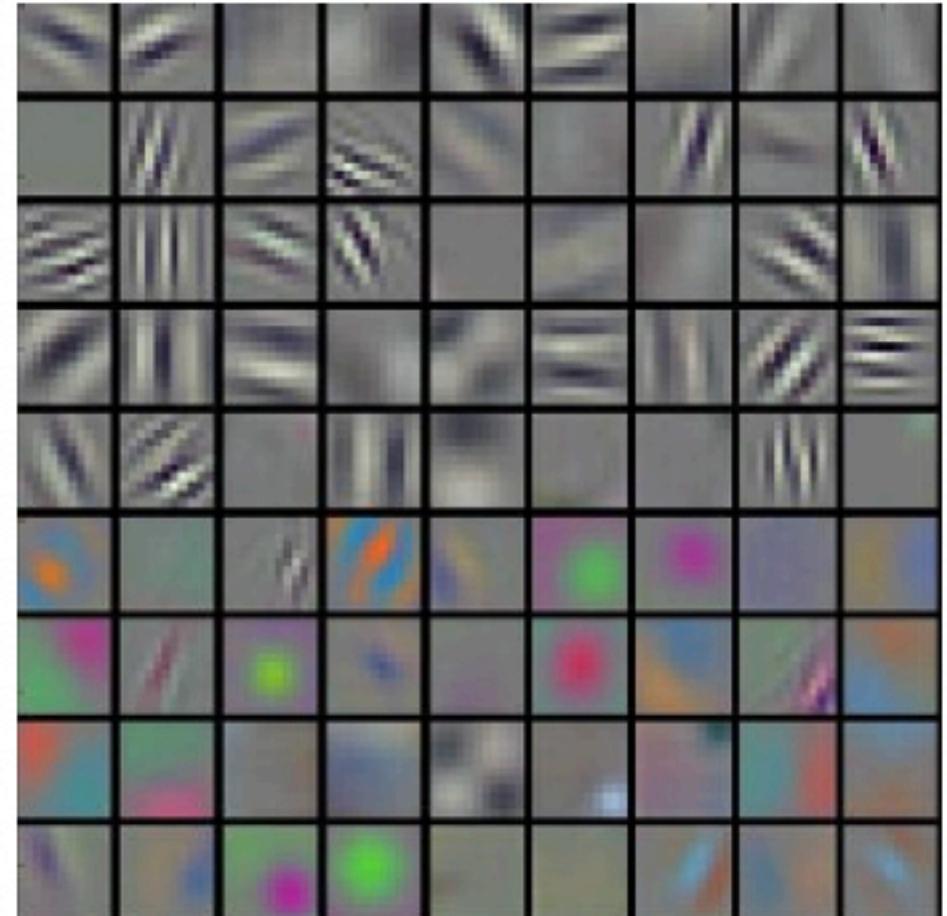
 Edit

Representation learning is concerned with training machine learning algorithms to learn useful representations, e.g. those that are interpretable, have latent features, or can be used for transfer learning.

Deep neural networks can be considered representation learning models that typically encode information which is projected into a different subspace. These representations are then usually passed on to a linear classifier to, for instance, train a classifier.

Representation learning can be divided into:

- **Supervised representation learning:** learning representations on task A using annotated data and used to solve task B



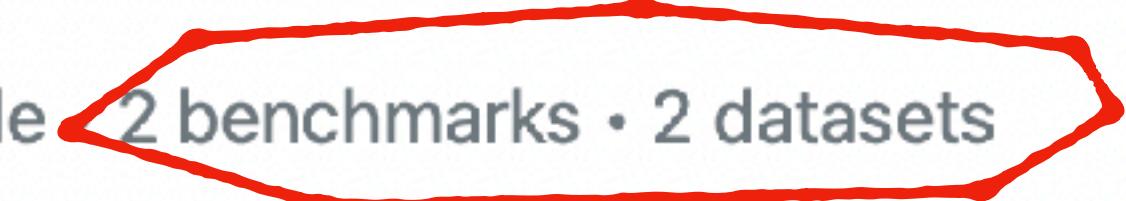
Content

Papers With Code

Caveat: sparsely populated

 Methodology

Representation Learning

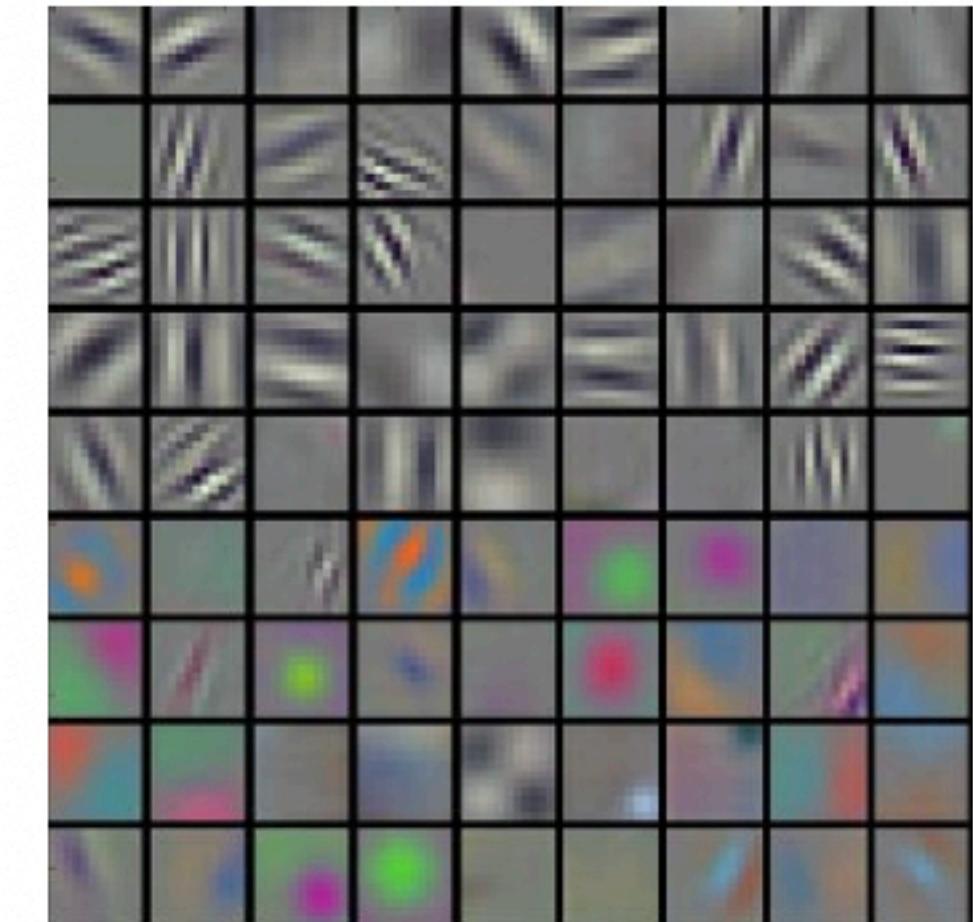
1643 papers with code  2 benchmarks • 2 datasets

Representation learning is concerned with training machine learning algorithms to learn useful representations, e.g. those that are interpretable, have latent features, or can be used for transfer learning.

Deep neural networks can be considered representation learning models that typically encode information which is projected into a different subspace. These representations are then usually passed on to a linear classifier to, for instance, train a classifier.

Representation learning can be divided into:

- **Supervised representation learning:** learning representations on task A using annotated data and used to solve task B



Content

Papers With Code

Caveat: sparsely populated

 Methodology

Representation Learning

 Edit

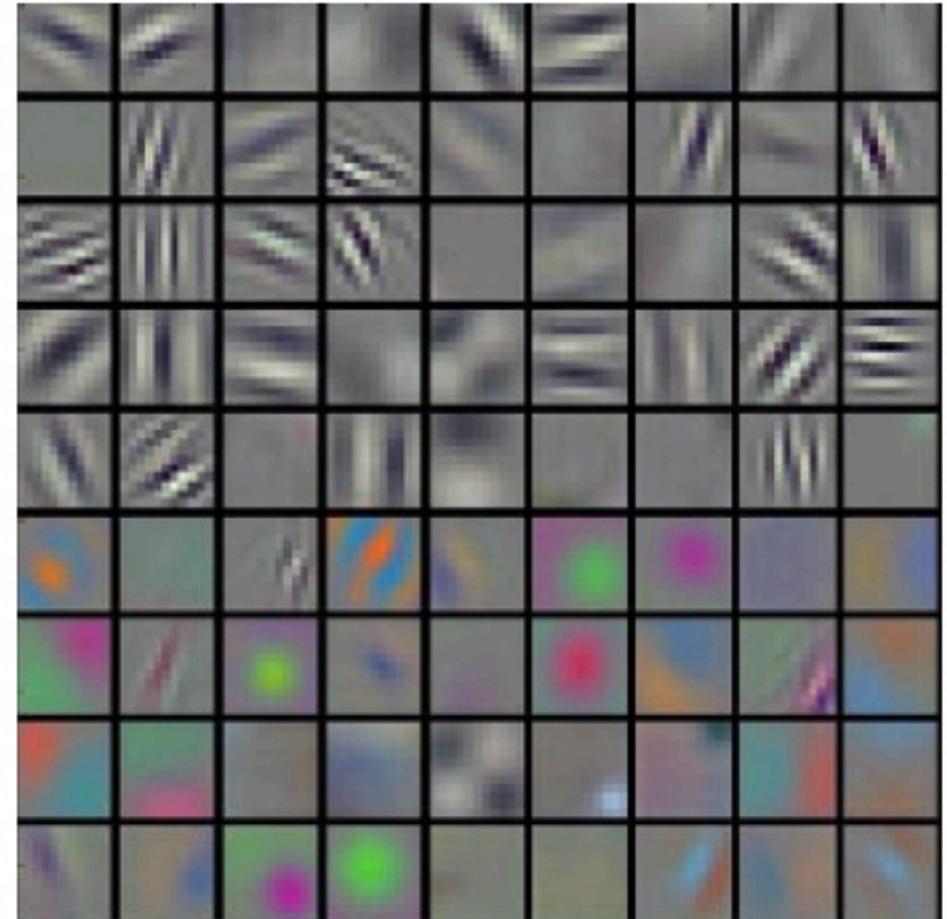
1643 papers with code  2 benchmarks • 2 datasets

Representation learning is concerned with training machine learning algorithms to learn useful representations, e.g. those that are interpretable, have latent features, or can be used for transfer learning.

Deep neural networks can be considered representation learning models that typically encode information which is projected into a different subspace. These representations are then usually passed on to a linear classifier to, for instance, train a classifier.

Representation learning can be divided into:

- **Supervised representation learning:** learning representations on task A using annotated data and used to solve task B



Content

Papers With Code

Caveat: sparsely populated

Methodology

Representation Learning

1643 papers with code 2 benchmarks • 2 datasets

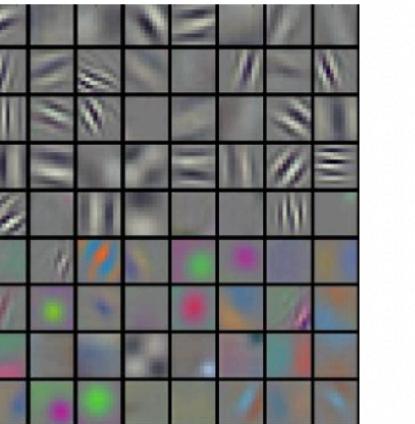
Representation learning is concerned with training machine learning algorithms to learn useful representations, e.g. those that are interpretable, have latent features, or can be used for transfer learning.

Deep neural networks can be considered representation learning models that typically encode information which is projected into a different subspace. These representations are then usually passed on to a linear classifier to, for instance, train a classifier.

Representation learning can be divided into:

- **Supervised representation learning:** learning representations on task A using annotated data and used to solve task B

Edit



Content

Papers With Code

Caveat: sparsely populated

Methodology

Representation Learning

1643 papers with code 2 benchmarks • 2 datasets

Representation learning is concerned with training machine learning representations, e.g. those that are interpretable, have latent feature learning.

Deep neural networks can be considered representation learning information which is projected into a different subspace. These representations are passed on to a linear classifier to, for instance, train a classifier.

Representation learning can be divided into:

- **Supervised representation learning:** learning representations on task A and used to solve task B

View

Benchmarks

Add a Result

These leaderboards are used to track progress in Representation Learning

Trend	Dataset	Best Model	Paper	Code	Compare
	Circle Data	🏆 Morphological Network			See all
	Sports10	🏆 Max Margin Contrastive			See all

Papers With Code

Caveat: sparsely populated

Methodology

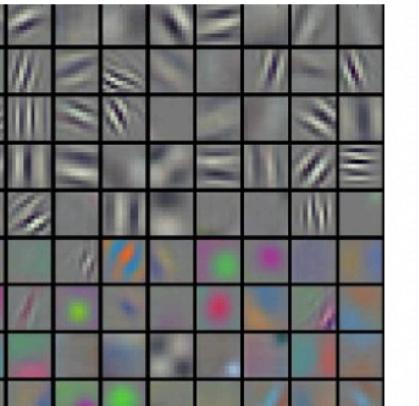
Representation Learning

1643 papers with code • 2 benchmarks • 2 datasets

Representation learning is concerned with training machine learning algorithms to learn useful representations, e.g. those that are interpretable, have latent features, or can be used for transfer learning.

Deep neural networks can be considered representation learning models that typically encode information which is projected into a different subspace. These representations are then usually passed on to a linear classifier to, for instance, train a classifier.

Edit



Benchmarks

Add a Result

These leaderboards are used to track progress in Representation Learning

Trend Dataset Best Model Paper Code Compare



Circle Data

🏆 Morphological Network



See all



Sports10

🏆 Max Margin Contrastive



See all

Papers With Code

Caveat: sparsely populated

Methodology

Representation Learning

1643 papers with code 2 benchmarks • 2 datasets

Representation learning is concerned with training machine learning algorithms to learn useful representations, e.g. those that are interpretable, have latent features, or can be used for transfer learning.

Deep neural networks can be considered representation learning models that typically encode information which is projected into a different subspace. These representations are then usually passed on to a linear classifier to, for instance, train a classifier.

Edit



Add a Result

Benchmarks

These leaderboards are used to track progress in Representation Learning

Trend Dataset Best Model Paper Code Compare



Circle Data

🏆 Morphological Network



See all

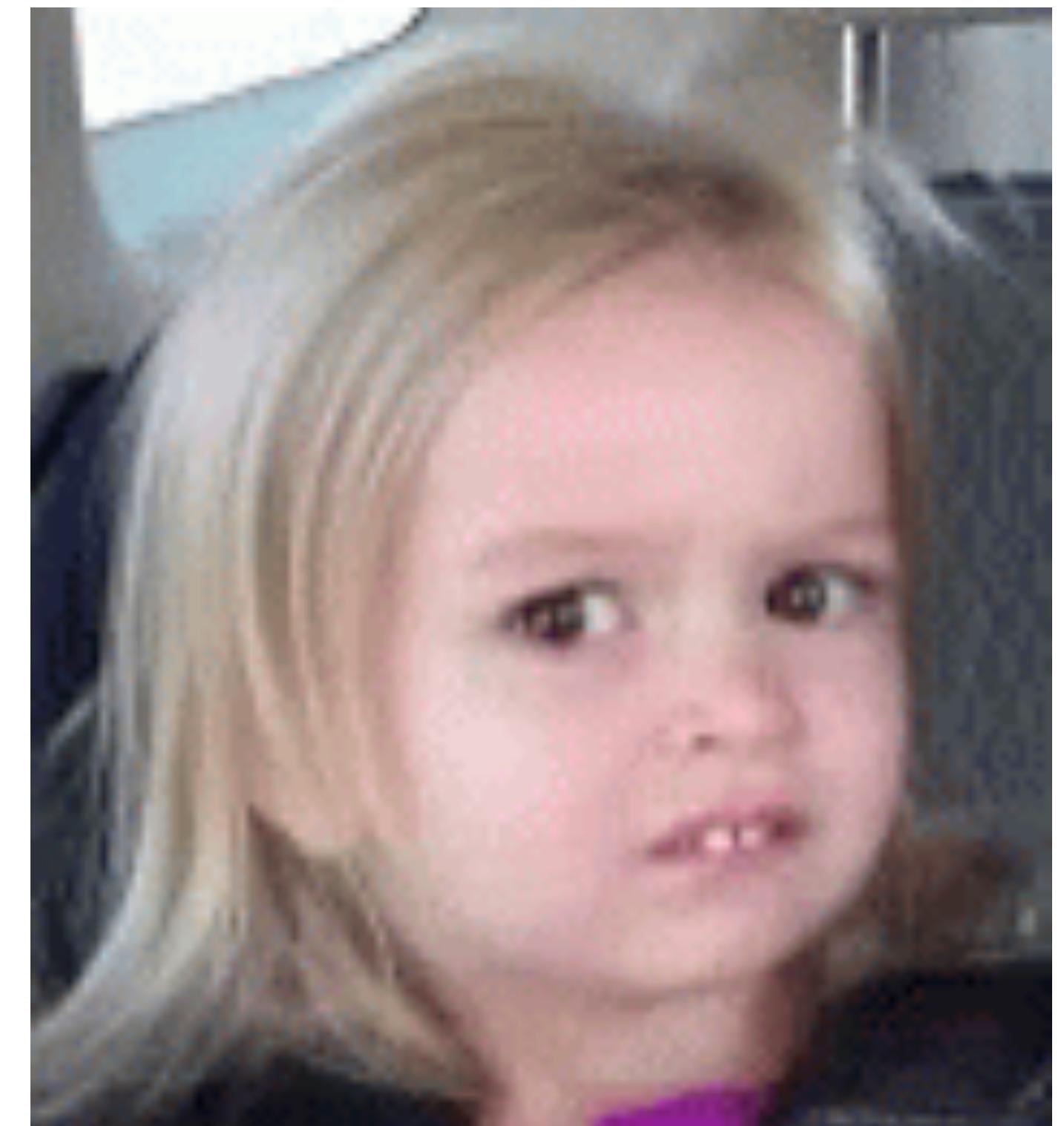


Sports10

🏆 Max Margin Contrastive



See all



**If both Kaggle & PapersWithCode fail
Then go to github + arXiv search**

If both Kaggle & PapersWithCode fail
Then go to github + arXiv search

- You might want to apply heuristics to decide which papers you should read first

If both Kaggle & PapersWithCode fail

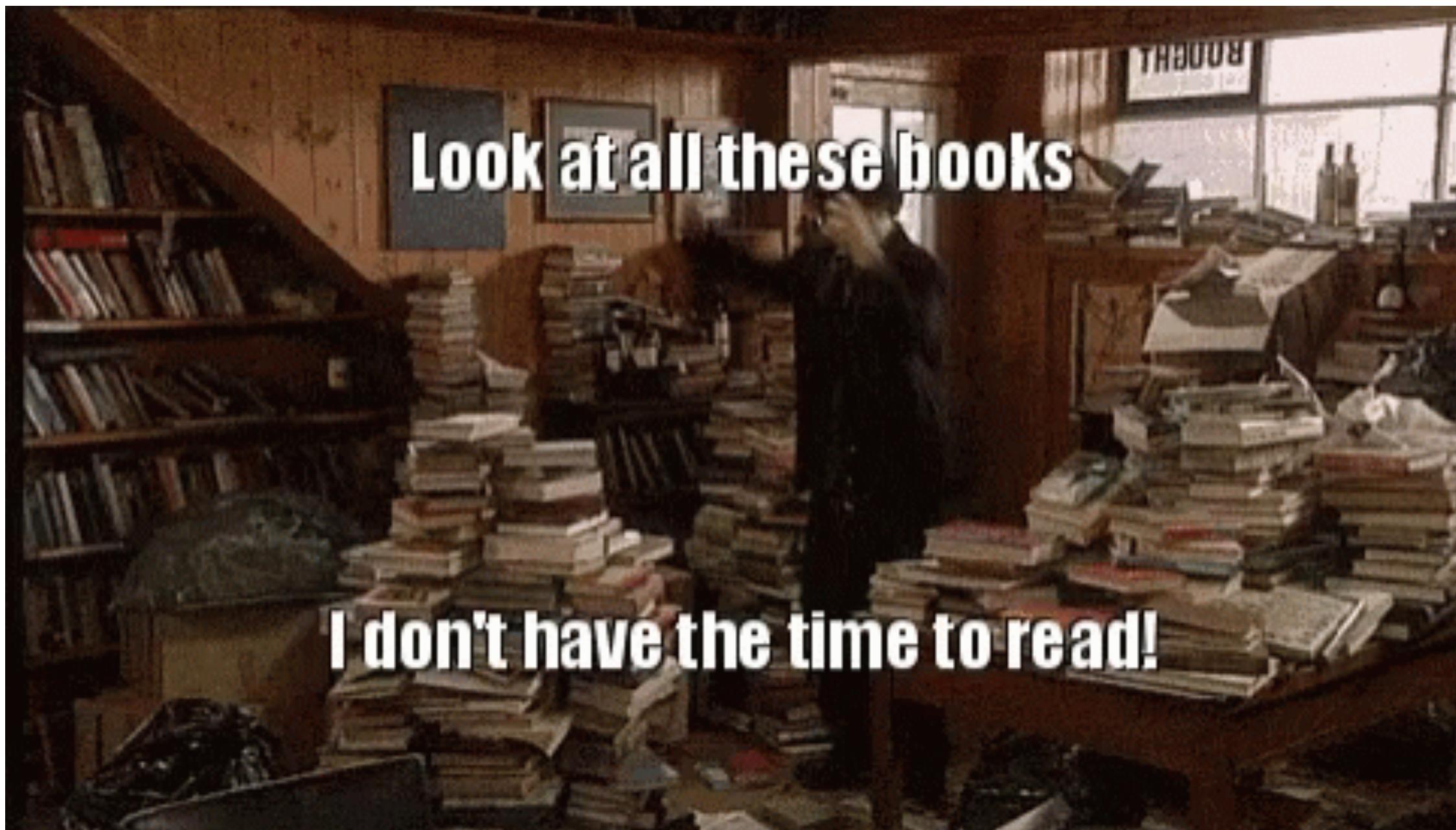
Then go to github + arXiv search

- You might want to apply heuristics to decide which papers you should read first
- But that is later, now let's assume you have some time...

If both Kaggle & PapersWithCode fail

Then go to github + arXiv search

- You might want to apply heuristics to decide which papers you should read first
- But that is later, now let's assume you have some time...



**You are new to the field
and would like to get familiar with it**

**You are new to the field
and would like to get familiar with it**

You are new to the field and would like to get familiar with it

- The best idea is to start with some modern course (e.g. fast.ai)

You are new to the field and would like to get familiar with it

- The best idea is to start with some modern course (e.g. fast.ai)
- Starting with dataset to play with (see Kaggle, PapersWithCode) is also a good idea.

You are new to the field and would like to get familiar with it

- The best idea is to start with some modern course (e.g. fast.ai)
- Starting with dataset to play with (see Kaggle, PapersWithCode) is also a good idea.
- Also a good idea is to start with a good benchmark paper

You are new to the field and would like to get familiar with it

- The best idea is to start with some modern course (e.g. fast.ai)
- Starting with dataset to play with (see Kaggle, PapersWithCode) is also a good idea.
- Also a good idea is to start with a good benchmark paper
- Or a good survey paper

You are new to the field and would like to get familiar with it

- The best idea is to start with some modern course (e.g. fast.ai)
- Starting with dataset to play with (see Kaggle, PapersWithCode) is also a good idea.
- Also a good idea is to start with a good benchmark paper
- Or a good survey paper
- But it is not always easy to find good survey paper, there many bad ones

How to spot good benchmark paper?

Citations

- It is used by community
(unless it is just published 2 days ago), i.e. people cite it

How to spot good benchmark paper?

Citations

- It is used by community (unless it is just published 2 days ago), i.e. people cite it

<https://arxiv.org> > cs ▾ [Přeložit tuto stránku](#)

[Revisiting Oxford and Paris: Large-Scale Image Retrieval ...](#)

autor: F Radenović · 2018 · Počet citací tohoto článku: 203 — In this paper we address issues with image retrieval benchmarking on standard and popular **Oxford 5k** and **Paris 6k** datasets...

How to spot good benchmark paper?

Citations

- It is used by community (unless it is just published 2 days ago), i.e. people cite it

<https://arxiv.org> › cs ▾ Přeložit tuto stránku

[Revisiting Oxford and Paris: Large-Scale Image Retrieval ...](#)

autor: F Radenović · 2018 · Počet citací tohoto článku: 203 — In this paper we address issues with image retrieval benchmarking on standard and popular **Oxford 5k** and **Paris 6k** datasets...

<https://research.google> › pub49052 ▾ Přeložit tuto stránku

[Google Landmarks Dataset v2 - A Large ... - Google Research](#)

autor: T Weyand · 2020 · Počet citací tohoto článku: 72 — We introduce the **Google Landmarks Dataset v2** (GLDv2), a new benchmark for large-scale, fine-grained instance recognition and...

How to spot good benchmark paper?

Citations

- It is used by community (unless it is just published 2 days ago), i.e. people cite it

<https://arxiv.org> › cs ▾ [Přeložit tuto stránku](#)

[Revisiting Oxford and Paris: Large-Scale Image Retrieval ...](#)

autor: F Radenović · 2018 · Počet citací tohoto článku: 203 — In this paper we address issues with image retrieval benchmarking on standard and popular **Oxford 5k** and **Paris 6k** datasets...

<https://research.google> › pub49052 ▾ [Přeložit tuto stránku](#)

[Google Landmarks Dataset v2 - A Large ... - Google Research](#)

autor: T Weyand · 2020 · Počet citací tohoto článku: 72 — We introduce the **Google Landmarks Dataset v2** (GLDv2), a new benchmark for large-scale, fine-grained instance recognition and...

<https://arxiv.org> › cs ▾ [Přeložit tuto stránku](#)

[Image Matching across Wide Baselines: From Paper to Practice](#)

autor: Y Jin · 2020 · Počet citací tohoto článku: 78 — Abstract: We introduce a comprehensive benchmark for local features and robust estimation algorithms, focusing on the downstream ta...

Cite as: [arXiv:2003.01587](#)

How to spot good benchmark paper?

Citations

- It is used by community (unless it is just published 2 days ago), i.e. people cite it

<https://arxiv.org> › cs ▾ [Přeložit tuto stránku](#)

[Revisiting Oxford and Paris: Large-Scale Image Retrieval ...](#)

autor: F Radenović · 2018 · Počet citací tohoto článku: 203 — In this paper we address issues with image retrieval benchmarking on standard and popular **Oxford 5k** and **Paris 6k** datasets...

<https://research.google> › pub49052 ▾ [Přeložit tuto stránku](#)

[Google Landmarks Dataset v2 - A Large ... - Google Research](#)

autor: T Weyand · 2020 · Počet citací tohoto článku: 72 — We introduce the **Google Landmarks Dataset v2** (GLDv2), a new benchmark for large-scale, fine-grained instance recognition and...

<https://arxiv.org> › cs ▾ [Přeložit tuto stránku](#)

[Image Matching across Wide Baselines: From Paper to Practice](#)

autor: Y Jin · 2020 · Počet citací tohoto článku: 78 — Abstract: We introduce a comprehensive benchmark for local features and robust estimation algorithms, focusing on the downstream ta...

Cite as: [arXiv:2003.01587](#)

<https://www.robots.ox.ac.uk> › vibes_ijcv2004 ▾ [PDF](#)

[A Comparison of Affine Region Detectors - University of Oxford](#)

autor: K MIKOLAJCZYK · Počet citací tohoto článku: 4188 — Abstract. The paper gives a snapshot of the state of the art in **affine covariant region detectors**, and compares their...

How to spot good benchmark paper?

Citations

- It is used by community (unless it is just published 2 days ago), i.e. people cite it

<https://arxiv.org> › cs ▾ Přeložit tuto stránku

Revisiting Oxford and Paris: Large-Scale Image Retrieval ...

autor: F Radenović · 2018 · Počet citací tohoto článku: 203 — In this paper we address issues with image retrieval benchmarking on standard and popular Oxford 5k and Paris 6k datasets...

<https://research.google> › pub49052 ▾ Přeložit tuto stránku

Google Landmarks Dataset v2 - A Large ... - Google Research

autor: T Weyand · 2020 · Počet citací tohoto článku: 72 — We introduce the **Google Landmarks Dataset v2** (GLDv2), a new benchmark for large-scale, fine-grained instance recognition and...

<https://arxiv.org> › cs ▾ Přeložit tuto stránku

Image Matching across Wide Baselines: From Paper to Practice

autor: Y Jin · 2020 · Počet citací tohoto článku: 78 — Abstract: We introduce a comprehensive benchmark for local features and robust estimation algorithms, focusing on the downstream ta...

Cite as: [arXiv:2003.01587](https://arxiv.org/abs/2003.01587)

<https://www.robots.ox.ac.uk> › vibes_ijcv2004 ▾ PDF

A Comparison of Affine Region Detectors - University of Oxford

autor: K MIKOLAJCZYK · Počet citací tohoto článku: 4188 — Abstract. The paper gives a snapshot of the state of the art in **affine covariant region detectors**, and compares their...

How to spot good benchmark paper?

Citations

- It is used by community (unless it is just published 2 days ago), i.e. people cite it

<https://arxiv.org> › cs ▾ Přeložit tuto stránku

Revisiting Oxford and Paris: Large-Scale Image Retrieval ...

autor: F Radenović · 2018 · Počet citací tohoto článku: 203 — In this paper we address issues with image retrieval benchmarking on standard and popular Oxford 5k and Paris 6k datasets...

<https://research.google> › pub49052 ▾ Přeložit tuto stránku

Google Landmarks Dataset v2 - A Large ... - Google Research

autor: T Weyand · 2020 · Počet citací tohoto článku: 72 — We introduce the Google Landmarks Dataset v2 (GLDv2), a new benchmark for large-scale, fine-grained instance recognition and...

<https://arxiv.org> › cs ▾ Přeložit tuto stránku

Image Matching across Wide Baselines: From Paper to Practice

autor: Y Jin · 2020 · Počet citací tohoto článku: 78 — Abstract: We introduce a comprehensive benchmark for local features and robust estimation algorithms, focusing on the downstream ta...

Cite as: [arXiv:2003.01587](https://arxiv.org/abs/2003.01587)

<https://www.robots.ox.ac.uk> › vibes_ijcv2004 ▾ PDF

A Comparison of Affine Region Detectors - University of Oxford

autor: K MIKOLAJCZYK · Počet citací tohoto článku: 4188 — Abstract. The paper gives a snapshot of the state of the art in **affine covariant region detectors**, and compares their...

How to spot good benchmark paper?

Citations

- It is used by community (unless it is just published 2 days ago), i.e. people cite it

<https://arxiv.org> › cs ▾ Přeložit tuto stránku

Revisiting Oxford and Paris: Large-Scale Image Retrieval ...

autor: F Radenović · 2018 · Počet citací tohoto článku: 203 — In this paper we address issues with image retrieval benchmarking on standard and popular Oxford 5k and Paris 6k datasets...

<https://research.google> › pub49052 ▾ Přeložit tuto stránku

Google Landmarks Dataset v2 - A Large ... - Google Research

autor: T Weyand · 2020 · Počet citací tohoto článku: 72 — We introduce the Google Landmarks Dataset v2 (GLDv2), a new benchmark for large-scale, fine-grained instance recognition and...

<https://arxiv.org> › cs ▾ Přeložit tuto stránku

Image Matching across Wide Baselines: From Paper to Practice

autor: Y Jin · 2020 · Počet citací tohoto článku: 78 — Abstract: We introduce a comprehensive benchmark for local features and robust estimation algorithms, focusing on the downstream ta...

Cite as: [arXiv:2003.01587](https://arxiv.org/abs/2003.01587)

<https://www.robots.ox.ac.uk> › vibes_ijcv2004 ▾ PDF

A Comparison of Affine Region Detectors - University of Oxford

autor: K MIKOLAJCZYK · Počet citací tohoto článku: 4188 — Abstract. The paper gives a snapshot of the state of the art in **affine covariant region detectors**, and compares their...

How to spot good benchmark paper?

Citations

- It is used by community (unless it is just published 2 days ago), i.e. people cite it

<https://arxiv.org> › cs ▾ Přeložit tuto stránku

Revisiting Oxford and Paris: Large-Scale Image Retrieval ...

autor: F Radenović · 2018 · Počet citací tohoto článku: 203 — In this paper we address issues with image retrieval benchmarking on standard and popular Oxford 5k and Paris 6k datasets...

<https://research.google> › pub49052 ▾ Přeložit tuto stránku

Google Landmarks Dataset v2 - A Large ... - Google Research

autor: T Weyand · 2020 · Počet citací tohoto článku: 72 — We introduce the Google Landmarks Dataset v2 (GLDv2), a new benchmark for large-scale, fine-grained instance recognition and...

<https://arxiv.org> › cs ▾ Přeložit tuto stránku

Image Matching across Wide Baselines: From Paper to Practice

autor: Y Jin · 2020 · Počet citací tohoto článku: 78 — Abstract: We introduce a comprehensive benchmark for local features and robust estimation algorithms, focusing on the downstream ta...

Cite as: [arXiv:2003.01587](https://arxiv.org/abs/2003.01587)

<https://www.robots.ox.ac.uk> › vibes_ijcv2004 ▾ PDF

A Comparison of Affine Region Detectors - University of Oxford

autor: K MIKOLAJCZYK · Počet citací tohoto článku: 4188 — Abstract. The paper gives a snapshot of the state of the art in **affine covariant region detectors**, and compares their...

How to spot good benchmark paper?

Year

How to spot good benchmark paper?

Year

- It should not be too old:
benchmark from 2004 is a little
bit obsolete
- Oxford-Affine is GREAT paper.
But it would not help you find
modern local features

How to spot good benchmark paper?

Year

- It should not be too old: benchmark from 2004 is a little bit obsolete
- Oxford-Affine is GREAT paper. But it would not help you find modern local features

https://www.robots.ox.ac.uk/vibes_ijcv2004 ▾ PDF

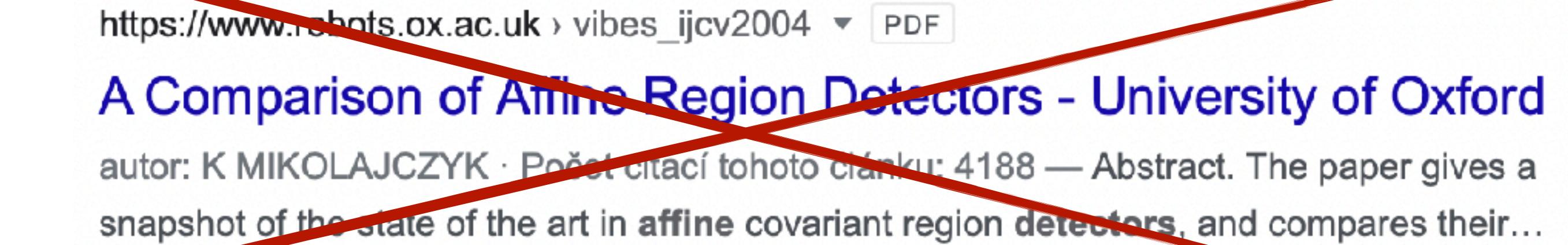
[A Comparison of Affine Region Detectors - University of Oxford](#)

autor: K MIKOLAJCZYK · Počet citací tohoto článku: 4188 — Abstract. The paper gives a snapshot of the state of the art in **affine covariant region detectors**, and compares their...

How to spot good benchmark paper?

Year

- It should not be too old: benchmark from 2004 is a little bit obsolete
- Oxford-Affine is GREAT paper. But it would not help you find modern local features



How to spot good benchmark paper?

It proposes new dataset (or new annotation)

How to spot good benchmark paper?

It proposes new dataset (or new annotation)

- There are TONS of papers, which just run methods from OpenCV on existing (and old) datasets

How to spot good benchmark paper?

It proposes new dataset (or new annotation)

- There are TONS of papers, which just run methods from OpenCV on existing (and old) datasets



International Journal of
Applied Mathematics,
Electronics and Computers
ISSN: 2147-8228
www.atscience.org/IJAMEC



A Comparative Evaluation of Well-known Feature Detectors and Descriptors[#]

How to spot good benchmark paper?

It proposes new dataset (or new annotation)

- There are TONS of papers, which just run methods from OpenCV on existing (and old) datasets



International Journal of
Applied Mathematics,
Electronics and Computers
ISSN: 2147-8228
www.atscience.org/IJAMEC



A Comparative Evaluation of Well-known Feature Detectors and Descriptors[#]

3.7. Effects of JPEG Compression

In this experiment, impacts of compression is examined in terms of comparison of results obtained from each method conducting on ubc dataset. For this purpose, the artifacts have introduced in Joint Photographic Experts Group (JPEG) compression by using a standard xv image browser with the image quality parameter changing from 40% to 2%.

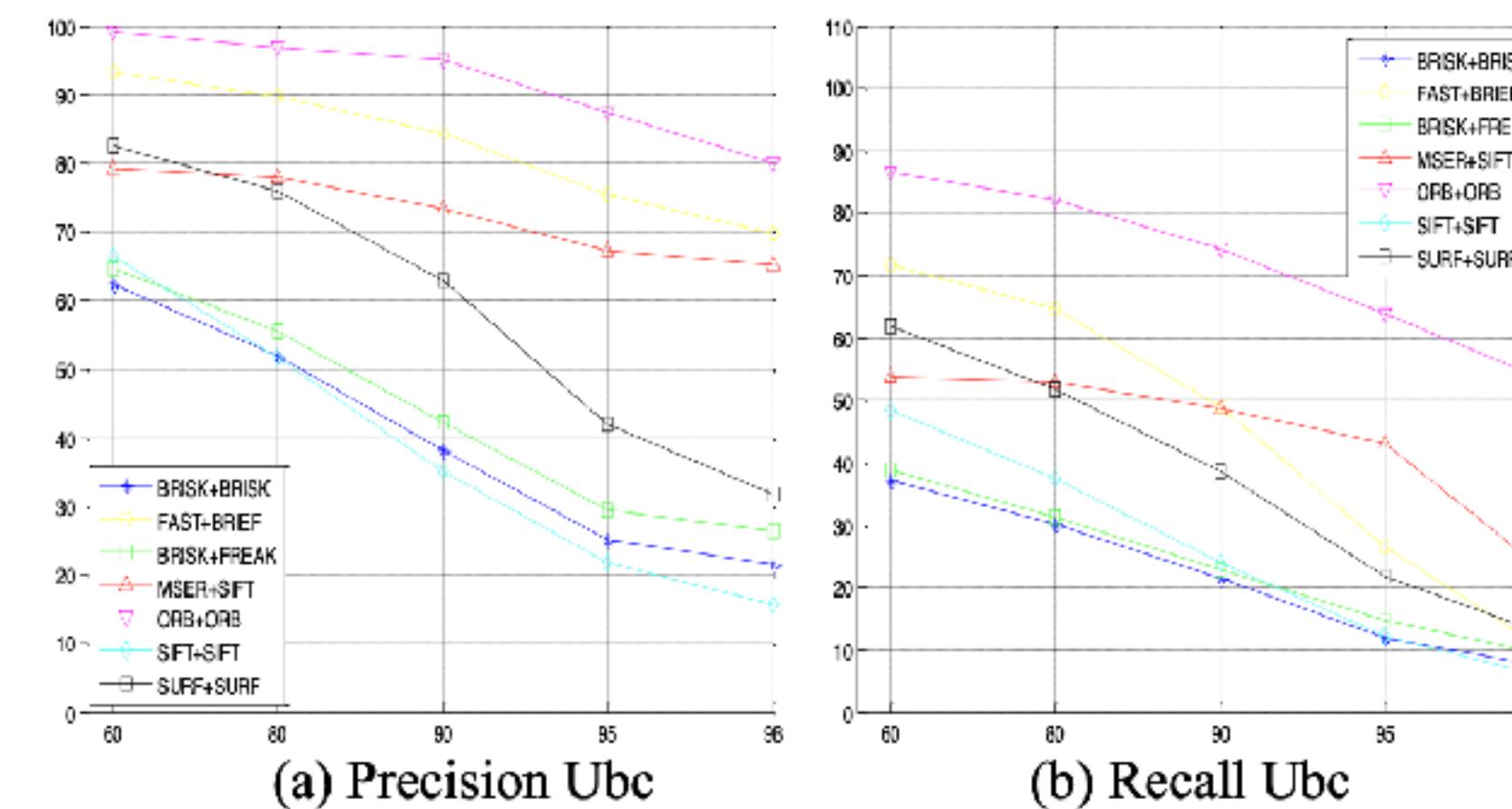
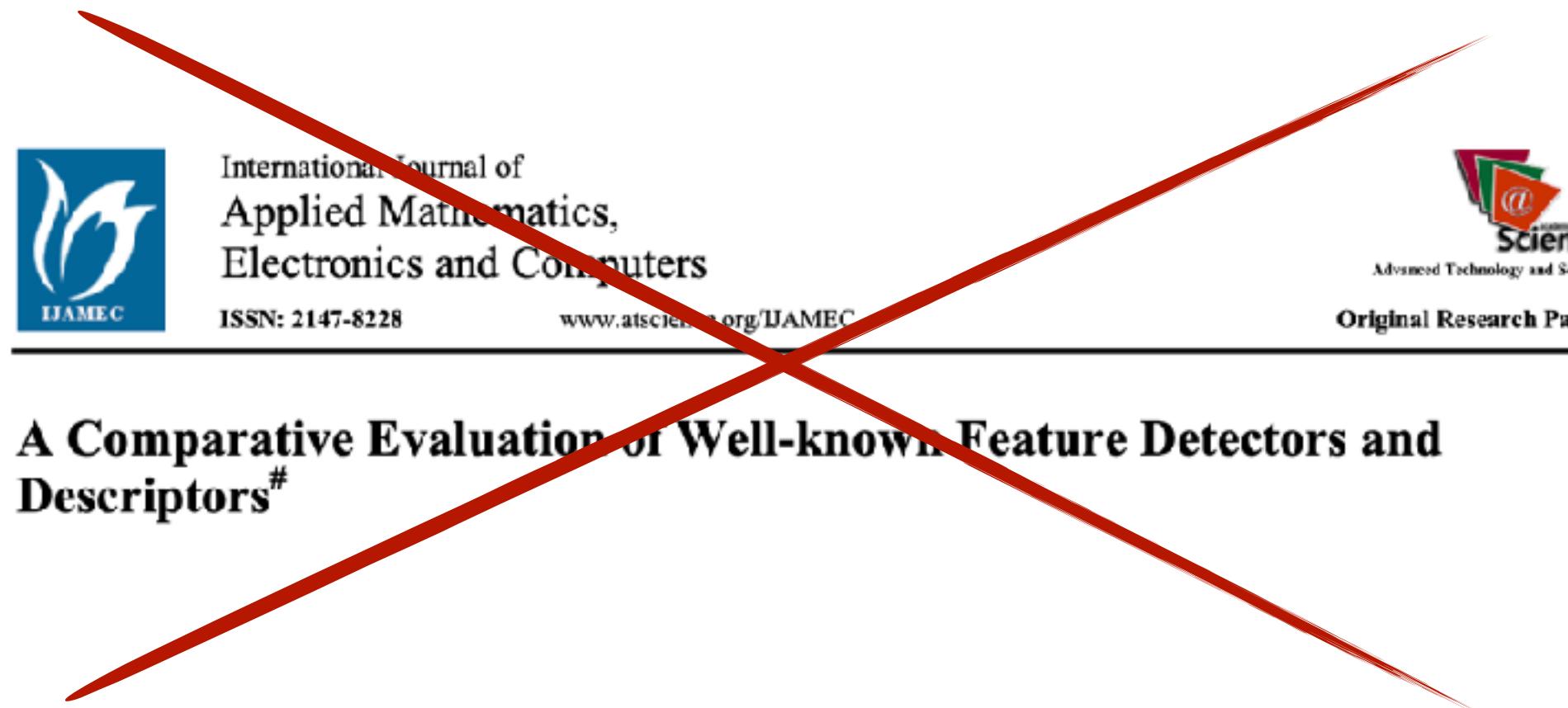


Figure 7: Precision and recall values for JPEG compression.

How to spot good benchmark paper?

It proposes new dataset (or new annotation)

- There are TONS of papers, which just run methods from OpenCV on existing (and old) datasets



3.7. Effects of JPEG Compression

In this experiment, impacts of compression is examined in terms of comparison of results obtained from each method conducting on ubc dataset. For this purpose, the artifacts have introduced in Joint Photographic Experts Group (JPEG) compression by using a standard xv image browser with the image quality parameter changing from 40% to 2%.

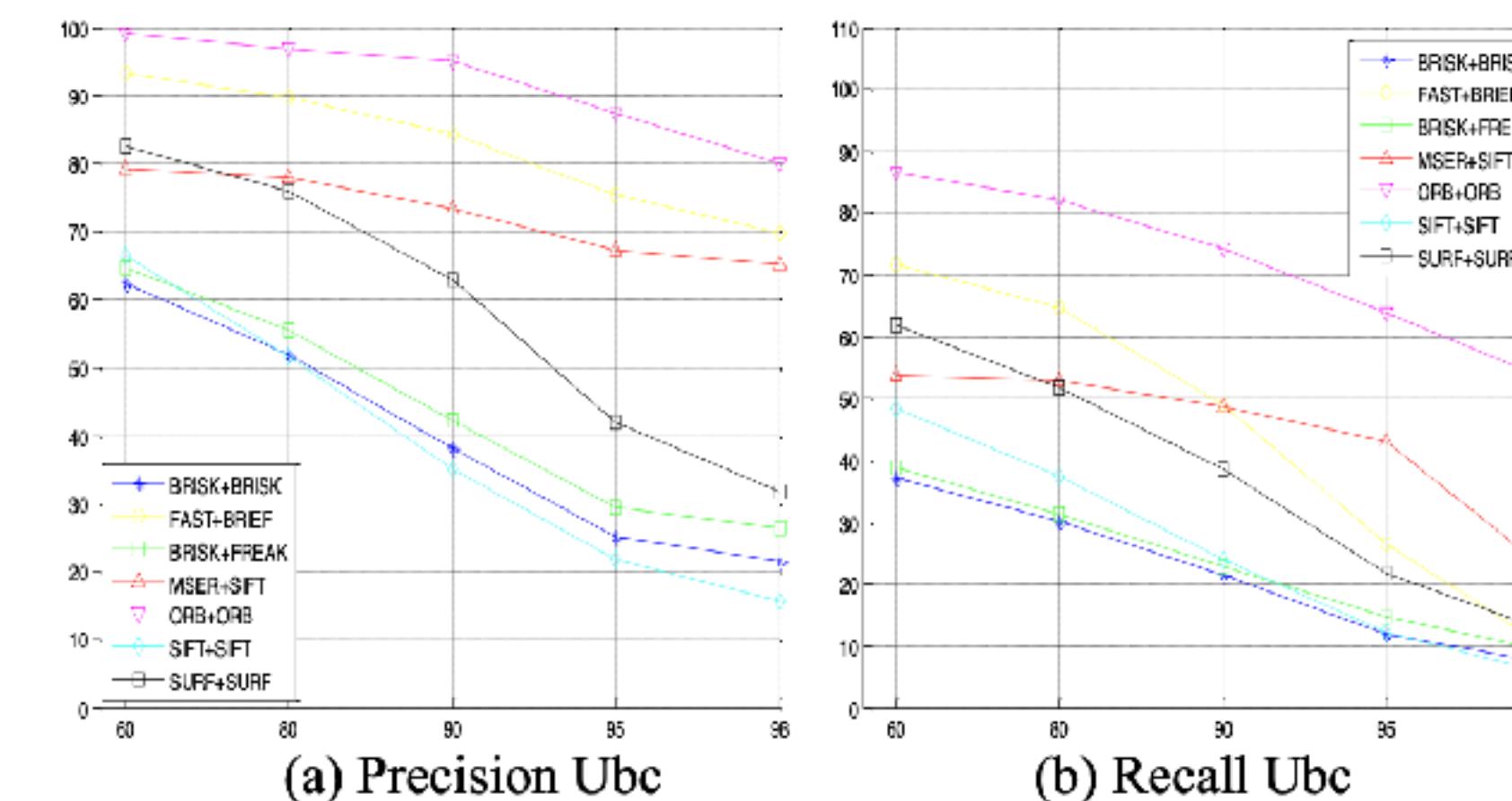
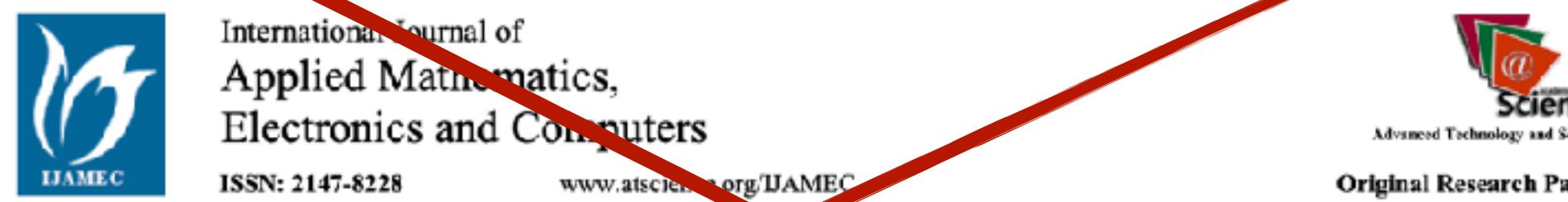


Figure 7: Precision and recall values for JPEG compression.

How to spot good benchmark paper?

It proposes new dataset (or new annotation)

- There are TONS of papers, which just run methods from OpenCV on existing (and old) datasets



A Comparative Evaluation of Well-known Feature Detectors and Descriptors[#]

3.7. Effects of JPEG Compression

In this experiment, impacts of compression is examined in terms of comparison of results obtained from each method conducting on ubc dataset. For this purpose, the artifacts have introduced in Joint Photographic Experts Group (JPEG) compression by using a standard xv image browser with the image quality parameter changing from 40% to 2%.

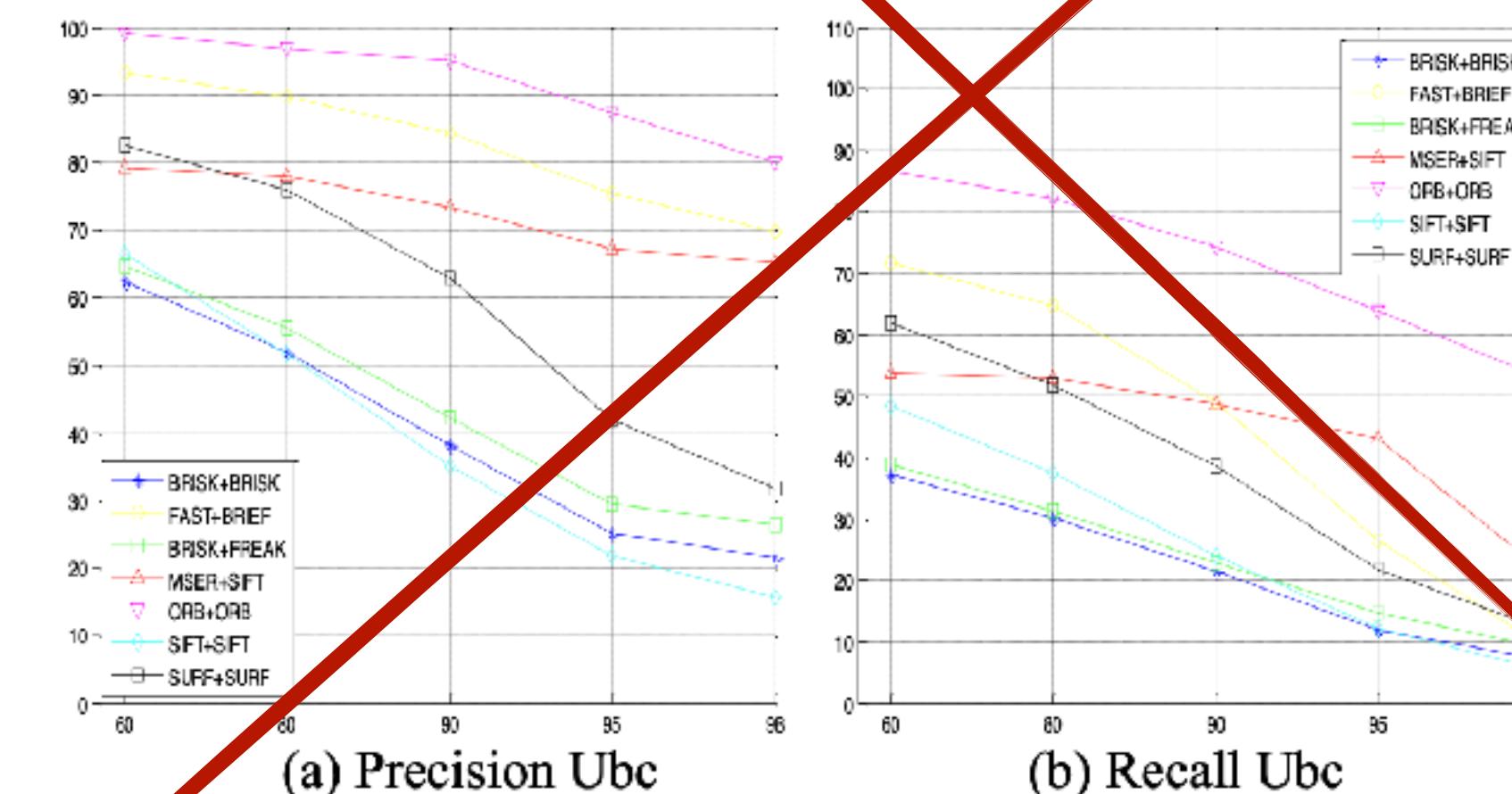


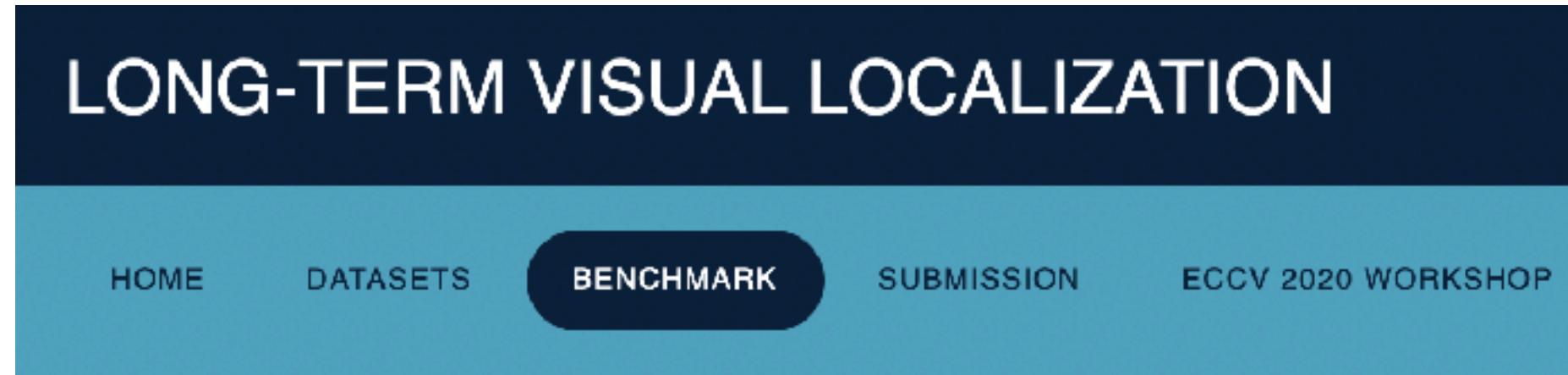
Figure 7: Precision and recall values for JPEG compression.

How to spot good benchmark paper?

It has a challenge linked to it, often – many years

How to spot good benchmark paper?

It has a challenge linked to it, often – many years



Benchmark

Listed below are the public results on the three benchmark datasets. The localization results are reported as the perc
translation and rotation thresholds, for each condition.

Aachen Day-Night dataset

Localization thresholds:

All conditions: (0.25m, 2°) / (0.5m, 5°) / (5m, 10°)

Method	day
Combined tuned HLoc	89.7 / 95.8 / 98
Hierarchical Localization - SuperPoint + SuperGlue	89.6 / 95.4 / 98
SuperGlue + Patch2Pix (HLoc)	89.2 / 95.5 / 98
ONavi-V2.0	88.6 / 95.1 / 99
hloc	89.3 / 95.9 / 98
KAPTURE-R2D2-FUSION	89.4 / 96.4 / 99

<https://www.visuallocalization.net/benchmark/>

<https://www.votchallenge.net>

How to spot good benchmark paper?

It has a challenge linked to it, often – many years

LONG-TERM VISUAL LOCALIZATION

HOME DATASETS **BENCHMARK** SUBMISSION ECCV 2020 WORKSHOP

Benchmark

Listed below are the public results on the three benchmark datasets. The localization results are reported as the perc translation and rotation thresholds, for each condition.

Aachen Day-Night dataset
Localization thresholds:
All conditions: (0.25m, 2°) / (0.5m, 5°) / (5m, 10°)

Method	day
<u>Combined tuned HLoc</u>	89.7 / 95.8 / 98
<u>Hierarchical Localization - SuperPoint + SuperGlue</u>	89.6 / 95.4 / 98
<u>SuperGlue + Patch2Pix (HLoc)</u>	89.2 / 95.5 / 98
<u>ONavi-V2.0</u>	88.6 / 95.1 / 99
<u>hloc</u>	89.3 / 95.9 / 98
<u>KAPTURE-R2D2-FUSION</u>	89.4 / 96.4 / 99

Benchmarks

Official results of past challenges are still available as a benchmark to researchers. We invite the visual tracking researchers to compare their methods with those results.



VOT2021 benchmark

The VOT2021 benchmark addresses short-term, long-term, real-time, RGB and RGBD trackers. Results were presented at the ICCV2021 VOT workshop.



VOT2020 benchmark

The VOT2020 benchmark addresses short-term, long-term, real-time, RGB, RGBT and RGBD trackers. Results were presented at the ECCV2020 VOT workshop.



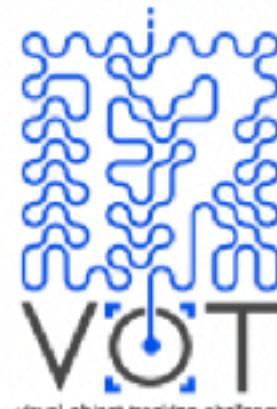
VOT2019 benchmark

The VOT2019 benchmark addresses short-term, long-term, real-time, RGB, RGBT and RGBD trackers. Results were presented at the ICCV2019 VOT workshop.



VOT2018 benchmark

The VOT2018 benchmark introduced a long-term subchallenge VOT-LT2018. Results were presented at the VOT workshop at ECCV2018.



VOT2017 benchmark

The VOT2017 benchmark introduced a refreshed a dataset and a real-time experiment. The winner will was determined on sequestered dataset. The results were presented at the VOT workshop at ICCV2017.



VOT2016 benchmark

The fourth challenge updated the dataset of 60 sequences with new annotations. The results were published in a joint paper presented at a workshop at ECCV2016.

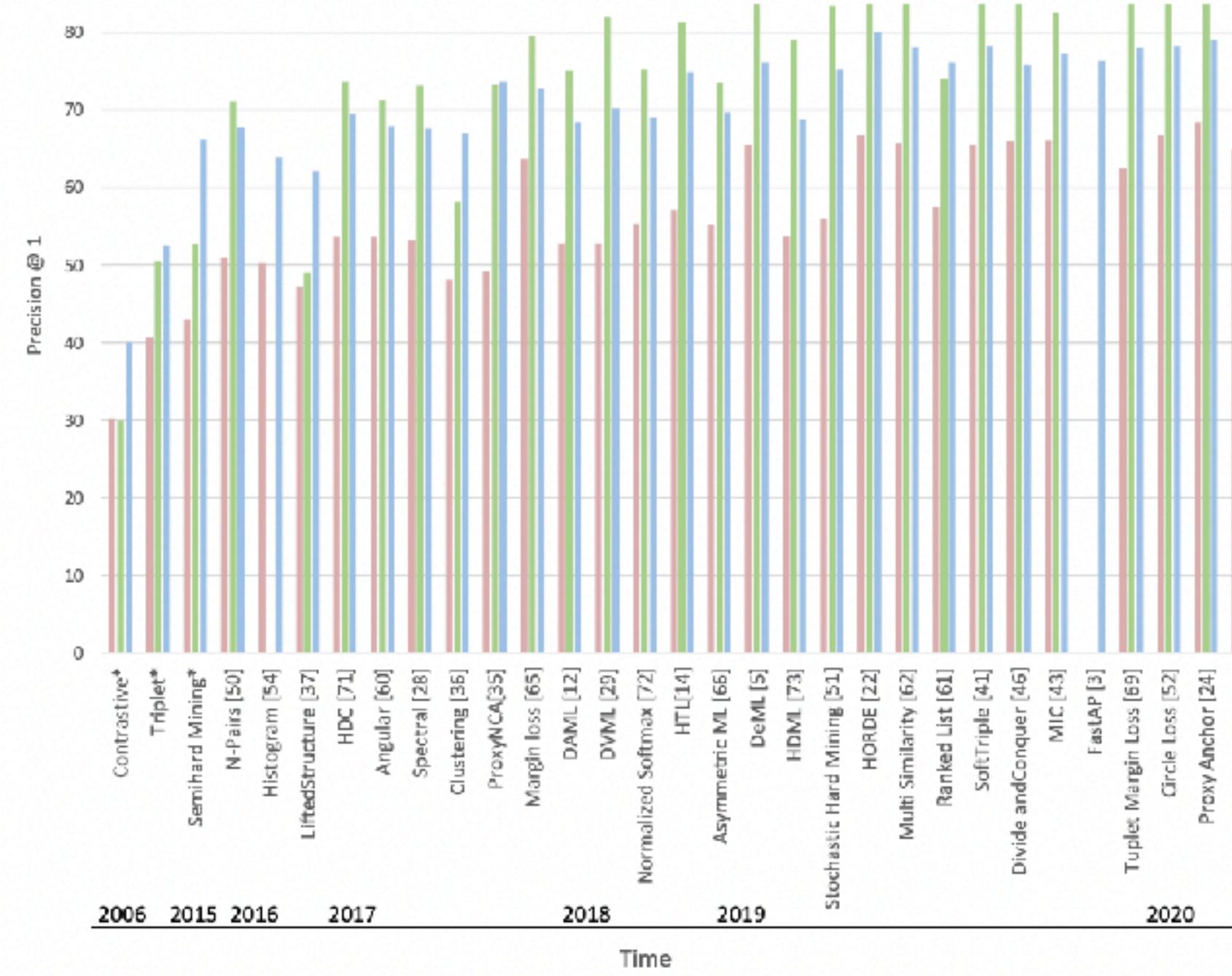
How to spot good benchmark paper?

It re-evaluates existing method on the common ground

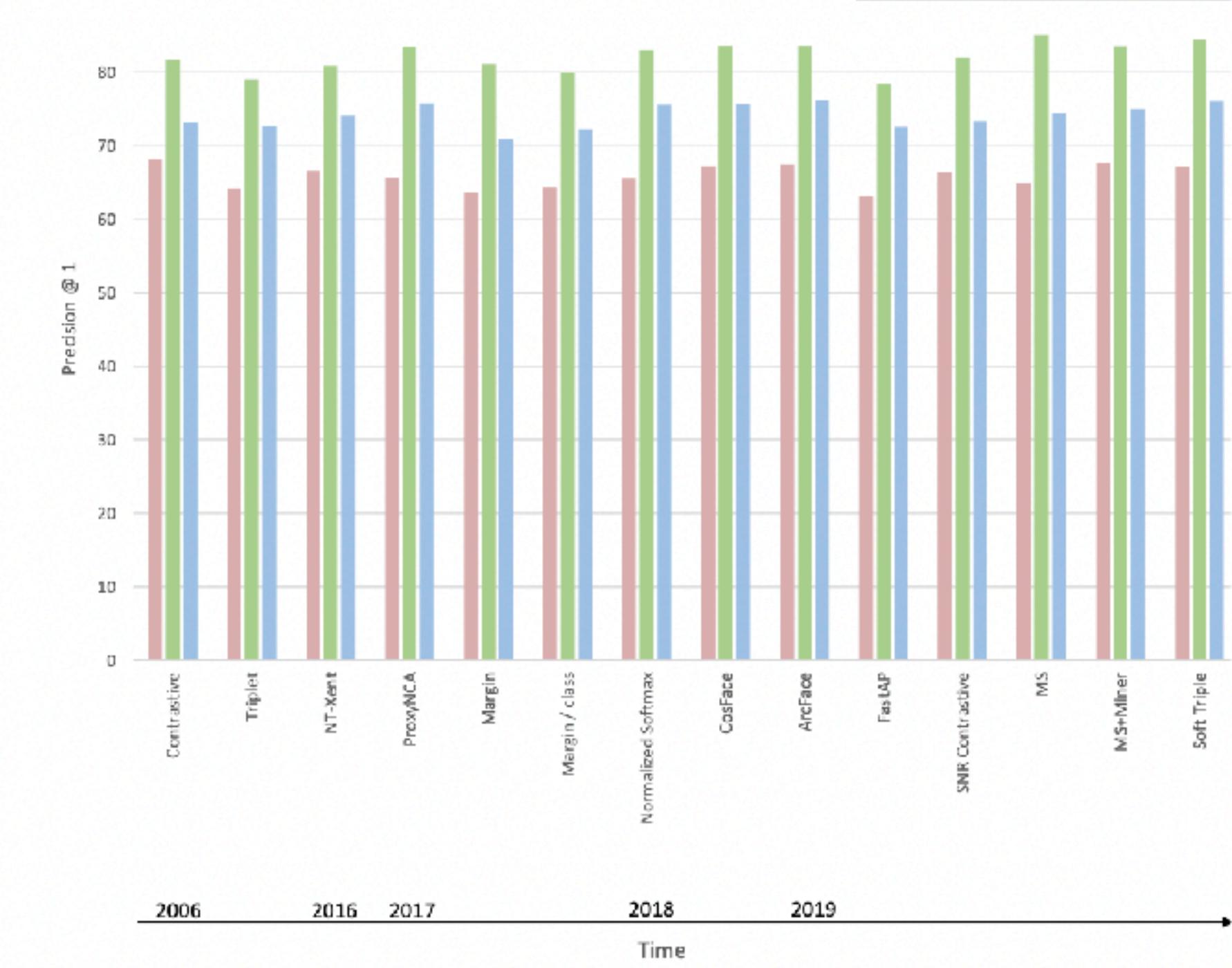
<https://arxiv.org> › cs ▾ Přeložit tuto stránku

[2003.08505] A Metric Learning Reality Check - arXiv

autor: K Musgrave · 2020 · Počet citací tohoto článku: 159 — Abstract: Deep metric learning papers from the past four years have consistently claimed great advances in accuracy, often...



(a) The trend according to papers



(b) The trend according to reality

How to spot good benchmark paper?

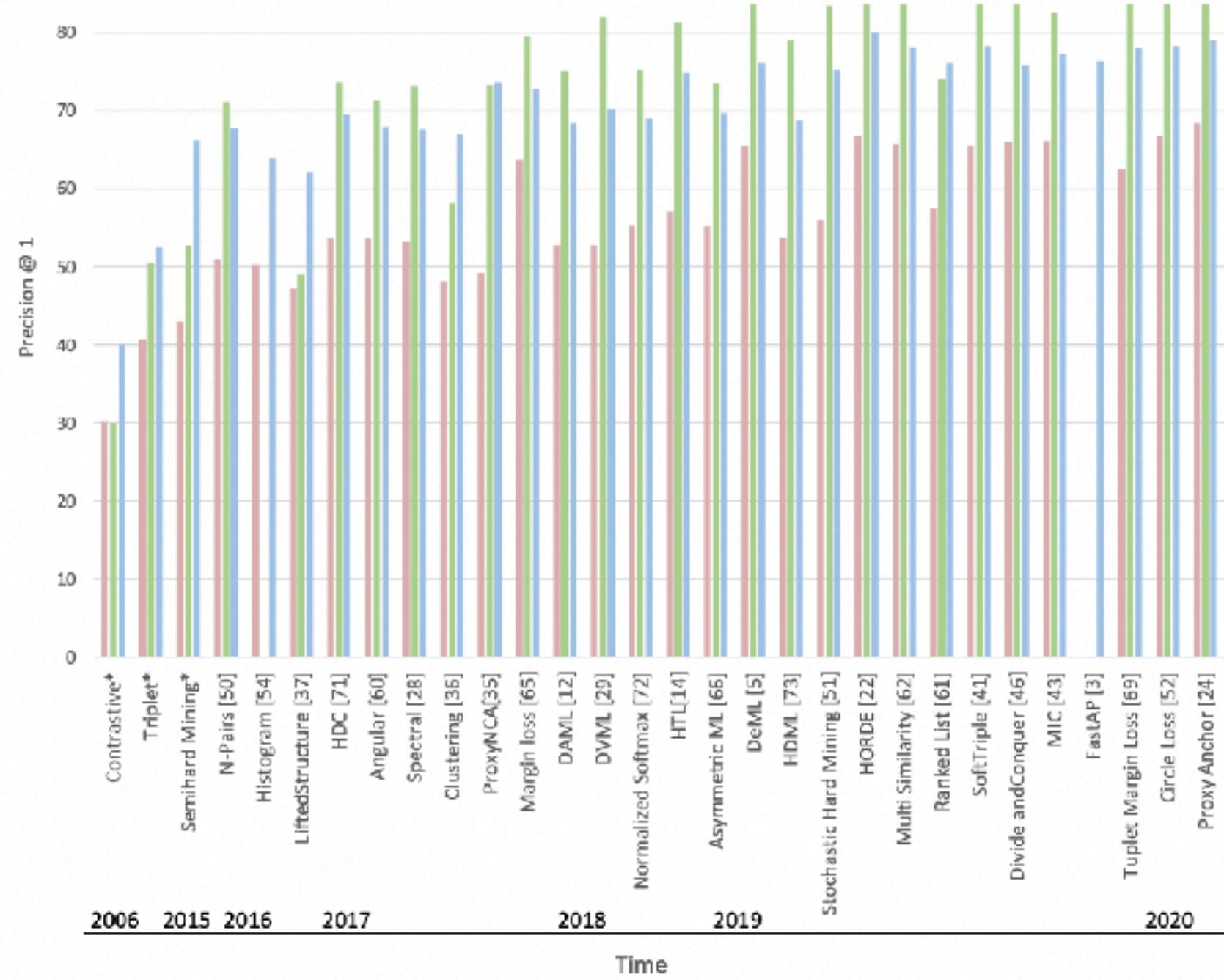
It re-evaluates existing method on the common ground

- Some papers do not propose new dataset, but they critically evaluate existing metrics

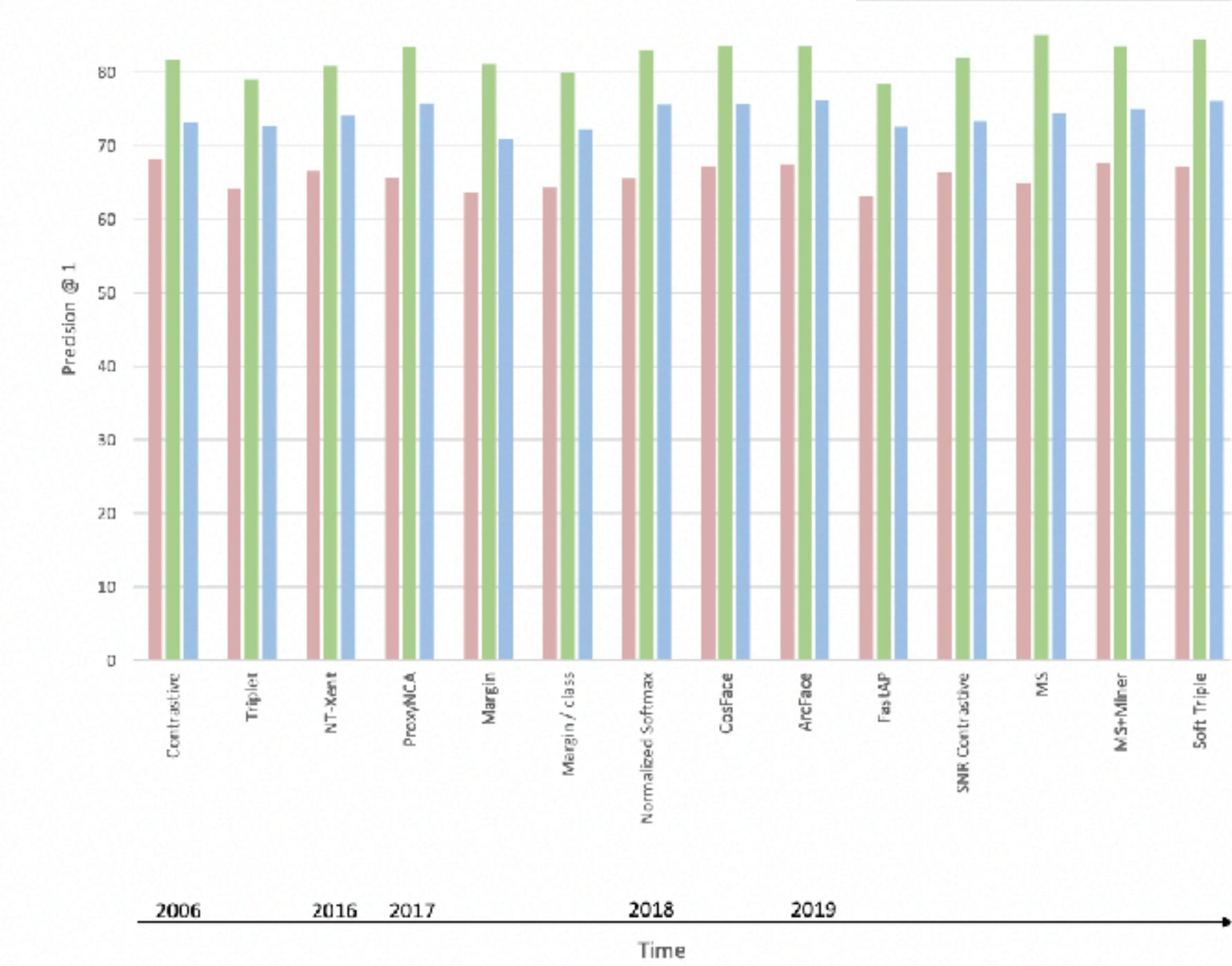
<https://arxiv.org> › cs ▾ Přeložit tuto stránku

[2003.08505] A Metric Learning Reality Check - arXiv

autor: K Musgrave · 2020 · Počet citací tohoto článku: 159 — Abstract: Deep metric learning papers from the past four years have consistently claimed great advances in accuracy, often...



(a) The trend according to papers



(b) The trend according to reality

How to spot good benchmark paper?

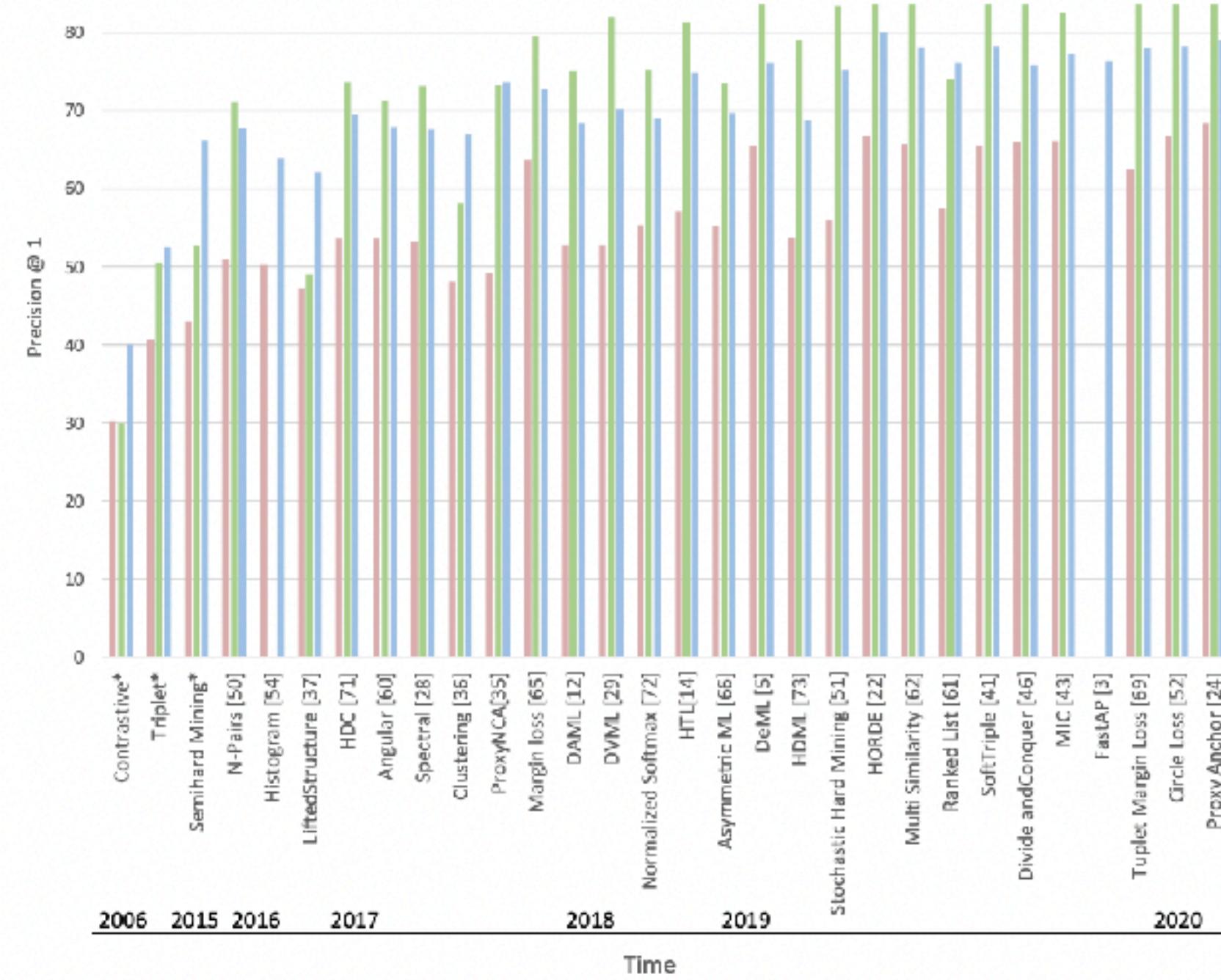
It re-evaluates existing method on the common ground

- Some papers do not propose new dataset, but they critically evaluate existing metrics
- And/or re-implement the methods from scratch on the common codebase

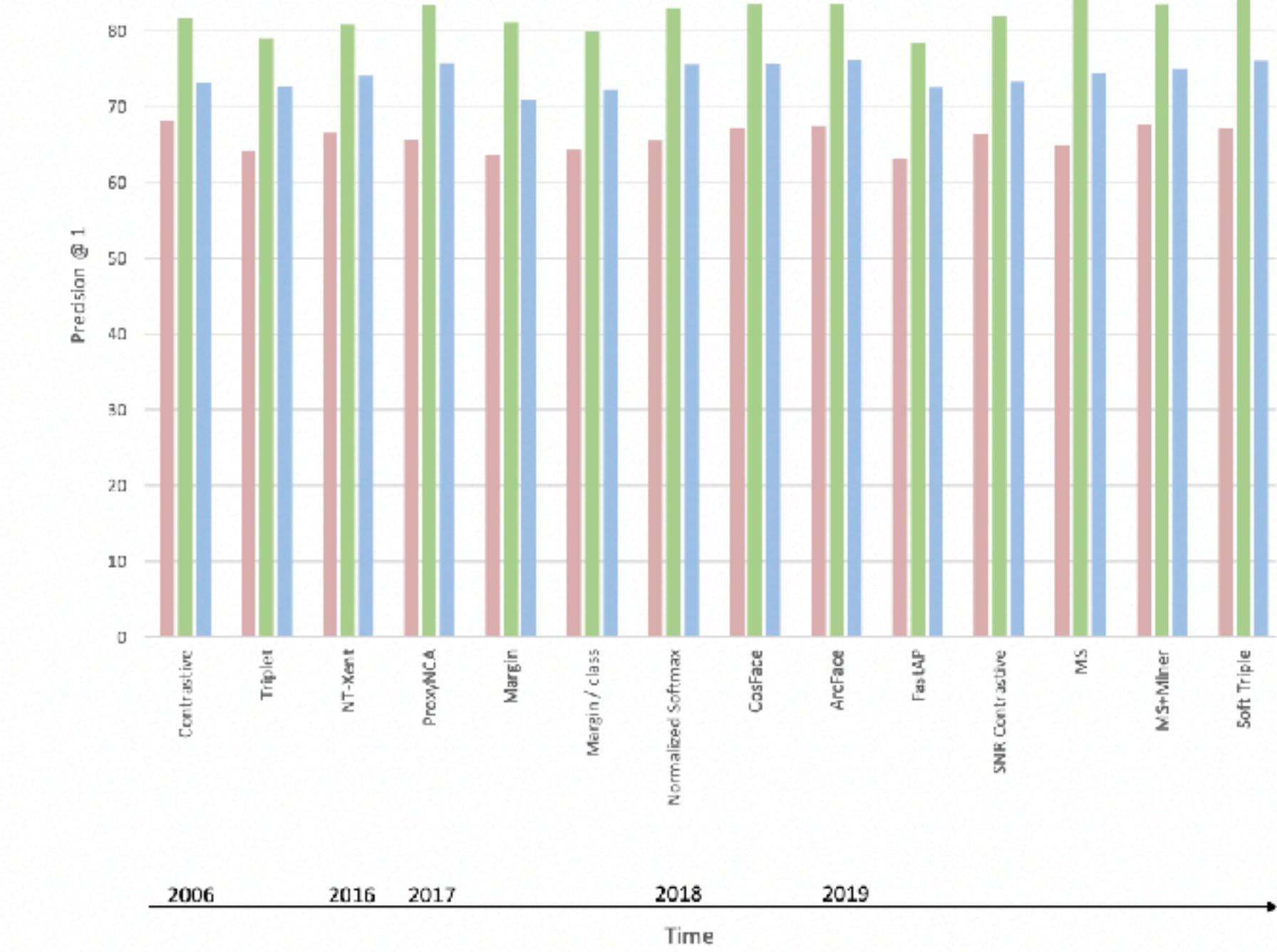
<https://arxiv.org> › cs ▾ Přeložit tuto stránku

[2003.08505] A Metric Learning Reality Check - arXiv

autor: K Musgrave · 2020 · Počet citací tohoto článku: 159 — Abstract: Deep metric learning papers from the past four years have consistently claimed great advances in accuracy, often...



(a) The trend according to papers



(b) The trend according to reality

Good Surveys: harder to distinguish

But there are couple of heuristics

Good Surveys: harder to distinguish

But there are couple of heuristics

- Good modern survey is getting updated

Good Surveys: harder to distinguish

But there are couple of heuristics

[arXiv:1704.05519 \(cs\)](#)

[Submitted on 18 Apr 2017 ([v1](#)), last revised 17 Mar 2021 (this version, v3)]

Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art

[Joel Janai](#), [Fatma Güney](#), [Aseem Behl](#), [Andreas Geiger](#)

Good Surveys: harder to distinguish

But there are couple of heuristics

- Good modern survey is getting updated

[arXiv:1704.05519 \(cs\)](#)

[Submitted on 18 Apr 2017 ([v1](#)), last revised [17 Mar 2021 \(this version, v3\)](#)]

Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art

[Joel Janai](#), [Fatma Güney](#), [Aseem Behl](#), [Andreas Geiger](#)

Good Surveys: harder to distinguish

But there are couple of heuristics

- Good modern survey is getting updated

arXiv:1704.05519 (cs)

[Submitted on 18 Apr 2017 (v1), last revised 17 Mar 2021 (this version, v3)]

Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art

Joel Janai, Fatma Güney, Aseem Behl, Andreas Geiger

[Submitted on 26 Jul 2018 (v1), last revised 14 Jun 2019 (this version, v3)]

From handcrafted to deep local features

Gabriela Csurka, Christopher R. Dance, Martin Humenberger

Good Surveys: harder to distinguish

But there are couple of heuristics

- Good modern survey is getting updated

arXiv:1704.05519 (cs)

[Submitted on 18 Apr 2017 (v1), last revised 17 Mar 2021 (this version, v3)]

Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art

Joel Janai, Fatma Güney, Aseem Behl, Andreas Geiger

[Submitted on 26 Jul 2018 (v1), last revised 14 Jun 2019 (this version, v3)]

From handcrafted to deep local features

Gabriela Csurka, Christopher R. Dance, Martin Humenberger

Good Surveys: harder to distinguish

But there are couple of heuristics

- Good modern survey is getting updated
- available on arXiv/author page

arXiv:1704.05519 (cs)

[Submitted on 18 Apr 2017 (v1), last revised 17 Mar 2021 (this version, v3)]

Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art

Joel Janai, Fatma Güney, Aseem Behl, Andreas Geiger

[Submitted on 26 Jul 2018 (v1), last revised 14 Jun 2019 (this version, v3)]

From handcrafted to deep local features

Gabriela Csurka, Christopher R. Dance, Martin Humenberger

Good Surveys: harder to distinguish

But there are couple of heuristics

- Good modern survey is getting updated
- available on arXiv/author page

arXiv:1704.05519 (cs)

[Submitted on 18 Apr 2017 (v1), last revised 17 Mar 2021 (this version, v3)]

Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art

Joel Janai, Fatma Güney, Aseem Behl, Andreas Geiger

[Submitted on 26 Jul 2018 (v1), last revised 14 Jun 2019 (this version, v3)]

From handcrafted to deep local features

Gabriela Csurka, Christopher R. Dance, Martin Humenberger

[Submitted on 14 Sep 2020 (v1), last revised 16 Sep 2020 (this version, v2)]

Efficient Transformers: A Survey

Yi Tay, Mostafa Dehghani, Dara Bahri, Donald Metzler

Good Surveys: harder to distinguish

But there are couple of heuristics

- Good modern survey is getting updated
- available on arXiv/author page
- bad survey is “published” in obscure journals or behind the paywall

arXiv:1704.05519 (cs)

[Submitted on 18 Apr 2017 (v1), last revised 17 Mar 2021 (this version, v3)]

Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art

Joel Janai, Fatma Güney, Aseem Behl, Andreas Geiger

[Submitted on 26 Jul 2018 (v1), last revised 14 Jun 2019 (this version, v3)]

From handcrafted to deep local features

Gabriela Csurka, Christopher R. Dance, Martin Humenberger

[Submitted on 14 Sep 2020 (v1), last revised 16 Sep 2020 (this version, v2)]

Efficient Transformers: A Survey

Yi Tay, Mostafa Dehghani, Dara Bahri, Donald Metzler

Good Surveys: harder to distinguish

But there are couple of heuristics

- Good modern survey is getting updated
- available on arXiv/author page
- bad survey is “published” in obscure journals or behind the paywall

<http://www.ijetc.com> › 2_NCETEC001_(p.9-17).... ▾ PDF

A SURVEY OF TRENDS IN LOCAL INVARIANT FEATURE ...

autor: JN Surekha · 2017 · Počet citací tohoto článku: 2 — Many computer vision algorithms use feature detection as the initial step, so as a result, a very large number of feature detectors...
9 stránek

arXiv:1704.05519 (cs)

[Submitted on 18 Apr 2017 (v1), last revised 17 Mar 2021 (this version, v3)]

Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art

Joel Janai, Fatma Güney, Aseem Behl, Andreas Geiger

[Submitted on 26 Jul 2018 (v1), last revised 14 Jun 2019 (this version, v3)]

From handcrafted to deep local features

Gabriela Csurka, Christopher R. Dance, Martin Humenberger

[Submitted on 14 Sep 2020 (v1), last revised 16 Sep 2020 (this version, v2)]

Efficient Transformers: A Survey

Yi Tay, Mostafa Dehghani, Dara Bahri, Donald Metzler

Good Surveys: harder to distinguish

But there are couple of heuristics

- Good modern survey is getting updated
- available on arXiv/author page
- bad survey is “published” in obscure journals or behind the paywall

~~[http://www.ijetc.com/2_NCETEC001_\(p.9-17\)....](http://www.ijetc.com/2_NCETEC001_(p.9-17)....)~~ PDF
A SURVEY OF TRENDS IN LOCAL INVARIANT FEATURE ...
autor: JN Surekha · 2017 · Počet citací: 1 · Počet clánu: 2 — Many computer vision algorithms use feature detection as the initial step, so as a result, a very large number of feature detectors...
9 stránek

arXiv:1704.05519 (cs)

[Submitted on 18 Apr 2017 (v1), last revised 17 Mar 2021 (this version, v3)]

Computer Vision for Autonomous Vehicles: Problems, Datasets and State of the Art

Joel Janai, Fatma Güney, Aseem Behl, Andreas Geiger

[Submitted on 26 Jul 2018 (v1), last revised 14 Jun 2019 (this version, v3)]

From handcrafted to deep local features

Gabriela Csurka, Christopher R. Dance, Martin Humenberger

[Submitted on 14 Sep 2020 (v1), last revised 16 Sep 2020 (this version, v2)]

Efficient Transformers: A Survey

Yi Tay, Mostafa Dehghani, Dara Bahri, Donald Metzler

Good Surveys: harder to evaluate at least without experience

- Cheat: you can use new edition (2021) of Szeliski book as a meta-survey.

Computer Vision: Algorithms and Applications, 2nd ed.

© 2021 [Richard Szeliski](#), The University of Washington



How to read benchmark/survey

How to read benchmark/survey paper if you are new to the field

How to read benchmark/survey paper if you are new to the field

- First, understand the task, by reading the intro and thinking. Write down:

How to read benchmark/survey paper if you are new to the field

- First, understand the task, by reading the intro and thinking. Write down:
- Who are the “users” of the algorithm?

How to read benchmark/survey paper if you are new to the field

- First, understand the task, by reading the intro and thinking. Write down:
- Who are the “users” of the algorithm?
 - Where is the result is going to be used?

How to read benchmark/survey paper if you are new to the field

- First, understand the task, by reading the intro and thinking. Write down:
 - Who are the “users” of the algorithm?
 - Where is the result is going to be used?
 - What is important for the downstream task?

How to read benchmark/survey paper if you are new to the field

- First, understand the task, by reading the intro and thinking. Write down:
 - Who are the “users” of the algorithm?
 - Where is the result is going to be used?
 - What is important for the downstream task?
 - How one can measure it?

How to read benchmark/survey paper if you are new to the field

- First, understand the task, by reading the intro and thinking. Write down:
 - Who are the “users” of the algorithm?
 - Where is the result is going to be used?
 - What is important for the downstream task?
 - How one can measure it?
 - Where the data comes from in production?

How to read benchmark/survey paper if you are new to the field

- First, understand the task, by reading the intro and thinking. Write down:
 - Who are the “users” of the algorithm?
 - Where is the result is going to be used?
 - What is important for the downstream task?
 - How one can measure it?
 - Where the data comes from in production?
 - Can we avoid solving this task at all?

How to read benchmark/survey paper if you are new to the field

- First, understand the task, by reading the intro and thinking. Write down:
 - Who are the “users” of the algorithm?
 - Where is the result is going to be used?
 - What is important for the downstream task?
 - How one can measure it?
 - Where the data comes from in production?
 - Can we avoid solving this task at all?

No name manuscript No.
(will be inserted by the editor)

Image Matching Across Wide Baselines: From Paper to Practice

Yubei Jin · Dmytro Mishkin · Anastasia Miskulinck · Jiri Matas · Pascal Fua ·
Kwang Meo Yi · Eduard Trulls

Received: date / Accepted: date

Abstract We introduce a comprehensive benchmark for local features and robust estimation algorithms, focusing on the downstream task – the accuracy of the reconstructed camera pose – as our primary metric. Our pipeline’s modular structure allows easy integration, configuration, and combination of different methods and heuristics. This is demonstrated by embedding dozens of popular algorithms and evaluating them, from seminal works to the cutting edge of machine learning research. We show that with proper settings, classical solutions may still outperform the *perceived state of the art*.

Besides establishing the *actual state of the art*, the conducted experiments reveal unexpected properties of Structure from Motion (SfM) pipelines that can help improve this work was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant ‘Deep Vision Geometry Machines’ (RGPIN-2018-0378), by grants supplied by Compute Canada, and by Google’s Visual Positioning Service. DM and JM were supported by OFVV funded project CZ.02.1.01/0.0/0/16_01/M000763 ‘Research Center for Informatics’. DM was also supported by CIU student grant SGS17/185/CFKCN37V1 and by the Austrian Ministry for Transport, Innovation and Technology, the Federal Ministry for Digital and Economic Affairs, and the Province of Upper Austria in the frame of the COMET center SCIL. AM was supported by the Swiss National Science Foundation.

V. Jin, K.M. Yi
Visual Computing Group, University of Victoria
E-mail: {yubiejia, kyi}@uvic.ca
A. Miskulinck, P. Fua
Computer Vision Lab, École Polytechnique Fédérale de Lausanne
E-mail: {anastasiia.miskulinck, pascal.fua}@epfl.ch
D. Mishkin, J. Matas
Center for Machine Perception, Czech Technical University in Prague
E-mail: {mishkin, matas}@cmp.felk.cvut.cz
E. Trulls
Google Research
E-mail: trulls@google.com

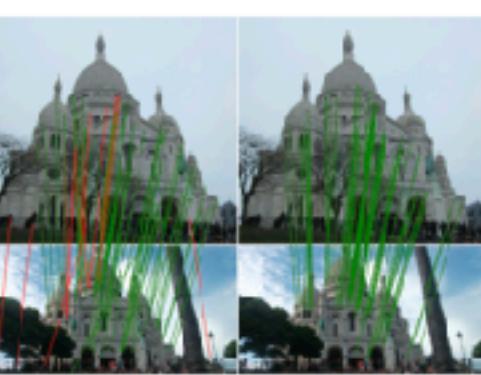


Fig. 1 Every paper claims to outperform the state of the art. Is this possible, or an artifact of insufficient validation? On the left, we show stereo matches obtained with D2-Net (20.9) [37]. A state-of-the-art local feature, using OpenCV RANSAC with its default settings. We color the inliers in green if they are coarse and in red otherwise. On the right, we show SIFT (1999) [34] with a carefully tuned MAGNET [15], so we see the latter performs much better. This illustrates our take home message: to correctly evaluate a method’s performance, it needs to be embedded within the pipeline used to solve a given problem and the different components in said pipeline need to be tuned carefully and jointly, which requires engineering and domain expertise. We fill this issue with a new, modular benchmark for sparse image matching, offering controls of built-in methods.

their performance, for both algorithmic and learned methods. Data and code are online¹, providing an easy-to-use and flexible framework for the benchmarking of local features and robust estimation methods, both *alongside* and *against* top-performing methods. This work provides a basis for the Image Matching Challenge².

¹ <https://github.com/vg-fau/image-matching-benchmark>

² <https://winter.univie.ac.at/image-matching-challenge/>

How to read benchmark/survey paper if you are new to the field

- First, understand the task, by reading the intro and thinking. Write down:
 - Who are the “users” of the algorithm?
 - Where is the result is going to be used?
 - What is important for the downstream task?
 - How one can measure it?
 - Where the data comes from in production?
 - Can we avoid solving this task at all?



Case study: image matching

Task understanding (1)

Noname manuscript No.
(will be inserted by the editor)

Image Matching Across Wide Baselines: From Paper to Practice

Yuhu Jin · Dmytro Mishkin · Anastasiia Mishchuk · Jiri Matas · Pascal Fua ·
Kwang Moo Yi · Eduard Trulls

Received: date / Accepted: date

Abstract We introduce a comprehensive benchmark for local features and robust estimation algorithms, focusing on the downstream task – the accuracy of the reconstructed camera pose – as our primary metric. Our pipeline’s modular structure allows easy integration, configuration, and combination of different methods and heuristics. This is demonstrated by embedding dozens of popular algorithms and evaluating them, from seminal works to the cutting edge of machine learning research. We show that with proper settings, classical solutions may still outperform the *perceived state of the art*.

Besides establishing the *actual state of the art*, the conducted experiments reveal unexpected properties of Structure from Motion (SfM) pipelines that can help improve

This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant “Deep Visual Geometry Machines” (RGPIN-2018-03738), by systems supplied by Compute Canada, and by Google’s Visual Positioning Service. DM and JM were supported by OP VVV funded project CZ.02.1.01/0.0/0.0/16_019/0000765 “Research Center for Informatics”. DM was also supported by CTU student grant SGS17/18/SOHRS/2/V/13 and by the Austrian Ministry for Transport, Innovation and Technology, the Federal Ministry for Digital and Economic Affairs, and the Province of Upper Austria in the frame of the COMET center SCCH. AM was supported by the Swiss National Science Foundation.

Y. Jin, K.M. Yi
Visual Computing Group, University of Victoria
E-mail: {yuhujin, kyi}@uvic.ca

A. Mishchuk, P. Fua
Computer Vision Lab, École Polytechnique Fédérale de Lausanne
E-mail: {anastasiia.mishchuk, pascal.fua}@epfl.ch

D. Mishkin, J. Matas
Center for Machine Perception, Czech Technical University in Prague
E-mail: {mishkin, matas}@cmp.felk.cvut.cz

E. Trulls
Google Research
E-mail: trulls@google.com

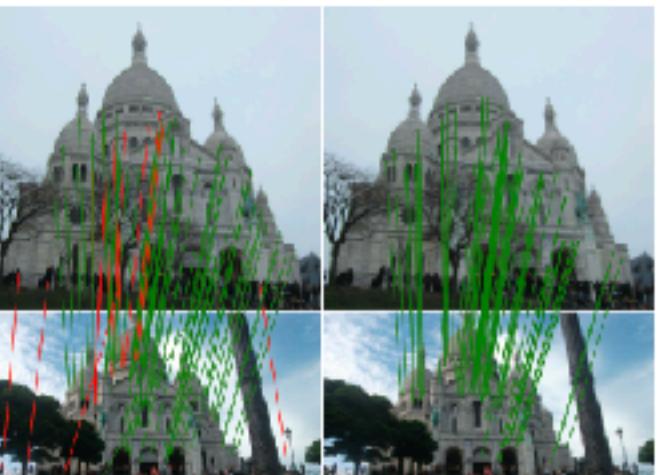


Fig. 1 Every paper claims to outperform the state of the art. Is this possible, or an artifact of inefficient validation? On the left, we show stereo matches obtained with D2-Net (2019) [17], a state-of-the-art local feature, using OpenCV RANSAC with its default settings. We color the inliers in green if they are correct and in red otherwise. On the right, we show SIFT (1999) [34] with a carefully tuned MAGSAC [15]—notice how the latter performs much better. This illustrates our take-home message: to correctly evaluate a method’s performance, it needs to be embedded within the pipeline used to solve a given problem, and the different components in said pipeline need to be tuned carefully and jointly, which requires engineering and domain expertise. We fill this need with a new, modular benchmark for sparse image matching, incorporating dozens of built-in methods.

their performance, for both algorithmic and learned methods. Data and code are online¹, providing an easy-to-use and flexible framework for the benchmarking of local features and robust estimation methods, both *alongside* and *against* top-performing methods. This work provides a basis for the Image Matching Challenge².

¹ <https://github.com/veg/uvic/image-matching-benchmark>
² <https://vision.uvic.ca/image-matching-challenge>

Case study: image matching

Task understanding (1)

- What is the task?

Noname manuscript No.
(will be inserted by the editor)

Image Matching Across Wide Baselines: From Paper to Practice

Yuhu Jin · Dmytro Mishkin · Anastasiia Mishchuk · Jiri Matas · Pascal Fua ·
Kwang Moo Yi · Eduard Trulls

Received: date / Accepted: date

Abstract We introduce a comprehensive benchmark for local features and robust estimation algorithms, focusing on the downstream task – the accuracy of the reconstructed camera pose – as our primary metric. Our pipeline’s modular structure allows easy integration, configuration, and combination of different methods and heuristics. This is demonstrated by embedding dozens of popular algorithms and evaluating them, from seminal works to the cutting edge of machine learning research. We show that with proper settings, classical solutions may still outperform the *perceived state of the art*.

Besides establishing the *actual state of the art*, the conducted experiments reveal unexpected properties of Structure from Motion (SfM) pipelines that can help improve

This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant “Deep Visual Geometry Machines” (RGPIN-2018-03738), by systems supplied by Compute Canada, and by Google’s Visual Positioning Service. DM and JM were supported by OP VVV funded project CZ.02.1.01/0.0/0.0/16_019/0000765 “Research Center for Informatics”. DM was also supported by CTU student grant SGS17/18/SOHRS/2/V/13 and by the Austrian Ministry for Transport, Innovation and Technology, the Federal Ministry for Digital and Economic Affairs, and the Province of Upper Austria in the frame of the COMET center SCCH. AM was supported by the Swiss National Science Foundation.

Y. Jin, K.M. Yi
Visual Computing Group, University of Victoria
E-mail: {yuhujin, kyi}@uvic.ca

A. Mishchuk, P. Fua
Computer Vision Lab, École Polytechnique Fédérale de Lausanne
E-mail: {anastasiia.mishchuk, pascal.fua}@epfl.ch

D. Mishkin, J. Matas
Center for Machine Perception, Czech Technical University in Prague
E-mail: {mishkin, matas}@cmp.felk.cvut.cz

E. Trulls
Google Research
E-mail: trulls@google.com

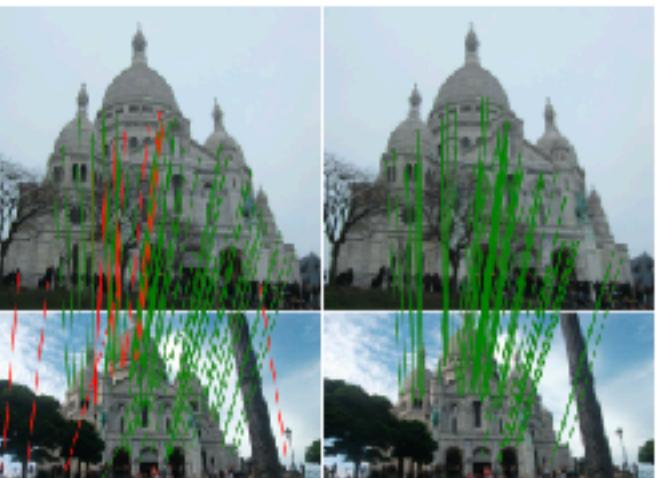


Fig. 1 Every paper claims to outperform the state of the art. Is this possible, or an artifact of inefficient validation? On the left, we show stereo matches obtained with D2-Net (2019) [17], a state-of-the-art local feature, using OpenCV RANSAC with its default settings. We color the inliers in green if they are correct and in red otherwise. On the right, we show SIFT (1999) [34] with a carefully tuned MAGSAC [15]—notice how the latter performs much better. This illustrates our take-home message: to correctly evaluate a method’s performance, it needs to be embedded within the pipeline used to solve a given problem, and the different components in said pipeline need to be tuned carefully and jointly, which requires engineering and domain expertise. We fill this need with a new, modular benchmark for sparse image matching, incorporating dozens of built-in methods.

their performance, for both algorithmic and learned methods. Data and code are online¹, providing an easy-to-use and flexible framework for the benchmarking of local features and robust estimation methods, both *alongside* and *against* top-performing methods. This work provides a basis for the Image Matching Challenge².

¹ <https://github.com/veg/uvic/image-matching-benchmark>
² <https://vision.uvic.ca/image-matching-challenge>

Case study: image matching

Task understanding (1)

- What is the task?

- Image matching.

Noname manuscript No.
(will be inserted by the editor)

Image Matching Across Wide Baselines: From Paper to Practice

Yuhe Jin · Dmytro Mishkin · Anastasiia Mishchuk · Jiri Matas · Pascal Fua ·
Kwang Moo Yi · Eduard Trulls

Received: date / Accepted: date

Abstract We introduce a comprehensive benchmark for local features and robust estimation algorithms, focusing on the downstream task – the accuracy of the reconstructed camera pose – as our primary metric. Our pipeline’s modular structure allows easy integration, configuration, and combination of different methods and heuristics. This is demonstrated by embedding dozens of popular algorithms and evaluating them, from seminal works to the cutting edge of machine learning research. We show that with proper settings, classical solutions may still outperform the *perceived state of the art*.

Besides establishing the *actual state of the art*, the conducted experiments reveal unexpected properties of Structure from Motion (SfM) pipelines that can help improve

This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant “Deep Visual Geometry Machines” (RGPIN-2018-03738), by systems supplied by Compute Canada, and by Google’s Visual Positioning Service. DM and JM were supported by OP VVV funded project CZ.02.1.01/0.0/0.016_0190000765 “Research Center for Informatics”. DM was also supported by CTU student grant SGS17/18/SOHRS/2/V/13 and by the Austrian Ministry for Transport, Innovation and Technology, the Federal Ministry for Digital and Economic Affairs, and the Province of Upper Austria in the frame of the COMET center SCCH. AM was supported by the Swiss National Science Foundation.

Y. Jin, K.M. Yi
Visual Computing Group, University of Victoria
E-mail: {yuhejin, kyi}@uvic.ca

A. Mishchuk, P. Fua
Computer Vision Lab, École Polytechnique Fédérale de Lausanne
E-mail: {anastasiia.mishchuk, pascal.fua}@epfl.ch

D. Mishkin, J. Matas
Center for Machine Perception, Czech Technical University in Prague
E-mail: {mishkin, matas}@cmp.felk.cvut.cz

E. Trulls
Google Research
E-mail: trulls@google.com

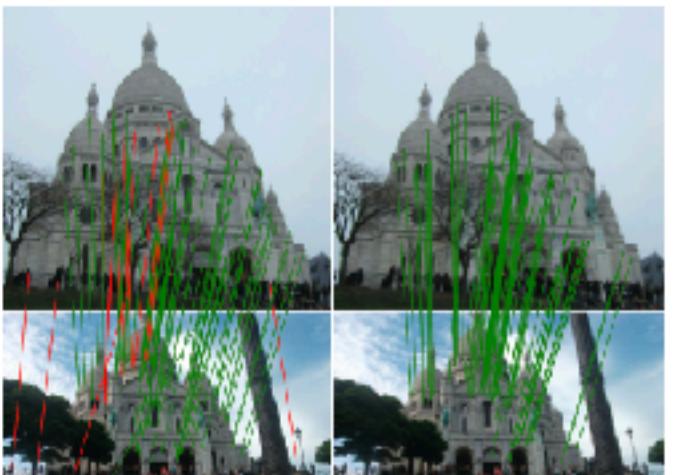


Fig. 1 Every paper claims to outperform the state of the art. Is this possible, or an artifact of inefficient validation? On the left, we show stereo matches obtained with D2-Net (2019) [17], a state-of-the-art local feature, using OpenCV RANSAC with its default settings. We color the inliers in green if they are correct and in red otherwise. On the right, we show SIFT (1999) [34] with a carefully tuned MAGSAC [15]—notice how the latter performs much better. This illustrates our take-home message: to correctly evaluate a method’s performance, it needs to be embedded within the pipeline used to solve a given problem, and the different components in said pipeline need to be tuned carefully and jointly, which requires engineering and domain expertise. We fill this need with a new, modular benchmark for sparse image matching, incorporating dozens of built-in methods.

their performance, for both algorithmic and learned methods. Data and code are online¹, providing an easy-to-use and flexible framework for the benchmarking of local features and robust estimation methods, both *alongside* and *against* top-performing methods. This work provides a basis for the Image Matching Challenge².

¹ <https://github.com/veg/uvic/image-matching-benchmark>
² <https://vision.uvic.ca/image-matching-challenge>

Case study: image matching

Task understanding (1)

- What is the task?
 - Image matching.
- What are the inputs and output of the task?

Noname manuscript No.
(will be inserted by the editor)

Image Matching Across Wide Baselines: From Paper to Practice

Yuhu Jin · Dmytro Mishkin · Anastasiia Mishchuk · Jiri Matas · Pascal Fua ·
Kwang Moo Yi · Eduard Trulls

Received: date / Accepted: date

Abstract We introduce a comprehensive benchmark for local features and robust estimation algorithms, focusing on the downstream task – the accuracy of the reconstructed camera pose – as our primary metric. Our pipeline’s modular structure allows easy integration, configuration, and combination of different methods and heuristics. This is demonstrated by embedding dozens of popular algorithms and evaluating them, from seminal works to the cutting edge of machine learning research. We show that with proper settings, classical solutions may still outperform the *perceived state of the art*.

Besides establishing the *actual state of the art*, the conducted experiments reveal unexpected properties of Structure from Motion (SfM) pipelines that can help improve

This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant “Deep Visual Geometry Machines” (RGPIN-2018-03738), by systems supplied by Compute Canada, and by Google’s Visual Positioning Service. DM and JM were supported by OP VVV funded project CZ.02.1.01/0.0/0.016_019/0000765 “Research Center for Informatics”. DM was also supported by CTU student grant SGS17/18/SOHRS/2/V/13 and by the Austrian Ministry for Transport, Innovation and Technology, the Federal Ministry for Digital and Economic Affairs, and the Province of Upper Austria in the frame of the COMET center SCCH. AM was supported by the Swiss National Science Foundation.

Y. Jin, K.M. Yi
Visual Computing Group, University of Victoria
E-mail: {yuhujin, kyi}@uvic.ca

A. Mishchuk, P. Fua
Computer Vision Lab, École Polytechnique Fédérale de Lausanne
E-mail: {anastasiia.mishchuk, pascal.fua}@epfl.ch

D. Mishkin, J. Matas
Center for Machine Perception, Czech Technical University in Prague
E-mail: {mishkin, matas}@cmp.felk.cvut.cz

E. Trulls
Google Research
E-mail: trulls@google.com

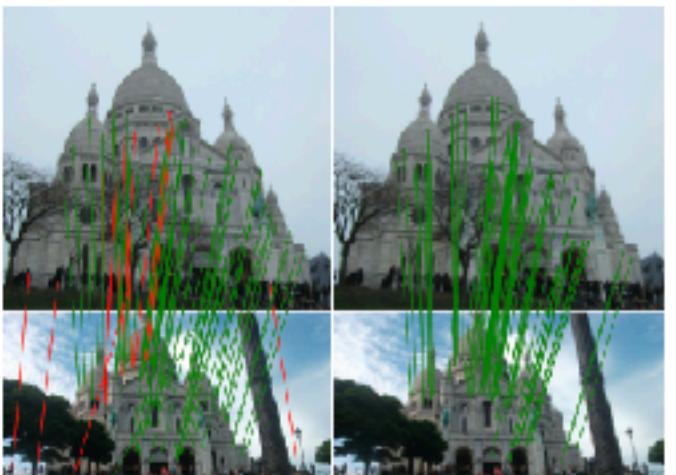


Fig. 1 Every paper claims to outperform the state of the art. Is this possible, or an artifact of inefficient validation? On the left, we show stereo matches obtained with D2-Net (2019) [17], a state-of-the-art local feature, using OpenCV RANSAC with its default settings. We color the inliers in green if they are correct and in red otherwise. On the right, we show SIFT (1999) [34] with a carefully tuned MAGSAC [15]—notice how the latter performs much better. This illustrates our take-home message: to correctly evaluate a method’s performance, it needs to be embedded within the pipeline used to solve a given problem, and the different components in said pipeline need to be tuned carefully and jointly, which requires engineering and domain expertise. We fill this need with a new, modular benchmark for sparse image matching, incorporating dozens of built-in methods.

their performance, for both algorithmic and learned methods. Data and code are online¹, providing an easy-to-use and flexible framework for the benchmarking of local features and robust estimation methods, both *alongside* and *against* top-performing methods. This work provides a basis for the Image Matching Challenge².

¹ <https://github.com/veg-uvic/image-matching-benchmark>
² <https://vision.uvic.ca/image-matching-challenge>

Case study: image matching

Task understanding (1)

- What is the task?
 - Image matching.
- What are the inputs and output of the task?
 - Missing in the intro! Need to google

Noname manuscript No.
(will be inserted by the editor)

Image Matching Across Wide Baselines: From Paper to Practice

Yuhu Jin · Dmytro Mishkin · Anastasiia Mishchuk · Jiri Matas · Pascal Fua ·
Kwang Moo Yi · Eduard Trulls

Received: date / Accepted: date

Abstract We introduce a comprehensive benchmark for local features and robust estimation algorithms, focusing on the downstream task – the accuracy of the reconstructed camera pose – as our primary metric. Our pipeline’s modular structure allows easy integration, configuration, and combination of different methods and heuristics. This is demonstrated by embedding dozens of popular algorithms and evaluating them, from seminal works to the cutting edge of machine learning research. We show that with proper settings, classical solutions may still outperform the *perceived state of the art*.

Besides establishing the *actual state of the art*, the conducted experiments reveal unexpected properties of Structure from Motion (SfM) pipelines that can help improve

This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant “Deep Visual Geometry Machines” (RGPIN-2018-03738), by systems supplied by Compute Canada, and by Google’s Visual Positioning Service. DM and JM were supported by OP VVV funded project CZ.02.1.01/0.0/0.0/16_019/0000765 “Research Center for Informatics”. DM was also supported by CTU student grant SGS17/18/SOHRS/2/V/13 and by the Austrian Ministry for Transport, Innovation and Technology, the Federal Ministry for Digital and Economic Affairs, and the Province of Upper Austria in the frame of the COMET center SCCH. AM was supported by the Swiss National Science Foundation.

Y. Jin, K.M. Yi
Visual Computing Group, University of Victoria
E-mail: {yuhujin, kyi}@uvic.ca

A. Mishchuk, P. Fua
Computer Vision Lab, École Polytechnique Fédérale de Lausanne
E-mail: {anastasiia.mishchuk, pascal.fua}@epfl.ch

D. Mishkin, J. Matas
Center for Machine Perception, Czech Technical University in Prague
E-mail: {mishkin, matas}@cmp.felk.cvut.cz

E. Trulls
Google Research
E-mail: trulls@google.com

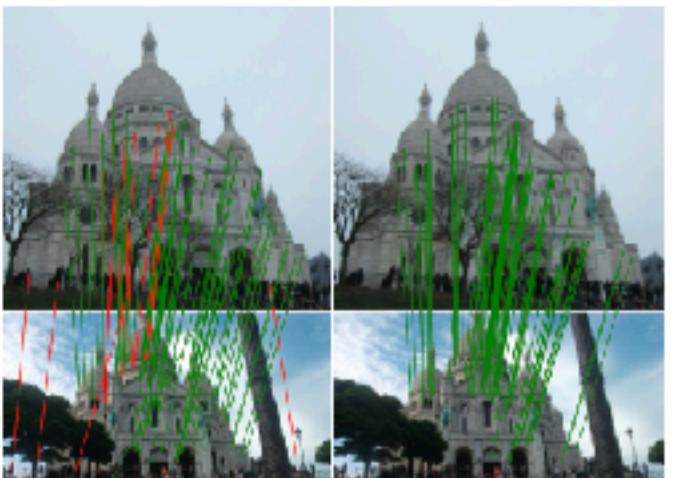


Fig. 1 Every paper claims to outperform the state of the art. Is this possible, or an artifact of inefficient validation? On the left, we show stereo matches obtained with D2-Net (2019) [17], a state-of-the-art local feature, using OpenCV RANSAC with its default settings. We color the inliers in green if they are correct and in red otherwise. On the right, we show SIFT (1999) [34] with a carefully tuned MAGSAC [15]—notice how the latter performs much better. This illustrates our take-home message: to correctly evaluate a method’s performance, it needs to be embedded within the pipeline used to solve a given problem, and the different components in said pipeline need to be tuned carefully and jointly, which requires engineering and domain expertise. We fill this need with a new, modular benchmark for sparse image matching, incorporating dozens of built-in methods.

their performance, for both algorithmic and learned methods. Data and code are online¹, providing an easy-to-use and flexible framework for the benchmarking of local features and robust estimation methods, both *alongside* and *against* top-performing methods. This work provides a basis for the Image Matching Challenge².

¹ <https://github.com/veg/uvic/image-matching-benchmark>
² <https://vision.uvic.ca/image-matching-challenge>

Case study: image matching

Task understanding (1)

- What is the task?
 - Image matching.
- What are the inputs and output of the task?
- Missing in the intro! Need to google

A screenshot of a Google search results page. The search query "what is image matching" is entered in the search bar. Below the search bar, there are several navigation links: "Vše" (All), "Obrázky" (Images), "Nákupy" (Shopping), "Videa" (Videos), "Zprávy" (News), and "Více" (More). The main search results area shows a snippet of text: "Image matching is an important concept in computer vision and object recognition. Images of the same item can be taken from any angle, with any lighting and scale. This as well as occlusion may cause problems for recognition. But ultimately, they still show the same item and should be categorized that way." At the bottom of the snippet, there is a link: "https://ai.stanford.edu/~syueung/cvweb/tutorial2".

Noname manuscript No.
(will be inserted by the editor)

Image Matching Across Wide Baselines: From Paper to Practice

Yuhu Jin · Dmytro Mishkin · Anastasiia Mishchuk · Jiri Matas · Pascal Fua ·
Kwang Moo Yi · Eduard Trulls

Received: date / Accepted: date

Abstract We introduce a comprehensive benchmark for local features and robust estimation algorithms, focusing on the downstream task – the accuracy of the reconstructed camera pose – as our primary metric. Our pipeline’s modular structure allows easy integration, configuration, and combination of different methods and heuristics. This is demonstrated by embedding dozens of popular algorithms and evaluating them, from seminal works to the cutting edge of machine learning research. We show that with proper settings, classical solutions may still outperform the *perceived state of the art*.

Besides establishing the *actual state of the art*, the conducted experiments reveal unexpected properties of Structure from Motion (SfM) pipelines that can help improve

This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant “Deep Visual Geometry Machines” (RGPIN-2018-03738), by systems supplied by Compute Canada, and by Google’s Visual Positioning Service. DM and JM were supported by OP VVV funded project CZ.02.1.01/0.0/0.0/16_019/0000765 “Research Center for Informatics”. DM was also supported by CTU student grant SGS17/18/SOHRS/2/V/13 and by the Austrian Ministry for Transport, Innovation and Technology, the Federal Ministry for Digital and Economic Affairs, and the Province of Upper Austria in the frame of the COMET center SCCH. AM was supported by the Swiss National Science Foundation.

Y. Jin, K.M. Yi
Visual Computing Group, University of Victoria
E-mail: {yuhujin, kyi}@uvic.ca

A. Mishchuk, P. Fua
Computer Vision Lab, École Polytechnique Fédérale de Lausanne
E-mail: {anastasiia.mishchuk, pascal.fua}@epfl.ch

D. Mishkin, J. Matas
Center for Machine Perception, Czech Technical University in Prague
E-mail: {mishkin, matas}@cmp.felk.cvut.cz

E. Trulls
Google Research
E-mail: trulls@google.com

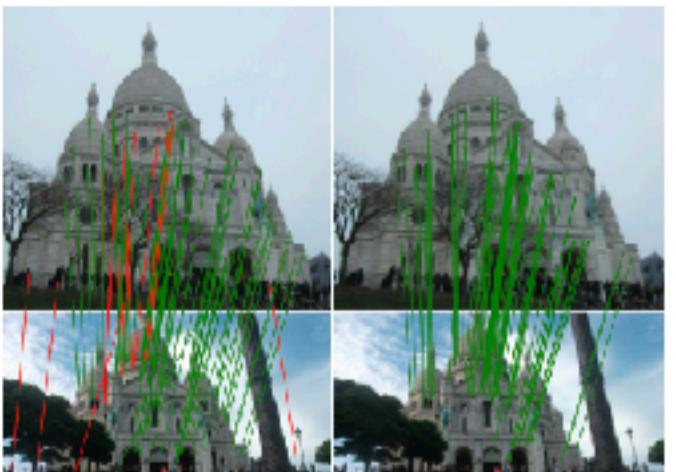


Fig. 1 Every paper claims to outperform the state of the art. Is this possible, or an artifact of inefficient validation? On the left, we show stereo matches obtained with D2-Net (2019) [17], a state-of-the-art local feature, using OpenCV RANSAC with its default settings. We color the inliers in green if they are correct and in red otherwise. On the right, we show SIFT (1999) [34] with a carefully tuned MAGSAC [13]—notice how the latter performs much better. This illustrates our take home message: to correctly evaluate a method’s performance, it needs to be embedded within the pipeline used to solve a given problem, and the different components in said pipeline need to be tuned carefully and jointly, which requires engineering and domain expertise. We fill this need with a new, modular benchmark for sparse image matching, incorporating dozens of built-in methods.

their performance, for both algorithmic and learned methods. Data and code are online¹, providing an easy-to-use and flexible framework for the benchmarking of local features and robust estimation methods, both *alongside* and *against* top-performing methods. This work provides a basis for the Image Matching Challenge².

¹ <https://github.com/veg/uvic/image-matching-benchmark>
² <https://vision.uvic.ca/image-matching-challenge>

Case study: image matching

Task understanding (1)

- Inputs: images of the “same item”.
- Output (after more googling): correspondences and relative camera pose
- Note: people can use different terms for the same thing:
 - “Image matching” == “Wide baseline stereo”

The wide multiple baseline stereo (WxBS) is a process of establishing a sufficient number of pixel or region correspondences from two or more images depicting the same scene to estimate the geometric relationship between cameras, which produced these images. Mar 27, 2020

<https://ducha-aiki.github.io> › 2020/03/27 › intro ::

The Role of Wide Baseline Stereo in the Deep Learning World



Case study: image matching

Task understanding (2)

Case study: image matching

Task understanding (2)

- Who are the “users” of the algorithm? Where is the result is going to be used?

Case study: image matching

Task understanding (2)

- Who are the “users” of the algorithm? Where is the result is going to be used?

1 Introduction

Matching two or more views of a scene is at the core of fundamental computer vision problems, including image retrieval [54, 7, 78, 104, 70], 3D reconstruction [3, 47, 90, 122], re-localization [85, 86, 57], and SLAM [68, 33, 34]. Despite

Case study: image matching

Task understanding (2)

- Who are the “users” of the algorithm? Where is the result is going to be used?
 - Image retrieval

1 Introduction

Matching two or more views of a scene is at the core of fundamental computer vision problems, including image retrieval [54, 7, 78, 104, 70], 3D reconstruction [3, 47, 90, 122], re-localization [85, 86, 57], and SLAM [68, 33, 34]. Despite

Case study: image matching

Task understanding (2)

- Who are the “users” of the algorithm? Where is the result is going to be used?
 - Image retrieval
 - 3D reconstruction

1 Introduction

Matching two or more views of a scene is at the core of fundamental computer vision problems, including image retrieval [54, 7, 78, 104, 70], 3D reconstruction [3, 47, 90, 122], re-localization [85, 86, 57], and SLAM [68, 33, 34]. Despite

Case study: image matching

Task understanding (2)

- Who are the “users” of the algorithm? Where is the result is going to be used?
 - Image retrieval
 - 3D reconstruction
 - visual localization

1 Introduction

Matching two or more views of a scene is at the core of fundamental computer vision problems, including image retrieval [54, 7, 78, 104, 70], 3D reconstruction [3, 47, 90, 122], re-localization [85, 86, 57], and SLAM [68, 33, 34]. Despite

Case study: image matching

Task understanding (2)

- Who are the “users” of the algorithm? Where is the result is going to be used?
 - Image retrieval
 - 3D reconstruction
 - visual localization
 - SLAM

1 Introduction

Matching two or more views of a scene is at the core of fundamental computer vision problems, including image retrieval [54, 7, 78, 104, 70], 3D reconstruction [3, 47, 90, 122], re-localization [85, 86, 57], and SLAM [68, 33, 34]. Despite

Case study: image matching

Task understanding (2)

- Who are the “users” of the algorithm? Where is the result is going to be used?
 - Image retrieval
 - 3D reconstruction
 - visual localization
 - SLAM
- Now we may, or may not need to check about this “users”

1 Introduction

Matching two or more views of a scene is at the core of fundamental computer vision problems, including image retrieval [54, 7, 78, 104, 70], 3D reconstruction [3, 47, 90, 122], re-localization [85, 86, 57], and SLAM [68, 33, 34]. Despite

Case study: image matching

Task understanding (2)

Case study: image matching

Task understanding (2)

- What is important for the downstream task?

Case study: image matching

Task understanding (2)

- What is important for the downstream task?
 - Image retrieval – if the result is relevant to query

Case study: image matching

Task understanding (2)

- What is important for the downstream task?
 - Image retrieval – if the result is relevant to query
 - 3D reconstruction – accuracy and completeness of the reconstruction

Case study: image matching

Task understanding (2)

- What is important for the downstream task?
 - Image retrieval – if the result is relevant to query
 - 3D reconstruction – accuracy and completeness of the reconstruction
 - visual localization – camera pose accuracy, robustness to nuisance

Case study: image matching

Task understanding (2)

- What is important for the downstream task?
 - Image retrieval – if the result is relevant to query
 - 3D reconstruction – accuracy and completeness of the reconstruction
 - visual localization – camera pose accuracy, robustness to nuisance
 - SLAM – camera pose accuracy, realtime speed, blur-robustness

Case study: image matching

Task understanding (2)

- What is important for the downstream task?
 - Image retrieval – if the result is relevant to query
 - 3D reconstruction – accuracy and completeness of the reconstruction
 - visual localization – camera pose accuracy, robustness to nuisance
 - SLAM – camera pose accuracy, realtime speed, blur-robustness

Is it relevant to me?

Case study: image matching

Task understanding (3)

Case study: image matching

Task understanding (3)

- We have already googled and checked other papers to answer the questions, which are necessary to understand the paper we are reading now

Case study: image matching

Task understanding (3)

- We have already googled and checked other papers to answer the questions, which are necessary to understand the paper we are reading now
- It is OK to deviate (a lot), when you are reading

Case study: image matching

Task understanding (3)

- We have already googled and checked other papers to answer the questions, which are necessary to understand the paper we are reading now
- It is OK to deviate (a lot), when you are reading
- Especially when you are new to the area

Case study: image matching

Task understanding (3)

- We have already googled and checked other papers to answer the questions, which are necessary to understand the paper we are reading now
- It is OK to deviate (a lot), when you are reading
- Especially when you are new to the area
- It is a good idea to “learn in public” and write answers to the question in a blog post.

Case study: image matching

How can we measure that?

- What is important for the downstream task?
 - Image retrieval – if the result is relevant to query
 - 3D reconstruction – accuracy and completeness of the reconstruction
 - visual localization – camera pose accuracy, robustness to nuisance
 - SLAM – camera pose accuracy, realtime speed, blur-robustness

Case study: image matching

How can we measure that?

- What is important for the downstream task?
 - Image retrieval – if the result is relevant to query
 - 3D reconstruction – accuracy and completeness of the reconstruction
 - visual localization – camera pose accuracy, robustness to nuisance
 - SLAM – camera pose accuracy, realtime speed, blur-robustness

Recall and precision

Case study: image matching

How can we measure that?

- What is important for the downstream task?
 - Image retrieval – if the result is relevant to query
 - 3D reconstruction – accuracy and completeness of the reconstruction
 - visual localization – camera pose accuracy, robustness to nuisance
 - SLAM – camera pose accuracy, realtime speed, blur-robustness

Recall and precision

Lots of metrics

Case study: image matching

How can we measure that?

- What is important for the downstream task?
 - Image retrieval – if the result is relevant to query
Recall and precision
 - 3D reconstruction – accuracy and completeness of the reconstruction
Lots of metrics
 - visual localization – camera pose accuracy, robustness to nuisance
Pose error, accuracy
 - SLAM – camera pose accuracy, realtime speed, blur-robustness

Case study: image matching

How can we measure that?

- What is important for the downstream task?
 - Image retrieval – if the result is relevant to query
Recall and precision
 - 3D reconstruction – accuracy and completeness of the reconstruction
Lots of metrics
 - visual localization – camera pose accuracy, robustness to nuisance
Pose error, accuracy
 - SLAM – camera pose accuracy, realtime speed, blur-robustness
Pose error, fps

Case study: image matching

Task understanding (4)

Case study: image matching

Task understanding (4)

- Where the data comes from in production?

Case study: image matching

Task understanding (4)

- Where the data comes from in production?
 - Image retrieval – user upload, can be anything

Case study: image matching

Task understanding (4)

- Where the data comes from in production?
 - Image retrieval – user upload, can be anything
 - 3D reconstruction – professional cameras, drones, smartphone, historical photos, etc

Case study: image matching

Task understanding (4)

- Where the data comes from in production?
 - Image retrieval – user upload, can be anything
 - 3D reconstruction – professional cameras, drones, smartphone, historical photos, etc
 - visual localization – robot camera, car, smartphone, internet photo (OSI)

Case study: image matching

Task understanding (4)

- Where the data comes from in production?
 - Image retrieval – user upload, can be anything
 - 3D reconstruction – professional cameras, drones, smartphone, historical photos, etc
 - visual localization – robot camera, car, smartphone, internet photo (OSI)
 - SLAM – robot, UAV, smart glasses, car

Back to the paper

What does the paper evaluate?

Case study: Image matching

- What does paper measure?
- Where the data in the paper comes from?
- Where the ground truth labels come from?

What does the paper evaluate?

Again, we are speaking purely about the task now

3 The Image Matching Challenge PhotoTourism Dataset

While it is possible to obtain very accurate poses and depth maps under controlled scenarios with devices like LIDAR, this is costly and requires a specific set-up that does not scale well. For example, Strecha’s dataset [95] follows that approach but contains only 19 images. We argue that a truly representative dataset must contain a wider range of transformations – including different imaging devices, time of day, weather, partial occlusions, etc. Phototourism images satisfy this condition and are readily available.

We thus build on 25 collections of popular landmarks originally selected in [47, 101], each with hundreds to thousands of images. Images are downsampled with bilinear interpolation to a maximum size of 1024 pixels along the long-side and their poses were obtained with COLMAP [90], which provides the (pseudo) ground truth. We do exhaustive image matching before Bundle Adjustment – unlike [92], which uses only 100 pairs for each image – and thus provide enough matching images for any conventional SfM to return near-perfect results in standard conditions.

What does the paper evaluate?

Again, we are speaking purely about the task now

- What does paper measure – mean average accuracy

3 The Image Matching Challenge PhotoTourism Dataset

While it is possible to obtain very accurate poses and depth maps under controlled scenarios with devices like LIDAR, this is costly and requires a specific set-up that does not scale well. For example, Strecha's dataset [95] follows that approach but contains only 19 images. We argue that a truly representative dataset must contain a wider range of transformations – including different imaging devices, time of day, weather, partial occlusions, etc. Phototourism images satisfy this condition and are readily available.

We thus build on 25 collections of popular landmarks originally selected in [47, 101], each with hundreds to thousands of images. Images are downsampled with bilinear interpolation to a maximum size of 1024 pixels along the long-side and their poses were obtained with COLMAP [90], which provides the (pseudo) ground truth. We do exhaustive image matching before Bundle Adjustment – unlike [92], which uses only 100 pairs for each image – and thus provide enough matching images for any conventional SfM to return near-perfect results in standard conditions.

What does the paper evaluate?

Again, we are speaking purely about the task now

- What does paper measure – mean average accuracy
- Where the data in the paper comes from?

3 The Image Matching Challenge PhotoTourism Dataset

While it is possible to obtain very accurate poses and depth maps under controlled scenarios with devices like LIDAR, this is costly and requires a specific set-up that does not scale well. For example, Strecha's dataset [95] follows that approach but contains only 19 images. We argue that a truly representative dataset must contain a wider range of transformations – including different imaging devices, time of day, weather, partial occlusions, etc. Phototourism images satisfy this condition and are readily available.

We thus build on 25 collections of popular landmarks originally selected in [47, 101], each with hundreds to thousands of images. Images are downsampled with bilinear interpolation to a maximum size of 1024 pixels along the long-side and their poses were obtained with COLMAP [90], which provides the (pseudo) ground truth. We do exhaustive image matching before Bundle Adjustment – unlike [92], which uses only 100 pairs for each image – and thus provide enough matching images for any conventional SfM to return near-perfect results in standard conditions.

What does the paper evaluate?

Again, we are speaking purely about the task now

- What does paper measure – mean average accuracy
- Where the data in the paper comes from?
 - YCC100M dataset, from flickr (user photos)

3 The Image Matching Challenge PhotoTourism Dataset

While it is possible to obtain very accurate poses and depth maps under controlled scenarios with devices like LIDAR, this is costly and requires a specific set-up that does not scale well. For example, Strecha’s dataset [95] follows that approach but contains only 19 images. We argue that a truly representative dataset must contain a wider range of transformations – including different imaging devices, time of day, weather, partial occlusions, etc. Phototourism images satisfy this condition and are readily available.

We thus build on 25 collections of popular landmarks originally selected in [47, 101], each with hundreds to thousands of images. Images are downsampled with bilinear interpolation to a maximum size of 1024 pixels along the long-side and their poses were obtained with COLMAP [90], which provides the (pseudo) ground truth. We do exhaustive image matching before Bundle Adjustment – unlike [92], which uses only 100 pairs for each image – and thus provide enough matching images for any conventional SfM to return near-perfect results in standard conditions.

What does the paper evaluate?

Again, we are speaking purely about the task now

- What does paper measure – mean average accuracy
- Where the data in the paper comes from?
 - YCC100M dataset, from flickr (user photos)
- Where the ground truth labels come from?

3 The Image Matching Challenge PhotoTourism Dataset

While it is possible to obtain very accurate poses and depth maps under controlled scenarios with devices like LIDAR, this is costly and requires a specific set-up that does not scale well. For example, Strecha’s dataset [95] follows that approach but contains only 19 images. We argue that a truly representative dataset must contain a wider range of transformations – including different imaging devices, time of day, weather, partial occlusions, etc. Phototourism images satisfy this condition and are readily available.

We thus build on 25 collections of popular landmarks originally selected in [47, 101], each with hundreds to thousands of images. Images are downsampled with bilinear interpolation to a maximum size of 1024 pixels along the long-side and their poses were obtained with COLMAP [90], which provides the (pseudo) ground truth. We do exhaustive image matching before Bundle Adjustment – unlike [92], which uses only 100 pairs for each image – and thus provide enough matching images for any conventional SfM to return near-perfect results in standard conditions.

What does the paper evaluate?

Again, we are speaking purely about the task now

- What does paper measure – mean average accuracy
- Where the data in the paper comes from?
 - YCC100M dataset, from flickr (user photos)
- Where the ground truth labels come from?
 - COLMAP (SIFT-based 3d reconstruction software)

3 The Image Matching Challenge PhotoTourism Dataset

While it is possible to obtain very accurate poses and depth maps under controlled scenarios with devices like LIDAR, this is costly and requires a specific set-up that does not scale well. For example, Strecha’s dataset [95] follows that approach but contains only 19 images. We argue that a truly representative dataset must contain a wider range of transformations – including different imaging devices, time of day, weather, partial occlusions, etc. Phototourism images satisfy this condition and are readily available.

We thus build on 25 collections of popular landmarks originally selected in [47, 101], each with hundreds to thousands of images. Images are downsampled with bilinear interpolation to a maximum size of 1024 pixels along the long-side and their poses were obtained with COLMAP [90], which provides the (pseudo) ground truth. We do exhaustive image matching before Bundle Adjustment – unlike [92], which uses only 100 pairs for each image – and thus provide enough matching images for any conventional SfM to return near-perfect results in standard conditions.

How the paper data & metric matches user?

Case study: image matching, metric

- User metrics:
 - Image retrieval – Recall and precision
 - 3D reconstruction – many metrics
 - visual localization – camera pose accuracy
 - SLAM – camera pose accuracy, realtime speed

How the paper data & metric matches user?

Case study: image matching, metric

- User metrics:
 - Image retrieval – Recall and precision
 - 3D reconstruction – many metrics
 - visual localization – camera pose accuracy
 - SLAM – camera pose accuracy, realtime speed
- Paper metrics:
 - mean average accuracy on camera pose

How the paper data & metric matches user?

Case study: image matching, metric

- User metrics:
 - Image retrieval – Recall and precision
 - 3D reconstruction – many metrics
 - visual localization – camera pose accuracy
 - SLAM – camera pose accuracy, realtime speed
 - Paper metrics:
 - mean average accuracy on camera pose
- 

How the paper data & metric matches user?

Case study: image matching, metric

- User metrics:
 - Image retrieval – Recall and precision
 - 3D reconstruction – many metrics
 - visual localization – camera pose accuracy
 - SLAM – camera pose accuracy, realtime speed

Note: not clear, why paper says
that image matching is used in
image retrieval

Add to the ToDo list to figure out
(low priority)



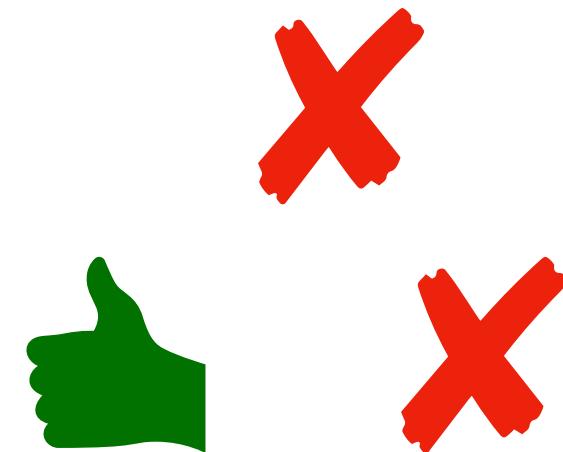
- Paper metrics:
 - mean average accuracy on camera pose

How the paper data & metric matches user?

Case study: image matching, metric

- User metrics:
 - Image retrieval – Recall and precision
 - 3D reconstruction – many metrics
 - visual localization – camera pose accuracy
 - SLAM – camera pose accuracy, realtime speed

Note: not clear, why paper says
that image matching is used in
image retrieval
Add to the ToDo list to figure out
(low priority)



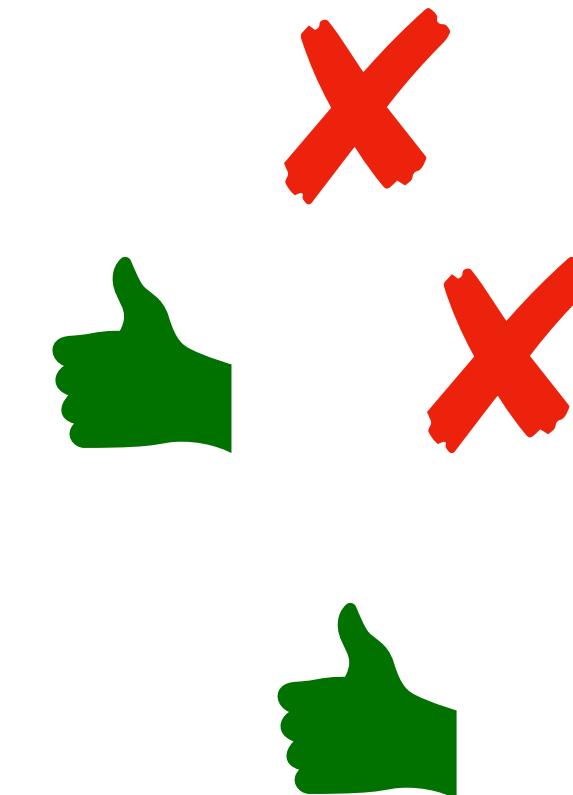
- Paper metrics:
 - mean average accuracy on camera pose

How the paper data & metric matches user?

Case study: image matching, metric

- User metrics:

- Image retrieval – Recall and precision
- 3D reconstruction – many metrics
- visual localization – camera pose accuracy
- SLAM – camera pose accuracy, realtime speed



Note: not clear, why paper says
that image matching is used in
image retrieval
Add to the ToDo list to figure out
(low priority)

- Paper metrics:

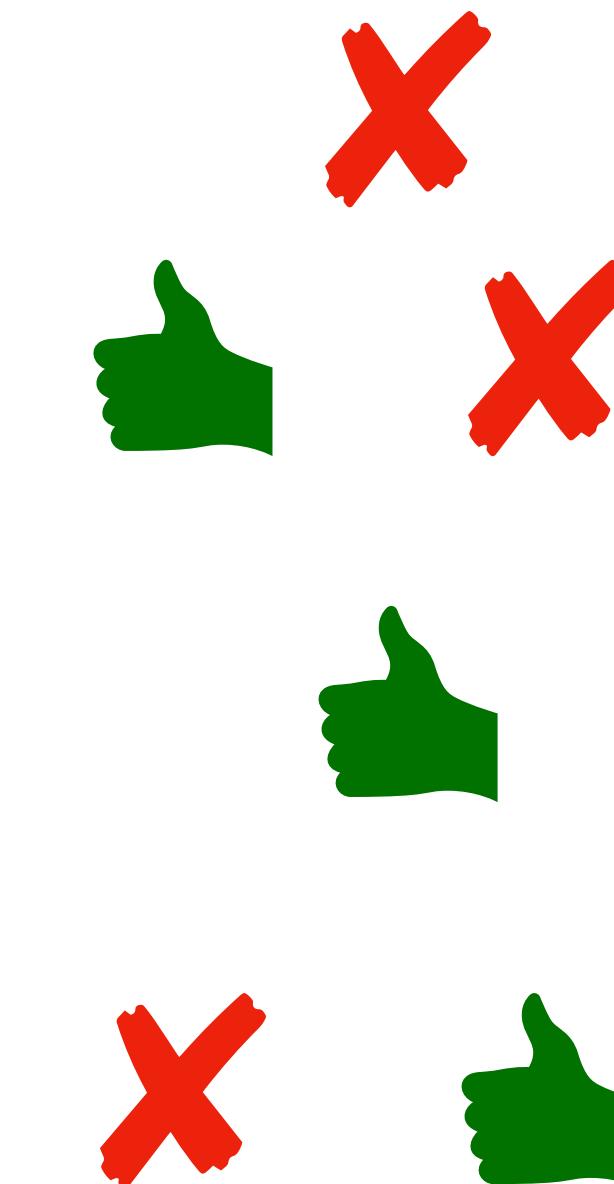
- mean average accuracy on camera pose

How the paper data & metric matches user?

Case study: image matching, metric

- User metrics:

- Image retrieval – Recall and precision
- 3D reconstruction – many metrics
- visual localization – camera pose accuracy
- SLAM – camera pose accuracy, realtime speed



Note: not clear, why paper says
that image matching is used in
image retrieval
Add to the ToDo list to figure out
(low priority)

- Paper metrics:

- mean average accuracy on camera pose

Case study: image matching

Case study: image matching, data

- Where the data comes from in production?
 - Image retrieval – user upload, can be anything
 - 3D reconstruction – professional cameras, drones, smartphone, historical photos, etc
 - visual localization – robot camera, car, smartphone, internet photo (OSI)
 - SLAM – robot, UAV, smart glasses, car

Case study: image matching

Case study: image matching, data

- Where the data comes from in production?
 - Image retrieval – user upload, can be anything
 - 3D reconstruction – professional cameras, drones, smartphone, historical photos, etc
 - visual localization – robot camera, car, smartphone, internet photo (OSI)
 - SLAM – robot, UAV, smart glasses, car
- Paper:
 - YCC100M dataset, from flickr (user photos)

Case study: image matching

Case study: image matching, data

- Where the data comes from in production?
 - Image retrieval – user upload, can be anything
 - 3D reconstruction – professional cameras, drones, smartphone, historical photos, etc
 - visual localization – robot camera, car, smartphone, internet photo (OSI)
 - SLAM – robot, UAV, smart glasses, car



- Paper:
 - YCC100M dataset, from flickr (user photos)

Case study: image matching

Case study: image matching, data

- Where the data comes from in production?
 - Image retrieval – user upload, can be anything
 - 3D reconstruction – professional cameras, drones, smartphone, historical photos, etc
 - visual localization – robot camera, car, smartphone, internet photo (OSI)
 - SLAM – robot, UAV, smart glasses, car



- Paper:
 - YCC100M dataset, from flickr (user photos)

Case study: image matching

Case study: image matching, data

- Where the data comes from in production?
 - Image retrieval – user upload, can be anything
 - 3D reconstruction – professional cameras, drones, smartphone, historical photos, etc
 - visual localization – robot camera, car, smartphone, internet photo (OSI)
 - SLAM – robot, UAV, smart glasses, car

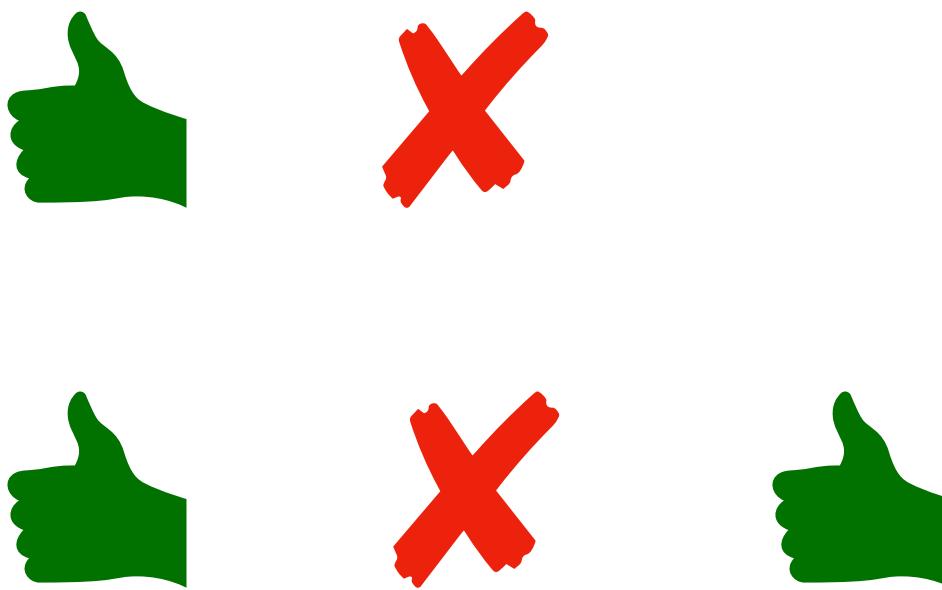


- Paper:
 - YCC100M dataset, from flickr (user photos)

Case study: image matching

Case study: image matching, data

- Where the data comes from in production?
 - Image retrieval – user upload, can be anything
 - 3D reconstruction – professional cameras, drones, smartphone, historical photos, etc
 - visual localization – robot camera, car, smartphone, internet photo (OSI)
 - SLAM – robot, UAV, smart glasses, car

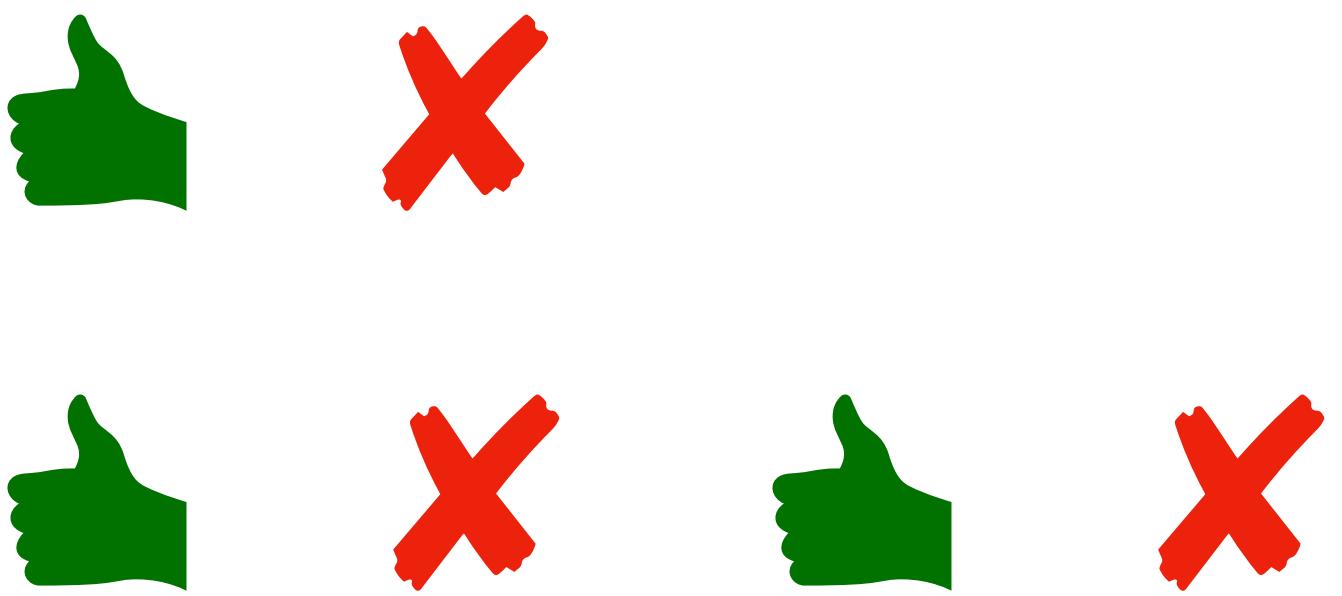


- Paper:
 - YCC100M dataset, from flickr (user photos)

Case study: image matching

Case study: image matching, data

- Where the data comes from in production?
 - Image retrieval – user upload, can be anything
 - 3D reconstruction – professional cameras, drones, smartphone, historical photos, etc
 - visual localization – robot camera, car, smartphone, internet photo (OSI)
 - SLAM – robot, UAV, smart glasses, car

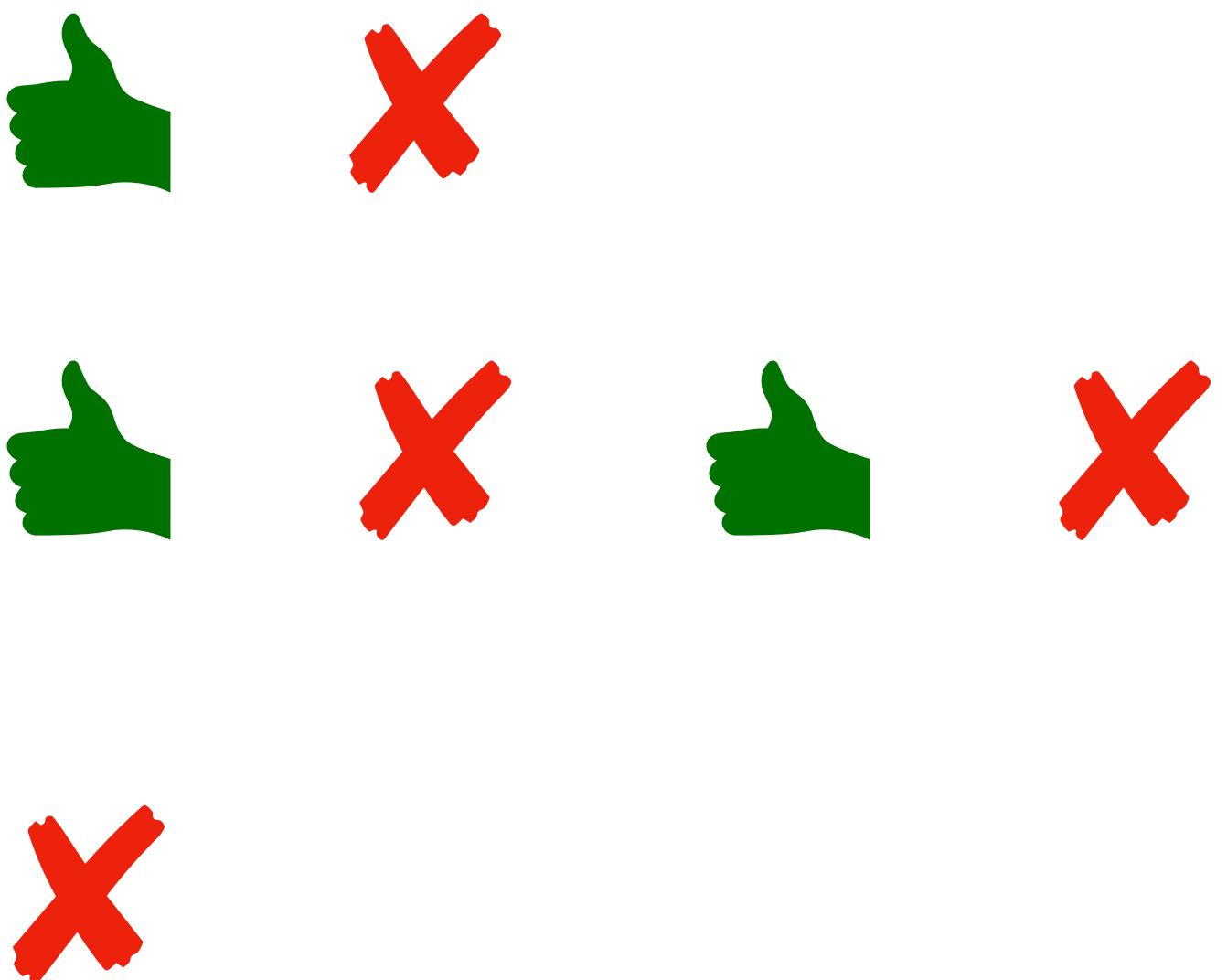


- Paper:
 - YCC100M dataset, from flickr (user photos)

Case study: image matching

Case study: image matching, data

- Where the data comes from in production?
 - Image retrieval – user upload, can be anything
 - 3D reconstruction – professional cameras, drones, smartphone, historical photos, etc
 - visual localization – robot camera, car, smartphone, internet photo (OSI)
 - SLAM – robot, UAV, smart glasses, car

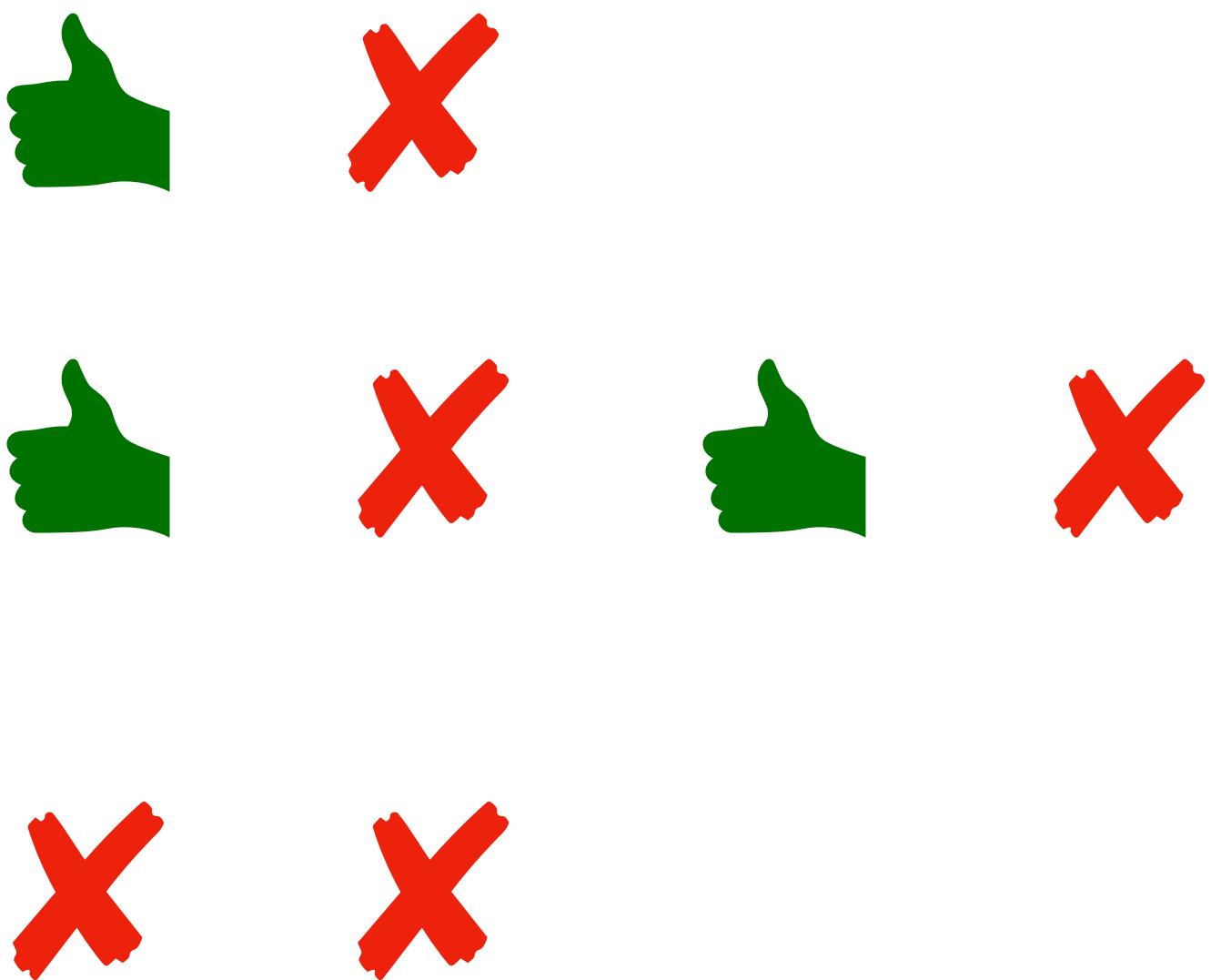


- Paper:
 - YCC100M dataset, from flickr (user photos)

Case study: image matching

Case study: image matching, data

- Where the data comes from in production?
 - Image retrieval – user upload, can be anything
 - 3D reconstruction – professional cameras, drones, smartphone, historical photos, etc
 - visual localization – robot camera, car, smartphone, internet photo (OSI)
 - SLAM – robot, UAV, smart glasses, car

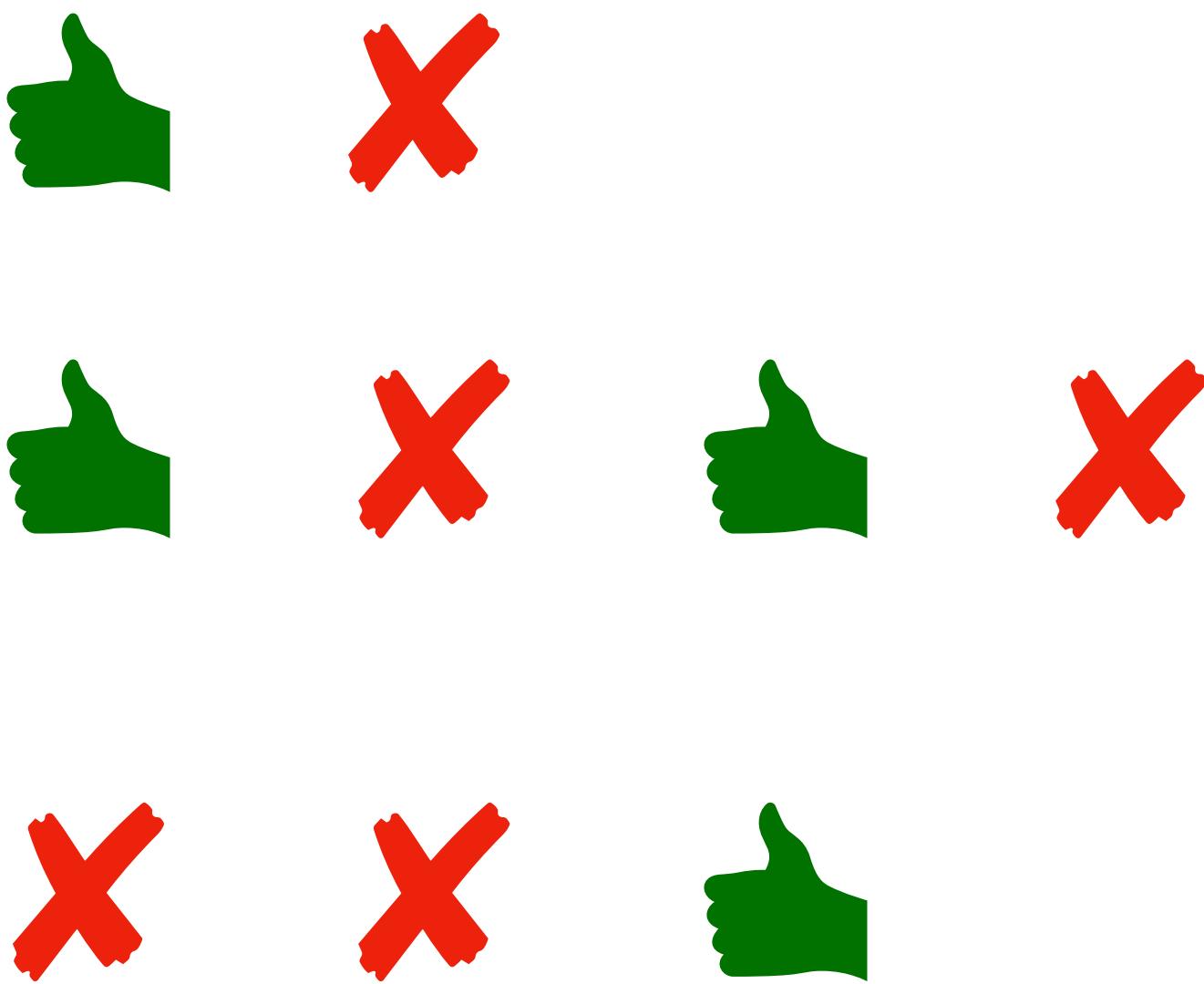


- Paper:
 - YCC100M dataset, from flickr (user photos)

Case study: image matching

Case study: image matching, data

- Where the data comes from in production?
 - Image retrieval – user upload, can be anything
 - 3D reconstruction – professional cameras, drones, smartphone, historical photos, etc
 - visual localization – robot camera, car, smartphone, internet photo (OSI)
 - SLAM – robot, UAV, smart glasses, car

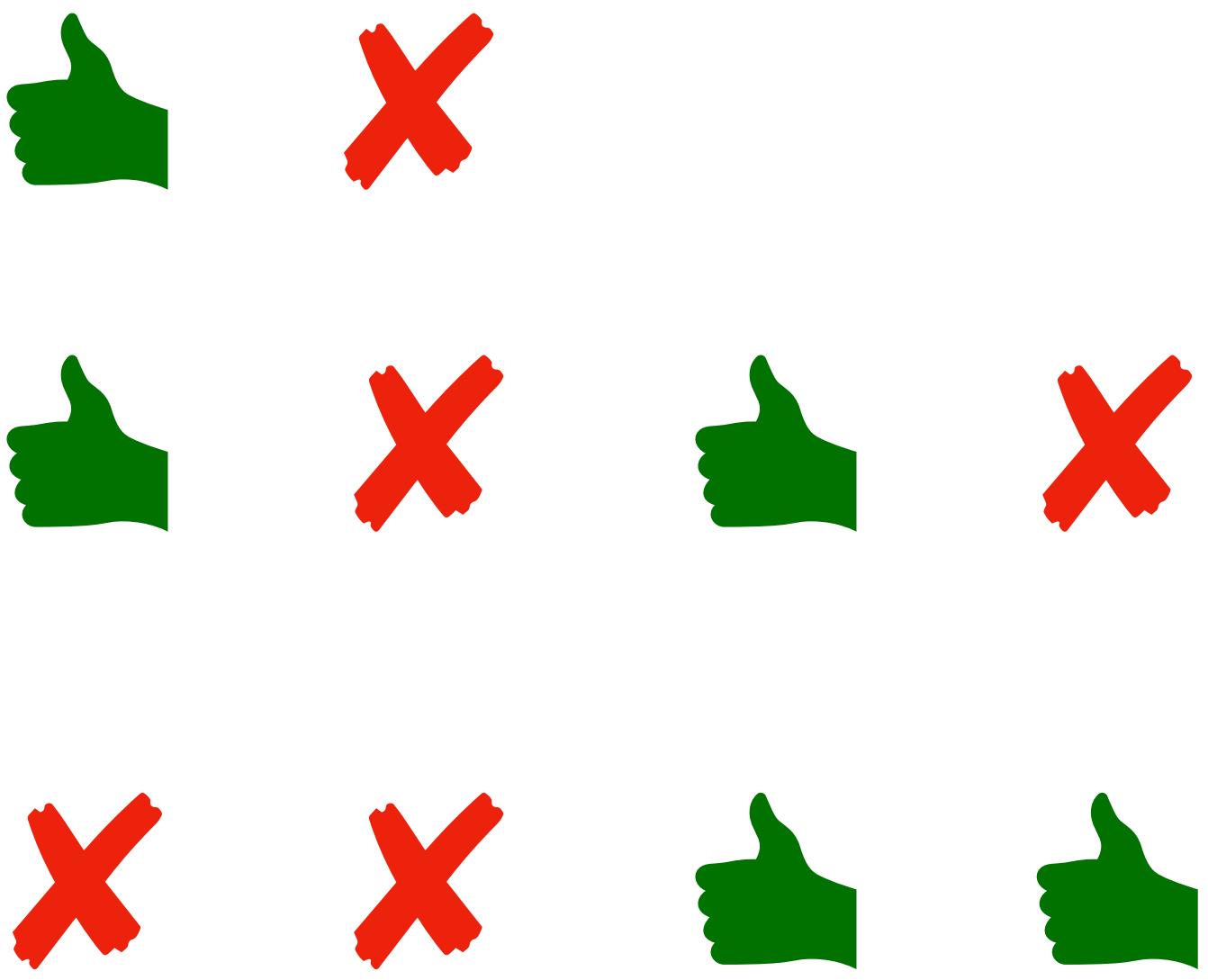


- Paper:
 - YCC100M dataset, from flickr (user photos)

Case study: image matching

Case study: image matching, data

- Where the data comes from in production?
 - Image retrieval – user upload, can be anything
 - 3D reconstruction – professional cameras, drones, smartphone, historical photos, etc
 - visual localization – robot camera, car, smartphone, internet photo (OSI)
 - SLAM – robot, UAV, smart glasses, car

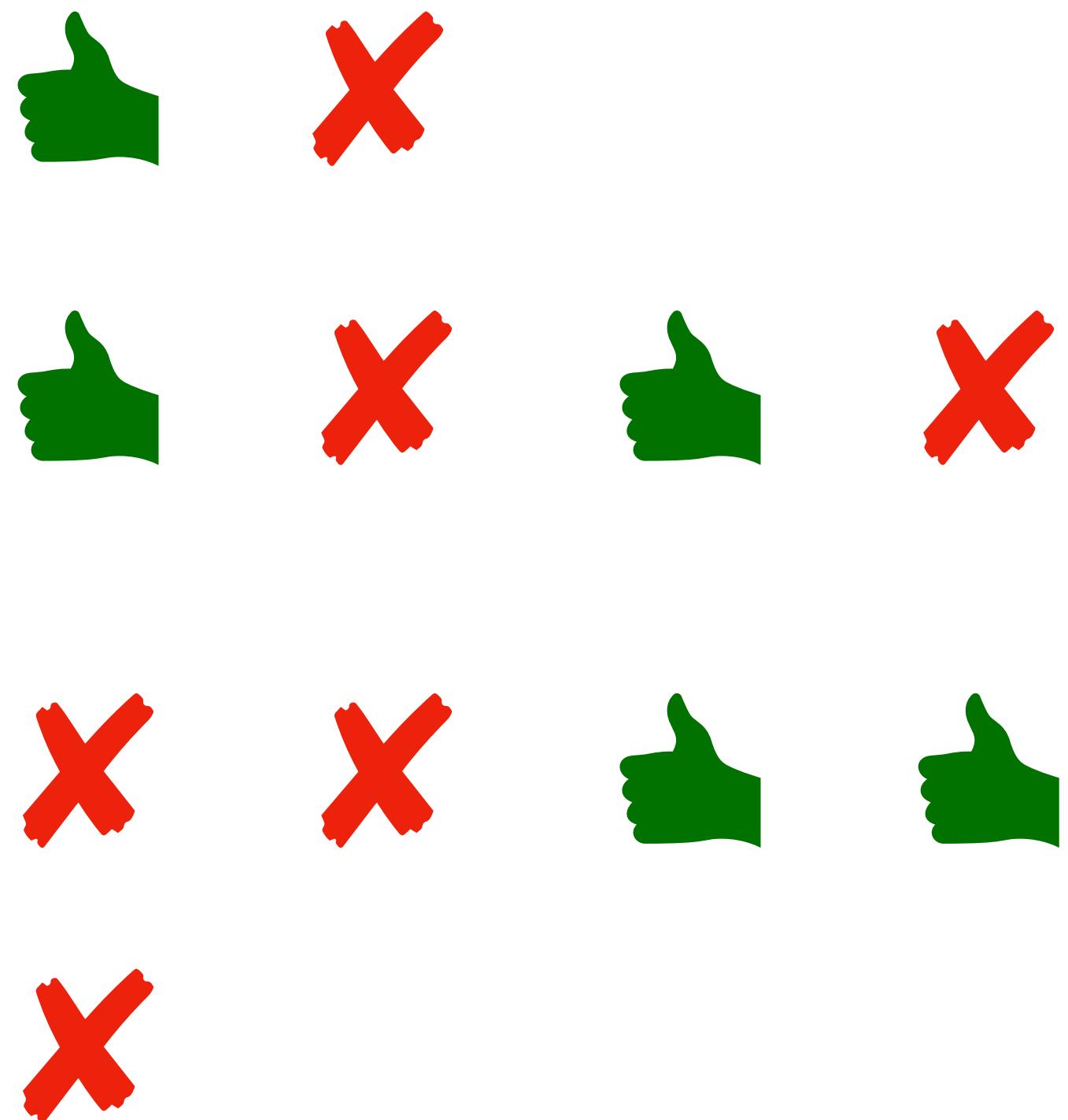


- Paper:
 - YCC100M dataset, from flickr (user photos)

Case study: image matching

Case study: image matching, data

- Where the data comes from in production?
 - Image retrieval – user upload, can be anything
 - 3D reconstruction – professional cameras, drones, smartphone, historical photos, etc
 - visual localization – robot camera, car, smartphone, internet photo (OSI)
 - SLAM – robot, UAV, smart glasses, car

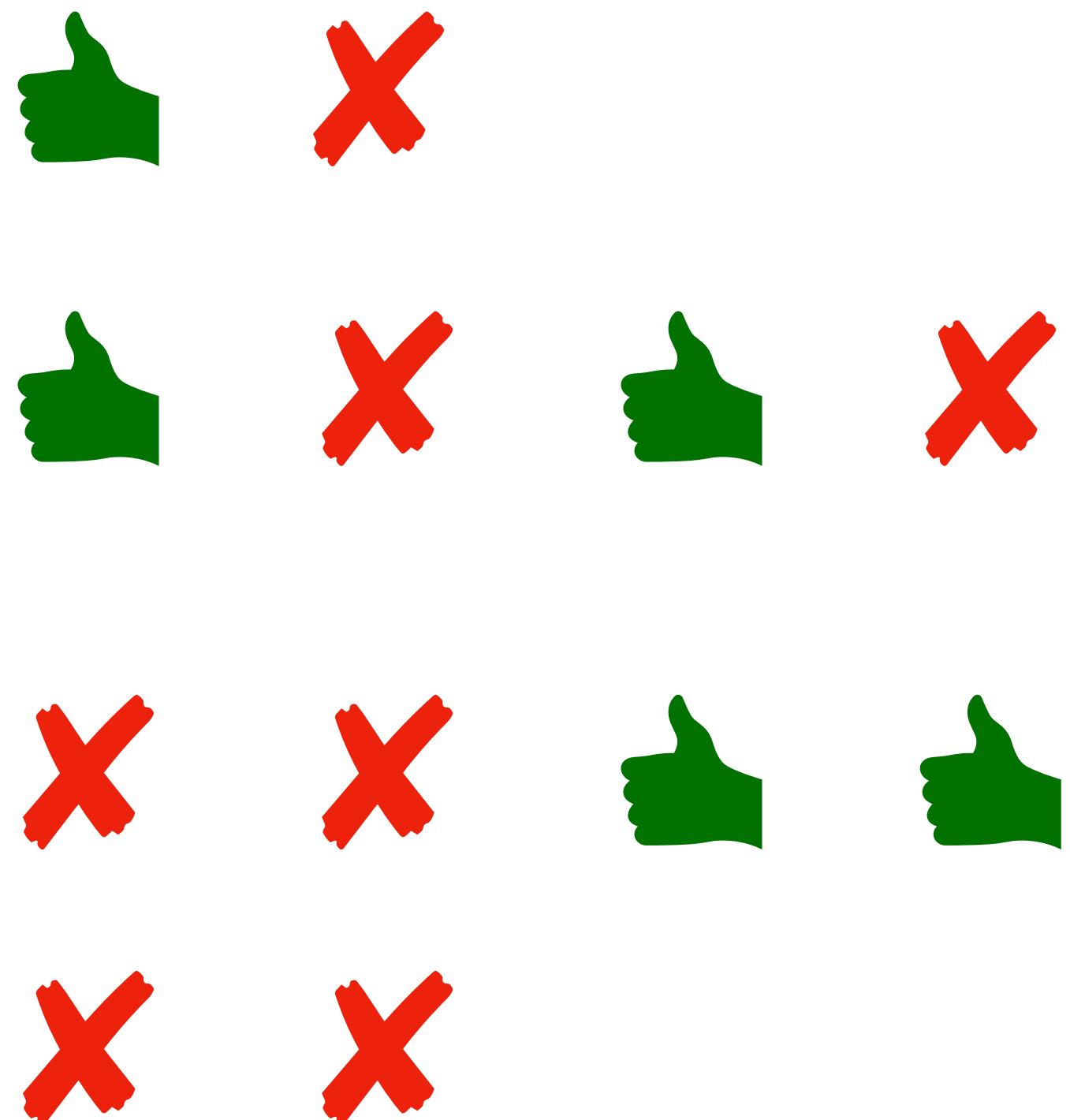


- Paper:
 - YCC100M dataset, from flickr (user photos)

Case study: image matching

Case study: image matching, data

- Where the data comes from in production?
 - Image retrieval – user upload, can be anything
 - 3D reconstruction – professional cameras, drones, smartphone, historical photos, etc
 - visual localization – robot camera, car, smartphone, internet photo (OSI)
 - SLAM – robot, UAV, smart glasses, car

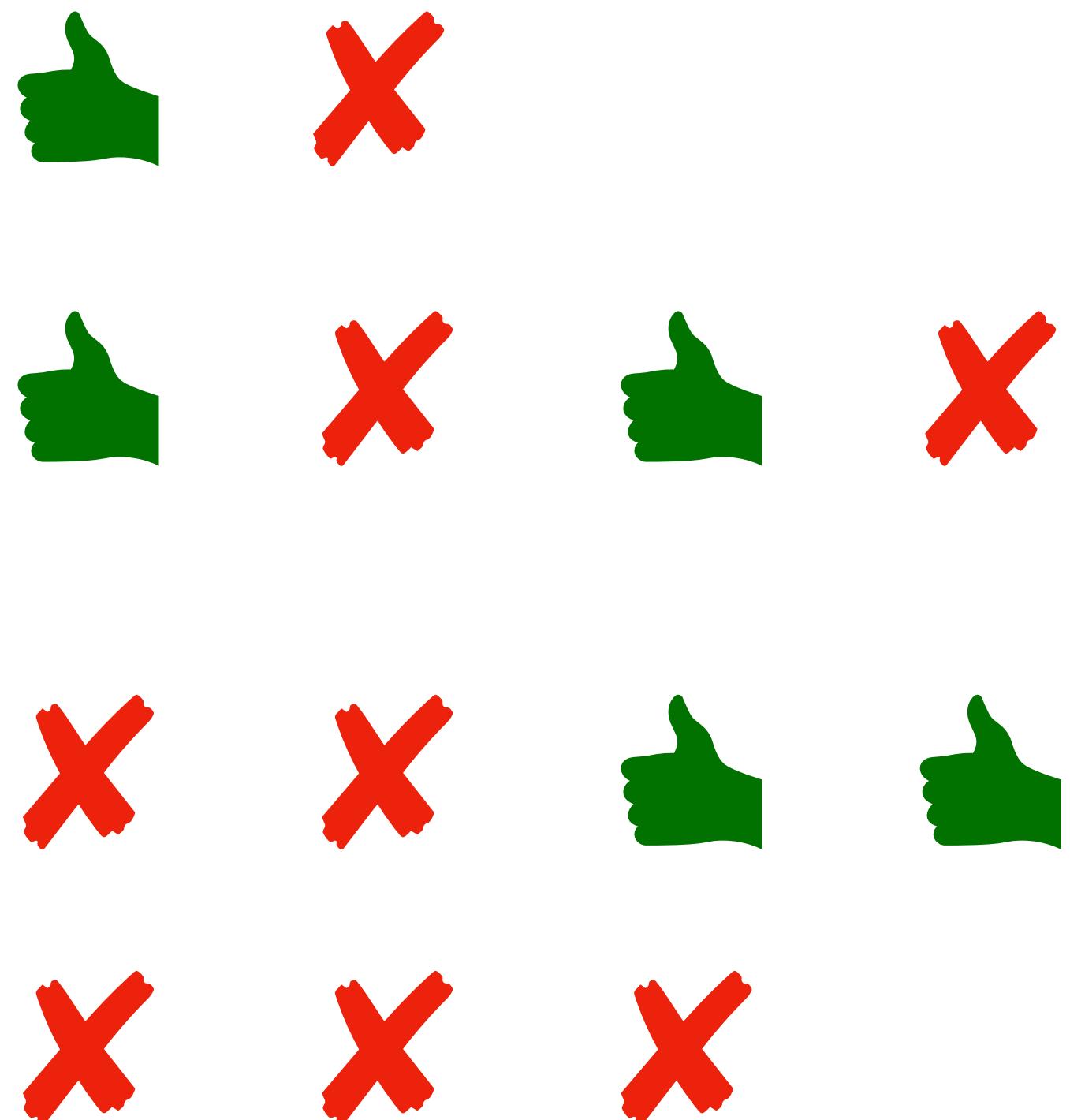


- Paper:
 - YCC100M dataset, from flickr (user photos)

Case study: image matching

Case study: image matching, data

- Where the data comes from in production?
 - Image retrieval – user upload, can be anything
 - 3D reconstruction – professional cameras, drones, smartphone, historical photos, etc
 - visual localization – robot camera, car, smartphone, internet photo (OSI)
 - SLAM – robot, UAV, smart glasses, car

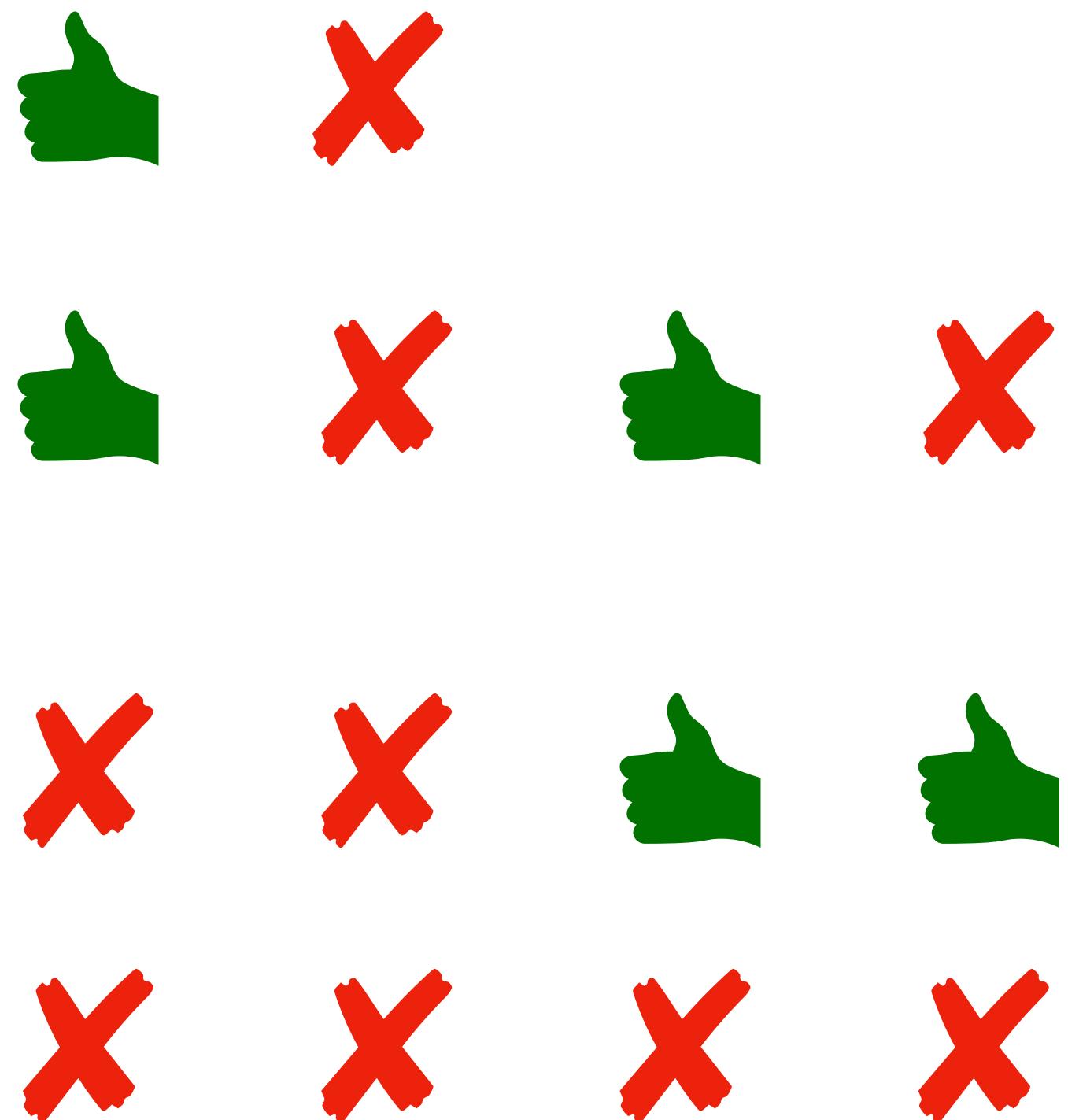


- Paper:
 - YCC100M dataset, from flickr (user photos)

Case study: image matching

Case study: image matching, data

- Where the data comes from in production?
 - Image retrieval – user upload, can be anything
 - 3D reconstruction – professional cameras, drones, smartphone, historical photos, etc
 - visual localization – robot camera, car, smartphone, internet photo (OSI)
 - SLAM – robot, UAV, smart glasses, car



- Paper:
 - YCC100M dataset, from flickr (user photos)

Case study: image matching

Case study: image matching, data

- How the paper results are relevant for

- Image retrieval

	Data	Metrics
• Image retrieval	Green thumbs up	Red X
• 3D reconstruction	Green thumbs up	Red X
• visual localization	Red X	Red X
• SLAM	Red X	Red X

- 3D reconstruction

- visual localization

- SLAM

Intermediate conclusion

- The paper findings are partially relevant for some use-cases in
 - visual localization
- 3D reconstruction
- Loosely relevant for the SLAM
- Not relevant at all for image retrieval

Noname manuscript No.
(will be inserted by the editor)

Image Matching Across Wide Baselines: From Paper to Practice

Yuhu Jin · Dmytro Mishkin · Anastasia Mishchuk · Jiri Matas · Pascal Fua ·
Kwang Moo Yi · Eduard Trulls

Received: date / Accepted: date

Abstract We introduce a comprehensive benchmark for local features and robust estimation algorithms, focusing on the downstream task – the accuracy of the reconstructed camera pose – as our primary metric. Our pipeline’s modular structure allows easy integration, configuration, and combination of different methods and heuristics. This is demonstrated by embedding dozens of popular algorithms and evaluating them, from seminal works to the cutting edge of machine learning research. We show that with proper settings, classical solutions may still outperform the *perceived state of the art*.

Besides establishing the *actual state of the art*, the conducted experiments reveal unexpected properties of Structure from Motion (SfM) pipelines that can help improve

This work was partially supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant “Deep Visual Geometry Machines” (RGPIN-2018-05788), by systems supplied by Compute Canada, and by Google’s Visual Processing Service. DM and JM were supported by OP VVV funded project CZ.02.1.01/0.0/0.0/16_019/0000765 “Research Center for Informatics”. DM was also supported by CTU student grant SG17/18/SOHSS2/19/13 and by the Austrian Ministry for Transport, Innovation and Technology, the Federal Ministry for Digital and Economic Affairs, and the Province of Upper Austria in the frame of the COMET center SICHT. AM was supported by the Swiss National Science Foundation.

Y. Jin, K.M. Yi
Visual Computing Group, University of Victoria
E-mail: {yuhujin, kyi}@uvic.ca

A. Mishchuk, P. Fua
Computer Vision Lab, École Polytechnique Fédérale de Lausanne
E-mail: {anastasia.mishchuk, pascal.fua}@epfl.ch

D. Mishkin, J. Matas
Center for Machine Perception, Czech Technical University in Prague
E-mail: {mishkin, matas}@cmp.felk.cvut.cz

E. Trulls
Google Research
E-mail: trulls@google.com

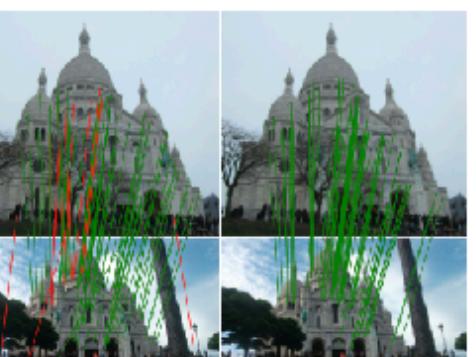


Fig. 1 Every paper claims to outperform the state-of-the-art, is this possible, or an artifact of insufficient validation? On the left, we show stereo matches obtained with D2-Net (2019)^[37], a state-of-the-art local feature, using OpenCV RANSAC with its default settings. We color the lines in green if they are correct and in red otherwise. On the right, we show SfI 1 (1999)^[34] with a carefully tuned MAISAC^[1]—notice how the latter performs much better. This illustrates our take-home message: to correctly evaluate a method’s performance, it needs to be embedded within the pipeline used to solve a given problem, and the different components in said pipeline need to be tuned carefully and jointly, which requires engineering and domain expertise. We fill this need with a new, modular benchmark for sparse image matching, incorporating dozens of built-in methods.

their performance, for both algorithmic and learned methods. Data and code are online¹, providing an easy-to-use and flexible framework for the benchmarking of local features and robust estimation methods, both *alongside* and *against* top-performing methods. This work provides a basis for the Image Matching Challenge².

¹ <https://github.com/vcg-uvic/image-matching-benchmark>
² <https://vision.uvic.ca/image-matching-challenge>

Why should I do all this crap?

Why should I do all this crap?

Why should I do all this crap?

If you need a truck,
do not watch racing car reviews



**Research sometimes deviates from practice
or often**

Research sometimes deviates from practice or often



Christoph Molnar
@ChristophMolnar

...

A lot of machine learning research has detached itself from solving real problems, and created their own "benchmark-islands".

How does this happen? And why are researchers not escaping this pattern?

A thread A small icon of a spool of purple thread with a loop of thread extending from it.

Research sometimes deviates from practice or often



Christoph Molnar
@ChristophMolnar

...

A lot of machine learning research has detached itself from solving real problems, and created their own "benchmark-islands".

How does this happen? And why are researchers not escaping this pattern?

A thread 

Research sometimes deviates from practice or often



Christoph Molnar
@ChristophMolnar

...

1. Establish a research topic. For example: Detect COVID-19 from x-ray images with machine learning.
2. First papers on this topic appear, and provide motivation for this research topic.
3. A precedent has been established: The initial papers pave the way for further research papers on the same topic.
4. The original motivation and assumptions for developing these predictive models are no longer questioned.
5. More researchers publish data and predictive models.

A lot of machine learning research has detached itself from solving real problems, and created their own "benchmark-islands".

How does this happen? And why are researchers not escaping this pattern?

A thread 

Research sometimes deviates from practice or often



Christoph Molnar
@ChristophMolnar

...

A lot of machine learning research has detached itself from solving real problems. and created their own

s not

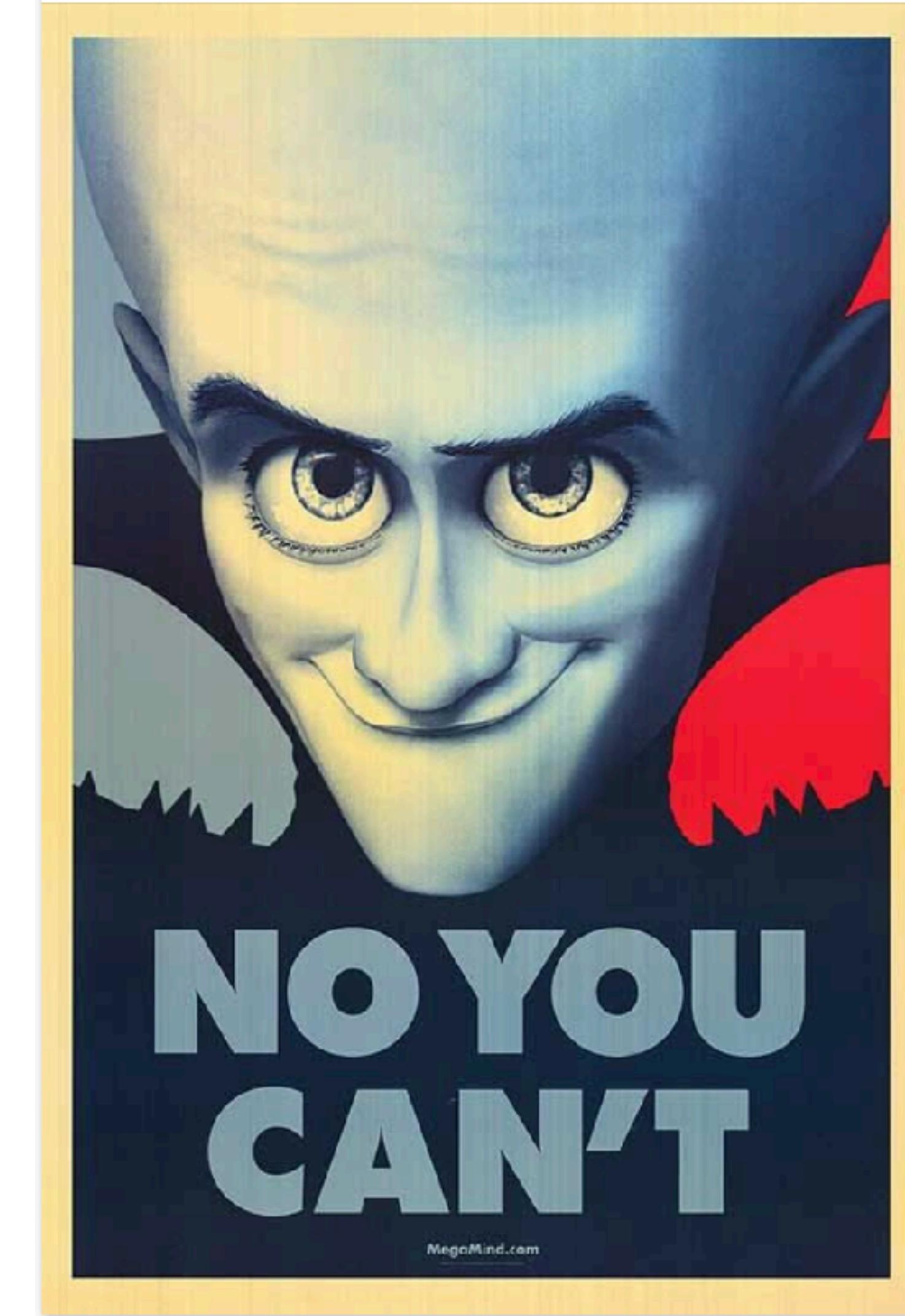
1. Establish a research topic. For example: Detect COVID-19 from x-ray images with machine learning.
2. First papers on this topic appear, and provide motivation for this research topic.
3. A precedent has been established: The initial papers pave the way for further research papers on the same topic.
4. The original motivation and assumptions for developing these predictive models are no longer questioned.
5. More researchers publish data and predictive models.
6. A community forms that cites and reviews each other.
7. Certain datasets become benchmarks and certain predictive models become state-of-the-art.
8. Predictive performance becomes the sole measure of progress, although improvements are becoming smaller.
9. Actual progress in solving the initial research questions has become irrelevant, even discouraged.
10. Decoupling from reality is complete.

OK, paper is relevant

OK, paper is relevant



OK, paper is relevant



Can we trust the paper?

*"The only correct attitude, when reading the paper is:
Why do these lying bastard lie to me?"*

David Forsyth

**Do the metrics in the paper make sense?
If you don't have experience, you can skip this**

Do the metrics in the paper make sense?

If you don't have experience, you can skip this

- Even if the metric is measuring a proper thing, it can still be misleading

Do the metrics in the paper make sense?

If you don't have experience, you can skip this

- Even if the metric is measuring a proper thing, it can still be misleading
 - Is it robust to outliers? E.g. 1 failure can kill average pose error over dataset

Do the metrics in the paper make sense?

If you don't have experience, you can skip this

- Even if the metric is measuring a proper thing, it can still be misleading
 - Is it robust to outliers? E.g. 1 failure can kill average pose error over dataset
 - Isn't it too robust to outliers? E.g. median error does not change if half of dataset is not solved at all

Do the metrics in the paper make sense?

If you don't have experience, you can skip this

- Even if the metric is measuring a proper thing, it can still be misleading
 - Is it robust to outliers? E.g. 1 failure can kill average pose error over dataset
 - Isn't it too robust to outliers? E.g. median error does not change if half of dataset is not solved at all
 - What is the “null” performance according to the metric?

Do the metrics in the paper make sense?

If you don't have experience, you can skip this

- Even if the metric is measuring a proper thing, it can still be misleading
 - Is it robust to outliers? E.g. 1 failure can kill average pose error over dataset
 - Isn't it too robust to outliers? E.g. median error does not change if half of dataset is not solved at all
 - What is the “null” performance according to the metric?

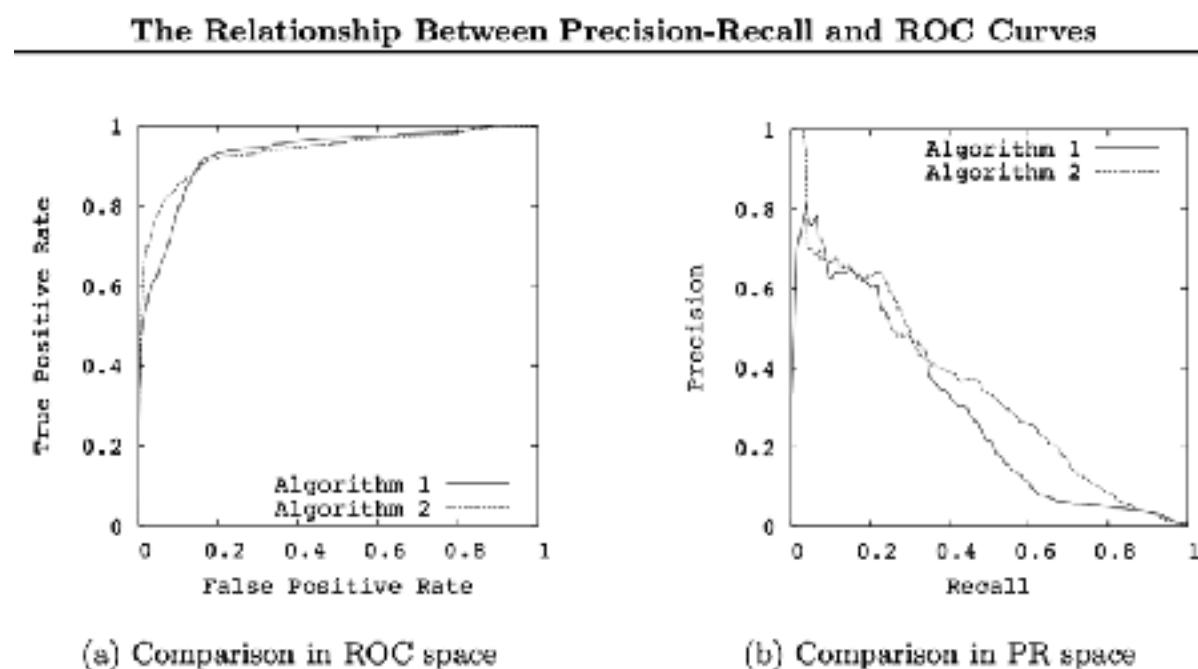


Figure 1. The difference between comparing algorithms in ROC vs PR space

<https://www.biostat.wisc.edu/~page/rocpr.pdf>

How the ground truth labels were obtained?

Noname manuscript No.
(will be inserted by the editor)

- Check if the annotation process makes sense
- Is it biased?

ImageNet Large Scale Visual Recognition Challenge

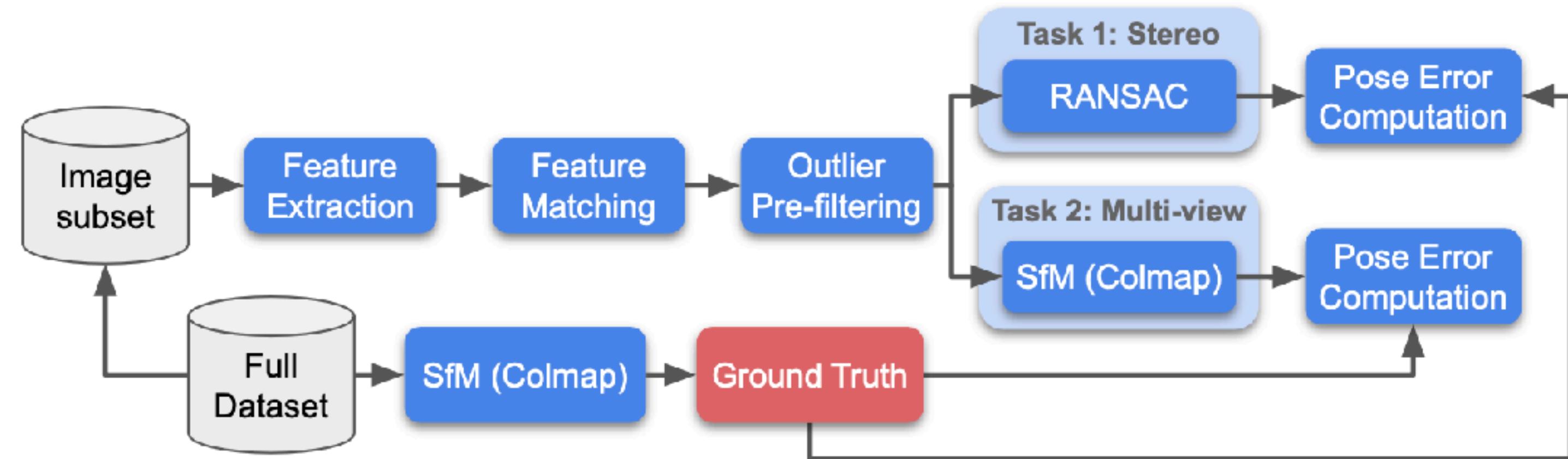
Olga Russakovsky* · Jia Deng* · Hao Su · Jonathan Krause ·
Sanjeev Satheesh · Sean Ma · Zhiheng Huang · Andrej Karpathy ·
Aditya Khosla · Michael Bernstein · Alexander C. Berg · Li Fei-Fei

1. **Drawing** A worker draws one bounding box around one instance of an object on the given image.
2. **Quality verification** A second worker checks if the bounding box is correctly drawn.
3. **Coverage verification** A third worker checks if all object instances have bounding boxes.

How the ground truth labels were obtained?

Case study: image matching

- Assumption: COLMAP reconstruction from the 1000 images outputs poses, which are good enough to evaluate stereo poses



How the ground truth labels were obtained?

Case study: image matching

- Assumption: COLMAP reconstruction from the 1000 images outputs poses, which are good enough to evaluate stereo poses
- Is it validated?

3.3 On the quality of the “ground-truth”

Our core assumption is that accurate poses can be obtained from large sets of images without human intervention. Such poses are used as the “ground truth” for evaluation of image matching performance on pairs or small subsets of images – a harder, proxy task. Should this assumption hold, the (relative) poses retrieved with a large enough number of images would not change as more images are added, and these poses would be the same regardless of which local feature is used. To validate this, we pick the scene “Sacre Coeur” and compute SfM reconstructions with a varying number of images: 100, 200, 400, 800, and 1179 images (the entire “Sacre Coeur” dataset), where each set contains the previous one; new images are being added and no images are removed. We run each reconstruction three times, and report the aver-

How the ground truth labels were obtained?

Case study: image matching

- Is assumption validated?

How the ground truth labels were obtained?

Case study: image matching

- Is assumption validated?

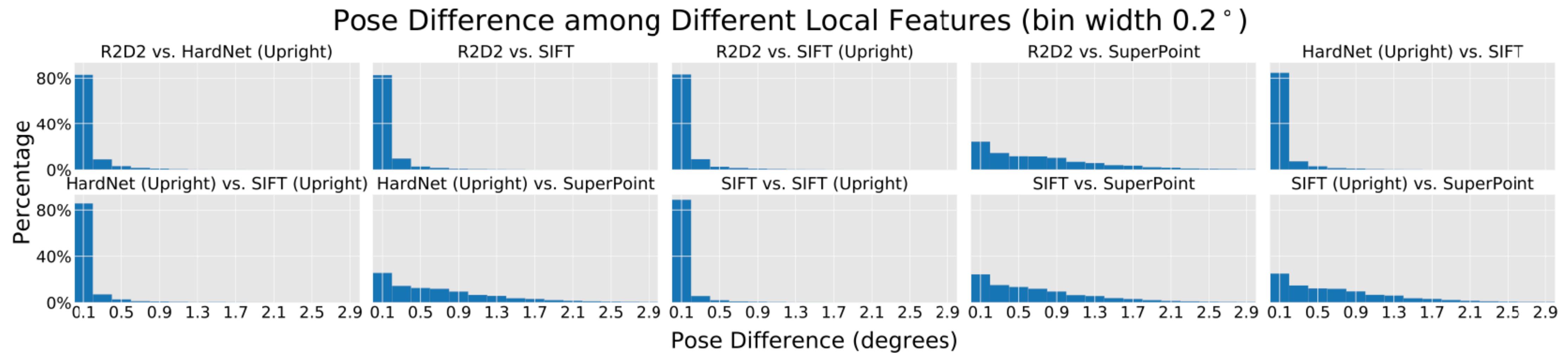


Fig. 7 Histograms of pose differences between reconstructions with different local feature methods. We consider five different local features – including rotation-sensitive and upright SIFT – resulting in 10 combinations. The plots show that about 80% percent of image pairs are within a 0.2° pose difference, with the exception of those involving SuperPoint.

How the ground truth labels were obtained?

Case study: image matching

- Is assumption validated?

How the ground truth labels were obtained?

Case study: image matching

- Is assumption validated?

Local featured type	Number of Images				
	100	200	400	800	all
SIFT [54]	0.06°	0.09°	0.06°	0.07°	0.09°
SIFT (Upright) [54]	0.07°	0.07°	0.04°	0.06°	0.09°
HardNet (Upright) [62]	0.06°	0.06°	0.06°	0.04°	0.05°
SuperPoint [34]	0.31°	0.25°	0.33°	0.19°	0.32°
R2D2 [80]	0.12°	0.08°	0.07°	0.08°	0.05°

Table 2 Standard deviation of the pose difference of three COLMAP runs with different number of images. Most of them are below 0.1° , except for SuperPoint.

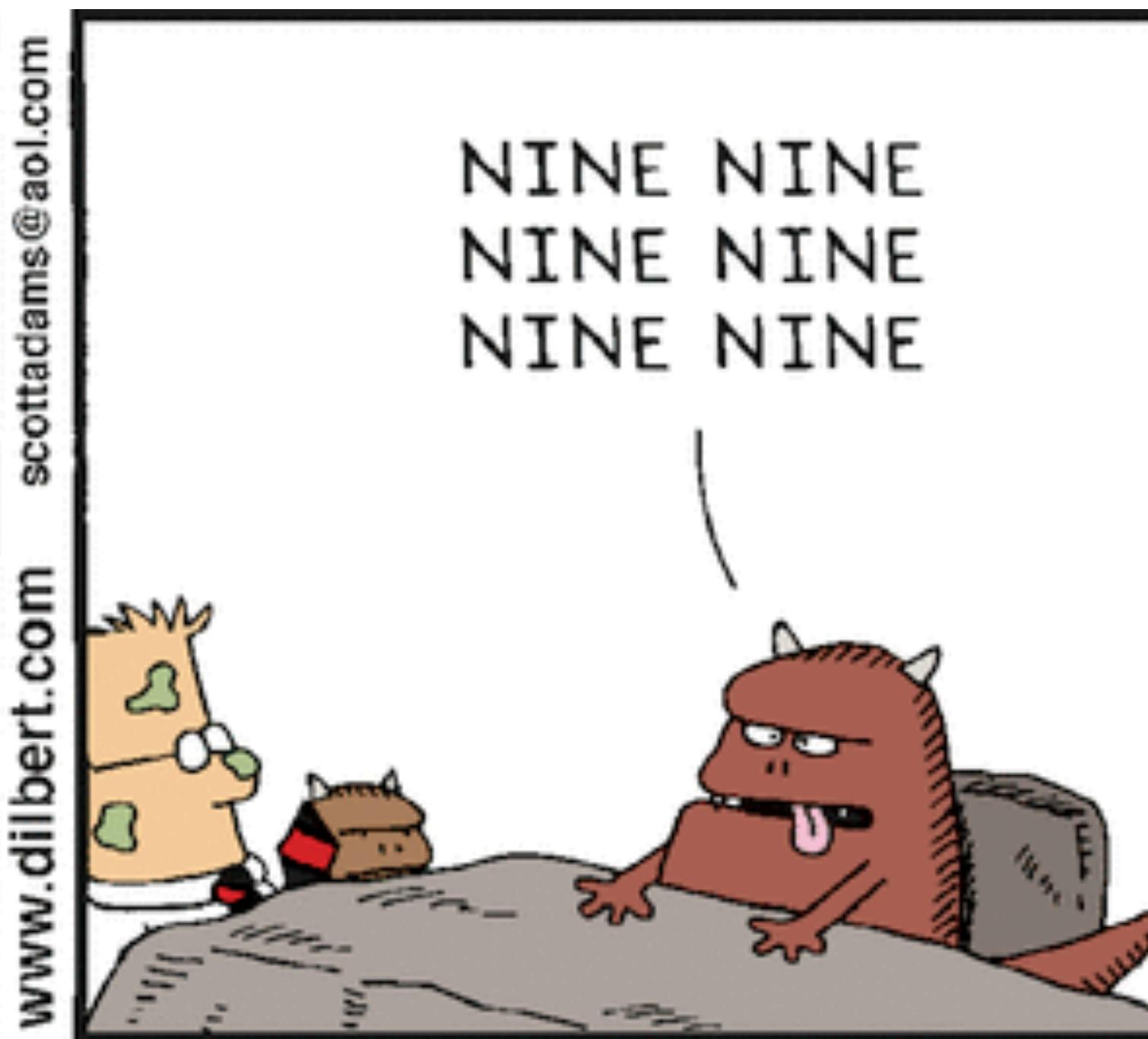
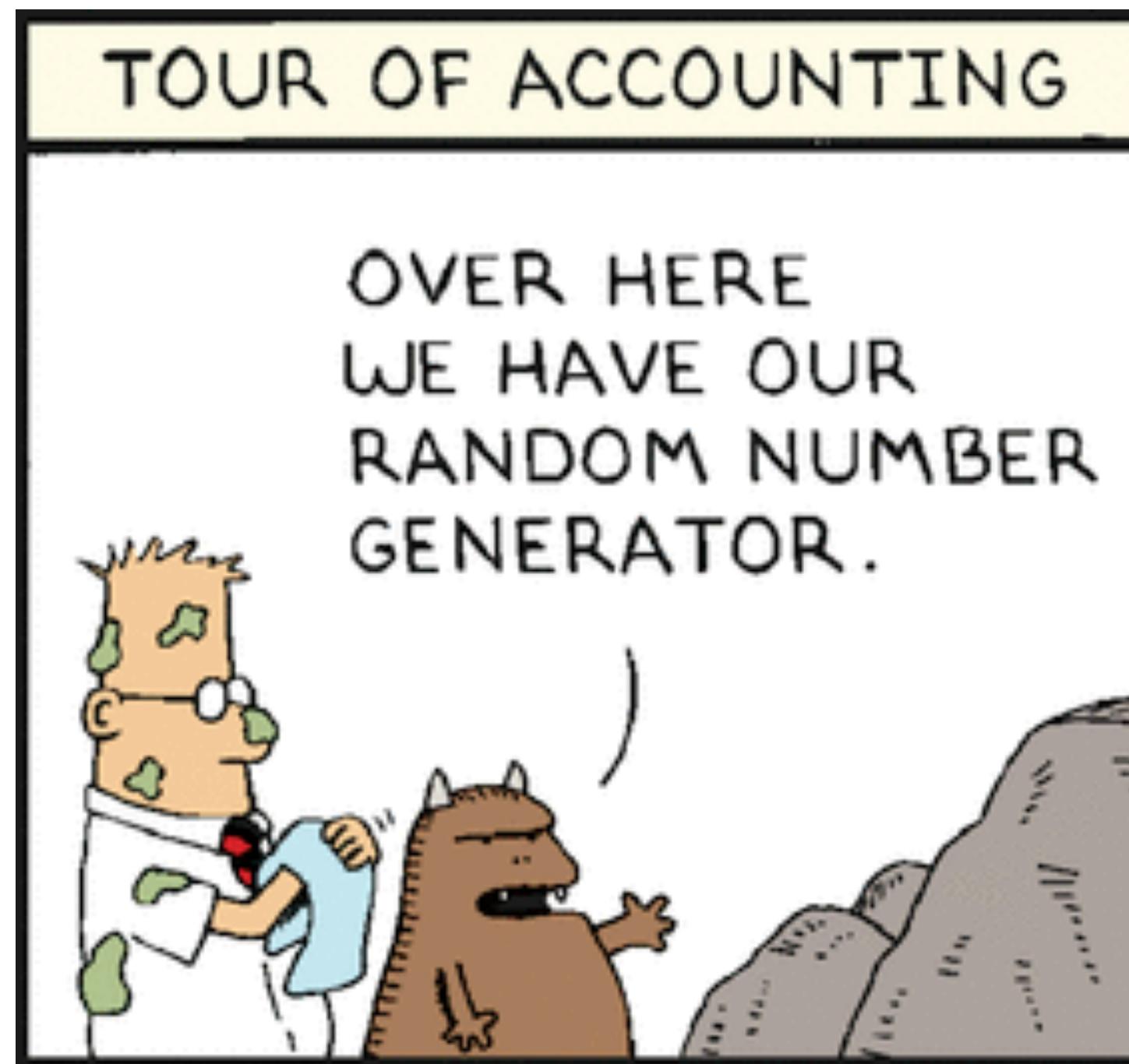
It seems that we can trust the paper...

It seems that we can trust the paper...

- Although we never can be sure

It seems that we can trust the paper...

- Although we never can be sure



**You should never be 100% sure
I am not joking**

You should never be 100% sure

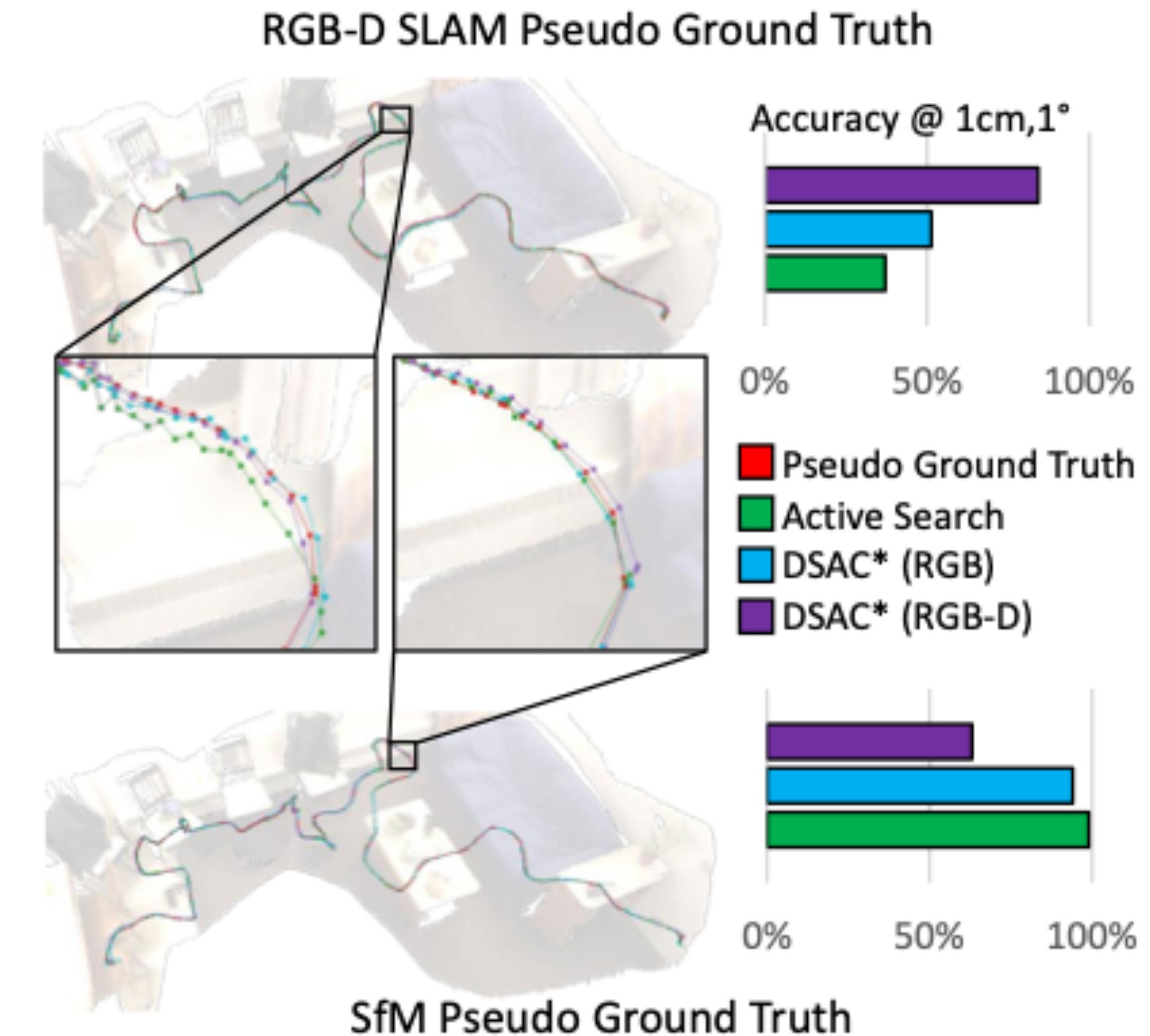
I am not joking

- On the Limits of Pseudo Ground Truth in Visual Camera Re-localisation. Brachmann et.al
ICCV 2021

You should never be 100% sure

I am not joking

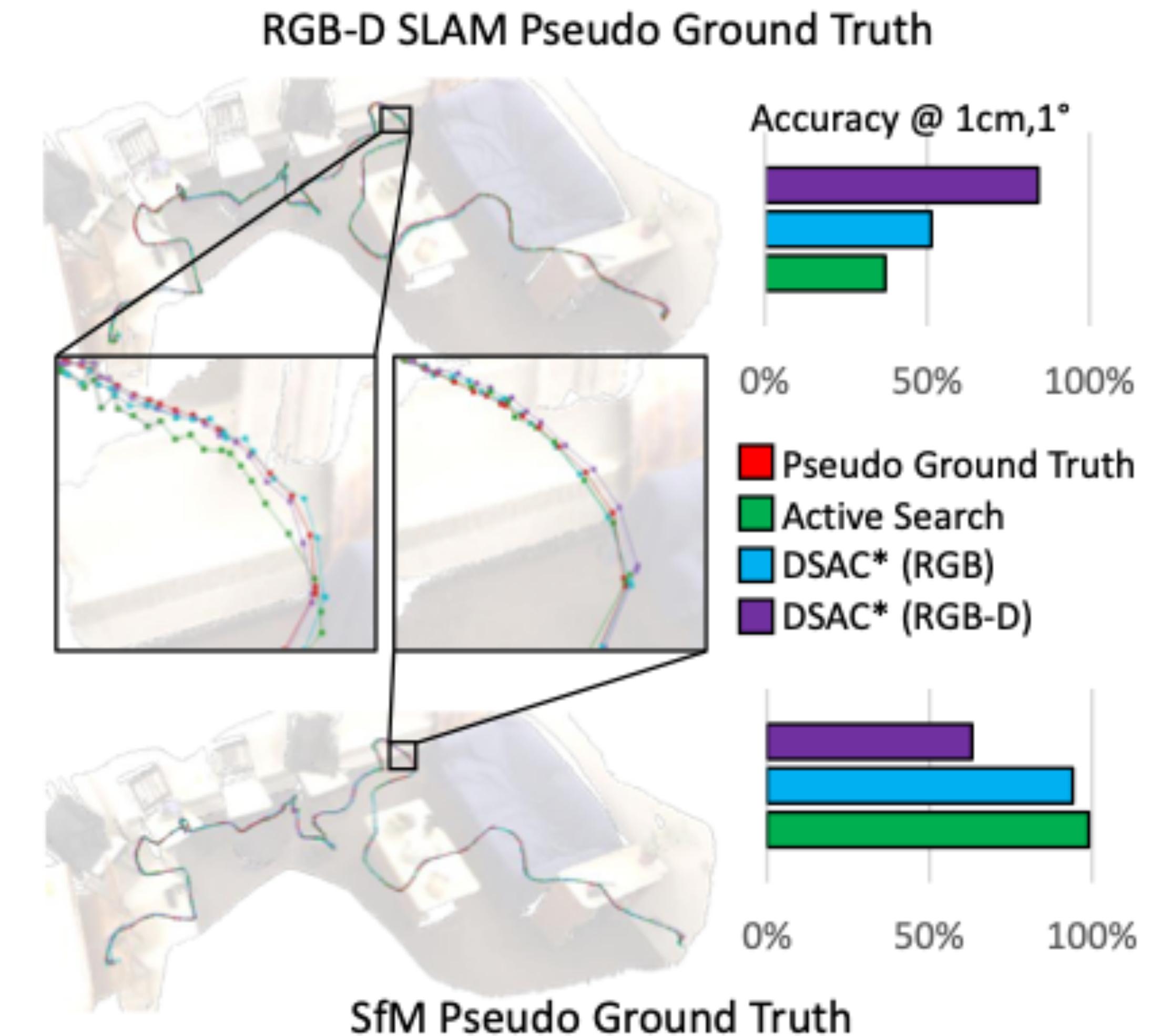
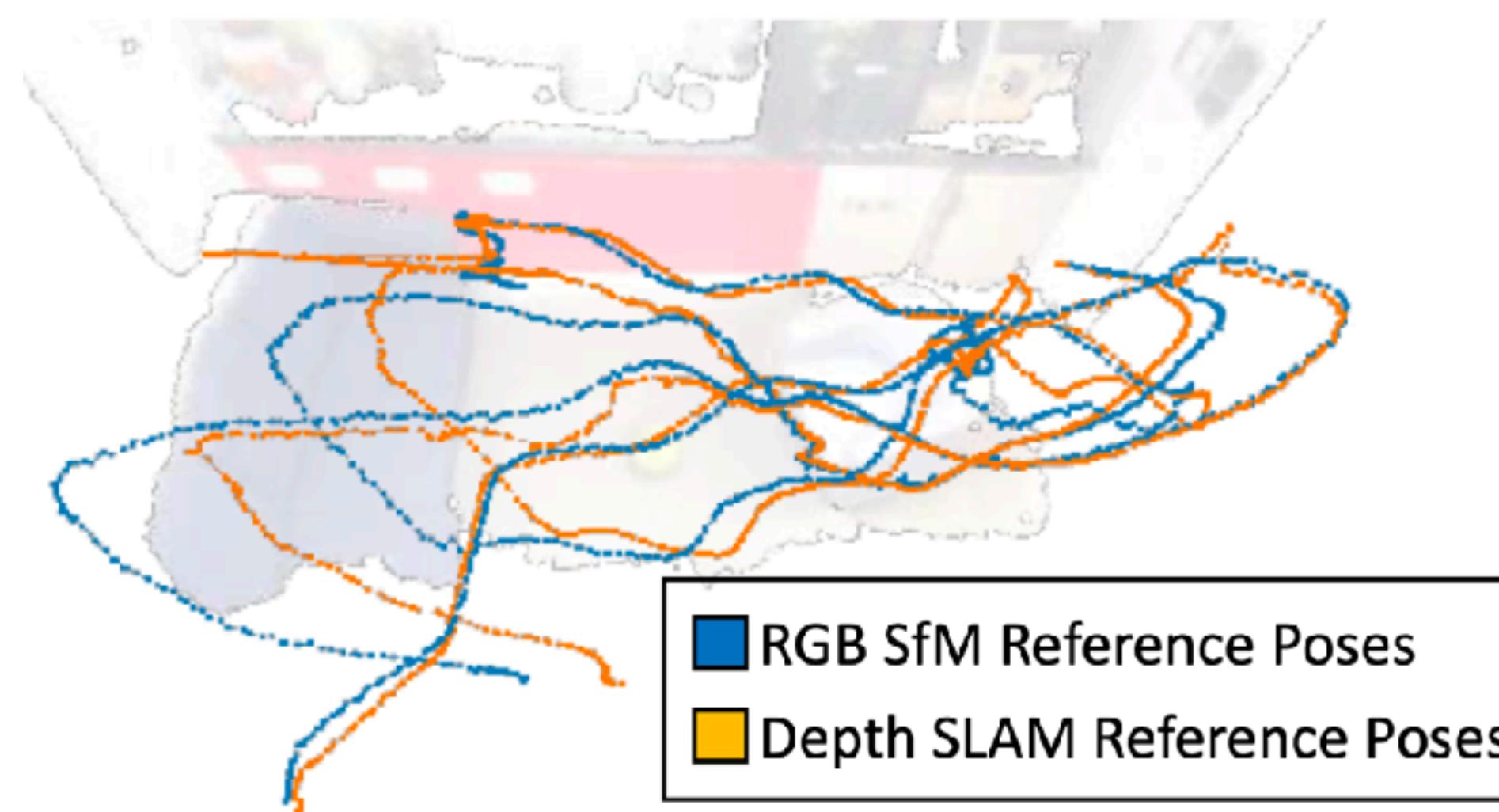
- On the Limits of Pseudo Ground Truth in Visual Camera Relocalisation. Brachmann et.al ICCV 2021



You should never be 100% sure

I am not joking

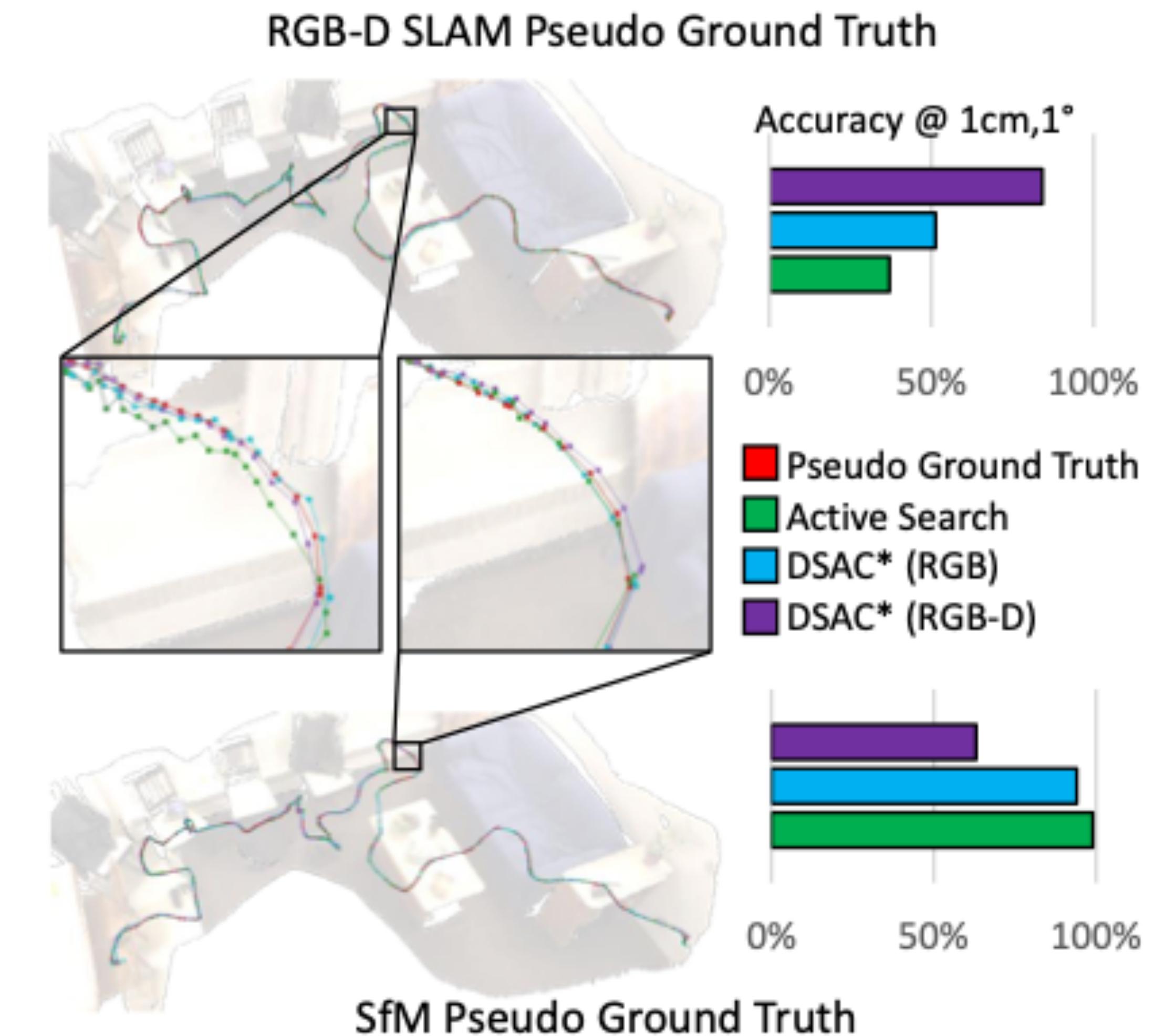
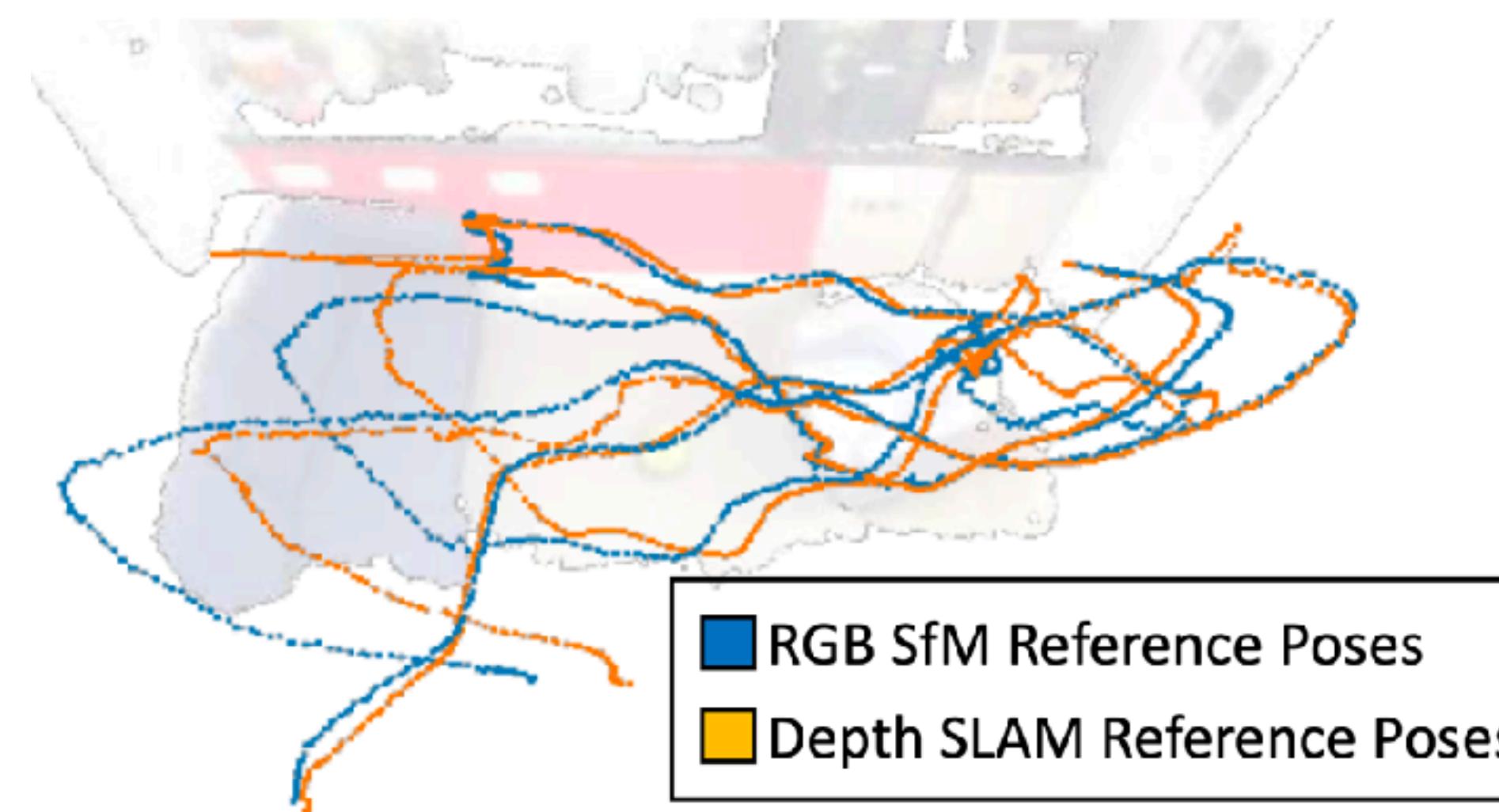
- On the Limits of Pseudo Ground Truth in Visual Camera Relocalisation. Brachmann et.al ICCV 2021



You should never be 100% sure

I am not joking

- On the Limits of Pseudo Ground Truth in Visual Camera Relocalisation. Brachmann et.al ICCV 2021



Paper is relevant and seems truthworthy

Paper is relevant and seems truthworthy



Paper is relevant and seems truthworthy



Checking results and methods

Start with baseline

Checking results and methods

Start with baseline

- Find the method, which is the established “gold standard” and/or the baseline

Checking results and methods

Start with baseline

- Find the method, which is the established “gold standard” and/or the baseline

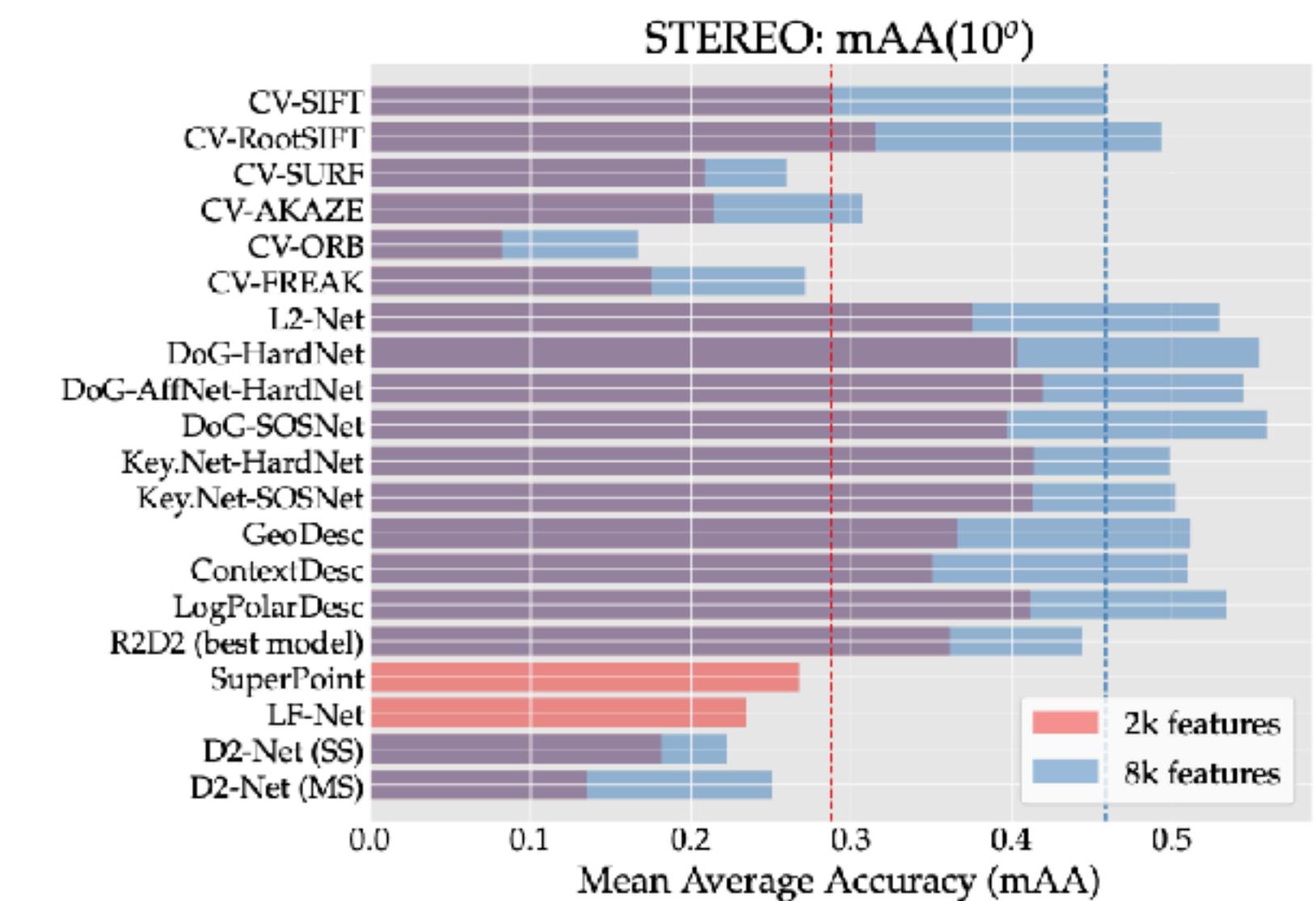


Fig. 17 Test – Stereo performance: 2k vs 8k features. We compare the results obtained with different methods using either 2k or 8k features – we use DEGENSAC, which performs better than other RANSAC variants under most circumstances. Dashed lines indicate SIFT’s performance. For LF-Net and SuperPoint we do not include results with 8k features, as we failed to obtain meaningful results. For R2D2, we use the best model for each setting.

Checking results and methods

Start with baseline

- Find the method, which is the established “gold standard” and/or the baseline
- For image matching it is SIFT

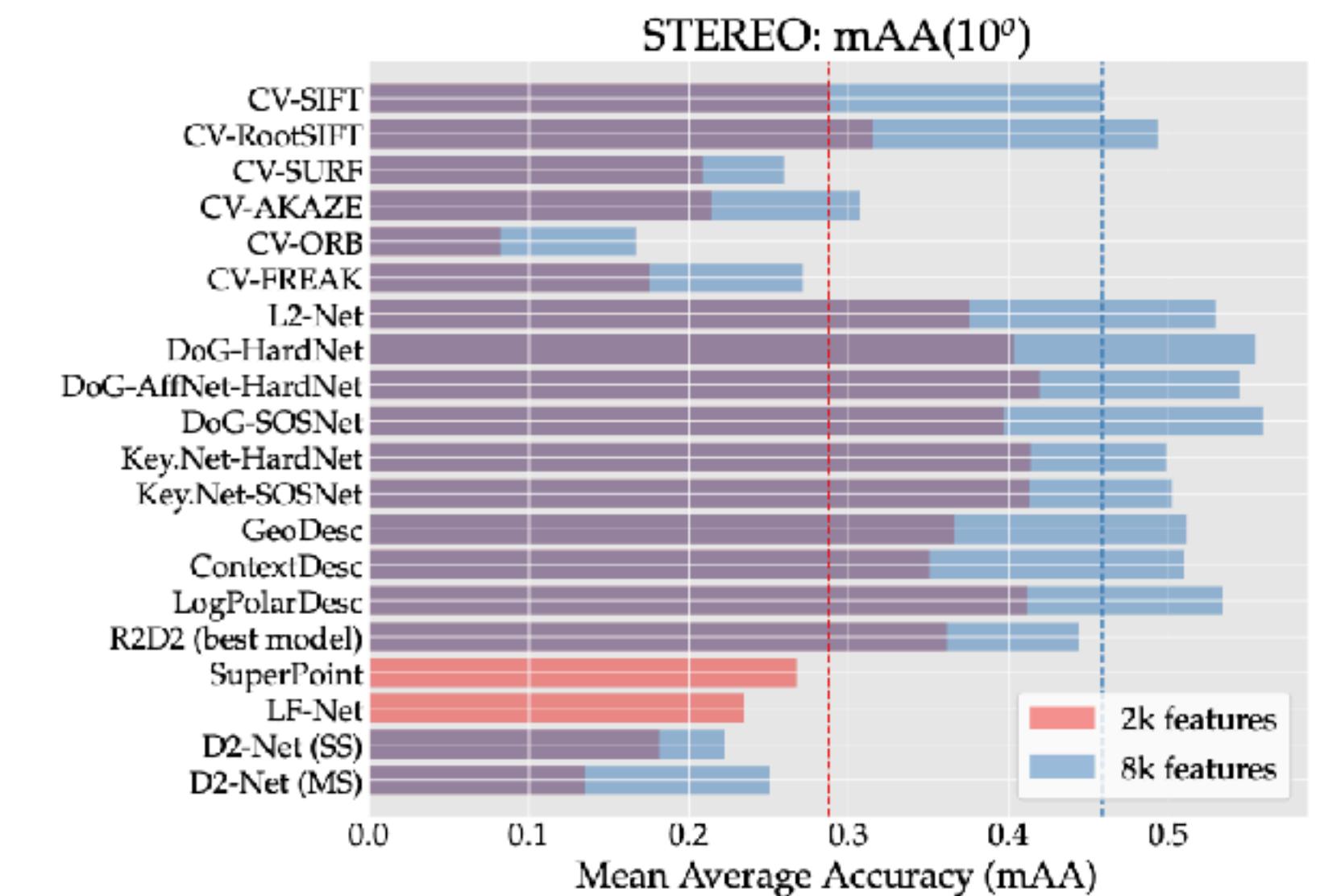


Fig. 17 Test – Stereo performance: 2k vs 8k features. We compare the results obtained with different methods using either 2k or 8k features – we use DEGENSAC, which performs better than other RANSAC variants under most circumstances. Dashed lines indicate SIFT’s performance. For LF-Net and SuperPoint we do not include results with 8k features, as we failed to obtain meaningful results. For R2D2, we use the best model for each setting.

Checking results and methods

Start with baseline

- Find the method, which is the established “gold standard” and/or the baseline
- For image matching it is SIFT
- Check how it works

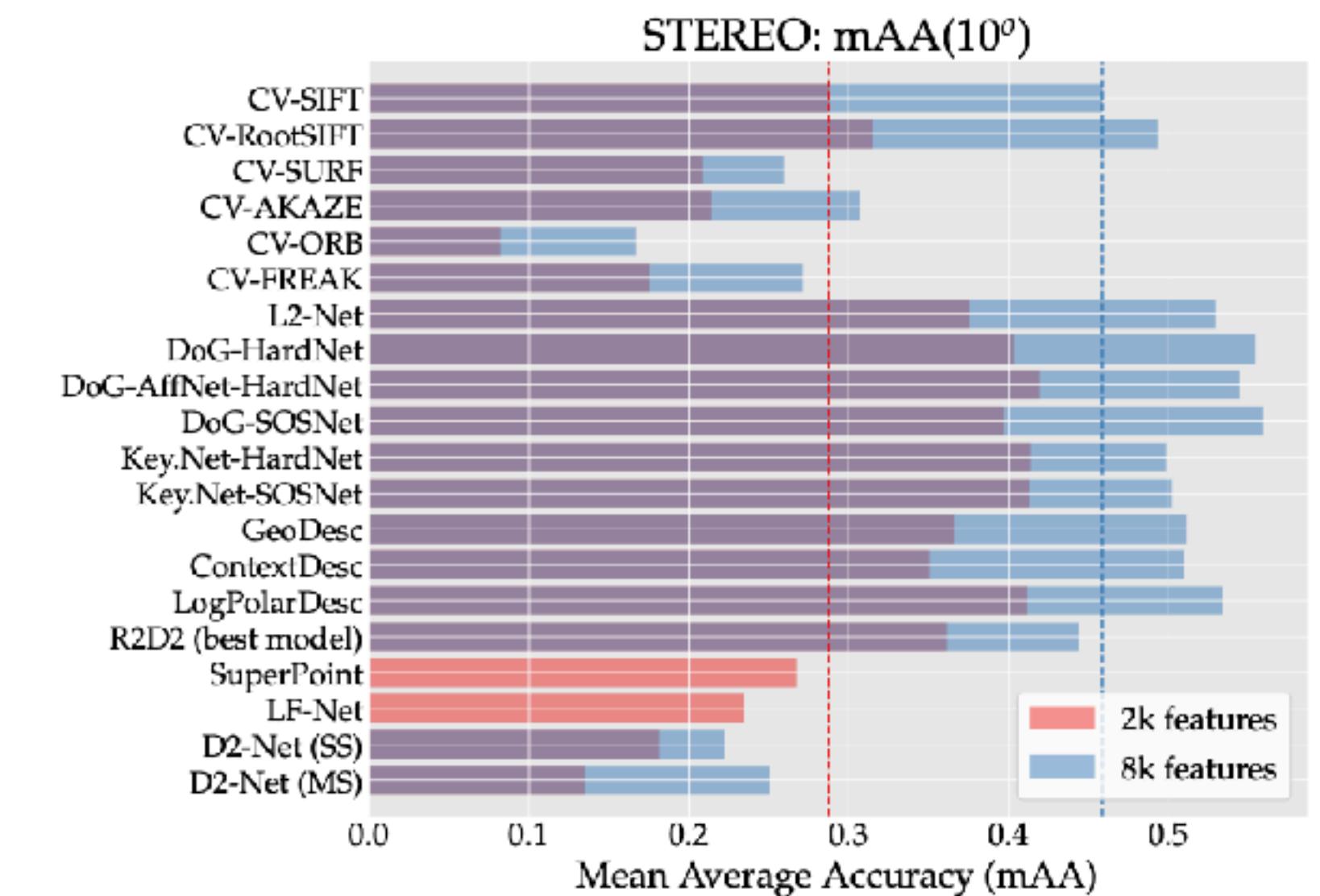


Fig. 17 Test – Stereo performance: 2k vs 8k features. We compare the results obtained with different methods using either 2k or 8k features – we use DEGENSAC, which performs better than other RANSAC variants under most circumstances. Dashed lines indicate SIFT’s performance. For LF-Net and SuperPoint we do not include results with 8k features, as we failed to obtain meaningful results. For R2D2, we use the best model for each setting.

Checking results and methods

Start with baseline

- Find the method, which is the established “gold standard” and/or the baseline
- For image matching it is SIFT
- Check how it works
 - Note, that the “gold standard” might be complex and highly engineered

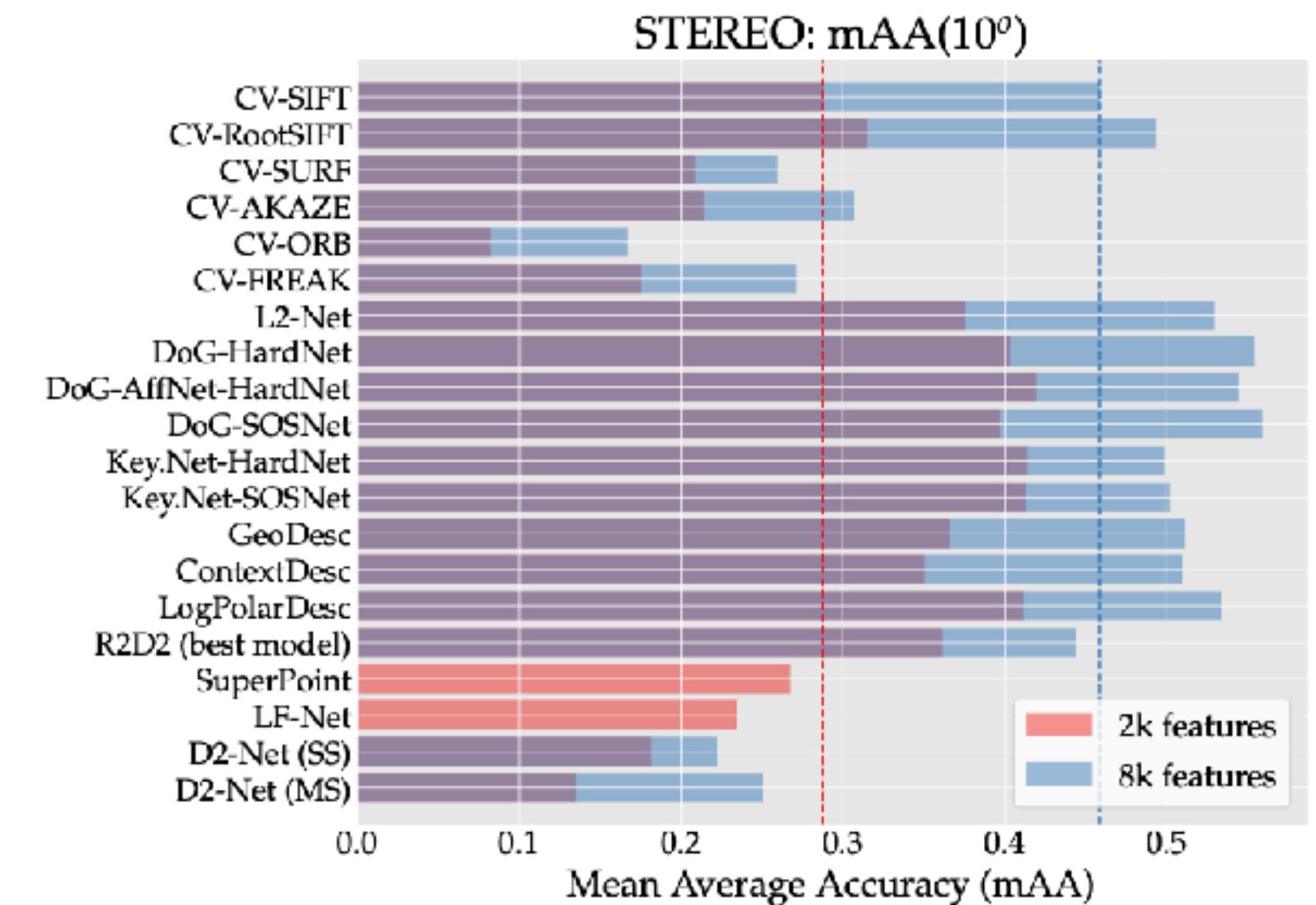


Fig. 17 Test – Stereo performance: 2k vs 8k features. We compare the results obtained with different methods using either 2k or 8k features – we use DEGENSAC, which performs better than other RANSAC variants under most circumstances. Dashed lines indicate SIFT’s performance. For LF-Net and SuperPoint we do not include results with 8k features, as we failed to obtain meaningful results. For R2D2, we use the best model for each setting.

Checking results and methods

Start with baseline

- Find the method, which is the established “gold standard” and/or the baseline
- For image matching it is SIFT
- Check how it works
 - Note, that the “gold standard” might be complex and highly engineered
 - New methods might be simpler to implement

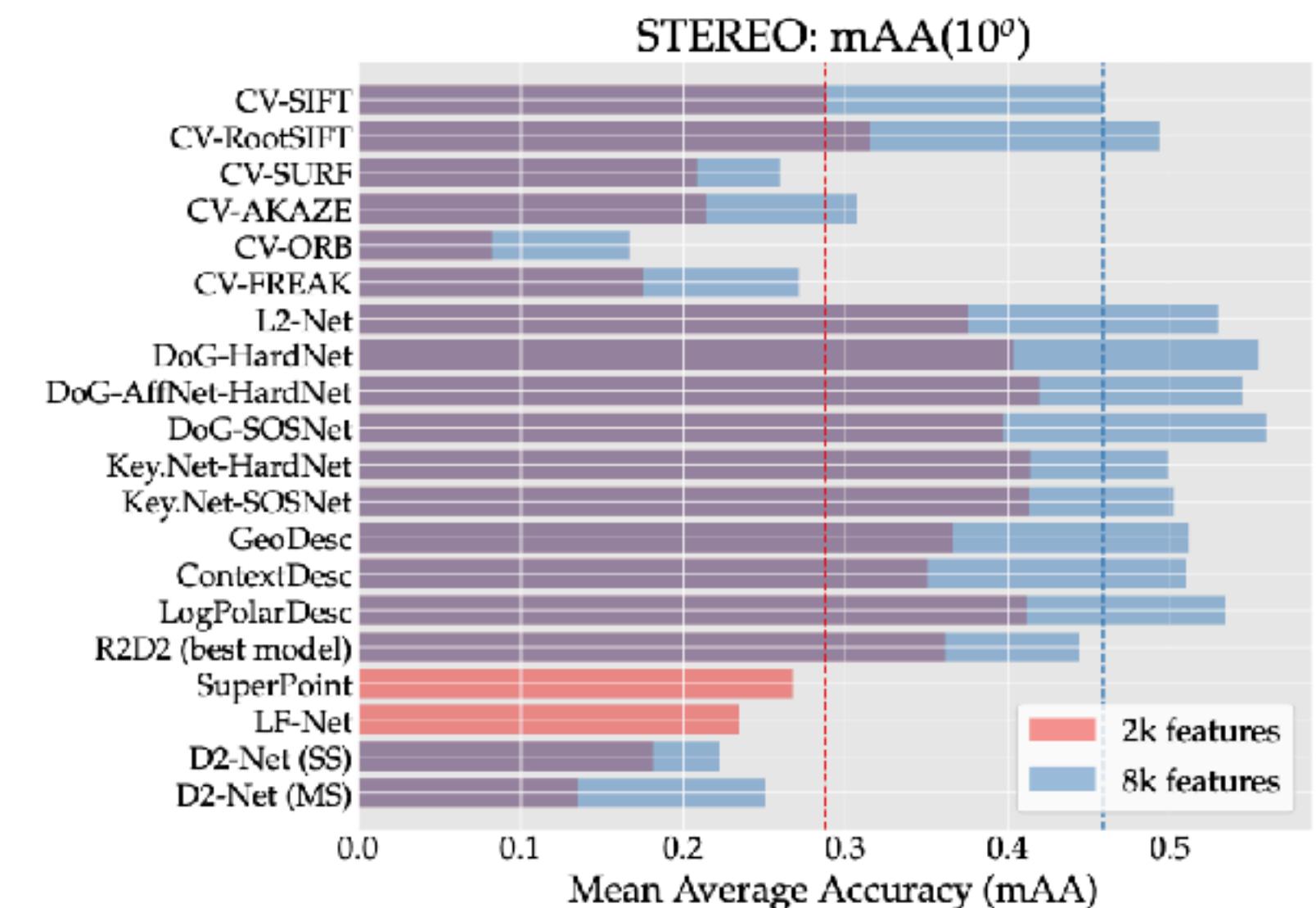


Fig. 17 Test – Stereo performance: 2k vs 8k features. We compare the results obtained with different methods using either 2k or 8k features – we use DEGENSAC, which performs better than other RANSAC variants under most circumstances. Dashed lines indicate SIFT’s performance. For LF-Net and SuperPoint we do not include results with 8k features, as we failed to obtain meaningful results. For R2D2, we use the best model for each setting.

Checking results and methods

Start with baseline

- Find the method, which is the established “gold standard” and/or the baseline
- For image matching it is SIFT
- Check how it works
 - Note, that the “gold standard” might be complex and highly engineered
 - New methods might be simpler to implement
- Play selected (good and simple) methods

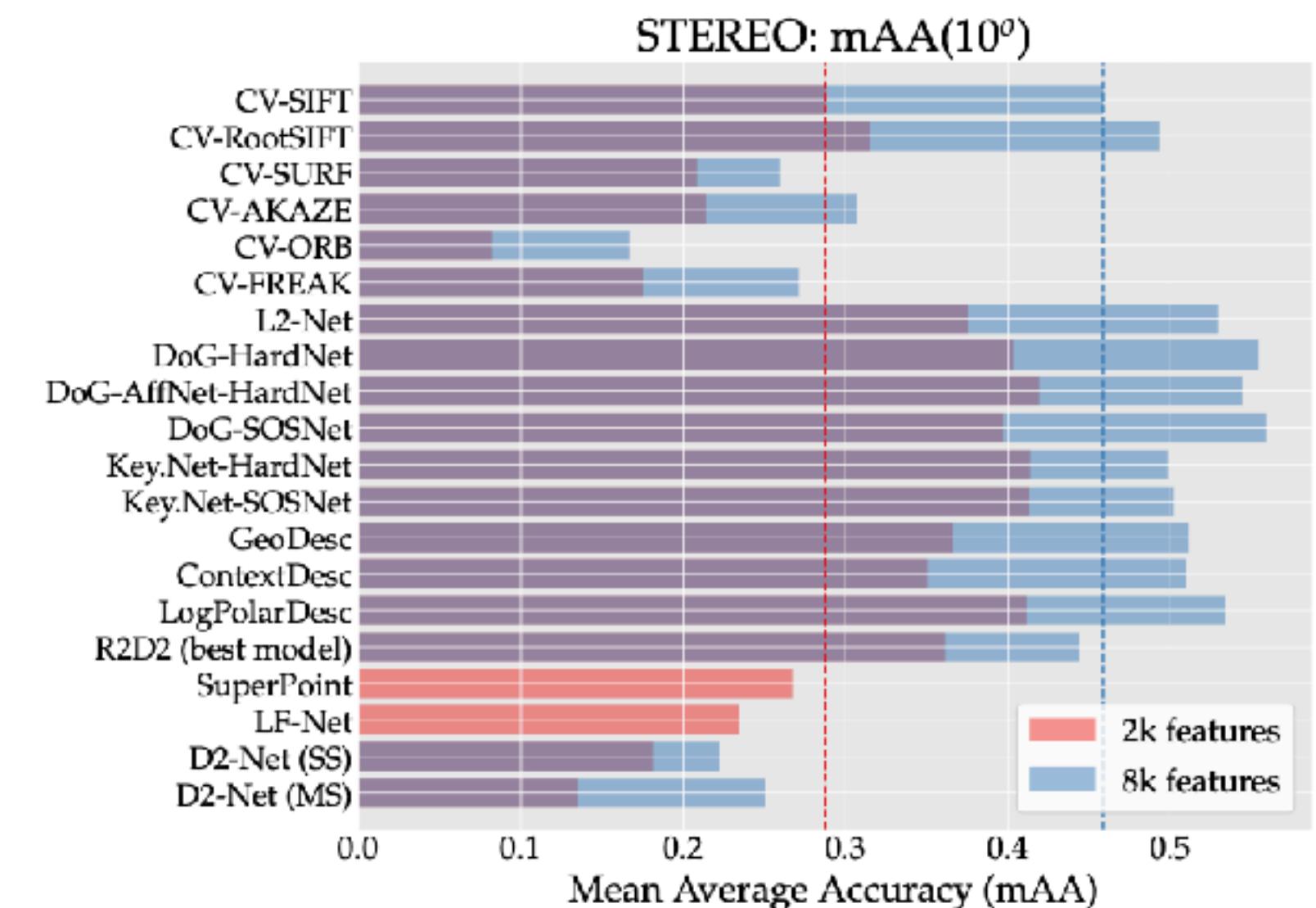


Fig. 17 Test – Stereo performance: 2k vs 8k features. We compare the results obtained with different methods using either 2k or 8k features – we use DEGENSAC, which performs better than other RANSAC variants under most circumstances. Dashed lines indicate SIFT’s performance. For LF-Net and SuperPoint we do not include results with 8k features, as we failed to obtain meaningful results. For R2D2, we use the best model for each setting.

Checking results and methods

Start with baseline

- Find the method, which is the established “gold standard” and/or the baseline
- For image matching it is SIFT
- Check how it works
 - Note, that the “gold standard” might be complex and highly engineered
 - New methods might be simpler to implement
- Play selected (good and simple) methods
 - Maybe add their implementation to your ToDo

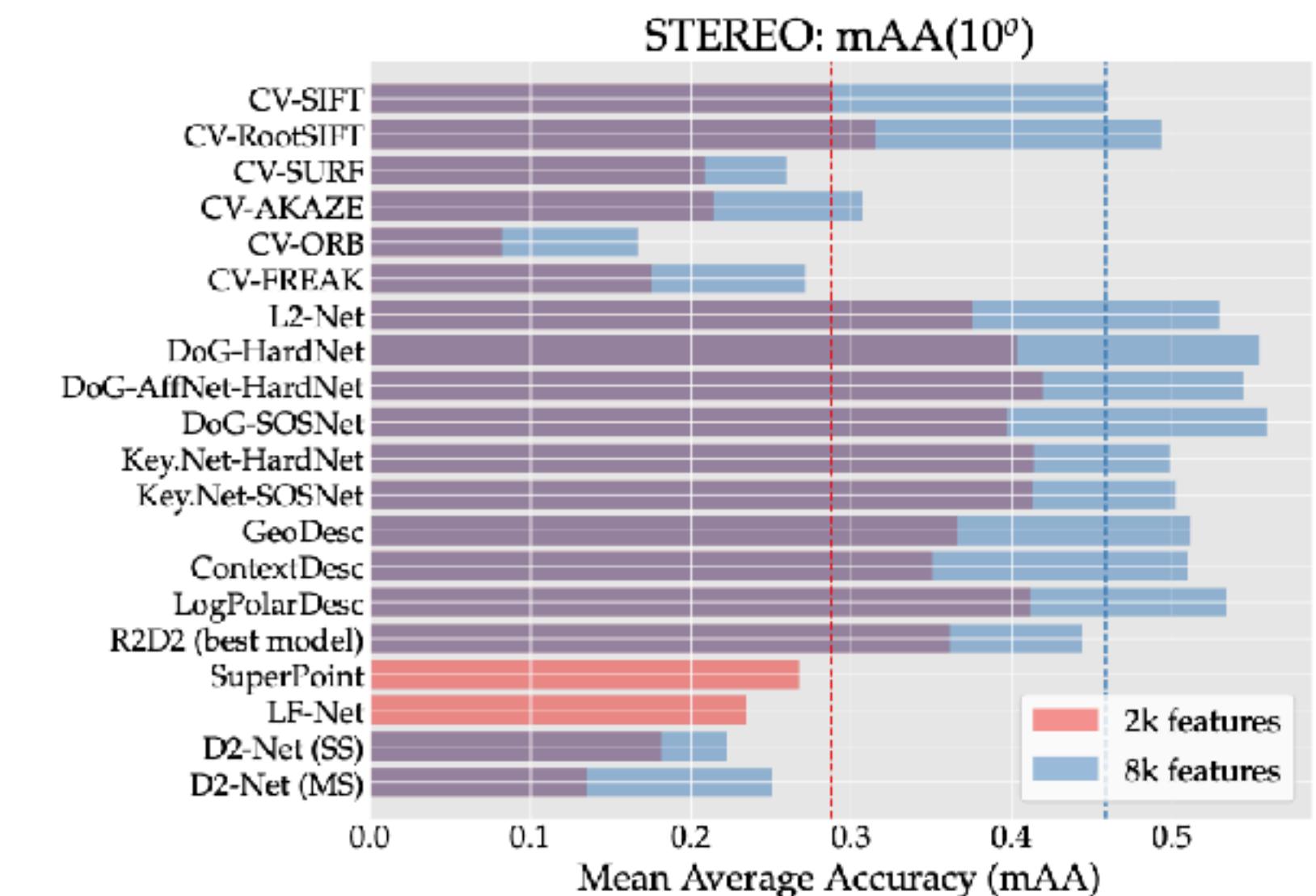


Fig. 17 Test – Stereo performance: 2k vs 8k features. We compare the results obtained with different methods using either 2k or 8k features – we use DEGENSAC, which performs better than other RANSAC variants under most circumstances. Dashed lines indicate SIFT’s performance. For LF-Net and SuperPoint we do not include results with 8k features, as we failed to obtain meaningful results. For R2D2, we use the best model for each setting.

How to “play with method”?

How to “play with method”?

- First, try to make it work

How to “play with method”?

- First, try to make it work
 - Visualize, visualize, visualize

How to “play with method”?

- First, try to make it work
 - Visualize, visualize, visualize
- Once it works, try to break it. Noise, transforms, data, etc.

How to “play with method”?

- First, try to make it work
 - Visualize, visualize, visualize
- Once it works, try to break it. Noise, transforms, data, etc.
 - that it is how understand its limits.

How to “play with method”?

- First, try to make it work
 - Visualize, visualize, visualize
- Once it works, try to break it. Noise, transforms, data, etc.
 - that it is how understand its limits.
- Try to understand the ideas it is based on

How to “play with method”?

- First, try to make it work
 - Visualize, visualize, visualize
- Once it works, try to break it. Noise, transforms, data, etc.
 - that it is how understand its limits.
- Try to understand the ideas it is based on
- Only then try to implement it on your own

Playing with method

Case study

Playing with method

Case study

- I was working on some image-matching method with co-authors

Playing with method

Case study

- I was working on some image-matching method with co-authors
- They sent me the training code, I had to plug it into the benchmark

Playing with method

Case study

- I was working on some image-matching method with co-authors
- They sent me the training code, I had to plug it into the benchmark
- Instead I first wrote simple script to visualize keypoints

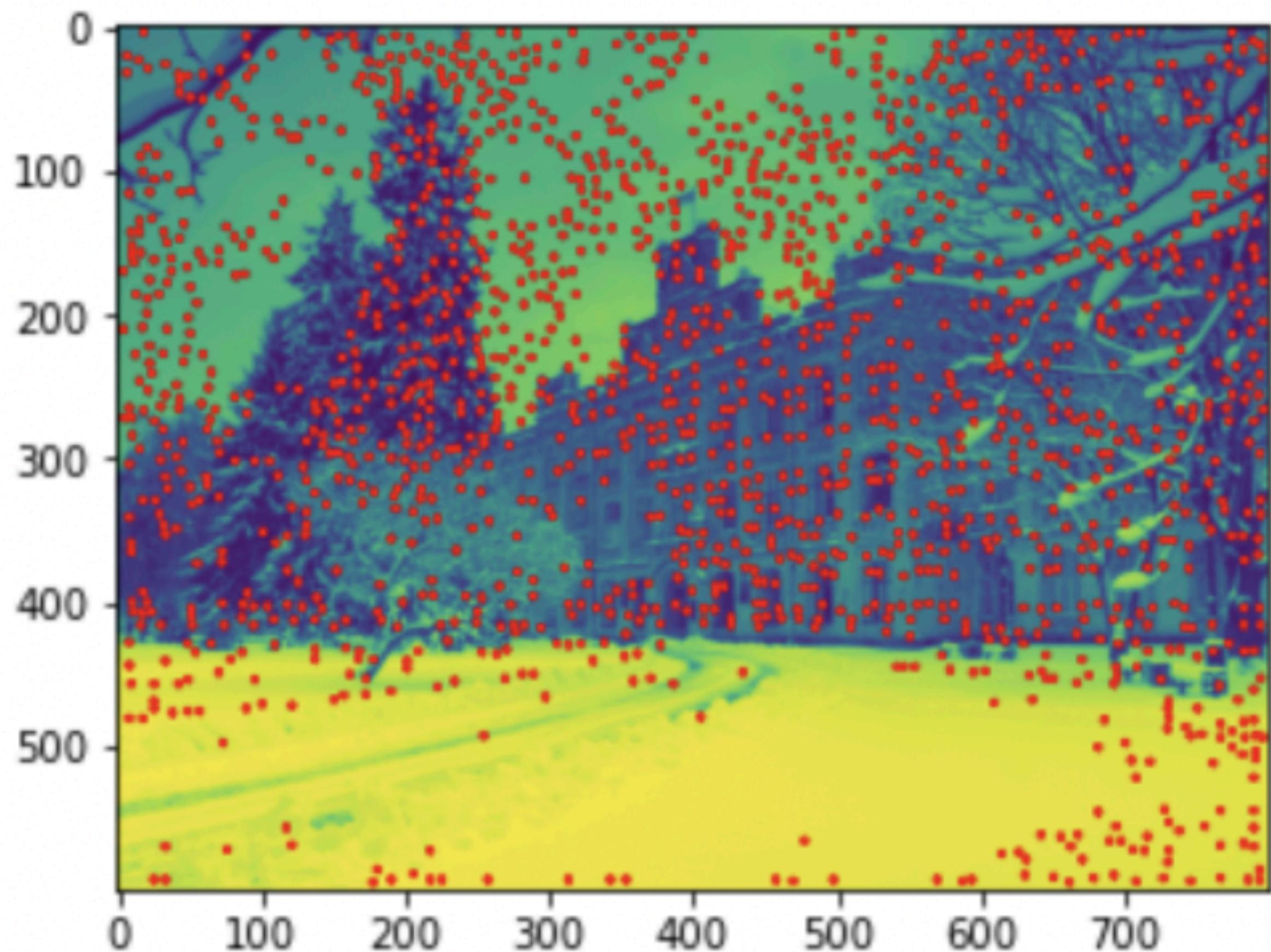
Playing with method

Case study

Playing with method

Case study

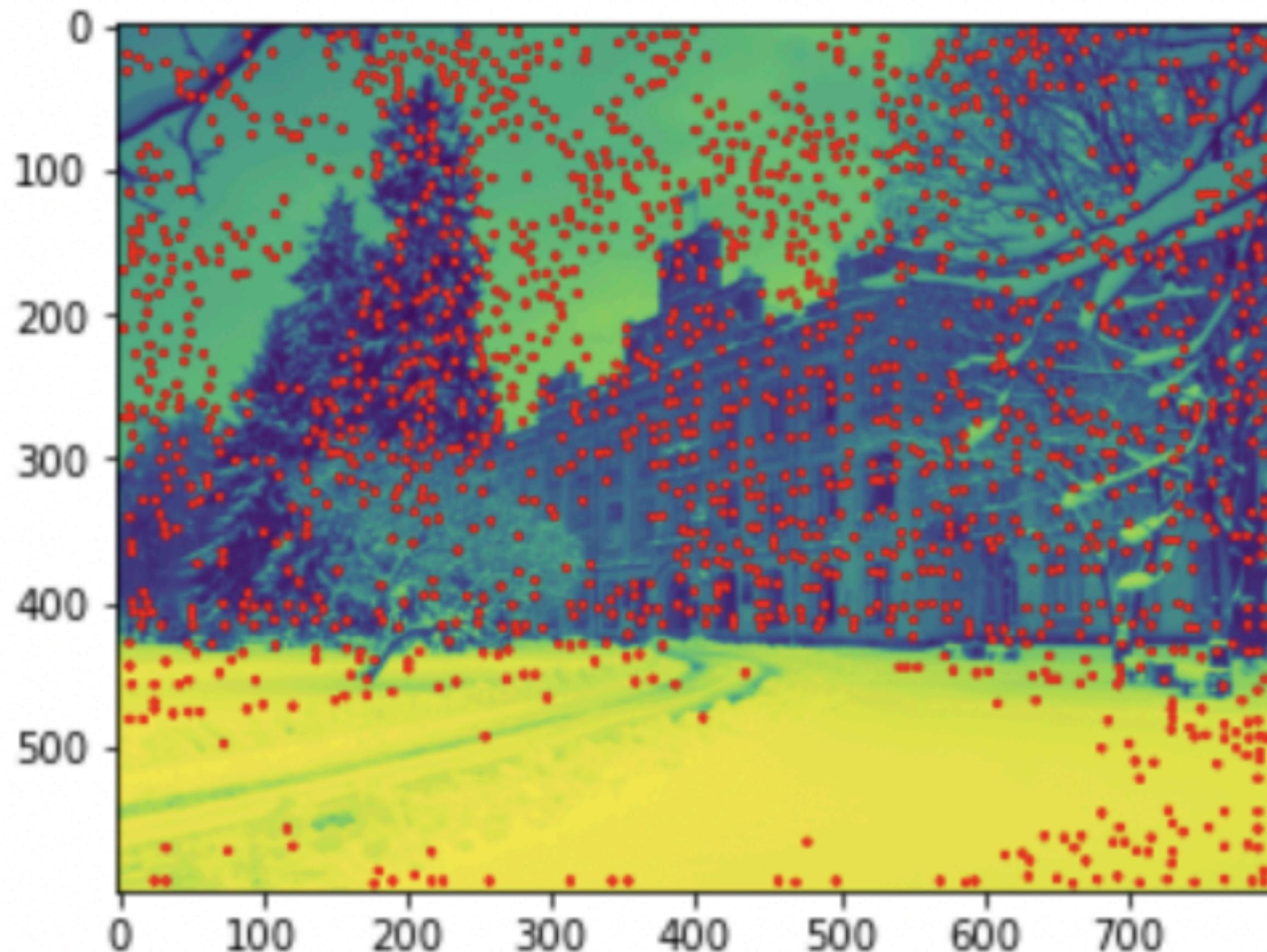
Model 1



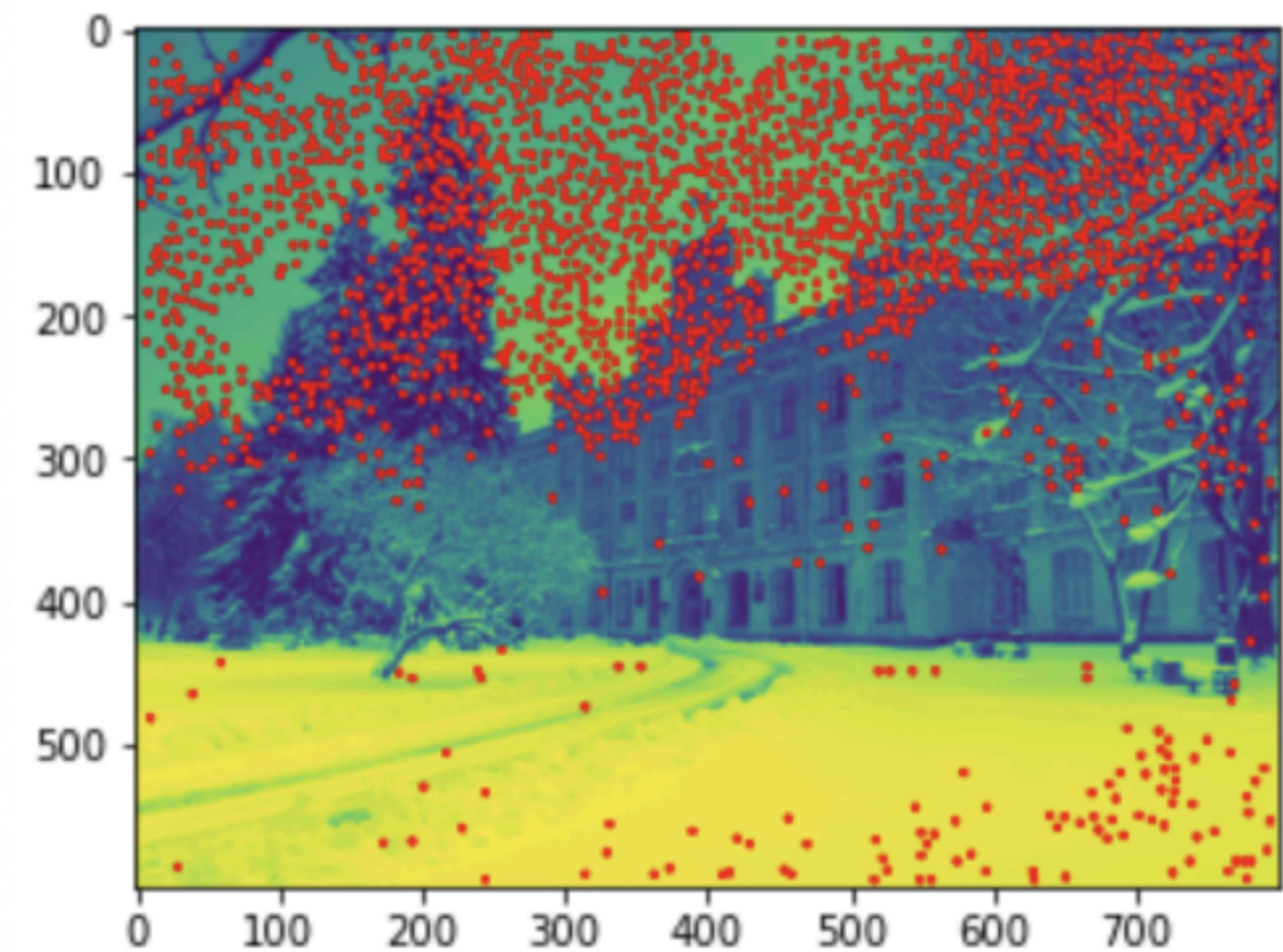
Playing with method

Case study

Model 1



Model 2

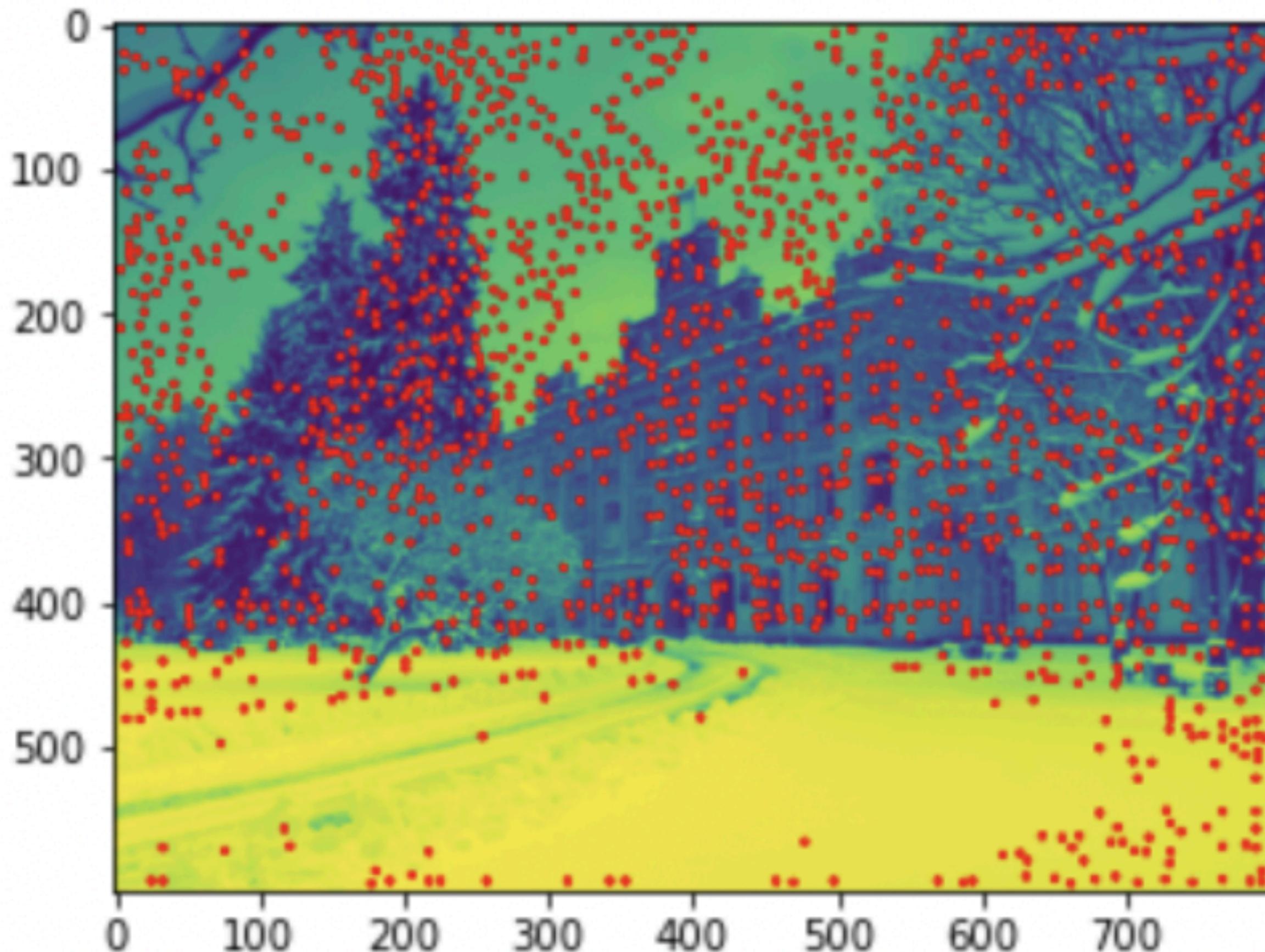


Playing with method

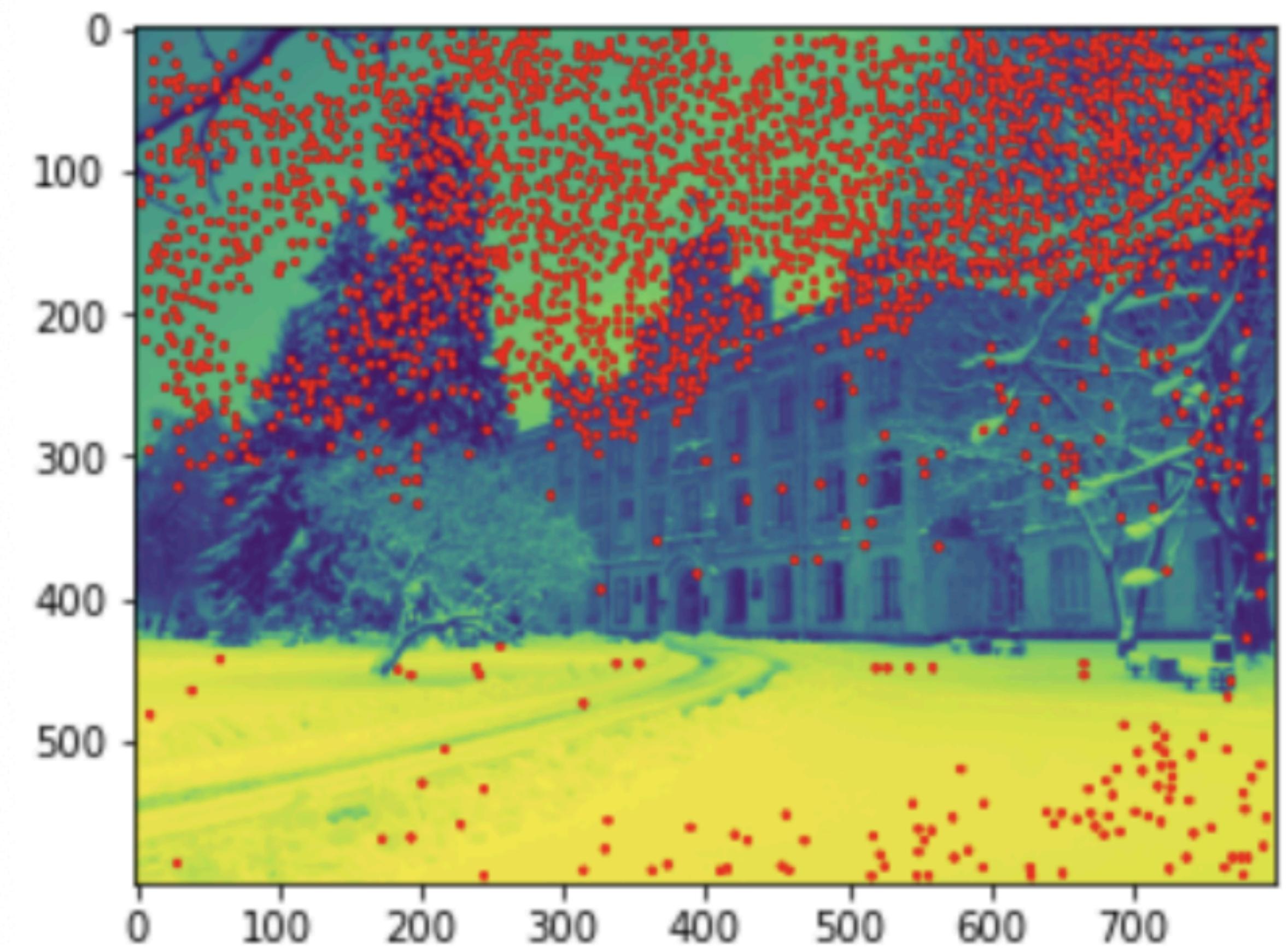
Case study

They were expected to look similar....

Model 1



Model 2

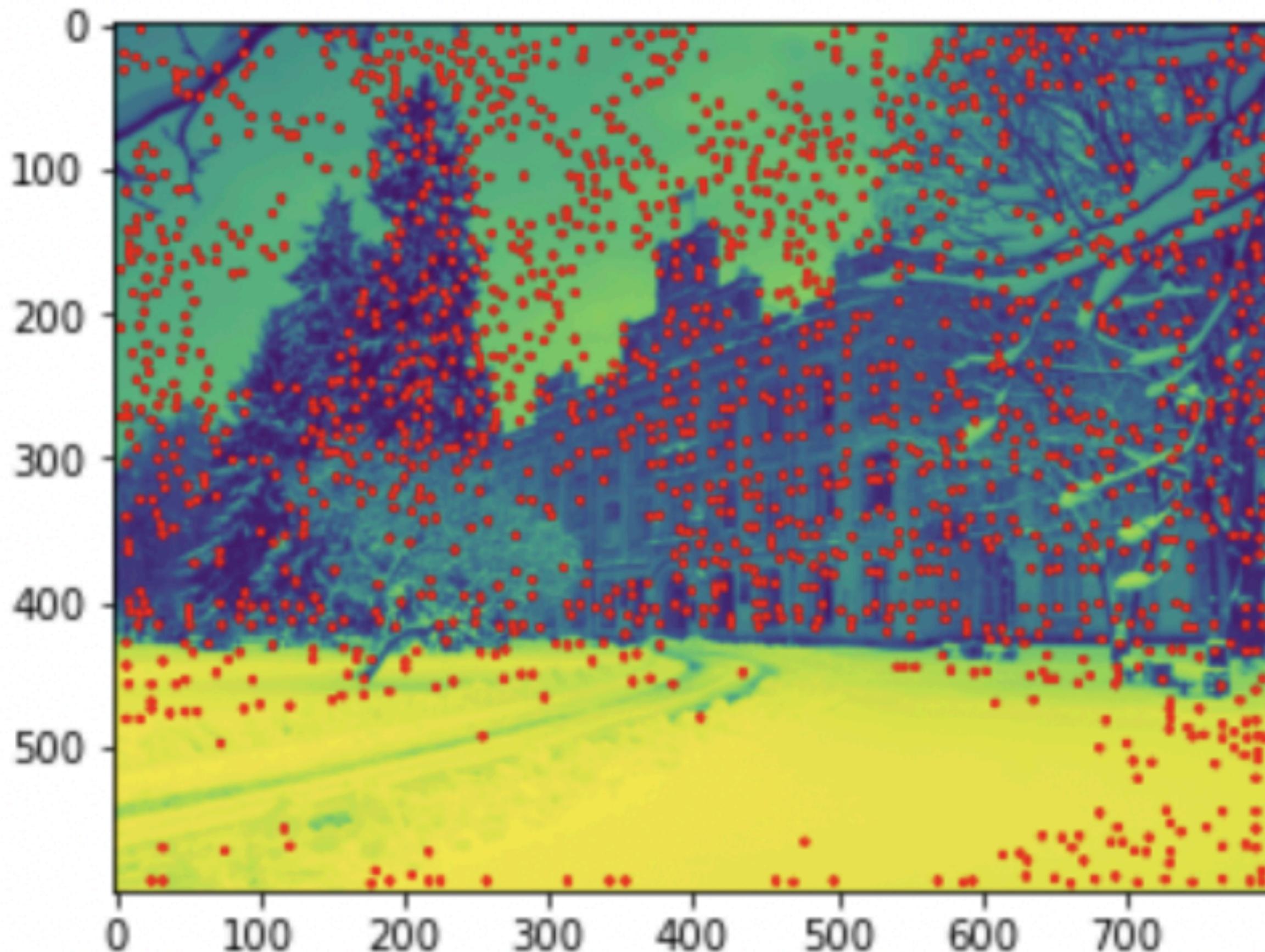


Playing with method

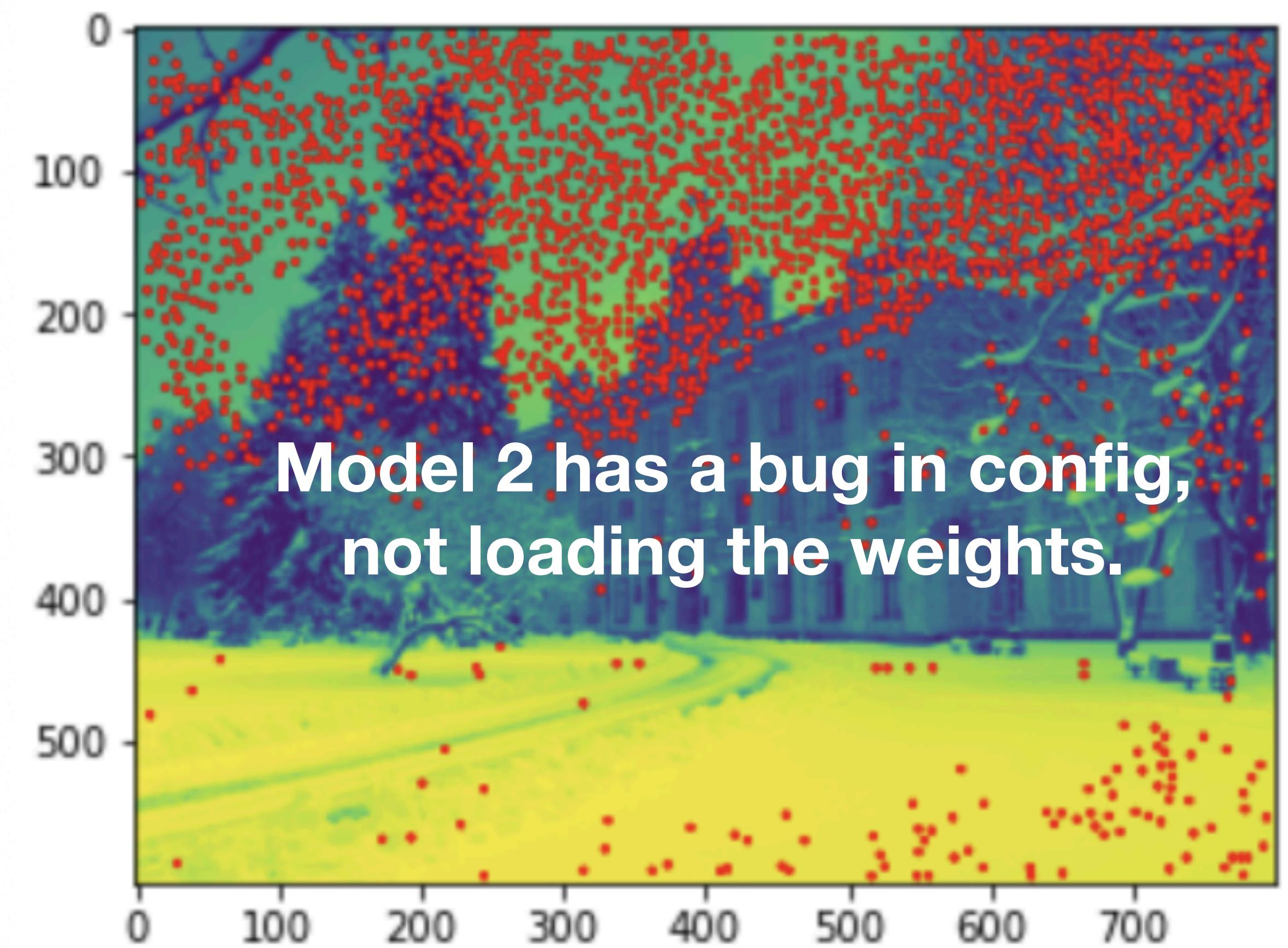
Case study

They were expected to look similar....

Model 1



Model 2



Playing with method

Case study

Playing with method

Case study

- If it is possible, gather some data yourself, not from benchmarks

Playing with method

Case study

- If it is possible, gather some data yourself, not from benchmarks
- Make it easy and fun to run the method on your data

Playing with method

Case study

- If it is possible, gather some data yourself, not from benchmarks
- Make it easy and fun to run the method on your data
- You can even build web demo for it

Playing with method

Case study

- If it is possible, gather some data yourself, not from benchmarks
- Make it easy and fun to run the method on your data
- You can even build web demo for it
- If it is not fun, it is not “playing”

Playing with method

Case study

- If it is possible, gather some data yourself, not from benchmarks
- Make it easy and fun to run the method on your data
- You can even build web demo for it
- If it is not fun, it is not “playing”
- You can think of this as “micro-benchmark”

Playing with method



Dmytro Mishkin
@ducha_aiki

R2D2 fails here, others are fine

9/

[Перекласти твіт](#)



3:47 пп · 22 лист. 2020 · Twitter Web App

https://twitter.com/ducha_aiki/status/1330495426865344515

...



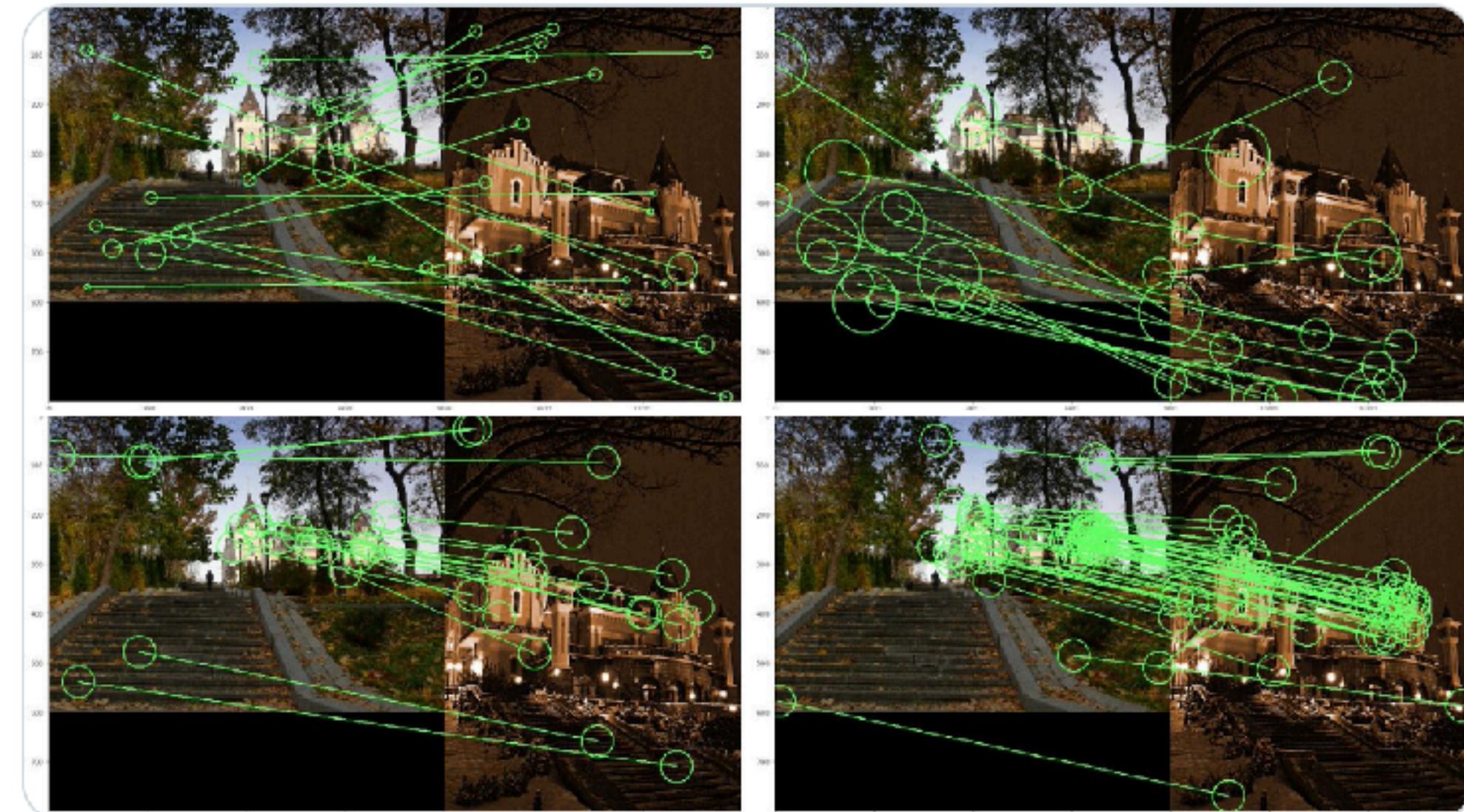
Dmytro Mishkin
@ducha_aiki

SIFT-HardNet and R2D2 fail here, SuperPoint rocks.

10/10

Overall, for random unknown upright pair of images with possible illumination change, I would go for SuperPoint, as least if we don't consider affine view synthesis.

[Перекласти твіт](#)



4:00 пп · 22 лист. 2020 · Twitter Web App

**You know area well enough
and want to keep yourself up-to-date**

You know area well enough and want to keep yourself up-to-date

- You want to maximize useful information flow

You know area well enough and want to keep yourself up-to-date

- You want to maximize useful information flow
- And minimize brain fuel & time spent

Where to look for papers on arXiv

arXiv itself?

- arxiv.org

Where to look for papers on arXiv arXiv itself?

- arxiv.org

The screenshot shows the arXiv.org homepage. At the top, there's a dark header with the Cornell University logo and a note about Simons Foundation support. Below this is a red navigation bar with the arXiv.org logo and search fields for "Search...", "All fields", and "Search". A "Login" button is also present. The main content area features a "COVID-19 Quick Links" box containing links to COVID-19 preprints on arXiv, medRxiv, and bioRxiv. It also includes a warning about the nature of e-prints. To the left, there's a "Subject search and browse" section for Physics, and below it, a "News" section with a link to the arXiv blog. The bottom part of the page lists categories like Astrophysics, Condensed Matter, and General Relativity and Quantum Cosmology.

arxiv.org

Cornell University

We gratefully acknowledge support from the Simons Foundation and member institutions.

arXiv.org

Search... All fields Search

Login

Help | Advanced Search

COVID-19 Quick Links

See COVID-19 SARS-CoV-2 preprints from

- arXiv
- medRxiv and bioRxiv

Important: e-prints posted on arXiv are not peer-reviewed by arXiv; they should not be relied upon without context to guide clinical practice or health-related behavior and should not be reported in news media as established information without consulting multiple experts in the field.

Subject search and browse:

Physics

Form Interface | Catchup

News

Read about recent news and updates on arXiv's blog. (View the former "what's new" pages here). Read robots beware before attempting any automated download.

Physics

- **Astrophysics (astro-ph new, recent, search)**
includes: [Astrophysics of Galaxies](#); [Cosmology and Nongalactic Astrophysics](#); [Earth and Planetary Astrophysics](#); [High Energy Astrophysical Phenomena](#); [Instrumentation and Methods for Astrophysics](#); [Solar and Stellar Astrophysics](#)
- **Condensed Matter (cond-mat new, recent, search)**
includes: [Disordered Systems and Neural Networks](#); [Materials Science](#); [Mesoscale and Nanoscale Physics](#); [Other Condensed Matter](#); [Quantum Gases](#); [Soft Condensed Matter](#); [Statistical Mechanics](#); [Strongly Correlated Electrons](#); [Superconductivity](#)
- **General Relativity and Quantum Cosmology (gr-qc new, recent, search)**

Where to look for papers on arXiv arXiv itself?

- [arxiv.org](#)

Computer Science

- Computing Research Repository ([CoRR new, recent, search](#))
includes (see detailed description): Artificial Intelligence; Computation and Language; Computational Complexity; Computational Engineering, Finance, and Science; Computational Geometry; Computer Science and Game Theory; Computer Vision and Pattern Recognition; Computers and Society; Cryptography and Security; Data Structures and Algorithms; Databases; Digital Libraries; Discrete Mathematics; Distributed, Parallel, and Cluster Computing; Emerging Technologies; Formal Languages and Automata Theory; General Literature; Graphics; Hardware Architecture; Human–Computer Interaction; Information Retrieval; Information Theory; Logic in Computer Science; Machine Learning; Mathematical Software; Multiagent Systems; Multimedia; Networking and Internet Architecture; Neural and Evolutionary Computing; Numerical Analysis; Operating Systems; Other Computer Science; Performance; Programming Languages; Robotics; Social and Information Networks; Software Engineering; Sound; Symbolic Computation; Systems and Control

Quantitative Biology

- Quantitative Biology ([q-bio new, recent, search](#))
includes (see detailed description): Biomolecules; Cell Behavior; Genomics; Molecular Networks; Neurons and Cognition; Other Quantitative Biology; Populations and Evolution; Quantitative Methods; Subcellular Processes; Tissues and Organs

Quantitative Finance

- Quantitative Finance ([q-fin new, recent, search](#))
includes (see detailed description): Computational Finance; Economics; General Finance; Mathematical Finance; Portfolio Management; Pricing of Securities; Risk Management; Statistical Finance; Trading and Market Microstructure

Statistics

- Statistics ([stat new, recent, search](#))
includes (see detailed description): Applications; Computation; Machine Learning; Methodology; Other Statistics; Statistics Theory

Where to look for papers on arXiv arXiv itself?

- arxiv.org

Computer Science

- Computing Research Repository ([CoRR new, recent, search](#))
includes (see detailed description): Artificial Intelligence; Computation and Language; Computational Complexity; Computational Engineering, Finance, and Science; Computational Geometry; Computer Science and Game Theory; Computer Vision and Pattern Recognition; Computers and Society; Cryptography and Security; Data Structures and Algorithms; Databases; Digital Libraries; Discrete Mathematics; Distributed, Parallel, and Cluster Computing; Emerging Technologies; Formal Languages and Automata Theory; General Literature; Graphics; Hardware Architecture; Human–Computer Interaction; Information Retrieval; Information Theory; Logic in Computer Science; Machine Learning; Mathematical Software; Multiagent Systems; Multimedia; Networking and Internet Architecture; Neural and Evolutionary Computing; Numerical Analysis; Operating Systems; Other Computer Science; Performance; Programming Languages; Robotics; Social and Information Networks; Software Engineering; Sound; Symbolic Computation; Systems and Control

Quantitative Biology

- Quantitative Biology ([q-bio new, recent, search](#))
includes (see detailed description): Biomolecules; Cell Behavior; Genomics; Molecular Networks; Neurons and Cognition; Other Quantitative Biology; Populations and Evolution; Quantitative Methods; Subcellular Processes; Tissues and Organs

Quantitative Finance

- Quantitative Finance ([q-fin new, recent, search](#))
includes (see detailed description): Computational Finance; Economics; General Finance; Mathematical Finance; Portfolio Management; Pricing of Securities; Risk Management; Statistical Finance; Trading and Market Microstructure

Statistics

- Statistics ([stat new, recent, search](#))
includes (see detailed description): Applications; Computation; Machine Learning; Methodology; Other Statistics; Statistics Theory

- CS.Computer Vision and Pattern Recognition
- CS.Artificial Intelligence
- CS.Neural and Evolutionary Computing
- CS.Machine Learning
- CS.Robotics
- Stat.Machine Learning

What you can do on arXiv.org

The screenshot shows the top navigation bar of the arXiv.org website. It includes a red header with the arXiv logo, a search bar with a placeholder 'Search...', a dropdown menu set to 'All fields', and a 'Search' button. Below the search bar are links for 'Help | Advanced Search'.

Artificial Intelligence

Authors and titles for recent submissions

- [Tue, 25 Jan 2022](#)
- [Mon, 24 Jan 2022](#)
- [Fri, 21 Jan 2022](#)
- [Thu, 20 Jan 2022](#)
- [Wed, 19 Jan 2022](#)

[total of 331 entries: 1–25 | 26–50 | 51–75 | 76–100 | ... | 326–331]

[showing 25 entries per page: [fewer](#) | [more](#) | [all](#)]

Tue, 25 Jan 2022 (showing first 25 of 79 entries)

[1] [arXiv:2201.09760](#) [pdf, other]

Multi-Graph Fusion Networks for Urban Region Embedding

[Shangbin Wu](#), [Xu Yan](#), [Xiaoliang Fan](#), [Shirui Pan](#), [Shichao Zhu](#), [Chuanpan Zheng](#), [Ming Cheng](#), [Cheng Wang](#)

Subjects: Artificial Intelligence (cs.AI)

[2] [arXiv:2201.09708](#) [pdf, other]

Towards Collaborative Question Answering: A Preliminary Study

[Xiangkun Hu](#), [Hang Yan](#), [Qipeng Guo](#), [Xipeng Qiu](#), [Weinan Zhang](#), [Zheng Zhang](#)

Subjects: Artificial Intelligence (cs.AI); Computation and Language (cs.CL)

[3] [arXiv:2201.09694](#) [pdf, other]

Scaling Up Knowledge Graph Creation to Large and Heterogeneous Data Sources

[Enrique Iglesias](#), [Samaneh Jozashoori](#), [Maria-Esther Vidal](#)

Subjects: Artificial Intelligence (cs.AI); Databases (cs.DB)

What you can do on arXiv.org

The screenshot shows the arXiv.org search interface. The top navigation bar is red with white text, showing the path "arXiv.org > cs > cs.AI". To the right is a search bar with a placeholder "Search...", a dropdown menu set to "All fields", and a "Search" button. Below the search bar are links for "Help | Advanced Search".

Artificial Intelligence

Authors and titles for recent submissions

- [Tue, 25 Jan 2022](#)
- [Mon, 24 Jan 2022](#)
- [Fri, 21 Jan 2022](#)
- [Thu, 20 Jan 2022](#)
- [Wed, 19 Jan 2022](#)

[total of 331 entries: 1–25 | 26–50 | 51–75 | 76–100 | ... | 326–331]

[showing 25 entries per page: [fewer](#) | [more](#) | [all](#)]

Tue, 25 Jan 2022 (showing first 25 of 79 entries)

[1] [arXiv:2201.09760](#) [pdf, other]

Multi-Graph Fusion Networks for Urban Region Embedding

[Shangbin Wu](#), [Xu Yan](#), [Xiaoliang Fan](#), [Shirui Pan](#), [Shichao Zhu](#), [Chuanpan Zheng](#), [Ming Cheng](#), [Cheng Wang](#)

Subjects: Artificial Intelligence (cs.AI)

[2] [arXiv:2201.09708](#) [pdf, other]

Towards Collaborative Question Answering: A Preliminary Study

[Xiangkun Hu](#), [Hang Yan](#), [Qipeng Guo](#), [Xipeng Qiu](#), [Weinan Zhang](#), [Zheng Zhang](#)

Subjects: Artificial Intelligence (cs.AI); Computation and Language (cs.CL)

[3] [arXiv:2201.09694](#) [pdf, other]

Scaling Up Knowledge Graph Creation to Large and Heterogeneous Data Sources

[Enrique Iglesias](#), [Samaneh Jozashoori](#), [Maria-Esther Vidal](#)

Subjects: Artificial Intelligence (cs.AI); Databases (cs.DB)

Dates, number of entries per page

What you can do on arXiv.org

The screenshot shows the arXiv.org search interface for the 'cs.AI' category. The top navigation bar includes a search bar with dropdown options for 'All fields' and 'Search', and links for 'Help | Advanced Search'. The main title 'Artificial Intelligence' is displayed in large bold letters. Below it, a section titled 'Authors and titles for recent submissions' lists five dates from January 2022: Tue, 25 Jan 2022; Mon, 24 Jan 2022; Fri, 21 Jan 2022; Thu, 20 Jan 2022; and Wed, 19 Jan 2022. A hand-drawn style arrow points from the date list towards the search bar above. Below the dates, there is a link to the first page of results: [total of 331 entries: 1-25 | 26-50 | 51-75 | 76-100 | ... | 326-331]. Further down, a specific submission is highlighted: [1] arXiv:2201.09760 [pdf, other]. The title is 'Multi-Graph Fusion Networks for Urban Region Embedding' by Shangbin Wu, Xu Yan, Xiaoliang Fan, Shirui Pan, Shichao Zhu, Chuanpan Zheng, Ming Cheng, Cheng Wang, with subjects in Artificial Intelligence (cs.AI). Other submissions listed include 'Towards Collaborative Question Answering: A Preliminary Study' and 'Scaling Up Knowledge Graph Creation to Large and Heterogeneous Data Sources'.

Dates, number of entries per page

What you can do on arXiv.org

The screenshot shows the arXiv.org search results for the category cs.AI. At the top, there's a red header bar with the arXiv.org logo, a search bar containing 'Search...', a dropdown menu set to 'All fields', and a 'Search' button. Below the header, the category 'Artificial Intelligence' is displayed. A large, hand-drawn style arrow points from the text 'Some papers appear on several lists' down towards the search results. The search results are titled 'Authors and titles for recent submissions' and show a list of dates from January 19 to 25, 2022. Below the dates, there are navigation links for page numbers and a link to see all entries. The first three results are listed below:

Tue, 25 Jan 2022 (showing first 25 of 79 entries)

[1] [arXiv:2201.09760 \[pdf, other\]](#)
Multi-Graph Fusion Networks for Urban Region Embedding
Shangbin Wu, Xu Yan, Xiaoliang Fan, Shirui Pan, Shichao Zhu, Chuanpan Zheng, Ming Cheng, Cheng Wang
Subjects: Artificial Intelligence (cs.AI)

[2] [arXiv:2201.09708 \[pdf, other\]](#)
Towards Collaborative Question Answering: A Preliminary Study
Xiangkun Hu, Hang Yan, Qipeng Guo, Xipeng Qiu, Weinan Zhang, Zheng Zhang
Subjects: Artificial Intelligence (cs.AI); Computation and Language (cs.CL)

[3] [arXiv:2201.09694 \[pdf, other\]](#)
Scaling Up Knowledge Graph Creation to Large and Heterogeneous Data Sources
Enrique Iglesias, Samaneh Jozashoori, Maria-Ester Vidal
Subjects: Artificial Intelligence (cs.AI); Databases (cs.DB)

Dates, number of entries per page

Some papers appear on several lists

What you can do on arXiv.org

The screenshot shows the arXiv.org search interface for the 'cs.AI' category. At the top, there's a red header bar with the arXiv.org logo, a search bar containing 'Search...', a dropdown menu set to 'All fields', and a 'Search' button. Below the header, the page title 'Artificial Intelligence' is displayed in large bold letters. Underneath, a section titled 'Authors and titles for recent submissions' lists five dates of submission: Tue, 25 Jan 2022; Mon, 24 Jan 2022; Fri, 21 Jan 2022; Thu, 20 Jan 2022; and Wed, 19 Jan 2022. A hand-drawn style arrow points from this list towards the first paper entry below. Below the dates, a note indicates a total of 331 entries and shows links for pages 1-25, 26-50, 51-75, 76-100, ..., 326-331, with 'more' and 'all' options. The main content area starts with a heading 'Tue, 25 Jan 2022 (showing first 25 of 79 entries)'. Three paper entries are listed:

[1] [arXiv:2201.09760](#) [pdf, other]
Multi-Graph Fusion Networks for Urban Region Embedding
Shangbin Wu, Xu Yan, Xiaoliang Fan, Shirui Pan, Shichao Zhu, Chuanpan Zheng, Ming Cheng, Cheng Wang
Subjects: Artificial Intelligence (cs.AI)

[2] [arXiv:2201.09708](#) [pdf, other]
Towards Collaborative Question Answering: A Preliminary Study
Xiangkun Hu, Hang Yan, Qipeng Guo, Xipeng Qiu, Weinan Zhang, Yang Zhang
Subjects: Artificial Intelligence (cs.AI); Computation and Language (cs.CL)

[3] [arXiv:2201.09694](#) [pdf, other]
Scaling Up Knowledge Graph Creation to Large and Heterogeneous Data Sources
Enrique Iglesias, Samaneh Jozashoori, Maria-Ester Vidal
Subjects: Artificial Intelligence (cs.AI); Databases (cs.DB)

Dates, number of entries per page

Some papers appear on several lists

What you can do on arXiv.org

The screenshot shows the arXiv.org search interface for the category cs.AI. At the top, there's a red header bar with the arXiv.org logo, a search bar containing 'Search...', a dropdown menu set to 'All fields', and a 'Search' button. Below the header, the category 'Artificial Intelligence' is selected. The main content area displays a list of recent submissions. A large arrow points from the text 'Some papers appear on several lists' to the list of authors and titles. Another large arrow points from the text 'You can see only title and authors' to the individual paper entries.

arXiv.org > cs > cs.AI

Search... All fields Search
Help | Advanced Search

Artificial Intelligence

Authors and titles for recent submissions

- Tue, 25 Jan 2022
- Mon, 24 Jan 2022
- Fri, 21 Jan 2022
- Thu, 20 Jan 2022
- Wed, 19 Jan 2022

[total of 331 entries: 1-25 | 26-50 | 51-75 | 76-100 | ... | 326-331]
[showing 25 entries per page: [fewer](#) | [more](#) | [all](#)]

Tue, 25 Jan 2022 (showing first 25 of 79 entries)

[1] [arXiv:2201.09760](#) [pdf, other]
Multi-Graph Fusion Networks for Urban Region Embedding
Shangbin Wu, Xu Yan, Xiaoliang Fan, Shirui Pan, Shichao Zhu, Chuanpan Zheng, Ming Cheng, Cheng Wang
Subjects: Artificial Intelligence (cs.AI)

[2] [arXiv:2201.09708](#) [pdf, other]
Towards Collaborative Question Answering: A Preliminary Study
Xiangkun Hu, Hang Yan, Qipeng Guo, Xipeng Qiu, Weinan Zhang, Yang Zhang
Subjects: Artificial Intelligence (cs.AI); Computation and Language (cs.CL)

[3] [arXiv:2201.09694](#) [pdf, other]
Scaling Up Knowledge Graph Creation to Large and Heterogeneous Data Sources
Enrique Iglesias, Samaneh Jozashoori, Maria-Ester Vidal
Subjects: Artificial Intelligence (cs.AI); Databases (cs.DB)

Dates, number of entries per page

Some papers appear on several lists

You can see only title and authors

What you can do on arXiv.org

The screenshot shows the arXiv.org search interface for the 'cs.AI' category. At the top, there's a red header bar with the arXiv.org logo, a search bar containing 'Search...', a dropdown menu set to 'All fields', and a 'Search' button. Below the header, the word 'Artificial Intelligence' is displayed in large bold letters. Underneath, a section titled 'Authors and titles for recent submissions' lists five dates from January 2022: Tue, 25 Jan 2022; Mon, 24 Jan 2022; Fri, 21 Jan 2022; Thu, 20 Jan 2022; and Wed, 19 Jan 2022. A hand-drawn style arrow points from this list towards the first paper entry. Below this list, there are navigation links for page numbers and a link to see all entries. The main content area shows three paper entries for Tuesday, January 25, 2022:

[1] arXiv:2201.09760 [pdf, other]
Multi-Graph Fusion Networks for Urban Region Embedding
Shangbin Wu, Xu Yan, Xiaoliang Fan, Shirui Pan, Shichao Zhu, Chuanpan Zheng, Ming Cheng, Cheng Wang
Subjects: Artificial Intelligence (cs.AI)

[2] arXiv:2201.09708 [pdf, other]
Towards Collaborative Question Answering: A Preliminary Study
Xiangkun Hu, Hang Yan, Qipeng Guo, Xipeng Qiu, Weinan Zhang, Meng Zhang
Subjects: Artificial Intelligence (cs.AI); Computation and Language (cs.CL)

[3] arXiv:2201.09694 [pdf, other]
Scaling Up Knowledge Graph Creation to Large and Heterogeneous Data Sources
Enrique Iglesias, Samaneh Jozashoori, Maria-Ester Vidal
Subjects: Artificial Intelligence (cs.AI); Databases (cs.DB)

Dates, number of entries per page

Some papers appear on several lists

You can see only title and authors

“other” means latex source, useful for copying tables, images, etc

What you can do on arXiv.org

The screenshot shows the arXiv.org search interface for the 'cs.AI' category. At the top, there's a red header bar with the arXiv.org logo, a search bar containing 'Search...', a dropdown menu set to 'All fields', and a 'Search' button. Below the header, the word 'Artificial Intelligence' is displayed in large bold letters. Underneath, a section titled 'Authors and titles for recent submissions' lists five dates from January 2022: Tue, 25 Jan 2022; Mon, 24 Jan 2022; Fri, 21 Jan 2022; Thu, 20 Jan 2022; and Wed, 19 Jan 2022. A hand-drawn style arrow points from this list towards the first paper entry. Below this list, there are navigation links for page numbers and a link to see all entries. The main content area shows three paper entries for Tuesday, January 25, 2022:

[1] arXiv:2201.09760 [pdf, other]
Multi-Graph Fusion Networks for Urban Region Embedding
Shangbin Wu, Xu Yan, Xiaoliang Fan, Shirui Pan, Shichao Zhu, Chuanpan Zheng, Ming Cheng, Cheng Wang
Subjects: Artificial Intelligence (cs.AI)

[2] arXiv:2201.09708 [pdf, other]
Towards Collaborative Question Answering: A Preliminary Study
Xiangkun Hu, Hang Yan, Qipeng Guo, Xipeng Qiu, Weinan Zhang, Yiqing Zhang
Subjects: Artificial Intelligence (cs.AI); Computation and Language (cs.CL)

[3] arXiv:2201.09694 [pdf, other]
Scaling Up Knowledge Graph Creation to Large and Heterogeneous Data Sources
Enrique Iglesias, Samaneh Jozashoori, Maria-Ester Vidal
Subjects: Artificial Intelligence (cs.AI); Databases (cs.DB)

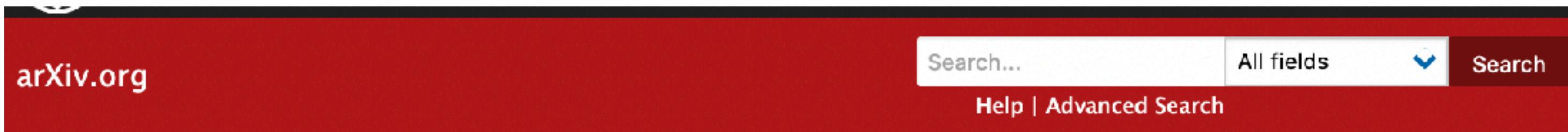
Dates, number of entries per page

Some papers appear on several lists

You can see only title and authors

“other” means latex source, useful for copying tables, images, etc

What you can do on arXiv.org “other” format



Format selector for arXiv:2201.09760v1

Set cookies: if your browser supports cookies you can [configure your default format](#).

PDF

Now includes fonts, see our [PDF help](#). [Download PDF](#)

Source

Delivered as a **gzipped tar** (.tar.gz) file if there are multiple files, otherwise as a **PDF** file, or a **gzipped TeX, DVI, PostScript or HTML** (.gz, .dvi.gz, .ps.gz or .html.gz) file depending on submission format. [[Download source](#)]

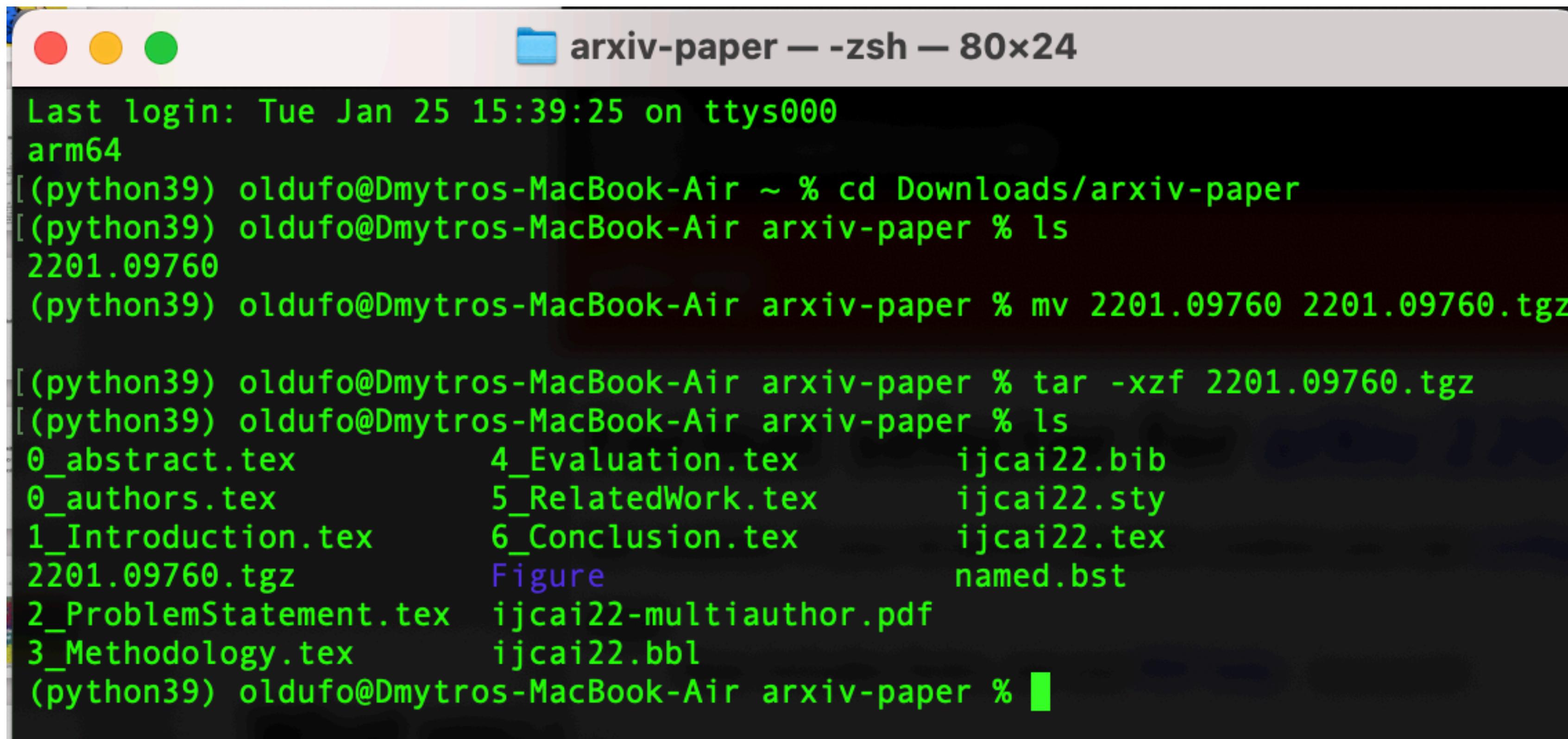
For more information consult the [viewing help](#).

Note: Many of the formats above are served gzipped (Content-Encoding: x-gzip). Your browser may silently uncompress after downloading so the files you see saved may appear uncompressed.

Please report any problems to [arXiv admins](#), being certain to state the paper identifier, explicit options you've selected and that you're using arxiv.org.

What you can do on arXiv.org

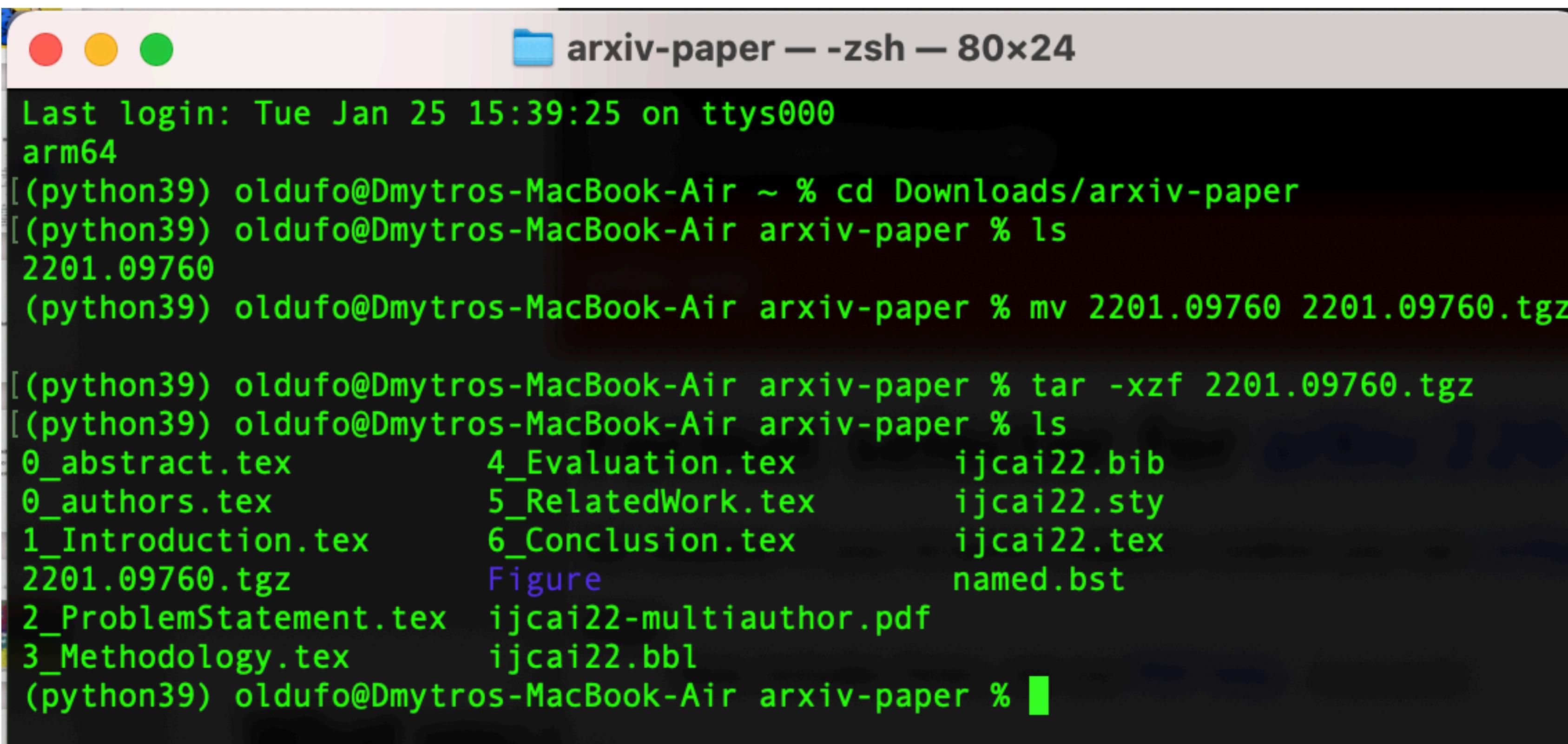
“other” format



Last login: Tue Jan 25 15:39:25 on ttys000
arm64
[(python39) oldufo@Dmytros-MacBook-Air ~ % cd Downloads/arxiv-paper
[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % ls
2201.09760
[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % mv 2201.09760 2201.09760.tgz
[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % tar -xzf 2201.09760.tgz
[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % ls
0_abstract.tex 4_Evaluation.tex ijcai22.bib
0_authors.tex 5_RelatedWork.tex ijcai22.sty
1_Introduction.tex 6_Conclusion.tex ijcai22.tex
2201.09760.tgz Figure named.bst
2_ProblemStatement.tex ijcai22-multiauthor.pdf
3_Methodology.tex ijcai22.bbl
[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % █

What you can do on arXiv.org

“other” format



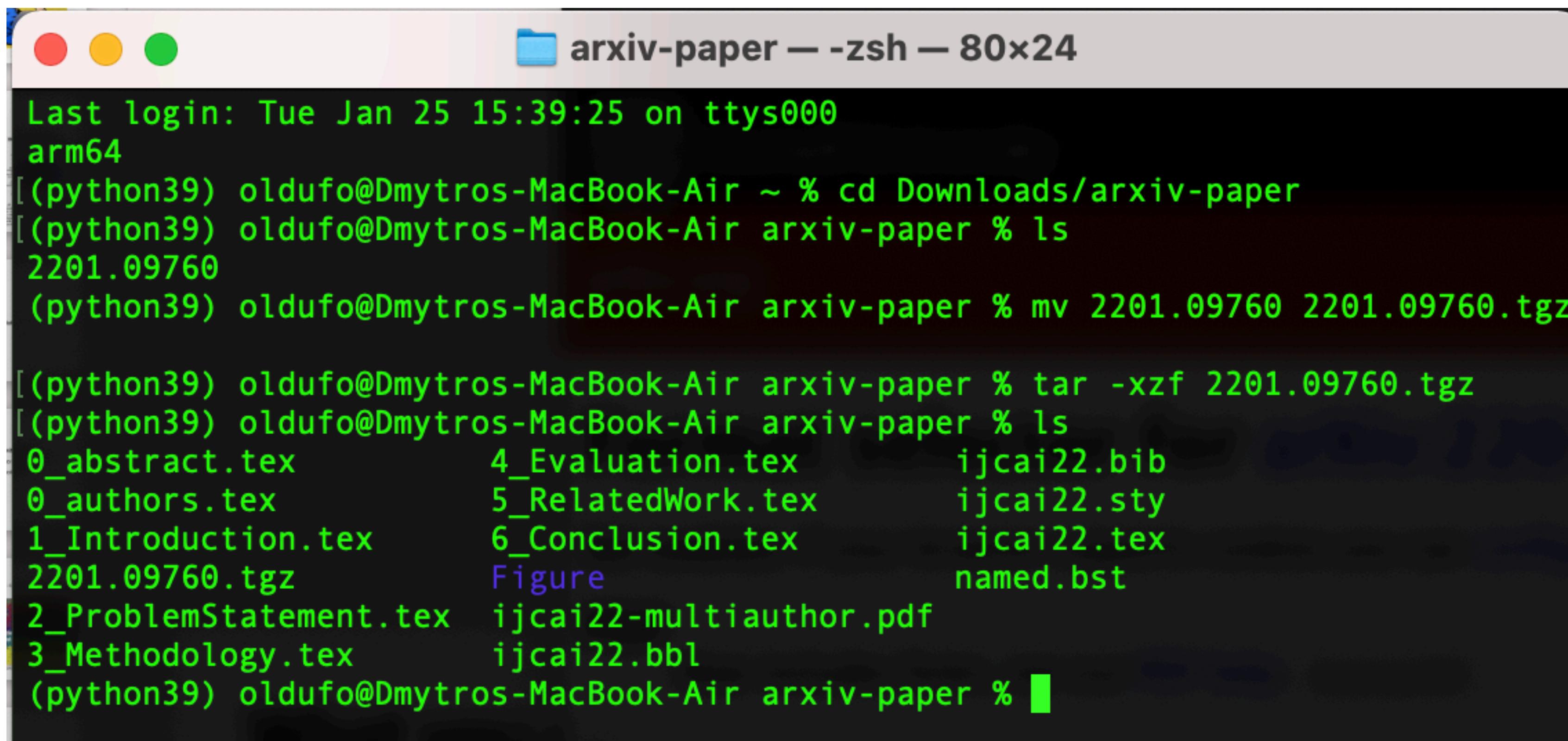
Last login: Tue Jan 25 15:39:25 on ttys000
arm64
[(python39) oldufo@Dmytros-MacBook-Air ~ % cd Downloads/arxiv-paper
[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % ls
2201.09760
[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % mv 2201.09760 2201.09760.tgz

[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % tar -xzf 2201.09760.tgz
[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % ls
0_abstract.tex 4_Evaluation.tex ijcai22.bib
0_authors.tex 5_RelatedWork.tex ijcai22.sty
1_Introduction.tex 6_Conclusion.tex ijcai22.tex
2201.09760.tgz Figure named.bst
2_ProblemStatement.tex ijcai22-multiauthor.pdf
3_Methodology.tex ijcai22.bbl
[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % █

- arXiv downloads “other” as a file without extension

What you can do on arXiv.org

“other” format



```
Last login: Tue Jan 25 15:39:25 on ttys000
arm64
[(python39) oldufo@Dmytros-MacBook-Air ~ % cd Downloads/arxiv-paper
[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % ls
2201.09760
[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % mv 2201.09760 2201.09760.tgz

[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % tar -xzf 2201.09760.tgz
[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % ls
0_abstract.tex          4_Evaluation.tex      ijcai22.bib
0_authors.tex            5_RelatedWork.tex    ijcai22.sty
1_Introduction.tex       6_Conclusion.tex    ijcai22.tex
2201.09760.tgz          Figure               named.bst
2_ProblemStatement.tex   ijcai22-multiauthor.pdf
3_Methodology.tex        ijcai22.bbl
(python39) oldufo@Dmytros-MacBook-Air arxiv-paper %
```

- arXiv downloads “other” as a file without extension
- it is actually tar.gz file

What you can do on arXiv.org

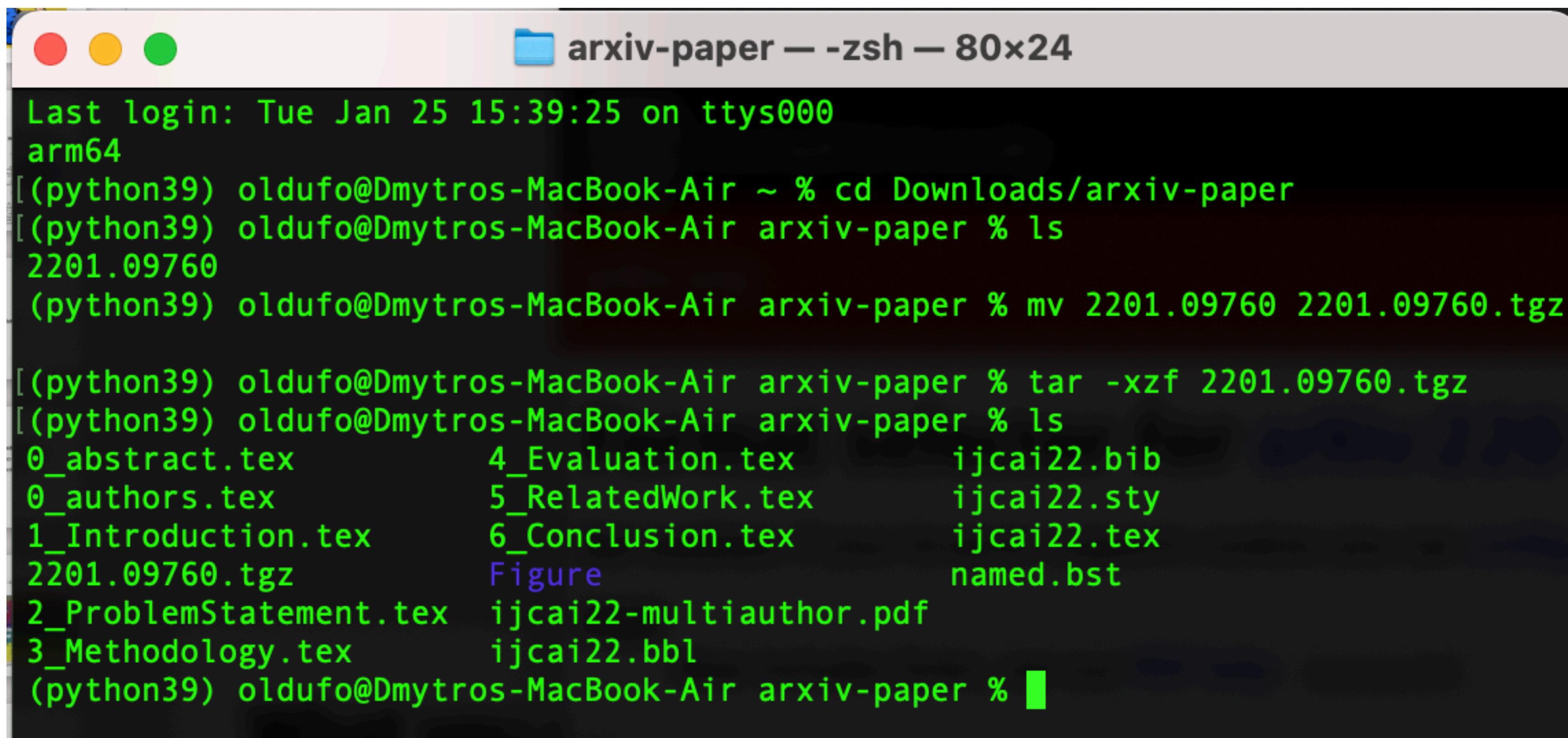
“other” format

```
Last login: Tue Jan 25 15:39:25 on ttys000
arm64
[(python39) oldufo@Dmytros-MacBook-Air ~ % cd Downloads/arxiv-paper
[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % ls
2201.09760
[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % mv 2201.09760 2201.09760.tgz
[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % tar -xzf 2201.09760.tgz
[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % ls
0_abstract.tex      4_Evaluation.tex      ijcai22.bib
0_authors.tex        5_RelatedWork.tex    ijcai22.sty
1_Introduction.tex   6_Conclusion.tex    ijcai22.tex
2201.09760.tgz      Figure                named.bst
2_ProblemStatement.tex ijcai22-multiauthor.pdf
3_Methodology.tex    ijcai22.bbl
(python39) oldufo@Dmytros-MacBook-Air arxiv-paper %
```

- arXiv downloads “other” as a file without extension
- it is actually tar.gz file
- It will unpack to the current dir, so better to create a new dir

What you can do on arXiv.org

“other” format

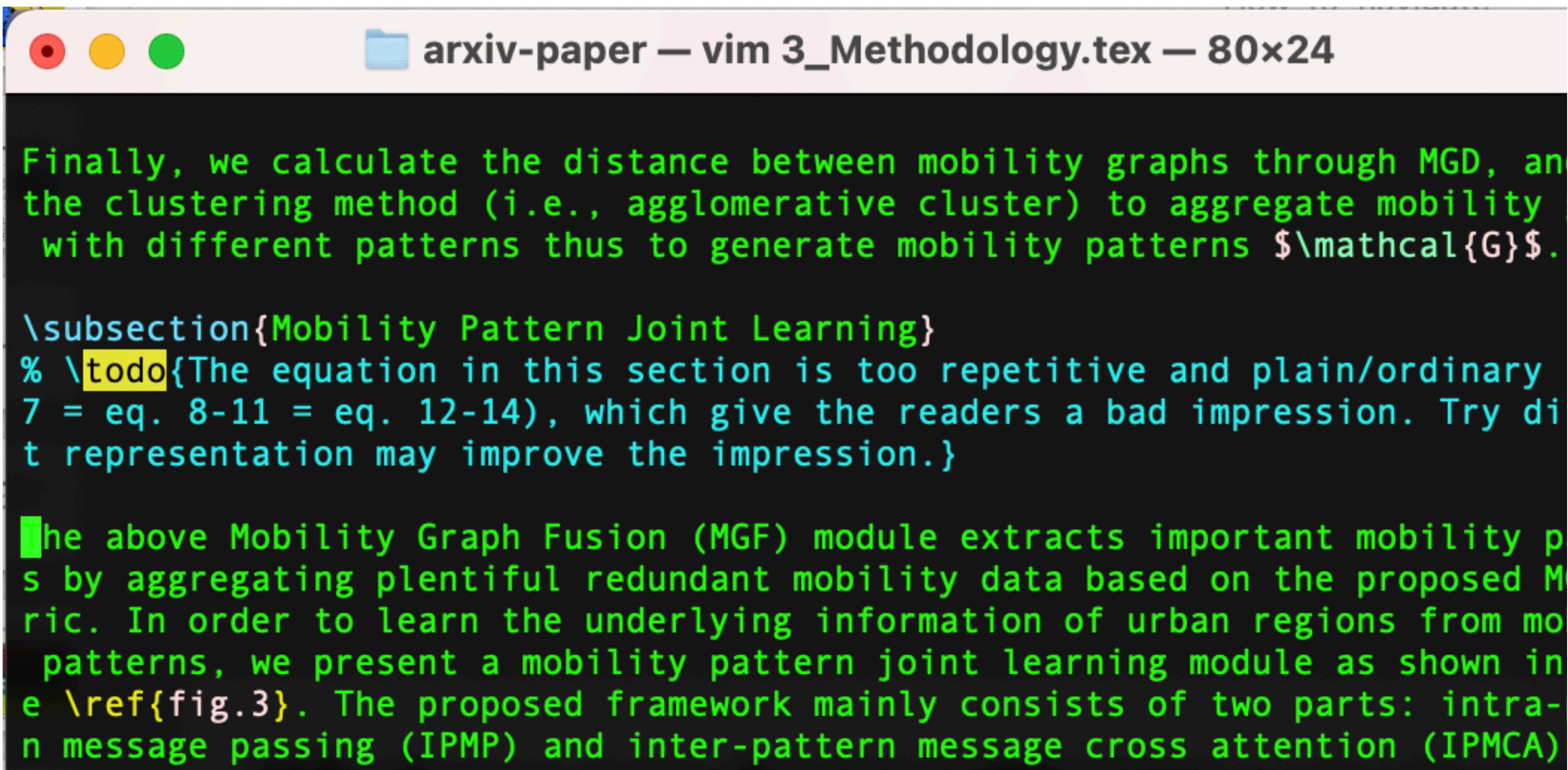


Last login: Tue Jan 25 15:39:25 on ttys000
arm64
[(python39) oldufo@Dmytros-MacBook-Air ~ % cd Downloads/arxiv-paper
[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % ls
2201.09760
[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % mv 2201.09760 2201.09760.tgz
[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % tar -xzf 2201.09760.tgz
[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper % ls
0_abstract.tex 4_Evaluation.tex ijcai22.bib
0_authors.tex 5_RelatedWork.tex ijcai22.sty
1_Introduction.tex 6_Conclusion.tex ijcai22.tex
2201.09760.tgz Figure named.bst
2_ProblemStatement.tex ijcai22-multiauthor.pdf
3_Methodology.tex ijcai22.bbl
[(python39) oldufo@Dmytros-MacBook-Air arxiv-paper %

- arXiv downloads “other” as a file without extension
- it is actually tar.gz file
- It will unpack to the current dir, so better to create a new dir
- You can find interesting points and todos there, if authors haven’t deleted them

What you can do on arXiv.org

“other” format



The screenshot shows a terminal window with a Vim editor open. The title bar says "arxiv-paper — vim 3_Methodology.tex — 80x24". The Vim status bar at the bottom shows "100% 1 100 100". The buffer content is as follows:

```
Finally, we calculate the distance between mobility graphs through MGD, an  
the clustering method (i.e., agglomerative cluster) to aggregate mobility  
with different patterns thus to generate mobility patterns  $\mathcal{G}$ .  
  
\subsection{Mobility Pattern Joint Learning}  
% \todo{The equation in this section is too repetitive and plain/ordinary  
7 = eq. 8-11 = eq. 12-14), which give the readers a bad impression. Try di  
t representation may improve the impression.  
  
The above Mobility Graph Fusion (MGF) module extracts important mobility p  
s by aggregating plentiful redundant mobility data based on the proposed M  
ric. In order to learn the underlying information of urban regions from mo  
patterns, we present a mobility pattern joint learning module as shown in  
e \ref{fig.3}. The proposed framework mainly consists of two parts: intra-  
n message passing (IPMP) and inter-pattern message cross attention (IPMCA)
```

- arXiv downloads “other” as a file without extension
- it is actually tar.gz file
- It will unpack to the current dir, so better to create a new dir
- You can find interesting points and todos there, if authors haven’t deleted them

What you can do on arXiv.org

Citations

The screenshot shows a search results page for arXiv.org. At the top, there's a header with the Cornell University logo and a message of thanks to the Simons Foundation. Below the header, a search bar with dropdown menus for 'Search...' and 'All fields' is visible, along with a 'Search' button and links for 'Help | Advanced Search'. The main content area displays a single research paper entry. The title is 'Multi-Graph Fusion Networks for Urban Region Embedding' by Shangbin Wu, Xu Yan, Xiaoliang Fan, Shirui Pan, Shichao Zhu, Chuanpan Zheng, Ming Cheng, Cheng Wang. It was submitted on 24 Jan 2022. The abstract discusses learning embeddings for urban regions from human mobility data to enable tasks like crime prediction. The right sidebar contains sections for 'Download' (PDF, Other formats, license), 'Current browse context' (cs.AI, with navigation links for < prev, next, new, recent, 2201), 'Change to browse by' (cs), 'References & Citations' (NASA ADS, Google Scholar, Semantic Scholar), 'Export Bibtex Citation', and 'Bookmark' (with icons for various platforms).

Computer Science > Artificial Intelligence

[Submitted on 24 Jan 2022]

Multi-Graph Fusion Networks for Urban Region Embedding

Shangbin Wu, Xu Yan, Xiaoliang Fan, Shirui Pan, Shichao Zhu, Chuanpan Zheng, Ming Cheng, Cheng Wang

Learning the embeddings for urban regions from human mobility data can reveal the functionality of regions, and then enables the correlated but distinct tasks such as crime prediction. Human mobility data contains rich but abundant information, which yields to the comprehensive region embeddings for cross domain tasks. In this paper, we propose multi-graph fusion networks (MGFN) to enable the cross domain prediction tasks. First, we integrate the graphs with spatio-temporal similarity as mobility patterns through a mobility graph fusion module. Then, in the mobility pattern joint learning module, we design the multi-level cross-attention mechanism to learn the comprehensive embeddings from multiple mobility patterns based on intra-pattern and inter-pattern messages. Finally, we conduct extensive experiments on real-world urban datasets. Experimental results demonstrate that the proposed MGFN outperforms the state-of-the-art methods by up to 12.35% improvement.

Subjects: Artificial Intelligence (cs.AI)

Cite as: arXiv:2201.09760 [cs.AI]

(or arXiv:2201.09760v1 [cs.AI] for this version)

Submission history

From: Shangbin Wu [view email]

[v1] Mon, 24 Jan 2022 15:48:50 UTC (1,281 KB)

arXiv.org

arXiv.org

What you can do on arXiv.org

Citations

Cornell University We gratefully acknowledge support from the Simons Foundation and member institutions

arXiv.org > cs > arXiv:2201.09760

Search... All fields Search Help | Advanced Search

Computer Science > Artificial Intelligence

(Submitted on 24 Jan 2022)

Multi-Graph Fusion Networks for Urban Region Embedding

Shangbin Wu, Xu Yan, Xiaoliang Fan, Shirui Pan, Shichao Zhu, Chuanpan Zheng, Ming Cheng, Cheng Wang

Learning the embeddings for urban regions from human mobility data can reveal the functionality of regions, and then enables the correlated but distinct tasks such as crime prediction. Human mobility data contains rich but abundant information, which yields to the comprehensive region embeddings for cross domain tasks. In this paper, we propose multi-graph fusion networks (MGFN) to enable the cross domain prediction tasks. First, we integrate the graphs with spatio-temporal similarity as mobility patterns through a mobility graph fusion module. Then, in the mobility pattern joint learning module, we design the multi-level cross-attention mechanism to learn the comprehensive embeddings from multiple mobility patterns based on intra-pattern and inter-pattern messages. Finally, we conduct extensive experiments on real-world urban datasets. Experimental results demonstrate that the proposed MGFN outperforms the state-of-the-art methods by up to 12.35% improvement.

Subjects: Artificial Intelligence (cs.AI)

Cite as: arXiv:2201.09760 [cs.AI]
(or arXiv:2201.09760v1 [cs.AI] for this version)

Submission history

From: Shangbin Wu [view email]
[v1] Mon, 24 Jan 2022 15:48:50 UTC (1,281 KB)

Download:

- PDF
- Other formats
(license)

Current browse context:
cs.AI
< prev | next >
new | recent | 2201

Change to browse by:
cs

References & Citations

- NASA ADS
- Google Scholar
- Semantic Scholar

Export Bibtex Citation

Bookmark

X Y Z Science WISE

What you can do on arXiv.org

Citations

The screenshot shows a detailed view of an arXiv.org page for a Computer Science paper. At the top, the Cornell University logo and a note of thanks to the Simons Foundation are visible. The navigation bar includes links for 'Search...', 'All fields', 'Help | Advanced Search', and 'Search'. The main content area displays the paper's title, authors, and abstract. A modal window titled 'Bibtex formatted citation' contains the following Bibtex code:

```
@misc{wu2022multigraph,  
    title={(Multi-Graph Fusion Networks for Urban Region Embedding)},  
    author=(Shangbin Wu and Xu Yan and Xiaoliang Fan and Shirui Pan and Shichao Zhu and Chuanpan Zheng  
and Ming Cheng and Cheng Wang),  
    year={2022},  
    eprint=(2201.09760),  
    archivePrefix=(arXiv),  
    primaryClass={cs.AI}  
}
```

Below the Bibtex code, a note says 'Data provided by: arXiv API'. To the right of the main content, a sidebar offers download options ('PDF', 'Other formats (license)'), a current browse context ('arXiv > cs > 201'), and citation-related links ('Citations', 'Cite as', 'arXivolar', 'Citation').

Subjects: Artificial Intelligence (cs.AI)
Cite as: arXiv:2201.09760 [cs.AI]
(or arXiv:2201.09760v1 [cs.AI] for this version)

Submission history
From: Shangbin Wu [view email]
[v1] Mon, 24 Jan 2022 15:48:50 UTC (1,281 KB)

Where to look for papers on arXiv

arxiv-sanity.com

Arxiv Sanity Preserver
Built in spare time by @karpathy to accelerate research.
Serving last 185571 papers from cs.[CV|CL|LG|AI|NE]/stat.ML
(New: Nov 27, 2021) Have a look at the new arxiv-sanity-lite.com

ducha.aiki@gmail.com log out Fork me on GitHub

most recent top recent top hype friends discussions recommended library

Only show v1

Showing most recent Arxiv papers:

CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving
Geunseob Oh, Huei Peng
1/24/2022 cs.LG | cs.AI | cs.RO

2201.09874v1 [pdf](#)
[show similar](#) | [discuss](#)

The task of predicting stochastic behaviors of road agents in diverse environments is a challenging problem for autonomous driving. To best understand scene contexts and produce diverse possible future states of the road agents adaptively in different environments, a prediction model should be probabilistic, multi-modal, context-driven, and general. We present Conditionalizing Variational AutoEncoders via Hypernetworks (CVAE-H); a conditional VAE that extensively leverages hypernetwork and performs generative tasks for high-dimensional problems like the prediction task. We first evaluate CVAE-H on simple generative experiments to show that CVAE-H is probabilistic, multi-modal, context-driven, and general. Then, we demonstrate that the proposed model effectively solves a self-driving prediction problem by producing accurate predictions of road agents in various environments.

Where to look for papers on arXiv

arxiv-sanity.com

The screenshot shows the homepage of arxiv-sanity.com. At the top, there's a red header bar with the text "Arxiv Sanity Preserver", "Built in spare time by @karpathy to accelerate research.", "Serving last 185571 papers from cs.[CV|CL|LG|AI|NE]/stat.ML", and "(New: Nov 27, 2021) Have a look at the new [arxiv-sanity-lite.com](#)". On the right of the header are "ducha.aiki@gmail.com", "log out", and a "Fork me on GitHub" button. Below the header is a search bar with a magnifying glass icon. Underneath the search bar is a navigation menu with buttons for "most recent", "top recent", "top hype", "friends", "discussions", "recommended", and "library". A "Only show v1" button is located below the navigation menu. The main content area displays a paper summary for "CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving" by Geunseob Oh, Huei Peng. The paper was submitted on 1/24/2022 and has categories cs.LG | cs.AI | cs.RO. It includes a PDF link (2201.09874v1 pdf), "show similar", "discuss", and a save icon. Below the summary is a green box containing a brief abstract of the paper.

Arxiv Sanity Preserver
Built in spare time by @karpathy to accelerate research.
Serving last 185571 papers from cs.[CV|CL|LG|AI|NE]/stat.ML
(New: Nov 27, 2021) Have a look at the new [arxiv-sanity-lite.com](#)

ducha.aiki@gmail.com log out Fork me on GitHub

most recent top recent top hype friends discussions recommended library

Only show v1

Showing most recent Arxiv papers:

CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving
Geunseob Oh, Huei Peng
1/24/2022 cs.LG | cs.AI | cs.RO

2201.09874v1 pdf
[show similar](#) | [discuss](#)

The task of predicting stochastic behaviors of road agents in diverse environments is a challenging problem for autonomous driving. To best understand scene contexts and produce diverse possible future states of the road agents adaptively in different environments, a prediction model should be probabilistic, multi-modal, context-driven, and general. We present Conditionalizing Variational AutoEncoders via Hypernetworks (CVAE-H); a conditional VAE that extensively leverages hypernetwork and performs generative tasks for high-dimensional problems like the prediction task. We first evaluate CVAE-H on simple generative experiments to show that CVAE-H is probabilistic, multi-modal, context-driven, and general. Then, we demonstrate that the proposed model effectively solves a self-driving prediction problem by producing accurate predictions of road agents in various environments.

- Developed by Andrew Karpathy

Where to look for papers on arXiv

arxiv-sanity.com

The screenshot shows the homepage of arxiv-sanity.com. At the top, there's a red header bar with the text "Arxiv Sanity Preserver", "Built in spare time by @karpathy to accelerate research.", "Serving last 185571 papers from cs.[CV|CL|LG|AI|NE]/stat.ML", "(New: Nov 27, 2021) Have a look at the new [arxiv-sanity-lite.com](#)", and user info "ducha.aiki@gmail.com" and "log out". A "Fork me on GitHub" button is also present. Below the header is a search bar with a magnifying glass icon. Underneath the search bar is a navigation menu with buttons for "most recent", "top recent", "top hype", "friends", "discussions", "recommended", and "library". A "Only show v1" button is located below the navigation menu. The main content area displays a single paper thumbnail for "CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving" by Geunseob Oh, Huei Peng, dated 1/24/2022, with categories cs.LG, cs.AI, and cs.RO. The thumbnail shows a grid of small images representing different driving scenarios. Below the thumbnail is a green summary box containing the first few paragraphs of the paper's abstract.

Arxiv Sanity Preserver
Built in spare time by @karpathy to accelerate research.
Serving last 185571 papers from cs.[CV|CL|LG|AI|NE]/stat.ML
(New: Nov 27, 2021) Have a look at the new [arxiv-sanity-lite.com](#)

ducha.aiki@gmail.com log out

Fork me on GitHub

most recent top recent top hype friends discussions recommended library

Only show v1

Showing most recent Arxiv papers:

CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving
Geunseob Oh, Huei Peng
1/24/2022 cs.LG | cs.AI | cs.RO

2201.09874v1 [pdf](#)
[show similar](#) | [discuss](#)

The task of predicting stochastic behaviors of road agents in diverse environments is a challenging problem for autonomous driving. To best understand scene contexts and produce diverse possible future states of the road agents adaptively in different environments, a prediction model should be probabilistic, multi-modal, context-driven, and general. We present Conditionalizing Variational AutoEncoders via Hypernetworks (CVAE-H); a conditional VAE that extensively leverages hypernetwork and performs generative tasks for high-dimensional problems like the prediction task. We first evaluate CVAE-H on simple generative experiments to show that CVAE-H is probabilistic, multi-modal, context-driven, and general. Then, we demonstrate that the proposed model effectively solves a self-driving prediction problem by producing accurate predictions of road agents in various environments.

- Developed by Andrew Karpathy
- Has uber-feature: page thumbnails

Where to look for papers on arXiv

arxiv-sanity.com

The screenshot shows the homepage of arxiv-sanity.com. At the top, there's a red header bar with the text "Arxiv Sanity Preserver", "Built in spare time by @karpathy to accelerate research.", "Serving last 185571 papers from cs.[CV|CL|LG|AI|NE]/stat.ML", "(New: Nov 27, 2021) Have a look at the new [arxiv-sanity-lite.com](#)", and user info "ducha.aiki@gmail.com log out". A "Fork me on GitHub" button is also present.

The main content area has a search bar and a navigation menu with links: "most recent", "top recent", "top hype", "friends", "discussions", "recommended", and "library". Below the menu is a button "Only show v1".

A section titled "Showing most recent Arxiv papers:" displays a grid of paper thumbnails. One specific paper is highlighted with a red oval: "CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving" by Geunseob Oh, Huei Peng, dated 1/24/2022, with categories cs.LG | cs.AI | cs.RO. The thumbnail shows a complex diagram of a neural network architecture.

The abstract of the highlighted paper reads:

The task of predicting stochastic behaviors of road agents in diverse environments is a challenging problem for autonomous driving. To best understand scene contexts and produce diverse possible future states of the road agents adaptively in different environments, a prediction model should be probabilistic, multi-modal, context-driven, and general. We present Conditionalizing Variational AutoEncoders via Hypernetworks (CVAE-H); a conditional VAE that extensively leverages hypernetwork and performs generative tasks for high-dimensional problems like the prediction task. We first evaluate CVAE-H on simple generative experiments to show that CVAE-H is probabilistic, multi-modal, context-driven, and general. Then, we demonstrate that the proposed model effectively solves a self-driving prediction problem by producing accurate predictions of road agents in various environments.

- Developed by Andrew Karpathy
- Has uber-feature: page thumbnails

Where to look for papers on arXiv

arxiv-sanity.com

Arxiv Sanity Preserver
Built in spare time by @karpathy to accelerate research.
Serving last 185571 papers from cs.[CV|CL|LG|AI|NE]/stat.ML
(New: Nov 27, 2021) Have a look at the new [arxiv-sanity-lite.com](#)

ducha.aiki@gmail.com log out Fork me on GitHub

most recent top recent top hype friends discussions recommended library

Only show v1

Showing most recent Arxiv papers:

CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving
Geunseob Oh, Huei Peng
1/24/2022 cs.LG | cs.AI | cs.RO

2201.09874v1 pdf
[show similar](#) | [discuss](#)

The task of predicting stochastic behaviors of road agents in diverse environments is a challenging problem for autonomous driving. To best understand scene contexts and produce diverse possible future states of the road agents adaptively in different environments, a prediction model should be probabilistic, multi-modal, context-driven, and general. We present Conditionalizing Variational AutoEncoders via Hypernetworks (CVAE-H); a conditional VAE that extensively leverages hypernetwork and performs generative tasks for high-dimensional problems like the prediction task. We first evaluate CVAE-H on simple generative experiments to show that CVAE-H is probabilistic, multi-modal, context-driven, and general. Then, we demonstrate that the proposed model effectively solves a self-driving prediction problem by producing accurate predictions of road agents in various environments.

- Developed by Andrew Karpathy
- Has uber-feature: page thumbnails
- Also shows paper abstract

Where to look for papers on arXiv

arxiv-sanity.com

Arxiv Sanity Preserver
Built in spare time by @karpathy to accelerate research.
Serving last 185571 papers from cs.[CV|CL|LG|AI|NE]/stat.ML
(New: Nov 27, 2021) Have a look at the new [arxiv-sanity-lite.com](#)

ducha.aiki@gmail.com log out Fork me on GitHub

most recent top recent top hype friends discussions recommended library

Only show v1

Showing most recent Arxiv papers:

CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving
Geunseob Oh, Huei Peng
1/24/2022 cs.LG | cs.AI | cs.RO

2201.09874v1 pdf
[show similar](#) | [discuss](#)

The task of predicting stochastic behaviors of road agents in diverse environments is a challenging problem for autonomous driving. To best understand scene contexts and produce diverse possible future states of the road agents adaptively in different environments, a prediction model should be probabilistic, multi-modal, context-driven, and general. We present Conditionalizing Variational AutoEncoders via Hypernetworks (CVAE-H); a conditional VAE that extensively leverages hypernetwork and performs generative tasks for high-dimensional problems like the prediction task. We first evaluate CVAE-H on simple generative experiments to show that CVAE-H is probabilistic, multi-modal, context-driven, and general. Then, we demonstrate that the proposed model effectively solves a self-driving prediction problem by producing accurate predictions of road agents in various environments.

- Developed by Andrew Karpathy
- Has uber-feature: page thumbnails
- Also shows paper abstract

Where to look for papers on arXiv

arxiv-sanity.com

Where to look for papers on arXiv

arxiv-sanity.com

The screenshot shows the homepage of arxiv-sanity.com. The top navigation bar includes a user account section with "ducha.aiki@gmail.com" and "log out", and a GitHub fork button. Below the header is a search bar and a navigation menu with links: "most recent", "top recent", "top hype", "friends", "discussions", "recommended", and "library". A "Only show v1" button is also present. The main content area displays a single paper summary:

CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving
Geunseob Oh, Huei Peng
1/24/2022 cs.LG | cs.AI | cs.RO

The paper's abstract is partially visible at the bottom:

The task of predicting stochastic behaviors of road agents in diverse environments is a challenging problem for autonomous driving. To best understand scene contexts and produce diverse possible future states of the road agents adaptively in different environments, a prediction model should be probabilistic, multi-modal, context-driven, and general. We present Conditionalizing Variational AutoEncoders via Hypernetworks (CVAE-H); a conditional VAE that extensively leverages hypernetwork and performs generative tasks for high-dimensional problems like the prediction task. We first evaluate CVAE-H on simple generative experiments to show that CVAE-H is probabilistic, multi-modal, context-driven, and general. Then, we demonstrate that the proposed model effectively solves a self-driving prediction problem by producing accurate predictions of road agents in various environments.

Where to look for papers on arXiv

arxiv-sanity.com

The screenshot shows the homepage of arxiv-sanity.com. At the top, there's a red header bar with the text "Arxiv Sanity Preserver", "Built in spare time by @karpathy to accelerate research.", "Serving last 185571 papers from cs.[CV|CL|LG|AI|NE]/stat.ML", and "(New: Nov 27, 2021) Have a look at the new [arxiv-sanity-lite.com](#)". On the right of the header are "ducha.aiki@gmail.com", "log out", and a "Fork me on GitHub" button. Below the header is a search bar with a magnifying glass icon. Underneath the search bar is a navigation menu with buttons for "most recent", "top recent", "top hype", "friends", "discussions", "recommended", and "library". A "Only show v1" button is located below the menu. The main content area displays a list of recent Arxiv papers. One paper is highlighted: "CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving" by Geunseob Oh, Huei Peng, dated 1/24/2022, with categories cs.LG | cs.AI | cs.RO. The paper has a version identifier "2201.09874v1", a PDF link, and links to "show similar" and "discuss". Below the paper abstract, there's a green box containing a summary of the research.

Arxiv Sanity Preserver
Built in spare time by @karpathy to accelerate research.
Serving last 185571 papers from cs.[CV|CL|LG|AI|NE]/stat.ML
(New: Nov 27, 2021) Have a look at the new [arxiv-sanity-lite.com](#)

ducha.aiki@gmail.com log out Fork me on GitHub

most recent top recent top hype friends discussions recommended library

Only show v1

Showing most recent Arxiv papers:

CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving
Geunseob Oh, Huei Peng
1/24/2022 cs.LG | cs.AI | cs.RO

2201.09874v1 [pdf](#)
[show similar](#) | [discuss](#)

The task of predicting stochastic behaviors of road agents in diverse environments is a challenging problem for autonomous driving. To best understand scene contexts and produce diverse possible future states of the road agents adaptively in different environments, a prediction model should be probabilistic, multi-modal, context-driven, and general. We present Conditionalizing Variational AutoEncoders via Hypernetworks (CVAE-H); a conditional VAE that extensively leverages hypernetwork and performs generative tasks for high-dimensional problems like the prediction task. We first evaluate CVAE-H on simple generative experiments to show that CVAE-H is probabilistic, multi-modal, context-driven, and general. Then, we demonstrate that the proposed model effectively solves a self-driving prediction problem by producing accurate predictions of road agents in various environments.

- You can save paper to the library (only 500 last)

Where to look for papers on arXiv

arxiv-sanity.com

The screenshot shows the homepage of arxiv-sanity.com. At the top, there's a red header bar with the text "Arxiv Sanity Preserver", "Built in spare time by @karpathy to accelerate research.", "Serving last 185571 papers from cs.[CV|CL|LG|AI|NE]/stat.ML", and "(New: Nov 27, 2021) Have a look at the new [arxiv-sanity-lite.com](#)". On the right of the header are "ducha.aiki@gmail.com", "log out", and a "Fork me on GitHub" button. Below the header is a search bar with a magnifying glass icon. Underneath the search bar is a navigation bar with tabs: "most recent" (which is selected), "top recent", "top hype", "friends", "discussions", "recommended", and "library". To the right of the tabs is a button labeled "Only show v1". A blue banner below the tabs says "Showing most recent Arxiv papers:". Below the banner, a paper card is displayed for "CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving" by Geunseob Oh, Huei Peng. The card includes the arXiv ID "2201.09874v1", a "pdf" link, "show similar", "discuss", and a save icon. Below the card is a green summary box containing a brief description of the paper's purpose and methodology.

Arxiv Sanity Preserver
Built in spare time by @karpathy to accelerate research.
Serving last 185571 papers from cs.[CV|CL|LG|AI|NE]/stat.ML
(New: Nov 27, 2021) Have a look at the new [arxiv-sanity-lite.com](#)

ducha.aiki@gmail.com log out Fork me on GitHub

most recent top recent top hype friends discussions recommended library

Only show v1

Showing most recent Arxiv papers:

CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving
Geunseob Oh, Huei Peng
1/24/2022 cs.LG | cs.AI | cs.RO

2201.09874v1 pdf
[show similar](#) | [discuss](#)

The task of predicting stochastic behaviors of road agents in diverse environments is a challenging problem for autonomous driving. To best understand scene contexts and produce diverse possible future states of the road agents adaptively in different environments, a prediction model should be probabilistic, multi-modal, context-driven, and general. We present Conditionalizing Variational AutoEncoders via Hypernetworks (CVAE-H); a conditional VAE that extensively leverages hypernetwork and performs generative tasks for high-dimensional problems like the prediction task. We first evaluate CVAE-H on simple generative experiments to show that CVAE-H is probabilistic, multi-modal, context-driven, and general. Then, we demonstrate that the proposed model effectively solves a self-driving prediction problem by producing accurate predictions of road agents in various environments.

- You can save paper to the library (only 500 last)
- Based on papers you save, arxiv-sanity recommends you new papers

Where to look for papers on arXiv

arxiv-sanity.com



The screenshot shows the homepage of arxiv-sanity.com. At the top, there's a dark red header with the text "Arxiv Sanity Preserver", "Built in spare time by @karpathy to accelerate research.", "Serving last 185571 papers from cs.[CV|CL|LG|AI|NE]/stat.ML", and "(New: Nov 27, 2021) Have a look at the new [arxiv-sanity-lite.com](#)". On the right side of the header are "ducha.aiki@gmail.com" and "log out" buttons, and a "Fork me on GitHub" link. Below the header is a search bar with a magnifying glass icon. Underneath the search bar is a navigation menu with buttons for "most recent", "top recent", "top hype", "friends", "discussions", "recommended", and "library". The "library" button is circled in red. To the right of the menu is a "Only show v1" button. Below the menu, a message says "Showing most recent Arxiv papers:". There's a list of papers, starting with "CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving" by Geunseob Oh, Huei Peng, dated 1/24/2022, with categories cs.LG | cs.AI | cs.RO. The paper has a PDF link (2201.09874v1 pdf), "show similar", and "discuss" links, and a save icon. A large preview image of the paper's first page is shown below the title. A green box at the bottom contains a summary of the paper's abstract.

Arxiv Sanity Preserver
Built in spare time by @karpathy to accelerate research.
Serving last 185571 papers from cs.[CV|CL|LG|AI|NE]/stat.ML
(New: Nov 27, 2021) Have a look at the new [arxiv-sanity-lite.com](#)

ducha.aiki@gmail.com log out Fork me on GitHub

most recent top recent top hype friends discussions recommended library

Only show v1

Showing most recent Arxiv papers:

CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving
Geunseob Oh, Huei Peng
1/24/2022 cs.LG | cs.AI | cs.RO

2201.09874v1 pdf
[show similar](#) | [discuss](#)

The task of predicting stochastic behaviors of road agents in diverse environments is a challenging problem for autonomous driving. To best understand scene contexts and produce diverse possible future states of the road agents adaptively in different environments, a prediction model should be probabilistic, multi-modal, context-driven, and general. We present Conditionalizing Variational AutoEncoders via Hypernetworks (CVAE-H); a conditional VAE that extensively leverages hypernetwork and performs generative tasks for high-dimensional problems like the prediction task. We first evaluate CVAE-H on simple generative experiments to show that CVAE-H is probabilistic, multi-modal, context-driven, and general. Then, we demonstrate that the proposed model effectively solves a self-driving prediction problem by producing accurate predictions of road agents in various environments.

- You can save paper to the library (only 500 last)
- Based on papers you save, arxiv-sanity recommends you new papers

Where to look for papers on arXiv

arxiv-sanity.com

The screenshot shows the homepage of arxiv-sanity.com. At the top, there's a red header bar with the text "Arxiv Sanity Preserver", "Built in spare time by @karpathy to accelerate research.", "Serving last 185571 papers from cs.[CV|CL|LG|AI|NE]/stat.ML", "(New: Nov 27, 2021) Have a look at the new [arxiv-sanity-lite.com](#)", and user information "ducha.aiki@gmail.com" and "log out". A GitHub fork button is also present.

The main interface features a search bar and a navigation menu with buttons for "most recent", "top recent", "top hype", "friends", "discussions", "recommended", and "library".

A message indicates "500 papers in your library:". Below this, a specific paper is highlighted:

Fixing the train-test resolution discrepancy
Hugo Touvron, Andrea Vedaldi, Matthijs Douze, Hervé Jégou
1/20/2022 (v1: 6/14/2019) cs.CV | cs.LG

The paper summary states: "Data-augmentation is key to the training of neural networks for image classification. This paper first shows that existing augmentations induce a significant discrepancy between the typical size of the objects seen by the classifier at train and test time. We experimentally validate that, for a target test resolution, using a lower train resolution offers better classification at test time. We then propose a simple yet effective and efficient strategy to optimize the classifier performance when the train and test resolutions differ. It involves only a computationally cheap fine-tuning of the network at the test resolution. This enables training strong classifiers using small training images. For instance, we obtain 77.1% top-1 accuracy on ImageNet with a ResNet-50 trained on 128x128 images, and 79.8% with one trained on 224x224 image. In addition, if we use extra training data we get 82.5% with the ResNet-50 train with 224x224 images. Conversely, when training a ResNeXt-101 32x48d pre-trained in weakly-supervised fashion on 940 million public images at resolution 224x224 and further optimizing for test resolution 320x320, we obtain a test top-1 accuracy of 86.4% (top-5: 98.0%) (single-crop). To the best of our knowledge this is the highest ImageNet single-crop, top-1 and top-5 accuracy to date."

- You can save paper to the library (only 500 last)
- Based on papers you save, arxiv-sanity recommends you new papers

Where to look for papers on arXiv

arxiv-sanity.com

The screenshot shows the homepage of arxiv-sanity.com. At the top, there is a red header bar with the text "Arxiv Sanity Preserver", "Built in spare time by @karpathy to accelerate research.", "Serving last 185571 papers from cs.[CV|CL|LG|AI|NE]/stat.ML", "(New: Nov 27, 2021) Have a look at the new [arxiv-sanity-lite.com](#)", and user information "ducha.aiki@gmail.com" and "log out". A "Fork me on GitHub" button is also present. Below the header, there is a search bar and a navigation bar with buttons for "most recent", "top recent", "top hype", "friends", "discussions", "recommended" (which is circled in red), and "library". A message "500 papers in your library:" is displayed. Below this, a paper summary for "Fixing the train-test resolution discrepancy" by Hugo Touvron, Andrea Vedaldi, Matthijs Douze, Hervé Jégou is shown, along with a snippet of its content and download links. At the bottom, there is a green box containing a detailed description of the paper's content.

Arxiv Sanity Preserver
Built in spare time by @karpathy to accelerate research.
Serving last 185571 papers from cs.[CV|CL|LG|AI|NE]/stat.ML
(New: Nov 27, 2021) Have a look at the new [arxiv-sanity-lite.com](#)

ducha.aiki@gmail.com log out

Fork me on GitHub

most recent top recent top hype friends discussions **recommended** library

500 papers in your library:

Fixing the train-test resolution discrepancy
Hugo Touvron, Andrea Vedaldi, Matthijs Douze, Hervé Jégou
1/20/2022 (v1: 6/14/2019) cs.CV | cs.LG

1906.06423v4 [pdf](#)
[show similar](#) | [discuss](#)

Data-augmentation is key to the training of neural networks for image classification. This paper first shows that existing augmentations induce a significant discrepancy between the typical size of the objects seen by the classifier at train and test time. We experimentally validate that, for a target test resolution, using a lower train resolution offers better classification at test time. We then propose a simple yet effective and efficient strategy to optimize the classifier performance when the train and test resolutions differ. It involves only a computationally cheap fine-tuning of the network at the test resolution. This enables training strong classifiers using small training images. For instance, we obtain 77.1% top-1 accuracy on ImageNet with a ResNet-50 trained on 128x128 images, and 79.8% with one trained on 224x224 image. In addition, if we use extra training data we get 82.5% with the ResNet-50 train with 224x224 images. Conversely, when training a ResNeXt-101 32x48d pre-trained in weakly-supervised fashion on 940 million public images at resolution 224x224 and further optimizing for test resolution 320x320, we obtain a test top-1 accuracy of 86.4% (top-5: 98.0%) (single-crop). To the best of our knowledge this is the highest ImageNet single-crop, top-1 and top-5 accuracy to date.

- You can save paper to the library (only 500 last)
- Based on papers you save, arxiv-sanity recommends you new papers

Where to look for papers on arXiv

arxiv-sanity.com

The screenshot shows the homepage of arxiv-sanity.com. At the top, there's a red header bar with the text "Arxiv Sanity Preserver", "Built in spare time by @karpathy to accelerate research.", "Serving last 185571 papers from cs.[CV|CL|LG|AI||NE]/stat.ML", and "(New: Nov 27, 2021) Have a look at the new [arxiv-sanity-lite.com](#)". On the right of the header is an email input field "ducha.aiki@gmail.com" and a "log out" button. A "Fork me on GitHub" button is also visible. Below the header is a search bar with a magnifying glass icon. Underneath the search bar is a navigation menu with buttons: "most recent", "top recent", "top hype", "friends", "discussions", "recommended" (which is circled in red), and "library". Below the menu are filters: "Only show v1", "Last day", "Last 3 days", "Last week", "Last month", "Last year", and "All time". A message below the filters says "Recommended papers: (based on SVM trained on tfidf of papers in your library, refreshed every day or so)". There's a list of papers, starting with "Avoiding Overfitting: A Survey on Regularization Methods for Convolutional Neural Networks" by Claudio Filipi Gonçalves dos Santos, João Paulo Papa, dated 1/10/2022, in categories cs.CV | cs.LG, with 27 pages. The paper summary discusses regularization methods for CNNs, mentioning ResNet and EfficientNet. At the bottom of the page, there's a green box containing a detailed description of the paper's content.

Several image processing tasks, such as image classification and object detection, have been significantly improved using Convolutional Neural Networks (CNN). Like ResNet and EfficientNet, many architectures have achieved outstanding results in at least one dataset by the time of their creation. A critical factor in training concerns the network's regularization, which prevents the structure from overfitting. This work analyzes several regularization methods developed in the last few years, showing significant improvements for different CNN models. The works are classified into three main areas: the first one is called "data augmentation", where all the techniques focus on performing changes in the input data. The second, named "internal changes", which aims to describe procedures to modify the feature maps generated by the neural network or the kernels. The last one, called "label", concerns transforming the labels of a given input. This work presents two main differences comparing to other available surveys about regularization: (i) the first concerns the papers gathered in the manuscript, which are not older than five years, and (ii) the second distinction is about reproducibility, i.e., all works referred here have their code available in public repositories or they have been directly implemented in some framework.

- You can save paper to the library (only 500 last)
- Based on papers you save, arxiv-sanity recommends you new papers

The main problem with arxiv-sanity

It is often offline

The screenshot shows the Arxiv Sanity Preserver website. At the top, there's a red header bar with the text "Arxiv Sanity Preserver", "Built in spare time by @karpathy to accelerate research.", "Serving last 185571 papers from cs.[CV|CL|LG|AI|NE]/stat.ML", "(New: Nov 27, 2021) Have a look at the new [arxiv-sanity-lite.com](#)", and user information "ducha.aiki@gmail.com" and "log out". A "Fork me on GitHub" button is also present.

The main interface features a search bar with a magnifying glass icon. Below it is a navigation bar with buttons for "most recent", "top recent", "top hype", "friends", "discussions", "recommended", and "library". A "Only show v1" button is located below the navigation bar.

A message "Showing most recent Arxiv papers:" is displayed above a list of papers. The first paper shown is titled "CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving" by Geunseob Oh, Huei Peng, dated 1/24/2022, with categories cs.LG | cs.AI | cs.RO. The paper has a version number "2201.09874v1" and links to "pdf", "show similar", and "discuss". There's also a "Save" button.

The abstract of the paper is partially visible, describing the task of predicting stochastic behaviors of road agents in diverse environments for autonomous driving, mentioning the use of Conditionalizing Variational AutoEncoders via Hypernetworks (CVAE-H).

It is often offline :(

Where to look for papers on arXiv

arxiv-sanity-lite.com

arxiv-sanity about stats profile

Rank by: time tags: pid: time_filter (days): svm_c: 0.01
skip_have: no Submit

Shortcuts: [recommend over last week](#) [recommend over last 3 days](#) [recent](#) [random last week](#)

0.81
[CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving](#)
Geunseob Oh, Huei Peng
Jan 24 2022 cs.LG, cs.AI, cs.RO

+ -

The task of predicting stochastic behaviors of road agents in diverse environments is a challenging problem for autonomous driving. To best understand scene contexts and produce diverse possible future states of the road agents adaptively in different environments, a prediction model should be probabilistic, multi-modal, context-driven, and general. We present Conditionalizing Variational AutoEncoders via Hypernetworks (CVAE-H); a conditional VAE that extensively leverages hypernetwork and performs generative tasks for high-dimensional problems like the prediction task. We first evaluate CVAE-H on simple generative experiments to show that CVAE-H is probabilistic, multi-modal, context-driven, and general. Then, we demonstrate that the proposed model effectively solves a self-driving prediction problem by producing accurate predictions of road agents in various environments.

[similar](#) [inspect](#)

0.81 [Transformers in Medical Imaging: A Survey](#)
Fahad Shamshad, Salman Khan, Syed Waqas Zamir, Muhammad Haris Khan, Munawar Hayat, Fahad Shahbaz Khan, Huazhu Fu
Jan 24 2022 eess.IV, cs.CV

+ -

Following unprecedented success on the natural language tasks, Transformers have been successfully applied to several computer vision problems, achieving state-of-the-art results and prompting researchers to reconsider the supremacy of convolutional neural networks (CNNs) as {de facto} operators. Capitalizing on these advances in computer vision, the medical imaging field

Where to look for papers on arXiv

arxiv-sanity-lite.com

The screenshot shows a search interface with a red header bar containing the text "arxiv-sanity" and links for "about", "stats", and "profile". Below the header is a search bar with a magnifying glass icon. Underneath the search bar are several input fields: "Rank by:" (set to "time"), "tags:", "pid:", "time_filter (days):" (set to 0.01), and "svm_c: 0.01". There is also a dropdown for "skip_have" set to "no" and a "Submit" button. Below these controls are "Shortcuts" links: "recommend over last week", "recommend over last 3 days", "recent", and "random last week".

0.81
[CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving](#)
Geunseob Oh, Huei Peng
Jan 24 2022 cs.LG, cs.AI, cs.RO
+ -
The task of predicting stochastic behaviors of road agents in diverse environments is a challenging problem for autonomous driving. To best understand scene contexts and produce diverse possible future states of the road agents adaptively in different environments, a prediction model should be probabilistic, multi-modal, context-driven, and general. We present Conditionalizing Variational AutoEncoders via Hypernetworks (CVAE-H); a conditional VAE that extensively leverages hypernetwork and performs generative tasks for high-dimensional problems like the prediction task. We first evaluate CVAE-H on simple generative experiments to show that CVAE-H is probabilistic, multi-modal, context-driven, and general. Then, we demonstrate that the proposed model effectively solves a self-driving prediction problem by producing accurate predictions of road agents in various environments.
[similar](#) [inspect](#)

0.81 [Transformers in Medical Imaging: A Survey](#)
Fahad Shamshad, Salman Khan, Syed Waqas Zamir, Muhammad Haris Khan, Munawar Hayat, Fahad Shahbaz Khan, Huazhu Fu
Jan 24 2022 eess.IV, cs.CV
+ -
Following unprecedented success on the natural language tasks, Transformers have been successfully applied to several computer vision problems, achieving state-of-the-art results and prompting researchers to reconsider the supremacy of convolutional neural networks (CNNs) as {de facto} operators. Capitalizing on these advances in computer vision, the medical imaging field

- Re-vamp of arxiv-sanity

Where to look for papers on arXiv

arxiv-sanity-lite.com

The screenshot shows the homepage of arxiv-sanity-lite.com. At the top, there is a red navigation bar with the text "arxiv-sanity" on the left and "about stats profile" on the right. Below the navigation bar is a search bar with a magnifying glass icon. Underneath the search bar are several input fields: "Rank by: time" (with a dropdown arrow), "tags:" (text input), "pid:" (text input), "time_filter (days):" (input field with value 0.01), and "svm_c: 0.01". There is also a "skip_have: no" dropdown and a "Submit" button. Below these controls are "Shortcuts" links: "recommend over last week", "recommend over last 3 days", "recent", and "random last week". The main content area displays two paper cards. The first card has a score of "0.81" and the title "CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving" by "Geunseob Oh, Huei Peng" (date: Jan 24 2022, categories: cs.LG, cs.AI, cs.RO). It includes a summary: "The task of predicting stochastic behaviors of road agents in diverse environments is a challenging problem for autonomous driving. To best understand scene contexts and produce diverse possible future states of the road agents adaptively in different environments, a prediction model should be probabilistic, multi-modal, context-driven, and general. We present Conditionalizing Variational AutoEncoders via Hypernetworks (CVAE-H); a conditional VAE that extensively leverages hypernetwork and performs generative tasks for high-dimensional problems like the prediction task. We first evaluate CVAE-H on simple generative experiments to show that CVAE-H is probabilistic, multi-modal, context-driven, and general. Then, we demonstrate that the proposed model effectively solves a self-driving prediction problem by producing accurate predictions of road agents in various environments." Below the summary are links "similar" and "inspect". The second card has a score of "0.81" and the title "Transformers in Medical Imaging: A Survey" by "Fahad Shamshad, Salman Khan, Syed Waqas Zamir, Muhammad Haris Khan, Munawar Hayat, Fahad Shahbaz Khan, Huazhu Fu" (date: Jan 24 2022, categories: eess.IV, cs.CV). It includes a summary: "Following unprecedented success on the natural language tasks, Transformers have been successfully applied to several computer vision problems, achieving state-of-the-art results and prompting researchers to reconsider the supremacy of convolutional neural networks (CNNs) as {de facto} operators. Capitalizing on these advances in computer vision, the medical imaging field". Both cards feature a blue "+" button and a red "-" button at the bottom left.

- Re-vamp of arxiv-sanity
- No page thumbnails

Where to look for papers on arXiv

arxiv-sanity-lite.com

The screenshot shows the homepage of arxiv-sanity-lite.com. At the top, there is a red navigation bar with the text "arxiv-sanity" on the left and "about stats profile" on the right. Below the navigation bar is a search bar with a magnifying glass icon. Underneath the search bar are several input fields: "Rank by: time" (with a dropdown arrow), "tags:", "pid:", "time_filter (days):" (with a slider), and "svm_c: 0.01". There is also a "skip_have: no" dropdown and a "Submit" button. Below these controls are "Shortcuts" links: "recommend over last week", "recommend over last 3 days", "recent", and "random last week". The main content area displays two research papers as cards. The first card, titled "CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving" by Geunseob Oh, Huei Peng, dated Jan 24 2022, has a score of 0.81. It includes a summary of the paper's purpose and methodology, and buttons for "+", "-", "similar", and "inspect". The second card, titled "Transformers in Medical Imaging: A Survey" by Fahad Shamshad, Salman Khan, Syed Waqas Zamir, Muhammad Haris Khan, Munawar Hayat, Fahad Shahbaz Khan, Huazhu Fu, dated Jan 24 2022, also has a score of 0.81 and similar buttons.

- Re-vamp of arxiv-sanity
- No page thumbnails
- You can control recommendations from tf-idf SVM

Where to look for papers on arXiv

arxiv-sanity-lite.com

The screenshot shows the homepage of arxiv-sanity-lite.com. At the top, there is a navigation bar with links for "about", "stats", and "profile". Below the navigation bar is a search bar with a magnifying glass icon. Underneath the search bar are several input fields: "Rank by:" (set to "time"), "tags:", "pid:", "time_filter (days)" (with a value of "0.01" highlighted with a red oval), and "skip_have:" (set to "no"). There is also a "Submit" button and some "Shortcuts" links: "recommend over last week", "recommend over last 3 days", "recent", and "random last week".

The main content area displays two research papers:

0.81
[CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving](#)
Geunseob Oh, Huei Peng
Jan 24 2022 cs.LG, cs.AI, cs.RO
+ -
The task of predicting stochastic behaviors of road agents in diverse environments is a challenging problem for autonomous driving. To best understand scene contexts and produce diverse possible future states of the road agents adaptively in different environments, a prediction model should be probabilistic, multi-modal, context-driven, and general. We present Conditionalizing Variational AutoEncoders via Hypernetworks (CVAE-H); a conditional VAE that extensively leverages hypernetwork and performs generative tasks for high-dimensional problems like the prediction task. We first evaluate CVAE-H on simple generative experiments to show that CVAE-H is probabilistic, multi-modal, context-driven, and general. Then, we demonstrate that the proposed model effectively solves a self-driving prediction problem by producing accurate predictions of road agents in various environments.
[similar](#) [inspect](#)

0.81 [Transformers in Medical Imaging: A Survey](#)
Fahad Shamshad, Salman Khan, Syed Waqas Zamir, Muhammad Haris Khan, Munawar Hayat, Fahad Shahbaz Khan, Huazhu Fu
Jan 24 2022 eess.IV, cs.CV
+ -
Following unprecedented success on the natural language tasks, Transformers have been successfully applied to several computer vision problems, achieving state-of-the-art results and prompting researchers to reconsider the supremacy of convolutional neural networks (CNNs) as (de facto) operators. Capitalizing on these advances in computer vision, the medical imaging field

- Re-vamp of arxiv-sanity
- No page thumbnails
- You can control recommendations from tf-idf SVM

Where to look for papers on arXiv

arxiv-sanity-lite.com

The screenshot shows the homepage of arxiv-sanity-lite.com. At the top, there is a navigation bar with links for 'about', 'stats', and 'profile'. Below the navigation bar is a search bar with a magnifying glass icon. Underneath the search bar are several input fields: 'Rank by:' (set to 'time'), 'tags:', 'pid:', 'time_filter (days)', and 'svm_c: 0.01'. The 'time_filter (days)' field is circled in red. Below these fields are buttons for 'skip_have:' (set to 'no') and 'Submit'. A row of 'Shortcuts' follows: 'recommend over last week', 'recommend over last 3 days', 'recent', and 'random last week'. The main content area displays two research papers:

0.81
[CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving](#)
Geunseob Oh, Huei Peng
Jan 24 2022 cs.LG, cs.AI, cs.RO
+ -
The task of predicting stochastic behaviors of road agents in diverse environments is a challenging problem for autonomous driving. To best understand scene contexts and produce diverse possible future states of the road agents adaptively in different environments, a prediction model should be probabilistic, multi-modal, context-driven, and general. We present Conditionalizing Variational AutoEncoders via Hypernetworks (CVAE-H); a conditional VAE that extensively leverages hypernetwork and performs generative tasks for high-dimensional problems like the prediction task. We first evaluate CVAE-H on simple generative experiments to show that CVAE-H is probabilistic, multi-modal, context-driven, and general. Then, we demonstrate that the proposed model effectively solves a self-driving prediction problem by producing accurate predictions of road agents in various environments.
[similar](#) [inspect](#)

0.81 [Transformers in Medical Imaging: A Survey](#)
Fahad Shamshad, Salman Khan, Syed Waqas Zamir, Muhammad Haris Khan, Munawar Hayat, Fahad Shahbaz Khan, Huazhu Fu
Jan 24 2022 eess.IV, cs.CV
+ -
Following unprecedented success on the natural language tasks, Transformers have been successfully applied to several computer vision problems, achieving state-of-the-art results and prompting researchers to reconsider the supremacy of convolutional neural networks (CNNs) as {de facto} operators. Capitalizing on these advances in computer vision, the medical imaging field

- Re-vamp of arxiv-sanity
- No page thumbnails
- You can control recommendations from tf-idf SVM
- “Relevancy score”

Where to look for papers on arXiv

arxiv-sanity-lite.com

The screenshot shows a web interface for searching arXiv papers. At the top, there's a red header bar with the text "arxiv-sanity" on the left and "about stats profile" on the right. Below the header is a search bar with a magnifying glass icon. Underneath the search bar are several input fields: "Rank by: time" (with a dropdown arrow), "tags:" (text input), "pid:" (text input), "time_filter (days):" (text input), and "svm_c: 0.01". There are also buttons for "skip_have: no" and "Submit". Below these controls are "Shortcuts" links: "recommend over last week", "recommend over last 3 days", "recent", and "random last week". The main content area displays two search results. The first result is for a paper titled "CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving" by Geunseob Oh, Huei Peng. It was posted on Jan 24 2022 under categories cs.LG, cs.AI, and cs.RO. A relevancy score of 0.81 is highlighted with a red oval. The abstract discusses the challenge of predicting stochastic behaviors of road agents and introduces CVAE-H, a conditional VAE that leverages hypernetworks. The second result is for a paper titled "Transformers in Medical Imaging: A Survey" by Fahad Shamshad, Salman Khan, Syed Waqas Zamir, Muhammad Haris Khan, Munawar Hayat, Fahad Shahbaz Khan, and Huazhu Fu. It was posted on Jan 24 2022 under categories eess.IV and cs.CV. The abstract notes the success of Transformers in various natural language tasks and their application to computer vision problems.

- Re-vamp of arxiv-sanity
- No page thumbnails
- You can control recommendations from tf-idf SVM
- “Relevancy score”

Where to look for papers on arXiv

arxiv-sanity-lite.com

The screenshot shows a search interface with a red header bar containing the text "arxiv-sanity". Below the header are navigation links for "about", "stats", and "profile". A search bar with a magnifying glass icon is present. Underneath the search bar are filter options: "Rank by: time", "tags:", "pid:", "time_filter (days):", and "svm_c: 0.01". There is also a "skip_have: no" dropdown and a "Submit" button. Below these are "Shortcuts" links: "recommend over last week", "recommend over last 3 days", "recent", and "random last week". The main content area displays two search results:

0.81
[CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving](#)
Geunseob Oh, Huei Peng
Jan 24 2022 cs.LG, cs.AI, cs.RO
[+ -]
The task of predicting stochastic behaviors of road agents in diverse environments is a challenging problem for autonomous driving. To best understand scene contexts and produce diverse possible future states of the road agents adaptively in different environments, a prediction model should be probabilistic, multi-modal, context-driven, and general. We present Conditionalizing Variational AutoEncoders via Hypernetworks (CVAE-H); a conditional VAE that extensively leverages hypernetwork and performs generative tasks for high-dimensional problems like the prediction task. We first evaluate CVAE-H on simple generative experiments to show that CVAE-H is probabilistic, multi-modal, context-driven, and general. Then, we demonstrate that the proposed model effectively solves a self-driving prediction problem by producing accurate predictions of road agents in various environments.
[similar](#) [inspect](#)

0.81 [Transformers in Medical Imaging: A Survey](#)
Fahad Shamshad, Salman Khan, Syed Waqas Zamir, Muhammad Haris Khan, Munawar Hayat, Fahad Shahbaz Khan, Huazhu Fu
Jan 24 2022 eess.IV, cs.CV
[+ -]
Following unprecedented success on the natural language tasks, Transformers have been successfully applied to several computer vision problems, achieving state-of-the-art results and prompting researchers to reconsider the supremacy of convolutional neural networks (CNNs) as {de facto} operators. Capitalizing on these advances in computer vision, the medical imaging field

- Re-vamp of arxiv-sanity
- No page thumbnails
- You can control recommendations from tf-idf SVM
- “Relevancy score”
- “Inspect” the terms used for relevancy score calculation

Where to look for papers on arXiv

arxiv-sanity-lite.com

The screenshot shows a search interface with a red header bar containing the text "arxiv-sanity". Below the header are navigation links for "about", "stats", and "profile". A search bar with a magnifying glass icon is present. Underneath the search bar are filter options: "Rank by: time" (selected), "tags:", "pid:", "time_filter (days):" (set to 0.01), and "svm_c: 0.01". There is also a "skip_have: no" dropdown set to "no" and a "Submit" button. Below these filters are "Shortcuts" links: "recommend over last week", "recommend over last 3 days", "recent", and "random last week".

0.81
[CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving](#)
Geunseob Oh, Huei Peng
Jan 24 2022 cs.LG, cs.AI, cs.RO
+ -
The task of predicting stochastic behaviors of road agents in diverse environments is a challenging problem for autonomous driving. To best understand scene contexts and produce diverse possible future states of the road agents adaptively in different environments, a prediction model should be probabilistic, multi-modal, context-driven, and general. We present Conditionalizing Variational AutoEncoders via Hypernetworks (CVAE-H); a conditional VAE that extensively leverages hypernetwork and performs generative tasks for high-dimensional problems like the prediction task. We first evaluate CVAE-H on simple generative experiments to show that CVAE-H is probabilistic, multi-modal, context-driven, and general. Then, we demonstrate that the proposed model effectively solves a self-driving prediction problem by producing accurate predictions of road agents in various environments.
[similar](#) [inspect](#)

0.81 [Transformers in Medical Imaging: A Survey](#)
Fahad Shamshad, Salman Khan, Syed Waqas Zamir, Muhammad Haris Khan, Munawar Hayat, Fahad Shahbaz Khan, Huazhu Fu
Jan 24 2022 eess.IV, cs.CV
+ -
Following unprecedented success on the natural language tasks, Transformers have been successfully applied to several computer vision problems, achieving state-of-the-art results and prompting researchers to reconsider the supremacy of convolutional neural networks (CNNs) as {de facto} operators. Capitalizing on these advances in computer vision, the medical imaging field

- Re-vamp of arxiv-sanity
- No page thumbnails
- You can control recommendations from tf-idf SVM
- “Relevancy score”
- “Inspect” the terms used for relevancy score calculation

Where to look for papers on arXiv

arxiv-sanity-lite.com

arxiv-sanity

about stats profile

CVAE-H: Conditionalizing Variational Autoencoders via Hypernetworks and Trajectory Forecasting for Autonomous Driving

Geunseob Oh, Huei Peng

Jan 24 2022 cs.LG, cs.AI, cs.RO

The task of predicting stochastic behaviors of road agents in diverse environments is a challenging problem for autonomous driving. To best understand scene contexts and produce diverse possible future states of the road agents adaptively in different environments, a prediction model should be probabilistic, multi-modal, context-driven, and general. We present Conditionalizing Variational AutoEncoders via Hypernetworks (CVAE-H); a conditional VAE that extensively leverages hypernetwork and performs generative tasks for high-dimensional problems like the prediction task. We first evaluate CVAE-H on simple generative experiments to show that CVAE-H is probabilistic, multi-modal, context-driven, and general. Then, we demonstrate that the proposed model effectively solves a self-driving prediction problem by producing accurate predictions of road agents in various environments.

The following are the tokens and their (tfidf) weight in the paper vector. This is the actual summary that feeds into the SVM to power recommendations, so hopefully it is good and representative!

0.34 cvae 0.21 road 0.19 variational autoencoders 0.17 driving 0.17 autoencoders 0.16 agents
0.16 multi modal 0.16 autonomous driving 0.15 environments 0.15 trajectory forecasting
0.15 modal 0.15 hypernetwork 0.14 probabilistic 0.14 problems like 0.14 variational
0.14 dimensional problems 0.14 tasks high 0.14 task predicting 0.14 possible future
0.13 model effectively 0.13 prediction problem 0.13 different environments 0.13 autonomous
0.12 diverse 0.12 accurate predictions 0.12 driven 0.12 oh 0.12 generative 0.12 self driving
0.12 prediction task 0.11 prediction model 0.11 context 0.11 solves 0.11 vae 0.11 general
0.10 challenging problem 0.10 producing 0.10 contexts 0.10 adaptively 0.09 extensively
0.09 forecasting 0.09 behaviors 0.09 peng 0.09 demonstrate proposed 0.09 proposed model
0.09 trajectory 0.09 leverages 0.08 conditional 0.08 states 0.08 high dimensional
0.08 understand 0.08 predicting 0.08 scene 0.08 performs 0.08 produce 0.07 stochastic
0.07 predictions 0.07 effectively 0.07 dimensional 0.07 possible 0.07 future 0.07 like
0.07 accurate 0.06 self 0.06 best 0.06 simple 0.06 evaluate

Where to look for papers on arXiv

arxiv-sanity-lite.com

2.00 Learning and Crafting for the Wide Multiple Baseline Stereo

Dmytro Mishkin

Dec 22 2021 cs.CV



This thesis introduces the wide multiple baseline stereo (WxBS) problem. WxBS, a generalization of the standard wide baseline stereo problem, considers the matching of images that simultaneously differ in more than one image acquisition factor such as viewpoint, illumination, sensor type, or where object appearance changes significantly, e.g., over time. A new dataset with the ground truth, evaluation metric and baselines has been introduced. The thesis presents the following improvements of the WxBS pipeline. (i) A loss function, called HardNeg, for learning a local image descriptor that relies on hard negative mining within a mini-batch and on the maximization of the distance between the closest positive and the closest negative patches. (ii) The descriptor trained with the HardNeg loss, called HardNet, is compact and shows state-of-the-art performance in standard matching, patch verification and retrieval benchmarks. (iii) A method for learning the affine shape, orientation, and potentially other parameters related to geometric and appearance properties of local features. (iv) A tentative correspondences generation strategy which generalizes the standard first to second closest distance ratio is presented. The selection strategy, which shows performance superior to the standard method, is applicable to either hard-engineered descriptors like SIFT, LIOPI, and MROGH or deeply learned like HardNet. (v) A feedback loop is introduced for the two-view matching problem, resulting in MODS -- matching with on-demand view synthesis -- algorithm. MODS is an algorithm that handles a viewing angle difference even larger than the previous state-of-the-art ASIFT algorithm, without a significant increase of computational cost over "standard" wide and narrow baseline approaches. Last, but not least, a comprehensive benchmark for local features and robust estimation algorithms is introduced.

[similar](#) [inspect](#)

Where to look for papers on arXiv

arxiv-sanity-lite.com

2.00 Learning and Crafting for the Wide Multiple Baseline Stereo

Dmytro Mishkin

Dec 22 2021 cs.CV



This thesis introduces the wide multiple baseline stereo (WxBS) problem. WxBS, a generalization of the standard wide baseline stereo problem, considers the matching of images that simultaneously differ in more than one image acquisition factor such as viewpoint, illumination, sensor type, or where object appearance changes significantly, e.g., over time. A new dataset with the ground truth, evaluation metric and baselines has been introduced. The thesis presents the following improvements of the WxBS pipeline. (i) A loss function, called HardNeg, for learning a local image descriptor that relies on hard negative mining within a mini-batch and on the maximization of the distance between the closest positive and the closest negative patches. (ii) The descriptor trained with the HardNeg loss, called HardNet, is compact and shows state-of-the-art performance in standard matching, patch verification and retrieval benchmarks. (iii) A method for learning the affine shape, orientation, and potentially other parameters related to geometric and appearance properties of local features. (iv) A tentative correspondences generation strategy which generalizes the standard first to second closest distance ratio is presented. The selection strategy, which shows performance superior to the standard method, is applicable to either hard-engineered descriptors like SIFT, LIOPI, and MROGH or deeply learned like HardNet. (v) A feedback loop is introduced for the two-view matching problem, resulting in MODS -- matching with on-demand view synthesis -- algorithm. MODS is an algorithm that handles a viewing angle difference even larger than the previous state-of-the-art ASIFT algorithm, without a significant increase of computational cost over "standard" wide and narrow baseline approaches. Last, but not least, a comprehensive benchmark for local features and robust estimation algorithms is introduced.

[similar](#) [inspect](#)

- Similarity search by tf-idf

Where to look for papers on arXiv

arxiv-sanity-lite.com

2.00 Learning and Crafting for the Wide Multiple Baseline Stereo

Dmytro Mishkin

Dec 22 2021 cs.CV



This thesis introduces the wide multiple baseline stereo (WxBS) problem. WxBS, a generalization of the standard wide baseline stereo problem, considers the matching of images that simultaneously differ in more than one image acquisition factor such as viewpoint, illumination, sensor type, or where object appearance changes significantly, e.g., over time. A new dataset with the ground truth, evaluation metric and baselines has been introduced. The thesis presents the following improvements of the WxBS pipeline. (i) A loss function, called HardNeg, for learning a local image descriptor that relies on hard negative mining within a mini-batch and on the maximization of the distance between the closest positive and the closest negative patches. (ii) The descriptor trained with the HardNeg loss, called HardNet, is compact and shows state-of-the-art performance in standard matching, patch verification and retrieval benchmarks. (iii) A method for learning the affine shape, orientation, and potentially other parameters related to geometric and appearance properties of local features. (iv) A tentative correspondences generation strategy which generalizes the standard first to second closest distance ratio is presented. The selection strategy, which shows performance superior to the standard method, is applicable to either hard-engineered descriptors like SIFT, LIOF, and MROGH or deeply learned like HardNet. (v) A feedback loop is introduced for the two-view matching problem, resulting in MODS -- matching with on-demand view synthesis -- algorithm. MODS is an algorithm that handles a viewing angle difference even larger than the previous state-of-the-art ASIFT algorithm, without a significant increase of computational cost over "standard" wide and narrow baseline approaches. Last, but not least, a comprehensive benchmark for local features and robust estimation algorithms is introduced.

[similar](#) [inspect](#)

- Similarity search by tf-idf

Where to look for papers on arXiv

arxiv-sanity-lite.com

2.00 [Learning and Crafting for the Wide Multiple Baseline Stereo](#)

Dmytro Mishkin

Dec 22 2021 cs.CV



This thesis introduces the wide multiple baseline stereo (WxBS) problem. WxBS, a generalization of the standard wide baseline stereo problem, considers the matching of images that simultaneously differ in more than one image acquisition factor such as viewpoint, illumination, sensor type, or where object appearance changes significantly, e.g., over time. A new dataset with the ground truth, evaluation metric and baselines has been introduced. The thesis presents the following improvements of the WxBS pipeline. (i) A loss function, called HardNeg, for learning a local image descriptor that relies on hard negative mining within a mini-batch and on the maximization of the distance between the closest positive and the closest negative patches. (ii) The descriptor trained with the HardNeg loss, called HardNet, is compact and shows state-of-the-art performance in standard matching, patch verification and retrieval benchmarks. (iii) A method for learning the affine shape, orientation, and potentially other parameters related to geometric and appearance properties of local features. (iv) A tentative correspondences generation strategy which generalizes the standard first to second closest distance ratio is presented. The selection strategy, which shows performance superior to the standard method, is applicable to either hard-engineered descriptors like SIFT, LIOPI, and MROGH or deeply learned like HardNet. (v) A feedback loop is introduced for the two-view matching problem, resulting in MODS -- matching with on-demand view synthesis -- algorithm. MODS is an algorithm that handles a viewing angle difference even larger than the previous state-of-the-art ASIFT algorithm, without a significant increase of computational cost over "standard" wide and narrow baseline approaches. Last, but not least, a comprehensive benchmark for local features and robust estimation algorithms is introduced.

[similar](#) [inspect](#)

- Similarity search by tf-idf

-69.66 [Digging Into Self-Supervised Learning of Feature Descriptors](#)

Iaroslav Melekhov, Zakaria Laskar, Xiaotian Li, Shuzhe Wang, Juho Kannala

Oct 10 2021 cs.CV



Fully-supervised CNN-based approaches for learning local image descriptors have shown remarkable results in a wide range of geometric tasks. However, most of them require per-pixel ground-truth keypoint correspondence data which is difficult to acquire at scale. To address this challenge, recent weakly- and self-supervised methods can learn feature descriptors from relative camera poses or using only synthetic rigid transformations such as homographies. In this work, we focus on understanding the limitations of existing self-supervised approaches and propose a set of improvements that combined lead to powerful feature descriptors. We show that increasing the search space from in-pair to in-batch for hard negative mining brings consistent improvement. To enhance the discriminativeness of feature descriptors, we propose a coarse-to-fine method for mining local hard negatives from a wider search space by using global visual image descriptors. We demonstrate that a combination of synthetic homography transformation, color augmentation, and photorealistic image stylization produces useful representations that are viewpoint and illumination invariant. The feature descriptors learned by the proposed approach perform competitively and surpass their fully- and weakly-supervised counterparts on various geometric benchmarks such as image-based localization, sparse feature matching, and image retrieval.

[similar](#) [inspect](#)

-72.01 [SIFT Matching by Context Exposed](#)

Fabio Bellavia

Aug 27 2021 cs.CV



This paper investigates how to step up local image descriptor matching by exploiting matching context information. Two main contexts are identified, originated respectively from the descriptor space and from the keypoint space. The former is generally used to design the actual matching strategy while the latter to filter matches according to the local spatial consistency. On this basis, a new matching strategy and a novel local spatial filter, named respectively blob matching and Delaunay Triangulation Matching (DTM) are devised. Blob matching provides a general matching framework by merging together several strategies, including rank-based pre-filtering as well as many-to-many and symmetric matching, enabling to achieve a global improvement upon each individual strategy. DTM alternates between Delaunay triangulation contractions and expansions to figure out and adjust keypoint neighborhood consistency. Experimental evaluation shows that DTM is comparable or better than the state-of-the-art in terms of matching accuracy and robustness. Evaluation is carried out according to a new benchmark devised for analyzing the matching pipeline in terms of correct correspondences on both planar and non-planar scenes, including several state-of-the-art methods as well as the common SIFT matching approach for reference. This evaluation can be of assistance for future research in this field.

[similar](#) [inspect](#)

How to read arXiv papers

Coarse-to-fine

How to read arXiv papers

Coarse-to-fine

- I skim through arxiv-sanity list (or [arXiv.org CS.CV](https://arxiv.org/CS.CV)) almost daily

How to read arXiv papers

Coarse-to-fine

- I skim through arxiv-sanity list (or [arXiv.org CS.CV](https://arxiv.org/CS.CV)) almost daily
- Check all titles

How to read arXiv papers

Coarse-to-fine

- I skim through arxiv-sanity list (or arXiv.org CS.CV) almost daily
- Check all titles
- Open in the new tab to check the abstract

How to read arXiv papers

Coarse-to-fine

- I skim through arxiv-sanity list (or arXiv.org CS.CV) almost daily
- Check all titles
- Open in the new tab to check the abstract
- Based on abstract open from 1 to 10 pdfs

Should read this paper?

Questions I ask myself (1)

Should read this paper?

Questions I ask myself (1)

- My areas? Yes → **read**.

Should read this paper?

Questions I ask myself (1)

- My areas? Yes → **read**.
- Opens a new (sub-)area of research? Yes → **read**. E.g. GAN in 2014, NERF

Should read this paper?

Questions I ask myself (1)

- My areas? Yes → **read**.
- Opens a new (sub-)area of research? Yes → **read**. E.g. GAN in 2014, NERF
- From the crowded area? Yes → **skip**. E.g GANs now

Should read this paper?

Questions I ask myself (1)

- My areas? Yes → **read**.
- Opens a new (sub-)area of research? Yes → **read**. E.g. GAN in 2014, NERF
- From the crowded area? Yes → **skip**. E.g GANs now
- Dataset or large-scale benchmark paper? Yes → **read**

Should read this paper?

Questions I ask myself (1)

- My areas? Yes → **read**.
- Opens a new (sub-)area of research? Yes → **read**. E.g. GAN in 2014, NERF
- From the crowded area? Yes → **skip**. E.g GANs now
- Dataset or large-scale benchmark paper? Yes → **read**
- Simple baseline? Yes → **read**, regardless of the area.

Should read this paper?

Questions I ask myself (1)

- My areas? Yes → **read**.
- Opens a new (sub-)area of research? Yes → **read**. E.g. GAN in 2014, NERF
- From the crowded area? Yes → **skip**. E.g GANs now
- Dataset or large-scale benchmark paper? Yes → **read**
- Simple baseline? Yes → **read**, regardless of the area.
- About understanding some aspect of machine learning? Yes → **read if have time**. E.g. padding, double descent, over-parametrization

Should read this paper?

Questions I ask myself (2)

Should read this paper?

Questions I ask myself (2)

- Theory paper? Yes → **skip**

Should read this paper?

Questions I ask myself (2)

- Theory paper? Yes → **skip**
- Relevant for me as a user? Yes → **read if have time.** E.g. new non-linearity, optimizer, etc.

Should read this paper?

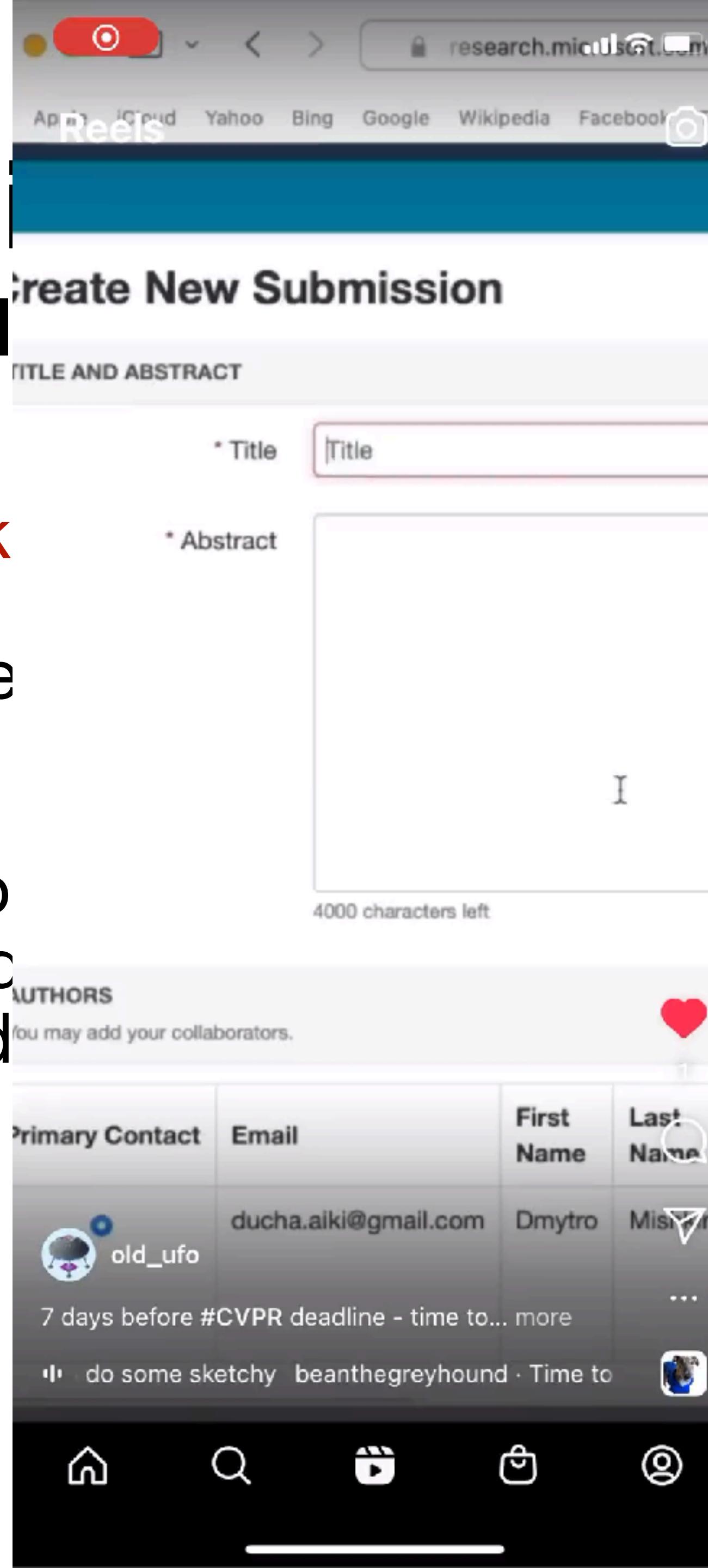
Questions I ask myself (2)

- Theory paper? Yes → **skip**
- Relevant for me as a user? Yes → **read if have time.** E.g. new non-linearity, optimizer, etc.
- Isn't the title over-keyworded and so on? E.g. "Adaptive DropBlock Enhanced Generative Adversarial Networks for Hyperspectral Image Classification". This kind of things are hard to verbalize.

Should read this

Questions I ask myself

- Theory paper? Yes → skip
- Relevant for me as a user optimizer, etc.
- Isn't the title over-keywordy? Enhanced Generative Adversarial Classification". This kind



me. E.g. new non-linearity,

Adaptive DropBlock
Hyperspectral Image
Segmentation.

Time for live example

Additional paper sources

Yannic Kilcher

The screenshot shows the YouTube channel page for 'Yannic Kilcher' with 118K subscribers. The 'VIDEOS' tab is selected. The channel features a profile picture of a dog and a red 'SUBSCRIBE' button.

Uploads ▾ SORT BY

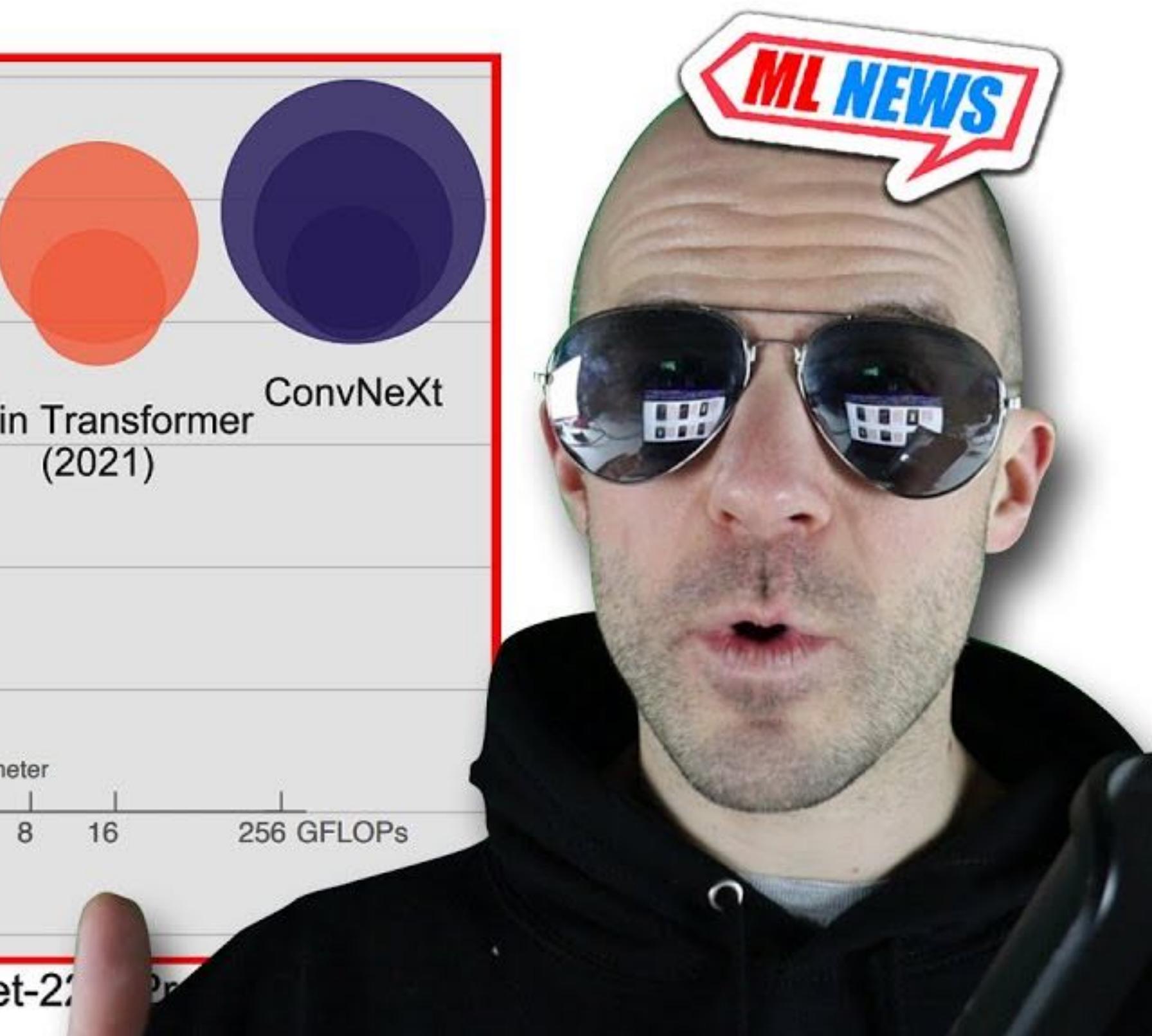
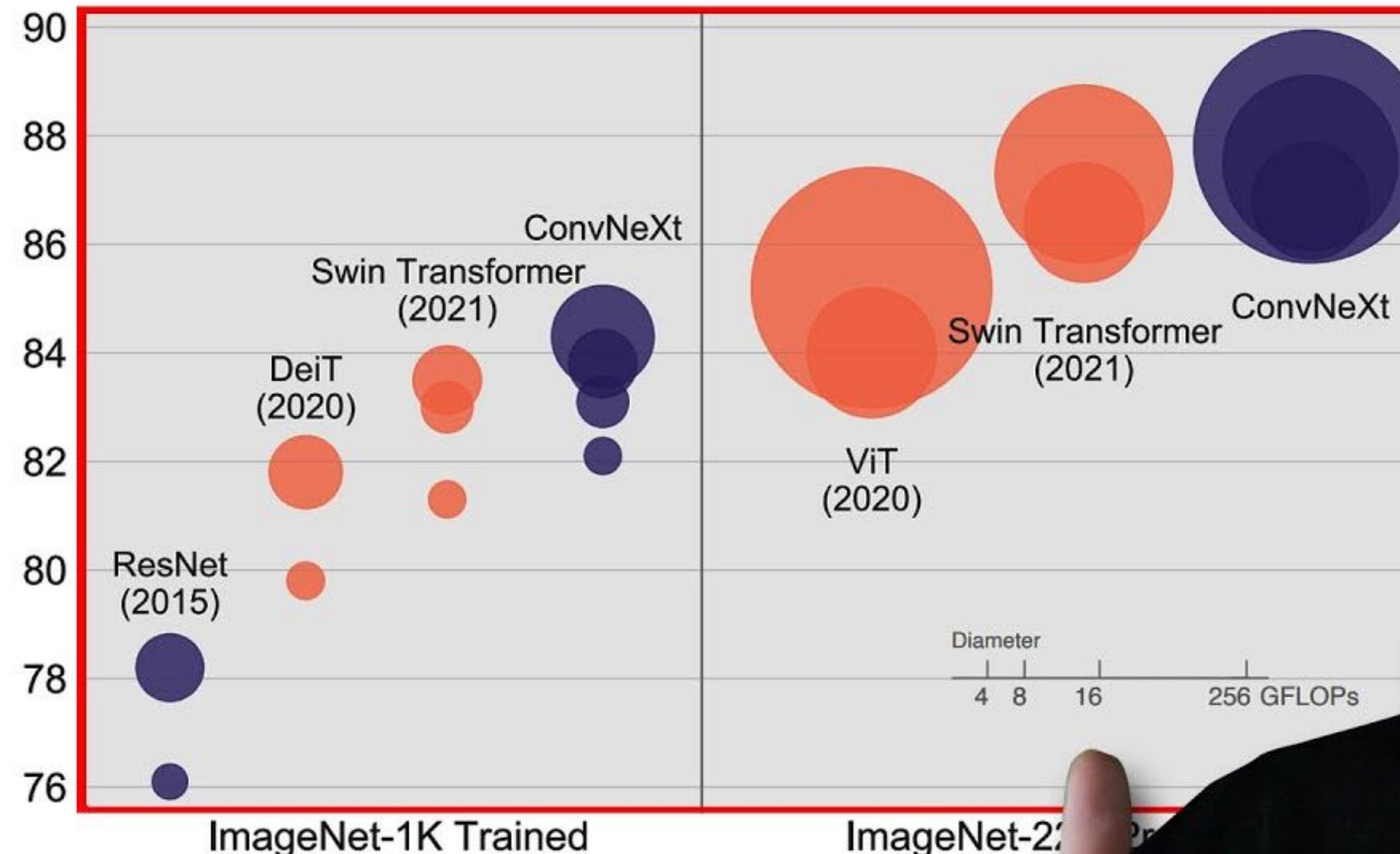
Thumbnail	Title	Views	Last Updated
	[ML News] ConvNeXt: Convolutions return China...	212 views	7 minutes ago
	Dynamic Inference with Neural Interpreters (w/...)	8.1K views	3 days ago
	Noether Networks: Meta-Learning Useful Conserved...	8.1K views	5 days ago
	This Team won the Minecraft RL BASALT Challenge!...	10K views	13 days ago
	Tesla FSD maybe next yea	41.5K views	41:51
	Player Of Games	14K views	54:11
	ML News LIVE STREAM On Air...	9.1K views	46:56
	AI generated Amazing Pictures!	13K views	42:17

Additional paper sources

Yannic Kilcher

Return of the Convolutions

ImageNet-1K Acc.



Twitter AK feed – general purpose interesting

@ak92501

AK
@ak92501

paper tweets, dms are open
Перекласти відомості про себе

Дата приєднання: квітень 2014

1 208 Ви читаете 45,9 тис. Читачі

Читають Нуен Соо Пак, Уфук Ефе і ще 134 з-поміж тих, хого ви читаете

Твіти Твіти й відповіді Медіафайли Вподобання

AK @ak92501 · 8 год
AlphaFold Accelerates Artificial Intelligence Powered Drug Discovery: Efficient Discovery of a Novel Cyclin-dependent Kinase 20 (CDK20) Small Molecule Inhibitor
abs: arxiv.org/abs/2201.09647

PandaOmics
Branches data analysis and visualization features:

- Compute z-scores for CMICs data
- Analyze OMICs data type
- Compute and evaluate novel drug targets
- Use new strategies to drug prioritizing

Chemistry42
Explore uncharted chemical spaces:

- Experiment the power of an automated and rapid machine learning platform
- Discover new drug targets from the analysis of internal datasets
- Generate structures and predict novel drug strategies

DARK TARGETS
With no reported crystal structure or small-molecule inhibitors

Molecule Synthesis

Bioassay Testing

Figure 1. Our approaches to combine AlphaFold with Insilico medicine end-to-end, AI-powered drug discovery platform PandaOmics and Chemistry42 in the drug discovery for hepatocellular carcinoma from target selection, hit generation to hit identification.

Twitter AK feed – general purpose interesting

@ak92501



AK
@ak92501
paper tweets, dms are open
Перекласти відомості про себе
Дата приєднання: квітень 2014
1 208 Ви читаєте **45,9 тис.** Читачі
Читають Hyun Soo Park, Ufuk Efe і ще 134 з-поміж тих, хого ви читаєте

Твіти **Твіти й відповіді** **Медіафайли** **Вподобання**

AK @ak92501 · 8 год
AlphaFold Accelerates Artificial Intelligence Powered Drug Discovery:
Efficient Discovery of a Novel Cyclin-dependent Kinase 20 (CDK20) Small
Molecule Inhibitor
abs: arxiv.org/abs/2201.09647

Figure 1. Our approaches to combine AlphaFold with In silico medicine end-to-end, AI-powered drug discovery platform PandaOmics and Chemistry42 in the drug discovery for hepatocellular carcinoma from target selection, hit generation to hit identification.

AK @ak92501 · 9 год
Description-Driven Task-Oriented Dialog Modeling
abs: arxiv.org/abs/2201.08904

Model	Pretrain. Model (# Params.)	MW2.1	MW2.2	MW2.3	MW2.4
Transformer-DST (Zeng and Nie, 2021)	BERT Base (110M)	55.35	-	-	-
SOM-DST (Kim et al., 2020)	BERT Base (110M)	51.2	-	55.5	66.8
TripPy (Heck et al., 2020)	BERT Base (110M)	55.3	-	63.0	59.6
SAVN (Wang et al., 2020)	BERT Base (110M)	54.5	-	58.0	60.1
SimpleTOD★ (Hosseini-Asl et al., 2020)	DistilGPT-2 (82M)	50.3/55.7	-	51.3	-
Seq2seq (Zhao et al., 2021)	T5 Base (220M)	52.8	57.6	59.3	67.1
DaP (seq) (Lee et al., 2021a)	T5 Base (220M)	-	51.2	-	-
DaP (ind) (Lee et al., 2021a)	T5 Base (220M)	56.7	57.6	-	-
D3ST (Base)	T5 Base (220M)	54.2	56.1	59.1	72.1
D3ST (Large)	T5 Large (770M)	54.5	54.2	58.6	70.8
D3ST (XXL)	T5 XXL (11B)	57.8	58.7	60.8	75.9

(a) JGA on MultiWOZ 2.1-2.4.

Model	Pretrain. Model (# Params.)	JGA	Intent	Req slot
SGD baseline (Rastogi et al., 2020)	BERT Base (110M)	25.4	90.6	96.5
DaP (ind) (Lee et al., 2021a)	T5 Base (220M)	71.8	90.2	97.8
SGP-DST (Ruan et al., 2020)	T5 Base (220M)	72.2	91.8	99.0
paDST■ (Ma et al., 2020)	XLNet Large (340M)	86.5	94.8	98.5
D3ST (Base)	T5 Base (220M)	72.9	97.2	98.9
D3ST (Large)	T5 Large (770M)	80.0	97.1	99.1
D3ST (XXL)	T5 XXL (11B)	86.4	98.8	99.4

(b) JGA, active intent accuracy and requested slot F1 on SGD.

Table 1: Results on MultiWOZ and SGD datasets with full training data. “-” indicates no public number is available. Best results are marked in bold. ★: SimpleTOD results are retrieved from the 2.3 website <https://github.com/lexmen318/MultiWOZ-coref>, in which two numbers are reported for 2.1 (one produced by the 2.3 author, the other by the original SimpleTOD paper). ■: No data pre-processing applied for MultiWOZ 2.1. ■: Data augmentation and special rules applied.



Twitter my feed: Image matching & hype

@ducha_aiki

Dmytro Mishkin
@ducha_aiki

Marrying classical CV and Deep Learning. I do things, which work, rather than being novel, but not working.

Перекласти відомості про себе

Ukraine ♂ dmytro.ai 📅 Дата приєднання: липень 2017

441 Ви читаете 5 852 Читачі

Твіти Твіти й відповіді Медіафайли Вподобання

Закріплений твіт

Dmytro Mishkin @ducha_aiki · 13 жовт. 2020

ArXiving before submission :

- helps professional identity building;
-protects against idea re-discovery/theft/gate-keeping
- facilitates open research distribution
- reduces inequality.

our refined statement, with @amy_tabb & Jiri Matas, now on arXiv

arXiv.org
ArXiving Before Submission Helps Everyone
We claim, and present evidence, that allowing arXiv publication before a conference or journal ...

9 44 160

Twitter my feed: Image matching & hype

@ducha_aiki

Dmytro Mishkin
@ducha_aiki

Marrying classical CV and Deep Learning. I do things, which work, rather than being novel, but not working.

Перекласти відомості про себе

Ukraine ♂ dmytro.ai 📅 Дата приєднання: липень 2017

441 Ви читаєте 5 852 Читачі

Твіти Твіти й відповіді Медіафайли Вподобання

Закріплений твіт

Dmytro Mishkin @ducha_aiki · 13 жовт. 2020 ArXiving before submission :

- helps professional identity building;
- protects against idea re-discovery/theft/gate-keeping
- facilitates open research distribution
- reduces inequality.

our refined statement, with @amy_tabb & Jiri Matas, now on arXiv

arXiv.org ArXiving Before Submission Helps Everyone We claim, and present evidence, that allowing arXiv publication before a conference or journal ...

9 44 160

Dmytro Mishkin @ducha_aiki · 7 січ. Towards realistic symmetry-based completion of previously unseen point clouds

Taras Rumezhak, @dobosevych, Rostyslav Hryniw, Vladyslav Selotkin, Volodymyr Karpiv, Mykola Maksymenko, @ucu_apps

tl;dr: good handcrafted method for point cloud completion

arxiv.org/abs/2201.01858

The diagram shows a flowchart of the point cloud completion process:

- Project the directions of normals on the unit sphere
- Perform PCA on the normal sphere
- Both first of principal components are symmetry planes
- Compute distances for
- Perform global registration
- Perform ICP
- Direct matching
- Fill the holes in the

Figure b) shows various 3D models with green arrows indicating normal directions. Figure c) shows the same models after processing. Figure d) shows a graph of CD × 10 vs Damage rate (%). The graph compares three methods: Ours (red), PCN (green), and Min-Nets (blue). The Ours method shows significantly lower CD values across all damage rates compared to the others.

1 7 15

Dmytro Mishkin @ducha_aiki · 7 січ. Bio-inspired Min-Nets Improve the Performance and Robustness of Deep Networks

Philipp Grüning, Erhardt Barth

tl;dr: $\text{out} = \min(\text{ReLU}(\text{convA}(\text{BN}(x))), \text{ReLU}(\text{convB}(\text{BN}((x))))$

Idea: both filters should fire for reliable detection, kind of opposite of maxout idea.

arxiv.org/abs/2201.02149...

The diagram illustrates a neural network architecture:

- Input x is processed by "Std Norm BN".
- The output is split into two parallel paths:
 - Path 1: "ReLU" followed by "ConvA(BN(x))".
 - Path 2: "ReLU" followed by "ConvB(BN((x)))".
- The outputs of both paths are combined via "Min" and then passed through "ReLU" and "BN+Scale".
- The final output is labeled T_2 .

**Register on twitter and follow
relevant people**

ResearchGate

way to share and request paywalled papers

The screenshot shows the ResearchGate homepage with two main research items displayed.

Top Navigation: Home, Questions, Jobs, Search bar: "Search for researchers, publications, and more", and a user profile icon.

Suggested research based on your interests:

Article 1: **Self-Supervised Keypoint Detection Based on Multi-Layer Random Forest Regressor** (March 2021 · IEEE Access · 90 Reads · 2 Citations)

Authors: Sangwon Kim, Mira Jeong, Byoungchul Ko

Actions: Download, Save, Follow, Recommend, Share

2 researchers recommend this research item.

Image: A diagram illustrating the self-supervised keypoint detection process using a multi-layer random forest regressor.

Article 2: **Learning Cross-Domain Descriptors for 2D-3D Matching with Hard Triplet Loss and Spatial Transformer Network** (September 2021 · 30 Reads)

Authors: Baiqi Lai, Weiquan Liu, Cheng Wang, Ming Cheng

Actions: Request full-text, Save, Follow, Recommend, Share

Jobs you may be interested in:

- University Professor (Salary Scale W3 NBesO) "Computer Engineering" (m/f/d)** at Technische Universität Clausthal
- Graduate teaching & research assistant, Graph neural networks** at University of Antwerp
- Research Officer in Behavioural and Attitudinal Change in respect of Ireland's Water Resources** at Economic and Social Research Institute

Google Scholar recommendations

Google Академія

Рекомендовані статті

Q-ViT: Fully Differentiable Quantization for Vision Transformer
Z Li, T Yang, P Wang, J Cheng
arXiv preprint arXiv:2201.07703 - 5 днів тому PDF

Poseur: Direct Human Pose Regression with Transformers
W Mao, Y Ge, C Shen, Z Tian, X Wang, Z Wang, A Hengel
arXiv preprint arXiv:2201.07412 - 5 днів тому PDF

Інші статті, опубліковані 5 днів тому

A graph-matching approach for cross-view registration of over-view and street-view based point clouds
X Ling, R Qin
ISPRS Journal of Photogrammetry and Remote Sensi... - 6 днів тому HTML

Google Scholar recommendations

Google Академія

The screenshot shows the Google Scholar interface. On the left, there's a sidebar titled "Рекомендовані статті" (Recommended articles) with three entries:

- Q-ViT: Fully Differentiable Quantization for Vision Transformer by Z Li, T Yang, P Wang, J Cheng. arXiv preprint arXiv:2201.07703 - 5 днів тому. PDF link.
- Poseur: Direct Human Pose Regression with Transformers by W Mao, Y Ge, C Shen, Z Tian, X Wang, Z Wang, A Hengel. arXiv preprint arXiv:2201.07412 - 5 днів тому. PDF link.
- A graph-matching approach for cross-view registration of over-view and street-view based point clouds by X Ling, R Qin. ISPRS Journal of Photogrammetry and Remote Sensi... - 6 днів тому. HTML link.

Below the sidebar, there's a blue button labeled "Інші статті, опубліковані 5 днів тому" (Other articles published within the last 5 days).

The main content area has a search bar with a magnifying glass icon and a menu icon. To the right of the search bar, it says "Сповіщення" (Notifications). Below this, it shows "Сповіщення для ducha.aiki@gmail.com" (Notifications for ducha.aiki@gmail.com) with two items:

- Andrej Karpathy – нові статті (Andrej Karpathy – new articles)
- "Learning discriminative affine regions via discriminability" – нові бібліографічні посилання (Learning discriminative affine regions via discriminability – new bibliographic links)
- "Mods: Fast and robust method for two-view matching" – нові бібліографічні посилання (Mods: Fast and robust method for two-view matching – new bibliographic links)

How to read a method paper?

The reading order of the method paper

Working hard to know your neighbor's margins: Local descriptor learning loss

The reading order of the method paper

Working hard to know your neighbor's margins: Local descriptor learning loss

- Abstract

The reading order of the method paper

Working hard to know your neighbor's margins: Local descriptor learning loss

- Abstract

Abstract

We introduce a loss for metric learning, which is inspired by the Lowe's matching criterion for SIFT. We show that the proposed loss, that maximizes the distance between the closest positive and closest negative example in the batch, is better than complex regularization methods; it works well for both shallow and deep convolution network architectures. Applying the novel loss to the L2Net CNN architecture results in a compact descriptor named HardNet. It has the same dimensionality as SIFT (128) and shows state-of-art performance in wide baseline stereo, patch verification and instance retrieval benchmarks.

The reading order of the method paper

Working hard to know your neighbor's margins: Local descriptor learning loss

- Abstract
- Method figure

Abstract

We introduce a loss for metric learning, which is inspired by the Lowe's matching criterion for SIFT. We show that the proposed loss, that maximizes the distance between the closest positive and closest negative example in the batch, is better than complex regularization methods; it works well for both shallow and deep convolution network architectures. Applying the novel loss to the L2Net CNN architecture results in a compact descriptor named HardNet. It has the same dimensionality as SIFT (128) and shows state-of-art performance in wide baseline stereo, patch verification and instance retrieval benchmarks.

The reading order of the method paper

Working hard to know your neighbor's margins: Local descriptor learning loss

- Abstract
- Method figure

Abstract

We introduce a loss for metric learning, which is inspired by the Lowe's matching criterion for SIFT. We show that the proposed loss, that maximizes the distance between the closest positive and closest negative example in the batch, is better than complex regularization methods; it works well for both shallow and deep convolution network architectures. Applying the novel loss to the L2Net CNN architecture results in a compact descriptor named HardNet. It has the same dimensionality as SIFT (128) and shows state-of-art performance in wide baseline stereo, patch verification and instance retrieval benchmarks.

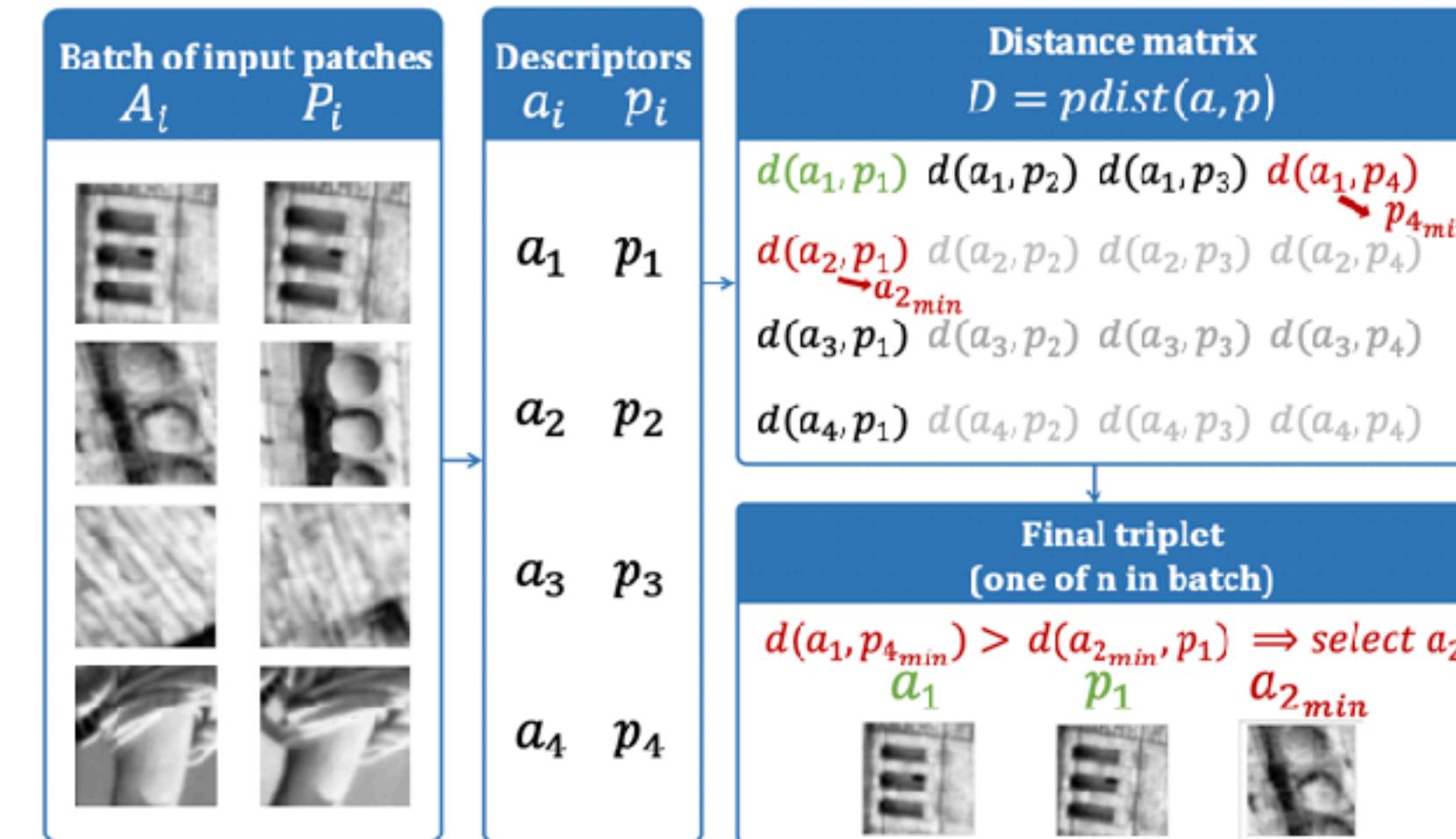


Figure 1: Proposed sampling procedure. First, patches are described by the current network, then a distance matrix is calculated. The closest non-matching descriptor – shown in red – is selected for each a_i and p_i patch from positive pair (green) respectively. Finally, among two negative candidates the hardest one is chosen. All operations are done in a single forward pass.

The reading order of the method paper

Working hard to know your neighbor's margins: Local descriptor learning loss

- Abstract
- Method figure

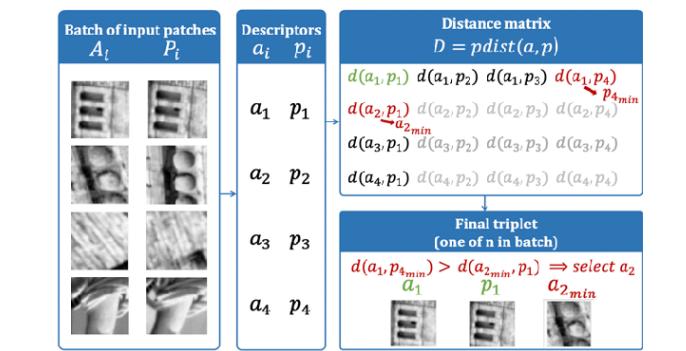


Figure 1: Proposed sampling procedure. First, patches are described by the current network, then a distance matrix is calculated. The closest non-matching descriptor – shown in red – is selected for each a_i and p_i patch from positive pair (green) respectively. Finally, among two negative candidates the hardest one is chosen. All operations are done in a single forward pass.

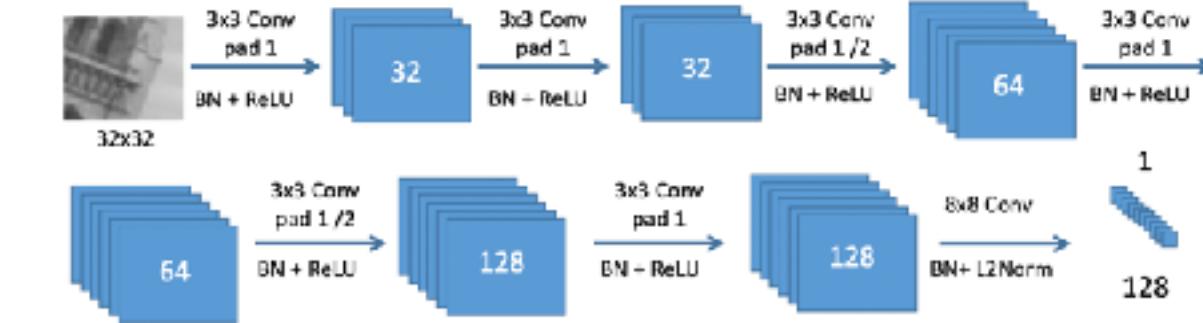


Figure 2: The architecture of our network, adopted from L2Net [24]. Each convolutional layer is followed by batch normalization and ReLU, except the last one. Dropout regularization is used before the last convolution layer.

The reading order of the method paper

Working hard to know your neighbor's margins: Local descriptor learning loss

- Abstract
- Method figure
- Main plot of results

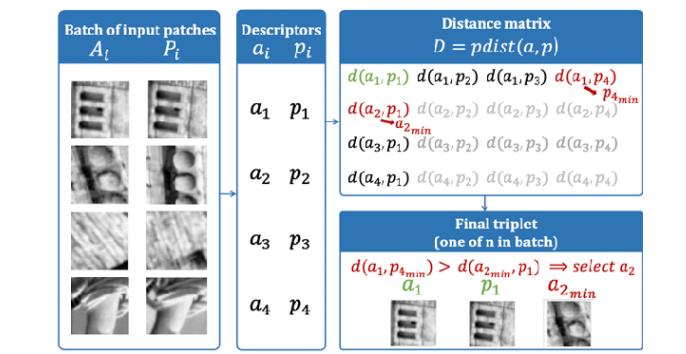


Figure 1: Proposed sampling procedure. First, patches are described by the current network, then a distance matrix is calculated. The closest non-matching descriptor – shown in red – is selected for each a_i and p_i patch from positive pair (green) respectively. Finally, among two negative candidates the hardest one is chosen. All operations are done in a single forward pass.

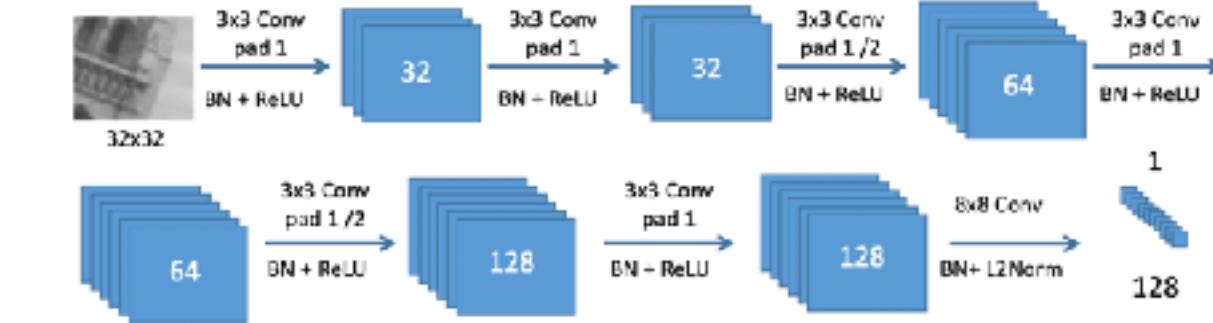


Figure 2: The architecture of our network, adopted from L2Net [24]. Each convolutional layer is followed by batch normalization and ReLU, except the last one. Dropout regularization is used before the last convolution layer.

The reading order of the method paper

Working hard to know your neighbor's margins: Local descriptor learning loss

- Abstract
- Method figure
- Main plot of results

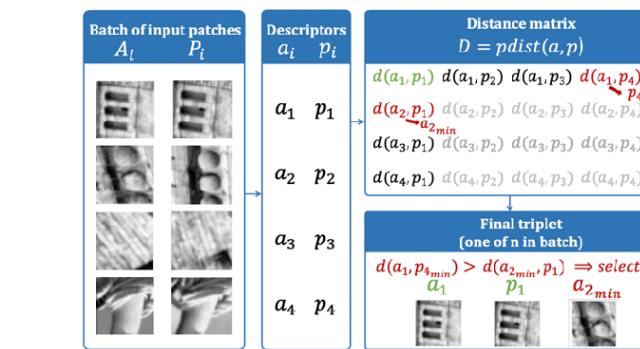


Figure 1: Proposed sampling procedure. First, patches are described by the current network, then a distance matrix is calculated. The closest non-matching descriptor – shown in red – is selected for each a_i and p_i patch from positive pair (green) respectively. Finally, among two negative candidates the hardest one is chosen. All operations are done in a single forward pass.

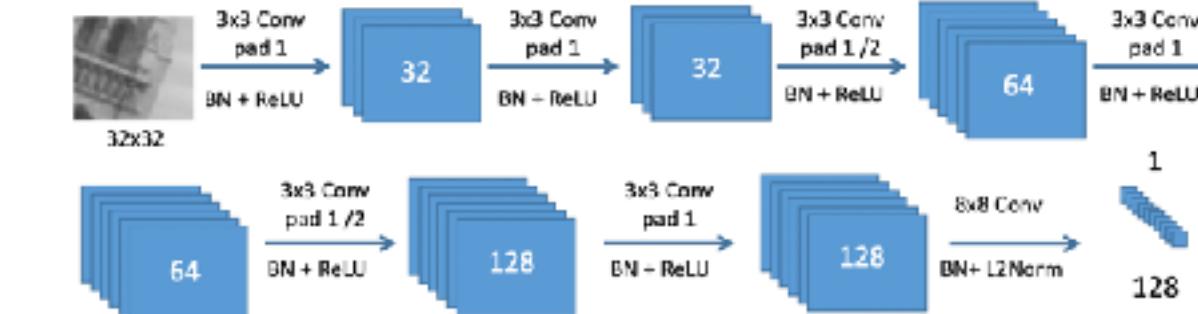
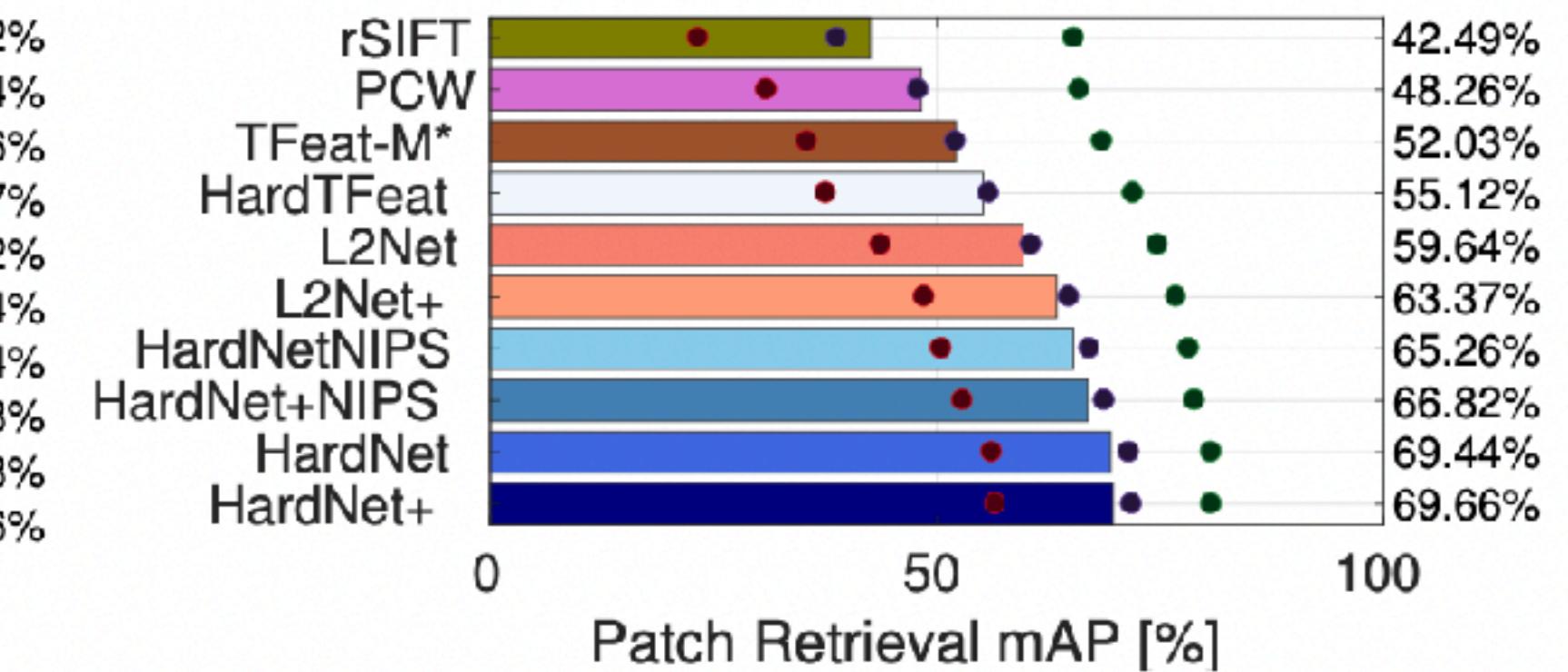
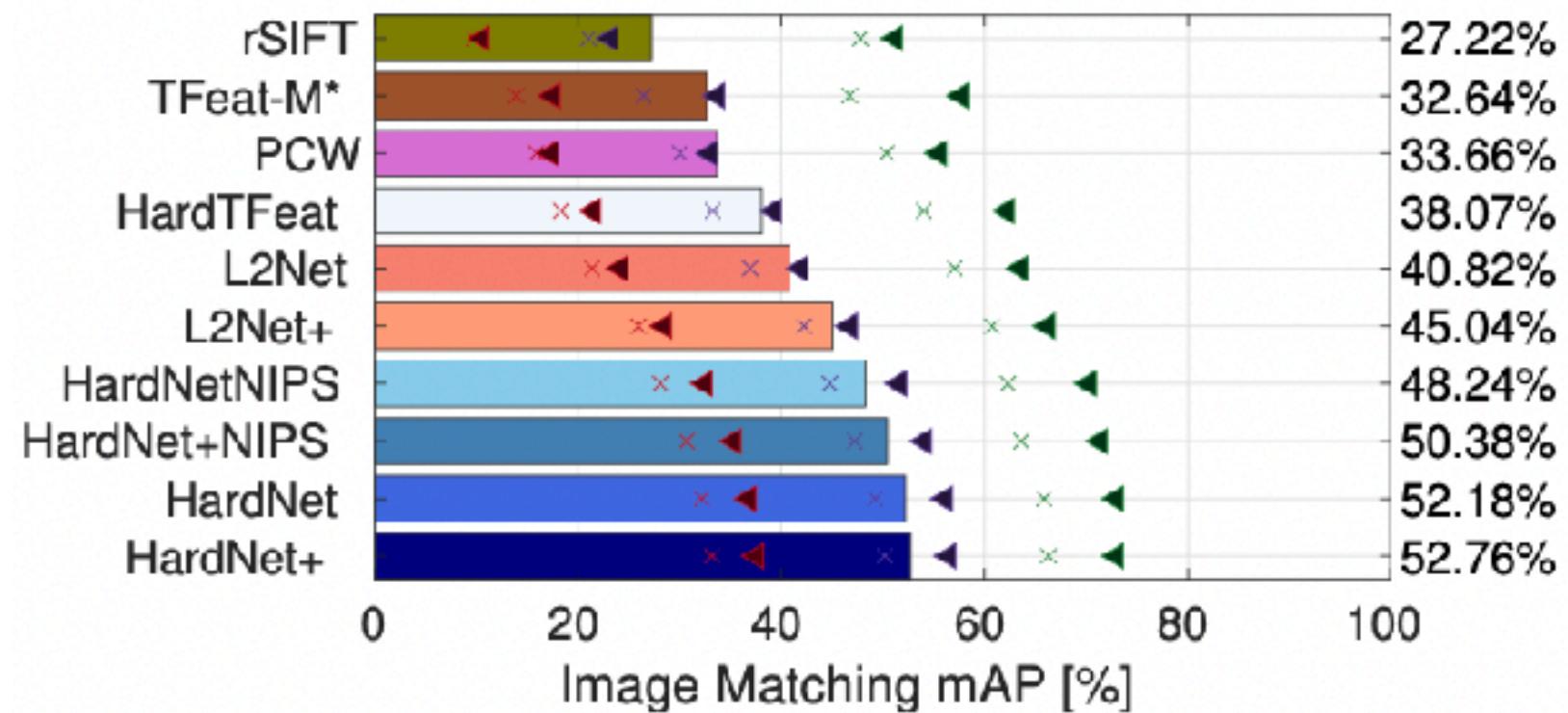


Figure 2: The architecture of our network, adopted from L2Net [24]. Each convolutional layer is followed by batch normalization and ReLU, except the last one. Dropout regularization is used before the last convolution layer.

■ EASY ■ HARD ■ TOUGH

◀ VIEWPT × ILLUM



The reading order of the method paper

Working hard to know your neighbor's margins: Local descriptor learning loss

- Abstract
- Method figure
- Main plot of results
- Conclusion

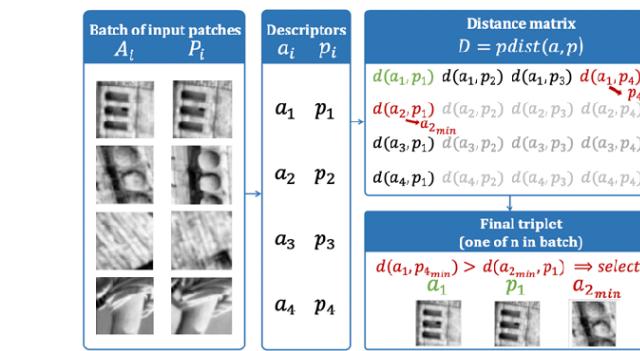


Figure 1: Proposed sampling procedure. First, patches are described by the current network, then a distance matrix is calculated. The closest non-matching descriptor – shown in red – is selected for each a_i and p_i patch from positive pair (green) respectively. Finally, among two negative candidates the hardest one is chosen. All operations are done in a single forward pass.

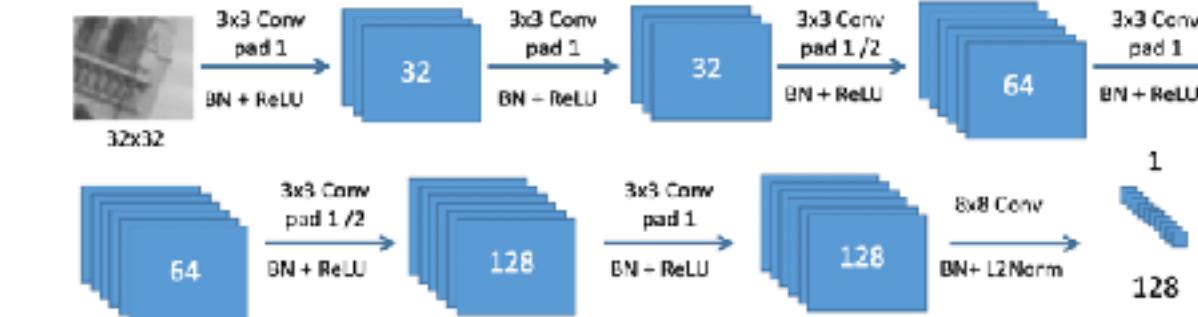
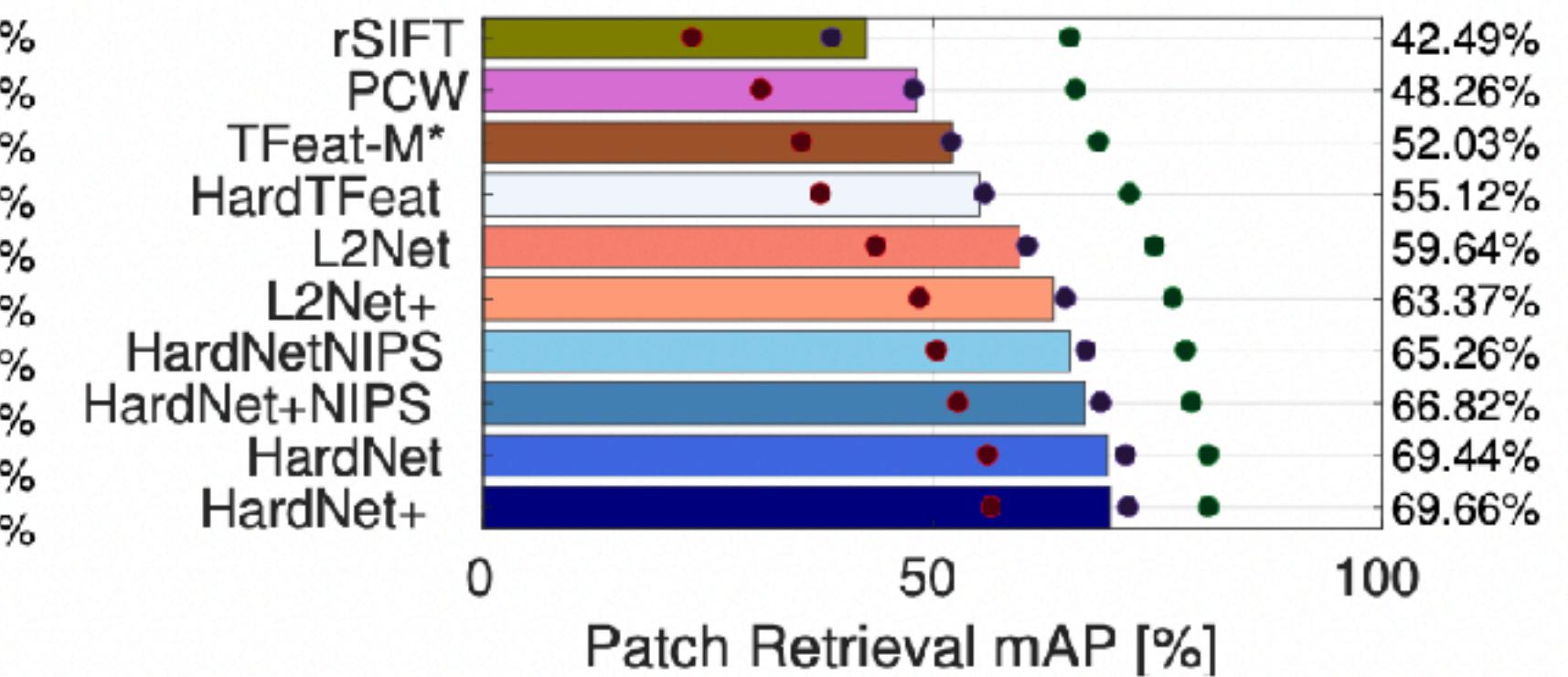
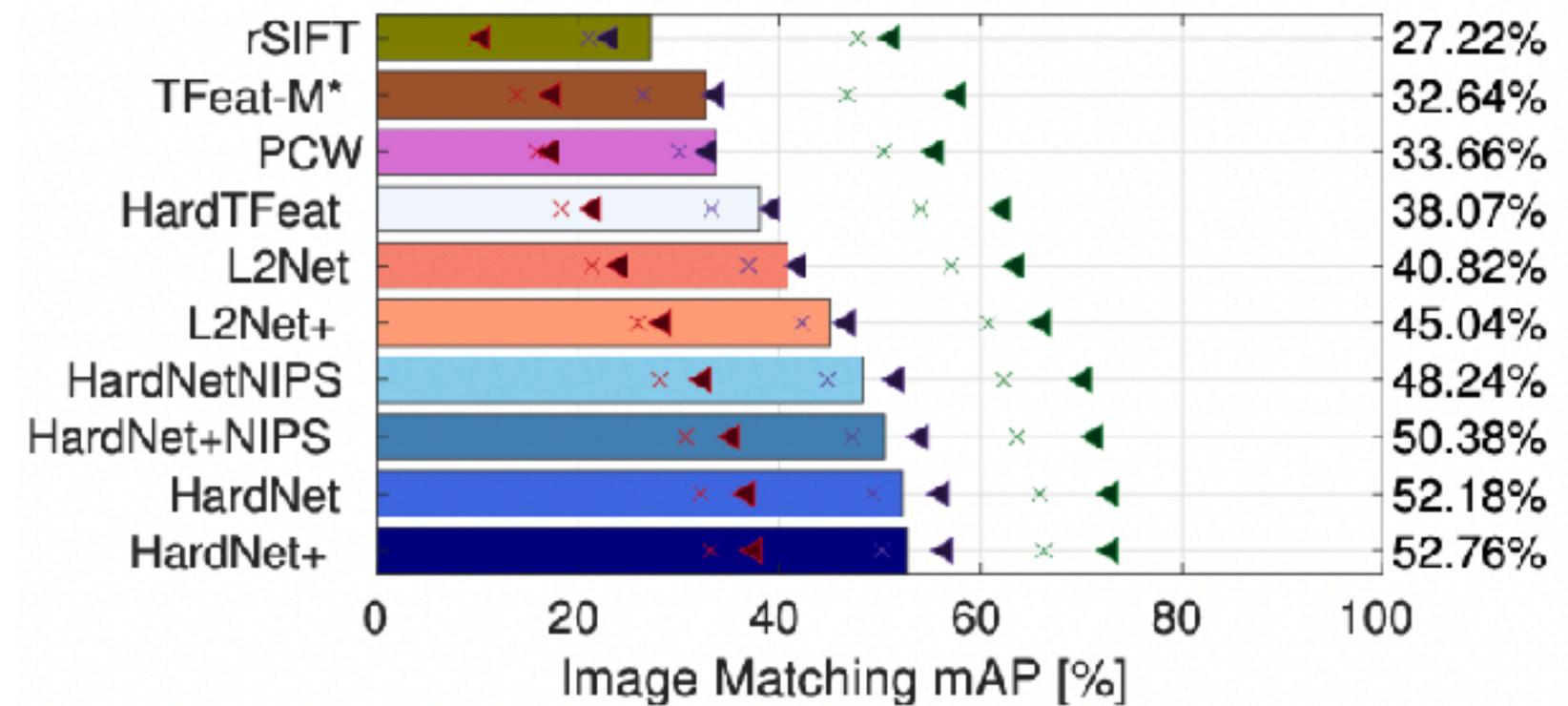


Figure 2: The architecture of our network, adopted from L2Net [24]. Each convolutional layer is followed by batch normalization and ReLU, except the last one. Dropout regularization is used before the last convolution layer.

■ EASY ■ HARD ■ TOUGH

◀ VIEWPT × ILLUM



The reading order of the method paper

Working hard to know your neighbor's margins: Local descriptor learning loss

- Abstract
- Method figure
- Main plot of results
- Conclusion
 - If it is not rephrased abstract

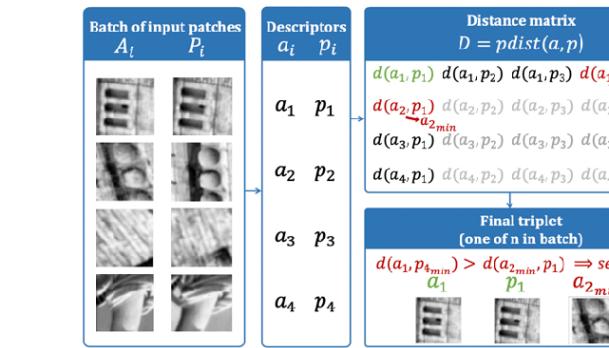


Figure 1: Proposed sampling procedure. First, patches are described by the current network, then a distance matrix is calculated. The closest non-matching descriptor – shown in red – is selected for each a_i and p_i patch from positive pair (green) respectively. Finally, among two negative candidates the hardest one is chosen. All operations are done in a single forward pass.

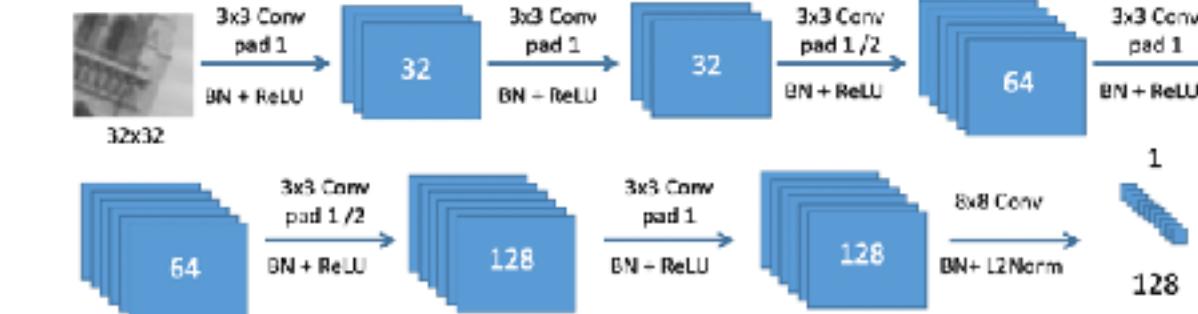
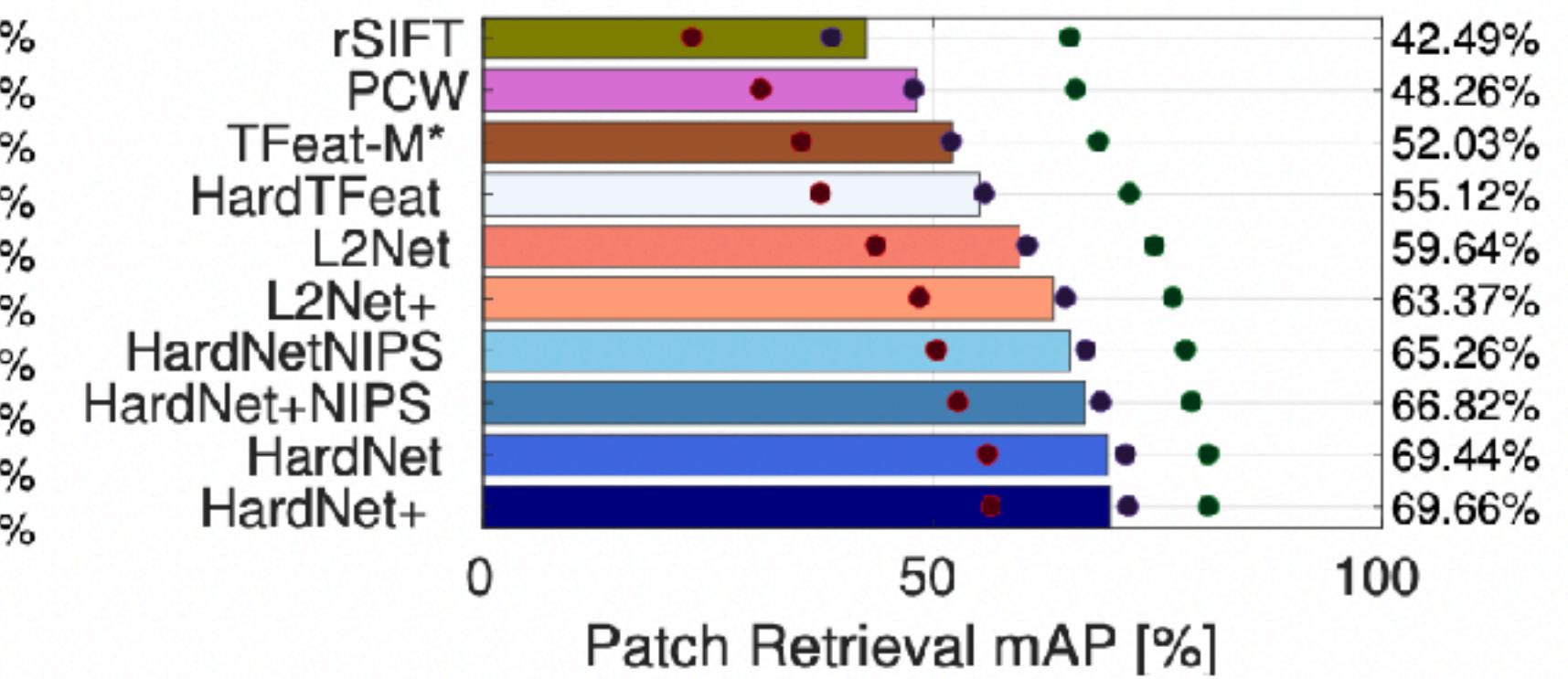
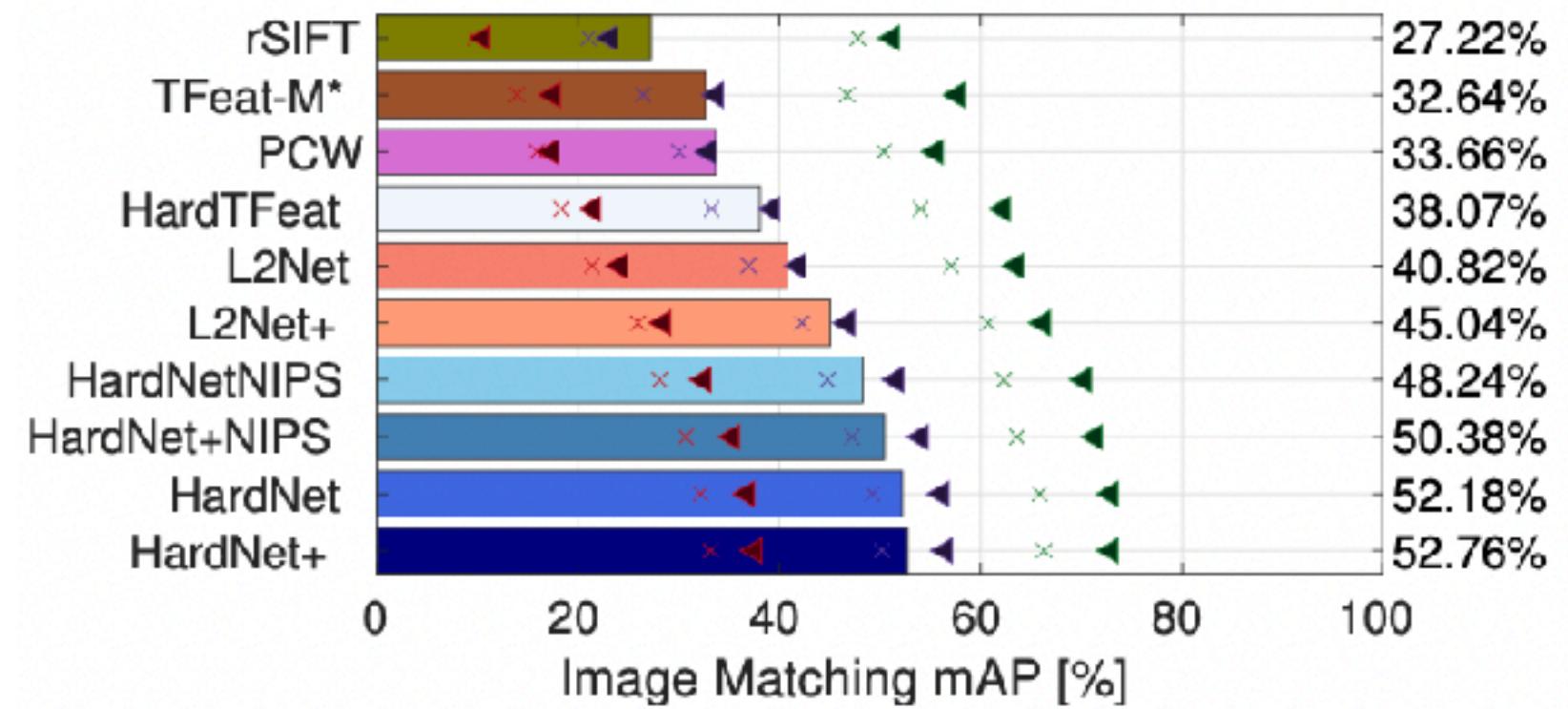


Figure 2: The architecture of our network, adopted from L2Net [24]. Each convolutional layer is followed by batch normalization and ReLU, except the last one. Dropout regularization is used before the last convolution layer.

■ EASY ■ HARD ■ TOUGH

◀ VIEWPT × ILLUM



The reading order of the method paper

Working hard to know your neighbor's margins: Local descriptor learning loss

- Abstract
- Method figure
- Main plot of results
- Conclusion
 - If it is not rephrased abstract
- Related work

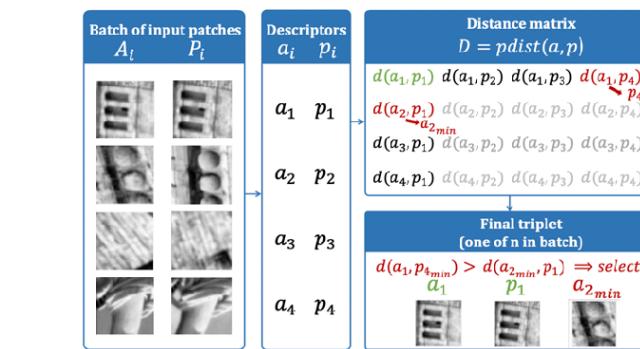


Figure 1: Proposed sampling procedure. First, patches are described by the current network, then a distance matrix is calculated. The closest non-matching descriptor – shown in red – is selected for each a_i and p_i patch from positive pair (green) respectively. Finally, among two negative candidates the hardest one is chosen. All operations are done in a single forward pass.

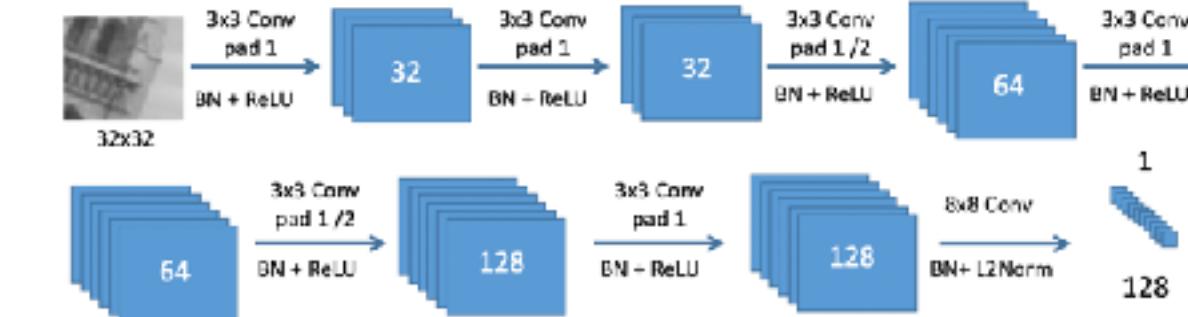
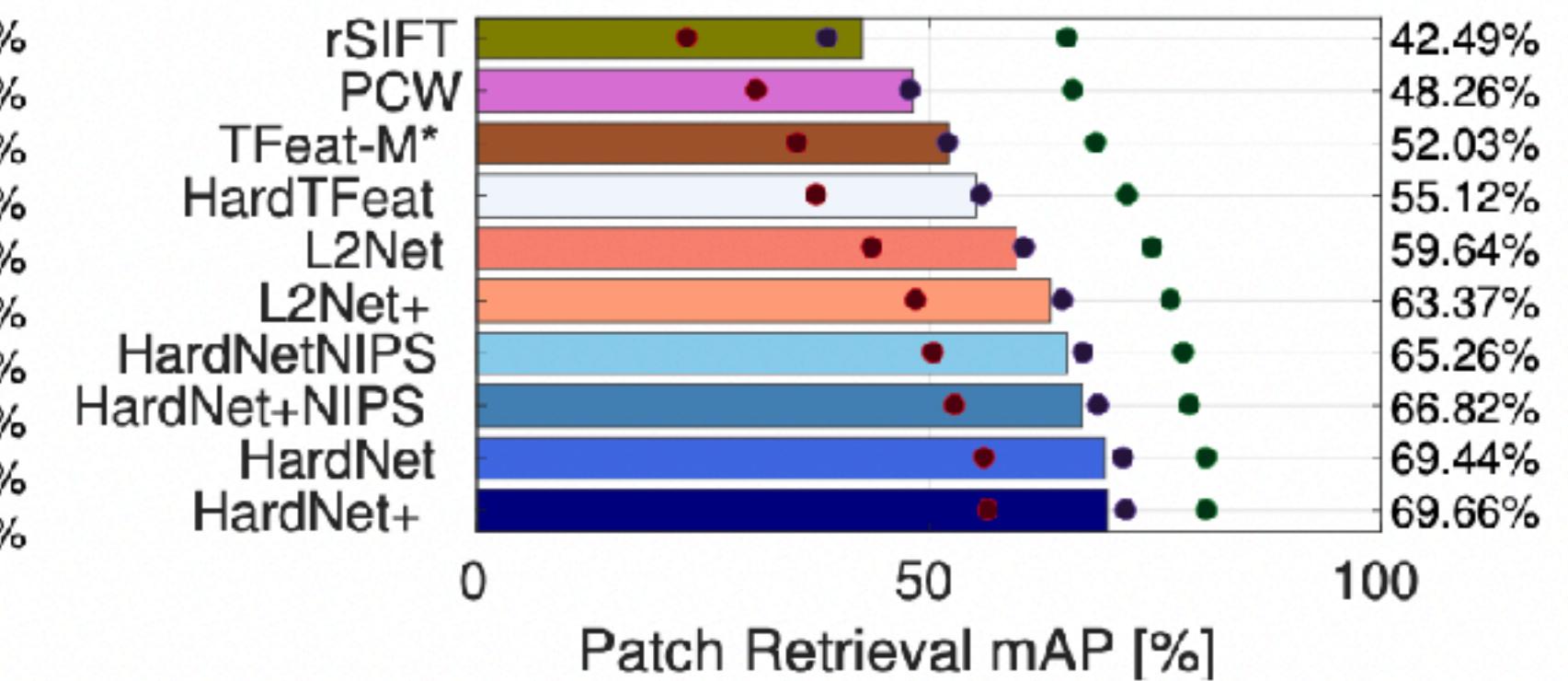
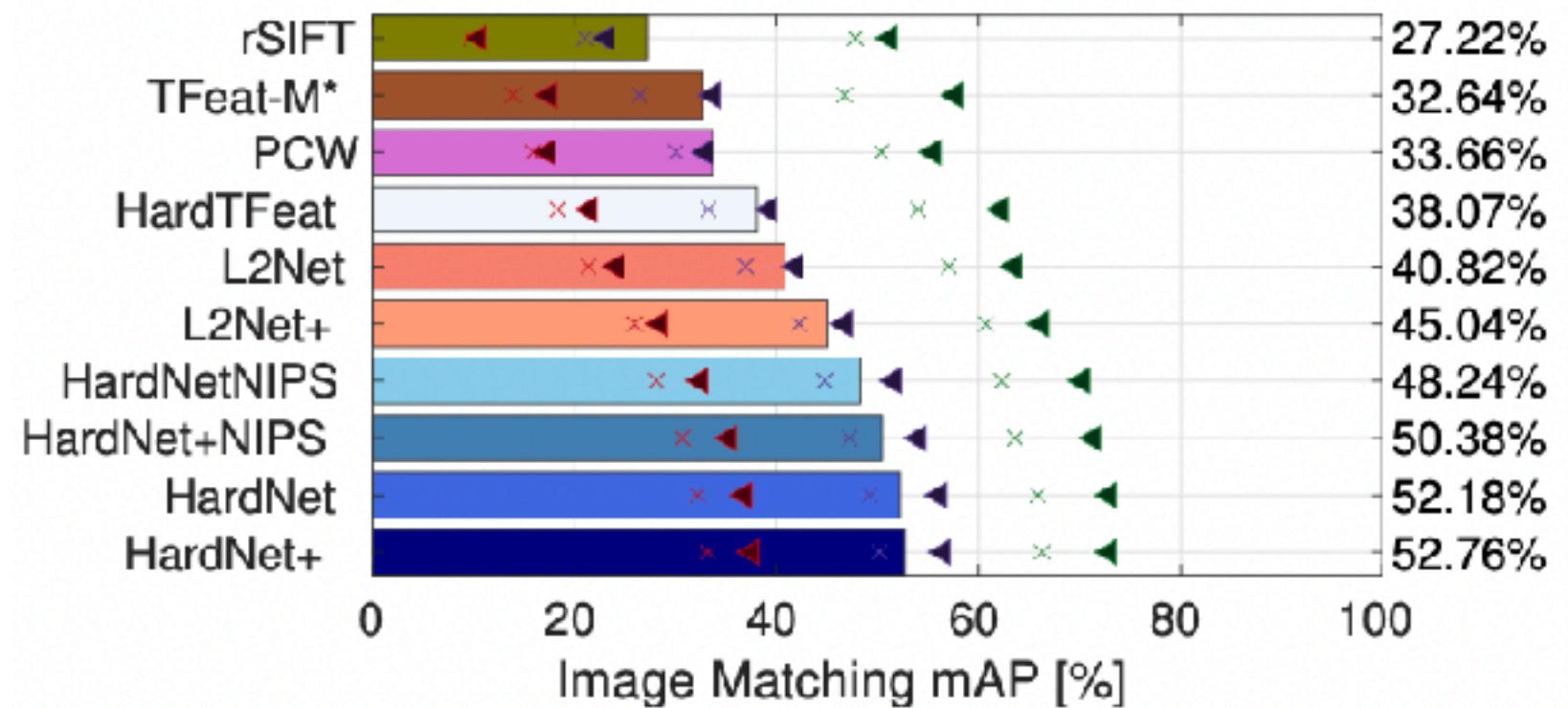


Figure 2: The architecture of our network, adopted from L2Net [24]. Each convolutional layer is followed by batch normalization and ReLU, except the last one. Dropout regularization is used before the last convolution layer.

■ EASY ■ HARD ■ TOUGH

◀ VIEWPT × ILLUM



Advice by Dushyant Mehta

From personal painful experience

Advice by Dushyant Mehta

From personal painful experience

- 1. Don't trust the story the paper tries to sell.**

Advice by Dushyant Mehta

From personal painful experience

1. Don't trust the story the paper tries to sell.
 - supposed intuition is often not true

Advice by Dushyant Mehta

From personal painful experience

1. Don't trust the story the paper tries to sell.
 - supposed intuition is often not true
 - Ask instead, what could be the other reasons for the claimed improvements.

Advice by Dushyant Mehta

From personal painful experience

- 1. Don't trust the story the paper tries to sell.**
 - supposed intuition is often not true
 - Ask instead, what could be the other reasons for the claimed improvements.
- 2. Learn to ask yourself about simple and obvious baselines.**

Advice by Dushyant Mehta

From personal painful experience

- 1. Don't trust the story the paper tries to sell.**
 - supposed intuition is often not true
 - Ask instead, what could be the other reasons for the claimed improvements.
- 2. Learn to ask yourself about simple and obvious baselines.**
 - What should have been the simple and obvious baselines for the paper you are reading. How would you solve the task building from first principles?

Advice by Dushyant Mehta

From personal painful experience

- 1. Don't trust the story the paper tries to sell.**
 - supposed intuition is often not true
 - Ask instead, what could be the other reasons for the claimed improvements.
- 2. Learn to ask yourself about simple and obvious baselines.**
 - What should have been the simple and obvious baselines for the paper you are reading. How would you solve the task building from first principles?
 - What experiments would you design to validate the claims of the paper?

Advice by Dushyant Mehta

From personal painful experience

Advice by Dushyant Mehta

From personal painful experience

- 3. Don't take the results at their face value**

Advice by Dushyant Mehta

From personal painful experience

3. Don't take the results at their face value

- Sure there is a big convenient table with all manner of methods conveniently compared.

Advice by Dushyant Mehta

From personal painful experience

3. Don't take the results at their face value

- Sure there is a big convenient table with all manner of methods conveniently compared.
- Are they really comparable? Same data, same compute budget, fair training, fair evaluation?

Advice by Dushyant Mehta

From personal painful experience

- 3. Don't take the results at their face value**
 - Sure there is a big convenient table with all manner of methods conveniently compared.
 - Are they really comparable? Same data, same compute budget, fair training, fair evaluation?
- 4. Don't discount the simple methods even if they are slightly worse than overly complex (claimed) SoTA.**

Advice by Dushyant Mehta

From personal painful experience

3. Don't take the results at their face value

- Sure there is a big convenient table with all manner of methods conveniently compared.
- Are they really comparable? Same data, same compute budget, fair training, fair evaluation?

4. Don't discount the simple methods even if they are slightly worse than overly complex (claimed) SoTA.

- There is immense advantage to the simple. Less finicky training, less faff in deployment, faster experiment iteration,

Advice by Dushyant Mehta

From personal painful experience

Advice by Dushyant Mehta

From personal painful experience

- 5. Do NOT start building on top of their *method***

Advice by Dushyant Mehta

From personal painful experience

5. Do NOT start building on top of their *method*

- At least without having determined to your satisfaction that the method is sound

Advice by Dushyant Mehta

From personal painful experience

5. Do NOT start building on top of their *method*

- At least without having determined to your satisfaction that the method is sound
- Reuse various parts from the codebase, without building on top of the method!

Advice by Dushyant Mehta

From personal painful experience

5. Do NOT start building on top of their *method*
 - At least without having determined to your satisfaction that the method is sound
 - Reuse various parts from the codebase, without building on top of the method!
6. Data, data, data! For a lot of the problems, not a lot beats good data.

Advice by Dushyant Mehta

From personal painful experience

5. Do NOT start building on top of their *method*
 - At least without having determined to your satisfaction that the method is sound
 - Reuse various parts from the codebase, without building on top of the method!
6. Data, data, data! For a lot of the problems, not a lot beats good data.
 - Not mindlessly collecting more data, but knowing the distribution of your data, and sampling smartly from it in a way that reflects your deployment setting.

Advice by Dushyant Mehta

From personal painful experience

5. Do NOT start building on top of their *method*
 - At least without having determined to your satisfaction that the method is sound
 - Reuse various parts from the codebase, without building on top of the method!
6. Data, data, data! For a lot of the problems, not a lot beats good data.
 - Not mindlessly collecting more data, but knowing the distribution of your data, and sampling smartly from it in a way that reflects your deployment setting.
 - Very simple/subtle sampling differences can cause a huge difference in your metrics, without anything about the method changing.

Advice by Dushyant Mehta

From personal painful experience

Advice by Dushyant Mehta

From personal painful experience

7. That brings us to the f**king metrics.

Advice by Dushyant Mehta

From personal painful experience

7. That brings us to the f**king metrics.

- If you mindlessly optimize for one number, look at leaderboards based on one number, you haven't understood the task.

Advice by Dushyant Mehta

From personal painful experience

7. That brings us to the f**king metrics.

- If you mindlessly optimize for one number, look at leaderboards based on one number, you haven't understood the task.
- What does that one number tell you?

Advice by Dushyant Mehta

From personal painful experience

7. That brings us to the f**king metrics.

- If you mindlessly optimize for one number, look at leaderboards based on one number, you haven't understood the task.
- What does that one number tell you?
- More importantly, what does it hide from you?

Advice by Dushyant Mehta

From personal painful experience

7. That brings us to the f**king metrics.

- If you mindlessly optimize for one number, look at leaderboards based on one number, you haven't understood the task.
- What does that one number tell you?
- More importantly, what does it hide from you?
- How relevant is it to the deployment scenarios for that task?

Advice by Dushyant Mehta

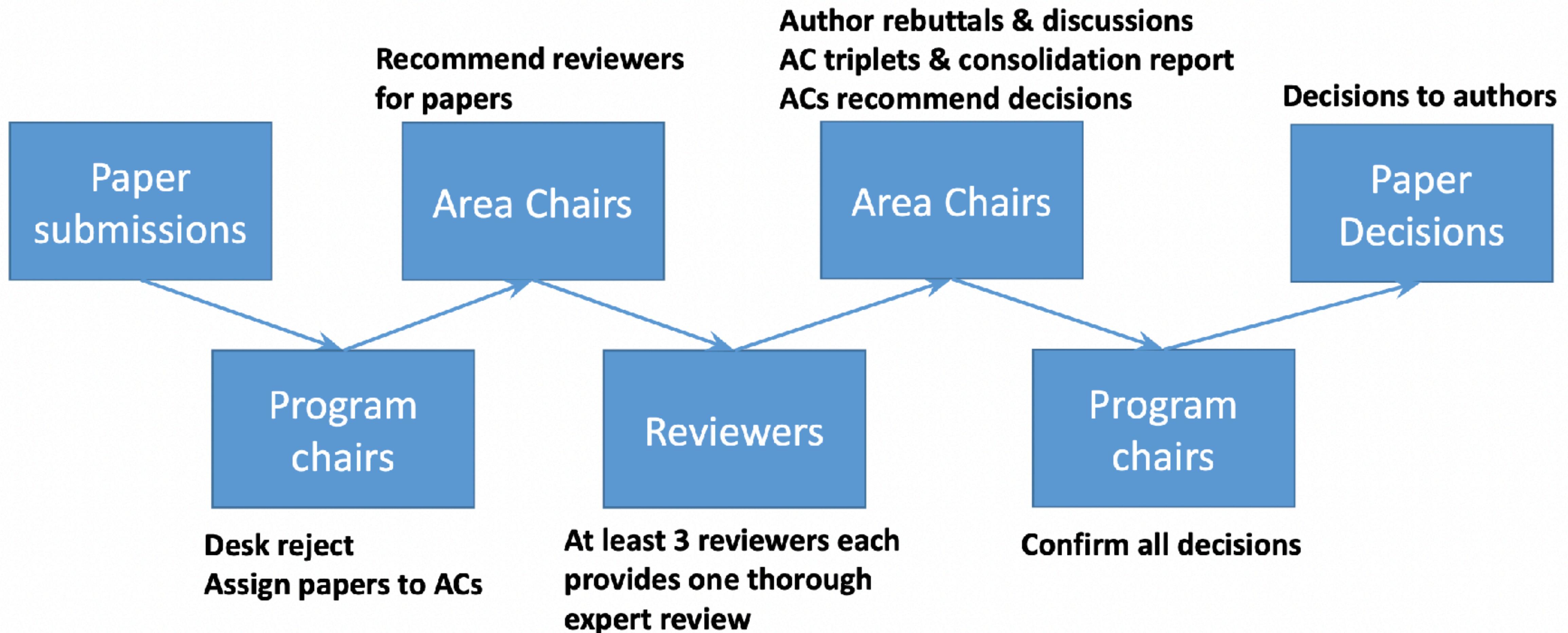
From personal painful experience

7. That brings us to the f**king metrics.

- If you mindlessly optimize for one number, look at leaderboards based on one number, you haven't understood the task.
- What does that one number tell you?
- More importantly, what does it hide from you?
- How relevant is it to the deployment scenarios for that task?
- What metrics would better reflect what matters the most for your deployment scenario?

What does peer review mean?

The process of accepting/rejecting a paper



The role of reviewers in the process

- Provide an independent, objective, critical, and comprehensive review
 - **Key:** What is the knowledge advancement in the paper?
- Discuss with AC and reviewer buddies to (hopefully) reach consensus
 - Explain clearly the basis of your review and recommendation
 - It is OK if the reviewers disagree with one another even after discussions
 - AC will form recommendations weighing in reviews, rebuttals, and discussions
- Make your final recommendations with solid justifications
 - Read the rebuttal and discussions. Do they change your position? Why?
 - This facilitates the ACs to make final recommendations for the paper

What paper should be accepted?

- Any paper that, in accordance with CVPR community standards,
 - presents **sufficient knowledge advancement** that is **well grounded**;
 - is of **sufficient interest** to some **CVPR audiences** who could **benefit** from it
- Note: CVPR is very inclusive
 - Historically rejection solely for out-of-scope is rather rare

What should be included in the review?

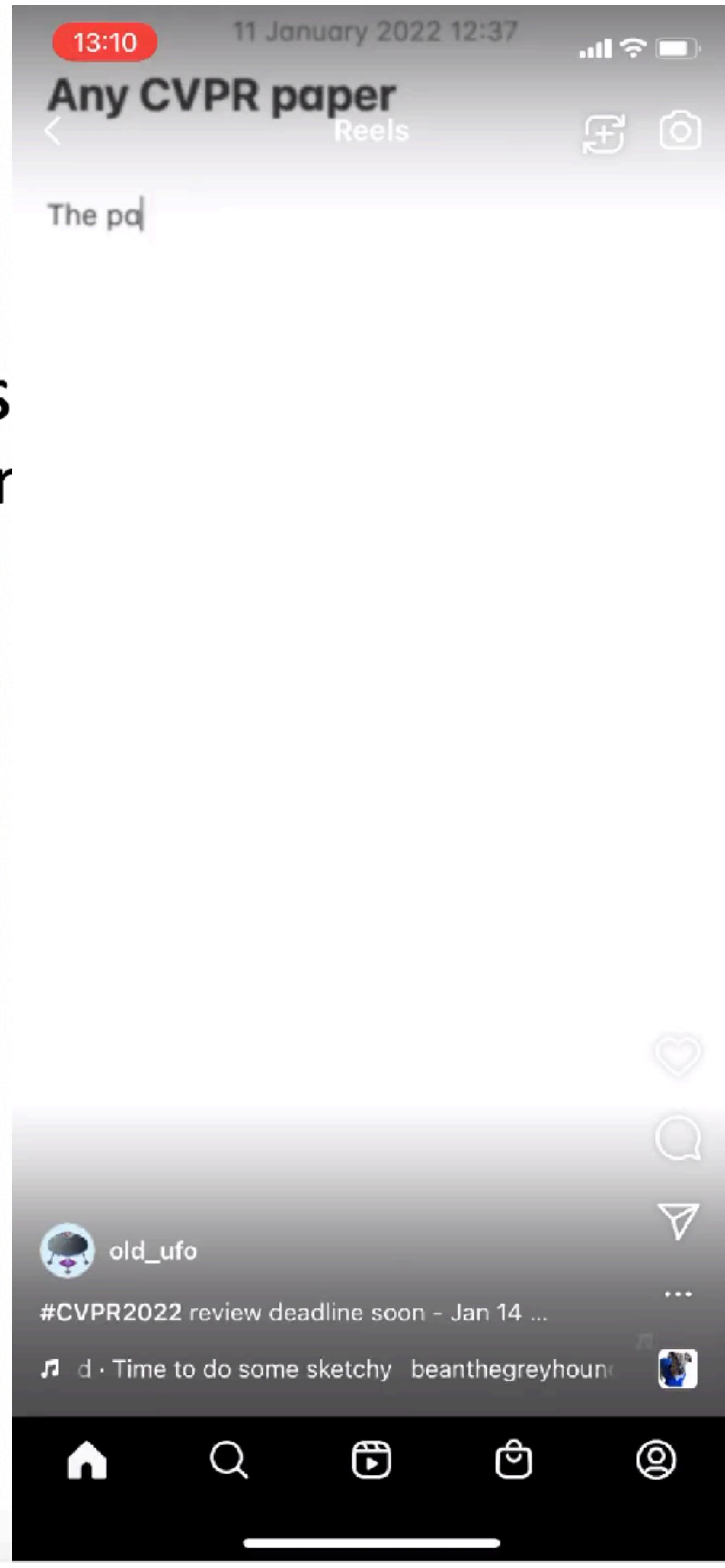
- A concise summary of the paper
 - What problem is addressed in the paper?
 - Is it a new problem? If so, why does it matter? If not, why does it still matter?
 - What is the key to the solution? What is the main contribution?
 - Do the experiments sufficiently support the claims?
- A clear statement of strengths and weaknesses
 - What are the key contributions and why do they matter?
 - What aspects of the paper most need improvement?
- A comprehensive check of potential fundamental flaws in the paper
 - Are the assumptions and theories (mathematically) sound?
 - Are the experiments scientifically sound and valid?
 - Is the problem addressed trivial?
 - Did the paper miss important prior work? Has it been done before? If yes, where?

What should be avoided in the review?

- Common mistakes in a sloppy review
 - Arrogance, ignorance, and inaccuracy
 - Be responsible!
 - Pure opinions
 - Be grounded!
 - Novelty fallacy
 - Be knowledge-driven!
 - Blank assertions
 - Be substantial!
 - Policy entrepreneurship
 - Be sensible!
 - Intellectual laziness
 - Be active!

What should be avoided in a review?

- Common mistakes in a review
 - Arrogance, ignorance,傲慢
 - Be responsible!
- Pure opinions
 - Be grounded!
- Novelty fallacy
 - Be knowledge-driven!
- Blank assertions
 - Be substantial!
- Policy entrepreneurship
 - Be sensible!
- Intellectual laziness
 - Be active!



Functions of the conferences

Why authors submit

Functions of the conferences

Why authors submit

- **Knowledge dissemination:** “I want people to know about the new knowledge I discovered.” The conference promises a certain minimum level of attention one’s work gets. In other words, if you don’t publish at top conferences, nobody would read your work.

Functions of the conferences

Why authors submit

- **Knowledge dissemination:** “I want people to know about the new knowledge I discovered.” The conference promises a certain minimum level of attention one’s work gets. In other words, if you don’t publish at top conferences, nobody would read your work.
- **Feedback:** I would like people to check my results through the review process and discussion.

Functions of the conferences

Why authors submit

- **Knowledge dissemination:** “I want people to know about the new knowledge I discovered.” The conference promises a certain minimum level of attention one’s work gets. In other words, if you don’t publish at top conferences, nobody would read your work.
- **Feedback:** I would like people to check my results through the review process and discussion.
- **Formal goodies:** checking boxes, required to defend Ph.D., for tenure package, performance review, to put into grant report, etc.

Functions of the conferences

Why authors submit

- **Knowledge dissemination:** “I want people to know about the new knowledge I discovered.” The conference promises a certain minimum level of attention one’s work gets. In other words, if you don’t publish at top conferences, nobody would read your work.
- **Feedback:** I would like people to check my results through the review process and discussion.
- **Formal goodies:** checking boxes, required to defend Ph.D., for tenure package, performance review, to put into grant report, etc.
- **Certification.** Publishing at CVPR is hard, therefore valuable.

Functions of the conferences

Why authors submit

- **Knowledge dissemination:** “I want people to know about the new knowledge I discovered.” The conference promises a certain minimum level of attention one’s work gets. In other words, if you don’t publish at top conferences, nobody would read your work.
- **Feedback:** I would like people to check my results through the review process and discussion.
- **Formal goodies:** checking boxes, required to defend Ph.D., for tenure package, performance review, to put into grant report, etc.
- **Certification.** Publishing at CVPR is hard, therefore valuable.
- **Reputation-building:** Listing certain conferences on C.V. as a way of building one’s name as a scientist.

Functions of the conferences

Why authors submit

- **Knowledge dissemination:** “I want people to know about the new knowledge I discovered.” The conference promises a certain minimum level of attention one’s work gets. In other words, if you don’t publish at top conferences, nobody would read your work.
- **Feedback:** I would like people to check my results through the review process and discussion.
- **Formal goodies:** checking boxes, required to defend Ph.D., for tenure package, performance review, to put into grant report, etc.
- **Certification.** Publishing at CVPR is hard, therefore valuable.
- **Reputation-building:** Listing certain conferences on C.V. as a way of building one’s name as a scientist.
- **Networking:** meeting with peers, potential employers, etc

Functions of the conferences

Why people attend

Functions of the conferences

Why people attend

- **Prefiltering:** time is limited, so we outsource the selection of what we are reading to the reviewers.

Functions of the conferences

Why people attend

- **Prefiltering:** time is limited, so we outsource the selection of what we are reading to the reviewers.
- **Certification:** time is limited, so we outsource the quality control and result check to the reviewers. We create a basic classifier: “If the paper is published by a top-conference, it is true.”

Functions of the conferences

Why people attend

- **Prefiltering:** time is limited, so we outsource the selection of what we are reading to the reviewers.
- **Certification:** time is limited, so we outsource the quality control and result check to the reviewers. We create a basic classifier: “If the paper is published by a top-conference, it is true.”
- **Special case of certification:** for people outside the field without the basic qualifications to select work that meets basic quality guarantees.

Functions of the conferences

Why people attend

- **Prefiltering:** time is limited, so we outsource the selection of what we are reading to the reviewers.
- **Certification:** time is limited, so we outsource the quality control and result check to the reviewers. We create a basic classifier: “If the paper is published by a top-conference, it is true.”
- **Special case of certification:** for people outside the field without the basic qualifications to select work that meets basic quality guarantees.
- **Authors promise to answer our questions** (symmetrical to “attention for the author from audience,” and audience gets the guarantee that questions about the work will be answered at the talk or poster session).

Thank you for your attention



Dmytro Mishkin, PhD

ducha.aiki@gmail.com

<http://dmytro.ai/>

 [ducha_aiki](#)

Thank you for your attention



Dmytro Mishkin, PhD

ducha.aiki@gmail.com

<http://dmytro.ai/>

 [ducha_aiki](#)

- Computer vision researcher at CTU in Prague.

Thank you for your attention



Dmytro Mishkin, PhD

ducha.aiki@gmail.com

<http://dmytro.ai/>

 [ducha_aiki](#)

- Computer vision researcher at CTU in Prague.

Thank you for your attention



Dmytro Mishkin, PhD

ducha.aiki@gmail.com

<http://dmytro.ai/>

 [ducha_aiki](https://twitter.com/ducha_aiki)

- Computer vision researcher at CTU in Prague.
- Member of the Expert Committee on Artificial Intelligence at Ministry of Digital Transformation Of Ukraine

Thank you for your attention



Dmytro Mishkin, PhD

ducha.aiki@gmail.com

<http://dmytro.ai/>

 [ducha_aiki](https://twitter.com/ducha_aiki)

- Computer vision researcher at CTU in Prague.
- Member of the Expert Committee on Artificial Intelligence at Ministry of Digital Transformation Of Ukraine

Thank you for your attention



Dmytro Mishkin, PhD

ducha.aiki@gmail.com

<http://dmytro.ai/>

 [ducha_aiki](#)

- Computer vision researcher at CTU in Prague.
- Member of the Expert Committee on Artificial Intelligence at Ministry of Digital Transformation Of Ukraine
- Computer vision and deep learning consultant

Thank you for your attention



Dmytro Mishkin, PhD

ducha.aiki@gmail.com

<http://dmytro.ai/>

 [ducha_aiki](#)

- Computer vision researcher at CTU in Prague.
- Member of the Expert Committee on Artificial Intelligence at Ministry of Digital Transformation Of Ukraine
- Computer vision and deep learning consultant

Thank you for your attention



Dmytro Mishkin, PhD

ducha.aiki@gmail.com

<http://dmytro.ai/>

 [ducha_aiki](https://twitter.com/ducha_aiki)

- Computer vision researcher at CTU in Prague.
- Member of the Expert Committee on Artificial Intelligence at Ministry of Digital Transformation Of Ukraine
- Computer vision and deep learning consultant
- Co-author of the first Ukrainian publications at CVPR an NeurIPS

Thank you for your attention



Dmytro Mishkin, PhD
ducha.aiki@gmail.com
<http://dmytro.ai/>
 [ducha_aiki](https://twitter.com/ducha_aiki)

- Computer vision researcher at CTU in Prague.
- Member of the Expert Committee on Artificial Intelligence at Ministry of Digital Transformation Of Ukraine
- Computer vision and deep learning consultant
- Co-author of the first Ukrainian publications at CVPR and NeurIPS
- CVPR 2021 Outstanding reviewer

Thank you for your attention



Dmytro Mishkin, PhD
ducha.aiki@gmail.com
<http://dmytro.ai/>
 [ducha_aiki](https://twitter.com/ducha_aiki)

- Computer vision researcher at CTU in Prague.
- Member of the Expert Committee on Artificial Intelligence at Ministry of Digital Transformation Of Ukraine
- Computer vision and deep learning consultant
- Co-author of the first Ukrainian publications at CVPR an NeurIPS
- CVPR 2021 Outstanding reviewer
- Kaggle Master