Research Skill Bootcamp

Literature Review and Paper Reading

Daochang Liu and Sirui Li

Why literature review?

Final Rating: 1: reject
Final Rating Justification:
The authors missed to address my concern regarding novelty compared to [3] (published in 2022 ECCV) and [4] (NeurIPS 2023).

An example: a review comment on one of my CVPR 2025 submissions:

One of the most common reasons to be rejected:

"You paper is not novel! There is already a similar paper!"

Such critics are usually difficult to answer.

It is better to avoid from very beginning.

Why literature review?

- Get a map of the known world
- To explore somewhere unknown
- Find the path from the known to unknown

This is the first step of your research journey!

By GPT-4o

When to do literature review?

80% of literature search

Get necessary context
Needs some iterations
Help refine your topic

10% of literature search

Know your related methods
Know the limitations
Get some inspirations

10% of literature search

Identify compared methods
Choose experiment settings
Plan your experiment table

1% of literature search

Find some handy evidence

General Reading

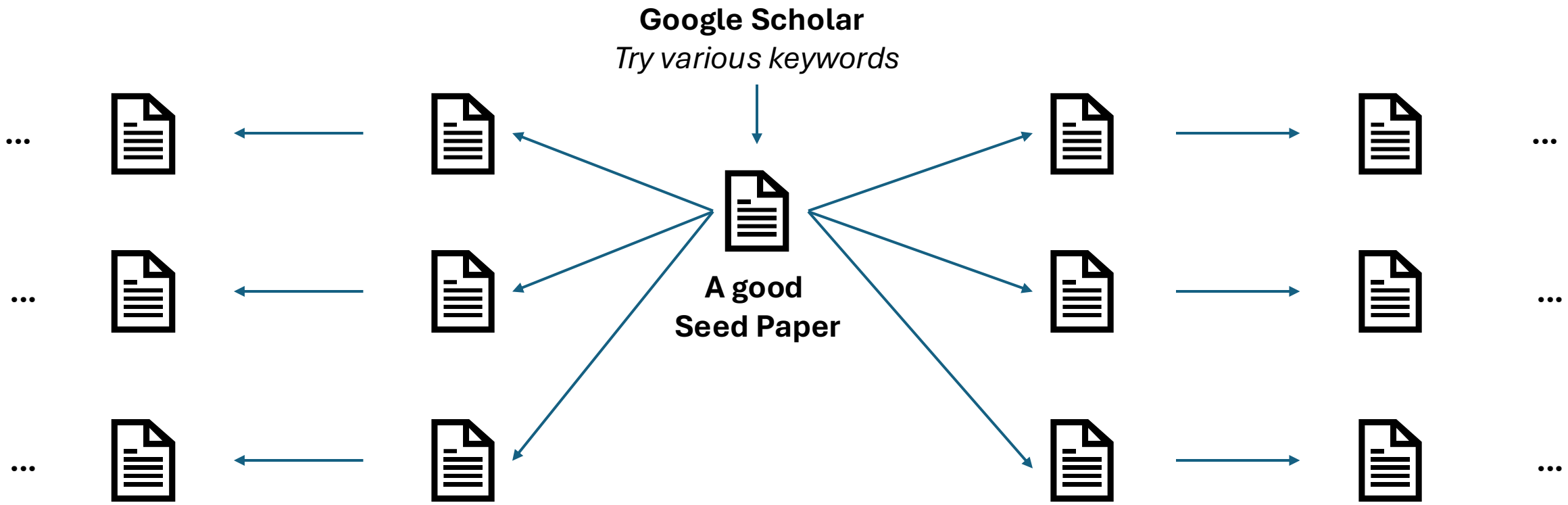
Untargeted Structure

Detailed Reading

Targeted Structure

How to do literature search?

BFS Graph Search



Find another seed paper and do again
Be patient … Until there is no new papers
(5 times are usually sufficient)

How to do literature search?

How to choose the seed paper?

- Datasets or Benchmarks: *All the method papers needs to cite the dataset*
- Survey Papers: *Not suitable for very new topic*

Vbench: Comprehensive benchmark suite for video generative models
Z Huang, Y He, J Yu, F Zhang, C Si… - Proceedings of the …, 2024 - openaccess.thecvf.com
… Then given two videos sampled from the same prompt, we ask human annotators to indicate preferences according to each **VBench** dimension respectively. We show that **VBench** …
☆ Save 99 Cite Cited by 234 Related articles All 7 versions »

Papers citing the dataset

Diffusion models in vision: A **survey**
FA Croitoru, V Hondru, RT Ionescu… - IEEE Transactions on …, 2023 - ieeexplore.ieee.org
… of articles on denoising **diffusion models** in computer vision. More precisely, we **survey** articles that fall in the category of generative **models** defined below. **Diffusion models** represent a …
☆ Save 99 Cite Cited by 1359 Related articles All 7 versions Web of Science: 398

Papers cited by the survey

How to do literature search?

Some correlated but not causal factors indicating a good paper:

- From top conferences / journals
  - CORE A*/A: NeurIPS, ICLR, ICML, ACL, EMNLP, NAACL, CVPR, ICCV, ECCV, …
  - CORE Rankings
- From top labs / researchers
- Fully open-sourced code, model, hyper-parameters
- Simple and effective
- Good writing, figure, tables, presentation
- High impact, High citation, A lot social media coverage

This is very subjective!

How to read a single paper? (General Reading)

Define/Choose Task → **Design Method** → **Experiments** → **Writing / Rebuttal**

80% of literature search

Get necessary context
Needs some iterations
Help refine your topic

10% of literature search

Know your related methods
Know the limitations
Get some inspirations

10% of literature search

Identify compared methods
Choose experiment settings
Plan your experiment table

1% of literature search

Find some handy evidence

General Reading

Untargeted Structure

Detailed Reading

Targeted Structure

How to read a single paper? (General Reading)

What is the task?

- What is the input? What is the output?
- What are the training/testing data?
- What is the ground truth? What are the evaluation metrics?

What is the motivation?

- What challenge does this task have?
- What high-level gap the method wants to fill?

What is the method?

- First answer this using very broad keywords, e.g., CNN, Transformer, RL, etc.

What is the result?

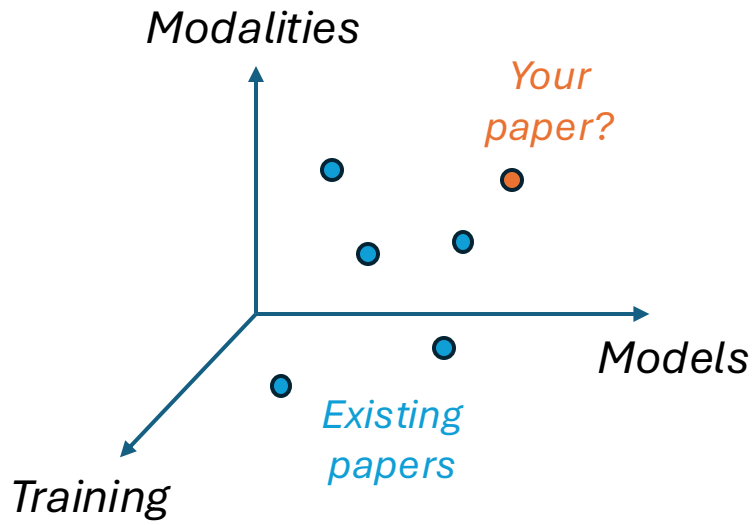
- What are the datasets? How good the performance is?
- How much GPUs are needed?

How to read a batch of paper? (Untargeted Structuring)

Decompose Task vs. Methodology: *These can be two separate scope to search and organize papers*

Grouping papers using different dimensions:

- **Chronologically**: How is the method improved over time?
- **Inputs / Outputs**: e.g., conditions in generative models
- **Modalities**: text, image, video, audio, etc.
- **Models**: autoregressive, diffusion, mask modeling, etc.
- **Networks architecture**: CNN, RNN, Transformers, etc.
- **Training Strategy**: supervised, semi-supervised, unsupervised, RL, etc.
- **Dataset:** ImageNet, CIFAR-100, LAION, etc.
- **Code Availability**



How to position your own paper in all these dimensions?

There can be many more dimensions, depending on the specific topic you are working on.

Focus on recent papers (<3 years, except for very basic papers like “Attention”)

When writing paper, you need to choose or adjust to fit in your story in Related Works (Targeted Structuring)

How to read a batch of paper? (Untargeted Structuring)

An example

Method	Venue	Representation	Model	Condition	Dataset
Action2Motion [7]	MM 2020	Rot.	VAE	Action class	[7], [87]
SA-GCN [88]	ECCV 2020	Kpts. (2D, 3D)	GAN	Action class	[87], [89]
ACTOR [90]	ICCV 2021	Rot.	VAE	Action class	[7], [87], [91]
Kinetic-GAN [92]	WACV 2022	Kpts. (3D)	GAN	Action class	[87], [89], [93]
ODMO [94]	MM 2022	Kpts. (3D)	VAE	Action class, Trajectory†	[7], [91]
PoseGPT [95]	ECCV 2022	Rot.	VAE	Action class, Duration, Past motion†	[7], [96], [97]
Cervantes *et al.* [98]	ECCV 2022	Rot.	Regression	Action class	[7], [87], [91]
MultiAct [99]	AAAI 2023	Kpts. (3D) / Rot.	VAE	Action class, Past motion†	[96]
MDM [14]	ICLR 2023	Kpts. (3D) / Rot.	Diffusion	Action class / Text	[3], [7], [91], [100]
MLD [101]	CVPR 2023	Kpts. (3D) / Rot.	Diffusion, VAE	Action class / Text	[3], [7], [91], [100]
Text2Action [102]	ICRA 2018	Kpts. (3D)	GAN	Text	[103]
JL2P [104]	3DV 2019	Kpts. (3D)	Regression	Text	[100]
Ghosh *et al.* [105]	ICCV 2021	Kpts. (3D)	GAN	Text	[100]
Guo *et al.* [3]	CVPR 2022	Kpts. (3D) / Rot.	VAE	Text, POS	[3], [100]
AvatarCLIP [106]	TOG 2022	Rot.	VAE	Text	[40]
MotionCLIP [107]	ECCV 2022	Rot.	Regression	Text	[96]
TEMOS [108]	ECCV 2022	Kpts. (3D) / Rot.	VAE	Text	[100]
TM2T [109]	ECCV 2022	Kpts. (3D) / Rot.	VAE	Text	[3], [100]
TEACH [110]	3DV 2022	Rot.	VAE	Text, Past motion†	[96]
FLAME [111]	AAAI 2023	Kpts. (3D) / Rot.	Diffusion	Text	[3], [96], [100]
T2M-GPT [112]	CVPR 2023	Kpts. (3D) / Rot.	VAE	Text	[3], [100]
OOHMG [113]	CVPR 2023	Rot.	VAE	Text	[40], [96]
UDE [114]	CVPR 2023	Rot.	Diffusion, VAE	Text / Music	[3], [115]
MoFusion [116]	CVPR 2023	Kpts. (3D)	Diffusion	Text / Music	[3], [96], [115]

Human Motion Generation: A Survey https://arxiv.org/pdf/2307.10894

How to read a single paper? (Detailed/Critical Reading)

80% of literature search

Get necessary context
Needs some iterations
Help refine your topic

10% of literature search

Know your related methods
Know the limitations
Get some inspirations

10% of literature search

Identify compared methods
Choose experiment settings
Plan your experiment table

1% of literature search

Find some handy evidence

General Reading

Untargeted Structure

Detailed Reading

Targeted Structure

How to read a single paper? (Detailed/Critical Reading)

What are the strengths?

- Is the motivation reasonable? Is the method novel? Are results significant? Is writing good?

What are the weaknesses?

- The opposite of above

How the method work in detail?

- What are the tensor shapes?
- What does each dimension mean intuitively?
- How the dimensions changes step by step?
- What does it mean intuitively?

Are the experiments solid?

- Is there any ablation study? Is the performance gain really from the contribution?
- How many datasets are used? Can the method generalize?
- Can visualizations well explain the results and the method?

What story the author is telling? What the author is actually doing?

Can you learn anything? Can you use it in your project? What inspiration can you have?

How to read a single paper? (Detailed/Critical Reading)

What are the strengths?

- Is the motivation reasonable? Is the method novel? Are results significant? Is writing good?

What are the weaknesses?

- The opposite of above

Think from the perspective of reviewers.

Look at reviewer guidelines from top conferences:

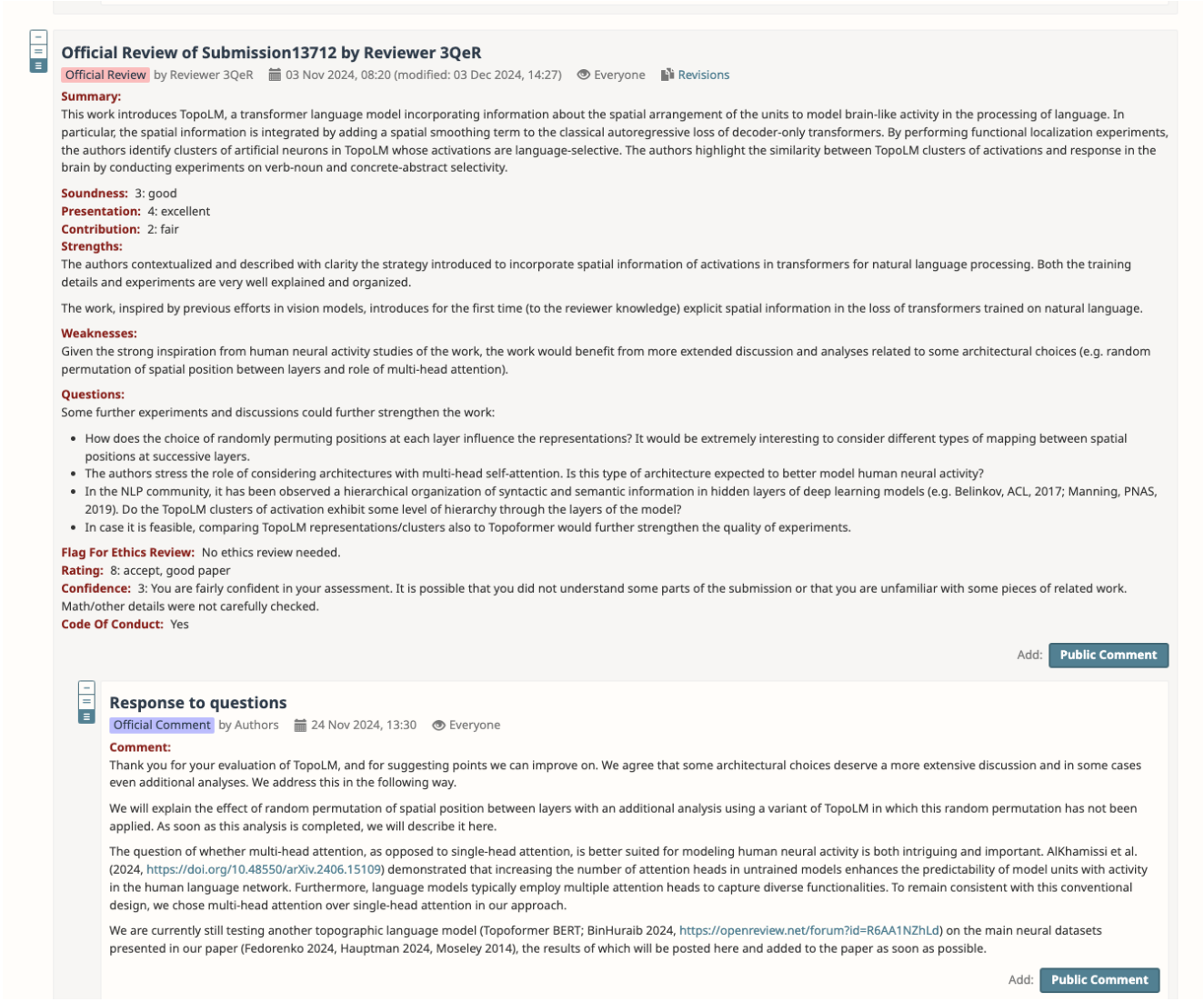
ICLR Reviewer Guide

NeurIPS Reviewer Guide

CVPR Reviewer Guide

2. **Strengths and Weaknesses:** Please provide a thorough assessment of the strengths and weakne
   a. *Originality:* Are the tasks or methods new? Is the work a novel combination of well-known tech
   adequately cited?
   b. *Quality*: Is the submission technically sound? Are claims well supported (e.g., by theoretical an
   work in progress? Are the authors careful and honest about evaluating both the strengths and w
   c. *Clarity*: Is the submission clearly written? Is it well organized? (If not, please make constructive
   paper provides enough information for an expert reader to reproduce its results.)
   d. *Significance:* Are the results important? Are others (researchers or practitioners) likely to use
   work? Does it advance the state of the art in a demonstrable way? Does it provide unique data, u
   You can incorporate Markdown and Latex into your review. See https://openreview.net/faq.

How to read a single paper? (Detailed/Critical Reading)



Official Review of Submission13712 by Reviewer 3QeR

Official Review by Reviewer 3QeR 03 Nov 2024, 08:20 (modified: 03 Dec 2024, 14:27) Everyone Revisions

Summary:
This work introduces TopoLM, a transformer language model incorporating information about the spatial arrangement of the units to model brain-like activity in the processing of language. In particular, the spatial information is integrated by adding a spatial smoothing term to the classical autoregressive loss of decoder-only transformers. By performing functional localization experiments, the authors identify clusters of artificial neurons in TopoLM whose activations are language-selective. The authors highlight the similarity between TopoLM clusters of activations and response in the brain by conducting experiments on verb-noun and concrete-abstract selectivity.

Soundness: 3: good
Presentation: 4: excellent
Contribution: 2: fair
Strengths:
The authors contextualized and described with clarity the strategy introduced to incorporate spatial information of activations in transformers for natural language processing. Both the training details and experiments are very well explained and organized.

The work, inspired by previous efforts in vision models, introduces for the first time (to the reviewer knowledge) explicit spatial information in the loss of transformers trained on natural language.

Weaknesses:
Given the strong inspiration from human neural activity studies of the work, the work would benefit from more extended discussion and analyses related to some architectural choices (e.g. random permutation of spatial position between layers and role of multi-head attention).

Questions:
Some further experiments and discussions could further strengthen the work:

- How does the choice of randomly permuting positions at each layer influence the representations? It would be extremely interesting to consider different types of mapping between spatial positions at successive layers.
- The authors stress the role of considering architectures with multi-head self-attention. Is this type of architecture expected to better model human neural activity?
- In the NLP community, it has been observed a hierarchical organization of syntactic and semantic information in hidden layers of deep learning models (e.g. Belinkov, ACL, 2017; Manning, PNAS, 2019). Do the TopoLM clusters of activation exhibit some level of hierarchy through the layers of the model?
- In case it is feasible, comparing TopoLM representations/clusters also to Topoformer would further strengthen the quality of experiments.

Flag For Ethics Review: No ethics review needed.
Rating: 8: accept, good paper
Confidence: 3: You are fairly confident in your assessment. It is possible that you did not understand some parts of the submission or that you are unfamiliar with some pieces of related work. Math/other details were not carefully checked.
Code Of Conduct: Yes

Add: Public Comment

Response to questions

Official Comment by Authors 24 Nov 2024, 13:30 Everyone

Comment:
Thank you for your evaluation of TopoLM, and for suggesting points we can improve on. We agree that some architectural choices deserve a more extensive discussion and in some cases even additional analyses. We address this in the following way.

We will explain the effect of random permutation of spatial position between layers with an additional analysis using a variant of TopoLM in which this random permutation has not been applied. As soon as this analysis is completed, we will describe it here.

The question of whether multi-head attention, as opposed to single-head attention, is better suited for modeling human neural activity is both intriguing and important. AlKhamissi et al. (2024, https://doi.org/10.48550/arXiv.2406.15109) demonstrated that increasing the number of attention heads in untrained models enhances the predictability of model units with activity in the human language network. Furthermore, language models typically employ multiple attention heads to capture diverse functionalities. To remain consistent with this conventional design, we chose multi-head attention over single-head attention in our approach.

We are currently still testing another topographic language model (Topoformer BERT; BinHuraib 2024, https://openreview.net/forum?id=R6AA1NZhLd) on the main neural datasets presented in our paper (Fedorenko 2024, Hauptman 2024, Moseley 2014), the results of which will be posted here and added to the paper as soon as possible.

Add: Public Comment

Open Review is good source of learning

Some conferences, e.g. ICLR, will release all reviews and rebuttals

How to read a single paper? (Detailed/Critical Reading)

How the method works in detail?

- What are the data structures (tensor shapes)?
- What does each dimension mean intuitively?
- How the dimensions changes step by step?
- What does it mean intuitively? *Intuition is useful in research*

An example in video models:

Tensor Shape: `[B, C, H, W, T]`

> `[Batch, Channel, Height, Width, Length]`

Suppose there are some affinity matrices in the attention layers:

`[HW, HW]`: *It might be a spatial attention*

`[T, T]`: *It might be a temporal attention*

How to read a single paper? (Detailed/Critical Reading)

What story the author is telling? What the method is actually doing?

Sometimes they can be different!

4.2.1 Topographical Embeddings

It is common to feed neural weather models multiple time-independent variables containing topographical information like sea-land mask [8]. Instead of manually selecting and preparing this kind of information, we use a novel technique of *topographical embeddings*, which allows the network to automatically discover relevant topographical information and store it in embedding. More precisely, we allocate a grid of embeddings with a stride of 4 km where each point is associated with 20 scalar parameters. For each input example, we calculate the topographical embedding of each input pixel center by bilinearly interpolating the embeddings from the grid. The embedding parameters are trained together with other model parameters similarly to embeddings used in NLP.

What story the author is telling?

"Topographical Embeddings"

This is the interpretation of the author

The interpretation is insightful in this case,
but it can be confusing in some other cases

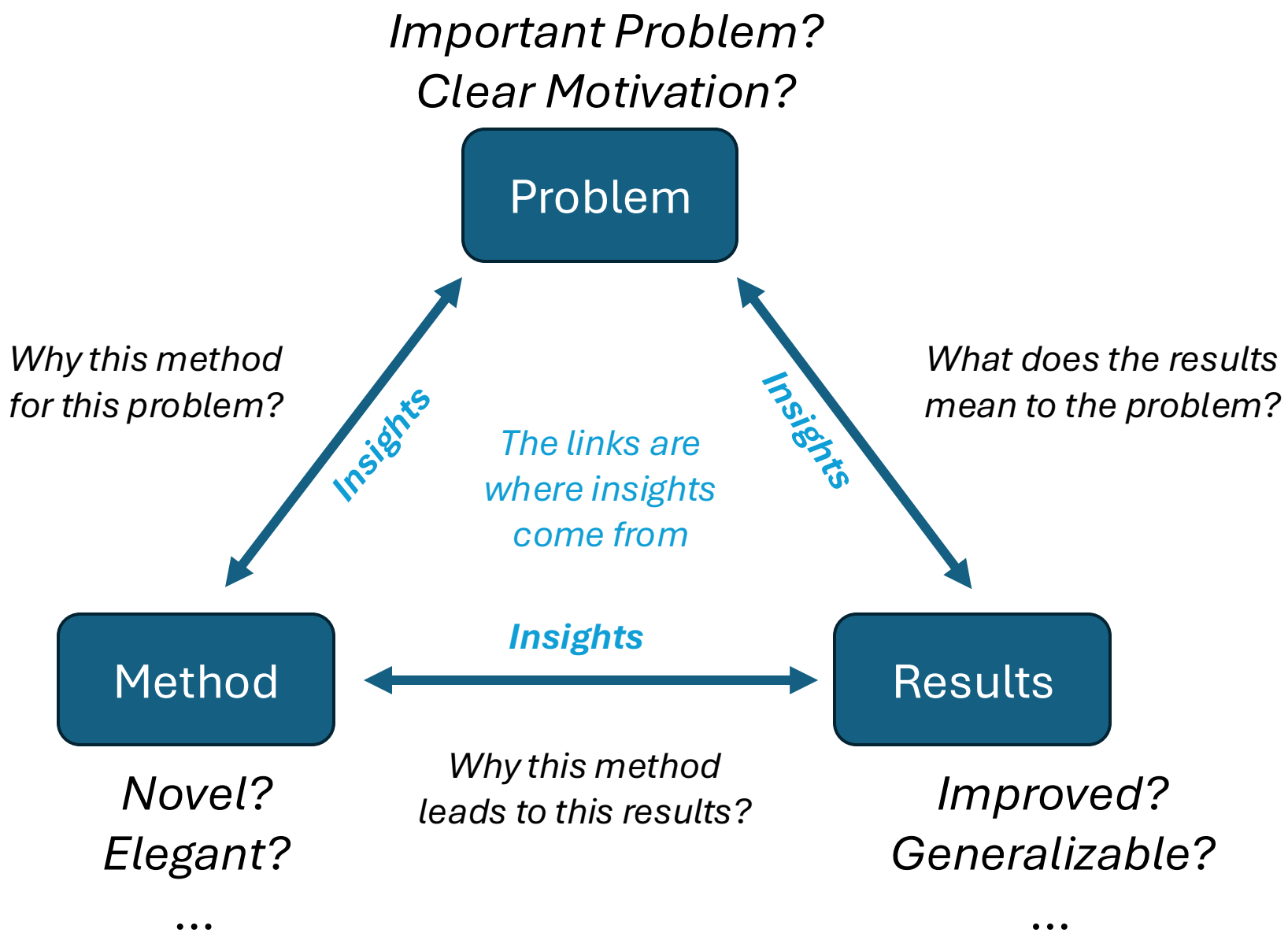
What the author is actually doing?

"Learnable positional embeddings"

This is some fact

Deep Learning for Day Forecasts from Sparse Observations https://arxiv.org/abs/2306.06079

How to read a single paper? (Detailed/Critical Reading)



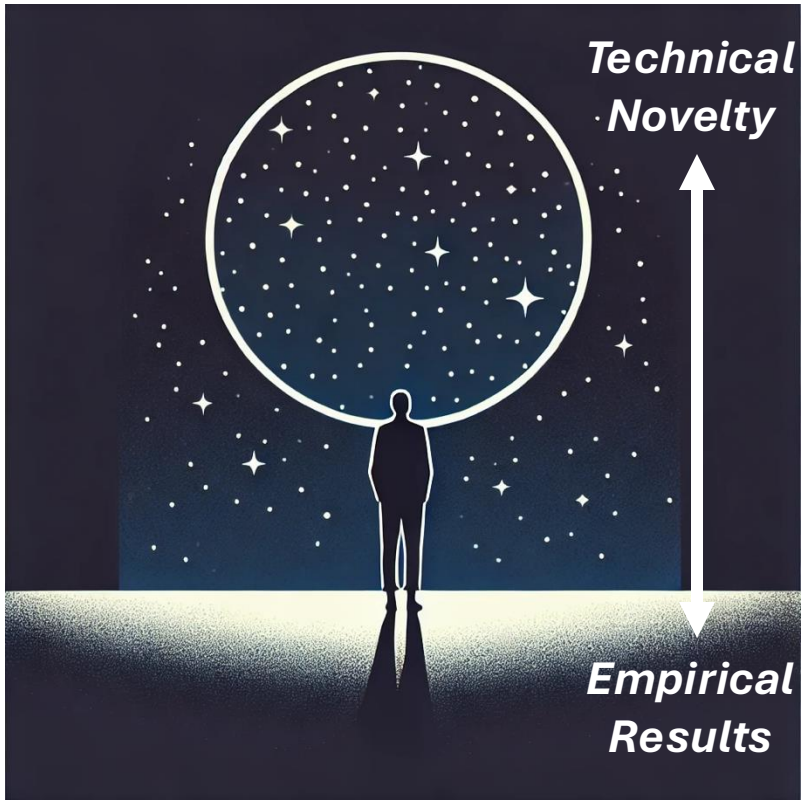
How to identifying research gaps?

Can you find any trends from your review?
What are the limitations of existing papers?

Two most common ways:

- Is there any new model not having been used in this task? Let's have a try. **(Top-down thinking)**
- Is there any pattern of failure cases of existing methods? Let's fix it. **(Bottom-up thinking)**

Be creative!



By GPT-4o

"Keep your feet on the ground and your eyes on the stars"

How to manage references?

Common Challenges in Research Writing:

- Forgetting sources
- Incorrect citations
- Formatting issues

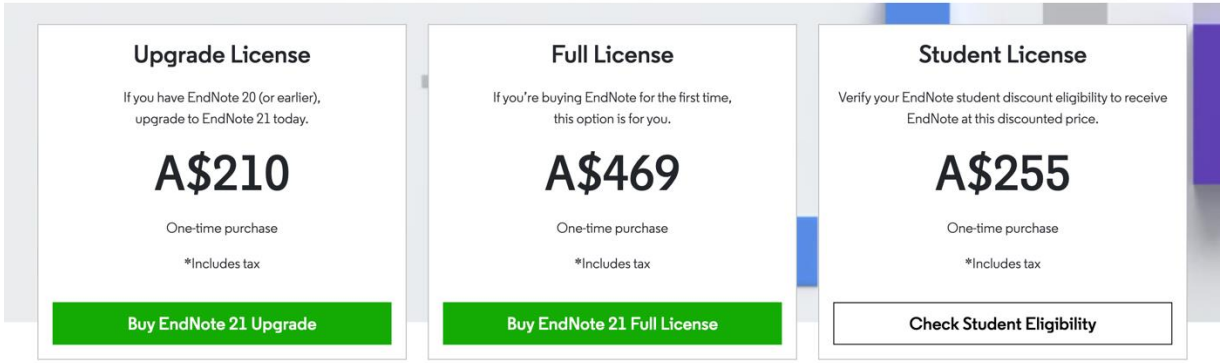
Solutions:

- EndNote: https://www.youtube.com/watch?v=XpqGuIJbP2I&t=144s
- Zotero: https://www.youtube.com/watch?v=JG7Uq_JFDzE
- Mendeley: https://www.youtube.com/watch?v=OzFHGFnAM2Q

They can directly export bib
Manage bibs on Overleaf

Endnote

- Paid Software



- Key features:
  - Reference Library – Store and manage all your sources in one place.
  - Browser Plugin: Available for **Firefox and Google Chrome** (EndNote Click)
  - Citation Plugin – Easily insert and format citations in **Microsoft Word, Google Docs and LibreOffice.**
  - PDF Annotation – Highlight, comment, and organise research papers.
  - Collaboration – Share and work on references with colleagues.
  - Cloud Sync – Access your references across multiple devices.

Zotero

- Free & Open-Source Software
- Key features:
  - Reference Library – Store and manage all your sources in one place.
  - Browser Plugin: Available for **Firefox, Google Chrome, and Edge**.
  - Citation Plugin – Easily insert and format citations in **Microsoft Word. Google Docs, and LibreOffice**.
  - PDF Annotation – Highlight, comment, and organise research papers.
  - Collaboration – Share and work on references with colleagues.
  - Cloud Sync – Access your references across multiple devices. **Free 300MB** of storage, paid options for more space.

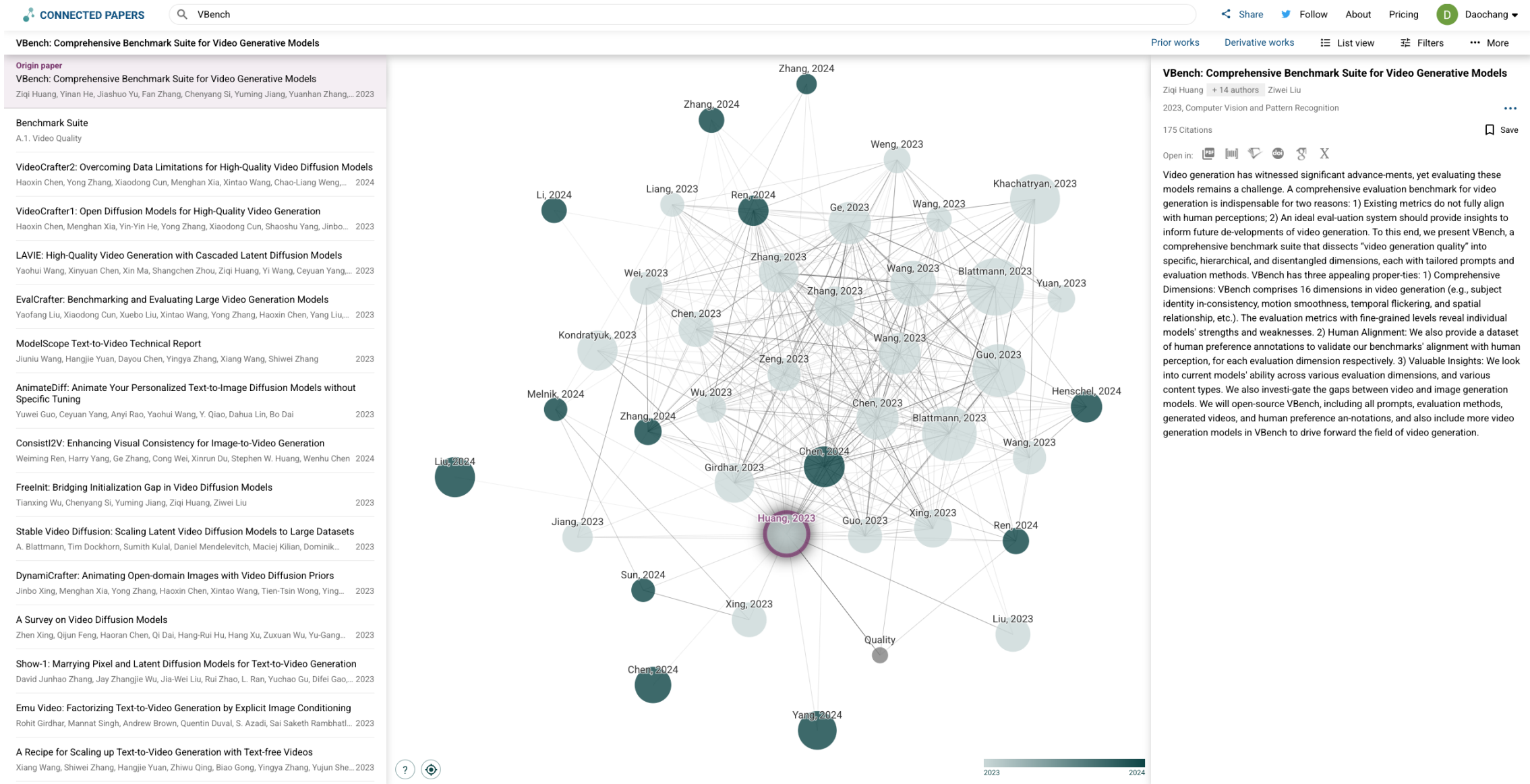
Mendeley

- Free with Premium Options
- Key features:
  - Reference Library – Store and manage all your sources in one place.
  - Browser Plugin: Available for **Firefox, Google Chrome, and Edge**.
  - Citation Plugin – Easily insert citations and bibliographies in **Microsoft Word and LibreOffice**.
  - PDF Annotation – Highlight, comment, and organise research papers.
  - Collaboration – Create shared libraries and collaborate with colleagues.
  - Cloud Sync – Sync your library across all devices, with up to **2GB** of free cloud storage. Paid options for additional storage.
  - Literature Search – Discover and access research articles directly within the platform.

How to manage references?

	EndNote	Zotero	Mendeley
Price	Paid (one-time or subscription)	Free (with paid cloud storage)	Free (with premium options)
Browser Plugin	Yes (Chrome & Firefox)	Yes (Chrome, Firefox, Edge)	Yes (for Chrome, Firefox, Edge)
Citation Plugin	Microsoft Word, Google Docs, LibreOffice	Microsoft Word, Google Docs, LibreOffice	Microsoft Word, LibreOffice
PDF Annotation	Yes	Yes	Yes
Cloud Sync	Yes	Yes (300MB free, paid for more)	Yes (2GB free, paid for more)
Collaboration	Yes (shared libraries)	Yes (group libraries)	Yes (shared groups)
Literature Search	Yes	No	Yes

CONNECTED PAPERS



Thanks!