# The PhD Playbook

## Lessons learned during my PhD experience that you can apply

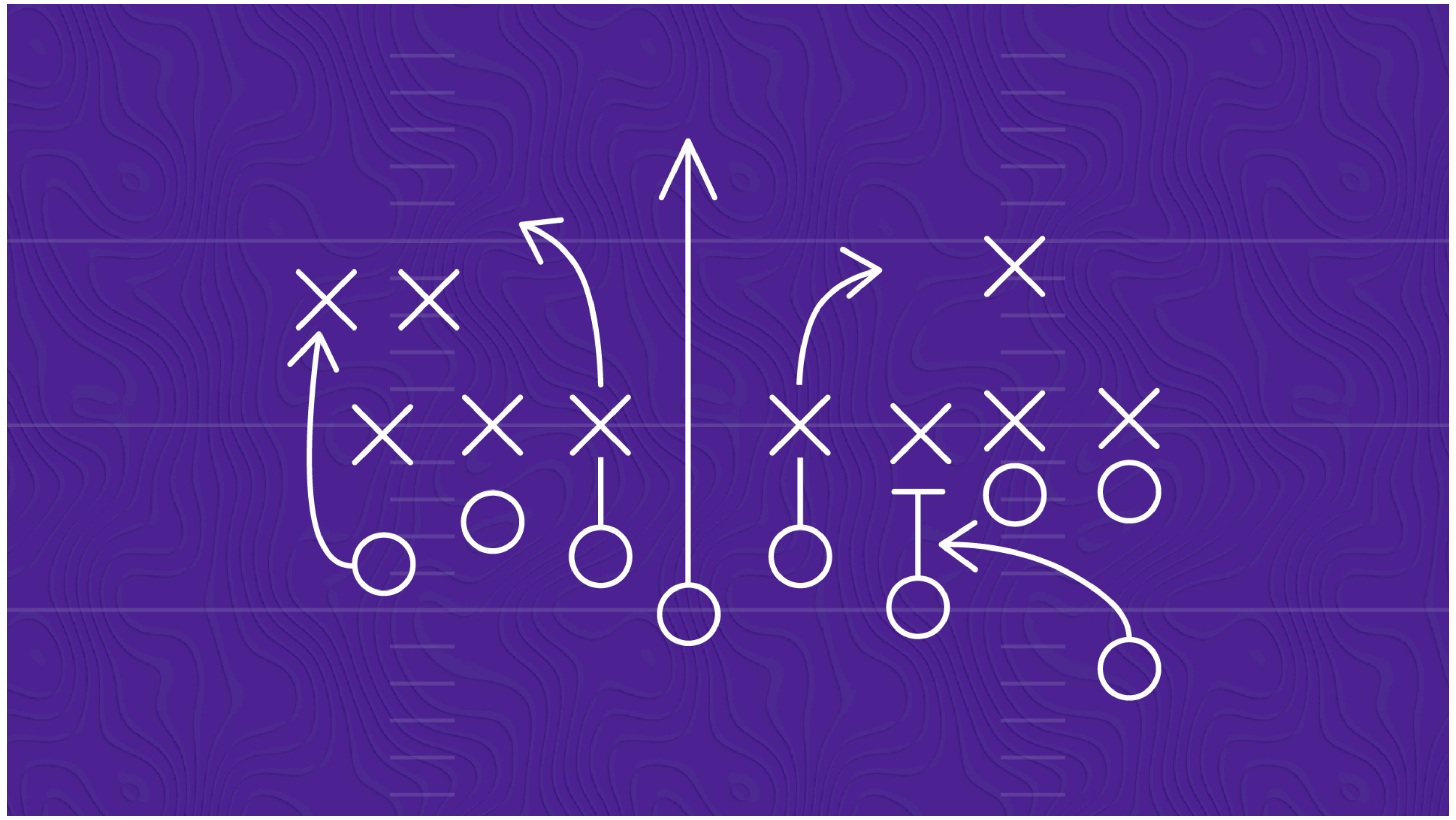
Mandela Patrick, 05/10/2021

# A little about me

- Born and grew-up in Trinidad and Tobago
- B.A honours Computer Science from Harvard College in 2018
- Won a Rhodes Scholarship to pursue PhD at Oxford
- Graduated in August with a PhD from the VGG group
- Research interests: multi-modal + self-supervised learning
- Currently, Machine Learning scientist at Piñata Farms.

# The PhD Playbook

- The set of strategies that made me successful during my PhD and can be helpful to others.



# The PhD Playbook

The playbook has been divided into the following sections:

1. ***"It takes a village"***: the importance of the people in your PhD journey
2. ***"The Paper Checklist"***: tips for a competitive submission
3. ***"What's next?"***: tips on what to do next after completing PhD

# It Takes a Village

## The importance of the people in your PhD journey

# What’s the most important section of my thesis?

# Acknowledgements

## Acknowledgements

To my co-supervisors, Andrea Vedaldi and João Henriques, thank you for your constant support, patience and guidance throughout this DPhil. Thank you for always being very generous with your knowledge and time, and giving me the space to tackle research problems that I am excited and passionate about. This thesis would not be possible without you two.

**Supervisors + Postdocs**

To my amazing collaborators, Yuki Asano, Ruth Fong, and Pa-Yao Huang, you have made each and every one of my research projects enjoyable, and intellectually enriching. I have learnt so much from each of you, and I would not be the researcher I am today if it was not for the countless hours we spent ideating, coding and brainstorming. Collaborators like you make a DPhil worth it.

**Collaborators**

More broadly, to my VGG family, I am so thankful to you for welcoming me as part of the lab, and always being a constant source of support. Andrew Brown, Lili Momeni, Daffy Afouras, Max Bain, Tendga Han, Christian Ruppreceht, and Dylan Campbell, I am gonna miss you all and cannot wait to see how you all impact this field in fundamental ways.

This DPhil would not be possible without the financial support of the Rhodes Trust. Winning the Rhodes Scholarship has been one of my biggest blessings, and being part of the Rhodes community has made my Oxford experience really special.

I also want to thank EPSRC AIMS CDT for funding my research, and in particular, Wendy Poole for taking care of every administrative task related to my DPhil, thus allowing me to focus on my research. Wendy, you are simply the best.

Thank you to Facebook for supporting my DPhil research. I would not have been able to push the limits of large-scale multi-modal self-supervision without access to the data, computing resources and amazing collaborators, Ishan Misra, Geoffrey Zweig, Florian Metze, Christoph Feichtenhofer, Polina Kuznetsova, Rose Kanjirathinkal and Dong Guo.

**Mentors + Sponsors**

To my friends in Oxford and London, Stephanie Ifayemi, Madeleine Chang, Samuel Liu, Alexander Thomas, Shaan Desai, Terrens Muradzikwa, Jelani Munroe, and Michael Chen, thank you for always supporting and pushing me throughout this journey.

**Friends**

You do not know how much it means to me to have you all in my life, and I have cherished every moment with you all.

Lastly, and most importantly, thank you to my family (Raymond, Hyacinth, and Nku) for always believing in me. You have been there for me during the highs and lows of this DPhil, and I can always count on you all to pick me up when I am down. This DPhil is for you.

**Family**

# Supervisors

- Talk to past and current students
  - Supervisory style: hands-on vs hands-off? Long-term vs short-term?
  - Research interests: does supervisor's interests overlap with your research passion?

Andrea Vedaldi

João Henriques

# Collaborators

- Collaborate, not compete: bounce ideas, pair program
- Share and rotate first-authorship
- Let everyone play to their strengths

**Language Models are Few-Shot Learners**

Tom B. Brown* Benjamin Mann* Nick Ryder* Melanie Subbiah*

Jared Kaplan† Prafulla Dhariwal Arvind Neelakantan Pranav Shyam Girish Sastry

Amanda Askell Sandhini Agarwal Ariel Herbert-Voss Gretchen Krueger Tom Henighan

Rewon Child Aditya Ramesh Daniel M. Ziegler Jeffrey Wu Clemens Winter

Christopher Hesse Mark Chen Eric Sigler Mateusz Litwin Scott Gray

Benjamin Chess Jack Clark Christopher Berner

Sam McCandlish Alec Radford Ilya Sutskever Dario Amodei

Yuki Asano

Bernie Huang

Ruth Fong

# Mentors

- You don't need to have to have all the answers
- "Your Personal Board of Directors": those who you go to for advice
  - Deciding between internship opportunities, research directions, post-grad

Ishan Misra

Maxine Williams

# Sponsors

- Sponsors: those in senior positions who **advocate** for you
- Try to establish such relationship during internships

Florian Metze

Geoffrey Zweig

# Friends + Family

- PhD is long and difficult journey, and family and friends play critical role in getting you through
  - Support you during the tough times, and celebrate the good times

Friends

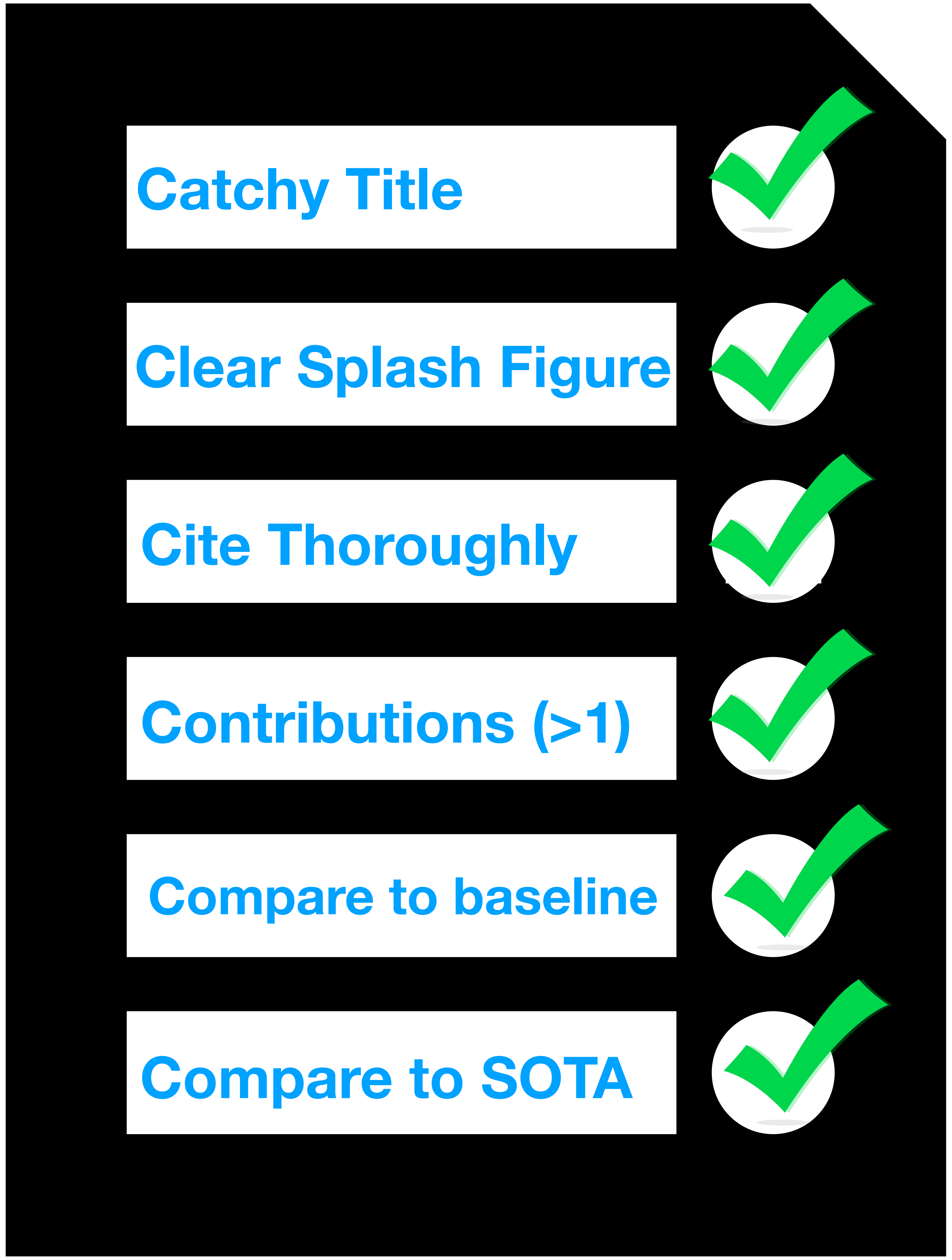

Family


IT TAKES A VILLAGE

# Paper checklist

## Tips for a competitive submission

# Checklist Overview (The 6 C's)



# Catchy Title

- Gives an idea about topic, but leaves the reader wanting to learn more

**Labelling unlabelled videos**
**from scratch with multi-modal self-supervision**

**Keeping Your Eye on the Ball:**
**Trajectory Attention in Video Transformers**

# Splash Figure

- Captures the method and/or the intuition of the approach very clearly

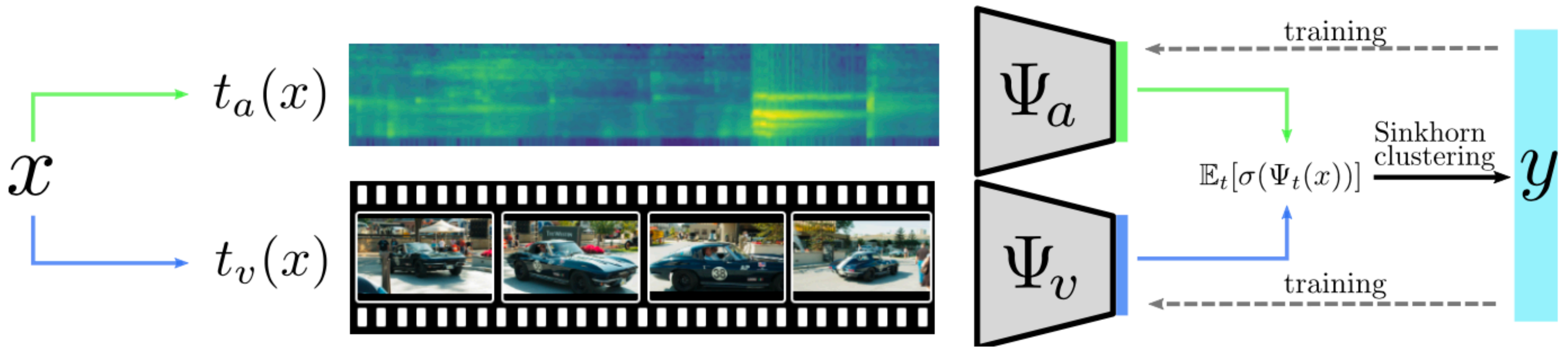
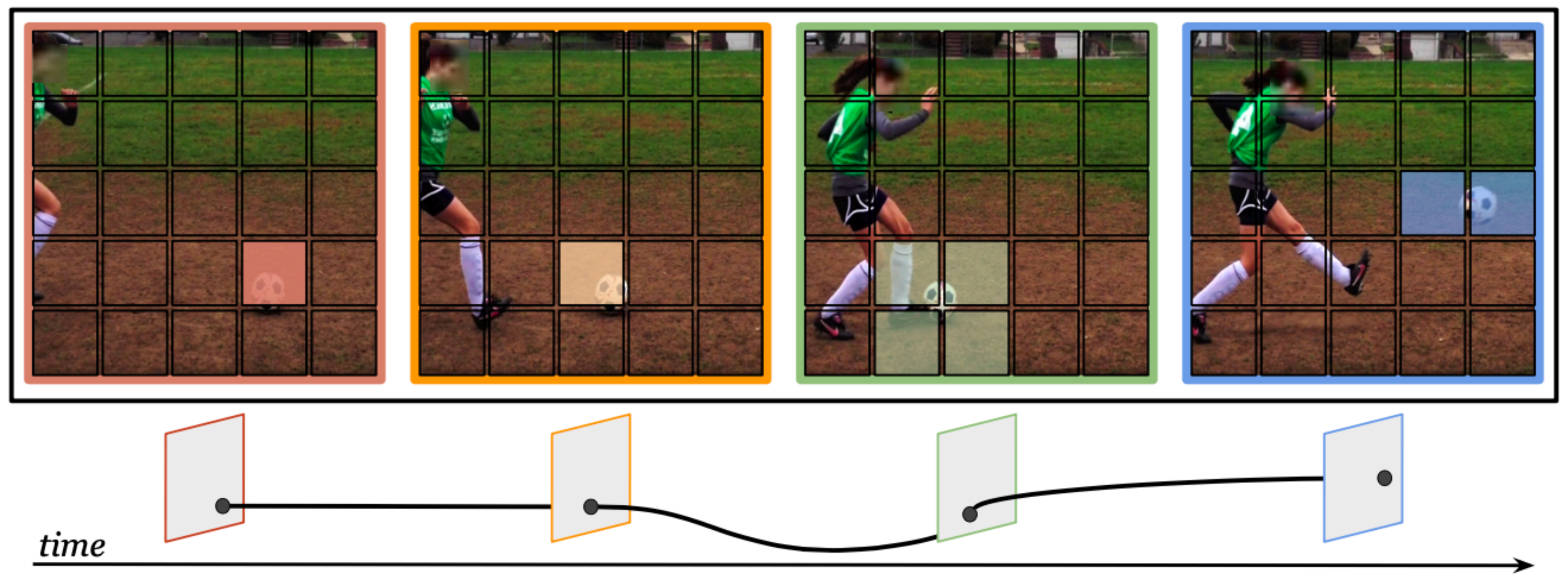
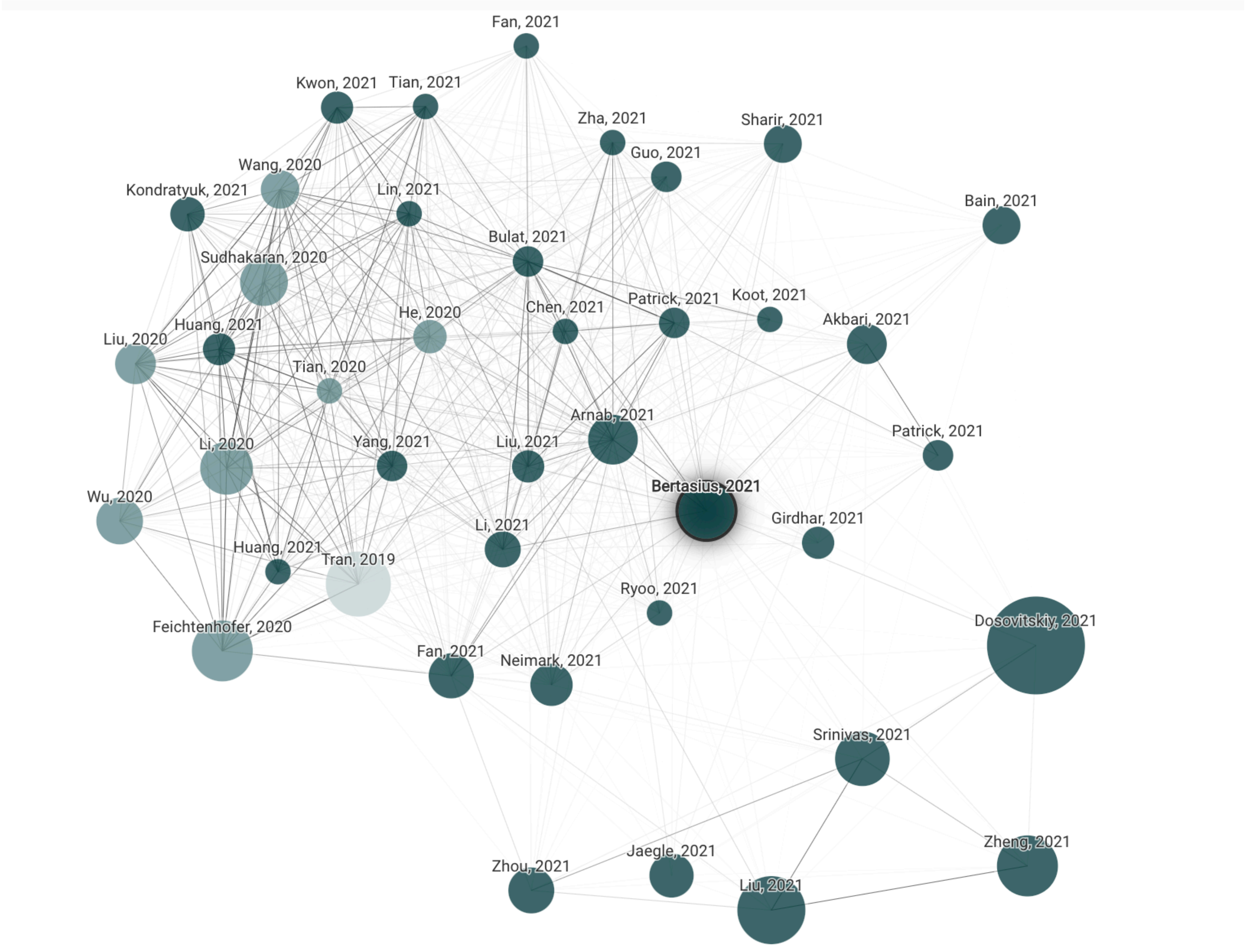



Figure 1: **Our model** views modalities as different *augmentations* and produces a multi-modal clustering of video datasets from scratch that can closely match human annotated labels.



# Related Works

- Be very thorough; cite as much relevant works as possible.
  - “Are you going to get as much citations on this work?” - supervisor
- Use websites such as ConnectedPapers, Semantic Scholar, PapersWithCode



# Method

- Clear, and simple language and writing (very easy-to-follow)
- More than 1 technical novel contribution (3 is ideal)

**3 Method**

**3.1 Non-degenerate clustering via optimal transport**

**3.2 Clustering with arbitrary prior distributions**

**3.3 Multi-modal single labelling**

**3 Trajectory Attention for Video Data**

**3.1 Video self-attention**

**3.2 Approximating attention**

# Results: Extensive Ablations

- Extensive ablations demonstrates the impact of your contribution clearly
- Anticipate any ablation requests from reviewers and add to main paper or appendix.
- Important: to define baseline.

Table 3: **Ablation** of multi-modality, Modality Alignment and Gaussian marginals. Decorrelated Heads. Models are evaluated at 75 epochs on the VGG-Sound dataset.

| Method | 🖼 | (🔊 📹) | MA? | G.? | DH? | **Acc** | **ARI** | **NMI** |
|---|---|---|---|---|---|---|---|---|
| (a) SeLa | ✓ | ✗ | – | – | – | 6.4 | 2.3 | 20.6 |
| (b) Concat | ✗ | ✓ | – | ✗ | ✗ | 7.6 | 3.2 | 24.7 |
| (c) SeLaVi | ✗ | ✓ | ✗ | ✗ | ✗ | 24.6 | 15.6 | 48.8 |
| (d) SeLaVi | ✗ | ✓ | ✗ | ✓ | ✓ | 26.6 | 18.5 | 50.9 |
| (e) SeLaVi | ✗ | ✓ | ✓ | ✗ | ✓ | 26.2 | 17.3 | 51.5 |
| (f) SeLaVi | ✗ | ✓ | ✓ | ✓ | ✗ | 23.9 | 14.7 | 49.9 |
| (g) **SeLaVi** | ✗ | ✓ | ✓ | ✓ | ✓ | 26.6 | 17.7 | 51.1 |

Table 4: **Attention ablations:** We compare trajectory attention with alternatives and ablate its design choices. We report GFLOPS and top-1 accuracy (%) on K-400 and SSv2. $\mathrm{Att}_T$: temporal attention, $\mathrm{Avg}_T$: temporal averaging, $\mathrm{Norm}_{ST}$: space-time normalization, $\mathrm{Norm}_S$: spatial normalization.

| Attention | $\mathrm{Att}_T$ | $\mathrm{Avg}_T$ | $\mathrm{Norm}_S$ | $\mathrm{Norm}_{ST}$ | GFLOPS | K-400 | SSv2 |
|---|---|---|---|---|---|---|---|
| Joint Space-Time | – | – | – | – | 180.6 | 79.2 | 64.0 |
| Divided Space-Time | – | – | – | – | 185.8 | 78.5 | 64.2 |
| | ✗ | ✓ | ✓ | ✗ | 180.6 | 76.0 | 60.0 |
| | ✓ | ✗ | ✗ | ✓ | 369.5 | 77.2 | 60.9 |
| Trajectory | ✓ | ✗ | ✓ | ✗ | 369.5 | **79.7** | **66.5** |

# Results: Comparison to State-of-Art

- Showing competitive performance compared to current state-of-the-art always helps your paper.
- Show comparisons across a number of datasets: 3 - 4 is ideal.
- Structure table to show other dimensions (FLOPs, memory, speed) that your approach excels in.

(a) **VGG-Sound.**

| Method | NMI | ARI | Acc. | ⟨H⟩ | ⟨$\mathbf{p}_{\max}$⟩ |
|---|---|---|---|---|---|
| Random | 10.2 | 4.0 | 2.2 | 4.9 | 3.5 |
| Supervised | 46.5 | 15.6 | 24.3 | 2.9 | 30.8 |
| DPC | 15.4 | 0.7 | 3.2 | 4.7 | 4.9 |
| XDC | 18.1 | 1.2 | 4.5 | 4.41 | 7.4 |
| MIL-NCE | 48.5 | 12.5 | 22.0 | 2.6 | 32.9 |
| **SeLaVi** | **55.9** | **21.6** | **31.0** | **2.5** | **36.3** |

(b) **AVE.**

| Method | NMI | ARI | Acc. | ⟨H⟩ | ⟨$\mathbf{p}_{\max}$⟩ |
|---|---|---|---|---|---|
| Random | 9.2 | 1.3 | 9.3 | 2.9 | 12.6 |
| Supervised | 58.4 | 34.8 | 50.5 | 1.1 | 60.6 |
| DPC | 18.4 | 5.0 | 15.1 | 2.7 | 17.5 |
| XDC | 17.1 | 6.0 | 16.4 | 2.6 | 19.1 |
| MIL-NCE | 56.3 | 30.3 | 42.6 | **1.2** | 57.1 |
| **SeLaVi** | **66.2** | **47.4** | **57.9** | **1.1** | **59.3** |

(c) **Kinetics.**

| Method | NMI | ARI | Acc. | ⟨H⟩ | ⟨$\mathbf{p}_{\max}$⟩ |
|---|---|---|---|---|---|
| Random | 11.1 | 0.2 | 1.8 | 5.1 | 3.3 |
| Supervised | 70.5 | 43.4 | 54.9 | 1.6 | 62.2 |
| DPC | 16.1 | 0.6 | 2.7 | 4.9 | 3.9 |
| XDC | 17.2 | 0.8 | 3.4 | 4.7 | 6.2 |
| MIL-NCE | **48.9** | **12.5** | **23.5** | **2.7** | **33.7** |
| **SeLaVi** | 27.1 | 3.4 | 7.8 | 4.8 | 9.4 |

(d) **Kinetics-Sound.**

| Method | NMI | ARI | Acc. | ⟨H⟩ | ⟨$\mathbf{p}_{\max}$⟩ |
|---|---|---|---|---|---|
| Random | 2.8 | 0.5 | 5.9 | 3.3 | 8.3 |
| Supervised | 81.7 | 66.3 | 75.0 | 0.5 | 85.4 |
| DPC | 8.8 | 2.2 | 9.6 | 3.1 | 13.6 |
| XDC | 7.5 | 1.9 | 9.4 | 3.1 | 13.6 |
| MIL-NCE | **47.5** | 24.0 | 37.8 | **1.5** | **51.0** |
| **SeLaVi** | **47.5** | **28.7** | **41.2** | 1.8 | 45.5 |

(a) **Something–Something V2**

| Model | Pretrain | Top-1 | Top-5 | GFLOPs ×views |
|---|---|---|---|---|
| SlowFast [25] | K-400 | 61.7 | - | 65.7×3×1 |
| TSM [46] | K-400 | 63.4 | 88.5 | 62.4×3×2 |
| STM [33] | IN-1K | 64.2 | 89.8 | 66.5×3×10 |
| MSNet [40] | IN-1K | 64.7 | 89.4 | 67×1×1 |
| TEA [45] | IN-1K | 65.1 | - | 70×3×10 |
| bLVNet [23] | IN-1K | 65.2 | 90.3 | 128.6×3×10 |
| VidTr-L [44] | IN-21K+K-400 | 60.2 | - | 351×3×10 |
| Tformer-L [7] | IN-21K | 62.5 | - | 1703×3×1 |
| ViViT-L [2] | IN-21K+K-400 | 65.4 | 89.8 | 3992×4×3 |
| MViT-B [22] | K-400 | 67.1 | 90.8 | 170×3×1 |
| **Mformer** | IN-21K+K-400 | 66.5 | 90.1 | 369.5×3×1 |
| **Mformer-L** | IN-21K+K-400 | **68.1** | **91.2** | 1185.1×3×1 |
| **Mformer-HR** | IN-21K+K-400 | 67.1 | 90.6 | 958.8×3×1 |

(b) **Kinetics-400**

| Method | Pretrain | Top-1 | Top-5 | GFLOPs×views |
|---|---|---|---|---|
| I3D [10] | IN-1K | 72.1 | 89.3 | 108×N/A |
| R(2+1)D [75] | - | 72.0 | 90.0 | 152×5×23 |
| S3D-G [84] | IN-1K | 74.7 | 93.4 | 142.8×N/A |
| X3D-XL [24] | - | 79.1 | 93.9 | 48.4×3×10 |
| SlowFast [25] | - | 79.8 | 93.9 | 234×3×10 |
| VTN [51] | IN-21K | 78.6 | 93.7 | 4218×1×1 |
| VidTr-L [44] | IN-21K | 79.1 | 93.9 | 392×3×10 |
| Tformer-L[7] | IN-21K | 80.7 | 94.7 | 2380×3×1 |
| MViT-B [22] | - | 81.2 | 95.1 | 455×3×3 |
| ViViT-L [2] | IN-21K | **81.3** | 94.7 | 3992×3×4 |
| **Mformer** | IN-21K | 79.7 | 94.2 | 369.5×3×10 |
| **Mformer-L** | IN-21K | 80.2 | 94.8 | 1185.1×3×10 |
| **Mformer-HR** | IN-21K | 81.1 | **95.2** | 958.8×3×10 |

(c) **Epic-Kitchens**

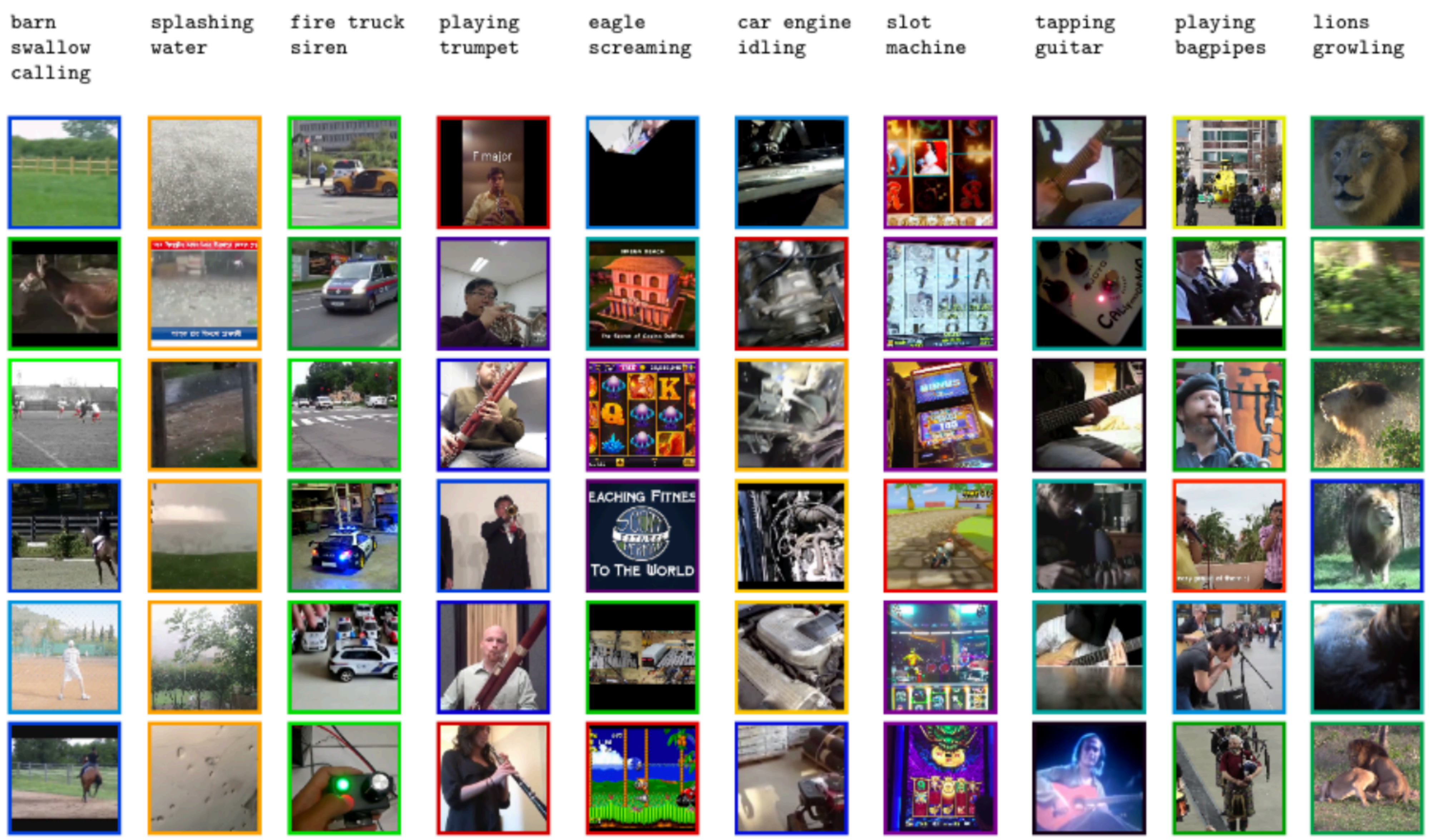
| Method | Pretrain | A | V | N |
|---|---|---|---|---|
| TSN [78] | IN-1K | 33.2 | 60.2 | 46.0 |
| TRN [86] | IN-1K | 35.3 | 65.9 | 45.4 |
| TBN [36] | IN-1K | 36.7 | 66.0 | 47.2 |
| TSM [46] | IN-1K | 38.3 | **67.9** | 49.0 |
| SlowFast [25] | K-400 | 38.5 | 65.6 | 50.0 |
| ViViT-L [2] | IN-21K+K-400 | 44.0 | 66.4 | 56.8 |
| **Mformer** | IN-21K+K-400 | 43.1 | 66.7 | 56.5 |
| **Mformer-L** | IN-21K+K-400 | 44.1 | 67.1 | 57.6 |
| **Mformer-HR** | IN-21K+K-400 | **44.5** | 67.0 | **58.5** |

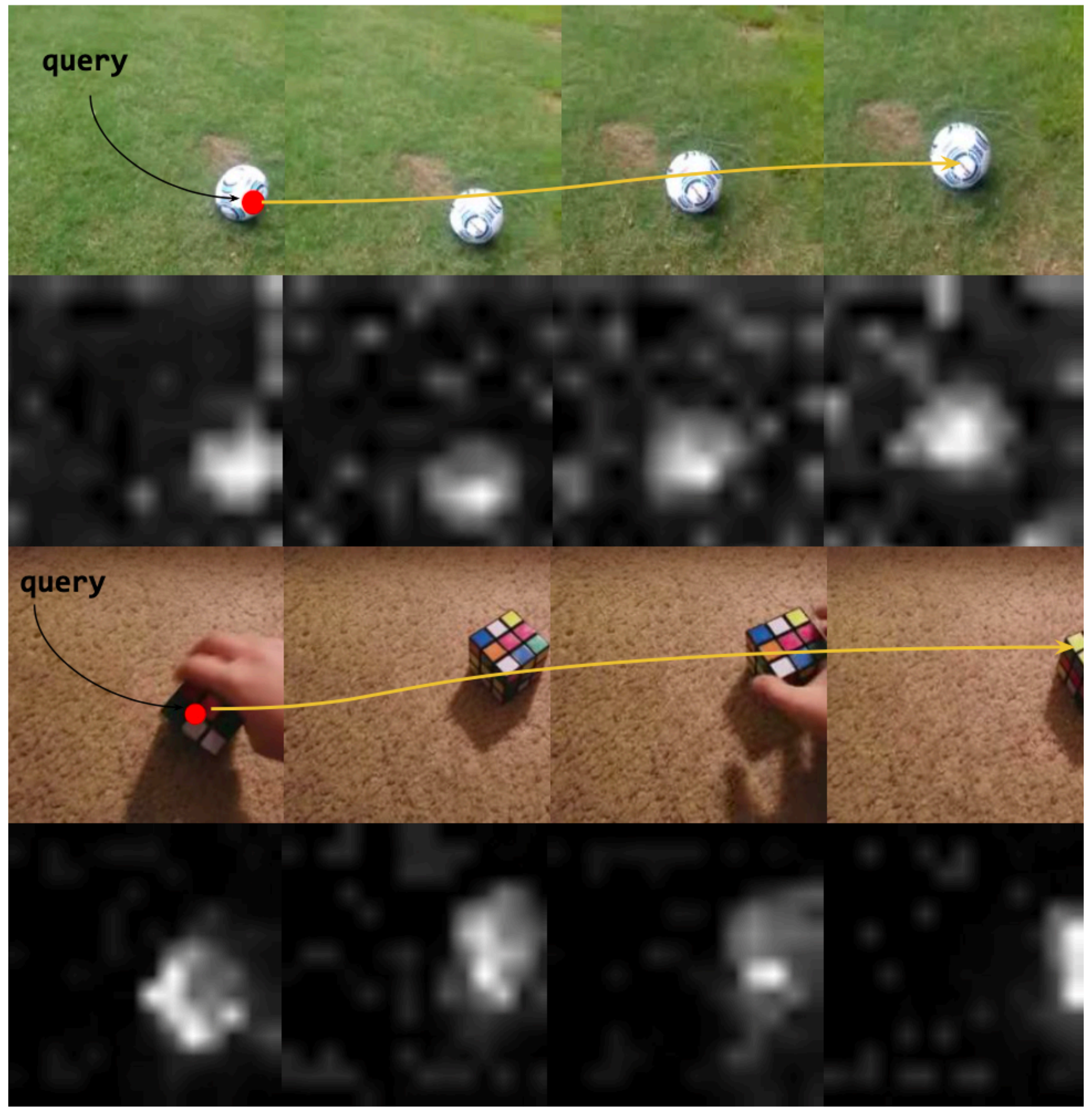
(d) **Kinetics-600**

| Model | Pretrain | Top-1 | Top-5 | GFLOPs ×views |
|---|---|---|---|---|
| AttnNAS [81] | - | 79.8 | 94.4 | - |
| LGD-3D [56] | IN-1K | 81.5 | 95.6 | - |
| SlowFast [25] | - | 81.8 | 95.1 | 234×3×10 |
| X3D-XL [24] | - | 81.9 | 95.5 | 48.4×3×10 |
| Tformer-HR [7] | IN-21K | 82.4 | 96.0 | 1703×3×1 |
| ViViT-L [2] | IN-21K | 83.0 | 95.7 | 3992×3×4 |
| MViT-B-24 [22] | - | **83.8** | **96.3** | 236×1×5 |
| **Mformer** | IN-21K | 81.6 | 95.6 | 369.5×3×10 |
| **Mformer-L** | IN-21K | 82.2 | 96.0 | 1185.1×3×10 |
| **Mformer-HR** | IN-21K | 82.7 | 96.1 | 958.8×3×10 |

# Results: Qualitative Figures

- Qualitative Figures complements your quantitative results by visually showing what your model is doing.



# Practical Tip 1: Choose venue wisely

- Every conference is different, and they each value different things
  - Theory vs applied? e.g. ICML vs. WACV
  - Preference for pushing state-of-the-art e.g. CVPR
  - Domain-specific vs domain-agnostic e.g. NeurIPS vs ICASSP

NEURAL INFORMATION PROCESSING SYSTEMS

ICML
International Conference On Machine Learning

# Practical Tip 2: Maintain experiment log

- Be very meticulous on maintaining experiment log
  - Very helpful rebuttals to find any requested experiments
  - Detect patterns in hyper-parameters for SOTA.
  - Reproducibility
- Spreadsheets or open-source tools (Mlflow, Neptune) are helpful for this.

| | SLURM ID | EXP DESC | ACC | MODEL | INIT | PATCH (P x P x T) | INPUT-SIZE | FRAMES | BATCH-SIZE | Attention Layer |
|---|---|---|---|---|---|---|---|---|---|---|
| K-400 | | | | | | | | | | |
| python3 run_with_submitit.py --num_shards 8 --partition priority --comment iccv-2021 --cfg configs/ICCV21/K_400/jointspacetimeformer_rgb_8x8.yaml --use_volta32 --job_dir /checkpoint/mandelapatrick/slowfast_k400_abl | 40426115 | | 78.90% | ViT-B (L=12, NH=12, d=3072) | IM-21K, ViT-B, 224x224 | 16 x 16 x 2 | 224 x 224 | 16 x 4 | 32 / NODE | Joint Space-Time |
| python3 run_with_submitit.py --num_shards 8 --partition priority --comment iccv-2021 --cfg configs/SOTA/K400/jointspacetimeformer_rgb_224_16x4_3D.yaml --use_volta32 --job_dir /checkpoint/mandelapatrick/neurips_sota | 40492435 | | 79.67% | ViT-B (L=12, NH=12, d=3072) | IM-21K, ViT-B, 224x224 | 16 x 16 x 2 | 224 x 224 | 16 x 4 | 32 / NODE | Joint Space-Time |
| python3 run_with_submitit.py --num_shards 8 --partition priority --comment iccv-2021 --cfg configs/SOTA/K400/timesformer_rgb_224_16x4_3D.yaml --use_volta32 --job_dir /checkpoint/mandelapatrick/neurips_sota | 40437031 | | 79.01% | ViT-B (L=12, NH=12, d=3072) | IM-21K, ViT-B, 224x224 | 16 x 16 x 2 | 224 x 224 | 16 x 4 | 32 / NODE | Divided Space-Time |
| python3 run_with_submitit.py --num_shards 8 --partition priority --comment iccv-2021 --cfg configs/SOTA/K400/spacetimeattendformer_rgb_224_16x4_3D.yaml --use_volta32 --job_dir /checkpoint/mandelapatrick/neurips_sota | 40437950 | RRC, no CJ, no RA | 79.79% | ViT-B (L=12, NH=12, d=3072) | IM-21K, ViT-B, 224x224 | 16 x 16 x 2 | 224 x 224 | 16 x 4 | 32 / NODE | Space-Time Motion |

# Practical Tip 3: Open-Source Early

- Open-sourcing code with pertained models soon after conference deadline:
  - Adds visibility / publicity to your work as others can easily build on it
  - Reproducibility of results by the community.

# What’s next?

## Tips on deciding on what’s next after wrapping up PhD

# What's next post-PhD?

## A professor, research scientist, and ML engineer walk into a bar

# The Post-PhD Job Matrix (At Graduation)

| | Prestige | Financial | Academic Freedom | Bureaucracy | Stability |
|---|---|---|---|---|---|
| **Industry Lab (FB, Google, DM)** | Medium | High | Medium | High | High |
| **Academic (Tenure Track)** | High | Low | High | High | High |
| **Startup (Seed / Series-A)** | Low | Medium | Low | Low | Low |

# Your preferences impacts the function

- The weights of this function depends on your preferences and circumstances.
- These weights may be positive or negative :)

$$\boxed{f = w_{prestige}prestige + w_{financial}financial + w_{people}people + w_{academicfreedom}academicfreedom + w_{bureaucracy}bureaucracy + w_{stability}stability}$$

# Your preferences vary with time

- As you get older, **what you value** changes.
    - e.g. One may value stability later on life, but not when younger

$$\boxed{f\boxed{(t)} = w_{prestige}\boxed{(t)}prestige + w_{financial}\boxed{(t)}financial + w_{academicfreedom}\boxed{(t)}academicfreedom + w_{bureaucracy}\boxed{(t)}bureaucracy + w_{stability}\boxed{(t)}stability}$$

# The variables vary with time

- The **variables of the function** usually **change value** over time.
  - e.g. salary, stability

$$f(t) = w_{prestige}(t)prestige(t) + w_{financial}(t)financial(t) + w_{academicfreedom}(t)academicfreedom(t) + w_{bureaucracy}(t)bureaucracy(t) + w_{stability}(t)stability(t)$$

## For industrial + academic path, the change of variables is known

- **How variables change** are a lot **more predictable** for academic and industrial jobs.

- **Salaries**:
  - University professor: publicly available online
  - Industrial jobs: websites are available e.g. Glassdoor, Levels.fyi

# For startups, there's a lot more unknowns

- As there is greater **information asymmetry and uncertainty** with startups, the **value of these variables can vary a lot** and is very startup-dependent.

- What are the questions to answer to **get the right information** to reduce this uncertainty when deciding on a startup?

# Joining a startup

- Does the mission excite you?
- Stage of startup?
- Do you like the people?
- What’s your role at the startup and how do you see it changing over time?
- What are your financial goals?
- Are you okay with doing more applied work?
- Who are the investors?
- What’s your risk appetite?

# In summary

- **Build the right village** to make you successful during PhD
- **Follow the checklist (6 C’s)** to have a competitive paper submission
- **Only you can decide** what you want to do after your PhD :)